Statistical Modeling of Daily Precipitation Process in the Context of Climate Change

By

Chenglong You (Yusif)

Department of Civil Engineering

McGill University

Montreal, Quebec, Canada

October 2024

A thesis submitted to the Graduate Postdoctoral Studies Office in partial fulfilments of requirements of the degree of Master of Science

© Chenglong You (Yusif) 2024

Abstract

Information on the variability of precipitation process is essential for the planning, design and management of various water resources systems. Furthermore, recent assessment reports on climate change have indicated a worldwide increase in the frequency of extreme storm events for the late 20th century because of global warming. Consequently, research on developing innovative approaches for limiting and adapting climate change impacts on water infrastructures is highly critical due to the substantial investments involved. Global Climate Models (GCMs) have been commonly used in various studies for assessing these potential impacts. However, outputs from these GCMs (generally greater than 200 km) are considered too coarse and hence are not suitable for climate change impact studies at a given site or over a catchment area. As a result, several downscaling techniques have been proposed to downscale these GCM outputs to the precipitation series at a given location of interest. Nevertheless, there is still no general agreement about which downscaling method is the best approach for describing accurately the observed precipitation characteristics at a given site in the climate change context, depending mainly on the study objectives and the climatology of the study area. The present study is therefore carried out in order to develop appropriate methods for improving the accuracy of precipitation estimation at a local site in the context of a changing climate. This study therefore proposes a new statistical model, herein referred to as SDGAM, using the Generalized Additive Models (GAM) to address the shortcomings of existing downscaling methods. The feasibility and accuracy of the proposed new approach were evaluated using the observed daily precipitation records available at two rain-gauge stations located in Quebec Province, and the National Center for Environmental Prediction (NCEP) re-

analysis data that are interpolated for two GCMs (Canadian CanESM2 and UK HadCM3). Results of this numerical application have indicated that the proposed SDGAM model was able to describe well many features of the daily precipitation process, including its amounts, occurrence frequency, intensity, and extremes. In addition, it has been demonstrated that the suggested SDGAM model could provide more accurate results than the popular Statistical Downscaling Model (SDSM) in the modeling of the daily precipitation process based on both numerical and graphical performance criteria. Finally, the proposed SDGAM can generate daily precipitation series for future periods under different climate change scenarios: RCP2.6, RCP4.5 and RCP8.5 for CanESM2, as well as A2 and B2 for HadCM3. These generated precipitation series are useful for various climate change impact studies in practice.

Résumé

Information sur la variabilité de précipitation est essentielle pour la planification, la conception et la gestion de divers systèmes de ressources hydriques. De plus, les rapports récents sur i'évaluation des impacts du changement climatique ont indiqué l'augmentation de la fréquence des événements d'orage extrêmes due au réchauffement global pour le tard du 20ième siècle. Par conséquent, les recherches sur le développement des approches innovatrices pour limiter et s'adapter aux impacts du changement climatique pour les structures hydrauliques sont hautement importantes à cause des investissements considérables concernés. En particulier, les modèles du climat global (MCG) avaient été fréquemment utilisés dans plusieurs d'études pour évaluer ces impacts potentiels. Toutefois, ces modèles fournissent des résultats aux très grandes résolutions (généralement plus grandes que 200 km) et alors ne sont pas appropriés pour les études d'impacts en un point ou à l'échelle d'un bassin versant. Plusieurs techniques de mise en échelle ont été alors proposées pour relier les sorties de ces modèles globaux aux précipitations en un point choisi. Mais il n'y a pas jusqu'à date un consensus sur le choix de la meilleure méthode qui est capable de représenter correctement les caractéristiques de précipitation en un point dans un contexte du changement climatique, dépendant des buts spécifiques des études et de la climatologie particulière de la région étudiée. La présente étude a alors pour objet d'élaborer des méthodes appropriées pour améliorer la précision d'estimation de précipitation en un point donné dans un contexte de changement climatique. Cette étude propose donc un nouveau modèle statistique, ci-après appelé SDGAM, en utilisant la méthode de modélisation additive généralisée (Generalized Additive Modeling, GAM) pour améliorer les faiblesses des méthodes de mise en échelle existantes.

On avait évalué la faisabilité et la précision de ce nouveau modèle en utilisant les données observées des séries des précipitations journalières disponibles en deux stations météorologiques au Québec et les données de réanalyse du National Center for Environmental Prediction (NCEP) qui sont interpolées pour deux MCG (CanESM2 du Canada et HadCM3 de UK). Les résultats de cette application numérique ont indiqué que le modèle SDGAM proposé sont capable de représenter correctement plusieurs propriétés statistiques et physiques du processus de précipitation journalière, comprenant la fréquence d'apparition, d'intensité et d'extrêmes. De plus, on avait démontré que le modèle SDGAM proposé peut fournir des résultats plus précis que les résultats du modèle populaire Statistical Downscaling Model (SDSM) en se basant sur divers critères numériques et graphiques de performance. Finalement, on avait effectué les projections des précipitations journalières pour les périodes dans le futur en utilisant le modèle SDGAM pour divers scénarios de changement climatique : RCP2.6, RCP4.5 et RCP8.5 pour CanESM2, et également A2 and B2 pour HadCM3. Ces projections seront utiles pour les études d'impacts en pratique.

Acknowledgements

I would like to show my profound and sincere gratitude to my supervisor, Prof. Van-Thanh-Van Nguyen for his guidance in my research work, his encouragement in overcoming difficulties, and the inspiration he gives me to gain wisdom in life.

I also express my sincere appreciation to my co-supervisor, Prof. Luis Mirando-Moreno. It is my honor to participate in his climate change and transportation research project. In addition, I wish to express my deep appreciation to my seniors, Hoang Lam Nguyen and Truong Huy Nguyen for their immense helps when I started this research work.

Furthermore, I am sincerely thankful to my friend Naheed Haidari, who has accompanied and encouraged me throughout the writing of my thesis and has witnessed my achievements in obtaining these research results. My thanks also extend to my friends Ziqi Hao and Xiaowen Han, without whom my life in Montreal would not have been as enriched and colorful.

Last but not least, I express my eternal love to my family. My parents always encourage me to study abroad and provide their financial support for my graduate study. My aunts and uncles always care about my life and studies. My sister always provides me with mental support. Without their help, it would be difficult for me to finish this research.

Table of Contents

Contents

List of Figures

List of Tables

List of Abbreviations

AMP	Annual Daily Maximum Precipitation
ANN	Artificial Neural Network
ASD	Automated Statistical Downscaling Model
BLUE	Best Linear Unbiased Estimator
CanAM4	Canadian Atmospheric Model version 4
CanESM2	Second-generation Canadian Earth System Model
CanRCM4	Canadian Regional Climate Model version 4
CDD	Maximum Consecutive Dry Days
DD	Dynamical Downscaling Method
ESD	Empirical/Statistical Downscaling Method
GAM	Generalized Additive Models
GCM	Global Climate Model
HadCM3	Hadley Centre Coupled Model, Version 3
LARS-WG	Long Ashton Research Station Weather Generator
ML	Machine Learning
MSE	Mean Squared Error
NAO	North Atlantic Oscillation Index
NCEP	National Center for Environmental Prediction
PRCP1	Proportion of Wet Days
Precp_m	Average of Daily Precipitation
Precp_sd	Standard Deviation of Daily Precipitation
Prec90p	90^{th} Percentile of Wet Day Amount

RCM	Regional Climate Model
RCP	Representative Concentration Pathway
RMSE	Root Mean Square Error
SDSM	the Statistical Downscaling Model
SDGAM	the Statistical Downscaling Model using GAM
SDII	Simple Daily Intensity Index
SSE	Sum of Squared Errors
SSP	Shared Socioeconomic Pathways
SST	Sea Surface Temperatures
SVM	Support Vector Machines
TAP	Total Annual Precipitation
VIF	Variance Inflation Factor
WGEN	the Weather Generator

1 Introduction

1.1 Statement of Problems

Investments in water infrastructure become increasingly crucial across various sectors. Such investments ensure the security and sustainability of harmonious coexistence between humans and the natural world, providing a foundation for future development and environmental protection. In Canada, multiple water infrastructure projects showcase the benefits of such investments, including the Chaudière-Appalaches Water Infrastructure Projects in Quebec (Government of Canada, 2020), the Site C Dam in British Columbia (BC Hydro, 2023), and the Lake Manitoba and Lake St. Martin Outlet Channels Project in Manitoba (Government of Manitoba, 2024). These infrastructures reduce the vulnerability of humans and the natural environment, but at the same time could make them more vulnerable to climate extremes, due to the lack of consideration of what might occur when the design is exceeded (Nguyen, 2022). Additionally, scientists suggest that there has been a global increase in the frequency of extreme weather events since the late 20th century. This trend, attributed to global warming, is expected to continue through the 21st century (Legg, 2021). Thus, it becomes imperative to integrate climate change into the design, planning, and management of water infrastructure projects to ensure that these systems are reliable enough to handle future challenges.

Recently, climate change has been recognized as having a profound impact on the hydrological cycle at different temporal and spatial scales (Nguyen and Giorgi, 2022). Precipitation is a vital component of the hydrologic cycle, serving as the primary mechanism through which water is transferred from the atmosphere

back to the Earth's surface. Reliable precipitation projection enables effective and robust planning, operation, and risk management for water infrastructure projects. Nowadays, Global Climate Models (GCMs) and Regional Climate Models (RCMs) can capture the main climate patterns at either global scale or regional scale. Furthermore, by simulating climate responses under various greenhouse gas emissions and policy scenarios, these climate models can predict potential climate trends and corresponding impacts. However, GCMs and RCMs typically operate at coarse spatial resolutions (larger than 200 km for GCMs and 20 to 50 km for RCMs) that limit their direct application in impact studies at a local scale. Hence, Empirical/Statistical downscaling (ESD) methods have been widely used for linking the large-scale predictors given by GCMs or RCMs to the observed precipitation series at a local site (Nguyen and Giorgi, 2022).

Among various Empirical/Statistical Downscaling methods, the Statistical Downscaling Model (SDSM) stands out as a widely adopted tool. SDSM has proven to be relatively effective across diverse geographical regions and climate types. This popularity is evidenced by its extensive application in numerous scientific studies (Wilby & Dawson, 2013). SDSM applies multiple linear regression techniques. However, multiple linear regression assumes a linear relationship between predictands and predictors, which may not adequately reflect the nonlinear relationships between observed precipitation and climate model outputs. This limitation necessitates the development of improved downscaling approaches that can capture the potential nonlinear relationships between climate model outputs and local precipitation.

1.2 Objectives of the Research

In view of the above-mentioned issues, this study develops a new statistical downscaling model, herein referred to as SDGAM, which relies on the Generalized Additive Models (GAM). This GAM approach allows for the modeling of the interactions between the predictors and the predictand through smooth functions, thus capturing the nonlinear relationships between these variables. In addition, GAM supports link functions, which can handle binary or other types of predictands, thereby facilitating a broader application range. Therefore, the proposed SDGAM in this study could provide a more accurate and more flexible approach to describing the relationships between large-scale climate model outputs and local-scale precipitation, thereby enhancing the reliability of precipitation projections in the future. More specifically, this study aims at the following objectives:

i. To propose a suitable method for selecting the most significant predictors for statistical downscaling models.
ii. To develop a new statistical downscaling model based on GAM that can capture the nonlinear relationship between large-scale predictors and local-scale precipitation.
iii. To propose a systematic evaluation procedure to assess the performance of various statistical downscaling methods.
iv. To generate reliable future precipitation series under different climate scenarios for high-quality climate change impact assessment studies in practice.

1.3 Organization of the Thesis

The current thesis is organized into five chapters. Chapter 1 provides the statement of problems and the main objectives of this research. Chapter 2 provides a literature review of previous works related to the downscaling methods. Chapter 3 introduces the

proposed SDGAM, also providing a detailed explanation of the mathematical principles behind GAM and an overview of the research data. Chapter 4 presents a comparative analysis between SDGAM and SDSM, as well as the precipitation projections under various climate scenarios. Finally, Chapter 5 presents the main conclusions of this research and the recommendations for further studies.

2 Literature Review

2.1 Overview of Downscaling Methods

GCMs have been extensively used to understand our planet's climate system and support policy decisions (Flato et al., 2014). In recent years, the reliability of these models has been significantly improved compared with those from the early 1990s (Nguyen, 2022). Recent GCMs can reasonably well describe the climate system at continental and hemispheric spatial scales (Legg, 2021). Despite significant progress, these models are still unable to provide reliable results for regional climate impact studies. The reasons for this limitation include:

1. Coarse Resolution: GCMs typically have a coarse spatial resolution, often around 100 to 300 kilometers. This level of resolution is insufficient for regional studies as it fails to capture the finer details and variability of local weather and climate processes. (Maraun et al., 2010).
2. Original Purpose: GCMs were not originally developed for climate change impact studies; they were initially developed to understand the climate system. Therefore, they are not well-suited for simulating regional hydrological variables at the catchment scale and specific locations (Maraun et al., 2010).
3. Inconsistency of Outputs: The outputs from different GCMs are inconsistent at the same regional scale, which complicates their utility for regional analyses (Arnbjerg-Nielsen et al., 2013).

To address these issues, the 'downscaling' method has emerged, which refines the GCMs' output to provide detailed and locally relevant climate predictions at a finer spatial resolution.

Generally, downscaling methods can be grouped into three categories:

1. Dynamical Downscaling Method (DD): This method involves the explicit solving of the physical dynamics of regional climate system (Mearns et al., 2003).
2. Empirical/Statistical Downscaling Method (ESD): This method establishes empirical/statistical relationships between coarse-scale predictors (e.g., GCM outputs) and local-scale predictands (e.g., precipitation) (Wilby et al., 2004).
3. Machine Learning Based Downscaling Method: This method keeps pace with the rapid advancements in Machine Learning (ML), and are increasingly becoming a new category of downscaling approaches (Vandal et al., 2019).

2.2 Types of Downscaling Methods

- *Dynamical Downscaling (DD)*

Dynamical Downscaling (DD) is a technique used to derive fine-scale climate information from coarser data. Generally, three different DD approaches have been used for climate change impact studies (Mearns et al., 2003):

1. Operating a regional-scale limited-area model using coarse GCM data as the boundary conditions.
2. Conducting global-scale experiments with high-resolution GCMs using coarse GCM data as initial conditions.
3. Using a variable resolution global model with the highest resolution over the area of interest.

The most common scheme of DD is to run a regional-scale limited-area model with the coarse GCM outputs as boundary conditions (Artlert & Chaleeraktrakoon, 2013). This method is typically referred to as Regional Climate Models (RCMs).

RCMs can capture resolutions ranging from 20-50 km, making their outputs ideally suited for integration with hydrologic models to assess the effects of climate change on hydrologic regimes. Additionally, RCMs can be used to assess the impacts of regional external forcings, such as changes in terrestrial ecosystems or atmospheric chemistry (Nguyen & Giorgi, 2022). While RCMs offer numerous advantages, they also come with several acknowledged limitations (Mearns et al., 2003). These limitations can be categorized as follows:

1. One primary limitation is the substantial demand for computing resources associated with RCMs, which restricts both the number of experiments that can be conducted and the duration of climate simulations.
2. The climate ensembles generated by RCMs are sensitive to the choice of boundary conditions.
3. For some hydrological application studies, it is necessary to further downscale the output from RCMs to even smaller scales.

- *Empirical/Statistical Downscaling (ESD)*

Empirical/Statistical Downscaling (ESD) methods are based on statistical relationships between large-scale climate predictors such as vorticity and humidity, along with local predictands such as temperature and precipitation (Pharasi, 2006). ESD methods can be classified into three categories: weather typing approaches (Bárdossy, 1997; Goodess & Palutikof, 1998); stochastic weather generators (Richardson & Wright, 1984); and regression methods (Kilsby et al., 1998; Wilks & Wilby, 1999; Harpham & Wilby, 2005; Hessami et al., 2008; Tareghian & Rasmussen, 2013). In general, three common assumptions are required for ESD methods (Nguyen & Giorgi, 2022):

1. The surface local predictands are a function of synoptic forcing.
2. The GCM used for deriving downscaled relationships is valid at the considered scale.
3. The derived relationship remains valid under climate change.

Weather typing methods involve linking the local surface variables with a subjectively defined weather pattern classification (Pharasi, 2006). Predictands are determined based on the corresponding conditional probability distributions of its daily weather pattern classification. However, this method is often criticized for its subjective nature to define the classification. Additionally, the changes in predictands derived from weather patterns often do not align with the changes predicted by the host GCM. Nevertheless, this method still retains its appeal because it provides a tangible understanding of the impact of complex climate systems on local weather patterns.

Rather than depending on predetermined weather types, stochastic weather generators simulate all predictands based on stochastic models, incorporating random variability to reflect natural fluctuations in weather patterns. Stochastic weather generators were pioneered by the introduction of the first-order Markov Chain renewal process (Gabriel & Neumann, 1962). Jones et al. (1993) extended the techniques by employing multiple-order Markov Chain renewal processes allowing for the simulation with memory characteristics. The most popularly used stochastic weather generators are WGEN (Richardson & Wright, 1984) and LARS-WG (Semenov & Barrow, 1997). WGEN models precipitation occurrence using a two-state, first-order Markov chain. On wet days, precipitation amounts are simulated with a gamma distribution. Meanwhile, temperature and radiation components are simulated using first-order autoregression, which is conditioned on

precipitation. LARS-WG is a stochastic weather generator widely used in agriculture. It employs a series of statistical distributions to simulate weather events. Stochastic weather generators can accurately reproduce many observed statistics, such as the mean, median and interquartile range. However, a significant challenge faced by stochastic weather generators is establishing the linkage between the parameters of these stochastic models and large-scale climate variables (e.g., GCM outputs).

Regression methods aim to establish regression relationships between large-scale predictors (e.g., GCM outputs) and local-scale predictands (e.g., precipitation and temperature). Examples of these methods includes basic linear regression (Kilsby et al., 1998; Wilby et al., 2002), polynomial regression (Conway et al., 1996), canonical correlation analysis (Karl et al., 1990), principal component regression (Benestad et al., 2015), and quantile regression (Tareghian & Rasmussen, 2013). More sophisticated regression downscaling methods, such as expanded downscaling (Bürger, 1996) and multifractal approaches (Xu et al., 2015), have also been used for modeling precipitation. Regression methods are computationally undemanding and capable of generating climate sets for conducting risk and uncertainty analysis, which highlights their overall effectiveness in climate research. Moreover, ESD methods directly use observational data from local weather stations, thereby ensuring that predictions of future climate are consistent with local station data to some extent. However, a notable limitation is the stationary assumption of the regression parameters, suggesting that the statistical relationships developed for the current climate are presumed to hold for the future climate as well (Yeo, 2014). Section 2.2 will discuss in more detail specific issues related to regression methods.

- *Machine Learning Approaches*

As machine learning (ML) becomes increasingly popular, downscaling techniques based on ML, such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), are also gaining traction. However, up to now, it has been found that the direct application of these state-of-the-art ML methods to downscaling has not provided a significant improvement over traditional regression methods (Vandal et al., 2019). These machine learning models are much more like a black box. Even with good calibration results, it is difficult to provide physical explanations and ensure applicability to other situations. Despite these challenges, machine learning methods have significant potential in effectively handling the complex and non-linear relationships of climate data and in the integration of various types of predictors (Serifi et al., 2021).

All downscaling approaches carry some underlying assumptions (Wilby et al., 2004). These assumptions highlight the need for careful selection of downscaling methods tailored to specific climatic conditions and research objectives. In conclusion, there is no general agreement on which method is most suitable approach for describing accurately the observed precipitation characteristic for a given study site, depending mainly on specific research objectives and specific climatology of the study area (Nguyen & Nguyen, 2008).

2.3 Regression-Based Downscaling Methods

After reviewing various downscaling methods, our focus will now shift to regression-based downscaling methods. These methods are widely recommended due to their practicality and effectiveness in climate assessment applications. For instance, the Statistical Downscaling Model (SDSM) (Wilby et al., 2002), as a regression

downscaling method, has been considered one of the most popular, as recommended by the Intergovernmental Panel on Climate Change (IPCC) (Nguyen, 2023). Next, we will explore the mechanisms, applications, and challenges of regression downscaling methods in the face of climate change.

2.3.1 Climate Predictors for Precipitation Downscaling

The relationship between precipitation and climate predictors is not always stationary. Nevertheless, the predictors selected as explanatory variables must significantly explain one or more features of the precipitation process, both from statistical and physical perspectives. The strongest predictor for the precipitation occurrence is the total shear vorticity (Z), which is highly correlated on a monthly or seasonal basis across different regions' sites (Gachon, 2005). Mean sea-level pressure and the components of the geostrophic wind flow (zonal and meridional) also play a crucial role in precipitation occurrence (Wilby, 1998). Precipitation amounts are more dependent on specific humidity at various levels besides the large-scale predictors already discussed. Low-frequency predictors, such as the North Atlantic Oscillation Index (NAO) and sea surface temperatures (SST), also assist researchers in better understanding long trends and periodicities in precipitation due to their inherent inter-decadal variability (Wilby, 1998; Conway et al., 1996). In terms of the selection of predictors, the Statistical Downscaling Model (SDSM) (Wilby & Dawson, 2004) uses partial correlation coefficients to choose predictors. Hessami (Hessami et al., 2008) implemented two methods based on backward stepwise regression and partial correlation coefficients in their automated statistical downscaling model (ASD) to select the predictors.

Large-scale atmospheric predictors inherently exhibit redundancy due to their physical meanings, and this redundancy cannot be resolved even through stepwise regression. This leads to multicollinearity and may cause researchers to eliminate certain explanatory variables which can be more successful in capturing the long-term climate (Wilby et al., 2004). Rotated principal components analysis (Cavazos & Hewitson, 2002; Choux, 2005; Benestad et al., 2015), priori correlation analysis (Wilby & Wigley, 2000) and ridge regression (Hessami et al., 2008) are explorations to solve these problems.

2.3.2 Key Characteristics, Advantages and Challenges

Autoregressive options are sometimes crucial for regression-based downscaling, as the precipitation occurrence and amounts exhibit temporal and spatial dependences. Temporally, the likelihood of rain on the following day not only depends on the large-scale atmospheric predictors but also on the weather conditions of the previous day. This dependency, informed by observational data and the physical mechanisms of precipitation, suggests that the probabilities of consecutive wet days (PII) or a dry day following a wet day (POI) offer a robust method for generating series of dry and wet days. Regarding precipitation amount, during wet periods, it is sometimes necessary to adjust the relationship between precipitation amounts and large-scale predictors on adjacent dates by incorporating lags of -1 or +1 (Wilby & Dawson, 2004). Spatially, the dependence among different weather stations is often overlooked, hence a multisite downscaling procedure is required to capture this spatial dependence (Khalili and Nguyen, 2017).

One of the well-known weaknesses of regression method is their tendency to underestimate the variance of daily precipitation.

Some previous studies have suggested inflating the variance of precipitation based on the predictors' variance itself (Karl et al., 1990; Huth, 1999; Gachon, 2005). Bürger (1996) introduced a canonical correspondence method termed "expanded downscaling" specifically designed to increase the variance. Von Storch (1999) introduced a 'randomized downscaling' approach to ensure the variance. The use of a 'Variance Inflation Factor' (VIF) to adjust the variance empirically has also been suggested by other studies (Wilby & Dawson, 2004; Hessami et al., 2008; Yeo et al., 2021).

Furthermore, the regression downscaling method for precipitation requires the prior determination of precipitation occurrence before the downscaling of precipitation amounts (Wilby et al., 2002). This approach has been included in the Statistical Downscaling Model (SDSM). The SDSM has successfully described the linkages between observational data with outputs from different GCMs. It is user-friendly, computation-economical, relatively reliable, and for these reasons, SDSM has become one of the most popular downscaling models. SDSM also exhibits global applicability: it is used extensively across North America, Europe, Asia and Africa. The model has been particularly prevalent in research in Canada, China, and the UK, with significant studies also stemming from areas that are relatively under-represented in downscaling activities, such as the South America and Australasia (Wilby & Dawson, 2013). Despite its widespread use and numerous advantages, SDSM is not without its limitations. Numerous studies have made further improvements based on the SDSM. Hessami (2008) applied ridge regression in the ASD model to address the issue of multicollinearity. Yeo et al. (2021) demonstrate through mathematical derivation that empirically increasing the variance of precipitation also raises the monthly mean precipitation. Therefore, a correction coefficient is added to correct the monthly

mean precipitation in the SDRain model. Furthermore, Yeo et al. (2021) and Khalili (Khalili et al., 2013) employed logistic regression to model the precipitation occurrence process, thus overcoming the issues by linear regression with binary variables. Nevertheless, these models still fail to capture the non-linear relationships between precipitation and large-scale climate predictors. Wilby et al. (1998) and George et al. (2016) attempted to use polynomial regression to describe nonlinear relationships, but higher-order polynomial regression models are susceptible to overfitting. This means that their polynomial regression models may perform well during the calibration period, but it might not perform as well beyond the calibration period. Additionally, a global polynomial may not effectively capture local variations or patterns present in the data due to its overfitting characteristic. Given these limitations, this thesis aims to propose a more effective and more robust downscaling method (SDGAM) based on the Generalized Additive Models (GAM).

GAM has been successfully employed in various water resources management studies. Villarini and Serinaldi (2012) used GAM to forecast seasonal precipitation in Romania. Jones et al. (2013) assessed the change of extreme daily precipitation in Northern Island using GAM. Laanaya et al. (2017) demonstrated that GAM outperforms logistic regression, residuals regression, and linear regression in modeling water temperature in the Sainte-Marguerite River catchment. These examples illustrate that GAM possesses exceptional capabilities in handling complex and nonlinear hydrological data, providing a strong justification for selecting GAM as the core of the statistical downscaling model proposed in this study.

3 Methodology

3.1 Constructing and Interpreting Generalized Additive Models (GAM)

3.1.1 From Linear Regression to GAM: an Evolution of Statistical Modeling Techniques

Traditionally, linear regression is used to model the linear relationship between a predictand (dependent variable) and one or more predictors (independent variable) as follows:

$$Y = \beta_0 + \sum_{j=1}^{m} \beta_j X_j + \varepsilon \tag{3-1}$$

where Y represents the predictand, X_j represents the predictors, β_0 is the intercept of the regression line, β_j is the coefficient of the linear regression model, and ε represents the error term. The error term accounts for the difference between the observed value and that predicted by the linear model. Multiple linear regression models rely on several key assumptions. Violating these assumptions can significantly compromise the accuracy of the model. Below are the key assumptions upon which multiple linear regression models are based on:

1. **Linearity:** The model assumes a linear relationship between the predictand and the combination of predictors. It cannot account for any nonlinear relationships.
2. **Independence of Errors:** It is assumed each observation's error should not be influenced by the errors of other observations.
3. **Normality of Errors and Homoscedasticity:** The residuals are assumed to follow a normal distribution with a mean of zero. This assumption validates that the least squares estimation

for linear regression is the Best Linear Unbiased Estimator (BLUE) and facilitates the various statistical test. Additionally, linear regression assumes the variance of the residuals is constant at all levels of the predictors.

4. **No Multicollinearity:** The model assumes that there is no high linear correlation between any of the predictors.

These assumptions ensure that linear regression is a simple yet accurate model. However, they also restrict its applicability in handling complex, real-world data where these assumptions do not hold. To address these limitations, Generalized Linear Model (GLM) were developed (Nelder & Wedderburn, 1972). The term 'generalized' refers to the ability of GLM to accommodate different types of predictands using various link functions and residual distributions. The equation of GLM can be expressed as follows:

$$g\big(E(Y|X)\big) = \beta_0 + \sum_{j=1}^{m} \beta_j X_j \tag{3-2}$$

where $E(Y|X)$ is the expected value of the predictand Y given the predictors X, $g(\cdot)$ is the link function, β_j are the model parameters, and X_j represents the predictors. GLM offers significant extensions to the assumptions of linear regression, providing flexibility to model a broader range of data types.

1. While linear regression assumes a linear relationship between predictors and the predictand, GLM incorporates link functions which connect the linear combination of predictors with the mean of predictand given the corresponding predictors. For instance, the logit link function is commonly used to model the binary predictand. Log link function is typically used to model non-negative integer predictand such as counts or rates.

2. GLM maintains the assumption of independent errors. However, for data with potential error correlation, GLM can be extended by incorporating autoregressive structures or similar correlation models to account for the dependence among observations.
3. Unlike linear regression, which requires normally distributed residuals and homoscedasticity, GLM is more flexible. This flexibility is primarily achieved by introducing link functions and conducting transformations. Furthermore, GLM can also be used in conjunction with Weighted Least Squares to handle heteroscedasticity.
4. Although GLM does not inherently solve multicollinearity, it can be combined with techniques such as ridge regression or principal component analysis to mitigate the effects of multicollinearity.

While GLM offers improvements over linear regression, it still only allows for the linear combination of predictors, which restricts its flexibility in describing nonlinear relationships. To overcome this limitation, Generalized Additive Models (GAM) (Wood, 2017) extends the GLM framework by using smooth functions on predictors. This structure not only retains all the advantages of GLM but also enhances the ability to handle nonlinear relationships. Additionally, unlike GLM, GAM allows for the modeling of interactions among predictors. A GAM that uses univariate splines function as smooth functions can be expressed as follows:

$$g\big(E(Y|X)\big) = \beta_0 + \sum_{j=1}^{m} f_j(X_j) \tag{3-3}$$

where $E(Y|X)$ is the expected value of the predictand Y given predictors X, $g(\cdot)$ is the link function, and $f_j(X_j)$ denotes the univariate spline function.

