

**REPUBLIC OF TÜRKİYE
ISTANBUL GELİSİM UNIVERSITY
INSTITUTE OF GRADUATE STUDIES**

Department of Civil Engineering

**EVALUATION OF BIAS CORRECTION AND
DOWNSCALING METHODS FOR THE OUTPUTS OF
GENERAL CIRCULATION MODELS**

Master Thesis

Nouamane MTARFI

Supervisor

Assist. Prof. Dr. Mustafa NURİ

Istanbul – 2026

THESIS INTRODUCTION FORM

Name and Surname : Nouamane MTARFI

Language of the Thesis : English

Name of the Thesis : Evaluation of Bias Correction and Downscaling Methods
for the Outputs of General Circulation Models

Institute : Istanbul Gelisim University Institute of Graduate Studies

Department : Civil Engineering

Thesis Type : Master

Date of the Thesis : 26.06.2026

Page Number : 92

Thesis Supervisors : Assist. Prof. Dr. Mustafa NURİ

Index Terms : General Circulation Models, CMIP6, Downscaling, Bias-
Correction, ANN, Linear scaling, Quantile Mapping

Turkish Anstract : Küresel sirkülasyon modellerinin çıktılarındaki
belirsizliğin düşürülmesi için farklı yöntemler
uygulanmıştır. Klasik yöntemlerin yanı sıra yapay-zeka
tabanlı yöntemler de uygulanmıştır. Günlük sıcaklık ve
yağış değerleri gözlenmiş verilerle mukayese edilmiştir.

Distribution List : 1. To the Institute of Graduate Studies of Istanbul
Gelişim University
2. To the National Thesis Center of YÖK (Higher
Education Council)

Signature

Nouamane Mtarfi

**REPUBLIC OF TÜRKİYE
ISTANBUL GELİSİM UNIVERSITY
INSTITUTE OF GRADUATE STUDIES**

Department of Civil Engineering

**EVALUATION OF BIAS CORRECTION AND
DOWNSCALING METHODS FOR THE OUTPUTS OF
GENERAL CIRCULATION MODELS**

Master Thesis

Nouamane MTARFI

Supervisor

Assist. Prof. Dr. Mustafa NURİ

Istanbul – 2026

DECLARATION

I hereby declare that in the preparation of this thesis, scientific ethical rules have been followed, the works of other persons have been referenced in accordance with the scientific norms if used, there is no falsification in the used data, any part of the thesis has not been submitted to this university or any other university as another thesis.

Nouamane MTARFI

26/06/2026

TO ISTANBUL GELISIM UNIVERSITY
THE DIRECTORATE OF INSTITUTE OF GRADUATE STUDIES

The thesis study of Nouamane MTARFI titled as Evaluation of Bias Correction and Downscaling Methods for the Outputs of General Circulation Models has been accepted as MASTER THESIS in the department of Civil Engineering by our jury.

Signature

Director

Assist. Prof. Dr. Sead Ahmed SALIH

Signature

Member

Assist. Prof. Dr. Mustafa NURİ

(Supervisor)

Signature

Member

Assist. Prof. Dr. Mesut BARIŞ

APPROVAL

I approve that the signatures above signatures belong to the aforementioned faculty members.

... / ... / 2026

Signature

Prof. Dr. İzzet GÜMÜŞ

Director of the Institute

SUMMARY

Simulations conducted with General Circulation Models (GCMs) under the Coupled Model Intercomparison Project Phase 6 (CMIP6) provide the primary input for climate change impact studies. However, due to their coarse spatial resolution and systematic biases, they cannot be directly used in assessments of regional climate change impacts and are frequently applied after bias correction. While the number of studies on bias correction methods is growing, there is still need to have a comprehensive and comparative study for various bias correction methods in daily and monthly scale.

In this study, we compared the performance of linear scaling, quantile mapping, and artificial neural network (ANN)-based bias correction methods on daily temperature and precipitation outputs from six CMIP6 models for the Florya station of Istanbul, Türkiye, using the European Centre for Medium-Range Weather Forecasts Reanalysis Version 5 (ERA5) dataset during the period 1985-2014. A wet-day quantile mapping strategy was also employed for precipitation. Corrected outputs were compared with NASA's NEX-GDDP-CMIP6 products, and future climate projections were analyzed under the SSP2-4.5 and SSP5-8.5 scenarios for the 2015-2100 period. The 4-point interpolation by using Inversed Distance Weighting (IDW) was applied to obtain the climatic variables time series for the station from gridded data. The comparison based on the statistical (RMSE, correlation etc) and graphical illustrations (Taylor diagram, Q-Q plot etc).

Overall, the best results for temperature correction were obtained with ANN, while the most realistic precipitation distribution was obtained with wet-day quantile mapping. None of the methods notably improved the daily precipitation correlation, highlighting a key limitation of point-scale precipitation correction. Locally calibrated corrections performed similarly to or better than the globally applied NASA benchmark, and future projections indicated continued warming and drying throughout the 21st century, particularly under the high-emission SSP5-8.5 scenario.

The findings of this study contribute to the understanding of bias correction method performance and provide useful information for future climate impact studies in the Florya station and the broader Istanbul region.

Keywords: General Circulation Models, CMIP6, Downscaling, Bias-Correction, ANN, Linear scaling, Quantile Mapping

ÖZET

CMIP6 (Coupled Model Intercomparison Project Phase 6) kapsamındaki küresel sirkülasyon modelleri (GCM'ler), iklim değişikliği tesirlerinin araştırılmasına temel kaynağını oluşturmaktadır. Ancak, düşük mekânsal çözünürlükleri ve sistematik yanlışlıkları nedeniyle bölgesel iklim çalışmaları için doğrudan kullanılamamakta ve genellikle düzeltme yöntemleri uygulanmaktadır. Düzeltme yöntemlerine yönelik çalışmalar artmasına rağmen, günlük ve aylık ölçekte çeşitli yanlışlık düzeltme yöntemleri için kapsamlı ve karşılaştırmalı bir çalışmaya hâlâ ihtiyaç duyulmaktadır.

Bu çalışmada, İstanbul'un Florya ilçesi için altı CMIP6 modelinden elde edilen günlük sıcaklık ve yağış verilerine uygulanan doğrusal ölçekleme (Linear Scaling), kantil eşleme (Quantile Mapping) ve yapay sinir ağı (ANN) tabanlı yanlışlık düzeltme yöntemlerinin performansları karşılaştırılmıştır. Analizlerde 1985-2014 dönemi için ERA5 yeniden analiz verisi referans olarak kullanılmış, yağış verileri için ayrıca yağışlı gün kantil eşleme yöntemi uygulanmıştır. Düzeltilmiş çıktılar NASA'nın NEX-GDDP-CMIP6 veri setleri ile karşılaştırılmış ve 2015-2100 dönemi için SSP2-4.5 ile SSP5-8.5 senaryoları altında gelecek iklim projeksiyonları değerlendirilmiştir. İstasyon için ızgara (grid) verilerinden iklimsel değişkenlerin zaman serilerini elde etmek amacıyla, Ters Mesafe Ağırlıklandırması (IDW) kullanılarak 4 noktalı enterpolasyon uygulanmıştır. Karşılaştırma, istatistiksel (RMSE, korelasyon vb.) ve grafiksel gösterimlere (Taylor diyagramı, Q-Q grafiği vb.) dayandırılmıştır.

Sonuçlar, sıcaklık düzeltmesinde en başarılı yöntemin ANN olduğunu, yağış dağılımının temsiline ise yağışlı gün kantil eşleme yönteminin en iyi performansı gösterdiğini ortaya koymuştur. Yöntemlerin hiçbirisi günlük yağış korelasyonunu önemli ölçüde iyileştirememiştir; bu durum noktasal ölçekte yağış düzeltmesinin temel sınırlamalarından biridir. Yerel olarak kalibre edilen yöntemler, NASA referans ürünleriyle benzer veya daha iyi sonuçlar vermiştir. Ayrıca projeksiyonlar, özellikle SSP5-8.5 senaryosu altında, 21. yüzyıl boyunca sıcaklık artışı ve kuraklaşma eğiliminin devam edeceğini göstermiştir.

Bu alıřma, Akdeniz kıyı iklim kořullarında yanlılık düzeltme yöntemlerinin performansının anlaşılmasına katkı sağlamakta ve Florya ile İstanbul bölgesindeki gelecekteki iklim etki alıřmalarına temel oluşturmaktadır.

Anahtar kelimeler: Küresel sirkülasyon modelleri, CMIP6, Ölçek küçültme, düzeltme, yapay sinir ağıları, lineer ölçeklendirme, Quantile Mapping

TABLE OF CONTENTS

SUMMARY	i
ÖZET.....	iii
TABLE OF CONTENTS.....	v
ABBREVIATIONS	viii
LIST OF TABLES	ix
LIST OF FIGURES	xi
PREFACE	12
INTRODUCTION.....	13

CHAPTER ONE THESIS STRUCTURE

1.1. Problem Statement.....	15
1.2. Research Objectives.....	16
1.3. Scope and Limitations	17
1.4. Organization of the Thesis.....	18

CHAPTER TWO LITERATURE REVIEW

2.1. Previous Studies on Climate Model Evaluation and Downscaling	19
2.2. Previous Studies on Bias Correction Methods	20
2.2.1. Studies on Linear Scaling	21
2.2.2. Studies on Quantile Mapping	21
2.2.3. Previous Studies on Artificial Neural Networks (ANNs)	22
2.3. Previous Studies on ERA5 and Observational Reference Datasets.....	24
2.4. Previous Studies on Regional Climate and Istanbul.....	24
2.5. Research Gap	25

CHAPTER THREE STUDY AREA AND DATA

3.1. Study Area	27
3.2. Data Sources	28
3.2.1. ERA5 Reanalysis Data	28
3.2.2. CMIP6 Climate Model Data.....	28
3.2.3. NASA NEX-GDDP Data	29
3.3. Data Preprocessing	30
3.3.1. Spatial Interpolation and Grid Cell Extraction	30
3.3.2. Calendar Harmonization.....	30
3.3.3. Unit Conversion and Quality Control.....	31

CHAPTER FOUR

METHODOLOGY

4.1. Overview of Methodology.....	32
4.2. Evaluation Metrics.....	33
4.2.1. Error Metrics.....	33
4.2.2. Additional Evaluation Metrics	34
4.2.3. Training and Testing Framework	34
4.3. Linear Scaling Bias Correction.....	35
4.4. Quantile Mapping Bias Correction.....	35
4.4.1. Temperature Quantile Mapping.....	36
4.4.2. Precipitation Quantile Mapping and Wet-day Correction.....	36
4.5. Artificial Neural Network Bias Correction.....	37
4.6. NASA NEX-GDDP-CMIP6 Comparison	37
4.7. Monthly Climate Analysis.....	38
4.8. Future Climate Projections	38

CHAPTER FIVE

RESULTS AND DISCUSSIONS

5.1. Evaluation of Raw CMIP6 Models.....	40
5.1.1. Temperature.....	40
5.1.2. Precipitation.....	42
5.2. Linear Scaling Results	44
5.2.1. Temperature.....	44
5.2.2. Precipitation.....	45
5.3. Quantile Mapping Results	47
5.3.1. Temperature Results	47
5.3.2. Precipitation.....	51
5.4. Artificial Neural Network Results.....	54
5.4.1. Temperature.....	54
5.4.2. Precipitation.....	58
5.5. Comparison of Bias-Correction Methods.....	61
5.5.1. Temperature.....	61
5.5.2. Precipitation.....	63
5.6. Comparison with NASA NEX-GDDP Data.....	66
5.6.1. Temperature.....	66
5.6.2. Precipitation.....	68
5.7. Monthly Climate Analysis.....	70
5.8. Future Projections.....	73
5.8.1. Temperature Projections.....	73
5.8.2. Precipitation Projections.....	75

CONCLUSION AND RECOMMENDATIONS	77
REFERENCES.....	80
RESUME.....	86

ABBREVIATIONS

ANN	:	Artificial Neural Network
BCSD	:	Bias Correction and Spatial Disaggregation
C3S	:	Copernicus Climate Change Service
CDF	:	Cumulative Distribution Function
CMIP6	:	Coupled Model Intercomparison Project Phase 6
CORDEX	:	Coordinated Regional Climate Downscaling Experiment
ECMWF	:	European Centre for Medium-Range Weather Forecasts
ERA5	:	Fifth Generation ECMWF Atmospheric Reanalysis
GCM	:	General Circulation Model
IDW	:	Inverse distance weighting
IPCC	:	Intergovernmental Panel on Climate Change
LS	:	Linear Scaling
MSE	:	Mean Squared Error
NASA	:	National Aeronautics and Space Administration
NCCS	:	NASA Center for Climate Simulation
NCSS	:	NetCDF Subset Service
NEX-GDDP	:	NASA Earth Exchange Global Daily Downscaled Projections
P95	:	95th Percentile
P99	:	99th Percentile
PDF	:	Probability Density Function
QM	:	Quantile Mapping
QQ	:	Quantile-Quantile
RCM	:	Regional Climate Model
RCP	:	Representative Concentration Pathway
RMSE	:	Root Mean Square Error
SSP	:	Shared Socioeconomic Pathway
STD	:	Standard Deviation
UHI	:	Urban Heat Island
Wet-QM	:	Wet-Day Quantile Mapping

