NRC·CNRC

Literature Guided Perspectives on Regional Low Flow Frequency Analysis for Ontario Streams

Report No.: NRC-OCRE-2021-TR-062
Date: December 21, 2021
Author: M.N. Khaliq

Ocean, Coastal and River Engineering Research Center

National Research Council Canada
Conseil national de recherches Canada

Canada

Cat. No. NR16-404/2022E-PDF

ISBN 978-0-660-45469-6

Acknowledgements

The work reported here was undertaken within the framework of an inter-departmental agreement between Ontario Ministry of the Environment, Conservation and Parks (MECP), and the National Research Council Canada (NRC). Project co-ordination support of Mr. Ted Belayneh, from the MECP, and technical assistance and review by Dr. Mohammad Sajjad Khan and Dr. Zhiping Yang also from the MECP is very much appreciated. Mr. Jason Mills, from the NRC, provided editorial support, which is graciously acknowledged. The technical detail and views expressed here are compiled based on the published literature on frequency analysis of hydrological variables, particularly low flows for both gauged and ungauged watersheds.

Executive Summary

Many water resources development projects require detailed information on the frequency and magnitude of extreme hydrological conditions such as low and high flows in order to inform various design, planning and management related activities. This study is concerned with low flow conditions and their probabilistic assessment. A low flow condition can be defined as a period, ranging from a day to several days, during which the average streamflow is a minimum for the entire year or for a selected seasonal period. Probabilistic characterization of low flow conditions is important for a number of engineering purposes and to satisfy many societal needs and natural environmental functions, such as the determination of minimum flow requirements downstream of a hydropower plant, design of water storage facilities, quantification of available water resources to inform municipal and industrial usages, management of water quality, determining effluent dilution capacity, assessing the impact of low flows on aquatic ecosystems and recreational facilities, etc.

Low flow studies often require estimation of the magnitude, frequency, and duration of low flow events. When duration of a low flow event is fixed, which generally is the case, the analysis then simply involves estimation of the magnitude and frequency of low flow events using statistical frequency analyses and observational records. These analyses can be carried out at a single site or at the level of a specific region of interest. The main purpose of the latter approach is to improve the quality of selected low flow indices at gauged locations and to facilitate a framework for the estimation of the same indices at ungauged locations, where streamflow data are either limited or nonexistent within the same region. In the literature on low flow hydrology, this approach is generally referred to as regional low flow frequency analysis (henceforth RLFA), which is the focal point of this report. In addition, a number of other techniques have also been developed for low flow estimation at ungauged locations.

Statistical frequency analysis involves fitting a probability distribution to a sample of low flow events. The output of this analysis gives an idea of the likelihood of future occurrences of a specified low flow event under some sensible assumptions. For such analyses, it is generally assumed that the magnitude of a low flow event can reliably be estimated if the sample size is reasonably long enough. In reality, historical records are often too short to estimate these flow magnitudes in a reliable manner. Among other reasons, this limitation has led to the development of regional frequency analysis (RFA) approaches, which involves identification of a homogeneous region based on physical, hydrologic, climatic or statistical homogeneity concepts and then pooling information from the entire region following established procedures and thereby estimating magnitudes of desired low flow indices. The regional approach not only improves the quality of low flow estimates at sites with short records, but also provides a basis for estimation of low flow indices at all ungauged locations within the target region of interest.

In Ontario, analysis of low flows was conducted in 1990s using observational records from nearly 340 gauging locations and a software package developed by Inland Waters Directorate (currently Water Survey of Canada) of Environment and Climate Change Canada. As the software has become almost obsolete and there is roughly 30 years of additional data, Ministry of the Environment, Conservations and Parks (MECP) desired to have the software redeveloped in a modern language with a user-friendly interface and all associated reports to be updated. The specific deliverables of the project were identified as: (1) an updated low flow frequency analysis (LFA) software, (2) a report pertaining to LFA of Ontario streams using most recent data, (3) development of a framework for undertaking LFA considering effects of future climate change, and (4) a documented review of regional LFA techniques, with a focus on ungauged locations. The National Research Council Canada (NRC) led this effort through an inter-departmental agreement between the MECP and the NRC. This report specifically documents a review of regional LFA (RLFA) techniques available in the literature, with the objective to estimate low flow indices at ungauged locations across Ontario. Throughout this effort, the focus has been on presenting a variety of existing and emerging techniques than presenting an exhaustive review of the subject, which undoubtedly is a daunting exercise. Where applicable, shortcomings are highlighted and recommendations are made for additional research in order to obtain improved estimates of low flow indices at gauged locations, which, in turn, will help improving the quality of similar estimates at ungauged locations.

This report is divided into five chapters and a section on references. The background information on at-site LFA, RLFA, the significance of flow duration curves in regional low flow analyses and some general information on the subject is provided in Chapter 1 in order to equip the reader with sufficient background on the topic. Objectives and limitations of the report are also discussed in this chapter. A short primer on low flow frequency analysis is provided in Chapter 2. Chapter 3 of the report provides a review of the literature on RLFA techniques within the realm of ungauged hydrology for transposition of low flow indices from gauged to ungauged locations; this is the major focus of this report. Chapter 4 pertains to regional analysis of flow duration curves, wherein perspectives on existing techniques and their applicability in Ontario are discussed. The final Chapter 5 explores avenues of future research, and discusses potential recommendations and steps necessary to be followed for developing regional low flow analysis outputs for Ontario. Some data sources for deriving watershed attributes are also discussed in this chapter since these attributes play a significant role in RLFFA. A list of references cited in all chapters is available at the end of the report.

The information provided in this report is expected to help pave the way forward for improving estimation of low flow indices at ungauged locations across Ontario, as well as for improving our understanding of geophysical and climatic controls on low flow regimes that form the critical basis for deriving statistical relationships required for estimating low flow indices at ungauged locations through information transposition from gauged to ungauged locations or through direct relationships based on watershed attributes, including topographic, geologic, soil and land use,

and climatic attributes. An effort has also been made to reflect on the present state of the knowledge in ungauged hydrology with respect to estimation of low flow indices. It is proposed that such reviews should occur on regular basis in order to strengthen and validate existing and emerging approaches based on refined and improved datasets of watershed attributes. These datasets are continuously being refined through dedicated national and regional level initiatives. Development and identification of an accurate method for transposition of low flow indices from gauged to ungauged locations using physiographic and climatic attributes still remains a significant challenge. It is hoped that the estimation of low flow indices at ungauged locations at the regional and national levels based on new technological developments, new analysis tools and improved scientific understanding of low flow regimes, as well as development of high quality geophysical and geospatial datasets, will continue in the future.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Description
CCA	Canonical Correlation Analysis
DEM	Digital Elevation Model
DHR	Delineation of Homogenous Regions
FDC	Flow Duration Curve
GCM	Global Climate Model
HYDAT	Hydrometric Database
IPCC	Intergovernmental Panel on Climate Change
LFA	Low Flow Analysis
LFFA	Low Flow Frequency Analysis
MECP	Ministry of Environment, Conservation and Parks
MOEE	Ministry of Environment and Energy
NRC	National Research Council Canada
OCRE	Ocean, Coastal and River Engineering
OLB	Outer Lower Bound
RBLI	Regression Based Logarithmic Interpolation
REM	Regional Estimation Method
RLFFA	Regional Low Flow Frequency Analysis
RMSE	Root Mean Square Error
RFA	Regional Frequency Analysis
ROI	Region of Influence
WMO	World Meteorological Organization

1 Introduction

1.1 Background

Many water resources development projects require detailed information on the frequency and magnitude of extreme hydrological conditions such as low and high flows to inform various design, planning and management related activities. In general, a low flow condition can be defined as a period, ranging from a day to several days, during which the average streamflow is a minimum for the entire year or for an entire season of interest. High flow conditions can also be defined in an analogous manner. This study is concerned with low flow conditions and their probabilistic assessment. Statistical characterization of low flow conditions are important for a number of engineering purposes and to satisfy many societal and natural environmental needs, such as the determination of minimum flow requirements downstream of a hydropower plant, quantification of available water resources to inform municipal and industrial usages, reservoir design and management, management of water quality, determining effluent dilution capacity, and assessing the impact of low flows on aquatic ecosystems (e.g. Riggs et al., 1980; Smakhtin, 2001; Gustard et al., 2004; Tallaksen and van Lanen, 2004; Laaha and Blöschl, 2007; WMO, 2008). Given such an importance of low flows in engineering and environment, many countries and jurisdictional regions have developed low flow estimation procedures for both gauged and ungauged locations. For example, see Holmes et al. (2002) for the UK, Aschwanden and Kan (1999) for Switzerland, England et al. (2006) for Norway, Ries (2002) for the USA, Henderson et al. (2005) for New Zealand, and Laaha and Blöschl (2007) for Austria. In Canada, each province has developed its own low flow estimation procedures and guidelines.

The low flow regime of a stream can be analyzed in various ways depending on the type of available data and the target application of the outputs of such an analysis (e.g. Tallaksen et al., 1997; Smakhtin, 2001; Tallaksen and van Lanen, 2004; Patel, 2007). In general, low flow studies often require estimation of the magnitude, frequency, and duration of low flow conditions. When duration of a low flow event is fixed based on physical constraints of a stream or due to operational and management related priorities, the analysis then simply involves estimation of the magnitude and frequency of low flow events using statistical analyses. Some studies also consider the concept of streamflow deficit and characterize spells of low flows to inform planning and management. Likewise, some studies have also used specific indices of flow duration curves for the same purpose. These analyses can be carried out at a single site or at the level of a specific region of interest. The main purpose of the latter approach is to improve the quality of selected low flow indices at gauged locations and to facilitate a framework for estimating the same indices at ungauged locations within the same region. In the literature on low flow hydrology, this approach is generally referred to as regional low flow frequency analysis (henceforth RLFFA), which is the focal point of this report. Further information on the RLFFA topic is provided below.

Statistical frequency analysis involves fitting a probability distribution to a sample of low flow events, which are derived from historical observations using either the block maxima or peaks-over-threshold sampling approach. In the former case only one low flow event from a given year or season is considered while in the case of the latter approach, more than one event from the same year can be included in the sample. The output of frequency analysis gives an idea of the likelihood of future occurrences of a specified low flow event. For such analyses, it is generally assumed that the magnitude of a low flow event can reliably be estimated if the sample size is reasonably long enough. In reality, historical records are often too short to estimate the magnitudes of low flow events in a reliable manner. Among other reasons, this limitation has led to the development of regional frequency analysis (RFA) approach, which involves identification of a homogeneous region based on physical and/or statistical homogeneity concepts and then pooling information from the entire region following established procedures and thereby estimating magnitudes of desired low flow events. The regional approach not only improves the quality of low flow estimates at sites with short records, but also provides a basis for low flow estimation at all ungauged locations within the target region of interest. The RFA approach can be applied to numerous hydrological variables, including mean, low and high flows. As this report is concerned with low flows, the RFA will facilitate estimation of low flow magnitudes corresponding to selected frequencies of streamflow being equal or less than the estimated value.

Any RFA procedure involves the following general steps: (1) collection of low flow data from gauging stations within the region of interest, following a selected sampling methodology; (2) screening the low flow data for gross errors, outliers or any other human/instrument-related causes that can make the data unsuitable for frequency analysis; (3) identifying homogeneous regions based on physical/climatic similarity, geographic proximity, statistical similarity or other understandings of similarity; (4) determining regional growth curves or standardized frequency factors; and (5) estimating low flow magnitudes of interest at all gauged and ungauged sites within the region of interest. The idea of RFA was proposed by Dalrymple (1960), which later was formalized by Hosking and Wallis (1997) using L-moments, which were developed by Hosking (1990). One of the fundamental assumptions of the RFA approach is that all sites within a homogeneous region shares a common probability distribution, except a scale factor, which is often assumed as the mean/median annual flood for high flow analyses and mean/median low flow magnitude for low flow analyses. This scale factor can be called an index flow to generalize the concept across various applications in hydrology and environmental sciences. When applying the RFA approach at ungauged locations within the same homogeneous region, an additional step is also required and that involves development of a regression relationship between the index flow and physiographical and climatological characteristics of watersheds included in the region. Several different approaches have been used in the literature to develop this relationship in a linear or nonlinear fashion. It is important to note that many variants of the RFA exist in the literature, however, the RFA approach, formalized by Hosking and Wallis (1997), has been well established in hydrology and other related disciplines. According to Shi et al. (2010), probably

Tallasken and van Lanen (2004) were the first who advocated application of Hosking and Wallis's RFA approach for regional low flow analysis.

In Ontario, analysis of low flows was conducted in the 1990s using observational data from nearly 340 gauging locations and a software package developed by Inland Waters Directorate (currently Water Survey of Canada) of Environment and Climate Change Canada. As the software has become obsolete and there is roughly 30 years of additional data, Ontario Ministry of the Environment, Conservation and Parks (MECP) desired to have the software redeveloped using a present day language and a user-friendly interface, and all associated reports to be updated. The specific deliverables of the project were identified as: (1) an updated low flow frequency analysis (LFFA) software, (2) a report pertaining to LFFA of Ontario streams using most recent data, (3) development of a framework for undertaking LFFA considering the effects of future climate change, and (4) a documented review of RLFFA techniques. The National Research Council Canada (NRC) led this effort through an inter-departmental agreement between the MECP and the NRC. This report specifically documents a review of RLFFA techniques available in the literature in the form of periodicals and technical reports, originating from private and government sectors. Throughout this effort, the focus has been on presenting a variety of existing and emerging techniques than presenting an exhaustive review of the subject.

Conventionally, all of the above analyses and investigations are performed using recorded streamflow data assuming a stationary climate. Due to climate change as projected by Global Climate Models (GCMs) and documented in various reports of the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2007, 2013), the assumption of a stationary climate has become questionable and therefore the applicability of low flow indices, derived from recorded historical observations, and their transposition at ungauged locations under the assumption of stationarity, has also become questionable. It is worth pointing out that many human related activities clearly affect the climate system. Most importantly, emissions of greenhouse gases, especially carbon dioxide and methane, are causing more heat to be trapped within earth's atmosphere. Therefore, the case for significant climate change is compelling in both the empirical observations and theoretical predictions. A warmer air mass can hold more water (i.e., warmer air has a higher saturation vapor pressure) and, therefore, it is reasonable to expect higher amounts of water vapor in the air, leading to intensification of the hydrologic cycle, with impacts ranging from one region to another and from one component of the hydrologic cycle to another (e.g. IPCC 2013; Khaliq, 2019). However, it is not so straightforward to consider the impacts of a changing climate when deriving low flow indices for ungauged locations within a target region of interest. Though recognized and acknowledged, the topic of non-stationary climate is not considered in this report with respect to estimation of low flow indices. However, a framework for the estimation of low flow indices under the influence of climate change is discussed and documented in a separate report, which is another deliverable of the project as indicated above.

1.2 Objectives

The objectives of this report are to:

- Document an overview of the hydrological aspects for estimating indices of low flows using regional frequency analysis approaches at ungauged locations in Ontario;
- Document strengths and weaknesses of the reviewed approaches and identify data needs of these approaches for Ontario-wide applications; and
- Carve a path forward for future research and development from a hydrological perspective in order to estimate indices of low flows at ungauged locations in Ontario and that, in turn, can inform service delivery, water extraction licensing and allocation targets of the MECP in Ontario's vast network of rivers, creeks and streams.

1.3 Organization of the Report

This report is divided into six chapters, including this introduction chapter, and a section on references. The background information on at-site LFFA, RLFFA, the significance of flow duration curves in regional low flow analyses and some general information on previous studies is provided in Chapter 1 in order to provide the reader with sufficient background on the topic. Objectives and limitations of the report are also discussed in this chapter. A short primer on at-site low flow frequency analysis is provided in Chapter 2 and that focuses on some basic information on the statistical concepts related to LFFA. Chapter 3 of the report provides a review of the literature on regional low flow analysis techniques covering both gauged and ungauged locations – the major focus of this report. Perspectives on flow duration curves and their estimation for ungauged watersheds is provided in Chapter 4. The final Chapter 5 explores avenues of future research, and discusses potential recommendations and steps necessary to be followed for developing regional low flow analysis approaches for Ontario. A list of references cited in the report is available at the end.

1.4 Convention on the Usage of Acronyms and Other Considerations

A number of acronyms are used in this report, which are devised based on various acronyms used previously in the literature. Some of the acronyms are chapter-specific, while others are utilized throughout the report. Therefore, to facilitate easy comprehension and smooth readability, the acronyms are reintroduced in their expanded form in each chapter so that each chapter can be read independently, without referring back and forth to other chapters.

In this report, the terms like river flow, streamflow, or simply flow, reflecting open channel flow conditions, are considered equal in terms of meanings. It was necessary to state it upfront since different terms are used in the literature on low flows. Selected percentiles of flow duration

curves from the lower portion of the curve and low flow magnitudes or quantiles corresponding selected return intervals or return periods are referred to as low flow indices in this report.

In Chapter 5, a number of different datasets are discussed, in addition to the hydrometric dataset pertaining to recorded streamflow. These datasets pertain to (1) topographic features; (2) soil and land use; (3) surficial geology; and (4) climatic features. A number of different attributes can be derived from each of these datasets to aid in RFA. In this report, for simplicity reasons, the attributes that can be derived from these datasets are referred to as watershed attributes.

1.5 Scope and Limitations

The review and discussions provided in this report are intended for individuals that have some basic understanding of runoff-generating mechanisms in riverine environments, methods pertaining to streamflow analysis and the statistical concepts involved in time series modelling and data analysis, estimation of flow duration curves and statistical frequency analyses of low flow sequences, in addition to many other hydrological analyses specific to gauged and ungauged locations at various temporal and spatial scales. To provide a broader perspective on the subject of RLFFA world-wide, both national and international sources are also cited. The documents and technical/scientific information sources considered for this report are mostly publicly available or available through dedicated publication portals. In this report, where applicable, references are also provided for obtaining additional information and details on various new methods that have emerged over the last several decades and that can be adapted for LFFA in Ontario.

The scope of this report is limited to only hydrological aspects. The environmental aspects that are equally important for assessing and defining low flow magnitudes to satisfy various environmental needs of streams are not discussed here. For a detailed account of these aspects, the reader is referred to appropriate published sources. To improve analysis and understanding of low flow regimes and their characteristics at gauged and ungauged locations, some avenues of future research are identified based solely on the review presented in this report and the developments reported in the literature on physical and stochastic hydrology. Detailed descriptions of theoretical aspects that underpin these new developments lie outside the scope of this report. For such descriptions, scientific articles and technical reports associated with these methods should be referred to.

When considering a region for frequency analysis, one can divide the research project into several stages. For example, (1) literature-guided pre-assessment; (2) data collection, preparation and quality control; (3) defining performance assessment metrics; (4) application and evaluation of potential techniques; and (5) recommendations for the target region of interest. The primary difference among these stages is the degree of analysis and the confidence that one can have in the performance of selected techniques, the primary product of the first four stages. Many of these stages are inter-dependent, follow a top-down approach and are supported by the literature-

guided pre-assessment. The review presented in this report falls under the first stage and therefore will act as a fundamental source of information for future studies on RLFFA in Ontario, as well as in other parts of Canada.

