Approaches for Nonstationary Frequency and Trend Analyses for Hydrometric Data

by

Nicole Lisa O'Brien

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following are the members who served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Jan F. Adamowski Associate Professor
Supervisor(s)	Donald H. Burn Professor
Internal Member	William K. Annable Associate Professor
Internal Member	Mahesh Pandey Professor
Internal-External member	Keith W. Hipel University Professor

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Statement of Contributions

Chapter 4 was completed in collaboration with Peter J. Thompson of GeoProcess Research Associates and William K. Annable of the Department of Civil and Environmental Engineering, University of Waterloo. Peter was instrumental in providing the event based data and mapping for this research. Prof. Annable helped with the final review and also suggested the use of the event based data.

Abstract

Peak hydrometeorological events have been the topic of much concern in recent years, chiefly their frequency and magnitude. According to the Fifth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC), land-surface temperatures have increased in the 20th and early 21st centuries and more extreme precipitation events are expect in this warmer climate. Furthermore, land-use changes can have a pronounced effect on the hydrologic regime. The subsequent changes in streamflow and precipitation data over time can introduce nonstationarity in the time series data, resulting in data that are not independently and identically distributed (IID). Significant trends, change-points, and serial correlation identified in time series data may invalidate the IID assumption, therefore, in these instances, more complex statistical techniques must be employed.

The first contribution of this thesis focuses on the development of a trend centered pooling approach. The most commonly used pooled frequency analysis techniques do not account for nonstationarity in the pooling group member data, however, nonstationarity must be accounted for if it is detected in the time series data. This approach pools data based on the form of trend found in the at-site data. Therefore all sites are assessed for trend before regionalization and each pooling group is created with members having similar forms of statistically significant trend (either increasing or decreasing). A case study is carried out on four homogeneous regions in different hydroclimatological regions in Canada. Using the trend centered technique and the four pooling groups, peak 50- and 100-year nonstationary quantiles are estimated. An uncertainty analysis of the examined regions indicates that there is less uncertainty in the nonstationary quantiles when compared to their stationary counterparts, thus reinforcing the effectiveness of this nonstationary regional frequency analysis technique.

This thesis additionally presents a peaks-over-threshold approach that makes use of covariate-dependent stationary and nonstationary thresholds. The generalized Pareto (GP) distribution is fit to threshold exceedance data using covariate-dependent location and scale parameters. Bivariate and multivariate models are used for both the threshold and GP models. The methodology is applied to precipitation data from 30 meteorological stations in coastal British Columbia (BC) due to the potential for climate change effects in the region. 50-year quantile estimates and their associated uncertainty are calculated, indicating that there is stronger evidence of stationarity, however numerous stations show nonstationary behaviour. A comparative trend analysis is also carried out on the frequency and magnitude of stationary threshold exceedance events for the winter and summer

seasons for the 1976-2014 and 1986-2014 time periods. The results from this analysis indicate that statistically and globally significant trends are present in the threshold exceedance and frequency of peak data.

The final contribution of this thesis focuses on the presence of serial correlation in hydrometeorological data. If autocorrelation is detected in time series data, standard nonparametric trend detection tests can produce erroneous results. Therefore, numerous techniques have been developed to account for detected serial correlation to improve trend detection accuracy. Several tests have previously undertaken the task of assessing the power and type I error rates of existing techniques, however, there has been little focus on negative serial correlation. Furthermore, a large number of the existing techniques assume the data to have an autoregressive lag-1 (AR(1)) memory structure, an assumption that is not consistently applicable. Accordingly, this research compares the power and type I error rates of various existing statistical techniques that account for positive and negative autocorrelation when used in combination with the Mann-Kendall nonparametric trend test. Using instantaneous peak data from a subset of southern Ontario watersheds, differences in the serial structure of the block maxima series (BMS) and peaks-over-threshold (POT) data are elucidated. It is determined that BMS data are more likely to have significant negative lag-1 serial correlation and are more likely to have AR(1) memory structures; positive significant autocorrelation is more commonly found in POT data. It is determined that the block bootstrap (BBS), the Hamed and Rao (1998) variance correction technique (VCCF1), and the sieve bootstrap (SBS) are the most powerful techniques in the presence of positive autocorrelation. Alternatively, the VCCF1 and BBS techniques are most appropriate when significant negative serial correlation is present. A comparative trend analysis is also undertaken on the same study area in which trends are assessed based on the land-use in the catchments of interest. The included land-use classifications include agricultural, natural, RHBN, and urban; trends in several hydrological characteristics are assessed. Differences in the detected trends using POT and BMS data are also revealed. Trends are identified in the natural and RHBN watersheds; however, there is a lack of consistency between those detected in the BMS and POT data. Furthermore, numerous statistically and globally significant trends are identified in the urban and agricultural catchments.

The overall contributions of this research include the development of a novel trend centered pooling approach for regional frequency analysis, the development of a covariate-dependent threshold selection methodology for at-site frequency analysis, and the comparison of the memory structure of BMS and POT data. Additionally, the effects of negative serial correlation on the type I error rates

and power of numerous statistical techniques developed for use with nonparametric trend tests are elucidated.

Acknowledgments

I would like to begin by thanking my supervisor, Prof. Don Burn who has been profoundly supportive, helpful, and caring along this journey. I am very fortunate to have been able to learn from such an accomplished researcher who gave me a great deal of flexibility to learn and explore my research topics. I am sincerely thankful for all his financial support and his overall kindness throughout my time at the University of Waterloo (UW).

I am deeply grateful for the amazing friends that I have made along the way. Special thanks go to Peter Thompson who has taught me so much over the years and has always gone above and beyond to help me whenever possible. I'm also very thankful for Andy Snowdon, Ben Plumb, and Mike Takeda who have all been wonderful friends and have all taught me so much about the amazing field of water resources. Last, but certainly not least, I am so thankful for my dear friend Steve Constable, who assured me that I wasn't the first person to have to do something 10,000 times. The level of expertise that I was able to attain programmatically over the years is directly associated with his selflessness and thoughtfulness. Words can't express how grateful I am for his kindness and how appreciative I am to have him as a friend.

During my time at UW, I had the pleasure of acting as a teaching assistant numerous times and am very grateful for each of those opportunities. Teaching was a very rewarding aspect of my degree program and I was able to learn so much from so many brilliant and caring instructors and professors. Very special thanks go to Dr. Bob McKillop and Dr. David Brush. Thank you both for believing in me and helping me to become the instructor that I am today. Furthermore, I'd like to acknowledge that my teaching was recognized twice at the departmental level with the Laurence Hamlin Memorial Award and a special recognition certificate for outstanding teaching.

I would additionally like to acknowledge Compute Canada/Calcul Canada given that a great deal of the code written for Chapter 4 was run in SHARCNET. Without this access, this research would not have been possible.

My warmest gratitude goes to my defense committee members, namely, Prof. Bill Annable, Prof. Keith Hipel, Prof. Mahesh Pandey, and Prof. Jan Adamowski. I'd additionally like to acknowledge my master's supervisor, Prof. Bruce Wilson, for being an amazing educator and providing me with so much support over the years.

I have been so fortunate to have an army of friends to have helped me throughout my PhD and I love them all dearly. My "tier" of best friends (Amanda Gibson, Rhonda Bell, Zoe MacKenzie, and

Jayme Stewart) have always been there to lend an ear. I am so utterly thankful for my husband Blair, who has been with me through thick and thin and given me unconditional love and support. I'm so blessed to have my sweet Eleanor to have helped me along the way; everyday with her is an adventure and it's a privilege to have her in my life.

My final acknowledgment is to my Mother. It's impossible to adequately describe the support that she's given me throughout my life and this degree. She has always been there to remind me that you never fail if you never stop trying. I'm indebted to her for all my success and promise to pay that same support forward to Eleanor.

Dedication

I dedicate this thesis to my Mother, Blair, Eleanor, and Sadie. All of my success has been a product of your unconditional love and support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
ACF	Autocorrelation function
AIC	Akaike information criterion
AMS	Annual maximum series
AR(1)	Autoregressive lag-1 model
ArkWSC	Extraction program for archived data
ARMA	Autoregressive-moving-average
AR(*p*)	Autoregressive lag-*p* model
BBS	Block bootstrap
BC	British Columbia
BCPA	Bayesian change point analysis
BCPW	Bias corrected prewhitening
BMS	Block maxima series
CCF	Cross-correlation function
CI	Confidence interval
ENSO	El Niño Southern Oscillation
ESS	Effective sample size
FFA	Flood frequency analysis
GCM	General circulation model
GEV	Generalized extreme value
GL	Generalized logistic
GN	Generalized normal
GP	Generalized Pareto
HYDAT	Hydrometric database
IACF	Inverse autocorrelation function
IF	Index flood
IID	Independently and identically distributed
IPACF	Inverse partial autocorrelation function
IPCC	Intergovernmental Panel on Climate Change
J-S	June to September
L-CV	Coefficient of L-variation
LIO	Land Information Ontario
MAICE	Minimum Akaike information criterion estimation
MK	Mann-Kendall
ML	Maximum likelihood
MTFPW	Modified trend-free prewhitening
NOAA	National Oceanic & Atmospheric Administration
NOAA-ESRL	National Oceanic & Atmospheric Administration – Earth System Research Library
O-D	October to December
OLS	Ordinary least squares
PACF	Partial autocorrelation function

PDO	Pacific Decadal Oscillation
POT	Peaks-over-threshold
PW	Prewhitening
Qdf	Flood-duration-frequency
Q-Q	Quantile plot
RHBN	Reference Hydrometric Basin Network
RMSE	Root mean square error
SBS	Sieve bootstrap
SRC	Spearman's Rank Correlation
TC	Threshold choice
TFPW	Trend-free prewhitening
US	United States
VCCF1	Variance correction technique 1
VCCF2	Variance correction technique 2
VCPW	Variance correction prewhitening
WSC	Water Survey of Canada

List of Symbols

Symbol	Description
a	Linear trend intercept estimate [m^3/s]
AIC	Akaike information criterion [-]
AIC_i	Akaike information criterion for site i [-]
AIC_j	Raw AIC [-]
b	Estimated slope [$m^3/s/year$]
b'	Modified slope estimator [$m^3/s/year$]
c	Number of hydrologic properties [-]
CF	Correction factor [-]
CF_1	Hamed and Rao (1998) correction factor [-]
CF_2	Yue and Wang (2004a) correction factor [-]
d_{ij}	Hydrologic response property for catchment i [-]
D_{ij}	Canberra dissimilarity metric [-]
d_{max}	Maximum geographic distance between any two catchments [km]
$E(S)$	Mann-Kendall test statistic mean [-]
e_t	White-noise from sample data [-]
H	Hosking and Wallis heterogeneity measure [-]
H_0	Null hypothesis [-]
H_1	Alternative hypothesis [-]
k	Lag [-]
K	Number of model parameters [-]
l_i	Record length of site i [-]
M	Number of sites in the region[-]
MD	Measure of the average time of occurrence of flood event [yearly day number]
$minAIC$	Minimum AIC
n	Number of observations [-]
N	Number of simulations [-]
n_i	Number of observations for site i [-]
N_{rej}	Number of simulated series in which the null hypothesis is rejected [-]
p	Probability of occurrence [-]
Q	Quantile [m^3/s]
$q(F,t)$	Nonstationary pooled growth curve [-]
$Q_i(F,t)$	Time-dependent quantile of site i [m^3/s]
Q_p	Generalized Pareto quantile [mm]
Q_T	Quantile of return period T [m^3/s]
$\bar{r}$	Measure of the spread of the data [-]

r_1	Lag-1 autocorrelation coefficient estimate [-]
r_1^*	Bias corrected lag-1 autocorrelation coefficient [-]
r_{1d}	Estimated lag-1 serial correlation coefficient from detrended series [-]
r_k^R	Lag-k autocorrelation coefficient of the ranks of the data [-]
R_{rej}	Rejection rate [-]
S	Mann-Kendall test statistic [-]
$s_{x_t'}^2$	Variance of x_t' [m^6/s^2]
$s_{y_t'}^2$	Variance of y_t' [m^6/s^2]
t	Time [years]
T	Return period [years]
$t^{(i)}$	Sample L-CV [-]
t_k	Number of ties [-]
t^R	Regional L-CV [-]
T_t	Linear trend [-]
u	Threshold [m^3/s][mm]
u_t	Nonstationary threshold at time t [mm]
V	Weighted variance statistic [-]
$V(S)$	Mann-Kendall test statistic variance [-]
$V^*(S)$	Modified Mann-Kenall test statistic variance [-]
w	Weighting factor that reflects the relative importance of d_{ij} and D_{ij} [-]
$w_j(AIC)$	AIC weight for site j [-]
x	IID random variable [mm]
$\bar{x}$	x-coordinate of the mean direction of flood dates [-]
x_i	Observations [mm]
$X(t)_i$	Data series at time t for site i [m^3/s]
x_t	Data at time t [m^3/s] [mm]
X_t	Simulated time series [-]
X_T	AR(1) model with the additional trend [-]
x_t'	Detrended series [m^3/s]
x_{t-1}	Observations of the original data at time $t-1$ [m^3/s]
X_{t-1}	Simulated time series at time $t-1$ [-]
x_{t-1}'	Detrended lagged time series [m^3/s]
x_t''	Final transformed series [m^3/s]
$\tilde{x}_t$	Transformed exponential time-dependent data [-]
Y_{ik}	Geographic distance between catchments i and k [km]
y_t	Prewhitened time series [m^3/s]
y_t'	Prewhitened detrended time series [m^3/s]

y_t''	Corrected trend-free prewhitened series [m^3/s]
$\bar{y}$	y-coordinate of the mean direction of flood dates [-]
Z	Standardized Mann-Kendall test statistic [-]
z_t	Prewhitened original series with detrended estimate lag-1 serial correlation coefficient [m^3/s]

Greek symbols

α	Significance level [-]
β_0	Intercept of the conditional regression quantile [mm]
β_1	Slope of the conditional regression quantile [mm/year]
γ	Probablity of type II error [-]
δ	Slope of linear trend [-]
$\Delta_j(AIC)$	Difference in AIC with respect to the minimum AIC [-]
ε_t	White-noise from simulated data [-]
ϵ_t	Transformed residual GP series at time t [-]
$\epsilon(t)_i$	Transformed residual GEV series at time t [-]
η	Amount to be added to block length [-]
θ_i	Angular value for the date of occurrence of event i [rad]
$\bar{\theta}$	Mean direction of the flood dates [rad]
ϑ	Vector of parameters [-]
ϑ_t	Vector of nonstationary parameters [-]
$\vartheta(t)$	Nonstationary distribution parameters [mm][-]
$\hat{\vartheta}_k^{(i)}$	Estimate of parameter k for site i [-]
$\hat{\vartheta}_k^R$	Regional estimate of parameter k [-]
κ	Stationary shape parameter [-]
$\kappa(t)$	Nonstationary scale parameter [-]
$\kappa_R(t)$	Time-invariant shape parameter [-]
λ	Rate parameter [threshold exceedances/year]
μ_i	Collection of m regional location parameters [-]
$\mu(t)$	Nonstationary location parameter [-]
μ_X	Mean of X_t [-]
$\mu_R(t)$	Time-dependent regional location parameter [-]
μ_V	Theoretical mean [-]
μ_ε	Mean of white-noise process[-]
ξ_i	Index-flood of site i [m^3/s]
ρ_1	Lag-1 serial correlation coefficient [-]
σ	Scale parameter [mm]
σ_i	Collection of m regional scale parameters [-]
$\sigma(t)$	Nonstationary scale parameter [-] [mm]
$\sigma_R(t)$	Time-dependent regional scale parameter [-]
σ_V	Theoretical standard deviation [-]
σ_X^2	Variance of X_t [-]
σ_ε^2	Variance of white-noise process [-]
τ	Regression quantile [mm]

τ_0 Selected nonstationary threshold/regression quantile [mm]

"Water is the driving force of all nature."

-Leonardo DeVinci

Chapter 1
Introduction

Extreme hydrological events can have a profound effect on human health and safety, in the form of extensive property/environmental damage and loss of life. The National Oceanic and Atmospheric Administration (NOAA) reports the 30 year flood loss averages of approximately $8 billion in damages and 82 deaths per year in the U.S. (NWS, 2015). In Europe, Hoyois and Guha Sapir (2003) reported a total of 264 flood disasters from 1973 to 2002, where they define a disaster by at least 10 deaths or 100 persons affected. The more recent flooding in the Canadian eastern Prairies in 2011 claimed one life, affected thousands and caused close to $1 billion in damages (EC, 2011). Although flooding events pose an obvious threat to those affected and the surrounding environment, the impacts can be minimized through effective flood risk planning and management. Commonly used methodologies rely on flood frequency analysis (FFA) techniques in order to estimate the scale of engineering design and flood protection measures. The methodologies in FFA have been developed on the premise of a stable climate, land use, and land cover, consequently producing hydrometeorological data free of inhomogeneities such as temporal trend.

It is widely accepted that the anthropogenic release of greenhouse gases into Earth's atmosphere is causing the mean global temperature to increase. We now live in a climate that is, on average, approximately 1ºC warmer than 50 years ago (Hartmann *et al*., 2013). In a warmer climate, the occurrences of flood events are projected to increase due to heavier precipitation amounts (Hartmann *et al*., 2013). In addition to climate change, land-use changes can have profound effects on the hydrological characteristics of a watershed. The 20th century has been characterized by intense land-use changes with respect to agricultural practices, urbanization, and forest management. These changes can cause shifts in hydrological and ecological systems and impact the rainfall-runoff relationship, thus affecting flood risk (Villarini *et al.*, 2009a). Also, the addition of impervious surfaces can lead to increased flood magnitude and decreased time-to-peak of extreme events. In the European Union, agricultural land has been lost to urbanization an average of 2% per decade, while Theobald *et al.* (2009) reported an estimate of an increase in impervious surface in the conterminous U.S. of 36.2% by 2030 (Kundzewicz *et al.*, 2010). Prior to the aforementioned anthropogenic forcings, hydroclimatalogical data were assumed to be independently and identically distributed (IID) meaning that the observations were independent of each other and all come from the same statistical

probability distribution. The assumption of IID data may no longer be valid due to the nonstationary character of the data. Hipel and McLeod (1994) define the stationarity of a stochastic process as "…being qualitatively interpreted as a form of statistical equilibrium. Therefore, the statistical properties of the process are not a function of time (p. 67)." Evidence of nonstationarity in the form of temporal trend has been found in peak flow data of catchments worldwide (Reed *et al.*, 1999; Jain and Lall, 2000; Zhang *et al.*, 2001; Burn and Hag Elnur, 2002; Hodgkins *et al.*, 2003; Burn and Cunderlik, 2004; Burn *et al.*, 2008; Burn *et al.*, 2010; Schmocker-Fackel and Naef, 2010; Bormann *et al.*, 2011; Hirsch, 2011). Many studies have focused on pristine catchments unaffected by land-use changes but these watersheds exhibit less apparent nonstationarities in the flow record and are sparsely populated. Areas affected by marked land-use change present the greater socio-economic risk; therefore, a focus of this thesis will be on both the analysis of watersheds affected by prominent land-use changes and potential climate change in addition to pristine catchments (Sugahara *et al.*, 2009; Vogel *et al.*, 2011).

The statistical techniques used in FFA have been developed for use with IID data. The return period of an event, or the average time between hydrological events of the same magnitude, is developed based on the IID assumption, therefore, in a changing climate with the increased occurrence of events, this characterization of these events becomes questionable. Within FFA, data can be analyzed on a site by site basis or through pooled frequency analyses. At-site flood frequency estimation is generally appropriate for watersheds with a streamflow record approximately twice the length of the return period of interest (Reed *et al.*, 1999). In the absence of this quantity of data, pooled frequency analysis can be used in lieu of at-site frequency estimation techniques. By pooling data from hydrologically similar sites, more accurate estimates can be determined. There is therefore a need to develop methods within the pooled flood frequency framework that account for possible nonstationarity in the pooling group member data.

Other types of nonstationarity aside from monotonically increasing/decreasing trends will also be explored. Cyclostationarity, caused by climatic teleconnections (i.e., low-frequency climatic oscillations) can, in some instances, be mistaken for increasing or decreasing trends in a series. Therefore, if cyclostationarity were detected and accounted for, the true nonstationarity within the data may then be detected.

The absence of IID data can be determined, in many instances, by the presence of an increasing or decreasing trend in the moments of statistical distributions fitted to the data. Modelling the time-dependent location and scale parameters has been carried out in a variety of contexts but the shape

parameters of extreme value distributions are more difficult to estimate (Coles, 2001). If a distribution is particularly skewed, this skewness may not be adequately represented in a limited sample of data (Hosking and Wallis, 1997). Thus, the uncertainty in estimating higher moments can be substantial and the significance of nonstationarities found in this data may be questioned. Due to potential bias in estimating the aforementioned parameter, the assumption of a constant shape parameter is used herein, while the assumption of a temporal dependence structure in the location and scale parameters has been employed (Katz *et al.*, 2002; Hanel *et al.*, 2009).

To ensure data are free of inhomogeneities, tests for trend and change-points are generally carried out, however, serial correlation (autocorrelation) within a data record can affect the results of this testing (von Storch, 1995; Yue *et al*., 2002b; Serinaldi and Kilsby, 2016). Numerous techniques have been developed for use with nonparametric trend testing methods to address the effects of serial correlation (von Storch, 1995; Hamed and Rao, 1998; Yue *et al*., 2002b; Yue and Wang, 2004a; Bayazit and Önöz, 2007; Khaliq *et al*., 2009; Önöz and Bayazit, 2012; Blain, 2013; Sonali and Kumar, 2013; Wang *et al*., 2015), though, many techniques make underlying assumptions that affect their accuracy.

Extreme hydrological data can be characterized in a variety of manners. There have, however, been several methods used more frequently than others. The block maxima and partial duration series, or peaks-over-threshold (POT) are commonly used for the purposes of FFA. Generally speaking, the block length employed for the block maxima is 1-year, resulting in the annual maximum series (AMS). The principal benefit of the block maxima series (BMS) is that the observations included in the series are most likely produced by separate hydroclimatological events, thus producing a series free of dependence between measurements. Issues may arise with the use of the BMS, as dry years are automatically incorporated into the hydrometric series, which may result in higher trend identification rates than would be found when using the POT series (Svensson et al., 2005). Therefore, the POT series may provide a more accurate representation of the true magnitudes of trends.

According to Milly *at al.* (2008), annual global investments in water infrastructure exceed U.S. $500 billion. Therefore, the risk to those affected by extreme hydrological events combined with the potential monetary loss of the overdesign of hydraulic structures are topics of concern. There is a need to develop more robust methods in FFA to deal with systematic changes in the extreme hydrological data (Sivapalan and Samual, 2009). Furthermore, employing risk-based techniques that consider both stationary and nonstationarity analyses, in addition to economic considerations, is a

valuable undertaking (Rosner *et al*., 2014). Projections for increasing or decreasing trends are largely dependent on the flood generating mechanisms of the catchment in question. Increasing flood magnitudes are expected where floods result from heavy rainfall and decreasing magnitudes are expected in areas where floods are characterized by snowmelt (Kundzewicz *et al*., 2010). Due to the dichotomy of the changing nature of flooding events, there is a need to characterize the uncertainty in their estimation for records displaying both increasing and decreasing temporal trend.

1.1 Objectives

When temporal dependence is apparent in the moments of a fitted distribution, the IID assumption no longer holds true and traditional statistical methods will produce results which may be significantly biased. In FFA, statistical distributions are fit to available data, which is carried out by determining the probability distribution function or cumulative distribution function (CDF) for a continuous random variable (x). The CDF is denoted by $F(x) = \Pr[Q \leq x]$, where $\Pr[Q \leq x]$ is the nonexceedance probability of the random variable, Q is the quantile (predetermined level or event of interest). FFA is concerned with the upper tail of the frequency distribution, primarily the quantile of return period T, Q_T, the inverse of the CDF, which is given by (Hosking and Wallis, 1997):

$$x(F) = Q_T = x(1 - 1/T) \tag{1-1}$$

where $F = 1 - 1/T$. It is clear from the above equation that the quantile is a function of the nonexceedance probability of an event. Therefore, the goal of flood frequency analysis is to estimate the Q- T relationship with minimal uncertainty (Burn 1989).

The overall objective of this thesis is to explore new methodologies for incorporating nonstationary data into extreme value analyses and to ascertain the effects of serial correlation on various types of peak event data. More specifically, the objectives of this research include:

1) The development of a trend centered pooling approach for regional frequency analysis of peak streamflow data based on the form of trend detected (either increasing or decreasing);
2) To develop a methodology for selecting covariate-dependent thresholds and generalized Pareto (GP) distribution parameters to determine those models that provide superior goodness-of-fit and less uncertainty for the purpose of quantile estimation;

3) To carry out a comparison of several well-known and other more recently developed techniques which account for serial correlation when the Mann-Kendall (MK) trend test is applied;
4) To determine the most appropriate techniques to account for autocorrelation in trend analyses when used on BMS and POT data; and
5) To provide an analysis of trends in southern Ontario watersheds based on land-use.

1.2 Thesis Organization

Chapters 2 and 3 of this thesis are provided in the form of manuscripts that have been published in scientific journals. Chapter 2 was published in the Journal of Hydrology (O'Brien and Burn, 2014) while Chapter 3 was published in the Canadian Water Resources Journal (O'Brien and Burn, 2018). Chapter 4 is presented as a thesis chapter; however, the intent is the submission of two manuscripts. Therefore, to facilitate the transition from Chapters 2 to 3 and Chapters 3 to 4, transition paragraphs are included to aid in the readability of this thesis.

The objective of Chapter 2 is the creation of a trend centered pooling approach based on the form of trend found in the at-site data (i.e., either decreasing or increasing). This chapter introduces the topic of pooled FFA and the index-flood technique. The methodology is outlined, along with a case study, showing the applicability and usefulness of the technique. This chapter also addresses the first objective listed above.

Chapter 3 presents a methodology for a covariate-dependent threshold and GP distribution model selection procedure, thus addressing the second objective of this thesis. The developed methodology for this technique, and an applicable case study are included.

The final three objectives are addressed by Chapter 4. Various well-known and more recent techniques that account for serial correlation when using nonparametric trend tests are examined. The power and type I error rates of these techniques in the presences of positive and negative autocorrelation are also examined. The chapter concludes with a comparative trend analysis based on land-use in a subset of watersheds in southern Ontario.

Chapter 5 presents the overall conclusions from this research and the potential for future research based on this work.

The list of references follows Chapter 5.

Chapter 2

A Nonstationary Index-Flood Technique for Estimating Extreme Quantiles for Annual Maximum Streamflow

This chapter is built upon the published article with the same title published in the Journal of Hydrology. Minor differences between the published paper and the chapter have been made to facilitate consistency and coherence.

O'Brien, N. L. and D. H. Burn. 2014. A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. *Journal of Hydrology*, 519: 2040-2048.

Summary

The magnitude and timing of peak streamflow events may be affected by land-use changes along with climate change, thus leading to nonstationarity in the records. Temporal trend, along with change-points, in peak flow records can affect the accuracy of quantile estimates; therefore, these issues should not be disregarded. Commonly used techniques for pooled flood frequency analysis do not account for nonstationarity found in the data recorded for members of a region. To overcome this shortcoming, the objective of this research is to introduce a trend centered pooling approach for regionalization in which pooling groups are created based on the form of trend found in the at-site data. The approach involves the formation of regions comprised entirely of sites exhibiting either statistically significant increasing or decreasing trends. Regional parameter estimates are determined using a maximum likelihood approach, which is carried out with the assumption of second-order nonstationarity. The technique was applied to four homogenous regions all located in differing hydroclimatological Canadian regions. The uncertainty of quantile estimates calculated through the implementation of this technique was established using a balanced regional vector resampling approach, which was found to be a comparatively advantageous approach (Burn, 2003). The results indicate that there is less uncertainty in quantile estimates found through the application of the trend centered pooling approach when compared to a regional stationary analysis of the same regions. The potential for overestimation/underestimation of design quantiles in the presence of significant regional nonstationarity (i.e., decreasing/increasing trends) is also elucidated.

2.1 Introduction

The 2013 flooding in Alberta, Canada and the 2013 Colorado, United States floods have sparked much discussion concerning adequate planning and protection from extreme hydrological events. Concern over flooding arises due to the substantial socioeconomic risks associated with these events. Research concerning the magnitude and timing of extreme hydrological events suggests that land-use changes along with climate change may lead to nonstationarity in peak streamflow records. Stationarity can be interpreted as a form of statistical equilibrium; therefore, the statistical properties of the process in question would not be time-dependent (Hipel and McLeod, 1994). Intensification of the hydrologic cycle is anticipated as the mean global temperature rises, which can affect the timing and magnitude of flooding events (Hartmann *et al*., 2013). In colder climates, the onset of spring snow-melt runoff is expected to shift from early spring to late winter thereby decreasing total runoff amounts while warmer climates may see an increase in total rainfall (Kundzewicz *et al*., 2010). In addition to these projected climate-change responses, land-use changes can have marked effects on the hydrologic characteristics of a watershed. The 20th century has been characterized by intense land-use changes with respect to agricultural practices, urbanization, and forest management. These changes can cause shifts in hydrological systems and impact the rainfall-runoff relationship, thus affecting flood risk (Villarini *et al*., 2009a). In particular, the addition of impervious surface can lead to increase flood amplitudes and decreased time-to-peak of flooding events. Evidence of these variations in runoff has been detected in peak flow data worldwide in the form of temporal trend (Robson and Reed, 1999; Jain and Lall, 2000; Zhang *et al*., 2001; Burn and Hag Elnur, 2002; Hodgkins *et al*., 2003; Burn and Cunderlik, 2003; Burn *et al*., 2010; Schmocker-Fackel and Naef, 2010; Bormann *et al*., 2011; Hirsch, 2011).

Modelling nonstationarity in extreme hydrometeorological data is a topic that has received a great deal of attention in recent years (Kharin and Zwiers, 2005; Khaliq *et a*l., 2006; El Adlouni *et al*., 2007; Kyselý *et al*., 2010, Westra *et al*., 2013). This may be due to the potential for systematic errors in quantile estimates in the presence of nonstationarity. For example, an increasing trend in the mean of a flow series could lead to underestimated at-site flood quantiles if standard statistical methods are applied, thus increasing the risk of failure of engineering structures. There is, however, less research with respect to the implications of nonstationarity in pooled FFA. Using pooled FFA allows for the incorporation of data from several sites, thus allowing for greater accuracy in estimating peak flows. Standard techniques used in pooled FFA were developed for independently and identically distributed (IID) data, meaning data are independent and belong to the same statistical probability distribution.

In the presence of nonstationarity, the IID assumption is no longer valid and results obtained using standard methodologies may be inaccurate. There is therefore a pressing need for the development of more precise pooled FFA techniques when time-dependence is exhibited in hydrological data.

Smith (1989) is an early reference that accounts for nonstationarity in a pooled frequency analysis. This study used IID data, which were not a function of time but of catchment characteristics. More recently, Cunderlik and Burn (2003) proposed a novel detrending approach for FFA for nonstationary data. Cunderlik and Ouarda (2006) developed a nonstationary approach to regional flood-duration-frequency (Qdf) modelling using the index-flood method. Hanel *et al*. (2009) propose a nonstationary index flood (IF) equation in which the pooled growth curve and index-flood both vary with time. The authors apply their methodology to the Rhine basin where regions are determined subjectively. The IF approach develop by Hanel *et al*. (2009) is also used by Roth *et al*. (2012) but transformed for the use of a peaks-over-threshold model with nonstationary parameters. Their methodology is applied to the Netherlands, which has been determined to be a suitably homogeneous region in previous studies. Other work has focused on modelling nonstationarity in regional frequency analysis using a Bayesian framework (Leclerc and Ouarda, 2007; Renard *et al*., 2006).

The central focus of this research is the development of a methodology for pooling peak streamflow data based on the form of trend detected (i.e., either increasing or decreasing trends). The standard techniques used in regional frequency analysis have been developed for data that are not temporally dependent. Therefore, if a region has members displaying nonstationarity, systematic errors may be introduced into the analysis. This becomes particularly important as more recent data are collected, potentially resulting in an increasing number of statistically significant trends in peak annual flow data in Canadian watersheds and in catchments worldwide (Bormann *et al*., 2011; Burn and Cunderlik, 2003; Villarini *et al*., 2009b). This trend centered pooling technique initially involves testing applicable data for trends using the Mann-Kendall test. Sites that display temporal trend in the mean tendency are then used for the creation of homogeneous regions. This methodology is employed under the premise of second-order nonstationarity, which results in the formation of regions comprised solely of sites exhibiting significant trends, on which the remainder of the analysis is focused.

2.2 Methodology

This section outlines the methodology used with respect to the trend centered pooling technique. The primary goal of this section is to describe the nonstationary index-flood model, the regionalization technique implemented, and the parameter estimation methodology used for this research.

2.2.1 Data Screening and Trend Detection

The proposed regionalization technique makes use of statistically significant trends found in individual site data. Trend detection is carried out using the non-parametric Mann-Kendall test, which is a rank-based test commonly used for analysis of hydrometric variables. The Mann-Kendall trend statistic allows for the determination of the significance of a trend found in the data (Mann, 1945; Kendall, 1975). Significant serial correlation in individual site data can have deleterious effects on the robustness of trend detection tests; therefore autocorrelation was addressed using trend-free prewhitening (TFPW). Following Yue *et al*. (2002), TFPW consist of fitting a linear trend to the time series, data are then pre-whitened and the final result involves blending the monotonic trend and the pre-whitened residual series. The use of this technique allows for the removal of serial correlation while maintaining any trend present in the data.

The power of the Mann-Kendall test can also be affected by change-points or shifts in the data. Therefore, the homogeneity of the data was assessed using Bayesian Change Point Analysis (BCPA), proposed by Barry and Hartigan (1992; 1993). BCPA is a parametric test allowing for the detection of multiple change-points in a time series. Regime shift testing is an important precursor for trend detection testing as change-points in the data may lead to the detection of trends when none exist.

2.2.2 Index Flood Model

The index flood technique (Dalrymple, 1960) provides a framework for determining design flood estimates that continues to be applied in a number of studies (Robson and Reed, 1999; Hanel *et al*., 2009; Roth *et al*., 2012; Ilorme and Griffis, 2013; Norbet *et al*., 2014; Wright *et al*., 2014). This method assumes that within an acceptably homogenous region, the flood response from all members is identically distributed aside from a site-specific scaling factor (i.e., the index flood). This model has been modified herein to include the nonstationarity of the data and can be described by the following model:

$$Q_i(F,t) = \xi_i q(F,t) \tag{2-1}$$

where $Q_i(F,t)$ is the time-dependent flood quantile of site i, ξ_i is the index-flood of site i, and $q(F,t)$ is the nonstationary pooled growth curve. The index-flood is taken as the mean of the at-site data. Under the assumption of second-order nonstationarity, distributional parameters can be modeled as a function of time, which can be expressed most generally as:

$$\begin{aligned} \mu_R(t) &= \mu_0 + \mu_1 t + \mu_2 t^2 + \cdots + \mu_m t^m \\ \sigma_R(t) &= \sigma_0 + \sigma_1 t + \sigma_2 t^2 + \cdots + \sigma_m t^m \\ \kappa_R(t) &= \kappa \end{aligned} \tag{2-2}$$

where $\mu_R(t)$ and $\sigma_R(t)$ are the time-dependent regional location and scale parameters, μ_i and σ_i are a collection of m regional parameters, and $\kappa_R(t)$ is the time-invariant shape parameter. When $m = 1$, Equation (2-2) reduces to a linear model. Although it is possible to model nonstationarity in all three parameters, the shape parameter is kept constant for the purposes of this research. Evidence of a constant shape parameter has been found in various hydroclimatological data series, even when time-dependence is identified in the location and/or scale parameters (Vinnikov and Robock, 2002; Kharin and Zwiers, 2005; Kyselý *et al*., 2010; Kay and Jones, 2012). Also, given the average length of streamflow record and the potential for large sample variability when estimating higher moments, a focus on the nonstationarity of the first two moments is appropriate (Cunderlik and Burn, 2003).

2.2.3 Regionalization

An agglomerative hierarchical clustering technique is used for region formation. This clustering technique is carried out by initially determining an interobject dissimilarity between objects, which is given by (Burn *et al*., 1997):

$$D_{ij}^{d} = \frac{D_{ij} + \frac{d_{ij}}{d_{max}} w}{1 + w} \tag{2-3}$$

where D_{ij} is the dissimilarity between catchments i and j in terms of the hydrologic response variables chosen; d_{ij} is the geographic distance between catchments i and j; $d_{\max}$ is the maximum geographic distance between any two catchments; and w is a weighting factor that reflects the relative importance of d_{ij} (scaled by $d_{\max}$, the maximum distance between any pair of catchments) and D_{ij}. The formation of pooling groups should not be solely based on spatial proximity but on a set of

attributes that effectively describes the hydrological response of the watersheds (Hosking and Wallis, 1997). With this consideration, the dissimilarity measure described above was implemented to allow for the formation of groups using hydrologic characteristics, while ensuring group formation within similar hydroclimatological regions. As per Burn *et al.* (1997), a weighting factor of 0.3 was employed. The final term in Equation (2-3), D_{ij} is the Canberra dissimilarity metric given by (Burn *et al.*, 1997):

$$D_{ij} = \frac{1}{c}\sum_{k=1}^{c} \frac{|Y_{ik} - Y_{jk}|}{Y_{ik} + Y_{jk}} \tag{2-4}$$

where Y_{ik} is the value of the k^{th} hydrologic response property for catchment i; and c is the number of hydrologic properties.

For the purposes of this research, hydrologic response properties concerning the timing and variability of peak flow events are used. Similarities in the time of occurrence of flood events can be quite useful for regionalization. Angular values of the date of occurrence are calculated following Bayliss and Jones (1993) and Burn (1997) by:

$$\theta_i = (Julian\ Date)_i \left(\frac{2\pi}{365}\right) \tag{2-5}$$

where θ_i is the angular value (in radians) for the date of occurrence of event i. It is often helpful to plot the mean flood dates, the coordinates of which are given by:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \cos(\theta_i)$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \sin(\theta_i) \tag{2-6}$$

where n is the number of peak events. With the $\bar{x}$- and $\bar{y}$-coordinates available, the mean direction of the flood dates is obtained by:

$$\bar{\theta} = \tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right) \tag{2-7}$$

The mean direction can be converted back into a day of the year using:

$$MD = \bar{\theta}\frac{365}{2\pi} \tag{2-8}$$

where MD is a measure of the average time of occurrence of the flood events for a given catchment. Using the $\bar{x}$- and $\bar{y}$-coordinates, a measure of the variability of the occurrences of peak flow events can also be determined through:

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \tag{2-9}$$

where $\bar{r}$ characterizes the spread of the data in a given catchment, which can range from 0 to 1. Values equal to unity would indicate that peak flows for a given catchment occur on the same day of the year, while values closer to zero would represent a great deal of variability in the flood dates. The hydrologic response properties used in Equation (2-4) included MD and $\bar{r}$. MD represents the average time of occurrence and $\bar{r}$ the variability in the timing of events, thus the use of both measures encompasses the timing and variability of peak events in selected sites.

Following the calculation of the dissimilarity between catchments, region formation is carried out using the agglomerative hierarchical clustering technique, which begins with the union of the two closest catchments in dissimilarity space. The dissimilarity is redefined between the new cluster and all other objects which allows for the inclusion of the next closest site. The algorithm continues until the desired number of sites have been included in the region.

Having applied the Mann-Kendall trend test to the data, the flow records may be separated into two subsets (i.e., negative and positive trends) and regionalization can then be carried out within each subset. The most common approach for testing regional homogeneity was developed by Hosking and Wallis (1997), which is applied after the desired number of stations have been included in the region. The Hosking and Wallis heterogeneity measure is initially calculated through the specification of a weighted variance statistic given by:

$$V = \left\{\sum_{i=1}^{M} l_i\left(t^{(i)} - t^R\right)/\sum_{i=1}^{M} l_i\right\}^{1/2} \quad (2\text{-}10)$$

where M is the total number of sites in the region, l_i is the record length of site i, $t^{(i)}$ is the at-site (sample) L-CV (coefficient of L-variation), and t^R is the regional L-CV. Simulation experiments are then carried out to determine the theoretical mean (μ_V) and standard deviation (σ_V) of V. This results in the following heterogeneity measure:

$$H = \frac{V - \mu_V}{\sigma_V} \quad (2\text{-}11)$$

A region can be considered acceptably homogeneous if $H < 1$, possibly heterogeneous if $1 \leq H \leq 2$, and definitely heterogeneous if $H > 2$ (Hosking and Wallis, 1997). The heterogeneity measure makes use of the sample and regional L-CV, which requires an ordered sample of data. As a result, potential bias may be present when applying this homogeneity test in the presence of nonstationarity (Tramblay *et al*., 2003). The creation of sufficiently homogenous regions is the precursor for parameter estimation but before parameter estimates can be determined, all at-site data were scaled by their stationary mean annual flood (Hosking and Wallis, 1997). It should be noted that this method is designed for cases where there is considerable overlap in the period of record for the sites in a pooling group.

2.2.4 At-Site Parameter Estimation

At-site parameter estimates are determined using the method of maximum likelihood (ML) as it provides flexibility for the incorporation of covariates into the estimates of distribution parameters. It is often less computationally intensive to take the logarithms of the likelihood function where the at-site log likelihood function $(\ell(\vartheta)_i)$ can be denoted as:

$$\ell(\vartheta_t)_i = \sum_t \ log\, f(x_t |\, \mu(t), \sigma(t), \kappa) \quad (2\text{-}12)$$

where ϑ_t is a vector of nonstationary parameters, $f(x_t |\, \mu(t), \sigma(t), \kappa)$ is the density function for site i, x_t are the time-dependent at-site data, $\mu(t)$ and $\sigma(t)$ are the nonstationary location and scale parameters, respectively, and κ is the stationary shape parameter. Due to the complexity of the log likelihood function, solutions are generally attained numerically through non-linear optimization

algorithms. The Nelder-Mead simplex algorithm was used for optimization herein (Nelder and Mead, 1965). In this study, three three-parameter distributions were considered as potential candidates for the regional growth curve, which include the generalized extreme value (GEV), generalized logistic (GL), and the generalized normal (GN) (Hosking and Wallis, 1997). Nonstationary parameter estimation was carried out using R, which facilitated the implementation of the Nelder-Mead algorithm through the use of the optim-function (Coles, 2001; R Core Team, 2016). Martins and Stedinger (2000) discuss how the ML estimation technique may lead to statistically/physically unreasonable estimates of skewness in small sample sizes. Therefore, shape parameter estimates were assessed for each site and at-site data that provided unrealistic results were removed from the analysis. For example, reasonable shape parameter estimates for the GEV distribution will likely fall between -0.3 and 0.3 (Martins and Stedinger, 2000).

2.2.5 Model Selection

For each site within a region, the aforementioned probability distributions are tested for best fit using the Akaike information criterion (AIC), which is given by (Akaike, 1974):

$$AIC_i = -2[\ell(\vartheta)_i] + 2K \tag{2-13}$$

where $\ell(\vartheta)_i$ is the at-site log likelihood for site i; and K is the number of model parameters. Given that goodness-of-fit of several distributions must be determined, the minimum AIC estimation (MAICE) procedure is implemented, where the distribution providing the lowest AIC value is chosen as the most suitable (Akaike, 1974). Therefore, the AIC_i values are summed for each distribution over all sites and the lowest overall value is chosen as the most suitable for the region.

