Regional Flood Frequency Analysis for Newfoundland and Labrador Using the L-Moments Index-Flood Method

By

Yang Lu

A thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

May 2016

St. John's **Newfoundland** **Canada**

ABSTRACT

The L-moments based index-flood procedure had been successfully applied for Regional Flood Frequency Analysis (RFFA) for the Island of Newfoundland in 2002 using data up to 1998. This thesis, however, considered both Labrador and the Island of Newfoundland using the L-Moments index-flood method with flood data up to 2013. For Labrador, the homogeneity test showed that Labrador can be treated as a single homogeneous region and the generalized extreme value (GEV) was found to be more robust than any other frequency distributions. The drainage area (DA) is the only significant variable for estimating the index-flood at ungauged sites in Labrador.

In previous studies, the Island of Newfoundland has been considered as four homogeneous regions (A, B, C and D) as well as two Water Survey of Canada's Y and Z sub-regions. Homogeneous regions based on Y and Z was found to provide more accurate quantile estimates than those based on four homogeneous regions. Goodness-of-fit test results showed that the generalized extreme value (GEV) distribution is most suitable for the sub-regions; however, the three-parameter lognormal (LN3) gave a better performance in terms of robustness. The best fitting regional frequency distribution from 2002 has now been updated with the latest flood data, but quantile estimates with the new data were not very different from the previous study.

Overall, in terms of quantile estimation, in both Labrador and the Island of Newfoundland, the index-flood procedure based on L-moments is highly recommended as it provided consistent and more accurate results than other techniques such as the regression on quantile technique that is currently used by the government.

ACKNOWLEDGEMENTS

I am grateful thankful to my parents for their endless love, understanding and support they gave me throughout my studies and my life.

I would also like to thank to my supervisor, Dr. Leonard M. Lye, who provided me with financial help and academic guidance in the process of conducting this research and writing this thesis.

Thanks also go the School of Graduate Studies of Memorial University and to the Institute for Biodiversity and Environmental Sustainability (IBES) for providing partial funding for the study.

I am also grateful to Ms. Jinghua Nie for her sincere help for my life and study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\overline{Q}$	Mean annual peak flow
τ	L-CV
μ	Mean
σ	Standard deviation
ξ	Location parameter of the distribution
α	Scale parameter of the distribution
ε	Error
τ_3	L-skewness (population)
τ_4	Standard deviation of sample regional L-kurtosis
τ_4^{DIST}	Distribution's L-kurtosis
βr	Population probability weighted moment
λr	Population L-moments
A_k	Coefficients of Polynomial Approximations
An	Site Characteristics
B	Bias
B_4	Bias of sample regional L-kurtosis
D_i	Discordancy measure
E	Expected value

F	Non-exceedance probability
F(x)	Cumulative distribution function
H	Heterogeneity measure
h	4^{th} parameter of the kappa distribution
k	Shape parameter of the distribution
In	Natural logarithm
lr	Sample L-moments
Nsim	Number of simulated regions
Q	Flow rate
q	Quantile function
t	Sample L-CV
t_3	Sample L-skewness
t_3^R	Regional average sample L-skewness
t_4	Sample L-kurtosis
t_4^R	Regional average sample L-kurtosis
t^R	Regional average sample L-CV
μv	Mean of simulated sites
V	Weighted standard deviation of at-site sample L-CVs
Z^{DIST}	Goodness-of-fit measure of the candidate distribution

LIST OF ACRNYMS

AARB	Absolute Relative Bias
ACLS	Lakes and Swamps
ARB	Average Relative Bias
AMEC	AMEC Environment & Infrastructure
DA	Drainage Area
DRD	Drainage Density
GEV	Generalized Extreme Value
GLO	Generalized Logistic
GPA	Generalized Pareto
IFM	Index-flood Method
IH	Institute of Hydrology
LAF	Lake Attenuation Factor
L-CV	Coefficient of L-variation
LN3	3-parameter Log Normal
LSF	Lakes and Swamps Factor
NE	Northeast
NERC	Natural Environment Research Council
NW	Northwest

PE3	Pearson Type III
PWMs	Probability Weighted Moments
RFFA	Regional Flood Frequency Analysis
RMSE	Root Mean Square Error
SE	Southeast
SEE	Standard Error of the Estimate
SMR	Regression Correlation Coefficient
SW	Southwest
USGS	United States Geological Survey
WSC	Water Survey of Canada

CHAPTER 1

INTRODUCTION

1.1 General

Accurate estimations of flood quantiles play a significant role in minimizing flood damage, specifically related to casualties, compensation related expenses and environmental damage, which are all caused by flooding. Furthermore, accurate estimations of flood frequencies can provide valuable information for designing and planning hydraulic structures and other flood protection schemes.

Flood frequency analysis was traditionally based on fitting a frequency distribution or probability model to the observed flood data at a single site. However, insufficient data often create a challenge for hydrologists to provide an accurate flood quantile. A preferable approach is to use regional flood frequency analysis (RFFA) to deal with this problem. RFFA uses data at neighboring sites in a defined homogeneous region to develop a model. Flood quantiles at any site within this region can then be derived. Multiple regression models and the index-flood method (IFM) are the prime methods for RFFA. The regression on quantile approach uses regression analysis to develop equations to relate climate and physiographic characteristics to the flow quantiles estimated from single-station flood frequency analysis in a homogeneous region. The index-flood method (IFM) establishes a relationship (growth curve) between the

scaled quantiles and the return period in a homogeneous region. Regionalization, substituting space for time, is regarded as the fundamental premise of RFFA. The L-moments based index-flood method is an advanced approach which has been widely used for flood studies. Recent studies include the regional flood frequency analysis in Sicily, Italy by Noto (2009); the regional flood estimation for ungauged basins in Sarawak, Malaysia by Lim & Lye (2003); and the regional flood frequency analysis for West Mediterranean Region of Turkey by Saf (2009). L-moments are the linear combination of PWMs. Because its parameters are less biased, it has the ability to estimate site characteristics in a simple way; in particular, to estimate distribution parameters. Detailed information about L-moments is presented in Chapter 2.

In general, the application of the IFM should satisfy two assumptions: 1) the data at each site are independent and identically distributed; 2) the frequency distribution at each site should be identical except for the scale factor. Based on Hosking and Wallis (1997), the index-flood method based on L-moments has the following steps:

1) Screening the data: The objective is to check for gross errors of the data and to make sure the data is continuously available over time. That is, there is no gap or missing data.

2) Identifying the homogeneous region: Deciding on which river basins can be grouped together as a homogeneous region. That is, flood data with approximate identical distribution except for scale.

3) Choosing a frequency distribution: Since the regional frequency distribution is essentially determined by the L-moment ratio diagram, a goodness-of-fit test will determine how well the selected distribution fit the data in the region. The application of robustness test can become necessary when there is more than one acceptable regional frequency distribution.

4) Estimating the frequency distribution: This process is designed to compute the flood quantiles for certain return periods at ungauged sites derived from the regional growth curve.

1.2 The application of RFFA for Newfoundland and Labrador

The first regional flood frequency analysis for Newfoundland was performed by Poulin (1971). Subsequent to this, regular updates by the provincial government of Newfoundland were carried out in 1984, 1990, 1999 and 2014 (Government of Newfoundland and Labrador, 1984; Government of Newfoundland and Labrador, 1990; Government of Newfoundland and Labrador, 1999 & AMEC, 2014). The regression method as described earlier, based on the observed data and sites characteristics, was the prime methodology used. However, this methodology, while easy to understand and apply often suffers from lack of consistent results and accuracy due to short historical data and other statistical issues.

The first RFFA for Newfoundland based on the L-moments index-flood procedure

was proposed by Pokhrel (2002). The regional divisions in this analysis were based on two references: 1) the division of four sub regions (A, B, C and D) used in the provincial government analysis in 1989, and 2) the division of two sub regions Y and Z suggested by the Water Survey of Canada (WSC). This research concluded that the WSC sub regions obtained more accurate quantile estimations than sub regions suggested in 1989. The generalized extreme value (GEV) distribution was also found to have a superior performance compared to the lognormal (LN3) distribution for the regions of 1989. The comparison between the at-site and regional estimates showed that the L-moments based IFM has the ability to provide more accurate quantile estimation for the ungauged sites than conventional regression models, and it obtained more accurate results than the study in 1989.

1.3 Rationale and objectives

As introduced in Section 1.2, the L-moments based RFFA had been successfully applied in in Newfoundland and Labrador, therefore, in this thesis, due to the excellent performance and the worldwide application of the L-moments based index-flood approach, the regional flood frequency analysis for the Island of Newfoundland will be updated with the latest data up to 2013. For Labrador, a RFFA using the L-moments based index-flood approach will be used for the first time to obtain flood quantile estimates for ungauged basins. These results will be compared

to those based on the regression based approach recently completed by AMEC (2014). The two main objectives of this thesis can thus be summarized as follows:

1) Update the quantile estimates at both gauged and ungauged sites for the Island of Newfoundland via the L-moments based index-flood procedure of RFFA. The updated results will be compared to those obtained by Pokhrel (2002) and those recently obtained by AMEC (2014).

2) Develop the first regional flood frequency analysis for Labrador using the L-moments based index-flood method and compare the results with those obtained using the regression method developed by AMEC (2014).

1.4 Outline

This thesis has six chapters. Chapter 1 provides a general introduction to regional flood frequency analysis in Newfoundland and Labrador and the methodologies used. It also provides objectives for the study and an outline of the thesis. Chapter 2 reviews recent and related research in the field of regional flood frequency analysis and application of RFFA in Newfoundland and Labrador. Popular methodologies used for RFFA, the application of index-flood procedure in hydrologic research, the commonly used methods of fitting frequency distribution models and methods for identifying homogeneous regions are also discussed. Chapter 3 describes the

methods used for the L-moments based index-flood procedure in a step by step manner. Chapter 4 presents the results of RFFA in Labrador and estimates the index flood at ungauged sites using a nonlinear regression model. A comparison with results from AMEC (2014) will also be presented. Updated results of the RFFA for the Island of Newfoundland will be shown in the Chapter 5, as well as the comparison of quantile results with those of Pokhrel (2002). Chapter 6 summarizes the results and provides conclusions and recommendations regarding the application of the L-moments based index-flood of RFFA in Newfoundland and Labrador. Limitations of this research are also discussed. A list of the references and programming codes used in this thesis are presented as appendices.

CHAPTER 2

LITERATURE REVIEW

2.1 General

This chapter first reviews some of the key steps in regional flood frequency analysis and some of the literature for each step. This is then followed by a brief review of RFFA that have been conducted in Newfoundland and Labrador. The estimation of regional flood frequency equations or curves has always been a popular topic among hydrologists and even among some statisticians. Two main methodologies in use today for RFFA are: 1) regional quantile regression approach which became popular with the advent of computers for performing multiple regression analysis; and 2) the index-flood approach which describes a regional quantile growth curve estimated graphically or by statistical methods. Since the development of the statistically sophisticated L-moments index-flood approach by Hosking and Wallis (1997), this approach is perhaps the most widely used worldwide today. Since the objectives of this thesis involve the use of the L-moments based index-flood procedure, the steps involved with this approach will be followed and literature pertaining to each step reviewed. RFFA using the L-moments based index-flood approach is carried out based on the following six steps:

1) Screen the data;

2) Define a homogeneous region;

3) Perform the homogeneity test for each proposed region;

4) Select a regional frequency distribution for each region and check for robustness;

5) Estimate flow quantiles for both gauged and ungauged sites; and

6) Verify and assess accuracy of quantile estimation.

In the next section, the literature review is based on the steps mentioned above.

2.2 Screening the data

Data screening is the first step to be taken in any data analysis. Before starting the work, one should ensure that the data is appropriate for the analysis. Questions to be asked include: 1) Are the environmental data of sufficient quality and quantity and do they follow the same frequency distribution? 2) Have the data changed over time? For example, the data used in this thesis may have been adjusted or corrected by the Water Survey of Canada (WSC). It is possible that the WSC did not just update the historical data to the year 2013; they may have also modified the record for every year at each station. For RFFA, Hosking and Wallis (1997) suggested three kinds of useful checks: 1) check the data individually to find gross errors (Wallis, et al. 1991); 2) check the data at each site for outliers and repeated values; 3) check for trends and

abrupt changes and compare the data between sites.

The existence of extreme values or outliers may bring bias to the estimation, but to simply discard the outliers may distort results (Kirby, 1974). Therefore, tests become necessary to screen out outliers and then to check whether they can be accepted within a homogeneous group. There are many tests for outliers. For example, the U.S. Water Resource Council (1981) used a statistical hypothesis test in flood frequency estimation, which compared the difference between the outliers and other values in a sample. A method based on a so-called “masking effect” was applied successfully by Barnett and Lewis (1994), which had an ability to distinguish multiple outliers. The sum of square statistics (Grubbs, 1950) and extreme-location statistics (e.g. Epstein, 1960a & 1960b) are other tests for outliers. Hosking and Wallis (1997) reported that double-mass plots or quantile-quantile plots are also well-known methods for detecting outliers which are easy to apply. Boxplot, histogram plot and dot plot provided in some statistical software can also work well for detecting outliers. Another alternative method is the L-moment ratios (Hosking and Wallis, 1997), which is designed to detect unusual sites from a group of sites by comparing their individual L-moment ratios with the regional L-moment ratios of a group. The detailed principle and application of L-moments will be discussed later in Chapter 3.

2.3 Definition of a homogeneous region

For RFFA it is common that the data of some sites are insufficient to provide reliable estimation. Therefore, identifying a homogenous region is a good way to transfer information from other available neighboring stations. Hosking and Wallis (1997) summarized some commonly used grouping methods to decide on a homogeneous region.

2.3.1 Geographical convenience

Delineating a homogeneous region based on geographical convenience is a direct and traditional method. The definition of geographical convenience usually means the administrative area (Natural Environment Research Council, 1975; Beable and McKerchar, 1982), political or physiographic boundaries. However, for larger areas the variability of physical or physiographic site characteristics may be large; therefore, the identification of a homogeneous region simply depending on geographical parameters is rarely used in recent studies. Attempts to define a homogeneous region based on geographical parameters are usually accompanied with a goodness-of-fit test or hypothesis test to make sure the defined sub-regions are reasonable and unbiased for RFFA.

2.3.2 Clustering techniques

Cluster analysis can be hierarchical and non-hierarchical (Downs & Barnard, 1992). It is a very developed and widely used technique of dividing a data set into groups or to combine several data sets into a group based on similar data vectors (site characteristics or at-site statistics) (Hosking and Wallis, 1997). This technique has been used in many hydrological studies worldwide. For example, Jingyi and Hall (2004) used Ward Linkage clustering, in addition to Fuzzy C-Means and Kohonen neural network to successfully delineate homogenous regions in the southeast of China. Hierarchical clustering analysis was used for regional estimation in Mexico (Quarda, T et al. 2008); and Bharath (2015) completed the delineation of homogeneous regions in India using wavelet-based global fuzzy cluster analysis. Burn (1989) identified homogeneous regions by combining cluster analysis and basin similarity measures. The hybrid-cluster and K-means algorithm was recommended by Rao & Srinivas (2006) for regionalization in Indiana, USA; and a fuzzy clustering approach was applied by Srinivas et al. (2008) to identify the regions of watersheds for flood frequency analysis. Noto and Loggia (2009) divided five regions using cluster analysis in Sicily, Italy; and Luis-Perez et al. (2011) applied two kinds of clustering techniques to delineate homogeneous regions in the Mexican-Mixteca region while Basu, B. and Srinivas, V. V. (2014) used kernel-based fuzzy clustering analysis to identify the homogeneous groups of watersheds in the U.S. Other related studies that used cluster analysis include Shu and Burn (2004), Wiltshire (1986),

Bhaskar and O'Connor (1989), Mosley (1981), Tasker (1982), Nathan and McMahon (1990), Richman and Lamb (1985), Kalkstein et al. (1987), Burn and Goel (2000), Lim and Lye (2003), and Fovell & Fovell (1993).

2.3.3 Subjective partitioning

This method of delineating homogeneous regions requires the sites to have similar site characteristics. Similar site variables include the amount of rainfall, drainage area, timing of floods, forested areas, etc. Gingras, Adamowski and Pilon (1994) used the time of year when the largest flood occurred as the parameter to delineate sub regions in Ontario and Quebec. De Coursey (1972) formed groups of basins with similar flood responses in Oklahoma. For RFFA, a heterogeneity test is usually carried out after using a subjective partitioning method. But as Hosking and Wallis (1997) mentioned, when the at-site statistics are used as the basis for subjective partitioning, the validity of the use of the heterogeneity measure may be affected in validating the regions.

2.3.4 Objective partitioning

This method is designed to divide sites into groups depending on whether their site characteristics exceed one or more threshold values. Mailhot et al. (2013) applied the

approach of Peak-Over-Threshold (POT) to estimate intense rainfall in southern Quebec. Pearson (1991b) used this procedure and successfully analyzed small basins' grouping in New Zealand. Similar to subjective partitioning, it is recommended that heterogeneity tests be carried out for the delineated regions when using this method of partitioning (Hosking and Wallis, 1997).

2.3.5 Other grouping methods

Other alternative methods of defining homogeneous regions include the method of residuals, canonical correlation analysis, and region-of-influence (ROI) (Basu, 2014), among others. For example, White (1975) grouped basins based on the factor analysis of the site characteristics in Pennsylvania; and Burn (1988) applied principal components analysis to group gauged sites, depending on which subjectively rotated set of principal components a site's annual maximum streamflow most closely resembled.

2.4 Homogeneity test for proposed regions

A homogeneous region is a fundamental requirement for quantile estimation. Once the regions or sub-regions are identified, a homogeneity test is needed to make sure that the delineated regions and subsequent analysis are appropriate and meaningful.

Multiple methods have been used for testing the degree of homogeneity of a region. Dalrymple (1960) proposed the first test that fitted the Gumbel distribution as the underlying distribution to every studied site. Chow (1964) tested the homogeneity by analyzing the sample coefficients of variation (Cv) and /or skewness (Cs). Lu (1991) used the L-moment ratios and normalized 10-year flood estimate to conduct a regional homogeneity test. Lu and Stedinger (1992) carried out a homogeneity test based on the sample variance and normalized 10-year flood quantile estimators. Fill and Stedinger (1995) compared the power of the Dalrymple test, normalized quantile test and a method of moment Cv test. Scholz and Stephen (1987) proposed the Anderson-Darling test (Anderson and Darling, 1954) for testing the homogeneity of samples. However, the most popular method is the L-moment ratios based heterogeneity test proposed by Hosking and Wallis (1993, 1997), which has been widely used in hydrological studies. For example, Gabriele and Chiaravalloti (2012) used this method to test the degree of homogeneity based on the rainfall sample data within regions. Abolverdi and Khalili (2010) tested the degree of homogeneity in southwestern Iran based on the regional rainfall annual maxima, among many other studies. The detailed formulation of this test will be discussed in Chapter 3.

2.5 Selection of regional frequency distribution

The appropriate selection of a regional flood distribution has a direct impact on the

quantile estimation at gauged and ungauged sites. A regional frequency distribution is fitted from a single site to other sites within the homogeneous region. As Hosking and Wallis (1997) mentioned that there may be more than one acceptable candidate regional distributions, and the best fitting distribution is the one with the ability to reflect the “true” distribution. Therefore, rather than identifying a “true” distribution, the aim is to determine a distribution which will provide the most approximate fit to the observed data and yield a more accurate quantile estimation for each single site.

Sveinsson (2001) compared the quantile estimation based on the population index flood fitted by the GEV distribution using Hosking and Wallis’s (1997) index flood regional PWM procedure. The Log Pearson Type (III) was used for peak flood discharge based on Bulletin 17 B in the U.S. (Lim and Voeller, 2009). Ashkar and Quarda (1996) discussed the use of the generalized Pareto distribution for flood frequency analysis. Griffis and Stedinger (2007) fitted LN3 distribution to the flood quantile estimation using the weighted Bulletin 17B procedure. Peel et al. (2001) compared multiple distributions based on two graphical different methods and found that using graphical methods with an L-moment ratio diagram can distort the choice of regional distribution of observed data. In a regional flood frequency analysis of the west Mediterranean region of Turkey, Saf (2009) found that Pearson type III distribution fitted well to the Antalya and lower-west Mediterranean, and that Generalized Logistic distribution was most suitable for the upper-west Mediterranean. In a Canada-wide study, Yue and Wang (2004a, b) fitted the generalized extreme

value (GEV) for the Pacific and southern British Columbia mountains, the 3-parameter lognormal distribution to the northwestern forest area, the Wakeby distribution to the Arctic tundra, the Pearson type Ⅲ to the Prairies, Northeastern forest, Great Lakes, and regions in St. Lawrence, Atlantic and Mackenzie. Atiem and Harmancioglu (2006) fitted five different distributions---generalized Pareto (GPAR), generalized extreme value, generalized logistic, generalized normal and PE3 to the annual maximum flood data for 14 sites in the Nile River tributaries based on the index-flood method.

The application of the moment-ratio diagram introduced by McCuen (1985) provides a quick and basic approach to judge how candidate distributions fit the data. Hosking (1990) recommended using L-moments which is a linear combination of the ranked observed data and exhibits less bias than the traditional moments. The L-moment ratio diagram is a simple plot of τ_4 against τ_3 (L-kurtosis and L-skewness) for commonly used distributions, and the at-site and regional average L-moment ratios can be plotted to compare with the population values of commonly used distributions (Hosking and Wallis, 1997). Since the use of the L-moment ratio diagram is a quick and basic approach to select a regional distribution, the final determination must rely on further goodness-of-fit and robustness tests. Goodness-of-fit tests include quantile-quantile plots, Kolmogorov-Smirnov, chi-squared and the most popular L-moments based tests introduced by Hosking and Wallis (1997).

The L-moment ratios based goodness-of-fit test is designed to test whether a given

regional distribution can provide a close fit to the data using a simulation process. Using parts of the approaches mentioned above, Malekinezhad (2011) determined that the generalized extreme value distribution was the best fit for flood estimation in the Namak-Lake basin. Atiem and Harmancioglu, (2006) found that the generalized logistic distribution provided the best fit for the data in the River Nile. Mkhandi and Kachroo (1997) found that the Pearson type Ⅲ was the most suitable distribution for regional flood in southern Africa. In another study, the Generalized Normal distribution was identified as the best fit for the flood data in the Mahi-Sabarmati Basin (Parida, et al. 1998). If there is more than one acceptable distribution, the robustness test (to be described in Chapter 3) is suggested when the underlying distribution is different from the selected one.

2.6 Quantile flow estimation for both gauged and ungauged sties

The index-flood procedure plays a key role in the estimation of flow quantiles. Once the studied region is found to be homogeneous and the regional frequency distribution has been determined, and it is assumed that the frequency distribution of all sites in the region is identical except for the site-special scaling factor known as the index flood. The index flood is usually the mean annual flood or the median annual flood. The flow quantiles can be estimated as the product of the index flood and regional growth curve or regional frequency distribution function. Early

applications of the index-flood procedure include Dalrymple (1960) and NERC (1975), Hosking et al. (1985), Jin and Stedinger (1989), Wallis and Wood (1985), Letttenmaier and Potter (1985). Cunnane (1988) and Pitlick (1994) demonstrated successful applications of the index-flood procedure for regional flood frequency analysis. Madsen et al. (1997) illustrated the advantages of the use of index-flood procedure in terms of both annual flood series and partial duration series. Portela and Dias (2005) described six homogeneous regions and used the data of annual maximum flood series of 120 Portuguese stream gauging stations in mainland Portugal using the index-flood method. Later, Hosking and Wallis (1997) successfully introduced the use of L-moments in the index-flood procedure and it was shown to be robust in the presence of any extreme values and outliers. Recent regional flood studies based on index-flood procedure include studies in the U.S.A. (Vogel et al. 1993; Vogel and Wilson, 1996), Malaysia (Lim and Lye, 2003), Australia (Pearson et al. 1991), Southern Africa (Mkhandi and Kachroo, 2000), New Zealand (Pearson, 1991, 1995; Madsen et al, 1997) and Turkey (Saf et al. 2009).

Estimation of flow quantiles at gauged sites with short records can be completed directly from an estimate of the index flood using the annual maximum or peaks-over-threshold values. For the ungauged sites where their index floods are not available, the most commonly used method is a regression model. The index flood of gauged sites is regressed against their respective catchment or site characteristics (e.g., basin area, length, basin slope, drainage density, etc.) to obtain a model relating

the index flood to basic characteristics. For example, IH (1999) developed regression models relating median flow to five different site characteristics. Brath et al (2001) reviewed different indirect methods of estimating the index flood at gauged sites and concluded that the regression method had a better performance than other approaches. The regression method was used for regional flood frequency analysis in Sicily (Noto and Loggia, 2009) and regional flood estimation for ungauged basins in Sarawak, Malaysia (Lim and Lye, 2003), among many others.

2.7 Verification and assessment of accuracy of quantile estimation

Accuracy of assessment is always needed for model evaluation. The factors that have an influence on accuracy of assessment are: 1) the regions are not adequately homogeneous, 2) the regional frequency distribution is not robust, and 3) the availability of data is limited. Assessment accuracy based on traditional statistics involves constructing confidence intervals for estimated parameters or quantiles on the assumption that all the statistical assumptions of the models are satisfactory. However, in practice, it is found that in most cases it is difficult to ensure that the models used are the "correct" ones (Hosking and Wallis, 1997). Although it is difficult to establish common criteria for model evaluation, the evaluation is still carried out in some studies based on some specific statistics such as sensitivity analysis and model calibration. For example, Gupta et al. (1999) calibrated

hydrological models using the shuffled complex evolution automatic procedure; Motovilov, et al. (1999) verified hydrological model ECOMAG with the use of standard meteorological and hydrological data in the NOPEX southern region; Van Liew et al. (2007) used two sub watersheds in the Little Washita River Experimental Watershed (LWREW) to calibrate the parameters of the Soil and Water Assessment Tool (SWAT) and Hydrologic Simulation Program-Fortran (HSPF) models in southwestern Oklahoma. Other similar research includes Santhi et al. (2001) and Singh et al. (2004). However, nobody used the acceptable ranges of values for each statistic until a review of the values for various statistics used was provided by Borah and Bera (2004).

A good model evaluation entails satisfying the following conditions: 1) it must be robust and acceptable to various constituents and climatic conditions, 2) be commonly used and recommended by various studies, and 3) be robust in model evaluation (Moriasi et al. 2007). Boyle et al. (2000) recommended the estimation of residual variance (the difference between the measure and simulated values) which can be estimated by the residual mean square or root mean square error (RMSE). Hosking and Wallis (1997) suggested the Monte Carlo simulation approach to assess the accuracy by calculating the RMSE when the region is not homogeneous enough, the regional frequency distribution is misspecified or the observed data are statistically dependent. Chapter 3 provides detailed information of the Monte Carlo simulation approach based on L-moments. The assessment of RMSE has been

widely applied in various studies; for example, Fill and Stedinger (1998) showed that the empirical Bayes estimator had the same or a better performance than the simpler normalized quantile regression estimator for sites with shorter records based on the results of RMSE; Saf (2009) developed a Monte Carlo simulation and evaluated the accuracy of the quantile estimates based on the relative root-mean-square error and relative bias; and Atiem and Harmancioglu (2006) evaluated the results of quantile estimation by assessing the RMSE% which is also based on the Monte Carlo simulation approach.

Other methods for example, the slope and y-intercept of the best-fit regression line, can indicate how the simulated data match the observed data on the assumption that the observed and simulated data are linearly related (Moriasi et al. 2007). The use of Pearson's correlation coefficient (r) and coefficient of determination (R^2) to measure the degree of linear collinearity between simulated and observed data are also popular; The index of agreement (d) (Willmott, 1984) that measures the degree of model prediction error and varies between 0 and 1 and Nash-Sutcliffe efficiency (NSE) that is designed to compare the relative magnitude of the residual variance ("noise") to the measured data variance ("information") (Nash and Sutcliffe, 1970) are also widely used ; Persistence model efficiency (PME)-a normalized model that evaluates the relative magnitude of the residual variance ("noise") to the variance of the errors obtained using a simple persistence model (Gupta et al, 1999) and the Prediction efficiency (Pe) (Santhi et al. 2001) that can determine how well the

simulated data can fit the observed data are also possible.

2.8 RFFA for Newfoundland

RFFA for the Island of Newfoundland was conducted in 1971, 1984, 1989, 1999, 2002 and 2014 by the Government of Canada, Newfoundland or its consultants (Poulin, 1971; Government of Newfoundland, 1984; Government of Newfoundland, 1990, Government of Newfoundland and Labrador, 1999; Pokhrel, 2002 & AMEC, 2014). The study in 2014 was the first to include the Labrador. The first provincial flood frequency research was by Poulin (Government of Canada, 1971) used the classical index-flood approach. In that study, the Island of Newfoundland was analyzed as one region with 17 gauged stations. The index flood which was the mean flows was used to develop a function relating the mean (Q) and the drainage area (DA).

In a subsequent study (Government of Newfoundland, 1984), the Island of Newfoundland was sub-divided into two regions, a North and a South region. Twenty one gauged stations were analyzed based on the regression on quantiles approach. Single-site flood frequency was performed for each station to obtain estimates of several key quantiles. Then these were regressed against site characteristics such as drainage area (DA) and latitude for the North region, and drainage area (DA), area controlled by lakes and swamps (ACLS) and slope in the South region. However,

Lye and Moore (1991) noted that the log-transformed model and the use of mean flow as the predictor value were not properly carried out which would lead to bias, and the variable latitude was not suitable for North region given that it is a very small area.

A new regional flood frequency analysis conducted by the provincial government in 1989 increased the number of gauged stations up to thirty-nine. This study divided the Island of Newfoundland into four regions (A-Avalon and Burin Peninsula; B-central region of the Island; C-Humber valley and northern peninsula; and D-the southwestern region of the island) taking into account the availability of data, the timing of regional floods and physiographic factors such as flood characteristics, amount of precipitation and results of regression analysis. The average record length was 21 years and the record was extended in some stations with short records. The drainage area (DA), lakes and swamps factor (LSF), drainage density and slope were included for the regression on quantiles.

In an updated study by the Government in 1999, the four sub regions of 1989 were renamed. The new names are -northwest (NW), northeast (NE), southeast (SE) and southwest (SW) and refer to the previous C, B, A, and D regions proposed in 1989. The drainage area (DA) and lake attenuation factor (LAF) were found to be significant predictors and the LSF was a significant variable only in the SW region.

Instead of the regression on quantiles approach, Pokhrel (2002) conducted a regional

flood frequency analysis of the Island of Newfoundland using annual peak flow until 1998 based on the L-moments index-flood procedure suggested by Hosking (1990). The Island of Newfoundland was sub-divided based on two kinds of regionalization -four regions as used by the Government of Newfoundland in 1989 and 1999, and the Y and Z regions suggested by the Water Survey of Canada (WSC). The determination of homogenous regions, selection of regional frequency distribution and quantile estimation were all based on the L-moments based index-flood method. It was found that for sub region Y, only the drainage area (DA) and drainage density (DRD) were significant at α=5% in terms of estimating the index flood. For sub region Z, in addition to DA and DRD, the lakes and swamps factor (LSF) was significant as well. The study also showed that the L-moments based index-flood approach with the Y and Z regions was superior to that of the regression on quantile approach and use of four sub-regions.

The latest provincial RFFA (AMEC, 2014) was conducted for both Newfoundland and Labrador. This study also used the regression on quantile approach, but with data up to 2012. Hence newly updated regression models were obtained. Seventy-eight gauged stations in Newfoundland and twelve gauged stations in Labrador were used in the study. Regression equations were obtained considering the Island of Newfoundland as a single homogeneous region and considering it as four sub hydrological homogeneous regions as proposed in 1999. Drainage area (DA) and lake attenuation factor (LAF) were significant for the NW, SE and NE sub regions,

whereas the lakes and swamps factor (LSF) and DA were shown to be significant for the SW region. Labrador was analyzed as a single homogeneous region and only DA was significant for developing regression equations. The study did not compare results to those obtained by Pokhrel (2002) nor were any robustness tests conducted.

CHAPTER 3

METHODOLOGY

3.1 General

The index-flood method (IFM), a widely used regression method for regional frequency flood analysis, was first proposed by the U.S. Geological Survey (Dalrymple, 1960). Many successful applications show that the IFM has the ability to define a more reliable homogeneous region in which the variability of the at-site data at gauged sites objectively exists. The quantile estimation at each gauged site can be derived directly from the regional flood quantile function, even for the ungauged sites. Hence determining the flood quantiles within a defined homogeneous region is possible.

The detailed modern procedures of the IFM suggested by Hosking (1990) and Hosking and Wallis (1997) will be introduced in this chapter. The L-moments, a modern and advanced mathematical statistics approach are involved to determine the homogenous regions, selection of the regional flood frequency distribution and quantile flows for both gauged sites and ungauged sites.

3.2 Regional flood frequency analysis

The development of a regional flood frequency analysis (RFFA) has proved to be an effective method for estimating flood quantiles at ungauged sites or sites with insufficient streamflow data using the flood information at neighboring sites within a homogeneous region. Compared to the traditional at-site estimations, the regional data can minimize the standard error of interest. Regional regression models and the index-flood procedure are commonly used for RFFA in previous and recent flooding studies. The regression approach develops regression equations to relate at-site climate and physiographic characteristics to flow quantiles from each single site within a homogeneous region. However, the regional regression approach sometimes has a limited ability to provide reliable estimations when the numbers of gauged stations are insufficient. Uncertainties and bias are inevitable. The modern index-flood procedure however successfully avoids these disadvantages. Instead, the flood quantiles at gauged sites can be achieved based on relationship between the quantile function of the regional frequency distribution and index flood at each site. Even for the ungauged sites, the quantiles estimation can be easily achieved using estimated index flood. According to Hosking and Wallis (1997) the quantile estimates can be obtained from:

$$Qi(F)=u_i * q(F) \qquad i=1, 2, 3\ldots N \qquad [3.1]$$

where u_i is the index flood at sites i in a homogenous region with N sites and q(F) is

the regional growth curve. The index flood at ungauged sites can be obtained by establishing a linear or nonlinear regression relationship between sites characteristics and the index flood at gauged sites within a homogeneous region. The application of the index-flood procedure follows an important assumption that all the sites in a defined region are distributed ideally except for a scale factor. Multiple recent studies show that index-flood approach can produce a more accurate and reliable quantile estimation than the regression on quantile approach (e.g. Pokhrel, 2002; Noto, 2009; Lim & Lye, 2003; Saf, 2009).

3.3 L-moments

L-moments are the linear combination of probability weighted moments (PWMs) which is widely used in fitting frequency distribution, estimating distribution parameters and hypothesis testing in flood frequency analysis. Greenwood et al. (1979) defined the PWMs as:

$$\beta r = E\{X\,[F(x)^r]\} \qquad [3.2]$$

where F(x) is the cumulative distribution function for X. X(F) is the inverse CDF of X evaluated at the probability F. βr equals to the mean stream flow when r=0.

Later, Hosking (1990) modified the “probability weighted moments” as:

$$\beta r=\frac{1}{n}\sum_{j=1}^{n-r}\left[\frac{\binom{n-j}{r}}{\binom{n-1}{r}}\right] x(j) \qquad r=0, 1, \dots , \qquad [3.3]$$

where x(j) is the ordered stream flow; βr is the probability weighted moments; n is the sample size, and j is the order of the observed steam flow.

The L-moments are generally defined as Eq. [3.4]. The first four L-moments are the mean of distribution, measure of scale, measures of skewness and kurtosis respectively, which are defined in Eq. [3.5].

$$\lambda_{r+1}=\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}\binom{r+k}{k}\beta_r \qquad r=0, 1, \dots , \qquad [3.4]$$

$$\lambda_1=\beta_0 \quad \lambda_2=2\beta_1-\beta_0 \quad \lambda_3=6\beta_2-6\beta_1+\beta_0 \quad \lambda_4=20\beta_3-30\beta_2+12\beta_1-\beta_0 \qquad [3.5]$$

Additionally, the dimensionless L-moments called L-moments ratios including L-CV, L-skewness and L-kurtosis shown in Eq. [3.6] also play key roles in the estimations of parameters of candidate distributions and the determination of the regional flood frequency distribution. In particular, the L-moment ratio diagram, plot of sample L-moment ratios, average L-moment ratios and theoretical L-moment ratios curves of candidate distributions on a single graph provides an essential visual tool to distinguish among the candidate distributions.

$$\text{L-CV}=\lambda_2/\lambda_1 \quad \text{L-skewness}(\tau_3)=\lambda_3/\lambda_2 \quad \text{L-kurtosis}(\tau_4)=\lambda_4/\lambda_2 \qquad [3.6]$$

The applications of L-moments show great advantages over conventional moments

(C-Moments). Hosking (1990) concluded that L-moments and L-moment ratios are nearly unbiased even for highly skewed observations. They have less sensitivity to the sample size and extreme observations and are more robust to outliers. The L-moment ratio diagram had been shown to be a useful tool to distinguish among candidate distributions by plotting sample L-moment ratios (L-skewness and L-kurtosis) and comparing them with theoretical L-moment ratios curves of candidate distributions. The theoretical L-moment ratios curves of commonly used candidate distributions on the L-moment ratio diagram are shown in Figure 3.1. GPA- generalized Pareto; GEV- generalized extreme-value; GLO- generalized logistic; LN3- lognormal; OLB- overall lower bound of τ_4 as a function of τ_3; and PE3- Pearson typeIII.

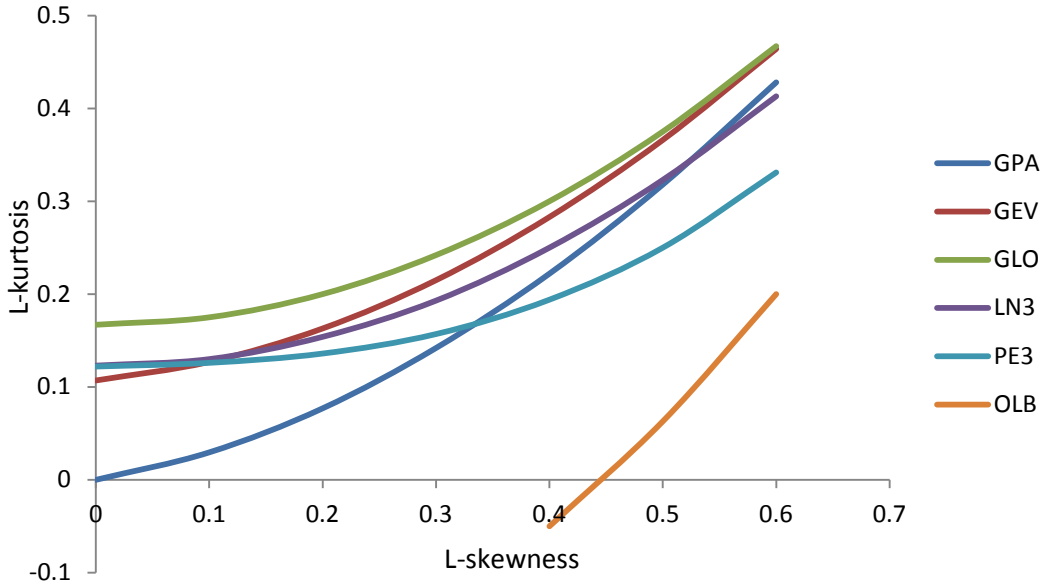


Figure 3.1 L-moment ratio diagram (after Hosking and Wallis, 1997)

Compared to PWMs, the applications of L-moments are more convenient and simpler for measuring the shape and scale of the observations. As noted in the introductory chapter, the applications of L-moments for the index-flood procedure are represented by the following steps (Hosking and Wallis, 1997):

1) Screening the data and use of discordancy measure;

2) Plotting sample L-moment ratios on Figure 3.1 to select a tentative regional frequency distribution;

3) Utilizing the regional homogeneity test based on Monte Carlo simulation to test the homogeneity of the region, and

4) Applying the goodness-of-fit test and robustness tests to determine the final regional flood frequency distribution.

