



University of Botswana
Faculty of Engineering and Technology
Department of Civil Engineering

A COMPARISON OF USE OF L- MOMENT AND LH-
MOMENT METHODS OF PARAMETER ESTIMATION FOR THE
REGIONAL FLOOD FREQUENCY ANALYSIS IN SEMI-ARID
AREAS: THE CASE OF LIMPOPO CATCHMENT, BOTSWANA

By

Solomon Shiferaw Yirba

A Dissertation submitted to the University of Botswana in partial fulfilment of the
Requirement for degree of

MASTER OF SCIENCE

Supervisor

Professor B. P. Parida

2016

APPROVAL

This research has been examined and is approved as meeting the required standards of scholarship for partial fulfilment of the requirements for the degree Master of Science (Civil Engineering).

Supervisor: _____

Internal Examiner: _____

External Examiner: _____

Dean, School of Graduate Studies: _____

STATEMENT OF ORIGINALITY

The work contained in this Dissertation was carried out by the author while a student at the University of Botswana between 2009 and 2016. It is original work except where due reference is made and neither has nor will be submitted for the award of a degree at any other university.

(Signature of the student)

Solomon Shiferaw Yirba

ID: 200908175

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to the University of Botswana for giving me the opportunity to study in the University. I would like to thank the Faculty of Engineering and Technology, specially the Department of Civil Engineering for allowing me to follow Water Resources and Environmental Engineering.

I am indebted to my supervisor, Professor B.P. Parida and Dr. P.T. Odirile, program coordinator, for their inspiration, guidance, critical analysis and substantial comments on the approach throughout the research period. Special thanks go to J.R.M. Hosking, IBM Research Division, USA, and Professor B.F. Alemaw for their guidance and assistance in completion of this Dissertation.

Thanks to my classmates in Water Resources and Environmental Engineering who have given me the joy of being a student again. Those classroom and practical days have created a friendly atmosphere one can never forget.

My sincere thanks go to the Department of Water Affairs and Department of Meteorological Services, Gaborone, for providing the rivers flow and rainfall data that were utilized for this research.

At last, I would like to thank all my family members for their constant support, encouragement and patience during my study.

ABSTRACT

Design flood estimation is an important task that is required in the planning and design of many hydraulic structures. This study was therefore aimed at developing the regional growth curves for estimation of floods of various return periods in the Limpopo catchment at either gauged or ungauged sites. Finding the most suitable distribution to flood sample and selecting the appropriate parameter estimation method are of great importance for flood frequency analysis. In this study, the newer methods of the L-moments and LH-moments using the Regional Flood Frequency Analysis (RFFA) technique have been used to characterise the flood data of this region. The study focussed on the RFFA of the Limpopo catchment, Botswana, which comprises of 13 hydrometric stations. The heterogeneity test has revealed that the Limpopo catchment using LH-moments has been identified as “acceptably homogeneous” through all levels $L_{\eta}, \eta = 1, 2, 3, 4$, that is, (L_1 to L_4); and therefore, this method was found the best technique to characterize the Limpopo flood data.

Three extreme value distributions, that is, generalized extreme value (GEV), generalized logistic (GLO) and generalized Pareto (GPA), through different levels of the LH-moments (L to L_4) have been applied to develop the regional parameters and describe the annual maximum flood data obtained from 13 sites in the Limpopo catchment. The Z-statistic criteria were used in the distribution selection, considering the respective LH-moments (L to L_4) and as such the GEV distribution using LH-moments at level 2 (L_2) has been found the best distribution of all other distributions. For final selection of the appropriate method of parameter estimates, the performances of GEV distribution using L-moments and LH-moments (L_2) have been assessed by evaluating the relative Root Mean Square Error (RMSE). The results of the RMSE showed that the LH-moments of level 2 (L_2) based on GEV has been found most suitable and more efficient with minimum RMSE for obtaining improved values of flood peaks than the L-moments.

Regional flood frequency relationships are developed for estimation of floods of various return periods for ungauged sites using the LH-moment (L_2) based on the GEV distribution. A general relationship between mean annual peak flood, drainage area and rainfall has been developed.

Keywords: L-moments; LH-moments; regional flood frequency analysis, Limpopo catchment, regional growth curve

TABLE OF CONTENTS

APPROVAL	ii
STATEMENT OF ORIGINALITY	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF FIGURES.....	ix
LIST OF TABLES	x
GLOSSARY OF ACRONYMS AND ABBREVIATIONS	xi
CHAPTER 1	1
1. INTRODUCTION	1
1.1 General	1
1.2 Background.....	2
1.3 Description of the Study Area	4
1.4 Statement of the Problem	7
1.5 Objective of the Research and Corresponding Research Questions	7
1.6 Significance of the Study.....	9
1.7 Scope of the Study	9
CHAPTER 2	11
2 LITERATURE REVIEW	11
2.1 Related Previous Studies	11
2.2 Probability Concepts and Theoretical Background	11
2.2.1 Random Variables and Distributions	11
2.2.2 Quantiles and Return Period.....	12
2.3 Choice of Frequency Distribution.....	13
2.4 Methods of Parameter Estimation.....	15
2.4.1 Historical Development of Different Methods	15
2.4.2 Conventional Methods of Parameter Estimation	18
2.4.3 Newer Methods of Parameter Estimation	22
2.4.3.1 L-Moments	22
2.4.3.2 LH-Moments.....	23
CHAPTER 3	25
3 METHODOLOGY	25
3.1 General	25
3.2 L-Moments	25

3.3	LH-Moments.....	27
3.4	Parameters Estimation.....	29
3.4.1	LH- and L-moments for the GEV distribution	29
3.4.1.1	L-Moments	30
3.4.1.2	LH-Moments.....	30
3.4.2	LH- and L-moments for the GLO distribution	31
3.4.2.1	L-Moments	32
3.4.2.2	LH-Moments.....	32
3.4.3	LH- and L-moments for the GPA distribution	33
3.4.3.1	L-Moments	33
3.4.3.2	LH-Moments.....	34
3.5	Steps Used for Regional Flood Frequency Analysis	34
3.5.1	Discordancy Measure.....	34
3.5.2	Regional Homogeneity (Heterogeneity Measure)	35
3.5.3	Goodness-Of-Fit Measure	38
3.5.4	Flood Quantile Estimates for the Region	39
3.6	Comparative Study of the Methods.....	43
CHAPTER 4		45
4	DATA ANALYSIS AND DISCUSSION OF RESULTS	45
4.1	Data Used	45
4.2	Analysis with L-Moments	45
4.2.1	L-moments Statistics.....	45
4.2.2	Discordancy Measure.....	47
4.2.3	Regional Homogeneity.....	48
4.2.4	Choice of Suitable Distribution	49
4.2.4.1	L-Moments Ratio Diagram.....	49
4.2.4.2	Z-Statistics Criteria	50
4.2.5	Parameters Estimation.....	51
4.2.6	Flood Quantiles Estimation and Development of Regional Growth Curves	51
4.2.7	Comparison of Observed and Estimated Floods.....	52
4.2.8	Summary.....	61
4.3	Analysis with LH-Moments	62
4.3.1	LH-Moments Statistics.....	62
4.3.2	Discordancy Measure.....	64
4.3.3	Regional Homogeneity.....	68
4.3.4	Choice of Suitable Distribution	69
4.3.4.1	LH-Moments Ratio Diagrams	69
4.3.4.2	Z-Statistics Criteria	71
4.3.5	Parameters Estimation.....	73
4.3.6	Flood Quantiles Estimation and Development of Regional Growth Curves	73
4.3.7	Comparison of Observed and Estimated Floods.....	74
4.3.8	Summary.....	83
4.4	Discussion of Results	84
4.5	Flood Estimation for Ungauged Basins	86

4.5.1	Procedures	86
4.5.2	Flood Estimation at Gauged Site Using Equation (102)	88
4.6	Reference to Previous Studies	89
CHAPTER 5	91
5	CONCLUSIONS AND RECOMMENDATIONS.....	91
5.1	Conclusions.....	91
5.2	Recommendations.....	92
REFERENCES	93

LIST OF FIGURES

Figure 1.1: Locality map of the study area.....	2
Figure 1.2: Limpopo Catchment and location of its stations.....	6
Figure 3.1: Theoretical plots of L-skew versus L-kurtosis diagram for statistical distributions (viz: Generalised Pareto (GPA), Generalised Extreme Values (GEV), Generalised Logistic (GLO), 3 Parameter Log-normal (LN3), Pearson Type 3 (PE3) Distribution) (Hosking and Wallis, 1997)	27
Figure 4.1: Plots of $L-C_v$ versus L-Skewness.....	48
Figure 4.2: L-moments ratio diagram.....	50
Figure 4.3: Regional growth curve using L-moments.....	52
Figure 4.4: Plot of observed and estimated floods at gauging stations No. 2411 and 4411, respectively	55
Figure 4.5: Plot of observed and estimated floods at gauging stations No. as shown	60
Figure 4.6: Plots of L_1-C_v , L_2-C_v , L_3-C_v and L_4-C_v versus L_1 -Skewness, L_2 -Skewness, L_3 -Skewness and L_4 -Skewness	67
Figure 4.7: LH-moments (L_1 to L_4) ratio diagrams.....	71
Figure 4.8: Regional growth curve for Limpopo region using LH-moment (L_2).....	74
Figure 4.9: Plot of observed and estimated floods at gauging stations No. 2411 and 4411, respectively	77
Figure 4.10: Plot of observed and estimated floods at gauging stations No. as shown	83
Figure 4.11: Regional growth curves using LH-moments (L_2) and L-moments	85
Figure 4.12: Regional growth curves by Parida (2004) and this study	90

LIST OF TABLES

Table 1.1: Specific Objectives and Research Questions	8
Table 3.1: Corresponding Coefficients of Equation (54) for different levels (η) of the LH-moment (Lee and Maeng 2003).....	31
Table 4.1: The sites of the study area and their corresponding information	45
Table 4.2: Values of Probability Weighted Moments using L-moments	46
Table 4.3: Sample L-moments and L-moment ratios.....	46
Table 4.4: L-moment ratios and discordancy measures	47
Table 4.5: Heterogeneity measure for L-moments	49
Table 4.6: Z -Statistics values	51
Table 4.7: Regional parameters of the distributions for L-moments	51
Table 4.8: Quantiles estimates for regional growth curves for L-moments at various probabilities of non-exceedance using GEV distribution.....	51
Table 4.9: Root Mean Square Error (RMSE) using L-moments	53
Table 4.10: Values of Probability Weighted Moments using LH-moments	62
Table 4.11: Sample LH-moments and LH-moment ratios for L_1	63
Table 4.12: Sample LH-moments and LH-moment ratios for L_2	63
Table 4.13: Sample LH-moments and LH-moment ratios for L_3	64
Table 4.14: Sample LH-moments and LH-moment ratios for L_4	64
Table 4.15: LH-moments ratios and discordancy measures for L_1	65
Table 4.16: LH-moments ratios and discordancy measures for L_2	65
Table 4.17: LH-moments ratios and discordancy measures for L_3	66
Table 4.18: LH-moments ratios and discordancy measures for L_4	66
Table 4.19: Heterogeneity measure for LH-moments.....	68
Table 4.20: Z -statistics values.....	72
Table 4.21: Regional parameters of the distributions for LH-moments, L_2 , $\eta=2$	73
Table 4.22: Quantiles estimates for regional growth curves of GEV distribution for LH-moments, L_2 , $\eta = 2$ at various probabilities of non-exceedance using GEV distribution	73
Table 4.23: Root Mean Square Error (RMSE) using LH-moment (L_2)	75
Table 4.24: Summary Root Mean Square Error (RMSE) for L- and LH-moments ...	86
Table 4.25: Regression results	87

GLOSSARY OF ACRONYMS AND ABBREVIATIONS

AMSL	Above Mean Sea Level
AWDR	The African Water Development Report
B.Sc.	Bachelor of Science Degree
CDF	Commutative Distribution Function
CGIAR	Consultative Group for International Agricultural Research
CPWF	Challenge Program on Water and Food
C_v	Coefficient of Variation
C_s	Coefficient of Skewness
C_k	Coefficient of Kurtosis
DWA	Department of Water Affairs
DMS	Department of Meteorological Services
EMWIS	Euro-Mediterranean Information System on Water Sector
FFA	Flood Frequency Analysis
km	Kilometre
km^2	Square Kilometre
M. Sc	Master of Science Degree
mm	Millimetre
m^2	Square Meter
m^3/s	Cubic Meters per Second
No.	Number
NWMPR	National Water Master Plan Review

PDF	Probability Density Function
RFA	Regional Frequency Analysis
RFFA	Regional Flood Frequency Analysis
SADC	Southern African Development Community
SD	Standard Deviation
SMEC	Snowy Mountain Engineering Corporation
SI No.	Serial Number
UB	University of Botswana
UNFCCC	United Nations Framework Convention on Climate Change
vs	Versus
WMO	World Meteorological Organization
WREE	Water Resources and Environmental Engineering

CHAPTER 1

1. INTRODUCTION

1.1 General

Flood is one of the most important disasters which can destroy the total physical and socio-economic environmental set up of the area and occurs almost in all part of the world. A study by Kundzewict (2003) showed that in recent decades, flood losses have increased worldwide. This study further indicated that an increase in flood risk is also foreseen for the future.

Flooding is also the most prevalent disaster in Africa (EMWIS, 2006); and another study by Conway (2009), revealed that Africa will suffer from floods with greater frequency. In the region of Southern Africa, represented by the Southern Africa Development Community (SADC), nations have experienced abnormally high rainfall and disastrous floods causing damage to infrastructure, loss of life and property (SADC, 2006). In another report by Holloway et al. (2013), about 14 million of the people in SADC region were affected due to the flood events from 2000-2012.

Botswana is arid to semi-arid with highly erratic rainfall. Because of this reason, most of the rivers originating in the country are ephemeral with an average flow over a period of between 10 to 70 days in a year (Parida, 2004). The mean annual rainfall ranges from over 650mm in the north-east to less than 250mm in the south-west. The national average rainfall is 475mm per year (MEWT, 2011). Most rain occurs in the months from October to April, and falls as localized showers and thunderstorms. Over 90% of the rainfall occurs in the summer months; and sometimes, 70% to 90% of the annual rainfall may occur in only one month (NWMPR, 2006). As a result of this, floods are experienced in the streams and rivers of Botswana very frequently, which have impacted on loss of life and economy.

The consequences of these floods thus should be taken in to account when any strategically important hydraulic structure is planned to be implemented at a site of interest.

1.2 Background

Major villages, towns and cities in Botswana fall within the Limpopo river system where major economic activities have been taking place. Accordingly, Limpopo catchment is the main potential water source to meet the current and future water needs to these demand centres.

Though many of the rivers in Limpopo catchment are gauged as shown in Figure 1.1, the gauging stations are largely far spaced from each other or are congested at certain localities.



Figure 1.1: Locality map of the study area
(Source: Department of Water Affairs, Botswana)

Data from some gauging stations are often discontinuous and with many gaps. In general, data records are short, with a range of being 12-39 years. Griffis and Stedinger (2007) in their work found that estimates of magnitude and frequency of floods using streamflow-gauging stations with shorter records of annual peak flow data will have higher standard errors or uncertainties when compared to estimates using stream gauges with longer annual peak flow records. On the other hand, when the data are scanty or short, the coefficient of variation, skewness and kurtosis of the distribution of the measured flood discharges will likely be higher than the corresponding coefficients of the parent flood distribution, and as such it distorts the statistical distribution (Vogel and Kroll, 1991). Because of this reason, it restricts the prediction period of the floods and poor estimates of the large flood quantiles (Rahman et al., 2014). This case always occurs in arid and semi-arid areas. Similarly, problems like data gap, measurement technique or errors, outliers, historical floods and parameter estimation methods can also influence the statistical analysis that could give poor estimates of design floods (for example Cong and Xu, 1987; Potter and Walker, 1981; Ben-Gal, 2004; Hosking and Wallis, 1997). The data gap could be attributed because of discontinuous peak-flow record due to missing or unavailable or destruction or removal of the gauge, incomplete peak-flow record because the gauge might not be operational for short period of time.

Data gap can result in erroneous peak-flow magnitude-frequency relations if missing peaks are simply excluded from the dataset and subsequent analysis (Doheny and Dillow, 1999). The same study advises that the retention, modification, or deletion of an outlier can substantially alter the statistical parameters computed from the dataset, especially if the peak-flow record is relatively short; and thus, the presence of low or high outliers must be determined, and appropriate adjustments to the relation must be made. The problem of historical data is subject to systematic error due to change of environmental conditions such as alterations in the infrastructure of the river's catchment.

Errors due to measurement techniques could be either due to a random error where measurable values being inconsistent when repeated measures of a constant attribute

or quantity are taken, or systematic error where the errors could not be determined by chance but are introduced by an inaccuracy inherent in the system.

While the above is true, choice of appropriate design flood estimation method is essential such that the errors associated with various circumstances mentioned above could be minimized. Though there are a number of routes of estimating design floods, the two main routes are the method of statistical flood frequency analysis (FFA) of peak flows and the method of unit hydrograph synthesis of the flood corresponding to a design storm (Sutcliffe, 1978). Unlike the later method, the FFA approach assumes that the flow statistical properties of the past will occur in the future and hence it is applicable for long, short or no flow records available at the design site, as in the case of the Limpopo catchment.

However, one of the major problems in hydrological design using the FFA approach is the estimation of maximum floods due to the fact that the magnitude of these events varies quite largely from year to year, and it is impossible to forecast in future times. Thus, estimating floods of large return periods is difficult because extreme events are by definition rare and the relevant data record is often short. Because of this hydrological problem, it is essential to consider observations from several independent sites with similar properties which can help to create a large sample through Regional Flood Frequency Analysis (RFFA) such that the accuracy of flood estimate can be improved. In other words, the longer the period of record, the better the likelihood of capturing the range of possible events. Moreover, RFFA is used to improve the record at regular measuring sites, and to provide estimates of frequency characteristics at sites where no data are available. This method derives a relation between flood magnitude and return period. Therefore, RFFA approach has been used in this study to estimate the design flood quantiles at desired recurrence intervals at gauged and ungauged sites in the Limpopo catchment.

1.3 Description of the Study Area

The Limpopo River Basin which discharges directly to the Indian ocean is apparently shared by four SADC Member States, i.e. Botswana, South Africa, Zimbabwe and Mozambique. However, this study area is the Limpopo Basin (*hereinafter called*

Limpopo Catchment) in Botswana. Accordingly, only stations within Botswana of this river system have been considered for RFFA. This catchment is located in eastern part of Botswana, and generally the elevation of the catchment declines from 1240m in the northeast to 542m in the east and rises to 1373m in the southeast. The upstream and downstream parts of the catchment are hilly areas, and the middle area is a flat lake catchment. This catchment has 13 stream gauging sites, all equipped with peak annual flood data collection capability. The locations of the study area and the 13 gauging sites used in the analysis are shown in Figure 1.2.

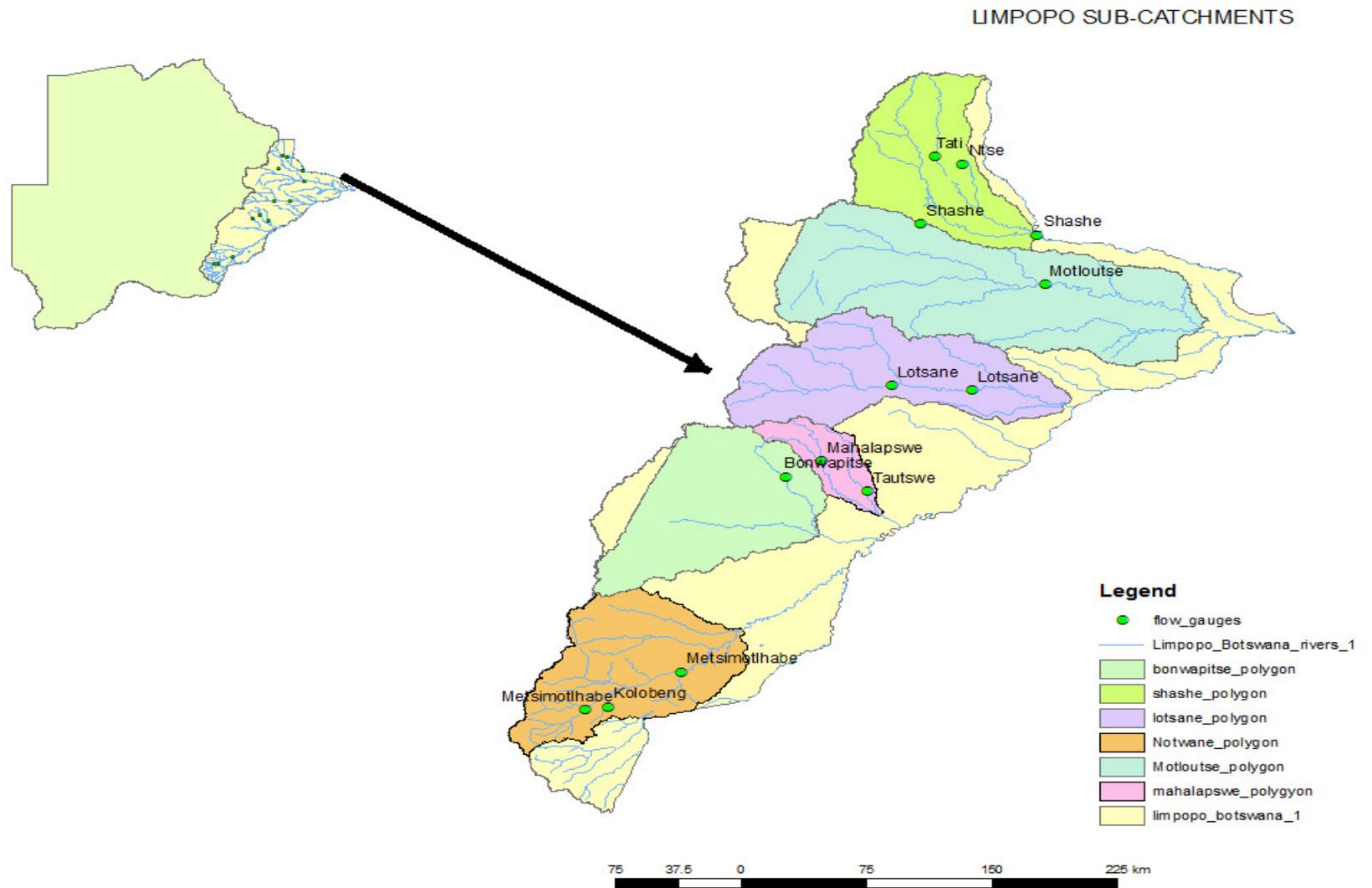


Figure 1.2
(Source: D

1.4 Statement of the Problem

The study area covers eastern corridor of Botswana where major economic activities take place. The Limpopo catchment contributes about 80% of the water needs in Botswana.

While this is the fact on the ground, there is limited data in the Limpopo catchment, which makes it difficult to understand the hydrological processes in this river system. Moreover, it could be understood that the density of gauging stations in the catchment is low, and the operation and maintenance of the stream gauging networks are difficult and annual flood series in the Limpopo catchment are too short to allow for a reliable estimation of extreme events or there is no flow record available at the site of interest.

Experiences have revealed that when the data are scanty (less than 30), it is often difficult to arrive at a proper choice of statistical distribution as well as method of parameter estimation. This situation occurs in arid and semi-arid areas like Botswana. Although it is not possible to forecast the flood events, it is possible to predict such events using newer parameter estimation techniques which yield least bias and minimum variance such that the statistical results assist with future planning in this catchment. This will be done by conducting a regional flood frequency analysis technique, based on index-flood procedures, which is a practical means of providing flood information at sites with little or no flow data available for the purposes of planning, design, construction and operation of water resources projects or decision process relating to hydraulic works or flood alleviation programs and in general for water resources management within the Limpopo catchment.

1.5 Objective of the Research and Corresponding Research Questions

The general objective of this Dissertation is to develop a regional growth curve that can be used for estimation of flood quantiles at desired recurrence intervals at the gauged sites. The regional growth curve can also be used with the catchment characteristics for design flood estimations at the desired recurrence interval for ungauged stations in the catchment.

The main research questions are also derived based on the specific objectives, which shall be addressed and solved throughout this study process.

With the above background and general objective of this study, the specific objectives of the research and the corresponding research questions are tabulated in Table 1.1:

Table 1.1: Specific Objectives and Research Questions

Sl No.	Specific Objectives	Research Questions
1	Identify geographical continuous regions for gauged sites where more than 10 years flow data have been recorded, and determine homogeneous regions,	What is the condition to satisfy a homogeneous region?
2	Select an appropriate distribution that can be used for parameters and quantiles estimation at desired recurrence intervals,	What are the procedures to be followed in selection of the best distribution to fit the observed data and give least bias quantile estimates and hydrologic risk across the Limpopo Catchment?
3	Develop regional growth curves based on the observed data of the gauged sites that can be used for estimation of flood quantiles at specified risks (design floods) at either gauged or ungauged sites, which can be used for design of hydraulic structures.	<ul style="list-style-type: none"> • Which newer methods should be applied to estimate the parameters and thus quantiles of the chosen distribution at desired recurrence intervals? • What procedure is to be used to develop a regional growth curve? • How the relationship between quantiles and physical characteristics will be developed for ungauged sites?

1.6 Significance of the Study

There are limited researches undertaken in the Limpopo catchment that could be due to limited information at the sites of interest and thus it has become a challenge to have the results of design floods at desired recurrence intervals for the implementation of hydraulic infrastructure. However, most of the major rivers in Botswana and water infrastructure assets such as dams are located within the Limpopo catchment and is viewed as one of the potential source to meet the current and future water needs such as domestic, mining, industrial, commercial and agriculture.

In order to have reliable design flood information, use of appropriate newer parameter estimation methods like L-moments and LH-moments is important. Hence, it is our motivation to carry out this study in this region because on one hand the LH-moments based on regional flood frequency analysis have recently emerged, and on the other hand relatively more data for longer period of time have been collected after the previous studies were conducted and as such the design events can be obtained with greater reliability. Therefore, the hydrological information (results) of this study is believed to be helpful for the relevant authorities for water resources project planning and flood alleviation programs within the river system.

1.7 Scope of the Study

The study basically covers the area in the eastern part of Botswana where most of the streams in the country are located, and where a number of hydraulic infrastructure like dams do exist, under construction or proposed to be constructed in the future. Accordingly, the maximum design flood at desired recurrence intervals for the streams within the Limpopo catchment is essential generally for water resources management within this catchment. For this purpose, 13 gauged stations within the Limpopo basin in Botswana have been considered based on annual maximum flood data observed between 1968 and 2008.

For estimation of the design floods at desired design risks (recurrence intervals), Regional Flood Frequency Analysis approach using the newer methods of L-moments and LH-moments procedures will be utilized to establish regional homogeneity as

well as to arrive at the quartile estimations and development of a regional growth curve.

