

738 YEARS OF GLOBAL CLIMATE MODEL SIMULATED STREAMFLOW IN  
THE NELSON-CHURCHILL RIVER BASIN

By  
Michael John Fernandes Vieira

A thesis submitted to the Faculty of Graduate Studies of  
The University of Manitoba  
In partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE

Department of Civil Engineering  
University of Manitoba  
Winnipeg, Manitoba

January 2016  
© 2016, by Michael John Fernandes Vieira

## ABSTRACT

Uncertainty surrounds the understanding of natural variability in hydrologic extremes such as droughts and floods and how these events are projected to change in the future. This thesis leverages Global Climate Model (GCM) data to analyse 738 year streamflow scenarios in the Nelson-Churchill River Basin. Streamflow scenarios include a 500 year stationary period and future projections forced by two forcing scenarios.

Fifty three GCM simulations are evaluated for performance in reproducing observed runoff characteristics. Runoff from a subset of nine simulations is routed to generate naturalized streamflow scenarios. Quantile mapping is then applied to reduce volume bias while maintaining the GCM's sequencing of events.

Results show evidence of future increases in mean annual streamflow and evidence that mean monthly streamflow variability has decreased from stationary conditions and is projected to decrease further into the future. There is less evidence of systematic change in droughts and floods.

## ACKNOWLEDGEMENTS

I sincerely thank my thesis advisor, Dr. Tricia Stadnyk, for all her mentorship and support. I also thank my advisory committee including Dr. Stadnyk, Dr. Peter Rasmussen and Dr. Ron Stewart for their patience and encouragement as I undertook graduate studies part time.

I thank my supervisor, Kristina Koenig, and colleagues Mark Gervais and Phillip Slota at Manitoba Hydro for their support, inspiration and consult. I also thank my manager Efrek Teklemariam for the opportunity to pursue graduate studies while working full time.

I thank the many mentors who helped with guidance, data collections, methodology refinement, and incorporation of good practices into my thesis. Namely, the support of Blaise Gauvin St-Denis and other colleagues at the Ouranos Consortium on Regional Climatology and Adaptation to Climate Change. Ouranos' support was instrumental in completing this thesis. I would also like to thank Dr. Nicholas Kouwen for providing code and support for implementing a modified version of his WATROUTE model.

I acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and I thank the climate modeling groups (listed in Table 5 of this thesis) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Finally, I would like to sincerely thank my family and friends for their continuous moral support and encouragement in my academic studies.

# TABLE OF CONTENTS

<b>Introduction.....</b>	<b>1</b>
1.1 Objectives .....	5
1.2 Thesis Organization .....	6
<b>Literature Review .....</b>	<b>7</b>
2.1 Traditional Assessment of Long Term Hydrology .....	9
2.1.1 Stochastic Approaches .....	9
2.1.2 Proxy Approaches .....	14
2.2 Climate Modeling .....	17
2.2.1 Developing Future Climate Scenarios .....	17
2.2.2 Climate Models, Evaluation and Bias Correction.....	24
2.3 Application of Climate Models in Assessment of Long Term Hydrology .....	37
2.3.1 River and Lake Routing of Climate Model Runoff Data.....	39
2.3.2 Future Runoff and Streamflow Scenarios from Climate Models .....	42
<b>Spatial and Temporal Domain.....</b>	<b>48</b>
<b>Methodology and Data Sources .....</b>	<b>52</b>
4.1 Evaluation of Climate Model Skill .....	52
4.2 River and Lake Routing.....	58
4.3 Bias Correction .....	67
4.4 Streamflow Scenarios and Time Series Analysis .....	70
<b>Results and Discussion.....</b>	<b>76</b>
5.1 Evaluation of Climate Model Skill .....	76
5.2 River and Lake Routing using WATROUTE <sub>MOD</sub> .....	81
5.3 Bias Correction of GCM-driven WATROUTE <sub>MOD</sub> Streamflow .....	89
5.4 Streamflow Scenarios and Time Series Analysis .....	94
5.5 Uncertainty.....	120
<b>Conclusions and Recommendations.....</b>	<b>124</b>
6.1 Evaluation of Climate Model Skill .....	124
6.2 River and Lake Routing .....	125
6.3 Bias Correction .....	127
6.4 Streamflow Scenario and Time Series Analysis.....	129
<b>References .....</b>	<b>136</b>

## LIST OF TABLES

Table 1 - Description of Representative Concentration Pathways (van Vuuren <i>et al.</i> , 2011).....	19
Table 2 - Nelson-Churchill River Basin and sub-basin mean annual flow and runoff for 1950-2005.....	49
Table 3 - CMIP5 temporal domains and additional temporal domains used in analysis.....	51
Table 4 - Skill metrics for GCM evaluation .....	52
Table 5 - GCM name, institution and land surface model name (adapted from Table 9.A.1 in Flato <i>et al.</i> , 2013).....	56
Table 6 - GCM simulation data availability from various CMIP5 experiments (as of January, 2014). For historic experiments, S denotes simulations containing a shorter record (1950-2005) and L denotes simulations containing a longer record (1861-2005).....	57
Table 7 - Physiographic comparison of WATROUTE (from Stadnyk and Newsom, 2013) and WATROUTE <sub>MOD</sub> models by NCRB sub-basin .....	59
Table 8 - Criteria for assigning degree of consistency with statistical evidence (DOC <sub>STAT</sub> ) and degree of consistency without statistical evidence (DOC <sub>NON-STAT</sub> ) .....	75
Table 9 - 1950-2005 observed (LTFD) values for skill evaluation performance metrics. ....	77
Table 10 - Physiographic comparison of the Winnipeg River Sub-Basin WATROUTE model and the Winnipeg River Sub-Basin in WATROUTE <sub>MOD</sub> . ....	86
Table 11 - Coefficients of Variation (CV) for mean annual NCRB streamflow in different periods. CV shown as $*10^{-3}$ . ....	97
Table 12 - Coefficients of Variation (CV) for mean monthly NCRB streamflow in different periods. CV shown as $*10^{-3}$ . ....	102
Table 13 - MW-U Non-exceedance probabilities for NCRB mean annual streamflow. Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations.....	105
Table 14 - MW-U Non-exceedance probabilities for NCRB multi-year drought severity. Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations.....	112
Table 15 - MW-U Non-exceedance probabilities for NCRB multi-year drought duration. Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations. ....	112
Table 16 - MW-U Non-exceedance probabilities for NCRB maximum monthly streamflow. Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations. ....	118
Table 17 - Future GCM projection summary for mean values of hydrological variables.....	120
Table 18 - Future GCM projection summary for extreme values of hydrological variables.....	120

## LIST OF FIGURES

Figure 1 - RCP CO <sub>2</sub> concentrations (left) and projected global surface warming (right). Data extracted on July 8, 2013 from KNMI Climate Explorer (climexp.knmi.nl; van Oldenborgh and Burgers, 2005).....	20
Figure 2 - Illustrative quantile mapping example for correcting monthly flow data in a reference period and future period. ....	33
Figure 3 - WATROUTE channel cross section .....	40
Figure 4 - The Nelson-Churchill River Basin and sub-basin spatial domains.....	48
Figure 5 - WATROUTE <sub>MOD</sub> Routing model domain for the NCRB .....	61
Figure 6 - GCM Skill in reproducing annual runoff volume and variability relative to the observed (LTFD) record for the NCRB (1950-2005).....	77
Figure 7 - RMSE portrait diagram of GCM skill for 1950-2005. Additional information on metrics and their units can be found in Table 4 and Table 9. ....	78
Figure 8 - Relative RMSE (RMSE') portrait diagram of GCM skill for 1950-2005. Additional information on metrics can be found in Table 4 and Table 9. ....	78
Figure 9 - GCM skill - mean monthly runoff for the aggregated NCRB. Left panel shows LTFD observations (1950-2005), the ensemble of 52 GCM simulations (1950-2005) and GRDC estimated observations (multiple year ranges considered). Right panel shows subset of nine GCM simulations selected for further analysis. ....	81
Figure 10 - WATROUTE <sub>MOD</sub> vs. WATROUTE at the Winnipeg River Basin outlet for a synthetic runoff event .....	82
Figure 11 - Uncorrected GCM-driven WATROUTE <sub>MOD</sub> mean annual hydrographs for CMIP5 historic period (1861-2005). Colours denote various locations in the NCRB. ...	89
Figure 12 - Bias correction adjustment values by NCRB sub-basin and GCM.....	90
Figure 13 - Observed and bias corrected mean monthly aggregated NCRB streamflow (1912-2005) for individual GCM simulations.....	92
Figure 14 - Time series of bias corrected mean annual NCRB aggregated streamflow (1362-2099) for individual GCM simulations.....	96
Figure 15 - Bias corrected piControl, historic and future monthly mean hydrographs for aggregated NCRB streamflow. Time periods are provided in Table 3.....	99
Figure 16 - Mean annual streamflow box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels. ....	103
Figure 17 - Drought severity box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels.....	108
Figure 18 - Drought duration box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels.....	109
Figure 19 - Maximum monthly flow box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels. ....	115

## LIST OF ACRONYMS AND SYMBOLS

AOGCM	Atmosphere-Ocean General Circulation Model
CMIP3	Coupled Model Intercomparison Project Phase 3
CMIP5	Coupled Model Intercomparison Project Phase 5
CO <sub>2</sub>	Carbon dioxide
CRCM	Canadian Regional Climate Model
CRD	Churchill River Diversion
°C	Degrees Celsius (temperature)
CV	Coefficient of Variation
DEM	Digital Elevation Model
DOC <sub>STAT</sub>	Degree of Consistency supported by statistical evidence
DOC <sub>NON-STAT</sub>	Degree of Consistency not supported by statistical evidence
DRT	Dominant River Tracing
ECDF	Empirical Cumulative Distribution Function
ESM	Earth System Model
GCM	Global Climate Model (individual model details listed in Table 5)
GWh	Gigawatt hours (energy)
km <sup>2</sup>	Squared kilometers (area)
Lake Winnipeg PIAO	Lake Winnipeg Partial Inflow Available for Outflow sub-basin
LTFD	Manitoba Hydro's Long Term Flow Data
m <sup>3</sup>	Cubic meters (volume)
m <sup>3</sup> /s	Cubic meters per second (flow rate)
mm	Millimeter (depth)
mrro	Total runoff (as per CMIP5 variable name in Taylor, 2013)
MW	Megawatt (power)
MW-U	Mann-Whitney-Wilcoxon U test
NCRB	Nelson Churchill River Basin
NRL	Nelson River Local sub basin
PDSI	Palmer Drought Severity Index
piControl	Preindustrial Control
pi1, pi2, pi3, etc.	Various 91-94 year periods from the 500 year preindustrial control period
ppm	Parts per million (concentration)
QM	Quantile Mapping
RCP	Representative Concentration Pathway
RMSE	Root Mean Squared Error
SRES	The IPCC's Special Report on Emission Scenarios
r1, r2, r3, etc.	Various GCM member runs
SPI	Standard Precipitation Index
SRB	Saskatchewan River sub-basin
UCR	Upper Churchill River sub-basin
W/m <sup>2</sup>	Watts per square meter (power per unit area)
WRB	Winnipeg River sub-basin

# CHAPTER 1

## Introduction

A reliable water supply is a basic requirement for civilization, important for many purposes including drinking, agriculture and energy production. Traditionally, hydrologic infrastructure which civilization relies on is planned, designed and operated based on observations (e.g., precipitation, streamflow), and is often designed to withstand extreme events (Olsen *et al.*, 2015). However, observations sample a finite period of time that may or may not adequately capture the full range of natural variability. While considerable information resides within observational records, the question of how observations compare to past occurrences and future projections remains. Traditionally, hydrologists use stochastic and paleo approaches to supplement the understanding of natural variability and past hydrology. Evolving climate science coupled with modeling is also used to enhance the understanding of hydrology and how hydrological processes are projected to change into the future.

Depending on the application, water supply may be defined using different variables and at various temporal and spatial scales. For example: meteorology and urban planning may consider hourly precipitation intensity or daily total precipitation at a specific location. Agricultural studies may consider seasonal precipitation, soil moisture and ground water levels at a regional scale. Frequency-based design for infrastructure may consider the magnitude and timing of daily peak streamflow (flood control infrastructure) or peak water levels (transportation infrastructure). Drinking water supply infrastructure may consider annual precipitation, watershed yield, and storage capacity to reliably supply daily demand. And for large hydropower systems with manageable storage, water supply parameters of interest include monthly and

annual inflows across large regions, or watersheds. Where many variables exist to define water supply (Koshida, Cohen and Mortsch, 2015), they are all intrinsically linked through the hydrological cycle. In many cases, streamflow is a fundamental parameter that represents area aggregated water supply within a basin of interest.

In Manitoba, Canada, Manitoba Hydro provides approximately 98% of the electric energy via 15 hydroelectric generating stations on five river systems (Manitoba Hydro, 2013a). Water supply for these generating stations is characterized in terms of monthly streamflow within the Nelson-Churchill River Basin (NCRB). The NCRB spans a diverse geographic area covering approximately 1.4 million square kilometres, draining into Hudson Bay in northern Manitoba. Over the observation period from 1912 to 2014, the NCRB has experienced water supply conditions that include both extended dry periods and large flood events. Periods of low streamflow are termed hydrologic droughts and are important considerations for the planning, design and operation of Manitoba's energy infrastructure. Drought is identified as a corporate risk with an estimated impact of up to \$2 billion (Manitoba Hydro, 2007). Manitoba Hydro is primarily concerned with energy drought (a form of socio-economic drought) instead of hydrologic drought; however, the two variables are closely related.

Changes in mean annual streamflow, mean monthly streamflow, maximum monthly streamflow and multiyear hydrologic droughts are of interest in the planning, design and operation of hydropower infrastructure. Mean streamflow can inform powerhouse capacity design whereas maximum monthly streamflow provides information that can help inform operations planning. A critical, multi-year, hydrologic drought can provide practical dependable energy limits for system planning. At Manitoba Hydro, the critical drought is currently set as the lowest water conditions on record, which occurred in Manitoba in the late 1930s to early 1940s.

Several other major drought events have occurred in the observed record and Manitoba Hydro acknowledges that a drought worse than the drought on record is possible (Manitoba Hydro, 2013b). Manitoba Hydro's regulators and regulatory process interveners have inquired about how climate change might impact the critical drought. As is the case for many global drought studies, no definitive answer currently exists (Trenberth *et al.*, 2014).

Anecdotal evidence such as entries in Hudson's Bay Company archives (Ball, 1983; Rannie, 1999), statistical evidence (Burn and DeWit, 1996; Akintuğ, 2006; Kubursi and Magee, 2010) and paleo evidence (Jones and Mann, 2004; Sauchyn, Vanstone, and Perez-Valdivia, 2011) suggest that past variability may have exceeded observed variability in the NCRB which may have produced more extreme droughts or floods. Further evidence also suggests that a warmer climate can intensify the hydrologic cycle (Huntington, 2006) and can reduce snow accumulation while accelerating snowmelt (Shrestha, Dibike and Prowse, 2012). This evidence along with statements that stationarity assumptions have been compromised (Milly *et al.*, 2008; Todhunter, 2013) raise further questions on the suitability of observed streamflow data for use in future hydrologic assessment. These questions place pressure on engineers and water managers to consider climate change impacts on planning, design and operations (Wood, Lettenmaier and Palmer, 1997). However, due to large uncertainties in projecting future hydrology (Chen *et al.*, 2011) and a lack of established standards to incorporate climate change impacts on future projects, little direction is currently available to guide industry (Olsen *et al.*, 2015). In many regions such as northern Manitoba, these concerns are accompanied with other issues that include shorter observational records and coarse spatial coverage (Coulibaly *et al.*, 2013).

Recent advances in Global Climate Models (GCMs) provide a potential source of information to explore natural variability and future projections. GCMs numerically simulate the

Earth's physics, including hydrological processes. Like many models, GCMs are not perfect and can contain bias when compared to observations (Trenberth, 1997; Flato *et al.*, 2013). Despite challenges in simulating observed patterns, GCMs are tools that can provide insight and enhance our understanding of natural climate variability and hydrologic response to increased greenhouse gas concentrations. Existing studies generally focus on coupling GCM temperature and precipitation projections with a calibrated hydrological model to analyze future streamflow projections (e.g., Wood *et al.*, 1997; Chen *et al.*, 2011; Brekke and Prairie, 2009; Hirabayashi *et al.*, 2008; Bohrn, 2012; Shrestha *et al.*, 2012). These existing studies typically focus on a 30 year baseline period (e.g., 1971-2000) and a 30 year future period (e.g., 2041-2070) and can require substantial computing resources. GCMs from the latest Coupled Model Intercomparison Project Phase 5 (CMIP5) offer finer spatial resolutions and improved land surface schemes with hydrologic schematic complexities similar to those in distributed hydrologic models (Sperna Weiland *et al.*, 2012b), nudging researchers to consider hydrological output from GCMs directly. The methods and results within this thesis directly couple the GCM's internally simulated runoff with a modified version of the WATROUTE (Kouwen, 2012) routing model to analyze 738 years of streamflow in three periods: a 500 year stationary period, a historic period spanning 1861 to 2005, and a future period spanning from 2006 to 2100. These streamflow scenarios leverage GCM data to enhance the understanding of hydrologic processes in the NCRB and analyze future streamflow projections including hydrologic droughts and floods.

## 1.1 Objectives

The overall objective of this research is to leverage 738 years of GCM simulated runoff data to assess how streamflow in the historic period (1912-2005) compares to natural variability and future projections with respect to key hydrological variables in the NCRB. This research focuses on naturalized streamflow in the NCRB over the period of 1362-2099 and relies primarily on simulated GCM data from CMIP5. The methods used in reaching this overall objective represent a simplified approach to addressing a complex issue. Acknowledging the simplifications and uncertainty inherent in this research, results should be viewed as initial academic findings. Results however, will contribute to an enhanced understanding of long term streamflow in the NCRB. In order to effectively address the overall objective, four sub-objectives are identified and used for thesis organization:

1. Evaluate GCM skill in simulating observed hydrological patterns;
2. Develop a model to route GCM runoff and produce streamflow at key locations;
3. Adapt existing bias correction methods, typically used for meteorological variables, for application to the correction of streamflow time series; and
4. Assess how streamflow in the historic period (1912-2005) compares to natural variability and future projections with respect to key hydrological variables using streamflow simulations from 1362-2099.

## **1.2 Thesis Organization**

Chapter two presents a review of literature related to this research and objectives. The literature review builds background information, introduces current understanding, approaches, results and gaps in the research.

Chapter three presents a description of the study area, the NCRB and various sub-basins considered and the various temporal domains (periods).

Chapter four presents the methodology and is divided into sections that align with the research objectives. Sections include methodology for evaluating GCM skill in reproducing observed hydrological patterns (Objective 1), development of a routing model based on a modified WATROUTE scheme (Objective 2), adaptation of existing bias correction methods for use in correcting streamflow time series (Objective 3), and time series analysis of bias corrected streamflow for key hydrological variables (Objective 4).

Chapter five presents results and discussion, following a similar format to chapter four, including subsections that align with each objective.

Chapter six presents conclusions and recommendations for further study.

## CHAPTER 2

### Literature Review

Extending knowledge about regional hydrology, beyond observational records, has been an area of interest in hydrologic literature for many years. The interest applies to extension of records backwards in time and also includes projecting hydrology into the future. A common goal in the scientific literature is to increase understanding of how observed hydrologic records compare to the potential range of natural variability and increase understanding of how hydrology is projected to change in the future. From an academic perspective, this understanding is important for improving climate science and modeling. From a practical perspective, this understanding is important for planning, design and risk assessment; especially in consideration of extreme droughts and floods.

Temporal windows used to define observed and future periods vary among studies but are generally defined as follows: The observed period coincides with a time window where direct measurements (typically instrumental) exist and the future period coincides with a window beyond observations. Theoretically, there is no end to the future period; however, when based on emission scenarios and GCM projections, the future period is limited by the modeled data. The potential range of natural hydrologic variability can be studied using several approaches, which generally require long term (e.g.,  $\geq 500$  years) records. Stochastic approaches, which use statistical models to generate hydrologic variables typically assume stationary conditions (statistical properties do not change with time) and do not have a time signature. Proxy approaches, which relate hydrologic variables to a measurable attribute, provide information that corresponds to specific points in time. Both stochastic and proxy approaches have been applied

in traditional assessments of long term hydrology and can provide information on how hydrology can vary naturally over time (e.g., Burn and DeWit, 1996; St. George, 2007). With recent advances in climate science and modeling, researchers have also turned to GCMs for assessment of long term hydrology (e.g., Milly, Dunne and Vecchia, 2005). As a result of the various approaches for assessing long term hydrology, a wide range of literature exists. Some studies seek to compare observed droughts and floods with stationary stochastic simulations or proxy data, whereas other studies seek to project future changes in average streamflow. Some studies focus on streamflow whereas others focus on precipitation, soil moisture or time integrated variables for drought analysis.

For organization, this literature review is divided into three sections: The first section presents a sample of literature that applies traditional methods for studying hydrologic droughts and floods. The second section presents background literature on climate modeling and the third section presents literature that employs GCMs for assessment of long term hydrology. The three part structure introduces key topics to provide background and understanding on the objectives and methods set in this thesis. The structure follows a somewhat chronological order that begins with traditional approaches, introduces newer climate science and ends with recent applications of newer climate science tools to answer traditional questions. The sections transition from views about the past to views about the future and identify gaps in existing literature to build an understanding for the methods chosen in this thesis. The spatial domains in the reviewed literature range from the global scale to smaller watershed scales. Where possible, a global perspective is presented. However, since many large northern basins have unique hydrological characteristics, such as snowmelt dominated streamflow regimes; this review also focuses on the literature pertaining to the NCRB and other large northern basins. Furthermore, where

information is available for the NCRB, this literature is highlighted as it is relevant to this thesis and its objectives.

## **2.1 Traditional Assessment of Long Term Hydrology**

Traditional assessments of long term hydrology typically follow stochastic approaches or proxy approaches. Stochastic approaches rely on statistical relationships derived from observations to generate synthetic time series of data for analysis. Proxy approaches rely on physically-based relationships between a variable of interest (e.g., streamflow) and a measurable quantity (e.g., tree ring width). Unlike stochastic approaches, proxy approaches have a corresponding time signature. In some studies, the two approaches are blended such that information obtained through proxy approaches helps enhance information obtained from observations to develop stochastic models (e.g., Henley *et al.*, 2011). For the purposes of this literature review, stochastic and proxy approaches are reviewed separately.

### **2.1.1 Stochastic Approaches**

Stochastic approaches have been applied extensively in hydrology, relying on statistical relationships to generate time series data from an assumed distribution. The distinction between stochastic and statistical analysis is that stochastic analyses incorporate a time component (i.e., generates a time series) whereas statistical analyses are used to assess magnitudes only (Caissie and El-Jabi, 1999). Stochastic approaches are practical and computationally inexpensive, which increases their appeal.

Stochastic models, of varying complexity have previously been explored for hydrologic analysis in the NCRB. Burn and DeWit (1996) performed drought analysis using 80,000 years (1000 simulations of 80 years each) of simulated streamflow from the SPIGOT multi-site stochastic model. The analysis used monthly streamflow observations from a period of 1912 to

1990 or from 1957 to 1990, depending on the location. Drought severity, duration and magnitude were considered, however, the authors reasoned that magnitude and duration were less attractive parameters. Burn and DeWit (1996) characterized the most severe observed drought ( $124 \cdot 10^9 \text{ m}^3$ ) as a one in 79 year event when using only the historic record, and as a one in 381 year event based on the stochastic simulations. Furthermore the authors noted that the worst stochastically simulated drought was nearly three times more severe than the observed event.

Akintuĝ (2006) studied system-wide energy drought in the NCRB following a similar approach to Burn and DeWit (1996) but incorporated a longer observed record (1912 to 1998) and tested nine stochastic modeling frameworks. Among the models were a traditional autoregressive model, and a Markov-Switching (MS) model. Autoregressive models were used to predict a given years' streamflow based on the previous year whereas MS models allowed for different behavior in dry and wet regimes. MS models have been used application in hydrology as early as 1975 and have been applied in simulating mean annual discharge for the Niagara River (Akintuĝ and Rasmussen, 2005). Both single site with spatial disaggregation models and multi-site models were applied in the NCRB by Akintuĝ (2006). The importance of maintaining spatial cross correlation was noted, and various generation and disaggregation schemes explored with uncertainties in statistical parameters and data (missing and estimated) considered. One million years were simulated (1000 sequences of 1000 year each) for the various stochastic models. Model results were compared, and advantages and disadvantages of each model were noted. Using the 1912 to 1998 data, return periods for the critical energy drought (3309 gigawatt hours; GWh) ranged from 200 years to 2380 years. Using a shorter period of 1930 to 1998, return periods varied from 120 years to 740 years. Akintuĝ (2006) noted that the single-site

model frameworks may not have been the best choice and also noted issues with the multi-site model's ability to accurately simulate autocorrelation.

Kubursi and Magee (2010) conducted an independent risk review for Manitoba Hydro including an assessment of drought using autoregressive models and extreme value models based on the generalized Pareto distribution. The generalized Pareto distribution was selected for its ability to better account for tail end risk (e.g., drought), which could be underestimated when using normal distributions. Reasoning that approximately 70% of Manitoba Hydro's risk is volumetric, the authors analyzed stochastic simulations of monthly streamflow and noted that long term predictions are valuable in making decisions on future generation investments. Instead of assigning return periods to specific events, the authors instead considered how their 94-year stochastic model simulations compared to observations. Results indicated that the observed minimum annual water flow, minimum five-year mean water flow and minimum monthly water flow was captured within the 95% intervals of the various simulations. However, the models indicated that more extreme (lower) minimum annual water flows and minimum five-year mean water flows were possible. The authors acknowledged that their methods assume stationarity and additional uncertainties exist due to climate change. However, they did not offer guidance on how to incorporate climate change.

Additional studies, not specific to the NCRB, exist that have applied different stochastic approaches. For example, Sadri and Burn (2014) combined statistical distributions from copula families and methods to substitute space for time in shorter records (regional frequency analysis) and analyzed joint probability distributions (severity and duration) of hydrologic droughts on the Canadian prairies. While regional studies have contributed to understanding drought risk at smaller spatial scales, they have offered limited information towards understanding basin-wide

drought risk in spatial domains such as the NCRB. In some cases, the application of methods from regional studies could be applied to enhance understanding in the NCRB.

Stochastic approaches have also been used to assess flood risk. Todhunter (2012) questioned the assumption of stationarity in flood frequency analysis for the 100 year flood on the Red River at Grand Forks, North Dakota and noted that examination of assumptions is required, especially in the future as climate change and other factors can lead to increased flood risk. Similarly, Ehsanzadeh, van der Kamp and Spence (2011) demonstrated monotonic and step-like nonstationary behaviour in the Lake Winnipeg watershed annual runoff, precipitation, maximum discharge, and runoff ratio records. The observed nonstationary behaviours were attributed to natural climate variability and climate change. Caissie and El-Jabi (1991) applied a stationary stochastic extreme value model (using an exponential distribution) and analyzed daily flood peaks across Canada. The authors described benefits of using a relatively simple exponential distribution and argued that more complex multi-parameter distributions may fit better but didn't guarantee a better representation of flood phenomena. The authors noted difficulties in applying the stochastic model consistently across Canada due to different flood driving mechanisms, climatic and physiographic characteristics. However, a majority of Alberta, Saskatchewan, Manitoba and northwestern Ontario were grouped into three regions.

Several studies have addressed non-stationarity and have explored methods for incorporating climate change through statistical approaches. For example, climate models, hydrological modeling and flood frequency analysis have been combined to assess changes to future extreme floods (Chernet, Alfredsen and Midttømme, 2014, in Norway; Camici *et al.*, 2014, in Italy). Salas and Obeysekera (2013) presented a different approach for incorporating non-stationarity into hydrologic frequency analysis where time variant parameters were used in

statistical distributions. Demonstrating limitations in assuming a stationary precipitation record, Kuo and Gan (2015) applied regional climate models for computation of statistical properties in various time periods, and found an increased risk of future extreme rainfall in central Alberta, Canada.

In a comprehensive review of floods in future climates, Whitfield (2012) provided insight into complexities surrounding the study of present floods and projection of future floods. Whitfield (2012) recognized the greater atmospheric water holding capacity in a warmer climate but balanced this view with inconsistent evidence that historical floods have increased in some regions and decreased in others. This behaviour was similar to global future water cycle studies, which projected wet areas to get wetter and dry areas to become drier (e.g., Trenberth *et al.*, 2014; Collins *et al.*, 2013); but it is important to also consider how driving flood mechanisms might change in the future. While recognizing that GCMs did not reproduce extreme events very well, it was proposed that multiple GCMs and hydrological models are required to properly address uncertainty. Whitfield (2012) also recognized that longer records, such as those available from proxy approaches, can provide interesting context in comparison of observed floods to long term climate variability and assessment of how extreme floods behave in warmer or cooler climates. In a more recent study, Burn and Whitfield (2015) addressed trends in Canadian rivers with different flood mechanisms. A majority of Canada, including the NCRB was characterized by snowmelt dominated floods (nival), which could decrease due to reduced snowpack in a warmer climate. However, flood risk for rivers with rainfall dominated flood regimes (pluvial) could increase due to increased extreme precipitation events.

### 2.1.2 Proxy Approaches

Proxy approaches are popular for their application in assessing long term meteorological and hydrological time series. Proxy approaches rely on relationships between a variable of interest (e.g., climatic variables such as temperature or hydrologic variables such as streamflow) and a measurable proxy (e.g., tree ring width, or physical and chemical composition of ice cores and lake sediments). Relationships and proxy records are then combined to produce reconstructed time series of the variable of interest. Studies relying on proxy data provide extended records into the past and are often given the paleo- prefix (e.g., paleoclimate or paleohydrology) denoting the ancient nature of the data or study. Since the spatial distribution of suitable proxy data can vary, most studies focussed on one proxy record, one variable of interest and one region. Compared to stochastic approaches, proxy approaches can produce time series that correspond to a specific point in time and can be useful in assessing how statistical properties have changed with time (non-stationarity). These properties make proxy approaches valuable in understanding stationarity and examining how small changes in mean conditions can result in larger changes to extremes (Milly *et al.*, 2008).

Within the NCRB, several tree-ring based studies exist that examined past meteorological and hydrological conditions on the Canadian Prairies. Sauchyn, Vanstone and Perez-Valdivia (2011) used tree rings in the upper Saskatchewan River basin to reconstruct mean streamflow from 1063-2007 and tested the hypothesis that observed record does not capture the entire range of experienced variability. The authors specifically noted that the wettest reconstructed period occurred in the late 19<sup>th</sup> century to the early 20<sup>th</sup> century but periods of multi-decadal droughts occurred prior to settlement and included droughts lasting 30 years up to nearly a century of below mean water conditions in the 14<sup>th</sup> century. Of particular interest, the authors correlated dry

and wet periods to large scale climate patterns such as the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO).

St. George and Nielson (2002) correlated tree ring records to summer precipitation anomalies in the Red River Basin since 1409. It was found that pre-settlement precipitation variability exceeded what is recorded in observations. Consequently, the authors cautioned that regional drought and flood planning based only on the observed record may not capture the potential range that is possible due to natural variability. A similar study by St. George (2007) correlated tree ring records to various streamflow and precipitation records in the Winnipeg River Basin. Weak correlations were found between tree rings and individual streamflow or climate variables but high growth was generally correlated with cool, wet summers and low growth was correlated with warm, dry summers. Overall, there was no clear evidence that the past variability in summer precipitation was more extreme prior to the observed record. St. George (2007) also noted that there is considerable spatial heterogeneity and as a result, it may not be appropriate to characterize large regions such as the Canadian Prairies as being either wet or dry.

Beyond the NCRB, Woodhouse *et al.* (2010) used tree rings to reconstruct 1,200 years of streamflow in the southwestern United States. The authors showed that the recent observed drought was exceeded in the past. In the Peace-Athabasca Delta (within the Mackenzie River Basin), Wolfe *et al.*, 2006 used lake sediment cores to reconstruct flood frequency for the past 300 years. The authors presented evidence of multi-decadal intervals between major floods and corroborated findings with anecdotal evidence such as Hudson Bay Company archives. Burn, Wychreschuk and Bonin (2004) combined 130 years of tree ring reconstructed annual streamflow records with observed records to better inform stochastic generation of streamflow in

the Athabasca River. The results indicated that estimation of drought quantiles could be improved by incorporating paleo data, helping to narrow the confidence intervals compared to using only the observed record. Brekke and Prairie (2009), Henley *et al.* (2011), Bonsal *et al.* (2013) and Ault *et al.* (2014) have also studied the incorporation of paleo data by blending it with observed records or future climate projections to obtain a more comprehensive picture of past, present and future water supply. Bonsal *et al.* (2013) considered multi-year meteorological droughts over the period of 1365-2100 and showed that longer droughts occurred in paleo records compared to observed records and that depending on which drought parameter was used (Simple Precipitation Index, SPI; or Palmer Drought Severity Index, PDSI), future droughts could become more or less frequent and last for longer or shorter durations. The authors highlighted the importance of combining precipitation and evaporation effects, as done in the PDSI method that indicated longer droughts that occurred more frequently into the future. Sheffield, Wood and Roderick (2012) and Trenberth *et al.* (2014) argued that temperature-based evaporative modeling (such as PDSI using the Thornwaite algorithm) can overestimate historic and future projections of global drought and suggested that a more physically based algorithm such as the Penman-Monteith equation is appropriate. Sheffield, Wood and Roderick (2012) also cautioned that paleoclimate drought reconstructions using tree rings could be susceptible to the Thornwaite algorithm's sensitivity to temperature changes and other environmental factors such as increased growth due to atmospheric carbon dioxide (CO<sub>2</sub>) concentrations. These special circumstances require further consideration when interpreting results from proxy approaches. Appropriate interpretation of proxy data also requires a proper understanding of the relationships between proxy variables and the climate.

## **2.2 Climate Modeling**

Numerical modeling of the earth's climate is a broad area of research, bringing together many different disciplines and areas of expertise (Trenberth, 1997). The Intergovernmental Panel on Climate Change (IPCC) was formed by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) in 1988 and acts as a scientific body for evaluating climate change research, providing guidance and producing assessment reports which summarize the current state of knowledge. The IPCC also plays a key role in preparing atmospheric forcing scenarios for use in the Coupled Model Intercomparison Projects led by the World Climate Research Programme's Working Group on Climate Modelling. Coupled models are a generation of climate models that specifically recognize the interactions between parts of the climate system. At a high level, the climate system can be defined as the atmosphere, hydrosphere, cryosphere, land surface, biosphere, and geosphere and the interactions between these various components (McGuffie and Henderson-Sellers, 2005). The realm of climate modeling can include very specific topics but for many researchers, climate models are tools and their output is used for impact assessments. This section covers some climate modeling basics, some specifics relating to climate model's representation of hydrology and the application of climate models in hydrological impact assessments.

### **2.2.1 Developing Future Climate Scenarios**

The IPCC and other literature provide good practise guidance for using climate models to develop future climate scenarios. Although the development of future climate scenarios for specific projects may require special consideration due to unique circumstances, there are several common considerations relevant to most studies. For example, several temporal domains are typically required to define baseline and future time periods. Other examples include the finite

number of future atmospheric forcing scenarios (sometimes referred to as greenhouse gas scenarios) available to drive climate models, the number of climate models with available data and the sources of uncertainty to consider when developing and assessing future climate scenarios. The sections below introduce some of the common considerations and include further detail as it pertains to the future climate scenarios used in this thesis

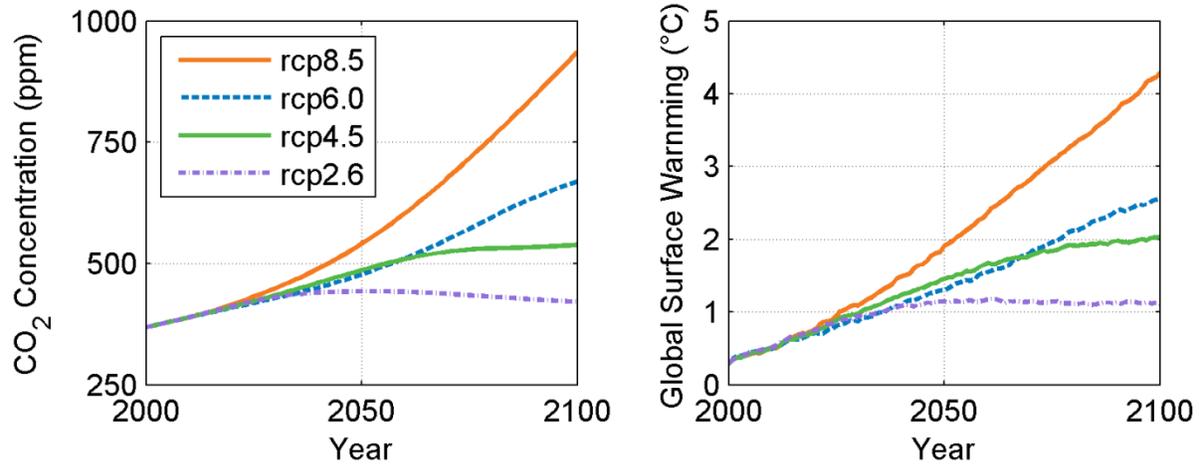
Selection of the temporal domain is an important early step in developing future climate scenarios. At the basic level, one fixed future period and one fixed baseline period is defined. In some instances, multiple baseline and future periods or continuous time spans are considered. The baseline period, sometimes referred to as the reference period, is generally selected for a historic period such that observations are available. A baseline period should be selected such that it covers a period of sufficient duration, captures a large portion of natural variability and includes various anomalies such as droughts and floods. Whereas in many cases, a typical 30 year period such as that used by the WMO might be sufficient, consideration should be given to incorporate longer periods in order to capture multidecadal variability. In many cases, such as those focusing on extreme events, a baseline period longer than 30 years is warranted (IPCC-TGICA, 2007). Similarly, Kundzewicz and Robson (2004) recommended the use of at least 50 years of record for climate change detection to avoid apparent trends and obscure climate change signals produced by natural climate variability. In the literature, future periods were typically of the same duration as the baseline periods. In the case of fixed 30 year domains future periods were often defined as the 2020s (2010-2039), 2050s (2040-2069) and 2080s (2070-2099).