LIST OF TABLES

Table 1: Summary of datasets used in this study.....	29
Table 2: Summary of the methodological framework applied in this study.....	33
Table 3: Baseline Temperature Performance Metrics of CMIP6 Models (1985–2014).	40
Table 4: Baseline Precipitation Performance Metrics of CMIP6 Models (1985–2014).	42
Table 5: Raw and bias-corrected temperature performance metrics under linear scaling (2006–2014).....	44
Table 6: Raw and bias-corrected precipitation performance metrics under linear scaling (2006–2014).....	46
Table 7: Raw and bias-corrected temperature performance metrics under quantile mapping (2006–2014).....	48
Table 8: Standard deviation and 5th percentile of raw and quantile mapping corrected temperature.	48
Table 9: 50th and 95th percentile of raw and quantile mapping corrected temperature.	49
Table 10: Raw, QM, and wet-day QM precipitation performance metrics for all six CMIP6 models (2006–2014).	51
Table 11: Standard deviation of raw, QM, and wet-day QM corrected precipitation compared to ERA5 (2006–2014).....	52
Table 12: 50th and 95th percentile of raw, QM, and wet-day QM corrected precipitation compared to ERA5 (2006–2014).	52
Table 13: ANN bias correction performance metrics for daily temperature across six CMIP6 models (2006–2014).	54
Table 14: ANN temperature extremes (cold and hot RMSE) across six CMIP6 models (2006–2014).....	55
Table 15: ANN temperature extremes (cold and hot Bias) across six CMIP6 models (2006–2014).....	55
Table 16: ANN bias correction performance metrics for daily precipitation across six CMIP6 models (2006–2014).	58
Table 17: ANN bias correction performance for precipitation extremes across six CMIP6 models (2006–2014).	59
Table 18: RMSE comparison of raw and bias-corrected daily temperature across all methods (2006–2014).....	61
Table 19: Bias comparison of raw and bias-corrected daily temperature across all methods (2006–2014).....	61
Table 20: Correlation comparison of raw and bias-corrected daily temperature across all methods (2006–2014).	61
Table 21: RMSE comparison of raw and bias-corrected daily precipitation across all methods (2006–2014).....	63
Table 22: Bias comparison of raw and bias-corrected daily precipitation across all methods (2006–2014).....	64

Table 23: Correlation comparison of raw and bias-corrected daily precipitation across all methods (2006–2014).	64
Table 24: NASA NEX-GDDP temperature performance metrics compared to ERA5 (2006–2014).	66
Table 25: NASA NEX-GDDP extreme temperature metrics compared to ERA5 (2006–2014).	67
Table 26: NASA NEX-GDDP precipitation performance metrics compared to ERA5 (2006–2014).	68
Table 27: NASA NEX-GDDP extreme precipitation metrics compared to ERA5 (2006–2014).	68
Table 28: Monthly temperature RMSE for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).	70
Table 29: Monthly temperature bias for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).	70
Table 30: Monthly temperature correlation coefficient for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).	70
Table 31: Monthly total precipitation RMSE for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).	71
Table 32: Monthly mean precipitation bias for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).	71
Table 33: Monthly precipitation correlation coefficient for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).	71
Table 34: Ensemble mean and extreme temperature projections across three future periods under SSP2-4.5 and SSP5-8.5 (2021–2100).	73
Table 35: Ensemble mean and extreme precipitation projections across three future periods under SSP2-4.5 and SSP5-8.5 (2021–2100)	75

LIST OF FIGURES

Figure 1: Location of the Florya study area, Istanbul, Turkey (40.975°N, 28.787°E). Basemap: OpenStreetMap contributors.....	27
Figure 2: Taylor diagram for the baseline temperature evaluation of six CMIP6 models against ERA5 observations (1985–2014).....	41
Figure 3: Taylor diagram for the baseline precipitation evaluation of six CMIP6 models against ERA5 observations (1985–2014).....	43
Figure 4: Mean Temperature Bias Before and After Linear Scaling Correction for ACCESS- CM2 (2006–2014).....	45
Figure 5: Daily precipitation for ACCESS-CM2 before and after linear scaling compared to ERA5 observations during a high-rainfall period (December 2009 – March 2010).	47
Figure 6: Quantile-quantile plot for ACCESS-CM2 daily temperature before and after quantile mapping correction against ERA5 observations (2006–2014).....	50
Figure 7: Quantile-quantile plot of ACCESS-CM2 daily precipitation before and after QM and wet-day QM correction against ERA5 observations (2006–2014).	53
Figure 8: Scatter plot of ANN-corrected versus ERA5 observed daily temperature for ACCESS-CM2 (2006–2014).	56
Figure 9: Taylor diagram comparing raw and ANN-corrected temperature for ACCESS- CM2 against ERA5 observations (2006–2014).	57
Figure 10: Representative time series of raw and ANN-corrected daily precipitation for ACCESS-CM2 against ERA5 observations during a high rainfall period (December 2009 – March 2010).	60
Figure 11: Taylor diagram comparing bias correction methods for ACCESS-CM2 daily temperature (2006–2014).....	62
Figure 12: Quantile-quantile plot comparing all bias correction methods for ACCESS-CM2 daily precipitation against ERA5 observations (2006–2014).....	65
Figure 13: Monthly temperature climatology benchmark comparing ANN-corrected ACCESS-CM2 and NASA NEX-GDDP against ERA5 observations (2006–2014).	67
Figure 14: Wet-day precipitation distribution benchmark comparing wet-QM corrected ACCESS-CM2 and NASA against ERA5 observations (2006–2014).....	69
Figure 15: Monthly ACCESS-CM2 temperature and precipitation before and after ANN correction compared to ERA5 observations (2006–2014).	72
Figure 16: Ensemble annual mean temperature projections under SSP2-4.5 and SSP5-8.5 (2015–2100).....	74
Figure 17: Projected changes in ensemble extreme temperature (P95) based on 10-year moving average under SSP2-4.5 and SSP5-8.5 (2015–2100).	74
Figure 18: Projected changes in ensemble mean daily precipitation (10-year moving average) under SSP2-4.5 and SSP5-8.5 (2015–2100).	76

PREFACE

This thesis was prepared as part of the Master of Science degree program in Civil Engineering at Istanbul Gelisim University and focuses on the evaluation and comparison of bias correction methods applied to outputs from CMIP6 climate models for the Florya district of Istanbul using ERA5 reanalysis observations.

Some difficulties arose in retrieving and processing the climate datasets used in this study, specifically in extracting location-specific information from large global archives.

I would like to thank my advisor, Assist. Prof. Dr. Mustafa Nuri, for his guidance and support throughout the preparation of this thesis.

Nouamane Mtarfi

Istanbul, June 2026

INTRODUCTION

Global climate change has wide ranging impacts on water resources, agriculture, ecosystems, infrastructure, and extreme weather events. Since global averages are too broad to be useful in practice, regional and local climate information is necessary for decision making and adaptation planning.

Climate and earth system models are the main tools to project future climate because they represent the complex biogeochemical processes underlying the climate system (IPCC, 2023). General Circulation Models (GCMs) simulate the past, present and future of the climate system (IPCC, 2023). The latest generation of climate models, including the Coupled Model Intercomparison Project Phase 6 (CMIP6), are GCMs and the main source of future climate projections under different levels of anthropogenic greenhouse gas emissions in a representative concentration pathway (Eyring et al., 2016). However, GCMs have a spatial resolution of 100km to 300km, and may fail to represent important local features relevant to climate, such as complex topography, land-sea contrast (coastal areas), or the surface (Teichmann et al., 2013). Thus, without post-processing, they are unsuitable for local climate impact analysis.

Statistical or dynamical downscaling methods based on GCM or reanalysis data can be used to provide regional or local scale climate projections (Maraun & Widmann, 2018). Climate downscaling methods bridge the scale between global climate system models and local climate variables that are required for impact analysis of regional climate change. In dynamical downscaling, regional climate models (RCMs) are used to generate high resolution climate fields (Giorgi and Mearns, 1991). Statistical downscaling methods use statistical relationships between large scale model output and station data, and are computationally less expensive (Wilby & Wigley, 1997).

Climate models are also subject to systematic errors, such as biases in the modeled temperature or precipitation (Maraun & Widmann, 2018). These biases bring into question the reliability of the resultant climate projections, and can obscure the understanding of climate impacts (Christensen et al., 2008). Accordingly, bias-correction methods are widely applied in climate downscaling studies (Piani et al., 2010).

Other methods differ by correcting either less or more than the mean. Methods that only correct the mean show low complexity, such as linear scaling (Teutschbein & Seibert, 2012), while other methods that correct the entire statistical distribution of modeled climate variables are more complex, such as quantile mapping (Piani et al., 2010; Gudmundsson et al., 2012). Recently, machine learning methods, such as artificial neural networks (ANNs), have been used for bias correction and statistical downscaling. These methods are more flexible in capturing nonlinear relationships between model outputs and observations. However, these models can be highly variable in their performance (Baño-Medina et al., 2020). A correction method that works well for temperature might do little for precipitation, and it can vary widely by region.

The application of bias correction methods relies on the existence of a high quality observational data set, which can be used to perform the bias correction, against well known atmospheric conditions. The ERA5 reanalysis, which consists of a combination of observations and numerical weather prediction, has been extensively used as an observational benchmark for climate model evaluation and bias correction (Hersbach et al., 2020). Although many studies discuss bias correction methods, few have focused on the Istanbul region in particular, indicating the need for region-specific analysis and the use of high-quality reference data to assess the performance.

CHAPTER ONE

THESIS STRUCTURE

1.1. Problem Statement

Compared to observations and reanalysis, the CMIP6 models still show biases in temperature and precipitation, although they provide useful indications of future climate. These biases exist for both mean and extreme values, and CMIP6 models cannot yet be used directly to inform regional climate assessments (Teichmann et al., 2013). Hydrological modeling, water resource management, land use change studies and climate impact studies require highly reliable and accurate local climate data (Teutschbein & Seibert, 2012). If the bias-corrected climate data are not used, bias is still present in the models used for climate impact studies, leading to incorrect and misleading results (Maraun & Widmann, 2018). A great many methods have been proposed for bias correction, which range from simple mean correction and variance correction to quantile mapping methods (Piani et al., 2010). The performance of bias corrections methods varies with the climate variable being corrected, the spatial domain and the skill metric used to evaluate the correction (Maraun & Widmann, 2018). Each of these approaches has strengths and weaknesses, and no one approach has been shown to be superior in all circumstances (White & Toumi, 2013).

Machine learning and artificial intelligence methods have received meaningful interest recently in the context of climate bias correction given their potential to represent complex nonlinear relationships between model output and observations (Baño-Medina et al., 2020). These methods are, however, still less consistent in their performance than standard statistical methods when it comes to spatial, variable and metric consistency, including for assessments of extreme and mean climate variables (Baño-Medina et al., 2020).

Therefore, raw CMIP6 output records should be assessed, and bias correction methods compared using a common observational reference data set where such studies are lacking. The coastal metropolitan district of Istanbul is one such region that is affected by the Marmara Sea, complex topography and urban heat island (UHI) effects. However, comparatively few studies have evaluated and compared these debiasing methods in terms of their effectiveness at correcting the biases of CMIP6

models. Here we evaluate and compare the effectiveness of linear scaling, quantile mapping and ANN-based methods, using ERA5 reanalysis products as observations, for correcting CMIP6 daily temperature and precipitation bias for local, regional and continental analyzes and future projection.

1.2. Research Objectives

Using ERA5 observations as a reference dataset, the primary goal of this study is to assess the performance of specific CMIP6 climate models for the Florya station in Istanbul and analyze the efficacy of various bias correction techniques in enhancing local temperature and precipitation simulations.

To achieve this main goal, the study pursues the following specific aims:

- Extract and preprocess daily temperature and precipitation data from selected CMIP6 climate models for the historical period of 1985-2014, and interpolate them over the Florya station by using IDW method.
- Compare the raw outputs of the CMIP6 models to ERA5 observations and evaluate their performance using statistical indicators such as the root mean square error (RMSE), bias, and correlation coefficient.
- Apply traditional bias correction methods, which include linear scaling and quantile mapping, and then evaluate their ability to reduce systematic errors in the models.
- Develop an artificial neural network (ANN) model for bias correction and compare its performance to that of traditional statistical methods.
- Evaluate the ability of different correction methods to reproduce climate variability, statistical distributions, and extreme temperature and precipitation events.
- Compare locally corrected CMIP6 data to NASA'S NEX-GDDP-CMIP6 climate products and assess their relative performance.
- Apply the selected correction framework to future climate projections and further analyze the potential changes in temperature and precipitation under different climate change scenarios

1.3. Scope and Limitations

This study focuses on the evaluation of correcting daily temperature and daily precipitation from six selected CMIP6 climate models for the Florya station, Istanbul. It covers the historical period of 1985-2014 in which the data is processed and analyzed, and then further compared with ERA5 which serves as an observational reference.

The six selected CMIP6 models were analyzed using several statistical indicators, such as the root mean square error (RMSE), the mean bias, and the correlation coefficient. To Improve the agreement between simulated and observed climate conditions, three main bias correction approaches were used: linear scaling, quantile mapping, and artificial neural networks (ANNs). Additionally, wet-day quantile mapping was also implemented for precipitation to improve the representation of dry and wet days.

The models' performance is mainly assessed on daily time periods in the study, with further analysis carried out at the monthly scale. In order to assess how well the various correction techniques replicate both average climatic conditions and extreme temperature and precipitation events, the analysis includes evaluations based on statistical distributions, percentile comparisons, and analyses of extreme events in addition to standard statistical indicators.

Future climate projections were also examined using a subset of climate models based on the SSP2-4.5 and SSP5-8.5 scenarios. The correction relationships established for the historical period were applied to future simulations to produce locally adjusted climate projections for the Florya station.