3.1.2 Understanding Smooth Function in GAM

Smooth functions are the core feature of GAM that enable flexible descriptions of nonlinear relationships between the predictand and predictors. B-spline functions are the most representative type of smooth functions. B-splines consist of a series of locally defined basis functions and can be expressed as the sum of weighted basis functions:

$$f(x) = \sum_{k=1}^{K} \beta_k b_k(x) \tag{3-4}$$

where $f(x)$ is a B-spline function, $b_k(x)$ is a basis function, β_k is a coefficient associated with basis functions, K is the number of knots. Basis functions are usually expanded by polynomial functions, as well as by sine and cosine functions. For polynomial expansions, $x^0 = 1$ is the constant term, x^1 is the linear term, x^2 is the quadratic term, x^3 is the cubic term and so forth. Each basis function affects the shape of the curve only within a localized range, ensuring the adjustments to local data do not impact other parts of the curve. In contrast, traditional polynomial fitting is a global fitting tool, meaning that minor changes anywhere within the entire data range can affect the entire output. Moreover, because B-splines are locally defined, they do not require high orders to fit complicated data and are less sensitive to noise in the data. B-splines adjust the curve by setting the number and placement of knots.

To prevent overfitting of the model, penalized maximum likelihood estimation is used to estimate the coefficient vectors β in the spline functions. Penalized maximum likelihood function can be expressed as follows:

$$L_P(\beta, f) = L(\beta, f) - \frac{1}{2}\sum_{j=1}^{J} \lambda_j \int [f_j''(x)]^2 dx \tag{3-5}$$

where $L_P(\beta, f)$ is the penalized maximum likelihood function, $L(\beta, f)$ is unpenalized likelihood function. λ_j is the smoothing parameter to define the trade-off between wiggliness (the degree of bending in the smooth functions) and closeness to the data. J represents the numbers of spline functions in a GAM. $\int [f''(x)]^2 dx$ is the integral of the square of the second derivative of the spline function, used to evaluate the wiggliness. The penalty term is often represented in matrix form in practical applications:

$$\int [f''(x)]^2 dx = \beta^T S \beta \tag{3-6}$$

where β is the coefficient vector, and S is the penalty matrix of the spline function $f(x)$. The dimensions of the penalty matrix S depend on the number of basis functions in the spline function. If $f(x)$ is represented through K basis functions, then S is a $K \times K$ matrix.

In conclusion, GAMs encompass a flexible class of models that allow for the estimation of nonlinear relationships. The R programming language provides strong support for implementing GAM through various packages. Among these, the 'mgcv' (Wood & Wood, 2015) package is particularly popular. SDGAM has been developed based on the 'mgcv' package.

3.2 Statistical Downscaling of Daily Precipitation Process (SDGAM): Data, Modeling, and Projection

3.2.1 Description of the Research Data

Quebec is located in eastern Canada, situated between latitudes 45°N and 62°N, and longitudes 57°W to 79°W. The province features diverse terrain: the northern part is predominantly

forest and tundra of the Canadian Shield; the central and western regions are traversed by the Appalachian Mountains; and the southern part consists of Saint Lawrence River Valley. This valley serves as a vital corridor connecting the two major cities—Montreal and Quebec City—and is not only the agricultural heart of Quebec but also a focal point of its economic and cultural activities. Extending from southwest to northeast, the Saint Lawrence River Valley is characterized by its relatively flat terrain, bordered by low mountains and hills. Due to the terrain's guiding effect, the valley frequently receives moist air masses from the southwest Great Lakes and the eastern Atlantic Ocean, resulting in higher precipitation, particularly increased snowfall in winter, compared to other regions in the province.

To evaluate the accuracy and feasibility of the proposed SDGAM model, a case study is conducted using observed daily precipitation data and the NCEP Reanalysis data. The observed data come from the MONTREAL/PIERRE ELLIOTT TRUDEAU INTL A and QUEBEC/JEAN LESAGE INTL weather stations, which have been providing complete and reliable precipitation daily observations since 1953 (Table 3-1). The NCEP Reanalysis data provide a comprehensive and corrected set of atmospheric and oceanic observations from around the world (Kalnay et al., 1996). In this study, two GCMs from Canada (CanESM2) and United Kingdom (HadCM3) are employed. CanESM2, developed by the Canadian Centre for Climate Modelling and Analysis, has contributed significantly to the IPCC Fifth Assessment Report (AR5). HadCM3, on the other hand, has been widely used since 1999 for climate research, especially in regions like the UK and Canada. The integration of NCEP data into the CanESM2 and HadCM3 grids results in the creation of tailored datasets, referred to as NCEP-CanESM2 data and NCEP-HadCM3 data. NCEP-CanESM2 data covers the period from 1961 to 2005, with the calibration period spanning

from 1961 to 1981, and the validation period from 1982 to 2005. NCEP-HadCM3 covers the period from 1961 to 2001, with the calibration period also from 1961 to 1981, and the validation period from 1982 to 2001. Figure 3-1 displays the location of rain-gauge stations S1 (Montreal/Pierre Elliot Trudeau International Airport) and S2 (Quebec/Jean Lesage International Airport), as well as their corresponding GCMs grids in North America.

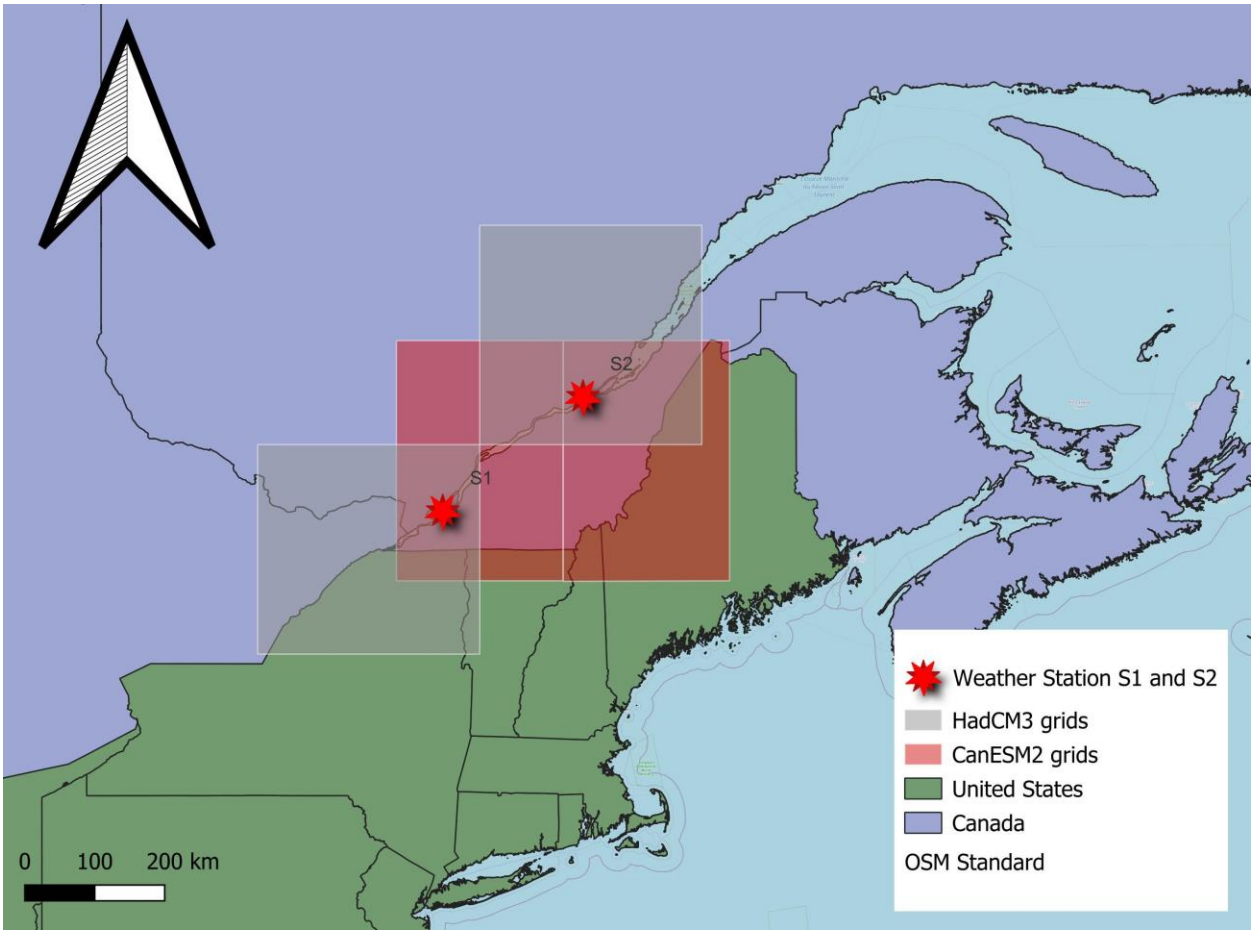

Figure 3-1. Locations of rain-gauge stations and GCMs grids in this study

Table 3-1. Information on rain-gauge stations in Quebec, Canada

Code	Site Name	Latitude	Longitude	Elevation	Staring
S1	MONTREAL/PIERRE ELLIOTT TRUDEAU INTL A	45.47°	-73.75°	36.00m	1951
S2	QUEBEC CITY/JEAN LESAGE INTL	46.81°	-71.38°	60.00m	1951

3.2.2 Selection of Large-scale Atmospheric Predictors

The selection of appropriate predictors in statistical downscaling is crucial, as a reasonable selection of predictors significantly enhances the model's sensitivity and adaptability to climate characteristics. Table 3-2 presents the 31 large-scale predictors from NCEP re-analysis data recommended by SDSM. The table categorizes each predictor by its level of measurement (surface, 500hPa, 800hPa) and indicates the availability of NCEP-CanESM2 and NCEP-HadCM3 data on the Government of Canada's website.

Table 3-2. List of large-scale predictors of NCEP re-analysis data in the GCMs grid box

Large-scale predictors	Level of measurement			NCEP-CanESM2	NCEP-HadCM3
Downward Shortwave Radiation	surface			unavailable	unavailable
Surface lifted index	surface			unavailable	unavailable
Mean sea level pressure	surface			available	available
Geostrophic airflow velocity	surface	500hPa	800hPa	available	available
Zonal velocity component	surface	500hPa	800hPa	available	available
Meridional velocity component	surface	500hPa	800hPa	available	available
Vorticity	surface	500hPa	800hPa	available	available
Wind direction	surface	500hPa	800hPa	available	available
Divergence	surface	500hPa	800hPa	available	available
Potential temperature		N/A		unavailable	unavailable
Precipitable water		N/A		unavailable	unavailable
Precipitation total	surface			available	unavailable
Relative humidity	surface	500hPa	800hPa	unavailable	available
Specific humidity	surface	500hPa	800hPa	available	surface available
Air temperature	surface			available	available
Geopotential height		500hPa	800hPa	available	available

Backward stepwise regression (McCuen, 2016) and partial correlation coefficients (Afifi et al., 2003) are widely used to select large-scale predictors. Backward stepwise regression begins by including all predictors in the model, then progressively eliminates the least significant ones until all the remaining predictors are statistically significant. The partial F-test, which is used to remove a predictor from a set including q predictors is:

$$F = \frac{(R_q^2 - R_{q-1}^2)(n-q-1)}{1-R_q^2} \tag{3-7}$$

where n is the sample size, R_q and R_{q-1} are the correlation coefficients of predictands with q and q-1 predictors respectively. If F is greater than critical F-value, the predictor is retained in the set; otherwise, the predictor is removed.

Partial correlation measures the correlation between two variables while controlling other variables. The partial correlation between the predictand (indicated by i) and j-th predictor while controlling for the third predictor k, is given by:

$$R_{ij,k} = \frac{R_{ij} - R_{ik}R_{jk}}{\sqrt{(1-R_{ik}^2)(1-R_{jk}^2)}} \tag{3-8}$$

where R_{ij} is the correlation coefficient between the predictand (indicated by i) and j-th predictor, R_{ik} is the correlation coefficient between the predictand (indicated by i) and k-th predictor, R_{jk} is the correlation coefficient between the j-th predictor and k-th predictor. In partial correlation method, if the t-value is greater than critical t-value, the predictor is

retained in the set; otherwise, it is removed. The t-value associated with partial correlation can be expressed as follows:

$$t = \frac{R(\sqrt{n-k-1})}{\sqrt{1-R^2}} \tag{3-9}$$

where R is the partial correlation coefficient, n is the sample size, k is the number of predictors. When controlling for more than one variable, a recursive algorithm is implemented to calculate high-order partial correlation coefficients.

In this study, scatter plots and spline function plots are used to detect nonlinear relationships. Backward stepwise regression and partial correlation coefficients are adopted to select predictors that exhibit significant relationships with predictand. To compare the performance of two statistical downscaling models, SDSM and SDGAM used the same predictors. More specifically, the large-scale predictors for CanESM2 are Zonal velocity component near the surface, Meridional velocity component near the surface, Vorticity at 500 hPa, 500 hPa geopotential height, Near surface specific humidity, and Precipitation total. The predictors for HadCM3 are Meridional velocity component near the surface, Vorticity near the surface, Meridional velocity component at 500 hPa, Zonal velocity component at 850 hPa, 500 hPa geopotential height, and Near surface relative humidity. The spline functions giving CanESM2 in S2 are displayed in Figure 3-2.

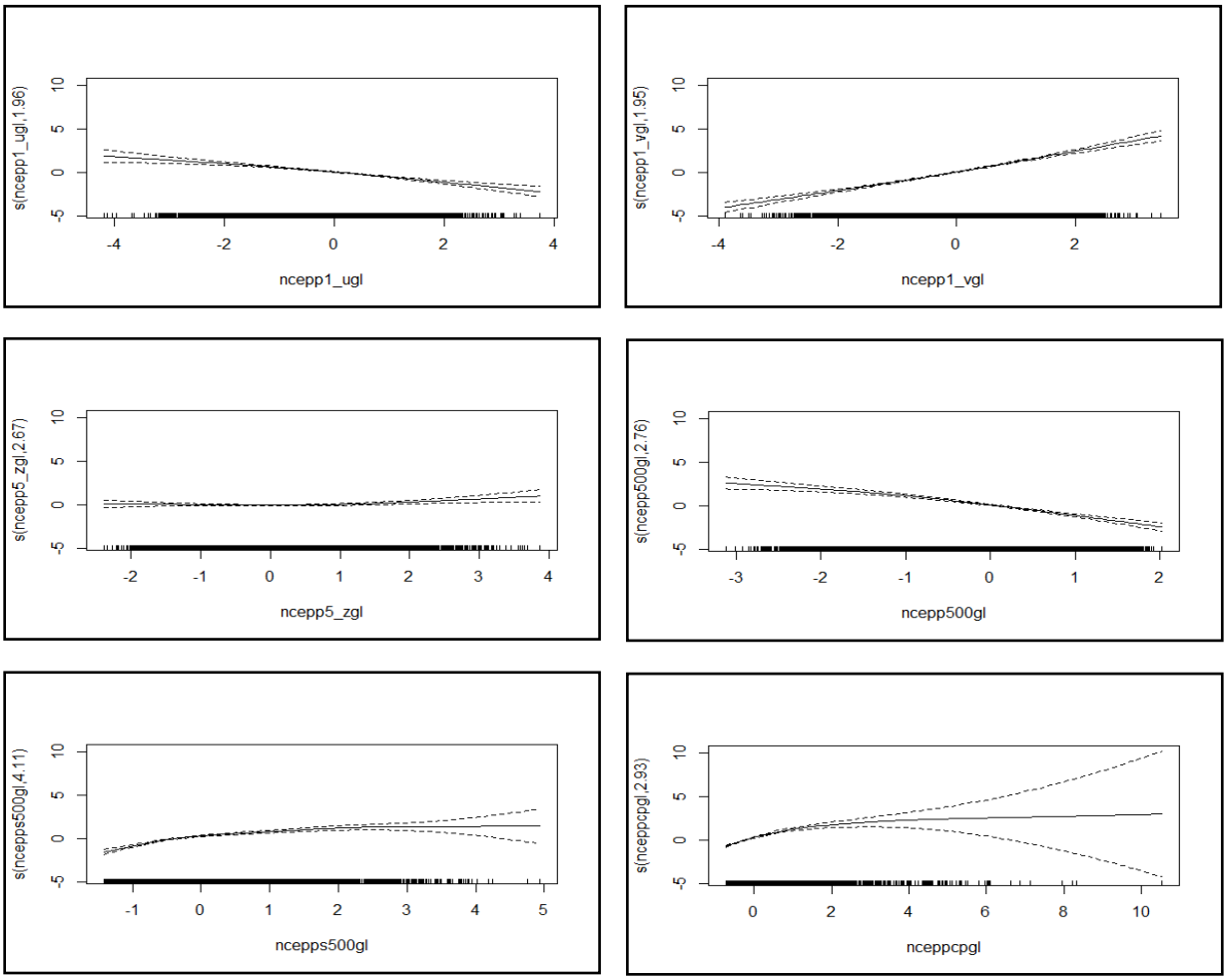

Figure 3-2. Plots of spline functions of each predictor in SDGAM precipitation occurrence process for S2 giving CanESM2

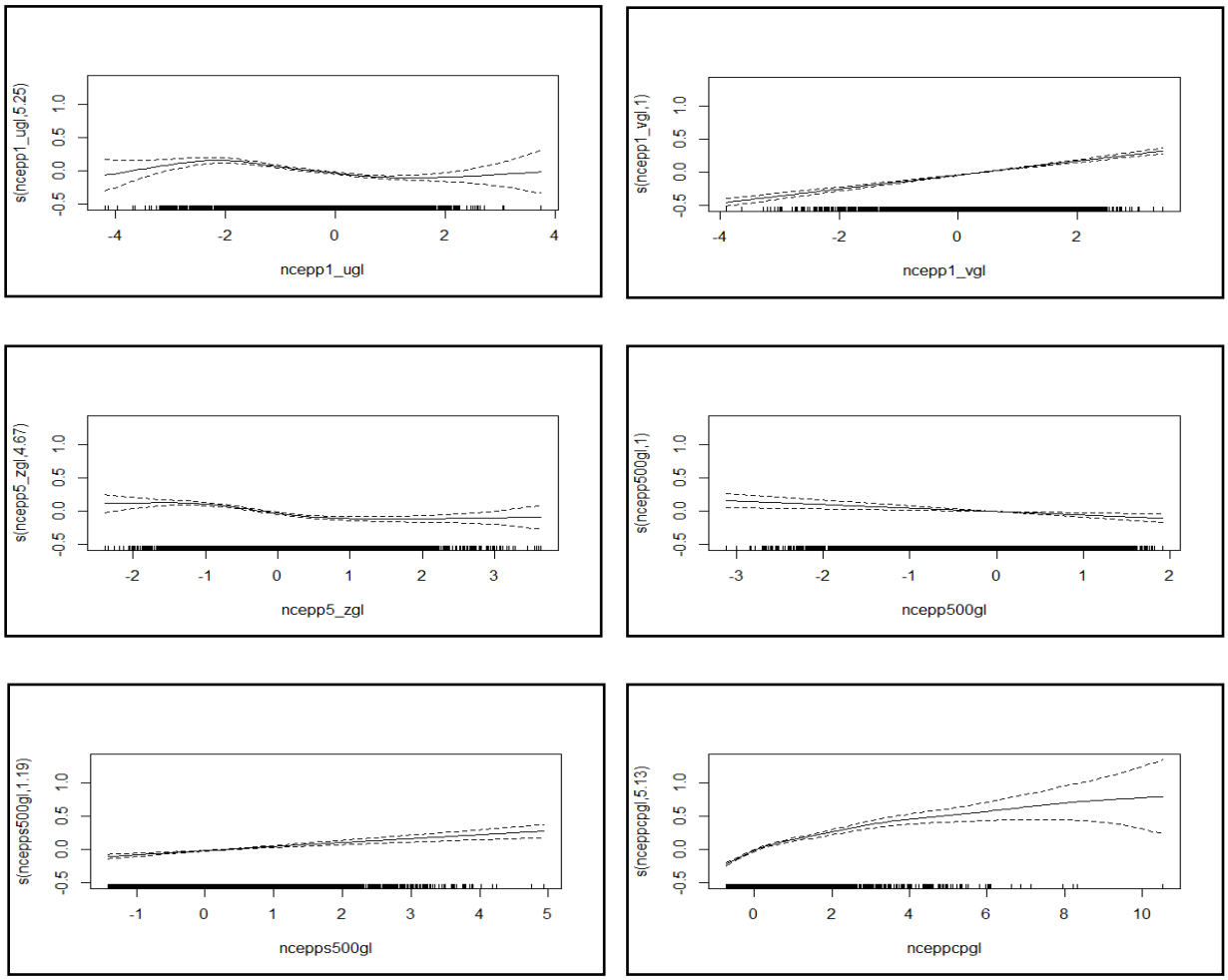

Figure 3-3. Plots of spline functions of each predictor in SDGAM precipitation amount process for S2 giving CanESM2

3.2.3 Modeling of the Daily Precipitation Process - SDGAM

3.2.3.1 Daily Precipitation Occurrence Process

Let O_i denote the random variable that represents the occurrence of daily precipitation, where $O_i = 0$ indicates a dry day and $O_i = 1$ indicates a wet day. The probability of a wet day (on day i) is π_i and the probability of a dry day is $1 - \pi_i$. Given that the precipitation occurrence on a certain day follows a Bernoulli Distribution, the variance can be expressed as: $Var(O_i) = \pi_i(1 - \pi_i)$. Clearly, the variance changes with different π, which violates the homoscedasticity assumption required by multiple linear regression. Moreover, the residual in multiple linear regression can be expressed as $\varepsilon_i = O_i - (\alpha_0 + \sum \alpha_i X_i)$, which has a high risk to violate normal residual assumption. Last, π_i estimated by multiple linear models may fall outside the 0 to 1 range. Given that the precipitation occurrence is a binary variable, and the spline function plots demonstrate non-linear relationships, the precipitation occurrence process in SDGAM can be expressed as follows:

$$\hat{\pi}_i = \frac{e^{[\alpha_O + \sum_{j=1}^{n} f_{Oj}(X_{ij})]}}{1 + e^{[\alpha_O + \sum_{j=1}^{n} f_{Oj}(X_{ij})]}} \tag{3-10}$$

where f_{Oj} represents the spline function of precipitation occurrence process, X_j is the j-th large-scale predictor, α_O is the intercept, $\hat{\pi}$ is the deterministic part of wet day probability, and i represents the day i. Alternatively, by logit transformation, (3-11) can be expressed as:

$$\ln\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = \alpha_O + \sum_{j=1}^{n} f_{Oj}(X_{ij}) \tag{3-11}$$

A uniformly distributed number $r_i\,(0 \le r_i \le 1)$ is generated to determine whether day i is a wet or dry day. For instance, if $r_i \le \hat{\pi}_i$, precipitation occurs on the day i.

In the RStudio environment (Team, 2021), the 'mgcv' package is used to calibrate the precipitation occurrence process. Calibration is conducted monthly using precipitation records, along with NCEP reanalysis data, to estimate $\hat{\pi}_i$. The threshold for determining a wet day is 0.2 mm.

3.2.3.2 Daily Precipitation Amounts Process

Daily precipitation amount is a non-zero and right-skewed distributed random variable. Common methods for transforming such data to more closely approximate a normal distribution include the fourth root transformation, natural log transformation, and Box-Cox transformation (Box & Cox, 1964). Additionally, the link functions in GAM such as 'Gamma' and 'inverse.gaussian' can be employed to handle non-zero and right-skewed data. In this study, the fourth root transformation was adopted for precipitation amount, as recommended in SDSM manual (Wilby & Dawson, 2004), which often yields better results compared to other transformations. The precipitation amounts process in SDGAM can be expressed as follows:

$$R_i = \left(\alpha + \sum_{j=1}^{n} f_j(X_{ij}) + \eta_i\right)^4 \qquad (3\text{-}12)$$

where f_j represents the spline function of precipitation amount process, X_j is the j-th large-scale predictor, α is the intercept, R_i represents the daily precipitation amounts, η represents the random term, and i denotes the day i. The fourth power represents the fourth root re-transformation. The random term η can be defined as follows:

$$\eta = Z \cdot MSE \tag{3-13}$$

where Z is a normally distributed number with a mean of 0 and a standard deviation of VIF_R:

$$Z \sim N(0, VIF_R{}^2) \tag{3-14}$$

VIF_R provides a method to manually adjust the standard deviation of modeled precipitation amounts to align with that of observed precipitation amounts. VIF_R is set to 1 during model calibration and validation. Changing VIF_R will adjust the standard deviation of precipitation amount; however, due to the fourth-root retransformation, it will also affect the mean of the series. In this case, it is necessary to introduce a coefficient C_{AK} to compensate for this effect.

MSE is the monthly Mean Squared Error, defined as follows:

$$MSE = \frac{SSE}{n-k} = \frac{\sum(y_i - \check{y}_i)^2}{n-k} \tag{3-15}$$

where y_i is the observed precipitation amount after fourth root transformation, $\check{y}_i$ is the fitted precipitation amount data in the fourth root form, n is the number of observations, and k is the number of large-scale predictors.

In the RStudio environment, the 'mgcv' package is used to calibrate precipitation amounts. Calibration is conducted monthly, using observed precipitation amounts on wet days along with NCEP reanalysis data.

3.2.4 Precipitation Performance Indices

The performance of SDGAM was evaluated using graphical and numerical indices as detailed in Table 3-3. These indices were selected to represent the significant characteristics of the

precipitation process, including the average and variance of precipitation, the frequency of precipitation occurrence, the intensity of precipitation amounts, and extreme events.

Table 3-3. List of Precipitation Indices

Categories	Indices	Definition	Unit	Time scale
Basic variable	Precp_m	Average of precipitation	mm/day	Month
	Precp_sd	Standard deviation of precipitation	mm/day	Month
Frequency	PRCP1	Proportion of wet days (>0.2mm)	/	Month
Extreme	AMP	Annual daily max precipitation	mm	Year
	CDD	Max consecutive dry days	days	Season
	Prec90p	90^{th} percentile of wet day amount	mm/day	Season
Intensity	SDII	Simple daily intensity index (Mean precipitation at wet day)	mm/day	Season
Annual	TAP	Total annual precipitation	mm	Year

The Root Mean Square Error (RMSE) is employed to evaluate the performance of the proposed model, and is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum (SI_{model} - SI_{observed})^2} \tag{3-16}$$

where n is the sample size, SI represents the precipitation indices. A lower RMSE indicates a better fit between the model's prediction and the observed data, with an RMSE of 0 representing an ideal case where the model predictions perfectly match the observations.

The study uses box plots to show downscaling models' performance. The horizontal line in a box plot represents the median, the size of the box represents the interquartile range, the length of the whiskers extends to 1.5 times the interquartile range, and points beyond the whiskers represent outliers.

3.2.5 Future Precipitation Projection

3.2.5.1 Bias Correction between Two Types of Downscaled Precipitation

For atmospheric large-scale predictors, differences exist between GCMs historical data and NCEP reanalysis data. These discrepancies can primarily be attributed to the different methodologies used to generate the two datasets. GCMs simulate the climate system based on physical principles, whereas NCEP reanalysis data are produced by assimilating observed data from multiple sources into one climate model to provide a coherent and consistent climatic record. In the calibration and validation phases, SDGAM uses NCEP reanalysis data, which are closer to the real atmospheric circulation data. However, for generating future precipitation under different scenarios, SDGAM employs data from GCMs because GCMs possess forward-looking capabilities, whereas reanalysis data are retrospective. Therefore, a bias correction between the two types of downscaled precipitation—one from NCEP data and the other from GCM data—is necessary.

The bias correction is first applied to precipitation occurrence process. The GCM historical data from the calibration period are processed through SDGAM to derive the corresponding PRCP1, denoted as P_{GCM_his}. Similarly, the NCEP reanalysis data from the same period are processed through SDGAM to derive the corresponding PRCP1, denoted as P_{NCEP}. $\hat{\pi}_i$ is the daily probability of a wet day (deterministic part) downscaled from GCM historical data. The corrected daily precipitation probability (deterministic part) can be expressed as:

$$\hat{\pi}_{i\_corr} = \hat{\pi}_i - (P_{GCM_{his}} - P_{NCEP}) \tag{3-17}$$

It is essential to limit the range of $\hat{\pi}_{i\_corr}$ between 0 and 1. Then, r_i is a uniformly distributed random number and if $r_i \leq \hat{\pi}_{i\_corr}$ on a given day i, then precipitation occurs on that day.

The bias correction is then applied to precipitation amount process. Both the GCM historical data and NCEP reanalysis data from the calibration period are processed through SDGAM to derive their SDII, denoted as S_{GCM_his} and S_{NCEP}, respectively. R_i is the daily precipitation amount downscaled from GCM historical data. The corrected daily precipitation amount can be expressed as:

$$R_{i\_corr} = R_i \cdot \frac{S_{NCEP}}{S_{GCM\_his}} \tag{3-18}$$

If, during the validation period, the PRCP1 and SDII of the bias-corrected GCM downscaled precipitation still align with those of the NCEP downscaled precipitation, it indicates that the bias correction performs well when encountering unseen data.

