School of Civil and Mechanical Engineering

CMIP5 Decadal Precipitation at Catchment Level and Its Implication to Future Prediction

Md Monowar Hossain

0000-0002-1814-2299

This thesis is presented for the Degree of

Doctor of Philosophy

of

Curtin University

April 2022

THESIS DECLARATION

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material, which has been accepted for the award of any other degree or diploma in any university.

Signature……

Date ……

DEDICATION

This thesis is dedicated to my adorable parents (Reanue Akter and Mr. Md Noyamia Akanda), my ever-loving and very caring wife (Mst Aslima Akter), and my beloved, riant, and very cute daughter (Mymunah Hossain).

This thesis is also dedicated to my maternal uncles who taught me and funded me as well at different levels of my education.

This thesis is dedicated to all of my teachers who have taught me at different levels of my institutional and non-institutional education.

For me, always trying to be a good practicing Muslim, a great believer in the wisdom of the Holy Quran, and an authentic follower of the last and final messenger Muhammad (peace be upon him).

ABSTRACT

Climate change has a significant impact on regional and local water resources. Different climate models (General Circulation Models, GCMs) have been developed across the world that provide the climate data of coarser spatial resolutions. Intergovernmental Panel on Climate Change (IPCC) has produced several assessment reports on climate change related issues using GCMs. Coupled Modelled Intercomparison Project phase-5 (CMIP5) in the IPCC's assessment report 5 (AR5) has predicted the climate data in decadal timescale (e.g., predicting ten years ahead) which attracted the climate research community due to its potential applications in many dimensions. However, the GCMs outputs contain systematic bias (drift) and their coarse spatial resolutions are inadequate for the local or catchment level applications. Most of the previous studies on decadal climate data were based on temperature or temperature-based climate indices. No study assessed CMIP5 decadal precipitation at a catchment level for a spatial resolution finer than 0.5^0 (50km × 50km). This is the first study that assessed the CMIP5 decadal hindcast monthly precipitation at a catchment level of 0.05^0 spatial resolution (5km x 5km) and showed its application for the prediction of precipitation ten years ahead (i.e., decadal-scale). For catchment level predictions, stochastic or statistical models including different forms of artificial neural networks (ANN) are commonly used in previous studies where only historical observed data are used to predict the future. Until now, no study used the GCMs derived decadal precipitation data together with the observed data for future prediction at the catchment level. This study predicted monthly precipitation for a ten-year period using GCMs and observed data through Facebook Prophet (FBP), ANN, and machine learning (ML) regression algorithms.

In CMIP5 experiments, there are two core sets of decadal data available such as 10-year and 30-year simulations and 10 GCMs contributed to reproducing the monthly hindcast precipitation in decadal scale. Out of these 10 GCMs, eight GCMs (MIROC4h, EC-EARTH, MRI-CGCM3, MPI-ESM-MR, MPI-ESM-LR, MIROC5, CMCC-CM, and CanCM4) were selected in this study based on their resolution pattern and both 10-year and 30-year simulations data were collected from CMIP5 data portal for the period 1961-2015. The observed gridded ($0.05^0 \times 0.05^0 \approx$ 5km×5km) data were collected from the Australian Bureau of Meteorology. The spatial resolution of these 8 GCMs datasets ranges from 0.5625^0 to 2.8125^0 (56.25km to 281.25km) which were then interpolated onto $0.05^0 \times 0.05^0$ (5km×5km) spatial resolution matching with the grids of observed data using eight spatial interpolation methods (linear,

bilinear, distance weighted average, inverse-distance weighted average, nearest neighbour, first-order conservative, second-order conservative, and bi-cubic). The wide range of interpolation methods were used to select the most suitable method for the use of decadal precipitation data at the catchment scale. The interpolated datasets were then subset for the Brisbane River catchment in Queensland, Australia. Brisbane River catchment was selected for this study because of its tropical climate nature and low to moderate yearly rainfall variability. All analyses were done for this selected catchment through several skill tests that include correlation coefficient (CC), Pearson correlation coefficient (PCC), anomaly correlation coefficient (ACC), index of agreement (IA), fractional skill scores (FSS), mean absolute error (MAE), root-mean-square error (RMSE), and total precipitation over time and space.

Based on the skill test results, the second-order conservative (SOC) was found as the most suitable spatial interpolation method for the GCMs derived gridded data as it conserves the precipitation flux during the interpolation process. The 10-year simulation data was found better than the 30-year simulation because of its shorter lead-time and higher number of ensembles. Hence, 10-year simulation data interpolated with SOC method were used in the subsequent analyses. The catchment level GCM data shows significant model bias (drift) which was quantified comparing with the observed data. This shows the necessity of drift correction of interpolated data for using it in the catchment scale for practical purposes. The model bias was investigated for monthly and seasonal (mean of three months considering four seasons in a year) data and the results revealed lower drifts for seasonal precipitation. Next, different drift correction alternatives (Nested bias correction, Standardization and re-scaling approach, Relative drift correction or linear scaling, and a Modified drift correction method) were investigated for seasonal data to select a suitable drift correction method. The results revealed that the new modified drift correction method proposed in this study shows better performances, especially for the models with higher drifts for seasonal data. However, there remains a necessity for developing a forecast model for monthly precipitation on a decadal scale. For this, at first, a method of formulation of suitable multi-model ensemble mean (MMEM) has been proposed in this study by categorizing the selected GCMs as there is no such formulation in the literature. Considering this suitable MMEM, forecast models for monthly precipitation in decadal-scale have been developed through FBP, ML Regression algorithms and a deep ANN (Bidirectional LSTM) using the CMIP5 decadal data and the corresponding observed values. In FBP, MMEM and all individual models were used as

additional regressor in addition to the observed values. In ANN and ML algorithms, MMEM was used as a feature and the corresponding observed values were used as target variables following a supervised training approach. The results revealed that the new ANN and FBP model developed in this study using the GCM and observed data could predict the monthly rainfall on a decadal scale comparatively better than the models forecasted using the observed data only. This was because of following a supervised training approach in the ANN model and multiplicative seasonality function along with other tuneable parameters in FBP models that enabled both models to reproduce the dry events more accurately. Comparing the prediction skills, FBP was found a little better than the ANN model whereas ANN was found comparatively better in reproducing the wet events. However, the overall methodology presented herein and the information on rainfall prediction in decadal-scale (ten years ahead) found in this research will be very useful for managing future water resources, agricultural practices, farming processes, and other water-related infrastructures. This study was conducted for one catchment only but the method developed in this study could be used in any catchment for predicting future rainfall for a decadal time scale.

ACKNOWLEDGEMENT

This thesis got its final shape with the guidance, cooperation, and support from several individuals who contributed from different sources.

Foremost, I would like to express my sincere gratitude to my principal supervisor A/Prof. A. H. M. Faisal Anwar for his patience, guidance, continuous support, motivation, enthusiasm, and critical feedback in every stage of my research activities. He is the first person who taught me how to do quality research and disseminate the research outcome to the audience. He also taught me how to address the comments from reviewers as well as to keep patience while addressing the comments.

My immense gratitude to my Associate supervisor Dr Mahesh Prakash, Senior Principal Research Scientist and team leader at CSIRO Data61, Melbourne. I appreciate his thoughtful advice on learning Python, machine learning, deep learning, and climate data operator (CDO) tools. I am thankful for his critical feedback on different steps of my research activities and for supporting me throughout my PhD journey.

I am deeply grateful to Dr Mohammed Bari, Supervising Chief Hydrologist at the Bureau of Meteorology, Western Australia. Bari has guided me with the research idea, given critical feedback to tune the research quality and presenting the outcome.

I owe my deepest gratitude to Dr Nikhil Garg, a research scientist at CSIRO Data61 who has taught me Python, initiated with the data processing, and provided critical feedback on presenting my research outcome. He also provided me all up to date information in the field of my PhD topic with different literature. I am grateful to Nikhil for pushing me to go forward and keeping me on my research track throughout the journey of my PhD study.

I also would like to thank my supervisory panel and industry collaborators for providing support and critical feedback on my research progress in the fortnightly online meeting held in WebEx.

I am deeply grateful and sincerely acknowledge my fund providers, Curtin University and Data61of Commonwealth Scientific and Industrial Research Organization (CSIRO) who made it possible to conduct this research at Curtin University in collaboration with CSIRO and Bureau of Meteorology, WA. I am thankful to all the academic and non-academic members of

the Department of Civil Engineering, Curtin University, for providing admirable research support and facilities. I am also thankful to the Australia Bureau of Meteorology for providing their enormous support along with the observed data and catchment shapefiles.

I am extremely grateful to my parents and my maternal uncles for their love, prayer, caring, and sacrifices for encouraging, educating, and preparing me for my future. I am also grateful to my beloved and ever-caring wife (Mst Aslima Akter), who sacrificed a lot throughout my study period in Curtin by giving me mental support, managing my family, and taking care of my health. She also played the role of my fellow to discuss my problems sharing my mental stress as well as other research-related issues.

Finally, I am very much grateful to my lord for his endless blessings and for providing me a good mental and physical health that powered me to go through my whole PhD study journey at Curtin University.

LIST OF PUBLICATIONS

Out of this research, two Q1 journal articles (IF: 4.07 and 3.38), one Q2 (IF: 3.12), one journal article in Indian Water Resources Society, and one conference article are published. One journal article is under review and one more draft will be submitted soon. The copyright permissions of published articles are provided in Appendix A. In all of these articles, I am the first author and the main contributor (Conceptualization; data curation; formal analysis; investigation; methodology; software; visualization; writing-original draft; writing review and editing). A/Prof. A. H. M. Faisal Anwar (Curtin University) and Dr Mahesh Prakash (Data61 CSIRO) are my principal and associate supervisors respectively who guided me all through these publications and hence they are the co-authors for all papers. As an industry collaborator, Dr Nikhil Garg (Data61 CSIRO) contributed and helped me for data processing and analysis and that is why he has been added as one of the co-authors. Dr Mohammed Bari of Bureau of Meteorology, WA is another industry collaborator and he has contributed by providing advice and critically reviewing the articles and giving the feedback. That is why Dr Bari was also a co-author for several papers. However, separate written statements (signed by each co-authors) are included in Appendix B.

Journal Articles:

(i) Hossain, M. M, Garg, N., Anwar, A.H.M.F., Prakash, M., 2021. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale, *Journal Indian Water Resources Society* 41, 2, 28-34. http://iwrs.org.in/journal/apr2021/5apr.pdf.

(ii) Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021. Drift in CMIP5 decadal precipitation at catchment level, *Stochastic Environmental Research and Risk Assessment 8, 5*. https://doi.org/10.1007/s00477-021-02140-8

(iii) Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2022. Intercomparison of drift correction alternatives for CMIP5 decadal precipitation, *International Journal of Climatology*, 42, 1015-1037. https://doi.org/10.1002/joc.7287 (First published online: 05 July 2021)

(iv) Hossain, M M, Anwar, A.H.M.F., Garg, N., Prakash, M., Bari, M., 2022. Evaluation of CMIP5 decadal precipitation at catchment level, *International Journal of Climatology* (Under review)

(v) Hossain, M.M., Anwar, A.H.M.F., Garg, N., Prakash, M., Bari, M., 2022. Monthly Rainfall Prediction at Catchment Level with the Facebook Prophet Model Using Observed and CMIP5 Decadal Data. Hydrology 9, 111. https://doi.org/10.3390/hydrology9060111

(vi) Hossain, M.M., Anwar, A.H.M.F., Garg, N., Prakash, M., Bari, M., 2022. Monthly rainfall prediction for decadal timescale using Bidirectional LSTM and CMIP5 near-term experiment data. *Journal of Advances in Modelling Earth Systems.* (To be submitted)

Conference proceedings:

(i) Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021. A comparative study on 10 and 30-year simulation of CMIP5 decadal hindcast precipitation at catchment level, in: Vervoort, R.W., Voinov, A.A., Evans, J.P. and Marshall, L. (Ed.), MODSIM2021, *24th International Congress on Modelling and Simulation.* Modelling and Simulation Society of Australia and New Zealand, pp. 609–615. https://doi.org/10.36334/modsim.2021.K5.hossain

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACC	:	Anomaly correlation coefficient
ANN	:	Artificial neural network
AWAP	:	Australian Water Availability Project
BoM	:	Bureau of Meteorology
BIC	:	Bi-cubic
BiLIN	:	Bi-linear
BiLSTM	:	Bidirectional Long-Short term Memory
BiLSTM*	:	Bidirectional Long-Short term Memory for observed data after removing outliers
CC	:	Correlation coefficient
CDF	:	Cumulative distribution function
CDO	:	Climate data operator
CMIP5	:	Coupled Modelled Intercomparison Project phase 5
DJF	:	December-January-February
DWA	:	Distance weighted average
FBP	:	Facebook Prophet Model
FOC	:	First-order conservative
FSS	:	Fractional skill scores
FSSa85	:	Fractional skill scores for above the 85 percentile values
FSSb15	:	Fractional skill scores for below the 85 percentile values

GCMs	:	General circulation models
IDW	:	Inverse distance weighted average
IA	:	Index of agreement
IPCC	:	Intergovernmental Panel on Climate Change
JJA	:	June-July-August
LGB	:	Light gradient boosting
LIN	:	linear
MAE	:	Mean absolute error
MAM	:	March-April-May
MDM	:	Modified method
MoD	:	Model
MMEM	:	Multi-model ensemble mean
MLP	:	Multi-layer perceptron
NBC	:	Nested bias correction method
NN	:	Nearest Neighbour
OK	:	Ordinary Kriging
PCC	:	Pearson correlation
Pr.	:	Precipitation
RAW	:	Model raw values
RBF	:	Radial basis function
RDF	:	Random forest regressor

RDT	:	Relative drift correction method
RCMs	:	Regional climate models
RMSE	:	Root mean squared error
SOC	:	Second-order Conservative
SON	:	September-October-November
STC	:	Stacking model
SVR	:	Support vector regression
XGB	:	Extreme gradient boosting

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Overview

The Coupled Model Intercomparison Project Phase 5 (CMIP5) is a standard experimental protocol established by the Working Group on Coupled Modelling (WGCM) under the World Climate Research Programme (WCRP). CMIP5 provides a wide range of future climate data derived through general circulation models (GCMs). To support the Intergovernmental Panel for Climate Change (IPCC) Fifth Assessment Report (AR5), the CMIP5 were designed with the three suites of experiments; (i) decadal hindcast and predictions simulations, (ii) 'long-term' simulations, and (iii) 'atmosphere-only' simulations for, especially computationally demanding models. There are two core sets of decadal hindcasts and predictions, (i) 10-year simulations (hindcasts or prediction) initialized based on the climate state in 1960, 1965 and thus every five years until 2015, (ii) 30-year simulations (hindcasts or prediction) initialized in 1960, 1980 and 2005 (Taylor et al., 2012). The observational records were prescribed as external forcing to the GCMs for generating the hindcasts data whereas the mid-range scenario (RCP4.5) was adopted for future predictions. Two common approaches of initialization are used in CMIP5 decadal experiments; full-field initialization and anomaly initialization. In full-field initializations, models' initial state is forced away from its equilibrium state to match as close as possible to its observed climate state while in the anomaly initialization, observed anomalies are added to the model climatology. Multiple runs carried out either for hindcasts or forecasts, of the same model with slightly different initialized conditions are referred to as ensemble members.

Precipitation is a very important hydrological aspect and a precious natural resource for all forms of animal beings and ecosystems. It influences our livelihood and agriculture in many dimensions. An early prediction of precipitation has many positive benefits from socioeconomic perspectives as it enables more efficient management of agriculture, water resources, power development, and planning and development of infrastructure (Apurv et al., 2015; Hansen et al., 2011; J.W. Jones et al., 2000; Mehta et al., 2013). However, prediction of this important climate variable has become a very challenging task, in terms of accuracy, due to its chaotic nature over time and space. Moreover, due to ongoing climate change, the

temporal and spatial variations of precipitation have been intensified in the past few decades. Over the past few years, precipitation prediction has been paid much attention from the climate research community (Ali et al., 2019; George et al., 2016; Hossain et al., 2020; Hung et al., 2009; Mekanik et al., 2011; Mislan et al., 2015; Ouyang et al., 2016). The precipitation prediction approaches are broadly classified into two main categories; (i) a knowledge-driven approach, and (ii) a data-driven approach. A knowledge-driven approach uses a scientific understanding on hydrological processes, thermodynamic balance, laws of physics, and the interaction between earth, atmosphere, and ocean. Climate models such as General Circulation Models (GCMs) use the knowledge-driven approach for the future prediction of climate variables. However, the knowledge-driven approach needs extensive data and intensive computational facility that sometimes becomes unavailable (Hong, 2008).

The stochastic, empirical, or statistical models use the data-driven approach for future precipitation prediction. Stochastic or statistical models are mainly based on different forms of regression analysis such as simple regression analysis (SRA), exponential smoothing, decomposition, and auto-regressive integrated moving average (ARIMA). Every individual stochastic or statistical method has its strengths and weaknesses. For instance, ARIMA is a popular stochastic model for time series prediction with greater flexibility. But, as a stochastic model, it needs stationarity of the data (Machiwal and Jha, 2012) and its presumed linear form of the associated data sometimes makes it inappropriate for complex nonlinear time series data like precipitation (Zhang, 2003). However, technological development in combination with the research innovations in this modern arena enhanced the computation facility that enabled higher accuracy of precipitation prediction of which Artificial Neural Networks (ANN) is the best example. Applications of the machine and deep learning algorithms, of which ANN of the different forms of architecture, have been popular for many time series predictions including time series of precipitation (Hung et al., 2009; Lee et al., 2018; Meinke et al., 2007; Mekanik et al., 2011; Mislan et al., 2015) because of its enhanced prediction accuracy. ANN is capable of modelling complex nonlinear real-world problems. Based on the level of complexity, ANN can be combined with different types of algorithms due to its highly flexible character. However, based on the requirements, researchers have come up with different research interests and periods for the application of ANN. Though the ANN is good to capture the nonlinear relationship of data, the presence of outliers in the time-series data can critically affect the reliability of ANN as it is a grey box model (Unnikrishnan and Jothiprakash, 2020). Thus ANN requires proper data pre-processing before its application especially for the climatic data

(Committee, 2000; Ramírez et al., 2006).Climate change is an ongoing dynamic process that depends on many factors and it will also continue to change in the future (IPCC, 2014). But the rate of change in future climate is not certain. This is why the GCMs' simulations of the future climate variables for longer timescales ahead may become more uncertain compared to the simulations of shorter timescales. Among the three suits of experiments of CMIP5, the "decadal hindcasts and predictions simulations" has been received huge attention from the climate research community (Barsugli et al., 2009; Means et al., 2010). It is due to its potential applications in policymaking, urban planning, infrastructure management, agriculture and agriculture-dependent business (Barsugli et al., 2009; Crawford et al., 2006; Kirtman et al., 2013; Means et al., 2010; Meehl et al., 2014; Smith et al., 2012).

1.2 Background and problem statement

Climate change is an ongoing dynamic process that is being changed continuously and will continue to change in the future. However, the rate of future climate change and its potential impact on precipitation is not certain as it depends on numerous factors. Due to the ongoing climate change, precipitation has been affected more compared to the other climate variables. Changes in precipitation patterns, seasonal shifting, longer dry spells, and extreme wet events along with overall reduction of total precipitation amount have been observed around the globe in the last few decades. In the last decades, these changes have been intensified due to the higher rate of ongoing climate change (IPCC, 2014). According to the IPCC report, the change in the future total precipitation and its extreme events (e.g., heavy precipitation, droughts) will be higher compared to the past depending on the geographical locations. As every year the climate condition is being changed and it would be intensified in the future, researchers should not rely only on the data-driven approach for the future precipitation prediction at the local level. However, application of both the knowledge-driven and data-driven approaches together in predicting the precipitation at the local level is not a common practice whereas application of CMIP5 decadal data for the same has not been seen yet.

Applications of ANN, of different forms of architecture, following the data-driven approach, have been seen in many previous pieces of research (Hung et al., 2009; Lee et al., 2018; Meinke et al., 2007; Mekanik et al., 2011; Mislan et al., 2015) where only the historically observed data were used as input for predicting the local precipitation. However, there is no study in which both the GCMs derived hindcast and corresponding historically observed data are used

as ANN model input to follow a supervised training approach. It may be due to the matter of research interest or unavailability of the decadal datasets at the local level.

The GCMs provide precipitation data for the global scale with spatial resolutions of approximately 100-250 km. These spatial resolutions are inadequate for the studies at the local level, such as the catchment level, because of the lack of regional information (Fowler et al., 2007; Grotch and MacCracken, 1991; Salathé, 2003). The use of the regional climate models (RCMs) for transferring the GCMs derived precipitation to the local level is prevalent nowadays but albeit computationally intensive as well as not easily available everywhere. For this reason, the application of spatial interpolations for re-gridding the GCMs derived precipitation is commonly used in practice (Amengual et al., 2012; Mehrotra et al., 2014; Miao et al., 2016). Another reason to use the spatial interpolation methods is, transforming the GCMs' different native spatial resolutions to a unique spatial resolution, for the assessments and skill comparison of the GCMs. However, in most cases, spatial interpolation methods are randomly selected and clarification behind selecting the method was not provided. For instance, bilinear interpolation has been used in many studies (Amengual et al., 2012; Kamworapan and Surussavadee, 2019; Miao et al., 2016) but the reason behind selecting the bilinear method was not well explained. As the precipitation shows high spatial variability in frequency and magnitude, Wagner et al. (2012) suggested using such a spatial interpolation that will consider the spatial distribution of the precipitation over the entire study area. Therefore, the selection of an appropriate spatial interpolation method is important to provide the accurate spatial distribution of the precipitation while transforming GCMs data from coarser to a finer spatial resolution. However, there is no study evaluating the performance of spatial interpolation methods for GCMs derived gridded precipitation at catchment level. To consider both the knowledge and data-driven approach in predicting future precipitation at the catchment level, selection of the GCMs, suitable data pre-processing, and sorting them for the best outcome through different assessments are the prerequisite works.

In the last decade, the CMIP5 decadal experiment has attracted climate researchers due to its potential applications in many dimensions. As a result, the evaluation of CMIP5 decadal prediction has been run far from the early stage based on different evaluation aspects such as different regions, different climate variables, and their different spatial and temporal resolutions. For instance, Choi et al. (2016) investigated the prediction skill of CMIP5 decadal hindcast near-surface air temperature for the global scale while other researchers investigated

other climate variables in continental or regional scales (Gaetani and Mohino, 2013; Lovino et al., 2018; McKellar et al., 2013). Lovino et al. (2018) evaluated decadal hindcast precipitation and temperature over northern Argentina and reported higher skills of models to reproduce the temperature as opposed to precipitation where precipitation skills were found remarkably lower. McKellar et al. (2013) investigated decadal hindcast maximum and minimum temperature over the state of California and reported the best performing model. Likewise, Gaetani and Mohino (2013) evaluated model performances to reproduce Sahelian precipitation and reported better performing models. However, these studies were for different geographical locations with coarser spatial resolutions for considered variables. For instance, the spatial resolution of models used by Kumar et al. (2013) and Choi et al. (2016) was 2.5^0, Gaetani and Mohino (2013) used models of more than 1.1^0, and Lovino et al. (2018) used precipitation data of 1.0^0 spatial resolution. At the regional level, Mehrotra et al. (2014) assessed the multi-model decadal hindcast of precipitation for different hydrological regions over Australia using 0.5^0 spatial resolution and reported lower skills for precipitation as opposed to temperature and geopotential height. Climate data of 0.5^0 spatial resolution covers a ground area equivalent to a square of 50 km length. Comparatively, a 50 km × 50 km area is very big where climate variabilities are also large and frequency and magnitude of precipitation vary in a few kilometers. As the precipitation shows higher spatial and temporal variations than temperature and the model performances vary region to region, therefore the GCMs derived precipitation should be assessed at the local level before any application. Numerous studies evaluated CMIP5 models (Bhend and Whetton, 2015; Choudhury et al., 2019; Flato et al., 2013; Mehrotra et al., 2014; Moise et al., 2015) but there is no study evaluating CMIP5 decadal precipitation either 10 or 30-year simulation at a catchment level with a spatial resolution finer than 0.50.

The fundamental problem with the CMIP5 decadal data is the drift (a time varying systematic bias) (Mehrotra et al., 2014), which is a long-term time-varying systematic bias generated by the GCMs while they revert to their equilibrium state from the forced initialized state. Now drift has been an important issue in decadal experiments where little systematic directional bias from model to model and/or region to region is seen (Gupta et al. 2012, Gupta et al. 2013). As 'drift' in the climate model (GCMs) outputs hinders the credible applications of models output, therefore, drift correction is an essential prerequisite step before the application of climate model forecasts (or hindcasts). ICPO (2011) recommended a mean drift correction method that has been employed and critically reviewed in previous studies (Choudhury et al., 2016; Mehrotra et al., 2014) but not for the precipitation. Taylor et al. (2012) recommended drift

correction by applying relatively sophisticated bias correction methods. To date, the drift of temperature and temperature-based climate indices have been paid much attention in many previous studies (Chikamoto et al., 2013; Choudhury et al., 2016; Hawkins et al., 2014; Kharin et al., 2012; Narapusetty et al., 2014) but drift in precipitation has been given a little attention (Gupta et al., 2013). It may be due to the higher prediction skill of models for reproducing the temperature (Masanganise et al., 2013; Meehl et al., 2014; Mehrotra et al., 2014) than the precipitation. However, no study was conducted on quantifying the drift of CMIP5 decadal precipitation and finding the drift correction alternatives so far at the catchment level.

GCMs are not perfect enough and their outputs contain systematic biases (Randall et al., 2007) that need rigorous correction before any application (Islam et al., 2011, 2014; Maurer and Hidalgo, 2008; Mehrotra and Sharma, 2010). To alleviate the biases, many researchers suggested using the multi-model ensemble mean (MMEM) while using GCMs data including the CMIP5 decadal experiment (Choudhury et al., 2016; Islam et al., 2014; Knutti et al., 2010; McSweeney et al., 2015). The use of MMEM may enhance the prediction performances (Kumar et al., 2013; Sheffield et al., 2013) by reducing the biases to some extent. However, there is no study on the ranking of the CMIP5 models, contributed to the decadal hindcast or prediction to reproducing the monthly precipitation, based on this how many and which of them should be considered to produce MMEM so that it could provide better outcome. Therefore the current research will be focusing on the assessment of the CMIP5 decadal precipitation at a catchment level and the outcome of the assessments will be employed to predict the future precipitation through multi-model approaches. At first, this assessment will sort out the best spatial interpolation method for precipitation and a better simulation (either 10 or 30-year) of the CMIP5 decadal experiment. Then it will quantify the model drifts and will investigate the drift correction alternatives. In the next step, this study will rank the models, which contributed to reproducing the monthly precipitation of CMIP5 decadal hindcast or prediction and optimize their number to form the best MMEM. Finally, the best MMEM along with the historically observed data will be employed in the machine and deep learning algorithms to predict the future precipitation at the catchment level.

1.3 Research objectives

The main objective of this research is to assess the monthly hindcast precipitation of CMIP5 decadal experiment at a catchment level for a spatial resolution of 0.05^0 and its application to the future prediction of precipitation using a combination of knowledge and data-driven

approach through the machine and deep learning algorithms. The following specific objectives will be carried out to achieve the overall goal and all analyses will be done at a catchment level for a spatial resolution of 0.05^0:

i. To select the suitable interpolation method by comparing the performance of different spatial interpolation methods for the monthly hindcast precipitation of CMIP5 decadal experiment at a catchment level.
ii. To select the suitable data from the available GCM simulation by comparing the performance of 10 and 30-year simulations for the monthly hindcast precipitation of CMIP5 decadal experiment at a catchment level.
iii. To investigate the drift in monthly and seasonal mean (aggregated from monthly values) hindcast precipitation of CMIP5 decadal experiment at a catchment level and assess the suitability of a mean-drift correction method.
iv. To investigate the drift correction alternatives for the seasonal precipitation of the CMIP5 decadal experiment at a catchment level.
v. To categorize the models contributed to the CMIP5 decadal experiment based on their performances for monthly hindcast precipitation at catchment level and identify their best combination that would provide a better performance as MMEM.
vi. To predict the monthly precipitation by Facebook Prophet Model using a combination of knowledge and a data-driven approach at a catchment level.
vii. To predict the monthly precipitation for a decadal timescale using CMIP5 decadal experiment data through a deep neural network at a catchment level.

1.4 Significance and novelty of this research

Decadal prediction of CMIP5 was the first attempt to examine the climate predictability and explore the prediction capabilities of the forecasting systems on decadal time scales. For its potential applications in many dimensions, it attracted huge attention from the climate research community (Crawford et al., 2006). So far, temperature and temperature-based climate indices from CMIP5 decadal prediction have been paid much attention whereas very limited attention is paid to the precipitation in general and no attention was paid to a catchment level and of a spatial resolution finer than 0.5^0. This study is the first attempt that assesses monthly precipitation of CMIP5 decadal hindcast data at a catchment level for a spatial resolution of 0.05^0. It comprises three different phases; Phase-I (objective i, and ii), Phase-II (objective; iii, iv, and v), and Phase-III (objective; vi, and vii).

In the previous studies, spatial interpolation methods were randomly selected and a similar method was applied for different climate variables (Amengual et al., 2012; Kamworapan and Surussavadee, 2019; Mehrotra et al., 2014; Miao et al., 2016) and the reason behind selections were not well explained. According to Wagner et al. (2012) and Jones, (1999), researchers should take care before selecting the interpolation methods, especially for the precipitation. The first phase of this study will help the potential users as well as researchers to select the suitable interpolation methods for the GCMs derived precipitation and simulation type of the decadal predictions.

The second phase will-

(i) quantify the drifts and assess the suitability of the mean drift correction method,
(ii) compare drift correction alternatives and propose a new drift correction method,
(iii) categorize the models based on their performances over the entire catchment and sort out the best combination of the selected models to form the best MMEM.

It will help the potential researchers to identify the uncertainty in the decadal prediction at the catchment level, finding the better approach reducing uncertainty, categorising the models and their best combination as MMEM that are important for model-based decision and policymaking for future water availability, climate impact assessments, and planning, design, and development of water supply infrastructures.

In the third phase, this study will develop a novel approach to predict monthly precipitation for decadal timescale using machine learning algorithms and a deep neural network where both the knowledge and data-driven approach will be employed. In the previous research, ANN was used following a data-driven approach where ANN was trained based on the observed data only (Hung et al., 2009; Lee et al., 2018; Meinke et al., 2007; Mekanik et al., 2011; Mislan et al., 2015). For the first time, this study will use MMEM of CMIP5 decadal hindcast precipitation and corresponding observed data for a supervised training approach of a deep neural network (Bidirectional LSTM). This will give a new insight into the future prediction of climate variables and may resolve some limitations of the existing bias correction methods as the technological developments in combination with the research innovations enhanced the computation facility in this modern arena of artificial intelligence.

The precipitation prediction in ten years ahead in this study will be beneficial for mitigating the floods, managing water resources, agriculture, agro-businesses, decision and policy-

making (Hansen et al., 2011; J. W. Jones et al., 2000) for infrastructure development, and for some other sectors such as retail industry, finance, insurance, fishery, transport, tourism, and others. It will be more beneficial for the Australian economy, as well as communities and government by informing decision-making in areas such as risk and natural asset management as Australia possesses the most variable climate and its water resources are highly vulnerable to climate change.

1.5 Organization of the thesis

This thesis has been organized with both published articles and unpublished works. Five articles have been published, four of them in journals and one in an international conference. One article is under review and one more will be submitted soon. All chapters are formatted as articles (either published and/or draft) and the sequence of the chapters will reflect the sequence of the specific objectives of this study. The status, either published or unpublished, of the articles are mentioned in the footnotes of the first page of each chapter. The organization of this entire thesis has been presented in Fig. 1-1. Note that specific objective-based literature reviews are provided in the introduction part of individual chapters. For this reason, no separated literature review chapter has been added in this thesis.

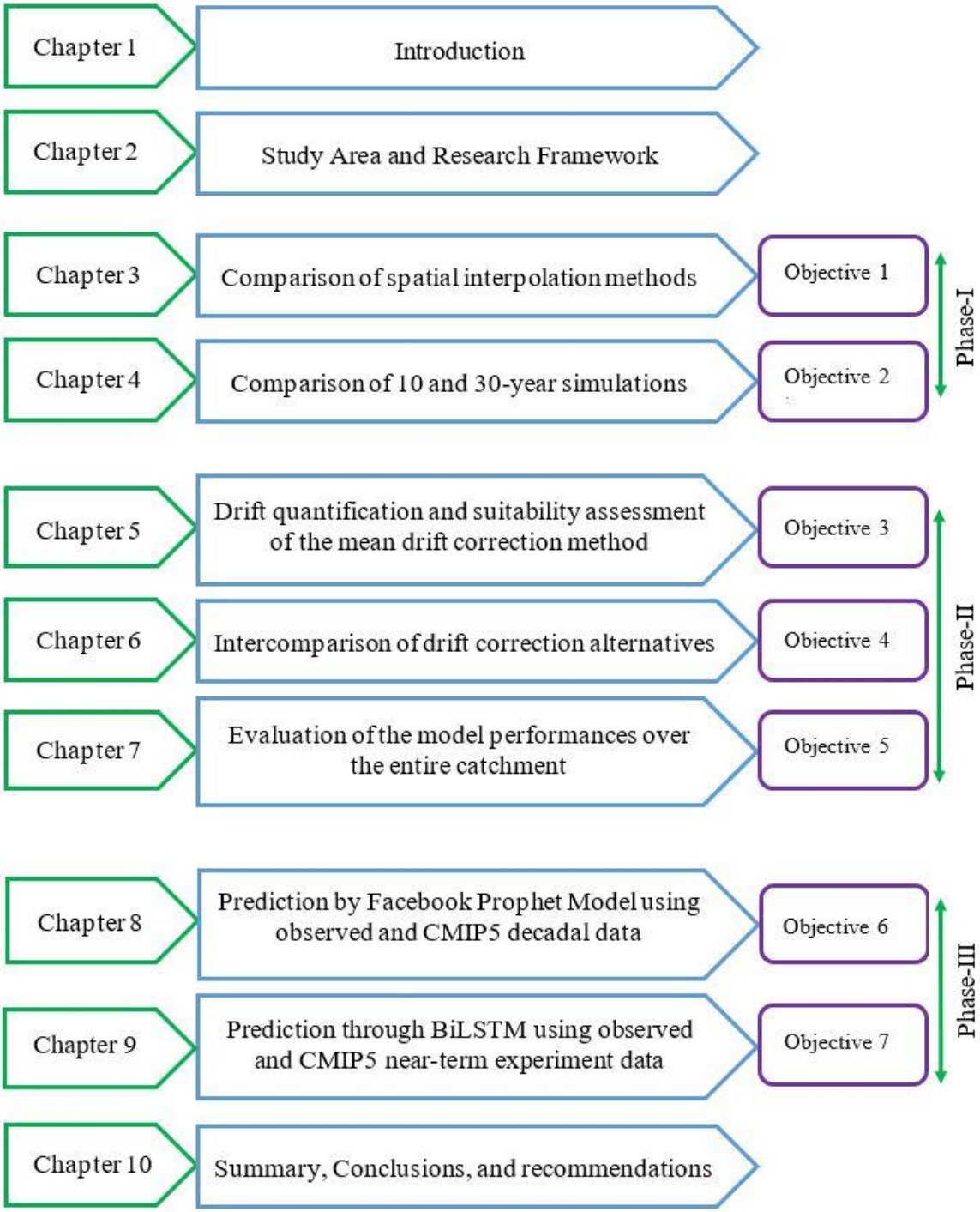


Fig. 1-1 Organization of the thesis

The first chapter provides the research background, identifies the problems, and outlines the research objectives. This chapter also provides the significance and novelty of this research and the organization of the thesis.

Chapter two describes the study area, data source and their processing steps used in this study. It also describes the research steps followed in the different phases of this study.

Chapter three compares the performance of different spatial interpolation methods, at a single grid and also over the entire catchment, and suggests the best method for re-gridding the GCMs derived precipitation. It also describes the number of models considered and the evaluation strategy, from both temporal and spatial perspectives, for sorting the best spatial interpolation method. The reason for being the best method for precipitation is explained in this chapter.

Chapter four compares the performance between 10 and 30-year simulations of the CMIP5 decadal hindcast precipitation and suggests a better simulation. It describes the importance, evaluation methodology, number of models used, and the reason for being a better simulation.

Chapter five quantifies the model drifts and assesses the suitability of the widely used mean drift correction method for the CMIP5 decadal hindcast precipitation at a catchment level. This chapter will also present the importance of drift quantification and its corrections at a catchment level. Suggestions for drift correction alternatives, a transformation of monthly data to seasonal mean, considering MMEM, and sorting the number of models to form MMEM are made in this chapter.

Chapter six, is a follow-up chapter of chapter five, investigates the drift correction alternatives for the seasonal mean precipitation and proposes a new drift correction method. It also explains the pros and cons of different alternatives and suggests no best method for the decadal dataset. It also suggests a further investigation to find a better drift correction method for the monthly dataset.

Chapter seven evaluates the performance of all selected models over the entire catchment and ranks them based on their performance metrics from both temporal and spatial perspectives. It also categorizes the models and optimizes the number of models to form a better MMEM that was suggested in chapter five.

Chapter eight predicts the monthly precipitation for a decadal timescale using Facebook Prophet Model and demonstrates that a combination of knowledge and a data-driven approach gives comparatively better prediction skills than only the data-driven approach. It also explains the importance of considering the knowledge and data-driven approach together instead of using only the data-driven approach and suggests the application of a deep neural network using a similar prediction approach.

Chapter nine also predicts the monthly precipitation for a decadal timescale using MMEM (categorised in chapter seven) and the corresponding observed values following a supervised training approach in a deep neural network (Bidirectional LSTM) to demonstrate an application of knowledge and a data-driven approach in the future prediction. Upon comparing the skills of predicted values and the MMEM for the monthly data, this study suggests that the supervised training approach can be considered as an alternative to the existing bias/drift correction methods that were suggested in chapter six.

Chapter ten presents the overall conclusions of this research. The conclusions are mainly derived from the conclusions drawn in different chapters (papers). It also outlines the limitations of this study and presents the recommendation for future study.

References

Ali, M., Deo, R.C., Downs, N.J., Maraseni, T., 2019. Monthly rainfall forecasting with markov chain monte carlo simulations integrated with statistical bivariate copulas, Handbook of Probabilistic Models. Elsevier Inc. https://doi.org/10.1016/B978-0-12-816514-0.00003-5

Amengual, A., Homar, V., Romero, R., Alonso, S., Ramis, C., 2012. A statistical adjustment of regional climate model outputs to local scales: Application to Platja de Palma, Spain. Journal of Climate 25, 939–957. https://doi.org/10.1175/JCLI-D-10-05024.1

Apurv, T., Mehrotra, R., Sharma, A., Goyal, M.K., Dutta, S., 2015. Impact of climate change on floods in the Brahmaputra basin using CMIP5 decadal predictions. Journal of Hydrology 527, 281–291. https://doi.org/10.1016/j.jhydrol.2015.04.056

Barsugli, J., Anderson, C., Smith, J.B., Vogel, J.M., 2009. Options for Improving Climate Modeling to Assist Water Utility Planning for Climate Change. Water Utility Climate Alliance 144.

Bhend, J., Whetton, P., 2015. Evaluation of simulated recent climate change in Australia. Australian Meteorological and Oceanographic Journal 65, 4–18. https://doi.org/10.22499/2.6501.003

BoM, 2020. Annual rainfall . _ State of the Environment (Department of Environment and Science).

Chikamoto, Y., Kimoto, M., Ishii, M., Mochizuki, T., Sakamoto, T.T., Tatebe, H., Komuro, Y., Watanabe, M., Nozawa, T., Shiogama, H., Mori, M., Yasunaka, S., Imada, Y., 2013. An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC. Climate Dynamics 40, 1201–1222. https://doi.org/10.1007/s00382-012-1351-y

Choi, J., Son, S.W., Ham, Y.G., Lee, J.Y., Kim, H.M., 2016. Seasonal-to-interannual prediction skills of near-surface air temperature in the CMIP5 decadal hindcast experiments. Journal of Climate 29, 1511–1527. https://doi.org/10.1175/JCLI-D-15-0182.1

Choudhury, D., Mehrotra, R., Sharma, A., Sen Gupta, A., Sivakumar, B., 2019. Effectiveness of CMIP5 Decadal Experiments for Interannual Rainfall Prediction Over Australia. Water Resources Research 55, 7400–7418. https://doi.org/10.1029/2018WR024462

Choudhury, D., Sharma, A., Sen Gupta, A., Mehrotra, R., Sivakumar, B., 2016. Sampling biases in CMIP5 decadal forecasts. Journal of Geophysical Research: Atmospheres 121, 3435–3445. https://doi.org/10.1002/2016JD024804

Climate-Data, 2020. Brisbane climate: Average weather, temperature, rainfall [Available at https://www.climatestotravel.com/climate/australia/brisbane].

Committee, A.S. of C.E.T., 2000. Artificial Neural Networks in Hydrology. II: Hydrologic Applications By the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Journal Of hydrologic engineering 5, 115–123.

Crawford, W., Martinez, R., Suga, T., 2006. List of possible applications of decadal prediction.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Anav, A., Andrews, T., Baehr, J., Bodas-salcedo, A., Catto, J., Sillmann, J., Simmons, A., 2013. Evaluation of Climate Models, in: Intergovernmental Panel on Climate Change (Ed.), Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge, pp. 741–866. https://doi.org/10.1017/CBO9781107415324.020

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling.

International Journal of Climatology 27, 1547–1578. https://doi.org/10.1002/joc.1556

Gaetani, M., Mohino, E., 2013. Decadal prediction of the sahelian precipitation in CMIP5 simulations. Journal of Climate 26, 7708–7719. https://doi.org/10.1175/JCLI-D-12-00635.1

George, J., Janaki, L., Parameswaran Gomathy, J., 2016. Statistical Downscaling Using Local Polynomial Regression for Rainfall Predictions – A Case Study. Water Resources Management 30, 183–193. https://doi.org/10.1007/s11269-015-1154-0

Grotch, S.L., MacCracken, M.C., 1991. The Use of General Circulation Models to Predict Regional Climatic Change. Journal of Climate 4, 286–303. https://doi.org/10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2

Gupta, A. Sen, Jourdain, N.C., Brown, J.N., Monselesan, D., 2013. Climate Drift in the CMIP5 Models*. Journal of Climate 26, 8597–8615. https://doi.org/10.1175/JCLI-D-12-00521.1

Gupta, A. Sen, Muir, L.C., Brown, J.N., Phipps, S.J., Durack, P.J., Monselesan, D., Wijffels, S.E., 2012. Climate drift in the CMIP3 models. Journal of Climate 25, 4621–4640. https://doi.org/10.1175/JCLI-D-11-00312.1

Hansen, J.W., Mason, S.J., Sun, L., Tall, A., 2011b. Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture 47, 205–240. https://doi.org/10.1017/S0014479710000876

Hawkins, E., Dong, B., Robson, J., Sutton, R., Smith, D., 2014. The interpretation and use of biases in decadal climate predictions. Journal of Climate 27, 2931–2947. https://doi.org/10.1175/JCLI-D-13-00473.1

Hong, W.C., 2008. Rainfall forecasting by technological machine learning models. Applied Mathematics and Computation 200, 41–57. https://doi.org/10.1016/j.amc.2007.10.046

Hossain, I., Rasel, H.M., Imteaz, M.A., Mekanik, F., 2020. Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics 132, 131–141. https://doi.org/10.1007/s00703-019-00679-4

Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K., 2009. An artificial neural network

model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences 13, 1413–1425. https://doi.org/10.5194/hess-13-1413-2009

ICPO, 2011. Data and bias correction for decadal climate predictions. CLIVAR Publication Series No. 150, 6 pp.

IPCC, 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, Core Writing Team, R.K. Pachauri and L.A. Meyer. https://doi.org/10.1017/CBO9781107415324.004

Islam, S.A., Bari, M., Anwar, A.H.M.F., 2011. Assessment of hydrologic impact of climate change on Ord River catchment of Western Australia for water resources planning: A multi-model ensemble approach, in: Chan, F., Marinova, D. and Anderssen, R.S. (Eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand (MSSANZ), Inc. https://doi.org/10.36334/modsim.2011.I6.islam

Islam, S.A., Bari, M.A., F. Anwar, A.H.M., 2014. Hydrologic impact of climate change on Murray-Hotham catchment of Western Australia: A projection of rainfall-runoff for future water resources planning. Hydrology and Earth System Sciences 18, 3591–3614. https://doi.org/10.5194/hess-18-3591-2014

Jones, J.W., Hansen, J.W., Royce, F.S., Messina, C.D., 2000. Potential benefits of climate forecasting to agriculture. Agriculture, Ecosystems & Environment 82, 169–184. https://doi.org/10.1016/S0167-8809(00)00225-5

Jones, P.W., 1999. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Kamworapan, S., Surussavadee, C., 2019. Evaluation of CMIP5 global climate models for simulating climatological temperature and precipitation for southeast Asia. Advances in Meteorology 2019. https://doi.org/10.1155/2019/1067365

Kharin, V. V., Boer, G.J., Merryfield, W.J., Scinocca, J.F., Lee, W.-S., 2012. Statistical adjustment of decadal predictions in a changing climate. Geophysical Research Letters

39. https://doi.org/10.1029/2012GL052647

Kirtman, B., Power, S.B., Adedoyin, J., Boer, G., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A., Kimoto, M., Meehl, G., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G., Vecchi, G., Wang, H., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P., Kirtman, B., 2013. Near-term Climate Change: Projections and Predictability, in: Intergovernmental Panel on Climate Change (Ed.), Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge, pp. 953–1028. https://doi.org/10.1017/CBO9781107415324.023

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G.A., 2010. Challenges in Combining Projections from Multiple Climate Models. Journal of Climate 23, 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Kumar, S., Merwade, V., Kinter, J.L., Niyogi, D., 2013. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. Journal of Climate 26, 4168–4185. https://doi.org/10.1175/JCLI-D-12-00259.1

Lee, J., Kim, C.G., Lee, J.E., Kim, N.W., Kim, H., 2018. Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea. Water (Switzerland) 10. https://doi.org/10.3390/w10101448

Lovino, M.A., Müller, O. V., Berbery, E.H., Müller, G. V., 2018. Evaluation of CMIP5 retrospective simulations of temperature and precipitation in northeastern Argentina. International Journal of Climatology 38, e1158–e1175. https://doi.org/10.1002/joc.5441

Machiwal, D., Jha, M.K., 2012. Hydrologic Time Series Analysis: Theory and Practice. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-1861-6

Masanganise, J., Chipindu, B., Mhizha, T., Mashonjowa, E., Basira, K., 2013. An evaluation of the performances of Global Climate Models (GCMs) for predicting temperature and rainfall in Zimbabwe 3, 1–11.

Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. Hydrology and Earth System Sciences 12, 551–563. https://doi.org/10.5194/hess-12-551-2008

McKellar, C., Cordero, E.C., Bridger, A.F.C., Thrasher, B., 2013. Evaluation of the CMIP5 Decadal Hindcasts in the State of California. Department of Meteorology and Climate Science. San José State University.

McSweeney, C.F., Jones, R.G., Lee, R.W., Rowell, D.P., 2015. Selecting CMIP5 GCMs for downscaling over multiple regions. Climate Dynamics 44, 3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Means, E., Laugier, M., Daw, J., Kaatz, L., Waage, M., Water Utility Climate Alliance, 2010. Decision Support Planning Methods: Incorporating Climate Change Uncertainties into Water Planning. Evaluation 113. https://doi.org/http://www.wucaonline.org/assets/pdf/pubs_whitepaper_012110.pdf

Meehl, G.A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., Van Oldenborgh, G.J., Vecchi, G., Yeager, S., 2014. Decadal climate prediction an update from the trenches. Bulletin of the American Meteorological Society 95, 243–267. https://doi.org/10.1175/BAMS-D-12-00241.1

Mehrotra, R., Sharma, A., 2010. Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. Water Resources Research 46, 1–17. https://doi.org/10.1029/2009WR008423

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. Journal of Hydrology 519, 2932–2951. https://doi.org/10.1016/j.jhydrol.2014.07.053

Mehta, V.M., Knutson, C.L., Rosenberg, N.J., Olsen, J.R., Wall, N.A., Bernadt, T.K., Hayes, M.J., 2013. Decadal Climate Information Needs of Stakeholders for Decision Support in Water and Agriculture Production Sectors: A Case Study in the Missouri River Basin. Weather, Climate, and Society 5, 27–42. https://doi.org/10.1175/WCAS-D-11-00063.1

Meinke, H., Sivakumar, M.V.K., Motha, R.P., Nelson, R., 2007. Preface: Climate predictions for better agricultural risk management. Australian Journal of Agricultural Research 58, 935–938. https://doi.org/10.1071/ARv58n10_PR

Mekanik, F., Lee, T.S., Imteaz, M.A., 2011. Rainfall modeling using Artificial Neural Network for a mountainous region in west Iran. MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty 3518–3524. https://doi.org/10.36334/modsim.2011.i5.mekanik

Miao, C., Su, L., Sun, Q., Duan, Q., 2016. A nonstationary bias-correction technique to remove bias in GCM simulations. Journal of Geophysical Research: Atmospheres 121, 5718–5735. https://doi.org/10.1002/2015JD024159

Mislan, Haviluddin, Hardwinarto, S., Sumaryono, Aipassa, M., 2015. Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia. Procedia Computer Science 59, 142–151. https://doi.org/10.1016/j.procs.2015.07.528

Moise, A., Wilson, L., Grose, M., Whetton, P., Watterson, I., Bhend, J., Bathols, J., Hanson, L., Erwin, T., Bedin, T., Heady, C., Rafter, T., 2015. Evaluation of CMIP3 and CMIP5 Models over the Australian Region to Inform Confidence in Projections. Australian Meteorological and Oceanographic Journal 65, 19–53. https://doi.org/10.22499/2.6501.004

Narapusetty, B., Stan, C., Kumar, A., 2014. Bias correction methods for decadal sea-surface temperature forecasts. Tellus, Series A: Dynamic Meteorology and Oceanography 66. https://doi.org/10.3402/tellusa.v66.23681

Ouyang, Q., Lu, W., Xin, X., Zhang, Y., Cheng, W., Yu, T., 2016. Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. Water Resources Management 30, 2311–2325. https://doi.org/10.1007/s11269-016-1288-8

Ramírez, M.C., Ferreira, N.J., Velho, H.F.C., 2006. Linear and Nonlinear Statistical Downscaling for Rainfall Forecasting over Southeastern Brazil. Weather and Forecasting 21, 969–989. https://doi.org/10.1175/WAF981.1

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E., 2007. Climate Models and Their Evaluation.

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the

simulation of streamflow in a rainshadow river basin. International Journal of Climatology 23, 887–901. https://doi.org/10.1002/joc.922

Sheffield, J., Camargo, S.J., Fu, R., Hu, Q., Jiang, X., Johnson, N., Karnauskas, K.B., Kim, S.T., Kinter, J., Kumar, S., Langenbrunner, B., Maloney, E., Mariotti, A., Meyerson, J.E., Neelin, J.D., Nigam, S., Pan, Z., Ruiz-Barradas, A., Seager, R., Serra, Y.L., Sun, D., Wang, C., Xie, S., Yu, J., Zhang, T., Zhao, M., 2013. North American Climate in CMIP5 Experiments. Part II: Evaluation of Historical Simulations of Intraseasonal to Decadal Variability. Journal of Climate 26, 9247–9290. https://doi.org/10.1175/JCLI-D-12-00593.1

Smith, D.M., Scaife, A.A., Kirtman, B.P., 2012. What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? Environmental Research Letters. https://doi.org/10.1088/1748-9326/7/1/015602

Syme, B., Barton, C., Rodgers, B., Ayre, R., Toombes, L., Diermanse, F., 2016. Brisbane River Catchment Flood Study (BRCFS) Technical Summary Report Hydrologic and Hydraulic Assessments.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Unnikrishnan, P., Jothiprakash, V., 2020. Hybrid SSA-ARIMA-ANN Model for Forecasting Daily Rainfall. Water Resources Management 34, 3609–3623. https://doi.org/10.1007/s11269-020-02638-w

Wagner, P.D., Fiener, P., Wilken, F., Kumar, S., Schneider, K., 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology 464–465, 388–400. https://doi.org/10.1016/j.jhydrol.2012.07.026

Zhang, P.G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

CHAPTER 2

STUDY AREA AND RESEARCH FRAMEWORK

2.1 Study Area

In this study, the Brisbane River catchment (Fig. 2-1) in Queensland has been selected as the study area which is located in the eastern states of Australia in between the latitudes 26.5^0S~28.15^0S and the longitudes 151.70E ~ 153.150E. It is bounded by the Great Dividing Range to the west and several smaller coastal ranges including the Brisbane, Jimna, D'Aguilar, and Conondale Ranges to the north and east. Most of the Brisbane River catchment lies to the west of the coastal ranges. The catchment topography is a mixture of natural forest, rural land, and urban development where the river system consists of the Brisbane River with small to large tributaries (Syme et al., 2016).

Brisbane River catchment has an area of 13549 square kilometers and a sub-tropical climate where most of the precipitation occurs during summer (December-January-February) and minimum precipitation in winter (June-July-August) (Climate-Data, 2020). Monthly observed precipitation (1911-2015) over the Brisbane River catchment varied from nil to 1360 mm with an annual average precipitation of 628 mm (BoM, 2020) and the number of upper and lower extremes are not quite small. Brisbane catchment was selected because of its tropical climate nature with low to moderate yearly precipitation variability. The variability of precipitation across the catchment is very important as it causes relative changes in the timing of floods from different tributaries. There are two flood mitigation dams, notably Wivenhoe and Somerset, both of which were built to supplement Brisbane's water supply. Flooding in the lower reaches is affected by tidal ranges. The annual and seasonal average precipitation is variable, affected by local factors that include topography, vegetation, and broader scale weather patterns, such as the El Niño– Southern Oscillation. The average annual temperature is 19°C. The average temperature range during summer, autumn, winter and spring are 21-29.8°C, 15-25°C, 11-21°C, and 15-25°C respectively.

There are 30 rain gauge stations across the catchments, operated by the Bureau of Meteorology (BoM), Australia. The observed rain gauge stations, including the stations across the Brisbane River catchment, were used to produce the gridded data for the entire Australia through the Water Resources Assessment Landscape model (AWRA-L V5) as described by Frost et al.

(2016). This study used an observation station located at latitude 27.48^0 S and longitude 153.04^0 E (marked in a circle in Fig. 2-1) as a reference station as it lies closest to a grid (latitude 27.5^0 S and longitude 153.05^0 E) of the observed gridded (5.0 km x 5.0 km) dataset. The BoM operated all rain gauge stations, Rivers, tributaries, water storages, Bioregions along with the catchment boundary are presented in Fig. 2-1.

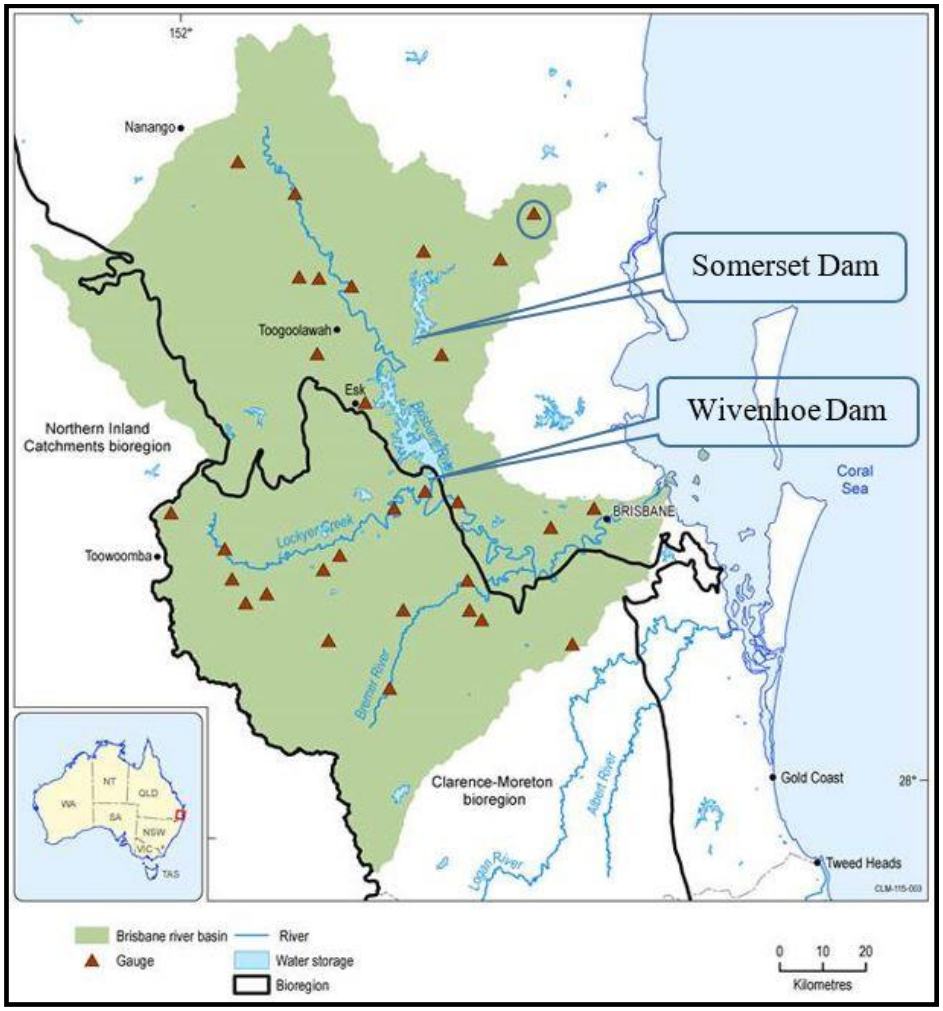


Fig. 2-1 Study Area, Brisbane River catchment (Source: Rassam et al., 2014)

2.2 Research framework

2.2.1 Data collection

CMIP5 involves 20 climate modelling groups with 40 GCMs around the globe. It includes historical simulations from 1850 to 2005, near-term projections until 2035, and long-term projection until 2100 (and beyond) considering four representative concentration pathways

(RCPs). However, for the monthly precipitation, only 10 models; MIROC4h, MRI-CGCM3, MPI-ESM-LR, MIROC5, CMCC-CM, HadCM3, EC-EARTH, MPI-ESM-MR, CanCM4, and IPSL-CM5A-LR contributed to the decadal experiment (near-term) for the years 1961-2005 and five of them (first five) contributed to 30-year simulation for future projection until 2035. For the different calendar systems and relatively coarser spatial resolution, HadCM3 (spatial resolution 3.75° x 2.5°) and IPSL-CM5A-LR (spatial resolution 3.75° x 1.89°) models were not considered in this study. Monthly decadal hindcasts precipitation from the rest eight GCMs are downloaded from CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). For the 10-year simulation, initialized every five years from 1960 to 2005, and for 30-year simulation, initialization years 1960, 1980, and 2005 are used in this study. The details of the selected models are given in Table 2-1.

Table 2-1: Name of the models, modeling group, and initialization years used in this study

Modelling Centre (or Group)	Model (Resolutions °lon × °lat))	Initialization Year (1960-2005)									
		60	65	70	75	80	85	90	95	00	05
		Number of ensembles									
EC-EARTH Consortium	EC-EARTH (1.125 X 1.1215)	14	14	14	14	14	14	14	14	10	18
Meteorological Research Institute	MRI-CGCM3* (1.125 X 1.1215)	06	08	09	09	06	09	09	09	09	06
Max Planck Institute for Meteorology	MPI-ESM-LR* (1.875 X 1.865)	10	10	10	10	10	10	10	10	10	10
	MPI-ESM-MR* (1.875 X 1.865)	03	03	03	03	03	03	03	03	03	03
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC4h (0.5625 X 0.5616)	03	03	03	05	05	05	05	05	05	05
	MIROC5* (1.4062 X 1.4007)	06	06	06	06	04	06	06	06	06	06
Canadian Centre for Climate Modelling and Analysis	CanCM4* (2.8125 X 2.7905)	20	20	20	20	20	20	20	20	20	20
Centro Euro-Mediterraneo per I Cambiamenti Climatici	CMCC-CM 0.75 X 0.748	03	03	03	03	03	03	03	03	03	03

(* indicates model has historical run until the initialization year 2010)

The observed gridded monthly precipitation of $0.05^0 \times 0.05^0$ (≈ 5km × 5km) was collected from the Australian Bureau of Meteorology (BoM).

2.2.2 Data processing

GCMs, provide climate data for the entire globe in netCDF format that requires suitable pre-processing depending on the users' interests. The common data processing steps along with tools used in this study are mentioned in Fig. 2-2. CMIP5 decadal gridded precipitation unit was in flux and the data processing steps it was converted to mm to match with the unit of the observed dataset. Python programming and climate data operator tool (CDO) are used in the data processing of this study. There are a few differences in data processing steps for individual chapters based on the specific objectives.

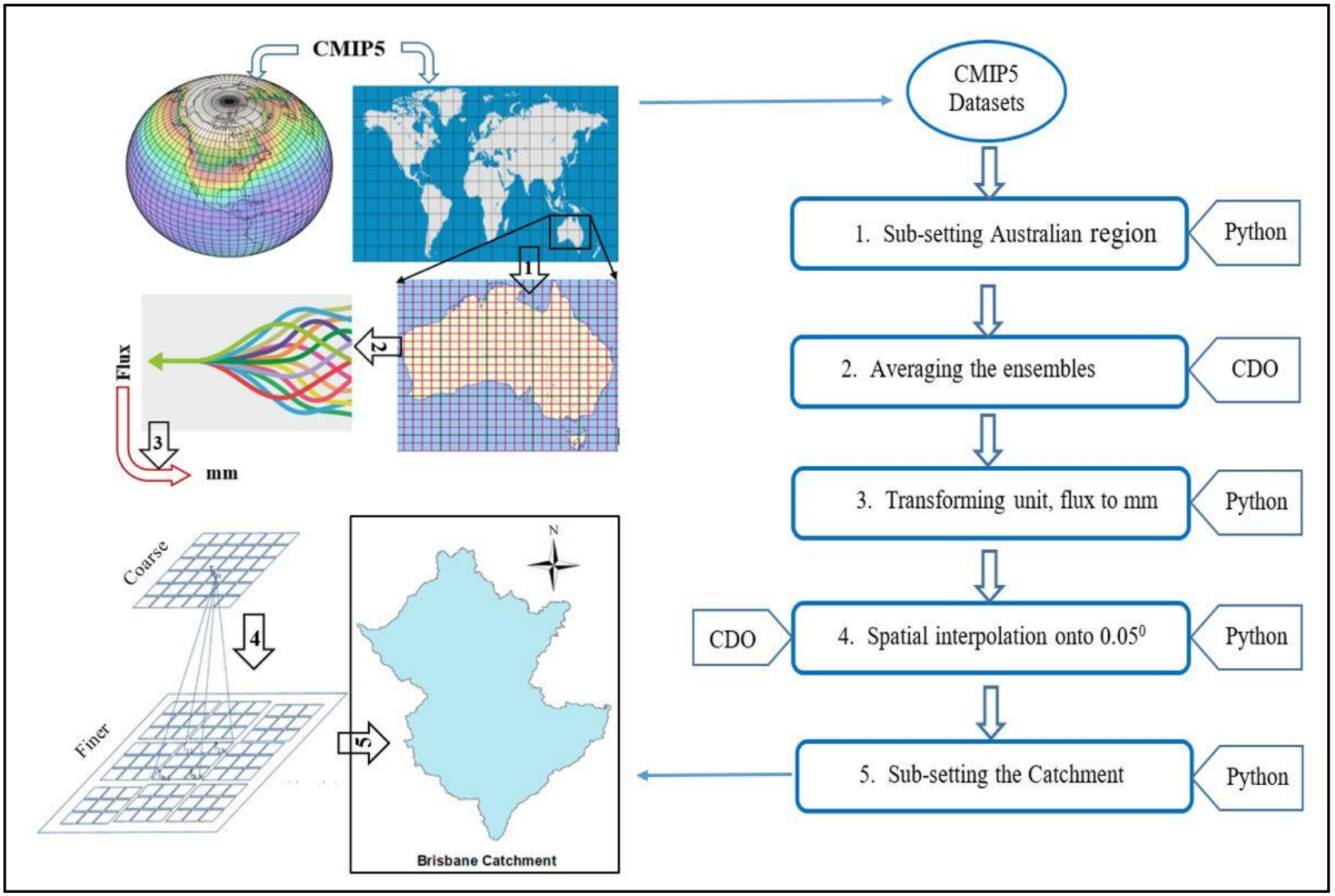


Fig. 2-2 Data processing flow diagram (Image source for steps 1-4: Internet)

2.2.3 Skill tests

Several quantitative performance metrics (skill tests) are used in all steps of assessing, measuring, and comparing the skills. The descriptions of these skill tests are given below.

Correlation Coefficient (CC):

The correlation coefficient is a good measure of linear association or phase error between two datasets. Statistically, it means, how well the values of two datasets correspond to each other. Its values range from -1 (no correlation) to perfect correlation, 1 (Rodwell et al., 2010).

$$CC = \frac{\sum(P-\overline{P})(O-\overline{O})}{\sqrt{\sum(P-\overline{P})^2}\sqrt{\sum(O-\overline{O})^2}} \qquad (2.1)$$

Here, P and O represent the modeled and observed values respectively and $\bar{P}, \bar{O}$ are the mean values, computed for an individual year.

Pearson correlation coefficient (PCC):

PCC also measures the linear correlation between two datasets. The basic difference between CC and PCC is in the calculation of mean values. In PCC, mean values ($\bar{P}, \bar{O}$) are calculated from the entire dataset whilst in CC, mean values are calculated for individual years of the considered datasets.

$$PCC = \frac{\sum(P-\overline{P})(O-\overline{O})}{\sqrt{\sum(P-\overline{P})^2}\sqrt{\sum(O-\overline{O})^2}} \qquad (2.2)$$

Anomaly Correlation Coefficient (ACC):

Wilks, (2011) suggested ACC, which shows how well the modeled anomalies correspond to the observed anomalies. It measures the correspondence between anomalies, calculated by subtracting the observed mean (mean over the entire time-span) from both the modeled and observed values. Its value ranges from -1 (no match) to 1 for perfect anomaly matching.

$$ACC = \frac{\sum\{(P-C)-(\overline{P-C})\}\times\{(O-C)-(\overline{O-C})\}}{\sqrt{\sum(P-C)^2}\sqrt{\sum(O-C)^2}} \qquad (2.3)$$

C represents the mean of the observed (BoM) data for the entire period.

Index of agreement (IA):

Wilmot, (1982) suggested an index of agreement, that measures the accuracy of model data. Index of agreement can be calculated as follows:

$$IA = 1 - \frac{\sum(P-O)^2}{\sum(|P-\bar{O}|+|O-\bar{O}|)^2} \quad (2.4)$$

Fractional Skill Score (FSS):

Fractional skill score directly compares the model and observed fractional coverage of the grid-box events (e.g., precipitation exceeding a certain threshold) for the entire catchment. FSS is a measure of how the spatial variability of the model values matches with the spatial variability of the observed values. FSS can be defined as:

$$FSS = 1 - \frac{\frac{1}{n}\sum_n\left(P_{f,m}-P_{f,o}\right)^2}{\frac{1}{n}\left[\sum_n P_{f,m}^2+\sum_n P_{f,o}^2\right]} \quad (2.5)$$

Where P_f indicates the calculated fraction, *n* indicates the number of events, and the subscript *m* and *o* indicate modeled and observed respectively. Fractions are calculated following Roberts and Lean (2008). In this study, the entire catchment is considered a whole unit, and the temporal averages are taken instead of the spatial averages. Two threshold values are considered; ≥85 percentile for the wet events (DJF) and <15 percentile for the dry events (JJA). To calculate the fractions for individual events, the number of grid points covered for a specified threshold value (for instance, 85 percentile of specific wet events) are counted and then divided by the total number of grids available in the Brisbane River catchment.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

MAE and RMSE are used to measure the average magnitude of the errors between model (drift corrected) and observed values. MAE is the average of the absolute values of the differences between forecasted and corresponding observed values and it is weighted equally in the average. The RMSE is a quadratic scoring rule which is squared before it is averaged and provides a relatively high weight to large errors. RMSE is useful when large errors are especially undesirable. The value of both RMSE and MAE ranges from 0 to ∞ where lower values indicate higher accuracy and vice versa.

$$MAE = \frac{1}{n}\sum(|P-O|) \quad (2.6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum(|P - O|)^2} \qquad (2.7)$$

Two correlations, two errors, two spatial skills (for extreme dry and wet events), one accuracy, and one anomaly correlation metric are used in this study. Every individual skill metric is different and presents skills from different perspectives. For instance, the mean of every individual year is used to calculate CC whilst the mean over a decade is used for PCC calculation. For this reason, CC will present a more robust correlation compared to PCC. Both of them are used in this study as PCC is well known and widely accepted. The anomaly correlation coefficient would present, how the models are able to reproduce the anomalies which are calculated by subtracting the observed mean from both the models and observed values. Similarly, both error metrics; MAE and RMSE are used considering the performance ranges (low to high) of the models. For instance, all models will not show similar performances and few of them may show big differences from the observed values. In that case, RMSE will provide comparatively higher error values due to its quadratic scoring rule which will not the case for MAE. For this, both error metrics are used to make the comparison easy among the models and draw conclusions. Similarly, the accuracy metric is used to show the accuracy of the models' predicted values compared to the observed values. One can assume the accuracy in percentage as its values range from 0 to 1.0.