3.4 Procedures for the index-flood based RFFA

Hosking and Wallis (1997) suggested that the application of index-flood based RFFA should follow the key assumption that all of the observations within a defined homogenous region are ideally distributed except for a scale factor (index flood). The procedure of index flood estimation uses the following steps:

1) Screening the data and discordancy measure.

2) Definition of a homogeneous region.

3) Selection of a regional frequency distribution, and

4) Quantile flow estimation for both gauged and ungauged sites.

3.4.1 Screening the data and discordancy measure

The data used for the RFFA is required to represent the true quantity being measured, and all of the observations should follow the same distribution. Basically, the purpose of screening the data is to satisfy three requirements: 1) the data collected for analysis are correct, 2) there are no extreme values or outliers, and 3) the data did not change over time. Hosking and Wallis (1993) first proposed the L-moments based discordancy measure (Eq. [3.7]) to identify unusual sites with different L-moment ratios from other sites within a region. The discordancy measure can be calculated using the Matlab program code (Appendix A-1).

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \qquad [3.7]$$

where D_i is the discordancy measure, $u_i = [t^{(i)}\ t_3^{(i)}\ t_4^{(i)}]^T$ is a vector of t, t_3 and t_4 for site i in a region with N sites and $\bar{u}$ is the unweighted group average which can be defined as:

$$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i \qquad [3.8]$$

And A is the covariance matrix ofu_i, given by

$$A = \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \qquad [3.9]$$

Applying this measure, the unusual sites with inconsistent L-moments ratios due to incorrect records or gross error can be screened out; then for the unusual sites, they might be removed or be included in another region based on the further investigation. Hosking and Wallis (1997) stated that the conclusion reached based on the discordancy measure largely depends on the number of sites in a region. Generally, the algebraic bound of D_i should satisfy:

$$D_i \leq (N-1)/3 \qquad [3.10]$$

The sites can be regarded as discordant from the remaining sites if the D_i value is larger than the critical value shown in Table 3.1. They also suggested that the $D_i \geq 3$ is only suitable for regions with 11 or more sites.

Table 3.1 Critical values of discordancy measure with N sites (Hosking and Wallis, 1997)

Number of sites in a region	Critical value	Number of sites in a region	Critical value
5	1.333	6	1.648
7	1.917	8	2.140
9	2.329	10	2.491
11	2.632	12	2.757
13	2.869	14	2.971
>15	3		

3.4.2 Delineation of homogeneous regions

The delineation of a homogeneous region is a prime step for regional flood frequency analysis. To determine whether a proposed region is homogeneous or not, Hosking and Wallis (1993) suggested a heterogeneity test which aims to assess the degree of homogeneity by comparing the between-site variations in sample L-moment ratios for the sites in a group with what the expected value would be in a definitely homogeneous region. The between-site variation of L-moment ratios is measured by calculating the standard deviation (Eq. [3.12]) of sample L-CVs.

The principle of heterogeneity test (Hosking and Wallis, 1997) can be described as: assume a region has N sites. Each site has the record length of n_i. $t^{(i)}$, $t_3^{(i)}$ and $t_4^{(i)}$ represent the sample L-moment ratios respectively, of which the weighted regional average L-moment ratios are defined as:

$$t^R=\sum_{i=1}^{N} n_i*t^{(i)}/\sum_{i=1}^{N} n_i \qquad [3.11]$$

$$t_3^R=\sum_{i=1}^{N} n_i*t_3^{(i)}/\sum_{i=1}^{N} n_i$$

$$t_4^R=\sum_{i=1}^{N} n_i*t_4^{(i)}/\sum_{i=1}^{N} n_i$$

The standard deviation of the at-site sample L-CVs is given by:

$$V= \{\sum_{i=1}^{N} n_i (t^{(i)}-t^R)^2/\sum_{i=1}^{N} n_i\}^{1/2} \qquad [3.12]$$

To calculate what would be expected in a homogeneous region, Hosking and Wallis (1997) recommended the use of the Monte Carlo simulation which is used to generate a large number of regions (Nsim=500). In the simulated regions each site is required to have the same record length as the sample sites. Then fit the four parameters kappa distribution to the simulated sites. Matlab program code (Appendix A-2) is employed to carry out the simulation process.

Thus, the heterogeneity measure H can be calculated as:

$$H= \frac{(V-\mu v)}{\sigma v} \qquad [3.13]$$

where μv and σv are the mean and standard deviation of simulated sites.

To determine whether a region is homogeneous, Hosking and Wallis (1997) provided the critical value of H which indicated that the region can be declared as “acceptably homogeneous” if $H<1$; “possibly heterogeneous” if $1\leq H<2$, and “definitely heterogeneous” if $H\geq 2$.

3.4.3 Selection of regional frequency distribution

After the homogeneous region is determined, the next step of the index-flood method is to select a regional frequency distribution based on the regional data. As proposed by Hosking and Wallis (1997) the delineation of regional frequency distribution can be completed with three steps: an L-moment ratio diagram, a goodness-of-fit test and a robustness test.

3.4.3.1 L-moment ratio diagram

The L-moment ratio diagram has the ability to provide an elementary visual judgement of a regional frequency distribution by plotting the sample L-moment ratios and average sample L-moment ratios (τ_3 and τ_4) or record length weighted average L-moment ratios (τ^R_3 and τ^R_4) as a scatterplot with theoretical curves of several candidate distributions in a L-skewness-L-kurtosis space. The selected distribution should give the closest approximation to the regional data. Compared to the two-parameter distribution with location and scale parameters, the three-parameter distribution curve is good at providing a more convenient and intuitive expression. The construction of a theoretical L-moments relationship for the common distributions is based on the polynomial approximations proposed by Hosking and Wallis (1991a), and can be summarized as:

$$\tau_4=\sum_{k=0}^{8} A_k\tau_3^k \qquad [3.14]$$

where A_k are the coefficients of polynomial approximations for several assumed distributions. Table 3.2 lists the A_k for key distributions:

Table 3.2 Polynomial approximations of τ_4 as a function of τ_3 (Hosking and Wallis, 1997)

	GPA	GEV	GLO	LN3	PE3	OLB
A_0	0	0.10701	0.16667	0.12282	0.12240	-0.25
A_1	0.20196	0.11090				
A_2	0.95924	0.84838	0.83333	0.77518	0.30115	1.25
A_3	-0.20096	-0.06669				
A_4	0.04061	0.00567		0.12279	0.95812	
A_5		-0.04208				
A_6		0.03763		-0.13638	-0.57488	
A_7						
A_8				0.11368	0.19383	

3.4.3.2 Goodness-of-fit test

After determining the regional frequency distribution based on the L-moment ratio diagram, the next step is to use the goodness-of-fit measure to test whether a selected distribution give the closest fit to the observed data. This measure also can identify the most suitable distribution when the acceptable regional distribution is more than one.

The goodness-of-fit test based on the L-moments (Hosking and Wallis, 1997) is

employed to measure how well the simulated L-moment ratios (τ_3 and τ_4) of a fitted distribution match the samples' average L-moment ratios.

The goodness-of-fit measure is then defined as:

$$Z^{DIST}= (\tau_4^{DIST}-\tau_4^{R} +B_4)/\sigma_4 \qquad [3.15]$$

where τ_4^{DIST} is the average L-kurtosis obtained from simulation of a fitted distribution; τ_4^{R} is the average L-kurtosis calculated from observed data in a given region; B_4 is the bias of τ_4^{R} and σ_4 is the standard deviation of L-kurtosis from simulation, and the B_4 and σ_4 are given by:

$$B_4=N_{sim}^{-1} \sum_{m=1}^{Nsim}(t_4^m-t_4^R) \qquad [3.16]$$

$$\sigma_4= [(Nsim-1)^{-1}\{\sum_{m=1}^{Nsim}(t_4^m-t_4^R)^2-NsimB_4^2\}]^{1/2} \qquad [3.17]$$

As suggested by Hosking and Wallis (1997), the acceptable value of $|Z^{DIST}|$ for any candidate distributions should be less than 1.64 and the selected distribution is required to have the $|Z^{DIST}|$ closer to zero.

To carry out the goodness-of-fit test, Hosking and Wallis (1997) suggested using the L-moments based Monte Carlo simulation. To simulate a large number of regions (Nsim=500) fitted by the kappa distribution. Each simulated region is designed to have the same number of sites. Each simulated site has the same record length as the

sample site in the real world. The simulated regions are required to have the same L-moment ratios as the regional average ones. The simulation process is summarized as:

1) Assume there is a homogeneous region with N sites and each site has a record length of n_i. The sample L-moment ratios are defined as $t^{(i)}$, $t_3^{(i)}$ and $t_4^{(i)}$, and the regional weighted average L-moment ratios are t^R, t_3^R and t_4^R.

2) Simulate a large number of regions with the same number of sites and record length as the sample data. Fit the four parameters kappa distribution to the simulated data in the simulated regions.

3) Calculate the basis B_4 of t_4^R and standard deviation of L-kurtosis from simulation.

The Matlab program code (Appendix A-3) is employed to complete this simulation.

3.4.3.3 Robustness test

Hosking and Wallis (1997) mentioned that even though the best fitted regional frequency distribution is determined, there is no guarantee that the chosen distribution can match to future data. A robust distribution should have the ability to

yield accurate quantile estimates when mis-specification of the distribution or the region is not homogeneous, even when the true at-site frequency distribution deviates from the chosen one.

The robustness test for the candidate distribution is achieved by comparing the bias and root mean square error (RMSE) of the estimated extreme quantiles when 1) the chosen distribution is the same as the underlying distribution, 2) the chosen distribution differs from the underlying distribution.

The bias, B and the RMSE are defined as:

$$B=E(Q_{Test}-Q_T) \qquad [3.18]$$

$$RMSE=[E(Q_{Test}-Q_T)^2]^{1/2} \qquad [3.19]$$

where Q_{Test} is the regional estimated quantiles by fitting with candidate distribution and Q_T is the true at-site quantiles. In practice, Q_T is unknown and it can only be obtained by fitting the underlying distribution to the observed data.

Following Hosking and Wallis (1997), Pokhrel and Lye (2002) summarized the simulation procedures of calculating the B and the RMSE in the following steps:

1) Assume there is a region having the same number of sites and record length as observed sites in a given homogeneous region;

2) Fit the underlying distribution to the observed data and calculate the at-site parameters of the underlying distribution by using the sample L-moments; and

3) Calculate the at-site quantiles based on the at-site frequency distribution.

Generate 1000 simulated regions.

For each region:

1) Generate a series of sample data having the same length of records as the observed data and fit them with underlying distribution;

2) Calculate at-site and regional L-moment ratios for all of the sites in the simulated region;

3) Fit the candidate distribution to the simulated data. Then describe the regional growth curve and calculate the at-site quantiles fitted by candidate distribution;

4) Calculate the average relative bias and RMSE of the estimated at-site quantiles of all of the sites in the simulated region; and

5) Calculate the regional average relative bias (ARB), average absolute relative bias (AARB) and relative root mean square error (RMSE) of the quantiles of all of

the sites in the region.

The Matlab program code (Appendix A-4) is employed to complete this calculation.

3.4.4 Quantile estimation

Hosking and Wallis (1997) provided the definition of the quantile estimates based on the regional L-moment algorithm: Suppose that a homogeneous region has N sites. Each site i has the record length of n_i and observed data Q_{ij}, j=1,…, n_j. All of the sites in the homogeneous region are ideally distributed except for a site-specific scaling factor (index flood). Then, the estimates of quantile with non-exceedance probability F is

$$Q_i(F)=\mu_i\, q(F) \qquad i=1:N \qquad [3.20]$$

where $Q_i(F)$ is the flood quantile with T return years. μ_i is the site-dependent scale factor known as the index flood and q(F) is the regional quantile of non-exceedance probability F which can be obtained from regional growth curve.

The construction of the regional growth curve provides the values of the regional quantile of non-exceedance probability F for different return periods. Hosking and Wallis (1997) summarized the procedures in the following steps:

1) Calculate the sample L-moments and L-moment ratios of each site in a given

homogeneous region;

2) Calculate the regional record length weighted average L-moment ratios;

3) Obtain the parameters of regional frequency distribution using the formula provided by Hosking and Wallis (1997) with the relationship with the weighted average L-moments; and

4) Plot the regional frequency distribution function with the Gumbel variate of non-exceedance probability (-log (-log (F)), where F represents different return periods.

The construction of regional growth curve can easily be carried out with the use of the Matlab program code (Appendix A-5).

3.4.5 Index flood estimation at ungauged sites

To estimate the quantile flow at ungauged sites, the index flood is required. However, in reality, it is impossible to obtain the index flood for ungauged sites, or it is not available to be estimated directly using the index flood at gauged sites. Therefore, under these circumstances the common approach used is to establish a linear or nonlinear regression model to relate the physiographic site characteristics to the index flood at gauged sites within a homogeneous region. In this thesis, the

development of a nonlinear regression model is based on the least-square approach which is defined as:

$$\overline{Q} = \alpha_0 A_1^{\alpha 1} A_2^{\alpha 2} \ldots\ldots A_n^{\alpha n} + \varepsilon_0 \qquad [3.21]$$

where $\overline{Q}$ is the annual peak flow of each site; A_1 is the site characteristics; α_0, α_1 are the model parameters; and ε_0 is the error.

3.4.6 Assessment of estimation accuracy

The bias and error due to the uncertainties or other factors in hydrology studies are impossible to avoid. In this thesis, the assessment of accuracy is carried out by plotting the at-site and estimated regional quantiles to see how well the regional quantiles model match observed data. The procedures are summarized as follows:

1) Select a number of gauged sites which are not used for the RFFA due to the short record length within a given homogeneous region;

2) Calculate the L-moments and L-moment ratios of each site;

3) Calculate the at-site parameters of regional frequency distribution using sample L-moment ratios and obtain the at-site quantiles at each site;

4) Obtain quantile estimates at each site using regional quantile estimates model; and

5) Draw a straight line with y=x and plot the at-site quantiles and regional quantiles respectively to see how well the regional quantile model match the observed data.

In the next chapter, the methodology described in this chapter will be applied to the flood data from Labrador and Newfoundland.

CHAPTER 4

DATA ANALYSIS AND RESULTS FOR LABRADOR

4.1 General

Labrador is on the mainland in the northeastern part of Canada with an area of 294,330 square kilometers. The only regional flood frequency analysis (RFFA) for Labrador was competed by AMEC (AMEC Environment & Infrastructure) in 2014 using the regression on quantile approach authorized by the Government of Newfoundland and Labrador. Although there has been only one documented flood in Labrador, accurate RFFA will become more significant in the face of changing climate in the future (Flood Risk and Vulnerability Analysis Project, 2012).

The RFFA based on the regression on quantile approach conducted by AMEC (2014) can be described as follows:

1) Twelve gauged sites were selected with a record length of at least 10 years. Sample flood data were obtained from the HYDAT database updated in April, 2014 with data up to 2012 from the Water Survey of Canada (WSC).

2) Labrador was treated as a single hydrologically homogeneous region.

3) A three parameter lognormal distribution (LN3) was fitted to each of the 12 sites.

Flood quantile regression equations relating site characteristics to selected flood quantiles for single stations and the entire region were developed respectively.

4) One parameter-Drainage area (DA) and two parameters-Drainage area (DA) and Lake Attenuation Factor (LAF) were used to develop regression models and to estimate selected flood quantiles at ungauged sites respectively.

The regression equations developed by AMEC (2014) are shown in Table 4.1.

Table 4.1 Regression equations and goodness-of-fit developed by AMEC (2014) for Labrador

One Parameter Equations	SMR	SEE	Two Parameters Equations	SMR	SEE
$Q_2=0.495*DA^{0.837}$	0.968	0.120	$Q_2=0.581*DA^{0.845}*LAF^{-0.053}$	0.969	0.125
$Q_5=0.617*DA^{0.838}$	0.965	0.127	$Q_5=0.685*DA^{0.843}*LAF^{-0.034}$	0.965	0.133
$Q_{10}=0.692*DA^{0.839}$	0.962	0.131	$Q_{10}=0.746*DA^{0.842}*LAF^{-0.025}$	0.962	0.138
$Q_{20}=0.761*DA^{0.839}$	0.960	0.135	$Q_{20}=0.800*DA^{0.842}*LAF^{-0.017}$	0.960	0.142
$Q_{50}=0.847*DA^{0.840}$	0.958	0.139	$Q_{50}=0.866*DA^{0.841}*LAF^{-0.008}$	0.958	0.147
$Q_{100}=0.909*DA^{0.840}$	0.956	0.142	$Q_{100}=0.914*DA^{0.840}*LAF^{-0.002}$	0.956	0.150
$Q_{200}=0.970*DA^{0.840}$	0.954	0.145	$Q_{200}=0.959*DA^{0.840}*LAF^{0.04}$	0.954	0.153

SMR: Regression correlation coefficient (R^2)
SEE: Standard error of the estimate
Q_n; Flood quantile with T return years

Hosking and Wallis (1997) indicated that variability will be effectively reduced by using a set of sample or other related samples drawn from similar probability

distribution rather than using a single sample. Quantile estimates become more accurate by using regional data.

In this thesis, the scope of work in Labrador is summarized as the follows:

1) Choose gauged stations with the record length equal to or more than 15 years. Collect the annual peak flow at each site using the HYDAT database with the flood data available until 2013 from the Water Survey of Canada (WSC);

2) Carry out the discordancy measure (D_i) and heterogeneity test (H) to ensure that the region is homogeneous for quantile estimation;

3) Determine the regional frequency distribution and develop the regional quantile function based on the best fitted regional frequency distribution;

4) Estimate index flood at ungauged sites by developing a nonlinear regression relationship between index flood and site characteristics at gauged sites in a given homogeneous region; and

5) Assess the accuracy of the estimations and compare the results to those based on AMEC (2014) equations.

4.2 Screening the data and discordancy measure

Labrador was documented as having 32 gauged stations. Because 15 years of record is the basic requirement in this study, only 10 stations were finally selected. The annual peak flow at each gauged site is collected from the HYDAT CDROM database of the Water Survey of Canada (WSC) with record available from 1954 to 2013. Figure 4.1 illustrates the locations of each station on the map.

The discordancy measure (D_i) proposed by Hosking and Wallis (1993) aims to screen out the unusual sites from other sites in a group by comparing their L-moment ratios. The definition of the discordancy measure is given in Chapter 3.

Table 4.2 summarizes the statistics of each site including the station number, station name, drainage area, length of records, L-moment ratios and results of the discordancy measure.

It is observed that site 03OC003 has a higher D_i value than the critical value (D_i=2.491). Figure 4.2 shows the boxplot of the flood data of site 03OC003. It can be seen that despite the data being positive skewed, neither outliers nor unusual data are detected. Therefore, the higher D_i may be due to its higher L-sk and L-ku and shorter length of record. The positions of sample L-moment ratios in the region are scattered as expected (Figure 4.3), and the results of the discordancy measure for other sites are shown to be satisfactory.

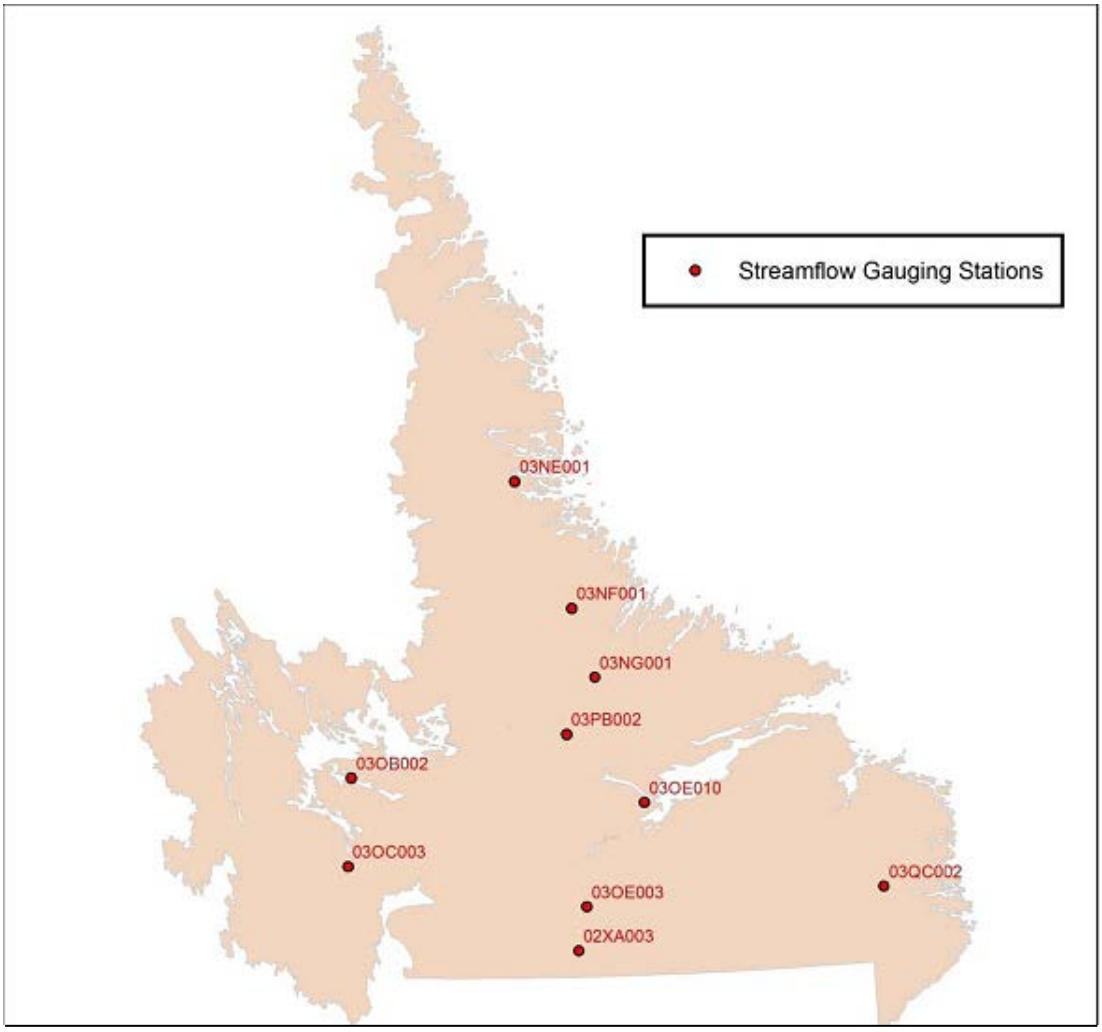


Figure 4.1 Location of studied sites in Labrador (Cited and modified based on the report conducted by AMEC, 2014)

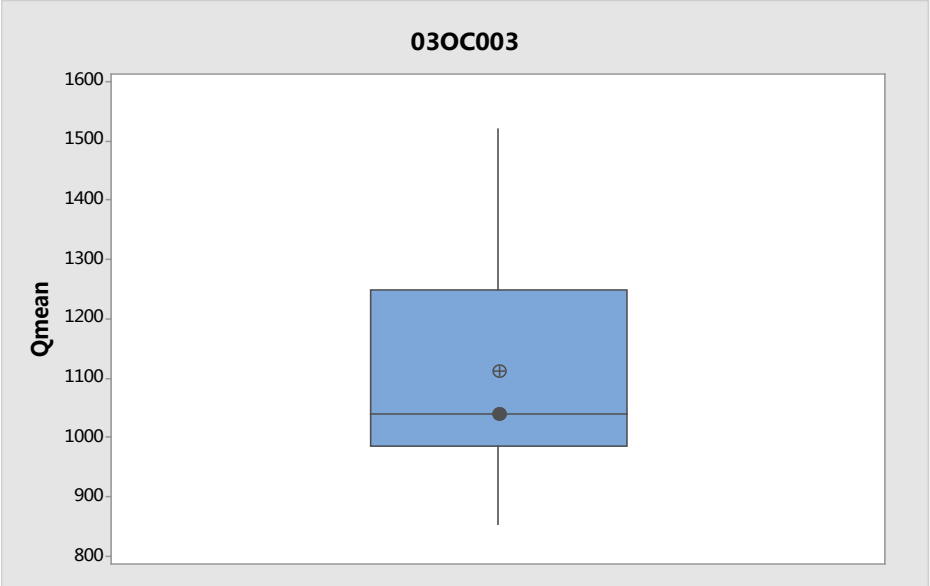


Figure 4.2 Boxplot of site 03OC003

Table 4.2 Summary statistics, L-moment ratios and discordancy measure (D_i) for 10 sites in Labrador

Station Number	Station Name	Years of Record	Mean Max Flow (m^3/s)	Drainage Area (km^3)	Range of Year	L-CV	L-SK	L-Ku	D_i
02XA003	Little Mecatine River above Lac Fourmont	28	653.8	4540	1979-2013	0.148	0.097	0.164	0.148
03NF001	Ugjoktok River below Harp Lake	23	1151.0	7570	1979-2013	0.186	0.165	0.018	1.736
03OC003	Atiktonak River above Panchia Lake	15	1111.1	15100	1999-2013	0.099	0.264	0.157	2.807*
03OE003	Minipi River below Minipi Lake	28	234.0	2330	1979-2013	0.154	0.132	0.135	0.175
03PB002	Naskaupi River below Naskaupi Lake	25	467.6	4480	1978-2011	0.141	-0.057	-0.002	1.898
03QC001	Eagle River above Falls	38	1834.0	10900	1967-2013	0.172	0.112	0.065	0.523
03QC002	Alexis River near Port Hope Simon	32	524.6	2310	1978-2013	0.154	0.024	0.137	0.286
03OB002	Churchill River at Flour Lake	16	2558	33900	1955-1970	0.136	0.064	0.279	1.622
03OE010	Big Pond Brook below Big Pond	20	15.04	71.4	1994-2013	0.150	0.050	0.153	0.194
03OE001	Churchill River above upper Muskrat Falls	54	4560	92500	1954-2012	0.135	0.019	0.068	0.612

* The D_i of this site is higher than the critical value of 2.491

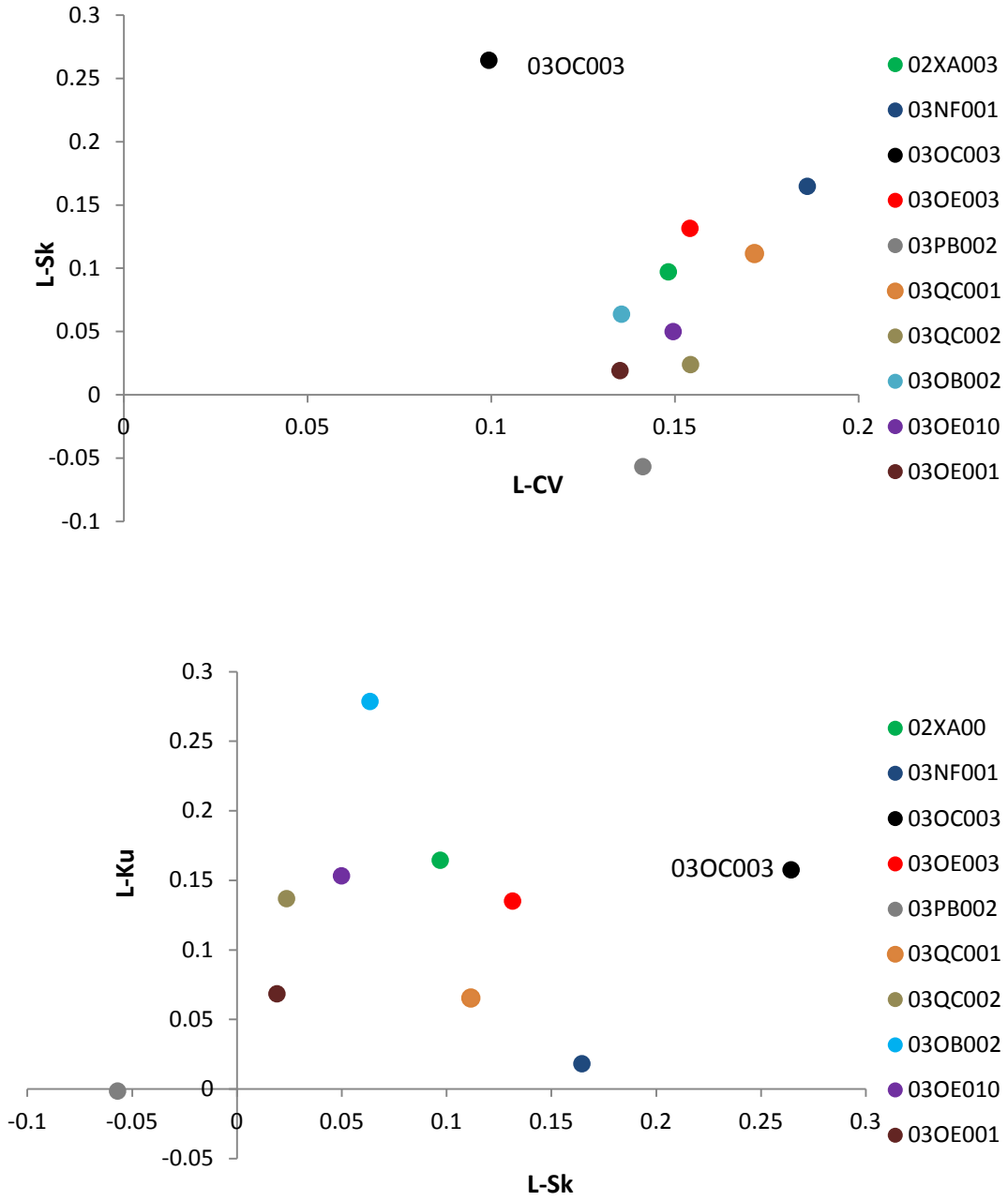


Figure 4.3 L-moment ratios in Labrador

After removing site 03OC003, the rest of sites within the region are shown to have lower D_i value (Table 4.3) than the critical value (2.329). However, whether site

03OC003 should be kept or removed needs a further heterogeneity test.

Table 4.3 Results of discordancy measure after removing site 03OC003

Station Number	Di	Station Number	Di	Station Number	Di
02XA003	0.3526	03QC001	0.5086	03OB002	1.4265
03NF001	1.5355	03QC002	1.0461	03OE010	0.2413
03OE003	0.8360	03PB002	1.7367	03OE001	1.3166

4.3 Delineation of homogeneous regions

As discussed in Chapter 3, the completion of the index flood procedure must be premised on the basis that the supposed region is homogeneous except for the at-site scale factor. Hosking and Wallis (1993) proposed that the degree of heterogeneity can be derived by comparing the between-site variations in sample L-moment ratios for the sites in a group with the expected value that would be in a definitely homogeneous region.

To determine the degree of heterogeneity of a region, Hosking and Wallis (1993) recommended Monte Carlo simulation based on the L-moments described in Chapter 3.

Table 4.4 shows the regional weighted average L-moment ratios, kappa parameters,

mean and standard deviation of simulated V and the result of the heterogeneity measure for 10 gauged sites in Labrador.

Table 4.4 Weighted L-moment ratios, kappa parameters, μv, σv and H value of Labrador

t^R	t_3^R	t_4^R	V	ξ
0.14943	0.07448	0.10457	0.01936	0.86602
α	k	h	μv	σv
0.27823	0.22190	0.16578	0.0141	0.0108
		H=0.48		

Hosking and Wallis (1997) suggested that the region can be regarded as “acceptably homogeneous” if H<1, “possibly heterogeneous” if 1≤H≤2, and “definitely heterogeneous” if H>2. From Table 4.5 it can be observed that the degree of homogeneity has improved by removing site 03OC003, but in order to avoid to fit an outdated frequency distribution, therefore, it is decided to keep it for quantile estimation despite its discordancy measure being unsatisfactory. Thus, it can be concluded that the Labrador is homogenous and the next step of index flood estimation can proceed.

Table 4.5 Weighted L-moment ratios, kappa parameters, μv, σv and H value in the absence of site 03OC003

t^R	t_3^R	t_4^R	V	ξ
0.15227	0.06369	0.10157	0.01567	0.86464
α	k	h	μv	σv
0.28853	0.24371	0.16885	0.0149	0.011
		H=0.07		

4.4 Selection of regional frequency distribution

4.4.1 L-moment ratio diagram

The L-moment ratio diagram is usually used as the first visual inspection tool for selecting a regional frequency distribution from sample data of a region. As discussed in Section 3.4.3.1, sample L-moment ratios, regional average L-moment ratios and theoretical L-moment ratios curves of candidate distributions are plotted in an L-sk and L-ku space. The theoretical plotting positions of the candidate distributions are described based on the polynomial approximations (Hosking and Wallis, 1997). Figure 4.4 shows the L-moment ratio diagram in Labrador.

It can be observed that the regional average L-moment ratios τ_3-τ_4 (the black square) is located on the GEV, LN3 and PE3 distributions. And most of the sample L-moment ratios follow along one of these three distributions. Therefore, it can be initially judged that a regional frequency distribution might be generated from those three distributions.

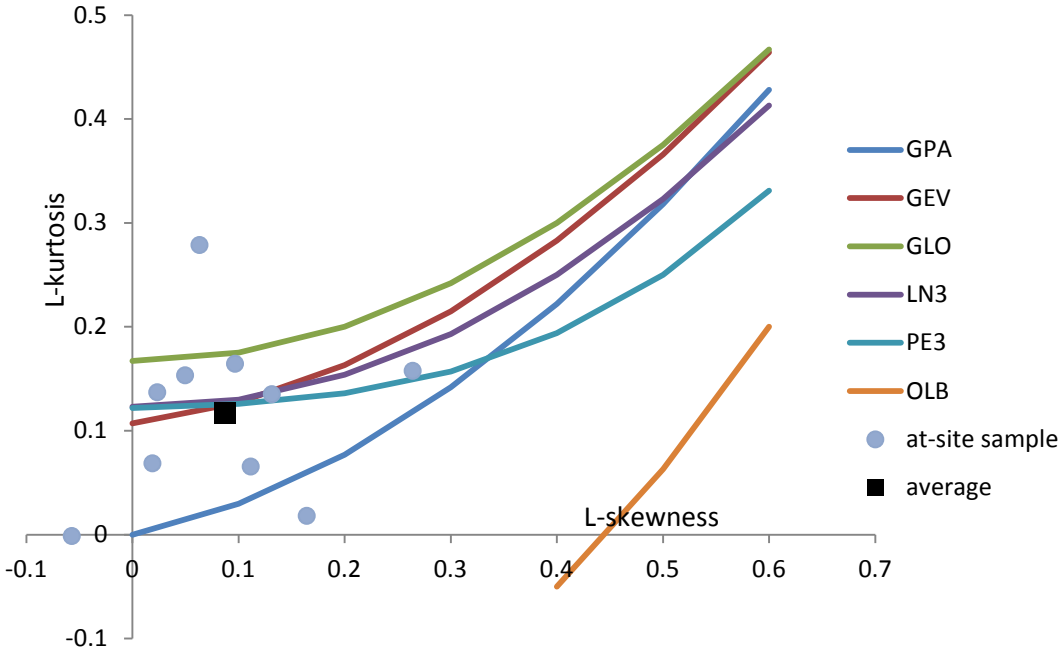


Figure 4.4 L-moment ratio diagram in Labrador

4.4.2 Goodness-of-fit test

The purpose of the goodness-of-fit test is to determine the best fitting frequency distribution by computing the difference of the L-kurtosis between the sample data and the fitted distribution, especially when there is more than one acceptable distribution. Common candidate regional frequency distributions include generalized GLO, GEV, GPA, LN3 and PE3. As discussed in Chapter 3, the goodness-of-fit test can be calculated using the Monte Carlo simulation method. Table 4.6 present the standard deviation of L-kurtosis for each candidate distribution and the results of their goodness-of-fit measure.

Table 4.6 Results of goodness-of-fit measure of five candidate distributions

	GLO	GEV	LN3	PE3	GPA
τ_4^{DIST}	0.0200	0.119949	0.127124	0.12410	0.020281
Z^{DIST}	3.349968	0.787433	1.145547	0.994625	-4.186857

Hosking and Wallis (1997) declared that the reasonably critical value of $| Z^{DIST} |$ should be less than 1.64. An adequate fit can be declared if $| Z^{DIST} |$ is close to zero. Therefore, based on this standard, it can be said that the distributions of GEV, LN3 and PE3 are satisfactory because of their $| Z^{DIST} |$ being less than 1.64.

4.4.3 Robustness test

The purpose of robustness test is to select and verify the most robust regional frequency distribution when there is more than one distribution that provides an adequate fit. From the results of goodness-of-fit measure, the most robust regional frequency distribution will be generated among the GEV, LN3 and PE3 distribution. However, for the PE3 distribution, the distribution function cannot be defined and also it is inconvenient to be used by practitioners, hence only the LN3 distribution and GEV distribution will be used in the robustness test. A robustness test is carried out to compare the regional average relative bias (ARB) and relative root mean square error (RMSE) of the estimated extreme quantiles when 1) the chosen distribution is LN3 but the underlying distribution is GEV; 2) when the chosen

distribution is GEV but the underlying distribution LN3; 3) when the chosen distribution is the same as the underlying distribution. Table 4.7 shows the results of regional average relative bias (ARB), average absolute relative bias (AARB) and relative root mean square error (RMSE) of the estimated extreme quantiles based on the cases mentioned above.

From the results shown in Table 4.7 it can be observed that when the chosen and underlying distributions are GEV, the values of ARB, AARB and RMSE are significantly lower than those when the chosen or underlying distribution is LN3. The differences of ARB, AARB and RMSE of estimated quantiles for 100-year quantiles are lower in LN3-GEV than in the case of GEV-LN3. Therefore, it can be concluded that the GEV distribution is the most robust regional frequency distribution for Labrador.

Table 4.7 Results of robustness test in Labrador

	Quantiles						Difference for 100-year event
	0.9	**0.99**	**0.999**	**0.9**	**0.99**	**0.999**	
	GEV-GEV			**GEV-LN3**			
ARB	-2.86	-5.91	-8.34	16.62	71.27	8.59	77.18
AARB	4.03	8.07	13.04	16.62	71.27	80.37	63.20
RMSE	0.30	0.76	1.72	2.20	109.24	85.63	108.48
	LN3-LN3			**LN3-GEV**			
ARB	17.80	75.73	13.30	2.01	2.08	3.85	-73.65
AARB	17.80	75.73	87.08	3.57	7.61	14.37	-68.12
RMSE	2.41	116.76	93.59	0.28	0.65	1.69	-116.11

ARB: Average relative bias
AARB: Average absolute relative bias
RMSE: Relative root mean square error

4.5 Quantile Estimation

After the regional frequency distribution is determined, the flow quantiles at each site with T return periods within a homogeneous region can be estimated based on the equation [3.20] proposed by Hosking and Wallis (1997). It is repeated here for continuity as Eq. [4.1].

$$Q_i(F)=\mu_i\, q(F) \qquad i=1:N \qquad [4.1]$$

where $Q_i(F)$ is the quantile function of fitted distribution at site i; μ_i is the site-dependent scale factor; and q(F) is the regional quantile of non-exceedance probability F which can be obtained from regional growth curve.

4.5.1 Regional growth curve

The regional growth curve, or the regional frequency distribution function for Labrador is described by the quantile function of the GEV distribution (Eq. [4.2]) based on the L-moments provided by Hosking and Wallis (1997).

$$q(F)=x(F)=Q_T/Q_{mean}=\begin{cases}\xi+\alpha\{1-(-\log F)^k\}/k, & k\neq 0\\ \xi-\alpha\log(-\log F), & k=0\end{cases} \qquad [4.2]$$

where Q_T is the maximum flow quantile and Q_{mean} is the at-site mean peak discharge; here, the F is replaced by 1/T. ξ, α and k are the parameters of GEV distribution.