2.2.6 Regional Parameter Estimation

Having determined the at-site nonstationary parameter estimates, pooled growth curve estimates can be calculated. Within an acceptably homogeneous region, parameter estimates are well described by the regional average. Therefore, at-site parameter estimates are combined in accordance with Hosking and Wallis (1997) to obtain regional parameter estimates:

$$\hat{\vartheta}_k^R = \frac{\sum_{i=1}^{M} n_i \hat{\vartheta}_k^{(i)}}{\sum_{i=1}^{M} n_i} \tag{2-14}$$

where $\hat{\vartheta}_k^R$ is the regional estimate of parameter k, n_i is the number of observations at site i, and $\hat{\vartheta}_k^{(i)}$ is the at-site estimate of parameter k. Calculated values of the parameter estimates (i.e., $\mu_R(t)$, $\alpha_R(t)$, and κ_R) can now be substituted into the distribution in question, allowing for the determination of $q(F,t)$, the pooled growth curve.

2.2.7 Uncertainty

One of the benefits of the maximum likelihood approach is the convenience of estimating standard errors with respect to the parameters of the distribution through the inversion of the information matrix of the likelihood function. But with small samples, bootstrap resampling procedures may be of more use in assessing the standard errors of the parameter estimates and design values (Katz *et al.*, 2002). Therefore, a balanced resampling approach was implemented, as per Burn (2003). This approach recommends creating 999 copies of the original sample then concatenating these copies into a new longer sample. This concatenated series is then randomly permutated and divided back into the original number of copies. The result is 999 bootstrapped resamples that can be used to determine the uncertainty in quantiles estimates.

Given the nonstationarity found in the flow records used herein, data were transformed, resulting in a residual IID series (Katz, 2002). Equation (2-15) below demonstrates this transformation using the GEV distribution:

$$\epsilon(t)_i = \{1 - \kappa_i[X(t)_i - \mu_i(t)]/\sigma_i(t)\}^{1/\kappa_i} \tag{2-15}$$

where $\epsilon(t)_i$ is the transformed residual series at time t, for site i; $X(t)_i$ is the original data series at time t, for site i; $\mu_i(t)$ and $\sigma_i(t)$ are the time-dependent location and scale parameters for site i; and κ_i is the stationary shape parameter for site i. Final samples were then created by converting the bootstrapped residuals into their original form through the inverse of Equation (2-15).

To preserve the spatial dependence between regional members, a vector resampling approach is used to determine uncertainty in pooled quantile estimates. This is carried out by ensuring that for any given year, all sites having a flow value for that year were included in the bootstrapped samples. The balanced resampling approach is then applied to each vector of streamflow data. Similar to the above noted procedure, 999 copies of the original sample are concatenated, permutated, then divided back into 999 bootstrapped samples. 95% confidence intervals (CIs) are then calculated through the use of the resampled quantiles. The vector resampling approach is also used to assess the uncertainty

in quantile estimates using a stationary methodology, which allows for a comparison between the two techniques.

2.3 Application

The nonstationary pooled flood frequency model described in this study is demonstrated on four pooling groups from different physiographic Canadian regions. Data from the Water Survey of Canada's HYDAT database were used for region formation. Site selection excludes stations potentially affected by anthropogenic land-use change; those included are in the Reference Hydrometric Basin Network (RHBN), which is comprised of catchments with relatively pristine or stable land-use conditions (Harvey *et al*., 1999).

Peak annual daily streamflow data from unregulated sites with at least 15 years of record were assessed for trend using the Mann-Kendall test at the 10% significance level. Only data extending to at least 2006 were retained for analysis. Uncharacteristically low or high flow values at the beginning or end of the flow record can cause erroneous results when applying the Mann-Kendall test, therefore all records were screened for such occurrences with neighboring sites to ensure the legitimacy of the flow values.

The application of the clustering approach described in section 2.2.3 resulted in the formation of four pooling groups, which are shown in Figure 2-1. Three of the regions are comprised of sites exhibiting negative trend (Regions 1-3) and one with positive trend (Region 4). It has been suggested that pooling group formation be founded on the 5T guideline, where T is the return period of interest (Reed *et al*., 1999). The guideline suggests the inclusion of 5 times more station years in a region than the return period of interest to account for potential spatial dependence between members of the pooling group. Therefore, the accuracy of the quantiles of the return period of interest is dependent on the number of station years in the region. Sites were formed with the intention of adhering to this guideline but limitations in the number of significant at-site trends restricted the size of the pooling groups, which was only a concern for the 100-year return period quantiles. Although the regions formed in this analysis were smaller than desired (i.e., less than 5T number of station years), it is believed that valuable information may still be garnered from their use (Reed *et al*., 1999). Due to the potential impacts of spatial correlation among sites in the regions (Castellarin *et al*., 2008), intersite dependence was assessed following the methodology of Douglas *et al*. (2000). Through this analysis, it was determined that cross-correlation amongst the members of all the regions was not statistically significant at the 5% significance level.

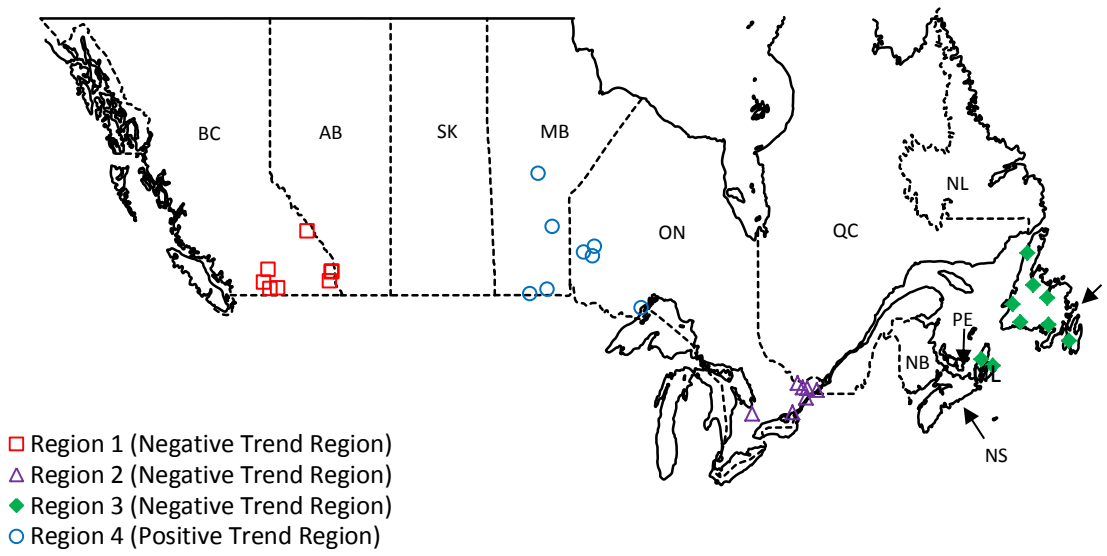

Figure 2-1: Location of nonstationary pooling groups.

The regions comprise four distinct physiographic areas. Region 1, located in the South British Columbia Mountains Climate Region, is characterized by heavy snowfall, leading to peak events predominantly from spring and early summer meltwaters. The hydrologic regime of Region 2, situated in the Great Lakes-St. Lawrence Region is characterized by spring snowmelt events. Region 3, located in the Atlantic Canada Region is also characterized by spring freshet events. Lastly, peak flood events in Region 4, located in the Prairie Region, are most commonly generated from spring snowmelt (Hare and Thomas, 1979). The seasonality measures for Regions 1-4 are shown in Figure 2-2, where each point is plotted using the $\bar{x}$- and $\bar{y}$-coordinates defined in Equation (2-6). Table 2-1 shows the regional members of the pooling groups created using sites displaying significant negative trend, their respective number of station years and Hosking and Wallis heterogeneity measure. Table 2-2 is similar to Table 2-1 but outlines the information from the region created from sites displaying statistically significant positive trends.

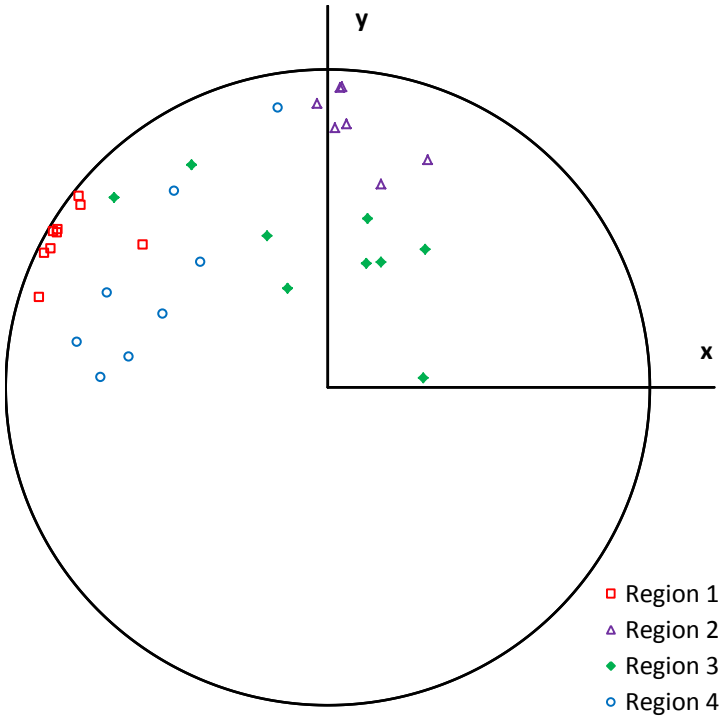

Figure 2-2: Seasonality measure for Regions 1-4.

Table 2-1: Nonstationary negative regional information.

Negative Trend Regions			
Region	**Station IDs**	**Station Years**	**Heterogeneity Measure**
1	08LG016 08NB016 08NK002 08NK016 08NK018 08NK022 08NL004 08NL024 08NM171	430	1.94
2	02ED014 02KF011 02LA007 02LB007 02LB020 02MC001 02HM005	287	1.11

Negative Trend Regions			
Region	**Station IDs**	**Station Years**	**Heterogeneity Measure**
3	01FB001 01FJ002 02YC001 02YJ001 02YK005 02YO008 02ZC002 02ZF001 02ZK004	301	0.99

Table 2-2: Nonstationary positive regional information.

Positive Trend Regions			
Region	**Station IDs**	**Station Years**	**Heterogeneity Measure**
4	02AB019 05OB021 05OD001 05QC003 05QE008 05QE012 05RB003 05UA003	276	1.93

2.4 Results

For the purposes of this study, six models were assessed for goodness-of-fit and are listed in Table 2-3. Nonstationarity was modeled in either the location and/or scale parameters of the three-parameter distributions used herein. Due to limitations in the record length of the pooling group members, linear trends were used to model the nonstationarity in the distributional parameters. Given that a trend in the location parameter does not necessitate the same form of trend in the scale parameter, the form of trend in the scale parameters is assessed in each region. For example, a site may have an increasing trend in the location parameter and either an increasing or decreasing trend in the scale parameter. If a mixture of increasing and decreasing trends are present in the at-site scale parameter estimates in a pooling group, the final regional estimates may not be representative of the form of trend in the at-site data. In all the included regions, there is a mixture of trends in the scale parameters (i.e., both increasing and decreasing). Therefore, the scale parameter is kept constant for the remainder of this research.

Table 2-3: Summary of models used.

Models	**Location/Scale Parameters of the Models**
GEV-1 GL-1 GN-1	$\mu(t) = \mu_0 + \mu_1 t, \quad \sigma(t) = \sigma$
GEV-2 GL-2 GN-2	$\mu(t) = \mu_0 + \mu_1 t, \quad \sigma(t) = \sigma_0 + \sigma_1 t$

Maximum likelihood estimates are then determined for each regional member and an overall ranking was assigned to each distribution. An overview of the ranked distributions with nonstationary location parameters can be found in Table 2-4. Having determined the most suitable regional growth curve for each of the four regions, nonstationary regional parameters are then calculated. This was done for each of the distributions highlight in Table 2-4 and summarized in Table 2-5.

Table 2-4: Overview of goodness-of-fit (ranking).

	Overall Rank / Goodness-of-Fit			
Distribution	**Region 1 (AIC)**	**Region 2 (AIC)**	**Region 3 (AIC)**	**Region 4 (AIC)**
GEV-1	2 (394.5)	2 (204.4)	2 (185.6)	***1 (408.4)***
GL-1	3 (402.8)	3 (211.6)	3 (191.2)	3 (424.7)
GN-1	***1 (393.6)***	***1 (203.1)***	***1 (182.7)***	2 (415.8)

*Bold and italicized cells represent the distribution with the best fit.
*Bracketed figures represent the regional total AIC values.

Table 2-5: Summary of the pooled growth curve parameters.

		Pooled Growth Curve Parameters			
		Nonstationary		Stationary	
		Location Parameters		Scale	Shape
Region	**Distribution**	μ_0	μ_1	σ	k
1	GN-1	1.0630	-0.0058	0.3560	-0.3978
2	GN-1	1.1679	-0.0102	0.3158	-0.2530
3	GN-1	1.1113	-0.0103	0.2917	-0.3746
4	GEV-1	0.4927	0.0183	0.4187	0.1102

The 50- and 100-year return period quantiles are estimated using Equation (2-1) for the sites used as the basis for the trend centered regionalization approach. Both the nonstationary and stationary quantiles are shown in Figures 2-3(a-d) and 2-4(a-d), for the 50-year and 100-year quantiles, respectively. Upon comparison of the time-dependent and stationary quantiles, it is clear there is potential for underestimation in the case of increasing trends with time and vice versa with regards to decreasing quantiles. For example, if the trend for Region 4 (Figure 2-4(d)) was projected to 2040, the nonstationary 100-year quantiles would have been underestimated by approximately 70 m^3/s (~ 40%). An examination of Figure 2-3 shows that the stationary quantiles do not exclusively lie close to the mid-point of the nonstationary quantiles. A possible explanation may be attributed to the

difference between the at-site slope estimates and the regional slope estimate. That is, if the site in question had a slope close to the regional estimate, it would be more likely for the stationary and nonstationary quantiles to intersect mid-way.

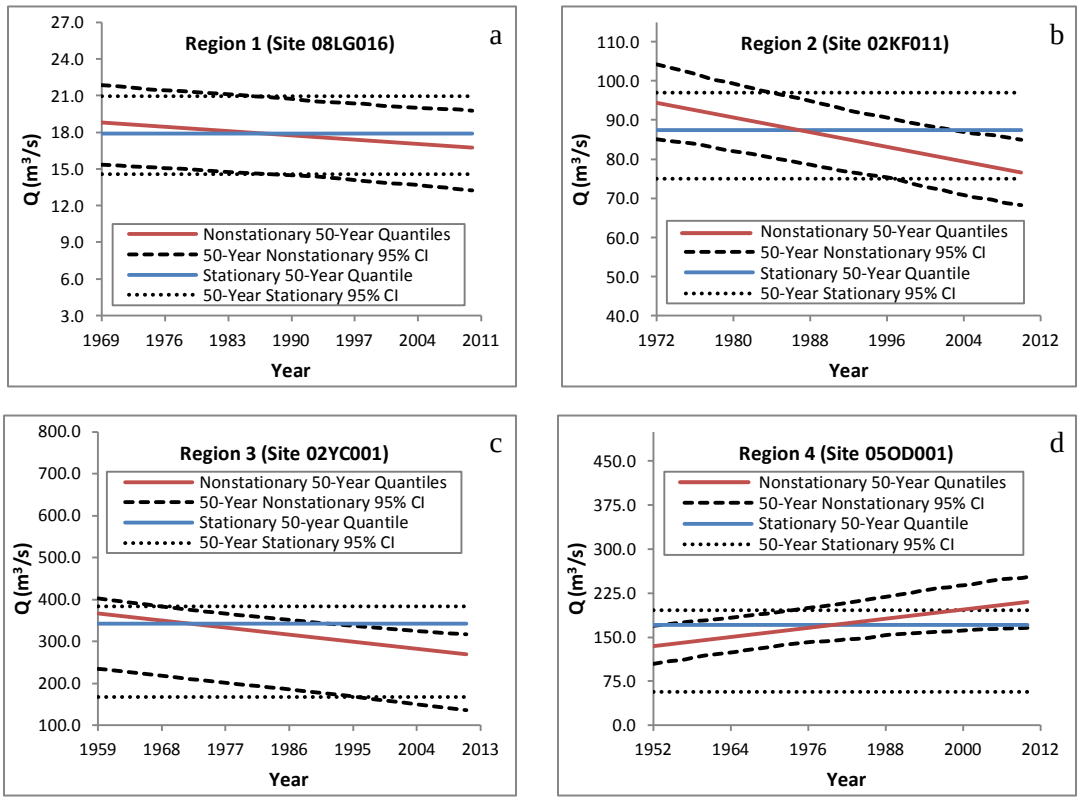

Figure 2-3: 50-year stationary and nonstationary quantiles and their associated 95% confidence intervals; (a) Region 1 (b) Region 2 (c) Region 3 (d) Region 4.

An uncertainty analysis for both the 50- and 100-year stationary and nonstationary quantiles is carried out using the vector resampling approach, described in section 2.2.6. Uncertainty is estimated in both cases to allow for comparison. The final time-dependent 95% confidence intervals for Regions 1-4 are shown in Figures 2-3(a-d) and 2-4(a-d). One site was chosen within each region for Figures 2-3 and 2-4 for graphical purposes.

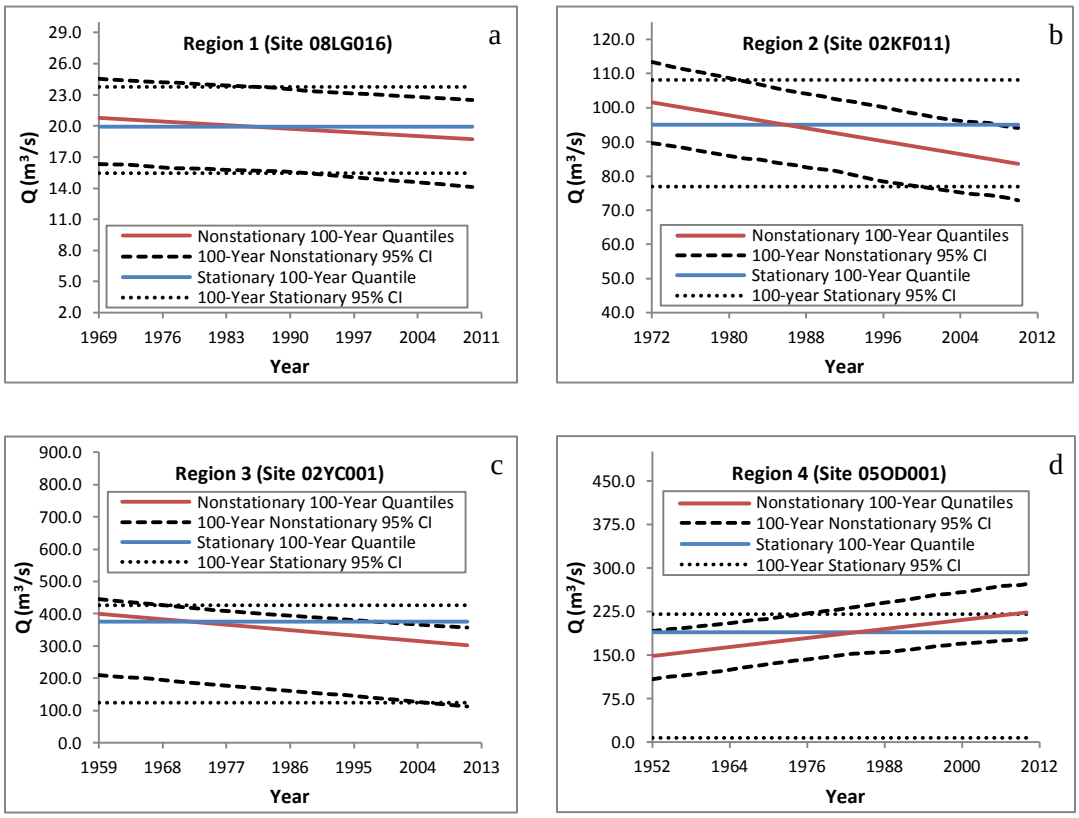

Figure 2-4: 100-year stationary and nonstationary quantiles and their associated 95% confidence intervals; (a) Region 1 (b) Region 2 (c) Region 3 (d) Region 4.

The implementation of the trend centered pooling approach allows for a comparison between the goodness-of-fit of the proposed technique to the stationary regional frequency analysis procedure, the results of which are shown in Table 2-6. The Hosking and Wallis (1997) Z^{dist} is the standard goodness-of-fit technique used in regional FFA, and is used in this comparison for the stationary analysis. Table 2-6 reinforces the need for the inclusion of nonstationarity in pooled FFA, as there are apparent differences in the stationary and nonstationary goodness-of-fit results. In the case of Region 1, for example, if the best fitting distribution were selected using a stationary methodology, the GEV distribution would have been chosen. Given that this distribution was not the highest ranking for the nonstationary goodness-of-fit, quantiles determined through its use may be erroneous. There appear to be similarities in the goodness-of-fit of both the stationary and nonstationary techniques in the remaining regions but this may not hold true if a larger data set were available.

Table 2-6: Goodness-of-fit comparison of the proposed methodology and the Hosking and Wallis (1997) Z^{dist} measure.

	Overall Rank / Goodness-of-Fit			
Distribution	**Region 1**	**Region 2**	**Region 3**	**Region 4**
GEV-1	2	2	2	1
GEV (stationary)	1	1	2	2
GL-1	3	3	3	3
GL (stationary)	3	3	3	3
GN-1	1	1	1	2
GN (stationary)	2	2	1	1

2.5 Discussion and Conclusions

Pooled FFA is a valuable technique due to its use of data from hydrologically similar sites, thus allowing for more accurate quantile estimation. As climate change effects progress and land-use changes become more prominent, time-dependence in peak streamflow records may become increasingly common. Therefore, the need for robust flood estimation techniques in the presence of nonstationarity is a pressing concern.

A trend centered pooling approach is developed herein to account for statistically significant trends found in peak annual flow records using an index-flood approach. Due to the limited number of statistically significant trends found in the HYDAT database, regional formation was restricted to four pooling groups. The uncertainty analysis carried out for the 50-year and 100-year quantile estimates for Regions 1-4 confirmed that the methodology presented in this paper is more effective than a stationary analysis when significant regional nonstationarity is present. It is also determined that in the presence of significant regional nonstationarity, there is potential for overestimation/underestimation of design quantiles dependent on the form of trends found in the at-site data.

Transition Paragraph A

Nonstationarity in hydrometeorological data may result in inaccuracies in peak quantile estimates. Time-dependent data may become more common as the mean global temperature increases, thus resulting in the likely intensification of the hydrologic cycle (Hartmann *et al*., 2013). The impacts of global climate change have been documented throughout Canada and amplify existing climate hazards (Walker *et al*., 2008). However, the socioeconomic risks of these hazards are most pronounced in densely populated regions. Coastal BC is one such area as it includes the Greater Vancouver Regional District (Metro Vancouver) in which approximate 75% of the total provincial populace reside (Walker *et al*., 2008). Substantial evidence of increasing temperatures in the region in recent decades (Zhang *et al*., 2000; Whitfield *et al*., 2002; Mote, 2003; Barrow *et al*., 2004; Meyn *et al*., 2013;) are expected to exacerbate hydroclimatological risks (BCME, 2007). Urban flash flooding is a prominent consequence of extreme precipitation in the Metro Vancouver area, furthermore, the municipality of Delta, BC has experienced marked flooding from the adjacent Fraser River (Baron *et al.*, 2012). Other extreme precipitation related hazards in the area include landslides and mudflows due to the proximity to the mountains (NVD, 2015). In the previous chapter, it was demonstrated that nonstationarity in peak streamflow data can result in inaccurate pooled quantile estimates. This chapter[1] also focuses on peak quantile estimation, however, the context shifts from streamflow to precipitation estimates. Natural sources of climate variability have a pronounced effect on the western Canadian climate, the most prominent being the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO). Statistically significant extreme precipitation trends may be the result of these climatic variations; therefore, employing a methodology that accounts for these oscillatory patterns is needed for accurate trend identification. The objective of this chapter[1] is to develop a methodology through which extreme precipitation quantiles are estimated and the results of which are compared to standard trend detection results. The developed technique allows the reader to visually assess existing trends in peak precipitation quantile estimates, whereby, climatic variability has been accounted for in the central tendency and variability of the data. This chapter focuses on daily precipitation, while the preceding chapter focused on annual maximum streamflow data.

[1]O'Brien, N. L. and D. H. Burn. 2018. A Nonstationary Peaks-Over-Threshold Approach for Modelling Daily Precipitation in Coastal British Columbia. *Canadian Water Resources Journal*, DOI: 10.1080/07011784.2018.1455538

Chapter 3

A Nonstationary Peaks-Over-Threshold Approach for Modelling Daily Precipitation in Coastal British Columbia

This chapter is built upon the published article with the same title published in the Canadian Water Resources Journal. Minor differences between the published paper and the chapter have been made to facilitate consistency and coherence.

O'Brien, N. L. and D. H. Burn. 2018. A Nonstationary Peaks-Over-Threshold Approach for Modelling Daily Precipitation in Coastal British Columbia. *Canadian Water Resources Journal*, DOI: 10.1080/07011784.2018.1455538

Summary

The estimation of extreme precipitation events is a topic of growing interest and concern, particularly in highly urbanized areas. The Fifth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC) reported that land-surface air temperatures have increased and that more extreme precipitation events are expected in a warmer climate. Hydrometeorological data often display nonstationary characteristics in a changing climate, prompting numerous studies worldwide. This research proposes a peaks-over-threshold approach using both stationary and covariate-dependent nonstationary thresholds (bivariate and multivariate models). The GP distribution is fit to threshold exceedance data, which accounts for potential nonstationarity in the variability of the extreme events. The analysis herein focuses on coastal British Columbia (BC), which includes the Greater Vancouver Regional District (Metro Vancouver), due to the potential impacts of climate change in the region. Results from quantile estimates and an uncertainty analysis indicate that in the winter and summer months, there is stronger evidence of stationarity in 50-year quantiles. A trend analysis is also performed on the magnitude and frequency of POT events for winter and summer for two time periods (1976-2014 and 1986-2014). Results of this analysis indicate that globally significant trends are found in the threshold exceedance and frequency of peak event data.

3.1 Introduction

Intensifications of the magnitude and/or frequency of severe precipitation events could result in more extensive and frequent flooding and can have far-reaching socioeconomic consequences in highly populated or urbanized areas. For example, the widespread 2013 Alberta, Canada floods affected approximately 100,000 residents, resulted in four deaths, and caused US$5.7 million in damages (EM-DAT, 2016). The Fifth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC) states that global land-surface air temperatures have increased and that increases in extreme precipitation are consistent with a warmer climate (Hartmann *et al*., 2013). Hartmann *et al*. (2013) note that there is convincing evidence of increases in the frequency and intensity of heavy precipitation in North America. Therefore, the potential for nonstationarity in the magnitude and frequency of extreme precipitation events is of paramount concern. Additionally, with the assertion by Milly *et al*. (2008) that stationarity is dead, the time-independence of the statistics of historical rainfall records is brought into question.

Hydrometeorological quantile estimates for low-frequency, high-magnitude events are essential for civil infrastructure design. For instance, several Canadian provinces have imposed guidelines requiring that stormwater infrastructure adequately contain/convey runoff from a 100-year rainfall event (Stephens *et al*., 2002; Ontario Ministry of the Environment [OME], 2008). Standard statistical techniques assume data to be independently and identically distributed (IID), meaning the data arise from the same statistical probability distribution (Katz *et al*., 2002). This assumption is not valid in the presence of nonstationarity thus establishing a need for covariate-dependent approaches in a changing climate. Estimating nonstationary hydrometeorological extremes has henceforth become a topic which has gained a great deal of interest (Kharin and Zwiers, 2005; Nadarajah, 2005; El Adlouni *et al*., 2007; Hundecha *et al*., 2008; Sugahara *et al*., 2009; Rajagopalan, 2010; Villarini *et al*., 2010; Beguería *et al*., 2011; Burn *et al*., 2011; Ouarda and El-Adlouni, 2011; Villarini *et al*., 2011; Gilroy and McCuen, 2012; Roth *et al*., 2012; Tramblay *et al*., 2013; Li and Tan, 2015).

The extremes within hydrological data can be characterized in a variety of manners. Two commonly used techniques include the block maxima series and the partial duration series (peaks-over-threshold). The block maxima series is comprised of the largest peak value of a data set in a predetermined block length, typically selected to be 1-year and resulting in the annual maximum series (Katz *et al*., 2002). This research adopts the POT technique, in which all values that fall above a suitably high threshold are retained for analysis (Coles, 2001). In addition to the peak annual data, this method facilitates the inclusion of supplementary extreme events in wet years. The timing of

exceedances of a sufficiently high threshold are frequently modelled using a homogenous Poisson process, while excesses over the threshold are commonly modelled using the GP distribution, typically referred to as the Poisson-GP model (Khaliq *et al*., 2006). The GP distribution parameters may, however, be characterized by temporal covariates (Renard *et al*., 2006; Sugahara *et al*., 2009; Kyselý *et al*., 2010; Beguería *et al*., 2011; Roth *et al*, 2012; Tramblay *et al*., 2013; Sungwook *et al*., 2016) including the use of a nonhomogeneous Poisson process with the inclusion of a covariate-dependent intensity parameter (Katz, 2010; Beguería *et al*., 2011). A number of studies have focused on modelling nonstationarity in the selected threshold (Kyselý *et al*., 2010; Northrop and Jonathan, 2011; Roth *et al*., 2012; Jonathan *et al*., 2013a; Jonathan *et al*., 2013b; Jonathan *et al*., 2014; Roth *et al*., 2014), but to the authors' knowledge, none has developed a goodness-of-fit technique for selecting the most appropriate stationary or nonstationary threshold/GP model combination. Furthermore, this study addresses the confounding issue of time-dependence and large-scale ocean-atmosphere phenomena in nonstationary threshold selection.

The purpose of this study is to assess nonstationarity in both threshold models and GP distribution parameters to determine those models that provide superior goodness-of-fit and less uncertainty for the purpose of quantile estimation. Through this POT analysis a combination of bivariate and multivariate (linear and quadratic) covariate-dependent threshold models are evaluated. This research focuses on coastal British Columbia (BC), Canada. Extreme precipitation is associated with numerous hazards in the area, including flooding and landslides (Walker *et al*., 2008). With approximately 75% of BC's population living in coastal BC, primarily in Metro Vancouver (Walker *et al*., 2008), research of this nature is required to determine the magnitude and potential changes in extreme precipitation events that may be expected in one of Canada's most populated regions.

The Rocky Mountains act as a barrier to arctic air masses, giving rise to mild seasonal temperatures in coastal BC (Hare and Thomas, 1979). Although the area is one of the highest seismic activity zones in Canada (Chang *et al*., 2012), extreme weather events are the foremost hazard, affecting British Columbians more than any other climate hazard (Walker *et al*., 2008; Moore *et al*., 2010). The climate in southern BC is influenced by the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO), both of which are large-scale ocean-atmosphere phenomena (Walker *et al*., 2008). Several indices have been used to describe ENSO; the Niño 3.4 index is used herein. The Niño 3.4 index is a measure of the area average sea surface temperature (SST) from 5°S - 5°N and 170°E - 120°W (Rayner *et al*., 2003). The PDO Index is the leading principal

component of the North Pacific monthly sea surface variability (poleward of 20°N) (Mantua *et al*., 1997).

3.2 Methodology

A POT approach is developed for the analysis of hydrometeorological data using stationary and nonstationary thresholds in which the resulting threshold exceedance data are modelled using the GP distribution. Trend detection is carried out on stationary threshold exceedance data using the Mann-Kendall nonparametric trend test (Mann, 1945; Kendall, 1975) on total daily precipitation and rainfall frequency.

3.2.1 Climate Oscillation Selection

Due to the potential influence of both ENSO and PDO on coastal BC's climate, the correlation between both of these climate indices and the precipitation data is assessed (Gershunov and Barnett, 1998; Walker *et al*., 2008). The Niño 3.4 Index is obtained from the National Oceanic & Atmospheric Administration – Earth System Research Library (NOAA-ESRL) Physical Sciences Division (https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.anom.data), while the PDO Index is retrieved from NOAA National Climate Data Center (http://www.ncdc.noaa.gov/teleconnections/pdo/). An analysis of the correlation between total monthly precipitation at each observation station and the Niño 3.4 and PDO indices is carried out through a cross-correlation analysis.

3.2.2 Generalized Pareto Distribution Model

The most prominent benefit of the POT approach over the block maxima approach is that more data are retained for analysis as opposed to using only one peak value in a given block length. The Balkema-de Haan-Pickands theorem states that for a series of IID data, excesses over a sufficiently large threshold can be approximated using the GP distribution (Balkema and de Hann, 1974; Pickands, 1975). Also, in the POT framework it can be demonstrated that the occurrence of exceedances over a high threshold (u) follow a Poisson distribution with rate parameter λ.

For a sequence of IID random variables $x_1, x_2, \dots, x_n$, where $x_i > u$, the cumulative distribution function, $F(x)$, of $x_i - u$ for a suitably large u is given by (Coles, 2001):

$$F(x) = 1 - \left(1 + \frac{\kappa}{\sigma}x\right)^{-1/\kappa} \qquad for\ \sigma > 0, x > 0, 1 + \kappa(x/\sigma) > 0 \tag{3-1}$$

where σ and κ are the scale and shape parameters, respectively. In the event that $\kappa = 0$, the distribution reduces to the exponential distribution, which is not employed in this analysis. The quantile function, Q_p of the GP distribution is required to calculate the *T*-year event, which is an event exceeded, on average, once every $T = 1/(1-p)$, where $0 < p < 1$. The quantile function is commonly expressed in terms of the scale and shape parameters, rate parameter, and *T*-year event (Coles, 2001):

$$Q_p = u + \left(\frac{\sigma}{\kappa}\right)[(\lambda T)^{\kappa} - 1] \quad for\ \kappa \neq 0 \tag{3-2}$$

In the event that data are nonstationary and therefore not IID, the GP distribution can be expressed in terms of covariates. If distributional parameters are, for example, a function of time, the cumulative distribution function of the GP distribution may be expressed as (Coles, 2001):

$$F(x_t) = 1 - \left(1 + \frac{\kappa(t)}{\sigma(t)} x_t\right)^{-1/\kappa(t)} \tag{3-3}$$

where x_t are the time-dependent data; $\sigma(t)$ and $\kappa(t)$ are the time-varying scale and shape parameters, respectively, and are expressed most generally by (O'Brien and Burn, 2014):

$$\begin{aligned} \sigma(t) &= \sigma_0 + \sigma_1 t + \sigma_2 t^2 + \cdots + \sigma_m t^m \\ \kappa(t) &= \kappa_0 + \kappa_1 t + \kappa_2 t^2 + \cdots + \kappa_m t^m \end{aligned} \tag{3-4}$$

when $m = 1$, Equation (3-4) reduces to a linear model, whereas, if $m = 2$, Equation (3-4) assumes a quadratic form. Evidence of a constant shape parameter has been found in various hydrometeorological studies and this assumption is used hereafter (Vinnikov and Robock, 2002; Kharin and Zwiers, 2005; Kyselý *et al*., 2010; Kay and Jones, 2012).

If nonstationarity is present in the mean of the original data, the use of a stationary threshold will result in a nonhomogeneous Poisson process for the excesses. This inhomogeneity can be modelled using a nonstationary rate parameter (Katz, 2010; Beguería *et al*., 2011); however the use of a nonstationary threshold may result in a homogenous Poisson process. The latter approach is used herein to model potential changes in the central tendency of the data.

3.2.3 Threshold Selection

The distribution of exceedances over a threshold asymptotically approaches the GP distribution, given that a suitably high threshold is chosen. Therefore, threshold selection must be done with care to avoid undue bias or, in the case of too high a threshold, high variance (Coles, 2001). To ascertain whether change-points exist in the data, Pettitt's test (Pettitt, 1979) is applied prior to both stationary and nonstationary threshold selection. The following sections discuss the stationary and nonstationary thresholds selection criteria that are applied for this research.

3.2.4 Stationary Threshold

Stationary thresholds are selected through the use of several exploratory plots (Lang *et al*., 1999; Mailhot *et al*., 2013; Osman *et al*., 2013; Roth *et al*., 2016; Sungwook *et al*., 2016). Assessing the validity of the homogeneous Poisson process assumption is carried out through the dispersion index plot (Cunnane, 1979), where an optimally selected threshold has a dispersion index equal to 1 (Cunnane, 1979; Lang *et al*., 1999). The mean residual life plot depicts the mean excess above a given threshold versus a range of threshold values. The GP distribution provides a suitable fit to excesses over the selected threshold when linearity is apparent in the mean residual life plot. Complimentary threshold choice (TC) plots involve fitting the GP distribution to a range of threshold exceedances and assessing the stability of the resulting distributional parameters. These include the scale and shape parameters for the GP distribution (Coles, 2001). An overview of the threshold selection process is provided in Figure 3-1 for the Mission West Abbey summer season data. Figure 3-1(a) demonstrates that an appropriate threshold range falls somewhere between 14 and 17 mm. A final threshold of 17 mm was selected as it provides slightly more linearity in Figure 3-1(b), which is the mean residual life plot. The modified scale and shape parameters are also suitably stable after this point, and are depicted in Figure 3-1(c) and (d), respectively.

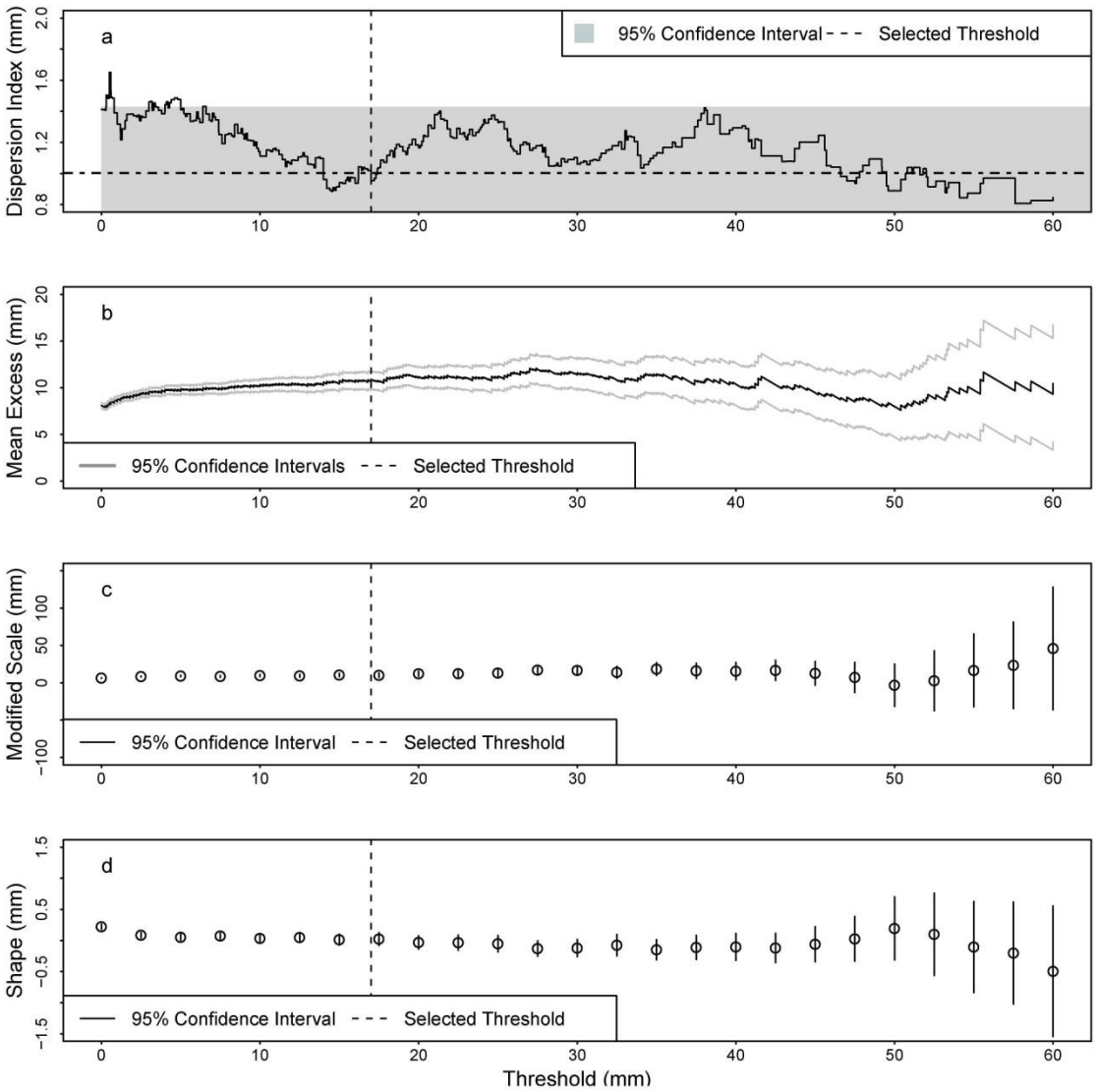

Figure 3-1: Stationary threshold selection plots for the Mission West Abbey station summer season data; (a) dispersion index plot (b) mean residual life plot (c) modified scale parameter plot (d) shape parameter plot. - - - Indicates a dispersion index of 1.

3.2.5 Nonstationary Threshold

Quantile regression is used to determine nonstationary thresholds, which relies on the fact that sample quantiles can be determined by the solution of a linear optimization problem. Given a set of observations (x_i), the τ^{th} regression quantile is the solution of (Koenker and Basset, 1978; Koenker, 2005):

$$\min_{\beta \in \mathbb{R}} \sum_i \rho_\tau(x_i - \beta) \tag{3-5}$$

where:

$$\rho_\tau(u) = \begin{cases} u(\tau - 1), & u < 0 \\ u\tau, & u \geq 0 \end{cases}$$

In the event that the covariate in question varies linearly with time (t), for example, the τ^{th} conditional quantile function becomes:

$$Q_x(\tau|t) = \beta_0(\tau) + t\beta_1(\tau) \tag{3-6}$$

which can be estimated by:

$$\min_{\beta_0, \beta_1 \in \mathbb{R}} \sum_t \rho_\tau(x_t - \beta_0 - t\beta_1) \tag{3-7}$$

where x_t is a set of time-dependent observations; and β_0 and β_1 are the intercept and the slope components of the regression quantiles, respectively. For a more detailed explanation of quantile regression, please refer to Koenker (2005).

This research follows the methodology of Roth *et al*. (2014), which provides a basis for time-dependent threshold selection. That is, for a selected time-dependent threshold, the TC plot of the scale parameter for the selected threshold (τ_0) should be approximately constant and the mean residual life plot should be approximately linear for all $\tau > \tau_0$. In addition to the mean residual life

plot and TC plots, dispersion index plots were examined for all $\tau > \tau_0$ to assess the potential of a homogeneous Poisson process for threshold exceedances.

Four bivariate quantile regression models are used for nonstationary threshold modelling along with two multivariate models, all of which incorporate a climate oscillation index and/or time as covariates.