FSSb15 and FSSa85 are used to compare the model performances in reproducing the extreme dry and wet events respectively over the entire catchment. These two skills will play as the key indicators for sorting the models from the hydrological viewpoint and extreme dry and wet events as well. Note that, the number of skill tests will vary in different chapters depending on the research objective.

2.2.4 Research steps

This study comprises three different phases; Phase-I, Phase-II, and Phase-III. Each phase has multiple objectives. Brief descriptions of all three phases are given below.

Phase-I

This phase investigated the best spatial interpolation methods and a better simulation of the CMIP5 decadal experiments for monthly precipitation. For finding the best spatial interpolation method, three datasets of the initialization year 1990 from three models (EC-EARTH, MIRCO4h, and MPI-ESM-LR) of different spatial resolutions are selected. The datasets are interpolated onto 0.05^0 spatial resolution from their native grids using eight interpolation methods: Linear (LIN), Bi-linear (BiLIN), Nearest Neighbor (NN), Distance Weighted Average (DWA), Inverse Distance weighted Average (IDW), First-order conservative (FOC), Second-order conservative (SOC), and Bi-Cubic (BIC). Several skill tests are then used to measure the skills of the interpolation methods at the selected location as well as over the entire Brisbane River catchment. Based on the skill tests, this study finds the SOC method outperformed all the selected methods as it conserves the precipitation flux while re-gridding the dataset.

Then the SOC method was applied to 10 and 30-year simulations for all the available initialization years of the five models (MIROC4h, MRI-CGCM3, MPI-ESM-LR, MIROC5 and CMCC-CM). The interpolated datasets are then compared with the corresponding observed values of 0.05^0 spatial resolutions through the multiple skill tests over the entire catchment. Results show that 10-year simulation contains lower bias that corresponds to the higher skills as opposed to 30-year simulation. As a result, 30-year simulation datasets are discarded from this stage and 10-year simulation data of 0.05^0 spatial resolution using SOC method are used for the next phases of this study.

Phase-II

In this phase, at first, model drifts were quantified for the monthly and seasonal mean precipitation of individual models along with their MMEM and then assessed the suitability of the widely used mean-drift correction method recommended from ICPO (2011) to alleviate the drift. Next, it measured the skills of both the models' raw (interpolated) and drift corrected values and finds insignificant skill improvements of the models after the drift correction. It also finds lower drift in seasonal mean precipitation compared to the monthly values and lowest drift in their respective MMEMs. It suggests further investigation for drift correction alternatives and finding the optimum number of models to form a better MMEM.

In the second step, this phase investigated the drift correction alternatives and proposed a new drift correction method for the seasonal mean precipitation. After going through seven skill tests from different perspectives, this study suggested no best method for the decadal datasets rather suggests using the prudency of the users based on their demands. Finally, it recommends further investigation on drift correction approaches for different time scales e.g., monthly precipitation, and their application to individual ensembles.

In the last step of this phase, this study evaluates the models based on their performance measured from both the temporal and spatial perspective over the entire catchment. It also divided the models into three different categories and found the optimum number of models to form the best MMEM for the selected catchment.

Phase-III

In this phase, monthly precipitation was predicted for a decadal timescale using both the GCMs derived precipitation and the corresponding observed values. This phase has two different sections. In the first section, Facebook Prophet (FBP) model was used to demonstrate incorporating GCMs derived precipitation along with the observed values gives comparatively better prediction accuracy than the prediction based on the observed data only. For this, at first monthly precipitation was predicted for a decade (2006-2015) based on the historically observed data only. In the second step, the best MMEM (sorted from the second phase), and in the third step all individual models were used as additional regressor in FBP in addition to the observed data to predict monthly precipitation for the same period. To justify the prediction skill of the FBP model, six different Machine Learning (ML) regression algorithms were used to predict the monthly precipitation for the same decade where MMEM was used as a feature and corresponding observed values were used as target variables. Upon comparing the prediction skills, it was observed that incorporating GCMs derived values along with the observed values for the local level prediction improves the prediction accuracy where FBP showed better skills than the ML regression models. The first section suggests employing a deep neural network for the future prediction using a combination of GCMs and the corresponding observed data.

In the second section, this study uses a Bidirectional LSTM (a deep neural network) model for the prediction of monthly precipitation in a decadal timescale (2006-2015) using the best MMEM (as features) and corresponding observed values (as targets) in a supervised training

approach. Then it compares the prediction skills over the skills of MMEM to demonstrate the skill improvements of the predicted values by minimizing the bias through the learning process of the deep neural network during its training period.

The complete research steps mentioning the individual objectives of this study and the links among the steps along with objectives are presented in Fig. 2-3.

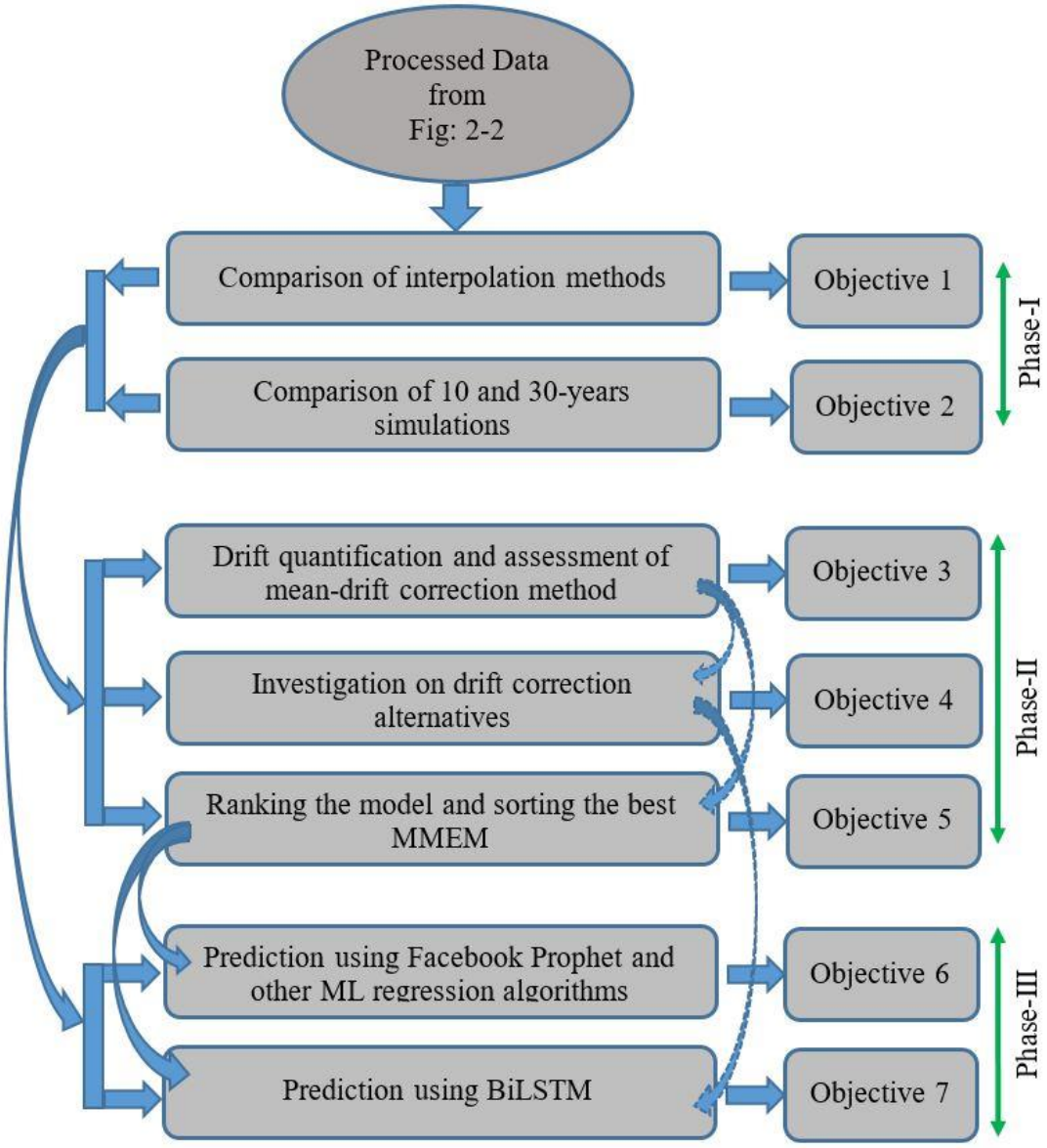


Fig. 2-3 Research steps. The curved arrows on the left (with solid line) indicate the application of the outcome and the curved arrows on the right side (with broken lines) indicate the recommended studies

2.3 Concluding Remarks

This chapter described the study area and marked a reference point that has been mentioned in all subsequent chapters (as selected grid). It also provided a clear overview of data collection and processing steps for the GCMs derived and observed values and provided detailed descriptions of the skill tests, which are employed in all subsequent chapters for measuring, assessing and comparing the skills to address the specific objectives mentioned in the former chapter. The research steps, reflecting the sequential order of the individual specific objectives, have been figured out in this chapter that includes links among the individual steps and their follow-up studies. However, based on the research objectives and their outcomes, research steps have been grouped into three different phases and their descriptions are provided here. Every individual research objective and its outcomes are organized as an individual chapter in this thesis. Depending on the research objective, data processing steps and the number of skill tests may be different but detailed descriptions are mentioned in individual chapters.

References

BoM, 2020. Annual rainfall . _ State of the Environment (Department of Environment and Science).

Climate-Data, 2020. Brisbane climate: Average weather, temperature, rainfall [Available at https://www.climatestotravel.com/climate/australia/brisbane].

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

ICPO, 2011. Data and bias correction for decadal climate predictions. CLIVAR Publication Series No. 150, 6 pp.

Rassam, D., Raiber, M., McJannet, D., Janardhanan, S., Murray, J., Gilfedder, M., Cui, T., Matveev, V., Doody, T., Hodgen, M., Ahmad, M., 2014. Context statement for the Clarence-Moreton bioregion. Product 1.1 from the Clarence-Moreton Bioregional Assessment.

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Monthly Weather Review 136, 78–97. https://doi.org/10.1175/2007MWR2123.1

Rodwell, M.J., Richardson, D.S., Hewson, T.D., Haiden, T., 2010. A new equitable score suitable for verifying precipitation in numerical weather prediction. Quarterly Journal of the Royal Meteorological Society 136, 1344–1363. https://doi.org/10.1002/qj.656

Syme, B., Barton, C., Rodgers, B., Ayre, R., Toombes, L., Diermanse, F., 2016. Brisbane River Catchment Flood Study (BRCFS) Technical Summary Report Hydrologic and Hydraulic Assessments.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

CHAPTER 3

COMPARING SPATIAL INTERPOLATION METHODS FOR CMIP5 MONTHLY PRECIPITATION AT CATCHMENT SCALE

Abstract

The use of Regional Climate Models (RCMs) is prevalent in downscaling the large-scale climate information from the General Circulation Models (GCMs) to the local scale. But it is computationally intensive and requires the application of a numerical weather prediction model. For more straightforward computation, spatial interpolation is commonly used to re-gridding the GCM data to local scales. There are many interpolation methods available, but mostly they are chosen randomly, especially for GCM data. This study compared eight interpolation methods (linear, bi-linear, nearest neighbor, distance weighted average, inverse distance weighted average, first-order conservative, second-order conservative, and bi-cubic interpolation) for re-gridding of CMIP5 decadal experimental data to a catchment scale. For this, CMIP5 decadal precipitation data from three GCMs were collected and subset for Australia and then re-gridded to 0.05^0 spatial resolution matching with the observed gridded data. The re-gridded data were subset for Brisbane catchment in Queensland, Australia, and several skill tests (root mean squared error, mean absolute error, correlation coefficient, and index of agreement) were conducted for a selected observed grid to check the performances of different interpolation methods. Additionally, temporal skills were computed over the entire catchment and compared. Based on the skill tests over the study area, the second-order conservative (SOC) method was found to be an appropriate choice for interpolating the gridded dataset.

Keywords— Comparison, Interpolation, Precipitation, Spatial and Catchment

This chapter has been published as: Hossain, M. M., Garg, N., Anwar, A.H.M.F., Prakash, M., 2021. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale. *Journal Indian Water Resources Society* 41, 2, 28-34. http://iwrs.org.in/journal/apr2021/5apr.pdf. However, few textual changes have been made to address the examiners' comments.

3.1 Introduction

General Circulation Models (GCMs) are widely used to assess climate change and its potential impacts at different temporal and spatial scales, but their coarse spatial resolution (100-250 km) is inadequate for their application at a local scale due to lack of spatial details (Fowler et al., 2007; Grotch and MacCracken, 1991; Salathé, 2003). The Regional Climate Models (RCMs) are often used to down-scale the large-scale climate information from GCMs to a local scale; however, RCMs are complicated, computationally intensive, and time-consuming. To avoid this complexity, in practice, spatial interpolations are applied (Homsi et al., 2020; Jain et al., 2019; McKellar et al., 2013; Yang et al., 2015) to re-grid the coarser-resolution climate model data onto a finer resolution. However, in most cases, spatial interpolation methods are randomly used. For instance, bilinear interpolation has been used in many studies (Amengual et al., 2012; Kamworapan and Surussavadee, 2019; Miao et al., 2016) but the reason behind selecting the bilinear method was not well explained. Climate variables such as precipitation show high spatial variability in frequency and magnitude, where, understanding the spatial distribution of precipitation at different spatial scales is important for water resource management, hydrological modeling, agricultural industries, and urban planning. Therefore, the selection of an appropriate spatial interpolation method is important to provide the accurate spatial distribution of the precipitation when transforming from a relatively coarser to a finer spatial resolution.

Various spatial interpolation techniques ranging from simple to complex have already been used for remapping data to the desired finer resolution (Choudhury et al., 2015; McKellar et al., 2013). For interpolating the rain gauge station data at small and medium scale catchments (or basins), Nearest Neighbour (NN), Inverse Distance Weighting (IDW), Thiessen polygons, Spline and different forms of Kriging are frequently used (Tomczak, 1998). Many studies have compared the performance of these spatial interpolation methods for the rainfall data at different temporal and spatial scales. For instance, da Silva et al. (2019) compared seven interpolation methods for the monthly precipitation over Pernambuco, Brazil, and reported trend surface analysis to be the best followed by natural NN, IDW, and Kriging. Yang et al. (2015) compared four methods with the model generated daily precipitation data and reported that IDW performed slightly better than Spline, Kriging, and ANUDEM (Jones et al., 2009). Dirks et al. (1998) didn't find any advantage of using Kriging over IDW, Thiessen, or areal-mean while gridding rainfall data from 13 rain gauges stations on Norfolk Island.

Consequently, Wu et al. (2019) evaluated several spatial interpolation methods for mapping the bathymetry of the lowermost Mississippi River, which includes IDW, Ordinary Kriging (OK), Universal Kriging (UK), Radial Basis Function (RBF), local Polynomial, and anisotropic form of Elliptical IDW, and OK and found that both the RBF and anisotropic form of OK performed best. Zhang et al. (2016) compared OK, co-Kriging with elevation as covariate (Cok-elevation), and co-Kriging with covariates with precipitation data from tropical rainfall measuring mission (Cok_TRMM) to interpolate precipitation data from 39 rain gauge stations in the Tibetan Plateau. They reported that Cok-TRMM is more effective than the other two that was also confirmed by Foehn et al. (2018). Note, the performance of the interpolation methods depends on several factors, in particular, temporal and spatial resolution of the considered data and the study region. Degré et al. (2015) reviewed several spatial interpolation methods from a different perspective and concluded that, for annual and monthly rainfall, geo-statistical interpolation methods (different mode of Kriging) seem preferable to the deterministic methods (Thiessen, NN, IDW, etc.), but for the daily rainfall, geo-statistical methods and IDW can be a better option. Most of the aforementioned studies interpolated the rain gauge station data and evaluated the interpolation methods at selected grid within the study area by using error metrics; root mean squared error (RMSE), mean absolute error, mean standard errors. To get an idea of which method produces better interpolation at the catchment level, it is essential to apply the methods for the entire study area in addition to a single grid. Wagner et al. (2012) also suggested that evaluation of the interpolation methods should include the spatial distribution over the study area for precipitation data. Therefore, the objective of this study is to evaluate the different interpolation methods for the application of GCM data at a catchment level. This study will consider a single grid measure as well as an entire catchment for the spatial distribution of precipitation data. In addition, this study also assesses the performance of interpolation methods due to the change in spatial resolution of the selected data sets.

3.2 Materials and methods

3.2.1 Data collection

CMIP5 experiments (e.g., decadal) provide global climate data for a wide range of climate variables generated from several climate models. The decadal simulation, once initialized, generates climate data for ten years and longer in some cases (Taylor et al., 2012). Monthly precipitation data from three GCMs; MIROC4h, EC-EARTH, and MPI-ESM-LR was

downloaded from the CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). To observe the performance of interpolation methods over different spatial resolutions, three models of different spatial resolutions are selected here. Details of the models and the data are given in Table 3-1.

The gridded monthly precipitation data with a spatial resolution of 0.05^0 was collected from the Australian Bureau of Meteorology (BoM). The gridded observed data of BoM were produced by the Australian Water Availability Project (AWAP) and the details can be found here (Frost, A. J., Ramchurn, A., and Smith, 2016).

Table 3-1 Model used in this study

Modelling Centre (or Group)	Model Name	(Atmospheric Resolutions in degree)	Time span
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC4h	(0.5625 X 0.5616)	10 Year; From January 1991 to December 2000
Meteorological Research Institute	EC-EARTH	(1.125 X 1.1215)	
Max Planck Institute for Meteorology	MPI-ESM-LR	(1.875 X 1.865)	

3.2.2 Data processing

Firstly, the model datasets were subset for the Australian region, thereafter, all the available ensembles (i.e., multiple runs of the same model with slightly perturbed initial conditions) of the individual models are averaged to produce a single dataset for each model. These datasets were then interpolated from their native grids onto 0.05 x 0.05 degree matching with the grid of the observed dataset. Finally, the interpolated data were subset for the selected Brisbane catchment (i.e., longitude from 151.70 E to 153.150 E and latitude from 26.50 S to 28.150 S) in Queensland, Australia.

3.2.3 Interpolation methods

In this study, eight different interpolation methods were evaluated. The six methods; Bi-linear (BiLIN), Nearest Neighbor (NN), Distance Weighted Average (DWA), First-order conservative (FOC), Second Order conservative (SOC), and Bi-Cubic (BIC) interpolation were performed by the Climate Data Operator (CDO) (Schulzweida, 2019) tool, whilst Linear (LIN)

and Inverse-Distance Weighted Average (IDW) were performed by the Matplotlib and Scipy libraries in Python. It is worth noting that DWA is also an IDW method, where four nearest neighbor grids (by default) are used, whilst in the Scipy based IDW method, only three nearest neighbor grids are considered.

Linear interpolation is the concatenation of linear interpolants between each pair of data points. But the "LinearTriInterpolator" from Matplotlib performs linear interpolation on a triangular grid. Each triangle is represented by a plane so that interpolated values lie on that plane of the triangle containing the interpolants. For the Inverse-Distance Weighted Average, Scipy spatial algorithm described by Maneewongvatana and Mount (2001) is used to locate the neighboring points for a given set of points.

CDO uses adapted interpolation methods from the SCRIP library. SCRIP is a software package. It computes the addresses and weights for remapping and interpolating variables between grids on the spherical coordinates. Initially, it was written for remapping the fields to desired grids in a coupled climate model but can also be used for other applications.

3.2.4 Performance Assessment

The observed dataset has 496 grids (5.0 km X 5.0 km) within the Brisbane catchment, and the skill tests are performed at the grid (latitude 27.50 S and longitude 153.050 E) located closest to an AWS (Automated weather stations) rain gauge (the observed grid at latitude 27.480 S and longitude 153.040 E) operated by the Bureau of Meteorology, Australia. To assess the performance, five skill tests: root mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (CC), anomaly correlation coefficient (ACC) according to Wilks, (2011), and index of agreement (IA) suggested by Wilmot, (1982) were used.

$$(i) \quad RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2} \tag{3.1}$$

$$(ii) \quad MAE = \frac{1}{N}\sum_{i=1}^{N}|F_i - O_i| \tag{3.2}$$

$$(iii) \quad CC = \frac{\sum(F-\overline{F})(o-\overline{O})}{\sqrt{\sum(F-\overline{F})^2}\sqrt{\sum(o-\overline{O})^2}} \tag{3.3}$$

$$(iv) \quad ACC = \frac{\sum\{(F-C)-(\overline{F-C})\}\times\{(O-C)-(\overline{O-C})\}}{\sqrt{\sum(F-C)^2}\sqrt{\sum(O-C)^2}} \tag{3.4}$$

$$(v) \quad IA = 1 - \frac{\sum_{i=1}^{n}(F_i-O_i)^2}{\sum_{i=1}^{n}(|F_i-\bar{O}|+|O_i-\bar{O}|)^2} \tag{3.5}$$

F and O present modeled (interpolated) and observed values respectively whilst $\bar{F}, \bar{O}$ present their annual mean, and C is the decadal mean of the observed (BoM) data.

3.3 Result and analyses

The monthly precipitation data from three CMIP5 models are evaluated against the observed data, and the results are presented in Table 3-2. These models with three different spatial resolutions were chosen (see Table 3-1) to assess the effect on skills of the interpolation methods due to the variations of atmospheric spatial resolution (before interpolation) of the interpolant dataset. The results for the interpolation methods aren't significantly different, but to some extent, some of them are slightly better than others. Overall, the DWA method has comparatively lower errors for all three selected models with varying values for the skill tests; CC, ACC, and IA. However, the performances of the interpolation methods are sensitive to the choice of models; and the spatial resolution of the interpolant dataset. DWA has the lowest errors (RMSE and MAE) and IA, whereas LIN has the highest values for CC and ACC for the coarse spatial resolution model followed by DWA (see Table 3-2), whilst SOC and FOC performed poorly on all skill tests except IA.

Overall, the DWA method has the lowest errors for all three models and outperforms all methods on all temporal skills for the MIROC4h model at a single grid. With the change in spatial resolutions, the skill specifically CC, ACC, and IA varied a little with little to no change in RMSE and MAE.

Table 3-2 Skill comparison of interpolation methods at the selected grid

Interp.	MIROC4h				EC-EARTH				MPI-ESM-LR			
	RMSE	CC	ACC	IA	RMSE	CC	ACC	IA	RMSE	CC	ACC	IA
BiLIN	80.991	0.368	0.354	0.539	79.377	0.437	0.362	0.458	77.432	0.338	0.306	0.414
LIN	80.948	0.368	0.355	0.539	79.377	0.437	0.362	0.458	77.407	0.343	0.307	0.414
NN	82.634	0.345	0.334	0.520	79.207	0.438	0.364	0.462	80.083	0.325	0.289	0.437
IDW	80.998	0.366	0.353	0.536	79.221	0.438	0.364	0.461	77.994	0.334	0.301	0.423
DWA	79.744	0.380	0.365	0.540	79.060	0.436	0.363	0.458	77.284	0.339	0.306	0.405
FOC	82.303	0.350	0.338	0.525	79.204	0.438	0.364	0.462	80.100	0.325	0.289	0.437
SOC	82.307	0.350	0.338	0.525	79.207	0.438	0.364	0.462	80.083	0.325	0.289	0.436
BIC	81.414	0.362	0.349	0.536	79.480	0.438	0.363	0.458	78.085	0.332	0.300	0.424

This study also compared the spatial variations of these temporal skills over the entire catchment for all three models, but only IA (Fig. 3-1) and RMSE (Fig. 3-2) for the MIROC4h models are presented here. From the spatial comparison, it is evident that NN along with conservative methods found little better in CC, ACC (not shown), and RMSE whilst DWA outperforms other methods for IA. An overview of the spatial comparison of all three models based on the specified thresholds of individual skills is presented in Table 3-3. From this comparison, it is evident that NN, DWA, and the conservative methods perform better than others with little variations in skills over model types where SOC found more consistent, followed by FOC in better performance than DWA and NN.

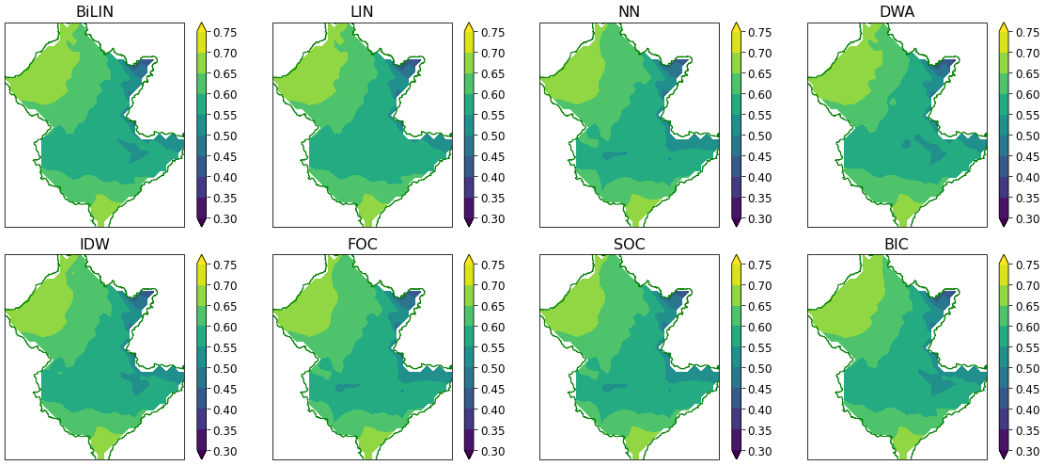


Fig. 3-1 Spatial comparison of Index of Agreement (IA) of different interpolation methods (MIROC4h) over the catchment. Labels on the right of each plot indicate more the brightest area higher the performance of the interpolation methods

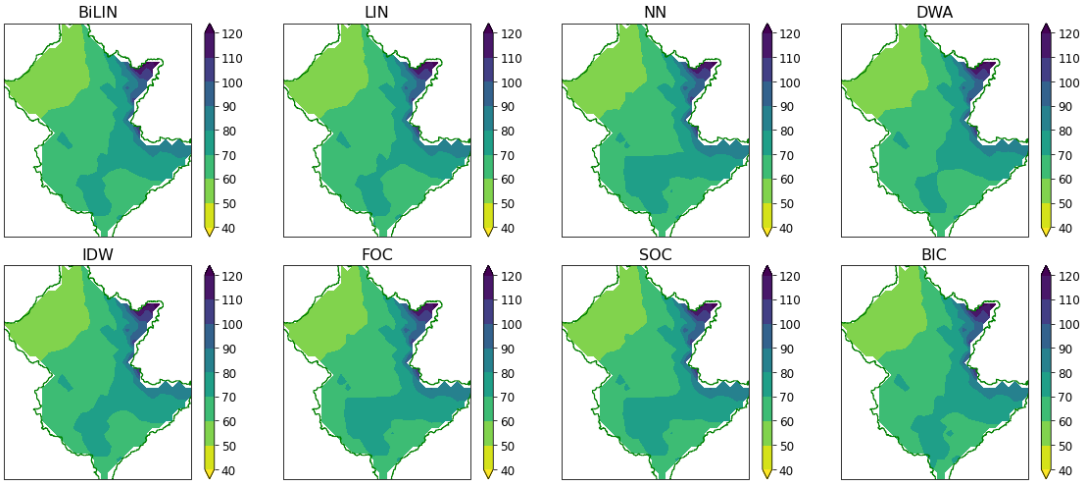


Fig. 3-2 Comparison of the spatial variations of Root Mean Squared Error (RMSE) of different interpolation methods (MIROC4h) over the catchment. Labels on the right of each plot indicate more the brightest area higher the performance of the interpolation methods

Table 3-3 Number of grids covered by the interpolation methods for the specific thresholds of the skills. Selected models, Skills, and corresponding thresholds are presented in the first, second, and third row respectively. Higher the number in the respective columns presents better the performance of the interpolation methods over the catchment and vice versa

Interp	MIROC4h				EC-EARTH				MPI-ESM-LR			
	RMSE	CC	ACC	IA	RMSE	CC	ACC	IA	RMSE	CC	ACC	IA
	<=	>=	>=	>=	<=	>=	>=	>=	<=	>=	>=	>=
	55	0.55	0.5	0.65	55	0.55	0.45	0.6	55	0.45	0.35	0.6
BiLIN	47	5	11	102	22	9	0	5	20	3	8	87
LIN	45	9	16	102	22	11	0	5	20	2	2	84
NN	52	14	20	88	56	0	3	22	19	1	16	85
IDW	50	6	12	92	19	0	0	3	19	3	2	79
DWA	45	2	8	103	20	11	1	6	19	11	27	86
FOC	52	14	21	88	56	0	3	21	19	1	15	85
SOC	52	14	21	88	58	0	3	22	19	1	14	87
BIC	51	11	19	112	23	12	0	5	17	3	5	97

3.4 Discussion and Conclusion

This study compared different spatial interpolation methods at a catchment scale, where the temporal errors and skills were evaluated at an observed grid within the catchment and spatial comparison of temporal skills for the whole catchment. Preliminary results show no significant difference among the interpolation methods when compared at observed stations. This may be due to not capturing the spatial variations at a single grid because of using regularly gridded

interpolant datasets which demonstrate considering the entire study area for the comparison. The difference among the interpolation methods may appear even at a single grid for interpolating the irregularly distributed data like point rain gauge stations. For instance, IDW (k=3) and DWA (k=4) followed a similar interpolation approach but their different performances were evident due to considering the different number of neighbouring points. This may be because the interpolant datasets are regularly gridded as opposed to the irregularly distributed point rain gauge stations. For irregular datasets such as point rain gauge stations, the difference in skills for the interpolation methods may appear, even for IDW (k=3) and DWA (k=4), where the only difference between these two methods is the number of neighboring grids used for the interpolation. Upon comparison of the errors and skills at a single grid within the catchment, DWA was found to be better than other interpolation methods, also reported in other studies (Chen and Liu, 2012; Hsieh et al., 2006; Yang et al., 2015). Note, Chen and Liu, (2012) and Hsieh et al. (2006) used rainfall data from the rain gauge stations, whereas Yang et al. (2015) used regional climate models' generated data, but all reported that IDW performs better.

For the sake of brevity, when comparing temporal skill over the catchment, only RMSE, CC, ACC, and IA are considered. For the spatial comparison, a specific threshold for skill values is set, and the number of grids covered for the thresholds is counted. The spatial comparison reveals that the conservative methods performed much better than the other five interpolation methods, with SOC outperforming FOC. It appears that maintaining the spatial distribution of the precipitation by interpolating conservatively is the main reason behind the better spatial skills of these conservative methods. In conservative methods, the precipitation flux is conserved when interpolated from the source grid onto an interpolated grid. The conservation of flux while interpolating spatially is important, especially for the discontinuous variable like precipitation and due to its high temporal and spatial variability. For instance, if a few grids have no precipitation while others have large values, then bilinear interpolation can make all grids zero, including the large values as it uses four grids nearest the 2-degree target grid. In this case, conservative interpolation would be a good approach. During the spatial interpolation, it is presumed that an accurate approximation of the flux on a source grid leads to a more accurate remapping, as evidenced by the use of SOC. In the SOC method, the area-weighted distance from the source cell centroid is considered as the gradient of flux for the interpolated cell (Jones, 1999). Jones (Jones, 1999) compared first and second-order conservative with the other different interpolation methods and found that conservative

methods perform much better for the dataset on a regular rectangular grid, where second-order conservative shows an order-of-magnitude improvement over the first order.

Wagner et al. (2012) suggested that the spatial skills of the interpolation methods must be considered rather than the skill measured at points/grids. Maintaining the spatial distribution is more important to assess climate variability at a local scale, especially for the precipitation. said the results revealed that the conservative methods would suit better for spatial interpolation of precipitation as they maintain the spatial distribution of the interpolated variables by conserving the flux. Furthermore, SOC may be the best option for spatial interpolating the gridded precipitation dataset like those from the GCMs as found in this study. This finding is in line with the previous study (Jones, 1999), where the second-order conservative (SOC) method was found to be an appropriate choice for interpolating the gridded dataset. For the cross-validation, similar studies at other catchments/regions are recommended.

List of symbols

C	:	Mean of the observed values for the total time span of each dataset
F	:	Model predicted/forecasted precipitation values
$\overline{F}$	:	Yearly Mean of the model predicted/forecasted precipitation values
O	:	Observed precipitation values
$\bar{O}$	:	Yearly Mean of the observed values
$F - C$	:	Model anomaly
$\overline{F - C}$	:	Mean of the model anomalies
$O - C$	:	Observed anomaly
$\overline{O - C}$	:	Mean of the observed anomalies

References

Amengual, A., Homar, V., Romero, R., Alonso, S., Ramis, C., 2012. A statistical adjustment of regional climate model outputs to local scales: Application to Platja de Palma, Spain. Journal of Climate 25, 939–957. https://doi.org/10.1175/JCLI-D-10-05024.1

Chen, F.W., Liu, C.W., 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy and Water Environment 10, 209–222. https://doi.org/10.1007/s10333-012-0319-1

Choudhury, D., Sharma, A., Sivakumar, B., Sen Gupta, A., Mehrotra, R., 2015. On the predictability of SSTA indices from CMIP5 decadal experiments. Environmental Research Letters 10, 074013. https://doi.org/10.1088/1748-9326/10/7/074013

da Silva, A.S.A., Stosic, B., Menezes, R.S.C., Singh, V.P., 2019. Comparison of interpolation methods for spatial distribution of monthly precipitation in the state of pernambuco, Brazil. Journal of Hydrologic Engineering 24, 1–11. https://doi.org/10.1061/(ASCE) HE.1943-5584.0001743

Degré, A., Tech, G.A., Passage, S.S., 2015. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale : a review PoPuPS | Different methods for spatial interpolation of rainfall data for ... Base 17, 1–10.

Dirks, K.N., Hay, J.E., Stow, C.., Harris, D., 1998. High-resolution studies of rainfall on Norfolk Island. Journal of Hydrology 208, 187–193. https://doi.org/10.1016/S0022-1694(98)00155-3

Foehn, A., García Hernández, J., Schaefli, B., De Cesare, G., 2018. Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. Journal of Hydrology 563, 1092–1110. https://doi.org/10.1016/j.jhydrol.2018.05.027

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27, 1547–1578. https://doi.org/10.1002/joc.1556

Frost, A. J., Ramchurn, A., and Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

Grotch, S.L., MacCracken, M.C., 1991. The Use of General Circulation Models to Predict Regional Climatic Change. Journal of Climate 4, 286–303. https://doi.org/10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2

Homsi, R., Shiru, M.S., Shahid, S., Ismail, T., Harun, S. Bin, Al-Ansari, N., Chau, K.W., Yaseen, Z.M., 2020. Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. Engineering Applications of Computational Fluid Mechanics 14, 90–106. https://doi.org/10.1080/19942060.2019.1683076

Huey-hong Hsieh, Shin-jen Cheng, Jing-ye Liou, Shin-chang Chou, B.S., 2006. Characterization of spatially distributed summer daily rainfall. J Chin Agric Eng 52, 47–55.

Jain, S., Salunke, P., Mishra, S.K., Sahany, S., 2019. Performance of CMIP5 models in the simulation of Indian summer monsoon. Theoretical and Applied Climatology 137, 1429–1447. https://doi.org/10.1007/s00704-018-2674-3

Jones, D.A., Wang, W., Fawcett, R., 2009. AWAP_Jones_2009. Australian Meteorological and Oceanographic Journal 58 58, 233–248.

Jones, P.W., 1999. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Kamworapan, S., Surussavadee, C., 2019. Evaluation of CMIP5 global climate models for simulating climatological temperature and precipitation for southeast Asia. Advances in Meteorology 2019. https://doi.org/10.1155/2019/1067365

Maneewongvatana, S., Mount, D.M., 2001. On the efficiency of nearest neighbor searching with data clustered in lower dimensions. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2073, 842–851. https://doi.org/10.1007/3-540-45545-0_96

McKellar, C., Cordero, E.C., Bridger, A.F.C., Thrasher, B., 2013. Evaluation of the CMIP5 Decadal Hindcasts in the State of California. Department of Meteorology and Climate Science. San José State University.

Miao, C., Su, L., Sun, Q., Duan, Q., 2016. A nonstationary bias-correction technique to remove bias in GCM simulations. Journal of Geophysical Research: Atmospheres 121, 5718–5735. https://doi.org/10.1002/2015JD024159

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. International Journal of Climatology. https://doi.org/10.1002/joc.922

Schulzweida, U., 2019. CDO User Guide 1–206.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An Overview of CMIP5 and the Experiment Design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Tomczak, M., 1998. Spatial Interpolation and its Uncertainty Using (IDW) - Cross-Validation / Jackknife Approach. Journal of Geographic Information and Decision Analysis 2, 18–30.

Wagner, P.D., Fiener, P., Wilken, F., Kumar, S., Schneider, K., 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology 464–465, 388–400. https://doi.org/10.1016/j.jhydrol.2012.07.026

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

Wu, C.Y., Mossa, J., Mao, L., Almulla, M., 2019. Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River. Annals of GIS 25, 133–151. https://doi.org/10.1080/19475683.2019.1588781

Yang, X., Xie, X., Liu, D.L., Ji, F., Wang, L., 2015. Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region. Advances in Meteorology 2015. https://doi.org/10.1155/2015/563629

Zhang, X., Lu, X., Wang, X., 2016. Comparison of spatial interpolation methods based on rain gauges for annual precipitation on the Tibetan plateau. Polish Journal of Environmental Studies 25, 1339–1345. https://doi.org/10.15244/pjoes/61814

CHAPTER 4

A COMPARATIVE STUDY ON 10 AND 30-YEAR SIMULATION OF CMIP5 DECADAL HINDCAST PRECIPITATION AT CATCHMENT LEVEL

Abstract

Early prediction of precipitation has many positive benefits as it enables a longer time for proper planning and decision making especially for the water managers, agricultural stakeholders, and policy and decision-makers. However, due to ongoing climate change along with the chaotic nature of precipitation, a too early prediction may lead to inefficient planning and decision making due to higher uncertainty and poor skills of the predicted data as the climate models are imperfect replicas that need continuous improvement to predict future change. To investigate the difference between the short (a decade) and near-term (30 years) time simulation, this study aimed to compare the performance of 10 and 30-year simulation of CMIP5 decadal hindcast data of 0.05^0 spatial resolution at catchment level. For this, monthly hindcast precipitation of five general circulation models (GCMs): MIROC4h, MRI-CGCM3, MPI-ESM-LR, MIROC5, and CMCC-CM were downloaded from the CMIP5 data portal. Firstly, the model data were cut for the Australian region and then the unit of the GCMs data was converted to the millimetre. In the next step, the GCMs data were spatially interpolated onto 0.05^0 spatial resolution using the second-order conservative method by Climate Data Operator (CDO) tool. Monthly observed gridded data of 0.05^0 spatial resolution were collected from the Australian Bureau of Meteorology (BoM). In the last step, both the observed and GCMs data were cut for the Brisbane River catchment in Queensland, Australia. Models' performances are assessed compared with the corresponding observed values through four skill tests; mean bias, mean absolute error, anomaly correlation coefficient, and index of agreement. The results show that 30-year simulations have comparatively higher mean bias and lower skills than 10-year simulated data that seems related to the higher number of ensembles in 10-

This chapter has been published as: Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021. A comparative study on 10 and 30-year simulation of CMIP5 decadal hindcast precipitation at catchment level, in: Vervoort, R.W., Voinov, A.A., Evans, J.P. and Marshall, L. (Ed.), MODSIM2021, *24th International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand, pp. 609–615. https://doi.org/10.36334/modsim.2021.K5.hossain. However, few textual changes have been made to address the examiners' comments.

year simulations and the external forcing from increasing GHGs that models were not able to capture due to longer simulation period.

Keywords: Comparison, decadal, hindcast, precipitation, catchment

4.1 Introduction

The Coupled Model Intercomparison Project phase 5 (CMIP5) includes two types of modelling experiments; (i) long-term, which were usually designed for century timescales, and (ii) near-term, which were usually designed for 10-30 years' timescale called as decadal experiments (Meehl et al., 2009). In the near-term, there are two core sets of experiments; (i) 10-year hindcasts or predictions initialized in 1960, 1965, 1970 and thus every 5 years to 2005, (ii) 30-year simulation initialized in 1960, 1980, and 2005 and ending simulation by an additional 20 year (Taylor et al., 2012). Both the 10 and 30-year hindcasts predictions were initialized from the similar observed climate states, for a particular initialization year, but predicted for a different time span (either 10 or 30). However, based on the number of ensembles, multiple runs of a model with slightly different initialization conditions, the initialization conditions may be slightly different. In addition to the slightly different initialization conditions due to the different number of ensembles, the external forcing from increasing GHGs may dominate more the model response for 30-year simulation compared to 10-year (Taylor et al., 2012).

Due to the potential applications in many dimensions, decadal experiments have been paid much attention in the past decade in which temperature and temperature-based climate indices have been paid more attention compared to precipitation. However, precipitation is an important climate variable and hydrological aspect that has been significantly affected around the globe due to ongoing climate change. High temporal and spatial variability along with chaotic nature made it difficult for the climate models to project the change in the future precipitation than temperature (Zeke Hausfather, 2018). Climate change is an ongoing dynamic process that is being continuously changed and will continue in the future. However, the rate of change of future climate and its potential impact on precipitation is not certain. According to the IPCC report (IPCC, 2014), the change in the precipitation amount and its extreme events (e.g., heavy rainfall, droughts) will be higher in the future compared to the past depending on the geographical locations. As every year the climate condition is being changed, that would be intensified in the future, models' projected precipitation for longer timescale may become with higher uncertainty compared to shorter timescale predictions. However, there are always

some agreements and disagreements among the models as they are imperfect replicas but they are continuously improving to project the change in future precipitation (Zeke Hausfather, 2018).

Information of local climate features, especially precipitation, are very important for local water managers, developers of water supply infrastructures, and water-related other stakeholders. That is why research on the local level's climate variables and their potential applications are high in demand for transferring research-based scientific knowledge to increase the resilience of the society to climate change (Kumar et al., 2013). Since climate change and its impact on precipitation varies from region to region, therefore, it is important to assess the models' predicted precipitation for every individual region and for finer spatial resolutions for the regions where the most variable climate exists, like in Australia. Few studies (Gaetani and Mohino, 2013; Lovino et al., 2018; Mehrotra et al., 2014) are carried out to assess the CMIP5 decadal precipitation around the globe but no study was at the catchment level and finer than 0.5-degree spatial resolution. Early prediction of precipitation allows longer time for proper planning and decision-making process for managing water resources, assessing future water availability, agricultural planning, and large-scale investment for infrastructure development (Hansen et al., 2011; Jones et al., 2000; Mehta et al., 2013). However, too early may lead to inefficient planning and decision making due to higher uncertainty in the models predicted longer timescale precipitation data. The reason behind this is, the chaotic nature of precipitation over time and space as well as the climate models are not perfect enough. To examine that, this study aimed to compare the performance of 10 and 30 years simulated precipitation for CMIP5 decadal hindcast precipitation at a catchment level of 0.05^0 spatial resolution.

4.2 Data collection and processing

Monthly observed precipitation of 5km × 5km gridded data, produced through the Water Resources Assessment Landscape model (AWRA-L V5), was collected from the Australian Bureau of Meteorology (BoM). A detailed description of the observed data is available here (Frost et al., 2016).

Monthly hindcast precipitation data from five GCMs (Table 4-1), who have both the 30 and 10 years simulations, are collected from the CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). There are three initialization years; 1960 (1961-1990), 1980

(1981-2010), and 2005 (2006-2035) which have 30-year simulations. However, in this study, the initialization years 1960 and 1980 for 30-year simulation were selected as the observed data available until 2020 only. For better comparison, 30-year datasets were divided into three equal decades thus matching with the time span of 10-year simulation data initialized in 1960, 70, 80, 90, and 2000. Firstly, all available ensembles were averaged to produce a single dataset for each initialization and then the averaged datasets were subset for the Australian region. Secondly, the precipitation unit was converted to the millimetres and the datasets were spatially interpolated onto 0.05° × 0.05° (5km × 5km) grids matching with the grids of observed data. The second-order conservative (SOC) method was employed through Climate Data Operator (CDO) tool as SOC was found comparatively better than other commonly used spatial interpolation methods (Hossain et al., 2021a). Finally, both the observed and GCMs' data of 5km × 5km spatial resolution were cut for Brisbane River catchment, Queensland, Australia.

Table 4-1 List of models used in this study

Models	Resolutions (°lon × °lat)	10-year simulation.					30-year simulation.	
		1961-70	1971-80	1981-90	1991-00	2001-10	1961-90	1981-10
		Number of ensembles						
MIROC4h	(0.5625 X 0.5616)	03	03	06	06	06	03	04
MRI-CGCM3	(1.125 X 1.1215)	06	09	06	09	09	03	03
MPI-ESM-LR	(1.875 X 1.865)	10	10	10	10	10	03	03
MIROC5	(1.4062 X 1.4007)	06	06	04	06	06	06	04
CMCC-CM	0.75 X 0.748	03	03	03	03	03	03	03

4.3 Study area

Brisbane River catchment is in Queensland, the eastern state of Australia. It lies in between the latitudes 26.50S and 28.150S and the longitudes 151.70E and 153.150E. It has an area of 13549 square kilometres and a sub-tropical climate where maximum precipitation occurs during summer (December-January-February) and minimum precipitation in winter (June-July-August) (Climate Data, 2020).

4.4 Methodology

The comparisons between the 10 and 30-year simulations were performed based on the corresponding observed values through four quantitative performance metrics; Mean Bias, Mean Absolute Error (MAE), Anomaly Correlation Coefficient (ACC), and Index of Agreement (IA) which are usually referred to as skill tests.

4.4.1 Mean Bias

Choudhury et al., (2017) presented the difference between the raw ensembles' mean and the corresponding observed values as mean bias. As this study used the mean of all available ensembles, the mean bias can be obtained from the absolute differences between the models' raw ensembles' mean and the corresponding observed values [henceforth the mean bias will be referred to as bias].

$$Bias = |P_i - O_i| \tag{4.1}$$

Where P and O present models' raw and observed values respectively and the subscript i varies from 1 to n where n is the number of months in each dataset. These notations are the same also for the following skills.

4.4.2 Mean Absolute Errors (MAE)

As the name suggests, MAE presents the average magnitude of the differences between modelled and observed values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{4.2}$$

4.4.3 Anomaly Correlation Coefficient (ACC)

The centered ACC suggested by Wilks, (2011) measures the correspondence between the anomalies of model-predicted and observed values. A higher ACC value does not represent the higher accuracy of the predicted data but the anomalies of the predictions.

$$ACC = \frac{\sum\{(P_i-\bar{O})-\overline{(P_i-\bar{O})}\} \times \{(O_i-\bar{O})-\overline{(O_i-\bar{O})}\}}{\sqrt{\sum(P_i-\bar{O})^2}\sqrt{\sum(O_i-\bar{O})^2}} \tag{4.3}$$

Here, $\bar{O}$ presents the decadal mean of the observed values and the bar over the anomalies presents the mean of them.

4.4.4 Index of Agreement (IA)

IA suggested by Wilmot (1982), measures the accuracy of models' predicted values corresponding to observed values. IA is bounded between 0 and 1 where, the closer the value to 1, the more efficient the prediction is

$$IA = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i - O'| + |O_i - O'|)^2} \quad (4.4)$$

Here O' presents the mean of every individual year of the predicted period.

4.5 Results and discussion

In the Brisbane River catchment, there are 496 grids of 5km × 5km spatial resolution. The aforementioned skills tests are performed for every individual grid of all the selected models and initialization years. For simplicity, the results are presented here for a single grid point (27.50S and 153.050E), which is closest to a BoM operated automated weather station (AWS, located at 27.480S and 153.040E which is in the northern-east of the Wivenhoe). The bias was calculated for the monthly data and accumulated into yearly values for the sake of brevity in presentation. Fig. 4-1 presents the yearly total bias at the selected grid of the MRI-CGCM3 model. It is evident that 30-year simulation data shows comparatively higher bias as opposed to 10-year simulations and similar results were found for other models. However, the magnitude of the bias varies over the models, initialization years, and simulation periods. For instance, in 30 years simulation initialized in 1980 [henceforth, referred to as 1980(30)] all models showed comparatively lower bias as opposed to initialized in 1960 [henceforth, referred to as 1960(30)] with few exceptions during 1981-1990. In 10 years simulation, models also showed comparatively lower bias during 1991-2010 as compared to other initialization years. Taylor et al., (2012) mentioned that models might show higher biases at the beginning of the simulation period as compared to the other times. However, in this study for decadal timescale, models showed higher bias in the first decade for 1980(30) and in the second decade for 1960(30) except for MIROC4h, which showed higher bias in the first decade of both 30 years simulations.

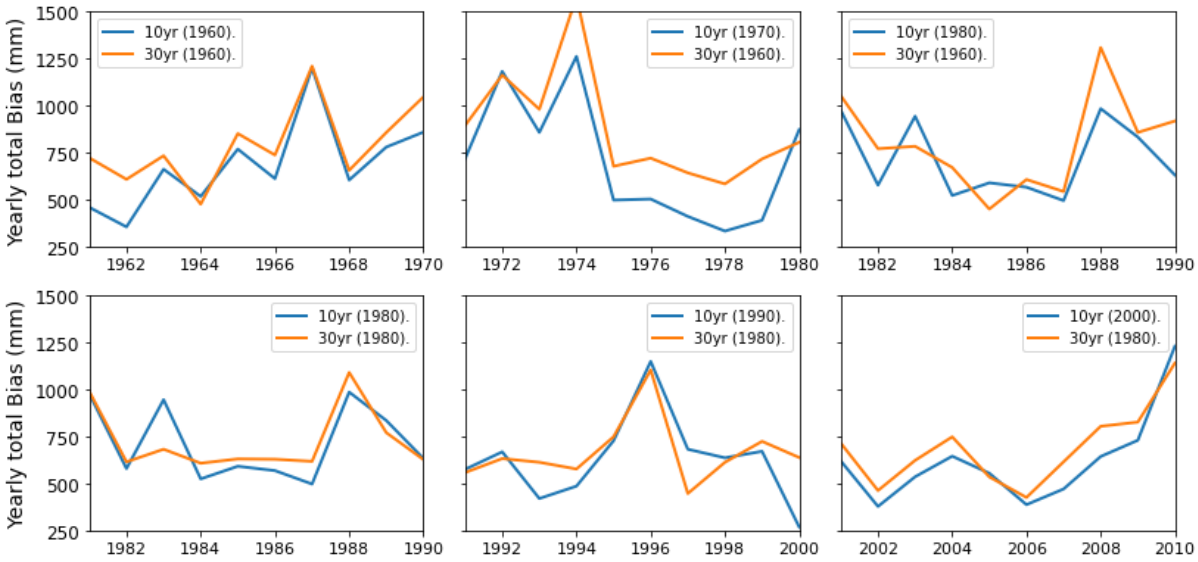


Fig. 4-1 Yearly total bias (obtained from monthly bias) comparisons of 10 and 30 years simulation for MRI-CGCM3 model. The vertical axis presents the yearly total of bias and the horizontal axis presents lead time (in a year). The initialization years are mentioned in the parenthesis of labels

For 10-year simulations, higher biases were observed after 2-3 years from the starting of the simulation of all selected models. During 1981-1990 all models showed their highest bias in the case of both 30 and 10 years simulations (Table 4-2) of individual models which was also evident in the skill tests. As individual models, MIROC4h and MRI-CGCM3 showed comparatively lower bias along with higher skills where MRI-CGCM3 is a little ahead of MIROC4h for both the 10 and 30-year simulations except in the last decade of 30 years simulations. On the contrary, CMCC-CM and MIROC5 showed higher bias along with lower skills where CMCC-CM is a little ahead during 1991-2010 and behind during 1971-1990 than MIROC5. Lower bias along with higher skills of models seems highly relevant to models' atmospheric resolutions where higher resolutions may reason comparatively lower bias and vice-versa that was also reported in previous studies (Jain et al., 2019; Lovino et al., 2018). Though having finer atmospheric resolutions, CMCC-CM showed higher bias and lower skills, it may be due to the different principle of CMCC-CM for predicting precipitation (Sakamoto et al., 2012) at the local level but may show better performance for different locations and variables (Lovino et al., 2018).

Table 4-2 Comparison of total bias at the selected grid

Models	10 years simulation			30 years simulation (1960)			10 years simulation			30 years simulation (1980)		
	1961-70	1971-80	1981-90	1961-70	1971-80	1981-90	1981-90	1991-00	2001-10	1981-90	1991-00	2001-10
MIROC4h	7986.7	7582.1	7387.3	7986.7	7629.8	7935.1	7387.3	6295.3	6840.1	8154.6	6762.9	6649.3
MRI-CGCM3	6830.2	7039.4	7128.3	7910.0	8775.2	7975.3	7128.3	6276.4	6185.8	7250.2	6646.2	6886.0
MPI-ESM-LR	7912.7	8356.9	7234.4	8593.6	8827.3	8045.1	7234.4	6930.5	7646.4	7145.3	7470.9	7141.8
MIROC5	7401.8	8250.5	7903.1	7743.3	8811.1	8530.2	7903.1	7716.8	7814.2	7903.1	7407.1	8900.8
CMCC-CM	8150.6	9063.1	8212.5	8150.6	9008.2	8093.3	8212.4	6454.2	7112.9	8212.4	6703.6	7046.1

Table 4-3 Comparison of MAE at the selected grid

Models	10 years simulation			30 years simulation (1960)			10 years simulation			30 years simulation (1980)		
	1961-70	1971-80	1981-90	1961-70	1971-80	1981-90	1981-90	1991-00	2001-10	1981-90	1991-00	2001-10
MIROC4h	66.55	63.18	61.56	66.55	63.58	66.12	61.56	52.46	57.0	67.95	56.36	55.41
MRI-CGCM3	56.91	58.66	59.40	65.91	73.12	66.46	59.40	52.30	51.55	60.42	55.38	57.38
MPI-ESM-LR	65.93	69.64	60.28	71.61	73.56	67.04	60.28	57.75	63.72	59.54	62.26	59.51
MIROC5	61.68	68.75	65.85	64.52	73.42	71.08	65.86	64.31	65.12	65.86	61.72	74.17
CMCC-CM	67.92	75.52	68.43	67.92	75.06	67.44	68.43	53.78	59.27	68.43	55.86	58.72

Table 4-4 Comparison of ACC at the selected grid

Models	10 years simulation			30 years simulation (1960)			10 years simulation			30 years simulation (1980)		
	1961-70	1971-80	1981-90	1961-70	1971-80	1981-90	1981-90	1991-00	2001-10	1981-90	1991-00	2001-10
MIROC4h	0.11	0.37	0.28	0.11	0.47	0.16	0.28	0.34	0.35	0.21	0.30	0.42
MRI-CGCM3	0.30	0.43	0.26	0.22	0.10	0.24	0.26	0.27	0.28	0.23	0.22	0.22
MPI-ESM-LR	0.24	0.20	0.26	0.10	0.27	0.20	0.26	0.28	0.17	0.32	0.29	0.29
MIROC5	0.23	0.30	0.27	0.24	0.19	0.04	0.27	0.17	0.20	0.14	0.22	0.20
CMCC-CM	0.15	0.13	0.06	0.15	0.10	0.16	0.06	0.28	0.20	0.06	0.19	0.15

Table 4-5 Comparison of IA at the selected grid

Models	10 years simulation			30 years simulation (1960)			10 years simulation			30 years simulation (1980)		
	1961-70	1971-80	1981-90	1961-70	1971-80	1981-90	1981-90	1991-00	2001-10	1981-90	1991-00	2001-10
MIROC4h	0.34	0.51	0.48	0.34	0.58	0.40	0.48	0.52	0.56	0.43	0.51	0.62
MRI-CGCM3	0.41	0.44	0.45	0.41	0.24	0.42	0.44	0.40	0.39	0.39	0.39	0.38
MPI-ESM-LR	0.41	0.30	0.41	0.34	0.43	0.42	0.41	0.43	0.37	0.50	0.48	0.51
MIROC5	0.35	0.33	0.30	0.30	0.31	0.35	0.30	0.31	0.35	0.30	0.31	0.35
CMCC-CM	0.15	0.11	0	0.15	0.06	0.25	0	0.35	0.3	0	0.25	0.19

This study also compared the total bias over the entire catchment and found that MRI-CGCM3 showed the lowest total bias during 1981-2010 and 1960-1990 in 10-year case only (Fig. 4-2). This may be due to capturing temporal variations comparatively better than other models.

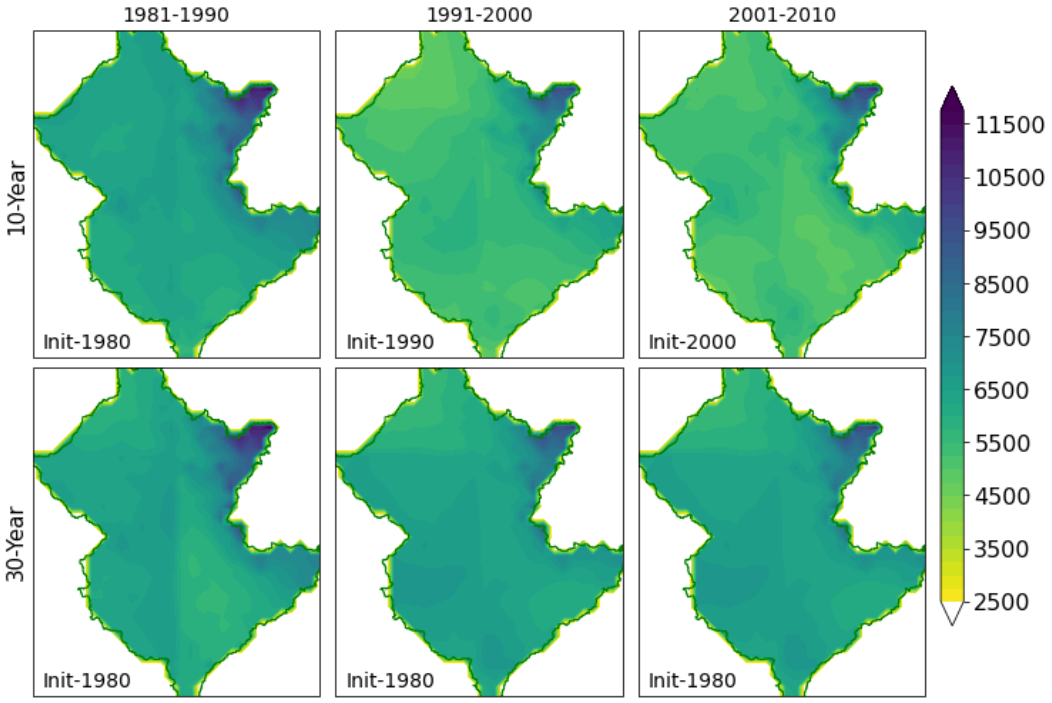


Fig. 4-2 Comparison of total (sum over 120 months) bias (in mm) between 10 and 30-year simulations across the catchment for the period of 1981-2010 of MRI-CGCM3 model. The periods are mentioned on the top of individual columns the initialization years are mentioned at the bottom left corner of individual plots

For 1960(30), MIROC4h showed a little higher bias than MPI-ESM-LR. In the case of longer time simulations (e.g., 30-year), only MIROC5 showed higher bias over the entire catchment during the ending of the predicting periods (not presented) whereas other models showed either the first or in the second decade. CMCC-CM showed the highest bias during the last decade of 1960 (30) only. Other skills; MAE, ACC, and IA for the selected grid point are presented in Table 4-3, 4-4, and 4-5 respectively. From the skill tests results, it is revealed that the model's skills correspond to the magnitude of bias and MAE. Higher bias resulted in lower skills and vice versa. During 1981-1990, models show the lowest skills and the highest errors whilst higher skills are observed during 1991-2010 which was also reported in a previous study (Hossain et al., 2021b). It is assumed that during 1961-1970 and 1981-1990, for both the 30 years and 10 simulations, models had the same initialization conditions except slight perturbation for different ensembles. However, 30 years simulations showed comparatively

higher bias and lower skills compared to corresponding 10 years simulations that may be due to the lower number of ensembles in 30-year simulation as opposed to 10-year. Higher bias in 30-year simulation during 1971-1990 and 1991-2010 due to the different initialization conditions where 10-year simulation had updated climatic condition for models' initialization compared to the 30-year. This may indicate that models may not capture the dynamic change of precipitation over time for a longer lead time. In addition, the models were dominated by the external forcing from increasing GHGs that may result in more time-varying bias which is referred to as drift (Choudhury et al., 2017; Mehrotra et al., 2014) encountered more in 30-year simulation than the 10-year. Though a 30-year time is not long enough in the climate studies perspective but compared to a decade it is longer.

4.6 Conclusion

This study aimed to compare the performance of 10 and 30-year simulations of CMIP5 decadal hindcast precipitation of 0.05-degree spatial resolution at catchment level. The skills of two 30-year simulation data were assessed and compared with their corresponding 10-year simulations. However, this study considered only two 30-year simulations; 1960(30) and 1980(30), and did not consider the 30-year simulation which was initialized in 2005 as the observed data until 2035 yet to observe. The performances are compared based on the calculated skills; bias, MAE, ACC, and IA. Based on the presented skills at the observed station and total bias over the entire catchment, this study finds comparatively higher bias and lower skills of 30-year simulation compared to 10-year simulations. Though the differences of bias are not significant, stakeholders may require prudence before making model-based decisions and planning. However, this study was limited to only the Brisbane River catchment that is why further investigation for other catchments at other locations is highly encouraged.

Acknowledgments

This study is financially supported by CIPRS and Data61 student scholarship of Curtin University and CSIRO (Commonwealth Scientific and Industrial Research Organisation) respectively which is jointly provided to the first author for pursuing his Ph.D. study at Curtin University, Australia. Authors are thankful to different working groups of CMIP5 for producing and making the model data available. The authors are also thankful to the Australian Bureau of Meteorology for providing their kind support and cooperation.

List of symbols

P	:	Model predicted values
O	:	Observed values
$\bar{O}$	:	Mean of the observed values for a decade
O'	:	Mean of the observed values for individual years
i	:	Number of events (months)

References

Choudhury, D., Sen Gupta, A., Sharma, A., Mehrotra, R., Sivakumar, B., 2017. An Assessment of Drift Correction Alternatives for CMIP5 Decadal Predictions. Journal of Geophysical Research: Atmospheres 122, 10282–10296. https://doi.org/10.1002/2017JD026900

Climate Data, 2020. Liberia climate: Average Temperature, weather by month, Liberia weather averages.

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

Gaetani, M., Mohino, E., 2013. Decadal prediction of the sahelian precipitation in CMIP5 simulations. Journal of Climate 26, 7708–7719. https://doi.org/10.1175/JCLI-D-12-00635.1

Hansen, J.W., Mason, S.J., Sun, L., Tall, A., 2011. Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture 47, 205–240. https://doi.org/10.1017/S0014479710000876

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., 2021a. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale. Indian Water Resources Society I, 285.

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021b. Intercomparison of drift correction alternatives for CMIP5 decadal precipitation. International Journal of Climatology joc.7287. https://doi.org/10.1002/joc.7287

IPCC, 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 1–112. https://doi.org/10.1017/CBO9781107415324

Jain, S., Salunke, P., Mishra, S.K., Sahany, S., 2019. Performance of CMIP5 models in the simulation of Indian summer monsoon. Theoretical and Applied Climatology 137, 1429–1447. https://doi.org/10.1007/s00704-018-2674-3

Jones, J.W., Hansen, J.W., Royce, F.S., Messina, C.D., 2000. Potential benefits of climate forecasting to agriculture. Agriculture, Ecosystems and Environment 82, 169–184. https://doi.org/10.1016/S0167-8809(00)00225-5

Kumar, S., Merwade, V., Kinter, J.L., Niyogi, D., 2013. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. Journal of Climate 26, 4168–4185. https://doi.org/10.1175/JCLI-D-12-00259.1

Lovino, M.A., Müller, O. V., Berbery, E.H., Müller, G. V., 2018. Evaluation of CMIP5 retrospective simulations of temperature and precipitation in northeastern Argentina. International Journal of Climatology 38, e1158–e1175. https://doi.org/10.1002/joc.5441

Meehl, G.A., Goddard, L., Murphy, J., Stouffer, R.J., Boer, G.J., Danabasoglu, G., Dixon, K., Giorgetta, M.A., Greene, A.M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., Stockdale, T., 2009. Can It Be Skillful? Bulletin of the American Meteorological Society 90, 1467–1486. https://doi.org/10.1175/2009BAMS2778.I

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. Journal of Hydrology 519, 2932–2951. https://doi.org/10.1016/j.jhydrol.2014.07.053

Mehta, V.M., Knutson, C.L., Rosenberg, N.J., Olsen, J.R., Wall, N.A., Bernadt, T.K., Hayes, M.J., 2013. Decadal Climate Information Needs of Stakeholders for Decision Support in Water and Agriculture Production Sectors: A Case Study in the Missouri River Basin. Weather, Climate, and Society 5, 27–42. https://doi.org/10.1175/WCAS-D-11-00063.1

Sakamoto, T.T., Komuro, Y., Nishimura, T., Ishii, M., Tatebe, H., Shiogama, H., Hasegawa, A., Toyoda, T., Mori, M., Suzuki, T., Imada, Y., Nozawa, T., Takata, K., Mochizuki, T.,

Ogochi, K., Emori, S., Hasumi, H., Kimoto, M., 2012. MIROC4h-A new high-resolution atmosphere-ocean coupled general circulation model. Journal of the Meteorological Society of Japan 90, 325–359. https://doi.org/10.2151/jmsj.2012-301

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An Overview of CMIP5 and the Experiment Design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

Zeke Hausfather, 2018. Explainer: What climate models tell us about future rainfall. Carbon Brief. Available at https://www.carbonbrief.org/explainer-what-climate-models-tell-us-about-future-rainfall/

CHAPTER 5

DRIFT IN CMIP5 DECADAL PRECIPITATION AT CATCHMENT LEVEL

Abstract

Over the last few years, the decadal prediction has been paid much attention for its potential applications in agriculture, hydrology and for assessing the climate impact on the various aspects of human life. Though the fidelity of decadal prediction specifically the hindcasts experiments through Coupled Model Inter-comparison Project Phase 5 (CMIP5) has been examined for many climate variables and at different temporal and spatial scales, the drift in CMIP5 decadal precipitation at a local scale remains unknown. Drift is the long-term time-varying systematic bias generated by GCMs while they revert to their equilibrium state from the forced initialized state. This study used seven general circulation models (GCMs) from five different modelling groups to examine the drift in monthly and seasonal mean precipitation from the CMIP5 decadal hindcasts for Brisbane River catchment, Australia. Drifts of individual model's ensemble mean (IMEM) and multi-model ensembles' mean (MMEM) at monthly and seasonal time scales were quantified and examined using four different skill tests. Results revealed that the magnitudes of drifts are higher in monthly precipitation than the seasonal mean precipitation. Next, the drift in hindcast precipitation was corrected using the mean drift correction method and found that the mean drift correction method is not sufficient to alleviate the drift in CMIP5 decadal precipitation. This suggests further research for an appropriate drift correction method for decadal precipitation. Comparing the drift and skill test results over the entire catchment, this study finds, MMEM showing the lowest drifts and outperformed in all models in skill tests.

Keywords: CMIP5, drift, precipitation, decadal, monthly, seasonal, catchment

---

This chapter has been published as: Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021. Drift in CMIP5 decadal precipitation at catchment level. *Stochastic Environmental Research and Risk Assessment* 8, 5. https://doi.org/10.1007/s00477-021-02140-8. However, few textual changes have been made to address the examiners' comments.

5.1 Introduction

In recent years, decadal climate predictions of Coupled Model Inter-comparison Project phase 5 (CMIP5) have attracted climate scientists due to their potential applications in many different fields such as agriculture, hydrology and socio-economic impacts due to climate hazards. A reliable prediction of climate variables (such as precipitation and temperature) from season to decade is very important especially for agriculture and agro-related businesses in order to plan for future water resources requirements (Hansen et al., 2011; Jones et al., 2000). Effective rainfall/streamflow prediction for longer time scales can help policymakers and water managers in reducing the impact of hydro-climatic extremes such as droughts and floods (Apurv et al., 2015; Mehta et al., 2013).