3.2.5.2 Future Precipitation Projections under Different Scenarios

GCMs require specific scenarios as inputs to simulate different future climate conditions. These scenarios are based on various assumptions, including socio-economic development, population growth, energy usage and policy decisions. The primary types of scenarios include Emission Scenarios, Representative Concentration Pathways (RCPs), and Shared Socioeconomic Pathways (SSPs).

Emission scenarios are employed in HadCM3 to simulate the climate system's response to varying concentrations of greenhouse gases. The IPCC's Special Report on Emission Scenarios (SRES) includes scenarios such as A1, A2, B1, and B2. The A2 scenario portrays a world where countries develop independently with ongoing population growth and a focus on regional economies, whereas the B2 scenario depicts a world where both population and

economic growth are moderate, focusing on sustainable solutions (Nakicenovic et al., 2000). The RCPs are a new methodology introduced in the IPCC's Fifth Assessment Report (AR5) and are key inputs in CanESM2. RCPs describe four main pathways based on radiative forcing levels, including RCP2.6, RCP4.5, RCP6.0, and RCP8.5 (Pachauri et al., 2014). These RCP numbers indicate the potential radiative forcing values by the year 2100, measured in watts per square meter (W/m^2).

The study continues to use equations (3-17) and (3-18) for bias correction in future precipitation generation. Differently, in this case, $\hat{\pi}_i$ represents the daily probability of a wet day (deterministic part) downscaled from GCM future data, while R_i is the daily precipitation amount downscaled from GCM future data. For CanESM2, precipitation projections under the RCP2.6, RCP4.5 and RCP8.5 scenarios are generated. For HadCM3, precipitation projections under the A2 and B2 scenarios are generated. Each precipitation projections under a certain scenario consists of 100 sets.

This study also calculated precipitation indices derived from the regional climate model CanRCM4 and compare those with indices obtained from SDGAM. The parent model of CanRCM4 is CanAM4 (von Salzen et al., 2019), which forms the atmospheric component of CanESM2 (Arora et al., 2011). CanRCM4 uses a rotated latitude-longitude grid system as its coordinate system, with two common resolutions: 0.22° horizontal grid resolution (approximately 25 km) and 0.44° horizontal grid resolution (approximately 50 km). Precipitation series corresponding to the station grids can be directly extracted from CanRCM4. Lastly, the scheme of SDGAM is illustrated in Figure 3-3.

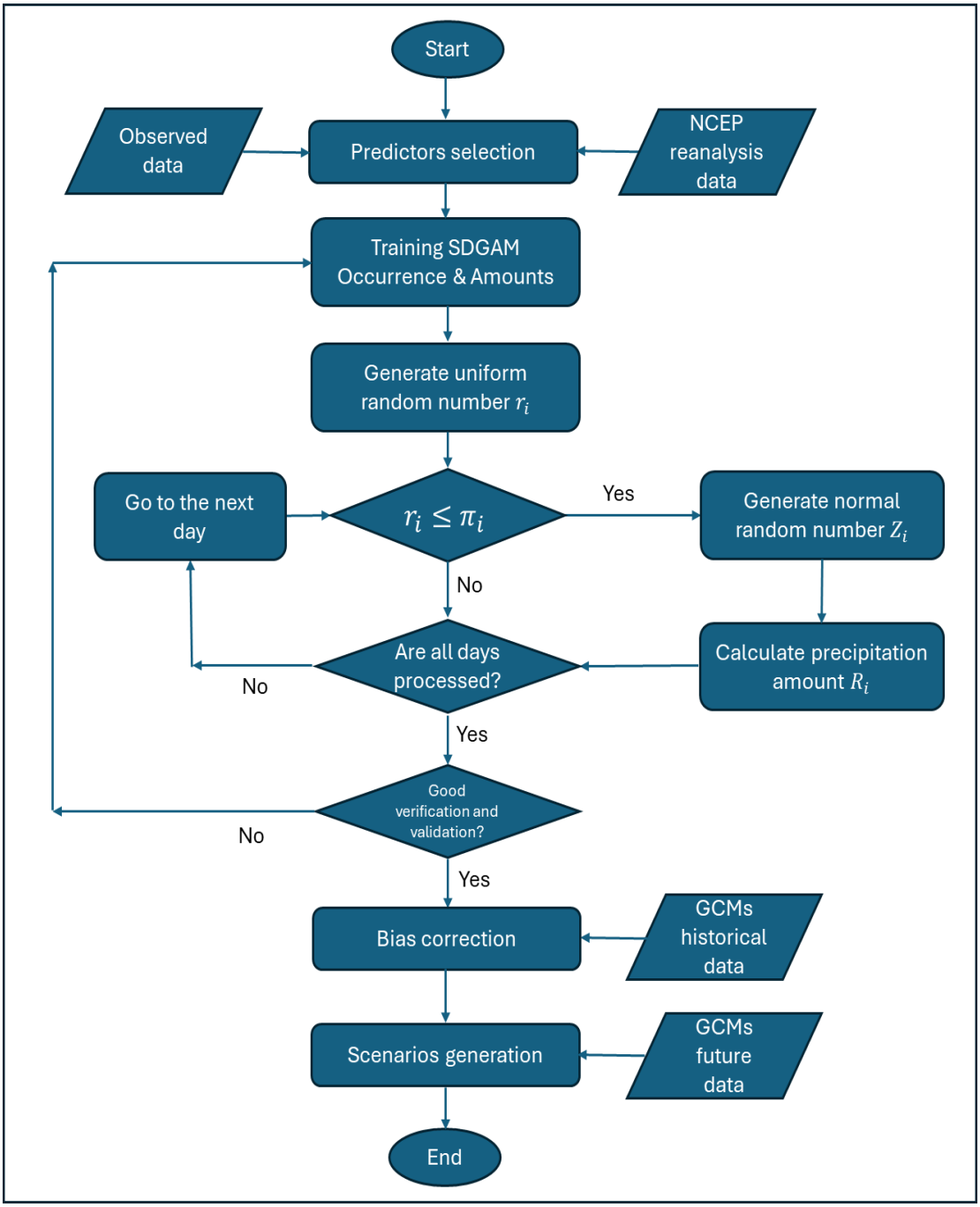

Figure 3-4. Scheme of SDGAM

4 Results

4.1 Verification and Validation of SDGAM and SDSM Results

4.1.1 Numerical Analysis

The feasibility and accuracy of the proposed SDGAM are assessed by numerical analysis using the different numerical performance criteria as described in the previous chapter. This numerical analysis was conducted using the 100 sets of daily precipitation series generated by the SDGAM. In addition, the performance of the SDGAM was also compared to the performance of the popular SDSM using the 100 sets of daily precipitation series generated by the SDSM. The RMSE values listed in the following tables represent the median value of the RMSEs calculated from the 100 sets of precipitation series for each index. Bold values in these tables denote the cases where the RMSE value of SDGAM is higher than that of SDSM.

More specifically, Tables 4-1 to 4-4 present the results for monthly mean of precipitation (Precp_m) for both stations S1 and S2 and giving CanESM2 and HadCM3. The RMSE values tend to be higher during the summer months for both SDGAM and SDSM, indicating more challenges in accurately modeling precipitation during these months. In addition, for nearly all cases, the SDGAM exhibits slightly lower RMSE values than the SDSM during both calibration and validation periods. This suggests that the SDGAM can provide more accurate estimation of the Precp_m than the SDSM.

Tables 4-5 to 4-8 show the results for monthly standard deviation of precipitation (Precp_sd). It can be observed that there is a trend of higher RMSE values during the summer and early

autumn months (June to September). For Montreal (S1) station when giving NCEP-CanESM2 and NCEP-HadCM3, and for Quebec City (S2) station when using NCEP-HadCM3, the SDGAM and SDSM perform comparably in capturing the variability of precipitation. However, for Quebec City (S2) station giving NCEP-CanESM2, the monthly standard deviations of precipitation given by SDGAM are noticeably closer to the observed values as compared to those from SDSM.

Tables 4-9 to 4-12 provide the RMSE values for monthly proportions of wet days (PRCP1). It was found that with the application of logistic transformation and spline functions, SDGAM consistently outperforms SDSM in nearly all months for both stations. In particular, the results for the validation period have clearly indicated the superior performance of SDGAM as compared to SDSM.

Tables 4-13 and 4-14 report the RMSE values for the maximum number of consecutive dry days (CDD) for each season. This index is regarded as one of the most difficult physical characteristics to be captured in the modeling of the daily precipitation process (Yeo, 2014). For Montreal (S1) station giving NCEP-CanESM2 and NCEP-HadCM3, SDGAM generally outperforms SDSM across all seasons. For Quebec City (S2) station giving NCEP-HadCM3, SDGAM consistently shows better performance than SDSM in all seasons except for winter. During the calibration period for Quebec City (S2) station giving NCEP-CanESM2, the RMSE values obtained with SDGAM are consistently lower than those obtained with SDSM across all seasons. However, in the validation period, the RMSE values derived from SDGAM exceed those from SDSM for all seasons.

Tables 4-15 and 4-16 indicates that during the calibration period, SDGAM more accurately reflects the actual 90^{th} percentile of rain day amount (Prec90p) compared to SDSM. In the validation

period, however, SDGAM and SDSM show comparable performance across both locations. Tables 4-17 and 4-18 show that SDGAM consistently yields lower RMSE values during the calibration period for both locations, suggesting it is generally more accurate in modeling the mean precipitation amounts on wet days. The performance is less consistent in the validation period, where specifically for Montreal (S1) station, SDGAM yields results that are closer to the observed SDII, whereas for Quebec City (S2) station, SDSM produces results that more closely match the observed SDII.

Table 4-19 provides the results of evaluation of annual daily maximum precipitation (AMP) for both calibration and validation periods. Both SDGAM and SDSM could provide a comparable performance in the estimation of the AMP values. Specifically, when giving NCEP-CanESM2 data, the AMP derived from SDGAM is closer to the observed values, while AMP derived from SDSM using NCEP-HadCM3 data more closely matches the observed values. Finally, Table 4-20 shows the results for the estimation of total annual precipitation (TAP). In general, SDGAM demonstrates better performance for both stations.

In summary, the proposed SDGAM can accurately describe various precipitation features for both calibration and validation periods for both Montreal (S1) station and Quebec City (S2) station. Across the indices of Precp_m, PRCP1, CDD, Prec90p, SDII, and TAP, the precipitation downscaled using SDGAM is closer to the observed precipitation than that downscaled using SDSM. For the indices Precp_sd and AMP, the performance of both SDGAM and SDSM is comparable.

Table 4-1 RMSEs of monthly mean of precipitation of Montreal (S1) station giving NCEP-CanESM2

Month	Montreal (S1) NCEP-CanESM2 Precp_m			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	0.65	0.81	0.94	1.04
Feb	0.74	0.84	0.78	0.86
Mar	0.75	0.77	0.95	0.96
Apr	0.95	1.05	1.16	1.18
May	**0.85**	**0.82**	**1.06**	**1.03**
Jun	1.16	1.21	**1.30**	**1.29**
Jul	**1.57**	**1.53**	**1.67**	**1.63**
Aug	**1.28**	**1.27**	**1.61**	**1.57**
Sep	1.69	1.74	**1.42**	**1.35**
Oct	0.95	0.99	**1.23**	**1.17**
Nov	0.98	1.01	1.28	1.30
Dec	0.76	0.89	0.88	0.91

Table 4-2 RMSEs of monthly mean of precipitation of Montreal (S1) station giving NCEP-HadCM3

Month	Montreal (S1) NCEP-HadCM3 Precp_m			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	0.87	0.88	**1.26**	**1.22**
Feb	0.93	0.94	0.84	0.88
Mar	0.86	0.98	**1.15**	**1.10**
Apr	0.99	1.01	1.27	1.30
May	0.90	0.91	1.07	1.15
Jun	**1.22**	**1.20**	**1.20**	**1.19**
Jul	1.59	1.73	1.43	1.44
Aug	1.29	1.33	**1.59**	**1.56**
Sep	1.84	1.86	1.32	1.35
Oct	1.01	1.08	1.18	1.18
Nov	1.00	1.07	1.37	1.45
Dec	0.92	0.94	1.13	1.27

Table 4-3 RMSEs of monthly mean of precipitation of Quebec City (S2) station giving NCEP-CanESM2

Month	Quebec City (S2) NCEP-CanESM2 Precp_m			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	1.09	1.33	0.90	0.99
Feb	1.01	1.19	0.87	0.97
Mar	0.96	1.02	**0.96**	**0.93**
Apr	0.99	1.08	1.28	1.30
May	1.00	1.04	1.47	1.53
Jun	1.57	1.59	1.62	1.62
Jul	1.52	1.63	**1.81**	**1.71**
Aug	1.78	1.80	**1.82**	**1.76**
Sep	1.71	1.80	1.61	1.79
Oct	1.01	1.06	1.08	1.30
Nov	1.05	1.16	1.12	1.19
Dec	1.00	1.20	1.25	1.37

Table 4-4 RMSEs of monthly mean of precipitation of Quebec City (S2) station giving NCEP-HadCM3

Month	Quebec City (S2) NCEP-HadCM3 Precp_m			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	1.15	1.19	**1.10**	**1.09**
Feb	1.12	1.16	**1.06**	**1.03**
Mar	1.03	1.10	**1.11**	**1.08**
Apr	**1.14**	**1.11**	1.31	1.34
May	**1.18**	**1.15**	1.46	1.54
Jun	1.58	1.63	1.68	1.73
Jul	1.82	1.82	**1.77**	**1.76**
Aug	**1.87**	**1.83**	**1.64**	**1.61**
Sep	1.81	1.91	**1.79**	**1.66**
Oct	1.14	1.19	1.19	1.20
Nov	1.11	1.19	1.34	1.43
Dec	1.18	1.25	**1.28**	**1.27**

Table 4-5 RMSEs of monthly standard deviation of precipitation of Montreal (S1) station giving NCEP-CanESM2

Month	Montreal (S1) NCEP-CanESM2 Precp_sd			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	1.59	2.45	2.09	2.42
Feb	1.86	2.07	1.85	2.02
Mar	**2.20**	**2.19**	**2.39**	**2.34**
Apr	2.23	2.48	**2.62**	**2.56**
May	**1.89**	**1.86**	**2.35**	**2.33**
Jun	2.85	2.98	**3.94**	**3.73**
Jul	3.31	3.32	**3.67**	**3.52**
Aug	**3.44**	**3.28**	**3.74**	**3.53**
Sep	**4.69**	**4.63**	3.95	3.95
Oct	2.15	2.40	**3.85**	**3.03**
Nov	2.17	2.51	3.75	3.98
Dec	2.03	2.75	1.93	2.61

Table 4-6 RMSEs of monthly standard deviation of precipitation of Montreal (S1) station giving NCEP-HadCM3

Month	Montreal (S1) NCEP-HadCM3 Precp_sd			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	**2.15**	**2.07**	**2.74**	**2.63**
Feb	**2.01**	**2.00**	**2.06**	**1.98**
Mar	2.35	2.37	**2.65**	**2.35**
Apr	**2.35**	**2.20**	**3.21**	**2.86**
May	**1.89**	**1.85**	2.65	2.67
Jun	2.90	2.92	**3.85**	**3.77**
Jul	3.33	3.44	**3.26**	**3.23**
Aug	3.28	3.29	**3.91**	**3.61**
Sep	4.88	5.02	**3.37**	**3.25**
Oct	2.21	2.23	**3.66**	**3.55**
Nov	**2.19**	**2.14**	3.86	4.06
Dec	**2.16**	**2.06**	2.21	2.36

Table 4-7 RMSEs of monthly standard deviation of precipitation of Quebec City (S2) station giving NCEP-CanESM2

Month	Quebec (S2) NCEP-CanESM2 Precp_sd			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	2.15	3.71	1.91	2.59
Feb	2.06	2.43	2.04	2.11
Mar	2.60	2.90	2.74	2.80
Apr	2.24	2.39	2.32	2.32
May	2.41	2.41	2.99	2.99
Jun	4.30	4.32	3.59	3.73
Jul	3.36	3.52	**3.80**	**3.72**
Aug	**3.77**	**3.66**	3.84	3.87
Sep	4.37	4.88	4.34	5.16
Oct	2.62	2.85	2.95	3.28
Nov	2.41	2.73	2.40	2.92
Dec	2.28	3.14	2.65	3.19

Table 4-8 RMSEs of monthly standard deviation of precipitation of Quebec City (S2) station giving NCEP-HadCM3

Month	Quebec (S2) NCEP-HadCM3 Precp_sd			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	**2.39**	**2.36**	**2.39**	**2.19**
Feb	**2.19**	**2.12**	**2.37**	**2.27**
Mar	2.99	3.05	**2.64**	**2.54**
Apr	**2.44**	**2.40**	**2.33**	**2.30**
May	2.53	2.58	2.96	2.96
Jun	4.23	4.35	3.89	3.89
Jul	**3.81**	**3.67**	**3.70**	**3.69**
Aug	**3.85**	**3.82**	**3.78**	**3.45**
Sep	4.71	4.84	**4.77**	**3.95**
Oct	2.98	2.98	2.67	2.72
Nov	2.51	2.55	**2.57**	**2.50**
Dec	**2.72**	**2.54**	**2.48**	**2.32**

Table 4-9 RMSEs of monthly proportion of wet days of Montreal (S1) station giving NCEP-CanESM2

Month	Montreal (S1) NCEP-CanESM2 PRCP1			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	0.098	0.117	0.111	0.114
Feb	0.105	0.111	0.111	0.113
Mar	0.099	0.118	0.090	0.093
Apr	0.089	0.102	0.109	0.124
May	0.104	0.107	0.097	0.112
Jun	0.091	0.096	0.110	0.114
Jul	0.104	0.112	**0.141**	**0.137**
Aug	0.113	0.113	0.125	0.129
Sep	0.103	0.124	0.113	0.117
Oct	0.104	0.110	0.103	0.107
Nov	0.099	0.109	0.094	0.106
Dec	0.090	0.107	0.128	0.137

Table 4-10 RMSEs of monthly proportion of wet days of Montreal (S1) station giving NCEP-HadCM3

Month	Montreal (S1) NCEP-HadCM3 PRCP1			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	0.105	0.112	**0.105**	**0.102**
Feb	0.106	0.109	0.094	0.104
Mar	0.110	0.112	0.094	0.102
Apr	0.098	0.112	0.119	0.120
May	**0.108**	**0.107**	0.115	0.122
Jun	0.102	0.111	0.126	0.127
Jul	0.104	0.115	**0.144**	**0.140**
Aug	0.108	0.115	**0.129**	**0.128**
Sep	0.112	0.127	0.110	0.115
Oct	0.114	0.122	0.107	0.113
Nov	0.106	0.109	0.097	0.102
Dec	0.096	0.106	0.136	0.138

Table 4-11 RMSEs of monthly proportion of wet days of Quebec City (S2) station giving NCEP-CanESM2

Month	Quebec (S2) NCEP-CanESM2 PRCP1			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	0.090	0.107	0.104	0.113
Feb	0.100	0.113	0.128	0.136
Mar	0.090	0.097	0.112	0.116
Apr	0.096	0.107	0.096	0.100
May	0.089	0.105	0.124	0.137
Jun	0.111	0.118	0.100	0.102
Jul	0.116	0.117	0.122	0.125
Aug	0.102	0.102	0.094	0.096
Sep	0.088	0.098	0.105	0.108
Oct	0.110	0.117	0.109	0.115
Nov	0.119	0.125	0.108	0.118
Dec	0.108	0.119	0.106	0.107

Table 4-12 RMSEs of monthly proportion of wet days of Quebec City (S2) station giving NCEP-HadCM3

Month	Quebec (S2) NCEP-HadCM3 PRCP1			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Jan	0.101	0.109	0.103	0.113
Feb	0.107	0.111	0.123	0.133
Mar	0.094	0.097	0.115	0.118
Apr	0.106	0.109	0.102	0.114
May	0.099	0.105	0.129	0.142
Jun	0.101	0.109	0.106	0.110
Jul	0.113	0.115	0.119	0.124
Aug	**0.096**	**0.092**	0.091	0.095
Sep	0.092	0.098	0.107	0.112
Oct	0.115	0.116	0.107	0.114
Nov	0.115	0.127	0.124	0.132
Dec	0.111	0.114	0.100	0.105

Table 4-13 RMSEs of seasonal maximum consecutive dry days of Montreal (S1) station giving NCEP-CanESM2 and NCEP-HadCM3

Month	Montreal (S1) NCEP-CanESM2 CDD			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	7.06	7.25	4.25	4.42
Spring	**5.28**	**5.23**	3.58	3.68
Summer	3.43	3.66	**6.92**	**6.87**
Autumn	4.29	4.44	4.12	4.18
Month	**Montreal (S1) NCEP-HadCM3 CDD**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	6.90	7.26	4.80	4.94
Spring	5.63	5.71	3.29	3.61
Summer	3.52	3.68	**7.19**	**7.17**
Autumn	4.58	4.66	3.54	3.54

Table 4-14 RMSEs of seasonal maximum consecutive dry days of Quebec City (S2) station giving NCEP-CanESM2 and NCEP-HadCM3

Month	Quebec (S2) NCEP-CanESM2 CDD			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	4.90	5.44	**4.26**	**3.97**
Spring	3.75	4.02	**5.21**	**5.00**
Summer	3.33	3.46	**3.55**	**3.43**
Autumn	3.64	3.84	**3.03**	**2.99**
Month	**Quebec (S2) NCEP-HadCM3 CDD**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	**5.76**	**5.57**	**3.98**	**3.86**
Spring	3.98	3.99	5.43	5.46
Summer	3.45	3.57	3.39	3.39
Autumn	3.58	3.86	2.89	3.02

Table 4-15 RMSEs of seasonal 90th percentile of rain day amount of Montreal (S1) station giving NCEP-CanESM2 and NCEP-HadCM3

Month	Montreal (S1) NCEP-CanESM2 Prec90p			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	4.76	6.15	**6.64**	**4.76**
Spring	3.77	4.64	**4.39**	**3.86**
Summer	5.92	7.07	**6.86**	**6.10**
Autumn	3.36	4.91	4.18	3.79
Month	**Montreal (S1) NCEP-HadCM3 Prec90p**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	4.79	4.91	**6.91**	**6.72**
Spring	3.74	4.00	4.75	4.82
Summer	5.78	6.03	5.86	6.31
Autumn	3.80	3.91	**5.59**	**5.17**

Table 4-16 RMSEs of seasonal 90th percentile of rain day amount of Quebec City (S2) station giving NCEP-CanESM2 and NCEP-HadCM3

Month	Quebec (S2) NCEP-CanESM2 Prec90p			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	5.28	5.85	5.57	5.78
Spring	4.58	4.76	**4.98**	**4.78**
Summer	7.33	7.58	6.59	6.86
Autumn	4.02	4.56	3.95	4.16
Month	**Quebec (S2) NCEP-HadCM3 Prec90p**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	5.78	6.01	**5.57**	**5.15**
Spring	4.79	4.90	**4.46**	**4.44**
Summer	7.50	7.55	**7.00**	**6.83**
Autumn	4.39	4.55	4.12	4.17

Table 4-17 RMSEs of seasonal simple daily intensity of Montreal (S1) station giving NCEP-CanESM2 and NCEP-HadCM3

Month	Montreal (S1) NCEP-CanESM2 SDII			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	0.77	0.97	1.14	1.27
Spring	1.20	1.23	1.33	1.38
Summer	1.76	1.81	2.22	2.30
Autumn	1.83	1.95	**1.90**	**1.72**
Month	**Montreal (S1) NCEP-HadCM3 SDII**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	0.96	0.96	**1.64**	**1.56**
Spring	1.29	1.41	1.37	1.40
Summer	1.69	1.84	2.20	2.36
Autumn	1.86	1.97	1.85	1.85

Table 4-18 RMSEs of seasonal simple daily intensity of Quebec City (S2) station giving NCEP-CanESM2 and NCEP-HadCM3

Month	Quebec (S2) NCEP-CanESM2 SDII			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	1.11	1.30	1.12	1.23
Spring	1.34	1.38	**1.48**	**1.46**
Summer	2.19	2.31	**2.20**	**2.13**
Autumn	1.43	1.55	1.54	1.72
Month	**Quebec (S2) NCEP-HadCM3 SDII**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Winter	**1.19**	**1.16**	1.25	1.27
Spring	**1.48**	**1.46**	**1.38**	**1.34**
Summer	2.22	2.31	**2.11**	**2.08**
Autumn	1.50	1.67	**1.81**	**1.64**

Table 4-19 RMSEs of annual daily maximum precipitation *of Montreal (S1) station and Quebec City (S2) station giving NCEP-CanESM2 and NCEP-HadCM3*

Month	Montreal (S1) NCEP-CanESM2 AMP			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	20.22	25.71	26.11	29.42
Month	**Montreal (S1) NCEP-HadCM3 AMP**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	**21.10**	**19.83**	**23.16**	**20.81**
Month	**Quebec (S2) NCEP-CanESM2 AMP**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	21.72	32.23	23.60	33.38
Month	**Quebec (S2) NCEP-HadCM3 AMP**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	**21.72**	**19.59**	**22.53**	**20.19**

Table 4-20 RMSEs of total annual precipitation *of Montreal (S1) station and Quebec City (S2) station giving NCEP-CanESM2 and NCEP-HadCM3*

Month	Montreal (S1) NCEP-CanESM2 TAP			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	131.80	136.03	123.67	127.04
Month	**Montreal (S1) NCEP-HadCM3 TAP**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	137.28	152.20	137.70	154.93
Month	**Quebec (S2) NCEP-CanESM2 TAP**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	146.24	163.98	156.31	157.15
Month	**Quebec (S2) NCEP-HadCM3 TAP**			
	Calibration period		Validation period	
	SDGAM	SDSM	SDGAM	SDSM
Annual	153.90	171.36	**167.34**	**162.02**

4.1.2 Graphical Analysis

For the purposes of graphical comparison, box plots are used to evaluate both the accuracy (how closely the model's estimated median value aligns with the observed data) and the dispersion of the model's results (indicated by the size of the Inter-Quartile Range box). The following box plots are constructed using 100 sets of daily precipitation series from SDGAM and 100 sets from SDSM. Blue points indicate the index of observed data, and box plots indicate model results. Figures 4-1 to 4-4 show the box plots for monthly mean of precipitation (Precp_m) for both S1 and S2 stations. During the calibration period, the median values obtained through SDGAM during calibration period are very close to the observed Precp_m, showing a higher level of accuracy as compared to the SDSM. During the validation period, the closeness of the median values derived from the SDGAM model to the observed Precp_m is comparable to that of the SDSM. Regarding the dispersion of models' results, the width of the box plots for SDGAM is also similar to that of SDSM. Additionally, it is noteworthy that the precipitation series downscaled by both SDGAM and SDSM tend to overestimate the monthly mean precipitation for the months July, August, and September.

In Figures 4-5 to 4-8, both SDGAM and SDSM can adequately describe the standard deviation of precipitation. However, it is evident that compared to SDSM, the median values in the box plots from SDGAM are closer to the observed Precp_sd. Figures 4-9 to 4-12 indicate that during the calibration period, the median of PRCP1 obtained from SDGAM closely aligns with the observed PRCP1, nearly overlapping with it. The median of PRCP1 derived from SDSM is also very close to the observed PRCP1. In the validation period, despite the closeness of both downscaling PRCP1 to the observed PRCP1, the

observed values often fall outside the box plots due to the small interquartile ranges.

Figures 4-13 to Figure 4-24 show the box plots for the seasonal indices: CDD, and Prec90p, and SDII, respectively. Figures 4-13 to 4-16 indicates that both SDGAM and SDSM adequately describe the observed precipitation's Consecutive Dry Days (CDD). In particular, the CDD values in the spring and summer of southern Quebec are slightly higher than in the autumn and winter seasons. However, both downscaling models tend to underestimate CDD by approximately two days during both the calibration and validation periods. Figures 4-17 to 4-20 show that SDGAM slightly outperforms SDSM in representing the 90^{th} percentile of rain day amounts. The captured 90th percentile precipitation values by both models are higher in the summer than in the spring, autumn, and winter. Throughout calibration and validation periods, both models tend to underestimate Prec90p by approximately 2-3 mm. Figures 4-21 to 4-24 indicates that SDGAM, compared to SDSM, more accurately represents the observed Simple Daily Intensity Index (SDII). During the calibration period, the observed SDII all falls within the corresponding boxes of SDGAM, whereas the observed SDII often appears on the whiskers of the box plots of SDSM. Both downscaling models capture the observed pattern where the SDII is higher in summer than in spring, followed by autumn, and lowest in winter. During the validation period, both SDGAM and SDSM tend to overestimate the summer SDII and significantly underestimate the winter SDII at Station S1 giving NCEP-HadCM3.

Figures 4-25 and 4-32 show the box plots for the annual indices: AMP and TAP. Figures 4-25 to 4-28 show that both SDGAM and SDSM describe well the trends of the annual maximum daily precipitation for both Montreal and Quebec City stations, which typically ranges from 30mm to 100mm. Moreover, with SDSM, the AMP

in certain years significantly exceeds the observed values, whereas SDGAM more effectively addresses this issue. Figures 4-29 to 4-32 indicate that both SDGAM and SDSM accurately capture the annual precipitation ranges, with Montreal's typically between 800mm to 1200mm and Quebec's between 1000mm to 1400mm. Both downscaling models also describe adequately the annual variability of the precipitation.