The study comes with several limitations that must be mentioned. First, the analysis is limited to one geographic area, thus, the results cannot be generalized to other regions which could have different climate conditions. Second, only temperature and precipitation are processed, other variables such as humidity, wind speed, etc. are not included. Third, CMIP6 models have a coarse and large grid cells that can cover more than 100km, whereas Florya represents a specific point. Consequently, local factors and small-scale geographic features can influence observed conditions but are not fully revolved by the models. Fourth, the study focuses on a limited number of

climate models as well as correction techniques. Finally, the evaluation of the results relies on ERA5 reanalysis data, which despite its quality, remains a product of reanalysis rather than direct observations from weather stations.

1.4. Organization of the Thesis

This thesis is structured into four main chapters, in addition to the introduction and conclusion sections.

The introduction presents the general context of the study, defines the research problem, and outlines the objectives, as well as describing the scope and limitations of the work. Chapter 1 includes the structure of the thesis.

Chapter 2 is devoted to a review of the literature on climate change, General Circulation Models (GCMs), statistical downscaling techniques, bias correction methods, and the applications of machine learning in climate science. This chapter also highlights previous work and identifies the scientific gap that this thesis seeks to address.

Chapter 3 presents the study area, the datasets used, and the data preparation procedures. It describes the ERA5 reanalysis data and the selected CMIP6 climate models, along with the various preprocessing steps performed before the analyses.

Chapter 4 outlines the methodology adopted for this study. This chapter details the evaluation framework, the performance indicators used, and the bias correction and statistical downscaling methods applied, which are linear scaling, quantile mapping, wet-day quantile mapping, and artificial neural networks (ANNs).

Chapter 5 presents and discusses the results obtained. The performance of the raw climate models and the various correction methods is evaluated using distribution-based analyses, statistical indicators, and assessments of extreme events. This chapter also includes comparisons between the methods, as well as the best performing method with NASA NEX-GDDP climate products. Moreover, it includes an additional monthly evaluation of two models and an analysis of future climate projections.

Finally, the Conclusion and Recommendations section summarizes the main results of the study, highlights its contributions and limitations, and then offers recommendations and avenues for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1. Previous Studies on Climate Model Evaluation and Downscaling

Eyring et al. (2016) established how CMIP6 models have been designed to evaluate past, present and future climate change that arises from naturally occurring climate fluctuations and external forcing mechanisms, which made it increasingly important to evaluate not only the projected climate signals produced by these models, but also the biases that may affect their historical simulations and future projections.

While CMIP6 provides the model basis for climate projection studies, its output still faces scale-related limitations when applied regionally. Teichmann et al. (2013) investigated the link between global and regional climate simulations over many CORDEX (Coordinated Regional Climate Downscaling Experiment) domains using the MPI-ESM global model and the REMO regional climate model. They pointed out that although GCMs are frequently used to evaluate the current and future climate on a global scale, their direct usefulness for regional impact assessments is limited by their coarse resolution of roughly 100 to 300 km. Regional climate models, which employ large-scale GCM outputs as boundary forcing and dynamically refine them within a constrained domain, are frequently used to provide higher-resolution regional climate information in order to solve this (Teichmann et al., 2013).

Model performance also varies strongly by region. For example, Babaousmail et al. (2021) evaluated how well 15 CMIP6 models replicated rainfall variability over North Africa. They discovered that although the models accurately replicated mean annual climatology, some models exhibited systematic overestimation or underestimation between dry and wet months. According to their findings, regional impact studies can produce more dependable results by using a multi-model ensemble of the top-performing models.

Several studies have since examined the performance of downscaling techniques in addressing this limitation. Xu et al. (2019) examined several regional downscaling techniques, such as GCM bias correction, quantile-quantile correction, and bias correction with physical constraints. They found that, despite the various assumptions

and limitations associated with each technique, bias correction generally increased the accuracy of downscaled regional climate simulations.

The importance of bias correction was further underlined by Christensen et al. (2008), in which systematic biases in temperature and precipitation across an ensemble of thirteen regional climate models were evaluated. They noted that model biases could increase in the context of global warming. Their results demonstrated that while systematic precipitation biases may contribute to either an overestimation or an underestimation of anticipated precipitation changes depending on the model, warmer conditions typically result in an exacerbated warm bias, leading to an overestimation of projected warming. They concluded that bias correction must be applied before utilizing regional climate model outputs for impact assessments since doing so will probably lessen the dispersion among projected climate change signals.

Overall, these studies demonstrate that regional climate applications encounter two interrelated issues: the recurrent biases that persist in temperature and precipitation simulations and the scale mismatch between GCM outputs and local climate processes. Although prior research indicates that downscaled or regional outputs are still susceptible to the quality of the driving model data, downscaling can enhance the spatial relevance of climatic information. Consequently, subsequent research has increasingly concentrated on both producing regional climate forecasts and assessing and adjusting model bias prior to the use of those projections in impact assessments.

2.2. Previous Studies on Bias Correction Methods

Since climate models are typically biased and frequently run at resolutions lower than necessary for local impact assessments, bias correction techniques are widely used by users of climate model data (Maraun, 2016). These techniques range from simple mean-based adjustments to more complex distribution-based and machine learning approaches, each with distinct assumptions and limitations (Teutschbein & Seibert, 2012).

More recently, Dhawan et al. (2024) conducted a comprehensive comparison of multiple bias correction methods applied to ERA5-Land data across different temporal resolutions, underscoring that method performance depends strongly on the variable and timescale considered.

2.2.1. Studies on Linear Scaling

Linear scaling is generally considered a simple mean-bias correction method. Teutschbein and Seibert (2012) describe it as a procedure that adjusts climate model outputs so that the simulated monthly mean agrees with the corresponding observed monthly mean. For temperature, this is usually done through an additive correction, while for precipitation it is commonly applied through a multiplicative factor. The method therefore does not attempt to modify the full distribution of the variable but only corrects the average difference between simulated and observed values.

Due to its simplicity, linear scaling is useful as a baseline correction method. It is straightforward to implement, requires relatively limited data, and produces corrections that are easy to interpret. Zhang et al. (2024) also demonstrated that linear scaling can preserve the original climate change signal better than some distribution-based methods, as it does not strongly reshape the simulated distribution, making it particularly useful in studies where maintaining the broad model-projected trend is a priority.

In a similar study, Azman et al. (2022) used linear scaling to correct precipitation and temperature data after statistically downscaling them from the climate models. They found that this reduces the error in the rainfall data by at least 13%, but the correlation between the corrected rainfall data and observed rainfall was only moderate. When the seasons are wet, the correlation coefficients of the corrected rainfall data were rather poor. This indicates that linear scaling may not always be sufficient for reproducing the precipitation across seasons.

2.2.2. Studies on Quantile Mapping

Quantile mapping, which transforms the entire statistical distribution of the climate parameter and not just the mean, is probably the most common and advanced method. Besides quantile mapping, Piani et al. (2010) evaluated statistically based bias correction methods using a fitted histogram equalization function of daily precipitation and temperature. The authors found that quantile mapping can be used for downscaling any variable and is suitable for hydrological applications. Gudmundsson et al. (2012) followed up on this by comparing several quantile mapping methods applied to downscaling RCM precipitation for three different types

of distribution derived, parametric and nonparametric transforms. They found that, with many methods available, nonparametric transformations had the highest skill in removing precipitation bias in the returned data.

Berg et al. (2012) applied a distribution-based bias correction to high-resolution regional climate model output, correcting the full distribution of temperature and precipitation rather than the mean alone.

Considerable research has been conducted to compare quantile mapping methods in certain regions. Enayati et al. (2020) compared many quantile mapping methods in Iran's Karkheh River basin, with high variance in performance depending on transformation function and topography; some methods performed worse than raw model data alone. In a different study by Ngai et al. (2017), quantile mapping was applied to daily precipitation and temperature fields across the Southeast Asia region. The authors found that the quantile mapping method considerably reduced biases for both fields, with more improvement in the temperature field. Irwandi et al. (2023) used the ERA5 reanalysis data with subsequent quantile mapping to describe climate in Indonesia at the regional level, finding a statistically meaningful increase in temperature and precipitation over forty years.

While quantile mapping is an improvement over linear scaling, it is highly sensitive to the nature of the underlying observations. Quantile mapping alters important aspects of the observations (e.g. spatial and temporal structure and variability), it overcorrects for drizzle, and it overestimates extremes (Maraun 2013). Cannon et al. (2015) showed that standard quantile mapping can create artifacts in future model projected trends such as those in precipitation extreme indices, while quantile delta mapping avoids this issue by preserving the relative change in quantiles. Overall, there appears to be an improvement in bias correction when using quantile mapping compared to mean based bias correction techniques, but careful selection and understanding of limitations is still needed.

2.2.3. Previous Studies on Artificial Neural Networks (ANNs)

The limitations of customary bias correction approaches have led to the use of machine learning methods, such as artificial neural networks, that allow for more flexibility in the representation of nonlinear relationships between model outputs and

observations. A recent review by Baño-Medina et al. (2020) on recent deep learning approaches for continental scale statistical downscaling across Europe found that, despite the limited added-value of CNNs for temperature extremes, they outperform customary methods for most criteria and are feasible for large-scale climate applications.

Since then, multiple neural network methods have been proposed for bias correction of individual climate variables. Dutta and Bhattacharjya (2022) proposed a GRNN for bias correction of near-surface temperature over India. The trained model corrected GCM biases, which were strongly correlated between corrected and observed temperature. Moghim and Bras (2017) applied a feedforward neural network to correct temperature-precipitation bias in GCMs over northern South America. Their findings revealed that ANN models outperformed linear regression while correcting for the variables' probabilistic structure, even if the quantile information was not specifically included in the models. Yang et al. (2022) used a bidirectional long short-term memory network to rectify satellite precipitation bias in Malaysia. They found important improvements in correlation coefficients and reductions in root mean square error relative to historic distribution-based bias correction methods.

Further work on deeper networks for simultaneous bias correction and downscaling shows that deep learning models reduce space, time and extreme biases relative to the quantile delta mapping (QDM) (Wang and Tian, 2022) in CMIP6 daily temperature simulations, while also better reproducing intervariable dependences that were not modeled under customary DC methods. Despite being a promising advance, ANN based methods have produced varied results, depending on domain, the variable of interest, and the performance measure utilized. Moreover, their application to climate change projections has raised concerns of extrapolation beyond conditions met during training (Baño-Medina et al, 2020), and no single correction method has been able to outperform others in all scenarios.

The distinction between linear and nonlinear approaches has long been recognized; Weichert and Bürger (1998) compared the two for statistical downscaling and reported that nonlinear techniques could offer modest improvements, particularly for non-normally distributed variables such as precipitation.

2.3. Previous Studies on ERA5 and Observational Reference Datasets

ERA5 is an ECMWF reanalysis conducted by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Union's Copernicus Climate Change Service (C3S). With a spatial resolution of 31 km and hourly output from 1950 onwards, it is the successor to the previous generation ECMWF reanalysis, ERA-Interim. Hersbach et al. (2020) gave an overview of the setup, evaluation and performance of ERA5. ERA5 gives a better fit to the observations according to various metrics of temperature, wind, humidity and precipitation. For instance, the correlation of global mean monthly precipitation improved from 67% to 77%, giving a good observation-based reference.

Multiple studies have evaluated the performance of ERA5 in representing precipitation. Lavers et al. (2022) compared ERA5 precipitation to the in-situ precipitation of over 5600 stations around the world and found that the least error was in the winter extratropics and the most in the tropics, concluding that ERA5 is a useful dataset for tracking extratropical precipitation. According to Jiang et al. (2021) when ERA5 precipitation over the Chinese Mainland was evaluated, the ERA5 dataset performed well spatially, but light precipitation was overestimated and heavy precipitation was underestimated. Performance varied in different climate zones.

In a related finding, ERA5 and ERA5-Land precipitation data for Spain over 70 years generally matched observations well, but to different extents when related to climatic region, precipitation intensity, and orography (Gomis-Cebolla et al., 2023). The Mediterranean coast was one of the most difficult regions in Spain to achieve a good model fit using ERA5. The research also identified a wet bias in coastal regions, which is also relevant for the coastal Mediterranean context of Istanbul. Despite these shortcomings, ERA5 is the most complete and widely used reanalysis product for climate model evaluation and bias correction (Hersbach et al., 2020).

2.4. Previous Studies on Regional Climate and Istanbul

The Mediterranean region has often been identified as among the most vulnerable regions to climate change, with many reports projecting warming and drier conditions. Seker and Gumus (2022) projected for the Mediterranean Sea precipitation decreases of 15-20% and temperature increases from 1.0-3.1°C, based on a multi-

model ensemble of 22 CMIP6 models. Projected precipitation decreases are larger along the coastline of the Mediterranean.

Studies exist on national scale in Türkiye concerning climate projection and bias-correction. Oruc (2022) concluded in a study comparing three bias correction methods applied to five CMIP6 models in Türkiye that the model performance was affected mainly by the region and reference dataset, rather than the bias correction method applied. Gumus et al. (2023) analyzed Türkiye's climate extremes for CMIP6 using ERA5-Land and reported that dryness and extreme temperatures were common, particularly in the Marmara region. Yilmaz et al. (2024) applied six CMIP6 models with ERA5 and an ANN-based ensemble over the Turkish dam basin, revealing a temperature increase of up to 5.64 degrees and a precipitation decrease of up to 43% in the SSP5-8.5 scenario.