CMIP5 includes a set of decadal climate prediction experiments conducted using different general circulation models (GCMs) that provide both hindcasts and future predictions of climate variables (Taylor et al., 2012). Decadal hindcasts can be used to quantify the model biases and other systematic errors in model output that will help to correct the predicted future climate variables (e.g. precipitation). This will also enable us to use the corrected future rainfall variables as model input for hydrological modelling to predict the river streamflow (Islam et al., 2014). Prediction of river streamflow in a decadal scale will provide very important information for future infrastructure development. In CMIP5, the decadal hindcasts are from two core sets of experiments; the first set consists of 10-yr hindcasts initialized in 1960, 1965, and every five years to 2005 and the other set is the 30-yr simulations initialized in 1960, 1980 and 2005 (Taylor et al., 2012). Two common approaches of initialization are used in CMIP5 decadal experiments; full-field initialization and anomaly initialization. In full-field initializations, models' initial state is forced away from its equilibrium state to match as close as possible to its observed climate state while in the anomaly initialization, observed anomalies are added to the model climatology. Multiple runs carried out either for hindcasts or forecasts, of the same model with slightly different initialized conditions are referred to as ensemble members. As the climate models are the imperfect replica of the real-world phenomena, after running from an initialized state, the model tends to revert to its equilibrium state from the forced initialized states (Mehrotra et al., 2014). This results in spurious long-term linear or non-linear systematic errors in climate models (GCMs) which are referred to as "drift".

The model drift and its evolution over time can depend on various factors such as the variables of concern and their temporal and spatial variabilities, the models' type and the study regions. For instance, lower drifts have been reported in CMIP5 models than the CMIP3 models (Gupta et al., 2013). It has also been reported that the drift generally becomes less important at a global scale or over a larger area, and when considering average across multiple models (Gupta et al., 2012), but at a local scale it might be important when internal variability is large. Therefore, drift should not be neglected in predicting climate variables (Taylor et al., 2012). In terms of time scale, drift in decadal prediction experiments may affect most of the variables in a more complicated way than the long-term runs. Now drift has been an important issue in decadal experiments where little systematic directional bias from model to model and/or region to region is seen (Gupta et al. 2012, Gupta et al. 2013). As 'drift' in the climate model (GCMs) outputs hinders the credible applications of models output, therefore, drift correction is an essential prerequisite step before the application of climate model forecasts (or hindcasts). Taylor et al. (2012) recommended drift correction by applying relatively sophisticated bias correction methods. International Climate and Ocean: Variability, Predictability and Change Project Office (ICPO) has recommended removing lead time-dependent mean bias for drift correction of CMIP5 decadal prediction (ICPO, 2011). This method has been so far applied for the sea-surface temperature or temperature-based climate indices for drift correction. Kharin et al. (2012) used the CMIP5 decadal near-surface temperature of the CanCM4 model where they reported that the mean drift correction method would introduce errors due to the initialization date if the observation and models have different long-term trends. Other studies (Ho et al. 2012; Meehl et al. 2014; Meehl and Teng 2012) also argued that drift correction itself may introduce additional errors if proper data and methods are not considered. Choudhury et al., (2016) investigated sampling effects of two differently initialized CMIP5 decadal sea surface temperature (SST) and found that the mean drift correction introduces large biases in the considered variables.

To date, the drift of temperature and temperature-based climate indices have been paid much attention in many previous studies (Chikamoto et al., 2013; Choudhury et al., 2016; Hawkins et al., 2014; Kharin et al., 2012; Narapusetty et al., 2014). It may be because of the higher prediction skill of models found for temperature (Masanganise et al., 2013; Meehl et al., 2014; Mehrotra et al., 2014) than the precipitation. To the best of our knowledge, drift in precipitation has been given little attention (Gupta et al., 2013) and no studies conducted so far for CMIP5

decadal precipitation at the local scale. In Australia, only Mehrotra et al. (2014) used mean drift correction for precipitation along with two other atmospheric variables and reported lower skills of models for precipitation than the air temperature and geopotential height. However, Mehrotra et al. (2014) used 0.5^0 spatial resolution that covers 50 km × 50 km area over Australian ground which is a large spatial resolution for a region where climate variabilities are high. As precipitation is an important climate variable that has direct economic, environmental and social impacts especially for the agriculture-dependent countries, it is essential to check the drift in models' precipitation before taking any decision based on the model output. For this, precipitation drift should be studied at a local scale with finer spatial resolutions especially for a country like Australia where the precipitation shows high temporal and spatial variability (Gupta et al., 2013). Therefore, the objective of this study is to investigate the characteristics of drift in precipitation of CMIP5 decadal experiments by quantifying the drift and assessing the suitability of mean drift correction for CMIP5 decadal precipitation at a catchment level (0.050 × 0.050 ≈ 5km×5km). To assess the applicability of time-aggregated data for practical applications, this study uses monthly and seasonal mean precipitation and their skill tests results are presented for individual model's ensemble mean (IMEM) and multi-model ensembles' mean (MMEM).

5.2 Data collection and processing

5.2.1 Data collection

For monthly hindcasts decadal precipitation, there are ten GCMs available in the CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). Among them, seven GCMs (MIROC4h, EC-EARTH, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, MIROC5 and CanCM4) were selected for this study based on their spatial resolution and preliminary investigation. Relatively coarser spatial resolutions were found for IPSL-CM5A-LR (3.75° × 1.89°) and HadCM3 (3.75° × 2.5°) and hence they were excluded. Though CMCC-CM has a relatively finer resolution (0.75° × 0.75°) its simulation period was taken 121 months instead of 120 months for a decadal experiment. Because of this longer period of data set compared to other models, CMCC-CM was also not considered in this study. The data with a simulation length of 10 years (120 months) and models simulated with full-filled initialization for every five years during 1960-2005 was selected for this study. The number of ensembles for different models are different

and all available ensembles are averaged to produce their respective ensemble mean for each initialization. The name of models, their modelling groups, the available historical runs, and the ensemble number of individual models are given in Table 5-1.

Monthly gridded observed precipitation of $0.05^0 \times 0.05^0$ (5km × 5km) spatial resolution for entire Australia was collected from the Bureau of Meteorology (BoM), Australia. The gridded observed data was produced by BoM using the water resources assessment landscape model (AWRA-L V5) (Frost et al., 2016).

Table 5-1 List of models, their spatial resolutions and number of ensembles used in this study

Modelling Centre (or Group)	Model Name (Resolutions in degree)	Initialization Year (1960-2005)									
		60	65	70	75	80	85	90	95	00	05
		Number of ensembles									
EC-EARTH Consortium	EC-EARTH (1.125 X 1.1215)	14	14	14	14	14	14	14	14	10	18
Meteorological Research Institute	MRI-CGCM3 (1.125 X 1.1215)	06	08	09	09	06	09	09	09	09	06
Max Planck Institute for Meteorology	MPI-ESM-LR (1.875 X 1.865)	10	10	10	10	10	10	10	10	10	10
	MPI-ESM-MR (1.875 X 1.865)	03	03	03	03	03	03	03	03	03	03
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC4h (0.5625 X 0.5616)	03	03	03	06	06	06	06	06	06	06
	MIROC5 (1.4062 X 1.4007)	06	06	06	06	04	06	06	06	06	06
Canadian Centre for Climate Modelling and Analysis	CanCM4 (2.8125 X 2.7905)	10	10	10	10	10	10	10	10	10	10

5.2.2 Data processing

GCMs' spatial resolutions (e.g., 100-250km grids) are inadequate for regional studies as they have lack of information at local scales (Fowler et al., 2007; Grotch and MacCracken, 1991; Salathé, 2003). The regional climate models (RCMs) typically use dynamic downscaling for GCMs' large-scale information to local scales that are computationally intensive and it may introduce further bias of their own. Spatial interpolations of GCMs data from their native grids to finer resolutions are also common practice (Amengual et al., 2012; Mehrotra et al., 2014;

Miao et al., 2016) in climate research. This study used a spatial interpolation method for re-gridding the GCM data onto its finer spatial resolution. Skelly and Henderson-Sellers (1996) suggested that GCMs derived precipitation data should be treated as areal or grid box quantities. They also stated that spatial interpolation does not create any new information for the grid box quantities but only increases the spatial precisions. One can sub-divide the grid box in almost any manner until the original volume remains the same. For this, GCMs' data were subset for the Australian region and interpolated onto 0.05° × 0.05° spatial resolution using the second-order conservative (SOC) method in this study. Jones (1999) suggested that the precipitation flux must be remapped in a conservative manner in order to maintain the water budget of the coupled climate system. While sub-gridding the GCM data using the SOC method, the total precipitation volume of the original grid of GCM data are basically conserved over subsequent grids. The SOC method was selected here as it maintains a conservative manner (conserves precipitation flux) while sub-gridding the GCMs derived precipitation from their native grids to subsequent grids (Jones, 1999). Moreover, the SOC method was found suitable for spatial interpolation of precipitation data of regular rectangular grids (Hossain et al., 2021; Jones, 1999). For the suitability of direct comparison, spatial interpolation grids were set similar to the grids of observed data. Before interpolation, the precipitation unit was converted into millimetres.

For the seasonal precipitation, all ensemble members for each initialization were averaged, and the ensemble means were temporally aggregated to seasonal mean for the Australian seasons; Summer (December-January-February [DJF]), autumn (March-April-May [MAM]), winter (June-July-August [JJA]) and spring (September-October-November [SON]). Each dataset spans ten years, starting from January (1st year) and ending in December (10th Year). To make the complete seasons, January and February from the 1st year and December from the 10th year were discarded, thus resulting in the reduction of one DJF season in each dataset. To produce the multi-model ensembles' mean (MMEM), ensembles' mean of every individual initialization of all selected models were averaged to produce a single dataset. Finally, both monthly and seasonal mean precipitation data were subset for the selected Brisbane River catchment in Queensland, Australia (Fig. 5-1). There are 496 grids (at 0.050 × 0.050) within the Brisbane River catchment. It has an area of 13549.2 square kilometres and a sub-tropical climate where maximum rainfall occurs during summer (December-January-February) and minimum rainfall in winter (June-July-August) (Climate Data, 2020). This low to moderate

yearly rainfall variability along with the tropical climate nature is the main reason for selecting this catchment.

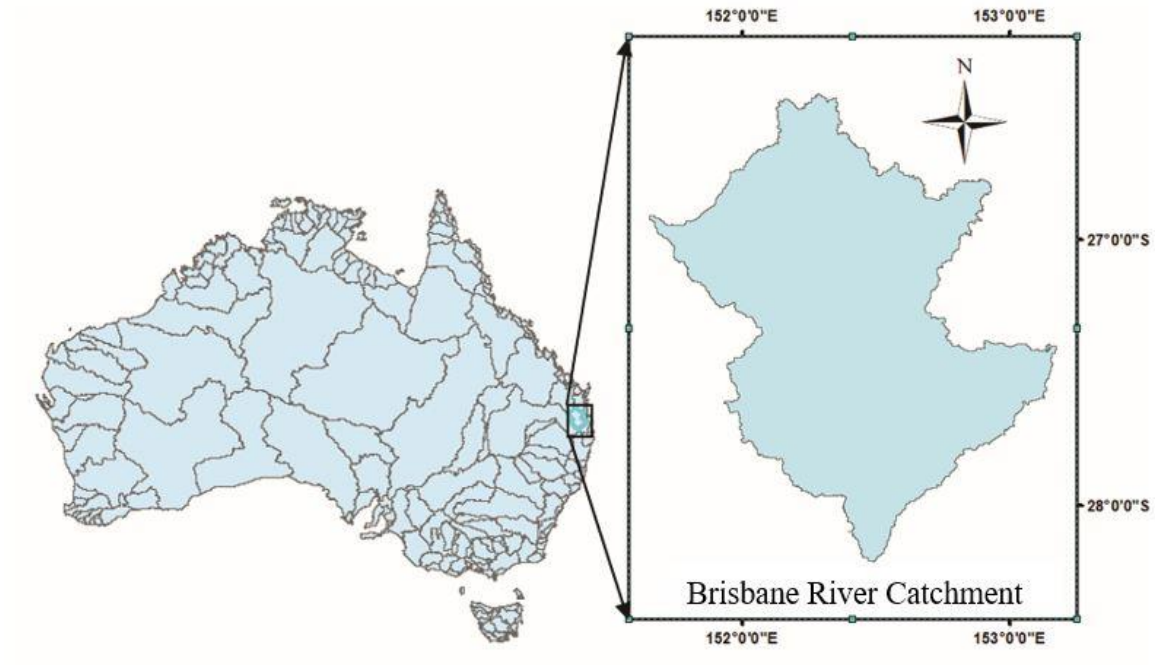


Fig. 5-1 Location of the Brisbane River catchment (inset)

5.3 Methods

5.3.1 Model drifts

The climate variables simulated in CMIP5 decadal experiments show systematic differences between the model hindcast and corresponding observed data. With time, models' simulated data initialized with the historical climate conditions tend to drift towards the model climatology and it produces systematic bias. This makes it necessary to correct the model drift in order to validate or interpret the predicted variables. To correct the drift of CMIP5 decadal experiments output, a simple drift correction method (known as mean drift correction) is suggested by ICPO (2011) which has also been used by Mehrotra et al. (2014) and Choudhury et al. (2016). This method assumes that the drift in monthly prediction is just a function of lead-time and is not affected by the external radiative forcing or model state. According to ICPO

(2011) method, the average value of the model's prediction ($\bar{F}_t$) and the corresponding average of observed data ($\bar{O}_t$) at a lead-time t are calculated as follows:

$$\bar{F}_t = \frac{1}{n}\sum_{i=1}^{n} F_{it} \qquad (5.1)$$

$$\bar{O}_t = \frac{1}{n}\sum_{i=1}^{n} O_{it} \qquad (5.2)$$

where F_{it} and O_{it} are the model's predicted and observed data respectively at lead-time t initialized at $i = 1, 2, \ldots n$. Here, n is the initialization years (n=10, corresponding to 1960, 1965 …. to 2005). For monthly precipitation, the lead-time $t = 1, 2, \ldots \ldots 120$ and for seasonal precipitation, lead-time $t = 1, 2, \ldots . 39$. The lead-time dependent drift (d_t) is the difference between $\bar{F}_t$ and $\bar{O}_t$ for the lead-time t.

$$d_t = \bar{F}_t - \bar{O}_t \qquad (5.3)$$

Quantitatively drift (Eq 5.3) is the mean of the biases over time (e.g., different initialization years). In Equations 5.1-5.3, the drift is calculated as the arithmetic mean of all biases over different initialization years (n value). That means the drift would be the same for all initialization years for each model.

Finally, the drift corrected values ($\hat{F}_{it}$) are calculated by subtracting the estimated drift (d_t) from the model's predicted values:

$$\hat{F}_{it} = F_{it} - d_t \qquad (5.4)$$

In this study, the model drifts in both monthly and seasonal mean precipitation are calculated at all 496 grids of the study area. For all initialization years, the model drifts were quantified for IMEM and their MMEM respectively.

5.3.2 Skill tests

The skills may be assessed by comparing the models' predicted historical values with the corresponding observations. At initialization year i and lead-time t, the model predicted decadal precipitation (F_{it}) and the drift corrected precipitation ($\hat{F}_{it}$) are assessed against the observed precipitation (O_{it}) using the following four statistical skill tests. These skill tests are performed

at each grid for all IMEM and MMEM, of all individual initialization years. As skill tests are carried out for all initialization years, the model predicted and drift corrected values are referred, in general, as F_t and the corresponding observed precipitation is referred as O_t in all skill tests.

Pearson correlation coefficient (PCC):

PCC measures the linear correlation between two datasets. PCC is used to measure the linear correlation between the modelled (and drift corrected) and observed values. Its value varies between -1 and 1 (perfect correlation).

$$PCC = \frac{\sum_{t=1}^{N}(F_t-\bar{F})(O_t-\bar{O})}{\sqrt{\sum_{t=1}^{N}(F_t-\bar{F})^2}\sqrt{\sum_{t=1}^{N}(O_t-\bar{O})^2}} \qquad (5.5)$$

where $\bar{F}$ and $\bar{O}$ represent the mean of all F_t and O_t respectively. Here, t = 1, 2, ……to N. N is the maximum lead-time (e,g., the maximum number of months-120 or the maximum number of seasons -39).

Index of Agreement (IA):

Wilmot (1982) suggested IA measures the accuracy of predicted data with respect to the observed data. IA values bounded between 0 and 1, where, the value closer to 1 presents the more efficient prediction (drift correction).

$$IA = 1 - \frac{\sum_{t=1}^{N}(F_t-O_t)^2}{\sum_{t=1}^{N}(|F_t-O'|+|O_t-O'|)^2} \qquad (5.6)$$

Here O' presents the mean of every individual year of the predicted period.

Root mean squared error (RMSE) and mean absolute error (MAE):

The RMSE and MAE both are used to measure the average magnitude of errors, the differences between the modelled and observed values. The RMSE is a quadratic scoring rule measuring the average magnitude of the errors and provides a relatively high weight to large errors because of squaring the errors before taking the average. This means RMSE is useful when large errors are undesirable. The MAE measures the average of the errors in a set of predicted and observed values. The MAE provides a linear score, which means it is weighted equally in

the average. RMSE and MAE values range from 0 to ∞, where the lower value indicates higher accuracy and vice versa.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(F_t - O_t)^2} \qquad (5.7)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|F_t - O_t| \qquad (5.8)$$

5.4 Data analyses and results

The model drifts are calculated at all 496 grids of the study area. Because of too many results produced for all grids, the results are presented here for the most representative single grid point (lat. -27.5 and lon. 153.05) which is closest to the observed rain gauge station (latitude -27.48 and longitude 153.04) of the Bureau of Meteorology, Australia. This might have less effect produced by the diffusive characteristics of the interpolation method. The drifts are calculated for IMEM and MMEM respectively for both monthly and seasonal data. Individual ensembles initialized in 1990 and their mean for MIROC4h, and IMEMs along with MMEM of the same initialization, for monthly precipitation, are presented in Fig. 5-2 and for seasonal mean precipitation are presented in Fig. 5-3 respectively.

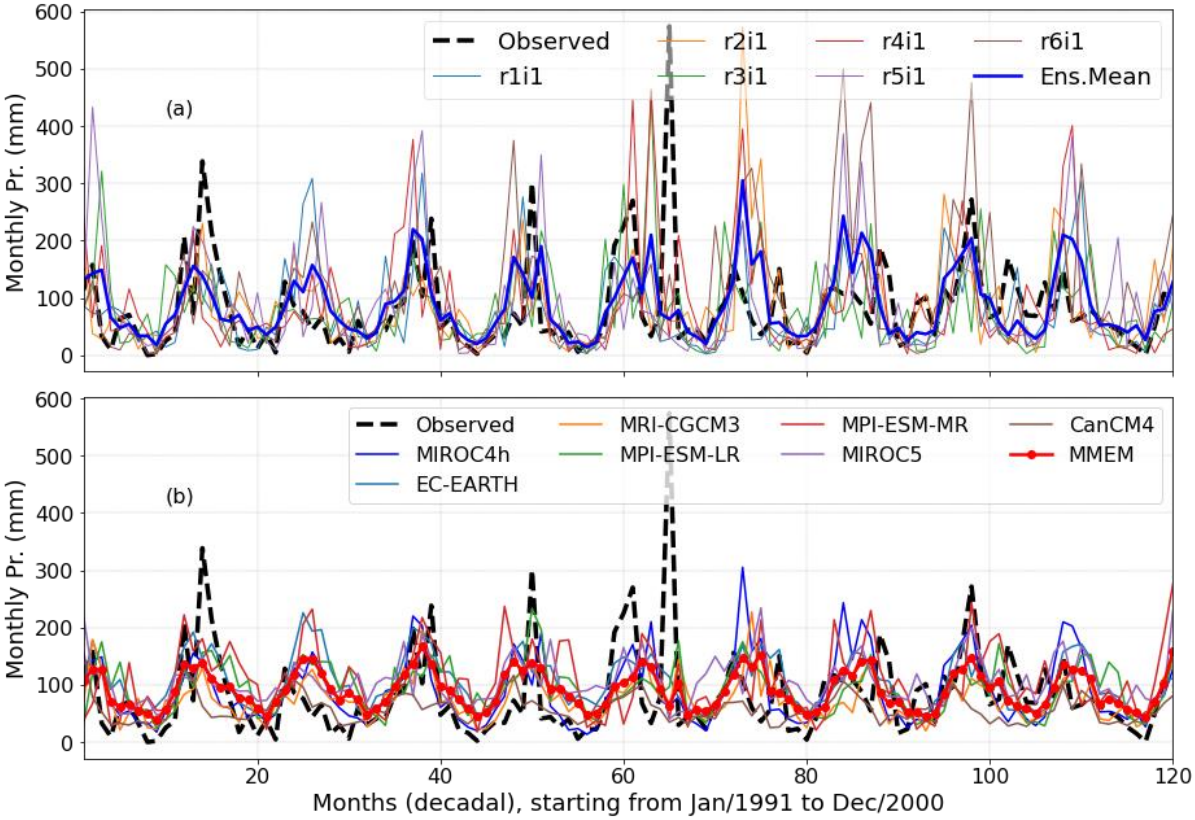


Fig. 5-2 Monthly precipitation of (a) individual ensembles (initialized in 1990) and their mean for MIROC4h and (b) the IMEMs along with MMEM of the same initialization

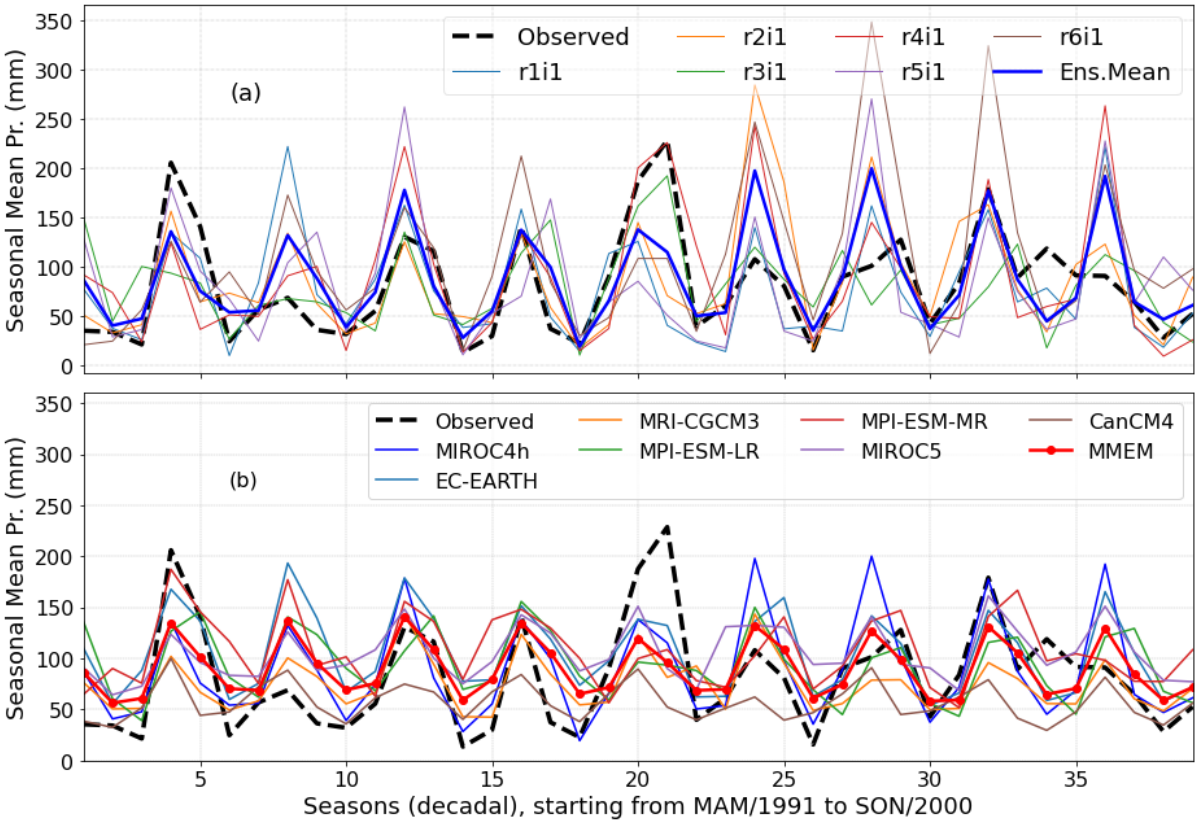


Fig. 5-3 Seasonal mean precipitation of (i) individual ensembles (initialized in 1990) and their mean for MIROC4h, and (ii) the IMEMs along with MMEM of the same initialization

The overall results considering all initialization years for all individual models show that their ensembles' means represent better with the observed values than their individual ensembles. The results for MMEMs were found showing better resemblances than the IMEMs.

5.4.1 Model drifts

Monthly precipitation

Drifts for monthly precipitation were calculated for all available grids and the results of the selected grid are shown in Fig. 5-4. All models show variations of drifts over time but the frequency and magnitude of variations are different for different models. Over the period of 120 months, the magnitude and frequency of negative drifts are found more pronounced albeit dependent on the model type. For instance, the CanCM4 model has a higher number of negative drifts while EC-EARTH shows more positive drifts.

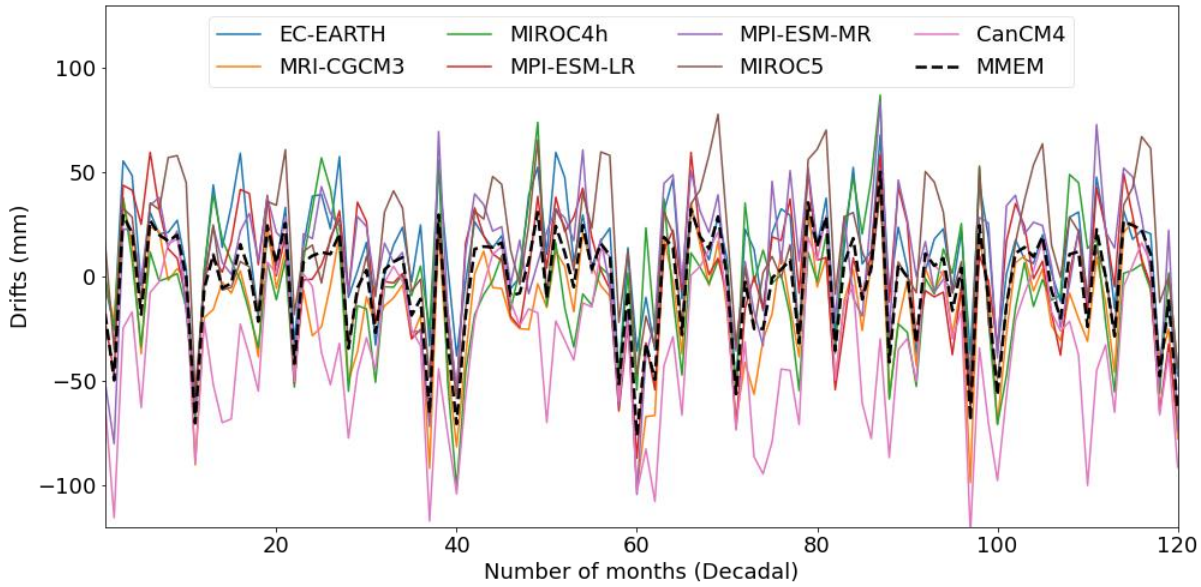


Fig. 5-4 Drifts of different models over 120 months (decade) at the selected grid point for monthly precipitation

In drift calculation, negative (or positive) drifts mean models tendency of under (or over) prediction. This under (or over) prediction varies over periods and seasons within the data. For monthly precipitation, almost all models showed higher negative drifts in the months of 10-12, 36-39, 57-60, and 97-100, with MIROC4h showing lower drifts for months 57-60 compared to other models.

As the climate models are not perfect to simulate reality, their output will differ from the observed values. In full-field initialization, the models are initialised based on the observations and they will be initially forced away from the equilibrium to match with the observed climate states and thus will revert to its equilibrium states over a period of time (Mehrotra et al., 2014; Taylor et al., 2012). This resulting spurious linear or nonlinear transition referred as drift (in this paper) depends on the time scale and the variables of concern being assessed in the models (Gupta et al., 2012). In this study, variations of peaks of calculated drift were observed along the time span and over the models. All models show multiple higher peaks of drifts for the months of 36-100 except CanCM4, which showed a high drift in the second month of the model run.

To investigate how the drift varies over the decadal time span, this study compared total drifts at the observed point as well as over the entire catchment after splitting the time span into three equal spells; 1st 40 (1-40 months), 2nd 40 (41-80 months) and 3rd 40 (81-120 months). The

higher drift was found in the 1st 40 months for EC-EARTH and CanCM4, 2nd 40 months for MRI-CGCM3, MIROC5, and MPI-ESM-MR and the 3rd 40 months for MIROC4h respectively. However, the MPI-ESM-LR showed an almost similar drift in all three spells. MMEM also showed a higher drift in the 3rd 40 months spell while the lowest drift was found in the 2nd 40 months.

The spatial comparison of mean drifts (average of the absolute values over each time spell) for monthly precipitation of MIROC4h and the MMEM are presented in Fig. 5-5 and all other models are presented in the supplementary materials in Fig. 5-S1. The lowest drift was found in the 2nd 40 months for MIROC4h and CanCM4 whilst EC-EARTH showed the lowest drift in the 3rd 40 months. MMEM showed an almost similar drift in all three spells. At the observed station, MMEM showed the lowest total drift followed by IMEM of MIROC4h, EC-EARTH and MRI-CGCM3 while CanCM4 showed the highest total followed by the IMEM of MIROC5 and MPI-ESM-MR respectively.

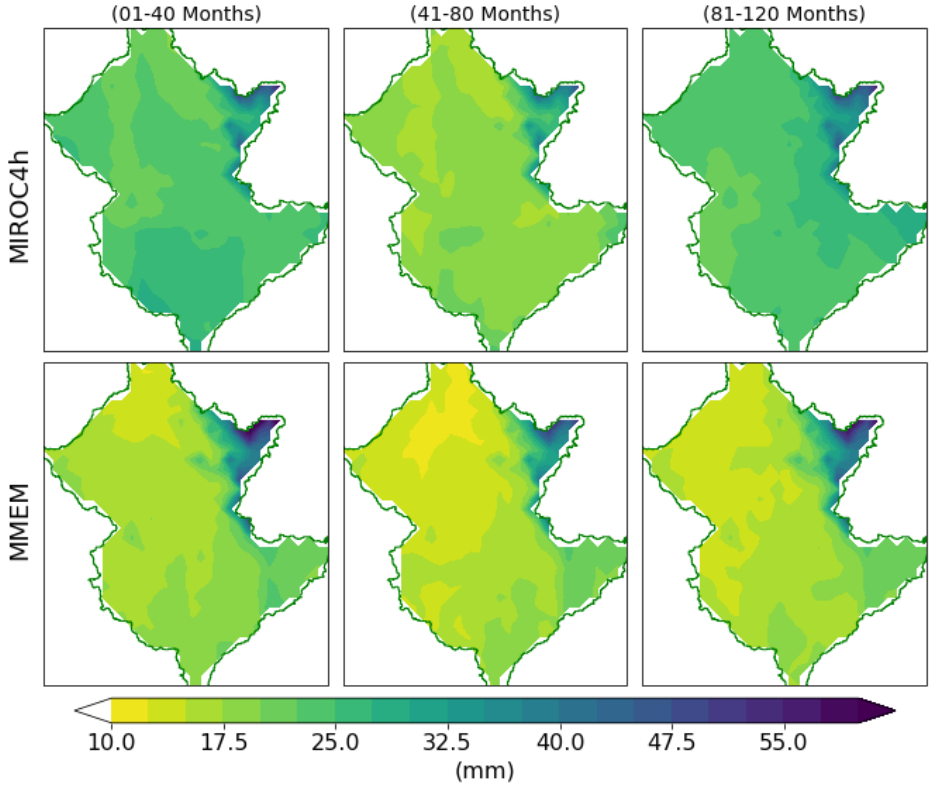


Fig. 5-5 Spatial variations of temporal mean drifts of MIROC4h (1st row) and MMEM (2nd row) for monthly precipitation. The temporal mean of the 1st, 2nd and 3rd spell, each of 40 months, are presented in the 1st, 2nd and 3rd column respectively

Nevertheless, over the entire catchment, MMEM showed the lowest total drift followed by IMEM of MPI-ESM-MR, MIROC4h, and EC-EARTH while MIROC5 showed the highest total followed by the IMEM of CanCM4. The drift comparison of IMEM and MMEM shows that the drift values of CanCM4 and MRI-CGCM3 are always lying underneath the drift values of MMEM's, whereas the drift in EC-EARTH and MIROC5 always remain above the drift in MMEM. Models MIROC4h, MPI-ESM-LR, and MPI-ESM-MR show almost similar types of drifts, which are more (less) than the MMEM drift in case of positive (negative) drift. However, the drifts observed in monthly precipitation of CMIP5 decadal experiments vary over time but depend on the model types (Fig. 5-4) which are similar to Gupta et al. (2013).

Seasonal mean precipitation

The drifts of all selected models and MMEM for seasonal mean precipitation at the selected grid are presented in Fig. 5-6. The spatial variations of mean drift (average of the absolute values over each time spell) over the catchment for seasonal precipitation of MIROC4h and MMEM are presented in Fig. 5-7 and all other models are presented in Fig. 5-S2. The results show that the higher drifts in seasonal mean precipitation are corresponding to the months showing higher drifts in monthly precipitation. However, there is a dampening in the overall magnitude and the frequency of the drift because of seasonal aggregations. For all models, the highest negative drifts were observed in the 19^{th} season and the second-highest in the 12^{th} season. Some other peaks were also observed but weren't as pronounced as in the monthly precipitation.

Similar to monthly precipitation, the highest negative drifts were found in CanCM4 whereas the higher positive drifts were found in EC-EARTH. The spatial variations of the drift in seasonal mean precipitation were very similar to monthly variations but with considerably lower magnitude (Fig. 5-S2).

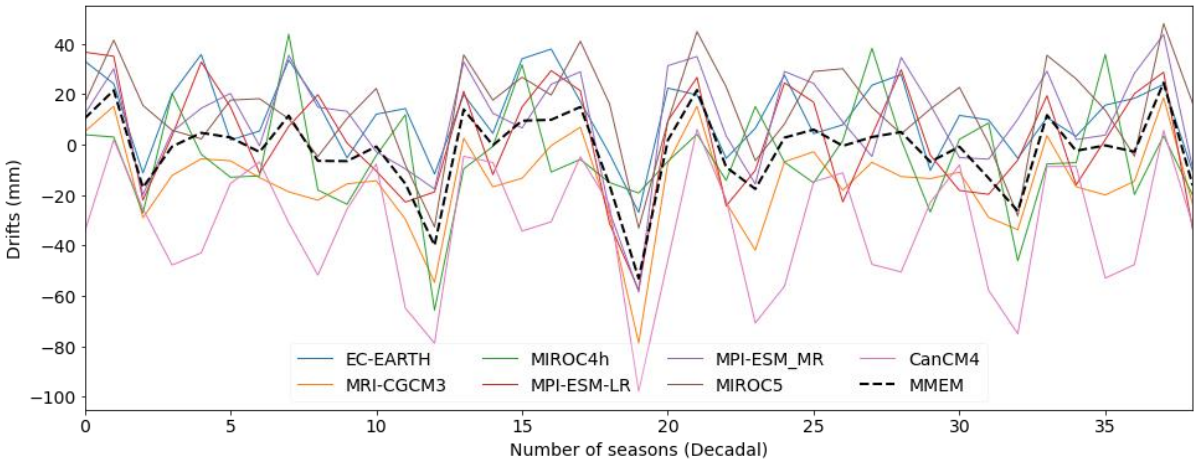


Fig. 5-6 Drifts of different models over the time span of 39 seasons at the selected grid point for seasonal precipitation

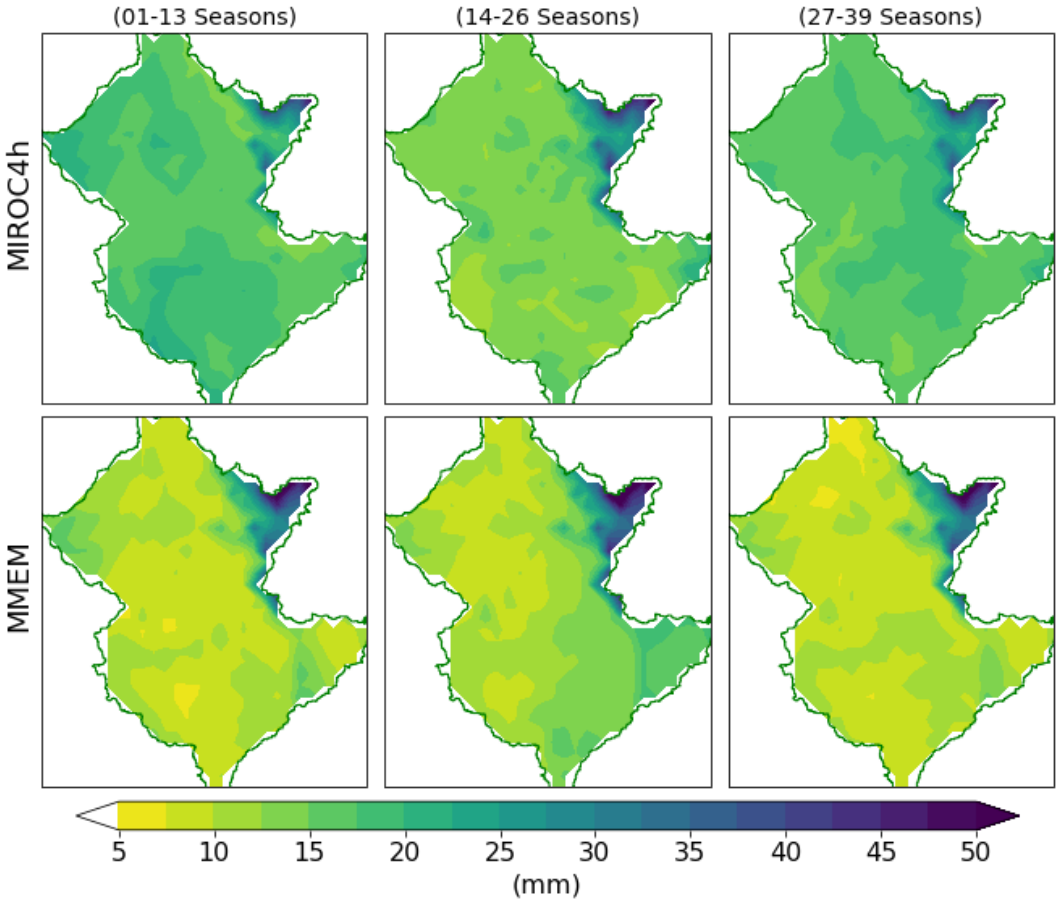


Fig. 5-7 Spatial variations of temporal mean drifts of MIROC4h (1st row) and MMEM (2nd row) for seasonal precipitation. The temporal mean of the 1st, 2nd and 3rd spell each of 13 seasons are presented in the 1st, 2nd and 3rd column respectively

5.4.2 Drift correction

Monthly precipitation

Drift corrections are performed by subtracting the calculated drift from the individual ensembles of the models, IMEM and MMEM for all initialization years using the mean drift correction method. Drift corrected individual ensembles and their mean of MIROC4h (the initialization year 1990) and drift corrected IMEM along with the MMEM of same initialization are presented in Fig. 5-8.

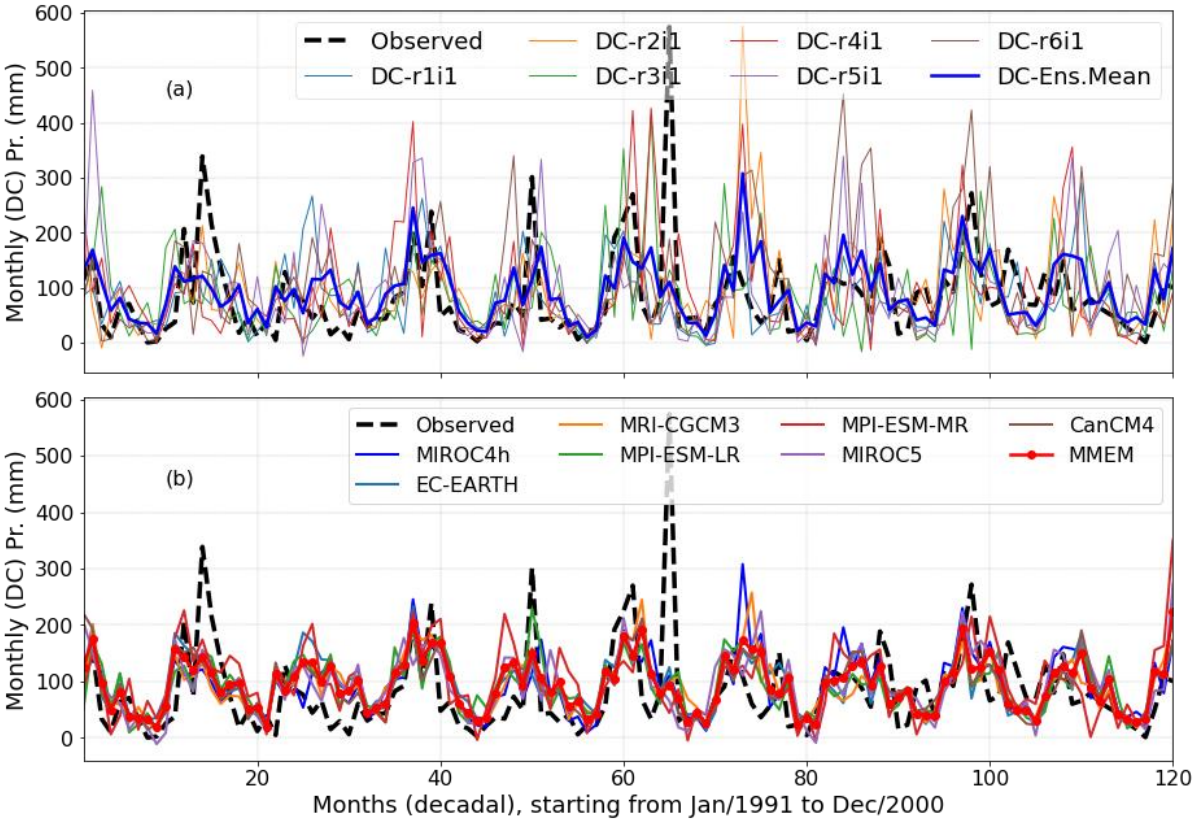


Fig. 5-8 Drift corrected (DC); (a) individual ensembles initialized in 1990 and their mean of MIRCO4h and (b) IMEMs along with MMEM for the same initialization of monthly precipitation

The drift corrected and uncorrected ensembles' mean precipitation are compared with the corresponding observed values for all initialization years of all models. It is observed that drifts are reduced after applying mean drift correction but also introduce additional errors for some models that were evident in skill tests. The magnitude of drift reduction varies over model

types. However, the overall drift reduction seems insufficient for the monthly precipitation as evidenced by the results of the skill tests.

Seasonal mean precipitation

Drift corrected individual ensembles of MIROC4h, IMEMs, and MMEM for seasonal mean precipitation are shown in Fig. 5-9. Comparison of the seasonal mean precipitation before and after drift correction shows improvements in the reduction of drift albeit the improvement is insignificant ($p>0.05$). It is worth noting that the drift in both monthly and seasonal precipitation was also assessed at other randomly selected grids within the Brisbane catchment but no noticeable changes were seen.

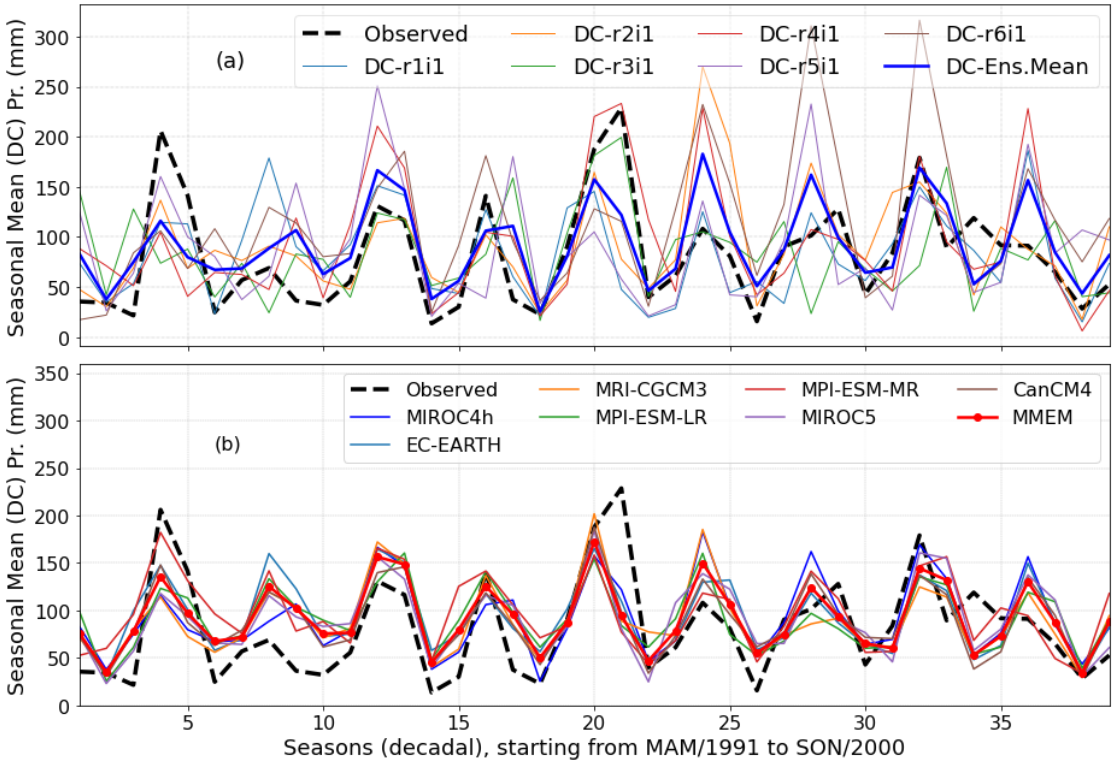


Fig. 5-9 Drift corrected (DC); (a) individual ensembles initialized in 1990 and their mean of MIRCO4h, and (b) IMEMs along with MMEM for the same initialization of seasonal mean precipitation

5.4.3 Skill test analysis

Model skills are analysed based on the skill tests results before and after drift correction. The results of four skill tests such as RMSE, MAE, PCC, and IA are presented in this section. These skill tests are performed for the ensemble mean of each initialization for each model for both

monthly and seasonal mean precipitation. The higher values of PCC and IA but lower values of errors (e,g., MAE and RMSE) represent the higher skill of models.

Monthly precipitation

The skill assessment of monthly precipitation before and after drift correction was performed and the change in skills (PCC, IA, MAE and RMSE) is shown in Fig. 5-10. The skill scores of PCC, IA, MAE, and RMSE before and after drift correction for monthly precipitation are shown in Fig. 5-S3 and S4 respectively in the supplementary materials. The results revealed that all models show lower skills in all performance metrics in the initialization years 1965 and 1980 while the second-highest RMSE is in 1970. As an individual model, MIROC4h shows higher IA followed by EC-EARTH and MPI-ESM-LR whereas CanCM4 shows the lowest skill in IA and PCC but it is almost similar to MPI-ESM-MR for MAE and RMSE. The highest RMSE and MAE were observed at the initialization year 1965 and the lowest was observed at 1975 for almost all models. MMEM showed the highest performance skills and outperformed all selected individual models in all skill tests and it remained almost similar after drift correction. After drift correction, the differences between IMEMs and observed values are minimized resulting reduction in errors (RMSE and MAE) and raising in skills (PCC and IA).

From Fig. 5-10, it is noted that models with higher skills before drift corrections are showing little to no improvement after drift correction and vice versa (see in Fig. 5-S3). For instance, MMEM, MIROC4h and EC-EARTH showed comparatively better skill before drift correction than the other models but did not improve much after drift correction. Consequentially, CanCM4 and MIROC5 showed the lowest performance before drift correction and benefitted enormously from the drift correction as evidenced by the notable improvements in PCC and IA. Except for MIROC4h, all models showed significant change ($p<0.05$) in PCC and IA whilst no model showed significant change ($p>0.05$) in errors (MAE and RMSE) after drift correction of monthly precipitation. In addition, MIRCO4h, MRI-CGCM3 and CanCM4 showed a drop of skills in some initialization years after the drift correction.

Interestingly, at the selected grid the temporal variation of PCC, IA, MAE and RMSE can be noted, where the models performed poorly (higher values of RMSE) in the earlier years especially before 1980 (Fig. 5-S3). Change in skills before and after drift correction was also studied for the entire catchment and observed that the improvements in model skill varied over

the catchment and initialization years where larger improvements in skill were noted in the earlier years than the latter. The changes of skills over the catchment for monthly precipitation of MIROC4h and CanCM4 (initialization 1990) are shown in Fig. 5-11. The results revealed that the grids with higher skills before drift correction were found with little improvement and/or drop-in skills after drift correction and vice versa.

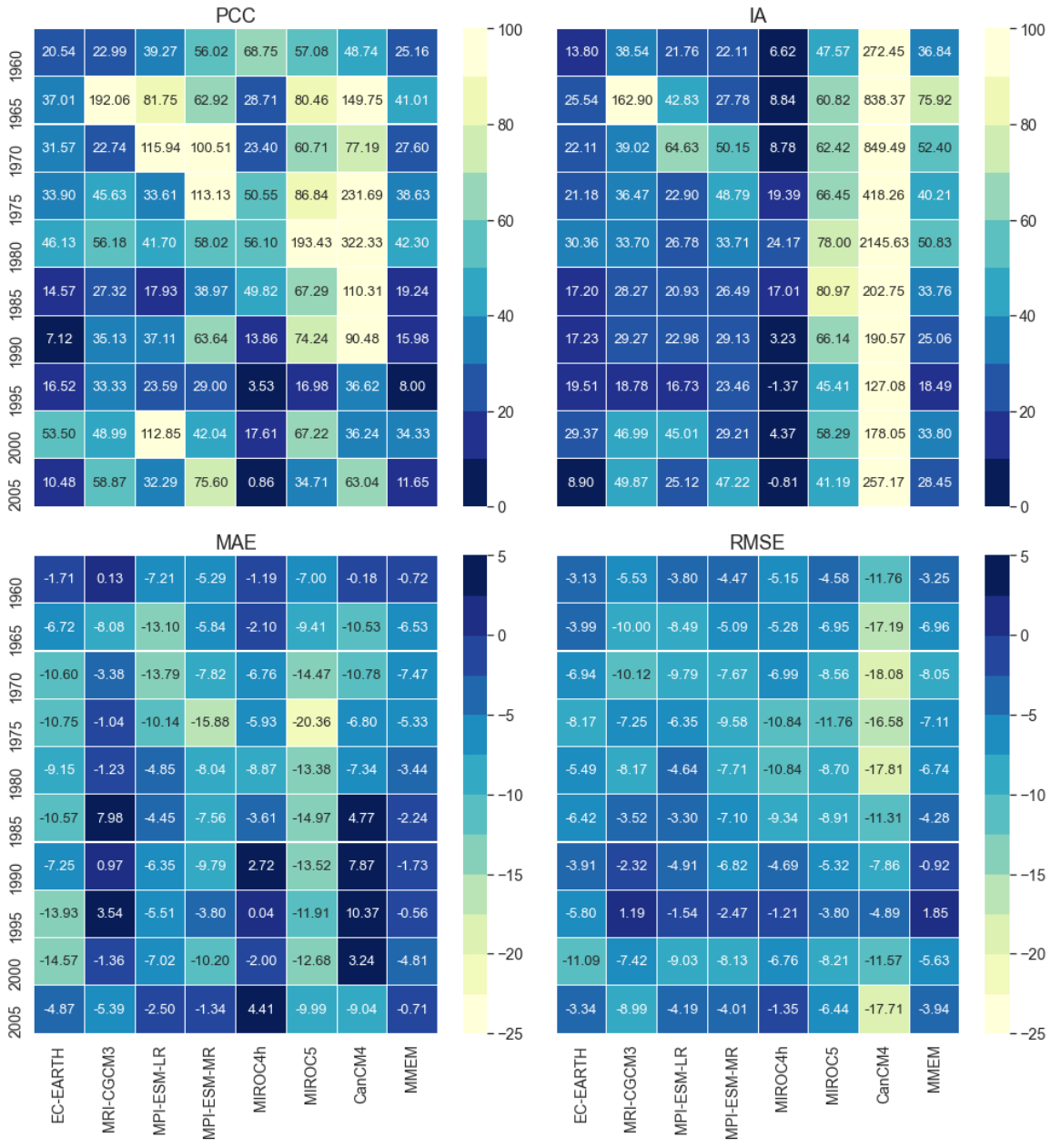


Fig. 5-10 Change in skills; PCC, IA, MAE and RMSE in between before and after drift correction of the models' monthly precipitation at the selected grid. The positive value indicates an increase in skills and vice-versa

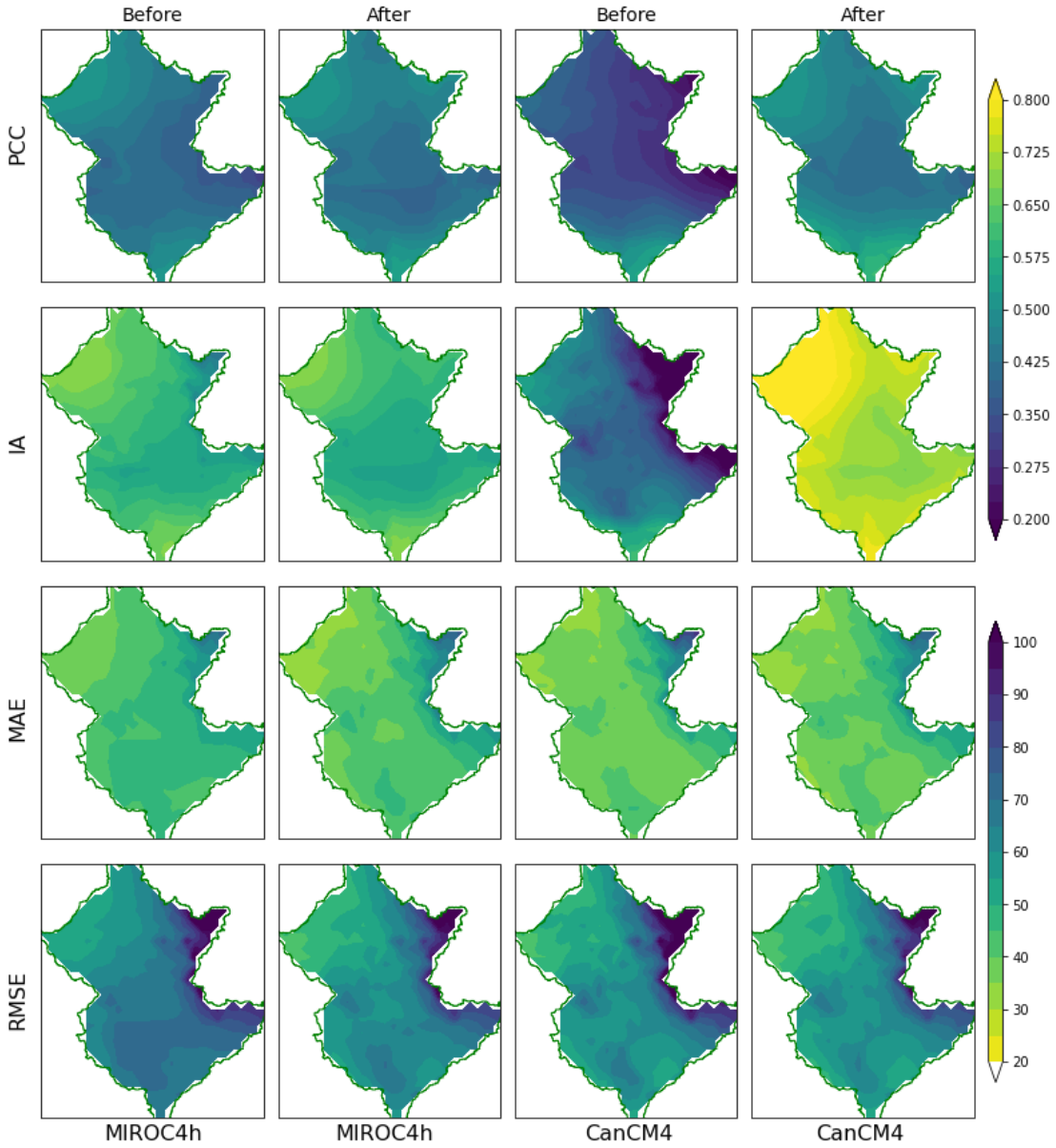


Fig. 5-11 Change in skills before and after drift correction for monthly precipitation of MIROC4h and CanCM4 initialized in 1990

Seasonal mean precipitation

Skill assessment of seasonal precipitation before and after drift correction at the selected grid was performed and the results are presented in Fig. 5-S5 and Fig. 5-S6 respectively and their changes are presented in Fig. 5-12. The results revealed that all models, except CanCM4, show higher skills for seasonal precipitation as compared to their corresponding monthly values before drift correction. It is due to minimizing the drift by the aggregation of monthly precipitation to the seasonal mean. By this aggregation, drifts of individual months were

averaged over the considered months, of respective seasons, hence lowering the drift of individual months to seasons and consequently resulting in higher skills. However, the value of IA for CanCM4 model precipitation dropped but no such reduction was noticed for the corresponding monthly values. After the drift correction, the skills of all models including CanCM4 improved (see Fig. 5-S6). Fig. 5-12 shows CanCM4 having the largest improvement, which is similar to that of monthly precipitation.

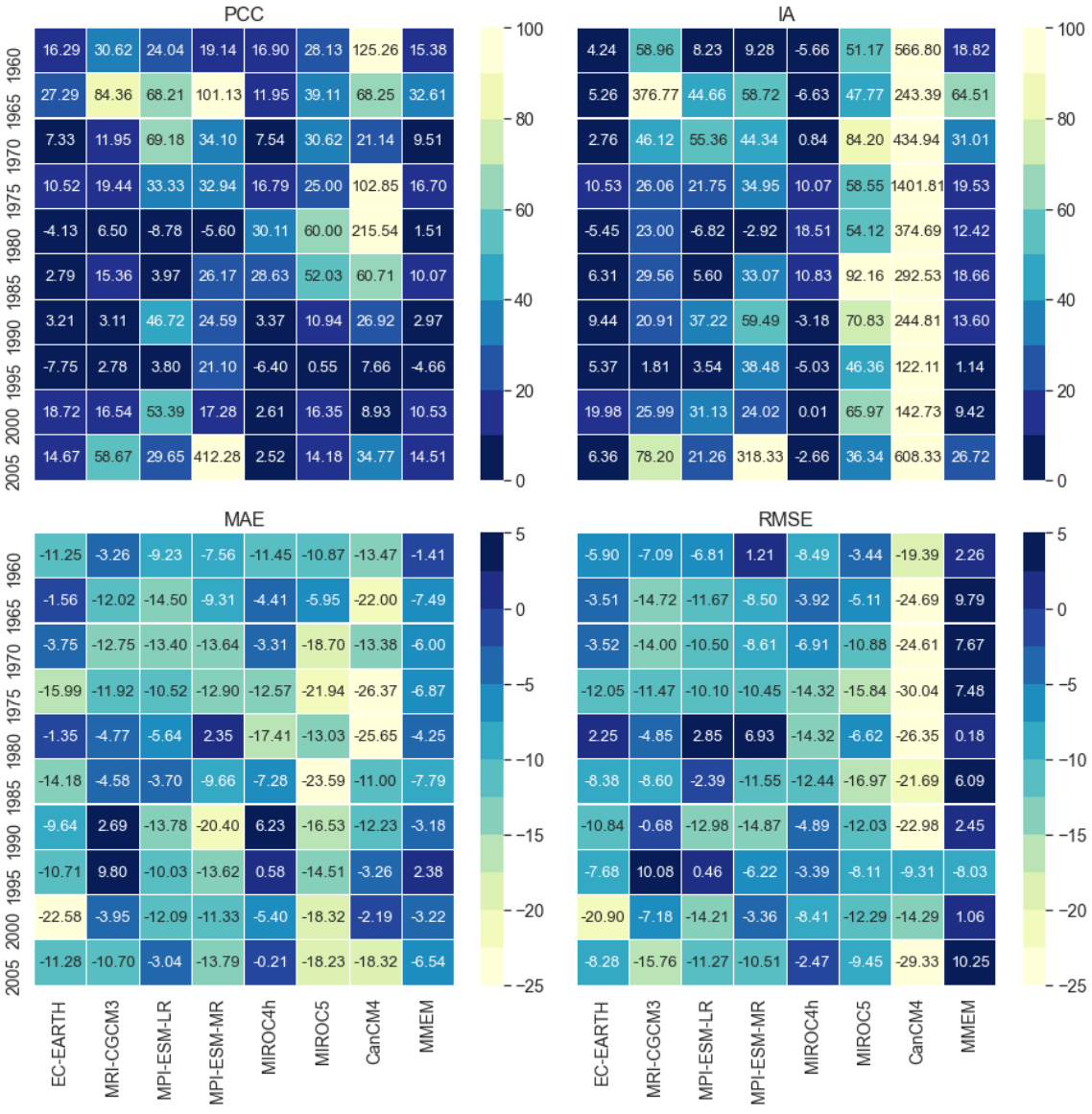


Fig. 5-12 Change in skills; PCC, IA, MAE and RMSE in between before and after drift correction of the models' seasonal mean precipitation. The positive value indicates an increase in skills and vice-versa

All models, except MIROC4h and EC-EARTH, showed significant change ($p<0.05$) in PCC and IA at the selected grid for seasonal precipitation while no model showed significant change ($p>0.05$) in errors (MAE and RMSE) after drift correction. Similar to monthly precipitation, changes in skills before and after drift corrections over the catchment for seasonal precipitation of MIROC4h and CanCM4 (initialization 1990) are shown in Fig. 5-13.

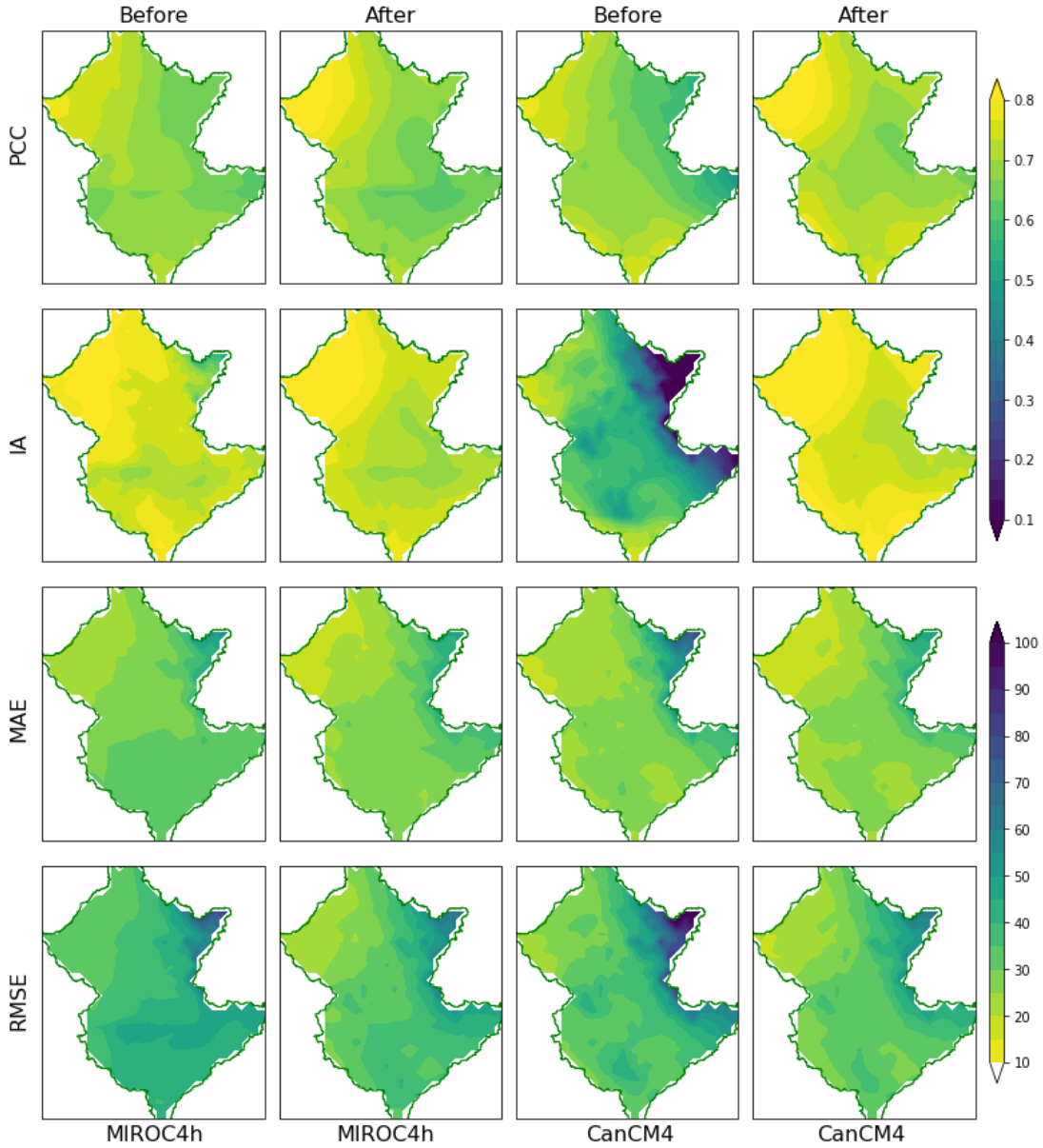


Fig. 5-13 Change in skills before and after drift correction for seasonal precipitation of MIROC4h and CanCM4 initialized in 1990

The changes of skills were found similar to monthly precipitation where the grids with higher skills before drift correction were found with little improvement and/or drop-in skills after drift

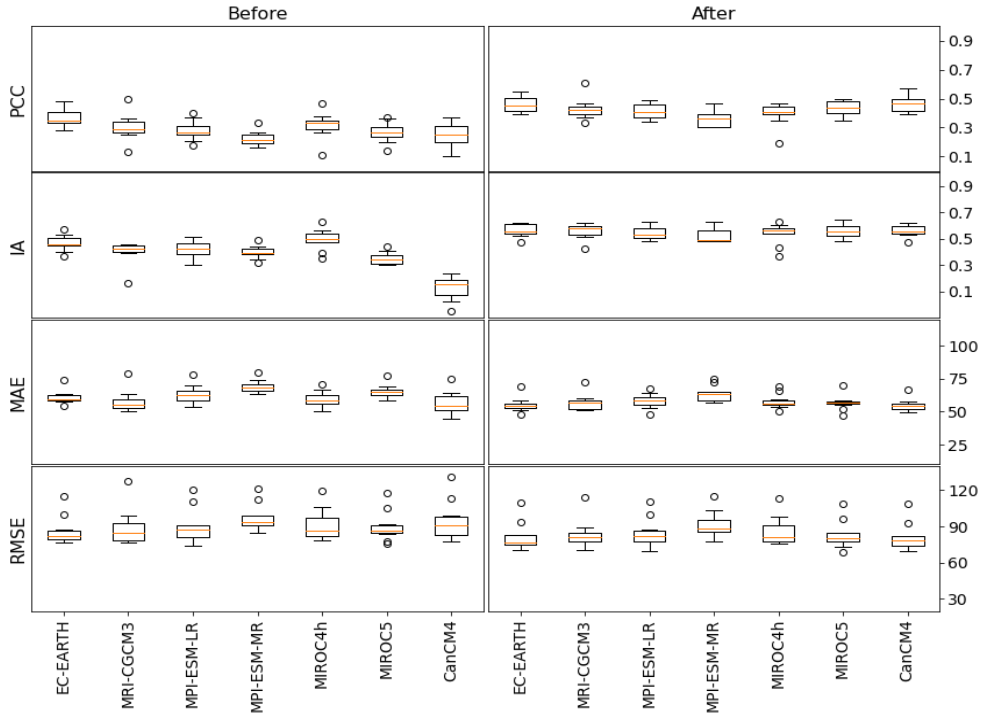
correction and vice versa. Upon comparing, the changes in skills of drift corrected monthly and seasonal precipitation; it was found that few models showed a drop of skill score after drift correction. For instance, the negative values of IA of MIROC4h and positive values of MAE of MRI-CGCM3, MIROC4h and CanCM4 (Fig. 5-10) show a drop of skills for monthly precipitation. For the seasonal mean precipitation, this drop of skills is more pronounced (see Fig. 5-12), which is an indication of introducing additional errors by the mean drift correction method that was also reported in other studies (Meehl et al. 2014; Meehl and Teng 2012).

