

**DEVELOPMENT OF MULTI-MODEL ENSEMBLES FOR CLIMATE
DOWNSCALING IN ONTARIO, CANADA**

**DEVELOPMENT OF MULTI-MODEL ENSEMBLES FOR CLIMATE
DOWNSCALING IN ONTARIO, CANADA**

By

XINYI LI

B.Sc.

Faculty of Engineering
Department of Civil Engineering

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfilment of the Requirements

For the Degree

Master of Applied Science

McMaster University

© Copyright by Xinyi Li, December 2019

MASTER OF APPLIED SCIENCE (2019)
(Civil Engineering)

McMaster University
Hamilton, Ontario

TITLE: Development of Multi-model Ensembles for
Climate Downscaling in Ontario, Canada

AUTHOR: Xinyi Li
B.Sc. (Beijing Normal University)

SUPERVISOR: Professor Zhong Li

NUMBER OF PAGES: xvi, 114

Abstract

Climate change has widespread impacts on the environment, economy, and municipal planning. Thus, generating accurate, high resolution climate predictions could aid in the assessment of the impacts of climate change on a local scale. Multi-model ensembles have been proven to improve the accuracy of climate prediction, and machine learning techniques are a promising tool for temperature downscaling. The thesis investigates machine learning and statistical methods in the development of multi-model ensembles for climate downscaling.

Firstly, three neural network algorithms are used to develop multi-model ensembles for daily mean temperature downscaling, including Multi-layer Perceptron (MLP), Time-lagged Feed-forward Neural Network (TLFN) and Nonlinear Auto-Regressive Network with exogenous inputs (NARX). The inputs and outputs are the simulated daily mean temperatures obtained from six Regional Climate Models (RCMs) collected from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) archive and observed daily mean temperatures collected from the Digital Archive of Canadian Climatological Data, respectively. A case study of Big Trout Lake in Ontario, Canada is carried out as a preliminary study to evaluate the performance of the proposed neural network models. The results show that the neural network based ensembles outperformed each of the individual regional climate models and generated predictions with smaller

fluctuations.

Secondly, the thesis investigates and compares the applicability and performance of machine learning and statistical methods in developing multi-model ensembles for downscaling long-term daily temperature. The machine learning methods include Long Short-Term Memory (LSTM) networks and Support Vector Machine (SVM) and the statistical methods include arithmetic ensemble mean (EM) and Multiple Linear Regression (MLR). These ensembles share the same input and output variables with the preliminary study. The performance of the proposed machine learning and statistical ensembles are evaluated at twelve meteorological stations across Ontario, Canada. The results show that multi-model ensembles with machine learning or statistical techniques all performed well at downscaling daily temperature, and had similar performance with relatively high accuracy. This is the first attempt to apply advanced machine learning techniques and compare them with statistical methods in developing multi-model ensembles for downscaling in Canadian communities. The results provide a technical basis for applying statistical and machine learning methods to generate long-term high-resolution daily temperature projections. The generated climate projections will also provide useful information to support climate adaptation in Ontario.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. Zhong Li for her invaluable guidance, support, encouragement and thoughtful suggestions through the Master's process. It is an honor to work with her.

I would also like to thank my examination committee, Dr. Yiping Guo and Dr. Sonia Hassini for their guidance, and advice.

I would like to express my gratitude to my co-authors and collaborators, Wendy Huang, Pengxiao Zhou, and Qianqian Zhang, for all of our discussions that directed this work.

I gratefully acknowledge the funding from the Natural Sciences and Engineering Research Council of Canada, and the Department of Civil Engineering at McMaster University.

Lastly, I would like to thank my family for their love, support, and encouragement.

Publication List

This thesis consists of the following papers:

Paper I

Li, X., Li, Z., Zhang, Q., Zhou, P., and Huang, W. (2019). Prediction of Long-Term Near-Surface Temperature Based on NA-CORDEX Output. *Journal of Environmental Informatics Letters*, 2 (1), 10-18.

Paper II

Li, X., Li, Z., Huang, W., and Zhou, P., Performance of statistical and machine-learning ensembles for daily temperature downscaling. Submitted to *Theoretical and Applied Climatology* in October 2019.

Co-Authorship

This thesis was prepared on the “sandwich” format in accordance with the guidelines provided by the School of Graduate Studies, McMaster University and was co-authored.

For the paper presented in Chapter 3, the model development and data analysis were conducted by X. Li in consultation with Dr. Z. Li, Q. Zhang, and P. Zhou. The paper was written by X. Li and edited by Dr. Z. Li and Dr. W. Huang.

For the paper presented in Chapter 4, the model development and data analysis were completed by X. Li in consultation with Dr. Z. Li and P. Zhou. The paper was written by X. Li and edited by Dr. Z. Li and Dr. W. Huang.

Table of Contents

Abstract.....	iii
Acknowledgments	v
Publication List	vi
Co-Authorship.....	vii
Table of Contents.....	viii
List of Tables.....	xii
List of Figures.....	xiii
List of Symbols	xv
1. Introduction.....	1
1.1 Background	1
1.2 Objective and Scope	3
1.3 Organization of the Thesis	4
2. Literature Review	5
2.1 Regional Climate Models	5
2.2 Statistical Downscaling.....	6
2.2.1 Multiple Linear Regression.....	6
2.2.2 Neural Networks	7

2.2.3	Support Vector Machines	11
2.3	Multi-model Ensembles	12
2.4	Downscaling Studies in Ontario	14
3.	Prediction of Long-term Near-surface Temperature based on NA-CORDEX	
Output	16
	Abstract	16
3.1	Introduction	17
3.2	Methodology	20
3.2.1	Multi-layer Perceptron (MLP)	20
3.2.2	Time-lagged Feed-forward Neural Network (TLFN)	22
3.2.3	Nonlinear Auto-Regressive Networks with Exogenous Inputs (NARX) ..	
	24
3.3	Study Area and Data Collection.....	25
3.4	Neural Network Design and Training	27
3.5	Results and Discussion	28
3.5.1	Neural Networks Performance.....	28
3.5.2	Comparison between RCMs and Neural Networks	32
3.5.3	Comparison of Performance between MLP, TLFN, and NARX	35
3.6	Conclusions.....	36

3.7	References.....	38
4.	Performance of Statistical and Machine Learning Ensembles for Daily Temperature Downscaling.....	45
	Abstract.....	45
4.1	Introduction.....	47
4.2	Study Area and Data	53
4.3	Methodology	58
4.3.1	Long Short-Term Memory Networks	58
4.3.2	Support Vector Machine.....	60
4.3.3	Statistical Methods.....	62
4.3.4	Design and Training of Downscaling Models	62
4.4	Results and Discussion	64
4.4.1	Downscaling Performance of Long Short-Term Memory networks	64
4.4.2	Downscaling Performance of Support Vector Machine.....	70
4.4.3	Downscaling Performance of Statistical Methods	71
4.4.4	Comparison Between the Downscaling Performance of Machine Learning and Statistical Techniques.....	74
4.4.5	Current Difficulties in Predicting Extreme Values.....	78
4.4.6	Performance Comparison among 12 Stations over Ontario.....	80

4.5	Conclusions.....	83
4.6	References.....	85
5.	Conclusions.....	97
	References.....	100

List of Tables

Table 3-1 Comparison of performance between RCMs and neural networks	33
Table 4-1 Information on selected stations and their corresponding nearest RCM grid....	55
Table 4-2 GCM and RCM combinations	57
Table 4-3 Statistical performance of LSTM ensemble	66
Table 4-4 Statistical performance of SVM ensemble	71
Table 4-5 Statistical performance of mean ensemble	72
Table 4-6 Statistical performance of MLR ensemble	74

List of Figures

Fig. 3-1 Structure of MLP with one input layer, one hidden layer, and one output layer. .21	
Fig. 3-2 Structure of TLFN with one input layer, one hidden layer, and a delay-line with memory depth of k . (z^{-1} is an operator that delays the input by one sample). (Dibike and Coulibaly, 2006).....23	
Fig. 3-3 Structure of NARX with one input layer, one hidden layer, and one output layer (z^{-1} denotes delay for one time step).25	
Fig. 3-4 Time series plot of observed and predicted daily near-surface temperature values obtained by MLP with 12 neurons.29	
Fig. 3-5 Scatter plots of observed and predicted daily near-surface temperature values obtained by MLP with 12 neurons.29	
Fig. 3-6 Time series plot of observed and predicted daily near-surface temperature values obtained by TLFN with 5 neurons and a time lag of 3 days.30	
Fig. 3-7 Scatter plots of observed and predicted daily near-surface temperature values obtained by TLFN with 5 neurons and a time lag of 3 days.31	
Fig. 3-8 Time series plot of observed and predicted daily near-surface temperature values obtained by NARX with 15 neurons and a time lag of 3 days.32	
Fig. 3-9 Scatter plot of observed and predicted daily near-surface temperature values obtained by NARX with 15 neurons and a time lag of 3 days.32	

Fig. 3-10 Time series plot of 6 RCMs and neural networks of (a - c) winter and (d - f) summer in 1989.....	35
Fig. 4-1 Locations of the 12 selected meteorological stations.....	54
Fig. 4-2 A LSTM memory cell with input, forget, and output gates.....	60
Fig. 4-3 Scatter plots of the first 4 stations for testing datasets	67
Fig. 4-4 Scatter plots of the middle 4 stations for testing datasets.....	68
Fig. 4-5 Scatter plots of the last 4 stations for testing datasets	69
Fig. 4-6 A time series plot of ensembles of the Toronto Pearson International Airport station in 1987	75
Fig. 4-7 Case 1- observation beyond the maximum and minimum values of RCMs.....	78
Fig. 4-8 Case 2 - observation within the maximum and minimum values of RCMs.....	79
Fig. 4-9 Histogram of 12 stations (the blue dotted line represents the range of training data)	81
Fig. 4-10 RMSE to SD ratios of machine ensembles and statistical ensembles for 12 stations	83

List of Symbols

ANN	Artificial Neural Network
CCCma	Canadian Centre for Climate Modeling and Analysis
CanESM	Canadian Earth System Model
CanRCM	Canadian regional climate model
CRCM	Canadian regional climate model
ECCC	Environment and Climate Change Canada .
EC-EARTH	European community Earth-System Model
EM	Arithmetic Ensemble Mean
GCM	General Circulation Model
LSTM	Long Short-Term Memory Networks
MLR	Multiple Linear Regression
MLP	Multi-Layer Perceptron
MME	Multi-Model Ensemble
MPI-ESM	Max Planck Institute Earth System Model
NARX	Nonlinear Auto-Regressive Network with exogenous inputs
NA-CORDEX	North American Coordinated Regional Downscaling Experiment
R	The correlation coefficient
RBF	Radial Basis Function

RCA	Rosby Centre's Regional Atmosphere Model
RCM	Regional Climate Model
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
R^2	The coefficient of determination
SDSM	Statistical DownScaling Model
SVM	Support Vector Machine
TLFN	Time-lagged Feedforward Neural Network
Tmax	Maximum Temperature
Tmin	Minimum Temperature

1. Introduction

1.1 Background

The earth's climate has changed throughout history, and the accelerated rate of temperature change in modern history and the impacts of temperature on modern infrastructure has led to a need for climate adaptation. According to the IPCC Special Report on Global Warming of 1.5°C, observed global mean surface temperature for the decade 2006–2015 was approximately 0.87°C higher than the average over the 1850–1900 period (IPCC, 2018). The increased temperature has dramatic adverse impacts on ecosystem, agriculture, water resources, and infrastructure (Calzadilla et al., 2013; Grimm et al., 2016; Neumann et al., 2014; Wallach et al., 2016; Wheeler and Von Braun, 2013), which have led to significant cultural, economic, and environmental losses (Adger et al., 2013; Adger et al., 2012; Bouwer, 2013; Stern, 2016). Rising temperatures and climate variations have also resulted in an increase in the frequency and intensity of extreme events, such as heat waves and droughts (Rahmstorf and Coumou, 2011; Trenberth et al., 2014), which have had adverse impacts to public health and ecosystems (Naughton et al., 2002; Walther et al., 2002). To better adapt to the changing climate, there is a need for the investigation of climate patterns and generation of reliable, long term climate projections.

In the late 1960s, General Circulation Models (GCMs) were developed to simulate the global climate. A GCM is a complex mathematical representation of the major climate

system components (atmosphere, land surface, ocean, and sea ice) and their interactions (Goosse et al., 2010). Simulating the climate at a global scale is extremely computational demanding, thus, GCMs have a low resolution at both spatial and temporal scales, with a horizontal resolution of between 150-300 km and a temporal resolution ranging from hourly to monthly (Jalota et al., 2018). A number of GCMs have been developed by different institutions around the world. They are composed of different atmospheric and physical ocean components. For instance, Canadian Earth System Model (CanESM2) was developed by the Canadian Centre for Climate Modeling and Analysis (CCCma) (Fyfe et al., 2013; Hua et al., 2015). European community Earth-System Model (EC-EARTH) was developed by the European EC-Earth consortium and Max Planck Institute Earth System Model (MPI-ESM) was developed by Max-Planck-Institut für Meteorologie from Germany. As the resolution of GCMs is too coarse to provide detailed information for impact assessment and planning at a local scale, downscaling methods are utilized to bridge the gap.

Dynamical and statistical downscaling are the two main approaches for downscaling. Dynamical downscaling uses a Regional Climate Model (RCM) driven by boundary conditions from a GCM to derive finer-scale information. RCMs are capable of better representing the local landscape and possibly local atmospheric processes. RCMs run on a regional sub-domain area and generally have a high-resolution of 10 to 50 km (Leung et

al., 2003). Similar to GCMs, RCMs use physical principles to reproduce local climates. Thus, they are computationally intensive, which is the main limitation of both GCMs and RCMs. Although the resolution of RCMs is higher than that of GCMs, the grid size of RCMs is still too coarse for local climate impact studies. Thus, post-processing is required to obtain high-resolution information. For instance, statistical methods are used to further downscale RCM outputs.

1.2 Objective and Scope

The objective of this research is to investigate the performance of machine learning and statistical methods in developing multi-model ensembles for daily temperature downscaling. This entails the following tasks:

- (1) collecting North American Coordinated Regional Downscaling Experiment (NA-CORDEX) data to provide RCM outputs to feed the proposed models;
- (2) developing three neural network models including Multi-Layer Perceptron (MLP), Time-lagged Feedforward Neural Network (TLFN) and Nonlinear Auto-Regressive Network with exogenous inputs (NARX) models for daily temperature downscaling;
- (3) evaluating the performance of neural network models using a case study of the Big Trout Lake station in Ontario, Canada;
- (4) applying a number of statistical and machine learning techniques including Long Short-

Term Memory (LSTM) networks, Support Vector Machine (SVM), arithmetic ensemble mean (EM), and Multiple Linear Regression (MLR) for developing multi-model climate ensembles;

(5) evaluating the performance of machine learning and statistical ensembles for daily temperature downscaling at 12 meteorological stations over Ontario, Canada.

1.3 Organization of the Thesis

The thesis consists of five chapters. Chapter 1 introduces the background and research objectives. Chapter 2 provides a review of temperature downscaling and ensemble modeling techniques. In Chapter 3, three neural network methods are applied in temperature downscaling as a preliminary study. The results of this preliminary study show that neural networks perform well for temperature downscaling, which provides a foundation for applying other machine learning methods to generate reliable temperature downscaling. In Chapter 4, machine learning and statistical methods are investigated and compared to develop multi-model ensembles for temperature downscaling. In Chapter 5, the main conclusions are summarized and the potential for future research is discussed.

2. Literature Review

This chapter reviews several widely used RCMs, statistical downscaling methods, multi-model ensemble techniques and downscaling studies in Ontario. In particular, various statistical methods and machine learning methods applied for downscaling are introduced.

2.1 Regional Climate Models

Various RCMs have been developed to simulate the local scale climate. CanRCM4 and CRCM5 are two Canadian regional climate models developed by the Canadian Centre for Climate Modelling and Analysis (Ben Alaya et al., 2019). CanRCM4 was developed by employing a novel philosophy of coordinating RCMs and GCMs (Scinocca et al., 2016). It has a closer association with its parent GCM, the fourth generation of the Canadian Atmospheric Global Climate Model (von Salzen et al., 2013). Another Canadian regional climate model (CRCM) was developed by the University of Quebec (Martynov et al., 2012) and has been widely used for climate simulation (Laprise, 2008; Plummer et al., 2006). Different from CanRCM4, CRCM5 represents an independent RCM development and application (Scinocca et al., 2016). The horizontal resolution of CRCM5 and CanRCM4 can be configured with 25km (0.22°) or 50km (0.44°) (Diro et al., 2014). Both two RCMs use the same dynamical core, while their physics packages are independent.

RCMs developed by European institutions have also been commonly used for climate

studies, such as the Rossby Centre’s Regional atmosphere model (RCA4) and HIRHAM. RCA4 was developed by the Swedish Meteorological and Hydrological Institute (Strandberg et al., 2015). It is based on HIRLAM, a numerical weather prediction model (Cavazos et al., 2019; Samuelsson et al., 2016) and is generally set up and run at a horizontal resolution of 0.44° . Another RCM, HIRHAM, was developed by the Danish Meteorological Institute (Lucas-Picher et al., 2012). It is based on the dynamics of HIRLAM (Eerola, 2006; Undén et al., 2002) and the physical parameterization schemes of ECHAM, a GCM (Roeckner et al., 2003). Given that RCMs are computationally demanding, statistical downscaling is used to downscale GCMs and RCMs to finer resolutions.

2.2 Statistical Downscaling

2.2.1 Multiple Linear Regression

Statistical downscaling refers to using statistical methods to downscale climate data from GCMs or RCMs. As it is based on the assumption that statistical relationship between large- and small-scale climate variables are stationary, the prediction performance of statistical downscaling under changing climate condition would be affected, which leads to its main limitation. However, given that statistical downscaling is a data-driven modeling process, it is straightforward to implement and is typically computationally less intensive.

Linear regression is one of the most commonly used techniques for statistical downscaling. Predictors and predictands are fitted with linear regression and the method has been widely applied in temperature downscaling (Mahmood and Babel, 2012). Schoof and Pryor (2001) applied the least-squares Multiple Linear Regression (MLR) for downscaling daily maximum temperature (T_{max}) and minimum temperature (T_{min}) with synoptic indices set as predictors. Huth (2004) applied an MLR model with gridded data and another MLR model of principal components to predict the change in daily temperature using 500- and 1000-hPa heights, as well as 850-hPa temperature. Jeong et al. (2012) developed a multivariate multi-site statistical downscaling model based on MLR, which performed well in downscaling daily T_{max} and T_{min} from a 400-km resolution to station scale. Khalili et al. (2013) proposed an approach based on MLR and the spatial moving average process for downscaling extreme temperatures. The method could accurately describe T_{max} and T_{min} characteristics compared to CRCM. MLR has the simplest model structure and is the least computationally demanding downscaling technique. However, the drawback of MLR is that the assumption that relationship between the predictors and predictands is linear, which is not valid for most downscaling applications (Casson and Farmer, 2014).

2.2.2 Neural Networks

In the past three decades, machine learning methods have been widely applied in

downscaling and have shown promising results. Specifically, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) are two categories of commonly used techniques. The concepts of ANNs were proposed by McCulloch and Pitts (1943), who were inspired by the human brain and developed algorithms to emulate its function (Tripathi et al., 2006). ANNs have been widely applied in many disciplines due to their capability to capture both nonlinear and linear relationships (Cannon and Whitfield, 2002; Snell et al., 2000; Tang et al., 2000). Jeong et al. (2011) compared MLR with ANN for downscaling Tmax, Tmin, and precipitation, and found that monthly MLR, which was calibrated on monthly series, performed better than annual ANN and annual MLR, which were calibrated on annual series, respectively. Goyal and Ojha (2012) compared the performance of MLR and ANN in downscaling mean monthly Tmax and Tmin from GCM outputs, and found that ANN-based models were statistically superior to MLR based models. It is still controversial whether neural networks outperform linear regression for downscaling (Gaitan et al., 2013; Huth et al., 2008; Miksovsky and Raidl, 2005).