2 A Primer on Low Flow Frequency Analysis

2.1 General

The main focus of this report is to review regional frequency analysis (RFA) approaches for the estimation of low flow indices at ungauged locations. Before discussing RFA of low flows for any region of interest, it is important to know some basic statistical terminology and how at-site frequency analysis of low flows is conducted. These concepts and procedures are described below in this chapter. The RFA procedure described in Chapter 3 of this report is closely tied with the L-moments approach of Hosking and Wallis (1997) and therefore some fundamental concepts, procedures involved in estimating L-moments, and their interpretations with reference to at-site low flow data are also discussed here.

In general, for frequency analysis of any hydrological variable, including low flows, the collected data from a site must be a true representation of the associated watershed conditions and must be drawn from the same probability distribution, which is the fundamental requirement for conducting statistical frequency analysis. For this analysis, it is also assumed that the data values are random and independent and constitute a homogeneous sample. Though several statistical tests are available to verify these assumptions, these tests are not included in this report. It is also worth mentioning that many of these assumptions are relaxed when performing non-stationary frequency analyses (e.g. Coles, 2001; Strupczewski et al., 2001a, 2001b;, Khaliq et al., 2006; Sushama et al., 2006; Mudersbach and Jensen, 2010; Tramblay et al., 2013; Salas and Obeysekera, 2014;, Xiong et al., 2015; Tan and Gan, 2015; Šraj et al., 2016; Salas et al., 2018; and Wu and Xue, 2018). The non-stationary aspects are also not covered in this report.

2.2 Low Flow Frequency Analysis (LFFA)

For conducting statistical frequency analysis of low flows at a given site, where continuous streamflow observations are available, it is important to begin with by extracting a sample of low flows. A low flow event can be defined as the annual minimum daily flow or it can also be defined in terms of an annual average low flow value that can persist over a period of d days, where d could be 3, 5, 7, 15 or any other discrete number of days. The choice of d depends on the regulatory norms, mandated by watershed management authorities, and also on the objectives of the study, among many other factors. However, averaged flows for d > 1-day are believed to be less sensitive to measurement errors (Shi et al., 2010). On an annual scale, there will be as many number of d-day low flows as there is the number of years of streamflow records at a given gauging station. For conducting at-site frequency analysis, some statistical notions need to be defined and understood with respect to the phenomenon of low flow occurrences in rivers, streams and creeks of a geographic region.

Let X be a random variable that represents a low flow value at a given site (here the notation X is used to generalize the statistical concepts; in many practical applications it is replaced with Q). The variable X can take on values that are real numbers. In the case of low flows, X can take on any value between zero and infinity. The relative frequency with which these X values occur over the length of recorded observations defines the frequency distribution or probability distribution of X and that is specified by the cumulative distribution function (CDF) $F(x)$, which is the probability that the random variable X is at most x or less than and equal to x. The $F(x)$ is expressed as:

$$F(x) = \text{Prob}[X \leq x]. \tag{2.1}$$

For continuous probability distributions, an associated concept is the probability density function, which is defined as $f(x) = \frac{d}{dx}F(x)$. The $F(x)$ is an increasing function of x, and its value is always between zero and unity. The inverse function of the CDF is called the quantile function, which is denoted by $x(F)$ and expresses the magnitude of an event in terms of its non-exceedance probability F, i.e. the value of the random variable X such that the probability that X does not exceed $x(F)$ is F (the F notation is used here to simplify the idea). In risk and reliability engineering as well as in hydrologic and environmental applications, a quantile is usually expressed in terms of its return period. The quantile of return period T, XT, is an event such that it has a probability of $1/T$ of being equal or falling below this value in a given year. It is important to note that for low flow frequency analysis, the quantiles are specified with reference to the lower tail. In this case, the quantile XT is specified as: $XT = x(1/T)]$. In other words, it is assumed that $F(XT) = 1/T$. Estimation of low flow quantiles XT corresponding to several return periods of interest (e.g. 2, 5, 10, 20, 50 years) is the main goal of LFFA. In LFFA area, return periods larger than 50 years are rarely used. A typical low flow frequency curve is shown in Figure 2.1.

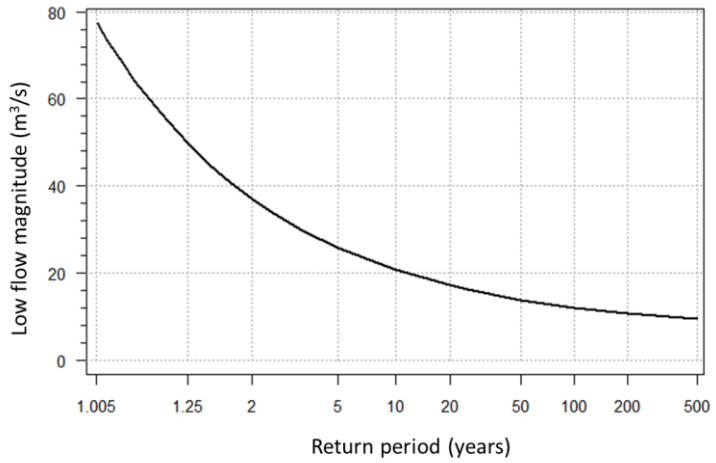

Figure 2.1: A typical low flow frequency curve.

2.2.1 Choice of a Distribution Function and Parameter Estimation Method

Conventionally, this analysis is generally performed at all gauged locations using a preferred or a range of probability distribution functions. Among the distribution functions that have been used previously for LFFA include the Weibull, gamma, log-normal, Pearson Type III, and Generalized Extreme Value distributions. Both two and three parameter Weibull distributions have often been used for LFFA.

For estimating parameters of these distributions from observed samples of low flows, several methods can be used, including the method of moments (MOM), method of maximum likelihood (MML), method of L-moments (MLM). The MOM is easy to apply compared to the MML, however, the MML is statistically the most efficient method (i.e. it provides asymptotically minimum variance estimators). Compared to the MOM, the MML generally involves non-linear equations which often require numerical solutions or optimization techniques. Due to some attractive properties in terms of robustness for small samples, the MLM has become popular for frequency analysis. As the main focus of this review is on RFA, with attention to L-moments approach of Hosking and Wallis (1997), some background information on L-moments is provided below. For estimating distribution parameters, theoretical moments or moment ratios in terms of distribution parameters are equated to their corresponding estimates from the data sample and the resulting equations are solved simultaneously or iteratively. This procedure is applicable for both the MOM and the MLM.

Theoretical L-moments

L-moments were derived by Hosking (1990) from probability weighted moments (PWMs), which were introduced by Greenwood et al. (1979). The PWMs can be defined as:

$$M_{p,r,s} = E[X^p\{F(X)\}^r\{1-F(X)\}^s]. \quad (2.2)$$

where p, r and s are real numbers and $M_{p,r,s}$ exists for all $r, s \geq 0$ if and only if $E|X|^p$ exists. For a distribution that has the quantile function $x(u)$, two special cases of the PWMs can be described as:

$$M_{1,0,r} = \alpha_r = \int_0^1 x(u)\,(1-u)^r\,du, \text{ and} \quad (2.3)$$

$$M_{1,r,0} = \beta_r = \int_0^1 x(u)\,(u)^r\,du. \quad (2.4)$$

These equations are similar to the conventional moments, which are defined as:

$$E(X^r) = \int_0^1 \{x(u)\}^r\,du \quad (2.5)$$

At the time when the PWMs were defined, some published studies wherein the PWMs were used directly for estimating parameters of probability distribution functions include Landwehr et al. (1979a, b) and Hosking and Wallis (1987). Though some investigators still use PWMs for the same purpose, it is difficult to make direct interpretations in terms of scale and shape of the

distribution functions. Hosking (1990) considered certain linear combinations of the PWMs α_r and β_r and defined L-moments as:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^{r} P^*_{r,k}\, \alpha_k = \sum_{k=0}^{r} P^*_{r,k}\, \beta_k. \quad (2.6)$$

For a random variable X with quantile function $x(u)$, L-moments can also be described as:

$$\lambda_r = \int_0^1 x(u)\, P^*_{r-1}(u) du \quad (2.7)$$

where $r = 0,2,3, ...,$ and

$$P^*_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2 (r-k)!} \text{ and} \quad (2.8)$$

$$P^*_r(u) = \sum_{k=0}^{r} P^*_{r,k}\, u^k. \quad (2.9)$$

The first four L-moments in terms of PWMs can be written as:

$$\lambda_1 = \alpha_0 = \beta_0 \quad (2.10)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \quad (2.11)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (2.12)$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (2.13)$$

Analogous to conventional moment ratios, i.e. coefficient of skewness and coefficient of kurtosis, L-moment ratios are dimensionless versions of L-moments. These ratios are obtained by dividing the higher order L-moments by λ_2. The L-moment ratios, L-CV, L-skewness and L-kurtosis are respectively defined as:

$$\tau = \lambda_2/\lambda_1 \text{ (L-CV)}, \quad (2.14)$$

$$\tau_3 = \lambda_3/\lambda_2 \text{ (L-skewness) and} \quad (2.15)$$

$$\tau_4 = \lambda_4/\lambda_2 \text{ (L-kurtosis).} \quad (2.16)$$

The first L-moment λ_1 is a measure of central tendency and is equivalent to the mean of the distribution function, whereas λ_2 is a measure of dispersion. Thus, λ_2/λ_1 is equivalent to commonly used coefficient of variation, i.e. σ/μ, where σ is the standard deviation and μ is the mean. The L-moment ratios are easy to interpret as they are analogous to the conventional moments. Their popularity for RFA has grown exponentially over the years because they are less biased than the conventional moments, resistant to outliers and have better ability to discriminate between competing distribution functions (e.g. Cunnane, 1989; Hosking, 1990; Hosking and Wallis, 1997; Peel et al., 2001).

Sample L-moments

Theoretical relationships of L-moments for many commonly used distributions have already been determined by Hosking and Wallis (1997). For distribution fitting purposes, it is necessary to estimate L-moments from a finite sample of data. For this purpose, let us consider that

$x_1, x_2, x_3, \ldots, x_n$ is a sample of size n and let the ordered sample be $x_{1:n} \leq x_{2:n} \leq x_{3:n} \leq \cdots \leq x_{n:n}$. An unbiased estimator of the PWM β_r is given by:

$$b_r = n^{-1} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_{j:n} \quad (2.17)$$

Analogous to Equations (2-10) to (2-13), the sample L-moments can be defined as follows:

$$l_1 = b_0 \quad (2.18)$$
$$l_2 = 2b_1 - b_0 \quad (2.19)$$
$$l_3 = 6b_2 - 6b_1 + b_0 \quad (2.20)$$
$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \quad (2.21)$$

Similar to Equations (2.14) to (2.16), sample L-moment ratios can be defined. Once an appropriate distribution is fitted to a sample of low flows following the above described procedure, desired quantiles of interest corresponding to selected return periods can be determined following $XT = x(1/T)$ relationship. From a set of candidate distributions, the most appropriate distribution can be selected based on some goodness-of-fit measures, e.g. Anderson-Darling (e.g. Millington et al., 2011), Kolmogorov-Smirnov (Haan, 1977; Helsel and Hirsch, 2002; Eris et al., 2019) or L-moment based guidance (Hosking and Wallis, 1997; Yue and Pilon, 2005; Yue and Wang, 2004a, 2004b). Some investigators also prefer to use L-moment ratio diagrams as a visual tool to select an appropriate distribution function as explained below. Previous studies demonstrate that for a given dataset there is no single best distribution, but a set of credible distributions with similar fit, requiring an uncertainty analysis on the goodness-of-fit of a distribution function.

L-moment Ratio Diagram

A convenient way of representing L-moments of different distribution functions is the L-moment ratio diagram, which is created by plotting L-skewness against L-kurtosis. A two-parameter distribution would plot as a single point on this diagram, while the three-parameter distributions as a curve and distributions with more than three-parameters would generally cover two-dimensional areas on this diagram (Hosking and Wallis, 1997). Figure 2.2 shows a typical L-moment ratio diagram for some commonly used distribution functions.

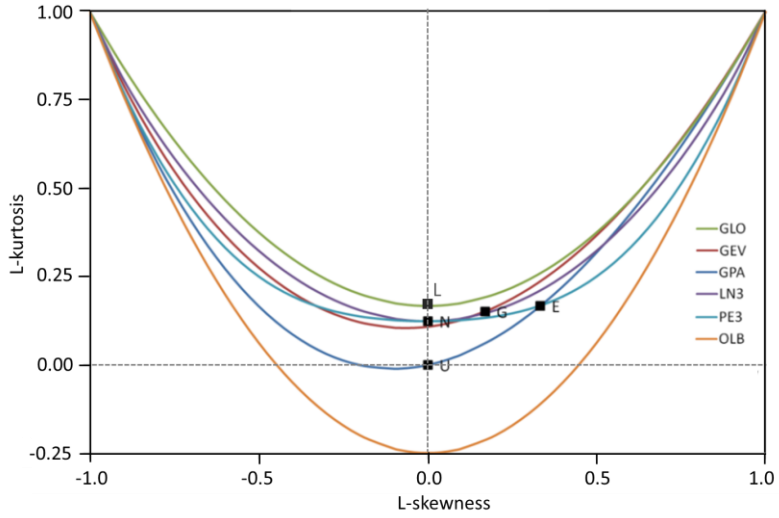

Figure 2.2: L-moment ratio diagram for commonly used distribution functions, i.e. exponential (E), Gumbel (G), Logistic (L), Normal (N), uniform (U), Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Pareto (GPA), three parameter lognormal (LN3), and Pearson Type III (PE3). OLB is the outer lower bound. Source: Hosking and Wallis (1997).

Hosking and Wallis (1997) stated that it is more convenient to express L-kurtosis as a function of L-skewness and therefore polynomial approximations of this relationship were developed for many distribution functions. According to Hosking and Wallis (1997), these relationships take the following general form:

$$\tau_4 = \sum_{k=0}^{8} A_k \tau_3^k. \qquad (2.22)$$

The values of coefficients A_k are available in Hosking and Wallis (1997) and those were used to develop the L-moment ratio diagram. For a given sample of low flows, the position of L-skewness and L-kurtosis point on this diagram will indicate the most appropriate candidate distribution for that sample. However, due to sampling variability, deviations from this behavior cannot be ruled out and therefore careful judgements are often exercised, supported with formal statistical inferences.

2.3 Concluding Remarks

In this chapter a basic introduction to at-site frequency analysis of low flows is presented in order to provide the reader an overview of the statistical terminology and procedures that underpin estimation of low flow quantiles at sites where continuous streamflow data is available for extracting low flow values following an established methodology, e.g. annual or seasonal approaches. Though the general procedure of distribution selection and parameter estimation methods is discussed, a detailed account is presented only for the case of L-moments. This is due to the reason that the RFA approach of Hosking and Wallis (1997), which has been well established in hydrology, is closely tied with L-moments. The RFA approach not only improves

the quality of at-site quantiles but also provides a reasonable basis for estimating low flow quantiles at ungauged locations. For at-site frequency analysis, distribution fitting using L-moments is a reasonable choice compared to the approach of conventional product moments. Additional information on the topic of RFA is provided in Chapters 3 and 4 in the context of RFA of low flows and flow duration curves.

3 Regional Frequency Analysis of Low Flows for Ungauged Locations

3.1 General

Perspectives from the literature on regional frequency analysis (RFA) are presented here, with emphasis on low flows. The RFA procedure using the index flood method was introduced by Dalrymple (1960) and that later was formalized by Hosking and Wallis (1997) using L-moments. This approach has been well established in hydrology and its use has grown exponentially over the last several years. The review reported here starts off with the generalized approach proposed by Hosking and Wallis (1997), but also builds on the work of many other researchers and knowledge generated through several applications of this and other relevant approaches in different parts of the world. The L-moments based RFA approach is applicable for a number of variables not only from hydro-meteorology but also from other scientific disciplines (e.g. high and low flow magnitudes and volumes, minimum and maximum temperature values, variables driving environmental pollution, wind fields, storm surge, extreme water levels, ice and snow loads, precipitation extremes, etc.). As pointed out before in the introduction chapter that the RFA approach helps improve the reliability of estimated quantiles at sites with short records within a region of interest, but also facilitates a reasonable framework for estimating desired quantiles at all ungauged locations within the same region. For RFA of low flows, often annual minimum values of 1-, 7-, 10-, and 15-day averaged flows are considered. In situations, where low flow occurrences are dominated by more than one generating mechanism, seasonal analyses of low flows are also considered. For example, low flows could occur due to frozen conditions during winter in cold environments and/or they could also occur during summer due to lack of precipitation and high evaporation demands. A large number of studies exist on RFA of flood flows compared to other variables (cf. Shi et al., 2010). However, the underlying principles as discussed below in this chapter are similar, irrespective of the variable of interest.

The objective of regional low flow frequency analysis (RLFFA) is to estimate quantiles of low flows (e.g. 7Q10, the average 7-day low flow magnitude associated with 10-year return period, or 7Q2, the average 7-day low flow magnitude associated with 2-year return period, or similar other quantiles) at all gauged and ungauged locations within the target region of interest. The general procedure for implementing the RFA approach for low flows (or any other variable) can be divided into the following steps: (1) data screening and quality control; (2) delineation of homogeneous regions; (3) testing regional homogeneity; (4) selection of a regional probability distribution function; and (5) estimation of desired quantiles at both gauged and ungauged locations. Additional insights on these steps are provided below. A major portion of this chapter is devoted to the RFA approach of Hosking and Wallis (1997) because it has been well accepted by practicing engineers, hydrologists, environmentalists, conservationists and many other professionals.

Regression-on-quantiles is another regionalization approach that is also popular and practiced in many countries, e.g. the UK (Gustard et al., 1992; Patel, 2007) and the USA (Ries, 2002). In this approach, after forming homogeneous regions, the quantity of interest (e.g. 7Q10 or Q95, the flow that is exceeded 95% of the time from the flow duration curve) are directly regressed on watershed attributes. This approach is almost similar to the estimation of index flow in the RFA approach of Hosking and Wallis (1997). Here, this approach is addressed in a separate section. In addition to the above mentioned approaches, Machine Learning (ML) techniques are being increasingly used for developing nonlinear relationships between index flow, selected quantiles or percentiles of flow duration curves and watershed attributes. In the literature, these approaches have also been used to identify groups of similar sites to form homogeneous regions. Some insights and potential applications of these approaches are provided in another separate section in this chapter. It is expected that future studies will consider applications of these approaches for regional analysis of low flow indices in Ontario. In the literature on ungauged hydrology, there are some studies where the concept of regional homogeneity has not been exploited but the statistical approaches used for transposition of known information from gauged to ungauged locations are very similar to regression-on-quantiles (e.g. Bond and Kennard, 2017).

Below the steps involved in the RFA approach of Hosking and Wallis (1997) are presented first followed by regression-on-quantiles and ML approaches. As mentioned before, it is not the intention of this report to present an exhaustive review of this subject, rather to present a variety of approaches that can be experimented for RFA of low flows in Ontario.