To describe the regional growth curve and confidence limits, Hosking and Wallis (1997) recommended a Monte Carlo simulation which can be carried out with Matlab program code (Appendix A-5) based on the following steps:

1) Calculate sample L-moment ratios and regional L-moment ratios respectively.

2) Compute the at-site and regional parameters of GEV distribution using the at-site and regional L-moment ratios.

3) Simulate a large number of realizations (Nsim=1000). Each simulated region is required to have the same number of sites. The length of record at each site is required

to be the same as the observed sites. Then plot the regional growth curve with the relationship of q(F) and Gumbel variate of non-exceedance probability (-log (-log (F)), and

4) Obtain the quantile estimation from Eq. [4.1].

The regional parameters of the GEV distribution and regional GEV quantile function are shown in Table 4.8. Figure 4.5 describes the regional growth curve for Labrador with 90% confidence intervals.

Table 4.8 Regional GEV parameters and GEV quantile function in Labrador

ξ (location)	α (scale)	k (shape)
0.8977	0.2229	0.1342
GEV Quantile Function x(F)=0.8977+0.2229{1-(-logF)*0.1342}/0.1342		

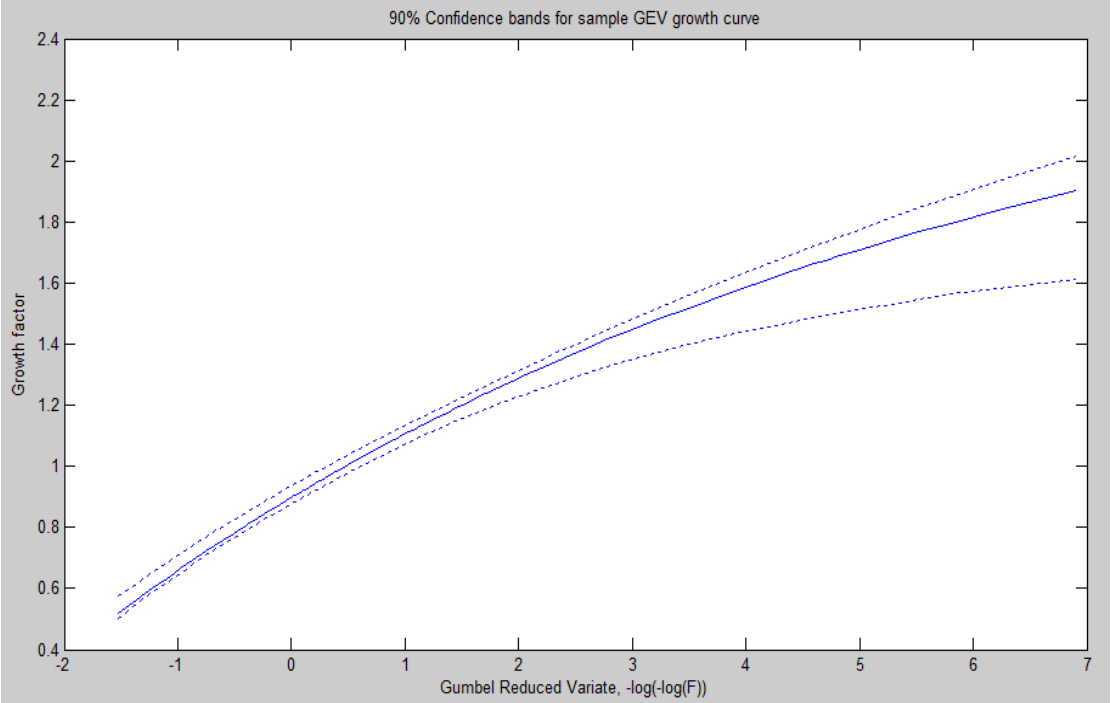


Figure 4.5 Regional GEV growth curve for Labrador with 90% confidence intervals

Figure 4.6 shows the relationship between the at-site Q_T/Q_{mean} and regional growth factor q(F) for T return years. The empirical distribution of the at-site data can be obtained from the Cunnane plotting position formula $p_i(j)=(j-0.4)/(n_i+0.2)$ (Cunnane, 1978). The observed peak flow is arranged in an ascending order starting from 1 to the number of records—n at each site. From Figure 4.6 it can be seen that there is a good agreement between the estimated regional quantile function and empirical at-site value. Table 4.9 list the values of q(F) for different return period (2, 10, 20, 50, 100 and 200) with 90% confidence intervals in Labrador.

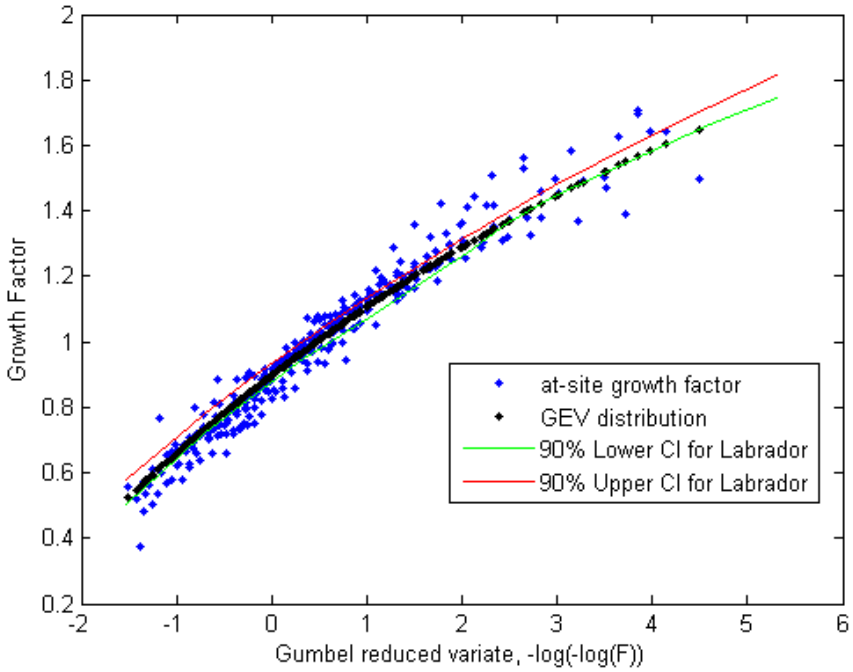


Figure 4.6 Regional quantile function fit observed data with 90% confidence limits in Labrador

Table 4.9 Return period growth factor with 90% confidence intervals in Labrador

Return Period (Years)	Annual Exceedence Probability (AEP)	Reduced Gumbel Variate	Observed growth factor	Lower 90% Confidence Interval	Upper 90% Confidence Interval
2	0.5	0.367	0.997	0.952	1.011
10	0.1	2.260	1.331	1.262	1.356
20	0.05	2.963	1.444	1.346	1.477
50	0.02	3.916	1.575	1.435	1.621
100	0.01	4.6	1.663	1.488	1.720
200	0.005	5.301	1.532	1.532	1.815

4.5.2 Results of quantile estimation

Table 4.10 compares the results of quantile estimates with the return periods of 50 and 100 years between the at-site and regional quantile estimates.

Table 4.10 Results of comparison between at-site and regional quantile estimates in Labrador

Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02XA003	28	1063.59	1128.93	1029.73	1087.27	-3.29	-3.83
03NF001	23	2308.56	2518.39	1812.83	1914.11	-27.35	-31.57
03OC003	15	1725.38	1900.76	1749.98	1847.76	1.41	-2.87
03OE003	28	395.92	425.45	368.55	389.14	-7.43	-9.33
03PB002	25	678.07	694.48	736.47	777.62	7.93	10.69
03QC001	38	3200.15	3431.05	2888.55	3049.94	-10.79	-12.50
03QC002	32	823.88	858.72	826.25	872.41	0.29	1.57
03OB002	16	3937.02	4128.36	4028.85	4253.95	2.28	2.95
03OE010	20	23.78	24.92	23.69	25.01	-0.39	0.37
03OE001	54	6819.02	7076.21	7182.00	7583.28	5.05	6.69
					Absolute average	**6.62**	**8.24**

Site 03NF001 shows the biggest percentage of difference between at-site and regional analysis, these results may be as a results of its highest τ_3 value.

4.5.3 Comparison with the regression on quantile results

As mentioned at the beginning of the thesis, the RFFA (AMEC, 2014) was conducted

based on the regression on quantile approach. Labrador was treated as a single homogeneous region. Each gauged site or the whole region was fitted by the three-parameter lognormal (LN3) distribution. The research (AMEC, 2014) compared results of quantile estimates between single site and regional regression models. Table 4.11 represents the differences of quantile flows between at-site and regional estimation based on the index-flood procedure and regression on quantile approach when the return periods are 50 and 100 years. It can be seen that the percentage of difference between at-site and regional analysis based on the index-flood procedure are significantly less than those obtained from quantile on regression models. The regional quantile function provides a better fit to the observed data.

Some studies (Kjeldsen and Jones, 2007; Robson & Reed, 1999; Institute of Hydrology, 1999 & Lim and Lye, 2003) stated that using the median as the index flood rather than the mean can result in a more accurate estimation; thus, in order to verify this suggestion the quantile flow in Labrador is calculated again using the annual median peak flow. Table 4.12 shows the results of the quantile estimation in Labrador based on index-flood procedure using the median and the mean respectively.

Table 4.11 Results of comparison between at-site and regional quantile estimations in Labrador

Station Number	**L-moments-Current Study % Difference**		**Regression-AMEC (2014) % Difference**	
	Q50	Q100	Q50	Q100
02XA003	-3.29	-3.83	-3.19	-3.32
03NF001	-27.35	-31.57	-34.24	-35.34
03OC003	1.41	-2.87	44.23	45.71
03OE003	-7.43	-9.33	26.41	26.43
03PB002	7.93	10.69	23.54	23.64
03QC001	-10.79	-12.50	NA	NA
03QC002	0.29	1.57	-53.17	-52.62
03OB002	2.28	2.95	NA	NA
03OE010	-0.39	0.37	20.00	24.24
03OE001	5.05	6.69	NA	NA
02XA004	NA	NA	-23.96	-25.23
03NE001	NA	NA	3.13	5.71
03NG001	NA	NA	-18.97	-20.77
03OD007	NA	NA	-38.74	-39.05
03OE011	NA	NA	1.33	1.23
Absolute Average	**6.62**	**8.24**	**24.24**	**25.27**

*NA means the site is not available in the analysis.

From Table 4.12 it can be seen that using the mean as the index flood results in a better performance than the median in terms of quantile estimation, at least in Labrador.

Table 4.12 Results of comparison of quantile estimation based on the median and mean

Station Number	Median (m^3/s)	% difference			
		Median		Mean	
		Q50	Q100	Q50	Q100
02XA003	631.5	-6.94	-7.51	-3.29	-3.83
03NF001	1050.0	-31.17	-36.24	-27.35	-31.57
03OC003	1040.0	-5.34	-9.92	1.41	-2.87
03OE003	226.5	-10.99	-12.97	-7.43	-9.33
03PB002	503.0	14.39	16.96	7.93	10.69
03QC001	1840.0	-10.44	-12.14	-10.79	-12.50
03QC002	549.0	1.42	2.69	0.29	1.57
03OB002	2555.0	2.16	2.83	2.28	2.95
03OE010	14.6	-22.24	-21.34	-0.39	0.37
03OE001	4550	4.83	6.46	5.05	6.69
Absolute average		**10.99**	**12.91**	**6.62**	**8.24**

4.5.4 Index flood estimation at ungauged sites

The index flood is required for estimating the quantile flows at ungauged sites. However, the observed data at ungauged sites cannot be obtained in a direct way; therefore, an indirect approach is used by developing a linear on nonlinear regression model between site characteristics and index flood at gauged sites, which worked successfully in the previous studies. The AMEC (2014) determined the drainage area (DA) to be the only significant parameter for the regression model. Zadeh (2012) successfully developed the regression equations using the drainage area (DA) in terms of the low flow quantiles estimation in Labrador. Therefore, in this study a nonlinear

least square regression equation relating the drainage area (DA) to the annual mean peak flow (Q) was used. Figure 4.7 illustrates the regression relationship between drainage area (DA) and annual peak flow (Q) in Labrador.

The data is log-transformed, and the coefficient of regression equation R^2=96.6%. The regression relationship between Q and DA is given by

$$Q_{mean}=0.6470*DA^{0.8081} \qquad [4.3]$$

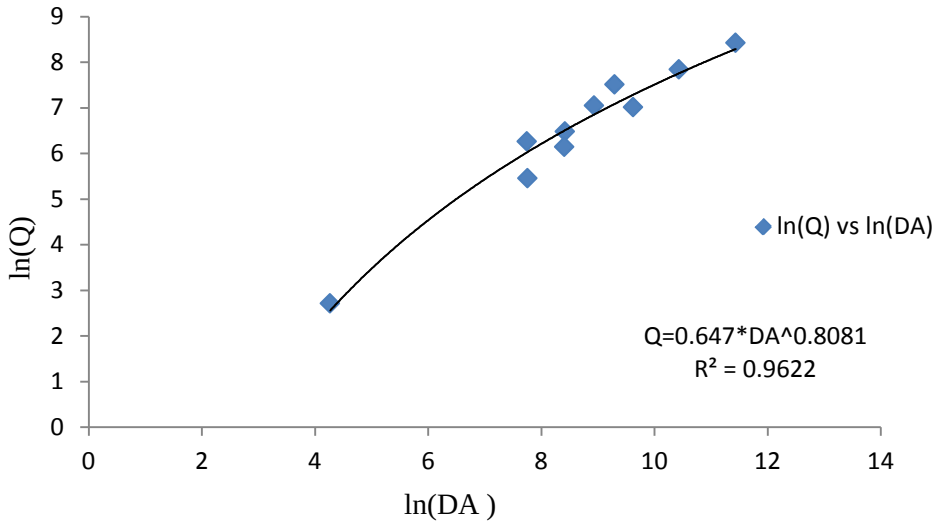


Figure 4.7 Regression relationship of Drainage Area (DA) vs. Peak Flow (Q)

4.6 Assessment of estimation accuracy

As discussed in Section 3.4.5, to assess the estimation accuracy, the method used in this thesis is to plot the at-site and estimated regional quantiles to see how well the

regional quantiles model match sample data. Seven sites in Labrador are selected for testing whether the regional quantile model can provide a good agreement with observed data. The basic information from tested stations is listed in Table 4.13. The at-site and regional growth factors of each tested site are plotted in Figure 4.8.

Table 4.13 Basic information of the stations for verification in Labrador

Station Number	**Station Name**	**Years of Records**	**Drainage Area (km^3)**
02XA004	Riviere Joir Near Provincial Boundary	12	2060
03NG001	Kanairiktok River Below Snegamook Lake	13	8930
03OD007	East Metchin River	13	1750
03OE011	Pinus River	14	779
03NE001	Reid Brook at Outlet of Reid Pond	14	75.7
03PB001	Nashaupi River at Fremont Lake	14	8990
03NE002	Camp Pond Brook below Camp Pond	14	24.3

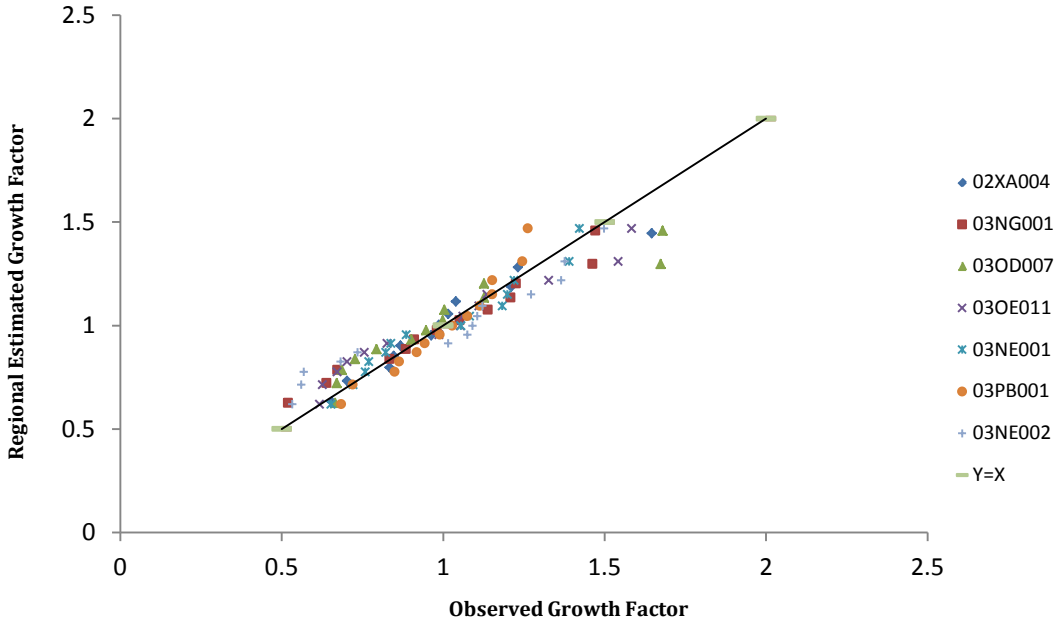


Figure 4.8 Regional growth factor has a good agreement with the observed data

It can be seen that the regional quantile model agrees well with the observed data.

Figures 4.9-4.10 show the comparison of quantile estimates between at-site and regional quantile estimates for Q50 and Q100 at tested sites based on the index-flood procedure and regression models (AMEC, 2014) in Labrador, respectively.

For the approach of L-moments based index-flood procedure, the plot can be completed as the follows:

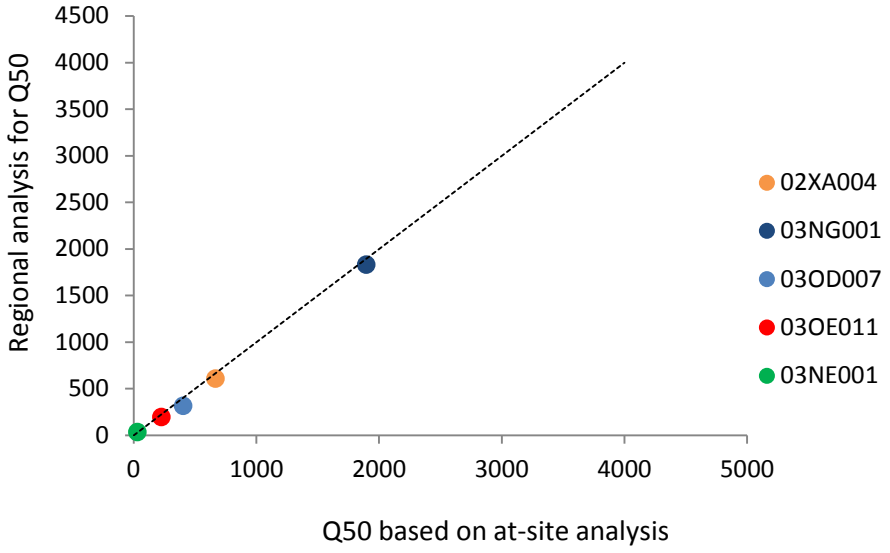
1) Fit the LN3 distribution to at-site and regional data of studied sites.

2) Obtain regional and at-site parameters of the LN3 distribution, respectively.

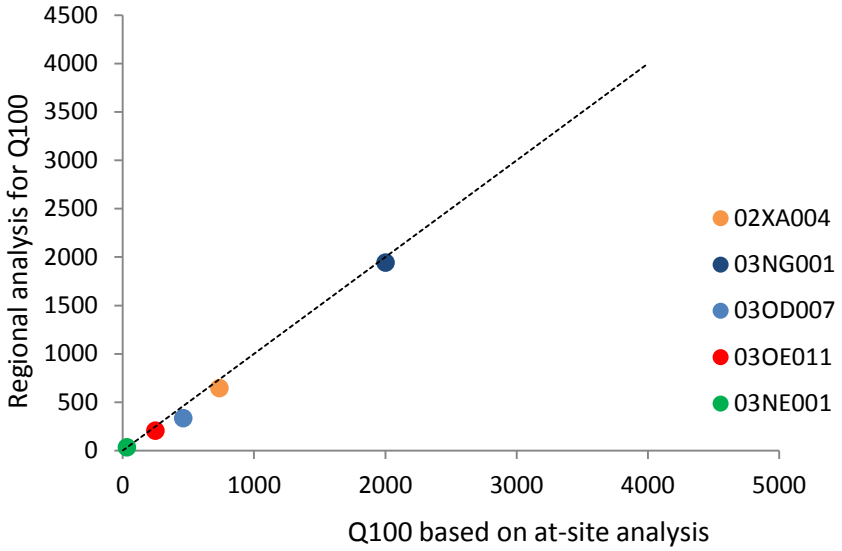
3) Develop regional and at-site quantile functions based on the LN3 distribution.

4) Plot at-site and regional quantile flows of Q50 and Q100 respectively.

Due to the regional quantile estimates at site 03PB001 and site 03NE002 are not available in the study of AMEC (2014), thus five sites are finally used to assess the accuracy of results.

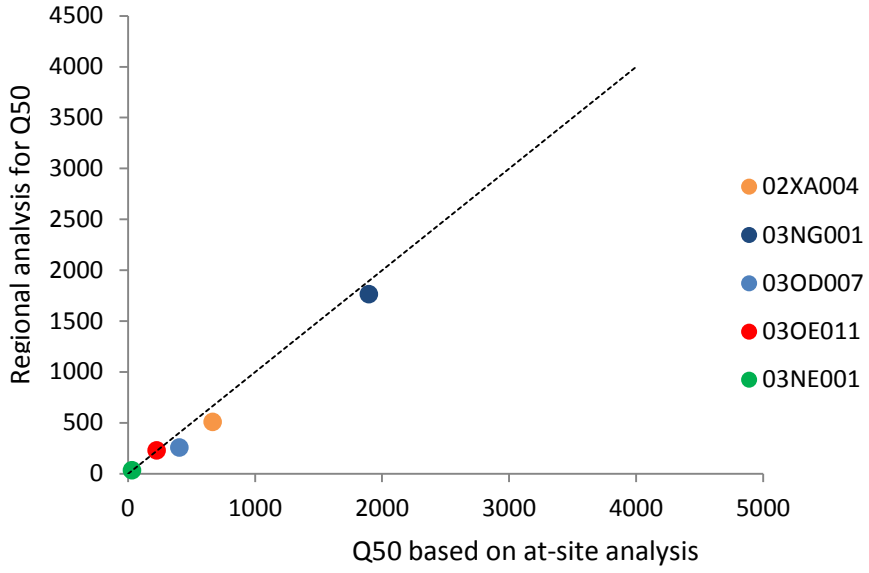


(a)

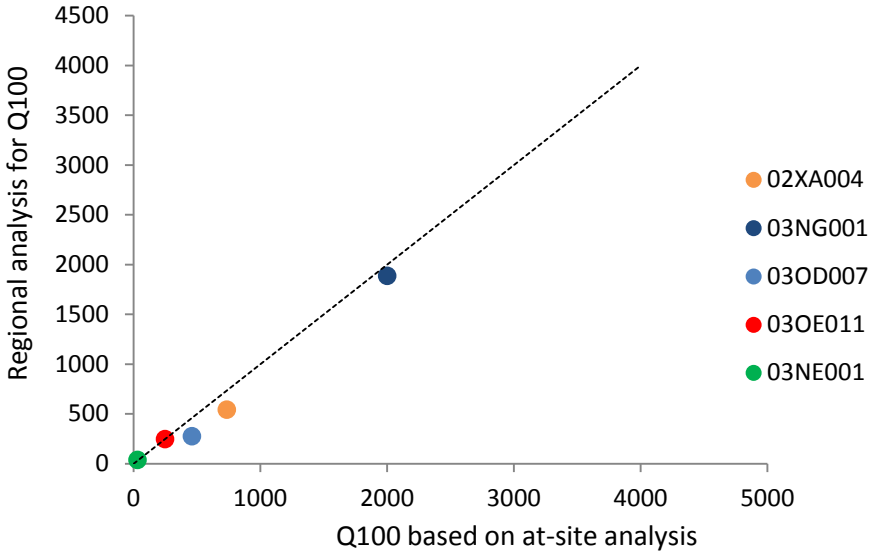


(b)

Figure 4.9 Comparison of quantile estimates for Q50 and Q100 between at-site and regional analysis based on the L-moments based index-flood procedure



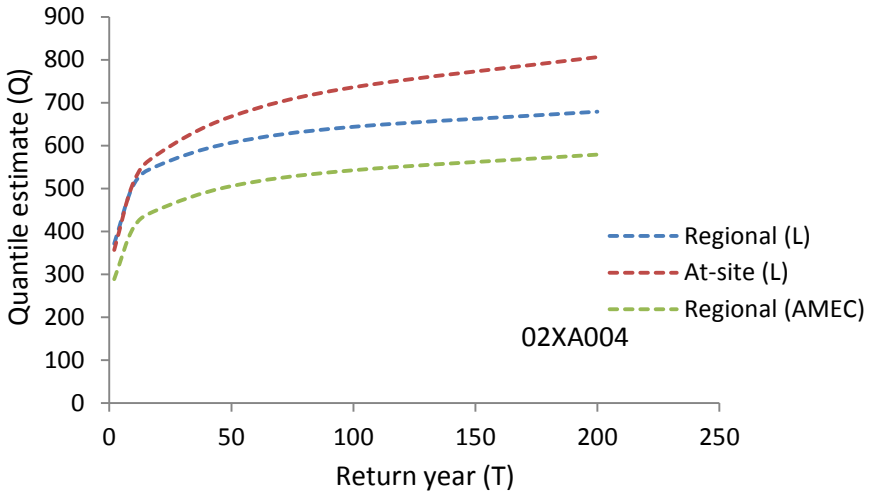
(c)

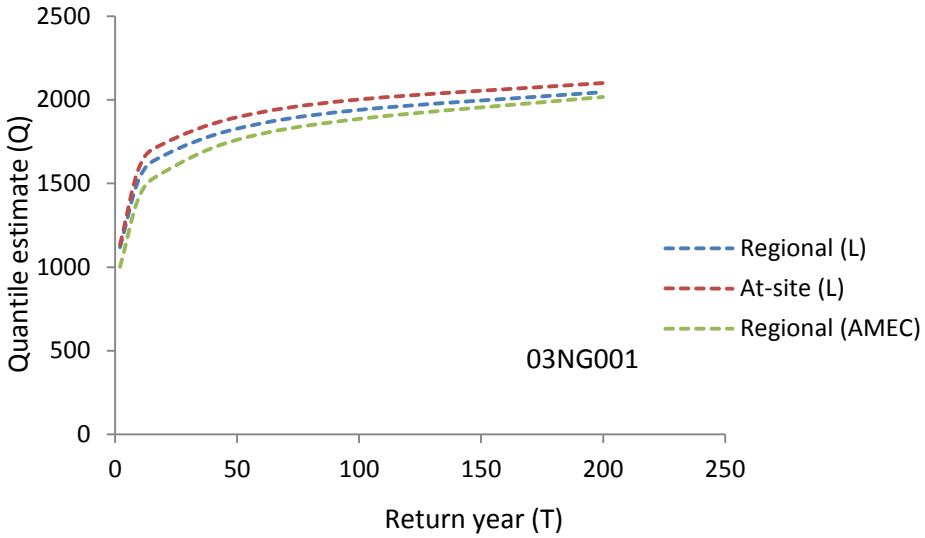


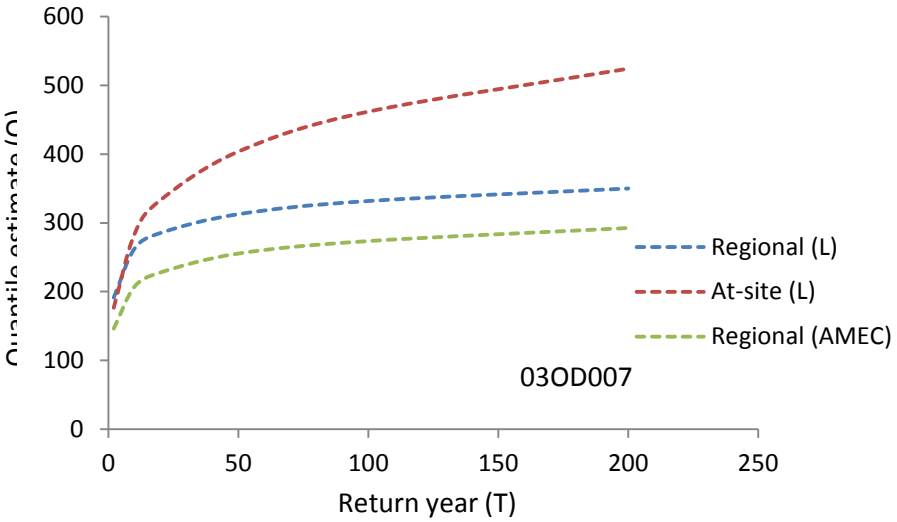
(d)

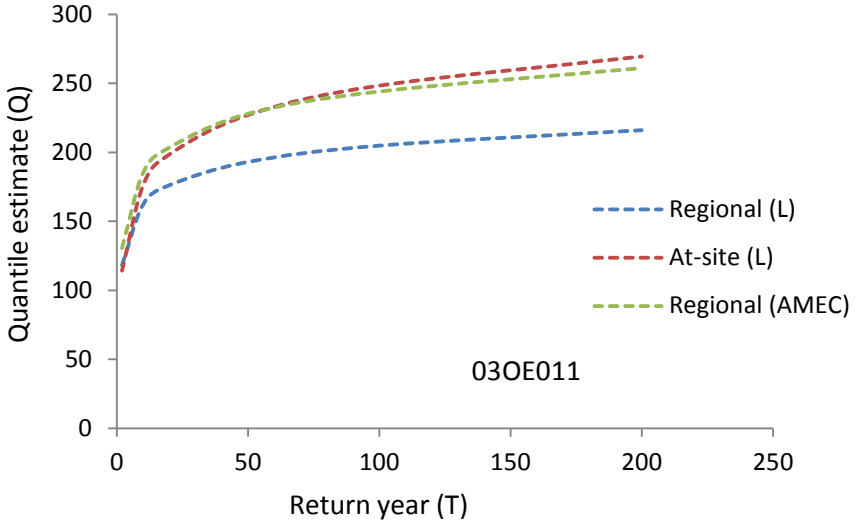
Figure 4.10 Comparison of quantile estimates for Q50 and Q100 between at-site and regional analysis based on the regression on quantiles method obtained from AMEC (2014)

In Figures 4.11, the regional and at-site quantile estimates obtained from the index-flood procedure and the regional quantile estimates based on the regression models from AMEC (2014) are plotted for each single tested site with different return years (T).


900
800
700
600
500
400
300
200
100
0
Quantile estimate (Q)
0
50
100
150
200
250
Return year (T)
Regional (L)
At-site (L)
Regional (AMEC)
02XA004


2500
2000
1500
1000
500
0
Quantile estimate (Q)
0
50
100
150
200
250
Return year (T)
Regional (L)
At-site (L)
Regional (AMEC)
03NG001


600
500
400
300
200
100
0
0
50
100
150
200
250
Return year (T)
Regional (L)
At-site (L)
Regional (AMEC)
03OD007


Quantile estimate (Q)
300
250
200
150
100
50
0
0
50
100
150
200
250
Return year (T)
Regional (L)
At-site (L)
Regional (AMEC)
03OE011

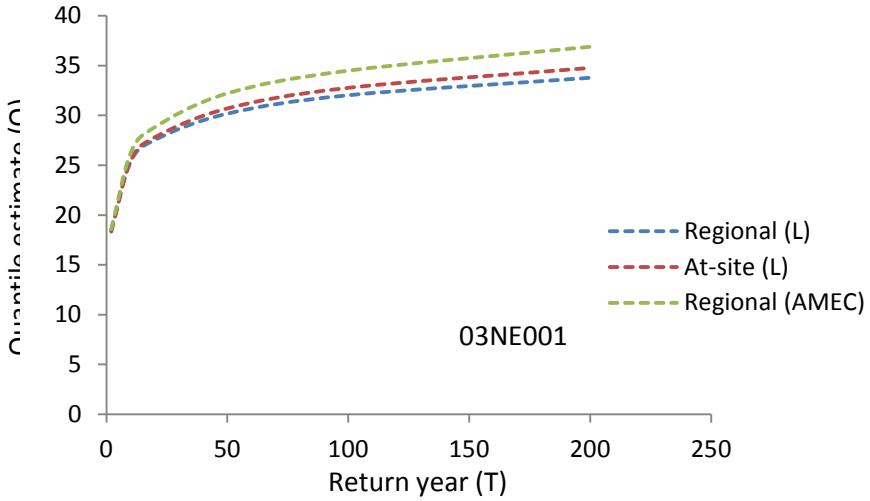


Figure 4.11 Comparison of quantile estimates between L-moments based index-flood procedure and quantile regression models (AMEC, 2014) in Labrador

From the figures shown above, it can be seen that for most of the tested sites the regional quantile estimates obtained from index-flood procedure show better agreement with the at-site flood quantiles. The estimation of quantile flows are based on the available gauged sites in a homogenous region, and the quantile models are limited due to the sample size, available observed flow and other factors, therefore, it is true that the quantile models cannot work perfectly on each sites such as the site 03OE011. Although the LN3 distribution is not the best fitted regional frequency distribution to regional data, the comparisons of quantile estimates show that the method of index flood provide more accurate results than the approach of regression on quantile in Labrador. The quantile models will be improved in the future studies because of more available sites information.

CHAPTER 5

DATA ANALYSIS AND RESULTS FOR THE ISLAND OF NEWFOUNDLAND

5.1 General

In the past 44 years, there were five editions of regression on quantile based RFFA for the Island of Newfoundland. The latest being by AMEC (2014). The first RFFA in Newfoundland using the L-moments based index-flood procedure was conducted by Pokhrel (2002) with the observed data until 1998. Pokhrel (2002) showed that the index-flood procedure provided more accurate quantile results than regression approach. Therefore, the objective of this chapter is to conduct the quantile estimates for rivers on the Island of Newfoundland with the latest data up to 2013 and to compare the results to those obtained from Pokhrel and the latest regression on quantile approach by AMEC (2014). Data used for the AMEC (2014) study were up to 2012 only.

Since the RFFA in 1989, the Island of Newfoundland was treated as four sub homogeneous regions (A, B, C and D) taking into account data availability and regional sites characteristics. See Figure 5.1. Later, two sub homogeneous regions (Y and Z) that were proposed by the Water Survey of Canada (WSC) were applied successfully in Pokhrel's research. See Figure 5.2. Therefore, in this chapter, the RFFA

for the Island of Newfoundland will be conducted based on both regionalization schemes mentioned above, but using the latest data from HYDAT which is the Archived Hydrometric Database of the Water Survey of Canada (WSC). The quantile estimates will then be compared to those from Pokhrel's research and AMEC (2014). This chapter is organized as follows:

1) Conduct the RFFA for Newfoundland based on the index-flood procedure using L-moments.

2) Obtain the regional quantile functions based on the best fitted regional frequency distribution.

3) Determine the regionalization schemes by testing each scheme for homogeneity.

4) Compare the results of the quantile estimates to those obtained from Pokhrel (2002) and AMEC (2014) respectively.

5) Assess the accuracy of the estimations for each approach and regionalization scheme.

5.2 RFFA for four sub regions

5.2.1 Data screening and discordancy measure

The annual peak flow data are collected from HYDAT which is the Archived Hydrometric Database of Water Survey of Canada (WSC). In the latest update in 2014, the Water Survey of Canada not only updated the historical data up to 2013, but also modified many of the previous records that were used in AMEC (2014) and Pokhrel (2002).

To consider the accuracy of the analysis, the sites studied are required to have at least 15 years of records. In this work, 53 gauged sites in Newfoundland are selected and they are divided into the four sub regions (A, B, C and D). Tables 5.1-5.4 list their basic information including station number, station name, length of record, drainage area (DA) and L-moment ratios. Figure 5.1 illustrates the locations of selected sites on the map. The L-moment ratios L-CV vs. L-skewness and L-skewness vs. L-kurtosis for each sub region are plotted in Figure 5.3.