5.5 Discussion

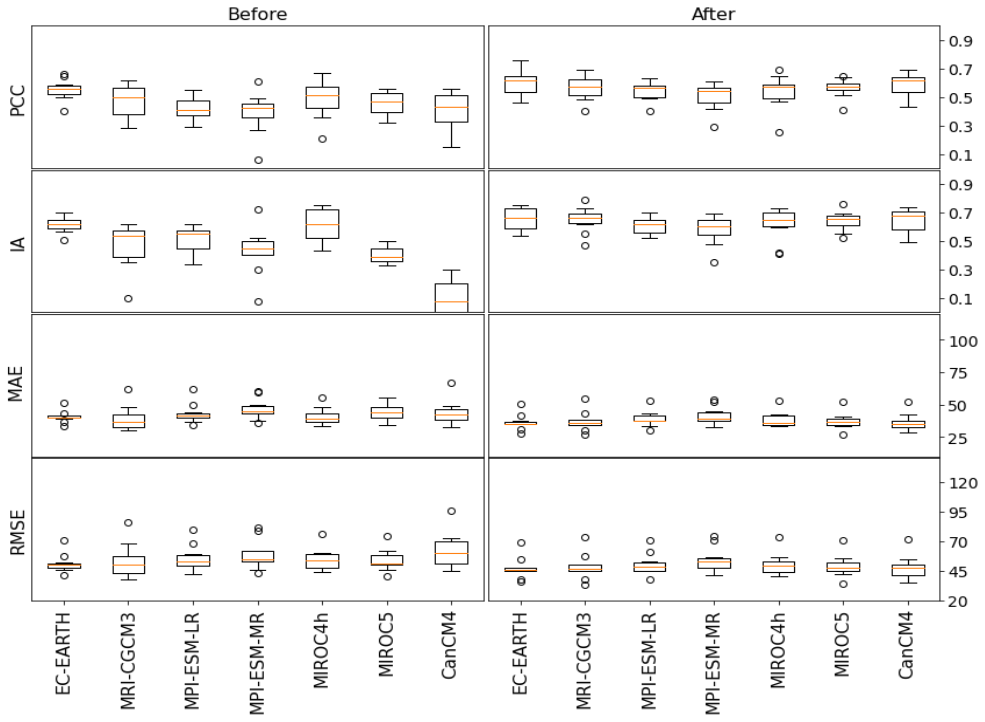
This study reports drift in monthly and seasonal mean precipitation from CMIP5 decadal experiments at a catchment level. The efficacy of applying the mean drift correction method to individual models as well as MMEM precipitation was assessed through four skill tests. In this study, MMEM showed the lowest drifts (highest skills) prior to drift correction and remained similar after drift correction which is also supported by previous studies (Jain et al., 2019; Lovino et al., 2018; Mehrotra et al., 2014). Gupta et al., (2013) mentioned that the drift may be negligible if MMEM is considered. However, the selection of models to produce MMEM is very challenging as averaging the model's output may further lead to loss of signals (Knutti et al., 2010) and the combination of all models may not provide better results always. For instance, MIROC4h and EC-EARTH showed comparatively lower drifts and higher skills while CanCM4 and MIROC5 showed higher drifts (lowest skills) and the models with higher drifts will influence the lower skill of MMEM. In this case, ignoring CanCM4 and MIROC5 to produce MMEM may result in better skills than the skills shown in Fig. 5-S3 to 5-S6. However, prioritising the models to produce better MMEM was not investigated in this study.

The results in Fig. 5-S3 to 5-S6 revealed that the skills of seasonal mean precipitation are higher than the corresponding monthly precipitation. From the comparison of drift in monthly and seasonal mean precipitation (before and after drift correction), the monthly precipitation was found with relatively higher drift. It is due to the high temporal frequency of monthly data where drifts are encountered for individual events and causes lower skills. However, the reduction in drift by the mean drift correction method seems insufficient for few models as evident from the skill tests results. Caution and care must be taken about the drift before selecting the temporal frequency of the precipitation data. The drift can be reduced by

considering the MMEM and/or aggregating the monthly precipitation to a seasonal mean. Aggregating to seasonal mean will give a reasonable reduction compared to the reduction from the mean drift correction of the monthly precipitation but in that case, the temporal frequency of the data will be different that may change the suitability of its applications. In addition to considering MMEM and aggregating to seasonal mean, further drift reduction was attained by applying the mean drift correction method but it did not show a significant change in skills for CMIP5 decadal precipitation for seasonal time scale at a catchment level. The comparison of skills before and after drift corrections at a selected grid point over all the models and initialization years for monthly and seasonal mean precipitation is shown in Fig. 5-14 by a Box-Whisker plot. From this comparison, it is observed that the overall skills of models' improved after the drift correction but this improvement were not much for monthly precipitation. For instance, IA and PCC are two important indicators that measure the accuracy and linear correlation respectively, but their lower scores were noted even after drift correction. In Fig. 5-14(a), the averages (over the models) of PCC and IA of monthly precipitation after the drift correction are 0.42 and 0.55 respectively but these were 0.285 and 0.38 respectively before the drift correction. This means, these values are raised by 30% after drift correction but their drift corrected values are still rather low. Similarly, average errors (MAE and RMAE) of monthly precipitation after the drift correction have been dropped by 6-7% but their values (57.7 and 84.5 respectively) are relatively still high (see Fig. 5-14a). For the seasonal precipitation (Fig. 5-14b), the improvements in the average (over the models) PCC and IA before and after drift correction are slightly higher than the improvements observed in the average errors (MAE and RMSE). However, the drift corrected overall skills for seasonal precipitation are considerably higher than their counter monthly values.



(a) Monthly precipitation



(b) Seasonal precipitation

Fig. 5-14 Comparison of skills for monthly and seasonal mean precipitation before and after drift corrections at a selected grid point over all the initialization years

Several researchers have shown that the rainfall/and or streamflow predictions at the longer lead time have positive benefits both from an environmental and social perspective (Hansen et al., 2011; Jones et al., 2000; Shams et al., 2018). It has a direct and indirect influence on the national economy for both developing and developed countries, especially those who experience large variations in climates, and where agriculture is a significant contributor to the national economy (Carberry et al., 2002; Jones et al., 2000). The climate variability can have drastic consequences on the Australian economy, where a typical major drought in a season can cause about a 10% reduction of agricultural production and about 1% reduction of the gross national product (White, 2000). Seasonal climate prediction plays an important role in agricultural management activities like crops type selections, seeding, irrigation planning, harvesting and so on (Jones et al., 2000). Furthermore, seasonal climate prediction may result in higher public benefit than private benefit (Mjelde et al., 2000). The importance, benefits and potential application of seasonal climate prediction have been reviewed by Paull (2002) and others, where conclusions were based on different case studies, surveys and model-based outcomes. Careful consideration of the spatial and temporal scale of climate prediction is important for any practical application. For instance, predicting climatic variables at a lead time of several months to a couple of seasons are important for the farmers and other stakeholders in making their decision for farming or agricultural management at a local scale to a district level (Paull, 2002). Moreover, water resources managers, groups of wholesalers, agronomic business organizations, processors of agronomic products, different public and private investors (e.g., insurance companies) demand predictions with longer lead times (Paull, 2002). On one hand, farmers who form the majority of the potential users of climate prediction have identified increased confidence in using climate predictions at high spatial resolutions as a research priority (White, 2000). On the other hand, groups handling agronomic products and those who play role in planning and decision-making demand climate prediction beyond the seasons to annual time scale. CMIP5 decadal precipitation prediction can be a good choice for the above-mentioned stakeholders who are interested in future precipitation data for the decadal timescale with a finer spatial resolution at the catchment scale. This paper presents the application of CMIP5 decadal precipitation at the catchment scale and identifies the presence of significant drift within the decadal model data. The drift may be reduced by using MMEM instead of IMEM and seasonal mean precipitation instead of monthly precipitation, which may be further reduced by the application of a suitable drift correction method. The application of

the mean drift correction method shown in this study provides its inadequacy in correcting the drift in model precipitation that suggests investigating the alternative drift correction methods for correcting the drift of decadal precipitation of CMIP5 experiments.

However, decadal prediction in CMIP5 was the first attempt to examine climate predictability and explore the predictive capabilities of forecast systems on decadal time scales. Boer et al. (2016) reported that the models participating in CMIP5 may show considerably lower skills in reproducing precipitation using different initialization methods and the lessons learned from CMIP5 have been taken into consideration in the Decadal Climate Prediction Project (DCPP), which has contributed to the sixth Coupled Model Intercomparison Project (CMIP6). Decadal data of CMIP6 includes more frequent hindcast start dates (yearly instead of 5-yearly) with larger ensembles of hindcasts dates than CMIP5 but there was no study using CMIP5 decadal precipitation for an Australian catchment in a finer resolution. The current study was undertaken as a first step to investigate the drift characteristics of CMIP5 decadal precipitation for an Australian catchment. As a follow-up study, further research is recommended to compare the drift produced by the CMIP6 decadal precipitation with that of CMIP5 model output in a catchment scale. This would provide more robust understandings about the characteristics of drifts for the practical uses of models' predicted precipitation data.

5.6 Conclusion

This paper investigates the characteristics of drifts in the precipitation of CMIP5 decadal experiments. Prediction of climate variables using CMIP5 decadal experiments is an emerging research area for longer timescales (e.g., beyond the season to annual time-scales). Most of the previous researches used CMIP5 data for temperature or temperature-based indices but none of the previous studies investigated the CMIP5 decadal precipitation for a finer resolution of $0.05^0 \times 0.05^0$ (5km × 5km) at catchment level. This study made the first attempt to use CMIP5 decadal precipitation data at this finer resolution of $0.05^0 \times 0.05^0$ matching with the observed data for an Australian catchment (Brisbane catchment, Queensland). First, the CMIP5 decadal precipitation data was interpolated to this resolution using the second-order conservative method and next, drift was evaluated comparing with the observed data. The results from four skill tests show significant model drifts in CMIP5 output for decadal precipitation. Analysis of seasonal and monthly precipitation revealed that MMEM could produce lesser drift than IMEM

or individual ensembles. MMEM was calculated using all seven GCMs in this study but the lesser drift also could be obtained if the MMEM was estimated after prioritising the GCMs based on their individual model drift. However, drift cannot be eliminated using MMEM only and hence there is a necessity of correcting the model drift by the use of a suitable drift correction method. A simple drift correction method (mean drift correction) was used in this study for both monthly and seasonal precipitation but the results were found not very promising which indicates for further investigations to select from an alternative drift correction methods.

Acknowledgements

This project is supported by CIPRS scholarship of Curtin University and Data61 student scholarship of CSIRO (Commonwealth Scientific and Industrial Research Organisation) which is jointly provided to the first author for his PhD study at Curtin University, Australia. The authors would like to thank the working groups of CMIP5 for producing the model data and making it available for the researchers. Authors also acknowledge the support of Australian Bureau of Meteorology for providing the observed gridded data and catchment's shape files. The authors also thank three anonymous reviewers for their constructive and professional suggestions which improved the paper.

List of symbols

F_{it}	:	Model forecasted precipitation at lead time t of initialization year i
t	:	Lead time
i	:	Initialization year
O_{it}	:	Observed Precipitation (corresponding of F_{it})
d_τ	:	Drift
$\hat{F}_{it}$	:	Drift corrected model precipitation at lead time t of initialization year i
F_t	:	Model forecasted precipitation at lead time *t* (Raw/Drift corrected, where applicable)
O_t	:	Observed precipitation at lead time *t*

$\bar{O}$: Mean (decadal) of observed precipitation

O' : Mean of individual year of observed data

$\bar{F}$: Mean (decadal) of forecasted precipitation

N : Number of months/ Seasons

References

Amengual, A., Homar, V., Romero, R., Alonso, S., Ramis, C., 2012. A statistical adjustment of regional climate model outputs to local scales: Application to Platja de Palma, Spain. Journal of Climate 25, 939–957. https://doi.org/10.1175/JCLI-D-10-05024.1

Apurv, T., Mehrotra, R., Sharma, A., Goyal, M.K., Dutta, S., 2015. Impact of climate change on floods in the Brahmaputra basin using CMIP5 decadal predictions. Journal of Hydrology 527, 281–291. https://doi.org/10.1016/j.jhydrol.2015.04.056

Boer, G.J., Smith, D.M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G.A., Msadek, R., Mueller, W.A., Taylor, K.E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., Eade, R., 2016. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. Geoscientific Model Development 9, 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Carberry, P.S., Hochman, Z., McCown, R.L., Dalgliesh, N.P., Foale, M.A., Poulton, P.L., Hargreaves, J.N.G., Hargreaves, D.M.G., Cawthray, S., Hillcoat, N., Robertson, M.J., 2002. The FARMSCAPE approach to decision support: farmers', advisers', researchers' monitoring, simulation, communication and performance evaluation. Agricultural Systems 74, 141–177. https://doi.org/10.1016/S0308-521X(02)00025-2

Chikamoto, Y., Kimoto, M., Ishii, M., Mochizuki, T., Sakamoto, T.T., Tatebe, H., Komuro, Y., Watanabe, M., Nozawa, T., Shiogama, H., Mori, M., Yasunaka, S., Imada, Y., 2013. An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC. Climate Dynamics 40, 1201–1222. https://doi.org/10.1007/s00382-012-1351-y

Choudhury, D., Sharma, A., Sen Gupta, A., Mehrotra, R., Sivakumar, B., 2016. Sampling biases in CMIP5 decadal forecasts. Journal of Geophysical Research: Atmospheres 121, 3435–3445. https://doi.org/10.1002/2016JD024804

Climate Data, 2020. Liberia climate: Average Temperature, weather by month, Liberia weather averages.

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27, 1547–1578. https://doi.org/10.1002/joc.1556

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

Grotch, S.L., MacCracken, M.C., 1991. The Use of General Circulation Models to Predict Regional Climatic Change. Journal of Climate 4, 286–303. https://doi.org/10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2

Gupta, A. Sen, Jourdain, N.C., Brown, J.N., Monselesan, D., 2013. Climate Drift in the CMIP5 Models*. Journal of Climate 26, 8597–8615. https://doi.org/10.1175/JCLI-D-12-00521.1

Gupta, A. Sen, Muir, L.C., Brown, J.N., Phipps, S.J., Durack, P.J., Monselesan, D., Wijffels, S.E., 2012. Climate drift in the CMIP3 models. Journal of Climate 25, 4621–4640. https://doi.org/10.1175/JCLI-D-11-00312.1

Hansen, J.W., Mason, S.J., Sun, L., Tall, A., 2011. Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture 47, 205–240. https://doi.org/10.1017/S0014479710000876

Hawkins, E., Dong, B., Robson, J., Sutton, R., Smith, D., 2014. The interpretation and use of biases in decadal climate predictions. Journal of Climate 27, 2931–2947. https://doi.org/10.1175/JCLI-D-13-00473.1

Ho, C.K., Stephenson, D.B., Collins, M., Ferro, C.A.T., Brown, S.J., 2012. Calibration Strategies: A Source of Additional Uncertainty in Climate Change Projections. Bulletin of the American Meteorological Society 93, 21–26. https://doi.org/10.1175/2011BAMS3110.1

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., 2021. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale. Journal I, 285.

ICPO, 2011. Data and bias correction for decadal climate predictions. CLIVAR Publication

Series No. 150, 6 pp.

Islam, S.A., Bari, M.A., F. Anwar, A.H.M., 2014. Hydrologic impact of climate change on Murray-Hotham catchment of Western Australia: A projection of rainfall-runoff for future water resources planning. Hydrology and Earth System Sciences 18, 3591–3614. https://doi.org/10.5194/hess-18-3591-2014

Jain, S., Salunke, P., Mishra, S.K., Sahany, S., 2019. Performance of CMIP5 models in the simulation of Indian summer monsoon. Theoretical and Applied Climatology 137, 1429–1447. https://doi.org/10.1007/s00704-018-2674-3

Jones, J.W., Hansen, J.W., Royce, F.S., Messina, C.D., 2000. Potential benefits of climate forecasting to agriculture. Agriculture, Ecosystems & Environment 82, 169–184. https://doi.org/10.1016/S0167-8809(00)00225-5

Jones, P.W., 1999. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Kharin, V. V., Boer, G.J., Merryfield, W.J., Scinocca, J.F., Lee, W.-S., 2012. Statistical adjustment of decadal predictions in a changing climate. Geophysical Research Letters 39. https://doi.org/10.1029/2012GL052647

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G.A., 2010. Challenges in Combining Projections from Multiple Climate Models. Journal of Climate 23, 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Lovino, M.A., Müller, O. V., Berbery, E.H., Müller, G. V., 2018. Evaluation of CMIP5 retrospective simulations of temperature and precipitation in northeastern Argentina. International Journal of Climatology 38, e1158–e1175. https://doi.org/10.1002/joc.5441

Masanganise, J., Chipindu, B., Mhizha, T., Mashonjowa, E., Basira, K., 2013. An evaluation of the performances of Global Climate Models (GCMs) for predicting temperature and rainfall in Zimbabwe 3, 1–11.

Meehl, G.A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G.J., Vecchi, G., Yeager,

S., 2014. Decadal Climate Prediction: An Update from the Trenches. Bulletin of the American Meteorological Society 95, 243–267. https://doi.org/10.1175/BAMS-D-12-00241.1

Meehl, G.A., Teng, H., 2012. Case studies for initialized decadal hindcasts and predictions for the Pacific region. Geophysical Research Letters 39. https://doi.org/10.1029/2012GL053423

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. Journal of Hydrology 519, 2932–2951. https://doi.org/10.1016/j.jhydrol.2014.07.053

Mehta, V.M., Knutson, C.L., Rosenberg, N.J., Olsen, J.R., Wall, N.A., Bernadt, T.K., Hayes, M.J., 2013. Decadal Climate Information Needs of Stakeholders for Decision Support in Water and Agriculture Production Sectors: A Case Study in the Missouri River Basin. Weather, Climate, and Society 5, 27–42. https://doi.org/10.1175/WCAS-D-11-00063.1

Miao, C., Su, L., Sun, Q., Duan, Q., 2016. A nonstationary bias-correction technique to remove bias in GCM simulations. Journal of Geophysical Research: Atmospheres 121, 5718–5735. https://doi.org/10.1002/2015JD024159

Mjelde, J.W., Penson, J.B., Nixon, C.J., 2000. Dynamic Aspects of the Impact of the Use of Perfect Climate Forecasts in the Corn Belt Region. Journal of Applied Meteorology 39, 67–79. https://doi.org/10.1175/1520-0450(2000)039<0067:DAOTIO>2.0.CO;2

Narapusetty, B., Stan, C., Kumar, A., 2014. Bias correction methods for decadal sea-surface temperature forecasts. Tellus, Series A: Dynamic Meteorology and Oceanography 66. https://doi.org/10.3402/tellusa.v66.23681

Paull, C.J., 2002. The value and benefits of using seasaonal climate forecasts in making business decisions: a review 42.

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. International Journal of Climatology 23, 887–901. https://doi.org/10.1002/joc.922

Shams, M.S., Faisal Anwar, A.H.M., Lamb, K.W., Bari, M., 2018. Relating ocean-atmospheric climate indices with Australian river streamflow. Journal of Hydrology 556, 294–309. https://doi.org/10.1016/j.jhydrol.2017.11.017

Skelly, W.C., Henderson-Sellers, A., 1996. Grid box or grid point: What type of data do GCMs deliver to climate impacts researchers? International Journal of Climatology 16, 1079–1086. https://doi.org/10.1002/(sici)1097-0088(199610)16:10<1079::aid-joc106>3.0.co;2-p

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

White, B., 2000. Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems, Atmospheric and Oceanographic Sciences Library. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-015-9351-9

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

Supplementary Information for chapter 5

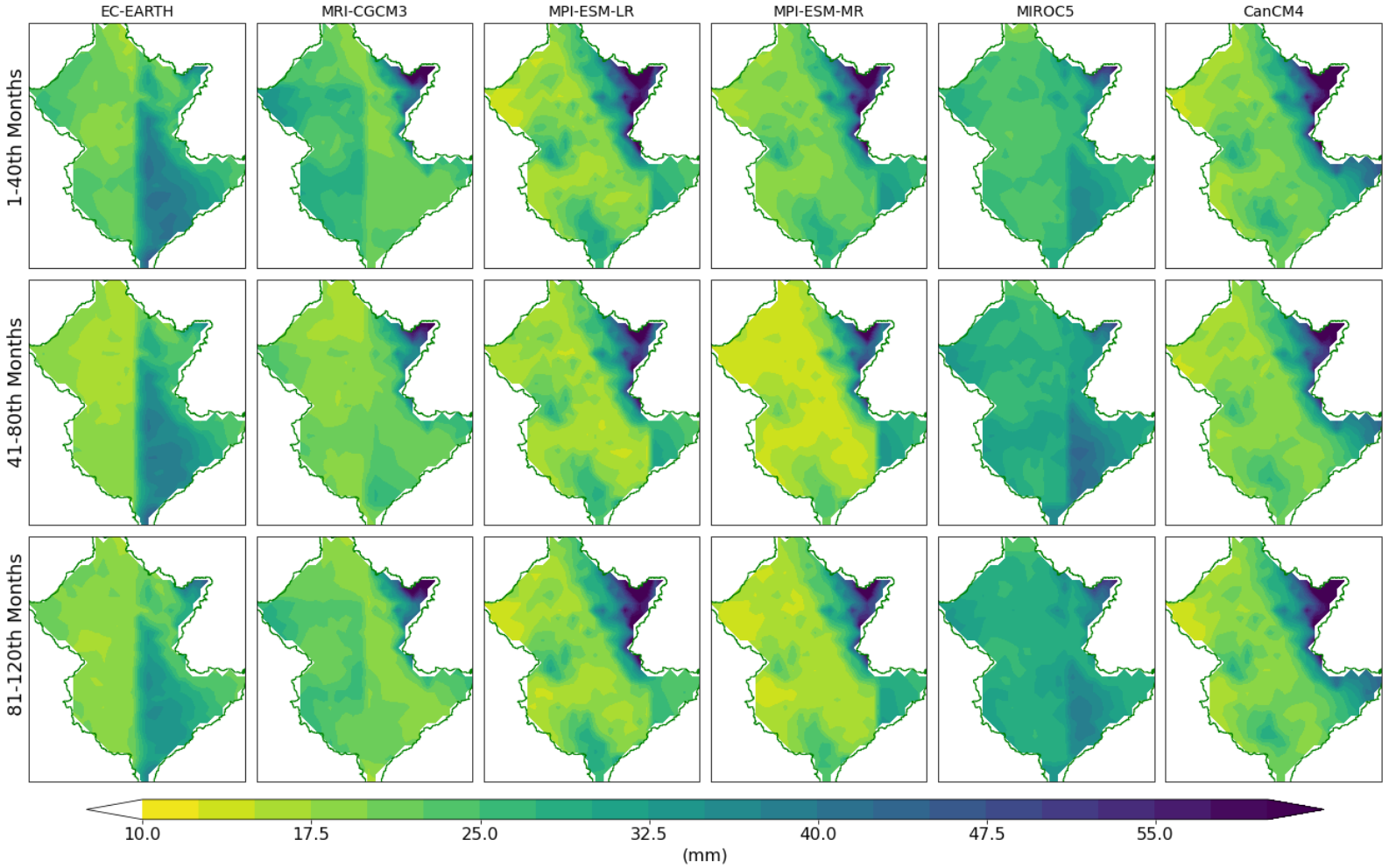


Fig. 5-S1 Spatial comparison of the drifts for monthly precipitation of all selected models except MIROC4h. Temporal mean of the 1st, 2nd and 3rd spell each of 40 months are presented in the 1st, 2nd and 3rd row respectively

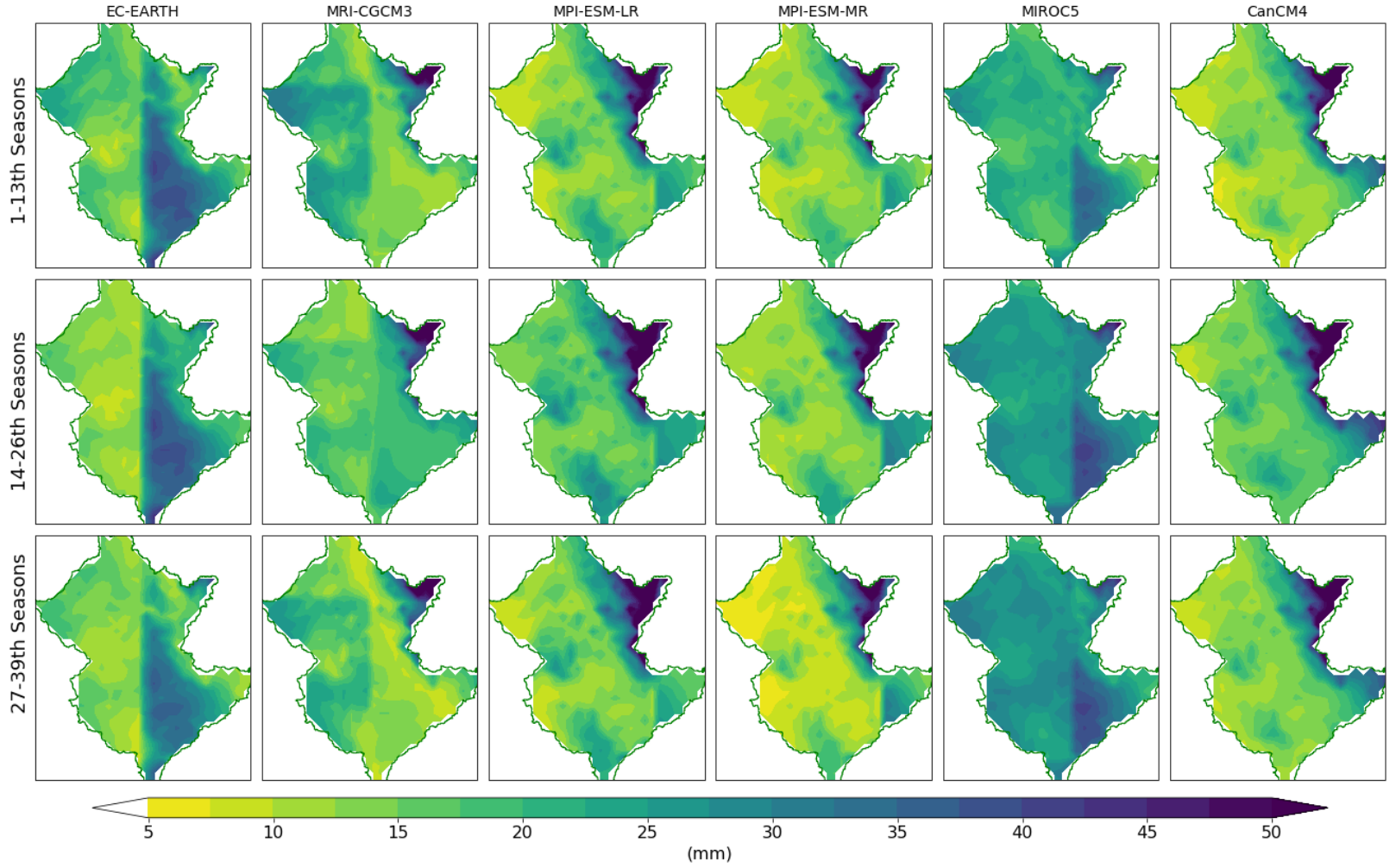


Fig. 5-S2 Spatial comparison of the drifts for seasonal mean precipitation of all selected models except MIROC4h. Temporal mean of the 1st, 2nd and 3rd spell each of 13 seasons are presented in the 1st, 2nd and 3rd row respectively

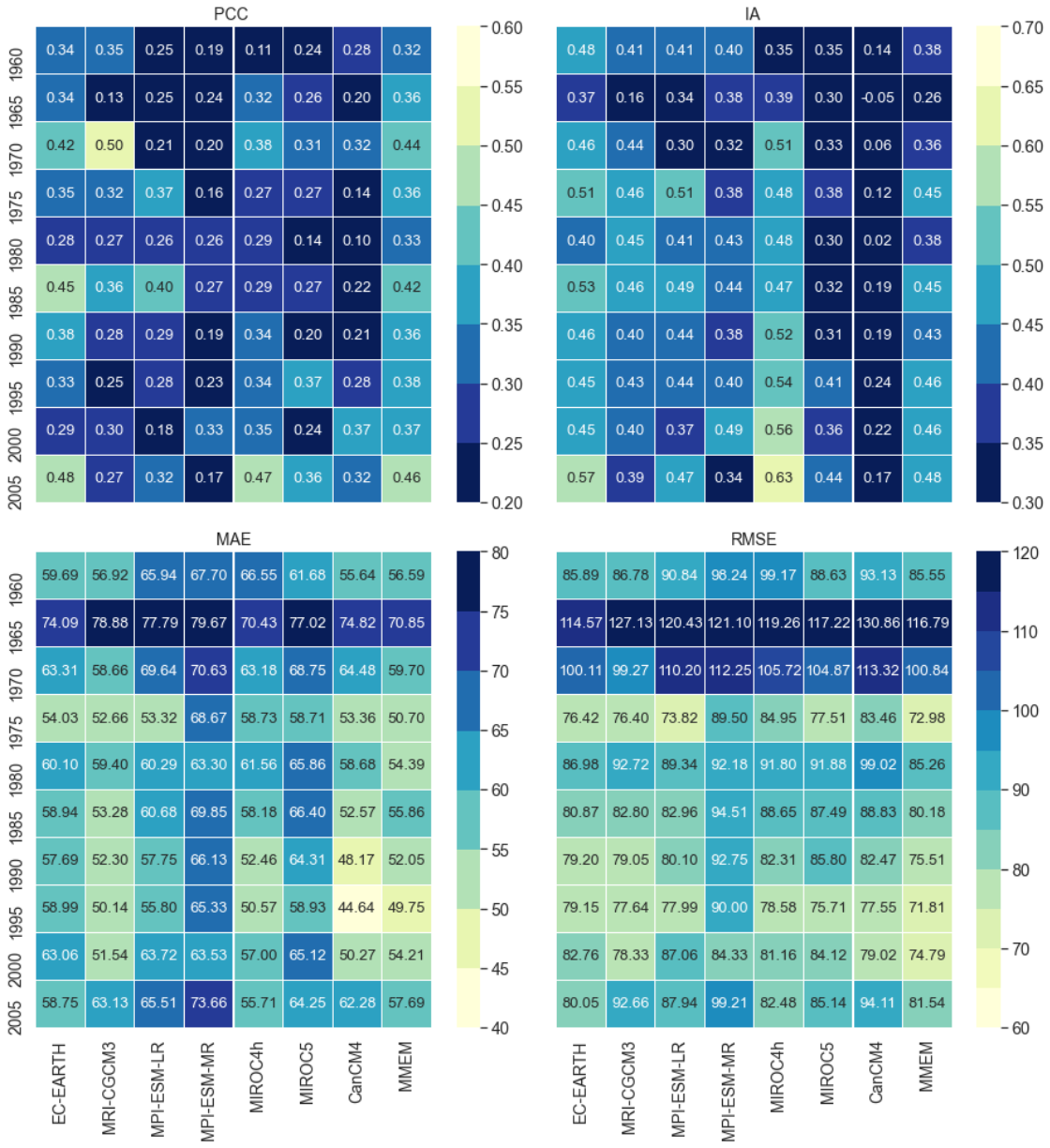


Fig. 5-S3 Time series skills; PCC, IA, MAE and RMSE for monthly precipitation at the selected grid point before drift correction

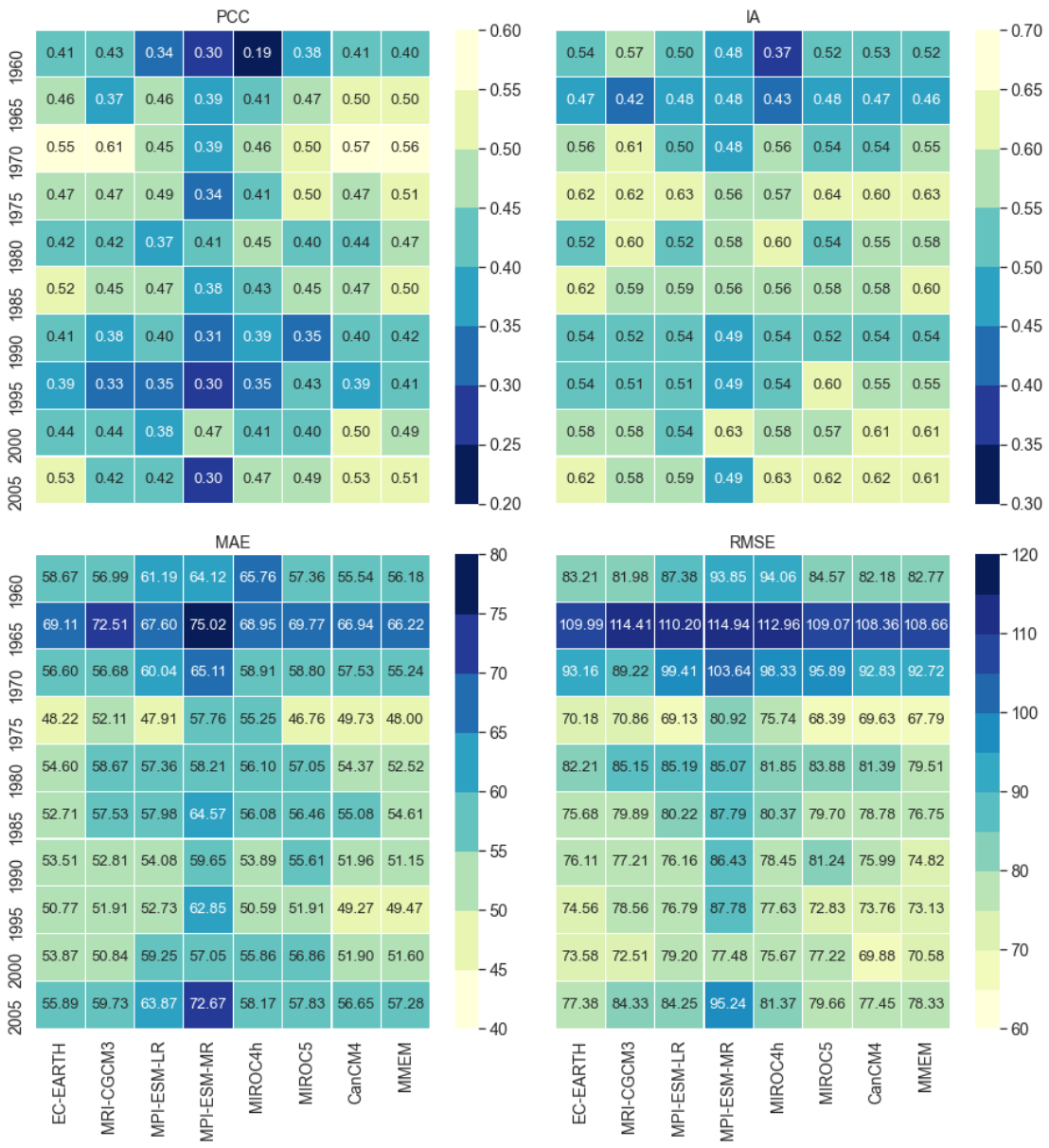


Fig. 5-S4 Time series skills; PCC, IA, MAE and RMSE for drift corrected monthly precipitation at the selected grid point

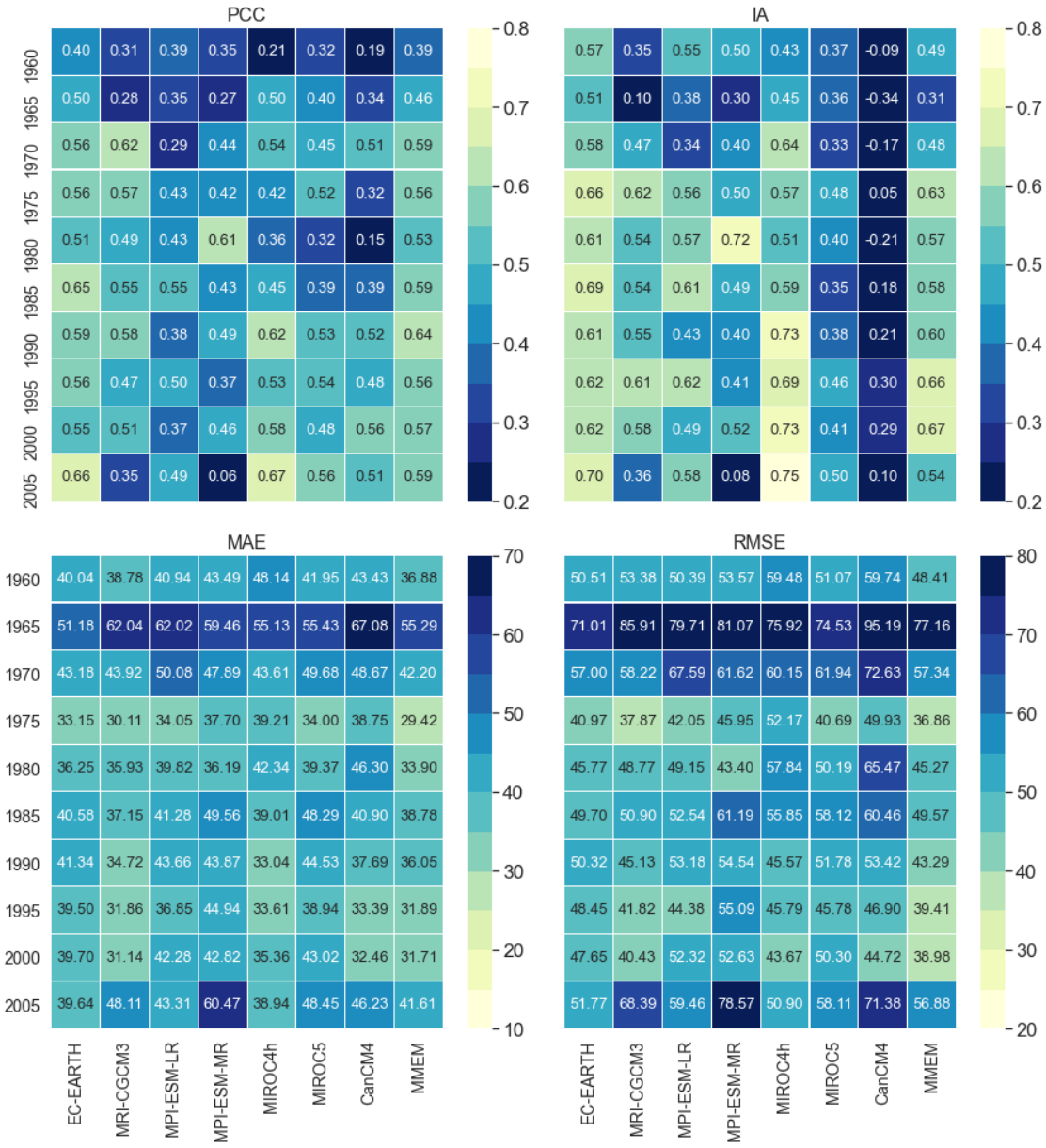


Fig. 5-S5 Time series skills; PCC, IA, MAE and RMSE of models' for seasonal mean precipitation before drift correction at the selected grid point

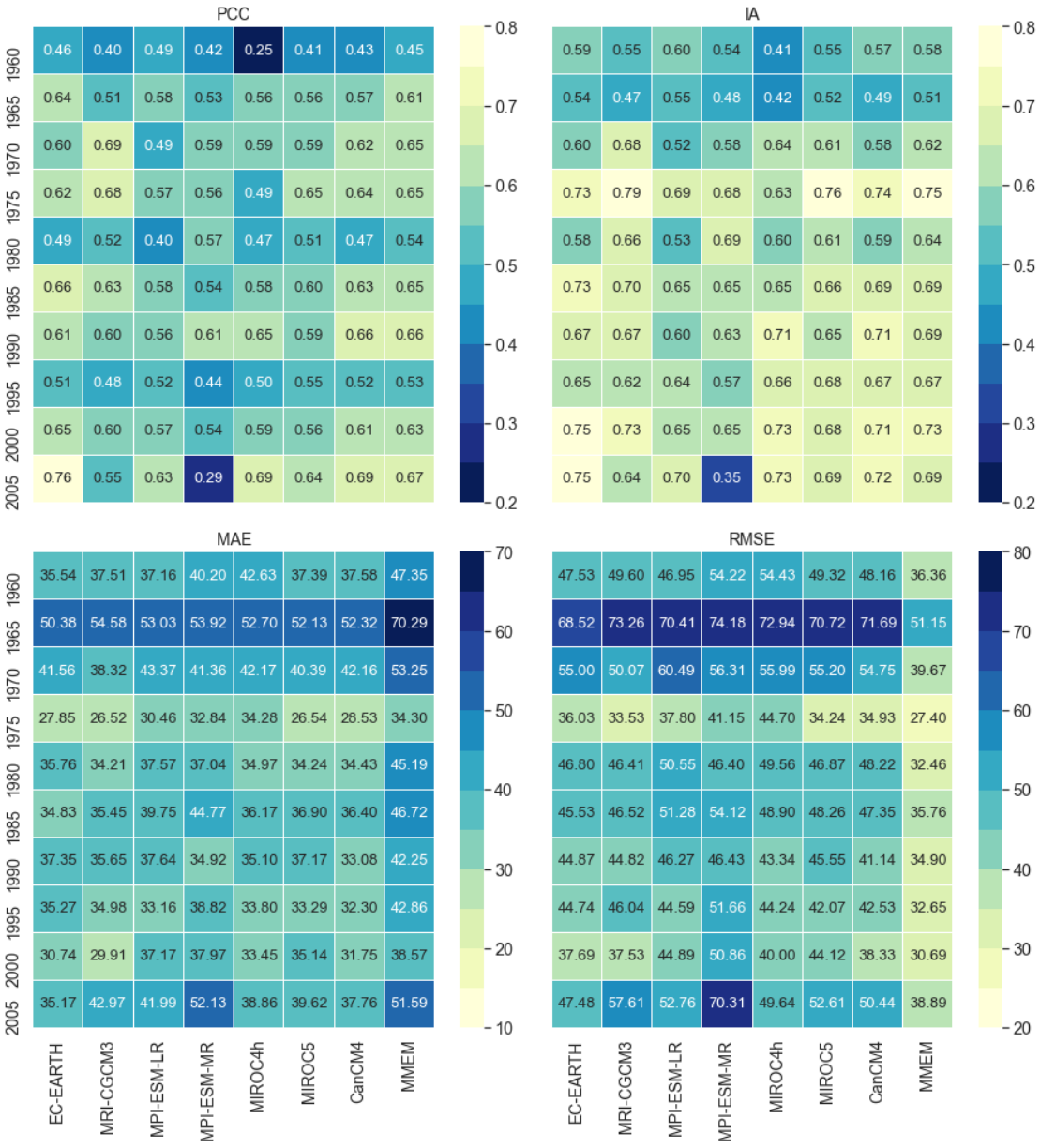


Fig. 5-S6 Time series skills; PCC, IA, MAE and RMSE of models for drift corrected seasonal mean precipitation at the selected grid point.

CHAPTER 6

INTERCOMPARISON OF DRIFT CORRECTION ALTERNATIVES FOR CMIP5 DECADAL PRECIPITATION

Abstract

Decadal experiments' output of the coupled model inter-comparison project phase-5 (CMIP5) contains significant model drift. For practical use of CMIP5 decadal climate variables, it is necessary to correct this model drift. In previous studies, drift correction of CMIP5 decadal data of temperature and temperature-based climate indices have been investigated, but there is no study that investigated the drift correction of decadal precipitation at a catchment scale. This study investigates the practical use of CMIP5 decadal precipitation data on the seasonal scale using different drift correction alternatives for the Brisbane catchment in Australia. For this, decadal monthly precipitation data from eight GCMs (EC-EARTH, MRI-CGCM3, MPI-ESM-LR, MPI-ESM-MR, MIROC4h, MIROC5, CMCC-CM, and CanCM4) were cut for the Australian region. By using the bi-linear interpolation, the GCM data were re-gridded to $0.05^0 \times 0.05^0$ spatial resolution matching with the observed gridded precipitation data collected from the Australian Bureau of Meteorology. Both model and observed data were subset for the Brisbane catchment and aggregated into seasonal means for Australian seasons. Four drift correction alternatives including one new modified method were employed for the selected GCM models and the models were categorized based on their performances using different skill scores. The results revealed that the performances of drift correction alternatives vary among different models and initialization years. Though some drift correction methods showed better performances than others for a given model it is still difficult to suggest the most suitable method for seasonal precipitation. Drift correction approaches for other time scales such as monthly precipitation, and their application for individual ensemble members may be investigated further to better understand the implications of alternative drift corrections for decadal forecasting.

This chapter has been published as: Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021. Intercomparison of drift correction alternatives for CMIP5 decadal precipitation. *International Journal of Climatology* 42, 1015-1037. https://doi.org/10.1002/joc.7287 (First published online: 16 July 2021). However, few textual changes have been made to address the examiners' comments.

Keywords: CMIP5, drift correction, decadal, precipitation, catchment

6.1 Introduction

In order to assess the near-term forecasting, the Coupled Model Intercomparison Project-Phase 5 (CMIP5) has experimented with the forecast of inter-annual to decadal-scale. The CMIP5 includes a set of decadal climate experiments for both hindcasts and predictions (Taylor et al., 2012). The experiment designs were finalized with three suites of experiment among which near-term (or decadal) experiments aimed to improve our understanding of the physical climate system and our capability to predict the climate parameters for longer timescales from season to decades (Taylor et al., 2012). There are two sets of core near-term integration; a set of 10-year hindcasts initialized from observed climate states near the years 1960, 1965, and every five years to 2005, and the other is 30-year hindcasts initialized in 1960, 1980 and 2005. From these two sets of integrations, the decadal experiments explored the potential of decadal predictability of the climate parameters (precipitation, temperature, humidity, wind speed, and others) on 10 to30-year time scales (Meehl et al., 2009) through General Circulation Models (GCMs). Climate models like GCMs are the imperfect replica of the real-world phenomena and contain systematic biases (Randall et al., 2007) which require post-processing for further uses (Islam et al., 2011, 2014; Maurer and Hidalgo, 2008; Mehrotra and Sharma, 2010). This also includes the uncertainty involving the teleconnection between climate indices and catchment hydrology (Shams et al., 2018). Depending on the time scales and type of variables, the magnitude of biases may be different. It has been reported by Sun et al. (2007) that GCMs tend to overestimate the number of wet days and underestimate the more intense events. This was also confirmed by Stephens et al. (2010), who found model precipitation is approximately double the observed value. In order to assess the climate change impacts on water resources, it is necessary to correct the model biases (Liang et al., 2008).

Currently, there is no standard approach for bias (or drift) correction (Taylor et al., 2012). The bias correction method may be different depending on locations, climate variables, time scales, and the application field of the data. (Chen et al., 2013; Gangstø et al., 2013; Kruschke et al., 2016). Some bias correction techniques use transformation functions (Ines and Hansen, 2006) such as downscaling, quantile mapping (QM), histogram equalizing, and rank matching (Bates et al., 1998; Charles et al., 2004; Piani et al., 2010; Wood et al., 2004). Others use stochastic

and/or dynamic downscaling (Bates et al., 1998; Charles et al., 2004; Mehrotra and Sharma, 2006; Wilby et al., 1998). Maraun, (2016) has presented a critical review on commonly used bias correction methods for simulating climate change such as the Delta change approach, simple mean bias correction, and QM. The QM technique developed by Wood et al., (2004) and also used by other researchers (Ines and Hansen, 2006; Maurer and Hidalgo, 2008; Sharma et al., 2007) demonstrates an improvement of bias correction of precipitation on daily, monthly, and annual scale. However, traditional QM assumes model biases are stationary over time which is not supported for a longer time scale (Chen et al., 2015) and thus it may result in poorer performance of bias-corrected data. For instance, Ines and Hansen (2006) applied QM to daily precipitation for predicting wheat yields in Africa and found this method could not change the length of wet and dry spells. On the contrary, after using the traditional QM on daily precipitation, Maurer and Pierce, (2014) suggested that QM improves the correspondence with observed changes but does not represent a future, especially where natural variability is much smaller than that due to external forcing. They also suggested that the influence of QM on seasonal precipitation trends does not systematically degrade projected differences at least for the next several decades. The stationarity limitation of traditional QM has been overcome by an updated nonstationary cumulative-distribution-function-matching (CNCDFm) technique developed by Miao et al., (2016), equidistance quantile-matching method developed by Li et al., (2010), a new quantile-quantile calibration method developed by Amengual et al., (2012). Cardoso Pereira et al., (2020) used CNCDFm technique for daily precipitation and Viceto et al., (2019) used the new quantile-quantile calibration method (Amengual et al., 2012) for daily maximum and minimum temperature and both of these research reported good performances. Piani et al.(2010) and Hempel et al., (2013) proposed bias correction for daily precipitation where the number of dry and wet days are considered while Mehrotra and Sharma, (2016, 2015) and Johnson and Sharma, (2012) proposed the method for a wide range of time scales. But till now, there hasn't been a study on the bias correction of seasonal precipitation. Simply, the bias in climate models is the difference between the model outputs and the observed values but there are systematic errors in climate models arising from different factors including limited spatial resolution, simplified physics, and thermodynamic processes, numerical schemes, or incomplete understanding of climate systems. Every climate model has an equilibrium climatology that differs from the real-world phenomena. Models start from an observed climate

state at the beginning of the simulation and tend to converge to their equilibrium state thus resulting in systematic time-varying biases also referred to as "drift"(Choudhury et al., 2017).

In decadal predictions, the drift is more pronounced than the longer-term projections (Taylor et al., 2012). The impact of drift may be different depending on the initial conditions, variables, regions, and time scales (Fučkar et al., 2014; Taylor et al., 2012). For the decadal climate prediction, some drift correction method has been recommended for CMIP5 GCMs outputs by Boer et al., (2016) and Taylor et al., (2012). But the appropriate choice of the drift correction methods depends on the temporal and spatial scale of the variables and finally, the applications of drift corrected variables and their acceptance. In recent years, the drift correction of temperature and temperature-based indices has been paid more attention than the precipitation for CMIP5 decadal experiments (Choudhury et al., 2017, 2016, 2015; Kim et al., 2012). It may be due to the additional complexity of atmospheric precipitation than that of temperature (Doblas-Reyes et al., 2011) or better performance of climate models in simulating temperature as opposed to precipitation (Kumar et al., 2013; Lovino et al., 2018; Masanganise et al., 2013; Mehrotra et al., 2014). In Australia, recently Mehrotra et al., (2014) used the CMIP5 decadal hindcast precipitation ($0.5° \times 0.5° \approx 55km \times 55km$) to assess GCMs skill using a drift correction, suggested by International Climate and Ocean: Variability, Predictability and Change (CLIVAR) Project Office (ICPO), (CLIVAR, 2011) in different hydrological regions and found greater skill for temperature and geopotential height than for precipitation. Notably, rainfall shows more temporal and spatial variability than temperature, therefore, climate model outputs (especially rainfall) at higher resolution (finer) are more useful than those at the coarser resolutions are. For this, CMIP5 decadal experiments' precipitation data might be a good opportunity to forecast precipitation for longer time scales (seasons to a decade) in finer resolution (Boer et al., 2016; Taylor et al., 2012). That is why a suitable drift correction method for a longer lead-time precipitation forecast has become essential to investigate.

To date, there is no study investigating the drift correction alternatives for the CMIP5 decadal precipitation at a finer spatial resolution for an Australian catchment. Therefore, the objective of this study is to assess the drift correction alternatives for seasonal mean precipitation of CMIP5 decadal data in a finer spatial resolution of $0.05^0 \times 0.05^0$ ($\approx 5km \times 5km$) for an Australian catchment (Brisbane catchment, Queensland, Australia). Four drift correction methods, including a new proposed modified method, are investigated in this study and their

performances are evaluated using different skill tests. The alternative drift correction methods are also used to investigate the performance of different GCMs through these skill scores.

6.2 Data collection and processing

6.2.1 Data collection

Monthly hindcasts precipitation data of decadal time scale from eight (EC-EARTH, MIROC4h, MRI-CGCM3, MPI-ESM-LR, MPI-ESM-MR, MIROC5, CanCM4, and CMCC-CM) out of ten available GCMs were downloaded from CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). For the initialization period 1960-2005 data for 10-year periods, initialized every five years are selected (few models have historical run until 2010). The other two models, HadCM3 (spatial resolution 3.75° x 2.5°) and IPSL-CM5A-LR (spatial resolution 3.75° x 1.89°) were not considered in this study because of their relatively coarser spatial resolution and different calendar system (HadCM3). The models selected in this study have spatial grids with resolutions smaller than 2° except CanCM4, which has a resolution smaller than 3° (see Table 1). However, the number of ensembles for selected models varied from 3 to 18, and all available ensembles of each initialization year are used in this study. The name of models, their group number, spatial resolutions, and the available historical run including ensemble numbers of individual models are given in Table 6-1.

Table 6-1 Selected GCMs used in this study

Modelling Centre (or Group)	Model (Resolutions °lon × °lat))	Initialization Year (1960-2005)									
		60	65	70	75	80	85	90	95	00	05
		Number of ensembles									
EC-EARTH Consortium	EC-EARTH (1.125 X 1.1215)	14	14	14	14	14	14	14	14	10	18
Meteorological Research Institute	MRI-CGCM3* (1.125 X 1.1215)	06	08	09	09	06	09	09	09	09	06
Max Planck Institute for Meteorology	MPI-ESM-LR* (1.875 X 1.865)	10	10	10	10	10	10	10	10	10	10
	MPI-ESM-MR* (1.875 X 1.865)	03	03	03	03	03	03	03	03	03	03
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC4h (0.5625 X 0.5616)	03	03	03	05	05	05	05	05	05	05
	MIROC5* (1.4062 X 1.4007)	06	06	06	06	04	06	06	06	06	06
Canadian Centre for Climate Modelling and Analysis	CanCM4* (2.8125 X 2.7905)	20	20	20	20	20	20	20	20	20	20
Centro Euro-Mediterraneo per I Cambiamenti Climatici	CMCC-CM 0.75 X 0.748	03	03	03	03	03	03	03	03	03	03

(* indicates model has historical run until the initialization year 2010)

The observed monthly gridded rainfall of 0.05° × 0.05° (5km × 5km) spatial resolution for entire Australia are collected from the Australian Bureau of Meteorology (BoM). The BoM has produced the gridded data using the Australian Water Resources Assessment Landscape model (AWRA-L V5) (Frost et al., 2016).

6.2.2 Data processing

The GCMs' resolution of approximately 100-250km is inadequate for regional studies because of the lack of regional information at catchment scales (Fowler et al., 2007; Grotch and MacCracken, 1991; Salathé, 2003). The use of the regional climate models (RCMs) for downscaling large-scale climate information from GCMs to local scales, is prevalent nowadays, albeit a computationally intensive approach. Conversely, this study uses spatial interpolation to re-grid the GCM data onto 0.05°× 0.05° resolution from the spatial resolution of respective models.

First, the GCM outputs were subset for the Australian region and then bi-linearly interpolated onto 0.05°×0.05° spatial resolution thus matching with the grid used in observed data. Bi-linear

interpolation was used here because of its suitability for monthly precipitation, as there was no zero precipitation values in the interpolant dataset, and it has been routinely used in similar studies (Jain et al., 2019; Kamworapan and Surussavadee, 2019; Liang et al., 2008). Secondly, all model calendars were transformed to Proleptic_Gregorian and the precipitation unit were converted to the millimetre. Afterward, the ensemble mean of different GCM ensembles for every initialization year was computed and aggregated to the seasonal mean for the Australian seasons. In Australia, four seasons; summer (December-January-February, (DJF)), autumn (March-April-May, (MAM)), winter (June-July-August, (JJA)), and spring (September-October-November (SON)) contrary to the Northern Hemisphere. Each dataset is for ten years, starting from January (1st year) and ending at December (10th Year). For a continuous seasonal dataset, the first two months (January and February) at the start of a 10-year period and the last one-month (December) at the end of the said period were discarded. Finally, the processed data were subset for the selected catchment. In this study, Brisbane catchment in Queensland, Australia was selected for investigating the seasonal precipitation on the decadal time scale (Fig. 6-1). This catchment was selected because of its tropical climate nature and having low to moderate yearly rainfall variability.

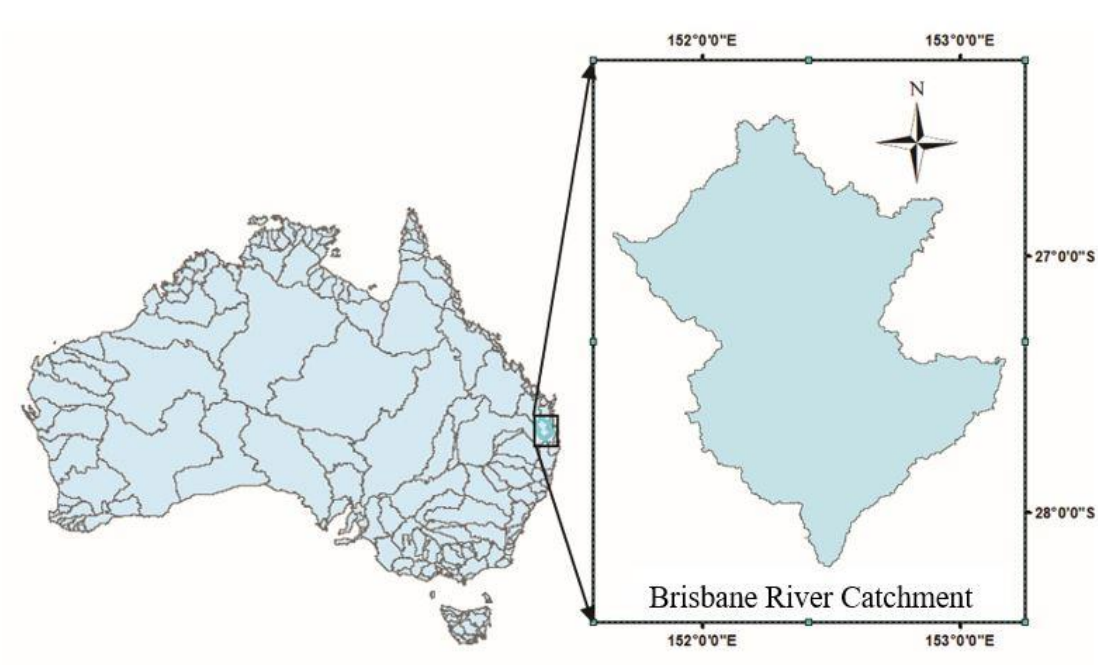


Fig. 6-1 Map of the Australian catchments with the Brisbane catchment in the inset

6.3 Methods

6.3.1 Drift correction

In this study, the first three drift correction methods are employed, namely Nested Bias Correction (NBC), Standardization and re-scaling, and Relative drift (linear scaling). Standardization and linear scaling methods are commonly used in statistical drift correction approaches (Graham et al., 2007; Hawkins et al., 2013; Sperna Weiland et al., 2010). Additionally, a modified method combining standardization and delta method is also suggested in this study.

Nested Bias Correction (NBC):

The NBC method was proposed by Johnson and Sharma, (2012) and works well to correct means, standard deviations, and lag-1 autocorrelations when the biases/drift in GCMs are not too large. In the first step, the model values were standardized as

$$p_{it} = \frac{P_{i,t}^{m} - \mu_t^m}{\sigma_t^m} \qquad (6.1)$$

Where, P, p = precipitation, μ = mean, σ = standard deviation, the subscript, t = lead time in year, i = individual season of time t, the superscript m represents model, and o represents observed (in the following equations). Hereafter the same notations will be used.

Then, the lag-1 autocorrelations (r_i^m) present in the standardized time series of the model data, were replaced with the lag-1 autocorrelations (r_i^o) of the observed data as per equation 6.2.

$$\tilde{p}_{i,t}^{m} = r_i^o * \tilde{p}_{i-1,t}^{m} + \sqrt{1-(r_i^o)^2} * \left(\frac{p_{i,t}^{m} - r_i^m * p_{i-1,t}^{m}}{\sqrt{1-(r_i^m)^2}} \right) \qquad (6.2)$$

Afterward, it was rescaled by the observed mean and standard deviation to obtain a nested time series on a seasonal scale.

$$\tilde{P}_{i,t}^{m} = \tilde{p}_{i,t}^{m} * \sigma_t^o + \mu_t^o \qquad (6.3)$$

Note, Johnson and Sharma, (2012) used monthly precipitation values in Eq. 6.1-6.3, but in this study seasonal mean precipitation values are used instead. In the second step, the nested

seasonal precipitation $\left(\tilde{P}_{i,t}^{m}\right)$ were aggregated to annual scale (P_t^m) followed by standardization and lag-1 removal using Eq. 6.4-6.6.

For standardization, $p_t^m = \frac{P_t^m - \mu^m}{\sigma^m}$ and for lag-1 autocorrelation, the following equation was used.

$$\tilde{p}_t^m = r^o * \tilde{p}_{t-1}^m + \sqrt{1-(r^o)^2}\left(\frac{p_t^m - r^m * p_{t-1}^m}{\sqrt{1-(r^m)^2}}\right) \qquad (6.4)$$

Afterward, the nested annual time series precipitation $(\tilde{p}_t^m)$ was rescaled by observed mean and standard deviations.

$$\tilde{P}_t^m = \tilde{p}_t^m * \sigma^o + \mu^o \qquad (6.5)$$

Finally, the drift corrected seasonal precipitation was obtained using equation (6.6).

$$P_{cor,i,t} = \tilde{P}_{i,t}^m * \left(\frac{\tilde{P}_t^m}{P_t^m}\right) \qquad (6.6)$$

Standardization and Re-scaling approach (STD)

Ho (2010) suggests that if two datasets are from the same (normal) distribution with the same shape but different mean and variances then the following equations can map a point between the two distributions which effectively equates the normalized difference between the points.

$$\frac{X_1 - M_1}{\sigma_1} = \frac{X_2 - M_2}{\sigma_2} \qquad (6.7)$$

Re-arranging this gives,

$$X_2 = M_2 + \frac{\sigma_2}{\sigma_1}(X_1 - M_1) \qquad (6.8)$$

Where, X_1, X_2 are two points, M_1, M_2 are the mean and σ_1, σ_2 are the standard deviations of two data sets respectively. Here, Eq. 8 is used for the drift correction, where observed and modeled data are the two different datasets. The datasets were transformed to log-normal distribution to make them form the same distribution, where the normality (log-normal) of both datasets were confirmed by the Shapiro-Wilk test, D'Agostino's K^2 test, and Anderson-Darling test. Re-arranging Eq. 8 gives

$$P_{cor,it} = \mu_t^o + \frac{\sigma_t^o}{\sigma_t^m}\left(P_{i,t}^m - \mu_t^m\right) \qquad (6.9)$$

Relative drift correction (RDT) or linear scaling

The simplest way of drift correction is to add (for temperature) or multiply (for precipitation) the change of mean to the observed data of base line. To avoid encountering negative values in precipitation, multiplicative approach is used. It is widely used for preparing climate change scenario at local scale and is defined as -

$$P_{cor,it} = P_{it} * \frac{\mu_t^o}{\mu_t^m} \qquad (6.10)$$

Modified Method (MDM)

It was noticed that the STD method overestimated the peaks of maximum precipitation with higher ranges of extremes as well as number of outliers while RDT methods fluctuated around the mean thus giving a lower range of spread with no or minimal outliers. To minimize the range of over estimation (both upper and lower) and number of outliers, the values from these two methods are averaged and used in this study as MDM. The average equation can be written as-

$$P_{cor,it} = 0.5 * \left(P_{it}^m * \frac{\mu_t^o}{\mu_t^m} + \frac{P_{i,t}^m - \mu_t^m}{\sigma_t^m} * \sigma_t^o + \mu_t^o\right) \qquad (6.11)$$

Note, the mean and standard deviations in the equations for all drift correction methods are calculated for individual years. It is also worth noting that all the drift correction methods are applied at all grids in the catchment and grid cells are assumed spatially independent. To compare the performance of different drift correction methods, the skill tests are performed at few randomly selected grids distributed over the catchment.

6.3.2 Skill assessment

To evaluate the different drift correction methods, seven skill tests are performed. These include, correlation coefficient, anomaly correlation coefficient, index of agreement, root mean squared error, mean absolute error and fractional skill scores.

Correlation Coefficient (CC)

Correlation coefficient is a good measure of linear association or phase error. Statistically it means, how well the model (drift corrected) values correspond to the observed values, whereas

visually it illustrates how close the scatter plot points are to a diagonal line. The value of CC ranges from -1 (no correlation) to perfect correlation, 1 (Rodwell et al., 2010).

$$CC = \frac{\sum(F-\overline{F})(O-\overline{O})}{\sqrt{\sum(F-\overline{F})^2}\sqrt{\sum(O-\overline{O})^2}} \qquad (6.12)$$

Here, F and O represent the seasonal model (drift corrected) and observed values respectively. Whereas $\overline{F}$, $\overline{O}$ are the mean of the respective season values for each individual year. Note, the mean value is calculated for each individual year, and the seasonal anomaly for each individual year is the difference between the modelled/ observed values and its mean value.

Anomaly Correlation Coefficient (ACC)

ACC shows how well the bias corrected anomalies correspond to the observed anomalies. It measures the correspondence (or phase difference) between the model (drift corrected) and observed values by subtracting the decadal mean. It is frequently used for the verification of output from the numerical weather prediction models. However, it is not sensitive to the bias i.e., a high anomaly correlation coefficient does not represent the accuracy of bias corrected values but the accuracy of the anomalies of the bias corrected values. Its value ranges from -1 (no match) to 1 for perfect anomaly matching. Wilks, (2011) reported centered and uncentered ACC. Centered ACC is computed according to Pearson correlation, which operates on the number of grid pairs of forecast and observations referring to the climatological mean of each grid. The uncentered ACC differs from centered ACC where map-mean anomalies are not subtracted. However, we used centered ACC in this study to calculate the temporal anomalies over the decadal time span.

$$ACC = \frac{\sum\{(F-C)-(\overline{F-C})\}\times\{(O-C)-(\overline{O-C})\}}{\sqrt{\sum(F-C)^2}\sqrt{\sum(O-C)^2}} \qquad (6.13)$$

Where *C* represents decadal mean of the observed (BoM) data. The ACC was calculated for each grid but the climatological mean was taken as the decadal mean (C) for each grid. Thus ACC may also be referred to spatio-temporal correlation if each grid is compared with corresponding years. However, ACC was calculated on a temporal scale only for each individual grids in this study.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)

MAE and RMSE are used to measure the average magnitude of the errors between model (drift corrected) and observed values. MAE is the average of the absolute values of the differences between forecasted and corresponding observed values and it is weighted equally in the average. The RMSE is a quadratic scoring rule that is squared before it is averaged and provides a relatively high weight to large errors. RMSE is useful when large errors are especially undesirable. The value of both RMSE and MAE ranges from 0 to ∞ where lower values indicate higher accuracy and vice versa.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|F_i - O_i| \qquad (6.14)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2} \qquad (6.15)$$

Index of agreement (IA)

To visualize the errors with respect to climatology, Wilmot, (1982) suggested an index of agreement that is directly related to the accuracy of drift-correction methods. Index of agreement can be calculated as follows:

$$IA = 1 - \frac{\sum_{i=1}^{n}(F_i - O_i)^2}{\sum_{i=1}^{n}(|F_i - \bar{O}| + |O_i - \bar{O}|)^2} \qquad (6.16)$$

The IA is bounded between 0 and 1 ($0 \leq IA \leq 1$), where, the closer the value is to 1 the more efficient is the drift correction. The observed mean and the difference between model (drift corrected) values and the observed mean are calculated for individual years in this study.

Fractional Skill Score (FSS)

Fractional Skill Score directly compares the model (drift corrected) and observed fractional coverage of the grid-box events (e.g., rain exceeding a certain threshold) for the entire catchment. FSS is a measure of how the spatial variability of the drift corrected values matches with the spatial variability of the observed values. FSS can be defined as:

$$FSS = 1 - \frac{\frac{1}{N}\sum_{N}\left(P_{f,m} - P_{f,o}\right)^2}{\frac{1}{N}\left[\sum_{N} P_{f,m}^2 + \sum_{N} P_{f,o}^2\right]} \qquad (6.17)$$

Where P_f indicates the calculated fraction, N indicates the number of years, and the subscript m and o indicate modelled and observed respectively. Fractions are calculated following Roberts and Lean (2008). In this study, the entire catchment is considered a whole unit, and the temporal averages (from every season) are taken instead of the spatial averages. Two threshold values are considered; ≥85 percentile for the wet seasons (DJF) and <15 percentile for the dry seasons (JJA). To calculate the fractions for individual seasons, the number of grids covered for a specified threshold value (for instance, 85 percentile of a specific wet season) are counted and then divided by the total number of grids within the Brisbane catchment. The differences between the fractions (the numerator of the equation) of models are calculated for individual wet/dry seasons. The FSS computed here will be a temporal average score for the entire catchment, as every dataset contains nine wet and ten dry seasons. FSS ranges from 0 to 1 (perfect match).

In addition to the FSS, we performed a spatial comparison of drift correction methods based on another fraction. It was computed as the number of grids covered by the models' simulated precipitation for a certain threshold (precipitation > 85 percentile and <=15 percentile of observed values for wet and dry events respectively) divided by the number of grids covered by the observed values for the same threshold.

In addition to the above skill tests, the distribution of both corrected and un-corrected precipitation were checked and visually compared using Box-Whisker plots.