3.2 Data Screening

The first step in RFA is to check that the available sequences of low flows (or samples of low flows) from all sites within the target region are appropriate for statistical frequency analyses. As discussed in Chapter 2, it is important to ensure that the samples of low flows consist of random and independent values and are homogeneous and have come from the same probability distribution. The errors or unexpected behaviours/patterns in observed low flows could arise from instrument malfunctioning or observer biases, systematic changes that had occurred over time, related to station location and instrument type, flow regulation influences, or any combination of these and other factors. The influence of these and other relevant factors may lead to outliers, non-homogeneities, serial dependence, trends or even upward and downward jumps. As a result of these causes, the reliability of RFA reduces considerably. Several statistical tests exist to identify these issues and those have been implemented in many commercial and public domain software packages (e.g. Matlab, Minitab, SAS, R, Python, etc.) in addition to LFA (low flow frequency analysis) and CFA (consolidated frequency analysis) software tools, developed by Environment and Climate Change Canada.

For initial screening purposes, Hosking and Wallis (1997) introduced a discordancy measure, which is a measure of discordancy between the observed L-moment ratios of a site and the

average L-moment ratios of a group of sites, i.e. the target region of interest. Based on this test, sites with gross errors or unusual behaviours can be flagged within the group of sites. To demonstrate this test, let $u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$ be the vector of at-site L-moment ratios for the region and let $\bar{u} = \sum_{i=1}^{N} u_i / N$, where $\bar{u}$ is the unweighted regional average value. The discordancy measure for site *i* in the region is then defined as:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}), \tag{3.1}$$

where *A* is the matrix of sums of squares and cross-products and that is given by:

$$A = (N-1)^{-1} \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T. \tag{3.2}$$

Hosking and Wallis (1997) suggested that a site should be considered as discordant if D_i value is large. The critical value to determine how small/large this value should be depends on the number of sites in the region (i.e. *N*). In general a site can be considered discordant if D_i is greater than 3 (for all *N* values greater than and equal to 15). However, this critical value can be as small as 1.917 for *N* = 7 and 2.971 for *N* = 14. In practice, the sites having high D_i values are either removed from the set of available sites for the region or moved to a different adjoining region. In addition to this statistical assessment, the removal of site should also be supported based on physical reasons and instrumental or historical evidence associated with the apparent discordancy.

3.3 Delineation of Homogeneous Regions

Identification of homogeneous regions is a challenging task for RFA. In many situations one can also end up taking subjective decisions. The objective of this step is to delineate groups of sites/stations such that their higher order statistics (e.g. L-coefficient of variation, L-skewness or L-kurtosis) are nearly identical, except for a site-specific scale factor and that is often taken as the mean/median of the variable of interest (i.e. low flows in the present context). Several methods have been proposed for grouping sites into homogeneous regions for RFA and that can roughly be grouped into the following categories:

Geographical proximity: In many studies, regions are defined based on geographic proximity and convenience and sometimes owing to the constraints imposed by administrative units of a larger country, province or watershed (e.g. Beable and McKerchar, 1982; CEH, 1999; Hortness and Berenbrock, 2001) or major physiographic and administrative regions of a large river basin (e.g. Matalas et al, 1975; Grandry et al., 2012; Masud et al., 2016a, 2016b), especially international rivers (e.g. Great Lakes watershed). According to a few earlier investigations by Wiltshire (1986a, 1986b) and Acreman and Sinclair (1986), geographic proximity cannot guarantee statistical and hydro-climatologic similarity as some watersheds within the same geographic region could be very different from the viewpoint of large scale atmospheric mechanisms responsible for extreme hydrologic conditions, such as low and high flows. In certain cases, some investigators have to exercise sound judgements as to the similarity of runoff controlling mechanisms for low and high flow situations and then ensuring similarity of derived

statistics in subsequent analyses (e.g. Parida et al., 1998; Kachroo et al., 2000; Shi et al., 2010; Mladjic et al., 2011). According to Salinas et al. (2013), the main rational behind geographic groupings is that the watersheds that are geographically close to each other may exhibit similar low flow generating processes.

The notion of subjective and objective considerations: Sites can also be grouped together subjectively by inspecting dominant characteristics of target sites, especially for small scale studies (e.g. sub-basins of a major watershed). Consequently, the resulting regions may or may not be geographically contiguous and may or may not share same physical characteristics. For example, sites located in sporadic and continuous permafrost regions could be considered as two distinct groups, but they may lack statistical homogeneity. The regions so formed need to be formally tested using heterogeneity measures. Gingras et al. (1994) formed regions for annual maximum streamflow analysis in Ontario and Quebec by grouping sites according to the time of occurrence of the largest flood (i.e. based on seasonality measures). Similarly, Laaha and Blöschl (2007) considered seasonality measures to divide Austria into eight homogeneous regions based on seasonality of low flow occurrences. Contrary to the subjective considerations, regions can also be formed objectively by assigning sites to one of two groups depending on whether a chosen site characteristic does or does not exceed a pre-defined threshold. This threshold is generally chosen to minimize within-group heterogeneity. Wiltshire (1985) used a few basin characteristics to group them together. In an iterative fashion, the optimum size of the region can be defined by minimizing within-group variability of the targeted statistics. Pearson (1991a, 1991b) applied a similar procedure and used within-group variation of sample L-moments as the objective criterion. Hosking and Wallis (1997) described this procedure as an effective approach and also proposed a homogeneity test for testing the homogeneity of regions so formed.

Cluster analysis: It is a standard method of statistical multivariate analysis and is used for dividing a collection of sites into distinct groups. This method has been used in several studies to form regions for RFA. In this method, sites are represented by a vector of site characteristics and the sites are grouped according to the similarity within the space of these characteristics. De Coursey (1973) was the first to apply this method to form groups of sites having similar peak flow response. In addition, Acreman and Sinclair (1986), Burn (1989), Guttman (1993), and Lim and Lye (2003) also used this method for identifying homogeneous regions for RFA. Grandry et al. (2012) used cluster analysis to form homogeneous regions within the Walloon region of Belgium using 25 climatic and physiographic characteristics and data from 59 gauging sites. Following the work of Hosking and Wallis (1997), who favoured cluster analysis for the delineation of homogeneous regions or groups of stations, the use of cluster analysis exploded for RFA, especially for flood frequency and rainfall frequency analyses. This could also be due to the fact that all necessary software modules were freely made available by Hosking and Wallis (1997). Thus, it was just a matter of running these modules on new datasets. These authors have also provided additional guidance on the use, constraints, and limitations of this method, in

addition to the insights into the maximum and minimum number of sites required in the homogeneous region to achieve a reasonable level of performance of the RFA approach.

In summary, an objective procedure for delineating homogeneous regions and specific procedures for confirming homogeneity of already identified geographic or climatic regions or regions formed from other perspectives need to be adopted. This is important for satisfying theoretical assumptions that underpin RFA approach. One such procedure, which Hosking and Wallis (1997) developed based on L-moments, is described below. Additional information on forming non-contiguous homogeneous regions based on the concept of nearest-neighbours or neighbourhoods is discussed in Chapter 4.

3.4 Regional Homogeneity Tests

After deciding on regions based on preferred site characteristics or other features of interest, it is important to verify whether a given region or group of sites is statistically homogeneous or it requires to be divided further into sub-regions in order to facilitate application of RFA procedure. The hypothesis of regional homogeneity is based on the assumption that the at-site probability distribution functions in a homogeneous region are identical, except for a site-specific scale factor, which is commonly taken as the mean or median flow–the index flow. The test of significance is generally developed by selecting a few at-site statistics. For example, Dalrymple (1960) tested the regional homogeneity based on 10-year flood magnitudes estimated from the Gumbel distribution. Thus, the regional distribution was pre-supposed in this case.

Wiltshire (1986 a, b) also proposed two statistical hypothesis tests for testing regional homogeneity. The first test involved testing regional homogeneity based on the coefficient of variation of the standardized annual maximum series, while the second test involved distribution based testing that used the geometry of the cumulative distribution function of the dimensionless regional distribution. A non-parametric jack-knife procedure was used to estimate at-site distributions in order to evaluate the regional homogeneity. Chowdhury et al. (1991) also suggested a statistical test based on L-moments; their test was more powerful than existing tests at the time. Hosking and Wallis (1993; 1997) proposed a regional homogeneity test based on sample L-moment ratios, which seems to be holding firm ground until today (cf. Castellarin et al., 2001; Lim and Lye, 2003). This test compares the variability of L-moment ratios of all sites within a region with the expected random variability, obtained through Monte Carlo simulation experiments.

In general, it is useful to start off with a larger region and successively form and test smaller homogeneous units. In certain cases, it is also possible to merge, for example, two already defined regions into a larger region. This is generally done to increase the number of available gauged locations within the region for a defensible analysis. According to Hosking and Wallis

(1997), a region's homogeneity can be tested based on L-CV, L-skewness and L-kurtosis. In the case of L-CV, one calculates the V statistic as follows:

$$V = \frac{\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^{N} n_i} \tag{3.3}$$

where t^R is the sample size weighted regional average L-CV. Similarly, the V static corresponding to sample size weighted regional average L-skewness (t_3^R) and L-kurtosis (t_4^R) can be calculated. The notation using R as a superscript emphasises on the notion of a region. A four parameter kappa distribution is fitted using weighted average L-moment ratios, i.e. 1, t^R, t_3^R and t_4^R. Additional detail on the kappa distribution and statistical reasoning for the choice of this distribution can be found in the work of Hosking and Wallis (1997).

With the fitted kappa distribution, a large number of homogeneous kappa regions, say N_{sim}, are simulated, each region having the same number of sites with exactly the same record lengths as their original counterparts. For all simulated regions, the V statistic is calculated and based on N_{sim} values of V, the mean μ_V and the standard deviation σ_V are calculated. After that, the heterogeneity measure H is calculated as:

$$H = \frac{(V - \mu_V)}{\sigma_V}. \tag{3.4}$$

Hosking and Wallis (1997) suggested considering a region as "acceptably homogeneous" if $H < 1$, "possibly heterogeneous" if $1 \leq H \leq 2$, and "definitely heterogeneous" if $H > 2$. Some authors, including Robson and Reed (1999), considered a relatively relaxed criterion, i.e. a region could be considered heterogeneous if $2 \leq H \leq 4$ and strongly heterogeneous if $H > 4$. In practice, some compromises are made as it is difficult to attain combined homogeneity based on H values associated with L-CV, L-skewness and L-kurtosis (e.g. see Mladjic et al., 2011). To obtain reliable values of μ_V and σ_V, the number of simulations N_{sim} should be large. Hosking and Wallis (1997) recommended 500 simulations, however, a larger number could also be used.

3.5 Selection and Fitting of a Regional Distribution

After verifying statistical homogeneity of a region, a single probability distribution is selected for the whole region through distribution fitting and evaluation procedure. Some investigators tend to evaluate the candidate distribution functions by assessing their ability to accurately estimate at-site quantiles on the basis of bias and root mean square error or some other similar measures. There is no unique and universally accepted probability distribution function for low flow analyses and therefore several families of distributions can be potential choices for RFA. From a practical standpoint, there could be a range of return periods for which quantile estimates are important for a given region. In that sense, the probability distribution that outperforms other distributions for this specific set of quantiles can be selected. Leave-one-out cross validation can also be used to select an appropriate distribution that can provide minimum errors for the

selected set of quantiles (see, for example, Lahaa and Blöschl, 2007). It must be noted that many distributions generally differ in the estimation of most extreme quantiles. In addition to a few notable earlier researchers (e.g. Matalas and Wallis, 1973; Kroll and Vogel, 2002), many other investigators (e.g. Caruso, 2000; Hewa et al., 2007; Liu et al., 2011) also corroborate the same idea, i.e. competing distribution functions that fit the observed data satisfactorily may differ significantly in the tails. When selecting a regional distribution, Cunnane (1989) suggested 'robustness' to be an important property of a distribution function for RFA. Different regional distribution functions were used in several studies on RFA. For example, CEH (1999) recommended the Generalized Extreme Value distribution for high flow analysis in the UK, while Durrans and Tomic (1996) recommended the log-Pearson Type III (LP3) distribution for regional low flow frequency analysis in the USA. Kroll and Vogel (2002) also recommended the LP3 distribution for low flows from intermittent sites and the three-parameter lognormal distribution for non-intermittent sites in the USA. Similarly, Chen et al. (2006) developed low flow frequency analyses for southern China using the three-parameter lognormal distribution as the most appropriate regional distribution, while Shi et al. (2010) used the Generalized Logistic (GLO) distribution for low flow analysis in Wujiang River basin in southwest China. Eris et al. (2019) also used the three-parameter lognormal distribution for at-site low flow analysis within entire watersheds in Turkey. Several studies on RFA have considered only the GLO, GNO (Generalized Normal), GEV (Generalized Extreme Value), PE3 (Pearson Type III), and GPA (Generalized Pareto) distributions for distribution fitting and evaluation purposes. The main reason behind this trend/practice is that these distribution choices are readily available in a software package, developed by Hosking and Wallis (1997), in addition to their flexibility in fitting complex datasets. For low flow analysis, different forms of the Weibull, Gumbel, PE3, and log-normal distributions have frequently been used (Smakhtin, 2001). In Zaidman et al. (2003), the GEV, PE3, GPA, and GLO distributions were evaluated using annual minimum flows from the UK. They found that for averaging intervals less than 60 days, data from high storage watersheds were best fitted to the GLO and GEV distributions, while that from low storage watersheds were best described by the PE3 or the GEV distribution.

Hosking and Wallis (1997) noted that distributions with three to five parameters are appropriate candidates for RFA because they yield relatively less biased estimates of extreme quantiles. As mentioned above, it is very likely that more than one distribution functions will fit the regional data adequately. In this situation, the best choice would be the one that provides the most robust estimates of quantiles (i.e. quantile estimates with minimum standard errors). Additionally, Hosking and Wallis (1997) and Minocha (2003) also suggested that the final choice should be made based on some formal 'goodness-of-fit' measures. They formulated an approach based on regionally averaged values of sample L-moments. For a three-parameter distribution, the goodness-of-fit can be judged by how closely the L-kurtosis of the fitted candidate distribution matches its regionally averaged counterpart from the observed data from all sites (Pandey et al., 2001; Yue and Wang, 2004a, 2004b; Yue and Pilon, 2005).

McCuen (1985) suggested the use of moment ratio diagram by plotting observed moments and visually selecting the distribution that is judged to be the most appropriate choice. The basic advantage of using the moment ratio diagram is that a single diagram can visually compare the fit of several candidate distributions for a given set of observed data. In the regional context, the regionally averaged dimensionless moments are plotted on the diagram for the choice of a regional distribution. Similar to the moment ratio diagram, Hosking (1990) introduced the idea of L-moment ratio diagram. According to Vogel and Fennessy (1993) and Peel et al. (2001), L-moment ratio diagrams are more accurate than the moment ratio diagrams in discriminating the (so called) true distribution from a set of candidate distributions. Since their introduction, the L-moment ratio diagrams are being routinely used in hydrology, most importantly the one involved L-skewness and L-kurtosis (see Chapter 2). However, Hosking and Wallis (1997) did indicate that the L-moment ratio diagrams provide just a visual guidance and therefore the choice should be backed with formal statistical testing and robustness of the distribution. The latter idea was also supported by Cunnane (1989). The ultimate objective of these assessments is to assist the analyst in the choice of a distribution that is most suitable for RFA.

Several methods are available in the literature for testing the goodness-of-fit of a distribution. Among these, Hosking and Wallis's (1997) Z goodness-of-fit test is commonly used in conjunction with L-moment ratio diagrams in RFA-oriented studies. Some insights on the L-moment ratio diagram are already presented in Chapter 2. In the case of RFA, regionally weighted average values of L-moments are plotted on this diagram and the location of this point helps identify a potential distribution. This is formally complemented with the results from the Z good-of-fit test. For implementing the Z-test, each candidate distribution is fitted to the regional average L-moment ratios and then the theoretical L-kurtosis of the fitted distribution is calculated based on the theoretical relationship (i.e. τ_4^{DIST} is computed). Afterwards, the same Monte Carlo simulation procedure as for estimating the heterogeneity measures is followed to simulate a large number of kappa regions. For each simulated region, the regional average L-kurtosis, $\tau_4^{R(m)}$, is calculated. The Z goodness-of-fit measure for each candidate distribution is obtained from the following equation:

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4)/\sigma_4 \tag{3.5}$$

where B_4 is the bias of t_4^R and σ_4 is the standard deviation of t_4^R. The bias B_4 and the standard deviation σ_4 are respectively defined as:

$$B_4 = \sum_{m=1}^{N_{sim}} (t_4^{R(m)} - t_4^R)\,/N_{sim}, \text{ and} \tag{3.6}$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1}\left\{\sum_{m=1}^{N_{sim}} (t_4^{R(m)} - t_4^R)^2 - N_{sim} B_4^2\right\}\right]^{1/2}. \tag{3.7}$$

The candidate distribution being tested is declared to offer an adequate fit if Z^{DIST} is sufficiently small. Hosking and Wallis (1997) suggested that a reasonable criterion to use would be $|Z^{DIST}|$ being ≤ 1.64. It is likely that more than one distribution may satisfy this criterion. In that

situation, one possibility is to select the distribution with the lowest Z^{DIST}. Still another possibility would be to invoke the concept of robustness, i.e. evaluate root mean square error (RMSE) associated with the specific set of required quantiles and select the distribution that is associated with the lowest RMSE.

3.6 Estimation of Low Flow Magnitudes

After delineating a potential homogeneous region and ensuring its statistical homogeneity and the selection of a suitable regional distribution, a reginal growth curve (i.e. $q \sim T$ or $q \sim F$ relationship, where q is a dimensionless low flow quantity) is developed, which is used to derive various regional growth factors corresponding to desired return periods / return frequencies. For example, if T is the desired return period then the growth factor can be represented in terms of the return period T (i.e. $q(T)$) or in terms of the non-exceedance probability F (i.e. $q(F)$). For any site, the low flow quantile $Q(F)$ can be derived using the following relationship:

$$Q(F) = \mu q(F) \tag{3.8}$$

where μ is the index flow, i.e. at-site mean or median low flow (i.e. $\bar{Q}$ or Q_{med}). In a given region, there will be as many values of μ as there are the number of sites.

Estimation of Index Flow

For estimating a T-year return period low flow quantile at any ungauged site within the homogeneous region of interest, an estimate of the index flow μ (e.g. the mean annual low flow) is required. Since observed low flow data is not available at ungauged sites, the at-site mean at the ungauged site cannot be computed. To have an estimate of the index flow at ungauged locations within the region, it is necessary to establish a relationship between the index flows from gauged locations and physiographic and climatic attributes of associated watersheds within the homogeneous region. Ideally, watershed attributes should represent a broad range of features representing hydrologic, hydraulic, geologic, and meteorological factors that could influence the quantity of low flows in a watershed. Both linear and/or non-linear regression relationships have often been considered for developing these relationships. A generalized and perhaps a convenient relationship that could be optimal in some situations can be of the following form:

$$\mu = f(W_1, W_2, \dots, W_m) = \alpha_0 W_1^{\alpha_1} W_2^{\alpha_2} \dots W_m^{\alpha_m} = \alpha_0 \prod_{i=1}^{m} W_i^{\alpha_i} \tag{3.9}$$

where, $W_1, W_1, \dots, W_m$ are the m watershed attributes and $\alpha_0, \alpha_1, \alpha_1, \dots, \alpha_m$ are regression parameters/coefficients. Log transformation of the above equation can produce a relation that is homoscedastic, i.e. the standard error is the same throughout the ranges of the independent variables (Riggs, 1990). The selected relationship is then transposed to ungauged locations, where low flow quantiles are required, using the same physiographic and climatic attributes from those locations. After having estimated the index flow through transposed relationships, the desired low flow quantiles can be obtained following the above relationship (Equation (3.8)). It

is important to note that the growth factors derived from the regional distribution are applicable across the entire homogeneous region under the RFA assumptions. Additional insights on the development of regional regression relationships and attribute selection are provided in the section to follow and are also discussed in Chapter 6. With respect to regression relationships, it is useful to diagnose regression residuals to examine patterns that might be undesirable and require additional investigations.