The Multi-Layer Perceptron (MLP) is the simplest model of ANN, which consists of at least three layers of nodes: an input layer, one (or more) hidden layer(s), and an output layer. It uses a back propagation algorithm to find the optimal values of weight vectors in order to minimize the error between the MLP outputs and target values. Haylock et al. (2006) compared statistical downscaling methods including MLP and dynamical

downscaling models for downscaling seasonal indices of heavy precipitation, such as mean precipitation and precipitation intensity. The results showed that MLP performed the best at modeling the inter-annual variability of the indices.

As MLP does not consider time series effects, more advanced neural networks have been further developed based on MLP for solving time series problems, including Time-Lagged Feed-Forward Network (TLFN) and Recurrent Neural Network (RNN). The input layer of TLFN is a memory structure that contains predictors of previous time steps. Coulibaly et al. (2005) applied TLFN in predicting daily precipitation, Tmax and Tmin, and found that TLFN performed well in prediction these time series. TLFN was also found to outperform the Statistical DownScaling Model (SDSM) based on MLR (Dibike and Coulibaly, 2006). Instead of incorporating past predictor samples as input, the Nonlinear Autoregressive with exogenous inputs (NARX) network, a type of RNN, feeds the predicted past values of the exogenous series back to the neural network by a delay line (Chen et al., 1990). NARX is suitable for predicting dynamic time series because of its long-term dependencies (Wang et al., 2019b). In the past few years, NARX has been widely applied in water resource prediction. Kronenberg et al. (2013) applied NARX in predicting water balance and NARX outperformed the coupled RNN and distributed watershed model. Aribarg et al. (2017) used NARX to predict monthly discharge under the impacts of climate change. NARX has shown great performance in time series prediction; however, it has not

been thoroughly investigated for climate downscaling.

Another RNN, Long Short-Term Memory (LSTM) network, has been widely applied and proven to perform well in solving time series prediction problems. The LSTM network was introduced by Hochreiter and Schmidhuber (1997) to solve the gradient exploding and vanishing problems in RNN. LSTM is known to capture long-term dependencies and can selectively memorize past information, which provides a strong basis for time series prediction. It has been widely used in sequence learning (Sutskever et al., 2014) and translation in natural language processing (Wen et al., 2015). In recent years, it has been applied for downscaling climate projections. Tran Anh et al. (2019) proposed an LSTM model and a Feedforward Neural Network (FNN) model to downscale monthly precipitation data from five GCMs obtained from CMIP5 in Vietnam and found that the correlation between observed and predicted values of both models was good. The correlation coefficient (R) of the LSTM model was above 0.9, while those of FNN ranged between 0.60 and 0.85. Misra et al. (2017) compared the LSTM model to Deep Neural Network (DNN) and other regression methods, and found that LSTM performed the best in downscaling precipitation with CEP/NCAR predictor variables. Mouatadid et al. (2017) found that LSTM had high accuracy in downscaling daily mean temperature from reanalysis data compared to ANN and extreme learning machine. Salman et al. (2018) used single- and four-layer LSTM to predict visibility with intermediate data (temperature,

pressure, humidity and dew point) and showed that multi-layer LSTM outperformed the single-layer LSTM.

2.2.3 Support Vector Machines

SVM (Cortes and Vapnik, 1995) is another widely applied machine learning technique in downscaling. It implements the structural risk minimization principle, which attempts to minimize bounds on the generalization error (Anandhi et al., 2009). Thus, it has excellent generalization performance and can avoid getting trapped in local minima, addressing one of the weaknesses of neural networks. It can also capture both linear and nonlinear relationships between predictors and predictands by using different kernels. On the other hand, the limit choice of kernel and the selection of the kernel function parameters are the major limitations of SVM. Tripathi et al. (2006) applied SVM in downscaling precipitation and SVM was shown to be superior to MLP. Anandhi et al. (2009) downscaled monthly Tmax and Tmin using SVM and the results showed that SVM was a feasible downscaling technique. Aksornsingchai and Srinilta (2011) studied SVM with Radial Basis Function (RBF) kernel and polynomial (POL) kernel, and found that SVM-RBF was the most accurate model compared to SVM-POL and MLR in downscaling monthly average rainfall and temperature. Duhan and Pandey (2014) found SVM performed slightly better than ANN and MLR in downscaling monthly Tmax and Tmin in India. Srinivas et al. (2014)

applied SVM to obtain daily Tmax and Tmin with reanalysis data. The results showed that SVM performed well at pattern recognition and time series analysis. SVM has been used as a traditional method, but it has not been compared with advanced techniques. Now that there are many new methods such as the LSTM network and NARX suitable for time series prediction, the performance of SVM needs to be re-evaluated.

2.3 Multi-model Ensembles

As the accuracy of downscaling relies on the input predictors obtained from RCMs or GCMs, Multi-Model Ensembles (MMEs) are introduced to post-process RCMs to obtain those variables with higher accuracy. MME refers to combining predictions generated by different modeling systems (Doblas-Reyes et al., 2005). The concept originated from combining multiple subjective forecasters in weather forecasting (Sanders, 1963). It was then extended to an objective multi-model prediction system and has been shown to be superior to a single model (Clemen and Murphy, 1986). By combining multiple models using ensemble techniques, MME was established as a new approach (Harrison et al., 1995). The simplest MME is combining all ensemble members and assigning each model with the same weight (Hagedorn et al., 2005). More complex methods can also be applied to determine the weights of ensemble models (Rajagopalan et al., 2002). Krishnamurti et al. (1999) used a linear regression method to determine the weights and found the ensemble

outperformed all individual models for weather and hurricane forecasting (Krishnamurti et al., 2000). Robertson et al. (2004) developed an MME with weights of GCMs determined by an improved Bayesian optimal weighting scheme and the proposed MME outperformed the ensemble mean for seasonal prediction. Kharin and Zwiers (2002) found that ensemble mean and regression-improved ensemble mean could generate the most skillful forecasts in Tropics (30°S–30°N) and extratropics (Pacific–North America sector, 20°–80°N, 180°–45°W), respectively. Li et al. (2016) found the weighted mean and reliability ensemble averaging methods showed better skill in simulating precipitation than the ensemble mean. Over the past few years, MME has also been investigated in other areas. Wang et al. (2019a) used GCMs as ensemble members and compared eight methods for assigning weights to GCMs based on their ability to represent hydrological simulations. Based on these results, the ensemble mean was recommended for climate impact studies. The rationale of MME performing better than a single model could be explained by the fact that MME includes information from all models (Hagedorn et al., 2005). Moreover, given that the performance of climate models varies with region and season, MME would be a pragmatic approach to generate an optimum forecast for a specific area during a specific period of time (Weigel et al., 2008). Particularly, the ensemble mean is the least computationally demanding method. However, with the increased application of advanced techniques such as machine learning in developing MMEs, the performance of the ensemble mean method needs to be

reassessed and compared with those advanced techniques.

In recent years, machine learning methods have been applied in MME because of their strength in capturing nonlinear relationships among different ensemble members. Anderson and Lucas (2018) applied random forests and a multi-resolution perturbed parameter ensemble to predict high-resolution precipitation. Xu et al. (2019) applied the wavelet support vector machine (WSVM), wavelet random forest (WRF) and a traditional method (quantile mapping, QM) to downscale the North American multi-model ensemble forecasts at a local scale. The results showed that WSVM and WRF could improve the downscaling accuracy relative to QM. Wang et al. (2018) compared two machine learning methods (random forest and SVM) and two statistical methods (Bayesian model averaging and the arithmetic ensemble mean) in developing MME for reproducing monthly temperature and precipitation. Machine learning MME has shown to have higher accuracy than statistical MME. However, its performance and reliability in temperature downscaling at a higher temporal resolution (e.g. daily) have not been investigated. Specifically, advanced machine learning methods such as LSTM networks and NARX have not been applied in developing MME for temperature downscaling.

2.4 Downscaling Studies in Ontario, Canada

Over the past five decades, the province of Ontario has suffered from the effects of

climate change, including a rise in average temperatures and increase in the frequency and intensity of extreme weather events such as flooding and heatwaves. As the most populated province with the largest economy in Canada, climate change and extreme weather events can have considerable social, economic, and environmental impacts across the province (Lemmen et al., 2008). Downscaling research has been carried out to obtain local climate information for policymaking in Ontario to better adapt to climate change. Deng et al. (2017) proposed a novel method combining the Ensemble Optimal Interpolation and bias correction techniques for daily temperature and precipitation downscaling with multiple GCMs. Zhai et al. (2018) employed a stepwise clustered downscaling model to downscale multiple GCMs projections for the city of Ottawa, Ontario. The projection results all show that Ontario will experience significant warming trends over the century (Wang et al., 2015a; Wang et al., 2015b). Samouly et al. (2018) developed MME based on mean and median to simulate monthly temperature in Ontario. The mean ensemble outperformed the median ensemble as well as all individual RCMs. However, most previous downscaling studies used multiple GCMs as inputs, while RCM outputs with a higher spatial resolution have yet to be used for building MME for Ontario. Meanwhile, machine learning methods have not been investigated for downscaling climate projections of Ontario. Thus, it is necessary to investigate machine learning and statistical methods in development multi-model ensembles for climate downscaling in Ontario.

3. Prediction of Long-term Near-surface Temperature based on NA-CORDEX Output

Xinyi Li, Zhong Li, Qianqian Zhang, Pengxiao Zhou and Wendy Huang

Abstract

Temperature is one of the most important parameters in climate modeling, as it has significant impacts on various geophysical processes such as evaporation and precipitation. Applying multiple climate models for prediction generally outperforms the use of individual climate models, and neural networks perform well at capturing nonlinear relationships, which can provide more reliable temperature projections. In this study, three neural network algorithms, including Multi-layer Perceptron (MLP), Time-lagged Feed-forward Neural Network (TLFN) and Nonlinear Auto-Regressive Network with exogenous inputs (NARX), were used to develop data-driven models for predicting daily mean near-surface temperature based on North American Coordinated Regional Downscaling Experiment (NA-CORDEX) output. A case study of Big Trout Lake in Ontario, Canada was carried out to demonstrate the applications and to evaluate the performance of the proposed neural network based models. The results showed that MLP, TLFN, and NARX performed well in generating accurate daily near-surface temperature predictions with the coefficient of determination (R^2) values above 0.84. The three neural network based models had similar performance with no significant difference in terms of root mean square error

and R^2 . Neural network based climate prediction models outperformed each of the individual regional climate models and generated smoother predictions with less fluctuation. This study provides a technical basis for generating reliable predictions of daily temperature using neural networks based model.

Keywords: Neural Networks, Temperature Prediction, Regional Climate Model, NA-CORDEX, Ontario

3.1 Introduction

Temperature changes have significant impacts on natural processes and human activities (Karl et al., 2009), for instance, biological changes (Parmesan and Yohe, 2003) and construction sensibility (Xia et al., 2012). Thus, predicting temperature precisely is of vital importance. Multiple climate models, such as Global Climate Model (GCM) and Regional Climate Model (RCM), have been developed and can be applied to temperature simulations and predictions, which provide support for climate impact analysis (Li et al., 2016; Thomson et al., 2006; Wagner et al., 2017). These models were developed by different institutions and their temperature predictions are not always consistent with one another. Although these models have errors in certain processes (e.g., cloud formation), they can provide plausible estimations for future variations in climate (Huo and Li, 2012; Ragone et al., 2015).

Applying dynamic downscaling to drive RCM is computationally costly and time-consuming (Spak et al., 2007). Moreover, the uncertainties in the modeling system lead to an increase in forecast errors with increasing forecast length (Kumar et al., 2012). Using statistical methods to post-process multiple RCMs would help to better generate predictions with higher accuracy than an individual RCM (Barfus and Bernhofer, 2014; Palmer et al., 2005). For example, Samouly et al. (2018) used mean and median values of multi-model ensembles for monthly temperature predictions, which showed better prediction performance than using a single RCM. However, as each model generates a different range of predictions and errors, the mean value calculated by allocating the same weight to each RCM may not be enough to fully take advantage of each prediction model.

Artificial neural networks (ANNs), which are more powerful than the regression-based techniques, have been widely applied in climate prediction because of their high potential for complex, nonlinear and time-varying input-output mapping (Von Storch et al., 2000). For instance, ANNs have been widely applied in statistical downscaling for temperature and precipitation prediction (Wilby and Wigley, 1997; Wilby et al., 1998). Previous studies suggest that using computer-based learning algorithms, such as ANNs, to develop accurate prediction models can profoundly reduce the long-term dependency (Caswell, 2014; Sfetsos, 2000; Shen and Chang, 2013; Siegelmann, 1997). Moreover, current and future temperatures have a close connection with the temperatures of previous days. Incorporating

both concurrent and antecedent predictor values as input could improve the accuracy of temperature prediction (Coulibaly et al., 2005). Various types of neural networks have an internal memory structures that can store information about past variables. Time-lagged feed-forward networks (TLFNs) and recurrent neural networks (RNNs) are the two major groups of dynamic neural networks that are commonly used in time series analysis (Coulibaly et al., 2001a; Dibike and Coulibaly, 2006). A TLFN simply replaces the neurons in the input layer of a Multi-layer perception (MLP) with a memory structure. It is less complex than the RNNs and has similar capability for processing temporal patterns (Dibike et al., 1999). TLFN is an efficient method for downscaling both daily precipitation as well as daily maximum and minimum temperature series (Coulibaly et al., 2005). The Nonlinear Auto-Regressive Networks with exogenous inputs (NARX) model is a dynamic network that has been widely used for time series prediction (Dhussa et al., 2014). It can learn the behavior of a system in an effective way. It also converges much faster and generalizes better than other networks (Çoruh et al., 2014; Lin et al., 1996). It has been demonstrated that NARX is capable of capturing the dynamics of nonlinear complex systems (Chan et al., 2015; Diaconescu, 2008). Moreover, NARX performs favorably on long-term dependencies (Rahimi et al., 2018). Thus, NARX is particularly useful for time series modeling.

Considering that neural networks perform well at grasping the nonlinear relationships

between predictors and predictands, MLP, TLFN, and NARX models will be applied to simulate daily mean near-surface temperature and generate predictions basing on multiple RCMs. The goal of this study is to develop, validate and evaluate the performance of neural networks for daily mean near-surface temperature prediction with multiple RCMs in the province of Ontario, Canada. This entails the following: (1) collecting North American Coordinated Regional Downscaling Experiment (NA-CORDEX) data to provide inputs for the proposed neural network based models; (2) developing MLP, TLFN, and NARX models to generate daily mean near-surface temperature; (3) evaluating the performance of MLP, TLFN and NARX using a case study of the Big Trout Lake station in Ontario, Canada.

3.2 Methodology

3.2.1 Multi-layer Perceptron (MLP)

MLP is a widely-used ANN model which usually consists of an input layer, one or more hidden layers, and an output layer (Fig. 3-1). Each layer includes some neurons (Jiang et al., 2018). The numbers of neurons in the input and output layers are determined by the numbers of elements in the external input array and output array of the network, respectively (Osman and Abdellatif, 2016). The number of neurons in the hidden layers are determined by the trial and error (Hammerstorm, 1993) for the best performing model. Different layers are connected with weights and biases. The connections between the layers

allow information flow forward towards the output layer. The neuron network first computes the weighted sum of the inputs, z , and feeds z into the neurons in the hidden layer (Eq. 3.1). A nonlinear activation function $f(\cdot)$, is applied to z to get the output a of the neuron (Eq. 3.2). The network repeats the same process to the hidden layer (Eqs. 3.3 & 3.4). Rectified linear unit (ReLU), conventional sigmoids function, hyperbolic tangent function, and logistic function are examples of commonly used activation functions. The ultimate goal of training a MLP is to minimize the cost function (Eq. 3.5), which measures the errors between observations and predictions for training data. A back-propagation algorithm is used to find the minimum cost function using the chain rule of differentiation to calculate the partial derivative or gradient of the cost corresponding to the weights (Zhang et al., 2018). Back-propagation calculates the error-derivative for the weight of each neuron to minimize the cost function.

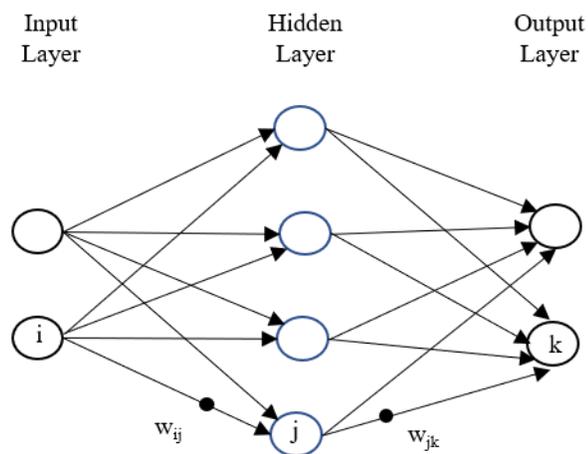


Fig. 3-1 Structure of MLP with one input layer, one hidden layer, and one output layer.

$$z_j = \sum_i^n w_{ij}x_i + b_j \quad (3.1)$$

$$a_j = f(z_j) \quad (3.2)$$

$$z_k = \sum_j^m w_{jk}a_j + b_k \quad (3.3)$$

$$y_k = f(z_k) \quad (3.4)$$

$$C_k = \frac{1}{2}(y_k - t_k)^2 \quad (3.5)$$

$$\frac{\partial C}{\partial z_k} = \frac{\partial C}{\partial y_k} \times \frac{\partial y_k}{\partial z_k} \quad (3.6)$$

$$\frac{\partial C}{\partial y_k} = y_k - t_k \quad (3.7)$$

where x_i is the i^{th} input, a_j is the output of the j^{th} neuron; w_{ij} and w_{jk} represent the weight of j^{th} neuron in the hidden layer and k^{th} neuron in the output layer, respectively; and b is the bias. C refers to the cost of the cost function, y_k is the predicted output and t_k is the observed true value. The error-derivative for the weight w_{jk} on the connection from unit k is $a_j (\partial C) / (\partial z_k)$. The error-derivative for the weight w_{ij} on the connection from unit j is $x_i (\partial C) / (\partial z_j)$ (Eq. 3.6). (LeCun et al., 2015). Eq. 3.7 shows the partial derivative of the cost function corresponding and activation function.

3.2.2 Time-lagged Feed-forward Neural Network (TLFN)

TLFN is formulated based on MLP and replace the neurons in the input layer with a

memory structure, which is sometimes called a tap delay-line, as shown in Fig. 3-2 (Coulibaly et al., 2005). TLFN uses delay-line processing elements (PEs) by holding past samples of the input signal. The output $y(n)$ of TLFN with one hidden layer is shown as Eq. 3.8 (Coulibaly, 2004).

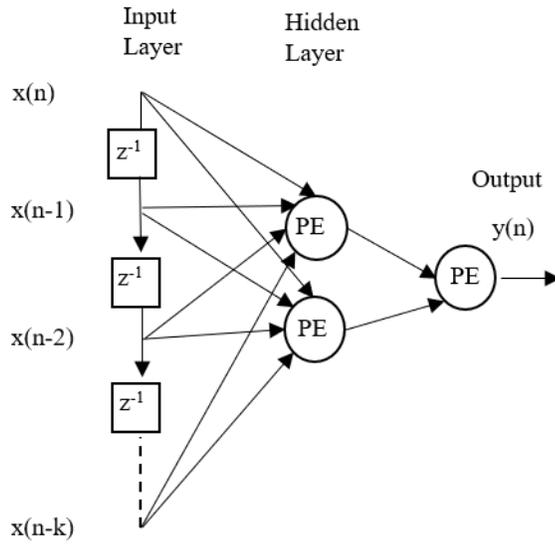


Fig. 3-2 Structure of TLFN with one input layer, one hidden layer, and a delay-line with memory depth of k . (z^{-1} is an operator that delays the input by one sample). (Dibike and Coulibaly, 2006).

$$y(n) = f_1\left(\sum_{j=1}^m w_j y_j(n) + b_o\right) = f_1\left\{\sum_{j=1}^m w_j f_2\left[\sum_{l=0}^k w_{jl} x(n-l) + b_j\right] + b_o\right\} \quad (3.8)$$

Where m is the size of the hidden layer, n is the time step, w_j is the weight vector for the connection between the output layer and the hidden layer, and w_{jl} is the weight matrix for the connection between the hidden layer and the input layer. f_1 and f_2 are the active functions at the output layer and hidden layer, respectively. b_o and b_j are the bias terms.