3.2.6 Trend Detection

There exist comparable trend detection tests in terms of their power of trend detection (Yue *et al.*, 2002) yet the Mann-Kendall test is widely applied and is therefore used for this research (Mann, 1945; Kendall, 1975). The presence of positive/negative serial correlation in hydrometric data can increase/decrease the likelihood of finding a significant trend (von Storch, 1995). Although the data used for this analysis were declustered to ensure independence of events, any potential residual serial correlation was accounted for using the nonparametric BBS technique (Kundzewicz and Robson, 2000). This approach is carried out by resampling data in blocks, the length of which is selected based on the number of contiguous lags of significant autocorrelation. After resampling is carried out a large number of times (2000 is implemented for this research), the empirical distribution of the results can be used to assess the significance of trends in the data set. The benefit in using the BBS technique is that the autocorrelation structure of the data is preserved, whereas most techniques require the adoption of a lag-one autoregressive model [AR(1)]. For more information regarding the BBS technique, please refer to Khaliq *et al*. (2009) or Önöz and Bayazit (2012).

To determine whether significant (local) trends found using the Mann-Kendall test have occurred by chance, the field (global) significance is determined. Field significance is employed to determine if an identified statistically significant trend occurred by chance. Field significance is determined using a group block bootstrapping approach (GBBS) which assesses the significance of increasing and decreasing trends separately (Yue *et al.*, 2003; Khaliq *et al.*, 2009). The goal of this technique is to determine the number of significant trends that occur by chance for a selected number of stations while maintaining the serial structure of the original data. Vector resampling of years is carried out in blocks that preserve the serial structure of the original data. Each station having data for the block bootstrapped year vector is added into a resampled data set and this continues until the desired number of station years are met (i.e., the resampled data set contains the same number of station-years of data as the original data set). This process is repeated for a large number of times (1000 herein),

where trend detection is carried out for each resampled data set. The empirical distribution of trends and subsequent p-value are then determined, which are then employed to determine field significance.

3.2.7 Parameter Estimation

The method of maximum likelihood (ML) is used for parameter estimation due to the ease of incorporation of covariates into distributional parameters. In the case of a stationary model, the log-likelihood function, $\ell(\vartheta)$, is expressed as (Coles, 2001):

$$\ell(\vartheta) = logL(\vartheta) = \sum_{i=1}^{n} log\, f(x_i|\vartheta) \tag{3-8}$$

where ϑ is a vector of parameters; $f(x_i|\vartheta)$ is the probability density function; and x_i are the time-independent series of data. In the event that parameter estimation is to be carried out in a nonstationary (time-dependent) context, the log-likelihood function is denoted by (O'Brien and Burn, 2014):

$$\ell(\vartheta_t) = \sum_t \ log\, f\big(x_t\big|\vartheta(t)\big) \tag{3-9}$$

where ϑ_t is a vector of nonstationary parameters; $f(x_t|\,\vartheta(t))$ is the density function; x_t are the time-dependent at-site data; and $\vartheta(t)$ represents the nonstationary distribution parameters. Due to the complex nature of the log-likelihood function, parameter estimates are commonly determined numerically through the use of non-linear optimization techniques; the Nelder-Mead simplex algorithm is used for this analysis (Nelder and Mead, 1965).

3.2.8 Model Selection

Figure 3-2 outlines the two stage approach that is applied in choosing the most suitable model for each data set. The first stage involves calculating AIC weights that are used as goodness-of-fit criteria for the (non)stationary GP model selection. After the most suitable GP model is selected for each threshold, the best fitting threshold models are selected by means of the root mean square error (RMSE).

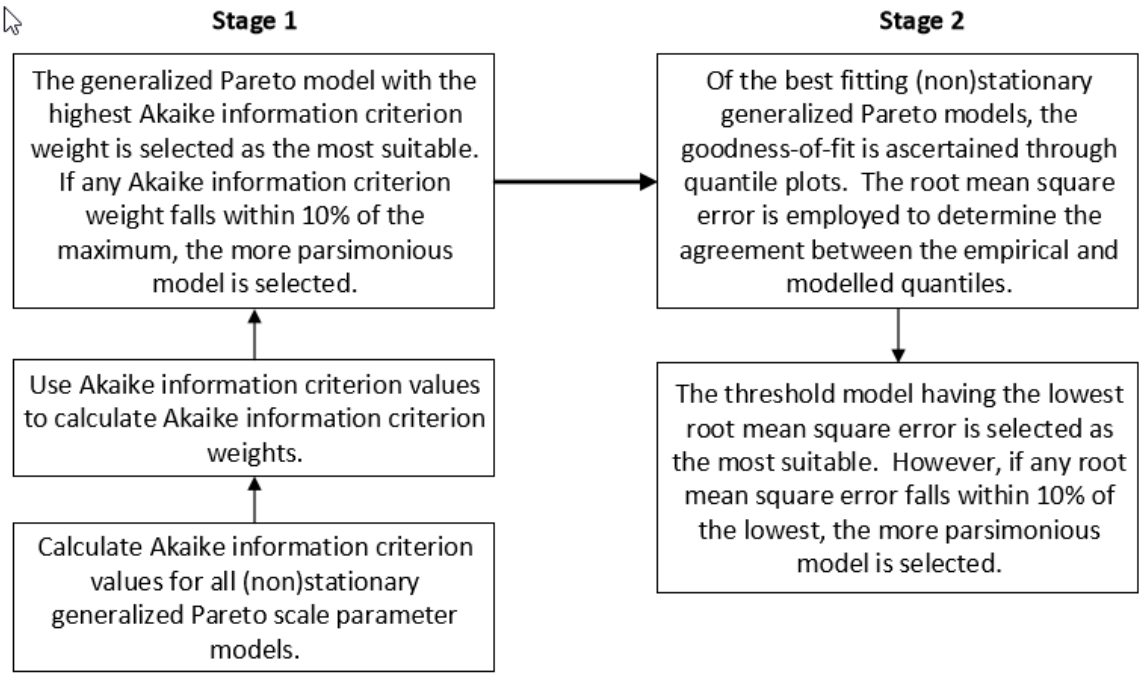

Figure 3-2: Schematic of model selection process.

Akaike weights are initially employed as a goodness-of-fit measure to facilitate the selection of the stationary or nonstationary scale parameter models for each threshold. The AIC is given by (Akaike, 1974):

$$AIC = -2[\ell(\vartheta)] + 2K \tag{3-10}$$

where K is the number of parameters. Using raw AIC values, the difference between the minimum AIC value and those of the other models is determined (Burnham and Anderson, 2002):

$$\Delta_i(AIC) = AIC_i - minAIC \tag{3-11}$$

where $\Delta_j(AIC)$ is the difference in AIC with respect to the minimum AIC; AIC_j is the raw AIC for each model; and $minAIC$ is the minimum AIC value. Akaike weights are then calculated by (Burnham and Anderson, 2002):

$$w_j(AIC) = \frac{exp\{-0.5\Delta_j(AIC)\}}{\sum_{j=1}^{M} exp\{-0.5\Delta_m(AIC)\}} \tag{3-12}$$

where $w_j(AIC)$ is the AIC weight for model j and $\sum_{j=1}^{M} w_j(AIC) = 1$. The GP distribution model with the largest AIC weight is selected as the most suitable. If any AIC weight falls within 10% of the maximum, the more parsimonious model is selected. Estimates of the GP distribution model parameters are determined using the 'ismev' package in R (Heffernan and Stephenson, 2009; R Core Team, 2016). The 'gp.fit' function in this package allows for the estimation of stationary and nonstationary GP distribution parameters through ML estimation, with the option of adding a trend component into selected parameters.

Having selected the best fitting GP model for each threshold (stationary or nonstationary), the goodness-of-fit of the threshold models is ascertained through a quantile plot (Q-Q) comparison. In the event of a nonstationary scale parameter and/or nonstationary threshold, a residual quantile plot is produced by transforming data to a standard exponential distribution (Coles, 2001), where (for example) time-dependent parameters are included:

$$\tilde{x}_t = \frac{1}{\kappa} ln\left\{1 + \kappa\left(\frac{x_t - u_t}{\sigma(t)}\right)\right\} \tag{3-13}$$

where $\tilde{x}_t$ is the transformed exponential time-dependent data; u_t is the nonstationary threshold at time t. The selection of the best fitting threshold was determined using the RMSE, through which the agreement between the empirical and modelled quantiles is assessed. The threshold model having the lowest RMSE is selected, although, if any threshold model falls within 10% difference of the lowest RMSE, the model with the smaller number of parameters is selected to preserve model parsimony.

3.2.9 Uncertainty

A balanced resampling approach is implemented for the calculation of quantile confidence intervals (CI). This technique involves first creating a large number of copies (999 is used herein) of the original sample data and concatenating the copies. This concatenated vector of data is then randomly

permutated and broken back into the original number of copies. The resulting 999 bootstrapped resamples can then be used to estimate the uncertainty present in the quantile estimates (Burn, 2003).

In the event that nonstationarity is present in the scale parameter of the original data set of interest, data are transformed to remove the covariate-dependence before resampling. Given that the GP distribution is applied for this research and using time as an example, nonstationary data were transformed to IID data using the following (Coles, 2001):

$$\epsilon_t = [1/\kappa] ln\{1 + \kappa[(x_t - u_t)/\sigma(t)]\} \tag{3-14}$$

where ϵ_t is the transformed residual series at time t; x_t is the original data series at time t; and u_t is the time-dependent threshold. Data are transformed before bootstrapping and final samples are calculated using the inverse of Equation (3-14).

3.3 Application

3.3.1 Case Study Area

Total daily precipitation data, retrieved from the National Climate Data Archive of Environment and Climate Change Canada, are used for this research. Data from 30 gauges throughout coastal BC are used for the implementation of a POT analysis. Before threshold selection is carried out, data are separated into winter season events, spanning from mid-October to mid-April, and summer season events, encompassing the remainder of the year. Winter storms in BC are dominated by midlatitude cyclonic activity, resulting in substantial precipitation amounts. The summer climate is characterized by a subtropical high pressure system that results in less frequent and intense storms (Walker *et al.*, 2008). Therefore, seasonal partitioning of the data is necessary to reflect the different storm generating mechanisms.

Figure 3-3 shows the locations of 30 observation stations used in the analysis, six of which are located in Metro Vancouver. Table 3-1 provides a spatial summary of the station characteristics in three geographical areas, which include the North Coast, the South Coast, and Vancouver Island. The elevations and locations of the meteorological stations have a discernible impact on the mean annual precipitation recorded at these locations. Substantial variability in mean annual precipitation is evident throughout coastal BC. Windward areas and those at higher elevations receive greater annual

precipitation amounts. Also, the rain shadow of Vancouver Island produces drier conditions in the south coast interior (Moore *et al*., 2010).

The precipitation data used for this analysis were initially screened for obvious transcription errors and uncharacteristically low or high values. Observation stations for this analysis were retained for analysis if they have no more than four consecutive years of missing observations. All stations with daily data west of the mainland Coast Mountains meeting these criteria were included in the analysis. Table 3-1 provides a summary of the percentage of missing data at each of the 30 retained meteorological stations. All the stations, aside from North Pender Island, have less than 20% missing observations from the period of record; 27 of the stations have less than 10% missing data. It is acknowledged that bias may be present in the precipitation measurements due to variations in wind speed and differences in instrumentation. Furthermore, the current analysis employs a two day separation time between maxima as a means of retaining peak precipitation from individual events (Coles, 1993; Caires *et al*., 2006; Kyselý and Beranová, 2009; Roth *et al*., 2012). All data were also found to be free of step-changes prior to parameter estimation.

Table 3-1: Summary of stations used for analysis.

Station Name	Station ID	Start Year	End Year	Latitude (°N)	Longitude (°W)	Elevation (m)	Mean Annual Precipitation (mm)	Percentage of Missing Data (%)
			North Coast					
(N1) Kitimat 2	1064321	1966	2014	49.34	121.76	16.8	2629	3.5
(N2) Kitimat Townsite	1064320	1954	2014	54.05	128.63	98.0	2120	4.2
(N3) Terrace PCC	1068131	1968	2014	54.50	128.62	67.0	1099	3.4
			South Coast					
(S1) Burnaby Simon Fraser U	1101158	1965	2014	49.28	122.92	365.8	1732	8.7
(S2) Coquitlam Como Lake Ave.	1101889	1972	2014	49.27	122.87	160.0	1626	13.4
(S3) Delta Tsawwassen Beach	1102425	1971	2014	49.01	123.09	2.4	902	1.9
(S4) Gibsons Gower Point	1043152	1961	2014	49.39	123.54	34.0	1313	2.4
(S5) Mission West Abbey	1105192	1962	2014	49.15	122.27	197.0	1823	2.3
(S6) N Vancouver Wharves	1105669	1962	2014	49.31	123.12	7.0	1579	8.7
(S7) Richmond Nature Park	1106PF7	1977	2014	49.17	123.09	3.0	1171	5.0
(S8) Sumas Canal	1107785	1957	2014	49.11	122.11	9.0	1544	11.3
(S9) Vancouver Harbour CS	1108446	1925	2014	49.30	123.12	2.5	1377	9.8
			Vancouver Island					
(V1) Alberni Robertson Creek	1030230	1961	2014	49.34	124.98	73.8	2068	2.7
(V2) Courtenay Grantham	1021988	1960	2014	49.76	125.00	81.0	1285	9.6
(V3) Cowichan Lake Forestry	1012040	1949	2014	48.82	124.13	176.8	2126	1.6
(V4) Gabriola Island	1023042	1967	2014	49.15	123.73	46.0	850	8.1
(V5) Galiano North	10130MN	1975	2014	48.99	123.57	6.0	859	6.8
(V6) Gold River Townsite	1033232	1975	2014	49.78	126.06	119.0	2524	8.0
(V7) Lake Cowichan	1012055	1960	2014	48.83	124.05	171.0	1747	9.4
(V8) Metchosin	1015105	1968	2014	48.37	123.56	140.0	967	4.3
(V9) North Pender Island	1015638	1972	2014	48.76	123.29	98.0	666	21.5
(V10) Port Alice	1036240	1924	2014	50.39	127.46	21.0	3117	1.4
(V11) Qualicum R Fish Research	1026565	1962	2014	49.39	124.62	7.6	1268	2.0
(V12) Quatsino	1036570	1900	2014	50.53	127.65	3.4	2534	4.4
(V13) Quinsam River Hatchery	1026639	1975	2014	50.02	125.30	45.7	1555	3.0
(V14) Saanichton CDA	1016940	1914	2014	48.62	123.42	61.0	858	0.7
(V15) Saltspring ST Mary's L	1016995	1976	2014	48.89	123.55	45.7	956	1.2
(V16) Shawnigan Lake	1017230	1911	2014	48.65	123.63	159.0	1216	1.0
(V17) Tofino A	1038205	1959	2014	49.08	125.77	24.5	3236	0.8
(V18) Ucluelet Kennedy Camp	1038332	1964	2014	48.95	125.53	40.0	3309	2.3

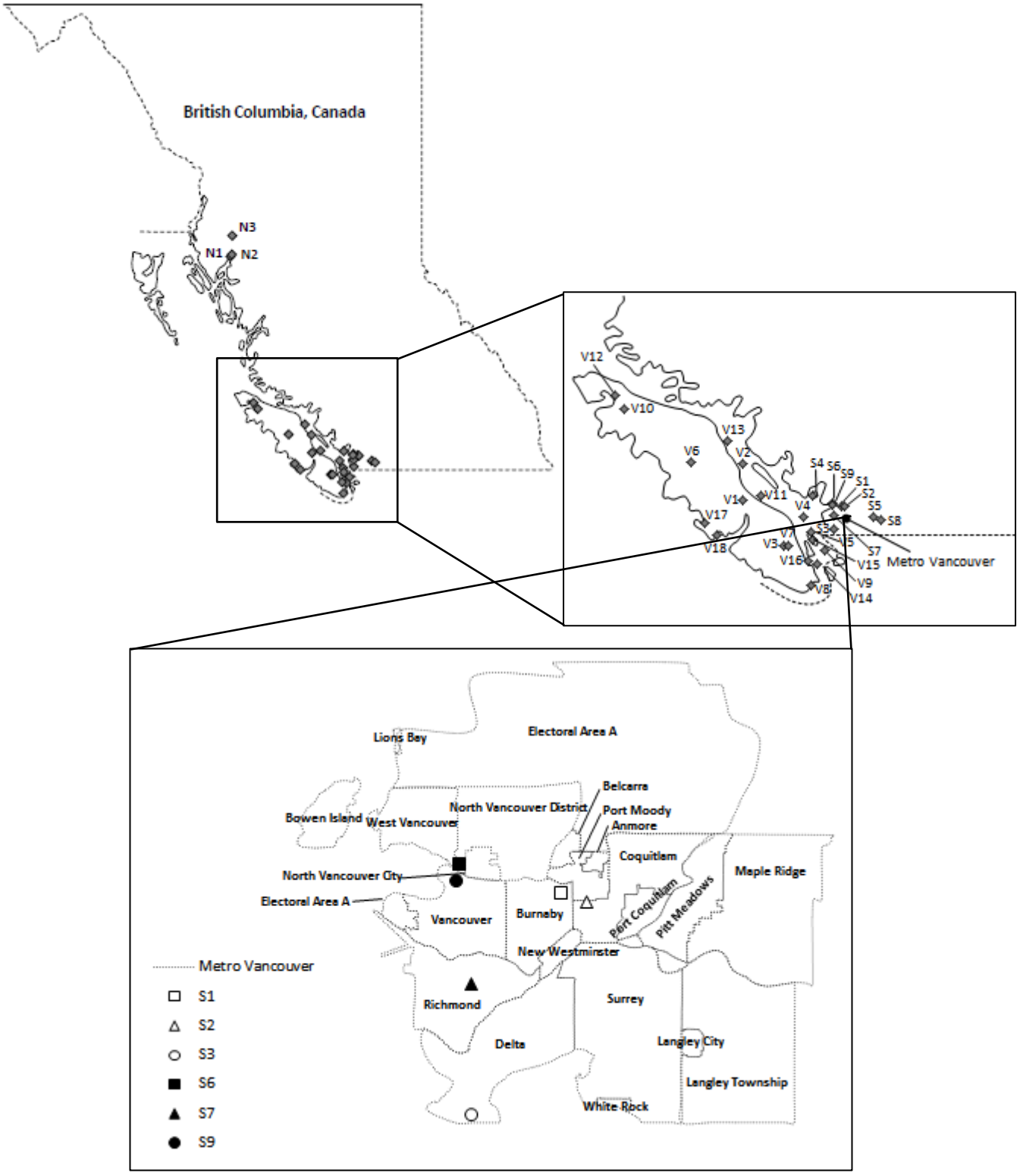

Figure 3-3: Location of 30 meteorological stations used for the analysis.

3.4 Results

3.4.1 Threshold Selection and Models

Correlations between climate indices and monthly precipitation amounts were assessed for positive and negative lags of cross-correlation function (CCF) plots at the 5% significance level for the Niño 3.4 and PDO indices. It is determined that the PDO index is more strongly correlated with the monthly precipitation amounts in both the summer (73% of cases for PDO/27% for Niño 3.4) and winter seasons (76% of cases for PDO/24% for Niño 3.4). For this reason, the PDO index and/or time are used as covariates for the univariate and bivariate thresholds and GP models used for this analysis. Table 3-2 outlines the nonstationary threshold models employed.

Table 3-2: Quantile regression threshold models.

$Q_x(\tau\|t) = \nu_0(\tau) + t \cdot \nu_{t1}(\tau)$	(TH_1)
$Q_x(\tau\|PDO_t) = \nu_0(\tau) + PDO_t \cdot \nu_{PDO1}(\tau)$	(TH_2)
$Q_x(\tau\|t, PDO_t) = \nu_0(\tau) + t \cdot \nu_{t1}(\tau) + PDO_t \cdot \nu_{PDO1}(\tau)$	(TH_3)
$Q_x(\tau\|t) = \nu_0(\tau) + t \cdot \nu_{t1}(\tau) + t^2 \cdot \nu_{t2}(\tau)$	(TH_4)
$Q_x(\tau\|PDO_t) = \nu_0(\tau) + PDO_t \cdot \nu_{PDO1}(\tau) + PDO_t^2 \cdot \nu_{PDO2}(\tau)$	(TH_5)
$Q_x(\tau\|t, PDO_t) = \nu_0(\tau) + t \cdot \nu_{t1}(\tau) + t^2 \cdot \nu_{t2}(\tau) + PDO_t \cdot \nu_{PDO1}(\tau) + PDO_t^2 \cdot \nu_{PDO2}(\tau)$	(TH_6)

3.4.2 GP Distribution – Model Selection

The GP distribution is fit to threshold exceedance data established through the use of stationary and nonstationary thresholds; the shape parameter was kept constant in all instances. The stationary and nonstationary GP distribution models used for this analysis are outlined in Table 3-3. Using the log-likelihood estimates provided by the 'gp.fit' function, AIC weights are calculated for the stationary and nonstationary threshold exceedance data. Results from the Mission West Abbey observation station are provided in Table 3-4 for the summer season data as an example of the model selection process and are typical of all results. The GP scale parameter model with the highest weight is selected as the most appropriate model for all thresholds except for TH_2 and TH_6, where TH_2 is the linear, time-dependent threshold and TH_6 is the quadratic time- and PDO-dependent threshold. In these two instances, there was a tie for the highest weight; therefore, the more parsimonious model was selected for both threshold models.

Table 3-3: POT data and GP distribution quantile functions.

Stationary GP Dist. Quantile Function	$Q_p = u + \left(\frac{\sigma_0}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{St})
Nonstationary GP Dist. Quantile Function	$Q_p(t) = u/u_{cov} + \left(\frac{\sigma_0 + \sigma_{t1} t}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{NonSt1})
	$Q_p(PD0_t) = u/u_{cov} + \left(\frac{\sigma_0 + \sigma_{PDO1} PDO_t}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{NonSt2})
	$Q_p(t, PDO_t) = u/u_{cov} + \left(\frac{\sigma_0 + \sigma_{t1} t + \sigma_{PDO1} PDO_t}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{NonSt3})
	$Q_p(t) = u/u_{cov} + \left(\frac{\sigma_0 + \sigma_{t1} t + \sigma_{t2} t^2}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{NonSt4})
	$Q_p(PD0_t) = u/u_{cov} + \left(\frac{\sigma_0 + \sigma_{PDO1} PDO_t + \sigma_{PDO2} PDO_t^2}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{NonSt5})
	$Q_p(t, PDO_t) = u/u_{cov} + \left(\frac{\sigma_0 + \sigma_{t1} t + \sigma_{t2} t^2 + \sigma_{PDO1} PDO_t + \sigma_{PDO2} PDO_t^2}{\kappa}\right)[(\lambda T)^{\kappa} - 1]$	(GP_{NonSt6})

*u/u_{Cov} denotes a stationary threshold (u) or one of the six nonstationary thresholds (u_{Cov}) models in Table 3-2.

Table 3-4: AIC weights for the Mission West Abbey observation station – summer season.

Stationary/ Nonstationary GP Dist.	Stationary Threshold (u)	Nonstationary Threshold Models (u_{Cov})					
		TH_1	TH_2	TH_3	TH_4	TH_5	TH_6
GP_{St}	0.005	0.009	0.039	0.040	0.032	0.025	0.004
GP_{NonSt1}	0.012	0.011	0.033	0.046	0.041	0.020	0.006
GP_{NonSt2}	0.200	***0.297***	***0.291***	***0.250***	***0.259***	***0.352***	***0.281***
GP_{NonSt3}	0.246	0.221	0.291	0.201	0.232	0.201	0.281
GP_{NonSt4}	0.030	0.053	0.091	0.187	0.165	0.071	0.031
GP_{NonSt5}	***0.374***	0.344	0.184	0.167	0.162	0.235	0.269
GP_{NonSt6}	0.133	0.064	0.072	0.109	0.110	0.097	0.128

- Bold and italicized cells represent the distribution with the best fit (highest AIC weight).

The best fitting threshold model was then determined using the model that provided the highest AIC weight. This is carried out through the comparison of quantile or residual quantile plots (for nonstationary models), where the agreement between the model and empirical quantiles is measured using the RMSE. An example is provided in Figure 3-4, showing the results for the summer season quantile plots at the Mission West Abbey observation station. Figure 3-4(c) was chosen as the most suitable model as it is the most parsimonious within 10% difference of the minimum RMSE of all the threshold models. Tables 3-5 and 3-6 outline the final selection of models, the associated parameter

estimates of the GP distribution models, and the applicable stationary or nonstationary threshold parameters.

3.4.3 Trend Detection Summary

A summary of the trend analysis of the magnitude and frequency of the POT data is presented in Figures 3-5 through 3-8. Two analysis periods are examined (39- and 29-years) for all 30 stations and for the winter and summer seasons at the 5% and 10% significance levels.

Figure 3-5 indicates that the winter season threshold exceedance data are dominated by decreasing trends. Trends are observed at the 5% and 10% significance levels for threshold exceedances for both analysis periods at several sites in the North Coast and Vancouver Island regions. Although a number of statistically significant decreasing trends were identified, only those for the 1976-2014 (39-year) and 1986-2014 (29 year) period at the 10% significance level were determined to be field significant. A globally significant increasing trend is also identified in Figure 3-6 for the 1986-2014 (29-year) analysis period at the 5% and 10% significance levels at the Metchosin (V8) station only.

Results of the Mann-Kendall trend test for the summer season threshold exceedance and frequency data (Figures 3-7 and 3-8, respectively) indicate that there are a number of both increasing and decreasing trends, which are located primarily in the South Coast and Vancouver Island region. It should be noted, however, that globally significant trends were only found in the peak data (Figure 3-7), which include 39-year increasing trends at the 10% significance level and the 29-year decreasing trends at the 10% significance level. These results highlight the importance of trend identification in different analysis periods.

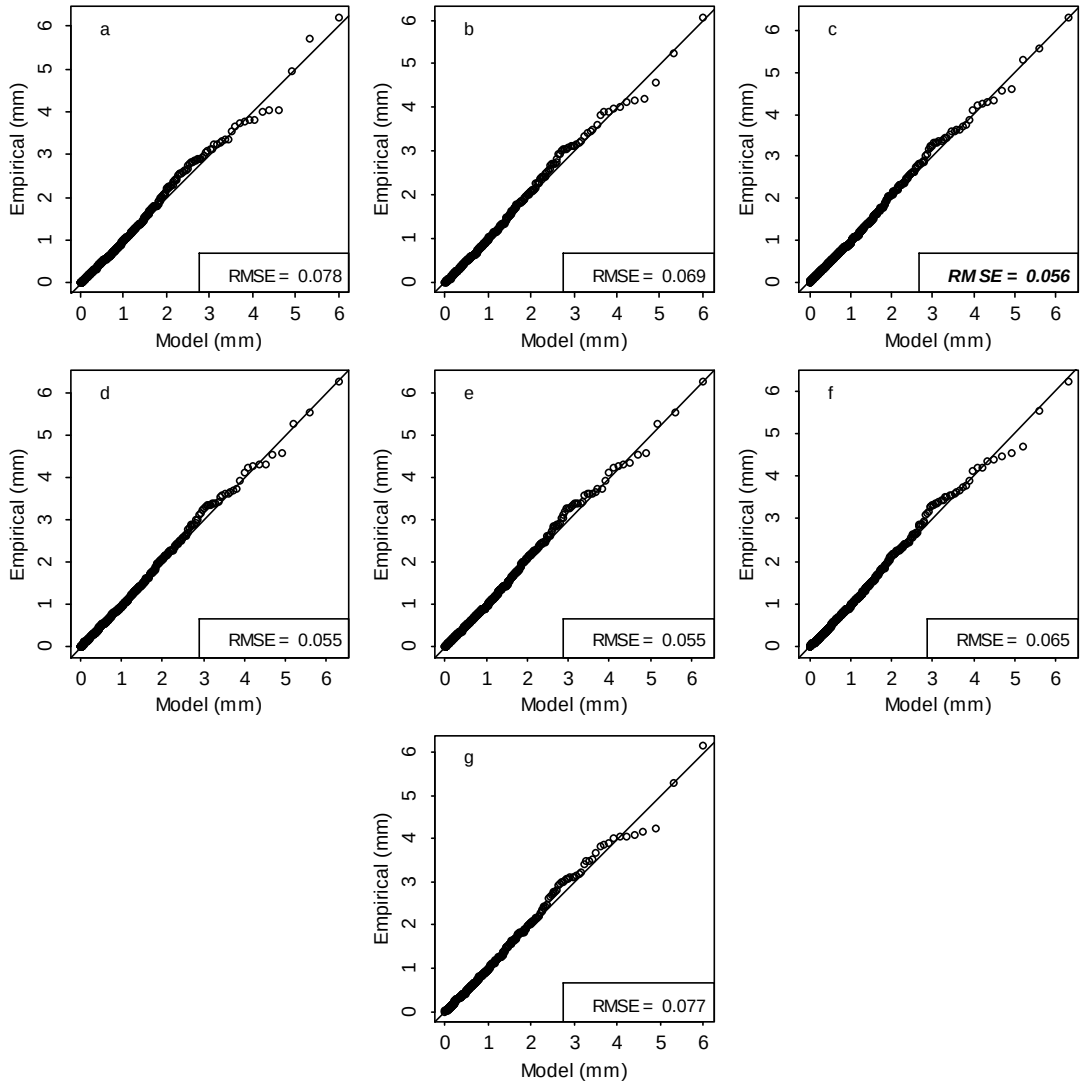

Figure 3-4: Comparison of residual quantile plots of above threshold data from the summer season at observation station Mission West Abbey; (a) stationary threshold and quadratic PDO-dependent scale parameter (b) Linear time-dependent threshold and linear PDO-dependent scale parameter (c) Linear PDO-dependent threshold and linear PDO-dependent scale parameter (d) Linear time and PDO-dependent threshold and linear PDO-dependent scale parameter (e) Quadratic time-dependent threshold and linear PDO-dependent scale parameter (f) Quadratic PDO-dependent threshold and linear PDO-dependent scale parameter (g) Quadratic time and PDO-dependent threshold and linear PDO-dependent scale parameter. Note: The RMSE for the selected model is bolded and italicized.

Table 3-5: Final threshold selection and model parameters for the winter season data.

	Threshold Model Parameter(s)					GP Dist. Model Parameters					
						Scale Parameter(s)					Shape Parameter
	ν_0	ν_{t1}	ν_{t2}	ν_{PDO1}	ν_{PDO2}	σ_0	σ_{t1}	σ_{t2}	σ_{PDO1}	σ_{PDO2}	κ
North Coast											
(N1) Kitimat 2	34.76	-0.054	-	-	-	20.85	0.4947	-0.0133	-	-	-0.034
(N2) Kitimat Townsite	29.19	-0.115	0.001	-0.790	0.580	16.50	-	-	-	-	0.023
(N3) Terrace PCC	10.00	-	-	-	-	12.90	-0.0582	-	-	-	0.041
South Coast											
(S1) Burnaby Simon Fraser U	24.19	-0.052	-	-1.547	-	14.78	-	-	-	-	0.017
(S2) Coquitlam Como Lake Ave.*	21.00	-	-	-	-	14.72	-	-	-	-	0.078
(S3) Delta Tsawwassen Beach*	9.90	-	-	-	-	8.29	-	-	-	-	-0.008
(S4) Gibsons Gower Point*	13.60	-	-	-	-	10.34	-	-	-	-	-0.085
(S5) Mission West Abbey*	23.39	-	-	-	-	12.84	-	-	-	-	0.081
(S6) N Vancouver Wharves*	22.00	-	-	-	-	13.53	-	-	-	-	0.018
(S7) Richmond Nature Park	14.01	-0.010	-	-	-	9.90	-	-	-	-	0.004
(S8) Sumas Canal	23.60	-	-	-	-	14.00	0.0614	-	-	-	-0.016
(S9) Vancouver Harbour CS	23.38	0.094	-0.001	-	-	9.80	0.1573	-0.0014	-	-	0.063
Vancouver Island											
(V1) Alberni Robertson Creek*	26.70	-	-	-	-	21.50	-	-	-	-	-0.090
(V2) Courtenay Grantham	22.61	-0.041	-	-	-	14.95	-	-	-	-	-0.043
(V3) Cowichan Lake Forestry*	33.20	-	-	-	-	20.85	-	-	-	-	-0.026
(V4) Gabriola Island	13.23	-0.010	-	-	-	9.66	-	-	-	-	-0.013
(V5) Galiano North	12.66	-	-	0.417	-	10.97	-	-	-	-	-0.059
(V6) Gold River Townsite	35.60	-	-	-	-	28.85	-0.2162	-	-	-	-0.040
(V7) Lake Cowichan	30.18	-	-	0.175	-	21.05	0.1540	-	-	-	-0.052
(V8) Metchosin	15.73	0.010	-	-1.503	-	7.63	0.5492	-0.0104	-1.5826	0.0625	0.075
(V9) North Pender Island	11.85	-	-	-0.090	-	10.55	-	-	-	-	-0.007
(V10) Port Alice	46.29	0.092	-	2.148	-	17.93	0.0747	-	1.3582	-	0.158
(V11) Qualicum R Fish Research*	21.10	-	-	-	-	14.16	-	-	-	-	0.069
(V12) Quatsino	30.72	-	-	1.096	-	16.90	-	-	-	-	-0.002
(V13) Quinsam River Hatchery	18.02	-	-	-0.615	-	16.29	-	-	-	-	-0.146
(V14) Saanichton CDA	15.87	0.023	-	-	-	8.40	0.0394	-	-	-	0.022
(V15) Saltspring ST Mary's L*	13.00	-	-	-	-	10.57	-	-	-	-	-0.030
(V16) Shawnigan Lake*	22.10	-	-	-	-	13.65	-	-	-	-	-0.019
(V17) Tofino A*	42.20	-	-	-	-	22.94	-	-	-	-	0.058
(V18) Ucluelet Kennedy Camp*	40.96	-	-	-	-	24.74	-	-	-	-	0.024

*denotes those sites having a stationary threshold and stationary GP dist. model.

Table 3-6: Final threshold selection and model parameters for the summer season data.

	Threshold Model Parameter(s)					GP Dist. Model Parameters					
						Scale Parameter(s)					Shape Parameter
	ν_0	ν_{t1}	ν_{t2}	ν_{PDO1}	ν_{PDO2}	σ_0	σ_{t1}	σ_{t2}	σ_{PDO1}	σ_{PDO2}	κ
North Coast											
(N1) Kitimat 2*	18.00	-	-	-	-	16.81	-	-	-	-	0.051
(N2) Kitimat Townsite	18.27	0.037	-	-	-	13.32	-	-	-	-	0.112
(N3) Terrace PCC*	10.40	-	-	-	-	7.40	-	-	-	-	0.158
South Coast											
(S1) Burnaby Simon Fraser U	12.97	0.024	-	0.713		14.18	-0.0697	-	-	-	-0.043
(S2) Coquitlam Como Lake Ave.	9.69	-	-	0.589	0.573	14.69	-	-	-0.5935	-0.9711	-0.144
(S3) Delta Tsawwassen Beach*	5.51	-	-	-	-	7.08	-	-	-	-	-0.091
(S4) Gibsons Gower Point	9.85	-0.026	-	-	-	9.19	-	-	-	-	-0.100
(S5) Mission West Abbey	14.46	-	-	0.116	-	12.65	-	-	1.1586	-	-0.047
(S6) N Vancouver Wharves	21.88	0.006	-	-0.182	-	12.69	-0.0903	-	-	-	-0.032
(S7) Richmond Nature Park*	6.80	-	-	-	-	9.09	-	-	-	-	-0.089
(S8) Sumas Canal	16.69	0.018	-	-	-	10.09	0.0728	-	-	-	0.024
(S9) Vancouver Harbour CS*	15.80	-	-	-	-	9.90	-	-	-	-	0.032
Vancouver Island											
(V1) Alberni Robertson Creek*	13.10	-	-	-	-	12.45	-	-	-	-	0.053
(V2) Courtenay Grantham	10.82	-	-	1.266		7.10	-	-	-	-	0.175
(V3) Cowichan Lake Forestry	12.94	0.008	-	-	-	11.86	-	-	-	-	0.048
(V4) Gabriola Island	5.80	-	-	-	-	7.05	-	-	0.268	-0.428	-0.034
(V5) Galiano North*	6.93	-	-	-	-	5.50	-	-	-	-	0.108
(V6) Gold River Townsite*	19.00	-	-	-	-	13.43	-	-	-	-	0.158
(V7) Lake Cowichan*	11.30	-	-	-	-	10.11	-	-	-	-	0.183
(V8) Metchosin	6.28	0.026	-	-	-	5.55	-	-	-	-	0.155
(V9) North Pender Island	4.84	0.027	-	-	-	5.18	-	-	-	-	0.037
(V10) Port Alice	27.38	-0.006	-	0.397	-	21.94	-	-	-	-	0.009
(V11) Qualicum R Fish Research	9.43	-	-	0.844	-	7.45	-	-	-	-	0.035
(V12) Quatsino	20.53	-0.032	-	0.508	-	11.66	-	-	0.425	0.714	0.087
(V13) Quinsam River Hatchery	10.59	-	-	-	-	7.48	-	-	-0.639	-	0.125
(V14) Saanichton CDA	7.92	0.019	-	-	-	5.45	-	-	-	-	0.117
(V15) Saltspring ST Mary's L	5.98	-	-	0.519	-	5.31	-	-	-	-	0.073
(V16) Shawnigan Lake	7.27	0.010	-	-	-	7.37	-0.017	-	-	-	0.081
(V17) Tofino A*	25.78	-	-	-	-	17.17	-	-	-	-	0.054
(V18) Ucluelet Kennedy Camp	29.02	-0.423	0.008	-	-	17.99	-	-	-	-	0.084

* denotes those sites having a stationary threshold and stationary GP dist. model.

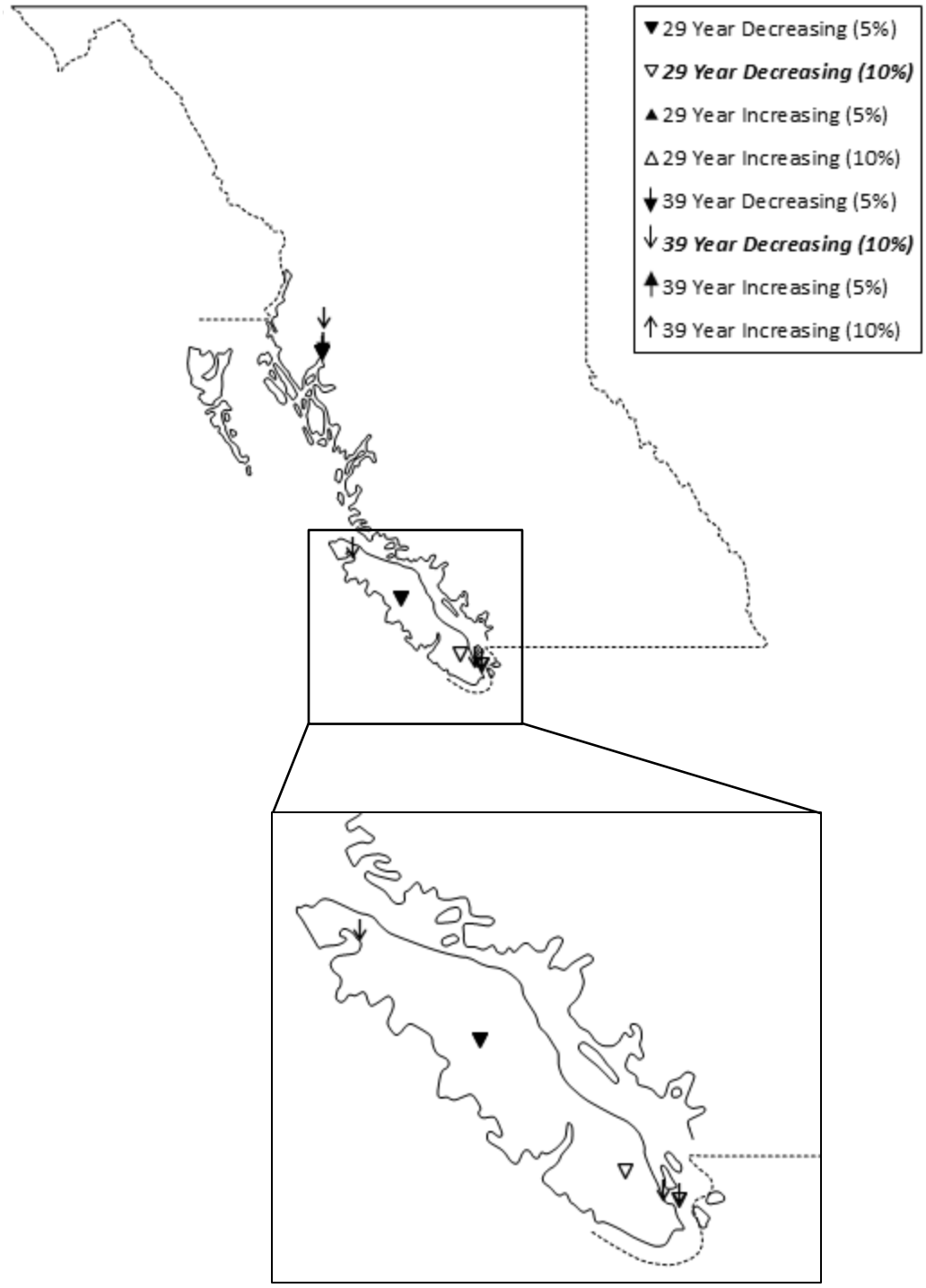

Figure 3-5: Winter season stationary threshold exceedance trend locations. Bold and italicized legend fonts indicate trends that are globally significant. Note: Only locations showing trends are shown. Depicted trends may be either for a 29- or 39-year analysis periods at the 5% and 10% significance levels.

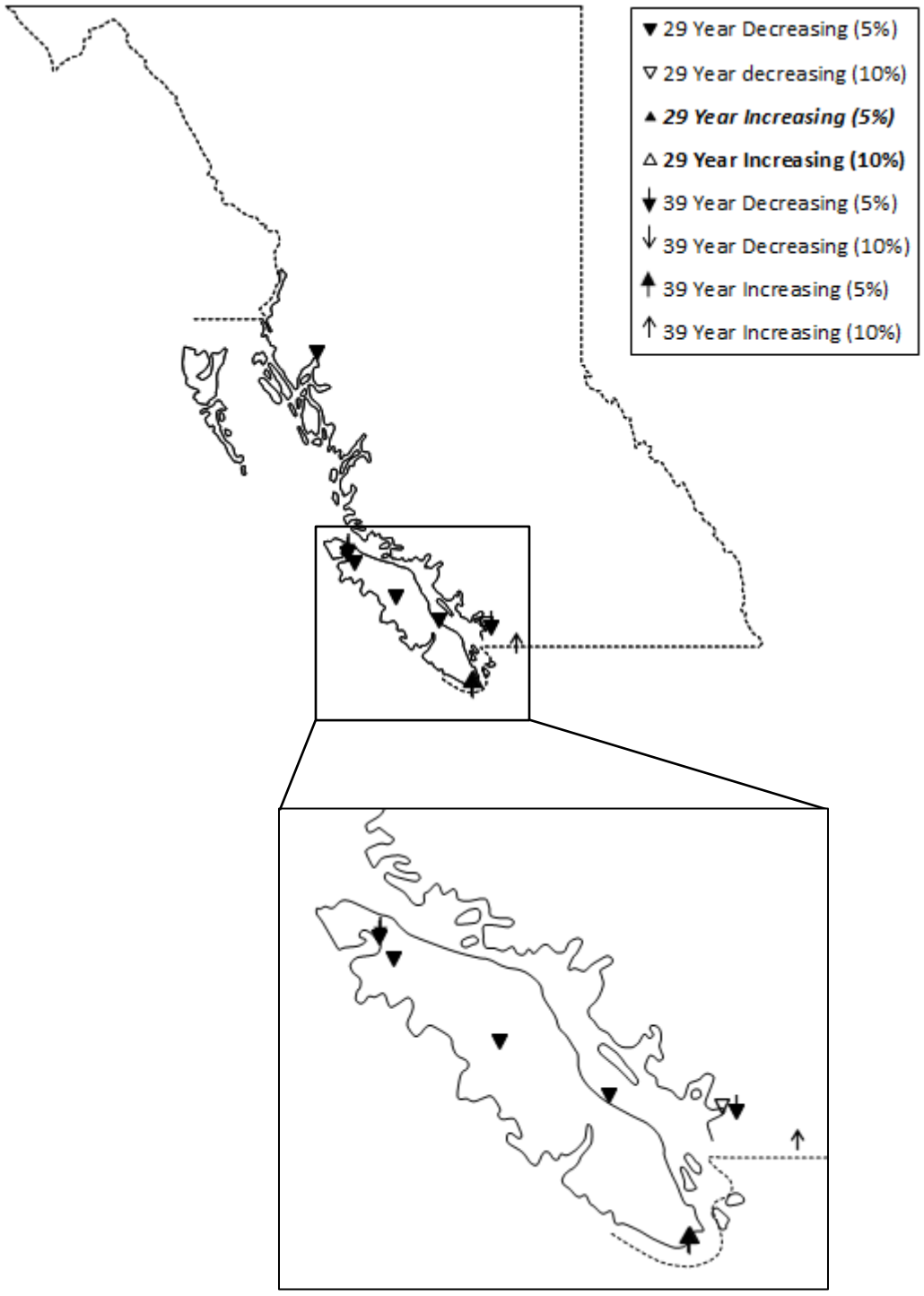

Figure 3-6: Winter season frequency of stationary threshold exceedance trend locations. Bold and italicized legend fonts indicate trends that are globally significant. Note: Only locations showing trends are shown. Depicted trends may be either for a 29- or 39-year analysis periods at the 5% and 10% significance levels.