3.7 Regression-on-Quantiles as a Regionalization Approach

The regression-on-quantiles is another popular approach that is used for regionalization of low flows in many parts of the world. It involves establishing regression relationships between at-site low flow indices and watershed attributes. For example, the Q95 percentile flow from the flow duration curve and selected quantiles (i.e. 7Q10 or 7Q2) from low flow frequency analyses have often been used for developing these relationships. The Q95 is the discharge that is exceeded 95% of the time and 7Q10 (7Q2) is the 7-day average low flow quantile corresponding to 10-year (2-year) return period. The low flow quantiles at all gauged locations within the target region are estimated using at-site frequency analysis (see Chapter 2). For the USA, the US Geological Survey has developed regional multiple regression equations between low flow quantiles and watershed attributes representing terrain, land cover, soil type, and precipitation (e.g. Tasker, 1987; Barnes, 1986; Ries, 2002). Along similar lines, the state of Idaho was divided into eight regions by a cluster analysis of watershed attributes, and then separate regression equations were developed for each region (Hortness and Berenbrock, 2001). In the UK, a pooled regionalization strategy was adopted by Holmes et al. (2002). In their approach, Q95 low flow indices were estimated by a weighted average of standardized Q95 values of 10 gauged reference watersheds that were most similar in terms of soil classes. For Austria, Lahaa and Blöschl (2007) also used the Q95 index in an effort to develop national low flow estimation procedures. This index was selected because it was widely used in Europe and was chosen because of its relevance for numerous water resources management applications (e.g. Kresser et al., 1985; Gustard et al., 1992; Smakhtin, 2001). For their study area, Q95 was found to be correlated with the mean annual minimum flow, but was deemed more robust to data errors. The whole country was divided into eight homogeneous regions from low flow seasonality perspectives and 31 physiographic attributes (without any transformation) were considered for regional regression analyses. In addition, they also evaluated the residual pattern approach (e.g. Hayes, 1992; Aschwanden and Kan, 1999), weighted cluster analysis (Nathan and McMahon, 1990), and regression trees (Breiman et al., 1984) to form homogeneous groupings of watersheds. To minimise inter-variable correlations and multicollinearity, a stepwise regression approach was adopted and Mallow's C_p (Weisberg, 1985) was used as the criterion of optimality. However, care must be exercised in using stepwise regression as some investigators caution on the limitations of this approach and those are often neglected in practical applications (e.g. Whittingham et al., 2006).

The regression-on-quantiles as a low flow regionalization approach utilizes regional regression relationships of specific low flow indices (e.g. Q95/7Q10 divided by the drainage area) and watershed attributes, developed independently for each region of a larger study area (e.g. a whole country); each region is assumed homogeneous in terms of low flow processes and predominant characteristics of low flow sequences. Thus, delineation of homogeneous regions for the implementation of regression-on-quantiles is an important pre-requisite. Stating it differently, some form of homogeneity is ensured so that the developed equations can be applied at all ungauged locations within the entire region. It can also be stated that the regression-on-quantiles approach consists of coupling homogeneous region delineation methods with regression analyses. Whether to standardize Q95/7Q10 or not for regression purposes varies from one study to another. If the drainage area is considered as an independent predictor then the Q95/7Q10 (or another index) can be directly regressed on the drainage area, along with other watershed attributes. Alternatively, drainage area standardized values can also be regressed on watershed attributes, excluding the drainage area (e.g. see Grandry et al., 2012). Some investigators (e.g. Barnes, 1986) have found this latter option more suitable for multiple regression relationships. In principle, the regression-on-quantiles approach is very similar to the estimation of index flow in Hosking and Wallis's (1997) RFA approach. Although regressions with watershed attributes were considered in many countries and jurisdictions, many other aspects of the regional estimation procedures differed significantly from one study to another. For example, the procedures involved in delineating homogeneous regions.

When regression-on-quantiles is considered as a regionalization approach, record augmentation techniques are often adopted. This is due to the reason that this approach is heavily dependent on reliable estimates of at-site low flow indices because estimates from short records not only deviate significantly from long-term behaviours due to climatic and other sources of variability, they are also associated with large uncertainties. Record augmentation can be accomplished using information purely from the neighbourhood of a target site using weighted transposition of recorded observations from donor sites or using drainage area based proportional techniques where applicable (see Chapter 4 for more information on this procedure). The regression-on-quantiles can take on a number of different forms, including linear and nonlinear relationships. It is possible that a number of such regression relationships can be equally likely candidates for a given region. In that situation, leave-one-out cross validation can be used to select a reasonable regression relationship. When applying this approach, one withholds low flow indices and watershed attributes from a particular site/location, makes an estimate of these indices at that site using the regional regression developed from rest of the sites and then compares the estimated values with at-site indices. This procedure is repeated for every site within the region. The performance of each regression equation can be assessed in terms of root mean square error or other similar measures (e.g. Nash-Sutcliffe Efficiency criterion; Nash and Sutcliff, 1970). This strategy of leave-one-out cross validation fully emulates ungauged cases. The advantage of cross-validation over other assessment techniques is its robustness and its ability to evaluate the

performance of various competing approaches even if the underlying assumptions are not fully satisfied (Efron and Tibshirani, 1993; Lahaa and Blöschl, 2007; Khaliq et al., 2009; Mladjic et al., 2011). The regression relationship with the smallest cross-validation error can be selected and recommended as a national regionalization approach for the estimation of low flow indices at ungauged locations. For estimating cross-validation errors, k-fold cross validation procedure can also be used if the number of available sites permits to do so.

For the implementation of regression-on-quantiles approach for a given ungauged location, one needs to develop quantitative estimates of watershed attributes and that may not be feasible for many consulting projects. To overcome such difficulties, regional maps of low flow indices, developed through non-linear interpolation techniques, can be very useful. For example, Engeland et al. (2006) recommended that a low flow map be developed to be used as a national guidance.

3.8 Emerging Techniques

With reference to the information presented above in Sections 3.2 to 3.7, estimation of low flow quantiles at ungauged locations can be seen as a regression problem. Therefore, estimation of low flow quantiles at ungauged locations can also be conducted within the framework of rapidly expanding ML approaches (Breiman, 2001; Lesmeister, 2015; Theobald, 2017; Gupta et al., 2020), such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbours (KNNs), Boosted Regression Trees (BTR), and various forms of the Artificial Neural Networks (ANNs). It is believed that the conventional regression or ML models when applied within homogeneous or hydrologically similar regions can significantly improve estimates of low flows at ungauged locations, compared to their application to a random group of sites (Smakhtin, 2001; Laaha and Bloschl, 2007; Vezza et al., 2010; Tsakiris et al., 2011). According to the comparative assessment conducted by Salinas et al. (2013), the regional regressions always performed better than the global regressions. ML models learn from complex data patterns and inter-variable relationships to predict the target variable of interest. In addition to hydrology and water resources engineering and management related areas, machine learning models have also shown considerable success in other scientific disciplines, such as economics, automobile, remote sensing, speech recognition, image processing, medical sciences, among several other fields. It is expected that with enough tuning and training on sufficient observational and/or experimental data, ML models will be able to provide accurate estimates/predictions of the target variable of interest. Over the last 10 years or so, the use of ML approaches to solve applied problems has seen an exponential growth and this trend is expected to continue in the future.

The ANNs have become very popular in hydrology for modelling a number of hydrologic variables and processes, e.g. low flows, flood magnitudes, streamflow forecasting, sediment transport, water quality, etc. The ANNs can describe the relationship between inputs and outputs, without making explicit assumptions on the model parameters and the system being modelled

(e.g. Govindaraju and Rao, 2010; Tanty and Desmukh, 2015; Oyebode and Stretch, 2018). In addition to RF, SVM, KNNs and BRT, several variants of ANN-based models can also be considered as candidate member models in ensemble learning frameworks and therefore can be very effective for regression oriented problems (e.g. Green and Ohlsson, 2007; Zaier et al., 2010). It is expected that an ensemble framework can perform better than any of the individual member models. This framework also facilitates a reasonable course for the quantification of prediction uncertainties. In ensemble learning frameworks, a combiner is generally required for integrating results from individual models and for that purpose an ANN-based combiner can be quite useful compared to a simple averaging procedure. Thus, ensemble based modelling is preferable from an application perspective (e.g. Shu and Ouarda, 2007; Alam et al., 2020). Ouarda and Shu (2009) used this type of approach for modelling quantiles of summer and winter low flows at ungauged locations in Quebec. In their study, the ANN-based modelling showed better performance than the commonly used regression analyses.

Compared to linear and non-linear regression analyses, pre-processing of input and output data are important for ANN-based modelling using techniques like linear transformation–such as linear scaling and normalization, and nonlinear transformations–such as logarithmic and Box-Cox transformations (Box and Cox, 1964). Such a treatment is helpful in rapid convergence of the training algorithm and for constraining the problem within reasonable bounds. However, the amount of pre-processing does depend on the requirements of a given ML model and the nature of available data (e.g. Ouarda and Shu, 2009; Govindaraju and Rao, 2010). In developing ML models, it is a common practice to divide the available data into training, validation and testing portions. Model training is accomplished on the training set, with feedbacks from the validation set, and model testing is accomplished on the test set, which is kept unseen to the model. In some applications, merely training and testing strategy is used, with 70/30 percent split of the available data.

Recently, Alobaidi et al. (2021) demonstrated application of ML approaches for estimating low flow quantiles at ungauged locations within an ensemble framework considering data from a group of near natural stations from southern Quebec. However, they did not evaluate whether those stations form a homogeneous region or not. Low flow quantiles of 2-, 5- and 10-year return periods for 7- and 30-day durations for winter and summer seasons were modelled separately. Seven physiographical and meteorological attributes were considered, including drainage area, percent of the watershed area covered by forests and lakes, annual mean degree days less than zero degrees Celsius, annual mean number of days with temperature above 27 °C, summer seasonal mean liquid precipitation, and curve number, representing soil characteristics. After pre-processing input and output variables (i.e. watershed attributes and low flow quantiles), an ensemble of ANN-based models was trained and tested (in a prediction mode) using a jack-knife approach. They concluded that the ensemble modelling approach significantly reduces the errors associated with quantile estimates compared to those obtained from a single model. This was a reasonable attempt to explore the strengths of ML models for estimating low flows at ungauged

locations. However, the reader is cautioned that a number of misconceptions and misinterpretations were also noticed in their published study. Bond and Kennard (2017) also studied extrapolation of a large number of river flow metrics pertaining to central tendency, intermittency of high and low flows, seasonality and variability from locations with known values to those locations where no such information was available using four different ML approaches (i.e. RF, BRT, and Multivariate Adaptive Regression Splines (Friedman, 1991)) and several watershed attributes reflecting climate and physiographic features (i.e. attributes related to land use, water use, runoff, substrate, terrain, and vegetation cover). They found that these models delivered relatively high predictive performance relative to simple spatial transpositions and thus these models offer a practical alternative for generating river flow metrics for ungauged watersheds.

ML approaches can be integrated with the Hosking and Wallis's (1997) RFA approach for the estimation of index flow and as a competitive alternative to the regression-on-quantiles approach. However, it must be noted that like the conventional multiple linear or nonlinear regressions, where manual calculations are readily feasible if one wishes to do so, it is not possible to have an explicit functional form of low flow quantiles and watershed attributes in the case of ML approaches and hence computer-based computations are indispensable. Some investigators may find this reality as an unattractive feature and a barrier to practical applications of ML based regression models. In essence, ML approaches can be very powerful compared to conventional approaches to solve regression problems involving complex inter-variable relationships. In certain applications, for example streamflow forecasting, ML techniques in combination with conventional modelling approaches can be very powerful and the resulting framework can produce superior outputs compared to individual models.

3.9 Concluding Remarks

The index flood based RFA procedure was first introduced by Dalrymple (1960), in which the observed annual peak flows at each site were standardized by dividing by the sample mean (commonly known as the index flood), and then all standardized observations were pooled together to fit a probability distribution function which in turn facilitated a dimensionless frequency curve, which is also known as "growth curve". The desired site-specific quantiles within the homogeneous region were estimated by multiplying the dimensionless quantiles from the regional growth curve by the at-site sample mean. This procedure is called index flood procedure due to its original application to flood peaks. However, to generalize the application of this concept across multiple variables, index flood can be referred to as index flow. The index flow based procedure for RFA is very popular among practicing engineers and hydrologists, and the same has been adopted for conducting several RFA studies world-wide.

The main goal of RFA is to improve the reliability of estimated quantiles at gauged sites, especially those with shorter records because shorter records are unable to capture the natural

variability of low flows, and to be able to estimate desired quantiles of interest at ungauged sites where no records exist. In the latter case, it is important to have the index flow available at ungauged sites. In many applications of the RFA approach, regional linear or nonlinear regression relationships were developed by regressing the index flow onto watershed attributes (e.g. Lim and Lye, 2003; Mostofi-Zadeh et al., 2012). These relationships were then transposed to ungauged locations for the estimation of various quantiles. Other variants of this approach do exist in the literature. For example, the US Geological Survey (Thomas, 1987; Tasker, 1987) developed separate relationships for every quantile of interest and the watershed attributes and those relationships were transposed to ungauged locations within the region of interest, by estimating same watershed attributes at the location of interest. This method has also been used in many parts of the world and is commonly known as "regression-on-quantiles" method in the literature on RFA. Compared to the index flow procedure, this method avoids finding a regional distribution (and hence the regional growth curve). However, with the introduction of L-moments and formalization of the RFA approach by Hosking and Wallis (1993, 1997), the index flow method has become very popular and is well established in hydrology and other disciplines. Most of the earlier applications of the index flow based methodology were for flood frequency analysis. Overtime, its applications to several other variables in many engineering and scientific disciplines have exploded, including applications for regional low flow analysis (e.g. Pearson, 1995; Durrans and Tomic, 1996; Tate et al., 2000; Kroll and Vogel, 2002; Yurekli et al., 2005; Chen et al., 2006; Modarres, 2008; Shi et al., 2010; Dodangeh et al., 2014, among several others).

Various extensions of the Hosking and Wallis's (1997) approach have been proposed overtime. Among these, the concept of non-contiguous homogeneous regions is important to mention. The studies developed around this concept do not require the homogeneous regions to be geographically contiguous and therefore consider a neighbourhood or a group of nearest-neighbours as the target homogeneous region for any site in question. Consequently, each site is characterised by its own neighbourhood. The sites in the neighbourhood are identified based on similarity within the space of selected watershed attributes (e.g. Burn, 1990a, 1990b; Tasker et al., 1996). After forming neighbourhoods, it is also possible to ensure their statistical homogeneity based on the tests proposed by Hosking and Wallis (1997). Two such approaches that were also applied in Canada are the Canonical Correlation Analysis (Ouarda et al., 2001) and Region of Influence (Burn, 1990a, 1990b). These approaches are discussed in Chapter 4 in the context of regional analysis of flow duration curves.

Apart from the procedures for delineating homogeneous regions, the regional low flow estimation for ungauged locations can be considered as a regression problem and therefore ML approaches can also be used, in addition to conventional linear or nonlinear multiple regression approaches. ML approaches are becoming popular in many scientific disciplines, including hydrology, and their use in solving applied problems has grown exponentially over the last several years. These approaches learn from complex data patterns and inter-variable relationships

to predict the target variable of interest. Additionally, these approaches has shown considerable skill in modelling overly complex and non-linear problems. These approaches can easily be integrated with the L-moments based RFA and can also be considered as a reasonable alternative to the regression-on-quantiles approach. ML approaches have not been explored for low flow estimation at ungauged locations in Ontario and therefore could be a potential avenue of research for future studies.

Another approach that has received little attention is the use of process-based modelling and continuous streamflow simulation for generating low flow indices at ungauged locations. Such models are extensively used for streamflow forecasting and simulation to inform water resources development and management related projects. Specifically, the distributed versions of such models are ideal for generating low flow information at all ungauged locations across the entire watersheds. Though these models may require considerable amount of time and effort in model setups and involve large execution time, their role in generating low flow information cannot be overstated. Engeland et al. (2006) attempted this approach using a gridded version of the HBV model in southwestern Norway using two partitions of the region based on low flow seasons, i.e. winter and summer. They concluded that the low flow indices derived from transposition of regression relationships of low flow indices and catchment characteristics were better than those derived from HBV model simulations. Though this was an interesting conclusion, the importance of process-based models for simulating several other watershed functions is unequivocal. Therefore, the research along the lines initiated by Engeland (2006) continued. A modified version of the HBV model, namely MAC-HBV, was developed at the Water Resources and Hydroclimatic Modelling Lab of McMaster University in partnership with Ontario Ministry of Natural Resources and Forestry to estimate continuous flows and their characteristics at gauged and ungauged watersheds in Ontario (https://www.hydrology.mcmaster.ca/?at=machbv).

4 Regional Analysis of Flow Duration Curves for Ungauged Locations: State of Practice

4.1 General

The frequency with which a specific streamflow value is expected to be exceeded at a given location over a longer period of time is important for characterizing low flow regimes of a watershed. The flow duration curve (FDC), which is generally derived from continuous streamflow records, provides a graphical representation of streamflow variability and expected frequencies of streamflow values at a given location (see Figure 4.1). It is straightforward to derive FDCs at locations where continuous streamflow data is available. Compared to this, estimation of FDCs at ungauged locations is accomplished by a number of indirect means, including direct transposition of FDCs from gauged to ungauged locations, using empirical scaling or regression-based statistical methods, and also through process-based hydrologic modelling. For water resources development purposes, information about the entire FDC is generally required. However, for water extraction, licensing and waste load allocation purposes, the concentration is mainly focused on the lower portion of the FDC that typically reflects the behaviour of low flow regime of a watershed.

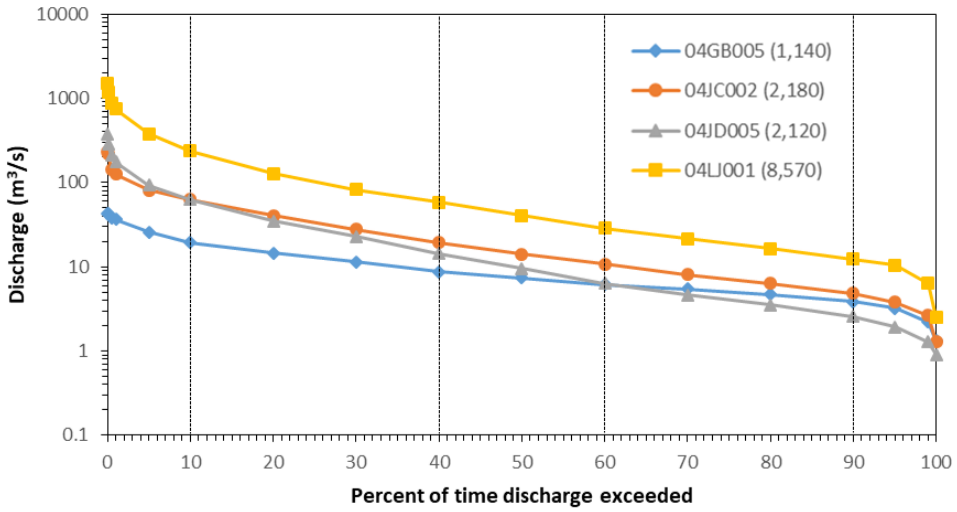

Figure 4.1: Flow duration curves for four sample streamflow recording stations of Environment and Climate Change Canada for the 1981–2010 period, reflecting the diversity in streamflow regimes. Watershed drainage areas (in km^2) are shown in the legend and arbitrary divisions in terms of different (high, intermediate, low and transitionary) flow regimes are also shown.