The input pattern $x(n)$ has multiple inputs of size p (Eq. 3.9) and $X(n)$ is the combined input at time step n , whose delay line with memory depth k (Eq. 3.10). $x(n-1)$ is obtained by delaying $x(n)$ by one sample.

$$x(n) = (x_1(n), x_2(n), \dots, x_p(n)) \quad (3.9)$$

$$X(n) = [x(n), x(n-1), \dots, x(n-k+1)] \quad (3.10)$$

3.2.3 Nonlinear Auto-Regressive Networks with Exogenous Inputs (NARX)

Networks that use feedback connections, enabling information flow laterally or backwards within the network, are called RNNs. NARX is a special type of RNN that creates a relationship between the current value of a time series and predicted past values of the exogenous series, and the outputs are fed back to the input by a delay line (Haykin, 1998). As shown in Fig.3-3, the structure of the NARX model is similar to the traditional multi-layered perceptron (MLP) model. The NARX model can be expressed as in Eq. 3.11 (Lin et al., 1996).

$$y_t = f(y_{t-1}, \dots, y_{t-dy}; x_t, x_{t-1}, \dots, x_{t-dx}) \quad (3.11)$$

where x_t and y_t represent the input and output of the network at time t , respectively. $f(\cdot)$ is a nonlinear function, which can be approximated by a standard MLP network. dx and dy are the time lags for the input and output series (Lin et al., 1998).

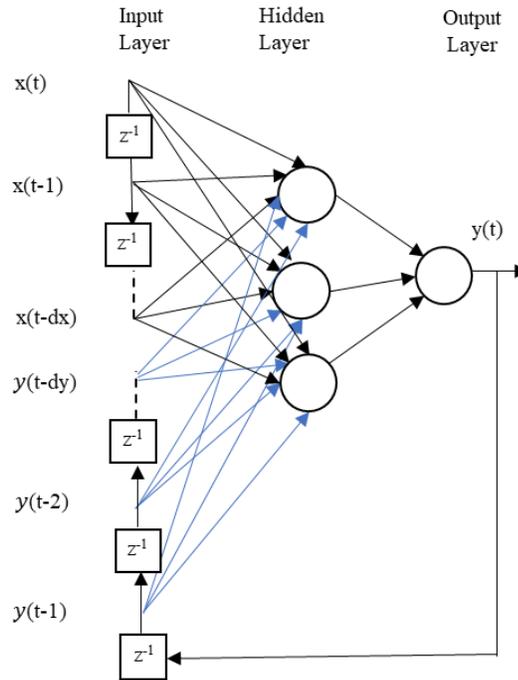


Fig. 3-3 Structure of NARX with one input layer, one hidden layer, and one output layer (z^{-1} denotes delay for one time step).

3.3 Study Area and Data Collection

Big Trout Lake in Northern Ontario, Canada was chosen to test the performance of the proposed methods. According to Canada’s Changing Climate Report 2019 (Bush et al., 2019), Northern Canada has warmed and will continue to warm at even more than double the global rate. Between 1948 and 2016, the observed changes ($^{\circ}\text{C}$) in annual temperature in Northern Ontario were higher than in Southern. The Big Trout Lake station (53.83°N , 89.87°W) is located in the far northwestern region of Ontario and south of Hudson Bay. It is classified as having a subarctic climate, which includes year-round precipitation, short

and cool summers, and long and cold dry winters (Tam et al., 2018), resulting in high annual variation in temperature. The average temperature and yearly precipitation of Big Trout Lake are $-2.7\text{ }^{\circ}\text{C}$ and 609.1 mm. The average monthly temperature ranges from -23.7 to $16.2\text{ }^{\circ}\text{C}$. The minimum and maximum recorded temperature of the Big Trout Lake station were $-47.8\text{ }^{\circ}\text{C}$ (January 1951) and $35.6\text{ }^{\circ}\text{C}$ (July 1955). The highest historical daily precipitation occurred in August 1955, reaching 84.1 mm. Considering the region's high variation in temperature and climate sensitivity, the Big Trout Lake station was chosen for evaluating the performance of neural networks methods.

The study used daily mean temperature simulation data obtained from six RCMs and observation data of the Big Trout Lake station from 1979 to 1989. The six RCMs are each driven by different GCM models. They are 1) CanRCM4, CRCM5, and RCA4 driven by CanESM2; 2) HIRHAM5 and RCA4 driven by EC-EARTH. 3) CRCM5 driven by MPI-ESM-LR. The grid resolution for each RCM is $0.44^{\circ} \times 0.44^{\circ}$. The simulated daily mean temperature data were downloaded from NA-CORDEX archive (Mearns et al., 2017), a branch of the International CORDEX Initiative (Giorgi, 2009; Lucas-Picher et al., 2012). The observed temperature data of the Big Trout Lake were downloaded from the Digital Archive of Canadian Climatological Data provided by Environment and Climate Change Canada (ECCC).

3.4 Neural Network Design and Training

The neural network models in this study were developed with net functions in MATLAB (version R2014b). The Levenberg-Marquardt backpropagation algorithm was applied for training the models, as it is one of the fastest backpropagation algorithms for feedforward networks (Hagan and Menhaj, 1994; Lee et al., 2016).

Inputs to the neural networks were the simulated daily mean temperature of six RCMs while the output was daily mean near-surface temperature observed at the Big Trout Lake station. RCM outputs at the closest grid point to the Big Trout Lake station (53.76°N, 89.84°W) were used as inputs for the prediction models. The first 70% of the dataset (January 1979 - September 1986) were used for training the models. Then, the following 15% of the dataset (September 1986 – May 1988) were used to validate those models, which verified the applicability of the model. The last 15% of the dataset (May 1988 – December 1989) were used for testing, which assessed the generalization ability of the model. The different parameters of each model were adjusted during calibration to obtain the best statistical agreement between observed and simulated mean temperature and were assessed using mean square error (MSE).

The structure of the networks used in the study consisted of one input layer, one output layer, and one hidden layer. MLP was trained with the number of neurons ranging from 5 to 20 and the MLP with 12 neurons was selected as it generated the best performing network.

Both TLFN and NARX were trained with lag time (time delay) ranging from 1 to 3 days and the number of neurons ranging from 5 to 20. The TLFN model with 5 neurons and a time lag of 3 days and NARX with 15 neurons and a time lag of 3 days were selected as they generated the best performing network.

Performance of three neural network models was evaluated by comparing predicted results with observed temperature values. Statistical criteria, such as root mean square error (RMSE) and coefficient of determination (R^2), were used for performance evaluation.

3.5 Results and Discussion

3.5.1 Neural Networks Performance

3.5.1.1 MLP Performance

The time-series plot and the scatter plot of the observation and prediction of daily mean temperature obtained by MLP are shown in Fig. 3-4 and 3-5, respectively. The time-series plot shows that MLP could predict the seasonal pattern of daily mean near-surface temperature. RMSE and R^2 of testing were 6.537 °C and 0.843, respectively. The small RMSE and high R^2 values indicates that MLP performed well and could predict mean near-surface temperature with relatively high accuracy.

In addition, while the observed temperature of all datasets varied from -38.9 to 26.1 °C, MLP could generate predictions ranging from -25.6 to 17.5 °C. For observations ranging

from -30 to -16°C , MLP tended to give prediction values of around -20°C . For observations ranging from 12 to 26.1°C , MLP generated prediction ranging from 10 to 17°C , which implies that MLP could not capture the extreme values precisely. This may be due to the tendency of neural networks sacrificing variance to gain high RMSE.

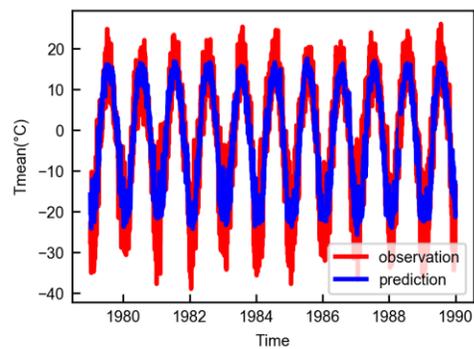


Fig. 3-4 Time series plot of observed and predicted daily near-surface temperature values obtained by MLP with 12 neurons.

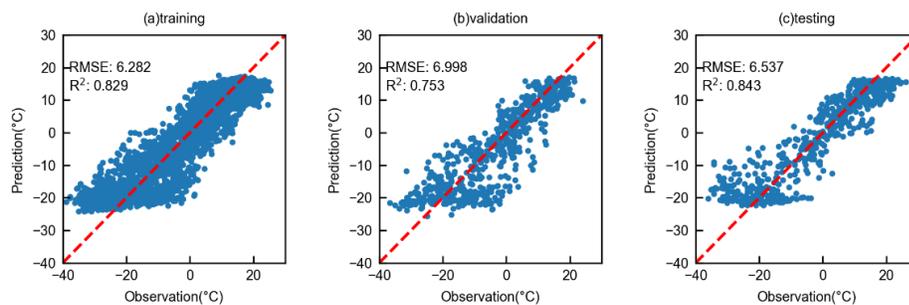


Fig. 3-5 Scatter plots of observed and predicted daily near-surface temperature values obtained by MLP with 12 neurons.

3.5.1.2 TLFN Performance

Fig. 3-6 and 3-7 show the statistical performance and time series plot of TLFN with a time lag of 3 days and 5 neurons. TLFN had similar performance with MLP and had small improvement, with RMSE decreased to 6.363 °C and R^2 increased to 0.854. This indicates that TLFN is an efficient model for capturing the changing pattern and predicting daily mean near-surface temperature. Compared with the MLP model, TLFN generated predictions scattered more closely with observations and had a smaller range of temperature prediction from -23.9 to 16.5°C. For observations ranging from -30 to -20°C, the overestimated prediction errors of TLFN were smaller than MLP. Similar to MLP, TLFN did not capture the extreme values well.

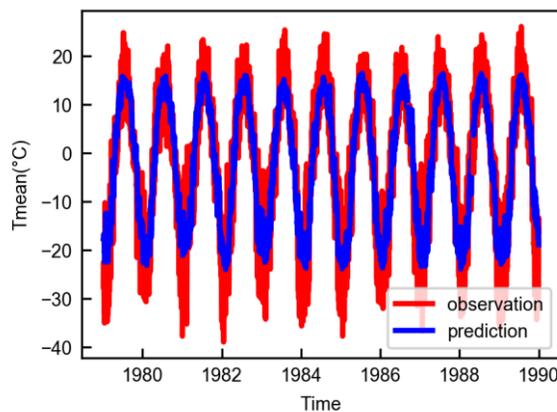


Fig. 3-6 Time series plot of observed and predicted daily near-surface temperature values obtained by TLFN with 5 neurons and a time lag of 3 days.

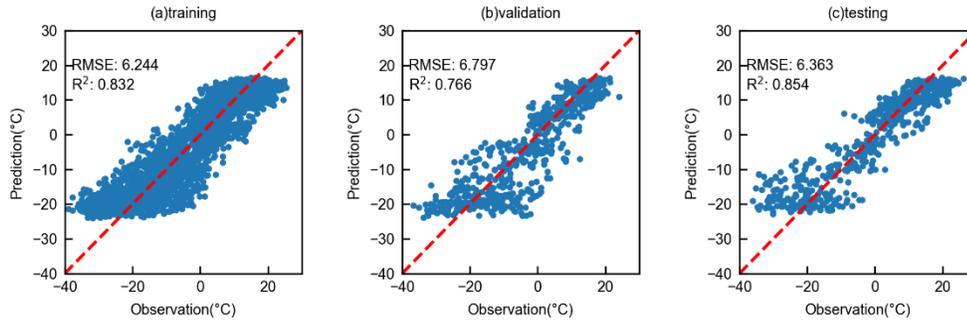


Fig. 3-7 Scatter plots of observed and predicted daily near-surface temperature values obtained by TLFN with 5 neurons and a time lag of 3 days.

3.5.1.3 NARX Performance

Fig. 3-8 and 3-9 show that NARX performs well at generating mean temperature prediction and could accurately predict the changes of daily mean near-surface temperature with a low RMSE of 6.345 °C and high R² of 0.856. This suggests that NARX performs the most effectively among the three in predicting daily mean near-surface temperature. The prediction range generated by NARX was from -25.4 to 15.7°C; the maximum value was lower than the predictions generated by MLP and TLFN. From the scatter plot of observed and predicted temperatures shown in Fig. 3-9, the points are scattered more densely along the diagonal line than MLP and TLFN, indicating that the error of prediction and observation values were smaller than that of MLP and TLFN. However, the accuracy of NARX for prediction extreme temperature values was similar to MLP and TLFN.

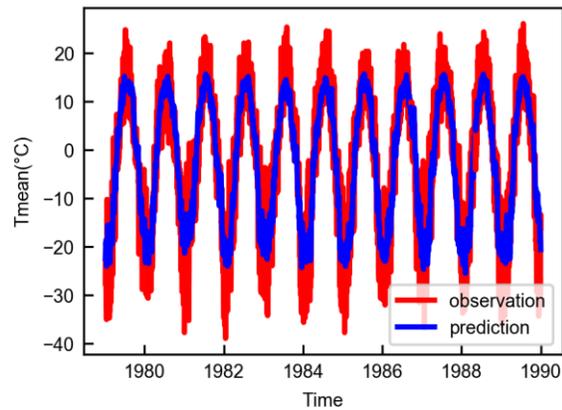


Fig. 3-8 Time series plot of observed and predicted daily near-surface temperature values obtained by NARX with 15 neurons and a time lag of 3 days.

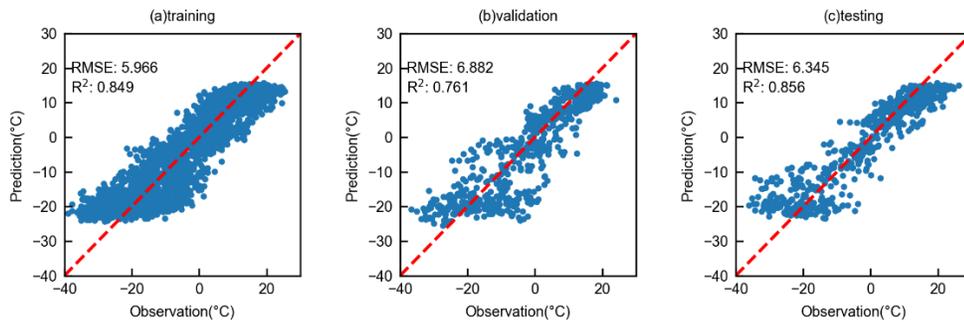


Fig. 3-9 Scatter plot of observed and predicted daily near-surface temperature values obtained by NARX with 15 neurons and a time lag of 3 days.

3.5.2 Comparison between RCMs and Neural Networks

Table 3-1 shows the statistical performance of six RCMs and three neural network models for training, validation, and testing. The RMSE of six models ranged from 8.253°C

to 9.239°C and the R^2 ranged from 0.691 to 0.732. Among the six models, CRCM5 derived by MPI_ESM_LR performed the best while CanRCM4 derived by CanESM2 had the lowest R^2 and the highest RMSE. Neural network based models outperformed each individual RCM, with RMSE decreased by approximately 2 °C and R^2 increased from 0.7 to 0.85.

Table 3-1 Comparison of performance between RCMs and neural networks

GCM	RCM	RMSE(°C)			R^2		
		training	validation	testing	training	validation	testing
	CRCM5	8.682	8.800	8.861	0.717	0.638	0.728
CanESM2	CanRCM4	9.402	9.124	9.239	0.665	0.617	0.691
	RCA4	9.029	8.709	9.008	0.690	0.653	0.708
EC-EARTH	HIRHAM5	8.349	8.411	8.547	0.695	0.659	0.722
	RCA4	9.304	10.475	8.676	0.667	0.565	0.707
MPI-ESM-LR	CRCM5	8.321	9.406	8.253	0.705	0.612	0.732
Neural	MLP	6.282	6.998	6.537	0.829	0.753	0.843
Networks	TLFN	6.244	6.797	6.363	0.832	0.766	0.854
	NARX	5.966	6.882	6.345	0.849	0.761	0.856

Fig. 3-10 shows the time series plot of six RCMs and neural networks of winter (January and February) and summer (July and August) 1989 predictions. The observation values fall within the range of the RCMs, while the neural networks tended to predict temperatures of -20°C and 15°C for winter and summer, respectively, with very little variance and fluctuation. For winter, three RCMs driven by CanESM2 predicted relatively well with observations aligning closely to the RCMs predicted values. RCA4 driven by EC-EARTH performed the worst which tended to overestimate the low temperature and underestimated the high temperature in winter. CRCM5 driven by MPI-ESM-LR performed well with low error in winter. All RCMs performed better in the summer months than in the winter months, with observation values falling between the maximum and minimum prediction of 6 RCMs. However, the changing pattern of neural network based models was smoother than RCMs with smaller fluctuation. As neural network based models generate predictions with smooth variation pattern, they tend to have smaller RMSE and higher R^2 than RCMs.

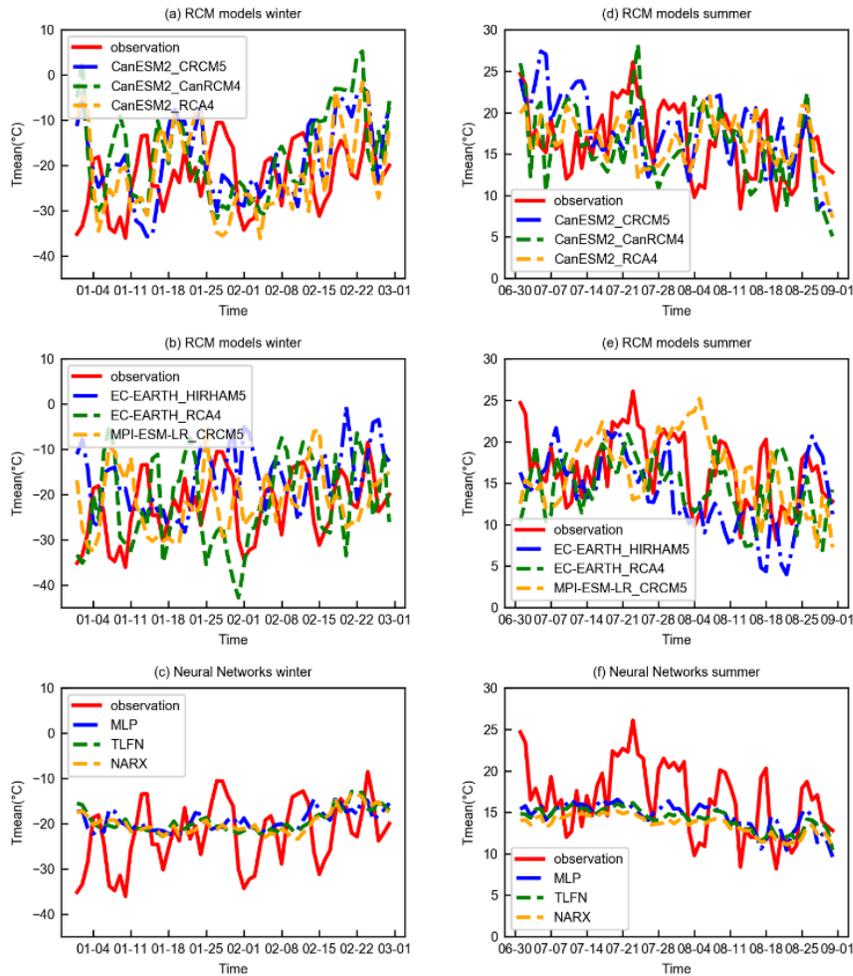


Fig. 3-10 Time series plot of 6 RCMs and neural networks of (a - c) winter and (d - f) summer in 1989.

3.5.3 Comparison of Performance between MLP, TLFN, and NARX

The RMSE for all three methods ranged from 6.345 °C to 6.537 °C and R^2 were above 0.84, indicating that the three neural networks could provide reliable temperature forecasts for Big Trout Lake. The nonlinear transfer function associated with each hidden and output node allows ANNs to approximate highly nonlinear relationships without a prior

assumption, which leads to relatively high accuracy in prediction.