One of the major inputs to climate modeling, and an important consideration for developing future climate scenarios is the atmospheric forcing used to drive the model. Atmospheric forcing scenarios prescribe the changing conditions that the climate responds to. In

the past, atmospheric forcing was prescribed by the IPCC's Special Report on Emission Scenarios (SRES; Nakicenovic *et al.*, 2000). The latest climate models in CMIP5 use atmospheric forcing prescribed by Representative Concentration Pathways (RCPs). RCPs are used to prescribe the levels of various anthropogenic induced forcing agents (e.g., GHGs and aerosols) in the atmosphere. RCPs include assumptions about societal evolution and represent different demographic, social, economic, regulatory, technological, and environmental developments. RCPs are used in GCMs to simulate the evolution of climate over time in response to changes in atmospheric forcing agents and can be useful in exploration of uncertainty due to future GHG scenarios. RCPs were developed by the research community, independent of the IPCC and were named according to their radiative forcing in 2100. An overview of the four RCPs can be found in van Vuuren *et al.* (2011) and are summarized in Table 1. Figure 1 illustrates RCP projections of CO<sub>2</sub> concentrations and modeled global surface warming. Sanford *et al.* (2014) showed that emission rates are currently tracking just above RCP8.5.

**Table 1 - Description of Representative Concentration Pathways (van Vuuren *et al.*, 2011)**

<b>RCP</b>	<b>Description</b>	<b>CO<sub>2</sub> equivalent</b>
RCP8.5	Rising radiative forcing pathway leading to 8.5 W/m <sup>2</sup> by 2100	~1370 ppm
RCP6.0	Stabilization without overshoot pathway to 6 W/m <sup>2</sup> at stabilization after 2100	~850 ppm
RCP4.5	Stabilization without overshoot pathway to 4.5W/m <sup>2</sup> at stabilization after 2100	~650 ppm
RCP2.6	Peak radiative forcing at ~3W/m <sup>2</sup> before 2100 and then a decline to 2.6 W/m <sup>2</sup> by 2100	~490 ppm



**Figure 1 - RCP CO<sub>2</sub> concentrations (left) and projected global surface warming (right). Data extracted on July 8, 2013 from KNMI Climate Explorer ([climexp.knmi.nl](http://climexp.knmi.nl); van Oldenborgh and Burgers, 2005).**

In ideal situations, modelling centers and researchers conducting impact assessments would use all available atmospheric forcing scenarios. However, due to resource limitations, it is often not possible to assess all RCP scenarios and researchers must make decisions on which RCP(s) to consider. At a basic level, one RCP can be used. However, this decision hinders a complete assessment of uncertainty since future greenhouse gas and aerosol emissions are a primary source of uncertainty (IPCC-TGICA, 2007). The North American Regional Climate Change Assessment Program (NARCCAP) used only one emission scenario however, the program was designed to primarily focus on climate model uncertainties (Mearns *et al.*, 2012). If a considerable amount of certainty in a single RCP projection existed, this might provide some justification for only using one RCP scenario. In most cases, researchers consider two or more RCPs to capture a range of potential future outcomes.

The end use of the RCP is to drive a climate model which simulates the temporal evolution of climate parameters as a result of the specified atmospheric forcing. From these climate models, variables of interest such as temperature and precipitation can be examined. Not all climate models are equal as they can use various physical schemes, parameterizations,

resolutions and numerical solution methods. In some cases coupled climate models might share an ocean model but use different atmospheric models. Furthermore, not all climate models contain output for all RCPs and for all variables at desired temporal resolutions which could limit the number of models available for assessment. A more detailed description of climate models is provided in Section 2.2.2, but at this point it is important to recognize that climate models are the primary source of climate change data for impact assessment studies. Climate model data generally requires post processing to be useful in impact assessments and as such, further work is required to develop future climate scenarios from climate model data.

Several methods and guidelines exist for developing future climate scenarios that can help guide researchers in selecting RCPs, climate models and post-processing methods. The IPCC and other researchers have published guidance on: assessing and combining multi-model projections (Knutti *et al.*, 2010a; Knutti *et al.*, 2010b; Pierce *et al.*, 2009), treatment and communication of uncertainties (Mastrandrea *et al.*, 2010) and the use of scenario data for impact assessments (IPCC-TGICA, 2007; Mote *et al.*, 2011).

Evaluation of climate model performance is an important consideration for impact assessments. Ideally, climate models would produce perfect representations of observed climate, which may increase the confidence in the climate model's ability to project future climate. However, for many reasons (e.g., coarse resolution, parameterizations, model drift, natural climate variability, theoretical understanding of the climate) today's climate models still contain bias (Reichler and Kim, 2008). Bias correction of climate model data is a form of post-processing and is discussed later in this chapter in Section 2.2.2.3.

Researchers have found that combining multiple models often leads to improved baseline performance. Considerable effort has gone into assessing climate model performance in attempts

to find a preferred subset of climate models from the entire suite (ensemble) of available models. Knutti *et al.* (2010b) and Pierce *et al.* (2009) showed that performance of the ensemble continued to improve for up to five models after which the performance gains become less. The authors also suggested that the mean of a few good performing models may outperform the entire multi model ensemble which may include models with poor performance. However, the climate state derived by averaging models may not be physically realistic and presentation of a multi-model mean may be misleading without further discussion. As an alternative to taking the mean of a subset of good performing models, some researchers have assigned weights based on performance. However, care should be given such that multiple simulations from the same model are not artificially assigned greater weights (Mote *et al.*, 2011).

Although no clear directive exists on a preferred subset of climate models to select, the following general guidance is useful in preparing future climate scenarios: Researchers may select a subset of models for a particular analysis but should document the reasons why (Knutti *et al.*, 2010a). It has also been suggested that researchers may choose to discard a particular climate model due to poor regional performance (Knutti *et al.*, 2010a) or due to lack of available data (as done in Reichler and Kim, 2008). Though some authors have suggested using a multi-model mean as a best estimate (Reichler and Kim, 2008), other authors have suggested presenting all model results without combination as a useful choice that benefits adaptation planning (Knutti *et al.*, 2010b). In addition to considering multiple models, Mote *et al.* (2011) emphasised the recognition and understanding of various uncertainty sources including GCM selection, future greenhouse gas emission scenarios and natural climate variability. It is also important to consider the regional and time varying importance of each of these uncertainty sources (Hawkins and Sutton, 2011).

Similar to how good performance may increase confidence in a climate model's ability to simulate the future; agreement among multiple climate models' future projected changes may also increase confidence in a particular result. This multi model agreement provides greater evidence regarding the direction or magnitude of change into the future. In some cases with strong agreement on direction and magnitude of change, such as projected temperature increases, researchers might assign confidence intervals to the magnitude of the projection. In other cases with strong agreement on direction but less agreement on magnitude, researchers might assign confidence to the direction of change (e.g., increasing future precipitation). It is important to note, however, that while model agreement can increase confidence in a particular result, it does not necessarily infer greater likelihood (Knutti *et al.*, 2010a). Confidence can be expressed as a qualitative measure (e.g., low, medium, high) as a function of model agreement and the amount of evidence. Mastrandrea *et al.* (2010) provides criteria where likelihood may be assigned. One such criteria states that when *“a likelihood or probability can be determined for a variable, for the occurrence of an event, or for a range of outcomes (e.g., based on multiple observations, multiple ensemble runs, or expert judgment): Assign a likelihood for the event or outcomes, for which confidence should be ‘high’ or ‘very high’.*” The use of the IPCC's likelihood scale must be adapted for individual studies but provides a good means for communicating results. The scale also provides a good tool and language for treatment of uncertainty as is important in all climate change studies. Olsen *et al.* (2015) reiterated the IPCC's emphasis on the importance of communicating the difference between confidence and likelihood, and noted that likelihood should only be assigned when confidence in a projected change was high.

Since the confident application of future climate scenarios intrinsically relies on climate models that perform reasonably well, it is important to consider details on various models and

evaluation techniques. Furthermore, it is important to consider techniques to correct bias identified in the climate model performance evaluation.

## **2.2.2 Climate Models, Evaluation and Bias Correction**

The overall development and assessment of future climate scenarios should follow the general best practise guidance described in the previous section. In addition, climate scenario development is limited by available climate models, their output, and should follow established methods for evaluation and post processing techniques. A brief introduction to climate models and climate model evaluation is provided below along with specific examples of established bias correction methods.

### **2.2.2.1 Climate Model Types and Components**

There are several types of climate models with varying degrees of complexity. Climate models have evolved with time from coarse models of the atmosphere only to finer resolution coupled models such as Atmosphere-Ocean General Circulation Models (AOGCMs) that simulate the atmosphere and ocean and can also include the land surface, ice processes and other earth components. Earth Systems Models (ESMs) are the most recent type of climate model which include AOGCM components as well as biogeochemical processes such as the carbon cycle (Flato *et al.*, 2013). Despite the intricate differences among the various types of climate models, they can be generally classified into one of two types depending on their spatial domain. Global Climate Models (GCMs) refer to climate models that simulate the entire earth whereas Regional Climate Models (RCMs) refer to climate models that simulate a limited area of the earth.

RCMs are occasionally referred to as limited area models or nested models and require boundary layer conditions which are typically provided by GCMs. The process of driving an RCM with GCM boundary conditions is also known as dynamical downscaling, where

downscaling refers to the process of taking coarse resolution data and bringing it down to a finer resolution. RCMs are developed using finer spatial resolutions compared to GCMs (e.g., 50 km in Mearns *et al.*, 2012) which allows for greater topographical definition. In certain areas, such as the head waters of the Columbia River Basin, topographical definition was found to be important and influenced the results of hydrologic climate change impact assessments (Gao *et al.*, 2011). However, in areas with low topographic relief, such as the Canadian prairies, or for longer time-averaged periods (e.g., monthly), the added value from RCMs was less apparent (Di Luca, de Elía and Laprise R, 2012).

In general, GCMs and RCMs contain at least an atmospheric model and a Land Surface Model (LSM) which is often referred to as a land surface scheme. A simple schematic showing the interaction between the atmosphere and land surface can be found in Verseghy (2000). Modern GCMs also contain an ocean model but many RCM experiments (e.g., Mearns *et al.*, 2012) do not simulate the ocean, and instead require an ocean boundary layer to be prescribed. For many climate impact assessments, model outputs from the atmosphere and land surface are of particular interest. Variables of interest from the atmospheric model include surface air temperature, precipitation, humidity, surface pressure, zonal and meridional wind speed, short wave radiation and precipitable water (in an atmospheric column). Atmospheric variables provide some of the inputs required by LSMs which then compute the water and energy balances and return other variables including upward moisture flux (evapotranspiration), runoff, snow water equivalent, ground temperature, and soil moisture. Typically, LSMs also require information about vegetation and soil properties and can include a river and lake routing component. Routing is important to delay the delivery of runoff into the oceans that can affect

salinity and circulation (Arora and Boer, 1999). However, this routing may be very conceptual and routed runoff is not a standard climate model output (Taylor, 2013).

Differentiating between LSMs and traditional hydrological models is important for understanding differences in model behaviour. Haddeland *et al.* (2011) defined LSMs as models using energy balance approaches to solve vertical exchanges of heat and water at small time steps (e.g., hourly). LSMs may utilize information on rainfall, snowfall, temperature, specific humidity, surface pressure, radiative fluxes, and wind speed. LSMs are well suited for coupling with atmospheric models since most of the required data is readily available. Alternatively, Haddeland *et al.* (2011) suggested that hydrological models are more focussed on solving lateral water exchanges and often use approximations to solve the energy balance. Since hydrological models generally operate on longer time steps (e.g., daily) and rely on fewer variables (some only require temperature and precipitation), hydrological models are common in practical applications with limited computational resources and available data. These definitions for LSMs and hydrological models are fairly simple and it is important to recognize that complexity varies for both LSMs and hydrological models. Several researchers have characterized the additional uncertainty introduced by the LSM or hydrological model in climate impact assessments (Chen *et al.*, 2011; Prudhomme *et al.*, 2014; Hagemann *et al.*, 2013; Eum, Dibike and Prowse, 2014) but recognized that this uncertainty was often masked by larger uncertainties such as GCM selection. Furthermore, although LSMs and hydrological models often include a river and lake routing routine, it is common for the routing model to be separate from the LSM or hydrological model.

Another important feature of climate models is their ability to simulate the chaotic nature of the climate. This chaos is inherent in the real climate system (Trenberth, 1997) and plays a

role in natural climate variability (internal variability). Other causes of natural climate variability include ocean circulation patterns (e.g., ENSO, PDO) and changes to the earth's radiative forcing (e.g., due to the sun's solar activity or volcanic eruptions). Most GCM experiments do not include natural variability due to changes in radiative forcing and natural variability due to ocean circulation is internally modeled (not forced). However, many GCM experiments are designed to explore natural climate variability due to unknowns in initial conditions. These experiments use identical GCM configurations and forcing but initialize the model with slightly different initial conditions. Due to the chaotic nature of the climate, small perturbations in the initial conditions can result in different climate evolutions. GCM simulations from these experiments are referred to as runs or members (e.g., run1, run2, run3). Deser *et al.* (2012a) used a 40 run ensemble and demonstrated that natural climate variability affected several decades of future regional temperature projections. The authors also noted that precipitation was subject to even greater natural variability and advances in climate modeling were not expected to overcome uncertainties due to the chaotic nature of the climate. Although natural climate variability has been shown to contribute large uncertainty to shorter term climate projections (less than 50 years), at longer time horizons, other sources of uncertainty such as emission scenario and GCM selection play a more significant role (Hawkins and Sutton 2011).

#### **2.2.2.2 Evaluation of Climate Model Skill**

There are many ways to evaluate climate model skill and many climate change studies include some form of model evaluation. One common skill assessment approach is to calculate the error (bias) in simulating climatology (mean conditions) over an observed reference period. Comparison of climate statistics such as mean conditions or quantiles is preferred due to the chaotic nature and natural variability simulated by climate models (e.g., Gao *et al.*, 2011). Many

different performance metrics and error terms have been used. Temperature and precipitation on various temporal and spatial scales were common performance metrics but some researchers have looked at specific variables at specific locations over specific periods of time. Statistical measures have become more common in recent model evaluation studies (Flato *et al.*, 2013) but it is important to note that currently, no all-purpose metric has been found that identifies the best model (Knutti *et al.*, 2010a; Reichler and Kim, 2008) and a climate model that performs well with one metric may perform poorly in other metrics. Furthermore, a climate model that has good skill in simulating observed climatological conditions does not ensure it is robust for projecting future climate, and for certain variables, observational uncertainties in the reference climate may be quite large. Regardless of issues with quantifying climate model performance and searching for a best model to project the future, it is important to consider a large set of models for analysis, and only disregard a model if its performance is very poor with respect to a relevant metric.

Gleckler, Taylor and Doutriaux (2008) ranked CMIP3 GCMs based on spatial mean monthly climatologies for 22 metrics including temperature, precipitation, wind, specific humidity and precipitable water content. Root Mean Square Error (RMSE) was the statistical measure used to calculate the error. In this study, RMSE was averaged spatially (all grid points for a given region) and temporally for the annual cycle (monthly means) over the period of 1980 to 1999. Although other statistical measures exist including the climate prediction index (Reichler and Kim, 2008), the probability plot correlation coefficient (Hirabayashi *et al.*, 2008) or several other measures presented in Schoetter *et al.* (2012) the space-time RMSE was presented in Gleckler *et al.* (2008) as a comprehensive statistical measure that provides an overall depiction of GCM performance. Gleckler *et al.* (2008) also ranked GCMs using relative

RMSE, which compared how individual GCMs perform with respect to the GCM with the median performance. For individual performance metrics, relative RMSE provided information on the best, median and worst performing models, relative to the entire GCM ensemble. Results were presented using “portrait” diagrams that summarize and help to visualize individual model performance by metric, using colours to indicate relative RMSE. Gleckler *et al.* (2008) found that averaging results among multiple GCMs often performed better than any individual GCM.

Sillmann *et al.* (2013) evaluated GCM performance for 27 extreme indices based on temperature and precipitation data. Following similar methods in Gleckler *et al.* (2008), RMSE and relative RMSE were used to evaluate 31 CMIP5 models for the period of 1981 to 2000. RMSE was, however, averaged spatially on annual metrics instead of monthly. Sillmann *et al.* (2013) also used portrait diagrams, found that the GCM median climatology outperformed individual GCMs and noted improvement in the ability of CMIP5 GCMs to capture extremes in comparison to CMIP3 GCMs. The authors highlighted the importance of considering observational uncertainties, noting large differences in observed (reanalysis) datasets.

Sheffield *et al.* (2013) evaluated 17 CMIP5 models over North America for the period of 1979 to 2005 on multiple metrics including runoff using offline LSMs. In some regions, runoff was underestimated and peaked earlier in the spring while in dry regions and high latitudes runoff was overestimated. However, the authors noted that spatial variability in total annual runoff was generally replicated. Over the NCRB, the GCM mean performed quite well with respect to mean annual runoff and runoff to precipitation ratio when compared to the average of two LSMs. It is important to note that this assessment relied on the accuracy of other models, the LSMs, and their forcing data that can introduce additional uncertainty. The frequency of occurrence of meteorological and agricultural droughts and excess moisture events were also

assessed spatially. Many models reasonably captured spatial patterns but underestimated the frequency of occurrence of drought.

Other studies have evaluated climate model skill using different metrics, statistical analyses to quantify error, and spatial domains. Watterson, Bathols and Heady (2014) found that CMIP5 outperformed CMIP3 GCMs and suggested that the improvement is primarily a function of model formulation upgrades and increased resolution and not necessarily the inclusion of Earth system components. Reichler and Kim (2008) analyzed spatially averaged annual metrics and also found increased performance in CMIP3 when compared against previous model generations. Improvements were attributed primarily to finer resolutions and more realistic parameterizations. Recognizing that sampling uncertainty combined with natural climate variability can affect performance and detect-ability of climate change, there is ongoing work to incorporate measures of natural climate variability into climate change assessments (Deser *et al.*, 2012b). Ault *et al.* (2012) assessed CMIP5 simulations of precipitation at the decadal and multi-decadal scale over the 1850-2005 period. Overall the CMIP5 ensemble mean underestimated decadal variability in several key regions which could have an important influence on the assessment of drought risk. The key regions include northern Africa, Australia, Western North America and the Amazon but the authors' concerns were not as pronounced for the NCRB.

Sushama *et al.* (2006) and Poitras *et al.* (2011) included an evaluation of routed Canadian Regional Climate Model (CRCM) runoff in the NCRB. CRCM was shown to overestimate streamflow volume and overestimate the monthly cycle of streamflow in the Nelson and in the Churchill rivers. This outcome is partially explained by the studies not considering diversion flows and other forms of regulation, as well as concerns with basin delineation (i.e., the Nelson River delineation included the neighbouring Hayes River Basin).

Overall, climate model data users should not expect perfect performance from climate models but should recognize that evaluation is important for appropriate interpretation of results. Climate model errors may result from several sources including shortcomings of the climate models or even perceived errors due to uncertainty in observational data. Once errors are better understood, users can account for them through bias correction to increase the utility of climate model data in impact studies.

### **2.2.2.3 Bias Correction of Climate Model Data**

Similar to the many ways of evaluating climate model skill, there are many ways to correct the bias present in the climate models. Bias correction techniques range from simple methods that only account for changes in monthly mean climate to more complex methods that correct for the daily statistical distributions. For climate models, bias correction has traditionally been limited to the correction of temperature and precipitation fields where observational data is more reliable over large spaces for longer periods of time. Some studies (Shrestha *et al.*, 2012; Wilke, Mendlik and Gobiet, 2013; Cheng *et al.*, 2014), however, have bias corrected other variables such as solar radiation, relative humidity, wind speed and surface air pressure when observational data is available. Other studies such as (Hagemann *et al.*, 2013) chose only to bias correct some GCM variables and used other variables directly. Hashino, Bradley and Schwartz (2007) applied bias correction techniques to streamflow data in shorter-term forecasting applications to improve forecast skill.

Studies such as Mpelasoka and Chiew (2009), Watanabe *et al.* (2012), Chen *et al.* (2013) and Rätty, Räisänen and Ylhäisi (2014) have summarized and compared common techniques for post-treating climate model data. Generally, there are two categories of techniques. Perturbation

methods (e.g., the delta method) scale the observed climate time series based on changes between climate modeled future and reference simulations:

$$Future\ Scenario = Observed + (GCM_{Future} - GCM_{Reference}) \quad (1)$$

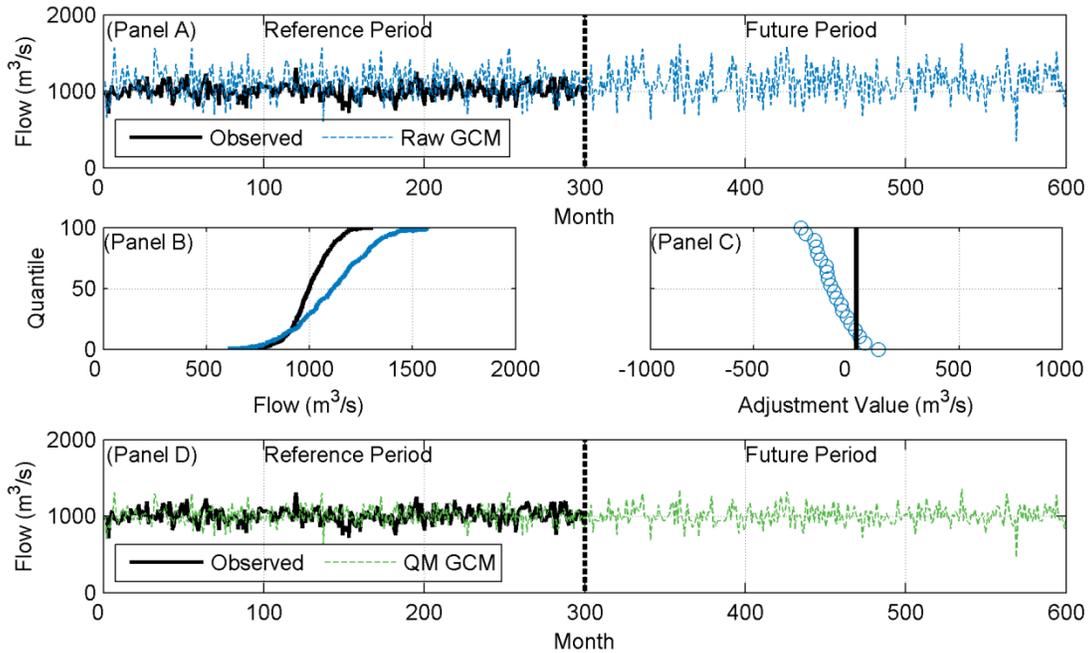
The other category is referred to as bias correction methods which correct the climate model's generated time series based on bias in a reference period (e.g., daily translation):

$$Future\ Scenario = GCM_{Future} + (Observed - GCM_{Reference}) \quad (2)$$

Each technique also varies in its complexity ranging from those that correct the mean climate to those that correct the entire statistical distribution. Although perturbation techniques offer more realistic sequencing of events and can better match observed climate variability, since they are based on observations, it is often of interest to look at bias correction techniques to study climate change impacts that depend on potential changes to sequencing of events, such as long term droughts. Shrestha *et al.* (2012) suggested that since bias correction methods allow for changes in future variability, these methods are more appropriate for climate impact assessments.

Of the available bias correction techniques, distribution-based methods such as quantile mapping (QM) of empirical distribution functions performed better than mean-based methods as they correct the entire distribution instead of just changes in the mean (Chen *et al.*, 2013). Various types of QM techniques also exist including techniques that depend on an assumption of the statistical distribution and techniques that also incorporate correction of wet and dry days (for precipitation). Adjustment values computed in QM can be applied for various numbers of quantiles and either as additive values (i.e., for variables that can be negative such as temperature) or as multiplicative factors (i.e., for variables that must be greater than zero such as precipitation). Using fictitious data, Figure 2 illustrates how QM can be applied using empirical distribution functions for 20 quantiles in an additive fashion. Panel A shows the observed and

raw GCM time series. Panel B compares the empirical distribution functions for observed and GCM in the reference period. Panel C shows the adjustment values computed at 20 quantiles between the Panel B distribution functions. Panel D shows the observed and quantile mapped GCM time series which has had the adjustment values applied.



**Figure 2 - Illustrative quantile mapping example for correcting monthly flow data in a reference period and future period.**

Despite QM being a commonly preferred approach compared to alternatives, there does not appear to be one method that universally performs well (Räty *et al.*, 2014) and there are some notable issues: QM assumes that bias is constant in the reference period and future periods. For empirical distributions, the bin size for adjustment value computations should be sized appropriate to available data. For example, 20 years of data may not contain enough information to define adjustment values for 100 quantiles (Chen *et al.*, 2013). Separate QM of individual variables can lead to physical inconsistencies, for example when minimum daily temperature and maximum daily temperature are corrected individually (Thrasher *et al.*, 2012). QM retains the GCM's strengths and weaknesses such as the ability to accurately simulate sequencing of events

which may underestimate or overestimate observed climate variability (Wilke *et al.*, 2013). When QM using empirical distributions does not capture the range of extremes in the future, assumptions are required such as holding adjustment values constant for extreme minimum and maximum quantiles (Wilke *et al.*, 2013; Chen *et al.*, 2013). It should also be recognized that bias correction can add uncertainty to future climate assessments, however, uncertainty in bias correction techniques is typically small compared to other sources of uncertainty such as GCM selection (Mpelasoka and Chiew, 2009; Chen *et al.*, 2011).

Although many researchers have either separately applied raw GCM runoff data or bias corrected GCM atmospheric data, the process of bias correcting gridded runoff fields from climate models has not received a lot of attention. Several reasons likely contribute to this gap in the literature. For one, gridded GCM runoff is generally coarse and may not be suitable for direct use in smaller watersheds. Secondly, the largest effort has gone into analyzing, understanding and applying atmospheric fields (i.e., temperature and precipitation) from climate models and there is greater support for bias correcting these variables. Thirdly, hydrologists may be more inclined to use a hydrological model that has been set up for a specific purpose, in a specific region, calibrated to observations and proven to perform reasonably well. Finally, there are no long term, reliable, gridded observed runoff datasets at fine temporal resolutions. Gridded runoff products are either a function of observed streamflow coupled with a water balance model such as the World Meteorological Associations' (WMO) monthly gridded Global Runoff Data Center (GRDC; Fekete, Vörösmarty and Grabs, 2000) or based on output from an LSM driven by observed precipitation and temperature. González-Zeas *et al.* (2012) bias corrected RCM runoff data via a simple annual correction factor derived from the GRDC dataset. The authors showed

considerable improvement in RCM runoff performance, however, the use of annual correction factors were noted as a shortcoming since RCM runoff seasonality was not corrected.

Many gridded runoff products from LSMs exist, however there are documented performance issues with these products (Reichle *et al.*, 2011; Schwalm *et al.*, 2014) and using their values as observations for bias correction is not be an ideal solution. Koshida *et al.*, 2015 recognized that gridded runoff products, such as those developed from global hydrological models or LSMs typically overestimated global terrestrial runoff and disagreed on runoff fractions. Errors in the forcing precipitation data, evapotranspiration schemes (Zhou, *et al.*, 2012) and representation of snow water equivalent (Pietroniro *et al.*, 2007) are among the major sources for the LSM bias. Zhou *et al.* (2012) also showed that 14 LSMs overestimated mean annual runoff in the St. Lawrence, Mississippi and Nelson River Basin for the 1986-1995 period. In a case study on hydrological model uncertainty in the Athabasca watershed, Canada, Eum *et al.* (2014) showed that most locations were more sensitive to precipitation forcing in comparison to soil and routing parameters. However, Vaze *et al.* (2010) showed that the hydrologic state (wet or dry) and duration of time (10 to 40 years) used to calibrate hydrologic models can impact their performance in a climate of the opposite hydrologic state. Haddeland *et al.*, (2011) found mixed results where some LSMs overestimated runoff and other LSMs underestimated runoff in Arctic basins (Mackenzie River and Lena River). However, the authors noted that the models generally overestimated runoff in arid and semiarid basins. Some of the identified causes of the differences include how models handle snow accumulation/melting and lakes. Haddeland *et al.*, (2011) who also assessed LSM differences in runoff timing, cautions that since the LSMs simulate naturalized conditions, it is not appropriate to compare directly to observed discharge at sub annual time scales.

In a recent conference paper, Duong, Tachikawa and Yorozu (2015) used gridded runoff from an LSM to bias correct GCM runoff for use in a routing model. The authors recognized the absence of observed gridded runoff data and used an LSM to approximate observations. Two Japanese river basins, less than 3,000 km<sup>2</sup>, were simulated at 20 km resolution. To correct bias, QM was applied at each grid for individual months. The authors noted that bias correction improved river discharge simulations but also noted several concerns. One concern was the LSM generated runoff sensitivity to precipitation input data. Two possible precipitation datasets were considered and one was selected based on how well the generated LSM streamflow compared to observed streamflow distributions. However, no emphasis was placed on temporal distribution patterns. The authors also noted that further work is needed to address spatial correlation between neighbouring grid cells for bias correction. Successful future applications of similar works will rely heavily on LSM ability to reproduce realistic gridded runoff (spatially and temporally) and will also require sophisticated bias correction techniques to preserve spatial patterns. Similar obstacles have been addressed in statistical downscaling of precipitation data where advanced methods are now available (e.g., Werner and Cannon, 2015) and potentially applicable for bias correction of gridded runoff in future work. Another potential approach is to use statistical up-scaling of observed streamflow to generate gridded runoff estimates, as done by Gudmundsson and Seneviratne (2015) for the continent of Europe.

Overall, there is consensus in the literature that considerable differences exist between LSM simulated evapotranspiration and runoff, thereby introducing uncertainty into estimates of observed gridded runoff. These uncertainties are an important consideration in future climate impact studies especially when considering methods for bias correcting gridded runoff. While it is important to invest in improving estimates of observed gridded runoff, it is also important to

move forward with evaluating alternative bias correction techniques for GCM runoff for applications in assessment of long term hydrology.

### **2.3 Application of Climate Models in Assessment of Long Term Hydrology**

Climate models are recognized as sophisticated tools used to assess global response to increased greenhouse gases and other atmospheric forcing (Reichler and Kim, 2008; Gao *et al.*, 2011). It can also be argued that due to their physically-based processes, climate models can be useful in understanding long term natural variability. Although it is important to recognize limitations in modeling the climate (e.g., bias in simulating current climate) one should not discount the usefulness of climate models. Instead, as suggested by Beven (2011), decision makers should be precautionary and use available climate science in understanding a range of potential future conditions. Similarly, Trenberth (1997) suggested that despite uncertainties, climate model projections provided information that is preferred over the alternative of declaring ignorance.

There are many ways in which climate models can be used in the assessment of long term hydrology. In some cases, future climate model projections are combined with physical understanding of historic climate processes to aid in interpreting future impacts. Trenberth *et al.* (2014) demonstrated this process by accepting that future temperature will increase but rationalized that it may not translate into increased drought risk, as drought is a complex phenomenon. Instead, the authors cautioned that climate change trends could be masked with natural variability and the understanding of climate change impacts on natural droughts such as those caused by ENSO (a common cause of episodic droughts around the world) contain outstanding issues. The authors concluded that climate change may not initiate future droughts, but could accelerate their onset, severity and duration. Contrary to using climate change projections to physically rationalize future impacts, some studies have relied directly on climate

model output to portray the future climate. These studies involved post processing and assessment of climate model data and are discussed below from a hydrological assessment standpoint.

Hydrological analyses with respect to natural climate variability and climate change varies, depending on the parameter of interest. And often, multiple parameters are considered in the same study. For streamflow, some studies have considered changes in mean streamflow at different temporal scales (e.g., annual, seasonal, monthly, daily), whereas other studies have looked at more specific events such as changes to timing of the spring freshet, changes to the one day high flow event, changes to intra-annual drought events at various time scales and changes to multi-annual drought events. Presentation and analysis of averaged hydrographs (e.g., Huziy *et al.*, 2013; Chen *et al.*, 2011; Nohara, *et al.*, 2006; Sperna Weiland *et al.*, 2012a; Poitras *et al.*, 2011) and continuous time series hydrographs (Woodhouse *et al.*, 2010; Alkama *et al.*, 2013) were common approaches to answer many of the studies` questions. Other studies have used box plots and cumulative distribution functions to present results (Burn *et al.*, 2004; Bonsal *et al.*, 2013; Sadri and Burn, 2014; Burn and DeWit, 1996; Sperna Weiland *et al.*, 2012a; Sushama *et al.*, 2006). Events which span extended lengths of time such as multi-year hydrological droughts have typically been defined using the theory of runs (Yevjevich, 1967), beginning and ending when the streamflow crosses a specified threshold. Thresholds were subjective values but were typically taken as the mean, or a time varying mean such as the monthly mean. In studies where a hydrological model was not used to produce streamflow hydrographs, river and lake routing has been applied.

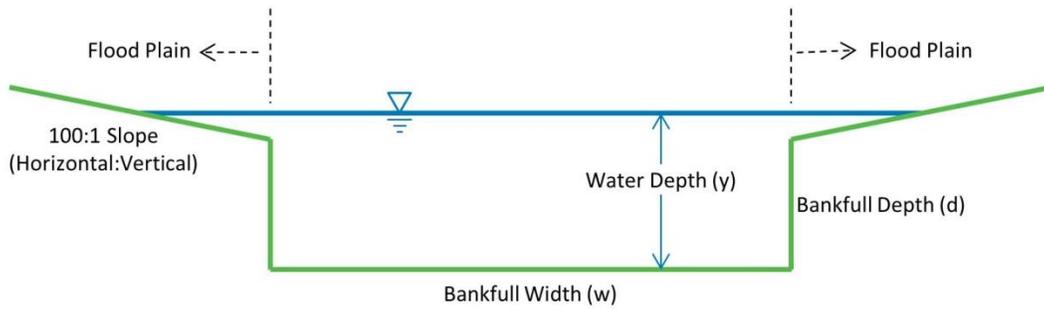
### 2.3.1 River and Lake Routing of Climate Model Runoff Data

River and lake routing is used to produce streamflow from runoff. In some climate models, a basic routing scheme is used to return freshwater generated on land surface to the ocean, and to assist with GCM validation through hydrograph comparison (Arora, Chiew & Grayson, 1999). However, since routed runoff is not a standard climate model output (Taylor, 2013), and GCM's operate on coarse spatial grids, offline routing is typically a post processing exercise. Various routing schemes are available ranging from simpler hydrologic routing methods that use lumped transfer functions (e.g., Dingman, 2008) to more complex hydraulic routing methods that solve the complete Saint-Venant equations to describe shallow water flow in a distributed environment (e.g., Rousseau *et al.*, 2015; Arora *et al.*, 2001).

Using an offline routing model allows for customization to a specific region and is a common method for evaluating climate model runoff. River and lake routing in a gridded domain (i.e., GCM domains) typically requires more detailed physiographic information including elevation, river drainage network, river width and depth, lake storage and outflow relationships and river roughness characteristics. WATROUTE is a popular routing model that has been coupled with both hydrological models (Kouwen, 2012; Pietroniro *et al.*, 2007), GCMs (Arora *et al.*, 2001) and RCMs (discussed below).

WATROUTE uses a storage-routing algorithm and falls somewhere in the middle of routing model complexities. Kouwen (2012) noted that more sophisticated models may not provide increased accuracy, especially in larger basins. Although storage-routing methods are often associated with simplified transfer functions (Li, *et al.*, 2013), WATROUTE assumes steady, uniform flow, a parallel energy grade and channel bottom (friction slope equals channel slope) and falls within the Saint-Venant equation category as a Kinematic Wave routing method.

Compared to solving the complete shallow water equations, Kinematic Wave routing neglects acceleration, and pressure terms in the momentum equation and has limitations such as producing null discharge in completely flat terrain (Rousseau *et al.*, 2014). Assuming a rectangular river cross section area (Figure 3) and prescribing a minimum channel slope, WATROUTE combines the continuity equation (Equation 3) and Manning's equation (Equation 4 and Equation 5) to solve for storage changes and outflow changes in each grid cell. WATROUTE also allows wetland routing (when coupled with the hydrological model WATFLOOD) and lake routing using power functions or polynomial functions as relationships between lake storage and outflow.



**Figure 3 - WATROUTE channel cross section**

$$\frac{Inflow_1 + Inflow_2}{2} - \frac{Outflow_1 + Outflow_2}{2} = \frac{Storage_2 - Storage_1}{\Delta t} \quad (3)$$

where subscripts ( $_1$  and  $_2$ ) denote the beginning and end of a time step.

$$Q = \frac{1}{n} AR^{\frac{2}{3}} S^{\frac{1}{2}} \quad (4)$$

where  $Q$  denotes streamflow,  $A$  denotes area,  $R$  denotes hydraulic radius,  $S$  denotes channel slope and  $n$  denotes Manning's roughness coefficient.

Assuming that  $A$  equals river width ( $w$ ) multiplied by water depth, and the channel is wide such that  $R$  is approximately equal to  $A/w$ , Equation 4 is simplified to Equation 5.

$$Q = \frac{1}{n} \frac{1}{w^{0.667}} A^{1.667} S^{0.5} \quad (5)$$

WATROUTE can be run in multiple configurations depending on available runoff data from the hydrological model or land surface scheme. The minimum requirement is gridded total runoff but it can also incorporate gridded recharge (lower zone flow) and gridded leakage (potentially from a groundwater model). Using the minimum requirement routes total runoff as surface flow directly into the channel network and is common when only total runoff is available (Kouwen, 2012; Arora *et al.*, 2001).

Other routing models have been used to route climate model runoff. For example, Falloon *et al.*, 2011 implemented the Total Runoff Integrated Pathways (TRIP) model to validate GCM runoff. Sperna Weiland *et al.*, 2012b applied Kinematic Wave routing for two GCMs at 0.5° resolution in six large basins including the Mississippi River Basin. Koirala *et al.* (2014) used the Catchment-based Macro-scale Floodplain Model (CaMa-Flood) global river routing model for daily GCM runoff at 0.25° resolution. Arora (2001) used a variable velocity flow routing algorithm from Arora and Boer (1999) to assess GCM generated streamflow. Comparing the variable velocity flow routing algorithm to WATROUTE for the Mackenzie River Basin, Arora *et al.*, (2001) showed that differences in routing schemes did not introduce major differences in simulated streamflow and that spatial resolution can contribute to more substantial differences. In a comparison of routing methods Rousseau *et al.* (2014) also found Kinematic Wave routing to be appropriate for hydrological applications. Some of the Kinematic Wave limitations discussed in Rousseau *et al.* (2014) are overcome by implementing a minimum channel slope, as is the case in WATROUTE.

