Investigation of Flood Frequency Analysis across Canada

By

Ziyang Zhang

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Civil Engineering

University of Manitoba

Winnipeg, Manitoba, Canada

December 2018

Abstract

The National Science and Engineering Research Council of Canada (NSERC) funded FloodNet Strategic Network Project is currently developing a flood estimation manual for standardized use of flood frequency analysis (FFA) in Canada. This thesis is a part of that project and emphasizes two important research topics related to FFA. The first is to identify a preferred statistical distribution for at-site FFA in Canada. Annual maximum and instantaneous peak flow data across Canada are used to fit four popular candidate distributions, and then the fits are assessed based on goodness-of-fit analyses. Results indicate the Generalized Extreme Value distribution is most preferred for at-site FFA across Canada.

As homogeneous flood regions are an important component for the index flood regional FFA, the second contribution investigates five distinct flood-related attributes for their relevance in developing homogeneous regions Canada-wide. The relevance of each attribute is assessed based on the number of homogenous regions produced and the accuracy of regional flood estimation. Amongst the five study attributes, the attribute of precipitation pattern is preferred for general use in Canada. The attributes of flood seasonality, physiographic variables, temperature pattern display good potential to be used in selected regions. Findings from this research are considered valuable input to the development of a Canadian flood estimation manual by the FloodNet Strategic Network.

Acknowledgements

First and foremost, I would like to thank my advisor Dr. Trish Stadnyk, for inspiring me to swim in the meandering creek (undergraduate hydrology class), for challenging me to boat through rugged and uneven river valley (watershed processes class), and for guiding me to sail on the ocean of research (hydrological research). I am deeply grateful for her tremendous support and positive encouragement throughout my student career and my personal ups and downs.

I would like to dedicate this study to the memory of Professor Dr. Peter Rasmussen, who had been my advisor since undergraduate thesis till halfway of this master's thesis. I truly miss the days that I had Peter in his class, in his office, and in his place. Though those days were short and memories are unstoppably getting lost, I try to hold those memories with every pieces to last as long as possible for the rest of my life. I would like to acknowledge the contribution of Dr. Peter Rasmussen that this research journey could not be made without him.

Sincere thanks is given to Dr. Donald Burn – professor at the University of Waterloo. Dr. Burn has provided very valuable expertise in my research and in edits of manuscripts. I truly value the knowledge and skills that I have learned from his counsel.

I would like to express my gratitude to the committee members – Dr. Masoud Asadzadeh and Dr. Shirley Thompson for providing valuable comments.

I would like to express my appreciation to the funding organizations, including NSERC FloodNet and the University of Manitoba. Your monetary support has made this research come true.

I am grateful for the assistance of Erika Klyszejko at Environment and Climate Change Canada for the support of research data.

I also would like to thank Ameer Muhammad and Micha Roemer as my best friends in school. I am very thankful that our paths have crossed and we have always been super supportive to each other in both life and work. I am grateful for the friendship and memories.

A special appreciation is giving to my loving wife Nan Zheng, who supports me throughout all stages of my graduate study. I am incredibly thankful for the love, the kindness, the understanding, and the patience she has given to me.

Lastly, I would like to thank my beloved family members, including my parents Chao Zhang and Cheng Li, my grandparents Yuexi Zhang, Aihua Sun, Wengao Li, and Ying Chen, and my relatives Yong Zhang, Lina Liu, Jinwei Zhang, Jack Chen, and Lily Li. Many of them live half world apart and do not understand the English I put here, but my work could not be made without their support. Especially, I would like to thank my parents for their unstinting love so that I can have a fantastic life in Canada.

Table of Contents

List of Figures

List of Tables

Contributions of Co-Authors

This thesis is immensely aided by of Dr. Tricia Stadnyk, Dr. Peter Rasmussen, and Dr. Donald Burn. However, it should be stated that overall contents are made of my own work.

Chapter 3

Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. **(Zhang, Z., Stadnyk, T., and Burn, D.)**

Dr. Tricia Stadnyk provided lots of guidance on writing this paper, including the definition of research objective, framework of paper structure, as well as discussion of figures and tables. She also carried out edits of the manuscript before submission to the journal. Dr. Donald Burn provided guidance on particularly the technical component of this paper. He provided valuable forefront expertise to elevate this study, including the selection of research methods and methods for the presentation of results. In addition, he assisted the editing of the paper before submission to the journal. The paper is currently under revision with the Canadian Water Resources Journal.

Chapter 4

Investigation of attributes that use for developing homogeneous flood regions for regional flood frequency analysis in Canada. **(Zhang, Z., and Stadnyk, T.)**

Dr. Tricia Stadnyk provided guidance for the completion of this research. She provided valuable inputs in both writing and technical parts of this study. She posed challenging questions to improve the value of this study. She provided edits to the manuscript before submission to the journal.

Chapter 1 Introduction

1.1 Study background and motivation

The ability to estimate flooding potential is of paramount importance for engineering and hydrological applications, particularly related to the design of hydraulic structures and proactive management of flood risk. Improving our capacity to estimate flooding has become increasingly important for Canada – a water rich country that is increasingly more impacted by flooding at a variety of magnitudes and scales in recent decades (Burn and Whitfield 2016; Burn et al. 2016).

Flood Frequency Analysis (FFA) has been widely recognized as an effective method to estimate flooding potential in terms of flood quantile at different return periods. FFA considers using a statistical distribution to model historical observed flood data, and thus, the flood quantile of a given return period (i.e., probability of occurrence) can be estimated from the statistical distribution.

Proactive comprehensive research on the specific method or application of FFA, which is tailored specifically for an area of interest, can considerably improve the accuracy of quantile estimation. In fact, a number of countries have conducted nation-wide research with the express purpose of producing a national standard for flood estimation, including flood estimation manuals as guidance for operational use (U.S. Water Resources Council 1982; Robson and Reed 1999; Salinas et al. 2014b; Ball et al. 2016; England Jr. et al. 2018). These published manuals provide recommended guidance on detailed methodology and techniques concerned with flood quantile estimation using FFA.

In Canada, little guidance for FFA can be found, and the methods of application are therefore used inconsistently across the nation (Cunnane 1989; Watt 1989; Pilon and Adamowski

1992; Aucoin et al. 2011; Faulkner et al. 2016; Lu 2016). This is partly because Canada is a sizable country and methods related to flood management have been traditionally directed at the provincial level (Jakob and Church 2011; Faulkner et al. 2016). To address some of these inconsistency issues, the Natural Sciences and Engineering Research Council of Canada (NSERC) funded the FloodNet Strategic Network (FloodNet 2018). FloodNet is a nation-wide research network with more than a hundred contributors from a dozen Canadian universities, various levels of government agencies, and many industrial partners. It was designed to provide operational tools for flood management and forecasting for Canadian use. FloodNet Theme 1-6 was tasked with undertaking the development of a standardized flood estimation manual and guideline for Canada. Similar to flood estimation manuals in other countries, such a manual would be designed to provide rigorous guidance on methods, techniques, and procedures for FFA, and would be tailored specifically to Canadian hydrological conditions and watersheds. This would be a valuable asset, particularly for practitioners who work on flood management, water resources infrastructure, or hydraulic engineering projects.

This research is affiliated with FloodNet Theme 1-6 and provides insight into the application of FFA across Canada, contributing toward the development of a consistent Canadian guideline. Specifically, the

1. Selection of an adequate statistical distribution for modelling historical observed flood data (Cunnane 1989; Bobée and Rasmussen 1995). A common selection approach in operational practice is to construct the flood frequency plot and visually inspect goodness-of-fit between the distribution's modelled curve and the observed data at a particular site of interest; however, this method is substantially subjective and lacks statistical reasoning.

2. Absence of long-term historical flood data. In classic at-site FFA, the record length of the at-site gauging data needs to exceed the return period of interest for reasonable flood estimation (Robson and Reed 1999). For desirable estimation, the record length should be double the return period of interest. For example, in estimating a 100-year design flood (typically used for major highway bridges and culvert crossings in Canada; (Watt 1989; Moudrak and Feltmate 2017), a reasonable record length is 100 years, and the desirable record length is 200 years. In Canada, fewer than 1% of hydrometric stations (i.e., 51 out of 6329) have record lengths achieving 100 years or more (Environment Canada Data Explorer - HYDAT, accessed July, 2018), meaning that at-site FFA is generally not applicable in Canada for design flood estimation. Alternatively, regional FFA provides a means to deal with this problem by "trading space for time" (Hosking and Wallis 1997). Regional FFA expands the record length by the development of a homogeneous region, which comprises of a set of hydrologically similar sites. One fundamental problem that needs to be addressed for regional FFA is the development of such homogeneous regions (Bobée and Rasmussen 1995), which is typically solved using similarity measures of flood-related attributes, which are poorly documented and researched in Canada.

1.1 Research objective and scope

The over-arching goal of this research is to provide quantitative measures of applying FFA across Canada for the purposes of contributing towards a Canadian flood estimation manual. There are two specific objectives defined to improve FFA within Canada:

1. Identify a preferred flood frequency distribution for at-site FFA in Canada based on the goodness-of-fit analysis.

2. Investigate different flood-related attributes, and to develop techniques that assist with the delineation of homogenous flood regions for Canada.

Findings from this research provide an important basis for the development of a Canadian flood estimation manual by the FloodNet Strategic Network.

1.2 Thesis outline

This thesis is comprised of six chapters.

Chapter 1: the introduction of the thesis, including the introduction of the affiliated project - NSERC Canadian FloodNet, and the over-arching goal as well as the specific objectives of the thesis.

Chapter 2: provides theory related to at-site and regional FFA. It also describes some of the key steps related to distribution selection and development of homogeneous flood regions.

Chapter 3: the first manuscript that addresses the first objective of the thesis. This manuscript has been submitted to the Canadian Water Resources Journal:

Chapter 4: the second manuscript that addresses the second objective of this thesis. This manuscript is currently in preparation for submission to an academic journal.

Chapter 5: a summary of thesis research which outlines major conclusions and recommendations for future work.

Chapter 2 Theory and relevant literature

This chapter provides theory and relevant literature related to some of the key steps in our research work. Section 2.1 briefly describes the concept of at-site FFA. Section 2.2 introduces commonly used flood frequency distributions as well as parameter estimation method in current status of research. Section 2.3 reviews the goodness-of-fit tests that have been commonly used to support the selection of flood frequency distribution. Sections 2.1, 2.2, and 2.3 provide the background to support the first objective of this thesis. Similarly, Sections 2.4 and 2.5 provide the background supporting the second objective of the thesis by introducing the basis of regional FFA, with a specific focus on the index flood method (Section 2.4); and summarizing commonly used flood-related attributes used to support the development of homogeneous flood regions (Section 2.5).

2.1 At-site Flood Frequency Analysis

Estimating the flood potential of future events at a site is required for a variety of engineering and hydrological works. For example, the (1) design of hydraulic structures such as culverts, dams, and spillways where the size of the structure must be adequate to pass or with stand flood level flows without excessive cost due to over design; and, (2) floodplain analysis for urban planning, flood mitigation, insurance assessment, and other hydrologic research.

For the site where gauged flow records (i.e., flood data) are available, at-site FFA is one of the most direct, simple, and reasonable methods to estimate flooding potential. At-site FFA is a statistical method used to estimate flood quantile at a given site based on historical flood data. The method is not subject to forecasting the timing and magnitude of a specific flooding event, instead, it aims to estimate the expected flood magnitude as specified by the frequency of occurrence. Statistically, the expected flood magnitude is denoted as a flood quantile (Q_T), with the frequency of occurrence specified as return period (T). The return period can be converted to the probability of non-exceedance (P) using, $P = 1 - T^{-1}$. For instance, if the 100-year flood quantile at a site is 80 m^3/s, then the chance that an 80 m^3/s flood is expected to occur in any year is 1% (i.e., 0.99 non-exceedance probability).

The true population of flood frequency at a site is always unknown. However, this true population of flood frequency can be estimated by statistical modelling based on historical flood data. The historical flood data, a statistical distribution, and a parameter estimation method are the basic requirements for at-site FFA. Historical flood data are applied to the parameter estimation method to estimate parameters of the statistical distribution. This process is also referred to as fitting the distribution to the flood data. The distribution with estimated parameters is termed the sample distribution that can be used to model the flood sample and to extrapolate flood quantiles for different return periods. More details of the theory of flood frequency analysis can be found in Cunnane (1989), Hosking and Wallis (1997), Robson and Reed (1999), and Rao and Hamed (2000).

The most used historical flood data is the flood sample of annual maximum flow (AMF), which comprises a set of annual maximum observed flows obtained from a gauging site. The size of the flood sample (i.e., the number of observed values) is denoted as record length. For at-site FFA, the record length has to be sizable as the sample needs to be large enough to reflect the true population. Robson and Reed (1999, p46) have stated that a minimum AMF record length for at-site FFA is 14-station-years. For reasonable estimation, the record length of the at-site gauged data needs to exceed the return period of interest. For desirable estimation, the record length should be double the return period of interest (Robson and Reed 1999, p46). In addition, the flood values recorded must be independent and identically distributed (i.e. IID requirement) (Rao and Srinivas 2006). Independence requires that recorded flood data are generated from independent flooding events. Identical distribution means that observed values are stationarity with time, such that there are no significant trends in magnitude over time. The IID criteria are used to ensure that flood data have an identical population of flood frequency.

2.2 Distributions for FFA and parameter estimation

The statistical distribution for FFA is a continuous cumulative distribution function (CDF) (i.e., not discrete distributions like Poisson or Binomial), which define the relationship between flood quantile (Q_T) and the frequency of occurrence presented as return period (T), or probability of non-exceedance (P). Distributions used in modelling flood events typically contain two to four parameters. Defining the values of the parameters is a way to adjust the scale and shape of the distribution. For flood frequency analysis, parameters are estimated through historical flood data via a parameter estimation method. Classic distributions for FFA analysis include 2-parameter Gumbel (EV1), Frechet (EV2), and log-normal (LN2), as well as 3-parameter Generalized Pareto (GPA), Generalized Extreme Value (GEV), and log Pearson Type III (LP3). Conventional parameter estimation methods include method of product moments, maximum likelihood method, and the L-Moments method.

Using different distributions in FFA results in different quantile estimations, therefore, the choice of a suitable flood frequency distribution is relevant for design flood estimation. A milestone survey of frequently used distributions for FFA was conducted by Cunnane (1989). This survey, published in the 1989 Operational Hydrology Report of the World Meteorological Organization, revealed that 2-parameter EV1, EV2, Weibull (EV3), gamma (G2), LN2, as well as 3-parameters GEV, Pearson Type III (PE3), LP3, and gamma (G3) distributions have been extensively used by 54 government agencies in 28 countries. The 2-parameter EV1 distribution has been the most used distribution amongst the survey recipients. Since the 1990s, however, flood frequency distributions using three parameters have become increasingly popular relative to 2-parameter distributions (Karim and Chowdhury 1993; Vogel et al. 1993; Önöz and Bayazit 1995). Bobée et al. (1993) stated that flood frequency distributions should have at least three parameters to sufficiently modeling the wide variety of flood data. Vogel and Wilson (1996) and Salinas et al. (2014b) have indicated that the GEV distribution has become the most used flood frequency distribution across the world. The GEV distribution is the most acceptable partly because its mathematical form (i.e., CDF) covers mathematical formulation of the EV1, EV2, and EV3 distributions, which have been historically popular (Cunnane 1989). In addition, the GEV distribution has a strong theoretical justification for flood frequency modeling (Coles 2001). Based on the Extreme Value Theory, the distribution of independent extreme events should resemble the GEV distribution family (Coles 2001). Another commonly used distribution in current practice is the LP3 distribution, potentially because it is the default distribution used in the U.S. over the past several decades (U.S. Water Resources Council 1982). It is also one of the default distributions used in Australia (Ball et al. 2016).

In terms of parameter estimation methods, the L-Moment estimators (Hosking 1990) have replaced the classic methods (i.e., product moments and maximum likelihood) for their simplicity, robustness, and the nearly unbiased estimation (Vogel and Fennessey 1993; Hosking and Wallis 1997). Thus, the L-Moment method can be treated as the default parameter estimation method for nearly all distributions. The only exception appears to be the LP3 distribution, where parameters are normally estimated based on product moments with a regional weighted skewness method, in the same manner as the in the U.S. (U.S. Water Resources Council 1982; England Jr. et al. 2018).

In Canada, flood frequency distributions are used inconsistently across the country (Watt 1989), however, studies related to a preferred distribution for FFA have been conducted in selected geographical regions. Fortin et al. (1997) has demonstrated a rational approach of using nonparametric Bayesian simulation to compare a set of distributions based on flood flow data at seven stations in the Quebec and Ontario regions. It has been reported that the use of 2-parameter normal, EV1, and G2 distributions have been generally preferable to those of 3-parameter GEV, PE3, LP3, and G3. For flood data with low coefficients of variation and positive skewness, the GEV, PE3, and LP3 have been preferred. Yue and Wang (2004) identified appropriate flood distributions for ten climatic regions of Canada based on L-Moment techniques. They compared ten distributions and revealed that GEV, 3-parameter log Normal (LN3), PE3, and LP3 have been preferable for most regions. Aucoin et al. (2011) applied GEV and LN3 distributions for at-site flood frequency analyses at 56 hydrometric stations in and around the Province of New Brunswick. Three goodness-of-fit analyses, including negative log-likelihood value, AD test, and flood frequency plot, have been used to evaluate the two distributions at each station. Results revealed that the majority of stations accepted the fit of both GEV and LN3 distributions, with slightly more favouring the GEV distribution. Lu (2016) utilized L-Moment techniques to explore the best-fit regional flood frequency distribution for Labrador and Newfoundland. It was found that the use of GEV was preferred for Labrador, while the use of LN3 was favoured for Newfoundland.

The above studies have provided insights into some geographical regions in Canada. Some studies have been funded by the government agencies, indicating the pressing need for guidance on suitable flood frequency distributions. Yet, no definitive conclusion regarding a preferred distribution for Canada has yet been made.

2.3 Goodness-of-fit analysis

The first objective of this study is to identify the preferred flood frequency distribution in Canada for at-site FFA. The goodness-of-fit analysis is the most standard method to assess whether the sample distribution is accepted or rejected for frequency modelling of a given flood sample (Ahmad et al. 1988a; Chowdhury et al. 1991; Heo et al. 2013). In terms of selecting a preferred distribution for a sizeable region, since a universal distribution that can adequately fit all data for every station does not exist, a preferred distribution is generally the one that is acceptable for most stations (Önöz and Bayazit 1995; Robson and Reed 1999).

In the past, goodness-of-fit analysis based on empirical CDF (i.e., empirical goodness-of-fit test) has played an important role in evaluating the fit of a flood frequency distribution (Cunnane 1985; Cunnane 1989). The empirical CDF is non-parametric and plotted as a step function in the CDF plot. According to Ahmad et al. (1988a), for a given AMF series comprised of n years observed record, let $x_1 \leq x_2 \leq \cdots \leq x_n$ denote the ordered values. The empirical CDF $F_n(x)$ for a given variable x is defined as

$$F_n(x) = 0 \quad if \quad x < x_1 \tag{1}$$

$$F_n(x) = \frac{i}{n} \quad if \quad x_i \leq x < x_{i+1} \tag{2}$$

$$F_n(x) = 1 \quad if \quad x_n \leq x \tag{3}$$

where i is an integer ranged from 1 to $n-1$.

Let $F(x)$ denotes the CDF of the sample distribution estimated based on observed flood data. The statistic assessing the different between $F_n(x)$ and $F(x)$ is the statistic of an empirical goodness-of-fit test.

Kolmogorov-Smirnov (KS) test, chi-square (CS) test, Cramer von Mises (CVM) test and Anderson-Darling (AD) tests have been the popular empirical goodness-of-fit tests in the past, and have been widely used in applications of selecting distributions for FFA (Ahmad et al. 1988a; Chowdhury et al. 1991; Haktanir 1991; Mutua 1994; Önöz and Bayazit 1995; Stedinger and Lu 1995). Statistics for empirical goodness-of-fit tests can be as simple as the Kolmogorov-Smirnov test, which takes the largest distance between $F_n(x)$ and $F(x)$, denoted as

$$D = \underset{1 \leq i \leq n}{max} |F_n(x) - F(x)| \tag{4}$$

Or, as complex as assigning proper weighting function to data in the two tails of the CDF, such as the AD test, denoted as

$$A_n^2 = \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 \big[F(x)\big(1 - F(x)\big)\big]^{-1} dF(x) \tag{5}$$

The power of some of the empirical goodness-of-fit tests, however, is weak (e.g., K-S test (Önöz and Bayazit 1995), and it is common that several empirical goodness-of-fit tests are employed together to draw a more broad picture in a goodness-of-fit analysis (Chowdhury et al. 1991; Karim and Chowdhury 1995; Önöz and Bayazit 1995; Rahman et al. 2013).

Ahmad et al. (1988a) has justified that FFA should only interest quantiles in the upper tail of the CDF, therefore, he has proposed a modified Anderson-Darling (MAD) test, which assigns more weights the discrepancies between $F_n(x)$ and $F(x)$ on the upper tail. Studies of the MAD test for GLO and GEV distributions in a wide range of scale parameters can be found in Shin et al. (2012), and Heo et al. (2013). The MAD test is employed in the first research objective of this thesis, with further details of the MAD test presented in Chapter 3.

Since the development of the L-Moments technique in the 1990s (Hosking 1990; Hosking and Wallis 1993; Hosking and Wallis 1997), the use of L-Moments goodness-of-fit analyses (i.e., L-Moment ratio diagram and L-Moment Z^{DIST}) have gradually replaced empirical goodness-of-fit tests for distribution selection in FFA (Pearson 1991; Pilon and Adamowski 1992; Vogel and Wilson 1996; Peel et al. 2001; Yue and Wang 2004; Lu 2016). In particular, the UK flood estimation manual (Robson and Reed 1999) used the L-Moment goodness-of-fit analysis alone to identify the preferred distribution for UK practitioners.

The L-Moments goodness-of-fit analysis has a strong statistical reasoning for assessing the fit of a sample distribution. The L-Moment statistic constitutes L-Moment ratios of mean, L-CV, L-Skewness, and L-Kurtosis. For a given flood sample, sample L-Moment ratios are unbiased statistics of the flood sample (Vogel and Fennessey 1993). For given 2- or 3-parameter distributions, there is a fixed relationship between distribution's L-Skewness and L-Kurtosis (Hosking and Wallis 1997). When fitting the distribution, such as a 3-parameter distribution to the flood sample, the location, scale, and shape parameters can be estimated to match the mean, L-CV, and L-Skewness of the flood data. This leaves the L-Kurtosis as a checking variable, where a smaller difference between the sample L-Kurtosis and distribution's L-Kurtosis indicates a better goodness-of-fit. The L-Moment ratios diagram is a simple visual tool to identify which distributions fit better (i.e. smaller difference in L-Kurtosis). For giving a definitive result as to whether the fitting distribution should be rejected or not, however, L-Moment Z^{DIST} provides the best avenue on the basis of Monte-Carlo simulation.

Based on Hosking and Wallis (1997), procedures to estimate L-Moment statistics for a flood sample (i.e., sample L-Moments ratios) are presented as follows. Let $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ be the ordered sample and n is the sample size. First, the estimator of probability weighted moment β_r is calculated as:

$$\beta_r = n^{-1} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_{j:n} \tag{6}$$

Then, the sample L-Moments are estimated:

$$\ell_1 = b_0 \tag{7}$$

$$\ell_2 = 2b_1 - b_0 \tag{8}$$

$$\ell_3 = 6b_2 - 6b_1 + b_0 \tag{9}$$

$$\ell_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \tag{10}$$

Last, the L-Moment statistics (i.e., L-Moment ratios) of a sample are defined:

Mean: ℓ_1 (11)

L-CV: $t = \frac{\ell_2}{\ell_1}$ (12)

L-Skewness: $t_3 = \frac{\ell_3}{\ell_2}$ (13)

L-Kurtosis: $t_4 = \frac{\ell_4}{\ell_2}$ (14)

The L-Moment Z^{DIST} test is employed in the first objective of this thesis, and details of the L-Moment Z^{DIST} test are presented in Chapter 3.

2.4 Regional FFA and the index flood method

The biggest problem for at-site FFA is that it is only applicable for sites with sizeable record lengths. Once the record length is limited, or the return period of interest is large, the accuracy of at-site estimation is often questionable. The method of regional FFA is considered an enhanced version of at-site FFA, and is built to address the issues of short records and/or larger return periods. Regional FFA considers "trading space for time" (Hosking and Wallis 1997), obtaining flood characteristics from hydrological similar sites to refine the flood frequency estimation of the target site.

The regional weighted skewness method used particularly for the LP3 distribution is one classic approach for regional FFA. This method takes flood data from geographically contiguous sites to improve the distribution through a weighted adjustment of the skewness parameter. A Monte-Carlo experiment applied in Griffis and Stedinger (2007) demonstrated the adequacy of the regional weighted skewness method used with the LP3 distribution.

The most popular method for regional FFA is the index flood method (Dalrymple 1960). Hosking et al. (1985) and Lettenmaier and Potter (1985) documented the index flood method for its accuracy in quantile estimation. Wallis and Wood (1985) and Potter and Lettenmaier (1990) demonstrated the index flood method is superior to the regional weighted skewness method for predicting flood quantiles. The UK also considers the index flood method as the default method for UK flood estimation (Robson and Reed 1999). In the ongoing development of Canadian flood estimation manual, the FloodNet Theme 1-6 will develop methods and techniques for regional FFA primarily based on the index flood method.