In terms of the structure of neural networks, TLFN and NARX were built based on the structure of MLP. These two methods incorporate antecedent predictor values as input to improve forecasting. Although all three methods had similar performance with no significant differences in terms of RMSE and R^2 , TLFN and NARX had a smaller error in prediction than MLP. Thus, incorporating antecedent predictor values as input would slightly improve the performance of the neural network. When compared with TLFN, NARX not only incorporates previous RCM data into the network but also considers previously predicted values. However, the time required to train the NARX model and generate predictions was much longer than that of TLFN. As TLFN has similar capability to process and predict temporal patterns as RNN while having a less complex structure being less computationally demanding, TLFN is recommended for the prediction of temperature values in areas where the climate is similar to the study area.. This finding is consistent with the results from previous studies on using neural networks for temperature predictions (Coulibaly et al., 2001b; Coulibaly et al., 2005).

3.6 Conclusions

The study investigated the applicability of three neural networks (MLP, TLFN, and NARX) for daily mean near-surface temperature prediction using NA-CORDEX

simulation output. A case study of Big Trout Lake in Ontario, Canada was carried out to demonstrate the applicability and performance of the three models. Daily mean temperatures simulated by six RCMs from 1979 to 1989 were applied for training, validation, and testing. The temperature values predicted by MLP, TLFN, and NARX were compared with the observations from the Big Trout Lake monitoring station. The performance of neural network models was compared with six individual RCMs.

The results show that MLP, TLFN, and NARX are effective methods for predicting daily mean temperature. Based on the RMSE and R^2 , all three methods had similar performance, with RMSE ranged from 6.345°C to 6.537°C and R^2 above 0.84. It is worth mentioning that the differences in prediction performance among these three models were not significant in terms of RMSE and R^2 . Neural-network based temperature prediction models outperformed individual RCMs, with RMSE decreased by about 2°C and R^2 increased from 0.7 to 0.85. Neural network models generated smoother predictions with less fluctuation than RCMs. It was also found that MLP, TLFN, and NARX could not capture ‘extreme’ values below -20°C accurately. Those values appeared during a similar time period each year (i.e., winter). Thus, further work could be done to develop prediction models for a seasonal time period which have different temperature range.

3.7 References

- Barfus, K., and Bernhofer, C. (2014). Assessment of GCM performances for the Arabian Peninsula, Brazil, and Ukraine and indications of regional climate change. *Environmental Earth Sciences*, 72(12), 4689-4703.
- Bush, E., Lemmen, D. S., and editors. (2019). Canada's Changing Climate Report. *Government of Canada, Ottawa, ON, 444 p.*
- Caswell, J. M. (2014). A Nonlinear Autoregressive Approach to Statistical Prediction of Disturbance Storm Time Geomagnetic Fluctuations Using Solar Data. *Journal of Signal and Information Processing*, 05(02), 42-53.
- Chan, R. W., Yuen, J. K., Lee, E. W., and Arashpour, M. (2015). Application of Nonlinear-Autoregressive-Exogenous model to predict the hysteretic behaviour of passive control systems. *Engineering Structures*, 85, 1-10.
- Çoruh, S., Geyikçi, F., Kılıç, E., and Çoruh, U. (2014). The use of NARX neural network for modeling of adsorption of zinc ions using activated almond shell as a potential biosorbent. *Bioresource Technology*, 151, 406-410.
- Coulibaly, P. (2004). Downscaling daily extreme temperatures with genetic programming. *Geophysical Research Letters*, 31(16).
- Coulibaly, P., Anctil, F., Aravena, R., and Bobée, B. (2001a). Artificial neural network modeling of water table depth fluctuations. *Water Resources Research*, 37(4), 885-

896.

Coulibaly, P., Anctil, F., and Bobee, B. (2001b). Multivariate reservoir inflow forecasting using temporal neural networks. *Journal of Hydrologic Engineering*, 6(5), 367-376.

Coulibaly, P., Dibike, Y. B., and Anctil, F. (2005). Downscaling Precipitation and Temperature with Temporal Neural Networks. *Journal of Hydrometeorology*, 6(4), 483-496.

Dhussa, A. K., Sambhi, S. S., Kumar, S., Kumar, S., and Kumar, S. (2014). Nonlinear Autoregressive Exogenous modeling of a large anaerobic digester producing biogas from cattle waste. *Bioresource Technology*, 170, 342-349.

Diaconescu, E. (2008). The use of NARX neural networks to predict chaotic time series. *Wseas Transactions on computer research*, 3(3), 182-191.

Dibike, Y. B., and Coulibaly, P. (2006). Temporal neural networks for downscaling climate variability and extremes. *Neural Networks*, 19(2), 135-144.

Dibike, Y. B., Solomatine, D., and Abbott, M. B. (1999). On the encapsulation of numerical-hydraulic models in artificial neural network. *Journal of Hydraulic Research*, 37(2), 147-161.

ECCC. Digital Archive of Canadian Climatological Data. Retrieved from http://climate.weather.gc.ca/historical_data/search_historic_data_e.html

Giorgi, F., Jones, C., & Asrar, G. R. . (2009). Addressing climate information needs at the

- regional level: the CORDEX framework. *World Meteorological Organization (WMO) Bulletin*, 58(3), 175.
- Hagan, M. T., and Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6), 989-993.
- Hammerstorm, D. (1993). Working with neural networks. *IEEE Spectrum.*, 46–53.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*: Prentice Hall PTR.
- Huo, A., and Li, H. (2012). Assessment of climate change impact on the stream-flow in a typical debris flow watershed of Jianzhuangcuan catchment in Shaanxi Province, China. *Environmental Earth Sciences*, 69(6), 1931-1938.
- Jiang, W., He, G., Long, T., Ni, Y., Liu, H., Peng, Y., Lv, K., and Wang, G. (2018). Multilayer Perceptron Neural Network for Surface Water Extraction in Landsat 8 OLI Satellite Images. *Remote Sensing*, 10(5), 755.
- Karl, T. R., Melillo, J. M., and Peterson, T. C. (2009). Global Climate Change Impacts in the United States. *Cambridge University Press*.
- Kumar, A., Mitra, A., Bohra, A., Iyengar, G., and Durai, V. J. M. A. (2012). Multi-model ensemble (MME) prediction of rainfall using neural networks during monsoon season in India. *19(2)*, 161-169.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lee, C. C., Sheridan, S. C., Barnes, B. B., Hu, C., Pirhalla, D. E., Ransibrahmanakul, V.,

and Shein, K. (2016). The development of a non-linear autoregressive model with exogenous input (NARX) to model climate-water clarity relationships: reconstructing a historical water clarity index for the coastal waters of the southeastern USA. *Theoretical and Applied Climatology*, 130(1-2), 557-569.

Li, Z., Huang, G., Wang, X., Han, J., and Fan, Y. (2016). Impacts of future climate change on river discharge based on hydrological inference: A case study of the Grand River Watershed in Ontario, Canada. *Science of the Total Environment*, 548, 198-210.

Lin, T., Horne, B. G., and Giles, C. L. (1998). How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. *Neural Networks*, 11(5), 861-868.

Lin, T., Horne, B. G., Tino, P., and Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks*, 7(6), 1329-1338.

Lucas-Picher, P., Somot, S., Déqué, M., Decharme, B., and Alias, A. (2012). Evaluation of the regional climate model ALADIN to simulate the climate over North America in the CORDEX framework. *Climate Dynamics*, 41(5-6), 1117-1137.

Mearns, L., McGinnis, S., Korytina, D., and Arritt, R. (2017). The NA-CORDEX dataset, version 1.0. NCAR Climate Data Gateway, Boulder CO.

Osman, Y. Z., and Abdellatif, M. E. (2016). Improving accuracy of downscaling rainfall by

combining predictions of different statistical downscale models. *Water Science*, 30(2), 61-75.

Palmer, T. N., Doblus-Reyes, F. J., Hagedorn, R., and Weisheimer, A. (2005). Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philos Trans R Soc Lond B Biol Sci*, 360(1463), 1991-1998.

Parmesan, C., and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918), 37-42.

Ragone, F., Lucarini, V., and Lunkeit, F. (2015). A new framework for climate sensitivity and prediction: a modelling perspective. *Climate Dynamics*, 46(5-6), 1459-1471.

Rahimi, Z., Shafri, H., and Norman, M. (2018). A GNSS-based weather forecasting approach using Nonlinear Auto Regressive Approach with Exogenous Input (NARX). *Journal of Atmospheric and Solar-Terrestrial Physics*, 178, 74-84.

Samouly, A. A., Luong, C. N., Li, Z., Smith, S., Baetz, B., and Ghaith, M. (2018). Performance of multi-model ensembles for the simulation of temperature variability over Ontario, Canada. *Environmental Earth Sciences*, 77(13).

Sfetsos, A., & Coonick, A. H. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, 68(2), 169-178.

Shen, H. Y., and Chang, L. C. (2013). Online multistep-ahead inundation depth forecasts by recurrent NARX networks. *Hydrology and Earth System Sciences*, 17(3), 935-

945.

Siegelmann, H. T., Horne, B. G., & Giles, C. L. (1997). Computational capabilities of recurrent NARX neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 27(2), 208-215.

Spak, S., Holloway, T., Lynn, B., and Goldberg, R. (2007). A comparison of statistical and dynamical downscaling for surface temperature in North America. *Journal of Geophysical Research*, 112(D8).

Tam, A., Gough, W. A., Kowal, S., and Xie, C. (2018). The Fate of Hudson Bay Lowlands Palsas in a Changing Climate. *Arctic, Antarctic, and Alpine Research*, 46(1), 114-120.

Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A. P., and Palmer, T. N. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, 439(7076), 576-579.

Von Storch, H., Hewitson, B., and Mearns, L. (2000). *Review of empirical downscaling techniques. In Regional climate development under global warming.* Paper presented at the General technical report no. 4. Conference proceedings regclim spring meeting, Norway.

Wagner, T., Themeßl, M., Schüppel, A., Gobiet, A., Stigler, H., and Birk, S. (2017). Impacts of climate change on stream flow and hydro power generation in the Alpine region.

Environmental Earth Sciences, 76(1), 4.

Wilby, R. L., and Wigley, T. M. L. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography: Earth and Environment*, 21(4), 530–548.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., and Wilks, D. S. (1998). Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research*, 34(11), 2995-3008.

Xia, Y., Chen, B., Weng, S., Ni, Y.-Q., and Xu, Y.-L. (2012). Temperature effect on vibration properties of civil structures: a literature review and case studies. *Journal of Civil Structural Health Monitoring*, 2(1), 29-46.

Zhang, D., Lindholm, G., and Ratnaweera, H. (2018). Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *Journal of Hydrology*, 556, 409-418.

4. Performance of Statistical and Machine Learning Ensembles for Daily Temperature Downscaling

Xinyi Li, Zhong Li, Wendy Huang and Pengxiao Zhou

Abstract

Temperature changes have widespread impacts on the environment, economic activity, and municipal services. Generating accurate climate prediction at finer spatial resolution through downscaling could help better assess the future effects of climate change on a local scale. Ensembles of multiple climate models have been proven to improve the accuracy of temperature prediction. Meanwhile, machine learning techniques have shown high performance in solving various predictive modeling problems which make them a promising tool for temperature downscaling. This study investigated the performance of machine learning methods (Long Short-Term Memory, LSTM, networks and Support Vector Machine, SVM) and statistical methods (arithmetic ensemble mean, EM, and Multiple Linear Regression, MLR) in developing multi-model ensembles for downscaling long-term daily temperature. A case study of twelve meteorological stations across Ontario, Canada was conducted to evaluate the performance of the proposed machine learning and statistical ensembles. The results showed that both machine learning and statistical techniques performed well at downscaling daily temperature with multi-model ensembles

and had similar performance with relatively high accuracy. The R^2 of 12 stations ranged between 0.756 and 0.820 and RMSE ranged between 4.318°C and 7.063°C. Both machine learning and statistical ensembles for downscaling had difficulty in predicting extreme values for temperature below -10°C and above 20°C. Since machine learning ensembles are computationally demanding, the mean ensemble can be used instead to downscale long-term daily temperature projections. The results provided a technical foundation for using statistical and machine learning methods to generate high-resolution daily temperature prediction.

Keywords: Machine learning, Multi-model Ensemble, Temperature Downscaling, Regional Climate Model, Ontario

4.1 Introduction

Temperature changes have widespread impacts on the environment, economic activity, and municipal services (Bartos and Chester, 2015; Neumann et al., 2014). As infrastructure and agricultural activities are sensitive to temperature thresholds (Gornall et al., 2010; Hatfield and Prueger, 2015), it is essential to generate reliable long-term daily temperature predictions. Climate models have been developed to obtain accurate results of climate predictions. For large spatial climate change, General Circulation Models (GCMs) are conducted to simulate the global climate, which has a relatively coarse resolution of typically 150–300 km (Kendon et al., 2010). For the practical planning of local issues, governments require information on a more local scale. Regional Climate Models (RCMs) were then introduced to provide small scale information by increasing the resolution of the GCM output in a small, limited area of interest with a typical horizontal resolution of 25 - 50 km. Although RCMs have a finer spatial resolution than GCMs, it is still too coarse to support local climate impact analysis (Wang et al., 2013). Hence, downscaling methods are in need to provide climate projection with finer spatial resolution and support local climate adaptation.

To improve the accuracy of downscaling, Multi-Model Ensemble (MME) is used to provide reliable climate models inputs for downscaling with higher accuracy. MME is a combination of multiple numerical models. It has the potential to provide more information

for practical forecasting which would help to decrease the uncertainty of forecasting (Kumar et al., 2012). Previous studies have shown that MME outperforms the prediction of a single numerical model, which improves the prediction quality and decreases their mean square errors (Krishnamurti et al., 2009; Kumar et al., 2012; Mitra et al., 2011). Hence, MME applied for downscaling may help to improve the downscaling accuracy as it provides more reliable input for downscaling. Because of the outstanding performance of MME, it is gaining more popularity and is widely applied in climate prediction and weather forecasting (Rozante et al., 2014). Specifically, the weights of multi-models should be well considered (Christensen et al., 2010).

In general, there are two categories of ensemble techniques: statistical and machine learning methods. Statistical methods have been widely used in MMEs. The simplest method of statistical ensembles is the mean ensemble, which entails merging multi-models with equal weights (Hagedorn et al., 2005; Jarsjö et al., 2017; Wallach et al., 2016). The mean ensemble outperforms any single model (Hagedorn et al., 2005; Li et al., 2016; Tang et al., 2016). Another commonly used MME technique is linear regression (Krishnamurti et al., 2000). It was first introduced by Krishnamurti et al. (1999a) and originally applied in a seasonal climate forecast. In recent studies, it is also applied in daily precipitation forecast (Krasnopolsky and Lin, 2012) and has shown to be an efficient ensemble method for temperature and precipitation prediction (Feng et al., 2010).

Meanwhile, machine learning techniques have gained a lot of attention because of their high performance in solving various predictive modeling problems. For example, the Support Vector Machine (SVM) has been applied to obtain temperature and precipitation prediction with high accuracy (Devak et al., 2015; Mellit et al., 2013; Pour et al., 2018). It is a highly effective model in solving nonlinear problems even with small quantities of training data (Zhao and Magoulès, 2012). SVM is a promising alternative to conventional methods for statistical downscaling (Tripathi et al., 2006a). SVM is also found to outperform Multi-Layer Perceptron (MLP) for short-term daily maximum temperature prediction (Radhika and Shashi, 2009). Artificial neural networks (ANNs) have been widely applied in climate prediction because of their high potential for complex, nonlinear and time-varying input-output mapping. This leads to ANNs being more powerful than the other regression-based techniques (Von Storch et al., 2000). ANN models have also been utilized in predicting long-range changes in climatological time series in recent decades (Tangang et al., 1998). Krasnopolsky and Lin (2012) found that neural network ensembles improved upon conservative ensemble and Multiple Linear Regression (MLR) ensemble when applied in 24h precipitation forecasts. Furthermore, the temporal resolutions of previous studies were generally low. Most models were only validated with a monthly or even yearly time step. For instance, Kisi and Sanikhani (2015a, 2015b) found that long-term monthly temperatures of any site can be successfully estimated by ANNs, Support

Vector Regression (SVR) and Gene Expression Programming (GEP) using geographical inputs. As climate prediction is a time series problem, time delay should be considered (Coulibaly et al., 2005). Thus, methods suitable for temporal sequences processing are recommended (Dibike and Coulibaly, 2006).

More recently, Long Short-Term Memory (LSTM) networks have been proven to be a powerful tool in processing long term time series data (Kratzert et al., 2018; Zhang et al., 2018). LSTM networks are designed to handle sequence dependency, and have been widely applied in machine translation (LeCun et al., 2015). Such advanced techniques have great potential in climate prediction. However, LSTM networks have not yet been applied to long-term daily temperature prediction; they have only been applied in downscaling with gridded reanalysis data and weather forecasting. Mouatadid et al. (2017) used gridded reanalysis data for downscaling and found that LSTM networks could generalize the daily mean temperatures well at different locations and have higher downscaling accuracy than MLR and Extreme Learning Machines. For weather forecasting, LSTM networks give substantial results with high accuracy (Fente and Singh, 2018).

The comparison of performance between machine learning and statistical methods are in dispute. Some researchers believe their effectiveness and efficiency for building predictive models are similar. For example, Sharda and Patil (1992) compared a neural network with a sophisticated forecasting method and found that the neural network

performed similarly with the conventional forecasting method. Makridakis et al. (2018) found the performance of statistical methods in solving univariate prediction problem is even better than machine learning methods. On the other hand, some researchers claim that machine learning methods perform better at solving predictive problems. Cakir et al. (2013) used a multi-layer perceptron to predict the near-surface temperature in Turkey and found that MLP performed better than the simple bias-corrected ensemble mean. Wang et al. (2018) compared the capacity of four different MME methods including random forest (RF), SVM, Bayesian model averaging (BMA) and the arithmetic ensemble mean (EM) in reproducing observed monthly rainfall and temperature with 33 GCMs and found that the RF and SVM demonstrated a significant improvement over EM and BMA in terms of performance criteria. The results also showed that machine learning ensembles could be efficient and useful with improved accuracy in reproducing historical climate variables (Wang et al., 2018). There is no consensus on which category is better for climate projection. Given that machine learning's effectiveness and efficiency have not been tested or compared with traditional statistical ensemble techniques for daily temperature downscaling, their performance needs to be further investigated.

In recent decades, severe rain, ice and wind storms, prolonged heat waves and milder winters have become more common occurrences. Climate change has a great impact on people across the province of Ontario – especially Northern communities – and all sectors

of the economy, which leads to more costs in addressing the impacts of climate change (Ministry of the Environment, 2018). To better quantify climate change, reliable climate projections at finer resolutions over the domain of Ontario are required, which offers information for policymakers in the assessment of the plausible future effects of climate change (Wang et al., 2014). Many of the previous studies for Ontario were based on a single climate model. For instance, Wang et al. (2013) proposed a downscaling technique based on a stepwise cluster analysis method to obtain high-resolution climate projections for the City of Toronto. A high-resolution projection of near-surface air temperature over Ontario was developed with The Providing Regional Climates for Impacts Studies (PRECIS) system (Wang et al., 2014) and was applied to generate future climate projects with dynamical statistical approach (Wang et al., 2015). In terms of MME, only multi-model mean and median ensembles based on monthly temperature data are generated for climate predictions, which outperformed than an individual model (Samouly et al., 2018). Machine learning and statistical ensembles for daily temperature downscaling have not been thoroughly investigated in Ontario.

Therefore, the objective of this study is to develop a number of statistical and machine learning techniques for establishing climate ensembles and compare their performance for downscaling long-term daily temperature. A case study of 12 meteorological stations over Ontario will be conducted to evaluate the performance of machine learning and statistical

ensembles for daily temperature downscaling. The input is the simulated daily mean temperature obtained from six RCM models collected from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) archive, while the output is the observed daily mean temperature collected from the Digital Archive of Canadian Climatological Data. LSTM networks and SVM are applied to develop machine learning ensembles while MLR and EM are used to develop statistical ensembles. This will be the first attempt to introduce LSTM networks for building a climate ensemble. It will also be the first attempt to compare various ensemble techniques for Canadian communities. The results provide technical foundations for using statistical and machine learning methods to generate near-surface air temperature projections with high temporal resolution. Applications of the developed approach will provide useful information to support climate adaptation and social development in Ontario.