Overall, it is important to recognize that many different routing models exist and many have been applied in GCM studies. The WATROUTE model has been tested and accepted in

hydrologic studies and balances computational efficiency with data requirements of large domain watersheds, such as the NCRB.

### **2.3.2 Future Runoff and Streamflow Scenarios from Climate Models**

Typical future climate assessments combine future greenhouse gas scenarios with GCMs and some form of downscaling technique and/or bias correction technique to produce temperature and precipitation projections. Temperature and precipitation projections are then used to force calibrated hydrological models which are then used for impact assessments of future streamflow. Most of these GCM studies compared 20 year or 30 year climates in a baseline period (e.g., 1971 to 2000) to a future period such as the 2050s (e.g., 2041 to 2070). Studies have focussed on temperature, precipitation, streamflow as well as drought variables. Recent studies such as Milly *et al.* (2005), Nohara *et al.* (2006), Meehl *et al.* (2007), Sperna Weiland *et al.* (2012a), Sperna Weiland, *et al.* (2012b), Collins *et al.* (2013) and Koirala *et al.* (2014) have explored the use of direct GCM hydrologic data across large areas whereas Sushama, *et al.* (2006), Frigon, Music and Slivitzky (2010), Music *et al.* (2012), Poitras, *et al.* (2011), Clavet-Gaumont *et al.* (2013) and Huziy *et al.* (2013) have explored the use of direct Regional Climate Model (RCM) hydrologic data in smaller basins. Comparing RCM projections to their driving GCM, Gao *et al.* (2011) found that annual runoff changes were generally consistent among the two model types. For seasonal changes, the authors showed that GCMs exhibited larger inter-model variability than RCMs. The authors also noted that differences between GCM and RCM projected runoff changes can be important in areas of high topographic relief and large snow accumulations. Most existing studies focused on the CMIP3 ensemble of GCMs.

Around the time of the IPCC's Second Assessment Report, it was recognized that hydrologic variables such as runoff were poorly reproduced by GCMs (Wood *et al.*, 1997). Since

then, several studies have demonstrated how modeling advances have increased the hydrologic representation in GCMs. Sperna Weiland *et al.* (2012a) investigated the suitability of using GCM generated runoff fields for hydrologic impact studies and found that when GCM runoff was tuned with discharge observations and a routing scheme is added, direct GCM runoff can be as suitable as discharge derived from runoff calculated by an offline hydrological model for large scale studies. Additionally, some GCMs use land surface schemes that have benefited from coupling with land surface schemes in more detailed hydrological models (Soulis *et al.*, 2000). Suitable results obtained using runoff from the Coupled Model Intercomparison project Phase 3 (CMIP3) GCMs in large basins ( $>170,000 \text{ km}^2$  in Sperna Weiland *et al.*, 2012a) with different climates show promise for the use of GCM runoff in other large basins. Huziy *et al.*, 2013 proposed that GCMs and RCMs are well suited to evaluate the climate change impacts on streamflow because of their physical basis which includes a closed water budget. Furthermore, Guay, Minville and Braun (2015) suggested that climate models may produce more reliable estimates of actual evapotranspiration since they are based on radiative balances instead of empirical relationships using temperature data. Use of direct GCM runoff data leverages the immense investment in computer processing involved in climate modeling (reducing off-line processing) while still accounting for many of the sources of uncertainty in climate change studies and producing results that are reasonable for use in hydrologic impact assessment. The use of GCM direct runoff data captures the largest sources of uncertainty previously assessed in other studies such as Kay *et al.* (2009) and Chen *et al.* (2011).

Many studies which assessed climate change impacts on hydrology considered a global domain, in which information about specific regions can be extracted. Of the available global studies, Milly *et al.* (2005), Meehl *et al.* (2007), Collins *et al.* (2013), Nohara *et al.* (2006),

Sperna Weiland *et al.* (2012a) generally illustrated increasing mean annual runoff volume in the NCRB as a whole. However, greater changes and greater model agreement were projected for northern parts of the basin and some decreases were projected for southwestern parts of the basin. Alkama *et al.* (2013) showed increasing annual runoff for North America from an ensemble of CMIP5 GCMs. Kumar *et al.* (2014) assessed monthly water availability (precipitation minus evapotranspiration) in the CMIP5 ensemble and found future water availability to increase in the winter and decrease in the summer in North America under a more severe future greenhouse gas scenario. Koirala *et al.*, (2014) also assessed global runoff from an ensemble of 11 CMIP5 GCMs, opting to assess relative changes without bias correction. The authors presented the multi-model mean which projected increasing mean and low flows with decreasing high flows in the NCRB. However, there was weak to moderate GCM agreement that low flows will increase and less evidence for changes in mean and high flows.

Inconsistent results can be found in the literature that reduces confidence in how future hydrology is projected to change. In a study on changing flood and drought risk, Hirabayashi *et al.* (2008) used a single GCM and analyzed future hydrological projections over 30 large basins. For the Nelson River's future, the authors projected annual runoff to decrease by 17.8%, the frequency of the 100 year flood to become a 133,000 year flood, and three times more drought days. However, it should be noted that this study only used one GCM simulation, which projected relatively large temperature increases (+6.57°C) and an overall very dry future in the NCRB. Interestingly, the authors noted that an increase in annual discharge does not always correspond to an increase in 100 year flood frequency and that future changes in drought should be correlated to changes in long term water availability (i.e., precipitation minus evapotranspiration) and not precipitation alone. In a similar study, Milly *et al.* (2002) analyzed

annual maximum of monthly-mean flows from a climate model under an idealized quadrupled CO<sub>2</sub> scenario to assess changes in 100 year flood events. Downstream of Lake Winnipeg at Bladder Rapids on the Nelson River, mean annual discharge was projected to increase by 76%, a future 100 year event was projected to increase by 36% and the baseline 100 year event could become a 9 year event in the future. The quadruple CO<sub>2</sub> is quite extreme and the authors found that the frequency of floods having shorter return periods did not change significantly.

Using the CRCM, Poitras *et al.* (2011) combined CRCM runoff with the WATROUTE routing scheme in Western Canada to examine climate change impacts on seasonal streamflow including low and high flow events. This study considered the Nelson River Basin, Churchill River Basin and isolated the North Saskatchewan and South Saskatchewan River sub-basins. The four river basins all showed increases in mean annual flows (12% to 17%), increases in mean winter flows, and earlier snow melt with a higher peak. Sushama *et al.* (2006) analyzed CRCM runoff routed using a variable-lag flow algorithm, in the Nelson and Churchill River Basins. Different CRCM model versions and future emission scenarios were used, showing that the annual change in runoff (-6% to +9% for the Nelson River and -1% to +13% for the Churchill River) was sensitive to model and forcing scenario. Both Nelson and Churchill rivers showed small increases to 7-day low flow events, increases to late-winter flows and decreases to snowmelt peaks but insignificant changes to high flow events.

Overall, studies that applied direct GCM or RCM runoff presented similar results as studies that coupled GCM climate forcing to hydrological models. Hagemann *et al.* (2013) showed increases in a majority of NCRB but also indicated decreasing water availability or very little change in some arid regions. These authors noted the importance of considering seasonality, supported the conclusion that wet (dry) areas are projected to get wetter (drier) and

suggested that while bias correction reduced the spread of projections, it was not a large source of uncertainty compared to other sources. Haddeland *et al.* (2014) used GCMs and hydrological models to assess the cumulative impacts of anthropogenic impact (irrigation withdrawals, reservoir operation) and a two degree increase in global mean temperature on global annual runoff. At northern latitudes, including the NCRB and Mackenzie River Basin, there was a clear increase in runoff and the anthropogenic impacts were small. However, in the Mississippi and Colorado River basins, the anthropogenic effect outweighed the climate change effect. The authors noted that in all cases, water withdrawals increased with increasing temperatures but climate change can cause increases or decreases in water availability therefore amplifying or damping the effect of increased withdrawals. Prudhomme *et al.* (2014) assessed global hydrologic drought, measured by the number of days where runoff less than 10<sup>th</sup> percentile. When using the multi model ensemble (5 GCMs and 7 hydrological models) most of NCRB showed 5-10% increase in drought days. However there is less agreement among GCMs and hydrological models in North America including one model that showed decreased future drought (-5%).

Limitations in hydrological model studies include the absence of feedback from the hydrological model to the atmosphere (Hagemann *et al.*, 2013) and the absence of dynamic vegetation schemes in most hydrological models. Dynamic vegetation can be important since increased CO<sub>2</sub> can reduce stomatal openings in plants and reduce transpiration. Prudhomme *et al.* (2014) showed that a hydrological model with dynamic vegetation projected decreased future drought risk in some areas which is partially attributed to reduced plant transpiration.

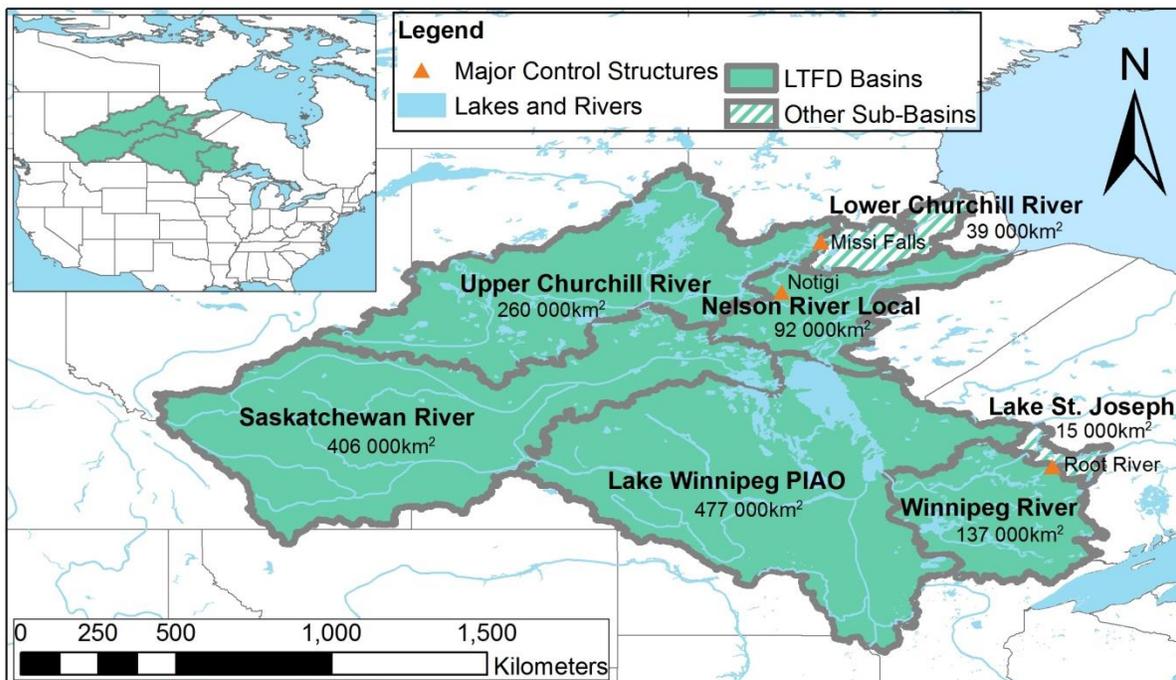
Cohen, Koshida and Mortsch (2015) addressed the infrequent application of raw climate model runoff data in past studies but suggested that there is an opportunity to explore ways to

implement routed climate model runoff data to complement current practise. As demonstrated in this chapter, considerable effort has led to improved GCMs, LSMs and applications of their runoff data directly in climate impact studies. Also noteworthy are the limitations in existing studies that used relatively short 30 year time periods to characterize climate and reliance on hydrological models. The use of shorter time periods can increase sampling uncertainty which can be important for analysis of extremes in hydrologic records with low frequency oscillations. Limitations due to reliance on hydrological models include the absence of feedback from hydrological model to climate model, temperature based evapotranspiration schemes, computational resources and other resource requirements to set up calibrate and validate the models. Although the use of GCM runoff data in climate impact studies comes with its own set of limitations, results can provide researchers and practitioners with a more balanced view using multiple assessment approaches. This thesis intends to help build that balanced view by applying tools and methods for assessment of long term hydrology to long term simulations from the latest CMIP5 GCMs and their runoff variable.

## CHAPTER 3

### Spatial and Temporal Domain

This chapter characterizes the spatial and temporal domain of study for the proposed objectives. In alignment with Manitoba Hydro interests, the spatial domain is the Nelson-Churchill River Basin (NRCB) and is shown in Figure 4. The NRCB is separated into five sub-basins, consistent with Manitoba Hydro's Long Term Flow Dataset (LTFD) used in planning studies. The five LTFD sub-basins are the: Upper Churchill River Sub-Basin (UCR), Nelson River Local Sub-Basin (NRL), Saskatchewan River Sub-Basin (SRB), Lake Winnipeg Partial Inflow Available for Outflow (PIAO) and Winnipeg River Sub-Basin (WRB).



**Figure 4 - The Nelson-Churchill River Basin and sub-basin spatial domains**

The NRCB covers approximately 1.4 million km<sup>2</sup> spanning west to the Rocky Mountains, east to Lake Superior, stretching as far south as South Dakota, United States and nearly to the northern borders of Manitoba and Saskatchewan. For the 1950-2005 period, Table 2 presents

annual mean flows calculated using LTFD, drainage areas (Agriculture and Agri-food Canada, 2012) and approximate annual mean runoff (mm/day) calculated by dividing mean flow by the effective drainage area. Similar to Ehsanzadeh, *et al.* (2011), effective drainage area is used to account for non-contributing drainage areas (such as prairie potholes), which do not contribute to streamflow under certain climatic conditions. Effective drainage areas presented in Agriculture and Agri-food Canada (2012) are representative of average hydrological conditions (i.e., a 1 in 2 year event). Annual mean runoff values shown in Table 2 are similar to values reported in Statistics Canada (2010) for the period of 1971-2004.

**Table 2 - Nelson-Churchill River Basin and sub-basin mean annual flow and runoff for 1950-2005**

Basin	Mean Annual Flow		Gross Drainage Area (km <sup>2</sup> )	Effective Drainage Area (km <sup>2</sup> )	Mean Annual Runoff (mm/day)
	(m <sup>3</sup> /s)	% of NCRB			
UCR	972	27.3%	260,000	252,000	0.333
NRL	425	12.0%	92,000	92,000	0.399
SRB	575	16.2%	406,000	231,000	0.215
PIAO	647	18.2%	477,000	313,000	0.179
WRB	935	26.3%	137,000	135,000	0.598
NCRB	3555	100.0%	1,372,000	1,023,000	0.300

Two additional NCRB sub-basins are notable in this thesis. The Lower Churchill River Sub-Basin stretches from Missi Falls, the natural outlet of Southern Indian Lake, to the mouth of the Churchill River Basin at Hudson Bay. Together the Upper and Lower Churchill River Basins naturally drained the Churchill River into the Hudson Bay. Starting in the late 1970s, a portion of the Churchill River flow was diverted at Southern Indian Lake, through the Notigi control structure and into the Local Nelson River Sub-Basin to assist power production on the Nelson River. This system is referred to the Churchill River Diversion (CRD; Newbury, McCullough and Hecky, 1984) and is operated by Manitoba Hydro. Similar to the CRD, water is also diverted at Lake St. Joseph into the Winnipeg River Sub-Basin through the Root River control structure. The Lake St Joseph Diversion has been in operation since the late 1950s and is currently

operated by Ontario Power Generation for power production purposes (Lake of the Woods Control Board, 2014). Manitoba Hydro's generating stations downstream from the diversion also benefit from the additional diverted water. Other diversions exist within the NCRB but are not considered in this thesis due to their small volumes or transfer of water within the larger basin boundary.

Manitoba Hydro operates 15 hydroelectric generating stations in the NRCB with a total capacity of approximately 5,200 megawatts (MW), producing approximately 30,000 GWh of energy each year. Of these 15 generating stations, six are located in the WRB (579 MW total), one is located in the SRB (479 MW), two are located in a small sub-basin of the UCR (10 MW total) and the remaining six are located in the NRL (4,149 MW total). Keeyask, a seventh generating station is currently being constructed in the NRL, which will add approximately 695 MW of capacity and 4,400 GWh of annual average energy.

Manitoba Hydro operates several reservoirs in the NCRB, the largest of which being Lake Winnipeg which covers an area of 24,400 km<sup>2</sup>. Manitoba Hydro also operates Southern Indian Lake, Cedar Lake, and a series of smaller reservoirs. Operation of these reservoirs can affect the natural flow regime in the rivers, but it should be noted that since a large portion of the NCRB lies upstream of Manitoba Hydro's operations, the rivers' natural flows are generally augmented by other organisations (e.g., for power production, flood relief, water supply and agricultural use). As such, much of the flow into Manitoba Hydro's reservoirs is affected by anthropogenic influenced to some degree.

Three general temporal windows are considered, however some variation is required for different steps of the analysis. Temporal windows are a combination of available data and CMIP5 experimental design (Taylor, Stouffer and Meehl, 2012), which include a historic period,

a preindustrial control (piControl) period, and a future period. Based on available GCM data, the historic period used in this study spans from 1861 to 2005 and the future period continues from 2006 to 2099. For the purposes of analysis and plotting of data, the 500 year piControl period is nominally assigned a date range of 1361 to 1860. This range is arbitrary because the piControl period corresponds to a 500 year un-dated, quasi-equilibrium GCM experiment where atmospheric forcing is held constant (i.e., atmospheric CO<sub>2</sub> concentration is held constant at approximately 280 parts per million; ppm) and the GCM's are allowed to simulate unforced natural climate variability.

Two additional reference periods are used from within the historic period. For the evaluation of model skill, a reference period of 1950 (earliest date when GCM data is commonly available for the large GCM ensemble) to 2005 is selected. For bias correction, a reference period of 1912 (earliest observed LTFD entry) to 2005 is selected. For streamflow scenarios and time series analysis, the piControl and historic periods are further divided into quasi-equal blocks of time (91 to 94 years), one of which (piControl 6; pi6) spans both the piControl and Historic periods. A summary of the temporal domains are provided in Table 3.

**Table 3 - CMIP5 temporal domains and additional temporal domains used in analysis**

<b>CMIP5</b>		<b>Streamflow Scenario Analysis</b>		<b>Additional Reference Periods</b>	
<b>Period</b>	<b>Dates</b>	<b>Period</b>	<b>Dates</b>	<b>Period</b>	<b>Dates</b>
piControl	1361-1860	piControl 1 (pi1)	1362-1452	Reference for Evaluation of Model Skill	1950-2005
		piControl 2 (pi2)	1453-1543		
		piControl 3 (pi3)	1544-1634	Reference for Bias Correction	1912-2005
		piControl 4 (pi4)	1635-1726		
		piControl 5 (pi5)	1727-1818		
		piControl 6 (pi6)	1819-1911		
Historic	1861-2005	Historic	1912-2005		
Future	2006-2099	Future	2007-2099		

## CHAPTER 4

### Methodology and Data Sources

This chapter outlines the methods and data sources used in this thesis, divided into sections according to the four thesis sub-objectives.

#### 4.1 Evaluation of Climate Model Skill

Climate model skill is first evaluated in order to better understand how well different GCMs represent existing hydrologic conditions in the NCRB. Initially, 16 skill metrics were chosen but this was later reduced to nine as many of the errors among the metrics were found to be well correlated. The nine metrics considered in this thesis are shown in Table 4 and selected such that they are relatively simple and statistically robust in accordance with Knutti *et al.* (2010a) and appropriate for the objectives of this thesis in accordance with Johnson *et al.* (2011). Daily total runoff is the only GCM variable used in this thesis and was selected based on routing model requirements and data availability. Daily total runoff represents the total runoff leaving a land grid cell, including drainage through the base of the soil and is identified by the “mrro” output variable name (Taylor, 2013).

**Table 4 - Skill metrics for GCM evaluation**

<b>Skill Metric Identifier</b>	<b>Description (units)</b>
Mean	Mean annual flow (mm/day)
Min	Minimum of mean annual flow (mm/day)
Max	Maximum of mean annual flow (mm/day)
Var	Variance of mean annual flow (mm/day)
AR1	Annual autocorrelation lag-1
Slope	Slope (percent of mean flow)
Cum. Wet	Mean number of cumulative wet years (years)
Cum. Dry	Mean number of cumulative dry years (years)
Cross Corr.	Spatial cross correlation (15 basin-to-basin combinations)

Metrics are based on daily total runoff averaged throughout a calendar year; therefore the timing of runoff is not evaluated. Runoff timing is not incorporated in the skill analysis because in the absence of routing (as is the case for this preliminary skill assessment), it is difficult to compare GCM runoff to available observed streamflow. Second, no reliable, long term, observed gridded runoff dataset exists. Finally, observed streamflow records often contain considerable anthropogenic influence, such as reservoir operations that can alter hydrograph timing. Use of annual average flows focuses the skill assessment on runoff volume instead of timing, as in Zhou *et al.*, 2012. However, even when considering annual runoff, interannual water storage or consumptive use (volume withdrawal) in the observed record is not accounted for.

The observed dataset that GCMs are evaluated against is Manitoba Hydro's LTFD. Streamflow records from 1912 to current are available for the outlets of the five sub-basins described in Chapter 3 (consistent with Manitoba Hydro's resource planning) and are adjusted to represent present day use. LTFD also includes some smaller sub-basins within the NRL but for the purposes of this thesis, these smaller basins are aggregated into one, larger, sub-basin. In some instances, it is also useful to aggregate all sub-basins as done in Table 2 to produce a measure of system-wide NCRB inflows. For the skill evaluation, daily runoff is estimated from LTFD by dividing mean annual streamflow by the effective drainage area of each of the sub-basins.

Since the skill evaluation is performed at the outlet of each sub-basin, GCM runoff must be aggregated to represent total sub-basin response. To uniformly evaluate all GCMs, which operate at various spatial resolutions, GCM runoff is re-gridded to a common grid. This is a common step in skill evaluation for multiple GCMs (Sillmann *et al.*, 2013; Gleckler *et al.*, 2008). A  $0.25^\circ$  latitude x  $0.25^\circ$  longitude grid (approximately 25 km x 25 km) is used based on the

resolution selected for the routing model, and balancing of spatial detail with computational demand. Re-gridding is performed using a nearest neighbour approach which conserves runoff volume within a GCM grid. For example, if a 1° x 1° GCM grid produced 1 mm of runoff in a given day, each of the 16 0.25° x 0.25° grids within the original grid would also produce 1 mm of runoff on that same day. Other interpolation methods such as bilinear interpolation in Koirala *et al.* (2014) may produce more realistic spatial patterns but can introduce additional errors that affect the volume of water. In a comparison of methods, González-Zeas *et al.* (2012) showed that the nearest neighbour interpolation performed better than an inverse distance squared interpolation.

Many skill assessments consider 20 or 30 year periods, however, given the potential for low frequency climate patterns to impact runoff climatology, a longer 56 year period (1950-2005) was selected. 1950-2005 represents a period in which all analyzed GCMs have daily runoff data. Skill is computed for each metric in Table 4 based on 56 year climatological means using Root Mean Squared Error (RMSE):

$$RMSE_{Metric} = \sqrt{\langle (GCM_{Metric} - Observed_{Metric})^2 \rangle} \quad (6)$$

where  $GCM_{Metric}$  and  $Observed_{Metric}$  denote the 56 year climatologies for each metric and  $\langle \rangle$  denotes the spatial mean calculation

Similar to Sillmann *et al.* (2013), RMSE is calculated as a spatial mean, covering the five sub-basins plus the entire NCRB (six domains total). In the case of spatial cross correlation, the spatial mean is calculated based on 15 individual basin-to-basin cross correlations, including the aggregated NCRB. The entire NCRB is included in the nine metrics because some GCMs may perform better at larger spatial scales instead of individual sub-basin scales. Two additional error terms from Sillmann *et al.* (2013) are also considered: the Relative Root Mean Squared Error (RMSE'; Equation 7) and the Mean Relative Root Mean Square Error (RMSE'\_{ALL}; Equation 8):

$$RMSE'_{Metric} = \frac{RMSE_{Metric} - RMSE_{Median(Metric)}}{RMSE_{Median(Metric)}} \quad (7)$$

$$RMSE'_{ALL} = \frac{RMSE'_{Metric1} + RMSE'_{Metric2} + \dots + RMSE'_{Metric9}}{9} \quad (8)$$

RMSE' provides information on how individual models rank with respect to the ensemble, neglecting absolute errors. For every metric there will be a GCM simulation whose performance is the median of all simulations. For each skill metric, an RMSE' value of zero is assigned to the simulation with the median performance. Models performing better (worse) than the median simulation will have negative (positive) values of RMSE'. RMSE'\_{ALL} is an overall score for each GCM simulation based on its relative performance for all nine metrics. The best (worst) performing simulation, overall, will have the smallest (largest) RMSE'\_{ALL}.

GCM data for the skill evaluation comes from the output of 23 CMIP5 models with daily runoff available for the historic period. From the 23 GCMs, 53 simulations were evaluated, which include GCMs with multiple runs. GCM data were extracted in January 2014 and provided by the Ouranos Consortium on Regional Climatology and Adaptation to Climate Change. Since CMIP5 databases are continually updated as modeling centers make their data available, it is possible that more daily runoff simulations have been made available since the extraction date for this study. Table 5 (adapted from Flato *et al.*, 2013) shows the GCMs used in this thesis along with the institution name and the name of their LSM. Table 6 shows the individual GCM simulations and available data for piControl (at least 500 years), historic (1950-2005) and future (2006-2099) periods. Some historic GCM simulations also contained longer historical periods starting in 1861.

**Table 5 - GCM name, institution and land surface model name (adapted from Table 9.A.1 in Flato *et al.*, 2013)**

<b>Model Name</b>	<b>Institution</b>	<b>Land Surface Model Name</b>
BCC-CSM1.1	Beijing Climate Center	BCC-AVIM1.0 (based on CLM3)
BNU-ESM	Beijing Normal University	ColM
CanESM2	Canadian Center for Climate Modeling and Analysis	CLASS 2.7
CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici	From ECHAM5 GCM
CMCC-CMS		
CNRM-CM5	Centre National de Recherches Meteorologiques and Centre Europeen de Recherche et Formation Avancees en Calcul Scientific	ISBA (through SURFEX)
CSIRO-Mk3.6.0	Queensland Climate Change Centre of Excellence and Commonwealth Scientific and Industrial Research Organization	Included within Atmospheric Model
CSIRO-Mk3L 1.2		
FGOALS-g2	LASG (Institute of Atmospheric Physics)- CESS (Tsinghua University)	CLM3
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory	LM3
GFDL-ESM2G		LM3
GFDL-ESM2M		LM3
INM-CM4	Russian Institute for Numerical Mathematics	Included (Simple)
MIROC-ESM-CHEM	University of Tokyo, National Institute for Environmental Studies and Japan Agency for Marine-Earth Science and Technology	MATSIRO
MIROC-ESM		
MIROC4h		
MIROC5		
MPI-ESM-LR	Max Planck Institute for Meteorology	JSBACH
MPI-ESM-MR		
MPI-ESM-P		
MRI-CGCM3	Meteorological Research Institute	HAL
MRI-ESM1		
NorESM1-M	Norwegian Climate Centre	CLM4

**Table 6 - GCM simulation data availability from various CMIP5 experiments (as of January, 2014). For historic experiments, S denotes simulations containing a shorter record (1950-2005) and L denotes simulations containing a longer record (1861-2005).**

Simulation #	Model and Run	CMIP5 Experiment						Simulation #	Model and Run	CMIP5 Experiment										
		piControl	Historic	Rep26	Rep45	Rep60	Rep85			piControl	Historic	Rep26	Rep45	Rep60	Rep85					
1	BCC-CSM1.1 r1		S	✓	✓	✓	✓	30	MIROC-ESM-CHEM r1		S	✓	✓	✓	✓					
2	BCC-CSM1.1 r2		S					31	MIROC-ESM r1		S	✓	✓	✓	✓					
3	BCC-CSM1.1 r3		S					32	MIROC-ESM r2		S									
4	CanESM2 r1*	✓	L	✓	✓		✓	33	MIROC-ESM r3		S									
5	CanESM2 r2*		L	✓	✓		✓	34	MIROC4h r1		S		✓							
6	CanESM2 r3*		L	✓	✓		✓	35	MIROC5 r1*	✓	L	✓	✓	✓	✓					
7	CanESM2 r4		L	✓	✓		✓	36	MIROC5 r2*		L	✓	✓		✓					
8	CanESM2 r5		L	✓	✓		✓	37	MIROC5 r3*		L	✓	✓		✓					
9	CMCC-CM r1		S		✓		✓	38	MIROC5 r4		L		✓		✓					
10	CMCC-CMS r1		L		✓		✓	39	MIROC5 r5		L		✓		✓					
11	CNRM-CM5 r1		S		✓		✓	40	MPI-ESM-LR r1		S	✓	✓		✓					
12	CSIRO-Mk3.6.0 r1		S	✓	✓	✓	✓	41	MPI-ESM-LR r2		S	✓	✓		✓					
13	CSIRO-Mk3.6.0 r10		S					42	MPI-ESM-LR r3		S	✓	✓		✓					
14	CSIRO-Mk3.6.0 r7		S					43	MPI-ESM-MR r1		S	✓	✓		✓					
15	CSIRO-Mk3.6.0 r8		S					44	MPI-ESM-MR r2		S		✓							
16	CSIRO-Mk3.6.0 r9		S					45	MPI-ESM-MR r3		S		✓							
17	CSIRO-Mk3L 1.2 r1		L					46	MPI-ESM-P r1		S									
18	CSIRO-Mk3L 1.2 r2		L					47	MPI-ESM-P r2		S									
19	CSIRO-Mk3L 1.2 r3		L					48	MRI-CGCM3 r1		S	✓	✓	✓	✓					
20	FGOALS-g2 r1		S	✓	✓		✓	49	MRI-ESM1 r1		S				✓					
21	FGOALS-g2 r3		S					50	NorESM1-M r1*	✓	L	✓	✓	✓	✓					
22	GFDL-CM3 r1	✓	L					51	NorESM1-M r2		L									
23	GFDL-CM3 r2		L					52	NorESM1-M r3		L									
24	GFDL-CM3 r3		L					N/A	BNU-ESM r1		S	✓	✓		✓					
25	GFDL-CM3 r4		L					<b>Total Simulation Count:</b>						6	53	21	31	9	29	
26	GFDL-CM3 r5		L																	
27	GFDL-ESM2G r1*	✓	L	✓	✓	✓	✓													
28	GFDL-ESM2M r1*	✓	L		✓	✓	✓													
29	INM-CM4 r1		S		✓		✓													

\* Denotes simulations that were selected for routing, bias correction and streamflow time series analysis.

## 4.2 River and Lake Routing

The river and lake routing methods used in this thesis are largely based on WATROUTE. However, since WATROUTE is typically set up in conjunction with WATFLOOD using pre-processors such as Green Kenue (CHC, 2010), some additional data and pre-processing were required. Additionally, some adjustments were made to the original code to allow for implementation in the MATLAB software package (MATLAB, 2012), and to enable the code for use with the GCM daily runoff data. MATLAB was chosen for its common availability within academic and industry communities, for straight forward interfacing with GCM data and for the researcher's familiarity with the software. The routing method developed for this thesis will be henceforth referred to as WATROUTE<sub>MOD</sub>, a modified version of the original WATROUTE code. A description of WATROUTE can be found in Kouwen (2012) and was summarized in Chapter 2. This Chapter contains a description of the unique pre-processing features and code implementation used for WATROUTE<sub>MOD</sub> in this thesis.

For comparison, Table 7 summarizes basic physiographic properties in NCRB sub-basins for the WATFLOOD/WATROUTE model set up by Stadnyk and Newsom (2013) and WATROUTE<sub>MOD</sub>. Several differences exist between these two models and it should be noted that neither model was calibrated in a traditional sense and is subject to change. Of particular note is WATFOOD's consideration of water coverage within grids (determined by land cover data), which inherently affects channel properties in WATROUTE. Channel width and depths are typically calculated using geomorphological relationships but in grids with considerable water coverage, WATFLOOD internally increases channel widths accordingly. Further comparison between WATROUTE<sub>MOD</sub> and a separate calibrated WATFLOOD/WATROUTE model for the WRB with finer resolution is discussed in Chapter 5.

**Table 7 - Physiographic comparison of WATROUTE (from Stadnyk and Newsom, 2013) and WATROUTE<sub>MOD</sub> models by NCRB sub-basin**

	<b>Parameter</b>	<b>WATROUTE</b>	<b>WATROUTE<sub>MOD</sub></b>
	Spatial Resolution	0.14°lat x 0.18°lon	0.25°lat x 0.25°lon
<b>UCR</b>	Mean channel slope	0.0011	0.0017
	Mean bankfull channel geometry (width/depth in m)	2010.8 / 1.9*	44.3 / 1.0
	Mean Manning's n (in-channel/flood plain)	0.003 / 0.400	0.010 / 0.400
<b>NRL</b>	Mean channel slope	0.0007	0.0014
	Mean bankfull channel geometry (width/depth in m)	1454.2 / 1.7*	105.8 / 1.9
	Mean Manning's n (in-channel/flood plain)	0.009 / 0.400	0.010 / 0.400
<b>SRB</b>	Mean channel slope	0.0024	0.0035
	Mean bankfull channel geometry (width/depth in m)	589.9 / 3.7*	60.5 / 1.3
	Mean Manning's n (in-channel/flood plain)	0.013 / 0.400	0.010 / 0.400
<b>PIAO</b>	Mean channel slope	0.0014	0.0015
	Mean bankfull channel geometry (width/depth in m)	1399.7 / 3.0*	40.8 / 0.9
	Mean Manning's n (in-channel/flood plain)	0.005 / 0.317	0.010 / 0.400
<b>WRB</b>	Mean channel slope	0.0008	0.0011
	Mean bankfull channel geometry (width/depth in m)	2534.6 / 2.7*	55.7 / 1.1
	Mean Manning's n (in-channel/flood plain)	0.019 / 0.400	0.010 / 0.400

\*Values show WATFLOOD's internally corrected bankfull channel geometry.

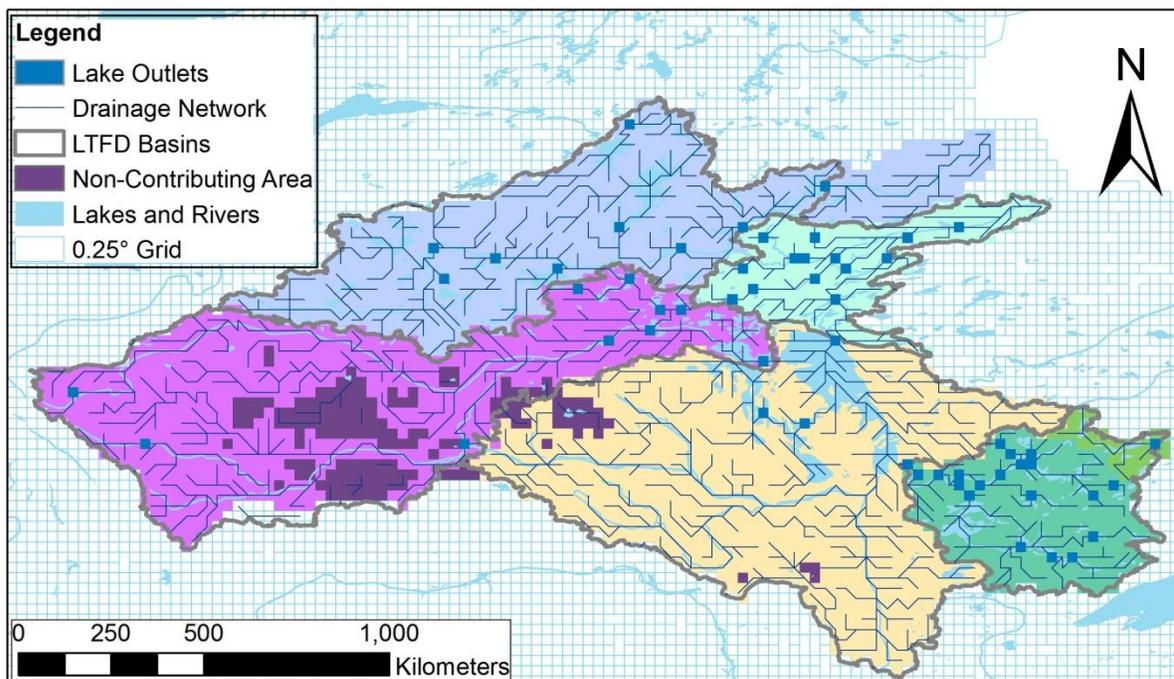
Gridded routing models require several pieces of information within each grid, a large portion of which is derived from Digital Elevation Models (DEMs). Considerable work already exists where elevation data from DEMs have been processed into information, such as drainage directions, required for routing models. This thesis leverages that existing work, largely based on the publically available, global coverage HydroSHEDS (Lehner, Verdin and Jarvis, 2008) and Hydro1k (U.S. Geological Survey, 2000) DEM products.

The river network and its drainage directions are critical inputs to the routing model. Wu *et al.* (2011) presented an automated method, dominant river tracing (DRT), which upscaled river networks in Hydro1k to coarser resolutions (2°, 1°, 0.5°, 0.25°, 0.125° and 0.0625°) for use in macroscale hydrological modeling. Selecting an appropriate resolution requires balancing detail and computational requirements. The authors compared the various resolutions and showed that a resolution of 0.25° (grid size of about 900km<sup>2</sup>) was reasonable for capturing basin shape and area for basins as small as the Salmon River, Idaho, (35,502km<sup>2</sup>) in 35 grids. The data

product in Wu *et al.* (2012) is used in this thesis and is an updated version of Wu *et al.* (2011) using the HydroSHEDS product. Flow direction and flow distance (i.e., river length in a grid) at 0.25° resolution from Wu *et al.* (2012) are used in WATROUTE<sub>MOD</sub>. A resolution of 0.25° results in grid cells with an average area of 462 km<sup>2</sup> and discretizes the NRCB domain into 3,321 grids, not including grids with non-contributing drainage areas. At 0.25° resolution, WATROUTE<sub>MOD</sub> takes approximately 45 minutes to simulate one year of streamflow on an Intel Xenon processors operating at 3.47 Gigahertz with 24 Gigabytes of RAM on a 64-bit operating system. WATROUTE and WATROUTE<sub>MOD</sub> are both linearly programmed using upstream to downstream, cell by cell, computation. Processing speed improvements could be realized by making the model run parallel when possible (e.g., by simulating upstream basins simultaneously), but this was outside the scope of this thesis.

