Impact of climate change on urban water consumption: a case study for Greater Montreal

Niousha RasiFaghihi

A Thesis

in

The Department

of

Building Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Civil Engineering) at

Concordia University

Montréal, Québec, Canada

December 2019

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Niousha RasiFaghihi**

Entitled: **Impact of climate change on urban water consumption: a case study for Greater Montreal**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Civil Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair and Examiner
Dr. Mazdak Nik-Bakht

_____ External Examiner
Dr. Marius Paraschivoiu

_____ Examiner
Dr. Ali Nazemi

_____ Co-supervisor
Dr. S. Samuel Li

_____ Co-supervisor
Dr. Fariborz Haghighat

Approved by _____
Ashutosh Bagchi, Chair
Department of Building Civil and Environmental Engineering

_____ 2019

_____
Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Impact of climate change on urban water consumption: a case study for Greater Montreal

Niousha RasiFaghihi

Smart cities need a sustainable plan and management of urban water consumption (UWC). A reliable long-term forecast of UWC is one of the key tasks to ensure water security and to achieve a balance between future water demand and supply. Long-term forecasts of UWC inevitably contain uncertainties. The uncertainties can associate with: 1) historical data of UWC; 2) existing observations of hydrological/climate variables as potential drivers of UWC; 3) the dependence of UWC on the potential drivers; and 4) projections of future climate change. The purpose of this research is to improve our understanding of the climate-change impact on UWC and of feasible ways to handle the foregoing uncertainties. Specifically, this research seeks answers to two key questions: 1) What quantity of water will be needed in the long-term? 2) To what extent will long-term UWC be affected by climate change? This research took the probabilistic approach to the problem of UWC forecast and made an application to the City of Brossard in the Greater Montreal metropolitan area. The methodologies involve Bayesian statistics as well as cluster analyses, which are a frequently used technique in machine learning. The analyses were performed on multiple year records of daily water consumption (DWC) in the city as well as field measurements of climate variables from the Montreal area. The analyses produced results of decomposed base water consumption (BWC) and seasonal water consumption (SWC). The DWC time-series was shown to have a transition from BWC to SWC at a threshold air temperature. The BWC was independent of climate change but subject to weekend effects, being higher on a weekend than weekdays. The SWC was a function of daily minimum air-temperature, daily maximum air-temperature, and daily total precipitation. The SWC forecasts allowed for inherent uncertainties in climate variables. Markov Chain Monte Carlo

was used as a sampling method in approximating the posterior distribution of regression parameters. The results from Bayesian linear regression gave a probability distribution of DWC. To quantify the impact of climate change on UWC, future projections of air temperature and precipitation were obtained from 21 General Circulation Models and downscaled for the city. The downscaled daily air temperature and precipitation corresponded to two scenarios of levels of greenhouse gas concentrations. Using quantile mapping methods, bias corrections were made to the downscaled daily minimum temperature, daily maximum temperature and daily total precipitation. These data gave input to the Bayesian linear regression model and produced SWC forecasts for the next three decades. The SWC was shown to display a positive trend over time in response to changing climate. The methodologies discussed in this thesis can be applied to other cities, producing results useful for upgrade and/or construction planning of water supply infrastructures.

Acknowledgments

I would like to thank my supervisors, Dr. S. Samuel Li and Dr. Fariborz Haghighat, and express my great appreciation for their guidance, patience, and support during these two years of my Master studies. Their feedback has been invaluable in shaping this work. Studying under their supervision and working within their research group have been a golden opportunity for me. My sincere thanks must also go to the members of my thesis advisory and exam committee: Dr. Ali Nazemi, Dr. Mazdak Nik-Bakht and Dr. Marius Paraschivoiu. They generously gave their time to offer me valuable comments toward improving my work.

I send my sincerest regards and emotions to my husband, Amir Saman Jalali Afshar, for his unconditional trust, timely encouragement, and endless patience. I also express my gratitude to my parents and brother for their unfailing emotional support. My family have been the most significant support for me throughout my journey at Concordia University.

Special thanks to my friends and peers, especially my best friend, Golnoosh Karimipourfard, who went through hard times with me, cheered me up, and celebrated each accomplishment.

I want to thank Mr. J. P. Richard of the City of Longueuil for providing data of daily water use and Mr. T. Aldea for his help in the data acquisition process.

List of Symbols

θ	Daily minimum temperature (°C)
Θ	Daily maximum temperature (°C)
Θ_0	Daily mean temperature (°C)
p	Daily precipitation (mm)
q	Daily urban water consumption ($m^3/capita.day$)
μ	Mean value
σ	Standard deviation
Θ_t	Temperature threshold (°C)
Z_i	Z-score
S	Silhouette coefficient
β_0	Regression coefficient
β_1	Regression coefficient
β_2	Regression coefficient
β_3	Regression coefficient
ϵ	Regression error term
H_{Gf}	Daily future GCM projection
H_{Gd}	Daily downscaled future data
$R(x)$	Change factor function for downscaling

List of Acronyms

UWC	Urban Water Consumption
DWC	Daily Water Consumption
SWC	Seasonal Water Consumption
BWC	Base Water Consumption

Contents

List of Figures

List of Tables

Chapter 1

Thesis Overview

1.1 Background

It is challenging to achieve sustainable management of UWC. The challenges are due to urbanism, climate change and limited freshwater resources. Traditionally, water shortages have been particularly problematic for water-scarce countries. However, countries with abundant freshwater resources like Canada face the same issue in the past few decades. Canada has freshwater resources in different forms of water bodies. However, sustainable management is a significant issue in major cities and municipalities. Thus, it is crucial to study sustainable strategies for UWC.

Long-term forecasting is necessary for enhancing water security and keeping the balance between supply and demand. In the years following World War II, as the water was considered an inexpensive commodity, excess capacity was developed to avoid risks of water shortage Rinaudo [2015]. However, economic and population growth continued, and water resources were not readily available. On the other hand, in some parts of the world (the western US states), an unanticipated decline in per capita water use led to costly oversized water supply systems Rinaudo [2015]. Therefore, researches in this field followed the path through water demand forecasting. Long-term forecasting is useful for infrastructure and capital planning and can provide valuable information for determining efficient pricing Buck, Soldati, and Sunding [2015]. Accurate forecasts of UWC lead a profound understanding of past till present correlations of water consumption drivers with water consumption for water resources planning and management in the long run.

Water demand projections are subject to a large uncertainty about the future, as well as the causes of historical and recent trends in water consumption patterns. The uncertainties can associate with: 1) historical data of UWC, 2) existing observations of hydrological/climate variables as potential drivers of UWC, 3) the dependence of UWC on the potential drivers, and 4) projections of future climate change Tiwari and Adamowski [2014]. Inaccurate forecasts would result in costs to water utilities, water consumers, and the environment. For instance, the associated uncertainties in the estimates of water demand in California by 2040 were addressed by increasing total demand up to 10% Miro, Groves, Catt, Miller, and Social [2018]. Recently, researches developed multiple scenarios and complex uncertainty analyses to provide insights into the range of possible future outcomes.

Climate change impacts have consequences for human health, assets and livelihood. The impacts on each city depend on the actual changes in climate, such as higher annual temperature or increased precipitation, and vary from place to place. Also, the water consumption pattern of households can be influenced by changes in outdoor weather condition. According to Canada's Changing Climate Report (CCCR) 2019, between 1948 and 2016, the best estimate of mean annual temperature increase is $1.7°$C for Canada as a whole and $2.3°$C for northern Canada Zhang et al. [2019]. Also, there is high confidence that daily extreme precipitation is projected to increase in future Zhang et al. [2019]. UWC is not equally distributed over time and space as households and services tend to demand more water in hot and dry periods. There is a particular need to manage water consumption, including the control of both peak and daily-averaged water demands, and help plan future consumption for growing cities under a changing climate.

1.2 Objectives

This research explores the impact of climate change on DWC in the next three decades. A stochastic methodology is developed to forecast residential water demand taking associated uncertainty into account. The application of the proposed model is examined in the City of Brossard (Greater Montreal) in the Canadian Province of Quebec. The main objectives of this thesis are as follows:

- To reveal the effects of climate change on UWC in long-term.
- To develop a prediction model for forecasting of DWC addressing uncertainties.

1.3 Major contributions

The contributions of this research are:

- Decomposition of UWC into main components that are affected by distinct variables using a machine learning approach.
- Quantification of uncertainties associated with random fluctuations of the climatic variable in future by using an ensemble model of General Circulation Models (GCMs).
- Development of a stochastic model for predicting DWC. In this model, DWC is a function of weather variables.
- Analysis of historical and future weather data from GCMs. The data is spatially downscaled to find the future projection for the area of study.
- Forecast of DWC as a probability distribution by 2050. The probability distribution embodies uncertainties in weather variables as well as the dependence of UWC on the potential drivers.

1.4 Organization of the thesis

To provide the relevant context, the rest of the thesis is organized as follows:

- Chapter 2 provides a comprehensive background and literature review on UWC forecasting, as well as the science of climate change assessment. Various forecasting approaches along with their application are also described.
- Chapter 3 describes the collected data and presents the methodologies. It includes the UWC decomposition and explains the applied algorithm. The steps toward building the prediction model, downscaling the GCMs and forecasting DWC are also provided.
- Chapter 4 presents the results of UWC decomposition, spatial downscaling and the predictor model. This chapter provides discussion of uncertainties in the long-term forecast of UWC.

- Chapter 5 concludes the thesis and provides some directions for future research.

Chapter 2

Background and Literature Review

In this chapter, a review of UWC forecasting is given. The main elements of the science of climate change impact assessment are also addressed. Besides, various methodologies for water consumption forecasting along with their limitations are discussed.

2.1 UWC components

UWC is affected by various climatic and socio-economic factors that are different and specific from place to place G. Singh, Goel, and Choudhary [2015]. It is far from being straightforward to distinguish the influence of climatic factors from that of the socio-economic. Some of the early studies decomposed UWC into two components: a) BWC; and b) SWC Eslamian, Li, and Haghighat [2016]; Gato, Jayasuriya, Hadgraft, and Roberts [2005]; Gato, Jayasuriya, and Roberts [2007]; Wong, Zhang, and Chen [2010]. In the early studies, the indoor water use was taken as BWC, whereas the outdoor water use was taken as SWC. BWC was thought to be insensitive to climate effects Eslamian et al. [2016]; Gato et al. [2007]; House-Peters, Pratt, and Chang [2010]; Syme, Shao, Po, and Campbell [2004]; Wong et al. [2010]; Zhou, McMahon, Walton, and Lewis [2000], but sensitive to the socio-economic factors. The socio-economic factors include population, household income, and water price (or consumption payment). SWC was considered as dependent on changes in air temperature and precipitation. In most of the cases, the forecast results reflect the conditions of application regions and hence are site-specific.

It is worth to mention that in the majority of studies, the actual amount of households water use in individual were not available; therefore, similar to other studies Chang, Praskievicz, and Parandvash [2014]; Eslamian et al. [2016]; Hamlet et al. [2010]; Parandvash and Chang [2016], water used per capita per day was estimated by dividing daily total supply of water for the city by its population.

2.1.1 Transition from BWC to SWC

Some studies have simply assumed that BWC is equal to water consumption in winter House-Peters et al. [2010]; Praskievicz and Chang [2009]; Syme et al. [2004]. Eslamian et al. [2016] discussed two alternative methods for determining BWC. One approach is to identify a BWC curve based on water use during the winter months. If the curve exhibits an upward or downward trend over time, the curve is described by a polynomial function of time Eslamian et al. [2016]; Wong et al. [2010]; Zhou et al. [2000]. Then, the transition from BWC to SWC is determined based on the calendar date. Given that the transition point can be quite sensitive to weather conditions in some regions, the alternative method is to define threshold values of air temperature and precipitation, at which the transition occurs, and water consumption starts to be dependent on climatic variables Gato et al. [2007]. Discerning the transition is more robust in the second than the first method. The reason is that the second method uses physical variables (air temperature and precipitation), as opposed to calendar date applied in the first method.

A number of studies tried to identify the temperature threshold level to discriminate BWC and SWC. Eslamian et al. [2016] plotted daily water use against daily maximum temperature and suggested 10°C as the daily maximum temperature threshold in Montreal, Canada. Gato et al. [2007] failed to identify the temperature threshold by regressing the daily water use against daily maximum temperature. Therefore, they proposed a polynomial function of daily maximum temperature against the reciprocal of the corresponding daily water use. By taking the derivative of the polynomial function and setting the derivative to zero yields, Gato et al. [2007] determined the daily maximum temperature threshold as 15.27°C. In another study, the use of the wavelet analysis gave a daily mean temperature threshold of 12°C in Montreal, Canada Tiwari and Adamowski [2013].

2.2 Climate change impact assessment

As mentioned before, SWC is sensitive to the changes in weather condition. Parandvash and Chang [2016] suggested that any change in climate can affect UWC in the long-term. Several researchers Haque, Rahman, Hagare, and Kibria [2014]; Khatri and Vairavamoorthy [2009]; Parandvash and Chang [2016]; Ruth, Bernier, Jollands, and Golubiewski [2007] have reported the impact of climate variabilities on water demand. It is essential to account for uncertainties due to a wide range of possible greenhouse gases emission scenarios, the variability of different global climate models, downscaling, and uncertainties in hydrological and impact models Khatri and Vairavamoorthy [2009]. Commonly, economic and demographic factors are used in long-term forecast of water demand, whereas climatic factors are used in the short-term forecast Khatri and Vairavamoorthy [2009].

2.2.1 General circulation models

There are various General Circulation Models (GCMs) that are developed around the globe to look at the past and future evolutions in the global climate. GCMs represent the three-dimensional climate system by describing movement of energy, momentum and the conservation of mass and water vapour Alexander et al. [2009]. GCMs give the notion of future climate in a coarse-scale, and they are usually presented under different greenhouse gasses emission scenarios known as Representative Concentration Pathways (RCPs) Meinshausen et al. [2011]. The GCMs cover the atmosphere by coarse grids. The horizontal spatial resolutions are typically 250 to 600 km, and the vertical resolutions are 10 to 20 layers. GCMs may produce projections that are globally accurate but locally biased in their descriptive statistics (i.e., mean, variance, and so forth).

2.2.2 Representative concentration pathways

The Coupled Model Intercomparison Project Phase 5 provides new pathways for projecting future climate that is called RCPs and provide four scenarios for various levels of greenhouse gas concentrations. The four pathways also refer to the amount of total radiative forcing that is experienced until the year 2100. These four different climate scenarios have been labelled RCP 2.6,

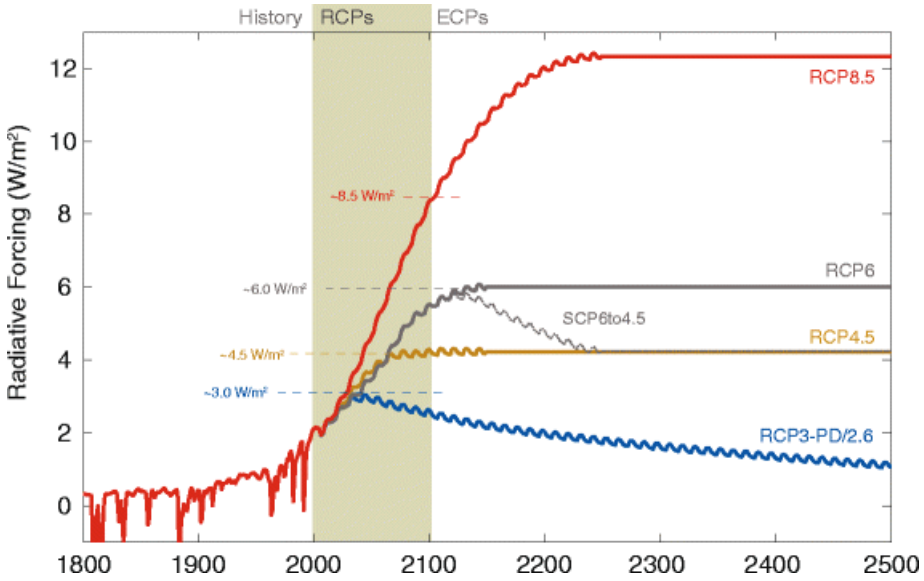

Figure 2.1: *RCPs based on radiative forcing Meinshausen et al. [2011]*

RCP 4.5, RCP 6, and RCP 8.5, and are based on the change in the radiative forcing compared to pre-industrial conditions with the rate of +2.6, +4.5, +6.0, and +8.5 W/m^2, respectively. Fig. 2.1 displays the behaviour of all future scenarios of RCPs. RCP 4.5 is a stabilization scenario with policies for limiting emissions and radiative forcing Thomson et al. [2011], whereas RCP 8.5 is the worst-case scenario Riahi et al. [2011].

Several recent studies have addressed the impacts of climate change on water resources and drought risk. The vulnerability-based assessment of global freshwater availability of Koutroulis et al. [2019] considered climate change for RCP 8.5. Zhuang, Li, Nie, Fan, and Huang [2018] proposed a method for evaluating climate change impacts on water resources using multi-ensemble GCMs. J. Liu et al. [2017]; Prusty, Das, and Patra [2018] analyzed water management and allocation on the scale of a river basin considering the impact of climate change under RCP 2.6, RCP 4.5 and RCP 8.5. Wang, Duan, Liu, Li, and Feng [2019] assessed drought response to climate change. They forecast drought tendency using three GCMs for RCP 4.5 and RCP 8.5. Ahmadalipour, Moradkhani, Castelletti, and Magliocca [2019] investigated drought risk by quantifying drought hazard using an ensemble of 10 regional climate models for RCP 4.5 and RCP 8.5.

2.2.3 Downscaling of climate model projections

GCMs grid resolution is too coarse to be used at the local and/or regional scale. Impact assessment needs to be done at a much finer spatiotemporal resolution. A standard method to derive the high resolution information needed by the impact model from the coarse-scale resolution is downscaling Boé, Terray, Habets, and Martin [2007]. Downscaling the process of relating data at relatively coarse spatial and temporal scales to desired products at finer spatial and temporal scales. There are two types of downscaling, spatial and temporal. Temporal downscaling is the derivation of fine-scale temporal data from coarser-scale temporal information. Spatial downscaling refers to transforming simulated climate patterns at a coarse grid to a finer spatial resolution of local interest H. Li, Sheffield, and Wood [2010]. Spatial downscaling can be divided into two main approaches, namely dynamical and statistical downscaling. Dynamical downscaling refers to using regional climate models and uses the GCMs as boundary conditions to increase the resolution.

2.2.4 Bias correction downscaling

The bias correction downscaling method is one of the statistical downscaling methods. This mathematical approach is not only straightforward and fast but also, can be preferable for large weather data sets (e.g. for 30 years and more). In this method, the gridded observation parameters are aggregated to the GCM grid scale. Then, using the quantile mapping, bias in the GCM data is removed. Bennett et al. [2014] have shown that quantile mapping is effective at correcting climate model biases across a range of values and variables. Fig. 2.2 shows the procedure for performing the quantile mapping method. The quantile mapping method is defined based on the one-by-one mapping between the cumulative distribution functions (CDF) of the historical GCM data and the observed data. Bias correction methods have been developed in different literature Boé et al. [2007]; H. Li et al. [2010]; Schmidli, Frei, and Vidale [2006].

2.2.5 NASA Earth Exchange Global Daily Downscaled Projections

NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP) data set is comprised of downscaled climate scenarios for the globe that are derived from the GCM runs conducted

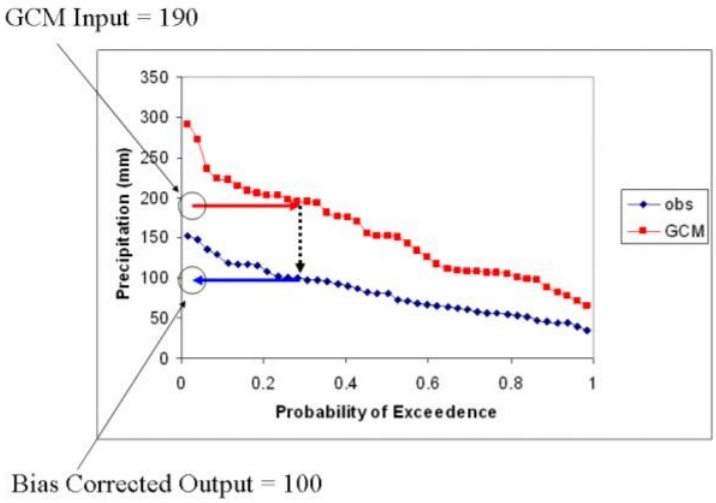

Figure 2.2: *The procedure of bias correction quantile mapping downscaling method Hamlet et al. [2010]*

under Coupled Model Intercomparison Project Phase 5. This data set is an archive of downscaled climate scenarios for the conterminous United States at a 0.25-degree (25 km in 25 km) spatial resolution.The archived climate scenarios were derived from the GCMs implemented under the Coupled Model Intercomparison Project Phase 5 Taylor, Stouffer, and Meehl [2012] and developed in support of the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. The archived results are downscaled projections for RCP 4.5 and RCP 8.5 from the 21 models and comprise daily maximum temperature, daily minimum temperature and precipitation from the time period of 1950 to 2005 and of 2006 to 2099.