4.2 Study Area and Data

Ontario was chosen as the study area to evaluate the performance of the proposed ensembles. Ontario is the second-largest province in Canada, located in the east-central area of Canada and covers more than 106 km². As shown in Fig.4-1, Ontario is bounded by Quebec to the east, Manitoba to the west, Hudson Bay and James Bay to the north, and the Great Lakes to the south (Wang et al., 2014). The climate in Ontario can typically be

considered as humid continental, except for parts of Northern Ontario under the influence of Hudson’s Bay, which have a more maritime climate (Perera et al., 2011). In summer, temperatures in Ontario can soar above 30°C, whereas in winter they can drop below -40°C (Ministry of the Environment, 2011). The annual mean temperature has increased 1.3°C for the Ontario region over the period 1948–2016. The trend is strongest in winter, with an increase of 2.0°C, and weakest in autumn, at 1.0°C. Annual precipitation has increased by 9.7% during the period of 1948–2012, with seasonal trends ranging from 5.2% in winter to 17.8% in fall (Bush et al., 2019).



Fig. 4-1 Locations of the 12 selected meteorological stations

Ontario has made significant progress to address climate change. Using the Paris Agreement baseline year, 2005, as a benchmark, the province of Ontario’s total greenhouse

gas (GHG) emissions has decreased by 22%. According to *A Made-in-Ontario Environment Plan* released on November 29, 2018, Ontario commits to reducing emissions to 30 percent below 2005 levels by 2030 (MECP, 2018).

Table 4-1 Information on selected stations and their corresponding nearest RCM grid

Station Name	Short Name	Station		RCM grid		Elevation
		Latitude	Longitude	Latitude	Longitude	
Big Trout Lake	BTL	53.83°N	89.87°W	53.76°N	89.84°W	224.1m
London						
International Airport	LA	43.03°N	81.15°W	42.98°N	81.04°W	278.0m
Moosonee	MUA	51.27°N	80.65°W	51.34°N	80.60°W	9.1m
North Bay Airport	NB	46.36°N	79.42°W	46.28°N	79.50°W	370.3m
Ottawa						
International Airport	OMIA	45.32°N	75.67°W	45.40°N	75.76°W	222.2m
Sault Ste Marie	SSMA	46.48°N	84.51°W	46.50°N	84.56°W	192.0m

Airport						
Sioux Outlook	SLA	50.12°N	91.90°W	50.02°N	91.82°W	294.7m
Airport						
Timmins Victor	TVPA	48.57°N	81.38°W	48.48°N	81.48°W	383.4m
Power Airport						
Toronto Island	TIA	43.63°N	79.40°W	43.64°N	79.50°W	173.4m
Airport						
Toronto Pearson						
International	TPIA	43.68°N	79.63°W	43.64°N	79.72°W	76.8m
Airport						
Warton Airport	WTA	44.75°N	81.11°W	44.74°N	81.04°W	114.0m
Windsor Airport	WSA	42.28°N	82.96°W	42.32°N	83.02°W	189.6m

In this study, twelve representative stations in Ontario were selected to validate the performance of the proposed ensembles. These twelve stations are located throughout Ontario, as shown in Fig.4-1 and Table 4-1. Among the 12 stations, six stations are in Southern Ontario and six are located in Northern and Central Ontario. The observation data of these twelve stations were downloaded from the Digital Archive of Canadian Climatological Data provided by Environment and Climate Change Canada (ECCC). For

the stations located in Northern Ontario, the Big Trout Lake station, Moosonee and Sioux Lookout Airport station, the standard deviation (SD) of training and testing set are above 14°C. The stations located in Southern Ontario have smaller variance with SD ranges between 9.759°C and 12.485°C.

The study used six RCMs for the temperature ensemble. As shown in Table 4-2, the six GCM and RCM combinations were developed by different institutions. The grid resolution for each RCM is $0.44^\circ \times 0.44^\circ$. The simulated daily mean temperature data were downloaded from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) archive (Mearns et al., 2017), a branch of the International CORDEX Initiative (Giorgi, 2009; Lucas-Picher et al., 2012).

Table 4-2 GCM and RCM combinations

GCM	RCM	Grid	Modeling Institution	Institution Full Name
	CanRCM4	0.44°	CCCma	Canadian Centre for Climate Modelling and Analysis
CanESM2	CRCM5	0.44°	UQAM	Université du Québec à Montréal
	RCA4	0.44°	SMHI	Swedish Meteorological and Hydrological Institute

	HIRHAM5	0.44°	DMI	Danish Meteorological Institute
EC-EARTH	RCA4	0.44°	SMHI	Swedish Meteorological and Hydrological Institute
MPI-ESM-LR	CRCM5	0.44°	UQAM	Université du Québec à Montréal

4.3 Methodology

This study investigates the performance of MMEs in predicting daily temperature with two representative machine learning methods, which are widely applied in temperature downscaling including LSTM networks and SVM, and two statistical methods including EM and MLR.

4.3.1 Long Short-Term Memory Networks

The LSTM network is a special Recurrent Neural Networks (RNNs) structure that has been proven to be stable and powerful for modeling long-range dependencies in various previous studies (Graves, 2013; Hochreiter and Schmidhuber, 1997; Sutskever et al., 2014). Compared with feedforward neural networks, RNNs allow forward and backward connections between time steps, which makes them well suited for processing sequential data (Mouatadid et al., 2017). The major innovation of LSTM is the memory cell, in which the information is selectively accumulated. The cell is accessed, written and cleared by

several self-parameterized controlling gates (Shi et al., 2015). Every time when the memory cell has new inputs; the forget gate f_t decides whether the past cell status c_{t-1} will be “forgotten”. The output of f_t ranges between 0 and 1 where a 1 represents that piece of information in the corresponding component of c_{t-1} will be kept and a 0 represents it will be “forgotten”. Information of new inputs will accumulate in the cell if the input gate i_t is activated. The input modulation \tilde{c}_t modulates the information of the input gate (Kong et al., 2018; Kratzert et al., 2018). The output gate o_t controls whether the latest cell state c_t will be propagated to the hidden state h_t (Kim et al., 2017; Shi et al., 2015).

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (4.1)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (4.2)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (4.3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4.4)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (4.5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (4.6)$$

Where \odot denotes element-wise multiplication of two vectors and σ is the logistic sigmoid function. The input gate i , forget gate f , output gate o , cell c and cell input activation vectors are the same size as the hidden vector h . W is the weight matrix, for instance, W_{hi} is the hidden-input gate matrix (Graves, 2013). b denotes the bias vector, for example, b_i is the input gate bias vector (Donahue et al., 2017).

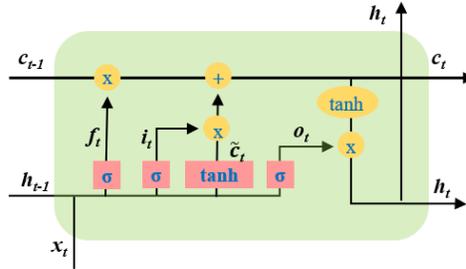


Fig. 4-2 A LSTM memory cell with input, forget, and output gates

4.3.2 Support Vector Machine

SVM was developed for binary classification and then extended to regression problems, which was called Support Vector Regression (SVR) (Drucker et al., 1997; Zhao and Magoulès, 2012). The SVM model estimates the regression based on a series of kernel functions, which are able to convert the original, lower-dimensional input data to a higher-dimensional feature space implicitly (Fan et al., 2018). The estimation of the regression model could be expressed as Eq. 4.7.

$$\hat{y} = \langle w, x \rangle + b \quad (4.7)$$

where w is the vector of feature weights, angle brackets denote a dot product and b is a bias term (Li et al., 2009).

SVR integrates loss function to minimize the prediction error and intends to create a boundary to include as many samples for reliability as possible (Wang et al., 2019). Following regularization theory (Haykin, 2003), the parameters w and b are estimated by minimizing the cost function as shown in Eq. 4.8 (Tripathi et al., 2006b).

$$\text{Min} : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (4.8)$$

where ξ_i and ξ_i^* are positive slack variables which capture the magnitude of residuals beyond the prescribed tolerance ε and serve to guarantee a solution for all ε . C is a regularization term that determines the degree of the linear penalty applied to the residual excess (Jain et al., 2014).

Accordingly, Eq. 4.8 is subject to the following constraints:

$$y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \quad (4.9)$$

$$\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \quad (4.10)$$

$$\xi_i^* \geq 0, \xi_i \geq 0 \quad (4.11)$$

where x_i is a data point in the input space \mathbf{X} and y_i is the corresponding output (Kisi and Sanikhani, 2015a, 2015b).

The commonly applied kernel functions include linear, polynomial, radial basis and sigmoid functions. The Gaussian radial basis function (RBF) is one of the most widely used kernel functions which could generalize nonlinear functions and have performed well at large datasets. The RBF kernel is formulated as follows:

$$\varphi(x, x') = \exp(-\gamma |x - x'|^2), \gamma > 0 \quad (4.12)$$

where γ is the kernel parameter. Intuitively, γ defines the radius of influence for each data point.

4.3.3 Statistical Methods

This study used mean and MLR methods to represent the performance of statistical methods in daily temperature downscaling with MMEs. A mean ensemble refers to merging all the input models with equal weights, which is known as the EM (Zhang et al., 2015).

$$Y(t) = \frac{1}{N} \sum_{n=1}^N P_n(t) \quad (4.13)$$

where $Y(t)$ is an ensemble mean for time t , N is the total number of RCMs and $P_n(t)$ is the prediction of the n^{th} RCM for time t .

MLR has an equation of the form

$$Y(t) = b_0 + \sum_{n=1}^N b_n P_n(t) \quad (4.14)$$

where b_0 and b_n are regression coefficients, and b_n is a weighting for model n (Krishnamurti et al., 1999b). What differentiates it from the EM is that MLR allocates unequal weights to predictors. The regression coefficients are obtained by fitting the equation with observations and the Regional Climate Model (RCM) inputs in the training set. Then, the fitted values of regression coefficients are applied to downscaling the daily temperature in the testing data (Fumo and Rafe Biswas, 2015).

4.3.4 Design and Training of Downscaling Models

The study used 10 years of RCM daily temperature data and observed daily

temperature data to evaluate the performances of ensembles. The first 70% of the dataset (1980 - 1986) was used for training the models. The remaining 30% of the dataset (1987 - 1989) was used for testing the models, which was used to assess the generalization capability of the model. Inputs to the ensemble models were the simulated daily mean temperature of six RCMs outputs at the closest grid point to 12 meteorological stations. While the output was daily mean temperature observed at meteorological stations. The information of the 12 meteorological stations and their corresponding closest RCM grid points are shown in Table 4-1.

The structure of LSTM networks used in the study consisted of one input layer, one hidden layer, and one output layer. The LSTM network was trained with lag time (time delay) ranging from 1 to 7 days and the number of neurons ranging from 10 to 60. Data from the Toronto Pearson International Airport station was used to find the best parameter suitable for the LSTM ensemble and SVM ensemble. The best parameters of machine learning methods were chosen with mean square error (MSE) of observed and simulated temperature. The LSTM ensemble structure with 30 neurons and a time lag of 3 days was selected as it generated the best performing network for the Toronto Pearson International Airport Station. The generalized parameters of the LSTM ensemble and SVM ensemble were applied for the other 11 stations.

The structure of the SVM ensemble used in the study was selected based on the results

of 5-fold cross-validation on the training set. SVM with different combinations of C and γ were tested to find the optimal parameters. C was set as 0.1, 0.5, 1 and 5 and γ was set as 0.0001, 0.0005, 0.001, 0.01, 0.1 and 1. The optimal value of the SVM ensemble for C and γ were 1 and 0.0001, respectively.

In terms of the MLR ensemble, the coefficient of each RCM for each station was obtained by fitting with the corresponding training data of that station. In the mean ensemble, the predicted value was generated by taking the average of six RCM inputs. Performances of machine learning and statistical ensembles were evaluated by comparing predicted results with observed temperature values. Statistical criteria, such as root mean square error (RMSE), coefficient of determination (R^2) and the ratio of RMSE to the standard deviation (SD) were used for performance evaluation.

4.4 Results and Discussion

In this study, a total of four statistical and machine learning techniques (LSTM networks, SVM, EM and MLR) were used to build the ensembles for downscaling. The models were trained using data from 1980 to 1986, and tested using data from 1987 to 1989.

4.4.1 Downscaling Performance of Long Short-Term Memory networks

Table 4-3 shows the statistical performance of the LSTM ensemble for downscaling.

The R^2 of 12 stations range between 0.779 and 0.820. The RMSE of 12 stations ranges between 4.318°C and 6.540°C . Comparing with the temperature ranges of 12 stations, the ratios of RMSE to SD of the corresponding station ranges between 0.426 and 0.471. This indicates that the error of LSTM models for 12 stations is relatively small. As shown in Fig. 4-3 to 4-5, downscaled temperature of LSTM ensembles are closely concentrated around the best fit line. The high R^2 value and low RMSE value indicate that the LSTM technique performs well at building a relationship between daily temperature observation and simulation data from the six RCMs, and could predict daily temperature with relatively high accuracy. However, for all 12 stations, the R^2 of the training dataset is larger than the testing dataset and the RMSE of the training dataset is smaller than the testing dataset, which indicates that LSTM is prone to overfitting.

The observation and prediction values scattered more closely at Toronto Island Airport station with high accuracy. The temperature has a small variance for that station and temperature mostly lies between -10°C and 25°C without extreme low temperatures below -20°C . For the Big Trout Lake station, prediction value scattered more tightly for temperature above -10°C than temperature below -10°C . This indicates that the LSTM ensemble could not predict the temperature below -10°C with high accuracy. That is because most of the temperature data are above -10°C and neural networks tend to sacrifice variance to gain high RMSE.

Table 4-3 Statistical performance of LSTM ensemble

Station	RMSE(training) (°C)	RMSE(testing) (°C)	R ² (training)	R ² (testing)	RMSE/SD (training)	RMSE/SD (testing)
BTL	6.005	6.540	0.842	0.820	0.400	0.426
LA	4.560	4.961	0.820	0.789	0.430	0.461
MUA	6.126	6.289	0.813	0.799	0.437	0.452
NB	5.183	5.784	0.821	0.790	0.426	0.460
OMIA	4.695	5.465	0.849	0.810	0.392	0.438
SSMA	4.806	5.378	0.814	0.779	0.434	0.470
SLA	5.891	6.426	0.828	0.805	0.417	0.439
TVPA	5.752	6.384	0.816	0.787	0.433	0.465
TIA	3.985	4.318	0.841	0.817	0.408	0.430
TPIA	4.460	4.965	0.825	0.790	0.422	0.458
WTA	4.621	4.930	0.800	0.781	0.454	0.471
WSA	4.480	4.975	0.830	0.792	0.417	0.460

