Assessment of Regional Flood Frequency Analysis Frameworks for Canada

by

Tianshuo Zhou

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering - (Water)

Waterloo, Ontario, Canada, 2019

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Floods are a devastating natural disaster causing economic loss and even threatening human lives. A better understanding of floods can lead to improved design of flood mitigation measures. The Natural Sciences and Engineering Research Council of Canada (NSERC) funded the FloodNet Strategic Network to identify ways to improve the resilience of Canadian society to flooding. This thesis is a part of the NSERC FloodNet project on the development of a flood estimation manual for Canada. The objective of this thesis is to assess and improve regional flood frequency analysis (RFFA) frameworks developed in the NSERC FloodNet project.

The annual maximum (AMAX) RFFA framework is compared with at-site flood frequency analysis (FFA) for 1114 Water Survey of Canada hydrometric stations across Canada. Different flood-related attributes were explored to find quantitative measures to identify stations for which the AMAX RFFA performed poorer than at-site FFA. The results show the AMAX RFFA framework performs better than at-site FFA for 81.69% of sites across Canada. The average basin slope, lake effect, and L-moment ratios are effective quantitative measures to identify sites where at-site FFA performs better than AMAX RFFA.

The peaks-over-threshold (POT) RFFA framework is compared with the AMAX RFFA framework for 89 sites having more than 50-years of record across Canada. For 78 of 89 long record sites, the POT RFFA performs better than the AMAX RFFA. For sites where the peaks per year of the POT flow series is greater than 1.5, the POT RFFA performs better than the AMAX RFFA.

The Rocky Mountain region was identified as a problematic region where both RFFA frameworks did not perform well. An alternative RFFA framework, the POT-Rocky RFFA, was proposed for the Rocky Mountain region. The POT-Rocky framework performs better or at the same level as the AMAX and POT RFFA frameworks for most of the long record sites within the Rocky Mountain Region.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Donald H. Burn for his continual guidance, teaching, financial support and encouragement for pursuing the Master Program at the University of Waterloo. Special thanks go to Dr. Shabnam Mostofi Zadeh, Department of Water Resources, Government of Newfoundland and Labrador for providing the data for this research and sharing her ideas.

I am thankful to NSERC FloodNet for funding my research and giving me this opportunity to work with people with different backgrounds.

I also acknowledge the group of individuals in Dr. Burn's research group. I learned a lot from every one one of you.

I would like to thank Angie Fossland for editing this thesis.

Finally, I need to thank my family, especially my parents Min Zhou and Qing Tian. My work could not be made without their support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The frequency and intensity of extreme weather events are a paramount concern when designing infrastructure and ensuring the safety of human. From 1998 to 2017, floods ranked as the most frequent natural disaster in the world, affecting more than 2 billion people (UN, 2017). The Insurance Bureau of Canada has identified a trend of increasing payouts for property and casualty insurance across Canada. The average payout rose from \$405 million per year between 1983 and 2008 to \$1.8 billion between 2009 and 2017. Water damage is the key driver for this trend (IBCBAC, 2018). In the last two decades, multiple large floods hit major cities in Canada, causing severe damage. In 1997 "the flood of the century" hit Winnipeg, causing \$1.2 billion in losses. In 2011, southern Manitoba was again struck by severe flood, leading to \$1 billion in losses. In June 2013, excess precipitation caused unprecedented flooding in the Bow and Elbow rivers, with an estimated cost of \$6 billion (Eccles et al., 2017). Thus, there is an increasing need to improve precision in flood estimation for Canada.

Although floods are an inevitable natural hazard, proper mitigation can reduce their impact on people and society. A solid understanding of flood frequency is essential for design and risk assessment of flood mitigation measures. Flood quantile estimation is the most critical step for many flood mitigation measures (Zaman et al., 2012). Flood frequency analysis (FFA) has been widely used and proven to be an effective tool to obtain design flood quantiles in infrastructure design. Flood frequency analysis is a statistical tool that estimates flood quantiles by using statistical distributions fitted to historical records. At-site FFA is the most direct method of estimating floods; however, a long flood record is required. For instance, to estimate a 100-year flood, which is the common design criteria for highway bridges, the recommended record length from the hydrometric station is 200 years (FEH 1999). However, in Canada, only 0.8% of Water Survey of Canada hydrometric stations have a record length greater than 100 years. A lack of long-term records reduces the reliability of at-site flood frequency analysis. Instead, regional flood frequency analysis (RFFA) is a powerful method for short-record sites. The RFFA method takes advantage of information from statistically similar sites. In other words, the RFFA method trades space for time (Hosking & Wallis, 1997).

Many countries have conducted nation wide research on improving and standardizing flood frequency analysis. Some have standardized their flood estimation manual, such as Guidelines for Determining Flood Flow Frequency Bulletin 17C in the United States (IACWD, 2018), Flood Estimation Handbook Volume 3 (FEH, 2008) in the UK and Peak

Discharge Estimation Book 3 in Australia (ARR Australia, 2015). All three handbooks use RFFA analysis as a method to obtain flood quantiles. In Canada, there is no national flood estimation guideline. Each province uses its preferred procedure with little cross-validation between provinces since flood management is traditionally conducted at the provincial level. To address these issues, the Natural Science and Engineering Council of Canada (NSERC) funded the FloodNet Strategic Network. FloodNet is a nation-wide multi-disciplinary research network focused on improving the resilience of Canadian society to floods. The network consists of 13 Canadian universities, multiple federal and provincial agencies and industrial partners (FloodNet, 2019). The research presented in this thesis is under FloodNet theme 1 Flood regimes in Canada: Learning from the past and preparing for the future. The product of theme 1 will be software designed for Canadian hydrological conditions and watersheds and a flood estimation manual for Canada. These products will be valuable assets for industrial partners and government organizations in managing and mitigating the impact of floods.

1.2 Study Background

This research assesses the performance of the currently developed RFFA frameworks for Canada from NSERC FloodNet project 1-1. Then, this study explores sites where the current RFFA frameworks perform worse than the at-site flood frequency analysis. Ultimately, this research offers recommendations to improve the RFFA frameworks for Canada.

Two RFFA frameworks for Canada were developed: (1) the annual maximum (AMAX) RFFA (Mostofi Zadeh & Burn, 2019) and (2) the peaks-over-threshold (POT) RFFA

(Mostofi Zadeh et al., 2019). The major difference between the AMAX RFFA framework and the POT RFFA framework is the input data. The AMAX RFFA estimates the flood quantile by using the annual maximum flood from the gauging record. The POT RFFA uses flood time series constructed by choosing floods above a certain site-specific threshold value (Mostofi Zadeh & Burn, 2019; Mostofi Zadeh et al., 2019). The key ideas for both frameworks are similar: to identify and group statistically similar hydrometric stations into a pooling group or region. If all the hydrometric stations within the pooling group are statistically similar, that pooling group is homogeneous. Then, a statistical distribution is selected for the pooling group through a goodness-of-fit test. Last, the flood quantile is estimated through the selected distribution.

In this research, the AMAX RFFA framework was compared with the most widely used at-site FFA model for 1114 Canadian catchments. The width of the confidence interval for the flood quantile was used as the performance measure for identifying the superior framework. Any site for which the at-site FFA performs better than the AMAX RFFA framework is named a problematic site. Different factors were explored to find effective quantitative measures to identify problematic sites. The first factor is the flood-related attributes of average basin slope and the lake attenuation effect. The second factor is the effect of the outlier site within the pooling group and heterogeneous pooling group, and the third is the effect of L-moment ratios for the flow series.

Finally, an alternative RFFA framework is proposed for the Rocky Mountain region (POT-Rocky RFFA framework). The previous study identified the Rocky Mountain region as a problematic region for both RFFA frameworks for Canada. The performance of the POT-Rocky RFFA framework was compared with both the AMAX RFFA framework and the POT RFFA framework for long-record hydrometric stations within the Rocky Mountain region. This alternative framework improves the currently developed frameworks for the Rocky Mountain region.

1.3 Research Objectives and Scope

The overarching goal for this research is to assess and improve RFFA frameworks for Canada. Three objectives were defined to accomplish this goal:

- Assess currently developed RFFA frameworks to find quantitative measures to identify problematic sites
- Compare the performance of the AMAX RFFA framework and the POT RFFA framework for long-record sites and identify ways to select between the two frameworks
- Formulate recommendations to improve currently developed RFFA frameworks for Canada

1.4 Organization of the Thesis

This thesis is organized into five chapters.

Chapter 1 gives an overview of the topic, including research motivation, a broad introduction to the study background, and specific research objectives.

Chapter 2 is the literature review relevant to the thesis objectives. This chapter discusses the existing research for two flood frequency analysis techniques, at-site FFA and RFFA, and introduces two RFFA frameworks for Canada; the AMAX RFFA and the POT RFFA.

Chapter 3 describes the methodology used in this research, including factors that can be used to identify problematic sites. The chapter discusses the comparison between the two developed RFFA frameworks for Canada and an alternative RFFA framework for the Rocky Mountain region.

Chapter 4 presents the results of the research. This chapter lists quantitative measures for selection among frameworks and the performance of the alternative RFFA framework for the Rocky Mountain region.

Chapter 5 presents the summary and conclusions of this thesis and recommendations for the flood estimation manual for Canada.

Chapter 2

Literature Review

This chapter introduces different flood frequency analysis techniques that are related to this thesis. The flood frequency techniques introduced in this chapter include: 1) The most widely used at-site flood frequency analysis with annual maximum flow series (at-site FFA); 2) Two techniques extending the information for flow series, paleoflood hydrology, which extends the record period by using indicators left by major floods that occurred prior to the systematic gauging period, and the peaks-over-threshold (POT) method, which extracts multiple flood events from one year; and 3) A general framework for regional flood frequency analysis (RFFA), and 2 RFFA frameworks designed specifically for Canadian catchments, the annual maximum regional flood frequency analysis (AMAX RFFA) framework for Canada and the peaks-over-threshold regional flood frequency analysis (POT RFFA) framework for Canada.

2.1 Flood Frequency Analysis

Flood magnitudes for a design flood can be obtained through physically based hydrologic models or a statistically-based model. Since flood generation mechanisms are complex and drainage basin specific, the uncertainty associated with physically-based models is magnified by the number of processes in the model. Therefore the statistical model, flood frequency analysis, is widely used in industry. Hosking and Wallis (1997) stated "Frequency analysis is the estimation of how often a specified event will occur." In hydrology applications, this method is known as flood frequency analysis. Flood frequency analysis uses the fitted distribution from historical data to estimate the magnitude and occurrence of floods in the future, relating return periods to an associated flood by magnitude using

$$P_r(Q > Q_T) = \frac{1}{T} \tag{2.1}$$

$$F(Q_T) = 1 - \frac{1}{T} \tag{2.2}$$

where T is the return period in years, $\mathrm{P_r}$ is the exceedance probability for a given discharge $\mathrm{Q_T}$, and $\mathrm{F(Q_T)}$ is the cumulative distribution function (CDF) for the fitted statistical distribution. Equations (2.1) and (2.2) are the two governing equations for flood frequency analysis (FEH, 1999). The magnitude of design flood ($\mathrm{Q_T}$) is often the objective for flood frequency analysis. Two types of flood frequency analysis are used in the field, at-site flood frequency analysis and regional flood frequency analysis. Both methods are used in this research and explained in the following subsections.

2.1.1 At Site Flood Frequency Analysis

At-site flood frequency analysis uses data from a single hydrometric station. This approach is well established and easy to apply, therefore it is widely accepted in industry. In at-site flood frequency analysis, the historical flood record is used to calculate parameters for a statistical distribution. This process is also known as fitting a distribution to the flood data. The statistical distribution is selected based on either goodness of fit test or recommended by regulators for a given jurisdiction. Various goodness of fit techniques have been recommended in different flood frequency analysis manuals, such as the plotting position method recommended in the United States (England, 2019) and Bayesian calibration and L-moments used in Australia (Micevski & Kuczera, 2009). The fitted distribution is used to extrapolate flood quantiles for the desired return period. Flood frequency analysis is a statistical tool and the flood record for a hydrometric station is considered to be a sample from the true flood population. Therefore the larger the sample is the more reliable the estimated quantiles are. To obtain reliable estimates of flood quantiles, a long flood record at the target site is required. A general guideline for the required record length is that for estimating a flood with a return period of T years, the record length at the target site should be between T and 2T years (FEH, 2008). More details for at-site frequency analysis can be found in England (2019), FEH (2008), Geoscience Australia et al. (2016) and Cunnane (1989).

In reality, however, the flood record from a single site is usually shorter than the return period required for designing infrastructure. For example, design criteria for highway bridges in Ontario is 50-year flood or 100-year flood for freeway bridges (MTO, 1979) and 100-year flood or 200-year flood for bridges in British Columbia (BC MoIT, 2016). However, as stated in the last chapter, only 0.8 % of Water Survey of Canada hydrometric stations have flood a record over 100 years. In this situation, two approaches can be used to improve the accuracy of flood frequency analysis, flood record extension and regional flood frequency analysis. Both methods will be discussed in the following sections.

2.1.2 Flood Record Extension

Flood record extension for a single site can be achieved by use of a partial duration series or by paleoflood hydrology. The method using partial duration series is known as peaks-over-threshold (POT). Unlike the annual maximum method that only uses the highest annual discharge event, the POT model can include multiple flood events from the same year, thus increasing the available information. Paleoflood hydrology uses evidence of high flood stages from prior to the systematic gauging record in the process of estimating flood frequencies.

2.1.2.1 Paleoflood Hydrology

Paleoflood hydrology studies floods before human record, by geologic and physical evidence left on the flood plain (Baker, 1986). Paleoflood data have two types, paleostage indicators, and nonexceedance bound information. Paleostage indicators are used to determine the magnitude and nonexceedance bound is for determining the return period (Benito & Thorndycraft, 2005). Most common paleostage indicators are shown in Figure 2.1.

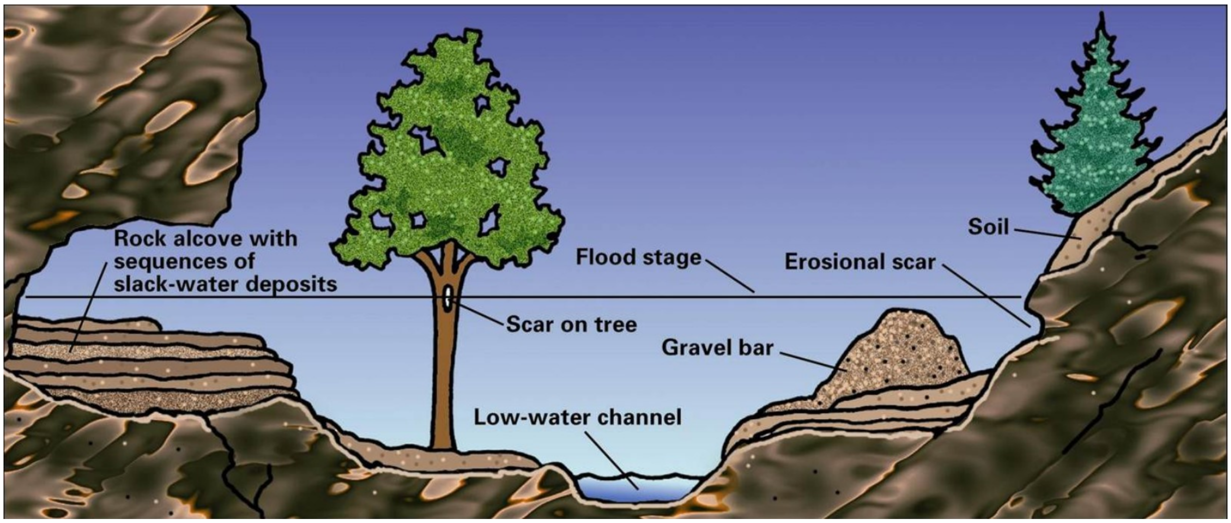

Figure 2.1: Diagram of a section showing typical paleoflood features used as paleostage indicators (from Jarrett & England, 2002).

The United States Geological Survey (2019) recommends "Paleoflood information should be obtained and documented whenever possible, particularly where the systematic record is relatively short and (or) the annual exceedance probability of interest is less than 0.01." Paleoflood information is used to construct the flow interval for pre-record years and is used in fitting a distribution to the flood data. The Expected Moments Algorithm (EMA) combines information from the record period, historical flood information and paleoflood and botanical information to estimate parameters for the distribution. In this approach floods from the historical period and years before present (paleoflood) are represented by flow interval ($Q_{Y,\text{lower}}$, $Q_{Y,\text{upper}}$); where $Q_{Y,\text{lower}}$ and $Q_{Y,\text{upper}}$ are based on observations, written records and physical evidence. By adding information from paleofloods, a better fit of the probability curve can be achieved, especially for high return period floods. Although information from paleofloods is valuable for large floods, it requires extensive research and geographic survey on major catchments. However, this is difficult to achieve in Canada under current circumstances. Therefore the POT model, which extracts more information from the gauging record, is more applicable for Canada.

2.1.2.2 Peaks-Over-Threshold Model

The POT method has been used in a number of studies (Langbein, 1949; Cunnane, 1973 & Lang et al., 1999). The POT model identifies events above a prescribed threshold value to create flood sequences for frequency analysis. The two most important aspects of the POT model are: 1) Identification of an appropriate flood threshold; and 2) Identification of independent exceedances of the threshold that do not correspond to the same event (Madsen et al., 1997b). Both aspects are illustrated in Figure 2.2 where S is the identified threshold level and θ should be long enough to prevent overlapping of hydrographs for floods X_1 and X_2. In practice, however, the POT model is not widely used, since the selection of the threshold level is difficult to automate and hard to standardize (Mostofi Zadeh et al, 2019).

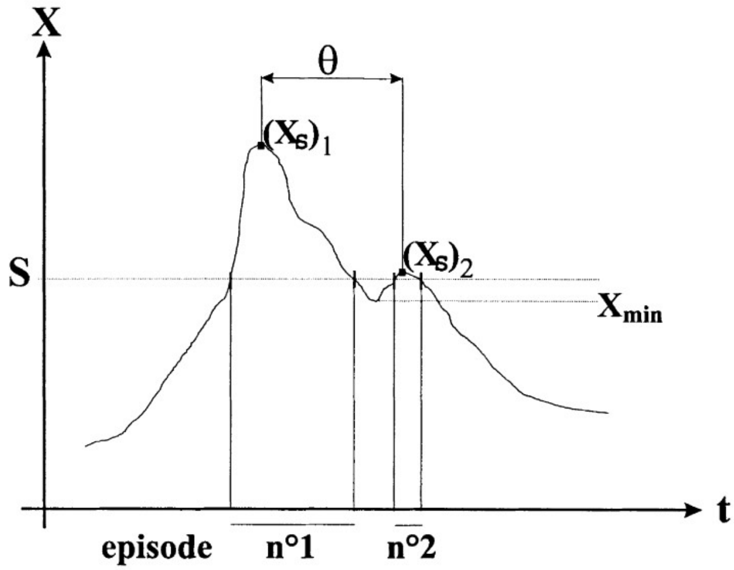

Figure 2.2: Inter-flood duration criteria. (from Lang et al, 1999).

Several methods have been proposed in the literature to verify the hypothesis of independent flood peaks. The Water Resource Council (USWRC, 1976) stated that successive floods should be separated by a number of days equal to five plus the natural logarithm of the basin area in square miles. This is in addition to the requirement that the intermediate flows between two consecutive peaks must drop below 75% of the lowest of these two flood events. Cunnane (1979) suggested two other criteria: 1) The gap between two selected peaks should be greater than $3T_p$, where T_p is the average time to peak from the first five clean hydrographs; 2) The intermediate flows between two consecutive peaks must drop below 66.7% of the earlier of these two flood events. Miquel (1984) proposed to test the lag 1 and lag 2 autocorrelation coefficients of peak values. If the autocorrelation coefficients are below a given significance level, the selected peaks are considered to be independent.