As per Jain, Salunke, Mishra, Sahany, and Choudhary [2019], compared to Coupled Model Intercomparison Project Phase 5, the NEX-GDDP data captures the spatial patterns of seasonal mean temperatures and precipitation with highest accuracy and the inter-annual variations in precipitation are closer to the observations. According to NEX-GDDP data set, the combined effects of precipitation and temperature variations reveals severe future drought in the western United States Ahmadalipour, Moradkhani, and Svoboda [2017]. Mandapaka and Lo [2018] conducted a study on the assessment of future changes of precipitation in Southeast Asian using the NEX-GDDP data set. They indicated the substantial increases in mean and extreme precipitation by the end of 21st century under RCP 4.5 and RCP 8.5. Raghavan, Hur, and Liong [2018] evaluated NEX-GDDP data set, compared to gridded observations on daily scales in Southeast Asia. They concluded that

NEX-GDDP are in better agreement with monthly observations over the historical rather than daily scale.

2.3 UWC forecasting

The topic of forecasting UWC is of importance to proper planning and sustainable development of urban centres. Early researchers have attempted to forecast UWC as a function of a variety of influence factors, including climatic as well as socio-economic factors. Various methods in literature assess the correlation between UWC and different variables. These methods are ranging from the most straightforward methods, such as linear regression to the most sophisticated ones. Although complex models provide accurate predictions, they require fine-scale data. In the previous literature, researchers demonstrated that water consumption has a stochastic non-stationary pattern; hence, they came up with various ideas and methodologies to interpret stochastic water demand time series.

There is a particular need for forecasting the future water demand addressing uncertainties associated with 1) individual influence factors (a lack or incompleteness of data; data outliers; mathematical models; parameters subject to future changes); 2) a combination of some of these factors or all of them. Forecasting models can be categorized into deterministic and probabilistic, wherein the deterministic approaches the stochastic nature of predictor variables are not considered Almutaz, Ajbar, Khalid, and Ali [2012]; Stoker and Rothfeder [2014]. In probabilistic or stochastic approaches, the uncertainties associated with the influencing variables themselves, including climatic and socio-economic variables, and with their correlation are considered; accordingly, the results can be potentially adequate and efficient. In the literature, Monte Carlo simulation is the only stochastic approach in the area of water demand forecasting which was applied by a number of studies Almutaz et al. [2012]; Haque et al. [2014]; Khatri and Vairavamoorthy [2009]. Moreover, bootstrap is the other methodology that was proved to be capable of addressing uncertainties Khatri and Vairavamoorthy [2009]; Tiwari and Adamowski [2013, 2014].

2.4 UWC forecasting using machine learning approaches

In this section, it is tried to collect the most recent papers that implied any of machine learning approaches to make prediction or forecasting of a variable in different subjects. Apparently, there are many studies in the area of building energy consumption in the last decade that is developed based on machine learning algorithms (Table 2.1). In Table 2.1, eight papers have been reviewed and categorized into four groups based on their objectives. For each paper, applied tools and objectives have been listed. The most well-known and conventional machine learning approaches are included in separate columns.

Table 2.1: *Summary of machine learning approaches in buildings'energy consumption*

Reference	Tools	Objectives	Application of decision tree	Application of BN	Application of ARM	Application of ANN	Application of clustering
Liu et al. (2018); Ashouri et al. (2018)	Clustering, ARM, ANN	Finding the sources of high and low energy consumption	N/A	N/A	Assessing correlations	Predicting energy consumption	Grouping buildings considering weather conditions and finding energy consumption of industries
Singh & Yassine (2018); Yu et al. (2012)	Clustering, ARM, BN	Examining correlations among building operational data	N/A	Predicting multiple appliances usage in future	Assessing correlations	N/A	Uncovering appliance-to-time association
Yu et al. (2011a); Yu et al. (2011b)	Clustering, decision tree, ARM	Examining influence of occupants' behavior on building energy consumption	Indicating whether the supply hot water load is high or low	N/A	Assessing correlations among building operational data	N/A	Examining effects of occupants' behavior on building energy consumption
Yu et al. (2013); Yu et al. (2010)	Clustering, decision tree, ARM	Predicting energy demand and occupants' behaviour	Predicting building energy demand	N/A	Assessing correlations among building operational data	N/A	Examining the effects of occupants' behavior on building energy consumption

As per Table 2.1, one can notice that there is a substantial similarity between water and energy consumption correlation with their influencing factors. Since both contexts consist of a massive amount of data, the approaches to forecast demand of these utilities can be considered similar. Interestingly, researchers in the field of building energy consumption have included variables such as weather and socio-economic factors in their forecasting models.

Two well-known terminologies in the machine learning context are supervised and unsupervised learning. Supervised learning is considered as a synonym for classification; whereas, unsupervised learning can be a synonym for clustering Han, Pei, and Kamber [2011]. In supervised learning, the training data set has a labeled example; while in unsupervised learning, the input examples are not class labeled; therefore, classes within the data need to be discovered Han et al. [2011]. In the following subsections, various supervised and unsupervised methodologies are discussed. Also, approaches that are not classified in any of the learning categories are covered.

2.4.1 Decision tree

Decision tree is one of the commonly used classification methodologies or supervised training techniques in many scientific fields. It is simple to use and capable of generating an accurate model that is understandable and interpretable Z. Yu, Haghighat, Fung, and Yoshino [2010]. Decision tree consists of three kinds of nodes: root, internal, and leaf. A root node and internal node represents a binary split test on an attribute, while a leaf node denotes an outcome of the classification, and thus holding a categorical target label. Decision tree is generated in two main steps; learning and classification. In classification, the data set split into training and test data; then, the training data is analyzed by the algorithm to generate decision tree. In the next step, the accuracy of the obtained decision tree is estimating using test data. Although, decision tree is capable of possessing both categorical and numerical data; however, they are more appropriate in predicting categorical variables Z. Yu, Fung, Haghighat, Yoshino, and Morofsky [2011].

2.4.2 Bayesian Networks

Bayesian Networks (BNs) are the graphical model which are based on Bayes' theorem and are capable of modeling probabilistic relationships among a set of variables Heckerman [1997]. This

methodology is considered as a classification approach and can also be used for prediction Sriram [n.d.]. One of the most significant characteristics of BNs is their ability to account for the uncertainty associated with inaccurate and incomplete databases. In a BN, variables of interest are represented by nodes, and the links between them indicate informational or causal dependencies among them Ismail, Sadiq, Soleymani, and Tesfamariam [2011]. Constructing a BN consists of two steps, 1) structure learning: finding a graphical structure of dependencies between nodes and 2) parameter learning: defining conditional probability distribution among nodes.

2.4.3 Associated Rule Mining

Associated Rule Mining (ARM) is a data mining approach that is capable of representing the patterns of parameters that are frequently associated together G. Liu, Yang, Hao, and Zhang [2018]; Z. Yu et al. [2011]. In this unsupervised training approach, there are two common-used terminologies; support and confidence to denote the validity and certainty of an association rule. ARM is an unsupervised learning process which is implemented in items frequently associated and represents the frequency of two items happen together Ashouri, Haghighat, Fung, Lazrak, and Yoshino [2018]. Technically, in performing ARM, the value of quantitative attributes requires to be classified into categorical values Ashouri et al. [2018]; Z. Yu et al. [2011].

2.4.4 Artificial Neural Network

Artificial Neural Network (ANN) has become a feasible, multipurpose methodology with potential influence on any discipline. ANNs are considered as predictive and unsupervised learning models that are built based on inputs and outputs of historical data and is able to predict outputs of new input Fayyad, Piatetsky-Shapiro, and Smyth [1996]. Even though, ANNs are capable of solving highly nonlinear and complex problems; however, they have the shortcoming of being black boxes, meaning that there is no chance to interpret the process and calculations inside of them. This data-driven approach is capable of modeling the nonlinear relationship among the factors influencing water demand and implementing the trained ANN to forecast future water demand Ghiassi, Zimbra, and Saidane [2008].

2.4.5 Clustering analysis

Cluster analysis is the process of dividing the observations into classes or clusters so that objects in the same cluster have a high similarity, while objects in different clusters have low similarity. Clustering can be regarded as a form of classification that creates labeling of objects with cluster labels derive from data Z. J. Yu, Haghighat, Fung, Morofsky, and Yoshino [2011]. Hence this methodology might be referred to as unsupervised classification. Fundamental clustering methods can be classified as of the four categories: partitioning methods, hierarchical methods, density-based methods, and grid-based methods Han et al. [2011].

2.4.6 Coupled models

There are several developed approaches which adopted various machine learning approaches and coupled them to benefit from a more advanced and potential tool for machine learning. Z. Yu et al. [2011] proposed clustering, decision tree and ARM to study occupant's behaviour in residential buildings. Two years later, they applied an integrated models including decision tree, clustering and ARM for the same purpose as the previous study. Decision tree was developed as a predictive model to predict building energy demand; clustering adopted to examine only the effect of occupants' behaviour on building energy consumption; eventually, the associations and correlations were addressed using ARM. S. Singh and Yassine [2018] proposed clustering analysis, ARM, and BN to analyze and forecasted the energy time series to extract various temporal energy consumption patterns. Initially, the appliance-to-appliance association was uncovered using ARM; then, the appliance-to-time association was evaluated through clustering. Then, they utilized BN to forecast multiple appliance usage in the future. Ashouri et al. [2018] developed a machine learning model contacting clustering, ARM and ANN. They aimed to reduce the effect of weather condition by clustering, finding the correlation among the appliances by ARM and eventually, build ANN models based on the rules to make the prediction.

2.5 UWC studies

In this section, the aim is to give an overview of studies in the area of UWC prediction, forecast and management considering various climatic and socio-economic factors. As it was discussed in ection 2.3, there is a considerable amount of uncertainty associated with influence factors in the future and their impact on UWC. In addition, there is a limited number of studies that propose a method to forecast UWC. Table 2.2 summarizes the early literature that addressed uncertainties and forecast UWC. Likewise, Table 2.3 covers studies that do not consider uncertainties, but forecast UWC and Table 2.4 embody researches that only predict UWC without communicating uncertainties. In each table, tool(s), the purpose of study, predictor variables and contribution(s), are listed. Papers are descending ordered based on the date with the newest records listed first.

Table 2.2: *Summary of tools, purpose, predictor variables and contribution(s) of UWC studies that consider uncertainties and forecast UWC*

Reference	Tool(s)	Purpose	Predictor variables	Major contribution(s)
Tiwari & Adamowski (2017)	Coupled model: wavelet-bootstrap-neural	UWC short-term forecasting	T[1] and P[2]	A coupled model to account for uncertainties
Haque et al. (2014)	Monte Carlo simulation	UWC long-term forecasting for single and multiple dwelling residential sectors	T, P, water price water savings from conservation programs, and water restriction savings	A probabilistic long-term water demand forecasting and the effects of climate change in water demand
Tiwari & Adamowski (2013)	Coupled model: wavelet-bootstrap-neural	UWC short-term forecasting	T and P	A coupled model to account for uncertainties
Almutaz et al. (2012)	Mont Carlo simulation	UWC long-term forecasting	Household size, T, and household income	A probabilistic long-term water demand forecasting
Khatri & Vairavamoorthy (2009)	Mont Carlo simulation, regression	UWC forecasting considering climate change, population and economic growth	T, P and population	Scenario-based approaches to forecast future population, future climate change and future water demand addressing uncertainties

[1] Air temperature, [2] Precipitation

Table 2.3: *Summary of tools, purpose, predictor variables and contribution(s) of UWC studies that do not consider uncertainties but forecast UWC*

Reference	Tool(s)	Purpose	Predictor variables	Major contribution(s)
Mouatadid & Adamowski (2017)	ELM, ANN, SVM, MLR	UWC short-term forecasting	T and P	Machine learning approaches in UWC forecasting
Parandvash & Chang (2016)	Time series regression	UWC long-term forecasting considering the effects of climate change in urban versus suburban areas	T, P, population and unemployment rate	UWC forecasting considering the impacts of climate change and socio-economic vairables
Al-Zahrani & Abo-Monasar (2015)	Coupled model: Time series-ANN	UWC short-term forecasting	T, P, humidity and wind speed	A coupled model to forecast UWC
Adamowski et al. (2012)	Coupled model: wavelet-neural network	UWC short-term forecasting	T and P	A coupled model to forecast UWC
Adamowski & Karapataki (2010)	Regression analysis and ANN	Comparing multiple linear regression and three types of ANN to forecast short-term weekly UWC	T and P	Comparison between two commonly-used approaches in water demand prediction
Ghiassi et al. (2009)	ANN	UWC forecasting	N/A	Non-stationary stochastic analysis
Ruth et al. (2007)	Regression	UWC forecasting considering the effects of climate change	T, P, humidity, wind speed, gust speed and population	UWC forecast considering different scenarios for population and climate projection

Table 2.4: *Summary of tools, purpose, predictor variables and contribution(s) of UWC studies that predict UWC without communicating uncertainties*

Reference	Tool(s)	Purpose	Predictor variables	Major contribution(s)
Eslamian et al. (2016)	Multi regression	Water consumption behaviours and water use prediction	T, P, population, household income and water price	Decomposing UWC into BWC and SWC with a reliable estimate of BWC
Romano et al. (2014)	Linear mixed-effects model	Residential water demand considering variety of the drivers	Population, altitude, T and P, annual residential expenditure, household income, ownership, and location	Unusual socio-economic variables effects on UWC
Chang et al. (2014)	Autoregressive integrated moving average	Relationship between SWC and climate variables at multiple temporal scales	T, P, and population	Sensitivity of SWC to climate variables at multiple temporal scales
Stoker & Rothfeder (2014)	Ordinary least square regression	UWC prediction models for four different urban land use types	T, P, number of bedrooms, household characteristics and number of families	Water demand in different types of urban land uses

Table 2.4 – continued

Reference	Tool(s)	Purpose	Predictor variables	Major contribution(s)
Panagopoulos (2013)	Factor analysis	Influence of drivers on UWC	Water price, population, income, total number of active connections, T and P	Frequency domain model to employ demographic, socio-economic and hydrological data
Adamowski et al. (2013)	Wavelet transform	Relationships between temporal patterns in UWC and meteorological records	T and P	Changes in temporal pattern of UWC and its meteorological drivers
Wong et al. (2010),	Regression	Influence of drivers on UWC	T and P	Decomposing UWC to three components; BWC, SWC and calendrical consumption
House-Peters et al. (2010)	Least square multi regression	Influence of climate, spatical structure and sociodemographic on UWC in single-families	T, P, household income, education, population age and household characteristics	The effects of urban spatial structure, socio-economic and climatic variables on UWC

Table 2.4 – continued

Reference	Tool(s)	Purpose	Predictor variables	Major contribution(s)
Polebitski & Palmer (2009)	Regression	Influence of drivers on UWC	T, P, water price, policy during droughts and household characteristics	Household characteristics effects on UWC
Praskievicz & Chang (2009)	Regression	Influence of weather variables on SWC	T, P, wind speed, relative humidity, daylight length, and cloud cover	Variety of climatic variables
Kenney et al. (2008)	Regression	Statistical analysis of influence of weather variables on UWC	Water price, length of bill period, household contribution in sustainable water consumption, T, P, economic and household characteristics	Household characteristics effects on UWC
Gato et al. (2007)	Regression	A prediction model for UWC	T and P	Decomposing UWC into BWC and SWC
Gato et. al. (2005)	Time series regression	A prediction model for UWC	T and P	Decomposing UWC into BWC and SWC

The early studies of UWC can be grouped into four categories as per their objectives and the capability of the proposed methodologies. In the first category, the methodologies are suitable for estimates of UWC under the current socio-economic and/or climatic conditions Adamowski, Adamowski, and Prokoph [2013]; Adamowski, Fung Chan, Prasher, Ozga-Zielinski, and Sliusarieva [2012]; Adamowski and Karapataki [2010]; Arturo, Alvarez-Chavez, Ramos-Corella, and Soto-Hernandez [2017]; Chang et al. [2014]; Eslamian et al. [2016]; Gato et al. [2005, 2007]; Kenney, Goemans, Klein, Lowrey, and Reidy [2008]; Panagopoulos [2014]; Polebitski and Palmer [2009]; Praskievicz and Chang [2009]; Stoker and Rothfeder [2014]. Overall, regression is the most frequently used method to estimate UWC. Other proposed approaches are linear mixed effects model, factor analysis and wavelet transform. Literature in the first category are not concerned about the future usage of water.

In the secondary category, UWC forecast takes the econometric approach. Most commonly, UWC forecasting includes end-use forecast, time series forecast, and econometric forecast Almutaz et al. [2012]; Khatri and Vairavamoorthy [2009]; Zhou et al. [2000]. The end-use forecast gives water demand based on a forecast of water uses. The time series forecast determines water consumption directly, without considering other influence factors. The econometric approach estimates the historical relationships between independent variables, and the dependent variable, under the assumption that the relationships continue in the future. Parandvash and Chang [2016] obtained estimates of future per capita water demand, by assuming a consistent difference between the average weather effects on the demand over the historical and future periods. Mouatadid and Adamowski [2017] performed econometric forecast, using an artificial neural network, support vector machine, and regression models. Several other researchers used regression models for econometric forecast Al-Zahrani and Abo-Monasar [2015]; Parandvash and Chang [2016]; Ruth et al. [2007].

In the third category, UWC forecast uses stochastic methodologies. Only a limited number of researchers have developed probabilistic forecast models to account for the stochastic nature of water consumption. They performed Mont Carlo simulations, and assigned probability distributions to variables in order to communicate uncertainties Almutaz et al. [2012]; Haque et al. [2014]; Khatri and Vairavamoorthy [2009]. Haque et al. [2014] and Khatri and Vairavamoorthy [2009]

considered future climate variations. Other researchers have used the bootstrap simulation methodology to address uncertainties Al-Zahrani and Abo-Monasar [2015]; Haque et al. [2014]; Tiwari and Adamowski [2014].

In the fourth category, UWC forecast considers climate change. Available GCMs predict future climate at a coarse-scale under different greenhouse gases emission scenarios. Parandvash and Chang [2016] compared the impacts of climate change on urban and suburban water demand. They used three GCMs to represent a set of 30 years of historical climate projections and 30 years of future projection. Using future projection from rescaled models under two emission scenarios, Ruth et al. [2007] analyzed the impact of climate change on water consumption in the long-term. Very few studied have been carried out to forecast future UWC using the probabilistic approach and GCMs estimate long-term climate variables.

2.6 Applied methodologies in urban water consumption studies

2.6.1 Regression

Regression is a widely utilized statistical technique in various disciplines and is well-known due to its simplicity. As opposed to the classification that predicts categorical labels, regression can model continuous-valued functions. Besides, it is capable of predicting missing or unavailable numerical data values rather than (discrete) class labels Han et al. [2011]. There are different applications of regression in the literature. According to Gato et al. [2005] and Parandvash and Chang [2016], structural time series regression model was adopted to represent the demand for water, based on its drivers to predict future water demand. Ordinary least square (OLS) multiple linear regression is a statistical method for studying the relationship between a single dependent variable and one or more independent variables Allison [1999]. Ordinary differentiates the simplest and the most complicated of least square, least square indicate the approach to determine regression equation, and linear serves to describe linear regression equation and multiple refers to two or more independent variables. Therefore, different combinations of these words might be seen in the literature, such as multiple regression, linear regression, least square regression and, so forth.

2.6.2 Linear Mixed Models

In linear models, it is assumed that observations are drawn from the same population and are independent; whereas in linear mixed models, observations are not independent, and they handle the data in a complex multilevel or hierarchical structure Romano, Salvati, and Guerrini [2014]. In this approach, observations are grouped in various levels (clusters), which are dependent within the same level and independent among different levels. Linear mixed models consider random effects, hierarchical effects, repeated measures and spatial correlations in various applications to address sources of variations. As per Romano et al. [2014], two linear mixed models are fitted, one to estimate consumption and the other to measure tariff independently.

2.6.3 Wavelet Transform

Wavelet Transform (WT) is an approach to transform time series data into the frequency domain to remove the influence of time in the data. In order to perform this task, various shapes and sizes of short filtering functions called wavelets have been applied. Hence, unlike the Fourier transform, where a sinusoidal wave function has been frequently used as the basis of decomposition, in WT, other basis functions can be selected according to the attributes of the variables Kim and Melhem [2004]. WT is a potential method for detecting patterns in UWC both in terms of the wavelength of cycles and the time of occurrence of them Adamowski et al. [2013]. Adamowski et al. [2013] adopted WT to identify the association between UWC and two climatic variables; precipitation and temperature.

2.6.4 Support Vector Machine

Besides ANNs, Support Vector Machines (SVMs) are another machine learning technique that has been used in UWC forecasting Shabani, Yousefi, and Naser [2017]. Similar to decision tree and BN, SVMs are under the category of classification training algorithms and supervised training, which develops a classifier to predicate the class of the new sample Sriram [n.d.]. SVM implements a nonlinear mapping to transform the original training data into a higher dimension Han et al. [2011]. Then, considering the new dimension, it searches for the linear optimal separating decision

boundary in order to separate the tuples of one class from another Han et al. [2011]. Since SVMs are able to model complex nonlinear decision boundaries, they are highly accurate; however, their training time is slow Han et al. [2011].

2.6.5 Monte Carlo simulation

Monte Carlo simulation is a potential methodology in the probabilistic analysis due to its capability of considering uncertainty in modeling. Monte Carlo simulation uses random samples from known populations of simulated data to track a statistic's behaviour Mooney [1997]. Apparently, in each experiment, the population, including values of the random input variables, are simulated; then, using a computational method, the statistic concept is developed. Eventually, based on the sampling distribution, the output variables are calculated Mahadevan [1997].