6.4 Results and analyses

6.4.1 Assessment of models

Modelled seasonal precipitation values are compared with the corresponding observed precipitation and their skills are assessed. These comparisons are performed for all models; for all initialization years and the skills are assessed before and after the drift correction. As the initialization method for the same year is determined by each model group, one may assume problematic while comparing the models. However, this study mainly compares the drift correction methods for each model and their initialization years to find the suitable drift correction method for decadal GCM data. Comparing drift correction methods for the same model for the same year, in this case, the initialization condition would not be a problem. Model

performances are compared based on the quantitative performance metrics with the assumption that higher performance skills will present the lower models' drift and vice versa. Seven skill tests (CC, ACC, IA, FSSa85, FSSb15, RMSE, and MAE) are used and the skill test results are used to compare different models and different drift correction methods. Mehrotra et al., (2014) assessed CMIP5 multi-model decadal hindcasts using performance skill measures such as RMSE, BIAS, and correlation. These skill measures were evaluated over each grid within their study area. They used BIAS as the averaged absolute difference of observations and predictions and correlation as the Pearson correlation. The RMSE provides a simple, transparent quantitative measure of the difference between the models (corrected) and observed values. However, we used CC as the measure of the correlation between observed and forecasted values in the annual time scale while ACC measures the correlation of their anomalies for the decadal time scale.

The metrics CC, ACC, IA, RMSE, and MAE are used to assess skill at a single grid whereas FSSa85 and FSSb15 are used for the entire catchment. The observed dataset has 496 grids (5.0 km X 5.0 km) within the Brisbane catchment and the skill tests are performed for several randomly selected grids distributed over the entire catchment. Here, the results of the temporal skill test (CC, ACC, IA, RMSE, and MAE) and their comparison over time are presented for a single grid (latitude 27.5^{o}S and longitude 153.05^{o}E) which is closest to an automated weather station (AWS, located at latitude 27.48^{o}S and longitude 153.04^{o}E) operated by BoM. Although skill is generally estimated for a large regional area and for slow varying variables such as temperature, it is equally important to have skill assessment over a small area (i.e., for each region and at different spatial scales) for precipitation as it exhibits high temporal and spatial variability.

In general, the skill of all models improved with time with exception of initialization years 1980 and 1985, where the skill of all models dropped significantly. This decline was especially pronounced for the year 1980. During 1981-1990, for both 1980 and 1985**,** the lower performance is attributed to the higher than observed peak values in the modeled precipitation with a shifting (early peak) of seasons relative to the observed precipitation. However, the highest skill was noted for the initialization year 1990 for almost all the models. MPI-ESM-MR showed the lowest skill in the initialization year 2005 which persisted even after the drift correction. The reason behind lower skill in 1980 and highest skill 1990 is not well studied

here. However, Meehl et al. (2015) investigated the effect of volcanic eruption on the decadal hindcast skill of Pacific sea surface temperature and reported that Agung (erupted in 1963) and El Chichón (1982) did not significantly degrade the hindcast skill but Fuego (1974) and Pinatubo (1991) degraded the decadal hindcasts skill. However, Fuego was a smaller eruption with a lower amplitude climate forcing compared to Mount Pinatubo and this was the reason for lower degradation of skill in the mid-1970s whilst a significantly higher reduction in the mid-1990s (Meehl et al., 2015). In this study, models' higher skills in the 1990s and lower skills in the 1980s, seems to be irrelevant to the volcanic eruption but either the post-eruption sequences may be affected the observed precipitation or the quality and coverage of ocean observation to initialize the model was not as much realistic as other initialization years.

Models' performances without drift correction

Fig. 6-2 presents models' skills of all initialization years before drift corrections at the selected location where EC-EARTH, MIROC4h, MPI-ESM-LR, and MRI-CGCM3 showed the highest skill throughout the time span of the data and consequently, MPI-ESM-MR, MIROC5, CanCM4, and CMCC-CM had the lower skill. Overall, highest skill was noticed for the EC-EARTH whereas CMCC-CM has the lowest skill for all initialization years (see Fig. 6-2). EC-EARTH performed best on all skill tests except Fractional Skill Score above 85 percentile (FSSa85) and had the lowest values for RMSE and MAE. The lower value of FSSa85 indicates the poorer skill of EC-EARTH in reproducing the extreme events during the wet seasons, whereas MRI-CGCM3, MPI-ESM-LR, and MPI-ESM-MR had a comparatively higher skill. MIROC4h lags behind MRI-CGCM3 and MPI-ESM-LR for CC and FSSa85, however, perform considerably better on ACC, IA, FSSb15, RMSE, and MAE. In terms of accuracy and the reproducibility of anomalies, MIROC4h lags only EC-EARTH, and it surpasses both MRI-CGCM3 and MPI-ESM-LR on IA, ACC, FSSb15, RMSE, and MAE (not shown). When categorizing models on their average skill scores for the time period of initialization from 1960 to 2005, EC-EARTH, MIROC4h, and MRI-CGCM3 will be in the first category, and MPI-ESM-LR, MPI-ESM-MR, and MIROC5 will be in second, whereas, CanCM4 and CMCC-CM will be in the third category. Note, this categorization is based on the skill assessment of models before drift corrections. This will likely change after drift corrections.

This study also compared the differences between observed and models' hindcasts raw precipitation at the selected grid and the comparisons show that the models exhibit large

differences from the observed precipitation. Although models can reproduce the temporal variations but are unable to reproduce the extreme wet and dry values. Models' response to temporal variations was relatively higher for the initialization years 1990 and onward compared to the initialization years 1960 to 1980 (not shown) that could also explain the lower differences noted from 1990 onwards. For the time period 1981-1990, models showed shifting of seasonal peaks (early) and wider difference from observed precipitation in both initialization years 1980 and 1985 (not shown).

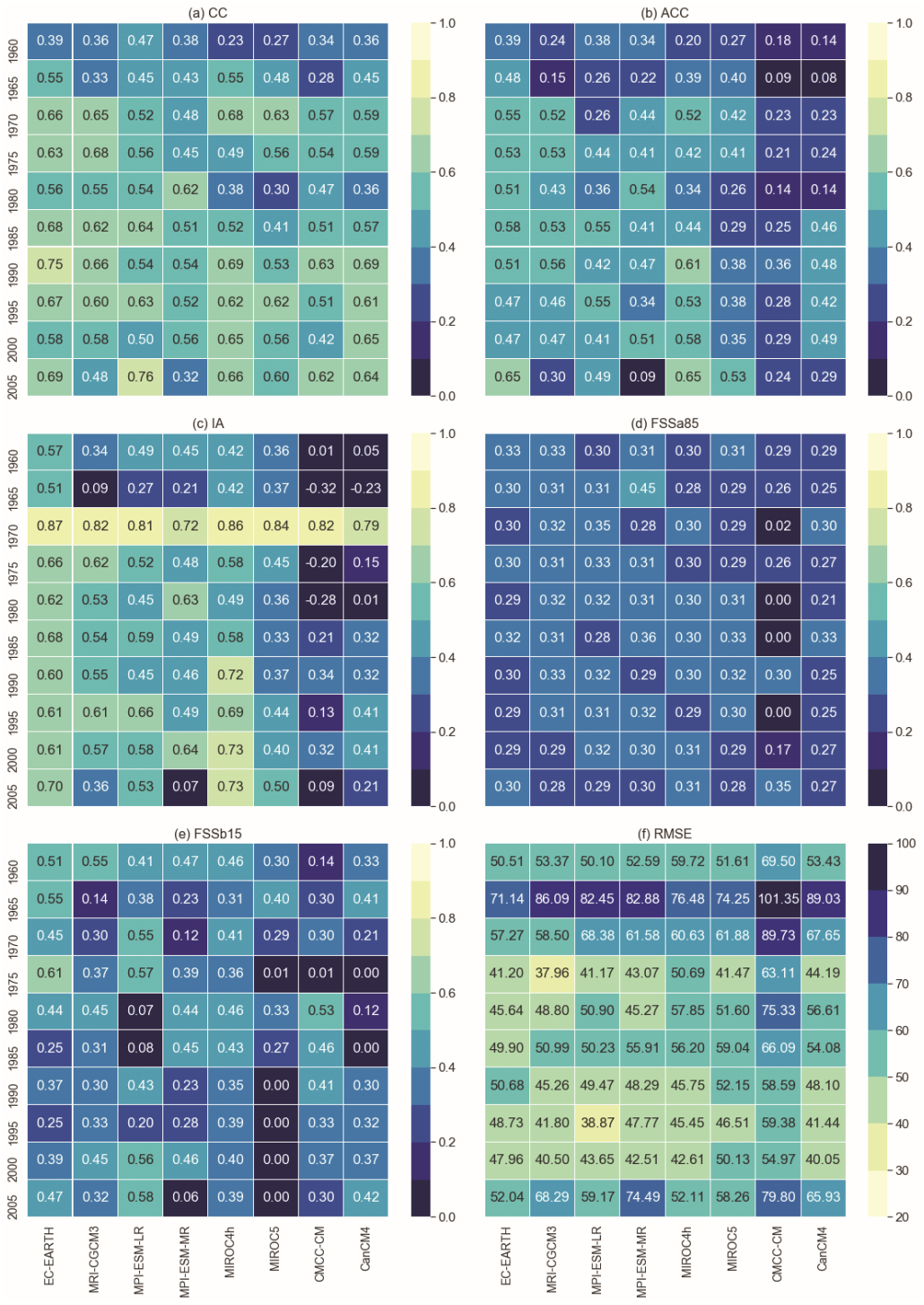


Fig. 6-2 Skill test results of different models prior to drift corrections. Fig. (a) presents Correlation Coefficient (CC), (b) Anomaly Correlation Coefficient (ACC), (c) Index of Agreement (IA), (d) Fractional Skill Score above 85 percentile of the observed precipitation (FSSa85), (e) Fractional Skill Score below 15 percentile of the observed values (FSSb15), and (f) Root Mean Squared Error (RMSE) of models raw precipitation

Models' performance after drift corrections

Models' performances after drift corrections are assessed based on the same skill tests used to assess the models' performance prior drift corrections. Fig. 6-3 presents the models' skills after drift correction (NBC method) at the selected location and the results indicate the improvements varied for different models and for the different drift correction methods. A two-sided t-test was performed and found more than 95% of all skills have changed significantly ($p< 0.05$) for all models. For all drift correction methods, EC-EARTH and MRI-CGCM3 show similar scores on all skill assessments except FSS where EC-EARTH performed better than MRI-CGCM3. From the skill tests results, it is noticed that EC-EARTH has the highest skill among the selected models. Before the drift correction, MIROC4h followed EC-Earth in all skill assessments, however, after the drift corrections, the skill of MRI-CGCM3 improved considerably, and was behind EC-EARTH, whereas MIROC4h followed MRI-CGCM3. Relative to other models, MIROC5 has the lowest skill, thus positioned behind CanCM4 and CMCC-CM. It doesn't mean that after drift correction the performance of MIROC5 declined, rather the improvements in other models are much more than in MIROC5. It is worth noting that the model skill was evaluated for all the drift correction methods, however, for the sake of brevity, only the NBC method is presented. Using the skill assessment (Fig. 6-2, 6-3), eight models are categorized into two types; Category-I (models which showed comparatively better performance and consistently were in the top four positions in all initialization years) and Category-II (comparatively lower skill than Category-I). EC-EARTH, MRI-CGCM3, MIROC4h, and MPI-ESM-LR are found in Category-I, while MPI-ESM-MR, CanCM4, CMCC-CM, and MIROC5 are in the Category-II. Hereafter, these two categories will be referred to in the following sections. The skill assessment results of individual models for different drift correction methods are shown in the following section.

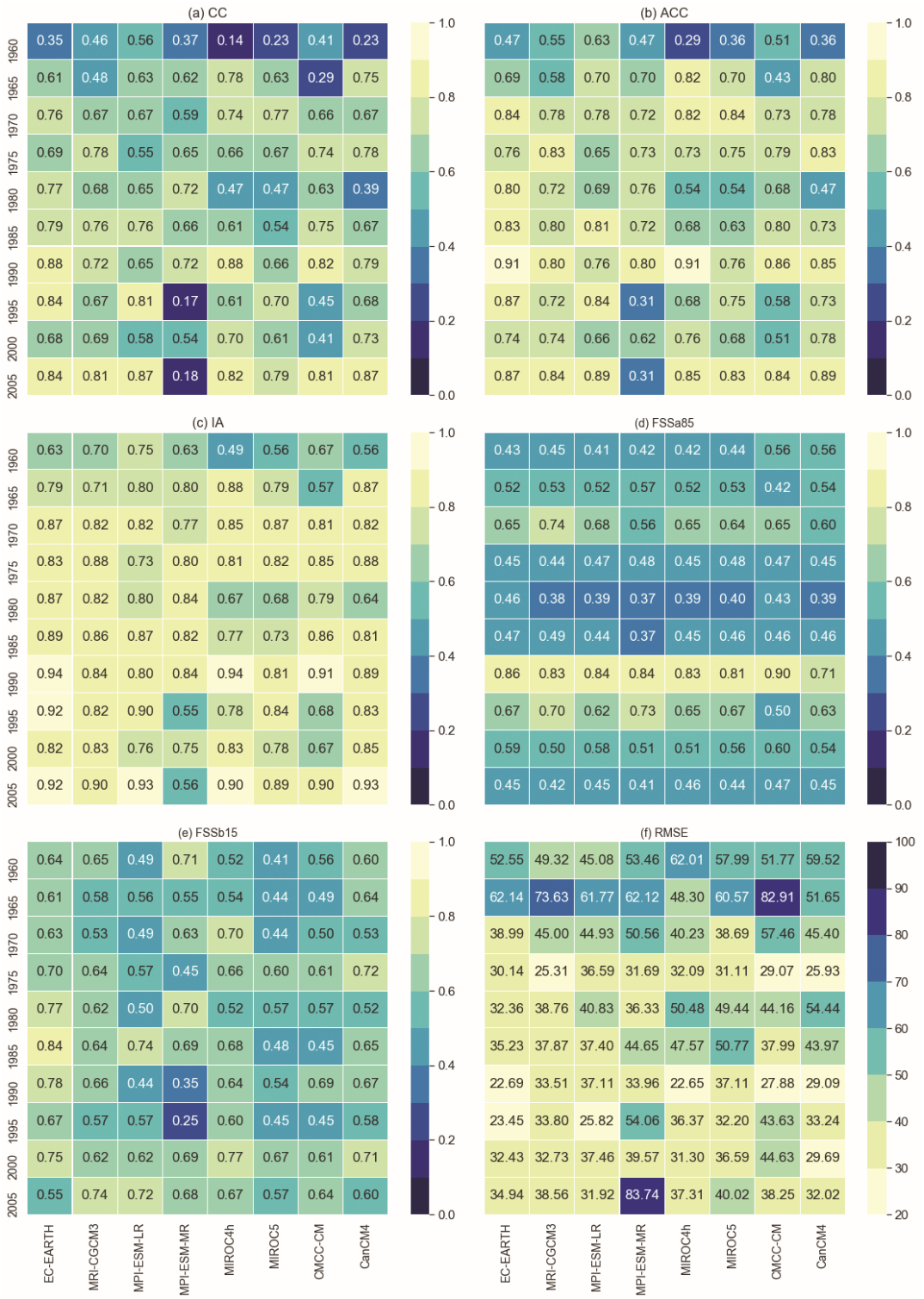


Fig. 6-3 Skill test results of different models after drift corrections (NBC method). Fig. (a) Presents Correlation Coefficient (CC), (b) Anomaly Correlation Coefficient (ACC), (c) Index of Agreement (IA), (d) Fractional Skill Score above 85 percentile (FSSa85) of observed values, (e) Fractional Skill Score below 15 percentile (FSSb15) of observed precipitation, and (f) Root Mean Squared Error

6.4.2 Performance of drift correction methods

Modelled precipitation are drift corrected using four methods: NBC, STD, RDT, and MDM are compared with the observed precipitation for all initialization years and it is noticed that all drift corrected precipitations can reproduce the observed precipitation anomalies, but differences exist for different drift correction methods and the model categories. For instance, the STD method overestimates the wet extremes in both model categories, but, the range and the number of overestimations were higher in the Category-II models than in the Category-I models. NBC and MDM showed fewer overestimations for the wet extremes, but, for dry extremes (dry seasons) MDM performed better than NBC. The skill comparison of drift correction approaches for the EC-EARTH model presented in Fig. 6-4 whilst skills for the CanCM4 model presented in Fig. 6-5 and spatial comparison of the drift correction approaches, for RMSE and initialization year 1990, is presented in Fig. 6-6.

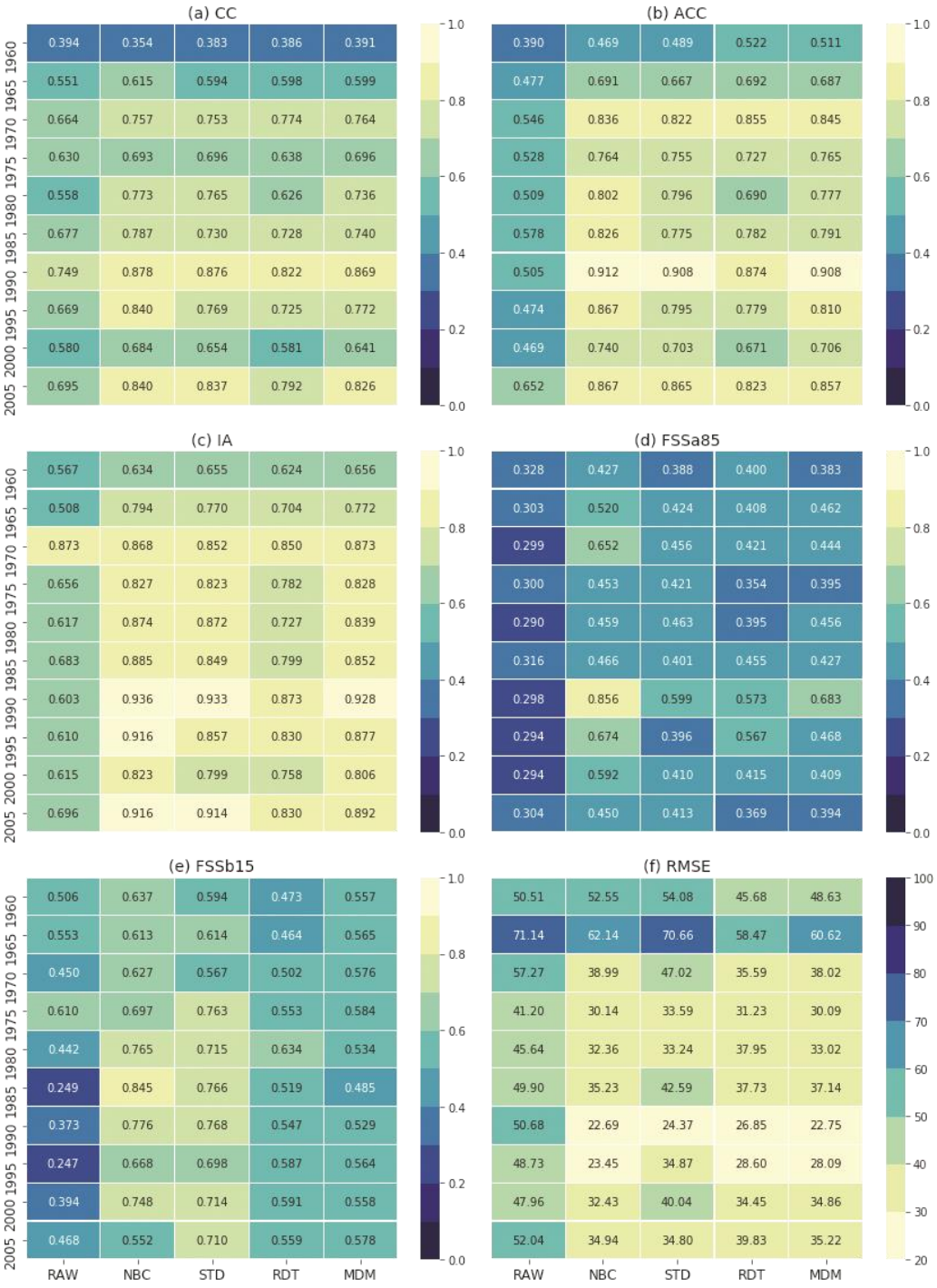


Fig. 6-4 Skill comparison of different drift correction methods, obtained from EC-EARTH. The vertical axis presents initialization years and the horizontal axis are presenting different drift correction methods including model (RAW) values

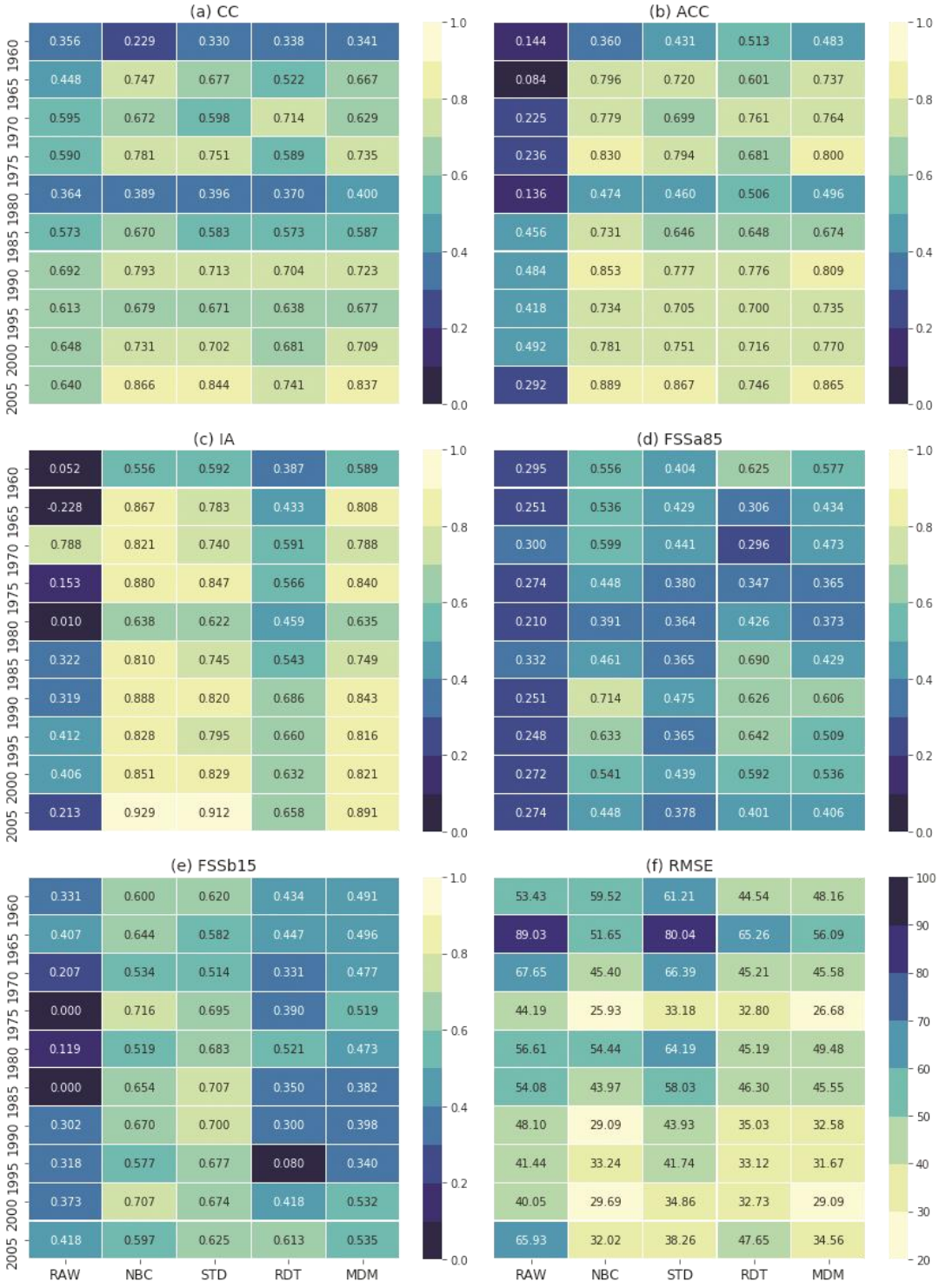


Fig. 6-5 Skill comparison of different drift correction methods, obtained from CanCM4. The vertical axis presents initialization years and the horizontal axis are presenting different drift correction methods including model (RAW) values

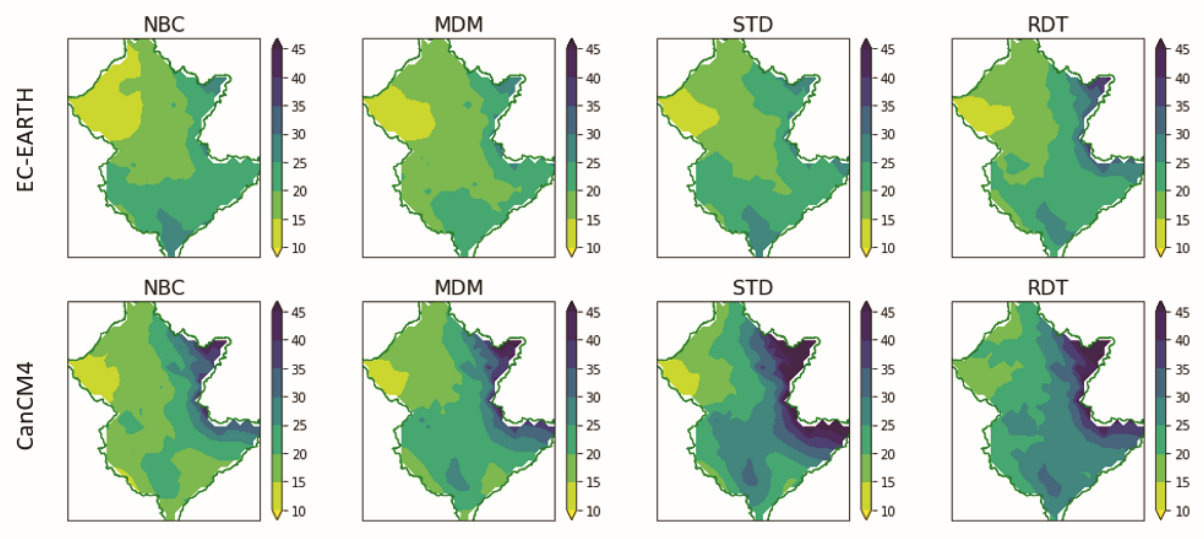


Fig. 6-6 Comparison of drift correction approaches for the skill (RMSE) over the catchment. The first row is presenting the spatial distribution of RMSE of the EC-EARTH model (initialization year 1990) after drift correction and the second row is presenting the spatial distribution of RMSE of the CanCM4 model (initialization year 1990) after drift correction

For all selected models, among the considered drift correction methods, NBC has the highest score on all skill tests except FSSb15. For FSSb15, NBC follows the STD method, however, still outperforms all other drift correction methods. This is in line with the lower performance of NBC in reproducing dry extremes in dry seasons, also evident from Fig. 6-7. It should be noted that NBC improves the reproducibility of wet extremes in wet seasons (Fig. 6-8). Note, the fractional skill scores are computed only for dry (JJA) and wet seasons (DJF).

STD and MDM perform similarly on all skill tests, albeit with slight differences for different model categories. For EC-EARTH, MRI-CGCM3, and MPI-ESM-LR models, the STD method outperforms MDM on CC, ACC, and IA but for the other skill tests, the reverse is true. For FSSb15, the STD method outperforms all other drift correction methods but lags MDM for RMSE and MAE.

This suggests that the STD method can reproduce dry seasons better than all other methods, albeit with larger errors than MDM. For models other than EC-EARTH, MRI-CGCM3, and MPI-ESM-LR, MDM outperforms the STD method, thus implying that MDM is more suited for Category-II models. From the different skill tests and drift correction methods used here, it is difficult to conclude which method has the best overall performance. However, it was found that STD is more suitable for Category-I, whereas, MDM is suitable for Category-II models.

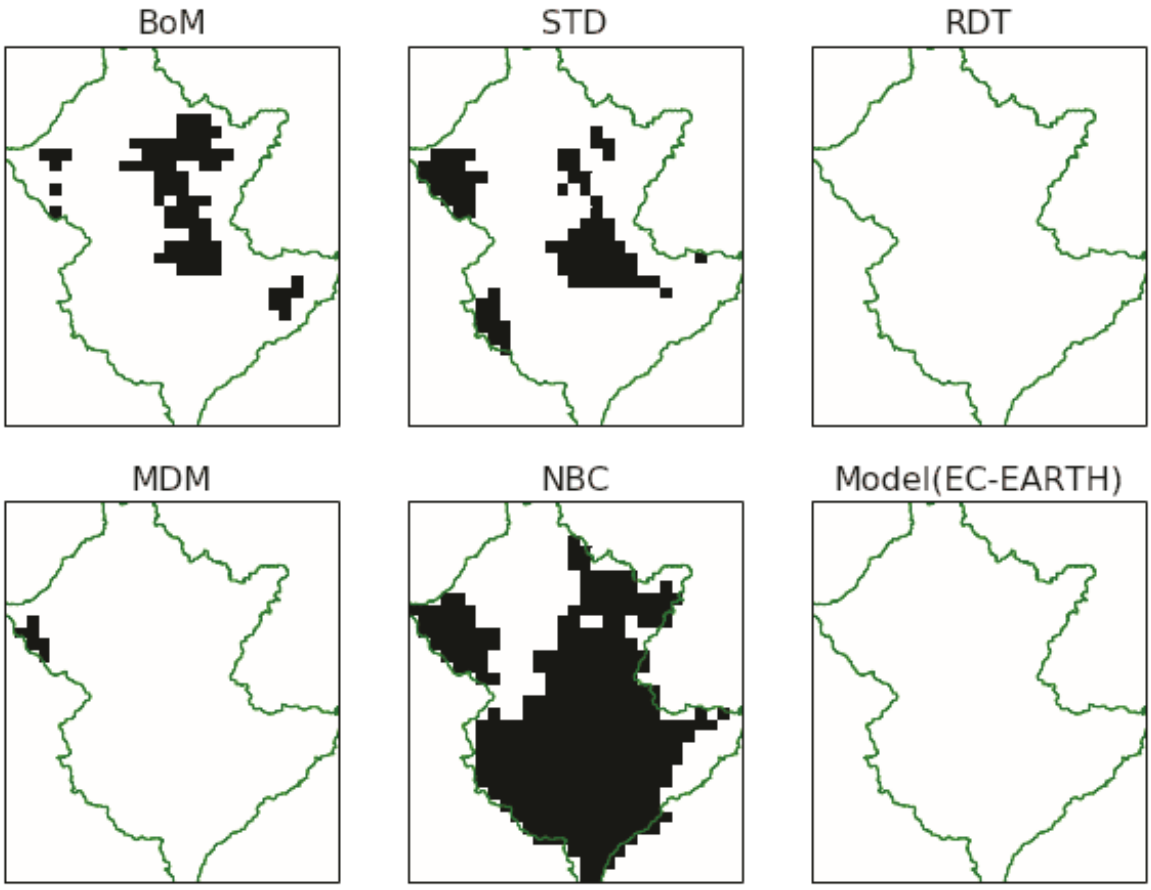


Fig. 6-7 Spatial variability comparison for an example dry season before and after correction (Initialization year 1990, Season JJA 1992). The black-coloured grids in different drift correction methods (STD, RDT, MDM, and NBC), model raw values (EC-EARTH), and the observed data (BoM) present precipitation values below 15 percentile (23.6 mm) of BoM (observed).

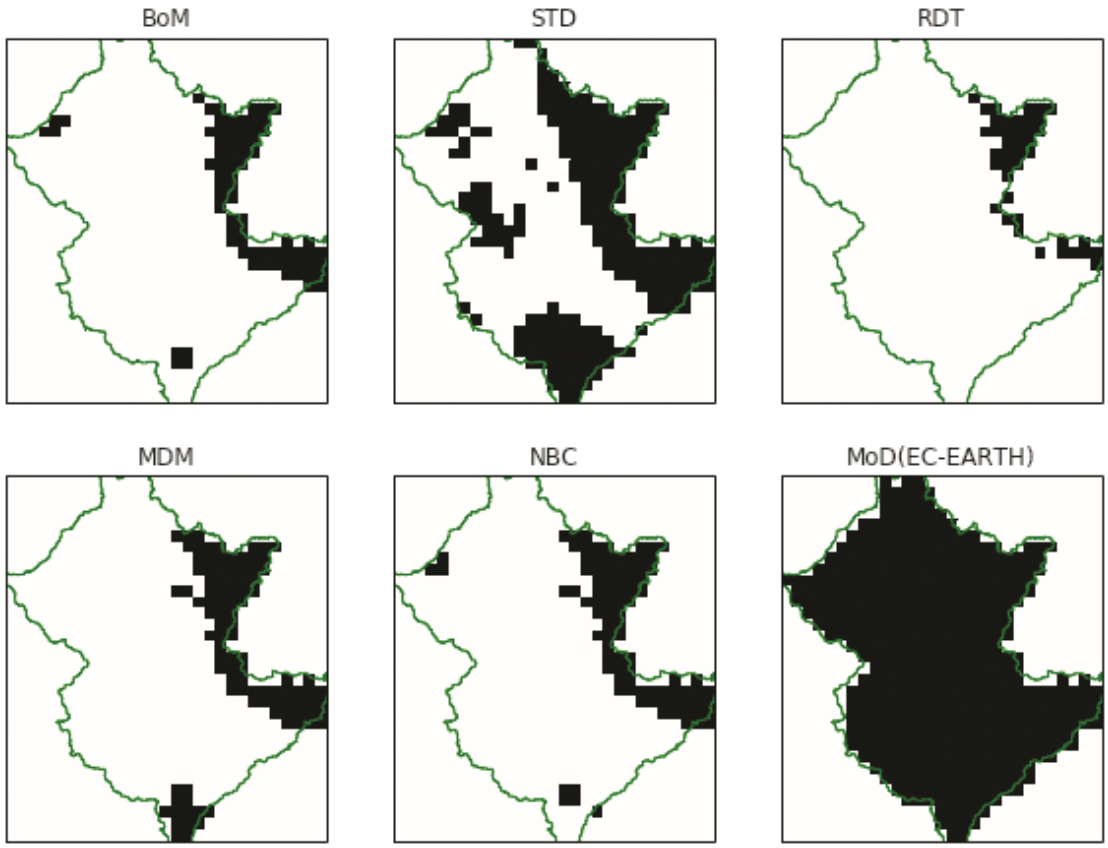


Fig. 6-8 Spatial variability comparison of different drift correction methods for an example wet season (initialization year 1990, Season DJF composed of December 1993 & January-February of 1994). The black-coloured grids in different drift correction methods (STD, RDT, MDM, and NBC), model raw values (EC-EARTH), and the observed data (BoM) present precipitation above 85 percentile values (100.2 mm) of BoM (observed) data.

RDT is worst for all skill assessments of all models except for total annual rainfall and accumulated rainfall over five years or over a decade, where it performs the best. For the accumulated rainfall, all four drift correction methods performed well, with minimal differences when compared to the observed rainfall with RDT being the best. From the comparison of accumulated rainfall over a given time period, it is noticed that the RDT method can reproduce the accumulated rainfall accurately for almost all the models and all of the initialization years, but is unable to improve the temporal and spatial variability, anomalies and the phase differences, as noted for all the other methods.

CC measures correspondence between the drift corrected values and the observed values, ACC measures the correspondence between the anomalies of drift corrected values and observations, and IA measures the accuracy of the drift correction methods. It can be said that these three

skill tests measure the accuracy of the drift corrections over a given time span for each initialization year and at each grid individually. Consequently, fractional skill score; both FSSa85 during the wet season and FSSb15 during dry season, measures the spatial and temporal accuracy of the modeled precipitation (corrected or uncorrected) across the catchment. Fig. 6-7 and Fig. 6-8 exemplify the spatial variability of the precipitation extremes for a single dry and wet season of a specific year respectively, while the spatial variability of the total rainfall during a single wet season across the entire catchment is shown in Fig. 6-9. It is evident model EC-EARTH has a wet bias during the wet season (Fig. 6-8) and dry bias during the dry seasons (Fig. 6-7), similar characteristics are exhibited by all of the models. However, for different initialization years or different lead-time for the same initialization year, the model skill prior to and after the drift correction varies. Moreover, FSS also provides the temporal average scores where the performance of different models and of different drift correction methods for individual years (or lead-time) is overlooked.

The spatial variability from different drift correction methods for individual years is evaluated by investigating the number of grids with the two thresholds. To that end, four models (i.e., EC-EARTH, MRI-CGCM3, MPI-ESM-LR, and CanCM4) and two initialization years (1990 and 2000) are chosen. It is compared by the ratio of the number of grids for different models, both before and after drift correction, to the number of grids with the said threshold in the observed precipitation. Fig. 6-10 and Fig. 6-11 compare the ratios for individual years and for the four models, where, the models were initialized in 1990. It is obvious that, before the drift correction, models exhibit a wet (dry) bias in the wet (dry) season.

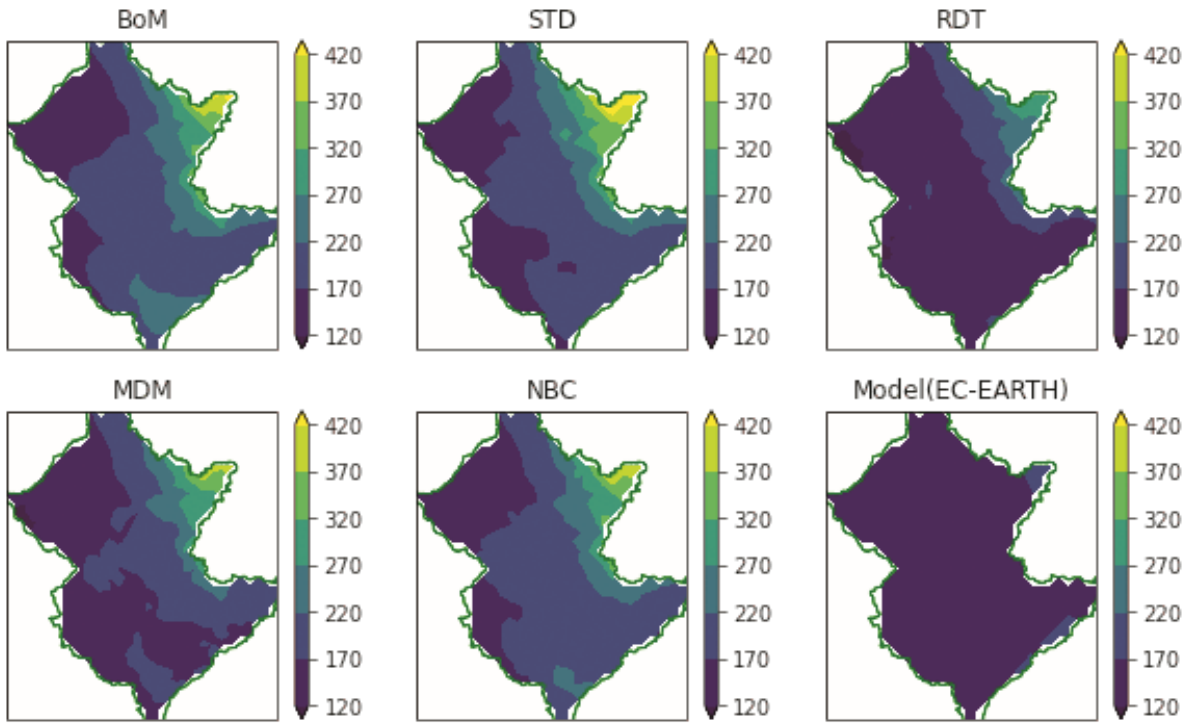


Fig. 6-9 Comparison of spatial variability of total precipitation of a sample wet season of EC-EARTH model (initialization year 1990, season=DJF,1991[Dec] &1992[Jan & Feb]) of different drift correction methods. The color bar on the right of each plot presents the precipitation in millimetres

However, during the wet season, over-prediction is more pronounced than under-prediction in the dry season. After the drift correction, the dry and wet bias in model outputs decreased with an improved representation of spatial variability. NBC corrected data is closest to 1.0, followed by the MDM method (Fig. 6-10). Consequently, the STD method outperforms other methods in reproducing dry extremes (Fig. 6-11), also noted from FSSb15 scores.

Qualitatively, similar results were found at other grids, albeit with quantitative differences in skill scores. The improvement from STD and MDM vary for different model categories and different skill tests. For instance, when comparing CC and IA, both STD and MDM methods show similar improvements. When comparing ACC, MDM performs best for Category-II models, with similar improvements Category-I models. Finally, it is worth noting that for the other skill tests: FSSa85, RMSE, and MAE, MDM outperforms for Category-II models, whereas STD is best for Category-I models. As there are 7 skill tests, 4 drift correction methods, 8 models, and 10 initialization years used in this study, there might be too many results to present here.

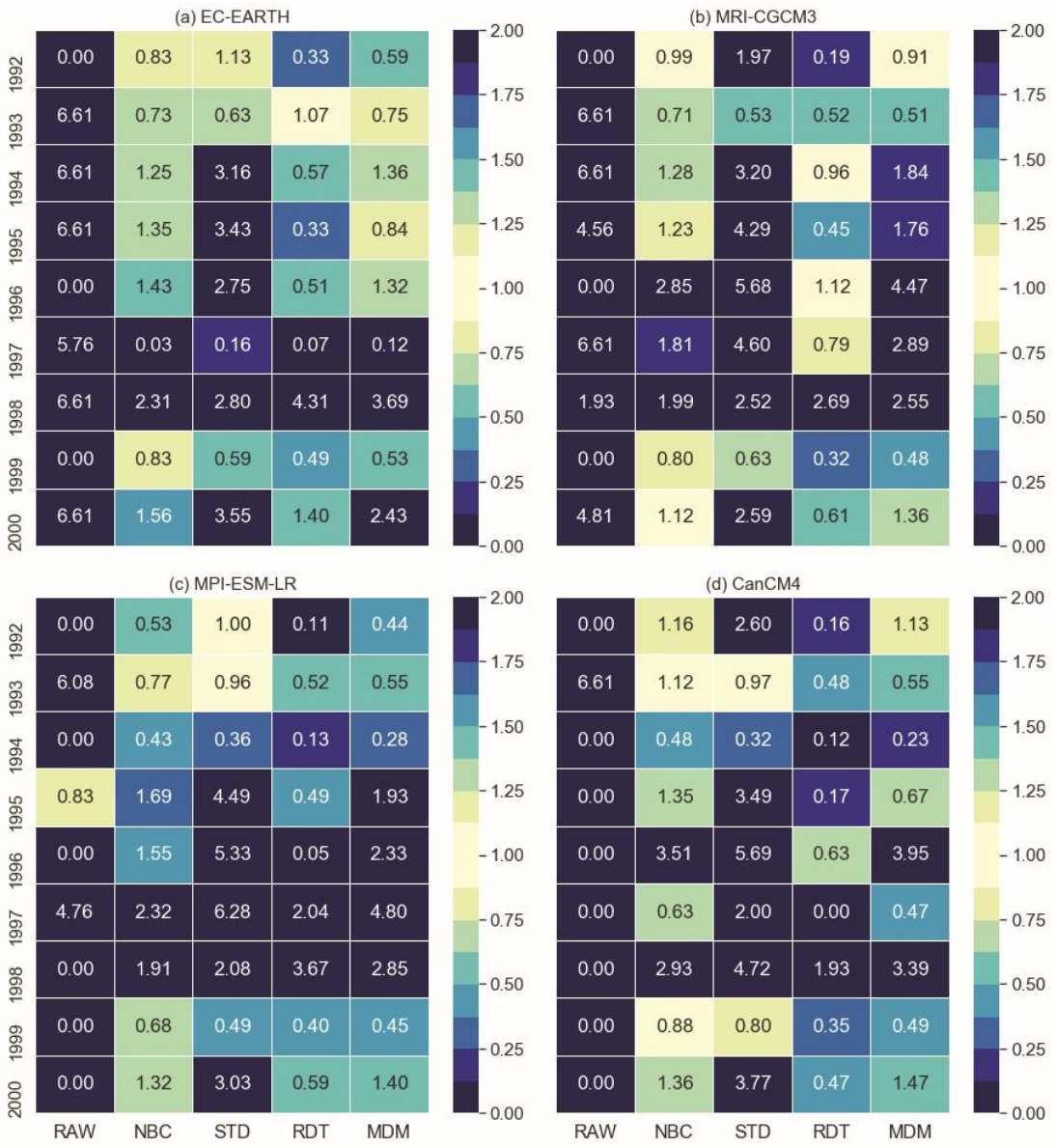


Fig. 6-10 Example comparison of reproducing extreme wet events among different drift correction methods. This comparison is based on the ratio of the number of grids covered by the models' simulated (initialized in 1990) values to the number of grids covered by the observed values. These ratios are for the threshold equal and above 85 percentile of the observed data (in the wet season, DJFs only). Values 1.0 presents the exact correspondence whilst values more and less than 1.0 indicate over and underestimation by the drift correction methods (models)

The change of skills for all initialization years and drift correction methods for two representative models (EC-EARTH and CanCM4; one from each category) after drift correction are given in Table 6-S1 and Table 6-S2 in supplementary materials. Reduction of error (such as average RMSE) of different models and the relative change for different drift correction methods are presented in Table 6-2.

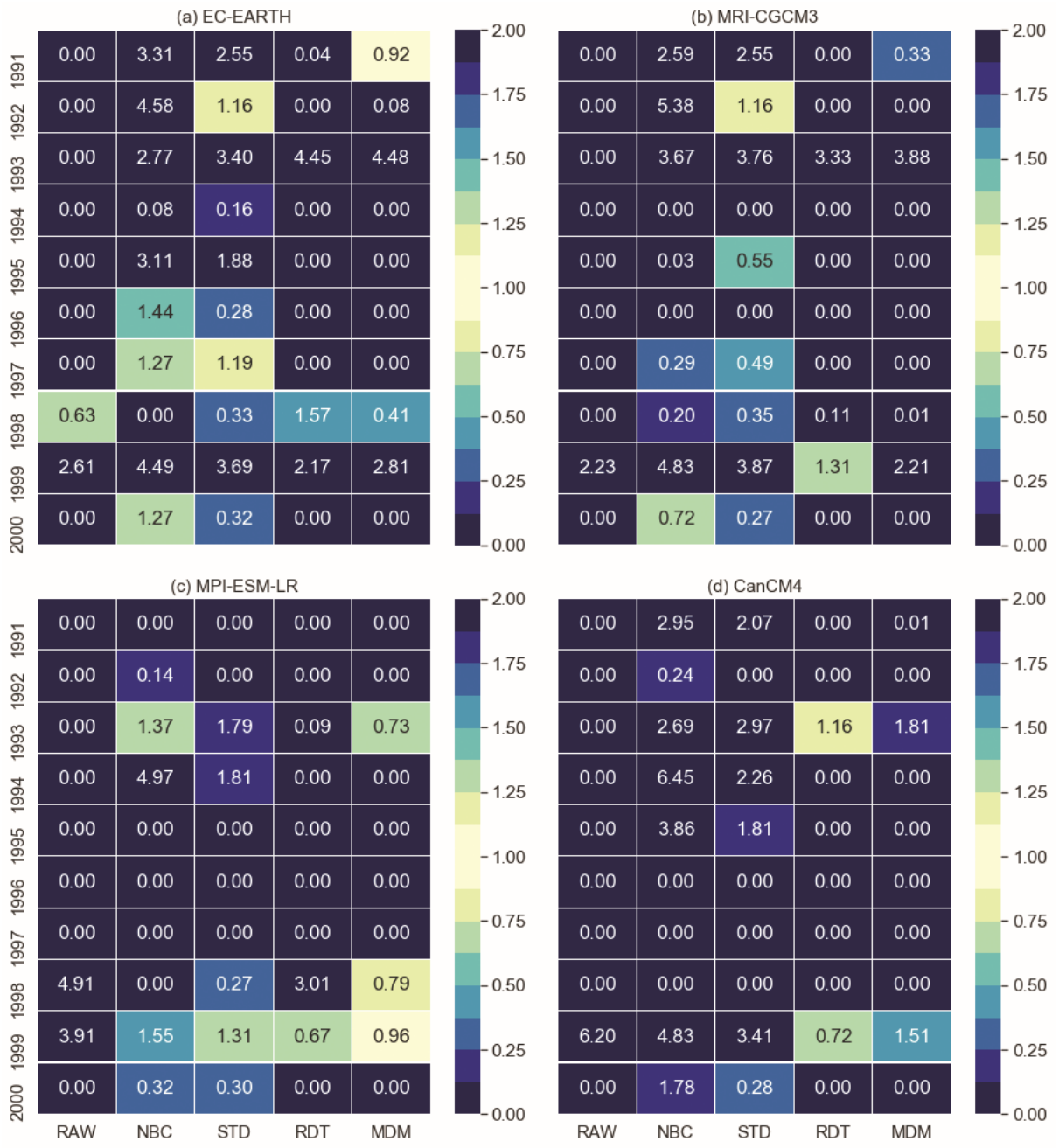


Fig. 6-11 Example comparison of reproducing extreme dry events among different drift correction methods. This comparison is based on the ratio of the number of grids covered by the models' raw (initialized in 1990) value to the number of grids covered by the observed data (in the dry season, JJAs only). The values 1.0 present the exact correspondence whilst values above and below 1.0 indicate over and underestimation by the drift correction methods (models)

Table 6-2 Reduction in average RMSE in percent of individual drift corrections. Negative values indicate the reduction in RMSE values of model data without drift correction (RAW). Relative changes of individual drift corrections are in the right columns of respective drift correction methods. Average RMSE means, the average of RMSE values of all initialization years started from 1960 to 2005 at the selected grid

Models	RAW	NBC	Change in %	STD	Change in %	RDT	Change in %	MDM	Change in %
EC-EARTH	51.51	36.49	-29.15	41.53	-19.38	37.64	-26.93	36.84	-28.47
MRI-CGCM3	53.16	40.85	-23.15	49.75	-6.40	40.56	-23.70	41.57	-21.79
MPI-ESM-LR	53.44	39.89	-25.35	45.69	-14.50	39.78	-25.56	39.74	-25.63
MIROC4h	54.75	40.83	-25.42	49.52	-9.56	45.17	-17.49	44.93	-17.94
MPI-ESM-MR	55.44	49.01	-11.58	56.09	1.18	44.29	-20.11	46.33	-16.42
MIROC5	54.69	43.45	-20.56	59.58	8.93	43.92	-19.69	45.21	-17.33
CMCC-CM	71.79	45.78	-36.23	50.03	-30.31	51.11	-28.81	48.52	-32.41
CanCM4	56.05	40.49	-27.76	52.18	-6.90	42.78	-23.67	39.95	-28.73

In general, all drift correction methods show a significant reduction in RMSE for all initialization years, but the reductions for individual models varied for different drift correction methods. For instance, the NBC method reduces RMSE of EC-EARTH by 29.15%, whereas, MDM, STD, and RDT reduce it by 28.47%, 19.38%, and 26.93% respectively. NBC outperforms all other methods in reducing RMSE, followed by MDM, similar reductions are noted for MAE (not described here). Furthermore, RDT performed better than STD in reducing RMSE and MAE. EC-EARTH and CMCC-CM had the highest reduction in RMSE for all drift correction methods, however, the reduction in RMSE/ MAE varied for MIROC5. Contrary to expectation, for the STD method, the overall RMSE in MIROC5 and MPI-ESM-MR increased. This unusual increase in RMSE was not seen for other models, as RMSE was reduced for all other models. When comparing the relative change in RMSE values for different drift correction methods, minimal changes are noted for MIROC5 and MPI-ESM-MR as opposed to the other models. Upon further investigation, the STD method was found to perform better in reproducing the observed wet extremes, especially for MIROC5, where, the wet peaks were reproduced for almost all initialization years. For the MIROC5 model, the range between wet and dry peaks is comparatively higher, with more outliers than in other models, (see Fig. 6-12).

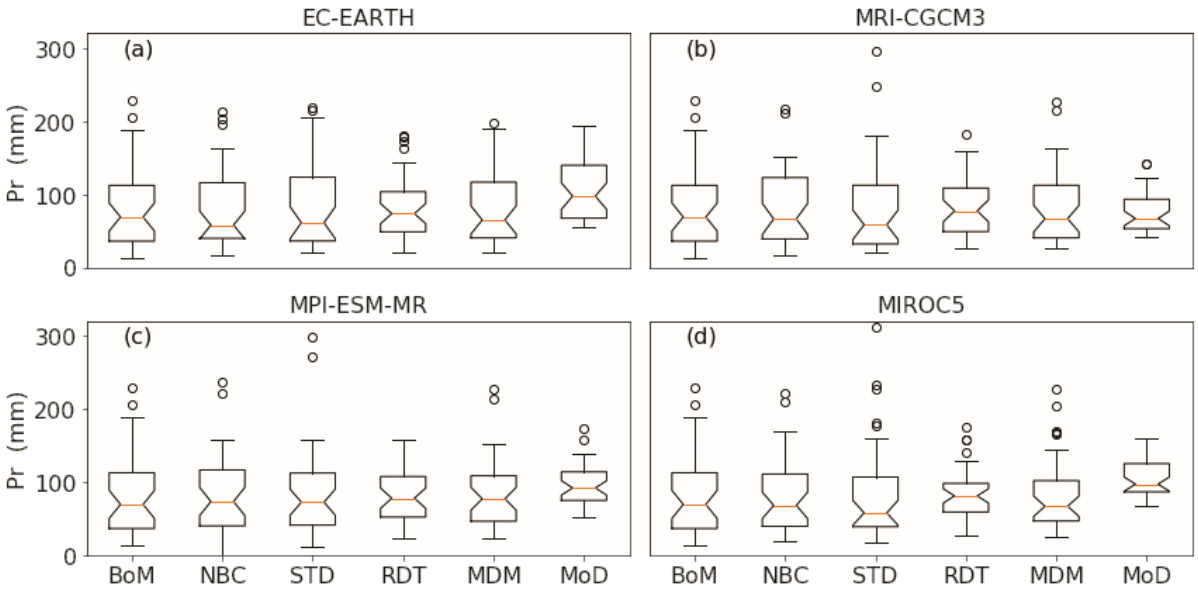


Fig. 6-12 Box Whisker plot of four selected models, initialization year 1990 (January 1991-December 2000). Y-axis presents precipitation in millimeters and the x-axis presents drift correction methods including model and BoM data

6.5 Discussion

This study assesses the drift correction alternatives for CMIP5 decadal experiments' hindcast precipitation (seasonal mean) from different models at the catchment scale. Three existing drift correction methods namely NBC, STD, RDT, and one modified method (MDM) were employed and their skills were assessed for the individual models and all initialization years. Previous studies compared the performance of the drift/bias correction methods on their efficacy in reducing errors like RMSE or MAE (Purwaningsih and Hidayat, 2016) only. Some other studies used ACC (Choi et al., 2016) and CC along with RMSE (Kamworapan and Surussavadee, 2019) but their applications were for different spatial scales. In this study, we used a comprehensive list of skill tests developed from different perspectives, thus enabling skill assessment in resembling both temporal and spatial variability. Moreover, these methods are applied at catchment scales and a higher spatial resolution of 0.05-degree. Here, the ensembles' mean (of all available ensembles) for every initialization year of all models were used, as in chapter 5 it has been seen that the ensemble mean has higher skill than the individual ensembles which was also confirmed in previous studies (Choudhury et al., 2015; Kim et al., 2012; Schepen et al., 2014). All the available ensembles are considered regardless of the initialization conditions; full field or anomaly. Furthermore, the monthly precipitation was

aggregated to a seasonal mean. This overlooks the drift in individual ensemble members and over different months. However, when comparing different drift correction methods, the overall effects of a method are of more relevance.

The results presented in this study showed how the skill of drift correction methods and of different models varied over the initialization years. After going through a comprehensive assessment, it is revealed that different drift correction methods have their own advantages and disadvantages. Overall, the NBC method performed best unless the drift is not too high, whereas improvements from other methods varied over different skill tests and for different models. For instance, the STD method performed better than MDM on CC and FSSb15 while lagging on RMSE and MAE. When comparing for different model categories, MDM showed better improvement in ACC and FSSa85 for the Category-II models but similar improvements for category-I models. However, both STD and MDM methods showed similar improvements in the case of IA. The lower ACC and FSSa85 for STD may be attributed to the overestimation of higher wet peaks and a larger range of outliers not noticed for the MDM method.

When comparing models for different initialization years, it is noticed that, after the drift correction, the skill was higher for those initialization years where the models performed better before drift correction. This suggests that models, which perform better prior to drift correction, do so even after the drift correction. It is also noticeable that the models could reproduce both the anomalies and extreme events of observed precipitation after the drift correction while it could not do so before the drift correction. The model skill in reproducing extreme events and anomalies was higher for the initialization year from 1990 to 2005 than for the earlier initialization years, i.e., 1960 to 1980 (not shown).

From both qualitative and quantitative skill measurements, it was found that models with higher skill before drift correction also have higher skill after the drift correction. But the improvement from drift correction differed for different models. For instance, in all drift correction methods, MRI-CGCM3 showed the highest improvement in CC while MPI-ESM-LR showed the highest in ACC and IA, and similar results were found for other models for other skill tests. But highest improvement does not necessarily mean the highest performance of the model, rather it is the change in the skill of a given model before and after the correction. Actually, model performance depends on various factors such as variables affecting earth-climate interaction, geographical locations (Choi et al., 2016; Homsi et al., 2020; Purwaningsih

and Hidayat, 2016), the spatial resolution of model structure (Jain et al., 2019; Lovino et al., 2018), the temporal resolution of simulating variables, and spatial and temporal scales of the considered variables (Sheffield et al., 2013; Ta et al., 2018). Comparison of model performances of different categories (as identified in this study) revealed that the models in category-I are more skilful than models in Category-II for decadal hindcast precipitation. This may be because of the models with an atmospheric finer spatial resolution (see Table 1) could reproduce local climate features better than the coarser one (Jain et al., 2019; Lovino et al., 2018). However, the skill of the CMCC-CM model for precipitation was found low which is similar to the findings of Lovino et al. (2018). But this model performs better for temperature forecasting as reported by Lovino et al. (2018). Among the individual models, EC-EARTH outperforms all models both before and after the drift corrections.

For simulating future climate, it is reasonable to assume that the drift of hindcast and future simulated data would not change as the observed and modelled statistics of current climate are used to adjust for the future model result (Johnson and Sharma, 2012). This assumption may have a shift of statistics but allows to have a similar relationship (e.g. drift) between the current and future climate. In this study, this assumption seems more reasonable for decadal data where the drift may not change drastically within a ten years' time period as climate change is a slow evolving process.

The assessment and applicability of forecasted rainfall at a higher spatial resolution is important for the local stakeholders. In Australia, the climate shows high year-on-year variability and can have an enormous economic impact. For instance, a typical major drought in a season can reduce agricultural production by about 10% and gross national product by 1% (White, 2000). The rainfall forecasting can be beneficial for agriculture and agriculture depended businesses, decision and policymaking (Hansen et al., 2011; Jones et al., 2000) for water resource management, and other sectors like the retail industry, finance, insurance, fishery, transport, tourism, and others. The stakeholders from these sectors have different spatial and temporal requirements for such data. For instance, in agriculture sector, farmers demand rainfall data at local scale or even at district level (Paull, 2002), with a lead time of few months or couple of seasons, to aid in their decision making process for farming and agricultural management. While, wholesalers, retailers of production, grain/fibre handling and marketing organizations, processors, forward sellers and purchasers of agronomic products, water resources managers,

different public and private investors like insurance companies demand more accurate forecasts with longer lead times (Paull, 2002). These stakeholders need future rainfall forecast at local level and for longer time span for their businesses. For the practical use of CMIP5 data, the outcomes of this study of using alternative drift correction methods will be very useful for taking any informed decision in order to expand or plan for their businesses in future.

6.6 Conclusions

Forecasting precipitation of high spatial resolution at the local level for a longer time span (e.g. seasons to decades) has great societal importance. CMIP5 decadal experiment data explored climate features for decadal timescale that has attracted stakeholders for their suitable applications. However, drift correction is required to improve the quality of outputs from the GCMs of CMIP5 decadal experiments before they can be used. This study employed four drift correction methods on CMIP5 decadal precipitation and compared the improvement in individual skills from different drift corrections for individual models. Based on the comprehensive assessment undertaken here, this study concludes that the NBC method performs best for CMIP5 decadal precipitation unless the drifts are not too high. However, the advantages and disadvantages of STD and MDM made it difficult to conclude which one of these two methods is better, thus requiring prudence when selecting a method. Depending on the desired skill and the potential applications, drift correction methods and models can be selected. For instance, when the accuracy of dry seasons is weighed higher than that of wet seasons, the STD method can be a better option, while in instances where the accuracy of wet seasons and the seasonal anomalies is desired, MDM might be a better option. However, when only considering the reduction in error (both RMSE and MAE) like in past studies, then MDM is always superior to the STD method. While considering the total accumulated precipitation, the RDT method outperformed all other methods and it can be the best option to choose for relevant uses. Finally, it should be noted that the statistical drift correction methods employed here are not enough to decide the best drift correction approach for the decadal hindcast precipitation at seasonal timescale and at high spatial resolution. Further investigation on drift correction approaches for different time scales e.g., monthly precipitation, and their application to individual ensembles is recommended.

Acknowledgments

We thank the working group of the World Climate Research Program who are responsible for CMIP5 and making available their models' output used in this study (see Table-1). We also would like to thank the Bureau of Meteorology, Australia for providing the gridded observed data. Authors gratefully acknowledge the financial support of the CIPRS scholarship of Curtin University and Data61 student scholarship of CSIRO (Commonwealth Scientific and Industrial Research Organisation) jointly provided to the first author for his Ph.D. study at Curtin University, Australia. The authors also thank the anonymous reviewers for their valuable comments on the first draft of this paper.

List of symbols

C	:	Mean (over the total time span) of the observed values
F	:	Model forecasted (corrected) values
O	:	Observed values
$F - C$	:	Anomaly of the model values
$\overline{F - C}$	:	Mean of the model anomalies
$O - C$	:	Observed anomaly
$\overline{O - C}$	:	Mean of the observed anomalies
P *(or p)*	:	Precipitation
P_f	:	Calculated fraction in FSS
μ	:	Mean
σ	:	Standard deviation,
t (subscript)	:	Lead time in year
i (subscript)	:	individual season of time *t*

m (superscript):		Represents the model
o (superscript):		Represent the observed
p_{it}	:	Standardized time series (seasonal)
$\tilde{p}_{i,t}^{m}$	:	Standardized time series (seasonal) after replacing models' lag-1 autocorrelation by the corresponding observed values
$\tilde{P}_{i,t}^{m}$	:	Nested time series (seasonal)
p_{t}^{m}	:	Standardized time series (yearly)
$\tilde{p}_{t}^{m}$	:	Standardized time series (yearly) after replacing models' lag-1 autocorrelation by the corresponding observed values
$\tilde{P}_{t}^{m}$	:	Drift corrected seasonal precipitation (NBC)
$P_{cor,it}$	:	Drift corrected seasonal precipitation (for STD, RDT and MDM)
r_{i}^{m}	:	Lag-1 autocorrelations

References

Amengual, A., Homar, V., Romero, R., Alonso, S., Ramis, C., 2012. A statistical adjustment of regional climate model outputs to local scales: Application to Platja de Palma, Spain. Journal of Climate 25, 939–957. https://doi.org/10.1175/JCLI-D-10-05024.1

Bates, B.C., Charles, S.P., Hughes, J.P., 1998. Stochastic downscaling of numerical climate model simulations. Environmental Modelling and Software 13, 325–331. https://doi.org/10.1016/S1364-8152(98)00037-1

Boer, G.J., Smith, D.M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G.A., Msadek, R., Mueller, W.A., Taylor, K.E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., Eade, R., 2016. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. Geoscientific Model Development 9, 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Cardoso Pereira, S., Marta-Almeida, M., Carvalho, A.C., Rocha, A., 2020. Extreme

precipitation events under climate change in the Iberian Peninsula. International Journal of Climatology 40, 1255–1278. https://doi.org/10.1002/joc.6269

Charles, S.P., Bates, B.C., Smith, I.N., Hughes, J.P., 2004. Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. Hydrological Processes 18, 1373–1394. https://doi.org/10.1002/hyp.1418

Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resources Research 49, 4187–4205. https://doi.org/10.1002/wrcr.20331

Chen, J., Brissette, F.P., Lucas-Picher, P., 2015. Assessing the limits of bias-correcting climate model outputs for climate change impact studies. Journal of Geophysical Research 120, 1123–1136. https://doi.org/10.1002/2014JD022635

Choi, J., Son, S.W., Ham, Y.G., Lee, J.Y., Kim, H.M., 2016. Seasonal-to-interannual prediction skills of near-surface air temperature in the CMIP5 decadal hindcast experiments. Journal of Climate 29, 1511–1527. https://doi.org/10.1175/JCLI-D-15-0182.1

Choudhury, D., Sen Gupta, A., Sharma, A., Mehrotra, R., Sivakumar, B., 2017. An Assessment of Drift Correction Alternatives for CMIP5 Decadal Predictions. Journal of Geophysical Research: Atmospheres 122, 10282–10296. https://doi.org/10.1002/2017JD026900

Choudhury, D., Sharma, A., Sen Gupta, A., Mehrotra, R., Sivakumar, B., 2016. Sampling biases in CMIP5 decadal forecasts. Journal of Geophysical Research: Atmospheres 121, 3435–3445. https://doi.org/10.1002/2016JD024804

Choudhury, D., Sharma, A., Sivakumar, B., Sen Gupta, A., Mehrotra, R., 2015. On the predictability of SSTA indices from CMIP5 decadal experiments. Environmental Research Letters 10, 074013. https://doi.org/10.1088/1748-9326/10/7/074013

CLIVAR, 2011. Decadal and bias correction for decadal climate predictions. CLIVAR Publication Series No.150, 6pp 1–3.

Doblas-Reyes, F.J., Balmaseda, M.A., Weisheimer, A., Palmer, T.N., 2011. Decadal climate prediction with the European Centre for Medium-Range Weather Forecasts coupled forecast system: Impact of ocean observations. Journal of Geophysical Research Atmospheres 116, 1–13. https://doi.org/10.1029/2010JD015394

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27, 1547–1578. https://doi.org/10.1002/joc.1556

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

Fučkar, N.S., Volpi, D., Guemas, V., Doblas-Reyes, F.J., 2014. A posteriori adjustment of near-term climate predictions: Accounting for the drift dependence on the initial conditions. Geophysical Research Letters 41, 5200–5207. https://doi.org/10.1002/2014GL060815

Gangstø, R., Weigel, A.P., Liniger, M.A., Appenzeller, C., 2013. Methodological aspects of the validation of decadal predictions. Climate Research 55, 181–200. https://doi.org/10.3354/cr01135

Graham, L.P., Andreáasson, J., Carlsson, B., 2007. Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods - A case study on the Lule River basin. Climatic Change 81, 293–307. https://doi.org/10.1007/s10584-006-9215-2

Grotch, S.L., MacCracken, M.C., 1991. The Use of General Circulation Models to Predict Regional Climatic Change. Journal of Climate 4, 286–303. https://doi.org/10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2

Hansen, J.W., Mason, S.J., Sun, L., Tall, A., 2011. Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture 47, 205–240. https://doi.org/10.1017/S0014479710000876

Hawkins, E., Osborne, T.M., Ho, C.K., Challinor, A.J., 2013. Calibration and bias correction of climate projections for crop modelling: An idealised case study over Europe. Agricultural and Forest Meteorology 170, 19–31. https://doi.org/10.1016/j.agrformet.2012.04.007

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction – The ISI-MIP approach. Earth System Dynamics 4, 219–236. https://doi.org/10.5194/esd-4-219-2013

Ho, C., 2010. Projecting extreme heat-related mortality. Ph.D Thesis. College of Engineering,

Mathematics and Physical Sciences, University of Exeter ; Met Office.