Lang et al. (1999) summarized several approaches and tests for appropriate threshold selection. Many researchers proposed to use a fixed number of peaks-per-year as the basis for determining the threshold level, where the number of peaks-per-year is specific to a climatic or geographic region (Taesombut & Yevjevich, 1978; Konecny & Nachtnebel, 1985; FEH, 1999; Bacova-Mitkova & Onderka, 2010; Bezak et al, 2014). Gottschalk and Krasovskaia (2002) suggested a threshold level equal to the flow that is three standard deviations larger than mean daily flow. Thresholds have also been defined as a quantile for significant floods, such as 50-year flood or 100-year flood. The quantile is estimated using the distribution fit to the daily flow series (Solari et al, 2012). Davison and Smith (1990) recommended using the threshold such that the mean-exceedance above the threshold is a linear function of threshold level. Ashkar and Rousselle (1987) proposed the use of the adequacy of the Poisson Process as an indicator to select the threshold. The threshold selected should have a dispersion index within the 95% confidence limit for the Poisson Process. The threshold can also be determined by a graphical method that combines the tests introduced above. Plots used in the graphical method are: (1) The mean residual life plot, which is a plot of the mean flood exceedances versus a range of threshold values; (2) Plots of scale and shape parameters of the Generalized Pareto distribution versus the threshold level; and (3) A dispersion index plot (Lang et al. 1999; Coles 2001; Burn et al., 2016). The threshold is selected based on all plots implying that human inspection is required in the graphical method (Burn et al., 2016).

Durocher et al. (2018) developed an automated threshold selection method. This approach is particularly useful when a threshold must be selected for a large number of sites for which manual threshold selection by the graphical method is tedious. In the automated threshold selection method, the declustering method introduced by Lang et al. (1999) is used to ensure no significant auto-correlation for extracted flows. The automated POT selection method assumes that exceedances above a well-chosen threshold follow the Generalized Pareto distribution with a constant shape parameter (Durocher et al., 2018). The p-value of the Anderson-Darling (AD) goodness of fit test is evaluated for a large range of threshold values ordered from lowest to highest. The first threshold with a p-value passing a critical p-value (0.25) is selected as the first candidate. This test statistic ensures that the extracted peaks follow a Poisson Process. To avoid the selection of a low threshold that violates the independence assumption, a second candidate threshold is introduced. The second candidate threshold is selected based on a fixed number of exceedances per year. The final threshold is selected from the two candidate thresholds. For a site where the flood quantile estimate from the first candidate is consistent with the second candidate, the first candidate is accepted as the final threshold. Otherwise, the higher threshold from the first and second candidate is selected as the final threshold.

2.1.3 Regional Flood Frequency Analysis

Another well known method for dealing with short data records is regional flood frequency analysis. The most commonly used regional flood frequency analysis model is the index flood procedure proposed by Dalrymple (1960). In the index flood method, the quantile for the site of interest can be calculated as the product of the index flood and the regional growth curve (Hosking & Waills, 1997):

$$Q_i(F) = \mu_i q(F) \tag{2.3}$$

where $Q_i(F)$ is the flood quantile for the i^{th} site, μ_i is the index flood, and q(F) is the regional growth curve. The most commonly used index flood is the mean value from the flow sequence. The regional growth curve is a dimensionless function that can be applied for every site within a pooling group, or region.

Screening of data is the first step for any statistical data analysis. For frequency analysis, the sample used in the analysis must represent the true population and have been drawn from the same distribution (Hosking & Wallis, 1997). Tests for outliers and trends in the data are well established in the literature (e.g, Barnett & Lewis, 1994; Kendall, 1975), such as the Mann-Kendall trend test and the quantile-quantile plot. Hosking and Wallis (1997) suggested using a discordancy measure to identify outliers. The discordancy measure evaluates differences between the L-moment ratios for a site and the average L-moment ratio for the pooling group. The discordancy measure (D_i) is used to identify outliers within the pooling group, and is defined as (Hosking & Wallis, 1997):

$$u_i = [t^{(i)}\ t_3^{(i)}\ t_4^{(i)}]^T \tag{2.4}$$

$$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i \tag{2.5}$$

$$A = \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \tag{2.6}$$

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \tag{2.7}$$

where u_i is a vector containing the L-moment ratios for site i, $\bar{u}$ is a vector of the unweighted group average L-moment ratios, and D_i measures the dispersion of L-moment ratios for site i with respect to its group. If the D_i the value for a site exceeds the critical value, that site is considered an outlier site. The critical value for the discordancy measure depends on the number of sites within the group and is defined as follows:

$$D_{i-critical} = \begin{cases} \frac{N-1}{3} & \text{if } N \leq 15 \\ 3 & \text{if } N > 15 \end{cases} \tag{2.8}$$

A popular method of estimating the parameters of a statistical distribution in regional flood frequency analysis is the method of L-moments. L-Moments are an alternative system describing the shape of a distribution, where L stands for linear combinations of probability weighted moments (PWMs). This method was developed by Grenwood et al (1979) and modified by Hosking and Wallis (1997); L-moments can be calculated from PWMs as follows:

$$b_r = n^{-1} \binom{n-1}{r}^{-1} \sum_{j=n+1}^{n} \binom{j-1}{r} x_{j:n} \tag{2.9}$$

$$l_1 = b_0 \tag{2.10}$$

$$l_2 = 2b_1 - b_0 \tag{2.11}$$

$$l_3 = 6b_2 - 6b_1 + b_0 \tag{2.12}$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \tag{2.13}$$

where $x_{j:n}$ is the vector of ordered sample values, n is the sample length, b_r are PWMs and l_1 to l_4 are called L-moments. L-moment ratios are defined as follows:

$$Mean = l_1 \tag{2.14}$$

$$L-CV = \frac{l_2}{l_1} \tag{2.15}$$

$$L-Skewness = \frac{l_3}{l_2} \tag{2.16}$$

$$L-Kurtosis = \frac{l_4}{l_2} \tag{2.17}$$

Similarly to conventional moments, the method of L-Moments can be used to estimate parameters of a distribution by equating sample L-moments to the theoretical L-moments for a particular distribution (Hosking & Wallis, 1997).

The next step of regional frequency analysis is forming the pooling group. In regional flood frequency analysis, pooling groups are sets of hydrometric stations that are considered to be similar. Within a homogeneous pooling group, all individual hydrometric stations share the same distribution (Hosking & Wallis, 1997; Burn, 1997). Two popular approaches for forming pooling groups are: 1) Forming a fixed pooling group based on flood-related attributes through cluster analysis; and 2) Forming a region for a specific site through the Region of Influence method. Cluster analysis is a well developed multivariate analysis for dividing a dataset into groups. The K-means algorithm is the most frequently used clustering method in practice (Burn, 1989, Chang et al., 2008; Dikbas et al., 2013). Hosking and Wallis (1997) advocate cluster analysis of site characteristics as the most practical method of forming regions from a large data set. The Region of Influence (ROI) method allows forming a unique pooling group for each station (Burn, 1990). A pooling group is formed by evaluating a similarity measure with respect to the target site. The ROI approach, and its modifications, have been extensively applied as a regionalization approach in flood frequency analysis (e.g. Zrinji and Burn, 1994; Burn, 1990; 1997; Castellarin et al., 2001; Grover et al., 2002; Merz and Blöschl, 2005; Faulkner et al, 2016; Burn et al., 2016).

The step following pooling group formation is assessing regional homogeneity. This is a critical step in regional flood frequency analysis. While many homogeneity tests have been proposed in the literature, the Hosking and Wallis (1997) heterogeneity measures are now commonly used by hydrologists (Castellarin et al., 2008). The Hosking and Wallis (1997) heterogeneity measures compare the dispersion of L-moment ratios for the pooling group with the expected dispersion, obtained from Monte-Carlo simulation. Three dispersion measures for L-moments are proposed by Hosking and Wallis (1997).

– Dispersion Measure for L-CV

$$V_1 = \sqrt{\frac{\sum_{i=1}^{N}(n_i(t^{(i)} - t^{(R)})^2)}{\sum_{i=1}^{N}(n_i)}} \tag{2.18}$$

– Dispersion Measure for both L-CV and L-skewness

$$V_2 = \frac{\sum_{i=1}^{N}(n_i((t^{(i)} - t^{(R)})^2 + (t_3^{(i)} - t_3^{(R)})^2)^{0.5})}{\sum_{i=1}^{N}(n_i)} \tag{2.19}$$

– Dispersion Measure for both L-skewness and L- kurtosis

$$V_3 = \frac{\sum_{i=1}^{N}(n_i((t_3^{(i)} - t_3^{(R)})^2 + (t_4^{(i)} - t_4^{(R)})^2)^{0.5})}{\sum_{i=1}^{N}(n_i)} \tag{2.20}$$

where t^R, $t_3{}^R$ and $t_4{}^R$ are the regional average L-CV, L-skewness, and L-kurtosis, respectively; t^i, $t_3{}^i$ and $t_4{}^i$ and n_i are the values for L-CV, L-skewness, L-kurtosis, and the sample size for site i respectively; and N is the number of sites in the pooling group. The heterogeneity measure, H_k is then calculated as:

$$H_k = \frac{V_k - \mu_{V_k}}{\sigma_{V_k}} \tag{2.21}$$

where μ_{V_k} and σ_{V_k} are the expected mean and standard deviation for the corresponding dispersion measure V_k. These two values are calculated through the Monte-Carlo simulation. Hosking and Wallis (1997) suggested that a group of sites may be regarded as acceptably homogeneous if H_k is less than 1, possibly heterogeneous if H_k is between 1 and 2, and definitely heterogeneous if H_k is greater than 2. It has been observed (Hosking & Wallis,

1993) that higher-order L-moments tend to be more homogenous in space than the lower-order ones. Therefore heterogeneity measure by using L-CV ratio is the most widely used heterogeneity measure.

Based on the results of the homogeneity test, revision of the pooling group may be required. Burn and Goel (2000) discussed several characteristics that regions should possess to ensure effective information transfer and therefore the efficient estimation of extreme flow quantiles. 1) Homogeneity of the pooling group. 2) The proper size of the pooling group. FEH (1999) suggests a pooling group should contain 5T station-years of data to estimate a flood event with a return period of T. If the initially constructed region is unacceptably heterogeneous or lacks a sufficient amount of station-years of data, revision to the region is needed. The goal of revising regions is to create homogeneous groups by moving catchments to a new region (usually based on the discordancy statistic), removing a catchment from a region and not reassigning it, merging or splitting regions, or replicating regions (Burn & Goel, 2000).

The next step in regional flood frequency analysis is to select the proper statistical distribution for the final pooling group. The best fit distribution is selected using a goodness-of-fit test (Hosking & Wallis, 1997). Some available approaches for the goodness-of-fit test are quantile-quantile plots, chi-squared test, Kolmogorov-Smirnov test, Cramer von Mises test, Anderson-Darling test, and tests based on moments or L-moments (Hosking & Wallis, 1997; Liao 2004). After the development of the L-moments technique in the 1990s (Hosking 1990; Hosking & Wallis, 1993, 1997), the method of L-moments became popular in flood frequency analysis. For the method of L-Moments, the quality of fit is assessed by:

$$Z^{dist} = \frac{\tau_4^R - \tau_4^{dist} + B_4}{\sigma_4} \tag{2.22}$$

where Z^{dist} is the goodness of fit measure; ${\tau_4}^R$ is the regional L-kurtosis; ${\tau_4}^{dist}$ is the theoretical L-kurtosis for the selected distribution; B_4 is the bias correction term and σ_4 is the standard deviation for the ${\tau_4}^R$. B_4 and σ_4 are computed from pooling groups having the same number of sites and same record lengths where pooling groups are generated by Monte-Carlo Simulation. The fit is declared to be adequate if Z^{dist} is sufficiently close to zero, a reasonable criterion being $|Z^{dist}|$ less than 1.64 while the lowest magnitude Z^{dist} is indicative of the distribution having the best fit to the data (Hosking & Wallis, 1997).

The final step in regional flood frequency analysis is to estimate parameters for the fitted statistical distribution. Some well-developed methods for estimating statistical parameters are the method of moments, the method of L-moments, and maximum likelihood. The method of moments involves equating the sample statistical moments to the distribution moments (Hosking & Wallis, 1997). Similar to the method of moments, the method of L-moments matches sample L-moments to population L-moments. The key idea for the maximum likelihood (MLE) is to estimate the parameter values that maximize the likelihood function, where the likelihood function is the joint probability distribution of the random sample of given observations (Haddad & Rahman, 2010). The MLE can produce unbiased solutions with high efficiency; however it often requires numerical computation and non-linear optimization. Since the available data records for flood frequency analysis are typically short, the method of L-moments can achieve similar performance to the MLE with less computation and is thus a popular method in regional flood frequency analysis.

2.2 Proposed Flood Frequency Analysis Framework for Canada

This research is based on a flood frequency analysis framework for Canada developed by (Mostofi Zadeh & Burn, 2019) and (Mostofi Zadeh et al., 2019). This section introduces two regional flood frequency analysis frameworks for Canada; a regional annual maximum flood frequency analysis (AMAX RFFA) and a regional peaks-over-threshold flood frequency analysis (POT RFFA). Both frameworks have the following components: screening of data; formation of pooling groups; assessment and revision of pooling groups; estimation of the regional growth curve.

2.2.1 Regional Annual Maximum Flood Frequency Analysis Framework for Canada

2.2.1.1 Data Preparation and Screening

Data used for the annual maximum flood frequency analysis (AMAX RFFA) framework are annual maximum flow series from Water Survey of Canada hydrometric stations. The data used in flood frequency analysis must initially be screened to ensure they satisfy the independent identically distributed (IID) assumption for the analysis. Two tests were selected for data screening, temporal trend analysis and change-point-analysis. The Mann-Kendall nonparametric trend test is commonly used in trend analysis for hydrometric data and is based on calculating at test statistic S, through:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \tag{2.23}$$

$$sign(\Theta) = \begin{cases} +1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases} \tag{2.24}$$

where S is the Mann-Kendall test statistic, n is the sample length and $\mathrm{x_i}$, $\mathrm{x_j}$ are values from the flow series (Kendall, 1975). Significant serial correlation in a data series can impair the robustness of trend detection (Wang et al., 2015) given the assumption of serial independence of data by the Mann-Kendall test (Őnőz & Bayazit, 2012). To overcome the effect of serial correlation, the block bootstrap sampling (BBS) technique is applied. The procedure for BBS Mann-Kandell trend test is summarized as follows (Khaliq et al., 2009): First, apply the Mann-Kendall trend test to the original flow series. Second, determine the number of significant serial correlations k. Third, resample the original data in blocks with size of k + η and estimate the Mann-Kendall test statistic for each simulation. Fourth, construct the distribution of Mann-Kendall test statistic based on the simulation results. Fifth, compare the Mann-kendall trend test statistics from the original flow series to the distribution from step four. The trend in the original flow series is significant if the original Mann-Kendall trend test statistics is located in the tail of the simulated distribution.

Shifts in a data series also affect the power of the Mann-Kendall test (OBrien & Burn, 2014). Change-point analysis is used to detect abrupt changes in the moments of the data. Tan and Gan (2015) indicated that the Pettitt test is the best option among other tests.

The nonparametric rank-based Pettitt test is given by (Pettitt, 1979):

$$K_T = max|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sign(x_j - x_i)| \tag{2.25}$$

where $\mathrm{K_T}$ is the change-point for the given flow series; the change-point is detected by using the 5% significant level. If a change-point is detected in the flow series, the series is first divided into two subseries (before and after change-point) and then trend analysis is performed on the subseries. In the AMAX RFFA framework, both trend and change-point analyses are performed in which the former is done after the latter.

2.2.1.2 Forming the Pooling Group

2.2.1.2.1 Super-Region Approach 1114 hydrometric stations were selected after data screening. The selected stations are located across Canada. Due to the high variability of physiographic and climate characteristics across the country, Mostofi Zadeh and Burn (2019) proposed a hierarchical process of creating super-regions from cluster analysis. The watershed mean annual precipitation and drainage area are used as attributes in the cluster analysis. Figure 2.3 shows the 6 super-regions derived from cluster analysis. The ROI approach was used to form station specific pooling groups within each super-region.

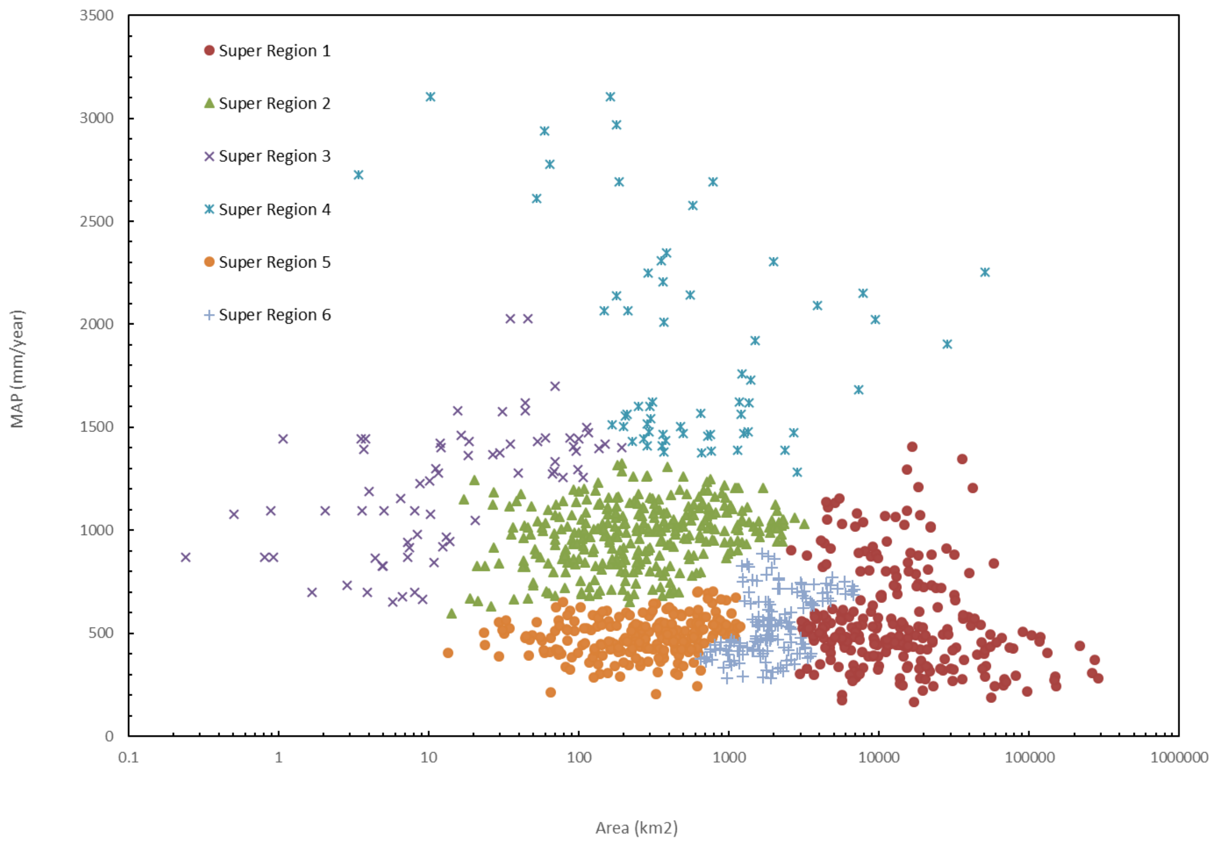

Figure 2.3: Catchment characteristics of 771 Canadian catchments (from Mostofi Zadeh & Burn, 2019).

2.2.1.2.2 Region-of-Influence Approach The key idea in the ROI approach is calculating the similarity measure between a target site and other sites within the super-region containing the target site. The similarity measure is an essential requirement for the formation of pooling groups (Burn, 1997). Flood seasonality measures are used to define the similarity measure. In this analysis, the Julian date of occurrence of peak flow for a flood event i is converted to an angular value using (Burn, 1997):

$$\theta_i = (\text{Julian Date})_i \frac{2\pi}{\text{length for given year}} \tag{2.26}$$

where θ_i is the angular representation for flood event i. January 1$^{\text{st}}$ is day 1 and December 31$^{\text{st}}$ is day 365 (or 366 in a leap year). The average Julian day for a flood sequence with n events can be calculated as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} cos(\theta_i) \tag{2.27}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} sin(\theta_i) \tag{2.28}$$

The mean event date (MD) represents the average occurrence of floods for given site. It can be defined as:

$$MD = arctan(\frac{\bar{y}}{\bar{x}})(\frac{\text{length of the year}}{2\pi}) \tag{2.29}$$

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \tag{2.30}$$

where $\bar{\mathrm{r}}$ is dimensionless and ranges from zero to one. A value close to one represents perfect regularity and a value close to zero represents low regularity.