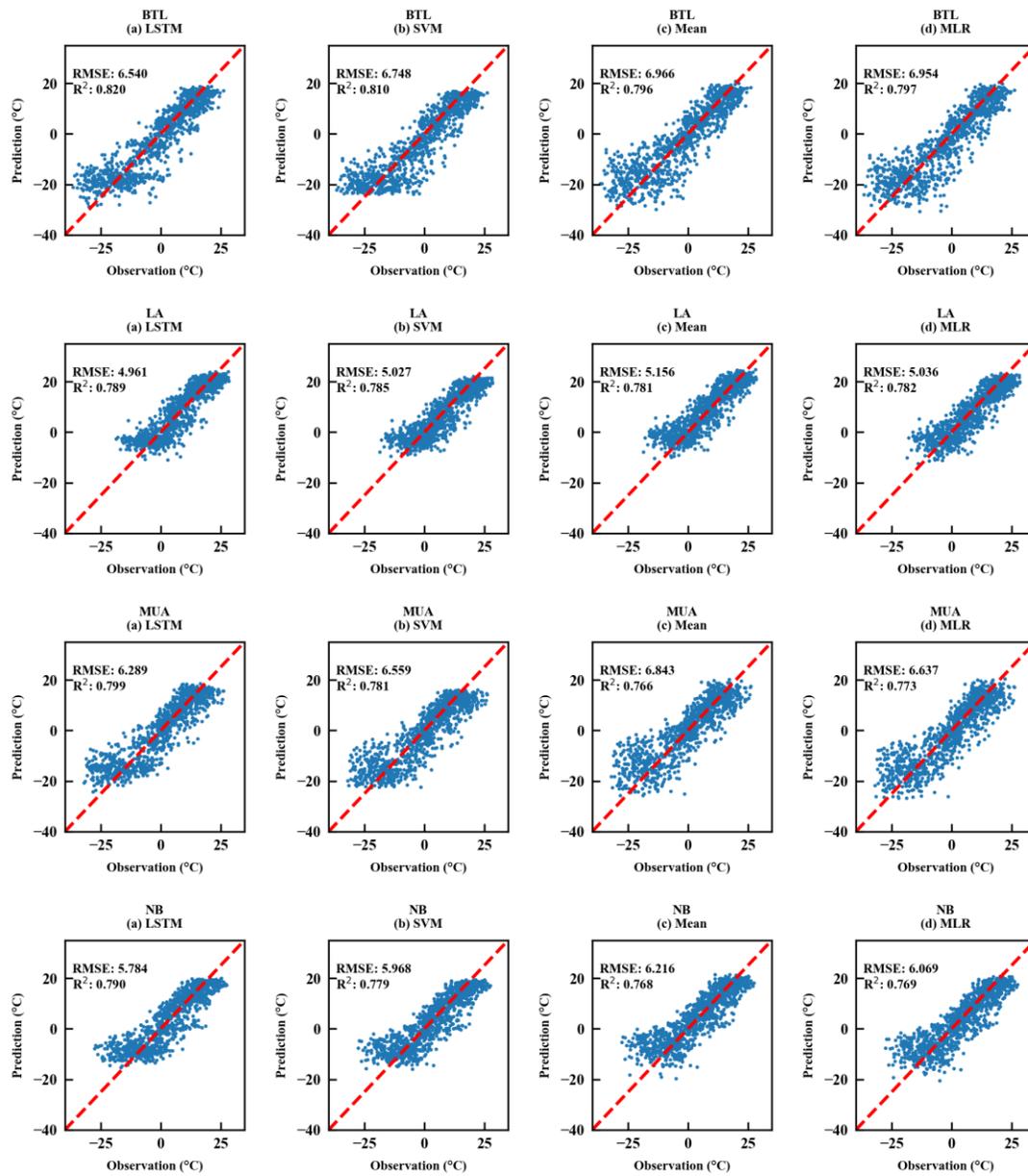


Fig. 4-3 Scatter plots of the first 4 stations for testing datasets

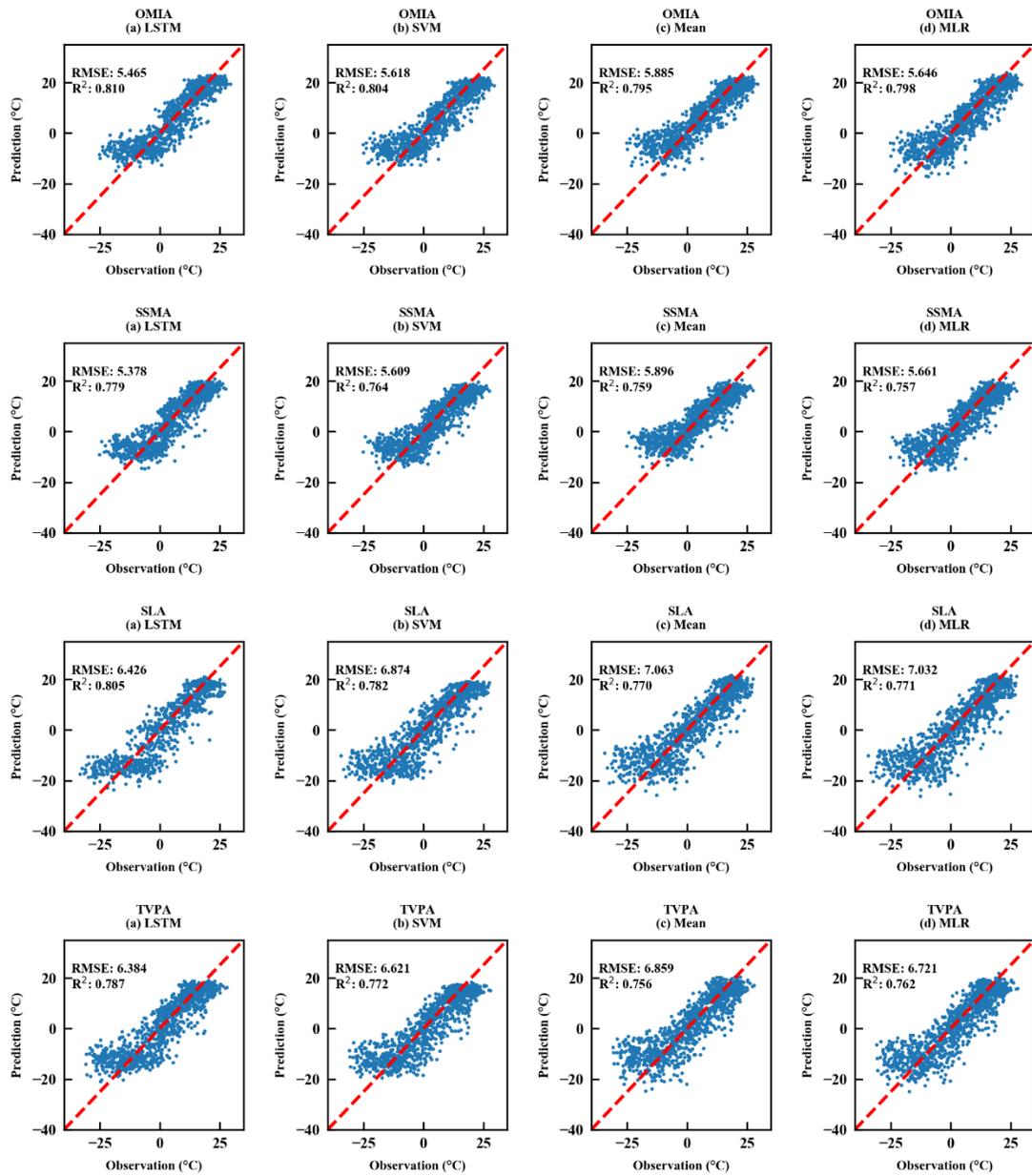


Fig. 4-4 Scatter plots of the middle 4 stations for testing datasets

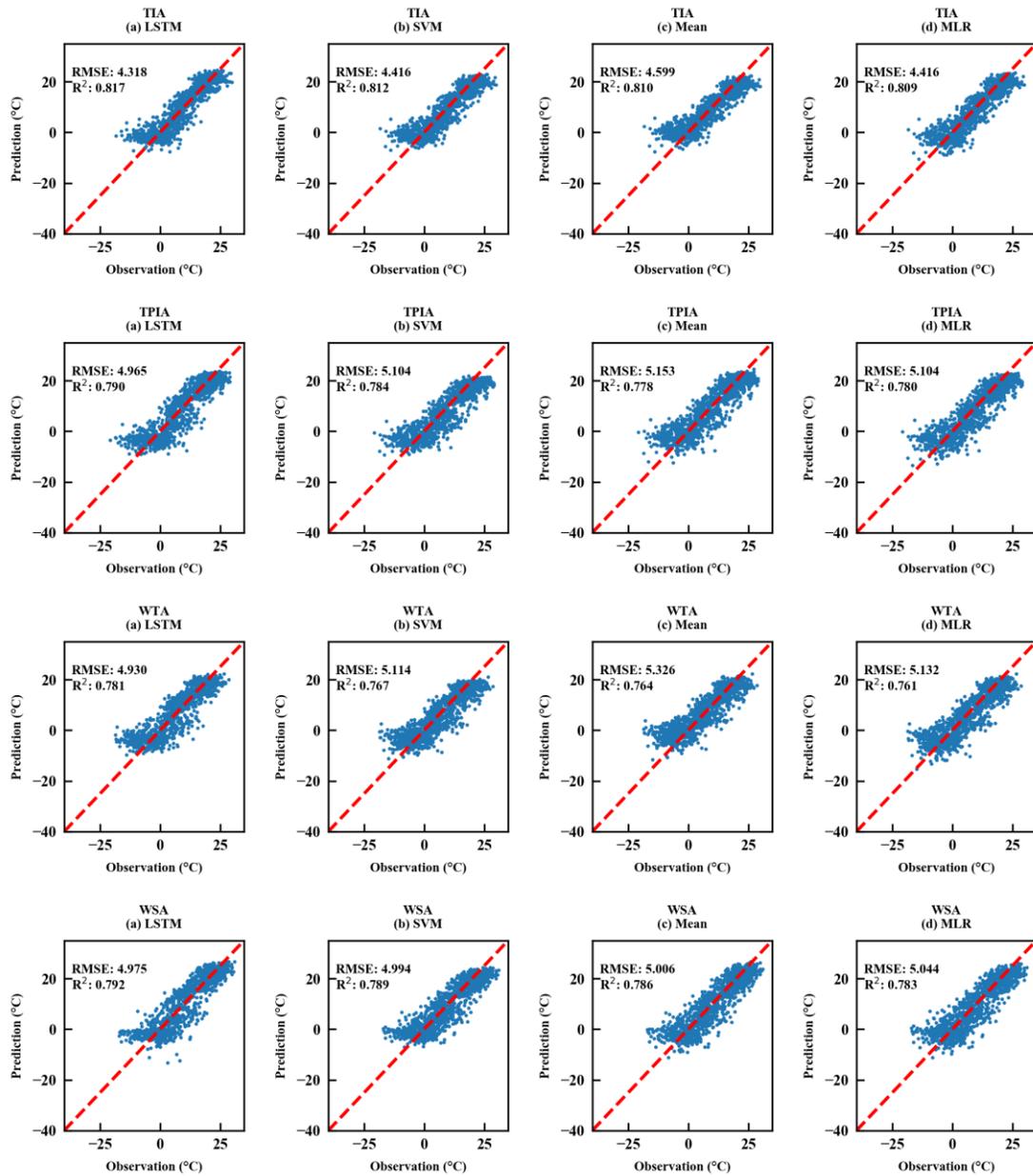


Fig. 4-5 Scatter plots of the last 4 stations for testing datasets

4.4.2 Downscaling Performance of Support Vector Machine

Table 4-4 shows the statistical performance of the SVM ensemble for downscaling. The R^2 of 12 stations ranges from 0.764 to 0.812. The RMSE values are similar to that of LSTM ensemble and range from 4.416°C to 6.874°C. The ratio of RMSE to SD of testing data ranges from 0.439 to 0.490, which is also similar to the LSTM ensemble and demonstrates that SVM performs well at downscaling daily mean temperature. Compared to the results of the LSTM ensemble, the downscaled temperature of the SVM ensemble is scattered not as closely as the LSTM ensemble around the observation data, especially for temperature lower than -10°C.

As shown in Fig. 4-3 to 4-5, the prediction values scatter closely around observation values for temperatures above 0°C, which indicates that SVM models perform well at predicting temperature above 0°C. However, for stations with high variance, SVM tends to generate downscaled temperature ranges from -20°C to -10°C for observation ranges from -30°C to -10°C (the Big Trout Lake station, the Moosonee UA station, the Sioux Lookout Airport station, and the Timmins Victor Power Airport station). For other stations, the SVM ensemble tends to generate downscaled temperature ranges from -10°C to 0°C for temperature ranges from -20°C to 0°C, which demonstrates the SVM ensemble could not capture extreme low temperatures well.

Table 4-4 Statistical performance of SVM ensemble

Station	RMSE(training) (°C)	RMSE(testing) (°C)	R ² (training)	R ² (testing)	RMSE/SD (training)	RMSE/SD (testing)
BTL	6.540	6.748	0.810	0.810	0.436	0.439
LA	5.006	5.027	0.778	0.785	0.472	0.467
MUA	6.801	6.559	0.767	0.781	0.485	0.471
NB	5.837	5.968	0.771	0.779	0.480	0.475
OMIA	5.404	5.618	0.799	0.804	0.451	0.450
SSMA	5.321	5.609	0.772	0.764	0.480	0.490
SLA	6.517	6.874	0.789	0.782	0.461	0.470
TVPA	6.478	6.621	0.764	0.772	0.487	0.482
TIA	4.357	4.416	0.804	0.812	0.446	0.440
TPIA	4.984	5.104	0.779	0.784	0.472	0.471
WTA	5.046	5.114	0.757	0.767	0.495	0.488
WSA	4.970	4.994	0.787	0.789	0.463	0.462

4.4.3 Downscaling Performance of Statistical Methods

Table 4-5 shows the statistical performance of the mean ensemble for downscaling. The R² of 12 stations ranges from 0.756 to 0.810. The RMSE is a bit higher than those of the

LSTM and SVM ensembles, and range from 4.599°C to 7.063°C. The ratio of RMSE to SD of testing data ranges from 0.453 to 0.515, which is a bit higher than the LSTM and SVM ensemble. However, the difference between mean ensemble and machine learning ensemble is not significant. As shown in Fig. 4-3 to 4-5, the downscaled temperature value scattered closely and evenly around 45° line for temperatures above 0°C, while for observations under 0 °C, the scatter points spread more widely, which indicates that the mean ensemble could predict temperature above 0 °C with relatively high accuracy. The mean ensemble also tends to overestimate low temperatures.

Table 4-5 Statistical performance of mean ensemble

Station	RMSE(training) (°C)	RMSE(testing) (°C)	R ² (training)	R ² (testing)	RMSE/SD (training)	RMSE/SD (testing)
BTL	7.030	6.966	0.786	0.796	0.468	0.453
LA	5.287	5.156	0.768	0.781	0.498	0.479
MUA	7.201	6.843	0.744	0.766	0.513	0.491
NB	6.156	6.216	0.754	0.768	0.506	0.495
OMIA	5.681	5.885	0.782	0.795	0.474	0.471
SSMA	5.691	5.896	0.763	0.759	0.514	0.515
SLA	6.754	7.063	0.773	0.770	0.478	0.483

TVPA	6.797	6.859	0.743	0.756	0.511	0.500
TIA	4.541	4.599	0.799	0.810	0.465	0.458
TPIA	5.171	5.153	0.767	0.778	0.490	0.475
WTA	5.361	5.326	0.746	0.764	0.526	0.508
WSA	5.116	5.006	0.774	0.786	0.477	0.463

Table 4-6 shows the statistical performance of MLR ensemble for downscaling. The R^2 of 12 stations ranges from 0.757 to 0.809 with the highest and lowest R^2 . The RMSE is similar to that of mean ensemble and ranges from 4.416°C to 7.032°C and the ratio of RMSE to SD of testing data ranges from 0.440 to 0.495. Similar to mean ensemble, the downscaled temperature of MLR ensemble is also scattered closely and evenly around the observation data. It demonstrates that statistical ensembles have similar performance in downscaling daily temperature and could generate reliable daily downscaled temperature. As shown in Fig.4-3 to 4-5, the scatter points spread more widely for temperature below 0°C than above 0°C, which implies that statistical techniques could not capture low temperatures as precisely as they could for higher temperatures.

Table 4-6 Statistical performance of MLR ensemble

Station	RMSE(training) (°C)	RMSE(testing) (°C)	R ² (training)	R ² (testing)	RMSE/SD (training)	RMSE/SD (testing)
BTL	6.773	6.954	0.796	0.797	0.451	0.453
LA	5.048	5.036	0.774	0.782	0.476	0.468
MUA	6.951	6.637	0.755	0.773	0.495	0.477
NB	5.943	6.069	0.761	0.769	0.488	0.483
OMIA	5.524	5.646	0.788	0.798	0.461	0.452
SSMA	5.333	5.661	0.768	0.757	0.482	0.495
SLA	6.65	7.032	0.779	0.771	0.470	0.481
TVPA	6.61	6.721	0.753	0.762	0.497	0.489
TIA	4.361	4.416	0.800	0.809	0.447	0.440
TPIA	5.033	5.104	0.773	0.780	0.477	0.471
WTA	5.051	5.132	0.754	0.761	0.496	0.490
WSA	5.007	5.044	0.782	0.783	0.466	0.467

4.4.4 Comparison Between the Downscaling Performance of Machine Learning and Statistical Techniques

Tables 4-3 to 4-6 show the statistical evaluation metrics for machine learning and

statistical downscaling including R^2 , RMSE and the ratios of RMSE to SD. For all 12 stations tested, the performance of machine learning and statistical ensembles are similar, with R^2 ranging between 0.756 and 0.820, RMSE ranging between 4.318°C and 7.063°C, and the ratios of RMSE to SD range between 0.426 and 0.515.

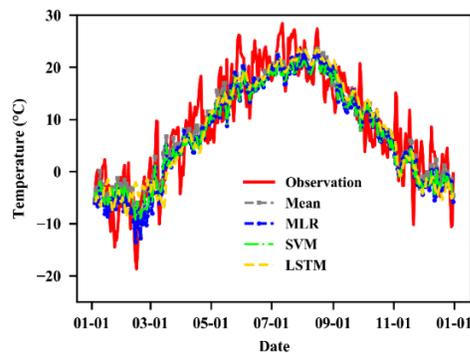


Fig. 4-6 A time series plot of ensembles of the Toronto Pearson International Airport station in 1987

Fig. 4-6 shows the time series plot of machine learning and statistical ensembles of the Toronto Pearson International Airport station in 1987. All ensembles could grasp the general increasing and decreasing trend of daily temperature. The observation of 1987 ranges between -18.7°C and 28.4°C . The downscaled temperature ranges generated by LSTM and SVM are $[-8.7^{\circ}\text{C}, 23.5^{\circ}\text{C}]$ and $[-9.2^{\circ}\text{C}, 21.8^{\circ}\text{C}]$, respectively. While for statistical ensembles, the downscaled temperature ranges of mean and MLR ensembles are $[-12.3^{\circ}\text{C}, 23.6^{\circ}\text{C}]$ and $[-13.5^{\circ}\text{C}, 22.9^{\circ}\text{C}]$, respectively. The downscaled temperature values

for all ensembles lie within the range of observation values. The lower bounds of statistical ensembles are lower than those of machine learning ensembles. For instance, during the end of February 1987, prediction of MLR and mean ensemble are lower than LSTM and SVM ensembles. The upper bounds of statistical ensembles are similar to those of machine learning ensembles. However, all ensembles could not generate downscaled temperatures above 25°C and below -14°C and the variance between daily temperature could not be accurately captured by all ensembles.

Machine learning and statistical ensembles perform well at downscaling daily temperature and have similar performance. In terms of R^2 and RMSE, machine learning ensembles perform slightly better than statistical ensembles. However, it is worth mentioning that the difference is not significant. The improved performance of the LSTM ensemble can be explained by its ability to learn long-term dependencies in sequential datasets. In the case of other ensemble methods, the inputs in the time series are assumed to be independent of each other and are not treated as sequential datasets. In contrast, the inputs are processed in the sequence in the LSTM ensemble and the outputs are generated based on the previous computations. In other words, LSTMs have a "memory" which enables them to identify and capture the previously calculated useful information, and pass them along to the next iteration. Moreover, the daily temperature has a close connection with the temperature of previous days. Incorporating antecedent predictor values could

improve the accuracy of temperature prediction (Coulibaly et al., 2005). The slightly improved performance of SVM can be explained by the framework and principles of SVM. SVM implements structural risk minimization (SRM) (Vapnik, 2013), uses margin-based loss function and solves the problem in high dimensional feature space (Al-Anazi and Gates, 2010; Samui, 2008). Besides, SVM can achieve good generalization by simultaneously minimizing both empirical error and model complexity (Yoon et al., 2016).

Machine learning methods have been proven to outperform statistical methods in many fields, for instance, speed prediction for traffic control (Jiang et al., 2016). However, in this study, the difference between machine learning and statistical ensembles is not as significant as other studies, which is attributed to having the same variable type of input and output. The variable type of both the inputs and output is daily mean temperature, which is categorized as the same class. Furthermore, the range and distribution of inputs and observation values are similar. Previous studies have shown that for univariate forecasting, statistical methods such as AutoRegressive Integrated Moving Average (ARIMA) outperform machine learning methods (Makridakis et al., 2018).

In general, statistical ensembles have similar capabilities in processing and predicting temporal patterns as machine learning ensembles for downscaling, while having much simpler structures and being less computationally demanding and time-consuming. Thus, statistical ensembles could be used as a substitute for downscaling daily temperature for

those lacking computational resources. Among them, the mean ensemble is recommended as it is the simplest ensemble and is capable of generating reliable downscaled daily temperature in all stations tested in this study.

4.4.5 Current Difficulties in Predicting Extreme Values

The reasons for low accuracy in predicting ‘extreme’ values (above 20°C and below -10°C) is due to two different circumstances: one, when the observation values exceed the maximum or minimum values of six RCMs; and two, when the observation value is located between the maximum values and minimum values of six RCMs. Samples of RCMs and observation from February 1987 and 1988 at the Big Trout Lake station are chosen to demonstrate the two circumstances of predicting extreme values as an example.

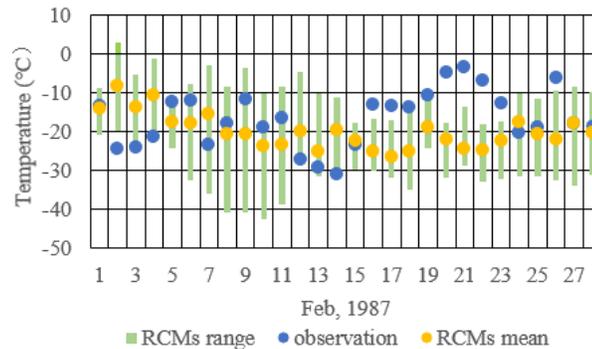


Fig. 4-7 Case 1- observation beyond the maximum and minimum values of RCMs

Fig.4-7 shows that the observation values exceed the range of maximum and minimum values of six RCMs. The maximum and minimum values of six RCMs range between -18.2°C and 2.8°C and between -42.4°C and -16.3°C, respectively, while the observation value ranges between -30.9°C and -3.3°C. The observation values exceed the minimum value of input data for 10 days of the month, particularly between Feb. 20th to 23rd. In this case, the ensembles need to generate a downscaled temperature that exceeds the minimum value of input RCMs.

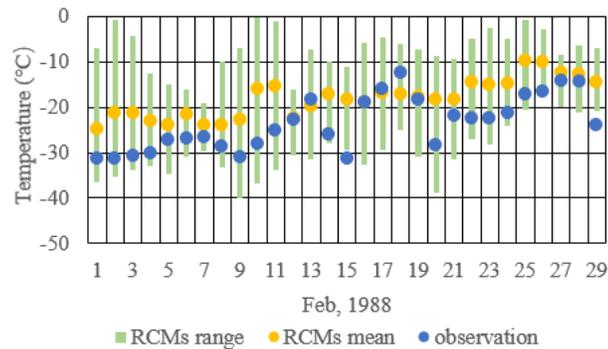


Fig. 4-8 Case 2 - observation within the maximum and minimum values of RCMs

Fig. 4-8 shows that the observed values lie between the maximum and minimum values of the six models. The maximum and minimum values of six RCMs range [-19.0°C, -0.1°C] and [-40.0°C, -16.3°C], respectively, while the observation value ranges between -31.3°C and -12.6°C. Most of the observation values lie between the maximum and minimum values of six RCM inputs.

As ensembles are built based on both circumstances, the parameters and the structures of ensembles are the same for both cases. Specifically, for the LSTM ensemble, the weight of neurons and activate functions are the same for those two cases. For case one, the downscaled temperature value should either greater than the maximum value or lesser than the minimum value while for case two, the downscaled temperature should be between the maximum and minimum value. Thus, machine learning methods and statistical methods may have difficulty in identifying the scenario to be case one or two, which leads to low accuracy in predicting extreme values.

4.4.6 Performance Comparison among 12 Stations over Ontario

Fig. 4-9 shows the histogram of training and testing of 12 stations. The distribution of training and testing dataset is similar for all stations, which provides a foundation for both machine learning and statistical ensembles generating reliable downscaled temperature.