Overall, the precipitation indices obtained from SDGAM can describe more accurately the observed precipitation indices as compared to those from SDSM. During the calibration period, the indices Precp_m, PRCP1, and SDII from SDGAM are very close to the observed values. Meanwhile, in the validation period, they are only slightly better than those from SDSM.

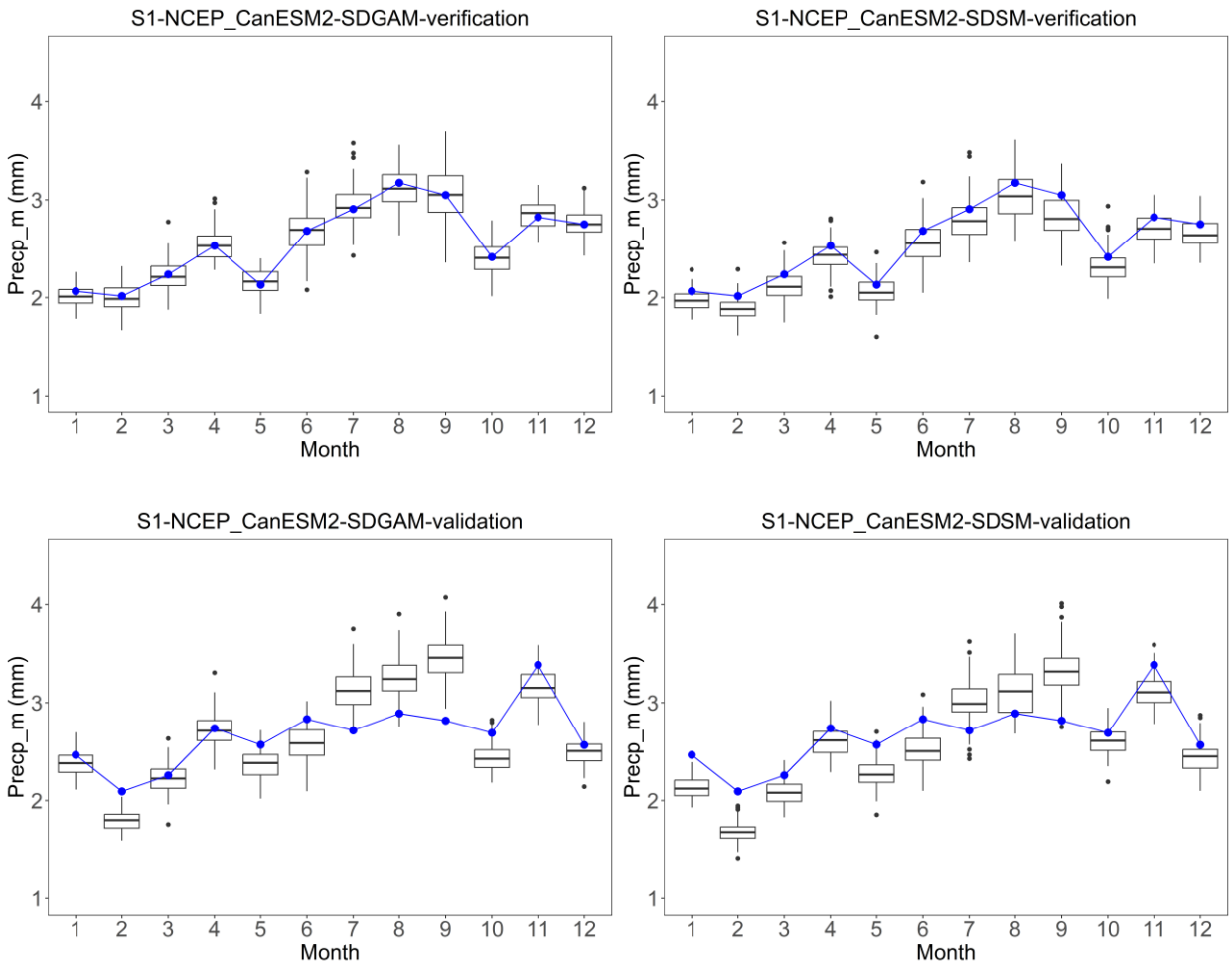

Figure 4-1. Box plots for monthly means of precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

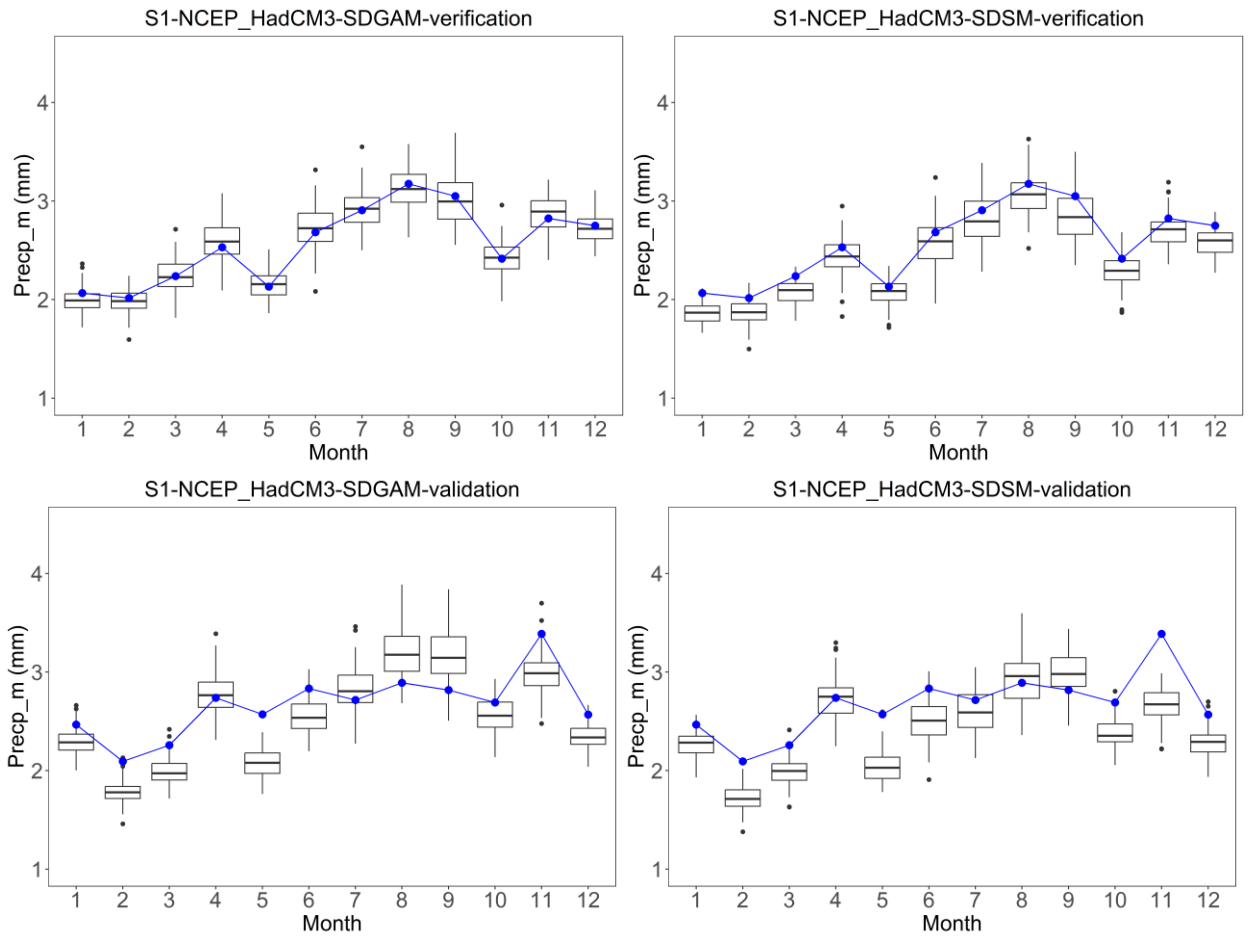

Figure 4-2. Box plots for monthly means of precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

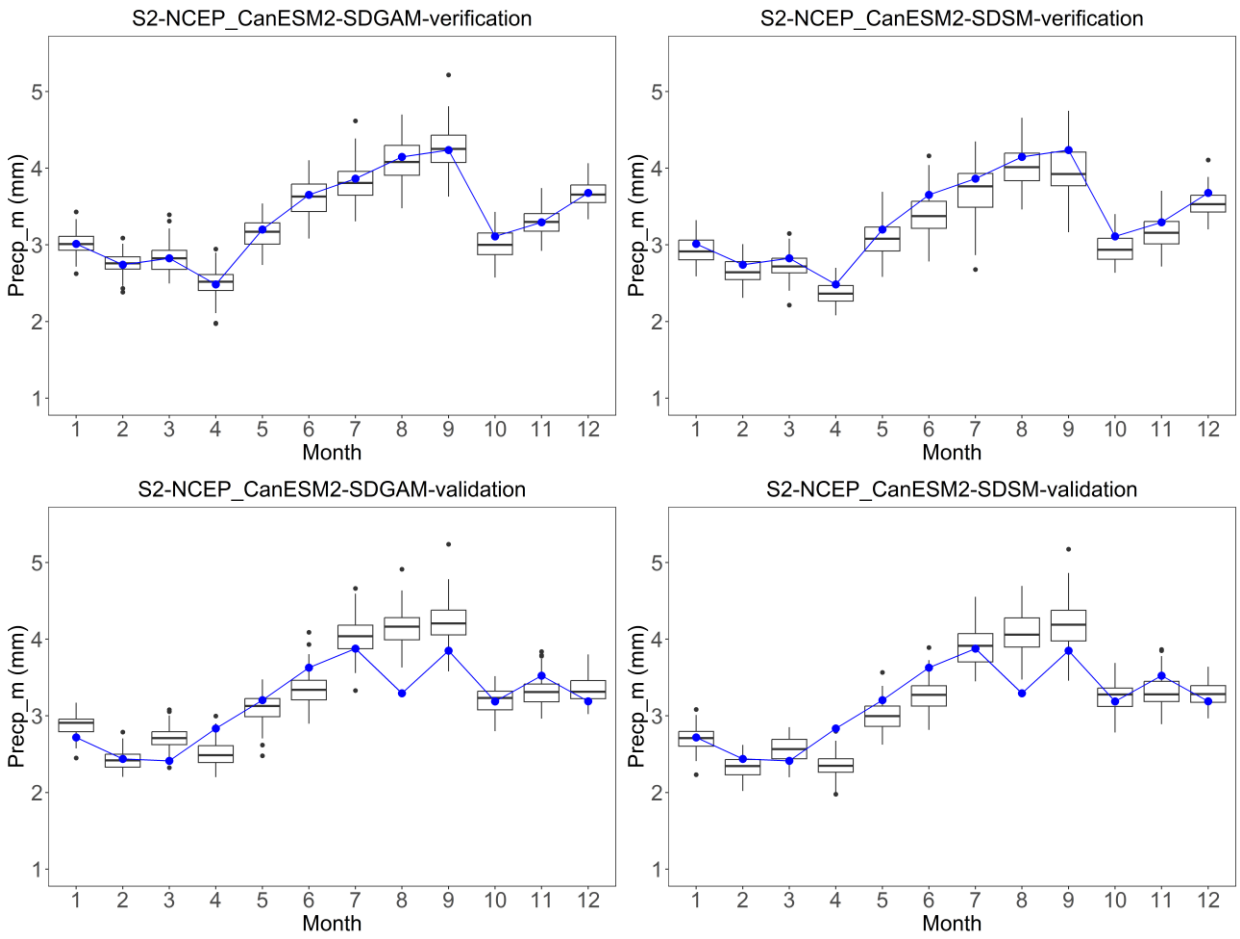

Figure 4-3. Box plots for monthly means of precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

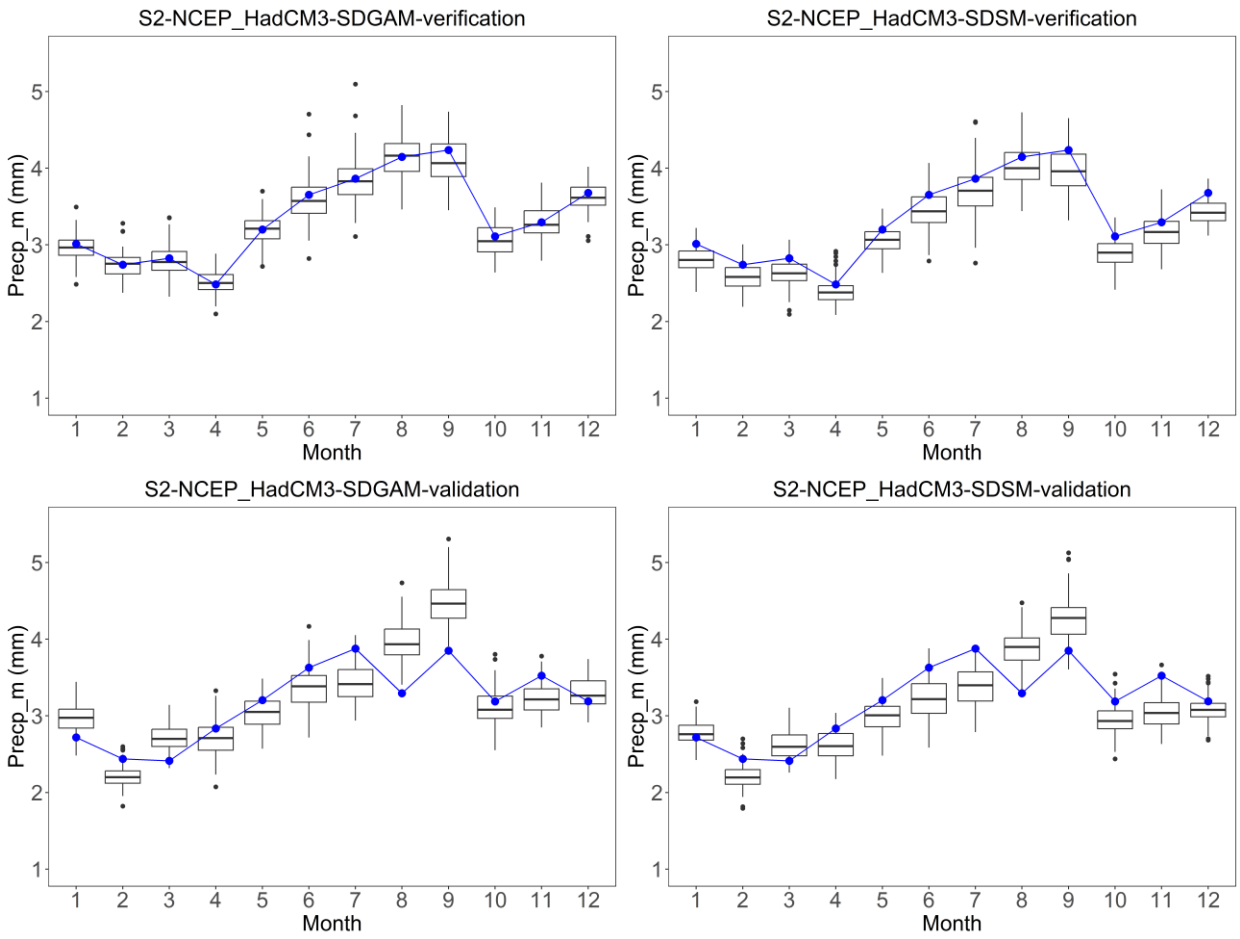

Figure 4-4. Box plots for monthly means of precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

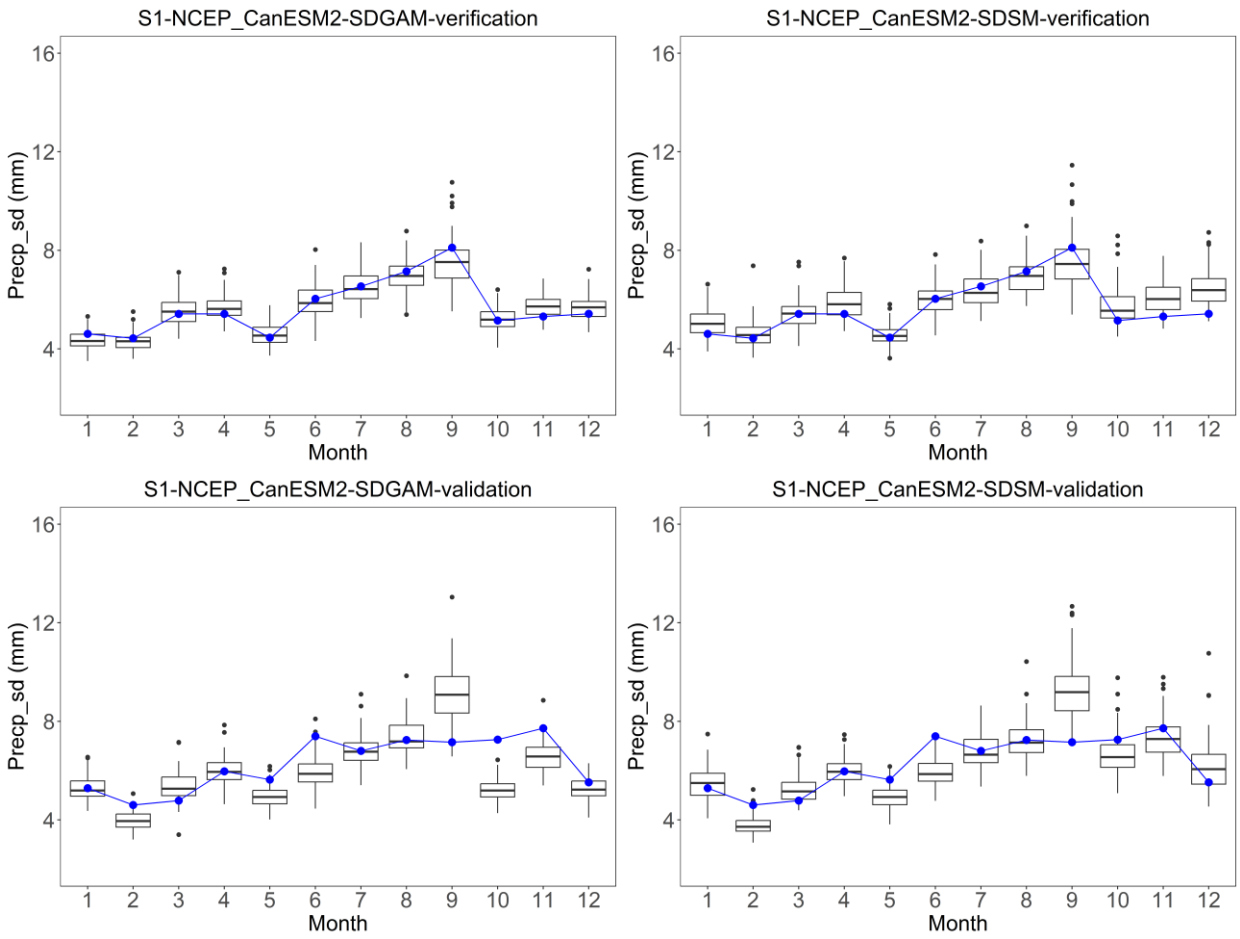

Figure 4-5. Box plots for monthly standard deviation of precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

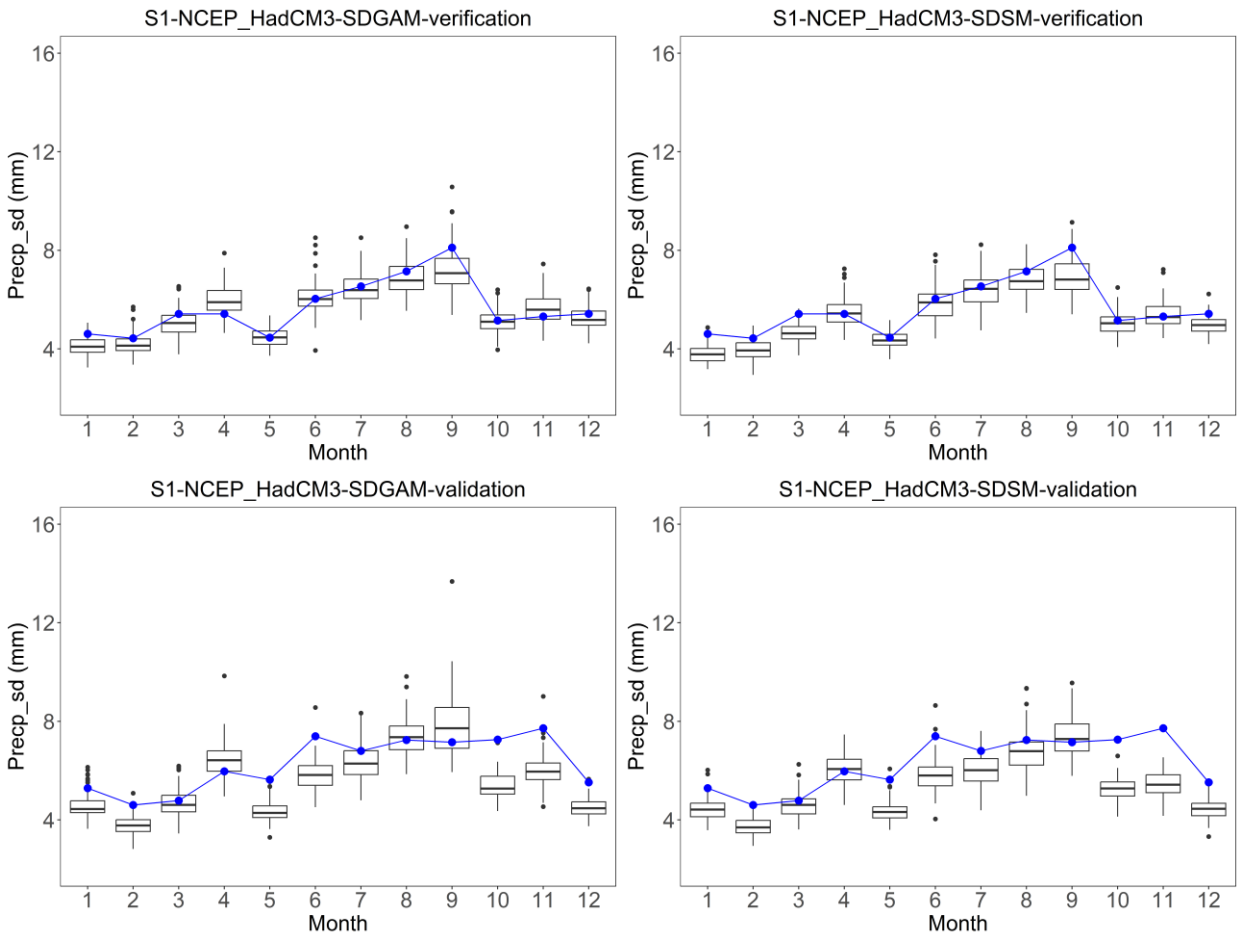

Figure 4-6. Box plots for monthly standard deviation of precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

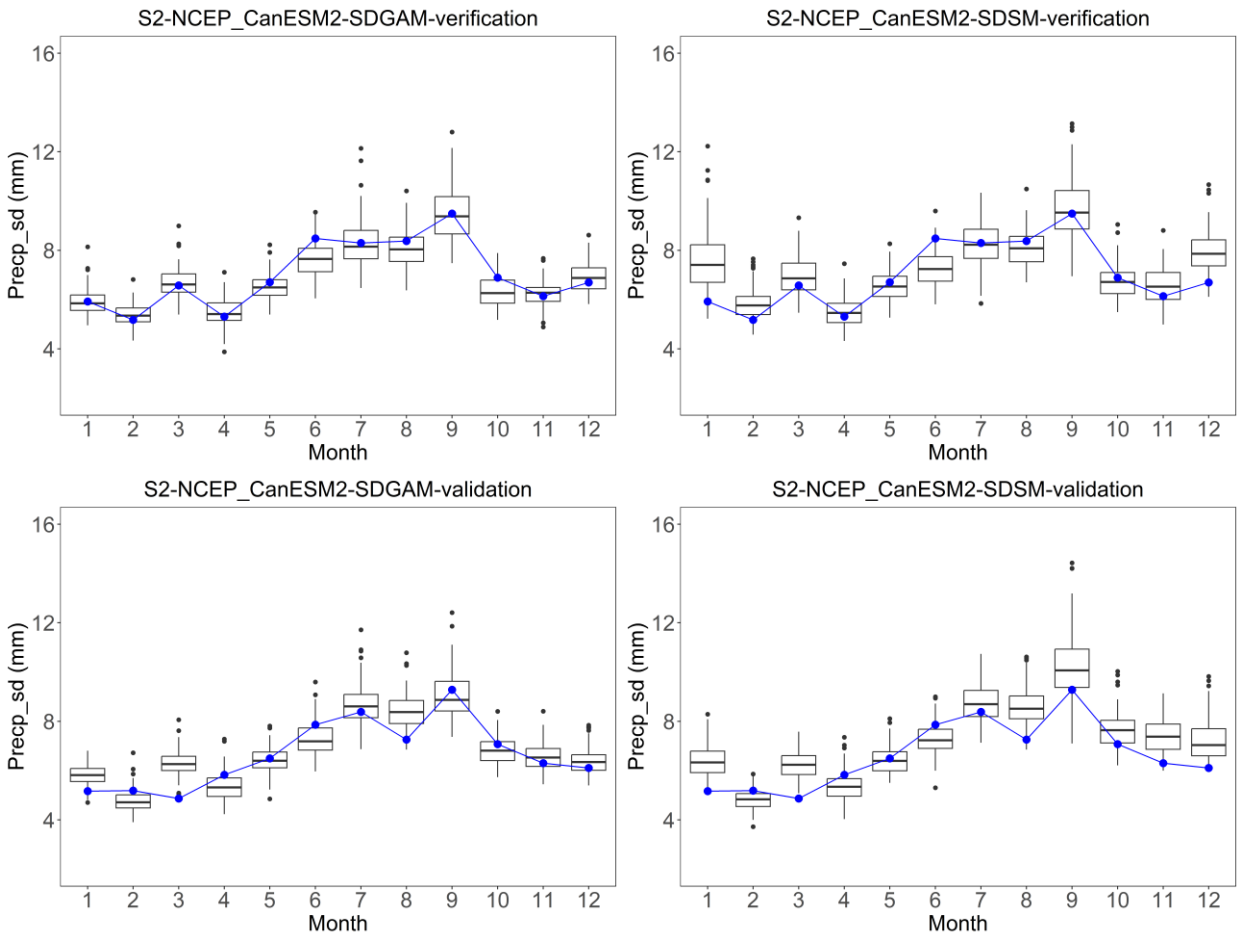

Figure 4-7. Box plots for monthly standard deviation of precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

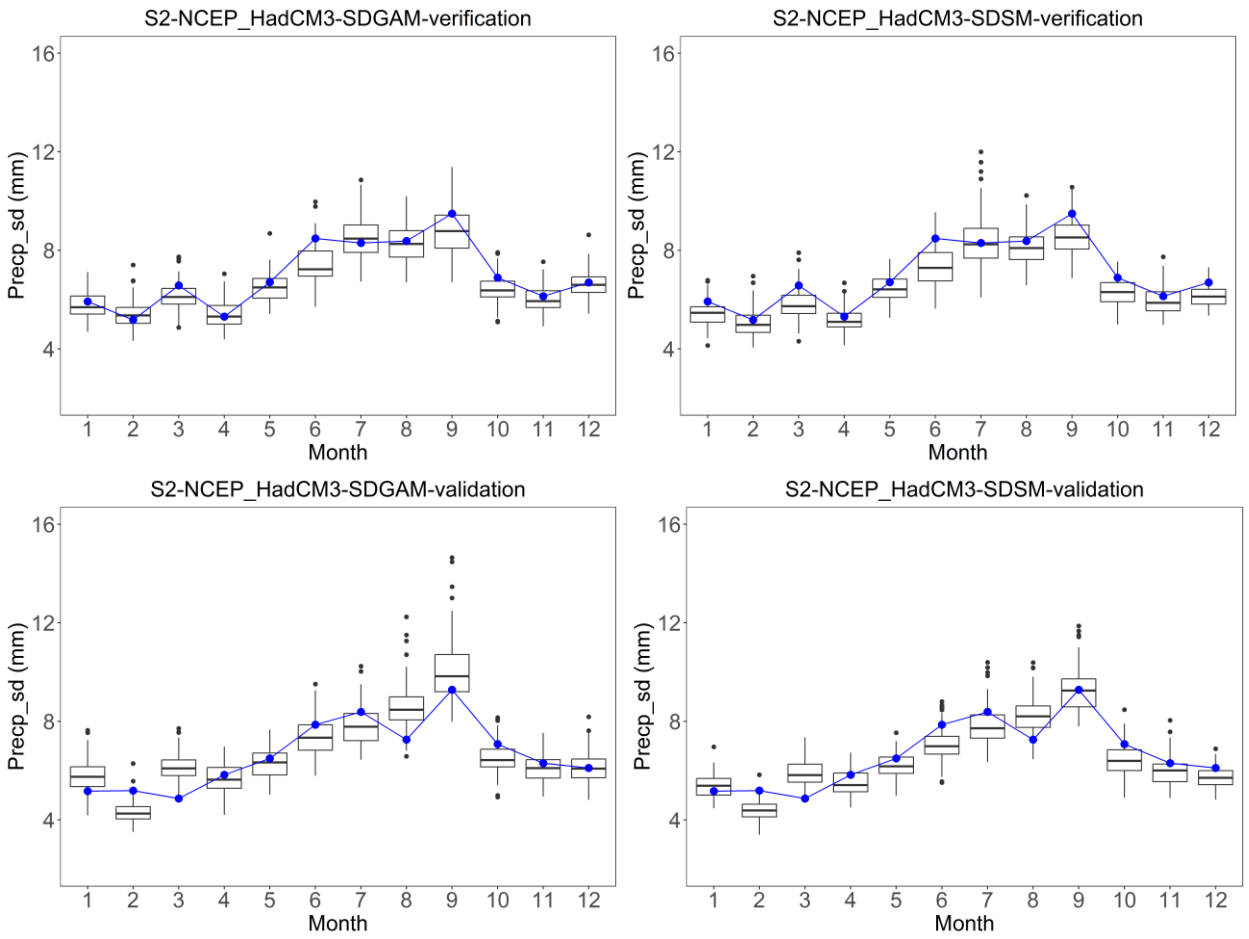

Figure 4-8. Box plots for monthly standard deviation of precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