Seasonality characteristics of the flood regime can be calculated by Equations (2.29) and (2.30). A nival flood regime usually has high regularity and a pluvial flood regime usually has lower regularity.

Chen et al. (2013) proposed a modified flood seasonality measure including the magnitude of each flood. Details for this approach are described as follows:

$$\bar{x}' = \frac{\sum_{i=1}^{n} q_i cos(\theta_i)}{\sum_{i=1}^{n} q_i} \tag{2.31}$$

$$\bar{y}' = \frac{\sum_{i=1}^{n} q_i sin(\theta_i)}{\sum_{i=1}^{n} q_i} \tag{2.32}$$

where q_i is the magnitude for flood event i.

2.2.1.2.3 Forming the Initial Pooling Group The similarity measure between catchments can be obtained through numerical values measuring the difference for certain attributes affecting floods. In the literature, distance metrics have been used to form hydrological neighborhoods (e.g. Castellarin et al. 2001; Lance & Williams 1966; Tasker 1982; Mostofi Zadeh & Burn, 2019). The similarity measure used in this framework is the Euclidian distance between catchments for certain basin characteristics (Mostofi Zadeh & Burn, 2019) and can be calculated as:

$$D_{ij} = \sqrt{\sum_{m=1}^{M} (x_m^i - x_m^j)^2} \tag{2.33}$$

where D_{ij} is the distance between site i and j, x_m^i are the flood attributes and M is the number of attributes. A smaller value of D_{ij} represents a higher similarity. Different combination of flood seasonality measures were tested, including $\bar{x}$ and $\bar{y}$; MD and $\bar{r}$; $\bar{x}'$ and $\bar{y}'$; MD' and $\bar{r}'$. The seasonality measure based on MD and $\bar{r}$ was recommended (Mostofi

Zadeh & Burn, 2019). Hosking and Wallis (1997) found that no substantial benefit is gained when forming regions with more than 20-25 sites. Therefore, the first 24 sites with the highest similarity with respect to the target site were combined with the target site to form the initial pooling group.

2.2.1.3 Assessing and Revising of the Pooling Group

This analysis uses the most common approach for testing regional homogeneity developed by Hosking and Wallis (1997) and described in Section 2.1.3. A modified homogeneity measure that takes into account cross-correlation between sites within the pooling group is used in the annual maximum regional flood frequency analysis framework (Castellarin et al., 2008). The modified homogeneity statistics is calculated as:

$$H_{1,adj} = H_1 + C\bar{\rho}^2(R-1) \tag{2.34}$$

where $\mathrm{H}_{1,\mathrm{adj}}$ is the adjusted value for the homogeneity measure, H_1 is the homogeneity measure introduced by Hosking and Wallis (1997), C is an empirical corrector with a value of 0.122, $\bar{\rho}^2$ is the averaged correlation of concurrent flows and R is the number of sites within the pooling group.

If the adjusted homogeneity measure passes the criteria of homogeneity stated in section 2.1.3, then the initial constructed pooling group is considered as the final pooling group. Otherwise, a revision process is required. In the revision process, the site having the highest effect on the modified homogeneity statistics is removed one site at time until the pooling group is homogeneous.

2.2.1.4 Estimating the Regional Growth Curve

The sample L-moment ratios of annual maximum flow series from N sites in the pooling group are used to derive the regional growth curve. The frequency curve for each site within the pooling group is derived from the regional growth curve by the index flood method, described in section 2.1.3.

Regional L-moment ratios are calculated by taking the weighted average of L-moment ratios for all individual sites in the pooling group. Detailed information on L-moment ratios and how to calculate them are discussed in Section 2.1.3. The regional L-Moment ratios are defined as:

$$t^R = \frac{\sum_{i=1}^{N} w_i t^i}{\sum_{i=1}^{N} w_i} \tag{2.35}$$

where i represents each individual site within the pooling group, w_{i} is the weighted coefficient for the i^{th} site, $\mathrm{t}^{(\mathrm{i})}$ represents L-moment ratios for the i^{th} site. The weight coefficient uses both the station record length and the site similarity with the target site and is given as (FEH, 1999; Burn, 2003)

$$w_i = s_i n_i \tag{2.36}$$

where $\mathrm{n_i}$ is record length for site i and $\mathrm{s_i}$ is the similarity rank for site i. The similarity rank comes from the similarity measure in the pooling group forming process decribed in Section 2.2.1.2.3.

The six distributions selected as candidates for fitting are Generalized extreme value distribution (GEV), Generalized normal distribution (GNO), Generalized logistics distribution

(GLO), Pearson type III distribution (PE3), Generalized Pareto distribution (GPA) and Wakeby distribution (WKB). The fit of different distributions and their suitability as candidates can be evaluated by the goodness-of-fit statistic described in Section 2.1.3. The L-moment ratio diagram is used as a graphic tool to inspect the goodness-of-fit.

Once a distribution is identified for the final pooling group, parameters for the fitted distribution are estimated by the method of L-moments. Details for sample L-moments are discussed in Section 2.1.3.

2.2.1.5 Uncertainty Analysis

The selected measure of uncertainty for the proposed annual maximum regional flood frequency analysis framework is the width of the confidence interval (CI) for the estimated quantile. The method used for computing confidence intervals was developed by Hosking and Wallis (1997). In this approach, the confidence interval is estimated through Monte-Carlo simulation where the realization of the final pooling group is generated using L-moments and the fitted distribution from the final pooling group. For sites without a long data record, the width of the confidence interval provided valuable information about the performance of the regional framework.

2.2.2 Regional Peaks-Over-Threshold Flood Frequency Analysis Framework for Canada

2.2.2.1 Data Preparation and Screening

Data used for the peaks-over-threshold (POT) flood frequency analysis framework are daily flow records from Water Survey of Canada hydrometric stations. Determining the appropriate threshold value is the initial step in POT flood frequency analysis. As described in Section 2.1.2.2, the POT series should have the following characteristics: 1) No significant auto-correlation within POT series; and 2) POT series should follow the Poisson Process. The automated threshold selection method introduced in Section 2.1.2.2 is used in the POT RFFA framework to generate POT flow series.

The POT series is extracted based on the selected threshold. The diagnostic checks used in the annual maximum flood frequency analysis framework are also applied to ensure the independent identically distributed assumption. This is described in section 2.2.1.1.

2.2.2.2 Forming the Pooling Group

The procedure for region forming is exactly the same as for the annual maximum flood frequency analysis framework. It is described in section 2.2.1.2.

2.2.2.3 Assessing and Revising the Pooling Group

The procedure for assessing the pooling group is mostly the same as for the annual maximum flood frequency analysis framework, except that cross-correlation is not considered within the pooling group. The basic assumption for cross-correlation between different sequences is that all sequences share a common time interval. This assumption is hard to fulfill for POT series within the pooling group. Therefore the homogeneity test used in POT regional flood frequency analysis is the Hosking and Wallis (1997) homogeneity measure described in section 2.1.3. The revision process for the pooling group is the same as the annual maximum flood frequency analysis framework, introduced in section 2.2.1.3.

2.2.2.4 Estimating the Regional Growth Curve

Unlike the annual maximum flood frequency analysis framework introduced in section 2.2.1.4, only one distribution, the Generalized Pareto distribution, is selected for distribution fitting. Following the procedure described in section 2.2.2.1, extracted POT series follow the Poisson Process, which can be described with the following probability function:

$$P(N_t = n) = \frac{(\lambda t)^n}{n!} exp(-\lambda t) \tag{2.37}$$

where λ equals the expected number of exceedances per year. The exceedance magnitudes, which are the difference between flow magnitudes and the selected threshold, are assumed to be independent and identically distributed following the Generalized Pareto distribution. The cumulative distribution function of Generalized Pareto distribution with the scale and shape parameter, α and κ respectively are:

$$\begin{cases} F(x) = 1 - exp(-\frac{x}{\alpha}) & \text{if } \kappa = 0 \\ F(x) = 1 - (1 - \kappa\frac{x}{\alpha})^{\frac{1}{\kappa}} & \text{if } \kappa \neq 0 \end{cases} \tag{2.38}$$

when ϰ=0, the Generalized Pareto distribution is reduced to the exponential distribution. The range of x is $(0 \leq x < \infty)$ for negative shape parameters. For positive shape parameters, an upper limit $(0 \leq x < \frac{\alpha}{\kappa})$ exists. The T-year event, $\mathrm{x_T}$, is defined as the $(1 - 1/\lambda T)$ quantile in the distribution of exceedances. The flow magnitude $\mathrm{x_T}$ can be obtained by following equation:

$$\begin{cases} x_T = F^{-1}(1 - 1/\lambda T) = \alpha ln(\lambda T) & \text{if } \kappa = 0 \\ x_T = F^{-1}(1 - 1/\lambda T) = \frac{\alpha}{\kappa}(1 - (1/\lambda T)^{\kappa}) & \text{if } \kappa \neq 0 \end{cases} \tag{2.39}$$

Scale (α) and shape (κ) parameters, can be estimated using L-moments.

$$\hat{\alpha} = \hat{\lambda_1}(\frac{1}{\hat{\tau}_2} - 1) \tag{2.40}$$

$$\hat{\kappa} = \frac{1}{\hat{\tau}_2} - 2 \tag{2.41}$$

where λ_1, τ_2 are first and second L-moment ratios, respectively. The method used to obtain flood quantile is the same as the annual maximum flood frequency analysis framework described in section 2.2.1.4.

2.2.2.5 Uncertainty Analysis

The procedure for uncertainty analysis is the same as for the annual maximum flood frequency analysis framework, described in section 2.2.1.5.

Chapter 3

Methodology

At-site annual maximum flood frequency analysis remains the most popular method to estimate design flood magnitudes in Canada due to its simplicity. The two proposed frameworks are both based on RFFA, using different input data. To create a flood estimation manual for Canada based on RFFA frameworks, a detailed comparison between the three frameworks is required. The most widely used method in the industry, at-site annual maximum flood frequency analysis, is used as a benchmark to assess the regional flood frequency analysis frameworks.

This methodology chapter consists of three sections. The first section compares at-site AMAX flood frequency analysis and regional annual maximum flood frequency analysis for Canada. This section seeks quantitative flood-related measures that can be used to determine sites where the AMAX RFFA framework is expected to perform worse than the at-site flood frequency analysis model. The second section compares the AMAX RFFA framework with the POT RFFA framework for sites with relatively long record lengths. The third section proposes an alternative flood frequency analysis framework for the Rocky Mountain region, a region for which the two regional flood frequency analysis frameworks do not perform well. The steps of the analysis are detailed in the following sections.

3.1 Assessing Performance of Annual Maximum Flood Frequency Analysis Framework for Canada

Unlike a physically-based surface water model, which models daily runoff for an entire water year, validation is missing for statistically-based frequency analysis models. Because the magnitude of different return period floods are unknown, the typical method to assess performance for flood frequency analysis models is uncertainty analysis. The method used in this research compares the width of the confidence intervals for site flood quantiles. The detailed procedure is described in Section 2.2.1.5. A narrower confidence interval represents better performance. This assessment is applied for all 1114 Water Survey of Canada hydrometric stations used for the AMAX RFFA framework for Canada. After identifying problematic sites, two studies were applied. First, different flood-related attributes were explored to find quantitative measures used to identify problematic sites. Second, the relationship between the performance of RFFA and regional L-moment ratios was explored.

3.1.1 Flood-Related Attributes

Flood-related attributes are critical factors in a region forming process. The key idea in a region forming process is to group similar sites, and flood-related attributes are used to calculate catchment similarity measures. Therefore, flood-related attributes are potential candidates for indicators used to identify problematic sites. In other words, if problematic sites share a common flood-related attribute, that flood-related attribute can be used as an indicator to identify problematic sites. Many flood-related attributes have been recommended by different researchers. Hosking and Wallis (1997) proposed the use of geographical and physiographic proximity, site elevation, mean annual precipitation, ratio of minimum average two-month precipitation to maximum average two-month precipitation, and beginning month of minimum average two-month precipitation and beginning month of maximum average two-month precipitation in a region forming process. Burn (1997) defined the flood seasonality measure for the site similarity measure. Faulkner et al. (2016) recommend using catchment area and lake attenuation factors as two variables to estimate design flood. Three flood-related attributes considered in this study are (1) geographic proximity, (2) basin elevation and average slope, and (3) lake attenuation.

3.1.1.1 Geographic Proximity

For geographic proximity, catchments closer to each other generally share similar hydrological and physiographical characteristics. Therefore, catchments with close geographical proximity are more likely to share the same flood regime. In this study, geographical proximity was estimated by calculating the distance between catchment centroids within the final pooling group. In addition, the Canadian ecoregions were used in the analysis. Ecoregions were developed based on macro-features of climate and vegetation (Bailey, 2005). In the developed RFFA framework, each individual station will have its own pooling group. Therefore, the site of interest for each pooling group will have the highest importance. The geographical proximity analysis for pooling group was conducted using ArcGIS and the following steps:

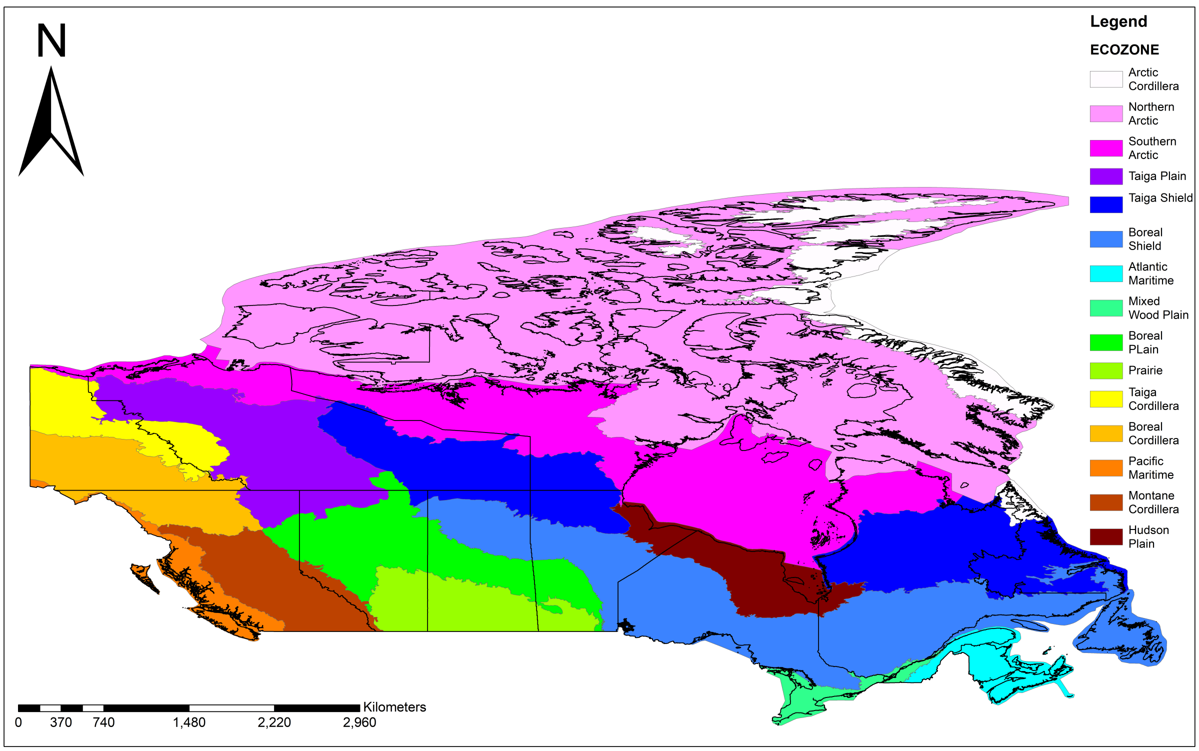

Figure 3.1: Canadian Ecoregions

– Calculate geographic distance between the site of interest and any other sites within the pooling group

– Apply penalty for sites located outside the ecoregion for the site of interest; the general rule is light penalty for neighboring ecoregions and heavy penalty for nonadjacent ecoregions

– Estimate average weighted geographic distance for the pooling group

3.1.1.2 Basin Elevation and Average Slope

The data source used to extract basin elevation was the archived Canadian Digital Elevation Model (CDEM) with 1.5 arc-second resolution. Basin elevation was estimated as the elevation at the centroid of the catchment. Basin slope is the average slope for the main reach from the headwater to the hydrometric station. The procedure to obtain basin elevation and basin slope is listed as follows:

– Identify site of interest

– Form a large CDEM covering the entire catchment by merging multiple 1.5 arc-second resolution CDEMs

– Delineate watershed for the site of interest and generate flow reach by using ArcGIS

– Validate watershed boundary and simulated reach by overlaying major catchments and hydro-network map for Canada

– Extrapolate basin elevation and average slope for the main reach

This analysis was applied to all sites where the pooling group was possibly homogeneous or definitely heterogeneous. This analysis was not extended to all 1114 sites because of the significant work required for the full-scale analysis.

3.1.1.3 Lake Attenuation Factor

Lakes and reservoirs within a watershed can significantly affect the flood regime. Lakes act as storage in the watershed and tend to flatten peaks in a hydrograph. Previous studies in Canada developed various indices to measure the lake effect on flood peaks. Rollings (1999) applied the lake attenuation factor, which is an additive index for individual lakes within the catchment, accounting for the open water area for lakes, the area draining into each lake and the total watershed area. In this study, a similar measure for total lake effect was used: flood attenuation by reservoirs and lakes (FARL). This method was originally developed in the UK Flood Estimation Handbook (FEH, 1999).

As the product of the attenuation index for each lake and reservoir within a watershed, FARL is defined by equation (3.1):

$$FARL = \prod_{i=1}^{n} \alpha_i \tag{3.1}$$

where n is the total number of lakes within the watershed and α_{i} is the attenuation index for the i^{th} lake. The index for an individual lake is calculated as follows:

$$\alpha = 1 - \sqrt{r}^{w} \tag{3.2}$$

where r is relative lake surface area to its tributary area,

$$r = \frac{surface\ area}{subcatchment\ area} \tag{3.3}$$

and w is the weight coefficient, which reflects the relative importance of the lake within the watershed (FEH, 1999):

$$w = \frac{subcatchment\ area}{catchment\ area} \tag{3.4}$$

Faulkner et al. (2016) conducted a nation-wide study of Canadian catchments to develop a regression model to estimate floods in Canadian catchments. In their study, a total of 2469 lakes with a surface area greater than 20 km^2 were used to calculate FARL. In the Faulkner et al. study, FARL was useful; however, many Canadian catchments are not included in their study. Moreover, for small drainage basins, even a small lake can have a significant lake effect. Therefore, in this study, the FARL index was estimated for 118 Canadian catchments. The global lakes and wetland database, developed by McGill University, provided the lake surface area. In this study, there was no constraint for minimum lake surface area. ArcGIS and GreenKenue were used to delineate the watershed and sub-catchments for each lake. Five lakes with the highest lake effect were used to compute FARL.

3.1.2 Heterogeneous Pooling Group

The discordancy measure (D_i) for the target site was computed for all 1114 Water Survey of Canada stations. The discordancy measure and adjusted Hosking and Wallis (HW) homogeneity measure (H_{adj}) were combined to determine the problematic sites. Moreover, for final pooling groups, which are not homogeneous, D_i was used to identify outliers within the pooling group. The details of obtaining D_i were introduced in Section 2.1.3. Outlier sites were compared with all the other sites within the pooling group. Moreover, the discordancy of the site of interest was estimated for all final pooling groups to identify potential problematic sites. Finally, common characteristics shared by outlier sites were identified.