2.6.6 Factor analysis

Factor analysis is known as the data reduction method, which is capable of identifying interconnection among various components of a system. This approach assumes that the observed factors (variables) are linear combinations of some underlying source factors and tries to exploit this correspondence to make conclusions about the factors (Kim et al., 1978). Accordingly, Factor analysis provides more interpretable insights into the data set so as to recognize the relationship among data components Panagopoulos [2014]. Panagopoulos [2014] adopted Factor analysis to identify interconnection between the demographic, socio-economic and hydrological variables influence UWC.

2.6.7 Coupled models applied in UWC

Altunkaynak and Nigussie [2017] developed a WT in combination with ANNs to predict monthly water consumption. Adamowski et al. [2012] proposed a method based on coupling discrete model wavelet-neural in order to forecast short-term UWC. They claimed that because data-based methods, including ANN, have limitations with nonstationary water consumption data, WT can effectively resolve this issue. Moreover, they suggested using bootstrap method to account for the uncertainty which was not considered in the proposed method. One year later, a paper published

to suggest a coupled model, wavelet-bootstrap-neural network, to study short-term UWC forecasting Tiwari and Adamowski [2013]. Then, Tiwari and Adamowski published another research, and they adopted wavelet-bootstrap-neural network again to implement mid-term UWC forecast Tiwari and Adamowski [2014]. A few years later, they applied their developed methodology in the other cities in Canada with the same purpose as before Tiwari and Adamowski [2017]. In all of those papers, they made efforts to test their methodology by comparing it with different approaches, including traditional WT, ANN, bootstrap and various forms of autoregression. All in all, even though they proposed a potentially applicable methodology for short-term and mid-term UWC forecasting, their approach was limited to climatic variables and socio-economic variables were not addressed. All of the papers verified the effectiveness of the proposed coupled model.

The other coupled model in the literature is the combination of times series and ANN in water demand prediction Al-Zahrani and Abo-Monasar [2015]. In the first stage, they used moving average and autoregressive techniques for time series forecasting. Afterwards, ANN prediction and combination of both mentioned methods were adopted as second and third stages, respectively. Results indicated that the coupled model is more capable of simulating daily water consumption pattern and seasonal trend of water consumption.

2.7 Shortcoming of the existing literature

A couple of noteworthy drawbacks exist in the previous literature, which is necessary to be considered in order to enhance the level of accuracy and application of UWC studies.

- The transition from BWC to SWC is determined based on the calendar date. The transition point can be quite sensitive to weather conditions in some regions; therefore, the alternative method is to define threshold values of air temperature and precipitation, at which the transition occurs, and water consumption starts to be dependent on climatic variables.
- The majority of studies that aimed to forecast UWC applied the techniques which are not capable of addressing uncertainties associated to the future changes in the variables. These approaches include regression approach, linear mixed effect models, factor analysis, wavelet transform, and, so forth.

- UWC was estimated by assuming a consistent difference between the average weather effects on water consumption over the historical and future periods. Hence, the future projection does not consider future uncertainties associated with climate change effects on future weather data.

Chapter 3

Methodologies

3.1 Data collection

The city of Brossard, Quebec, Canada, is chosen as a study site. The main land-use types in Brossard are residential and commercial with many parks scattered through the city Eslamian et al. [2016]. The city is a part of the metropolitan area of Montreal on the south shore of the Saint Lawrence River, which is the source of drinking water for the city Fig. 3.1.

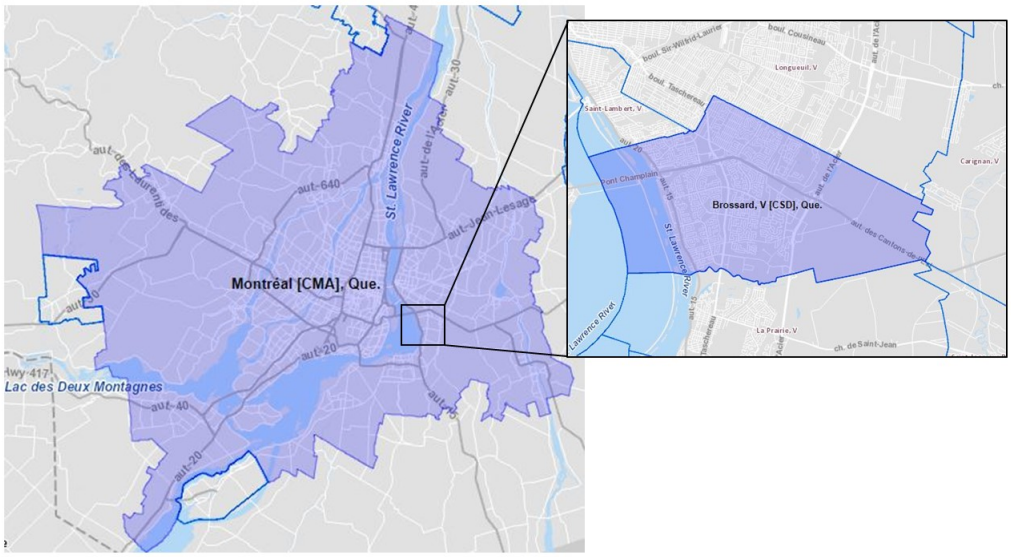

Figure 3.1: *Location map of city of Brossard in the census metropolitan area (CMA) of Montreal, Quebec Statistics-Canada [2019]*

According to the last two census, city of Brossard population in 2011 is $79,273$ and in 2016 is $85,721$, representing a percentage growth of 8.13% from 2011 Statistics-Canada [2019]. As per the city by-law Brossard [2019a], watering with sprinklers not equipped with a timer is permitted for even-numbered addresses on Wednesdays, Fridays and Sundays, while for odd numbered addresses permitted days are Tuesdays, Thursdays and Saturdays Brossard [2019b]. This example points to the need to manage both peak and daily-averaged water demands.

Time series of total DWC in cubic meter per day are obtained from the city of Brossard from January 2011 to October 2015. To eliminate the effect of population variation, for each year of the time period, values of total DWC are divided by the population for the year. The population from 2012 to 2015 are obtained by interpolation. Thus, values of DWC q have the units of cubic meter per capita per day. Time series of q is shown in Fig 3.2a.

For the same time period as UWC data, time series of climatic variables are also obtained: daily minimum temperature θ, daily maximum temperature Θ, daily mean temperature θ_0, and total daily precipitation p. Measurements of these climatic variables were made from an Environment Canada station in the Pierre Elliott Trudeau International Airport (YUL), located at ($45°28'11.06$" N, $73°44'41.71$" W). Time series of daily mean temperature, and total daily precipitation are plotted in Fig 3.2b and Fig 3.2c, respectively.

3.2 Overall approach

The proposed methodology of this research is covered in this chapter. Fig 3.3 shows the general steps of the work.

Total DWC and the climatic variables are obtained for the same period of time in data collection. The first step is data pre-processing where values are normalised, and missing values are replaced to prepare the suitable input for clustering. This step also includes outlier detection and replacement. Then, UWC is decomposed to SWC and BWC. Finally, SWC and downscaled data are used as the input for the predictor model. The model output and the estimated BWC are merged to forecast long-term UWC. Further details of each step are elaborated in the following sections.

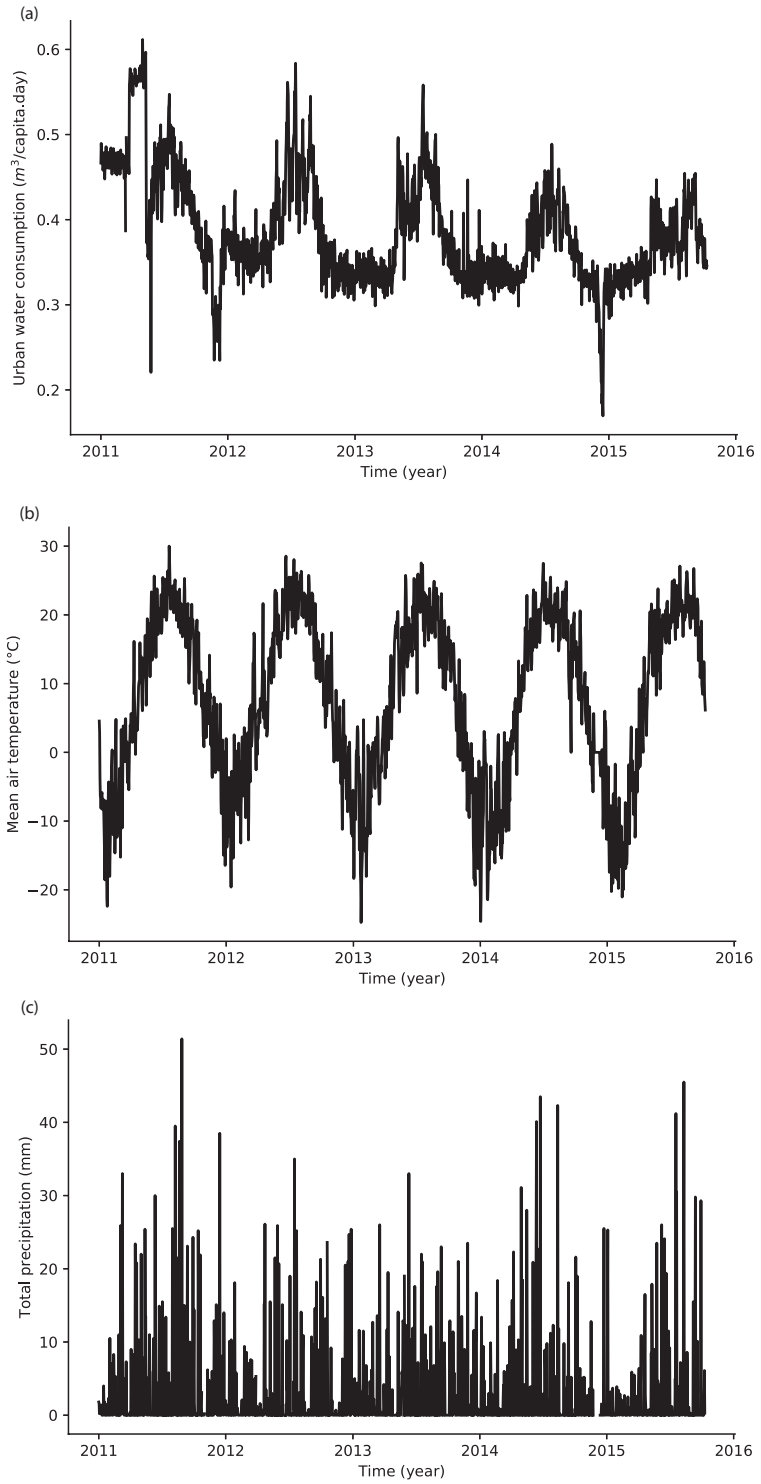

Figure 3.2: *Time series of (a) metered DWC (cubic meter per capita per day) in the city of Brossard in the metropolitan area of Montreal, Quebec; (b) observed daily mean temperature (°C); and (c) observed total precipitation (mm). The temperature and precipitation observations were made in the Montreal area. The starting date is January first,* 2011 *and the ending date is October* 10*th,* 2015*.*

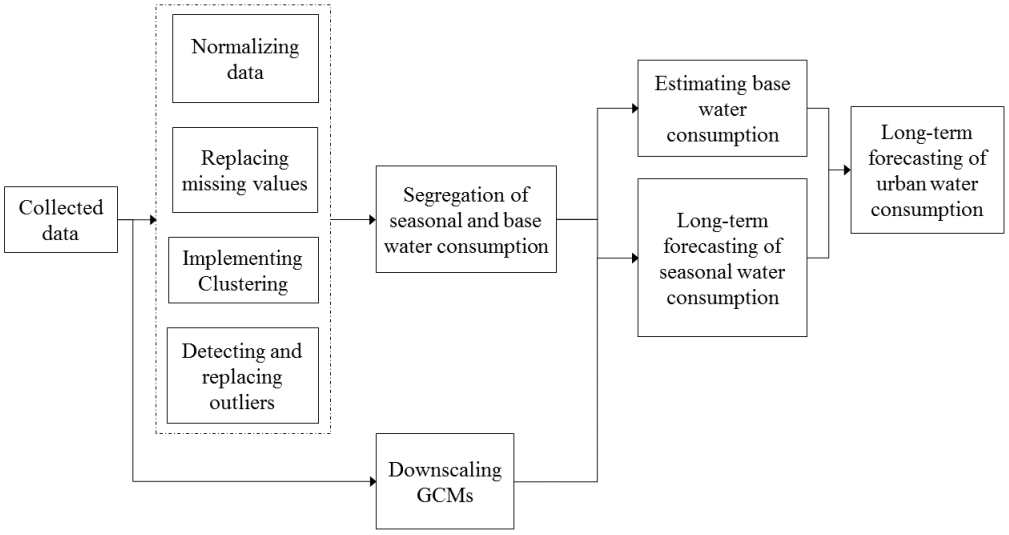

Figure 3.3: *Framework to forecast long-term UWC*

3.3 Normalisation of raw data

Normalising the raw data is one of the steps of data pre-processing. There are many methodologies such as various clustering algorithms that are based on the distance of data points; therefore, they require normalised data as the input.

Let x_i denote the i^{th} value of a raw series of the variable in question, μ denotes the mean of the data series, and σ denotes the standard deviation. The values are converted into a z-score, Z_i, using the z-transform method, given by

$$Z_i = \frac{x_i - \mu}{\sigma}, \quad for \quad i = 1, 2, 3, \ldots, n, , \tag{1}$$

Eq. (1) is applied to time series of q = (q_1, q_2, q_3, ... , q_n); θ = (θ_1,θ_2 ,θ_3 , ... ,θ_n); and Θ = (Θ_1,Θ_2 ,Θ_3 , ... ,Θ_n). Outputs of z-score or normalised time series are used as input to clustering analysis. The purpose of clustering analysis is:

- to detect data point outliers Ashouri et al. [2018]; Z. Yu, Fung, and Haghighat [2013].
- to decompose UWC into BWC and SWC.

The suitability of the minimum-maximum method for data normalisation was tested. It was

found that using the results from this method as input, clustering analysis did not yield proper differentiation between BWC and SWC (results not shown).

3.4 Cluster analysis

Cluster analysis divides data points of q, Θ and θ into classes or clusters such that objects in the same cluster have high similarity, while objects in different clusters have low similarity. This form of classification gives objects with cluster labels derived from data, known as unsupervised classification. A distance matrix is used as clustering criteria and hierarchical clustering is implemented by agglomerative nesting. This algorithm begins with assigning each data point to a cluster by its own. Therefore, initially, the total number of clusters is equal to the number of data points. Then, it merges the data points of the least dissimilarity until all the data points belong to the same cluster. In this research, the agglomerative nesting is based on Euclidian distance Taleb and Kaddour [2017], which is a measure of similarity levels.

The presence of missing values in a data set is very typical in real life data. According to literature, there are some methodologies including Bayesian statistics that are robust to the missing data. However, some algorithms including clustering, fail in handling missing data. Before performing the clustering analysis, the missing values of q, Θ and θ are replaced by interpolation.

The performance of the clustering algorithm is evaluated by using the Silhouette coefficient Ashouri et al. [2018] given by

$$S = \frac{b - a}{max(a, b)} \tag{2}$$

where a is the mean Euclidian distance between a sample and all other data points in the same class; and b is the mean Euclidian distance between a sample and all other points in the next nearest cluster. The coefficient S is the mean of all samples in the dataset. The coefficient reveals the capability of the algorithm in grouping the objects and satisfying intra class similarity and interclass dissimilarity. The values of S range from -1 to 1, with S = 1 corresponding to a high quality of clustering, and S = -1 to false clustering.

In the clustering analysis of q, the effect of socio-economic variables is ignored and the focus

is on climatic variables. This helps isolate the influence of temperature changes during a year on UWC patterns. The attributes used in the clustering are Θ, θ and θ_0. The reason for selecting Θ and θ is the wide range of daily air temperature for the city of Brossard. Through the clustering analysis, hidden groups in data points is found and the influence of seasonality (air temperature) on q are verified. At the same time, data point outliers or data points that behave in an unexpected manner are detected.

3.5 Detection and replacement of data point outliers

Outliers are values that behave differently from expectation Han et al. [2011]. Clustering-based methods and statistical methods are available for outlier detection. The statistical methods assume a normal distribution of data points. Therefore, values in a low probability region are considered as outliers. The clustering-based methods accept that outliers might belong to a small or sparse cluster or might be far from the clusters to which they are closest. Both the clustering-based and statistical methods are used to detect unusual patterns and outliers. The concept of maximum likelihood Rousseeuw and Hubert [2011] in the statistical methods is applied to those clusters that contain values with abnormal behaviours. Values of q outside the range of $\mu \pm 2\sigma$ are labeled as an outlier. Note that $\mu \pm 2\sigma$ contains 95% of data under the assumption of normal distribution.

The presence of data outliers in a cluster potentially misleads the training process and reduce the accuracy of models to be developed for forecasting UWC. In this research, upon detecting an outlier, it is replaced by the average between the two values for the dates that precede and follow the date of the outlier. After the replacement of outliers, so-called clean and consistent data sets are obtained for the development of forecast model for UWC.

3.6 Forecast model

Assume that the transition from BWC to SWC is sensitive to weather. The temperature threshold is determined, θ_t, for the transition. When $\theta_0 \leq \theta_t$, UWC is no longer sensitive to the actual values of θ_0 and p Eslamian et al. [2016]; Gato et al. [2007]. The DWC, q, is considered as a numerical target, with two components: BWC and SWC. In this research, a forecast model is developed by

combining Bayesian statistics Heckerman [1997] with a predictor model. One advantage is an enhanced accuracy of the long-term forecast and reduced uncertainties. The developed method uses a Bayesian linear regression model Bańbura, Giannone, and Reichlin [2010]; Raftery, Madigan, and Hoeting [1997]; Yuan et al. [2017] and downscaled values of air temperature and precipitation.

3.6.1 Temperature threshold

To determine the temperature threshold, θ_t, for the transition from BWC to SWC, scatter plots of q versus θ_0 are made for each year of the records and for the entire time period of records. Separate straight lines are fitted for SWC and BWC, respectively. The slopes of these lines are estimated using a non-parametric method known as Sen's estimator Sen [1968]. The intersection of the two lines is considered as the transition point. The lines are merged to obtain the best fit of q versus θ_0. The temperature associated with the elbow of the best fit is θ_t for the year in question or for the entire time period of records.

3.6.2 BWC estimation

Some early studies Eslamian et al. [2016]; Gato et al. [2007] reported a negative or positive trend of BWC over time. For convenience, this research considers BWC as stationary. In fact, nonstationary features of water use records are captured in SWC and are lumped into variations affected by climate change.

3.6.3 Bayesian statistics

Bayesian statistics represent the probability by quantifying uncertainties. Uncertainties in model parameters are expressed by a probability distribution over possible parameter values. Bayes theorem Heckerman, Geiger, and Chickering [1995] is the key rule in the Bayesian statistics. If H is a hypothesis and D is observed data, the theorem is as follows

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D)} \tag{3}$$

where $P(H|D)$ represents the posterior probability of hypothesis given the condition that D occurs. $P(H)$ is the prior probability of hypothesis, $P(D)$ is the marginal (total) probability of observed data and is effectively a normalising constant, and $P(D|H)$ is the probability (likelihood) of data given hypothesis.

3.6.4 Bayesian linear regression

Bayesian statistics is a powerful technique for probabilistic modeling that have been adopted in a wide range of statistical modeling, including linear regression models to make prediction about a system Han et al. [2011]; Mudgal, Hallmark, Carriquiry, and Gkritza [2014]; Raftery et al. [1997]; Yuan et al. [2017]. A linear regression model is expressed as

$$q = \beta_0 + \beta_1\theta + \beta_2\Theta + \beta_3 p + \epsilon, \tag{4}$$

where β_0 is the intercept; β_1, β_2 and β_3 are the weights (known as the model parameters); θ, Θ and p are called the predictor variables; and ϵ is an error term representing random noise or the effect of variables not included in the model equation. Eq. (4) can be rewritten as $q = \boldsymbol{X}\boldsymbol{\beta}+\epsilon$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$'; and $\boldsymbol{X} = (1, \theta, \Theta, p)$ Han et al. [2011]; Taleb and Kaddour [2017]. The linear regression is formulated using probability distributions rather than point estimates to predict q. Therefore, q is not estimated as a single value but is assumed to be drawn from a probability distribution.

The model for Bayesian linear regression with the response sampled from a normal (Gaussian) distribution N is $q \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2)$. The output, q, is generated from a normal distribution characterised by a mean and variance. The mean for the normal distribution is the regression coefficient matrix ($\boldsymbol{\beta}$) multiplied by the predictor matrix ($\boldsymbol{X}$). The variance is the square of the standard deviation, σ.