Since not all stream reaches are gauged perhaps partially due to lack of resources, it becomes important to have reasonable estimates of FDCs at all ungauged locations within a region of interest. Irrespective of the spatial extent of a country or a specific region and the amount of data to be processed, estimates of FDCs at ungauged locations need to be obtained in an efficient and

robust manner, with ideally minimal manual interventions. To achieve this objective, regionalization approaches, which being quite promising, are often used. Through these approaches, one could estimate FDCs at all ungauged locations in a target region of interest.

In general, two primary aspects are given consideration for regionalization: (1) the nature of the streamflow characteristic that needs to be estimated at the target ungauged location (e.g. annual or seasonal low flows, percentiles of FDCs, annual or seasonal high flows, etc.), and (2) the selection of a suitable technique for regionalization of selected streamflow characteristics based on data from gauged locations. This partitioning of the methodology is essential since most of the regionalization approaches are tied with the hydrologic variable being estimated at the targeted ungauged locations. Moreover, the above mentioned second step consists of two additional independent steps, i.e. (1) delineation of homogenous regions (DHR), which classifies or groups a number of source data sites (i.e. gauged locations) that exhibit similarity in terms of some selected features of interest (i.e. climatological, geographical, geological, statistical or other characteristics), and (2) selection of a regional estimation method (REM), which is employed to transfer the required information from source sites to the target ungauged site. Many of these aspects have already been discussed in Chapter 3 in the context of regional frequency analysis (RFA) of low flows. Here, the DHR and REM steps are elaborated further from slightly a different perspective, with a focus on regionalization of FDCs. However, the reader may find some resemblance between the concepts presented here and those discussed in Chapter 3.

As a large part of the Ontario's river and stream network is ungauged, reasonable methods from both national and international literature on ungauged hydrology need to be considered, tested and refined for Ontario-wide applications. Therefore, a number of selected studies on the estimation of streamflow characteristics at ungauged locations are reviewed, in addition to methods pertaining to transposition of FDCs from gauged to ungauged locations. During transposition of FDCs to ungauged locations, the impacts of river ice, ice jams or ice cover on the estimation of stream flows are not explicitly considered. In addition, studies related to the use of deterministic hydrologic models for the estimation of continuous streamflow at ungauged locations are not considered for this review. These models are often used for streamflow forecasting and warning purposes. Though computationally expensive and time consuming, distributed hydrologic modelling is a reasonable approach for generating FDCs at ungauged locations.

4.2 Delineation of Homogeneous Regions (DHR) or Neighbourhoods

The regionalization framework for estimating any streamflow index at ungauged locations is based on the principle/understanding that the sites with recorded data, which are more similar to the ungauged site within the space of selected watershed attributes, are the best possible

predictors of streamflow indices at the target site. Therefore, such sites should be considered in defining a homogeneous region or neighbourhood. The watershed attributes could be selected from observed streamflow statistics or could be derived from climatic and physiographic characteristics of watersheds. A mixture of these characteristics is expected to produce better results. However, the process of identifying homogeneous regions is generally tied with the variable of interest, e.g. high flows or low flows. Regions identified as homogeneous for high flow analysis may not be the same as those identified for low flow analysis. The main physical reason behind this disparity is that the underlying mechanisms that govern low flow occurrences may not be similar to those that govern high flow generating processes within various watersheds of a large geographic area. For the case of FDCs, considerable difficulty arises because the FDC represents the entire flow regime of a watershed, ranging from low flows to high flows, as well as flows in-between these extremes. With respect to FDCs, Dingman (1978) and Searcy (1959) noted that the lower flow ranges of a FDC are controlled less by climactic drivers than by basin geology and physiography, whereas in a runoff-dominated watershed, the local climate has a significant impact on higher flow ranges of an FDC. Consequently, regionalization of FDCs is a challenging task. Generally, precipitation amounts and temperature and evaporation patterns can affect streamflow patterns on large regional scales, while physical properties of watersheds (i.e. geology, land use, and presence or absence of surface water bodies) can affect streamflow patterns on local scales (Homes et al., 2002).

There are many possible approaches that can be used for delineating homogeneous regions. One of the most popular and the easiest to comprehend approach is to delineate geographically contiguous homogeneous regions based on the geographic proximity concept. If an ungauged site falls within a homogeneous geographic region then the characteristics of that region as a whole (from all sites within the region) are used to estimate target streamflow indices at the ungauged location. Acres Consulting Services Limited (1984a) was the first study in Canada that identified hydrologic homogeneous regions at the national scale. In that study 12 hydrologic homogeneous regions were identified. This was done by first identifying a number of predefined physiographic regions within Canada and then sub-dividing them by the presence or absence of permafrost and based on the differences in regional climatic parameters (Acres Consulting Services Limited 1984b). Among other regional studies wherein a similar approach was used is Gingras et al. (1994), who identified nine homogeneous regions in Ontario and Quebec based on statistical characteristics of flood flows. In a similar manner, the Ontario Ministry of Environment and Energy (MOEE) delineated homogeneous regions within the province of Ontario based on various characteristics of low and high flows (Chang et al., 2002). For instance, to predict low flows at ungauged locations, homogeneous regions were delineated within different administrative regions of the Ministry (MOEE, 1995), while for predicting high flows, 12 different regions were delineated (Moin and Shaw, 1985, 1986). In addition to these studies, there are other provincial and regional studies wherein homogeneous regions were identified and used (e.g. Loukas and Quick, 1995; Wang, 2000; Eaton et al., 2002, etc.).

Cluster analysis is commonly used for delineating homogeneous regions (e.g. Tasker, 1982; Nathan and McMahon, 1990; Leboutillier and Waylen, 1993). This technique identifies different groups or clusters of sites/stations based on similarity of statistical, climatic, geophysical and hydrologic attributes. For the case of FDCs, clustering is generally based on similarity of several attributes. Once homogeneous regions are identified, desired percentiles of FDCs at ungauged locations within the identified homogeneous regions can be estimated. For marginal cases, when ungauged locations are suspected to belong to more than one cluster, a weighting scheme is generally adopted. In the literature, this approach is also referred to as fractional membership technique (e.g. Acreman and Wiltshire, 1989; Srinivas et al, 2008; Satyanarayana and Srinivas, 2011; Asong et al., 2014). Based on studies conducted in Canada and elsewhere (e.g. Tasker, 1982; Leboutillier and Waylen, 1993; Ottawa Engineering Limited, 1997; Acres Consulting Services Limited, 1984a; and Natural Resources Canada, 2004a, 2004b), the use of geographically contiguous homogeneous regions seems to be the most popular approach. In certain circumstances, especially when the number of gauging stations is very limited, delineation and use of non-contiguous homogeneous regions are also reported in the hydrologic literature. Two such techniques that have been employed in Canada and other parts of the world are described below.

4.2.1 The Region of Influence (ROI) Approach

The ROI approach is used to identify homogeneous regions or neighbourhoods that are not necessarily geographically contiguous. The gauging sites (or donor sites) included in such regions share similarity within the space of selected hydrologic, climatic or physiographic attributes. Each site is assumed to be associated with a specific region. This technique was proposed originally by Acreman and Wiltshire (1989) for the UK, but has been used in many other parts of the world, including Canada (e.g. Burn, 1990a, 1990b; Eng et al., 2005). Though not necessary, a weighting function is used to weight individual sites depending upon a similarity/dissimilarity measure in the form of Euclidian distance, calculated for a set of attributes within the attribute space. The biggest advantage of the ROI approach is that it allows formation of homogeneous regions that can contain a large number of sites and that in turn can be useful to obtain robust estimates of desired quantiles. If a geographic homogeneous region contains a smaller number of sites then the ROI approach can be useful to expand the number of neighbouring sites for that region and to obtain relatively more reliable estimates of desired quantiles. Tasker et al. (1996) and many others (e.g. Burn 1990a, 1990b) have used this approach for RFA. Holmes et al. (2002) used the ROI approach to estimate FDCs at ungauged locations.

4.2.2 Canonical Correlation Analysis (CCA)

The CCA approach is a multivariate statistical technique that permits establishment of interrelationships between two groups of variables by determining linear combinations of one group that are most correlated to linear combinations of the second group. The CCA technique

has been employed as a regionalization method for flood frequency analysis, where one group of variables represents flood characteristics and the second group represents physical and climatological characteristics of watersheds. The principle being that by knowing the second the first can be predicted (Bobée et al., 1996). For a given gauging site, a homogeneous region or a group of sites can be identified by examining the proximity of the site to other gauging sites within the canonical space of attributes. A chi-squared distance measure is used to identify neighbouring sites for each ungauged location. This procedure has been applied for flood frequency analysis in Quebec (Ribeiro-Correa et al., 1995) and Ontario (Ouarda et al., 2001). The applications of this method for determining FDCs at ungauged locations are relatively limited in the literature, but are emerging slowly (e.g. Bomhof, 2013).

4.3 Regional Estimation Methods (REMs)

A number of REMs from the literature that have shown some promise for estimating FDCs at ungauged locations are discussed below.

4.3.1 Index Flood Method

This method was proposed by Dalrymple (1960) for regional flood frequency analysis, but can also be used for other variables of interest as has already been discussed in Chapter 3. The principle is that the at-site flood frequency curves in a homogenous hydrologic region are identical, except a scale factor (e.g. at-site mean or median flow) that can be described in terms of watershed attributes (e.g. climatic, physiographical or other characteristics). First, this method requires identification of homogenous regions and then determining a standardized (or normalized) flood frequency curve, commonly known as ‘growth curve’, for the entire homogeneous region (see Chapter 3). The growth curve is assumed to be applicable across all sites within the region, including ungauged locations. In the original application of this method, delineation of geographic regions and pooling of standardized flood flows to derive the regional growth curve were considered. Derivation of site-specific flood flow quantiles is discussed in Chapter 3, where one multiplies the growth factors obtained from the growth curve corresponding to given return periods with the at-site index flood. This concept from regional flood frequency analysis was borrowed by some investigators for determining FDCs at ungauged locations. Acres Consulting Services Limited (1984a) employed this concept and normalized FDCs using the 2-year return flow that needed to be estimated at the target site of interest. Natural Resources Canada (Natural Resources Canada 1984a, 1984b) determined a representative FDC for a region of interest and normalized it using the mean annual flow. The mean annual flow was estimated at the target location by employing specific runoff values from published national maps and by calculating the related drainage area at the target location. Some variants of this approach were also implemented in Smakhtin et al. (1997) and Smakhtin and Masse (2000), wherein the normalization was conducted using the mean annual flow.

4.3.2 Drainage Area Ratio Method

This is one of the simplest methods for estimating streamflow values or any derived index (whether pertaining to low flows, high flows or FDCs) at ungauged locations. In this method the target index from a source site is scaled based on the ratio of drainage areas, as shown below:

$$Q_u = Q_g \left(\frac{A_u}{A_g}\right)^m \tag{4.1}$$

where Q_u and Q_g are respectively the streamflow indices for the ungauged and gauged locations, A_u and A_g are respectively the upstream drainage areas at the ungauged and gauged locations, and *m* is a calibration parameter that accounts for non-linearity of the relationship. The parameter *m* requires calibration, but is often taken as unity for simplicity reasons (Mohamoud, 2008; NRC-CHC, 2008; Shu and Ouarda, 2012). Using this method, a complete FDC can be generated at an ungauged site either by estimating continuous streamflow data or by estimating selected percentiles of the FDC. However, caution is generally warranted when using this method as the relationship between streamflow and drainage area is affected by a number of physiographic, climatic and other factors. The strength of the relationship drops off quickly as the drainage area ratio diverges significantly from unity (Copeland et al., 2000; McCuen and Levy, 2000). A number of studies have employed this method (e.g. Gulliver and Murdock, 1993; Mohamoud, 2008; NRC-CHC, 2008). Drainage area differences of more than 25–50% have been considered as the applicability limits of this method (McCuen and Levy, 2000; Durand et al., 2002). In addition, this method performs better when the ungauged site is located on the same river, upstream or downstream of the gauging site. Mohamoud (2008) proposed modified drainage area ratio methods for the US Mid-Atlantic region, but their modifications have seen very limited applications as they have never been evaluated in later studies on ungauged hydrology.

4.3.3 Parametric Characterization of FDCs

In parametric characterization of FDCs, FDC is assumed to be represented by analytical relationships. These relationships could be in the form of polynomial or exponential relationships. The parameters of the relationship are estimated through regional analyses. One of the first FDC regionalization and transposition study was conducted by Quimpo et al. (1983), who was the first to propose parametric characterization of FDCs. Though the approach is parsimonious in nature, it makes the FDC inflexible due to constraining the shape of the FDC. Applications of this method are rare in the hydrologic literature. Franchini and Suppo (1996) also proposed a parametric technique for estimating the FDC by fitting a curve to three selected percentiles of the FDC. The authors proposed two possible relationships for describing the lower portion of the FDC. This technique was later extended by Castellarin et al. (2004), who considered four percentiles instead of three. This method regionalizes streamflow percentiles rather than parameters as a function of watershed attributes. Mandal and Cunnane (2009) also presented a parametric method, wherein they considered lower three-quarters of the FDC (i.e. the

range from 25th to 99.99th percentiles) and proposed a parametric method for regionalization of the FDC by relating parameters of the functional from with physiographic and climatological characteristics of watersheds.

4.3.4 Statistical Characterization of FDCs

Statistical characterization of FDCs involves describing the FDC in the form of a probability distribution function. Leboutillier and Waylen (1993) conducted a streamflow regionalization study in British Columbia by fitting a two-component, two-parameter lognormal mixture distribution to the FDC, resulting in a five parameter FDC. The values of each of the five parameters were clustered into seven regional clusters using a two-stage density linkage cluster analysis. The generated clusters showed distinct contiguous regions within the province of British Columbia. Averaged parameter values were then determined and representative FDCs were generated for each of the identified regions. Predictive capabilities of this method, however, have not been investigated in the literature (Jenkinson, 2010).

4.3.5 Graphical Characterization of FDCs

This technique was proposed by Smakhtin et al. (1997) and was further developed in Castellarin et al. (2004), as a “graphical” FDC transposition method. In this method, FDCs from gauged sites were normalized using an index flow and the regional FDC was determined by averaging percentiles of the normalized FDCs within the region. The index flow values at the ungauged sites were estimated using a linear regression technique. Shu and Ouarda (2012) stated that the distinctive characteristic of this technique was that the method made no assumptions about the shape of the FDC as is often done in parametric and statistical distribution based methods. The entire FDC at the ungauged site was derived from observed FDCs at other locations. This technique is advantageous if the entire FDC is required at the site of interest, or if a specific region of the FDC is required that cannot easily be represented by other methods. Along similar lines, Mohamoud (2008) estimated 15 percentiles of the FDC by employing step-wise regression and grouping these percentiles into low, median, and high flow ranges, with five percentiles in each, and determining unique predictors for each of the three ranges. The selection of source sites was based on pre-defined landscape classifications. Shu and Ouarda (2012) expanded these techniques, proposed by Smakhtin et al. (1997) and Mohamoud (2008), and considered 17 different percentiles to represent the FDC. A separate regression relationship was developed for each of the 17 percentiles by employing a step-wise regression analysis and using climatic and watershed characteristics. FDCs at ungauged locations were estimated using various distance weighting schemes and employing area, positional, physiographic and climatic data from multiple sites. Logarithmic interpolation technique was used for obtaining values lying in-between any two predicted percentiles, where required. The authors found that the FDC technique outperformed the simple area ratio method and that the inclusion of multiple source sites consistently improved predictive capability of the method.

4.3.6 Other REMs

A few other methods have also been proposed for estimating FDCs at ungauged locations and these methods were found to be similar to the above mentioned REMs. For example, non-linear spatial interpolation technique of Hughes and Smakhtin (1996). These authors developed a nonlinear technique for infilling missing data at proximal gauges, and generating continuous streamflow time series using the FDC as a transfer function. Although this technique was developed primarily to fill-in missing data, the procedure resembles that of FDC transposition at ungauged locations. Monthly FDCs using daily streamflow data for each calendar month for both the target and source sites were employed for constructing streamflow time series on a monthly basis. Smakhtin (1999) and Smakhtin and Masse (2000) also developed a technique for ungauged locations where the FDC at the target site was unknown. These authors suggested to normalize FDCs using an index flow, and then determining the target FDC, along with the index flow at the target location. The FDC at the source site was represented as a discharge table for fixed percentage points and then data between points was interpolated using a logarithmic interpolation technique. The source sites were weighted based on similarity of the source sites to the target location. The authors recommended that up to five sites could be used as source sites. The normalized FDCs for ungauged locations were then de-normalized with the respective index flow determined through regional regression analyses. The authors also suggested that one should avoid direct use of drainage area and should prefer the use of mean annual flow in the regression analyses. They also suggested using 20 to 25 years of data for applying this method. Metcalfe et al. (2005) reviewed this method favourably for generating streamflow regimes at ungauged locations in Ontario.

Inspired by Smakhtin (1999) and Smakhtin and Masse (2000), Shu and Ouarda (2012) suggested Regression Based Logarithmic Interpolation (RBLI) technique for generating FDCs at ungauged locations. They transposed 17 different percentiles of the FDC from source sites to ungauged locations using multiple regression based on climatic and physiographical attributes, but without normalizing the FDC as was the case in Smakhtin (1999) and Smakhtin and Masse (2000). Like the studies of Smakhtin (1999) and Smakhtin and Masse (2000), in-between percentiles of the FDCs were obtained through a logarithmic interpolation technique, when generating continuous time series of streamflow at ungauged locations. The authors evaluated the effect of considering single and multiple source sites in the RBLI approach and variants of the area ratio method for transposition of FDCs using data from Quebec. The RBLI method performed better than the area ratio method and multiple source sites option was found to show substantial improvement over the single source site option in most cases. For the case of multiple source sites, geographic distance based weighting scheme was found to perform better compared to the weighting scheme based on physiographic attributes.

4.4 Concluding Remarks

Compared to several regionalization studies which are available in the literature on flood flows and precipitation extremes, regionalization of FDCs is not very common. Though the underlying regionalization principles are about the same, the number of studies available in the literature remains quite small. This could be due to the complexity that different parts of the FDC are governed by different streamflow generating mechanisms. For example, spring high flows under Canadian conditions are generally associated with snowmelt and/or rain-on-snow events, while winter low flows occur due to frozen conditions compared to summer low flows, which are associated with lack of precipitation and high evaporation demands. In spite of this complexity, some investigators have tried to model FDCs and have developed methods for their transposition on ungauged locations. From the perspectives presented above in this chapter on the estimation of FDCs at ungauged sites, several observations can be made and those are summarized below.