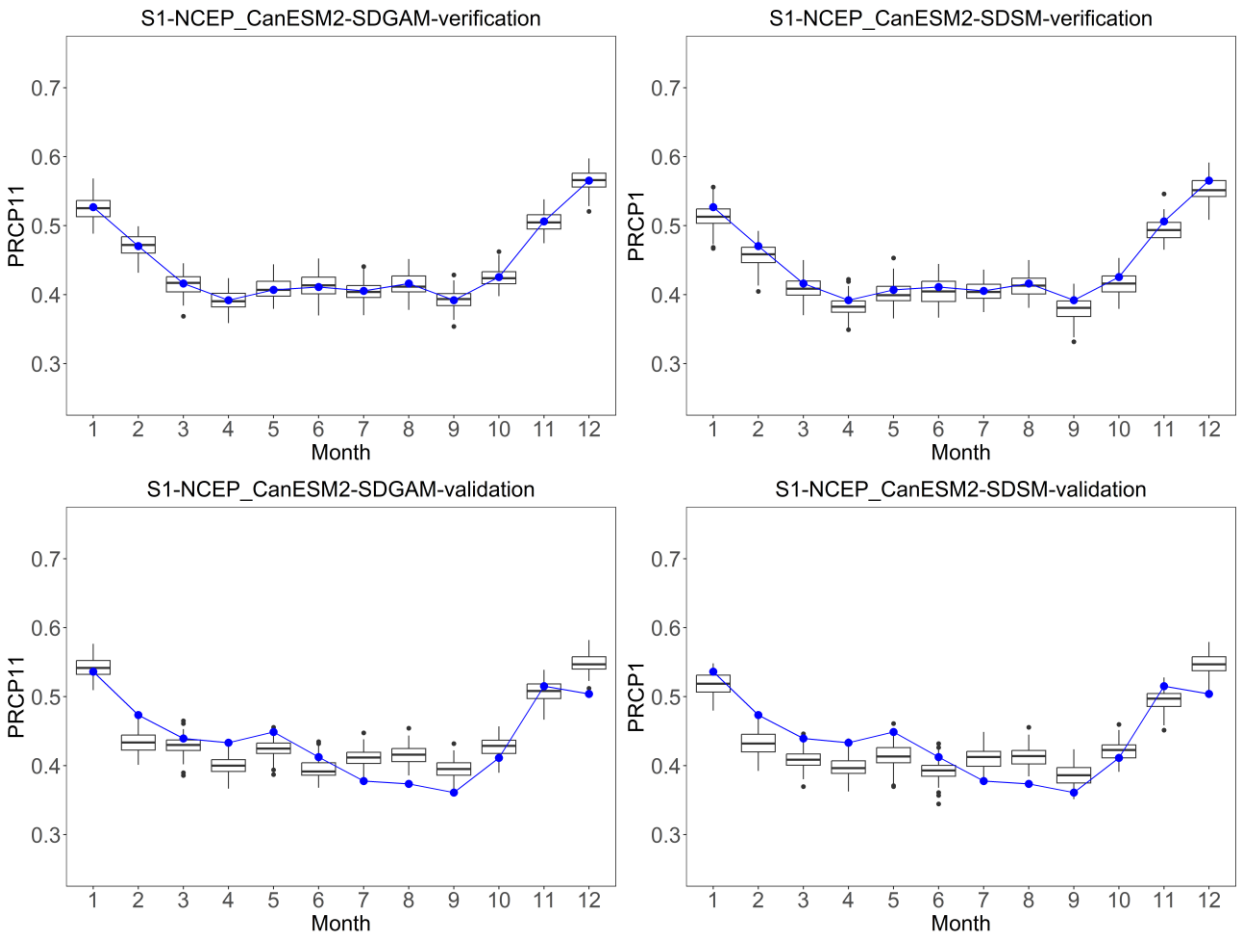

Figure 4-9. Box plots for monthly proportion of wet days of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

.

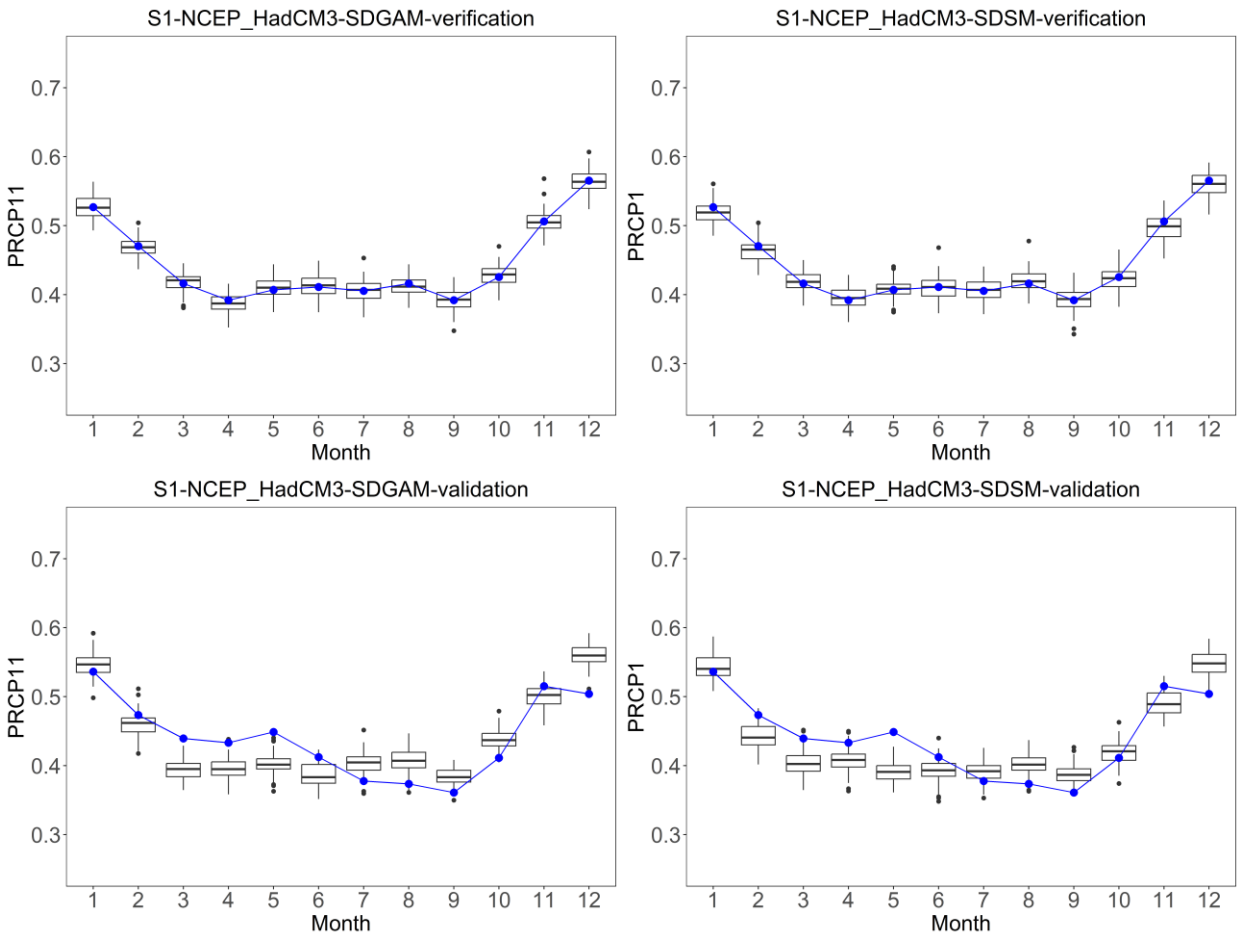

Figure 4-10. Box plots for monthly proportion of wet days of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

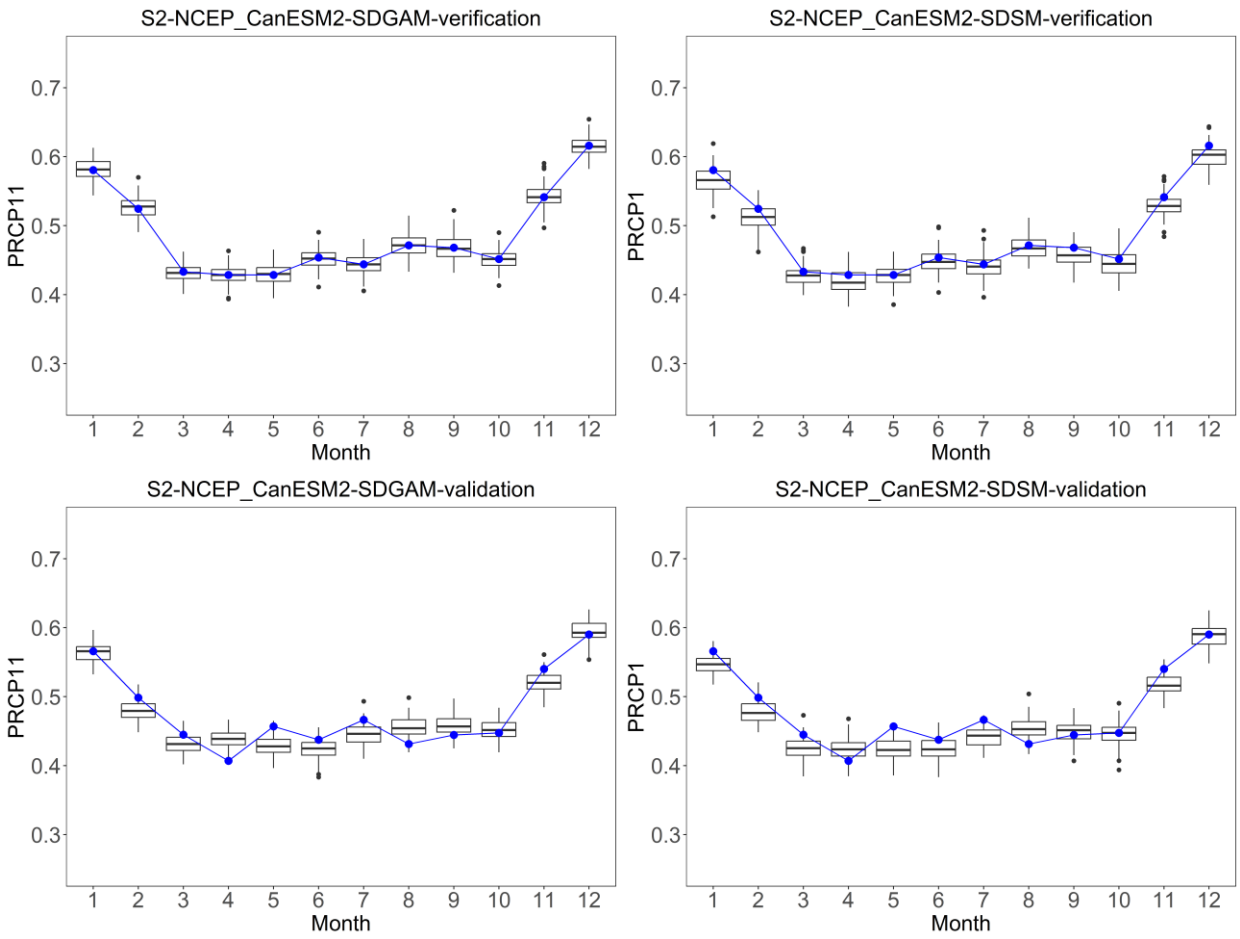

Figure 4-11. Box plots for monthly proportion of wet days of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

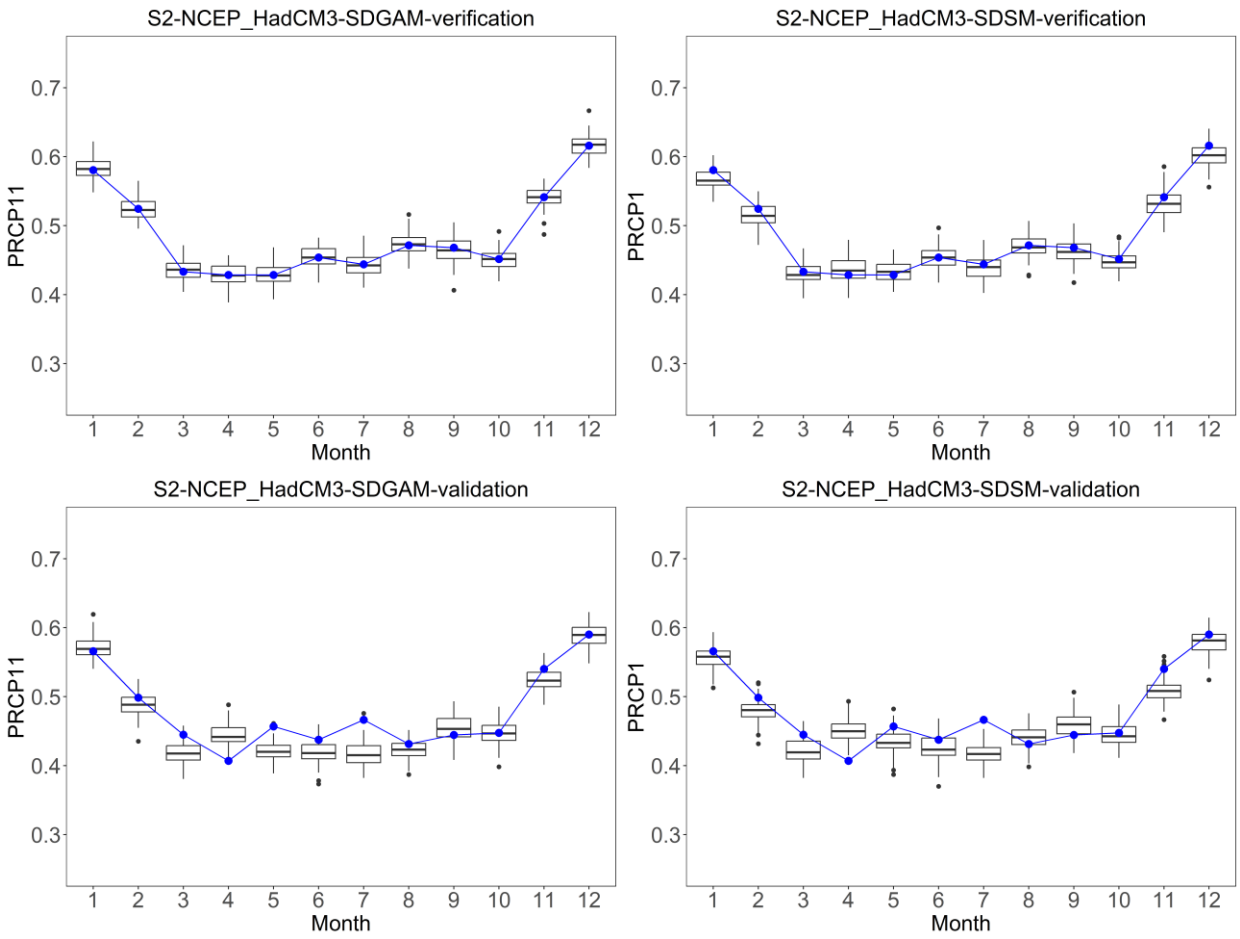

Figure 4-12. Box plots for monthly proportion of wet days of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

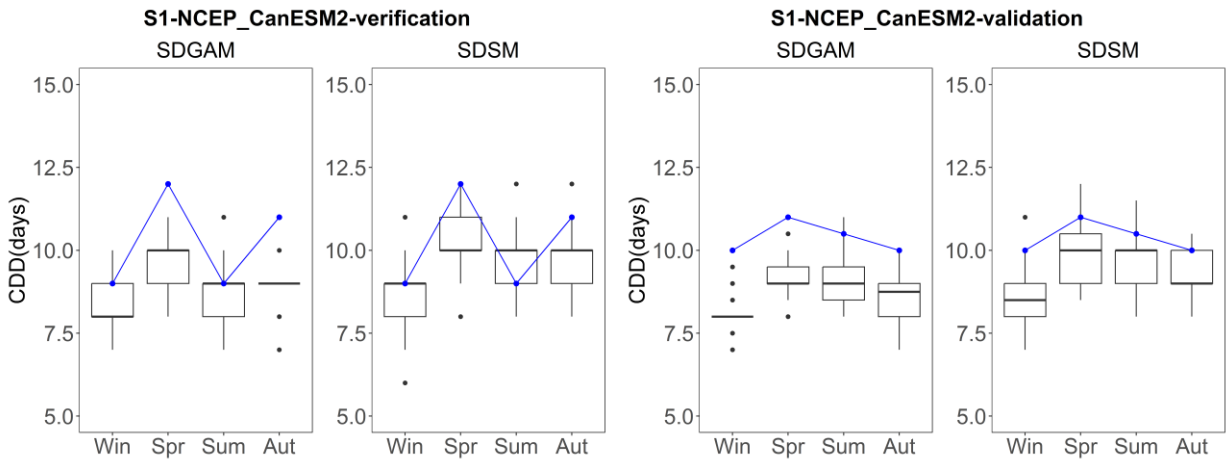

Figure 4-13. Box plots for seasonal maximum consecutive dry days of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

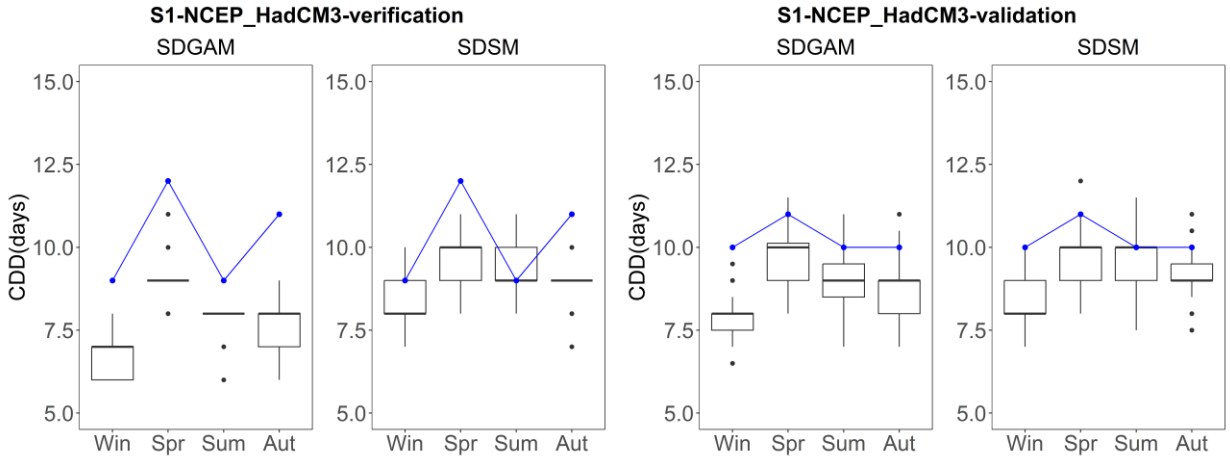

Figure 4-14. Box plots for seasonal maximum consecutive dry days of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

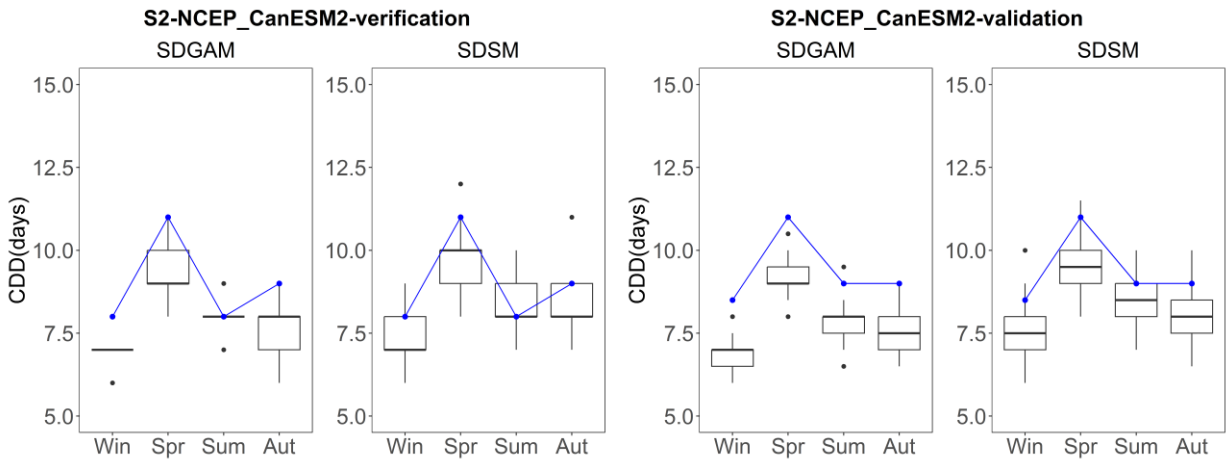

Figure 4-15. Box plots for seasonal maximum consecutive dry days of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

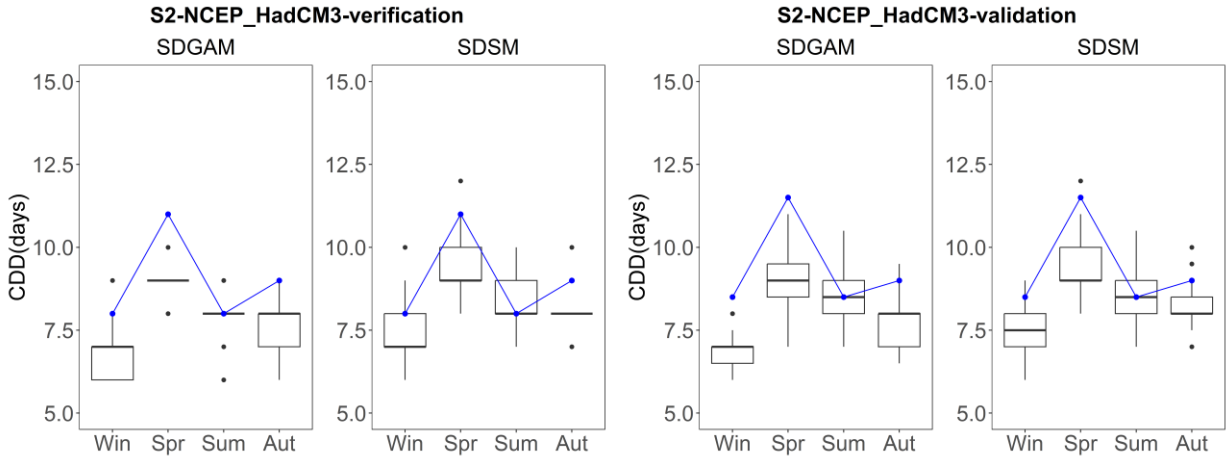

Figure 4-16. Box plots for seasonal maximum consecutive dry days of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

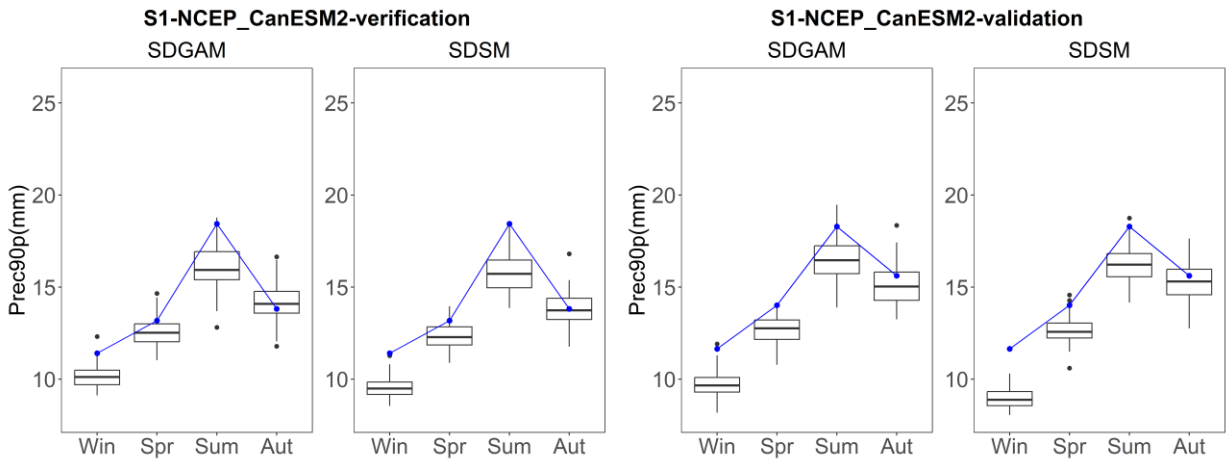

Figure 4-17. Box plots for seasonal 90th percentile of rain day amount of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

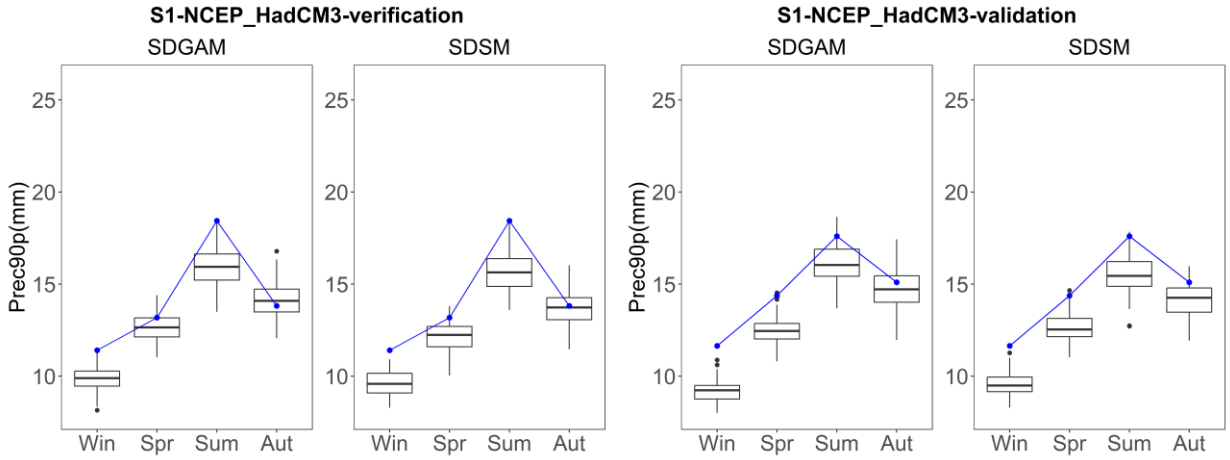

Figure 4-18. Box plots for seasonal 90th percentile of rain day amount of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

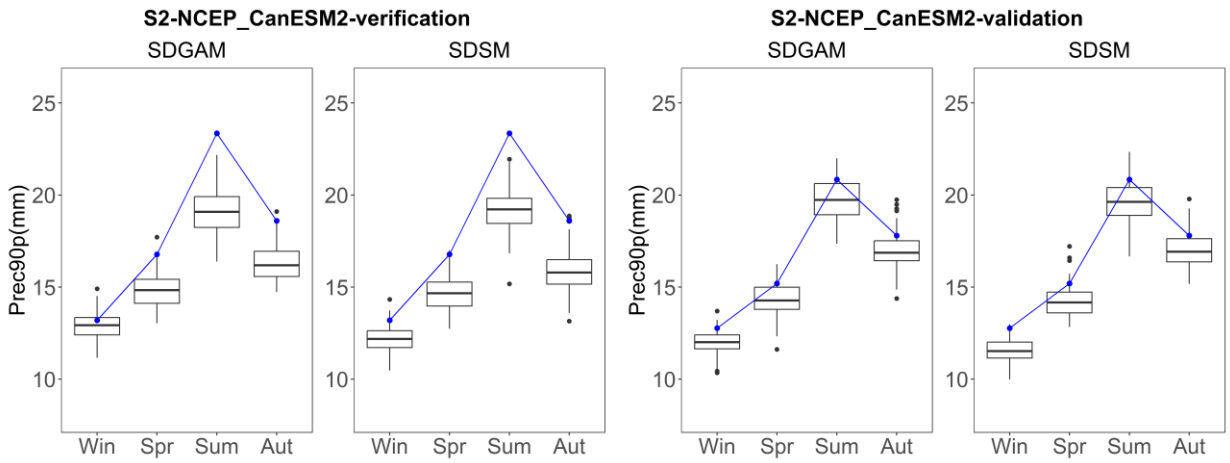

Figure 4-19. Box plots for seasonal 90th percentile of rain day amount of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