This index flood method requires a developed homogeneous flood region, which constitutes a set of sites of similar flood characteristics. The index flood method explicitly assumes that flood samples taken from a flood region are considered to have an identical flood frequency distribution, with deviations only in terms of a scaling factor (the index event). Thus estimated flood quantiles at each site can be calculated by using the identical distribution multiplied by the scaling factor.

The development of L-Moment techniques (Hosking 1990; Hosking and Wallis 1993; Hosking and Wallis 1997) greatly assist in solving several key procedures related to the index flood method. As previously mentioned, L-Moment statistics have been largely used to identify appropriate flood frequency distributions, and to estimate distribution parameters. Also, the L-Moment homogeneity test has been intensively used to assess the homogeneity of a flood region (Zrinji and Burn 1994; Fill and Stedinger 1995; Burn 1997; Yang et al. 2010).

2.5 Flood-related attributes for the development of homogeneous flood regions

The L-Moment technique has standardized some key procedures of the index flood method, however, there is no general consensus within the hydrological community on how to best develop homogeneous flood regions. Determining a preferred method for Canada remains a key research priority, and is being addressed by FloodNet Theme 1-6. The developed methods should be easy to use and tailored specifically for Canadian hydrological conditions.

The development of homogeneous regions is typically aided by similarity measures of a flood-related attribute. The attribute should be relevant to the flood regime, and readily obtainable for the watershed (Burn 1989). The attribute of each site is presented in a number so that similarity measures of two sites can be carried out by calculating the distance between values. Sites with smaller distances are considered as having similar flood regimes, and thus would be more likely to form a homogeneous region.

Commonly used flood-related attributes are generally classified into four types. The first type is based on geographical proximity. The rationale for this is that sites that are geographically contiguous generally encompass similar physiographical and meteorological characteristics. Therefore, these sites are more likely to exhibit similar flood regimes and thus, able to form a homogeneous region. Geographical proximity is advantaged by convenience in practice (Natural Environment Research Council 1975; U.S. Water Resources Council 1982). However, for regions where gauging sites are scattered and sparsely distributed, geographical proximity might be ineffective as hydrological variability general increases with increasing spatial distance. In addition, geographical proximity is hampered when a small region contains multiple hydrologically distinct regimes (Greis and Wood 1981).

The second type of flood attribute constitutes a set of physiographic and/or meteorological variables that can have a clear influence on annual flood regimes. Using this type of attribute enables basins that are hydrologically similar but not necessarily geographical contiguous to be considered within the same group (Wiltshire 1985). It is generally difficult to determine which variables are most influential, however, as floods are driven by a variety of factors (Burn 1989), and the importance of these factors might vary from site to site and flood to flood. Generally influential variables are rainfall pattern, basin slope, soil type, and drainage area (Burn 1989).

The third type is related to the information of the flood sample, such us correlation coefficient of coincident flood values from two sites (Burn 1988), and the statistics of the flood sample. The latter one is generally avoided as a flood-related attribute because flood statistics are typically employed for assessing region homogeneity, and the repeated use of flood statistics during the region formation process might call into question the validity of the homogeneity assessment (Hosking and Wallis 1997).

Flood seasonality is another attribute that relates to the flood sample. Flood seasonality refers to the regularity of the dates of the observed flood values. This attribute is easy to obtain as it can be derived from available flood information and without the use of flood magnitude data, which can be used for the homogeneity test. Burn (1997), Merz et al. (1999), and Castellarin et al. (2001) have demonstrated flood seasonality yields suitably homogeneous regions for regional FFA.

Attributes of previous three can also be combined together as the composite attribute. Studies have shown that a composite attribute is reasonably effective for grouping homogeneous flood regions in selected regions of Canada (Burn 1990b; Zrinji and Burn 1996), and the world (Burn and Goel 2000; Merz and Blöschl 2005; Eslamian and Hosseinipour 2010). In particular, Merz and Blöschl (2005) investigated various distinct attributes for basin regionalization in Austria, and the results favoured composite attributes. The selection of variables to integrate into the composite attribute, however, requires engineering judgment, such as which variable is more important, and the proper influence for each variable (i.e., the weights of each variable).

A preferred attribute which effectively describes the general flood regime in Canada remains unknown. As the country has huge landmass and diverse hydrological characteristics, a composite attribute constituting several influential flood-related variables is preferred, but the derivation of this attribute may require tremendous efforts. It is worthwhile investigating some of the basic attributes so that those influential variables can be uncovered and adopted for future research.

Chapter 3 Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada

Ziyang Zhang[a*], Tricia A. Stadnyk[a], Donald H. Burn[b]

[a]*Civil Engineering, University of Manitoba, Winnipeg, Canada*

[b]*Civil and Environmental Engineering, University of Waterloo, Waterloo, Canada*

*Corresponding author: E-mail: umzha494@myumanitoba.ca

Floods are the most frequent natural disaster in Canada. Projected variations in precipitation and temperature will further intensify extreme events, necessitating improved flood planning and water resource management. The Natural Sciences and Engineering Research Council funded FloodNet project is developing a standardized flood estimation manual for Canada; such a nation-wide manual would make flood frequency application more effective, consistent, and reproducible, reducing the need for subjective judgement. This research investigates a preferred at-site flood frequency distribution for Canadian annual peak flow dataset. Four 3-parameter distributions: Generalized Logistic, Generalized Extreme Value, Pearson Type III, and Log Pearson Type III are compared both quantitatively and qualitatively using two goodness-of-fit tests (i.e., Modified Anderson-Darling test and L-Moment Z^{DIST} test). Results show the Generalized Extreme Value distribution is the preferred distribution amongst the four considered distributions for at-site flood frequency analysis across Canada. Three flood frequency plots are selected as examples to characterize the predictive performance of distributions in a visual context. The findings of this work will assist with the development of a Canadian Flood Estimation Manual.

Key words: at-site flood frequency analysis; annual peak flow; statistical distribution; goodness-of-fit test.

3.1 Introduction

The ability to estimate flood quantiles is essential for water resources management and engineering design. Flood frequency analysis (FFA) is one of the standard methods to reliably model flood quantiles for different return periods. One pending issue of FFA in Canada is the absence of a preferred statistical distribution, which is likely due to the sizeable geographical extent of the country where water resources management has traditionally been administrated by provincial jurisdictions (Jakob and Church 2011; Faulkner et al. 2016). Consequently, various distributions are in use for FFA across the country (Watt 1989; Pilon and Adamowski 1992; Aucoin et al. 2011; Lu 2016).

The FloodNet Strategic Network funded by the Natural Sciences Engineering Research Council of Canada (NSERC) is a national collaboration with universities, government agencies, and industries with the aim of enhancing hydrologic forecasting and flood management at the national scale (FloodNet 2018). It is designed to provide practical flood management tools and software for operational use. Theme 1-6 undertakes the task of developing a unified national flood estimation manual, which provides recommended procedures of FFA for operational users. As part of Theme 1-6, this study aims to identify the preferred statistical distribution that best-fits station flood data across Canada. Despite regional FFA having received growing attention in the hydrological community, this study is framed under the at-site approach as a beginning investigation for the establishment of Canadian flood estimation manual. Findings of this study provide a benchmark for the developing manual and support the expansion of other techniques that will be further developed in the FloodNet Theme 1-6.

The choices of the statistical distribution and the parameter estimation method have historically been a major research topic for FFA where investigators focused on finding the optimal combination of both subjects for the region of interest (Cunnane 1985; Cunnane 1989; Bobée et al. 1993; Bobée and Rasmussen 1995). Since the development of L-Moments statistics (Hosking 1990), however, it has been proven that parameter estimation methods should employ L-Moment methods for most flood distribution modelling due to its simplicity, unbiasedness, and robustness (Vogel and Fennessey 1993; Bobée and Rasmussen 1995; Hosking and Wallis 1997). Therefore, the scope of this research is confined to the choice of distribution.

Distributions for flood modelling usually have two to four parameters. The increase in the number of parameters of a distribution improves the flexibility of modelling fit, but it also enlarges uncertainty bounds for estimation (Bobée and Rasmussen 1995). Although 2-parameter distributions were historically popular (Cunnane 1989), Bobée et al. (1993) stated that distributions for flood modelling should have at least three parameters to provide sufficient flexibility to address the large variety of flood data. Since the early 1990s, a growing number of studies (Table 1) have recommended the use of 3-parameter distributions. This phenomenon was more pronounced for distributions that can be found in the national flood estimation manuals of several countries (Robson and Reed 1999; Salinas et al. 2014b; Ball et al. 2016; England Jr. et al. 2018). Therefore, 3-parameter distributions have been widely recognized as the optimal trade-off between flexibility and uncertainty for flood modelling.

Table 1: Previous studies of the preferred distribution(s) for FFA in different geographical areas of the world.

Reference	Area	Preferred distributions	Goodness-of-fit measure	At-site or Regional FFA
Haktanir 1991	Turkey	GLO; LP3	CS; KS	At-site
Mutua 1994	Kenya	LN3	AIC; CS	At-site
Karim and Chowdhury 1995	Bangladesh	GEV	PPCC; LMRD; Root mean square deviation;	At-site
Önöz and Bayazit 1995	Worldwide	GEV	AD; CS; KS; LMRD; PPCC; Expected number of exceedances	At-site
Vogel and Wilson 1996	US	GEV; LN3; LP3	LMRD	At-site
Robson and Reed 1999	UK	GLO	LMRD; LMZT	Regional
Rahman et al. 2013	Australia	LP3; GEV; GPA	AD; AIC; BIC; KS; LMRD	At-site
Pilon and Adamowski 1992	NS, Canada	GEV	LMRD	Both
Fortin, Bobée, and Bernier 1997	ON and QC, Canada	No conclusive preference	Nonparametric Bayesian simulation	At-site
Yue and Wang 2004	Canada	GEV for the Pacific and southern BC mountains; LN3 for YT and northern BC; LP3 for northwestern forest; W3 for Arctic Tundra; PE3 for Prairies, northeastern forest, Great Lakes and	LMRD; LMZT	At-site

		St. Lawrence, Atlantic, and Mackenzie regions.		
Aucoin et al. 2011	NB, Canada	GEV	AD; Negative log-likelihood value; Visual assessment of quantile-quantile plots	Both
Lu 2016	NL, Canada	LN3 for Newfoundland; GEV for Labrador	LMRD; LMZT	Regional

Abbreviations used in Table 1 as follows:

Distribution: GEV: Generalized Extreme Value; GLO: Generalized Logistics; LN3: 3-parameter log Normal; LP3: log Pearson Type III; PE3: Pearson Type III; W3: 3-parameter Weibull.

Goodness-of-fit measures: AD: Anderson-Darling test; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; CS: Chi-Squared test; KS: Kolmogorov-Smirnov test; LMRD: L-Moment ratio diagram; LMZT: L-Moment Z^{DIST} test; PPCC: Probability plot correlation coefficient.

Canadian provinces and territories: BC: British Columbia; NB: New Brunswick; NL: Newfoundland and Labrador; NS: Nova Scotia; ON: Ontario; QC: Quebec; YT: Yukon.

It is well established that there is no universal distribution that adequately fits flood data for every station record. Published studies have shown that the preferred distribution for a region of interest was generally the one that was acceptable for most of the stations (Önöz and Bayazit 1995; Robson and Reed 1999; Rahman et al. 2013), however, for regions with diverse flood characteristics, one distribution is possibly insufficient (Haktanir 1991; Rahman et al. 2013; Salinas et al. 2014b). A comprehensive survey result of the preferred distribution(s) for 55 government agencies in 27 countries was given from the Operational Hydrology Report of the World Meteorological Organization (Cunnane 1989). Reported in the year of 1989, the 2-parameter Gumbel distribution was the most used distribution across the survey respondents. However, this situation was changed in the early 1990s. Table 1 presents some studies published since the early 1990s that have investigated the preferred distribution(s) for FFA in different geographical regions of the world. Findings in these studies suggest that the 3-parameter Generalized Extreme Value (GEV) distribution has been more frequently preferred across the world for FFA. In addition, the GEV distribution can be considered as an enhanced version of the Gumbel distribution because the Gumbel distribution is one of the three special cases of the GEV distribution.

Goodness-of-fit analysis has been proven as the standard technique to evaluate whether applied distributions adequately fit the station flood sample within a certain degree of confidence (Cunnane 1985). The goodness-of-fit analysis can be as simple as a visual examination of the flood frequency plot, or as complex as a statistical test that reduces the need for subjective judgment. The latter is typically in the form of a hypothesis test that the result of the applied distribution is either accepted or rejected. One conventional type of hypothesis testing is the empirical goodness-of-fit test (Stephens 1974), which uses the empirical cumulative distribution function (CDF) to avoid the utility of a fixed form distribution. The empirical CDF is a non-parametric CDF derived from the sample data and presented as a step function. The test statistic measures the differences between the step function and the fitted distribution model. The Kolmogorov-Smirnov (KS) test (Kolmogorov 1933; Smirnov 1948) and the Anderson-Darling (AD) test (Anderson and Darling 1952; Anderson and Darling 1954) are frequently used empirical tests for assessing the fit of a distribution for flood modelling (Ahmad et al. 1988a; Haktanir 1991; Önöz and Bayazit 1995; Aucoin et al. 2011; Rahman et al. 2013).

Another popular goodness-of-fit analysis is based on the L-Moment statistic (Hosking and Wallis 1997), which has been extensively utilized to identify a 3-parameter distribution used in FFA (Pearson 1991; Nathan and Weinmann 1991; Pilon and Adamowski 1992; M. A. Karim and Chowdhury 1995; Önöz and Bayazit 1995; Vogel and Wilson 1996; Robson and Reed 1999;

Yue and Wang 2004; Rahman et al. 2013; Lu 2016). The theoretical reasoning of this approach is, when estimating a 3-parameter distribution, the location, scale and shape parameters are estimated by matching the first three L-Moments (i.e., L-Location, L-Scale, and L-Skewness) of the sample. This leaves the fourth L-Moment, L-Kurtosis, as a validating variable. If the sample L-Kurtosis is close to the distribution's theoretical L-Kurtosis, it is suggested that the estimated distribution is appropriate in modelling the sample. The L-Moment goodness-of-fit analysis can be implemented by two schemes. The first scheme is the L-Moment ratio diagram that provides a visual tool to compare multiple distributions and to indicate a relatively better distribution. The second scheme, the L-Moment Z^{DIST} test, which has the framework of a statistical hypothesis test, is considered an enhanced version of L-Moment ratio diagram. The Z^{DIST} test considers the effect of sampling variability (simulated by a 4-parameter Kappa distribution) and provides a definitive test conclusion in terms of whether the distribution fitting is accepted or not. In addition, the Z^{DIST} test statistic directly expresses the quality of fit, therefore, multiple distribution comparison can be simply made by comparing test statistics of the distributions. Such comparison cannot be efficiently made in empirical goodness-of-fit tests that require additional transformation from test statistic to probability value (i.e., p value).

In this study, four widely used 3-parameter distributions, i.e., Generalized Logistic (GLO), GEV, Pearson Type III (PE3), and log Pearson Type III (LP3), are compared both quantitatively and qualitatively based on their fit to flood data from a number of Canadian stations. Distributions are evaluated quantitatively via goodness-of-fit tests comprised of the modified Anderson-Darling (MAD) test and the L-Moment Z^{DIST} test. The number of accepted samples for the distributions is determined, with the distribution having the largest number of accepted samples providing an indication of a general preferred distribution. Then flood samples for which all candidate distributions are acceptable under both goodness-of-fit tests are selected to perform the qualitatively analyses. At each sample, the qualities of fit of candidate distributions are compared based on the Z^{DIST} statistic and distributions are ranked based on the comparison result. The ranking frequency of each distribution reflects the qualitative performance of each considered distribution. Consequently, the distribution that outperforms in both quantitative and qualitatively measure is identified as the generally preferred distribution Canada-wide.

3.2 Data

In Canada, flood data are available as annual maximum flow (AMF) and instantaneous peak flow (IPF). In each calendar year, the AMF records the largest mean daily flow, while IPF records the largest instantaneous flow magnitude. In this study, two annual peak flow series, i.e., AMF series and IPF series, are utilized to assess the fit of the distributions. Flood samples from these series' were obtained from Canadian hydrometric stations in the Reference Hydrometric Basin Network (RHBN) (Brimley et al. 1999; Harvey et al. 1999). Developed by Water Survey of Canada (WSC), the RHBN is a sub-set of Canadian hydrometric gauging stations developed with the intent of examining the hydrologic effects of climate change. Each RHBN station is described as a near pristine basin with stable land-use conditions, high quality of flow measurement records, and absence of anthropogenic influences (Whitfield et al. 2012; Burn et al. 2012). Flood data provided from RHBN stations generally satisfy the requirements for FFA.

There are 227 RHBN hydrometric stations with basin area ranging between 3.63km^2 and 146,400km^2. The number of stations with available flood data was 223 at the time of project launch. Among these stations, 19 stations located in Quebec had records available for the AMF only. Stations with fewer than 14-station-years were removed in the analysis because short record length is typically deemed unsuitable for at-site FFA (Robson and Reed 1999). Figure 1 shows the location of RHBN stations with at least 14-station-years of AMF record. The figure reveals a non-uniform station distribution across Canada with a majority located in the southern region of the country. Stations that contain 0 m^3/s in its flood samples are screened out in this study because the existence of 0 m^3/s precludes the normal procedure of fitting distributions, and requires additional adjustment of conditional probability distribution (Jennings and Benson 1969). For simplicity, these stations are removed. It is worthwhile mentioning that all such stations were located in the Palliser triangle (Figure 1), the driest area in the Canadian Prairies.

Consequently, the AMF series was comprised of 211 stations, while the IPF series contained 186 stations. The maximum, mean, and minimum record length for the AMF series were 103, 49, and 15 years, respectively; and were 83, 39, and 14 years, respectively for the IPF series. More than 86% of AMF samples and 77% of IPF samples had station record length over 30 years.

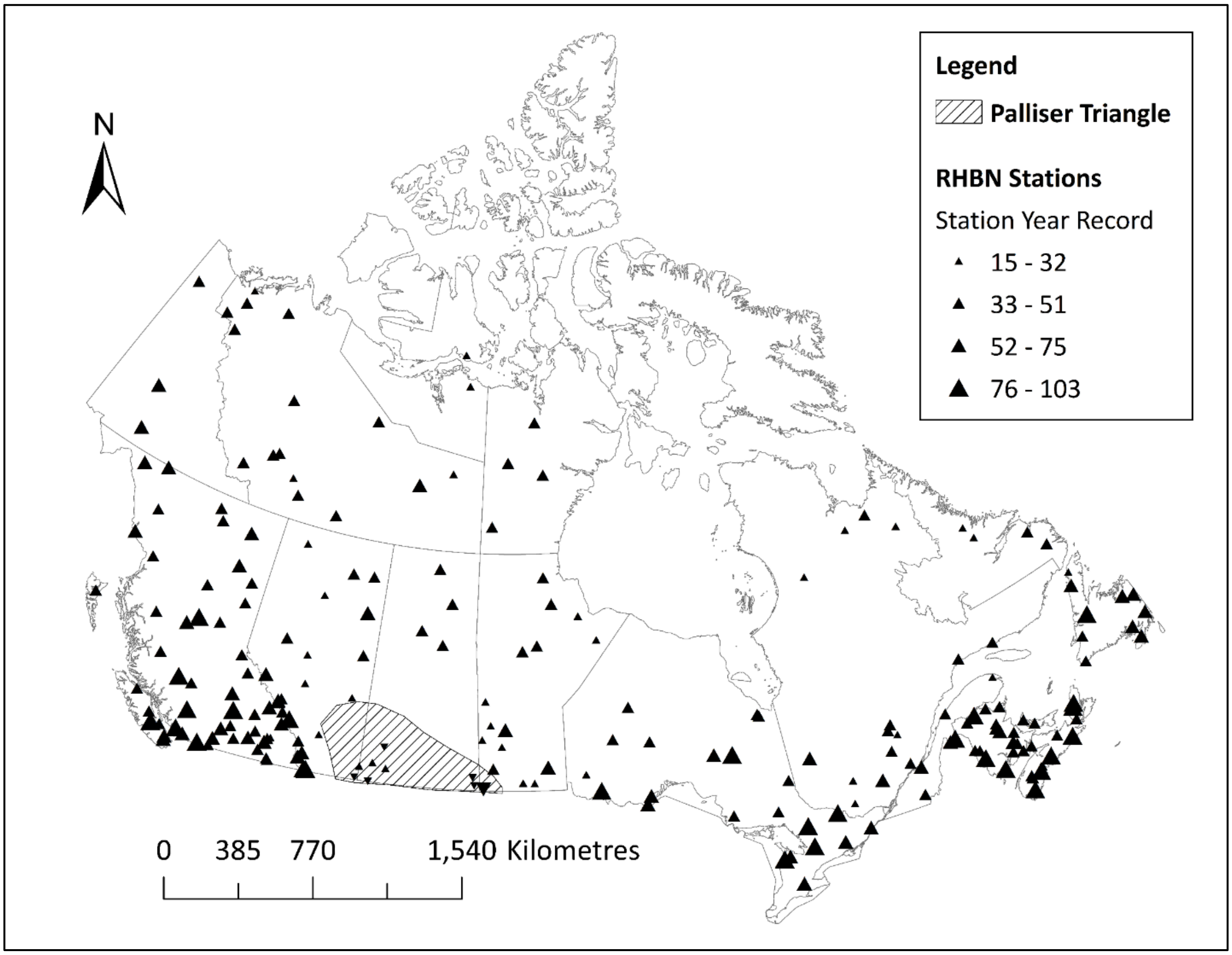

Figure 1: Location of RHBN stations with at least 14-year station record. Size of the upward pointing arrow is proportional to the station record length of the AMF series; downward arrow indicates station having 0 m^3/s in the annual flood series.

3.3 Candidate Distributions

3.3.1 Generalized Logistic Distribution (GLO)

The GLO distribution is the reparametrized version of log-logistic distribution introduced by Ahmad et al. (1988b). When the skewness of a dataset is positive (common for flood data), GLO's sample distribution is lower bounded and upper unbounded. This feature is well suited for flood frequency modelling because flooding potential should not be limited by a maximum flow value. Other commonly used distributions may obtain a sample distribution with an upper bound, which is considered unrealistic in flood response because the maximum discharge of a river should not be constrained to an upper limit.

The GLO has the following cumulative distribution function:

$$F(x) = \left\{1 + \left[1 - \frac{k(x-\mu)}{\alpha}\right]^{\frac{1}{k}}\right\}^{-1}, \quad k \neq 0 \tag{15}$$

$$F(x) = \left\{1 + \exp\left[-\frac{(x-\mu)}{\alpha}\right]\right\}^{-1}, \quad k = 0 \tag{16}$$

where μ, α, and k are the location, scale and shape parameters, respectively. For $k > 0$ ($k < 0$), the variable x is upper (lower) bounded to $\mu + \frac{\alpha}{k}$. For $k = 0$, the variable x is unbounded.

3.3.2 Generalized Extreme Value Distribution (GEV)

The GEV distribution (Jenkinson 1955) is included because it is a recommended flood frequency distribution in many countries, such as Australia, Austria, Germany, Italy and Spain (Salinas et al. 2014b). Its CDF covers the 2-parameter distributions of Gumbel (EV1), Frechet (EV2), and Weibull (EV3). The GEV distribution is considered to have a strong justification for flood modelling. According to extreme value theory, the distribution of independent rare events should resemble a distribution that belongs to the family of GEV distributions (Coles 2001).

The GEV has the following cumulative distribution function:

$$F(x) = \exp\left\{-\left[1 - \frac{k(x-\mu)}{\alpha}\right]^{\frac{1}{k}}\right\}, \quad k \neq 0 \tag{17}$$

$$F(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\alpha}\right)\right], \qquad k = 0 \tag{18}$$

where μ, α, and k are the location, scale and shape parameters, respectively. For $k > 0$ ($k < 0$), the variable x is upper (lower) bounded to $\mu + \frac{\alpha}{k}$. For $k = 0$, the variable x is unbounded. The GEV distribution becomes Gumbel, Frechet, and Weibull for shape parameters $k = 0$, $k > 0$, and $k < 0$, respectively.

3.3.3 Pearson Type III (PE3) and Log Pearson Type III (LP3) Distributions

PE3 and LP3 (i.e., PE3 with log transformation) distributions are of particular importance in hydrologic applications and have been extensively used for flood modelling (Cunnane 1989). PE3 has been applied in various parts of Canada (Watt 1989) and is considered in this study. LP3 has commonly been used in FFA (Griffis and Stedinger 2007b; Griffis and Stedinger 2007a), with both the US and Australia considering LP3 as their national standard distribution (U.S. Water Resources Council 1982; Ball et al. 2016). Since both countries have sizeable geographical extent with diverse hydrology (similar to Canada), LP3 has been included in this study.

For PE3 distribution, the cumulative distribution function, adopted from Hosking and Wallis (1997), is shown below. Let μ, σ, and γ denote the mean, standard deviation and skewness of the original data. If $\gamma \neq 0$, let $\alpha = 4/\gamma^2$, $\beta = 0.5\sigma|\gamma|$, and $\xi = \mu - 2\sigma/\gamma$. The CDF function becomes

$$F(x) = \frac{G\left(\alpha, \frac{x-\xi}{\beta}\right)}{\Gamma(\alpha)}, \qquad \gamma > 0 \tag{19}$$

$$F(x) = 1 - \frac{G\left(\alpha, \frac{\xi - x}{\beta}\right)}{\Gamma(\alpha)}, \qquad \gamma < 0 \tag{20}$$

where $G(\alpha, x)$ is the incomplete gamma function and $\Gamma(.)$ is the gamma function with

$$G(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt \tag{21}$$

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{22}$$

The variable x is bounded by $\xi \leq x < \infty$ when $\gamma > 0$ and $-\infty < x \leq \xi$ when $\gamma < 0$. For $\gamma = 0$, the distribution is normal with unbounded x.