3.1.3 L-Skewness and L-Kurtosis

The HW heterogeneity measure is used in both AMAX and POT RFFA frameworks. The details are described in Section 2.1.3. However, Viglione et al. (2007) found the power of the HW heterogeneity test drops for high L-skewness pooling groups. Lilienthal et al. (2017) found regional skewness and the HW homogeneity test rejection rate have a positive correlation through a simulation study. The above research indicates pooling groups with a high L-skewness defined as homogeneous through the HW heterogeneity measure may actually be heterogeneous. This study compares regional skewness and regional kurtosis with respect to confidence interval ratios. Further, the study explores L-moment ratios for the target site compared to its pooling group. The following steps were used:

- Compute regional L-moment ratios for all 1114 final pooling groups
- Calculate dispersion of L-moment ratios (number of standard deviations away from the regional mean) for the target site with respect to its pooling group

– Plot confidence interval ratio versus L-moment ratios (width of confidence interval from regional framework and at-site framework) to identify patterns and problematic sites

3.2 Compare Performance of Peak-Over-Threshold and Annual Maximum Flood Frequency Analysis Frameworks for Canada

In the last section, the problematic sites from the AMAX RFFA framework were identified and different aspects were explored. Theoretically, a properly extracted POT flow series contains more information than an AMAX series; therefore, the POT framework should perform better than the AMAX framework. However, for some sites, the AMAX framework performs better than the POT framework. Thus, the study compared the AMAX and the POT RFFA frameworks and developed quantitative measures to identify sites where the POT framework performs better than the AMAX framework. The comparison between the two frameworks was achieved by checking the following properties. First, the performance of the automated threshold selection technique described in Section 2.2.2.1. Second, the average extracted peaks per year for the POT flow series. Third, the spatial distribution of stations within the pooling group.

3.2.1 Check Performance of Auto-Threshold Algorithm

The most important step in POT extraction is to determine the flood threshold. The POT series from a bad threshold value will lead to poor performance. The first step in comparing the AMAX and POT frameworks is to check the validity of the threshold value from the POT framework. All individual events within a properly extracted POT flow series should be independent and follow the Poisson process. The generalized Pareto distribution was chosen as the distribution for POT extraction. Traditionally, the threshold value is determined through a series of statistical tests described in Section 2.1.2.2. The POT framework developed for Canada selects a threshold value based on an automated threshold selection algorithm, which is covered in Section 2.2.2.1. The automated threshold selection algorithm is effective for a dataset having many hydrometric stations since no manual judgment is required. Durocher et al., (2018) compare the outcome from the automatic threshold selection method with the manually selected threshold for different flow regimes across Canada. In this study, the comparison between manual and automatic threshold selection was extended to 261 Water Survey of Canada stations.

The graphic method introduced in Section 2.1.2.2 was used to manually estimate threshold values. Four plots were used in the graphic method: the mean residual life plot, plots for shape and scale parameters of the generalized Pareto distribution and the dispersion index plot. In addition, equation 3.5 was applied to check the assumption that all events within the extracted POT are independent (Lang et al. 1999):

$$\Theta < 5 + log(A) \\ or \\ X_{min} > 0.75min[Q_1, Q_2] \tag{3.5}$$

where Θ is the minimum separation in days between any two events, A is the watershed area in mi^2 and x_{min} is the lowest flow value between two successive flow events, Q_1 and Q_2.

A sufficient threshold should have the following properties. First, the residual life plot after the selected threshold has a linear trend. Second, the generalized Pareto shape and scale parameters are constant after the selected threshold. Third, the dispersion index after the selected threshold is 1. Fourth, the average peaks per year is greater than 1. The threshold values from both the manual and automatic methods were compared in terms of magnitude and peaks per year.

3.2.2 Peaks per Year

Peaks per year has been used as a quantitative measure to select the better model between POT and AMAX in many studies. Cunnane (1973) examined the performance between the AMAX model and POT model using generalized extreme value distribution (POT-GEV model) and concluded the POT-GEV model performs better than AMAX if peaks per year for the POT-GEV model exceed 1.65. Jin and Stedinger (1989) and Wang (1991) compared the performance for the AMAX model fitting generalized extreme value distribution and

POT model using generalized Pareto distribution (POT-GP model) and found both models perform equally well when peaks per year from the POT-GP model are above 1. In this study, the peaks per year from the POT series was compared with the performance of the POT model for the corresponding site to determine the recommended peaks per year to select from the AMAX and the POT framework.

The performance of the regional AMAX and regional POT models was estimated only by using stations that were stationary and had more than 50 years of record within Canada. For the sites with record lengths greater than 50 years, the flood quantiles of 5-year, 10-year, 25-year, and 50-year events from the at-site FFA is believed reliable. The flood quantile from the regional AMAX and regional POT models were compared with the quantile estimated from the at-site FFA model, and a smaller dispersion recommends better performance. The RMSE was used to measure the dispersion and was calculated using equation (3.6) (FEH, 1999 and Mostofi Zadeh et al., 2019):

$$RMSE_T = \sqrt{\frac{\sum_{i=1}^{N_{long}}(lnq_{T_i} - lnq^{p}_{T_i})^2}{N_{long}}} \tag{3.6}$$

where $\mathrm{RMSE_T}$ is the measure of uncertainty for a given return period, T, $\mathrm{N_{long}}$ is the total number of long-record sites, $\mathrm{q_{T_i}}$ is the T year flood quantile for the site i, and $\mathrm{q^p{}_{T_i}}$ is the T year flood quantile from the regional method for site i.

The better model between the AMAX and POT frameworks has a smaller $\mathrm{RMSE_T}$. The peaks per year for each station's POT series was then compared to the model performance.

3.2.3 Spatial Distribution of the Stations in a Pooling Group

The spatial distribution of the stations is also an important factor to consider when assessing the AMAX and POT frameworks for Canada. Section 3.1.1 described some flood-related attributes comparing the AMAX RFFA framework with at-site FFA. Those flood-related attributes were used to find effective quantitative measures to choose the better framework. A GIS study was used to evaluate flood-related attributes, including the lake effect and geographic proximity within the final pooling group for each target station.

3.3 Investigation of Alternative RFFA Framework for Rocky Mountain Region

The result of comparing the AMAX framework with the at-site FFA and the comparison between the AMAX and the POT frameworks had similar conclusions. Two problematic regions, the Canadian Prairie and the Rocky Mountains, were identified. The Canadian Prairie has a complex flood generation mechanism. A detailed discussion is presented in the results section. For the Rocky Mountains, stations for which the AMAX framework outperformed the POT framework did not have common flood-related attributes. An alternative POT framework is proposed for the Rocky Mountain region (POT-Rocky framework), considering its unique characteristics. The detailed procedure for the proposed POT framework is described in the following sections.

3.3.1 Data Preparation and Screening

The preprocessing steps for the POT-Rocky framework are the same as the POT framework covered in section 2.2.2.1.

3.3.2 Forming the Pooling Group

In the POT-Rocky framework, the Rocky Mountain region consists of three Canadian ecoregions: Montane Cordillera, Boreal Cordillera and Taiga Cordillera. The total number of stations in the Rocky Mountain region is 261, as Figure 3.2 illustrates. Similar to the AMAX and POT frameworks, the ROI approach was used to rank the similarity between stations. In the POT-Rocky framework, four flood-related attributes were used to calculate catchment similarity:

- The modified flood seasonality measure
- Elevation at the outlet of each catchment
- Mean annual precipitation at the centroid of each catchment
- The geographic proximity between stations

The modified flood seasonality measure was introduced in Section 2.2.1.2.2 and was normalized by average melt day from the site of interest. The 2 m air temperature from National Centers for Environmental Prediction North American Regional Reanalysis data from 1979 to 2017 was used in the analysis. A high order polynomial was fitted to the

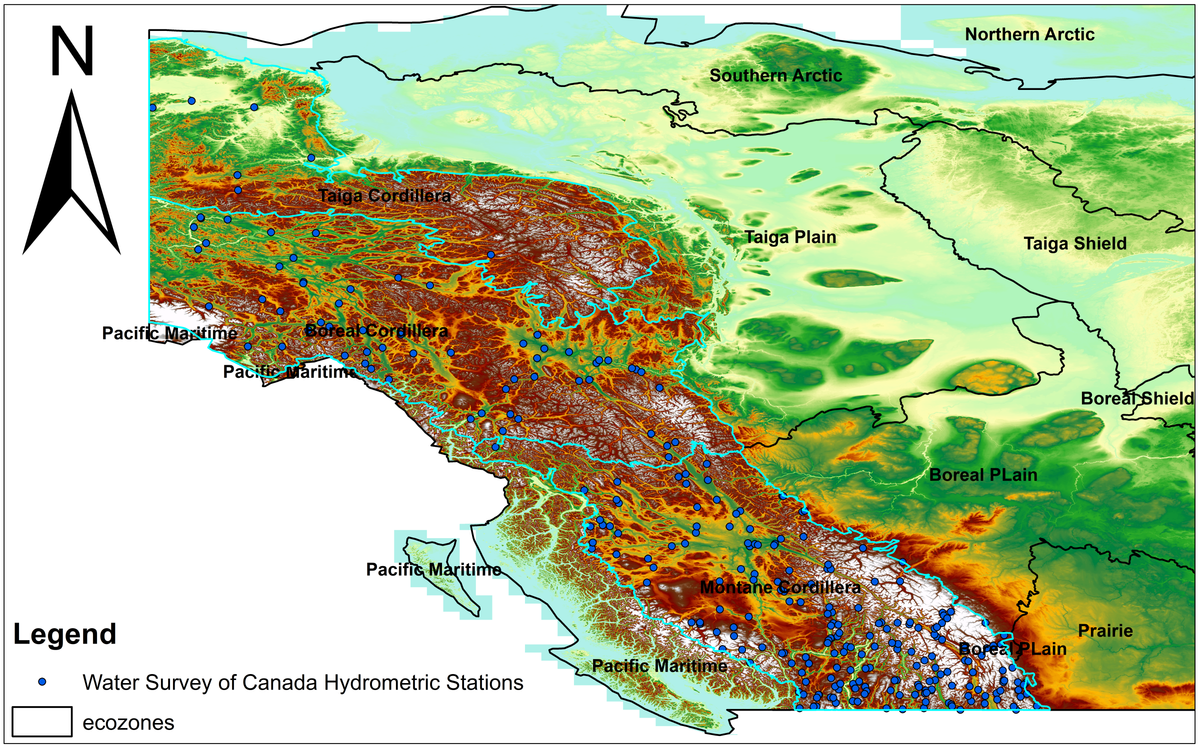

Figure 3.2: Water Survey of Canada Hydrometric Stations for the POT-Rocky framework

temperature data at the centroid of each catchment. The fitted polynomial is used to estimate the average melt day for each catchment within the watershed. The average melting day for all 261 catchments was then used in normalizing the Julian day of the flood event. The difference in the average melting day between the target site and other sites was used to correct the Julian day of each flood event in the corresponding POT flow series and can be described as follows:

$$(JulianDate)_{adjusted,i} = (JulianDate)_i + (MD_i - MD_{target}) \tag{3.7}$$

where Julian $\text{Date}_{\text{adjusted,i}}$ is the adjusted Julian day series for station i, MD_{i} is the average melt day for site i and $\text{MD}_{\text{target}}$ is the average melt day for the target site. The adjusted Julian day was used to estimate the flood seasonality measure discussed in Section 2.2.1.2.2. The flood seasonality measure was estimated using the adjusted Julian day. Details for estimating the flood seasonality measure are described in Section 2.2.1.2.2 and Section 2.2.1.2.3.

The elevation effect is a unique and important flood-related attribute for the mountainous region. Therefore, elevation at the outlet of each catchment was selected as a similarity measure to account for the elevation effect. The data and steps to extract station elevation were introduced in Section 3.1.1.2. Closer elevation leads to higher catchment similarity.

Mean annual precipitation (MAP) at the centroid of each catchment was used in AMAX and POT RFFA frameworks for Canada to form super-regions (Mostofi Zadeh et al, 2019). MAP has been tested as an effective way of grouping similar catchments. Therefore mean annual precipitation was included as a similarity measure in this work. Data used to estimate catchment mean annual precipitation was provided by Environment and Climate Change Canada. Similar catchments usually have similar mean annual precipitation.

Last, the distance between flow gauges was used to calculate geographic proximity. The smaller distance between stations leads to higher catchment similarity.

Since all four similarity measures had different units and different scales, normalization was required before estimating the overall similarity between catchments. The overall similarity measure between stations i and j was calculated using:

$$D_{total} = \sum_{k=1}^{N} W_k D_{ij}^k \tag{3.8}$$

where W_k is the weight coefficient. In this study, the four similarity measures were treated equally. D_{ij}^k is the normalized k^{th} similarity measure. The initial pooling group consists of the target site and 24 stations with the highest similarity to the target site.

3.3.3 Assessing and Revising the Pooling Group

The assessing and revising processes were the same as the POT framework introduced in Section 2.2.2.3. The HW homogeneity measure was used to assess the pooling group homogeneity. For the heterogeneous pooling group, the site with the largest discordance measure was dropped and another candidate site was chosen based on catchment similarity until the pooling group was homogenous.

3.3.4 Estimating the Regional Growth Curve

The procedure for fitting the regional growth curve for the POT-Rocky framework was the same as the POT framework for Canada. The generalized Pareto distribution was fitted to the final pooling group. The method of L-moments was used to estimate the shape and scale parameters for the generalized Pareto distribution. The procedure is explained in Sections 2.1.3 and 2.2.2.4.

3.3.5 Uncertainty Analysis

The procedure for uncertainty analysis is described in Section 3.2.2. This analysis was achieved by comparing the estimated flood quantile from at-site FFA and POT-Rocky framework for stations with more than 50 years record. The 5-year, 10-year, 25-year, and 50-year events were used to estimate the degree of discordancy between flood quantile from at-site FFA and POT-Rocky RFFA. The performance of the POT-Rocky framework was compared with the AMAX framework and POT framework for Canada. The best model for the Rocky Mountain region has the smallest $\mathrm{RMSE}_{\mathrm{T}}$.

Chapter 4

Results

This chapter presents the results of the AMAX and POT RFFA framework evaluation and the performance of the alternative POT-Rocky framework. The methodology introduced in Chapter 3 was adopted in this research. This chapter is further divided into three sections. Section 4.1 discusses the performance of the AMAX RFFA framework for Canada and some quantitative measure for detecting sites where the AMAX RFFA framework performs worse than at-site FFA. Section 4.2 compares the AMAX and POT RFFA frameworks to find quantitative measures that can be used to select a better RFFA framework, and Section 4.3 proposes an alternative RFFA framework for the Rocky Mountain region and assesses its performance.

4.1 Performance of AMAX Regional Frequency Analysis Framework for Canada

This section provides a general assessment for the AMAX RFFA framework for all 1114 sites. Comparing the flood quantile for the regional method and at-site method is well developed and recommended in several flood estimation manuals. However, this method works for long-record sites. For short-record sites, the estimated quantile from the at-site model is not reliable. Since the majority of the 1114 stations had a record length of less than 40 years, the width of the confidence interval for the flood quantile function was selected as the evaluation metric to assess the performance of the AMAX RFFA framework over Canada. The methodology was discussed in Section 2.2.1.5, where the performance of the AMAX RFFA framework was compared with at-site FFA. The problematic site is defined as a site for which at-site FFA gives a narrower confidence interval than the AMAX RFFA framework. The confidence interval (CI) ratio is the width of the confidence interval for the flood quantile between AMAX RFFA and at-site FFA; the confidence interval for 50-year flood and 100-year flood were used in this analysis. The problematic site is defined as the site at which the CI ratio is greater than 1. Figures 4.1 and 4.2 depict the geographic distribution of the six super regions and the problematic sites.

Overall, the AMAX RFFA framework performed better than the at-site FFA for many Canadian catchments. However, two regions were identified as problematic: the Rocky Mountain region and the Prairie. The percent of problematic sites for each geographic region is provided in Table 4.1, where similar Canadian ecoregions were merged into seven geographic regions. The performance for each super-region is displayed in Table 4.2. Super-regions 5 and 6 have the highest percent of problematic sites. These regions mostly consist of stations in the Rocky Mountain region and the Canadian Prairie. The Canadian Prairie has a flat terrain and complex flood generation mechanism. The tributary area for a watershed within the Prairie region changes depending on the soil moisture and the amount of ponded water on the ground. Since its system is not stationary, the statistical model does not perform well in this region (Dumanski et al, 2015). Therefore, the Canadian Prairie was excluded in this study. For the Canadian Prairie, the multiple input time series model and physical-based hydrologic model are recommended. The multiple input time series model combines more information in the analysis, such as temperature and snow accumulation, leading to more accurate estimations of floods. The physically-based model, which is able to simulate the flood generation mechanism, can produce accurate flood estimations. However, the physically-based model is specific to catchments and requires users to have an in-depth understanding of the flood generation mechanism.

Table 4.1: Performance of the AMAX RFFA framework for seven geographic regions

Geographic Region	Percent Problematic Site (%)	Number of Site
Rocky Mountain	43.77	281
Canadian Prairie	40.57	106
Arctic	31.58	19
Taiga Plain	28.75	80
Mixed Wood Plain	22.32	112
Boreal Plain	21.13	355
Maritime	11.92	151

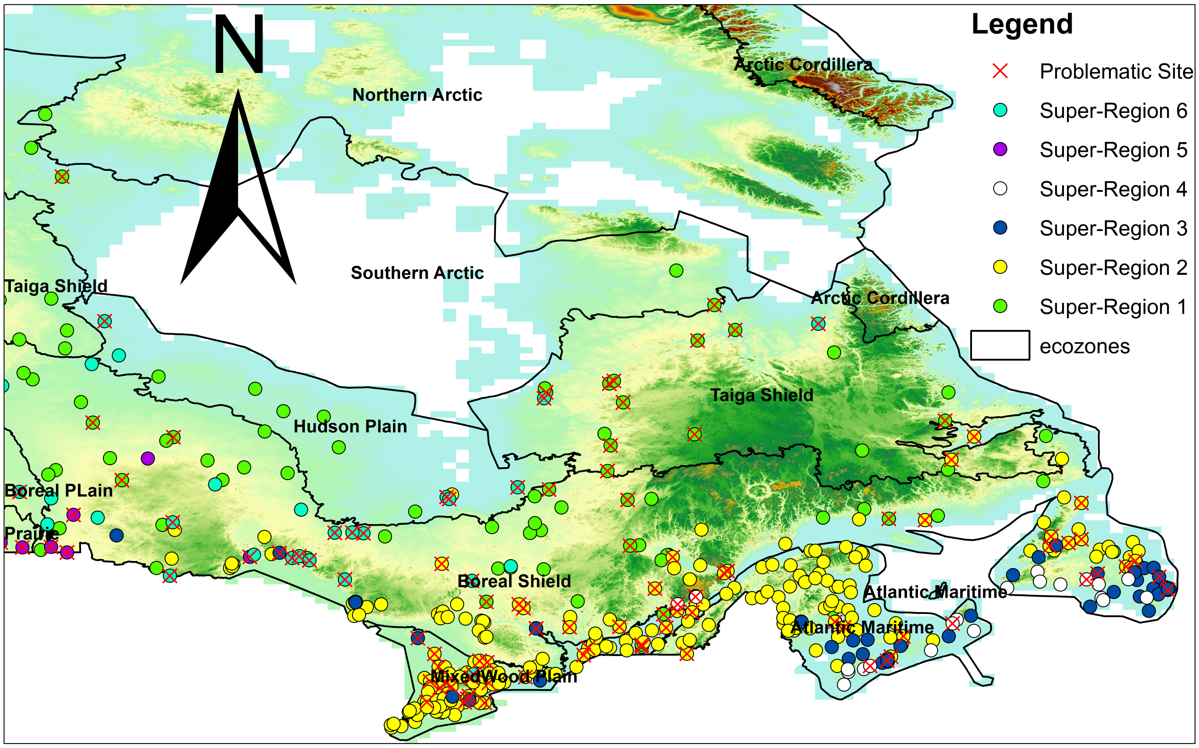

Figure 4.1: Super-regions and problematic sites for eastern Canada

Table 4.2: Performance for the AMAX RFFA framework for six super-regions

Super-region	Percent Problematic Site (%)	Number of Site
1	21.32	258
2	18.43	331
3	18.06	72
4	24.19	62
5	45.85	229
6	45.68	162