Bayesian linear regression model provides the representation of the uncertainties in predictor variables and determine the posterior distribution for the model parameters. Not only is the response generated from a probability distribution, but the model parameters are assumed to come from a distribution as well. The posterior probability of the model parameters comes from Bayes theorem

[Eq. (5)]:

$$(\boldsymbol{\beta}|q, \boldsymbol{X}) = \frac{P(q|\boldsymbol{\beta}, \boldsymbol{X}) \times P(\boldsymbol{\beta}|\boldsymbol{X})}{\int P(q, \boldsymbol{X}|\beta_i) d\beta_i}, \tag{5}$$

where $P(\boldsymbol{\beta}|q,\boldsymbol{X})$ is the posterior probability distribution of the model parameter given the predictors and the dependent variable; $P(q|\boldsymbol{\beta},\boldsymbol{X})$ is the likelihood of data; $P(\boldsymbol{\beta}|\boldsymbol{X})$ is the prior probability of parameters and the denominator of the equation is the marginal probability that can be found as per the law of total probability.

3.6.5 Marcove Chain Monte Carlo

In practice, it is difficult to compute the marginal likelihood for continuous values; it is intractable to calculate the exact posterior distribution. As a solution, a sampling method Markov Chain Monte Carlo is involved in approximating the posterior without computing the marginal likelihood Godsill [2001]; Q. Li, Gu, Augenbroe, Wu, and Brown [2015]; Yuan et al. [2017]. Monte Carlo is a general technique of drawing random samples, and Markov Chain means the next sample drawn is based only on the previous sample value. By bringing more samples, the approximation of the posterior will eventually converge on the actual posterior distribution for β_1, β_2 and β_3.

As the starting point in applying Markov Chain Monte Carlo, parameter space is defined covering all the possible values for β_1, β_2 and β_3. Then, following the study of Davidson-Pilon [2015] and other similar studies Bańbura et al. [2010]; Bishop and Tipping [2003]; Ghosh, Basu, and O'Mahony [2007]; Lambert [2018],the prior probability is defined for each of the parameters as a normal distribution. Next, the likelihood for each possible parameter is computed. Last, prior $\times$ likelihood for any given point in parameter space is computed.

If we have domain knowledge, or a guess for what the model parameters should be, informative prior distributions should be incorporated into the analysis. In the absence of any estimates ahead of time, we can use non-informative priors for the parameters that express uncertainty about the parameters and also represents the priori constraints Bishop and Tipping [2003]; Raftery et al. [1997]. There are a large range of distributions available to use as prior distribution. The popular choices of prior distribution for regression coefficients are Normal, Student-t and cauchy Lambert [2018].

According to Lambert [2018], these three distributions are suitable for continuous unconstrained parameters. Also, the distributions are characterized by a mean, the location parameters, and a variance, the scale parameter. For the purpose of simplicity and following the similar studies Bańbura et al. [2010]; Bishop and Tipping [2003]; Ghosh et al. [2007], the prior probability for each of the parameters is defined as a normal distribution.

Markov Chain Monte Carlo is implemented using the No-U-Turn algorithm Hoffman and Gelman [2014]. This algorithm is efficient when the variables involved are continuous. There is no need to set the number of steps. This is an advantage over Hamiltonian Monte Carlo, which takes a series of steps informed by first-order gradient information, and is sensitive to the number of steps.

A set of parameter values for accepted moves are generated (if the proposed move is not accepted, the previous value is repeated) and after many repetitions, the empirical approximation of the distribution is found. Eventually, the result of performing Bayesian Linear Regression is the probability density function of possible model parameters based on the data and the prior.

3.6.6 Model performance measure

The entire data set is splitted into training and testing sets. The training set contains 75% of data and used to build the model; while, the testing set contains 25% and is utilized for verifying the accuracy of prediction. Also, two goodness-of-fit measures are used to evaluate the performance of models developed in this study. These measures are Mean Absolute Error and Root Mean Squared Error. Mean Absolute Error is the average of the absolute value of the difference between predictions and the actual values [Eq. (7)], and Root Mean Squared Error is the square root of the average of the squared differences between the predictions and the actual values [Eq. (6)].

$$RMSE = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n}}, \tag{6}$$

where y_i is the observed value for the i^{th} observation, $\hat{y}_i$ is the predicted value and n is the number of observations.

$$MAE = \frac{\sum |\hat{y}_i - y_i|}{n}, \tag{7}$$

3.7 Future climate projection

This research uses the NASA earth exchange global daily downscaled projections. The archived results comprise daily θ, Θ and p from the time period of 1950 to 2005 and of 2006 to 2050. Table 3.1 shows all the 21 models. All these models are downscaled into a unique $0.25^\circ \times 0.25^\circ$ grid resolution.

The closest grid to the climate station (station ID: 30165) is found and the past and future GCMs data are downloaded from https://cds.nccs.nasa.gov/nex-gddp/. To do so, we click on NCCS THREDDS from data access section. This link directs us to the catalog including minimum temperature, maximum temperature and precipitation. This data is available for all 21 GCMs and in form of historical data, future data under RCP 4.5 and RCP 8.5.

3.7.1 Implementation of bias correction

Bias correction is applied to obtain fine-scale weather data. To make bias corrections, the downscaled projections is compared to observations. Assume that the GCM bias is stationary. The correction algorithm for current climate conditions are also valid for future conditions. The performance of corrections requires historical/current climate data at large scale and fine scale. The historical climate condition at large scale is extracted from the GCMs at the stations close to Montreal and the fine scale data is the observed historical temperature and precipitation collected from YUL. Then, the same correction algorithm is applied on the future climate data from the GCMs to obtain the fine scale future climate variables data.

The correction algorithm is quantile mapping. The quantile relationship between simulated GCM outputs and climate observations remains unchanged. Following Bennett et al. [2014], p is corrected by multiplying correction factors ('multiplicative quantile mapping'), θ and Θ by adding correction factors ('additive quantile mapping').

In multiplicative quantile mapping, the change factor is calculated at each percentile as

$$R(x) = F_O^{-1}(x)/F_G^{-1}(x), \tag{8}$$

where R is the change factor function. The right hand side is the ratio of inverse function F_O^{-1} to inverse function F_G^{-1}. The argument of R is a function itself, given by x = F_{Gf} (H_{Gf}). F_O^{-1} is the inverse function of the empirical cumulative distribution function of the historical daily climate observation; F_G^{-1} is the inverse cumulative distribution function of the GCM future data; the argument H_{Gf} is the daily future GCM projection and F_{Gf} is the cumulative distribution function of the GCM future data.

For multiplicative quantile mapping, the GCM future data is multiplied by R to give the downscaled future data H_{Gd} as H_{Gd} = H_{Gf} $\times$ R. For additive quantile mapping, the change factor is calculated at each percentile as

$$R(x) = F_O^{-1}(x) - F_G^{-1}(x), \tag{9}$$

R is calculated for all given percentile levels, made corrections to the future daily GCM projection as H_{Gd} = H_{Gf} + R. The corrections are applied to each of the calendar months.

3.8 Ensemble GCM model

Following making bias correction, the corrected future fine-scale weather variables are used in the forecast of long-term UWC up to the year of 2050. There is uncertainty in climate predictions resulting from structural differences in the GCMs as well as uncertainty due to variations in initial conditions or model parameterisations Semenov and Stratonovitch [2010].

To avoid the challenge of the uncertainty in climate projections, Intergovernmental Panel on Climate Chang guidelines is followed for handling uncertainties in climate projections, and treated the four GCMs using the ensemble approach Stocker et al. [2014]. Therefore, all the 21 models are downscaled for RCP 4.5 and RCP 8.5. Specifically, all available 21 GCMs are obtained from NASA earth exchange global daily downscaled projections for RCP 4.5 and RCP 8.5 in terms of minimum

temperature, maximum temperature and total precipitation. The temporal variations of the three climate variables are plotted and identified the lower and the upper bound of the variations. The four GCMs are selected to give the lower bound, upper bound and representative temporal variations between the lower and upper bounds, and used the average of their outputs. Then, Ensemble GCM model is used as the input for the Bayesian linear regression model.

Table 3.1: *Description of the 21 IPCC-CMIP5 climate models included in the NEX-GDDP downscaled climate scenarios Jaramillo and Nazemi [2018]*

Models name	Institution	Latitude resolution (°)	Longitude resolution (°)
ACCESS1-0	CSIRO (Commonwealth Scientific and Industrial Research Organization, Australia), and BOM (Bureau of Meteorology, Australia)	1.25	1.875
CSIRO-MK3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with the Queensland Climate Change Centre of Excellence	1.8653	1.875
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	2.7906	2.8125
BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration	2.7906	2.8125
GFDL-CM3	NOAA's Geophysical Fluid Dynamics Laboratory	2	2.5
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	2.7906	2.8125
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	2.7906	2.8125
GFDL-ESM2G	NOAA's Geophysical Fluid Dynamics Laboratory	2.0225	2
MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	1.4008	1.40625
CanESM2	Canadian Centre for Climate Modelling and Analysis	2.7906	2.8125
GFDL-ESM2M	NOAA's Geophysical Fluid Dynamics Laboratory	2.0225	2.5
MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	1.8653	1.875
CCSM4	National Center for Atmospheric Research	0.9424	1.25
INMCM4	Institute for Numerical Mathematics, Moscow, Russia	1.5	2
MPI-ESM-MR	Max Planck Institute for Meteorology (MPI-M)	1.8653	1.875
CESM1-BGC	National Science Foundation, Department of Energy, National Center for Atmospheric Research	0.9424	1.25
IPSL-CM5A-LR	Institut Pierre-Simon Laplace	1.8947	3.75
MRI-CGCM3	Meteorological Research Institute	1.12148	1.125
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancées en Calcul Scientifique	1.4008	1.40625
IPSL-CM5A-MR	Institut Pierre-Simon Laplace	1.2676	2.5
NorESM1-M	Norwegian Climate Centre	1.8947	2.5

Chapter 4

Results and discussion

We performed data analysis tasks using Python code, including some libraries (Numpy, Scipy, Seaborn, Pandas, Matplotlib, Scikit-learn, and Pymc3). The results from the analysis are discussed in this chapter.

4.1 Clustering UWC based on θ and Θ

The z-scores [Eq. (1)] of daily water consumption q, minimum air temperature θ and maximum air temperature Θ each contain 1744 records (or objects). The transformed q, θ and Θ values are in the range of -3.4 to 3.5, -2.6 to 1.8, and -2.7 to 1.8, respectively. For example, on January 1, 2012, the transformed values are: q = -0.22, θ = -0.42 and Θ = -0.52, compared to q = 1.32, θ = 1.17 and Θ = 1.10 for July 1, 2012. The January and July objects are significantly different and thus are expected to belong to different clusters. In order to determine how many clusters into which the objects should be grouped, a hierarchical clustering analysis of the three variables was performed and grouped the q records by weather (air temperature) conditions in order to identify the temporal variations in q. The results are shown in Fig. 4.1 as a dendrogram (a tree diagram).

The results may be interpreted as follows:

- Choice A: The 1744 objects can possibly be grouped into two clusters (namely 0 and 1). Cluster 0 contains 811 objects, whereas cluster 1 contains the remaining 933 objects. The sizes of the two clusters are somewhat different. The Euclidean distance between the two

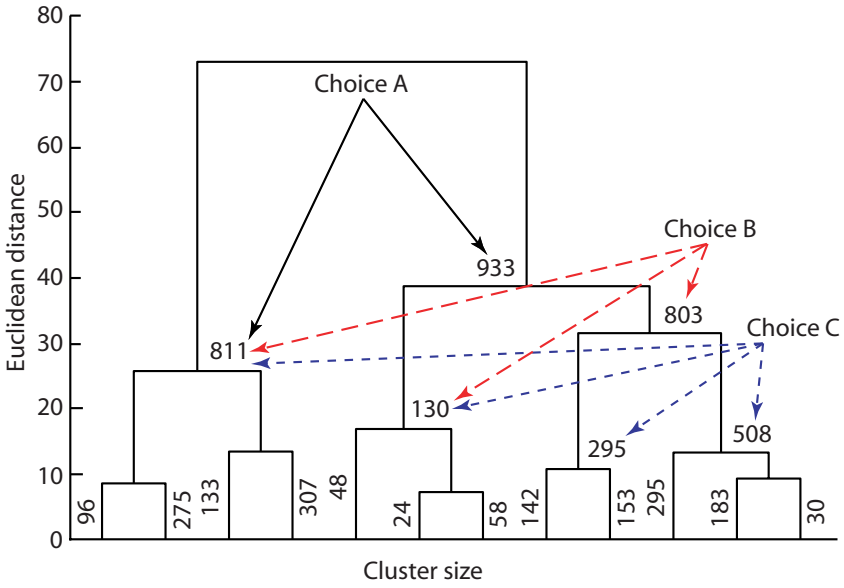

Figure 4.1: *Dendrogram showing the Euclidean distances for possible data clusters. The number in of objects within the cluster in question is indicated.*

clusters is 74. The distance is a measure of dissimilarity between the two clusters.

- Choice B: If the 1744 objects are to be grouped into three clusters (namely 0, 1, and 2), one way to do so is to keep the 811 objects in cluster 0, but to split the 933 objects into two clusters. The split creates new cluster 1 with 130 objects and cluster 2 with 803 objects. Between clusters 1 and 2, the Euclidean distance will be 39.

- Choice C: If the objects are to be grouped into four clusters (namely 0, 1, 2, and 3), one may further slip the 803 objects into two clusters. The further split gives new cluster 2 with 295 objects and cluster 3 with 508 objects. The Euclidean distance between these two clusters is smaller than 39.

There is a need to validate the consistency within clusters of q observations, by examining the associated values of Silhouette coefficient S [Eq. (2)]. Silhouette plots for choices A, B and C are shown in Fig. 4.2a, Fig. 4.3a, and Fig. 4.4a, respectively. In all the three choices, for each of objects, the coefficient is calculated using Eq. (2), and the coefficients for the objects are sorted based on the actual S values from the calculations. For each of the objects, the figures show a straight horizontal line, displaying an S value in the horizontal axis. For example, for cluster 0 in choice A, there

is a total of 811 horizontal lines stacked vertically (Fig. 4.2a), adjacent lines being close and thus appearing to merge. Similarly, for the objects within clusters 1, 2 and 3, the sorted values of S are shown Fig. 4.2a.

The S values for all the objects (or samples) in choice B give an average S of 0.469, higher than an average S of 0.467 in choice A, and an S average of 0.444 in choice C. These averages are marked in Fig. 4.2a, Fig. 4.3a, and Fig. 4.4a by the vertical dashed lines. The data points of q are showed within clusters 0 and 1 of choice A in Fig. 4.2a as a time series and in Fig. 4.2c as a q versus θ_0 scatter. The data points are shown within clusters 0, 1 and 2 for choice B (Fig. 4.3b and Fig. 4.3c) and those within clusters 0, 1, 2, and 3 for choice C (Fig. 4.4b and Fig. 4.4c) in the same way.

Choice A effectively separates SWC, the green data points (Fig. 4.2b) within cluster 1 (Fig. 4.2a) from BWC, the black data points (Fig. 4.2b) within cluster 0 (Fig. 4.2a). In Fig. 4.2b, the SWC data points are from the warm months (May to October), whereas the BWC data points are from the cold months (November to April). In Fig. 4.2c, there exists a good correlation between q and θ_0 during the warm season (the green data points), and the vast majority of the black data points show a narrow range of BWC. However, cluster 0 contains data points of abnormally high or low BWC. Thus, choice A is not suitable.

Choice B also effectively separates SWC, the light blue data points (Fig. 4.3b) within cluster 1 (Fig. 4.3a) from BWC, the black data points (Fig. 4.3b) within cluster 0 (Fig. 4.3a). At the same time, choice B groups a small number of abnormal data points (Fig. 4.3b, the red data points from January to May 2011) into cluster 2 (Fig. 4.3a). These data points are outliers. Overall, choice B produces desirable clustering results:

- SWC (cluster 1) separates from BWC (cluster 0);
- SWC and temperature are well correlated (Fig. 4.3c, the light blue data points); and
- Data outliers (cluster 2) are identified and removed from the data set.

Choice C produces the above-mentioned desirable results (Fig. 4.4) as choice B. However, choice C splits BWC into two clusters (Fig. 4.4a, clusters 1 and 3; Fig. 4.4b, the yellow data points

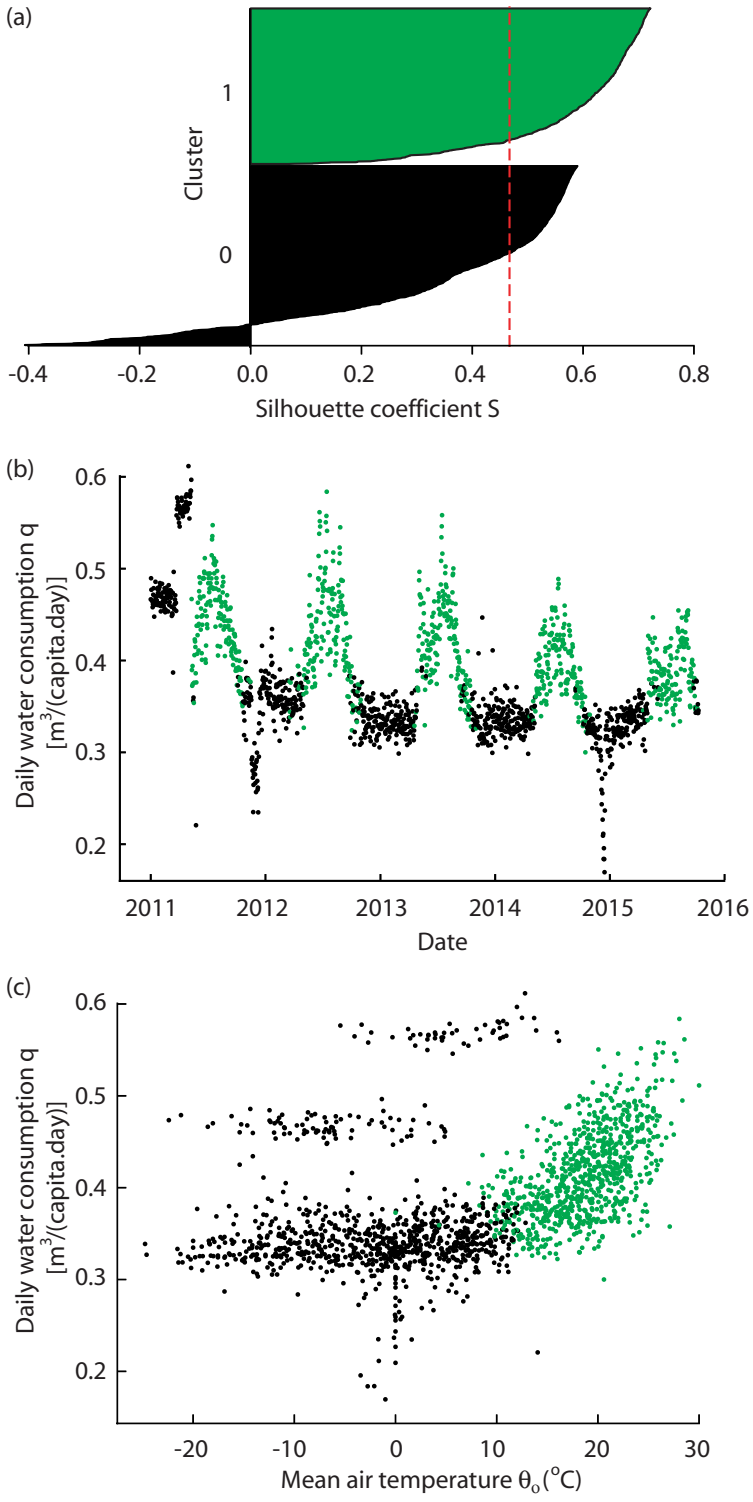

Figure 4.2: *Results of hierarchical clustering of daily water consumption observation: (a) Silhouette plot showing two clusters; (b) observations of daily water consumption; (c) scatter plot of daily water consumption vs. mean air temperature.*