For LP3 distribution, variable x is transformed into log space (i.e., $\log x$) and then proceeding into equation (21) and (22). Parameters of μ, σ, and γ denote the mean, standard deviation, and skewness of the logarithm data, respectively. Base-10 logarithm is employed in this study as guided by the US flood estimation manual - Bulletin 17B (U.S. Water Resources Council 1982).

3.4 Method of Parameter Estimation

In this study, the method of L-Moments is used for parameter estimation for the GEV and GLO distributions. As previously mentioned, L-Moment estimator has been reported to have the advantages of being effective, accurate, approximately unbiased, and robust to the presence of outliers (Vogel and Fennessey 1993; Bobée and Rasmussen 1995; Hosking and Wallis 1997). Such advantages have resulted in the L-Moment method being extensively applied for GEV and GLO distributions (Ahmad et al. 1988a; Önöz and Bayazit 1995; Robson and Reed 1999; Shin et al. 2012; Heo et al. 2013). Details of the L-Moment parameter estimation for GLO and GEV distributions are given by Hosking and Wallis (1997).

Despite the popularity of L-Moment approach, the method of moments remains the most widely used parameter estimation method for PE3 and LP3 applications (Griffis and Stedinger 2007a), partly because of its good performance, and based on robustness and convenience. Review of England Jr. et al. (2018) describing recent development of the US flood estimation manual - Bulletin 17C considers the use of the method of moments as the base approach for parameter estimation for the LP3 distribution. In this study, the method of moments is used to estimate parameters for PE3 and LP3 distributions. The regional skewness coefficient method, a technique to refine the method of moment described in both Bulletin 17B and 17C, is not incorporated in this study. The regional skew coefficient considers using of regionalized information, while this study is focused on using at-site flood record only.

3.5 Feasibility of fit

The feasibility of fit is the first assessment to be addressed when a sample distribution is estimated. Only distributions that can be feasibly fit to the data for a given station are considered for evaluating goodness-of-fit. The feasibility of fit is determined by whether or not the fitted data are inside the boundary of the sample distribution. In terms of estimation of a 3-parameter sample distribution, the sample distribution is either upper bounded or lower bounded except when the shape parameter is estimated to a value of zero. This particular exception rarely occurs when modelling skewed data (i.e., flood data). A lower (upper) bounded sample distribution is denoted as feasibly fit sample distribution, when the minimum (maximum) fitted data value is larger (smaller) than the boundary limit; otherwise, it is denoted as an infeasibly fit sample distribution. The infeasibly fit sample distribution is not considered in further modelling as it is unable to fit the original dataset. In this study, the feasibility of fit is assessed before the goodness-of-fit analysis, where only feasibly fit distributions are considered for goodness-of-fit analysis.

3.6 Goodness-of-fit

3.6.1 Modified Anderson-Darling test

The AD test is considered one of the most powerful goodness-of-fit tests and has been widely used in hydrological studies (Kottegoda 1984; Önöz and Bayazit 1995; Baldassarre et al. 2009; Aucoin et al. 2011; Rahman et al. 2013). The test statistic includes a quadratic function and a stronger tail weighting form that gives greater sensitivity to discrepancies in the tails of the distribution. For flood frequency applications, however, only high flow values are of interest. Therefore Ahmad et al. (1988a) proposed a modified version of AD test that measured discrepancies only on the upper tail of the distribution.

AU^2, denoted as the upper tail MAD statistic, can be computed as (Shin et al. 2012):

$$AU^2 = \frac{n}{2} - 2\sum_{i=1}^{n} F(x_i) - \sum_{i=1}^{n}\left(2 - \frac{2i-1}{n}\right) \log[1 - F(x_i)] \quad (23)$$

where n is the sample size (i.e., the number of observations of the sample); x_i is the i^{th} observational value of ordered observations ($x_1 \leq x_2 \ldots \leq x_n$); and $F(x)$ is the CDF to be evaluated.

To decide whether a distribution is accepted, critical values for each considered distribution are needed. Critical values are the cut-off point used to compare to the test statistic, and thus to determine whether the null hypothesis is rejected. In both AD and MAD tests, the null hypothesis is that the estimated distribution is not rejected (i.e., accepted). If the test statistic is greater than the specified critical value, the null hypothesis is rejected.

Although critical values for AD test have been provided from previous studies, critical values for MAD test are different from those for AD test; to our knowledge, only three studies analyzed critical values for MAD test with considerations of GLO and GEV distributions only (Ahmad et al. 1988a; Shin et al. 2012; Heo et al. 2013). Shin et al. (2012) and Heo et al. (2013) illustrated that critical values are correlated with the shape parameter of a given sample distribution, and that Monte-Carlo simulation experiments can be carried out to generate a critical values table based on the shape parameter(s). This study followed the procedure developed by Heo et al. (2013) that the required critical values of each flood sample under each candidate distribution are generated through Monte-Carlo simulation experiments.

For each sample distribution, the procedure followed is:

(1) Form a new inverse CDF (invCDF) of the specified distribution by setting the scale and location parameters to 1 and 0, respectively. Match the shape parameter of the invCDF to the shape parameter of the sample distribution.

(2) Use the invCDF to generate 10,000 random samples with sample size equivalent to the original fitting sample (Shin et al. 2012; Heo et al. 2013). The number of sample generations regulates the trade-off between the precision of critical values and the efficiency of computation. Indicated in Shin et al. (2012) and Heo et al. (2013) studies,

the 10,000 number of sample generations provided reasonably precise critical values for the MAD test with high efficiency of computation.

(3) Each generated sample is fit to the specified distribution and then the AU^2 statistic is calculated.

(4) Sort the 10,000 AU^2 statistics in ascending order.

(5) Critical value of 0.1 significance level corresponding to the 9000th value of the sorted AU^2 statistics is selected to determine the acceptance of the sample distribution.

3.6.2 L-Moment Z^{DIST} test

L-Moment Z^{DIST} test is also used for assessing goodness-of-fit. Similarity to the MAD test, the Z^{DIST} test is able to evaluate whether the fitted distribution is accepted at each station record. In addition, the Z^{DIST} statistic is capable of comparing multiple distributions in terms of the quality of fit.

The Z^{DIST} statistics is expressed as (Hosking and Wallis, 1997):

$$Z^{DIST} = (\tau_4^{DIST} - t_4^S + B_4)/\sigma_4 \tag{24}$$

where τ_4^{DIST} is the theoretical L-Kurtosis of the fitted distribution; t_4^S is the L-Kurtosis of the sample; B_4 is a bias correction term; and σ_4 represents the standard deviation of t_4^S.

B_4 and σ_4 are calculated through a series of simulation experiments. First, a 4-parameter Kappa distribution is estimated based on fitting the original sample through L-Moment parameter estimation. Second, N_s number of samples are generated based on the estimated Kappa distribution ($N_s = 1000$ has been used here). The record length of each generated sample equals the record length of the original sample. Third, the L-Moment ratios of each generated samples are calculated. Thus, B_4 and σ_4 are calculated:

$$B_4 = \frac{1}{N_s}\sum_{m=1}^{N_s}\left(t_4^{[m]} - t_4^S\right) \tag{25}$$

$$\sigma_4 = \left[\frac{1}{N_s - 1}\sum_{m=1}^{N_s}\left(t_4^{[m]} - t_4^S - B_4\right)^2\right]^{\frac{1}{2}} \tag{26}$$

where $t_4^{[m]}$ is the sample L-Kurtosis for the m^{th} generated sample.

At 0.1 significance level, $|Z^{DIST}| \leq 1.64$ indicates the acceptance of fit. For multi-distribution comparisons, the distribution with the smallest $|Z^{DIST}|$ is considered to have the best quality of fit.

It should be noted that the 4-parameter Kappa distribution is commonly selected in Monte-Carlo simulation experiments for flood value generation, while the L-Moment test is implemented for 2- or 3-parameter distributions (Hosking and Wallis 1997). The Kappa distribution can generates a wide variety of flood values, and thus making the statistical test more robust for testing the 2- or 3-parameter distributions.

The two goodness-of-fit methods selected vary in terms of theoretical reasoning, therefore, evaluating the distributions using the selected methods should provide a robust measure of the appropriateness of a distribution. The MAD test focuses on measuring the upper tail discrepancy between the observed data and the modelled distribution. Deviation at the tail dataset provides greater sensitivity to the MAD test result. In terms of L-Moment Z^{DIST} test, the theoretical L-Kurtosis of a given distribution corresponds to the shape of the modelled distribution; focusing on evaluating the shape of the modelled distribution against the statistical characteristic of the fitted data.

3.7 Results and Discussion

Figure 2 presents the MAD test results of candidate distributions at 0.1 significance level for (a) the AMF series and for (b) the IPF series. The number of infeasibly fit samples, rejected samples, and accepted samples are displayed as individual categories in the four distribution columns.

For either type of data, the GEV distribution is accepted by the majority of the flood samples, approximately 87% and 89% acceptance rate for AMF and IPF series, respectively. GLO and LP3 obtained a very similar number of accepted samples, with about 82% to 87% acceptance rate. The PE3 distribution is the weakest amongst the four candidates, obtaining acceptance rates of 67% and 65% for AMF and IPF series, respectively. Overall, the MAD test indicates that the GEV distribution is preferred for modelling Canadian annual peak flow series over other candidates.

Analyzing the amount of infeasibly fit samples and rejected samples assists in examining the performance of a distribution. For the GEV distribution, only one sample is infeasibly fit in the IPF series. This proves that the GEV distribution has a very flexible shape, which is able to cover a wide variety of flood data. The GLO distribution is also subjected to a small number of infeasibly fit samples, while it has the largest number of rejected samples compared with other candidates. For the PE3 distribution, a substantial number of samples are infeasibly fitting, resulting in a considerably lower acceptance rate. The employment of a log transformation (i.e., LP3) considerably improves the problem of infeasibly fit samples for the PE3 distribution, resulting in a much improved acceptance rate, with the number of rejected samples remaining relatively constant.

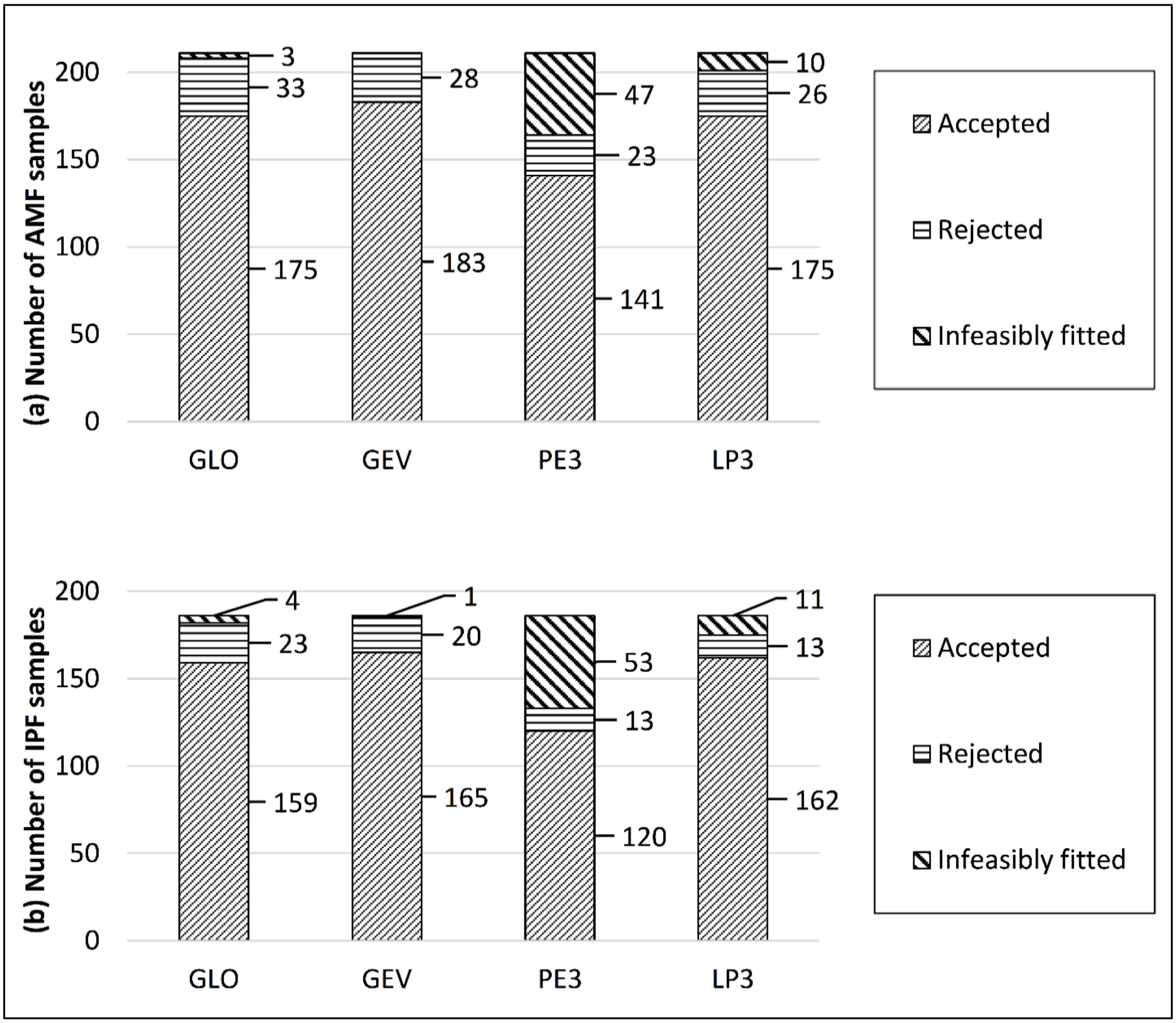

Figure 2: MAD test result at 0.1 significance level for (a) 211 AMF samples and (b) 186 IPF samples.

Figure 3 reports results of the Z^{DIST} test at the 0.1 significance level for (a) the AMF series and for (b) the IPF series. Samples in each series are again categorized into accepted, rejected, and infeasibly fit categories. In addition, three AMF samples and two IPF samples are removed from the LP3 analysis. These samples are precluded from the Z^{DIST} test because their sample data with logarithm transformation have been identified as infeasibly fitting the Kappa distribution, and hence their Z^{DIST} statistics cannot be computed.

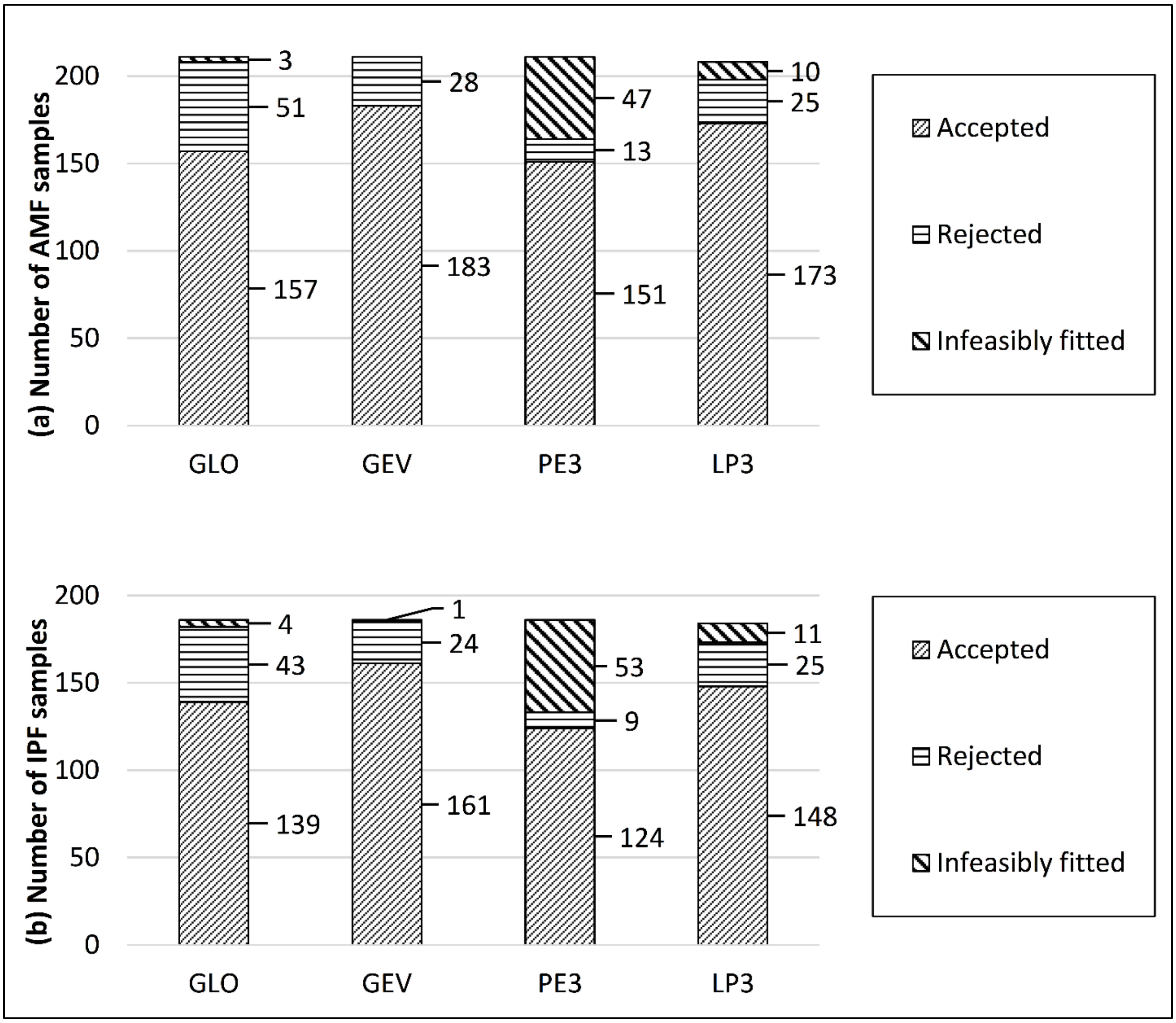

Figure 3: Z^{DIST} test result at 0.1 significance level for (a) 211 AMF samples and (b) 186 IPF samples.

Similar to the MAD test results, the GEV distribution receives the largest number of accepted samples for both AMF and IPF series with about an 87% acceptance rate. The GEV is followed by the LP3 distribution with an acceptance rate of 82% for AMF and 80% for IPF; and then the GLO distribution with an acceptance rate of 74% for AMF and 75% for IPF, and lastly the PE3 distribution that accepts 72% AMF and 67% IPF samples. The number of infeasibly fit samples for each distribution is the same as those shown in the MAD results. The ranking order derived from the Z^{DIST} test is similar to the ranking order derived from the MAD test. However, the number of accepted samples is different between the two tests such that the Z^{DIST} test tends to produce lower acceptance rate, in particular for the IPF series.

The Z^{DIST} statistic is further utilized to compare the quality of fit between candidate distributions at each sample. In total, 88 AMF and 72 IPF samples are selected to carry out this analysis. All these samples are accepted by both MAD and Z^{DIST} tests across four candidate distributions. Based on the distribution's Z^{DIST} statistics for each sample, distributions are ranked from level I to level IV. Level I refers to the best quality of data fitting, while level IV refers to the least quality of data fitting. The frequencies for distributions for each level are collected and summarized for 88 AMF samples (Figure 4a) and 72 IPF samples (Figure 4b).

Based on results in the AMF series, it is clear that the GEV distribution outperforms other distributions. GEV frequencies in level I and II are considerably higher than other distributions and frequency in level IV is much lower. The GEV is followed, in order, by PE3 and LP3 distributions, for which the frequencies of both distributions are about equally distributed across the four levels. The GLO distribution is considered the weakest because more than 50% of samples falls into level IV. For the IPF series, little discrepancy between distribution frequencies is found in level I, therefore, the distribution frequencies in level II to level IV are the focus. The

GEV distribution outperforms the others again because of its higher frequencies in level II and III, while only two records appear as level IV. The GEV is followed by PE3 and LP3, for which similar frequencies are obtained in level II to level IV. The GLO performs the least satisfactory since the majority of samples are level IV. Overall, the multiple model comparison reveals that the GEV distribution performs the best in terms of quality of fit, while GLO performs the worst.

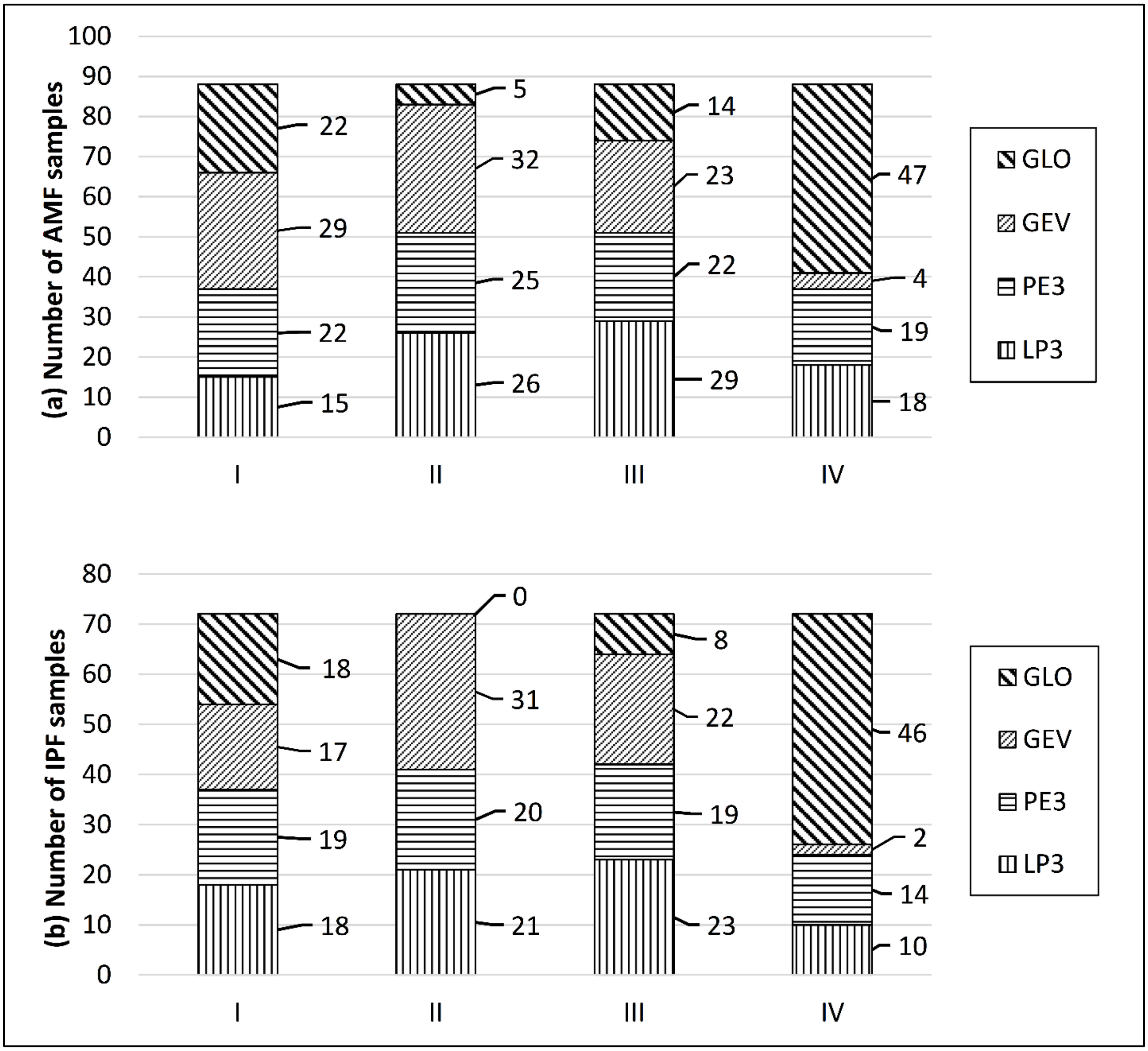

Figure 4: Qualitative goodness-of-fit measure using Z^{DIST} test for (a) 92 AMF samples and (b) 80 IPF samples.

The above goodness-of-fit analyses shows that the GEV distribution is the preferred candidate (over GLO, PE3 and LP3 distributions) for Canada. First, the quantitative measure indicates that the GEV distribution is accepted at most Canadian RHBN stations with respect to both AMF and IPF datasets. Second, the qualitative measure reveals that the GEV distribution is superior to other candidates in terms of quality of fit. This finding corresponds with the general consensus based on the literature overview listed in Table 1 where the GEV distribution was the most popular distribution for modelling flood frequency data.

The second preferred candidate is the LP3 distribution, which performs satisfactorily on both quantitative and qualitative measures. On the other hand, the GLO and PE3 are considered poor in the goodness-of-fit analyses. GLO has been rejected by more than a quarter of the samples for both datasets using the Z^{DIST} test, and has performed the weakest in terms of the qualitative analyses. PE3 is unable to estimate a feasibly fit sample distribution for a considerable number of samples.

The use of either AMF or IPF datasets in the analysis has no influence on the ranking of desirability of the four distributions. For a given considered distribution, its performance (relative to other considered distributions) for the AMF series is generally the same as for the IPF series, however, a generally lower acceptance rate for the IPF series is found. This result is in response to the different configurations of AMF and IPF series. For a given gauging station, flood values in the IPF sample are generally more wide spread (i.e., covering a wider range) compared to flood values in the AMF sample, and therefore, modelling the IPF sample requires the distribution having more robustness to fit a wider range of values.