CHAPTER 2

2 LITERATURE REVIEW

2.1 Related Previous Studies

Previous studies by Farquharson et al. (1992) developed regional growth curves for sample arid areas in the world where Botswana was also included. The regional growth so developed could be suitable for large-scale application, but for small-scale application for the purpose of designing strategic hydraulic structures in Limpopo catchment, a specific regional growth curve for this particular region should be developed. Another study was undertaken by SMEC (NWMPR, 2006) on surface water resources of Botswana to perform certain specified hydrological analysis that are required for hydraulic design purposes. However, the study was carried out with limited flood information. A study on the flood characteristics of selected rivers in Botswana using L-moments was also carried out by Parida (2004), which was addressed based on limited data.

2.2 Probability Concepts and Theoretical Background

Events that cannot be predicted precisely are often called random. Many if not most of the inputs to, and processes that occur in, water resources systems are to some extent random.

Suppose a random variable X may take k different values, with the probability that $X = x_i$ defined to be $P(X = x_i) = p_i$. The probabilities p_i must satisfy the following:

$$0 \leq p_i \leq 1 \text{ for each } i \quad (1)$$

$$p_1 + p_2 + \dots + p_k = 1 \quad (2)$$

2.2.1 Random Variables and Distributions

Let X denote a continuous random variable, and x a possible value of that random variable X . For any real-valued random variable X , its cumulative distribution

function (CDF) $F_X(x)$, equals the probability that the value of X is less than or equal to a specific value or threshold x :

$$F_X(x) = P(X \leq x) \quad (3)$$

$F_X(x)$ is the nonexceedance probability for the value x .

The probability density function (PDF) is the derivative of the CDF, so that:

$$f_X(x) = \frac{dF_X(x)}{dx} \geq 0 \quad (4)$$

2.2.2 Quantiles and Return Period

According to Loucks and Beek (2005), the simplest approach to describing the distribution of a random variable is to report the value of several quantiles. If X is a continuous random variable (example river/stream flow), then in the region here $f_X(x) \geq 0$, the quantiles are uniquely defined and are obtained by solution of

$$F_X(x_p) = p \quad (5)$$

The p th quantile is also the $100p$ percentile. In floodplain management and the design of flood control structures, the 100-year flood $x_{0.99}$ is a commonly selected design value (Loucks and Beek, 2005).

In a general way, Maidment (1993) states that the $100p$ percentile is often called the $100(1-p)$ percent exceedance events because it will be exceeded with probability $1-p$.

The return period or recurrence interval is often specified rather than the exceedance probability. For example, the annual maximum flood-flow exceeded with a 1 percent probability in any year, or chance of 1 in 100, is called the 100 year flood. In many hydraulic engineering applications, it may be necessary to determine the probability of occurrence of extreme flood events. Accordingly, the relationship between probability of occurrence -probability of exceedance (P) - and the recurrence interval (T) is given by:

$$T = \frac{1}{P} \text{ or } P = \frac{1}{T} \quad (6)$$

where T is expressed as occurrence of a given flood event on an average once in T -years or may not occur at all in the T -years but may occur say 2 times in the period of next $2T$ -years and likewise. Equation (6) can also be written in terms of probability of non-exceedance (say $p=1-P$) as:

$$p = F(x) = 1 - P = 1 - \frac{1}{T} \quad (7)$$

Here, there are two ways that return period can be understood (Maidment, 1993):

- (i) In a fixed T -year period, the expected number of exceedances of the T -year event is exactly 1 if the distribution of floods does not change over that period; thus on average one flood greater than the T -year flood level occurs in a T -year period.
- (ii) Alternatively, if floods are independent from year to year, the probability that the first exceedance of level x_p occurs in year k is the probability of $(k-1)$ years without an exceedance followed by a year in which the value of X exceeds x_p :

$$P(\text{exactly } k \text{ years until } X \geq x_p) = p^{k-1}(1-p) \quad (8)$$

The above is also useful in assessing the risk involved while designing a hydrologic structure. Or in other words, to compute the probability of occurrence of at least one occurrence of the designed flow during the life time of a structure which also can be expressed as:

$$Risk = (1 - e^{-L/T}) \text{ or } [1 - (1 - P)^n] \quad (9)$$

where, $L = n =$ Design life of the structure and $T=$ Design recurrence interval in years= $1/p$.

Taking various considerations of hydrologic risk and life span of the hydraulic structure, T is suggested to lie between 50 and 1000 years (Parida, 1998).

2.3 Choice of Frequency Distribution

Among other difficulties in FFA are related to the identification of the appropriate statistical distribution for describing data and to the estimation of the parameters of a

selected distribution. Accordingly, collective evidence about distributional behaviour of floods on many rivers needs to be studied (WMO, 1989). Several fundamental techniques have been used in the past for evaluating the suitability of different distributions for annual maximum flood series such that both the descriptive and predictive abilities are met. The two approaches viz: behavioural analysis and robustness procedures have been utilized in recent days for the choice of the appropriate distribution (details are given in WMO (1989)). Nevertheless, the robustness approach has been followed in our analysis of selecting the appropriate distribution because it tests whether a distribution and method of parameter estimation, considered jointly are insensitive to departures from assumptions made in their use. Several fundamental issues arise when selecting a distribution (Maidment, 1993) such as (1) the true distribution from which the observations are drawn, (2) robust estimates of design quantiles and hydrologic risk, and (3) consistency of the proposed distribution with the available data for a site. Because there is no firm theoretical basis for choice between distributions, goodness of fit tests are often used to select the appropriate distribution which best fits the flood data (Sutcliffe, 1978). He further indicated in his work that earlier investigations suggested that the three parameter distributions were found to be better or more flexible than the two parameter distributions. According to WMO (1989), robustness studies indicate that quantile estimates using two parameter distributions suffer more from bias than those based on three parameter ones.

Though various probability distributions have been assumed to characterize various random variables in hydrology (Dingman, 2008), thus, the most commonly used three parameter distributions have been summarized by Hosking and Wallis (1997) as GPA, GEV, GLO, LN3 and PE3. Apparently, the true distribution is probably too complex to be of practical use – rather, the newer methods of parameter estimation techniques like L-moments and LH-moments skewness-kurtosis and CV-skewness ratio diagrams are good for investigating what simple families of distributions are consistent with available data sets for a region. Standard goodness-of-fit criteria such as Z-statics have been used to see how well a member of each family of distribution can fit a sample. The distribution that best fit each data set is considered robust, which is used for parameter estimates and yields reliable flood quantile and risk estimates.

A recent study by Parida (2004) on Limpopo/Makgadikgadi region using L-moments reported that GEV distribution fitted well along with another possible distribution viz: Generalized Logistic (GLO). Wang (1997) identified that the GEV distribution by using LH-moments was formulated and found that GEV fitted well to characterize the upper part of distributions and larger events in data. Similarly, other studies (Gheidari, 2013; Bhuyan, et al., 2009; and Shabri, 2008) reported that GEV distribution along other distribution functions (namely: GPA and GLO) using LH-moments performed better than other distributions. However, the LH-moment has been developed for only three most commonly used distributions viz: the GLO, GEV and GPA, while the L-moment has been developed for most of the common distributions including the above three distributions. In line with this, the frequency distribution analysis in this study will be carried out using the aforementioned three distributions to compare the performance of the two newer parameter estimation methods and thus for the selection of the least bias candidate.

2.4 Methods of Parameter Estimation

2.4.1 Historical Development of Different Methods

The flood estimation procedure generally emphasises on two aspects viz: on the appropriate choice of a statistical model and a robust method of parameter estimation, such that both descriptive and predictive aspects are well covered (Cunnane, 1987). In other words, the descriptive property relates to the requirement that the chosen distribution shape resembles the observed sample distribution of floods and that random samples drawn from the chosen model distribution must be statistically similar to the properties of real flood series; and that of descriptive property relates to the requirement that quantile estimates are robust with small bias and standard error (WMO, 1989). While this is true, one of the first steps towards obtaining a coherent analysis is the detection of outliers in the observations. Detected outliers are candidates for aberrant data that may otherwise adversely lead to model misspecification, biased parameters estimation and incorrect results (Ben-Gal, 2004). Incorrect data values either during recording and/ measuring and the circumstances under which the data so collected may have changed over time (Hosking and Wallis, 1997) and as such these cast doubt on estimation of the probability distribution.

Similar studies by Hosking and Wallis (1997) illustrated that the measurement error tends to inflate the coefficient of variation and coefficient of skewness of the observed data and may be expected to lead to overestimation of the upper-tail quantiles of the frequency distribution. Even though the data may be reliable, it is still important to check for errors or outliers prior to modelling and analysis (Hosking & Wallis, 1997; Meshgi & Khalili, 2007b; Gheidari, 2013; Shabri, 2002; Walfish, 2006; and Ben-Gal, 2004). Potter and Walker (1981) studies have shown that the effect of modelling error because of the measurement error is far more damaging than the sampling error. While this is understood, a study by WMO (1989) has shown that outliers have only a small effect if an efficient method of parameter estimation is used - say L-moments and LH-moments.

In view of the above, the extreme value nature and the need for frequency analysis of annual flood peaks have over the years motivated researchers to explore the merits of a number of probability distribution functions. The parameters of probability distribution functions can be estimated by various methods. The oldest and widely understood technique for fitting frequency distributions to observed data are method of moments (MOM) or product moments, and the method of maximum likelihood (ML) (Vogel and Fennesey, 1993). Through conducting experiments based on daily streamflow, they found that both skewness and coefficient of variation exhibit remarkable bias for highly skewed populations, and hence the product moments were found to be of little value for discriminating among potential candidate distributions. For many distributions, the inverse functions of their distribution functions using MOM and ML cannot be explicitly derived (Greenwood et al, 1979). They also presented that ML estimates of the parameter values are not easily obtained. However, because of the generally small sample sizes available for characterizing hydrologic time series in the Limpopo catchment, estimates of the third and higher product moments are usually very uncertain. Greenwood et al. (1979) presented the probability weighted moments (PWMs) as an alternative to the more conventional methods of the aforementioned techniques and Hosking et al. (1985) showed that the PWM method is superior to the ML method when the GEV distribution is used for longer return periods, that is, return period of more than 100 years. Greenwood et al. (1979) found that many distributions may be explicitly defined as both the distribution

function and their inverse functions using PWMs. They showed that all higher-order PWMs are simply linear combinations of the order statistics; and thus they declared that the relations between the PWMs and the distributions' parameters are of simpler analytical structure than those between the conventional moments and parameters.

Researchers further worked on the development of procedures for utilization of the linear combinations of the PWMs and, identified as the linear moments. Hosking (1990) thus introduced L-moments as a linear combination of PWMs and he used L-moment ratio diagram to identify the underlying parent distributions and the L-moment ratios for testing goodness-of-fit of different distributions. Since sample estimators of L-moments are always linear combinations of the ranked observations (Vogel and Fennesey (1993)), they are subject to less bias than ordinary product moment estimators. The variance and skewness (Equations (17) and (18), respectively) require squaring and cubing the observations in the conventional moments, which causes them to give greater weight to the observations far from the mean, resulting in substantial bias and variance. Hosking and Wallis (1997) investigated that L-moments have the theoretical advantages over conventional moments of being able to characterize a wider range of distributions and when estimated from a sample, of being more robust to the presence of outliers in the data.

Wang (1997) on the other hand introduced the LH-moments as a generalization or modified form of the L-moments, which is intended to improve parameter estimation when higher level of the liner moments are utilized. Statistical analysis of extremes is often conducted for predicting large return period events. Thus more relevant to the analysis are the upper part of distributions and more extreme sample events. Wang (1997) in his work thus compared the L-moments and LH-moments that distribution curves fitted by using L-moments are influenced too much by small annual maximum flows, leading to poor prediction of large return period events; and in contrast he found that the curves fitted by the LH-moments better capture the trends shown by the larger flows and thus the LH-moments estimates of large return period events are less influenced by the small annual maximum flows. These findings were further investigated and supported by different researchers (Gheidari, 2013, Bhuyan et al, 2009; Lee & Maeng, 2003; Meshgi & Khalili, 2009).

The letter ‘H’ in LH-moments denotes higher order L-moments, the L_1 -, L_2 -, L_3 - and L_4 -moments denote the first, second, third and fourth higher order LH-moments, respectively. The zero-order LH-moments or L_0 is equivalent to simple L-moments. Development of the LH-moments by Wang (1997) was extended to four levels (L_1 to L_4), with the L-moments (L) considered as the special case. He concentrated only on the GEV distribution while Meshgi and Khalili (2007a, 2007b) developed the LH-moments for the generalised Pareto (GPA) and generalised logistic (GLO) distributions in the RFFA of Karkhe watershed, located in western Iran. Gheidari (2013), Bhuyan et al (2009), Lee and Maeng (2003) did their work using RFFA based on GPA, GLO and GEV distributions and presented that the results by LH-moments are more efficient and robust for obtaining improved values of flood peaks than the L-moments. Unlike the LH-moments, the L-moment was developed for many distributions. Accordingly, the L-moment method is also a robust method of frequency distribution parameter estimations and has been widely used for many hydrological processes. With the above research findings and while the theory of LH-moments is relatively new and is not widely used for RFFA for estimation of design floods at desired recurrence intervals, the methods of the L-moments and LH-moments will be used for regional flood frequency analysis of the Limpopo catchment and a comparative study will be made between these two newer methods.

2.4.2 Conventional Methods of Parameter Estimation

Product Moments of Distribution (MOM)—these are also called *theoretical moments* - A common approach to describing a distribution’s centre, spread and shape is by reporting the moments of a distribution. From a set of observations (X_1, \dots, X_n), the distribution’s centre, spread and shape of the moments of a distribution can be estimated. Accordingly, the statistical estimators of the first moment about the origin (the mean of X) (Chow et al., 1988), the second moment (the variance), and the square root of the variance (the standard deviation) are given by equations (10), (11) and (12), respectively.

$$\mu_x = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx \quad (10)$$

$$\sigma_X^2 = \text{Var}(X) = E[(X - \mu_X)^2] \quad (11)$$

$$\sigma_X = \sqrt{\text{Var}} \quad (12)$$

While the mean μ_X is a measure of the central value of X and also known as the location parameter, the standard deviation σ_X is a measure of the spread of the distribution of X about μ_X .

Another measure of the variability in X is the coefficient of variation given by:

$$Cv_X = \frac{\sigma_X}{\mu_X} \quad (13)$$

The coefficient of variation expresses the standard deviation as a proportion of the mean. It is useful for comparing the relative variability of the flow in rivers of different sizes, or of rainfall variability in different regions when the random variable is strictly positive.

The third moment about the mean, denoted λ_X , measures the asymmetry, or skewness, of the distribution and is given by:

$$\lambda_X = E[(X - \mu_X)^3] \quad (14)$$

Typically, the dimensionless coefficient of skewness γ_X is reported rather than the third moment λ_X . The coefficient of skewness -a measure of asymmetry- is the third moment rescaled by the cube of the standard deviation so as to be dimensionless and hence unaffected by the scale of the random variable:

$$\gamma_X = \frac{\lambda_X}{\sigma_X^3} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3} \quad (15)$$

The coefficient of kurtosis which describes the thickness or peakedness of a distribution's tails is given by:

$$\beta_X = \frac{E[(X - \mu_X)^4]}{\sigma_X^4} \quad (16)$$

Sample Estimators – it is also called *sample moments* - When the distribution of a random variable is not known, but a set of observations $\{x_1, \dots, x_n\}$ is available, the

moments of the unknown distribution of X can be estimated based on the sample values using the following equations:

The sample estimate of the mean:

$$\hat{\mu}_X = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (17)$$

The sample estimate of the variance:

$$\hat{\sigma}_X^2 = S_X^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (18)$$

The sample estimate of skewness:

$$\hat{\lambda}_X = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (X_i - \bar{X})^3 \quad (19)$$

The sample estimate of the coefficient of variation:

$$\hat{C}_{v_x} = \frac{S_x}{\bar{X}} \quad (20)$$

The sample estimate of the coefficient of skewness:

$$\hat{\gamma}_X = G = \frac{\hat{\lambda}_X}{S_X^3} = \frac{n}{(n-1)(n-2)S_X^3} \sum_{i=1}^n (X_i - \bar{X})^3 \quad (21)$$

All of these sample estimators provide only estimates of actual or true values. Unless the sample size n is very large, the difference between the estimators and the true values of μ_x , σ_x^2 , λ_x , C_{v_x} , and γ_x may be large.

The sample estimate of kurtosis:

$$\hat{\beta}_X = \left[\frac{n^2}{\{(n-1)(n-2)(n-3)\}} \right] (X_i - \bar{X})^4 \quad (22)$$

The sample estimate of the coefficient of kurtosis:

$$C_k = \frac{\hat{\beta}_X}{S_X^4} = \left[\frac{n^2}{\{(n-1)(n-2)(n-3)\}S_X^4} \right] \sum (X_i - \bar{X})^4 \quad (23)$$

Using these moments, the parameters of a distribution can be determined. Hydrologists often favour this method for its simplicity. However, moments of higher

order than these four are not commonly used in the statistical analysis of hydrologic data because most hydrologic data do not have sufficiently long length of record and thus cannot warrant reliable estimates of the moments of higher order (Chow, 1964). This is because as the exponent of the standard deviation increases, the error increases by the same exponent.

Probability Weighted Moments (PWMs) - This method which was introduced by Greenwood *et al.* (1979) involves moment estimation from the linear weighted combination of the ordered statistics and use them for the estimation of parameters (Dingman, 2002). The statistics are:

$$\beta_0 = \frac{1}{n} \sum_{i=1}^n x_i \quad (24)$$

which is equivalent to the mean.

$$\beta_1 = \frac{1}{n} \sum_{i=1}^n x_i \frac{(i-1)}{(n-1)} \quad (25)$$

$$\beta_2 = \frac{1}{n} \sum_{i=1}^n x_i \frac{(i-1)(i-2)}{(n-1)(n-2)} \quad (26)$$

and in general,

$$\beta_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)(i-3)\dots(i-r)}{(n-1)(n-2)(n-3)\dots(n-r)} X_r \quad (27)$$

where, β_r = PWM of order r, such that r = 1, 2, ...; n = sample size and i = rank of the ordered data (X_i) in the ascending order.

This method is useful for distributions with explicitly defined inverse form, that is,

If $F(x)$ exists, then $x(F)$ also exists.

It also works well for situations where records are extremely short and stream flow samples are highly skewed and highly kurtotic. This situation often arises when the sample size at each site is small (Kumar and Chander, 1987).

2.4.3 Newer Methods of Parameter Estimation

A statistical distribution can be misled or distorted due to the short record, presence of outliers, data gaps and measurement errors. However, researchers like Hosking and Wallis (1997), Parida (1999), Meshgi and Khalili (2007a), Bhuyan et al. (2009), Gheidari (2013), Lee and Maeng (2003) and Murshed et al. (2013) have shown that the use of the L- and LH-moments can overcome the above problems to a great extent, while being able to identify a suitable distribution and also producing unbiased estimates due to use of linear moments of ordered statistics. L-and LH-moments are currently being used widely for parameter and quantile estimations. These two methods including their relative advantages are also discussed in brief in the subsequent sections.

2.4.3.1 L-Moments

In many hydrologic applications, sample estimators of L-moments (L) are linear combinations of the ranked observations, and thus do not involve squaring or cubing the observations as do the product-moment estimators. As a result, L-moment estimators of the dimensionless coefficients of variation and skewness are almost unbiased and have very nearly a normal distribution. On the other hand, the product-moments estimators of the coefficient of variation and of skewness are both highly biased and highly variable in small samples (Maidment, 1993). For further discussion about these issues, refer Hosking and Wallis (1997).

The following are specific advantages of L-moments approach over ordinary product moments (Zafirakou-Koulouris et al., 1998):

1. L-moment ratio estimators of location, scale, and shape are nearly unbiased, regardless of the probability distribution from which the observations arise (Hosking, 1990);
2. L-moment ratio estimators such as L-CV, L-skewness, and L-kurtosis can exhibit lower bias than conventional product moment ratios, especially for highly skewed samples;

3. The L-moment ratio estimators of L-CV and L-skewness do not have bounds, which depend on sample size as do the ordinary product moment ratio estimators of CV and skewness;
4. L-moment estimators are linear combinations of the observations and thus are less sensitive to the largest observations in a sample than product moment estimators, which square or cube the observations;
5. L-moment ratio diagrams are particularly good at identifying the distributional properties of highly skewed data, whereas ordinary product moment diagrams are almost useless for this task (Vogel and Fennessey, 1993).
6. L-moment provides better flood estimates for low return periods.

In this method, no apriori choice of distribution is made and is prompted by the use of L-Skewness (τ_3) and L-Kurtosis (τ_4) diagram based on the location of the plot of L-Skewness and L-Kurtosis value using the observed data.

However, when L-moment is compared with LH-moments, the specific disadvantage of this method is that distribution curves fitted by L-moments are influenced too much by small maximum data leading poor prediction of large return period events. In other words, L-moments are oversensitive to the lower part of distributions and give insufficient weight to large sample values that actually contain useful information on the upper distribution.

Computed values of τ_3 and τ_4 are then plotted on to the theoretical L-moment ratio diagram which suggests the likely/possible statistical distribution which could be used for analysis.

2.4.3.2 LH-Moments

LH-moments (L_1 to L_4), a generalization of L-moments, are introduced by Wang (1997) for characterizing the upper part of distributions and larger events in data. Being newer method of parameter estimating technique, LH-moments are more useful than L-moments for characterizing distributions, for interpreting data, and for regional analysis, just like L-moments are more useful and easier to interpret than PWMs. However, LH-moment has been developed for only three common distributions viz: GLO, GEV and GPA distributions.

While conventional moments fail to converge to a solution due to their mathematical complexity, LH-moments overcome some of these difficulties. Therefore, LH-moments are found unbiased and robust for characterizing the data and estimation of the parameters and quantiles. LH-moments thus have gained popularity in recent years; and therefore, regional flood frequency analysis is also applied by the LH-moments method again for the Limpopo Basin. The same procedure of the L-moments has been followed to construct the LH-moments ratio diagram to suggest the possible statistical distribution.

There are a number of specific advantages of LH-moments approach, in RFFA literature including: Wang, 1997; Meshgi and Khalili (2007a); Bhuyan et al. (2009); Gheidari (2013); Lee and Maeng (2003); and Murshed et al. (2013); some of which are:

1. Unlike L-moment, the method of LH-moment reduces undesirable influences that small sample events may have on the estimation of large return period events.
2. LH-moment diminishes the influence of small sample values as η increases.
3. LH-moment overcomes complex mathematical solution.
4. LH-moment fits the distribution well.
5. LH-moment increases efficiency for higher quantile estimations yet with less bias.
6. LH-moment is useful in avoiding undue influence that more frequent observations have on less frequent observations.
7. LH-moment provides improved values of flood peaks for higher return periods.

However, when LH-moment is compared with L-moments, the following are specific disadvantages of this method:

1. There is limited use of LH-moments in flood frequency analysis due to the difficulty of formation of homogeneous region and development of LH-moments for other commonly used distributions.
2. LH-moment is less efficient than L-moments for shorter return periods.

Computed values of τ_3 and τ_4 are then plotted on to the theoretical L-moment ratio diagram which suggests the likely/possible statistical distribution which could be used for analysis.

CHAPTER 3

3 METHODOLOGY

3.1 General

The method of L-moments introduced by Hosking (1990) has the advantage over the previous conventional methods in providing parameter estimates that are nearly unbiased and highly efficient and thus are better suited for use in constructing moment diagrams. Wang (1997) introduced LH-moments, which are modified forms of the L-moments, to characterize the upper part of a distribution. Though LH-moments are advocated by some researchers as a better parameter and quantile estimations method over the L-moments, the methodology of the two procedures are described in the subsequent sections to compare their performance using the Limpopo real flood data.

3.2 L-Moments

Hosking (1990) introduced L-moments as a linear combination of PWMs. The PWMs, defined by Greenwood et al. (1979) for a non-negative integer may be given as:

$$\beta_r = E[x\{F(x)\}^r] \quad (28)$$

which can be written as:

$$\beta_r = \int_0^1 x(F)F^r dF, r = 0, 1, 2, \dots \quad (29)$$

where, $F = F(x)$ is the cumulative distribution function (CDF) for x , $x(F)$ is the inverse CDF of x evaluated at the probability F . When $r = 0$, β_0 is equal to the mean of the distribution $\mu = E[x]$.

Hosking (1990) defined r^{th} L-moments related to the r^{th} PWMs as:

$$\lambda_{r+1} = \sum_{\kappa=0}^r \beta_{\kappa} (-1)^{r-\kappa} \binom{r}{\kappa} \binom{r+\kappa}{\kappa} \quad (30)$$

Using the ranked data (in the ascending order), an unbiased estimate of sample Probability Weighted Moments can be computed from the following equation:

$$\beta_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)(i-3)\dots(i-r)}{(n-1)(n-2)(n-3)\dots(n-r)} X_r \quad (31)$$

where, β_r = PWM of order r, such that r = 1, 2, n= sample size and i = rank of the ordered data (X_i) in the ascending order.

Since the L-moments are related to PWMs, the first four L-moments can be calculated from:

$$\lambda_1 = \beta_0$$

$$\lambda_2 = 2\beta_1 - \beta_0$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (32)$$

These four moments are analogous to the first four conventional moments of X, such that the L-coefficient of variance, $L - C_v(\tau_2)$; L-skewness $L - C_s(\tau_3)$; and L-kurtosis, $L - C_\kappa(\tau_4)$ can be written as:

$$\tau_2 = \frac{\lambda_2}{\lambda_1} \quad (33)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad (34)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \quad (35)$$

Computed values of τ_3 and τ_4 are then plotted on to the Theoretical L-Moment ratio diagram (shown as +) which suggests the likely/possible statistical distribution which could be used for analysis. The theoretical L-moment ration diagram for five commonly used distributions of three parameters is presented in Figure 3.1.