At the regional scale, Kara et al. (2016) investigated extremes in precipitation over eastern Istanbul, Türkiye, using a multi-model ensemble of regional models and found increases in extreme precipitation in winter and spring over this region. This study is also the first attempt to bias correct climate data for Istanbul, as the previous study was before CMIP6 and was not systematic in doing so. The Florya station is located next to the Sea of Marmara, and there is an urban heat island in the region, but there are not many climate data from the region.

2.5. Research Gap

Despite the growing volume of literature on bias correction of climate models and regional climate modeling, the following gaps are addressed in this study. Although there are studies using linear scaling, quantile mapping, and artificial neural networks (ANNs) in different regions, the systematic comparison of bias correction methods for daily temperature and precipitation for the Istanbul region in CMIP6 models has not been addressed. Previous studies calculated bias correction factors for Türkiye (Oruc, 2022; Gumus et al. 2023), but their estimates are only applicable to the national level, and cannot be used at site scale. Kara et al. (2016) report estimates of the climate change impact on Istanbul, but their study predates CMIP6 and their bias correction model is different from the bias correction applied in this study.

While the ERA5 reanalysis product is frequently used as a reference dataset in climate research, its application for CMIP6 bias correction in the Istanbul region remains limited. This is particularly relevant for the Florya station, whose climate is influenced by the Marmara Sea, urbanization, and local geographic conditions.

This study seeks to fill this gap by comparing the performance of linear scaling, quantile mapping, and ANN-based bias correction methods for the downscaling of daily temperature and precipitation data in the Florya station in Istanbul using the ERA5 dataset as a reference. The performance of the bias correction methods was tested on six CMIP6 models between 1985 and 2014. This study seeks to determine the most appropriate method for the local climate and future climate assessments in Florya, which has received little attention.

CHAPTER THREE

STUDY AREA AND DATA

3.1. Study Area

In this thesis the daily precipitation and temperature data from Florya station located in Florya of the Istanbul Province of Türkiye on the European side of Istanbul at 40.975°N 28.787°E (Figure 1). Istanbul is one of the biggest cities in the world, located between Europe and Asia. To the north of the city is the Black Sea and to the south lies the Marmara Sea with the Bosphorus Strait between the two; Florya is a seaside neighborhood of the Sea of Marmara. Factors such as proximity to the sea, coastal wind patterns, urbanization, and the varying built environment have influenced the development of microclimates and the urban heat island effect. Local microclimates of Istanbul are influenced by a Mediterranean climate with hot dry summers and cool wet winters, which differ considerably from larger regional climate signals derived from global climate models. These complex local effects mean that Florya would make a particularly good place to evaluate and climate models although few studies have been performed on Florya's climate localized from global models.

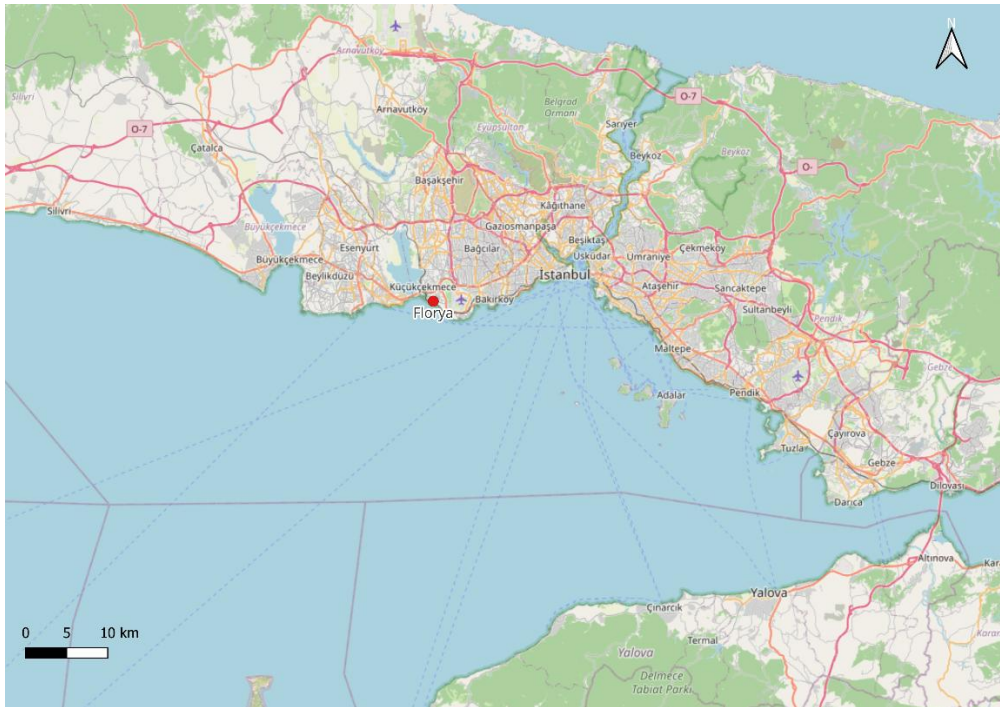


Figure 1: Location of the Florya study area, Istanbul, Türkiye (40.975°N, 28.787°E). Basemap: OpenStreetMap contributors.

3.2. Data Sources

This study uses three main datasets: ERA5 reanalysis data as the observational reference for bias correction, CMIP6 climate model outputs as the primary source of climate projections to be corrected, and NASA's NEX-GDDP-CMIP6 (Thrasher et al., 2022) data as an additional benchmark for evaluating the performance of the corrected model outputs. A summary of the key characteristics of all datasets used in this study is provided in Table 1.

3.2.1. ERA5 Reanalysis Data

ERA5 is the fifth generation global atmospheric reanalysis developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the Copernicus Climate Change Service (C3S) (Hersbach et al., 2020). Combining historical in situ and satellite observed data with numerical weather prediction models allows ERA5 to form a spatial and temporal record of atmospheric state. It provides a horizontally enhanced resolution of approximately 31km. Daily temperature and precipitation data were obtained from the ERA5 daily dataset at a spatial resolution of 24km, accessed via Climate Engine for the Florya location for the historical period of 1985-2014. ERA5 was selected as the observational reference dataset due to its high temporal resolution, global coverage, and previous demonstrated reliability for climate model evaluation and bias correction applications.

3.2.2. CMIP6 Climate Model Data

Climate model simulations used in this study were obtained from the Coupled Model Intercomparison Project Phase 6 (CMIP6), an international framework for climate and climate change projections (Eyring et al., 2016). It provides modeling outputs of several General Circulation Models (GCMs) produced by research institutes around the world. CMIP6 forms the scientific basis for the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).

The CMIP6 datasets used in this study were downloaded from Copernicus Climate Data store. Daily near-surface air temperature (tas) and precipitation (pr) data for the historical period of 1985-2014 were extracted. Six CMIP6 models were selected which produced a diverse range of climate responses: ACCESS-CM2, MPI-

ESM1-2-LR, NorESM2-MM, EC-Earth3-CC, HadGEM3-GC31-LL, and GFDL-ESM4. Additionally, projections of future climate using the Shared Socioeconomic Pathways (SSP) SSP2-4.5 and SSP5-8.5 were used for the temperature and precipitation projections for three selected models (ACCESS-CM2, GFDL-ESM4, NORESM2-MM). SSP2-4.5 is a Representative Concentration Pathway (RCP) for intermediate levels of global emissions, while SSP5-8.5 assumes a high level of greenhouse gas emissions, driven by continued development based on fossil fuels (O'Neill et al., 2016).

CMIP6 models have been used for analyzing present and future climate. However, due to their coarser spatial resolution, they are unsuitable to be used for local scale climates, and they also exhibit systematic biases compared with the measured temperature and precipitation. For this reason, CMIP6 simulations were compared to the ERA5 reanalysis data and bias corrected, before being used for further research.

Table 1: Summary of datasets used in this study.

Dataset	Source	Period	Variables	Grid Resolution
ERA5 Reanalysis	Climate Engine (ERA5, ECMWF/C3S)	1985–2014	Temperature, Precipitation	24 km
CMIP6 GCMs	Copernicus Climate Data Store	Historical (1985–2014), SSP2-4.5, SSP5-8.5	Temperature, Precipitation	100-200 km
NASA NEX-GDDP-CMIP6	NASA Center for Climate Simulation (NCCS)	1985–2014	Temperature, Precipitation	$0.25^{\circ} \times 0.25^{\circ}$

3.2.3. NASA NEX-GDDP Data

The NASA Earth Exchange Global Daily Downscaled Projections dataset (NEX-GDDP-CMIP6) was used as an external benchmark for evaluating the performance of the bias-corrected CMIP6 model outputs. These data are global daily downscaled climate projections with a spatial resolution of 0.25 degrees. The dataset is based on day-to-day simulations from CMIP6 Global Climate Model (GCM) projections downscaled using the daily variant of the Bias Correction and Spatial Disaggregation (BCSD) method (Thrasher et al., 2022). Daily temperature and precipitation records from the 1985-2014 historical period from six CMIP6 models were downloaded from the NASA Center for Climate Simulation (NCCS).

3.3. Data Preprocessing

To maintain consistency and compatibility, all datasets were preprocessed using Python before analysis.

3.3.1. Spatial Interpolation and Grid Cell Extraction

Climate models operate on coarse spatial grids, meaning there is no individual climate model for precise locations. Florya station is located at 40.9758°N, 28.7865°E. For example, CMIP6 models have a spatial resolution of about 100 to 200 km, and the nearest grid cell is biased with respect to the station location because the grid cell does not just cover the station location, but the climate of the center of the grid cell, surrounded by other grid cells. To overcome this problem, four grid cells surrounding the station were selected for each model. The values for the four surrounding grid cells were averaged for each day to form a single time series that represented the climate for Florya that day. This method uses surrounding grid cells to reduce the climate variability and is widely used in climate studies. This process was replicated for future projection data under SSP2-4.5 and SSP5-8.5 across the three selected models.

For the NASA NEX-GDDP-CMIP6 dataset, the data was directly requested from the NASA NetCDF Subset Service (NCSS) archive and stored only for the bounding box around Florya (40.8-41.2°N, 28.5-29.1°E) to avoid the need to download the entire global dataset. Daily data was concatenated to build continuous time series for each model and each year. Model naming conventions, ensemble identifiers, and file structures were standardized during data acquisition to bring the datasets into agreement.

The ERA5 data were extracted for Florya as a single point time series and therefore did not need to be spatially interpolated.

3.3.2. Calendar Harmonization

In the selected CMIP6 models, ACCESS-CM2, EC-Earth3-CC, and MPI-ESM1-2-LR use a proleptic Gregorian calendar, while GFDL-ESM4 and NorESM2-MM use a no-leap calendar with every year being 365 days long. HadGEM3-GC31-

LL uses a 360-day calendar where every year contains twelve months that are each 30 days long.

Such differences lead to temporal offsets between model output and observations from ERA5, which uses the Gregorian calendar system. To synchronize all model output with records from ERA5, the outputs were converted to a common datetime format, and records were aligned accordingly. This required special consideration for HadGEM3-GC31-LL, as its use of a 360-day calendar produced some dates that cannot correspond to dates in the Gregorian calendar. These dates were therefore identified and removed during its pre-processing. After the datasets were harmonized, the dates were adjusted so that the model evaluation and bias-correction algorithms would operate on the same days for each of the datasets.

3.3.3. Unit Conversion and Quality Control

After extraction, global climate model temperatures from the CMIP6 and NASA NEX-GDDP-CMIP6 datasets were converted from Kelvin to degrees Celsius, and precipitation values were converted from kilograms per square meter per second to millimeters per day to match the units used by the ERA5 reference dataset. After the unit conversion, each dataset was checked for missing values and any differences in the time dimension were corrected.

CHAPTER FOUR

METHODOLOGY

4.1. Overview of Methodology

A series of bias correction, data processing, and climate projection methods were applied to the outputs of the CMIP6 climate models for the Florya station in Istanbul. The daily outputs of the six different CMIP6 climate models were compared to the ERA5 reanalysis data for the historical period of 1985-2014. The raw model simulations have subsequently been evaluated against a range of statistical, distributional and graphical evaluation metrics.

Three commonly used methods of bias correction were then applied to the climate model output to reduce the systematic model errors: Linear Scaling (LS), Quantile Mapping (QM) and Artificial Neural Networks (ANN). The corrected data were then verified and compared against the model output to test the performance of each bias correction approach for temperature and precipitation.

In addition to the bias-correction experiments, the NASA NEX-GDDP-CMIP6 datasets were analyzed in an external comparison experiment to validate the models' performance. Monthly climate characteristics were also analyzed for two selected models in an experiment to determine the ability of the bias-correction method to reproduce seasonal characteristics in the climate data. Finally, projections during future climate scenarios (SSP2-4.5 and SSP5-8.5) were determined according to CMIP6 model selection to evaluate the impact of climate change on temperature and precipitation. An overview of the methodological steps, including the methods applied, variables considered, and reference periods used, is provided in Table 2.

Table 2: Summary of the methodological framework applied in this study.

Step	Method	Variable	Reference Period
Baseline Evaluation	RMSE, Bias, Correlation	Temperature, Precipitation	1985–2014
Bias Correction	Linear Scaling	Temperature, Precipitation	Train: 1985–2005, Test: 2006–2014
Bias Correction	Quantile Mapping	Temperature, Precipitation	Train: 1985–2005, Test: 2006–2014
Bias Correction	Wet-Day QM	Precipitation	Train: 1985–2005, Test: 2006–2014
Bias Correction	ANN	Temperature, Precipitation	Train: 1985–2005, Test: 2006–2014
Benchmark Comparison	NASA NEX-GDDP	Temperature, Precipitation	1985–2014
Future Projections	QM + Wet-QM	Temperature, Precipitation	2015–2100

4.2. Evaluation Metrics

4.2.1. Error Metrics

To assess the capability of the climate models and bias correction methods to produce credible outputs, three common statistical assessments were employed: the Root Mean Square Error (RMSE), Bias, and Pearson correlation coefficient (R). These were selected since they can provide complementary information about model errors, bias, and ability of the model to reproduce temporal variability.