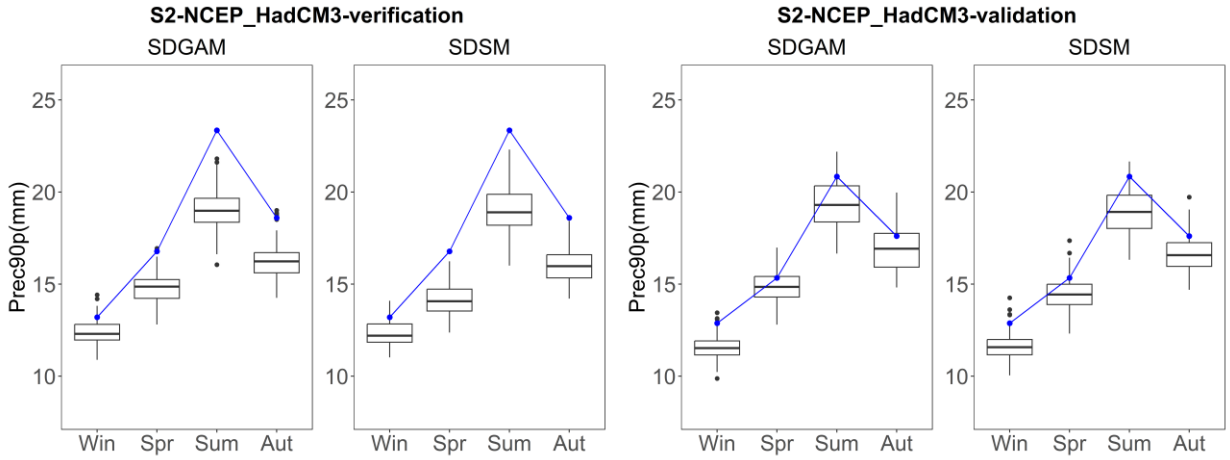

Figure 4-20. Box plots for seasonal 90th percentile of rain day amount of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

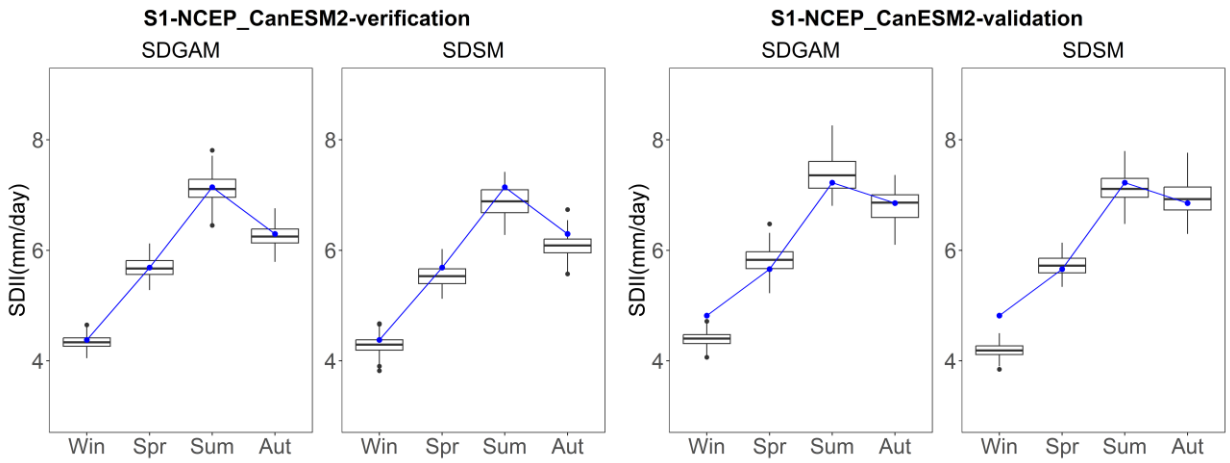

Figure 4-21. Box plots for seasonal mean precipitation at wet days *of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods*

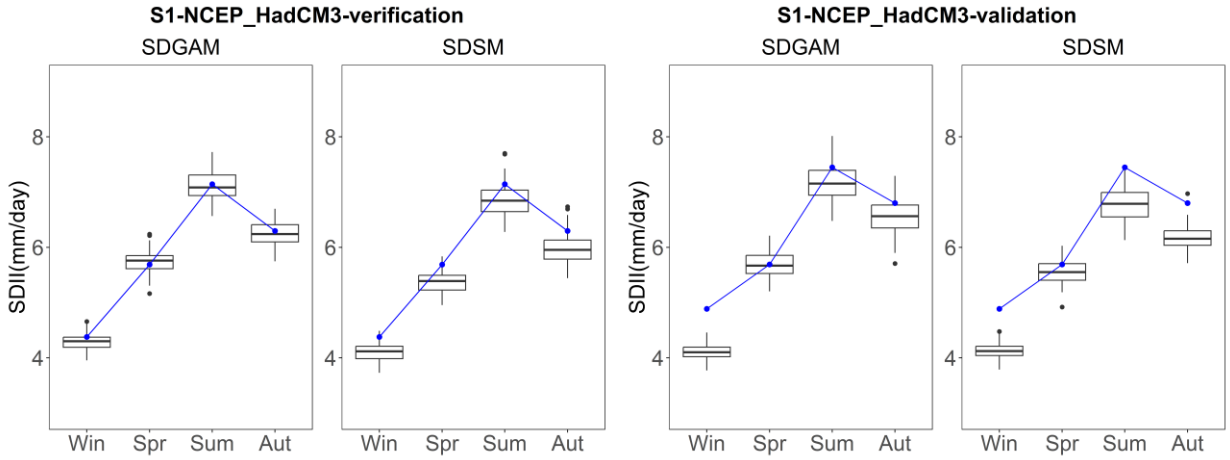

Figure 4-22. Box plots for seasonal mean precipitation at wet days *of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods*

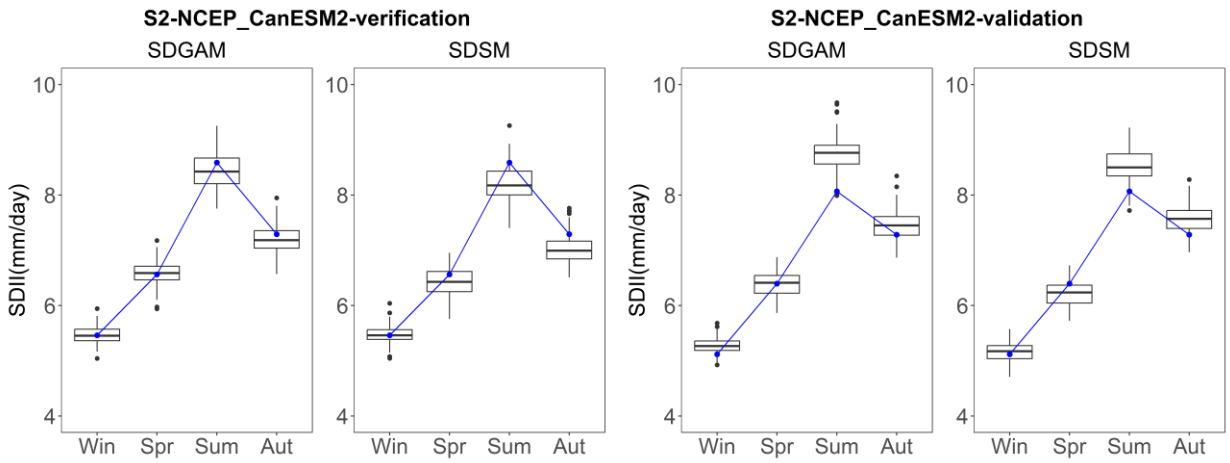

Figure 4-23. Box plots for seasonal mean precipitation at wet days *of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods*

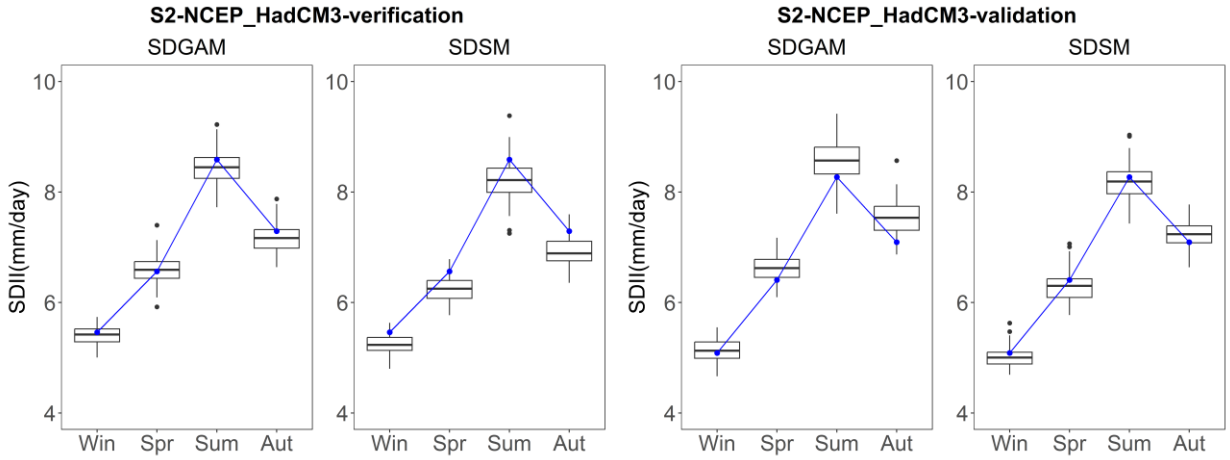

Figure 4-24. Box plots for seasonal mean precipitation at wet days *of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods*

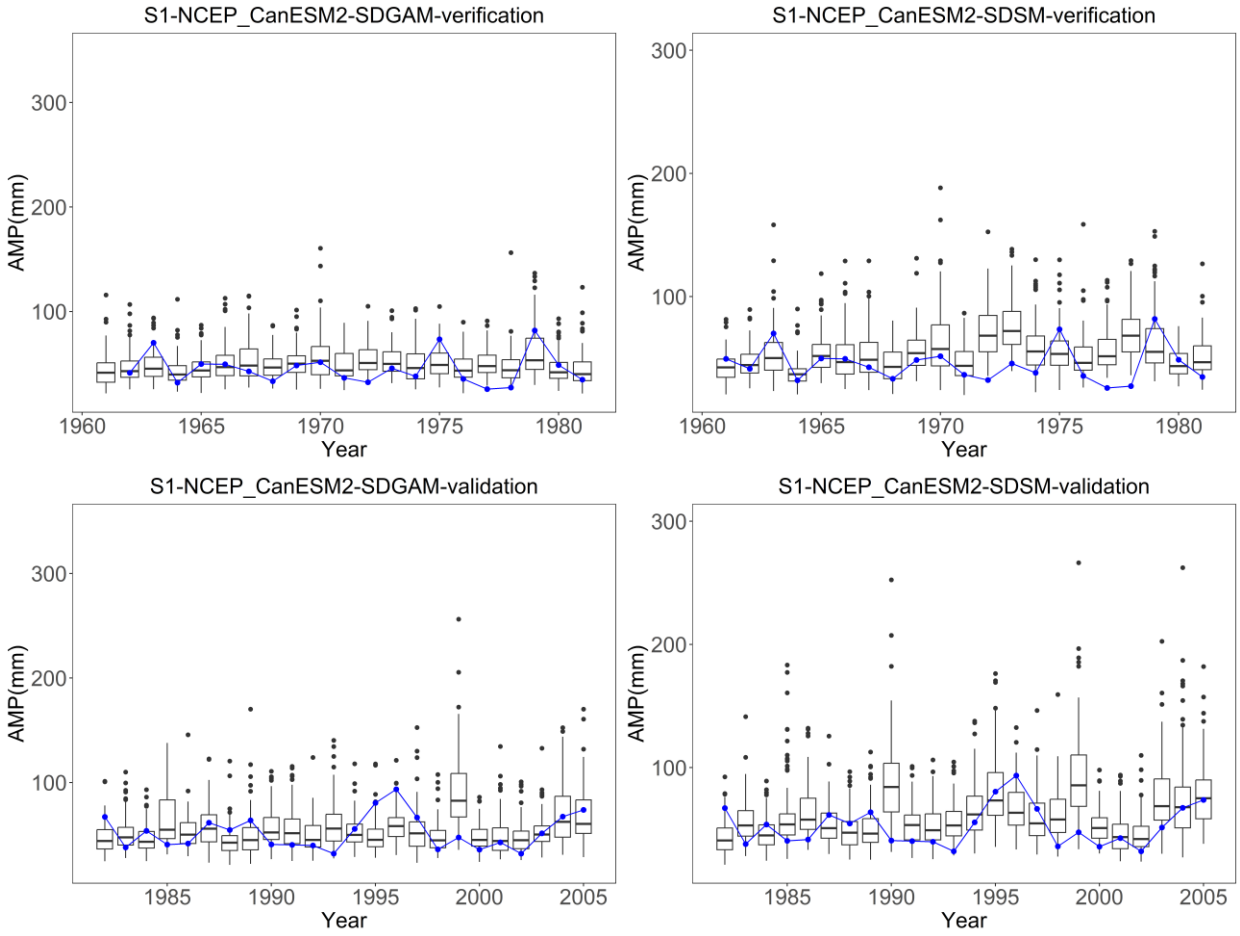

Figure 4-25. Box plots for annual maximum precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

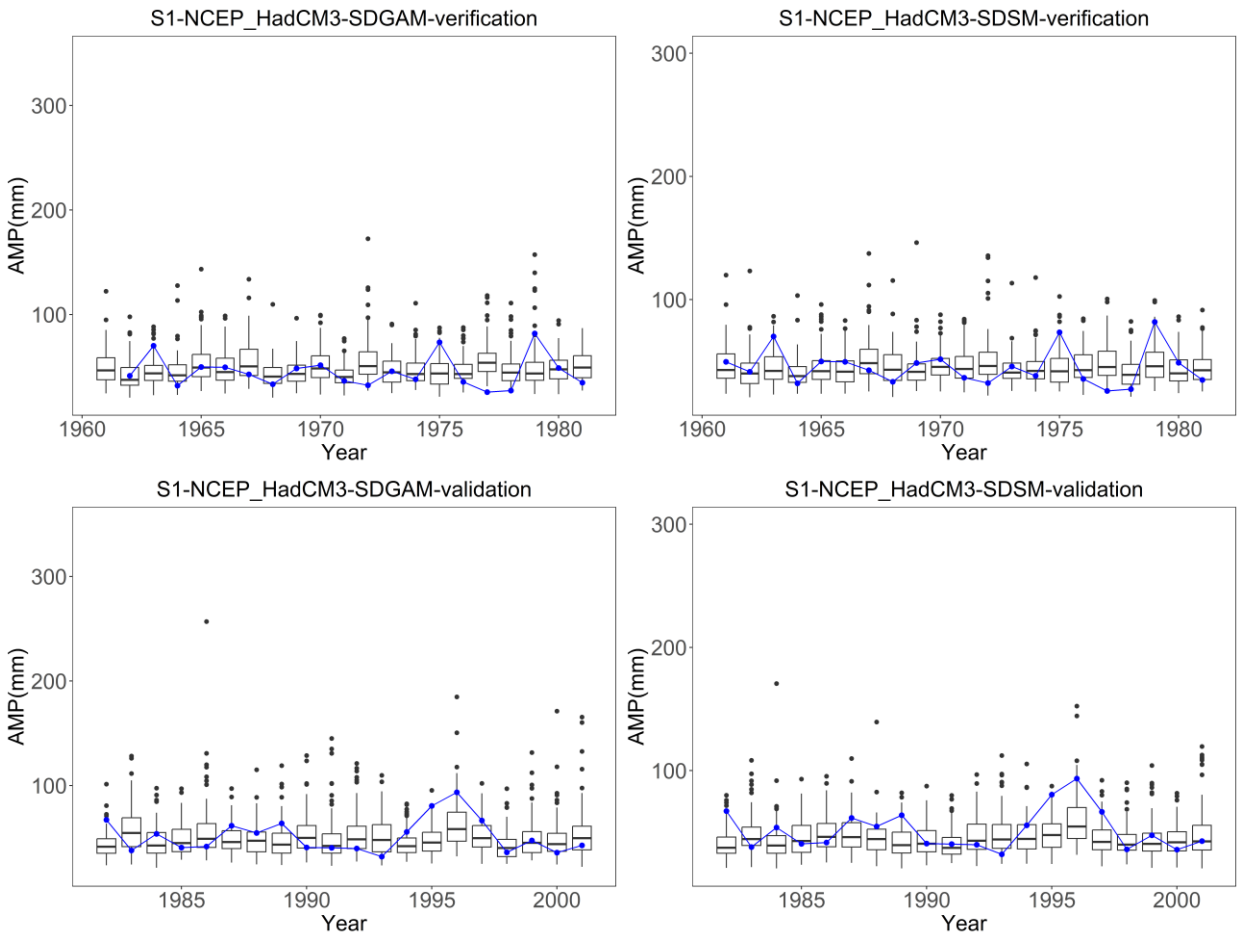

Figure 4-26. Box plots for annual daily maximum precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

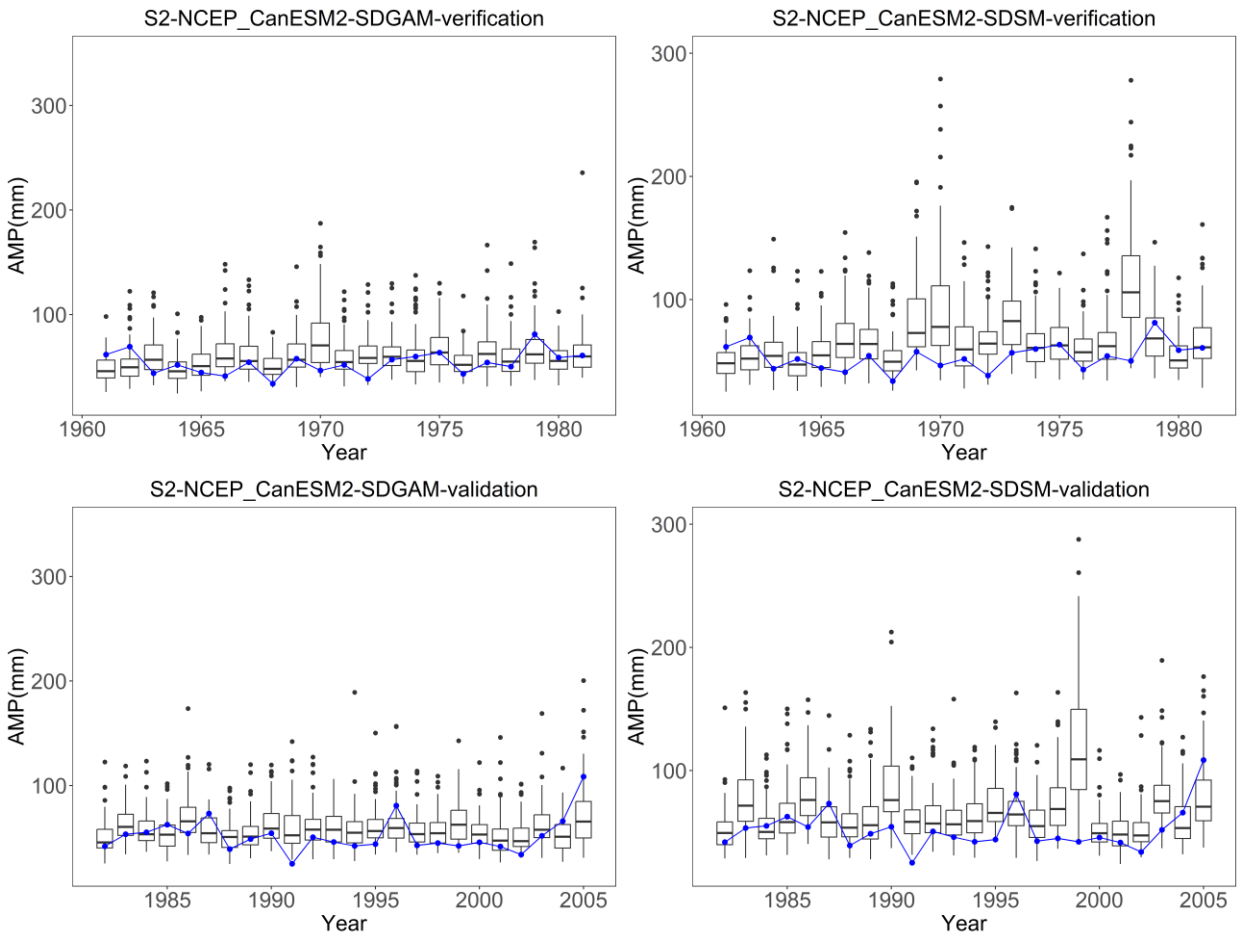

Figure 4-27. Box plots for annual daily maximum precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

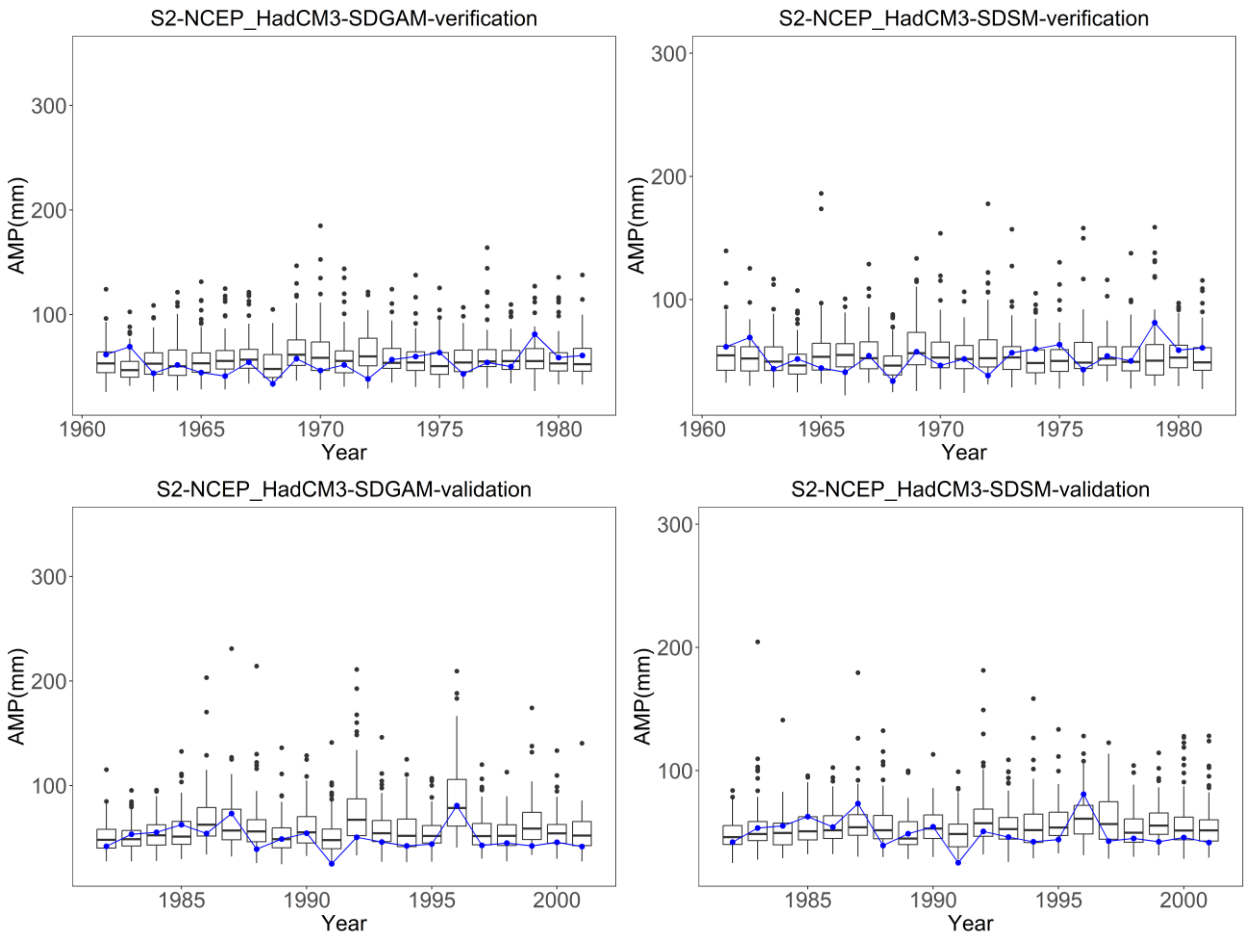

Figure 4-28. Box plots for annual daily maximum precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

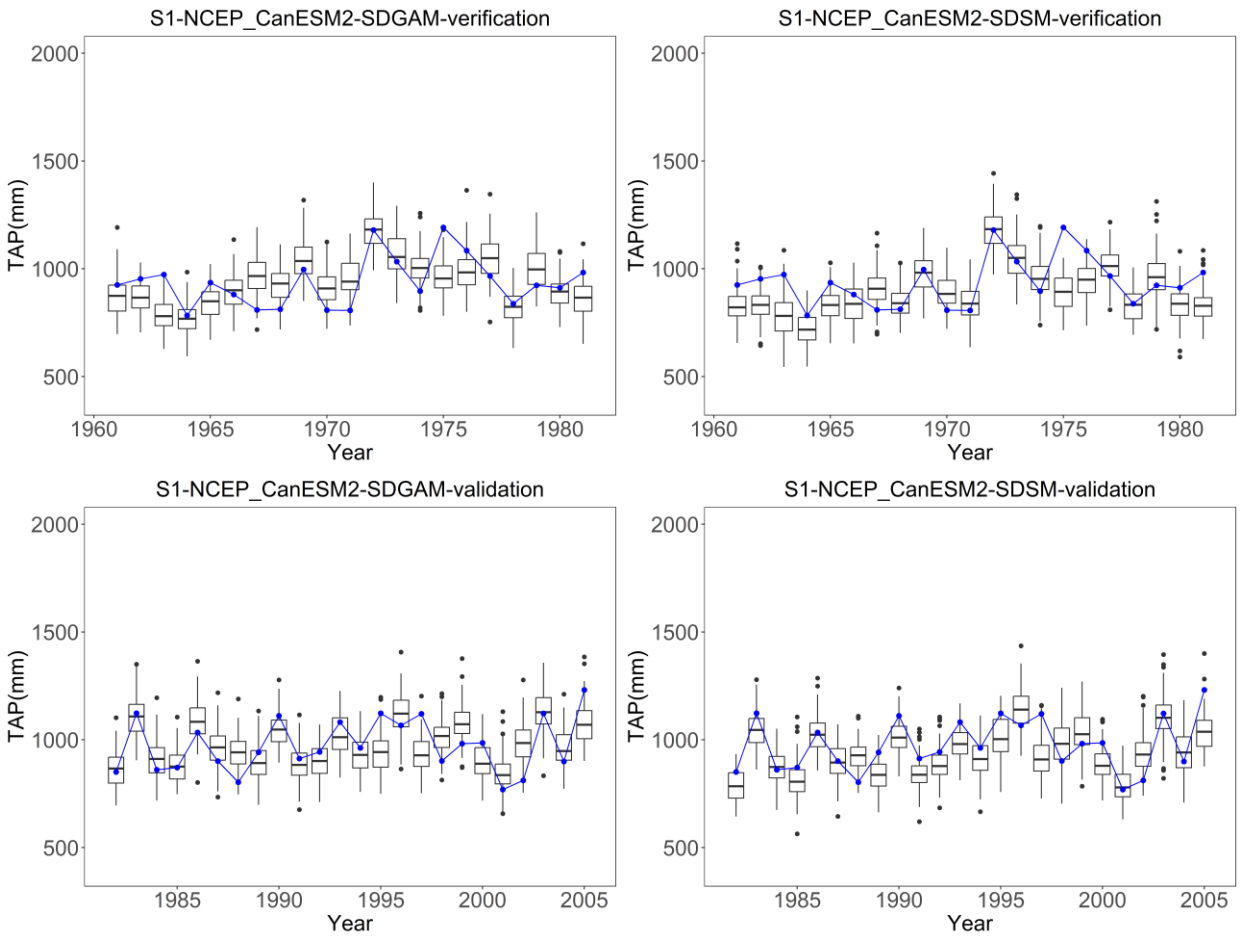

Figure 4-29. Box plots for total annual precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-CanESM2 over calibration and validation periods