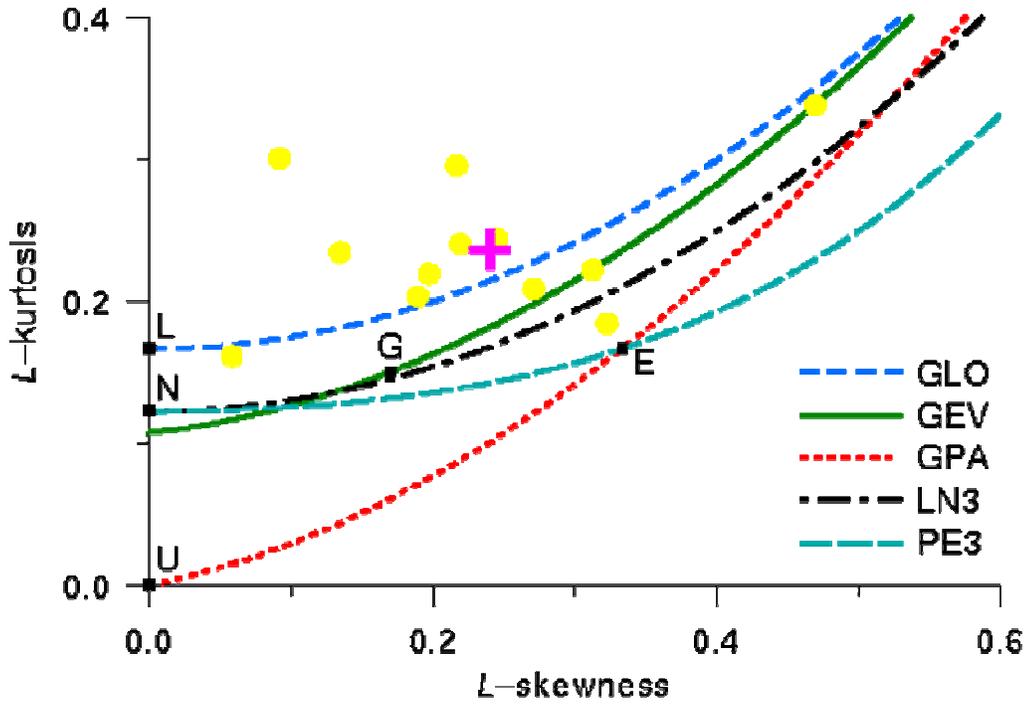


Figure 3.1: Theoretical plots of L-skew versus L-kurtosis diagram for statistical distributions (viz: Generalised Pareto (GPA), Generalised Extreme Values (GEV), Generalised Logistic (GLO), 3 Parameter Log-normal (LN3), Pearson Type 3 (PE3) Distribution) (Hosking and Wallis, 1997)

3.3 LH-Moments

Wang (1997) introduced LH-moments as a linear combination of higher PWMs. The fundamental concepts of the LH-moments are essentially the same as the L-moments, varying only in the corresponding coefficients used as multiplication terms (Meshgi and Khalili, 2007a). The first four LH-moments as defined by Wang (1997) are then given below:

$$\lambda_1^0 = (X_{\eta+1:\eta+1})$$

$$\lambda_2^\eta = \frac{1}{2}E(X_{\eta+2:\eta+2} - X_{\eta+1:\eta+2})$$

$$\lambda_3^\eta = \frac{1}{3}E(X_{\eta+3:\eta+3} - 2X_{\eta+2:\eta+3} + X_{\eta+1:\eta+3})$$

$$\lambda_4^\eta = \frac{1}{4}E(X_{\eta+4:\eta+4} - X_{\eta+3:\eta+4} - 3X_{\eta+2:\eta+4} + X_{\eta+1:\eta+4}) \quad (36)$$

Wang (1997) defined the sample LH-moments as follows:

$$\lambda_1^\eta = \beta_\eta$$

$$\lambda_2^\eta = \frac{1}{2}(\eta + 2)[\beta_{\eta+1} - \beta_\eta]$$

$$\lambda_3^\eta = \frac{1}{3!}(\eta + 3)[(\eta + 4)\beta_{\eta+2} - 2(\eta + 3)\beta_{\eta+1} + (\eta + 2)\beta_\eta]$$

$$\lambda_4^\eta = \frac{1}{4!}(\eta + 4)[(\eta + 6)(\eta + 5)\beta_{\eta+3} - 3(\eta + 5)(\eta + 4)\beta_{\eta+2}$$

$$+ 3(\eta + 4)(\eta + 3)\beta_{\eta+1} - (\eta + 3)(\eta + 2)\beta_\eta] \quad (37)$$

where β_r is the r th root sample PWM for $r=1, 2, \dots$ and $\eta=0, 1, 2, \dots$, which was defined by Greenwood et al. (1979) as:

$$\beta_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)(i-3)\dots(i-r)}{(n-1)(n-2)(n-3)\dots(n-r)} X_{i:n} \quad (38)$$

in which,

$$\beta_r = \frac{\int_0^1 x(F)F^r dF}{\int_0^1 F^r dF} = (r + 1) \int_0^1 x(F)F^r dF = (r + 1)\beta_r \quad (39)$$

The LH-moment ratios LH-coefficient of variation ($LH - C_v, \tau_2^\eta$), LH-coefficient of skewness ($LH - C_s, \tau_3^\eta$) and LH-coefficient of kurtosis ($LH - C_k, \tau_4^\eta$) are defined as:

$$\tau_2^\eta = \lambda_2^\eta / \lambda_1^\eta \quad (40)$$

$$\tau_3^\eta = \lambda_3^\eta / \lambda_2^\eta \quad (41)$$

$$\lambda_4^\eta = \lambda_4^\eta / \lambda_2^\eta \quad (42)$$

As η increases, LH-moments reflect more and more the characteristics of the upper part of distributions and larger events in data.

As discussed in Chapter 1, the LH-moments were investigated and developed only for the three distributions of the GEV, GLO and GPA (Gheidari, 2013). Accordingly, details of the parameter and quantile estimations for these three distributions are provided in Section 3.4. Accordingly, the L- and LH-moments diagrams will be generated for the commonly used three parameter distributions of the GPA, GEV, GLO distributions to select the robust distribution model for parameter and quantile estimations.

3.4 Parameters Estimation

As indicated in Section 2.4, the LH-moments method was developed only for the GEV, GLO and GPA distributions. The regional average LH-moment ratios $\tau_{\eta,2}^R$ and $\tau_{\eta,3}^R$ together with $\hat{\lambda}_1^\eta$, for $L_\eta, \eta=0, 1, 2, 3, 4$, that is, (L to L₄) will be used for estimates of the parameters of the above distributions as proposed by Hosking and Wallis (1997) for L-moments and Bhuyan et al (2009) for LH-moments. Accordingly, the regional average LH-moment ratios $\tau_{2,\eta}^R$ and $\tau_{3,\eta}^R$ together with $\hat{\lambda}_1^\eta=1$, for $\eta = 0,1,2,3,4 \dots$ will be used for estimates of the parameters of GEV, GLO and GPA distributions.

Let X be a real-valued ordered random variable of size n , such that $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ with cumulative distribution function $F(x)$ and quantile function $x(F)$, then the L-moments (Hosking, 1990) and LH-moments of X (Wang, 1997) can be defined in the following sections.

3.4.1 LH- and L-moments for the GEV distribution

The probability density function of the GEV is given by:

$$f(x) = \alpha^{-1} e^{-(1-\kappa)y - e^{-y}}, y = \begin{cases} -\kappa^{-1} \ln \left\{ 1 - \frac{\kappa(x-\xi)}{\alpha} \right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases} \quad (43)$$

The cumulative distribution function of the GEV can be written as:

$$F(x) = e^{-e^{-y}} \quad (44)$$

$$F(x) = \exp\left\{-\left[1 - \frac{\kappa}{\alpha}(x - \xi)\right]^{\frac{1}{\kappa}}\right\}, \text{ if } \kappa \neq 0 \quad (45)$$

$$= \exp\left\{-\exp\left[\frac{1}{\alpha}(x - \xi)\right]\right\}, \text{ if } \kappa = 0 \quad (46)$$

The inverse of the CDF will give the quantile function of the GEV distribution for the required return periods and is given by:

$$q(F) = x(F) = \xi + \frac{\alpha}{\kappa}\{1 - [-\ln F]^\kappa\}, \kappa \neq 0 \quad (47)$$

$$= \xi - \alpha \ln(-\ln F), \kappa = 0 \quad (48)$$

where ξ is a location parameter, α is a scale parameter and κ is a shape parameter

In Equation (47), $-\infty < X \leq \xi + \alpha/\kappa$ for $\kappa > 0$ and $\xi + \alpha/\kappa \leq X < +\infty$ for $\kappa < 0$.

3.4.1.1 L-Moments

Hosking and Wallis (1997) estimated the parameters of the GEV distribution for the case of L-moments as:

$$c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3} \quad (49)$$

$$\hat{\kappa} \approx 7.8590c + 2.9554c^2 \quad (50)$$

$$\hat{\alpha} = \frac{\hat{\kappa}\lambda_2}{\Gamma(1+\hat{\kappa})(1-2^{-\hat{\kappa}})} \quad (51)$$

$$\hat{\xi} = \lambda_1 - \frac{\hat{\alpha}}{\hat{\kappa}}[1 - \Gamma(1 + \hat{\kappa})] \quad (52)$$

where $\Gamma(\cdot)$ is a Gamma function

3.4.1.2 LH-Moments

Equations (53) – (56) were developed by Lee and Maeng (2003) for the parameters of the GEV distribution at different LH-moments levels as:

$$c = \frac{(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_{\eta}}{(\eta+3)\beta_{\eta+2} - (\eta+1)\beta_{\eta}} - \frac{\ln(\eta+2) - \ln(\eta+1)}{\ln(\eta+3) - \ln(\eta+1)} \quad (53)$$

$$\hat{\kappa} = a_1 c + a_2 c^2 \quad (54)$$

Table 3.1 gives the coefficient values of Equation (54) for different LH-moments levels.

Table 3.1: Corresponding Coefficients of Equation (54) for different levels (η) of the LH-moment (Lee and Maeng 2003)

H	a_1	a_2
0	7.8589	2.9534
1	11.9082	2.7787
2	15.9316	2.7301
3	19.9455	2.7072
4	23.9546	2.6936

$$\hat{\alpha} = \frac{\hat{\kappa}[(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_{\eta}]}{\Gamma(1+\hat{\kappa})[(\eta+1)^{-\hat{\kappa}} - (\eta+2)^{-\hat{\kappa}}]} \quad (55)$$

$$\hat{\xi} = (\eta+1)\beta_{\eta} - \frac{\hat{\alpha}}{\hat{\kappa}} [1 - (\eta+1)^{-\hat{\kappa}} \Gamma(1+\hat{\kappa})] \quad (56)$$

3.4.2 LH- and L-moments for the GLO distribution

The probability density function of the GLO is given by:

$$f(x) = \frac{\alpha^{-1} e^{-(1-\kappa)y - e^{-y}}}{(1+e^{-y})^2}, y = \begin{cases} -\kappa^{-1} \ln \left\{ 1 - \frac{\kappa(x-\xi)}{\alpha} \right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases} \quad (57)$$

The cumulative distribution function of the GLO as provided by Hosking and Wallis (1997) can be written as:

$$F(x) = 1 / (1 + e^{-y}) \quad (58)$$

Equation (58) was rewritten by Shabri (2012) as:

$$F(x) = \left[1 + \left[1 + \left\{ 1 - \frac{\kappa}{\alpha} (x - \xi) \right\}^{\frac{1}{\kappa}} \right] \right]^{-1}, \text{ if } \kappa \neq 0 \quad (59)$$

The inverse of the CDF will give the quantile function of the GLO distribution for the required return periods and is given by Hosking and Wallis (1997) as:

$$x(F) = \xi + \frac{\alpha}{\kappa} \left\{ 1 - \left[\frac{1-F}{F} \right]^{\kappa} \right\}, \kappa \neq 0 \quad (60)$$

$$= \xi - \alpha \ln\{(1 - F)/F\}, \kappa = 0 \quad (61)$$

3.4.2.1 L-Moments

The parameters of the GLO distribution were given using L-moments by Hosking (1990) as:

$$\hat{\kappa} = -\frac{\lambda_3}{\lambda_2} \quad (62)$$

$$\hat{\alpha} = \frac{\lambda_2}{\Gamma(1+\hat{\kappa})\Gamma(1-\hat{\kappa})} \quad (63)$$

$$\hat{\xi} = \lambda_1 + \frac{\lambda_2 - \hat{\alpha}}{\hat{\kappa}} \quad (64)$$

3.4.2.2 LH-Moments

Meshgi and Khalili (2007b) used the LH-moments method in different levels (η) to estimate the parameters of the GLO distribution. In their work, κ , ξ and α values were estimated by Equations (65)-(67) for different levels of the LH-moments:

$$\hat{\kappa} = \frac{(\eta+3)(\eta+2)\beta_{\eta+2} - [(\eta+2)^2 + (\eta+2)(\eta+1)]\beta_{\eta+1} + (\eta+1)^2\beta_{\eta}}{(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_{\eta}} \quad (65)$$

$$\hat{\alpha} = \frac{\Gamma(\eta+2)[(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_{\eta}]}{\Gamma(\eta+1-\hat{\kappa})\Gamma(1+\hat{\kappa})} \quad (66)$$

$$\hat{\xi} = (\eta + 1)\beta_{\eta} - \frac{\hat{\alpha}}{\hat{\kappa}} \left[1 - \frac{\Gamma(\eta+1-\hat{\kappa})\Gamma(1+\hat{\kappa})}{\Gamma(\eta+1)} \right] \quad (67)$$

3.4.3 LH- and L-moments for the GPA distribution

The probability density function of the GPA is given by:

$$f(x) = \alpha^{-1} e^{-(1-\kappa)y}, y = \begin{cases} -\kappa^{-1} \ln \left\{ 1 - \frac{\kappa(x-\xi)}{\alpha} \right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases} \quad (68)$$

The cumulative distribution function of the GPA as provided by Hosking and Wallis (1997) can be written as:

$$F(x) = 1 - e^{-y} \quad (69)$$

Equation (69) is rewritten and given by Maidment (1993) as:

$$F(x) = 1 - \left[1 - \kappa \left(\frac{x-\xi}{\alpha} \right) \right]^{\frac{1}{\kappa}}, \text{ for } > 0 \quad (70)$$

The inverse of the CDF will give the quantile function of the GPA distribution for the required return periods and is given by Hosking and Wallis (1997) as:

$$x(F) = \xi + \frac{\alpha}{\kappa} \{1 - [1 - F]^\kappa\}, \kappa \neq 0 \quad (71)$$

$$= \xi - \alpha \ln(1 - F), \kappa = 0 \quad (72)$$

3.4.3.1 L-Moments

The parameters of this distribution for the case of L-moments were estimated by Hosking (1990) as:

$$\hat{\kappa} = \frac{4}{\hat{\tau}_3 - 3} \quad (73)$$

$$\hat{\alpha} = \lambda_2 [(\hat{\kappa} + 1)(\hat{\kappa} + 2)] \quad (74)$$

$$\hat{\xi} = \lambda_1 + \lambda_2(\hat{\kappa} + 2) \quad (75)$$

3.4.3.2 LH-Moments

Meshgi and Khalili (2007b) estimated the parameters of the GPA distribution using LH-moments for different levels by Equations (76)-(78):

$$\hat{\kappa} = \frac{-5-2\eta+((\eta+3)[(\eta+3)\beta_{\eta+2}-(\eta+1)\beta_{\eta}])/((\eta+2)\beta_{\eta+1}-(\eta+1)\beta_{\eta})}{-1+((\eta+3)\beta_{\eta+2}-(\eta+1)\beta_{\eta})/((\eta+2)\beta_{\eta+1}-(\eta+1)\beta_{\eta})} \quad (76)$$

$$\hat{\alpha} = \frac{\hat{\kappa}\Gamma(\eta+3+\hat{\kappa})\Gamma(\eta+2+\hat{\kappa})[(\eta+2)\beta_{\eta+1}-(\eta+1)\beta_{\eta}]}{(\eta+1)!\Gamma(1+\hat{\kappa})[(\eta+2)\Gamma(\eta+2+\hat{\kappa})-\Gamma(\eta+3+\hat{\kappa})]} \quad (77)$$

$$\hat{\xi} = (\eta + 1)\beta_{\eta} - \frac{\hat{\alpha}}{\hat{\kappa}} \left[1 - \frac{(\eta+1)\Gamma(\eta+1)\Gamma(1+\hat{\kappa})}{\Gamma(\eta+2+\hat{\kappa})} \right] \quad (78)$$

3.5 Steps Used for Regional Flood Frequency Analysis

The procedure suggested by Hosking and Wallis (1997) precisely involves evaluation of three statistical measures using L-moments, which can also be used for LH-moments, which are:

- (i) Discordancy measure (D_i),
- (ii) Heterogeneity measure (H), and
- (iii) Goodness-of-fit measure (Z).

3.5.1 Discordancy Measure

Hosking and Wallis (1997) proposed a discordancy measure (D_i) based on L-moments to recognize those sites that are grossly discordant with the group as a whole. The same discordancy measure is used for different LH-moments, $L_{\eta}, \eta = 1, 2, 3, 4$, that is, from L_1 to L_4 with the required modification of the components of u_i vector in the proposed formula of Hosking and Wallis (1997). The discordancy measure is used to assist in identifying those sites whose LH-moment (L to L_4) ratios are discordant (markedly different) relative to LH-moment (L to L_4) ratios for the collection of sites.

It is important to note that low or high outliers should not be removed from the datasets as these are expected outcome in large samples from multiple sites. The ‘‘apparent’’ outliers are important indicators about the natural variability of the phenomenon and the frequency of occurrence of low or high values. It is therefore

suggested that if no obvious cause of discordant measure, the site(s) shall be kept with the proposed group of sites and delay decision until heterogeneity measure (H) is computed during regional analysis (<http://www.mgsengr.com/LRAP/Download/LMoments.pdf>). In conducting the RFFA, the heterogeneity measure is the primary indicator for accepting or rejecting a proposed region (group of sites). The discordancy measures for the various sites provide a secondary indicator to consider whether a discordant site should be removed to another region.

The discordancy measure is defined as:

$$D_i^\eta = \frac{1}{3}(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}), \quad (79)$$

Where u_i = vector of $LH-C_v$, $LH-C_s$ and $LH-C_k$ for site i ;

S = covariance matrix of u_i ;

\bar{u} = mean of vector u_i .

A given site is declared discordant if $D_i > 3.0$.

3.5.2 Regional Homogeneity (Heterogeneity Measure)

In RFFA, identification of homogeneous regions within the study area is the first step of the analysis. The procedure proposed by Hosking and Wallis (1997) for L-moments (L) with extension of all higher order LH-moments groups (L_1 to L_4) have been used for test of regional homogeneity. The homogeneity measure basically compares the between-site variation in sample LH-moments (L to L_4) for the group of sites with what would be expected for a homogeneous region. For this purpose, the LH-moments (L to L_4) and the parameters of Kappa distribution developed by Meshgi and Khalili (2007a) will be used for 500 simulations in each of the LH-moments $L_\eta, \eta = 0, 1, 2, 3, 4$ that is from L to L_4 .

Suppose a proposed region has N sites, with site i having record length n_i and sample LH-moment ratios $\tau_2^i, \tau_3^i, \text{ and } \tau_4^i$. Let $\tau_2^R, \tau_3^R, \text{ and } \tau_4^R$ be the regional averages of LH-

Cv, LH-skewness and LH-kurtosis; then the weighted proportionally to the sites' record length are given by:

$$\tau_{\eta,2}^R = \frac{\sum_{i=1}^N n_i \tau_2^i}{\sum_{i=1}^N n_i} \quad (80)$$

$$\tau_{\eta,3}^R = \frac{\sum_{i=1}^N n_i \tau_3^i}{\sum_{i=1}^N n_i} \quad (81)$$

$$\tau_{\eta,4}^R = \frac{\sum_{i=1}^N n_i \tau_4^i}{\sum_{i=1}^N n_i} \quad (82)$$

The weighted standard deviation of the at-site sample LH-CVs, LH-skewness and LH-kurtosis are given as:

$$V_{\eta,1} = \left\{ \frac{\sum_{i=1}^N n_i (\tau_2^i - \tau_2^R)^2}{\sum_{i=1}^N n_i} \right\}^{\frac{1}{2}} \quad (83)$$

$$V_{\eta,2} = \left\{ \frac{\sum_{i=1}^N n_i (\tau_2^i - \tau_2^R)^2 + (\tau_3^i - \tau_3^R)^2}{\sum_{i=1}^N n_i} \right\}^{\frac{1}{2}} \quad (84)$$

$$V_{\eta,2} = \left\{ \frac{\sum_{i=1}^N n_i (\tau_3^i - \tau_3^R)^2 + (\tau_{34}^i - \tau_4^R)^2}{\sum_{i=1}^N n_i} \right\}^{\frac{1}{2}} \quad (85)$$

For the simulation purpose, an L-MOMENTS program written in FORTRAN version 3.0-1 written and revised by Hosking (2015) will be basically used to determine the statistical characteristics of homogeneous regions with the same number of stations and the same record lengths as those in the analysis. The program generates a homogeneous region with total number of observations equal to the sum of observations in the stations being analyzed. These will be divided such that each station of the simulated region will have the same number of observations as corresponding stations in the analysis.

Heterogeneity measure, H_I , is an important parameter used for identification of homogeneous region based on observed and simulated dispersion of LH-moments (LH-C_v) for a group of sites under consideration. This can be computed from Equation (86) as:

$$H_1^\eta = \frac{V_1 - \mu_{V_1}}{\sigma_{V_2}} \quad (86)$$

$$H_2^\eta = \frac{V_2 - \mu_{V_2}}{\sigma_{V_2}} \quad (87)$$

$$H_3^\eta = \frac{V_3 - \mu_{V_3}}{\sigma_{V_3}} \quad (88)$$

where, V =weighted standard deviation of τ_2 values,

μ_V, σ_V = the mean and standard deviation of N_{sim} values of V , and

N_{sim} = no. of simulations

A region is declared as “acceptably homogeneous” if $H_i < 1$, “possibly heterogeneous” if $1 \leq H_i \leq 2$, and “definitely heterogeneous” if $H_i > 2$. To avoid committing to a particular two or three-parameter distribution, simulation will be undertaken using a four-parameter Kappa distribution (Hosking and Wallis, 1997 for L-moments; and Meshgi and Khalili, 2007a for LH-moments) and the number of simulation is kept at least at 500 to arrive at reliable estimates of μ_V and σ_V .

In addition to the above, two additional measures viz: H_2 and H_3 based on LH-C_v/LH-C_k and LH-C_s/LH-C_k (L to L₄) distances respectively are also considered. The measure H_2 indicates whether at-site and regional estimates will be close to each other, while H_3 indicates whether the at-site and regional estimates will be in agreement. A large value of H_2 usually indicates a large deviation between regional and at-site estimates, whereas a large value of H_3 indicates a large deviation between at-site estimates and observed data. In practice, the H measure for the observed variability in LH-C_v (L to L₄) has been found to be very useful measure. Conversely the high level of natural variability in sample values of LH-skewness and LH-kurtosis (L to L₄) result in the H_2 and H_3 measures having low discriminatory power. Therefore, the heterogeneity measure H_1 for the level of variability in at-site values of LH-C_v (L to L₄) becomes the de-facto measure for assessing the relative level of heterogeneity for the proposed region (<http://www.mgsengr.com/LRAP/Download/LMoments.pdf>).

Nevertheless, a small amount of heterogeneity does not negate the advantage of regional flood frequency analysis (Cunnane, 1988). Similarly, previous studies by Lettenmaier and Potter (1985), Lettenmaier et al. (1987), Hosking and Wallis (1988) and Potter and Lettenmaier (1987) suggested that even though a region may be moderately heterogeneous, regional analysis will still yield more accurate quantile estimates than at-site analysis.

3.5.3 Goodness-Of-Fit Measure

(i) LH-Moments Ratio Diagram (L to L₄)

The L-moments (L) ratio diagram proposed by Hosking and Wallis (1997) is extended to each LH-moments $L_\eta, \eta = 1, 2, 3, 4$ that is (L₁ to L₄) and used to select a best fitting probability distribution for the region. In LH-moments ratio diagram, the theoretical curves of the three parameter distributions mentioned in Section 2.5 for each level of LH-moments as well as the regional average LH-skewness and LH-kurtosis will be plotted on the same graphs for selecting the best fit distribution. The method of generating the L -skewness versus L -kurtosis curve by plotting polynomial approximations (Hosking and Wallis, 1997, pp. 207-208) is extended to generating the LH-moments curve. Nonetheless, one or more candidates might qualify for the proposed distribution and therefore it is difficult to select the best fitting for the region on the basis of LH-moments ratio diagram only. Hence, the Z -statistic criteria depicted below will be considered for selecting the best fitting distribution.

(ii) Z-Statistic Criteria

Z -statistics criteria for L-moments proposed by Hosking and Wallis (1997) and LH-moments, $L_\eta, \eta = 1, 2, 3, 4$ that is from L₁ to L₄ proposed by Meshgi and Khalili (2007a) are used to select the best fitting distribution among the distribution candidates. For this purpose, as indicated earlier, the LH-moments (L to L₄) of the kappa function will be developed. This measure, Z , judges how well the LH-skewness and LH-kurtosis (L to L₄) of the fitted distribution (simulated) matches the regional average LH-skewness and LH-kurtosis (L to L₄) of the observed data. Z for the chosen distribution Z^{DIST} is defined by:

$$Z_{\eta}^{DIST} = \frac{\tau_4^{DIST} - \tau_4^R + \beta_4}{\sigma_4} \quad (89)$$

where, τ_4^R = regional average LH-kurtosis value computed from the data of a given region,

$$\beta_4 = \text{bias of } \tau_4^R,$$

τ_4^{DIST} = average LH-kurtosis value computed from simulation for a fitted distribution,

$$\sigma_4 = \text{standard deviation of LH-kurtosis value (from simulation).}$$

The bias of τ_4^R is given by:

$$\beta_{\eta,4} = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} (\tau_4^m - \tau_4^R) \quad (90)$$

The standard deviation of α_4 is given by:

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \{ \sum_{m=1}^{N_{sim}} (\tau_4^m - \tau_4^R)^2 - N_{sim} \beta_4^2 \} \right]^{1/2} \quad (91)$$

where, τ_4^m = regional average LH-kurtosis value computed from simulation.

The values of $|Z_{\eta}^{DIST}|$ -statistics for the candidate distributions based on L, L₁, L₂, L₃ and L₄-moments can be calculated; and hence, a given distribution is declared a good fit if $|Z_{\eta}^{DIST}| \leq 1.64$. While a number of distributions may qualify for the goodness-of-fit measure criteria, the most potential will be the one that has the minimum $|Z_{\eta}^{DIST}|$ value.

3.5.4 Flood Quantile Estimates for the Region

The index-flood procedure proposed by Dalrymple (1960) will be used for quantile estimates in this study of RFFA approach.

Index-Flood Procedure

The key assumption of an index-flood procedure is that the sites form a homogeneous region, that is, that the frequency distributions of the N sites are identical apart from a

site-specific scaling factor, the *index flood* (Hosking and Wallis, 1997). In other words, when the distribution functions are scaled with their respective index-flood, the resulting dimensionless of all sites in the region can be assumed to have the same shape, which is independent of drainage area and of any other site characteristics (Brutsaert, 2005). According to Hosking and Wallis (1997), frequency distributions at different sites are identical apart from a scale factor, *the index-flood*. Accordingly, this method comprises of two components:

(1) A regional flood frequency curve

The index-flood procedure proposed by Dalrymple (1960) will be used for quantile estimates $\hat{Q}_i(F)$ at site i , which is given as:

$$\hat{Q}_i(F) = \hat{\lambda}_{1,i}^\eta \hat{q}^\eta(F), \eta = 0, 1, 2, 3, 4, \dots \quad (92)$$

where, $\hat{\lambda}_{1,i}^\eta$ is the index-flood for site i and $\hat{q}^\eta(F)$ is the regional quantile estimates or regional growth curve common to every site.

To derive this curve, first the flood distribution curve of each stream flow gauging site in the region is made dimensionless, that is normalized, by dividing the flow rates by the index-flood of the site. The regional flood frequency curve is then constructed as the average curve of the available dimensionless curves.