Root Mean Square Error (RMSE) describes model error as the root mean squared average of the deviation between observed and model predictions. A lower value reflects better agreement between the model and observations, while a higher value indicates greater disparity.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - O_i)^2} \quad (1)$$

where M_i and O_i are the model-simulated and observed values at time i , respectively, and n is the total number of observations.

Bias is a measure of how much on average the model predictions are above or below the observed values. Positive bias indicates overestimation, while negative bias indicates underestimation.

$$Bias = \frac{1}{n} \sum_{i=1}^n (M_i - O_i) \quad (2)$$

The Pearson correlation coefficient (R) was calculated to quantify the degree to which the models and the observed time series varied together. R can take on values from -1 to 1. Values closer to 1 indicate a greater positive correlation and a better representation of temporal variability.

$$R = \frac{\sum_{i=1}^n (M_i - \bar{M}) (O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (3)$$

where M and O denote the mean simulated and observed values.

4.2.2. Additional Evaluation Metrics

Additional measures were used when appropriate to give a more focused evaluation of the model's performance. The standard deviation, 5th percentile (P5), P50, and 90th percentile (P90) were calculated to describe the accuracy with which the bias correction method was able to reproduce the statistical distribution of the climate variables. Extreme-specific metrics include hot and cold RMSE, and bias. These metrics were used to assess model performance at each temperature range and heavy precipitation event. Extreme-specific metrics selectively depending on the characteristics of each correction method, and were not necessarily applied to all methods.

Taylor diagrams were also used to visualize the correlation, standard deviation, and the RMSE value all at once (Taylor, 2001). Quantile-quantile plots and cumulative distribution functions (CDFs) were utilized to visually compare the distribution of the corrected data and observed data.

4.2.3. Training and Testing Framework

In order to evaluate the generalizability of the bias correction methods, the historical (1985-2014) period was split into a training period (1985-2005) and a testing period (2006-2014), which corresponds to approximately 70% and 30% of the data, respectively. Model parameters and correction factors were only fitted on the training period and applied on the testing data.

This is preferable, as it better reflects model performance than calibrating and testing using the same data, consistent with split-sample validation framed previously described by Klemeš (1986). This training-testing approach was used for all bias-correction techniques considered in this study.

4.3. Linear Scaling Bias Correction

The simplest bias correction method implemented was linear scaling, whereby the model output was scaled so that the mean of the simulations matches the observed mean over the training period. For temperature, the additive approach was preferred because model bias is, to a first order, a constant offset.

$$T_{corr} = T_{mod} + (\bar{T}_{obs} - \bar{T}_{mod}) \quad (4)$$

For precipitation, the multiplicative approach was preferred because bias is, to a first order, proportional to precipitation.

$$P_{corr} = P_{mod} \times \frac{\bar{P}_{obs}}{\bar{P}_{mod}} \quad (5)$$

where T_{corr} and P_{corr} are the corrected temperature and precipitation values, respectively; T_{mod} and P_{mod} are the original model-simulated values; and $\bar{T}_{obs}$, $\bar{P}_{obs}$, $\bar{T}_{mod}$, and $\bar{P}_{mod}$ represent the mean observed and modeled values calculated from the training period.

The correction factors computed for the training period were directly applied with no adjustment to the testing data period. Linear scaling is a simple operation that eliminates the systematic mean bias and preserves the temporal variability. However, it cannot compensate for the higher moments of the distribution (Lenderink et al., 2007; Teutschbein & Seibert, 2012).

4.4. Quantile Mapping Bias Correction

In quantile mapping, the full statistical distribution of the climate variable is bias corrected. Instead of only correcting the mean, quantile mapping takes the value of the modeled variable, looks up its position in the modeled CDF, and replaces it by the corresponding value in the observed CDF at the same quantile position. This allows the model to correct not only for mean, but also for the variance, the shape of the distribution and the extremes of the modeled data (Piani et al., 2010; Themeßl et al.,

2012). Thus, an empirical quantile mapping was chosen, where the correction was calculated directly from the sorted observed and modeled values without assuming any theoretical distribution. These correction functions were calculated over the whole training period and were then applied to the testing period.

4.4.1. Temperature Quantile Mapping

For temperature, empirical quantile mapping was used, where the CDFs of the modeled and observed daily temperature data were built from the training period, and quantile mapping was applied to the modeled temperature data. For each modeled temperature value, the quantile of that value was searched for in the modeled CDF, and the corrected temperature was the corresponding observed temperature.

$$T_c = F_o^{-1}(F_m(T_m)) \quad (6)$$

where T_c is the corrected temperature, T_m is the modeled temperature, F_m is the model CDF, and F_o^{-1} is the inverse observed CDF. This transformation ensures that the corrected temperature distribution matches the observed distribution over the training period.

4.4.2. Precipitation Quantile Mapping and Wet-day Correction

For precipitation, The same standard quantile mapping procedure was applied, although it presents additional challenges because standard climate models typically simulate drizzle days (days with little precipitation) that are not strictly dry days, but are also not wet days. Later analysis had to be carried out using wet-day quantile mapping with a 0.1mm per day threshold, below which days would be classed as dry. In order to make the model wet-day frequency distribution match that of the observed wet-day frequency distribution, the modeled wet days having the driest precipitation were recast as dry days, until the distribution became equal (Schmidli et al. 2006). QM was then applied to the remaining wet-day values.

$$P_c = F_{o,w}^{-1}(F_{m,w}(P_m)) \quad (7)$$

where P_c is the corrected precipitation, P_m is the modeled precipitation, $F_{m,w}$ is the wet-day model CDF, and $F_{o,w}^{-1}$ is the inverse wet-day observed CDF. This allows for correcting both the frequency and intensity of precipitation events, which is a well-

known limitation of standard quantile mapping applied to precipitation (Gudmundsson et al., 2012).

4.5. Artificial Neural Network Bias Correction

An artificial neural network (ANN) was also trained as a machine learning alternative to the statistical bias correction methods outlined above. The ANN was in the form of a feedforward multilayer perceptron with two hidden layers, consisting of 16 and 8 neurons, respectively, and one output neuron. The hidden layers used the Rectified Linear Unit (ReLU) activation function. The output layer used a linear activation function. For optimization, the Adam optimizer was used to minimize the mean squared error (MSE) loss function (Kingma & Ba, 2015). Deep learning methods have also been applied to climate downscaling and bias correction, as they are capable of capturing complex nonlinear relationships between model output and observations (Rampal et al., 2022; Baño-Medina et al., 2020).

A separate ANN was trained for each climate model and for each variable, so there are separate temperature and precipitation models for each of the six CMIP6 climate models. The input to the ANN is the raw daily climate model output for the variable which is to be mapped to the ERA5 observation. All input and target variables were standardized prior to training using the standard scaler fitted only on the training data to avoid data leakage. In addition, precipitation outputs from the ANN were clipped at zero to avoid physically unrealistic negative precipitation.

The maximum number of epochs was set at 100 with a batch size of 32. An early stopping method was used in which the validation loss that had not improved for a set number of epochs caused the training to stop, then the training of the model was rolled back to the weights of the epoch with the best loss. L1 and L2 regularization were tested but showed no improvement. All models were trained and tested on the same training and testing period (Section 3.2.3).

4.6. NASA NEX-GDDP-CMIP6 Comparison

To independently measure the performance of the bias correction methods developed in this study, we used the NASA NEX-GDDP-CMIP6 historical climate projections as a benchmark to evaluate the output of the method that generated the best

bias correction results over the 1985-2014 period in conjunction with ERA5 data as the reference value. For temperature, the ANN corrected output was used, while for precipitation, wet day quantile mapping was used. The performance of each output was evaluated by the standard error metrics and the evaluation metrics for the extreme events. This allowed for the NASA downscaled product to be compared with the local bias correction methods in terms of their mean climate and extreme events.

4.7. Monthly Climate Analysis

To provide understanding into the seasonality of the bias correction methods, the ACCESS-CM2 and GFDL-ESM4 models were evaluated on a monthly basis. Monthly mean temperature and monthly total precipitation were calculated for the original output and the output of the three bias correction methods and compared to the ERA5 reanalysis dataset by RMSE, bias, and correlation coefficient. However, this analysis was only used as a supplementary evaluation to determine if applied corrections were able to apply realistic seasonal cycles.

4.8. Future Climate Projections

Future climate projections were developed using three CMIP6 models (ACCESS-CM2, GFDL-ESM4 and NorESM2-MM) , and two emissions pathways (SSP2-4.5 and SSP5-8.5) , spanning the period 2015-2100. Notably, the same grid cell averaging procedure for temperature and precipitation data was used for the future scenarios as was used for the historical data.

The climate models were bias-corrected using the bias correction relationships established for the historic period using all the climate model data. For temperature, an extrapolating quantile mapping was used, to avoid incorrect clipping of future temperatures to the observed range. For precipitation, wet-day quantile mapping was applied with 0.1 mm/day threshold, so that dry days remained unchanged and only wet-day precipitation was corrected, balancing correction stability and avoiding unrealistic precipitation extrapolation behavior.

The future period was divided into three sub-periods, near future (2021-2040), mid-century (2041-2070), and late century (2071-2100), and summary statistics (mean temperature and precipitation, and indices of extremes) were computed for the near

future, mid-century, and late-century periods under each of the three climate change scenarios. The outputs of the three models were combined to create ensemble projections to provide a multi-model representation of projected future climate conditions for the Florya station.

All data used in the study are gridded data and we applied an Inversed Distance Weighting (IDW) interpolation method to obtain the time series for the Florya station (Panigrahi 2023).

CHAPTER FIVE

RESULTS AND DISCUSSIONS

5.1. Evaluation of Raw CMIP6 Models

RMSE, mean bias, and Pearson correlation coefficient were used to compare the raw outputs of the six CMIP6 models to ERA5 observations for the historical period 1985–2014 before bias correction.

5.1.1. Temperature

Table 3 summarizes the performance of the six CMIP6 models against ERA5 observations for daily temperature during the historical period (1985–2014).

Table 3: Baseline Temperature Performance Metrics of CMIP6 Models (1985–2014).

Rank	Model	RMSE (°C)	Bias (°C)	Correlation (r)
1	MPI-ESM1-2-LR	3.70	0.21	0.857
2	NorESM2-MM	3.81	-0.32	0.854
3	EC-Earth3-CC	3.99	0.75	0.858
4	HadGEM3-GC31-LL	3.88	-0.79	0.852
5	GFDL-ESM4	4.16	-0.83	0.850
6	ACCESS-CM2	6.42	-2.86	0.815

The baseline results found that the six CMIP6 models performed reasonably well in simulating the observed past daily temperature variability, as the correlation coefficients of the model runs and the observed time series varied between 0.815 and 0.858. The RMSE values of five of the six models were less than 4.2°C. There were differences in the models' bias values, with some models having a tendency to overestimate and other models tending to underestimate the temperature. The models were rated based on their overall performance based on the evaluation of RMSE, bias, and correlation. This revealed discernible variations in simulation accuracy within the selected CMIP6 ensemble.

Overall, MPI-ESM1 was the model with the lowest RMSE of 3.70°C, a high correlation coefficient of 0.857, and a negligible bias of 0.21°C. The next best performing model was NorESM2, with a similarly strong performance. At the other

end of the spectrum, ACCESS-CM2 is the poorest performing model, with the highest RMSE of 6.42°C , and the largest negative temperature bias of -2.86°C , indicating a systematic and relatively large underestimate of temperature. Even with the lowest performance, the correlation coefficients were above 0.81, indicating that the temporal structure of temperature variation was well captured over the study duration. Figure 2 presents the Taylor diagram for the baseline temperature evaluation.