Homsi, R., Shiru, M.S., Shahid, S., Ismail, T., Harun, S. Bin, Al-Ansari, N., Chau, K.W., Yaseen, Z.M., 2020. Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. Engineering Applications of Computational Fluid Mechanics 14, 90–106. https://doi.org/10.1080/19942060.2019.1683076

Ines, A.V.M., Hansen, J.W., 2006. Bias correction of daily GCM rainfall for crop simulation studies. Agricultural and Forest Meteorology 138, 44–53. https://doi.org/10.1016/j.agrformet.2006.03.009

Islam, S.A., Bari, M., Anwar, A.H.M.F., 2011. Assessment of hydrologic impact of climate change on Ord River catchment of Western Australia for water resources planning: A multi-model ensemble approach, in: Chan, F., Marinova, D. and Anderssen, R.S. (Eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand (MSSANZ), Inc. https://doi.org/10.36334/modsim.2011.I6.islam

Islam, S.A., Bari, M.A., Anwar, A.H.M.F., 2014. Hydrologic impact of climate change on Murray–Hotham catchment of Western Australia: a projection of rainfall–runoff for future water resources planning. Hydrology and Earth System Sciences 18, 3591–3614. https://doi.org/10.5194/hess-18-3591-2014

Jain, S., Salunke, P., Mishra, S.K., Sahany, S., 2019. Performance of CMIP5 models in the simulation of Indian summer monsoon. Theoretical and Applied Climatology 137, 1429–1447. https://doi.org/10.1007/s00704-018-2674-3

Johnson, F., Sharma, A., 2012. A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. Water Resources Research 48, 1–16. https://doi.org/10.1029/2011WR010464

Jones, J.W., Hansen, J.W., Royce, F.S., Messina, C.D., 2000. Potential benefits of climate forecasting to agriculture. Agriculture, Ecosystems and Environment 82, 169–184. https://doi.org/10.1016/S0167-8809(00)00225-5

Kamworapan, S., Surussavadee, C., 2019. Evaluation of CMIP5 Global Climate Models for Simulating Climatological Temperature and Precipitation for Southeast Asia. Advances in Meteorology 2019, 1–18. https://doi.org/10.1155/2019/1067365

Kim, H.-M., Webster, P.J., Curry, J.A., 2012. Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. Geophysical Research Letters 39, n/a-n/a. https://doi.org/10.1029/2012GL051644

Kruschke, T., Rust, H.W., Kadow, C., Müller, W.A., Pohlmann, H., Leckebusch, G.C., Ulbrich, U., 2016. Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms. Meteorologische Zeitschrift 25, 721–738. https://doi.org/10.1127/metz/2015/0641

Kumar, S., Merwade, V., Kinter, J.L., Niyogi, D., 2013. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. Journal of Climate 26, 4168–4185. https://doi.org/10.1175/JCLI-D-12-00259.1

Li, H., Sheffield, J., Wood, E.F., 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. Journal of Geophysical Research 115, D10101. https://doi.org/10.1029/2009JD012882

Liang, X.Z., Kunkel, K.E., Meehl, G.A., Jones, R.G., Wang, J.X.L., 2008. Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections. Geophysical Research Letters 35, 1–5. https://doi.org/10.1029/2007GL032849

Lovino, M.A., Müller, O. V., Berbery, E.H., Müller, G. V., 2018. Evaluation of CMIP5 retrospective simulations of temperature and precipitation in northeastern Argentina. International Journal of Climatology 38, e1158–e1175. https://doi.org/10.1002/joc.5441

Maraun, D., 2016. Bias Correcting Climate Change Simulations - a Critical Review. Current Climate Change Reports 2, 211–220. https://doi.org/10.1007/s40641-016-0050-x

Masanganise, J., Chipindu, B., Mhizha, T., Mashonjowa, E., Basira, K., 2013. An evaluation of the performances of Global Climate Models (GCMs) for predicting temperature and rainfall in Zimbabwe 3, 1–11.

Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. Hydrology and Earth System Sciences 12, 551–563. https://doi.org/10.5194/hess-12-551-2008

Maurer, E.P., Pierce, D.W., 2014. Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. Hydrology and Earth System Sciences 18, 915–925. https://doi.org/10.5194/hess-18-915-2014

Meehl, G.A., Goddard, L., Murphy, J., Stouffer, R.J., Boer, G.J., Danabasoglu, G., Dixon, K., Giorgetta, M.A., Greene, A.M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., Stockdale, T., 2009. Can It Be Skillful? Bulletin of the American Meteorological Society 90, 1467–1486. https://doi.org/10.1175/2009BAMS2778.I

Meehl, G.A., Teng, H., Maher, N., England, M.H., 2015. Effects of the Mount Pinatubo eruption on decadal climate prediction skill of Pacific sea surface temperatures. Geophysical Research Letters 42, 10840–10846. https://doi.org/10.1002/2015GL066608

Mehrotra, R., Sharma, A., 2016. A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: implications for downscaling. Journal of Climate 29, 3519–3539. https://doi.org/10.1175/JCLI-D-15-0356.1

Mehrotra, R., Sharma, A., 2015. Correcting for systematic biases in multiple raw GCM variables across a range of timescales. Journal of Hydrology 520, 214–223. https://doi.org/10.1016/j.jhydrol.2014.11.037

Mehrotra, R., Sharma, A., 2010. Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. Water Resources Research 46, 1–17. https://doi.org/10.1029/2009WR008423

Mehrotra, R., Sharma, A., 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. Journal of Geophysical Research Atmospheres 111, 1–16. https://doi.org/10.1029/2005JD006637

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. Journal of Hydrology 519, 2932–2951. https://doi.org/10.1016/j.jhydrol.2014.07.053

Miao, C., Su, L., Sun, Q., Duan, Q., 2016. A nonstationary bias-correction technique to remove bias in GCM simulations. Journal of Geophysical Research: Atmospheres 121, 5718–5735. https://doi.org/10.1002/2015JD024159

Paull, C.J., 2002. The value and benefits of using seasaonal climate forecasts in making

business decisions: a review 42.

Piani, C., Haerter, J.O., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theoretical and Applied Climatology 99, 187–192. https://doi.org/10.1007/s00704-009-0134-9

Purwaningsih, A., Hidayat, R., 2016. Performance of Decadal Prediction in Coupled Model Intercomparisson Project Phase 5 (CMIP5) on Projecting Climate in Tropical Area. Procedia Environmental Sciences 33, 128–139. https://doi.org/10.1016/j.proenv.2016.03.064

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E., 2007. Climate Models and Their Evaluation.

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Monthly Weather Review 136, 78–97. https://doi.org/10.1175/2007MWR2123.1

Rodwell, M.J., Richardson, D.S., Hewson, T.D., Haiden, T., 2010. A new equitable score suitable for verifying precipitation in numerical weather prediction. Quarterly Journal of the Royal Meteorological Society 136, 1344–1363. https://doi.org/10.1002/qj.656

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. International Journal of Climatology 23, 887–901. https://doi.org/10.1002/joc.922

Schepen, A., Wang, Q.J., Robertson, D.E., 2014. Seasonal Forecasts of Australian Rainfall through Calibration and Bridging of Coupled GCM Outputs. Monthly Weather Review 142, 1758–1770. https://doi.org/10.1175/MWR-D-13-00248.1

Shams, M.S., Faisal Anwar, A.H.M., Lamb, K.W., Bari, M., 2018. Relating ocean-atmospheric climate indices with Australian river streamflow. Journal of Hydrology 556, 294–309. https://doi.org/10.1016/j.jhydrol.2017.11.017

Sharma, D., Gupta, A. Das, Babel, M.S., 2007. Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. Hydrology and Earth System Sciences 11, 1373–1390. https://doi.org/10.5194/hess-11-1373-2007

Sheffield, J., Camargo, S.J., Fu, R., Hu, Q., Jiang, X., Johnson, N., Karnauskas, K.B., Kim, S.T., Kinter, J., Kumar, S., Langenbrunner, B., Maloney, E., Mariotti, A., Meyerson, J.E., Neelin, J.D., Nigam, S., Pan, Z., Ruiz-Barradas, A., Seager, R., Serra, Y.L., Sun, D., Wang, C., Xie, S., Yu, J., Zhang, T., Zhao, M., 2013. North American Climate in CMIP5 Experiments. Part II: Evaluation of Historical Simulations of Intraseasonal to Decadal Variability. Journal of Climate 26, 9247–9290. https://doi.org/10.1175/JCLI-D-12-00593.1

Sperna Weiland, F.C., Van Beek, L.P.H., Kwadijk, J.C.J., Bierkens, M.F.P., 2010. The ability of a GCM-forced hydrological model to reproduce global discharge variability. Hydrology and Earth System Sciences 14, 1595–1621. https://doi.org/10.5194/hess-14-1595-2010

Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J.C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., Haynes, J., 2010. Dreary state of precipitation in global models. Journal of Geophysical Research Atmospheres 115, 1–14. https://doi.org/10.1029/2010JD014532

Sun, Y., Solomon, S., Dai, A., Portmann, R.W., 2007. How often will it rain? Journal of Climate 20, 4801–4818. https://doi.org/10.1175/JCLI4263.1

Ta, Z., Yu, Y., Sun, L., Chen, X., Mu, G., Yu, R., 2018. Assessment of Precipitation Simulations in Central Asia by CMIP5 Climate Models. Water 10, 1516. https://doi.org/10.3390/w10111516

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Viceto, C., Cardoso Pereira, S., Rocha, A., 2019. Climate Change Projections of Extreme Temperatures for the Iberian Peninsula. Atmosphere 10, 229. https://doi.org/10.3390/atmos10050229

White, B., 2000. The Importance of Climate Variability and Seasonal Forecasting to the Australian Economy. Springer, Dordrecht, pp. 1–22. https://doi.org/10.1007/978-94-015-9351-9_1

Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J., Wilks, D.S., 1998. Statistical downscaling of general circulation model output: A comparison of

methods. Water Resources Research 34, 2995–3008. https://doi.org/10.1029/98WR02577

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic Change 62, 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Supplementary information for chapter 6

Table 6-S1 Change of Skills after drift correction of EC-EARTH. The Positive values indicate increase in skills and vice versa

Initialization year	ACC Change in %				IA Change in %				FSSa85 Change in %				RMSE Change in %			
	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT
1960	20.27	31.12	25.44	33.90	11.77	15.75	15.60	10.04	30.12	16.76	18.15	21.74	4.0	-3.7	7.1	-9.5
1965	44.81	43.99	39.84	45.03	56.24	51.83	51.45	38.42	71.40	52.18	39.70	34.64	-12.6	-14.8	-0.7	-17.8
1970	53.10	54.75	50.58	56.60	-0.55	0.00	-2.41	-2.70	117.60	48.36	52.31	40.67	-31.9	-33.6	-17.9	-37.9
1975	44.63	44.91	42.90	37.73	26.19	26.35	25.48	19.29	51.12	31.61	40.24	17.95	-26.9	-27.0	-18.5	-24.2
1980	57.63	52.72	56.45	35.53	41.72	36.03	41.35	17.94	58.07	56.86	59.39	36.02	-29.1	-27.6	-27.2	-16.9
1985	42.89	36.81	34.08	35.16	29.57	24.65	24.21	16.98	47.66	35.32	26.87	44.19	-29.4	-25.6	-14.7	-24.4
1990	80.59	79.89	79.85	73.00	55.26	53.83	54.76	44.81	187.57	129.56	101.17	92.44	-55.2	-55.1	-51.9	-47.0
1995	82.81	70.89	67.65	64.36	50.23	43.95	40.63	36.13	128.82	58.88	34.36	92.60	-51.9	-42.4	-28.4	-41.3
2000	57.87	50.68	50.03	43.12	33.88	31.09	29.97	23.22	101.37	39.21	39.28	41.15	-32.4	-27.3	-16.5	-28.2
2005	33.08	31.56	32.71	26.39	31.57	28.14	31.22	19.15	47.79	29.49	35.87	21.42	-32.9	-32.3	-33.1	-23.5

Initialization year	MAE Change in %				FSSb15 Change in %				CC Change in %			
	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT
1960	-1.52	-6.94	-10.99	-10.99	25.88	10.03	17.39	-6.47	-10.19	-0.95	-3.02	-2.16
1965	-15.91	-13.32	-14.87	-14.87	10.74	2.15	10.93	-16.16	11.64	8.01	7.87	8.50
1970	-32.03	-31.96	-36.33	-36.33	39.26	27.97	25.93	11.64	13.96	13.12	13.46	16.62
1975	-29.11	-28.41	-24.66	-24.66	14.29	-4.19	25.06	-9.25	10.01	9.50	10.47	1.29
1980	-28.01	-25.56	-13.44	-13.44	73.07	20.75	61.80	43.45	38.53	24.12	37.05	12.05
1985	-39.00	-31.24	-23.02	-23.02	239.64	95.01	207.77	108.79	16.16	8.41	7.85	7.50
1990	-57.08	-57.41	-46.10	-46.10	107.90	41.65	105.59	46.48	17.16	13.83	16.97	9.79
1995	-54.53	-46.00	-43.04	-43.04	171.00	128.69	183.30	137.90	25.64	13.41	15.03	8.49
2000	-35.66	-32.17	-31.95	-31.95	89.79	41.51	81.17	50.02	17.88	9.45	12.76	0.20
2005	-37.19	-	-21.05	-21.05	17.99	23.45	51.73	19.52	20.81	15.83	20.36	13.91

Table 6-S2 Change of skills after drift correction of CanCM4. The positive values indicate increase in skills and vice versa

Initialization year	ACC Change in %				IA Change in %				FSSa85 Change in %				RMSE Change in %			
	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT
1960	150.24	236.10	199.84	257.26	980.01	1043.34	1048.61	651.10	88.23	95.51	36.78	111.57	11.4	-9.8	14.5	-16.6
1965	847.30	778.02	756.99	615.82	480.59	454.54	443.56	289.97	113.68	73.09	71.05	22.14	-41.9	-37.0	-10.1	-26.7
1970	246.40	239.79	210.86	238.56	4.10	0.00	-6.19	-25.06	99.82	57.90	47.04	-1.23	-32.8	-32.6	-1.8	-33.1
1975	252.41	239.53	236.79	189.22	474.10	448.21	452.51	269.23	63.36	33.16	38.69	26.38	-41.3	-39.6	-24.9	-25.7
1980	247.60	263.67	237.28	270.88	6248.58	6217.14	6084.87	4469.50	85.84	77.23	73.14	102.57	-3.8	-12.6	13.3	-20.1
1985	60.39	47.83	41.76	42.25	151.35	132.60	131.33	68.62	38.89	29.27	10.00	108.12	-18.7	-15.7	7.2	-14.3
1990	76.19	67.20	60.46	60.38	178.78	164.72	157.31	115.25	184.76	141.42	89.46	149.42	-39.5	-32.2	-8.6	-27.1
1995	75.52	75.86	68.58	67.36	100.94	98.09	93.00	60.12	155.69	105.65	47.59	159.36	-19.8	-23.57	0.73	-20.1
2000	58.76	56.55	52.63	45.59	109.68	102.24	104.39	55.67	98.63	96.72	61.28	117.30	-25.9	-27.35	-12.9	-18.3
2005	204.15	195.93	196.57	155.32	336.68	318.86	328.71	209.35	63.38	48.20	37.69	46.10	-51.44	-47.58	-41.9	-27.7

Initialization year	MAE Change in %				FSSb15 Change in %				CC Change in %			
	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT	NBC	MDM	STD	RDT
1960	6.90	-12.01	0.13	-12.67	81.31	48.49	87.48	31.10	-35.51	-4.25	-7.19	-4.88
1965	-38.99	-31.00	-20.95	-21.08	58.26	22.04	43.00	9.87	66.83	49.02	51.30	16.71
1970	-28.01	-22.42	-2.49	-23.15	157.36	130.19	148.13	59.62	13.04	5.79	0.51	20.15
1975	-42.13	-37.49	-32.04	-25.91	71.60	51.86	69.49	38.96	32.20	24.40	27.12	-0.29
1980	-10.64	-10.61	3.53	-13.43	334.25	296.46	471.72	336.68	6.69	9.90	8.68	1.54
1985	-22.90	-16.20	-7.42	-8.04	65.37	38.21	70.67	35.02	16.88	2.36	1.71	-0.01
1990	-41.23	-33.31	-22.10	-24.83	121.64	31.76	131.60	-0.58	14.59	4.60	3.10	1.84
1995	-16.06	-14.77	-0.93	-11.07	81.32	6.77	112.82	-74.74	10.77	10.44	9.43	4.04
2000	-36.22	-28.58	-21.59	-14.31	89.71	42.66	80.90	12.16	12.91	9.53	8.49	5.14
2005	-44.19	-38.63	-37.99	-20.00	42.66	27.81	49.41	46.46	35.28	30.81	31.77	15.70

CHAPTER 7

EVALUATION OF CMIP5 DECADAL PRECIPITATION AT CATCHMENT LEVEL

Abstract

Coupled Model Inter-comparison Project Phase-5 (CMIP5) performed the decadal experiments for a wide range of GCMs. Previous studies using CMIP5 decadal data were conducted mainly for temperature or temperature-based climate indices. Very few studies were conducted using CMIP5 decadal precipitation but none of them evaluated the performances of GCMs for CMIP5 decadal precipitation at a catchment level. Evaluation of CMIP5 decadal precipitation is an important step to perform reliable estimation for future water availabilities at a local level. This study evaluates the performances of CMIP5 decadal precipitation for eight selected GCMs (GCMs; MIROC4h, EC-EARTH, MRI-CGCM3, MPI-ESM-MR, MPI-ESM-LR, MIROC5, CMCC-CM, and CanCM4) for the Brisbane River catchment in Queensland, Australia. First, the dataset was subset for the entire Australia and then interpolated onto a finer resolution of $0.05^0 \times 0.05^0$ (5 km×5 km), using the second-order conservative method, matching with the grids of observed data. Secondly, the interpolated datasets are cut for the Brisbane River catchment. Next, models' outputs are evaluated for temporal skills, dry and wet periods, and total precipitation based on the observed values. Correlation coefficient (CC), anomaly correlation coefficient (ACC), and index of agreement (IA) are used to measure the temporal skills whereas fractional skill scores (FSS) are used to measure the dry and wet periods. To measure the model skills for total precipitation over the entire catchment, the field-sum and total-sum are used. These skills are measured at individual grids and for the entire catchment. Based on the skill scores, models are divided into three categories (Category-I: MIROC4h, EC-EARTH and MRI-CGCM3; Category-II: MPI-ESM-LR and MPI-ESM-MR; and Category-III: MIROC5, CanCM4, and CMCC-CM) and suggestions are made for the formation of suitable multi-model ensembles' mean (MMEM). Three MMEMs are formed using the arithmetic mean of models in Category-I (MMEM1), Category-I and II (MMEM2), and all eight models

This chapter has been submitted as: Hossain, M M, Anwar, A.H.M.F., Garg, N., Prakash, M., Bari, M., 2022. Evaluation of CMIP5 decadal precipitation at catchment level. *International Journal of Climatology* (Under review).

(MMEM3). The performance of all of these three MMEMs are also assessed using the same skill tests and found MMEM2 performed better.

Keywords: Performance, CMIP5, decadal, precipitation, catchment, multi-model ensemble

7.1 Introduction

Performance evaluation of General Circulation Models (GCMs) has become a very important task to measure the models' credibility on future prediction of climate variables. Evaluation of models' predicted historical data based on their corresponding observed values determines how well the GCMs represent historical climate and thus forms an integral part of the confidence-building exercise for climate predictions. It is assumed that the better performance of models over the historical period will lead to developing more confidence in their future predictions. As the GCMs are used to explore the future climate variabilities and potential impacts on the Earth, evaluation of GCMs has been a growing need in the climate research community. However, depending on the requirements, available resources, geographical locations, and variables considered to assess the model performances, the evaluation strategies become different. Since the change of climate and its potential impact on this planet varies from region to region, it is important to evaluate the models based on different regions and spatial scales though the evaluation of climate models and their ensembles is crucial in climate studies (Flato et al., 2013). Researches on regional or local climate variability and their potential impacts are high in demand for transferring research-based scientific knowledge to increase the resilience of the society to climate change. This will help in planning the future development of the infrastructures of a region (Kumar et al., 2013).

Coupled Model Inter-comparison Project Phase-5 (CMIP5) provides an unprecedented collection of global climate data of different time scales including decadal experiments which were produced by a wide range of GCMs (Taylor et al., 2012). Evaluation of CMIP5 decadal prediction has been run far from the early stage based on different evaluation aspects such as different regions, different climate variables, and their different spatial and temporal resolutions. For instance, Choi et al. (2016) investigated the prediction skill of CMIP5 decadal hindcast near-surface air temperature for the global scale while other researchers investigated other climate variables in continental or regional scales (Gaetani and Mohino, 2013; Lovino et al., 2018; McKellar et al., 2013). Lovino et al. (2018) evaluated decadal hindcast precipitation

and temperature over northern Argentina and reported higher skills of models to reproduce the temperature as opposed to precipitation where precipitation skills were found remarkably lower. McKellar et al. (2013) investigated decadal hindcast maximum and minimum temperature over the state of California and reported the best performing model. Likewise, Gaetani and Mohino (2013) evaluated model performances to reproduce Sahelian precipitation and reported better performing models. However, these studies were for different geographical locations with coarser spatial resolutions for considered variables. For instance, the spatial resolution of models used by Kumar et al. (2013) and Choi et al. (2016) was 2.5^0, Gaetani and Mohino (2013) used models of more than 1.1^0, and Lovino et al. (2018) used precipitation data of 1.0^0 spatial resolution. At a regional level, Mehrotra et al. (2014) assessed the multi-model decadal hindcast of precipitation for different hydrological regions over Australia using 0.5^0 spatial resolution and reported lower skills for precipitation as opposed to temperature and geopotential height. Climate data of 0.5^0 spatial resolution covers a ground area equivalent to a square of 50 km length over the Australian region. Comparatively, a 50 km × 50 km area is very big where climate variabilities are also large and frequency and magnitude of precipitation vary in a few kilometres (such as in Australia). As the precipitation shows more spatial and temporal variability than temperature and the model performances vary region to region, therefore the model performances at the local level for finer spatial resolution is essential for precipitation.

Numerous studies evaluated CMIP5 models over Australia (Bhend and Whetton, 2015; Choudhury et al., 2019; Flato et al., 2013; Mehrotra et al., 2014; Moise et al., 2015) but study on evaluating CMIP5 decadal precipitation at catchment scale can hardly be found. After Mehrotra et al. (2014), who assessed the CMIP5 decadal hindcast precipitation over different hydrological regions ($0.5^0 \times .0.5^0$) in Australia, recently Hossain et al. (2021a, 2021b) used the CMIP5 decadal precipitation at a further finer resolution of $0.05^0 \times 0.05^0$ (5km×5km) for Brisbane catchment Australia for the first time. Hossain et al. (2021a, 2021b) compared the model performances for investigating the model drift and their subsequent correction using alternative drift correction methods for both the monthly and seasonal mean precipitation. However, they compared the model performances at a single grid point within the Brisbane River catchment. On the contrary, Mehrotra et al. (2014) used only a multi-model approach but did not consider individual models finer than 0.5^0. Local climate variables of finer temporal and spatial resolution, especially for precipitation, are very important for water managers for

planning and developing infrastructures as well as decision making for local businesses and agricultures. To maintain sustainable development with effective future planning based on the models' projected precipitation, it is important to evaluate the performance of the CMIP5 models' hindcasts precipitation.

Many researchers have suggested using MMEM (Choudhury et al., 2016; Islam et al., 2014; Knutti et al., 2010; McSweeney et al., 2015) while using GCMs data to reduce the model biases. The use of MMEM may enhance the model performances (Kumar et al., 2013; Sheffield et al., 2013) by reducing the biases to some extent but there is no information available on the ranking of GCM models and based on this, which and how many models should be considered to produce MMEM so that it could provide better outcome. This is more essential for CMIP5 decadal precipitation because of its wide range in spatial and temporal variability in providing the model output ten years ahead. That is why the objective of this paper is, first, to categorize the models based on their performances at catchment level with a spatial resolution of 0.05^0 and next, to identify the best combination of different models that would provide better performance. This would help the water managers and policymakers to sort out models depending on their specific needs while assessing the future water availability based on the GCMs derived precipitation on a decadal scale through CMIP5.

7.2 Data collection and processing

7.2.1 Data collection

CMIP5 decadal experiment provides 10 and 30-year long ensemble predictions from multiple modelling groups [henceforth mentioned as CMIP5 decadal hindcasts, (Meehl and Teng, 2014)]. Monthly decadal hindcasts precipitation from eight GCMs (out of ten); MIROC4h, EC-EARTH, MRI-CGCM3, MPI-ESM-MR, MPI-ESM-LR, MIROC5, CMCC-CM, and CanCM4 for which decadal hindcast precipitation are downloaded from CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). The other two models, HadCM3 (spatial resolution 3.75° x 2.5°) and IPSL-CM5A-LR (spatial resolution 3.75° x 1.89°) were not considered in this study because of their relatively coarser spatial resolution and different calendar system (HadCM3). For the initialized period 1960-2005, data simulated over 10 years that are initialized every five years during this period are selected for this study as they were found comparatively better than

the 30-year simulation (Hossain et al., 2021d). The details of the selected models are given in Table 7-1.

The observed gridded monthly precipitation of 0.050 ×.0.050 (≈ 5km × 5km) was collected from the Australian Bureau of Meteorology (BoM). This data was produced using the Australian Water Resources Assessment Landscape model (AWRA-L V5) (Frost et al., 2016).

Table 7-1 Selected models with the initialization year 1960-2005

Model name (Modelling center or group) Resolutions:°lon × °lat	Initialization Year (1960-2005)									
	60	65	70	75	80	85	90	95	00	05
	Number of ensembles									
EC-EARTH (EC-EARTH Consortium) 1.125 X 1.1215	14	14	14	14	14	14	14	14	10	18
MRI-CGCM3 (Meteorological Research Institute) 1.125 X 1.1215	06	08	09	09	06	09	09	09	09	06
MPI-ESM-LR (Max Planck Institute for Meteorology) 1.875 X 1.865	10	10	10	10	10	10	10	10	10	10
MPI-ESM-MR (Max Planck Institute for Meteorology) 1.875 X 1.865	03	03	03	03	03	03	03	03	03	03
MIROC4h (AORI-Tokyo University, NIES and JAMEST)* 0.5625 X 0.5616	03	03	03	06	06	06	06	06	06	06
MIROC5 (AORI-Tokyo University, NIES and JAMEST)* 1.4062 X 1.4007	06	06	06	06	04	06	06	06	06	06
CanCM4 (Canadian Centre for Climate Modelling and Analysis) 2.8125 X 2.7905	20	20	20	20	20	20	20	20	20	20
CMCC-CM (Centro Euro-Mediterraneo per I Cambiamenti Climatici) 0.75 X 0.748	03	03	03	03	03	03	03	03	03	03

*Atmosphere and Ocean Research Institute-The University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology)

7.2.2 Data processing

The GCMs' resolutions (100-250 km) are found inadequate for regional studies due to lack of information at catchment levels (Fowler et al., 2007; Grotch and MacCracken, 1991; Salathé, 2003). The regional climate model (RCM) is useful to transfer the coarse spatial GCMs' data to local scale but it needs a wide range of climate variables as well as rigorous efforts to develop. For this reason, GCMs data are spatially interpolated onto $0.05^0 \times 0.05^0$ spatial resolution using the second-order conservative (SOC) method matching with the grids of observed data. For the gridded precipitation data, the SOC method was found comparatively better than other commonly used spatial interpolation methods (Hossain et al., 2021a). Skelly and Henderson-Sellers (1996) suggested GCM derive gridded precipitation to consider as areal quantities and spatial interpolation will not create any new information except the spatial precision of the data. Skelly and Henderson-Sellers (1996) also suggested that researchers could sub-divide the grid box in almost any manner until the original volume remains the same. On the contrary, Jones (1999) suggested that precipitation flux must be remapped in a conservative manner to maintain the water budget of the coupled climate system. While sub-gridding the GCM data using the SOC method, it conserves precipitation flux from their native grids to subsequent grids (Jones, 1999). For this reason, this study used the SOC method for spatial interpolation as it was followed in other research (Hossain et al., 2021c). Brisbane catchment was selected for this study because of its tropical climate that produces low to moderate variability of annual precipitation values.

7.3 Evaluation methodology

A simple and direct approach for the model evaluation is to compare the model output with the observations and analyze the differences. In this study, models are evaluated for temporal skills, dry and wet periods, and total precipitation based on the observed values. Here, CC, ACC, and IA are used to measure the temporal skills, FSS are used to measure the skills over dry and wet periods, field-sum and total-sum are used to measure the skills for total precipitation. There are 496 grids in the Brisbane River catchment with a spatial resolution of 5.0 km × 5.0 km. The descriptions of the skills are given below.

7.3.1 Correlation Coefficient (CC)

CC measures the linear association and presents the scale of temporal agreement between predicted and observed values. Statistically, it measures how much closer the scatter plot points to a straight line. CC ranges from -1 to 1 for no to perfect correlation respectively.

$$CC = \frac{\sum(F_i-\overline{F})(O_i-\overline{O})}{\sqrt{\sum(F_i-\overline{F})^2}\sqrt{\sum(O_i-\overline{O})^2}} \qquad (7.1)$$

Here, F and $\bar{F}$, represent models' predicted and their mean whereas O and $\bar{O}$ represent observed precipitation and their mean respectively. In the following skill tests, these notations will remain the same. Note that the mean is calculated for every individual year.

7.3.2 Anomaly Correlation Coefficient (ACC)

ACC was suggested by Wilks, (2011) to measure the temporal correlation between anomalies of the observed and predicted values. For the verification of numerical weather models' prediction ACC is frequently used. Its value ranges from -1 to 1 for no to perfect anomaly matching.

$$ACC = \frac{\sum\{(F_i-C)-(\overline{F-C})\}\times\{(O_i-C)-(\overline{O-C})\}}{\sqrt{\sum(F_i-C)^2}\sqrt{\sum(O_i-C)^2}} \qquad (7.2)$$

Here, C represents the mean of the entire time-span (ten years) of the observed (BoM) data. The higher value of ACC will indicate the higher performance in reproducing the monthly anomalies.

7.3.3 Index of agreement (IA)

Wilmot, (1982) suggested IA to measure the accuracy of predictions. The index of agreement can be calculated as follows.

$$IA = 1 - \frac{\sum_{i=1}^{n}(F_i-O_i)^2}{\sum_{i=1}^{n}(|F_i-\bar{O}|+|O_i-\bar{O}|)^2} \qquad (7.3)$$

The index is bounded between 0 and 1 ($0 \leq IA \leq 1$). The value closer to 1 indicates the most efficient predicting of the models.

7.3.4 Fractional Skill Score (FSS)

FSS is a grid-box event that directly compares the fractional coverage of models' predicted and observed values for the entire catchment. It measures how the spatial variability of models' predicted values corresponds to the spatial variability of the observed values. FSS can be obtained as:

$$FSS = 1 - \frac{\frac{1}{N}\sum_N \left(P_{f,m} - P_{f,o}\right)^2}{\frac{1}{N}\left[\sum_N P_{f,m}^2 + \sum_N P_{f,o}^2\right]} \qquad (7.4)$$

Where P_f and N refers to calculated fraction and number of years respectively. The subscript *m* and *o* present modelled and observed respectively. In this study, fractions are calculated according to Roberts and Lean (2008) but considered entire catchment as a whole unit, and the temporal averages (for considered months) are taken instead of the spatial averages. For doing this, threshold values; ≥85 percentile for the months of wet seasons (December to February - DJF) and <15 percentile for the months of the dry seasons (June to August - JJA) are considered. To get the fractions (say for January), the number of grid points covered for a specified threshold value is counted and then divided by the total number of grids within the catchment. The differences between predicted and observed fractions (the numerators of equation 4) are calculated for individual months. The FSS will be a temporal average score for the catchment for each considered month. It ranges from 0 to 1 for no to perfect match respectively.

7.3.5 Field-sum and total-sum

Models' ability to reproduce the total precipitation over the entire catchment is considered as the spatial skills of the models. Field-sum is the sum of precipitation over the entire catchment for individual time steps and the total-sum is the field-sum over the total time span. Field-sum and total-sum of the models' precipitation are compared with the corresponding observed values.

7.4 Results and analysis

7.4.1 Evaluation for temporal skills

The temporal skills are computed at every individual grid (total 496 grids) of the catchment for all initialization years of each model. Spatial variations of model temporal skills across the catchment for the initialization year 1990 (1991-2000) are presented in Fig. 7-1. The models are evaluated from the spatial perspective by counting the number of grids covered by different models for different threshold values of CC, ACC, and IA (Fig. 7-2). The higher number of grids represents the higher spatial skill of models across the catchment. From the comparison of temporal skills, it is evident that model performance varies over the initialization years and also across the catchment. From the initialization year 1990 and onward, all models show a comparatively higher number of grids for the same thresholds of CC, ACC, and IA and the lowest skill observed in 1980. With the increase of threshold values, the number of grids declines for all models in all three temporal skills except CMCC-CM in ACC as it shows no grid in all three threshold of ACC presented in Fig. 7-2. Compared to other selected models, MIROC4h, EC-EARTH, and MRI-CGCM3 show a higher number of grids for all thresholds in which MIROC4h is much ahead of EC-EARTH and MRI-CGCM3. It means temporal agreement, the resemblance of anomalies, and the prediction accuracy of MIROC4h and EC-EARTH spatially higher than other models. This study also checked the number of grids for the threshold >=0.6 for CC and ACC but no model could reproduce CC and ACC >=0.6 at any grid. However, MIROC4h, EC-EARTH, and MRI-CGCM3 show a significant number of grids for the IA threshold >=0.6 where MIROC4h outperformed EC-EARTH and MRI-CGCM3 (Fig. 7-2). Comparing the models, MIROC4h shows higher temporal skills from the spatial perspective, followed by EC-EARTH and MRI-CGCM3 whilst MPI-ESM-MR, MIROC5, and CMCC-CM show low to lowest temporal skills. Over the catchment MIROC5, MPI-ESM-MR, CanCM4 show little better scores than CMCC-CM.

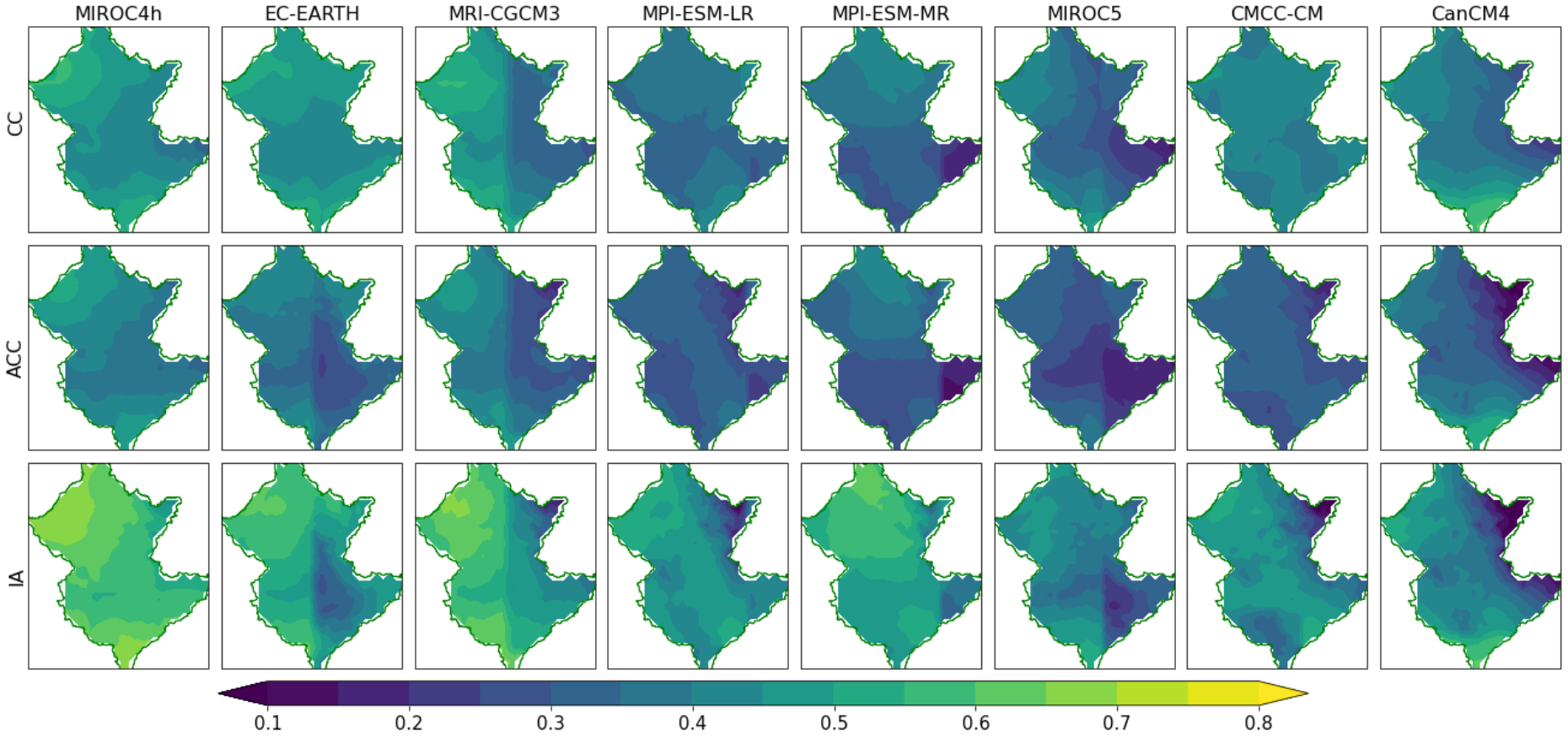


Fig. 7-1 Spatial variations of temporal skills (CC, ACC, and IA) of the models initialized in 1990 (period; 1991-2000) over the Brisbane River catchment

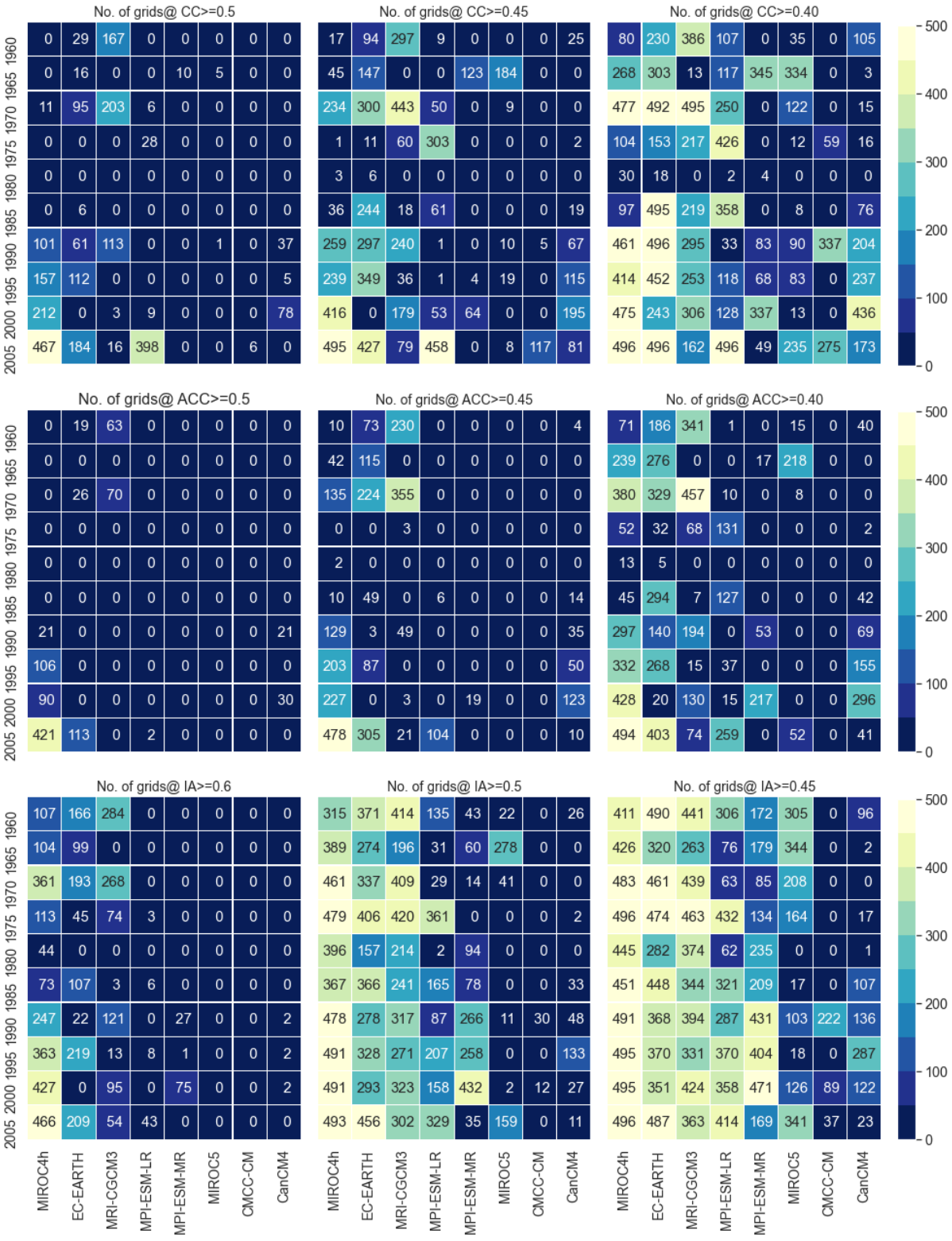


Fig. 7-2 Number of grids covered by different models for different thresholds of CC, ACC, and IA. The vertical axis presents the initialization years and the horizontal axis presents the model name. Threshold values are provided on the top of each subplot

7.4.2 Evaluation for dry and wet periods

Skills to reproduce the dry and wet events are assessed at the selected grid and also over the entire catchment. For the selected grid all months are considered against four different thresholds (25, 50, 75, and 90 percentiles correspond to 25, 60, 110, and 175 mm respectively) whereas for the entire catchment, FSS are used for the months of dry (JJA) and wet (DJF) periods only.

At the selected grid

A comparison to reproduce the dry and wet events based on the selected precipitation thresholds at the selected grid is presented in Fig. 7-3. This comparison was based on the ratio of the number of months of respective precipitation thresholds (mentioned on the top of the individual plot in Fig. 7-3) in model data to observed data. It is observed that EC-EARTH and MIROC5 could reproduce no dry events (Pr<=25mm) whilst CMCC-CM overestimates the number of dry events which is almost double the dry events in observed data. Meanwhile, MIROC4h performed better to produce dry events as well as 50 and 75 percentile values as compared with the other models. However, MIROC4h is a little behind the MPI-ESM-MR for the extreme wet events (Pr>=175mm). It means MPI-ESM-MR can reproduce extreme wet events better than the other models. EC-EARTH, MPI-ESM-LR, and MPI-ESM-MR underestimated the events of threshold Pr<=60mm whereas overestimated the events of threshold Pr>=110mm which is an indication of models' tendency to reproduce a higher number of wet events than opposed to dry. However, MRI-CGCM3 performed similarly to MIROC4h in reproducing the number of events for the threshold of <= 60mm but underestimated the number of events thresholds of >= 110mm. To reproduce the extreme wet events (Pr>=175mm), all models show underestimation in which MPI-ESM-MR and MIROC4h showed considerably higher skills. The CMCC-CM and CanCM4 showed poorest, and no skill respectively for extreme wet events.

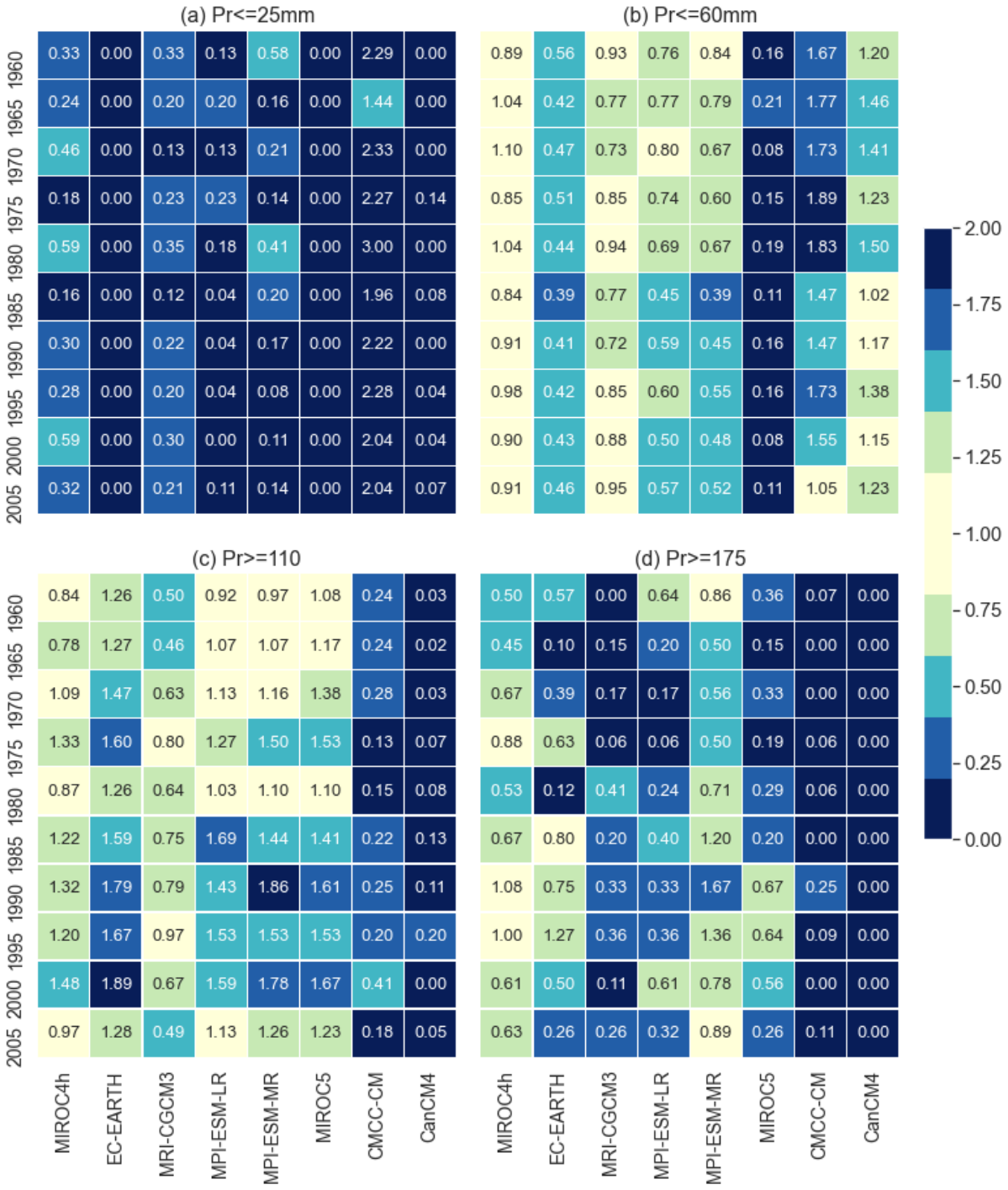


Fig. 7-3 Comparison of model skills to reproduce dry and wet events at a selected grid point. Values 1.0 present perfect matching whilst values below and above 1.0 present under and over prediction respectively

Over the entire catchment

FSSs are calculated for the months of winter (dry) and summer (wet) seasons only. FSS of all the initialization year of all models are shown in Fig. 7-4.

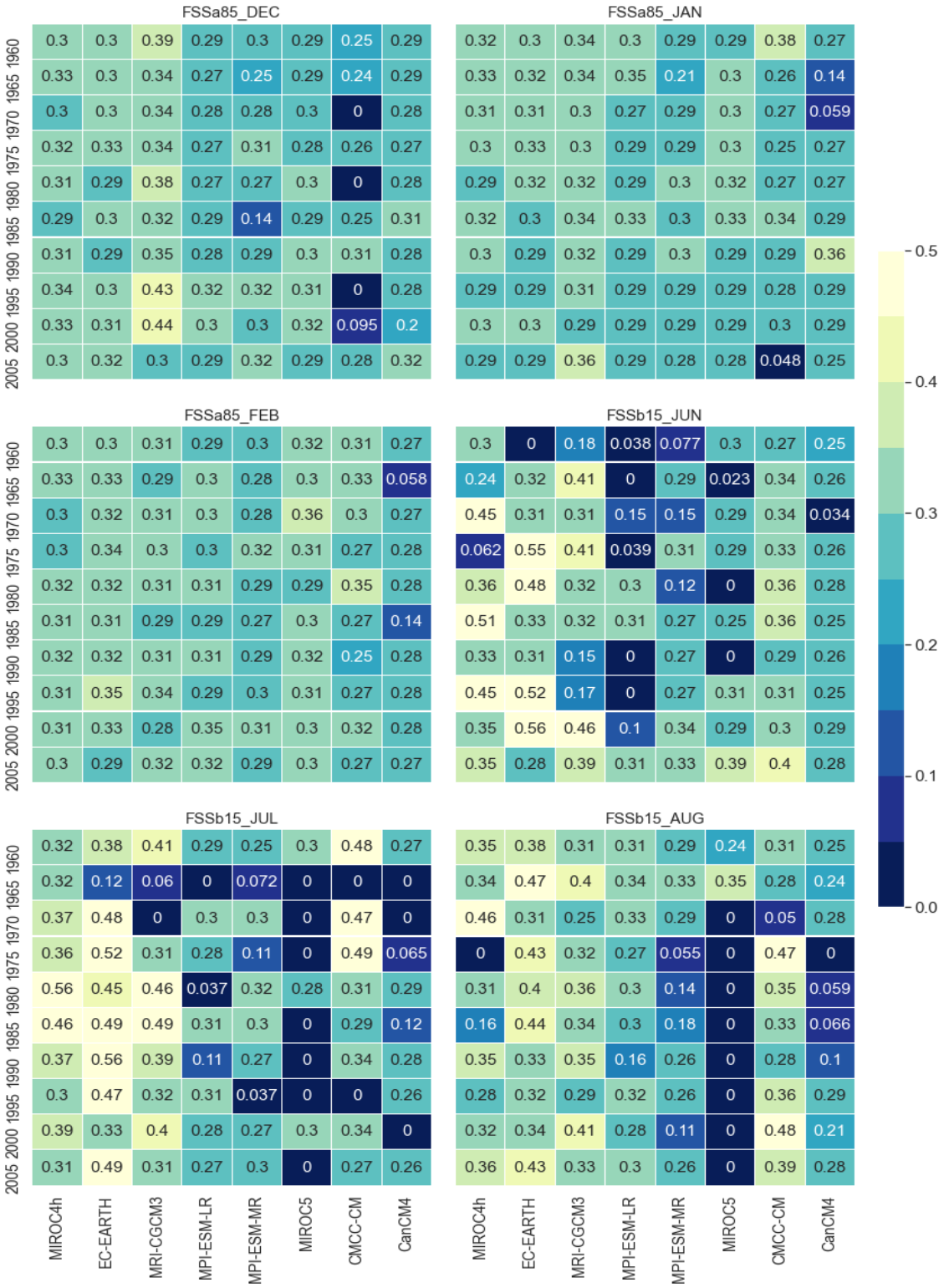


Fig. 7-4 Fractional skill score for the months of winter and summer seasons

Results show that for the months of summer seasons (DJF), MRI-CGCM3 shows higher skills in December and January but little behind than EC-EARTH in February. On the contrary,

CMCC-CM shows the lowest skill in December but shows similar skill with other models in January and February. However, except higher skill of MRI-CGCM3 and the lowest skill of CMCC-CM in December, all other models show similar skill scores with few variations in winter seasons. This indicates different models' skills are almost similar to reproducing wet events. In the dry season, MIROC5 shows the lowest skill while EC-EARTH shows the higher skills, which is even higher than MIROC4h and MRI-CGCM3. The FSSb15 scores of EC-EARTH, MIROC4h, and MRI-CGCM3 are much better than the score obtained for FSSa85. This reveals that these models are better to reproduce dry events as opposed to wet events and the reverse is true for MIROC5, MPI-ESM-MR, and CanCM4 respectively.

7.4.3 Evaluation for total precipitation

At the selected grid

To evaluate the model performances in reproducing the total precipitation, models' cumulative (over time) precipitation at several randomly selected grids (evenly distributed across the catchment) within the catchments and total precipitation over the entire catchment are compared. The cumulative sum of monthly precipitation of different models at the selected grid for different initialization years is presented in Fig. 7-5. The model skills show both temporal and spatial variations in predicting accumulated precipitation but no model could reproduce the accumulated precipitation as observed. However, only a few models (MIROC4h, MPI-ESM-LR, and MPI-ESM-MR) could reproduce the accumulated precipitation close to the observed accumulation. Nevertheless, CMCC-CM, CanCM4, and MRI-CGCM3 underestimated the accumulated precipitation whilst EC-EARTH and MIROC5 overestimated the accumulated values. With the change of grid locations, model performances may change but the relative performances among the models will remain the same.

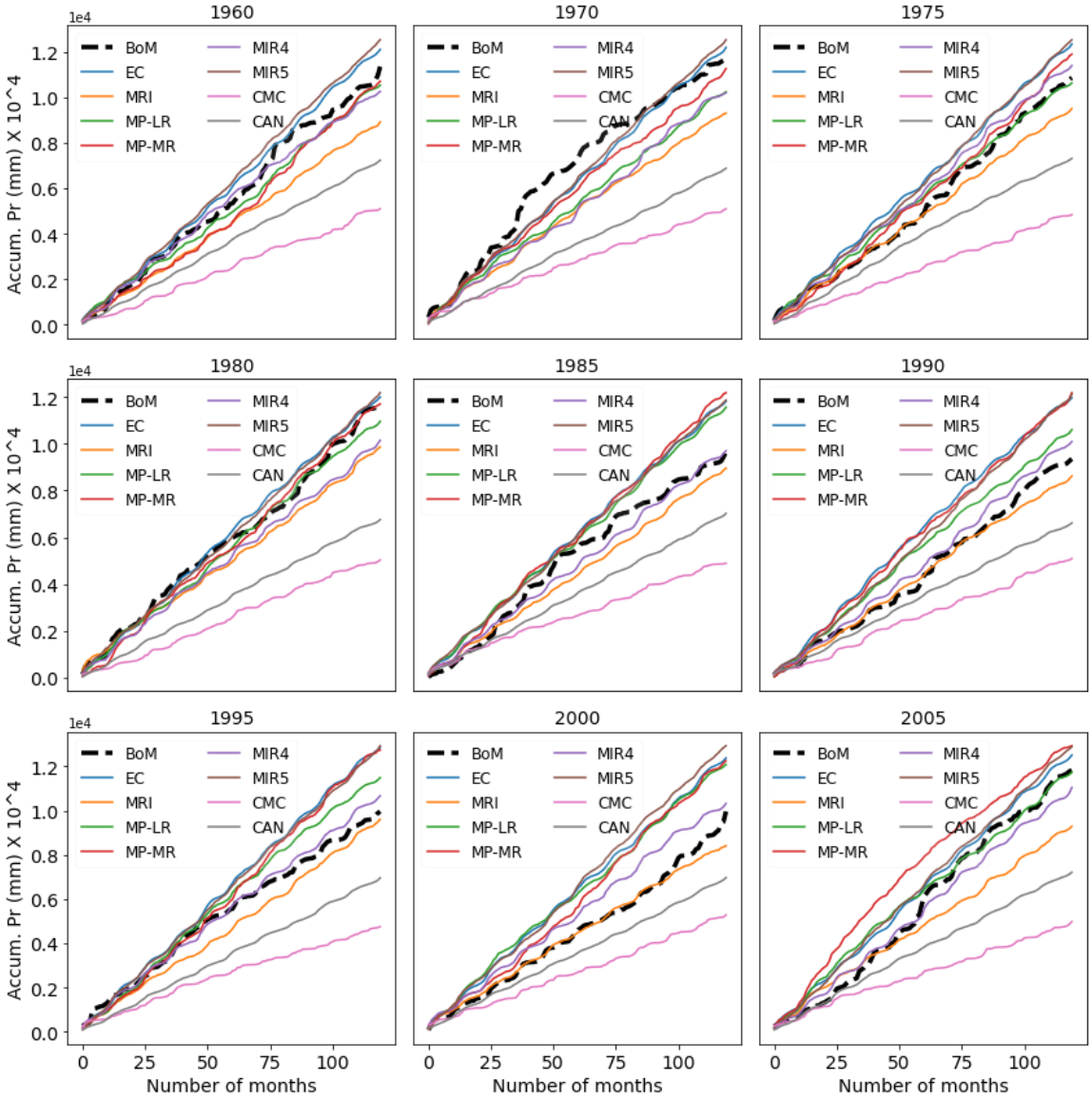


Fig. 7-5 Cumulative sum of monthly precipitation of different models at the selected grid point in different initialization years. The vertical axis presents accumulated precipitation and the horizontal axis presents the number of months over the decade.

Over the catchment

For comparing the model performances on total precipitation over the entire catchment, this study calculated the field-sum of the models and observed values then assessed through the temporal skills as shown in Fig. 7-6. The total sum of the models and observed values are also calculated and assessed through the ratio between model and observed values (Fig. 7-6). From the comparison, it is observed that the field sum of MIROC4h, EC-EARTH, and MRI-CGCM3 show comparatively higher accuracy (IA), temporal agreement (CC), and the resemblance of

anomalies (ACC) with the field-sum of the observed precipitation. The model performances on reproducing the total precipitation vary over the initialization years (Fig. 7-6.d). Before and after 1985, MRI-CGCM3 and MPI-ESM-MR showed comparatively better resemblance with the observed total precipitation followed by MIROC4h and EC-EARTH. On the contrary, CMCC-CM showed the lowest performance to reproduce total-sum precipitation throughout all initialization years.

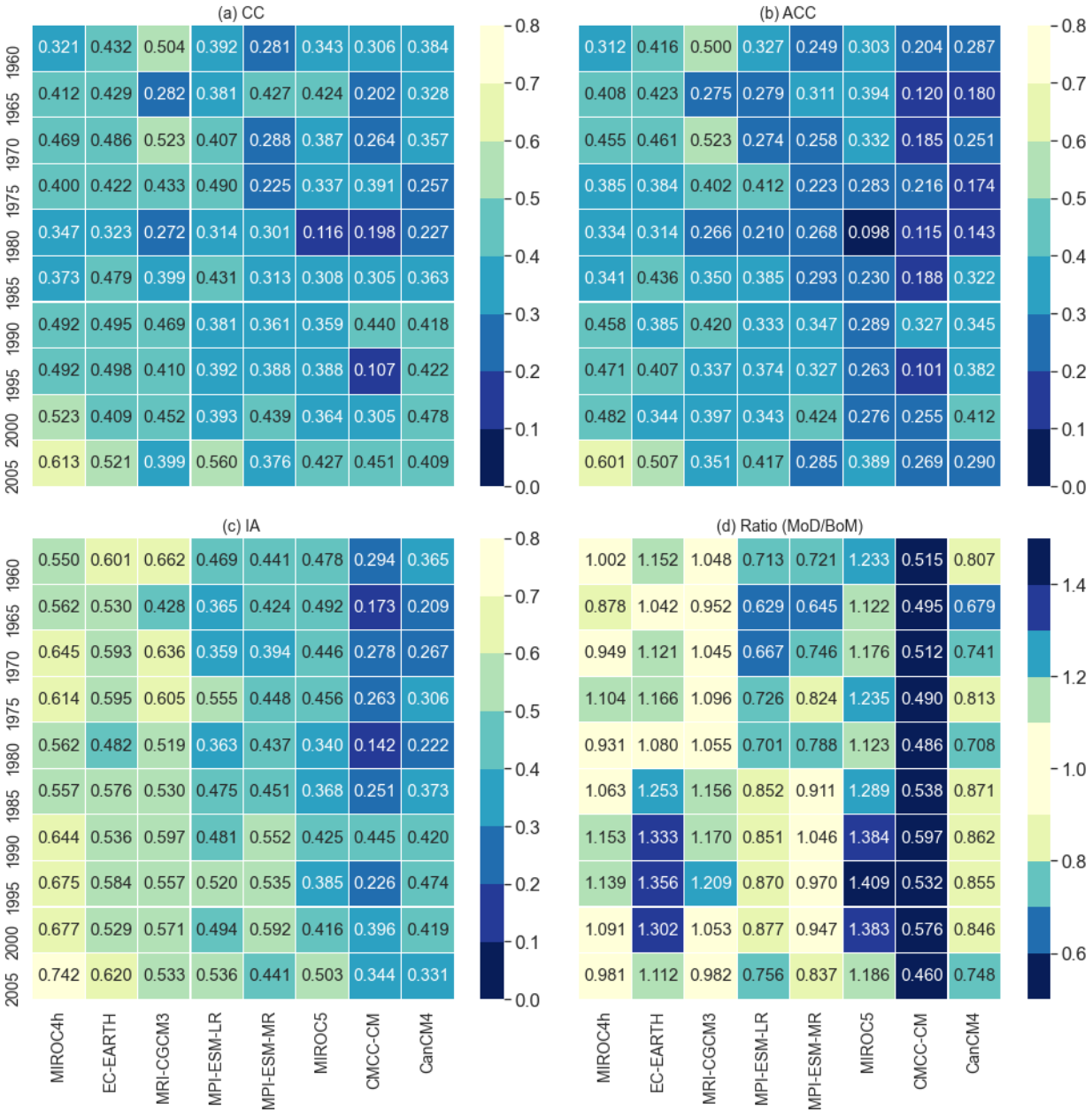


Fig. 7-6 Performance indicators of the models to reproduce the total precipitation of the entire catchment

From the skill assessments, it is revealed that the MIROC4h surpasses other models in almost all performance indicators followed by EC-EARTH and MRI-CGCM3 whilst MPI-ESM-LR and MPI-ESM-MR show medium skill scores. Lower skill scores were observed for MIROC5, CanCM4, and CMCC-CM respectively. MIROC4h was also marked as the best model to reproduce precipitation in other studies (Jain et al., 2019; Lovino et al., 2018) though they did not use the decadal experiments data. It may be due to the finer resolution of the atmospheric component of MIROC4h that enhanced its ability to capture the more realistic climate features (Jain et al., 2019; Sakamoto et al., 2012) at the local level.

The overall skill assessment results revealed that all models show the lowest skill in all performance indicators for the initialization year 1980 but the highest performance was noticed for the initialization year 1990. However, all models show comparatively better skills from the initialization year 1990 and onward as compared to 1960-1985.

7.4.4 Model categorisation and formulation of MMEM

Based on the skill comparisons, this study divided the models into three different categories; Category-I, Category-II, and Category-III. While categorizing the models based on their skills at the selected grid and over the catchment, MIROC4h, EC-EARTH, and MRI-CGCM3 fall in the first category (Category-I) as they consistently performed in the top three and their performance metrics were found very close to each other. Similarly, MPI-ESM-LR and MPI-ESM-MR are in the second (Category-II) category as they have shown medium skill scores in all skill tests over the initialization years. Lastly, MIROC5, CanCM4, and CMCC-CM fall in Category-III.

GCMs' outputs indeed contain uncertainties and biases which will cause the lower skill score but multi-model ensembles mean (MMEM) may enhance the models' skills (Hossain et al., 2021c; Islam et al., 2014; Kumar et al., 2013; Sheffield et al., 2013) by reducing uncertainties (Hossain et al., 2021c; Islam et al., 2014; Knutti et al., 2010; McSweeney et al., 2015). In this study, the skill tests are employed on the ensembles' mean of individual models' interpolated raw values only. Here the arithmetic mean of multiple models has referred to as MMEM. The performances of MMEMs were also assessed based on the similar skill tests that are employed on individual models and the results are summarised below. To form the MMEMs, three different combinations are considered. The arithmetic mean of Category-I models is referred

to as first MMEM (MMEM1), the arithmetic mean of the Category-I and Category-II models is referred to as the second MMEM (MMEM2) and finally arithmetic mean of all models is referred to as the third MMEM (MMEM3).

7.4.5 Performance of MMEMs

The temporal skills at individual grids of the different thresholds, temporal skills along with the ratios of the field-sum, and skill on reproducing dry and wet events of different thresholds for MMEMs are presented in Fig. 7-7, 7-8 and 7-9 respectively. In general, MMEMs show better performance than the individual models for comparatively lower thresholds of the performance metrics. For instance, the MIROC4h model showed the highest number of grids for CC and ACC at the threshold 0.5 (Fig. 7-2) but no MMEMs could reproduce this number of grids at the same threshold (Fig. 7-7). The same results were also observed for IA at the threshold 0.6 (see Fig. 7-7i) but for the lower thresholds, MMEM2 shows better skill than MIROC4h in CC and ACC but not in IA. Among the three combinations, MMEM2 surpasses the other two combinations in reproducing CC and ACC. Nevertheless, in the case of IA, MMEM2 is little behind than MMEM1.

Similar results are evident for performance indicators obtained from the field-sum of MMEM and the observed values (Fig. 7-8), where MMEM2 shows best for the CC and ACC but both MMEM2 and MMEM1 show similar skills for IA. However, to produce the dry and wet events, MMEMs show lower performance as compared to individual models. For instance, MIROC4h, MRI-CGCM3, and MPI-ESM-MR captured some dry events (Pr<=25mm) at the selected grid points (Fig. 7-3) but no combination could capture it (Fig. 7-9) whilst for the wet events, MMEM shows very poor skills.

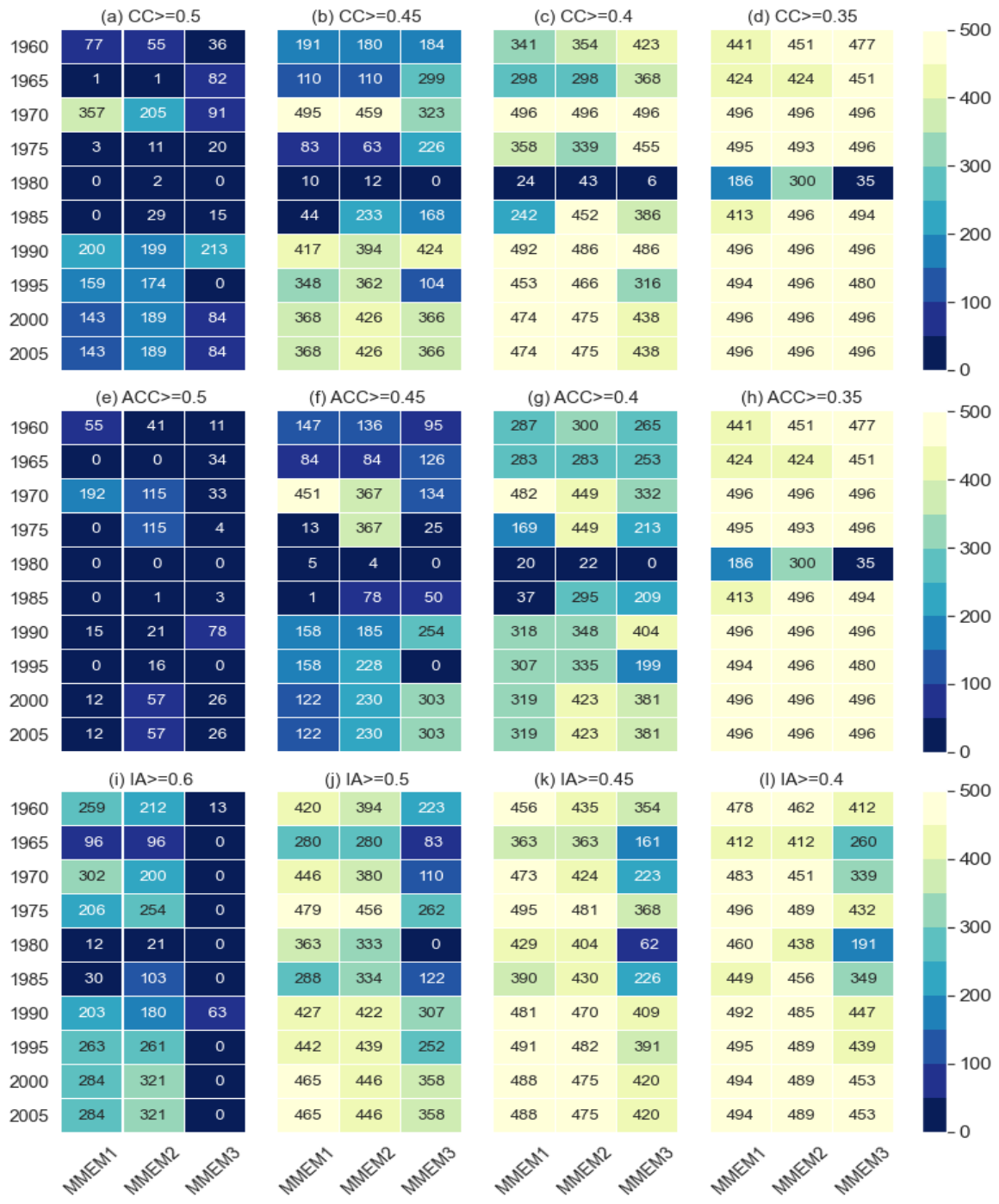


Fig. 7-7 Number of grids covered by different combinations of models for different threshold values of performance metrics. Thresholds and the performance indicators are mentioned on the top of the individual blocks

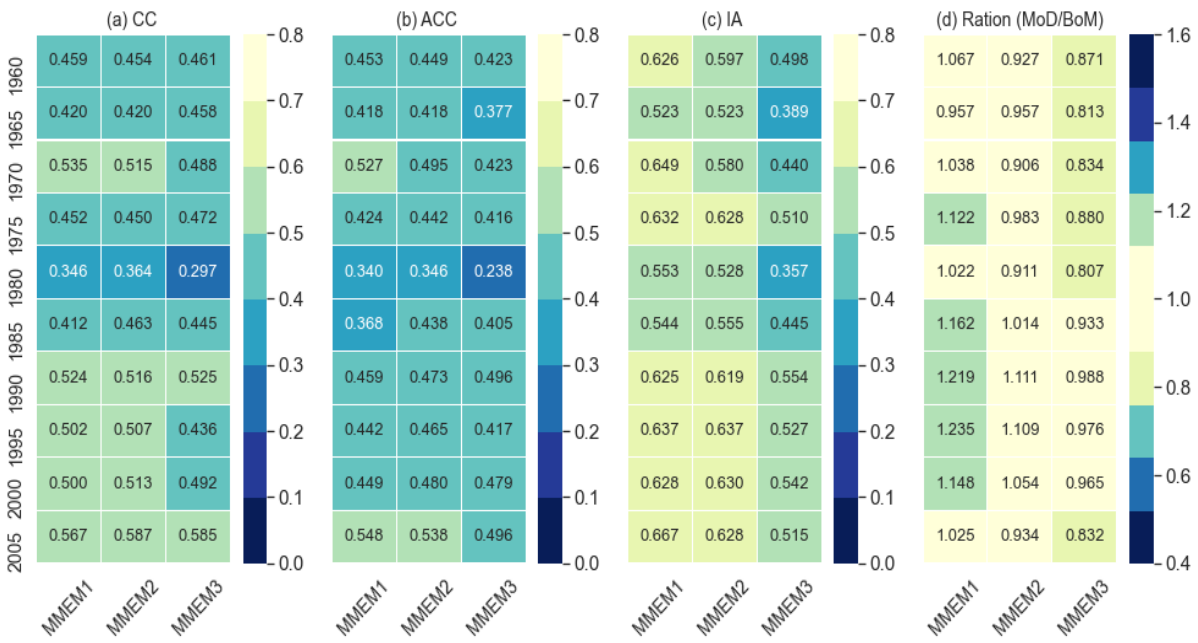


Fig. 7-8 Performance indicators obtained from the field-sum of different MMEMs and corresponding observed values.