Wu *et al.* (2012) noted drainage direction discrepancies in flat areas, including an area overlapping the NCRB. Pietroniro *et al.* (2007) also recognized difficulties in resolving coherent drainage directions from DEMs and noted that adjustments are often required. For quality assurance, the flow direction and flow distance data were manually examined and some corrections were made. Three major drainage network edits were made: (1) along the border between the UCR and SRB, near the NRL, (2) in the NRL near the Churchill River Diversion (CRD), and (3) delineation and drainage of SRB boundary. For (1), an incorrect drainage direction in one grid originally drained approximately half of the UCR into the SRB, which was manually corrected to properly drain the UCR. For (2), a small portion of the NRL northwest corner that originally drained into the UCR was manually corrected to drain into the NRL. And for (3), seven grids were edited to drain into the SRB instead of their original drainage to the neighbouring Athabasca river basin. These types of error are common in coarse flow direction

datasets. For example, the Nelson River Basin as portrayed in Poitras *et al.* (2011) appears to incorrectly include the Hayes River, which in fact drains directly into Hudson Bay very close to the Nelson River's mouth. Figure 5 shows the discretized routing model domain over the NCRB including original flow paths, non-contributing areas (closed basins) and locations of lake outlets. Lake outlets represent lakes exceeding 4 km<sup>2</sup> in area that are included in Manitoba Hydro's individual (sub-basin) WATROUTE models.



**Figure 5 - WATROUTE<sub>MOD</sub> Routing model domain for the NCRB**

Since drainage areas can behave dynamically and the routing model is stationary, no changes were made to non-contributing drainage areas. This is a potential area for further improvement and is discussed in Chapter 6. In some instances, where flow distance from Wu *et al.*, (2012) was (presumably) inaccurately reported as 1 m, this was recomputed as the average flow distance from eight neighbouring cells. The average value of eight neighbouring cells samples flow distance within the nearby physiographic region for a reasonable estimate.

Elevation data to determine river slope were derived directly from the HydroSHEDS DEM. The finer resolution DEM was averaged over the 0.25° grid, and slopes were computed as per the corrected drainage directions. In some instances, calculated slopes were very small and in rare instances, were negative. Negative slopes can occur because mean elevations at 0.25° resolution may not accurately portray the river bed profile. Arora and Boer (1999) encountered a similar problem and prescribed a minimum slope of 0.001. As such, a minimum slope of 0.001 is applied across the entire domain in WATROUTE<sub>MOD</sub>. A maximum slope of 0.235 was calculated in a western grid of the SRB, near the Rocky Mountains. A mean slope of 0.0026 was calculated for the entire NCRB.

In order to solve Manning's equation in each grid cell, river properties such as bankfull width, depth and roughness (Manning's *n*) are required. Some studies (Arora and Boer, 1999; Arora, Chiew and Grayson, 1999; Arora, 2001) have used geomorphological relationships to estimate these properties. In WATROUTE, bankfull area is a function of geomorphological relationships, and the river width to depth ratio is user specific for each river class. Recently, Andreadis, Schumann and Pavelski (2013) combined geomorphological relationships with HydroSHEDS hydrographic data to improve the estimates and produce a global dataset. Gridded bankfull widths and depths from Andreadis, Schumann and Pavelski (2013) were incorporated into WATROUTE<sub>MOD</sub>. In most large scale studies, Manning's *n* is estimated. A common range of 0.03 to 0.04 for natural streams is suggested by Arora *et al.* (1999) but a wider range of values exist in specific literature such as 0.01 (Rousseau *et al.*, 2014), 0.012 (Aldridge and Garrett, 1973) and 0.4 for overland flow in Li *et al.* (2013). In-channel and overbank Manning's *n* values in WATROUTE<sub>MOD</sub> were estimated as 0.01 and 0.4, respectively, using NCRB spatially averaged values from Stadnyk and Newsom (2013).

WATROUTE applies a dynamic computational time step as a function of channel storage and flow. The time step is selected such that water in the fastest draining grid does not completely drain. This dynamic computational time step is similar to the Courant criterion which aids numerical stability.

$$C = \left| \frac{\Delta t * v}{\Delta x} \right| \leq 1 \quad (9)$$

where  $C$  denotes the Courant criterion,  $\Delta t$  denotes the time step,  $v$  denotes the velocity and  $\Delta x$  denotes the spatial distance of a grid.

In WATROUTE, the minimum time step is user defined and typically set at five minutes, which coincides well with WATFLOOD's generation of runoff at 15 minute intervals. WATFLOOD is also typically run with much finer spatial resolutions than  $0.25^\circ$  that require smaller routing time steps. Testing was done to find a suitable, constant, time step for the routing model used in this thesis. Numerical instabilities were noted at the GCM's archived 24 hour time step but a one hour time step ensured computational stability. As such a constant one hour time step was used in WATROUTE<sub>MOD</sub>. The Courant criteria dependant time step applied in WATROUTE likely contributes to improved computational performance compared to WATROUTE<sub>MOD</sub>.

Lake routing is coded into WATROUTE<sub>MOD</sub> for 57 lake outlets, capturing 86.8% of the lake surface area that is currently captured in existing WATROUTE models for the NCRB. In WATROUTE, lakes are coded using stage-discharge through power law or polynomial functions, which can be calibrated to capture natural lakes as well as regulated lakes. Except for the Lake St. Joseph Diversion, WATROUTE<sub>MOD</sub> does not account for regulation, and therefore lakes are coded using stage-storage curves and naturalized rating curves. Where stage-storage data is available (e.g., Lake of the Woods Control Board, 2014), it is used directly. In other cases, stage-storage is estimated based on lake surface area, similar to the process followed in

WATROUTE. Naturalized rating curves (with no regulation decisions) are approximated using the broad-crested weir equation:

$$Q = CLH^{1.5} \quad (10)$$

where  $Q$  is flow in  $m^3/s$ ,  $C$  is assumed to be constant at 1,  $L$  is the sill length in metres and  $H$  is the water height above the sill.

The broad crested weir equation largely resembles other power functions used to approximate naturalized lake rating curves such as those in Pietroniro *et al.* (2007) for the Great Lakes. In order to best estimate the natural flow conditions, the sill length ( $L$ ) is estimated by reorganizing the equation, assuming average flow conditions, average lake levels above the sill and solving for  $L$ . In cases where flow and height data are not available, average flow from a nearby gauge is used and height above the sill is arbitrarily set to 1 m. WATROUTE also allows for a Manning's roughness ( $n$ ) multiplier to account for lakes that are present within a grid but not resolved by WATROUTE. Since spatially uniform roughness values were implemented in WATROUTE<sub>MOD</sub>, a roughness multiplier, which accounts for grids with considerable water area, was not used. As such, WATROUTE<sub>MOD</sub> may provide less flow damping in certain sections when compared to an equivalent WATROUTE model.

Although WATROUTE<sub>MOD</sub> is developed to look at naturalized flow conditions, it is important to consider the Lake St. Joseph Diversion as it adds a considerable amount of volume ( $80 m^3/s$  on average) to the WRB, which is reflected in the LTFD historic record. Rather than coding a complex decision algorithm into WATROUTE<sub>MOD</sub>, a simple three-state decision was coded. This decision operates on the limits that the maximum diversion outflow is  $245 m^3/s$ , a minimum preferred diversion outflow is  $10 m^3/s$  and a minimum mean annual outflow into the Albany River is  $18 m^3/s$  (Lake of the Woods Control Board, 2014). Accordingly, WATROUTE<sub>MOD</sub> partitions Lake St. Joseph flows in accordance with the following rules:

$$\text{If } Q_{L.St.J.} \geq 263 (245 + 18)$$

$$Q_{Diversi\text{on}} = 245$$

$$\text{If } Q_{L.St.J.} \geq 28 (10 + 18) \text{ AND } Q_{L.St.J.} < 263$$

$$Q_{Diversi\text{on}} = Q_{L.St.J.} - 18;$$

$$\text{If } Q_{L.St.J.} < 28$$

$$Q_{Diversi\text{on}} = Q_{L.St.J.}/2$$

where  $Q_{L.St.J.}$  is the naturalized Lake St. Joseph outflow, as calculated from the naturalized rating curve.  
 $Q_{Diversi\text{on}}$  is the flow diverted into the WRB

It is recognized that this representation of Lake St. Joseph diversions does not properly account for wet conditions where the diversion is reduced due to high flows on the Winnipeg River. However, the upper diversion limit remains at 245 m<sup>3</sup>/s and for average years and dry years the relationships performs reasonably.

Ice can play an important role in river and lake routing as it can naturally restrict flow, raise water levels, and cause rapidly changing water conditions when ice dams break. WATROUTE accounts for ice either through a degree day method driven by temperature, or through the use of different rating curves that can be hard-coded to switch on a specific date. Manitoba Hydro also accounts for ice effects through the use of outlet performance curves. Outlet performance curves are a function of time and show that some reservoir outlet capacities can be reduced to ~70% of their normal capacity. Although it is possible to set up a conceptual outlet performance model at each location driven by GCM temperature, the intent of this thesis is not to look at the effect of ice. Instead, a synthetic curve was created, based on outlet performance curves from Manitoba Hydro and existing WATROUTE relationships at various lakes. This synthetic curve was then applied consistently in WATROUTE<sub>MOD</sub> space and is held stationary for every year of simulation. Similarly, seasonal ice corrections were also applied in Pietroniro *et al.* (2007) for modeling the Great Lakes with WATROUTE.

Since some GCMs can produce negative runoff, it is important to handle the occurrence of negative runoff appropriately. Negative runoff represents water coming out of storage, as is the case for lakes when there is water deficit but lake volume is assumed to be constant. This is

the case for land surface schemes based on the Community Land Model (CLM; Oleson *et al.*, 2004). Negative storage is handled in WATROUTE<sub>MOD</sub> by allowing water to be removed from channel and lake storage for a particular grid. The channel or lake storage is bound at zero and will not become negative.

Although it is difficult to test routing models in highly regulated basins like the NCRB (Sushama *et al.*, 2006), a simple test was designed to compare WATROUTE to WATROUTE<sub>MOD</sub> in the WRB. This test is not intended to compare WATROUTE<sub>MOD</sub> to reality, but rather to ensure that its performance is consistent with an established WATROUTE model. The WRB WATROUTE model used for comparison was provided by Manitoba Hydro in October, 2014. The WRB WATFLOOD model was originally set up by WATFLOOD developer Dr. Nicholas Kouwen and has undergone continuous improvement and calibration at Manitoba Hydro. Past versions of the WRB WATFLOOD model were implemented in Master of Science theses at the University of Manitoba (Wruth, 2013; Slota, 2013).

A synthetic runoff time series applying a uniform 0.2 mm/day with a single annual peak of 4.7 mm/day in May was generated to perform this test. This time series was repeated for five years to allow routing models to spin up. To help with consistency, both models were run without Lake St. Joseph diversion and WATROUTE<sub>MOD</sub> used a Manning's *n* of 0.01 (in channel) and 0.4 (overbank) to better account for in-stream lakes simulated in WATROUTE. While this comparison serves as a good test for WATROUTE<sub>MOD</sub>, one should recall the differences between the two models including discretized drainage area, Manning's *n*, number of lakes, lake routing relationships, channel slopes, and channel geometry. Results from this comparison can be found in Chapter 5.

### 4.3 Bias Correction

In the absence of vetted bias correction methods for GCM generated streamflow, a QM method (daily translation; Mpelasoka and Chiew, 2009) is adapted for correcting WATROUTE<sub>MOD</sub> time series. Limitations with this approach are acknowledged, and are further discussed as potential improvements in Chapter 6. For the objectives of this thesis, however, the QM approach was chosen because it corrects for streamflow volume errors while maintaining the GCM's sequencing of events. A similar method was also tested in Hashina *et al.*, (2007) for correcting shorter term streamflow forecasts. Although some GCM runoff studies have skipped bias correction, reasoning that bias will cancel out when looking at changes in GCM baseline and future simulations (Alkama *et al.*, 2013; Kumar *et al.*, 2014), this thesis presents bias corrected results that may better reflect realistic magnitudes for analysis of results. To better understand how QM is applied in this thesis, an overview of the underlying daily translation method is provided, and then differences between daily translation and the implementation of QM in this thesis are discussed.

Daily translation can be described in three simplified steps. Step 1 compares observed empirical cumulative distribution functions (ECDFs) with GCM distributions in a reference period. In step 2, adjustment values from the ECDFs are computed at various quantiles. Step 3 applies adjustment values to GCM simulations to produce bias corrected time series. QM can be adaptable to specific applications, with differences in approaches observed in literature. For example, the reference period may include 30 years of daily data (Räty *et al.*, 2014) or 25 years of monthly data (Watanabe *et al.*, 2012). Some studies may compute adjustment values at each percentile (100 quantiles in Räty *et al.*, 2014) or might group percentiles to compute adjustment values (20 quantiles in Mpelasoka and Chiew, 2009). And some studies may compute adjustment

values for each month (Hashino *et al.*, 2007), or group months together and compute adjustment values for seasons (Mpelasoka and Chiew, 2009).

This thesis applies QM of ECDFs to grouped monthly streamflow data in a 94 year reference period (1912-2005) to compute adjustment values for 100 quantiles. The result is one adjustment value set for each sub-basin and each GCM. Using ECDFs essentially ranks the data from the smallest value to the largest value and in the case of 94 years of monthly data, 1,128 values are ranked and define the ECDF. Adjustment values for 100 quantiles based on 1,128 values therefore results in each adjustment value representing one correction for approximately 11 data points. Alternatively, the use of a continuous distribution would alleviate the need for quantile-specific adjustments but would require further assumptions about the underlying statistical distribution. Empirically defining adjustments is common in literature and the use of 11 data points per quantile is within the range used in published literature. Use of a 94 year reference period helps reduce sampling uncertainty associated with shorter time periods and may be particularly important in streamflow analysis with low frequency fluctuations.

Grouping all months together, as opposed to independently calculating adjustment values for each month, avoids the undesirable effect of correcting streamflow timing. Correcting streamflow timing is undesirable because it would rely on QM to correct for more than just the systematic GCM volume bias. The additional timing related errors include anthropogenic influence in the observed record, routing model errors, and the GCM's timing of runoff generation. The approach used in this thesis assumes that streamflow time series generated by WATROUTE<sub>MOD</sub> include monthly flow distributions that are similar to the observed record, but allows peak and low flow months to occur at different times of the year. This assumption is realistic for the occurrence of high and low flow events. For example, during an extreme high

flow event, a naturalized model will pass flow through lakes providing some natural attenuation. In reality a reservoir might first be filled prior to increasing outflows, but could still produce similar peak flows to the naturalized environment. In the case of an extreme low flow event where reservoir storage has been depleted, it is conceivable that an operator may set reservoir outflow equal to inflow, producing minimum flows similar to a naturalized environment. There may be more discrepancy in the median flow range where operations can have the greatest impact. For this thesis, differences in the median flow range are accepted as the objectives are primarily concerned with variability in high and low flow events.

Once adjustment values are computed for the reference period (1912-2005), they are applied individually to de-trended piControl, historic and future periods. The de-trending and individual application helps ensure that a potentially more extreme future (high or low) does not oversample adjustment values from the tail ends of the distribution. After adjustment values are applied, the linear trend is added back into the time series. Whereas the de-trending step can be important for scenarios with large trends, it will not have a substantial impact on scenarios that are more stationary, including the piControl periods.

Since QM of monthly data can alter the GCM's original simulated annual change signal, similar to an issue noted by Maurer and Pierce (2014), an additional step was considered to preserve the GCM's original annual change. The step was conceptualized as a secondary adjustment, applied uniformly across all months in a given year, to preserve the percent departure in GCM annual flow from the average annual GCM flow. After some testing, it was found that the differences were not large and this secondary adjustment was not performed.

#### 4.4 Streamflow Scenarios and Time Series Analysis

Bias corrected GCM streamflow scenarios from WATROUTE<sub>MOD</sub> provide simulations of streamflow under stationary conditions (piControl), historic conditions, and various future conditions as a result of specific RCPs. Since the bias correction does not account for all types of errors, and we accept that the streamflow scenarios are representative of more naturalized conditions, it is best to compare simulations from individual GCMs. In other words, results will be based on the comparison of GCM piControl and future periods to GCM historic periods (as opposed to comparison with observation). Where appropriate, results will include 1912-2005 observations to help illustrate residual bias.

Due to computational requirements and limited GCM simulations with piControl, historic and future periods, a subset of simulations listed in Table 6 were selected (marked with an asterisk; \*). Nine simulations from five GCMs were selected that contained adequate data in piControl and historic periods, and contained projections for RCP4.5 and RCP8.5 scenarios. Three simulations (run1, run2 and run3) from CanESM2 and MIROC5 are included to explore uncertainty in natural climate variability for historic and future periods. These runs are driven by the exact same GCM under the exact same atmospheric forcing but are initialized with a slightly different set of initial conditions, causing their climates to evolve differently over time. The subset of simulations selected for this analysis is similar to other studies such as Koirala *et al.* (2014) who only considered RCP4.5 and RCP8.5 and Prudhomme *et al.* (2014) who assessed five GCMs.

From a practical perspective, it is of interest to consider how the streamflow climatology may have been observed in different 94 year windows, outside of the 1912-2005 reference period. The use of longer samples is also consistent with literature which has recognized that

longer samples can aid the assessment of extreme events (Huziy *et al.*, 2013; Maloney *et al.*, 2014). Available data from 1361 to 2099 were split into eight quasi-equal segments of 91 to 94 years (Table 3). These segments include five piControl periods, one period that bridges piControl and historic periods, one 1912-2005 reference period, and one future period. The period that bridges piControl and historic periods is included to maximize use of available information but one should note that this period is likely discontinuous at the year 1861. Without knowledge of how modeling institutions used piControl simulations to initialize historic runs, it is not possible to consider a continuous time series at 1861. Furthermore, due to WATROUTE<sub>MOD</sub> spin-up, the first year of each CMIP5 period (1361, 1861 and 2006) is discarded from analysis. Future periods could be further divided and considered individually as there are two different realizations depending on the RCP.

Monthly mean flows generated by WATROUTE<sub>MOD</sub> are analyzed using different approaches across the aggregated NCRB domain. The NCRB as a whole is considered to better understand the entire system as one unit instead of individual sub-basins. Considering the aggregated NCRB also aligns with Manitoba Hydro's hydraulic generation system, which is primarily located at the downstream end of the Nelson River.

To compare average conditions, mean hydrographs of the annual cycle are calculated and presented to analyze streamflow climatology and extract information regarding monthly mean flows and timing of runoff. Annual average flows are also analyzed to better understand variability in the annual water budget. Monthly mean and annual mean streamflow are among the most common hydrological indicators used in climate change literature (Koshida *et al.*, 2015). For these indicators, changes in hydrologic variability are determined using the coefficient of variation (CV):

$$CV_Q = \frac{\sigma_Q}{\mu_Q} \quad (11)$$

where  $CV$  is coefficient of variation,  $\sigma$  is the standard deviation and  $\mu$  is the mean.  $Q$  may be monthly streamflow or annual streamflow.

$CV$  has been used in past studies to compare model agreement and spread in multi-model ensembles (Haddeland *et al.*, 2011; Hagemann *et al.*, 2013; Koirala *et al.*, 2014) but is applied in this thesis to determine changes in hydrologic variability was done in Gudmundsson, Tallaksen and Stahl (2011).  $CV$  calculated on mean annual streamflow provides information of inter-annual streamflow variability.  $CV$  calculated on mean monthly streamflow provides information on intra-annual flow variability.

To better understand extreme events in stationary and future climates, droughts and floods are also analyzed. Using time series of annual average flow, multi-annual drought severity ( $m^3$ ) and duration (years) are considered. Consistent with the theory of runs (Yevjevich, 1967), a drought begins when the annual average flow drops below a threshold and ends when the flow rises above the same threshold. For simplicity, the threshold is set as the mean 1912-2005 observed NCRB streamflow, common for all GCMs. A common threshold enables comparability among GCMs and since bias corrected mean flows are within  $90 m^3/s$  of the observed mean, the results are not overly sensitive to this decision. The maximum of monthly mean flows within a year are also used to examine floods. In large basins, the maximum of monthly mean flows are well correlated to instantaneous peak flows (Milly *et al.*, 2002), providing a suitable indicator for changes in flood events.

The statistical tests and methods presented in Koirala *et al.* (2014) are applied in this thesis to quantify the significance of change and assign confidence to projected change. A brief summary is provided below with further details found in Koirala *et al.* (2014). The method combines the two-sample Mann-Whitney-Wilcoxon U (MW-U) non-parametric rank-based test

with bootstrap resampling. The two samples are comprised of historic, piControl, and future streamflow scenarios. One sample always comes from a historic simulation while the other sample comes from either a piControl or future simulation. GCMs with one historic simulation (run1) produce six piControl pairs and two future pairs. GCMs with three historic simulations (run1, run2 and run3) produce 18 piControl pairs and six future pairs. For the five GCMs assessed, a total of 54 piControl pairs and 18 future pairs are tested for statistically significant differences between historic and stationary conditions (piControl) and differences between historic and future projections (RCP4.5 and RCP8.5). These tests assume that all simulations including both future scenarios are equally plausible. The MW-U tests the null hypothesis that both samples have equal means. Therefore, the results inform on changes in mean annual flow, mean drought severity, mean drought duration and mean of maximum monthly flow. In Koirala *et al.*, (2014) bootstrap resampling is used to filter out effects of outliers in the studies' relatively short 30 year periods. Although the 91-94 year periods used in this thesis provide a larger sample, the sample size of multi-year droughts is less than 91-94, therefore bootstrap resampling can be beneficial. In addition, bootstrapping is good practise for statistical analysis of hydrologic data (Kundzewicz and Robson, 2004) since it does not require assumptions about the underlying statistical distribution. Bootstrap resampling based on random number generation with replacement is therefore applied consistently to all MW-U tests.

The procedure to test statistical significance is briefly described as follows. A pair of data (e.g., CanESM2 run1 historic and CanESM2 run1 RCP8.5) is used to compute an original MW-U test statistic ( $MW-U_0$ ). Ten thousand equal-length bootstrap samples are generated for each the historic and future data (normalized to the original sample as in Koirala *et al.*, 2014 supplementary information) and 10,000 MW-U test statistics are generated ( $MW-U_1$  to  $MW-$

$U_{10,000}$ ). The bootstrapped test statistics are then sorted in ascending order. The rank of  $MW-U_0$  is then used to determine the non-exceedance probability using a plotting position that is in accordance with Cunnane (1978) recommendations:

$$p_0 = \frac{R_0 - 0.4}{N_b + 0.2} \quad (12)$$

where  $p_0$  is the non-exceedance probability,  $R_0$  is the rank of  $MW-U_0$  among bootstrap resampled  $MW-U$ s and  $N_b$  is the number of bootstrap samples (10,000 in this case).

At the two-sided 5% significance level, a non-exceedance probability less than 0.025 (greater than 0.975) rejects the null hypothesis and suggests a statistically significant increase (decrease). Non-exceedance probabilities between 0.025 and 0.975 accept the null hypothesis.

Similar to Koirala *et al.* (2014), the degree of consistency (DOC) is used to determine agreement among climate simulations regarding the direction of future climate change. In this thesis, DOC is also used to determine agreement among climate simulations regarding natural climate variability by comparing historic and piControl simulations. A strong DOC demonstrates consensus among model simulations regarding the direction of change. When coupled with statistically significant evidence of change, a strong DOC identifies robust changes which reflect greater confidence in results.

DOC is also used for characterizing differences in CV; however, no statistical evidence is presented to support these results. While a statistical test for different variances could have been applied, there is little literature currently available for detecting changes in second order statistics (Kundzewicz and Robson, 2000). One potential test is the two sample f-test, but this test assumes that the samples are normally distributed and can be sensitive to outliers. In the absence of a reliable and appropriate test for change in variability, DOC is presented for CV without statistical evidence. The differentiation is made by  $DOC_{STAT}$  for DOC supported by statistical

evidence and  $DOC_{NON-STAT}$  for DOC not supported by statistical evidence. Table 8 contains criteria for assigning DOC.

**Table 8 - Criteria for assigning degree of consistency with statistical evidence ( $DOC_{STAT}$ ) and degree of consistency without statistical evidence ( $DOC_{NON-STAT}$ )**

	<b>DOC</b>	<b>Criteria</b>
<b>DOC<sub>STAT</sub></b>	<b>Strong</b>	≥89% scenarios statistically significant and agree on direction of change
	<b>Moderate</b>	≥72% scenarios statistically significant and agree on direction of change
	<b>Weak</b>	≥56% scenarios statistically significant and agree on direction of change
	<b>Negligible</b>	<56% scenarios statistically significant and don't agree on direction if change
<b>DOC<sub>NON-STAT</sub></b>	<b>Strong</b>	≥89% scenarios agree on direction of change
	<b>Moderate</b>	≥72% scenarios agree on direction of change
	<b>Weak</b>	≥56% scenarios agree on direction of change
	<b>Negligible</b>	<56% agree on direction of change

The DOC criteria presented as percentages depend on the number of pairs being tested. For future projections (18 pairs total), the percentages correspond to 16 pairs for strong DOC, 13 pairs for moderate DOC and 10 pairs for weak and negligible DOC. For natural climate variability (54 pairs total), the percentages correspond to 48 scenarios for strong DOC, 39 scenarios for moderate DOC and 30 scenarios for weak or negligible DOC.

Other qualitative assignments of evidence and agreement have been used in literature. The IPCC offers some guidance on this topic and promotes the use of summary terms, which are subjective. In Mastrandrea *et al.* (2010), qualitative terms such as “low”, “medium”, “high”, “limited” and “robust” are used to describe evidence and agreement to a specific conclusion. Mastrandrea *et al.* (2010) also suggests that a quantitative confidence level should be presented only when there is high agreement and robust evidence.

## CHAPTER 5

### Results and Discussion

This chapter presents the results obtained following the methods and datasets described in Chapter 4. As with Chapter 4, this chapter is organized by sections according to thesis objectives.

#### **5.1 Evaluation of Climate Model Skill**

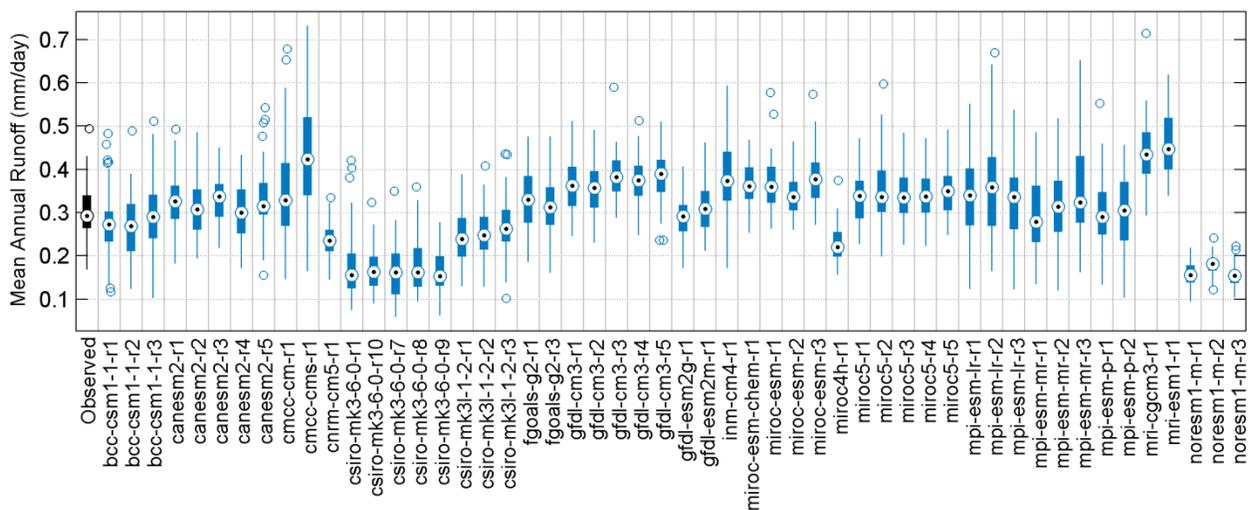
GCM skill evaluation provides the foundation for interpretation of subsequent results. Of the 53 available GCM simulations listed in Table 6, one simulation is discarded from the analysis (and is not shown). Daily runoff data from BNU-ESM run1 contains values that were orders of magnitude larger than other simulations, and much larger than realistically expected (>300 mm/day). While this runoff could be model error, it could also be an error in the way the model's daily runoff data was archived since the monthly runoff data from the same simulation is more realistic. It is important to note that runoff errors in GCMs could result from more than just GCM LSM error. For example, runoff errors could be the result of poor representations of precipitation as discussed in Chapter 2 (Sections 2.2.2.1 and 2.2.2.3).

Table 9 shows observed (LTFD) values to which GCM performance is compared and Figure 6 to Figure 9 illustrate GCM skill. Figure 6 illustrates how GCM simulations perform with respect to observed mean runoff and runoff variability in the NCRB. Annual average runoff over the 1950-2005 period are shown. Generally, simulations from the same GCM have similar characteristics with differences attributed to natural climate variability. Figure 6 helps identify models that produce: relatively higher (e.g., CMCC-CMS) and lower (e.g., CSIRO-Mk3.6.0 and NorESM1-m) mean annual runoff, greater interannual variability (e.g., CMCC-CMS), and less interannual variability (e.g., NorESM1-m). Some performance aspects of certain GCMs can be

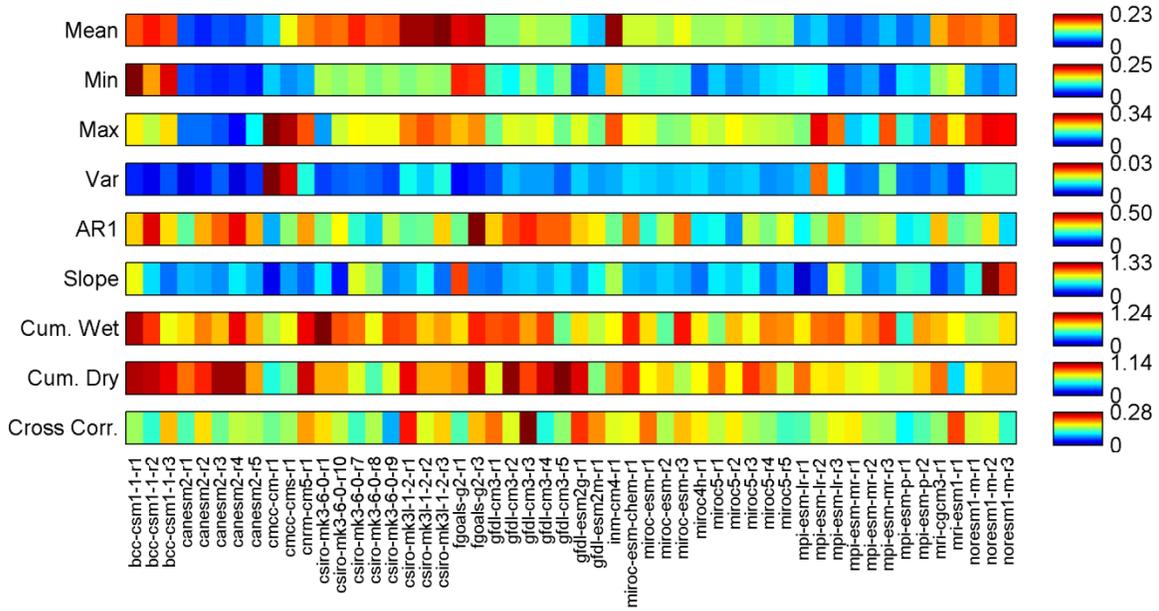
attributed to their LSM. For example, models such as BCC-CSM1.1, FGOALS-g2 and NorESM1-m whose land surface models are based on Community Land Model (CLM; Table 5) may produce lower runoff since CLM allows runoff to be negative (Oleson *et al.*, 2004). This issue was discussed in Chapter 4 (Section 4.2) of this thesis. Although Alkama *et al.* (2013) noted that BCC-CSM1.1 largely underestimated global runoff, its performance seems reasonable with respect to mean annual observed runoff in the NCRB.

**Table 9 - 1950-2005 observed (LTFD) values for skill evaluation performance metrics.**

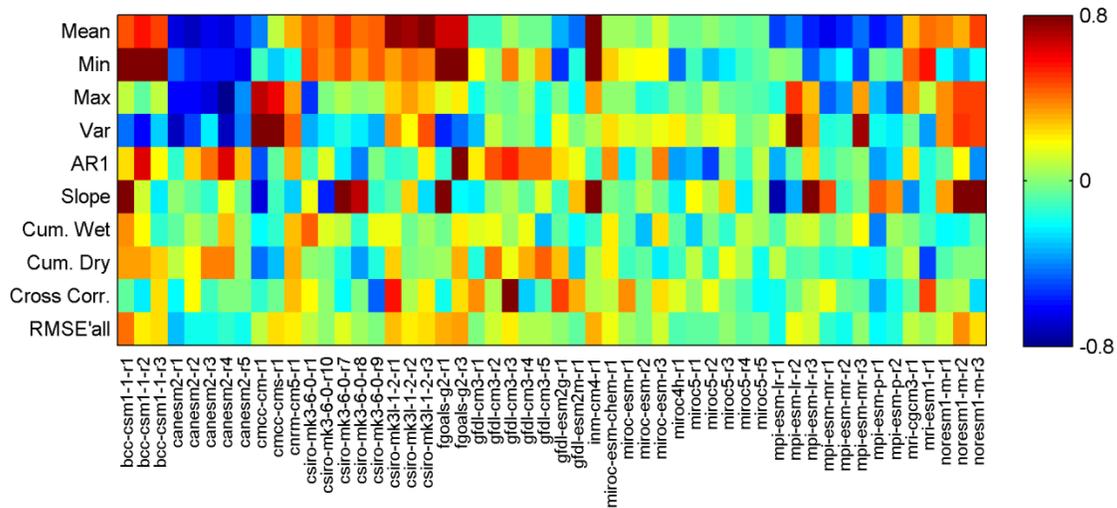
Metric (units)	UCR	LNR	SRB	PIAO	WRB	NCRB	
Mean (mm/day)	0.334	0.401	0.216	0.179	0.597	0.300	
Min (mm/day)	0.218	0.161	0.109	0.036	0.292	0.169	
Max (mm/day)	0.536	0.801	0.387	0.344	0.902	0.494	
Var (mm/day) <sup>2</sup>	0.004	0.016	0.004	0.007	0.026	0.004	
AR1	0.454	0.300	0.414	0.310	0.215	0.326	
Slope (% of mean)	-0.424	0.101	-0.672	0.067	-0.012	-0.204	
Cum. Wet (years)	3.625	2.077	2.273	2.273	2.636	2.600	
Cum. Dry (years)	3.375	2.417	3.100	2.818	2.455	3.000	
Cross Correlation	UCR	1.000	0.536	0.381	0.229	0.142	0.576
	LNR	-	1.000	0.257	0.404	0.444	0.684
	SRB	-	-	1.000	0.385	0.291	0.622
	PIAO	-	-	-	1.000	0.637	0.829
	WRB	-	-	-	-	1.000	0.772
	NCRB	-	-	-	-	-	1.000



**Figure 6 - GCM Skill in reproducing annual runoff volume and variability relative to the observed (LTFD) record for the NCRB (1950-2005)**



**Figure 7 - RMSE portrait diagram of GCM skill for 1950-2005. Additional information on metrics and their units can be found in Table 4 and Table 9.**



**Figure 8 - Relative RMSE (RMSE') portrait diagram of GCM skill for 1950-2005. Additional information on metrics can be found in Table 4 and Table 9.**

Figure 7 illustrates RMSE for the 52 GCM simulations with respect to the nine metrics listed in Table 4. Whereas Figure 6 and Figure 7 contain similar information, RMSE in Figure 7 is spatially averaged such that the GCM's performance in each sub-basin and the NCRB as a whole is reflected. Therefore, Figure 6 might indicate that a simulation performs well across the

aggregated NCRB domain, but Figure 7 may suggest that GCM performance in individual sub-basins is not as good. An example of this behaviour is seen for BCC-CSM1erf.1 simulations where mean annual runoff performs well for the aggregated NCRB, but is largely overestimated in the SRB, and underestimated in the WRB and NRL, resulting in a poorer RMSE in Figure 7. Figure 8 illustrates similar information as Figure 7 except  $RMSE'$  and  $RMSE'_{ALL}$  is shown instead of RMSE, reflecting how individual GCMs perform relative to the median simulation ( $RMSE'$  of zero) within the ensemble of 52 GCM simulations. A great deal of information is presented in Figure 7 and Figure 8 but RMSE alone does not make the distinction of whether a simulation over or under predicts a given metric. Although this limitation is important for metrics such as slope, a RMSE value of (or near) zero is still preferred and therefore the metric is included in the analysis.