The procedure for developing the regional flood frequency curve is as follow:

First four LH-moments (L to L₄) computed at each station from among the set of stations qualifying the homogeneity tests are made dimensionless by dividing them with their respective λ_1^η values. Weighted values of these dimensionless LH-moments (L to L₄) are used to obtain the regional standard LH-moments (L to L₄) estimates ($\hat{\lambda}_{r,\eta}^R$).

$$\hat{\lambda}_{r,\eta}^R = \frac{\sum_{i=1}^N n_i \hat{\lambda}_{r,\eta}^i}{\sum_{i=1}^N n_i} \quad (93)$$

where, $\hat{\lambda}_{r,\eta}^R$ = Regional standardized LH-moments (L to L₄) of order r ,

n_i = No. of years of observation at station i ,

$\hat{\lambda}_{r,\eta}^i$ = Standardized LH-moments (L to L₄) of order r at station i ,

N = Number of homogeneous stations in the identified region

Substituting Equation (93) into $\hat{q}^\eta(F)$ of Equation (92), it will give the estimated regional flood frequency curve of Equation (94) as:

$$\hat{q}^\eta(F) = \hat{q}(F; \hat{\lambda}_{r,\eta}^R), \eta = 0, 1, 2, 3, 4, \dots \quad (94)$$

Equation (92) is used to calculate the quantiles for each site once the regional flood frequency curve is developed according to Equation (94).

The RFF curve (regional growth curve) is the curve of the quantile versus the probability of non-exceedance. For the plotting purpose, the probability of non-exceedance can be written in terms of Gumbel reduced Variate as given by:

$$y = -\ln[-\ln F(x)] \quad (95)$$

$F(x)$ is the probability of non-exceedance as provided in Equation (9).

Substituting Equation (7) into Equation (95), we will have:

$$y_T = -\ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \quad (96)$$

where, y_T = reduced Gumbel Variate for return period T .

In general, for a given return period, the corresponding flood quantile can be determined from the regional growth curve. Firstly, the reduced Gumbel variate for the given return period is determined using Equation (96), and for the particular reduced variate, the standardized flood quantile can be read off the growth curve. To change this to the actual flood value as provided in Equation (92), the standardized quantile is multiplied by the mean of discharge at the site of interest.

(2) A relationship between the magnitude of the index-floods, catchment characteristics:

The two major catchment characteristics variables which are responsible to floods generation are the drainage area and rainfall. In this case, floods information will be transferred from gauged to ungauged sites within a pre-specified hydrologically homogeneous region. The end products of the analysis of the available flow data are a dimensionless regional frequency curve and a regression equation relating the index event with drainage area and rainfall. These three relationships can then be used to predict the frequency curve for any ungauged sites. Here the index event is first estimated from the drainage area and rainfall of the ungauged sites and a regional flood frequency curve will be used to estimate quantiles for ungauged sites.

The linear multiple regression equation is given by:

$$\hat{\lambda}_1^{\eta} = c + aA + bR \quad (97)$$

where, $\hat{\lambda}_1^{\eta}$ = Dependent variable (mean discharge of ungauged sites),

A = Independent variable (catchment area),

R = Independent variable (Mean annual rainfall),

c = Regression constant (intercept),

a = Regression coefficient – slope (area coefficient),

b = Regression coefficient – slope (rainfall coefficient)

Using the least square method, the discharge is regressed upon area and rainfall to determine the slope and the intercept of the multiple regression line, and a relationship between these and the catchment area and rainfall is used to determine the mean or index-flood discharge of an ungauged site. The relationship between the ungauged index-flood ($\hat{\lambda}_{1,i}^{\eta}$), catchment area (A) and rainfall (R) in log form is given by:

$$\hat{\lambda}_{1,i}^{\eta} = cA^aR^b \quad (98)$$

By substituting Equation (98) into Equation (92), it will give the quantiles estimate for ungauged sites. Therefore, for any given catchment within the homogeneous region, flood quantiles for a given return period can be determined by the use of the regional growth curve. Since, the value of discharge determined from the regional growth curve is a standardized value, it is to be multiplied by the $\hat{\lambda}_{1,i}^{\eta}$ (average flood value). For gauged catchments, $\hat{\lambda}_{1,i}^{\eta}$ can be estimated from the available flood data, while for ungauged catchments, $\hat{\lambda}_{1,i}^{\eta}$ can be found using Equation (97) or Equation (98).

3.6 Comparative Study of the Methods

A comparative study will be performed between regional flood frequency analysis by L-moments and LH-moments (selected level) to select the appropriate order of LH-moments via comparison and analysis of the frequency of observed and estimated annual maximum floods. For this purpose, the relative Root Mean Square Error (RMSE) as suggested by Hosking and Wallis (1997) will be used to evaluate the precision of the methods of parameter or of quantile estimates of the selected distributions using L-moments and LH-moments. The relative RMSE can be expressed as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - x_{F_i}}{x_i} \right)^2} \quad (99)$$

where, x_i = order set observed values and x_{F_i} = computed observation values for a given value of F_i . For a complete data series ($0 \leq F \leq 1$), the non-exceedance probabilities (F_i) will be determined based on sample size data and Cunnane's (1980) plotting position formula given by:

$$PP = F_i = \frac{i-0.4}{n+0.2} \quad (100)$$

where PP = plotting position

i = rank number

n = total number of observations

The steps for undertaking the PP analysis are as follow:

1. Arrange the observed data in an ascending order.
2. Assign rank to each of these observations as said in step 1.
3. Using the rank number (say i), compute the PP for each observation using Equation (100).
4. Using the values of non-exceedance as calculated in step 3, calculate the estimated quantiles using the selected distributions of the L-moments and LH-moments.
5. Plot observed flood quantiles versus estimated flood quantiles.

Using Equation (99), the RMSE for the region will be calculated for both L-moments and LH-moments (selected level) for the selected distributions. Finally the distribution plus the method that gives the minimum RMSE of the L-moments and LH-moments will be considered as the best parameter and quantile estimation method.

CHAPTER 4

4 DATA ANALYSIS AND DISCUSSION OF RESULTS

4.1 Data Used

The annual maximum floods observed at 13 gauged sites in the Limpopo catchment have been chosen for the study. The details and location of these stations which are well spread across the Limpopo region have been given in Table 4.1 and Figure 1.2. The observed data with more than 10 years of record have been considered for the study.

Table 4.1: The sites of the study area and their corresponding information

Name of the River/Stream	Name of the Station	Station No.	Latitude (°)	Longitude (°)	Catchment Area (km ²)	Period of Observation	Average Rainfall (mm)
Metsimotlhabe	Morwa	2411	24.45	26.07	3400	1975-1996 (22 years)	525
Metsimotlhabe	Thamaga Bridge	2421	24.7	25.55	982	1983-2001 (19 years)	525
Kolobeng	Kumakwane	2511	24.68	25.67	120	1978-2002 (25 years)	550
Bonwapitse	Ntswaneng	3111	23.27	24.73	1125	1985-2002 (18 years)	450
Tautswe	Bodungwe Hill	3121	23.25	27.07	630	1985-1996 (12 years)	425
Mahalaptswa	Madiba	3221	23.05	26.82	840	1971-1994 (24 years)	430
Lotsane	Maunatlala	3321	22.58	27.63	6385	1979-2007 (29 years)	500
Lotsane	Palapye	3331	22.55	27.20	3815	1971-1995 (25 years)	450
Motloutse	Tobane	4121	21.88	28.02	8400	1969-1996 (28 years)	425
Shashe	Lower Shashe	4321	21.55	27.98	7810	1970-2000 (31 years)	410
Shashe	Mooke Weir	4361	21.48	27.35	2500	1968-2002 (35 years)	430
Ntse	Ntse Weir	4411	21.08	27.58	800	1970-2008 (39 years)	425
Tati	Tati Weir	4511	21.03	27.43	570	1970-2008 (39 years)	450

4.2 Analysis with L-Moments

4.2.1 L-moments Statistics

First the probability weighted moments (PWMs) using Equation (31) have been calculated for each station as tabulated in Table 4.2.

Table 4.2: Values of Probability Weighted Moments using L-moments

Name of Site	Station No.	Sample Size	β_0 (mean)	β_1	β_2	β_3
Metsimothabe	2411	22	32.7	26.0	21.9	19.1
Metsimothabe	2421	19	38.5	29.6	24.3	20.8
Kolobeng	2511	25	52.7	47.7	43.9	40.8
Bonwapitse	3111	18	15.1	13.3	12.1	11.2
Tautswe	3121	12	41.6	37.8	35.6	34.0
Mahalaptswa	3221	24	3.8	3.0	2.5	2.2
Lotsane	3321	29	42.4	36.0	32.3	29.7
Lotsane	3331	25	41.6	35.9	32.5	30.3
Motloutse	4121	28	175.1	122.6	97.0	81.4
Shashe	4321	31	390.4	301.3	250.8	216.8
Shashe	4361	35	168.6	134.3	116.4	105.1
Ntse	4411	39	56.1	42.9	36.4	32.4
Tati	4511	39	79.1	61.7	53.2	47.9

Since the L-moments are related to PWMs, the first four L-moments and L-moment ratios have been calculated using Equations (32-35) as shown in Table 4.3.

Table 4.3: Sample L-moments and L-moment ratios

Name of Site	Station No.	Sample Size	Average Flow (m ³ /s)	λ_1	λ_2	λ_3	λ_4	L-C _v (τ_2)	L-Skewness (τ_3)	L-Kurtosis (τ_4)
Metsimothabe	2411	22	32.712	32.712	19.194	8.105	4.777	0.587	0.422	0.249
Metsimothabe	2421	19	38.549	38.549	20.557	7.320	2.250	0.533	0.356	0.109
Kolobeng	2511	25	52.674	52.674	42.743	29.806	19.755	0.811	0.697	0.462
Bonwapitse	3111	18	15.062	15.062	11.441	7.996	6.335	0.760	0.699	0.554
Tautswe	3121	12	41.562	41.562	34.103	28.116	25.457	0.821	0.824	0.746
Mahalaptswa	3221	24	3.823	3.823	2.128	1.081	0.770	0.557	0.508	0.362
Lotsane	3321	29	42.403	42.403	29.664	20.096	14.519	0.700	0.677	0.489
Lotsane	3331	25	41.598	41.598	30.276	21.250	18.540	0.728	0.702	0.612
Motloutse	4121	28	175.050	175.050	70.218	21.135	14.162	0.401	0.301	0.202
Shashe	4321	31	390.358	390.358	212.311	87.209	37.164	0.544	0.411	0.175
Shashe	4361	35	168.586	168.586	100.002	61.264	52.752	0.593	0.613	0.528
Ntse	4411	39	56.115	56.115	29.680	17.355	12.802	0.529	0.585	0.431
Tati	4511	39	79.135	79.135	44.339	28.104	23.051	0.560	0.634	0.520

As explained in Sections 2.4.2 and 2.4.3, the primary use of PWMs (and related L-moments) is in the estimation of parameters for a probability distribution. However, the results of Table 4.2 have been used to calculate the sample L-moments of Table 4.3, which will help for the tests of whether the Limpopo region is homogeneous as

proposed by Hosking and Wallis (1997). In their work, most of the tests involve combination of the L-CV, L-skewness and L-kurtosis, and as explained also in detail in Section 3.5 of this study. It shall be noted that in a homogeneous region all sites have the same population L-moment ratios; however, their sample L-moment ratios could be different owing to sampling variability as it can be seen in Table 4.3 and will be discussed in Section 4.2.3.

4.2.2 Discordancy Measure

The discordancy test has been carried out using equation (79) for all sites and the discordancy values are given in Table 4.4.

Table 4.4: L-moment ratios and discordancy measures

Name of Site	Station No.	Sample Size	λ_1 (m ³ /s)	L-CV	L-Skewness	L-Kurtosis	D_i
Metsimotlhabe	2411	22	32.7	0.587	0.422	0.249	0.89
Metsimotlhabe	2421	19	38.5	0.533	0.356	0.109	1.18
Kolobeng	2511	25	52.7	0.811	0.697	0.462	1.51
Bonwapitse	3111	18	15.1	0.760	0.699	0.554	0.46
Tautswe	3121	12	41.6	0.821	0.824	0.746	1.43
Mahalaptswa	3221	24	3.8	0.557	0.508	0.362	0.11
Lotsane	3321	29	42.4	0.700	0.677	0.489	0.57
Lotsane	3331	25	41.6	0.728	0.702	0.612	0.73
Motloutse	4121	28	175.1	0.401	0.301	0.202	2.08
Shashe	4321	31	390.4	0.544	0.411	0.175	0.82
Shashe	4361	35	168.6	0.593	0.613	0.528	0.52
Ntse	4411	39	56.1	0.529	0.585	0.431	1.52
Tati	4511	39	79.1	0.560	0.634	0.520	1.18
Weighted means				0.607	0.565	0.413	

It can be seen from Table 4.4 that the D_i values for all sites in the region are less than the critical value 2.869 (given by Hosking and Wallis, 1997) for the thirteen sites in the study area; and thus there is no discordant site found in this region.

The plots of sample cumulative variance and skewness of the L-moments are shown in Figure 4.1.

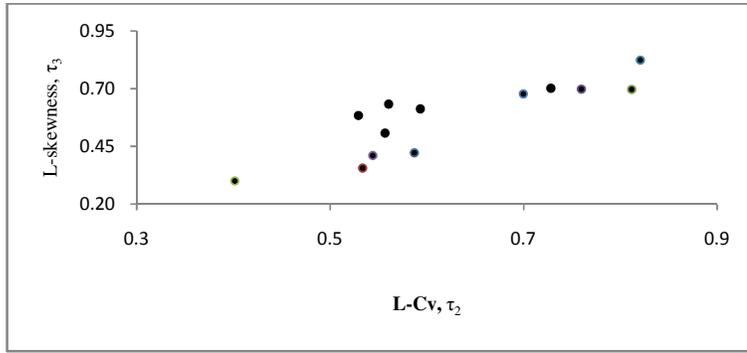


Figure 4.1: Plots of $L-C_v$ versus L-Skewness

Figure 4.1 illustrates the L-moments diagram (plots of the $L-C_v$ versus the $L-C_s$ of the observed data) under the assumption of one homogeneous region. As indicated in this figure, rather considerable dispersion levels are observed in the computed $L-C_v$ and $L-C_s$ values. The L-moments diagram of Figure 4.1 helps in visual identification on whether the sites of a homogeneous group provide similar $L-C_v$ and $L-C_s$ values. Hence, it has been found that further evaluation of the study area using heterogeneity tests is required.

4.2.3 Regional Homogeneity

An L-MOMENT program written in FORTRAN version 3.0-1 written and revised by Hosking (2015) has been used for simulations to determine the statistical characteristics of homogeneous regions. In line with this, initially the entire catchment was assumed as one homogeneous region and homogeneity evaluations were performed for the L-moment. Then, simulation with the kappa distribution was performed to conduct the H_i tests. The heterogeneity measures H_i , ($i=1, 2, 3$) were calculated for L-moments based on the simulations of the observed standard deviation (SD) of $L-C_v$, simulated mean of SD of $L-C_v/L-C_s$ and simulated standard deviation of SD of $L-C_v/L-C_k$ and 500 sets of simulated data. As discussed in Section 3.5.2, the heterogeneity measure (H_1) which is related to $L-C_v$ is most important and has larger effect than variation in $L-C_s$ or $L-C_k$ for assessing the relative level of heterogeneity for the proposed region. On the other hand, H_2 and H_3 which are related to $L-C_s$ and $L-C_k$ have low discriminatory power as they are influenced by sample size and

presence of extraordinary flows. The results of the heterogeneity tests are presented in Table 4.5.

Table 4.5: Heterogeneity measure for L-moments

L-moments	H_1	H_2	H_3
	1.27	0.73	0.28

From Table 4.5, it can be observed that the H_1 value of 1.27 is slightly greater than the criteria value of “acceptably –homogeneous” value, that is 1, and is considered as “possibly-homogeneous”. On the other hand the values of H_i , ($i=2, 3$) are less than 1 and identified as the homogeneous region. Since, the value of H_1 is not far from the homogeneous criterion value of 1, the whole Limpopo region can be considered as homogeneous.

4.2.4 Choice of Suitable Distribution

4.2.4.1 L-Moments Ratio Diagram

In the L-moments ratio diagram (Figure 4.2), the theoretical curves of GEV, GLO and GPA distributions as well as the regional averages L-skewness and L-kurtosis are plotted for selecting the suitable fit distribution.

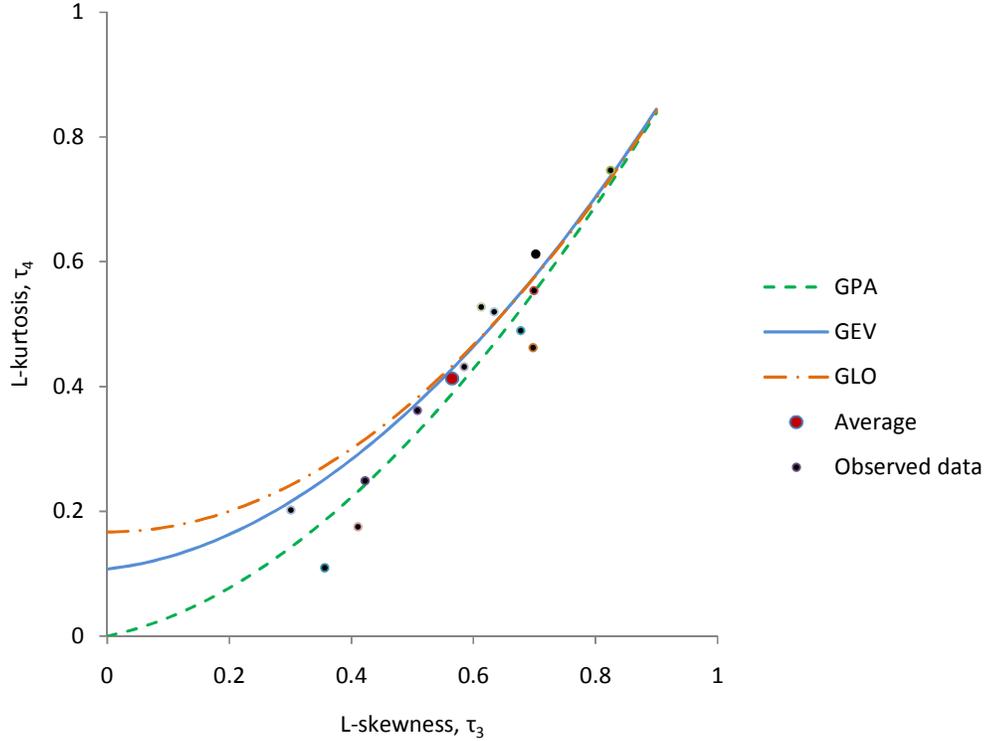


Figure 4.2: L-moments ratio diagram

With close observation, it can be seen from Figure 4.2 that L-moment identifies GEV distribution as the suitable fit distribution for the Limpopo region flood data. The L-moments ratio diagram can also show that there is a better scattering of data around all the three distributions. Therefore, it is difficult to select the best fitting for a region on the basis of L-moment ratios diagram only. Hence, the Z-statistic criteria should be considered for selecting the best fitting distribution.

4.2.4.2 Z-Statistics Criteria

As indicated in Section 3.5.2.1 of item II, the Z^{DIST} requires evaluation of τ_4^{DIST} , which has been obtained by performing simulation with the kappa distribution. The values of |Z|-statistics for the GEV, GLO and GPA based on the L-moments have been calculated and tabulated in Table 4.6.

Table 4.6: |Z|-Statistics values

	GLO	GEV	GPA
L-moments	-1.32	-1.23	-2.1

It can be seen from Table 4.6 that values of |Z|-statistic for GEV and GLO distributions are less than the critical value 1.64 and qualify for distribution candidates. However, the value for GEV distribution is the smallest and GEV has been identified as possibly the true regional distribution.

4.2.5 Parameters Estimation

Based on the goodness-of-fit criteria discussed above, the GEV using L-moments is identified as the best fitting distribution among the three distributions. Accordingly, Equations (49-52) for GEV based on the regional average L-moment ratios τ_2^R and τ_3^R together with $\hat{\lambda}_1$ are used for estimates of the parameters of the GEV distribution and the results are tabulated in Table 4.7.

Table 4.7: Regional parameters of the distributions for L-moments

	Distributions	Parameters		
		Location	Scale	Shape
L-moments	GEV	0.413	0.400	-0.479

4.2.6 Flood Quantiles Estimation and Development of Regional Growth Curves

The quantiles of the regional growth curves of GEV distribution are calculated for the L-moments by using values of the regional parameters given in Table 4.7 and Equation (47) and the results are presented in Table 4.8.

Table 4.8: Quantiles estimates for regional growth curves for L-moments at various probabilities of non-exceedance using GEV distribution

Reccurence Interval, T (Years)	2	10	20	50	100	1000
Probability of non-exceedance (F)	0.5	0.9	0.95	0.98	0.99	0.999
Gumbel Reduced Variate (Y)	0.37	2.25	2.97	3.90	4.60	6.91
Estimated Standardized Quantiles, L-GEV, $x(F)$	0.57	2.03	3.04	5.00	7.15	22.46

For the study region, the estimated standardized quantiles at specified recurrence intervals (Gumbel Reduced Variates) have been computed and the regional growth curve developed for the L-moments as shown in Figure 4.3.

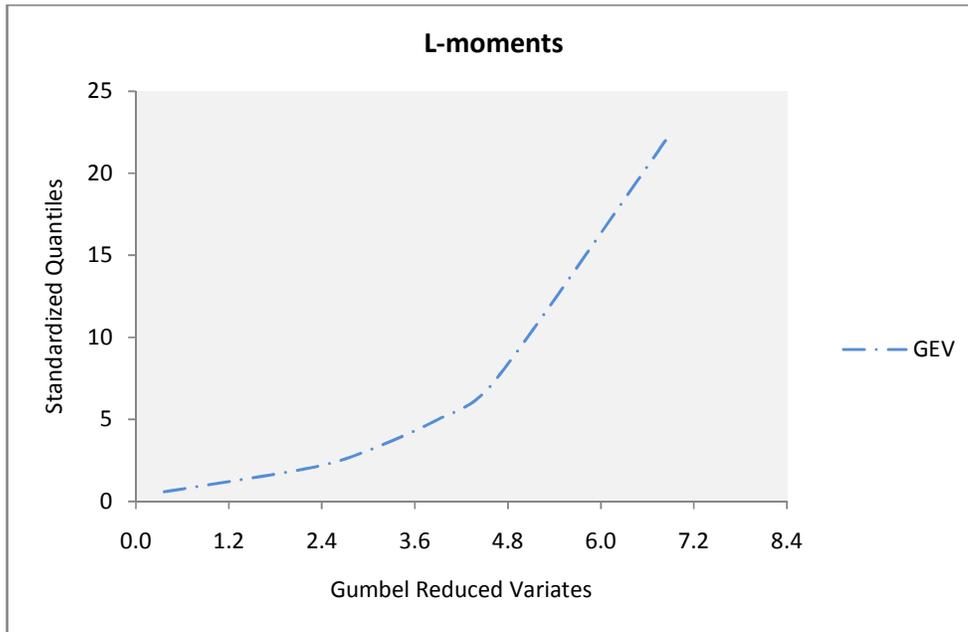


Figure 4.3: Regional growth curve using L-moments

Hence, using the regional growth curve of Figure 4.3 and the mean annual maximum flood in Table 4.2, the design flood for the site of interest can be estimated using the prediction Equation (92) for desired recurrence intervals.

4.2.7 Comparison of Observed and Estimated Floods

To assess the descriptive ability of the suggested distribution using L-moments, the relative Root Mean Square Error (RMSE) has been calculated using Equation (99) and the results are presented in Table 4.9.

Table 4.9: Root Mean Square Error (RMSE) using L-moments

Sl. No.	Name of the River/Stream	Station No.	Observed Mean Flow (m ³ /s)	Estimated Mean Flow (m ³ /s)	RMSE*
1	Metsimotlhabe	2411	32.71	28.94	0.96
2	Metsimotlhabe	2421	38.55	33.79	0.21
3	Kolobeng	2511	52.67	46.97	11.08
4	Bonwapitse	3111	15.06	12.97	1.71
5	Tautswe	3121	41.56	35.19	7.40
6	Mahalaptswa	3221	3.82	3.40	0.21
7	Lotsane	3321	42.40	38.13	0.60
8	Lotsane	3331	41.60	37.09	1.44
9	Motloutse	4121	175.05	157.09	0.36
10	Shashe	4321	390.36	352.24	0.24
11	Shashe	4361	168.59	153.06	0.22
12	Ntse	4411	56.11	52.56	0.23
13	Tati	4511	79.13	74.13	0.29
Average RMSE					1.92

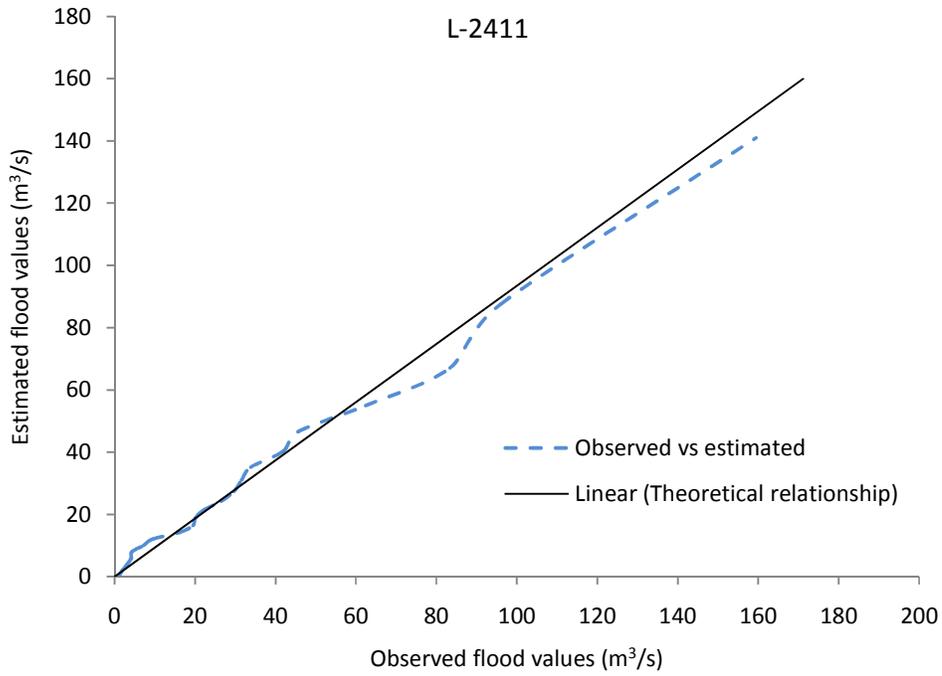
*RMSE calculated based on all the observed and computed values at each station.

Table 4.9 shows the RMSE values for each site and the regional average using the L-moment procedure. It is important to see that the RMSE value for Kolobeng is distinctively too high. However, the D_i for this site is 1.51 (Table 4.4) which is less than the critical value, 3, which has also fairly high L-CV as shown on the same table. It might be worthwhile to remove this site from Limpopo region if there are physical grounds for doing so, but there is no evidence of gross error in the data.