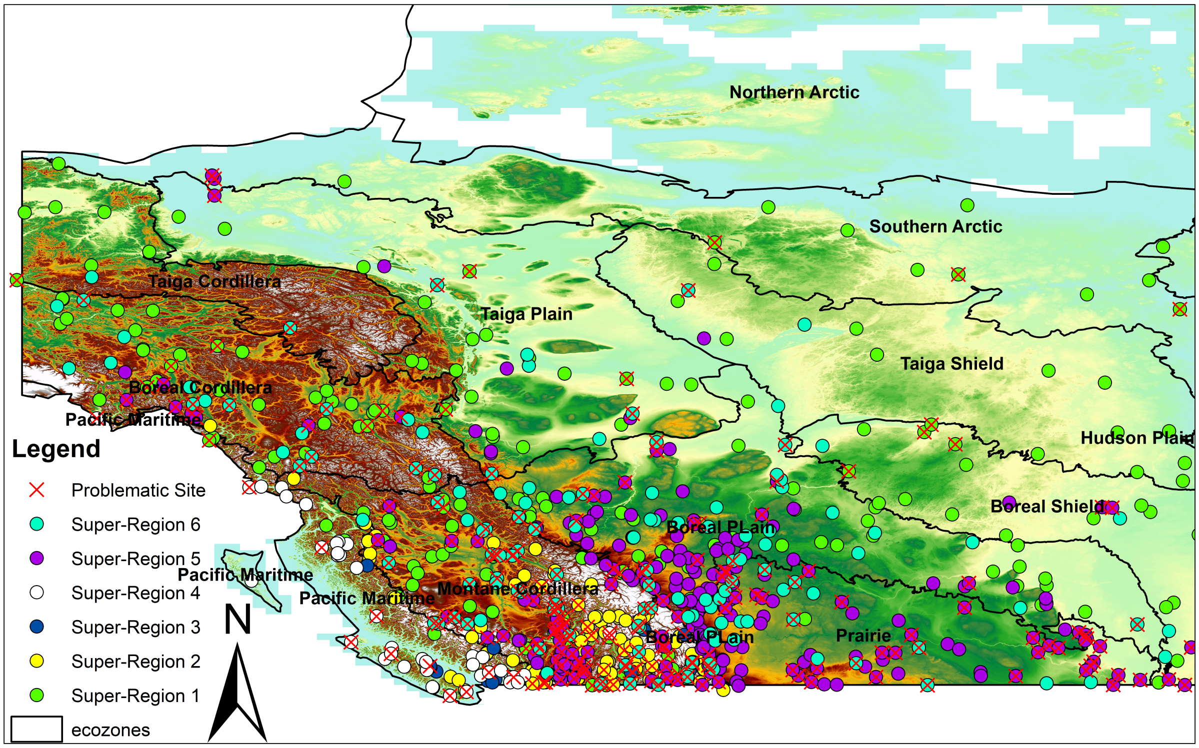

Figure 4.2: Super-regions and problematic sites for western Canada

4.1.1 Flood-related Attributes

In this section, two flood-related attributes, drainage basin average slope and lake effect, are explored. Due to the complexity of extracting the average basin slope and the lake attenuation factor, not all Water Survey of Canada stations were selected. The homogeneity of the final ROI pooling group was used as the basis for selecting subsets of the station set for analysis. Although the homogeneous final pooling group does not guarantee the final performance, it remains the most important indicator in the pooling group revision process. Of 1114 stations, 118 were selected to explore the lake effect and the basin average slope effect. The 118 selected ROI pooling groups were a possibly homogeneous region (1 $<H_{adj}$ <2) or a definitely heterogeneous region ($H_{adj} >2$). Figure 4.3 illustrates the geographic distribution of all 118 sites. Most of the stations are in the Rocky Mountain region, the Canadian Prairie, Boreal Plain and the Mixed Wood plain.

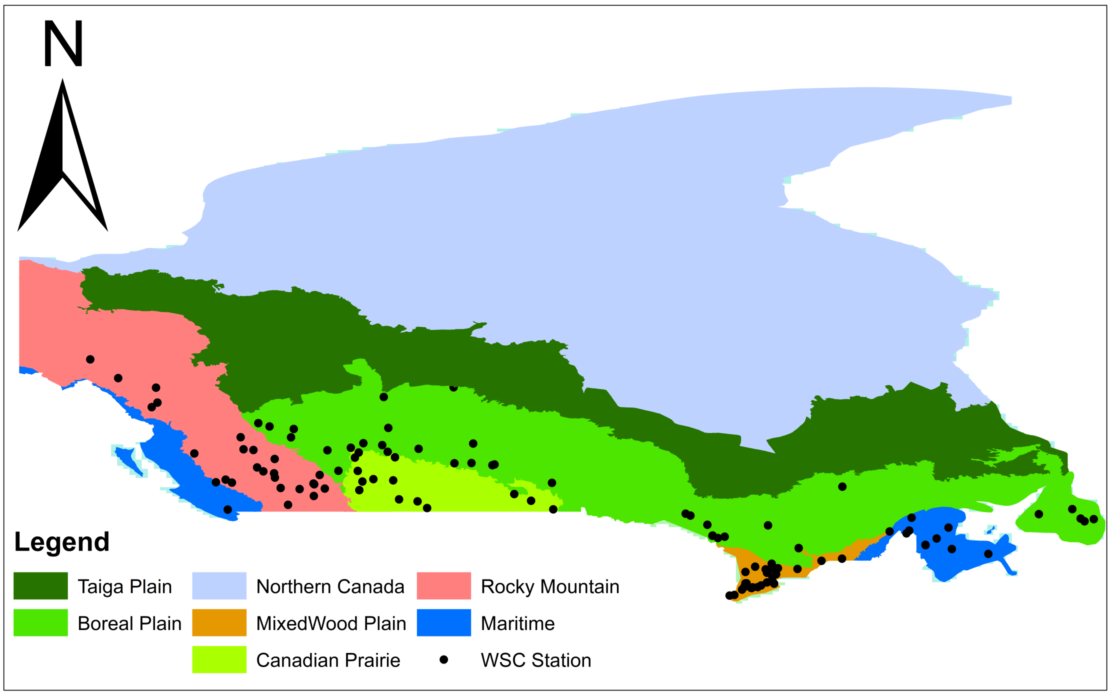

Figure 4.3: Selected Water Survey of Canada Stations

4.1.1.1 Average Basin Slope

Figure 4.4 depicts the relationship between the average basin slope and the confidence interval ratio (CI ratio). The result indicates no direct connection between the average basin slope and the CI ratio; however, a basin with a slope steeper than 15° is more likely to be a problematic site. This finding is subjective and needs future justification on more hydrometric stations.

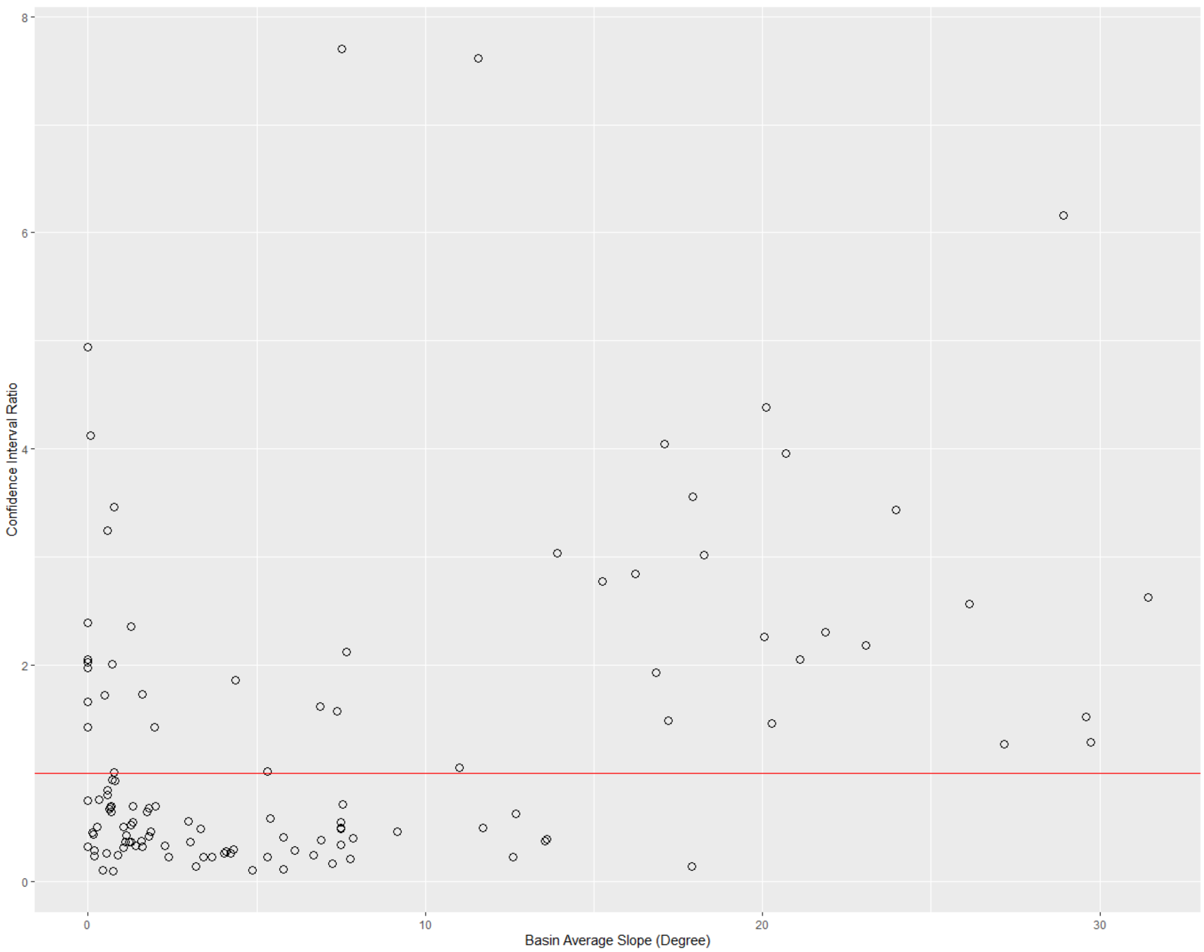

Figure 4.4: AMAX RFFA versus at-site FFA, with respect to average basin slope

4.1.1.2 Lake Effect

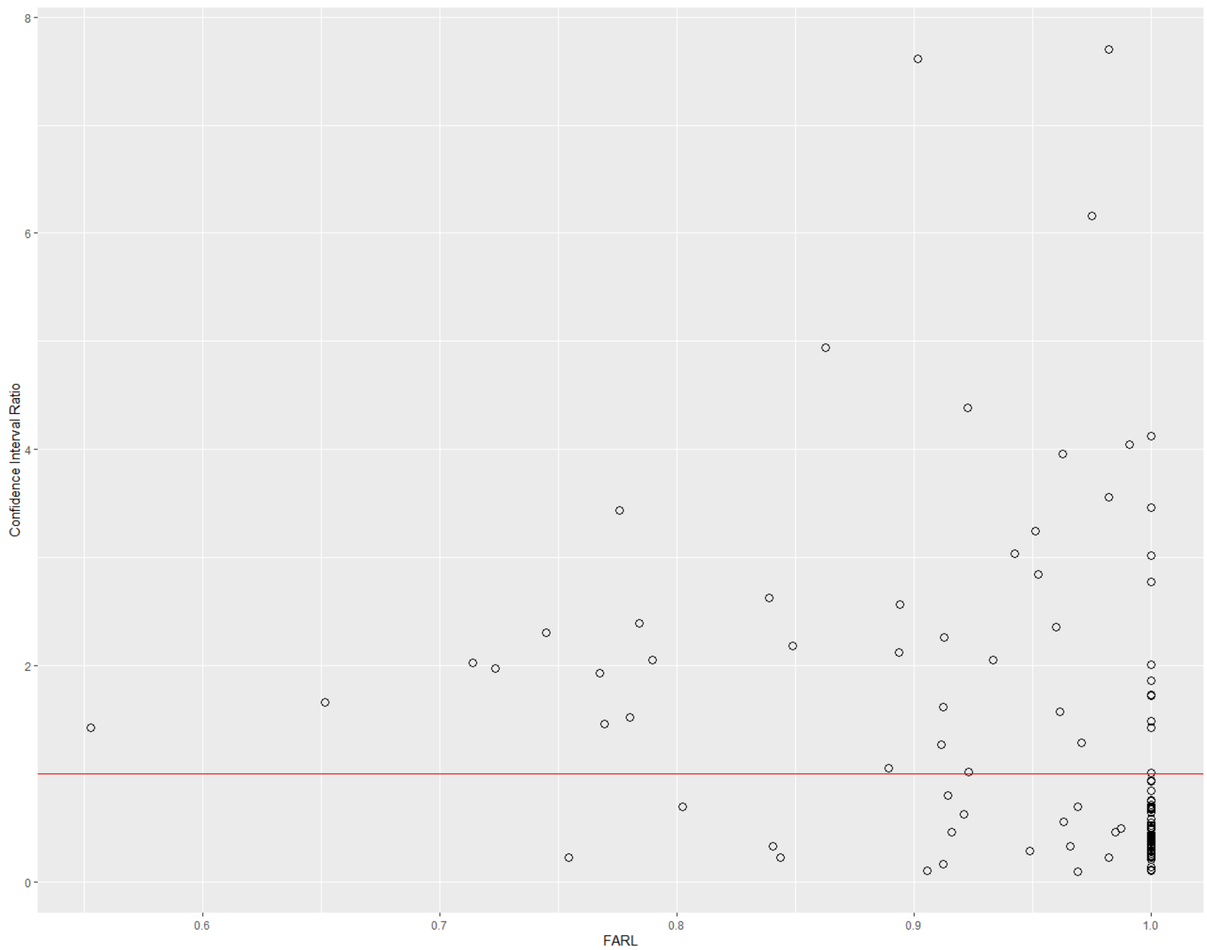

Figure 4.5: AMAX RFFA versus at-site FFA, with respect to lake effect

Figure 4.5 displays the relationship between the lake effect and the CI ratio. The FARL factor represents the basin lake effect, smaller FARL represents a strong lake effect while a watershed with no lake effect has a FARL factor of 1. The results indicate FARL is not an effective quantitative measurement for selecting the better model between the AMAX

RFFA and at-site FFA. However, the FARL index remains useful in some unique cases, such as when the target site has a significant lake effect but other sites within the pooling group have no significant lake effect. Moreover, the FARL index is sensitive in the Rocky Mountain region and the Boreal plain.

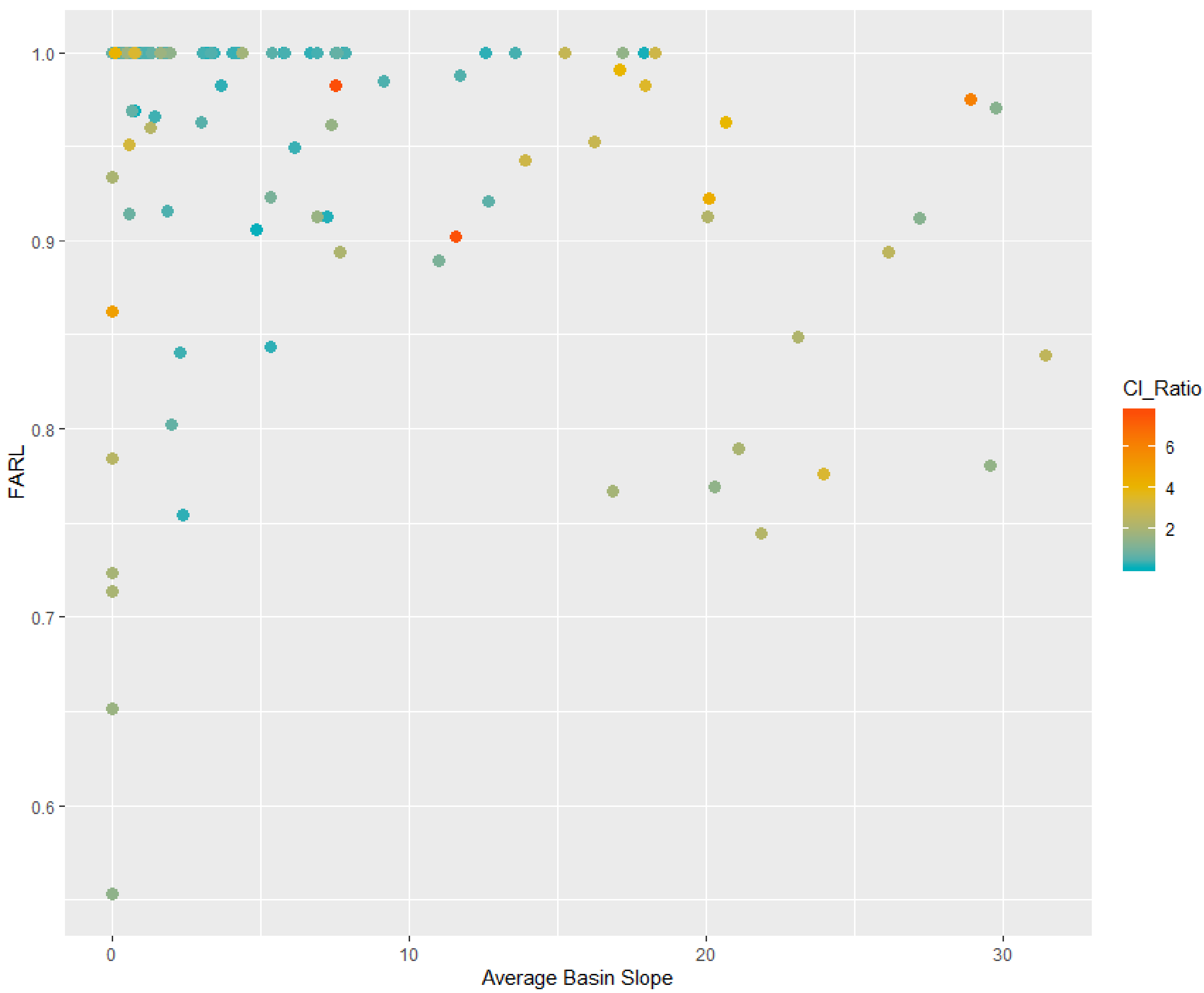

Figure 4.6: AMAX RFFA versus at-site FFA, with respect to both lake effect and average basin slope.

Considering average basin slope and lake effect individually are not strong enough. Figure 4.6 presents the result of combining the average basin slope and the FARL index. The blue dots represent stations for which AMAX RFFA provides a narrower confidence interval. Watersheds with average slopes above 15° with or without the lake effect are problematic. For watersheds with low average basin slopes, lake effect is an effective quantitative measure. A watershed with a FARL less than 0.8 is likely to be a problematic site.

In the next two sections, the AMAX RFFA performance is assessed with the homogeneity and L-moment ratios of its corresponding final pooling group. All 1114 Water Survey of Canada stations were used in the analysis.

4.1.2 Heterogeneous Pooling Group

Figure 4.7 provides an overview of the results of combining the discordancy measure for the target site ($\mathrm{D_i}$) and the homogeneity measure for the final pooling group ($\mathrm{H_{adj}}$). The stations with discordancy measure greater than 3 or homogeneity measure greater than 2 were defined as heterogeneous. However, there is no direct link between the confidence interval ratio and the degree of heterogeneity for the final pooling group. Sites with homogeneity measures exceeding 2 are more likely to have a confidence interval ratio greater than 1. For all sites with confidence interval greater than 2, the discordancy of the target sites was greater than 2. Moreover, the degree of dispersion for the target site concerning its final pooling group was related to the confidence interval ratio. Afterward, the dispersion of L-moments for the target site to its final pooling group was checked.