Meanwhile, MMEMs show better performance indicators (CC, ACC, and IA) for the total precipitation of the entire catchment (field-sum) which is even better than the individual models. Nevertheless, MMEM is a little behind the MIROC4h and MRI-CGCM3 for the ratio of total-sum (sum over total time span and catchment) model combinations over the corresponding observed values (see Fig. 7-8. d)

Fig. 7-9 Skill comparison of three MMEMs to reproduce dry and wet events at the selected grid point. This comparison was based on the ratio, obtained from the number of months of respective precipitation thresholds (mentioned on the top of the individual plot) in model data to the number of months of observed values for different initialization years (Y-axis)

7.5 Discussion

This study evaluated the performance of eight selected GCMs simulation of CMIP5 decadal precipitation at a catchment level of 0.05-degree spatial resolution. Different skill metrics were employed from both temporal and spatial perspectives in this evaluation assessment. The performance metrics; CC, ACC, and IA measured the temporal skills of the models. The number of grids corresponding to individual metrics' thresholds represents the spatial skills of the models. These metrics are also calculated for the spatial sum (sum over the entire catchment) of the precipitation for all models. In addition to these, FSSa85 and FSSb15 presented the spatial skill of the models for wet and dry seasons respectively. The CC and ACC measured the phase and correspondence (or anomalies) of the model time series concerning the observed values. The models showed a wide range of performance scores over the initialization years as well as across the catchments. It may be due to the difference in understanding of models on local climate features or the precipitation data of finer temporal and spatial resolutions or the combination of both.

Indeed, the model performances are dependent on the model assumptions or basic principle on understanding the earth climate system, its processes, and interactions among atmosphere,

oceans, land, and ice-covered regions of the planet. Besides them, decadal prediction skill also depends on the method of model initialization, and quality and coverage of the ocean observations (Taylor et al., 2012). Different initializations also may cause models' internal variability that is still open for further discussion. For the decadal prediction, one of the most important aspects is the model drift and its correction (Mehrotra et al., 2014). However, to evaluate the performances of models' derived raw data, neither the drifts were investigated nor any drift correction methods are employed. The reason is, the drift correction method itself may introduce additional errors that may not reflect the real performance of the models (Hossain et al., 2021c, 2021b). Based on the understanding of physical, chemical, and biological mechanisms of earth systems, different modelling groups have come up with different models with reproducing capabilities of climate variables that may vary over different regions (Choi et al., 2016; Homsi et al., 2020; Purwaningsih and Hidayat, 2016) and climate variables (Kamworapan and Surussavadee, 2019; Kumar et al., 2014, 2013). For instance, Kumar et al. (2013) analyzed the precipitation and temperature trends of the twentieth century from nineteen CMIP5 models and reported that the models' relative performances are better for temperature as opposed to precipitation trends. Generally, models show lower skill to simulate precipitation than they do for temperature. This is because that the temperature is obtained from a thermodynamic balance, while precipitation results are from simplified parameterizations approximating actual processes (Flato et al. 2013; see also references therein). In addition, temporal and spatial scale (considered area) of the considered variables including seasons of the year (Sheffield et al., 2013; Ta et al., 2018) may also be the reason to vary the model performances. For instance, few models can reproduce winter precipitation very well but the other may not and vice versa. Likewise, Lovino et al. (2018) evaluated performances of CMIP5 model for the decadal simulation and reported the best models for the different climate variables (precipitation and temperatures). They also suggested that the MMEM could reproduce large-scale features very well but fail to replicate the smaller scale spatial variability of the observed annual precipitation pattern. These show clear evidence that there is a spatial variation in the climate model performances across the globe as they are developed by different organizations (Chen et al., 2017). This study noticed the highest skill in the initialization year of 1990 and the lowest skill in the initialization year 1980, but the reason behind the highest and lowest skill remains unknown. However, Meehl et al. (2015) reported that the consequences of Fuego (in 1974) and Pinatubo (1991) eruption degraded the decadal

hindcasts skill of Pacific sea surface temperature in the mid-1970s to mid-1990s respectively. As Fuego was smaller than Mount Pinatubo and a lower degrade of skill in the mid-1970s and higher degrade of skill in the mid-1990s were evident but no degrade on the hindcast skill was evident due to Agung (erupted in 1963) and El Chichón (1982) (Meehl et al., 2015). In this study, models' higher and lower skills of initialization 1990s and 1980s, seem neither relevant to volcanic eruption nor the post-eruption sequences. Nevertheless, the observed precipitation or coverage of the ocean observed state to initialize the models have been affected.

The CC and ACC values of all the selected models in all initialization years remained under the threshold>= 0.6, which was marked as the threshold of significant level in previous studies (Choi et al., 2016; Lovino et al., 2018) though those studies were for coarser spatial resolutions and one of them for different climate variables. Lovino et al. (2018) compared CMIP5 model performances over two variables at the local level and reported higher skill scores for the temperature than precipitation of the same models where the skill scores for precipitation were remarkably lower than the scores for temperature. Similar results were also reported by Jain et al. (2019). In this sense, it seems precipitation data with higher spatial resolution may be the reason for not capturing the significant level of skills on linear association (CC) and phase differences or anomalies (ACC). However, few models show that the level of significance (threshold>=0.6 if we say) for the performance metric IA, which is a measure of the predicting accuracy that seems promising predictive skill of the models. But the studies that mentioned 0.6 as the level of significance for CC and ACC, used either coarser resolution data (Lovino et al., 2018) or different climate variables (Choi et al., 2016). For the local or regional level as well as models' raw precipitation data of higher spatial and temporal resolution, 0.50 seems a significant score, which is also the same for the similar performance metrics for the case of total precipitation.

This study also investigated the model performances to reproduce the summer and winter precipitation. Upon comparing the model skills to reproduce the extreme wet (>=85 percentile of the observed values) and dry events (<15 percentile of the observed values) across the catchment and also at the selected grid, this study reveals that except CMCC-CM, all models show almost similar skills to reproduce the summer precipitation but exhibits some variations to produce the winter precipitation. Similar skills are also noted for other intermediate thresholds. It is due to the maximum and minimum precipitation occurring in Brisbane during

summer and winter respectively. This means that models' responses to reproduce summer precipitation are better than the winter with the tendency to overestimate higher precipitation events. However, the Category-I model comparatively performed better to capture the dry events (Fig. 7-4) than the wet events, but this may vary for different regions around the globe. For instance, MRI-CGCM3 showed very good skills and has been marked as the first category model in this study but to reproduce the Sahelian precipitation, MRI-CGCM3 showed insignificant or no skills whilst MPI-ESM-LR and MIROC5 are categorized as the second and third category model but were marked as improved skilled models for Sahelian precipitation (Gaetani and Mohino, 2013).

Previous studies (Jain et al., 2019; Lovino et al., 2018) reported that MMEM improves the models' skills to reproduce climate variables but the selection of models to form MMEM is very challenging as the arithmetic means of the models' output may further lead to loss of signal (Knutti et al., 2010). This study also examined the performance of MMEM and revealed that MMEM improves the performance metrics to some extent but not always and the performances are highly dependent on models' combination to form MMEM. For instance, MMEM2 shows better performance metrics than the other two combinations in reproducing the extremely dry and wet events where MMEM3 showed worse performance (Fig. 7-9). On contrary for the highest thresholds of individual metrics where few individual models were found better than MMEM3. Similar results were also reported in some other studies (Kumar et al., 2013; McKellar et al., 2013) where individual models were found better to some extent than the MMEM. However, lower skills of CMIP5 models for decadal precipitation as compared to temperature is also true for the MMEM which was also reported by Mehrotra et al. (2014).

In addition to understanding the climate system, models' configuration structuring spatial and temporal resolutions of the simulating variables also play a vital role in determining the model performance (Sakamoto et al., 2012). In this study except for CMCC-CM, models with finer atmospheric resolutions performed better than the coarser resolutions' models (see Table 7-1, Category-I model). It means, models of finer atmospheric resolutions can reproduce local climate features better than the models of coarser spatial resolutions and similar results were also reported in previous studies (Jain et al., 2019; Lovino et al., 2018). However, the lower skill of CMCC-CM may be due to the difference in understanding or geographical locations.

However, for different climate variables like temperature, the performance of CMCC-CM may be different (Lovino et al., 2018). This study will help the water manager, infrastructure developers, agricultural stakeholders to sort out the models before taking any decision in planning and developing infrastructures based on the models' predicted future precipitation. Findings of this study will also help the researchers for hydrological modelling, and other relevant stakeholders to increase the resilience of the society to climate change in relation to future water availability and uncertainty.

7.6 Conclusion

In order to use CMIP5 decadal precipitation at the catchment level, performances of eight selected GCMs were assessed in this study using different performance matrices (skill tests). Model performances to reproduce the hindcast data are necessary to check their credibility for the projection of future water availabilities. However, assessments of the decadal predictions have run long away in the last decade but no attention was paid to precipitation at the catchment level. For the CMIP5 decadal hindcast monthly precipitation, this is the first attempt that assessed the model performances at a catchment level with finer spatial resolution. Models are categorized based on the performance of individual models for temporal skills, dry and wet periods, and total precipitation (over time and space) at a selected grid and over the entire catchment. In addition, this study assessed the performance of different MMEMs formed from the combinations of different model categories. Considering a wide range of skill tests from both the temporal and spatial perspectives, the following conclusions are drawn.

- Models with higher atmospheric resolutions show comparatively better performances as opposed to the models of coarse spatial resolutions.
- Model performances vary over the initialization years and across the catchment. From 1990 onward, the skills of all models improved across the catchment where MIROC4h shows the highest skills followed by EC-EARTH and MRI-CGCM3 respectively. The internal structure of high atmospheric resolutions may be the main reason for MIROC4h reproducing the local climate variables comparatively better than the other.
- To reproduce the dry events and total precipitation over the entire catchment, EC-EARTH and MRI-CGCM3 respectively outperformed all models whilst CMCC-CM shows the lowest scores in all forms of skills. For capturing the wet periods, all models

showed almost similar skills with little exceptions for CMCC-CM and CanCM4 but for the dry periods, models show a range of skill scores.

- Based on the performance skills, the GCM models were ranked into three categories in ascending order: Category-I (MIROC4h, EC-EARTH, and MRI-CGCM3), Category-II (MPI-ESM-LR and MPI-ESM-MR), and category-III (MIROC5, CanCM4, and CMCC-CM). MMEMs were formulated as MMEM1 of Category-I models, MMEM2 combining Catergory-I and Category-II models, and MMEM3 as the combination of all three categories. Out of these three different MMEMs, MMEM2 was found performing better than other MMEMs based on the overall skills but MMEM1 performed relatively better for the case of extreme wet events. This shows the necessity of forming suitable MMEM for practical purposes of GCM data use especially for the decadal precipitation.

The outcomes presented in this study are based on one catchment in Australia only, but the process could be carried out in any catchment that has the availability of observed gridded data.

Acknowledgements

The authors are thankful to the working groups of the World Climate Research Program who made the CMIP5 decadal experiment data available for the researchers. Authors also would like to thank the Australian Bureau of Meteorology for providing the gridded observed rainfall data and the catchment's shape file. Authors gratefully acknowledge the financial support from Curtin University and CSIRO Data61 that was jointly provided to the first author to conduct his Ph.D. study at Curtin University, Australia.

List of symbols

C	:	Mean (over the total time span) of the observed values
F	:	Model forecasted (corrected) values
O	:	Observed values
$F - C$	:	Anomaly of the model values
$\overline{F - C}$	:	Mean of the model anomalies

$O - C$	:	Observed anomaly
$\overline{O - C}$	:	Mean of the observed anomalies
$Pr.$	:	Precipitation
P_f	:	Calculated fraction in FSS

References

Bhend, J., Whetton, P., 2015. Evaluation of simulated recent climate change in Australia. Australian Meteorological and Oceanographic Journal 65, 4–18. https://doi.org/10.22499/2.6501.003

Chen, J., Brissette, F.P., Lucas-Picher, P., Caya, D., 2017. Impacts of weighting climate models for hydro-meteorological climate change studies. Journal of Hydrology 549, 534–546. https://doi.org/10.1016/j.jhydrol.2017.04.025

Choi, J., Son, S.W., Ham, Y.G., Lee, J.Y., Kim, H.M., 2016. Seasonal-to-interannual prediction skills of near-surface air temperature in the CMIP5 decadal hindcast experiments. Journal of Climate 29, 1511–1527. https://doi.org/10.1175/JCLI-D-15-0182.1

Choudhury, D., Mehrotra, R., Sharma, A., Sen Gupta, A., Sivakumar, B., 2019. Effectiveness of CMIP5 Decadal Experiments for Interannual Rainfall Prediction Over Australia. Water Resources Research 55, 7400–7418. https://doi.org/10.1029/2018WR024462

Choudhury, D., Sharma, A., Sen Gupta, A., Mehrotra, R., Sivakumar, B., 2016. Sampling biases in CMIP5 decadal forecasts. Journal of Geophysical Research: Atmospheres 121, 3435–3445. https://doi.org/10.1002/2016JD024804

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Anav, A., Andrews, T., Baehr, J., Bodas-salcedo, A., Catto, J., Sillmann, J., Simmons, A., 2013. Evaluation of Climate Models, in: Intergovernmental Panel on Climate Change (Ed.), Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge, pp. 741–866. https://doi.org/10.1017/CBO9781107415324.020

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27, 1547–1578. https://doi.org/10.1002/joc.1556

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

Gaetani, M., Mohino, E., 2013. Decadal prediction of the sahelian precipitation in CMIP5 simulations. Journal of Climate 26, 7708–7719. https://doi.org/10.1175/JCLI-D-12-00635.1

Grotch, S.L., MacCracken, M.C., 1991. The Use of General Circulation Models to Predict Regional Climatic Change. Journal of Climate 4, 286–303. https://doi.org/10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2

Homsi, R., Shiru, M.S., Shahid, S., Ismail, T., Harun, S. Bin, Al-Ansari, N., Chau, K.W., Yaseen, Z.M., 2020. Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. Engineering Applications of Computational Fluid Mechanics 14, 90–106. https://doi.org/10.1080/19942060.2019.1683076

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., 2021a. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale. Journal I, 285.

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021b. Intercomparison of drift correction alternatives for CMIP5 decadal precipitation. International Journal of Climatology joc.7287. https://doi.org/10.1002/joc.7287

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021c. Drift in CMIP5 decadal precipitation at catchment level. Stochastic Environmental Research and Risk Assessment 8, 5. https://doi.org/10.1007/s00477-021-02140-8

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021d. A comparative study on 10 and 30-year simulation of CMIP5 decadal hindcast precipitation at catchment level, in: Vervoort, R.W., Voinov, A.A., Evans, J.P. and Marshall, L. (Ed.), MODSIM2021, 24th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 609–615. https://doi.org/10.36334/modsim.2021.K5.hossain

Islam, S.A., Bari, M.A., F. Anwar, A.H.M., 2014. Hydrologic impact of climate change on Murray-Hotham catchment of Western Australia: A projection of rainfall-runoff for future water resources planning. Hydrology and Earth System Sciences 18, 3591–3614. https://doi.org/10.5194/hess-18-3591-2014

Jain, S., Salunke, P., Mishra, S.K., Sahany, S., 2019. Performance of CMIP5 models in the simulation of Indian summer monsoon. Theoretical and Applied Climatology 137, 1429–1447. https://doi.org/10.1007/s00704-018-2674-3

Jones, P.W., 1999. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Kamworapan, S., Surussavadee, C., 2019. Evaluation of CMIP5 global climate models for simulating climatological temperature and precipitation for southeast Asia. Advances in Meteorology 2019. https://doi.org/10.1155/2019/1067365

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G.A., 2010. Challenges in Combining Projections from Multiple Climate Models. Journal of Climate 23, 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Kumar, D., Kodra, E., Ganguly, A.R., 2014. Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation. Climate Dynamics 43, 2491–2518. https://doi.org/10.1007/s00382-014-2070-3

Kumar, S., Merwade, V., Kinter, J.L., Niyogi, D., 2013. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. Journal of Climate 26, 4168–4185. https://doi.org/10.1175/JCLI-D-12-00259.1

Lovino, M.A., Müller, O. V., Berbery, E.H., Müller, G. V., 2018. Evaluation of CMIP5 retrospective simulations of temperature and precipitation in northeastern Argentina. International Journal of Climatology 38, e1158–e1175. https://doi.org/10.1002/joc.5441

McKellar, C., Cordero, E.C., Bridger, A.F.C., Thrasher, B., 2013. Evaluation of the CMIP5 Decadal Hindcasts in the State of California. Department of Meteorology and Climate Science. San José State University.

McSweeney, C.F., Jones, R.G., Lee, R.W., Rowell, D.P., 2015. Selecting CMIP5 GCMs for

downscaling over multiple regions. Climate Dynamics 44, 3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Meehl, G.A., Teng, H., 2014. CMIP5 multi-model hindcasts for the mid-1970s shift and early 2000s hiatus and predictions for 2016-2035. Geophysical Research Letters 41, 1711–1716. https://doi.org/10.1002/2014GL059256

Meehl, G.A., Teng, H., Maher, N., England, M.H., 2015. Effects of the Mount Pinatubo eruption on decadal climate prediction skill of Pacific sea surface temperatures. Geophysical Research Letters 42, 10840–10846. https://doi.org/10.1002/2015GL066608

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. Journal of Hydrology 519, 2932–2951. https://doi.org/10.1016/j.jhydrol.2014.07.053

Moise, A., Wilson, L., Grose, M., Whetton, P., Watterson, I., Bhend, J., Bathols, J., Hanson, L., Erwin, T., Bedin, T., Heady, C., Rafter, T., 2015. Evaluation of CMIP3 and CMIP5 Models over the Australian Region to Inform Confidence in Projections. Australian Meteorological and Oceanographic Journal 65, 19–53. https://doi.org/10.22499/2.6501.004

Purwaningsih, A., Hidayat, R., 2016. Performance of Decadal Prediction in Coupled Model Intercomparisson Project Phase 5 (CMIP5) on Projecting Climate in Tropical Area. Procedia Environmental Sciences 33, 128–139. https://doi.org/10.1016/j.proenv.2016.03.064

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Monthly Weather Review 136, 78–97. https://doi.org/10.1175/2007MWR2123.1

Sakamoto, T.T., Komuro, Y., Nishimura, T., Ishii, M., Tatebe, H., Shiogama, H., Hasegawa, A., Toyoda, T., Mori, M., Suzuki, T., Imada, Y., Nozawa, T., Takata, K., Mochizuki, T., Ogochi, K., Emori, S., Hasumi, H., Kimoto, M., 2012. MIROC4h-A new high-resolution atmosphere-ocean coupled general circulation model. Journal of the Meteorological Society of Japan 90, 325–359. https://doi.org/10.2151/jmsj.2012-301

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. International Journal of Climatology

23, 887–901. https://doi.org/10.1002/joc.922

Sheffield, J., Camargo, S.J., Fu, R., Hu, Q., Jiang, X., Johnson, N., Karnauskas, K.B., Kim, S.T., Kinter, J., Kumar, S., Langenbrunner, B., Maloney, E., Mariotti, A., Meyerson, J.E., Neelin, J.D., Nigam, S., Pan, Z., Ruiz-Barradas, A., Seager, R., Serra, Y.L., Sun, D., Wang, C., Xie, S., Yu, J., Zhang, T., Zhao, M., 2013. North American Climate in CMIP5 Experiments. Part II: Evaluation of Historical Simulations of Intraseasonal to Decadal Variability. Journal of Climate 26, 9247–9290. https://doi.org/10.1175/JCLI-D-12-00593.1

Skelly, W.C., Henderson-Sellers, A., 1996. Grid box or grid point: What type of data do GCMs deliver to climate impacts researchers? International Journal of Climatology 16, 1079–1086. https://doi.org/10.1002/(sici)1097-0088(199610)16:10<1079::aid-joc106>3.0.co;2-p

Ta, Z., Yu, Y., Sun, L., Chen, X., Mu, G., Yu, R., 2018. Assessment of Precipitation Simulations in Central Asia by CMIP5 Climate Models. Water 10, 1516. https://doi.org/10.3390/w10111516

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

CHAPTER 8

MONTHLY PRECIPITATION PREDICTION AT CATCHMENT LEVEL BY FACEBOOK PROPHET MODEL USING OBSERVED AND CMIP5 DECADAL DATA

Abstract

Early prediction of precipitation is important for the planning of agriculture, water infrastructure, and other socio-economic developments. Near-term prediction (e.g., 10 years) of hydrologic data is a recent development in GCM (General Circulation Model) simulations such as CMIP5 (Coupled Modelled Intercomparison Project phase-5) decadal experiments. The prediction of monthly precipitation on a decadal time scale is an important step for catchment management. Previous studies considered stochastic models using the observed time-series data only for precipitation prediction but no studies used the GCM decadal data together with the observed data at the catchment level. This study used the Facebook prophet (FBP) model and six machine-learning (ML) regression algorithms for the prediction of monthly precipitation on the decadal time scale for the Brisbane river catchment in Queensland Australia. Monthly hindcast decadal precipitation data of eight GCMs (EC-EARTH, MIROC4h, MRI-CGCM3, MPI-ESM-LR, MPI-ESM-MR, MIROC5, CanCM4, and CMCC-CM) were downloaded from the CMIP5 data portal, and the observed data were collected from the Australian Bureau of Meteorology. At first, the FBP has been used for the predictions based on; (i) the observed data only, and (ii) a combination of observed and CMIP5 decadal data. In the next step, predictions are performed through ML regressions where CMIP5 decadal data are used as features and corresponding observed data are used as target variables. The prediction skills are assessed through several skill tests including Pearson Correlation Coefficient (PCC), Anomaly Correlation Coefficient (ACC), Index of Agreement (IA), and Mean Absolute Error (MAE). Upon comparing the skills, this study finds that predictions based on the combination of observed and CMIP5 decadal data through FBP provides better skills

This chapter has been published as: Hossain, M.M., Anwar, A.H.M.F., Garg, N., Prakash, M., Bari, M., 2022. Monthly Rainfall Prediction at Catchment Level with the Facebook Prophet Model Using Observed and CMIP5 Decadal Data. Hydrology 9, 111. https://doi.org/10.3390/hydrology9060111. However, few textual changes have been made to address the examiners' comments.

than the predictions based on the observed data only. The performance of FBP showing higher skills, especially for the dry periods is mainly due to its multiplicative seasonality function.

Keywords: Facebook Prophet, precipitation, prediction, monthly, knowledge-driven, data-driven.

8.1 Introduction

Rainfall is a very important climate variable and precious natural resource, which affects our livelihood and agriculture in many dimensions. An early and accurate prediction of precipitation enables more efficient management of floods, agriculture, water resources, power development, and planning and development of infrastructure (Apurv et al., 2015; Hansen et al., 2011; Jones et al., 2000; Mehta et al., 2013). However, the prediction of this most important hydrological aspect has become a very challenging task in terms of accuracy due to its peculiar variations over time and space. Due to ongoing climate change, the temporal and spatial variations of precipitation have been intensified in the past few decades. Over the past few years, precipitation prediction has become a greater concern to the climate research community (Ali et al., 2019; George et al., 2016; Hossain et al., 2020; Hung et al., 2009; Mekanik et al., 2011; Mislan et al., 2015; Ouyang et al., 2016). The precipitation prediction approaches are broadly classified into two main categories; (i) a knowledge-driven approach, and (ii) a data-driven approach. Knowledge-driven approaches use scientific understanding, thermodynamic balance, and physical mechanisms of hydrological processes such as General Circulation Models (GCMs). GCMs predict climate variables of coarse spatial resolutions on a global scale. However, the knowledge-driven approach needs extensive data and computational facility that sometimes becomes unavailable (Hong, 2008). The data-driven approach is the stochastic and/or empirical statistical modelling approach that is widely used in precipitation prediction at the local level based on the observational relationship of the predictand variable. The data-driven approaches have some limitations and all approaches could not perform well in predicting for longer time spans as they cannot capture the non-linearity and dynamic behaviour of precipitation over time (Rajeevan, 2001; Zhang, 2003). Several statistical/stochastic methods have been used for precipitation prediction and most of them are based on regression analysis such as simple regression analysis (SRA), exponential smoothing, decomposition, and auto-regressive integrated moving average (ARIMA). Every individual

method has its strengths and weaknesses. For instance, ARIMA is a popular stochastic model for time series prediction with greater flexibility. But, as a stochastic model, it needs stationarity of data (Machiwal and Jha, 2012) and its presumed linear form of the associated data sometimes makes it inappropriate for complex nonlinear time series data like precipitation (Zhang, 2003). This is why, a better output from ARIMA heavily depends on the expertise of the modeller (Machiwal and Jha, 2012). Dastorani et al., (2016) compared different forms of the ARIMA model and concluded that the model parameters need to be tuned to get a certain level of accuracy based on location and data type.

Applications of machine learning algorithms, of which artificial neural networks (ANN) of the different forms of architecture, have been popular for many time series predictions including time series of precipitation and enhanced the prediction accuracy (Hung et al., 2009; Lee et al., 2018; Lin et al., 2022; Meinke et al., 2007; Mekanik et al., 2011; Mislan et al., 2015; Shen et al., 2022). According to the level of complexity of the dataset, ANN can be combined with different types of algorithms due to its highly flexible characteristics. However, based on the need and opportunities, different researchers have come up with different research interests and time scales with the application of ANN. For instance, Wu et al., (2001) predicted monsoon precipitation in China over 10 years ahead whereas Chakraverty and Gupta, (2008) predicted Indian Monsoon precipitation 6 years in advance. To predict the summer monsoon of India for 1 year in advance, Chattopadhyay and Chattopadhyay (2008) used 129 years of historical data. Though the ANN is good to capture the nonlinear relationship of data, the presence of outliers in the time-series data can critically affect the reliability of ANN as it is a grey box model. Thus ANN requires proper data pre-processing before its application, especially for the climatic data (Committee, 2000; Ramírez et al., 2006). Some other hybrid models also came into existence and showed very good skills in precipitation predictions (Khandelwal et al., 2015; Unnikrishnan and Jothiprakash, 2020; Zhang, 2003). However, all of the above-mentioned data-driven approaches used the historically observed data, and based on the historical relationship they performed the predictions for several years ahead assuming the climatic conditions remain the same in the historical and prediction period.

Compared to the other climatic variables, precipitation has been affected mostly due to ongoing climate change. Over the past few decades, temporal change and shifting of precipitation patterns, extreme precipitation during wet periods, extreme longer dry spells during dry

periods, an overall reduction of total precipitation amount have been very common phenomena around the globe. In the last decades, these changes have been intensified due to the ongoing climate change (IPCC, 2014). Climate change does exist that will continue to change, but the rate of change may be higher in the future. The future higher rate of climate change may adversely affect the future precipitation and its level of impact may be significant (IPCC, 2014). Therefore, researchers should not rely only on data-driven approaches (based on the historical data only) for future precipitation prediction. For this reason, this study aimed to predict future precipitation for decadal time scales in a combination of both the knowledge and data-driven approach where both the GCMs derived precipitation and historically observed data are employed. To do this, this study used Facebook Prophet (description provided later) model where historically observed precipitation was used as an input variable and GCMs derived precipitation data from the decadal experiment of Coupled Model Intercomparison Project Phase-5 (CMIP5) (description provided in the data collection section) data were used as an additional regressor to guide Prophet in the prediction process. Though the application of Prophet in time series prediction is not new (Samal et al., 2019; Subashini et al., 2019; Toharudin et al., 2020) but it is rarely found in the literature for predicting precipitation.

8.2 Study area, data, and methods

8.2.1 Study area

The Brisbane River catchment in Queensland was selected as the study area that lies in the eastern states of Australia in between the latitudes 26.50S ~28.150S and the longitudes 151.70E ~153.150E (Fig. 8-1). It has an area of 13549 square kilometers and a sub-tropical climate where maximum precipitation occurs during summer (December-January-February) and minimum precipitation in winter (June-July-August) (Climate-Data, 2020). Monthly observed precipitation (1911-2015) over the Brisbane River catchment varied from nil to 1360 mm with an annual average precipitation of 628 mm (BoM, 2020) and the number of upper and lower extremes are not quite small.

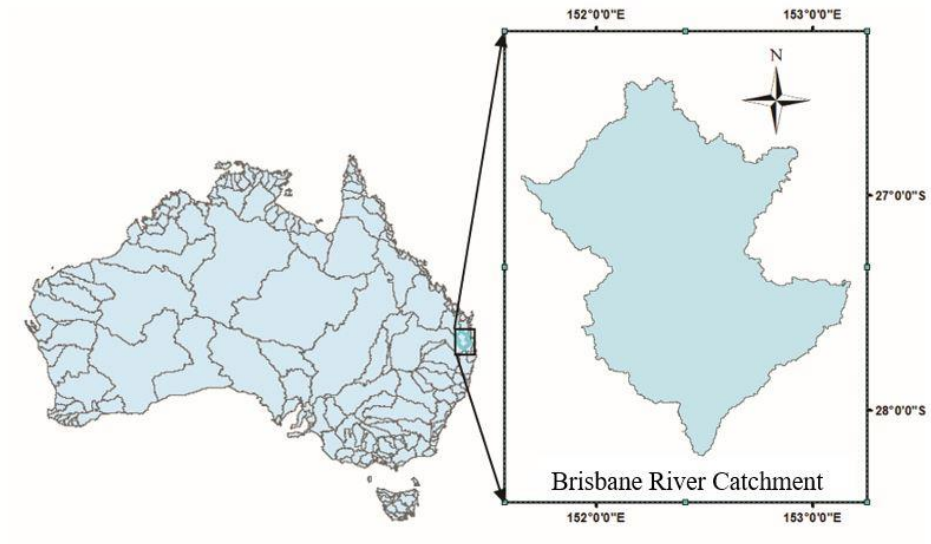


Fig. 8-1 Study Area

8.2.2 Data collection

The observed monthly gridded precipitation of 0.05° × 0.05° (5km × 5km) spatial resolution for entire Australia were collected from the Australian Bureau of Meteorology (BoM). The BoM has produced the gridded data using the Australian Water Resources Assessment Landscape model (AWRA-L V5) (Frost et al., 2016). Monthly hindcasts precipitation (precipitation) data of decadal time scale from eight (EC-EARTH, MIROC4h, MRI-CGCM3, MPI-ESM-LR, MPI-ESM-MR, MIROC5, CanCM4, and CMCC-CM) GCMs were downloaded from CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/) for the period 1960-2005; initialized at 1960, 1965, 1970, 1975… thus every five years up to 2005. The name of models, spatial resolutions, and the available historical run are given in Table 8-1.

Table 8-1 List of models (GCMs) used as additional regressors in this study

Model	Resolutions (°lon × °lat))	Initialization Year (1960-2005)									
		60	65	70	75	80	85	90	95	00	05
		Number of ensembles									
EC-EARTH	(1.125 X 1.1215)	14	14	14	14	14	14	14	14	10	18
MRI-CGCM3	(1.125 X 1.1215)	06	08	09	09	06	09	09	09	09	06
MPI-ESM-LR	(1.875 X 1.865)	10	10	10	10	10	10	10	10	10	10
MPI-ESM-MR	(1.875 X 1.865)	03	03	03	03	03	03	03	03	03	03
MIROC4h	(0.5625 X 0.5616)	03	03	03	06	06	06	06	06	06	06
MIROC5	(1.4062 X 1.4007)	06	06	06	06	04	06	06	06	06	06
CanCM4	(2.8125 X 2.7905)	20	20	20	20	20	20	20	20	20	20
CMCC-CM	0.75 X 0.748	03	03	03	03	03	03	03	03	03	03

8.2.3 Data processing

In the first step, all the available ensembles of individual initializations were averaged to produce a single dataset and then were subsets for the Australian region. Secondly, the averaged ensembles were spatially interpolated, using the second-order conservative (SOC) method, onto 0.05°×0.05° spatial resolution thus matching with the grid used in the observed data. This study used SOC as it conserves the precipitation flux while sub-gridding the GCMs data (Jones, 1999) and marked it as the most suitable spatial interpolation method, especially for the GCMs derived gridded dataset (Hossain et al., 2021a). Then both the models' and observed datasets were subsets for the Brisbane River catchment. Every initialization spans a dataset of 10 years that overlaps five years with the dataset of the next initializations. In the third step, the last five years of each initialization, except 2005, were discarded and the first five years were combined to produce a single time series from 1961 to 2015. For the initialization year 2005 (2006-2015), the whole dataset was taken instead of the first five years to make the dataset longer.

8.2.4 Model description

This study used FBP Model to predict monthly precipitation for a decade (2006/Jan-2015/Dec) and then the performance of FBP's predicted values was compared with the predictions from six different machine learning regression models; Multi-Layer Perceptron (MLP), Epsilon-Support Vector Regression (SVR), Light Gradient Boosting (LGB), Extreme Gradient Boosting (XGB), Random Forest (RDF) and the combination of these five models. The descriptions of all models are given below.

Facebook Prophet (FBP)

FBP is a fully automatic open-sourced time-series forecasting library developed by Facebook's Core Data Science team. Though Prophet was built for business purposes, it works for observed hourly, daily, weekly, and monthly time series data that has strong seasonality. It predicts time series as a generalized additive model combining the trend function, seasonality function, holiday effects, and an error term as given in equation 8.1.

$$Y(t) = g(t) + s(t) + h(t) + \in_t \quad (8.1)$$

Where $g(t)$ and $s(t)$ represent the trend and seasonality respectively whilst $h(t)$ presents the holiday effect and $\in_t$ is the error term. As this study uses the monthly precipitation data as an input variable, therefore, holiday effect will be invalid here. FBP provides a decomposition regression model that is extendable and easy to use for time series forecasting with a wide range of tunable parameters. FBP has many default parameters' values that maintain its fully automatic nature. However, a little change of its parameters values does not make a big difference in the prediction process. It has functionality for cross-validation to measure the forecasting errors and provision to include additional regressor and customize the seasonality. The additional regressor feature enhances forecasting accuracy, makes the prediction process more transparent, and helps to tune the prediction process. The additional regressor must be a separately forecasted variable that should be available for both the training and prediction periods.

Prophet can handle outliers, without any requirement for imputation, and missing data but the best way to handle the outliers is to remove them. Taylor and Letham (2018) described further information about Prophet on simulating historical forecasting. Compared to the other data-

driven approaches, Prophet has two main advantages; (i) Prophet automatically detects changes in trends by selecting change points from the historical data and it is much more straightforward to create a reasonable, accurate forecast, (ii) Its predictions are customizable in ways that are intuitive to non-expert users and does not need rigorous data pre-processing. It is easy to use and the components are easily explainable. Its predictions are decent, however, in some cases, certain parameters need to be tweaked compared to the default setting and that can be easily done.

Multi-Layer Perceptron (MLP) Regressor

MLP is a class of feedforward artificial neural networks (ANN) that utilizes a supervised learning algorithm. It learns by training the dataset using backpropagation with no activation function in the output layer.

Epsilon-Support Vector Regression (SVR)

SVR is also a supervised learning algorithm that acknowledges the presence of non-linearity of the data and provides a proficient prediction using the similar principle of Support Vector Machines (SVMs). The basic idea of SVR is to find the best-fit line that has the maximum number of points. To fit the best line within a threshold, SVR tries to minimize the errors between the real and predicted values.

Gradient Boosting

Boosting is a strategy that combines several simple models into a composite single model. Gradient boosting is a type of boosting and a very popular supervised machine learning technique for regression problems. Light Gradient Boosting (LGB) uses histogram-based learning algorithms following a leaf-wise splitting approach whilst (Extreme Gradient Boosting) XGB uses a level-wise tree growth approach. XGB is a more regularized form of gradient boosting that delivers a more accurate prediction by using the strengths of the second-order derivative of the loss function.

Random Forest Regressor (RDF)

RDF is a supervised learning algorithm that uses the ensemble-learning method for regression problems. RDF builds multiple decision trees during the training period and merges them to

get a more stable and accurate prediction. To control the overfitting problem, a bootstrap technique was used in RDF.

In addition to the five regression models, another combined regression model was developed by stacking the above-mentioned five regression models (referred to as STC) and was used to predict the precipitation for the same period, 2006-2015.

For the predictions, at first, all models were trained from 1911/1961 to 2005 and then predicted for 2006-2015. This study considered two different cases: (i) only the observed data to demonstrate the data-driven approach (Case-I), (ii) observed data along with GCMs derived precipitation data as the additional regressor (Case-II) to demonstrate the combination of data and knowledge-driven approach. For ease of comparison, each of these two cases is further divided into two different sub-cases; Case-I (a, b) and Case-II (a, b). For Case-I, the FBP model was trained from 1911 [referred to as Case-I (a)] and 1961 [referred as Case-I (b)]. For Case-II, two different modes of additional regressor are added to FBP. The first additional regressor is the arithmetic mean of the best five GCMs; MIRCO4h, EC-EARTH, MRI-CGCM3, MPI-ESM-LR, and MPI-ESM-MR [henceforth referred as MMEM] among the eight considered GCMs in Table 8-1. The arithmetic mean of these five models showed comparatively better performance in separate research in chapter 7 (of the same authors). FBP with MMEM as an additional regressor is referred to as case-II (a) whilst FBP with all the eight GCMs (in Table 8-1) as eight individual regressor (together) is referred to as case-II (b). As the CMIP5 decadal data are available since 1961, Case-II and all regression models were trained from 1961 to 2005. For training the regression models, GCMs derived precipitation (MMEM) was used as the independent variable (feature) and corresponding observed data was used as the dependent variable (target variable). After training the regression model from 1961 to 2005, GCMs derived data from 2006 to 2015 was provided to the trained regression models to predict the dependent variable (the observed data).

However, to train the model, the most important task is to optimize the model parameters. To optimize the models' parameters, a wide range of parameter values were given and the best parameters combinations were chosen based on the minimum mean absolute errors (FBP) using the Scikit-Learn Parameter Grid function. For FBP, optimization of the parameters' values was performed at a single grid point (latitude 27.50S and longitude 153.050E, henceforth referred to as Point-I). The same parameter values were then applied to the two other points as FBP is

automatic and a little change in parameter values does not make a big difference. In addition, this study assumed there will not be a big change in the optimized parameters' values at the selected points as the study area was small. However, it can be considered as a limitation of this study. A multiplicative seasonality function with cap and floor values 600 mm and 0 mm respectively are used for PFB prediction. For the regression models, parameters were optimized for all three points and MMEM for the training and used prediction purposes. In the training process, the regression models eventually developed transfer functions (GCMs to observed values) using the best combination of the used parameters (see Table S1) based on the minimum mean squared errors after going through 10-times cross-validation. These transfer functions, obtained from the training period, were then used to transfer the GCMs data (2006-2015) to the target variable (referred to as predated data).

As Prophet performs better without outliers, the monthly precipitation values above 250 mm in the observed dataset were set to 250 and zero values were replaced by 1.0 mm. Note that Prophet shows predicted values from its training periods to predicted period. In addition to the predicted values, it also provides upper and lower limits of the predicted values along with other statistical parameters. Preliminary results revealed that FBP could not reproduce the upper (in summer) and lower extreme precipitation (in winter). For this reason, few correction factors are employed. For instance, a factor of 0.85 for the months of July and August, 1.15 for December, and an average of the upper limit, and the raw predicted values were employed for January and February. These correction factors were obtained based trial and error basis from the comparisons between the observed and the raw predictions at several randomly selected locations within the Brisbane River catchment.. The final predicted values were then examined using four different skills such as Pearson Correlation Coefficient (PCC), Anomaly Correlation Coefficient (ACC), Index of Agreement (IA), and Mean Absolute Error (MAE). A brief description of the skill tests is given below and the detailed descriptions can be found in (Hossain et al., 2021b).

8.2.5 Skill tests

Pearson correlation coefficient (PCC):

PCC is a very commonly used performance metric that measures the linear correlation between two datasets. Here, it is used to measure the linear correlation between the predicted and observed values. Its value varies between -1 and 1 (perfect correlation).

$$PCC = \frac{\sum_{t=1}^{N}(P_t-\bar{P})(O_t-\bar{O})}{\sqrt{\sum_{t=1}^{N}(P_t-\bar{P})^2}\sqrt{\sum_{t=1}^{N}(O_t-\bar{O})^2}} \quad (8.2)$$

Where P and O present the predicted and observed values respectively and this notation will be the same for the following skill tests also. A bar over the predicted ($\bar{P}$) and observed ($\bar{O}$) represent the mean of the predicted and observed values respectively. N is the maximum lead-time (e,g., the maximum number of months-120).

Anomaly Correlation Coefficient (ACC):

ACC was suggested by Wilks, (2011) for measuring the correlation between the anomalies between two datasets. Here, ACC is used to measure the temporal anomaly correlation between the anomalies of predicted and observed values. Anomalies are calculated by subtracting the mean (C, mean of the observed values over the entire prediction period) from both the predicted and corresponding observed values.

$$ACC = \frac{\sum\{(P-C)-(\overline{P-C})\}\times\{(O-C)-(\overline{O-C})\}}{\sqrt{\sum(P-C)^2}\sqrt{\sum(O-C)^2}} \quad (8.3)$$

ACC values range from zero to 1.0 and higher values of ACC do not represent the higher accuracy of the prediction values but the anomalies.

Index of Agreement (IA):

Wilmot (1982) suggested IA for measuring the accuracy of predicted data based on the corresponding observed values. IA values bounded between 0 and 1, where, the value closer to 1 presents the more efficient prediction.

$$IA = 1 - \frac{\sum_{t=1}^{N}(P_t-O_t)^2}{\sum_{t=1}^{N}(|P_t-O'|+|O_t-O'|)^2} \quad (8.4)$$

Here O' presents the mean of every individual year of the predicted period.

Mean Absolute Error (MAE):

MAE measures the average magnitude of errors, the differences between the predicted and observed values. MAE values range from 0 to ∞, where the lower value indicates higher accuracy and vice versa.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|P_t - O_t| \qquad (8.5)$$

8.3 Results and discussion

There are 496 grids, of 5.0 x 5.0 km spatial resolution, available in the Brisbane River catchment. This study predicted monthly precipitation for a decade at three different locations (see

Table 8-2), which are closest to the automated weather stations operated by the Australian Bureau of Meteorology. Precipitation predictions for a few months, seasons, or to some extent for a few years are commonly seen in the literature. This study predicted precipitation for a decade because of using the additional regressor for a decadal time scale that was derived through the GCMs contributed to the decadal experiment of CMIP5. In the first section, FBP's predictions of different cases are compared and assessed through different skill tests. In the second part, the performances of FBP for the monthly precipitation predictions are compared with the performance of six different regression models.

8.3.1 Prediction using FBP

Fig. 8-2 presents the comparison between observed and FBP predicted monthly precipitation values of different cases at Point-I where Case-I presents the data-driven, MMEM presents the knowledge-driven and Case-II presents the combination of knowledge and data-driven approach. From the comparison, it is evident that FBP can reproduce the seasonal variability with better performance in producing dry events but none of the cases of FBP could reproduce the extreme peak values of the observed precipitation. However, Case-II(a) shows comparatively better performance to catch the upper peaks followed by Case-I(a). For the dry events, Case-II shows a considerably better resemblance with the observed values compared to Case-I. It means prediction skills improve, especially in dry events, when the combination of

knowledge and data-driven approach is employed. These improvements are also observed in the skill tests of all considered cases.

Table 8-2 presents the different skill test results and the percentage of over and under prediction of total precipitation, cumulative sum over the different periods, of different cases of FBP along with MMEM at three selected points.

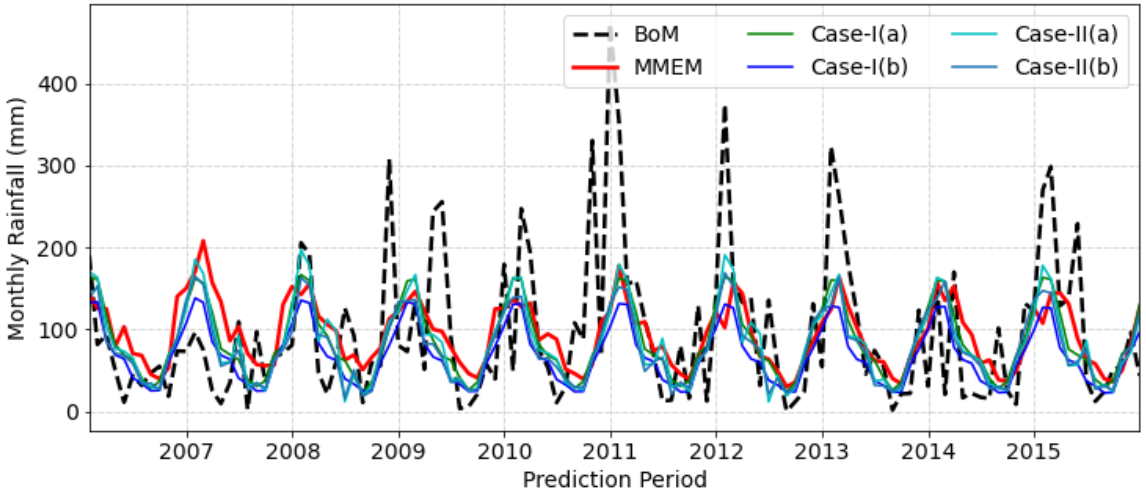


Fig. 8-2 Comparison of FBP predicted monthly precipitation of different cases with the corresponding observed values

Comparing the predicted values and their skill tests at all three points, it is evident that Case-I(a) shows comparatively better skills and comparatively a lower percentage of under/over prediction of total precipitation than Case-I(b). This is due to the higher training period (1911-2015) of Case-I(a) which is about double the training period of all other cases. This means, with the higher training period, FBP can come up with better prediction performance. Among the cases of similar training periods, Case-II(a) shows comparatively better skill scores at Point-I, where FBP model parameter values were optimized. However, at points II and III, Case-II(b) showed the highest skills than Case-II(a). Better performance of case-II(b) may be due to involving of a range of climate responses provided by the different GCMs as individual regressors or their higher skills than of MMEM. The better skills of case-II(a) at Point-I may be either due to better skills of MMEM (Table 8-2) that guided FBP as additional regressor or tuning the FBP parameters (whereas at other two points FBP models' parameters values were not tuned).

Table 8-2 Comparison of skills and total precipitation prediction among the different cases of FBP models

Location (Lon/Lat)	Cases	Skills				Under and overestimation of total Precipitation (%)			
		MAE	PCC	ACC	IA	1Y	3Y	5Y	8Y
Point-I (153.05E/ 27.50S)	I-(a)	53.6	0.549	0.536	0.615	35.9	14.6	-7.2	-11.2
	I-(b)	55.9	0.526	0.418	0.491	11.9	-5.94	-24.6	-28.7
	II-(a)	54.9	0.533	0.517	0.622	33.5	15.1	-8.34	-12.8
	II-(b)	55.1	0.528	0.488	0.577	24.8	5.25	-16.0	-19.3
	MMEM	58.11	0.434	0.433	0.510	48.6	35.6	5.64	-3.1
Point-II (152.0E/ 27.0S)	I-(a)	40.8	0.497	0.496	0.603	50.2	26.5	2.12	-10.4
	I-(b)	41.0	0.484	0.484	0.581	50.7	27.1	2.46	-6.4
	II-(a)	40.8	0.489	0.486	0.593	47.3	26.3	0.53	-8.4
	II-(b)	39.8	0.519	0.517	0.611	38.5	22.7	-1.42	-8.2
	MMEM	41.4	0.494	0.493	0.612	58.2	39.3	13.8	-5.6
Point-III (152.05E / 27.30S)	I-(a)	46.1	0..491	0.490	0.588	54.4	24.1	3.4	-6.7
	I-(b)	48.1	0.471	0.470	0.583	65.5	32.1	10.6	-0.15
	II-(a)	46.9	0.464	0.460	0.567	51.7	23.1	1.1	-9.9
	II-(b)	45.2	0.490	0.485	0.580	44.6	18.7	-1.6	-10.2
	MMEM	44.7	0.489	0.474	0.571	48.7	23.9	-0.8	-14.2

It is difficult to attach here any valid reason behind why case-II(b) is performing better in the other two points but not at Point-I as no comparison are conducted on the performance of MMEM and all selected GCMs between the points. Anyway, case-II (either a or b) is the combination of knowledge and data-driven approach that provided better prediction skills than only the data-driven approaches. From the predicted values and their skill comparisons at all three selected points, this study reveals that the skill improvement of FBP predictions with the combination of the knowledge and data-driven approach was due to better capturing the dry events that were even better than the MMEM (Fig. 8-3). For reproducing the peak values (upper extremes), the knowledge-driven approach (MMEM) has been found better than FBP predictions. From Fig. 8-3, it can be observed that for the dry events (lower values), FBP predictions show a comparatively better resemblance with the observed values and worse for the upper peaks (wet events).

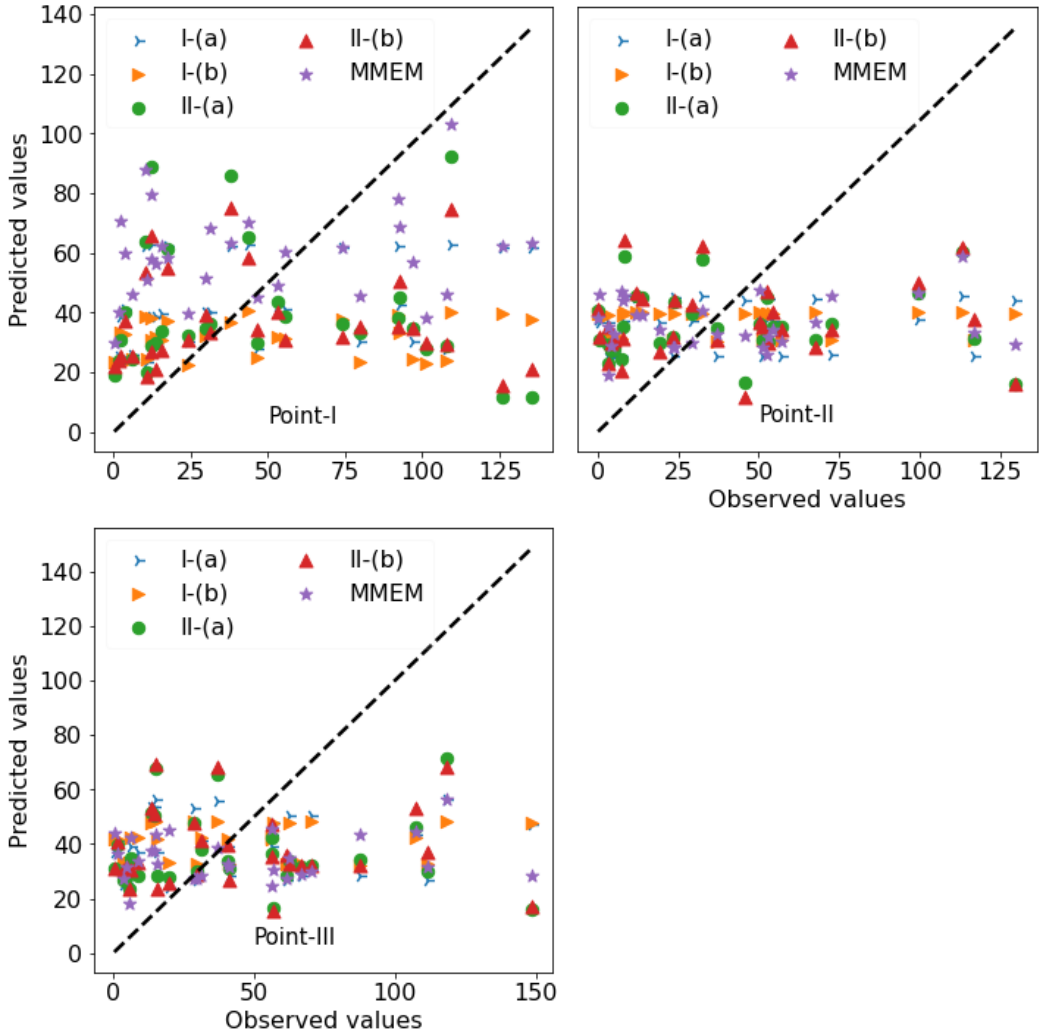


Fig. 8-3 Comparison of the resemblance of dry and wet events produced by FBP and MMEM

For capturing the dry events, the combination of knowledge and data-driven approaches was found better than any of the individual approaches where case-II(b) little better than II(a) in terms of all skills, as well as total precipitation predictions over different time spans.

8.3.2 Prediction using regression models

Comparison of monthly precipitation prediction at Point-I by five supervised machine-learning regression algorithms and their stacking models (STC) using MMEM as the feature is presented in Fig. 8-4. From the comparison it is observed that the regression models are also able to reproduce seasonal variations with very little improvement in reproducing the lower extremes than the MMEM. This study also used all selected models as independent variables, but the results were not as good as MMEM. For this reason, only the skills of regression models using

MMEM as a feature are presented here. From the comparisons, it is observed that all regression models, except RDF, show similar skills in monthly precipitation prediction. Among them MLP and SVR are comparatively better than others (see Table 8-3).

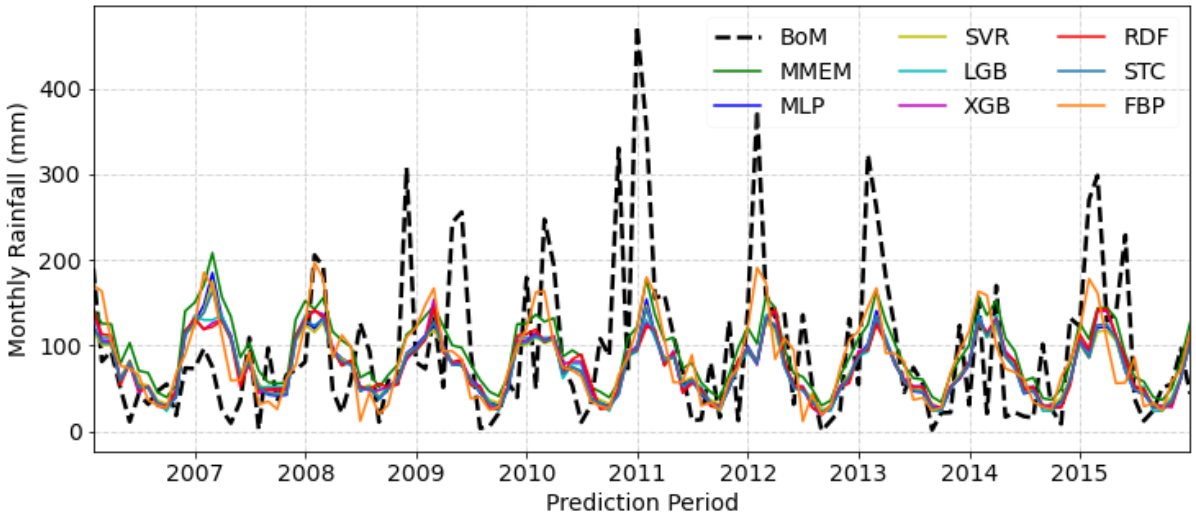


Fig. 8-4 Monthly precipitation predicted from different regression models

Table 8-3 Skill comparison of different regression models

Models	Point-I				Point-II				Point-III			
	MAE	PCC	ACC	IA	MAE	PCC	ACC	IA	MAE	PCC	ACC	IA
MLP	57.1	0.430	0.371	0.445	39.3	0.480	0.450	0.515	43.5	0.476	0.427	0.494
SVR	57.6	0.430	0.361	0.418	39.3	0.481	0.447	0.516	43.5	0.478	0.430	0.487
LGB	56.6	0.432	0.374	0.442	39.5	0.469	0.432	0.510	43.7	0.466	0.425	0.493
XGB	57.2	0.427	0.370	0.439	39.7	0.451	0.417	0.503	44.1	0.444	0.410	0.484
RDF	57.2	0.426	0.369	0.441	39.9	0.427	0.372	0.433	44.0	0.421	0.359	0.412
STC	57.1	0.434	0.365	0.435	39.1	0.483	0.425	0.475	43.4	0.464	0.405	0.464
FBP(II-a)	54.9	0.533	0.517	0.622	40.9	0.489	0.486	0.593	46.9	0.464	0.460	0.567
MMEM	58.1	0.434	0.433	0.510	41.4	0.494	0.493	0.612	44.7	0.489	0.474	0.571

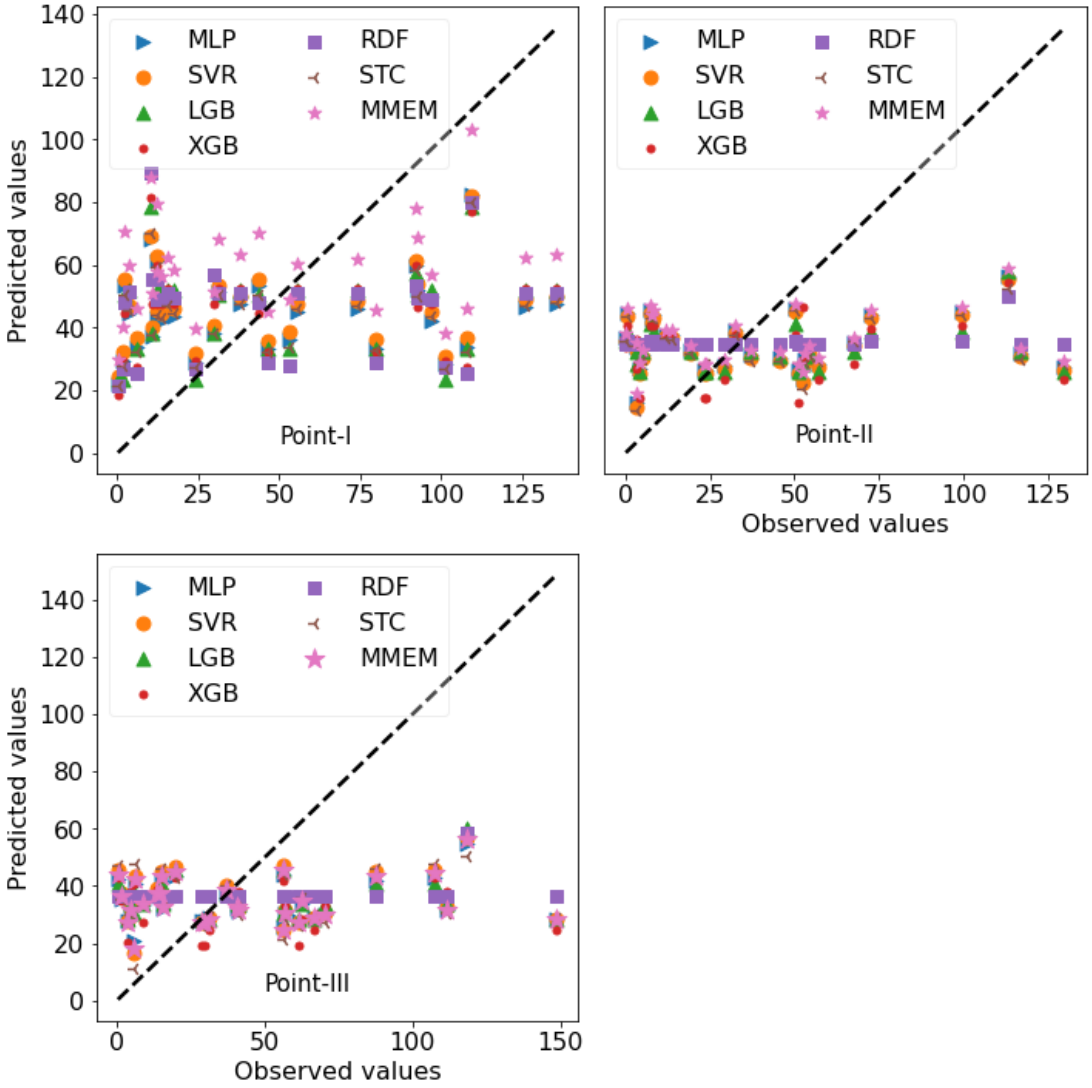


Fig. 8-5 Comparison of the resemblance of dry and wet events produced by regression models

From the comparison of the predicted monthly precipitation values and the skill tests, this study finds the regression models showing little improvement in reproducing the lower extremes than the MMEM and the reverse in reproducing the peak values (upper-extremes).

In reproducing the lower extremes, FBP showed better performance than the regression models and it was the same even for the upper extremes. Note that correction factors are employed in FBP predictions but no such correction factors are considered for the regression models. However, the reason behind the better performance of FBP to catch the dry events is due to

employing the correction factor or due to choosing the multiplicative seasonality function or combination of both that enabled FBP to reproduce wider seasonal variations compared to the regression models (Kourentzes, 2014). The skill test results show little weak prediction where IA values varied between 0.5 to 0.62 and MAE values varied 40 to 50. The main reason behind these weak skills is very frequent extreme peaks in the observed values (see Fig. 8-2 or 8-4) which are used to measure the skills. Another reason may be, comparatively shorter training period (1960-2005) where FBP was not familiar with the rainfall values above 250 mm in the target variable during the training process as they were set to 250 mm to remove the outliers from the observed datasets. Upon comparing Fig. 8-2, 8-3, 8-4, and 8-5, it can be seen that FBP and the regression models show a comparatively better resemblance of the lower extreme values (dry events) than the MMEM. Note that GCMs are imperfect replicas of real-world phenomena and contain systematic biases (Randall et al., 2007). They intend to overestimate the wet events and underestimate the dry events (Stephens et al., 2010; Sun et al., 2007). GCMs output needs rigorous correction before applications (Islam et al., 2014, 2011; Maurer and Hidalgo, 2008; Mehrotra and Sharma, 2010). This study reveals that precipitation prediction at the local level using any time-series models that will enable to use a combination of both GCMs and observed data would enhance the overall prediction accuracy. It may provide better skill in reproducing the dry events and a longer training period with a longer period of GCMs derived hindcast data will enhance the prediction accuracy that has been seen in the first case (Case-I).

Note that, this study used two different types of datasets (GCMs and observed) of the same time span and followed a supervised training approach where known observed values are used for both the training and prediction period (for comparison with the predicted values). As no real future data was predicted here, one can consider the training period as verification and assessed prediction skills as validation. Every individual future prediction model, either ANN or any of ML algorithms, is different and shows different performance depending on their functions, tuning parameters, and variables considered for predictions. In addition, every individual precipitation time series are different at different geographical locations. For the prediction purpose, researchers used different models with different input variables, and data pre-processing techniques. Then performances are assessed based on the corresponding observed values. That is why it is difficult to compare and contrast different types of models for different regions and periods with the results obtained in this study.

Early prediction of upper and lower extremes can help in managing floods and droughts respectively. However, in this study, it is observed that neither the FBP nor the ML regression model could reproduce extreme wet events. Rather, they could reproduce the dry events considerably better than the wet events. Reproducing the dry events is also important to agriculture-dependent countries like Australia, where the most variable climate exists. In Australia, a typical major drought in a season may reduce agricultural production by about 10% and gross national product by 1% (White, 2000). This study will be beneficial for water resources managers for assessing the future water availability, managing agriculture, agricultural-dependent business, and other water-related stakeholders for planning and development of infrastructure.

8.4 Conclusions

Precipitation prediction is highly important from both social and economic perspectives. Predictions through GCMs and other time-series models have been seen individually, but the application of GCMs derived precipitation together with the observed values in a time-series model has not been found yet. On the contrary, as a relatively new time-series forecasting library, the application of Facebook Prophet in precipitation prediction can hardly be found in the literature. Time-series data, which has strong seasonality, Facebook Prophet works well for their future prediction. As the climate variables show seasonality over the cycle of a year, this study aimed to predict monthly precipitation, using the combination of both GCMs derived and observed data through the Facebook Prophet Model. In doing so, this study used historically observed monthly precipitation as input and GCMs derived monthly precipitation of CMIP5 decadal experiment as the additional regressor. Multiple additional regressors are implemented and compared with the performance of a single additional regressor. Few correction factors are introduced for the predicted values of different months that enabled FBP to provide better prediction accuracy. From the comparison of skills, this study finds that the combination of both GCMs derived and observed values gives better prediction accuracy compared to the prediction based on the observed data only. It is due to using GCMs derived data as the additional regressor that guided the FBP model in future prediction. GCMs derived data includes not only the scientific understanding but also the historical records that guided FBP to come up with higher prediction accuracy. Based on the outlined skill assessments the following conclusions are drawn

(i) FBP can reproduce the dry events considerably better than the wet events that may be due to a better understanding of FBP over the dry periods through the training and its multiplicative seasonality function.

(ii) Following the combination of GCMs derived data (as additional regressor) and the corresponding observed values FBP would be able to reproduce the future precipitation with more prediction accuracy than only the prediction based on the observed values.

(iii) A higher number of regressor will provide comparatively better prediction accuracy than a single additional regressor. In this case, a longer period of GCMs hindcast data would enhance the higher prediction accuracy.

However, this study compared the performance of FBP with six regression models, for the same places and same datasets, and finds FBP outperformed. This study highly encourages cross-validation of a similar approach by using different forms and architecture of deep ANN that may increase the prediction accuracy by utilizing their different tunable features.

Acknowledgments

This project was supported by the CIPRS scholarship of Curtin University and Data61 student scholarship of CSIRO (Commonwealth Scientific and Industrial Research Organisation) which is jointly provided to the first author for his Ph.D. study at Curtin University, Australia. The authors would like to thank the working groups of CMIP5 for producing the model data and making it available for the researchers. The authors also acknowledge the support of the Australian Bureau of Meteorology for providing the observed gridded data and catchment's shapefiles.

List of symbols

C	:	Mean (over the total time span) of the observed values
P	:	Model predicted values
O	:	Observed values
$P - C$	:	Anomaly of the model values
$\overline{P - C}$	:	Mean of the model anomalies
$O - C$	:	Observed anomaly
$\overline{O - C}$	:	Mean of the observed anomalies

References

Ali, M., Deo, R.C., Downs, N.J., Maraseni, T., 2019. Monthly rainfall forecasting with markov chain monte carlo simulations integrated with statistical bivariate copulas, Handbook of Probabilistic Models. Elsevier Inc. https://doi.org/10.1016/B978-0-12-816514-0.00003-5

Apurv, T., Mehrotra, R., Sharma, A., Goyal, M.K., Dutta, S., 2015. Impact of climate change on floods in the Brahmaputra basin using CMIP5 decadal predictions. Journal of Hydrology 527, 281–291. https://doi.org/10.1016/j.jhydrol.2015.04.056

BoM, 2020. Annual rainfall . _ State of the Environment (Department of Environment and Science).

Chakraverty, S., Gupta, P., 2008. Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Computing and Applications 17, 187–192. https://doi.org/10.1007/s00521-007-0093-y

Chattopadhyay, S., Chattopadhyay, G., 2008. Comparative study among different neural net learning algorithms applied to rainfall time series. Meteorological Applications 15, 273–280. https://doi.org/10.1002/met.71

Climate-Data, 2020. Brisbane climate: Average weather, temperature, rainfall [Available at https://www.climatestotravel.com/climate/australia/brisbane].

Committee, A.S. of C.E.T., 2000. Artificial Neural Networks in Hydrology. II: Hydrologic Applications By the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Journal Of hydrologic engineering 5, 115–123.

Dastorani, M., Mirzavand, M., Dastorani, M.T., Sadatinejad, S.J., 2016. Comparative study among different time series models applied to monthly rainfall forecasting in semi-arid climate condition. Natural Hazards 81, 1811–1827. https://doi.org/10.1007/s11069-016-2163-x

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

George, J., Janaki, L., Parameswaran Gomathy, J., 2016. Statistical Downscaling Using Local Polynomial Regression for Rainfall Predictions – A Case Study. Water Resources Management 30, 183–193. https://doi.org/10.1007/s11269-015-1154-0

Hansen, J.W., Mason, S.J., Sun, L., Tall, A., 2011. Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture 47, 205–240. https://doi.org/10.1017/S0014479710000876

Hong, W.C., 2008. Rainfall forecasting by technological machine learning models. Applied Mathematics and Computation 200, 41–57. https://doi.org/10.1016/j.amc.2007.10.046

Hossain, I., Rasel, H.M., Imteaz, M.A., Mekanik, F., 2020. Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics 132, 131–141. https://doi.org/10.1007/s00703-019-00679-4

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., 2021a. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale. Indian Water Resources Society I, 285.