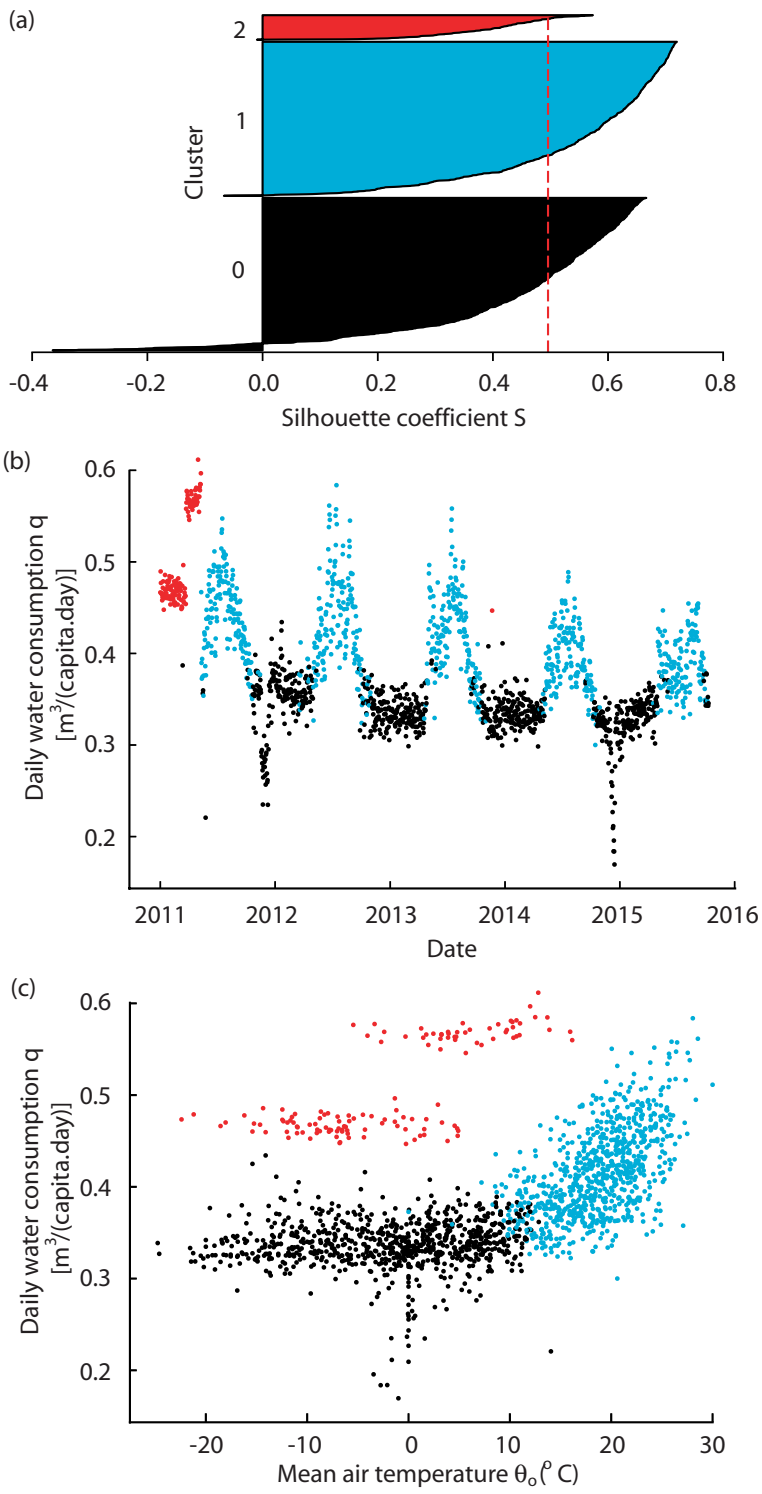

Figure 4.3: *Results of hierarchical clustering of daily water consumption observation: (a) Silhouette plot showing two clusters; (b) observations of daily water consumption; (c) scatter plot of daily water consumption vs. mean air temperature.*

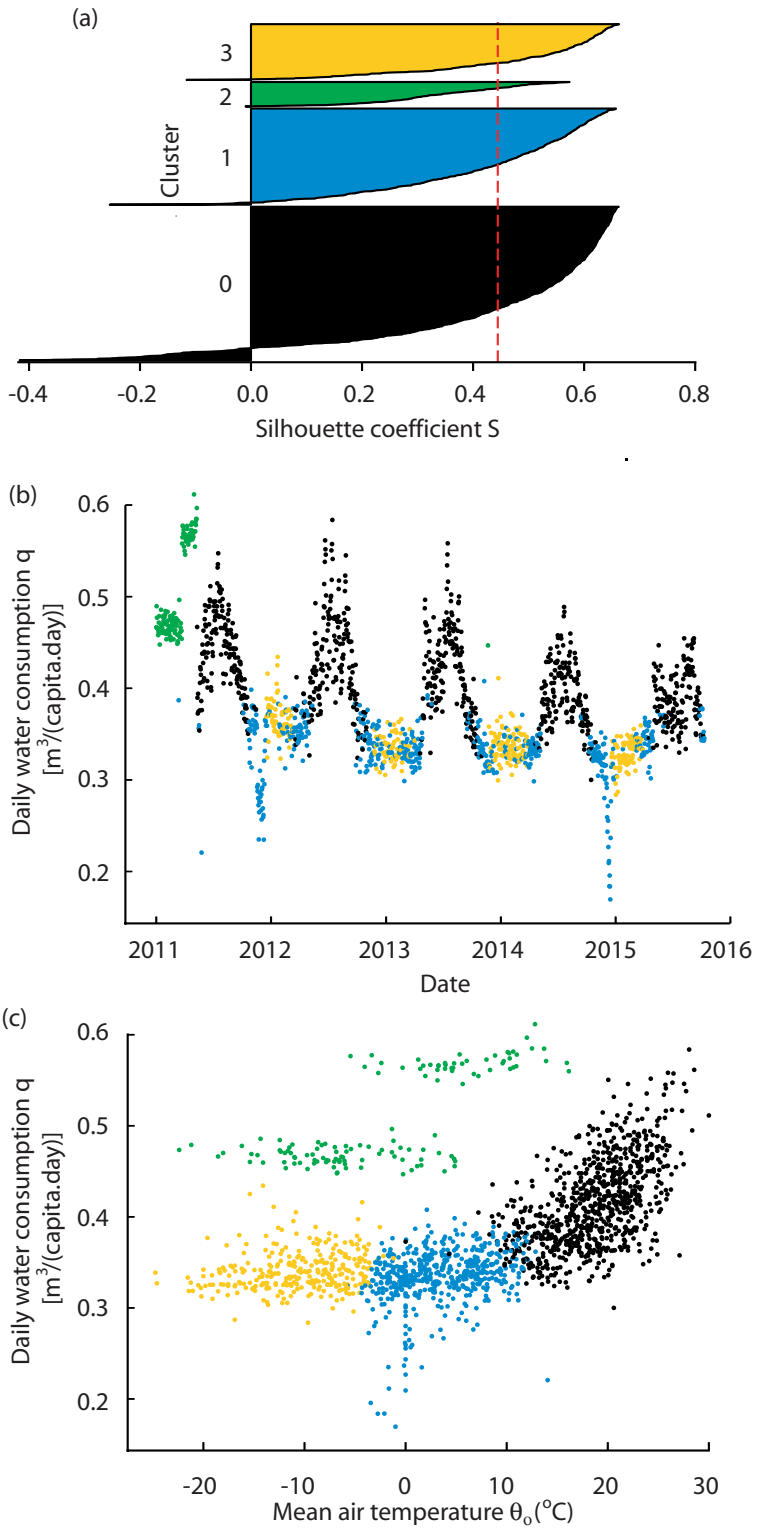

Figure 4.4: *Results of hierarchical clustering of daily water consumption observation: (a) Silhouette plot showing two clusters; (b) observations of daily water consumption; (c) scatter plot of daily water consumption vs. mean air temperature.*

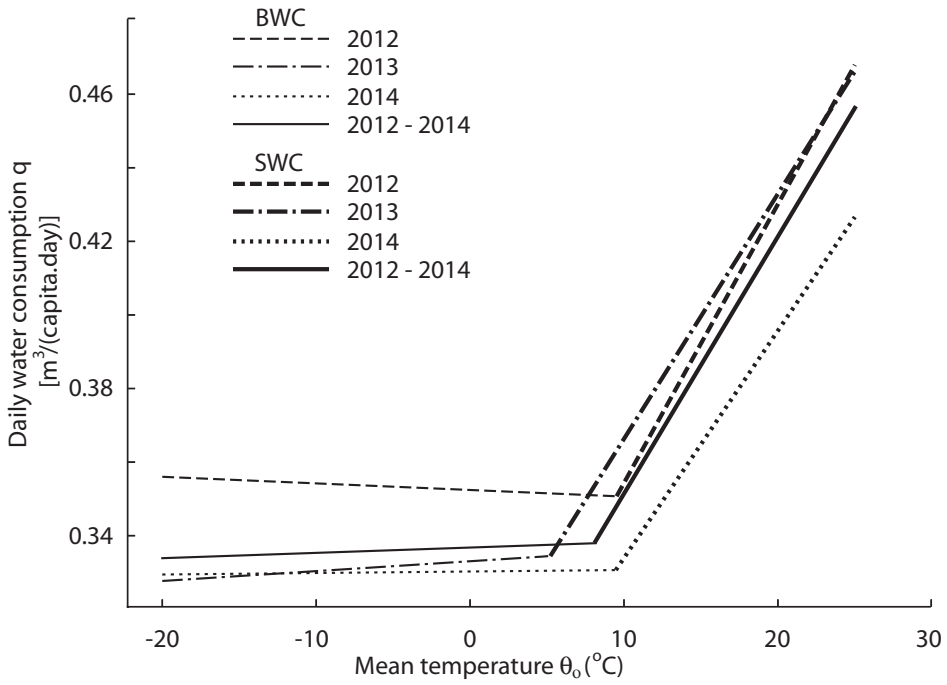

Figure 4.5: *Lines of best-fit to observations of daily water consumption versus mean air temperature.*

and the light blue data points). Cluster 1 represents BWC for the season of $\theta_0 \leq$ -3°C and cluster 3 contains BWC for the season of -3°C $\leq \theta_0 \leq$ 10°C. Thus, choice C is not suitable because of the split of BWC into two clusters. For further analysis, it is much convenient to have BWC in one cluster only, as is the case in choice B. Thus, the preference is choice B.

4.2 Threshold temperature

To determine the threshold temperature, θ_t, q is regressed against θ_0 for BWC and SWC. In Fig. 4.5, lines of best fit to scatter data of q vs. θ_0 are plotted for the individual years of 2012, 2013, and 2014 and for the entire record period of 2012 to 2014. The corresponding lines for BWC and SWC intersect at certain values of θ_0. These values are the threshold temperature . The results show that θ_t is equal to 9.5°C for 2012 and 2014, and 5.2°C for 2013. For 2013, the transition from BWC to SWC took place at lower temperature, and more data points fall into the SWC cluster. Given that other random weather variables can affect the transition, for practical purposes of accommodating uncertainties, θ_t is taken as 9°C.

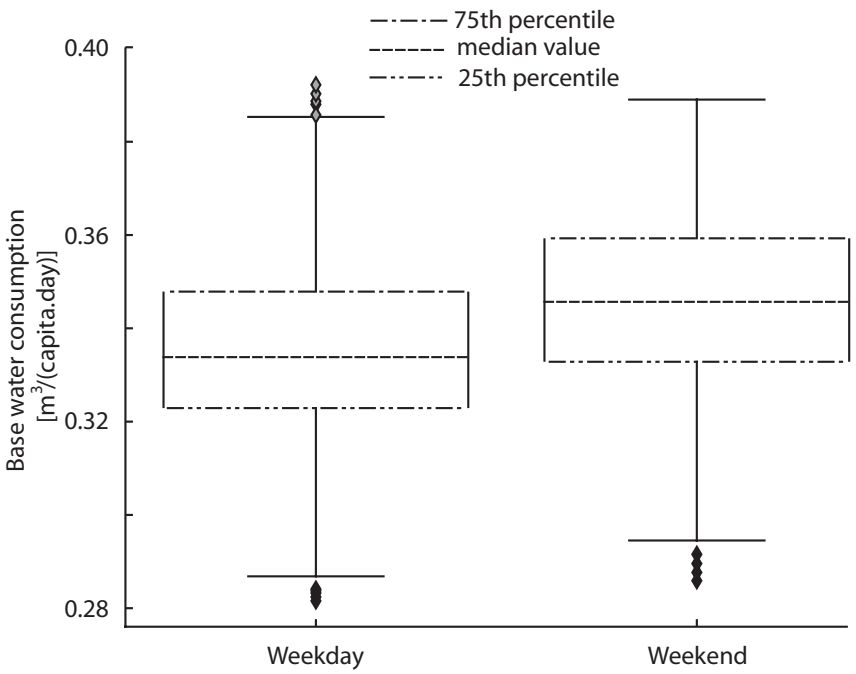

Figure 4.6: *Base water use for weekends and weekdays for the period of 2011 to 2015.*

4.3 BWC calculations

Some early studies Eslamian et al. [2016]; Gato et al. [2007] reported a negative or positive trend of BWC over time. For convenience, this research considers BWC as stationary. In fact, nonstationary features of water use records are captured in SWC and are lumped into variations affected by climate change. In Fig. 4.6, it is shown that BWC ranges from 0.33 to 0.36 m^3 per capita per day for weekend, and from 0.32 to 0.35 m^3 per capita per day for weekdays. Although the BWC levels differ between weekends and weekdays, but the lower and upper values of the ranges give the same interval. This interval is 0.03 m^3 per capita per day. It is assumed that this interval is valid for the future three decades.

4.4 SWC calculations

Model Eq. (4) gives predictions of q as the dependent variable, with θ, Θ, and p as predictors, and with β_1, β_2 and β_3 as model parameters. In Bayesian statistics, the determination of β_1, β_2 and β_3 requires posterior of the parameters, resulting in a probability distribution for each parameter.

The probability density functions of the posterior for intercept β_0, parameters β_1, β_2 and β_3 (for θ, Θ, and p, respectively) and standard deviation σ [Eq. (5)] are derived using Monte Carlo Markov Chain.

The results are illustrated in Fig. 4.7 as posterior histograms for β_0, β_1, β_2 and β_3, and σ. The histograms show 95% Highest Posterior Density (HPD), which is a credible interval for the parameters. A credible interval in Bayesian statistics is an equivalent of a confidence interval Curran [2005]. The probability of $0.034 \leq \sigma \leq 0.037$ is 95%.

Before using the predictor model [Eq. (4)] to forecast UWC, its accuracy is assessed in terms of two goodness of fit measures: Root Mean Squared Error and Mean Absolute Error. The assessments indicate that the two errors are 0.0450 and 0.0378, respectively. Therefore, it is concluded that the predictor model has a good accuracy.

4.5 Ensemble GCM model

All the 21 GCMs are downscaled for RCP 4.5 and RCP 8.5. The comparison between observation data, future coarse-scale GCM and future downscaled data in the station of interest in the form of cumulative distribution function (CDF) are demonstrated under two emission scenarios. Fig. 4.8 and Fig. 4.9 show this comparison for the three variables for RCP 4.5 and RCP 8.5.

Based on Fig. 4.8 and Fig. 4.9, the future bias-corrected or downscaled CDF curves is in the middle of observation CDF curve and future coarse-scale GCM which is precisely in accordance with the technical concept of empirical quantile bias-correction. Apparently, in precipitation plots, it is shown that observation and future coarse-scale GCMs are close; hence, downscaled CDF does not give a clear demonstration.

To make the comparison between the models' outputs, annual minimum temperature, annual maximum temperature and annual precipitation are plotted for RCP 4.5 and 8.5 by 2050. For the purpose of visualization, 21 models' outputs of the three variables are broken into three graphs and are shown in Fig. 4.10 to Fig. 4.15.

Comparing the downscaled results by 2050 in Fig. 4.10 to Fig. 4.15, four GCMs: CanEMS2, MIROC_ESM, CNRM_CM5, GFDL_ESM2M for RCP 4.5 and RCP 8.5, are selected to cover upper

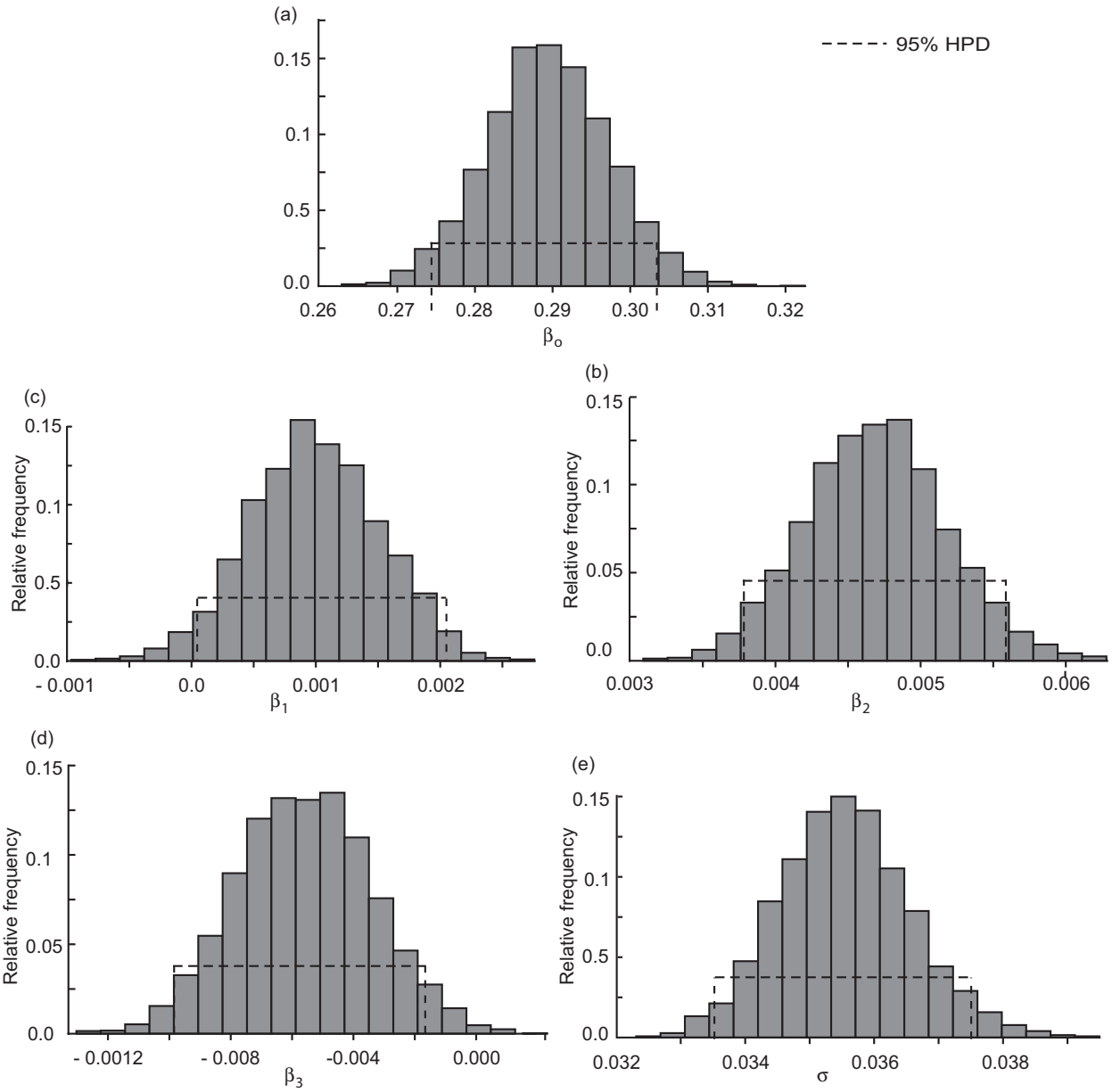

Figure 4.7: *Histograms of posteriors for: (a) intercept β_0; (b) regression coefficient $\beta_1$for θ; (c) regression coefficient β_2 for Θ; (d) regression coefficient $\beta_3$for p; (d) standard deviation σ*

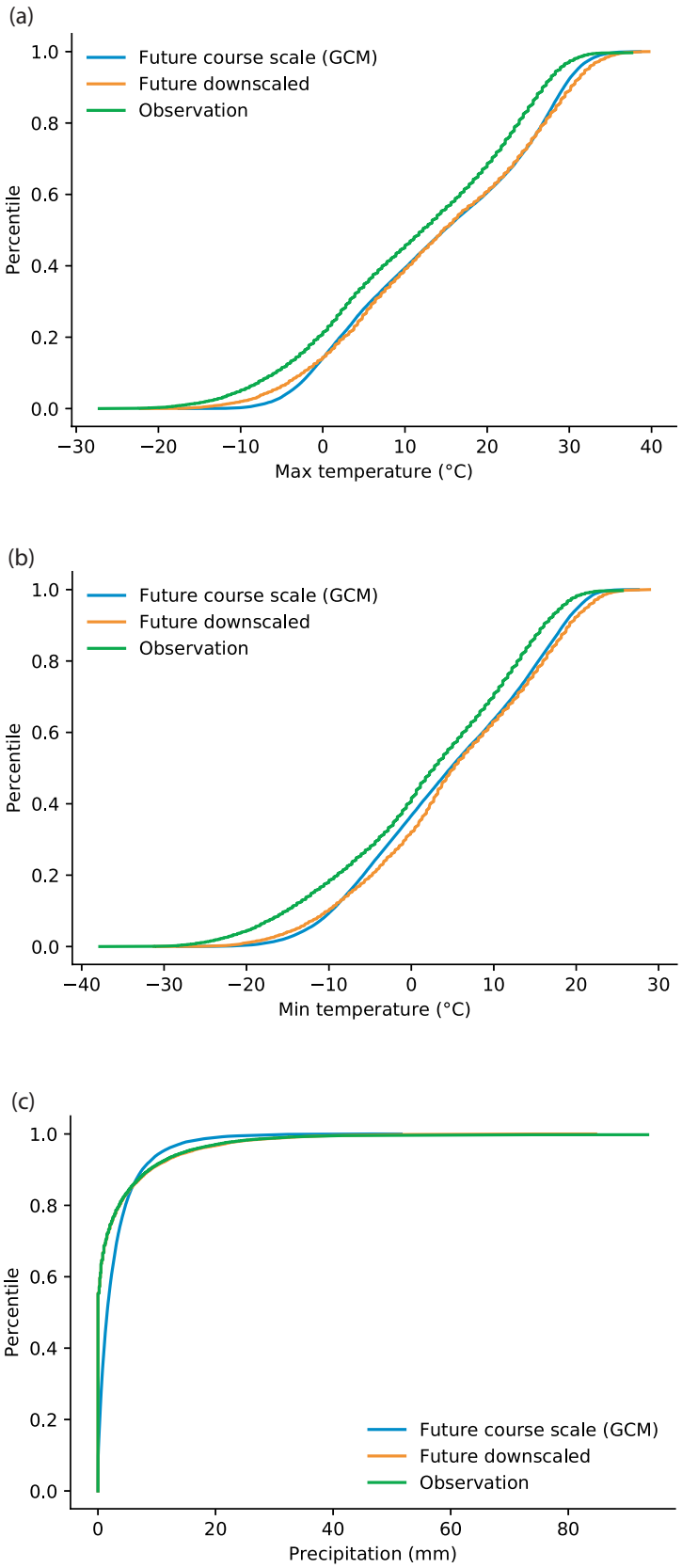

Figure 4.8: *Observation, future downscaled and future coarse-scale of a) maximum temperature, b) minimum temperature and c) precipitation for RCP* 4.5