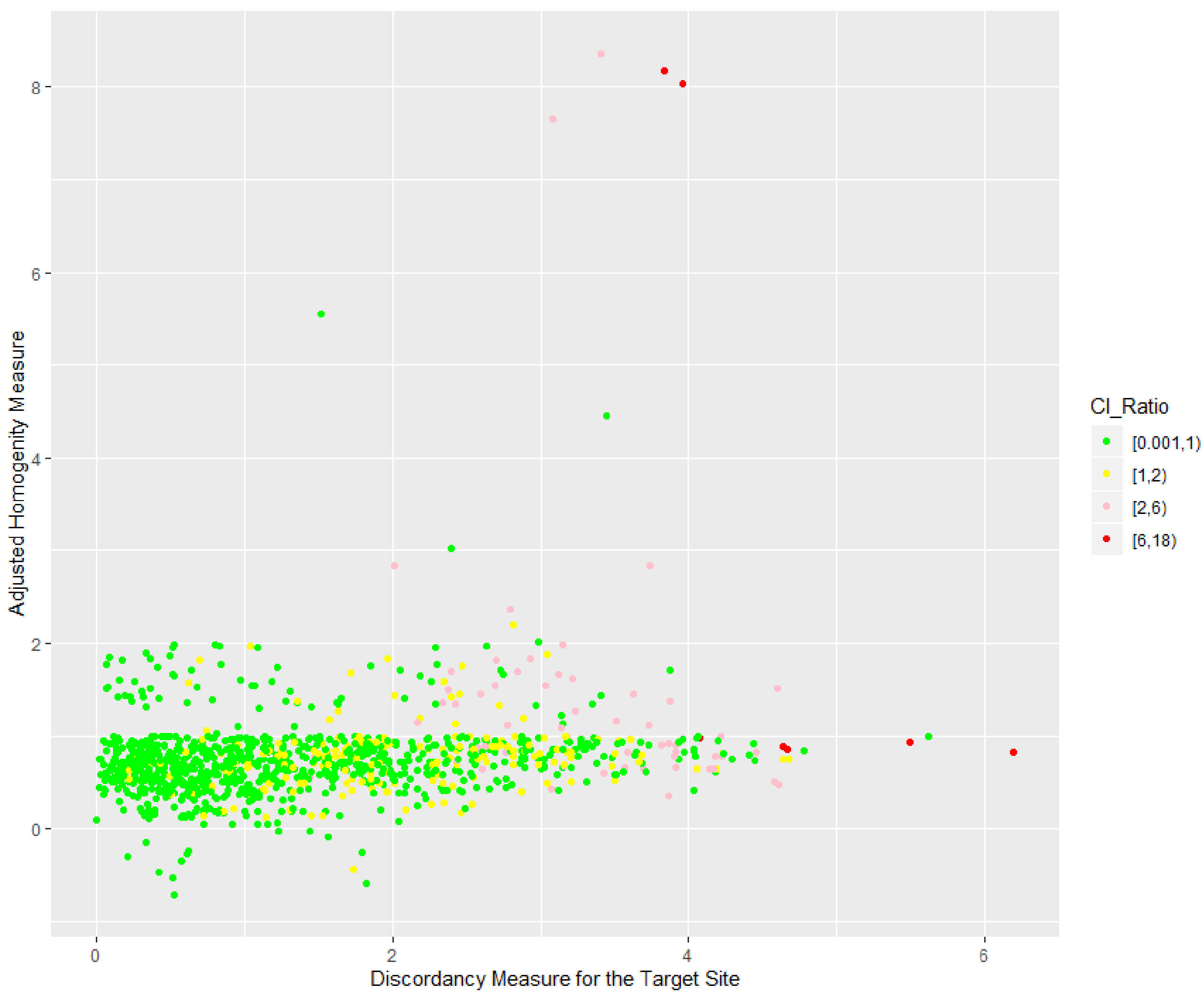

Figure 4.7: AMAX RFFA versus at-site FFA, with respect to discordancy measure for the target site and adjusted homogeneity measure

4.1.3 L-Skewness and L-Kurtosis

In this section, the confidence interval ratios are compared with the discordancy of L-moment ratios of the sites of interest. Figures 4.8, 4.9 and 4.10 illustrate the cross-comparison between three L-moment ratios: LCV ratio, L-skewness and L-kurtosis. L-moment ratios were found to be effective quantitative measures to select the best model from the AMAX RFFA and at-site FFA frameworks. The results indicate the problematic sites share some common characteristics. The discordancy of the L-CV ratio and L-skewness are more effective than L-kurtosis. Generally, when the L-CV ratio or L-skewness from the target site was 2 standard deviations lower than the pooling group average value, that site was more likely to be a problematic site. In other words, if the flow series from the target site has a low average flow and skews to the left compared to all the other flow series within its pooling group, at-site FFA is recommended for that specific site.

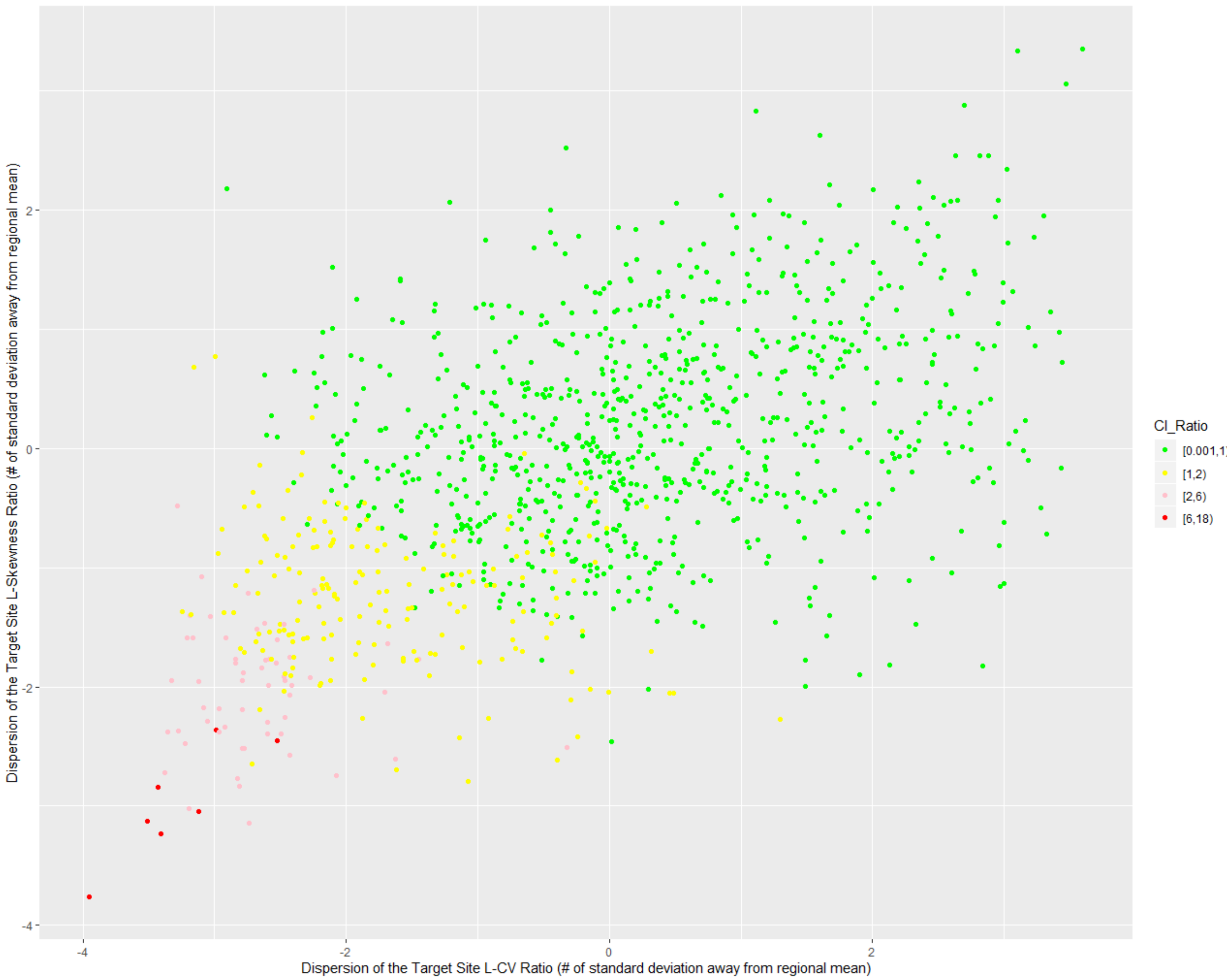

Figure 4.8: AMAX RFFA versus at-site FFA, with respect to the spread of L-skewness and L-CV ratio for the target site

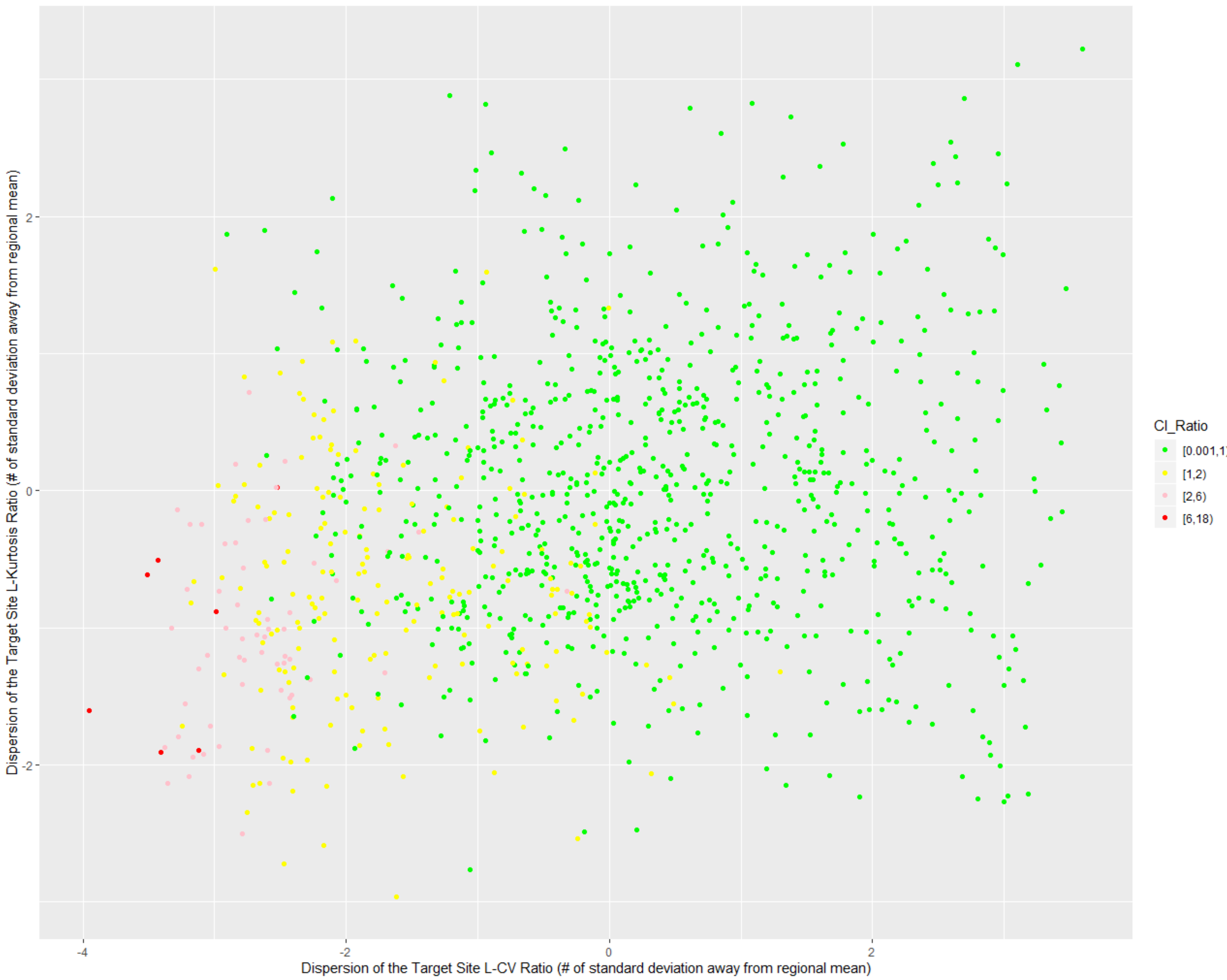

Figure 4.9: AMAX RFFA versus at-site FFA, with respect to the spread of L-kurtosis and L-CV ratio for the target site

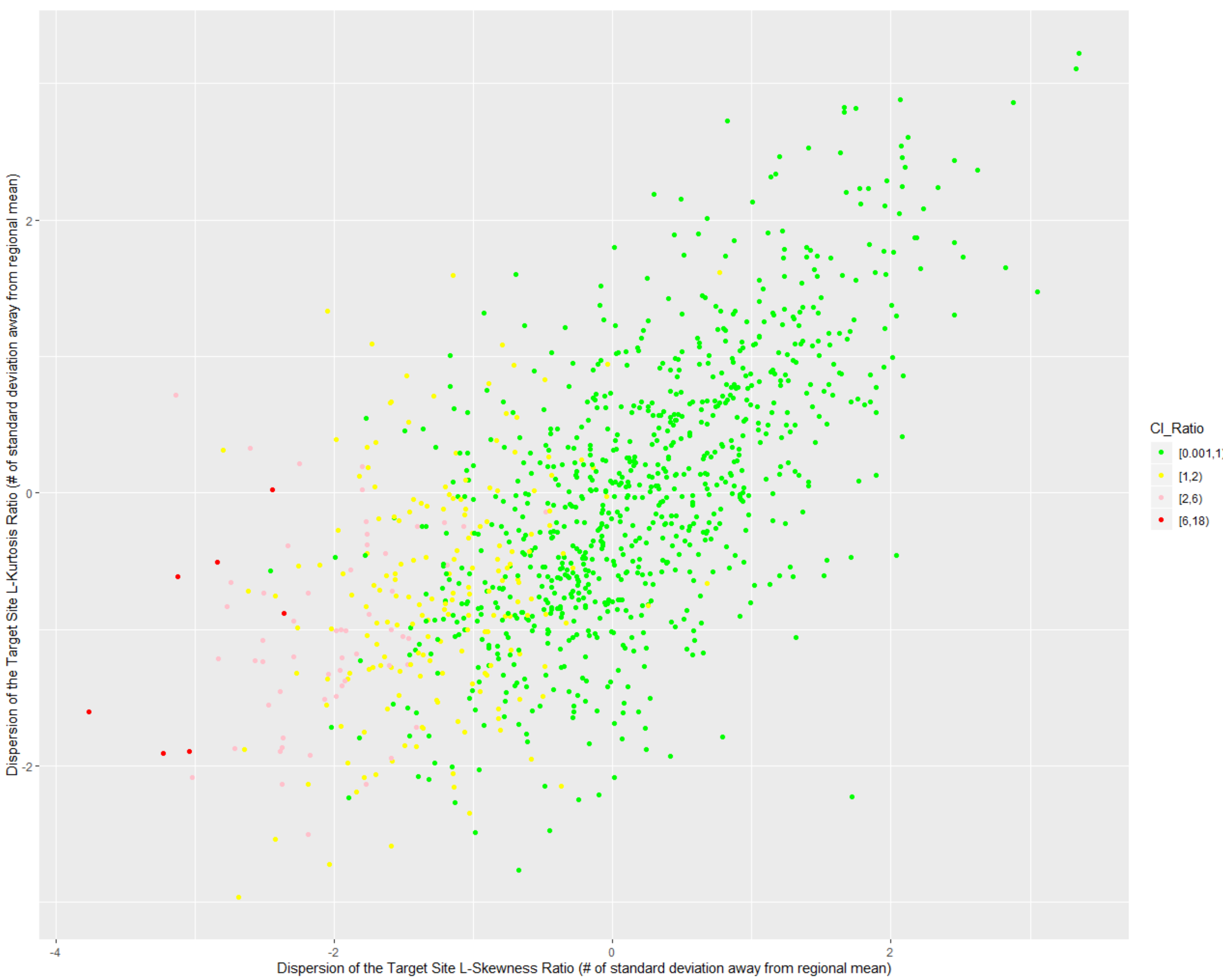

Figure 4.10: AMAX RFFA versus at-site FFA, with respect to the spread of L-skewness and L-kurtosis for the target site

4.2 The Annual Maximum Regional Flood Frequency Analysis Framework Versus The Peaks-Over-Threshold Regional Flood Frequency Analysis Framework

The performance measure for the confidence interval is easy to apply without any constraint on record length. Therefore, this performance measure was used (Section 4.1) to provide a comprehensive analysis for the performance of the AMAX RFFA framework across Canada. The results indicate the AMAX RFFA framework performs better than at-site FFA for most stations. In this section, the AMAX RFFA framework is compared to the POT RFFA framework using another performance measure. Flood quantiles from the AMAX RFFA framework and the POT RFFA framework are compared with the quantile from at-site FFA and the smaller difference represents better performance. Since this performance is only valid for sites at which the quantile from at-site FFA is reliable, only long-record sites with more than 50-year records were considered. The assessment procedure was introduced in Section 3.2.2. For sites where the difference between flood quantiles from the AMAX RFFA and the POT RFFA is less than 15%, the two frameworks are considered to perform the same.

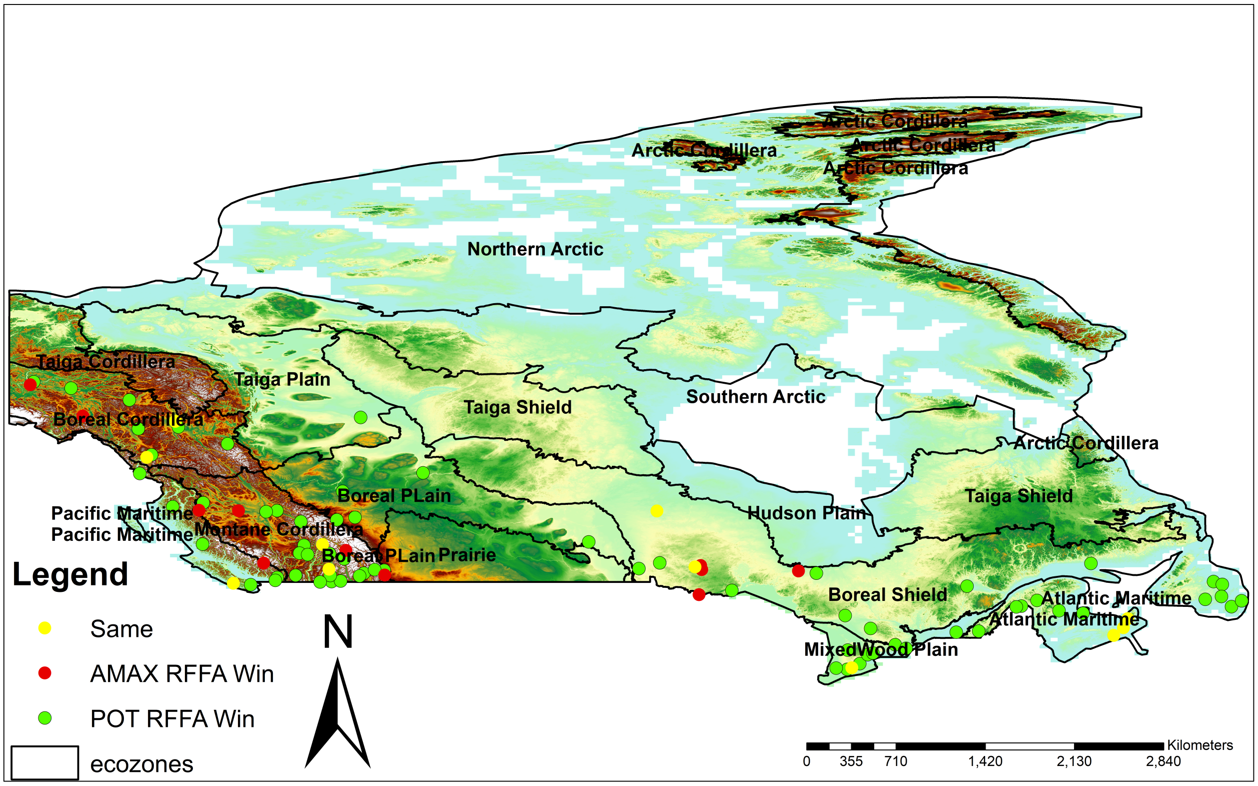

Figure 4.11: AMAX RFFA versus POT RFFA, for long record stations

Table 4.3: Performance of the AMAX RFFA and POT RFFA frameworks for long-record sites

	Number of Station
POT Win	67
Same	11
AMAX Win	11

Table 4.3 and Figure 4.11 present the performance for the AMAX RFFA and POT RFFA frameworks. The POT RFFA framework is superior to the AMAX RFFA framework for most of the long-record sites except some stations in the Rocky Mountain region and

the lake-rich region in the Boreal Shield. The Rocky Mountain region remains the most problematic region. This finding matches the results from Section 4.1. Theoretically, a successfully extracted POT flow series with proper pooling technique contains more information than the AMAX flow series; therefore, POT RFFA should perform better than AMAX RFFA for long-record sites. In the following sections, the automatically generated threshold value is compared with the manually selected threshold. The number of peaks per year from the POT flow series and the corresponding performance is compared with the recommended peaks per year from the literature. The spatial distribution for sites where AMAX RFFA performs better than POT RFFA is explored.