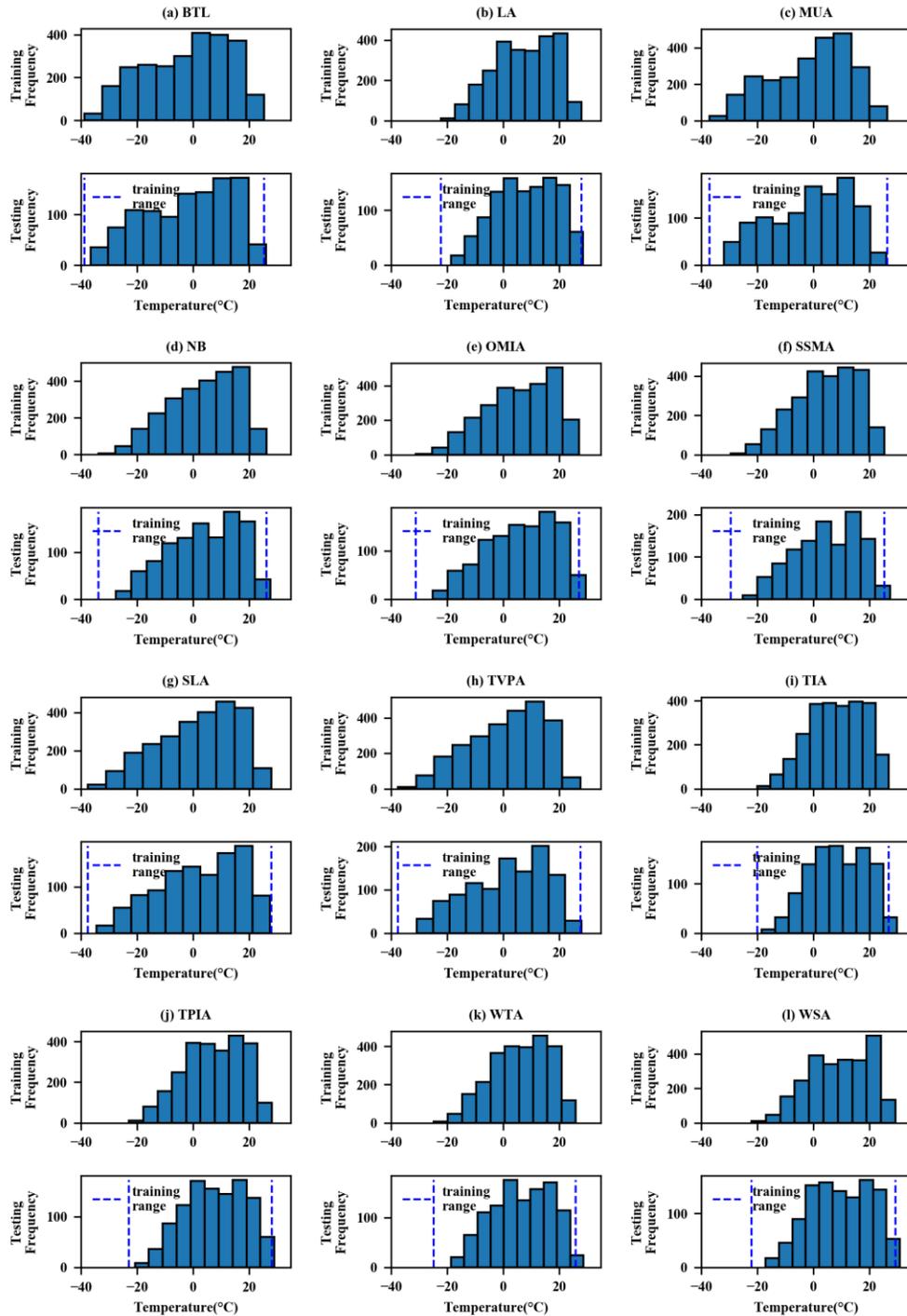


Fig. 4-9 Histogram of 12 stations (the blue dotted line represents the range of training data)

Fig. 4-10 shows the ratio of RMSE to SD of training and testing for LSTM, SVM, mean and MLR ensemble. For all four ensembles, twelve stations have similar performance, with the ratio of RMSE to SD ranging between 0.426 and 0.515. Moreover, the rank of 12 stations for all ensembles are also similar. Big Trout Lake station, Toronto Island Airport station, and Ottawa International Airport station are the best 3 performers overall, having low RMSE to SD ratio values for both machine learning and statistical ensembles. This may be because the training data at these three stations are more representative for predicting the patterns of the testing period. The Sault Ste Marie Airport station, the Wiarton Airport, and the Timmins Victor Power Airport station performed the poorest with a high RMSE to SD ratio for all ensembles. Although stations located in Northern Ontario tend to have higher variance, the ratio between RMSE and SD shows no significant difference with southern stations. Moreover, though the machine learning ensembles were calibrated using training data of the Toronto Pearson International Airport station, both LSTM ensemble and SVM ensembles perform well across all 12 stations, with a low RMSE to SD ratio. The good performance of all four ensembles indicated that both machine learning and statistical ensembles are stable and reliable for generating downscaled daily temperature.

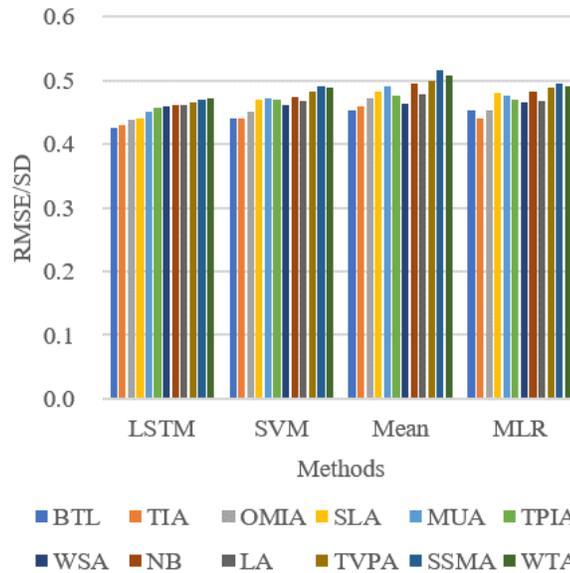


Fig. 4-10 RMSE to SD ratios of machine ensembles and statistical ensembles for 12 stations

4.5 Conclusions

This study investigated the performance of machine learning methods (LSTM networks and SVM) and statistical methods (EM and MLR) in developing multi-model ensembles for downscaling daily temperature in Ontario, Canada. The input used the simulated daily mean temperature obtained from six RCMs collected from NA-CORDEX and the output was the observed daily mean near-surface air temperature obtained from ECCC. Twelve meteorological stations over Ontario were chosen to evaluate the downscaling performance of machine learning and statistical ensembles. Data from 1980 to 1989 were used for training and testing the ensembles. The results of ensembles were compared with the

observation data, and the downscaling performance of ensembles was evaluated by RMSE, R^2 , and ratio of RMSE to SD.

The study showed that both machine learning and statistical ensembles performed well at downscaling daily mean temperature. Machine learning and statistical ensembles had a similar downscaling performance with relatively high accuracy. The R^2 of 12 stations ranged between 0.756 and 0.820, RMSE ranged between 4.318°C and 7.063°C, and the ratio of RMSE to SD ranged between 0.426 and 0.515. The high R^2 and low RMSE of 12 stations indicated both machine learning and statistical ensembles could generate stable and reliable downscaled daily temperatures. Considering that machine learning ensembles are computationally demanding and time-consuming, the mean ensemble could be used as a less computationally demanding method for generating downscaled daily temperature and developing near/long-term scenarios of regional climate change for the future with similar accuracy to machine learning ensembles.

Machine learning and statistical techniques both have difficulty predicting extreme values. Specifically, machine learning ensembles showed a trend for under-estimating the high observed temperature above 20°C and an over-estimating trend for the low observed temperature below -10°C in both the calibration and validation periods. The relationship between observation and RCMs could be divided into two cases: observation lying between or exceeding the maximum and minimum value of six RCMs. Ensembles could not

recognize which circumstance to apply in downscaling temperature, leading to low accuracy in predicting extreme values. Thus, further study should be focusing on solving the difficulty of predicting extreme values.

Acknowledgement:

This study is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the McMaster Engineering Big Ideas Initiative of McMaster University.

4.6 References

- Al-Anazi, A. F., and Gates, I. D. (2010). Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study. *Computers & Geosciences*, 36(12), 1494-1503.
- Bartos, M. D., and Chester, M. V. (2015). Impacts of climate change on electric power supply in the Western United States. *Nature Climate Change*, 5(8), 748-752.
- Bush, E., Lemmen, D. S., and editors. (2019). Canada's Changing Climate Report. *Government of Canada, Ottawa, ON, 444 p.*
- Cakir, S., Kadioglu, M., and Cubukcu, N. (2013). Multischeme ensemble forecasting of surface temperature using neural network over Turkey. *Theoretical and Applied*

Climatology, 111(3), 703-711.

Christensen, J. H., Kjellström, E., Giorgi, F., Lenderink, G., and Rummukainen, M. (2010).

Weight assignment in regional climate models. *Climate Research*, 44(2-3), 179-194.

Coulibaly, P., Dibike, Y. B., and Anctil, F. (2005). Downscaling Precipitation and

Temperature with Temporal Neural Networks. *Journal of Hydrometeorology*, 6(4),

483-496.

Devak, M., Dhanya, C., and Gosain, A. (2015). Dynamic coupling of support vector

machine and K-nearest neighbour for downscaling daily rainfall. *Journal of*

Hydrology, 525, 286-301.

Dibike, Y. B., and Coulibaly, P. (2006). Temporal neural networks for downscaling climate

variability and extremes. *Neural Networks*, 19(2), 135-144.

Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko,

K., and Darrell, T. (2017). Long-Term Recurrent Convolutional Networks for Visual

Recognition and Description. *IEEE Trans Pattern Anal Mach Intell*, 39(4), 677-691.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., and Vapnik, V. (1997). *Support vector*

regression machines. Paper presented at the Advances in neural information

processing systems.

ECCC. Digital Archive of Canadian Climatological Data. Retrieved from

http://climate.weather.gc.ca/historical_data/search_historic_data_e.html

- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., and Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology*, 263, 225-241.
- Feng, J., Lee, D.-K., Fu, C., Tang, J., Sato, Y., Kato, H., McGregor, J. L., and Mabuchi, K. (2010). Comparison of four ensemble methods combining regional climate simulations over Asia. *Meteorology and Atmospheric Physics*, 111(1-2), 41-53.
- Fente, D. N., and Singh, D. K. (2018). *Weather Forecasting Using Artificial Neural Network*. Paper presented at the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT).
- Fumo, N., and Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, 47, 332-343.
- Giorgi, F., Jones, C., & Asrar, G. R. . (2009). Addressing climate information needs at the regional level: the CORDEX framework. *World Meteorological Organization (WMO) Bulletin*, 58(3), 175.
- Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., and Wiltshire, A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. *Philos Trans R Soc Lond B Biol Sci*, 365(1554), 2973-2989.
- Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint*

arXiv:1308.0850.

Hagedorn, R., DOBLAS-REYES, F. J., and Palmer, T. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting–I. Basic concept. *Tellus A*, 57(3), 219-233.

Hatfield, J. L., and Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10, 4-10.

Haykin, S. (2003). Neural Networks: A comprehensive foundation. *Fourth Indian Reprint*, Pearson Education, Singapore, pp. 842.

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.

Jain, R. K., Smith, K. M., Culligan, P. J., and Taylor, J. E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, 123, 168-178.

Jarsjö, J., Törnqvist, R., and Su, Y. (2017). Climate-driven change of nitrogen retention–attenuation near irrigated fields: multi-model projections for Central Asia. *Environmental Earth Sciences*, 76(3).

Jiang, H., Zou, Y., Zhang, S., Tang, J., and Wang, Y. (2016). Short-Term Speed Prediction Using Remote Microwave Sensor Data: Machine Learning versus Statistical Model.

Mathematical Problems in Engineering, 2016, 1-13.

Kendon, E. J., Jones, R. G., Kjellström, E., and Murphy, J. M. (2010). Using and Designing GCM–RCM Ensemble Regional Climate Projections. *Journal of Climate*, 23(24), 6485-6503.

Kim, S., Hong, S., Joh, M., and Song, S. K. (2017). DeepRain: ConvLSTM Network for Precipitation Prediction using Multichannel Radar Data. *arXiv:1711.02316*.

Kisi, O., and Sanikhani, H. (2015a). Modelling long-term monthly temperatures by several data-driven methods using geographical inputs. *International Journal of Climatology*, 35(13), 3834-3846.

Kisi, O., and Sanikhani, H. (2015b). Prediction of long-term monthly precipitation using several soft computing methods without climatic data. *International Journal of Climatology*, 35(14), 4139-4150.

Kong, Y.-L., Huang, Q., Wang, C., Chen, J., Chen, J., and He, D. (2018). Long Short-Term Memory Neural Networks for Online Disturbance Detection in Satellite Image Time Series. *Remote Sensing*, 10(3).

Krasnopolsky, V. M., and Lin, Y. (2012). A Neural Network Nonlinear Multimodel Ensemble to Improve Precipitation Forecasts over Continental US. *Advances in Meteorology*, 2012, 11.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). Rainfall–runoff

modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005-6022.

Krishnamurti, T., Kishtawal, C., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S. (1999a). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, 285(5433), 1548-1550.

Krishnamurti, T., Kishtawal, C., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S. J. S. (1999b). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, 285(5433), 1548-1550.

Krishnamurti, T., Mishra, A., Chakraborty, A., and Rajeevan, M. (2009). Improving global model precipitation forecasts over India using downscaling and the FSU superensemble. Part I: 1–5-day forecasts. *Monthly Weather Review*, 137(9), 2713-2735.

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., and Surendran, S. (2000). Multimodel Ensemble Forecasts for Weather and Seasonal Climate. *Journal of Climate*, 13(23), 4196-4216.

Kumar, A., Mitra, A., Bohra, A., Iyengar, G., and Durai, V. (2012). Multi-model ensemble (MME) prediction of rainfall using neural networks during monsoon season in India. *Meteorological Applications*, 19(2), 161-169.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

- Li, P., Kwon, H., Sun, L., Lall, U., and Kao, J. (2009). A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan. *International Journal of Climatology*, 30(8), 1256-1268.
- Li, Q., Wang, S., Lee, D.-K., Tang, J., Niu, X., Hui, P., Gutowski, W. J., Dairaku, K., McGregor, J. L., Katzfey, J., Gao, X., Wu, J., Hong, S.-Y., Wang, Y., and Sasaki, H. (2016). Building Asian climate change scenario by multi-regional climate models ensemble. Part II: mean precipitation. *International Journal of Climatology*, 36(13), 4253-4264.
- Lucas-Picher, P., Somot, S., Déqué, M., Decharme, B., and Alias, A. (2012). Evaluation of the regional climate model ALADIN to simulate the climate over North America in the CORDEX framework. *Climate Dynamics*, 41(5-6), 1117-1137.
- Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One*, 13(3), e0194889.
- Mearns, L., McGinnis, S., Korytina, D., and Arritt, R. (2017). The NA-CORDEX dataset, version 1.0. NCAR Climate Data Gateway, Boulder CO.
- MECP. (2018). Preserving and Protecting our Environment for Future Generations: A Made-in-Ontario Environment Plan. Retrieved from <https://www.ontario.ca/page/made-in-ontario-environment-plan>

- Mellit, A., Pavan, A. M., and Benghanem, M. (2013). Least squares support vector machine for short-term prediction of meteorological time series. *Theoretical and Applied Climatology*, *111*(1), 297-307.
- Ministry of the Environment, C. a. P. (2011). Climate action: Adapting to change, protecting our future. *PIBS 8291*, 16 pp.
- Ministry of the Environment, C. a. P. (2018). A Made-in-Ontario Environment Plan.
- Mitra, A., Iyengar, G., Durai, V., Sanjay, J., Krishnamurti, T., Mishra, A., and Sikka, D. (2011). Experimental real-time multi-model ensemble (MME) prediction of rainfall during monsoon 2008: Large-scale medium-range aspects. *Journal of earth system science*, *120*(1), 27-52.
- Mouatadid, S., Easterbrook, S., and Erler, A. R. (2017). *A machine learning approach to non-uniform spatial downscaling of climate variables*. Paper presented at the 2017 IEEE International Conference on Data Mining Workshops (ICDMW).
- Neumann, J. E., Price, J., Chinowsky, P., Wright, L., Ludwig, L., Streeter, R., Jones, R., Smith, J. B., Perkins, W., Jantarasami, L., and Martinich, J. (2014). Climate change risks to US infrastructure: impacts on roads, bridges, coastal development, and urban drainage. *Climatic Change*, *131*(1), 97-109.
- Perera, A. H., Euler, D. L., and Thompson, I. D. (2011). *Ecology of a managed terrestrial landscape: patterns and processes of forest landscapes in Ontario*: UBC Press.

- Pour, S. H., Shahid, S., Chung, E.-S., and Wang, X.-J. (2018). Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmospheric Research*, 213, 149-162.
- Radhika, Y., and Shashi, M. (2009). Atmospheric temperature prediction using support vector machines. *International journal of computer theory and engineering*, 1(1), 55.
- Rozante, J., Moreira, D., Godoy, R., and Fernandes, A. (2014). Multi-model ensemble: technique and validation. *Geosci Model Dev Discuss*, 7(3), 2933-2959.
- Samouly, A. A., Luong, C. N., Li, Z., Smith, S., Baetz, B., and Ghaith, M. (2018). Performance of multi-model ensembles for the simulation of temperature variability over Ontario, Canada. *Environmental Earth Sciences*, 77(13).
- Samui, P. (2008). Slope stability analysis: a support vector machine approach. *Environmental Geology*, 56(2), 255-267.
- Sharda, R., and Patil, R. B. (1992). Connectionist approach to time series prediction: an empirical test. *Journal of Intelligent Manufacturing*, 3(5), 317-323.
- Shi, X., Chen, Z., Hao, W., Yeung, D. Y., and Woo, W. C. (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*. Paper presented at the International Conference on Neural Information Processing Systems.

- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). *Sequence to sequence learning with neural networks*. Paper presented at the Advances in neural information processing systems.
- Tang, J., Li, Q., Wang, S., Lee, D.-K., Hui, P., Niu, X., Gutowski, W. J., Dairaku, K., McGregor, J., Katzfey, J., Gao, X., Wu, J., Hong, S.-Y., Wang, Y., and Sasaki, H. (2016). Building Asian climate change scenario by multi-regional climate models ensemble. Part I: surface air temperature. *International Journal of Climatology*, 36(13), 4241-4252.
- Tangang, F. T., Hsieh, W. W., and Tang, B. (1998). Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *Journal of Geophysical Research: Oceans*, 103(C4), 7511-7522.
- Tripathi, S., Srinivas, V., and Nanjundiah, R. S. (2006a). Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*, 330(3-4), 621-640.
- Tripathi, S., Srinivas, V. V., and Nanjundiah, R. S. (2006b). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, 330(3-4), 621-640.
- Vapnik, V. (2013). *The nature of statistical learning theory*: Springer science & business

media.

Von Storch, H., Hewitson, B., and Mearns, L. (2000). *Review of empirical downscaling techniques. In Regional climate development under global warming.* Paper presented at the General technical report no. 4. Conference proceedings regclim spring meeting, Norway.

Wallach, D., Mearns, L. O., Ruane, A. C., Rötter, R. P., and Asseng, S. (2016). Lessons from climate modeling on the design and use of ensembles for crop modeling. *Climatic Change, 139*(3-4), 551-564.

Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A., and Yu, Q. (2018). Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *International Journal of Climatology, 38*(13), 4891-4902.

Wang, W., Hong, T., Xu, X., Chen, J., Liu, Z., and Xu, N. (2019). Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. *Applied Energy, 248*, 217-230.

Wang, X., Huang, G., Lin, Q., and Liu, J. (2014). High-resolution probabilistic projections of temperature changes over Ontario, Canada. *Journal of Climate, 27*(14), 5259-5284.

Wang, X., Huang, G., Lin, Q., Nie, X., Cheng, G., Fan, Y., Li, Z., Yao, Y., and Suo, M.

- (2013). A stepwise cluster analysis approach for downscaled climate projection – A Canadian case study. *Environmental Modelling & Software*, 49, 141-151.
- Wang, X., Huang, G., Liu, J., Li, Z., and Zhao, S. (2015). Ensemble Projections of Regional Climatic Changes over Ontario, Canada. *Journal of Climate*, 28(18), 7327-7346.
- Yoon, H., Hyun, Y., Ha, K., Lee, K.-K., and Kim, G.-B. (2016). A method to improve the stability and accuracy of ANN- and SVM-based time series models for long-term groundwater level predictions. *Computers & Geosciences*, 90, 144-155.
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., and Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, 561, 918-929.
- Zhang, X., Xiong, Z., Zhang, X., Shi, Y., Liu, J., Shao, Q., and Yan, X. (2015). Using multi-model ensembles to improve the simulated effects of land use/cover change on temperature: a case study over northeast China. *Climate Dynamics*, 46(3-4), 765-778.
- Zhao, H., and Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), 3586-3592.