- The regionalization framework for developing FDCs at ungauged locations is very similar to that used for high flows or low flows, i.e. identification of a homogeneous region or a neighbourhood first and then developing methods for the estimation of FDCs at ungauged locations. In the studies reported in this chapter, homogeneous regions were developed within a geographic space and neighbourhoods were identified within the space of watershed attributes. After implementing these steps, some studies have utilized a normalized regional FDC, while others have utilized separate regression relationships between selected percentiles of FDCs and watershed attributes for developing FDCs at ungauged locations. In the former case, regression relationships were also developed between the normalizing index flow (e.g. mean annual flow) and watershed attributes.
- The area ratio method embodies a simple and quick approach for estimating not only continuous streamflow time series, but also the complete FDC at an ungauged location. Though simple, this method has certain limitations. For example, this method is specifically suitable for ungauged locations within the same watershed and is not so when applied across different watersheds in a larger geographic region. In spite of such limitations, the area ratio method has been used for developing complete FDCs in a few previous studies (e.g. NRC-CHC 2008; Mohamoud, 2008; Shu and Ouarda, 2012). When applied within a regionalization context, a weighting scheme based on similarity measures in terms of Euclidean distance can also be integrated with this method to improve accuracy of predicted FDCs.
- Transposition of graphical FDCs method was developed by Hughes and Smakhtin (1996) for generating streamflow sequences at ungauged locations. The same method was adapted by Metcalfe (2005) for transposition of FDCs and was developed further in Shu and Ouarda (2012) based on the neighbourhood concept by defining nearest-neighbours within the geographic space. CCA-based regionalization was also used to predict 17 percentiles of FDCs at ungauged locations in a study by Bomhof (2013).
- When transposing FDCs from donor sites to ungauged sites within the regionalization framework, information on various watershed attributes as predictors of FDCs is required for

both gauged and ungauged sites. These watershed attributes can be derived from several datasets including climate data, digital elevation data, soil and land use maps, geological data, etc. The attributes that can be derived from these datasets and used in developing regression relationships for generating information on ungauged sites are described in Chapter 5.

- For developing FDCs at ungauged sites within the six low flow regions of Ontario, some of the reviewed approaches can be evaluated. For example, those approaches which utilize direct regression relationships of selected percentiles of FDCs and watershed attributes can be quite useful. However, it will be beneficial to implement these methods by identifying smaller neighbourhoods within the individual low flow regions of Ontario or even using sites from other regions, where necessary. For the identification of neighbourhoods, both the CCA-based approach and the ROI approach can be used for Ontario as the results from both will be complementary. The performance of these combined approaches can be compared with the simple area ratio method to establish baseline benchmarks.

5 Future Considerations and Research Avenues for Regional Analysis of Low Flows and Flow Duration Curves in Ontario

5.1 General

For the estimation of low flow indices in a given region, availability of long-term observational records play a critical role. The low flow indices that are often derived from observations and considered for low flow assessments in riverine environments are selected quantiles from low flow frequency curves (e.g. 7Q10) and/or selected percentiles (e.g. Q95) from flow duration curves (FDCs). Here, the term "low flow indices" is used in a generalized context and it collectively refers to both low flow quantiles and lower percentiles of FDCs. Though the use of low flow quantiles is more common, percentiles of FDCs are often used to complement these quantiles, but are commonly used in European countries to characterise low flow conditions in riverine environments. It is well known that the majority of the Canadian stream network is ungauged and recorded observations are available more frequently in southern parts of the country and much less so for northern regions. Ontario's stream network is also not heavily gauged due to continuous suspension of flow monitoring stations overtime. In the absence of recorded observations, low flow indices cannot be derived at ungauged locations and therefore these indices are obtained through indirect means, e.g. by transposing known or processed information from gauged to ungauged locations following an established methodology, e.g. some forms of the RFA approach (cf. Chapter 3) or transposition of functional relationships of low flow indices and watershed attributes (cf. Chapter 3).

Numerous techniques abound in the literature for the estimation of low flow quantiles and percentiles of FDCs at ungauged locations through direct scaling procedures or using regression-based functional relationships. Some of these approaches have already been discussed in Chapters 3 and 4 of this report. For the case of FDCs, these techniques include drainage area ratio methods (e.g. Mohamoud, 2008), parametric characterization of FDCs (e.g. Castellarin et al., 2004), graphical characterization of FDCs (e.g. Castellarin et al., 2004; Smakhtin and Masse, 2000; and Shu and Ouarda, 2012), and various variants of the regression framework, developed mainly on the basis of hydrologic homogeneous regions or groups of watersheds with similar attributes of interest. For the case of low flow quantiles, the regression framework has mostly been employed. An important feature of most of these techniques is that one seeks regression relationships between low flow indices (e.g. low flow quantiles) and watershed attributes (e.g. drainage area, mean annual precipitation, etc.) from gauged locations and then transposes those relationships to target ungauged locations with known attributes. However, recent research has shown that if a set of nearest-neighbours (i.e. gauged sites with watershed attributes similar to those of the target ungauged site within the geographic space or within the attribute space) can be identified then the reliability of estimated low flow indices at ungauged sites can be improved

(e.g. Burn, 1990a, 1990b; Zrinjti and Burn, 1994; Tasker et al., 1996; Ouarda et al., 2000, 2001; Cavadias et al., 2001; Eng et al., 2005; Shu and Ouarda, 2012). It is important to note that there is no consensus on the use of geographic space or attribute space for defining hydrologic similarity and therefore it remains an open question in statistical hydrology. Furthermore, lack of available gauged or so called donor sites within a given geographic region remains a serious challenge in achieving hydrologic similarity solely within the geographic space (Khaliq et al., 2015).

For Ontario, regionalization of low flow characteristics was conducted in early 1990s (MOEE, 1995). In this study, the following five predictive models were considered: (i) Mapped Isolines, (ii) Graphical Index Method, (iii) Statistical Index Method, (iv) Regression Method, and (v) Proration Method. Of these five methods, Mapped Isolines approach was found to perform better in relative terms than the other four methods. Later in early 2000s, an automated implementation of the first four methods was incorporated in the Ontario Flow Assessment Tool (OFAT). This tool was developed by the Ontario Ministry of Natural Resources and Forestry (Chang et al., 2002) and was used internally at the time within the Ontario Government by various water resources practitioners (personal communication with MECP, October 2021). In late 2000s, OFAT was made public over the internet, along with the most recent digital elevation model of the province for delineating watersheds and calculation of watershed attributes in an automated fashion. Of the above mentioned five methods, only Graphical Index and Regression Methods for low flow predictions were included in the publicly accessible version of OFAT (https://www.lioapplications.lrc.gov.on.ca/OFAT/index.html?viewer=OFAT.OFAT&locale=en-ca). The outcomes of this report would be helpful in improving these earlier methods and their implementation in OFAT in the future.

For the regional analysis of low flow indices, some homogeneous regions were identified within the province in MOEE (1995). Since the completion of this study, several more years of data is now available and several new insights on regionalization of low flows have emerged overtime. It will be beneficial from both scientific and application viewpoints to re-evaluate the homogeneity of those regions using information from longer samples and new guidance from the literature. With the additional data included, it is likely to have some sites falling in more than one homogeneous region. In that case, the approach based on partial membership concept can be explored (e.g., see Srinivas et al., 2008; Satyanarayana and Srinivas, 2011; Asong et al., 2014). Additionally, the steps involved in the RFA approach of Hosking and Wallis (1997), documented in Chapter 3, can be followed. For estimating low flow quantiles at ungauged locations within the RFA setting, it will be necessary to develop regression relationships between index flows and watershed attributes. Some guidelines on improving the quality and reliability of these relationships is provided below in this chapter in order to set the stage for future research on regionalization of low flows in Ontario's rivers and streams. In addition to the established RFA approach, it will also be useful to explore the idea of nearest-neighbours to form non-contiguous homogenous regions, and estimate target low flow indices at ungauged sites for comparison

purposes. To the author's knowledge, none of the existing studies have assessed low flows along these lines for Ontario. Some information on conducting such analyses is also provided below to instigate new studies. Additionally, some guidance on the development of non-linear regression relationships between low flow indices and watershed attributes using rapidly emerging Machine Learning (ML) approaches is also discussed.

5.2 Perspectives on Watershed Attributes and Regression Relationships

For delineating homogeneous regions or neighbourhoods for regionalization of low flow indices and developing functional relationships between low flow indices and watershed attributes for transposing known information from gauged to ungauged locations, one should consider several attributes reflecting the influence of climate, landscape features, geologic formations and soil characteristics on low flow occurrences. These characteristics can be derived from digital elevation data, soil characteristics data, land cover data, surficial geology data, and climatic/meteorological data. The list of attributes that can be derived from these datasets is provided below and the definitions of these attributes are provided in Table 5.1. Both pieces of information should be read simultaneously.

- From the Canadian or Ontario's digital elevation data, it is possible to derive and experiment with MinElev, MaxElev, MeanElev, MedElev, and StdDevElev.
- Regarding soil characteristics, one could consider DrainageIndex, KSAT, and KP0.
- From the land use maps, some dominant land use types can be derived, e.g. portion of the watershed covered by Lakes, Barren, Developed, Shrublands, Wetlands, Grasslands, Croplands, Forest, etc.
- From the Canadian surficial geology data, one could consider GeoGlaciers, GeoLakeMud, GeoLakeSand, GeoMud, GeoPeat, GeoRock, GeoSandGravel, GeoTill, GeoWater, etc.
- Climatic indicators in the form of AnnPrecRain, AnnPrecSnow, AnnPrecTotal, MinAnnTemp, MaxAnnTemp, MeanAnnTemp, PE, GSS, GSE, GSL, GDD0, GDD5, GDD10, and GDD15 could be explored to uncover linkages between climatic attributes and low flow regimes of Ontario's rivers and streams.

Table 5.1: Watershed attributes for finding nearest-neighbours and developing regression relationships of low flow indices.

Attribute	Description	Attribute	Description
DrainageArea	Drainage area [km^2]	MeanAnnTemp	Mean annual temperature [°C]
Perimeter	Watershed perimeter [m]	PE	Potential evapotranspiration [mm]
CentroidLat	Latitude of the centroid of the watershed [°N]	GSS	Growing season start date [Julian day]
CentroidLong	Longitude of the centroid of the watershed [°E]	GSE	Growing season end date [Julian day]
MinElev	Minimum elevation of the watershed [m]	GSL	Length of growing season [days]

MaxElev	Maximum elevation of the watershed [m]	GDD0	Growing degree days above 0°C [.]
MeanElev	Mean elevation of the watershed [m]	GDD5	Growing degree days above 5°C [GDD]
MedElev	Median elevation [m]	GDD10	Growing degree days above 10°C [GDD]
StdDevElev	Standard deviation of elevation [m]	GDD15	Growing degree days above 15°C [GDD]
Lakes	Proportion of watershed containing lakes [.]	DrainageIndex	Scaled drainage index [categorical]
Barren	Proportion barren land [.]	KSAT	Saturated hydraulic conductivity [m/day]
Developed	Proportion developed land [.]	KP0	Soil permeability
Shrublands	Proportion shrublands [.]	GeoGlaciers	Geological class glaciers [.]
Wetlands	Proportion wetlands [.]	GeoLakeMud	Geological class lake mud [.]
Grasslands	Proportion grasslands [.]	GeoLakeSand	Geological class lake sand [.]
Croplands	Proportion crop lands [.]	GeoMud	Mud type geological classes [.]
Forest	Proportion of watershed with forests [.]	GeoPeat	Geological class peat [.]
AnnPrecRain	Amount of mean annual rain [mm]	GeoRock	Rock type geological classes [.]
AnnPrecSnow	Amount of mean annual snow [mm]	GeoSandGravel	Combined sand and gravel classes [.]
AnnPrecTotal	Total mean annual precipitation [mm]	GeoTill	Geological class thick and continuous till [.]
MinAnnTemp	Minimum annual temperature [°C]	GeoWater	Geological class water [.]
MaxAnnTemp	Maximum annual temperature [°C]		

These attributes can be employed for identifying hydrological neighbourhoods based on the canonical correlation (CCA) or region of influence (ROI) approach and subsequently developing regression relationships for estimating various indices of low flows at all ungauged streams in Ontario. It is important to note that according to the CCA/ROI approach each target stream reach is associated with its own neighbourhood or a group of nearest-neighbours. Thus, it is reasonable to expect that the size as well as the formation of the neighbourhoods will vary spatially from one location to the next within a larger area of interest. These approaches can provide parallel estimates of low flow quantiles and indices of FDCs at ungauged locations, in addition to the ones obtained through RFA for already identified fixed low flow regions of Ontario. For developing robust regression relationships it is necessary to screen these attributes for various reasons, e.g. reducing the influence of multicollinearity and outliers on the estimated regression parameters.

From statistical and practical viewpoints, having a larger set of watershed attributes does not necessarily guarantee that the resulting regression or information transposition relationships or groups of nearest-neighbours will bear higher degree of reliability because it is very likely that many of the attributes could be mutually correlated. Therefore, all attributes need to be screened individually through pair-wise correlation plots and/or on the basis of Variance Inflation Factors (VIFs) (Eng et al., 2005; Fox, 2008). To have statistically meaningful attributes, it is important to

consider just one attribute from a pair of strongly correlated attributes in order to avoid the influence of multicollinearity, which may lead to irrational and hard to explain regression coefficients. For example, in a similar investigation, Khaliq et al. (2015) found that the watershed drainage area and watershed perimeter for Canadian watersheds are highly correlated, with the correlation value approximated at 0.98. Thus, it is illogical to consider both the watershed drainage area and perimeter in a regression relationship. In the literature on ungauged hydrology, drainage area was commonly used as an independent predictor in regional analyses and therefore drainage area should be preferred over watershed perimeter.

The *VIF* approach can expedite the process of selecting independent attributes from a given set of attributes compared to the pair-wise correlation analysis. Following Eng et al. (2005), the *VIF* is defined as:

$$VIF = \frac{1}{1-R_{VIF}^2} \tag{5.1}$$

where R_{VIF}^2 is the coefficient of determination obtained when the predictor variable of interest (i.e. a selected attribute) is regressed on the remaining predictor variables. A high correlation among the predictor variable of interest and the other variables will result in a large value of R_{VIF}^2 and that, in turn, will lead to a large value of *VIF*, and vice versa for low correlations. According to Montgomery et al. (2001), a value of the *VIF* greater than 10 would be indicative of significant multicollinearity. For stringent requirements, a smaller cut-off threshold can also be used. However, it is advisable to seek physical justifications before eliminating any candidate attribute.

After screening all attributes and identifying a potential set of candidate attributes, it is reasonable to develop functional relationships for each of the selected low flow quantiles or percentiles of FDCs. For developing these relationships, it is important to avoid observational records with zero flow values. When finalizing these relationships, it is also important to retain only those attributes that are statistically significant at a chosen significance level, which is commonly taken as 5%. Inclusion or exclusion of an attribute in the regression relationship can be guided partially using the step-wise regression technique. An analysis of standardized partial correlation coefficients (McCuen, 2003) can also shed additional light on the importance of certain attributes.

5.3 Perspectives on the Identification of Neighbourhoods or Nearest-Neighbours

Many approaches are available in the literature for identifying neighbourhoods or nearest-neighbours for estimating target variables of interest (e.g. low flow indices) at an ungauged location. In the past, a CCA-based approach has been used in Canada for flood frequency analysis (Ouarda et al., 2001), estimation of mean monthly flows (Khaliq et al., 2015) and

percentiles of FDCs (e.g. Bomhof, 2013). Another comparable approach is based on the concept ROI and that has also been used in many studies world-wide (e.g. Burn, 1990a, 1990b; Eng et al, 2005). There are many conceptual similarities between the two approaches regarding identification of nearest-neighbours. Here, the CCA approach is discussed in a greater detail compared to the ROI approach.

The CCA approach finds groups of similar watersheds by correlating a group of low flow indices (i.e. one set of, so called, dependent variables) with watershed attributes (i.e., a second set of variables, obtained through data screening procedures). More specifically, the CCA approach simplifies such a multidimensional dataset so that all of the original variables are represented by new canonical variables, which are made from linear combinations of the original normalized variables such that the correlation of the canonical variables is maximized. If the correlation between the canonical variables is high then it is assumed that one set of variables will be useful for estimating the other set of variables and vice versa (Cavadias et al., 2001). For low flow analysis, one set of variables could be selected percentiles of FDCs or low flow quantiles, and the other set could be all watershed attributes. In order to find nearest-neighbours for a target ungauged site, the location of the site is determined in the canonical space based on site's attributes, and nearest-neighbours are identified using the Mahalanobis distance measure and an extreme upper quantile of the chi-squared distribution (taken as a cut-off value) corresponding to a selected exceedance probability 'alpha'. Smaller (larger) values of alpha would lead to more (less) nearest-neighbours in the neighbourhood of a target ungauged site. In practice, there is a trade-off between achieving a higher degree of similarity, with having only a small number of neighbours in the neighbourhood, and the desired robustness of the relationships derived on the basis of those neighbours. The value of the parameter alpha can be optimized through a cross validation approach by evaluating a set of assumed alpha values. In the cross validation approach, a site from the group of available sites is systematically removed and the CCA approach is applied to the remaining sites to find nearest-neighbours for the removed site. The neighbours found so are used to develop regression relationship to estimate the target low flow index at the removed station. The estimated low flow indices at all sites are assessed by calculating an assessment criterion (e.g. root mean square error). Graphical plots of the chosen assessment criterion against the range of alpha values can be used to select a suitable value of the parameter alpha for each of the target low flow indices (e.g. selected percentiles of the FDC or other low flow indices of interest).

Following the attribute screening procedure and development of a baseline relationship between target indices and watershed attributes, it is possible to identify some attributes that are common across all selected percentiles of FDCs. For example, one could decide to retain only those attributes which are found significant for at least (say) five out of (say) 10 selected percentiles of FDCs. This arbitrary criterion can help in reducing the undue noise due to relatively less influential attributes, with the aim to achieve a higher degree of similarity within the neighbourhoods. Since not all selected percentiles of the FDCs will have the same number of

significant variables in the regression relationships, the above approach is also helpful in overcoming such non-uniform scenarios.

The ROI approach works on a somewhat similar concept. In this case, the nearest neighbours are identified based on the Euclidean distance measure within the attribute space. Based on this measure various sites are ranked from the closets to the farthest. In practice, a certain number of closest sites are identified and used in developing regression relationships for estimating unknown low flow indices at the target ungauged site.

5.4 Perspectives on the Estimation of Low Flow Indices at Ungauged Locations

Following the screening of watershed attributes and identification of neighbourhoods or nearest-neighbours, various indices of low flows are estimated through developing regression relationships. These relationships could be developed in a linear or nonlinear manner. When developing these relationships, it is important to maximize the contributions of nearest-neighbours in these relationships. Below, some guidance on these aspects is provided under different headings.