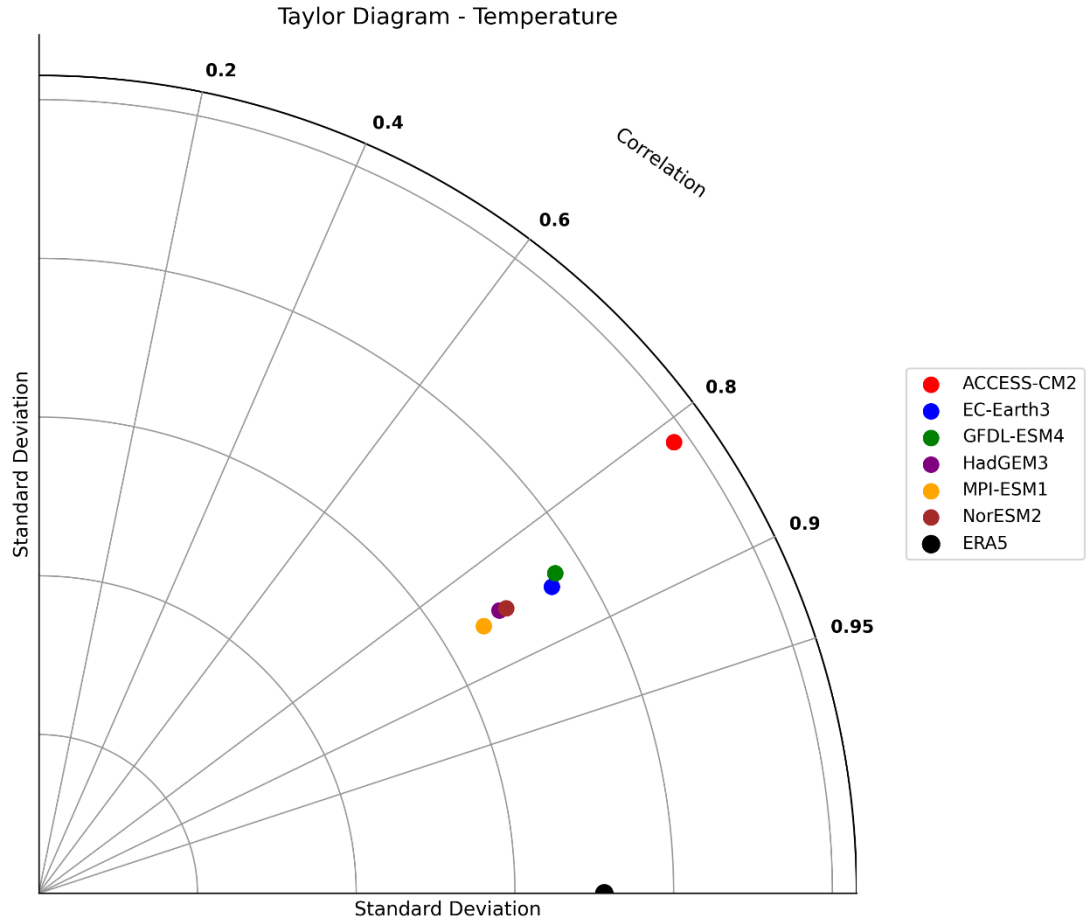


Figure 2: Taylor diagram for the baseline temperature evaluation of six CMIP6 models against ERA5 observations (1985–2014).

The Taylor diagram expands on the model evaluation above by providing a quantitative measure of correlation, standard deviation, and overall agreement with the ERA5 reference dataset. all six CMIP6 models are in a narrow region of the Taylor diagram, indicating a similar ability in reproducing temperature variability. The five models closest to the ERA5 reference are MPI, NorESM, EC-Earth, GFDL, and HadGEM, and are therefore characterized with the highest agreement to the

observations. ACCESS is the model furthest from the reference point as it has the highest RMSE value and is thus ranked the lowest, consistent with Table 3.

The reasonable temperature performance of the six CMIP6 models is consistent with other studies that have assessed climate models' performance. Babausmail et al. (2021) found CMIP6 models to adequately represent mean annual temperature climatology, though performance varied between models and domains. The relatively weaker performance of ACCESS-CM2 in particular is consistent with the conclusion that model skill is highly sensitive to model configuration as well as the climate of the region studied (Oruc, 2022) .

5.1.2. Precipitation

Table 4 summarizes the performance of the six CMIP6 models against ERA5 observations for daily precipitation during the historical period (1985–2014).

Table 4: Baseline Precipitation Performance Metrics of CMIP6 Models (1985–2014).

Rank	Model	RMSE (°C)	Bias (°C)	Correlation (r)
1	MPI-ESM1-2-LR	5.35	0.10	0.081
2	EC-Earth3-CC	5.39	-0.20	0.058
3	HadGEM3-GC31-LL	6.28	0.09	0.045
4	NorESM2-MM	5.30	-0.31	0.030
5	GFDL-ESM4	5.76	0.29	0.042
6	ACCESS-CM2	6.04	0.11	0.015

Unlike temperature, the six models' projected precipitation was poorly correlated, showing Pearson correlation coefficients of 0.015-0.081, and failed to resolve the timing of daily precipitation events at the grid cell scale. MPI-ESM1 ranked first with an RMSE of 5.35 mm/day and a bias of 0.1 °C among the six models, but the differences were small. The ability to resolve precipitation at daily temporal resolution is a general challenge for climate models rather than a characteristic of a particular model. Figure 3 presents the Taylor diagram for the baseline precipitation evaluation.

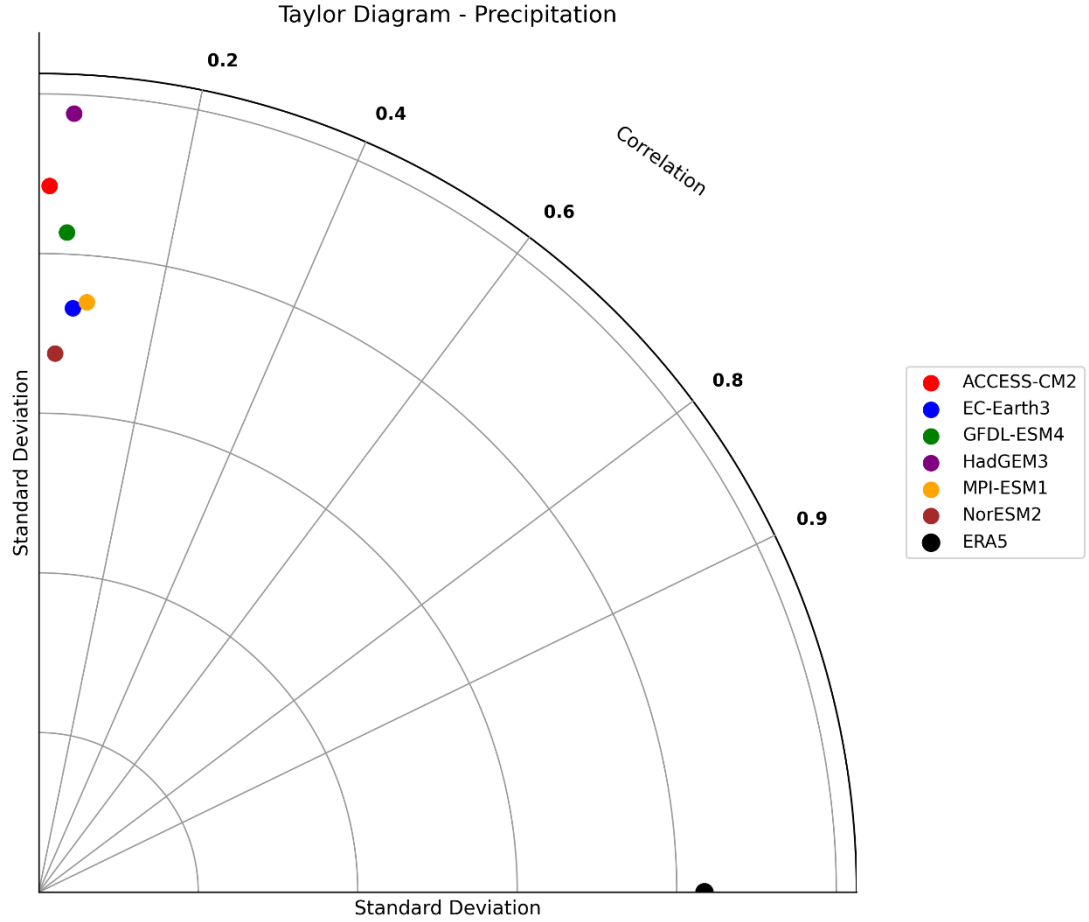


Figure 3: Taylor diagram for the baseline precipitation evaluation of six CMIP6 models against ERA5 observations (1985–2014).

The Taylor diagram confirms the conclusions of Table 4, showing that all six models are far away from the mean at the ERA5 point and have low correlation coefficients. The overall pattern suggests that precipitation was far more difficult to replicate than temperature across the CMIP6 ensemble, even though small variations in standard deviation are seen among the models.

The low correlation coefficients of precipitation for all six models are compatible with those found in previous studies for the CMIP6 models. For example, Babausmail et al. (2021) found low correlation coefficients of CMIP6 precipitation over North Africa and postulated that the CMIP6 models are unable to reproduce local precipitation variability. This behavior relates to the fine scale and local nature of precipitation processes, which occur at scales far smaller than the grid scale of general circulation models (Maraun & Widmann, 2018).

5.2. Linear Scaling Results

5.2.1. Temperature

Linear scaling was applied to the temperature outputs of all six CMIP6 models using the training period 1985–2005, with performance evaluated over the independent testing period 2006–2014. Table 5 presents the raw and corrected performance metrics for all six models.

Table 5: Raw and bias-corrected temperature performance metrics under linear scaling (2006–2014).

Model	RMSE (°C)		Bias (°C)		Correlation	
	Raw	Corrected	Raw	Corrected	Raw	Corrected
MPI-ESM1-2-LR	3.7076	3.7140	0.0495	-0.2244	0.8594	0.8594
NorESM2-MM	3.7214	3.6949	-0.5834	-0.3789	0.8667	0.8667
EC-Earth3-CC	3.8519	3.7802	0.7397	-0.0114	0.8643	0.8643
HadGEM3-GC31-LL	3.5919	3.5520	-0.5994	0.2724	0.8709	0.8709
GFDL-ESM4	4.0295	3.9134	-0.9860	-0.2228	0.8603	0.8603
ACCESS-CM2	6.5558	5.7907	-3.0910	-0.3280	0.8298	0.8298

Linear scaling reduced the mean bias for all six models, with bias after correction approaching 0, but HadGEM had the best overall performance as it had the lowest corrected RMSE of 3.55°C and the highest correlation coefficient of 0.87 of the six models. NorESM followed closely with a corrected RMSE of 3.69°C. ACCESS-CM2 had the biggest absolute error improvement (6.56°C to 5.79°C), but remained the model with the lowest skill after bias correction. Correlation coefficients before and after bias correction were always the same in all model structures, indicating that linear rescaling preserved the temporal variability of the raw model output without modifying it. Note, however, that for MPI the near-zero raw bias of 0.05°C was overcorrected slightly to -0.22°C, indicating that linear scaling methods may introduce small errors when the raw bias is small. Figure 4 shows the mean temperature bias of ACCESS-CM2 during the testing period (2006–2014) as a representative model.

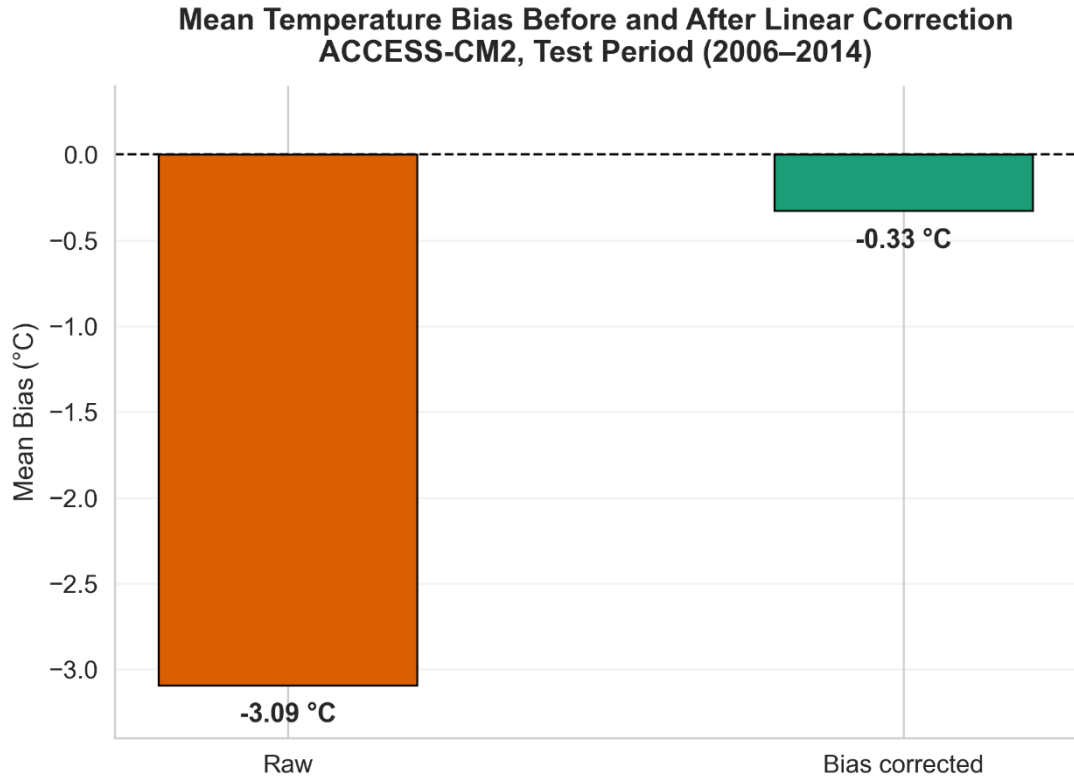


Figure 4: Mean Temperature Bias Before and After Linear Scaling Correction for ACCESS-CM2 (2006–2014).

These results correspond with past research showing that linear scaling is effective in removing systematic mean bias while respecting temporal variability (Teutschbein & Seibert, 2012; Lenderink et al., 2007). It is unsurprising that the RMSE improvement in the majority of models is small, given that mean-based bias correction methods do not account for errors in daily variability or the shape of the distribution (Azman et al., 2022). Models with larger biases at initialization, including ACCESS-CM2, tend to show larger reductions in RMSE, as seen in the work of Christensen et al. (2008) on regional climate model bias correction.

5.2.2. Precipitation

Following the same framework applied for temperature, linear scaling was applied to the precipitation outputs of all six CMIP6 models. Table 6 presents the raw and corrected performance metrics for the testing period 2006–2014.

Table 6: Raw and bias-corrected precipitation performance metrics under linear scaling (2006–2014).