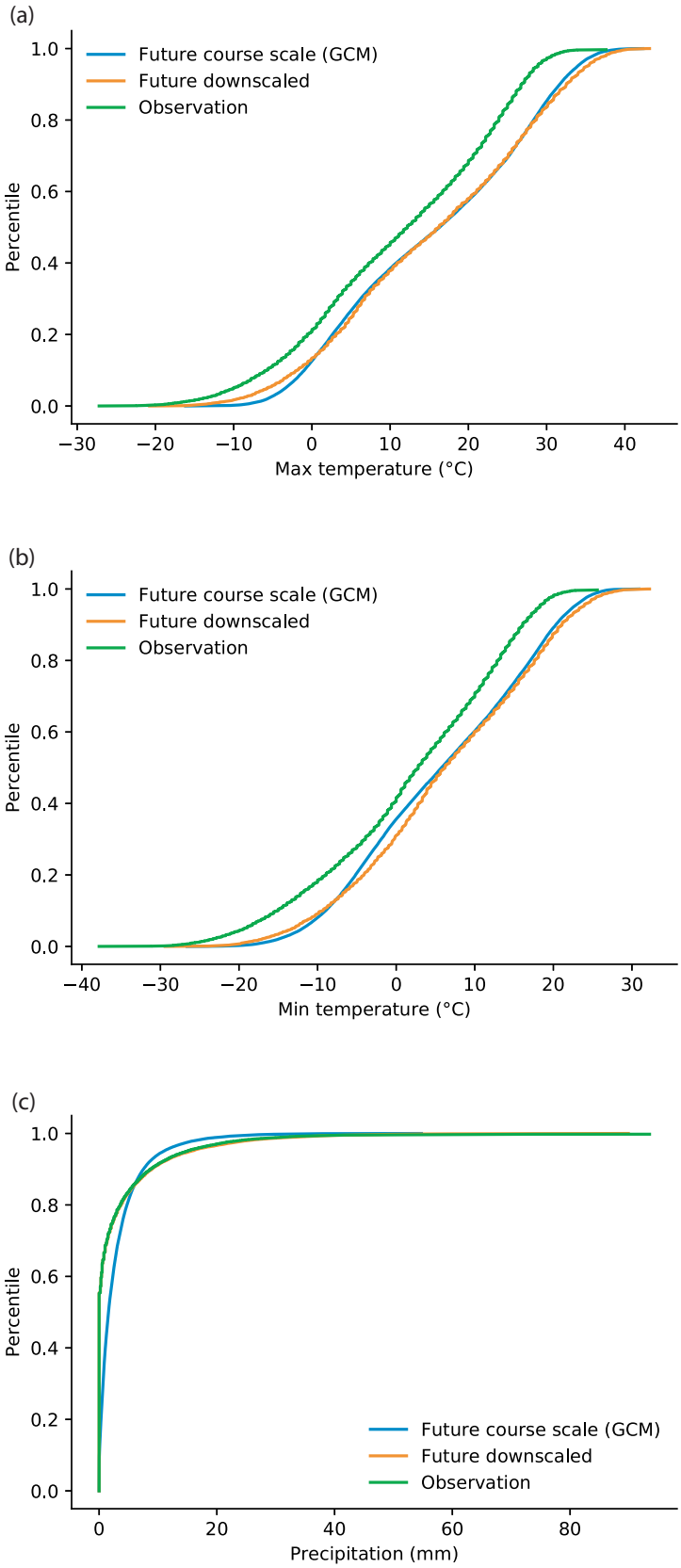

Figure 4.9: *Observation, future downscaled and future coarse-scale of a) maximum temperature, b) minimum temperature and c) precipitation for RCP* 8.5

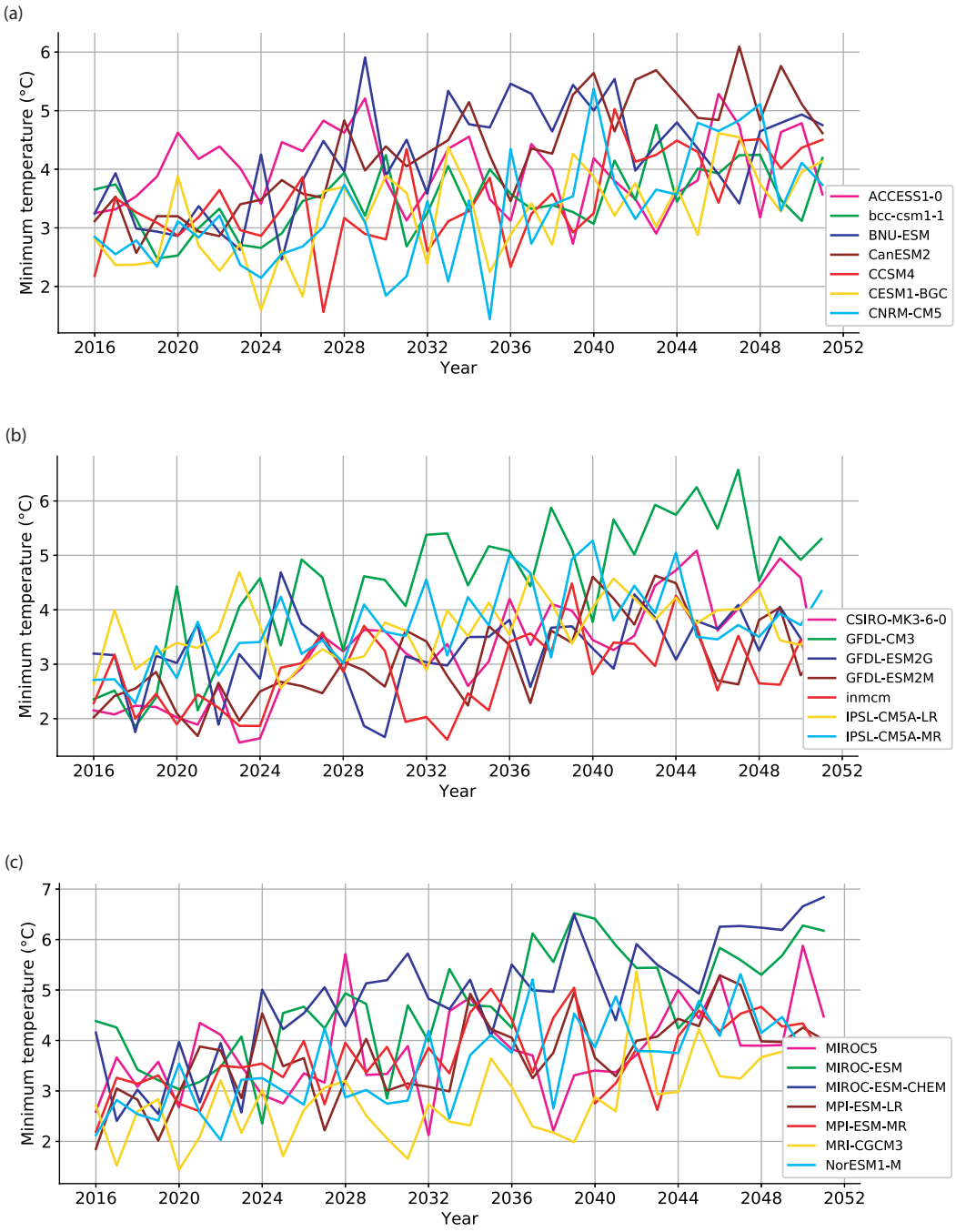

Figure 4.10: *Annual values of annual minimum temperature a) first series; b) second series; c) third series of GCMs outputs for RCP* 4.5 *from* 2015 *to* 2050*. The annual values are plotted at the end of the year.*

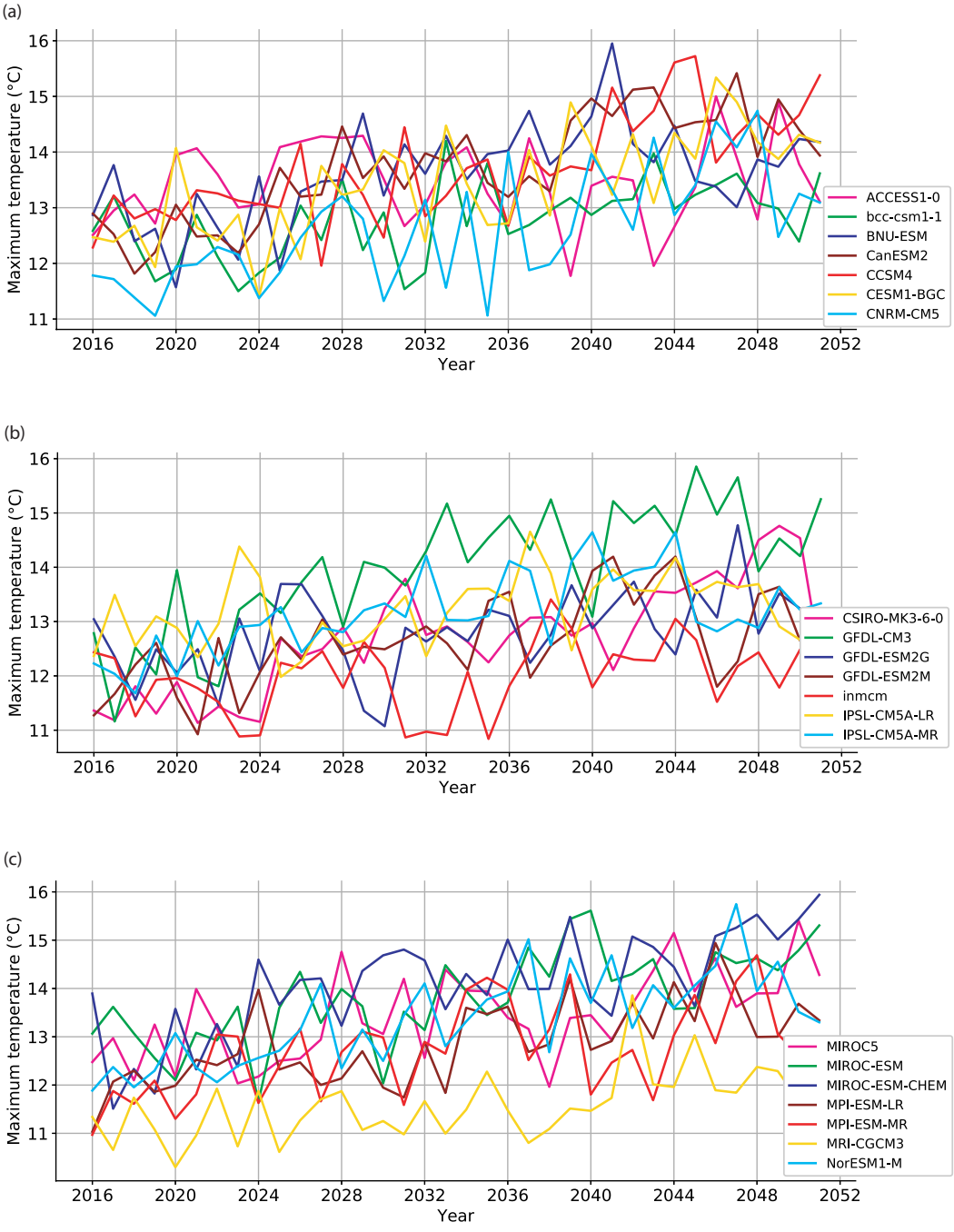

Figure 4.11: *Annual values of annual maximum temperature a) first series; b) second series; c) third series of GCMs outputs for RCP* 4.5 *from* 2015 *to* 2050*. The annual values are plotted at the end of the year.*

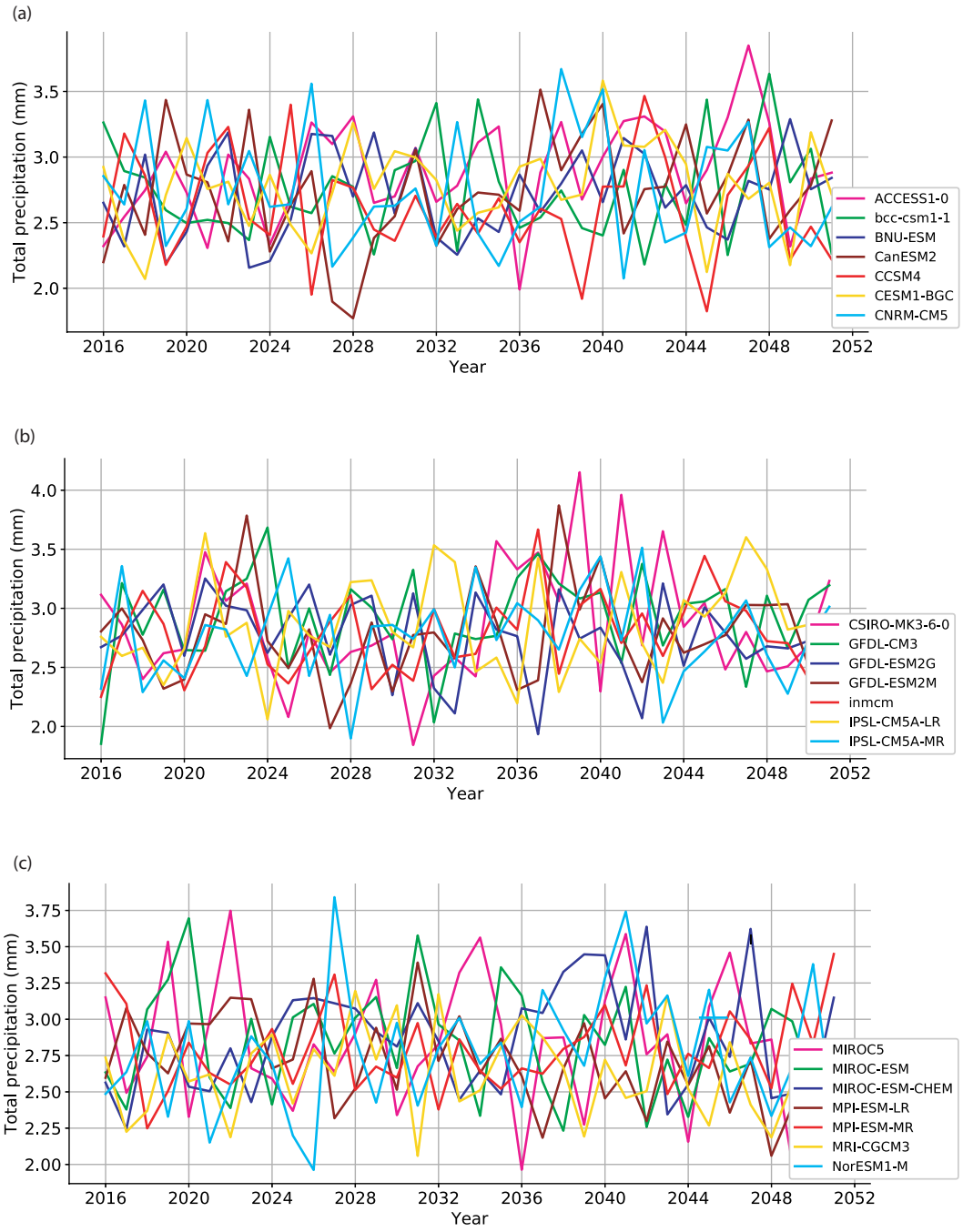

Figure 4.12: *Annual values of annual total precipitation a) first series; b) second series; c) third series of GCMs outputs for RCP* 4.5 *from* 2015 *to* 2050. *The annual values are plotted at the end of the year.*

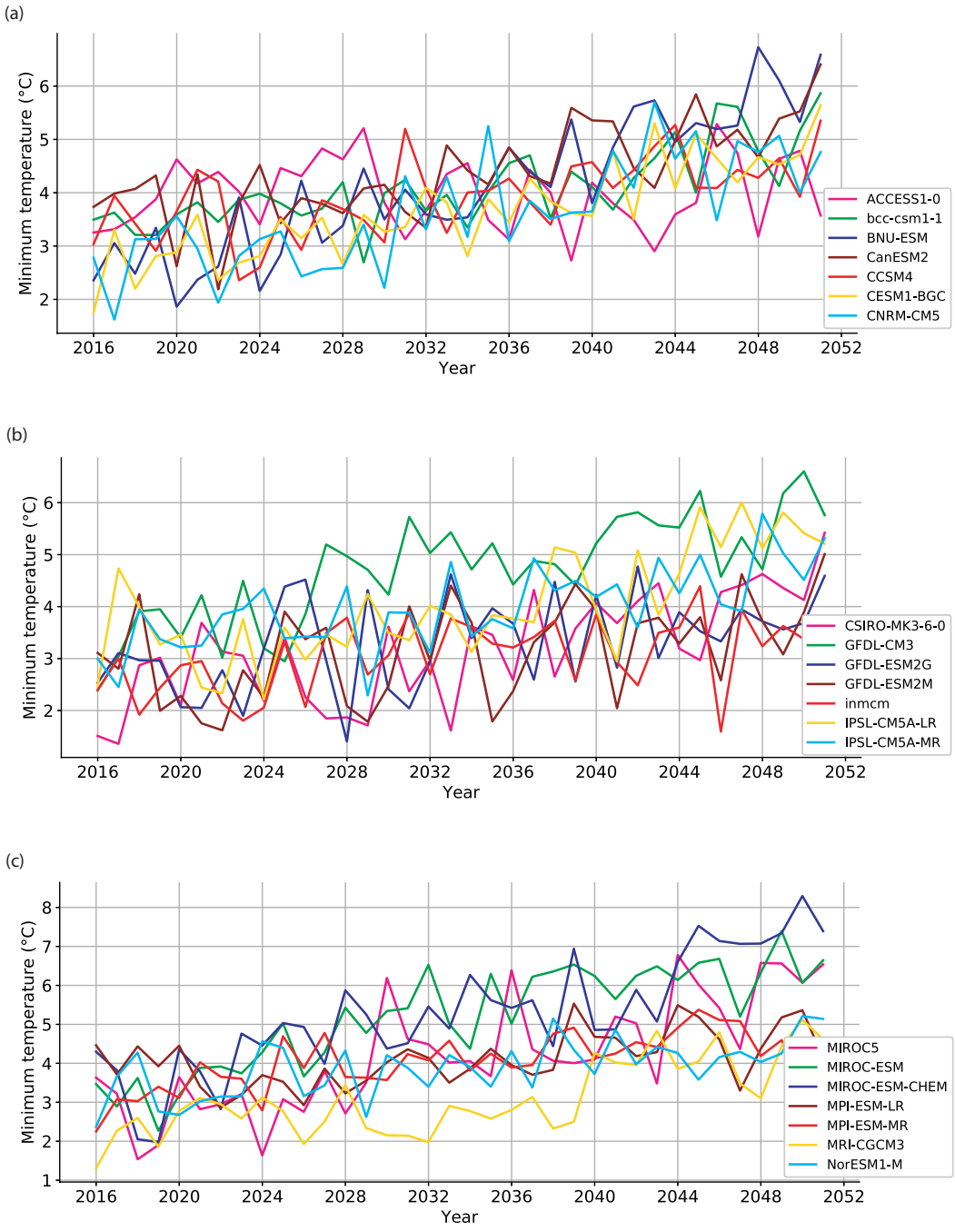

Figure 4.13: *Annual values of the a) first series, b) second series and c) third series of GCMs outputs for annual minimum temperature for RCP* 8.5 *from* 2015 *to* 2050*. The annual values are presented at the end of the year.*