5.4.1 Weighting of Nearest-Neighbours

In the neighbourhood of a target ungauged site, it is very likely that some neighbours are relatively more similar to the target site than others in the attribute space and therefore it is reasonable to adopt a weighting scheme such that more similar neighbours will receive higher weights than less similar neighbours in the estimation of low flow indices. In the case of CCA-assisted neighbourhoods, Mahalanobis distance measure of each neighbour from the target site can be used to weight various neighbours for developing functional relationships. The following weighting function, which has some similarity to the one used in some earlier studies (e.g. Burn 1990b) can be used:

$$w_i = 1, \text{if } d_i \le d_L \text{ else } w_i = 1 - \left(\frac{d_i - d_L}{d_U - d_L}\right)^{\eta} \tag{5.2}$$

where w_i is the weight and d_i is the Mahalanobis distance measure of the *i*th neighbour in the neighbourhood of a target site; d_L and d_U are respectively the lower and upper thresholds and η is called the weighting exponent. It is possible to select d_L from a smaller group of percentiles (e.g. 5th, 10th, 20th and 25th) of the Mahalanobis distance measure and d_U is generally taken as the maximum value of the distribution of Mahalanobis distance measure, which in turn depends on the value of a parameter (say alpha) that controls the size of the neighbourhood. A smaller (larger) value of alpha would lead to more (less) nearest-neighbours in the neighbourhood of a target ungauged site. This weighting function ensures higher weights to be assigned to the closest neighbours and lower weights to the distant neighbours, with rapidly decaying values for smaller values of η (e.g. 0.05). For real world applications, all weights need to be normalized. For the

case of ROI approach, the parameter *d* can be replaced with a measure of the Euclidean distance within the space of watershed attributes (see Burns, 1990b).

5.4.2 Transformation of Attributes

In addition to the above considerations, it is also important to investigate which transformation of a given attribute is relatively more suitable for the overall relationship. For example, the drainage area and the low flow indices could be highly correlated with each other in the logarithmic domain. Thus, it will be advisable to regress logarithmically transformed low flow indices on to the logarithmically transformed drainage areas. In addition to the logarithmic transformation, square root, cube root, or other suitable transformations can be used. A simple way of identifying a suitable transformation is to regress logarithmically transformed low flow indices against the selected attributes, separately for each of the selected transformations and select the one that produces the highest value of the coefficient of determination. Some investigators prefer to use logarithmic transformations of all attributes due to convenience reasons. Some guidance on this aspect is also available in Engeland (2006).

5.4.3 Other Considerations

When developing and finalizing regression relationships, it is important to examine regression diagnostics in order to verify if the underlying theoretical assumptions are satisfied. For example, normality of residuals, homogeneity of variance, independence of residuals, absence of outliers, uncorrelated predictors, etc. From a practical standpoint, it used to be a difficult task to do so, but that is not the case anymore. Almost all statistical packages (such as Matlab, SAS, R platform, Minitab, etc.) provide ready to use tools to produce these diagnostics. In some situations when the values of certain predictors are close to zero, it is likely that coefficients of these predictors may end up being equal to undefined flags (e.g. NA in R). Such attributes need to be explicitly removed from the regression relationships. In some software packages, this issue is also flagged as ‘rank deficient problem’. It is also a good practice to check finiteness of the p-value of the regression model (Walpole et al., 2011). When analysing large amounts of data in an automated fashion, it is possible that such problems can go unnoticed.

In modelling low flow indices as a function of watershed attributes, it is also useful to look at standardized regression coefficients. This is important because all watershed attributes do not share the same scale, i.e. they are at different scales. Irrespective of the original scale, a standardized regression coefficient of 1 for a given attribute means that an increase in its value of 1 standard deviation will produce a corresponding 1 standard deviation increase in the dependent variable. Consequently, if an attribute A has a larger absolute value of the standardized regression coefficient than the attribute B, then the attribute A has a stronger relationship with the dependent variable (i.e. a low flow index). For any attribute, the standardized regression coefficient is equal to ‘the product of its coefficient and standard deviation divided by the standard deviation of the dependent variable’. Absolute values of all standardized regression

coefficients should be ≤ 1. Some investigators express this assessment in terms of model rationality (e.g. McCuen, 2003).

When developing regression relationships, it is likely that a reduced model consisting of most important attributes need to be favoured in situations where the condition on the required minimum number of independent pairs of data is not satisfied. This situation can be encountered in the neighbourhood based regression relationships. Regarding the choice of the reduced model, the above analysis of standardized regression coefficients can be very insightful. In certain situations, drainage area alone can be the most favourable choice.

5.5 Final Remarks

In addition to the statistical aspects discussed above in this chapter to improve regression relationships, it is also important to carefully inspect quality and reliability of watershed attributes before using them in regression relationships. In general, the reliability of any model, including regression relationships, depends on the quality and accuracy of the underlying data used for calibrating the target model. When one desires to estimate various low flow indices at ungauged locations by identifying neighbourhoods within the attribute space, each ungauged location is expected to have its own neighbourhood. Therefore, even within the same larger homogeneous hydrologic region, there could be a smaller group of donor sites, from where the known information is drawn, that is relatively more similar to the ungauged location in question compared to the rest of the sites of the region. Consequently, the notion of spatially varying neighbourhoods is quite appealing than just using a constant neighbourhood for all ungauged locations within the geographic boundaries of a larger hydrologic region.

For multiple regression relationships, a number of factors play a crucial role in developing reliable and theoretically defensible relationships such as: (1) relevance of predictor variables and their selection procedures; (2) inter-dependence of predictor variables; (3) how the predictor variables are introduced in the regression relationships (e.g. in their original form or using log-transformation or square root transformation, etc.); and (4) tests of diagnostics. Guidelines on these aspects are readily available in many text books on applied statistics (e.g. Montgomery, 2001; Helsel and Hirsch, 2002; McCuen, 2003; and Walpole et al., 2011) and therefore these aspects should not be ignored. Another important aspect that is often neglected when developing multiple regression relationships concerns the number of available independent data pairs and the number of unknown regression parameters. Ideally, the former should be considerably larger than the latter to develop sound relationships. McCuen (2003) recommended to have independent number of data pairs more than four times the number of unknown regression parameters. This aspect is very important for neighbourhood-based regression relationships, since it is very unlikely to find sufficient number of nearest-neighbours at least for certain ungauged locations. In those situations, it is better to try a relationship with a smaller number of attributes. For example, by selecting the three most important attributes that are able to explain

the majority of the variability of the dependent variable than using all available attributes. This will avoid having some locations with indeterminate estimates. However, the reader is reminded that this strategy may not work under all situations.

In the literature on hydrologic regionalization, several different methods are available and these can be used to delineate neighbourhoods or nearest-neighbours when estimating unknown low flow indices at ungauged locations from the corresponding known indices available at gauged locations. The ROI approach is quite common in the international literature, while the CCA-based approach has also been used in some national studies to identify neighbourhoods. The principles that underpin these two approaches allow consideration of gauged locations from adjoining or distantly located hydrologic regions or geographic areas. Thus, geographic or political boundaries are not considered a limitation for applying these approaches. However, some investigators do object to such definition of neighbourhoods due to considerable differences in associated atmospheric mechanisms that influence regional climate and local weather patterns. In order to reconcile both school of thoughts, perhaps it is useful to consider larger climatic or hydrologic regions and apply the ROI or the CCA approach to identify nearest-neighbours within the same larger region. Such an approach is advantageous from climatological, hydrological and statistical viewpoints and also ensures to some extent the physical proximity of the target location and nearest-neighbours. Canada has been divided into 11 large climatic regions (e.g. Plummer et al., 2006; Mladjic et al., 2011) and those regions can be used as a basis to develop both ROI and CCA-based regionalization approaches. Similar concepts can also be applied across Ontario. For certain situations, the results from both ROI and CCA-based approaches could be very different and therefore, it will be useful to apply both approaches together within the same climatic region. This will help in reaping the benefits of both approaches and ultimately to have better estimates of low flow indices at ungauged locations within a target region of interest.

In order to improve quality of estimated low flow indices at ungauged locations, it is important to improve quality of various physiographic and climatic attributes that play a fundamental role in the estimation of these quantities at ungauged locations. As discussed in Chapter 3, applications of ML approaches to solve applied problems are becoming increasingly popular. Therefore, it will be interesting to apply these approaches in the regionalization of low flows across Ontario by developing nonlinear regression relationships between low flow indices and watershed attributes. It is expected that these approaches will provide at least as good estimates as those obtained through conventional multiple regression based techniques.

Apart from low flow frequency analyses and evaluating percentiles of FDCs, it will be beneficial to explore the behaviour of low flow spells in order to advance our understanding of low flow characteristics of Ontario streams. To investigate how long streamflow will be below a certain flow level, how large the deficit volume is, and how intense the extreme deficit is, statistical analysis of low flow spells is conducted using, e.g., theory of runs or joint frequency analyses

within a multivariate framework since these characteristics are inter-dependent. Low flow spells can be found by assuming environmental instream minimum flows (e.g. Q95 or a specified percentage of the mean annual flow) as a threshold and processing entire time series of daily flows. All periods below these thresholds are considered as low flow spells, which can be characterized in terms of spell duration (in days), deficit volume (in cubic meters), peak intensity (in m^3/sec), and intensity of deficit (deficit volume divided by duration). According to some investigators, this approach provides rather a more complete picture of the low flow regime of a stream compared to the approaches based on frequency analysis of fixed duration low flow events and percentiles of FDCs.

Lastly, this chapter was specifically devoted to inspire new studies on low flow analysis from a regionalization perspective. Several suggestions have been made to improve estimates of low flow indices at ungauged locations based on the perspectives collected from the literature on ungauged hydrology and regional frequency analysis approaches. It is hoped that the future studies and analyses that can be initiated based on the information provided in this report will complement and advance our existing understanding about the low flow regimes of Ontario's rivers and streams and will therefore enable better strategic decision-making for agriculture, water management, health, environment, and several other water-sensitive sectors.

6 References

Acreman MC, Sinclair CD, 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. Journal of Hydrology 84(3–4): 365–380.

Acreman M, Wiltshire S, 1989. The regions are dead. Long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. IAHS-AISH publication 187:175–188.

Acres Consulting Services Limited, 1984a. Hydrologic design methodologies for small scale hydro at ungauged sites – phase I: Applications manual. Technical report, Environment Canada, Inland Waters Directorate.

Acres Consulting Services Limited, 1984b. Hydrologic design methodologies for small scale hydro at ungauged sites – phase I: Study documentation report. Technical report, Environment Canada, Inland Waters Directorate.

Alam KMR, Siddique N, Adeli H, 2020. A dynamic ensemble learning algorithm for neural networks. Neural Computing and Applications 32: 8675–8690.

Alobaidi MH, Ouarda TBMJ, Prashanth RM, Chebana F, 2021. Diversity-driven ANN-based ensemble framework for seasonal low-flow analysis at ungauged sites. Advances in Water Resources 147, 103814.

Aschwanden H, Kan C, 1999. Die Abflussmenge Q347—Eine Standortbestimmung. Hydrologische Mitteilungen 27, Landeshydrologie und -geologie, Bern, Switzerland.

Asong ZE, Khaliq MN, Wheater HS, 2014. Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes. Stochastic Environment Research and Risk Assessment, DOI: 10.1007/s00477-014-0918-z.

Barnes, CR, 1986. Method for estimating low flow statistics for ungauged streams in the Lower Hudson River Basin, New York. Water-Resources Investigations Report 85-4070, US Geological Survey.

Beable ME, McKerchar AI, 1982. *Regional flood estimation in New Zealand.* National Water and Soil Conservation Organization, Water and Soil Division, Ministry of Works and Development, Wellington, New Zealand.

Bobee B, Mathier L, Perron H, Trudel P, Rasmussen PF, Cavadias G, Bernier J, Nguyen V-T-V, Pandey G, Ashkar F, Ouarda TBMJ, Adamowski K, Alila Y, Daviau JL, Gingras D, Liang GC, Rousselle J, Birikundavyi S, Ribeiro-Corra J, Roy R, Pilon PJ, 1996. Presentation and review of some methods for regional flood frequency analysis. Journal of Hydrology 186 (1-4): 63–84.

Bomhof J, 2013. Estimating flow, hydraulic geometry, and hydrokinetic power at ungauged locations in Canada. MSc thesis, University of Ottawa, Ottawa, Ontario.

Bond NR, Kennard MJ, 2017. Prediction of hydrologic characteristics for ungauged catchments to support hydroecological modeling. Water Resources Research, DOI; 10.1002/2017WR021119.

Box GEP, Cox DR, 1964. An analysis of transformations. Journal of the Royal Statistical Society, Series B. 26 (2): 211–252.

Breiman L, Friedman JH, Olshen R, Stone CJ, 1984. Classification and regression trees. Wadsworth, California, USA.

Breiman L, 2001. Random forests. Machine Learning 45: 5–32.

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

Burn DH, 1990a. An appraisal of the "region of influence' approach to flood frequency analysis. Hydrological Sciences Journal 35(2): 149–165.

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

Caruso B S, 2000. Evaluation of low-flow frequency analysis methods. Journal of Hydrology 39 (1): 19–47.

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

Castellarin A, Galeati G, Brandimarte L, Montanari A, Brath A, 2004. Regional flow-duration curves: reliability for ungauged basins. Advances in Water Resources 27(10): 953–965.

Cavadias GS, Ouarda TBMJ, Bobee B, Girard C, 2001. A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins. Hydrological Sciences Journal 46(4): 499–512.

CEH, 1999. Flood Estimation Handbook: 1-5. Center for Ecology and Hydrology Natural Environment Research Council, London, UK.

Chang C, Ashenhurst F, Damaia S, Mann W, 2002. Ontario Flow Assessment Techniques Version 1.0 User’s Manual. Northeast Science and Information Section, Ontario Ministry of Natural Resources. NESI Technical Manual TM-011.

Chen YD, Huang G, Shao Q, Xu C, 2006. Regional analysis of low flow using L-moments for dongjiang basin, south China. Hydrological Sciences Journal 51(6): 1051–1064.

Chowdhury JU, Stedinger JR, Lu L, 1991. Goodness-of-fit tests for regional generalized extreme value flood distributions. Water Resources Research 27(7): 1765–1776.

Coles S, 2001. An Introduction to Statistical Modeling of Extreme Values. Springer Publishers, Great Britain.

Copeland RR, Biedenharn DS, Fischenich JC, 2000. Channel-forming discharge. Technical Report ERDC/CHL CHETN-VIII-5, US Army Corp of Engineers (USACE).

Cunnane C, 1989. Statistical Distributions for Flood Frequency Analysis. WMO- No. 718. Operational Hydrology Report No. 33. World Meteorological Organization, Geneva.

Dodangeh E, Soltani S, Sarhadi A, Shiau J-T, 2014. Application of L-moments and Bayesian inference for low flow regionalization in Sefidroud basin, Iran. Hydrological Processes 28: 1663–1676.

Dalrymple T, 1960. Flood-frequency analyses. U.S. Geological Survey, Washington D.C.

De Coursey DG, 1973. Objective regionalization of peak flow rates in floods and droughts. Proceedings of the Second International Symposium in Hydrology, Fort Collins, Colorado, pp 395-05.

Dingman SL, 1978. Synthesis of flow-duration curves for unregulated streams in New Hampshire. Journal of the American Water Resources Association 14(6): 1481–1502.

Durrans SR, Tomic S, 1996. Regionalization of low-flow frequency estimates: an Alabama case study. Water Resources Bulletin 32(1): 23–37.

Durand N, Bourban SE, Crookshank N, 2002. Development of a Toolkit to Estimate the Concentration of Substances Released into Rivers and Streams - Scientific Literature Review. Technical report, Canadian National Research Council - Canadian Hydraulics Centre.

Eaton B, Church M, Ham D, 2002. Scaling and regionalization of flood flows in British Columbia, Canada. Hydrological Processes 16(16): 3245–3263.

Efron B, Tibshirani RJ, 1993. An introduction to the bootstrap. Monographs on Statistics and Applied Probability, Volume 57. Chapman & Hall, New York, USA.

Eng K, Tanker GD, Milly PCD, 2005. An analysis of region-of-influence methods for flood regionalization in the Gulf-Atlantic rolling plains. Journal of American Water Resources Association, 41: 135–143.

Engeland K, Hisdal H, Beldring S, 2006. A comparison of low flow estimates in ungauged catchments using regional regression and the HBV-model. NVE Report, Norwegian Water Resources and Energy Directorate, Oslo, Norway.

Eris E, Aksoy H, Onoz B, Cetin M, Yuce MI, Selek B, Aksu H, Burgan HI, Esit M, Yildirim I, Karakus EU, 2019. Frequency analysis of low flows in intermittent and non-intermittent rivers from hydrological basins in Turkey. Water Supply, DOI: 10.2166/ws.2018.051.

Friedman JH, 1991. Multivariate adaptive regression splines. The Annals of Statistics 19(1): 1–67.

Fox J, 2008. Applied Regression Analysis and Generalized Linear Models, 2nd Edition, Sage Publications.

Franchini M, Suppo M, 1996. Regional analysis of flow duration curves for a limestone region. Water Resources Management 10(3): 199–218.

Gingras D, Adamowski K, Pilon PJ, 1994. Regional Flood Equations for the Province Of Ontario and Quebec. Journal of the American Water Resources Association 30(1): 55–67.

Govindaraju, RS, Rao AR, 2010. Artificial Neural Networks in Hydrology. Springer Publishing Company, Incorporated.

Green M, Ohlsson M, 2007. Comparison of standard resampling methods for performance estimation of artificial neural network ensembles. Third International Conference on Computational Intelligence in Medicine and Healthcare, pp 25–27.

Greenwood JA, Landwehr JM, Matalas NC, Wallis JR, 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research 15(5): 1049–1054.

Gulliver JS, Murdock RU, 1993. Prediction of river discharge at ungauged sites with analysis of uncertainty. Journal of Water Resources Planning and Management 119(4): 473–487.

Gupta S, Ansari RA, Sarkar D, 2020. Deep Learning with R Cookbook. Packt Publishing Ltd., Birmingham, UK.

Gustard A, Bullock A, Dixon JM, 1992. Low flow estimation in the United Kingdom. Institute of Hydrology, Report No. 108, Wallingford, UK.

Gustard A, Young AR, Rees G, Holmes MGR, 2004. Operational hydrology. In: Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater. LM Tallaksen and HAJ van Lanen (Editors). Developments in Water Science 48, Elsevier, The Netherlands, pp 455–484.

Guttman NB, 1993. The use of L-moments in the determination of regional precipitation climates. Journal of Climate 6(12): 2309–2325.

Grandry M, Gailliez S, Sohier C, Verstraete A, Degre A, 2012. A method for low flow estimation at ungauged sites, case study in Wallonia (Belgium). Hydrol. Earth System Science Discussions 9: 11583–11614.

Haan CT 1977. Statistical Methods in Hydrology. The Iowa University Press, Iowa, USA.

Hayes DC, 1992. Low flow characteristics of streams in Virginia. US Geological Survey, Water Supply Paper No. 2374.

Hortness JE, Berenbrock C, 2001. Estimating Monthly and Annual Streamflow Statistics at Ungaged Sites in Idaho. Water-Resources Investigations Report 01–4093. US Geological Survey. Idaho, USA.