Figure 5.1 Locations of studied sites in Newfoundland (Cited and modified based on the research conducted by Zadeh , 2012)

Figure 5.2 Boundary of sub regions Y and Z in Newfoundland (Cited and modified based on the research conducted by Zadeh , 2012)

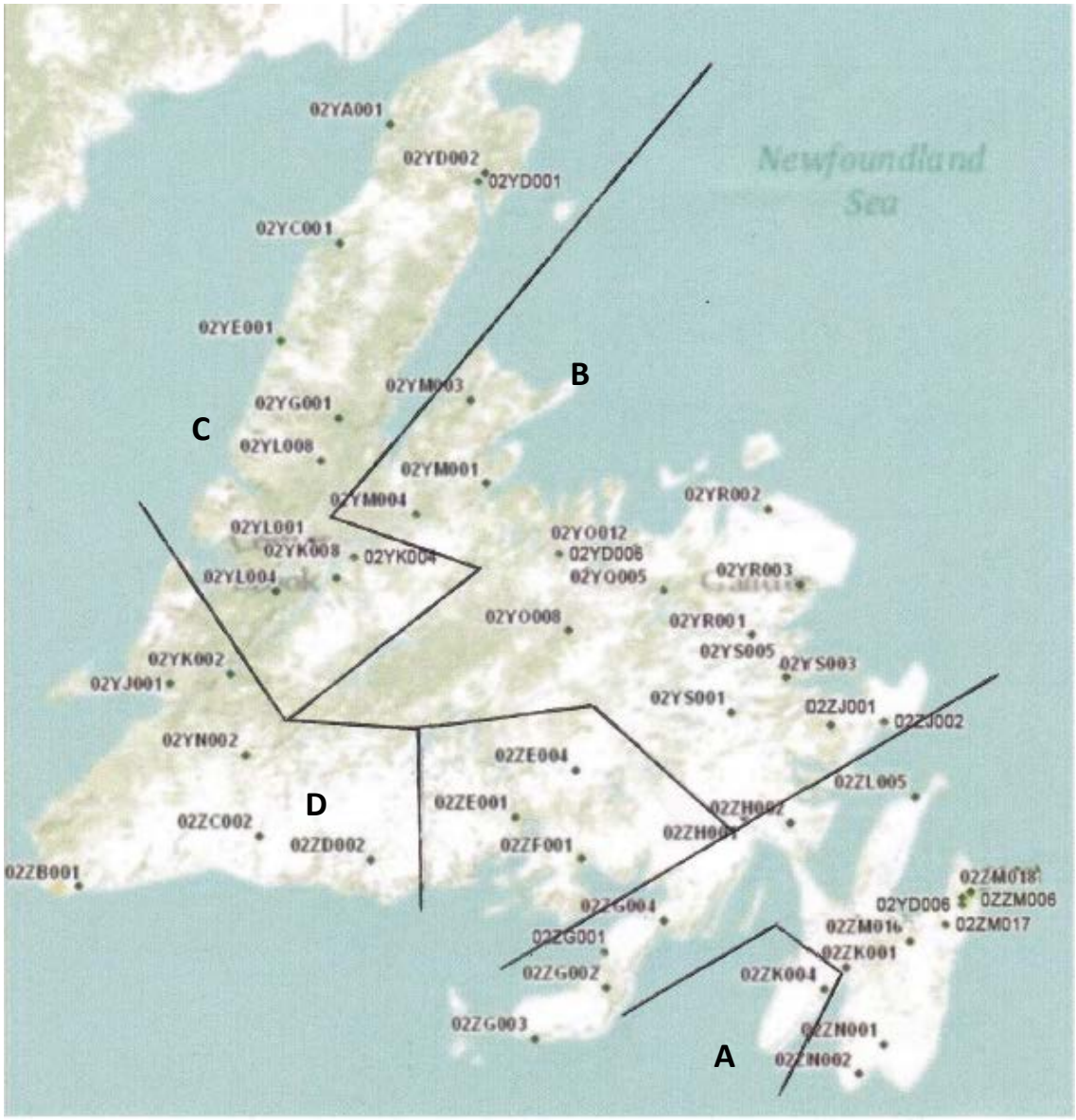


Figure 5.3 Boundaries of sub regions A, B, C and D in Newfoundland (Cited and modified based on the research conducted by Zadeh , 2012)

Table 5.1 Basic information and L-moment ratios for sub region A in Newfoundland

Region A								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**Range of Year**	**L-CV**	**L-SK**	**L-Ku**
02ZG001	Garnish River Near Garnish	49	66.64	205	1959-2013	0.275	0.403	0.265
02ZG002	Tides Brook Below Freshwater Pond	19	50.01	166	1977-1996	0.233	0.259	0.289
02ZG003	Salmonier River Near Lamaline	32	72.45	115	1980-2013	0.226	0.144	0.156
02ZG004	Rattle Brook Near Boat Harbour	31	40.80	42.7	1981-2013	0.262	0.366	0.317
02ZH002	Come By Chance River Near Goobies	41	32.76	43.3	1971-2013	0.214	0.041	0.090
02ZK001	Rocky River Near Colinet	61	156.73	301	1948-2013	0.221	0.202	0.135
02ZL004	Shearstown Brook At Shearstown	29	16.97	28.9	1983-2013	0.276	0.275	0.169
02ZL005	Big Brook At Lead Cove	29	5.95	11.2	1985-2013	0.256	0.271	0.239
02ZM006	Northeast Pond River At Northeast Pond	44	3.74	3.63	1970-2013	0.211	0.151	0.077
02ZM009	Seal Cove Brook Near Cappahayden	35	29.52	53.6	1979-2013	0.128	0.243	0.243
02ZM010	Waterford River At Mount Pearl	15	18.47	16.6	1981-1995	0.223	0.183	-0.071
02ZM016	South River Near Holyrood	31	11.59	17.3	1983-2013	0.200	0.153	0.153
02ZM017	Leary Brook At St. John’s	15	13.24	15.3	1983-1997	0.178	0.143	0.219
02ZM018	Virginia River At Pleasantville	28	9.69	10.7	1984-2013	0.179	0.098	0.069
02ZN001	Northwest Brook At Northwest Pond	28	38.85	53.3	1966-1996	0.167	0.024	0.022
02ZN002	St. Shotts River Near Trepassey	18	10.79	15.5	1985-2013	0.223	0.130	0.079

Table 5.2 Basic information and L-moment ratios for sub region B in Newfoundland

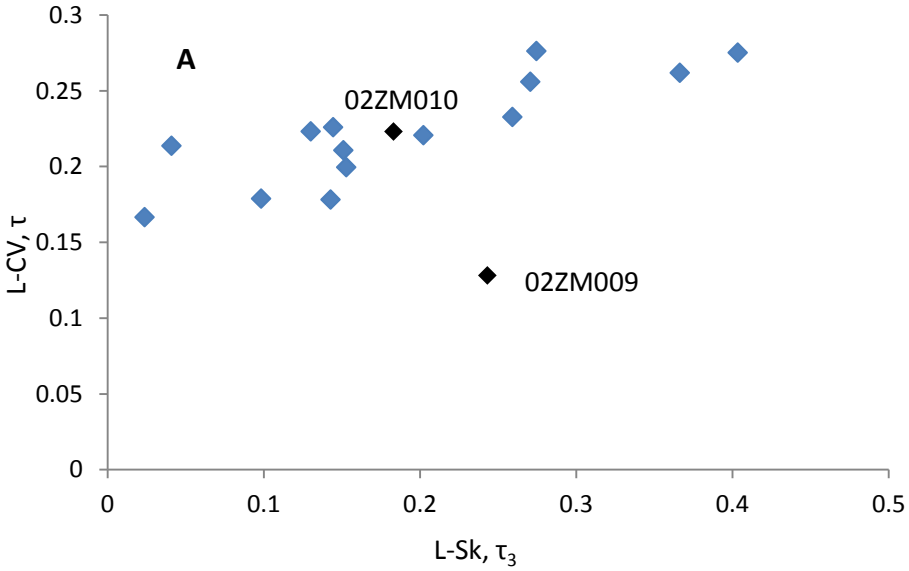
Station Number	Station Name	Years of Record	Mean Max Flow (m^3/s)	Drainage Area (km^3)	Range of Year	L-CV	L-SK	L-Ku
02YM001	India Brook At Indian Falls	40	147.62	974	1955-1995	0.150	0.050	0.200
02YM003	South West Brook Near Baie Verte	31	41.04	93.2	1980-2013	0.247	0.213	0.133
02YO006	Peters River Near Botwood	32	47.83	177	1981-2013	0.224	0.342	0.310
02YO008	Great Rattling Brook Above Tote River Cnfluence	22	214.4	773	1985-2013	0.178	0.143	0.073
02YO012	Southwest Brook At Lewisporte	24	16.25	58.7	1989-2013	0.215	0.188	0.197
02YQ005	Salmon River Near Glenwood	21	40.06	80.8	1991-2013	0.280	0.242	0.116
02YR001	Middle Brook Near Gambo	50	30.11	275	1961-2013	0.179	0.134	0.159
02YR002	Ragged Harbour River Near Musgrave Harbour	17	67.38	399	1978-1997	0.160	0.310	0.237
02YR003	Indian Bay Brook Near Northwest Arm	31	61.98	554	1981-2013	0.185	0.122	0.074
02YS001	Terra Nova River At Eight Mile Bridges	30	182.7	1290	1953-1983	0.167	0.181	0.127
02YS003	Southwest Brook At Terra Nova National Park	43	14.62	36.7	1968-2013	0.205	0.229	0.212
02YS005	Terra Nova River At Glovertown	29	223.1	2000	1985-2013	0.183	0.023	0.076
02ZH001	Pipers Hole River At Mothers Brook	57	241.6	764	1952-2013	0.245	0.167	0.151
02ZJ001	Southern Bay River Near Southern Bay	32	26.63	67.4	1977-2011	0.306	0.444	0.382
02ZJ002	Salmon Cove River Near Champneys	22	13.91	73.6	1983-2013	0.175	0.151	0.201

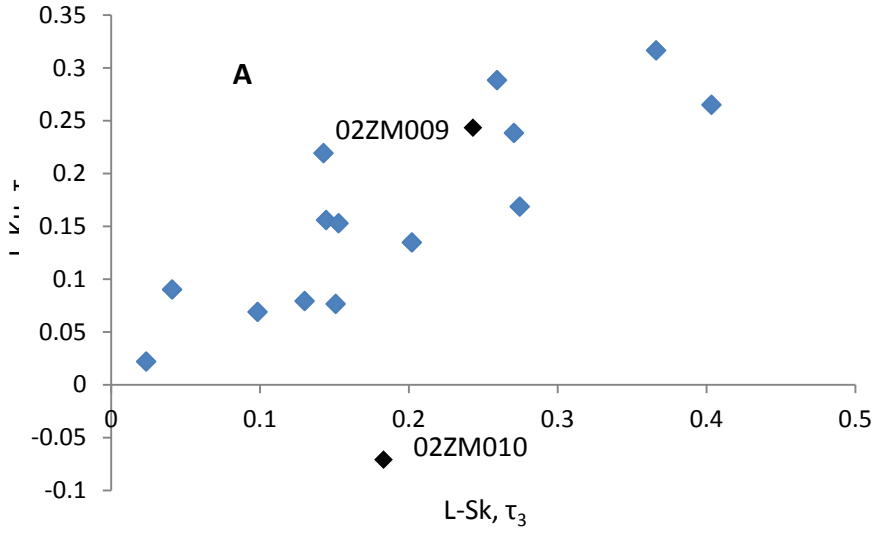
Table 5.3 Basic information and L-moment ratios for sub region C in Newfoundland

Region C								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**Range of Year**	**L-CV**	**L-SK**	**L-Ku**
02YA001	Ste. Genevieve River Near Forresters Point	25	33.15	306	1970-1996	0.186	0.251	0.080
02YC001	Torrent River At Bristol's Pool	53	184.77	624	1959-2013	0.189	0.195	0.171
02YD001	Beaver Brook Near Roddickton	19	103.95	237	1960-1978	0.185	0.224	0.199
02YD002	Northeast Brook Near Roddickton	33	40.32	200	1980-2013	0.128	0.120	0.175
02YE001	Greavett Brook Above Portland Creek Pond	26	46.13	95.7	1985-2013	0.180	0.244	0.141
02YG001	Main River At Paradise Pool	26	314.9	627	1986-2013	0.140	-0.023	-0.024
02YK004	Hinds Brook Near Grand Lake	22	94.13	529	1957-1978	0.141	0.060	-0.006
02YK008	Boot Brook At Trans-Canada Highway	28	10.35	20.4	1985-2013	0.266	0.273	0.179
02YL001	Upper Humber River Near Reidville	85	595.1	2110	1929-2013	0.130	0.127	0.146
02YL004	South Brook At Pasadena	29	44.89	58.5	1983-2013	0.264	0.500	0.364
02YL008	Upper Humber River Above Black Brook	25	251.6	471	1988-2013	0.141	0.087	0.150
02YM004	Indian Brook Diversion Above Birchy Lake	24	38.94	238	1990-2013	0.095	-0.080	0.104

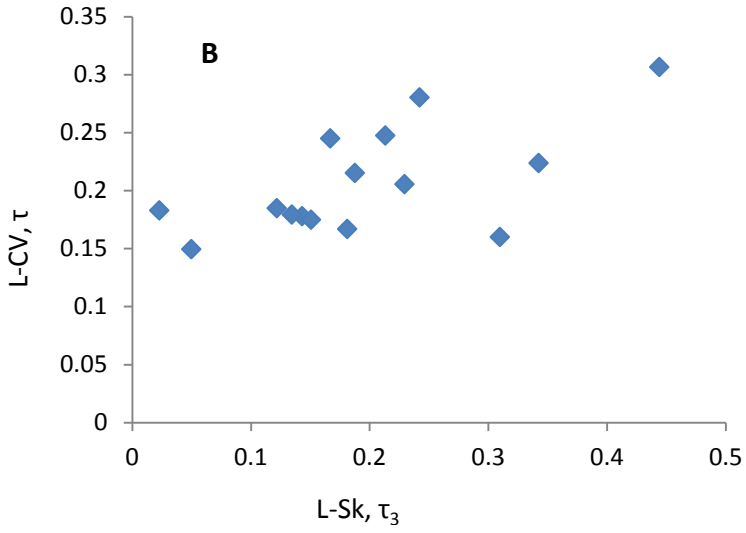
Table 5.4 Basic information and L-moment ratios for sub region D in Newfoundland

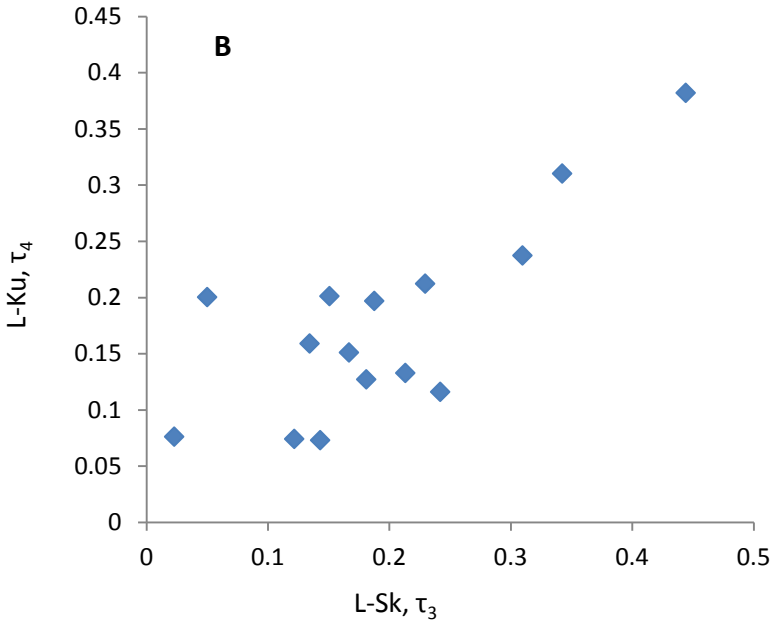
Region D								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**Range of Year**	**L-CV**	**L-SK**	**L-Ku**
02YJ001	Harrys River Below Highway Bridge	42	311.9	640	1969-2013	0.206	0.187	0.153
02YK002	Lewaseechjeech Brook At Little Grand Lake	50	111.15	470	1954-2013	0.176	0.143	0.150
02YN002	Lloyds River Below King George Iv Lake	33	182	469	1981-2013	0.220	0.226	0.169
02ZB001	Isle Aux Morts River Below Highway Bridge	51	373.7	205	1962-2013	0.261	0.224	0.102
02ZC002	Grandy Brook Below Top Pond Brook	28	365.7	230	1984-2013	0.176	0.192	0.144
02ZD002	Grey River Near Grey River	32	872.3	1340	1970-2013	0.242	0.153	0.073
02ZE001	Salmon River At Long Pond	16	292.2	2640	1950-1965	0.157	-0.014	-0.112
02ZE004	Conne River At Outlet of Conne River Pond	25	42.56	99.5	1989-2013	0.177	0.134	0.057
02ZF001	Bay Du Nord River At Big Falls	61	210.2	1170	1951-2013	0.208	0.255	0.254
02ZK004	Little Salmonier River Near North Harbour	31	89.06	104	1983-2013	0.237	0.241	0.129


A
02ZM010
02ZM009
L-CV, τ
L-Sk, τ3
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5

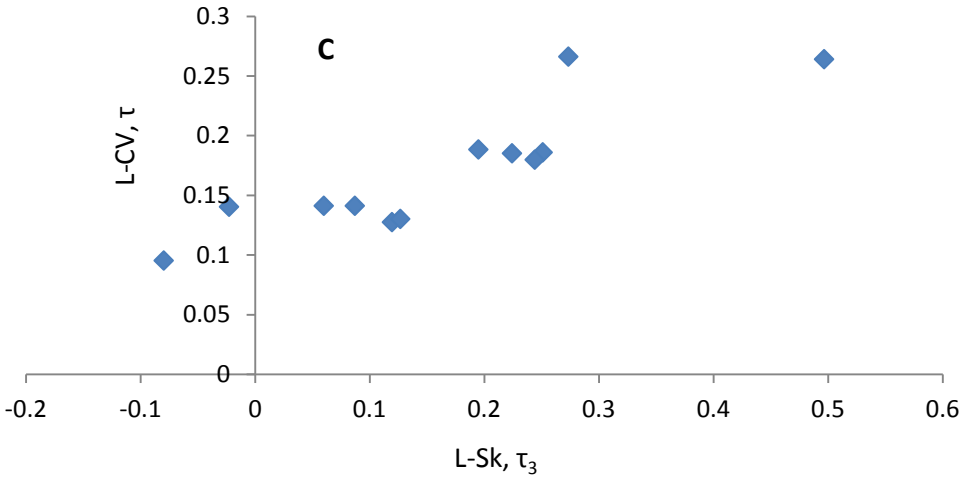

A
02ZM009
02ZM010
L-Sk, τ3
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
0
0.1
0.2
0.3
0.4
0.5

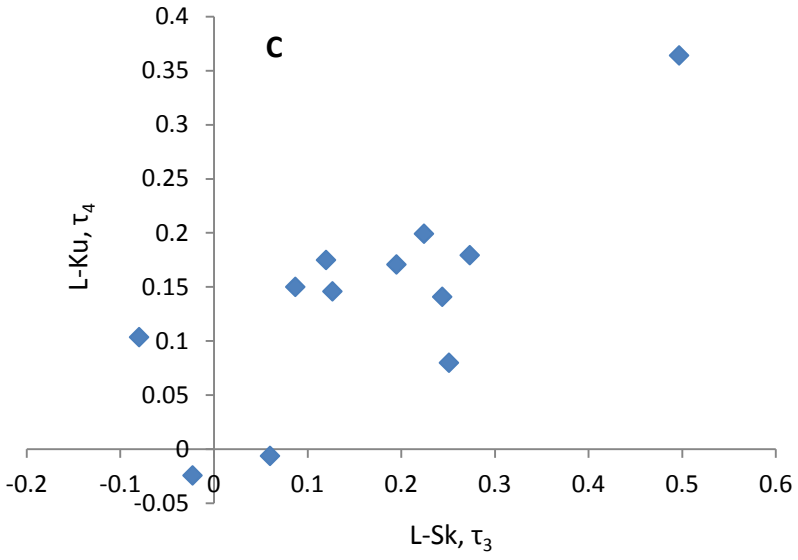
(a)


B
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
L-CV, τ
L-Sk, τ3

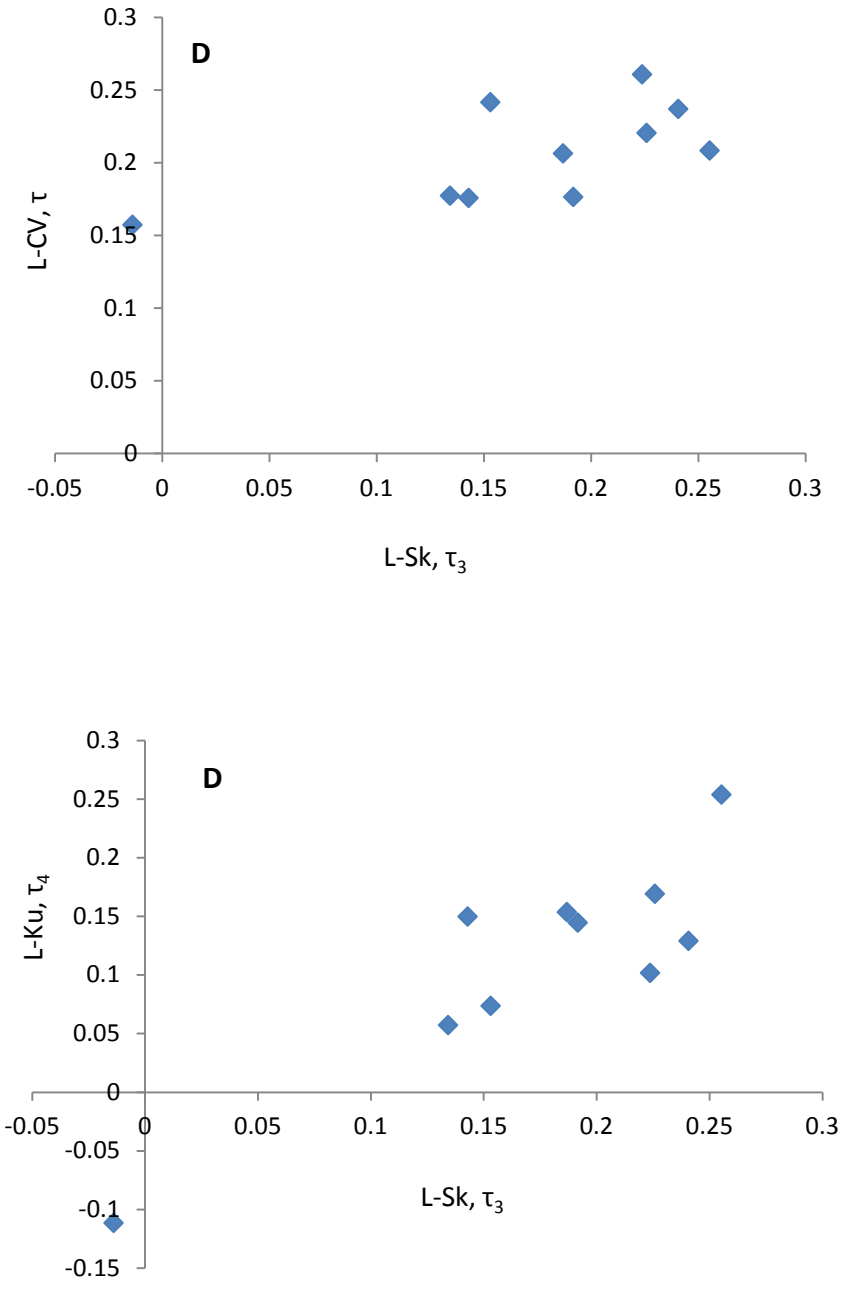

B
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
L-Ku, τ4
L-Sk, τ3

(b)


C
L-CV, τ
L-Sk, τ3
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6


C
L-Ku, τ4
L-Sk, τ3
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6

(c)



(d)

Figure 5.4 L-moment ratios plots in Newfoundland

Using Eq. [3.7] and Matlab program code (Appendix A-1), the results of the discordancy measure (D_i) for four sub regions are shown in Tables 5.5-5.8. It can be

concluded that all of the sites in sub region B, C and D are not discordant from the other sites. But in sub region A, site 02ZM009 has a slightly higher D_i value than the critical value of 3. The positions of L-moment ratios of site 02ZM009 are well scattered, but the boxplot (Figure 5.4) shows that it still has one outlier after taking logarithms. For site 02ZM010, the flood data are well distributed (Figure 5.5), thus its high D_i value may be as a result of low value of L-ku. Removing site 02ZM009, site 02ZM010 still has a high D_i value (2.9609) (Table 5.9). Removing 02ZM010, site 02ZM009 also shows a high D_i value (3.3551) (Table 5.10). Although all of the other sites in region A are shown to be not discordant in the absence of site 02ZM009 and site 02ZM010, whether sites 02ZM009 and 02ZM010 should be kept or removed needs to be further investigated.

Table 5.5 Results of discordancy measure (D_i) of studied sites in sub region A in Newfoundland

Region A								
ID	**Station Number**	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
1	02ZG001	Garnish River Near Garnish	49	66.64	0.275	0.403	0.265	1.5186
2	02ZG002	Tides Brook Below Freshwater Pond	19	50.01	0.233	0.259	0.289	0.6856
3	02ZG003	Salmonier River Near Lamaline	32	72.45	0.226	0.144	0.156	0.3987
4	02ZG004	Rattle Brook Near Boat Harbour	31	40.80	0.262	0.366	0.317	1.0903
5	02ZH002	Come By Chance River Near Goobies	41	32.76	0.214	0.041	0.090	1.4068
6	02ZK001	Rocky River Near Colinet	61	156.73	0.221	0.202	0.135	0.0443
7	02ZL004	Shearstown Brook At Shearstown	29	16.97	0.276	0.275	0.169	0.7942
8	02ZL005	Big Brook At Lead Cove	29	5.951	0.256	0.271	0.239	0.4709
9	02ZM006	Northeast Pond River At Northeast Pond	44	3.738	0.211	0.151	0.077	0.1974
10*	02ZM009*	Seal Cove Brook Near Cappahayden*	35	29.52	0.128	0.243	0.243	3.5876*
11*	02ZM010*	Waterford River At Mount Pearl	15	18.47	0.223	0.183	-0.071	3.1631*
12	02ZM016	South River Near Holyrood	31	11.594	0.200	0.153	0.153	0.1115
13	02ZM017	Leary Brook At St. John’s	15	13.24	0.178	0.143	0.219	0.8388
14	02ZM018	Virginia River At Pleasantville	28	9.694	0.179	0.098	0.069	0.4227
15	02ZN001	Northwest Brook At Northwest Pond	28	38.85	0.167	0.024	0.022	0.9736
16	02ZN002	St. Shotts River Near Trepassey	18	10.794	0.223	0.130	0.079	0.2957

*Means that the site has higher D_i value than the critical value of 3.

Table 5.6 Results of discordancy measure (D_i) of studied sites in sub region B in Newfoundland

Region B								
ID	**Station Number**	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
1	02YM001	India Brook At Indian Falls	40	147.62	0.150	0.050	0.200	2.1822
2	02YM003	South West Brook Near Baie Verte	31	41.04	0.247	0.213	0.133	0.5884
3	02YO006	Peters River Mear Botwood	32	47.83	0.224	0.342	0.310	0.9255
4	02YO008	Great Rattling Brook Above Tote River Cnfluence	22	214.4	0.178	0.143	0.073	0.8715
5	02YO012	Southwest Brook At Lewisporte	24	16.25	0.215	0.188	0.197	0.1433
6	02YQ005	Salmon River Near Glenwood	21	40.06	0.280	0.242	0.116	1.6723
7	02YR001	Middle Brook Near Gambo	50	30.11	0.179	0.134	0.159	0.1814
8	02YR002	Ragged Harbour River Near Musgrave Harbour	17	67.38	0.160	0.310	0.237	2.2374
9	02YR003	Indian Bay Brook Near Northwest Arm	31	61.98	0.185	0.122	0.074	0.5764
10	02YS001	Terra Nova River At Eight Mile Bridges	30	182.7	0.167	0.181	0.127	0.7295
11	02YS003	Southwest Brook At Terra Nova National Park	43	14.62	0.205	0.229	0.212	0.0776
12	02YS005	Terra Nova River At Glovertown	29	223.1	0.183	0.023	0.076	1.0723
13	02ZH001	Pipers Hole River At Mothers Brook	57	241.6	0.245	0.167	0.151	0.6337
14	02ZJ001	Southern Bay River Near Southern Bay	32	26.63	0.306	0.444	0.382	2.6801

Table 5.7 Results of discordancy measure (D_i) of studied sites in sub region C in Newfoundland

Region C								
ID	**Station Number**	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
1	02YA001	Ste. Genevieve River Near Forresters Point	25	33.15	0.186	0.251	0.080	1.3808
2	02YC001	Torrent River At Bristol's Pool	53	184.77	0.189	0.195	0.171	0.1034
3	02YD001	Beaver Brook Near Roddickton	19	103.95	0.185	0.224	0.199	0.1317
4	02YD002	Northeast Brook Near Roddickton	33	40.32	0.128	0.120	0.175	0.6755
5	02YE001	Greavett Brook Above Portland Creek Pond	26	46.13	0.180	0.244	0.141	0.4630
6	02YG001	Main River At Paradise Pool	26	314.9	0.140	-0.023	-0.024	1.2213
7	02YK004	Hinds Brook Near Grand Lake	22	94.13	0.141	0.060	-0.006	0.9749
8	02YK008	Boot Brook At Trans-Canada Highway	28	10.35	0.266	0.273	0.179	2.5440
9	02YL001	Upper Humber River Near Reidville	85	595.1	0.130	0.127	0.146	0.5054
10	02YL004	South Brook At Pasadena	29	44.89	0.264	0.500	0.364	1.9990
11	02YL008	Upper Humber River Above Black Brook	25	251.6	0.141	0.087	0.150	0.3178
12	02YM004	Indian Brook Diversion Above Birchy Lake	24	38.94	0.095	-0.080	0.104	1.6833

Table 5.8 Results of discordancy measure (D_i) of studied sites in sub region D in Newfoundland

Region D								
ID	**Station Number**	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
1	02YJ001	Harrys River Below Highway Bridge	42	311.9	0.206	0.187	0.153	0.2312
2	02YK002	Lewaseechjeech Brook At Little Grand Lake	50	111.15	0.176	0.143	0.150	1.1493
3	02YN002	Lloyds River Below King George Iv Lake	33	182	0.220	0.226	0.169	0.1630
4	02ZB001	Isle Aux Morts River Below Highway Bridge	51	373.7	0.261	0.224	0.102	1.1774
5	02ZC002	Grandy Brook Below Top Pond Brook	28	365.7	0.176	0.192	0.144	0.9955
6	02ZD002	Grey River Near Grey River	32	872.3	0.242	0.153	0.073	1.4245
7	02ZE001	Salmon River At Long Pond	16	292.2	0.157	-0.014	-0.112	2.2177
8	02ZE004	Conne River At Outlet of Conne River Pond	25	42.56	0.177	0.134	0.057	0.6588
9	02ZF001	Bay Du Nord River At Big Falls	61	210.2	0.208	0.255	0.254	0.9552
10	02ZK004	Little Salmonier River Near North Harbour	31	89.06	0.237	0.241	0.129	1.0275

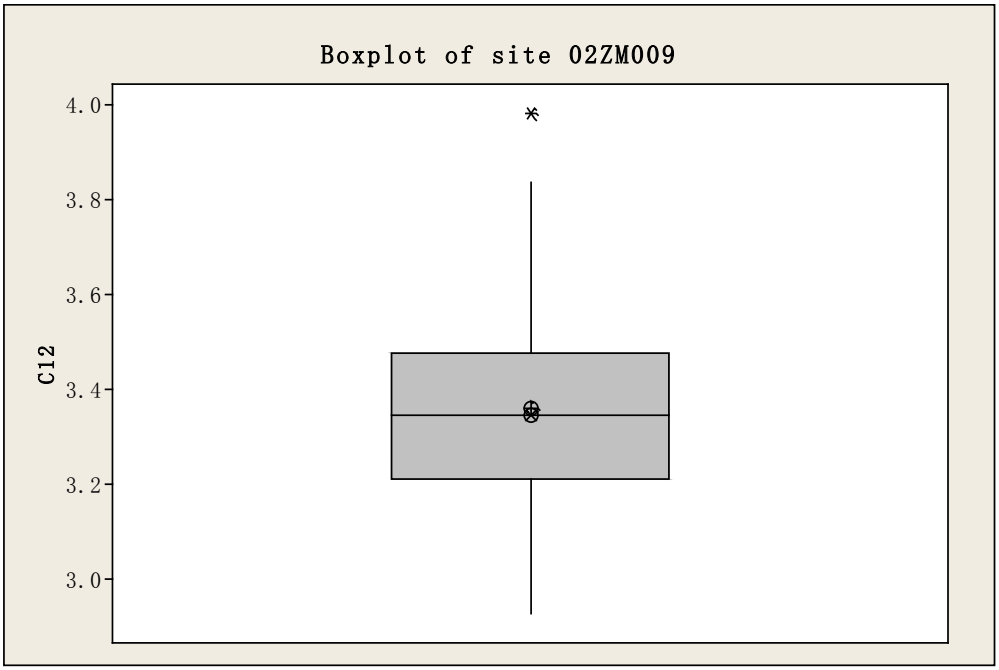


Figure 5.5 Boxplot logged data of site 02ZM009

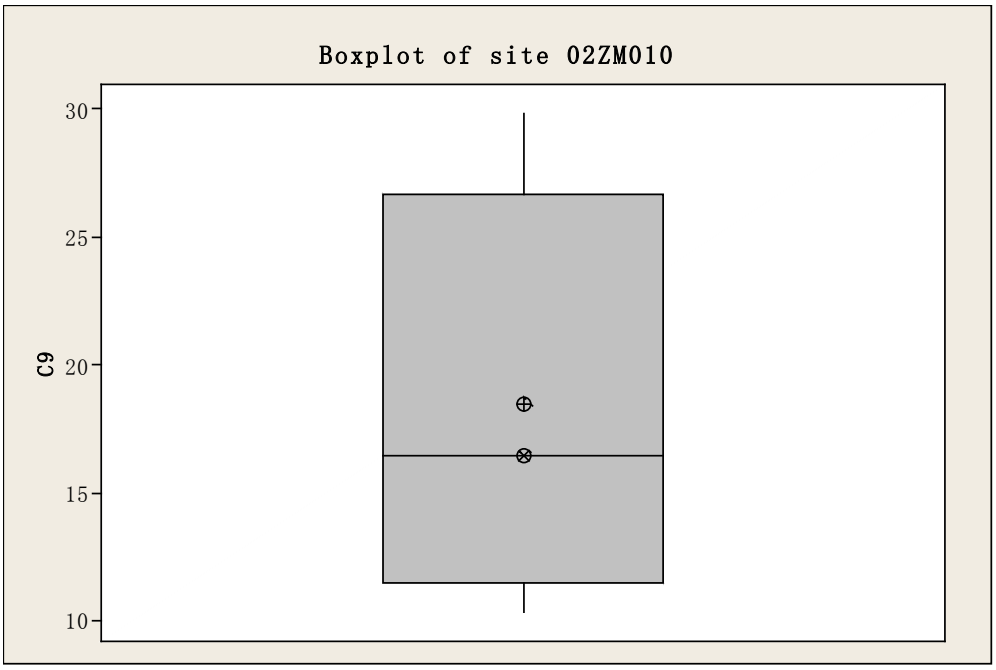


Figure 5.6 Boxplot of site 02ZM010

Table 5.9 Results of discordancy measure (D_i) of 15 selected sites excluding 02ZM009 in sub region A

ID	Station Number	Station Name	Years of Record	Mean Max Flow (m^3/s)	L-CV	L-SK	L-Ku	D_i
1	02ZG001	Garnish River Near Garnish	49	66.64	0.275	0.403	0.265	1.5010
2	02ZG002	Tides Brook Below Freshwater Pond	19	50.01	0.233	0.259	0.289	0.7133
3	02ZG003	Salmonier River Near Lamaline	32	72.45	0.226	0.144	0.156	0.5115
4	02ZG004	Rattle Brook Near Boat Harbour	31	40.80	0.262	0.366	0.317	1.1272
5	02ZH002	Come By Chance River Near Goobies	41	32.76	0.214	0.041	0.090	2.0722
6	02ZK001	Rocky River Near Colinet	61	156.73	0.221	0.202	0.135	0.0829
7	02ZL004	Shearstown Brook At Shearstown	29	16.97	0.276	0.275	0.169	1.2916
8	02ZL005	Big Brook At Lead Cove	29	5.951	0.256	0.271	0.239	0.4778
9	02ZM006	Northeast Pond River At Northeast Pond	44	3.738	0.211	0.151	0.077	0.1832
10	02ZM010*	Waterford River At Mount Pearl	15	18.47	0.223	0.183	-0.071	2.9609*
11	02ZM016	South River Near Holyrood	31	11.594	0.200	0.153	0.153	0.2697
12	02ZM017	Leary Brook At St. John’s	15	13.24	0.178	0.143	0.219	1.6200
13	02ZM018	Virginia River At Pleasantville	28	9.694	0.179	0.098	0.069	0.7526
14	02ZN001	Northwest Brook At Northwest Pond	28	38.85	0.167	0.024	0.022	1.0229
15	02ZN002	St. Shotts River Near Trepassey	18	10.794	0.223	0.130	0.079	0.4139

*Means that the site has higher D_i value than the critical value of 3.

Table 5.10 Results of discordancy measure (D_i) of 15 selected sites excluding 02ZM010 in sub region A

ID	Station Number	Station Name	Years of Record	Mean Max Flow (m^3/s)	L-CV	L-SK	L-Ku	D_i
1	02ZG001	Garnish River Near Garnish	49	66.64	0.275	0.403	0.265	1.7158
2	02ZG002	Tides Brook Below Freshwater Pond	19	50.01	0.233	0.259	0.289	1.1261
3	02ZG003	Salmonier River Near Lamaline	32	72.45	0.226	0.144	0.156	0.4118
4	02ZG004	Rattle Brook Near Boat Harbour	31	40.80	0.262	0.366	0.317	1.0579
5	02ZH002	Come By Chance River Near Goobies	41	32.76	0.214	0.041	0.090	1.4276
6	02ZK001	Rocky River Near Colinet	61	156.73	0.221	0.202	0.135	0.2652
7	02ZL004	Shearstown Brook At Shearstown	29	16.97	0.276	0.275	0.169	0.9625
8	02ZL005	Big Brook At Lead Cove	29	5.951	0.256	0.271	0.239	0.4568
9	02ZM006	Northeast Pond River At Northeast Pond	44	3.738	0.211	0.151	0.077	0.6868
10	02ZM009*	Seal Cove Brook Near Cappahayden	35	29.52	0.128	0.243	0.243	3.3551*
11	02ZM016	South River Near Holyrood	31	11.594	0.200	0.153	0.153	0.0840
12	02ZM017	Leary Brook At St. John’s	15	13.24	0.178	0.143	0.219	1.2680
13	02ZM018	Virginia River At Pleasantville	28	9.694	0.179	0.098	0.069	0.6326
14	02ZN001	Northwest Brook At Northwest Pond	28	38.85	0.167	0.024	0.022	1.0783
15	02ZN002	St. Shotts River Near Trepassey	18	10.794	0.223	0.130	0.079	0.4716

*Means that the site has higher D_i value than the critical value of 3.

5.2.2 Delineation of homogeneous regions

The heterogeneity measure Eq. [3.13] for the Island of Newfoundland is carried out for four sub regions respectively using Matlab program code (Appendix A-2). Due to the uncertainty of sites 02ZM009 and 02ZM010, the tests in sub region A will be carried out based on 16 sites , 15 sites (without 02ZM009 or without 02ZM010) and 14 sites (without sites 02ZM009 and 02ZM010), respectively. According to the simulation procedures discussed in Section 3.4.2, simulate 1000 regions, and then fit the kappa distribution to every regional L-moment ratios. Calculate mean and standard deviation of simulated sites. The results of the heterogeneity measure, the parameters of kappa distribution and regional weighted L-moments ratios are presented in Table 5.11.

The results of the heterogeneity measure indicate that sub regions A, B, C and D in the Newfoundland area are “possible homogeneous” as declared by Hosking and Wallis (1997). For sub region A, the region is shown to be more homogeneous in the presence of site 02ZM010, although the result of discordancy measure shows to be unsatisfactory. Therefore, in order to reach more accurate quantile estimations, site 02ZM010 will be kept for the following analysis and site 02ZM009 is removed.