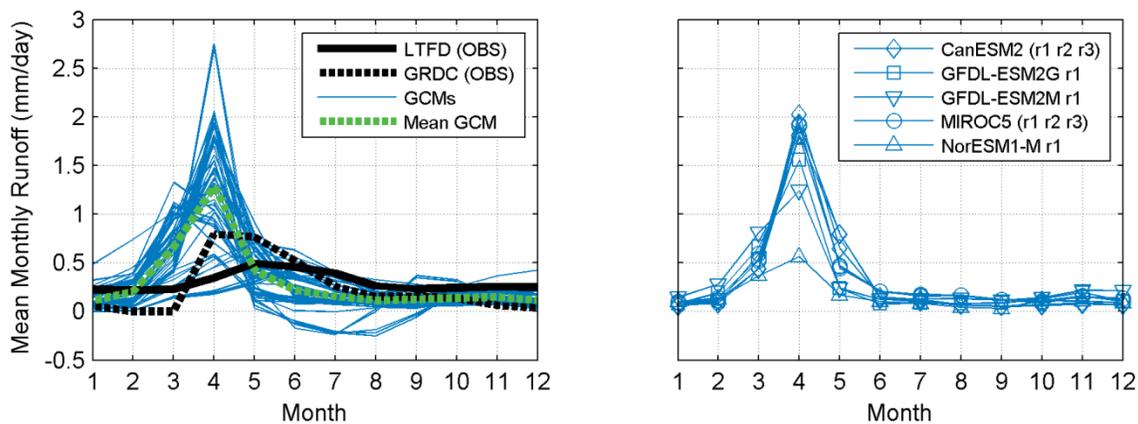
No one GCM stands out as having exceptional performance with respect to lag-1 autocorrelation (AR1) or the cross correlation (cross corr.) metrics. For AR1, RMSE ranges from 0.13 to 0.50 with a mean of 0.29; however, there is also considerable variability in observed LTFD estimates of AR1. To test sampling uncertainty, AR1 in the observed LTFD record is calculated for a 56 year moving window from 1912-2005 showing an AR1 range of 0.33 to 0.54 across the NCRB. The lowest AR1 of 0.33 is calculated in the 1950-2005 observed period which is the reference period in this skill evaluation. Using the same aggregated NCRB data, Akintuğ (2006) reported an AR1 value of 0.48 for the period of 1912-1998. Similar sampling uncertainty is possibly evident in cross corr. although this was not specifically tested. Colours shown on Figure 7 and Figure 8 suggest that the GCMs do not accurately capture the cum. wet and cum. dry metrics. However, y-axis bounds on Figure 7 show that RMSE is generally less than 1.24

years. In other words, most GCMs simulate average durations of wet and dry periods with an RMSE less than 1.24 years of the average duration of observed wet and dry periods.

Alkama *et al.* (2013) found that most GCMs simulated mean runoff reasonably well over North America. Similarly, simulations evaluated in this thesis indicate that most GCMs perform well in simulating mean annual runoff across the NCRB. Results above also show that GCMs are capable of reasonably simulating other annual runoff metrics such as minimum annual flow, maximum annual flow, variance and mean durations of cumulative wet and dry periods. Past studies such as Ault *et al.* (2014) have suggested that CMIP5 GCMs do not capture long term persistence in precipitation fields over certain areas such as the southwestern United States. Johnson *et al.* (2011) also assessed CMIP3 GCM skill in capturing long term persistence and showed that performance varied considerably among models. The use of 56 year climatologies in this thesis aligns with Johnson *et al.* (2011) who noted the importance of considering longer periods in assessing GCM skill. In this thesis, persistence is partially represented by the AR1 skill metric. And although there are considerable differences between observed and GCM simulated values, uncertainty in observations may play a significant role in the RMSE values. Furthermore, persistence of low flow conditions only addresses one component of drought risk. Another component is drought severity which can be important for short duration, very low flow events even in the absence of long term persistence. Overall, based on  $RMSE'_{ALL}$ , the subset of nine simulations selected for further analyses capture a range in model performance, including those that perform 10% worse (NorESM1-m r1) to 30% better (CanESM2 r1) than the median model.

Figure 9 compares monthly GCM runoff to monthly LTFD observed streamflow, illustrating the importance of using a routing model to capture the time lag between runoff

generation and streamflow. Monthly mean composite runoff fields from the Global Runoff Data Center (GRDC; Fekete, Vörösmarty and Grabs, 2002) are shown as supplementary. Since GRDC methodology uses streamflow records with as few as 12 years of observations, the period does not correspond to 1950-2005. In the GCMs and GRDC, normal peak runoff correlates to snow melt, typically occurring in March, April or May. In reality, snow melt might begin in March or April, but realization of the freshet (i.e., runoff from the melted snowpack) in the downstream streamflow record may not occur until May or later, depending on the basin. Figure 9 also illustrates GCM simulations that frequently generate negative runoff in the summer. This behaviour requires special handling in WATROUTE<sub>MOD</sub> and was discussed in Chapter 4 (Section 4.2).



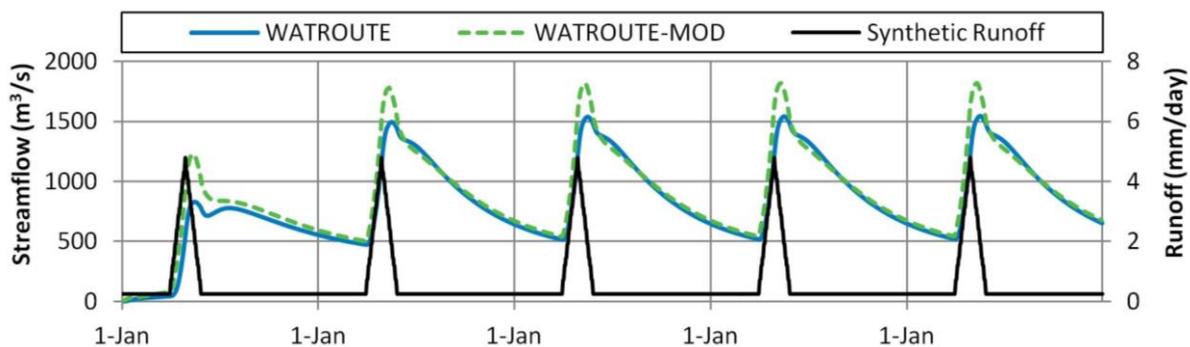
**Figure 9 - GCM skill - mean monthly runoff for the aggregated NCRB. Left panel shows LTFD observations (1950-2005), the ensemble of 52 GCM simulations (1950-2005) and GRDC estimated observations (multiple year ranges considered). Right panel shows subset of nine GCM simulations selected for further analysis.**

## 5.2 River and Lake Routing using WATROUTE<sub>MOD</sub>

Results and discussion for WATROUTE<sub>MOD</sub> output include a comparison with WATROUTE for the WRB and non-bias corrected (raw) output at key locations within the NCRB, corresponding to LTFD sub-basin outlets. Comparison between WATROUTE<sub>MOD</sub> and WATROUTE verifies that WATROUTE<sub>MOD</sub> performs in a similar manner as the original WATROUTE code. Although

it is possible to conduct a more comprehensive set of tests, including additional synthetic runoff events, this simple test illustrates how the two routing models respond to a single event and allows for assessment of the rising and falling limbs of the hydrographs. It is also important to note that this is a comparison of model results to model results, with no comparison to the truth (i.e., the observed record).

Although a comparison to observation would be ideal, there are challenges due to the absence of gridded runoff observations and imperfect knowledge of anthropogenic influence on observed records. An alternative test may have used WATFLOOD generated gridded runoff for a set of observed years (multiple years required to capture antecedent conditions) in place of a synthetic runoff event. In this case, both WATROUTE and WATROUTE<sub>MOD</sub> would be subject to errors in WATFLOOD's generated runoff which may artificially favor the performance of WATROUTE since WATFLOOD and WATROUTE were calibrated in parallel. The use of a synthetic runoff event with no comparison to observations reduces bias towards the calibrated model and allows for a more direct model to model comparison with differences that are easier to diagnose. A similar approach was used in Arora *et al.* (2001) to compare two routing methods in a hypothetical basin. Figure 10 shows the results from the test in the WRB.



**Figure 10 - WATROUTE<sub>MOD</sub> vs. WATROUTE at the Winnipeg River Basin outlet for a synthetic runoff event**

Hydrograph timing is similar in both routing models, especially for the receding limb. However, WATROUTE<sub>MOD</sub> responds with a slightly earlier rising limb that peaks 279 m<sup>3</sup>/s higher. WATROUTE produces a more attenuated hydrograph response with some other subtle differences resulting from model setup and parameterization. The addition of lakes and lake area when running WATROUTE<sub>MOD</sub> has one of the greatest impacts to the simulated hydrograph in the test basin. There are 58 lakes programmed into WATROUTE in the WRB, covering an area of approximately 12,600 km<sup>2</sup>. Originally, WATROUTE<sub>MOD</sub> was programmed with only five lakes (not including Lake St. Joseph) and during preliminary testing, peaked 1000 m<sup>3</sup>/s greater than the current version. Twenty-two additional lakes were added to the original WATROUTE<sub>MOD</sub> version; however some lakes were combined into one with the same lake outlet as they were too small to be discretized into their own grid at the 0.25° resolution. Together, lakes in WATROUTE<sub>MOD</sub> now account for 83.1% of the original lake area in the WRB WATROUTE model. This realization also led to increasing the number of lakes coded in the NCRB domain, which went from an original 15 to the 57 lake outlets embedded in the current version of the model. In addition, the rating curves used to route water through the lakes also varied substantially as a result of the increasing grid resolution in WATROUTE<sub>MOD</sub>, resulting in differences in flow attenuation and runoff travel time.

WATROUTE uses a finer grid resolution of approximately 0.10° x 0.07° and discretizes the shape and area of the WRB slightly differently (into 3039 grids). Differences in resolution and discretized area can affect both streamflow amplitude and volume. In a study of runoff model sensitivity to resolution, Arora *et al.* (2001) found that a coarser routing model overestimated high flows, helping to explain differences between WATROUTE<sub>MOD</sub> and WATROUTE. Differences in streamflow volume are partially an attribute of differences in the

discretized basin shape and area. The WRB in WATROUTE<sub>MOD</sub> is 2.5% larger than in WATROUTE (137,720 km<sup>2</sup> vs. 134,354 km<sup>2</sup>), and the domain is discretized into only 274 grids. Whereas one would also expect mean annual streamflow to be 2.5% greater (given the larger area), WATROUTE<sub>MOD</sub> produces mean annual streamflow that is 4.6% greater than WATROUTE. In this test, WATROUTE<sub>MOD</sub> produces mean annual streamflow within 0.08% of the expected value (964 m<sup>3</sup>/s for WATROUTE<sub>MOD</sub>; mean annual runoff of 0.605 mm/day over an area of 137,720 km<sup>2</sup>), while WATROUTE underestimates the expected runoff (941 m<sup>3</sup>/s; mean annual runoff of 0.605 mm/day over an area of 134,354 km<sup>2</sup>) by 1.9%. WATROUTE's underestimation of flow volume may be related to the internal treatment of grid cells, where only a fraction of their area contributes to the WRB or due to the numerical solution method. Fractionalized grid cells are not applied in WATROUTE<sub>MOD</sub>. Instead, grids either contribute completely to the basin or contribute nothing. Both WATROUTE<sub>MOD</sub> and WATROUTE use numerical solution schemes that allowed 20 iterations or convergence to within three percent. However, WATROUTE<sub>MOD</sub> uses a constant one hour time step whereas WATROUTE uses a dynamic time step to increase computational efficiency. It is possible that the one hour time step in WATROUTE<sub>MOD</sub> performs better; meeting a 3% convergence criterion before the 20 iteration limit and therefore provides more accurate streamflow estimates. However, this was not directly tested.

WATROUTE's parameterization, specifically channel geometry, slope, length, and Manning's n, differ from WATROUTE<sub>MOD</sub> partially due to different grid resolutions. This can impact streamflow attenuation. WATROUTE's channel bankfull area is based on user specified geomorphological relationships as a function of upstream drainage area. WATROUTE then applies user specified width to depth ratios to translate channel bankfull area to bankfull width

and bankfull depth. Bankfull area versus drainage area relationship parameters and width to depth ratios are set by river class. Of the 16 river classes in the WRB WATROUTE model, user specified width to depth ratios vary from 20:1 to 57:1 with a mean value of 25:1. In WATROUTE<sub>MOD</sub>, which uses bankfull channel width and depths directly from Andreadis, *et al.* (2013), width to depth ratios vary from 25:1 to 50:1 with a mean value of 40:1. In a secondary step, WATROUTE internally incorporates land cover information to adjust channel widths to correspond to the total water coverage in a grid. In grids where land cover classified as water exceed the area determined by channel width multiplied by channel length; channel width is adjusted to match the grid's land cover derived water coverage. This adjustment accounts for grids containing lakes that are too small to be discretized but are important for routing. However, since this adjustment only affects channel width and not depth, the resulting width to depth ratio is exaggerated. As a result, WATROUTE uses much larger bankfull areas and width to depth ratios than WATROUTE<sub>MOD</sub>. Table 10 details select physiographic differences between Manitoba Hydro's finer resolution WRB WATROUTE model and the WRB sub-basin in this thesis' NCRB WATROUTE<sub>MOD</sub> model.

Channel slope in WATROUTE ranges from 0.010% to 1.950% (mean of 0.115%); whereas in WATROUTE<sub>MOD</sub>, channel slope ranges from 0.100% to 0.944% (mean of 0.113%). For channel lengths, WATROUTE<sub>MOD</sub> is based on dominant river tracing (DRT; Wu *et al.*, 2012; see Chapter 4), where WATROUTE approximates channel length according to grid dimensions and drainage direction. For drainage directions flowing east-west or north-south, channel length is set equal to the grid length (7.48 km in WRB). For diagonal drainage directions (e.g., northeast to southwest) channel length is set to the grid length multiplied by 1.41. Due to

differences in spatial resolution, it is not meaningful to compare gridded channel lengths between WATROUTE and WATROUTE<sub>MOD</sub> and as such channel length is excluded from Table 10.

**Table 10 - Physiographic comparison of the Winnipeg River Sub-Basin WATROUTE model and the Winnipeg River Sub-Basin in WATROUTE<sub>MOD</sub>.**

Parameter	WATROUTE	WATROUTE <sub>MOD</sub>
Spatial Resolution	0.07°lat x 0.10°lon	0.25°lat x 0.25°lon
Minimum channel slope	0.00010	0.00100
Mean channel slope	0.00115	0.00110
Maximum channel slope	0.01950	0.00940
Minimum bankfull channel geometry (width/depth in m)	1.6 / 0.1*	5.5 / 0.2
Mean bankfull channel geometry (width/depth in m)	976.3 / 1.1*	55.7 / 1.1
Maximum bankfull channel geometry (width/depth in m)	10,575.6 / 14.9*	165.0 / 3.1
Minimum bankfull width to depth ratio	20:1*	25:1
Mean bankfull width to depth ratio	1673:1*	40:1
Maximum bankfull width to depth ratio	30,839:1*	53:1
Minimum Manning's n (in-channel/flood plain)	0.010 / 0.400	0.010 / 0.400
Mean Manning's n (in-channel/flood plain)	0.022 / 0.400	0.010 / 0.400
Maximum Manning's n (in-channel/flood plain)	0.042 / 0.400	0.010 / 0.400

*\*Values show WATFLOOD's internally corrected bankfull channel geometry.*

User specified Manning's n for WATROUTE varies from 0.01 (in channel) to 0.4 (overbank). Similar to WATROUTE's adjustment of bankfull channel width, in-channel Manning's n can also be internally adjusted to compensate for small lakes. The adjustment is a function of a grid's water coverage derived from land cover data as well as a user specified in-grid lake multiplier. In the WRB model, the in-grid lake multiplier ranges from 1 to 3 with an average value of 2 but this option was turned off for testing purposes. Other differences such as resulting drainage directions when the channel network is delineated across the discretized domain may also affect results.

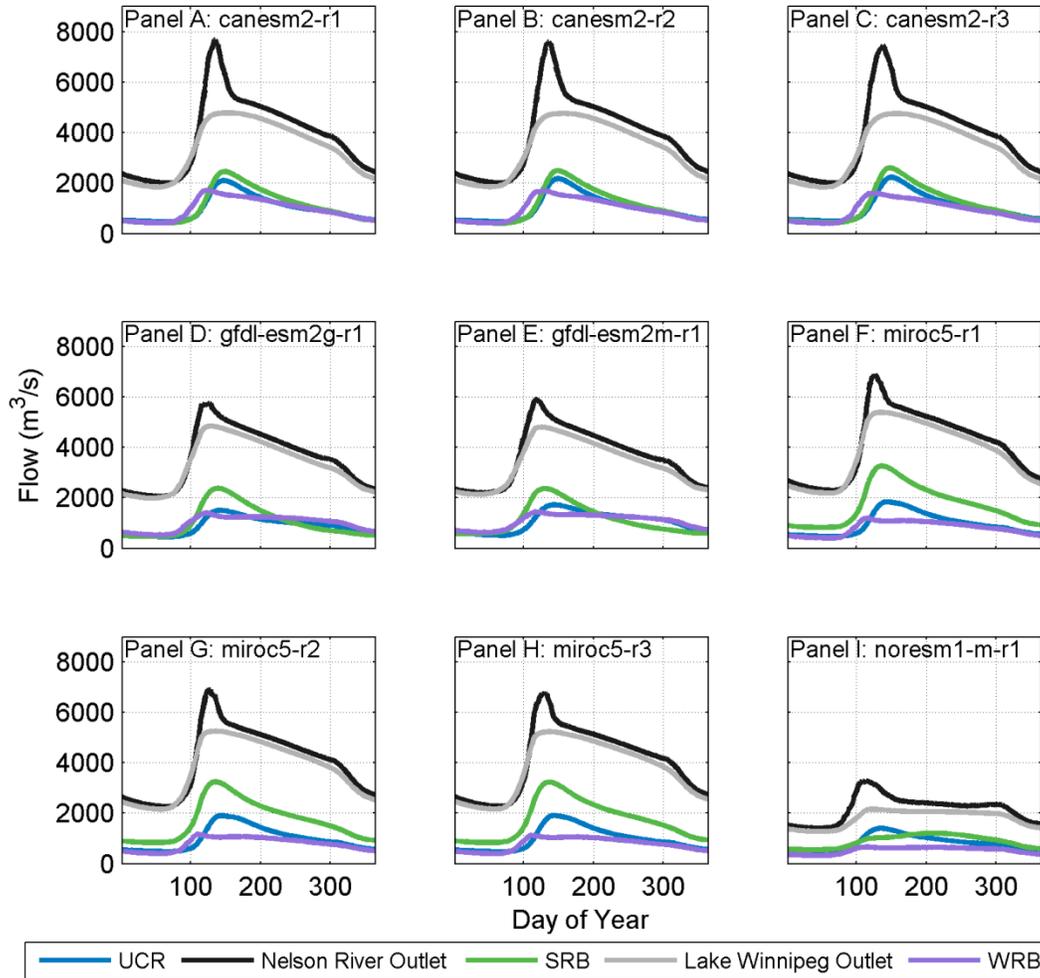
Despite the differences between WATROUTE<sub>MOD</sub> and WATROUTE, both models show skill in translating a runoff event that is uniformly distributed (spatially) within the river basin to streamflow at the basin outlet. Without observed gridded runoff data and a streamflow record with anthropogenic influence removed, it is not possible to determine which routing model performs more accurately. One consideration for future research may include further study and

testing to improve WATROUTE<sub>MOD</sub>. For the purposes of this thesis WATROUTE<sub>MOD</sub> facilitates interaction with GCM daily runoff data, serves as a routing model with a spatial domain covering the entire NCRB and performs adequately with respect to a finer resolution WATROUTE model. The adequate performance of WATROUTE<sub>MOD</sub> shown in this thesis supports other publications that found WATROUTE and modified versions of WATROUTE to perform well when coupled with GCMs (Arora *et al.*, 2001), RCMs (Poitras *et al.*, 2011; Huziy *et al.*, 2013; Clavet-Gaumont *et al.*, 2013) and LSMs (Soulis *et al.*, 2000; Pietroniro *et al.*, 2007).

Given the importance of the Lake St Joseph Diversion in the WRB (and other sub-basins in the NRCB), it is important to ensure that WATROUTE<sub>MOD</sub> internally handles this diversion in a realistic manner. To test this, WATROUTE<sub>MOD</sub> driven by GCM historic simulations are considered and the volume of diverted water is assessed. According to Lake of the Woods Control Board (2014), the mean annual diversion flow is 80 m<sup>3</sup>/s, typically varying from 70 m<sup>3</sup>/s in late winter and early spring to 90 m<sup>3</sup>/s in fall, with historic diversion flows peaking as high as 195 m<sup>3</sup>/s. In the nine historic GCM simulations, the mean annual hydrographs from 1861-2005 show mean annual diversion flows ranging from 64 m<sup>3</sup>/s to 100 m<sup>3</sup>/s, typically peaking in April or May. All GCM-driven WATROUTE<sub>MOD</sub> simulations show daily diversion flows that reach zero cubic meters per second, and all GCM-driven WATROUTE<sub>MOD</sub> simulations except NorESM1-m reach the maximum daily diversion outflow capacity of 245 m<sup>3</sup>/s. Despite some timing issues, the WATROUTE<sub>MOD</sub> Lake St. Joseph diversion seems to function in a reasonable manner and generally allocates the correct annual volume of water into the WRB. Timing issues are consistent with the overall unregulated methodology followed in this thesis. A more comprehensive rule set for Lake St. Joseph diversion that includes upper limits based on

downstream flow, downstream lake levels and season dependant preferences would likely improve the routing model's performance.

Annual average hydrographs (1861-2005 daily flows) based on WATROUTE<sub>MOD</sub> driven by the nine GCM simulations are shown in Figure 11. Various basin outlets (coloured lines) are included to compare flow within the NCRB, however, these basin flows differ from the LTFD sub-basin flows used as the observed record in this thesis. For example, Lake Winnipeg and Nelson River Outflow on Figure 11 include all upstream contributions instead of local flows corresponding to Lake Winnipeg PIAO or NRL from the LTFD observed record. For bias correction, WATROUTE<sub>MOD</sub> output is used to calculate local sub-basin flows to better match with the observed record. For example, Lake Winnipeg PIAO is calculated by subtracting SRB flow, WRB flow and back-routing Lake Winnipeg Outflow through Lake Winnipeg. Within the hydrographs, one will notice that CanESM2 produces the greatest amount of local runoff in NRL, as seen by the difference between Nelson River Outflow (black line) and Lake Winnipeg Outflow (grey line). One can also note that the SRB flows have a wet bias for most GCM simulations and that the NorESM1-m simulation is much drier and produces relatively flat hydrographs.

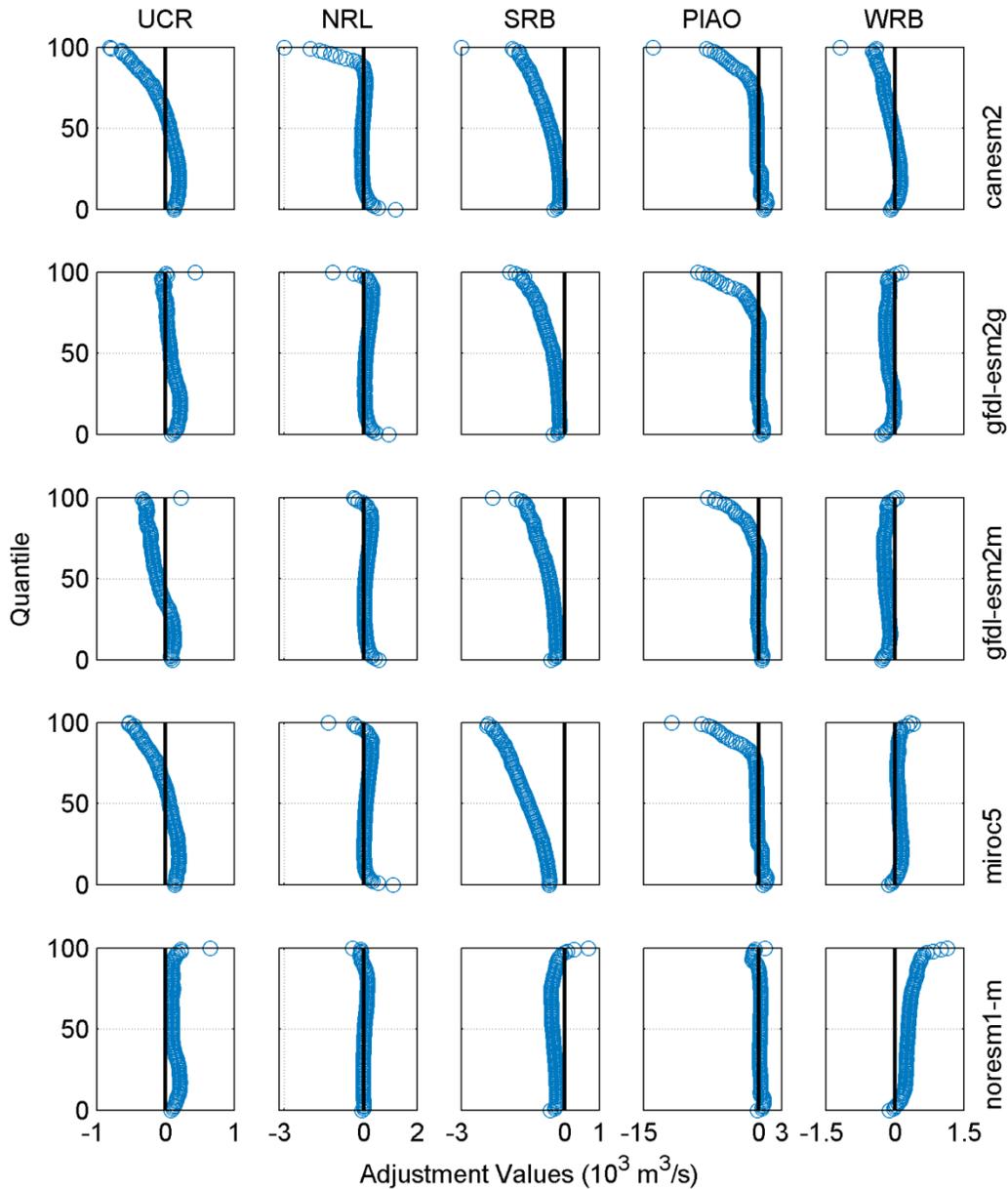


**Figure 11 - Uncorrected GCM-driven WATROUTE<sub>MOD</sub> mean annual hydrographs for CMIP5 historic period (1861-2005). Colours denote various locations in the NCRB.**

### **5.3 Bias Correction of GCM-driven WATROUTE<sub>MOD</sub> Streamflow**

Results from the bias correction step are summarized in plots of adjustment values (Figure 12). The plots illustrate the magnitude of adjustment, at each quantile, that are applied to monthly WATROUTE<sub>MOD</sub> streamflow after de-trending the time series and prior to re-trending. A set of adjustment values is derived for each sub-basin and GCM combination. For CanESM2 and MIROC5, three runs (run1 run2 and run3) are combined during the reference period of 1912-2005 to help reduce sampling uncertainty, resulting in a total of 25 sets of adjustment values

(five GCMs and five sub-basins). Going forward, each set of adjustment values will be referred to as a quantile map (QM).



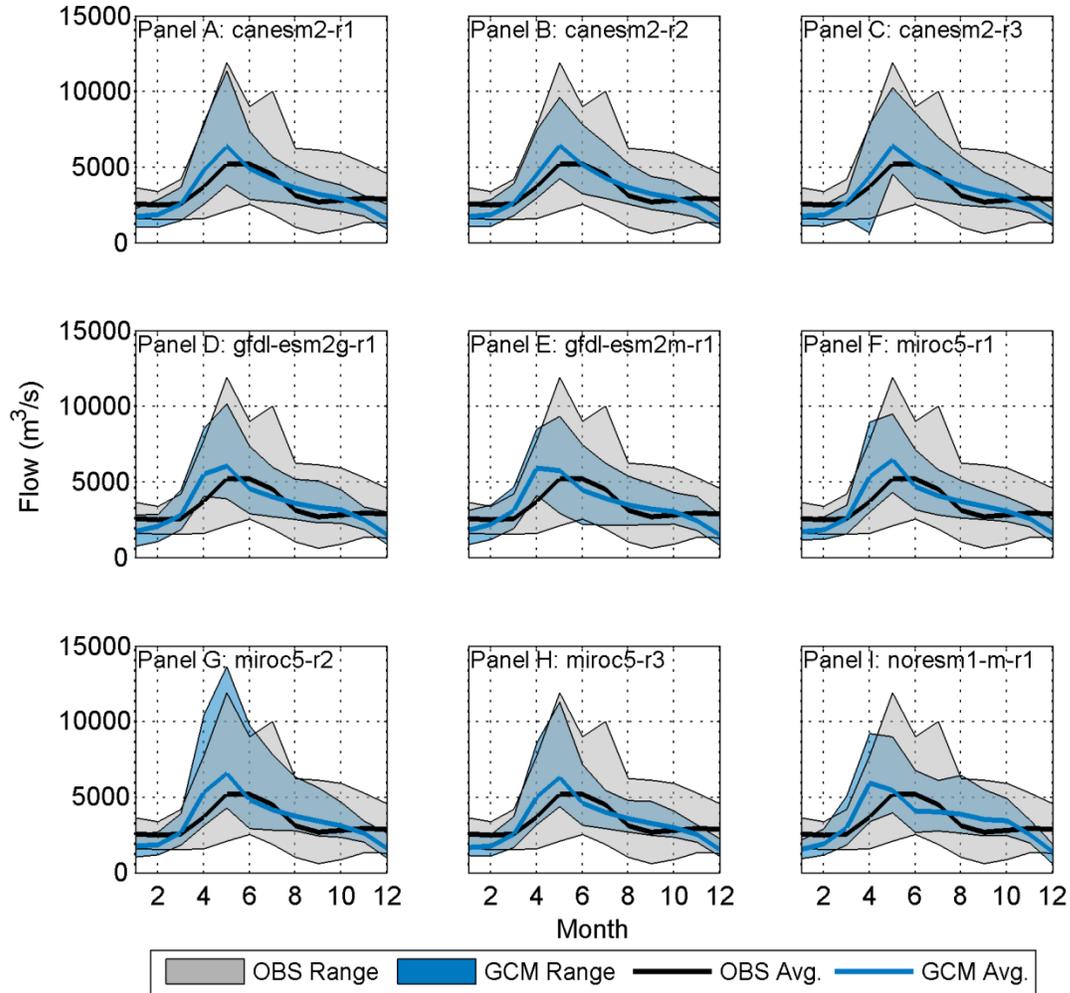
**Figure 12 - Bias correction adjustment values by NCRB sub-basin and GCM**

Within each QM there are 100 points. The y-axis represents quantiles from 0 to 100, and the x-axis shows the adjustments for a given quantile (Figure 12). If the GCM's monthly flow distribution matches the observed distribution perfectly (recall timing of flow is not considered), then all adjustment values within the QM would be zero. Negative values indicate the GCM

over-predicts a quantile, and positive values indicate the GCM under-predicts a quantile. Adjustment values that lie further away from zero indicate larger adjustments are required, suggesting the GCM runoff combined with WATROUTE<sub>MOD</sub> is not able to properly capture the given streamflow quantile. It is acknowledged that the bias correction procedure is accounting for errors in both the GCM runoff, and errors in WATROUTE<sub>MOD</sub>'s ability to simulate realistic flow response to runoff events.

Among the sub-basins, Lake Winnipeg PIAO requires application of the largest adjustment value, with most CGMs over-predicting monthly flows at higher quantiles. Part of this over-prediction is attributed to many small lakes within the sub-basin that are not considered in WATROUTE<sub>MOD</sub> as well as large snowpack accumulation due to positive bias in simulating winter precipitation (Sheffield *et al.*, 2013). For most GCMs and most quantiles, SRB flows are over-predicted, which is partly attributed here to reasons listed above for Lake Winnipeg PIAO and the absence of water withdrawals in WATROUTE<sub>MOD</sub>. For UCR, NRL and WRB the majority of quantiles show adjustment values near zero, with some exceptions at higher and lower quantiles. MIROC5 produces notably higher SRB flows in comparison to other GCMs (Figure 11 and Figure 12). With respect to high flows, most simulations show a mix of over and underestimation in the various sub-basins. CanESM2 however, shows consistent overestimation of high flows in all sub-basins, while NorESM1-m shows nearly consistent underestimation of high flows in all sub-basins. Overall, NorESM1-m produces monthly flow distributions closest to the observed record and, therefore, shows the smallest adjustment values. It is important to note that Nor-ESM1-m also produces the flattest hydrographs and has an overall dry bias for mean annual runoff.

Figure 13 illustrates the bias corrected monthly flows for the aggregated NCRB. Dark lines represent the 1912-2005 observed mean (black) and bias corrected WATROUTE<sub>MOD</sub> mean (blue), while shaded areas show the range. The mean bias corrected WATROUTE<sub>MOD</sub> flows generally peak earlier and produce lower winter flows, compared to flatter observed hydrographs that contain anthropogenic influence. Ranges in monthly flows are generally well captured.



**Figure 13 - Observed and bias corrected mean monthly aggregated NCRB streamflow (1912-2005) for individual GCM simulations**

In theory, a QM method which uses a larger number of quantiles could perfectly replicate the observed distribution. One must be weary, however, of over-fitting the data as is the case in some statistical distribution fitting studies (Caissie and El-Jabi, 1991). Furthermore, a more

detailed QM method could better match the observed hydrograph timing if monthly maps are used. This approach would lump the correction of additional bias sources (e.g., GCM timing of runoff generation, routing model errors, anthropogenic influence, etc.) with the correction of GCM runoff volume bias, adding another layer of complexity that can complicate interpretation of results. Bias correction of runoff timing is beyond the scope of this thesis which intends to correct for GCM runoff volume bias, alone, and allows for streamflow simulation of naturalized conditions. Naturalized conditions are preferred for this thesis given the long time frame (738 years) and unknowns about past and future anthropogenic effects.

The QM method applied in this thesis successfully reduces bias in the entire distribution of streamflow in NCRB sub-basins by implementing techniques developed for bias correction of other climatic variables. The success is evident through comparison of bias corrected hydrograph ranges in Figure 13 with adjustment values in Figure 12. Issues such as overestimation of high flows in CanESM2 and underestimation of high flows in NorESM1-m are resolved and the results from both GCMs are more comparable after QM. The QM method applied in this thesis also uses a 94 year baseline period which helps reduce sampling uncertainty that can result from use of shorter time periods. Bias corrected streamflow scenarios facilitate analysis of GCM output by increasing comparability among GCMs as well as comparability with observed records, which is important for meeting the objectives of this thesis and for practitioners utilizing the results. Building on previous work that applied bias correction techniques to non-traditional GCM variables (e.g., Wilke *et al.*, 2013; Cheng *et al.*, 2014), this thesis demonstrates the utility of simple methods for bias correcting variables other than traditional climatic variables such as temperature and precipitation. This thesis also builds on previous work by Koirala *et al.* (2014) which did not correct for bias in GCM derived streamflow scenarios. In line with many

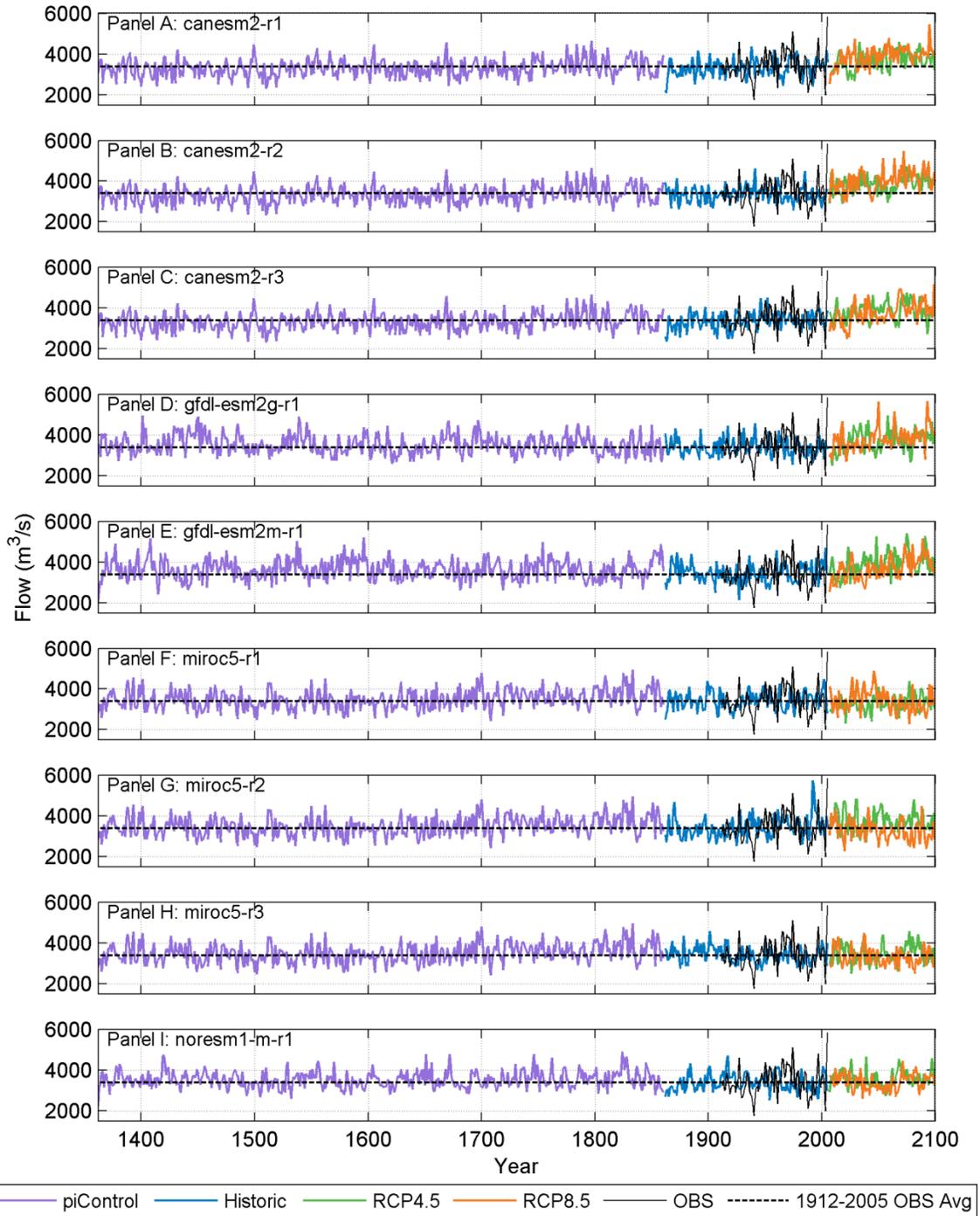
hydrologic studies (e.g., Eum, *et al.*, 2014; Pietroniro *et al.*, 2007; Haddeland *et al.*, 2011; Haddeland *et al.*, 2014; van Huijgevoort *et al.*, 2014; Koirala *et al.*, 2014), this thesis simulates naturalized conditions as opposed to regulated conditions. The QM method facilitates the approach of modeling naturalized conditions by grouping all months in each quantile map instead of bias correcting individual months. Although this approach does not directly account for issues with multi-year storage, it does account for intra-annual storage which is more common in the relatively small reservoirs in the NCRB.