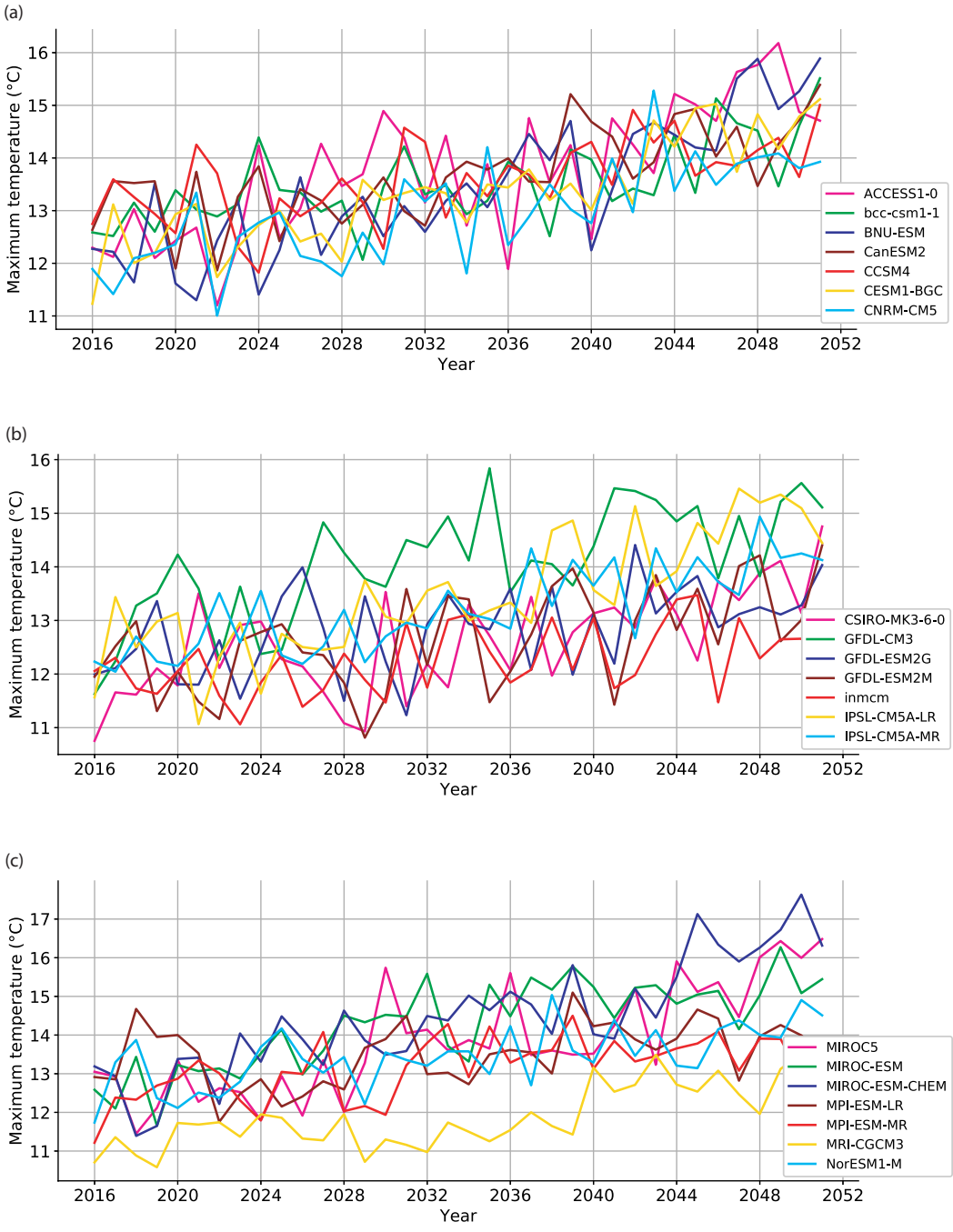

Figure 4.14: *Annual values of the a) first series, b) second series and c) third series of GCMs outputs for annual maximum temperature for RCP* 8.5 *from* 2015 *to* 2050*. The annual values are presented at the end of the year.*

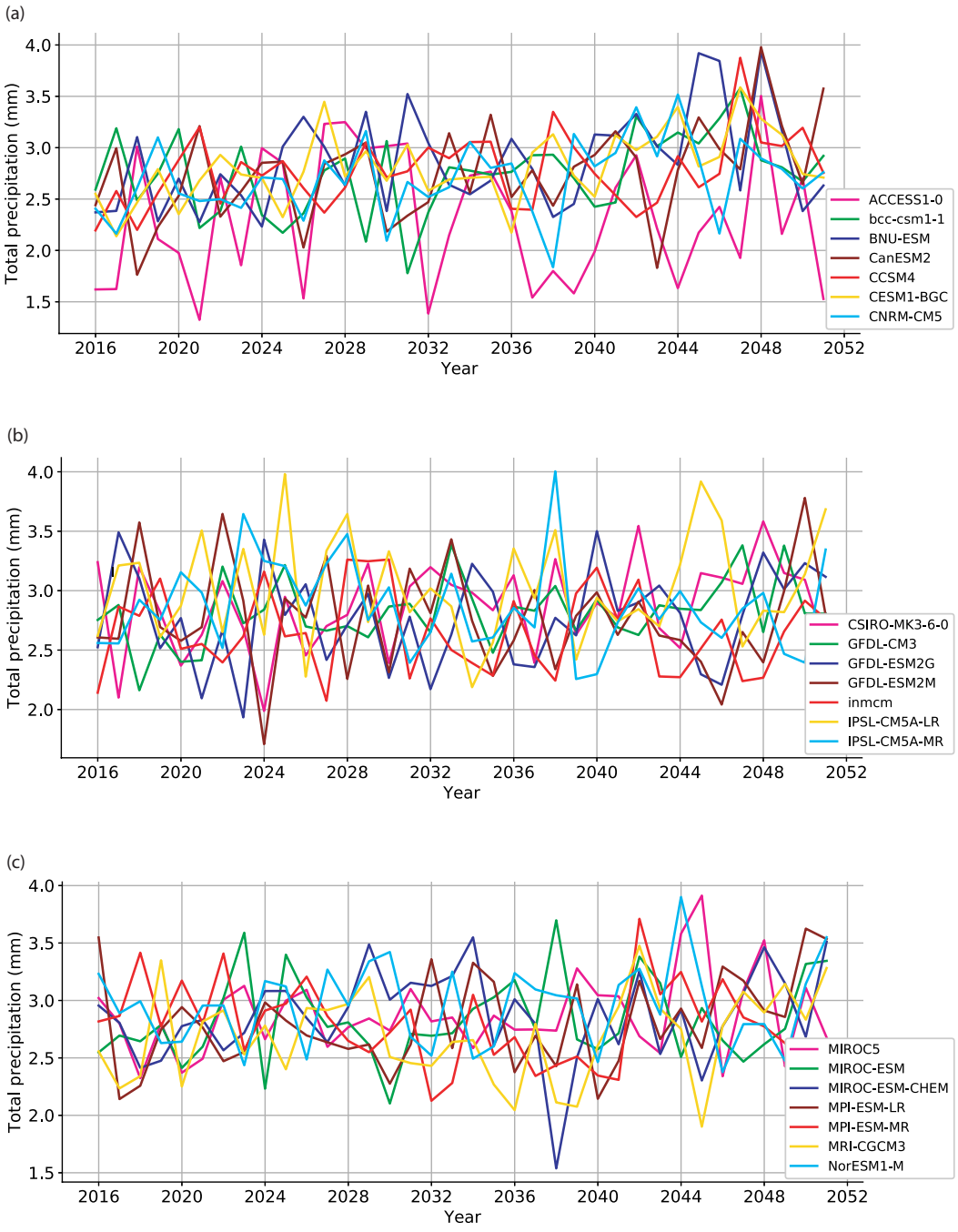

Figure 4.15: *Annual values of the a) first series, b) second series and c) third series of GCMs outputs for annual precipitation for RCP* 8.5 *from* 2015 *to* 2050*. The annual values are presented at the end of the year.*

bound and lower bound of variations. The ensemble GCM is the average of the outputs of the four selected models. Then, one can obtain downscaled future projections of minimum temperature, maximum temperature and precipitation up to the year of 2050. Fig. 4.16 shows the trend lines for the three variables in the next 30 years.

Fig. 4.16 shows that both minimum temperature and maximum temperature increase by time in future for RCP 4.5 and RCP 8.5. Maximum temperature increase by the rate 0.072 degree per year for both RCP 4.5 and RCP 8.5. Minimum temperature increase by the rate 0.072 degree per year for RCP 8.5 and 0.037 degree per year for RCP 4.5. Also, daily precipitation gradually increase in the next 30 years. These increasing trends have higher slope for RCP 8.5 in all three variables which is in accordance with the expectation.

4.6 Long-term forecast of UWC under changing climate

This research gives forecast of UWC for the next three decades. Bias corrections were made to future daily data of θ, Θ, and p for the time period of 2015 to 2050. The results of corrections are such that $R(x)$ ranges from 0 to 1.6 for p [Eq. (8)] and from -3.3 to 1.9 for θ [Eq. (9)] and from -5.7 to 1.3 for Θ [Eq. (9)]. For example, for July 1, 2040, $R(x)$ for p, θ and Θ are 0.26, 0.99, and 0.39, respectively.

The output from Bayesian linear regression is the DWC in the form of probability distributions. The output accuracy between two distinct cases are compared: a) the mean of the probability density function; b) the mean of 95% HPD. The comparison shows consistent results between the two cases. Therefore, for simplicity, the mean of the probability density function is selected as the average DWC (Fig. 4.17).

With respect to temperature threshold, if $\theta_0 \leq 9°$C, UWC is categorised as BWC. The average weekend value of BWC is 0.35 cubic metre per capita per year, and the average weekday value is 0.33 cubic metre per capita per year (Fig. 4.6). In Fig. 4.17, the BWC oscillates over time due to the assumption of stationary BWC for the next three decades. The SWC varies symmetrically about the middle of the season. The variations follow the seasonal temperature patterns. This is the case for both RCP 4.5 and RCP 8.5. The peak water consumption occurs in the month of July when the

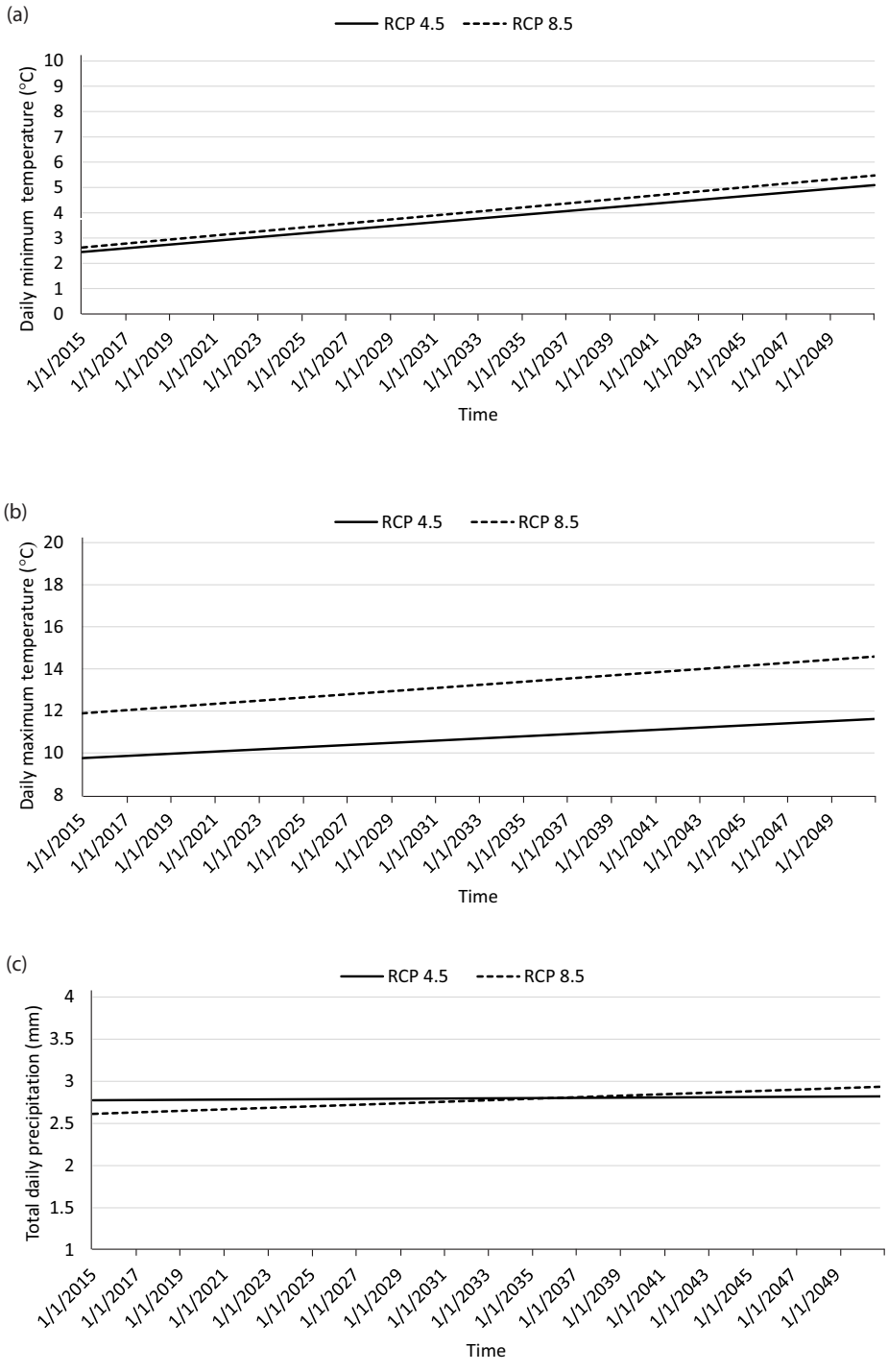

Figure 4.16: *The trends for annual a) minimum temperature, b) maximum temperature, and c) precipitation projection as the results of ensemble model by* 2050

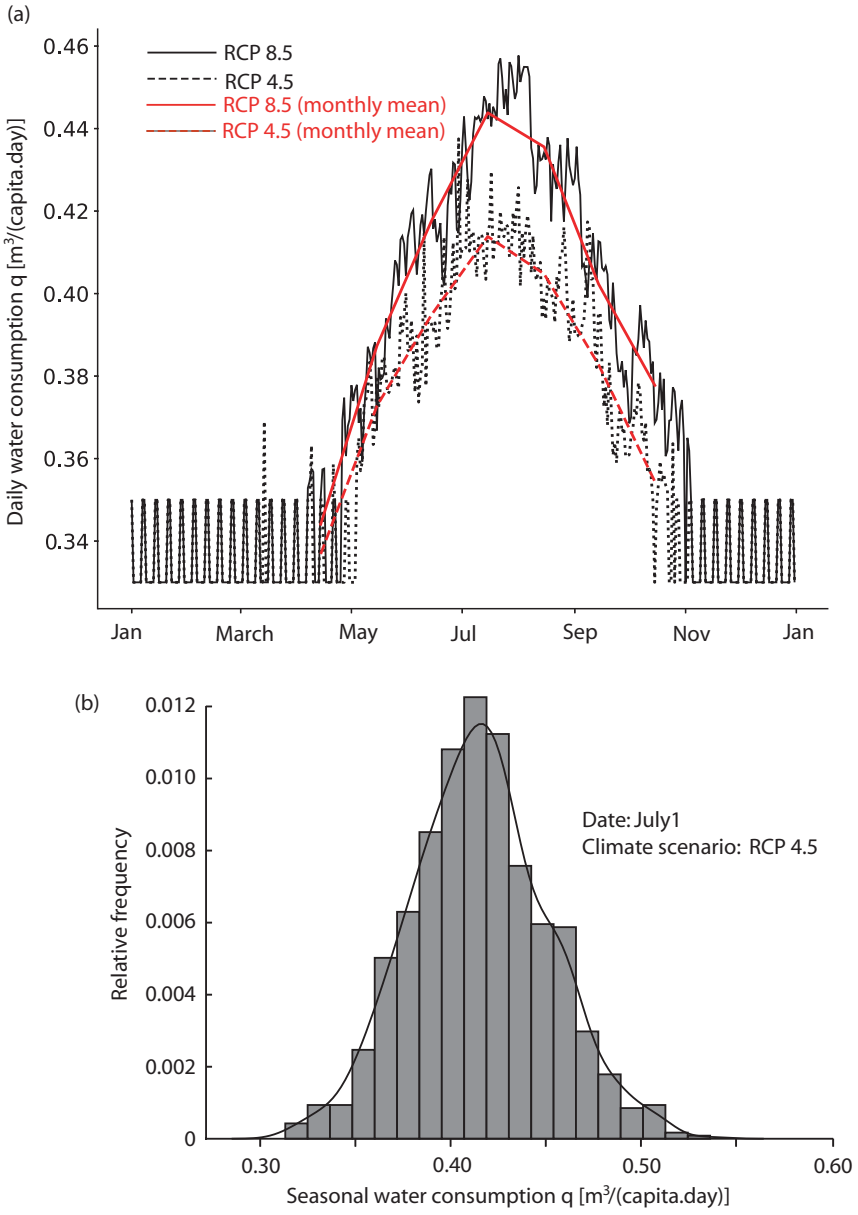

Figure 4.17: *(a) Typical daily urban water consumption using $RCP4.5$ and RCP 8.5 scenarios as input to the Bayesian linear regression model. Daily values of water consumption are the mean of the probability density function of water consumption (for the purpose of generating this graph, the year 2040 has been chosen). (b) An example of posterior predictive distribution of water consumption.*

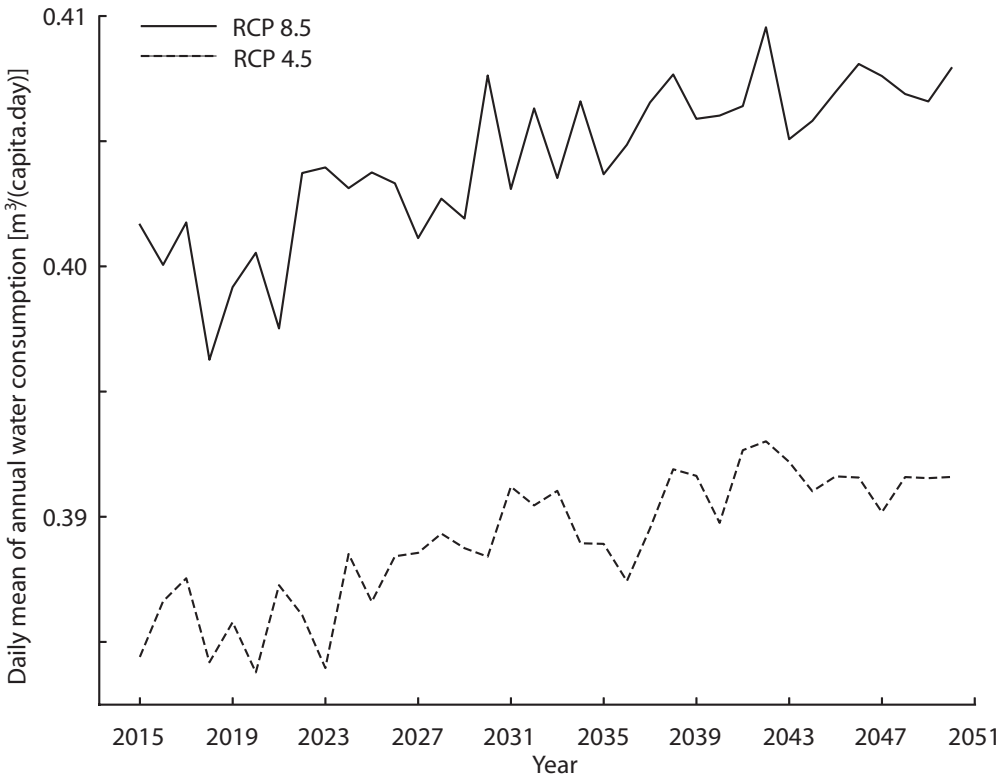

Figure 4.18: *Annual urban water consumption forecasted for the time period of* 2016 *to* 2050.

temperature is the highest over the year. RCP 8.5 has a greater impact on DWC than RCP 4.5, as expected. It is concluded that the worst-case scenario of carbon dioxide emission (RCP 8.5) will potentially lead to an increase in DWC.

The forecasted DWC over the year to obtain annual UWC are integrated, which is useful for long-term planning of water supply. The annual UWC (Fig. 4.18) shows a trend of increasing water consumption for the next three decades. Compared to RCP 4.5, RCP 8.5 is shown to cause a 4% increase in annual UWC.

It is important to note that uncertainties exist in the long-term forecast of SWC. The uncertainties are illustrated in Figure 5.10, showing the results of SWC against its predictors under the impact of RCP 8.5. The results are the possible ranges of 100 samples of the target variable drawn from the posterior [Eq. (5)]. The posterior of SWC comprises of 100 linear regressions in a three dimension space including θ, Θ and p. The lower bounds of θ and Θ for SWC are determined on the basis of θ_t as well as the value ranges of bias-corrected weather data for the time period of 2015 to 2050. The

distributed lines in Fig. 4.19 demonstrate the levels of uncertainties in SWC forecast considering the influence of any of the predictors. In Fig. 4.19a, SWC is shown to have the least variation with Θ. In Fig. 4.19b and Fig. 4.19c, the proposed Bayesian linear regression is less certain about the influence of θ and p on forecasted SWC for the next three decades.