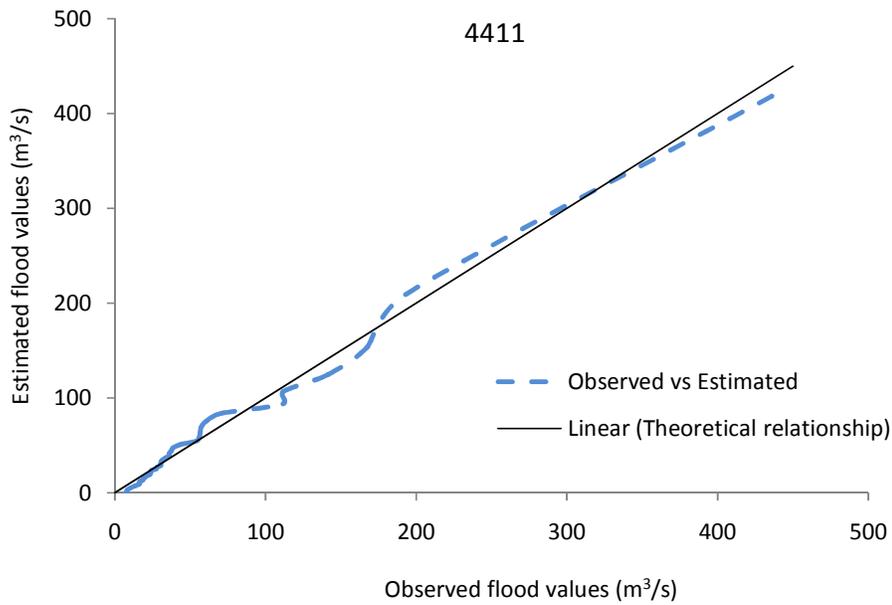
The regional average RMSE measures the overall deviation of estimated quantiles from true quantiles. It is the criterion to which it is given most weight in judging whether the estimation procedure is superior to another. Accordingly, the regional average RMSE using the L-moment will be compared with the regional average RMSE using the LH-moment (L_2) in Section 4.4 to conclude the suitability of the methods.

Similarly, graphical comparisons between the observed and estimated flood values at corresponding probabilities of non-exceedance, at different sites of the region have been undertaken. For this, Cunnane's (1980) plotting position formula (Equation 100)

has been used in order to maintain unbiasedness and minimum variance in computation of probabilities of non-exceedance. Examination of the observed and estimated floods at different sites showed that a good agreement, in general, can be seen from two typical plots as given in Figures 4.4a and b.



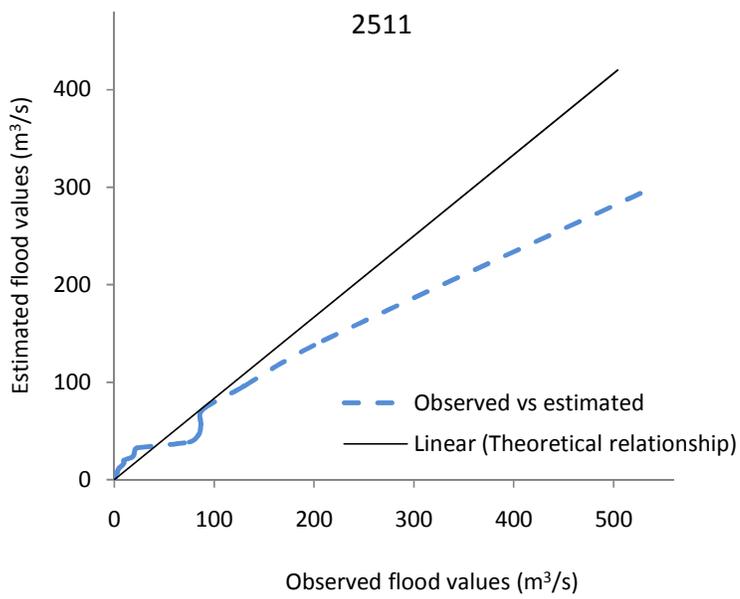
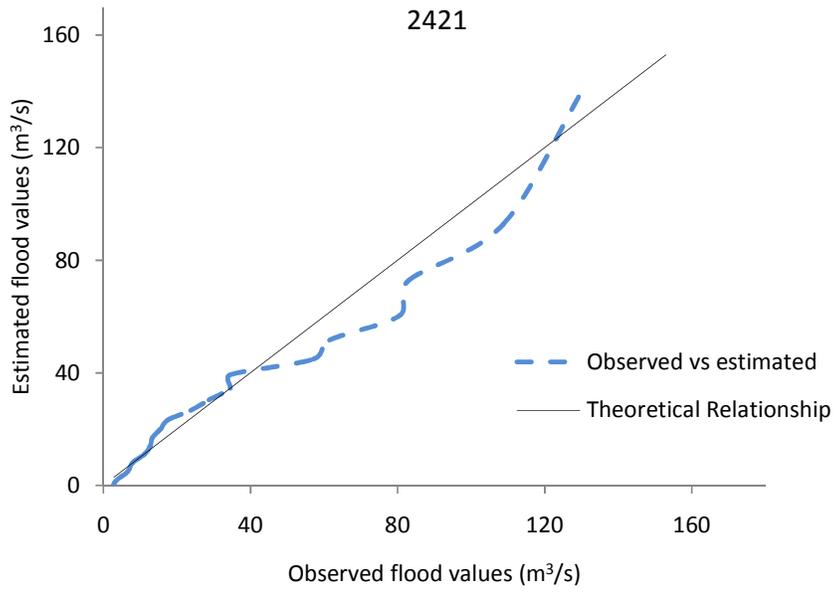
(a)

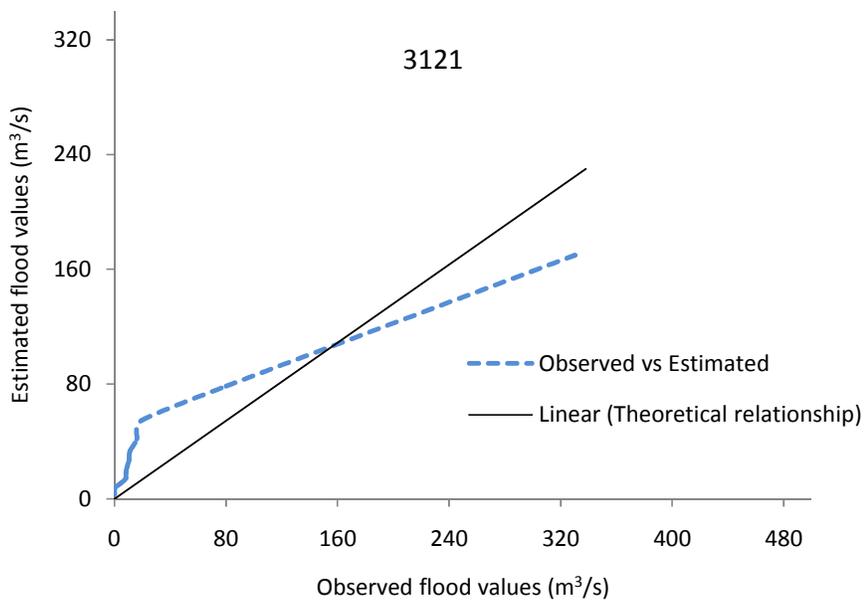
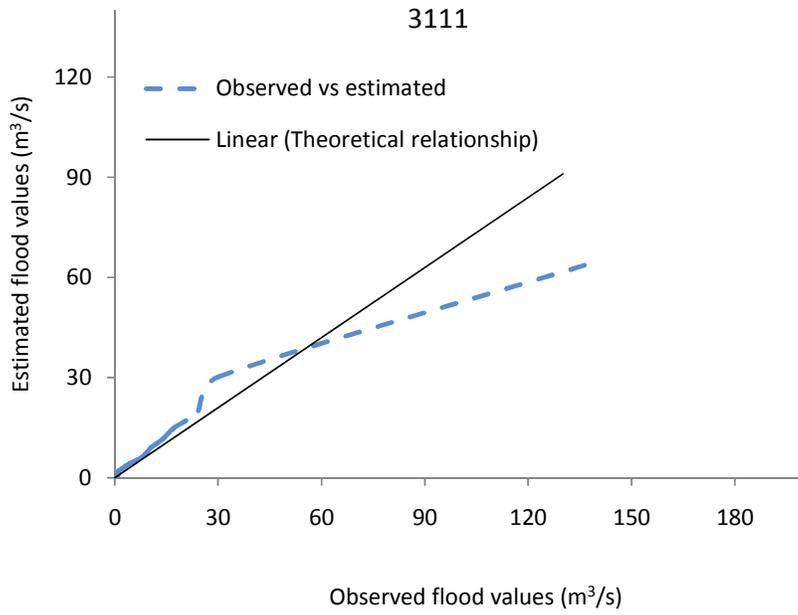


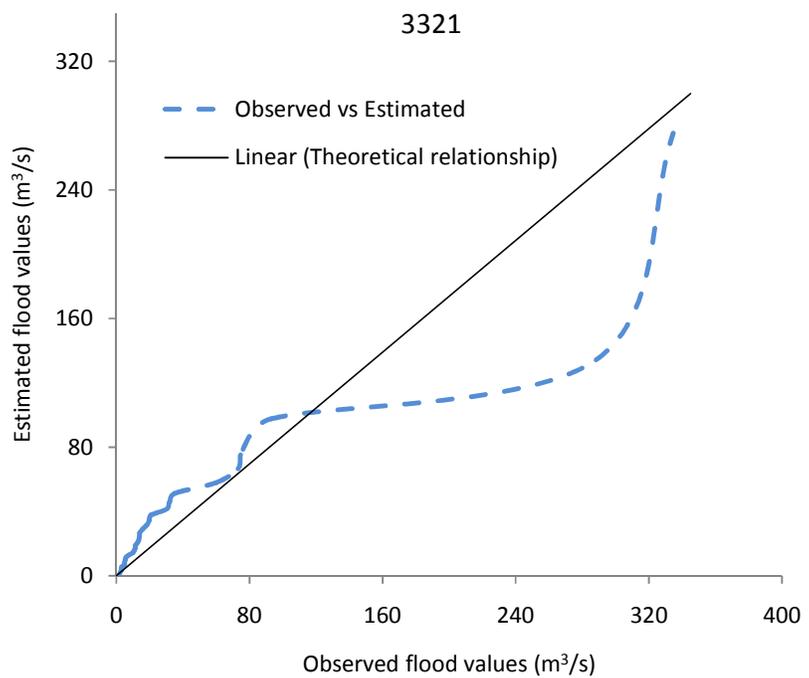
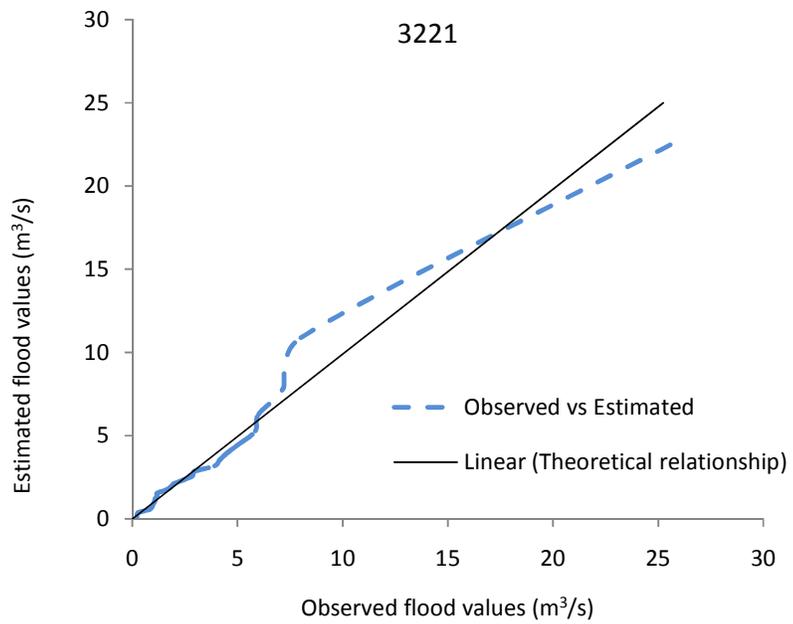
(b)

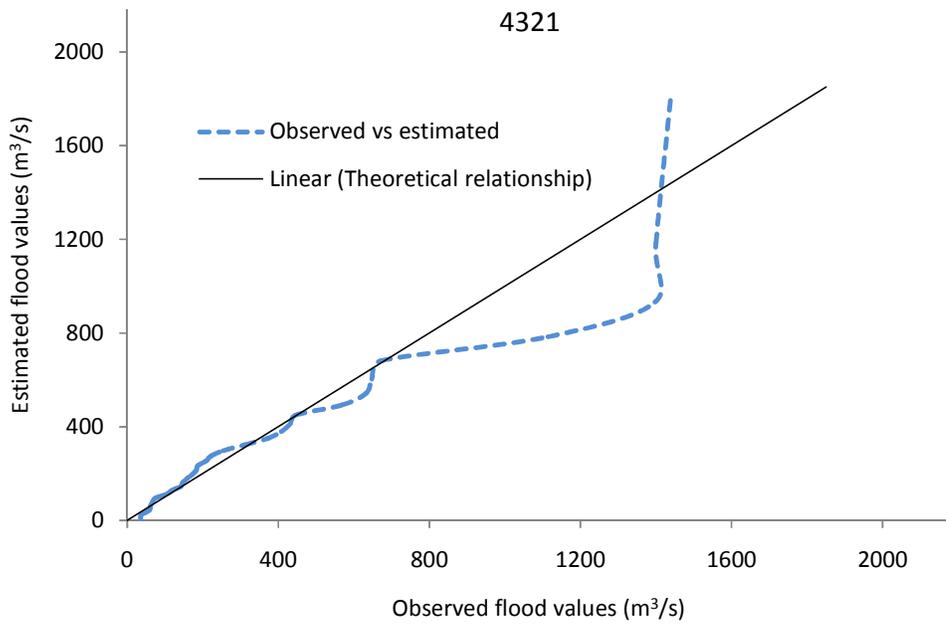
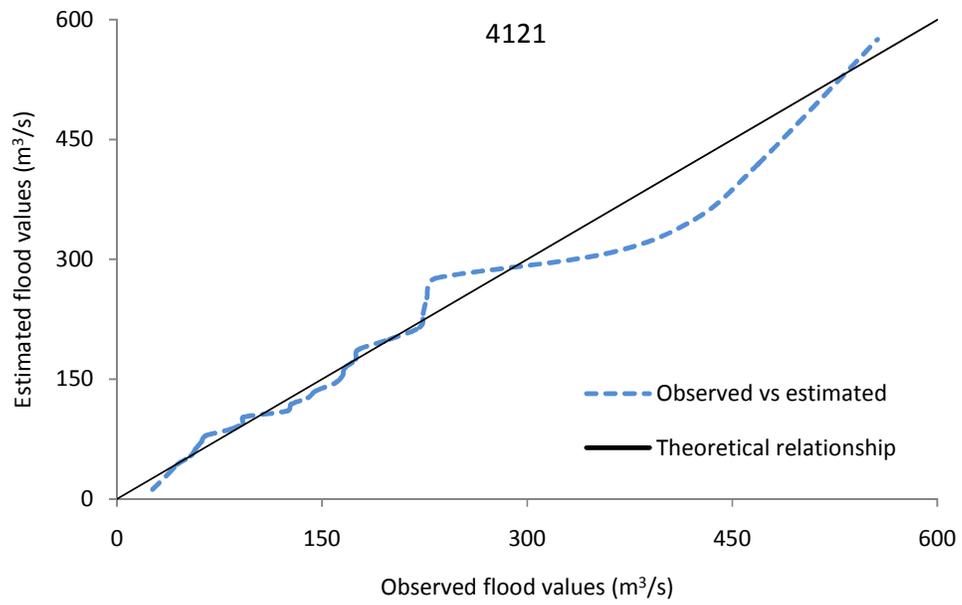
Figure 4.4: Plot of observed and estimated floods at gauging stations No. 2411 and 4411, respectively

The same plots for other sites have been provided in Figure 4.5.









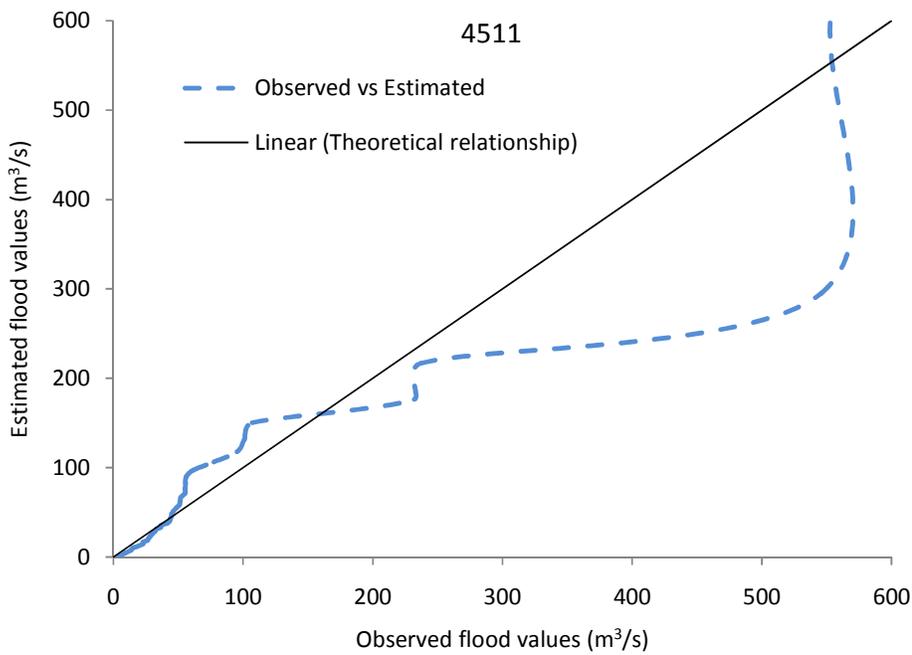
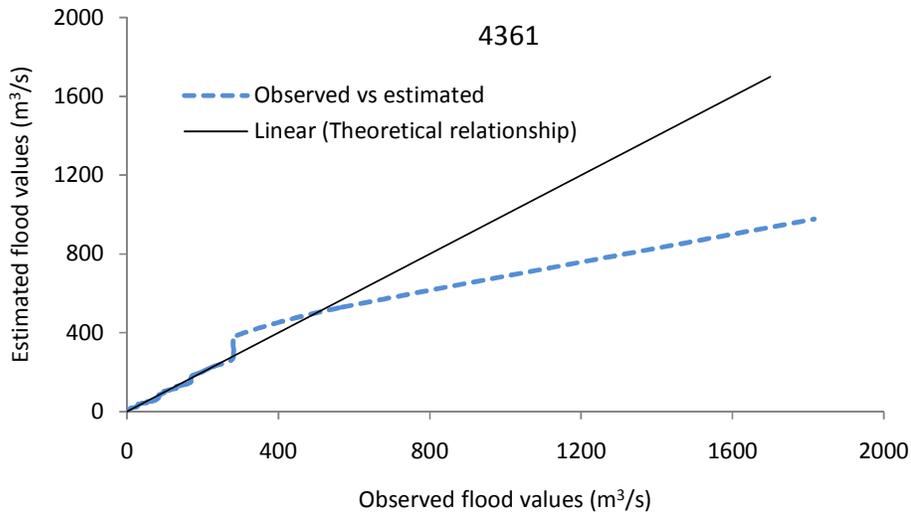


Figure 4.5: Plot of observed and estimated floods at gauging stations No. as shown

It can be seen from Figure 4.5 that observed floods and estimated floods using L-moment for some sites at some record periods are not in agreement. In line with Section 1.2, this could be attributed due to mainly the measurement and extreme

sampling error of the flows at those sites. The same situation was also generally indicated in NWMPR (2006) and Farquharson (1992) as a caution. Hence, the estimated flows at large return periods may not be reliable. Accordingly, special attention shall be given when a flood of higher return period is considered.

4.2.8 Summary

The Regional Flood Frequency Analysis based on the GEV distribution using the L-moments can be summarized as follow:

1. In the step of initial screening of the data, the discordancy measure is used for L-moments. The discordancy measure shows that data of all gauging sites of this study area are suitable for using regional flood frequency analysis by the L-moments.
2. As a second step, the heterogeneity test was undertaken. The heterogeneity measure shows that the Limpopo region has been found to be homogeneous.
3. The regional flood frequency analysis in this study was performed for L-moments by using the three parameters distributions viz: GLO, GEV and GPA. The L-moment ratios diagram has been used to identify the suitable distribution for the region. Again the $|Z|$ -statistic criteria have been used to identify the best suitable distribution for the study area. From Table 4.6, it is observed that the $|Z|$ -statistic value of GEV distribution in L-moment is the smallest among all the $|Z|$ -statistic value. Therefore, the GEV distribution with L-moment is identified as most suitable for regional flood frequency analysis of the Limpopo region.
4. Based on the GEV regional parameters, the regional quantile estimates with non-exceedance probability F have been calculated as presented in Table 4.8 and a regional growth curve has been developed as shown in Figure 4.3.
5. The Relative Mean Square Error (RMSE) has been used to compare the observed and estimated floods to assess the descriptive ability of the suggested distribution using L-moments. According to the result, it has been found that the regional average RMSE is 1.92.

4.3 Analysis with LH-Moments

4.3.1 LH-Moments Statistics

First the probability weighted moments (PWMs) using Equation (38) have been calculated for each station as tabulated in Table 4.10.

Table 4.10: Values of Probability Weighted Moments using LH-moments

Name of Site	Station No.	Sample Size	β_0 (mean)	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Metsimothabe	2411	22	32.7	26.0	21.9	19.1	17.1	15.5	14.3	13.3
Metsimothabe	2421	19	38.5	29.6	24.3	20.8	18.2	16.3	14.7	13.4
Kolobeng	2511	25	52.7	47.7	43.9	40.8	38.3	36.2	34.5	32.9
Bonwapitse	3111	18	15.1	13.3	12.1	11.2	10.6	10.1	9.7	9.3
Tautswe	3121	12	41.6	37.8	35.6	34.0	32.8	31.8	31.0	30.2
Mahalaptswa	3221	24	3.8	3.0	2.5	2.2	2.0	1.9	1.7	1.6
Lotsane	3321	29	42.4	36.0	32.3	29.7	27.7	26.0	24.6	23.4
Lotsane	3331	25	41.6	35.9	32.5	30.3	28.6	27.3	26.3	25.5
Motloutse	4121	28	175.1	122.6	97.0	81.4	70.7	62.9	56.8	51.9
Shashe	4321	31	390.4	301.3	250.8	216.8	191.8	172.3	156.7	143.7
Shashe	4361	35	168.6	134.3	116.4	105.1	97.1	91.2	86.4	82.6
Ntse	4411	39	56.1	42.9	36.4	32.4	29.5	27.2	25.5	24.0
Tati	4511	39	79.1	61.7	53.2	47.9	44.1	41.2	38.8	36.8

Since the LH-moments (L_1 to L_4) are related to PWMs, the first four LH-moments and LH-moment ratios have been calculated using Equation (37) and Equations (40-42), respectively, as shown in the following tables (Table 4.11 to Table 4.14).

Table 4.11: Sample LH-moments and LH-moment ratios for L_1

Name of Site	Station No.	Sample Size	λ_1^1	λ_2^1	λ_3^1	λ_4^1	L_1-C_V (τ_2)	L_1 - Skewness (τ_3)	L_1 - Kurtosis (τ_4)
Metsimothabe	2411	22	26.0	26.6	9.6	5.4	1.026	0.361	0.204
Metsimothabe	2421	19	29.6	28.7	7.7	1.8	0.972	0.268	0.064
Kolobeng	2511	25	47.7	60.1	35.6	22.4	1.260	0.592	0.373
Bonwapitse	3111	18	13.3	16.3	10.0	7.7	1.233	0.613	0.473
Tautswe	3121	12	37.8	50.0	36.4	29.8	1.322	0.728	0.595
Mahalaptsw	3221	24	3.0	3.1	1.3	0.9	1.039	0.423	0.299
Lotsane	3321	29	36.0	42.9	24.4	16.0	1.190	0.568	0.372
Lotsane	3331	25	35.9	43.7	27.4	22.7	1.217	0.625	0.520
Motloutse	4121	28	122.6	107.0	24.3	15.9	0.872	0.227	0.148
Shashe	4321	31	301.3	300.4	95.3	34.5	0.997	0.317	0.115
Shashe	4361	35	134.3	147.8	77.9	62.5	1.100	0.527	0.423
Ntse	4411	39	42.9	45.0	20.8	14.3	1.048	0.462	0.318
Tati	4511	39	61.7	67.1	34.8	25.5	1.087	0.519	0.380

Table 4.12: Sample LH-moments and LH-moment ratios for L_2

Name of Site	Station No.	Sample Size	λ_1^2	λ_2^2	λ_3^2	λ_4^2	L_2-C_V (τ_2)	L_2 - Skewness (τ_3)	L_2 - Kurtosis (τ_4)
Metsimothabe	2411	22	21.9	32.6	10.8	5.8	1.493	0.332	0.178
Metsimothabe	2421	19	24.3	34.6	7.7	1.6	1.422	0.223	0.046
Kolobeng	2511	25	43.9	75.6	40.6	25.1	1.722	0.537	0.331
Bonwapitse	3111	18	12.1	20.8	12.0	9.1	1.720	0.575	0.438
Tautswe	3121	12	35.6	65.0	43.9	33.4	1.825	0.675	0.515
Mahalaptsw	3221	24	2.5	3.9	1.5	1.1	1.529	0.397	0.285
Lotsane	3321	29	32.3	54.2	27.9	16.9	1.676	0.515	0.313
Lotsane	3331	25	32.5	56.0	33.2	26.7	1.719	0.593	0.477
Motloutse	4121	28	97.0	131.4	27.7	15.9	1.355	0.211	0.121
Shashe	4321	31	250.8	365.5	99.5	30.3	1.457	0.272	0.083
Shashe	4361	35	116.4	187.6	93.5	71.9	1.611	0.498	0.383
Ntse	4411	39	36.4	56.6	23.9	15.5	1.553	0.423	0.273
Tati	4511	39	53.2	85.2	40.7	27.2	1.600	0.478	0.319

Table 4.13: Sample LH-moments and LH-moment ratios for L_3

Name of Site	Station No.	Sample Size	λ^3_1	λ^3_2	λ^3_3	λ^3_4	L_3 -Cv (τ_2)	L_3 -Skewness (τ_3)	L_3 -Kurtosis (τ_4)
Metsimothabe	2411	22	19.1	37.6	11.8	6.0	1.972	0.313	0.160
Metsimothabe	2421	19	20.8	39.2	7.6	1.4	1.880	0.194	0.037
Kolobeng	2511	25	40.8	89.6	45.1	27.8	2.193	0.504	0.310
Bonwapitse	3111	18	11.2	24.9	13.8	10.4	2.216	0.556	0.419
Tautswe	3121	12	34.0	79.1	50.6	36.6	2.323	0.640	0.463
Mahalaptsw	3221	24	2.2	4.5	1.8	1.3	2.026	0.389	0.283
Lotsane	3321	29	29.7	64.1	30.8	17.6	2.160	0.481	0.274
Lotsane	3331	25	30.3	67.4	38.8	30.4	2.226	0.576	0.451
Motloutse	4121	28	81.4	150.0	30.2	15.2	1.844	0.201	0.101
Shashe	4321	31	216.8	416.9	100.7	25.4	1.923	0.241	0.061
Shashe	4361	35	105.1	223.0	108.2	81.1	2.121	0.485	0.364
Ntse	4411	39	32.4	66.4	26.7	16.5	2.051	0.403	0.248
Tati	4511	39	47.9	100.8	45.7	28.2	2.104	0.453	0.280

Table 4.14: Sample LH-moments and LH-moment ratios for L_4

Name of Site	Station No.	Sample Size	λ^4_1	λ^4_2	λ^4_3	λ^4_4	L_4 -Cv (τ_2)	L_4 -Skewness (τ_3)	L_4 -Kurtosis (τ_4)
Metsimothabe	2411	22	17.1	41.9	12.5	6.1	2.457	0.298	0.146
Metsimothabe	2421	19	18.2	42.8	7.4	1.4	2.346	0.174	0.033
Kolobeng	2511	25	38.3	102.4	49.4	30.6	2.671	0.482	0.299
Bonwapitse	3111	18	10.6	28.8	15.7	11.7	2.716	0.545	0.408
Tautswe	3121	12	32.8	92.5	56.8	39.2	2.817	0.614	0.424
Mahalaptsw	3221	24	2.0	5.1	2.0	1.4	2.528	0.388	0.285
Lotsane	3321	29	27.7	73.1	33.3	18.0	2.642	0.455	0.246
Lotsane	3331	25	28.6	78.2	44.3	34.0	2.734	0.566	0.434
Motloutse	4121	28	70.7	165.1	32.0	14.1	2.335	0.194	0.085
Shashe	4321	31	191.8	458.8	99.7	19.8	2.392	0.217	0.043
Shashe	4361	35	97.1	255.5	122.4	90.3	2.630	0.479	0.353
Ntse	4411	39	29.5	75.0	29.2	17.4	2.547	0.389	0.232
Tati	4511	39	44.1	114.8	49.8	28.6	2.601	0.434	0.249

4.3.2 Discordancy Measure

The discordancy test has been carried out using equation (79) for all sites and the discordancy values are given in the following tables (Table 4.15 to Table 4.18).