4.2.1 Performance of the Automatic Threshold Algorithm

The automatic threshold selection algorithm is used in the POT RFFA framework to generate the threshold. Therefore, the first step in the analysis was to check the performance of the automatic-threshold selection algorithm. In total, 261 stations in the Rocky Mountain region were used in the analysis. The manual threshold selection method is the graphical method introduced in Section 3.2.1. The threshold value and the peaks per year from both methods were compared to estimate the performance of the automatic threshold selection algorithm.

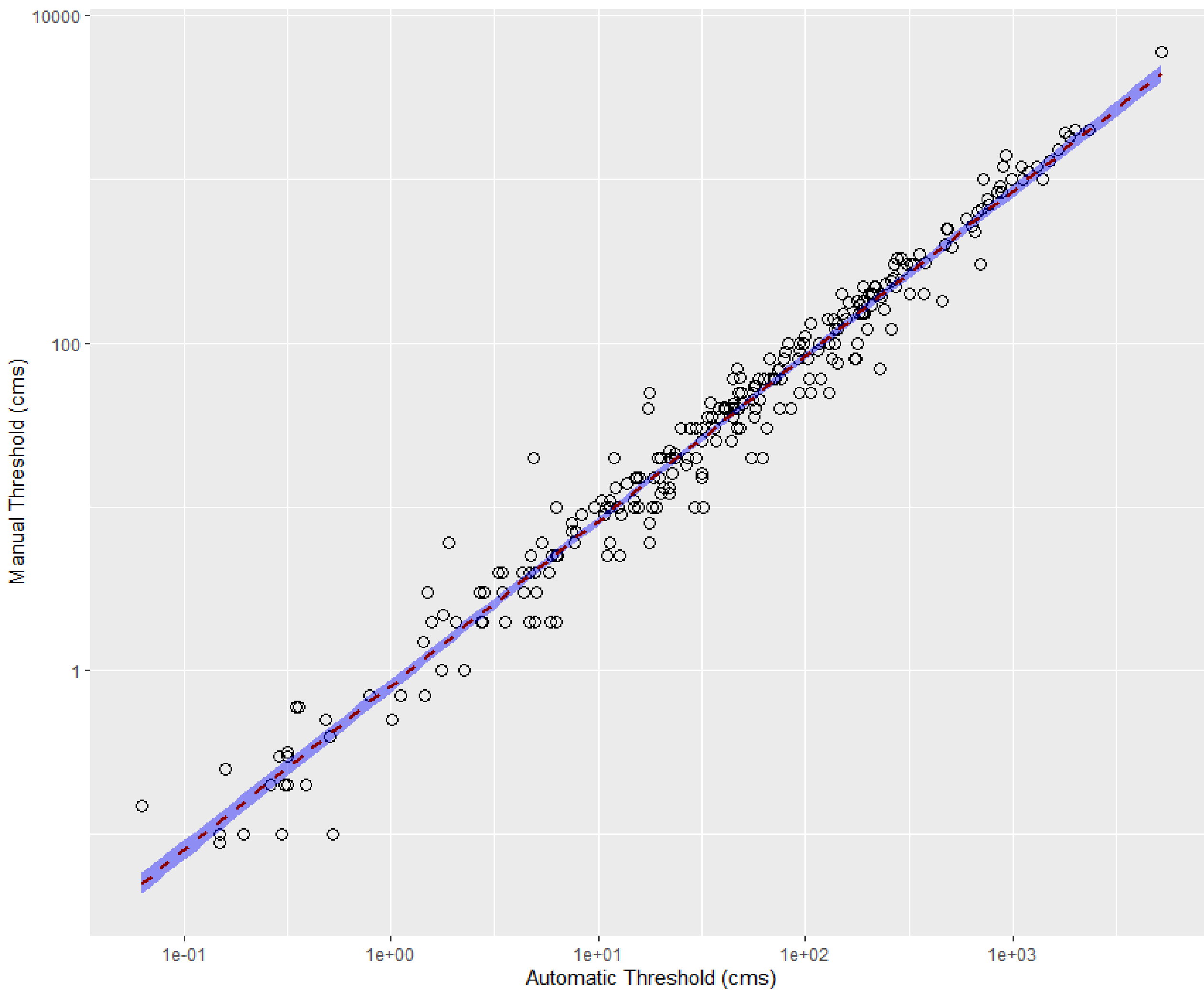

Figure 4.12: Automatic-threshold versus manual threshold, threshold value

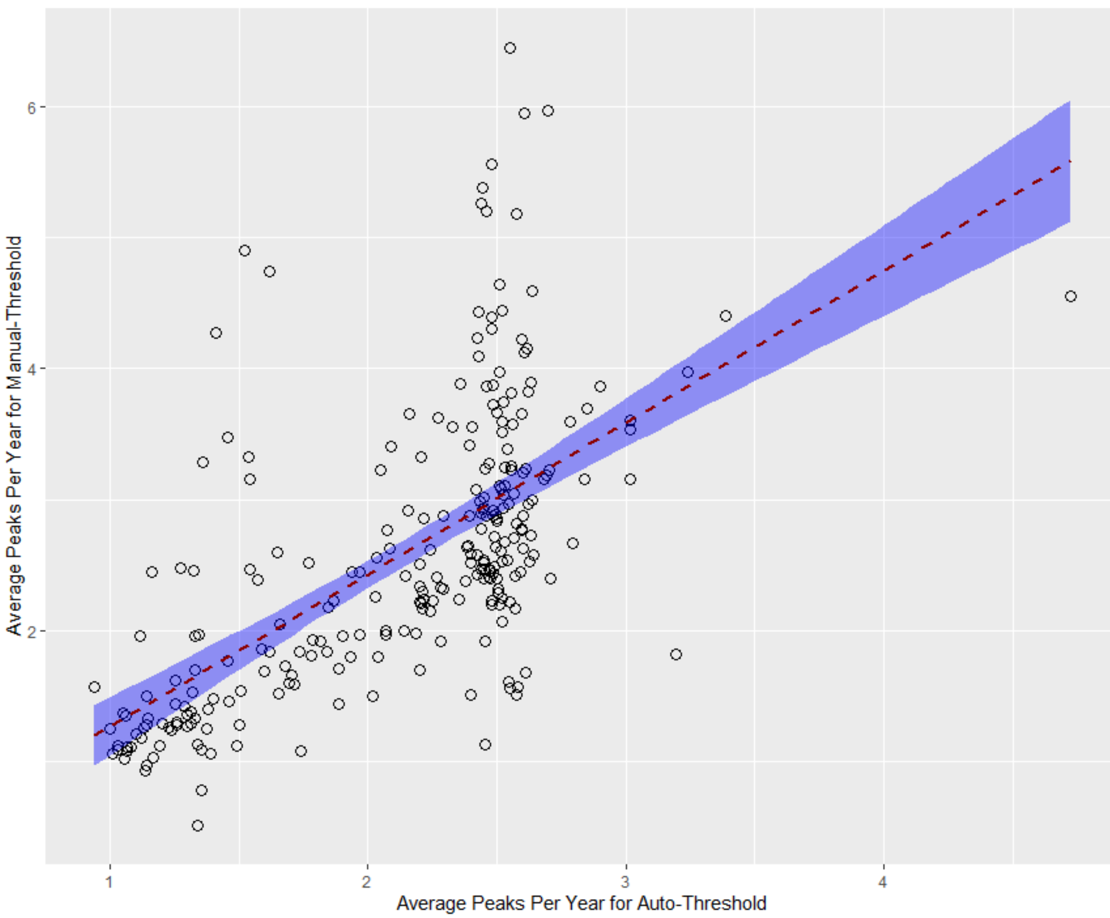

Figure 4.13: Automatic-threshold versus manual threshold, peaks per year

Figures 4.12 and 4.13 illustrate the performance of the automatic threshold selection method. The dashed line represents the trend for all data points and the shaded area is the 95% confidence interval. Both plots demonstrate the automatic threshold selection method can generate threshold values close to the manual selected threshold for 261 hydrometric stations within the Rocky Mountain region. Therefore, the threshold values used in the POT RFFA framework for the Rocky Mountain region were valid. Thus, the reason AMAX

RFFA performed better than the POT RFFA for some sites is not strongly related to the poor threshold values within the Rocky Mountain region.

4.2.2 Peaks Per Year

Figure 4.14 depicts the performance of the AMAX RFFA and the POT RFFA for the peaks per year for long-record sites across Canada. Previous literature (Jin & Stedinger, 1989 and Wang, 1991) concluded that when the peaks per year from the POT flow series ranges from 1 to 1.65, AMAX FFA is recommended (Section 3.2.2). This study supports that conclusion as most of the sites where the AMAX RFFA framework performed better than the POT RFFA framework had peaks per year less than 1.5 for the POT flow series. Peaks per year for the POT flow series is an effective way to select between the POT and AMAX RFFA frameworks. For a site with peaks per year less than 1.5, the AMAX RFFA framework is recommended. In other words, for hydrometric stations having a nival flood regime, the AMAX RFFA framework is recommended.

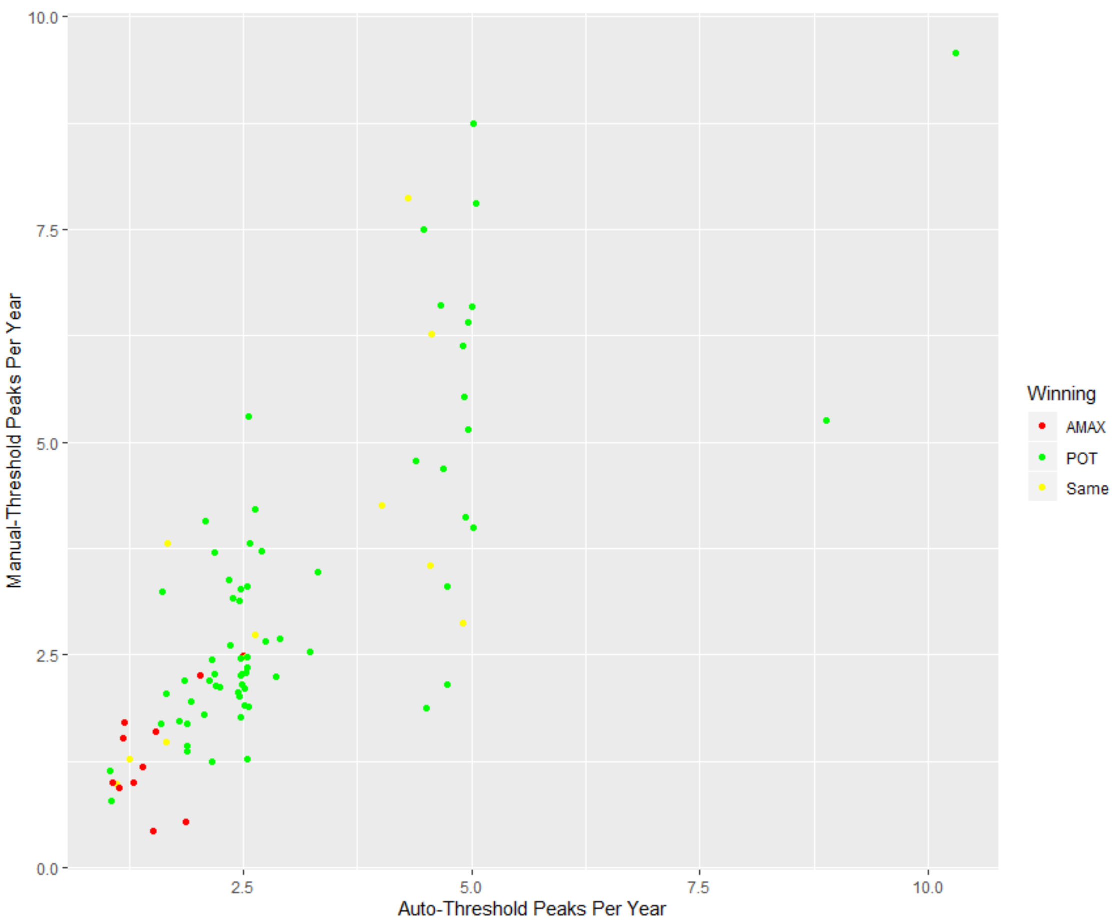

Figure 4.14: Peaks per year for long record sites across Canada

4.2.3 Spatial Distribution of Stations

A GIS study was conducted for sites where the AMAX RFFA framework performed better. No clear flood-related attributes were found. Regarding the spatial distribution of six super-regions in the Rocky Mountain region, no dominate super-region within the Rocky Mountain region exists. Therefore when forming the pooling group, geographically close stations with high similarity to the target site might have been excluded because those stations were from different super-regions. Moreover, a critical flood-related attribute, elevation, was not included in the similarity measure in region forming. In the next section, an alternative POT RFFA framework is proposed for the Rocky Mountain region.

4.3 Performance of the Alternative POT RFFA Framework for the Rocky Mountain Region

The alternative POT RFFA framework introduced in Section 3.3 was applied to 261 Water Survey of Canada stations. The results and the performance of the alternative POT RFFA framework for the Rocky Mountain region (POT-Rocky RFFA framework) are described in the following sections.

4.3.1 Data Preparation and Screening

Data used in this analysis was a subset of the stations from the POT RFFA framework located within the Rocky Mountain region. In total, 282 hydrometric stations were included in this framework. The automatic threshold algorithm was applied to all stations to select the threshold value. Seasonal stations were excluded. The trend test was applied for each POT flow series to remove all nonstationary flow series. This analysis followed the procedure introduced in Section 3.3.1. Finally, 261 of 282 stations were used in the POT-Rocky RFFA framework.

4.3.2 Form Pooling Group and Quantile Estimate

The procedure used in this section was described in Section 3.3.2.

Table 4.4: Homogeneity measure and station-years for long-record site from POT-Rocky RFFA framework

Station ID	H1	Station-Years (yr)	Station ID	H1	Station-Year (yr)
05AA022	0.88	652	05AA023	0.89	968
08CD001	0.02	2277	08CE001	0.68	1465
08ED002	**2.30**	**523**	08EE008	1.90	2421
08JB003	1.37	594	08KA004	0.76	2566
08KA005	0.97	2523	08KB001	1.46	2184
08LB047	1.00	2675	08LB064	0.70	2699
08LD001	0.91	2759	**08NA002**	**2.38**	**2138**
08ND012	1.81	2098	08ND013	1.46	2630
08NE039	1.79	2568	08NE074	0.94	2804
08NE077	0.77	890	08NE087	0.92	2592
08NG065	0.03	2158	08NH119	1.78	2489
08NH120	1.69	2675	08NK016	1.70	2394
08NL007	0.28	2337	08NL024	0.75	2780
09AC001	1.37	1178	09AE003	0.46	1965
09BA001	0.68	2237	09BC001	1.37	2290
09CD001	0.99	1047	09DD003	1.08	2151
10AA001	0.79	2131	10AB001	1.16	2143
10BE004	1.41	1683			

Table 4.4 presents the results for stations with more than 50-year records within the Rocky Mountain region; bolded stations indicate the final pooling group is definitely heterogeneous. Two final pooling groups were categorized as definitely heterogeneous; the other 33 pooling groups were homogeneous or possibly-homogeneous. The station years from the final pooling group ranges from 523 years to 2804 years ensuring the information provided by the pooling group is sufficient to estimate 100-year floods. Overall, the

POT-Rocky framework is able to delineate homogeneous pooling groups with long station years within the Rocky Mountain region.

The Generalized Pareto distribution was fitted for each final pooling group. The logic and necessary steps were discussed in Section 2.2.2.4.

4.3.3 Performance of the POT-Rocky RFFA Framework

The performance of the POT-Rocky RFFA framework was evaluated by comparing the flood quantile from the POT-Rocky RFFA framework to the quantile from at-site FFA for long-record sites. The cross-comparison of the AMAX RFFA, POT RFFA, and POT-Rocky RFFA is displayed in Table 4.5 and Figure 4.15.

Table 4.5: Performance measure for three RFFA frameworks at long-record sites in the Rocky Mountain region

Station ID	$RMSE_f$			Station ID	$RMSE_f$		
	AMAX	POT	POT-Rocky		AMAX	POT	POT-Rocky
05AA022	0.2459	0.4468	**0.1309**	08NE077	**0.1037**	**0.1144**	0.1612
05AA023	0.0933	**0.0463**	0.1798	08NE087	0.1486	0.0340	**0.0183**
08CD001	0.2252	**0.0363**	0.0510	08NG065	0.7081	**0.0009**	0.0541
08CE001	**0.1130**	**0.1177**	**0.1009**	08NH119	0.5030	0.1012	**0.0396**
08ED002	0.1204	0.3097	**0.0397**	08NH120	0.1118	0.0576	**0.0395**
08EE008	0.2393	0.0956	**0.0735**	08NK016	0.3823	0.1714	**0.0870**
08JB003	**0.1153**	0.4194	0.1984	08NL007	0.7317	0.1165	**0.0538**
08KA004	1.0239	0.2309	**0.1337**	08NL024	1.1142	0.1748	**0.1109**
08KA005	3.3487	0.2863	**0.1018**	09AC001	**0.1117**	0.1811	**0.0949**
08KB001	0.6929	0.2171	**0.1183**	09AE003	0.0782	**0.0328**	**0.0390**
08LB047	1.3676	**0.0504**	0.0766	09BA001	0.1425	0.0544	**0.0106**
08LB064	1.0724	**0.1107**	**0.1201**	09BC001	0.1556	**0.0632**	**0.0574**
08LD001	1.0822	7.1434	**0.0593**	09DD003	0.1503	0.0235	**0.0051**
08ND012	**0.1193**	**0.1147**	**0.1098**	10AA001	0.2096	**0.0350**	0.0629
08ND013	0.9773	0.2158	**0.0691**	10AB001	0.2093	0.2639	**0.0756**
08NE039	1.4323	0.2490	**0.0924**	10BE004	0.1437	0.0708	**0.0435**
08NE074	1.5779	0.1933	**0.1250**	09CD001	**0.1167**	0.1825	**0.1052**
08NA002	0.9773	0.2158	**0.0691**				

Winning frameworks are bolded[1]

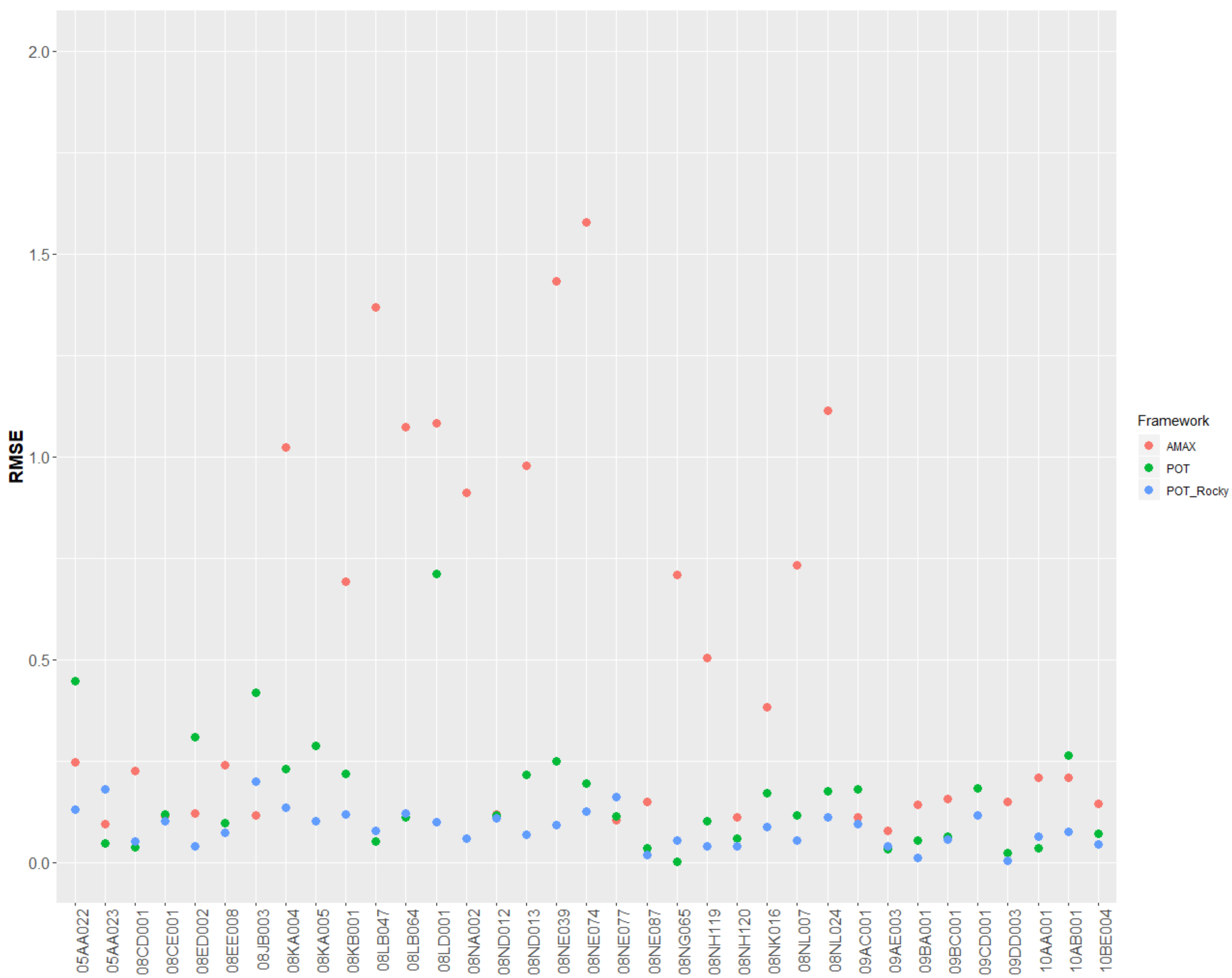

Figure 4.15: Performance measure for three RFFA frameworks at long-record sites in the Rocky Mountain region.