#### **5.4 Streamflow Scenarios and Time Series Analysis**

Bias corrected WATROUTE<sub>MOD</sub> streamflow for all periods are shown on Figure 14 (annual time series) and Figure 15 (monthly mean hydrographs) as NCRB aggregated streamflow. Figure 16 through to Figure 19 present boxplots of mean annual flow, drought severity, drought duration and maximum monthly flow. Since droughts are defined as multi-year events ( $\geq 1$  year) and each analysis period is 91 to 94 years long, each boxplot represents a limited number of events (e.g., 14 to 20 drought events during the historic period). The small sample of drought events within a period can create odd distributions, which should be considered when interpreting boxplots. For all figures, the various colours represent the piControl (purple), historic (blue) and future (green as RCP4.5, and orange as RCP8.5) periods; with grey representing LTFD observations from 1912-2005. LTFD observations (OBS) are incorporated for comparison purposes, but results are drawn from GCM data only. The rationale for doing so is that comparison of a GCM's future or piControl simulation to the same GCM's historic simulation removes the reliance on the bias correction method to remove all bias in all hydrologic variables. Drawing results by comparing a GCM's future simulation to the same GCM's historic simulation is standard practice in climate model studies (e.g., van Huijgevoort *et al.*, 2014; Hageman *et al.*, 2013; Arnell and Gosling,

2013; Chen *et al.*, 2011; Prudhomme *et al.*, 2014; Shrestha *et al.*, 2011; Gao *et al.*, 2011; Sushama *et al.*, 2006; Poitras *et al.*, 2011). Results are presented in Figure 14 to Figure 19, Table 11 to Table 16, and discussed in the text. Summaries of future projected changes in mean and extreme conditions are included at the end of this section in Table 17 and Table 18.

Each panel within Figure 14 shows the annual average streamflow time series of aggregated NCRB flow from 1362-2099, with discontinuities at 1861 and 2006. The dotted black line represents the 1912-2005 observed mean annual flow (3,399 m<sup>3</sup>/s), which is used as the constant drought threshold. Although a time variable drought threshold (e.g., monthly thresholds in Burn and DeWit, 1996) captures additional information such as droughts lasting fractions of one year, there are complexities associated with using stationary sub-annual thresholds in the context of climate change. For example, van Huijgevoort *et al.* (2014) found that climate change induced shifts in hydrograph timing could lead to unintentional classification of sub-annual droughts, especially in cold climates. In a future scenario that produced an earlier spring freshet with earlier hydrograph recession, the authors showed that summer and fall months could be classified as droughts even if annual average streamflow increased. Although one could argue that these months are indeed experiencing drought (as defined in the study), water managers in systems with multi-month storage (such as the NCRB) might have options to manage the shift in timing without the resulting consequences of a typical multi-annual drought event.



**Figure 14 - Time series of bias corrected mean annual NCRB aggregated streamflow (1362-2099) for individual GCM simulations**

Overall, Figure 14 shows that GCM year-to-year variability is underestimated relative to variability in the observed record; however, the GCMs are capable of simulating periods of sustained low flow and sustained high flow conditions representative of the observed record.

Very little trend in the time series is apparent up until 2006, where CanESM2 and the GFDL GCMs appear to exhibit an increasing trend.

Coefficients of Variation (CV; Table 11) support the finding that GCMs underestimate year-to-year mean annual streamflow variability compared to observations which produce a CV of 0.20. Comparing each historic simulation to piControl simulations (54 pairs), 33 pairs (61.1%) show that mean annual streamflow variability in piControl periods exceed variability in the historic period, and 20 pairs (37.0%) show that mean annual streamflow variability in historic period exceeds variability in piControl periods. One pair (1.9%) shows the same variability in historic and piControl periods. Comparing each historic simulation to future projections (18 pairs), 11 pairs (61.1%) show that mean annual streamflow variability in the future exceeds variability in the historic period, and seven pairs (38.9%) show that mean annual streamflow variability in the historic period exceeds variability in future projections.

**Table 11 - Coefficients of Variation (CV) for mean annual NCRB streamflow in different periods. CV shown as  $*10^{-3}$ .**

Simulation	piControl						Historic			OBS	RCP4.5			RCP8.5		
	pi1	pi2	pi3	pi4	pi5	pi6	r1	r2	r3		r1	r2	r3	r1	r2	r3
CanESM2	124	123	115	120	134	131	124	113	116	200	128	106	119	121	131	145
GFDL ESM2g	138	125	131	128	127	130	121	n/a	n/a		142	n/a	n/a	143	n/a	n/a
GFDL ESM2m	157	126	139	112	138	139	133	n/a	n/a		145	n/a	n/a	153	n/a	n/a
MIROC5	130	128	124	144	104	150	125	155	112		121	132	126	159	148	131
NoRESM1-m	107	111	107	110	110	137	126	n/a	n/a		114	n/a	n/a	114	n/a	n/a

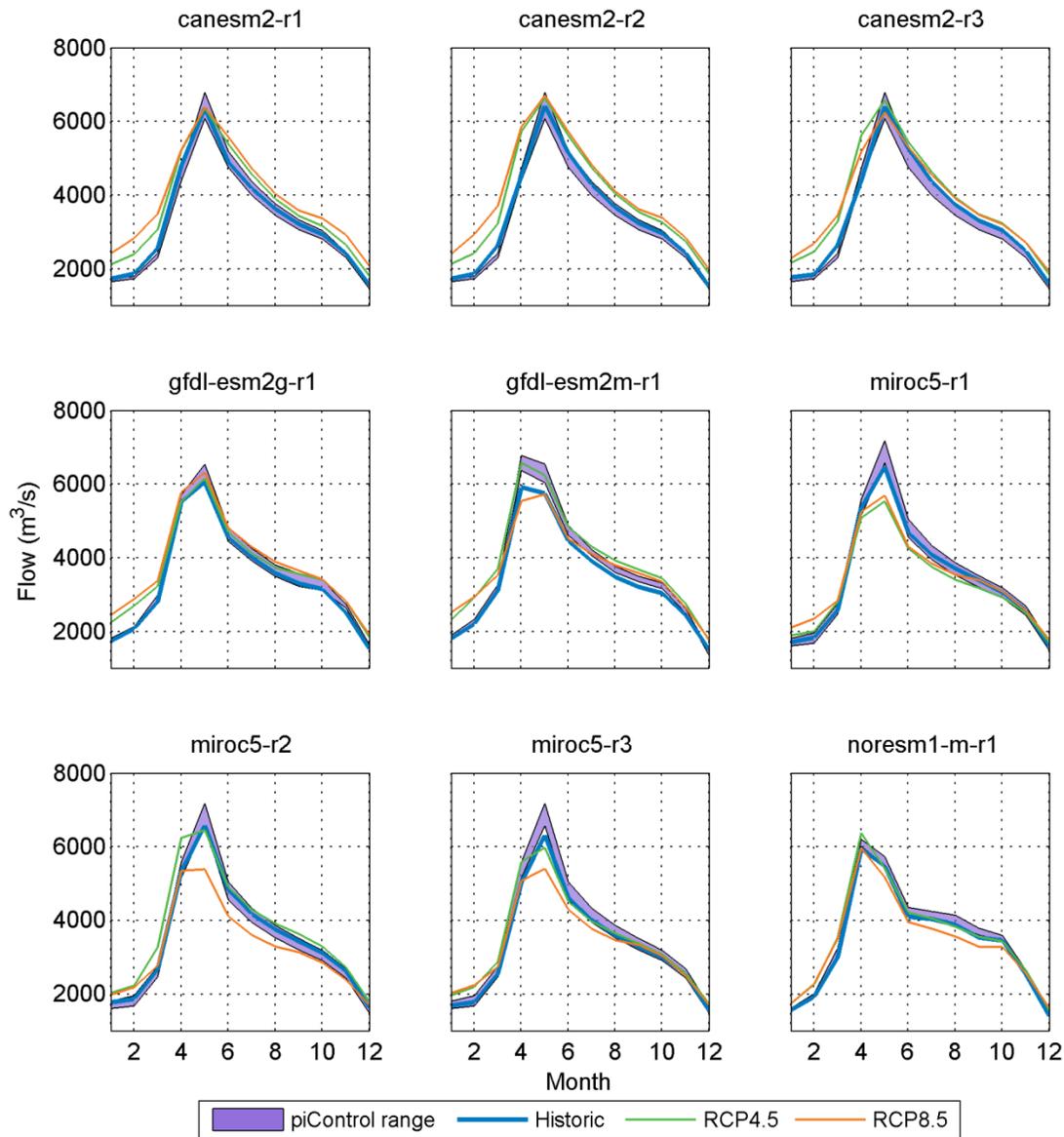
These results show weak evidence ( $DOC_{NON-STAT}$ ) that variability in the historic period underestimates potential variability in a stationary climate (natural climate variability) for mean annual streamflow. These results also show weak evidence ( $DOC_{NON-STAT}$ ) for future changes, that annual mean streamflow variability is projected to exceed variability in the historic period. It is important to note that these results simply compare the agreement on direction of change in

CV and are not supported by statistical evidence. One should note that there are no dramatic differences in mean annual streamflow variability for piControl, historic and future periods.

Mean monthly hydrographs of NCRB streamflow are shown in Figure 15 for each GCM simulation. To keep figures simple, the six piControl periods (Table 3) are shown as a shaded range and the observed record is omitted. Overall, the GCMs show tight ranges in monthly cycles for the piControl periods, which are similar to the historic period. Some differences are noted between piControl and historic periods, such as the GFDL-ESM2m historic simulation simulating lower mean monthly flows than the piControl range during April to October.

All GCMs agree that normal winter flows in the NCRB are projected to increase into the future, exceeding normal flows in both historic and piControl periods. Future increases in winter flows are a common finding in northern climates where the hydrological regime is largely influenced by snowmelt (Kundzewicz and Gerten, 2015). For example, Sushama *et al.* (2006), Poitras *et al.* (2011) and Shrestha *et al.* (2011) showed increased winter flows for the Nelson River and Churchill River basins using RCMs. Kumar *et al.* (2014) presented increased winter water availability for North America which is consistent with Arnell and Gosling (2013), Sperna Weiland *et al.* (2012a) and Hagemann *et al.* (2013) who showed that winter runoff in a majority of Canada was projected to increase. Arnell and Gosling (2013) also showed much higher consistency among 21 GCMs in winter runoff projections. Similar results were shown for the Colorado River Basin which is situated at lower latitudes but is also influenced by snowmelt (Gao *et al.*, 2011). CanESM2 and GFDL-ESM2g generally simulate futures with monthly mean flows greater than historic for both RCPs. GFDL-ESM2m, MIROC5 and NorESM1-m show future monthly flows that transition from wetter conditions in winter, to drier conditions in the spring and summer. Many of the studies that projected increased winter streamflow (noted

above) also projected decreased summer flows (e.g., Sperna Weiland *et al.*, 2012a; Hagemann *et al.*, 2013; Kumar *et al.*, 2014; Arnell and Gosling, 2013; Gao *et al.*, 2011). However, the agreement among GCM's was typically less for summer decreases compared to winter increases (Arnell and Gosling, 2013).



**Figure 15 - Bias corrected piControl, historic and future monthly mean hydrographs for aggregated NCRB streamflow. Time periods are provided in Table 3.**

No clear relationship seems to exist between RCP4.5 and RCP8.5 simulations. In some cases, RCP8.5 flows are very similar to RCP4.5 and in other cases the two RCPs produce

different mean monthly flow cycles. Differences also occur within a single model, as illustrated by the three runs of CanESM2 and MIROC5 that project RCP4.5 or RCP8.5 to behave differently depending on the run. These results suggest that increased carbon dioxide concentrations can alter the monthly mean flow patterns, but the degree of change may also be influenced by natural climate variability. Past research (Koirala *et al.*, 2014) showed stronger signals from higher emission scenarios such as RCP8.5 compared to lower emission scenarios such as RCP4.5. However, many studies only considered a single run from each GCM, potentially missing information regarding the role of natural climate variability in future climates. Differences between RCP4.5 and RCP8.5 may also be underemphasized in this thesis, which considers a 93 year future period of 2007-2099 selected for its relevance in continuous planning studies. Most studies considered 30 year future periods (e.g., 2070-2099) where greater differences between RCP4.5 and RCP8.5 exist (e.g., Figure 1).

Visual detection of changes to the timing of spring freshet requires substantial shifts in the monthly hydrographs. Figure 15 shows some indication of earlier freshet, as indicated by larger March and April flows accompanied by reduced or constant May flows; but it remains difficult to resolve these changes at the monthly scale. Assessment of bias corrected daily averaged hydrographs may have revealed a more pronounced and quantifiable change in freshet timing. However daily analysis is outside of this thesis' scope and it remains questionable whether coarse GCM data is suitable to resolve daily streamflow details, especially for extreme flows (Hirabayashi *et al.* (2008). Nohara *et al.* (2006) examined monthly hydrographs to assess the issue of timing and noted earlier spring freshets due to earlier snowmelts at high latitude basins. Hirabayashi *et al.* (2008) found that the month of maximum streamflow was projected to occur earlier at higher latitudes and cause a lower peak discharge. Sperna Weiland *et al.* (2012a)

also showed peak streamflow to occur approximately 0.5 months earlier in the NRCB. Explanation for this behaviour (earlier peak with reduced magnitude) can be attributed to snowpack evolution. In a warmer climate, snowpack accumulation may start later in the season and experience intermittent melting throughout the winter. Snowmelt may also begin earlier due to warmer temperatures before the snowpack has reached a higher water equivalent that would have been possible in a colder climate.

Coefficients of Variation (CV) (Table 12) suggest that the bias corrected streamflow scenarios overestimate monthly streamflow variability compared to observations producing a CV of 0.41. The overestimation may be a result of the unregulated simulated data compared to the regulated observed data which represents a flatter hydrograph under normal operating conditions which is also visible in Figure 13. Comparing each historic simulation to piControl simulations (54 pairs), 48 pairs (88.9%) show that mean monthly streamflow variability in piControl periods exceed variability in the historic period, and 6 pairs (11.1%) show that mean monthly streamflow variability in historic period exceeds variability in piControl periods. Comparing each historic simulation to future projections (18 pairs), all 18 pairs (100.0%) show that mean monthly streamflow variability in the historic period exceeds variability in the future period. These results show strong evidence ( $DOC_{NON-STAT}$ ) that variability in the historic period underestimates potential variability in a stationary climate (natural climate variability) for mean monthly streamflow. These results also show strong evidence ( $DOC_{NON-STAT}$ ) for future change, indicating that mean monthly streamflow variability is projected to decrease. It is important to note that these results simply compare the agreement on direction of change and are not supported by statistical evidence. These future projections are consistent with literature, which

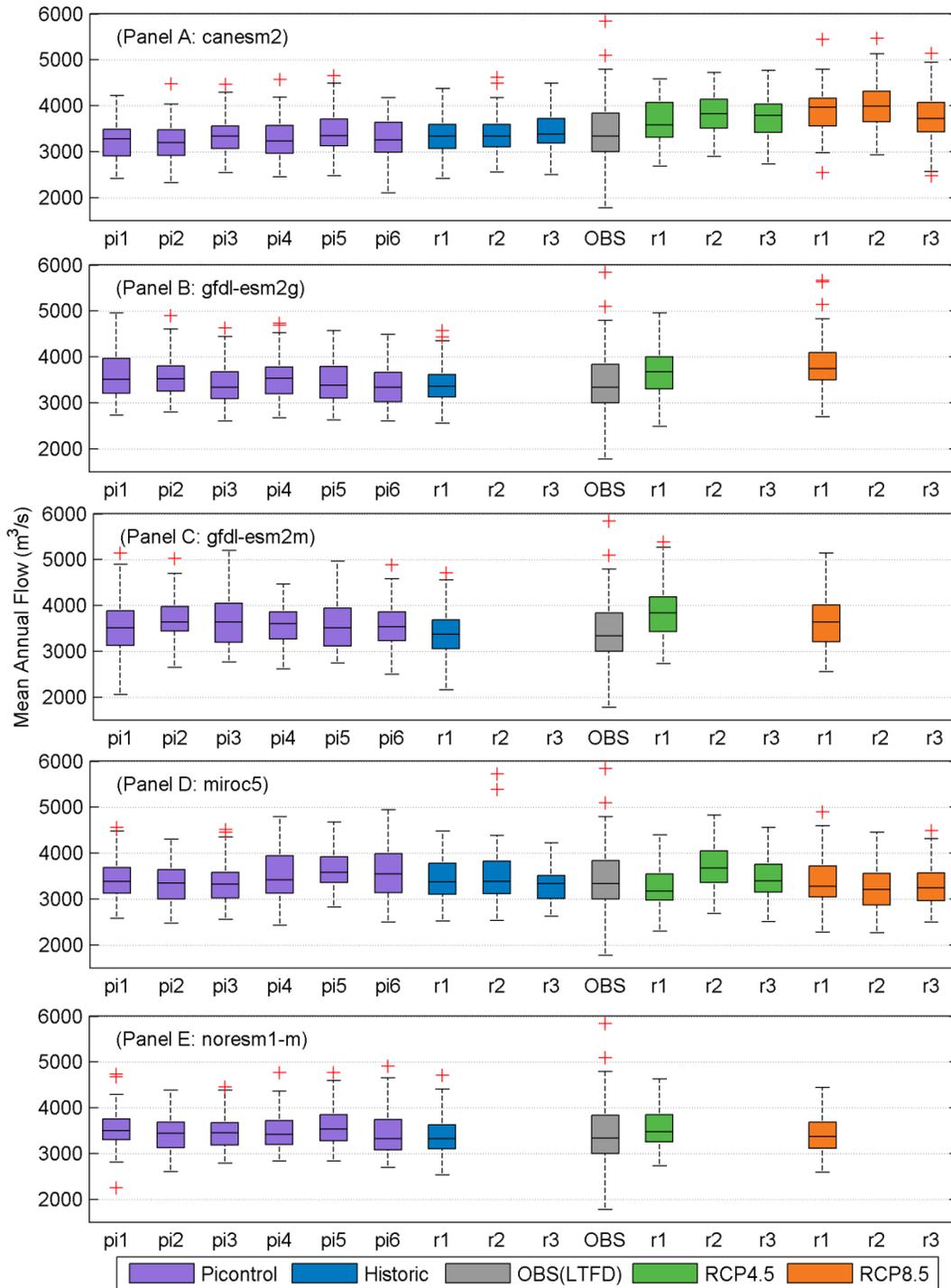
suggests decreasing CV in northern latitudes (Gudmundsson *et al.*, 2011) and increased low flows combined with decreased high flows presented in Koirala *et al.* (2014).

**Table 12 - Coefficients of Variation (CV) for mean monthly NCRB streamflow in different periods. CV shown as  $*10^{-3}$ .**

Simulation	piControl						Historic			OBS	RCP4.5			RCP8.5		
	pi1	pi2	pi3	pi4	pi5	pi6	r1	r2	r3		r1	r2	r3	r1	r2	r3
CanESM2	485	472	481	487	497	482	465	464	463	410	421	422	419	378	409	395
GFDL ESM2g	466	452	452	452	449	460	447	n/a	n/a		398	n/a	n/a	400	n/a	n/a
GFDL ESM2m	493	482	479	449	470	469	445	n/a	n/a		419	n/a	n/a	373	n/a	n/a
MIROC5	506	487	501	510	476	500	465	479	456		404	440	423	415	430	404
NoRESM1-m	430	431	428	439	435	449	442	n/a	n/a		419	n/a	n/a	402	n/a	n/a

Changes in water supply seasonality can have important implications that depend on the seasonality of water demand. From a hydropower perspective, increased winter flows may correspond well to periods of high energy demand that can be beneficial for hydropower operators. A reduction in intra-annual variability also flattens the hydrograph which might facilitate regulators who seek to anthropogenically create a flatter hydrograph (e.g., Kundzewicz and Gerten, 2015). However, lower summer flows projected by some simulations, might put additional strain on the agriculture industry with high summer water demands or fish needs which might have specific environmental flow requirements. It is also important to note that there are many other considerations (e.g., water licensing, changes in energy demand and other social factors) that are not accounted for in this thesis.

Figure 16 presents simulated mean annual streamflow distributions across the various time periods for the aggregated NCRB streamflow. Each panel (A through E) presents results from an individual GCM. GCM historic simulations reasonably simulate median values of mean annual flow, but underestimate the observed range (i.e., observed minimum of 1,776 m<sup>3</sup>/s to the observed maximum of 5,846 m<sup>3</sup>/s).



**Figure 16 - Mean annual streamflow box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels.**

Comparing piControl and historic simulations from individual GCMs, the median and range of mean annual streamflow is fairly stationary and does not vary substantially with time. However, there are exceptions that illustrate the potential impact of natural climate variability. For example, Figure 16 (Panel C) and Table 11 demonstrate that the pi1 simulation from GFDL-ESM2m produces greater variability and a wider range in mean annual streamflow (2,053 m<sup>3</sup>/s to 5,144 m<sup>3</sup>/s) compared to the other stationary piControl periods (pi2 to pi6). This simulation also contains the minimum single-year mean annual streamflow (lowest leg of the boxplot) among piControl, historic and future periods. Similarly, the historic run2 from MIROC5 (Figure 16 Panel D and Table 11) produces greater variability and a wider range in mean annual streamflow compared to other historic runs (historic r1 to historic r3). MIROC5 historic run2 contains the maximum single-year mean annual streamflow (highest point on the boxplot) among piControl, historic and future period. These results suggest that although a 90-94 year sampling period is typically well suited to characterize mean annual streamflow, there are exceptions even in a stationary climate that can produce notably higher or lower mean annual streamflow. The potential for sampling uncertainty has important implications for water resource planners interested in annual water supply, especially if observational records are short term.

The three future runs (available from CanESM2 and MIROC5) generally show consistent results within each GCM and RCP combination. Although there are some minor differences such as MIROC5 RCP4.5 run2 showing a slightly higher median than MIROC5 RCP4.5 run1 and run3 (Figure 16, Panel D), most distributions among the three runs are fairly similar. This result suggests that for the analysis of mean annual streamflow, one 93 year future run may reasonably capture a large portion of natural climate variability. However, since piControl and historic simulations can exhibit greater natural climate variability, further testing of multiple future runs

should be explored to support this finding. For comparison, Chen *et al.* (2011) used five runs from one GCM as part of an overall uncertainty assessment and found that natural climate variability was the greatest source of uncertainty for timing of flood events but ranked third for overall uncertainty, behind GCM selection and emission scenario. This display of natural climate variability underscores the importance of considering GCMs with multiple runs.

Table 13 presents non-exceedance probabilities for the two-sample bootstrapped MW-U test. Statistically significant changes (at the 5% level) are shown in bold font (red for decreasing flow, blue for increasing flow). Comparing each historic simulation to piControl simulations (54 pairs), 16 pairs (29.6%) shows statistically significant greater mean annual streamflow in piControl simulations compared to historic simulations. Eight pairs (14.8%) show statistically significant greater mean annual streamflow in the historic simulations compared to piControl simulations, and 30 pairs (55.6%) were statistically insignificant. Comparing each historic simulation to future projections (18 pairs), 13 pairs (72.2%) project statistically significant increases in future mean annual streamflow and two pairs (11.1%) project statistically significant decreases in the future. Three pairs (16.7%) were statistically insignificant.

**Table 13 - MW-U Non-exceedance probabilities for NCRB mean annual streamflow. Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations.**

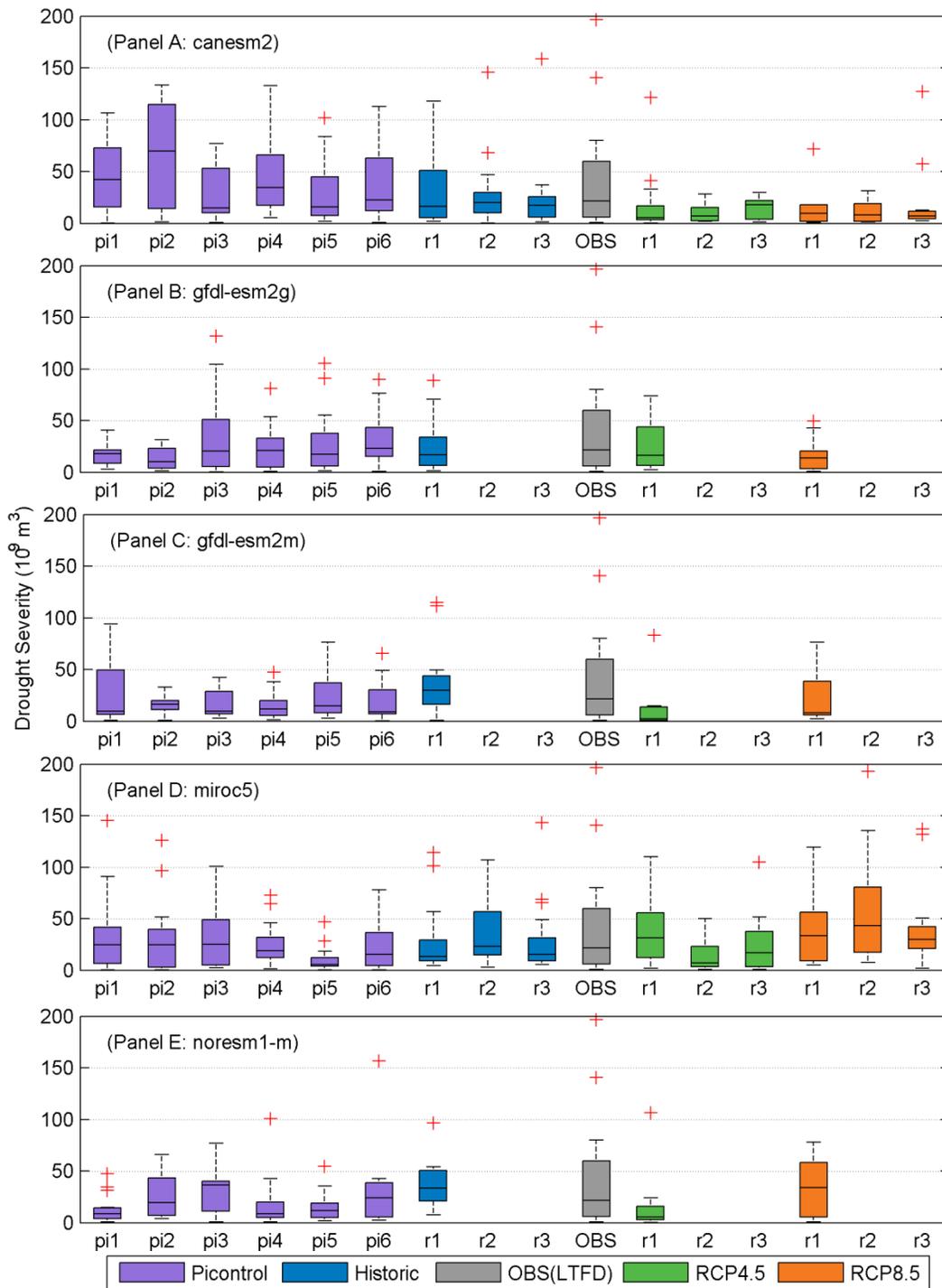
Simulation	Historic vs. piControl						Historic vs. Future					
	pi1	pi2	pi3	pi4	pi5	pi6	RCP4.5			RCP8.5		
							r1	r2	r3	r1	r2	r3
CanESM2 r1	0.951	<b>0.996</b>	0.582	0.932	0.102	0.715	<b>0.000</b>	n/a	n/a	<b>0.000</b>	n/a	n/a
CanESM2 r2	<b>0.985</b>	<b>0.999</b>	0.721	0.969	0.155	0.864	n/a	<b>0.000</b>	n/a	n/a	<b>0.000</b>	n/a
CanESM2 r3	<b>0.997</b>	<b>1.000</b>	0.886	<b>0.993</b>	0.359	0.937	n/a	n/a	<b>0.000</b>	n/a	n/a	<b>0.000</b>
GFDL ESM2g r1	<b>0.001</b>	<b>0.003</b>	0.429	<b>0.016</b>	0.160	0.732	<b>0.000</b>	n/a	n/a	<b>0.000</b>	n/a	n/a
GFDL ESM2m r1	<b>0.019</b>	<b>0.000</b>	<b>0.000</b>	<b>0.003</b>	<b>0.013</b>	<b>0.005</b>	<b>0.000</b>	n/a	n/a	<b>0.000</b>	n/a	n/a
MIROC5 r1	0.424	0.920	0.911	0.103	<b>0.000</b>	0.014	<b>0.996</b>	n/a	n/a	0.618	n/a	n/a
MIROC5 r2	0.752	<b>0.988</b>	<b>0.982</b>	0.351	<b>0.016</b>	0.123	n/a	<b>0.000</b>	n/a	n/a	<b>0.999</b>	n/a
MIROC5 r3	0.056	0.554	0.469	<b>0.010</b>	<b>0.000</b>	<b>0.001</b>	n/a	n/a	<b>0.022</b>	n/a	n/a	0.701
NoRESM1-m r1	<b>0.007</b>	0.331	0.193	0.061	<b>0.003</b>	0.216	<b>0.009</b>	n/a	n/a	0.576	n/a	n/a

Results show negligible evidence ( $\text{DOC}_{\text{STAT}}$ ) of systematic changes within piControl and historic periods for mean annual streamflow. Although most pairs are statistically insignificant, all GCMs contain at least one pair with a significant change. The presence of statistically significant changes suggest that even in a stationary climate, significant increases or decreases are possible among the 90-94 year periods. In the case of GFDL-ESM2m, the historic simulation is significantly drier than piControl and future simulations, and all pairs show statistically significant higher flows than the historic simulation. In the absence of multiple runs from GFDL-ESM2m with daily runoff data in the historic period, it is difficult to discern whether the drier historic period is a function of natural variability or a true climate signal when transitioning from a stationary climate (~280 ppm  $\text{CO}_2$  in piControl periods) to a climate with increasing  $\text{CO}_2$  (historic period). Since GFDL-ESM2m future periods (with further increases in  $\text{CO}_2$ ) indicate statistically significant wetter periods, it is likely that the dry historic period is part of a naturally induced “dip” in the streamflow record instead of an atmospheric  $\text{CO}_2$  concentration induced climate signal.

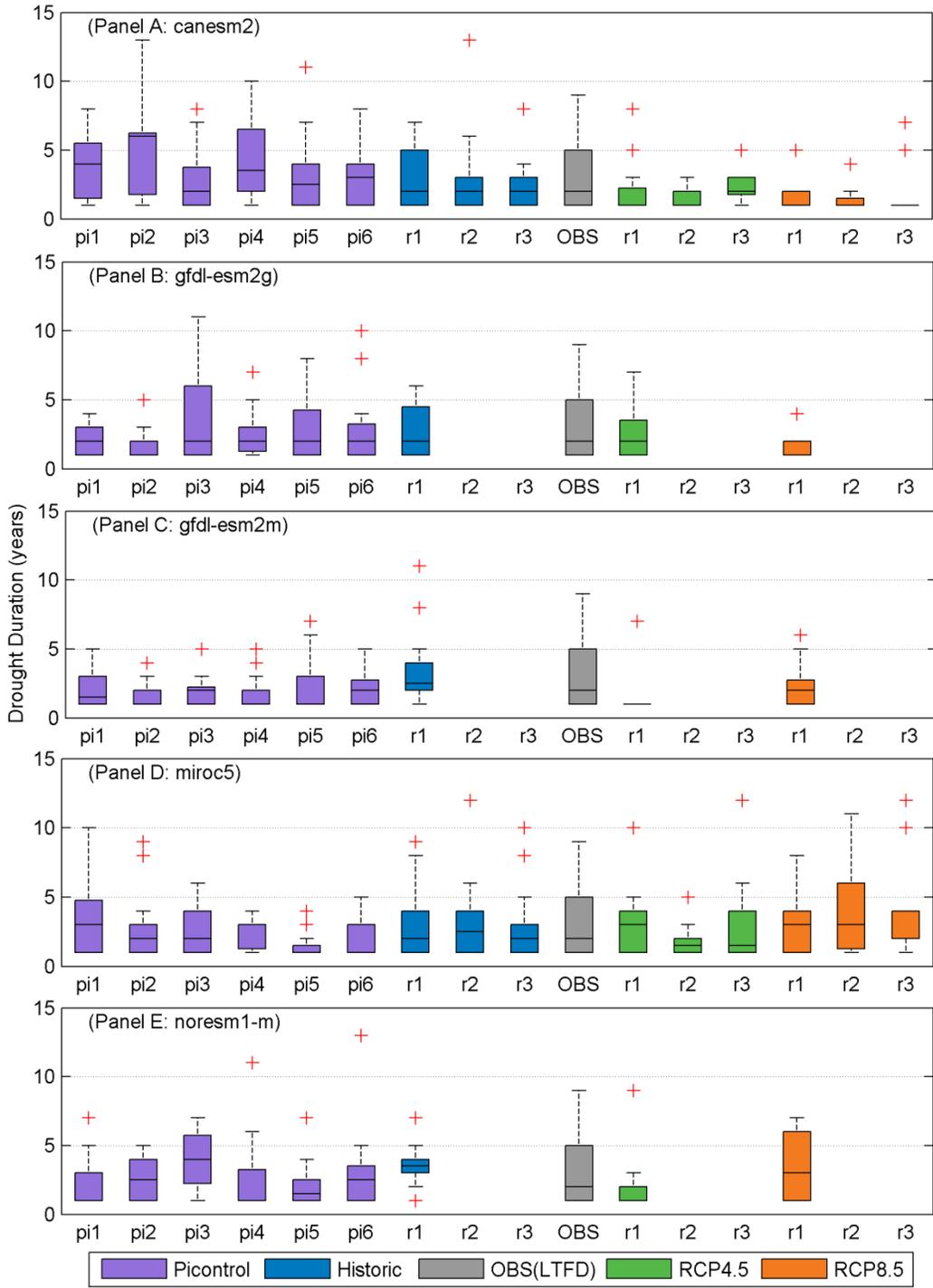
Results also indicate moderate evidence ( $\text{DOC}_{\text{STAT}}$ ) for future changes, indicating that mean annual streamflow is projected to increase in the future. These future projections are consistent with literature, suggesting increasing mean annual streamflow at northern latitudes (Arnell and Gosling, 2013; Koirala *et al.*, 2014). A summary of changes in all hydrologic variables is shown in Table 17. The average of all GCM simulations projects future mean annual streamflow to increase by 228  $\text{m}^3/\text{s}$ . Individual projections range from a decrease of 253  $\text{m}^3/\text{s}$  to an increase of 637  $\text{m}^3/\text{s}$ . Increasing mean annual streamflow in the future provides greater water supply on average. Arnell and Gosling (2013) found similarities between changes in mean annual runoff and changes in droughts (defined as the 10 year return period of minimum annual

runoff), however, the authors did not consider how multi-year drought severity or duration were projected to change, as done in this thesis.

Figure 17 and Figure 18 present multi-year drought severity and multi-year drought duration distributions across the various time periods from the aggregated NCRB streamflow simulations. GCM historic simulations reasonably simulate median drought severities and durations but underestimate the largest observed drought severity ( $196.7 \cdot 10^9 \text{ m}^3$ ). All GCMs display capability in simulating drought severity similar to the second largest observed drought ( $140.6 \cdot 10^9 \text{ m}^3$ ). All GCMs also display capability in simulating droughts longer than the longest observed drought of nine years which occurred from 1936-1944.



**Figure 17 - Drought severity box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels.**



**Figure 18 - Drought duration box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels.**

Unlike GCM simulations of mean annual streamflow, which are approximately stationary for piControl and historic periods, GCM simulated median drought severity and duration show greater variability. This variability is seen in Figure 17 and Figure 18, illustrating larger sampling uncertainty, even in a stationary climate. For example, Figure 17 and Figure 18 (Panel B) illustrate that GFDL-ESM2g pi1 and pi2 simulations represent wetter periods with less severe and shorter droughts followed by a drier pi3 period with more severe and longer droughts. Similarly, the CanESM2 pi2 simulation (Figure 17 and Figure 18, Panel A) represents a period with more severe median droughts compared to historic and future periods. However, the relatively wetter CanESM2 historic periods (less severe and shorter duration median droughts) are capable of producing severe and long droughts, exceeding the most severe drought in pi2, illustrated by the upper (red crosses) on the box plots. Results suggest that a 90-94 year sampling period is not particularly ideal for characterizing multi-year drought severity and duration. The potential for sampling uncertainty has important implications for water resource planners interested in determining the dependable flow, especially if observational records are short term.

Three future runs (available from CanESM2 and MIROC5; Panel A and Panel D of Figure 17 and Figure 18) echo the variability noted above, showing different drought distributions within each GCM and RCP combination. Median drought conditions and extreme drought conditions (upper legs and red crosses on boxplots) show variability from one future run to the next. This result suggests that for the analysis of future multi-year drought, it is important to consider GCMs with multiple runs in the future period and caution should be exercised when interpreting results from single-run models. For example, results presented in Hirabayashi *et al.* (2008) are based on a single GCM, emission scenario and run project 3.34 times more drought days (number of days below 10<sup>th</sup> percentile flow) in the Nelson River Basin for the period of

2071-2100 compared to 1901-2000. Based on results presented in this thesis, it is possible that another run of the same model and emission scenario would have produced conflicting results to those presented by Hirabayashi *et al.* (2008). Many studies now use multiple runs for analysis of hydrologic extremes, such as Maloney *et al.* (2014) who analyzed persistent drought.

Table 14 and Table 15 present non-exceedance probabilities for the two-sample bootstrapped MW-U test results. Statistically significant changes (assessed at the 5% level) are shown in bold font (red for more severe or longer droughts, blue for less severe or shorter droughts). Comparing each historic simulation to piControl simulations (54 pairs) for drought severity: 11 pairs (20.4%) show statistically significant lower mean drought severity in piControl simulations compared to historic simulations. Two pairs (3.7%) show statistically significant lower mean drought severity in the historic simulations compared to piControl simulations, and 41 pairs (75.9%) were statistically insignificant. Comparing each historic simulation to future projections (18 pairs), eight pairs (44.4%) project statistically significant decreases in future mean drought severity and one pair (5.6%) projects statistically significant increases in the future. Nine pairs (50.0%) were statistically insignificant.

Comparing each historic simulation to piControl simulations (54 pairs) for drought duration: 13 pairs (24.1%) show statistically significant lower mean drought duration in piControl simulations compared to historic simulations. One pair (3.7%) shows statistically significant lower mean drought duration in the historic simulations compared to piControl simulations, and 40 pairs (74.1%) were statistically insignificant. Comparing each historic simulation to future projections (18 pairs), eight pairs (44.4%) project statistically significant decreases in future mean drought severity and no pairs (0.0%) project statistically significant increases in the future. Ten pairs (55.6%) were statistically insignificant.