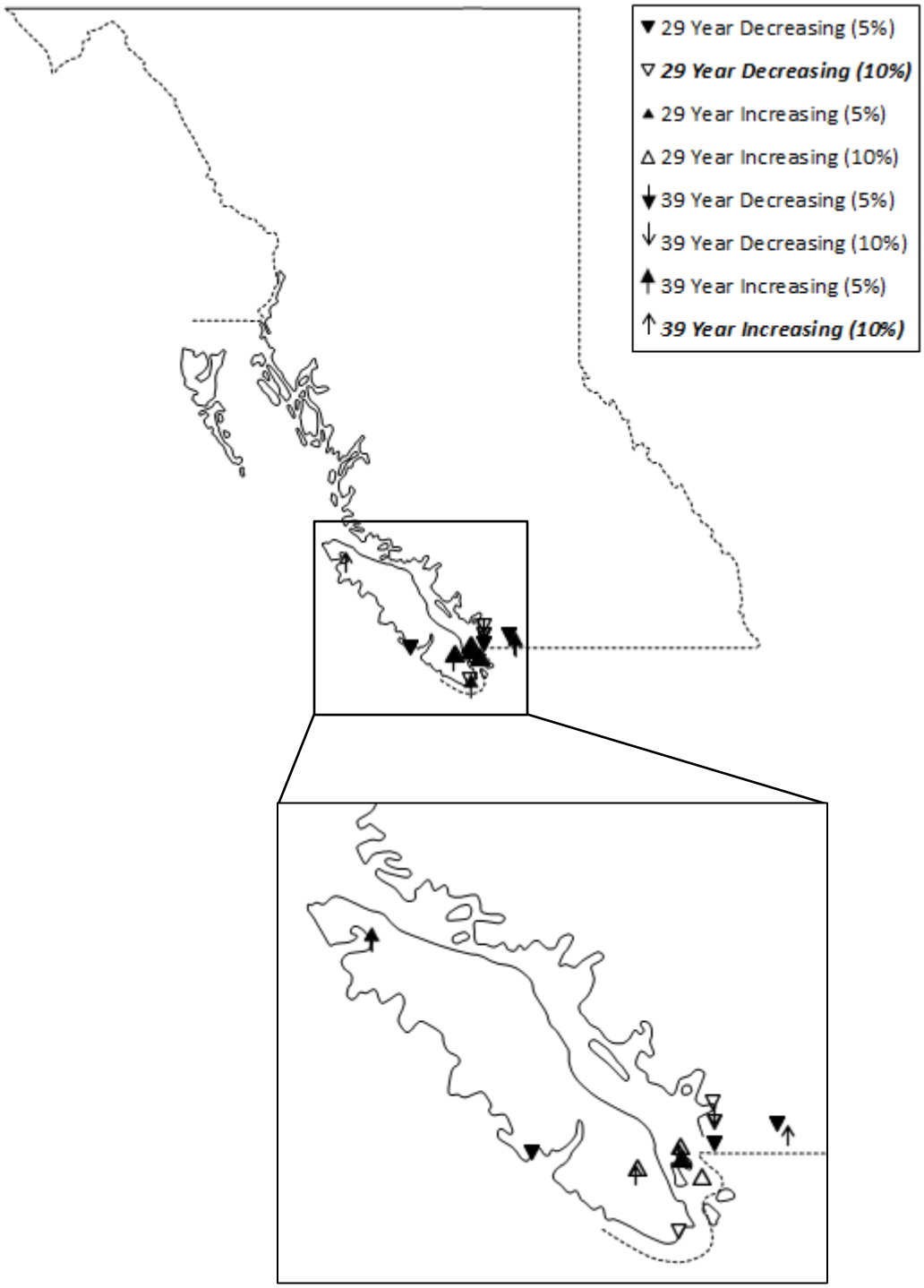

Figure 3-7: Summer season stationary threshold exceedance trend locations. Bold and italicized legend fonts indicate trends that are globally significant. Note: Only locations showing trends are shown. Depicted trends may be either for a 29- or 39-year analysis periods at the 5% and 10% significance levels.

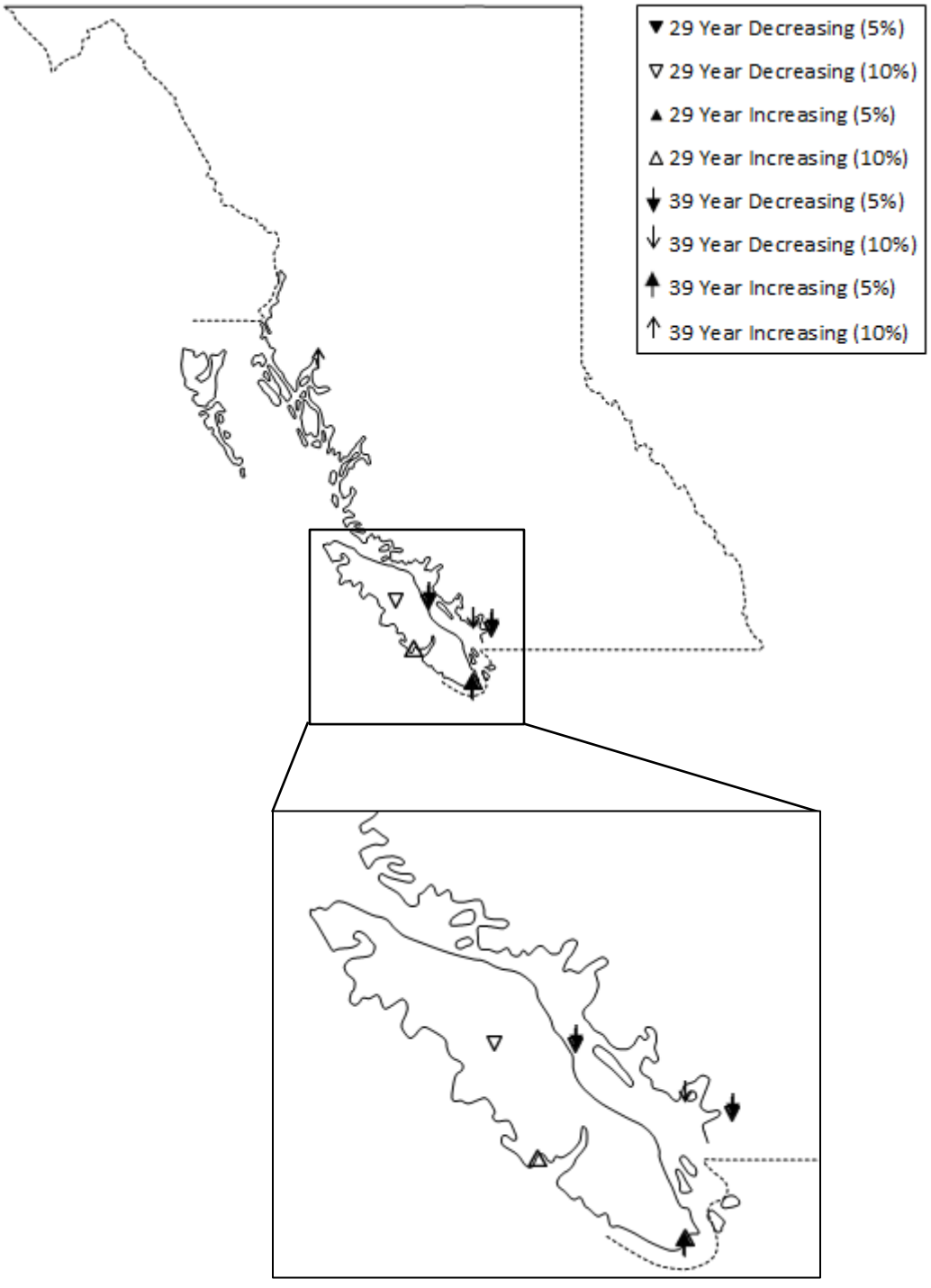

Figure 3-8: Summer season frequency of stationary threshold exceedance trend locations. Note: Only locations showing trends are shown. Depicted trends may be either for a 29- or 39-year analysis periods at the 5% and 10% significance levels.

3.4.4 Quantile Estimation Summary

Quantile estimates are calculated for the 50-year events for each of the winter and summer season data using Equation (3-2) and are illustrated in Figures 3-9 and 3-10, respectively. If a stationary model is determined to be the best fit, a comparison with nonstationary quantiles is not warranted and is therefore not presented. For all the remaining data sets, a comparison between the best-fitting stationary and nonstationary model quantiles is provided. Due to the inclusion of PDO as a covariate, the quantile estimates are not consistently smooth, particularly in the case of a PDO-dependent GP scale parameter. Trends identified through the Mann-Kendall trend test are included in the top right corner of Figures 3-9 and 3-10 as a means of comparison.

For the winter season data, 12 of the 30 sites are identified as stationary through the model selection technique. All of the sites in the North Coast region are identified as nonstationary (decreasing), whereas 56% (5 of 9) and 39% (7 of 18) of sites in the South Coast and Vancouver Island regions are, respectively. Of the remaining four stations on the South Coast, two (S1 and S7) display very mild nonstationarity, while the remaining two (S8 and S9) reveal increasing 50-year quantiles. Within the Vancouver Island region, seven of the stations have stationary quantiles (V1, V3, V11, V15, V16, V17, and V18), while seven of the remaining stations show nominal changes in quantiles with respect to time (V2, V4, V5, V8, V9, V12, and V13). The four remaining sites exhibit discernable nonstationarity, three of which are increasing (V7, V10, and V14) and one of which is decreasing (V6). A decreasing trend was also identified at station V6 at the 5% significance level for the 29 year period (1986-2014). An inconsistent decreasing trend is identified at station V14 but this could be due to the marked difference in the period of record of the station data and the 29/39 year time frames. The magnitude of uncertainty in the quantile estimates is similar for both stationary and nonstationary models, aside from site V10 which has smaller stationary CIs. Given that the stationary threshold and GP model provides substantially less uncertainty for this particular site, confidence in these nonstationary estimates is low.

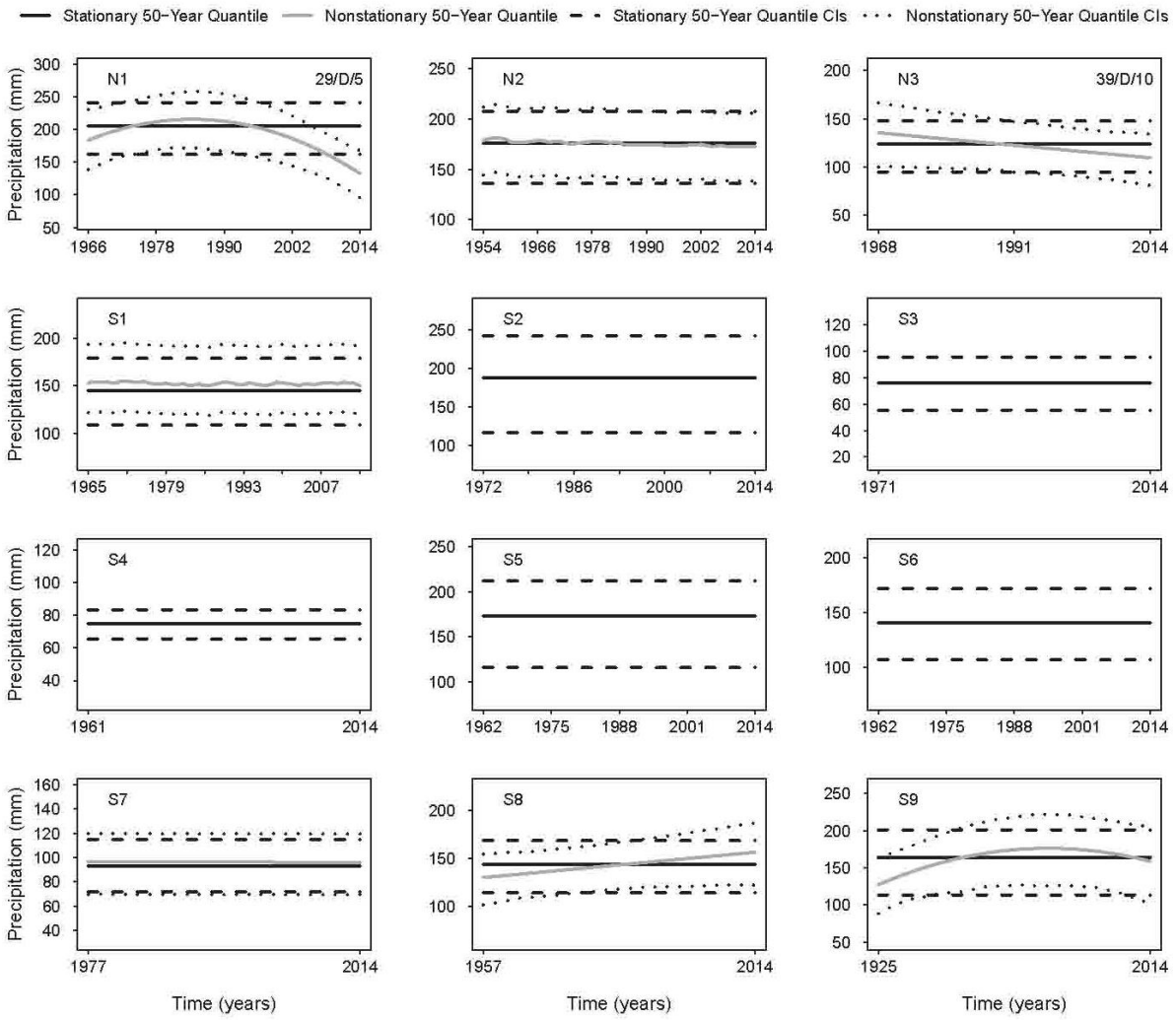

Figure 3-9 (N1-S9): 50-year nonstationary and/or stationary quantiles and their associated 95% confidence intervals - winter season. Notes: The notation in the top left of each plot is the associated station from Table 3-1; the symbol in top right corner indicates trend results for exceedance data for the 29- or 39-year periods of record, D indicates a decreasing trend, I an increasing trend, and 5/10 indicate the significance level (%). The threshold exceedance trend results are provided as a means of interpretation of model fit and comparison.

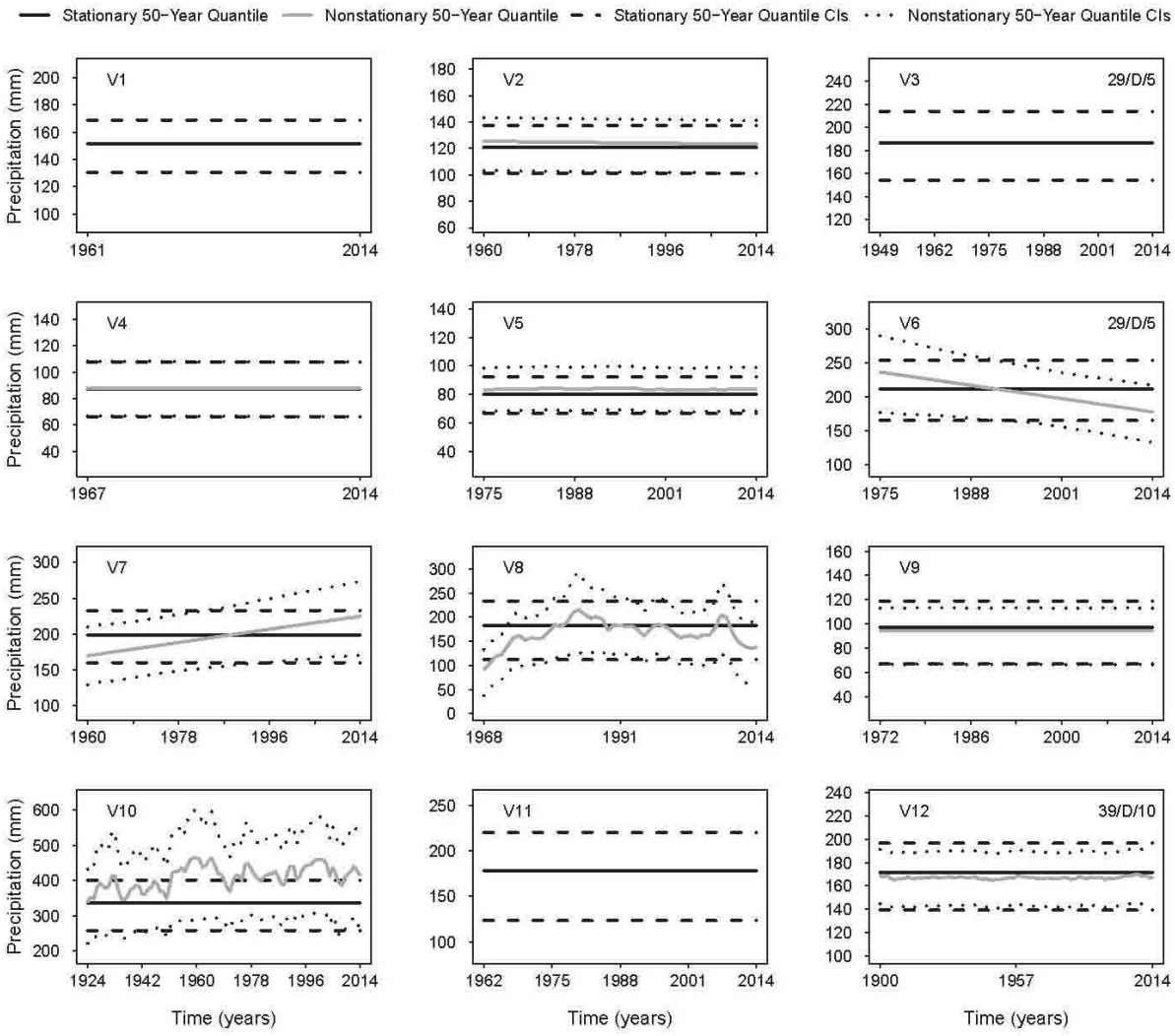

Figure 3-9 (V1-V12): 50-year nonstationary and/or stationary quantiles and their associated 95% confidence intervals - winter season. Notes: The notation in the top left of each plot is the associated station from Table 3-1; the symbol in top right corner indicates trend results for exceedance data for the 29- or 39-year periods of record, D indicates a decreasing trend, I an increasing trend, and 5/10 indicate the significance level (%). The threshold exceedance trend results are provided as a means of interpretation of model fit and comparison.

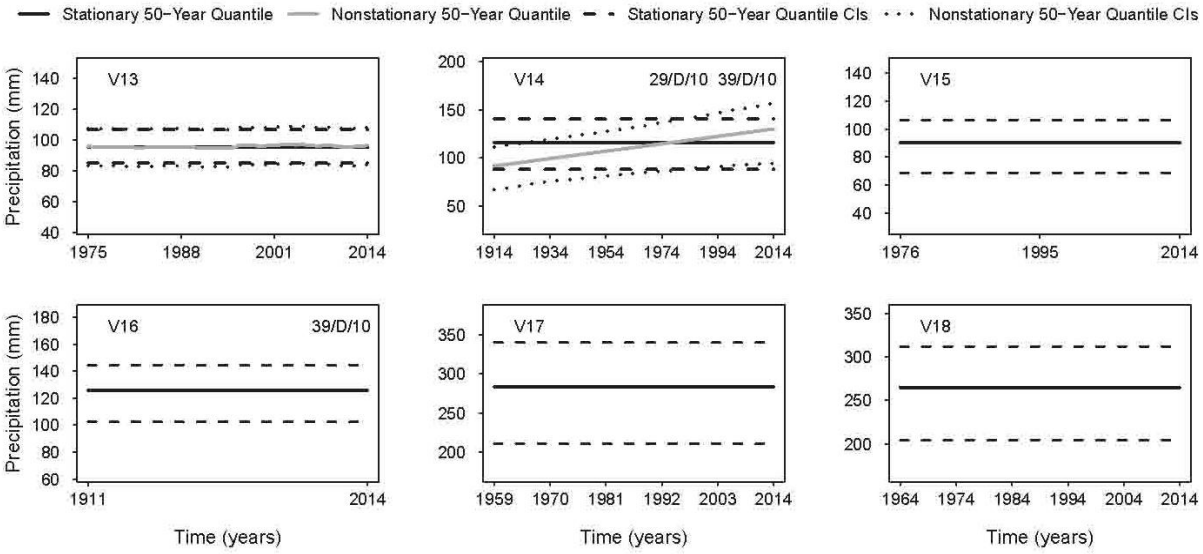

Figure 3-9 (V13-V18): 50-year nonstationary and/or stationary quantiles and their associated 95% confidence intervals - winter season. Notes: The notation in the top left of each plot is the associated station from Table 3-1; the symbol in top right corner indicates trend results for exceedance data for the 29- or 39-year periods of record, D indicates a decreasing trend, I an increasing trend, and 5/10 indicate the significance level (%). The threshold exceedance trend results are provided as a means of interpretation of model fit and comparison.

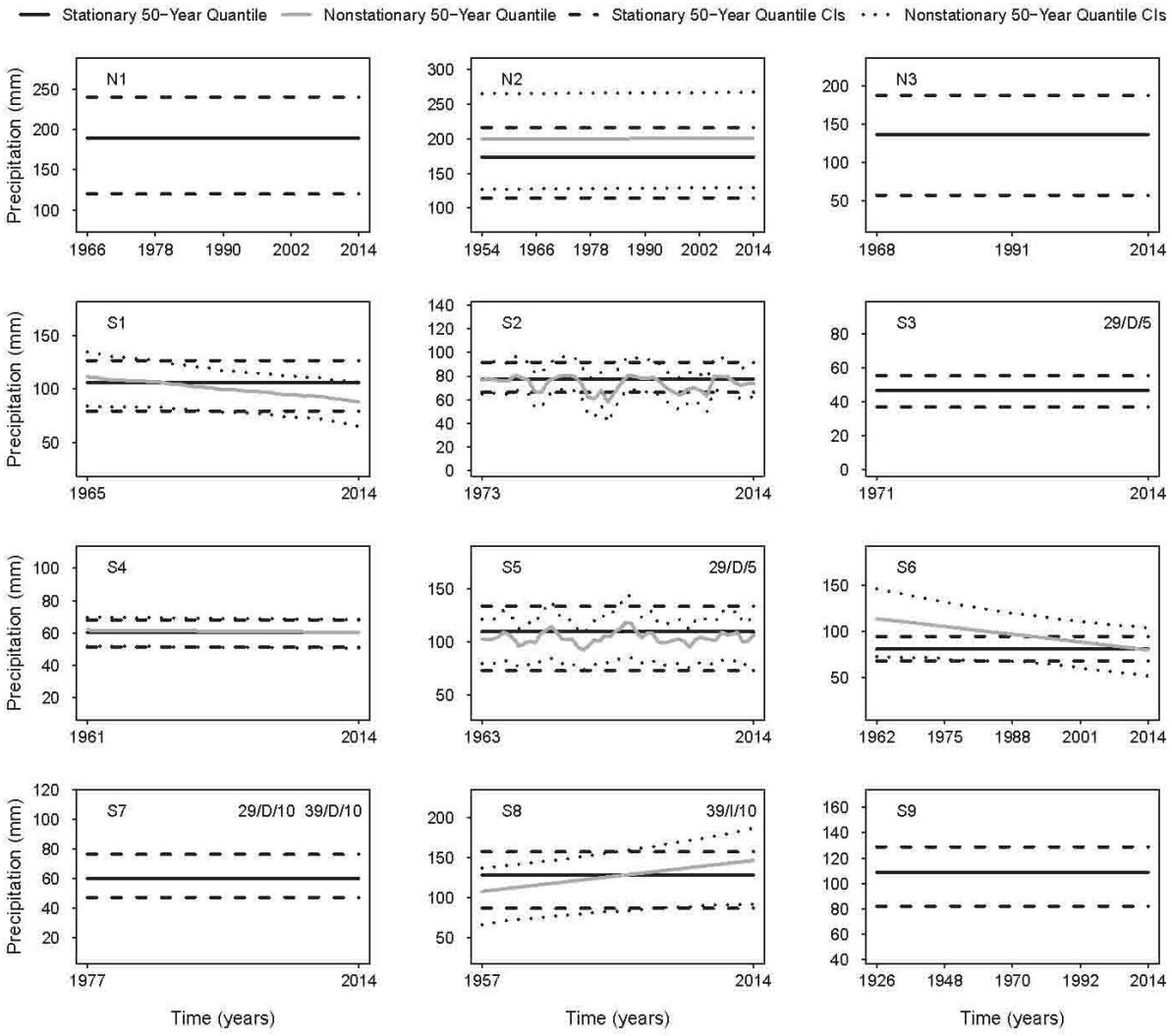

Figure 3-10 (N1-S9): 50-year nonstationary and/or stationary quantiles and their associated 95% confidence intervals - summer season. Notes: The notation in the top left of each plot is the associated station from Table 3-1; the symbol in top right corner indicates trend results for exceedance data for the 29- or 39-year periods of record, D indicates a decreasing trend, I an increasing trend, and 5/10 indicate the significance level (%). The threshold exceedance trend results are provided as a means of interpretation of model fit and comparison.

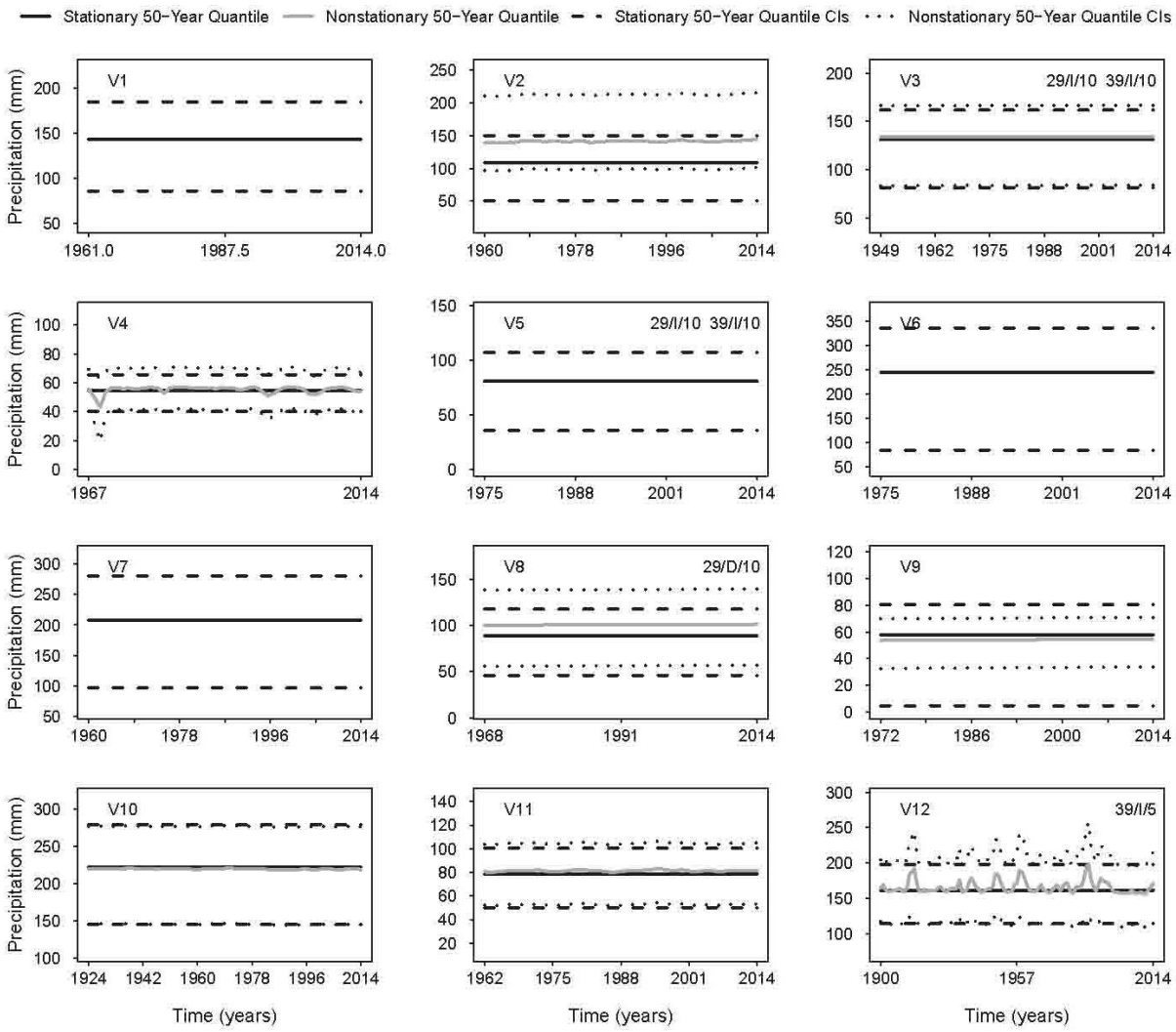

Figure 3-10 (V1-V12): 50-year nonstationary and/or stationary quantiles and their associated 95% confidence intervals - summer season. Notes: The notation in the top left of each plot is the associated station from Table 3-1; the symbol in top right corner indicates trend results for exceedance data for the 29- or 39-year periods of record, D indicates a decreasing trend, I an increasing trend, and 5/10 indicate the significance level (%). The threshold exceedance trend results are provided as a means of interpretation of model fit and comparison.

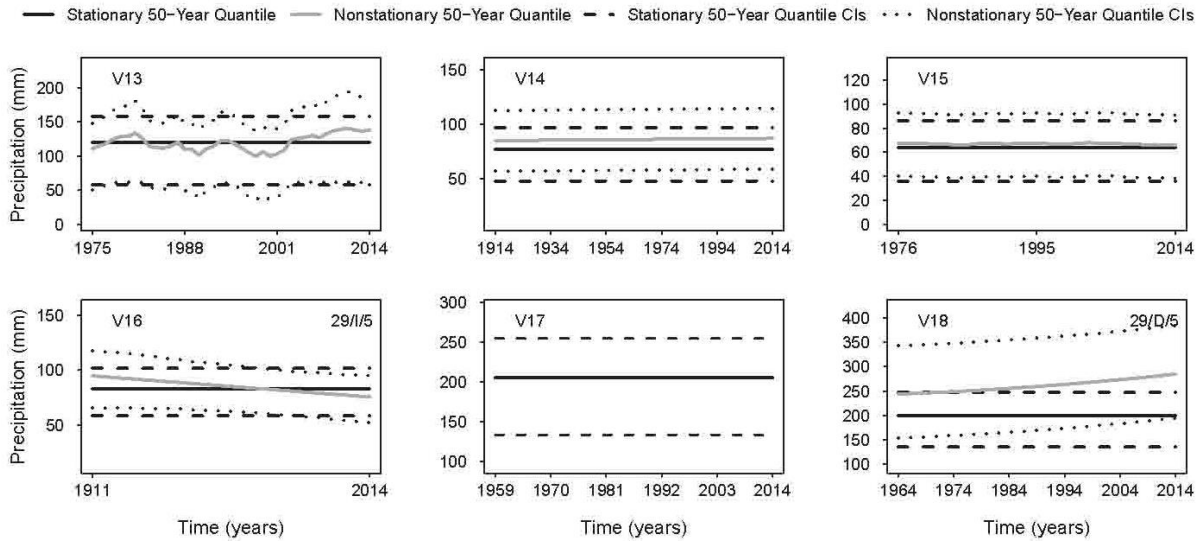

Figure 3-10 (V13-V18): 50-year nonstationary and/or stationary quantiles and their associated 95% confidence intervals - summer season. Notes: The notation in the top left of each plot is the associated station from Table 3-1; the symbol in top right corner indicates trend results for exceedance data for the 29- or 39-year periods of record, D indicates a decreasing trend, I an increasing trend, and 5/10 indicate the significance level (%). The threshold exceedance trend results are provided as a means of interpretation of model fit and comparison.

The summer season data show slightly less stationarity, with 10 of the 30 sites having time-independent quantiles. In the North Coast region, two of the three sites are stationary, with the remaining nonstationary site (N2) showing little discernable change over the period of record. Three of the nine sites in the South Coast region are stationary (S3, S7, and S9) and three have minimal nonstationarity or fluctuate around a constant mean value (S2, S4, and S5). Of the remaining three stations, two show decreasing and one increasing 50-year quantiles, although nonstationary estimates for station S6 have slightly more uncertainty than their stationary counterparts. Finally, Vancouver Island contains five stationary sites, but of those remaining, only two (V16-decreasing and V18-increasing) show apparent changes in 50-year peak return period precipitation estimates. In agreement with the winter season findings, the uncertainty in the stationary and nonstationary 50-year quantile estimates are relatively comparable, aside from those of site V18. There is noticeably more uncertainty in the nonstationary quantiles for this station; therefore, there is minimal confidence in these estimates.

3.4.5 Regional Pattern Overview

Figure 3-9 demonstrates that the North Coast, winter season 50-year quantile estimates are generally decreasing, although one site (N2) shows very little change over the period of record. Furthermore, the trend analysis identified a decreasing trend at N1 for the 1986-2014 period at the 5% significant level, and a decreasing trend for the 1976-2014 period was detected at site N3 at the 10% significance level. One trend was also identified in the North Coast region frequency data in the winter season, a variable for which the trend results are field significant (Figure 3-6). Although the detected trends in the magnitude of the POT data were not determined to be field significant, these findings suggest overall decreasing peak precipitation in the winter months. An analysis of Figure 3-10 reveals that there is minimal indication of nonstationarity in the summer peak quantile estimates, as well as no trends being identified at those sites. The GP and threshold model parameters in the winter season for the North Coast displayed both PDO- and time-dependence, although the models selected for the summer months are dominated by stationarity (Tables 3-5 and 3-6). Furthermore, divergent nonstationary behaviour is observed in the winter quantile estimates at stations N1 and N2, although these two stations are very closely located. Differences in the nonstationary quantile estimates at these two stations may be due to other factors such as changes in instrumentation and location over time.

Figures 3-5 and 3-6 illustrate that Vancouver Island is dominated by decreasing trends in above threshold occurrences and the frequency of those occurrences in the winter season. Globally significant decreasing trends were identified for the above threshold occurrences for the 39-year period decreasing trends at the 10% significance level. Additionally, one field significant increasing trend was identified (Figure 3-6) for the winter 1986-2014 period at the 5% and 10% significance levels. A combination of increasing and decreasing trends were found in the summer months in the magnitude and frequency of POT events; however, only increasing threshold exceedance trends for the 39-year period (10% significance level) and decreasing trends for the 29-year period (10% significance level) were found to be field significant. The winter quantile estimates have somewhat contradictory results, given that a number of sites are discernably increasing. Also, the summer 50-year quantile estimates were predominantly stationary, yet a number of significant trends were identified through the trend analysis. Finally, an analysis of the threshold models from Tables 3-5 and 3-6 selected for Vancouver Island suggests that the central tendency of the winter and summer season data is influenced by both time and PDO. An analysis of the GP scale parameter estimates reveals that the winter variability is more likely stationary.

The South Coast region is characterized by relatively stationary quantiles for both seasons, although several fluctuate about a constant mean. However, two sites have noticeably increasing 50-year quantile estimates in the winter and summer quantiles indicate that several sites have discernably increasing and decreasing quantiles as well. Two significant decreasing trends were identified in the winter season, while summer season trends for both the frequency and magnitude of POT events were generally decreasing. Additionally, the South Coast region shows a mixture of time and PDO-dependence in the summer in all model parameters, while the winter season appears to be rather stationary.

3.5 Discussion

This section includes a comparison of this analysis with other similar studies. Such comparison is a challenging task due to differences in methodologies, but is nonetheless a valuable undertaking.

In an analysis of daily precipitation data throughout Canada, Stone *et al*. (2000) found decreasing trends in heavy precipitation frequency in winter months in coastal BC, which is consistent with the findings of this work. Increases in the frequency of peak summer events were also identified by Stone *et al*. (2000); however, their research did not find globally significant trends in summer frequency data. Zhang *et al*. (2000) found statistically significant increases in gridded daily precipitation in summer months in the southwestern portion of BC for the 1900-1998 and 1950-1998 periods. The findings of the quantile analysis for the South Coast region herein are in agreement with these findings. Zhang *et al*. (2000) also identified significant increasing trends in an area similar to that of the North Coast region for the 1900-1998 period, which is generally consistent with the quantile analysis of this area. In a Canada-wide analysis, Vincent and Mekis (2006) examined numerous precipitation and temperature indices from 1950 to 2003. They found an increase in the number of days with snow and rain in coastal BC, although these finding were not statistically significant. However, significant increasing trends in very wet days ($\geq 95^{th}$ percentile) were primarily identified on Vancouver Island and in the North Coast region, which is inconsistent with the results of this research, given that the detected trends in frequency of over threshold events were generally decreasing (significant or otherwise). Burn *et al*. (2011) used hourly precipitation to carry out a POT analysis of events of varying duration in BC. The results of this analysis for the summer season found statistically and globally significant trends in the frequency of above threshold events for the 24-hour duration storms (40-year period). However, no globally significant trends were found in the summer frequency data herein. Also, the winter season frequency data of Burn *et al.* (2011) are

dominated by increasing trends, although none were field significant. These trend results are somewhat consistent with those of the winter trend analysis carried out herein, whereby a globally significant increasing trend was found for the winter frequency data. The winter peak magnitude data herein are dominated by decreasing trends. However, some evidence of increasing winter quantiles was found in the Vancouver Island and South Coast regions.

Mote (2003) found predominantly increasing trends in daily precipitation for the winter season from 1950 to 2000 in southern coastal BC, which is consistent with the quantile estimates from the South Coast region. Mote (2003) also found a mixture of increasing and decreasing trends on Vancouver Island, which is also consistent with the results herein. In an analysis of Canadian daily total rainfall, Vincent and Mekis (2009) found increasing trends in summer and decreases in winter precipitation in southwestern BC. These results are somewhat similar to those found for the North Coast region in the winter and the South Coast region in the summer. Although the work of Mote (2003) and Vincent and Mekis (2009) used daily precipitation indices, they did not assess peak precipitation trends; therefore, a direct comparison between these two studies is not possible.

General circulation model (GCM) projections for coastal BC indicate the potential for wetter winter seasons and drier conditions in the summer months (British Columbia Ministry of the Environment [BCME], 2007; Walker *et al*., 2008). The results established through the trend analysis are somewhat consistent with the above noted projections for the winter season given that globally and statistically significant increasing trends in frequency were found for the 29-year analysis period at both significance levels. Also in agreement, a number of trends in the summer season were found to be decreasing. It should also be noted that the results of the quantile analysis provided better agreement with likely climate change projections.

Whitfield *et al*. (2010) describe the PDO as a powerful post-hoc explanatory concept, limiting the forecasting potential with this index. Therefore, care should be taken in extrapolating the quantile estimates past the period of record, particularly in instances with strong quadratic behaviour. One of the anonymous reviewers suggested that concave quadratic behaviour may possibly be due to a period of predominant La Niña between approximately 1990 and 2012, which led to a long period of cold PDO SSTs.

3.6 Conclusions

This paper incorporates existing nonstationary extreme value approaches to develop a methodology for the selection of stationary/nonstationary thresholds. Nonstationary threshold selection is carried out through quantile regression, which allows for the inclusion of bivariate and multivariate models. A two-stage selection process is proposed, in which AIC weights are initially employed to determine the best-fitting (non)stationary GP distribution model for each threshold. Threshold/GP distribution models are then compared through the use of Q-Q plots. This methodology is developed as a complementary approach to trend detection which allows the user to visually identify potential trends in peak quantile estimates.

There is more evidence of stationarity in the North Coast winter and summer threshold exceedance data, which is confirmed by the quantiles and trend analyses. Similar results are revealed for the South Coast region, although there is some evidence of increasing quantiles in the winter. Vancouver Island shows the most potential for nonstationary increasing peak 50-year quantiles in both seasons, although there is some evidence of decreases in the winter. Through the trend analysis, it was revealed that a number of significant trends were found in the magnitude and frequency of POT data for both seasons.

In general, the results from the quantile and trend analyses are in agreement with previously published work. Furthermore, dissimilarities in the trend and quantiles analyses reinforce the usefulness of investigating trends through various means and highlight the importance of comparisons within various periods of record. Based on this historical data trend analysis, there is less risk of serious precipitation related winter events; however, the frequency of these events is increasing in recent years. Furthermore, the quantile analysis suggests that there is moderate risk of increases/decreases in winter or summer hydrometeorological hazards in coastal BC.

Transition Paragraph B

Hydrometeorological time series commonly exhibit serial correlation due to climatological and hydrological phenomena. Serial correlation occurs when observations at a given time are dependent on previous observations. For example, antecedent soil moisture conditions may lead to the detection of autocorrelation in a data set. The presence of significant serial correlation may have deleterious effects on nonparametric trend tests applied to the data. It has been demonstrated that the existence of positive autocorrelation leads to increased rates (type I error) of false detection of trends when none exists (von Storch, 1995). Conversely, negative serial correlation decreases the likelihood of detecting significant trend when one does indeed exist. Accurate identification of existing trends is essential before the application of FFA techniques to ensure that this form of inhomogeneity is adequately addressed. Throughout the completion of Chapters 2 and 3, it was observed that there was a lack of research on the effects of negative serial correlation on existing techniques which account for autocorrelation for nonparametric trend testing. This has been largely omitted from the literature given the physical meaninglessness of negative memory structure (von Storch, 1995), however, it nonetheless exists and must be adequately addressed. Furthermore, differences in the serial structure of BMS and POT data have been largely overlooked. The purpose of Chapter 4 is to compare the power and type I error rates of various commonly used techniques that account for serial correlation. This analysis includes an assessment of both positive and negative serial correlation, the latter of which has not been thoroughly assessed in the literature. Furthermore, the differences in BMS and POT memory structures are analyzed, along with a land-use dependent trend analyses. While the preceding chapter focused on daily precipitation data, the following chapter employs an instantaneous dataset where block maxima and partial duration series data are used.