Model	RMSE		Bias		Correlation	
	Raw	Corrected	Raw	Corrected	Raw	Corrected
MPI-ESM1-2-LR	5.2811	5.2334	0.2500	0.2123	0.1071	0.1071
EC-Earth3-CC	5.3408	5.6909	-0.1300	0.0935	0.0698	0.0698
HadGEM3-GC31-LL	6.3332	6.2286	0.1698	0.1189	0.0568	0.0568
NorESM2-MM	5.2191	5.7165	-0.2299	0.1130	0.0167	0.0167
GFDL-ESM4	5.3450	4.9859	0.1852	-0.1159	0.0615	0.0615
ACCESS-CM2	6.0872	5.8869	0.0852	-0.0255	-0.0057	-0.0057

While linear scaling reduced the mean precipitation bias for most models, RMSE improvements were inconsistent. The most notably improved model was GFDL, which decreased RMSE from 5.34 mm/day to 4.99 mm/day. The RMSE slightly increased for EC-Earth3 and NorESM, indicating that the multiplicative scaling factor calculated during the training period did not generalize well to the testing period. In contrast, the correlation coefficients were effectively unchanged by linear scaling, and were also near-zero, indicating that daily precipitation events are not represented better in time with linear scaling. ACCESS shows weak negative correlations both with and without correction suggesting that precipitation is still hard to represent at the local scale. Overall these results suggest that while linear scaling fixes some aspects of the mean precipitation amount, this correction is not sufficient to address the distributional and timing errors in raw CMIP6.

Figure 5 shows the daily precipitation for ACCESS before and after linear scaling during a high-rainfall period, which emphasizes the corrected output's ongoing failure to represent observed precipitation peaks.

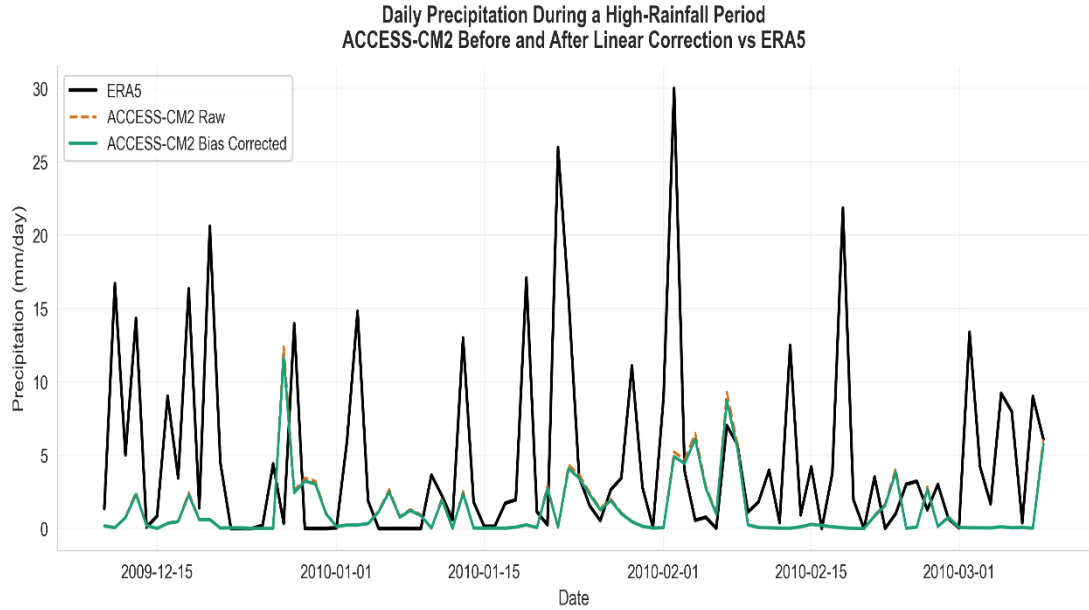


Figure 5: Daily precipitation for ACCESS-CM2 before and after linear scaling compared to ERA5 observations during a high-rainfall period (December 2009 – March 2010).

The modest improvement in precipitation metrics following linear scaling is consistent with previous studies on bias correction. Teutschbein and Seibert (2012) found that linear scaling is only able to remove the mean bias and cannot correct the bias in variability or in distribution shape. Azman et al. (2022) had also come to a similar conclusion that linear scaling works poorly for the correction of precipitation (at the local level), especially when there is high variability of the bias during the wet seasons. All of these results point to the inadequacy of mean-based correction techniques for precipitation bias correction, which encourages the use of more complex distribution-based strategies like quantile mapping.

5.3. Quantile Mapping Results

5.3.1. Temperature Results

Table 7 presents the quantile mapping bias correction results for daily temperature across all six CMIP6 models, comparing raw and corrected performance metrics for the testing period 2006–2014.

Table 7: Raw and bias-corrected temperature performance metrics under quantile mapping (2006–2014).

Model	RMSE		Bias		Correlation	
	Raw	Corrected	Raw	Corrected	Raw	Corrected
MPI-ESM1-2-LR	3.7080	3.7680	0.0490	-0.1890	0.8590	0.8640
NorESM2-MM	3.7210	3.6960	-0.5830	-0.4010	0.8670	0.8700
EC-Earth3-CC	3.8520	3.6340	0.7400	-0.0890	0.8640	0.8670
HadGEM3-GC31-LL	3.5920	3.5250	-0.5990	0.2540	0.8710	0.8750
GFDL-ESM4	4.0300	3.7480	-0.9860	-0.3010	0.8600	0.8610
ACCESS-CM2	6.5560	4.1860	-3.0910	-0.4730	0.8300	0.8380

Quantile mapping led to a decrease in RMSE of temperature for all the six models. Corrected RMSE values ranged from 3.53 to 4.19°C, in contrast to raw values that varied between 3.59 and 6.56°C. Unlike linear scaling, quantile mapping improved modeled temperature correlation for all six models, meaning the distribution of modeled temperature aligned better with the distribution of ERA5. HadGEM model had the lowest RMSE and highest correlation coefficient (0.875) after correction, with an RMSE of 3.53. ACCESS had the highest order of absolute change from correction, with its RMSE improving from 6.56 to 4.19°C. Bias correction performed well for all models, although there was a slight overcorrection with the MPI model, in which the raw bias of +0.05°C shifted to -0.19°C. Tables 8 and 9 present the distributional statistics for temperature outputs before and after quantile mapping correction, compared against ERA5 observations.

Table 8: Standard deviation and 5th percentile of raw and quantile mapping corrected temperature.

Model	STD			P5		
	Raw	QM	ERA5	Raw	QM	ERA5
MPI-ESM1-2-LR	6.6470	7.2380	7.1930	4.4610	3.5550	4.1150
NorESM2-MM	7.0390	7.2390	7.1910	3.4220	3.5000	4.1200
EC-Earth3-CC	7.3140	6.7910	7.1930	4.4040	4.4720	4.1150
HadGEM3-GC31-LL	6.5760	6.7840	7.1750	4.1820	4.8370	4.1460
GFDL-ESM4	7.5400	6.9370	7.1910	2.9180	4.4670	4.1200
ACCESS-CM2	10.1310	7.4030	7.1930	-4.2140	3.1840	4.1150

Table 9: 50th and 95th percentile of raw and quantile mapping corrected temperature.

Model	P50			P95		
	Raw	QM	ERA5	Raw	QM	ERA5
MPI-ESM1-2-LR	14.9090	14.6380	14.6090	25.0250	25.3360	25.6490
NorESM2-MM	14.3110	14.7040	14.6210	25.5030	25.2480	25.6500
EC-Earth3-CC	15.2790	14.6040	14.6090	26.8270	24.8530	25.6490
HadGEM3-GC31-LL	14.7950	15.4250	14.6250	24.5160	25.4040	25.6270
GFDL-ESM4	13.3710	14.5920	14.6210	25.9470	25.1700	25.6500
ACCESS-CM2	12.2370	14.5060	14.6090	26.5900	25.0810	25.6490

Quantile mapping significantly enhanced the depiction of temperature variability beyond mean correction alone, according to the distributional statistics. The standard deviation values obtained with quantile mapping were closer to the observed standard deviation value of 7.19°C in ERA5 than the raw values for most of the models. Post-correction P5 values of the lower tail were closer to the ERA5 (4.1°C) reference in all models, with ACCESS-CM2 showing the largest improvement in P5 values. P50 values were well-modeled in all models under quantile mapping, with a median difference of 0.05°C from the ERA5 (14.6°C) reference. For upper tail, P95 corrected values also came closer to the ERA5 reference of 25.6°C, although again certain models were a little off. Relative to historical conditions, these results show that quantile mapping corrects for the mean and shape of the distribution of temperature, something that simple linear scaling cannot do. These results show that quantile mapping effectively corrects not only the mean but also the distributional shape of temperature. Figure 6 presents the quantile-quantile plot for ACCESS-CM2 temperature before and after quantile mapping correction.

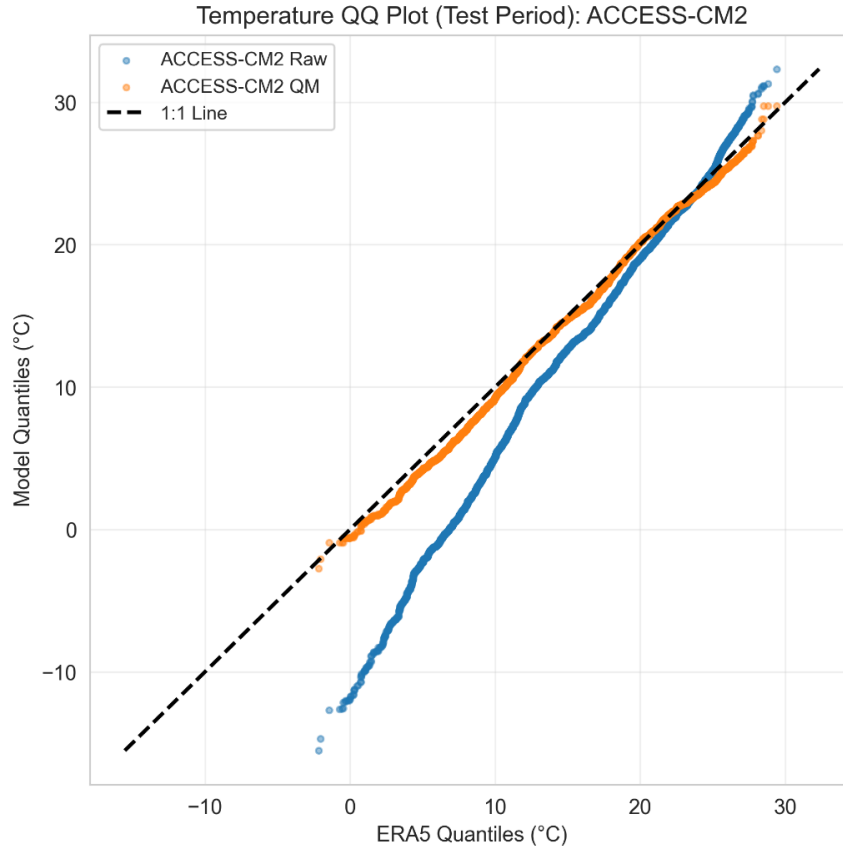


Figure 6: Quantile-quantile plot for ACCESS-CM2 daily temperature before and after quantile mapping correction against ERA5 observations (2006–2014).

As can be seen in figure 6, the raw simulation deviated substantially from the 1:1 line (which represents perfect agreement between the model and ERA5 quantiles), especially at low quantiles, indicating a cold bias. After the correction, the QM curve aligns closely with the 1:1 line, and the simulation is much better at both the lower and upper ends of temperature. This confirms the ability of Quantile Mapping to correct distributional biases in temperature simulations.

The improvement in temperature distribution following quantile mapping is consistent with findings from Tong et al. (2021), who found that quantile mapping reduced systematic bias in temperature in regional climate model experiments over China during the validation period. Themeßl et al. (2012) reported that empirical quantile mapping reduced temperature distribution biases by about an order of magnitude among European regional climate models. These studies show that quantile mapping is a clear improvement over mean-based correction methods for temperature.

5.3.2. Precipitation

Table 10 presents the quantile mapping and wet-day QM bias correction results for daily precipitation across all six CMIP6 models, comparing raw and corrected performance metrics for the testing period 2006–2014.

Table 10: Raw, QM, and wet-day QM precipitation performance metrics for all six CMIP6 models (2006–2014).

Model	RMSE			Bias			Correlation		
	Raw	QM	Wet-QM	Raw	QM	Wet-QM	Raw	QM	Wet-QM
MPI-ESM1-2-LR	5.28	5.78	5.74	0.25	0.23	0.38	0.11	0.10	0.11
EC-Earth3-CC	5.34	5.75	5.60	-0.13	0.10	0.07	0.07	0.07	0.07
HadGEM3-GC31-LL	6.33	5.82	5.90	0.17	0.10	0.14	0.06	0.06	0.06
NorESM2-MM	5.22	5.72	5.49	-0.23	0.10	-0.11	0.02	0.02	0.02
GFDL-ESM4	5.35	5.31	5.72	0.19	-0.14	0.47	0.06	0.06	0.06
ACCESS-CM2	6.09	5.96	6.18	0.09	-0.02	0.26	-0.01	-0.01	0.01

Quantile mapping and wet-day quantile mapping produced mixed results for precipitation across the six models. While both corrections generally reduced bias, with corrected values closer to zero than raw outputs, RMSE improvements were inconsistent. The most notable RMSE reduction was observed for GFDL under standard QM, while MPI and EC-Earth3 showed slight RMSE increases after correction, suggesting that the correction functions derived during the training period did not generalize well to the testing period. Correlation coefficients remained near zero across all models and correction methods, confirming that neither QM nor wet-day QM improved the temporal representation of daily precipitation events. ACCESS-CM2 exhibited a near-zero or slightly negative correlation before and after correction, consistent with the difficulty of simulating precipitation timing at the local scale observed under linear scaling. Both methods reduced mean bias but did not substantially affect RMSE and correlation, reflecting the inherent challenge of correcting daily precipitation at the station scale. Tables 11 and 12 present the distributional statistics for precipitation, comparing raw, QM, and wet-day QM corrected outputs against ERA5 observations in terms of standard deviation and percentile indices (P50, P95) respectively.