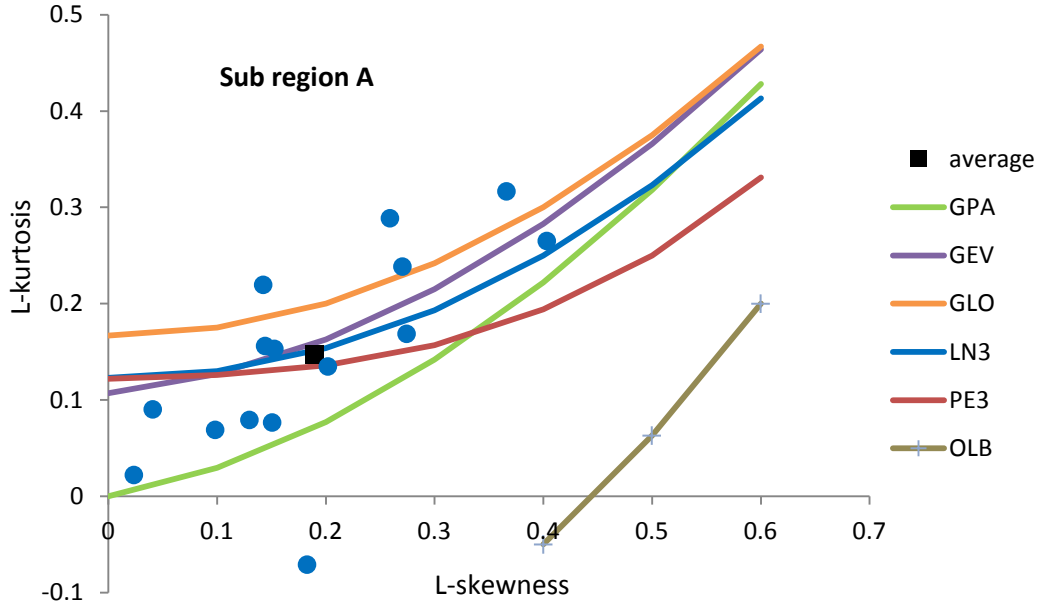
Table 5.11 Kappa parameters and results of the heterogeneity measure in Newfoundland

	Region A (Exclude site 02ZM009)	**Region A** (Exclude site 02ZM010)	**Region A** (16 sites)	**Region A** (14 sites)	**Region B**	**Region C**	**Region D**
$\mathbf{t^R}$	0.2257	0.2188	0.2189	0.22579	0.20768	0.16663	0.21056
$\mathbf{t_3^R}$	0.1968	0.2005	0.2000	0.19725	0.18930	0.16455	0.19052
$\mathbf{t_4^R}$	0.1517	0.1651	0.1581	0.15906	0.17780	0.14603	0.13602
V	0.0324	0.0405	0.0399	0.04359	0.04332	0.04856	0.03085
ξ	0.7812	0.8159	0.7994	0.79926	0.86743	0.85806	0.76440
α	0.3386	0.2970	0.3141	0.31956	0.24787	0.2467	0.35487
k	-0.0063	-0.0539	-0.0292	-0.03269	-0.09953	0.01672	0.05239
h	0.1277	-0.0272	0.0647	0.03565	-0.30416	0.02864	0.27108
μv	0.0173	0.0184	0.0173	0.01860	0.02130	0.0200	0.01640
σv	0.0133	0.0139	0.0132	0.01380	0.01600	0.0147	0.01240
H	**1.14**	**1.60**	**1.71**	**1.82**	**1.38**	**1.94**	**1.17**

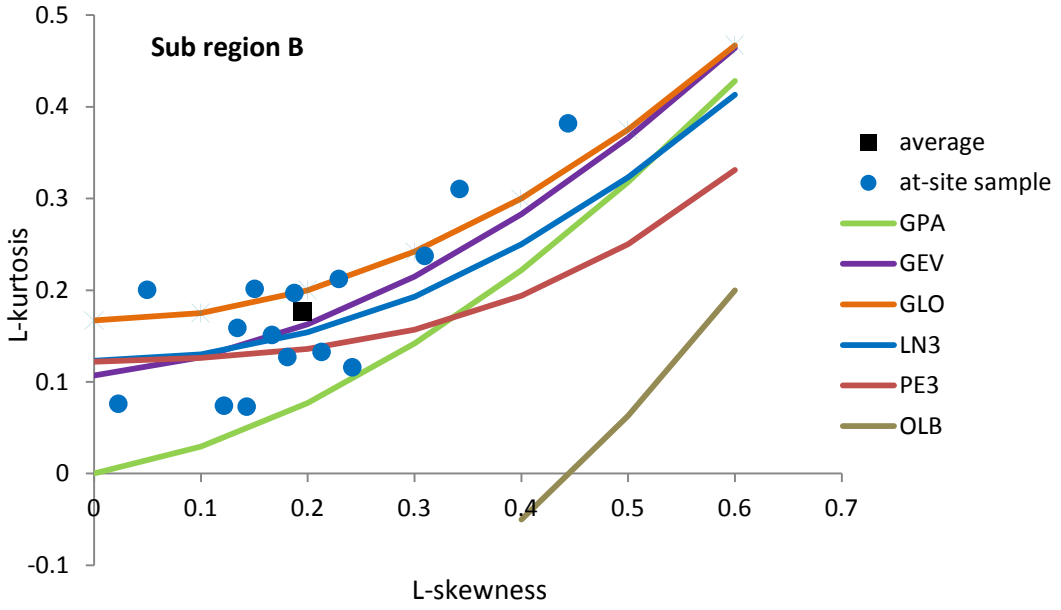
5.2.3 Selection of regional frequency distribution for four sub regions

5.2.3.1 L-moment ratio diagram

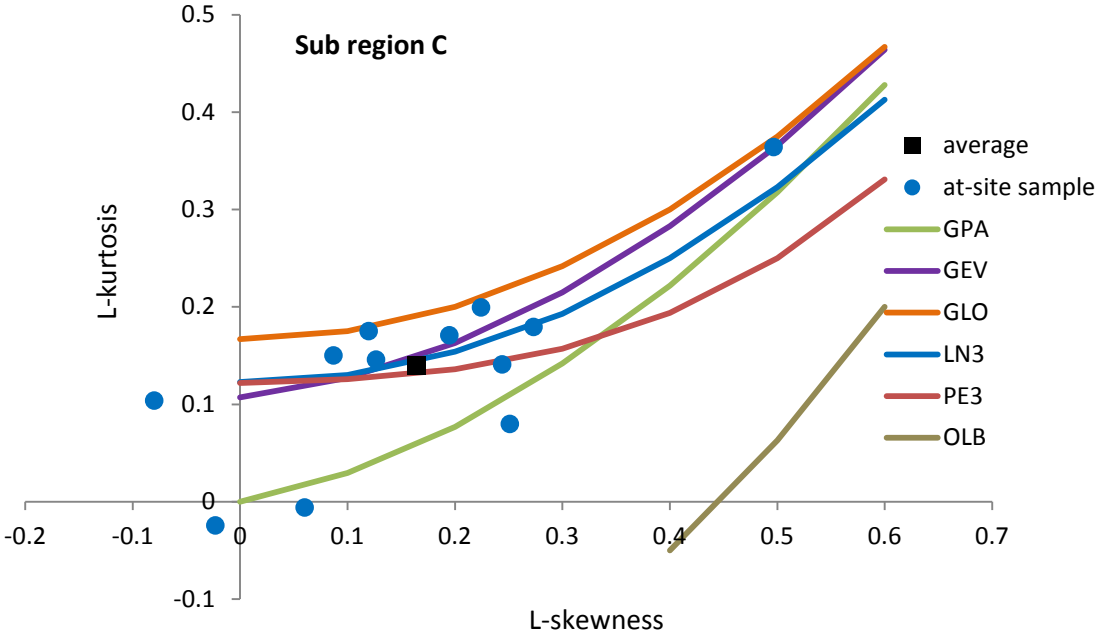
In order to make a simple and intuitive judgment from candidate distributions, the sample L-moment ratios, regional average L-moment ratios (black squares) of four sub regions in Newfoundland and theoretical curves of candidate distributions based on L-moment ratios are plotted in Figure 5.6.


Sub region A
L-kurtosis
L-skewness
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
average
GPA
GEV
GLO
LN3
PE3
OLB

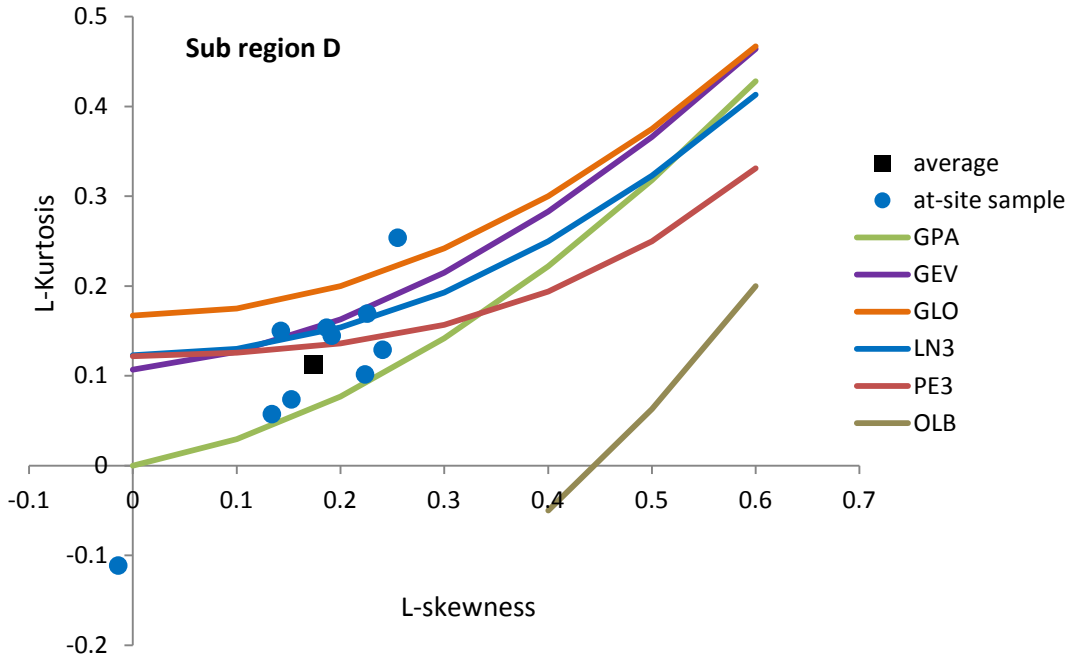
(a)


Sub region B
L-kurtosis
L-skewness
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
average
at-site sample
GPA
GEV
GLO
LN3
PE3
OLB

(b)



(c)



(d)

Figure 5.7 L-moment ratio diagram and regional L-moment ratios in Newfoundland

From the positions of regional average L-moment ratios it can be observed that in sub region A, B and C, the regional frequency distribution could be generated from the GEV distribution and lognormal distribution (LN3), and for region D the Pearson type Ⅲdistribution is probably a good choice.

5.2.3.2 Goodness-of-fit test

As discussed in Section 3.4.3.2, the goodness-of-fit test determines the best fitted regional frequency distribution by comparing the regional weighted L-kurtosis with that of the candidate distributions. Follow the procedures discussed in Chapter 3, Matlab program code (Appendix A-3) is employed to complete the computation. Table 5.12 provides the results of the goodness-of-fit measure of each sub region.

The bias and standard deviation of regional L-kurtosis of four sub regions (from A to D) are -0.0012, 0.0195; 0.0043, 0.0208; 0, 0.0194; -0.0001, 0.0209, respectively. From Table 5.12 it can be seen that for region A, the distributions-GEV, LN3 and PE3 are acceptable; however, only LN3 can be accepted as a result of its minimum $|Z^{DIST}|$. Equally, the GEV distribution is selected as the best fit regional frequency distribution for the sub regions A, B and C. In the sub region D the PE3 distribution is more acceptable than other distributions. However, the Pearson typeⅢis rarely applied in recent hydrological research and not convenient to use because of its mathematical complexity. In addition, the sample size in sub region D is small and the positions of

samples L-moment ratios of sub region D are scattered (Figure 5.2). Therefore, in order to achieve more accurate results the Pearson typelllwill not be considered to be the regional frequency distribution for sub region D. The next two with the best fit were used instead. These are the LN3 and GEV distributions.

Table 5.12 Results of goodness-of-fit measure for sub regions in Newfoundland

Region A					
	GLO	GEV	LN3*	PE3	GPA
τ_4^{DIST}	0.198945	0.161183	0.153019	0.135468	0.075427
Z^{DIST}	2.362758	0.424497	0.005473*	-0.895420	-3.977241
Region B					
	GLO	GEV*	LN3	PE3	GPA
τ_4^{DIST}	0.196533	0.157953	0.150751	0.134396	0.071296
Z^{DIST}	1.104365	-0.74787*	-1.093625	-1.878805	-4.908227
Region C					
	GLO	GEV*	LN3	PE3	GPA
τ_4^{DIST}	0.189234	0.147933	0.143897	0.131245	0.058340
Z^{DIST}	2.233141	0.099608*	-0.108886	-0.762440	-4.528563
Region D					
	GLO	GEV	LN3	PE3*	GPA
τ_4^{DIST}	0.196918	0.158471	0.151113	0.134566	0.071960
Z^{DIST}	2.911065	1.07139	0.719330	-0.072404*	-3.068072

*Represents the best fitted regional frequency distribution

5.2.3.3 Robustness test

The purpose of the robustness test is to recognize the most robust distribution when the number of acceptable distributions is more than one (Pokhrel, 2002). Misspecification of distribution could cause a large bias in quantile estimation (Hosking and Walls, 1997). Therefore, the robustness test is designed to compute the regional average relative bias (ARB) and relative root mean square error (RMSE) of the extreme quantiles when the distribution is correct and when the distribution is mis-specified. The bias is defined as Eq. [3.18] and the RMSE can be calculated from Eq. [3.19].

In sub regions A, B and C the robustness test is carried out between the best fitted distribution-GEV and the second best fitted distribution-LN3. In sub region D, as discussed in Section 5.2.3.2, the PE3 distribution will not be used for further analysis. Therefore, the robustness test is carried out between the GEV and LN3 distribution. Table 5.13 shows the results of ARB, RMSE and ARRB of extreme quantiles at each sub region when 1) The best fitted distribution is GEV but the underlying distribution is LN3; 2) The best fitted distribution is LN3 and the underlying distribution is GEV and 3) The best fitted and underlying distribution are the same.

Table 5.13 Results of robustness test for four sub regions in Newfoundland

Region A							
	Quantiles						**Difference for 100-year event**
	0.9	**0.99**	**0.999**	**0.9**	**0.99**	**0.999**	
	GEV-GEV			**GEV-LN3**			
ARB	2.45	4.19	7.53	2.20	5.36	9.37	1.17
AARB	4.10	18.03	32.49	4.76	17.33	28.28	-0.70
RMSE	0.66	3.10	9.01	0.65	2.75	6.73	-0.35
	LN3-LN3			**LN3-GEV**			
ARB	2.48	-1.97	-3.15	2.02	4.92	8.71	6.89
AARB	3.24	11.72	19.62	4.64	17.07	27.85	5.35
RMSE	0.49	1.54	3.60	0.63	2.70	6.64	1.16
Region B							
	Quantiles						**Difference for 100-year event**
	0.9	**0.99**	**0.999**	**0.9**	**0.99**	**0.999**	
	GEV-GEV			**GEV-LN3**			
ARB	2.36	2.51	3.73	1.66	2.86	4.61	0.35
AARB	5.84	16.81	29.18	6.10	15.84	24.21	-0.97
RMSE	0.75	2.97	7.86	0.65	2.48	5.53	-0.49
	LN3-LN3			**LN3-GEV**			
ARB	2.39	2.29	3.22	7.89	4.27	3.22	1.98
AARB	5.81	16.70	29.07	8.68	17.32	29.21	0.62
RMSE	0.68	2.82	7.49	1.01	2.80	6.96	-0.02

Table 5.13 Cont.

Region C							
	Quantiles						Difference for 100-year event
	0.9	0.99	0.999	0.9	0.99	0.999	
	GEV-GEV			GEV-LN3			
ARB	-24.43	-29.74	-34.41	-0.02	-0.32	0.84	-29.42
AARB	24.43	29.74	34.45	7.25	21.51	32.77	-8.23
RMSE	3.29	5.39	8.19	0.68	3.64	8.14	-1.75
	LN3-LN3			LN3-GEV			
ARB	5.23	3.87	5.26	5.42	3.38	5.05	-0.49
AARB	8.43	22.33	34.10	7.65	23.66	39.30	1.33
RMSE	0.84	3.87	8.54	0.90	4.23	10.82	0.36
Region D							
	Quantiles						Difference for 100-year event
	0.9	0.99	0.999	0.9	0.99	0.999	
	GEV-GEV			GEV-LN3			
ARB	100.90	166.70	234.48	2.33	6.56	10.85	-160.14
AARB	100.90	166.70	234.48	5.23	12.83	20.10	-153.87
RMSE	68.3257	252.61	862.69	0.56	2.21	5.46	-250.40
	LN3-LN3			LN3-GEV			
ARB	7.74	7.89	9.39	63.84	53.92	40.69	46.03
AARB	8.04	13.05	19.68	63.84	53.92	63.91	40.87
RMSE	0.82	2.10	4.45	73.29	81.61	106.23	79.51

ARB: Average relative bias
AARB: Average absolute relative bias
RMSE: Relative root mean square error

From the comparison of extreme quantiles for 100-year event, it can be seen that the LN3 distribution is more robust than the GEV distribution for all sub regions in Newfoundland. Therefore, the LN3 distribution is determined to be the regional frequency distribution for further quantile analysis.

5.2.4 Quantile estimation for four sub regions

As all of sub regions are tested to be homogeneous, the regional frequency distributions are determined, therefore, the estimation of quantile flow can take place. Based on the equation [3.20] discussed in Chapter 3, the quantile estimates are defined as the regional growth factor q(F) multiplied by the index flood, of which q(F) can be obtained from the regional growth curve fitted by the regional frequency distribution. For sub regions A, B, C and D, the quantile function fitted by the LN3 distribution is defined as equation [5.1]. The parameters of the LN3 distribution and regional quantile functions are presented in Table 5.14.

$$q(F)=Q_T/Q_{mean}=\begin{cases}\xi+\alpha k^{-1}[1-\exp\{-k.\Phi^{-1}(F)\}] & k\neq 0\\ \xi+\alpha.\ \Phi^{-1}(F) & k=0\end{cases} \qquad [5.1]$$

Table 5.14 Regional parameters of the LN3 distribution and regional quantile functions for sub regions in Newfoundland

Region A		
ξ (location)	α (scale)	k (shape)
0.9215	0.3702	-0.4065
LN3 Quantile Function $x(F)=0.9215+0.3702/(-0.4065)\ \{1-\exp\{0.4065\Phi^{-1}(F)\}]$		
Region B		
ξ (location)	α (scale)	k (shape)
0.9289	0.3374	-0.4143
LN3 Quantile Function $x(F)=0.9289+0.3374/(-0.4046)\{1-\exp\{0.4046\Phi^{-1}(F)\}]$		
Region C		
ξ (location)	α (scale)	k (shape)
0.9498	0.2879	-0.3390
LN3 Quantile Function $x(F)=0.9498+0.2879/(-0.3390)\{1-\exp\{0.3390\Phi^{-1}(F)\}]$		
Region D		
ξ (location)	α (scale)	k (shape)
0.9349	0.3513	-0.3589
LN3 Quantile Function $x(F)=0.9349+0.3513/(-0.3589)\{1-\exp\{0.3589\Phi^{-1}(F)\}]$		

5.2.5 Comparison of quantile estimation

In this Section, the results of quantile estimation are compared between at-site and

regional analysis. Tables 5.15-5.18 show the results of comparison of quantile estimation between at-site and regional analysis for four sub regions in Newfoundland.

Table 5.15 Comparison of at-site and regional frequency estimation for sub region A

Region A							
Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02ZG001	49	188.31	231.02	140.54	156.94	-33.98	-47.20
02ZG002	19	113.60	129.89	105.47	117.77	-7.71	-10.29
02ZG003	32	147.29	161.89	152.80	170.62	3.61	5.12
02ZG004	31	108.27	130.15	86.05	96.08	-25.83	-35.46
02ZH002	41	59.65	63.65	69.09	77.15	13.67	17.50
02ZK001	61	327.03	365.32	330.54	369.10	1.06	1.03
02ZL004	29	43.35	50.35	35.79	39.96	-21.13	-25.98
02ZL005	29	14.35	16.56	12.55	14.02	-14.36	-18.18
02ZM006	44	7.37	8.10	7.88	8.80	6.54	8.12
02ZM010	15	38.25	42.49	38.95	43.50	1.81	2.31
02ZM016	31	22.43	24.59	24.45	27.30	8.26	9.93
02ZM017	15	23.88	25.95	27.92	31.18	14.48	16.78
02ZM018	28	17.21	18.51	20.45	22.83	15.85	18.92
02ZN001	28	66.67	70.61	81.94	91.49	18.63	22.82
02ZN002	18	21.51	23.53	22.77	25.42	5.50	7.43
Average absolute %difference						**12.83**	**16.47**

Table 5.16 Comparison of at-site and regional frequency estimation for sub region B

Region B							
Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02YM001	40	237.52	251.25	396.72	329.19	19.95	23.68
02YM003	31	93.01	104.98	82.49	91.52	-12.75	-14.70
02YO006	32	113.42	133.63	96.14	106.66	-17.98	-25.29
02YO008	22	386.82	420.35	430.94	478.11	10.24	12.08
02YO012	24	33.59	37.35	32.66	36.24	-2.85	-3.08
02YQ005	21	98.88	113.37	80.52	89.33	-22.80	-26.91
02YR001	50	53.85	58.37	60.52	67.15	11.02	13.08
02YR002	17	130.27	148.30	135.43	150.26	3.82	1.30
02YR003	31	111.66	120.82	124.58	138.22	10.37	12.59
02YS001	30	332.40	364.36	367.23	407.42	9.48	10.57
02YS003	43	30.09	33.79	29.39	32.60	-2.39	-3.64
02YS005	29	403.81	429.40	448.43	497.51	9.95	13.69
02ZH001	57	525.14	583.61	485.62	538.77	-8.14	-8.32
02ZJ001	32	83.50	105.24	53.53	59.39	-55.99	-77.22
02ZJ002	22	25.09	27.31	27.96	31.02	10.26	11.97
Average absolute %difference						**13.87**	**17.21**

Table 5.17 Comparison of at-site and regional frequency estimation for sub region C

Region C							
Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02YA001	25	66.27	74.60	59.80	65.27	-10.81	-14.30
02YC001	53	359.77	398.38	333.33	363.81	-7.93	-9.50
02YD001	19	202.95	226.36	187.53	204.68	-8.23	-10.59
02YD002	33	63.50	67.75	72.74	79.39	12.71	14.67
02YE001	26	90.06	100.94	83.22	90.83	-8.23	-11.13
02YG001	26	484.73	505.82	568.08	620.04	14.67	18.42
02YK004	22	145.61	153.70	169.81	185.34	14.25	17.07
02YK008	28	25.77	29.85	18.67	20.38	-38.04	-46.49
02YL001	85	948.35	1014.29	1073.56	1171.75	11.66	13.44
02YL004	29	132.88	170.43	80.98	88.39	-64.09	-92.82
02YL008	25	396.12	420.50	453.89	495.40	12.73	15.12
02YM004	24	50.38	51.58	70.25	76.67	28.28	32.73
Average absolute %difference						**19.30**	**24.69**

Table 5.18 Comparison of at-site and regional frequency estimation for sub region D

Region D							
Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02YJ001	42	629.35	698.09	623.80	689.30	-0.89	-1.28
02YK002	50	199.37	216.52	222.30	245.64	10.32	11.86
02YN002	33	390.44	439.93	364.00	402.22	-7.27	-9.38
02ZB001	51	873.60	991.69	747.40	825.88	-16.89	-20.08
02ZC002	28	683.60	752.54	731.40	808.20	6.61	6.89
02ZD002	32	1860.05	2058.10	1744.60	1927.78	-6.67	-6.76
02ZE001	16	458.05	479.21	584.40	645.76	21.62	25.79
02ZE004	25	75.69	82.00	85.12	94.06	11.08	12.82
02ZF001	61	444.45	503.95	420.40	464.54	-5.72	-8.48
02ZK004	31	201.85	229.55	178.12	196.82	-64.09	-92.82
Average absolute %difference						**10.04**	**12.00**

Like the analysis in Labrador, AMEC (2014) estimated single site and regional quantile flows respectively in Newfoundland. Tables 5.20-5.23 provide the results of comparison of the quantile flows between current analysis and results obtained from AMEC (2014) and Pokhrel (2002). The regression equations and goodness-of-fit developed by AMEC (2014) for the Island of Newfoundland are shown in Table 5.19. In the study of AMEC (2014), the regions NE, SE, NW and SW represent the regions B, A, C and D used in this study respectively.

Table 5.19 Regression equations and goodness-of-fit developed by AMEC (2014) in Newfoundland

Region A (SE)					
One Parameter Equations	SMR	SEE	Two Parameters Equations	SMR	SEE
$Q_2=1.464*DA^{0.762}$	0.901	0.145	$Q_2=3.820*DA^{0.715}*LAF^{-0.180}$	0.938	0.120
$Q_5=1.966*DA^{0.768}$	0.905	0.143	$Q_5=5.135*DA^{0.721}*LAF^{-0.181}$	0.942	0.117
$Q_{10}=2.293*DA^{0.772}$	0.904	0.144	$Q_{10}=5.993*DA^{0.725}*LAF^{-0.181}$	0.941	0.118
$Q_{20}=2.604*DA^{0.775}$	0.903	0.146	$Q_{20}=6.809*DA^{0.728}*LAF^{-0.181}$	0.939	0.121
$Q_{50}=3.005*DA^{0.778}$	0.900	0.149	$Q_{50}=7.861*DA^{0.731}*LAF^{-0.181}$	0.936	0.125
$Q_{100}=3.306*DA^{0.780}$	0.897	0.152	$Q_{100}=8.651*DA^{0.733}*LAF^{-0.181}$	0.932	0.128
$Q_{200}=3.608*DA^{0.782}$	0.894	0.155	$Q_{200}=9.443*DA^{0.735}*LAF^{-0.181}$	0.929	0.132
Region B(NE)					
One Parameter Equations	SMR	SEE	Two Parameters Equations	SMR	SEE
$Q_2=0.836*DA^{0.755}$	0.902	0.161	$Q_2=2.911*DA^{0.767}*LAF^{-0.285}$	0.964	0.102
$Q_5=1.271*DA^{0.733}$	0.882	0.173	$Q_5=4.746*DA^{0.745}*LAF^{-0.302}$	0.954	0.112
$Q_{10}=1.582*DA^{0.722}$	0.870	0.181	$Q_{10}=6.128*DA^{0.734}*LAF^{-0.310}$	0.947	0.119
$Q_{20}=1.895*DA^{0.712}$	0.858	0.187	$Q_{20}=7.568*DA^{0.725}*LAF^{-0.317}$	0.940	0.126
$Q_{50}=2.322*DA^{0.702}$	0.844	0.195	$Q_{50}=9.597*DA^{0.715}*LAF^{-0.325}$	0.931	0.134
$Q_{100}=2.658*DA^{0.695}$	0.834	0.200	$Q_{100}=11.243*DA^{0.708}*LAF^{-0.330}$	0.925	0.140
$Q_{200}=3.009*DA^{0.688}$	0.824	0.205	$Q_{200}=12.997*DA^{0.702}*LAF^{-0.335}$	0.918	0.145

Table 5.19 Cont.

Region C(NW)					
One Parameter Equations	SMR	SEE	Two Parameters Equations	SMR	SEE
$Q_2=0.611*DA^{0.875}$	0.778	0.241	$Q_2=3.959*DA^{0.883}*LAF^{-0.408}$	0.952	0.117
$Q_5=0.974*DA^{0.834}$	0.751	0.248	$Q_5=6.496*DA^{0.842}*LAF^{-0.415}$	0.942	0.125
$Q_{10}=1.242*DA^{0.812}$	0.734	0.253	$Q_{10}=8.416*DA^{0.820}*LAF^{-0.418}$	0.934	0.131
$Q_{20}=1.519*DA^{0.795}$	0.718	0.257	$Q_{20}=10.421*DA^{0.803}*LAF^{-0.421}$	0.925	0.138
$Q_{50}=1.905*DA^{0.775}$	0.699	0.262	$Q_{50}=13.256*DA^{0.783}*LAF^{-0.424}$	0.915	0.145
$Q_{100}=2.216*DA^{0.761}$	0.686	0.266	$Q_{100}=15.563*DA^{0.770}*LAF^{-0.426}$	0.906	0.151
$Q_{200}=2.544*DA^{0.749}$	0.673	0.270	$Q_{200}=18.024*DA^{0.757}*LAF^{-0.428}$	0.898	0.157
Region D(SW)					
One Parameter Equations	SMR	SEE	Two Parameters Equations	SMR	SEE
$Q_2=7.864*DA^{0.497}$	0.495	0.327	$Q_2=90.931*DA^{0.523}*LAF^{-4.825}$	0.887	0.164
$Q_5=10.853*DA^{0.492}$	0.462	0.346	$Q_5=141.407*DA^{0.519}*LAF^{-5.060}$	0.871	0.179
$Q_{10}=12.845*DA^{0.490}$	0.444	0.356	$Q_{10}=178.118*DA^{0.517}*LAF^{-5.183}$	0.863	0.188
$Q_{20}=14.762*DA^{0.488}$	0.430	0.365	$Q_{20}=215.518*DA^{0.516}*LAF^{-5.284}$	0.855	0.195
$Q_{50}=17.264*DA^{0.485}$	0.415	0.375	$Q_{50}=267.085*DA^{0.514}*LAF^{-5.399}$	0.846	0.204
$Q_{100}=19.163*DA^{0.484}$	0.405	0.382	$Q_{100}=308.149*DA^{0.513}*LAF^{-5.475}$	0.840	0.210
$Q_{200}=21.084*DA^{0.482}$	0.395	0.388	$Q_{200}=351.240*DA^{0.512}*LAF^{-5.544}$	0.835	0.215

Table 5.20 Comparison of at-site and regional quantile flows for sub region A

Region A							
Station Number	Years of record	% Difference Current Study		% Difference AMEC (2014)		% Difference Pokhrel (2002)	
		Q50	Q100	Q50	Q100	Q50	Q100
02ZG001	49	-33.98	-47.20	-3.00	-4.80	-14.05	-21.35
02ZG002	19	-7.71	-10.29	NA	NA	-15.85	-23.53
02ZG003	32	3.61	5.12	-15.20	-15.7	10.65	11.99
02ZG004	31	-25.83	-35.46	-43.20	-44.20	-15.73	-20.71
02ZH002	41	13.67	17.50	-21.70	-22.90	8.69	11.44
02ZK001	61	1.06	1.03	4.8	5.1	-2.49	-4.31
02ZL004	29	-21.13	-25.98	22.2	19.4	NA	NA
02ZL005	29	-14.36	-18.18	19.80	16.80	NA	NA
02ZM006	44	6.54	8.12	-1.70	-2.2	1.56	2.82
02ZM010	35	1.81	2.31	-16.80	-17.00	NA	NA
02ZM016	31	8.26	9.93	5.70	5.70	NA	NA
02ZM017	15	14.48	16.78	-32.80	-31.90	NA	NA
02ZM018	28	15.85	18.92	57.60	59.60	NA	NA
02ZN001	28	18.63	22.82	29.50	32.80	12.64	14.94
02ZN002	18	5.50	7.43	-6.10	-5.80	NA	NA
Average absolute %difference		**12.83**	**16.47**	**20.01**	**20.28**	**10.21**	**13.89**

NA means the site is not available in that research.

Table 5.21 Comparison of at-site and regional quantile flows for sub region B

Region B							
Station Number	Years of record	% Difference Current Study		% Difference AMEC (2014)		% Difference Pokhrel (2002)	
		Q50	Q100	Q50	Q100	Q50	Q100
02YM001	40	19.95	23.68	69.00	73.10	NA	NA
02YM003	31	-12.75	-14.70	-23.40	-24.70	-28.92	-15.42
02YO006	32	-17.98	-25.29	10.40	10.00	-75.48	-113.68
02YO008	22	10.24	12.08	-16.50	-15.80	NA	NA
02YO012	24	-2.85	-3.08	22.70	23.50	NA	NA
02YQ005	21	-22.80	-26.91	-33.40	-34.20	NA	NA
02YR001	50	11.02	13.08	2.80	2.50	11.15	14.89
02YR002	17	3.82	1.30	46.30	48.80	-17.61	-24.28
02YR003	31	10.37	12.59	23.60	23.00	7.21	8.77
02YS001	30	9.48	10.57	19.00	21.20	6.87	8.15
02YS003	43	-2.39	-3.64	27.60	29.70	17.21	21.19
02YS005	29	9.95	13.69	11.60	10.40	NA	NA
02ZH001	57	-8.14	-8.32	-24.20	-26.20	-0.42	1.13
02ZJ001	32	-55.99	-77.22	-28.10	-29.90	0.1	1.27
02ZJ002	22	10.26	11.97	-6.70	-7.60	NA	NA
Average absolute %difference		**13.87**	**17.21**	**24.35**	**25.37**	**18.33**	**23.20**

NA means the site is not available in that research.

Table 5.22 Comparison of at-site and regional quantile flows for sub region C

Region C							
Station Number	Years of record	% Difference Current Study		% Difference AMEC (2014)		% Difference Pokhrel (2002)	
		Q50	Q100	Q50	Q100	Q50	Q100
02YA001	25	-10.81	-14.30	3.50	2.20	NA	NA
02YC001	53	-7.93	-9.50	-36.90	-38.70	NA	NA
02YD001	19	-8.23	-10.59	4.00	3.10	-13.04	-15.84
02YD002	33	12.71	14.67	-5.50	-4.20	-7.89	-11.02
02YE001	26	-8.23	-11.13	-21.60	-20.80	NA	NA
02YG001	26	14.67	18.42	24.80	25.20	NA	NA
02YK004	22	14.25	17.07	-9.40	-9.80	8.98	12.14
02YK008	28	-38.04	-46.49	10.10	8.30	NA	NA
02YL001	85	11.66	13.44	9.20	7.90	5.10	5.82
02YL004	29	-64.09	-92.82	-41.10	-42.40	NA	NA
02YL008	25	12.73	15.12	-22.00	-21.70	NA	NA
02YM004	24	28.28	32.73	79.60	84.80	NA	NA
Average absolute %difference		**19.30**	**24.69**	**22.31**	**22.43**	**8.75**	**11.21**

NA means the site is not available in that research.

Table 5.23 Comparison of at-site and regional quantile flows for sub region D

Region D							
Station Number	Years of record	% Difference Current Study		% Difference AMEC (2014)		% Difference Pokhrel (2002)	
		Q50	Q100	Q50	Q100	Q50	Q100
02YJ001	42	-0.89	-1.28	-26.80	-27.10	-2.66	-5.01
02YK002	50	10.32	11.86	-2.50	-0.50	-6.56	-6.08
02YN002	33	-7.27	-9.38	-46.50	-47.30	-18.89	-25.34
02ZB001	51	-16.89	-20.08	-51.80	-52.90	-4.39	-4.95
02ZC002	28	6.61	6.89	54.00	57.20	8.94	7.89
02ZD002	32	-6.67	-6.76	17.60	17.60	NA	NA
02ZE001	16	21.62	25.79	36.80	37.70	13.67	18.33
02ZE004	25	11.08	12.82	52.70	54.00	NA	NA
02ZF001	61	-5.72	-8.48	-10.10	-11.10	-24.36	-33.56
02ZK004	31	-64.09	-92.82	-2.20	-3.40	NA	NA
Average absolute %difference		**10.04**	**12.00**	**30.10**	**30.88**	**11.35**	**14.45**

NA means the site is not available in that research.

The flood data used in Pokhrel (2002) had been changed. Therefore, it is difficult to determine which study has more accurate quantile estimates compared to the results from this study. However, it is seen that the index-flood procedure provide a better agreement with the observed data than the method of regression on quantile (AMEC, 2014).

5.2.6 Quantile estimation at ungauged sites

As the index flood at ungauged sites is not known, the estimation of quantile flow at ungauged sites largely depends on the development of a linear or nonlinear regression relationship between the index flood and sites characteristics at gauged sites within a homogeneous region. Based on the previous researches (AMEC, 2014 & Pokhrel, 2002), the selections of parameters are not limited to the Drainage Area (DA) and Lake Attenuation Factor (LAF). Another parameter-Lakes and Swamps Factor (LSF) is selected in order to achieve a higher R^2 value. The data is log transformed. The nonlinear least square regression equations and R^2 are presented in Table 5.24.

Table 5.24 Nonlinear regression equations and R^2 for sub regions in Newfoundland

Sub region	Regression equation	R^2
A	$Q=2.517*DA^{0.783}*LAF^{-0.117}$	0.96
B	$Q=0.3815*DA^{0.708}*LAF^{-0.336}$	0.96
C	$Q=0.2582*DA^{0.845}*LAF^{-0.393}$	0.94
D	$Q=8.1827*DA^{0.509}*LSF^{-3.62}$	0.82

5.3 RFFA for Y and Z Sub Regions

The regionalization of sub regions Y and Z was proposed by the Water Survey of

Canada (WSC). Pokhrel (2002) achieved more accurate results of quantile estimates based on this regionalization. Therefore, in the following sections, the results of RFFA based on the index flood procedures for sub regions Y and Z will be presented in a step by step manner.

5.3.1 Data screening and discordancy measure

Tables 5.25-5.26 list the basic information, L-moment ratios and results of discordancy measure of sub region Y and Z. 27sites are involved in sub region Y, and sub region Z contains 26 sites.

Hosking and Wallis (1997) suggested the critical value of D_i for a group with more than 15 sites should be equal to or less than 3. Therefore, from Tables 5.25-5.26 it can be observed that all of the sites are not discordant from other sites in sub region Y. However, in sub region Z, sites 02ZM009 and 02ZM010 still show higher Di value which match the results obtained from the discordancy measure in sub region A. Tables 5.27-5.29 show the discordancy measure after removing site 02ZM009 or site 02ZM010 and sites 02ZM009 and 02ZM010 respectively. After removing sites 02ZM009 and 02ZM010, all of the sites in sub region Z are shown to be not discordant.

Table 5.25 Summary of statistics and discordancy measure of sub region Y

Region Y								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
02YM001	Indian Brook At Indian Falls	40	147.62	974	0.150	0.050	0.200	1.902
02YM003	South West Brook Near Baie Verte	31	41.04	93.2	0.247	0.213	0.133	0.986
02YO006	Peters River Mear Botwood	32	47.83	177	0.224	0.342	0.310	1.323
02YO008	Great Rattling Brook Above Tote River Confluence	22	214.4	773	0.178	0.143	0.073	0.477
02YO012	Southwest Brook At Lewisporte	24	16.25	58.7	0.215	0.188	0.197	0.559
02YQ005	Salmon River Near Glenwood	21	40.06	80.8	0.280	0.242	0.116	2.305
02YR001	Middle Brook Near Gambo	50	30.11	275	0.179	0.134	0.159	0.161
02YR002	Ragged Harbour River Near Musgrave Harbour	17	67.38	399	0.160	0.310	0.237	1.945
02YR003	Indian Bay Brook Near Northwest Arm	31	61.98	554	0.185	0.122	0.074	0.361
02YS001	Terra Nova River At Eight Mile Bridges	30	182.7	1290	0.167	0.181	0.127	0.326
02YS003	Southwest Brook At Terra Nova National Park	43	14.62	36.7	0.205	0.229	0.212	0.234
02YS005	Terra Nova River At Glovertown	29	223.1	2000	0.183	0.023	0.076	1.179
02YA001	Ste. Genevieve River Near Forresters Point	25	33.15	306	0.186	0.251	0.080	2.038
02YC001	Torrent River At Bristol’s Pool	53	184.77	624	0.189	0.195	0.171	0.022
02YD001	Beaver Brook Near Roddickton	19	103.95	237	0.185	0.224	0.199	0.173

Table 5.25 Cont.

Region Y								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
02YD002	Northeast Brook Near Roddickton	33	40.32	200	0.128	0.120	0.175	0.869
02YE001	Greavett Brook Above Portland Creek Pond	26	46.13	95.7	0.180	0.244	0.141	0.752
02YG001	Main River At Paradise Pool	26	314.9	627	0.140	-0.023	-0.024	1.597
02YK004	Hinds Brook Near Grand Lake	22	94.13	529	0.141	0.060	-0.006	1.703
02YK008	Boot Brook At Trans-Canada Highway	28	10.35	20.4	0.266	0.273	0.179	1.253
02YL001	Upper Humber River Near Reidville	85	595.1	2110	0.130	0.127	0.146	0.705
02YL004	South Brook At Pasadena	29	44.89	58.5	0.264	0.500	0.364	2.878
02YL008	Upper Humber River Above Black Brook	25	251.6	471	0.141	0.087	0.150	0.481
02YM004	Indian Brook Diversion Above Birchy Lake	24	38.94	238	0.095	-0.080	0.104	2.397
02YJ001	Harrys River Below Highway Bridge	42	311.9	640	0.206	0.187	0.153	0.111
02YK002	Lewaseechjeech Brook At Little Grand Lake	50	111.15	470	0.176	0.143	0.150	0.051
02YN002	Lloyds River Below King George Iv Lake	33	182	469	0.220	0.226	0.169	0.213

Table 5.26 Summary of statistics and discordancy measure of sub region Z

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
02ZG001	Garnish River Near Garnish	49	66.64	205	0.275	0.403	0.265	1.605
02ZG002	Tides Brook Below Freshwater Pond	19	50.01	166	0.233	0.259	0.289	0.706
02ZG003	Salmonier River Near Lamaline	32	72.45	115	0.226	0.144	0.156	0.575
02ZG004	Rattle Brook Near Boat Harbour	31	40.80	42.7	0.262	0.366	0.317	0.941
02ZH002	Come By Chance River Near Goobies	41	32.76	43.3	0.214	0.041	0.090	1.946
02ZK001	Rocky River Near Colinet	61	156.73	301	0.221	0.202	0.135	0.033
02ZL004	Shearstown Brook At Shearstown	29	16.97	28.9	0.276	0.275	0.169	0.740
02ZL005	Big Brook At Lead Cove	29	5.951	11.2	0.256	0.271	0.239	0.402
02ZM006	Northeast Pond River At Northeast Pond	44	3.738	3.63	0.211	0.151	0.077	0.145
02ZM009*	Seal Cove Brook Near Cappahayden	35	29.52	53.6	0.128	0.243	0.243	3.955*
02ZM010*	Waterford River At Mount Pearl	15	18.47	16.6	0.223	0.183	-0.071	3.386*
02ZM016	South River Near Holyrood	31	11.594	17.3	0.200	0.153	0.153	0.170
02ZM017	Leary Brook At St. John’s	15	13.24	15.3	0.178	0.143	0.219	1.105
02ZM018	Virginia River At Pleasantville	28	9.694	10.7	0.179	0.098	0.069	0.336
02ZN001	Northwest Brook At Northwest Pond	28	38.85	53.3	0.167	0.024	0.022	0.896
02ZN002	St. Shotts River Near Trepassey	18	10.794	15.5	0.223	0.130	0.079	0.314

*****Means the site has a higher D_i value than the critical one.

Table 5.26 Cont.