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021b. Intercomparison of drift correction alternatives for CMIP5 decadal precipitation. International Journal of Climatology joc.7287. https://doi.org/10.1002/joc.7287

Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K., 2009. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences 13, 1413–1425. https://doi.org/10.5194/hess-13-1413-2009

IPCC, 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, Core Writing Team, R.K. Pachauri and L.A. Meyer. https://doi.org/10.1017/CBO9781107415324.004

Islam, S.A., Bari, M., Anwar, A.H.M.F., 2011. Assessment of hydrologic impact of climate change on Ord River catchment of Western Australia for water resources planning: A multi-model ensemble approach, in: Chan, F., Marinova, D. and Anderssen, R.S. (Eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand (MSSANZ), Inc. https://doi.org/10.36334/modsim.2011.I6.islam

Islam, S.A., Bari, M.A., Anwar, A.H.M.F., 2014. Hydrologic impact of climate change on Murray–Hotham catchment of Western Australia: a projection of rainfall–runoff for future water resources planning. Hydrology and Earth System Sciences 18, 3591–3614. https://doi.org/10.5194/hess-18-3591-2014

Jones, J.W., Hansen, J.W., Royce, F.S., Messina, C.D., 2000. Potential benefits of climate forecasting to agriculture. Agriculture, Ecosystems & Environment 82, 169–184. https://doi.org/10.1016/S0167-8809(00)00225-5

Jones, P.W., 1999. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Khandelwal, I., Adhikari, R., Verma, G., 2015. Time series forecasting using hybrid arima and ann models based on DWT Decomposition. Procedia Computer Science 48, 173–179. https://doi.org/10.1016/j.procs.2015.04.167

Kourentzes, N., 2014. Additive and multiplicative seasonality – can you identify them correctly? – Nikolaos Kourentzes.

Lee, J., Kim, C.G., Lee, J.E., Kim, N.W., Kim, H., 2018. Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea. Water (Switzerland) 10. https://doi.org/10.3390/w10101448

Lin, S.S., Zhang, N., Zhou, A., Shen, S.L., 2022. Time-series prediction of shield movement performance during tunneling based on hybrid model. Tunn. Undergr. Sp. Technol. 119,

104245. https://doi.org/10.1016/j.tust.2021.104245

Machiwal, D., Jha, M.K., 2012. Hydrologic Time Series Analysis: Theory and Practice. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-1861-6

Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. Hydrology and Earth System Sciences 12, 551–563. https://doi.org/10.5194/hess-12-551-2008

Mehrotra, R., Sharma, A., 2010. Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. Water Resources Research 46, 1–17. https://doi.org/10.1029/2009WR008423

Mehta, V.M., Knutson, C.L., Rosenberg, N.J., Olsen, J.R., Wall, N.A., Bernadt, T.K., Hayes, M.J., 2013. Decadal Climate Information Needs of Stakeholders for Decision Support in Water and Agriculture Production Sectors: A Case Study in the Missouri River Basin. Weather, Climate, and Society 5, 27–42. https://doi.org/10.1175/WCAS-D-11-00063.1

Meinke, H., Sivakumar, M.V.K., Motha, R.P., Nelson, R., 2007. Preface: Climate predictions for better agricultural risk management. Australian Journal of Agricultural Research 58, 935–938. https://doi.org/10.1071/ARv58n10_PR

Mekanik, F., Lee, T.S., Imteaz, M.A., 2011. Rainfall modeling using Artificial Neural Network for a mountainous region in west Iran. MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty 3518–3524. https://doi.org/10.36334/modsim.2011.i5.mekanik

Mislan, Haviluddin, Hardwinarto, S., Sumaryono, Aipassa, M., 2015. Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia. Procedia Computer Science 59, 142–151. https://doi.org/10.1016/j.procs.2015.07.528

Ouyang, Q., Lu, W., Xin, X., Zhang, Y., Cheng, W., Yu, T., 2016. Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. Water Resources Management 30, 2311–2325. https://doi.org/10.1007/s11269-016-1288-8

Rajeevan, M., 2001. Prediction of Indian summer monsoon: Status, problems and prospects. Current Science 81, 1451–1457.

Ramírez, M.C., Ferreira, N.J., Velho, H.F.C., 2006. Linear and Nonlinear Statistical Downscaling for Rainfall Forecasting over Southeastern Brazil. Weather and Forecasting 21, 969–989. https://doi.org/10.1175/WAF981.1

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E., 2007. Climate Models and Their Evaluation.

Samal, K.K.R., Babu, K.S., Das, S.K., Acharaya, A., 2019. Time series based air pollution forecasting using SARIMA and prophet model. ACM International Conference Proceeding Series 80–85. https://doi.org/10.1145/3355402.3355417

Shen, S.L., Zhang, N., Zhou, A., Yin, Z.Y., 2022. Enhancement of neural networks with an alternative activation function tanhLU. Expert Syst. Appl. 199, 117181. https://doi.org/10.1016/j.eswa.2022.117181

Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J.C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., Haynes, J., 2010. Dreary state of precipitation in global models. Journal of Geophysical Research Atmospheres 115, 1–14. https://doi.org/10.1029/2010JD014532

Subashini, A., K, S., Saranya, S., Harsha, U., 2019. Forecasting Website Traffic Using Prophet Time Series Model. International Research Journal of Multidisciplinary Technovation 56–63. https://doi.org/10.34256/irjmt1917

Sun, Y., Solomon, S., Dai, A., Portmann, R.W., 2007. How often will it rain? Journal of Climate 20, 4801–4818. https://doi.org/10.1175/JCLI4263.1

Taylor, S.J., Letham, B., 2018. Forecasting at Scale. American Statistician 72, 37–45. https://doi.org/10.1080/00031305.2017.1380080

Toharudin, T., Pontoh, R.S., Caraka, R.E., Zahroh, S., Lee, Y., Chen, R.C., 2020. Employing long short-term memory and Facebook prophet model in air temperature forecasting. Communications in Statistics: Simulation and Computation 0, 1–24. https://doi.org/10.1080/03610918.2020.1854302

Unnikrishnan, P., Jothiprakash, V., 2020. Hybrid SSA-ARIMA-ANN Model for Forecasting Daily Rainfall. Water Resources Management 34, 3609–3623. https://doi.org/10.1007/s11269-020-02638-w

White, B., 2000. The Importance of Climate Variability and Seasonal Forecasting to the Australian Economy. Springer, Dordrecht, pp. 1–22. https://doi.org/10.1007/978-94-015-9351-9_1

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

Wu, X., Hongxing, C., Flitman, A., Fengying, W., Guolin, F., 2001. Forecasting Monsoon Precipitation Using Artificial Neural Networks. Advances in Atmospheric Sciences 18, 949–958. https://doi.org/10.1007/bf03403515

Zhang, P.G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

CHAPTER 9

MONTHLY PRECIPITATION PREDICTION FOR DECADAL TIMESCALE USING BIDIRECTIONAL LSTM AND CMIP5 NEAR-TERM EXPERIMENT DATA

Abstract

Early prediction of precipitation enables efficiently managing floods, water resources, agriculture and harvesting, and infrastructure development by providing a longer time for the proper planning and action. Climate models such as general circulation models (GCMs) predict climate variables including precipitation for the global scale based on the scientific understanding whereas stochastic or statistical models including different forms of artificial neural network (ANN) predict precipitation at the local levels based on the historical observed data only. For the precipitation prediction at a local level, the applications of both GCMs derived precipitation and corresponding observed values through the ANN models are not in practice whilst the use of decadal precipitation data from the Coupled Model Intercomparison Project phase-5 (CMIP5) for the same has not been seen yet. This is the first study that aimed to predict monthly precipitation for a decadal timescale through the ANN model using both the CMIP5 decadal precipitation and corresponding observed values. For this, monthly hindcast precipitation from five GCMs (MIROC4h, EC-EARTH, MRI-CGCM3, MPI-ESM-LR, and MPI-ESM-MR) of CMIP5 decadal experiment are collected from the CMIP5 data portal, and observed monthly gridded precipitation data were collected from the Australian Bureau of Meteorology. The arithmetic mean of the selected models (MMEM) was used as a feature and the corresponding observed value was used as the target variable of a Bidirectional LSTM, a form of ANN models where supervised training approaches were employed. The predicted results were compared with the observed as well as MMEM of the CMIP5 through different skill tests. Results revealed that the combination GCMs derived and observed precipitation in the ANN model following the supervised training approach could give better prediction accuracy, especially for the dry events compared to MMEM of CMIP5. This study also reveals that GCMs derived hindcast data of longer training period, which contains more than 1000

This chapter will be submitted as: Hossain, M.M., Anwar, A.H.M.F., Garg, N., Prakash, M., Bari, M., 2022. Monthly rainfall prediction for decadal timescale using Bidirectional LSTM and CMIP5 near-term experiment data. *Journal of Advances in Modelling Earth Systems* (to be submitted).

events, following the supervised training approach of the BiLSTM model would enhance the prediction accuracy and can be considered as an alternative to existing bias correction methods.

Keywords: Precipitation, prediction, decadal, cmip5, Bidirectional, LSTM

9.1 Introduction

Early prediction of precipitation has many positive benefits though it is a very difficult task in terms of accuracy. The large variations of precipitation over time and space, due to the impacts of ongoing climate change, imposed additional complexity on the accurate prediction of this valuable natural resource. Prediction of precipitation has two main categories; (i) knowledge-driven approach, and (ii) data-driven approach. For the precipitation prediction, climate models such as General Circulation Models (GCMs) use the former approach whilst some stochastic, statistical, and empirical models use the data-driven approach. While predicting the future precipitation, GCMs consider laws of thermodynamic balance, physics, scientific understanding of the hydrological process, and the interaction among ocean, earth, and atmosphere. However, GCMs predict the climate variables for the entire globe and of coarse spatial resolutions. The spatial resolution of approximately 100-250 km is not adequate for local or catchment level studies because of the lack of information (Fowler et al., 2007; Grotch and MacCracken, 1991; Salathé, 2003). The use of the regional climate models (RCMs) for transferring the GCMs derived information to local levels is prevalent nowadays but computationally intensive and may not easily be available. In addition, GCMs are not perfect enough and their outputs contain systematic biases (Randall et al., 2007) that need rigorous correction before any application (Islam et al., 2014, 2011; Maurer and Hidalgo, 2008; Mehrotra and Sharma, 2010). Sun et al. (2007) reported that GCMs tend to overestimate the number of wet events and underestimate the intense events. This was also confirmed by Stephens et al. (2010), who found model precipitation is approximately double the observed value. To assess the future climate change impacts on water resources, it is necessary to correct the model biases (Liang et al., 2008).

On the contrary, some stochastic/statistical/empirical models and applications of machine learning/deep learning algorithms of which different forms and architecture of artificial neural networks (ANN) use the data-driven approach (Mekanik et al., 2011) for the prediction at the local or catchment levels. Stochastic or statistical models are mainly based on different forms

of regression analysis such as simple regression analysis (SRA), exponential smoothing, decomposition, and auto-regressive integrated moving average (ARIMA). Every individual stochastic or statistical method has its strengths and weaknesses. For instance, ARIMA is a popular stochastic model for time series prediction with greater flexibility. But, as a stochastic model, it needs stationarity of the data (Machiwal and Jha, 2012) and its presumed linear form of the associated data sometimes makes it inappropriate for complex nonlinear time series data like rainfall (Zhang, 2003). However, technological development in combination with the research innovations in this modern arena enhanced the computation facility that enabled higher accuracy of precipitation prediction of which ANN is the best example. Applications of the machine and deep learning algorithms, of which ANN of the different forms of architecture, have been popular for many time series predictions including time series of precipitation (Hung et al., 2009; Lee et al., 2018; Meinke et al., 2007; Mekanik et al., 2011; Mislan et al., 2015) because of its enhanced prediction accuracy. ANN is capable of modelling complex nonlinear real-world problems. Based on the level of complexity, ANN can be combined with different types of algorithms due to its highly flexible character. However, based on the need and opportunities, different researchers have come up with different research interests and periods with the application of ANN. Though the ANN is good to capture the nonlinear relationship of data, the presence of outliers in the time-series data can critically affect the reliability of ANN as it is a grey box model (Unnikrishnan and Jothiprakash, 2020). Thus ANN requires proper data pre-processing before its application, especially for the climate data (Committee, 2000; Ramírez et al., 2006). However, using only the historical data, after splitting it into training and test set (for validation), as the input of ANN is a very common practice (Abbot and Marohasy, 2012; Hossain et al., 2020; Mekanik et al., 2011). In this case, the ANN model will only learn the historical change, variability, and relationship between input datasets throughout the training period and that historical relationship may or may not reflect in the future as the future changing rate will be higher (IPCC, 2014) than the past.

Climate change is an ongoing dynamic process that is being changed continuously and will continue to change in the future. However, the rate of future climate change and its potential impact on precipitation is uncertain. Compared to the other climate variables, precipitation has been affected more due to the ongoing climate change. Changes in precipitation pattern, seasonal shifting, longer differences between the dry and wet periods, longer dry spells, and extreme wet events along with overall reduction of total precipitation amount have been

observed around the globe in the last few decades. In the last decades, these changes have been intensified due to the ongoing climate change (IPCC, 2014). According to the IPCC report, the change in the future precipitation amount and its extreme events (e.g., heavy rainfall, droughts) will be higher compared to the past depending on the geographical locations. As every year, the climate condition is being changed and it would be intensified in the future, precipitation prediction must not rely on the approaches based on the historically observed data only. Applications of ANN of different forms of architecture have been seen in many previous types of research where only the historically observed data were used as input and followed the data-driven approach only. Through Facebook Prophet (FBP) model, Chapter 8 demonstrated that the inclusion of GCMs derived precipitation together with the observed data provides better prediction accuracy compared to the prediction based on the observed data only. To the best of our knowledge, until now, no study used both the historically observed and GCMs derived hindcast (knowledge-driven) data as the input of the ANN model. However, Chapter 8 demonstrated that the inclusion of GCMs derived precipitation together with the observed data results in better prediction accuracy compared to the prediction based on the observed data only. Chapter 8 also suggested using a deep neural networks model to improve the prediction accuracy. To incorporate the GCMs derived precipitation in the local level prediction and following the suggestion made in chapter 8, this chapter aimed to predict the monthly precipitation in decadal timescale by ANN model using both the historically observed and CMIP5 decadal experiment precipitation data.

9.2 Data and methods

9.2.1 Data collection

The observed monthly gridded precipitation of 0.05° × 0.05° (5km × 5km) spatial resolution for entire Australia were collected from the Australian Bureau of Meteorology (BoM). The BoM has produced the gridded data using the Australian Water Resources Assessment Landscape model (AWRA-L V5) (Frost et al., 2016).

Monthly hindcasts precipitation data of decadal time scale from five GCMs (EC-EARTH MIROC4h, MRI-CGCM3, MPI-ESM-LR, and MPI-ESM-MR) were downloaded from CMIP5 data portal (https://esgf-node.llnl.gov/projects/cmip5/). The reason behind selecting these five models is their arithmetic mean was found as the best combination as multi-model ensembles

mean (MMEM) in chapter seven. Data for the period 1960-2005, initialized at 1960, 1965, 1975… thus every five years up to 2005 are used in this study. There are two core sets of CMIP5 decadal experiment; 10-year and 30-year simulation. In this study, only the 10-year simulation data are selected as it was found comparative better as opposed to 30-year simulation (Hossain et al., 2021b). The details of the selected models are presented in Table 9-1.

Table 9-1 List of models, their initializations, and number of ensembles used in this study

Model	Resolutions (°lon × °lat))	Initialization Year (1960-2005)									
		60	65	70	75	80	85	90	95	00	05
		Number of ensembles									
MIROC4h	(0.5625 X 0.5616)	03	03	03	06	06	06	06	06	06	06
EC-EARTH	(1.125 X 1.1215)	14	14	14	14	14	14	14	14	10	18
RI-CGCM3	(1.125 X 1.1215)	06	08	09	09	06	09	09	09	09	06
MPI-ESM-LR	(1.875 X 1.865)	10	10	10	10	10	10	10	10	10	10
MPI-ESM-MR	(1.875 X 1.865)	03	03	03	03	03	03	03	03	03	03

9.2.2 Data Processing

In the first step, all the available ensembles of individual initializations were averaged to produce a single dataset and then were subsets for the Australian region. Secondly, the averaged ensembles were spatially interpolated using the second-order conservative (SOC) method onto 0.05°×0.05° spatial resolution thus matching with the grids used in the observed data. This study used SOC as it conserves the precipitation flux while sub-gridding the GCMs data (Jones, 1999) and marked it as a better spatial interpolation method, especially for the GCMs derived gridded dataset (Hossain et al., 2021a). Then both the models' and observed datasets were subsets for the Brisbane River catchment. Every initialization spans a dataset of 10 years that overlaps five years with the dataset of the next initializations. In the third step, the last five years of each initialization except 2005 were discarded and the first five years from each initialization were combined to produce a single time series from 1961 to 2015. For the initialization year 2005 (2006-2015), the whole dataset was taken instead of the first five years to make the dataset longer. For the rest of this chapter, this dataset will be referred to as MMEM of CMIP5. A Box-cox power transformation, from Scikit-learn power transformation, was employed for both the MMEM of CMIP5 and the observed dataset.

9.2.3 Model description

Long Short-Term Memory (LSTM) is a special kind of Recurrent Neural Network (RNN) that is capable of learning long-term dependencies. RNNs are the first kind of neural network that can remember the previous input in memory and due to the vanishing gradient problem; they cannot handle the lengthy sequences. LSTM was initially proposed by Hochreiter and Schmidhuber (1997) and then was refined and popularized by many researchers (see Greff et al., 2017 and references therein). To avoid such long-term dependency problems faced by RNNs, LSTMs are explicitly designed to maintain information for longer periods in their memory cells. However, LSTMs are easy to overfit, sensitive to different random weight initialization, and require a longer dataset for the training.

A bidirectional LSTM (BiLSTM) is a redesign of traditional LSTMs. As the name suggests, BiLSTM offers to learn in both forward and backward directions. In recent years, a BiLSTM model has been investigated and found it provides better prediction accuracy by offering additional training capability through receiving information from past (backward) and future (forward) instances simultaneously (Abduljabbar et al., 2021; Siami-Namini et al., 2019). This study used Keras's sequential model of three BiLSTM (henceforth referred to as BiLSTM) hidden layers (Fig. 9-1) to predict the monthly precipitation for a decade (2006-2015) by using MMEM of CMIP5 decadal precipitation as a feature and the corresponding observed monthly values as the target variable. The reason behind choosing the BiLSTM is to get the higher prediction accuracy (Ezen-Can, 2020) from our comparatively shorter training dataset (Abduljabbar et al., 2021; Siami-Namini et al., 2019). Two sets of observed rainfall data are used here; (i) observed monthly precipitation without removing outliers (referred to as BiLSTM) and (ii) observed monthly precipitation after removing outliers (referred to as BiLSTM*).

The 'tanh' activation function is used in all three hidden layers with three different dropouts: 0.2, 0.1, and 0.05 for the first, second, and third hidden layers respectively. Learning rate of 0.005, 'Adam'optimizer, 'mean_squared_error' loss function with accuracy metrics are used for compiling the model. After going through numerous approaches of parameter selection, based on trial and error basis, batch sizes of 12 and 1000 epochs are employed.

Training the ANN model before the prediction is a mandatory task by which the ANN model learns the relationship between the datasets. In the ANN models, splitting the observed data into train and test sets is a very common practice (Abbot and Marohasy, 2012). However, instead of using only the observed data after splitting it into training and test set, this study used a supervised training approach where BiLSTM models were trained (1961/Jan-2005/Dec) for the GCMs derived monthly precipitation (as a feature) and corresponding observed values (as a target). From this training period, the BiLSTM models learned the historical relationship between the GCMs derived precipitation and corresponding observed values (Chakraverty and Gupta, 2008). This trained BiLSTM model was provided MMEM of CMIP5 for the period 2006-2015 as a test-set for the prediction. BiLSTMs predicted monthly rainfall for the period of test-set data that is presented in the result section.

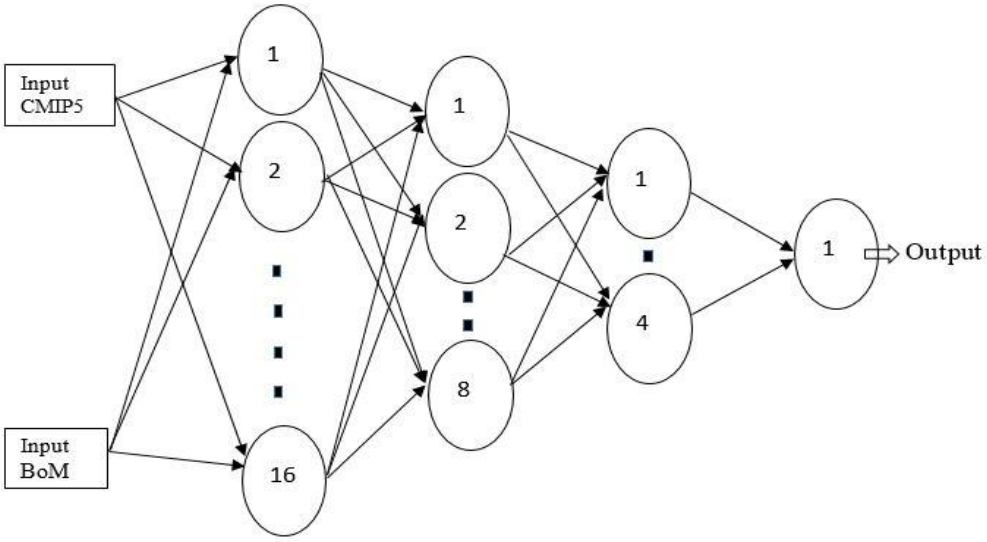


Fig. 9-1 The structure of the BiLSTM used in this study

9.2.4 Skill Tests

Pearson correlation coefficient (PCC)

PCC measures the linear correlation between two datasets. PCC is used to measure the linear correlation between the predicted and observed values. Its value varies between -1 and 1 (perfect correlation).

$$PCC = \frac{\sum_{i=1}^{N}(P_t-\bar{P})(O_t-\bar{O})}{\sqrt{\sum_{i=1}^{N}(P_t-\bar{P})^2}\sqrt{\sum_{i=1}^{N}(O_t-\bar{O})^2}} \qquad (9.1)$$

Where P and O present models' predicted and observed values respectively and the subscript i varies from 1 to N where N is the number of months in the prediction periods. These notations are the same also for the following skills.

Anomaly Correlation Coefficient (ACC)

A Wilks, (2011) suggested ACC that measures the correspondence between the anomalies of model-predicted and observed values. A higher ACC value does not represent the higher accuracy of the predicted data but the anomalies of the predictions.

$$ACC = \frac{\sum\{(P_i-\bar{O})-(\overline{P_i-\bar{O}})\}\times\{(O_i-\bar{O})-(\overline{O_i-\bar{O}})\}}{\sqrt{\sum(P_i-\bar{O})^2}\sqrt{\sum(O_i-\bar{O})^2}} \quad (9.2)$$

Here, $\bar{O}$ presents the decadal mean of the observed values and the bar over the anomalies presents the mean of them.

Index of Agreement (IA)

IA suggested by Wilmot (1982), measures the accuracy of model-predicted values corresponding to observed values. IA is bounded between 0 and 1 where, the closer the value to 1, the more efficient the prediction is

$$IA = 1 - \frac{\sum_{i=1}^{N}(P_i-O_i)^2}{\sum_{i=1}^{N}(|P_i-O'|+|O_i-O'|)^2} \quad (9.3)$$

Here O' presents the mean of every individual year of the predicted period.

Mean Absolute Errors (MAE)

As the name suggests, MAE presents the average magnitude of the differences between modelled and observed values.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \quad (9.4)$$

9.3 Data analysis and results

9.3.1 Training and test datasets

The hindcast precipitation from MMEM of CMIP5 and corresponding observed values are compared for both training (1961-2005) and prediction (test-set) period (2006-2015). Fig. 9-2

presents a comparison between MMEM of CMIP5 and the corresponding observed values for the training period. This comparison shows that MMEM could reproduce the seasonal variations with little to moderate variations in reproducing the wet events (events in December-January-February). However, MMEM was neither able to reproduce the dry events (events in June-July-August) nor the extreme wet events (events equal and above 200 mm). It was evident in the shape and distribution (Box-plot) along with the cumulative distribution function (CDF) of the data, used for both training (1961-2005) and test periods (2006-2015) for the BiLSTM model, presented in Fig. 9-3.

From the comparisons, the noticeable differences in the minimum, median, maximum, and interquartile ranges are evident between the observed and the MMEM of the CMIP5. In the observed data, a huge number of outliers and a long range of upper quartiles' values do exist whereas neither such outliers nor the lower and upper quartile values do exist in the MMEM of CMIP5. Both the Box-plot and the CDF show that MMEM was not able to reproduce the events of no and/or lower precipitation and the extreme higher precipitation events also (Fig. 9-3).

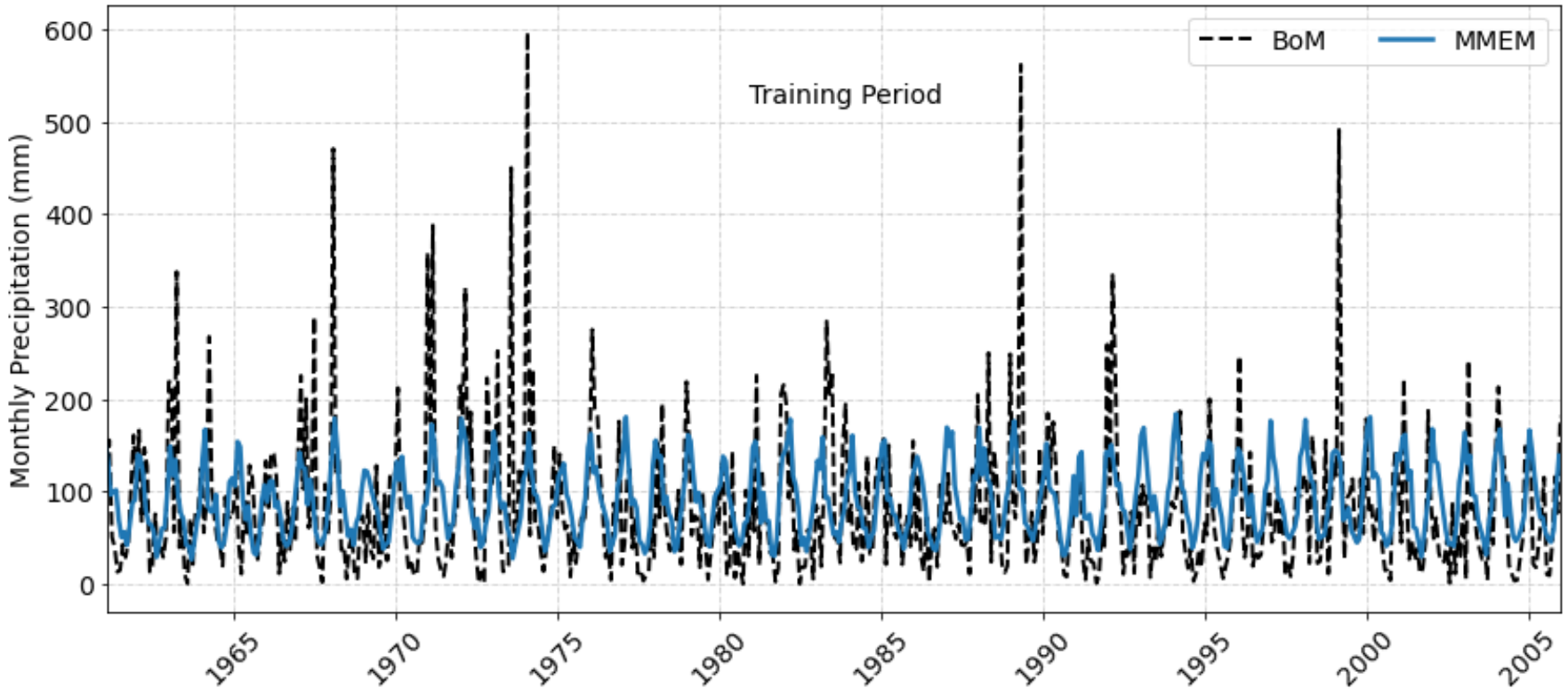


Fig. 9-2 Comparison of MMEM and corresponding observed values for the training period (January/1961-December/2005)

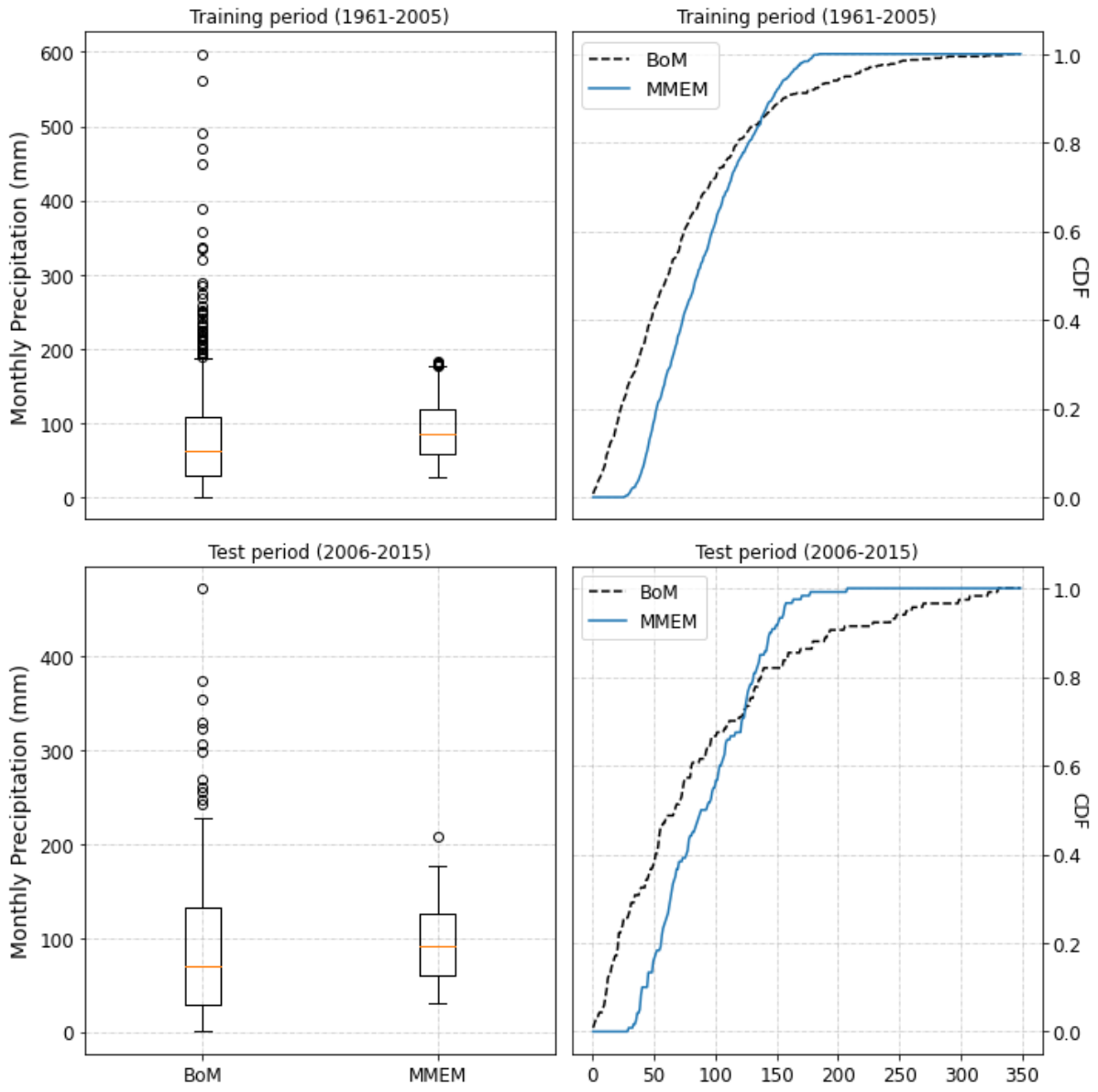


Fig. 9-3 Comparison of upper, lower quartiles along with interquartile ranges (Box plot) and the cumulative distribution function (CDF) for the training (1961-2005) and test set (2006-2015) of the data used in this study

9.3.2 BiLSTM vs MMEM of CMIP5 and observed values

The BiLSTM models' predicted values are compared mainly with the observed values and the MMEM of the CMIP5. The main reason to compare with the MMEM of CMIP5 is to investigate how the skills of MMEM of CMIP5 improve through the BiLSTM following a supervised learning approach. Fig. 9-4 presents the comparison between the predicted and corresponding observed values for the selected location. For clear visibility, prediction by using

BiLSTM is presented only in Fig. 9-4 but in other comparisons, both BiLSTM and BiLSTM* are presented

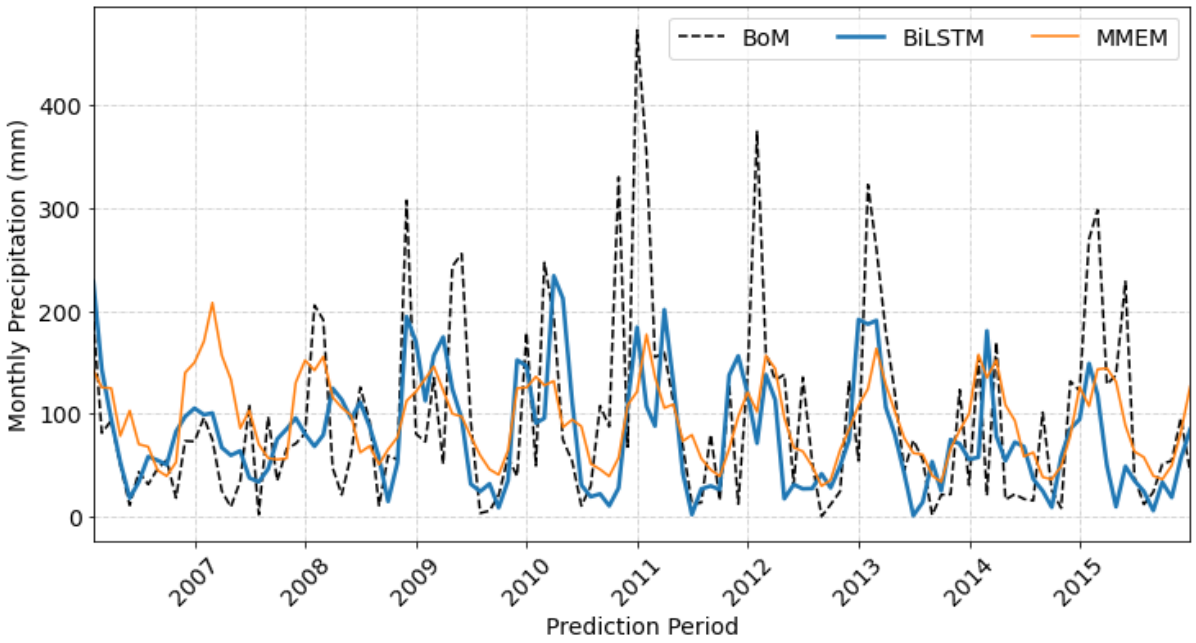


Fig. 9-4 Comparison of the predicted values by BiLSTM with the corresponding observed (BoM) and CMIP5 decadal experiment data (MMEM)

From the comparison, this study reveals that BiLSTM can reproduce the seasonal variations by capturing upper and lower extremes up to a certain limit. To reproduce the dry events, BiLSTM models are showing comparatively better resemblance with the observed values whereas no such resemblance was captured by the MMEM of the CMIP5 decadal dataset. In capturing the dry events, BiLSTM performed slightly better than BiLSTM* (not shown). The quartile ranges along with the CDF of the predicted values are presented in Fig. 9-5, where improvements in the interquartile ranges and the range of upper and lower quartiles values are evident. In this case, BiLSTM predicted values show a comparatively better resemblance with the observed values than the prediction from BiLSTM*. BiLSTM predicted values show almost similar lower quartile values and median as in the observed data but a little lower interquartile range with the lower number of outliers. However, both BiLSTM and BiLSTM* showed very similar performance for resembling the CDF (see Fig. 9-5).

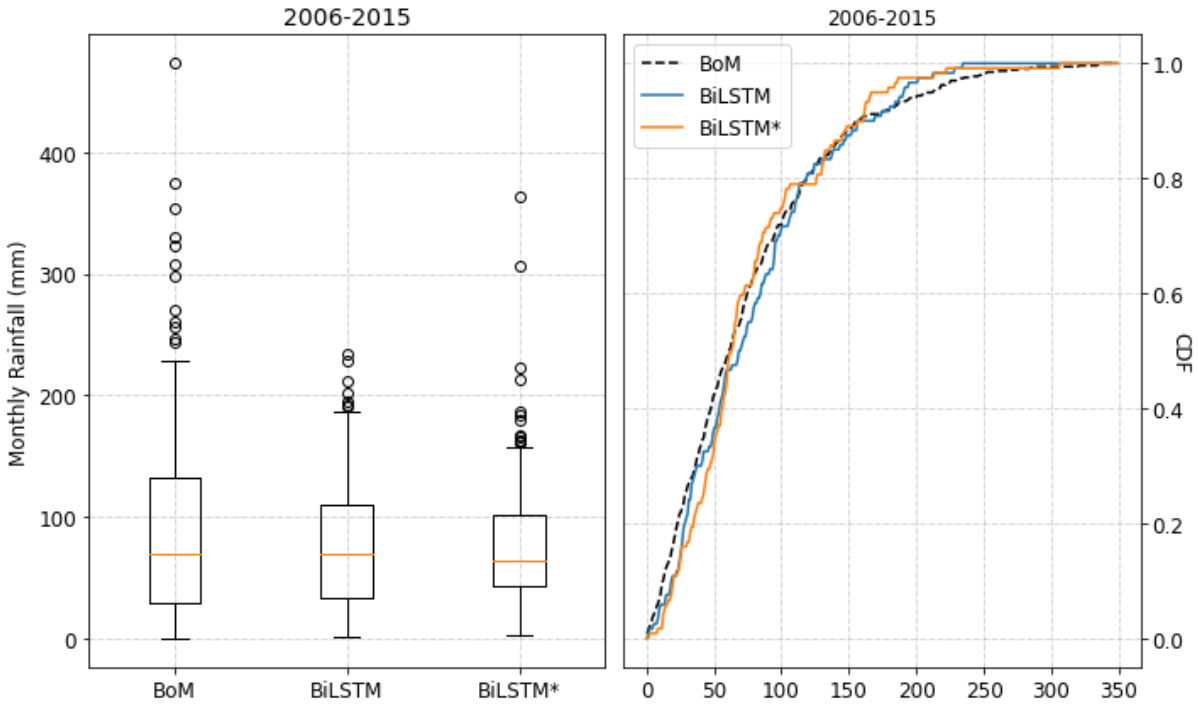


Fig. 9-5 Comparison of upper and lower quartiles along with interquartile range (Box plot) and cumulative distribution function (CDF) of the predicted and observed values

Over the prediction period, BiLSTM showed the best resemblance with the observed values in the first year whilst the lowest resemblance was observed in the last year. For this reason, this study assumes that the predicted values may show higher skills in the earlier years than the latter.

9.3.3 Comparison of skill tests

The aforementioned skill tests are performed for both the predicted values and the MMEM of CMIP5 decadal data based on the observed values. From the comparison of skills, it is observed that BiLSTM predicted values showed comparatively better resemblance with the observed values in the earlier years of the prediction period than the latter. To investigate this, this study calculated the skills for the first five years and for the entire prediction period (10 years) that are presented in Table 9-2. From the skill comparison of the predicted values, it can be seen that there is no significant difference between the skills of the five and 10-years predicted values whereas certain differences are observed between the skills of MMEM and the predicted values. While comparing the skill between the MMEM and the predicted values, noticeable differences can be noted for the first five years whereas no such difference can be seen for 10-

years skills except IA. In the first five years, MMEM of CMIP5 shows a comparatively lower skill score than the skill score for a total of ten years period. It seems, comparatively overestimations for the wet periods as well as wider gaps between MMEM and observed values in the dry events are the main reason for a lower-skill score in the first five years than the latter. Overall, both the BiLSTMs models improved the prediction skills where BiLSTM* was little better than BiLSTM. It may be due to removing the outliers from the training period.

Table 9-2 Skill comparison for the first five years and over the decadal time-span

Models	First 5 years				Total 10 years			
	PCC	ACC	IA	MAE	PCC	ACC	IA	MAE
MMEM	0.241	0.241	0.354	64.8	0.434	0.434	0.510	58.1
BiLSTM	0.426	0.425	0.585	57.4	0.425	0.425	0.572	59.8
BiLSTM*	0.437	0.433	0.612	58.8	0.437	0.447	0.603	59.8

9.3.4 Comparison for individual seasons

From the comparison between Fig. 9-3 and Fig. 9-5, it can be seen that the predicted values show noticeable improvement in the lower and upper quartiles with comparatively lower improvement in the interquartile range. The lower quartile values usually come from the dry events that occur during the winter (JJA) whilst the upper quartiles and the outliers come from wet periods that occur during the summer (DJF). To investigate the improvement in individual seasons, this study also compared the predicted values with the corresponding MMEM of CMIP5 through the scatter plots that are presented in Fig. 9-6. From the comparison, it is seen that the MMEM of CMIP5 overestimated the lower precipitation values during autumn and winter and underestimated the higher values. However, for the lower values or dry events, the predicted values from both datasets show considerably better correspondence with the observed values during autumn whereas no such improvement was observed for summer and spring (see Fig. 9-6).

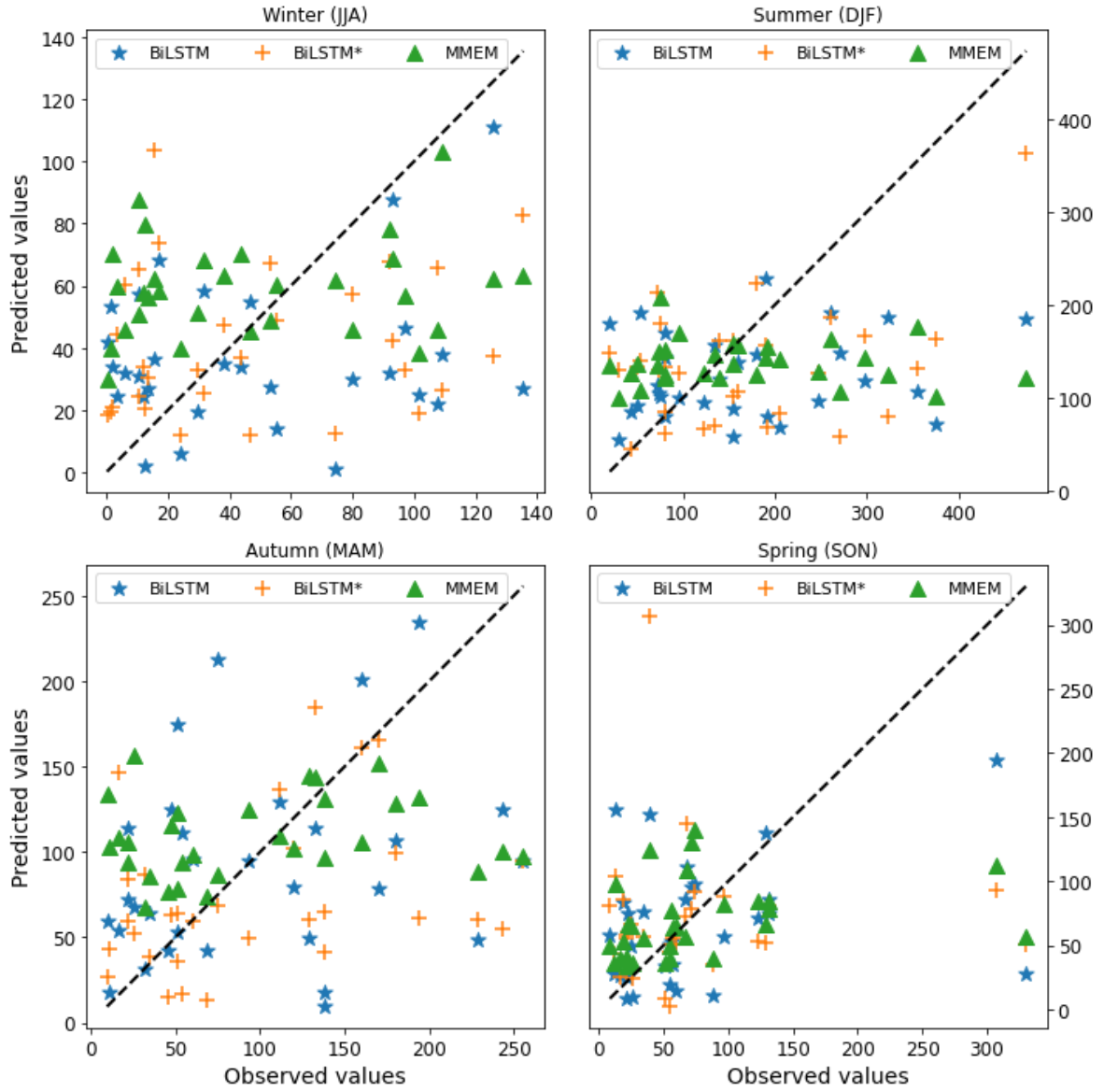


Fig. 9-6 Seasonal comparisons of the predicted values (monthly) with the corresponding MMEM of CMIP5

9.4 Discussion

This study predicted monthly precipitation for a decadal timescale using BiLSTM in which both the MMEM of CMIP5 decadal experiment data and the corresponding observed values are used as model input. The hindcast data of CMIP5 decadal monthly precipitation is available from 1961 until 2015 (55 years, 660-time steps). The First 45 Years (540-time steps) are used for training the BiLSTM models and the rest 10 years (2006-2015) are used for the test period. BiLSTM is a complex area of deep learning and a series of 540-time steps is not quite small

for the training period. However, a longer training period enables the ANN models better understand the relationship between the inputs that result in higher prediction accuracy. In this study, the prediction skills, PCC, and ACC values below 0.5, for both the five and ten-year prediction period, seem a bit lower scores. Numerous outliers from frequently extreme peak values, those were unable to catch by the BiLSTMs are the main reason for lower prediction skills. IA values around 0.6 along with a good response to dry and wet events are really promising, especially for the precipitation. In the first year of the prediction period, BiLSTM showed the best resemblance with the observed values whereas the reverse scenario is noted in the last year. This reveals that with the increasing length of the prediction period, the prediction skills may drop. Capturing dry events, which was not evident in the MMEM, would be very useful for the local level agriculture for dry periods and irrigation requirements, water managers for assessing the future water availability, and other stakeholders. However, every individual time series precipitation at different geographical locations is different. The researcher used different models with different input variables for rainfall prediction and their performances were assessed based on the observed values. Selection of model parameters, model architecture, and the period of the training and test dataset is open for discussion with the changing of locations. For this reason, it is difficult to compare and contrast different types of models for different regions and time series with the results obtained in this study.

In the previous studies, ANN models followed only the data-driven approach in which they were trained with only the observed data of two different periods after splitting into train and test sets (Abbot and Marohasy, 2012; Hossain et al., 2020; Mekanik et al., 2011). In that cases, ANN models were used to learn the temporal relationship between two different time-period of the observed values. This study is the first attempt in which the ANN model (BiLSTM) was trained with the GCMs derived precipitation (MMEM of CMIP5) and their corresponding observed values. The results reveal that BiLSTM is showing comparatively better prediction skills than the skill of MMEM of CMIP5. In other words, it can be said that BiLSTM reduced the prediction errors from MMEM that enhanced the prediction accuracy. In chapter 8 higher prediction skills were achieved through the FBP model when MMEM was used as an additional regressor, however, the best skills were achieved when all models were considered as eight additional regressor. This study (chapter 9) was limited to only MMEM as a feature where individual models were not considered. Upon comparing the prediction skills of FBP (in chapter 8) and BiLSTM, it is seen that, in all skill tests FBP is little better than BiLSTM. It

may be due to the correction factor that was employed in FBP but not in BiLSTM. These factors may be different for using FBP at different places for different datasets that would require cautions. In that case, users may need to utilize their prudence while using either FBP or BiLSTM or both.

As the GCMs contain systematic biases (Randall et al., 2007) that ANN models can read as the difference (or relation) between the GCMs' simulated and observed values. In this study, BiLSTM was trained with MMEM as a feature and corresponding observed values as a target that represents a supervised training approach. During the training period, BiLSTM learned the relationship (biases) between the MMEM of CMIP5 and the corresponding observed values, and based on this learning, BiLSTM predicted the future. From the obtained results and their corresponding skill tests, this study reveals that prediction skills of BiLSTM models are achieved by minimizing the biases in the MMEM of the CMIP5 dataset. The predicted values showed a comparatively better skill score than the MMEM of CMIP5, especially during the first five years of the prediction period where wider gaps between the observed and MMEM of CMIP5 exist (Fig. 9-4).

The fundamental problem with CMIP5 decadal simulation is the drift (Mehrotra et al., 2014) which was shown by Hossain et al. (2021c) for the precipitation at the Brisbane River catchment. Hossain et al. (2021c) quantified the drift of CMIP5 decadal hindcast precipitation and assessed the applicability of the widely used mean drift correction method proposed by ICPO (2011). Hossain et al. (2021c) reported that the mean drift correction method is not enough to alleviate the drift-bias (a time varying systematic bias of GCMs) and the improvement in the model skills after the drift correction is highly dependent on the model types. Taylor et al. (2012) recommended a sophisticated bias correction method to remove the drift-bias from CMIP5 decadal simulations. However, to correct the GCMs derived future rainfall, there is no standard bias correction approach (Taylor et al., 2012) but depending on locations, climate variables, time scales, and the application field of the data the correction approach may be different (Chen et al., 2013; Gangstø et al., 2013; Kruschke et al., 2016). Different transformation functions such as downscaling, quantile mapping (QM), histogram equalizing, and rank matching (Bates et al., 1998; Charles et al., 2004; Piani et al., 2010; Wood et al., 2004) stochastic and/or dynamic downscaling (Bates et al., 1998; Charles et al., 2004; Mehrotra and Sharma, 2006; Wilby et al., 1998) are commonly used in practice. Every

correction method has its strengths and weaknesses depending on its underlying assumptions. Among the many commonly used approaches (mentioned earlier), a statistical bias correction method proposed by Piani et al. (2010) has been used in many studies. Equalizing the cumulative distribution function (CDF) of the observed and simulated values was the main principle of that statistical method. In this study, the CDF of BiLSTM models' predicted values showed a very good resemblance with the CDF of the corresponding observed values (see Fig. 9-5) that indicate application of ANN models can be considered as a new dimension of drift correction approach for decadal data. In that case, the selection of the climate variable and their time span for the correction, ANN model types and their architecture, time span of training, and the test data set need further investigation.

9.5 Conclusion

The application of ANN models in precipitation prediction is an emerging research area as it enhanced the prediction accuracy by using the updated computational facility in this modern arena. Precipitation prediction for a couple of years using ANN models based on the historically observed values has been found in the literature. However, precipitation prediction for a decadal timescale is not common whereas incorporating GCMs derived data together with the observed values in the ANN model was not seen before. This study predicted for decadal timescale as it used decadal experiment data derived through GCMs in CMIP5. As GCMs follow the knowledge-driven approach considering many factors including future climate scenarios for the prediction, this study assumed that using the GCMs derived climate variables as features and corresponding observed values as target variables following a supervised training approach in ANN models will provide higher prediction accuracy by reducing the errors in GCMs derived values. This study also assumed that using BiLSTM would provide additional accuracy as it receives information from both the forward and backward direction simultaneously. Results revealed that using GCMs derived data together with the observed values through BiLSTM enhanced the prediction accuracy by reducing the gaps between the observed and GCMs values. This study also reveals that following a supervised training approach enables the BiLSTM models to reproduce the dry events better than the GCMs.

As the BiLSTM can capture the non-linear relationship between the features and the target variables through the supervised training, it can be used as a new approach of drift/bias

correction method. In that case, GCMs derived longer time hindcast data for training the BiLSTM would be useful but training period, correction period and model parameters need to be investigated further.

List of symbols

C	:	Mean (over the total time span) of the observed values
P	:	Model predicted values
O	:	Observed values
$P - C$	:	Anomaly of the model values
$\overline{P - C}$	:	Mean of the model anomalies
$O - C$	:	Observed anomaly
$\overline{O - C}$	:	Mean of the observed anomalies

References

Abbot, J., Marohasy, J., 2012. Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Advances in Atmospheric Sciences 29, 717–730. https://doi.org/10.1007/s00376-012-1259-9

Abduljabbar, R.L., Dia, H., Tsai, P.W., 2021. Unidirectional and bidirectional LSTM models for short-term traffic prediction. Journal of Advanced Transportation 2021. https://doi.org/10.1155/2021/5589075

Bates, B.C., Charles, S.P., Hughes, J.P., 1998. Stochastic downscaling of numerical climate model simulations. Environmental Modelling and Software 13, 325–331. https://doi.org/10.1016/S1364-8152(98)00037-1

Chakraverty, S., Gupta, P., 2008. Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Computing and Applications 17, 187–192. https://doi.org/10.1007/s00521-007-0093-y

Charles, S.P., Bates, B.C., Smith, I.N., Hughes, J.P., 2004. Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. Hydrological Processes 18, 1373–1394. https://doi.org/10.1002/hyp.1418

Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resources Research 49, 4187–4205. https://doi.org/10.1002/wrcr.20331

Choudhury, D., Sen Gupta, A., Sharma, A., Mehrotra, R., Sivakumar, B., 2017. An Assessment of Drift Correction Alternatives for CMIP5 Decadal Predictions. Journal of Geophysical Research: Atmospheres 122, 10282–10296. https://doi.org/10.1002/2017JD026900

Committee, A.S. of C.E.T., 2000. Artificial Neural Networks in Hydrology. II: Hydrologic Applications By the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Journal Of hydrologic engineering 5, 115–123.

Ezen-Can, A., 2020. A Comparison of LSTM and BERT for Small Corpus 1–12.

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27, 1547–1578. https://doi.org/10.1002/joc.1556

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

Gangstø, R., Weigel, A.P., Liniger, M.A., Appenzeller, C., 2013. Methodological aspects of the validation of decadal predictions. Climate Research 55, 181–200. https://doi.org/10.3354/cr01135

Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., Schmidhuber, J., 2017. LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems 28, 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Grotch, S.L., MacCracken, M.C., 1991. The Use of General Circulation Models to Predict Regional Climatic Change. Journal of Climate 4, 286–303. https://doi.org/10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hossain, I., Rasel, H.M., Imteaz, M.A., Mekanik, F., 2020. Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics 132, 131–141. https://doi.org/10.1007/s00703-019-00679-4

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., 2021a. Comparing Spatial Interpolation Methods for CMIP5 Monthly Precipitation at Catchment Scale. Indian Water Resources Society I, 285.

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021b. A comparative study on 10 and 30-year simulation of CMIP5 decadal hindcast precipitation at catchment level, in: Vervoort, R.W., Voinov, A.A., Evans, J.P. and Marshall, L. (Ed.), MODSIM2021, 24th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 609–615. https://doi.org/10.36334/modsim.2021.K5.hossain

Hossain, M.M., Garg, N., Anwar, A.H.M.F., Prakash, M., Bari, M., 2021c. Drift in CMIP5 decadal precipitation at catchment level. Stochastic Environmental Research and Risk Assessment 8, 5. https://doi.org/10.1007/s00477-021-02140-8

Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K., 2009. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences 13, 1413–1425. https://doi.org/10.5194/hess-13-1413-2009

ICPO, 2011. Data and bias correction for decadal climate predictions. CLIVAR Publication Series No. 150, 6 pp.

IPCC, 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, Core Writing Team, R.K. Pachauri and L.A. Meyer. https://doi.org/10.1017/CBO9781107415324.004

Islam, S.A., Bari, M., Anwar, A.H.M.F., 2011. Assessment of hydrologic impact of climate

change on Ord River catchment of Western Australia for water resources planning: A multi-model ensemble approach, in: Chan, F., Marinova, D. and Anderssen, R.S. (Eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand (MSSANZ), Inc. https://doi.org/10.36334/modsim.2011.I6.islam

Islam, S.A., Bari, M.A., Anwar, A.H.M.F., 2014. Hydrologic impact of climate change on Murray–Hotham catchment of Western Australia: a projection of rainfall–runoff for future water resources planning. Hydrology and Earth System Sciences 18, 3591–3614. https://doi.org/10.5194/hess-18-3591-2014

Jones, P.W., 1999. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Kruschke, T., Rust, H.W., Kadow, C., Müller, W.A., Pohlmann, H., Leckebusch, G.C., Ulbrich, U., 2016. Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms. Meteorologische Zeitschrift 25, 721–738. https://doi.org/10.1127/metz/2015/0641

Lee, J., Kim, C.G., Lee, J.E., Kim, N.W., Kim, H., 2018. Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea. Water (Switzerland) 10. https://doi.org/10.3390/w10101448

Liang, X.Z., Kunkel, K.E., Meehl, G.A., Jones, R.G., Wang, J.X.L., 2008. Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections. Geophysical Research Letters 35, 1–5. https://doi.org/10.1029/2007GL032849

Machiwal, D., Jha, M.K., 2012. Hydrologic Time Series Analysis: Theory and Practice. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-1861-6

Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. Hydrology and Earth System Sciences 12, 551–563. https://doi.org/10.5194/hess-12-551-2008

Mehrotra, R., Sharma, A., 2010. Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. Water Resources Research 46, 1–17. https://doi.org/10.1029/2009WR008423

Mehrotra, R., Sharma, A., 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. Journal of Geophysical Research Atmospheres 111, 1–16. https://doi.org/10.1029/2005JD006637

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. Journal of Hydrology 519, 2932–2951. https://doi.org/10.1016/j.jhydrol.2014.07.053

Meinke, H., Sivakumar, M.V.K., Motha, R.P., Nelson, R., 2007. Preface: Climate predictions for better agricultural risk management. Australian Journal of Agricultural Research 58, 935–938. https://doi.org/10.1071/ARv58n10_PR

Mekanik, F., Lee, T.S., Imteaz, M.A., 2011. Rainfall modeling using Artificial Neural Network for a mountainous region in west Iran. MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty 3518–3524. https://doi.org/10.36334/modsim.2011.i5.mekanik

Mislan, Haviluddin, Hardwinarto, S., Sumaryono, Aipassa, M., 2015. Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia. Procedia Computer Science 59, 142–151. https://doi.org/10.1016/j.procs.2015.07.528

Piani, C., Weedon, G.P., Best, M., Gomes, S.M., Viterbo, P., Hagemann, S., Haerter, J.O., 2010. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. Journal of Hydrology 395, 199–215. https://doi.org/10.1016/j.jhydrol.2010.10.024

Ramírez, M.C., Ferreira, N.J., Velho, H.F.C., 2006. Linear and Nonlinear Statistical Downscaling for Rainfall Forecasting over Southeastern Brazil. Weather and Forecasting 21, 969–989. https://doi.org/10.1175/WAF981.1

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman,

A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E., 2007. Climate Models and Their Evaluation.

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. International Journal of Climatology 23, 887–901. https://doi.org/10.1002/joc.922

Siami-Namini, S., Tavakoli, N., Namin, A.S., 2019. The Performance of LSTM and BiLSTM in Forecasting Time Series. Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019 3285–3292. https://doi.org/10.1109/BigData47090.2019.9005997

Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J.C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., Haynes, J., 2010. Dreary state of precipitation in global models. Journal of Geophysical Research Atmospheres 115, 1–14. https://doi.org/10.1029/2010JD014532

Sun, Y., Solomon, S., Dai, A., Portmann, R.W., 2007. How often will it rain? Journal of Climate 20, 4801–4818. https://doi.org/10.1175/JCLI4263.1

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Unnikrishnan, P., Jothiprakash, V., 2020. Hybrid SSA-ARIMA-ANN Model for Forecasting Daily Rainfall. Water Resources Management 34, 3609–3623. https://doi.org/10.1007/s11269-020-02638-w

Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J., Wilks, D.S., 1998. Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Research 34, 2995–3008. https://doi.org/10.1029/98WR02577

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed, International Geophysics. Elsevier, 676 pp.

Wilmot, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin American Meteorological Society 63, 1309–1313.

Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic Change 62, 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Zhang, P.G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

CHAPTER 10

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

10.1 Summary

CMIP5 introduced the decadal experiment, for the first time, to examine the predictive capabilities of the forecasting system on decadal timescales. Previous studies examined the fidelity of the decadal experiment over different climate variables of their different temporal and spatial scales. Most of the studies were, however, for the temperature and temperature-based climate indices. Quite limited research was conducted on decadal precipitation and no attention was paid to a catchment level for the spatial resolution finer than 0.5^0. This study assessed the CMIP5 decadal hindcast precipitation at a catchment level for a spatial resolution of 0.05^0 and then applied the assessment outcomes in the future prediction of monthly precipitation for a decadal timescale. It comprises three different phases in which a wide range of skill tests was employed for measuring, comparing, and assessing the skills. In the first phase, it was focused to investigate the suitable spatial interpolation method for re-gridding the GCMs derived precipitation from their native grids into the finer spatial resolution of 0.05^0 (chapter 3). It was also checked either 10 or 30-years simulations of the CMIP5 decadal experiment (chapter 4) work better for future prediction. In the second phase, it quantified the drifts of the models contributed to the CMIP5 decadal experiment, and their MMEM, for monthly and seasonal mean precipitation (aggregated from monthly data) and then assessed the suitability of widely used mean-drift correction method at a catchment level (chapter 5). Next, the drift correction alternatives were investigated for the seasonal mean precipitation and proposed a modified drift correction method for the seasonal data (chapter 6). In the last step of the second phase, models' performances were measured, categorized the models based on their performances over the entire catchment, and then sorted the optimum number of models for forming the best MMEM from both the temporal and spatial skills perspectives (chapter 7). In the third phase, this study predicted the monthly precipitation for a decadal timescale using Facebook Prophet, Machine Learning regression algorithms, and a Bidirectional LSTM where a combination of GCMs derived and observed data was employed (chapter 8 and 9). Prediction skills were measured and compared based on several skill tests. Each chapter has specific

conclusions that address the chapters' objectives. In this chapter, overall conclusions are drawn and some recommendations are made for further study.

10.2 Conclusions

Assessment of the GCMs' performances over every individual location is important to build confidence in their future predicted values and form their multi-model ensembles' mean. This study finds, GCMs with the higher atmospheric resolutions performed considerably better and the inclusion of GCMs derived values in the stochastic or time-series models along with the observed values for the future prediction enhances the prediction accuracy at the local or catchment levels. Individual chapters also presented specific conclusions while addressing the research questions. A summary of the conclusions is presented below.

- Spatial interpolations are commonly used in practice to transfer the GCMs derived coarse spatial resolution data to required finer spatial resolutions. The second-order conservative (SOC) method is an appropriate choice for re-gridding GCMs derived precipitation as it conserves the precipitation flux over the study area while re-gridding onto subsequent grids. It will help the researchers along with other potential users to sort out the best spatial interpolation methods, especially for the GCM derived precipitation in decadal scale.
- Due to the higher rate of ongoing climate change, GCMs predicted climate data of longer lead-time might contain higher uncertainty than the data of comparatively shorter lead-time. This study finds, the 10-year simulations of the CMIP5 decadal experiment show comparatively lower bias and higher skills as opposed to the 30-year simulations due to its shorter lead-time and higher number of ensembles. It reveals models' inability to adapt to the higher changing rate of future climate as well as a higher number of ensembles.
- Drift is the fundamental problem of the models for decadal predictions. This study finds drifts in monthly and seasonal (aggregated from the monthly values) hindcast precipitation of the CMIP5 decadal experiment and demonstrates that the mean-drift correction method is not sufficient to alleviate the drift. It refers to investigating drift correction alternatives. To minimize the drift up to a certain limit, aggregating monthly

values into seasonal mean and multi-model ensemble mean instead of individual models are suggested.

- Drift correction is the prerequisite work for the application of GCMs derived decadal climate variables. For the seasonal mean precipitation (aggregated from the monthly hindcast values), the Nested bias correction (NBC) method may work better if drifts are not too high but the modified drift correction (MDM) method proposed in this study found comparatively better for the models with higher drifts. The relative drift (RDT) correction method was found best while considering the correction for the total precipitation. Users need to utilize their prudence while selecting the drift correction methods based on their specific needs.
- Evaluation of GCMs' predicted historical data based on their corresponding observed values determines how well they represent the historical climate and thus develop confidence in their future predictions. This study finds that the ranking of models based on their skills and optimizing the number of models is an important requirement before forming a multi-model ensemble mean (MMEM). Models with higher atmospheric resolutions show comparatively better performance compared to the coarse spatial resolutions and MIROC4h outperformed all the selected models because of its high atmospheric resolutions. A combination of MIROC4h, MRICG-CM3, EC-EARTH, MPI-ESM-LR, and MPI-ESM-MR provides the best outcome as MMEM over the Brisbane River catchment. It will help the water manager, infrastructure developers, agricultural stakeholders in selecting the best models before making any decision in planning and developing infrastructures based on the models' predicted future values.
- Early and accurate predictions of precipitation have become very challenging as it is highly variable over time and space. For the precipitation prediction at the local level, considering GCMs derived data in addition to the observed values provide better prediction skills compared to the prediction based on the observed data only. Facebook Prophet (FBP) Model was found comparatively better over the Machine Learning regression models due to its wide range of tuneable features including seasonality functions.
- Application of GCMs derived precipitation as a feature and the corresponding observed values as target variables in a BiLSTM model following a supervised training approach enhance the prediction accuracy. The supervised training approach enabled the

BiLSTM model to reproduce the dry events comparatively better than the wet events. It reveals that a BiLSTM can be used as an alternate of drift/bias correction for a decadal timescale after following a supervised training approach for a longer training period.

Both the BiLSTM and FBP model showed comparatively better prediction skills than MMEM and the prediction based on the observed data only. Compared to wet events, both models performed considerably better in reproducing the dry events that may be due to following supervised training approach in BiLSTM, multiplicative seasonality function in FBP, and both models' capability to read the boundary limit (zero values) of dry events that was not possible for wet events. This study reveals a BiLSTM can be used as a sophisticated drift/bias correction method following a supervised training approach with a longer training period. Upon comparing the skills between FBP and BiLSTM, FBP was found little better BiLSTM. However, in reproducing the extreme wet events, BiLSTM seems performed better than FBP. These skills may change with the changes of locations for different datasets, different architecture, and features of models.

10.3 Limitations and the further recommendations

This study has been enclosed a comprehensive assessment through a wide range of skill tests for the CMIP5 decadal precipitation at the catchment level and predicted monthly precipitation for a decadal timescale using ANN and ML algorithms. For re-gridding the GCMs data from their native coarse to desired finer spatial resolutions, the spatial interpolation method was used. Though downscaling are prevalent, use of the spatial interpolation method is very common practice to avoid the complexity and time consuming. Downscaling requires rigorous computational facility as well as many other climate variables, which are not the scope of this study. Due to the limited scope and specified period, this study used spatial interpolation methods and was limited to only the Brisbane River catchment.

For the future prediction, Facebook Prophet and BiLSTM models are used and both of which have huge tuneable features. Though most of the parameter values were optimized, however, still there are a few time series models with different features, and other issues can be resolved in future studies which are listed below.

1. Using the downscaled data may provide a different presentation of the model performances over the catchment that can be considered in further study. A comparative study between downscaling and spatially interpolated precipitation data following the SOC method can be carried out to justify the model performance and their categories.
2. Other catchments at different hydrological regions in Australia and also in different locations around the globe can be carried out for the cross-validation of this study.
3. Recently, CMIP6 decadal data have been released that includes more frequent (yearly, instead of 5-yearly) hindcast start dates, larger ensembles of the hindcasts for individual start date, finer atmospheric resolution of the models, enhanced parameters of the cloud microphysical process, additional earth-system processes such as biogeochemical cycles and ice sheets. For the decadal prediction, CMIP5 used the mid-range future scenario (RCP4.5) based on 2100 radiative forcing whilst CMIP6 used socioeconomic pathways (SSPs), which is a more realistic scenario. This study recommends further research to compare the drift as well as models performances produced by the CMIP6 decadal precipitation with that of the CMIP5 model output at a catchment level. This would provide a more robust understanding of the characteristics of drifts and model performances for the practical uses of models' predicted precipitation data.
4. For future prediction, a few time series models such as Auto-Regressive Integrated Moving Average (ARIMA), Seasonal Auto-Regressive Integrated Moving Average (SARIMA), and Seasonal Auto-Regressive Integrated Moving Average with exogenous factors (SARIMAX) can be considered using both the GCMs derived precipitation and the corresponding observed data together.
5. For future prediction through BiLSTM, this study used only the Box-cox power transformation and 'tanh' activation function in all hidden layers of a BiLSTM model. A different form of architecture such as a higher number of hidden layers and different activation functions can be used for further investigation. In addition, different data processing such as 'Normalization', 'StandardScaler', and 'MinMaxScaler' transformation can also be investigated in further studies.