5. Conclusions

This thesis investigated the performance of statistical and machine learning methods in generating multi-model ensembles for temperature downscaling. RCM outputs obtained from NA-CORDEX were used as inputs for the proposed ensembles.

Chapter 1 provided background information and Chapter 2 presented a review on climate downscaling methods and ensemble modeling techniques.

In Chapter 3, a preliminary study was carried out to investigate the applicability and performance of neural network models for temperature downscaling. Multi-layer Perceptron (MLP), Time-lagged Feed-forward Neural Network (TLFN) and Nonlinear Auto-Regressive Network with exogenous inputs (NARX) were applied to develop multi-model ensembles and the performance of the proposed ensembles were evaluated using a case study of Big Trout Lake in Ontario, Canada. The results showed that MLP, TLFN, and NARX are effective methods for downscaling daily mean temperature and had similar performances based on root mean square error (RMSE) and coefficient of determination (R^2). Neural network based ensembles outperformed individual RCMs and generated predictions with smaller fluctuations. The results provide a foundation to further apply machine learning based multi-model ensembles for downscaling in larger areas.

Chapter 4 investigated the performance of statistical and machine learning based multi-model ensembles for downscaling daily temperature. Two statistical methods (arithmetic

ensemble mean, EM, and Multiple Linear Regression, MLR) and two machine learning methods (Long Short-Term Memory, LSTM, networks and Support Vector Machine, SVM) were chosen for multi-model ensemble development for temperature downscaling. A case study of twelve meteorological stations across Ontario, Canada was conducted to evaluate the performance of the proposed ensembles. Both machine learning ensembles and statistical ensembles have shown high accuracy in downscaling daily temperature and those ensembles had similar performance.

While the downscaling methods introduced in Chapters 3 and 4 performed well for daily temperature, there are still some limitations. One limitation is that all proposed ensembles have difficulties predicting extreme temperatures outside the range of -10°C to 20°C . One reason is that neural networks tend to grasp the trend of the majority of data, instead of a small portion of extreme events. Considering extreme temperatures appear during a similar time period each year (winter and summer), further research could consider splitting the dataset based on season and train one ensemble for each season. Bias correction could also be employed for improving the prediction of extreme values. Moreover, the performance of the ensembles could be further improved by including more input data. Given that machine learning methods are data-driven, their performance relies heavily on the quantity and quality of the input data. Further research could incorporate more RCMs driven by different GCMs and consider other relevant input variables such as daily

maximum temperature and daily minimum temperature.

References

- Adger, W. N., Barnett, J., Brown, K., Marshall, N., and O'brien, K. (2013). Cultural dimensions of climate change impacts and adaptation. *Nature Climate Change*, 3(2), 112.
- Adger, W. N., Quinn, T., Lorenzoni, I., Murphy, C., and Sweeney, J. (2012). Changing social contracts in climate-change adaptation. *Nature Climate Change*, 3(4), 330-333.
- Aksornsingchai, P., and Srinilta, C. (2011). *Statistical downscaling for rainfall and temperature prediction in Thailand*. Paper presented at the Proceedings of the international multiconference of engineers and computer scientists.
- Anandhi, A., Srinivas, V. V., Kumar, D. N., and Nanjundiah, R. S. (2009). Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology*, 29(4), 583-603.
- Anderson, G. J., and Lucas, D. D. (2018). Machine Learning Predictions of a Multiresolution Climate Model Ensemble. *Geophysical Research Letters*, 45(9), 4273-4280.
- Aribarg, T., Kimpan, C., and Supratid, S. (2017, 15-18 Nov. 2017). *Use of CMIP3 and CMIP5 Climate Models to Simulate Change Discharge in the Chao Phraya River*

Basin. Paper presented at the 2017 21st International Computer Science and Engineering Conference (ICSEC).

Ben Alaya, M. A., Zwiers, F., and Zhang, X. (2019). Evaluation and Comparison of CanRCM4 and CRCM5 to Estimate Probable Maximum Precipitation over North America. *Journal of Hydrometeorology*, 20(10), 2069-2089.

Bouwer, L. M. (2013). Projections of future extreme weather losses under changes in climate and exposure. *Risk Anal*, 33(5), 915-930.

Calzadilla, A., Rehdanz, K., Betts, R., Falloon, P., Wiltshire, A., and Tol, R. (2013). Climate change impacts on global agriculture. *Climatic Change*, 120(1-2), 357-374.

Cannon, A. J., and Whitfield, P. H. (2002). Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. *Journal of Hydrology*, 259(1), 136-151.

Casson, R. J., and Farmer, L. D. (2014). Understanding and checking the assumptions of linear regression: a primer for medical researchers. *Clin Exp Ophthalmol*, 42(6), 590-596.

Cavazos, T., Luna-Niño, R., Cerezo-Mota, R., Fuentes-Franco, R., Méndez, M., Pineda Martínez, L. F., and Valenzuela, E. (2019). Climatic trends and regional climate models intercomparison over the CORDEX-CAM (Central America, Caribbean, and Mexico) domain. *International Journal of Climatology*.

- Chen, S., Billings, S. A., and Grant, P. M. (1990). Non-linear system identification using neural networks. *International Journal of Control*, 51(6), 1191-1214.
- Clemen, R. T., and Murphy, A. H. (1986). Objective and subjective precipitation probability forecasts: some methods for improving forecast quality. *Weather Forecasting*, 1(3), 213-218.
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Coulibaly, P., Dibike, Y. B., and Anctil, F. (2005). Downscaling Precipitation and Temperature with Temporal Neural Networks. *Journal of Hydrometeorology*, 6(4), 483-496.
- Deng, Z., Liu, J., Qiu, X., Zhou, X., and Zhu, H. (2017). Downscaling RCP8.5 daily temperatures and precipitation in Ontario using localized ensemble optimal interpolation (EnOI) and bias correction. *Climate Dynamics*, 51(1-2), 411-431.
- Dibike, Y. B., and Coulibaly, P. (2006). Temporal neural networks for downscaling climate variability and extremes. *Neural Networks*, 19(2), 135-144.
- Diro, G. T., Sushama, L., Martynov, A., Jeong, D. I., Versegny, D., and Winger, K. (2014). Land-atmosphere coupling over North America in CRCM5. *Journal of Geophysical Research: Atmospheres*, 119(21), 11,955-911,972.
- Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N. (2005). The rationale behind the

- success of multi-model ensembles in seasonal forecasting — II. Calibration and combination. *Tellus A: Dynamic Meteorology and Oceanography*, 57(3), 234-252.
- Duhan, D., and Pandey, A. (2014). Statistical downscaling of temperature using three techniques in the Tons River basin in Central India. *Theoretical and Applied Climatology*, 121(3-4), 605-622.
- Eerola, K. J. H. N. (2006). About the performance of HIRLAM version 7.0. *51*, 93-102.
- Fyfe, J. C., von Salzen, K., Cole, J. N. S., Gillett, N. P., and Vernier, J. P. (2013). Surface response to stratospheric aerosol changes in a coupled atmosphere-ocean model. *Geophysical Research Letters*, 40(3), 584-588.
- Gaitan, C. F., Hsieh, W. W., Cannon, A. J., and Gachon, P. (2013). Evaluation of Linear and Non-Linear Downscaling Methods in Terms of Daily Variability and Climate Indices: Surface Temperature in Southern Ontario and Quebec, Canada. *Atmosphere-Ocean*, 52(3), 211-221.
- Goosse, H., Barriat, P.-Y., Loutre, M.-F., and Zunz, V. (2010). *Introduction to climate dynamics and climate modeling*: Centre de recherche sur la Terre et le climat Georges Lemaître-UCLouvain.
- Goyal, M. K., and Ojha, C. S. P. (2012). Downscaling of surface temperature for lake catchment in an arid region in India using linear multiple regression and neural networks. *International Journal of Climatology*, 32(4), 552-566.

Grimm, N. B., Groffman, P., Staudinger, M., and Tallis, H. (2016). Climate change impacts on ecosystems and ecosystem services in the United States: process and prospects for sustained assessment. In *The US National Climate Assessment* (pp. 97-109): Springer.

Hagedorn, R., DOBLAS-REYES, F. J., and Palmer, T. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting–I. Basic concept. *Tellus A*, 57(3), 219-233.

Harrison, M., Palmer, T., Richardson, D., Buizza, R., and Petroliaigis, T. (1995). *Joint ensembles from the UKMO and ECMWF models*. Paper presented at the ECMWF Seminar Proceedings: Predictability.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L., and Goodess, C. M. (2006). Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology*, 26(10), 1397-1415.

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.

Hua, W., Chen, H., Sun, S., and Zhou, L. (2015). Assessing climatic impacts of future land use and land cover change projected with the CanESM2 model. *International Journal of Climatology*, 35(12), 3661-3675.

- Huth, R. (2004). Sensitivity of Local Daily Temperature Change Estimates to the Selection of Downscaling Models and Predictors. *Journal of Climate*, 17(3), 640-652.
- Huth, R., Kliegrová, S., and Metelka, L. (2008). Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? *International Journal of Climatology*, 28(4), 465-477.
- IPCC. (2018). *Global warming of 1.5°C*. Retrieved from <https://www.ipcc.ch/sr15/>
- Jalota, S., Vashisht, B., Sharma, S., and Kaur, S. (2018). *Understanding Climate Change Impacts on Crop Productivity and Water Balance*: Academic Press.
- Jeong, D. I., St-Hilaire, A., Ouarda, T., and Gachon, P. (2012). A multivariate multi-site statistical downscaling model for daily maximum and minimum temperatures. *Climate Research*, 54(2), 129-148.
- Jeong, D. I., St-Hilaire, A., Ouarda, T. B. M. J., and Gachon, P. (2011). Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. *Stochastic Environmental Research and Risk Assessment*, 26(5), 633-653.
- Khalili, M., Van Nguyen, V. T., and Gachon, P. (2013). A statistical approach to multi-site multivariate downscaling of daily extreme temperature series. *International Journal of Climatology*, 33(1), 15-32.
- Kharin, V. V., and Zwiers, F. W. (2002). Climate Predictions with Multimodel Ensembles.

Journal of Climate, 15(7), 793-799.

Krishnamurti, T., Kishtawal, C., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C.

E., Gadgil, S., and Surendran, S. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, 285(5433), 1548-1550.

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E.,

Gadgil, S., and Surendran, S. (2000). Multimodel Ensemble Forecasts for Weather and Seasonal Climate. *Journal of Climate*, 13(23), 4196-4216.

Kronenberg, R., Barfus, K., Franke, J., and Bernhofer, C. (2013). On the Downscaling of

Meteorological Fields Using Recurrent Networks for Modelling the Water Balance in a Meso-Scale Catchment Area of Saxony, Germany. *Atmospheric and Climate Sciences*, 03(04), 552-561.

Laprise, R. (2008). Regional climate modelling. *Journal of Computational Physics*, 227(7),

3641-3666.

Lemmen, D. S., Warren, F. J., Lacroix, J., and Bush, E. (2008). From impacts to adaptation:

Canada in a changing climate. *Government of Canada, Ottawa, ON*.

Leung, L. R., Mearns, L. O., Giorgi, F., and Wilby, R. L. J. B. o. t. A. M. S. (2003). Regional

climate research: needs and opportunities. 84(1), 89-95.

Li, Q., Wang, S., Lee, D.-K., Tang, J., Niu, X., Hui, P., Gutowski, W. J., Dairaku, K.,

McGregor, J. L., Katzfey, J., Gao, X., Wu, J., Hong, S.-Y., Wang, Y., and Sasaki, H.

(2016). Building Asian climate change scenario by multi-regional climate models ensemble. Part II: mean precipitation. *International Journal of Climatology*, 36(13), 4253-4264.

Lucas-Picher, P., Wulff-Nielsen, M., Christensen, J. H., Aðalgeirsdóttir, G., Mottram, R., and Simonsen, S. B. (2012). Very high resolution regional climate model simulations over Greenland: Identifying added value. *Journal of Geophysical Research: Atmospheres*, 117(D2).

Mahmood, R., and Babel, M. S. (2012). Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. *Theoretical and Applied Climatology*, 113(1-2), 27-44.

Martynov, A., Sushama, L., Laprise, R., Winger, K., and Dugas, B. (2012). Interactive lakes in the Canadian Regional Climate Model, version 5: the role of lakes in the regional climate of North America. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1).

McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.

Miksovsky, J., and Raidl, A. J. N. P. i. G. (2005). Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures. *12(6)*, 979-991.

- Misra, S., Sarkar, S., and Mitra, P. (2017). Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and Applied Climatology*, 134(3-4), 1179-1196.
- Mouatadid, S., Easterbrook, S., and Erler, A. R. (2017). *A machine learning approach to non-uniform spatial downscaling of climate variables*. Paper presented at the 2017 IEEE International Conference on Data Mining Workshops (ICDMW).
- Naughton, M. P., Henderson, A., Mirabelli, M. C., Kaiser, R., Wilhelm, J. L., Kieszak, S. M., Rubin, C. H., and McGeehin, M. A. (2002). Heat-related mortality during a 1999 heat wave in Chicago. *American journal of preventive medicine*, 22(4), 221-227.
- Neumann, J. E., Price, J., Chinowsky, P., Wright, L., Ludwig, L., Streeter, R., Jones, R., Smith, J. B., Perkins, W., Jantarasami, L., and Martinich, J. (2014). Climate change risks to US infrastructure: impacts on roads, bridges, coastal development, and urban drainage. *Climatic Change*, 131(1), 97-109.
- Plummer, D., Caya, D., Frigon, A., Côté, H., Giguère, M., Paquin, D., Biner, S., Harvey, R., and De Elia, R. (2006). Climate and climate change over North America as simulated by the Canadian RCM. *Journal of Climate*, 19(13), 3112-3132.
- Rahmstorf, S., and Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, 108(44), 17905-17909.

- Rajagopalan, B., Lall, U., and Zebiak, S. E. (2002). Categorical Climate Forecasts through Regularization and Optimal Combination of Multiple GCM Ensembles. *Monthly Weather Review*, 130(7), 1792-1811.
- Robertson, A. W., Lall, U., Zebiak, S. E., and Goddard, L. (2004). Improved Combination of Multiple Atmospheric GCM Ensembles for Seasonal Prediction. *Monthly Weather Review*, 132(12), 2732-2744.
- Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornbluh, L., and Manzini, E. (2003). The atmospheric general circulation model ECHAM 5. PART I: Model description.
- Salman, A. G., Heryadi, Y., Abdurahman, E., and Suparta, W. (2018). Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. *Procedia Computer Science*, 135, 89-98.
- Samouly, A. A., Luong, C. N., Li, Z., Smith, S., Baetz, B., and Ghaith, M. (2018). Performance of multi-model ensembles for the simulation of temperature variability over Ontario, Canada. *Environmental Earth Sciences*, 77(13).
- Samuelsson, P., Jones, C. G., Will'En, U., Ullerstig, A., Gollvik, S., Hansson, U., Jansson, E., Kjellstro"m, C., Nikulin, G., and Wyser, K. (2016). The Rossby Centre Regional Climate model RCA3: model description and performance. *Tellus A: Dynamic Meteorology and Oceanography*, 63(1), 4-23.

- Sanders, F. (1963). On subjective probability forecasting. *Journal of Applied Meteorology*, 2(2), 191-201.
- Schoof, J. T., and Pryor, S. C. (2001). Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *International Journal of Climatology*, 21(7), 773-790.
- Scinocca, J. F., Kharin, V. V., Jiao, Y., Qian, M. W., Lazare, M., Solheim, L., Flato, G. M., Biner, S., Desgagne, M., and Dugas, B. (2016). Coordinated Global and Regional Climate Modeling*. *Journal of Climate*, 29(1), 17-35.
- Snell, S. E., Gopal, S., and Kaufmann, R. K. (2000). Spatial interpolation of surface air temperatures using artificial neural networks: Evaluating their use for downscaling GCMs. *Journal of Climate*, 13(5), 886-895.
- Srinivas, V. V., Basu, B., Nagesh Kumar, D., and Jain, S. K. (2014). Multi-site downscaling of maximum and minimum daily temperature using support vector machine. *International Journal of Climatology*, 34(5), 1538-1560.
- Stern, N. J. N. N. (2016). Economics: Current climate models are grossly misleading. *Nature News*, 530(7591), 407.
- Strandberg, G., Barring, L., Hansson, U., Jansson, C., Jones, C., Kjellström, E., Kupiainen, M., Nikulin, G., Samuelsson, P., and Ullerstig, A. (2015). *CORDEX scenarios for Europe from the Rossby Centre regional climate model RCA4*: SMHI.

- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). *Sequence to sequence learning with neural networks*. Paper presented at the Advances in neural information processing systems.
- Tang, B., Hsieh, W. W., Monahan, A. H., and Tangang, F. T. (2000). Skill Comparisons between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures. *13*(1), 287-293.
- Tran Anh, D., Van, S. P., Dang, T. D., and Hoang, L. P. (2019). Downscaling rainfall using deep learning long short-term memory and feedforward neural network. *International Journal of Climatology*, *39*(10), 4170-4188.
- Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., and Sheffield, J. (2014). Global warming and changes in drought. *Nature Climate Change*, *4*(1), 17-22.
- Tripathi, S., Srinivas, V. V., and Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, *330*(3-4), 621-640.
- Undén, P., Rontu, L., Jarvinen, H., Lynch, P., Calvo Sánchez, F. J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., and García-Moya, J. A. (2002). HIRLAM-5 scientific documentation.
- von Salzen, K., Scinocca, J. F., McFarlane, N. A., Li, J., Cole, J. N. S., Plummer, D.,

- Versegny, D., Reader, M. C., Ma, X., Lazare, M., and Solheim, L. (2013). The Canadian Fourth Generation Atmospheric Global Climate Model (CanAM4). Part I: Representation of Physical Processes. *Atmosphere-Ocean*, 51(1), 104-125.
- Wallach, D., Mearns, L. O., Ruane, A. C., Rötter, R. P., and Asseng, S. (2016). Lessons from climate modeling on the design and use of ensembles for crop modeling. *Climatic Change*, 139(3-4), 551-564.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., Fromentin, J.-M., Hoegh-Guldberg, O., and Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416(6879), 389.
- Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A., and Yu, Q. (2018). Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *International Journal of Climatology*, 38(13), 4891-4902.
- Wang, H. M., Chen, J., Xu, C. Y., Chen, H., Guo, S., Xie, P., and Li, X. (2019a). Does the weighting of climate simulations result in a better quantification of hydrological impacts? *Hydrol. Earth Syst. Sci.*, 23(10), 4033-4050.
- Wang, J., Wang, J., Shi, Y., Zhou, H., and Liao, L. (2019b). A Recursive Update Model for Estimating High-Resolution LAI Based on the NARX Neural Network and MODIS Times Series. *Remote Sensing*, 11(3).

- Wang, X., Huang, G., Lin, Q., Nie, X., and Liu, J. (2015a). High-resolution temperature and precipitation projections over Ontario, Canada: a coupled dynamical-statistical approach. *Quarterly Journal of the Royal Meteorological Society*, 141(689), 1137-1146.
- Wang, X., Huang, G., Liu, J., Li, Z., and Zhao, S. (2015b). Ensemble Projections of Regional Climatic Changes over Ontario, Canada. *Journal of Climate*, 28(18), 7327-7346.
- Weigel, A. P., Liniger, M., and Appenzeller, C. (2008). Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology*, 134(630), 241-260.
- Wen, T.-H., Gasic, M., Mrksic, N., Su, P.-H., Vandyke, D., and Young, S. (2015). Semantically conditioned lstm-based natural language generation for spoken dialogue systems. *arXiv:1508.01745*.
- Wheeler, T., and Von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341(6145), 508-513.
- Xu, L., Chen, N., Zhang, X., Chen, Z., Hu, C., and Wang, C. (2019). Improving the North American multi-model ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. *Climate Dynamics*, 53(1-2), 601-615.

Zhai, Y., Huang, G., Wang, X., Zhou, X., Lu, C., and Li, Z. (2018). Future projections of temperature changes in Ottawa, Canada through stepwise clustered downscaling of multiple GCMs under RCPs. *Climate Dynamics*, 52(5-6), 3455-3470.