Chapter 4

Trend Identification in the Presence of Serial Correlation: A Comparison of Block Maxima and Peaks-Over-Threshold Data

Summary

The existence of serial correlation in a hydrometeorological time series can have deleterious effects on trend detection tests. To account for significant autocorrelation detected in data sets, various techniques have been developed over time, each having their own assumptions and accuracy. Many of such techniques assume the data have an AR(1) memory structure, however, this assumption is not universally applicable. Furthermore, the existence of positive or negative serial correlation has dissimilar effects on these statistical techniques. This research compares the power and type I error rates of various well-known and several newer techniques to account for positive and negative serial correlation in combination with the Mann-Kendall nonparametric trend test. Through a case study of a subset of southern Ontario watersheds, it is determined that BMS data are more likely to have significant negative lag-1 serial correlation but are also more likely to have an AR(1) serial structure. POT data are predominantly positively autocorrelated, with a greater proportion of higher order memory structures. It is determined that in the case of positively serially correlated data, the BBS, the VCCF1, and the SBS are the most powerful. Alternatively, in the case of negative autocorrelation, the VCCF1 and BBS approaches are recommended.

Using peak event data parsed from instantaneous streamflow observations, a comparative trend analysis is carried out on 70 southern Ontario watersheds with mixed land-use. The land-use classifications were defined as urban, agricultural, natural, and RHBN. Furthermore, the differences in trends detected in POT and BMS data are highlighted. There is evidence of significant trends in the RHBN and natural catchments, however, these trends are not consistent in either the POT or BMS data sets. The results of the trend analysis highlight commonly detected trend patterns in the urban catchments, that is, statistically significant increases in peak discharge, volume, and frequency of events, along with decreasing duration, and time to peak are observed. Similar results are identified in the agricultural watersheds, with consistent significant trends found in the peak event magnitudes.

4.1 Introduction

There is a growing body of research focusing on temporal dependence in hydrometeorological event characteristics, most notably, accurate identification of trends in peak events. The Fifth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC) states that there is strong evidence to suggest that global land-surface temperatures have risen since the late 19th century and increases in extreme precipitation are anticipated in a warmer climate (Hartmann *et al*., 2013). Increased precipitation can have a cascading effect within the hydrologic cycle, thus prompting extensive research focused on trend identification in temperature, precipitation, and streamflow records, among other hydrometric time series worldwide (Lettenmaier, 1976; Hirsch *et al.* 1982; Hipel *et al*., 1988; Burn, 1994; Lins and Slack, 1999; Douglas *et al*., 2000; Zhang *et al*., 2001; Burn and Hag Elnur, 2002; Yue and Pilon, 2004; Svensson *et al*., 2005; Burn, 2008; Bormann *et al*., 2011; Burn *et al*., 2011; Noguchi *et al*., 2011; Sonali and Kumar, 2013). This chapter focuses on the effects of serial correlation on the power and type I error rates of positively and negatively autocorrelated data, and compares several well-known and more recent statistical techniques that account for serial correlation in hydrometeorological data. A further emphasis is placed on land-use dependent trend analyses using BMS and POT data.

4.1.1 Overview of the Background and Effects of Serial Correlation

Serial correlation is commonly detected in hydrometeorological data, for example, antecedent soil moisture conditions can lead to autocorrelation in streamflow measurements. Positive autocorrelation within a time series can affect trend detection tests by increasing the likelihood of detecting significant trend when none is present (von Storch, 1995). Therefore, if serial correlation is identified within a time series, it must be properly accounted for to accurately assess the significance of trends. There are a number of parametric and nonparametric tests that can be used for trend detection, although many commonly applied parametric tests (e.g. Student t-test) require data to be normally distributed. It has been documented that parametric tests are slightly more powerful than nonparametric tests when applied to Gaussian data, although this assumption may not hold true in all instances (Hamed and Rao, 1998; Yue and Pilon, 2004). The Mann-Kendall (MK) (Mann, 1945; Kendall, 1975) nonparametric test is the most commonly applied trend detection test; however, through extensive simulation, Yue *et al*. (2002a) documented the power of MK test and the Spearman

rank correlation (SRC) test (Lehmann, 1975; Sneyers, 1990) to be comparable for detecting trend through simulation experiments. These results may be inconsistent with practical applications of the nonparametric methods however (Khaliq *et al*., 2009; Sonali and Kumar, 2013).

Independence within a dataset is a fundamental assumption of all trend detection tests, one that is violated in the presence of serial correlation. When applying trend detection tests founded upon hypothesis testing, there are two potential sources of error. Type I error (with a probability equal to α or the significance level) occurs when the null hypothesis (H_0) of no trend is rejected, when there is no trend present in the data. The probability of rejecting the null hypothesis when it is true should be equal to the assigned (nominal) significance level. Type II error (with a probability equal to γ) arises when there is a trend present in the data but the alternative hypothesis (H_1) is incorrectly rejected. The power of the test is determined when a trend exists in the data and is correctly identified ($1-\gamma$) (Hipel and McLeod, 1994).

The work of von Storch (1995) was one of the first to highlight the effect of serially correlated data on the type I error rate of the MK test. The author suggests the use of prewhitening (PW) using an AR(1) model to account for autocorrelation before the MK test is applied. Through extensive simulation, Yue *et al*. (2002b) examine the effect of both positive and negative serial correlation (using and AR(1) model) on trend-free time series. For positive values of autocorrelation, the authors' results were in agreement with von Storch (1995) in that the type I error rate increases as a function of increasing lag-1 serial correlation (ρ_1). Also, it was found that negative serial correlation has the opposite effect, whereas, the type I error rate decreases as a function of decreasing ρ_1. Additionally, it was found that prewhitening effectively removes serial correlation, reducing the type I error close to the nominal value. When positive trend was superimposed onto a positive AR(1) process, it was determined that the PW approach removes a portion of the trend, thus making it less likely that a significant trend would be detected (Yue *et al*., 2002b). These results were in accordance with those of Yue and Wang (2002), who demonstrated that the PW approach may lead to inflation/deflation of negative/positive trend in time series. Due to the effect of prewhitening on trend magnitude, Yue *et al*. (2002b) suggested a revised procedure, trend-free prewhitening (TFPW), in which a linear trend is estimated and removed from the data before prewhitening. The residual prewhitened series and trend are then blended before trend detection. When applied to streamflow data, Yue *et al*. (2003) found that TFPW was able to detect more trends than PW, thus demonstrating the method's ability to more accurately limit the effects of serial correlation on data with trend. Yue and Wang (2002) and Bayazit and Önöz (2007) both suggested the use of selective prewhitening to

limit power loss of the technique, a suggestion that has been scrutinized by several authors (Bayazit and Önöz, 2004; Zhang and Zwiers, 2004; Hamed, 2009). Hamed (2009) argued that selective prewhitening is not a valid option and the technique could be corrected by simultaneously estimating ρ_1 and trend to avoid bias before applying the technique. More recently, in a study by Önöz and Bayazit (2012), the authors found the TFPW technique to have more power for detecting trend than the original MK test when applied to data with significant lag-1 serial correlation due to its propensity to inflate trend estimates and propose a modification to the procedure (MTFPW). MTFPW is similar to TFPW, the difference being that original series is prewhitened, as opposed to prewhitening the detrended series (both methods are explained in detail in section 4.3). Furthermore, Rivard and Vigneault (2009) demonstrate the TFPW technique may be more appropriate for negatively correlated data.

Several authors have demonstrated that autocorrelation in time series affects the variance of the MK test statistic (Hamed and Rao, 1998; Yue *et al*., 2002b) but not the asymptotic normality (Cabilio *et al*., 2013). Based on this observation and the work of Bayley and Hammersley (1946), Hamed and Rao (1998), and Yue and Wang (2004a) proposed a variance correction approach for the MK test to account for the increase/decrease in variance when positive/negative serial correlation is present. The authors propose correcting the variance with the use of an effective sample size (ESS) that accounts for the effects of autocorrelation (Hamed and Rao, 1998; Yue and Wang, 2004a). It has been established that both variance correction techniques have similar power in detecting trend (Khaliq *et al*., 2009; Blain, 2013). Wang *et al*. (2015) also investigate the effects of variance on the MK test but their methodology aims to correct the discrepancy between the variance of the original data and the residual noise after PW.

To account for serial correlation, Kundzewicz and Robson (2000, 2004) suggested the use of a BBS method for trend detection, whereby the resampling would replicate the serial dependence structure of the data by resampling in blocks. The suitability of the BBS technique for limiting the effects of serial correlation on the MK test was assessed by several authors, through which it was determined that the BBS approach had a comparable power for detecting trend as the variance correction approach when significant serial correlation was present in the data (Khaliq *et al*., 2009; Önöz and Bayazit, 2012). In a study by Khaliq *et al*. (2009), it was established that the TFPW approach identified more significant trends than the BBS approach. The authors attribute the behaviour to the Markovian dependence structure used in the TFPW technique, an assumption that may not be valid for all hydrometeorological time series. Önöz and Bayazit (2012) found the TFPW

approach to have a higher power for trend detection than the variance correction and BBS approaches but also note that the method had a larger type I error rate. Blain (2014) found similar results regarding TFPW and also determined that PW and the MTFPW approaches provides similar type I error rates but the MTFPW approach was more powerful.

Noguchi *et al*. (2011) suggest accounting for autocorrelation in a data series using an autoregressive SBS along with a statistical trend detection test. There is a considerable body of literature concerning the use and applicability of the SBS technique (Kreiss, 1992; Bühlmann, 2002; Politis, 2003; Chernick, 2008; Kreiss *et al*., 2011; Cabilio, 2013). This approach consists of fitting an autoregressive model of order p (AR(p)) to the data and bootstrapping the residuals of the model. The bootstrapped residuals are then converted into bootstrapped realizations using the same AR(p) model.

4.1.2 Introduction to the Effects of Land-Use on Watershed Hydrology

The hydrologic response of a watershed is related to the land cover and use within the catchment (Rosburg *et al*., 2017). Consequently, land-use changes are generally associated with marked changes in hydrologic response; however, the degree of hydrologic change is dependent on the extent and type of land-use modifications. These types of changes can be anthropogenic (urbanization, agricultural, deforestation, etc.) or natural (wildfire) but nonetheless result in similar hydrologic regime alterations. Given that a hydrograph is most commonly characterized by its peak, duration, and volume, the effects of hydromodifications are generally concentrated on these features.

According to Statistics Canada, the percentage of Canadians living in urban areas increased by 80% between 1911-2011 (Statistics Canada, 2011a). Urban expansion results in the addition of impervious surfaces and channelization of watercourses, with a subsequent increase in surface runoff. The hydrological consequences to the receiving waters are far reaching and include water quality and aquatic ecosystems degradation (Everard and Moggridge, 2012). Urbanization results in flashier streams with increased peak discharge, volume, and frequency of runoff events, while decreasing duration, time to peak, and additionally decreased baseflow in urban systems (Leopold, 1968; Seaburn, 1969; Konrad, 2003; Morgan *et al*., 2004; Burns *et al*., 2005; Chang, 2007; Hamdi *et al*., 2011). The goal of current storm water management (SWM) practices is to offset the effects of urban growth, however, runoff events may still tax these systems, resulting in similar results to those previously noted. Given that approximately 86% of Ontario's populace lives in urban centers

(Statistics Canada, 2011b), a focus on urban peak runoff event trends may help to mitigate the associated socioeconomic risks.

In addition to urbanization, agricultural land-use changes can have considerable hydrologic effects. It is estimated that over 50% of croplands in the United States (US) and Canada require artificial drainage for productive yields (Skagg *et al*., 1994). Surface ditches and subsurface tile drains are commonly installed for agricultural purposes to provide adequately dry, well-aerated soils (Irwin and Whiteley, 1983). Improved drainage results in substantial hydrological impacts; however, the manifestation of these changes are watershed specific (King *et al*., 2014; Apsīte *et al*., 2017). Parallel processes affect the hydrologic response of a tile drained agricultural watershed (Spaling, 1995). The addition of subsurface drainage allows water to move more quickly through the soil which could potentially increase peak flows (Robson, 1990). Conversely, tile drainage increases the depth to the groundwater table, creating greater storage capacity in the soil, thus increasing retention time, which may result in smaller peak flows (Robson, 1990; Spaling, 1995). Agricultural areas that depend on surface drainage may, however, see increases in peak events (King *et al*., 2014). The nature of the hydrological changes is, therefore, highly dependent on the type of artificial drainage (whether surface or subsurface), soil characteristics (well or poorly drained soils), and antecedent moisture conditions (Spaling, 1995). For these reasons, it is unsurprising that there have been conflicting findings regarding the effects of agricultural drainage on peak flow magnitudes in southern Ontario (Irwin and Whiteley, 1983; Serrano *et al*., 1985; Spaling, 1995; Fraser and Flemming, 2001) and worldwide (Madramootoo and Broughton, 1987; Madramootoo *et al*., 1988; King *et al*., 2014; Rahman *et al*., 2014; Muma *et al*. 2016; Apsīte *et al*., 2017; Rogger *et al*., 2017).

This research is a novel addition to the existing literature in that it extensively examines the effects of negative serial correlation on numerous existing statistical tests that account for autocorrelation when used with nonparametric trend tests. Furthermore, this research explores the differences in the memory structures of the BMS and POT dataset, thus allowing for the determination of the most appropriate technique for use on these types of data.

The purpose of this chapter is to provide a comparison of several well-known and other more recent techniques that account for short-term persistence when the MK trend test is applied. Similar research has previously been carried out; however, the results of this chapter indicate the need for a more critical analysis of the effects of negative serial correlation on existing techniques. The topic of negative serial correlation is one that has received limited attention in the literature, although, it is commonly found in practice. Furthermore, a comparison of the most appropriate techniques to

account for serial correlation for use on BMS and POT data is provided. This chapter additionally provides a comparative analysis of trends having various periods of record and land-use changes.

4.2 Overview of the Mann-Kendall Nonparametric Trend Test

Although there are numerous techniques for identifying trend in a time series, this research focuses on the Mann-Kendall nonparametric test as it is the more commonly applied technique. The MK test is a rank-based test used for determining the significance of a trend. For a series of data $x_1, x_2, \dots, x_n$, the MK test statistic (S) is given by (Mann, 1945; Kendall, 1975):

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \tag{4-1}$$

where x_j and x_i are sequential observations and:

$$sign(\omega) = \begin{cases} 1 & if\ \omega > 0 \\ 0 & if\ \omega = 0 \\ -1 & if\ \omega < 0 \end{cases} \tag{4-2}$$

When $n \geq 8$, the test statistic (S) is approximately normally distributed with mean ($E(S)$) equal to zero and variance given by:

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{k=1}^{n} t_k(k)(k-1)(2k+5)}{18} \tag{4-3}$$

where t_k is the number of ties and k is the extent of a tie. The standardized test statistic (Z) is as follows:

$$Z = \begin{cases} (S-1)/\sqrt{V(S)} & S > 0 \\ 0 & S = 0 \\ (S+1)/\sqrt{V(S)} & S < 0 \end{cases} \tag{4-4}$$

The standardized MK statistic follows a standard normal distribution; therefore, the null hypothesis of no trend is rejected when $abs(Z) \geq Z_{1-\alpha/2}$, where α is the adopted significance level of the test. It should also be noted that a positive value of S indicates an increasing trend while a negative value indicates a decreasing trend.

The MK test requires independence between observations; therefore, serial correlation within a time series can lead to erroneous results. The following section describes the approaches used herein, in conjunction with the MK test, to account for autocorrelation in a times series.

4.3 Statistical Approaches for Incorporating Serial Correlation

Hydrometeorological data commonly exhibit dependence between observations. Data displaying statistically significant positive serial correlation, for example, can increase the type I error rate of the MK test leading to potentially inaccurate results. The goal of this section is to provide a brief overview of several commonly used techniques for accounting for autocorrelation, which include: prewhitening (PW), trend-free prewhitening (TFPW), prewhitening with bias corrected lag-1 autocorrelation (BCPW), modified trend-free prewhitening (MTFPW), variance correction (VCCF1 and VCCF2), variance correction prewhitening (VCPW), block bootstrap (BBS), and sieve bootstrap (SBS). Please note that all slope estimates presented herein were determined through the use of Theil-Sen nonparametric slope estimation approach (Theil, 1950; Sen, 1968).

Prewhitening (PW)

The PW approach was initially suggested by Kulkarni and von Storch (1995) and von Storch (1995) to limit the effects of serial correlation in a time series. The approach assumes an AR(1) structure of dependence within the data in which the sample lag-1 autocorrelation coefficient (r_1) is estimated. The sample data are then prewhitened using the following formula (von Storch, 1995):

$$y_t = x_t - r_1 x_{t-1} \tag{4-5}$$

where y_t is the prewhitened time series; x_t is the original data at time t; $r_1$is the lag-1 autocorrelation coefficient estimate; and x_{t-1} are observations of the original data at time $t-1$. A trend detection test can then be applied to the residual series.

Trend-Free Prewhitening (TFPW)

Trend-free prewhitening was initially proposed by Yue *et al*. (2002b) as it was demonstrated through simulation that PW removes a portion of the trend magnitude in a time series (in the case of positively serially correlated data). The TFPW procedure is as follows:

1) The slope of a trend (b) in a sample of data is initially estimated using the Thiel-Sen slope estimator.
2) Any significant trend found within the time series is assumed to be linear and the data are detrended as follows:

$$x'_t = x_t - bt \tag{4-6}$$

where x'_t is the detrended series and b is the estimated slope.

3) $r_1$is then estimated from x'_t and the data are prewhitened, resulting in (y'_t):

$$y'_t = x'_t - r_{1d}x'_{t-1} \tag{4-7}$$

where x'_{t-1} is the detrended lagged time series and r_{1d} is the estimated lag-1 serial correlation coefficient of the detrended series.

4) The final step involves the blending of estimated trend (bt) and the residual series (y'_t):

$$y_t = y'_t + bt \tag{4-8}$$

where y_t is the blended series that can now be tested for trend.

Modified Trend-Free Prewhitening (MTFPW)

In an analysis of annual streamflow series in 17 Turkish rivers all with significant lag-1 autocorrelation, Önöz and Bayazit (2012) found the original MK test had more power for detecting trend than when TFPW was applied. As a result, the authors suggested a modified trend-free prewhitening approach. The three first steps of the MTFPW approach and TFPW are the same (i.e., the slope of trend is estimated, the series is detrended, and the lag-1 autocorrelation coefficient is estimated from the detrended series). The difference between the two methods occurs when the original series is prewhitened in the modified approach, as opposed to the detrended series in TFPW, which is given as:

$$z_t = x_t - r_{1d}x_{t-1} \tag{4-9}$$

where z_t is the prewhitened series. The residual prewhitened series can then be tested for significant trend.

Prewhitening with Bias Corrected lag-1 Autocorrelation (BCPW)

Hamed (2009) maintained that trend in a time series would interfere with the estimation of serial correlation, thus introducing bias into the estimate. As an alternative to detrending, the author suggests estimating the parameters of a Markovian model with linear trend given by:

$$x_t = r_1 x_{t-1} + a + bt + e_t \tag{4-10}$$

where a is the linear trend intercept estimate and e_t is a noise/residual term. Hamed (2009) recommends solving for r_1, a, and b simultaneously using the ordinary least squares (OLS) method as an alternative to detrending. The author compares several bias correction approaches for r_1 and suggests the use of the OLS correction of van Giersbergen (2005):

$$r_1^* = \frac{(nr_1+2)}{n-4} \tag{4-11}$$

where r_1^* is the bias corrected lag-1 autocorrelation coefficient and n is the sample length. With an estimate of r_1^* the series can then be prewhitened and resulting series tested for trend.

Variance Correction (VC)

The purpose of the variance correction technique is to correct the inflated/deflated variance of the MK test statistic due to positive/negative serial correlation. The modified variance $[V^*(S)]$ is given by:

$$V^*(S) = V(S) * CF \tag{4-12}$$

where $V(S)$ is the variance of the test statistic and CF is a correction factor based on the ESS. Two commonly applied approaches for calculating CF are those of Hamed and Rao (1998) and Yue and Wang (2004a), which are respectively as follows:

$$CF_1 = 1 + \frac{2}{n(n-1)(n-2)} \sum_{k=1}^{n-1} (n-k)(n-k-1)(n-k-2) r_k^R \tag{4-13}$$

$$CF_2 = 1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right) r_k \quad (4\text{-}14)$$

where CF_1 is the correction factor proposed by Hamed and Rao (1998); CF_2 is the correction factor proposed by Yue and Wang (2004a); k is the lag; r_k^R is the lag-k autocorrelation coefficient of the ranks of the data; and r_k is the lag-k autocorrelation coefficient of the data. A trend detection test can then be applied using the modified variance. The variance correction techniques are applied under the AR(1) assumption by $r_k = r_1^{|k|}$. Where appropriate the VCCF1 method is applied to all significant lags herein. The VCCF2 method is applied to all lags up to one-fourth of the series length as anything larger may be statistically unreliable (Box and Jenkins, 1970).

Variance Correction Prewhitening (VCPW)

Wang *et al*. (2015) proposed variance correction prewhitening based on two established variance correction techniques. The slope of the trend (b), is initially estimated and removed from the series using Equation (4-6). Using this detrended series, the lag-1 autocorrelation is estimated and if significant, the series is prewhitened using Equation (4-7). A corrected trend-free prewhitened series is then computed by:

$$y_t'' = y_t'\left(s_{x_t'}^2 / s_{y_t'}^2\right) \quad (4\text{-}15)$$

where $s_{x_t'}^2$ is the variance of x_t' and $s_{y_t'}^2$ is the variance of y_t'. A modified slope estimator (b') is then determined as follows for positive values of r_1:

$$b' = b/\sqrt{VIF}$$

where:

$$VIF \approx (1 + r_1)/(1 - r_1) \quad (4\text{-}17)$$

If $r_1 < 0$, $b' = b$ is considered. The final step recombines the modified trend ($b't$) and the corrected trend-free prewhitened series:

$$x_t'' = y_t'' + bt \tag{4-18}$$

where x_t'' is the final transformed series to be tested for trend using the MK test.

Block Bootstrap (BBS)

The block bootstrap is a nonparametric technique in which the data are bootstrapped in predefined blocks (Kundzewicz and Robson, 2000). Block length is determined based on the serial dependence structure of the data, that is, the contiguous number of significant lags of serial correlation. The BBS procedure is as follows:

1) Estimate the test statistic of the trend detection test from the original data.
2) Estimate the number (κ) of significant contiguous serial correlations.
3) Resample the original series in blocks of length $\kappa + \eta$ until a resampled series is created having a length the same as the original data. In the case of unequal block length, a smaller/larger block will not affect the result of the technique. Implement the resampling procedure a large number of times.
4) Estimate the trend detection test statistic for each resampled series. If the test statistic of the original data falls in the tails of the distribution of resampled test statistics, the result is considered significant.

Sieve Bootstrap (SBS)

Noguchi *et al*. (2011) suggest the use of the AR-sieve bootstrap and provide the methodology required to combine the technique with a trend detection test of choice. The authors' procedure for use with the MK test includes the following:

1) Estimate test statistic of the trend detection test from the original data.
2) Estimate the autoregressive structure [AR(p)] of the data by:

$$x_t = \sum_{k=1}^{p} r_k x_{t-k} + e_t \tag{4-19}$$

Resample the residuals from Equation (4-19) a large number of times, creating a resampled series the same length as the original for which the selected AR(p) filter has been re-added to the residuals of Equation (4-19).

3) Estimate the selected trend detection test statistic for each resampled series.
4) The p-value of the test is given by the number of times that the test statistic of the original data exceeds those calculated through the resampling approach. A p-value less than the pre-assigned significance level indicates a significant result.

4.4 Simulation Study

To determine the type I error rate and power of the aforementioned techniques, a Monte Carlo simulation study is carried out. The simulations include trials with and without serial correlation and also with and without a trend component. Descriptions of the simulation experiments are outlined in the following sections. Additionally, all the techniques are employed with the AR(1) assumption except the BBS approach that selects the block length based on the number of contiguous lags of significant serial correlation.

4.4.1 Series Without Trend

A simulation study is performed to ascertain the type I error rate of the techniques described in section 4.2. An AR(1) model is employed to assess type I error, where the simulated data were generated as follows:

$$X_t = \mu_X + \rho_1(X_{t-1} - \mu_X) + \varepsilon_t \tag{4-20}$$

where X_t and X_{t-1} is the simulated time series at times t and $t-1$, respectively; μ_X is the mean of X_t; and ε_t is a white-noise process with a mean of zero ($\mu_\varepsilon = 0$) and variance (σ_ε^2) equal to $\sigma_X^2(1-\rho_1^2)$.

The simulation results presented herein are comprised of 10,000 AR(1) time series with a mean (μ_X) of 1, coefficient of variation (C_v) of 0.25, 0.5, 0.75, 1, and 1.5 and length, n = 25, 50, 100, 150, and 200. Both positive and negative lag-1 autocorrelation coefficients are examined (ρ_1 = -0.9(0.1)0.9) including simulated time series with no autocorrelation. 2000 bootstrap resamples are generated for the bootstrapped based models (BBS and SBS), which was found to be suitable in

previous studies (Svensson *et al*., 2005; Önöz and Bayazit, 2012). Following the results from section 4.5.2.3, a value of $\eta = 1$ is employed.

Using the significance level, α = 0.05, the type I error rate is assessed for the simulated series. This is determined from the rejection rate (R_{rej}) of the null hypothesis of no trend, which is given as follows (Yue *et al*. 2002b):

$$R_{rej} = N_{rej}/N \tag{4-21}$$

where N_{rej} is the number of simulated series in which the null hypothesis is rejected and N is the total number of simulations.

4.4.2 Series With Trend

A Monte Carlo simulation is carried out using the techniques described in section 4.2 in order to assess the power of each method. An AR(1) model (Equation (4-20)) was used along with the addition of a linear trend ($T_t = \delta t$), superimposed as follows:

$$X_T = X_t + \delta t \tag{4-22}$$

where X_T is the AR(1) model with the additional trend. Linear trends with slopes δ = -0.01(0.002)0.01 were used for the simulated time series. The power of the techniques described in section 4.2 is assessed as the rejection rate of the null hypothesis (of no trend) when applied to series with superimposed trend.

4.5 Simulation Results

In the presence of serial correlation, a desirable characteristic of any statistical tests is the ability to maintain the nominal significance level to be considered a valid approach. Although a number of authors have emphasized the importance of type I error rates close to the assigned significance level (Zhang and Zwiers, 2004; Önöz and Bayazit, 2012), a balance between low type I error and high power must also be a consideration in selecting an appropriate technique. For the sake of brevity, only subsets of the simulated results are shown thus providing an overview of the behaviour of the statistical techniques.

4.5.1 Type I Error

The results of the type I error rates from the Monte Carlo simulations are depicted in Figure 4-1 for positive values of the lag-1 autocorrelation coefficient. As expected, the presence of increasing values of the lag-1 autocorrelation coefficient increase the likelihood of rejecting the null hypothesis of no trend, which is characterized by an increase in the type I error rate (von Storch, 1995; Hamed and Rao, 1998; Yue *et al.*, 2002b; Yue and Wang, 2002; Yue and Wang, 2004a) and changes in the length and C_v of the simulated series have little effect on this result. All the techniques are able to preserve the nominal significance level in the absence of serial correlation, which is particularly apparent with increased record lengths.

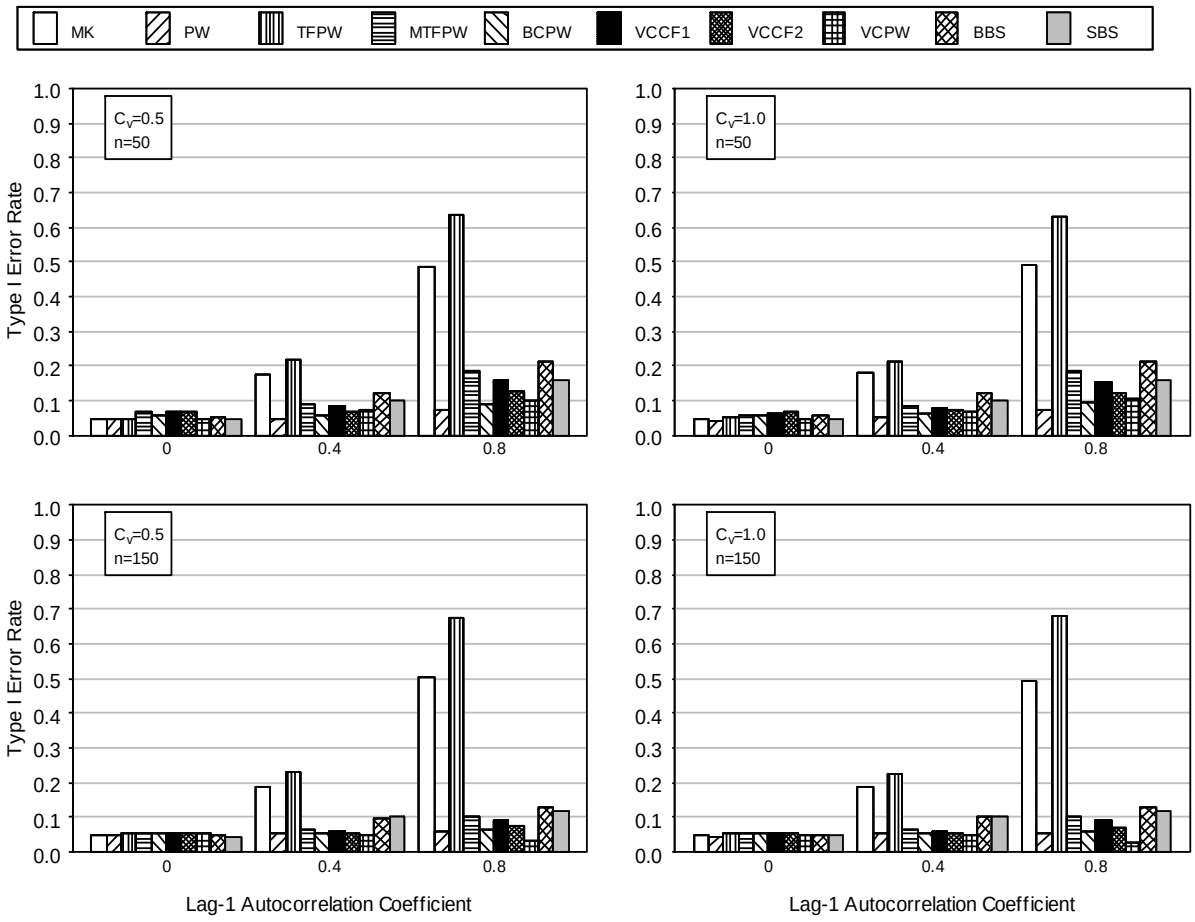

Figure 4-1: Type 1 error rate (rejection rate) attained from the MK test and nine complimentary serial correlation techniques for positive values of serial correlation.

Similar to previous research (Khaliq *et al*., 2009; Blain, 2014; Wang *et al*., 2015), TFPW is unable to preserve the nominal significance level in the presence of an increasing lag-1 serial correlation coefficient. The modified version (MTFPW), proposed by Önöz and Bayazit (2012), is capable of producing lower type I error rates than TFPW but diverges from the nominal significance level when the record length is shorter and as the value of the lag-1 autocorrelation coefficient increases. Analogous to the results of Yue *et al*. (2002b), Figure 4-1 demonstrates that the VCCF1 technique, proposed by Hamed and Rao (1998) produces rejection rates higher than the assigned significance level for shorter record lengths and larger values of the lag-1 autocorrelation coefficient. Several authors have maintained that the VCCF1 approach is incapable of preserving the nominal rejection rate due to its use of the ranks of the data to calculate the serial correlation coefficient (Yue *et al*., 2002b; Yue and Wang, 2004a), and as a result Yue and Wang (2004a) proposed a modified version in which ρ_1 is calculated directly from the data (VCCF2). It is apparent from Figure 4-1, and consistent with the results of Blain (2013), that the VCCF2 technique produces only slightly lower type I error rate than the VCCF1 method for moderate and mild values of ρ_1, independent of the length and $C_{v.}$ However, the rejection rates for the two techniques remain somewhat larger than desired for moderate to high values of ρ_1 for shorter records lengths. Additionally, when data are serially correlated, the two bootstrapping techniques invariably produce rejection rates higher than the assigned significance level. The VCPW technique provides rejection rates close to the nominal significant value, although, this approach results in slightly lower assessment results than the assigned rejection rate when n >= 150 and ρ_1 >= 0.6, for all included values of C_v. It should be noted that the results presented herein are not entirely consistent with those of Wang *et al*. (2015), although, the number of Monte Carlo simulations used in each analysis differ. This research employs 10,000 Monte Carlo simulations, whereas Wang *et al*. (2015) use 2000, therefore, it is likely the results presented herein more closely represent the true average performance of the VCPW technique. Also, in agreeance with several authors (Yue *et al*., 2002b; Blain, 2014; Wang *et al*., 2015) the PW approach results in type I error rates closest to the nominal significance level but at the cost of a marked decrease in power, which will be examined in section 4.5.2. Finally, the BCPW approach produces rejection rates very similar to those of the PW approach, however, the approach results in higher type I error rates when records lengths are shorter (n < 50).

Despite hydrometeorological data commonly exhibiting a positive serial dependence structure, several authors have detected negative values of the lag-1 autocorrelation coefficient (Hamed and

Rao, 1998; Yue and Wang, 2002, Khaliq *et al*., 2009; Blain, 2013). Significant lag-1 serial correlation is also found in this study (discussed in section 4.6.3.1) and is consequently included in the simulation experiments herein. Accordingly, this simulation experiment includes the effects of negative values of the lag-1 autocorrelation coefficient on type I error rates, which are displayed in Figure 4-2. It is evident that the length and C_v of the data have little effect on the rejection rate and that the MK test alone results in rejection rates that decrease in accordance with decreasing values of negative serial correlation. The TFPW, VCPW, BBS, and SBS techniques clearly decrease the type I error rate below the assigned significance level in the presence of negative serial correlation. The remaining approaches have a similar capacity for preserving the nominal significance level even for large values of ρ_1.

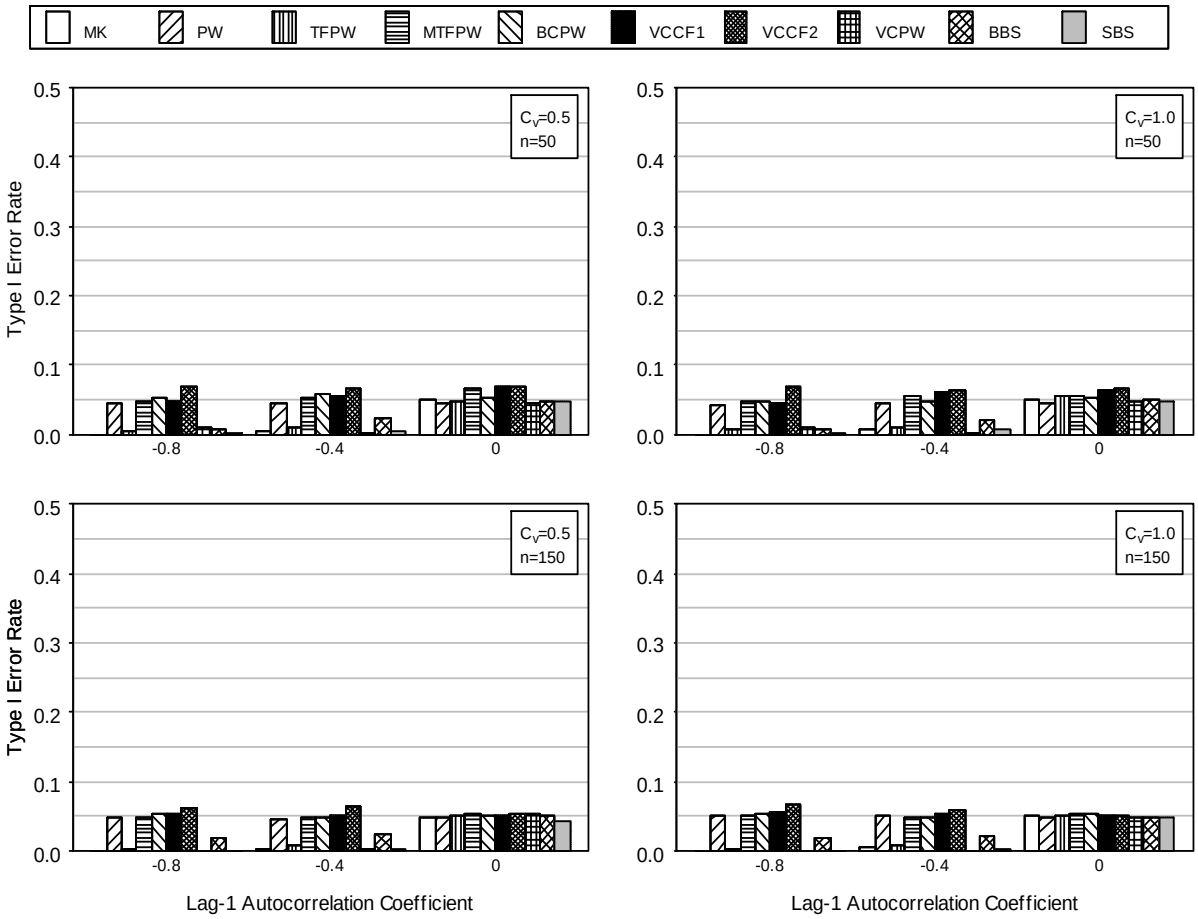

Figure 4-2: Type 1 error rate (rejection rate) attained from the MK test and nine complimentary serial correlation techniques for negative values of serial correlation. Notes: MK test is not shown due to low assessment results.

4.5.2 Power: Series with Positive Serial Correlation and Trend

Figure 4-3 illustrates the effects of the inclusion of increasing trend and positive values of the lag-1 autocorrelation coefficient on the simulated time series. By initially focusing on uncorrelated series (ρ_1 = 0), it is clear that the BBS approach provides higher power than the remaining techniques particularly when the slope and/or record length increase. The remaining approaches result in consistently lower rejection rates regardless of the coefficient of variation, slope, and/or record lengths, however, this result is likely due to the use of the AR(1) assumption when no significant serial correlation is present in the simulated data. Furthermore, in the absence of serial correlation,

when weak increasing trends are present in the data (β = 0.002), and record lengths are shorter (n <= 50), all the techniques have considerably low power. Therefore, accounting for positive serial correlation when none occurs in the data reduces the ability of any approach to detect a weak trend when one is present, as the removal of the AR(1) process effectively removes a portion of the trend (Yue and Wang, 2002). Changes in the coefficient of variation have minimal effect on the ability of any technique to reject the null hypothesis of no trend, again when β = 0.002, n <= 50, and ρ_1 = 0. In accordance with the results of Yue *et al*. (2002a) and Yue and Pilon (2004), in the absence of serial correlation, increases in the slope of trend cause associated increases in the power of all the examined techniques, although this effect is much more prominent when record lengths are longer (n >= 100).

As anticipated, all the statistical approaches have higher power when the record length increases and/or for higher slopes of trends. Figure 4-3 illustrates that when the slope of the trend is weak (β = 0.002) and the record length is shorter (n = 50), there is a direct relationship between increases in power of each assessed method and increasing amounts of serial correlation. In accordance with the results of Yue and Wang (2002), rejection rates deviate from this pattern (i.e., rejection rates decrease with increasing amounts of serial correlation) as the slopes and record lengths increase. Also, the resulting decrease in power associated with increases in C_v becomes more prominent as the record length increases and for higher slope magnitudes. It should be noted that the TFPW technique provides results that are an increasing function of serial correlation in relation to increasing ρ_1 coefficients yet deviations of this pattern are observed as the record length increases along with the slopes of trends. The addition of trend in the simulated series contaminates the estimates of the autocorrelation coefficient, supporting the assertion that trend produces spurious serial correlation (Yue and Wang, 2004b) and serial correlation in the data results in inflated trend estimates. A possible explanation for the increased power of the TFPW technique pertains to the above noted interaction of trend and serial correlation producing overestimated trend magnitudes. The removal of an inflated trend results in an underestimated lag-1 serial correlation coefficient, which is then used to prewhiten the data. Therefore, the data may not have been adequately prewhitened as the estimate of the lag-1 serial correlation coefficient is smaller than the true one. Furthermore, the inflated trend is then added back into the data before applying the MK test. The outcome of an overestimated trend and inadequate prewhitening produces type II error rates that are lower than those of other statistical techniques (Önöz and Bayazit, 2012; Blain, 2014).

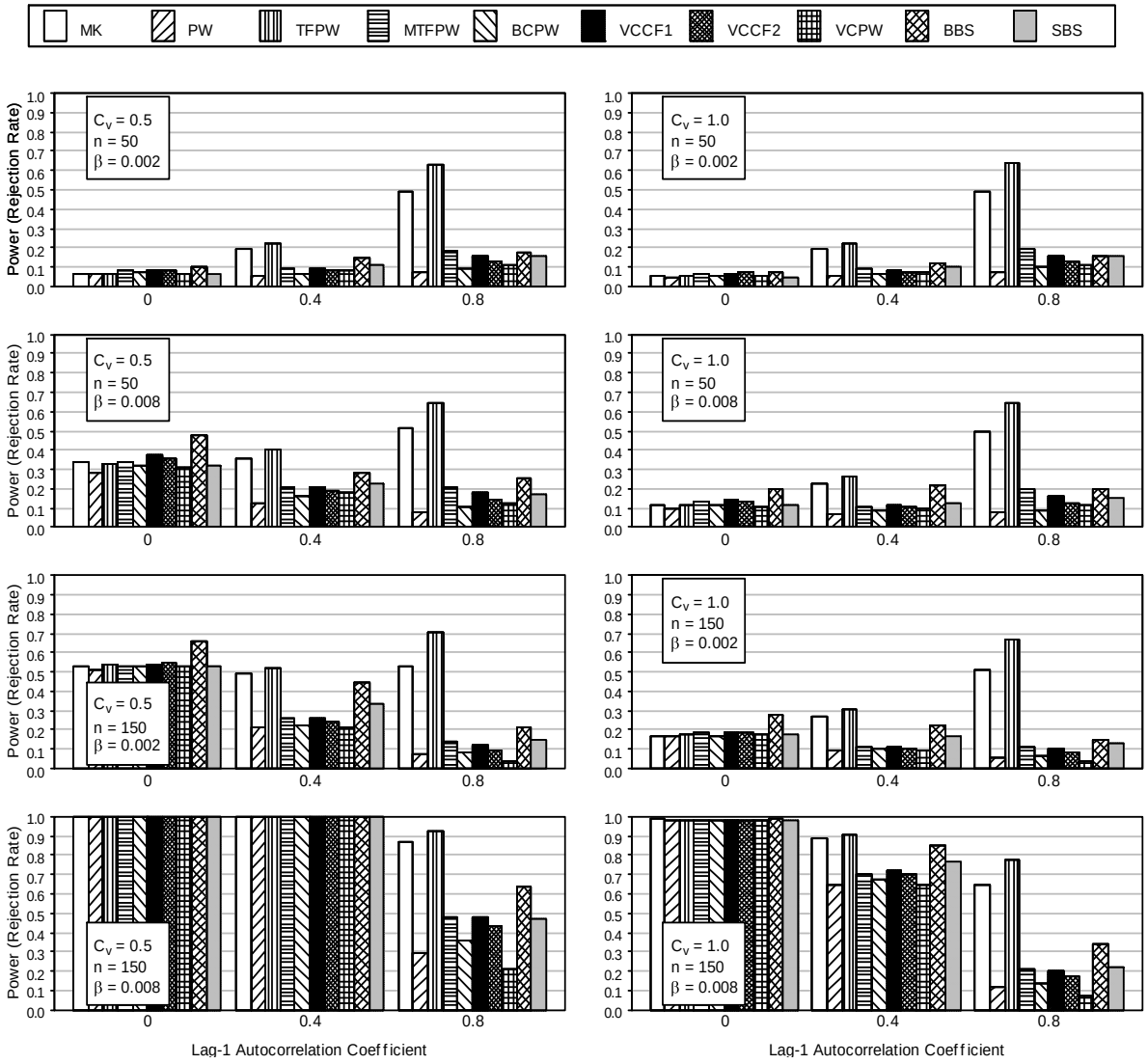

Figure 4-3: Power of the MK test and nine complimentary serial correlation techniques for positive values of serial correlation and increasing trends.