Table 4.15: LH-moments ratios and discordancy measures for L_1

Name of Site	Station No.	Sample Size	λ_1^1 (m ³ /s)	L_1-C_V (τ_2)	L_1 -Skewness (τ_3)	L_1 -Kurtosis (τ_4)	D_i
Metsimotlhabe	2411	22	26.0	1.026	0.361	0.204	0.39
Metsimotlhabe	2421	19	29.6	0.972	0.268	0.064	1.28
Kolobeng	2511	25	47.7	1.260	0.592	0.373	1.40
Bonwapitse	3111	18	13.3	1.233	0.613	0.473	0.61
Tautswe	3121	12	37.8	1.322	0.728	0.595	1.24
Mahalapswe	3221	24	3.0	1.039	0.423	0.299	0.19
Lotsane	3321	29	36.0	1.190	0.568	0.372	0.74
Lotsane	3331	25	35.9	1.217	0.625	0.520	0.86
Motloutse	4121	28	122.6	0.872	0.227	0.148	2.33
Shashe	4321	31	301.3	0.997	0.317	0.115	0.89
Shashe	4361	35	134.3	1.100	0.527	0.423	0.63
Ntse	4411	39	42.9	1.048	0.462	0.318	1.04
Tati	4511	39	61.7	1.087	0.519	0.380	1.39
Weighted means				1.091	0.470	0.322	

Table 4.16: LH-moments ratios and discordancy measures for L_2

Name of Site	Station No.	Sample Size	λ_1^2 (m ³ /s)	L_2-C_V (τ_2)	L_2 -Skewness (τ_3)	L_2 -Kurtosis (τ_4)	D_i
Metsimotlhabe	2411	22	21.9	1.493	0.332	0.178	0.21
Metsimotlhabe	2421	19	24.3	1.422	0.223	0.046	1.51
Kolobeng	2511	25	43.9	1.722	0.537	0.331	1.20
Bonwapitse	3111	18	12.1	1.720	0.575	0.438	0.71
Tautswe	3121	12	35.6	1.825	0.675	0.515	1.29
Mahalapswe	3221	24	2.5	1.529	0.397	0.285	0.44
Lotsane	3321	29	32.3	1.676	0.515	0.313	0.95
Lotsane	3331	25	32.5	1.719	0.593	0.477	0.90
Motloutse	4121	28	97.0	1.355	0.211	0.121	1.87
Shashe	4321	31	250.8	1.457	0.272	0.083	0.95
Shashe	4361	35	116.4	1.611	0.498	0.383	0.71
Ntse	4411	39	36.4	1.553	0.423	0.273	0.76
Tati	4511	39	53.2	1.600	0.478	0.319	1.50
Weighted means				1.580	0.432	0.282	

Table 4.17: LH-moments ratios and discordancy measures for L_3

Name of Site	Station No.	Sample Size	λ_1^3 (m ³ /s)	L_3 -C _V (τ_2)	L_3 -Skewness (τ_3)	L_3 -Kurtosis (τ_4)	D_i
Metsimotlhabe	2411	22	19.1	1.972	0.313	0.160	0.28
Metsimotlhabe	2421	19	20.8	1.880	0.194	0.037	1.81
Kolobeng	2511	25	40.8	2.193	0.504	0.310	0.83
Bonwapitse	3111	18	11.2	2.216	0.556	0.419	0.71
Tautswe	3121	12	34.0	2.323	0.640	0.463	1.49
Mahalapswe	3221	24	2.2	2.026	0.389	0.283	0.73
Lotsane	3321	29	29.7	2.160	0.481	0.274	1.14
Lotsane	3331	25	30.3	2.226	0.576	0.451	0.91
Motloutse	4121	28	81.4	1.844	0.201	0.101	1.45
Shashe	4321	31	216.8	1.923	0.241	0.061	0.99
Shashe	4361	35	105.1	2.121	0.485	0.364	0.71
Ntse	4411	39	32.4	2.051	0.403	0.248	0.67
Tati	4511	39	47.9	2.104	0.453	0.280	1.29
Weighted means				2.070	0.410	0.258	

Table 4.18: LH-moments ratios and discordancy measures for L_4

Name of Site	Station No.	Sample Size	λ_1^4 (m ³ /s)	L_4 -C _V (τ_2)	L_4 -Skewness (τ_3)	L_4 -Kurtosis (τ_4)	D_i
Metsimotlhabe	2411	22	17.1	2.457	0.298	0.146	0.30
Metsimotlhabe	2421	19	18.2	2.346	0.174	0.033	2.04
Kolobeng	2511	25	38.3	2.671	0.482	0.299	0.48
Bonwapitse	3111	18	10.6	2.716	0.545	0.408	0.74
Tautswe	3121	12	32.8	2.817	0.614	0.424	1.69
Mahalapswe	3221	24	2.0	2.528	0.388	0.285	1.02
Lotsane	3321	29	27.7	2.642	0.455	0.246	1.19
Lotsane	3331	25	28.6	2.734	0.566	0.434	0.92
Motloutse	4121	28	70.7	2.335	0.194	0.085	1.13
Shashe	4321	31	191.8	2.392	0.217	0.043	1.03
Shashe	4361	35	97.1	2.630	0.479	0.353	0.69
Ntse	4411	39	29.5	2.547	0.389	0.232	0.49
Tati	4511	39	44.1	2.601	0.434	0.249	1.28
Weighted means				2.561	0.395	0.241	

It can be seen from Table 4.15 to Table 18 that the D_i values for all sites in the region are less than the critical value 2.869 (given by Hosking and Wallis, 1997) for the

thirteen sites in the study area; and thus there is no discordant site found in the study region.

The plots of sample cumulative variance and skewness of the LH-moments are shown in Figure 4.5 for L_1 to L_4 .

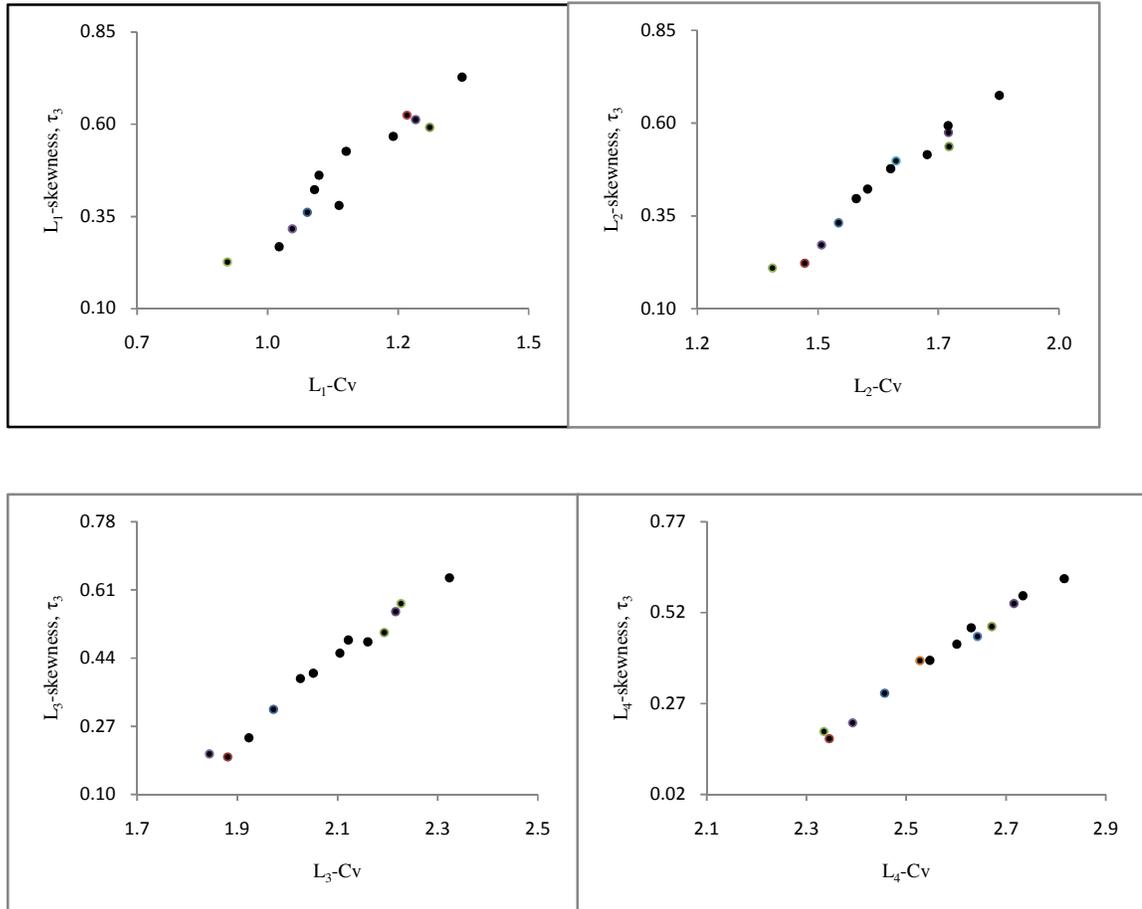


Figure 4.6: Plots of L_1 - C_v , L_2 - C_v , L_3 - C_v and L_4 - C_v versus L_1 -Skewness, L_2 -Skewness, L_3 -Skewness and L_4 -Skewness

Figure 4.6 illustrates the LH-moment diagrams (plots of the LH- C_v versus the LH- C_s from L_1 to L_4 of the observed data) under the assumption of one homogeneous region. However, LH- C_v and LH- C_s of level one (L_1) shows considerable dispersion of the data while level two to four (L_2 to L_4) of the same figure indicates more appropriate

grouping of data in the region. Hence, it has been found that further evaluation of the study area using heterogeneity tests is required.

4.3.3 Regional Homogeneity

Initially the entire catchment was assumed as one homogeneous region and homogeneity evaluations were performed for the LH-moments. Then, simulation with the kappa distribution was performed to conduct the H_i tests. The heterogeneity measures H_i , ($i=1, 2, 3$) were calculated for LH-moments based on the simulations of the observed standard deviation (SD) of LH-C_v, simulated mean of SD of LH-C_v/L-C_s and simulated standard deviation of SD of LH-C_v/L-C_k and 500 sets of simulated data. As discussed in Section 3.5.2, the heterogeneity measure (H_1) which is related to LH-C_v is most important and has larger effect than variation in LH-C_s or LH-C_k for assessing the relative level of heterogeneity for the proposed region. On the other hand, H_2 and H_3 which are related to LH-C_s and LH-C_k, respectively, have low discriminatory power as they are influenced by sample size and presence of extraordinary flows. The results of the heterogeneity tests based on the LH-moment are presented in Table 4.19.

Table 4.19: Heterogeneity measure for LH-moments

LH-moments	H_1	H_2	H_3
L_1	-0.15	-0.15	0.14
L_2	-0.08	-0.08	0.18
L_3	-0.05	-0.05	0.29
L_4	-0.22	-0.22	0.46

From Table 4.19, it has been observed that the H_i , ($i=1, 2, 3$) values for each LH-moments $L_{\eta}, \eta = 1, 2, 3, 4$, that is, (L_1 to L_4) are less than 1 and identified as the homogeneous region.

In general, since the value of H_1 1.27 for L-moments is not far from the “acceptably homogeneous” criterion value of 1 and based on other LH-moments, that is, L_1, L_2, L_3

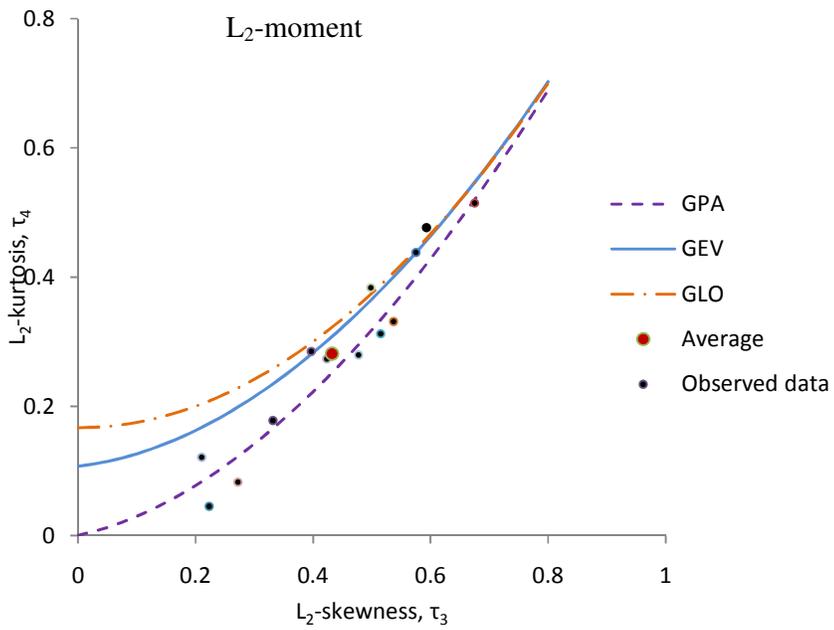
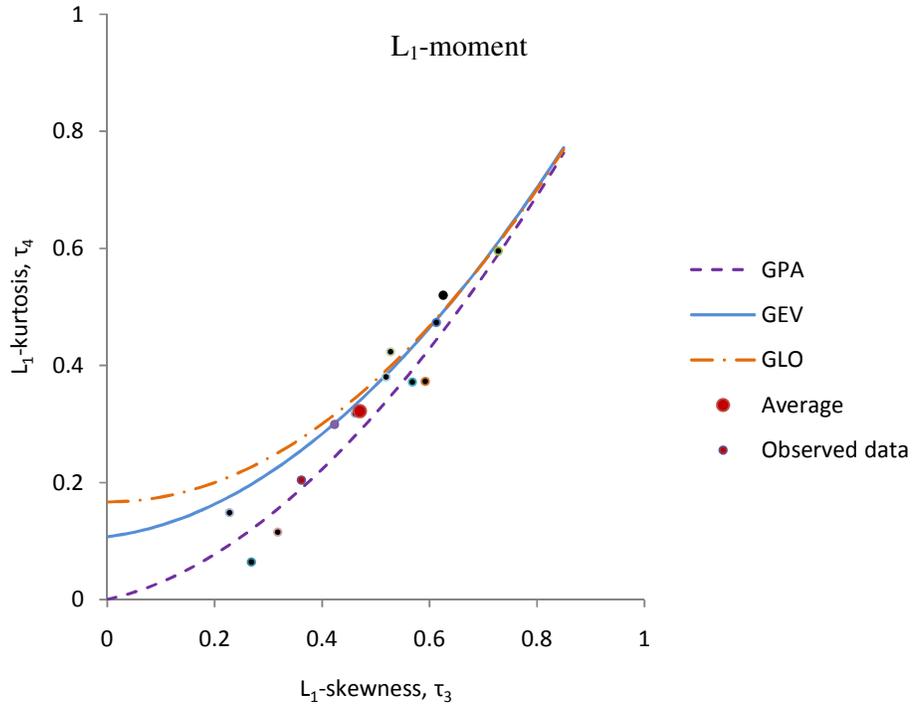
and L_4 the region has been identified as “acceptably homogeneous”, we can therefore consider the whole Limpopo region under study is homogeneous.

4.3.4 Choice of Suitable Distribution

4.3.4.1 LH-Moments Ratio Diagrams

In the LH-moments ratio diagram, the theoretical curves of GEV, GLO and GPA distributions for each level of LH-moments (L_1 to L_4) as well as the regional average LH-skewness and LH-kurtosis are plotted for selecting the best fit distribution (Figure 4.7).

The LH-moments ratio diagrams for the LH-moments L_η , $\eta = 1, 2, 3, 4$, that is, (L_1 to L_4) are shown in Figure 4.7. It has been observed from Figure 4.7 that L_1 and L_2 -moments identify GEV distribution as the best fit distribution for the Limpopo river system. Similarly, LH-moments of L_3 as well as L_4 identify GPA distribution as the best fitting distribution. On the other hand, Figure 4.7 indicates two rather different patterns for cases $\eta=1, 2$ and $\eta>2$. For the case of LH-moments $\eta=1, 2$, there is better scattering of data around GEV and GPA. However, for the case of $\eta>2$, most of the data points fall below the positions of GLO, but better scattered around the GPA distribution. Therefore, it is difficult to select the best fitting for a region on the basis of LH-moment ratio diagrams only. Hence, the Z-statistic criteria should be considered for selecting the best fitting distribution as discussed in the subsequent section.



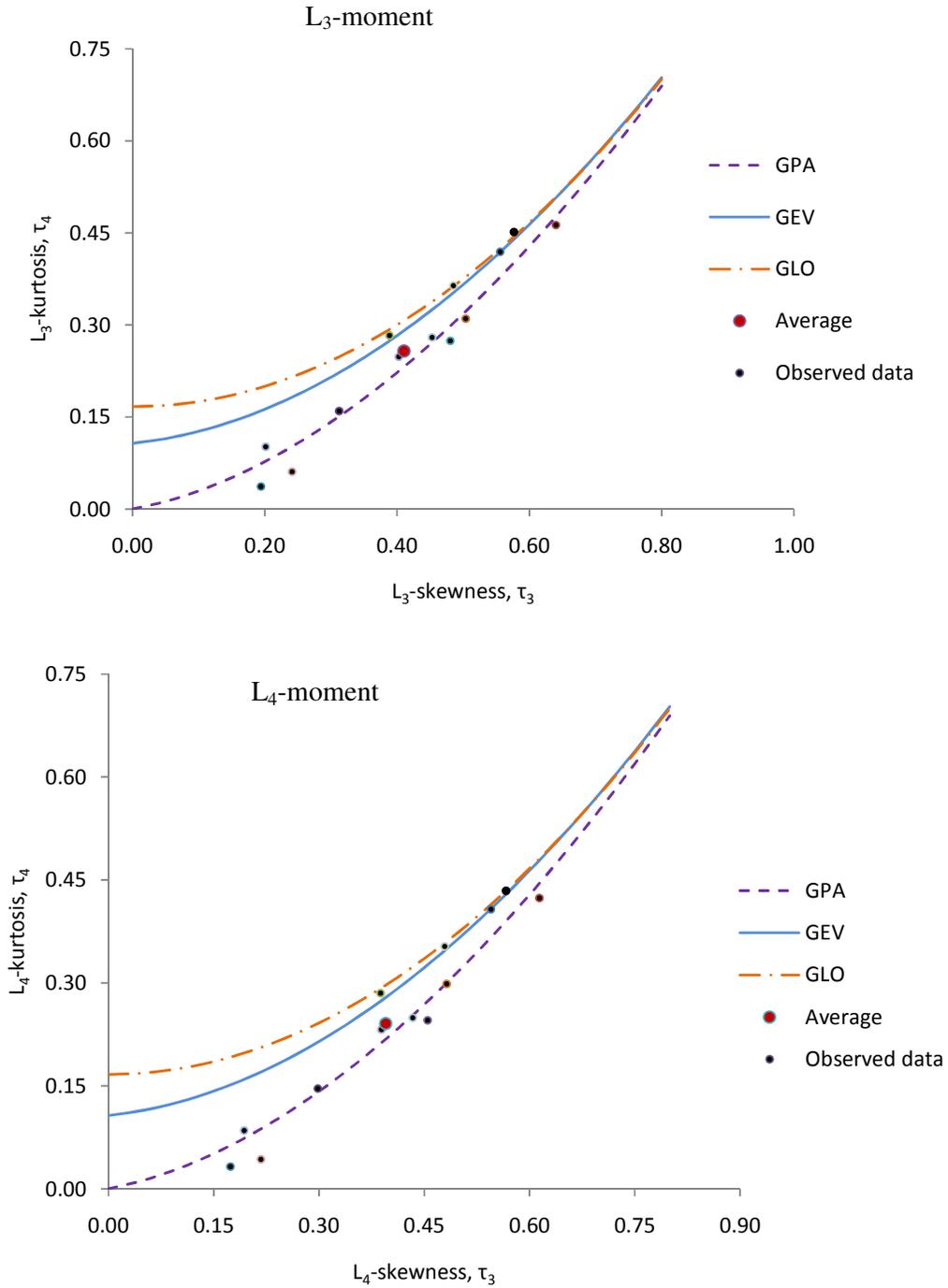


Figure 4.7: LH-moments (L_1 to L_4) ratio diagrams

4.3.4.2 Z-Statistics Criteria

As indicated in Section 3.5.2.1 of item II, the Z^{DIST} requires evaluation of τ_4^{DIST} , which has been obtained by performing simulation with the kappa distribution. The

values of $|Z|$ -statistics for the GEV, GLO and GPA based on L_1, L_2, L_3 and L_4 -moments were then calculated as shown in Table 4.20. It has been observed from Table 4.20 that values of $|Z|$ -statistic for GEV and GLO distributions are less than the critical value 1.64 for all the LH-moments $L_{\eta}, \eta = 1, 2, 3, 4$, that is, (L_1 to L_4). Similarly, in case of GPA distribution, the values of $|Z|$ -statistic are less than the critical value 1.64 for the LH-moments $L_{\eta}, \eta = 2, 3, 4$, that is, L_2, L_3 and L_4 , but the LH-moments $L_{\eta}, \eta = 1$, that is, L_1 gave values greater than the critical value 1.64. Nevertheless, L_2 -moment has given the smallest value for the GEV distribution, and therefore, the GEV distribution is identified as the best fitting distribution among the three distributions used for this study region based on L_2 -moments.

Table 4.20: $|Z|$ -statistics values

LH-moments	GLO	GEV	GPA
L_1	-0.35	-0.61	-1.77
L_2	0.26	-0.10	-1.50
L_3	0.69	0.26	-1.31
L_4	1.03	0.54	-1.16

Wang (1997) identified a reversing trend in improved performance of the GEV distribution at the LH-moments level 3 (L_3) during the goodness-of-fit test; and the same trend has been also observed in this study as shown in Table 4.20. Similarly, in the same table, the same trend has been also observed for GLO distribution. As for the case of the GPA distribution, an improved performance was observed for all levels moving from L_1 to L_4 . Then unlike the GEV distribution as identified by Wang (1997) (suggesting that additional investigations should be stopped at the L_2 level), it implies that it is possible to search for improved results by continuing on to the L_4 level of the GPA distributions (Meshgi and Khalili, 2007b).

4.3.5 Parameters Estimation

Based on the goodness-of-fit criteria discussed above, the GEV distribution using L_2 -moments has been identified as the best fitting distribution among the three distributions for fitting the data. Accordingly, Equations (53-56) for GEV based on the regional average LH-moment ratios $\tau_{2,2}^R$ and $\tau_{2,3}^R$ together with $\hat{\lambda}_1^2$, for $L_2, \eta=2$, are used for estimates the parameters of the GEV distribution and is presented in Table 4.21.

Table 4.21: Regional parameters of the distributions for LH-moments, $L_2, \eta=2$

LH-moments	Distribution	Parameters		
		Location ($\hat{\xi}$)	Scale ($\hat{\alpha}$)	Shape ($\hat{\kappa}$)
L_2 -moments	GEV	0.379	0.455	-0.434

4.3.6 Flood Quantiles Estimation and Development of Regional Growth Curves

The quantiles of the regional growth curves of GEV distribution are calculated for the LH-moments, $L_2, \eta = 2$ by using the regional parameters of Table 4.21 and Equation (47) and the results are presented in Table 4.22.

Table 4.22: Quantiles estimates for regional growth curves of GEV distribution for LH-moments, $L_2, \eta = 2$ at various probabilities of non-exceedance using GEV distribution

Reccurence Interval, T (Years)	2	10	20	50	100	1000
Probability of non-exceedance (F)	0.5	0.9	0.95	0.98	0.99	0.999
Gumbel Reduced Variate (Y)	0.37	2.25	2.97	3.90	4.60	6.91
Estimated Standardized Quantiles, L_2 -GEV, $x(F)$	0.56	2.11	3.14	5.03	7.05	20.31

For Limpopo region, the estimated standardized quantiles at specified recurrence intervals (Gumbel Reduced Variates) have been computed and growth curves developed for LH-moments (L_2) as shown in Figure 4.8

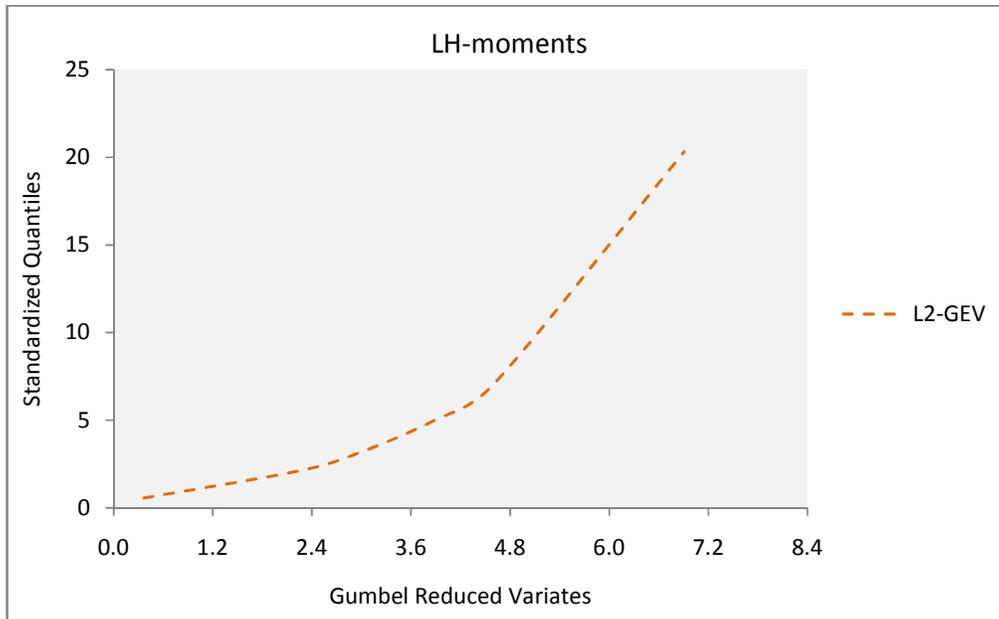


Figure 4.8: Regional growth curve for Limpopo region using LH-moment (L_2)

Hence, using the regional growth curve of Figure 4.8 and the mean annual maximum flood in Table 4.2, the design flood for the site of interest can be estimated using the predication Equation (92) for desired recurrence intervals.