**Table 14 - MW-U Non-exceedance probabilities for NCRB multi-year drought severity.**  
**Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations.**

Simulation	Historic vs. piControl						Historic vs. Future					
	pi1	pi2	pi3	pi4	pi5	pi6	RCP4.5			RCP8.5		
							r1	r2	r3	r1	r2	r3
CanESM2 r1	0.141	0.063	0.813	0.061	0.658	0.227	<b>0.995</b>	n/a	n/a	<b>0.986</b>	n/a	n/a
CanESM2 r2	0.081	0.064	0.620	<b>0.015</b>	0.396	0.113	n/a	<b>1.000</b>	n/a	n/a	<b>0.999</b>	n/a
CanESM2 r3	0.029	0.058	0.214	<b>0.004</b>	0.138	0.042	n/a	n/a	<b>0.984</b>	n/a	n/a	0.621
GFDL ESM2g r1	0.845	<b>0.990</b>	0.124	0.621	0.248	0.129	0.467	n/a	n/a	0.893	n/a	n/a
GFDL ESM2m r1	0.854	<b>0.999</b>	<b>0.997</b>	<b>1.000</b>	0.925	<b>0.982</b>	<b>0.999</b>	n/a	n/a	0.958	n/a	n/a
MIROC5 r1	0.195	0.452	0.441	0.671	<b>1.000</b>	0.809	0.042	n/a	n/a	0.125	n/a	n/a
MIROC5 r2	0.312	0.734	0.701	0.926	<b>0.998</b>	0.867	n/a	<b>0.997</b>	n/a	n/a	0.026	n/a
MIROC5 r3	0.208	0.462	0.476	0.780	<b>1.000</b>	0.844	n/a	n/a	0.687	n/a	n/a	<b>0.019</b>
NoRESM1-m r1	<b>1.000</b>	0.900	0.749	<b>0.999</b>	<b>0.999</b>	0.843	<b>1.000</b>	n/a	n/a	0.509	n/a	n/a

**Table 15 - MW-U Non-exceedance probabilities for NCRB multi-year drought duration.**  
**Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations.**

Simulation	Historic vs. piControl						Historic vs. Future					
	pi1	pi2	pi3	pi4	pi5	pi6	RCP4.5			RCP8.5		
							r1	r2	r3	r1	r2	r3
CanESM2 r1	0.206	0.106	0.869	<b>0.018</b>	0.179	0.804	<b>0.994</b>	n/a	n/a	<b>0.994</b>	n/a	n/a
CanESM2 r2	0.232	0.041	0.883	0.025	0.157	0.154	n/a	<b>0.999</b>	n/a	n/a	<b>0.999</b>	n/a
CanESM2 r3	0.038	0.038	0.105	0.027	0.163	0.160	n/a	n/a	0.172	n/a	n/a	0.881
GFDL ESM2g r1	0.909	0.929	0.149	0.152	0.114	0.159	0.880	n/a	n/a	0.935	n/a	n/a
GFDL ESM2m r1	<b>0.996</b>	<b>1.000</b>	<b>0.999</b>	<b>0.998</b>	<b>0.996</b>	<b>0.998</b>	<b>1.000</b>	n/a	n/a	<b>0.997</b>	n/a	n/a
MIROC5 r1	0.178	0.898	0.872	0.851	<b>0.999</b>	0.913	0.139	n/a	n/a	0.150	n/a	n/a
MIROC5 r2	0.204	0.832	0.832	0.803	<b>0.989</b>	<b>0.986</b>	n/a	<b>0.984</b>	n/a	n/a	0.238	n/a
MIROC5 r3	0.174	0.907	0.884	0.892	<b>0.999</b>	0.919	n/a	n/a	0.906	n/a	n/a	0.181
NoRESM1-m r1	<b>0.983</b>	<b>0.988</b>	0.289	0.726	<b>0.989</b>	0.753	<b>0.992</b>	n/a	n/a	0.697	n/a	n/a

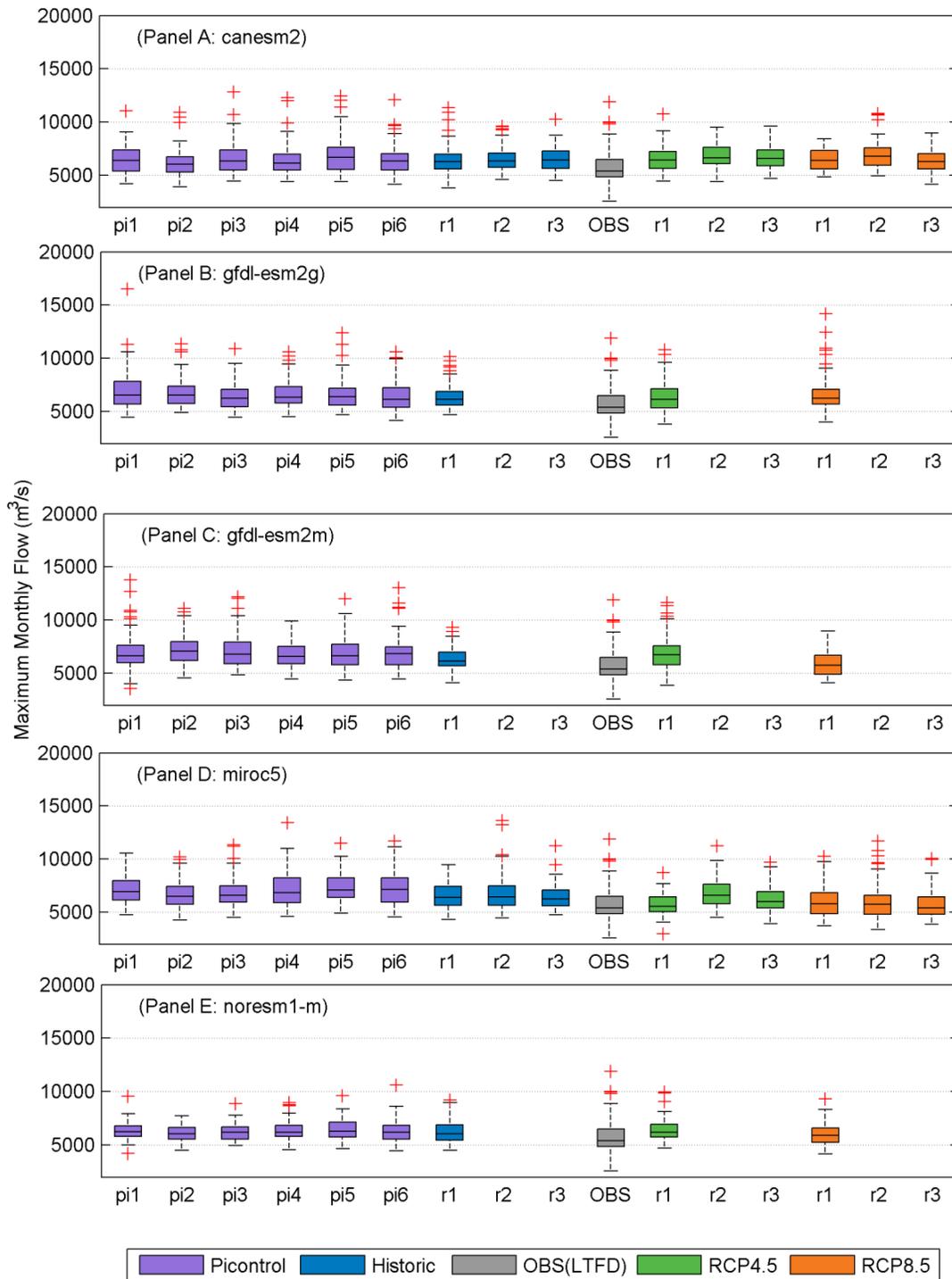
Results show negligible evidence ( $DOC_{STAT}$ ) of systematic changes within piControl and historic periods for mean drought severity and mean drought duration. Although a majority of pairs are statistically insignificant, most GCMs contain at least one pair with a significant change. The presence of statistically significant change suggests that even in a stationary climate, statistically significant increases or decreases are possible among the 90-94 year periods. Similar to results for mean annual streamflow, GFDL-ESM2m's historic simulation is significantly drier than many piControl and future simulations.

Similar to changes from piControl to historic periods, results also show negligible evidence ( $DOC_{STAT}$ ) that the mean drought severity and mean drought duration is projected to change in the future. Although drought indices vary from study to study, future projections presented in this thesis are generally consistent with the literature. Studies such as Sperna Weiland *et al.* (2012), Arnell and Gosling (2013) and Koirala *et al.* (2014) indicated a mean GCM tendency towards increased annual low flows in the future, but considerable spread among GCMs produced weak evidence of change. Although shorter term droughts (less than one year) were not considered in this thesis, other studies such as Poitras *et al.* (2011) showed increasing intra-annual low flow events (i.e., becoming wetter) in the Nelson River Basin, but these results corresponded to lower confidence.

A summary of mean changes (Table 17) and extreme changes (Table 18) is provided. The average of all GCM simulations (ensemble mean) projects future mean drought severity to decrease (i.e., become wetter) by  $6.5 \cdot 10^9 \text{ m}^3$  and mean drought duration to decrease (i.e., become shorter) by 0.7 years. Individual projections range from a mean drought severity decrease of  $27.3 \cdot 10^9 \text{ m}^3$  to an increase of  $23.4 \cdot 10^9 \text{ m}^3$  and a mean drought duration decrease of 2.0 years to an increase of 0.7 years. For changes in extremes, GCMs typically show the most severe and longest duration droughts (e.g., red crosses and highest box plots legs in Figure 17 and Figure 18) to decrease in the future, but this attribute was not statistically tested. The most extreme change in drought severity towards drier conditions is simulated by MIROC5 run2 under RCP8.5. This projection produces a future drought  $86.3 \cdot 10^9 \text{ m}^3$  (81%) more severe than the historic GCM simulation. Although this scenario represents the most extreme case in this thesis, the simulated drought is less severe than a stochastically simulated drought by Burn and DeWit (1996), which was three times worse than the observed drought of record. The extreme MIROC5

run2 RCP8.5 drought severity is associated with a shorter duration, suggesting that droughts can be more severe but this does not necessarily coincide with a longer duration. Although this sole projection should not be interpreted as truth, it is potentially useful as context for future drought risk discussion. Unlike paleo studies which have shown long periods of drought conditions (e.g., nearly a century of South Saskatchewan River drought in Sauchyn *et al.*, 2011) no multi-decadal droughts are simulated by the GCMs in this thesis for the NCRB. One reason for the lack of multi-decadal droughts might be the large NCRB compensating for dry conditions in one area with wetter conditions in another area, damping the effect of extreme drought. This behaviour is evident in MIROC5 run2 RCP8.5 where the SRB experiences more severe and longer duration droughts than the rest of the NCRB (not shown). Although this extreme simulation originates from RCP8.5, results do not show systematic differences between drought severities between RCPs, contradicting result from other studies such as Prudhomme *et al.* (2014) who suggested a systematic increase in drought severity with radiative forcing. Prudhomme *et al.* (2014) also discussed how some GCM feedbacks such as dynamic vegetation reduce projected drought risk.

Figure 18 presents simulated maximum monthly flow (flood) distributions across the various time periods for the aggregated NCRB streamflow. Each panel (A through E) presents results from an individual GCM. GCM historic simulations tend to overestimate the median and minimum floods. However, all GCMs are capable of simulating floods of similar magnitude to the greatest observed monthly flood (11,894 m<sup>3</sup>/s in May, 1974). It is important to reiterate that the floods presented here do not represent the streamflow at the NCRB outlet. Instead, the floods represent an aggregation of all naturalized inflows into the NCRB system. In reality, large flood peaks are attenuated by reservoirs and regulation as they make their way to the NCRB outlet.



**Figure 19 - Maximum monthly flow box plots for aggregated NCRB streamflow. GCM simulated piControl, historic and future periods plus LTFD observations shown on x-axis. Individual GCMs shown in separate panels.**

Comparing piControl and historic simulations from individual GCMs, Figure 19 shows that the median of maximum monthly streamflow is fairly stationary and does not vary substantially with time. However, there are exceptions, especially in extreme floods, illustrating the potential impact of natural climate variability. For example, Figure 19 (Panel B) demonstrates that the pi1 simulation from GFDL-ESM2g simulates an extreme flood (highest mark on the boxplot) of 16,492 m<sup>3</sup>/s; 4,102 m<sup>3</sup>/s greater than the next largest simulated flood in pi5. Similarly, pi1 in GFDL-ESM2m (Figure 19, Panel C) produces a wider range in monthly floods compared to other piControl periods. Results suggest that although a 90-94 year sampling period is typically well suited to characterize the median of maximum monthly streamflow, there are exceptions even in a stationary climate that can produce more extreme floods. The potential for sampling uncertainty has important implications for engineering design using maximum monthly streamflow as an input variable, especially if observational records are short term. However, many design approaches fit statistical distributions to observational data and then select a frequency based design value (*e.g.*, 1:100 year event). In these cases, fitting a statistical distribution and computing a 1:100 year event may smooth out sampling uncertainty and reduce the effect of differences in the absolute extremes (empirical maximums) that are presented in Figure 16. The impact on statistical distribution fitting is not within this thesis' scope.

The three future runs (available from CanESM2 and MIROC5) generally show consistent results within each GCM and RCP combination for median values of maximum monthly streamflow, with some variability in simulating extreme values. For example, MIROC5 run2 and run3 (Figure 16, Panel D) project future extreme floods of lower magnitude than historic extreme floods. However, run1 projects similar extreme floods in historic periods and under RCP4.5 and RCP8.5 projections. This suggests that it is important to consider GCMs with multiple runs in

the future period and that caution should be exercised when interpreting results from single-run models. For example, Hirabayashi *et al.* (2008); who used a single GCM, emission scenario and run; projected that the return period of a historic (1901-2000) defined 1:100 year flood event in the Nelson River Basin became a 1:133,000 year flood in 2071-2100. Based on results presented in this thesis, it is likely that another run of the same model and emission scenario could produce conflicting results to those presented in Hirabayashi *et al.* (2008). Multiple runs for analysis of hydrologic extreme events such as persistent drought were also used by Maloney *et al.* (2014). As for mean annual flow and multi-year droughts, the display of natural climate variability in maximum monthly streamflow underscores the importance of considering GCMs with multiple runs.

Table 16 presents non-exceedance probabilities for the two-sample bootstrapped MW-U test results. Statistically significant changes (assessed at the 5% level) are shown in bold font (red for decreasing flow, blue for increasing flow). Comparing each historic simulation to piControl simulations (54 pairs), 22 pairs (40.7%) show statistically significant greater mean of maximum monthly streamflow in piControl simulations compared to historic simulations. One pair (1.9%) shows statistically significant greater mean annual streamflow in the historic simulations compared to piControl simulations, and 31 pairs (57.4%) were statistically insignificant. Comparing each historic simulation to future projections (18 pairs): 3 pairs (16.7%) project statistically significant increases in future mean annual streamflow and five pairs (27.8%) project statistically significant decreases in the future. Ten pairs (55.6%) were statistically insignificant.

**Table 16 - MW-U Non-exceedance probabilities for NCRB maximum monthly streamflow. Bold font indicates statistically significant values where red (blue) denotes a drier (wetter) condition in piControl of future compared to historic simulations.**

Simulation	Historic vs. piControl						Historic vs. Future					
	pi1	pi2	pi3	pi4	pi5	pi6	RCP4.5			RCP8.5		
							r1	r2	r3	r1	r2	r3
CanESM2 r1	0.573	0.954	0.268	0.588	<b>0.021</b>	0.384	0.320	n/a	n/a	0.364	n/a	n/a
CanESM2 r2	0.673	<b>0.977</b>	0.379	0.677	0.031	0.501	n/a	<b>0.015</b>	n/a	n/a	<b>0.007</b>	n/a
CanESM2 r3	0.619	0.950	0.322	0.622	<b>0.021</b>	0.437	n/a	n/a	0.048	n/a	n/a	0.732
GFDL ESM2g r1	<b>0.002</b>	<b>0.008</b>	0.360	0.028	0.197	0.392	0.487	n/a	n/a	0.080	n/a	n/a
GFDL ESM2m r1	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.004</b>	<b>0.004</b>	<b>0.003</b>	<b>0.004</b>	n/a	n/a	<b>0.986</b>	n/a	n/a
MIROC5 r1	<b>0.005</b>	0.367	0.135	<b>0.002</b>	<b>0.000</b>	<b>0.003</b>	<b>1.000</b>	n/a	n/a	<b>0.999</b>	n/a	n/a
MIROC5 r2	0.030	0.626	0.352	<b>0.012</b>	<b>0.003</b>	<b>0.016</b>	n/a	0.366	n/a	n/a	<b>1.000</b>	n/a
MIROC5 r3	<b>0.000</b>	0.061	<b>0.010</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	n/a	n/a	0.889	n/a	n/a	<b>1.000</b>
NoRESM1-m r1	0.185	0.592	0.462	0.112	0.046	0.244	0.070	n/a	n/a	0.883	n/a	n/a

These results show negligible evidence ( $DOC_{STAT}$ ) of systematic changes within piControl and historic periods for the mean of maximum monthly streamflow. Although most pairs are statistically insignificant, four of five GCMs contain at least one pair with a significant change. The presence of statistically significant changes suggests that even in a stationary climate, statistically significant increases or decreases are possible among the 90-94 year periods.

Results also show a negligible evidence ( $DOC_{STAT}$ ) for future changes in the mean of maximum monthly streamflow. Although flood indices vary from study to study, the future projections presented in this thesis are generally consistent with the literature, showing that there is a lack of confidence in changes to future high flow events. Arnell and Gosling (2013) projected an increase in high flow events in most of Canada; however, Koirala *et al.* (2014) showed a decrease. These two studies along with others (Poitras *et al.*, 2011, Sushama *et al.*, 2006) reiterate the finding that there is generally low confidence in projections of future floods in the NCRB. Similar to Hirabayashi *et al.* (2008), results in this thesis do not indicate a consistent pattern between increases in mean annual flow and increases in flood events. As discussed in Whitfield (2012) who noted that a warming climate can have different impacts on different flood

driving mechanisms, uncertainty in projections of future floods may actually be the correct finding. This statement from Whitfield (2012) may be particularly relevant in the NCRB, which contains rivers with different flood driving characteristics and historic trends (Burn and Whitfield, 2015).

A summary of mean (Table 17) and extreme (Table 18) changes in all hydrologic variables is provided. The average of all GCM simulations (ensemble mean) projects the mean of future maximum monthly streamflow to decrease by 91 m<sup>3</sup>/s. Individual projections range from a mean decrease of 901 m<sup>3</sup>/s to an increase of 435 m<sup>3</sup>/s. For changes in hydrological variable extremes (Table 18) GCMs typically show that the absolute maximum monthly streamflow (e.g., red crosses and highest box plot legs in Figure 19) will decrease in the future, but this attribute was not statistically tested. The most extreme change in maximum monthly streamflow towards drier conditions is simulated by CanESM2 run1 under RCP8.5. This projection produces a future maximum monthly streamflow which is 2902 m<sup>3</sup>/s less (-25.6%) than its historic simulation. The most extreme change in maximum monthly streamflow towards wetter conditions is simulated by GFDL-ESM2g run1 under RCP8.5. This projection produces a future maximum monthly streamflow 4032 m<sup>3</sup>/s greater (+39.7%) than its historic simulation, but this simulation also has a drier historic period relative to piControl simulations which has the potential to exaggerate future projections. Although there are considerable amounts of uncertainty surrounding these two extreme case projections, they may be potentially useful as context for future flood risk discussions. Similar to the multi-year drought projections in this thesis, the large NCRB might compensate for dry conditions in one area with wetter or near-normal conditions in another area, damping the effect of extreme floods in the large basin.

**Table 17 - Future GCM projection summary for mean values of hydrological variables**

Simulation	Projected Changes to Mean							
	Annual Flow (m <sup>3</sup> /s)		Drought Severity (10 <sup>9</sup> m <sup>3</sup> )		Drought Duration (years)		Maximum Monthly Flow (m <sup>3</sup> /s)	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
CanESM2 r1	324	539	-13.4	-15.6	-1.1	-1.4	72	58
CanESM2 r2	459	637	-18.9	-16.8	-1.4	-1.3	338	435
CanESm3 r3	377	338	-8.0	-1.4	0.0	-0.5	282	-101
GFDL-ESM2g r1	268	397	0.8	-7.4	-0.1	-0.9	3	211
GFDL-ESM2m r1	469	268	-27.3	-15.5	-2.0	-1.2	496	-401
MIROC5 r1	-167	-25	9.2	11.5	0.2	0.1	-901	-581
MIROC5 r2	248	-253	-20.1	23.4	-1.5	0.7	76	-831
MIROC5 r3	126	-32	-4.0	12.4	0.0	0.5	-182	-653
NorESM1-m r1	146	-13	-23.2	-2.0	-1.6	-0.1	221	-172
Ensemble Mean	228		-6.5		-0.7		-91	

**Table 18 - Future GCM projection summary for extreme values of hydrological variables**

Simulation	Projected Change to Minimum Annual Flow (m <sup>3</sup> /s)	Projected Changes to Maximum Values						
		Drought Severity (10 <sup>9</sup> m <sup>3</sup> )		Drought Duration (years)		Maximum Monthly Flow (m <sup>3</sup> /s)		
		RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5
CanESM2 r1	267	134	4.0	-45.7	1.0	-2.0	-565	-2902
CanESM2 r2	335	374	-117.6	-114.4	-10.0	-9.0	-83	1230
CanESm3 r3	234	-21	-129.1	-31.3	-3.0	-1.0	-623	-1255
GFDL-ESM2g r1	-68	135	-15.2	-39.8	1.0	-2.0	671	4032
GFDL-ESM2m r1	574	404	-31.5	-38.1	-4.0	-5.0	2301	-371
MIROC5 r1	-222	-251	-4.4	4.8	1.0	-1.0	-766	800
MIROC5 r2	162	-265	-57.0	86.3	-7.0	-1.0	-2390	-1937
MIROC5 r3	-116	-129	-38.1	-5.8	2.0	2.0	-1569	-1204
NorESM1-m r1	199	58	9.8	-18.7	2.0	0.0	778	102
Ensemble Mean	100		-32.3		-2.0		-208	

## 5.5 Uncertainty

In addition to the address of uncertainty in results and discussion above, it is important to communicate a clear understanding on how uncertainty affects the interpretation of results and conclusions. In a recent publication, Kundzewicz and Gerten (2015) identified uncertainty as a

grand challenge related climate change assessment on water resources. There are several sources of uncertainty addressed in this thesis. The intent of this section is not to characterize magnitude of uncertainty or relative contributions from the sources. Instead, this section intends to discuss identify primary sources of uncertainty including uncertainty in: GCM selection, natural climate variability, and future forcing scenario (RCP). Other sources of uncertainty inherent in this research, but not explicitly considered, include the uncertainty introduced by the bias correction method, the WATROUTE<sub>MOD</sub> routing model, and uncertainty in the observed record. At a high level, different sources of uncertainty are understood to have different magnitudes of impact at different time scales. For example, in the NCRB, Hawkins and Sutton (2009) showed that natural climate variability dominate uncertainty in the first decade of projection, GCM uncertainty becomes increasingly important for the fourth decade and radiative forcing scenario (RCP) dominates uncertainty in the ninth decade. This decadal classification complicates the characterization of uncertainty in this thesis where the future period (2007-2099) spans all three decades considered in Hawkins and Sutton (2009).

Uncertainty in GCM selection can be explored through comparison of individual GCM results to one another. For example, if only MIROC5 run1 were assessed, results suggest drier conditions in the future; but if only CanESM2 run1 were assessed, results suggest a wetter future. Similarly, if only GFDL-ESM2m is considered (only 1 run available), all comparisons suggest that the piControl periods and future projections have increased streamflow due to the dry nature of the historic simulation. The subset of five GCMs sampled a range of models from different modeling centers, with different spatial resolutions, land surface models, and skill. This small subset tends towards projecting a wetter future (e.g., Table 13). However, these results should be confirmed through comparison with additional GCMs from a larger ensemble.

Uncertainty in natural climate variability can be explored through the comparison of the six pi-Control periods (natural climate variability in a stationary climate) and comparison of GCMs with multiple runs (natural climate variability due to different initial conditions). For example, comparison among piControl periods (e.g., Figure 16) illustrates six different plausible climates in quasi-equal (91-94 year) periods that could have been sampled. Similarly, each of the three runs in CanESM2 or MIROC5 could be valid representations of the historic, (1912-2005) period. Comparison of these historic periods (e.g., Figure 16 panels A and D) shows considerable differences in the hydrologic climate. If all piControl periods in GFDL-ESM2g are examined, pi1 and pi2 produce much wetter conditions than pi3 to pi6 (Figure 16 and Figure 17, Panel B). This realization complicates interpretation of future projections. Both future simulations (RCP4.5 and RCP8.5) from run1 show wetter conditions into the future. However, it is conceivable that the historic period sampled a drier climate and the future periods coincidentally sampled wetter climates that are part of the natural (modelled) cycle. Such natural climate variability may mask the climate change signal, and in some cases produce misleading results. Inclusion of the piControl period and GCMs with multiple runs help to demonstrate uncertainty due to natural climate variability, but this phenomenon should be further explored. Future studies should consider models with more runs, such as the 40 member CCSM3 ensemble or the 10,000 year control simulation using CCSM3's atmospheric model (Deser *et al.*, 2012a).

Uncertainty in future forcing scenarios can be explored through comparison of RCP4.5 with RCP8.5 results, but is also evident in the assumptions used to generate the RCP scenarios. In other words, one should note that the RCPs are conceptual pathways in which society, economy and technology evolves into the future and there are many uncertainties associated with the development of RCPs. For variables such as annual average global temperature, which is

well correlated with atmospheric carbon dioxide concentrations, there is a clear indication that RCP8.5 projections have greater temperature increases than RCP4.5 (Figure 1). However, this is not always the case for regional analyses or the more complex hydrologic variables in this thesis. If a strong correlation exists between hydrologic variables and atmospheric carbon dioxide concentrations, one might expect RCP8.5 and RCP4.5 projections to agree on direction and produce proportional changes in magnitude. Results in this thesis show that the two RCPs agree on the direction of change for most variables but the magnitude of change is not always proportional. There are also several instances where the RCPs disagree on the direction of change. Although RCP4.5 and RCP8.5 capture a wide range of possible future scenarios, further assessments should incorporate RCP2.6 and RCP6.0 to corroborate results and assess if a relationship exists between project hydrological change and RCP.

As suggested in Knutti *et al.* (2010a), the spread of results from an ensemble of opportunity (such as the subset used in this thesis) does not provide a direct measure of uncertainty but can help in characterizing uncertainty. In this thesis, agreement among GCM simulations is used as evidence regarding the projected future direction of change. Results in GCM agreement assume that each GCM simulation and each RCP is equally likely, which may or may not be the case. GCM agreement is reported to summarize results but further interpretation should acknowledge the various sources of uncertainty involved in analyzing hydroclimatic conditions in areas with large natural variability.

## CHAPTER 6

### Conclusions and Recommendations

From the results and discussion presented in Chapter 5, conclusions and recommendations are aligned with the four objectives and discussed in this chapter. Since a large amount of data was analyzed, including multiple GCM scenarios, sub-basins, performance metrics, and hydrological variables; conclusions are presented at a higher level compared to content presented in Chapter 5. In many cases, high level statements can be more robust than quantitative measures. Recommendations are presented to guide future work intending to improve and expand on the methods used in this thesis, and to suggest avenues for further study.

#### **6.1 Evaluation of Climate Model Skill**

The 52 GCM simulations show a wide range of skill in producing annual runoff characteristics in the NCRB. No single GCM stands out as the best performing model across all nine metrics, but some GCMs perform better and some worse, overall, when compared to the median GCM performance. There are cases where a single simulation displays very good performance for one metric and very poor performance for another metric. Many GCMs do a reasonable job of reproducing spatial patterns of mean annual runoff in the NCRB, including the Canadian Earth Systems Model, CanESM2. One GCM (BNU-ESM) appears to contain erroneous data, off by orders of magnitude, and is excluded from the analysis.

Past studies have critiqued GCM ability to reproduce long term persistence (e.g., Ault *et al.*, 2014). Results in this thesis show evidence that GCMs are capable of simulating persistent dry and persistent wet conditions but similar to Ault *et al.* (2014), this thesis found that most GCMs have some difficulty in reproducing exact patterns of observed long term persistence.

However, it is found that natural variability can play a major role in evaluating GCM skill, as is the case for autocorrelation. This finding suggests that further work is needed before blanket conclusions can be made about GCM ability to simulate long term hydrologic persistence in the NCRB. Future work should avoid the use of a single reference climate and focus on incorporating natural variability (e.g., a range of reference climates) into GCM skill assessment.

The GCM skill evaluation in this thesis is based on a single observed dataset in a specific time period, and as such reflects GCM performance with respect to the observed dataset and time period. The evaluation of GCM skill is supplementary information for the rest of this analysis and provides a means to identify good performing models, discard models with unrealistic performance and identify unique behaviours to aid in interpretation of results. Overall, the subset of nine simulations selected for further analysis samples GCMs with relatively poor and relatively good performance. This subset with a range in skill helps to ensure that this study samples fundamentally different GCMs that have been developed independently. Use of multiple independent GCMs produces more robust results in comparison to studies that use a single GCM or very similar GCMs with similar skill.

## **6.2 River and Lake Routing**

The  $\text{WATROUTE}_{\text{MOD}}$  routing scheme reasonably translates GCM simulated runoff into naturalized streamflow at various locations in the NCRB. In the absence of calibrated and validated  $\text{WATFLOOD}$  and  $\text{WATROUTE}$  models for the entire NCRB,  $\text{WATROUTE}_{\text{MOD}}$  provides a comparable routing model useful for GCM runoff.  $\text{WATROUTE}_{\text{MOD}}$  is particularly attractive for study domains comprising the entire NCRB and when GCM runoff is an input. Coding of  $\text{WATROUTE}_{\text{MOD}}$  into MATLAB software creates a product that is well positioned for GCM-based runoff studies and is capable of relatively quick processing speeds, allowing many

years of streamflow to be simulated in a timely fashion. Although GCM runoff can be adapted to WATROUTE input file formats, the use of MATLAB and WATROUTE<sub>MOD</sub> provides a more direct integration.

Further advancements of WATROUTE<sub>MOD</sub> are possible and can increase comparability with WATROUTE. Some improvements come with computational cost (e.g., using a finer spatial resolution) while others are modeller inputs (e.g., Manning's n, channel geometry, number of lakes coded). Through the comparison of WATROUTE<sub>MOD</sub> with WATROUTE, several WATROUTE functions are also identified for further investigation such as WATROUTE's internal handling of channel geometry and in-channel Manning's n when land classification within a grid contains a substantial portion of water coverage.

Some of the differences between WATROUTE and WATROUTE<sub>MOD</sub> are masked when evaluating mean monthly streamflow as opposed to mean daily streamflow. As such, WATROUTE<sub>MOD</sub> serves the purposes of this thesis, which analyzes mean monthly streamflow in the NCRB. Many areas for improvement can be considered but will increase the model's complexity and do not necessarily guarantee more accurate results. WATROUTE<sub>MOD</sub> used in this thesis relies on routing total runoff. In theory, total runoff could be partitioned into surface runoff, and groundwater components as in Poitras *et al.* (2011) and Sperna Weiland *et al.* (2012b) for example. Separate handling of groundwater would further attenuate the hydrograph, provide greater baseflow response, and increase the realism with minimal increases in computational cost. However, the additional data requirements (i.e., individual runoff components) might reduce the already small GCM ensemble and limit how many GCMs are analyzed. Other potential improvements include the use of a dynamic method for considering non-contributing drainage areas, the use of GCM temperature data to derive lake-outlet

performance curves (ice restricted) in individual years and the use of Regional Climate Model outputs for enhanced spatial detail. Non-contributing drainage area may behave differently in wet or dry years and could be an important consideration and impact adjustment values in higher quantiles. Furthermore, anthropogenic changes in land use and drainage systems over time may have also impacted hydrological response such as flooding in the Assiniboine River Basin as discussed in Szeto *et al.* (2015). Outlet performance based on GCM temperatures might add additional variability and uncertainty but could be important, especially when considering a warmer future. Finally, the use of finer resolution climate models (such as RCMs) provides enhanced spatial detail that may allow for analysis at a finer sub-basin scale.

### **6.3 Bias Correction**

Quantile mapping techniques developed for meteorological variables were adapted for use in this thesis to remove bias in the GCM-driven WATROUTE<sub>MOD</sub> streamflow simulations. The technique successfully corrects volume (not timing) and produces naturalized streamflow scenarios with similar characteristics as the observed record. The technique produces streamflow scenarios that facilitate comparison with observations, facilitate interpretation and provide more realistic patterns compared to non-bias-corrected simulations. The success of quantile mapping for streamflow simulations in this thesis illustrates potential for further implementation in similar studies and for adapting the method for use with other GCM variables.

Several improvements can be explored for future research. The technique in this thesis corrects mean annual flow and the range of monthly flows but the GCM's underestimation of inter-annual variability remains. A more advanced procedure could focus on correcting monthly and inter-annual variability simultaneously. Such a procedure might first correct the annual time series and then implement a second step to correct monthly time series. Guidance for this type of

improvement might be found in existing methods such as the Double Bias Corrected Constructed Analogue method, described in Werner and Cannon (2015). A preferred bias correction method might first apply a spatially distributed bias correction of gridded GCM runoff using pseudo-observed gridded runoff prior to routing (as explored in Duong *et al.*, 2015). This approach would require a reliable gridded runoff dataset with long term observations, which does not exist at this time.

Sub-basins with large quantile mapping adjustment values offer some guidance in identifying potential areas for improvement. In this thesis, the largest adjustment values were computed for the Lake Winnipeg PIAO sub-basin, which is the largest sub-basin with the fewest lakes coded into WATROUTE<sub>MOD</sub>. A potential improvement for bias correction in this sub-basin includes further separation into sub-basins such as the Red and Assiniboine river sub-basins. Improvements to WATROUTE<sub>MOD</sub> within Lake Winnipeg PIAO might also improve the bias correction. For example, additional lakes and better representation of non-contributing drainage area could reduce the bias due to the routing scheme. Consideration of the Portage diversion (which diverts a part of the Assiniboine River flow into Lake Manitoba) would also help attenuate peaks and reduce adjustment values at higher quantiles.

One limitation of the bias correction method is that it compares naturalized flow from WATROUTE<sub>MOD</sub> to the observed record, which is subject to anthropogenic influence. This thesis presents a unique bias correction method that merges naturalized streamflow simulations with regulated streamflow observations by grouping all months together and bias correcting for monthly volume instead of timing. However, some anthropogenic volume errors, such as water withdrawals in the SRB are unaccounted for. A more robust bias correction method is possible if naturalized streamflow records (with anthropogenic effects removed) were available. With

naturalized flow records, monthly bias correction methods could be explored without concern that bias correction is also accounting for anthropogenic influence. However, it is worthwhile noting that many recent studies have generally ignored anthropogenic influence (e.g., Hagemann *et al.*, 2013; Haddeland *et al.*, 2011), attesting to the difficulty in considering anthropogenic effects. Other studies that have considered anthropogenic influence show that climate change effects generally surpass anthropogenic effects in the NCRB (Haddeland *et al.*, 2014).

#### **6.4 Streamflow Scenario and Time Series Analysis**

Time series' of bias corrected streamflow scenarios were analyzed to characterize how observed variability (1912-2005) compares to natural variability and future projections in the NCRB. Four key hydrological variables were analyzed: mean streamflow, multi-year hydrologic drought severity, multi-year hydrologic drought duration and maximum monthly streamflow. Since bias correction only intended to correct monthly streamflow volume, the historic GCM simulations don't always capture every aspect of the observed record. Conclusions are therefore derived from comparison of GCM piControl and future simulations to GCM historic simulations. For future simulations, both RCP4.5 and RCP8.5 are considered equally probable. Based on results, it is preferred to draw conclusions by comparing GCM piControl and future simulations to GCM historic simulations as opposed to LTFD observations. The degree of consistency (DOC) criteria (Table 8) is used to summarize conclusions among all GCMs.  $DOC_{NON-STAT}$  is used to summarize changes in the coefficient of variability (CV) and  $DOC_{STAT}$  is used to summarize changes based on results from the MW-U statistical test.

Variability in mean streamflow is analyzed at annual and monthly resolutions. Results show weak evidence ( $DOC_{NON-STAT}$ ) that variability in the historic period underestimates potential variability in a stationary climate for mean annual streamflow; and weak evidence

(DOC<sub>NON-STAT</sub>) for future changes indicating that mean annual streamflow variability may increase. Despite the weak nature of the evidence, these results highlight the importance of considering more than just observations to fully understand risk due to hydrologic variability.

All GCMs tend to agree that normal winter flows in the NCRB are projected to increase into the future, exceeding normal flows in both historic and piControl periods. This represents an upward shift in months with typically lower natural flows. There is less agreement among GCM simulations regarding projected changes in other months, where some show decreases and other show increases. Overall, there is strong evidence (DOC<sub>NON-STAT</sub>) that variability in the historic period underestimates potential variability in a stationary climate for mean monthly streamflow and strong evidence (DOC<sub>NON-STAT</sub>) for future changes indicating that that mean monthly streamflow variability is projected to decrease further. Decreased future mean monthly streamflow variability produces a flattening of the naturalized hydrograph. The process of hydrograph flattening is typically done through regulation to aid water management. Here it is seen that a warming climate may naturally assist the hydrograph flattening process currently done by regulating reservoirs. Through modeling of naturalized conditions in this thesis, changes to mean monthly streamflow variability are potentially more pronounced whereas the inclusion of reservoir operation and water withdrawals could have masked the climate change signal if regulation were considered. Changes in mean monthly flows do not seem to be particularly sensitive to the future RCP scenario, suggesting that although increased carbon dioxide concentrations can alter the monthly mean flow patterns, the degree of change may also be influenced by natural climate variability. This result provides some rationale for considering a larger ensemble of RCP scenarios (e.g., RCP2.6, RCP4.5, RCP6.0 and RCP8.5) to increase the number of futures scenarios to assess and make best use of available GCM data.