The best models in Table 4.5 are bolded. If the difference in the $RMSE_f$ between frameworks is less than 20%, the performance of those frameworks are considered to be the same. The POT-Rocky RFFA framework performs better than or the same as the

POT RFFA and AMAX RFFA frameworks for most of the long-record sites in the Rocky Mountain region. The POT-Rocky framework considered the elevation effect, which is an important flood-related attribute in mountainous regions. Thus, this framework improves the performance in the Rocky Mountain region and, therefore is recommended for the end product from FloodNet theme 1, the flood manual for Canada.

Chapter 5

Conclusions

5.1 General Conclusions

In this research, the performance of the AMAX RFFA framework was compared to the well-known at-site FFA framework. The width of the confidence interval for the flood quantile was used as the evaluation measure; the model producing a narrower confidence interval is the superior model. For 1114 Water Survey of Canada stations, the AMAX RFFA framework provided a narrower confidence interval for 81.69% of stations. The sites at which at-site FFA gives narrower confidence intervals were named problematic sites. These sites concentrated in two geographic regions: the Rocky Mountain region and the Canadian Prairie. Different factors were explored to find effective quantitative measures to identify problematic sites. The first was the flood-related attributes geographic proximity, lake attenuation effect and average basin slope. The second was the effect of the heterogeneous pooling group and the high discordant target site to its pooling group. The third was the effect of L-moment ratios. For hydrometric stations with flat catchments, the lake effect is an important indicator to identify problematic sites. The average slope for a catchment is effective for identifying problematic sites for the mountainous region. Finally, if either the L-CV ratio or the L-skewness for the flow series from a hydrometric station is two standard deviations lower than the regional average for its pooling group, that hydrometric station is likely to be identified as a problematic site.

The AMAX RFFA framework was compared to the POT RFFA framework for sites with more than 50-year records. The quality of the threshold value obtained by the automatic threshold selection technique was compared with the well-developed manual threshold selection method for 261 stations within the Rocky Mountain region. The automatic threshold selection method can generate a suitable threshold value close to the manually selected threshold. The average peaks per year for the POT time series was checked as an effective quantitative measure to select the better framework between AMAX RFFA and POT RFFA. For sites with peaks per year from 1 to 1.5, AMAX RFFA performs the same as POT RFFA. Overall, the POT RFFA framework performs better than the AMAX RFFA framework for 78 of 89 long record stations across Canada. The POT framework has a better performance than the AMAX framework for most of Canada, except for the Rocky Mountain region and parts of the Boreal Shield with large lake coverage.

An alternative POT framework (POT-Rocky) is proposed for the Rocky Mountain region where the Rocky Mountains were treated as the same region. Four similarity measures were used to form the pooling group: geographic proximity, elevation at the centroid of each watershed, mean annual precipitation and modified flood seasonality measure. The pooling group for 94% of the long-record sites with more than 50-year flow records in the POT-Rocky framework is homogeneous or possibly homogeneous. The performance of the POT-Rocky framework was compared with the AMAX framework and the POT framework for long-record sites within the Rocky Mountain region. The POT-Rocky framework performs better than or the same as the AMAX and POT frameworks for most of the long record sites within the Rocky Mountain region. Thus, the POT-Rocky framework is recommended as the alternative RFFA framework for the Rocky Mountain region.

5.2 Recommendations for the Flood Estimation Manual for Canada

For hydrometric stations with the following characteristics, at-site FFA is recommended to be considered and compared with the regional frameworks to select the best model:

1 The catchment of the hydrometric station has an average basin slope beyond 15°

2 The catchment of the hydrometric station has an average basin slope close to 0° and has a FARL index lower than 0.9

3 Either the L-CV ratio or the L-skewness for the flow series from a hydrometric station is two standard deviations lower than the average for the pooling group

The AMAX RFFA framework is recommended for hydrometric stations with peaks per year less than 1.5.

The POT-Rocky framework is recommended as an alternative regional framework for the Rocky Mountain region.

References

Ashkar, F., & Rousselle, J. (1987). Partial duration series modeling under the assumption of a Poissonian flood count. *Journal of Hydrology*, *90*(1-2), 135–144. doi: 10.1016/0022-1694(87)90176-4

Bačová-Mitková, V., & Onderka, M. (2010). Analysis of extreme hydrological Events on THE danube using the Peak Over Threshold method. *Journal of Hydrology and Hydromechanics*, *58*(2), 88–101. doi: 10.2478/v10098-010-0009-x

Bailey, R. G. (2004). Identifying Ecoregion Boundaries. *Environmental Management*, *34*(S1), S14-S26. doi: 10.1007/s00267-003-0163-6

Baker, V. R. (1987). Paleoflood hydrology and extraordinary flood events. *Journal of Hydrology*, *96*(1), 79–99. doi: 10.1016/0022-1694(87)90145-4

Ball, J., Babister, M., Nathan, W., R; Weeks, Weinmann, P., Retallick, M., Testoni, I., ... Testoni, I. (2016). *Australian Rainfall and Runoff: A Guide to Flood Estimation.* Geoscience Australia.

Basu, B., & Srinivas, V. V. (2015). A recursive multi-scaling approach to regional flood frequency analysis. *Journal of Hydrology*, *529*(P1), 373–383. doi: 10.1016/j.jhydrol.2015.07.037

BC Ministry of Transportation and Infrastructure. (2016). *Bridge Standards & Procedures Manual - Volume 1 Supplement to CHBDC S6-14* (Tech. Rep.). Victoria: BC Ministry of Transportation and Infrastructure. Retrieved from https://www2.gov.bc.ca/gov/content/transportation/transportation-infrastructure/engineering-standards-guidelines/

structural/standards-procedures/volume-1

Benito, G., & Thorndycraft, V. R. (2005). Palaeoflood hydrology and its role in applied hydrological sciences. *Journal of Hydrology*, *313*(1-2), 3–15. doi: 10.1016/j.jhydrol.2005.02.002

Bezak, N., Brilly, M., & Šraj, M. (2013). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, *59*(5), 959–977. doi: 10.1080/02626667.2013.831174

Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, *202*(1-4), 212–230. doi: 10.1016/S0022-1694(97)00068-1

Burn, D. H. (2003). The use of resampling for estimating confidence intervals for single site and pooled frequency analysis. *Hydrological Sciences Journal*, *48*(1), 25–38. doi: 10.1623/hysj.48.1.25.43485

Burn, D. H. (2009). An appraisal of the region of influence approach to flood frequency analysis. *Hydrological Sciences Journal*, *35*(2), 149–165. doi: 10.1080/02626669009492415

Burn, D. H., Whitfield, P. H., & Sharif, M. (2016). Identification of changes in floods and flood regimes in Canada using a peaks over threshold approach. *Hydrological Processes*, *30*(18), 3303–3314. doi: 10.1002/hyp.10861

Chang, F. J., Tsai, M. J., Tsai, W. P., & Herricks, E. E. (2008). Assessing the ecological hydrology of natural flow conditions in Taiwan. *Journal of Hydrology*, *354*(1-4), 75–89. doi: 10.1016/j.jhydrol.2008.02.022

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, *18*(3-4), 257–271. doi: 10.1016/0022-1694(73)90051-6

Davison, A. C., & Smith, R. L. (2018). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425. doi: 10.1111/j.2517-6161.1990.tb01796.x

Dikbas, F., Firat, M., Koc, A. C., & Gungor, M. (2013). Defining Homogeneous Regions for Streamflow Processes in Turkey Using a K-Means Clustering Method. *Arabian Journal for Science and Engineering*, *38*(6), 1313–1319. doi: 10.1007/s13369-013-0542-0

Dumanski, S., Pomeroy, J. W., & Westbrook, C. J. (2015). Hydrological regime changes in a Canadian Prairie basin. *Hydrological Processes*, *29*(18), 3893–3904. doi: 10.1002/hyp.10567

Eccles, K. M., Checkley, S., Sjogren, D., Barkema, H. W., & Bertazzon, S. (2017). Lessons learned from the 2013 Calgary flood: Assessing risk of drinking water well contamination. *Applied Geography*, *80*, 78–85. doi: 10.1016/j.apgeog.2017.02.005

Faulkner, D., Warren, S., & Burn, D. (2016). Design floods for all of Canada. *Canadian Water Resources Journal*, *41*(3), 398–411. doi: 10.1080/07011784.2016.1141665

Gado, T. A., & Nguyen, V.-t.-v. (2016). Comparison of Homogenous Region Delineation Approaches for Regional Flood Frequency Analysis at Ungauged Sites. *Journal of Hydrologic Engineering*, *21*(3), 1–10. doi: 10.1061/(ASCE)HE.1943-5584.0001312.

Government of Canada. (2013). *Costs of flooding.* Retrieved from https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/costs-of-flooding.html#AB

Grover, P. L., Burn, D. H., & Cunderlik, J. M. (2002). A comparison of index flood estimation procedures for ungauged catchments. *Canadian Journal of Civil Engineering*, *29*(5), 734–741. doi: 10.1139/l02-065

Haddad, K., Johnson, F., Rahman, A., Green, J., & Kuczera, G. (2015). Comparing three methods to form regions for design rainfall statistics: Two case studies in Australia. *Journal of Hydrology*, *527*, 62–76. doi: 10.1016/j.jhydrol.2015.04.043

Haddad, K., & Rahman, A. (2011). Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. *Stochastic Environmental Research and Risk Assessment*, *25*(3), 415–428. doi: 10.1007/s00477-010-0412-1

Haddad, K., Rahman, A., & Ling, F. (2015). Regional flood frequency analysis method for Tasmania, Australia: A case study on the comparison of fixed region and region-of-influence approaches. *Hydrological Sciences Journal*, *60*(12), 2086–2101. doi: 10.1080/02626667.2014.950583

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, *29*(2), 271–281. doi: 10.1029/92WR01980

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis : an approach based on L-moments*. New York: Cambridge University Press.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*, *52*(10), 7548–7555. doi: 10.1002/2016WR019285

Jarrett, R. D., & England, J. F. (2011). Reliability of Paleostage Indicators for Paleoflood Studies. In P. Kyle House., Robert H. Webb., Victor R. Baker., & Daniel R. Levish.

(Eds.), *Ancient floods, modern hazards: Principles and applications of paleoflood hydrology, volume 5* (pp. 91–109). Washington DC: American Geophysical Union. doi: 10.1029/ws005p0091

John F. England Jr., Timothy A. Cohn., Beth A. Faber., Jery R. Stedinger., Wilbert O. Thomas Jr., Andrea G. Veilleux., ... Robert R. Mason, J. (2019). *Guidelines for determining flood flow frequencyBulletin 17C* (Tech. Rep.). Reston: U.S. Geological Survey. doi: 10.3133/tm4B5

Khaliq, M., Ouarda, T., Gachon, P., Sushama, L., & St-Hilaire, A. (2009). Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. *Journal of Hydrology*, *368*(1-4), 117–130. doi: 10.1016/J.JHYDROL.2009.01.035

Laio, F. (2004). Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*, *40*(9), 3497–3502. doi: 10.1029/2004WR003204

Lance, G. N., & Williams, W. T. (2015). Computer programs for hierarchical polythetic classification. *Journal of Software (Oulu)*, *10*(1), 1–1. doi: 10.1115/1.859971

Lang, M., Ouarda, T. B., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, *225*(3-4), 103–117. doi: 10.1016/S0022-1694(99)00167-5

Lilienthal, J., Fried, R., & Schumann, A. (2018). Homogeneity testing for skewed and cross-correlated data in regional flood frequency analysis. *Journal of Hydrology*, *556*, 557–571. doi: 10.1016/j.jhydrol.2017.10.056

Madsen, H., Pearson, C. P., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. *Water Resources Research*, *33*(4), 759–769. doi: 10.1029/96WR03849

Mami Mizutori., & Debarati Guha-Sapir. (1998). *Economic Losses,Poverty & DISASTERS 1998-2017* (Tech. Rep.). Brussels: UN Office for Disaster Risk Reduction. Retrieved from www.unisdr.org

Mediero, L., Kjeldsen, T. R., Macdonald, N., Kohnova, S., Merz, B., Vorogushyn, S., . . . órarinsson, . (2015). Identification of coherent flood regions across Europe by using the longest streamflow records. *Journal of Hydrology*, *528*, 341–360. doi: 10.1016/j.jhydrol.2015.06.016

Merz, R., & Blöschl, G. (2005). Flood frequency regionalisation - Spatial proximity vs. catchment attributes. *Journal of Hydrology*, *302*(1-4), 283–306. doi: 10.1016/j.jhydrol.2004.07.018

Micevski, T., & Kuczera, G. (2009). Combining site and regional flood information using a Bayesian Monte Carlo approach. *Water Resources Research*, *45*(4). doi: 10.1029/2008WR007173

Molotch, N. P., Barnard, D. M., Burns, S. P., & Painter, T. H. (2014). Measuring spatiotemporal variation in snow optical grain size under a subalpine forest canopy using contact spectroscopy. *Water Resources Research*, *52*(9), 7513–7522. doi: 10.1002/2016WR018954.Received

Mostofi Zadeh, S., & Burn, D. H. (2019). A Super Region Approach to Improve Pooled Flood Frequency Analysis. *Canadian Water Resources Journal*, *44*(2), 146–159. doi: 10.1080/07011784.2018.1548946

Mostofi Zadeh, S., Durocher, M., Burn, D. H., & Ashkar, F. (2019). Pooled flood frequency analysis: a comparison based on peaks-over-threshold and annual maximum series. *Hydrological Sciences Journal*, *64*(2), 121–136. doi: 10.1080/02626667.2019.1577556

Moudrak, N., Feltmate, B., Venema, H., & Osman, H. (2018). *Combating Canadas Rising Flood Costs: Natural infrastructure is an underutilized option* (Tech. Rep.). Insurance Bureau of Canada, Intact Centre on Climate Adaptation, University of Waterloo. Retrieved from http://www.ibc.ca/on/resources/studies/natural-infrastructure-is-an-underutilized-option

MTC. (1979). *M.T.C. Design Flood Criteria* (Tech. Rep.). Ministry of Transportation and Communications. Retrieved from http://www.mto.gov.on.ca/english/publications/pdfs/drainage/ministry-directive-b100.pdf

NSERC FloodNet. (n.d.). *FloodNet - NSERC Network - Enhanced flood forecasting and management capacity in Canada.* Retrieved from http://www.nsercfloodnet.ca/

Önöz, B., & Bayazit, M. (2012). Block bootstrap for Mann-Kendall trend test of serially dependent data. *Hydrological Processes*, *26*(23), 3552–3560. doi: 10.1002/hyp.8438

OBrien, N. L., & Burn, D. H. (2014). A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. *Journal of Hydrology*, *519*, 2040–2048. doi: 10.1016/j.jhydrol.2014.09.041

Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, *28*(2), 126. doi: 10.2307/2346729

Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., ... Auffray, A. (2013). Data-based comparison of frequency analysis methods: A general framework. *Water Resources Research*, *49*(2), 825–843. doi: 10.1002/wrcr.20087

Salinas, J. L., Castellarin, A., Kohnová, S., & Kjeldsen, T. R. (2014). Regional parent flood frequency distributions in Europe - Part 2: Climate and scale controls. *Hydrology and Earth System Sciences*, *18*(11), 4391–4401. doi: 10.5194/hess-18-4391-2014

Singh, L. P., Guo, V. P., Fang, S.-L., & Liu, B. (2013). A new method for identification of flood seasons using directional statistics. A new method for identification of flood seasons using directional statistics. *Hydrological Sciences Journal-Journal des Sciences Hydrologiques*, *58*(1), 28–40. doi: 10.1080/02626667.2012.743661

Solari, S., & Losada, M. A. (2012). A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resources Research*, *48*(10), 1–15. doi: 10.1029/2011WR011475

Tan, X., Gan, T. Y., Tan, X., & Gan, T. Y. (2015). Nonstationary Analysis of Annual Maximum Streamflow of Canada. *Journal of Climate*, *28*(5), 1788–1805. doi: 10.1175/JCLI-D-14-00538.1

Tasker, G. D. (1982). COMPARING METHODS OF HYDROLOGIC REGIONALIZATION. *Journal of the American Water Resources Association*, *18*(6), 965–970. doi: 10.1111/j.1752-1688.1982.tb00102.x

Tate Dalrymple. (1960). *Flood-frequency analyses, Manual of Hydrology: Part 3* (Tech. Rep.). WASHINGTON DC: UNITED STATES DEPARTMENT OF THE INTERIOR. doi: 10.3133/wsp1543A

Thomas R. Kjeldsen., David A. Jones., & Adrian C. Bayliss. (2008). *Improving the FEH statistical procedures for flood frequency estimation*. Bristol: Environment Agency.

Timothy A. Cohn., Nancy A. Barth., John F. England Jr., Beth A. Faber., Robert R. Mason, J., & Jery R. Stedinger. (2019). *Evaluation of Recommended Revisions* *to Bulletin 17B* (Tech. Rep.). Reston, VA: U.S. Geological Survey. doi: 10.3133/ofr20171064

Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, *43*(3). doi: 10.1029/2006WR005095

W. B. Langbein. (1949). Annual floods and the partialduration flood series. *American Geophysical Union*, *30*(6), 879–881. doi: 10.1029/TR030i006p00879

Wang, Q. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*, *129*(1-4), 263–280. doi: 10.1016/0022-1694(91)90054-L

Wang, W., Chen, Y., Becker, S., & Liu, B. (2015). Variance Correction Prewhitening Method for Trend Detection in Autocorrelated Data. *Journal of Hydrologic Engineering*, *20*(12), 04015033. doi: 10.1061/(ASCE)HE.1943-5584.0001234

Zaman, M. A., Rahman, A., & Haddad, K. (2012). Regional flood frequency analysis in arid regions: A case study for Australia. *Journal of Hydrology*, *475*, 74–83. doi: 10.1016/j.jhydrol.2012.08.054