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
02ZH001	Pipers Hole River At Mothers Brook	57	241.6	764	0.245	0.167	0.151	0.646
02ZJ001	Southern Bay River Near Southern Bay	32	26.63	67.4	0.306	0.444	0.382	2.156
02ZJ002	Salmon Cove River Near Champneys	22	13.91	73.6	0.175	0.151	0.201	0.813
02ZB001	Isle Aux Morts River Below Highway Bridge	51	373.7	205	0.261	0.224	0.102	0.674
02ZC002	Grandy Brook Below Top Pond Brook	28	365.7	230	0.176	0.192	0.144	0.661
02ZD002	Grey River Near Grey River	32	872.3	1340	0.242	0.153	0.073	0.527
02ZE001	Salmon River At Long Pond	16	292.2	2640	0.157	-0.014	-0.112	1.908
02ZE004	Conne River At Outlet of Conne River Pond	25	42.56	99.5	0.177	0.134	0.057	0.564
02ZF001	Bay Du Nord River At Big Falls	61	210.2	1170	0.208	0.255	0.254	0.478
02ZK004	Little Salmonier River Near North Harbour	31	89.06	104	0.237	0.241	0.129	0.319

Table 5.27 Results of discordancy measure in sub region Z excluding site 02ZM009

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
02ZG001	Garnish River Near Garnish	49	66.64	205	0.275	0.403	0.265	1.629
02ZG002	Tides Brook Below Freshwater Pond	19	50.01	166	0.233	0.259	0.289	0.712
02ZG003	Salmonier River Near Lamaline	32	72.45	115	0.226	0.144	0.156	0.623
02ZG004	Rattle Brook Near Boat Harbour	31	40.80	42.7	0.262	0.366	0.317	0.991
02ZH002	Come By Chance River Near Goobies	41	32.76	43.3	0.214	0.041	0.090	2.281
02ZK001	Rocky River Near Colinet	61	156.73	301	0.221	0.202	0.135	0.042
02ZL004	Shearstown Brook At Shearstown	29	16.97	28.9	0.276	0.275	0.169	0.882
02ZL005	Big Brook At Lead Cove	29	5.951	11.2	0.256	0.271	0.239	0.386
02ZM006	Northeast Pond River At Northeast Pond	44	3.738	3.63	0.211	0.151	0.077	0.129
02ZM010*	Waterford River At Mount Pearl	15	18.47	16.6	0.223	0.183	-0.071	3.249*
02ZM016	South River Near Holyrood	31	11.594	17.3	0.200	0.153	0.153	0.205
02ZM017	Leary Brook At St. John’s	15	13.24	15.3	0.178	0.143	0.219	1.335
02ZM018	Virginia River At Pleasantville	28	9.694	10.7	0.179	0.098	0.069	0.411
02ZN001	Northwest Brook At Northwest Pond	28	38.85	53.3	0.167	0.024	0.022	0.862
02ZN002	St. Shotts River Near Trepassey	18	10.794	15.5	0.223	0.130	0.079	0.375
02ZH001	Pipers Hole River At Mothers Brook	57	241.6	764	0.245	0.167	0.151	0.850
02ZJ001	Southern Bay River Near Southern Bay	32	26.63	67.4	0.306	0.444	0.382	2.062

Table 5.27 Cont.

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**D_i**
02ZJ002	Salmon Cove River Near Champneys	22	13.91	73.6	0.175	0.151	0.201	1.193
02ZB001	Isle Aux Morts River Below Highway Bridge	51	373.7	205	0.261	0.224	0.102	0.813
02ZC002	Grandy Brook Below Top Pond Brook	28	365.7	230	0.176	0.192	0.144	1.399
02ZD002	Grey River Near Grey River	32	872.3	1340	0.242	0.153	0.073	0.730
02ZE001	Salmon River At Long Pond	16	292.2	2640	0.157	-0.014	-0.112	1.835
02ZE004	Conne River At Outlet of Conne River Pond	25	42.56	99.5	0.177	0.134	0.057	0.843
02ZF001	Bay Du Nord River At Big Falls	61	210.2	1170	0.208	0.255	0.254	0.862
02ZK004	Little Salmonier River Near North Harbour	31	89.06	104	0.237	0.241	0.129	0.303

*Means the D_i value is higher than the critical value of 3.

Table 5.28 Results of discordancy measure in sub region Z excluding site 02ZM010

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**Di**
02ZG001	Garnish River Near Garnish	49	66.64	205	0.275	0.403	0.265	1.823
02ZG002	Tides Brook Below Freshwater Pond	19	50.01	166	0.233	0.259	0.289	0.894
02ZG003	Salmonier River Near Lamaline	32	72.45	115	0.226	0.144	0.156	0.618
02ZG004	Rattle Brook Near Boat Harbour	31	40.80	42.7	0.262	0.366	0.317	0.894
02ZH002	Come By Chance River Near Goobies	41	32.76	43.3	0.214	0.041	0.090	2.076
02ZK001	Rocky River Near Colinet	61	156.73	301	0.221	0.202	0.135	0.102
02ZL004	Shearstown Brook At Shearstown	29	16.97	28.9	0.276	0.275	0.169	0.828
02ZL005	Big Brook At Lead Cove	29	5.951	11.2	0.256	0.271	0.239	0.385
02ZM006	Northeast Pond River At Northeast Pond	44	3.738	3.63	0.211	0.151	0.077	0.278
02ZM009*	Waterford River At Mount Pearl	15	18.47	16.6	0.223	0.183	-0.071	3.795*
02ZM016	South River Near Holyrood	31	11.594	17.3	0.200	0.153	0.153	0.164
02ZM017	Leary Brook At St. John’s	15	13.24	15.3	0.178	0.143	0.219	1.409
02ZM018	Virginia River At Pleasantville	28	9.694	10.7	0.179	0.098	0.069	0.346
02ZN001	Northwest Brook At Northwest Pond	28	38.85	53.3	0.167	0.024	0.022	0.852
02ZN002	St. Shotts River Near Trepassey	18	10.794	15.5	0.223	0.130	0.079	0.326
02ZH001	Pipers Hole River At Mothers Brook	57	241.6	764	0.245	0.167	0.151	0.630
02ZJ001	Southern Bay River Near Southern Bay	32	26.63	67.4	0.306	0.444	0.382	2.075

Table 5.28 Cont.

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**Di**
02ZJ002	Salmon Cove River Near Champneys	22	13.91	73.6	0.175	0.151	0.201	0.930
02ZB001	Isle Aux Morts River Below Highway Bridge	51	373.7	205	0.261	0.224	0.102	0.937
02ZC002	Grandy Brook Below Top Pond Brook	28	365.7	230	0.176	0.192	0.144	0.704
02ZD002	Grey River Near Grey River	32	872.3	1340	0.242	0.153	0.073	0.6.4
02ZE001	Salmon River At Long Pond	16	292.2	2640	0.157	-0.014	-0.112	2.461
02ZE004	Conne River At Outlet of Conne River Pond	25	42.56	99.5	0.177	0.134	0.057	0.803
02ZF001	Bay Du Nord River At Big Falls	61	210.2	1170	0.208	0.255	0.254	0.477
02ZK004	Little Salmonier River Near North Harbour	31	89.06	104	0.237	0.241	0.129	0.588

*Means the D_i value is higher than the critical value of 3.

Table 5.29 Results of discordancy measure in sub region Z excluding sites 02ZM010 and 02ZM009

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**Di**
02ZG001	Garnish River Near Garnish	49	66.64	205	0.275	0.403	0.265	1.845
02ZG002	Tides Brook Below Freshwater Pond	19	50.01	166	0.233	0.259	0.289	0.883
02ZG003	Salmonier River Near Lamaline	32	72.45	115	0.226	0.144	0.156	0.664
02ZG004	Rattle Brook Near Boat Harbour	31	40.80	42.7	0.262	0.366	0.317	0.939
02ZH002	Come By Chance River Near Goobies	41	32.76	43.3	0.214	0.041	0.090	2.405
02ZK001	Rocky River Near Colinet	61	156.73	301	0.221	0.202	0.135	0.110
02ZL004	Shearstown Brook At Shearstown	29	16.97	28.9	0.276	0.275	0.169	0.948
02ZL005	Big Brook At Lead Cove	29	5.951	11.2	0.256	0.271	0.239	0.369
02ZM006	Northeast Pond River At Northeast Pond	44	3.738	3.63	0.211	0.151	0.077	0.259
02ZM016	South River Near Holyrood	31	11.594	17.3	0.200	0.153	0.153	0.195
02ZM017	Leary Brook At St. John’s	15	13.24	15.3	0.178	0.143	0.219	1.593
02ZM018	Virginia River At Pleasantville	28	9.694	10.7	0.179	0.098	0.069	0.419
02ZN001	Northwest Brook At Northwest Pond	28	38.85	53.3	0.167	0.024	0.022	0.818
02ZN002	St. Shotts River Near Trepassey	18	10.794	15.5	0.223	0.130	0.079	0.378
02ZH001	Pipers Hole River At Mothers Brook	57	241.6	764	0.245	0.167	0.151	0.828
02ZJ001	Southern Bay River Near Southern Bay	32	26.63	67.4	0.306	0.444	0.382	1.981
02ZJ002	Salmon Cove River Near Champneys	22	13.91	73.6	0.175	0.151	0.201	1.271

Table 5.29 Cont.

Region Z								
Station Number	**Station Name**	**Years of Record**	**Mean Max Flow (m^3/s)**	**Drainage Area (km^3)**	**L-CV**	**L-SK**	**L-Ku**	**Di**
02ZB001	Isle Aux Morts River Below Highway Bridge	51	373.7	205	0.261	0.224	0.102	1.043
02ZC002	Grandy Brook Below Top Pond Brook	28	365.7	230	0.176	0.192	0.144	1.425
02ZD002	Grey River Near Grey River	32	872.3	1340	0.242	0.153	0.073	0.783
02ZE001	Salmon River At Long Pond	16	292.2	2640	0.157	-0.014	-0.112	2.367
02ZE004	Conne River At Outlet of Conne River Pond	25	42.56	99.5	0.177	0.134	0.057	1.077
02ZF001	Bay Du Nord River At Big Falls	61	210.2	1170	0.208	0.255	0.254	0.836
02ZK004	Little Salmonier River Near North Harbour	31	89.06	104	0.237	0.241	0.129	0.563

5.3.2 Heterogeneity rest

In order to determine whether sites 02ZM009 and 02ZM010 should be removed or kept, the heterogeneity test will be carried out in region Z with 26sites, 25 sites and 24 sites respectively. Table 5.30 shows the results of the heterogeneity test for regions Y and Z.

Table 5.30 Results of heterogeneity measure for sub regions Y and Z

Region Y H=2.0	t^R	t_3^R	t_4^R	**V**	ξ
	0.18297	0.17019	0.15444	0.04238	0.85536
	α	**k**	**h**	**μv**	**σv**
	0.25558	-0.01484	-0.05241	0.0165	0.0130
Region Z (25sites, exclude site 02ZM009 H=1.26	t^R	t_3^R	t_4^R	**V**	ξ
	0.22624	0.20289	0.15425	0.0360802	0.77904
	α	**k**	**h**	**μv**	**σv**
	0.33663	-0.01536	0.13081	0.0184	0.014
Region Z (25sites, exclude site 02ZM010 H=1.59	t^R	t_3^R	t_4^R	**V**	ξ
	0.22223	0.20489	0.16194	0.0406556	0.79959
	α	**k**	**h**	**μv**	**σv**
	0.31209	-0.04262	0.042925	0.0188	0.0138
Region Z (24 sites) H=1.29	t^R	t_3^R	t_4^R	**V**	ξ
	0.22629	0.20325	0.15842	0.0364104	0.78915
	α	**k**	**h**	**μv**	**σv**
	0.32584	-0.030251	0.079516	0.0187	0.0138
Region Z (26 sites) H=1.57	t^R	t_3^R	t_4^R	**V**	ξ
	0.22225	0.20452	0.15788	0.0402997	0.78994
	α	**k**	**h**	**μv**	**σv**
	0.32224	-0.028276	0.094096	0.0183	0.014

From the results shown in Table 5.30, it can be concluded that sub region Y is "possible homogeneous" and that four tests in sub region Z with a different number of sites are found to be "possible homogeneous" as well. Although the result of the discordancy measure indicates that site 02ZM010 is discordant from other sites, the region Z shows more homogeneous when site 02ZM010 is included. Therefore, it is reasonable to keep site 02ZM010 for the further estimation. Site 02ZM009 is removed.

5.3.3 Selection of regional frequency distribution

As discussed in Chapter 3, in this thesis, the determination of regional frequency distribution is based on the results of the L-moment ratio diagram, the goodness-of-test and the robustness test.

5.3.3.1 L-moment ratio diagram

Figures 5.7-5.8 plot the sample L-moment ratios, regional average L-moment ratios and theoretical curves based on the L-moments of candidate distributions for regions Y and Z.

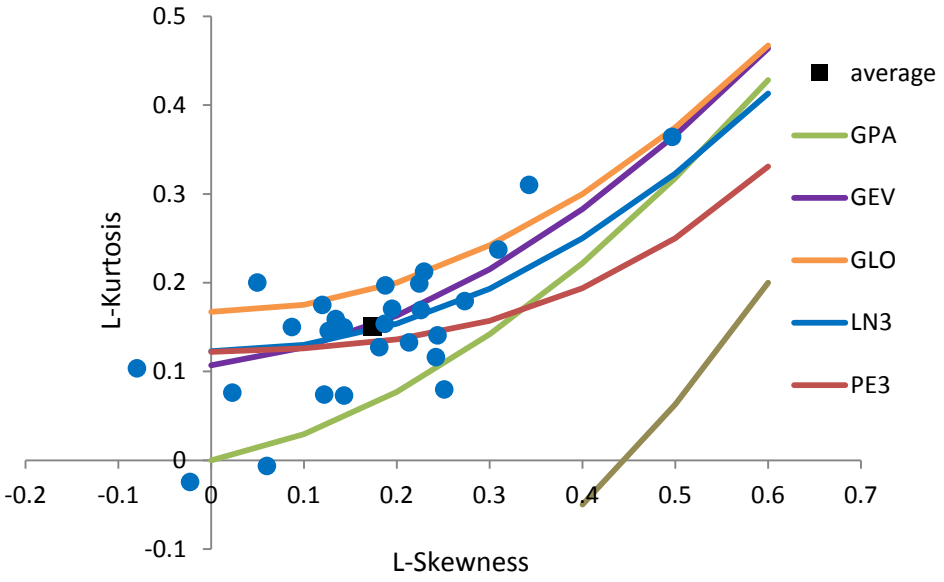


Figure 5.8 L-moment ratio diagram for sub region Y

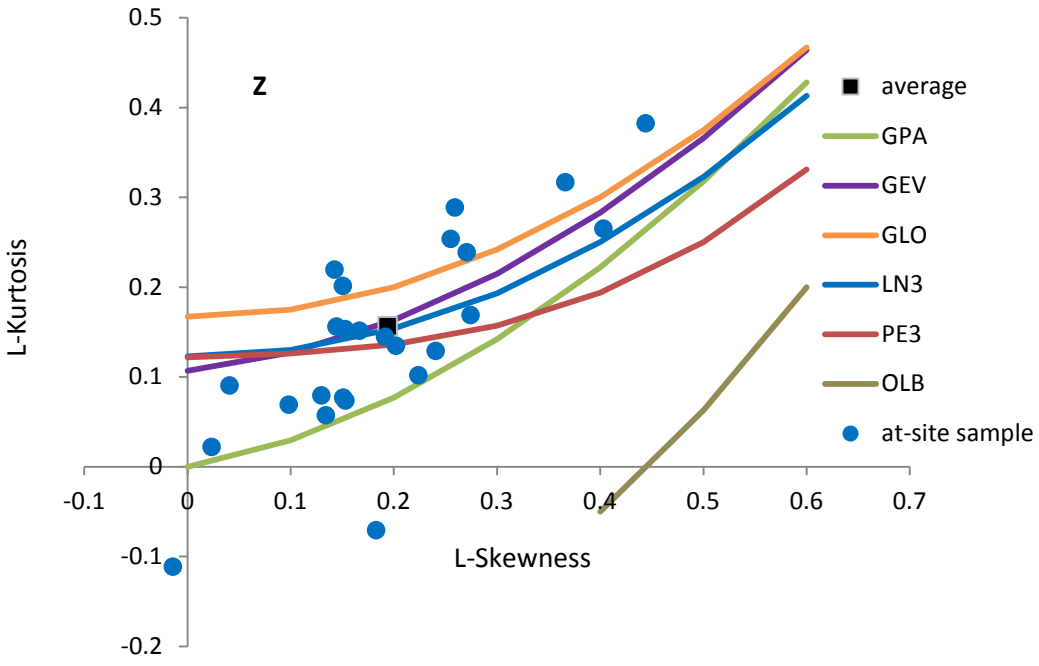


Figure 5.9 L-moment ratio diagram for sub region Z

From the Figures 5.7-5.8 it can be observed that in sub region Y the position of regional average L-moment ratios and most of the sample L-moment ratios are close to the GEV and LN3 distribution. In sub region Z the position of regional average L-moment ratios are closer to the LN3 distribution.

5.3.3.2 Goodness-of-fit test

The goodness-of-fit test is carried out for sub regions Y and Z respectively. The results shown in Table 5.31 illustrate that the three parameters lognormal distribution has a strong ability to provide a better fit frequency distribution to sample data for sub regions Y and Z, which supported the conclusions from the L-moment ratio diagrams. The bias and standard deviation of region L-Kurtosis for sub regions Y and Z are: 0.0022, 0.0132; 0.0035, 0.0147, respectively.

Table 5.31 Results of goodness-of-fit test for candidate distributions in sub regions Y and Z

Region Y					
	GLO	GEV	LN3	PE3	GPA
τ_4^{DIST}	0.192016	0.151801	0.146507	0.132430	0.063369
Z^{DIST}	3.298398	0.258057	-0.142121	-1.206381	-6.427466
Region Z					
	GLO	GEV	LN3	PE3	GPA
τ_4^{DIST}	0.200973	0.163874	0.154929	0.136381	0.078853
Z^{DIST}	3.109469	0.527723	-0.094806	-1.38557	-5.388966

5.3.3.3 Robustness test

The robustness test for sub region Y and Z are carried out to realize the more robust regional distribution between the generalized extreme value and three-parameter lognormal distribution. The test is organized to compare the bias and the root mean square (RMSE) of the extreme quantiles between these two distributions from the following aspects: 1) the underlying distribution is LN3 when the chosen distribution is GEV (GEV-LN3)-the underlying and chosen are GEV distribution (GEV-GEV); 2) the underlying distribution is GEV when the chosen distribution is LN3 (LN3-GEV)-the underlying and chosen are LN3 distribution (LN3-LN3). The results of average relative bias (ARB), average absolute relative bias (AARB) and relative root mean square error (RMSE) of extreme quantiles in sub regions Y and Z are presented in Table 5.32.

From the results shown in Table 5.32 it can be seen that in sub region Y the percentage of difference of bias and RMSE of extreme quantiles of GEV-LN3 for 100-year event is lower than those of GEV-GEV. Similarly, in sub region Z, the bias and RMSE of extreme quantiles of GEV-GEV is higher than in the GEV-LN3 case. Therefore, it can be conclude that the LN3 is the most robust regional frequency distribution for both Y and Z region, which matches the results obtained from the goodness-of-fit test.

Table 5.32 Results of robustness test for sub regions Y and Z

Region Y							
	Quantiles						**Difference for 100-year event**
	0.9	**0.99**	**0.999**	**0.9**	**0.99**	**0.999**	
	GEV-GEV			**GEV-LN3**			
ARB	0.32	-0.83	-0.64	0.12	-0.22	0.07	0.61
AARB	5.30	17.98	30.61	5.83	16.76	26.02	-1.22
RMSE	0.59	2.86	7.54	0.56	2.55	5.74	-0.31
	LN3-LN3			**LN3-GEV**			
ARB	5.49	-0.22	0.07	5.68	2.74	2.61	2.96
AARB	5.83	16.76	26.02	7.41	18.55	31.54	1.79
RMSE	0.56	2.55	5.74	0.75	3.01	7.84	0.46
Region Z							
	0.9	**0.99**	**0.999**	**0.9**	**0.99**	**0.999**	**Difference for 100-year event**
	GEV-GEV			**GEV-LN3**			
ARB	3.61	6.71	11.26	3.15	7.48	12.66	1.01
AARB	5.07	18.75	33.59	5.73	18.09	29.12	-0.65
RMSE	0.69	3.41	9.88	0.66	3.06	7.55	-0.30
	LN3-LN3			**LN3-GEV**			
ARB	9.03	8.27	9.84	9.53	7.60	8.69	-0.93
AARB	9.20	18.44	27.69	9.53	19.27	32.25	0.67
RMSE	1.04	2.99	6.44	1.12	3.37	8.75	0.28

ARB: Average relative bias
AARB: Average absolute relative bias
RMSE: Relative root mean square error

5.3.4 Quantile estimation

The regional parameters of the LN3 distribution and its quantile functions for sub regions Y and Z are shown in Table 5.33. Tables 5.34-5.35 present the results of comparison between at-site and regional analysis in regions Y and Z respectively.

Table 5.33 Regional parameters of LN3 distribution and LN3 quantile functions in sub regions Y and Z

Region Y		
ξ (location)	α (scale)	k (shape)
0.9411	0.3175	-0.3595
LN3 Quantile Function $x(F)=0.9411+0.3175/(-0.3595)\{1-\exp\{0.3595\Phi^{-1}(F)\}]$		
Region Z		
ξ (location)	α (scale)	k (shape)
0.9192	0.3684	-0.4193
LN3 Quantile Function $x(F)=0.9192+0.3684/(-0.4193)\{1-\exp\{0.4193\Phi^{-1}(F)\}]$		

Table 5.34 Results of comparison between at-site and regional analysis in region Y

Region Y							
Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02YM001	40	237.54	251.26	281.36	309.41	15.57	18.79
02YM003	31	93.00	104.98	78.22	86.02	-18.90	-22.04
02YO006	32	113.42	133.63	91.16	100.25	-24.42	-33.30
02YO008	22	386.80	420.35	408.65	449.38	53.47	6.46
02YO012	24	33.59	37.35	30.97	34.06	-8.46	-9.66
02YQ005	21	98.88	113.37	76.35	83.97	-29.51	-35.01
02YR001	50	53.85	58.37	57.39	63.11	6.17	7.51
02YR002	17	130.28	148.32	128.43	141.23	-1.44	-5.02
02YR003	31	111.66	120.82	118.13	129.91	5.48	6.99
02YS001	30	332.37	364.32	348.23	382.94	4.56	4.86
02YS003	43	30.09	33.79	27.87	30.64	-7.97	-10.28
02YS005	29	391.65	415.55	425.23	467.62	7.90	11.14
02YA001	25	66.27	74.60	63.18	69.48	-4.89	-7.37
02YC001	53	359.77	398.38	352.17	387.28	-2.16	-2.87
02YD001	19	202.94	226.35	198.13	217.88	-2.43	-3.89
02YD002	33	63.46	67.70	76.85	84.51	17.42	19.89
02YE001	26	90.06	100.93	87.92	96.69	-2.43	-4.39
02YG001	26	484.73	505.86	600.20	660.03	19.24	23.36
02YK004	22	145.60	153.68	179.41	197.3	18.85	22.11
02YK008	28	25.77	29.85	19.73	21.69	-30.61	-37.62
02YL001	85	948.35	1014.35	1134.26	1247.33	16.39	18.68
02YL004	29	132.88	170.44	85.56	94.09	-55.31	-81.15
02YL008	25	396.12	420.50	459.52	527.10	13.80	20.22
02YM004	24	50.39	51.58	74.22	81.62	32.11	36.80
02YJ001	42	403.60	698.13	594.48	653.74	32.11	-6.79
02YK002	50	199.35	216.50	211.85	232.97	5.90	7.07
02YN002	33	390.44	439.93	346.89	381.47	-12.55	-15.32
					Average	**15.02**	**17.73**

Table 5.35 Results of comparison between at-site and regional analysis in region Z

Region Y							
Station Number	Years of record	At-site		Regional		% Difference	
		Q50	Q100	Q50	Q100	Q50	Q100
02ZG001	49	188.59	231.24	141.21	158.00	-35.55	-46.35
02ZG002	19	113.52	130.03	105.97	118.57	-7.13	-9.66
02ZG003	32	147.07	161.56	153.52	171.78	4.20	5.95
02ZG004	31	108.12	130.15	86.46	96.74	-25.06	-34.54
02ZH002	41	59.62	63.55	69.42	77.67	14.11	18.18
02ZK001	61	327.57	365.18	332.11	371.61	1.37	1.73
02ZL004	29	43.27	50.40	35.96	40.24	-20.34	-25.26
02ZL005	29	14.34	16.54	12.61	14.11	-13.73	-17.25
02ZM006	44	7.36	8.07	7.92	8.86	7.03	8.90
02ZM010	15	38.23	42.48	39.14	43.79	2.31	2.99
02ZM016	31	22.38	24.58	24.57	27.49	8.92	10.59
02ZM017	15	23.83	25.95	28.06	31.39	15.05	17.33
02ZM018	28	17.16	18.52	20.54	22.98	16.47	19.44
02ZN001	28	66.82	70.71	82.32	92.11	18.83	23.24
02ZN002	18	21.48	23.53	22.87	25.59	6.09	8.06
02ZH001	57	524.27	584.67	511.95	572.83	-2.41	-2.07
02ZJ001	32	83.62	105.19	56.43	63.14	-48.18	-66.60
02ZJ002	22	25.04	27.26	29.48	32.98	15.05	17.33
02ZB001	51	874.46	990.30	791.87	886.04	-10.43	-11.77
02ZC002	28	683.86	753.34	774.92	867.07	11.75	13.12
02ZD002	32	1858.86	2058.63	1848.40	2068.22	-0.52	0.46
02ZE001	16	455.83	476.29	619.17	692.81	26.38	31.25
02ZE004	25	75.76	82.14	90.18	100.91	15.99	18.60
02ZF001	61	443.52	504.48	445.41	498.38	0.42	-1.22
02ZK004	31	202.17	229.77	188.72	211.16	-7.13	-8.81
					Average	**13.30**	**16.83**

Tables 5.36-5.37 present the results of the comparison of quantile flows for 50 and

100 return years in regions Y and Z.

Table 5.36 Comparison of at-site and regional frequency estimates between current research and Pokhrel's research (2002) in sub region Y

Region Y				
	% difference (Current Study)		**% difference (2002)**	
Station Number	**Q50**	**Q100**	**Q50**	**Q100**
02YM001	15.57	18.79	NA	NA
02YM003	-18.90	-22.04	-15.96	-15.87
02YO006	-24.42	-33.30	19.89	21.47
02YO008	53.47	6.46	NA	NA
02YO012	-8.46	-9.66	NA	NA
02YQ005	-29.51	-35.01	NA	NA
02YR001	6.17	7.51	13.32	15.89
02YR002	-1.44	-5.02	-15.36	-24.65
02YR003	5.48	6.99	21.84	27.49
02YS001	4.56	4.86	13.86	13.59
02YS003	-7.97	-10.28	-3.47	-4.91
02YS005	7.90	11.14	NA	NA
02YA001	-4.89	-7.37	NA	NA
02YC001	-2.16	-2.87	-6.50	-7.33
02YD001	-2.43	-3.89	-6.08	-7.50
02YD002	17.42	19.89	-1.86	-3.02
02YE001	-2.43	-4.39	NA	NA
02YG001	19.24	23.36	NA	NA
02YK004	18.85	22.11	20.84	23.44
02YK008	-30.61	-37.62	NA	NA
02YL001	16.39	18.68	14.82	17.70
02YL004	-55.31	-81.15	NA	NA
02YL008	17.40	20.26	NA	NA
02YM004	32.11	36.80	NA	NA
02YJ001	32.11	-6.79	-1.72	-1.56
02YK002	5.90	7.07	-1.15	-2.89
02YN002	-12.55	-15.32	NA	NA
Average	**15.02**	**17.73**	**11.20**	**13.38**

Table 5.37 Comparison of at-site and regional frequency estimates between current research and Pokhrel's research (2002) in sub region Z

Region Z				
	% difference (Current Study)		**% difference (2002)**	
Station Number	**Q50**	**Q100**	**Q50**	**Q100**
02ZG001	-35.55	-46.35	5.15	5.22
02ZG002	-7.13	-9.66	-3.58	-4.44
02ZG003	4.20	5.95	-9.61	-12.06
02ZG004	-25.06	-34.54	-2.25	-1.54
02ZH002	14.11	18.18	2.87	5.83
02ZK001	1.37	1.73	-2.11	-3.21
02ZL004	-20.34	-25.26	NA	NA
02ZL005	-13.73	-17.25	NA	NA
02ZM006	7.03	8.90	-5.07	-7.11
02ZM010	2.31	2.99	NA	NA
02ZM016	8.92	10.59	NA	NA
02ZM017	15.05	17.33	NA	NA
02ZM018	16.47	19.44	NA	NA
02ZN001	18.83	23.24	20.05	23.83
02ZN002	6.09	8.06	NA	NA
02ZH001	-2.41	-2.07	7.72	11.38
02ZJ001	-48.18	-66.60	1.67	1.13
02ZJ002	15.05	17.33	NA	NA
02ZB001	-10.43	-11.77	-20.78	-24.93
02ZC002	11.75	13.12	-0.89	-1.70
02ZD002	-0.52	0.46	NA	NA
02ZE001	26.38	31.25	25.68	30.05
02ZE004	15.99	18.60	NA	NA
02ZF001	0.42	-1.22	-13.29	-15.55
02ZK004	-7.13	-8.81	NA	NA
Average	**13.30**	**16.83**	**9.28**	**11.38**

Compared to the research of 2002 this research involves more gauged sites. From the

results shown in Tables 5.36-5.37 it can be seen that the percentage of difference between the at-site frequency estimation and regional estimation in the 2002 study is less than the result obtained from this research. Table 5.38 presents comparison of quantile flows between the current study and Pokhrel's research in 2002. However, the data used by Pokhrel had been changed by the Water Survey of Canada (WSC), so it is hard to compare the results of this study to those from Pokhrel (2002).

Table 5.38 compares the percentage of difference between at-site and regional quantile estimates based on results from this study and those obtained from AMEC (2014). The at-site and regional quantiles in this study are obtained based on the regional quantile functions in regions Y and Z respectively. From the results of comparison it can be concluded that the regional quantile function has a better agreement to the observed data than the regression on quantile approach. The index-flood procedure can get more accurate quantile estimates than the regression models conducted by AMEC (2014).

Table 5.38 Comparison of regional frequency estimates for sub regions Y and Z

	Regional Quantile Flow					
Station Number	**Current Study**		**Results of 2002**			
	Q50	**Q100**	**Q50**	**Q100**	**% d Q50**	**% d Q100**
02YM001	281.36	309.41	NA	NA	NA	NA
02YM003	78.22	86.02	78.30	86.30	0.10	0.12
02YO006	91.16	100.25	95.00	104.80	4.04	4.34
02YO008	408.65	449.38	NA	NA	NA	NA
02YO012	30.97	34.06	NA	NA	NA	NA
02YQ005	76.35	83.97	NA	NA	NA	NA
02YR001	57.39	63.11	54.80	60.40	-4.73	-4.49
02YR002	128.43	141.23	130.90	144.40	1.89	2.20
02YR003	118.13	129.91	107.60	118.60	-9.79	-9.54
02YS001	348.23	382.94	339.00	373.80	-2.72	-2.45
02YS003	27.87	30.64	25.90	28.50	-7.61	-7.51
02YS005	425.23	467.62	NA	NA	NA	NA
02YA001	63.18	69.48	NA	NA	NA	NA
02YC001	352.17	387.28	370.90	409.00	5.05	5.31
02YD001	198.13	217.88	192.30	212.10	-3.03	-2.73
02YD002	76.85	84.51	75.20	82.90	-2.19	-1.94
02YE001	87.92	96.69	NA	NA	NA	NA
02YG001	600.20	660.03	NA	NA	NA	NA
02YK004	179.41	197.3	171.80	189.40	-4.43	-4.17
02YK008	19.73	21.69	NA	NA	NA	NA
02YL001	1134.26	1247.33	1112.90	1227.20	-1.92	-1.64
02YL004	85.56	94.09	NA	NA	NA	NA
02YL008	479.55	527.35	NA	NA	NA	NA
02YM004	74.22	81.62	NA	NA	NA	NA
02YJ001	594.48	653.74	623.30	687.30	4.62	4.88
02YK002	211.85	232.97	235.30	259.50	9.97	10.22
02YN002	346.89	381.47	362.00	399.20	4.17	4.44
02ZG001	141.21	158.00	122.30	136.10	-15.46	-16.09
02ZG002	105.97	118.57	103.30	114.90	-2.58	-3.19
02ZG003	153.52	171.78	125.90	140.10	-21.94	-22.61

Table 5.38 Cont.

	Regional Quantile Flow					
Station Number	**Current Study**		**Results of 2002**			
	Q50	**Q100**	**Q50**	**Q100**	**% d Q50**	**% d Q100**
02ZG004	86.46	96.74	75.70	84.20	-14.21	-14.89
02ZH002	69.42	77.67	66.20	73.70	-4.86	-5.39
02ZK001	332.11	371.61	322.20	358.50	-3.08	-3.66
02ZL004	35.96	40.24	NA	NA	NA	NA
02ZL005	12.61	14.11	NA	NA	NA	NA
02ZM006	7.92	8.86	6.90	7.60	-14.78	-16.58
02ZM010	39.14	43.79	NA	NA	NA	NA
02ZM016	24.57	27.49	NA	NA	NA	NA
02ZM017	28.06	31.39	NA	NA	NA	NA
02ZM018	20.54	22.98	NA	NA	NA	NA
02ZN001	82.32	92.11	78.80	87.70	-4.47	-5.03
02ZN002	22.87	25.59	NA	NA	NA	NA
02ZH001	511.95	572.83	489.80	545.00	-4.52	-5.11
02ZJ001	56.43	63.14	47.90	53.30	-17.81	-18.46
02ZJ002	29.48	32.98	NA	NA	NA	NA
02ZB001	791.87	886.04	784.10	872.50	-0.99	-1.55
02ZC002	774.92	867.07	783.00	871.20	1.03	0.47
02ZD002	1848.40	2068.22	NA	NA	NA	NA
02ZE001	619.17	692.81	596.10	663.30	-3.87	-4.45
02ZE004	90.18	100.91	NA	NA	NA	NA
02ZF001	445.41	498.38	438.70	488.10	-1.53	-2.11
02ZK004	188.72	211.16	NA	NA	NA	NA
				Average	**6.12**	**6.41**

Table 5.39 Comparison of regional frequency estimates of studied sites in Newfoundland

	% difference (Current Study)		% difference (2014)	
Station Number	**Q50**	**Q100**	**Q50**	**Q100**
02YM001	15.57	18.79	69.00	73.10
02YM003	-18.90	-22.04	-23.40	-24.70
02YO006	-24.42	-33.30	10.40	10.00
02YO008	53.47	6.46	-16.50	-15.80
02YO012	-8.46	-9.66	22.70	23.50
02YQ005	-29.51	-35.01	-33.40	-34.20
02YR001	6.17	7.51	2.80	2.50
02YR002	-1.44	-5.02	46.30	48.80
02YR003	5.48	6.99	23.60	23.00
02YS001	4.56	4.86	19.00	21.20
02YS003	-7.97	-10.28	27.60	29.70
02YS005	7.90	11.14	11.60	10.40
02YA001	-4.89	-7.37	3.50	2.20
02YC001	-2.16	-2.87	-36.90	-38.70
02YD001	-2.43	-3.89	4.00	3.10
02YD002	17.42	19.89	-5.50	-4.20
02YE001	-2.43	-4.39	-21.60	-20.80
02YG001	19.24	23.36	24.80	25.20
02YK004	18.85	22.11	-9.40	-9.80
02YK008	-30.61	-37.62	10.10	8.30
02YL001	16.39	18.68	9.20	7.90
02YL004	-55.31	-81.15	-41.10	-42.40
02YL008	17.40	20.26	-22.00	-21.70
02YM004	32.11	36.80	79.60	84.80
02YJ001	32.11	-6.79	-26.80	-27.10
02YK002	5.90	7.07	-2.50	-0.50
02YN002	-12.55	-15.32	-46.50	-47.30
02ZG001	-35.55	-46.35	-3.00	-4.80
02ZG002	-7.13	-9.66	NA	NA
02ZG003	4.20	5.95	-15.20	-15.70

Table 5.39 Cont.

	% difference (Current Study)		% difference (2014)	
Station Number	**Q50**	**Q100**	**Q50**	**Q100**
02ZG004	-25.06	-34.54	-43.20	-44.20
02ZN002	6.09	8.06	-6.10	-5.80
02ZH001	-2.41	-2.07	-24.20	-26.20
02ZJ001	-48.18	-66.60	-28.10	-29.90
02ZJ002	15.05	17.33	-6.70	-7.60
02ZB001	-10.43	-11.77	-51.80	-52.90
02ZC002	11.75	13.12	54.00	57.20
02ZD002	-0.52	0.46	17.60	17.60
02ZE001	26.38	31.25	36.80	37.70
02ZE004	15.99	18.60	52.70	54.00
02ZF001	0.42	-1.22	-10.10	-11.10
02ZK004	-7.13	-8.81	-2.20	-3.40
Absolute average	**15.16**	**17.29**	**23.81**	**24.36**

5.4 Verification of the results

Similarly to the verification in Labrador, eight sites in sub region Y are selected to assess the accuracy of the regional flood frequency model. The basic information of the tested sites, including the station number, station name, length of record, drainage area and range of record is listed in Table 5.40. Figure 5.7 shows that the estimated regional quantiles agrees well with the observed data.