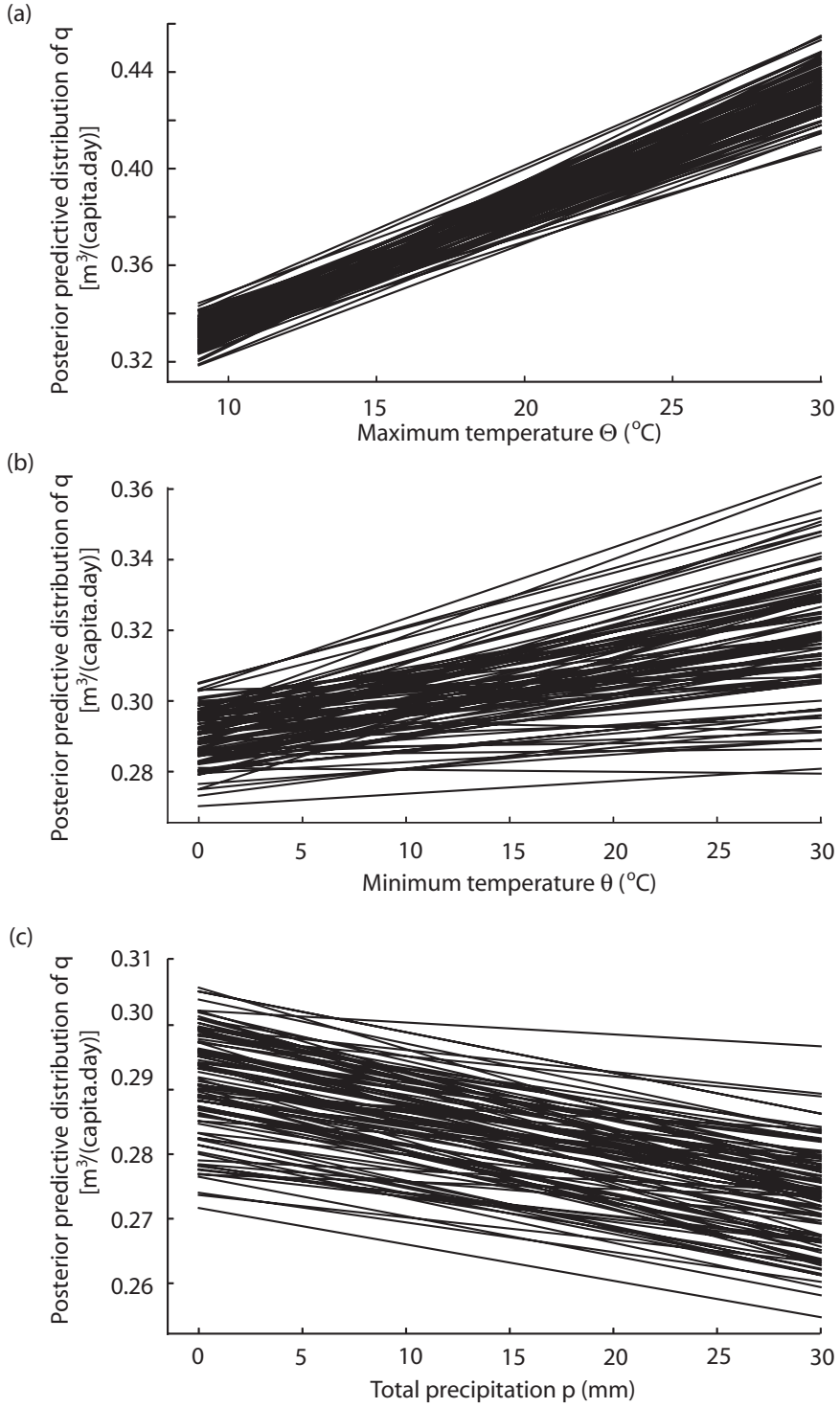

Figure 4.19: *Posterior predictive distributions of: (a) θ; b) Θ; c) p. The climate projections for* 2015-2050 *are based on the RCP 8.5 scenario.*

Chapter 5

Conclusion and Future Research Work

5.1 Conclusion remarks

This research took the stochastic approach to the problem of UWC forecast and made an application to the city of Brossard. The following conclusions have been reached:

(1) The proper choice of clustering UWC was determined based on θ and Θ. The choice B split the 933 objects into two clusters to create three clusters.

(2) Choice B was the preference in this study. It successfully separated SWC from BWC and showed that SWC and temperature were well correlated. At the same time, choice B grouped a small number of abnormal data points (from January to May 2011).

(3) It was suitable to determine the threshold temperature, θ_t, by regression of q against θ_0 for the individual years of 2012, 2013, and 2014 and the entire record period. The results showed that θ_t was equal to 9.5°C for 2012 and 2014, and 5.2°C for 2013. Given that other random weather variables could affect the transition, for practical purposes of accommodating uncertainties, it was suitable to take θ_t as 9°C.

(4) BWC was considered as stationary. Nonstationary features of water use records were captured in SWC and were lumped into variations affected by climate change. It was shown that BWC ranges from 0.33 to 0.36 m^3 per capita per day for the weekend, and from 0.32 to 0.35 m^3

per capita per day for weekdays. Although the BWC levels differed between weekends and weekdays, the lower and upper values of the ranges gave the same interval.

(5) The results of Bayesian linear regression gave as posterior histograms for β_0, β_1, β_2 and β_3, and σ. The histograms showed 95% Highest Posterior Density (HPD), which is a credible interval for the parameters. The good accuracy of the model was verified by Root Mean Squared Error and Mean Absolute Error, being 0.0450 and 0.0378, respectively.

(6) Both minimum temperature and maximum temperature increased with time in the future for RCP 4.5 and RCP 8.5. They increased at a rate of 0.072°C per year for RCP 8.5. Also, daily precipitation gradually increased in the next three decades. These increasing trends had a higher slope for RCP 8.5 in all three variables.

(7) The SWC varied symmetrically about the middle of the season, and the variations followed the seasonal temperature patterns. The peak water consumption occurred in the month of July when the temperature is the highest over the year. RCP 8.5 had a greater impact on DWC than RCP 4.5, as expected.

(8) The forecasted DWC over the year were integrated to obtain annual UWC. This is useful for long-term planning of water supply. The annual UWC showed a trend of increasing water consumption for the next three decades. Compared to RCP 4.5, RCP 8.5 was shown to cause a 4% increase in annual UWC.

(9) Uncertainties exist in the long-term forecast of SWC, considering the influence of any of the predictors. SWC was shown to have the least variation with θ. The proposed Bayesian linear regression was less certain about the influence of Θ and p on forecasted SWC for the next three decades.

5.2 Future work

Important future researches that this study motivates include:

- Investigate the effect of longer records of DWC on the estimate of temperature threshold, and the model parameters;
- Include a disaggregate socio-economic data set (household by household) for the estimation of BWC;
- Consider the impact of other weather variables such as wind speed and humidity on UWC.

One limitation of this study is that the available records of daily water consumption used for the development of Bayesian linear regression model are relatively short. Further efforts should be made to obtain longer records, which would improve the estimates of relevant model parameters. Also, including socio-economic data set and other weather variables can improve the completeness of the proposed predictor model.

References

Adamowski, J., Adamowski, K., & Prokoph, A. (2013). A spectral analysis based methodology to detect climatological influences on daily urban water demand. *Mathematical Geosciences*, *45*(1), 49–68.

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in montreal, canada. *Water Resources Research*, *48*(1).

Adamowski, J., & Karapataki, C. (2010). Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ann learning algorithms. *Journal of Hydrologic Engineering*, *15*(10), 729–743.

Ahmadalipour, A., Moradkhani, H., Castelletti, A., & Magliocca, N. (2019). Future drought risk in africa: Integrating vulnerability, climate change, and population growth. *Science of the Total Environment*, *662*, 672–686.

Ahmadalipour, A., Moradkhani, H., & Svoboda, M. (2017). Centennial drought outlook over the conus using nasa-nex downscaled climate ensemble. *International Journal of Climatology*, *37*(5), 2477–2491.

Alexander, L. V., Tapper, N., Zhang, X., Fowler, H. J., Tebaldi, C., & Lynch, A. (2009). Climate extremes: progress and future directions. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *29*(3), 317–319.

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

Almutaz, I., Ajbar, A., Khalid, Y., & Ali, E. (2012). A probabilistic forecast of water demand for

a tourist and desalination dependent city: Case of mecca, saudi arabia. *Desalination*, *294*, 53–59.

Altunkaynak, A., & Nigussie, T. A. (2017). Monthly water consumption prediction using season algorithm and wavelet transform–based models. *Journal of Water Resources Planning and Management*, *143*(6), 04017011.

Al-Zahrani, M. A., & Abo-Monasar, A. (2015). Urban residential water demand prediction based on artificial neural networks and time series models. *Water resources management*, *29*(10), 3651–3662.

Arturo, O. d. l. C., Alvarez-Chavez, C. R., Ramos-Corella, M. A., & Soto-Hernandez, F. (2017). Determinants of domestic water consumption in hermosillo, sonora, mexico. *Journal of cleaner production*, *142*, 1901–1910.

Ashouri, M., Haghighat, F., Fung, B. C., Lazrak, A., & Yoshino, H. (2018). Development of building energy saving advisory: A data mining approach. *Energy and Buildings*, *172*, 139–151.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of applied Econometrics*, *25*(1), 71–92.

Bennett, J. C., Grose, M. R., Corney, S. P., White, C. J., Holz, G. K., Katzfey, J. J., ... Bindoff, N. L. (2014). Performance of an empirical bias-correction of a high-resolution climate dataset. *International Journal of Climatology*, *34*(7), 2189–2204.

Bishop, C. M., & Tipping, M. E. (2003). Bayesian regression and classification. *Nato Science Series sub Series III Computer And Systems Sciences*, *190*, 267–288.

Boé, J., Terray, L., Habets, F., & Martin, E. (2007). Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*(12), 1643–1655.

Brossard. (2019a). *Municipal services, by-law.* Retrieved from http://www.ville.brossard.qc.ca/services-citoyens/Reglements/Reglements.aspx?lang=en-ca (Accessed: 2019-02-30)

Brossard. (2019b). *Municipal services, watering.* Retrieved from http://www.ville.brossard.qc.ca/services-citoyens/eau/Eau/Arrosage.aspx?lang=

en-ca (Accessed: 2019-02-30)

Buck, S., Soldati, H., & Sunding, D. L. (2015). *Forecasting urban water demand in california: Rethinking model evaluation* (Tech. Rep.).

Chang, H., Praskievicz, S., & Parandvash, H. (2014). Sensitivity of urban water consumption to weather and climate variability at multiple temporal scales: The case of portland, oregon. *International Journal of Geospatial and Environmental Research*, *1*(1), 7.

Curran, J. M. (2005). An introduction to bayesian credible intervals for sampling error in dna profiles. *Law, Probability and Risk*, *4*(1-2), 115–126.

Davidson-Pilon, C. (2015). *Bayesian methods for hackers: probabilistic programming and bayesian inference*. Addison-Wesley Professional.

Eslamian, S. A., Li, S. S., & Haghighat, F. (2016). A new multiple regression model for predictions of urban water use. *Sustainable Cities and Society*, *27*, 419–429.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37–37.

Gato, S., Jayasuriya, N., Hadgraft, R., & Roberts, P. (2005). A simple time series approach to modelling urban water demand. *Australasian Journal of Water Resources*, *8*(2), 153–164.

Gato, S., Jayasuriya, N., & Roberts, P. (2007). Temperature and rainfall thresholds for base use urban water demand modelling. *Journal of hydrology*, *337*(3-4), 364–376.

Ghiassi, M., Zimbra, D. K., & Saidane, H. (2008). Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management*, *134*(2), 138–146.

Ghosh, B., Basu, B., & O'Mahony, M. (2007). Bayesian time-series model for short-term traffic flow forecasting. *Journal of transportation engineering*, *133*(3), 180–189.

Godsill, S. J. (2001). On the relationship between markov chain monte carlo methods for model uncertainty. *Journal of computational and graphical statistics*, *10*(2), 230–248.

Hamlet, A., Carrasco, P., Deems, J., Elsner, M., Kamstra, T., Lee, C., ... others (2010). Final report for the columbia basin climate change scenarios project. *University of Washington, Climate Impacts Group, Seattle, Washington, DC*.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Haque, M. M., Rahman, A., Hagare, D., & Kibria, G. (2014). Probabilistic water demand forecasting using projected climatic data for blue mountains water supply system in australia. *Water resources management*, *28*(7), 1959–1971.

Heckerman, D. (1997). Bayesian networks for data mining. *Data mining and knowledge discovery*, *1*(1), 79–119.

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, *20*(3), 197–243.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

House-Peters, L., Pratt, B., & Chang, H. (2010). Effects of urban spatial structure, sociodemographics, and climate on residential water consumption in hillsboro, oregon. *JAWRA Journal of the American Water Resources Association*, *46*(3), 461–472.

Ismail, M. A., Sadiq, R., Soleymani, H. R., & Tesfamariam, S. (2011). Developing a road performance index using a bayesian belief network model. *Journal of the Franklin Institute*, *348*(9), 2539–2555.

Jain, S., Salunke, P., Mishra, S. K., Sahany, S., & Choudhary, N. (2019). Advantage of nex-gddp over cmip5 and cordex data: Indian summer monsoon. *Atmospheric Research*.

Jaramillo, P., & Nazemi, A. (2018). Assessing urban water security under changing climate: Challenges and ways forward. *Sustainable cities and society*, *41*, 907–918.

Kenney, D. S., Goemans, C., Klein, R., Lowrey, J., & Reidy, K. (2008). Residential water demand management: lessons from aurora, colorado 1. *JAWRA Journal of the American Water Resources Association*, *44*(1), 192–207.

Khatri, K., & Vairavamoorthy, K. (2009). Water demand forecasting for the city of the future against the uncertainties and the global change pressures: case of birmingham. In *World environmental and water resources congress 2009: Great rivers* (pp. 1–15).

Kim, H., & Melhem, H. (2004). Damage detection of structures by wavelet analysis. *Engineering Structures*, *26*(3), 347–362.

Koutroulis, A., Papadimitriou, L., Grillakis, M., Tsanis, I., Warren, R., & Betts, R. (2019). Global water availability under high-end climate change: A vulnerability based assessment. *Global*

and planetary change, *175*, 52–63.

Lambert, B. (2018). *A student's guide to bayesian statistics*. Sage.

Li, H., Sheffield, J., & Wood, E. F. (2010). Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres*, *115*(D10).

Li, Q., Gu, L., Augenbroe, G., Wu, C. J., & Brown, J. (2015). Calibration of dynamic building energy models with multiple responses using bayesian inference and linear regression models. *Energy Procedia*, *78*, 979–984.

Liu, G., Yang, J., Hao, Y., & Zhang, Y. (2018). Big data-informed energy efficiency assessment of china industry sectors based on k-means clustering. *Journal of cleaner production*, *183*, 304–314.

Liu, J., Yang, H., Gosling, S. N., Kummu, M., Flörke, M., Pfister, S., ... others (2017). Water scarcity assessments in the past, present, and future. *Earth's future*, *5*(6), 545–559.

Mahadevan, S. (1997). Monte carlo simulation. *Mechanical Engineering-New York and Basel-Marcel Dekker-*, 123–146.

Mandapaka, P. V., & Lo, E. Y. (2018). Assessment of future changes in southeast asian precipitation using the nasa earth exchange global daily downscaled projections data set. *International Journal of Climatology*, *38*(14), 5231–5244.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J.-F., ... others (2011). The rcp greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic change*, *109*(1-2), 213.

Miro, M. E., Groves, D. G., Catt, D., Miller, B., & Social, R. (2018). *Estimating future water demand for san bernardino valley municipal water district*. RAND.

Mooney, C. Z. (1997). *Monte carlo simulation* (Vol. 116). Sage Publications.

Mouatadid, S., & Adamowski, J. (2017). Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal*, *14*(6), 630–638.

Mudgal, A., Hallmark, S., Carriquiry, A., & Gkritza, K. (2014). Driving behavior at a roundabout: A hierarchical bayesian regression analysis. *Transportation research part D: transport and environment*, *26*, 20–26.

Panagopoulos, G. P. (2014). Assessing the impacts of socio-economic and hydrological factors on urban water demand: A multivariate statistical approach. *Journal of hydrology*, *518*, 42–48.

Parandvash, G. H., & Chang, H. (2016). Analysis of long-term climate change on per capita water demand in urban versus suburban areas in the portland metropolitan area, usa. *Journal of Hydrology*, *538*, 574–586.

Polebitski, A. S., & Palmer, R. N. (2009). Seasonal residential water demand forecasting for census tracts. *Journal of water resources planning and management*, *136*(1), 27–36.

Praskievicz, S., & Chang, H. (2009). Identifying the relationships between urban water consumption and weather variables in seoul, korea. *Physical Geography*, *30*(4), 324–337.

Prusty, R. M., Das, A., & Patra, K. C. (2018). Climate change impact assessment under cordex south-asia rcm scenarios on water resources of the brahmani and baitarini river basin, india.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.

Raghavan, S. V., Hur, J., & Liong, S.-Y. (2018). Evaluations of nasa nex-gddp data over southeast asia: present and future climates. *Climatic change*, *148*(4), 503–518.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., ... Rafaj, P. (2011). Rcp 8.5—a scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1-2), 33.

Rinaudo, J.-D. (2015). Long-term water demand forecasting. In *Understanding and managing urban water in transition* (pp. 239–268). Springer.

Romano, G., Salvati, N., & Guerrini, A. (2014). Estimating the determinants of residential water demand in italy. *Water*, *6*(10), 2929–2945.

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 73–79.

Ruth, M., Bernier, C., Jollands, N., & Golubiewski, N. (2007). Adaptation of urban water supply infrastructure to impacts from climate and socioeconomic changes: the case of hamilton, new zealand. *Water Resources Management*, *21*(6), 1031–1045.

Schmidli, J., Frei, C., & Vidale, P. L. (2006). Downscaling from gcm precipitation: a benchmark for dynamical and statistical downscaling methods. *International journal of climatology*, *26*(5), 679–689.

Semenov, M. A., & Stratonovitch, P. (2010). Use of multi-model ensembles from global climate models for assessment of climate change impacts. *Climate research*, *41*(1), 1–14.

Sen, P. K. (1968). Estimates of the regression coefficient based on kendall's tau. *Journal of the American statistical association*, *63*(324), 1379–1389.

Shabani, S., Yousefi, P., & Naser, G. (2017). Support vector machines in urban water demand forecasting using phase space reconstruction. *Procedia Engineering*, *186*, 537–543.

Singh, G., Goel, A., & Choudhary, M. (2015). An inventory of methods and models for domestic water demand forecasting: a review. *J. Indian Water Resources Society*, *35*(3), 34–45.

Singh, S., & Yassine, A. (2018). Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, *11*(2), 452.

Sriram, N. (n.d.). A study on machine learning techniques for data mining.

Statistics-Canada. (2019). *Focus on geography series'*. Retrieved from https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-cma-eng.cfm?LANG=Eng&GK=CMA&GC=462&TOPIC=1 (Accessed: 2019-08-1)

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., ... Midgley, P. (2014). Ipcc, 2013: Summary for policymakers in:climate change 2013: The physical science basis.

Stoker, P., & Rothfeder, R. (2014). Drivers of urban water use. *Sustainable Cities and Society*, *12*, 1–8.

Syme, G. J., Shao, Q., Po, M., & Campbell, E. (2004). Predicting and understanding home garden water use. *Landscape and Urban Planning*, *68*(1), 121–128.

Taleb, T., & Kaddour, M. (2017). Hierarchical agglomerative clustering schemes for energy-efficiency in wireless sensor networks. *Transport and Telecommunication Journal*, *18*(2), 128–138.

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*(4), 485–498.

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., ... others (2011). Rcp4. 5: a pathway for stabilization of radiative forcing by 2100. *Climatic change*, *109*(1-2), 77.

Tiwari, M. K., & Adamowski, J. (2013). Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resources Research*, *49*(10), 6486–6507.

Tiwari, M. K., & Adamowski, J. F. (2014). Medium-term urban water demand forecasting with limited data using an ensemble wavelet–bootstrap machine-learning approach. *Journal of Water Resources Planning and Management*, *141*(2), 04014053.

Tiwari, M. K., & Adamowski, J. F. (2017). An ensemble wavelet bootstrap machine learning approach to water demand forecasting: A case study in the city of calgary, canada. *Urban Water Journal*, *14*(2), 185–201.

Wang, Y., Duan, L., Liu, T., Li, J., & Feng, P. (2019). A non-stationary standardized streamflow index for hydrological drought using climate and human-induced indices as covariates. *Science of The Total Environment*, 134278.

Wong, J. S., Zhang, Q., & Chen, Y. D. (2010). Statistical modeling of daily urban water consumption in hong kong: Trend, changing patterns, and forecast. *Water resources research*, *46*(3).

Yu, Z., Fung, B. C., & Haghighat, F. (2013). Extracting knowledge from building-related data—a data mining framework. In *Building simulation* (Vol. 6, pp. 207–222).

Yu, Z., Fung, B. C., Haghighat, F., Yoshino, H., & Morofsky, E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and buildings*, *43*(6), 1409–1417.

Yu, Z., Haghighat, F., Fung, B. C., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, *42*(10), 1637–1646.

Yu, Z. J., Haghighat, F., Fung, B. C., Morofsky, E., & Yoshino, H. (2011). A methodology for identifying and improving occupant behavior in residential buildings. *Energy*, *36*(11), 6596–6608.

Yuan, X.-C., Sun, X., Zhao, W., Mi, Z., Wang, B., & Wei, Y.-M. (2017). Forecasting china's regional energy demand by 2030: A bayesian approach. *Resources, Conservation and Recycling*, *127*, 85–95.

Zhang, X., Flato, G., Kirchmeier-Young, M., Vincent, L., Wan, H., Wang, X., ... Kharin, V. (2019).

Changes in temperature and precipitation across canada; chapter 4 in bush e, lemmen ds.(eds.) canada's changing climate report. *Government of Canada, Ottawa, Ontario*, 112–193.

Zhou, S. L., McMahon, T. A., Walton, A., & Lewis, J. (2000). Forecasting daily urban water demand: a case study of melbourne. *Journal of hydrology*, *236*(3-4), 153–164.

Zhuang, X., Li, Y., Nie, S., Fan, Y., & Huang, G. (2018). Analyzing climate change impacts on water resources under uncertainty using an integrated simulation-optimization approach. *Journal of Hydrology*, *556*, 523–538.