Results of the two goodness-of-fit tests are plotted geographically for spatial pattern analysis. The original intent was to identify possible geospatial patterns that may reflect geographical preference/disfavor for a particular distribution. Such patterns could assist practitioners with selection guidelines for the choice of a preferred distribution. Partly because of the large extent of the country, and also the non-uniform distribution of stations, no particular spatial pattern has been detected among the considered distributions, Yet what has been revealed is a relatively high rejection rate within the Prairie region of western Canada. Here only the map of the GEV distribution for the AMF series is presented as it is considered the preferred distribution in the goodness-of-fit analyses (Figure 5). The majority of rejected stations are located in the southwest region of the country, extending from the Rocky Mountains to eastern Manitoba. A reduced number of rejected stations are shown for the Atlantic Provinces, and a lesser number of rejected stations are observed across the rest of the country. In general, the number of rejected stations in each sub-region is proportional to the total stations in each sub-region. The exception is the Prairie Provinces where a 32.6% rejection rate for the GEV distribution is found, while the rest of the country has a rejection rate of 14.5%.

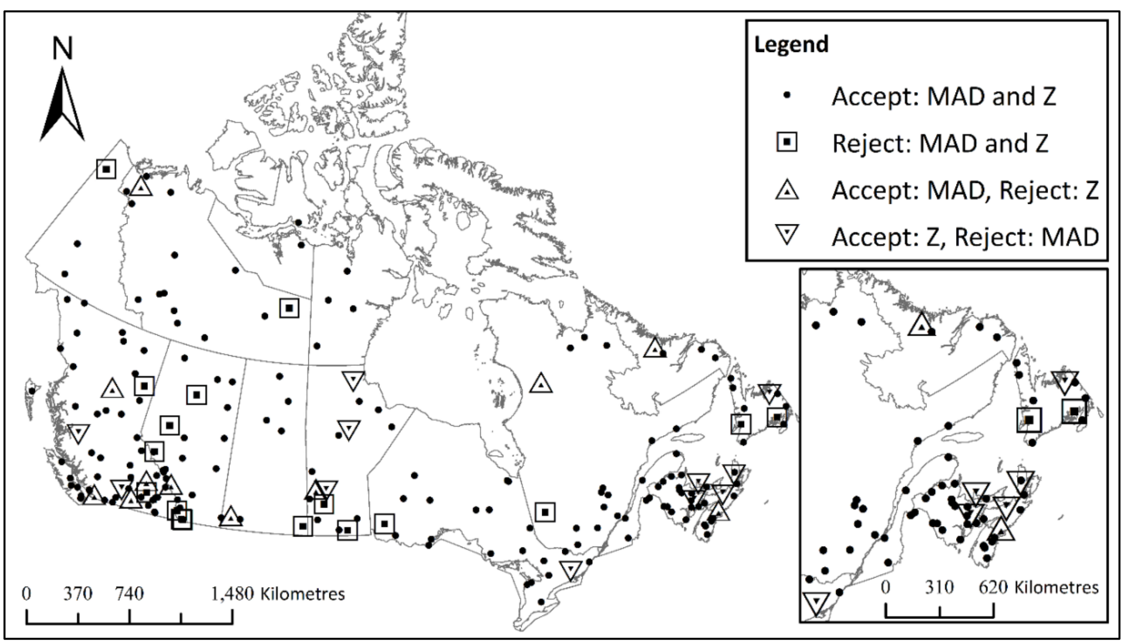

Figure 5: Geospatial analysis of GEV Goodness-of-fit test results for AMF samples at 0.1 significance level.

The modelling performance of the distribution at a site can be seen in a flood frequency plot. In the flood frequency plot, the closeness between the distribution's modelled curve and the observed points reflects the modelling performance of the distribution. Three sites are selected as examples, with the intention to visually illustrate the different fitting scenarios across Canada. The Weir River in northern Manitoba (WSC ID 05UH002) shows an example of mixed fitting results, comprised of acceptance, rejection, and infeasibly fit distributions. Athabasca River near the Municipality of Jasper in Alberta (WSC ID 07AA002) is an example where all candidate distributions are rejected, and the Atlin River near the Town of Atlin in British Columbia (WSC ID 09AA006) displays the opposite scenario, where all candidate distributions are accepted. For each flood frequency plot, flood data with units of cubic meters per second (m^3 s^{-1}) are plotted against the Gumbel reduced variate along with the modelled curves of considered distributions. The Gringorten plotting position (Gringorten 1963) is selected for its good performance regarding unbiased plotting of data (Cunnane 1978; Guo 1990).

Under the two goodness-of-fit tests, distributions of both GLO and GEV have been accepted for modeling the flood sample of site 05UH002 (Figure 6). This result can be visually confirmed in the flood frequency distributions for GLO and GEV, which satisfactorily fit the entire dataset with only a slight underestimation of the largest observed flows. The PE3 distribution has been tested as an infeasibly fit distribution, with the consequence that the PE3 modelled curve fails to capture the smaller observed flows. Although an infeasibly fit distribution is theoretically considered inapplicable for flood frequency analysis, this infeasibly fit PE3 distribution with a lower bound adequately fit the majority of the observed points, including the peak points (except for the largest observed point). Therefore, an infeasibly fit distribution with a lower bound may still be capable of extrapolating the quantile of floods. The two goodness-of-fit tests are not able to identify such situations. Here, the LP3 curve matches the majority of the observed points, however, fails to fit the largest observed point. The cause of this is possibly due to the use of log transformation. Log transformation makes model fitting less sensitive to larger observed values, and a larger deviation between observed points and the modelled curve is found in the upper tail.

At site 07AA002, all four candidate distributions are rejected by both MAD and Z^{DIST} analyses. The flood frequency plot shows that all distributions modelled deviate from the largest observed (peak) flows. Calculating the sample statistics reveals an irregularly low value of L-Kurtosis (0.05). According to the L-Moment ratio diagram (Hosking and Wallis 1997, 209), flood samples with L-Kurtosis below 0.1 are generally out of the acceptable range for distribution modelling. This distribution constraint is not only subject to the candidate distributions considered in this study, but also subject to other 3-parameter distributions which are suitable for AMF modelling. It is possible that a four-parameter distribution is robust enough to fit this particular sample.

For the Atlin River (09AA006), all distributions are accepted under both goodness-of-fit analyses (Figure 8). The modelled curves for GLO, GEV, PE3 and LP3 adequately approximate all the observed flows. For return periods exceeding the largest observed flow, GEV, PE3 and LP3 are all reasonably close to each other, while the GLO curve deviates, resulting in a higher estimated flood quantile. In this case, although all candidate distributions adequately depict the quantile range of observed points, the predictions of extreme value return periods can differ among distributions. This finding corresponds to that of Robson and Reed (1999, 46) who recommended the upper limit of quantile estimation should be below the at-site station record.

Quantile estimation greater than that should be avoided for an at-site approach; regional FFA is preferred in this situation.

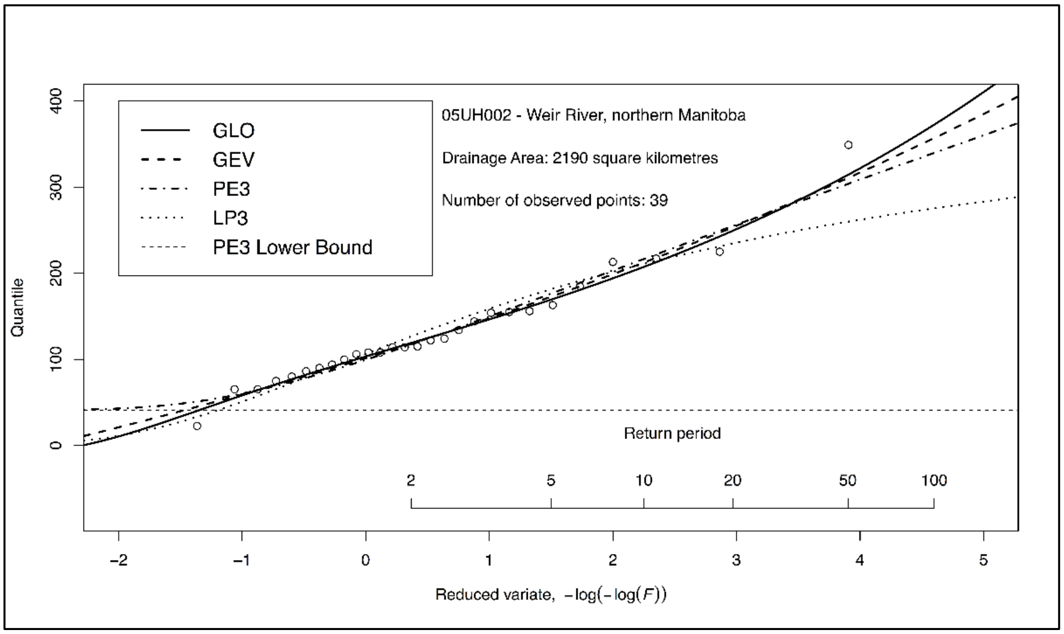

Figure 6: Flood frequency plot for AMF data of 05UH002, illustrating a mixture of fitting results.

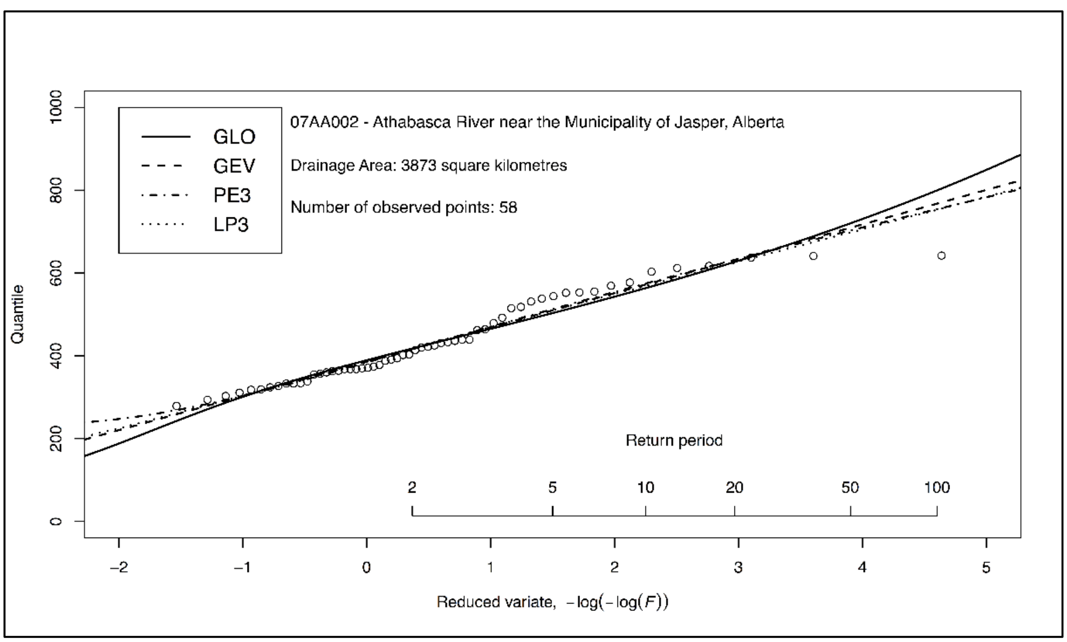

Figure 7: Flood frequency plot for AMF data of 07AA002, illustrating an example of all candidate distributions being rejected for flood modelling.

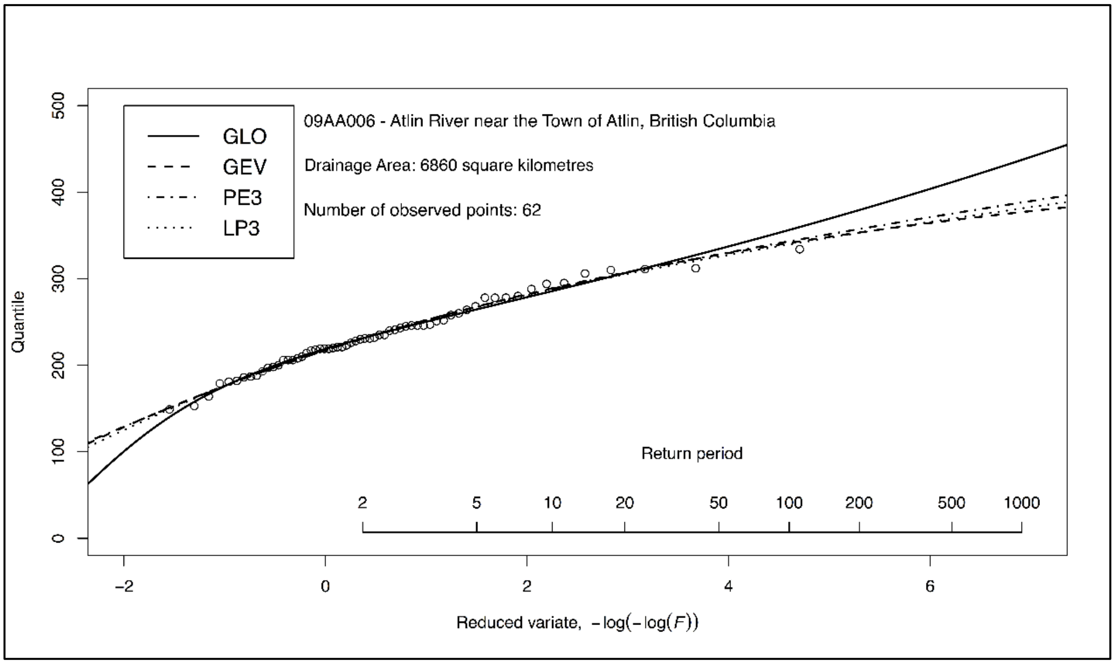

Figure 8: Flood frequency plot for AMF data of 09AA006, illustrating an example of all candidate distributions being accepted for flood modelling.

3.8 Conclusions

This study provides a basis for selecting the GEV distribution as the preferred distribution for at-site FFA in Canada. Based on the MAD and Z^{DIST} goodness-of-fit analyses over four candidate distributions, the GEV distribution is not only accepted with the largest number of Canadian annual peak flow series, but also presented the highest quality of data fitting.

The results of goodness-of-fit of all candidate distributions are plotted Canada-wide for spatial pattern detection. No particular geographical pattern is observed, except that a reduced number of accepted stations is clearly found within the Prairie Provinces.

Flood frequency plots reveal that an infeasibly fit distribution with a lower bound may still be appropriate for upper quantile modelling (i.e., flood modelling). This case cannot be identified by the two goodness-of-fit tests.

The flood frequency plot also reveals that at-site FFA should be used with caution for estimating quantiles where the return period greatly exceeds the station record length. In this situation, regional FFA is more appropriate because it is able to collect more flood data than at-site FFA (Bobée and Rasmussen 1995; Hosking and Wallis 1997). The work presented here is further applicable to find a preferred distribution for regional FFA on a Canada-wide/regional-wide scale.

3.9 Acknowledgements

This work was financially supported by the Natural Sciences and Engineering Research Council of Canada under the Canadian FloodNet project (Grant number – NETGP 451456) and the University of Manitoba under the GETS funding. Environment and Climate Change Canada and Water Survey of Canada are acknowledged for providing data used in this study. The authors would like to dedicate this study to the memory of Professor Peter Rasmussen who was a key contributor to the journey of this research.

Chapter 4 Investigation of attributes for developing homogeneous flood regions for regional flood frequency analysis in Canada

Ziyang Zhang[a*], and Tricia A. Stadnyk[a],

[a]*Civil Engineering, University of Manitoba, Winnipeg, Canada*

*Corresponding author: E-mail: umzha494@myumanitoba.ca

Researchers at the Natural Sciences Engineering Research Council of Canada (NSERC)-funded FloodNet project are developing a nation-wide flood estimation manual for flood frequency analysis (FFA) in Canada. As the homogeneous flood region is a basic requirement for regional index flood based FFA, this study investigates five distinct flood-related attributes (geographical proximity, flood seasonality, physiographic variables, precipitation pattern, and temperature pattern) for their relevance in developing homogeneous flood regions across Canada. For each attribute, the relevance for developing homogeneous regions is assessed based on the number of homogeneous regions that each attribute produces, and the accuracy of flood quantile prediction based on the developed homogeneous regions. Results show that homogeneous flood regions are readily formed for all attributes when gauging sites are spatially clustered. For regions where sites are scattered or sparsely distributed, homogeneous regions are generally difficult to produce regardless of which attribute is considered. Precipitation pattern is preferred for general use in Canada. Flood seasonality, physiographic variables, and temperature pattern are preferred for select regions in Canada. A combined use of flood seasonality and physiographic variables may benefit region formation in the interior region, mountains of British Columbia, and northern Canada. Findings from this study assist with the development of a regionalized approach for design flood estimation in Canada.

Keywords: regional flood frequency analysis; annual peak flow; index flood; flood-related attribute; attribute similarity measure.

4.1 Introduction

Flood frequency analysis is a commonly used method to predict flood potential in engineering design and water resources management. Classic at-site flood frequency analysis is favored for its simplicity, however, the accuracy of flood distributions is questionable when the record lengths of flood data at the target site are short; particularly when modelling larger return periods. Regional flood frequency analysis (RFFA) transfers flood information from "surrounding" sites to offset the data shortage of the target site, thus improving the accuracy of flood frequency estimation. The "surrounding" sites are not necessarily geographically contiguous, but should have similar flood characteristics (Burn and Goel 2000). Nonetheless, geographic contiguity has traditionally used for RFFA since physiographic and climatic variability tend to be smaller within smaller regions (U.S. Water Resources Council 1982; Robson and Reed 1999).

With more than half a century's development in RFFA (Dalrymple 1950; Dalrymple 1960), an increasing number countries have researched the best approach to RFFA for the purposes of achieving a standardized flood estimation guideline, which is beneficial for domestic uniformity in application (Bobée et al. 1993). For example, the U.S. Water Resources Council (1982) incorporated the weighted skewness method into Log Pearson Type III distribution to estimate regional flood quantile. Their published Bulletin 17 series of handbooks outlines default procedures for RFFA for US governmental agencies (U.S. Water Resources Council 1982; England Jr. et al. 2018). Documented in Britain Flood Studies Report (Natural Environment Research Council 1975) and Flood Estimation Handbook (Robson and Reed 1999), the index flood method for regional flood estimation has been accepted by UK governmental agencies. In the Australian guideline (Ball et al. 2016), distribution parameters estimated based on regression models of catchment characteristics form the basis of regionalized flood estimation. In Europe, recent studies have been carried out with the intent to develop unified RFFA techniques for a pan-European environment (Salinas et al., 2014a; b; Mediero et al. 2015). These published guidelines reflect the latest developments in RFFA, and provide rigorous guidance and improved methods for operational use, reducing the need for subjective interpretation by operational users.

In Canada, methods and standards related to FFA have been inconsistently used throughout the country, partly because of the sizeable geographic extent, so that flood management has been traditionally administrated by the provincial jurisdiction (Watt 1989; Faulkner et al. 2016; Moudrak and Feltmate 2017). Jakob and Church (2011) have discussed the need for national flood management in Canada. To address needs of regional flood estimation for practitioners, a unified flood estimation manual for Canada is currently being developed by the Natural Sciences Engineering Research Council of Canada (NSERC)-funded FloodNet Strategic Network. This manual aims to provide prescriptive methods and techniques related to RFFA that should be applied in Canada (FloodNet 2018). As a part of the FloodNet project, this study employs a national-scale application to investigate one key aspect of RFFA, which is the development of homogeneous flood regions based on flood-related attributes.

The index flood method (Dalrymple 1950; Dalrymple 1960) is one of the most credible methods for RFFA (Wallis and Wood 1985; Potter and Lettenmaier 1990; Stedinger and Lu 1995; Hosking and Wallis 1997; Robson and Reed 1999; Yang et al. 2010), and the homogeneous flood region is an essential prerequisite for index flood method. A homogeneous flood region collects a group of hydrologically similar sites so that flood information across these sites can be spatially transferred. The index flood method is a convenient method to transfer flood samples across these sites by considering that all flood samples collected in the region have the same shape of a flood frequency distribution while differ only in terms of an index factor.

The development of homogeneous flood regions is typically assisted by similarity measures of a flood-related attribute. The selected attribute has to be measurable by an attribute distance so that similarities between different catchments can be readily quantified. The flood-related attribute needs to be readily obtainable within a basin, and relevant to the flood regime (Burn 1989). As flood characteristics can change considerably across Canada, it is important to investigate the different flood-related attributes for their relevance in developing homogeneous regions.

In the past, several studies have been conducted that have related to the selection of flood-related attributes for region formation in Canada. Burn (1988) developed homogeneous regions for regional flood estimation in southern Manitoba. Coincident annual peak flood values for each site were treated as flood-related attributes. The similarity of two sites was measured based on the correlation coefficient of flood values from the two sites. The similarity measure provided principle indication of gauging site delineation; and the geographical distribution was also used as input in the delineation process. As a result, 41 sites in southern Manitoba were grouped into three homogeneous regions with an average region record length of 497 station-years.

In the same study area, Burn (1989) used a composite attribute including flood statistics and geographic location. The coefficient of variation (CV) of floods, mean annual flow divided by drainage area (QDA), and latitude and longitude of the gauging station were selected, and the weighting coefficients of each attribute were analyzed to optimize the homogeneity of the developed region. Results showed that flood statistics (i.e., CV and QDA) were relatively more effective than geographic proximity for the formation of homogeneous regions, and the CV attribute was found to be more informative than the QDA. The composite attribute is generally constituted by several key variables for which describe multiple aspects of hydrological characteristics, therefore, it is generally more powerful to divide catchments into different distinct flood regions. One challenge of using a composite attribute is the assignment of proper weighting values to each attribute during the aggregation process. Burn and Goel (2000) used optimization search techniques to determine the weighting coefficients as a more objective approach.

Burn (1997) introduced the attribute of flood seasonality for application across the Canadian Prairies. Similarity between sites is measured based on the timing and regularity of the dates of flooding. One advantage of flood seasonality is that the attribute value is convenient to derive compared to other attribute values, such as basin physiographic characteristics. It also avoids the use of statistics from flood data, which are preferred for validation of the homogeneity of the flood region. Burn and Whitfield (2016), Burn et al. (2016), and Durocher et al. (2018) used flood seasonality to discriminate flood behaviours of hundreds of Canadian sites into three principle regimes: nival, mixed, and pluvial. It was carried out based on the timing and regularity of flood dates in seasonality space, which reveals the benefit of using flood seasonality to classify flood behaviour and to identify hydrological similar sites in Canada.

The above studies have documented several applications for using flood attributes to form homogeneous flood regions in Canada, however, there is a lack of a consensus regarding which attribute is more relevant for Canadian flood characteristics, and thus more effective in developing homogeneous flood regions across Canada.

In this study, we investigate five distinct attributes for their relevance in developing homogeneous flood regions for RFFA Canada-wide. For each attribute, relevance is investigated based on productive and predictive performances, using the number of homogeneous regions that an attribute produces. An automatic region revision algorithm (ARRA) is created to support the development of homogeneous regions, and predictive performance is assessed based on the accuracy of flood quantile estimation. The findings of this study will support the FloodNet project, and contribute to the development of the Canadian flood estimation manual.

4.2 Methodology

4.2.1 Study attributes

The five attributes considered in this study are (1) geographical proximity, (2) physiographic variables, (3) flood seasonality, (4) precipitation pattern, and (5) temperature pattern.

Geographical proximity is based on the rationale that catchments closer to each other generally encompass similar hydrological and physiographical characteristics, and therefore, catchments with smaller geographical proximity are more likely to exhibit similar flood regimes, and thus to form a homogeneous region. The presence of large spatial variability in flood characteristics might question the use of geographical proximity, therefore, using physiographic variables, which have a key influence on the dominant flood generating mechanisms, provides another potential way to group sites with similar flood behavior. Geographic proximity and physiographic variables are the most common flood-related attributes for RFFA, and have been extensively used in applications worldwide (U.S. Water Resources Council 1982; Wiltshire 1985; Robson and Reed 1999; Burn and Goel 2000; Merz and Blöschl 2005), and in regions of Canada (Burn 1988; Burn 1989; Zrinji and Burn 1996).

As previously noted, flood seasonality has the advantage of convenience in attribute extraction. In addition, it has been previously applied and benefited flood studies in Canada in terms of catchment classification (Burn and Whitfield 2016; Burn et al. 2016; Durocher et al. 2018) and formation of homogeneous region (Burn 1997; Burn et al. 1997). Therefore, flood seasonality is included in this study.

Precipitation and temperature pattern are based on the 12 monthly average values of precipitation and temperature computed for the catchment, respectively. They are selected because we consider that each of these attributes has strong potential to describe flood characteristics in Canada. As noted previously, flood generating mechanisms in Canada are dominated by either rainfall (pluvial), snowmelt (nival), and rain-on-snow (mixed) events (Buttle et al. 2016; Durocher et al. 2018). The monthly temporal pattern of precipitation and temperature are considered containing key information concerning the flood generation process. For example, precipitation accumulation during winter months will dominate the magnitude of the spring melt event. Large precipitation values in summer and fall also suggest potential rainfall driven peak floods. Temperature values in the melt season affect the timing and magnitude of spring peak floods. Therefore, we explore the attributes of both precipitation and temperature pattern given their potential usefulness in mapping annual flood patterns.