An assessment of Figure 4-3 of the individual approaches demonstrates when $n <= 50$, $\rho_1 > 0$, and for all slope magnitudes, the TFPW approach consistently results in the highest power, although a reason for this has previously been provided and this technique will not be discussed further in this section. For the same record length and moderate values of the lag-1 autocorrelation coefficient ($0.3 > \rho_1 > 0.5$), the BBS technique is more capable of rejecting the null hypothesis of no trend than the remaining techniques, regardless of the magnitude of slope or C_v. The remainder of the statistical

tests produce similar rejection rates, although, the PW, BCPW, and VCPW techniques result in slightly lower rejection rates in comparison. Again when the record length is shorter but when ρ_1 > 0.6, the MTFPW approach has high power for detecting trend, although BBS, VCCF1, and SBS provide similar results. Additionally, the PW approach provides the lowest rejection rates as this approach has been shown to remove a portion of the existing trend (Yue and Wang, 2002).

It is evident that when the record length increases, the ability of all the approaches to detect trend does as well. When the record length increases (n >= 100), and when 0.3 > ρ_1 > 0.5 and β <= 0.004, similar results are observed for all the statistical techniques (aside from TFPW and original MK), with the exception of the BBS and SBS techniques which provide higher rejection rates. For the n >= 100, β <= 0.004, and ρ_1 > 0.6 the BBS remains the most powerful but the SBS and MTFPW techniques also provide high (albeit lower) rejection rates. The effects of C_v on the power of the tests are most obvious when the slope of trend is higher, whereby there is a decreased ability of each approach to detect trends when C_v = 1.0. It is clear that all the statistical techniques are powerful when strong slopes are present in data with long record lengths, although the power is affected by increasing values of ρ_1 and C_v. In agreeance with the results of Yue and Wang (2002), the effects of positive and negative trend magnitude produce similar results on rejection rates when positive serial correlation is present and have consequently been omitted.

Based on the results of the simulations with positive serial correlation, it is clear that the TFPW technique provides type I error estimates that are much greater than the nominal significance level, although it is certainly the most powerful technique. The remaining statistical techniques are reasonably capable of preserving the nominal significance level when the lag-1 autocorrelation coefficient is moderate or low, although, the BBS and SBS techniques consistently have slightly higher results. An examination of the power of the same techniques indicates, however, that the BBS and SBS techniques have higher power when serial correlation estimates are moderate to low.

4.5.3 Power: Series with Negative Serial Correlation and Trend

The impact of negative serial correlation and decreasing trend on the power of the statistical tests designed to account for the effect of serial correlation is presented in Figure 4-4. Regardless of the direction of the trend (positive or negative), in the absence of serial correlation the power of all the statistical techniques is similar, which is demonstrated in Figures 4-3 and 4-4 (Yue and Wang, 2002). When the record lengths are shorter (n <= 50) and in the presence of mild slopes (β <= -0.004), all

the techniques maintain similarly low power with increasing amounts of negative serial correlation (regardless of changes in the coefficient of variation). In accordance with the results of Yue and Wang (2002), the rejection rate of the MK test alone decreases as a function of decreasing serial correlation. It is also apparent that the TFPW, VCPW, and SBS exhibit similar behaviour. The performance of all statistical tests changes with decreasing slope magnitudes (β >= -0.006, n <= 50, C_v = 0.5), whereby, the power of the majority of the tests is considerably higher but there is an obvious decrease with an increase in C_v. Under the same conditions, several tests show increasing power as a function of decreasing lag-1 autocorrelation coefficient magnitudes (i.e., PW, MTFPW, and BCPW). The results of Yue and Wang (2002) demonstrate that the high rejection rates of the PW technique are a result of the trend being inflated by the removal of the Markovian model (i.e., a negative lag-1 serial correlation coefficient). The MTFPW, BCPW, VCCF1, VCCF2, and BBS techniques provide high power estimates for moderate amounts of serial correlation (-0.3 < ρ_1 < -0.5). The variance correction techniques are most powerful for moderate to low values of ρ_1 which are most commonly found in hydrometeorological data. The behaviour of the TFPW, VCPW, and SBS is similar to that of the MK test alone, whereby the likelihood of rejecting the null hypothesis decreases in proportion to decreasing values of the lag-1 autocorrelation coefficient. A possible explanation of the reduced power of the TFPW technique in the presence of negative lag-1 serial correlation could again be due to the contamination of the true slope of trend. If negative autocorrelation results in deflated slope trends then removal of this deflated trend will result in overestimated ρ_1 values. This interaction is reinforced by the results of Rivard and Vigneault (2009) where the authors show that prewhitening with a negative serial correlation coefficient results in a subsequent increase in the estimated slope. Prewhitening with a larger (more negative) ρ_1 will result in a re-inflated trend after which the deflated trend is re-added, resulting in reduced power. Rivard and Vigneault (2009) also demonstrated that the bias in the slope estimates when trend and negative serial correlation are present is lower than when positive serial correlation and trend are present. Therefore, the loss in power of the TFPW technique in the presence of negative serial correlation is less pronounced than the increase in power when positive serial correlation is present, which is demonstrated in Figures 4-3 and 4-4. Furthermore, the SBS technique has a marked decrease in power with increasing ρ_1 values. A possible explanation for this result may be again be due to the interaction of the slope of trend and ρ_1. The reduced power of the SBS technique may be a result of the re-addition of the AR(p) filter, where $r_1$is negative, having an overall deflationary effect on the final slope of trend.

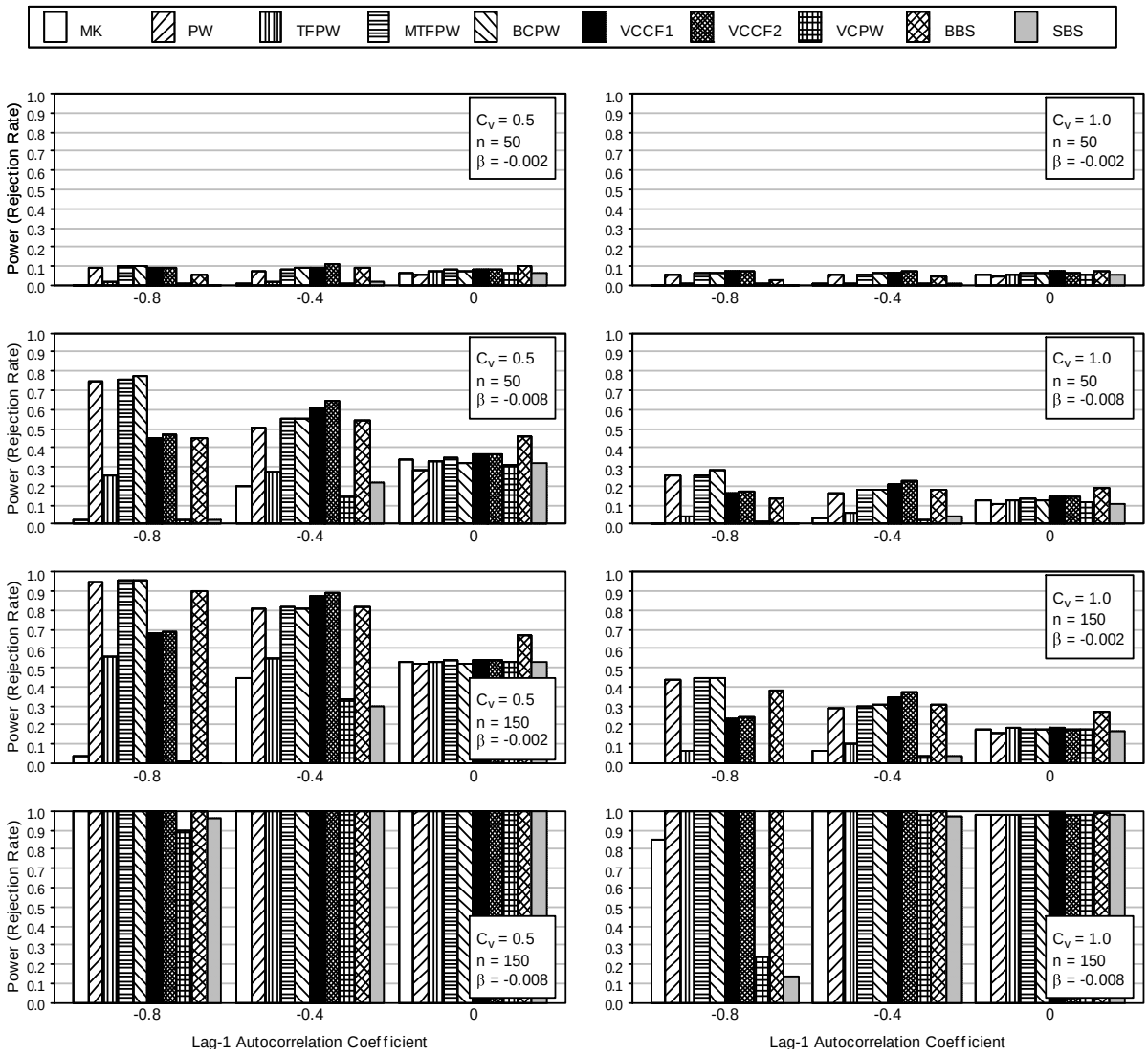

Figure 4-4: Power of the MK test and nine complimentary serial correlation techniques for negative values of serial correlation and decreasing trends.

When the record length increases (n >= 100), the power of the majority of the statistical tests increases along with it. The behaviour noted above is apparent when n >= 100 and β <= -0.004, although, the BBS technique now appears to have increasing power as a function of decreasing values of ρ_1. These characteristics are unchanged with an increase in C_v but a higher coefficient of variation does result in lower power estimates in all instances. Similar to the results presented in Figure 4-3, the impact of serial correlation on each statistical test's ability to detect trend is no longer an issue

when the record length is long and the slope of trend is high, although, when n = 150, β >= -0.006, ρ_1 > -0.7, and C_v >= 0.75, the SBS and VCPW techniques are less capable of detecting significant trend when one exists. Again, for the sake of brevity the results including negative serial correlation and increasing trends were omitted as they were similar to the results presented in Figure 4-4.

In the presence of negative serial correlation, it is established that the TFPW, VCPW, BBS, and SBS techniques cannot maintain the nominal significance level when no trend is present in the data. In the presence of mild trend and short record length, all the included approaches are very limited in their ability to detect the included trend. As the slope and/or record length increase, the power of many techniques does also, but the TFPW, VCPW, and SBS techniques comparatively provide lower rejection rates. When the record length is long and the magnitude of slope great, all techniques appear to be quite powerful, independent of the magnitude of the lag-1 serial correlation coefficient, although an increase in the coefficient of variation greatly affects the VCPW and SBS techniques for low ρ_1 values. It should be noted that, although the simulation experiments give insight into the behaviour of these statistical techniques, they are purely mathematical and may not adequately characterize the correlation structure of real data. Therefore, section 4.6 will examine how the statistical approaches perform on streamflow data gathered from a subset of gauging stations from southern Ontario.

4.5.4 BBS Approach - Selecting η

Khaliq *et al*. (2009) examined the optimal selection of η using annual mean daily flows in the Canadian Reference Hydrometric Basin Network (RHBN). The authors select the optimal block length based on the mean difference between the observed lag-1 autocorrelation coefficient and those estimates from the bootstrapped samples. The authors determined that there was a negligible difference in selecting a near optimal value of η and using η = 1. More recently, Önöz and Bayazit (2012) advocate selecting block length based on those that provide the minimum type I error. Also, the authors demonstrate through simulation of various AR(1) models, that type I error rates and power are not very sensitive to differences in block length. A simulation is carried out to determine if there is merit in choosing an optimal block length based on the minimum type I error over using η = 1 when data are generated by means of a Markovian model, the results of which are presented in Figures 4-5 and 4-6 for n = 50 and n = 150, respectively. Figure 4-5(a) illustrates the difference in block length when it is selected using the minimum type I error and using η = 1. It's clear that the block lengths differ with varying values of the lag-1 autocorrelation coefficient but the difference in

type I error, shown in Figure 4-5(b), is small. Figures 4-5 (c) and (d) elucidate the differences in power with β = 0.002 and β = 0.008, respectively, but again these differences are minor. Although using η = 1 does provides slightly higher type I error rates, it results in slightly higher power for trend detection. Similar results are illustrated in Figure 4-6 when n = 150. Therefore, if the serial structure of the data is indeed an AR(1) process, there appears to be little evidence to warrant the use of a near optimal η given the results from Figures 4-5 and 4-6. Furthermore, the determination of the block length which provides the minimum type I error is a computationally intensive endeavor. Figure 4-7 illustrates the differences in type I and type II error in five additional complex correlation structures when n = 50. Again, there appears to be a similar pattern of slightly higher type I error when using η = 1 accompanied by a nominally higher power when trend is present in the time series. The results for n = 150 are similar to those in Figure 4-7 and have been omitted for the sake of brevity.

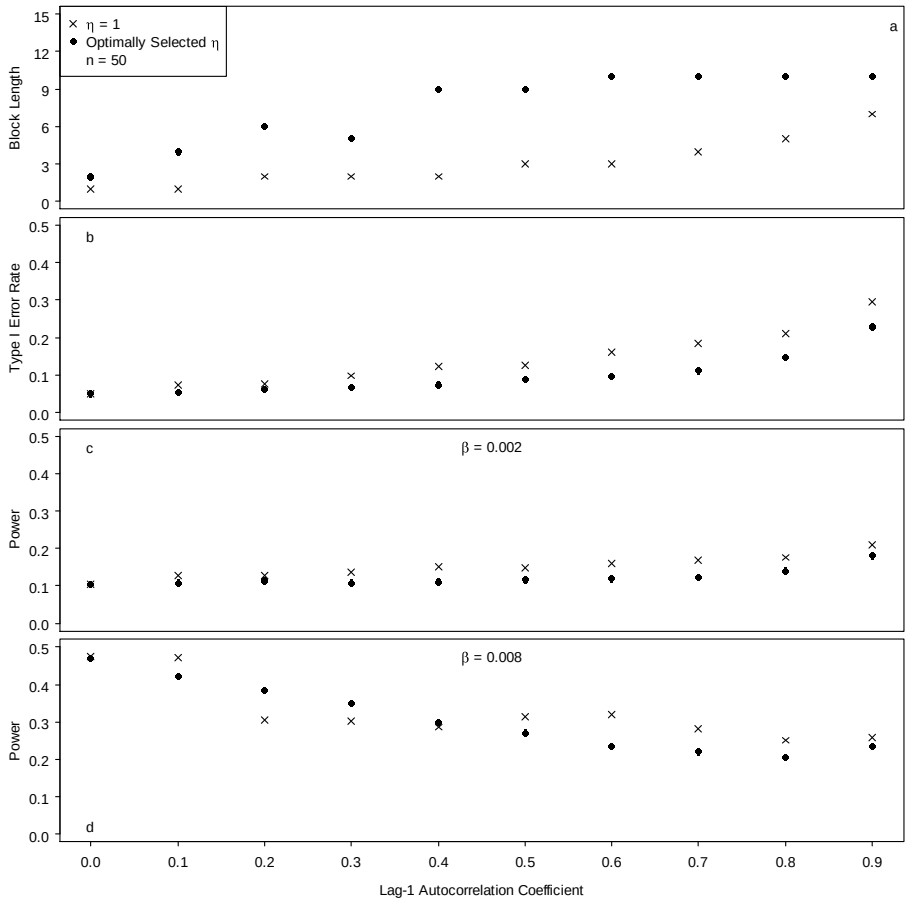

Figure 4-5: (a) Comparison of block length selected by means of the minimum type I error rate using η = 1 (b) Comparison of type I error rate associated with the block length from (a) and the type I error using η = 1 (c) Comparison of the power associated with the block length from (a) and power using η = 1 with $\beta = 0.002$ (d) Comparison of the power associated with the block length from (a) and power using η = 1 with $\beta = 0.008$. All plots have record length n = 50, C_v = 0.5, and data are generated with an AR(1) model.

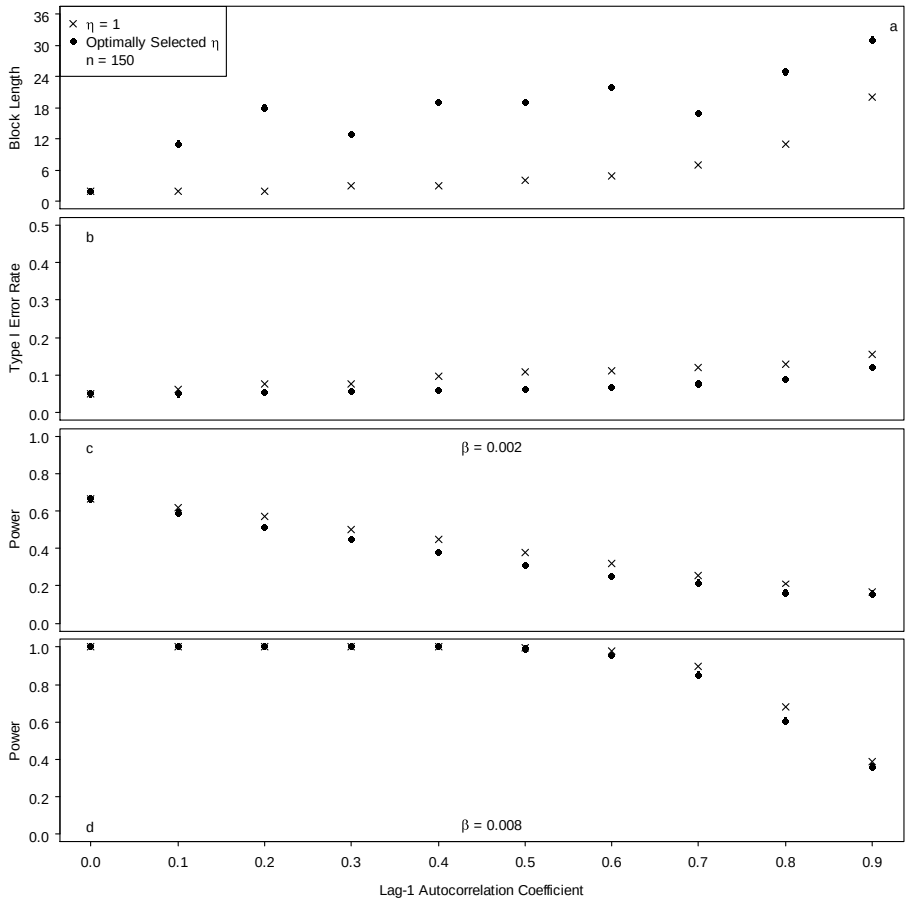

Figure 4-6: (a) Comparison of block length selected by means of the minimum type I error rate using η = 1 (b) Comparison of type I error rate associated with the block length from (a) and the type I error using η = 1 (c) Comparison of the power associated with the block length from (a) and power using η = 1 with $\beta = 0.002$ (d) Comparison of the power associated with the block length from (a) and power using η = 1 with $\beta = 0.008$. All plots have record length n = 150, C_v = 0.5, and data are generated with an AR(1) model.

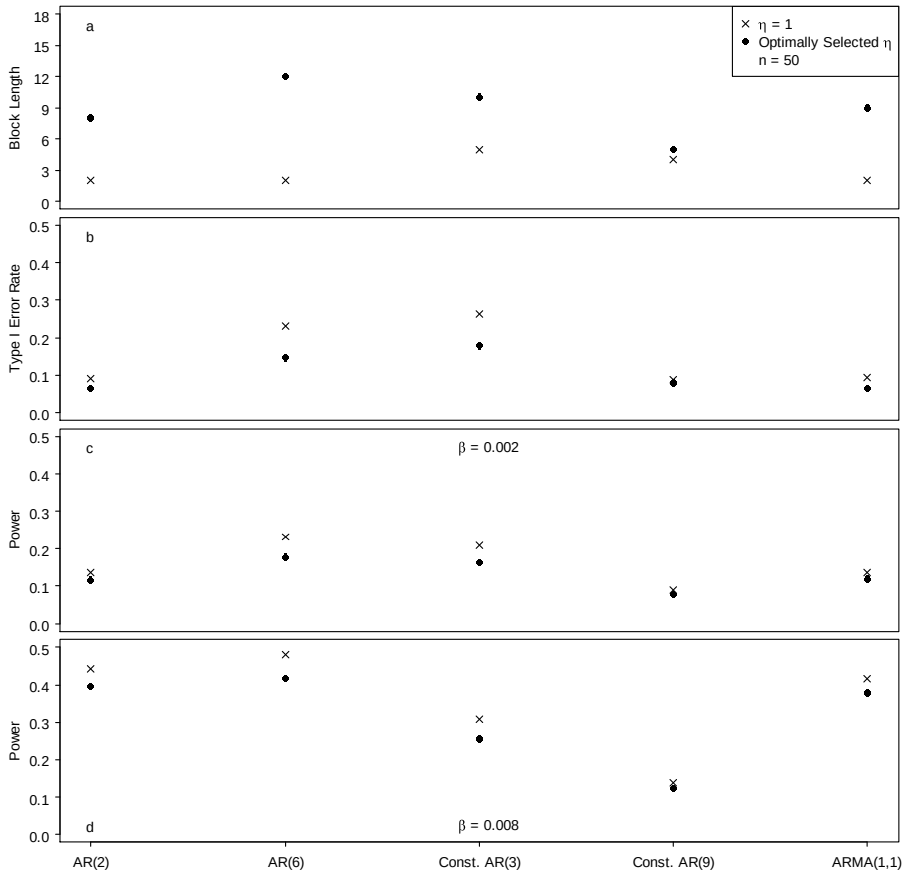

Figure 4-7: (a) Comparison of block length selected by means of the minimum type I error rate using η = 1 (b) Comparison of type I error rate associated with the block length from (a) and the type I error using η = 1 (c) Comparison of the power associated with the block length from (a) and power using η = 1 with $\beta = 0.002$ (d) Comparison of the power associated with the block length from (a) and power using η = 1 with $\beta = 0.008$. All plots have record length n = 150, C_v = 0.5, and data are generated with an AR(1) model. The time series models illustrated are generated through either an AR(p) process where $X_t = \sum_{k=1}^{p} \rho_p X_{t-k} + e_t$ or an ARMA(p,q) process where $X_t = \sum_{k=1}^{p} \rho_p X_{t-k} + \sum_{j=1}^{q} \theta_j e_{t-j} + e_t$. The selected model parameters are as follows: AR(2): $\rho_1 = 0.1$, $\rho_2 = 0.1$; AR(6): $\rho_1 = 0.022$, $\rho_2 = 0.123$, $\rho_3 = 0.12$, $\rho_4 = 0.07$, $\rho_5 = -0.019$, $\rho_6 = 0.219$, Constrained AR(3): $\rho_1 = 0.619$, $\rho_3 = 0.177$; Constrained AR(9): $\rho_1 = 1.325$, $\rho_2 = -0.605$, $\rho_3 = -0.13$; ARMA(1,1): $\rho_1 = 0.4$, $\theta_1 = -0.2$.

4.5.5 Simulation Conclusions

The preceding analyses provide simulation results that explore the power and type I error rates for cases where there is positive or negative lag-1 serial correlation in a dataset. The following conclusions can be drawn with respect to lag-1 autocorrelation present in time series data:

- Increasing positive values of ρ_1 are directly related to increasing type I error rates. In the case of negative values of ρ_1, the PW, MTFPW, BCPW, and VC techniques are more capable of preserving the nominal significance level;
- All the analyzed techniques are able to preserve the nominal significance level in the absence of serial correlation;
- TFPW suffers from increasingly high type I error rates with increasing values of positive lag-1 serial correlation, however, the technique is certainly powerful. The loss of power in the TFPW technique in the presence of negative lag-1 serial correlation is less pronounced than the increase when ρ_1 is positive;
- PW results in type I error rates very close to the nominal level in the case of positive ρ_1 values but at the cost of lower power. In the case of negative $\rho_1$values, PW adequately produces nominal significance levels and is quite powerful;
- In the case of positive serial correlation, all the techniques (aside from TFPW) are reasonably able to maintain the assigned significance level for mild to moderate values of the lag-1 serial correlation coefficient. However, the BBS technique is more powerful than the rest of the techniques regardless of the magnitude of ρ_1;
- The SBS technique has an obvious decrease in power when ρ_1 is negative; and
- Based on the simulation results, there is little evidence to warrant the use of near optimal η values and a value of $\eta = 1$ is sufficient.

4.6 Case Study

Using POT and BMS data from 70 gauging stations from southern Ontario, an analysis of the serial correlation structure and trend for each station is assessed using the techniques described herein. A description of the instantaneous data used for the analysis follows.

4.6.1 Overview of Data

The Water Survey of Canada's (WSC) hydrometric database (HYDAT) provides average daily flow data but knowledge of sub-daily processes, as they apply to the discharge hydrographs, can provide a more detailed understanding of streamflow characteristics. WSC introduced data processing techniques in 1968 that employed the use of computer programs allowing for hydrograph charts to be digitized at sub-hourly intervals. Digitized stream stage records, various rating curves, and any necessary corrections, were then archived in electronic form, thus providing all the required data to replicate the discharge hydrographs. Daily figures, available on the HYDAT database, were calculated using this instantaneous data and published. A computer program (ArkWSC) was written to extract and verify the archived data, which works in tandem with legacy WCS software, the result of which is instantaneous discharge data. Hourly and sub-hourly data from 1996 onward are stored in the WSC's internal CompuMOD database and were extracted and combined with the ArkWSC data. The final hydrometric record is comprised of hourly or sub-hourly (15 min) data spanning 1969-2016 (Thompson, 2013).

The increased temporal resolution of the instantaneous data allows for individual events to be identified on the hydrograph; an algorithm making use of watershed specific characteristics was developed by Thompson (2013) to parse the hydrograph into unique events. Pairs of valleys were identified in the data as local minima in the first derivative of the time series. The algorithm then recursively searches between found valleys for peaks above a predefined threshold. If a significant peak is found, the hydrograph between the corresponding valleys is taken as an event. The algorithm also distinguishes between secondary peaks that may occur during a given event. Event characteristics, including discharge, volume, duration, time to peak, and flashiness (peak event discharge divided by time to peak) can then be readily extracted. For more information regarding the WSC data extraction and the event parsing algorithm, please refer to Thompson (2013).

The observed deviation in the instantaneous and mean daily data is provided in Figure 4-8 through which the value of the instantaneous data is reinforced for peak event discharge. Peak discharges occur for short periods of time, which are not adequately captured in mean daily data. The deviation between the two datasets is most prominent for flashier urban systems but it is still apparent, to a lesser degree, for relatively pristine watersheds. The published WSC AMS corresponds to the instantaneous streamflow measurement, although, this resolution is only available for the peak annual events. Consequently, a great deal of information regarding the peak flow regime of the watersheds is not readily available, as there may be a number of low frequency, high magnitude events annually.

These data are well suited for POT analyses, as events are individually parsed, eliminating the need for declustering. A comparison of the performance of the modified statistical tests is carried out on BMS and POT data.

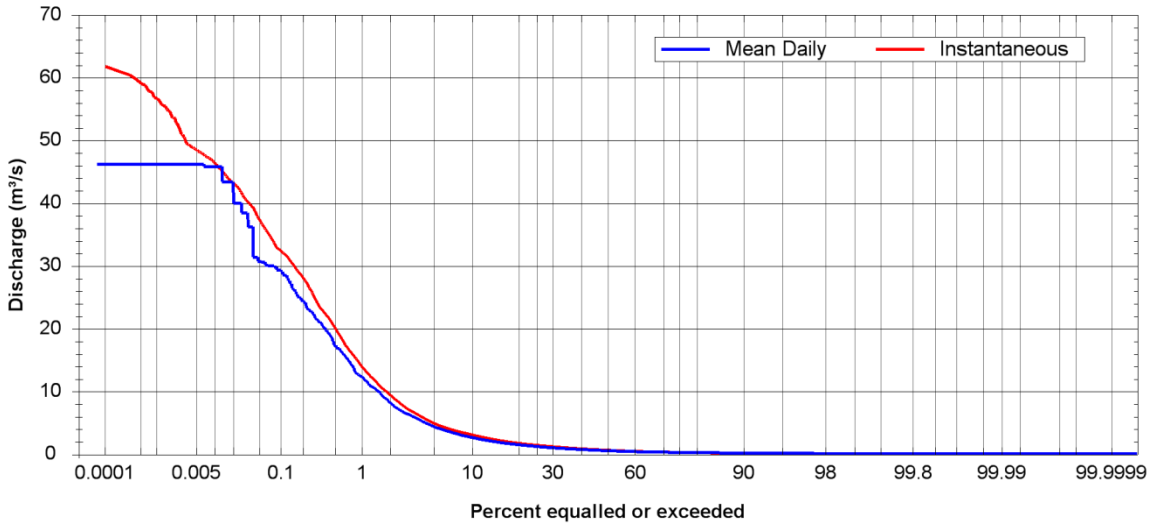

Figure 4-8: Observed occurrence of instantaneous vs. mean daily discharge at station 02HC009, having a drainage area of 191 km^2 (Thompson, 2013).

This analysis focuses on a subset of southern Ontario gauging stations falling within the Great Lakes-St. Lawrence climatic region (Hare and Thomas, 1979), all having a similar period of record. To account for the effects of seasonality in the data and to ensure similar flood generating mechanisms, each dataset is divided into three yearly time frames (January to April, June to September, and October to December). Yearly hydrographs of the data were analyzed to determine the timing of freshet events, which tend to occur in March or April in southern Ontario. As the instantaneous data do not necessarily include the peak freshet events due to ice cover, the January to April event data are omitted from the analysis. Convective storm events are common in the summer months and often result in intense but short periods of runoff, consequently, the June to September (J-S) time frame was selected to encompass such events. Lastly, frontal systems are prominent in the October to December (O-D) time period.

With the use of watershed delineations provided by the WSC and the Ontario Land Cover Compilation v.2.0, three broad land-cover classifications are generated, which include natural,

agricultural, and urban land-uses. The urban classification was carried out firstly, whereby, those catchments having more than 25% urban area were classified as such. Of the remaining stations, those having more than 30% agricultural land-use were classified accordingly. Due to the presence of vast amounts of undisturbed land, the remainder of the watersheds are classified as natural. Seven of the included catchments are part of the RHBN, comprised of watersheds having relatively stable or pristine land-use conditions (Harvey *et al*., 1999). An aerial analysis of each catchment is carried out to determine the level of regulation in each watershed in the study area. Those with highly regulated conditions, such as large reservoir systems and dams, were removed from the analysis. It should be noted, however, that major urban flood control systems may potentially be considered regulated (i.e., large detention storage ponds), although this consideration is encompassed in the urban station classification. Also, records having more than four consecutive years of missing data were removed from the analysis. Key characteristics of each watershed are provided in Table 4-1, including the number of years of data available, location, and land-use classification. A map of the included gauging station locations is provided in Figure 4-9, with reference numbers that correspond to Table 4-1. Threshold selection for the POT data is carried out using several exploratory plots, namely, mean residual life plots, complimentary threshold choice plots (for scale and shape parameters), and dispersion index plots (Cunnane, 1979; Lang *et al*., 1999). An extensive overview of the threshold selection methodology is presented in section 3.2.4.

Table 4-1: Summary of stations used for analysis.

Fig. 4-9 ID	Station ID	Station Name	Start Year	End Year	Latitude (°N)	Longitude (°W)	Area (km^2)	Agricultural (%)	Urban (%)	Natural (%)	Land-Use Classification
1	02DD015	Commanda Creek Near Commanda	1975	2016	45.95	79.61	103.8	3%	0%	97%	Natural
2	02EA005	North Magnetawan River Near Burk's Falls	1971	2015	45.67	79.38	328.8	3%	0%	97%	RHBN
3	02EA010	North Magnetawa River Above Pickerel Lake	1971	2015	45.70	79.31	155.1	2%	0%	98%	Natural
4	02ED003	Nottawasaga River Near Baxter	1969	2016	44.25	79.82	1230.6	73%	6%	21%	Agricultural
5	02ED007	Coldwater River at Coldwater	1969	2016	44.71	79.64	168.5	55%	7%	38%	Agricultural
6	02ED017	Hogg Creek Near Victoria Harbour	1989	2016	44.73	79.78	65.2	65%	3%	33%	Agricultural
7	02ED024	North River at the Falls	1988	2016	44.77	79.58	243.6	42%	8%	50%	Agricultural
8	02FA001	Sauble River at Sauble Falls	1969	2016	44.68	81.26	913.5	53%	3%	44%	Agricultural
9	02FA002	Stokes River Near Ferndale	1976	2016	45.04	81.34	50.5	24%	2%	74%	Natural
10	02FB007	Sydenham River Near Owen Sound	1969	2016	44.52	80.93	183.0	58%	3%	39%	RHBN
11	02FB009	Beaver River Near Clarksburg	1970	2016	44.52	80.47	587.1	58%	3%	39%	Agricultural
12	02FB010	Bighead River Near Meaford	1969	2016	44.57	80.65	298.0	65%	2%	33%	Agricultural
13	02FC001	Saugeen River Near Port Elgin	1972	2016	44.46	81.33	3953.5	68%	3%	29%	RHBN
14	02FC015	Teeswater River Near Paisley	1972	2016	44.27	81.27	669.9	67%	3%	31%	Agricultural
15	02FE003	Middle Maitland River at Listowel	1969	2016	43.73	80.97	73.4	80%	8%	12%	Agricultural
16	02FE009	South Maitland River at Summerhill	1969	2016	43.68	81.54	370.6	87%	3%	10%	Agricultural
17	02FF002	Ausable River Near Springbank	1969	2016	43.07	81.66	865.4	85%	4%	11%	Agricultural
18	02FF004	South Parkhill Creek Near Parkhill	1969	2016	43.16	81.73	42.7	89%	2%	8%	Agricultural
19	02FF007	Bayfield River Near Varna	1969	2016	43.55	81.59	460.4	86%	4%	9%	Agricultural
20	02FF008	Parkhill Creek Above Parkhill Reservoir	1973	2016	43.16	81.63	112.5	87%	2%	11%	Agricultural
21	02GA010	Nith River Near Canning	1970	2016	43.19	80.46	1034.3	81%	5%	14%	RHBN
22	02GA018	Nith River at New Hamburg	1969	2016	43.38	80.71	544.2	85%	4%	11%	Agricultural
23	02GA030	Alder Creek Near New Dundee	1969	2016	43.37	80.55	47.4	76%	12%	12%	Agricultural
24	02GB007	Fairchild Creek Near Brantford	1970	2014	43.15	80.15	388.6	67%	11%	21%	Agricultural
25	02GB008	Whitemans Creek Near Mount Vernon	1970	2016	43.13	80.38	385.9	77%	4%	19%	Agricultural
26	02GC002	Kettle Creek at St. Thomas	1969	2016	42.78	81.21	330.9	82%	7%	11%	Agricultural
27	02GC007	Big Creek Near Walsingham	1969	2016	42.69	80.54	566.7	76%	4%	19%	Agricultural
28	02GC010	Big Otter Creek at Tillsonburg	1969	2016	42.86	80.72	354.1	81%	5%	14%	Agricultural
29	02GC018	Catfish Creek Near Sparta	1969	2016	42.75	81.06	294.5	84%	6%	10%	Agricultural
30	02GC031	Dodd Creek Below Paynes Mills	1987	2016	42.79	81.27	99.6	87%	5%	9%	Agricultural
31	02GD004	Middle Thames River at Thamesford	1969	2016	43.06	80.99	306.0	83%	4%	13%	Agricultural
32	02GD021	Thames River at Innerkip	1978	2016	43.22	80.69	148.9	86%	4%	10%	Agricultural
33	02GG002	Sydenham River Near Alvinston	1969	2016	42.83	81.85	701.2	82%	5%	14%	Agricultural
34	02GG003	Sydenham River at Florence	1984	2016	42.65	82.01	1149.3	82%	4%	14%	Agricultural
35	02GG006	Bear Creek Near Petrolia	1969	2016	42.91	82.12	248.7	84%	3%	13%	Agricultural
36	02GG009	Bear Creek Below Brigden	1981	2016	42.81	82.30	535.6	83%	4%	13%	Agricultural
37	02GH002	Ruscom River Near Ruscom Station	1971	2016	42.21	82.63	125.0	94%	4%	3%	Agricultural
38	02GH003	Canard River Near Lukerville	1977	2016	42.16	83.02	159.0	86%	8%	6%	Agricultural
39	02GH004	Turkey Creek at Windsor	1982	2011	42.26	83.04	29.6	2%	94%	4%	Urban
40	02HA006	Twenty Mile Creek at Balls Falls	1970	2016	43.13	79.38	291.7	77%	9%	13%	Agricultural
41	02HB004	East Sixteen Mile Creek Near Omagh	1969	2016	43.50	79.78	193.0	68%	16%	16%	Agricultural
42	02HB012	Grindstone Creek Near Aldershot	1970	2016	43.30	79.87	77.9	57%	15%	28%	Agricultural

Table 4.1 (continued): Summary of stations used for analysis.

Fig. 4-9 ID	Station ID	Station Name	Start Year	End Year	Latitude (°N)	Longitude (°W)	Area (km^2)	Agricultural (%)	Urban (%)	Natural (%)	Land-Use Classification
43	02HB018	Credit River at Boston Mills	1982	2016	43.77	79.93	414.7	60%	12%	28%	Agricultural
44	02HB020	Credit River Erin Branch Above Erin	1983	2016	43.77	80.09	32.3	71%	7%	22%	Agricultural
45	02HC009	East Humber River Near Pine Grove	1969	2016	43.79	79.58	190.9	56%	21%	23%	Agricultural
46	02HC018	Lynde Creek Near Whitby	1969	2016	43.88	78.96	100.3	63%	17%	20%	Agricultural
47	02HC022	Rouge River Near Markham	1969	2016	43.86	79.23	181.3	38%	51%	12%	Urban
48	02HC024	Don River at Todmorden	1969	2016	43.69	79.36	318.5	12%	80%	9%	Urban
49	02HC025	Humber River at Elder Mills	1969	2016	43.81	79.63	296.3	60%	10%	30%	Agricultural
50	02HC027	Black Creek Near Weston	1969	2016	43.67	79.50	58.0	9%	87%	4%	Urban
51	02HC028	Little Rouge Creek Near Locust Hill	1969	2016	43.91	79.22	83.6	74%	14%	13%	Agricultural
52	02HC030	Etobicoke Creek Below Queen Elizabeth Highway	1969	2016	43.60	79.56	205.0	31%	63%	6%	Urban
53	02HC031	West Humber River at Highway No. 7	1969	2016	43.76	79.68	142.2	70%	20%	10%	Agricultural
54	02HC033	Mimico Creek at Islington	1969	2016	43.65	79.52	67.8	9%	88%	3%	Urban
55	02HC049	Duffins Creek at Ajax	1989	2016	43.85	79.06	257.5	64%	11%	25%	Agricultural
56	02HD003	Ganaraska River Near Osaca	1969	2016	44.02	78.44	67.3	51%	3%	45%	Agricultural
57	02HD008	Oshawa Creek at Oshawa	1969	2016	43.93	78.89	95.8	74%	11%	15%	Agricultural
58	02HD009	Wilmot Creek Near Newcastle	1969	2016	43.93	78.62	80.7	66%	8%	26%	Agricultural
59	02HD012	Ganaraska River Above Dale	1977	2016	43.99	78.33	241.9	58%	3%	39%	Agricultural
60	02HD013	Harmony Creek at Oshawa	1980	2012	43.89	78.82	42.9	41%	48%	10%	Urban
61	02HJ001	Jackson Creek at Peterborough	1969	2016	44.30	78.32	116.2	57%	11%	32%	Agricultural
62	02HK007	Cold Creek at Orland	1981	2016	44.13	77.79	160.5	58%	4%	39%	Agricultural
63	02HL003	Black River Near Actinolite	1969	2016	44.54	77.37	428.7	5%	1%	94%	Natural
64	02HL004	Skootamatta River Near Actinolite	1969	2016	44.55	77.33	677.7	3%	0%	97%	RHBN
65	02KF011	Carp River Near Kinburn	1971	2016	45.42	76.20	258.5	48%	14%	37%	Agricultural
66	02LA007	Jock River Near Richmond	1970	2016	45.25	75.79	526.1	45%	6%	50%	Agricultural
67	02LB006	Castor River at Russell	1969	2016	45.26	75.34	438.7	64%	9%	27%	Agricultural
68	02LB007	South Nation River at Spencerville	1971	2016	44.84	75.54	246.0	40%	3%	57%	RHBN
69	02LB008	Bear Brook Near Bourget	1977	2016	45.43	75.15	448.2	56%	8%	36%	Agricultural
70	02MC026	Riviere Beaudette Near Glen Nevis	1983	2016	45.27	74.49	132.6	56%	3%	41%	Agricultural

Figure 4-9: Location of the 70 gauging stations used for the analysis.

4.6.2 Serial Correlation Structure of the Data

The serial correlation structure of all POT and BMS datasets (extracted from the instantaneous data) is ascertained through standard ARMA (autoregressive-moving-average) modelling techniques. Preliminary model selection is carried out though an examination of the sample autocorrelation function (ACF) plot, partial autocorrelation function (PACF) plot, inverse autocorrelation function (IACF) plot, and the inverse partial autocorrelation function (IPACF) plot (Hipel and McLeod, 1994). Appropriate models are then selected with the use of the minimum Akaike information criterion (AIC) (Akaike, 1974). Several goodness-of-fit criteria are used to assess the suitability of the selected model to confirm the whiteness of residuals and the assumption of linearly weakly stationary data. Both the residual and squared residual ACF plots are examined for evidence of white noise. Due to the segmentation of each time series into seasonal periods, no evidence of seasonality is detected in any residual ACF plot. The goodness-of-fit of each ARMA model was assessed using the Ljung-Box Portmanteau test. Each model is also evaluated for autoregressive conditional heteroscedasticity (ARCH) and other nonlinearities using the McLeod-Li test (McLeod and Li, 1983). All hypothesis testing was carried out at the 5% significance level.

4.6.3 Trend Analysis

This section employs the previously noted techniques on both BMS and POT data. Data are analyzed for the J-S and O-D time periods separately and initially tested for change-points using Pettitt's test (Pettitt, 1979). Table 4-2 provides an overview of those sites having significant serial dependence structures (ARMA), the series length, coefficient of variation, standardized slope of trend, lag-1 autocorrelation coefficient, and the dependence structures of each site after detrending. The AR(p) dependence structures are added as they're needed for the implementation of the SBS technique. Furthermore, as recommended by Cabilio *et al*. (2013), the SBS and BBS techniques should be detrended before implementation when the precise model error is unknown. Therefore, the detrended serial structures of the data have been included in Table 4-2 as well.