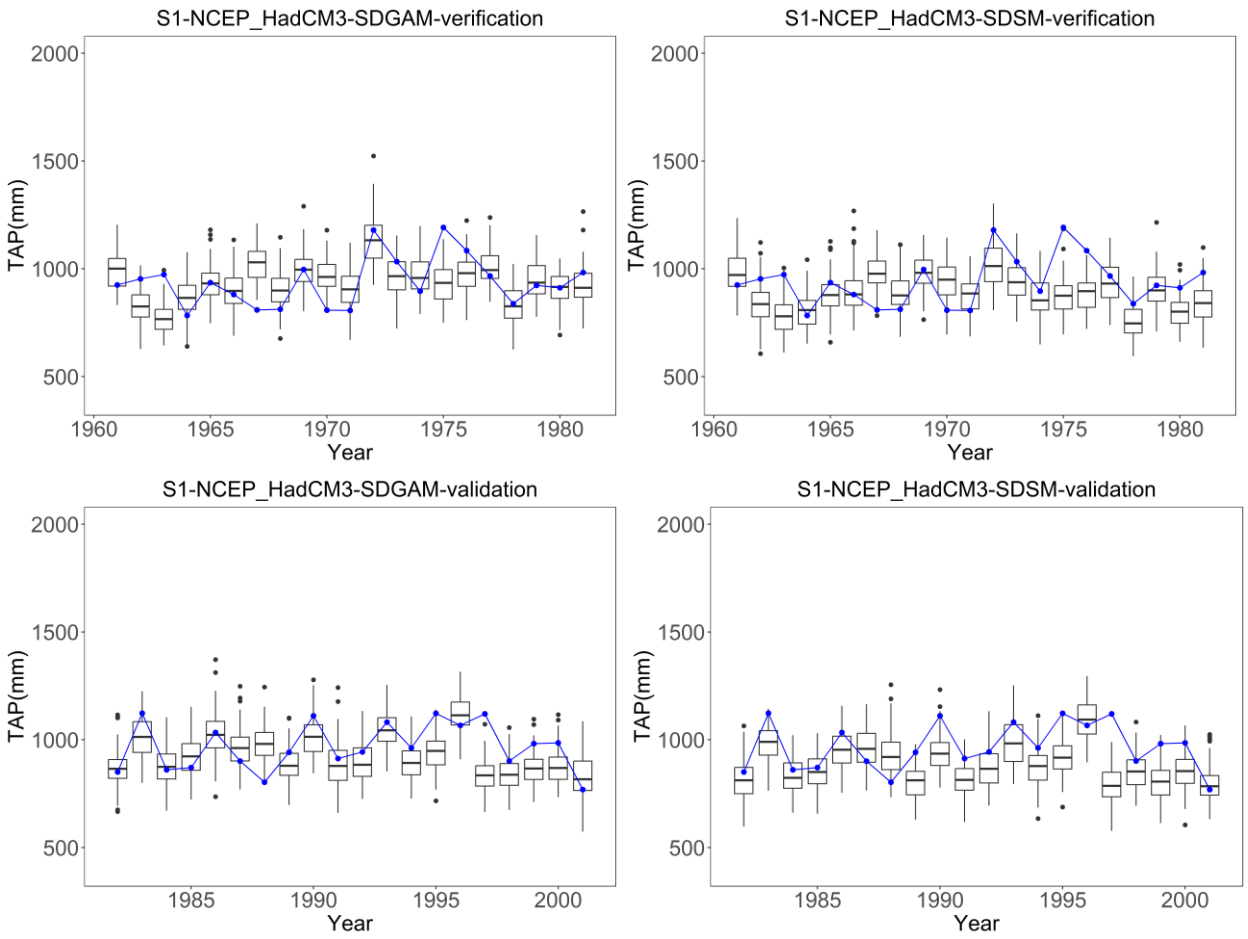

Figure 4-30. Box plots for total annual precipitation of SDGAM and SDSM of Montreal (S1) station giving NCEP-HadCM3 over calibration and validation periods

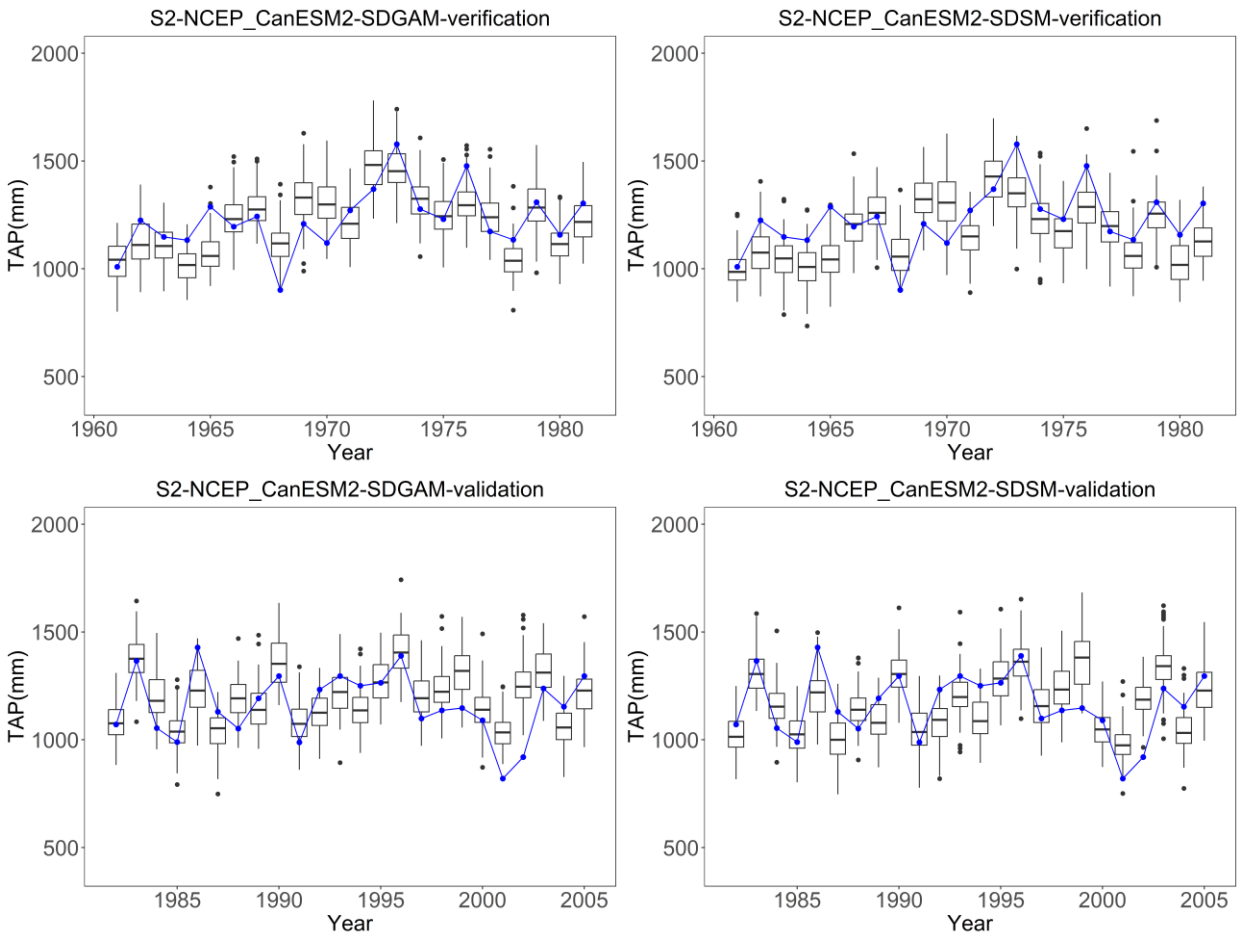

Figure 4-31. Box plots for total annual precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-CanESM2 over calibration and validation periods

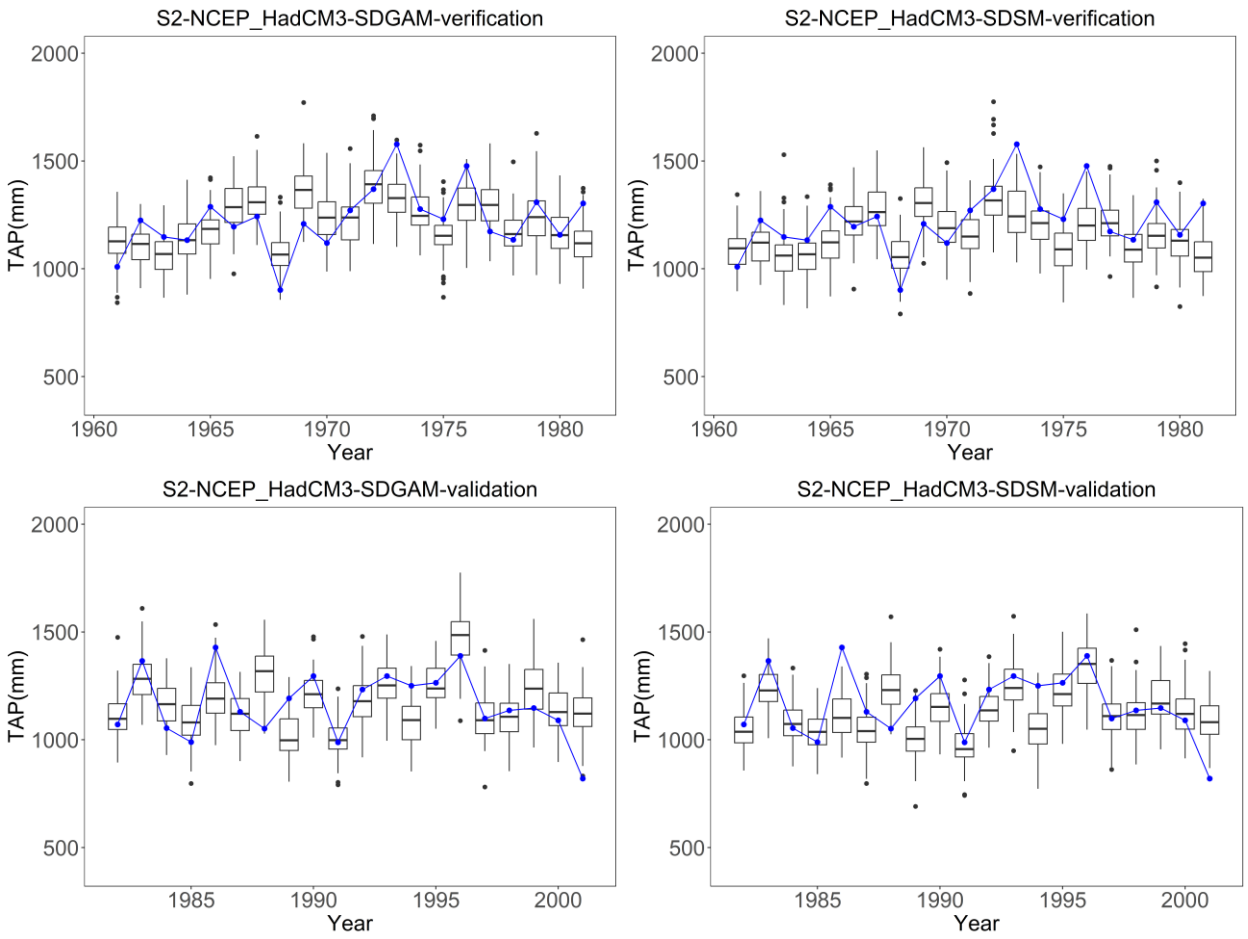

Figure 4-32. Box plots for total annual precipitation of SDGAM and SDSM of Quebec City (S2) station giving NCEP-HadCM3 over calibration and validation periods

4.2 Precipitation Projections by SDAGM and RCMs

4.2.1 Historical Precipitation Projections

The numerical and graphical analyses in the previous sections have indicated the feasibility and accuracy of the proposed SDGAM in capturing various characteristics of the observed daily precipitation data. The SDGAM also outperformed the SDSM in both calibration and validation periods giving CanESM2 and HadCM3. The next step is to conduct daily precipitation projections under different climate change scenarios. In this section the graphical displays are used to show the comparisons among the daily observed precipitation data, the downscaled precipitation series from historical GCMs by SDGAM (after bias correction), and the precipitation from the regional climate model CanRCM4. These graphical displays are presented for four representative indices: CDD, Prec90p, AMP and TAP. The median, maximum, and minimum, as well as the IQR of 100 SDGAM downscaling precipitation series sets, are calculated for the displays. The median of 100 different downscaled precipitation series smooths out individual series, which seems to lead to less variability compared to the observed data.

Figure 4-33 indicates that the observed Consecutive Dry Days (CDD) almost fall within the range bounded by the minimum and maximum values of CDD obtained from SDGAM. For Quebec City (S2) station, the CDD derived from downscaling is slightly less than the CDD from the observed daily precipitation data. Both the precipitation from CanRCM4 and downscaled from GCM historical data through SDGAM struggle to capture the observed CDD events approaching 30 days. This indicates that such long CDD events in the future would also be difficult for the models to predict. Figure 4-34 shows that the observed Prec90p fits well within the

range bounded by the minimum and maximum values of Prec90p obtained from SDGAM and aligns well with the median Prec90p from SDGAM. Figure 4-35 indicates that the observed AMP falls within the range bounded by the minimum and maximum values of AMP obtained from SDGAM and aligns well with the median AMP from SDGAM. Figure 4-36 shows that the TAP obtained from SDGAM for Montreal (S1) station mostly ranges between 700mm and 1300mm, and for Quebec City (S2) station, it mostly ranges between 800mm and 1600mm. This aligns well with the observed TAP.

The historical precipitation projections show that the precipitation series obtained from SDGAM are very close to the observed data and the precipitation from CanRCM4 in terms of CDD, Prec90p, AMP, and TAP. This lays a solid foundation for the future precipitation projections under different climate change scenarios.

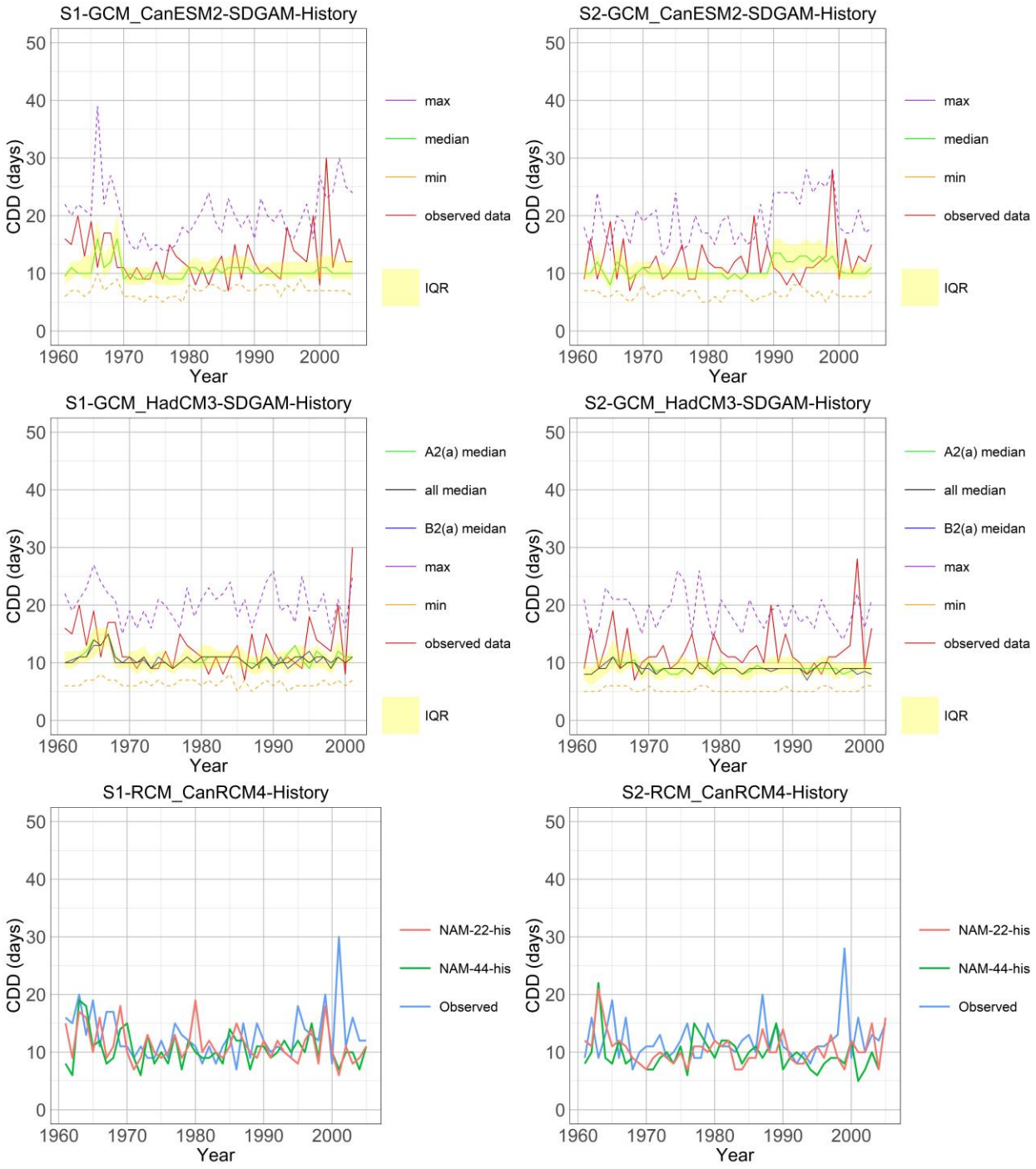

Figure 4-33. Time series plots of CDD from historical projection for Montreal (S1) station and Quebec City (S2) station

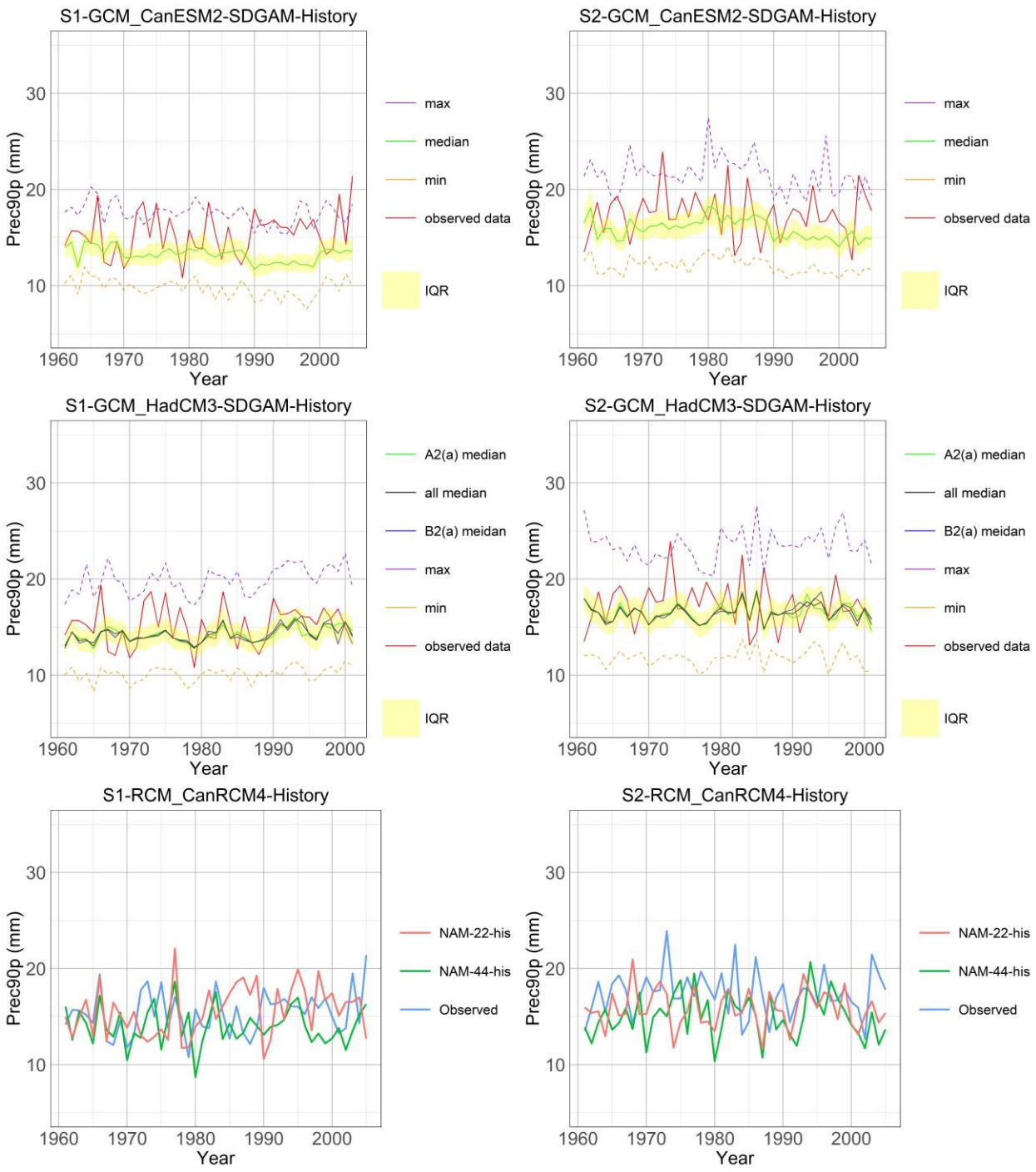

Figure 4-34. Time series plots of Prec90p *from historical projection for Montreal (S1) station and Quebec City (S2) station*

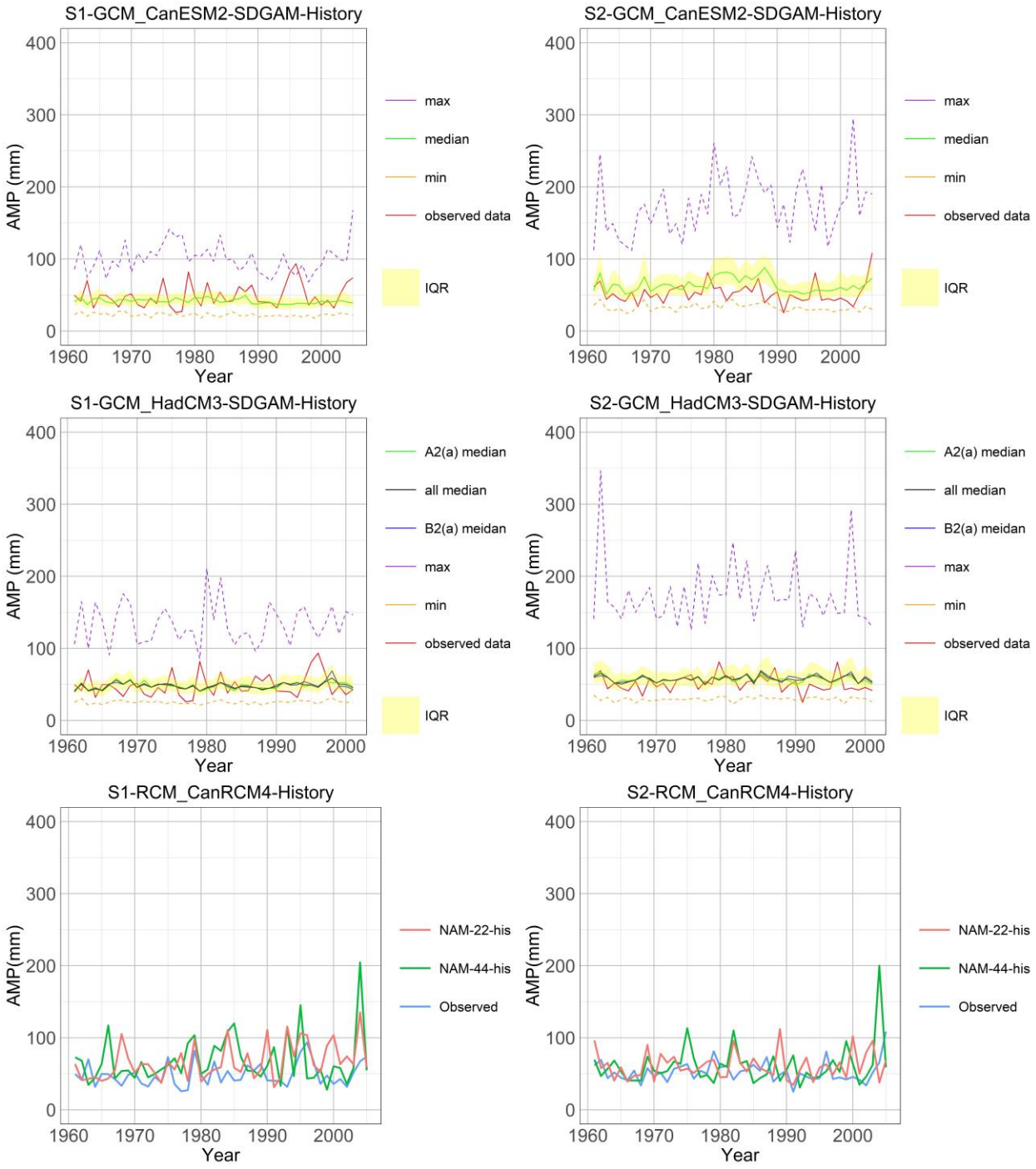

Figure 4-35. Time series plots of AMP from historical projection for Montreal (S1) station and Quebec City (S2) station

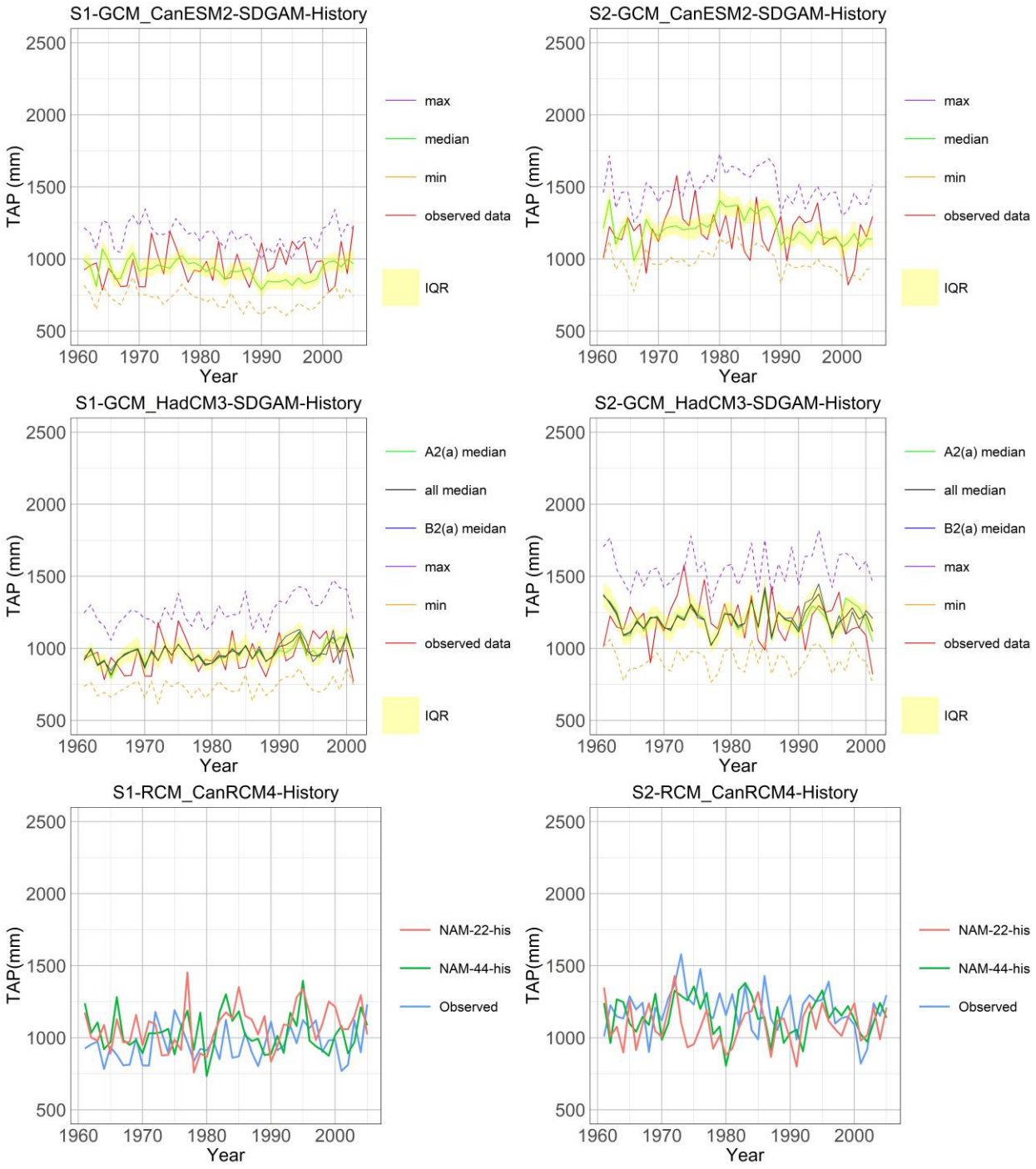

Figure 4-36. Time series plots of AMP from historical projection for Montreal (S1) station and Quebec City (S2) station

4.2.2 Precipitation Projections for Future Periods

In this section, the future precipitation projections are carried out under different climate change scenarios given by the CanESM2 (RCP2.6, RCP4.5, and RCP8.5) and the HadCM3 (A2 and B2) using the SDGAM (after bias correction). These projections are also compared with the precipitation from CanRCM4 and the results are presented for four representative indices: CDD, Prec90p, AMP and TAP.