Table 5.40 Basic information of verification stations in sub region Y in Newfoundland

Station Number	Station Name	Length of Years	Drainage Area	Range of Records(year)
02YF001	Cat Arm River Above Great Cat Arm	14	611	1969-1982
02YH001	Bottom Creek Near Rocky Harbour	12	33.4	1985-1997
02YJ003	Pinchgut Brook At Outlet Of Pinchgut Lake	11	119	1986-1996
02YK003	Sheffield River At Sheffield Lake	11	362	1956-1966
02YK007	Glide Brook Below Glide Lake	11	112	1984-1996
02YO007	Leech Brook Near Grand Falls	7	88.3	1987-1995
02YP001	Shoal Arm Brook Near Badger Bay	13	63.8	1982-1996
02YQ004	Northwest Gander River Near Gander Lake	10	2200	1985-1998

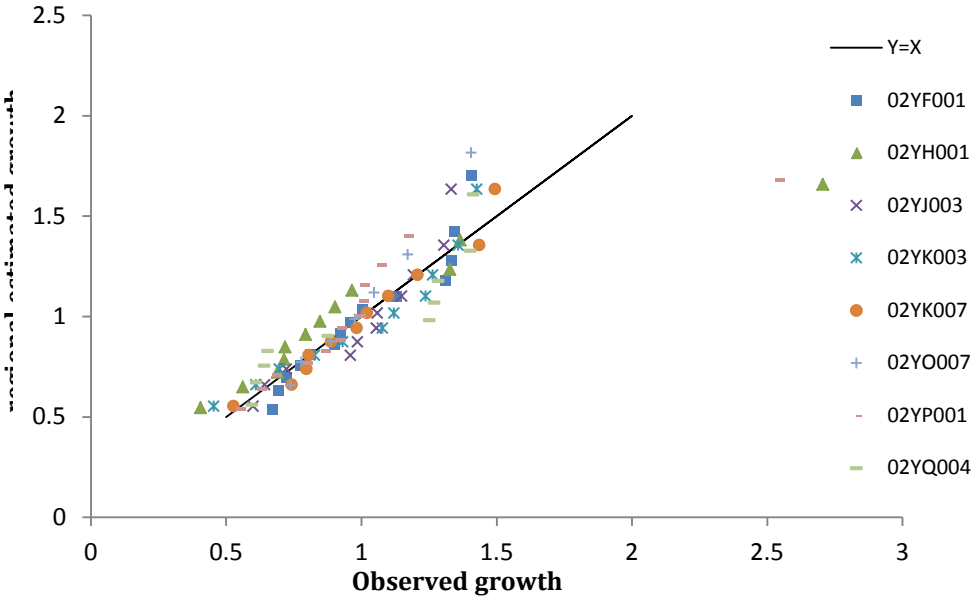


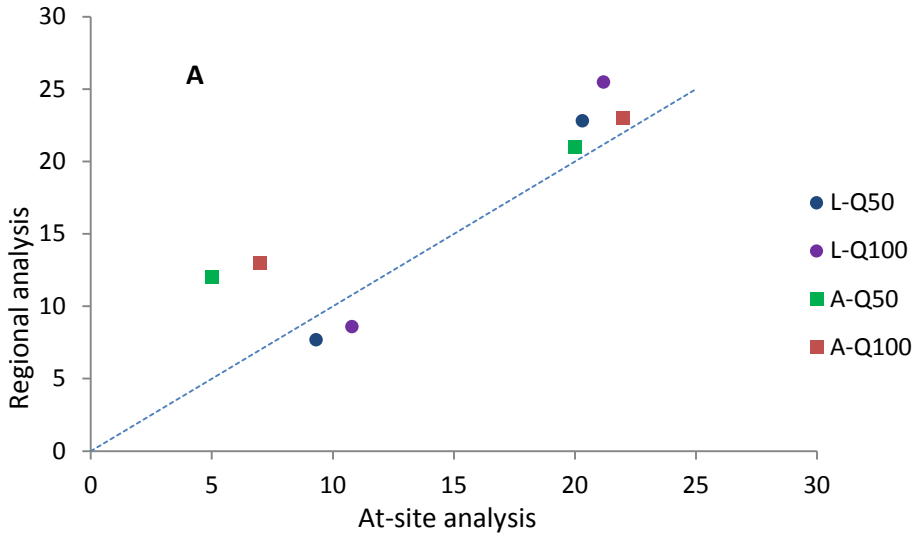
Figure 5.10 Regional frequency model has a good agreement with observed value

The quantile estimates between at-site and regional analysis based on the index flood procedure and regression models (AMEC, 2014) are compared using the data at gauged sites which are not included in this study due to short flood data. Table 5.41 lists the flood information of tested sites in four sub regions.

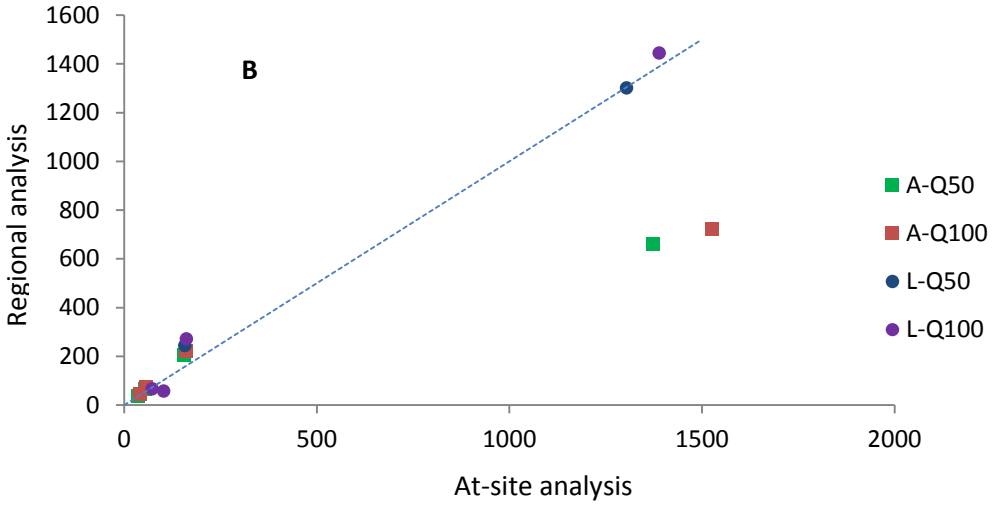
Table 5.41 Flood information of gauged sites for verification

Station Number	Station Name	Length of Years	Drainage Area	mean	Region
02ZM019	Virginia River at Cartwright Place	14	5.55	3.639	A
02ZM021	South Brook at Pearl Town Road	13	9.21	10.809	A
02ZG005	Little Barasway Brook near Molliers	6	28.2	24.03	A
02YN004	Star Brook Above Star Lake	14	276	121.49	B
02YO007	Leech Brook Near Grand Falls	7	88.3	29.81	B
02YP001	Shoal Arm Brook Near Badger Bay	13	63.8	25.68	B
02YQ004	Northwest Gander River Near Gander Lake	10	2200	647.4	B
02YF001	Cat Arm River Above Great Cat Arm	13	611	272.8	C
02YG002	Middle Arm Brook Below Flatwater Pond	10	224	49.77	C
02YK003	Sheffield River At Sheffield Lake	11	362	65.45	C
02YK007	Glide Brook Below Glide Lake	11	112	24.1	C
02YJ003	Pinchgut Brook At Outlet of Pinchgut Lake	11	119	30.04	D
02ZA003	Little Codroy River Near Doyles	14	139	161.9	D

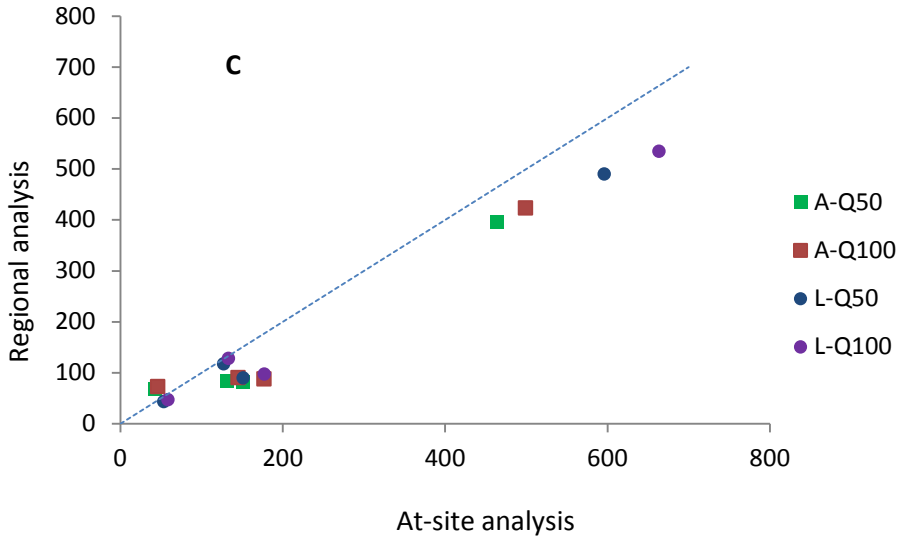
Figures 5.8a-d show the comparison of quantile estimates between at-site and regional quantile estimates for Q50 and Q100 at tested sites in sub regions A, B, C and D based on the index-flood procedure and regression models (AMEC, 2014) in the Island of Newfoundland respectively. “L-Q50” and “L-Q100” mean the quantile estimates for Q50 and Q100 obtained based on the index-flood procedure. “A-Q50” and “A-Q100” mean quantile estimates for Q50 and Q100 obtained from the study conducted by AMEC (2014).



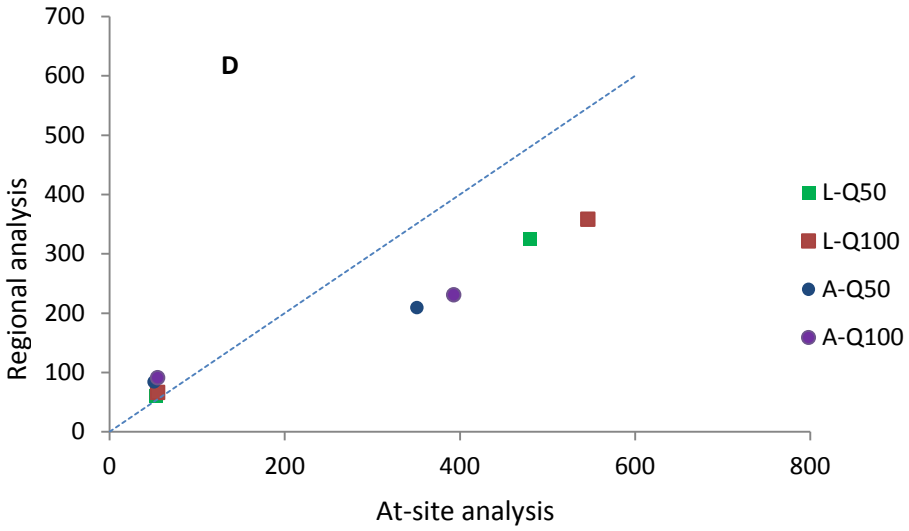
(a)


B
Regional analysis
At-site analysis
0
200
400
600
800
1000
1200
1400
1600
0
500
1000
1500
2000
A-Q50
A-Q100
L-Q50
L-Q100

(b)


C
Regional analysis
At-site analysis
0
100
200
300
400
500
600
700
800
0
200
400
600
800
A-Q50
A-Q100
L-Q50
L-Q100

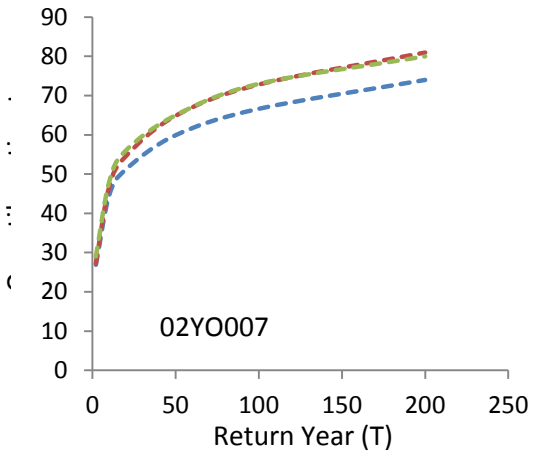
(c)

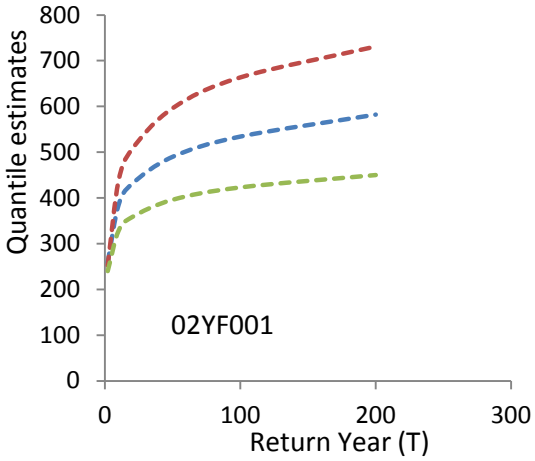


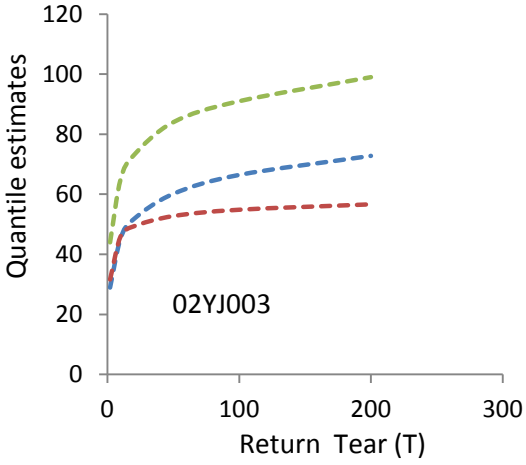
(d)

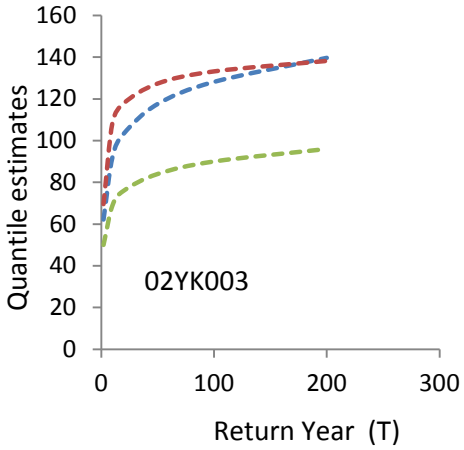
Figure 5.11 Comparison of quantile estimates for Q50 and Q100 between at-site and regional analysis in four sub regions in Newfoundland

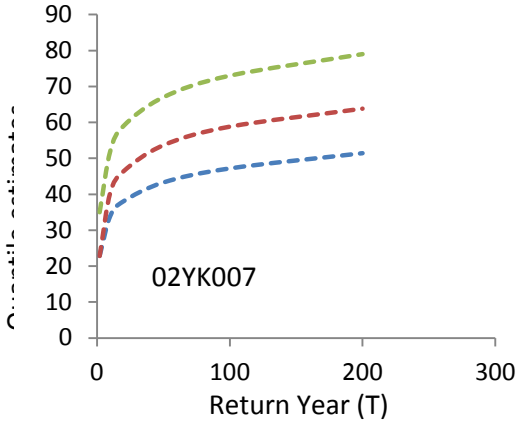
The quantile estimates when the return years are 5, 10, 20, 50, 100 and 200 for regional and at-site analysis based on the index-flood procedure and the regional analysis based on the regression models obtained from AMEC(2014) at each tested sites are plotted in Figures 5.9. The blue lines, red lines and green lines represent the regional and at-site analysis based on the index-flood procedure and regional analysis based on the regression-on-quantile approach respectively.

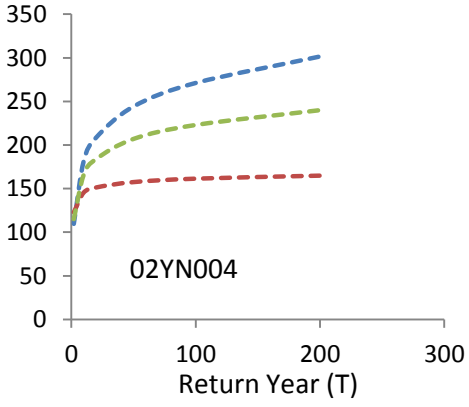

90
80
70
60
50
40
30
20
10
0
0
50
100
150
200
250
02YO007
Return Year (T)

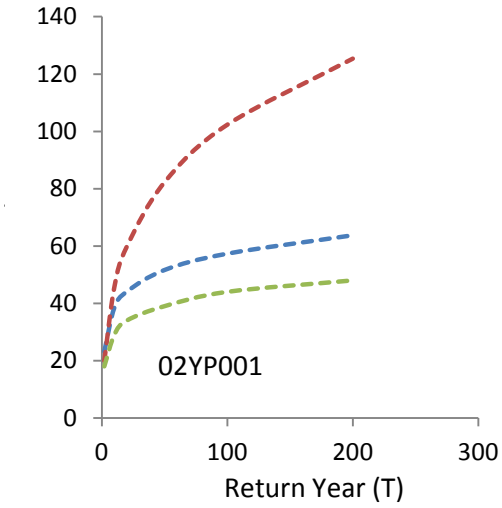

Quantile estimates
800
700
600
500
400
300
200
100
0
0
100
200
300
02YF001
Return Year (T)

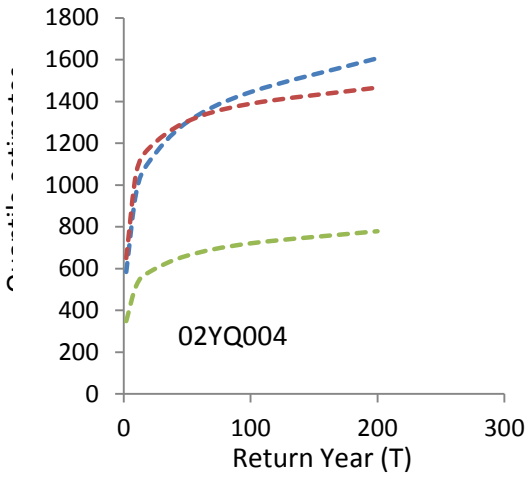

Quantile estimates
120
100
80
60
40
20
0
0
100
200
300
02YJ003
Return Tear (T)

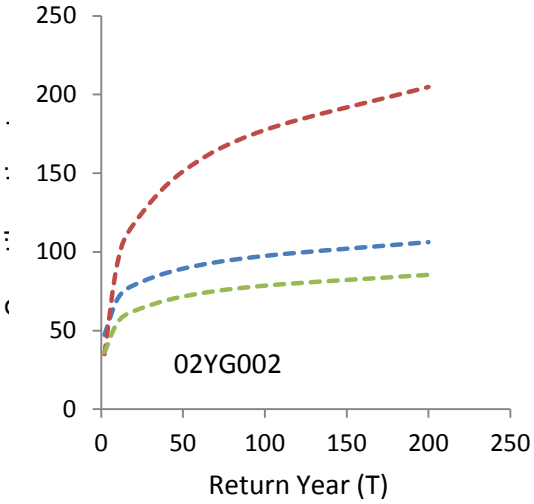

Quantile estimates
160
140
120
100
80
60
40
20
0
0
100
200
300
02YK003
Return Year (T)

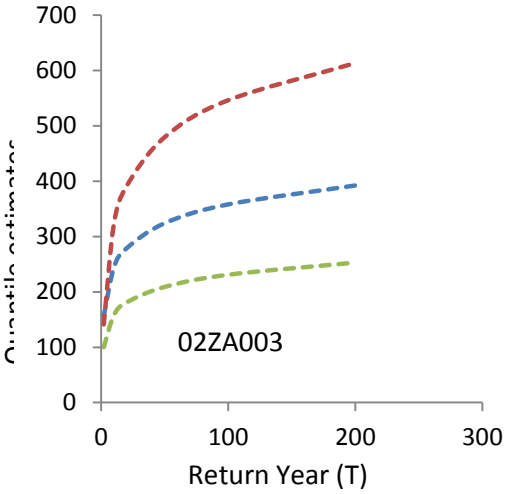

90
80
70
60
50
40
30
20
10
0
0
100
200
300
02YK007
Return Year (T)


350
300
250
200
150
100
50
0
0
100
200
300
02YN004
Return Year (T)


140
120
100
80
60
40
20
0
0
100
200
300
Return Year (T)
02YP001


1800
1600
1400
1200
1000
800
600
400
200
0
0
100
200
300
Return Year (T)
02YQ004


250
200
150
100
50
0
0
50
100
150
200
250
Return Year (T)
02YG002


700
600
500
400
300
200
100
0
0
100
200
300
Return Year (T)
02ZA003

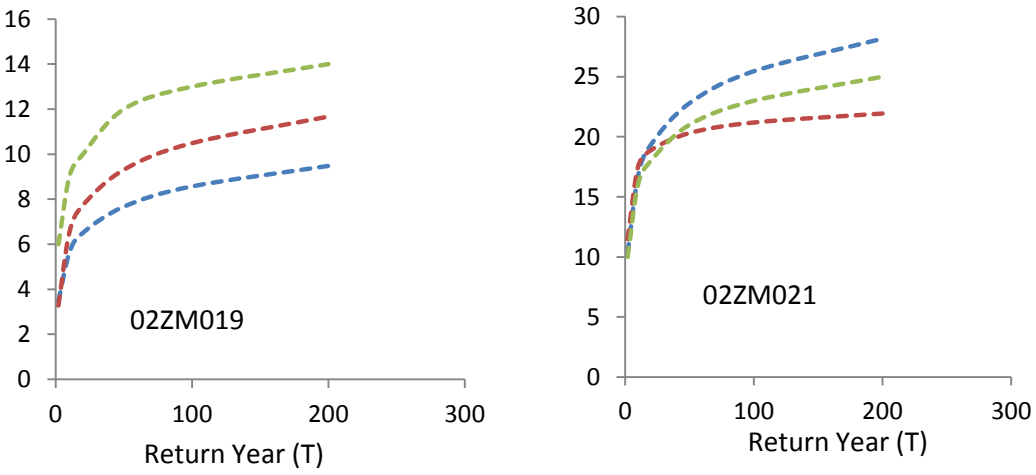


Figure 5.12 Comparison of quantile estimates between the index-flood procedure and regression models for each tested site in Newfoundland

The verification is also carried out in sub regions Y and Z. Fit regional and at-site LN3 parameters to the flood data at tested sites, then calculate the quantile estimates for different return years based on the index-flood procedure. Plot and compare the quantile estimates with the results obtained from regression models developed by AMEC (2014). Figures 5.10-5.11 plot the relationship between at-site and regional quantile estimates based on the index-flood method and the regression models for Q50 and Q100 respectively.

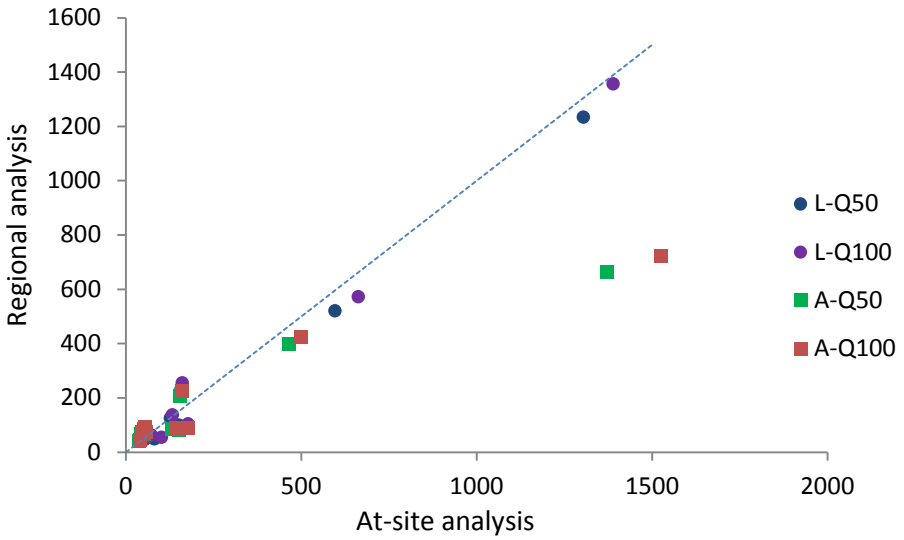


Figure 5.13 Comparison of quantile estimates for Q50 and Q100 between at-site and regional analysis in Region Y

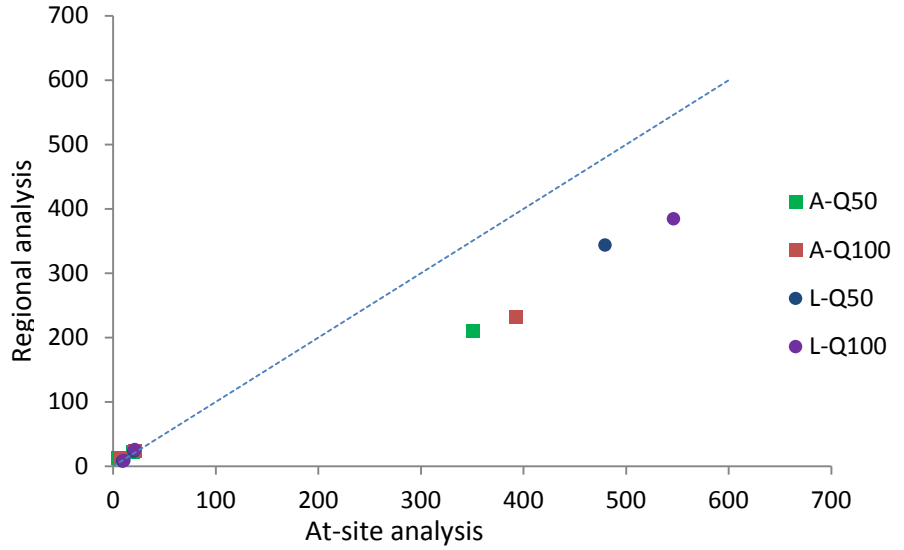
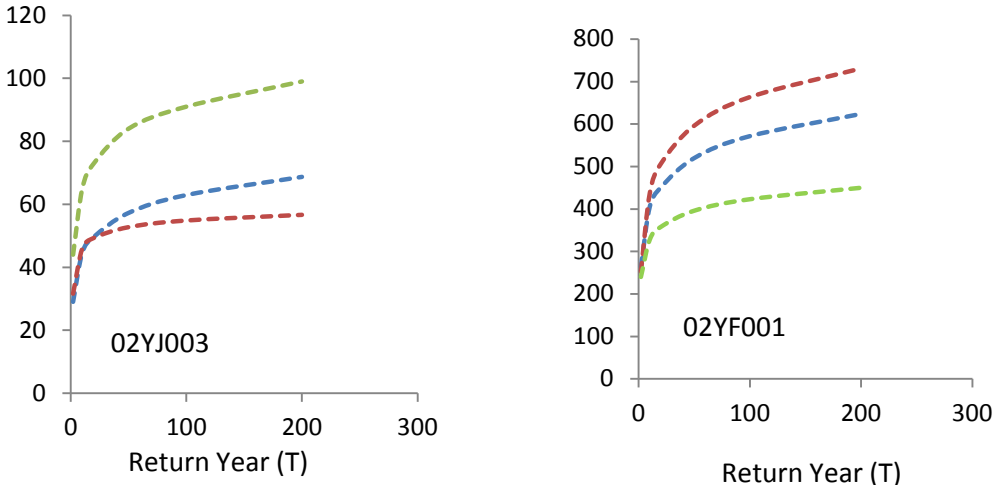
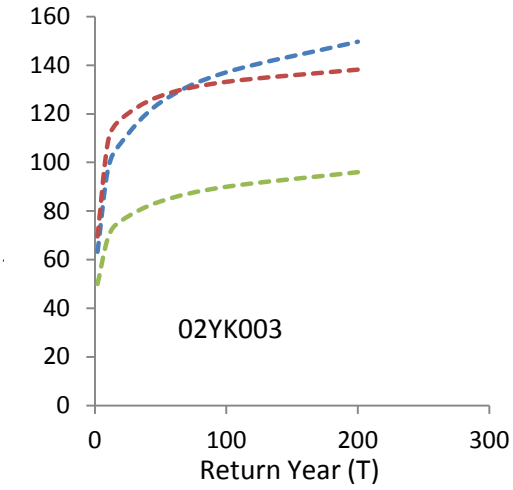
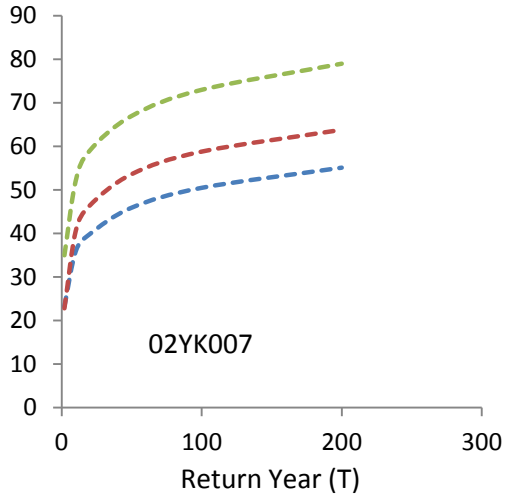


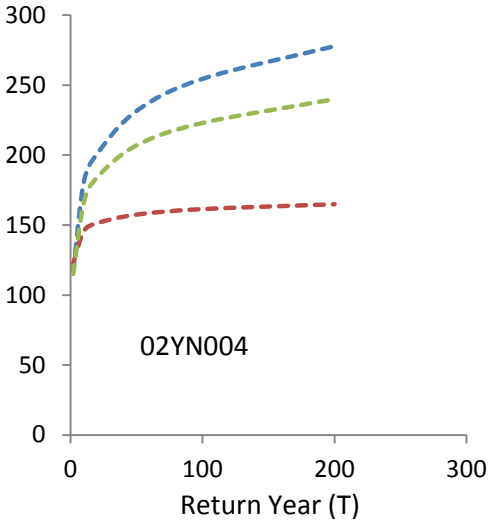
Figure 5.14 Comparison of quantile estimates for Q50 and Q100 between at-site and regional analysis in Region Z

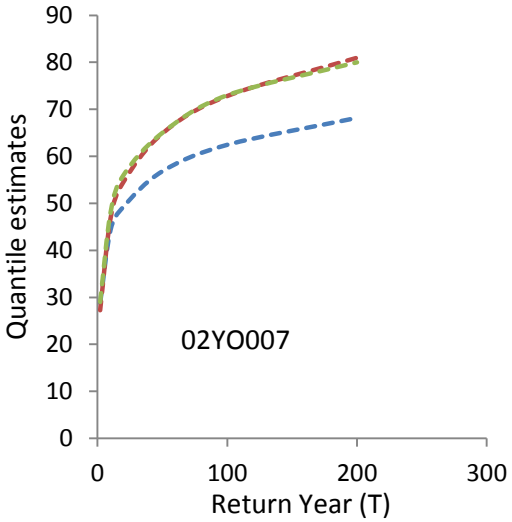
The plots in Figure 5.12 show the relationship between regional and at-site quantile estimates based on the index-flood method and the regional analysis based on the regression equations developed by AMEC (2014) at each single tested site. The blue lines, red lines and green lines represent the regional and at-site analysis based on the index-flood procedure and regional analysis based on the regression-on-quantile approach respectively.

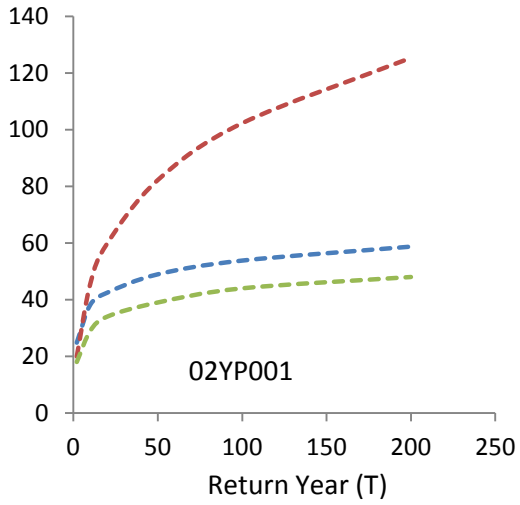


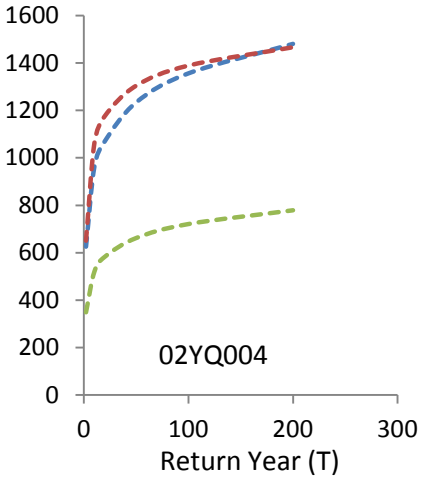

160
140
120
100
80
60
40
20
0
0
100
200
300
Return Year (T)
02YK003

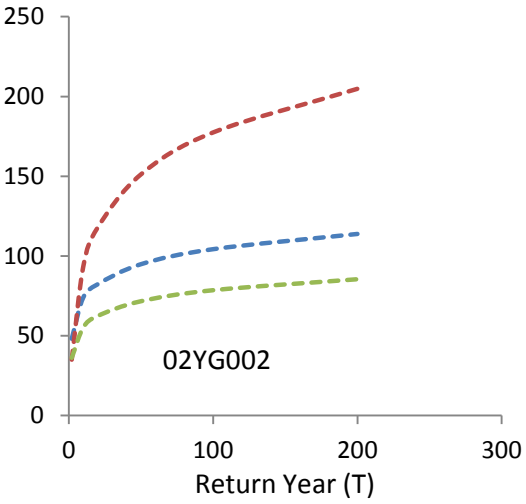

90
80
70
60
50
40
30
20
10
0
0
100
200
300
Return Year (T)
02YK007

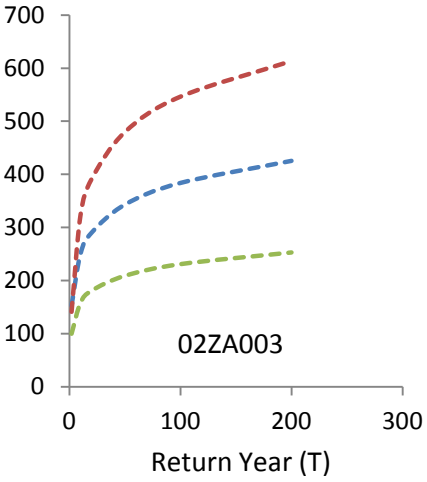

300
250
200
150
100
50
0
0
100
200
300
Return Year (T)
02YN004


Quantile estimates
90
80
70
60
50
40
30
20
10
0
0
100
200
300
Return Year (T)
02YO007


140
120
100
80
60
40
20
0
0
50
100
150
200
250
02YP001
Return Year (T)


1600
1400
1200
1000
800
600
400
200
0
0
100
200
300
02YQ004
Return Year (T)


250
200
150
100
50
0
0
100
200
300
02YG002
Return Year (T)


700
600
500
400
300
200
100
0
0
100
200
300
02ZA003
Return Year (T)

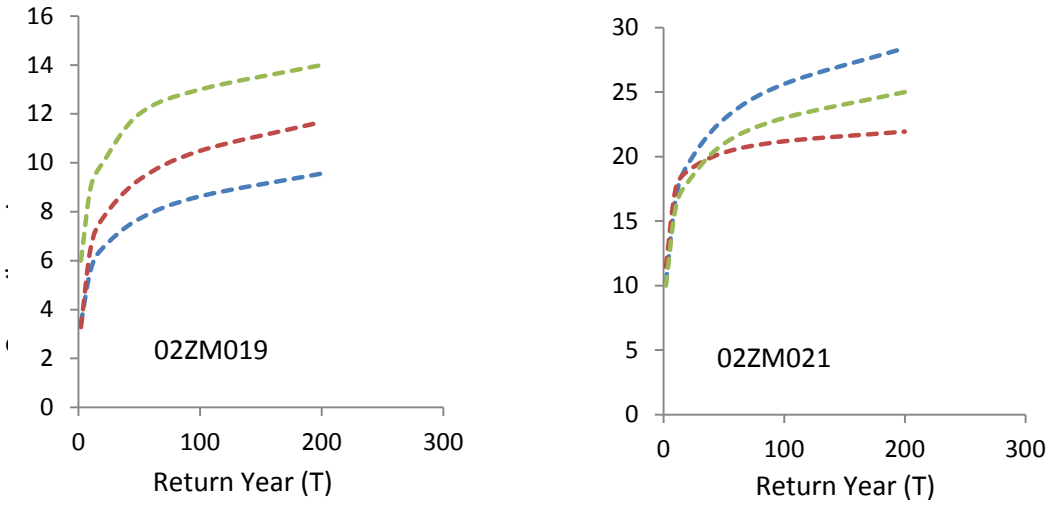


Figure 5.15 Comparison of quantile estimates between the index-flood procedure and regression models for each tested site in Newfoundland

From most of the figures shown above, it can be seen that the method of index-flood can provide more accurate quantile estimates than the method of regression models. And the regional quantile functions obtained from regions Y and Z provide better fit to the observed data than those from regions A, B, C and D.

For the estimation of index flood at ungauged sites, Table 5.41 gives the nonlinear regression equations relating the index flood and sites characteristics.

Table 5.42 Nonlinear regression equations and R^2 for sub regions Y and Z in Newfoundland

Sub Regions	Regression Equations	R^2
Y	$Q=2.76*DA^{0.8351}LAF^{-0.2379}DRD^{0.553}$	0.90
Z	$Q=4.344*DA^{0.8036}LAF^{-0.1469}DRD^{0.526}LSF^{-0.921}$	0.92

5.5 Newfoundland region

AMEC (2014) analyzed the regional quantile analysis when the Newfoundland was treated as a single homogeneous region. However, using L-moments based index-flood procedures the result of heterogeneity measure shows that the Newfoundland is definitely not a homogeneous region. Therefore, the quantile analysis when the Newfoundland is treated as a single region is not carried out based on the L-moments based index-flood procedure.

CHAPTER 6

SUMMARY OF RESULTS

6.1 General

Although the method of regression on quantile is still being used for regional flood frequency analysis in Newfoundland and Labrador, the increase and world wide applications of the index-flood procedure based on L-moments highlight its significant advantages over the regression method in yielding robust quantile estimates. The successful application of the index-flood method for the Island of Newfoundland by Pokhrel (2002) motivated this study to include Labrador. Compared to the traditional regression on quantile method which fit a probability distribution to a single station or a series of stations in a region and develop quantile regression models with site characteristics, the index-flood procedure however focuses on describing a regional growth curve by multiplying the index flood with the regional growth factor---q(F) for all of the available sites in a homogeneous region. To calculate the regional quantile estimates using the index-flood procedure based on the L-moments, the data should first be screened to make sure that there are no outliers and that all of the data are not discordant and will become a homogeneous region, then to delineate the regional growth curve based on the regional frequency distribution.

Pokhrel (2002) analyzed 39 gauged sites in Newfoundland and developed regional

frequency flood quantile functions based on two regionalization scheme using L-moments. The latest regional flood frequency by AMEC (2014) developed regional quantile regression models by fitting the LN3 distribution to the Island of Newfoundland and Labrador, respectively. The goal of this thesis is to develop regional flood frequency models and to obtain the quantile flows based on L-moments index-flood procedure using the latest flood data and to compare the quantile results with those obtained by Pokhrel (2002) and AMEC (2014). In summary, this thesis:

1) Conducted a regional flood frequency analysis based the on L-moments index-flood procedure for Labrador and the Island of Newfoundland, respectively.

2) Compared quantile floods with those from Pokhrel (2002) and AMEC (2014).

3) Developed nonlinear regression equations relating the index flood and sites characteristics for estimating the index flood at ungauged sites for Labrador and Newfoundland.

4) Confirmed and verified the accuracy of the results by testing the quantile estimates at gauged sites with flood records not used to develop the regional models.

6.2 Conclusions

6.2.1 Labrador

1) 10 gauged sites with at least 15 years of record are selected for the quantile estimates. Sites 03OC003 and 03OE010 were found to be discordant from other sites, but the results of heterogeneity test suggested that it was better to keep all of the sites for further study.

2) The results of the goodness-of-fit test confirmed that the three-parameter generalized extreme-value (GEV) is the best fitting regional frequency distribution as a result of its lowest $|Z^{DIST}|$ value.

3) Regional quantile estimates at each gauged sites were obtained using regional GEV quantile function.