Helsel DR, Hirsch RM, 2002. Statistical Methods in Water Resources. Available at http://water.usgs.gov/pubs/twri/twri4a3.

Henderson RD, Woods RA, Schmidt J, 2005. Low flow estimates for ungauged streams of New Zealand. CDROM and accompanying notes, National Institute of Water and Atmospheric Research, Christchurch, New Zealand.

Hewa GA, Wang Q J, McMahon TA, Nathan RJ, Peel MC, 2007. Generalized Extreme Value distribution fitted by LH moments for low-flow frequency analysis. Water Resources Research 43, W06301.

Holmes MGR, Young AR, Gustard A, Grew R, 2002. A region of influence approach to predicting flow duration curves within ungauged catchments. Hydrology and Earth System Science 6(4): 721–732.

Hosking JRM, Wallis JR, 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3): 339–349.

Hosking JRM, 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society, Series B, 52(1): 105–124.

Hosking JRM, Wallis JR, 1993. Some statistics useful in regional frequency analysis. Water Resources Research 29(2): 21–281.

Hosking JRM, Wallis JR, 1997. Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, Cambridge, New York.

Hughes DA, Smakhtin V, 1996. Daily flow time series patching or extension: A spatial interpolation approach based on flow duration curves. Hydrological Sciences Journal 41(6): 851–871.

IPCC, 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change DOI: 10.1038/446727a.

IPCC, 2013. Climate Change 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, and Midgley PM (eds)]. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jenkinson W, 2010. Assessment of Canada's Hydrokinetic Power Potential. Phase I Report: Methodology and Data Review. NRC Report No. CHC-TR-070. National Research Council of Canada, Ottawa, ON, Canada.

Kachroo RK, Mkhandi SH, Parida BP, 2000. Flood frequency analysis of southern Africa: I. delineation of homogeneous regions. Hydrological Sciences Journal 45(3): 437–447.

Khaliq MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B, 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. Journal of Hydrology 329: 534–552.

Khaliq MN, Ouarda TBMJ, Gachon P, Sushama L, St-Hilaire A. 2009. Identification of hydrological trends in the presence of serial and cross correlations: review of selected methods and their application to annual flow regimes of Canadian rivers. Journal of Hydrology 368: 117–130.

Khaliq MN, Jenkinson W, Bomhof J, Serrer M, Klyszejko E, 2015. Estimation of mean monthly flows at ungauged locations in the Maritimes and Pacific hydrologic regions. Canadian Society of Civil Engineering Conference, Montreal, Quebec, April 29–May 2.

Khaliq MN, 2019. An Inventory of Methods for Estimating Climate Change-Informed Design Water Levels for Floodplain Mapping. NRC Report No. OCRE-2019-TR-011. National Research Council Canada, Ottawa, ON.

Kresser W, Kirnbauer R, Nobilis F, 1985. Überlegungen zur Ermittlung von Niederwasserkenngrößen. Mitteilungsblatt des Hydrographischen Dienstes in Österreich 54: 13–47.

Kroll CN, Vogel RM, 2002. Probability distribution of low streamflow series. Journal of Hydrologic Engineering 7(2): 137–146.

Laaha, G. & Blöschl, G, 2007. A national low flow estimation procedure for Austria. Hydrological Sciences Journal 52(4): 625–644.

Landwehr JM, Matalas NC, Wallis JR, 1979a. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resources Research 15: 1055–1064.

Landwehr JM, Matalas NC, Wallis JR, 1979b. Estimation of parameters and quantiles of Wakeby distributions. Water Resources Research 15: 1361–79.

Leboutillier DW, Waylen PR, 1993. Regional variations in flow-duration curves for rivers in British Columbia, Canada. Physical Geography 14(4): 359–378.

Lesmeister C, 2015. Mastering Machine Learning with R. Packt Publishing Ltd., Birmingham, UK.

Lim YH, Lye LM, 2003. Regional flood estimation for ungauged basins in Sarawak, Malaysia. Hydrological Sciences Journal 48: 79–94.

Liu D, Guo S, Lian Y, Xiong L, Chen X, 2011. Climate informed low-flow frequency analysis using nonstationary modelling. Hydrological Processes 29: 2112–2124.

Loukas A, Quick MC, 1995. Comparison of six extreme flood estimation techniques for ungauged watersheds in coastal British Columbia. Canadian Water Resources Journal 20(1): 17–30.

Mandal U, C Cunnane, 2009. Low-flow prediction for ungauged river catchments in Ireland. Irish National Hydrology Seminar.

Masud MB, Khaliq MN, Wheater HS, 2016a. Projected changes to short- and long-duration precipitation extremes over the Canadian Prairie Provinces. Climate Dynamics 49 (5–6): 1597–1616.

Masud MB, Khaliq MN, Wheater HS, 2016b. Future changes to drought characteristics over the Canadian Prairie Provinces based on NARCCAP multi-RCM ensemble. Climate Dynamics, DOI: 10.1007/s00382-016-3232-2.

Matalas NC, Wallis JR, 1973. Eureka! It fits a Pearson type: 3 distribution. Water Resources Research 9(2): 281–289.

Matalas NC, Slack JR, Wallis JR, 1975. Regional skew in search of a parent. Water Resources Research, 11(6): 815–826.

McCuen RH, 1985. Statistical Methods for Engineers. Prentice-Hall, Englewood Cliffs, N. J.

McCuen RH, Levy BL, 2000. Evaluation of peak discharge transposition. J. Hydrologic Engineering 5(3): 278–289.

McCuen RH, 2003. Modeling Hydrologic Change: Statistical Methods. Lewis Publisher.

Metcalfe RA, Chang C, Smakhtin V, 2005. Tools to support the implementation of environmentally sustainable flow regimes at Ontario's waterpower facilities. Canadian Water Resources Journal 30(2): 97–110.

Millington N, Das S, Simonovic SP, 2011. The Comparison of GEV, Log-Pearson Type 3 and Gumbel Distributions in the Upper Thames River Watershed under Global Climate Models. Available at http://ir.lib.uwo.ca/wrrr/40/.

Minocha VK, 2003. Discussion of 'Probability distribution of low streamflow series' by CN Kroll and RM Vogel. Journal of Hydrologic Engineering 85: 297.

Mladjic B, Sushama L, Khaliq MN, Laprise R, Caya D, Roy R, 2011. Canadian RCM projected changes to extreme precipitation characteristics over Canada. Journal of Climate 24: 2565–2584.

Modarres R, 2008. Regional frequency distribution type of low flow in north of Iran by L-moments. Water Resources Management 22: 823–841.

MOEE, 1995. Regionalization of Low Flow Characteristics for Various Regions in Ontario. Ministry of Environment and Energy (MOEE), Ontario, Canada.

Mohamoud YM, 2008. Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. Hydrological Sciences Journal 53(4): 706–724.

Moin SMA, Shaw MA, 1985. Regional Flood Frequency Analysis for Ontario Streams, Volume 1: Single Station Analysis and Index Method. A study funded under the Canada/Ontario Flood Damage Reduction Program, Environment Canada.

Moin SMA, Shaw MA, 1986. Regional Flood Frequency Analysis for Ontario Streams, Volume 2: Multiple Regression Method. A study funded under the Canada/Ontario Flood Damage Reduction Program, Environment Canada.

Montgomery DC, Peck EA, Vining GG, 2001. Introduction to Linear Regression Analysis (3rd edition). John Wiley and Sons, Inc., New York, New York, pp 641.

Mostofi Zadeh S, Lye LM, Khan AA, 2012. Regional frequency analysis of low flow using L-moments for Labrador, Canada. Proceeding of the 125th CSCE Annual Conference, Edmonton, Canada.

Mudersbach C, Jensen J, 2010. Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline. Journal of Flood Risk Management 3(1): 52–62.

Nash, J.E., Sutcliffe, J.V. 1970. River flow forecasting through conceptual models part I – A discussion of principles. Journal of Hydrology 10(3): 282–290.

Nathan RJ, McMahon TA, 1990. Identification of homogeneous regions for the purpose of regionalization. Journal of Hydrology 121: 217–238.

Natural Resources Canada, 2004a. Clean Energy Project Analysis: RETScreen Engineering & Cases Textbook, Chapter Small Hydro Project Analysis. Natural Resources Canada.

Natural Resources Canada, 2004b. RETScreen Software Online User Manual: Small Hydro Project Model. Technical Report, Natural Resources Canada.

NRC-CHC, 2008. Methodology for the Assessment of Hydraulic Kinetic Energy in Rivers. Technical Report CHC-CTR-075, National Research Council Canada – Canadian Hydraulics Centre.

Ottawa Engineering Limited, 1997. Hydrological Method - RETScreen. Prepared for Natural Resources Canada.

Ouarda TBMJ, Hache M, Bruneau P, Bobée B, 2000. Regional flood peak and volume estimation in Northern Canadian Basin. Journal of Cold Regions Engineering 14(4): 176–191.

Ouarda TBMJ, Girard C, Cavadias GS, Bobée, B, 2001. Regional flood frequency estimation with canonical correlation analysis. Journal of Hydrology 254 (1–4): 157–173.

Ouarda TBMJ, Shu C, 2009. Regional low-flow frequency analysis using single and ensemble artificial neural networks. Water Resources Research 45: 1–16.

Oyebode O, Stretch D, 2018. Neural network modeling of hydrological systems: A review of implementation techniques. Natural Resource Modelling, DOI: 10.1002/nrm.12189.

Pandey MD, Gelder PH, Vrijling JK, 2001. Assessment of an L-kurtosis-based criterion for quantile estimation. Journal of Hydrologic Engineering 64: 284–291.

Parida BP, Kachroo RK, Shrestha DB, 1998. Regional flood frequency analysis of Mahi-Sabarmati Basin Subzone 3a using index flood procedure with L moments. Water Resources Management 12: 1–12.

Patel JA, 2007. Evaluation of low flow estimation techniques for ungauged catchments. Water and Environment Journal 21: 41–46.

Peel MC, Wang QJ, Vogel RM, McMahon TA, 2001. The utility of L-moment ratio diagrams for selecting a regional probability distribution. Hydrological Sciences Journal 46 (1): 147-155.

Pearson CP, 1991a. New Zealand regional flood frequency analysis using L-moments. Journal of Hydrology New Zealand 30: 53–64.

Pearson CP, 1991b. Regional flood frequency for small New Zealand basins, 2: flood frequency group. Journal of Hydrology New Zealand 30: 77–92.

Pearson C P, 1995. Regional frequency analysis of low flows in New Zealand. Journal of Hydrology New Zealand 33(2): 94–122.

Plummer DA and Coauthors, 2006. Climate and climate change over North America as simulated by the Canadian RCM. Journal of Climate 19: 3112–3132.

Quimpo RG, Alejandrino AA, McNally TA, 1983. Regionalized flow duration for Philippines. Journal of Water Resources Planning & Management 109(4): 320–330.

Ribeiro-Correa J, Cavadias GS, Clement B, Rousselle J, 1995. Identification of hydrological neighborhoods using canonical correlation analysis. Journal of Hydrology 173(1–4):71–89

Ries KG, 2002. STREAMSTATS: A US Geological Survey web site for stream information. In: Hydroinformatics 2002. Proceedings of the Fifth International Conference on Hydroinformatics, Cardiff, UK.

Riggs HC, Caffey JE, Orsborn JF, Schaake JC, Singh KP, Wallace JR, 1980. Characteristics of low flows. ASCE Journal of the Hydraulics Division 106: 717–731.

Riggs HC, 1990. Estimating flow characteristics at ungauged sites. In: Regionalisation in Hydrology, Proceedings of the Ljubljan Symposium, April 1990, IAHS Publication No. 191.

Robson A, Reed D, 1999. Flood Estimation Handbook, Volume 3: Statistical Procedures for Flood Frequency Estimation. Institute of Hydrology, Wallingford, UK.

Salas JD, Obeysekera J, 2014. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. Journal of Hydrologic Engineering 19(3): 554–568.

Salas JD, Obeysekera J, Vogel RM, 2018. Techniques for assessing water infrastructure for nonstationary extreme events: a review. Hydrological Sciences Journal 63(3): 325–352.

Salinas JL, Laaha G, Rogger M, Parajka J, Viglione A, Sivapalan M, Blöschl G, 2013. Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies. Hydrology and Earth System Science 17: 2637–2652.

Satyanarayana P, Srinivas VV, 2011. Regionalization of precipitation in data sparse areas using large scale atmospheric variables: a fuzzy clustering approach. Journal of Hydrology 405: 462–473.

Searcy JK, 1959. Flow duration curves. Water Supply Paper 1542-A, US Geological Survey.

Shi P, Chen X, Qu S-M, Zhang Z-C, Ma J-L, 2010. Regional frequency analysis of low flow based on L moments: Case study in Karst area, southwest China. Journal of Hydrologic Engineering 15(5): 370–377.

Shu C, Ouarda TBMJ, 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resources Research 43, W07438.

Shu C, Ouarda TBMJ, 2012. Improved methods for daily streamflow estimates at ungauged sites. Water Resources Research 48(2): 1–15.

Smakhtin VY, 1999. Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique. River Research and Applications 15(4): 311–323.

Smakhtin V U, 2001. Low flow hydrology: A review. Journal of Hydrology 240(3–4): 147–186.

Smakhtin VY, Masse B, 2000. Continuous daily hydrograph simulation using duration curves of a precipitation index. Hydrological Processes 14(6): 1083–1100.

Smakhtin VY, Hughes DA, Creuse-Naudin E, 1997. Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. Hydrological Sciences Journal 42(6): 919–936.

Šraj M, Viglione A, Parajka J, Blöschl G, 2016. The influence of non-stationarity in extreme hydrological events on flood frequency estimation. Journal of Hydrology and Hydromechanics 64(4): 426–437.

Srinivas VV, Tripathi S, Rao AR, Govindaraju RS, 2008. Regional flood frequency analysis by combined self-organizing feature map and fuzzy clustering. Journal of Hydrology 348: 148–166.

Strupczewski WG, Singh VP, Feluch W, 2001a. Non-stationary approach to at-site flood frequency modeling I. Maximum likelihood estimation. Journal of Hydrology 248: 123–142.

Strupczewski WG, Singh VP, Mitosek HT, 2001b. Nonstationary approach to at-site flood frequency modeling III. Flood analysis of Polish rivers. Journal of Hydrology 248: 152–167.

Sushama L, Laprise R, Caya D, Frigon A, Slivitzky M, 2006. Canadian RCM projected climate change signal and its sensitivity to model errors. International Journal of Climatology 26: 2141–2159.

Tallaksen LM, Madsen H, Clausen B, 1997. On the definition and modelling of streamflow drought duration and deficit volume. Hydrological Sciences Journal 42(1): 15–33.

Tallaksen LM, van Lanen HAJ, 2004. Hydrological Drought. Developments in water science, Elsevier, Amsterdam, The Netherlands.

Tan X, Gan TY, 2015. Nonstationary analysis of annual maximum streamflow of Canada. Journal of Climate 28(5): 1788–1805.

Tanty R, Desmukh TS, 2015. Application of Artificial Neural Network in Hydrology- A Review. International Journal of Engineering Research & Technology 4(06): 1–6.

Tasker GD, 1982. Comparing methods of hydrologic regionalization. Water Resources Bulletin 18(6): 965–970.

Tasker GD, 1987. A comparison of methods for estimating low flow characteristics of streams. Advances in Water Resources 23(6): 1077–1083.

Tasker GD, Hodge,SA, Barks CS, 1996. Region of influence regression for estimating the 50-year flood at ungauged sites. Water Resources Bulletin 32(1): 163–170.

Tate EL, Meigh JR, Prudhomme CP, McCartney MP, 2000. Drought Assessment in southern Africa Using River Flow Data. DFID Report 00/4. Institute of Hydrology, Wallingford, Oxfordshire, UK.

Theobald O, 2017. Machine Learning For Absolute Beginners. Packt Publishing Ltd., Birmingham, UK.

Thomas WO, 1987. Techniques used by U.S. Geological Survey in estimating the magnitude and frequency of floods. Proceeding of 18th Binghamton Geomorphology Symposium. Unwin and Hyman, London, pp 267–288.

Tramblay Y, Neppel L, Carreau J, Najib K, 2013. Non-stationary frequency analysis of heavy rainfall events in southern France. Hydrological Sciences Journal 58 (2): 280–294.

Tsakiris G, Nalbantis I, Cavadias G, 2011. Regionalization of low flows based on Canonical Correlation Analysis. Advances in Water Resources 34: 865–872.

Vezza P, Comoglio C, Rosso M, Viglione A, 2010. Low flows regionalization in North-Western Italy. Water Resources Management 24: 4049–4074.

Vogel RM, Fennessey NM, 1993. L-moment diagrams should replace product moment diagrams. Water Resources Research 29(6): 1745.

Walpole RE, Myers RH, Myers SL, Ye K, 2011. Probability & Statistics for Engineers & Scientists, 9th Edition. Prentice Hall, Boston, pp 812.

Wang Y, 2000. Development of Methods for Regional Flood Estimation in the Province of British Columbia. PhD Thesis. Department of Forest Resources. University of British Columbia, British Columbia, Canada, pp 214.

Weisberg S, 1985. Applied Linear Regression, Second Edition. Wiley, New York, USA.

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

Wiltshire SE, 1986a. Regional flood frequency analysis I: Homogeneity statistics. Hydrological Sciences Journal, 31(3): 321–333.

Wiltshire SE, 1986b. Regional flood frequency analysis II: Multivariate classification of drainage basins in Britain. Hydrological Sciences Journal 31(3): 335–346.

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP, 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 75: 1182–1189.

WMO, 2008. Manual on Low Flow: Estimation and Prediction. World Meteorological Organisation, Operational Hydrology Report 50, pp. 136 pp.

Wu Y, Xue L, 2018. Nonstationary modelling of annual discharge over the Tarim River headstream catchment, China. IOP Conference Series: Earth and Environmental Sciences 170, 022149.

Yue S, Wang CY, 2004a. Possible Regional Probability Distribution Type of Canadian Annual Streamflow by L-moments. Water Resources Management 18(5): 425–438.

Yue S, Wang C, 2004b. Determination of regional probability distributions of Canadian flood flows using L-moments.

Yue S, Pilon P, 2005. Probability distribution type of Canadian annual minimum streamflow. Hydrological Sciences Journal, DOI:10.1623/hysj.50.3.427.65021.

Yurekli K, Kurunc A, Gul S, 2005. Frequency analysis of low flow series from Çekerek Stream Basin. Journal of Agricultural Sciences 11(1): 72–77.

Xiong L, Du T, Xu CY, Guo S, Jiang C, Gippel CJ, 2015. Non-stationary annual maximum flood frequency analysis using the Norming Constants method to consider non-stationarity in the annual daily flow series. Water Resources Management, DOI: 10.1007/s11269-015-1019-6.

Zaidman M, Keller V, Young AR, Cadman D, 2003. Flow duration-frequency behavior of British rivers based on annual minima data. Journal of Hydrology 277: 195–213.

Zaier I, Shu C, Ouarda TBMJ, Seidou O, Chebana F, 2010. Estimation of ice thickness on lakes using artificial neural network ensembles. Journal of Hydrology 383: 330–340.

Zrinjti Z, Burn DH, 1994. Flood frequency analysis for ungauged sites using a region of influence approach. Journal of Hydrology 153: 1–21.