4.2.2 Dataset

Annual maxima flood samples taken from the Canadian Reference Hydrometric Basin Network (RHBN) are used for this application. Developed by Water Survey of Canada, the RHBN, which constitutes 223 gauging sites in total (Water Survey of Canada 2016), is only a small subset of Canadian hydrometric gauging network (6348 gauging with flow record sites in total) (Water Survey of Canada 2016). However, the RHBN sites have been identified as near pristine catchments, high quality of flow measurements, with an absence of anthropogenic effects (Brimley et al. 1999; Whitfield et al. 2012). These merits make their flood data ideal for RFFA. In addition to the 223 RHBN sites, only 186 sites have corresponding physiographic variables available, supported by Environment and Climate Change Canada (ECCC) (Klyszejko, Erika. E-mail message to author, May 2, 2016). Therefore, the total number of gauging sites considered in this study is 186, contributing a total of 186 annual maximum flood samples. Although RHBN stations generally have flow record that is greater than 20-year record length, some sites were seasonal operated so that not all calendar years are able to derive the annual maximum flood. The average station record length amongst our samples is 48 years, with a maximum 103 years and minimum of eight years. More than 80% samples have station record lengths greater than 30 years.

The geographical distribution of the 186 sites is presented in Figure 9. The record length distribution of the 186 sites is presented in Figure 10. The x-axis in the figure corresponds to the longitude degree from west to east with noted by provinces and territories. Figure 9 and 10 show that most study sites are located in British Columbia and Atlantic Provinces with relatively higher record lengths compared to other areas. The Prairie Provinces, particularly in Saskatchewan and Manitoba, have relatively fewer number of stations with relatively smaller record lengths. The three northern territories have the fewest number of gauging sites with about 40-year record length in average.

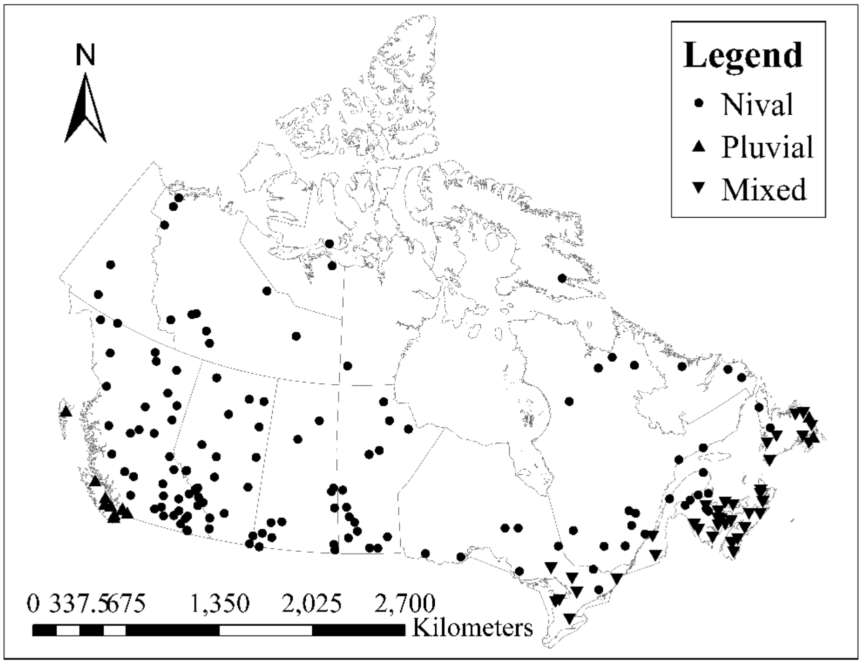

Figure 9: Geographical location of 186 study sites identifying primary cause of flood response.

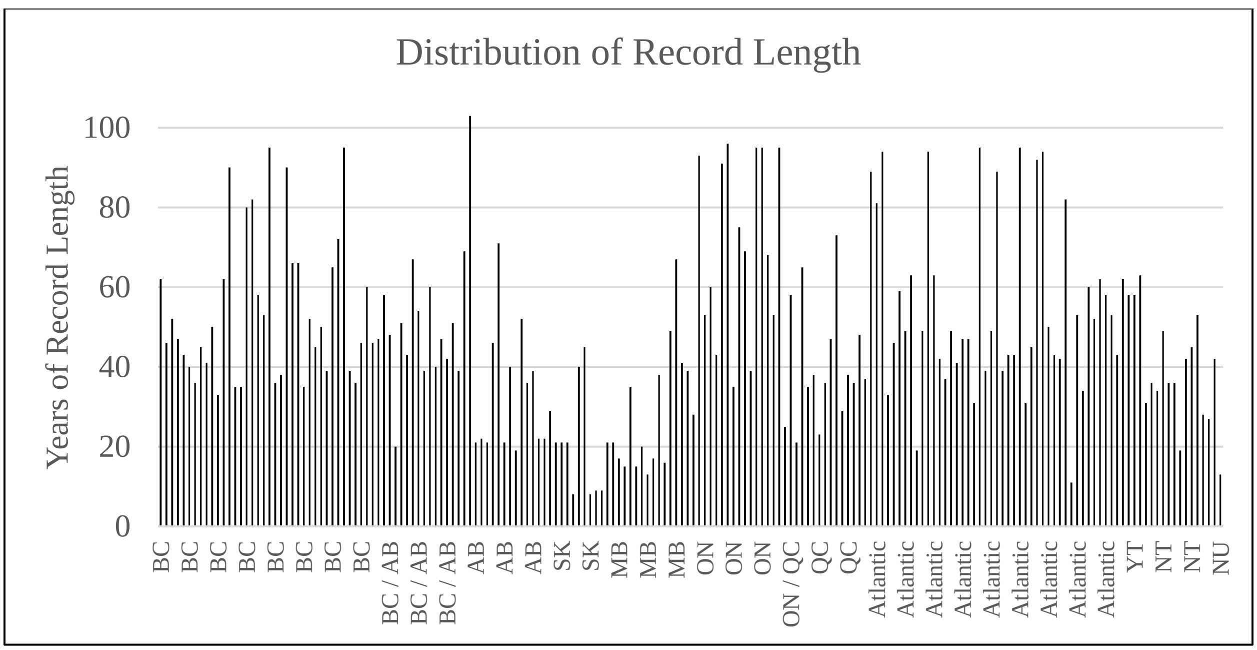

Figure 10: Distribution of record length of the 186 flood samples. Sites in the ten provinces are plotted in the order of longitude degree from west (left side of the figure) to east (right side of the plot). Sites in the three territories are plotted in the most right. (3 sites in northeast QC are embedded under the Atlantic Provinces)

4.2.3 Similarity distance attribute

Geographical proximity

The latitude and longitude of the gauging stations are used to calculate the similarity distance between two catchments. The altitude factor in distance calculation is ignored for simplicity for our Canada-wide scale application, as altitude variation is generally negligible with respect to spatial variation in latitude and longitude. Nevertheless, distance calculation without altitude consideration should be used in caution when sites are clustered in small area with large altitude variation. The similarity distance between catchment m and n is defined as:

$$d_{mn} = [(Lat_m - Lat_n)^2 + (Lon_m - Lon_n)^2]^{0.5} \quad (27)$$

where Lat_m and Lon_m are the latitude and longitude coordinates for the gauging site of catchment m.

Physiographic variables

The selection of physiographic variables is based on the stepwise regression method, which has been used to select flood-related attributes in several previous studies (Nathan and McMahon 1990; Noto and Loggia 2009; Eslamian and Hosseinipour 2010). The stepwise regression method is an automatic procedure to select explanatory variables based on the development of a multilinear regression model. Candidate variables are iteratively added and removed based on the use of statistical t-test until the predictive power of the regression model is optimized.

In this study, 66 different physiographic variables for each site have been provided by Environment of Climate Change Canada (ECCC) (Klyszejko, Erika. E-mail message to author, May 2, 2016) Because different variables have different units and scale, variables are normalized by their standard deviation. The dependent variable for the stepwise regression considers the median value of each flood sample, which corresponds to the 2-year return period flood on the annual maximum scale (Robson and Reed 1999). The median value is considered a robust indicator of annual maximum characteristics principally because it is not affected by the magnitude of exceptionally large flood events (Robson and Reed 1999; Faulkner et al. 2016). Consequently, the stepwise method screens all sets of physiographical attributes and selects a subset including: (1) catchment area, (2) waterbody area in the catchment, (3) standard deviation of elevation across the catchment, (4) average annual air temperature for the catchment, and (5) average annual precipitation for the catchment. Variables (2) and (3) were derived from the ECCC National Hydrology Network database. Variables (4) and (5) were computed based on 10-km historical gridded climate data from a 30-year period of record from 1981 to 2010. These data provided by ECCC were computed based on historical monthly climate grids for North America (McKenney et al. 2006; Natural Resources Canada 2018).

The similarity distance between catchment m and n is calculated based on a weighted Euclidean distance formula defined as:

$$d_{mn} = \left[\sum_{j=1}^{k} w_j \left(x_{mj} - x_{nj}\right)^2\right]^{0.5} \tag{28}$$

where k is the number of physiographic variables, w_j is the weighting factor for physiographic variable j, and x_{mj} is the standardized value for physiographic variable j of catchment m. w_j controls the relative importance of variable j. Here, 0.4 is assigned to the basin area and 0.15 is assigned to the remaining four variables. These selected weights reflect the coefficients output from the stepwise model, but are approximated to a 0.05 interval.

Flood seasonality

Similarity between catchments is measured using a unit polar coordinate system. A catchment is presented as a point in the polar coordinate system, and can be positioned by angular and radial values. The angular value reflects the average date of flood occurrence, where as the radial value reflects the variability of the occurrence dates of different floods. A larger radial value indicates smaller variability in flood occurrence date. A radial value of 1 indicates no variability in occurrence date, implying that all floods occur on the same day of the year, each year.

Based on the Burn (1997) definition, for a single flood event, the date of occurrence of the event is transformed from Julian date to an angular value, where Julian date 1 is January 1st and 365 is December 31st:

$$\theta_i = (Julian\ Date)_i(\frac{2\pi}{365}) \tag{29}$$

For a given catchment flood sample comprised k flooding events, its Cartesian coordinate $\bar{x}$ and $\bar{y}$ in the unit circle are calculated as:

$$\bar{x} = \frac{1}{k}\sum_{i=1}^{k}\cos(\theta_i) \tag{30}$$

$$\bar{y} = \frac{1}{k}\sum_{i=1}^{k}\sin(\theta_i) \tag{31}$$

Therefore, the similarity distance between catchment m and n is calculated as:

$$d_{mn} = [(\bar{x}_m - \bar{x}_n)^2 + (\bar{y}_m - \bar{y}_n)^2]^{0.5} \tag{32}$$

Followed by Durocher et al. (2018) classification, sites used in this study are classified into nival, pluvial, and mixed regimes, indicated in Figure 11. The nival sites are subject to regular flood dates in the spring snowmelt period, suggesting a domination of snowmelt driven flood. These sites are primarily located in interior and northern lands, and mountainous British Columbia. A smaller number of sites are exclusively driven by pluvial regimes with average flood dates occurring from fall to winter. These sites are located in southwest British Columbia and the Maritime Provinces. A substantial number of our study sites are also subject to mixed responses. These sites typically experience annual peak floods from March to May. The wide range of regularity suggests that annual peak floods are driven by multiple responses, such as snowmelt and rain-on-snow. These sites are mostly located in southeast Ontario, southern Quebec, and the Atlantic Provinces.

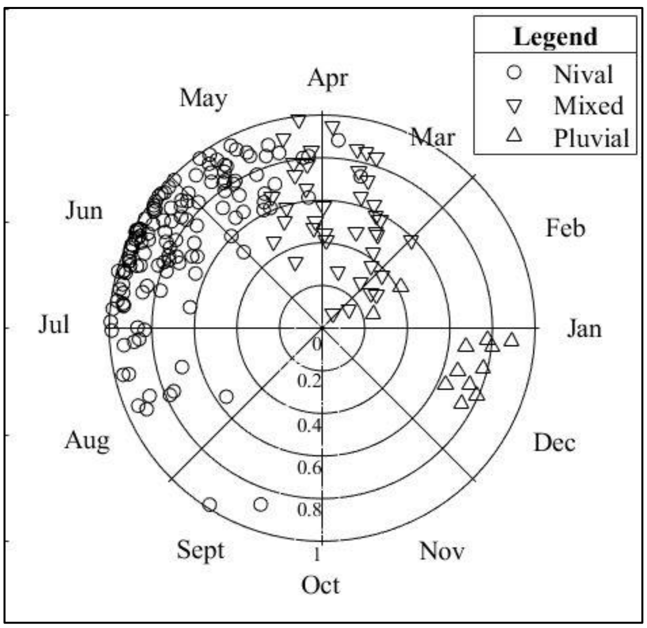

Figure 11: 186 study sites in flood seasonality space.

Precipitation pattern

Similarity measures based on precipitation pattern are attributed to the values of monthly average precipitation from January to December in each catchment. These values were computed from 10-km gridded climate data for a 30-year period between 1981 and 2010 (McKenney et al. 2006;

Natural Resources Canada 2018), and were provided by ECCC (Klyszejko, Erika. E-mail message to author, May 2, 2016). The correlation coefficient is selected to assist in measuring the similarity between two catchments. In contrast to Euclidean distance, correlation coefficient is considered more effective to characterize the pattern of two sets of data as it measures linearity of these data, while Euclidean distance measures the distance of two points in a specified matric space. The correlation coefficient between catchment n and m is described as:

$$r_{nm} = \frac{\sum_{i=1}^{12}(x_{ni} - \overline{x_n})(x_{mi} - \overline{x_m})}{\sqrt{\sum_{i=1}^{12}(x_{ni} - \overline{x_n})^2}\sqrt{\sum_{i=1}^{12}(x_{mi} - \overline{x_m})^2}} \tag{33}$$

where x_{ni} is the monthly average precipitation value for month i of catchment n, and $\overline{x_n}$ is the average of the 12 monthly average precipitation values for catchment n expressed as:

$$\overline{x_n} = \frac{1}{12}\sum_{i=1}^{12} x_{ni} \tag{34}$$

r_{nm} is ranged from -1 to 1 that being exactly 1 (-1) indicates a perfect positive (negative) linear relationship of the two datasets, and being exactly 0 indicates no linear relationship. For similarity measure of catchment m and n, r_{nm} closer to 1 indicates a stronger positive relationship for catchment m and n, therefore, the similarity distance based on correlation coefficient is computed as

$$d_{nm} = 1 - r_{nm} \tag{35}$$

Temperature pattern

Similarity measures based on temperature pattern are attributed to the values of monthly average temperature for each catchment. These values were also derived from 10-km gridded climate data for a 30-year period (1981 to 2010) (McKenney et al. 2006; Natural Resources Canada

2018), and were provided by ECCC (Klyszejko, Erika. E-mail message to author, May 2, 2016). Similarly, monthly average temperature of catchment n and m are input into equation (33) and (34), with equation (35) used to calculate the similarity distance between the two catchments.

4.2.4 Forming flood regions based on the region of influence approach

The region of influence (ROI) approach (Burn 1990a; b) is used to form flood regions for each study site. The ROI defines that each target site has a unique flood region. The addition of other sites to the region is in order of the shortest similarity distance to greatest. Determining the number of sites in a region is a trade-off between the size of the region and the quality of the region. A larger region size benefits flood estimation because of longer record lengths, and therefore the potential to estimate larger return period floods. The quality of the region (i.e., homogeneity in flood characteristics), however, generally decreases when more sites are added to a region. The 5T rule that region size (i.e., total station years of record) should be five times greater than the return period of interest (T) has been widely accepted as for an optimal trade-off point (Robson and Reed 1999; Burn and Goel 2000). The 5T rule is adopted in this study.

4.2.5 The Generalized Extreme Value (GEV) distribution

Although there is no universal distribution that is suitable for the modelling every flood sample, the Generalized Extreme Value (GEV) distribution with L-Moment parameter estimation method is one of the most robust distributions for FFA and is selected for this study. The GEV distribution has been recommended by many studies for RFFA (Noto and Loggia 2009; Salinas et al. 2014b; Ball et al. 2016), including for parts of Canada (Zrinji and Burn 1994; Aucoin et al. 2011; Lu 2016). In addition, a recent investigation showed that the GEV distribution is better than Generalized Logistic, Pearson Type III, and log Pearson Type III distribution for at-site fitting of both Canadian annual maximum data and instantaneous peak flow data (Zhang et al., submitted). The GEV distribution has a strong theoretical support for which the extreme value theory indicates that the distribution of extreme events should follow the mathematical form of the family of the extreme value distributions (Coles 2001). The L-Moment parameter estimation method is selected as it has become the default method to fit the GEV distribution for its simplicity, robustness, and nearly unbiased estimation (Vogel and Fennessey 1993; Hosking and Wallis 1997).

4.2.6 L-Moment homogeneity test

The homogeneity test aims to verify if sites in the flood region exhibit similar flood characteristics at an acceptable level of statistical significance. Since L-Moments are considered unbiased statistics of flood data, the L-Moment homogeneity test has received much attention in RFFA applications (Zrinji and Burn 1994; Zrinji and Burn 1996; Burn 1997; Robson and Reed 1999; Burn and Goel 2000; Merz and Blöschl 2005). Based on Hosking and Wallis (1997), the first step of the homogeneity test is to determine the regional L-Moment ratios t^R, t_3^R, and t_4^R, denoted as the regional L-CV, L-skewness, and L-kurtosis, respectively. For a region comprising N sites, the regional L-Moment t^R (similarly apply for t_3^R and t_4^R) is calculated as:

$$t^R = \sum_{i=1}^{N} n_i t^{(i)} / \sum_{i=1}^{N} n_i \tag{36}$$

where $t^{(i)}$ is the at-site L-Moment ratio for site i, and n_i is the record length for site i.

The dispersion can then be expressed as:

$$V = [\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2 / \sum_{i=1}^{N} n_i]^{0.5} \tag{37}$$

To assess if the dispersion V is within the limit of region homogeneity, two variables are required: μ_V - the expected mean of V; and σ_V - the expected standard deviation of V. μ_V and σ_V are estimated through a large number of reproductions of the original region. To do this, a Kappa distribution fitted by L-Moment ratios of 1, t^R, t_3^R, and t_4^R is used to reproduce N_{sim} number of the original regions (N_{sim}=1000 is used here). Each reproduced region has the same region size (N sites in the region) and the same record length n_i for site i with respect to the original region.

For each reproduced region, the dispersion V is calculated using equations (36) and (37). Based on the N_{sim} number of V values, the expected mean μ_V and the expected standard deviation σ_V can be obtained.

Lastly, the homogeneity statistic is defined as

$$H = \frac{V - \mu_V}{\sigma_V} \tag{38}$$

where H is the homogeneity statistic. A region is regarded as acceptably homogeneous if $H < 1$, possibly homogeneous if $1 \leq H < 2$, and definitely heterogeneous if $H \geq 2$. In this study, $H < 1$ is used for judging if the region is homogeneous.

4.2.7 Automatic region revision algorithm (ARRA)

For each attribute, the attribute similarity under the ROI approach develops an initial flood region for each study site. In many cases, the initially-formed region does not meet the homogeneity requirement (i.e., $H < 1$). As a result, a region revision process is required to revise the membership of the initial region and thus to reduce region heterogeneity. The revision process includes adding, deleting, and replacing site(s) in a region together with testing homogeneity after each progressive change. In the past, this was largely carried out through a heuristic process, meaning there was no set procedure regarding which stage should be processed first (Burn 1989; Burn and Goel 2000; Atiem and Harmancioğlu 2006; Yang et al. 2010). For our large-scale study, however, it is ineffective to proceed using a heuristic process for each region, therefore, an automatic region revision algorithm (ARRA) is developed with the intent to reduce region heterogeneity through automatic modification of the region membership.

When the ARRA is applied once, a heterogeneous region is input into the ARRA, and a revised region with improved homogeneity is output from the ARRA. If the output region does not meet the homogeneous criteria (i.e., $H < 1$), the ARRA can be re-applied to the region to further reduce heterogeneity. Every time the membership is modified by the ARRA, however, the homogeneity of the membership increases but attribute similarity decreases because the newly added sites have larger attribute distances compared to the site being replaced. As a region should be formed primarily based on attribute similarity, the number of ARRA applications should be constrained to a threshold value for an adequate trade-off between region homogeneity and attribute similarity. A sensitivity analysis using ARRA applications to revise 186 randomly formed initial regions, and then counting how many homogeneous regions can be produced after each ARRA application was performed. It has been determined that a maximum of 5 iterations of the ARRA should be applied for an appropriate trade-off.

Figure 12 illustrates the process flowchart of the ARRA in one application. The L-Moment homogeneity test is embedded in the ARRA and used to identify which region site should be removed and which new site should be added so that the homogeneity of the region can be improved by the greatest amount. The order of searching a newly added site depends on the attribute similarity such that higher attribute similarity sites are tested first. The process flowchart does not terminate until an improved region is formed along with satisfaction of the 5T region size rule.

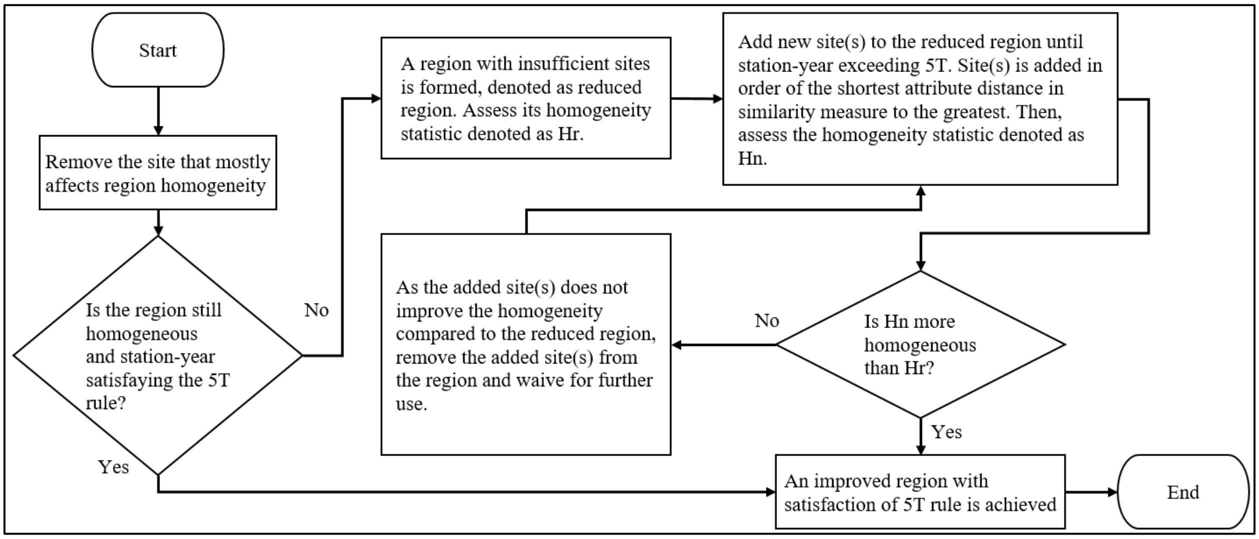

Figure 12: Flowchart of the ARRA process.

4.2.8 Method to assess the productive performance of each attribute

The productive performance of each attribute is assessed based on the number of homogeneous regions that the attribute can develop. For each of the five considered attributes, the process of producing homogeneous regions is demonstrated below. The first step is the development of the initial flood regions for each study site. This is carried out based on the ROI approach for which regions are formed based on attribute similarity. The region size is set to be 500 station years, which allows flood estimation up to a 100-year return period according to the 5T rule. The 100-year flood has been listed as one of the common design flood values for major highway bridges and culverts in Canada (Watt 1989; Moudrak and Feltmate 2017).

As each flood sample provides essential information of its flood regime, flood sample of the target site is normally a part of its flood region. In this study, the flood sample of the target site is not included in its region so that more rigorous testing of the developed region can be conducted. Leaving the target site out of the region also allows the method to be suitable for ungauged site analyses, a potential topic for future study.

The next step is to evaluate the homogeneity of the initial region using the L-Moment homogeneity test. If the initial region is heterogeneous, the ARRA is applied to revise the membership with maximum of five applications. The homogeneity of the revised region is evaluated again by the homogeneity test. Consequently, the number of homogeneous regions produced using each attribute is determined.

4.2.9 Method to assess the predictive performance of each attribute

The predictive performance of each attribute is assessed based on the accuracy of the regional estimate. The regional estimate is compared to a “true” flood quantile estimated by the at-site sample. It is common in practice that the “true” quantile can be reliably estimated based on at-site flood frequency analysis when the return period of interest is below half the at-site record length (i.e., a 2T requirement) (Robson and Reed 1999). Therefore, comparing the regional estimates to at-site estimates provides an avenue to assess the accuracy of the regional estimates.

There are only 11 sites having record lengths greater than 90 years, therefore, for the purpose of reliable at-site estimation, the return periods selected for comparison cannot be extreme, and are ranged from 20 to 45 years. For each return period T, the final selected sites are those for which are able to develop the 5T homogeneous regions across all attributes, and also have record length greater than 2T for reliable at-site estimation. A homogeneous region is easier to form when the region size is smaller, therefore, the number of sites that are available to analyze for each return period are different, and more sites are available for smaller return periods.

Table 2 lists the return periods considered for comparison, the number of sites available for each return period, and the required record lengths for adequate at-site and regional estimates. It is noteworthy that flood estimation of both at-site and regional approaches is subject to sampling uncertainty, and the uncertainty bound decreases with the decrease of estimation return period. Thus, relatively smaller return periods for comparison provides better reliability of assessing results. The Generalized Extreme Value (GEV) distribution with method of L-Moments is used to estimate flood quantiles for both at-site and regional estimates.

Table 2: Required record length for at-site and regional estimate for each comparison return period.