4.3.7 Comparison of Observed and Estimated Floods

To assess the descriptive ability of the suggested distribution using LH-moments (L_2), the relative Root Mean Square Error (RMSE) has been calculated using Equation (99) and the results are presented in Table 4.23.

Table 4.23: Root Mean Square Error (RMSE) using LH-moment (L_2)

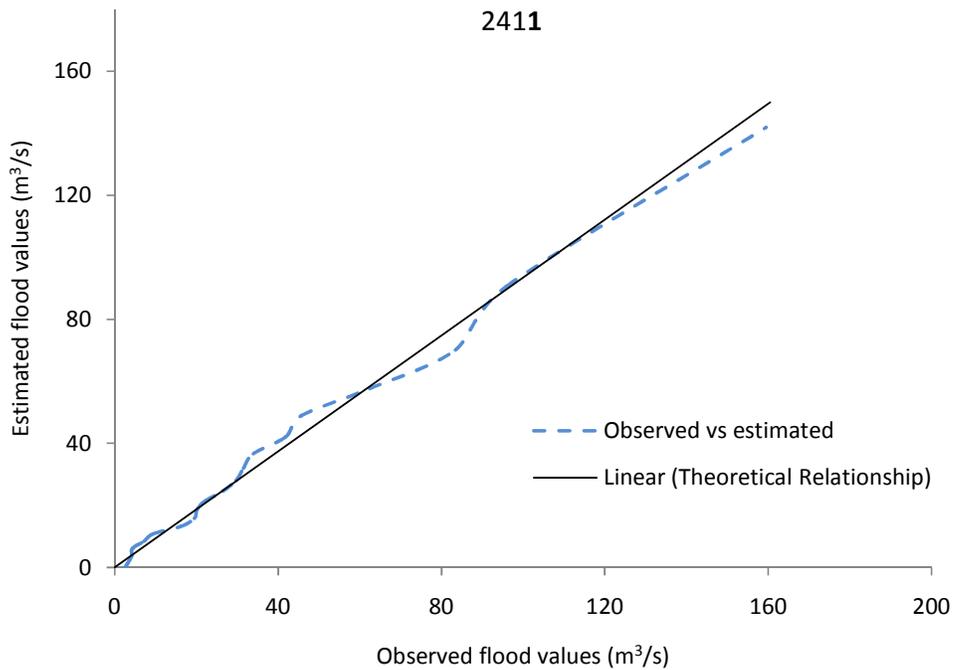
Sl. No.	Name of the River/Stream	Station No.	Observed Mean Flow (m ³ /s)	Estimated Mean Flow (m ³ /s)	RMSE*
1	Metsimothabe	2411	32.7	30.2	0.31
2	Metsimothabe	2421	38.5	35.5	0.23
3	Kolobeng	2511	52.7	48.6	8.51
4	Bonwapitse	3111	15.1	13.9	1.38
5	Tautswe	3121	41.6	38.2	4.22
6	Mahalapswe	3221	3.8	3.5	0.29
7	Lotsane	3321	42.4	40.7	0.54
8	Lotsane	3331	41.6	38.4	0.72
9	Motloutse	4121	175.1	168.2	0.38
10	Shashe	4321	390.4	374.3	0.21
11	Shashe	4361	168.6	161.4	0.27
12	Ntse	4411	56.1	53.6	0.33
13	Tati	4511	79.1	75.6	0.39
Average RMSE					1.37

*RMSE calculated based on all the observed and computed values at each station.

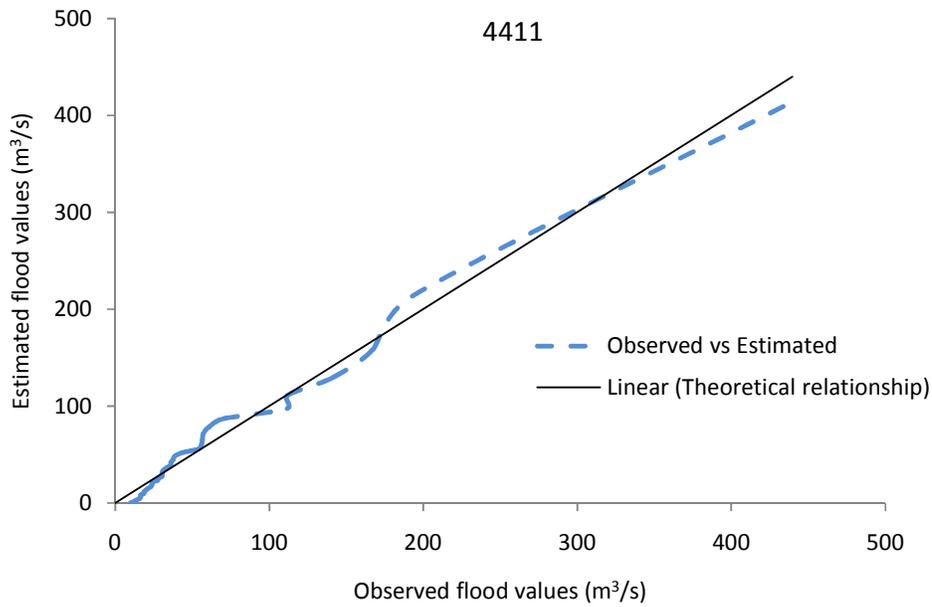
Table 4.23 shows the RMSE values for each site and the regional average using the LH-moment (L_2) procedure. Though generally the RMSE value for Kolobeng has dropped as compared to the one in L-moment, it is also important to see that the RMSE value is still distinctively high. In similar explanation to Section 4.2.7, there is no physical ground and evidence of gross error in the data as long as its D_i value of 0.83 is less than the critical value of 3.

The regional average RMSE measures the overall deviation of estimated quantiles from true quantiles. It is the criterion to which it is given most weight in judging whether the estimation procedure is superior to another. Accordingly, the regional average RMSE using the LH-moment (L_2) will be compared with the regional average RMSE using the L-moment in Section 4.4 to conclude the suitability of the methods.

Similarly, graphical comparisons between the observed and estimated flood values at corresponding probabilities of non-exceedance, at different sites of the region have been undertaken. For this, Cunnane's (1980) plotting position formula (Equation 100) has been used in order to maintain unbiasedness and minimum variance in computation of probabilities of non-exceedance. Examination of the observed and estimated floods at different sites showed that a good agreement, in general, can be seen from two typical plots as given in Figures 4.9a and b.



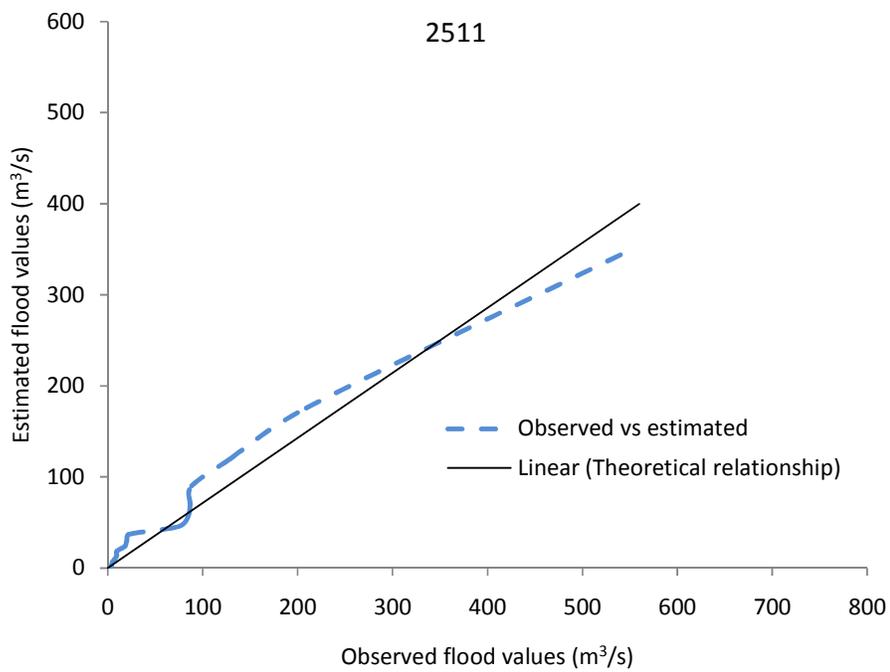
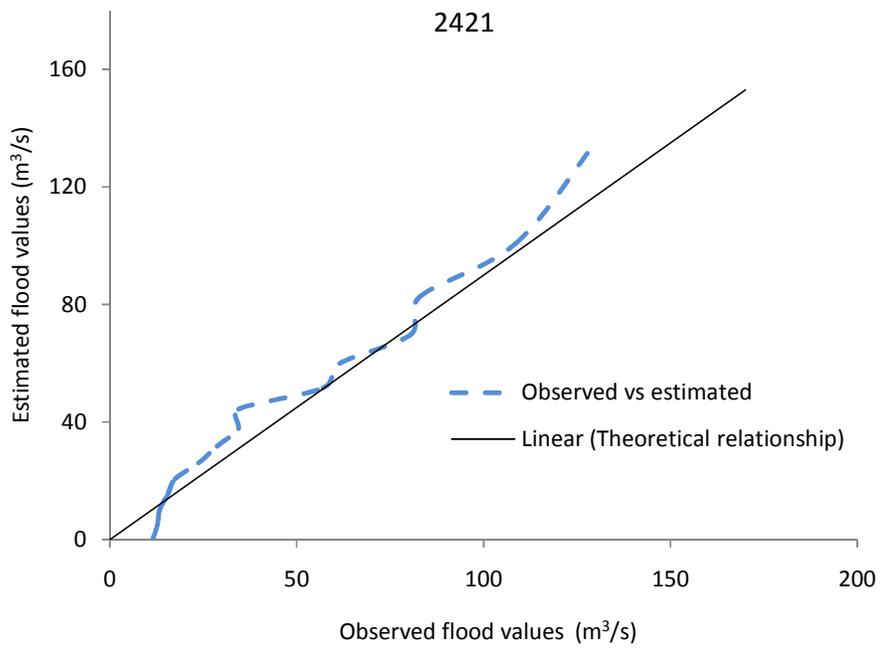
(a)

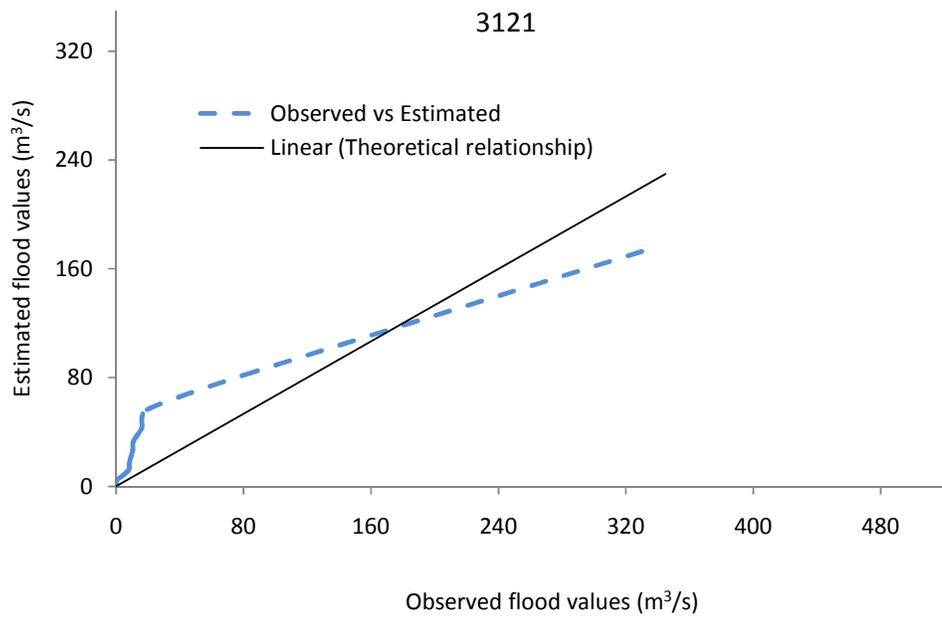
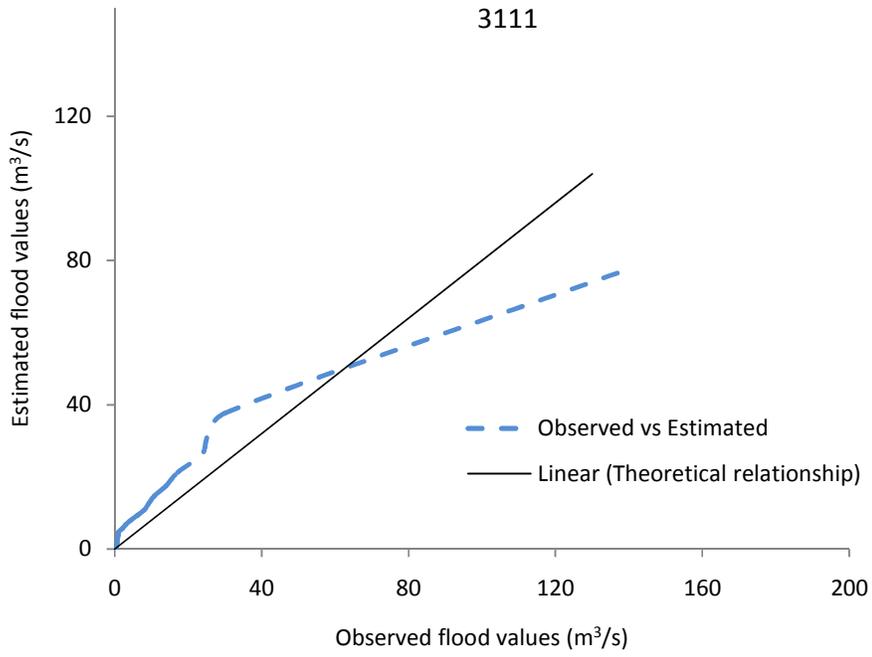


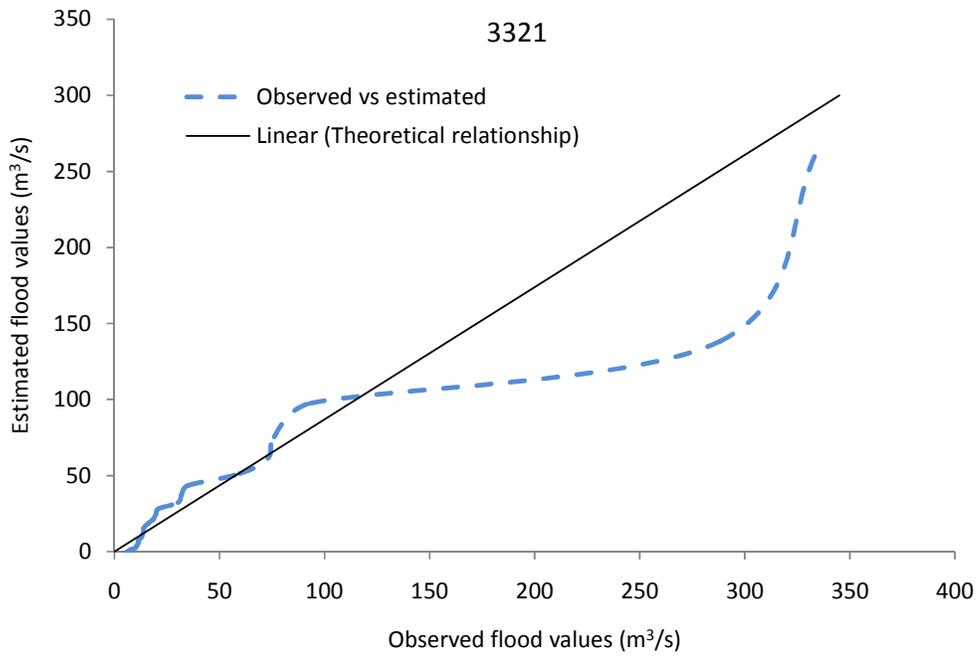
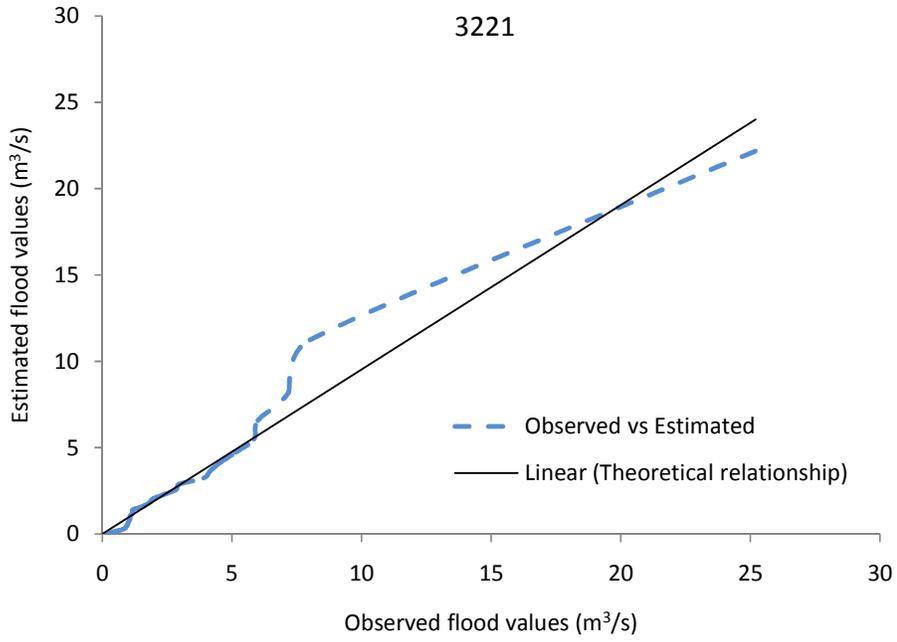
(b)

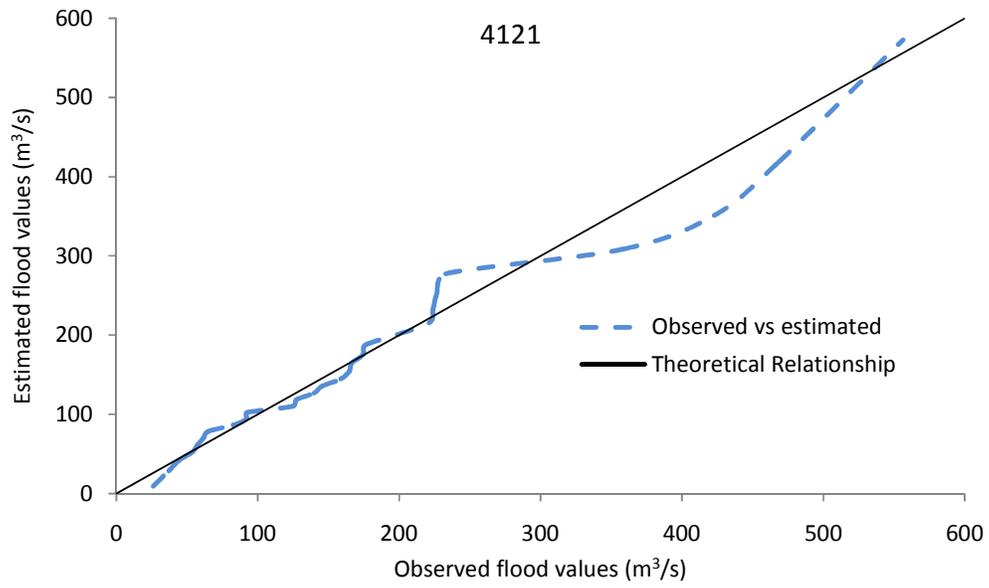
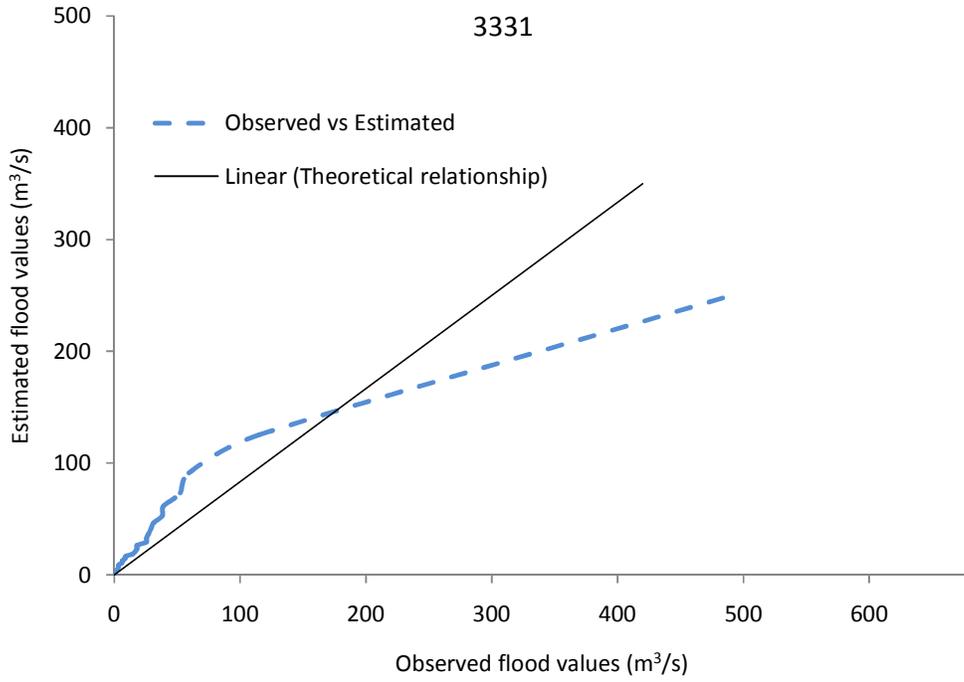
Figure 4.9: Plot of observed and estimated floods at gauging stations No. 2411 and 4411, respectively

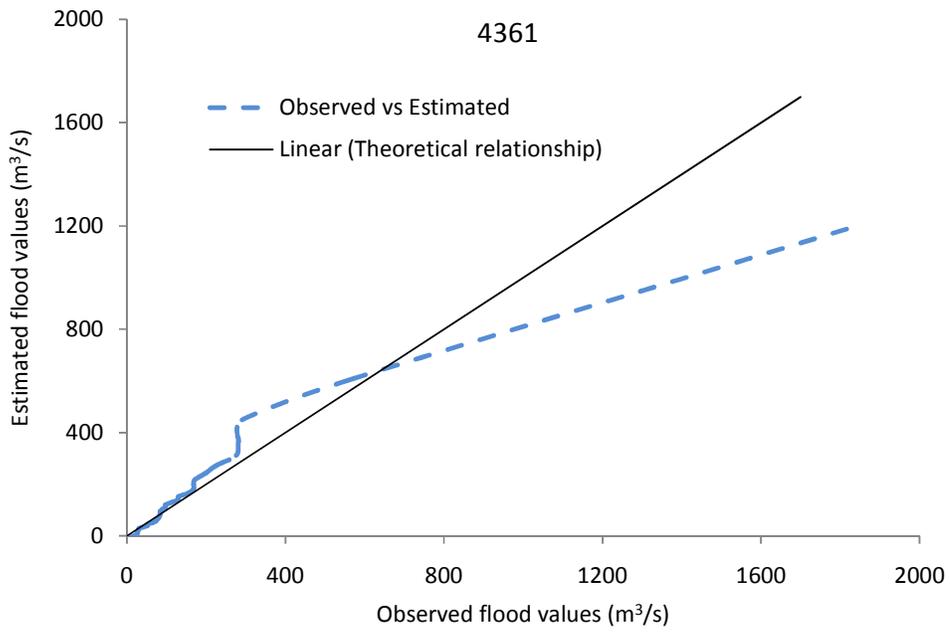
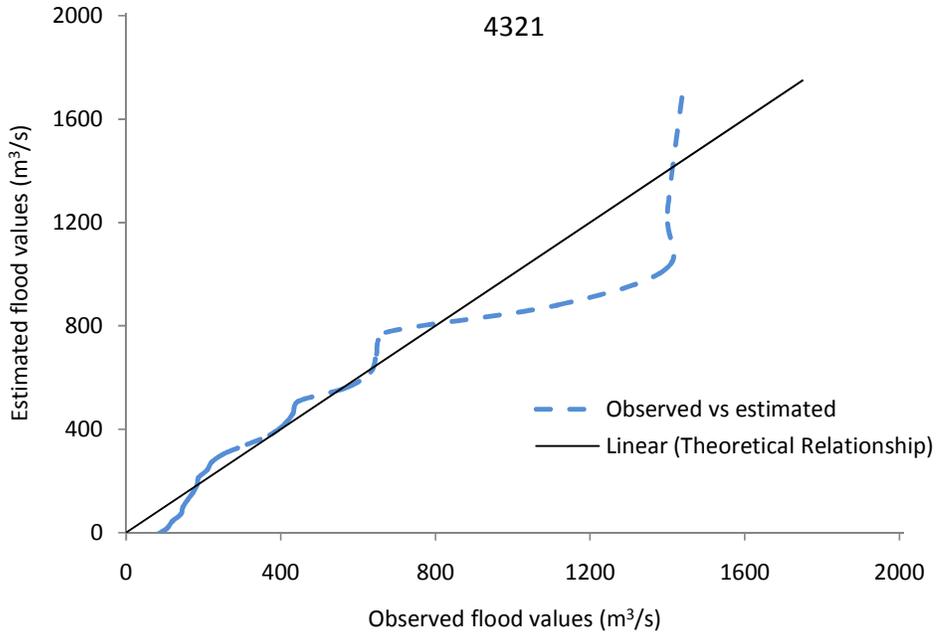
The same plots for other sites have been provided in Figure 4.10.