Table 4-2: Summary of BMS and POT characteristics.

	Season	Station ID	Dependence Structure	AR(*p*) Dependence Structure	n	C_V	Slope of Trend	r_1	Detrended Dependence Structure	Detrended AR(*p*) Dependence Structure
BMS	Jun.-Apr.	02ED024	AR(1)		29	0.568	-0.001	0.359	AR(1)	
		02GC010	AR(1)		41	0.734	0.005	0.365	AR(1)	
		02GG002	AR(1)		39	0.921	0.007	-0.378	AR(1)	
		02HA006	AR(1)		35	0.871	0.003	-0.471	AR(1)	
		02HC024	AR(1)		48	0.386	0.012	0.415	NS	NS
	Oct.-Dec.	02FA002	AR(1)		33	0.548	-0.007	-0.374	MA(1)	AR(1)
		02FF004	ARMA(1,1)	AR(1)	48	0.710	0.002	0.334	ARMA(1,1)	AR(1)
POT	Jun.-Apr.	02DD015	AR(1)		99	0.6	0.00071	0.195	AR(1)	
		02FA001	MA(1)	AR(1)	76	0.69	0.00139	0.311	MA(1)	AR(1)
		02FA002	AR(1)		60	0.93	0.00312	0.34	AR(1)	
		02FB007	AR(1)		172	0.8	-0.0002	0.199	AR(1)	
		02FB009	AR(1)		109	0.72	0.00084	0.301	AR(1)	
		02FC001	AR(1)		66	0.49	0.00016	0.234	AR(1)	
		02FC015	AR(1)		61	0.87	-0.0039	0.438	AR(1)	
		02FF002	AR(1)		85	0.89	0.00113	0.23	AR(1)	
		02FF004	ARMA(1,1)	AR(1)	84	0.98	-0.0006	0.241	AMRA(1,1)	AR(1)
		02GH003	AR(1)		44	0.49	0.01169	0.333	NS	NS
		02HB004	ARMA(1,1)	AR(1)	60	0.91	-0.0103	0.253	NS	NS
		02HC009	AR(2)		116	0.71	-0.0006	0.206	AR(2)	
		02HC018	AR(1)		116	0.84	0.00061	0.354	AR(1)	
		02HC022	MA(2)	AR(1)	84	0.68	-0.0002	0.474	MA(2)	AR(1)
		02HC025	AR(1)		113	0.78	0.00017	0.373	AR(1)	
		02HC027	AR(1)		145	0.41	0.00043	0.232	AR(1)	
		02HC028	AR(1)		51	0.52	6.8E-05	0.323	AR(1)	
		02HC030	AR(1)		65	0.27	0.0004	0.412	AR(1)	
		02HD013	AR(1)		55	0.25	0.00091	0.259	AR(1)	
	Oct.-Dec.	02FB009	AR(1)		48	0.46	-0.0023	0.267	AR(1)	
		02FE003	MA(1)	AR(1)	45	0.63	-0.0014	0.266	NS	NS
		02FF002	MA(1)	AR(1)	41	0.4	0.00476	0.356	MA(1)	AR(1)
		02HC025	AR(1)		91	0.67	0.00035	0.294	AR(1)	
		02HC027	AR(1)		91	0.43	0.00159	0.193	NS	NS
		02HC033	ARMA(1,1)	AR(1)	156	0.45	0.00113	0.146	NS	NS

- NS indicates that the series was found to have a non-significant serial structure.

It is clear that the BMS data are more likely to have AR(1) dependence structures but also the occurrence of negative serial correlation is limited to these data. The POT data are dominated by positive r_1 values and although the AR(1) model is quite common, these data are more likely to have more complex serial dependence structures. As expected, detrending the data affects the serial structure, in some cases removing any significant serial correlation the data initially may have had. All instances where this occurred are denoted by an 'NS' in Table 4-2.

Records with Significant Serial Correlation

Table 4-3 provides an overview of the BMS records in both the J-S and O-D time frames and the associated p-values of all the tests analyzed herein. Very low p-values would indicate a very powerful result. Four stations have positive r_1 values, which include, 02ED024 (J-S), 02GC010 (J-S), 02HC024 (J-S), and 02FF004 (O-D). The results of station 02HC024 (J-S) indicate the likely existence of a significant trend from several of the applied techniques. As expected, the original MK test is powerful in detecting the trend in the series, with the TFPW technique being slightly more powerful due to the likely inflation of the true slope of trend. The BBS approach also provides a similar p-value as the TFPW approach, indicating the technique's strength in detecting the trend at this site. The VCCF1 approaches are more powerful in detecting significant trend compared to the VCCF2 method under the AR(1) assumption, however the result after including multiple lags ((*n*-1)/4) is more powerful. The SBS technique is capable of detecting a significant trend at the 5% significance level, however the method's power to detect the trend is slightly lower than the aforementioned techniques. As expected, the PW approach is less powerful and fails to detect a trend at the 5% or 10% significance levels as a portion of the existing trend was likely removed. Additionally, the MTFPW, BCPW, and VCPW methods fail to detect a significant trend at the 5% significance level. The detrended BBS approach provides a similar result to that of the original BBS approach. Furthermore, detrending affects the magnitude of the lag-1 serial correlation coefficient, causing it to no longer be significant; therefore, the results of the detrended SBS approach are omitted. The results of site 02GC010 (J-S) are generally consistent with those of 02HC024 (J-S), although, no significant trend is found through the use of any technique. It is apparent, however, that the detrended SBS and BBS approaches provide very similar results to those of the original tests even when the slope of trend is relatively steep. An examination of 02ED024 (J-S) indicates that in the presence of a very mild slope, even when the autocorrelation structure is AR(1) numerous techniques provide similarly high p-values, indicating a lack of trend at this site. However, the PW, TFPW, MTFPW, BCPW, and VCPW methods provide lower assessment results than the remaining techniques. Again, similar results are found at site 02FF004 (O-D) notwithstanding its ARMA(1,1) dependence structure. As an alternative to AR(1) prewhitening, site 02FF04 (O-D) is also prewhitened with its true ARMA(1,1) dependence structure. The results of this modified prewhitening indicate that there is a slight increase in power when the true serial structure of the data is accounted for, however, these results are not wholly consistent with those of Table 4-4.

It is worth noting that almost half of the gauging stations listed in Table 4-3 have a significant negative r_1 value. Stations 02GG002 (J-S), 02HA006 (J-S), 02FA002 (O-D) have negative significant r_1 values in addition to true AR(1) processes. All these stations that have negative lag-1 autocorrelation have moderate to steep slope estimates. The original MK test is much less likely to detect a significant trend, if one does exist based on the high p-values of this method, which is the expected result when r_1 is negative. Lower p-values for the PW method are likely due to the addition of trend due to the negative lag-1 autocorrelation coefficient, although this result is inconsistent at site 02HA006 (J-S). Additionally, the MTFPW and BCPW approaches provide similar results to the PW technique. The results of the TFPW, VCPW, and SBS (original and detrended) techniques tend to be less powerful in detecting trend which is consistent with the simulation results. When the VCCF1 approach is applied to all significant lags the technique becomes slightly less powerful than when applied with the AR(1) assumption. This could, however, be due to the true serial structure of these sites being AR(1). There is an issue when applying the VCCF1 techniques with significant lags considered at site 02HA006 (J-S). It appears that several significant lags of negative serial correlation cause the correction factor to become negative. This in turn causes the MK variance to become negative and therefore, the technique fails in this instance. It should be noted that this error does not occur when the variance correction techniques are carried out under the AR(1) assumption. The VCCF2 technique when multiple lags are considered does not appear to suffer from this limitation in this instance, however, the p-value is comparatively very small. The low assessment result of the VCCF2 techniques when multiple lags are included is a potential indicator that this method may suffer from high Type I error rates, which was confirmed with simulation experiments (not included). Overall, the BBS and variance correction approaches provide the highest power for trend detection.

Table 4-3: Summary of techniques on BMS data.

	Station ID	MK	PW*	PW	TFPW*	MTFPW*	BCPW*	VCCF1*	VCCF1 (all sig. lags)	VCCF2 *	VCCF2 (up to lag (*n*-1)/4)	VCPW*	BBS	BBS Detrended	SBS	SBS Detrended
		p	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***
Jun.-Apr.	02ED024	0.985	0.621		0.594	0.621	0.707	0.990	0.987	0.990	0.974	0.621	0.955	0.979	0.988	0.987
	02GC010	0.363	0.807		0.616	0.771	0.880	0.375	0.363	0.524	0.108	0.771	0.377	0.391	0.453	0.468
	02GG002	0.418	0.078		0.131	0.092	0.097	0.133	0.375	0.218	***0.006***	0.152	0.254	0.258	0.392	0.384
	02HA006	0.589	0.700		0.594	0.700	0.700	0.356	-	0.382	0.094	0.553	0.465	0.404	0.549	0.548
	02HC024	***0.001***	0.169		***0.005***	0.061	0.081	***0.007***	***0.001***	***0.012***	***0.000***	0.061	***0.003***	***0.005***	***0.040***	-
Oct.-Dec	02FA002	0.588	0.372		0.486	0.372	0.390	0.452	0.588	0.426	0.157	0.549	0.481	0.467	0.574	0.582
	02FF004	0.776	0.942	0.880	0.985	0.942	0.898	0.825	0.842	0.839	0.695	0.942	0.816	0.829	0.777	0.795

*Indicates techniques applied under the AR(1) assumption.
-Indicates that no result was found.
Note: Significant results are bolded and italicized.

Table 4-4: Summary of techniques on POT data.

	Station ID	MK	PW*	PW	TFPW*	MTFPW*	BCPW*	VCCF1*	VCCF1 (all sig. lags)	VCCF2*	VCCF2 (up to lag (*n*-1)/4)	VCPW*	BBS	BBS Detrended	SBS	SBS Detrended
		p	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***	***p***
Jun.-Apr.	02DD015	0.551	0.476		0.425	0.496	0.450	0.639	0.644	0.624	0.123	0.496	0.600	0.565	0.548	0.517
	02FA001	0.417	0.431	0.587	0.328	0.431	0.498	0.508	0.417	0.554	0.059	0.453	0.432	0.422	0.363	0.363
	02FA002	0.379	0.539		0.340	0.530	0.539	0.530	0.493	0.533	***0.032***	0.530	0.433	0.452	0.368	0.352
	02FB007	0.598	0.663		0.612	0.661	0.720	0.669	0.596	0.666	0.200	0.663	0.607	0.634	0.576	0.578
	02FB009	0.271	0.646		0.430	0.635	0.612	0.367	0.398	0.421	0.180	0.658	0.325	0.294	0.241	0.236
	02FC001	0.868	0.816		0.843	0.808	0.816	0.879	0.838	0.896	0.754	0.808	0.878	0.819	0.863	0.850
	02FC015	0.139	0.259		0.118	0.254	0.293	0.255	0.272	0.340	0.076	0.270	0.193	0.170	0.176	0.217
	02FF002	0.497	0.929		0.820	0.923	0.923	0.542	0.447	0.590	0.281	0.929	0.522	0.488	0.513	0.502
	02FF004	0.799	0.648	0.719	0.626	0.665	0.665	0.807	0.799	0.841	0.804	0.665	0.784	0.797	0.798	0.808
	02GH003	***0.019***	0.070		***0.015***	0.053	0.070	0.054	***0.014***	0.064	***0.000***	0.062	***0.043***	***0.044***	***0.041***	
	02HB004	***0.000***	***0.025***	***0.003***	***0.001***	***0.004***	***0.008***	***0.000***	***0.000***	***0.000***	***0.000***	***0.005***	***0.001***	***0.000***	***0.002***	
	02HC009	0.343	0.950		0.835	0.958	0.950	0.337	0.394	0.436	0.407	0.965	0.339	0.335	0.585	0.581
	02HC018	0.468	0.858		0.742	0.858	0.950	0.495	0.436	0.615	0.119	0.881	0.482	0.489	0.414	0.438
	02HC022	0.892	0.906	0.844	0.838	0.906	0.931	0.911	0.857	0.935	0.867	0.912	0.901	0.894	0.867	0.887
	02HC025	0.738	0.645		0.723	0.645	0.649	0.807	0.792	0.820	0.593	0.635	0.789	0.759	0.703	0.717
	02HC027	0.296	0.480		0.404	0.476	0.510	0.367	0.296	0.408	***0.008***	0.482	0.331	0.324	0.302	0.287
	02HC028	0.981	0.854		0.854	0.854	0.920	0.984	0.987	0.986	0.980	0.854	0.980	0.978	0.978	0.985
	02HC030	0.768	0.876		0.931	0.876	0.940	0.826	0.814	0.848	0.555	0.876	0.754	0.839	0.747	0.745
	02HD013	0.504	0.788		0.709	0.788	0.893	0.532	0.504	0.606	0.146	0.788	0.518	0.506	0.530	0.527
Oct.-Dec.	02FB009	0.315	0.582		0.389	0.595	0.533	0.347	0.315	0.445	0.061	0.595	0.347	0.322	0.294	0.293
	02FE003	0.632	0.524	0.618	0.448	0.524	0.693	0.663	0.632	0.714	0.245	0.524	0.665	0.632	0.604	
	02FF002	0.103	0.279	0.141	0.213	0.289	0.395	0.239	0.202	0.242	***0.002***	0.289	0.142	0.157	0.193	0.195
	02HC025	0.696	0.770		0.851	0.786	0.764	0.696	0.660	0.772	0.486	0.780	0.680	0.704	0.706	0.723
	02HC027	0.124	0.499		0.309	0.460	0.495	0.160	0.221	0.191	***0.042***	0.460	0.138	0.148	0.244	
	02HC033	***0.010***	0.063	***0.023***	***0.029***	0.051	0.052	***0.015***	***0.007***	***0.021***	***0.001***	0.052	***0.020***	***0.009***	***0.025***	

*Indicates techniques applied under the AR(1) assumption.

Note: Significant results are bolded and italicized.

Table 4-4 provides a summary of the p-values of the POT data with significant dependence structures. Of the sites with true AR(1) serial dependence, few have strong slopes of trends. Several techniques detect trend significance at the 5% level at site 02GH003 (J-S). As expected, the TFPW approach is the most powerful, although the BBS (original and detrended), SBS, and variance correction techniques (with several lags considered) are also able to detect significant trends. The remainder of the techniques fail to detect significant trends at the 5% significance level. Of the sites with weaker trends, p-values of the PW method are consistently higher than those of the original MK test likely due to the removal of a portion of the existing trend (e.g., 02FB009 (J-S), 02HC009 (J-S), 02HC02d (O-D), etc.). Slightly lower p-values are provided by the TFPW method, indicating an increase in the technique's power to detect trend over the PW approach. Furthermore, the MTFPW and BCPW approaches appear to have similar power as the PW approach. Generally speaking, the VCCF2 method with lags (n-1)/4 considered is more powerful than when the AR(1) assumption is employed but again, this is likely due to the method's high type I error rates. Overall, the BBS, SBS, and variance correction techniques (including those that use the AR(1) assumption) are more powerful even when the true dependence structure is not AR(1).

The remainder of the sites in Table 4-4 have more complex serial correlation structures. Of those, 02HC009 (J-S) is the only site with higher order AR dependence. In this instance, the SBS techniques are outperformed by the VCCF1/VCCF2 techniques and BBS techniques. Noguchi *et al*. (2011) found reduced power for a higher order AR(p) model, stating that the slower decaying ACF leads to slower convergence rates of bootstrap approximations, a result that is also apparent herein. Significant trends were found at sites 02HB004 (J-S) at the 5% significance level using all the tested methods, likely due to the steepness of the estimated slope of trend at that site. The detrended BBS approach and all variance correction approaches are the most powerful in detecting the trend at this site, although all the techniques have considerably high power for trend detection. Similar results are found at site 02HC033 (O-D), however, the PW, MTFPW, BCPW, and VCPW methods fail to detect this weaker trend at the 5% significance level. For all the included gauging stations in Table 4-4, the detrended and original SBS and BBS techniques provide similar results. However, in accordance with Yue and Wang (2004a) the detrending procedure removes some of the existing autocorrelation, resulting in non-significant dependence structures at several sites. PW is applied to those sites having non AR(1) dependence structures, however, there is no consistent improvement in assessment results in these instances. This result may be due to sampling variations and bias in the estimates of the AR parameters (Hamed, 2009), resulting in residuals that may not be sufficiently white.

In all applicable instances, the SRC test was also used following the recommendations of Khaliq *et al*. (2009) and Sonali and Kumar (2013), however, these results are omitted for the sake of brevity. Very minor differences were found between the MK and SRC test results, thus, the overall conclusions of Tables 4-3 and 4-4 remain similar when either nonparametric test is employed. The following conclusions can be drawn from the serially correlated data:

- In agreeance with numerous authors (von Storch, 1995; Hamed and Rao, 1998; Yue *et al*., 2002b; Yue and Wang, 2004a; Bayazit and Önöz, 2007; Khaliq *et al*., 2009; Önöz and Bayazit, 2012; Blain, 2013; Sonali and Kumar, 2013; Wang *et al*., 2015), when the MK approach alone is applied to positively serially correlated data, there is a greater chance of detecting a non-significant or weak trend. The opposite is true when significant negative serial correlation is present in the data;
- VCCF1 doesn't perform as expected when all significant lags of serial correlation are included if numerous lags of negative serial correlation are present. Even when the true dependence structure is not AR(1) using the VCCF1 technique under this dependence structure is more appropriate in the presence of negative serial correlation. Using all significant lags of the autocorrelation function when $r_1$is positive offers comparable power to its AR(1) counterpart. In agreement with several authors (Khaliq *et al*, 2009; Blain, 2013), the VCCF1 and VCCF2 techniques under the AR(1) assumption provide comparable results, however, the VCCF1 technique is slightly more powerful for both positive and negative $r_1$values. Using the VCCF2 approach with multiple lags included is not recommended due to the method's high type I error rates. Furthermore, the variance correction techniques under the AR(1) assumption are highly powerful even when this is not the true memory structure of the data;
- Of the remaining methods which rely on the AR(1) assumption (PW, TFPW, MTFPW, BCPW, and VCPW), the TFPW approach tends to be the most powerful when the lag-1 serial correlation coefficient is positive due to the inflationary effect of the technique. It was shown through simulation, however, that this comes at the cost of high type I error rates. Conversely, when data are negatively correlated, TFPW is much less powerful than the other techniques. Furthermore, the propensity of the PW approach to remove (positive r_1) or add (negative r_1) a portion of the existing trend is observable through the assessment results in Tables 4-3 and 4-4. Given that the remainder of the techniques (i.e., MTFPW, BCPW, and VCPW) provide very similar results to the PW method, this may indicate that the implementation of these methods may affect the true slope of trend. Similar results are observed even when the true dependence structure of the data

is not AR(1). The results of the above noted techniques are a bit more varied when r_1 is negative; however, none appear to be the overall most appropriate in these instances;

- The detrended BBS approach provides very similar power as the standard approach. This is likely due to the method's insensitivity to small changes in block length. However, detrending may potentially remove any existing serial correlation, thus affecting the result of the SBS approach;
- The SBS technique is less effective on higher order AR dependence structures as well as having a marked decrease in power for negatively correlated data. Therefore, this technique should be used with caution on positively correlated data and avoided when r_1 is negative due to the technique's lack of power. Of the two bootstrap techniques analyzed, the SBS technique is much less computationally intensive; however, the technique is limited to linear weakly stationary data. Also, determining the appropriate memory structure for the SBS technique is somewhat time intensive;
- The lag-1 serial correlation coefficients of the BMS series are both negative and positive, however, the POT data are dominated by positive serial correlation. Furthermore, the serial structure of the BMS data is more likely to be AR(1) in comparison to the POT series. Additionally, in agreeance with the results of Svensson *et al*. (2005), lower slopes of trends are identified in the POT series. This is likely a results of the incorporation of dry years in the BMS series (which would likely be omitted from the POT data), which may cause steeper overall slope magnitudes;
- There are more instances of significant autocorrelation in the June to September data compared to those in October to December, likely due to antecedent groundwater conditions;
- Overall, for negatively serially correlated data, the VCCF1 approach under the AR(1) assumption and the BBS approach appear to be the most appropriate and powerful techniques. In the case where data have positive r_1 values, both VCCF1 techniques, BBS, and SBS techniques provide powerful results. The use of the SBS technique on higher order memory structures may result in a decreased ability to detect significant trend, however, the BBS and SBS techniques do have slightly higher type I error rates. Additionally, even when the AR(1) assumption is appropriate, the PW, TFPW, MTFPW, BCPW, and VCPW techniques fail to identify significant trends in several instances; and
- Overall, there is more significant serial correlation identified in the POT series, however, the BMS data are much more likely to have significant negative lag-1 serial correlation.

Records without Significant Serial Correlation

A comparison of the BBS(MK) and original MK techniques, as well as the BBS(SRC) and original SRC methods are carried out on data without significant lag-1 serial correlation. The results of the percentages of sites with significant trend are provided in Table 4-5. The differences between the four tested methods are small, indicating: (1) the BBS and original MK applications are very similar in their identified trend results; and (2) that the power for trend detection between the MK/SRC (original) and BBS(MK)/BBS(SRC) is similar but not identical. Therefore, in agreeance with the recommendations of Khaliq *et al*. (2009) and Sonali and Kumar (2013), more than one trend identification method should be employed. Additionally, and in accordance with the results of Svensson *et al*. (2005), more trends are found in the BMS data than in the POT data, likely due to the higher slopes of trends found in the latter. There are, in general, more trends found in the June to September period than in the October to December time frame, which is likely the result of more intense and frequent rainfall events during that time.

Table 4-5: Percentages of sites with significant trend - without significant serial structures.

Season	Series Type	Sig. Level	Percentage of Sig. Trends			
			BBS-MK	MK	BBS-SRC	SRC
Jun. - Sept.	BMS	0.05	20.0%	20.0%	16.9%	20.0%
		0.10	26.2%	26.2%	27.7%	26.2%
	POT	0.05	5.9%	3.9%	5.9%	3.9%
		0.10	17.6%	17.6%	17.6%	13.7%
Oct. - Dec.	BMS	0.05	10.3%	10.3%	8.8%	8.8%
		0.10	16.2%	14.7%	17.6%	13.2%
	POT	0.05	6.3%	7.8%	6.3%	7.8%
		0.10	10.9%	9.4%	10.9%	10.9%

4.6.4 Comparative Trend Analysis of Various Event Characteristics

The previous sections analyzed peak event data, using the full period of record. Using identical periods of record allows for a comparative analysis of trends and allows for the determination of field significance. This section additionally employs five flood variable, which include the duration, flashiness, frequency (for POT data), peak, time to peak, and volume. An overview of the global significance calculation methodology is provided in section 3.2.6. Given the results of section 4.6.3.1, the BBS(MK) will be employed for this task to account for significantly serially correlated series. Furthermore, as shown in section 4.6.3.2, the BBS approach provides similar results to the

original MK test in the event of uncorrelated data, therefore, this approach is appropriate for both types of data.

Tables 4-6 and 4-7 provide the trend analysis results for the June to September and October to December time periods, respectively. Each table displays the results from the four catchment types outlined in section 4.1, which include agricultural, natural, RHBN, and urban. For the sake of comparison, records having data for three periods of records are analyzed for trends, which consist of 1969-2016, 1979-2016, and 1989-2016. For each watershed type, the percentage of stations having significant trend at the 5% and 10% significance levels are presented, where 10% are bracketed and presented below the 5% results. Furthermore, positive (increasing) and negative (decreasing) trends are indicated. Field significant results are bolded and italicized and are presented separately for increasing and decreasing trends.

As anticipated, there are a substantial number of globally significant trends in the agricultural and urban catchments, which are likely due to the associated land-use changes. For the agricultural watersheds in the J-S time period (Table 4-6), there is a notable decrease in peak event duration through all the time periods. This may be the result of less baseflow contribution to stream flows due to the installation of tile drainage in the agricultural catchments. The tile drainage coverage of southern Ontario was retrieved from Land Information Ontario (LIO), which indicates that all of the agricultural watersheds in this analysis indeed have dense artificial drainage networks. This result could be due to extensive surface drainage improvement or the efficacy of the subsurface drainage. There is a general increasing tendency for peak event magnitudes over all the time periods and data types. Therefore, it would appear that the connection of previously unconnected depression storage and higher velocity subsurface flows have potentially caused increased peak flows. In the most recent time period, however, there appears to be both increasing and decreasing globally significant results, possibly indicating that the additional water storage capacity of the soil may play a role in decreasing peak flows in certain areas. The results of the time to peak are somewhat mixed and appear to be highly dependent on the data type and time periods. Therefore the field significant increasing flashiness trends are likely the results of increasing peak event magnitudes rather than decreasing times to peak. Event volume trends are highly dependent on the type of data used, for example, the BMS data indicates increasing globally significant trends, whereas the POT data show a combination of increasing and decreasing trends, with more of the former. Finally, there are increasing trends in the frequency of peak event magnitudes, however, the percentage of such decreases in the 1979/1989-2015 time periods.

There are similarly a large proportion of globally significant trends in urban station data, the results of which closely follow those noted in section 4.1. The addition of impervious surfaces over time has likely led to decreased baseflow and subsequent significant decreasing trends in duration. Trends in peak flow are greatly dependent on the type of data employed for analysis. Based on the BMS data, there are certainly increasing trends in peak flow over all the time periods, however, an analysis of the POT data indicates no trends in peak flows. The volume of events appears to be increasing, independent of the time period or type of data employed, although the POT trends are not globally significant in the 1979/1989-2014 time periods. Few field significant results are found in the time to peak data and those results are not consistent in the different times periods or the type of data. Regardless of the type of data analyzed or the time period, there is a consensus on a statistically significant increase in the flashiness of numerous catchments. Finally, it is unsurprising that field significant trends in the frequency of peak event flows were noted throughout all the examined time periods based on the nature of urban growth.

Given the extent of land-use changes in the agricultural and urban watersheds, no conclusions regarding climate variability could be garnered. To this end, an analysis of the RHBN and natural watershed is carried out. There are several field significant trends found in the RHBN data, of particular interest is that 50% of the sites for the peak streamflow data, in the 1969-2016 time period exhibit increasing trend at the 5% significance level. This could be an indicator of climate variability, however, no trends are found in the POT data for the same time period. Therefore, these trends in the BMS data could be due to the inclusion of dry years in the BMS series (Svensson *et al*., 2005). Moreover, both the natural and RHBN sites show globally significant increasing trends in frequency for the 1979-2016 time period. Finally, statistically significant increasing trends in flashiness were indicated for the POT data during the 1989-2016 time period, however, they are not accompanied by globally significant trend in the peak or time to peak event characteristics. Therefore, climate change signals may be confounded with land-use changes at the urban and agricultural watershed and the trends may not wholly be related to hydromodifications; however, this cannot be confirmed without further analysis of precipitation data over the same time periods.

Table 4-7 provides a similar analysis but presents the data from the O-D time period. In the land-use dominated catchments, there are noticeably less globally significant trends (particularly in shorter time periods) but this is likely a result of less rainfall at this time of year. The results in the agricultural catchments are somewhat similar to those found for the J-S time period for the duration, peak, and time to peak. The trend results for the watershed flashiness, volume, and frequency are

mixed and appear to be highly dependent on the type of data and time period. The urban catchments also follow a similar pattern when compared to the J-S results. The duration of events is similarly decreasing over all the time periods. There is, however, little evidence of increasing trends in peak event magnitudes aside from those found in the 1969-2016 time period, which could be a direct result of urbanization (Thompson, 2013). The volume trend results again appear to be highly dependent on the type of data and no overall consensus on this event characteristic can be determined. The proportion of time to peak and flashiness trends are generally decreasing with decreasing time periods. Lastly, there is no evidence of trends in the frequency of peak event magnitudes. An examination of the RHBN and natural sites also indicates that for the POT data during the 1969-2016 time period, 100% of the natural sites displayed globally significant increasing trends in flashiness and 100% of the RHBN sites showed increasing trends in frequency. Moreover, during the 1979-2015 time period, the POT data indicate that 33% of the RHBN sites exhibit increasing time to peak. The RHBN sites also display decreasing trends in duration during the 1989-2016 time period, however, this is not consistent with the POT data for the same time frame.

In regards to the natural and RHBN catchments, there is some evidence to suggest that there may be increasing peak events and frequency of those events. There is certainly no evidence of a decreasing tendency in these events, a result that has been found in numerous other studies (Zhang *et al*., 2001; Burn and Hag Elnur, 2002; Cunderlik and Ouarda, 2009; Burn *et al*., 2010; Burn *et al*., 2012). It should be noted however, that most peak analyses use published AMS data, the peaks of which are generally caused by freshet events in snowmelt dominated watersheds. Therefore, a direct comparison to these studies cannot be carried out as this analysis looks at secondary events which occur in the summer or fall seasons having dissimilar flood generating mechanisms.

It is apparent that the addition of impervious surfaces and dense drainage networks have resulted in various significant changes to the hydrologic regimes of the urbanized catchments. Results of the BMS trend analysis for peak magnitude for the J-S and O-D time periods are generally increasing, a result that is commonly found in the literature (Leopold, 1968; Rose and Peters, 2001; White and Greer, 2006; Hajazi and Markus, 2009; Rosburg *et al*., 2017). Similarly, the BMS and POT data show globally significant increasing trends in the volume of peak event and decreasing trends in duration, which again has been documented in several previous studies (Leopold, 1968; Boyd *et al*., 1993; Konrad, 2003; Roseburg *et al*. 2017). It is generally believed the urban growth decreases the time to peak (Leopold, 1968; Konrad, 2003); a result that is not consistently observed in the J-S time period but is, however, in the O-D data. Similar to the results of Roseburg *et al*. (2017), the

flashiness of the urban catchments during the J-S time period is increasing. During the O-D time period, a mix of increasing and decreasing significant trends are found. Finally, urbanization is commonly associated with an increase in the frequency of peak events (Rose and Peters, 2001; Konrad, 2003), a result which is also established in the summer (J-S) data. Conversely, the O-D data exhibit little evidence of trends in this characteristic.

The agricultural catchments have seen a notable increase in peak discharge over the three time periods and data types. This result is consistent with the findings of Turtola and Paajanen (1995) and Rozemeijer and Broers (2007). The reason for the increase in peak flows may be attributed to a number of basin characteristics, such as antecedent soil moisture conditions, soil type, crop type, climatic conditions, catchment slope, or density and type of subsurface drainage (Madramootoo, 1988; Fraser and Flemming, 2001). Serrano (1985) and Spaling (1995) found decreases in time to peak in agricultural catchments with artificial drainage. This is somewhat consistent with the results from this analysis, given that globally significant trends were found for this event characteristic, however, increasing trends were identified as well. Finally, the trend results for volume were somewhat mixed and were dependent on the time period and type of data (POT or BMS). Bengtson *et al*. (1988) found an increase in volume between drained and undrained plots in Louisiana, whereas Rahman *et al*. (2014) found decreases in flow volume in the Red River agricultural basin, located in the United States and southern Canada.

Table 4-6: Percentages of trends observed in the June – September period.

	June - September			
	1969-2015			
	Agricultural	**Natural**	**RHBN**	**Urban**
Number of Sites Included:	**29**	**1**	**2**	**5**
		BMS		
Duration	+0%/***-27.6%*** (+0%/***-34.5%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/***-60.0%*** (+0%/***-60.0%***)
Flashiness	***+27.6%***/-0% (***+37.9%***/-0%)	+0%/-0% (+100%/-0%)*	+50.0%/-0% (+50.0%/-0%)	***+80.0%***/-0% (***+100%***/-0%)
Peak	***+20.7%***/-0% (***+27.6%***/-0%)	+0%/-0% (+0%/-0%)	***+50.0%***/-0% (+50.0%/-0%)	***+80%***/-0% (***+100%***/-0%)
Time to Peak	+0%/-0% (***+6.9%***/-3.5%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+40.0%/-0% (***+40.0%***/-20.0%)
Volume	***+13.8%***/-0% (***+20.7%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+50.0%/-0%)	***+100.0%***/-0% (***+100.0%***/-0%)
		POT		
Duration	+3.5%/***-27.6%*** (+3.5%/***-31.0%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/***-60.0%*** (+0.0%/***-80.0%***)
Flashiness	***+6.9%***/-0% (***+6.9%***/***-3.4%***)	+0%/-0% (+0%/-0%)	+50.0%/-0% (+50.0%/-0%)	+20.0%/-0% (***+60.0%***/-0%)
Frequency	***+20.7%***/-0% (***+24.1%***/-3.4%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	***+100.0%***/-0% (***+100.0%***/-0.0%)
Peak	***+6.9%***/***-3.4%*** (***+6.9%***/***-10.3%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0.0% (+0.0%/-0.0%)
Time to Peak	+0%/-6.9% (+0%/***-13.8%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+20.0%/-20.0% (+20.0%/-20.0%)
Volume	+0%/***-6.9%*** (***+3.4%***/***-6.9%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	***+40.0%***/-0.0% (***+40.0%***/-0.0%)
	1979-2015			
	Agricultural	**Natural**	**RHBN**	**Urban**
Number of Sites Included:	**44**	**4**	**6**	**5**
		BMS		
Duration	+2.0%/***-34.1%*** (***+2.0%***/***-36.4%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-16.7%)	+0%/***-80.0%*** (+0%/***-100.0%***)
Flashiness	+11.4%/-0% (***+25.0%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+16.7***%***/-0%)	***+20.0%***/-0% (***+40.0%***/-0%)
Peak	+4.5%/-0% (***+18.2%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	***+20.0%***/-0% (***+40.0%***/-0%)
Time to Peak	+0%/-0% (+0%/***-5.0%***)	+0%/-0% (+0%/-0%)	+16.7%/-0% (+16.7***%***/-0%)	+0%/-20.0% (20.0%/-20.0%)
Volume	+5.0%/-0% (***+13.6%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+40.0%/-0% (***+60.0%***/-0%)

* Positive values indicate increasing trends and negative values indicate decreasing trends; Entries in bold/italics indicate field significant results; the first line of results are trend results at the 5% significance level, the second line in parentheses are for the 10% significance level, which applies to the field significant results as well. (*) identifies sites land-use classifications in which field significance could not be determined due to an inadequate number of stations.

Table 4-6: Percentages of trends observed in the June – September period (continued).

	June - September			
	1979-2015			
	POT			
Duration	+0%/***-18.2%*** (+0%/***-25.0%***)	+0%/-0% (+0%/-25.0%)	+0%/-16.7% (+0%/-16.7%)	+0%/***-60.0%*** (+0%/***-80.0%***)
Flashiness	***+5.0%***/-2.0% (***+5.0%***/***-2.0%***)	+25.0%/-0% (+25.0%/-0%)	+0%/-0% (+16.7%/-0%)	+20.0%/-0% (+40.0%/-0%)
Frequency	***+4.5%***/-0% (***+9.0%***/-0%)	+0%/-0% (***+25.0%***/-0%)	+0%/-0% (***+16.7%***/-0%)	+40.0%/-0% (***+40.0%***/-0%)
Peak	+0%/***-5.0%*** (***+6.8%***/***-5.0%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Time to Peak	+2.0%/-4.5% (+2.0%/***-6.8%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-20.0% (+20.0%/***-40.0%***)
Volume	+2.0%/***-6.8%*** (***+4.5%***/***-11.4%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+20.0%/-0%)
	1989-2015			
	Agricultural	**Natural**	**RHBN**	**Urban**
Number of Sites Included:	**52**	**4**	**6**	**8**
	BMS			
Duration	+0%/***-13.5%*** ***(+20.0%/-21.2%)***	+25.0%/-0% (+25.0%/-25.0%)	+0%/-0% (+0%/-0%)	+0%/***-37.5%*** (+0%/***-37.5%***)
Flashiness	***+11.5%***/-0% (***+19.2%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+25.0%/-0% (***+37.5%***/-0%)
Peak	+3.8%/-0% (***+15.4%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+25.0%/-0% (***+37.5%***/-0%)
Time to Peak	+2.0%/-0% (***+5.8%***/-2.0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-12.5% (+0%/-12.5%)
Volume	***+6.0%***/-0% (***+13.5%***/-0%)	+25.0%/-0% (+25.0%/-0%)	+0%/-0% (+0%/-0%)	***+25.0%***/-0% (***+37.5%***/-0%)
	POT			
Duration	+0%/***-11.5%*** (+0%/***-19.2%***)	+0%/-25.0% (+0%/-25.0%)	+0%/-0% (+0%/-0%)	***+12.5%***/***-37.5%*** (***+25.0%***/***-37.5%***)
Flashiness	***+7.7%***/-0% (***+9.6%***/-0%)	+25.0%/-0% (+25.0%/-0%)	+0%/-0% (***+33.3%***/-0%)	+25.0%/-0% (***+37.5%***/-0%)
Frequency	***+6.0%***/-0% (***+11.5%***/-0%)	+0%/-0% (+25.0%/-0%)	+0%/-0% (+16.7%/-0%)	***+25.0%***/-0% (***+50.0%***/-0%)
Peak	+0%/***-4.0%*** (***+5.8%***/***-5.8%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Time to Peak	+4.0%/***-8.0%*** (***+4.0%***/***-8.0%***)	+0%/-0% (+0%/-25.0%)	+0%/-16.7% (+0%/-16.7%)	+0%/-0% (+0%/-12.5%)
Volume	+0%/-2.0% (***+2.0%***/***-3.8%***)	+25.0%/-0% (+25.0%/-0%)	+0%/-0% (+0%/-0%)	+12.5%/-0% (+13.0%/-0%)

* Positive values indicate increasing trends and negative values indicate decreasing trends; Entries in bold/italics indicate field significant results; the first line of results are trend results at the 5% significance level, the second line in parentheses are for the 10% significance level, which applies to the field significant results as well.

Table 4-7: Percentages of trends observed in the October - December period.

	October - December			
	1969-2015			
	Agricultural	**Natural**	**RHBN**	**Urban**
Number of Sites Included:	**29**	**1**	**2**	**5**
		BMS		
Duration	***+10.3%*/*-24.2%*** (***+10.3%*/*-24.1%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-40.0% (+0%/***-60.0%***)
Flashiness	***+20.7%***/-3.4% (***+24.1%***/-3.4%)	+100%/-0%* (+100%/-0%)*	+0%/-0% (+0%/-0%)	***+60.0%***/-0% (***+60.0%***/-0%)
Peak	***+10.3%***/-0% (***+24.1%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	***+100.0%***/-0% (***+100.0%***/-0%)
Time to Peak	+0%/-0% (+0%/***-6.9%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-20.0% (+0%/***-40.0%***)
Volume	***+6.9%***/-0% (***+17.2%*/*-3.4%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+20.0%/-0% (+20.0%/-0%)
		POT		
Duration	+0%/-3.4% (+0%/***-17.2%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+20.0%/-0% (+0%/-20.0%)
Flashiness	+3.4%/-3.4% (***6.9%***/-3.4%)	+0%/-0% (+100%/-0%)*	+0%/-0% (+0%/-0%)	***+40.0%***/-0% (***+40.0%***/-0%)
Frequency	***+6.9%***/-0% (***+6.9%*/*-3.4%***)	+0%/-0% (+0%/-0%)	+50.0%/-0% (***+100%***/-0%)	+0%/-0% (0%/-0%)
Peak	+6.9%/-0% (***+10.3%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+20.0%/-0% (***+40.0%***/-0%)
Time to Peak	+0%/-6.9% (+0%/***-6.9%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-20.0% (0%/-20.0%)
Volume	+3.4%/-0% (***+13.8%*/*-3.4%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
	1979-2015			
	Agricultural	**Natural**	**RHBN**	**Urban**
Number of Sites Included:	**44**	**4**	**6**	**5**
		BMS		
Duration	+0%/***-29.5%*** (***+2.3%*/*-29.5%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/***-100.0%*** (+0%/***-100.0%***)
Flashiness	+4.5%/-2.3% ***(+6.8%/-2.3%)***	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Peak	+0%/-0% (***+2.3%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Time to Peak	+0%/-2.3% (+0%/***-11.4%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/***-80.0%*** (+0%/***-80.0%***)
Volume	+0%/***-6.8%*** (+0%/***-9.1%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)

* Positive values indicate increasing trends and negative values indicate decreasing trends; Entries in bold/italics indicate field significant results; the first line of results are trend results at the 5% significance level, the second line in parentheses are for the 10% significance level, which applies to the field significant results as well. (*) identifies sites land-use classifications in which field significance could not be determined due to an inadequate number of stations.

Table 4.7: Percentages of trends observed in the October - December period (continued).

	October - December			
	1979-2015			
	POT			
Duration	+0%/***-7.0%*** (***+2.3%***/***-18.2%***)	+0%/-0% (+0%/-0%)	+16.7%/-0% (+16.7%/-0%)	+0%/***-40.0%*** (+0%/***-60.0%***)
Flashiness	+2.3%/-2.3% (***+6.8%***/-2.3%)	+0%/-0% (25.0%/-0%)	+16.7%/-0% (+16.7%/-0%)	+20.0%/-0% (+20.0%/-0%)
Frequency	***+2.0%***/-0% (***+2.3%***/***-2.3%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Peak	+0%/-0% (***+4.5%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-16.7% (+0%/-16.7%)	+0%/-0% (+20.0%/-0%)
Time to Peak	+4.5%/-4.5% (***+6.8%***/***-4.5%***)	+0%/-0% (+0%/-0%)	+0%/-0% ***33.3%***/-0%	+0%/***-20.0%*** (+0%/***-20.0%***)
Volume	+0%/***-5.0%*** (***+2.3%***/***-9.1%***)	+0%/-0% (+0%/-0%)	+0%/-16.7% (+0%/-16.7%)	+0%/-0% (+0%/-0%)
	1989-2015			
	Agricultural	**Natural**	**RHBN**	**Urban**
Number of Sites Included:	**52**	**4**	**6**	**8**
	BMS			
Duration	+2.0%/-4.0% (+1.9%/***-7.7%***)	+0%/-0% (+0%/-25.0%)	+0%/-0% (+0%/***-33.3%***)	+0%/-25.0% (+0%/***-37.5%***)
Flashiness	+2.0%/-0% (***+9.6%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-12.5% (+0%/-12.5%)
Peak	+0%/-0% (***+1.9%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Time to Peak	+2.0%/-0% (***+3.8%***/-0%)	+0%/-0% (+0%/***-9.1%***)	+0%/-0% (+0%/-0%)	+0%/-12.5% (+0%/-25.0%)
Volume	+2.0%/-0% (***+3.8%***/-0%)	+0%/-0% (+4.5%/-0%)	+0%/-0% (+0%/-16.7%)	+0%/-0% (+13.0%/-0%)
	POT			
Duration	+0%/-1.9% (+0%/-6.0%)	+0%/-0% (+0%/-0%)	+0%/-16.7% (+0%/-16.7%)	+0%/-12.5% (+12.5%/-12.5%)
Flashiness	***+5.8%***/***-6.0%*** (***+7.7%***/***-10.0%***)	+0%/-0% (+0%/-0%)	+16.7%/-0% (+16.7%/-0%)	+0%/-12.5% (+0%/-12.5%)
Frequency	***+2.0%***/***-2.0%*** (***+1.9%***/***-3.8%***)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Peak	+0%/-0% (***+3.8%***/-0%)	+0%/-0% (+0%/-0%)	+0%/-16.7% (+0%/-16.7%)	+0%/-12.5% (+12.5%/-12.5%)
Time to Peak	***+3.8%***/-2.0% (***+9.6%***/***-2.0%***)	+0%/-25.0% (+0%/-25.0%)	+0%/-0% (+0%/-0%)	+0%/-0% (+0%/-0%)
Volume	***+3.8%***/-2.0% (***+5.8%***/***-4.0%***)	+0%/-0% (+0%/-0%)	+0%/-16.7% (+0%/-16.7%)	+0%/-0% (+0%/-0%)

* Positive values indicate increasing trends and negative values indicate decreasing trends; Entries in bold/italics indicate field significant results; the first line of results are trend results at the 5% significance level, the second line in parentheses are for the 10% significance level, which applies to the field significant results as well.