Return period for comparison	Required record length for at-site estimate	Number of sites available	Station-years of record for regional estimate
20	40	88	100
25	50	47	125
30	60	29	150
35	70	15	175
40	80	14	200
45	90	11	225

Bias and root mean square error (RMSE) measures are used to give a robust assessment of the regional estimates as the bias reflects the precision of estimation, and the RMSE reflects the spread (or uncertainty) with respect to the "true" quantile. Bias and RMSE are computed for each comparison between at-site and regional estimates, and then averaged for each attribute to reflect the generalized performance of each attribute. The statistics are computed as follow.

$$bias = \frac{1}{n}\sum_{i=1}^{n}\frac{Q_i - q_i}{q_i} \tag{39}$$

$$rmse = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_i - q_i}{q_i}\right)^2\right]^{0.5} \tag{40}$$

where Q_i is the quantile of regional estimate for site i, q_i is the quantile of at-site estimate for site i, and n is the number of available sites for analysis for each return period.

4.3 Results of productive performance of attributes

For each attribute, 186 initial flood regions are developed based on similarity measure of the attribute. These regions are then applied to the ARRA with different number of ARRA applications. The number of homogeneous regions produced is relative to the number of ARRA applications (Table 3). When the ARRA is not applied, meaning the region is formed based on attribute similarity alone, all attributes form homogeneous regions for less than 6% of the study sites, which is considerably low. Therefore, forming homogeneous regions based on attribute similarity alone is ineffective, and the use of the region revision process (i.e., the ARRA) is deemed necessary.

Table 3: The number of homogeneous regions produced for each attribute.

Number of ARRA applications	Considered Flood-Related Attributes					Alternative Series (initial regions randomly formed)
	Geographical proximity	Flood seasonality	Physiographic variables	Precipitation pattern	Temperature pattern	
0	***10***	6	5	6	***10***	0
1	***26***	22	17	23	21	1
2	49	43	35	50	***54***	9
3	70	50	52	69	***80***	22
4	83	66	69	82	***88***	43
5	89	78	83	***99***	94	63
6	97	98	97	***110***	97	74
7	106	110	104	***118***	105	98
8	106	116	106	***120***	109	112

Bold italicized numbers indicate the best outcome.

Once the ARRA is implemented, the number of homogeneous regions generated increases across all attributes. For two to four applications of the ARRA, the geographical proximity, precipitation pattern, and temperature pattern produce relatively more homogeneous regions than flood seasonality and physiographic variables. For five or more applications, precipitation pattern produces the most homogeneous regions. To determine a proper threshold for the number of applications of the ARRA, an alternative series comprised of 186 regions, for which membership is randomly formed without the use of attribute similarity, is used ranging from one to eight applications of the ARRA (last column, Table 3). Comparing productive results between the five attributes and the alternative series, it is found that after eight applications of the ARRA, attribute similarity is considered largely irrelevant to the production of homogeneous regions. When the ARRA is applied five times, homogeneous regions are produced for approximately half of the study sites across all attributes. At five applications, the number of regions associated with each attribute remains greater than the alternative series, suggesting reasonable preservation of attribute similarity. Therefore, a maximum of five applications of the ARRA appears to be a suitable balance between maintaining attribute similarity for a region, and leveraging the revision power of the ARRA.

With appropriate use of the ARRA (i.e. maximum five applications), about half of the study sites successfully develop homogeneous regions across all attributes, and these regions would be suitable for predicting the 100-year flood quantile. Precipitation pattern produces the largest number of homogeneous regions compared to other attributes, indicating it is generally more relevant to flood regimes in terms of hydrological similarity than the other attributes. This is potentially because peak floods in the Canadian domain are generally dominated by either rainfall (pluvial), snowmelt (nival), and rain-on-snow (mixed) mechanisms (Buttle et al. 2016; Durocher et al. 2018), and the temporal pattern of precipitation is highly relevant to discriminating among these flood generation mechanisms.

Temperature pattern performs second best amongst the five attributes, indicating that the temporal pattern of temperature is also effective for grouping sites with similar flood regimes.

The is possibly because that most of our study sites experience annual peak floods from March to July, and the temperature pattern would dictate the main flood driving mechanism between snowmelt or rain-on-snow response.

A large proportion of our study sites are clustered in small geographical regions, which resulted in geographical proximity ranking third for the attribute assessment. It should be noted that clusters of gauging sites are mostly presented in developed areas of southern Canada. For the majority of the Canadian landmass, gauging sites are sparse and scattered, where geographical proximity would be deemed unsuitable (Figure 13). Flood seasonality and physiographic variables produce fewer homogeneous regions than other candidate attributes based on total numbers. However, both attributes reveal their suitability in selected geographical regions of Canada. The spatial patterning of each attribute is considered.

For each attribute, the 186 study sites are plotted spatially (Figure 13) with sites that are capable of forming its unique homogeneous regions are identified in red circle. It should be noted that some sites are not able to develop its unique homogeneous regions, but they may be collected in other sites' homogeneous regions. These sites remain indicated in blue circle in the figure.

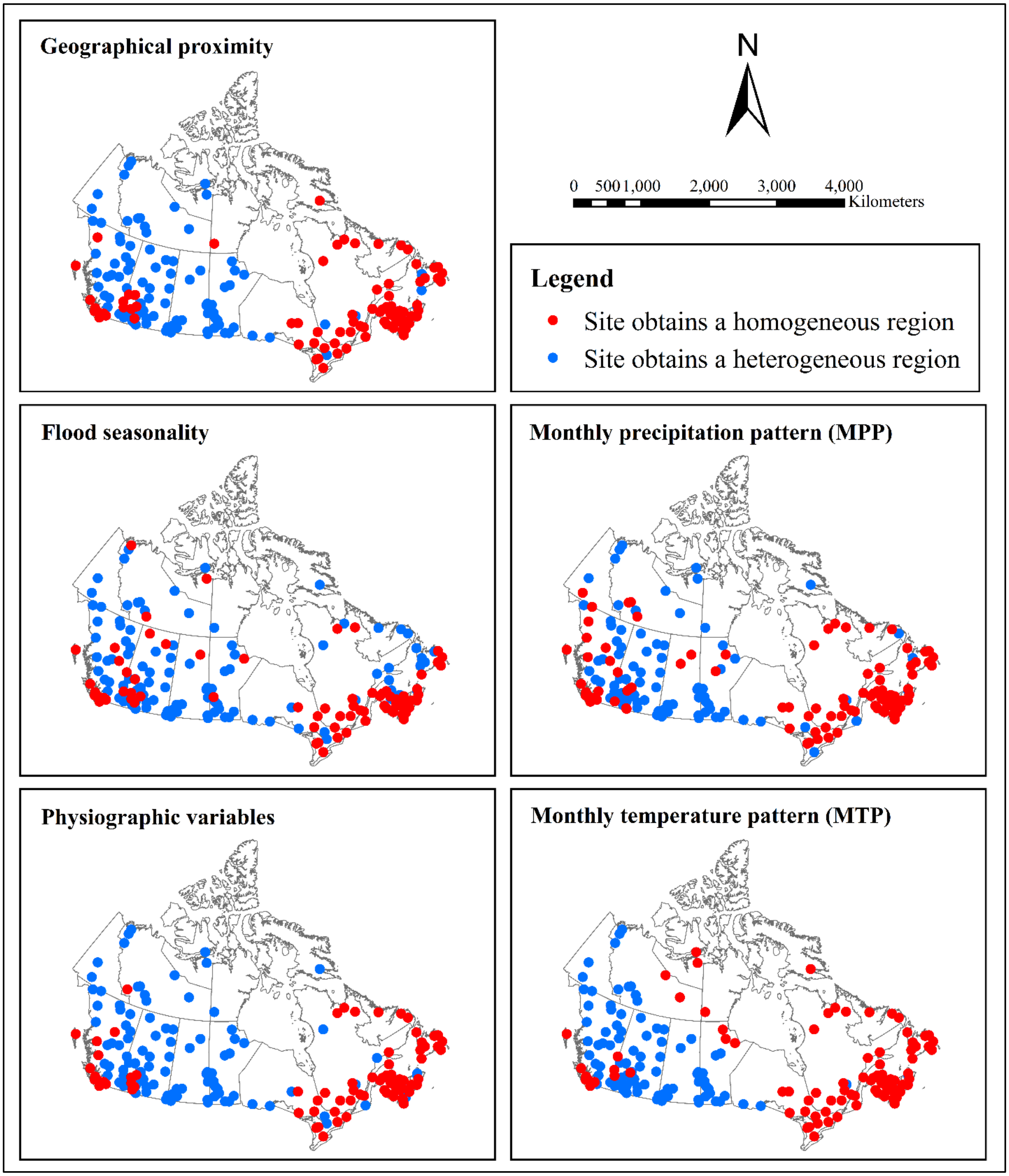

Figure 13: Sites which formed homogeneous regions identified on a map, considering all attributes.

In general, the mapping reveals that sites clustered in small geographical regions are more likely to form homogeneous regions than sites that are sparse or spatially scattered. Clusters of sites occur in southeast Canada, such as the Maritime Provinces, Newfoundland, southern Quebec, and southeast Ontario; as well as southwest Canada, from Vancouver Island to southeast Alberta. Sites in these regions are mostly able to develop homogeneous regions regardless of which attribute is considered.

For the interior and northern landmass, gauging sites are generally scattered or more sparsely distributed and the formation of homogeneous regions is, therefore, generally difficult due to the diversity of hydrological characteristics with increased spatial distance. Geographical proximity encounters difficulty, in particular, as spatial distance is poorly related to hydrological characteristics when sites are far away. The other four attributes reveal improved results relative to geographical proximity, but differ across attribute.

In the Arctic region of Canada, temperature pattern produces relatively more homogeneous regions compared to the other attributes. This is partly because annual peak floods in Arctic Canada are controlled by rate of snowmelt. The temporal pattern of temperature is effective for grouping sites based on different freshet behavior.

From the west coast to interior British Columbia, sites are dominated by mountain systems with rugged, steeply sloping terrain that results in considerable variability in precipitation patterns and physiographic characteristics (Buttle et al. 2016). As a result, not many sites are able to form homogeneous regions, regardless of which attribute is considered. Physiographic variables and precipitation pattern aide a few sites in developing homogeneous regions relative to the other three attributes.

The large interior and northern landmass is also problematic for forming homogenous regions among sites, regardless of which attribute is used. This is particularly true for the Prairie region. Annual peak floods at most sites are driven by nival responses. Some sites in the south are exposed to rain-on-snow behaviour or affected by summer thunderstorms (i.e., rainfall-driven floods). The runoff landscape is complicated by numerous potholes and wetlands that dynamically alter contributing area by (dis)connecting during (dry) flooding periods (Shook and Pomeroy 2011; Muhammad et al. 2018; Muhammad et al. 2019). As a result, flood behaviour is largely varied among catchments, resulting in difficulty when forming homogeneous regions. In addition, flood information in the Prairie region is generally limited in terms of the number of sites and record length of each site (Figure 10), which constrains the information that can be used to develop flood regions.

Although none of the attributes are satisfactory for the interior and Prairie region of Canada, flood seasonality aids a few sites in forming homogeneous regions. One successful example is taken from site WSC gauge ID 06DA004 - Geikie River below Wheeler River, located in northern Saskatchewan. A homogeneous region for this site is developed based on flood seasonality that consists of 12 sites. The geographical locations of region members are shown on Figure 14, where sites cover a wide spectrum of longitude with most located from mid to high latitude regions. The variabilities of physiography (i.e., tundra, boreal grass, and forest), topography (i.e., from flat to rolling plain), and climate (i.e., generally cold sub-arctic) are smaller in spatial scale, particularly compared to coastal and mountainous areas. The average dates and regularity of flood events of this regions members are presented on Figure 15. The flood seasonality space indicates that all sites are dominated by a nival response, and have good consistency based on annual date of occurrence.

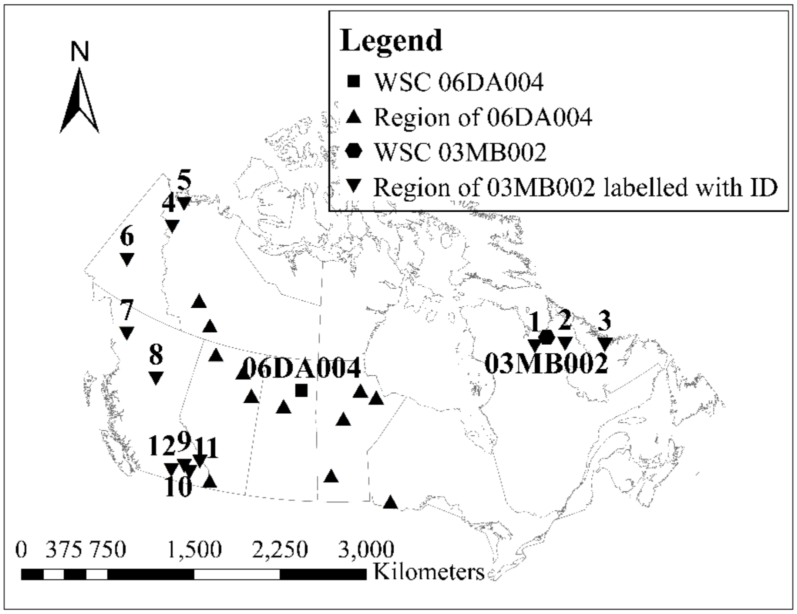

Figure 14: Developed regions based on flood seasonality for WSC sites 06DA004 and 03MB002.

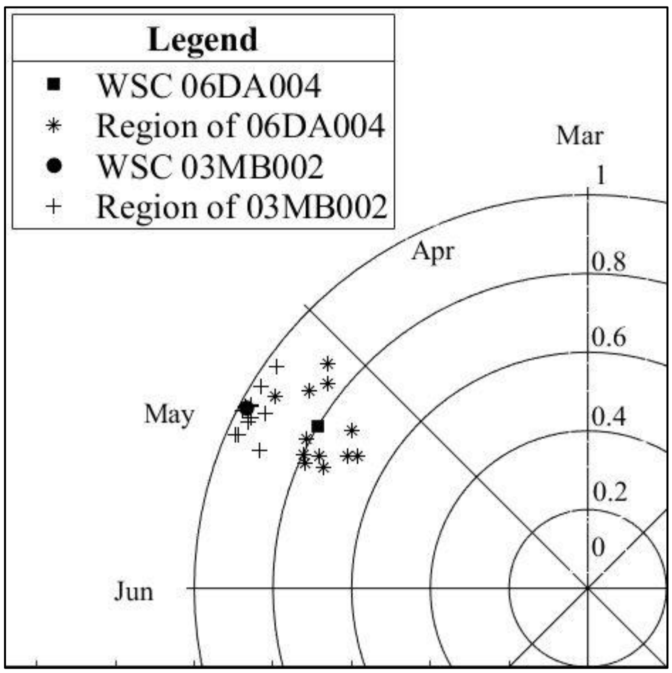

Figure 15: Positioning of WSC sites 06DA004 and 03MB002 in flood seasonality space.

The aforementioned spatial results show that sites from the western mountains to more interior regions also experience considerable difficulties in forming homogeneous regions.

Physiographic variables and flood seasonality obtain slightly better performance, and thus are considered as playing key roles in the flood response in these regions. Since each of the two attributes is associated with limited aspects of flood behaviour, the results are generally not satisfactory. Using the two attributes together, however, may potentially help to improve the identification of hydrologically similar sites.

An unsuccessful example of using flood seasonality is seen at WSC gauge ID 03MB002 – Whale River at 40.2 km from the Mouth, located in northern Quebec. A heterogeneous region is developed based on similarity of flood seasonality that is comprised of 12 sites located across Canada. Figure 14 and 15 present the 12 sites in terms of geographical location and flood seasonality space, respectively. The map indicates that all sites are spatially situation near the ocean, and correspondingly, all sites receive a substantial amount of mean annual precipitation. The site closest to the ocean (more southern) generally obtains the highest mean annual precipitation. The flood seasonality space indicates that all sites in the region share fairly similar annual average dates of flooding. The average peak flood dates correspond to the spring snowmelt period, suggesting most floods are driven by nival responses. Flood seasonality is not capable of developing a homogeneous regions in this case, as there are some noticeable differences of hydrological characteristics between these sites. These differences can be easily reflected by their physiographic variables presented in Table 3. Four sites in southeast British Columbia have high mean annual precipitation and noticeably higher mean annual temperature compared to other sites, suggesting potential rain-on-snow influence for these sites during the snowmelt period. A discrepancy in basin areas between these sites is revealed; five sites are small basins with areas $<500\ km^2$, whereas the 7 other sites, and the target site, have basin areas ranging between $3500\ km^2$ to $49000\ km^2$. Four of the five small basins are located in southeast

British Columbia. The basin compactness ratio: $Basin\ Area/Perimeter^2$ is a surrogate for time of peak flow concentration. It is found that the smaller basins are also subject to higher compactness ratio, indicating a much shorter period of water routing than in the larger basins. In addition, a wide spectrum of mean basin slopes across these sites is found. Some sites are also subject to Rocky Mountains topography and are much more steeply sloping than the other sites.

Table 4: Physiographic variables for WSC 03MB002 and its region formed based on flood seasonality.

WSC ID	Province	ID in Figure 14	Basin area	Basin perimeter	Compactness ratio	Mean basin slope	Mean annual precipitation	Mean annual temperature
			[km²]	[km]	Area/Perimeter² [%]	[%]	[mm]	[°C]
03MB002	QC	Target site	29124	1416.8	1.5	2.6	732.3	-4.6
03KC004	QC	1	39371	1901.0	1.1	3.3	654.9	-5.4
03MD001	QC	2	22440	1626.8	0.8	3.5	815.2	-4.4
03NF001	NL	3	7322	780.6	1.2	8.2	881.0	-3.8
10LA002	NT	4	18746	1173.0	1.4	18.1	385.9	-6.6
10ND002	NT	5	65	44.4	3.3	3.5	220.0	-8.7
09BC001	YT	6	48867	1751.5	1.6	15.9	456.5	-3.9
08CD001	BC	7	3555	488.9	1.5	7.4	562.1	-1.8
07EC002	BC	8	5559	597.8	1.6	23.2	648.7	0.2
08NE006	BC	9	330	103.3	3.1	45.4	1325.7	1.4
08NF001	BC	10	416	105.1	3.8	31.3	796.1	0.0
08NH005	BC	11	442	130.5	2.6	44.5	1217.7	1.2
08NN015	BC	12	233	100.3	2.3	12.1	941.7	2.1

The above analysis shows an example of using flood seasonality to develop a heterogeneous region, and then using physiographic variables to diagnose the reasons causing region heterogeneity. This process can be incorporated into the initial region formation and region revision process, where initial regions are formed using one attribute, and other attributes are used to revise region membership for regional improvement.

4.4 Results of predictive performance of attributes

The results of bias and RMSE are presented in Table 5, which statistically measure the regional quantile of each attribute to a "true" quantile at return periods between 20 to 45 years. In general, the biases of all attributes range from -0.6% to 3.7% relatively to the "true" quantile. As all these biases are reasonably close to zero, the accuracy of regional estimation is generally satisfactory. More positive biases are found across all attributes, suggesting that the RFFA generally overestimates the "true" flood quantile. Flood seasonality and physiographic variables have noticeably larger bias than geographical proximity, precipitation pattern, and temperature pattern. For the RMSE measures, RMSE generally increases with the increase in return period across all attributes. A similar RMSE among attributes is found for return periods equal to and below 35 years. For return periods of 40 and 45 years, attributes of seasonality and physiographic variables show noticeably larger RMSE than geographical proximity, precipitation pattern, and temperature pattern.

The overall measure of bias and RMSE show that geographical proximity, precipitation pattern, and temperature pattern produce regions that predict flood quantiles slightly better than flood seasonality and physiographic variables. But all attributes are deemed suitable for regional flood quantile estimation.

Table 5: The bias and RMSE measures for quantiles produced by regionalized estimates.

Statistic	Return period	Regionalization attribute				
		Geographic proximity	Flood seasonality	Physiographic variables	Precipitation pattern	Temperature pattern
Bias	20	***0.3%***	1.0%	0.9%	0.6%	0.6%
	25	0.8%	2.0%	0.5%	***0.4%***	1.0%
	30	***0.5%***	2.5%	3.3%	1.1%	1.9%
	35	1.1%	3.1%	2.8%	-0.6%	***-0.1%***
	40	0.9%	1.8%	2.3%	***0.2%***	0.8%
	45	0.1%	2.1%	3.7%	***0.004%***	-0.2%
RMSE	20	6.5%	6.7%	6.6%	***6.4%***	6.7%
	25	7.9%	7.7%	***7.6%***	7.7%	***7.6%***
	30	***7.5%***	7.9%	8.7%	8.5%	8.0%
	35	9.5%	8.9%	9.6%	8.4%	***8.1%***
	40	9.4%	9.7%	10.0%	8.8%	***8.7%***
	45	***8.6%***	11.1%	13.0%	9.9%	10.1%

Bold italicized numbers indicate the best outcome for the return period.

Geographical proximity, precipitation pattern, and temperature pattern perform slightly better than flood seasonality and physiographic variables possibly because regions developed based on the first three attributes often have a higher degree of geographical cohesion. Sites that are geographically neighbouring are often associated with a wide spectrum of hydrological similarities. Flood seasonality and physiographic measures often group sites across wide geographical extent, therefore, the degree of hydrological similarity between grouping sites might be lower than the other three attributes, resulting in a slightly poorer, but still acceptable, regional flood estimation.

4.5 Conclusions

This study presents a Canada-wide application to investigate five distinct attributes for their relevance in developing homogeneous regions for regional flood estimation. For each attribute, relevance is assessed based on the number of homogeneous regions that each attribute produces, and the accuracy of flood prediction estimated based on the region. The production of homogeneous regions is made based on the similarity measure in the ROI approach and the assistance of an ARRA. The developed homogeneous regions are applied with the L-Moment index flood method to estimate regional flood quantiles. These quantiles are compared to a "true" quantile for assessing the accuracy of regional estimation.

Spatial pattern analysis reveals that homogeneous regions are more readily formed when sites are clustered in small geographic regions, regardless of which attribute is used. For sites that are scattered or sparsely distributed, it is generally difficult to form homogeneous regions regardless of the attribute that is considered. Precipitation pattern is able to produce the greatest number of homogeneous regions in comparison with other attributes, at the Canada-wide scale. This is potentially because Canadian floods can be broadly classified into rainfall, snowmelt, and rain-on-snow driven events, and precipitation pattern is capable of discriminating among these drivers. Temperature pattern produces the second largest number of homogeneous regions, possibly because floods in most study sites in Canada are snowmelt or rain-on-snow driven, and the monthly temporal pattern of temperature is highly relevant to the timing and magnitude of a flood response. In addition, a good number of sites in Arctic Canada are able to develop homogeneous regions based on the temperature pattern attribute only. This finding suggests the strong connection between temperature pattern and nival regime flood response.

Geographic proximity has an advantage for sites that are spatially clustered, with this advantage also applying to precipitation and temperature pattern attributes as monthly temporal sequencing is typically more consistent across smaller spatial domains. These two attributes are considered superior to geographical proximity because they are additionally capable of grouping sites that are hydrologically similar but not geographically neighbouring.

Although the number of homogeneous regions produced in total is not desirable, flood seasonality and physiographic variables reveal good potential of use in selected geographic areas of Canada. Compared to the other three attributes, physiographic variables show higher relevance for flood response from western mountainous terrain; and flood seasonality has more relevance for the broad plains of the interior. It is important to explore new techniques that will produce desired output for these areas. A resulting case analysis inspires the combined use of flood seasonality and physiographic variables for the development of flood regions. Flood seasonality could be used to develop an initial flood region, and then apply physiographic variables to improve region homogeneity by further revising the membership. This proposal can be a future research topic.

The predictive assessment shows that all attributes are able to estimate accurate flood quantiles. Geographical proximity, precipitation pattern, and temperature pattern perform slightly better than flood seasonality and physiographic variables in terms of precision and uncertainty of estimation, possibly because regions developed by the first three attributes obtain a higher degree of geographical cohesion, making a higher degree of hydrological similarity in the flood region.

The following recommendations are made. Precipitation pattern is preferred for general use in Canada. In the circumstances that monthly precipitation pattern is ineffective to group homogeneous regions, the following attributes can be taken into consideration with certain conditions. For the Arctic regions where floods are exclusively driven by nival responses, temperature pattern should be considered. For sites in the interior plains, flood seasonality may be considered. For regions where floods are considerably influenced by local physiographic characteristics, such as in the mountains in British Columbia, using physiographic variables to group regions can be considered. Physiographic variables, however, should be selected for characterization of local flood regime.

It is important to mention that the number of sites used in this study is only a small fraction of the Canadian gauging network. Using more sites in future studies may verify the results of this study.

Findings and recommendations of this study are considered valuable input for the ongoing development of the Canadian Flood Estimation Manual. This study also supports potential future work in terms of expanding new techniques related to RFFA, such as techniques for ungauged basins, and development of composite attributes that are more relevant to annual peak flood regimes in Canada.

4.6 Acknowledgements

The authors would like to thank Professor Donald H. Burn for providing valuable advice and manuscript editing for this study. This work was financially supported by the Natural Sciences and Engineering Research Council of Canada under the Canadian FloodNet project (Grant number – NETGP 451456) and the University of Manitoba. Erika Klyszejko from Environment and Climate Change Canada is acknowledged for providing data used in this study. Water Survey of Canada is acknowledge for providing hydrometric data used in this study. The authors would like to dedicate this study to the memory of Professor Peter Rasmussen who was a key contributor to the journey of this research.

Chapter 5 Conclusions

5.1 General summary

Based on two national-scale research applications, this thesis provides new insights into at-site and regional FFA in Canada, with key findings that can be referenced as general guidance for Canadian use. These findings will make significant contributions to the development of the Canadian flood estimation manual by the FloodNet Strategic Network.