Figure 4-37 shows that the CDD for Montreal (S1) station and Quebec City (S2) station under different future scenarios is most likely to be around 10 days, with no significant overall trend until 2070s. However, it is noticed that from 2075 to 2100, Montreal's CDD under the RCP 8.5 scenario shows a significant increasing trend, potentially exceeding 30 days. Figure 4-38 shows that the future Prec90p of Montreal (S1) station is most likely around 15mm, and the Prec90p of Quebec City (S2) station is most likely around 18mm. The Prec90p obtained under different scenarios shows little difference. Additionally, there may be a slight increasing trend in Prec90p for both locations towards the end of 21st century. Figure 4-39 shows that the future AMP obtained from CanRCM4 frequently exceeds 100mm. However, the future AMPs obtained using SDGAM for Montreal (S1) station and Quebec City (S2) station are mostly between 40mm and 80mm. From the perspective of the maximum value (purple dash), the AMP can reach 150mm or even exceed 200mm. It should be noted that under RCP4.5 and RCP8.5 scenarios, some AMP values downscaled from CanESM2 exceed 300mm and even higher for Quebec City (S2) station after 2030s. This could be attributed to more frequent and extreme precipitation events due to climate change, or it could be ascribed to inaccuracies in the model, warranting further investigation. Figure 4-40 shows a slight potential increasing trend in TAP for

Montreal (S1) station and Quebec City (S2) station in the 21st century. The TAP obtained under different scenarios does not exhibit significant differences.

In conclusion, future precipitation projections through SDGAM can reflect the high level of uncertainty in climate simulations. In contrast, while projections from RCMs do not address these uncertainties, they have the advantage of being based on more realistic physical processes. Therefore, the use of both statistical and dynamic downscaling approaches is highly recommended in engineering practice. The future precipitation projections under different climate change scenarios are valuable information for various types of climate change impact and adaptation studies such as for studying future droughts, floods, water resource allocation in Quebec.

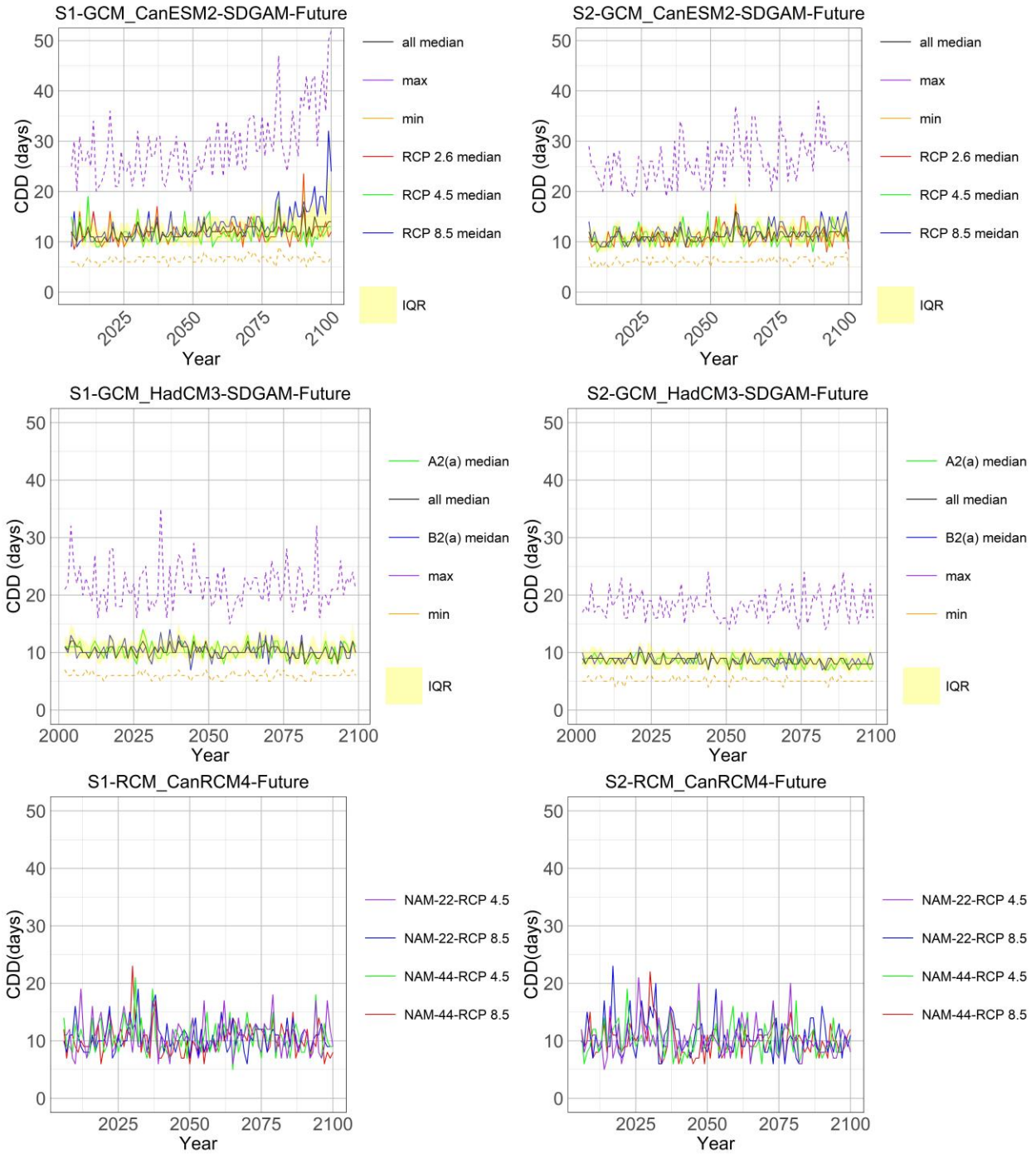

Figure 4-37. Time series plots of CDD from future projection for Montreal (S1) station and Quebec City (S2) station

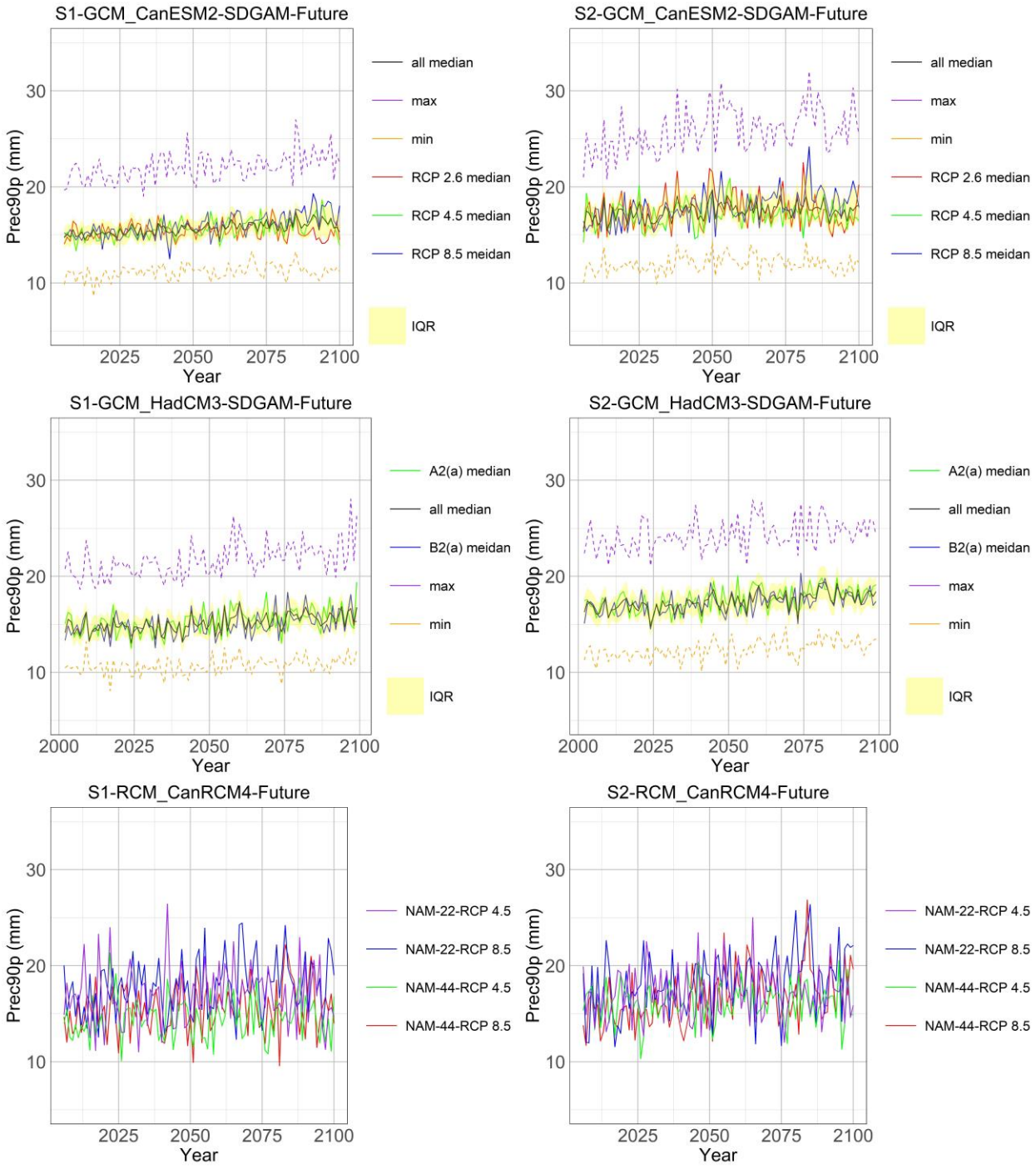

Figure 4-38. Time series plots of Pre90p *from future projection for Montreal (S1) station and Quebec City (S2) station*

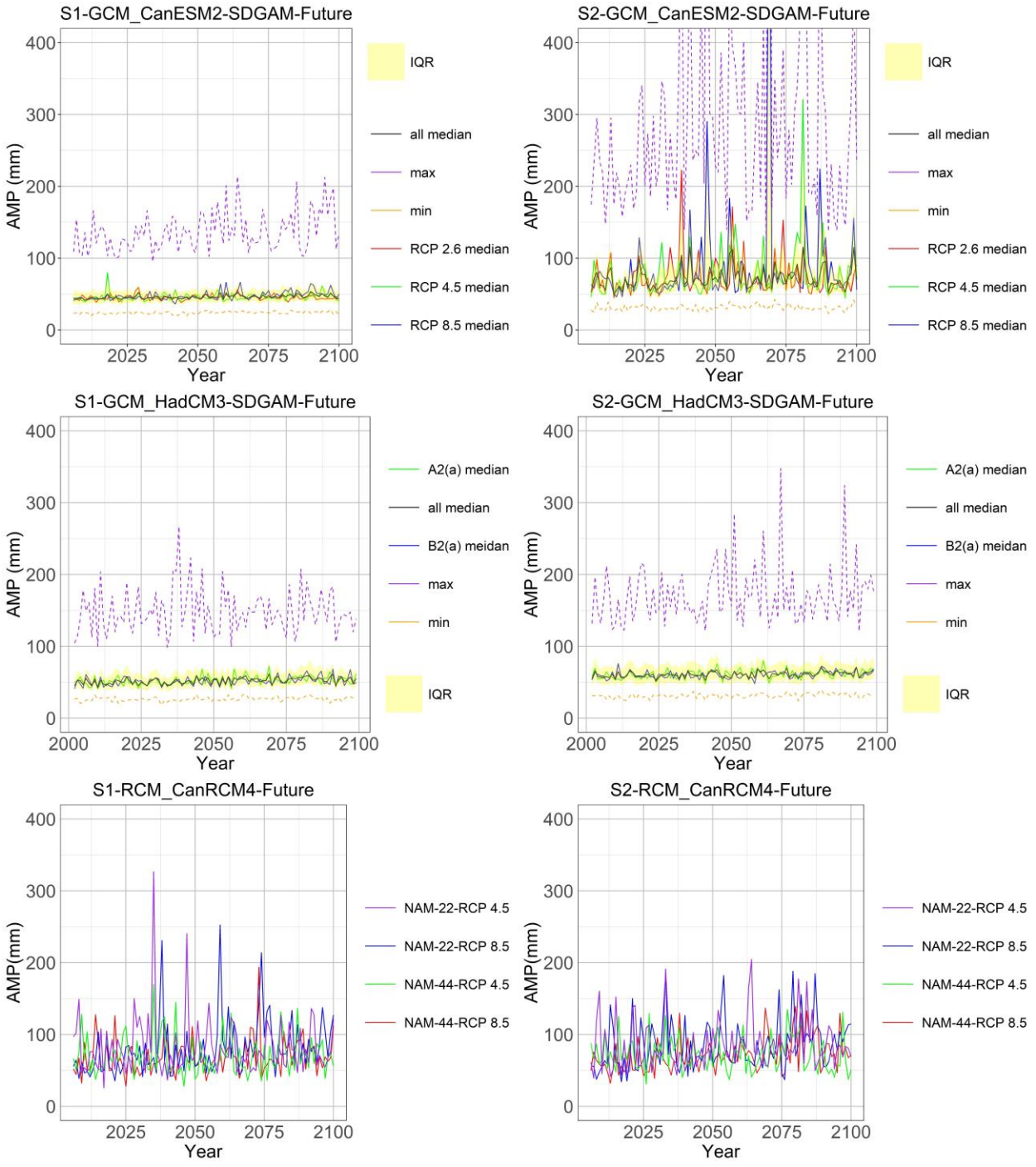

Figure 4-39. Time series plots of AMP from future projection for Montreal (S1) station and Quebec City (S2) station

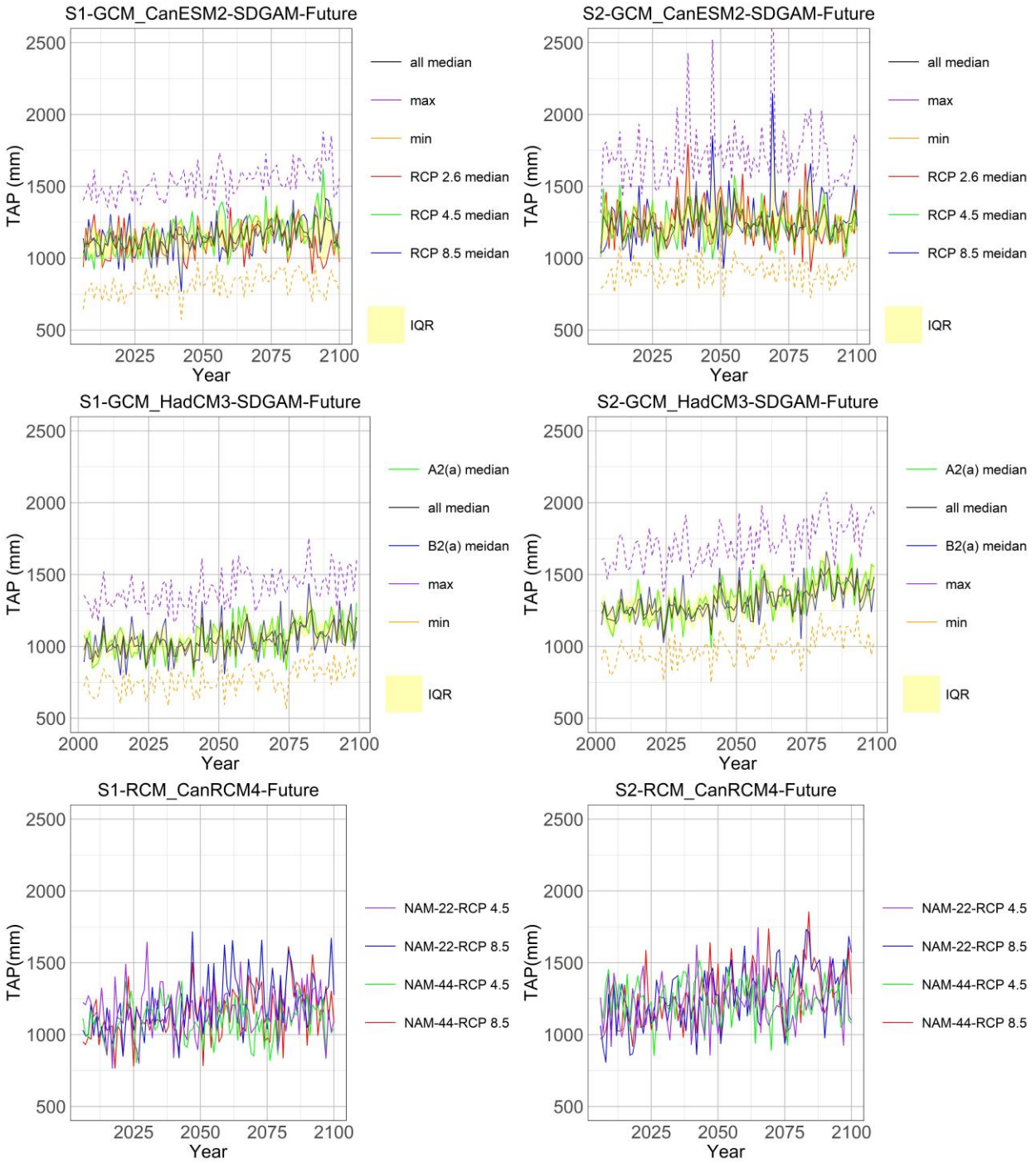

Figure 4-40. Time series plots of TAP from future projection for Montreal (S1) station and Quebec City (S2) station

5 Conclusions and Recommendations

5.1 Conclusions

The following main conclusions can be drawn from the present research:

1. This study proposed an improved method that integrates the partial correlation, backward stepwise regression, scatter plots, and spline function plots to select the significant large-scale predictors for downscaling of daily precipitation process at a given local site. Results of the numerical application using the precipitation records for two stations S1 and S2 in Quebec have indicated the feasibility of the proposed procedure. It has been demonstrated that the selected predictors can adequately and sufficiently explain observed local precipitation characteristics both statistically and physically.
2. This study developed a new statistical downscaling model (SDGAM) to simulate the daily precipitation process at a single site in the context of climate change. The proposed SDGAM was a combination of two components to describe the daily precipitation occurrence process and the daily precipitation amount process. The proposed SDGAM was based on the Generalized Additive Modeling (GAM) method to capture the nonlinear relationships between the large-scale climate predictors and the local precipitation. Results of an illustrative application using the observed daily precipitation records available at two rain-gauge stations located in Quebec and the National Center for Environmental Prediction (NCEP) re-analysis data that are interpolated for two GCMs (Canadian CanESM2 and UK HadCM3) have indicated the feasibility and accuracy of the proposed SDGAM.

3. A systematic evaluation procedure has been introduced in this study to assess the performance of the proposed SDGAM and the popular SDSM using various numerical and graphical performance criteria. It has been demonstrated that the SDGAM model was able to more accurately describe many features of the daily precipitation process, including its occurrence frequency, intensity, and extremes as compared to the SDSM.
4. The proposed SDGAM was used to make future precipitation projections for two locations in Quebec under different climate change scenarios given by two GCMs: Canadian CanESM2 and UK HadCM3. It was found that the precipitation characteristics, especially for the extremes, show some changes over the century for both locations. These precipitation projections under different climate change scenarios are valuable information for various types of climate change impact and adaptation studies in practice.

5.2 Recommendations for Future Studies

The future studies can be focus on the following aspects:

1. In this study, stepwise regression and partial correlation were used to select predictors. These methods are widely applied to identify predictors with significant linear relationships with the predictand. However, considering that SDGAM can capture nonlinear relationships between the predictand and predictors, the optimal selection of predictor sets for simulating local precipitation still warrants further investigation.
2. For Montreal (S1) station and Quebec City (S2) station, the assessment results have indicated the superior performance of SDGAM as compared to SDSM. Future research could be conducted to further examine the performance of

SDGAM for different regions with diverse climatic conditions in Canada and globally.

3. In this study, the most fundamental form of Generalized Additive Models (GAM) was employed. However, GAM is not limited to this basic application; it is a comprehensive collection of various mathematical tools. Future research can further explore the other forms of GAM to describe the relationship between local precipitation and large-scale atmospheric predictors.
4. Future research can consider the spatio-temporal correlation structure when using SDGAM for downscaling precipitations at different sites.
5. For further research on future precipitation projections, the Mann-Kendall test (Mann, 1945) can be used to examine trends in various precipitation indices. Additionally, future studies should focus on anticipated extreme weather events, such as over 30 CDD and more than 200mm AMP, to better understand their occurrence and impacts.

References

Afifi, A., May, S., & Clark, V. A. (2003). *Computer-aided multivariate analysis*. CRC press.

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., & Merryfield, W. J. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, *38*(5).

Artlert, K., & Chaleeraktrakoon, C. (2013). Modeling and analysis of rainfall processes in the context of climate change for Mekong, Chi, and Mun River Basins (Thailand). *Journal of Hydro-Environment Research*, 7(1), 2-17.

BC Hydro. (2023). *Site C Clean Energy Project*.

Benestad, R. E., Chen, D., Mezghani, A., Fan, L., & Parding, K. (2015). On using principal components to represent stations in empirical–statistical downscaling. *Tellus A: Dynamic Meteorology and Oceanography*, *67*(1), 28326.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), 211-243.

Bürger, G. (1996). Expanded downscaling for generating local weather scenarios. *Climate Research*, 7(2), 111-128.

Cavazos, T., & Hewitson, B. (2002). *Relative performance of empirical predictors of daily precipitation*.

Choux, M. (2005). Development of new predictor variables for the statistical downscaling of precipitation. *Degree Master of Engineering, Dept of Civil Engineering and Applied Mechanics, McGill University*.

Conway, D., Wilby, R. L., & Jones, P. D. (1996). Precipitation and air flow indices over the British Isles. *Climate Research*, 7(2), 169-183.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., & Eyring, V. (2014). Evaluation of climate models. In *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741–866). Cambridge University Press.

Gachon, P. (2005). FINAL REPORT "A first evaluation of the strength and weaknesses of statistical downscaling methods for simulating extremes over various regions of eastern Canada." *Meteorological Service of Canada: Toronto, ON, Canada*.

Goodess, C. M., & Palutikof, J. P. (1998). Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *18*(10), 1051–1083.

Government of Canada. (2020). *Canada and Quebec invest in water infrastructure to ensure adequate services and stimulate the economy in the Chaudière-Appalaches region*.

Government of Manitoba. (2024). *Lake Manitoba and Lake St. Martin Outlet Channels Project*.

Harpham, C., & Wilby, R. L. (2005). Multi-site downscaling of heavy daily precipitation occurrence and amounts. *Journal of Hydrology*, *312*(1-4), 235–255.

Hessami, M., Gachon, P., Ouarda, T. B., & St-Hilaire, A. (2008). Automated regression-based statistical downscaling tool. *Environmental Modelling & Software*, *23*(6), 813–834.

Huth, R. (1999). Statistical downscaling in central Europe: Evaluation of methods and potential predictors. *Climate Research*, *13*(2), 91–101.

Jones, M. R., Blenkinsop, S., Fowler, H. J., Stephenson, D. B., & Kilsby, C. G. (2013). Generalized additive modelling of daily

precipitation extremes and their climatic drivers. *National Center for Atmospheric Research, Colorado*.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., & Woollen, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472.

Karl, T. R., Wang, W.-C., Schlesinger, M. E., Knight, R. W., & Portman, D. (1990). A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *Journal of Climate*, *3*(10), 1053–1079.

Khalili, M., & Van Nguyen, V. T. (2017). An efficient statistical approach to multi-site downscaling of daily precipitation series in the context of climate change. *Climate Dynamics*, *49*(7–8), 2261–2278.

Khalili, M., Van Nguyen, V. T., & Gachon, P. (2013). A statistical approach to multi-site multivariate downscaling of daily extreme temperature series. *International Journal of Climatology*, *33*(1), 15–32.

Kilsby, C. G., Cowpertwait, P. S. P., O'connell, P. E., & Jones, P. D. (1998). Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *18*(5), 523–539.

Laanaya, F., St-Hilaire, A., & Gloaguen, E. (2017). Water temperature modelling: Comparison between the generalized additive model, logistic, residuals regression and linear regression models. *Hydrological Sciences Journal*, *62*(7), 1078–1093.

Lee, M. Y. (2013). *Statistical modeling of extreme rainfall processes in the context of climate change* [McGill University Libraries].

Legg, S. (2021). IPCC, 2021: Climate change 2021-the physical science basis. *Interaction*, *49*(4), 44–45.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, 245–259.

McCuen, R. H. (2016). *Modeling hydrologic change: Statistical methods*. CRC press.

Mearns, L. O., Giorgi, F., Whetton, P., Pabon, D., Hulme, M., & Lal, M. (2003). Guidelines for use of climate scenarios developed from regional climate model experiments. *Data Distribution Centre of the Intergovernmental Panel on Climate Change*, *38*.

Nakicenovic, N., Alcamo, J., Davis, G., Vries, B. de, Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. Y., & Kram, T. (2000). *Special report on emissions scenarios*.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Royal Statistical Society. Journal. Series A: General*, *135*(3), 370–384.

Nguyen, H. L. (2023). *Statistical Modeling of Precipitation Processes in the Context of Climate Change* [McGill University Libraries].

Nguyen, V. T. V., & Nguyen, T. D. (2008). A spatial-temporal statistical downscaling approach to estimation of extreme precipitations for climate-related impact studies at a local site. *World Environmental and Water Resources Congress 2008: Ahupua'A*, 1–10.

Nguyen, V.-T.-V., & Giorgi, F. (2022). Climate change projections for impact and adaptation studies at the urban watershed scale. *Hydrolink*, *3*, 72–76.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., & Dasgupta, P. (2014). *Climate change 2014: Synthesis report. Contribution of Working Groups I, II and III to the fifth*

assessment report of the Intergovernmental Panel on Climate Change. Ipcc.

Pharasi, S. (2006). *Development of statistical downscaling methods for the daily precipitation process at a local site* [McGill University Libraries].

Richardson, C. W., & Wright, D. A. (1984). WGEN: A model for generating daily weather variables. *Report ARS-8 August 1984. 83 p, 3 Fig, 12 Tab, 13 Ref, 4 App.*

Semenov, M. A., & Barrow, E. M. (1997). Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, *35*(4), 397–414.

Serifi, A., Günther, T., & Ban, N. (2021). Spatio-temporal downscaling of climate data using convolutional and error-predicting neural networks. *Frontiers in Climate*, *3*, 656479.

Tareghian, R., & Rasmussen, P. F. (2013). Statistical downscaling of precipitation using quantile regression. *Journal of Hydrology*, *487*, 122–135.

Team, Rs. (2021). *RStudio: Integrated development for R. RStudio, PBC, Boston, MA. 2020*.

Vandal, T., Kodra, E., & Ganguly, A. R. (2019). Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation. *Theoretical and Applied Climatology*, *137*, 557–570.

Villarini, G., & Serinaldi, F. (2012). Development of statistical models for at-site probabilistic seasonal rainfall forecast. *International Journal of Climatology*, *32*(14), 2197–2212.

Von Salzen, K., Scinocca, J. F., McFarlane, N. A., Li, J., Cole, J. N., Plummer, D., Verseghy, D., Reader, M. C., Ma, X., & Lazare, M. (2019). The Canadian fourth generation atmospheric global climate model (CanAM4). Part I: representation of physical processes. In *Data, Models and Analysis* (pp. 123–144). Routledge.

Von Storch, H. (1999). On the use of "inflation" in statistical downscaling. *Journal of Climate*, *12*(12), 3505-3506.

Wilby, R. L. (1998). Statistical downscaling of daily precipitation using daily airflow and seasonal teleconnection indices. *Climate Research*, *10*(3), 163-178.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., & Mearns, L. O. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting Material of the Intergovernmental Panel on Climate Change, Available from the DDC of IPCC TGCIA*, *27*.

Wilby, R. L., & Dawson, C. W. (2004). Using SDSM version 3.1—A decision support tool for the assessment of regional climate change impacts. *User Manual*, *8*, 1-7.

Wilby, R. L., & Dawson, C. W. (2013). The Statistical DownScaling Model: Insights from one decade of application. *International Journal of Climatology*, *33*(7), 1707-1719.

Wilby, R. L., Dawson, C. W., & Barrow, E. M. (2002). SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, *17*(2), 145-157.

Wilby, R. L., Hassan, H., & Hanaki, K. (1998). Statistical downscaling of hydrometeorological variables using general circulation model output. *Journal of Hydrology*, *205*(1-2), 1-19.

Wilby, R. L., & Wigley, T. M. L. (2000). Precipitation predictors for downscaling: Observed and general circulation model relationships. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *20*(6), 641-661.

Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: A review of stochastic weather models. *Progress in Physical Geography*, *23*(3), 329-357.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. CRC press.

Wood, S., & Wood, M. S. (2015). Package 'mgcv.' *R Package Version*, *1*(29), 729.

Xu, G., Xu, X., Liu, M., Sun, A. Y., & Wang, K. (2015). Spatial downscaling of TRMM precipitation product using a combined multifractal and regression approach: Demonstration for South China. *Water*, *7*(6), 3083–3102.

Yeo, M.-H. (2014). *Statistical modeling of precipitation processes for gaged and ungaged sites in the context of climate change* [McGill University Libraries].

Yeo, M.-H., Nguyen, H.-L., & Nguyen, V.-T.-V. (2021). Statistical tool for modeling of a daily precipitation process in the context of climate change. *Journal of Water and Climate Change*, *12*(1), 18–31.

Copyright Notice

© Chenglong You, 2024. All rights reserved.

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science at McGill University and is the result of my own work. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, without the prior written permission of the copyright holder, except in the case of brief quotations embodied in critical reviews and certain other non-commercial uses permitted by copyright law.