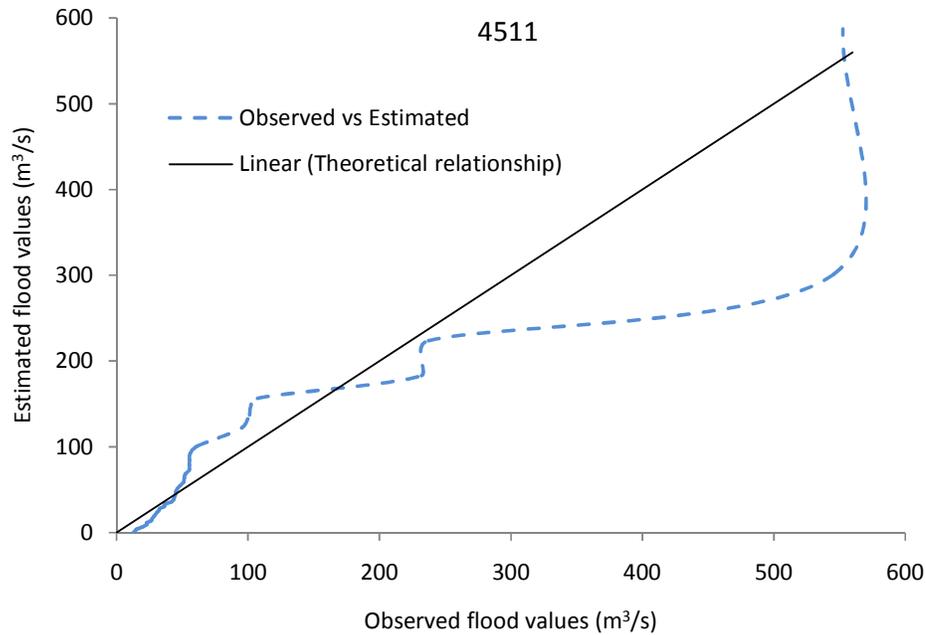


Figure 4.10: Plot of observed and estimated floods at gauging stations No. as shown

It can be seen from Figure 4.10 that observed floods and estimated floods using LH-moment (L_2) for some sites at some recording periods are not in agreement. In line with Section 1.2, this could be attributed due to mainly the measurement and extreme sampling error of the flows at those sites. The same situation was also generally indicated in NWMPR (2006) and Farquharson (1992) as a caution. Hence, the estimated flows at large return periods may not be reliable; and accordingly, special attention shall be given when a flood of higher return period is considered.

4.3.8 Summary

The Regional Flood Frequency Analysis based on the GEV distribution using the LH-moments can be summarized as follow:

1. In the step of initial screening of the data, the discordancy measure is used for all LH-moments (L_1 to L_4). The discordancy measure shows that data of all gauging sites of this study area are suitable for using regional flood frequency analysis by all the LH-moments.

2. For testing homogeneity of the region for all the LH-moments (L_1 to L_4), the L-moment based heterogeneity measure is used, and extended it to all the LH-moments. The heterogeneity measure shows that the Limpopo region has been found to be homogeneous for all the LH-moments.
3. The regional flood frequency analysis was performed for LH-moments (L_1 to L_4) by using the three parameters distributions viz: GLO, GEV and GPA. The LH-moment ratios diagram is used to identify the suitable distribution for each LH-moment levels for the region. Again the $|Z|$ -statistic criteria have been used to identify best suitable distribution for the study area. From Table 4.20, it is observed that the $|Z|$ -statistic value of GEV distribution in L_2 -moment is the smallest among all the $|Z|$ -statistic value. Therefore, the GEV distribution with L_2 -moment is identified as most suitable for regional flood frequency analysis of the Limpopo region.
4. Based on the GEV regional parameters, the regional quantile estimates with non-exceedance probability F have been calculated as presented in Table 4.22 and a regional growth curve has been developed as shown in Figure 4.8.
5. The Relative Mean Square Error (RMSE) has been used to compare the observed and estimated floods to assess the descriptive ability of the suggested distribution using LH-moment (L_2). According to the result, it has been found that the regional average RMSE is 1.37.

4.4 Discussion of Results

Based on the statistical analysis results with L-moments and LH-moments (L_2), the following summary can be provided:

1. Generally, L-moment as well as well as LH-moment of level 2 (L_2) have provided correlated quantile results based on GEV distribution as presented in Table 4.8 and Table 4.22, respectively.
2. L-moment in general shows a better fit with the developed GEV distribution for the data series ($0.5 \leq F \leq 0.95$) and is better than the results obtained using LH-moment of level 2 (L_2) based on GEV distribution as shown in the regional growth curves of Figure 4.11.

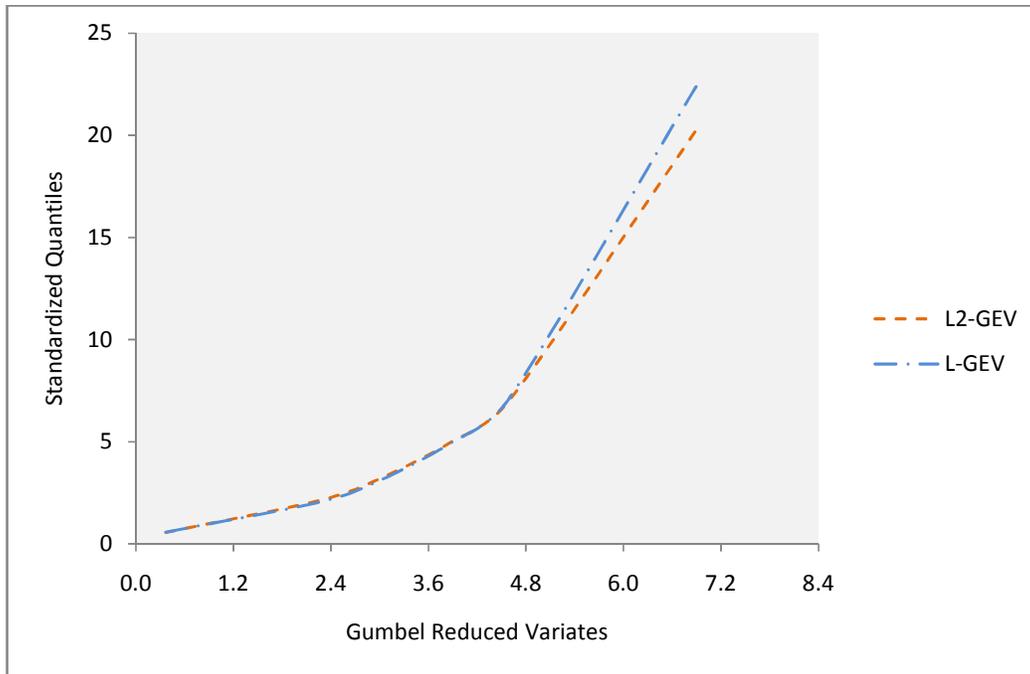


Figure 4.11: Regional growth curves using LH-moments (L_2) and L-moments

On the other hand, for the data series ($0.95 \leq F \leq 1$), it is clear that the GEV using L_2 -moments method is the best fit as compared to the L-moments method.

3. In line with item 2 above and Figure 4.11, the GEV/L-moment regional growth curve shows the highest deviation whereas GEV/ L_2 -moment is the least deviated fit at the large return periods.
4. For a complete data series ($0 \leq F \leq 1$), the same figure clearly shows that GEV distribution using LH-moments (L_2) is better than the L-moments to characterize the flood data of the Limpopo region.
5. Based on item 2 above, generally L-moment is more efficient than the LH-moment for shorter return periods; whereas LH-moments are more efficient than the L-moments for higher return periods.
6. Statistical analysis of extremes is often conducted for predicting large return period events and thus LH-moments are suitable methods to predict larger events.
7. The relative Root Mean Square Error (RMSE) based on the observed and estimated floods was calculated using the L-moment and LH-moment (L_2) as shown in Table 4.24.

Table 4.24: Summary Root Mean Square Error (RMSE) for L- and LH-moments

Distribution	Method	RMSE
GEV	L-moment	1.92
	LH-moment (L_2)	1.37

8. From Table 4.24, it can be seen that the RMSE for GEV distribution using LH-moment (L_2) is smaller than the L-moment based on the same distribution. This inference can also be seen from Table 4.9 and Table 4.23 that the design floods that have been derived using LH-moment (L_2) method tended to approach the design floods by the observed annual maximum floods. Thus, in summary, it can be said GEV distribution using LH-moment (L_2) provides a better fit to the annual maximum flood data of the Limpopo region.
9. Based on item 8 above and using LH-moments at level 2 (L_2), the prediction equation for estimation of flood quantiles at different non-exceedance probabilities using the location (ξ), scale (α) and shape (κ) parameters of Table 4.21 and Equation (47) can be given as:

$$\hat{Q}_i(F) = [-0.671 + 1.050(-\ln F)^{-0.434}] * \hat{\lambda}_{1,i}^2 \quad (101)$$

4.5 Flood Estimation for Ungauged Basins

Ungauged basins are those which lack sufficient streamflow data or data at all are not available which needs to be predicted using other available data / information from the relevant catchment or using other catchment. However, there are various methods for prediction of floods for such ungauged river basins like using catchment characteristics as discussed in the subsequent sections.

4.5.1 Procedures

Two major variables which are responsible for flood generation viz: drainage area (A in km^2) and mean annual rainfall (R in mm) were used for the estimation of the L_2 -moment ($\hat{\lambda}_1^2$) or the index flood in (m^3/s) at ungauged sites. For this purpose, the

iterative solving algorithm software SPSS ver. 16 – EQUiNOX has been used to develop best-fit prediction equation. Preference was originally given to power-form function equations but it has been found this form did not provide good results as measured by the coefficient of determination (R^2). The regression analysis has been therefore expanded to a linear relationship of the regression analysis between the observed average annual maximum floods (\bar{x} in m^3/s), drainage area (A in km^2) and mean annual rainfall (R in mm) at different sites of the region have been undertaken and the regression results are presented in Table 4.25.

Table 4.25: Regression results

Model Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	.716 ^a	.512	.415	80.29265	

a. Predictors: (Constant), rainfall, area

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	67663.433	2	33831.716	5.248	.028 ^a
	Residual	64469.101	10	6446.910		
	Total	132132.533	12			

a. Predictors: (Constant), rainfall, area
b. Dependent Variable: xbar

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	338.165	241.932		1.398	.192
	area	.021	.008	.582	2.564	.028
	rainfall	-.675	.508	-.302	-1.328	.214

a. Dependent Variable: xbar

Based on the statistical results of Table 4.25 and Equation (97), the mean annual maximum flood can be derived as:

$$\hat{\lambda}_1^2 = 338.165 + 0.021A - 0.675R \tag{102}$$

and the relationship yielded a coefficient of correlation equal to 0.5120 which is compared to critical coefficient of 0.4821 at 15 degrees of freedom and 5% significance, is reasonably high and can be accepted.

For development of the regional flood frequency relationship for estimation of floods of various return periods for ungauged basins, the regional flood frequency relationship given in Equation (101) can be coupled with the regional relationship given by Equation (102) and the following regional flood frequency relationship has been developed.

$$\hat{Q}_i(F) = -226.992 - 0.014A + 0.453R + (355.089 + 0.022A - 0.709R)(-\ln F)^{-0.434} \quad (103)$$

Therefore, for any given catchment within the homogeneous region, flood quantiles for a given return period can be determined by the use of Equation (103).

4.5.2 Flood Estimation at Gauged Site Using Equation (102)

Assuming the site at station 4411 has not been gauged. Let it be assumed also that a 100 years return period flood is required to be estimated. Then, knowing the area A and mean annual rainfall R from Table 4.1 as 800km² and 425mm, respectively, the mean annual maximum flood $\hat{\lambda}_1^2$ using Equation (102) gives 68.09m³/s. From the growth curve of Figure 4.8, for T=100 years return period, the standardised quantile \hat{q} is read off as 7.05. Then the estimated flood quantile is calculated as $\hat{Q}_i(F_{100}) = 7.05 * 68.090 = 479.77m^3/s$.

The observed annual maximum flood for the same region is 56.115m³/s and the standardised quantile \hat{q} for T₁₀₀ return period is 7.05. Then the estimated flood quantile for the same station is calculated using the regional Equation (101) as $\hat{Q}_i(F_{100}) = 7.05 * 56.12 = 395.39m^3/s$. Thus, when the estimated flood of 479.77m³/s is compared to 395.39m³/s as calculated from the regional equation, the percentage error is found to be 21%, which is acceptable for higher return period. For example, the percentage error for a return period of 50 years, the percentage error would be 14%. Therefore, due to the short time series available for most of the stations, the

flood frequency analysis results for longer recurrence intervals should be viewed with caution.

4.6 Reference to Previous Studies

As indicated in Section 2.1, there are some previous related studies undertaken in relation to the Botswana flood frequency analysis. However, in nearly all of such studies, the objectives were different from the main objectives in this study. Nevertheless, as a reference, below are some of the studies which were undertaken on floods in Botswana.

1. Farquharson et al. (1992) carried out a study to estimate mean annual flood (MAF) based on data from 42 catchments in Botswana and South Africa. Accordingly, they developed a regional growth curve which could be used for large-scale applications.
2. NWMPR (2006) undertook a study on the surface water resources of Botswana. This study did not develop a regional growth curve, and the short period of data was cited as the main hindrance.
3. Parida (2004) carried out a study on the flood characteristics of selected rivers in Botswana using L-moments method and regionalization technique, which was addressed based on limited data. The datasets used by Parida (2004) for analysis was based on flood records from the selected streams which varied between 10 and 34 years with an average of 25 years; whereas in this study the records varied between 12 and 39 years with an average of 27 years and thus in this study relatively more data for longer period of time have been collected after a study by Parida (2004) was conducted. And as such the longer the flow records the better improved design events. This argument is supported and can be seen in Figure 4.12 as the regional growth curve developed by this study is relatively the least deviated fit as compared with that of a study by Parida (2004). However, the average flow records by Parida (2004) and this study do not have as such significant differences and according to Figure 4.12 the regional growth curves

developed by Parida (2004) and by this study using the L-moment method are in agreement.

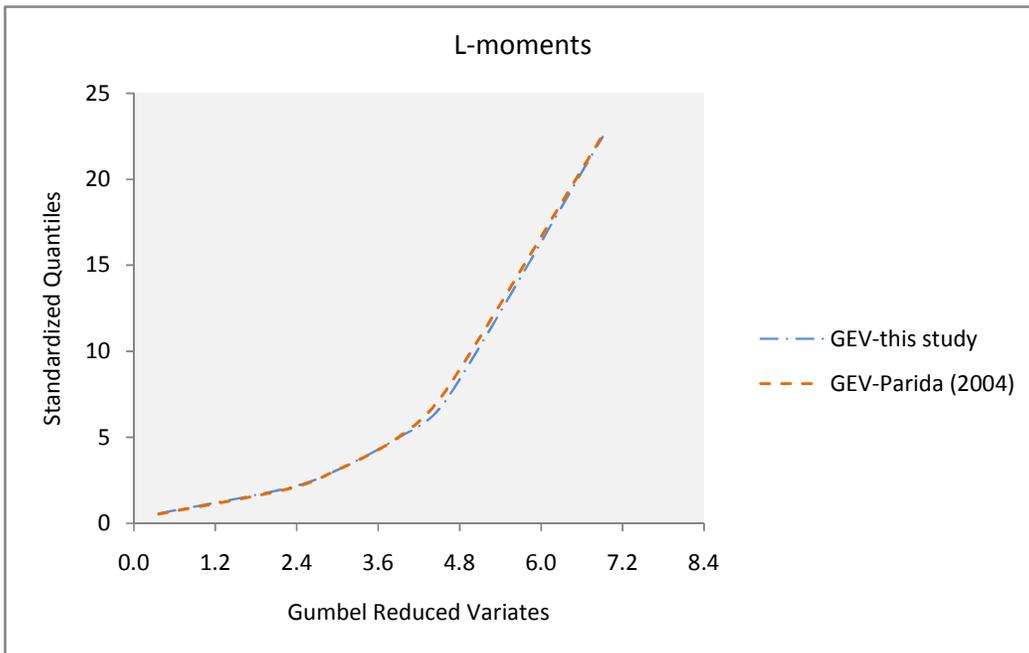


Figure 4.12: Regional growth curves by Parida (2004) and this study

CHAPTER 5

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study investigated and compared the newer available methods of parameter and quantile estimates viz: the L-moments and LH-moments using the commonly used three parameter distributions, that is, the generalized extreme value (GEV), generalized logistic (GLO) and generalized Pareto (GPA) distributions to develop regional growth curves that can be used for estimation of quantiles at either gauged or ungauged sites. The goodness-of-fit criteria using LH-moments show that generally GEV distribution well fitted the Limpopo flood data followed by GLO and GPA, respectively, except in L_1 in GPA as $|Z|$ -statistics value of 1.77 is slightly greater than the critical value of 1.64 (please refer Table 4.20).

The following conclusions can be drawn from the Regional Flood Frequency Analysis (RFFA) of this study area:

1. The RFFA was carried out based on real flood data from the selected gauged streams whose records varied between 12 and 39 years with an average of 27 years.
2. The heterogeneity test has revealed that all the 13 gauged sites in the Limpopo catchment have been found to be homogeneous for all LH-moments levels (L to L_4).
3. The goodness-of-fit criteria ($|Z|$ -statistics criteria) showed that GEV and GLO distributions fitted the data appropriately in both L-moments and LH-moments. Though GEV has been found to be the suitable distribution in L-moments, the GEV using LH-moments of level 2 (L_2) has been found to be the best suitable frequency distribution to characterize the flood data of the Limpopo region.
4. Further evaluation and comparison of the two methods (L-moments, L and LH-moments, L_2) using the relative Root Mean Square Error (RMSE) concluded that

again GEV using the LH-moments (L_2) with minimum RMSE found to be the best distribution that can model the floods of Limpopo region accurately.

5. Generally, it can be concluded that the LH-moments gave improved results compared to the L-moments as discussed in detail in Section 4.4.
6. For ungauged sites, a general equation between the average flood, catchment area and annual average rainfall has also been developed, which can be used in conjunction with values from the regional frequency curve (also known as the regional growth curve) for reliable estimation of design flood values of this study area.

5.2 Recommendations

The following recommendation can be drawn from the Regional Flood Frequency Analysis of this study area:

1. Flood frequency analysis is a dynamic process. Established methods are improved and changed with time and more reliable data for longer periods are collected from time to time. Therefore, because of the sensitivity of the issue of floods and moreover for water resources planning, development and management, the flood frequency of the Limpopo region should be revised any time and modifications shall be made to the established procedure, particularly if the same procedure would be used in the study of floods of the Limpopo region.
2. From the goodness-of-fit measure, several probability distributions emerged as possibly likely distributions for the area under study. The applicability of these to this study area should be investigated more to arrive at a more conclusive decision.
3. While the L-moments method has been developed for estimation of the parameters of many of the distributions, the LH-moments method (L_1 to L_4) has been developed for only three common distributions viz: GEV, GLO and GPA. As such many researchers have been investigating more on the parameter estimations of the LH-moments method; and therefore, it is recommended that the LH-moments method attempted in this Dissertation shall be revisited based on the findings on the subsequent developments of this method.

REFERENCES

- Ben-Gal, I. (2004). Outlier Detection. *Research Gate*. Retrieved from <http://www.researchgate.net/publication/226362876>. *Outlier Detection*.
- Bhuyan, A., Borah, M., and Kumar, R. (2009). Regional flood frequency analysis of North-Bank of the river Brahmaputra by using LH-moments. *Water resources management*, 24(9), 1779-1790.
- Brutsaert, W. (2005). Hydrology, an Introduction. *Cambridge University Press*.
- Chow, V.T. (ed.) (1964). Handbook of applied hydrology. *McGrawHill, Inc. New York*.
- Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). Applied hydrology. *McGrawHill, Inc. U.S.A.*
- Cong, S. and Xu, Y. (1987). Effects of discharge measurement error on the results of flood frequency analysis. *Application of Frequency and Risk in Water Resources*, V.P. Singh (Ed), 175 – 190.
- Conway, G. (2009). The science of climate change in Africa: impacts and adaptation. *Grantham Institute for Climate Change Discussion (1)*, Imperial College, London.
- Cunnane, C. (1987). Review of statistical models for flood frequency estimation. *Hydrologic Frequency Modelling*, D. Reidel Pub. Co., Holland, V.P. Singh (Ed), 49 – 95.
- Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *J. Hydrol.*, 100, 269 – 290.
- Dalrymple, T. (1960). Flood-Frequency Analysis. *Manual of Hydrology*, 3, United States Gov. Printing Office, Washington.

- Dingman, S.L. (2002). *Physical Hydrology. Second Edition University of New Hampshire.*
- Doheny, E.J. and Dillow, J.J.A. (1999). Adjustment to US Geological Survey Peak-Flow Magnitude-Frequency relations in Delaware and Maryland following Hurricane Floyd.
- EMWIS (2006). *The African Water Development Report.*
- Farquharson, F.A.K., Meigh, J.R. and Sutcliffe, J.V. (1992). Regional flood frequency analysis in Arid and Semi-Arid Areas. *Journal of Hydrology, 138, 472-501.*
- Gheidari, M.H.N. (2013). Comparison of the L- and LH-moments in the selection of the best distribution for regional flood frequency analysis in Lake Urmia Basin. *Department of Civil Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran, 30, 72-84.*
- Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability Weighted Moments. *Definitions and relation to parameters of several distributions expressible in inverse form, Wat. Res., 15 (5), 1049 – 1054.*
- Griffis, V. and Stedinger, J.R. (2007). The LP3 distribution and its application in flood frequency analysis, 2. Parameter estimation methods, *Journal of Hydrologic Engineering, 12(5), 492–500.*
- Holloway, A., Chasi V., de Waal, J., Drimie, S., Fortune, G., Mafuleka, G., Morojele, M., Penicela Nhambiu B., Randrianalijaona, M., Vogel, C. and Zweig, P. (2013). Humanitarian Trends in Southern Africa: Challenges and Opportunities. *Regional Inter-agency Standing Committee (RIASCO), Southern Africa. Rome, FAO.*
- Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics, 27(3), 251-261.*

- Hosking, J.R.M. and Wallis, J.R. (1988). The effect of intersite dependence on regional flood frequency analysis. *Water resource research*, 24, 588-600.
- Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society B*. 52, 105-124
- Hosking, J.R.M. (2015). Regional frequency analysis using L-moments, *Package 'lmomRFA', version 3.0-1*
- Hosking, J.R.M. and Wallis, J.R. (1997). Regional frequency analysis: An approach method based on L-moments. *Cambridge University Press, Cambridge*.
- Jing, D. and Rongfu, Y. (1988). The determination of probability weighted moments with the incorporation of extraordinary values into sample data and their application to estimating parameters for the Pearson Type three distribution. *Chengdu University of Science and Technology, Chengdu, Sichuan (P.R. of China)*.
- Kumar, A. and Chander, S. (1987). Statistical Flood Frequency Analysis: An overview, in Singh, V.P. (ed.) *Hydrologic Frequency Modelling*. D. Reidel Publishing Company. Dordrecht, Holland.
- Kundzewict, Z. (2003). Extreme precipitation and floods in the changing world. *IAHS-AISH publication (281): 32-39*
- Lee, S.H. and Maeng, J. (2003). Comparison and analysis of design floods by the change in the order of LH-moment methods. *Irrigation and drainage*, 52, 231-245.
- Lettenmaier, D.P. and Potter, K.W. (1985). Testing flood frequency estimation methods using a regional flood generation model. *Water resources research*, 21, 1903-14.

- Lettenmaier, D.P., Wallis, J.R., and Wood, E.F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water resources research*, 23, 313-23.
- Loucks, D.P. and van Beek, E. (2005). Water Resources Systems Planning and Management. *An Introduction to Methods, Models and Applications*. Published by United Nations, Educational Scientific and Cultural Organization, Delft Hydraulics, the Netherlands
- Maidment, D. (1993). Handbook of Hydrology. McGraw-Hill , New York.
- Meshgi, and Khalili, D. (2007a). Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. I. A re-visit to regional homogeneity. *Water engineering department, college of agriculture, Shiraz University, Fars, Shira, Iran, 23,119–135*
- Meshgi, and Khalili, D. (2007b). Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. II. Development of LH-moments parameters for the generalized Pareto and generalized logistic distributions. *Water engineering department, college of agriculture, Shiraz University, Fars, Shira, Iran, 23,119–135*.
- MEWT (2011). Second National Communication to the United Nations Framework Convention on Climate Change. *Botswana, Gaborone*.
- Murshed, Md. S., Seo, Y.A. and Park, J.S. (2013). LH-moment estimation of a four parameter kappa distribution with hydrologic applications. Department of Business Administration, Northern University Bangladesh, Dhaka 1209 Bangladesh; Department of Statistics, Chonnam National University, Gwangju 500-757, South Korea.
- NWMPR (2006). Surface Water Resources. *Final Report, 3, Department of Water Affairs, Ministry of Minerals, Energy and Water Resources*.

- Parida, B.P. (1999). Modelling of Indian summer monsoon rainfall using a four-parameter Kappa distribution. *International Journal of Climatology*, 19, 1389-1398.
- Parida, B.P., Kachroo, R.K., and Shrestha, D.B. (1998). Regional flood frequency analysis of Mahi-Sabarmati Basin (subzone 3-a) using Index Flood procedure with L-moments. *Water Resources Management*, 12, 1-12.
- Parida, B.P. (2004). Study of Flood characteristics of selected rivers in Botswana. *Department of Environmental Science, University of Botswana, Gaborone (109)*.
- Potter, K.W. and Walker, J.F. (1981). A model of discontinuous measurement error and its effect on the probability distribution of flood discharge measurement. *Wat. Res.*, 17(5), 1505 – 1509.
- Rahman, A.S., Haddad, K., Rahman, A. (2014). Impacts of Outliers in Flood Frequency Analysis: A Case Study for Eastern Australia. *School of Computing, Engineering and Mathematics, University of Western Sydney*.
- SADC (2006). Regional Flood Watch. *Regional Remote sensing Unit*.
- Shabri, A. (2002). Comparisons of the LH-moments and the L-moments. *Department of Mathematics University Technology Malaysia, 81310 UTM Skudai, Johor, Malaysia*, 18, 33–43.
- Sutcliffe J.V. (1978). Methods of flood estimation: *A guide to the flood studies report (49)*.
- Vogel, R.M. and Kroll, C.N. (1991). The value of streamflow record augmentation procedures in low-flow and flood-flow frequency analysis.
- Vogel, R.M., and Fennessey, N.M. (1993). L-moment diagrams should replace product moment diagrams. *Water Resources Research*, 29 (6), 1745-1752.
- Walfish, S. (2006). A review of statistical outlier methods.

- Wang, O.J. (1997). LH moments for statistical analysis of extreme events. *Water Res. Research*, 33 (12), 2841-2848.
- WMO (1989). Statistical Distributions for Flood Frequency Analysis. *Secretariat of the World Meteorological Organization-Geneva-Switzerland* (718).
- Zafirakou-Koulouris, A., Vogel, R.M., Craig, S.M., and Habermeier, J. (1998). L-moment diagrams for censored observations. *Water Resources: Res.*, 34(5), 1241-1249.