In the first application, the GEV distribution is identified as the preferred flood frequency distribution for at-site FFA in Canada on the basis of goodness-of-fit analyses. Compared to other candidate distributions (all commonly used for FFA), the GEV distribution is acceptable for the largest number of both Canadian annual maximum flow and instantaneous peak flow samples, and obtains better quality of sample fitting. In addition, Prairie streamflows are more difficult to model using any of the candidate distributions because of their wide-spanning in magnitude of the flood values and the candidate distributions are not flexible enough to model the wide-spanning. This is likely a direct result of the dynamic contributing areas of the Prairie Pothole region involved in generating floods, resulting in observed flood values that are no longer produced by an identical population of flood frequency.

In the second application, five distinct attributes are investigated in a national scale application for their relevance to form homogeneous flood regions for use for regional FFA. For each attribute, the number of homogeneous groups the attribute produced and the accuracy of regional flood estimates are assessed. It is found that homogeneous regions are more likely to be produced in geographic areas with a high density of gauging stations regardless of which attribute is used. For regions where gauging sites are sparse and scattered, homogeneous regions are generally difficult to form across all attributes. All attributes are able to satisfactorily estimate flood quantiles, although attributes of geographical proximity, precipitation pattern, and temperature pattern perform better than attributes of flood seasonality and physiographic variables in terms of bias and uncertainty. The precipitation pattern is preferred for general use in Canada partly because the attribute values are highly sensitive to identifying major flood response drivers across Canada (i.e., nival, pluvial, and mixed). In circumstances where monthly precipitation pattern is ineffective to form homogeneous regions, however, the following attributes can be taken into consideration with certain conditions. For sites with a high density of gauging stations in a small region, geographical proximity and temperature pattern may be considered. For the Arctic region where floods are exclusively driven by nival response, temperature pattern can be considered. For the massive interior and northern Canadian region where sites are sparse and scattered, flood seasonality is recommended. For the mountainous region, such as interior British Columbia, physiographic variables can be considered. In addition, results show that a combined use of flood seasonality and physiographic variables may be effective to develop homogeneous regions for sites in the interior and northern Canadian regions.

5.2 Recommendations for future research

Although this work provides important insight into applications of FFA across Canada, several future research directions can be considered to either increase the validity of this study, or develop new knowledge related to FFA in Canada.

- Since flood samples used in both studies are limited to RHBN stations, increasing the number of flood samples to include other unregulated hydrometric stations will increase the validity, and applicability of the findings from this study.

- Instrumenting more of northern Canada is highly recommended, since relatively few stations from high latitudes are available for FFA, significantly adding to the uncertainty of the results for these regions. It should be noted that significant large infrastructure (e.g., hydropower dams) reside in high latitude, sparsely gauged regions, which rely on extreme flood estimation (i.e., 100- to 500-year quantiles).
- Regional FFA has become increasingly popular for flood quantile estimation. A preferred regional flood frequency distribution in Canada is also worthy of investigation. Methodologies used in both applications of this thesis can be combined together to achieve this. First, applying techniques illustrated in the second application can develop homogeneous flood regions with sufficiently long records. Then goodness-of-fit analyses used in the first application can be employed within each region to assess the fit of candidate distributions. A preferred regional flood frequency distribution is one who's fits are acceptable for the largest number of homogeneous flood regions.
- The above technique could also be tested in high-latitude regions, specifically for extreme flood quantile estimation.
- Compared to other geographical regions in Canada, the Prairie region is more problematic in both finding a preferred distribution, and in developing homogeneous flood regions. A more in-depth FFA study for the Prairie region, using more gauging sites, or conducting individual site-by-site investigation, would be beneficial to uncovering the reasons for the underlying difficulties.
- In the second study, results show that the western mountainous region, interior, and northern lands of Canada have considerable difficulty in forming homogeneous regions regardless of which candidate attribute is used. The ability to address this problem is

important, as these areas cover the majority of the Canadian landmass. A case analysis shows that using flood seasonality and physiographic variables together may benefit the development of homogeneous regions. Using flood seasonality to develop the initial flood region, and then using physiographic variables to improve region homogeneity by revising region membership may be advantageous. In addition, researching a composite attribute that comprised of flood seasonality and physiographic variables should also improve the production of forming homogeneous regions, as more flood-related characteristics are used to measure the similarity between catchments. These ideas should be explored in future research.

- Regional FFA is also applicable for ungauged sites. Some of the flood-related attributes considered in the second application can be obtained for ungauged sites using GIS or remotely-sensed imagery. A regression model that constructs a relationship between the attribute(s) and estimated flood quantiles could be further investigated. Such a study would be highly beneficial for FFA in data-sparse or ungauged regions.

References

Ahmad, M.I., C.D. Sinclair, and B.D. Spurr. 1988. Assessment of flood frequency models using empirical distribution function statistics. *Water Resources Research* 24 (8): 1323–1328.

Ahmad, M.I., C.D. Sinclair, and A. Werritty. 1988. Log-logistic flood frequency analysis. *Journal of Hydrology* 98: 205–224.

Anderson, T.W., and D.A. Darling. 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics* 23 (2): 193–212.

Anderson, T.W., and D.A. Darling. 1954. A test of goodness of fit. *Journal of the American Statistical Association* 49 (268): 765–769.

Atiem, I.A., and N.B. Harmancioğlu. 2006. Assessment of regional floods using L-moments approach: The case of the River Nile. *Water Resources Management* 20 (5): 723–747.

Aucoin, F., D. Caissie, N. El-Jabi, and N. Turkkan. 2011. Flood frequency analyses for New Brunswick rivers. Canadian Technical Report of Fisheries and Aquatic Sciences 2920, xi + 77p.

Baldassarre, G. Di, F. Laio, and A. Montanari. 2009. Design flood estimation using model selection criteria. *Physics and Chemistry of the Earth, Parts A/B/C* 34 (10–12): 606–611.

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

Bobée, B., G. Cavadias, F. Ashkar, J. Bernier, and P. Rasmussen. 1993. Towards a systematic approach to comparing distributions used in flood frequency analysis. *Journal of Hydrology* 142 (1–4): 121–136.

Bobée, B., and P.F. Rasmussen. 1995. Recent advances in flood frequency analysis. *Reviews of*

Geophysics 33 (2 S): 1111–1116.

Brimley, B., J.F. Cantin, D. Harvey, M. Kowalchuk, P. Marsh, T.M.B.J. Ouarda, B. Phinney, et al. 1999. *Establishment of the Reference Hydrometric Basin Network (RHBN) for Canada*. Ottawa.

Burn, D.H. 1988. Delineation of groups for regional flood frequency analysis. *Journal of Hydrology* 104 (1–4): 345–364.

Burn, D.H. 1989. Cluster analysis as applied to regional flood frequency. *Journal of Water Resouces Planning and Management* 115 (5): 567–582.

Burn, D.H. 1990a. An appraisal of the “region of influence” approach to flood frequency analysis. *Hydrological Sciences Journal* 35 (2): 149–165.

Burn, D.H. 1990b. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research* 26 (10): 2257–2265.

Burn, D.H. 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology* 202 (1–4): 212–230.

Burn, D.H., and N.K. Goel. 2000. The formation of groups for regional flood frequency analysis. *Hydrological Sciences Journal* 45 (1): 97–112.

Burn, D.H., J. Hannaford, G.A. Hodgkins, P.H. Whitfield, R. Thorne, and T. Marsh. 2012. Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow. *Hydrological Sciences Journal* 57 (8): 1580–1593.

Burn, D.H., and P.H. Whitfield. 2016. Changes in floods and flood regimes in Canada. *Canadian Water Resources Journal* 41 (1–2): 139–150.

Burn, D.H., P.H. Whitfield, and M. Sharif. 2016. Identification of changes in floods and flood regimes in Canada using a peaks over threshold approach. *Hydrological Processes* 30 (18):

3303–3314.

Burn, D.H., Z. Zrinji, and M. Kowalchuk. 1997. Regionalization of catchments for regional flood frequency analysis 2 (2): 76–82.

Buttle, J.M., D.M. Allen, D. Caissie, B. Davison, M. Hayashi, D.L. Peters, J.W. Pomeroy, S. Simonovic, A. St-Hilaire, and P.H. Whitfield. 2016. Flood processes in Canada: Regional and special aspects. *Canadian Water Resources Journal* 41 (1–2): 7–30.

Castellarin, A., D.H. Burn, and A. Brath. 2001. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology* 241 (3–4): 270–285.

Chowdhury, J.U., J.R. Stedinger, and L. -H Lu. 1991. Goodness-of-fit tests for regional generalized extreme value flood distributions. *Water Resources Research* 27 (7): 1765–1776.

Coles, S.G. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

Cunnane, C. 1978. Unbiased plotting positions - a review. *Journal of Hydrology* 37 (3–4): 205–222.

Cunnane, C. 1985. Factors affecting choice of distribution for flood series. *Hydrological Sciences Journal* 30 (1): 25–36.

Cunnane, C. 1989. Statistical distributions for flood frequency analysis. In *Operational Hydrology Report No. 33, WMO - No. 718*. Geneva, Switzerland: World Meteorological Organization.

Dalrymple, T. 1950. Regional flood frequency. *Highway Research Board Research Report 11-B*, 4–20.

Dalrymple, T. 1960. Flood-Frequency Analyses. Manual of Hydrology Part 3. Flood-flow techniques. *Geological Survey Water-Supply Paper* 1543–A: 80.

Durocher, M., S.M. Zadeh, D.H. Burn, and F. Ashkar. 2018. Comparison of automatic procedures for selecting flood peaks over threshold based on goodness-of-fit tests. *Hydrological Processes* 32 (18): 2874–2887.

England Jr., J.F., T.A. Cohn, B.A. Faber, J.R. Stedinger, W.O. Thomas Jr., A.G. Veilleux, J.E. Kiang, and R.R. Mason Jr. 2018. Guidelines for determining flood flow frequency – Bulletin 17C. In *U.S. Geological Survey Techniques and Methods, Book 4, Chap. B*. Reston, Virginia: U.S. Geological Survey.

Eslamian, S.S., and E.Z. Hosseinipour. 2010. A modified region of influence approach for flood regionalization. In *World Environmental and Water Resources Congress 2010: Challenges of Change - Proceedings of the World Environmental and Water Resources Congress 2010*, 2388–2414. Providence, Rhode Island, United States.

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

Fill, H.D., and J.R. Stedinger. 1995. Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. *Journal of Hydrology* 166 (1–2): 81–105.

FloodNet. 2018. Floodnet – NSERC Network – Enhanced flood forecasting and management capacity in Canada.

Fortin, V., B. Bobée, and J. Bernier. 1997. Rational approach to comparison of flood distributions by simulation. *Journal of Hydrologic Engineering* 2 (3): 95–103.

Greis, N.P., and E.F. Wood. 1981. Regional flood frequency estimation and network design. *Water Resources Research* 17 (4): 1167–1177.

Griffis, V.W., and J.. R. Stedinger. 2007a. Log-Pearson type 3 distribution and its application in flood frequency analysis. II: Parameter estimation methods. *Journal of Hydrologic*

Engineering 12 (5): 492–500.

Griffis, V.W., and J.. R. Stedinger. 2007b. Log-Pearson type 3 distribution and its application in flood frequency analysis . I : Distribution characteristics. *Journal of Hydrologic Engineering* 12 (5): 482–491.

Gringorten, I.I. 1963. A plotting rule for extreme probability paper. *Journal of Geophysical Research* 68 (3): 813–814.

Guo, S.L. 1990. A discussion on unbiased plotting positions for the general extreme value distribution. *Journal of Hydrology* 121 (1–4): 33–44.

Haktanir, T. 1991. Statistical modelling of annual maximum flows in Turkish rivers. *Hydrological Sciences Journal* 36 (4): 367–389.

Harvey, K.D., P.J. Pilon, and T.R. Yuzyk. 1999. Canada's Reference Hydrometric Basin Network (RHBN). In *Partnerships in Water Resources Management, Proceedings of the CWRA 51st Annual Conference*. Nova Scotia.

Heo, J.-H., H. Shin, W. Nam, J. Om, and C. Jeong. 2013. Approximation of modified Anderson-Darling test statistics for extreme value distributions with unknown shape parameter. *Journal of Hydrology* 499: 41–49.

Hosking, J.R.M. 1990. L-Moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 52 (1): 105–124.

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

Hosking, J.R.M., and J.R. Wallis. 1997. *Regional Frequency Analysis - An Approach Based on L-Moments*. Cambridge, UK: Cambridge University Press.

Hosking, J.R.M., J.R. Wallis, and E.F. Wood. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*.

Jakob, M., and M. Church. 2011. The trouble with floods. *Canadian Water Resources Journal* 36 (4): 287–292.

Jenkinson, A.F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81 (348): 158–171.

Jennings, M.E., and M.A. Benson. 1969. Frequency Curves for Annual Flood Series with Some Zero Events or Incomplete Data. *Water Resources Research* 5 (1): 276–280.

Karim, A., and J.U. Chowdhury. 1993. Comparison of five distributions for suitability in flood frequency analysis in Bangladesh. In *International Conference on Hydrology and Water Resources,* 567–575. New Delhi, India.

Karim, A.M., and J.U. Chowdhury. 1995. A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrological Sciences Journal* 40 (1): 55–66.

Karim, M.A., and J.U. Chowdhury. 1995. A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrological Sciences Journal* 40 (1): 55–66.

Kottegoda, N.T. 1984. Investigation of outliers in annual maximum flow series. *Journal of Hydrology* 72 (1–2): 105–137.

Lettenmaier, D.P., and K.W. Potter. 1985. Testing flood frequency estimation methods using a regional flood generation model. *Water Resources Research* 21 (12): 1903–1914.

Lu, Y. 2016. Regional flood frequency analysis for Newfoundland and Labrador using the L-Moments index-flood method. *Memorial University of Newfoundland*. Memorial University of Newfoundland.

McKenney, D.W., J.H. Pedlar, P. Papadopol, and M.F. Hutchinson. 2006. The development of 1901-2000 historical monthly climate models for Canada and the United States. *Agricultural and Forest Meteorology* 138 (1–4): 69–81.

Mediero, L., T.R. Kjeldsen, N. Macdonald, S. Kohnova, B. Merz, S. Vorogushyn, D. Wilson, et al. 2015. Identification of coherent flood regions across Europe by using the longest streamflow records. *Journal of Hydrology* 528: 341–360.

Merz, R., and G. Blöschl. 2005. Flood frequency regionalisation - Spatial proximity vs. catchment attributes. *Journal of Hydrology* 302 (1–4): 283–306.

Merz, R., U. Piock-Ellena, G. Blöschl, and D. Gutknecht. 1999. Seasonality of flood processes in Austria. In *Proc. of IUGG 99 Symposium HS1 "Hydroiogical Extremes: Understanding, Predicting, Mitigating" IAHS Publ. No. 255*, 273–278.

Moudrak, N., and B. Feltmate. 2017. *Preventing Disaster before It Strikes: Developing a Canadian Standard for New Flood-Resilient Residential Communities*. *Prepared for Standards Council of Canada*.

Muhammad, A., G.R. Evenson, T.A. Stadnyk, A. Boluwade, S.K. Jha, and P. Coulibaly. 2018. Assessing the importance of potholes in the Canadian Prairie region under future climate change scenarios. *Water* 10 (11): 1–19.

Muhammad, A., G.R. Evenson, T.A. Stadnyk, A. Boluwade, S.K. Jha, and P. Coulibaly. 2019. Impact of model structure on the accuracy of hydrological modeling of a Canadian Prairie watershed. *Journal of Hydrology: Regional Studies* 21 (November 2018). Elsevier: 40–56.

Mutua, F.M. 1994. The use of the Akaike Information Criterion in the identification of an optimum flood frequency model. *Hydrological Sciences Journal* 39 (3): 235–244.

Nathan, R.J., and T.A. McMahon. 1990. Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology* 121 (1–4): 217–238.

Nathan, R.J., and P.E. Weinmann. 1991. Application of at-site and regional flood frequency analyses. In *International Hydrology and Water Resources Symposium*, 769–774.

Natural Environment Research Council. 1975. *Flood Studies Report*. London: Hydrological Studies.

Natural Resources Canada. 2018. Historical monthly climate grids for North America.

Noto, L. V., and G. La Loggia. 2009. Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water Resources Management* 23 (11): 2207–2229.

Önöz, B., and M. Bayazit. 1995. Best-fit distributions of largest available flood samples. *Journal of Hydrology* 167: 195–208.

Pearson, C.P. 1991. New Zealand regional flood frequency analysis using L-Moments. *Journal of Hydrology (New Zealand)* 30 (2): 53–64.

Peel, M.C., Q.J. Wang, R.M. Vogel, and T.A. McMahon. 2001. The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal* 46 (1): 147–155.

Pilon, P.J., and K. Adamowski. 1992. The value of regional information to flood frequency analysis using the method of L-moments. *Canadian Journal of Civil Engineering* 19 (1): 137–147.

Potter, K.W., and D.P. Lettenmaier. 1990. A comparison of regional flood frequency estimation methods using a resampling method. *Water Resources Research* 26 (3): 415–424.

Rahman, A.S., A. Rahman, M.A. Zaman, K. Haddad, A. Ahsan, and M. Imteaz. 2013. A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural Hazards* 69 (3): 1803–1813.

Rao, A.R., and K.H. Hamed. 2000. *Flood Frequency Analysis*. 1st Editio. Boca Raton, Florida: CPC Press.

Rao, A.R., and V. V. Srinivas. 2006. Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology* 318 (1–4): 57–79.

Robson, A., and D. Reed. 1999. *Statistical Procedures for Flood Frequency Estimation. Flood Estimation Handbook. Vol. 3. Institute of Hydrology*. Wallingford, U.K.

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

Salinas, J.L., A. Castellarin, A. Viglione, S. Kohnová, and T.R. Kjeldsen. 2014. Regional parent flood frequency distributions in Europe - Part 1: Is the GEV model suitable as a pan-European parent? *Hydrology and Earth System Sciences* 18 (11): 4381–4389.

Shin, H., Y. Jung, C. Jeong, and J.-H. Heo. 2012. Assessment of modified Anderson-Darling test statistics for the Generalized Extreme Value and Generalized Logistic distributions. *Stochastic Environmental Research and Risk Assessment* 26 (1): 105–114.

Shook, K.R., and J.W. Pomeroy. 2011. Memory effects of depressional storage in Northern Prairie hydrology. *Hydrological Processes* 25 (25): 3890–3898.

Stedinger, J.R., and L.H. Lu. 1995. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics* 9 (1): 49–75.

Stephens, M.A. 1974. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* 69 (347): 730–737.

U.S. Water Resources Council. 1982. *Guidelines for Determining Flood Flow Frequency. Bulletin 17B*. Washington, D.C.: Hydrology Committee.

Vogel, R.M., and N.M. Fennessey. 1993. L moment diagrams should replace product moment diagrams. *Water Resources Research* 29 (6): 1745–1752.

Vogel, R.M., T.A. McMahon, and F.H.S. Chiew. 1993. Floodflow frequency model selection in Australia. *Journal of Hydrology* 146: 421–449.

Vogel, R.M., and I. Wilson. 1996. Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *Journal of Hydrologic Engineering* 1 (2): 69–76.

Wallis, J.R., and E.F. Wood. 1985. Relative accuracy of log Pearson III procedures. *Journal of Hydraulic Engineering* 111 (7): 1043–1056.

Water Survey of Canada. 2016. Environment Canada Data Explorer - HYDAT Database. Environment and Climate Change Canada (ECCC).

Watt, W.E. 1989. *Hydrology of Floods in Canada - a Guide to Planning and Design - NRC Publications Archive - National Research Council Canada*. Ottawa, Canada: National Research Council Canada, Associate Committee on Hydrology.

Whitfield, P.H., D.H. Burn, J. Hannaford, H. Higgins, G.A. Hodgkins, T. Marsh, and U. Looser. 2012. Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal* 57 (8): 1562–1579.

Wiltshire, S.E. 1985. Grouping basins for regional flood frequency analysis. *Hydrological Sciences Journal* 30 (1): 151–159.

Yang, T., C.-Y. Xu, Q.-X. Shao, and X. Chen. 2010. Regional flood frequency and spatial patterns analysis in the Pearl River Delta region using L-Moments approach. *Stochastic Environmental Research and Risk Assessment* 24 (2): 165–182.

Yue, S., and C.Y. Wang. 2004. Possible regional probability distribution type of Canadian annual streamflow by L-Moments. *Water Resources Management* 18 (5): 425–438.

Zrinji, Z., and D.H. Burn. 1994. Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology* 153 (1–4): 1–21.

Zrinji, Z., and D.H. Burn. 1996. Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management* 122 (4): 245.

Appendix

The inverse cumulative distribution function (invCDF) of the 4-parameter Kappa distribution is used in Monte-Carlo simulation experiments for the L-Moment Z^{DIST} (Chapter 3) and L-Moment homogeneity test (Chapter 4).

According to Robson and Reed 1999, the inverse cumulative distribution function is defined as:

$$x = \mu + \frac{\alpha}{k}\left[1 - \left(\frac{1 - F^h}{h}\right)^k\right] \tag{41}$$

where F is the non-exceedance probability. x is the quantile variable. μ and α are the location and scale parameters, respectively. k and h are the two shape parameters.

The bounds for the Kappa distribution are as follow:

$$\mu + \frac{\alpha}{k}(1 - h^{-k}) \leq x \leq \mu + \frac{\alpha}{k} \qquad k > 0, \mathrm{h} > 0 \tag{42}$$

$$\mu + \frac{\alpha}{k}(1 - h^{-k}) \leq x \leq \infty \qquad k \leq 0, \mathrm{h} > 0 \tag{43}$$

$$-\infty \leq x \leq \mu + \frac{\alpha}{k} \qquad k > 0, \mathrm{h} \leq 0 \tag{44}$$

$$-\infty \leq x \leq \infty \qquad k = 0, \mathrm{h} \leq 0 \tag{45}$$

$$\mu + \frac{\alpha}{k} \leq x \leq \infty \qquad k < 0, \mathrm{h} \leq 0 \tag{46}$$