Mean annual streamflow, mean drought severity, mean drought duration and the mean of maximum monthly streamflow were analyzed statistically to assess how streamflow in the historic period (1912-2005) compares to natural variability and future projections. Results show negligible statistical evidence that mean annual streamflow has systematically changed within piControl and historic periods. However, all GCMs show some evidence that a statistically significant difference even in a stationary climate is possible. The presence of statistically significant differences suggests that notable increases or decreases are possible among the 90-94 year periods, perhaps due to very low frequency climatic patterns such as PDO. Results show moderate evidence ( $DOC_{STAT}$ ) for future changes indicating that mean annual streamflow is projected to increase. The average of all GCM simulations projects future mean annual streamflow to increase by  $228 \text{ m}^3/\text{s}$  with individual projections range from a decrease of  $253 \text{ m}^3/\text{s}$  to an increase of  $637 \text{ m}^3/\text{s}$ . Increasing mean annual streamflow in the future provides greater water supply on average. The worst case projection corresponds to an approximate 7% reduction in mean annual streamflow in the NCRB. Depending on where in the NCRB this reduction is projected to occur (not presented in this thesis), there are potential water resource management strategies to compensate for decreases. For example, one might use diversion infrastructure to move water to where it is needed.

Results show negligible evidence ( $DOC_{STAT}$ ) that the mean drought severity and mean drought duration has systematically changed within piControl and historic periods. However, similar to mean annual streamflow, most GCMs show some evidence that a statistically significant difference even in a stationary climate is possible. The presence of statistically significant differences suggests that different mean drought severities and durations are possible among the 90-94 year periods, perhaps due to very low frequency climatic patterns such as PDO.

Results also show negligible evidence ( $DOC_{STAT}$ ) that the mean drought severity and mean drought duration is projected to change in the future. The average of all GCM simulations (ensemble mean) projects future mean drought severity to decrease (i.e., become wetter) by  $6.5 \cdot 10^9 \text{ m}^3$  and mean drought duration to decrease (i.e., become shorter) by 0.7 years. The most extreme future projection (MIROC5 run2 RCP8.5) produced a future drought  $86.3 \cdot 10^9 \text{ m}^3$  more severe (approximately 81% more severe) than the historic GCM simulation. Despite being the worst case presented in this thesis, this simulated drought change is considerably less severe than a stochastically simulated a drought in Burn and DeWit (1996), which was three times worse than the drought of record. The worst projected change in drought severity in this thesis is associated with a shorter duration, suggesting that droughts can be more severe but this does not necessarily coincide with longer durations. Although there is a considerable amount of uncertainty surrounding this one projection, it is potentially useful as context for future drought risk discussions. These extreme projections may also partially serve as a tool to help address the grand global challenge for adaptation decision under high uncertainty proposed by Kundzewicz and Gerten (2015). The very dry MIROC5 run2 RCP8.5 projection does not show uniform drying throughout the NCRB and includes some regions such as the SRB which have much more drastic drought impacts (not shown).

Further investigation could consider a more detailed spatial breakdown to see if greater evidence exists for drought projections in specific regions as opposed to the entire NCRB which may have a damping effect on drought severity with wetter sub-basins offsetting drier sub-basins. The general finding that wet regions will become wetter and dry regions will become drier (Kundzewicz and Gerten, 2015) has important implications for the NCRB, which contains both wetter (e.g., WRB) and drier (e.g., SRB) regions. The balance of evidence in this thesis

suggests that the NCRB as a whole is tending towards slightly higher mean annual streamflow, which can be beneficial for hydropower generation near the basin outlet. However, the disaggregation of these changes into specific sub-basins may have important regional impacts.

These results show negligible evidence ( $DOC_{STAT}$ ) of systematic changes within piControl and historic periods for mean of maximum monthly streamflow. Results also show negligible evidence ( $DOC_{STAT}$ ) that the mean of maximum monthly streamflow is projected to increase. The average of all GCM simulations (ensemble mean) projects the mean of future maximum monthly streamflow to decrease by  $91 \text{ m}^3/\text{s}$ . Individual projections range from a mean decrease of  $901 \text{ m}^3/\text{s}$  to an increase of  $435 \text{ m}^3/\text{s}$ . The most extreme simulations project future absolute maximum monthly streamflow to decrease by as much as  $2902 \text{ m}^3/\text{s}$ , and increase by as much as  $2301 \text{ m}^3/\text{s}$ . Some of these projections produce alarming results at first, but it is important to consider that statistical analysis is usually combined with observed data in engineering design, and that the impact of different GCM streamflow samples on statistically derived low frequency return periods (e.g., 1:1,000 year floods) was not assessed in this thesis. This would make an interesting area of study, and is recommended in the future as a next step.

There are several key areas, in addition to those identified above, where further study would be beneficial to the science of climate change impacts on the NCRB water resources. In the case of aggregated NCRB flows, opposite changes among sub-basins may mask a more prominent pattern. This was discussed for drought severity, but is applicable to all other hydrological variables as well. Extending the existing analysis to individual sub-basins might identify whether spatial patterns, with greater GCM agreement, emerge. Additionally, to enhance understanding of large scale drivers of droughts and floods it would be interesting to isolate very extreme events produced by the GCMs and further analyze the GCM output. For example, one

could extract atmospheric and oceanic variables from CanESM2 in the piControl period (pi2) to understand what caused the drier conditions in this period. Other good candidate simulations for this study include the MIROC5 run2 RCP8.5 simulation, which produced severe future hydrologic droughts, and the GFDL-ESM2m model that produced relatively drier conditions in the historic period compared to both piControl and future periods. However, a comprehensive assessment should also consider multiple climate models, runs and RCPs.

For some of the more extreme hydrological variables (e.g., drought severity and maximum monthly streamflow), there is limited evidence with respect to systematic differences between the historic period and natural variability and how hydrologic conditions are projected to change into the future. This uncertainty creates challenges for implementing results in a practical manner. An alternative to implementing uncertain results in adaptation planning could be to further analyze output and develop climate change informed sensitivity analysis limits. For example, one could calculate percent changes in the most severe future drought, rank the GCM projections from smallest change to largest change and select an upper and lower limit (e.g., 5<sup>th</sup> and 95<sup>th</sup> percentiles) to inform a sensitivity analysis. From a general uncertainty and adaptation planning standpoint, it is important to stress the use of multiple GCMs, multiple member runs and multiple RCPs to sample various sources of uncertainty and produce more robust conclusions. Furthermore, it is important to stress that for some hydrologic variables, especially extremes, analyses should attempt to use longer time periods to limit the potential for under sampling the range in natural variability. These suggestions apply to future research, as well as practical applications.

Overall, this thesis presents information to leverage long term GCM simulations for assessment of how observed streamflow compares to natural variability and future projections. In

conclusion, results show negligible statistical evidence about systematic changes (historic vs. natural climate variability and historic vs. future projections) in drought severity, drought duration and maximum of mean monthly streamflow. Although there is some indication that statistically significant changes are possible, there is a lack of GCM agreement on the magnitude and direction of change. This is a common conclusion in the assessment of climate change impacts on extreme events and in light of conclusive information, some GCM-derived extreme changes are shown in this thesis to guide further discussion and efforts towards adaptation. Greater confidence exists in the assessment of mean annual streamflow and mean monthly streamflow. Although there is negligible statistical evidence that mean annual streamflow was different in a stationary climate compared to the historic climate, there is moderate statistically significant evidence ( $DOC_{STAT}$ ) that mean annual streamflow is projected to increase into the future. There is also some weaker non-statistically significant evidence ( $DOC_{NON-STAT}$ ) that mean annual streamflow was more variable in a stationary climate and is projected to increase in the future. Finally there is strong non-statistically significant evidence ( $DOC_{NON-STAT}$ ) that mean monthly streamflow was more variable in a stationary climate compared to the historic period and will continue to decrease into the future, demonstrating the tendency towards a flatter hydrograph in a warmer climate for the NCRB.

## References

- Agriculture and Agri-food Canada. (2012). ISO 19131 AAFC Watershed Project - 2012 - Data Product Specification. Revision A. Agriculture and Agri-Food Canada. 34pp.
- Akintuğ, B., & Rasmussen, P. F. (2005). A Markov switching model for annual hydrologic time series. *Water Resources Research*, 41, W09424.
- Akintuğ, B. (2006). Analysis of system drought for Manitoba Hydro using stochastic methods. Doctor of Philosophy Thesis, Department of Civil Engineering, University of Manitoba, Winnipeg, MB Canada. 300pp. Retrieved from MSpace (<http://hdl.handle.net/1993/7967>).
- Aldridge, B. N., & Garrett, J. M. (1973). Roughness coefficients for stream channels in Arizona. United States Department of the Interior Geological Survey. Tuscan, AZ. Open-file Report 73-3. Available online at: <http://pubs.er.usgs.gov/publication/ofr733>. 87pp.
- Alkama, R., Marchand, L., Ribes, A., & Decharme, B. (2013). Detection of global runoff changes: results from observations and CMIP5 experiments. *Hydrology and Earth Systems Sciences*, 17, 2967-2979.
- Andreadis, K. M., Schumann, G. J.-P., & Pavelski, T. (2013). A simple global river bankfull width and depth database. *Water Resources Research*, 49(1): 7164-7168.
- Arnel, N. W. & Gosling, S. N. (2013). The impacts of climate change on river flow regimes at the global scale. *Journal of Hydrology*, 486: 351-364.
- Arora, V. K. (2001). Streamflow simulations for continental-scale river basins in a global atmospheric general circulation model. *Advances in Water Resources*, 24: 775-791.
- Arora, V. K., Chiew, F. H. S., & Grayson, R. B. (1999). A river flow routing scheme for general circulation models. *Journal of Geophysical Research*, 104, 14,347-14,357.
- Arora V. K, Seglenieks, F., Kouwen, N., & Soulis, E. (2001). Scaling aspects of river flow routing. *Hydrological Processes*. 15(3): 461-477.
- Arora, V. K. & Boer, G. J. (1999). A variable velocity flow routing algorithm for GCMs. *Journal of Geophysical Research*, 104, 30,965-30,979.
- Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., & Meko, D. M. (2014). Assessing the risk of persistent drought using climate model simulations and paleoclimate data. *J. Climate*, 27: 7529-7549.
- Ault, T. R., Cole, J. E., & St. George, S. (2012). The amplitude of decadal to multidecadal variability in precipitation simulated by state-of-the-art climate models. *Geophysical Research Letters*, 39: L21705., doi:10.1029/2012GL053424.

- Ball, T. F. (1983). Climatic Change in Canada: A preliminary analysis of weather information from the Hudson's Bay Company Forts at York Factory and Churchill Factory, 1714-1850. Doctor of Philosophy thesis presented to the University of London Queen Mary College. Winnipeg, Manitoba.
- Beven, K. (2011). I believe in climate change but how precautionary do we need to be in planning for the future? *Hydrological Processes*, 25: 1517-1520.
- Bonsal, B.R., Aider, R., Gachon, P., & Lapp, S. (2013). An assessment of Canadian prairie drought: past, present, and future. *Climate Dynamics*, 41(2), 501-516.
- Bohrn, S. K. (2012). Climate change impact assessment and uncertainty analysis of the hydrology of a northern, data-sparse catchment using multiple hydrological models. Master of Science Thesis, Department of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada, 207pp. Retrieved from MSpace (<http://hdl.handle.net/1993/13692>).
- Brekke, L., & Prairie, J. (2009). Long-term planning hydrology based on various blends of instrumental records, paleoclimate, and projected climate information. Technical report prepared by Bureau of Reclamation, U.S. Department of the Interior Bureau of Reclamation. 139pp.
- Burn, D. H., & Whitfield, P. H. (2015). Changes in floods and flood regimes in Canada. *Canadian Water Resources Journal*, Published Online June 2, 2015. doi: 10.1080/07011784.2015.1026844
- Burn, D. H., Wychreschuk, J. & Bonin, D. V. (2004). An integrated approach to the estimation of streamflow drought quantiles, *Hydrological Sciences Journal*, 49(6). doi: 10.1623/hysj.49.6.1011.55727.
- Burn, D. H., & DeWit, W. J. (1996). Spatial characterization of drought events using synthetic hydrology. *Canadian Journal of Civil Engineering*, 23: 1231-1240.
- Caissie, D., & El-Jabi, N. (1991). A stochastic study of floods in Canada: frequency analysis and regionalization. *Canadian Journal of Civil Engineering*, 18: 225-236.
- Camici, S., Brocca, L., Melone, F., & Moramarco, T. (2014). Impact of climate change on flood frequency using different climate models and downscaling approaches. *Journal of Hydrologic Engineering*, 19(8): 04014002.
- Canadian Hydraulics Center, 2010. Green Kenue reference manual (September 2010). Canadian Hydraulics Center (CHC), National Research Council Canada. Ottawa, Ontario, Canada. 340pp.
- Chen, J., Brissette, F. P., Poulin, A. & Leconte, R. (2011). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research*, 47: W12509.

- Chen, J., Brissette, F. P., Chaumont, D. & Braun, M. (2013). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, 49: 4187-4205
- Cheng, C. S., Lopes, E., Fu, C., & Huang, Z. (2014). Possible impacts of climate change on wind gusts under downscaled future climate conditions: Updated for Canada. *Journal of Climate*, 27: 1255-1270.
- Chernet, H. H., Midttømme, G. H., & Alfredsen, K. (2014) Safety of hydropower dams in a changing climate. *Journal of Hydrologic Engineering*, 19(3): 569-582.
- Clavet-Gaumont, J., Sushama, L., Khaliq, M. N., Huziy, O., & Roy, R. (2013). Canadian RCM projected changes to high flows for Quebec watersheds using regional frequency analysis. *International Journal of Climatology*, 33(14): 2940-2955.
- Cohen, S., Koshida, G. & Mortsch, L. (2015). Climate and water availability indicators in Canada: Challenges and a way forward. Part III – Future scenarios. *Canadian Water Resources Journal*, 40(2): 160-172.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichet, T., Friedlingstein, P.,...Wehner, M. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J.,...Midgley, P. M. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 108pp.
- Coulibaly, P., Samuel, J., Pietroniro, A., & Harvey, D. (2013). Evaluation of Canadian National Hydrometric Network density based on WMO 2008 standards. *Canadian Water Resources Journal*, 38(2): 159-167.
- Cunnane, C. (1978). Unbiased plotting positions - A review. *Journal of Hydrology*, 37: 205–222.
- Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012a). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, 2: 775-779.
- Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012b). Uncertainty in climate change projections: the role of internal variability. *Journal of Climate Dynamics*, 38: 527-546.
- Di Luca, A., de Elía, R., & Laprise, R. (2012) Potential for added value in precipitation simulated by high-resolution nested Regional Climate Models and observations. *Climate Dynamics*, 38: 1229-1247.
- Dingman, S. L. (2008). *Physical Hydrology* (2nd ed.). Long Grove, Illinois, United States of America: Waveland Press, Inc. 646pp.

- Duong, D. T., Tachikawa, Y., & Yorozu, K. (2015, January 28-30). Application of a land surface model for bias correction of runoff generation data from MRI-AGCM3.2S Dataset. Paper presented at Thai Hydrologist Association (THA) 2015 International Conference on Climate Change and Water & Environment Management in Monsoon Asia, Bangkok, Thailand. 6pp.
- Ehsanzadeh, E., van der Kamp, G., & Spence, C. (2011). The Impact of climate variability and change in the hydroclimatology of Lake Winnipeg watershed. *Hydrological Processes*, 26(18), 2802-2813. DOI: 10.1002/hyp.8327.
- Eum, H.-I., Dibiki, Y., & Prowse, T. (2014). Uncertainty in modelling the hydrologic response of a large watershed: a case study of the Athabasca River basin, Canada. *Hydrological Processes*, 28: 4272-4293.
- Falloon, P., Betts, R., Wiltshire, A., Dankers, R., Mathison, C., McNeal, D.,...Trigg, M. (2011). Validation of river flows in HadGEM1 and HadCM3 with the TRIP River Flow Model., *Journal of Hydrometeorology*, 12: 1157-1180.
- Fekete, B. M., Vörösmarty, C. J., & Grabs, W. (2000). Global composite runoff fields based on observed river discharge and simulated water balances. Documentation for UNHGRDC Composite Runoff Fields, v.1.0, Global Runoff Data Center (GRDC) Koblenz, Germany. 120 pp.
- Fekete, B. M., Vörösmarty, C. J. & Grabs, W. (2002). High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, 16(3): 15-1 - 15-10.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W.,...Rummukainen, M. (2013). Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J.,...Midgley, P. M. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 741-866.
- Frigon, A., Music, B., & Slivitzky, M. (2010). Sensitivity of runoff and projected changes in runoff over Quebec to the update interval of lateral boundary conditions in the Canadian RCM. *Meteorologische Zeitschrift*, 19: 225-236.
- Gao, Y., Vano, J. A., Zhu, C., & Lettenmaier, D. P. (2011). Evaluating climate change over the Colorado River basin using regional climate models. *Journal of Geophysical Research*, 116: D13104.
- Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *J. of Geophysical Research*. 113: D06104. doi:10.1029/2007JD008972.
- González-Zeas, D., Garrote, L., Iglesias, A., & Sordo-Ward, A. (2012). Improving runoff estimates from regional climate models: a performance analysis in Spain. *Hydrology and Earth Systems Sciences*, 16: 1709-1723.

- Guay, C., Minville, M., & Braun, M. (2015). A global portrait of hydrological changes at the 2050 horizon for the province of Quebec. *Canadian Water Resources Journal*, 40(3): 285-302.
- Gudmundsson, L. & Seneviratne, S. L. (2015). Towards observation-based gridded runoff estimates for Europe. *Hydrology and Earth Systems Sciences*, 19: 2859-2879.
- Gudmundsson, L., Tallaksen, L. M. & Stahl, K. (2011). Projected changes in future runoff variability – a multi model analysis using the A2 emission scenario. Technical Report No. 49, WATCH Water and Global Change. 11pp.
- Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W.,...P. Yeh, P. (2011). Multimodel estimate of the global terrestrial water balance: Setup and first results. *Journal of Hydrometeorology*, 12: 869–884.
- Haddeland, I. D., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N.,...Wisser, D. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, 111(9): 3251-3256.
- Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I.,...Wiltshire, A. J. (2013). Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Systems Dynamics*, 4: 123-144.
- Hashino, T., A.A. Bradley, A. A., & Schwartz, S. S. (2007). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth Systems Sciences*, 11: 939-950.
- Hawkins, E. & Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37: 407-418.
- Henley, B. J., Thyer, M. A., Kuczera, G., & Franks, S. W. (2011). Climate-informed stochastic hydrological modeling: Incorporating decadal-scale variability using paleo data, *Water Resources Research*, 47: W11509.
- Hirabayashi, Y., Kanae, S., Emori, S., Oki, T., & Kimoto, M. (2008). Global projections of changing risks of floods and droughts in a changing climate, *Hydrological Sciences Journal*, 53(4): 754-772.
- Huntington, T. G., (2006). Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, 319: 83-95
- Huziy, O., Sushama, L., Khaliq, M. N., Laprise, R., Lehner, B. & Roy, R. (2013). Analysis of streamflow characteristics over Northeastern Canada in a changing climate. *Climate Dynamics*, 40: 1879-1901.

- IPCC-TGICA, (2007). General Guidelines on the use of scenario data for climate impact and adaptation assessment. Version 2. Prepared by T.R. Carter on behalf of the Intergovernmental Panel on Climate Change, Task Group on Data and Scenario Support for Impact and Climate Assessment. 66 pp.
- Johnson, F., Westra, S., Sharma, A., & Pitman, A. J. (2011). An assessment of GCM skill in simulating persistence across multiple time scales. *Journal of Climate*, 24: 3609-3623.
- Jones, P. D., & Mann, M. E. (2004). Climate over past millennia, *Rev. Geophys.*, 42: RG2002, doi:10.1029/2003RG000143.
- Kay, A. L., Davies, H. N., Bell, V. A. & Jones, R. G. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic Change*, 92: 41-63.
- Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P. J., Hewitson, B., & Mearns, L. (2010a). Good practice guidance paper on assessing and combining multi model climate projections. In: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M. & Midgley, P. M. (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland. 13pp.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010b). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23: 2739-2758.
- Koirala, S., Hirabayashi, Y., Mahendran R., & Kanae, S. (2014). Global assessment of agreement among streamflow projections using CMIP5 model outputs. *Environmental Research Letters*, 9: 1-11.
- Koshida, G., Cohen, S., & Mortsch, L. (2015). Climate and water availability indicators in Canada: Challenges and a way forward. Part I - Indicators. *Canadian Water Resources Journal*, 40(2): 135-145.
- Kouwen, (2012). WATFLOOD™ / WATROUTE Hydrological Model Routing & Flow Forecasting System (October 2012 Revision). University of Waterloo. Waterloo, Ontario, Canada. Accessed online December 1, 2012 at: <http://www.civil.uwaterloo.ca/watflood/downloads/downloads.htm>. 267pp.
- Kubursi, A., & Magee, L. (2010). Manitoba Hydro Risks: An Independent Review. Submitted to The Public Utilities Board of Manitoba. November 15, 2010. Accessed online May 4, 2015 at: <http://www.pub.gov.mb.ca/exhibits/km-2.pdf>. 350pp.
- Kumar, S., Lawrence, D.M., Dirmeyer, P.A., & Sheffield, J. (2014). Less reliable water availability in the 21<sup>st</sup> century climate projections. *Earth's Future*, 2(3), 152-160.
- Kundzewicz, Z. W. & Robson, A. J. (2000). Detecting trend and other changes in hydrological data. World climate program-data and monitoring, WMO/TD-No. 1013, World Meteorological Organization, Geneva. 158pp.

- Kundzewicz, Z. W., & Robson, A.J. (2004). Change detection in hydrological records – a review of the methodology. *Hydrological Sciences Journal*, 49(1): 7-19.
- Kundzewicz, Z.W., & Gerten, D. (2015). Grand challenges related to the assessment of climate change impacts on freshwater resources. *Journal of Hydrologic Engineering*, 20, SPECIAL ISSUE: Grand Challenges in Hydrology: A4014011-1.
- Kuo, C.-C. & Gan, T. Y. (2015). Risk of exceeding extreme design storm events under possible impact of climate change. *Journal of Hydrologic Engineering*, 20(12), 05015038.
- Lehner, B., Verdin, K. & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *EOS Transactions, American Geophysical Union*, 89(10): 93-94.
- Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., & Leung, R. (2013). A physically based runoff routing model for land surface and earth systems models. *Journal of Hydrometeorology*, 14: 808-828.
- Lake of the Woods Control Board. (2014). Lake St. Joseph and its diversion. Lake of the Woods Control Board Regulation Guide. Accessed online May 12, 2015.at: <http://lwcb.ca/reg-guide/rgp-PT2-LKSTJODIV.html>. Last edited November 17, 2014.
- Maloney, E. D., Camargo, S. J., Chang, E., Colle, B., Fu, R., Geil, K. L.,...Zhao, M. (2014). North American climate in CMIP5 experiments. Part III: Assessment of twenty-first-century projections. *Journal of Climate*, 27: 2230–2270.
- Manitoba Hydro, (2007). Manitoba Hydro-Electric Board 56<sup>th</sup> Annual Report. For the year ended March 31, 2007. Accessed online May 29, 2015 at [https://www.hydro.mb.ca/regulatory\\_affairs/electric/gra\\_08\\_09/information\\_requests/Appendix\\_15-Manitoba\\_Hydro-Electric\\_Board\\_56th\\_Annual\\_Report\\_March\\_31-2007.pdf](https://www.hydro.mb.ca/regulatory_affairs/electric/gra_08_09/information_requests/Appendix_15-Manitoba_Hydro-Electric_Board_56th_Annual_Report_March_31-2007.pdf)
- Manitoba Hydro, (2013a). Needs for and alternatives To: Chapter 5 – The Manitoba Hydro System, Interconnections and Export Market. August 2013. Accessed online May 4, 2015 at: [http://www.hydro.mb.ca/projects/development\\_plan/bc\\_documents/nfat\\_business\\_case\\_chapter\\_05\\_the\\_manitoba\\_hydro\\_system\\_interconnection\\_and\\_export\\_markets.pdf](http://www.hydro.mb.ca/projects/development_plan/bc_documents/nfat_business_case_chapter_05_the_manitoba_hydro_system_interconnection_and_export_markets.pdf)
- Manitoba Hydro, (2013b). Needs for and alternatives To: Chapter 10 – Economic Uncertainty Analysis – Probabilistic Analysis and Sensitivities. August 2013. Accessed online May 4, 2015 at: [http://www.hydro.mb.ca/projects/development\\_plan/bc\\_documents/nfat\\_business\\_case\\_chapter\\_10\\_economic\\_uncertainty\\_analysis\\_probabilistic\\_analysis\\_and\\_sensitivities.pdf](http://www.hydro.mb.ca/projects/development_plan/bc_documents/nfat_business_case_chapter_10_economic_uncertainty_analysis_probabilistic_analysis_and_sensitivities.pdf)
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J.,...Zwiers, F. W. (2010). Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. intergovernmental panel on climate change (IPCC). 4pp.

- Maurer, E. P., & Pierce, D. W. (2014). Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth Systems Science*, 18: 915-925.
- MATLAB and Statistics Toolbox Release R2012a, The MathWorks, Inc., Natick, Massachusetts, United States. <http://www.mathworks.com/products/matlab/>.
- McGuffie, K., & Henderson-Sellers, A. (2005). *A Climate Modelling Primer*, 3<sup>rd</sup> Edition. Chichester, West Sussex, England: John Wiley & Sons, Ltd. 288pp.
- Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S.,...Snyder, M. (2012). The Northern American regional climate change assessment program. Overview of Phase I Results. *Bulletin of the American Meteorological Society*, 93: 1337-1362.
- Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M.,...Zhao, Z.-C. (2007). Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B.,...Miller, H. L. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 747-845.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kunzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J., (2008). Stationarity is dead: whither water management? *Science*, 319: 573-574.
- Milly, P. C. D., Dunne, K. A., & Vecchia, A. V., (2005). Global pattern of trends in streamflow and water availability in a changing climate. *Letters to Nature*, 438: 347-350.
- Milly P. C. D., Wetherald, R. T., Dunne, K. A., & Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. *Nature*, 415: 514-517.
- Mote, P., Brekke, L., Duffy, P. B., & Maurer, E. (2011). Guidelines for constructing climate scenarios. *EOS, Transactions, American Geophysical Union*. 92(31): 257-264.
- Mpelasoka, F. S., & Chiew, F. H. S. (2009). Influence of rainfall scenario construction methods on runoff projections. *Journal of Hydrometeorology*, 10: 1168-1183.
- Music, B., Caya, D., Frigon, A., Musy, A., Roy, R., & Rodenhuis, D. (2012). Canadian regional climate model as a tool for assessing hydrological Impacts of climate change at the watershed scale. *Climate Change*: 157-165. doi: 10.1007/978-3-7091-0973-1\_12
- Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S.,...Dadi, Z. (2000). *Special Report on Emissions Scenarios*. Cambridge University Press. 599pp.
- Newbury, R. W., McCullough, G. K., & Hecky, R. E. (1984). The Southern Indian Lake Impoundment and Churchill River Diversion. *Canadian Journal of Fish Aquatic Sciences*, 41: 548-557.

- Nohara, D., Kitoh, A., Hosaka, M., & Oki, T. (2006). Impact of climate change on river discharge projected by multimodel ensemble. *Journal of Hydrometeorology*, 7: 1076-1089.
- van Oldenborgh, G. J., & Burgers, G. (2005). Searching for decadal variations in ENSO precipitation teleconnections. *Geophysical Research Letters*, 32(15): L15701,
- Oleson, K. W., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P.,...Zeng, X. (2004). Technical description of the community land model (CLM). NCAR Technical Note, NCAR/TN-461+STR, National Center for Atmospheric Research, Boulder, Colorado. 174pp.
- Olsen, J. R., Ayyub, B. M., Barros, A., Lei, W., Lombardo, F., Medina, M.,...Smith, J. (2015). Adapting infrastructure and civil engineering practise to a changing climate. *American Society of Civil Engineers*, Committee on Adaptation to a Changing Climate. Reston, Virginia, United States of America. 93pp.
- Pierce, D. W., Barnett, T. P., Santer, B. D., & Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Science*, 106(21): 8441-8446.
- Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davidson, B.,...Pellerin, P. (2007). Development of the MESH modeling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth Systems Science*, 11: 1279-1294.
- Poitras, V., Sushama, L., Seglenieks, F., Khaliq, M. N., & Soulis, E. (2011). Projected changes to streamflow characteristics over Western Canada as simulated by the Canadian RCM. *Journal of Hydrometeorology*, 12: 1395-1413.
- Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R.,...Wisser, D. (2014). Hydrological droughts in the 21<sup>st</sup> century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences*, 111(9): 3262-3267.
- Rannie, W. F., 1999. A survey of hydroclimate, flooding, and runoff in the Red River Basin prior to 1870. Geological Survey of Canada Open File 3705. Natural Resources Canada. 189pp.
- Räty, O., Räisänen, J., & Ihäisi, J. S. (2014). Evaluation of delta change and bias correction methods for future daily precipitation: intermodal cross-validation using ENSEMBLES simulations. *Climate Dynamics*, 42: 2287-2303.
- Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., & Touré, A. (2011). Assessment and enhancement of MERRA land surface hydrology estimates. *Journal of Climate*, 24: 6322-6338.

- Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, 89(3): 303-311.
- Rousseau, M., Cerdan, O., Delestre, O., Dupros, F., James, F., & Cordier, S. (2015). Overland flow modeling with the shallow water equations using a well-balanced numerical scheme: Better predictions of just more complexity. *Journal of Hydrologic Engineering*, 20(10): 10.1061/(ASCE)HE.1943-5584.0001171 , 04015012.
- Salas, J. D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19: 554-568.
- Sanford, T., Frumhoff, P. C., Luers, A. & Gullett, J. (2014). The climate policy narrative for a dangerously warming world. *Nature Climate Change*, 4: 164-166.
- Sauchyn, D., Vanstone, J., & Perez-Valdivia, C. (2011). Modes and forcing of hydroclimatic variability in the upper North Saskatchewan River Basin since 1063. *Canadian Water Resources Journal*, 36(3): 205-217.
- Schoetter, R., Hoffmann, P., Rechid, D., & Heinke Schlünzen, K. (2012). Evaluation and bias correction of regional climate model results using model evaluation measures. *Journal of Applied Meteorology*, 51: 1670-1684.
- Schwalm, C. R., Huntzger, D. N., Cook, R. B., Wei, Y., Baker, I. T., Nelson, R. P.,...Zeng, N. (2014). A model-data intercomparison of simulated runoff in the contiguous United States: results from the North America Carbon Regional and Continental Interim-Synthesis. *Biogeosciences Discussions*, 11: 1801-1826.
- Sheffield, J., Barrett, A. P., Colle, B., Fernando, D. N., Fu, R., Geil, K. L.,...Yin, L. (2013). North American Climate in CMIP5 Experiments. Part I: Evaluation of Historical Simulations of Continental and Regional Climatology. *Journal of Climate*, 26: 9209–9245.
- Sheffield, J., Wood, E. F., & Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature*, 491: 435-438.
- Shrestha, R. R., Dibike, Y. B., & Prowse, T.D. (2012). Modelling of climate-induced hydrologic changes in the Lake Winnipeg Watershed. *Journal of Great Lake Research*, 38: 83-94.
- Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extreme indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research, Atmospheres*, 118: 1716-1733. doi: 10.1002/jgrd.50203.
- Slota, P. (2013). Evaluation of operational lake evaporation methods in a Canadian Shield landscape. Master of Science Thesis, Department of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada, Retrieved from MSpace (<http://hdl.handle.net/1993/22103>). 198pp.

- Soulis E. D., Snelgrove, K. R., Kouwen, N., Seglenieks, F., & Verseghy, D. L. (2000). Towards closing the vertical water balance in Canadian atmospheric models: coupling of the land surface scheme class with the distributed hydrological model watflood. *Atmosphere-Ocean*, 38(1): 251-269.
- Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., & Bierrkens, M. F. P. (2012a). Global patterns of change in discharge regimes for 2100. *Hydrology and Earth System Science*, 16: 1047-1062.
- Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., & Bierrkens, M. F. P., (2012b). On the suitability of GCM runoff fields for river discharge modeling: A case study using model output from HadGEM2 and ECHAM5. *Journal of Hydrometeorology*, 13: 140-154.
- Stadnyk, T., & Newsom, D. (2013). Nelson-Churchill River Basin WATFLOOD MAP file Set-Up. Documentation in partial fulfilment of Environment Canada contract K3D35-13-1385, University of Manitoba, Winnipeg, MB. 24p.
- Statistics Canada, (2010). Human activity and the environment: Freshwater supply and demand in Canada. Environment Accounts and Statistics Division. Catalogue no. 16-201-X, ISBN 1913-6743. Ottawa, Canada. 59pp.
- Stewart, R., Lawford, R., & Boisvert, A.-A. (2011). The 1999-2005 Canadian Prairie Droughts: Science, Impacts and Lessons, DRI 2011, Department of Environment and Geography, University of Manitoba. 114pp.
- St. George, S., & Nielsen, E. (2002). Hydroclimatic change in southern Manitoba since A.D. 1409 inferred from tree rings. *Quaternary Research*, 58: 103-111.
- St. George R. S. (2007). Hydrological and peleo-drought variability in the Winnipeg River Basin, Canada and the Canadian Prairies. Doctor of Philosophy Thesis, Department of Geosciences, The University of Arizona, Tucson, AZ. Retrieved from UA Campus Repository at: <http://hdl.handle.net/10150/194832>. 157pp.
- Sushama, L., Laprise, R., Caya, D., Frigon, A., & Slivitzky, M. (2006). Canadian RCM projected climate-change signal and its sensitivity to model errors. *International Journal of Climatology*, 26(15): 2141-2159.
- Szeto, K., Brimelow, J., Gysbers, P., & Stewart, R. (2015). The 2014 Extreme Flood on the Southeastern Canadian Prairies. In "Explaining Extremes of 2014 from a Climate Perspective". *Bulletin of the American Meteorological Society*, 96(12), S20-S24.
- Taylor, K.E. (2013). CMIP5 Standard Output. August 15, 2013. 133p. Accessed online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/standard\\_output.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.pdf) on May 11, 2015.
- Taylor, K. E., Stouffer, R. J., Meehl, G. A. (2012). An Overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93: 485-498.

- Thrasher, B. L., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correction climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences Discussion*, 9: 5515-5529.
- Todhunter, P. E. (2012). Uncertainty of the assumptions required for estimating the regulatory flood: Red River of the North. *Journal of Hydrologic Engineering*, 17: 1011-1020.
- Trenberth, K. E., (1997). The use and abuse of climate models. *Nature*, 386: 131-133.
- Trenberth, K. E., Dai, A., van der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield, J. (2014). Global warming and changes in drought. *Nature Climate Change*, 4: 17-22.
- U.S. Geological Survey, (2000). HYDRO1k Elevation Derivative Database, Cent. for Earth Resour. Obs. and Sci., Sioux Falls, S. D. (Available at <http://edc.usgs.gov/products/elevation/gtopo30/hydro/>)
- van Huijgevoort, M. H. J., van Lanen, H. A. J., Teuling, A. J., & Uijlenhoet, R. (2014). Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge. *Journal of Hydrology*, 512: 421-434.
- van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K.,...Rosevan, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, 109: 5-31.
- Vaze, J., Post, D. A., Chiew, F. H. S., Parraud, J.-M., Viney, N. R., & Teng, J. (2010). Climate non-stationarity – validity of calibrated rainfall-runoff models for use in climate change studies. *Journal of Hydrology*, 394: 447-457.
- Verseghy, D. L. (2000). The Canadian land surface scheme (CLASS): Its history and future. *Atmosphere-Ocean*, 38(1): 1-13.
- Watanabe, S., Kanae, S., Seto, S., Yeh, P. J.-F., Hirabayashi, Y., & Oki, T. (2012). Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models. *Journal of Geophysical*, 117: D23114.
- Watterson, I. G., Bathols, J., & Heady, C. (2014). What influences the skill of climate models over the continents? *Bulletin of the American Meteorological Society*, 95: 689-700.
- Werner, A. T., & Cannon, A. J. (2015). Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth Systems Sciences Discussions*, 12: 6179-6239.
- Werner, A. T. (2011). BCSD Downscaled transient climate projections for eight select GCMs over British Columbia, Canada. Pacific Climate Impacts Consortium, University of Victoria, BC, 63pp.

- Whitfield, P. H. (2012). Floods in future climates: a review. *Journal of Flood Risk Management*.  
Doi: 10.1111/j.1753-318X.2012.01150.x
- Wilke, R. A. I., Mendlik, T., & A. Gobiet, A. (2013), Multi-variable error correction of regional climate models. *Climatic Change*, 120: 871-887
- Wolfe, B. B., Hall, R. I., Last, W. M., Edwards, T. W. D., English, M. C., Karst-Riddoch, T. L.,...Palmini, R. (2006). Reconstruction of multi-century flood histories from oxbow lake sediments, Peace-Athabasca Delta, Canada. *Hydrological Processes*, 20: 4131-4153.
- Wood, A. W., Lettenmaier, D. P., & Palmer, R.N. (1997). Assessing climate change implications for water resourced planning. *Climatic Change*, 37: 203-228.
- Woodhouse C. A., Meko, D. M., MacDonald, G. M., Stahle, D. W., & Cook, E. R. (2010). A 1,200 year perspective of 21<sup>st</sup> century drought in southwestern North America. *Proceedings of the National Academy of Science*, 107(50): 21283-21288.
- Wruth, S. G. (2013). Incorporation of non-stationary landcover into WATFLOOD climate change scenarios. Master of Science Thesis, Department of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada, Retrieved from MSpace (<http://hdl.handle.net/1993/23208>). 159pp.
- Wu, H., Kimball, J. S., Li, H., Huang, M., Leung, L. R., & Adler, R. F. (2012). A new global river network database for macroscale hydrologic modeling. *Water Resources Research*, 48(9): W09701.
- Wu, H., Kimball, J. S., Mantua, N., & Stanford, J. (2011). Automated upscaling of river networks for macroscale hydrological modeling. *Water Resources Research*, 47, W03517.
- Yevjevich, V., (1967). An objective approach to definitions and investigations of continental hydrologic droughts. *Paper 23*, Colorado State Univ., For Collins, CO.
- Zhou, X., Zhang, Z., Wang, Y., Zhang, H., Vaze, J., Zhang, L.,...Zhou, Y. (2012). Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. *Journal of Hydrology*, 470-471: 269-279.