4) The estimated index flood at ungauged sites was calculated from a nonlinear regression relationship between the index flood and site characteristics at gauged sites in Labrador.

5) The regional quantile functions were tested and found to agree well with the observed flood data.

6) The index-flood procedure based on the L-moments proved to have a better

performance than the method of regression on quantile in terms of estimating regional flood frequency. The regional quantile function produced more accurate and reasonable quantile estimates than those obtained by AMEC (2014).

6.2.2 Island of Newfoundland

1) Except for site 02ZM009 in sub region A, all of the sites in sub regions A, B, C and D were tested to be not discordant. The results of the heterogeneity test showed that excluding site 02ZM009 in sub region A, all of the regions were found to be homogeneous for quantile estimates.

2) The regional frequency distributions for each sub regions were selected according to the results of goodness-of-fit test, but the best fitted ones were determined based on the results of robustness test. The three-parameter LN3 distribution was better than the GEV distribution in sub regions A, B and C, and it worked well in sub region D which avoided the inconvenience of using the PE3 quantile function.

3) The quantile flows estimated from the index-flood procedure were found to provide better fit to the observed data than the regression on quantile method. Although the quantile estimates obtained from this thesis could not be compared directly with those from Pokhrel (2002) due to the changed flood data, their results

were still similar.

4) The nonlinear regression models relating the index flood and sites characteristics at gauged sites were used to estimate the index flood at ungauged sites in a homogeneous region. The Drainage Area (DA) and the Lake Attenuation Factor (LAF) were used to develop the regression models in sub regions A, B and C. In sub region D, the Drainage Area (DA) and the Lakes and Swamps Factor (LSF) were used. The lower R^2 may be due to its small sample size.

5) The sub regions Y and Z suggested by Water Survey of Canada were found to be homogeneous and results of the goodness-of-fit test and robustness test indicated that the three parameters lognormal (LN3) distribution was the best fitted regional distribution for regions Y and Z.

6) Compared to the regionalization scheme of four sub regions proposed in 1999, regions Y and Z were found to provide better fit to the observed data. And they were tested to be more robust for quantile estimates.

6.3 Recommendations

1) The regionalization scheme of Y and Z regions from Water Survey of Canada (WSC) used for Newfoundland are recommended. Compared to the regionalization

scheme of using four sub regions, the regional quantile functions in regions Y and Z provide better fit to the observed data and the larger sample size per region provided a more robust regional frequency distribution.

2) The use of gauged sites with shorter record that have not been used for model development for testing the accuracy of regional quantile models are recommended. AMEC (2014) estimated the regional quantile estimates based on the quantile estimates at single gauged sites. Verification of the accuracy of estimates using other gauged sites not used in model development was not attempted.

3) Future updates of regional flood frequency analysis by the government should be based on the L-moments index-flood approach as it has a rich statistical basis, used worldwide and has been shown to produce more accurate flood quantile estimates that the outdated regression on quantile approach.

REFERENCES

Abolverdi, J. and Khalili, D. (2010). Development of regional rainfall annual maxima for southwestern Iran by L-moments. Water Resources Management. Vol, 24, pp 2501-2526.

AMEC Environment & Infrastructure (2014). Regional flood frequency analysis for Newfoundland and Labrador 2014 update.

Anderson, T. W. and Darling, D. A. (1954). A test of goodness-of-fit. Journal of the American Statistical Association. Vol 49., pp. 765-769.

Ashkar, F. and Quarda, T.B.M.J. (1996). On some methods of fitting the generalized Pareto distribution. J. Hydrol., 177,117-141.

Atiem, A. and Harmancioglu, N.B. (2006). Assessment of regional floods using L-moments approach: the case of the River Nile. Water Resour Manag. 20:723-747.

Atiem, I.A. and Harmancioglu, N.B. (2006). Assessment of regional floods using L-moments approach: the case of the River Nile. Water Resources Management. 20:723-747.

Basu, B. and Srinivas, V. V. (2014). Regional flood frequency analysis using Kernel-based fuzzy clustering approach. Water Resour. Res. 50: 3295-3316.

Barnett, V. and Lewis, T. (1994). Outliers in statistical data, 3rd ed. Wiley, Chichester, U.K.

Beable, M.E. and McKerchar, A.I. (1982). Regional flood estimation in New Zealand. Water and Soil Technical Publication 20, Ministry of Works and Development, Wellington, N.Z.

Bharath, R. and Srinivas, V.V. (2015). Delineation of homogeneous hydrometeorological regions using wavelet based global fuzzy cluster analysis. International Journal of Climatology.

Bhaskar, N. R., and O'Connor, C. A. (1989). Comparison of method of residuals and

cluster analysis for flood regionalization, J. Water Resour. Plann. Manage., 115(6), 793–808.

Borah, D.K. and Bera, M. (2004). Watershel-scale hydrologic and nonpoint-source pollution models: Review of applications. Trans. ASAE 47(3):789-803.

Boyle, D.P., Gupta, H.V., and Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Res. 36(12): 3663-3674.

Brath, A., Castellarin, A., Franchini, M. and Galeati, G. (2001). Estimating the index flood using indirect methods. Hydrol. Sci. J. 46(3), 399-418.

Burn, D. (1989). Cluster analysis as applied to regional flood frequency. J. Water Resour. Plann. Manage. 115(5), 567-582.

Burn, D.H. (1988). Cluster analysis as applied to regional flood frequency. Journal of Water Resources Planning and Management, 115, 567-82.

Burn, D.H. and Goel, N.K. (2000). The formation of groups for regional flood frequency analysis. Journal of Hydrology Science. 45(1):97-112.

Chow, V.T. (Ed), (1964). Handbook of Hydrology. McGraw Hill, New York.

Cunnane, C. (1978). Unbiased plotting positions—a review. Journal of Hydrology, 37, 205-22.

Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. J Hydrol 100:.269-290.

Epstein, B. (1960a). Tests for the validity of the assumption that the underlying distribution of life is exponential: Part I. Technometrics, 2:83-101.

Epstein, B. (1960b). Tests for the validity of the assumption that the underlying distribution of life is exponential: Part II. Technometrics, 2:167-183.

Dalrymple, T. (1960). Flood frequency analysis, US Geology Survey on Water Supply Paper 1543-A, Reston, VA.

De Coursey, D.G. (1972). Objective regionalization by peak flow rates. Proc. Second Int. Symp, In Hydrology, Fort Collins, Colo, pp.385-405.

Downs, G.M. and Barnard, J.M. (1992). Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures. J. Chem. Inf. Comput. Sci. 1992, 32, 644-649.

Fill, H.D. and Stedinger, J.R. (1995). Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. Journal of Hydrology 166, 81-105.

Fill, H.D.and Stedinger, J.R. (1998). Using regional regression within index flood procedures and an emporocal Bayesian estimator. Journal of Hydrology 210, 128-145.

Flood Risk and Vulnerability Analysis Project (2012). AMEC Environment & Infrastructure a Division of AMEC Americas Limited

Fovell, R. G. and Fovell, M.-Y.C. (1993). Climate zones of the conterminous United States defined using cluster anlaysis. Journal of Climate, 6, 2103-35.

Gabriele, S. and Chiaravalloti, F. (2012). Searching regional rainfall homogeneity using atmospheric fields. Advances in Water Resources. Vol, 53, 163-174.

Gingras, D., Adamowski, K., and Pilon, P.J. (1994). Regional flood equations for the provinces of Ontario and Quebec. Water Resources Bulletin, 30, 55-67.

Griffis, V.W. and Stedinger, J.R. (2007). Evolution of flood frequency analysis with Bulletin 17. Journal of Hydrologic Engineering. pp, 283:297.

Government of Newfoundland and Labrador (1984). Regional flood frequency analysis for the Island of Newfoundland. Newfoundland Flood Damage Reduction Program, Department of Environment and Lands.

Government of Newfoundland and Labrador (1990). Regional flood frequency analysis for the Island of Newfoundland. Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division.

Government of Newfoundland and Labrador (1999). Regional flood frequency analysis for the Island of Newfoundland. Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. Annals of Mathematical Statistics, 21:27-58.

Gupta, H.V., Sorooshian, S. and Yapo, P.O. (1999). Status of automatic calibration for

hydrologic models: Comparison with multilevel expert calibration. J. Hydrologic Eng. 4(2)135-143.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., and Wallis, J.R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research, 15, 1049-54.

Hosking, J.R., 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. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society Series B, 52, 105-24.

Hosking, J.R.M. and Wallis, J.R. (1993). Some useful statistics in regional frequency analysis. Water Resources Research 29(2), 271-281.

Hosking, J.R.M. and Wallis, J.R. (1997). Regional frequency analysis. An approach based on L-moments. Cambridge University Press.

IH(Institute of Hydrology) (1999). Flood Estimation Handbook. IH, Wallingford, UK.

Jingyi, Z. and Hall, M. (2004). Regional flood frequency analysis for the gan-ming river basin in China. Journal of Hydrology. 296: 98-117.

Jin, M. and Stedinger, J.R. (1989). Flood frequency analysis with regional and historical information. Water Resour Res 25(5):925-936.

Kalkstein, L.S; Tan, G. and Skindlov, J.A. (1987) An evaluation of three clustering procedures for use in synoptical climatological classification. J. Clim. Appl. Meteorpl. 26,717-730.

Kjeldsen, T.R. and Jones, D. (2007). Estimation of an index flood using data transfer in the UK. Hydrological Sciences. 52(1):86-98.

Kirby, W. (1974). Flood estimation in the presence of outliers. Proc., Symp. Stat. Hydrol., Misc. Pub. 1275, U.S. Dep. Agric., Washington, D.C., pp.97-119.

Lettenmaier, D.P. and Potter, K.W. (1985). Testing flood frequency estimation methods using a regional flood generation model. Water Resour. Res. 21(12), 1903-1914.

Lewis, T. and Fieller, N. R. J. (1978). A recursive algorithm for null distribution for outliers: I Gamma samples. Technometrics, 21:371-376.

Lim, Y.H. and Lye, L.M. (2003). Regional flood estimation for ungauged basins in Sarawak, Malaysia. Hydrological Sciences Journal 48(1), 79-94.

Lim, Y.H. and Voeller, D.L. (2009). Regional flood estimations in Red River using L-moment-based index-flood and Bulletin 17B Procedures. Journal of Hydrologic Engineering. pp, 1002-1016.

Lu, L.H. (1991). Statistical methods for regional flood frequency investigations. Ph.D. Dissertation, Cornell University, Ithace, NY.

Lu, L.H. and Stedinger, J.R. (1992) Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. Journal of Hydrology 138, 223-245.

Luis-Perez, F. E., Cruz-Barbosa, R., Alvarez-Olguin,G. (2011). Regional flood frequency estimation for the Mexican Mixteca Region by clustering techniques.

Lye, L.M., and Moore, E. (1991). Discussion on "Instantaneous peak flow estimation procedures for Newfoundland streams. Water Resources Bulletin, 27(1): 125-127.

Madsen, H., Pearson, C.P., and Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series index-flood modeling. Water Resour Res. 33(4), 771-782.

Mailhot, A., Lachiance-Cloutier, S., Talbot, G. and Favre, A-C. (2013). Regional estimates of intense based on the Peak-Over-Threshold (POT) approach. Journal of Hydrology 476, 188-199.

Malekinezhad, H., Nachtnebel, H.P. and Klik, A. (2011). Comparing the index-flood and multiple-regression methods using L-moments. Physics and Chemistry of the Earth. 36:54-60.

McCuen, R.H, (1985). Statistical Methods for Engineers. Prentice-Hall, Englewood Cliffs, N.J.

Mkhandi, S. and Kachroo, S. (1997). Regional flood frequency analysis for Southern Afirca Southern African FRIEND. Technical Documents in Hydrology No. 15 UNESCO, Paris, France

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. American Society of Agricultural and Biological Engineers. Vol. 50(3):885-900.

Mosley, M. P. (1981). Delimitation of New Zealand hydrologic regions, J. Hydrol., 49, 173–192.

Motovilov, Y.G., Gottschalk, L., England, K. and Rodhe, A. (1999). Validation of distributed hydrological model against spatial observations. Agric. Forest Meteorology, 98-99:257-277.

Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual mdels: Part 1. A discussin of principles. J. Hydrology 10(3):282-290.

Nathan, R. J. and McMahon, T. A. (1990), Identification of homogeneous regions for the purposes of regionalization, J. Hydrol., 121, 217–238.

Natural Environment Research Council (1975). Flood Studies Report, vol. 1. Natural Environment Research Council, London.

NERC (Natural Environment Research Council) (1975) Flood Studies Report. NERC, London, UK.

Noto, L. V. and Loggia, G. L. (2009). Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. Water Resour Manage 23:2207-2229.

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.

Pearson, C.P. (1991). New Zealand regional flood frequency analysis using L-moments. The New Zealand hydrological society. J. Hydrol 30(2):53-64.

Pearson, C.P. (1991b). Regional flood frequency for small New Zealand basins, 2: Flood frequency groups. Journal of Hydrology (New Zealand), 30, pp.53-64.

Pearson, C.P. (1995). Regional frequency analysis of low flows in New Zealand rivers. The New Zealand hydrological society. J Hydrol 33(2):94-122.

Peel, M.C., Wang, O.J., Vogel, R.M. and McMahon, T.A.(2001). The utility of L-moment ratios diagram for selecting a regional probability distribution. Hydrological Sciences, 46(1). pp, 147-155.

Pitlick, J. (1994). Relation between peak flows, precipitation and physiography for five mountains regions in the Western USA. J Hydrol. 158, 219-240.

Pokhrel, J. and Lye, L.M. (2002). Regional flood frequency analysis for the island of Newfoundland, Canada using L-moments.

Portela, M.M. and Dias, A.T. (2005). Application of the index-flood method to the regionalization of flood peak discharges on the Portugal mainland. In: Brebbia CA. Antunes do Carmo JS(eds) River basin management Ⅲ. WIT. Southampton

Poulin, R.Y. (1971). Flood frequency analysis for Newfoundland Streams, Water Planning and Operations Branch, Department of Environment, Ottawa.

Quarda, T., Ba, K., Diza-Delgado, C., Carsteanu, A., Cholmani, K., Gingras, H., Quentin, E., Trujillo, E. and Bobee, B. (2008). Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. Journal of Hydrology. 348:40-58.

Rao, A. R. and Srinivas, V. V. (2006). Regionalization of watersheds by Hybrid cluster analysis. Journal of Hydrology 318(1-4), 37-56.

Richman, M. B. and Lamb, P. J. (1985). Climatic pattern analysis of three-and seven-day summer rainfall in the central United States: Some methodological considerations and a regionalization. Journal of Climate and Applied Meteorology, 24:1325-42.

Robon, A.J. and Reed, D.W. (1999). Statistical procedures for flood frequency estimation. Flood estimation handbook. Vol. 3, Institute of Hydrology, Wallingford, U.K.

Saf. B. (2009). Regional flood frequency analysis using L-moments for the West Mediterranean Region of Turkey. Water Resour Manage. 23:531-551.

Santhi, C., Arnold, J.G., Williams, J.R., Dugas, W.A., Srinivasam, R. and Hauck, L.M. (2001). Validation of the SWAT model on a large river basin with point and nonpoint sources. J. American Water Resources Assoc. 37(5):1169-1188.

Scholz, F.W. and Stephens, M.A. (1987). K-sample Anderson-Darling tests. Journal of American Statistical Association 82, 918-924.

Shu, C. and Burn, D. H. (2004). Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement, J. Hydrol., 291, 132–149.

Singh, J., Knapp, H.V. and Demissie, M. (2004). Hydrological modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08. Champaign, III.: IIIinois State Water Survey. Available at:www.sws.uiuc.edu/pubdoc/CR/ISWSCR2004-08.pdf.

Srinivas, V. V., Tripathi, S., Rao, A. R., and Govindaraju, R. S. (2008). Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. Journal of Hydrology 348: 148-166.

Sveinsson, O.G.B., Boes, D.C. and Salas, J.D. (2001). Population index flood method for regional frequency analysis. Water Resource Research, Vol. 37. 11, pp.2733-2748.

Tasker, G. D. (1982). Comparing methods of hydrologic regionalization, Water Resour. Bull., 18(6), 965–970.

U.S. Water Resources Council (1981). Guidelines for determining flood flow frequency. Bulletin 17B, Hydrology Committee, Washington, D.C.

Vogel, R.M. and Fennessey, N.M. (1993). L-moments should replace product moment diagrams. Water Resour Res. 29(6):1745-1752.

Van Liew, M.W., Veith, T.L., Bosch, D.D., and Arnold, J.G. (2007). Suitability of SWAT for the conservation effects assessment project: A comparison on USDA-ARS experimental watersheds. J. Hydrologic Eng. 12(2):173-189.

Vogel, R.M. and Wilson, I. (1996). Probability distribution of annual maximum, mean, and minimum streamflows in the United State. ASCE J Hydrol Eng 1(2):69-76.

Wallis, J.R. and Wood, E.F. (1985). Relative accuracy of log Pearson Ⅲ procedures. J Hydraul Div Am Soc Civ Eng 111(7):1043-1056.

Wallis, J.R., Lettenmaier, D.P., and Wood, E.F. (1991). A daily hydroclimatological data set for the continental United States. Water Resources Research, 27, 1657-63.

White, E.L. (1975). Factor analysis of drainage basin properties: classification of flood behavior in terms of basin geomorphology. Water Resour. Bull, 11(4):676-687.

Willmott, C.J. (1984). On the evaluation of model performance in physical geography. In Spatial Statistics and Mdels, 443-460.

Wiltshire, S. E. (1986). Regional flood frequency analysis. II. Multivariate classification of drainage basins in Britain, Hydrol. Sci. J., 31(3), 335–346.

Yue, S. and Wang, C.Y. (2004a). Regional probability distribution type of Canadian annual streamflow by L-moments. J Hydrol 43(1):59-73.

Yue, S and Wang, C.Y. (2004b). Possible regional probability distribution type of Canadian annual stramflow by L-moments. Water Resour Manag. 18:425-438.

Zadeh, S.M. (2012). Low flow frequency study for Newfoundland and Labrador., M. Eng. Thesis, Memorial University of Newfoundland

APPENDICES

A-1

(Translated FORTRAN code provided by Hosking and Wallis, 1997)

Matlab code for Discordancy Measure D_i

```
% File: Di_whole.m
% This macro computes the discordancy measures of the individual
% sites in the group
clear;
a=xlsread ('c:\users\lily\desktop\a.xlsx'); % File
% ratios (t, t3, t4) of the sites in the group
n=input('enter the number of sites in the group:');
ubar=[0;0;0];
for i=1:n
    ubar=ubar+1/n*a(i,1:3)';
end
A=zeros(3);
for i=1:n,
    A=A+(a(i,1:3)'-ubar)*(a(i,1:3)'-ubar)';
end
for i=1:n,
    Di(i)=1/3*n*(a(i,1:3)'-ubar)'*inv(A)*(a(i,1:3)'-ubar);
end
disp('===========');
disp('The Di Statistics follow');
disp('===========');
Di'
```

A-2

(Translated FORTRAN code provided by Hosking and Wallis, 1997)

Matlab code for Heterogenity Measure H

```
clc;
clear all;


v=input(' Enter the weighted sd of sample L-CVs for the region: ');
ns=input(' Enter the number of sites in this region: ');
nrg=input(' Enter the number of regions to be simulated: ');
eps=input(' Enter the location parameter of kappa distribution: ');
alpha=input(' Enter the scale parameter of kappa distribution: ');
k=input(' Enter the shape parameter of kappa distribution: ');
h=input(' Enter the 4th parameter of kappa distribution: ');


%open excel file with number of records at each site within the region in
%it (it should be in the same folder as this M-file
%sheet1 of this excel file contains the values
%
[type, sheets] = xlsfinfo('Sites_records.xlsx');
SitesMatrix = xlsread('Sites_records.xlsx', 'Sheet1');
disp ('simulating...please wait');
disp ('    ');
for k1=1:nrg,
      for k2=1:ns,
            nrec=SitesMatrix(k2);
            y=0;
            for i=1:nrec,
                  y(i)=eps+alpha/k*(1-((1-(rand)^h)/h)^k);
            end
            %mode='descend';
            %y_sort=sort(y,mode);
```

```
            y_sort=sort(y);
            x=y_sort/mean(y);
            x1=0;
            for j=1:nrec,
                  x1(j)=x(j)*(j-1);
            end
            x2=sum(x1)/(nrec*(nrec-1));    %b1
            x3=2*x2-mean(x);                    %l2=2*b1-b0
            x4(k2)=x3/mean(x);                  %l-CV=l2/l1
      end
      for k3=1:ns,
            x5(k3)=x4(k3)*SitesMatrix(k3);
      end
      x6=sum(x5)/sum(SitesMatrix);
      for l=1:ns,
            x7(l)=SitesMatrix(l)*((x4(1)-x6)^2)/sum(SitesMatrix);
      end
      x8(k1)=sqrt(sum(x7));
      k1
end
H=(v-mean(x8))/std(x8);
beep
disp ('Results:');
disp ('==============================');
disp ('      ')
if and(lt(H,1), ge(H,0))
      disp ('The region is homogeneous');
      disp ('    ');
elseif H<0
      disp ('The L-moments are correlated');
      disp ('    ');
elseif and (ge(H,1), lt(H,2))
      disp('The region is possibly heterogeneous');
      disp('    ');
else
      disp('The region is definitely heterogeneous: ');
      disp('    ');
```

```
end

fprintf ('The heterogeneity measure, H=%6.2f\n', H);
fprintf ('The meand of simulated regions is, mean=%6.4f\n', mean(x8));
fprintf( 'The standard deviation of simulated regions is, std=%6.4f\n', std(x8));
```

A-3

(Translated FORTRAN code provided by Hosking and Wallis, 1997)

Matlab code for Goodness-of-fit Test

```
clear all;
clc;

%this program calculates the goodness of fit measure 'z'
%in the first part it computes the bias and standard deviation of the
%sample regional L-Kurtosis.

%In the next part this program computes one part of calculations needed in
%goodness of fit test. (Calculating tau-4 for each candidate distribution)

%the candidate distribution names are as follow:
% GLO=Generalized Logistic Distribution
% GEV=Generalized Exterme Value Distribution
% LN3=Lognormal Distribution
% PE3=Pearson type III Distribtuion
% GPA=Generalized Pareto Distribuion

ns=input(' Enter the number of sites in this region: ');
nrg=input(' Enter the number of regions to be simulated: ');


eps=input(' Enter the location parameter of kappa distribution: ');
alpha=input(' Enter the scale parameter of kappa distribution: ');
k=input(' Enter the shape parameter of kappa distribution: ');
h=input(' Enter the 4th parameter of kappa distribution: ');
%distr=input('Enter the candidate distribution name:', 's');

Tau3=input(' Enter regional average L-Skewness tau3 for this region: ');
t4R=input(' Enter regional average L-Kurtosis for this region: ');
```

```
%open excel file with number of records at each site within the region in
%it (it should be in the same folder as this M-file
%sheet1 of this excel file contains the values
%
[type, sheets] = xlsfinfo('Sites_records.xlsx');
SitesMatrix = xlsread('Sites_records.xlsx', 'Sheet1');
disp ('simulating...please wait');
disp ('    ');
for k1=1:nrg,
      for k2=1:ns,
            nrec=SitesMatrix(k2);
            y=0;
            for i=1:nrec,
                  y(i)=eps+alpha/k*(1-((1-(rand)^h)/h)^k);
            end
            mode='descend';
            y_sort=sort(y,mode);
            x=y_sort/mean(y);
            x1=0;
            x2=0;
            x3=0;
            for j=1:nrec,
                  x1(j)=x(j)*(j-1);
                  x2(j)=x(j)*(j-1)*(j-2);
                  x3(j)=x(j)*(j-1)*(j-2)*(j-3);
            end
            b0=mean(x);
            b1=sum(x1)/(nrec*(nrec-1));
            b2=sum(x2)/(nrec*(nrec-1)*(nrec-2));
            b3=sum(x3)/(nrec*(nrec-1)*(nrec-2)*(nrec-3));

            l1=b0;
            l2=2*b1-b0;
            l3=6*b2-6*b1+b0;
            l4=20*b3-30*b2+12*b1-b0;

            t(k2)=l2/l1;
```

```
            t3(k2)=l3/l2;
            t4(k2)=l4/l2;
        end
         for i=1:k2,
              t4r(i)=SitesMatrix(i)*t4(i)/sum(SitesMatrix);
         end

         T4(k1)=sum(t4r)
end
%calculate the bias of t4R
for k1=1:nrg,
    b4(k1)=(T4(k1)-t4R)/nrg;
    b5(k1)=(T4(k1)-t4R)^2;
end
%bias for t4R
B4=sum(b4);
%standard deviation of t4R
B5=sum(b5);
sigma4=sqrt((B5-nrg*B4^2)/(nrg-1));
beep
disp('=======================================');
fprintf ('The Bias of regional L-Kurtosis, B4= %8.4f\n', B4);
fprintf ('The Standard deviation of regional L-Kurtosis, Sigma4= %8.4f\n', sigma4);

%if distr=='GLO'
      %Tau4distr=0.16667*Tau3^0+0.83333*Tau3^2;
%elseif distr=='GEV'
      %Tau4distr=0.10701*Tau3^0+0.11090*Tau3^1+0.84838*Tau3^2-0.06669*Tau3
^3+0.00567*Tau3^4-0.04208*Tau3^5+0.03763*Tau3^6;
%elseif distr=='LN3'
      %Tau4distr=0.12282*Tau3^0+0.77518*Tau3^2+0.12279*Tau3^4-0.13638*Tau3
^6+0.11368*Tau3^8;
%elseif distr=='PE3'
      %Tau4distr=0.12240*Tau3^0+0.30115*Tau3^2+0.95812*Tau3^4-0.57488*Tau3
^6+0.19383*Tau3^8;
%elseif distr=='GPA'
      %Tau4distr=0.20196*Tau3^1+0.95924*Tau3^2-0.20096*Tau3^3+0.04061*Tau3
```

```
^4;
%else
      %disp('wrong name was entered for candidate distribution');
%end

Tau4distr(1)=0.16667*Tau3^0+0.83333*Tau3^2;
Tau4distr(2)=0.10701*Tau3^0+0.11090*Tau3^1+0.84838*Tau3^2-0.06669*Tau3^3+
0.00567*Tau3^4-0.04208*Tau3^5+0.03763*Tau3^6;
Tau4distr(3)=0.12282*Tau3^0+0.77518*Tau3^2+0.12279*Tau3^4-0.13638*Tau3^6+
0.11368*Tau3^8;
Tau4distr(4)=0.12240*Tau3^0+0.30115*Tau3^2+0.95812*Tau3^4-0.57488*Tau3^6+
0.19383*Tau3^8;
Tau4distr(5)=0.20196*Tau3^1+0.95924*Tau3^2-0.20096*Tau3^3+0.04061*Tau3^4;

%distr(1)='GLO';
%distr(2)='GEV';
%distr(3)='LN3';
%distr(4)='PE3';
%distr(5)='GPA';
distr=['GLO';'GEV';'LN3';'PE3';'GPA'];

for j=1:5,
Zdist(j)=(Tau4distr(j)-t4R+B4)/sigma4;
 fprintf ('The L-Kurtosis of candidate distribution is: %8.6f\n', Tau4distr(j));
 fprintf ('The goodness of fit measure, Zdist of candidate distribution %-5.10s',
distr(j)), fprintf(' is: %8.6f\n', Zdist(j));
 %disp('The goodness of fit measure, Zdist of candidate distribution', distr(j), 'is=',
Zdist(j));
 if abs(Zdist(j))<= 1.64
       disp('The candidate distribution has accepted fit to the data');
 else
       disp('The candidate distribution does not give an adequate fit to the data');
 end
       disp('===============');
end

          beep
```

A-4

(Translated FORTRAN code provided by Hosking and Wallis, 1997)

Matlab code for Robustness Test GEV-LN3

```
% Test for robustness of GEV distribution when the underlying distribution
% is LN3

clear;
AA=xlsread ('c:\users\A\desktop\AA.xlsx'); %contains the sites' record lengths in the
region
sum_nrec=sum(AA);
ns=input('Enter no. of sites in the regions: ');
Nsim=input('Desired no. of simulated regions: ');
disp('AA');
% The parameters of the underlying distributin LN3 follow;
% there are 25 sites in this region
kp=[-0.859853864    -0.539058284    -0.29690173 -0.774778349    -0.083821874
    -0.41765089 -0.572114048    -0.563658075    -0.310502906    -0.504448466
    -0.314536617    -0.293502261    -0.201682941    -0.048388565
    -0.267193256    -0.343581688    -0.955814249    -0.310020389
    -0.46349257 -0.395871537    -0.315050575    0.028419262     -0.275899789
    -0.530709922    -0.499548854
];
alphap=[0.356810964     0.366683934     0.385805653     0.359923246
    0.373261594     0.358354204     0.431733136     0.395900982
    0.357851848     0.20252824  0.343075544     0.30119624  0.317241618
    0.335405635     0.382015357     0.418186582     0.368373875
    0.296272946     0.419679787     0.292813199     0.415747263
    0.282508269     0.296924738     0.324460169     0.381774571
];
epsp=[0.814463553   0.893649372     0.941409107     0.83748436  0.984211374
    0.92183467  0.865631313     0.879034406     0.943015155     0.945631441
    0.944714204     0.954838175     0.967684739     0.991789192
    0.947947692     0.925830772     0.7770411540.952988277      0.897376406
```

```
    0.939737085    0.932809107    1.003971733    0.958198202
    0.907446894    0.898459234
];
indxfld=[66.64  50.01   72.45   40.8    32.76   156.73  16.97   5.951   3.738
    29.52   11.594  13.24   9.694   38.85   10.794  241.6   26.63   13.91
    373.7   365.7   872.3   292.2   42.56   210.2   89.06
];
F=[0.9;0.99;0.999]; % The cumulative probabilities corresponding to 10, 100 and
1000 yr-return periods

xF=zeros(ns,length(F));
% estimate the true quantiles based on the underlying distribution at each
% site
for i=1:ns
    for j=1:length(F)
        xF(i,j)=epsp(i)+alphap(i)/kp(i)*(1-exp(-kp(i)*norminv(F(j))));
        qT(i,j)=1*xF(i,j);
    end
end

% Beginning of the regional simulation based on the underlying distribution

xF_SIM=zeros(ns,length(F));
XF_SIM=zeros(ns,length(F));
bias=zeros(ns,length(F));
Bias=zeros(ns,length(F));
BIAS=zeros(ns,length(F));
BIAS_SIM=zeros(ns,length(F));
relSE=zeros(ns,length(F));
relMSE=zeros(ns,length(F));
RELRMSE=zeros(ns,length(F));
RELMSE=zeros(ns,length(F));
for m=1:Nsim,
    for i=1:ns,
        nrec=AA(i);
        y=0;
        for i1=1:nrec
```

```
        y(i1)=epsp(i)+alphap(i)/kp(i)*(1-exp(-kp(i)*norminv(rand)));
    end
    x=sort(y);
    b0=mean(x);
    indxfld(i)=b0;
    x1=0; x2=0; x3=0; x4=0;
    for j=1:nrec,
        x1(j)=x(j)*(j-1);
        x2(j)=x(j)*(j-1)*(j-2);
        x3(j)=x(j)*(j-1)*(j-2)*(j-3);
        x4(j)=x(j)*(j-1)*(j-2)*(j-3)*(j-4);
    end
    b1=sum(x1)/(nrec*(nrec-1));
    b2=sum(x2)/(nrec*(nrec-1)*(nrec-2));
    b3=sum(x3)/(nrec*(nrec-1)*(nrec-2)*(nrec-3));
    b4=sum(x4)/(nrec*(nrec-1)*(nrec-2)*(nrec-3)*(nrec-4));
    l1(i)=b0;
    l2(i)=2*b1-b0;
    l3(i)=6*b2-6*b1+b0;
    l4(i)=20*b3-30*b2+12*b1-b0;
    l5(i)=70*b4-140*b3+90*b2-20*b1+b0;
    t(i)=l2(i)/l1(i);
    t3(i)=l3(i)/l2(i);
    t4(i)=l4(i)/l2(i);
end

for i=1:ns,
    l1r(i)=AA(i)*l1(i)/sum_nrec;
    l2r(i)=AA(i)*l2(i)/sum_nrec;
    l3r(i)=AA(i)*l3(i)/sum_nrec;
    l4r(i)=AA(i)*l4(i)/sum_nrec;
    tr(i)=AA(i)*t(i)/sum_nrec;
    t3r(i)=AA(i)*t3(i)/sum_nrec;
    t4r(i)=AA(i)*t4(i)/sum_nrec;
end

%regional average L-moments of the simulated series
```

```
    L1=sum(l1r);
    L2=sum(l2r);
    L3=sum(l3r);
    L4=sum(l4r);
    L5=sum(l4r);
    T=sum(tr);
    T3=sum(t3r);
    T4=sum(t4r);

    % regional GEV parameters (distribution under test)
    C=2/(3+T3)-log(2)/log(3);
    K=7.8590*C+2.9554*C^2;
    ALPHA=L2*K/(1-2^-K)*gamma(1+K);
    EPS=L1-ALPHA*(1-gamma(1+K))/K;

    % quantile estimation and computation of accuracy measures
    X_F=zeros(ns,length(F)); Bias=zeros(ns,length(F));

    for i=1:ns
        XF=zeros(ns,length(F)); bias=zeros(ns,length(F)); relSE=zeros(ns,length(F));
        for j=1:length(F)
            XF(i,j)=EPS+ALPHA/K*(1-(-log(F(j)))^K);
            QT=indxfld(i)*XF(i,j);
            bias(i,j)=(QT-qT(i,j))/qT(i,j)*100;
            relSE(i,j)=((QT-qT(i,j))/qT(i,j)).^2;
        end

        X_F=X_F+XF;
        Bias=Bias+bias;
        relMSE=relMSE+relSE;
    end
    XF_SIM=XF_SIM+1/Nsim*X_F;
    BIAS_SIM=BIAS_SIM+1/Nsim*Bias;
    RELMSE_SIM=RELMSE+1/Nsim*relMSE;
end
disp('=======================');
AV_BIAS=mean(BIAS_SIM); AAV_BIAS=mean(abs(BIAS_SIM));
```

```
AV_RELRMSE=(mean(RELMSE_SIM)).^1/2*100;
AV_XF=mean(XF_SIM);

disp('robustness of LN3 when GEV is the parent');
disp('non-exceedence prob, absolute bias and RMSE follow in the columns in the
order as shown:');
disp(' ');
disp('============');
[F';AV_BIAS;AAV_BIAS;AV_RELRMSE]
```

A-5

(Translated FORTRAN code provided by Hosking and Wallis, 1997)

Matlab code for GEV Growth Curve

```
% test for computing the GEV growth curves for Labrador region
clear;
Labrador=xlsread ('c:\users\Lily\desktop\Labrador.xlsx');
sum_nrec=sum(Labrador);
ns=10;%input('enter no. of sites in the region: ');
Nsim=1000; %input(Desired no. of simulated regions ：‘）;
disp('region Labrador');

%GEV parameters
kp=[-0.3346 -0.1349 0.0404 -0.2851 0.2124 -0.0497 -0.1572 -0.1515 0.0300 0.0269];
alphap=[0.2601 0.2917 0.3379 0.2680 0.3622 0.3035 0.3370 0.3143 0.3122 0.2951];
epsp=[0.7231 0.7871 0.8179 0.7415 0.8549 0.8092 0.7438 0.7636 0.8287 0.8374];

F=[0.01:.01:0.99 .991:.001:.999];
XF_sample=zeros(1,length(F));
k=-0.0312;alpha=0.3121;eps=0.8099; %Labrador region

for j=1:length(F)
XF_sample(j)=eps+alpha/k*(1-(-log(F(j)^k)));
end

u_L=zeros(length(F),1); %lower 95% conf. interval
u_U=zeros(length(F),1);
XF=zeros(Nsim, length(F));

%Beginning of the regional simulation

for m=1:Nsim,
for i=1:ns,
nrec=Labrador(i);
```

```
y=0;
for i1=1:nrec
y(i1)=epsp(i)+alphap(i)/kp(i)*(1-(-log(rand))^kp(i));
end
x=sort(y);
b0=mean(x);
indexfld(i)=b0;
x1=0;
x2=0;
x3=0;
x4=0;
for j=1:nrec,
x1(j)=x(j)*(j-1);
x2(j)=x(j)*(j-1)*(j-2);
x3(j)=x(j)*(j-1)*(j-2)*(j-3);
x4(j)=x(j)*(j-1)*(j-2)*(j-3)*(j-4);
end
b1=sum(x1)/(nrec*(nrec-1));
b2=sum(x2)/(nrec*(nrec-1)*(nrec-2));
b3=sum(x3)/(nrec*(nrec-1)*(nrec-2)*(nrec-3));
b4=sum(x4)/(nrec*(nrec-1)*(nrec-2)*(nrec-3)*(nrec-4));
l1(i)=b0;
l2(i)=2*b1-b0;
l3(i)=6*b2-6*b1+b0;
l4(i)=20*b3-30*b2+12*b1-b0;
l5(i)=70*b4-140*b3+90*b2-20*b1+b0;
t(i)=l2(i)/l1(i);
t3(i)=l3(i)/l2(i);
t4(i)=l4(i)/l2(i);
end

for i=1:ns,
l1r(i)=Labrador(i)*l1(i)/sum_nrec;
l2r(i)=Labrador(i)*l2(i)/sum_nrec;
l3r(i)=Labrador(i)*l3(i)/sum_nrec;
l4r(i)=Labrador(i)*l4(i)/sum_nrec;
```

```
tr(i)=Labrador(i)*t(i)/sum_nrec;
t3r(i)=Labrador(i)*t3(i)/sum_nrec;
t4r(i)=Labrador(i)*t4(i)/sum_nrec;
end
%regional average L-moments of the simulated seriers
L1=sum(l1r);
L2=sum(l2r);
L3=sum(l3r);
L4=sum(l4r);
L5=sum(l4r);
T=sum(tr);
T3=sum(t3r);
T4=sum(t4r);

%regional GEV parameters and growth curve
c=2/(3+T3)-log(2)/log(3);
K=7.859*c+2.9554*c*c;
ALPHA=L2*K/((1-2^-K)*exp(gamma(1+K)));
EPS=L1-ALPHA/K*(1-exp(gamma(1+K)));


%Quantile estimation
for j=1:length(F)
XF(m,j)=EPS+ALPHA/K*(1-(-log(F(j)^K)));
end
end
disp('====================');

%plot of regional growth curve
for i=1:length(F)
gum_var(i)=-log(-log(F(i)));%Gumbel reduced variate for plotting growth curves
end

% 95% confidence interval computation and plotting of regional growth curve

XF=sort(XF);
index_L=round(0.05*Nsim);
```

```
index_U=round(0.95*Nsim);
for j=1:length(F)
      u_L(j)=XF(index_L,j);
      u_U(j)=XF(index_U,j);
end
disp('Gumbel_Var Sample_growth curve for region Labrador:');
disp(' Growth Factor      Lower_5%    Upper_5%')
table=[gum_var(1) XF_sample(1) u_L(1) u_U(1);
    gum_var(41) XF_sample(41) u_L(41) u_U(41);
    gum_var(46) XF_sample(46) u_L(46) u_U(46);
    gum_var(49) XF_sample(49) u_L(49) u_U(49);
    gum_var(50) XF_sample(50) u_L(50) u_U(50);
    gum_var(55) XF_sample(55) u_L(55) u_U(55)]

disp('====================');
plot(gum_var,XF_sample,'b');
hold on;
plot(gum_var,u_L,'b:');
plot(gum_var,u_U,'b:');
hold off;
xlabel('Gumbel Reduced Variate, -log(-log(F))');
ylabel('Growth factor');
Title('90% Confidence bands for sample GEV growth curve');
```