The following pages provide key MATLAB scripts that have been used in the Chapter 4.

```
function [Best_Selec,H]=Develop_Flood_Regions( T,dis,mySta_Flood,NYr,...

    H_defined_homo,Lmax)
%{
Author: Ziyang Zhang.
Script completion date: Dec 16th, 2017.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
This script develop homogeneous flood regions for each imported site. Sites
with annual maximum flood data and distance of attribute similarity are
imported in this script. The homogeneous region for each imported site is
developed based on distance of attribute similarity and an Automatic Region
Revision Algorithm (ARRA).

See Hosking and Wallis (1997) for reference of the L-Moments Homogeneity
test.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Input variables:
(Assume the number of imported sites is n.)
---------------------------------------------------------------------------
T: A value. Desired region size of the flood region. If you want every
imported site to develop a flood region with 500-station-year record, then
T=500. The target site is not included in its flood region.
---------------------------------------------------------------------------
dis: A square matrix (n by n). Distance of attribute similarity between
the imported sites.

An example:
        Site 1  Site 2    Site 3   Site 4   ...    Site n
Site 1  0       0.0925    0.5180   0.6897   ...     ...
Site 2  0.0925  0         0.3091   0.4063   ...     ...
Site 3  0.5180  0.3091    0        0.0661   ...     ...
Site 4  0.6897  0.4063    0.0661   0        ...     ...
...
Site n  ...     ...          ...      ...     ...     0

So the distance between site 1 and site 2 is 0.0925. The distance between
site 1 and site 3 is 0.5108. The pattern is repeated for every two sites.
---------------------------------------------------------------------------
mySta_Flood: A matrix (n by col). This matrix contains flood samples of
imported sites. Each row contains flood values for a given site.
Total number of rows equals to the total number of imported sites.

An example:
        Value 1     Value 2     Value 3     Value 4...  Value 99  Value 100
Site 1  108         55.8        72.2        0        0           0
Site 2  161         127         120         92.9        110        141
Site 3  29.7        24.8        18.9        5.32        10.6       0
...
Site n ...          ...         ...         ...         ...          ..

If site 1 only have 3 flood values, then values starting at position 4 to
the end are all sets to be zeros. In here, site 3 has 99 flood values in
its flood sample.

The number of columns (col) of this matrix should be greater than the
maximum record length (i.e., number of flood values in the sample) of flood
samples of imported sites.
```

```
---------------------------------------------------------------------------
NYr: A vector (n by 1). This vector indicates the record length of each
imported site.

For the above mySta_Flood example, A(1)=3, A(2)=100, A(3)=99 ... A(n)=....
---------------------------------------------------------------------------
H_defined_homo: A value. The target homogeneity statistic (i.e., the H
value) for the flood region. If H=1, then this script aims to develop flood
regions with homogeneity statistic H less and equal to 1.
---------------------------------------------------------------------------
Lmax: A value (1 by 2). The maximum number of applications of Automatic
Region Revision Algorithm (ARRA).

ARRA is only activated when the initial developed region (i.e., region
developed based on distances of attribute similarity only) is
heterogeneous.

An example: Lmax: [5,30]. The value 5 sets the maximum number of sites
that can be revised (remove the site causing most region heterogeneity with
replacement by a new site).

Once the site for which causes most region heterogeneity is removed, a new
site is added to the original region and check if the new region is more
homogeneous than the previous stage. If yes, a new flood region (i.e.,
revised flood region) is produced. If not, the new added site is removed
and then another new site is added and repeat the above process . The value
30 sets the maximum number of repetitions. The order of addition of new
sites is based on distance of attribute similarity that site with smaller
distance is added first.
Both Lmax(1) and Lmax(2) should be smaller than the total number of
imported sites.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Output variables:
---------------------------------------------------------------------------
Best_Selec: A matrix (40 by n). This matrix returns the outputs of
developed flood regions for each imported site. Each column presents the
result of a imported sites, hence there are n number of columns.

An example:
Target site:                1   2   3   4   ... n
1st member in the region    2   11  14  7   ...
2nd member in the region    5   3   74  24  ...
3rd member in the region    9   14  54  31  ...
4th member in the region    0   15  84  65  ...
5th member in the region    0   0   45  63  ...
...
39th member in the region   0   0   0   0   ...

For the 1st target site, its flood region constitutes 3 members. Imported
site number 2, 5, and 9 are members in this region. This matrix is set to
have 40 rows, which means the number of maximum
members can be stored is 39.
---------------------------------------------------------------------------
H: A vector (n by 1). This vector outputs the homogeneity statistic
(i.e., the H value) of each developed region.
```

```
An example:
H(1)=1.4    H(2)=0.5    H(3)=2.4 ...H(n)
The homogeneity statistic for the first developed region is 1.4.
The homogeneity statistic for the first developed region is 0.5.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
The following scripts are needed in order to execute this script.
1. LHomogeneity_Test.m
2.      KappaInvCDF
3.      mySampleLMoment
4.      SPFitKap1
5. Revise_Pro_V4.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%}

NYr=NYr(:);
le=length(NYr);
distt=[dis, NYr];
Best_Selec=zeros(40,le);
CurReviseTimeL2=0;
H_Loop=zeros(le,3);
Txx=T;

for j=1:le
    ReviseTime_L1=Lmax(1);
    T=Txx+NYr(j);
    disp('Current Station');
    disp(j);
    sta_dis=distt(:,[j,end]);
    Total_kick_sta_ID=[];
    H(j)=199;
    [~,I]=sort(sta_dis(:,1));
    sta_dis_s=sta_dis(I,:);
    myns=sta_dis_s(:,2);
    sta_dis_s(:,3)=I;
    H1=H(j);
    sta_dis_s2=sta_dis_s;
    H_last=100;
    CurReviseTimeL1=0;
    while CurReviseTimeL1<=ReviseTime_L1
        disp(CurReviseTimeL1);
        Ac_NYr=cumsum(sta_dis_s2(:,2));
        sta_take=find(Ac_NYr>T,1);
        if isempty(sta_take)
            break
        end
        sta_index=sta_dis_s2((1:sta_take),3);
        ns=NYr(sta_index);
        Sta_Flood=mySta_Flood(sta_index,:);
        Best_Selec_G_ID=sta_index;
        [ H1,ifail ] = LHomogeneity_Test( Sta_Flood,ns );
        if ifail~=0
            H1=99.99;
        end
        if H1>H_defined_homo
            [ H1,Best_Selec_G_ID,kick_sta_ID,add_sta_id,h_flag,...
                CurReviseTimeL2 ] = Revise_Pro_V4(H1,H_defined_homo,...
                sta_index,mySta_Flood,sta_dis_s2,T,j,NYr,...
```

```
                CurReviseTimeL1,Lmax);
        else
            h_flag=1;
        end
        if h_flag==1
            H(j)=H1;
            Best_Selec(:,j)=zeros(40,1);
            Best_Selec(1:length(Best_Selec_G_ID),j)=Best_Selec_G_ID;
            break
        end
        if h_flag==0
            H_last=H1;
            H(j)=H_last;
            Best_Selec(:,j)=zeros(40,1);
            Best_Selec(1:length(Best_Selec_G_ID),j)=Best_Selec_G_ID;
        end
        disp('Loop 1 H');
        disp(H1);
        Total_kick_sta_ID= [Total_kick_sta_ID;kick_sta_ID];
        sta_id_do_not_use=[Total_kick_sta_ID; Best_Selec_G_ID];
        sta_id_do_not_use=unique(sta_id_do_not_use);
        sta_dis_s2(find(ismember(sta_dis_s2(:,3),sta_id_do_not_use)),:)=[];
        temp1=sta_dis_s(find(ismember(sta_dis_s(:,3),Best_Selec_G_ID)),:);
        sta_dis_s2=[temp1;sta_dis_s2];
        CurReviseTimeL1=CurReviseTimeL1+1;
    end
    H_Loop(j,:)=[j,CurReviseTimeL1,CurReviseTimeL2];
end
H=H(:);
end
```

```
function [ H,ifail ] = LHomogeneity_Test( x,ns )
%{
Author: Ziyang Zhang.
Script completion date: August 16th, 2016.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
This script determines L-Moment homogeneity test statistic (H value) based
on L-CV.

See Hosking and Wallis (1997) for reference of the L-Moments Homogeneity
test.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Input variables:
(Assume the number of imported sites is n.)
---------------------------------------------------------------------------
x: A matrix (n by col). This matrix contains flood samples of imported
sites. Each row contains flood values for a given site. Total number of
rows equals to the total number of imported sites.

An example:
        Value 1     Value 2     Value 3     Value 4...  Value 99  Value 100
Site 1  108         55.8        72.2        0        0           0
Site 2  161         127         120         92.9        110         141
Site 3  29.7        24.8        18.9        5.32        10.6        0
...
Site n ...          ...         ...         ...         ...          ..

If site 1 only have 3 flood values, then values starting at position 4 to
the end are all sets to be zeros. In here, site 3 has 99 flood values in
its flood sample.

The number of columns (col) of this matrix should be greater than the
maximum record length (i.e., number of flood values in the sample) of flood
samples of imported sites.
---------------------------------------------------------------------------
ns: A vector (n by 1). This vector indicates the record length of each
imported site.

For the above x example, A(1)=3, A(2)=100, A(3)=99 ... A(n)=....
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Output variables:
---------------------------------------------------------------------------
H: A vector (n by 1). This vector outputs the homogeneity statistic
(i.e., the H value) of each developed region.

An example:
H(1)=1.4    H(2)=0.5    H(3)=2.4 ...H(n)
The homogeneity statistic for the first developed region is 1.4.
The homogeneity statistic for the first developed region is 0.5.
---------------------------------------------------------------------------
ifail: A value. This value indicates if the Kappa distribution is
infeasibly fitted in the process of determining homogeneity test.

If ifail is 0: the Kappa distribution is feasibly fitted. The H statistic
is valid.

If ifail is not equal to 0. The script is returned. The homogeneity test is
```

```
not executed.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
The following scripts are needed in order to execute this script.
1. KappaInvCDF
2. mySampleLMoment
3. SPFitKap1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%}

[nrow]=length(ns);
tYr=sum(ns);
poolflo=[];
V=[];
t=[];
l12t34=[];
V_sim=[];V_m=[];V_std=[];H=-99;
for i=1:nrow
    flod=[];
    flod=x(i,1:ns(i));
    poolflo=[poolflo, (flod./median(flod))];
    [l(i,:),t(i,:)]=mySampleLMoment(flod);
end
l12t34(:,1:2)=l(:,1:2);
l12t34(:,3:4)=t(:,3:4);
lcvi=t(:,2);
lcvr=ns(:)'*lcvi./tYr;
V=((ns(:)'*(lcvr-lcvi).^2)./tYr).^0.5;
Region_lt=ns(:)'*l12t34./tYr;
NewFitKappa=[1, lcvr,Region_lt(3),Region_lt(4)];
[ para, ifail ] = SPFitKap1( NewFitKappa );
if ifail~=0
    return
end
ii=[0, cumsum(ns(:)')];
V_sim=-99.*ones(1,500);
for j=1:500
    gene=rand(tYr,1);
    Flod_sim=KappaInvCDF( gene,para );
    l_s=[];
    t_s=[];
    lcvi_sim=[];
    lcvr_sim=[];
    for k=1:nrow
        flod2=[];
        flod2=Flod_sim((ii(k)+1):ii(k+1));
        [~,t_s(k,:)]=mySampleLMoment(flod2);
    end
    lcvi_sim=t_s(:,2);
    lcvr_sim=ns(:)'*lcvi_sim./tYr;
    V_sim(j)=((ns(:)'*(lcvr_sim-lcvi_sim).^2)./tYr).^0.5;
end
V_m=mean(V_sim);
V_std=std(V_sim);
H=(V-V_m)./V_std;
end
```

```
function [ xF ] = KappaInvCDF( F,para )
%{
Author: Ziyang Zhang.
Script completion date: August 16th, 2016.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
This script is the inverse cumulative distribution function of the
four-parameter Kappa distribution.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Input variables:
---------------------------------------------------------------------------
para: A vector (1 by 4). This vector inputs the parameters of a Kappa
distributions.

Column 1: Location parameter e.
Column 2: Scale parameter a.
Column 3: Shape parameter k.
Column 3: Shape parameter h.

para can also be a matrix, which each row represents a set of parameters
of a Kappa distribution.

An example:
        e       a       k       h
Row 1:  24      1.5     2.3     4.5
Row 2:  45      2.6     2.4     0.1
...     ...     ...     ...     ...

So row 2 indicates a Kappa distribution with location parameter 45, scale
parameter 2.6, shape parameter k 2.4, and shape parameter h 0.1,
respectively.
---------------------------------------------------------------------------
F: A value. This value inputs the probability of non-exceedance (i.e.,
probability in inverse cumulative distribution function) that you want to
determine its quantile value.

F can also be a vector, which each row is corresponding to the same row in
para.

An example:
For a given para:
        e       a       k       h
Row 1:  24      1.5     2.3     4.5
Row 2:  45      2.6     2.4     0.1

And a given F:
F(1)=0.5, F(2)=0.75

With these input parameters, you are trying to obtain the two quantiles.
The first quantile corresponds to a non-exceedance probability of 0.5 in a
Kappa distribution with parameters of e=24, a=1.5, k=2.3, and h=4.5. The
second quantile corresponds to a non-exceedance probability of 0.75 in a
Kappa distribution with parameters of e=45, a=2.6, k=2.4, and h=0.1.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Output variables:
---------------------------------------------------------------------------
xf: A value (or a vector, depends on your input variables). This value
```

```
(or vector) outputs the quantile value(s) based on your request.
%}
F=F(:);
xF=para(:,1)+(para(:,2)./para(:,3)).*(1-((1-F.^para(:,4))./para(:,4)) ...
    .^para(:,3));
end
```

```
function [l,t]=mySampleLMoment(rawdata)
%{
Author: Ziyang Zhang.
Script completion date: Jan 29th, 2016.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
This script calculates sample L-Moments and L-Moment statistics for a given
flood sample.

See Hosking and Wallis (1997) for reference of L-Moments.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Input variables:
---------------------------------------------------------------------------
rawdata: A vector. This vector contains flood values of a given flood
sample.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Output variables:
---------------------------------------------------------------------------
l: A vector (1 by 4). This vector returns the sample L-Moments lambda
values (i.e., the l values).
l(1)=l1, l(2)=l2, l(3)=l3, and l(4)=l4.
---------------------------------------------------------------------------
t: A vector (1 by 4). This vector returns the sample L-Moment statistics.
t(1)=mean, t(2)=L-CV, t(3)=L-Skewness, and t(4)=L-Kurtosis.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%}
n=numel(rawdata);
x=sort(reshape(rawdata,[1,n]));
b0=mean(x);
b1=0;
for i=2:n
    b1=b1+((i-1)./(n-1)).*x(i);
end
b1=b1/n;
b2=0;
for j=3:n
    b2=b2+((j-1)*(j-2)/((n-1)*(n-2))).*x(j);
end
b2=b2/n;
b3=0;
for k=4:n
    b3=b3+((k-1)*(k-2)*(k-3)/((n-1)*(n-2)*(n-3))).*x(k);
end
b3=b3/n;
l(1)=b0;
l(2)=2*b1-b0;
l(3)=6*b2-6*b1+b0;
l(4)=20*b3-30*b2+12*b1-b0;
t(1)=l(1);
t(2)=l(2)/l(1);
t(3)=l(3)/l(2);
t(4)=l(4)/l(2);
end
```

```
function [ para, ifail ] = SPFitKap1( xmom )
%{
Author: Ziyang Zhang.
Script completion date: August 25th, 2016.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
This script aims to estimate the parameters of a Kappa distribution by a
given set of L-Moments l1, l2, t3, and t4.

This script is developed based on the following reference. This script is a
translation from Fortran script to MATLAB script.

Reference:
Hosking, J.R.M., 2005. LMOMENTS: Fortran routines for use with the method
of L-Moments. IBM RESEARCH REPORT RC20525. URL:
http://lib.stat.cmu.edu/general/lmoments.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Input variables:
---------------------------------------------------------------------------
xmom: A vector (1 by 4). This vector contains sample L-Moments l1(Lambda 1),
l2(Lambda 2), t3(L-Skewness), and t4(L-Kurtosis).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Output variables:
---------------------------------------------------------------------------
para: A vector (1 by 4). This vector returns the parameters of the
estimated Kappa distribution in the order e(location), a(scale), k(shape),
and h(shape).
---------------------------------------------------------------------------
ifail: A value. This value returns the output flag.
C                   0  SUCCESSFUL EXIT
C                   1  L-MOMENTS INVALID
C                   2  (TAU-3, TAU-4) LIES ABOVE THE GENERALIZED-LOGISTIC
C                      LINE (SUGGESTS THAT L-MOMENTS ARE NOT CONSISTENT
C                      WITH ANY KAPPA DISTRIBUTION WITH H.GT.-1)
C                   3  ITERATION FAILED TO CONVERGE
C                   4  UNABLE TO MAKE PROGRESS FROM CURRENT POINT IN
C                      ITERATION
C                   5  ITERATION ENCOUNTERED NUMERICAL DIFFICULTIES -
C                      OVERFLOW WOULD HAVE BEEN LIKELY TO OCCUR
C                   6  ITERATION FOR H AND K CONVERGED, BUT OVERFLOW
C                      WOULD HAVE OCCURRED WHEN CALCULATING XI AND ALPHA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%}
eps=1e-6;
maxit=20;
maxsr=10;
hstart=1.001;
big=10;
oflexp=170;
oflgam=53;
t3=xmom(3);
t4=xmom(4);
para=[0 0 0 0];

% Test for feasibility
if xmom(2)<=0
    ifail=1;
    return
```

```
end
if abs(t3)>=1 || abs(t4)>=1
    ifail=1;
    return
end
if t4<=((5.*t3.^2-1)./4)
    ifail=1;
    return
end
if t4>=((5.*t3.^2+1)./6)
    ifail=2;
    return
end
% Set starting values for N-R iteration
k=(1-3.*t3)./(1+t3);
h=hstart;
z=k+h.*0.725;
xdist=big;
% Start of Newton-Raphson iteration
for i=1:maxit
    for j=1:maxsr
        countt=1;
        if k>oflgam
            ifail=5;
            return
        end
        if h>0
            u1=exp(gammaln(1./h)-gammaln(1./h+1+k));
            u2=exp(gammaln(2./h)-gammaln(2./h+1+k));
            u3=exp(gammaln(3./h)-gammaln(3./h+1+k));
            u4=exp(gammaln(4./h)-gammaln(4./h+1+k));
        else
            u1=exp(gammaln(-1./h-k)-gammaln(-1./h+1));
            u2=exp(gammaln(-2./h-k)-gammaln(-2./h+1));
            u3=exp(gammaln(-3./h-k)-gammaln(-3./h+1));
            u4=exp(gammaln(-4./h-k)-gammaln(-4./h+1));
        end
        alam2=u1-2.*u2;
        alam3=-u1+6.*u2-6.*u3;
        alam4=u1-12.*u2+30.*u3-20.*u4;
        if alam2==0
            ifail=5;
            return
        end
        tau3=alam3./alam2;
        tau4=alam4./alam2;
        e1=tau3-t3;
        e2=tau4-t4;
        distt=max(abs(e1),abs(e2));
        if distt>=xdist
            del1=0.5.*del1;
            del3=0.5.*del3;
            k=xk-del1;
            h=xh-del3;
        else
            break
        end
```

```
    countt=countt+1;
end
if countt==maxsr
    ifail=4;
    return
end
% Test for convergence
if distt<eps
    ifail=0;
    para(4)=h;
    para(3)=k;
    temp=gammaln(1+k);
    if temp>oflexp
        ifail=6;
        return
    end
    gam=exp(temp);
    temp=(1+k).*log(abs(h));
    if temp>oflexp
        ifail=-6;
        return
    end
    hh=exp(temp);
    para(2)=xmom(2).*k.*hh./(alam2.*gam);
    para(1)=xmom(1)-para(2)./k.*(1-gam.*u1./hh);
    return
end
% Not converged: Calculate next step
xk=k;
xh=h;
xz=z;
xdist=distt;
rhh=1./(h.^2);
if h>0
    u1k=-u1.*psi(1./h+1+k);
    u2k=-u2.*psi(2./h+1+k);
    u3k=-u3.*psi(3./h+1+k);
    u4k=-u4.*psi(4./h+1+k);
    u1h=rhh.*(-u1k-u1.*psi(1./h));
    u2h=2.*rhh.*(-u2k-u2.*psi(2./h));
    u3h=3.*rhh.*(-u3k-u3.*psi(3./h));
    u4h=4.*rhh.*(-u4k-u4.*psi(4./h));
else
    u1k=-u1.*psi(-1./h-k);
    u2k=-u2.*psi(-2./h-k);
    u3k=-u3.*psi(-3./h-k);
    u4k=-u4.*psi(-4./h-k);
    u1h=rhh.*(-u1k-u1.*psi(-1./h+1));
    u2h=2.*rhh.*(-u2k-u2.*psi(-2./h+1));
    u3h=3.*rhh.*(-u3k-u3.*psi(-3./h+1));
    u4h=4.*rhh.*(-u4k-u4.*psi(-4./h+1));
end
dl2k=u1k-2.*u2k;
dl2h=u1h-2.*u2h;
dl3k=-u1k+6.*u2k-6.*u3k;
dl3h=-u1h+6.*u2h-6.*u3h;
dl4k=u1k-12.*u2k+30.*u3k-20.*u4k;
```

```
    dl4h=u1h-12.*u2h+30.*u3h-20.*u4h;
    d11=(dl3k-tau3*dl2k)/alam2;
    d12=(dl3h-tau3*dl2h)/alam2;
    d21=(dl4k-tau4*dl2k)/alam2;
    d22=(dl4h-tau4*dl2h)/alam2;
    det=d11*d22-d12*d21;
    h11= d22/det;
    h12=-d12/det;
    h21=-d21/det;
    h22= d11/det;
    del1=e1*h11+e2*h12;
    del3=e1*h21+e2*h22;
    % Take next N-R step
    k=xk-del1;
    h=xh-del3;
    z=k+h.*0.725;
    % Reduce step if K and H are outside the parameter space
    fa=1;
    if k<=-1
        fa=0.8.*(xk+1)./del1;
    end
    if h<=-1
        fa=min(fa,0.8.*(xh+1)./del3);
    end
    if z<=-1
        fa=min(fa,0.8.*(xz+1)./(xz-z));
    end
    if(h<=0 && (k.*h)<=-1)
        fa=min(fa,0.8.*(xk.*xh+1)./(xk.*xh-k.*h));
    end
    if fa~=1
        del1=del1.*fa;
        del3=del3.*fa;
        k=xk-del1;
        h=xh-del3;
        z=k+h.*0.725;
    end
end
ifail=3;
return

end
```

```
function [ H1,Best_Selec_G_ID,kick_sta_ID,add_sta_id,h_flag,...
    CurReviseTimeL2 ] = Revise_Pro_V4( H1,H_defined_homo,sta_index,...
    mySta_Flood,sta_dis_s2,T,j,NYr,CurReviseTimeL1,Lmax )
%{
Author: Ziyang Zhang.
Script completion date: Dec 16th, 2017.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
This script aims to executes the Automatic Region Revision Algorithm
(ARRA). This script is a part of the Develop_Flood_Regions.m script.

A flood region is inputed in this script and a revised region is outputed
from this script.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Input variables:
---------------------------------------------------------------------------
H1: A value. This value inputs the homogeneity statistic of the input flood
region.
---------------------------------------------------------------------------
H_defined_homo: A value. See its definition in Develop_Flood_regions.m.
---------------------------------------------------------------------------
sta_index: A vector. This vector inputs the station index of the input
flood region.
---------------------------------------------------------------------------
mySta_Flood: A matrix (n by col). See its definition in
Develop_Flood_regions.m.
---------------------------------------------------------------------------
sta_dis_s2: A matrix. This matrix contains the distances of attribute
similarity from target site to other sites. This matrix is produced by
Develop_Flood_regions.m.
---------------------------------------------------------------------------
T: A value. See its definition in Develop_Flood_regions.m.
---------------------------------------------------------------------------
j: A value. This value is used to display the site number of execution.
---------------------------------------------------------------------------
NYr: A vector (n by 1). See its definition in Develop_Flood_regions.m.
---------------------------------------------------------------------------
CurReviseTimeL1: A value. This value indicates the number of use of the
ARRA.
---------------------------------------------------------------------------
Lmax: A value (1 by 2). See its definition in Develop_Flood_regions.m.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Output variables:
---------------------------------------------------------------------------
H1: A value. This value indicates the homogeneity statistic of the new
flood region.
---------------------------------------------------------------------------
Best_Selec_G_ID: A vector. This vector returns site index of the new
flood region.
---------------------------------------------------------------------------
kick_sta_ID: A vector. This vector returns the site(s) removed from the
input flood region.
---------------------------------------------------------------------------
add_sta_id: A vector. This vector returns the site(s) added to the new
flood region.
---------------------------------------------------------------------------
h_flag: A value. This value is the output flag.
```

```
h_flag=0: Homogeneity improved. But the new region is not a homogeneous
region.
h_flag=1: Homogeneity improved. And a homogeneous region is developed.
h_flag=2: The site causing most region heterogeneous has been
removed. But no new site is added to the region.
---------------------------------------------------------------------------
CurReviseTimeL2: A value. This value shows the number of site that has been
checked for the adding to the flood region.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
The following scripts are needed in order to execute this script.
1. LHomogeneity_Test.m
2.      KappaInvCDF
3.      mySampleLMoment
4.      SPFitKap1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%}
H1_ori=H1;
h_flag=0;
add_sta_id=[];
CurReviseTimeL2=0;
ReviseTime_L2=Lmax(2);
rem_sta_total_yr=sum(NYr(sta_index));
Best_Selec_G_ID=sta_index;
group_remind_Sta_ID=sta_index;
kick_sta_ID=[];
while rem_sta_total_yr>=T
    le2=length(group_remind_Sta_ID);
    H_DEx=99.*ones(le2,1);
    for i=1:le2
        H1_DEx=[];
        site3=group_remind_Sta_ID;
        site3(i)=[];
        ns3=NYr(site3);
        Sta_Flood3=mySta_Flood(site3,:);
        [ H1_DEx,ifail ] = LHomogeneity_Test( Sta_Flood3,ns3 );
        if ifail~=0
            H1_DEx=99.99;
        end
        H_DEx(i)=H1_DEx;
    end
    H_DEx(1)=999;
    [min_H_DEx,min_loc]=min(H_DEx);
    kick_sta_ID_this=group_remind_Sta_ID(min_loc);
    group_remind_Sta_ID(find(group_remind_Sta_ID==kick_sta_ID_this))=[];
    rem_sta_total_yr=sum(NYr(group_remind_Sta_ID));
    kick_sta_ID=[kick_sta_ID;kick_sta_ID_this];
    if min_H_DEx<=H_defined_homo & rem_sta_total_yr>=T
        H1=min_H_DEx;
        Best_Selec_G_ID=group_remind_Sta_ID;
        h_flag=1;
        return
    end
end
year_need_add=T-rem_sta_total_yr;
loc_left_sta=find(ismember(sta_dis_s2(:,3),sta_index)==0);
ID_left_sta=sta_dis_s2([loc_left_sta],3);
```

```
while CurReviseTimeL2<=ReviseTime_L2
    new_cum_yr=cumsum(NYr(ID_left_sta));
    number_of_sta_add=find(new_cum_yr>=year_need_add,1);
    if isempty(number_of_sta_add)
        break
    end
    add_sta_id=ID_left_sta(1:number_of_sta_add);
    New_group_id=[group_remind_Sta_ID;add_sta_id];
    Sta_Flood2=mySta_Flood(New_group_id,:);
    ns2=NYr(New_group_id);
    [ H_D,ifail ] = LHomogeneity_Test( Sta_Flood2,ns2 );
    if ifail~=0
        H_D=99.99;
    end
    if H_D<=H_defined_homo
            h_flag=1;
            H1=H_D;
    Best_Selec_G_ID=New_group_id;
    return
    end
    if H_D<H1_ori
        H1=H_D;
        Best_Selec_G_ID=New_group_id;
        return
    else
        CurReviseTimeL2=CurReviseTimeL2+1;
        ID_left_sta(find(ismember(ID_left_sta,add_sta_id)))=[];
    end
        disp('Current H and loop');
    disp(H_D);
     disp([j CurReviseTimeL1 CurReviseTimeL2]);
end
h_flag=2
le3=length(group_remind_Sta_ID);
H_DEx2=99.*ones(le3,1);
for i=1:le3
    H1_DEx2=[];
    site4=group_remind_Sta_ID;
    site4(i)=[];
    ns4=NYr(site4);
    Sta_Flood4=mySta_Flood(site4,:);
    [ H22,ifail ] = LHomogeneity_Test( Sta_Flood4,ns4 );
    H_DEx2(i)=H22;
end
H_DEx2(1)=999;
[min_H22,min_loc2]=min(H_DEx2);
add_sta_id=[];
kick_sta_ID_this2=group_remind_Sta_ID(min_loc2);
kick_sta_ID=[kick_sta_ID;kick_sta_ID_this2];
group_remind_Sta_ID(min_loc2)=[];
Best_Selec_G_ID=group_remind_Sta_ID;
H1=min_H22;
end
```