4.7 Conclusions

This research examines the differences in serial correlation structure and trend prevalence in BMS and POT data. It was determined that BMS data are more likely to be negatively serially correlated and have an AR(1) memory structure, whereas the POT data were dominated by positive autocorrelation. It was established that the BBS technique and VCCF1 approach (under the AR(1) assumption) are most appropriate for negatively correlated data, whereas the BBS, VCCF1 techniques, and SBS approaches were found to be the most powerful for positively correlated data. Selecting the most appropriate ARMA(p,q) model may, however, be time consuming which makes the BBS approach attractive if the computational expense is not a concern.

According to Leopold (1968), urbanization is the most forceful type of land-use change in regards to altering the hydrologic regime. This statement is clearly exemplified through this analysis; however, it is also apparent that agricultural land-use changes have a marked effect on streamflow characteristics. Given that there have been increasing trends in air temperatures in southern Ontario (Nalley *et al.*, 2013) and more frequent precipitation events are expected in a warmer climate, trends in the natural and RHBN gauging data may be present due to anthropogenic climate change. Furthermore, trends in average annual flow have been detected in southern Ontario (Nalley et al. 2012) There are indications of significant changes in several event characteristics at the natural and RHBN sites, however, these changes are not consistent in the POT and BMS data or in the different time periods. Additionally, the trend results at the urban stations are quite representative of the effects of urbanization. The trend results at the agricultural stations, however, are strongly dependent on type of data and the time period used for the analysis. There is, however, a consistent pattern of increasing peak event flows, a result that is found less commonly in the literature. Therefore, future work should focus on quantifying the degree and type of artificial drainage within the agricultural catchments and determining if the increase in peak flows is attributable to land-use changes or climate change.

This research has highlighted the significant differences in identified trends when using BMS and POT data. Significant trends are more often detected in the BMS dataset likely due to the inclusion of dry years in these records, which causes increased slopes of trends. Therefore, the results of the POT analyses may be more valuable when addressing the significance of hydrometeorological trends in a given watershed. In practice, if serial correlation may be an issue, using the BBS approach is recommended due to the technique's robustness in the presence of both positive and negative serial correlation.

Chapter 5
General Conclusions

5.1 Introduction

In recent years, the seemingly larger than average number of peak hydrometeorological events has incited concern among the general public and scientific communities alike. There is evidence that suggests that these events are indeed becoming more frequent and of greater magnitude (Coumou and Rahmstorf, 2012). Therefore, the advancement of statistical techniques that incorporate nonstationarity are of great value in a changing climate. This thesis examines methodologies for the incorporation of nonstationarity in pooled and at-site FFA in addition to techniques for nonparametric trend identification in the presence of significant serial correlation. Thus, the overall contribution of this research aids in the development of nonstationary techniques in the face of temporal dependence, the key findings of which are outlined below.

5.2 Summary of Results and Conclusions

A trend centered pooling approach is developed in Chapter 2, which accounts for statistically significant trend found in peak annual streamflow records. Four homogeneous pooling groups were created based on the form of detected trend in the at-site data. The developed index-flood approach utilizes a time-dependent location parameter that is then incorporated into regional parameters. The following results and conclusions are drawn from Chapter 2:

- The potential for underestimation/overestimation of increasing/decreasing time-dependent quantiles was demonstrated, thus elucidating the need for nonstationary pooled FFA techniques;
- Through the use of the standard Hosking and Wallis (1997) goodness-of-fit technique, it was demonstrated that the best fitting distribution changes when the trend centered pooling approach is employed, compared to the standard index-flood methodology; and
- Less uncertainty was found in the nonstationary quantile estimates compared to their stationary counterparts, thus confirming that the methodology provides more accurate results than the standard index-flood approach.

The purpose of Chapter 3 is the development of a peaks-over-threshold approach that allows for the incorporation and selection of covariate-dependent thresholds and GP distribution parameters. This methodology uses both bivariate and multivariate models that account for temporal dependence in addition to large-scale ocean-atmosphere phenomena. Given the propensity for climate change effects in coastal BC, this area was used for the analysis. The area is divided regionally for comparative results, which include the North Coast, South Coast, and Vancouver Island. The overall results and conclusions of this chapter are as follows:

- The threshold and model selection methodology is robust. This was tested through the removal of the higher-order (quadratic) model, in which case, there was very little change in the selected models;
- There is limited overlap in trends detected through nonparametric means and those determined through the quantile modelling approach. This highlights (1) the need for alternative trend assessment approaches, and (2) when teleconnections have a strong influence on a region's climate, they should be accounted for to effectively address increasing or decreasing tendencies in peak hydroclimatological extremes; and
- The results of the analysis of the mainland areas (North Coast and South Coast) indicated that peak precipitation amounts are more likely stationary. The Vancouver Island analyses do, however, show a greater propensity for nonstationarity.

Chapter 4 provides a comparison of several techniques that account for serial correlation when used with nonparametric trend testing. The power and type I error rates of these methods were assessed for both positive and negative lag-1 autocorrelation coefficients. This research additionally addressed the differences in the serial structure of BMS and POT data and highlights which methods are most appropriate for use with each of these data types. Finally, this chapter provides a trend analysis of a subset of watersheds in southern Ontario based on land-use. The results and conclusions from Chapter 4 are presented below:

- In agreeance with the literature, the MK test alone is greatly affected by significant lag-1 serial correlation and should not be used in this instance;

- The memory structure of the BMS and POT data are quite different, whereby a great deal more significant serial correlation is identified in the POT data and the BMS data is more likely to have significant negative lag-1 serial correlation;
- When several lags of significant serial correlation are present, the VCCF1 approach does not perform as expected, and in this instance, it is recommended to use this approach under the AR(1) assumption regardless of the serial structure of the data. The VCCF2 approach has high type I error rates when larger lags are considered, therefore, using this technique, again under the AR(1) assumption, regardless of the memory structure of the data, is recommended. It was found, however, that the variance correction techniques were highly powerful when used under the AR(1) assumption in all instances;
- In agreeance with other research, the TFPW approach should be avoided due to its high type I error rates;
- The PW approach alone should be avoided due to its inflationary effect in the presence of positive lag-1 serial correlation;
- The BBS approach is similarly powerful to the original MK approach in the absence of serial correlation;
- Overall, the BBS, VCCF1 techniques, and SBS techniques are recommended for positively autocorrelated data, however, the SBS technique is less effective on higher order AR dependence structures;
- When data are negatively correlated, the VCCF1 approach under the AR(1) assumption and the BBS approach appear to be the most appropriate techniques;
- Significant trends and higher slopes of trends are more likely to be detected in BMS data, likely due to the incorporation of data in dry years;
- Numerous statistically and globally significant trends were found in the agricultural and urban land-use catchments, whereas, there is much less indication of trend significance in the RHBN and natural sites.

5.3 Future Research

In regards to Chapter 2, future work may focus on the creation of a regionalization technique based on the sign and magnitude of trends found in peak flow data. This may be carried out using a Monte Carlo simulation or through the use of a larger data set, with more pronounced time-dependence. Testing at-site data for trends in both the mean and variance before regionalization could also be a focus of future study. Given that the standard regional heterogeneity tests have been formulated for IID data, more research is also needed in this area to ascertain the effects of nonstationarity on this measure.

The application of the threshold and GP model selection approach presented in Chapter 3 could be applied to other types of hydrometeorological data, which would elucidate the flexibility of the methodology. Furthermore, the approach allows for the addition or removal of the complexity of the models included, so it could easily be used in different geographic areas.

The land-use trend analysis presented in Chapter 4 indicates a consistent pattern of increasing peak flow events in the agricultural watersheds. This result is less common in the literature and warrants further analysis. Climate variability should be ruled out using precipitation data for these catchments, after which, a soil analysis, and the extent of artificial drainage should be examined. Any trend patterns consistent with the installation of tile drainage could then be linked to those found in the watersheds or potentially to climate change effects.

References

Akaike, H. 1974. A new look at statistical-model identification. *IEEE Transaction on Automatic Control, AC* 19(6): 716-723.

Andreas, F. P., R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland. 2017. The future intensification of hourly precipitation extremes. *Nature Climate Change*, 7: 48-52, doi:10.1038/nclimate3168.

Apsīte, E., O. Nikodemus, G. Brūmelis, A. Lagzdiņš, D. Elferts, Z. Rendenieks, and L. Klints. 2017. Impact of climate variability, drainage, and land-cover changes on hemiboreal streamflow. *Hydrological Science Journal*, 62(15): 2558-2570.

Balkema, A.A. and L. de Hann. 1974. Residual Life Time at Great Age. *Annals of Probability*, 2(5): 792-804.

Baron, S., G. Canete, J. Carmichael, D. Flanders, E. Pond, S. Sheppard, and K. Tatebe. 2012. A climate change adaptation planning process for low-lying, communities vulnerable to sea level rise. *Sustainability*, 4: 2167-2208.

Barrow, E., B. Maxwell, and P. Gachon. 2004. Climate variability and change in Canada: past, present and future. *ACSD Science Assessment Series No. 2*, Meteorological Service of Canada, Environment Canada, Toronto, Ontario, 114 pp.

Barry, D. and J.A. Hartigan. 1992. Product Partition Models for Change Point Problems. *Annals of Statistics*, 20(1): 260-279.

Barry, D. and J.A. Hartigan. 1993. A Bayesian Analysis for Change Point Problems. *Journal of the American Statistical Association*, 88(421): 309-319.

Bayazit, M. and B. Önöz. 2004. Comment on "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test" by Shen Yue and Chun Yuan Wang. *Water Resources Research*, 40, W08801, doi: 10.1029/2002WR001925.

Bayazit, M. and B. Önöz. 2007. To prewhiten or not to prewhiten in trend analysis? *Hydrological Sciences Journal*, 52(4): 611-624.

Bayley, G. V. and J. M. Hammersley. 1946. The "effective" number of independent observations in an autocorrelated time series. *Supplement to the Journal of the Royal Statistical Society*, 8(2): 184-197.

Bayliss, A. C. and R.C. Jones. 1993. *Peaks-Over-Threshold Database: Summary Statistics and Seasonality. IH Report no. 121*, Institute of Hydrology, Wallingford, UK, 61 pp.

Beguería, S., M. Angula-Martínez, S. Vincente-Serrano, J.I. López-Moreno, and A. El-Kenawy. 2011. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006. *International Journal of Climatology*, 31: 2102-2114.

Bengtson, R. L, C. E. Carter, H. F. Morris, S. Bartliewicz. 1988. The influence of subsurface drainage practices on nitrogen and phosphorus losses in a warm, humid climate. *Transactions of the ASAE*, 31(3): 729-733.

Blain, G. C. 2013. The modified Mann-Kendall test: on the performance of three variance correction approaches. *Agrometeorology*, 72(4): 416-425.

Blain, G. C. 2014. Removing the influence of serial correlation on the Mann-Kendall test. *Revista Brasiliera de Meteorologia*, 29(2): 161-170.

Bormann, H., N. Pinter, and S. Elfert. 2011. Hydrological signatures of flood trends on German rivers: Flood frequencies, flood heights and specific stages. *Journal of Hydrology*, 404: 50-66.

Box, G. E. P. and G. M. Jenkins. 1970. Time Series Analysis Forecasting and Control. Holden-Day, San Fransisco, California, 712 pp.

Boyd, M. J., M. C. Bufill, and R. M. Knee. 1993. Pervious and impervious runoff in urban catchments. *Hydrological Sciences Journal*, 38(6): 463-478.
British Columbia Ministry of Environment (BCME). 2007. Environmental Trends in British Columbia: 2007, 352 pp.
Bormann, H., N. Pinter, and S. Elfert. 2011. Hydrological signatures of flood trends on German rivers: Flood frequencies, flood heights and specific stages. *Journal of Hydrology*, 404: 50-66.
Bühlmann, P. 2002. Boostraps for time series. *Statistical Sciences*, 17(1): 52-72.
Burn, D. H. 1989. Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management*, 115(5): 567-581.
Burn, D. H. 1994. Hydrologic effects of climate change in west-central Canada. *Journal of Hydrology*, 160: 53-70.
Burn, D. H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, 202: 212-230.
Burn, D. H. 2003. The use of resampling for estimating confidence intervals for single site and pooled frequency analysis. *Hydrological Sciences Journal*, 48(1): 25-38.
Burn, D. H. 2008. Climatic influences on streamflow timing in the headwaters of the Mackenzie River Basin. *Journal of Hydrology*, 352: 225-238.
Burn, D. H., Z. Zrinji, and M. Kowalchuk. 1997. Regionalization of catchments for regional flood frequency analysis. *Journal of Hydrologic Engineering*, 2(2): 76-82.
Burn, D. H., J. M. Cunderlik, and A. Pietroniro. 2003. Hydrological trends and variability in the Liard River basin. *Hydrological Sciences Journal*, 49(1): 53-67.
Burn, D.H. and M.A. Hag Elnur. 2002. Detection of hydrological trends and variability. *Journal of Hydrology*, 255: 107-122.
Burn, D. H., J. Hannaford, G. A. Hodgkins, P. H. Whitfield, R. Thorne, and T. Marsh. 2012. Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow. *Hydrological Sciences Journal*, 57(8): 1580-1593.
Burn, D.H., R. Mansour, K. Zhang, and P.H. Whitfield. 2011. Trends in variability in extreme rainfall event in British Columbia. *Canadian Water Resources Journal*, 36(1): 67-82.
Burn, D. H., M. Sharif, and K. Zhang. 2010. Detection of trends in hydrological extremes for Canadian watersheds. *Hydrological Processes*, 24: 1781-1791.
Burns, D., T. Vitvar, J. McDonnell, J. Hassett, J. Duncan, and C. Kendall. 2005. Effects on suburban development on runoff generation in the Croton River basin, New York, USA. *Journal of Hydrology*, 311: 266-281.
Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference. New York: Springer, 488 pp.
Cabilio, P., Y. Zhang, and X. Chen. 2013. Bootstrap rank tests for trend in time series. *Environmetrics*, 24: 537-549.
Caires, S., V.R. Swail, and X.L. Wang. 2006. Projection and analysis of extreme wave climate. *Journal of Climate*, 19: 5581-5605.
Castellarin, A., D. H. Burn, and A. Brath. 2008. Homogeneity testing: how homogeneous do heterogeneous cross correlated regions seem? *Journal of Hydrology*, 360: 67-76.
Chan, S. C., E. J. Kendon, N. M. Roberts, H. J. Fowler, and S. Blenkinsop. 2016. Downturn in scaling of UK extreme rainfall with temperature for future hottest days. *Nature Goescience*, 9: 24-28, doi:10.1038/ngeo2596.
Chang, H. 2007. Comparative streamflow characteristics in urbanizing basins in the Portland Metropolitan Area, Oregon, USA. *Hydrological Processes*, 21: 211-222.
Chang, S. E., M. Gregorian, K. Pathman, L. Yumagulova, and W. Tse. 2012. Urban growth and long-term changes in natural hazard risk. *Environment and Planning A*, 44: 989-1008.

Chernick, M. R. 2008. *Bootstrap methods: A Guide for Practitioners and Researchers*. Wiley, New Jersery, 369 pp.
Coles, S. 1993. Regional modelling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society: Series B*, 55(4); 797-816.
Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, United Kingdom, 208 pp.
Coumou, D. and S. Rahmstorf. 2012. A decade of weather extremes. *Nature Climate Change*, 2(7): 491-496. doi: 10.1038/NCLIMATE1452.
Cunderlik, J. M. and D. H. Burn. 2003. Non-stationary pooled flood frequency analysis. *Journal of Hydrology*, 276: 210-223.
Cunderlik, J. M. and T. B. M. J. Ouarda. 2006. Regional flood-duration-frequency modeling in the changing environment. *Journal of Hydrology*, 318: 276-291.
Cunderlik, J. M. and T. B. M. J. Ouarda. 2009. Trends in the timing and magnitude of floods in Canada. *Journal of Hydrology*, 375: 471-480.
Cunnane, C. 1979. A note on the Poisson assumption on partial duration series models. *Water Resources Research*, 15(2): 489-494.
Dalrymple, T. 1960. *Flood frequency analysis*. US Geological Survey Water-Supply Paper no. 1543-A, 80 pp.
Douglas, E. M., R. M. Vogel, and C. N. Kroll. 2000. Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology*, 240: 90-105.
El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée. 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43, W03410, doi:10.1029/2005WR004545.
Environment Canada (EC), 2011. Canada's Top Ten Weather Stories for 2011, Historic Flood Fights in the West. http://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=0397DE72-1.
The International Disaster Database–Center for Research on the Epidemiology of Disasters–CRED (EM-DAT). 2016. http://www.emdat.be/.
Everard, M. and H. L. Moggridge. 2012. Rediscovering the value of urban rivers. *Urban Ecosystems*, 15: 293-314.
Fraser, H. and R. Fleming. 2001. *Environmental benefits of tile drainage*. University of Guelph. https://www.ridgetownc.com/research/documents/fleming_drainage.pdf.
Gershunov, A. and T. P. Barnett. 1998. ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *Journal of Climate*, 11: 1575-1586.
Gilroy, K. L. and R. H. McCuen. 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *Journal of Hydrology*, 414-415: 40-48.
Hamdi, R., P. Termonia, and P. Baguis. 2011. Effects of urbanization and climate change on surface runoff of the Brussels Capitol Region: a case study using and urban soil–vegetation-atmosphere transfer model. *International Journal of Climatology*, 31: 1959-1974.
Hamed, K. H., 2009. Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of Hydrology*, 368: 143-155.
Hamed, K. H. and A. R. Rao. 1998. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204: 182-196.
Hanel, M., T. A. Buishand, and C. A. T. Ferro. 2009. A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *Journal of Geophysical Research*, 114, D15107, doi: 10.1029/2009JD011712.
Hare, F. K. and M. K. Thomas. 1979. *Climate Canada, 2nd Edition*, Wiley, New York, 230 pp.

Hartmann, D. L., A. M. G. Klein Tank, M. Rusticucci, L. V. Alexander, S. Brönnimann, Y. Charabi, F. J. Dentener, E. J. Dlugokencky, D. R. Easterling, A. Kaplan, B. J. Soden, P. W. Thorne, M. Wild, and P. M. Zhai. 2013. Observations: Atmosphere and Surface. In: Stocker, T. F., Qin, D., Platter, G. –K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and P. M. Midgley [edt] *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 30 pp.

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

Heffernan, J. E. and A. G. Stephenson. 2009. *An Introduction to Statistical Modeling of Extreme Values*, R package 'ismev'. http://cran.r-project.org/web/packages/ismev/ismev.pdf.

Hipel, K. W. and A. I. McLeod. 1994. *Time series modelling of water resources and environmental systems*. Elsevier, New York, 1013 pp.

Hipel, K. W., A. I. McLeod, and R. R. Weiler. 1988. Data analysis of water quality time series in Lake Erie. *Water Resources Bulletin*, 24(3): 533-544.

Hirsch, R. M., 2011. A perspective on nonstationarity and water management. *Journal of the American Water Resources Association*, 47(3): 436-446.

Hirsch, R. M., J. R. Slack, and R. A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18(1): 107-121.

Hodgkins, G. A., R. W. Dudley, and T. G. Huntington. 2003. Changes in the timing and high river flows in New England over the 20th Century. *Journal of Hydrology*, 278: 244-252.

Hosking, J. R. M. and J. R. Wallis. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, New York, 224 pp.

Hoyois, P. and D. Guha Sapir, 2003. Three decades of flood in Europe: a preliminary analysis of EMDAT data. Working draft prepared for the Centre for Research on the Epidemiology of Disasters. http://www.emdat.be/publications?tid=19.

Hundecha, Y., A. St-Hilaire, T. B. M. J. Ouarda, and S. El Adlouni. 2008. A nonstationary extreme value analysis for the assessment of changes in extreme annual wind speed over the Gulf of St. Lawrence, Canada. *Journal of Applied Meteorology and Climatology*, 47: 2745-2759, doi:10.1175/2008JAMC1665.1.

Ilorme F. and V. W. Griffis. 2013. A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology*, 492: 151–162.

Irwin, R. W. and H. R. Whiteley. 1983. Effects of land drainage on streamflow. *Canadian Water Resources Journal*, 8(2): 88-103.

Jain, S. and U. Lall. 2000. Magnitude and timing of annual maximum floods: Trends and large-scale climatic associations for the Blacksmith Fork River, Utah. *Water Resources Research*, 36(12): 3641-3651.

Jonathan, P., K. Ewans, and D. Randell. 2013a. Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, 79: 22-31.

Jonathan, P., D. Randell, Y. Wu, and K. Ewans. 2013b. Nonstationary conditional extremes of northern North Sea storm characteristics. *Environmentrics*, 25: 172-188 doi:10.1002/env.2262.

Jonathan, P., D. Randell, Y. Wu, and K. Ewans. 2014. Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88: 520-532.

Katz, R. W. 2010. Statistics of extremes in climate change. *Climatic Change*, 100: 71-76.

Katz, R. W., M. B. Parlange, P. Naveau. 2002. Statistics of extremes in hydrology. *Advances in Water Resources* 25: 1287-1304.

Kay, A. L. and D. A. Jones. 2012. Transient changes in flood frequency and timing in Britain under potential projections of climate change. *International Journal of Climatology*, 32: 498-502.
Kendall, M. G. 1975. *Rank Correlation Methods*, Griffin, London, 202 pp.
Khaliq, M. N., T. B. M. J. Ouarda, P. Gachon, L. Sushama, and A. St-Hilaire. 2009. Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. *Journal of Hydrology*, 368: 117-130.
Khaliq, M. N., T. B. M. J. Ouarda, J. -C. Ondo, P. Gachon, and B. Bobée. 2006. Frequency analysis of a sequence of dependent and/or non-sationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329: 534-552.
Kharin, V. V.. and F. W. Zwiers. 2005. Estimating extremes in transient climate change simulations. *Journal of Climate*, 18: 1156-1173.
King, K. W., N. R. Fausey, and M. R. Williams. 2014. Effect of subsurface drainage on streamflow in agricultural headwater watershed. *Journal of Hydrology*, 519: 438-445.
Koenker, R. 2005. *Quantile Regression*. Cambridge University Press, New York, 352 pp.
Koenker, R. and B. G. Basset. 1978. Regression Quantiles. *Econometrica*, 46(1): 33-50.
Konrad, C. P. 2003. *Effects of urban development on floods*. USGS fact sheet FS-076-033. https://pubs.usgs.gov/fs/fs07603/pdf/fs07603.pdf.
Kreiss, J. –P. 1992. Boostrap procedures for AR(∞) – processes. In Bootstrapping and Related Techniques, Trier. 1990 [K.–H. Jöckel, G. Rothe, and W. Sendler (eds.)]. *Lecture Notes in Economics and Mathematical Systems*, 376: 107-113.
Kreiss, J. –P., E. Paparoditis, and D. N. Politis. 2011. On the range of validity of the autoregressive sieve Bootstrap. *The Annals of Statist*ics, 39(4): 2103-2130.
Kulkarni, A. and H. von Storch. 1995. Monte carlo experiments on the effects of serial correlation on the MannKendall test of trends. *Meteorologische Zeitschrift N.F*, 4(2): 82-85.
Kundzewicz, Z. W., Y. Hirabayashi, and S. Kanae. 2010. River floods in the changing climate-observations and projections. *Water Resources Management*, 24: 2633-2646, doi:10.1007/s11269-009-9571-6.
Kundzewicz, Z. W. and A. J. Robson. 2000. Detecting trend and other changes in hydrological data. *World Climate Program-Data and Monitoring*. World Meteorological Organization, Geneva (WMO/TD-No. 1013).
Kundzewicz, Z. W. and A. J. Robson. 2004. Change detection in hydrological records – a review of the methodology. *Hydrological Sciences Journal*, 49(1): 7-19.
Kyselý, J. and R. Beranová. 2009. Climate-change effects on extreme precipitation in central Europe: Uncertainties of scenarios based on regional climate models. *Theoretical and Applied Climatology*, 95: 361-374.
Kyselý, J., J. Picek, and R. Beronová. 2010. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72: 55-68.
Lang, M., T. B. M. J. Ouarda, and B. Bobée. 1999. Towards operational guideline for over-threshold modeling. *Journal of Hydrology*, 225: 103-117.
Leclerc, M. and T. B. M. J. Ouarda. 2007. Non-stationary regional flood frequency analysis at ungauged sites. *Journal of Hydrology*, 343: 254-265.
Lehmann, E. L. 1975. *Nonparametrics: statistical methods based on ranks*. Holden-Day, Inc, California, 457 pp.
Lettenmaier, D. P. 1976. Detection of trends in water quality data from records with dependent observation. *Water Resources Research*, 12(5): 1037-1046.

Leopold, L. B. 1968. *Hydrology for urban land planning – A guidebook on the hydrologic effects of urban land use*. United States Department of the Interior. 18 pp.

Li, J. and S. Tan. 2015. Nonstationary flood frequency analysis for annual flood peak series, adopting climate indices and check dam index as covariates. *Water Resources Management*, 29: 5533-5550.

Lins, H. F. and J. R. Slack. 1999. Streamflow trends in the United States. *Geophysical Research Letters*, 26(2): 227-230.

Madramootoo, C. A. and R. S. Broughton. 1987. A computer simulation of surface and subsurface flow from agricultural areas. *Canadian Water Resources Journal*, 12(1): 30-43, doi:10.4296/cwrj1201030.

Madramootoo, C. A., P. Enright, V. T. V. Ngyyen, R. S. Broughton, and S. O. Prasher. 1988. Impact of the July 14, 1987 rainstorm on agricultural runoff in western Quebec. *Canadian Water Resources Journ*al, 13(2): 18-26, doi: 10.4296/cwrj1302018.

Mailhot, A., S. Lachance-Cloutier, G. Talbot, and A.-C. Favre. 2013. Regional estimates of intense rainfall based on peaks-over-threshold (POT) approach. *Journal of Hydrology*, 476: 188-199.

Mann, H. B. 1945. Nonparametric tests against trend. *Econometrica*, 13: 245-259.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis. 1997. A Pacific interdecadal climate oscillation with impact on salmon production. *Bulletin of the American Meteorological Society*, 78: 1069-1079.

Martins, E. S. and J. R. Stedinger. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3): 737-744.

McLeod, A. I. and W. K. Li. 1983. Diagnostic Checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4(4): 269-273.

Meyn, A., S. Schmidtlein, S. W. Taylor, M. P. Girardin, K. Thonicke, and W. Cramer. 2013. Precipitation-driven decrease in wildfires in British Columbia. *Regional Environmental Change*, 13: 167-177.

Milly, P. D. D., J. Betancourt, M. Falkenmark, R. Hirsch, W. Zbigniew, D. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer. 2008. Stationarity is dead: Whither water management? *Science*, 319(5863): 573-574.

Moore, R. D., D. L. Spittlehouse, P. H. Whitfield, and K. Stahl. 2010. Chapter 3: Weather and Climate. In: R. G. Pike, T. E. Redding, R. D. Moore, R. D. Winkler, and K. D. Bladon [edt]. *Compendium of forest hydrology and geomorphology in British Columbia*. BC Ministry of Forests and Range, Research Branch, Victoria, BC and FORREX Forest Research Extension Partnership, Kamloops, BC.

Morgan, A., B. Branfireun, and F. Csillag. 2004. And evaluation of the contributions of urbanization and climatic change to runoff characteristics in the Laurel Creek Watershed, Ontario. *Canadian Water Resources Journal*, 29(3): 171-182.

Mote, P. W. 2003. Twentieth-century fluctuations and trends in temperature, precipitation, and Mountain snowpack in the Georgia Basin – Puget Sound region. *Canadian Water Resources Journal*, 28(4): 567-585.

Muma, M., A. N. Rousseau, and S. L. Gumiere. 2016. Assessment of the impact of subsurface agricultural drainage on soil water storage and flows of a small watershed. *Water*, 8, 326, doi:10.3390/w8080326.

Nadarajah, S. 2005. Extremes of daily rainfall in west central Florida. *Climate Change*, 63: 325-342.

Nalley, D., J. Adamowski, B. Khalil, B. 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954-2008). *Journal of Hydrology*, 475: 204-228.

Nalley, D., J. Adamowski, B. Khalil, B. Ozga-Zielinski. 2013. Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967-2006 using the discrete wavelet transform. *Atmospheric Research*, 132-133: 375-398.

National Weather Service (NWS). 2015. Hydrologic Information Center – Flood Loss Data. http://www.nws.noaa.gov/hic/.

Nelder, J. A. and R. Mead. 1965. A simplex method for function minimization. *Computer Journal*, 7: 308-313.

Noguchi, K., Y. R. Gel, and C. R. Duguay. 2011. Bootstrap-based tests for trends in hydrological time series, with application to ice phenology data. *Journal of Hydrology*, 410: 150-161.

Norbet, J., M. Mugo, and H. Gadain. 2014. Estimation of design floods in ungauged catchments using a regional index flood method. A case study of Lake Victoria in Kenya. *Physics and Chemistry of the Earth*, 67-69: 4-11.

Northrop, P. J. and P. Jonathan. 2011. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7): 799-809, doi:10.1002/env.1106.

North Vancouver District (NVD). 2015. *Guide to living near steep slopes*. http://www.dnv.org/article.asp?c=1030.

O'Brien, N. L. and D. H. Burn. 2014. A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. *Journal of Hydrology*, 519: 2040-2048.

O'Brien, N. L. and D. H. Burn. 2018. A Nonstationary Peaks-Over-Threshold Approach for Modelling Daily Precipitation in Coastal British Columbia. *Canadian Water Resources Journal*, DOI: 10.1080/07011784.2018.1455538

Önöz, B. and M. Bayazit. 2012. Block bootstrap for Mann-Kendall trend test of serially dependent data. *Hydrological Processes*, 26: 3552-3560.

Ontario Ministry of the Environment (OME). 2008. Design Guidelines for Sewage Works. ISBN 978-1-4249-8438-1. https://www.ontario.ca/document/design-guidelines-sewage-works.

Osman, Y. Z., R. Fealy, and J. C. Sweeney. 2013. Downscaling extreme precipitation in Ireland using combined peaks-over-threshold generalized Pareto distribution model of varying parameters. *Journal of Water and Climate Change*, 4(4):21-47, doi:10.2166/wcc.2013.071.

Ouarda, T. B. M. J. and S. El-Adlouni. 2011. Bayesian nonstationary frequency analysis of hydrological variables. *Journal of the American Water Resources Association*, 47(3): 496-505.

Pettitt, A. N. 1979. A non-parametric approach to the change-point problem. *Journal of Applied Statistics*, 28: 126-135.

Pickands, J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1): 119-131.

Politis, D. N. 2003. The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2): 219-230.

Prein, A. F, M. R. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland. 2016. The future intensification of hourly precipitation extremes. *Nature Climate Change*, 7(1): 48-53, doi:10.1038/NCLIMATE3168.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.

Rahman, M. M., Z. Lin, X. Jia, D. D. Steele, and T. M. DeSutter. 2014. Impact of subsurface drainage drainage on streamflow. *Journal of Hydrology*, 511: 474-483.

Rajagopalan, B. 2010. Hydrologic frequency analysis in a changing climate. *Workshop on Nonstationarity, Hydrologic Frequency Analysis, and Water Management, Jan 13-15.*, Millennium Harvest House, Boulder, Colorado, U.S.A: 107-114.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108, 4407, doi.10.1029/2002JD002670.

Reed D. W., D. Jakob, and A. J. Robson. 1999. Choosing a pooling-group, chapter C6. In: *Flood Estimation Handbook, vol. 3, Statistical Procedures for Flood Frequency Estimation*. Institute of Hydrology, Wallingford, UK. 338 pp.

Renard, B., V. Gerreta, and M. Lang. 2006. An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42, W12422, doi:10.1029/2005WR004591.

Rivard, C. and H. Vigneault. 2009. Trend detection in hydrological series: when series are negatively Correlated. *Hydrological Processes*, 23: 2737-2743.

Robson, M. 1990. Impact of improved land drainage on river flows. *Institute of Hydrology*, Report no. 113.

Robson, A. J. and Reed, D. W., 1999. Flood Estimation Handbook, vol. 3, Institute of Hydrology, Wallingford, UK., 338 pp.

Rogger, M., M. Agnoletti, A. Alaoui, J. C. Bathurst, G. Bodner, M. Borga, V. Chaplot, F. Gallart, G. Glatzel, J. Hall, J. Holder, L. Holko, R. Horn, A. Kiss, S. Kohnová, G. Leitinger, B. Lennartz, J. Parajka, R. Perdigão, S. Peth, L. Plavcová, J. N. Quinton, M. Robinson, J. L. Slinas, A. Santoro, J. Szolgay, S. Tron, J. J. H. van den Akker, A. Viglione, and G. Blöschl. 2017. Land use change Impacts on floods at the catchment scale: Challenges and opportunities for future research. *Water Resources Research*, 52: 5209- 5219, doi: 10.1002/2017WR020723.

Rosburg, T. T., P. A. Nelson, and B. P. Bledsoe. 2017. Effects of urbanization on flow duration and stream flashiness: A study of Puget Sound streams, western Washington, USA. *Journal of the American Water Resources Association*, 53(2): 493-507.

Rose, S. and N.E. Peters. 2001. Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrological Processes*, 15: 1441-1457.

Rosner, A., R. M. Vogel, and P. H. Kirshen. 2014. A risk-based approach to flood management decisions in a nonstationary world. *Water Resources Research*, 50: 1928-1952. doi: 10.1002/2013WR014561.

Roth, M., T. A. Buishand, G. Jongbloed, A. M. G. Klein Tank, and J. H. van Zanten. 2012. A regional peaks-over-threshold model in a nonstationary climate, *Water Resources Research*, 48, W11533, doi:10.1029/2012WR012214.

Roth, M., T. A. Buishand, G. Jongbloed, A. M. G. Klein Tank, and J. H. van Zanten. 2014. Projections of precipitation extremes based on a regional non-stationary peaks-over-threshold approach: A case study for the Netherlands and north-western Germany, *Weather and Climate Extremes*, 4: 1-10.

Roth, M., G. Jongbloed, and T. A. Buishand. 2016. Threshold selection for regional peaks-over-threshold data. *Journal of Applied Statistics*, 43(7): 1291-1309.

Rozemeijer, J. C. and H. P. Broers. 2007. The groundwater contribution to surface water contamination in a region with intensive agricultural land use (Noord-Brabant, The Netherlands). *Environmental Pollution*, 148: 695-706.

Schmocker-Fackel, P. and F. Naef. 2010. More frequent flooding? Changes in flood frequency in Switzerland since 1850. *Journal of Hydrology*, 381: 1-8.

Seaburn, G. E. 1969. Effects of urban development on direct runoff to East Meadow Brook, Nassau County, Long Island, New York. *Geological Survey Professional Paper, 627-B*.

Sen, P. K. 1968. Estimates of the regression coefficient based on Kendall's Tau. *Journal of the American Statistical Association*, 63(324): 1379-1389.

Serinaldi, F. and C. G. Kilsby. 2016. The importance of prewhitening in change point analysis under persistence. *Stochastic Environmental Research and Risk Assessment*, 30(2): 763-777.
Serrano, S. E. 1985. Effects of agricultural drainage on streamflow in the Middle Thames River, Ontraio1949-1980. *Canadian Journal of Civil Engineering*, 12: 875-885.
Skaggs, R. W., M. A. Brevé, and J. W. Gilliam. 1994. Hydrologic and water quality impacts of agricultural drainage. *Critical Reviews in Environmental Science and Technology*, 24(1): 1-32.
Sivapalan, M. and J. M. Samual, 2009. Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment. *Hydrological Processes*, 23: 1671-1675.
Smith, J. A., 1989. Regional frequency analysis using extreme order statistics of the annual peak record. *Water Resources Research*, 25(2): 311-317.
Sneyers, R. 1990. *On the statistical analysis of series of observations*. World Meteorological Organization, Technical Note no. 143, WMO no. 415, 192 pp.
Sonali, P. and D. N. Kuman. 2013. Review of trend detection methods and their application to detect temperature changes in India. *Journal of Hydrology*, 476: 212-227.
Spaling, H. 1995. Analyzing cumulative environmental effects of agricultural land drainage in southern Ontario, Canada. *Agricultural Ecosystems and Environment*, 53: 279-292.
Statistics Canada. 2011a. Population, urban and rural, by province and territory (Canada). http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo62a-eng.htm.
Statistics Canada, 2011b. Population, urban and rural, by province and territory (Ontario). http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo62g-eng.htm.
Stephens, K. A., P. Graham, D. Reid. 2002. *Stormwater Planning: A Guidebook for British Columbia*. Province of British Columbia. http://www.env.gov.bc.ca/epd/epdpa/mpp/stormwater/guidebook/pdfs/stormwater.pdf.
Stone, D. A., A. J. Weaver, and F. W. Zwiers. 2000. Trends in Canadian Precipitation Intensity. *Atmosphere-Ocean*, 38(2): 321-347.
Sugahara, S., R. Porfírio da Rocha, and R. Silveira. 2009. Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology*, 29: 1339-1349.
Sungwook, W., J. B. Valdés, S. Steinschneider, and T. -W. Kim. 2016. Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. *Stochastic Environmental Research and Risk Assessment*, 30: 583-606.
Svensson, C., Z. W. Kundzewicz, and T. Maurer. 2005. Trend detection in river flow series: 2. Floods and low-flow index series. *Hydrological Sciences Journal*, 50(5): 811-823.
Theil, H. 1950. A rank invariant method of linear and polynomial regression analysis. I. *Proceedings of the Royal Netherlands Academy of Sciences*, 53: 386-392.
Theobald, D. M., S. J. Goetz, J. B. Norman, and P. Jantz. 2009. Watersheds at risk to increased impervious surface cover in the conterminous United States. *Journal of Hydrologic Engineering*, 14(4): 362-368.
Thompson, P. J. 2013. *Event based characterization of hydrologic change in urbanizing southern Ontario watersheds via high resolution stream gauge data*. (Unpublished master's thesis). University of Waterloo, Waterloo, ON, 259 pp.
Tramblay, Y., L. Neppel, J. Carreau, and K. Najiib. 2013. Non-stationary frequency analysis of heavy rainfall events in southern France. *Hydrological Sciences Journal*, 58(2): 280-294.
Turtola, E. and A. Paajanen. 1995. Influence of improved subsurface drainage on phosphorus losses and Nitrogen leaching from a heavy clay soil. *Agricultural Water Management*, 28: 295-310.
Villarini G., F. Serinaldi, J. A. Smith, J, and W. F. Krajewski. 2009a. On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, 45, W08417, doi: 10.1029/2008WR007645.

Villarini, G., J. A. Smith, and F. Napolitano. 2010. Nonstationary modeling of a long record of rainfall an temperature over Rome. *Advances in Water Resources*, 33: 1256-1267.

Villarini, G., J. A. Smith, A. A. Ntelekos, and U. Schwarz. 2011. Annual maximum and peaks-over-threshold analysis of daily rainfall accumulations for Austria. *Journal of Geophysical Research*, 116, D05103,doi: 10.1029/2010JD015038.

Villarini, G., J. A. Smith, F. Serinaldi, J. Bales, P. D. Bates, and W. F. Krajewski. 2009b. Flood frequency analysis for nonstationary annual peak records in an urban drainage area. *Advances in Water Resources*, 32: 1255-1266.

Vincent, L. A. and É. Mekis. 2006. Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. *Atmosphere-Ocean*, 44(2): 177-193.

Vincent, L. A. and É. Mekis. 2009. Discontinuities in due to joining precipitation station observations in Canada. *Journal of Applied Meteorology and Climatology*, 48: 156-166.

Vinnikov, K. Y. and A. Robock. 2002. Trends in moments of climate indices. *Geophysical Research Letters*, 29(2), doi:10.1029/2001GL014025.

Vogel, R. M, Yaindl, C., and Walter, M., 2011. Nonstationarity: flood magnification and recurrence reduction factors in the United States. *Journal of the American Water Resources Association*, 47(3): 464-474.

von Storch, V. H. 1995. Misuses of statistical analysis in climate research, In: *Analysis of Climate Variability: Applications of Statistical Techniques*, ed. von H. V. Storch and A. Navarra A. Springer-Verlag, Berlin: 11-26.

Walker, I., R. Sydneysmith, D. Allen, K. Bodtker, D. Bonin, B. Bonsal, S. Cohen, *et al*. 2008. Chapter 8: British Columbia. In: D. S. Lemmen, F. J. Warren, J. Lacroix, and E. Bush [edt]. *From Impacts to Adaptation: Canada in a Changing Climate 2007*. Government of Canada, Ottawa: 329-386.

Wang. W., Y. Chen, S. Becker, and B. Liu. 2015. Variance correction prewhitening method for trend detection in autocorrelated data. *Journal of Hydrologic Engineering*, 20(12): 04015033.

Westra, S., L. V. Alexander, and F. W. Zwiers. 2013. Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, 26(11): 3904-3918.

White, M. D. and K. A. Greer. 2006. The effects of watershed urbanization on the stream hydrology and riparian vegetation of Los Peñasquitos Creek, California. *Landscape and Urban Planning*, 74: 125-138.

Whitfield, P. H., K. Bodtker, and A. J. Cannon. 2002. Recent variations in seasonality of temperature and precipitation in Canada, 1976-95. *International Journal of Climatology*, 22: 1617-1644.

Whitfield, P. H., R. D. Moore, S. W. Flemming, and A. Zawadzki. 2010. Pacific decadal oscillation and the hydroclimatology of western Canada – review and prospects. *Canadian Water Resources Journal*. 35(1): 1-28.

Wright, M. J., C. M. Ferreira, and M. H. Houck. 2014. Evaluation of heterogeneity statistics as reasonable proxies of the error of precipitation quantile estimation in the Minneapolis-St. Paul region. *Journal of Hydrology*, 513: 457-466.

Yue, S. and P. Pilon. 2004. A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrological Sciences Journal*, 49(1): 21-37.

Yue, S., P. Pilon, and G. Cavadias. 2002a. Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, 259: 254-271.

Yue S., P. Pilon, and B. Phinney. 2003. Canadian streamflow trend detection: impacts of serial and cross-correlation, *Hydrological Sciences Journal*, 48(1): 51-63.

Yue, S., P. Pilon, B. Phinney, and G. Cavadias. 2002b. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological processes*, 16: 1807-1829.

Yue, S. and C. Y. Wang. 2002. Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test. *Water Resources Research*, 38(6), doi:10.1029/2001WR000861.

Yue, S. and C. Wang. 2004a. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management*, 18: 201-218.

Yue, S. and C. Wang. 2004b. Reply to comment by Mehmetcik Bayazit and Bihrat Önöz on "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test". *Water Resources Research*, 40, doi:10.1029/2003WR002555.

Zhang, X., K. D. Harvey, W. D. Hogg, and T. R. Yuzyk. 2001. Trends in Canadian streamflow. *Water Resources Research*, 37(4): 987-998.

Zhang, X., L. A. Vincent, W. D. Hogg, and A. Niitsoo. 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean*, 38(3): 395-429.

Zhang, X. and F. W. Zwiers. 2004. Comment on "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test" by Sheng Yue and Chun Yuan Wang. *Water Resources Research*, 40, W03805, doi:10.1029/2003WR002073.