Table 11: Standard deviation of raw, QM, and wet-day QM corrected precipitation compared to ERA5 (2006–2014).

Model	STD			
	Raw	QM	Wet-QM	ERA5
MPI-ESM1-2-LR	3.82	4.54	4.49	4.07
EC-Earth3-CC	3.75	4.35	4.15	4.07
HadGEM3-GC31-LL	5.10	4.42	4.53	4.06
NorESM2-MM	3.33	4.08	3.75	4.07
GFDL-ESM4	3.72	3.65	4.23	4.07
ACCESS-CM2	4.51	4.33	4.64	4.07

Table 12: 50th and 95th percentile of raw, QM, and wet-day QM corrected precipitation compared to ERA5 (2006–2014).

Model	P50				P95			
	Raw	QM	Wet-QM	ERA5	Raw	QM	Wet-QM	ERA5
MPI-ESM1-2-LR	0.19	0.11	0.20	0.09	10.45	10.40	10.82	9.83
EC-Earth3-CC	0.08	0.08	0.00	0.09	8.59	9.85	9.80	9.83
HadGEM3-GC31-LL	0.11	0.09	0.12	0.09	10.46	9.85	9.65	9.84
NorESM2-MM	0.04	0.11	0.00	0.09	8.66	10.08	9.63	9.84
GFDL-ESM4	0.37	0.08	0.40	0.09	9.33	8.80	11.33	9.84
ACCESS-CM2	0.17	0.08	0.22	0.09	10.14	9.44	10.17	9.83

The distributional characteristics given in Tables 11 and 12 show that both standard and wet-day QM made distributions closer to ERA5 observations. Corrected standard deviations were closer to the ERA5 reference of 4.07 mm/day for all models under both QM corrections, however NorESM was closest under standard QM. Both methods considerably improved the median P50 values, which were near the ERA5 reference value of 0.09 mm/day for nearly all cases compared to the values of 0.04 mm/day through to 0.37 mm/day without correction. The 95th percentile values generally better captured the extremes compared to ERA5's 9.83 mm/day, though some models exceeded the ERA5 reference after correction. Overall, the wet-day QM produced slightly better distributional agreement than standard QM for many metrics, including the median precipitation value. Figure 7 present quantile-quantile plot for ACCESS-CM2 precipitation before and after QM and wet-day QM correction.

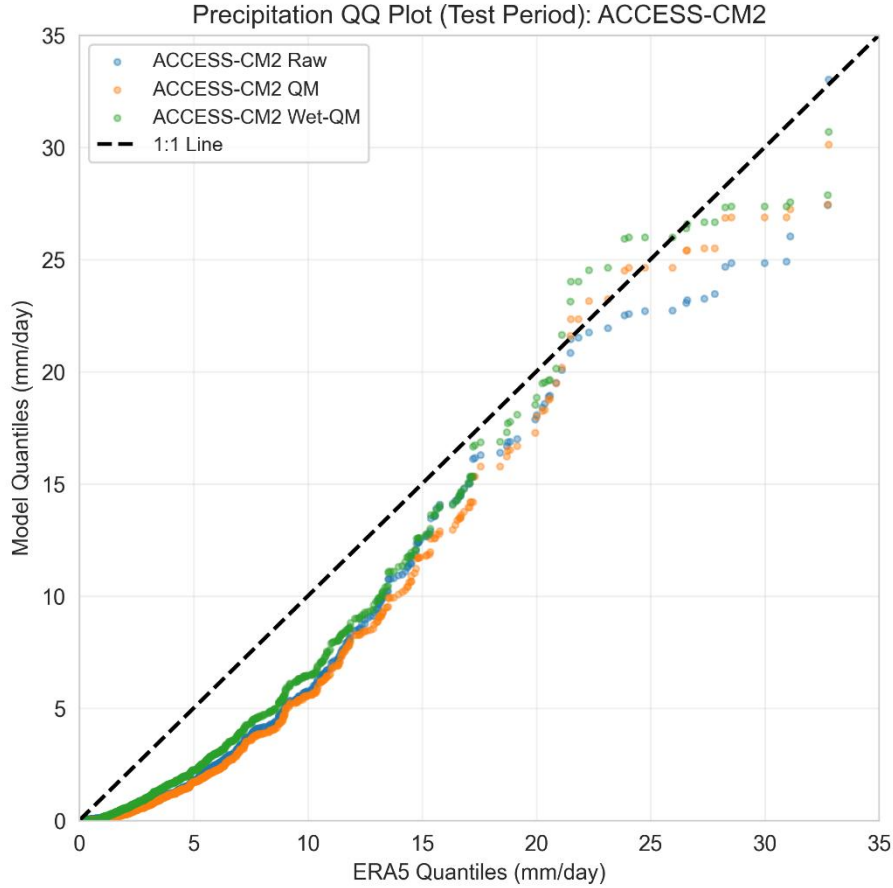


Figure 7: Quantile-quantile plot of ACCESS-CM2 daily precipitation before and after QM and wet-day QM correction against ERA5 observations (2006–2014).

The raw simulation systematically underestimates precipitation across all quantiles. Both QM and wet-day QM bring the model closer to ERA5 observations for the lower and middle quantiles compared to the raw simulation. The corrected outputs diverge from the 1:1 line at higher quantiles, indicating that extreme precipitation remains underestimated after correction. The wet-day QM curve follows ERA5 more closely at lower quantiles, likely due to its additional correction of wet-day frequency in addition to precipitation intensity.

The relatively small improvement of extreme precipitation quantiles after bias correction is consistent with the results of Lafon et al. (2013), who found that while the mean and variance of precipitation could be corrected with good robustness, correction of higher order moments such as extremes is much more sensitive to the choice of bias correction procedure and calibration period, which may explain the inconsistent behavior at higher quantiles. Standard QM also modifies extremes in precipitation trends when the correction function is extrapolated beyond its training

range (Cannon et al., 2015). This is consistent with Schmidli et al. (2006), whose findings suggest that the wet-day QM's closer match to ERA5 at lower quantiles compared to higher quantiles could be a result of correcting wet-day frequency alongside precipitation intensity, more effectively capturing the lower end of the distribution.

5.4. Artificial Neural Network Results

5.4.1. Temperature

Table 13 presents the ANN bias correction results for daily temperature across all six CMIP6 models, comparing raw and corrected performance metrics for the testing period 2006–2014.

Table 13: ANN bias correction performance metrics for daily temperature across six CMIP6 models (2006–2014).

Model	RMSE		Bias		Correlation	
	Raw	ANN	Raw	ANN	Raw	ANN
MPI-ESM1-2-LR	3.708	3.433	0.049	-0.451	0.859	0.881
NorESM2-MM	3.721	3.350	-0.583	-0.611	0.867	0.889
HadGEM3-GC31-LL	3.592	3.266	-0.599	-0.138	0.871	0.891
GFDL-ESM4	4.030	3.523	-0.986	-0.508	0.860	0.875
EC-Earth3-CC	3.852	3.465	0.740	-0.350	0.864	0.878
ACCESS-CM2	6.556	3.851	-3.091	-0.525	0.830	0.848

Bias correction with ANN reduced RMSE across all models, with corrected values ranging from 3.26°C to 3.85°C, as opposed to a raw model RMSE spread of 3.59°C to 6.55°C. For most models, the correlation improved slightly, with the highest being NorESM with a value of 0.891°C. HadGEM had the smallest RMSE after ANN bias correction, with its raw bias of -0.599°C being corrected to -0.138°C. ACCESS-CM2 retained the highest corrected RMSE (3.851°C) despite having the largest raw bias correction, from -3.091°C down to -0.525°C. Overall, biases were reduced across the ensemble but remained negative for several models, suggesting that the ANN tends to slightly underestimate observed temperature. Tables 14 and 15 present the ANN bias correction performance for temperature extremes across all six CMIP6 models,

reporting RMSE and bias under cold and hot conditions respectively for the testing period 2006–2014.

Table 14: ANN temperature extremes (cold and hot RMSE) across six CMIP6 models (2006–2014).

Model	Cold RMSE		Hot RMSE	
	Raw	ANN	Raw	ANN
ACCESS-CM2	6.239	6.322	3.342	4.451
EC-Earth3-CC	5.571	5.724	2.925	4.208
GFDL-ESM4	5.211	6.039	2.990	3.420
HadGEM3-GC31-LL	6.093	6.483	3.880	3.614
MPI-ESM1-2-LR	6.696	5.816	3.017	3.671
NorESM2-MM	5.516	5.618	3.194	3.710

Table 15: ANN temperature extremes (cold and hot Bias) across six CMIP6 models (2006–2014).

Model	Cold Bias		Hot Bias	
	Raw	ANN	Raw	ANN
ACCESS-CM2	-2.083	5.729	-1.972	-4.251
EC-Earth3-CC	4.120	5.432	-1.223	-4.039
GFDL-ESM4	3.743	5.804	-1.674	-3.192
HadGEM3-GC31-LL	4.787	6.094	-3.213	-3.128
MPI-ESM1-2-LR	5.727	5.501	-2.582	-3.527
NorESM2-MM	4.464	5.353	-2.171	-3.425

While mixed results were found for temperature extremes, the only model showing improved cold RMSE after correction was MPI-ESM1-2-LR, decreasing from 6.69°C to 5.81°C, while all remaining models saw cold RMSE increase. Hot RMSE similarly degraded across all models under ANN correction, most notably ACCESS-CM2, rising from 3.34°C to 4.45°C. As for bias, cold bias shifted to consistently positive values after correction for all models, with HadGEM reaching the highest corrected value of 6.094°C. Additionally, hot bias became more negative post-correction across the ensemble, with ACCESS-CM2 showing the largest magnitude at -4.25°C. The results demonstrate that although ANN decreases mean error, it fails to accurately capture temperature extremes, overestimating cold circumstances and underestimating hot conditions. this pattern aligns with regression to the mean, a recognized weakness of ANN models optimized for mean squared error. Figure 8 shows the scatter plot for the ACCESS-CM2 model, used as a representative model.

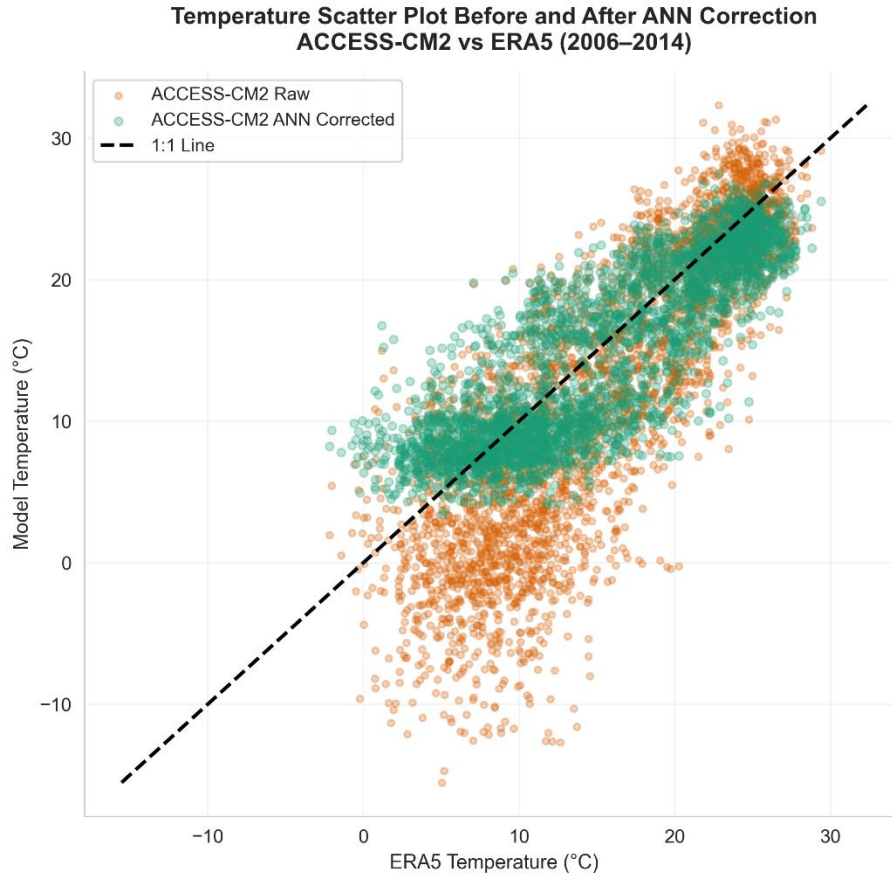


Figure 8: Scatter plot of ANN-corrected versus ERA5 observed daily temperature for ACCESS-CM2 (2006–2014).

In the scatter plot, ANN-corrected ACCESS-CM2 temperatures are much more tightly clustered around the 1:1 line than the uncorrected model output. The raw ACCESS-CM2 values systematically fall below the 1:1 line, indicate clear cold bias. This offset is largely removed following the ANN correction, with the largest residual scatter seen at the lower and upper tails, consistent with the regression to the mean behavior noted earlier. Figure 9 shows the Taylor diagram for the ACCESS-CM2 model.

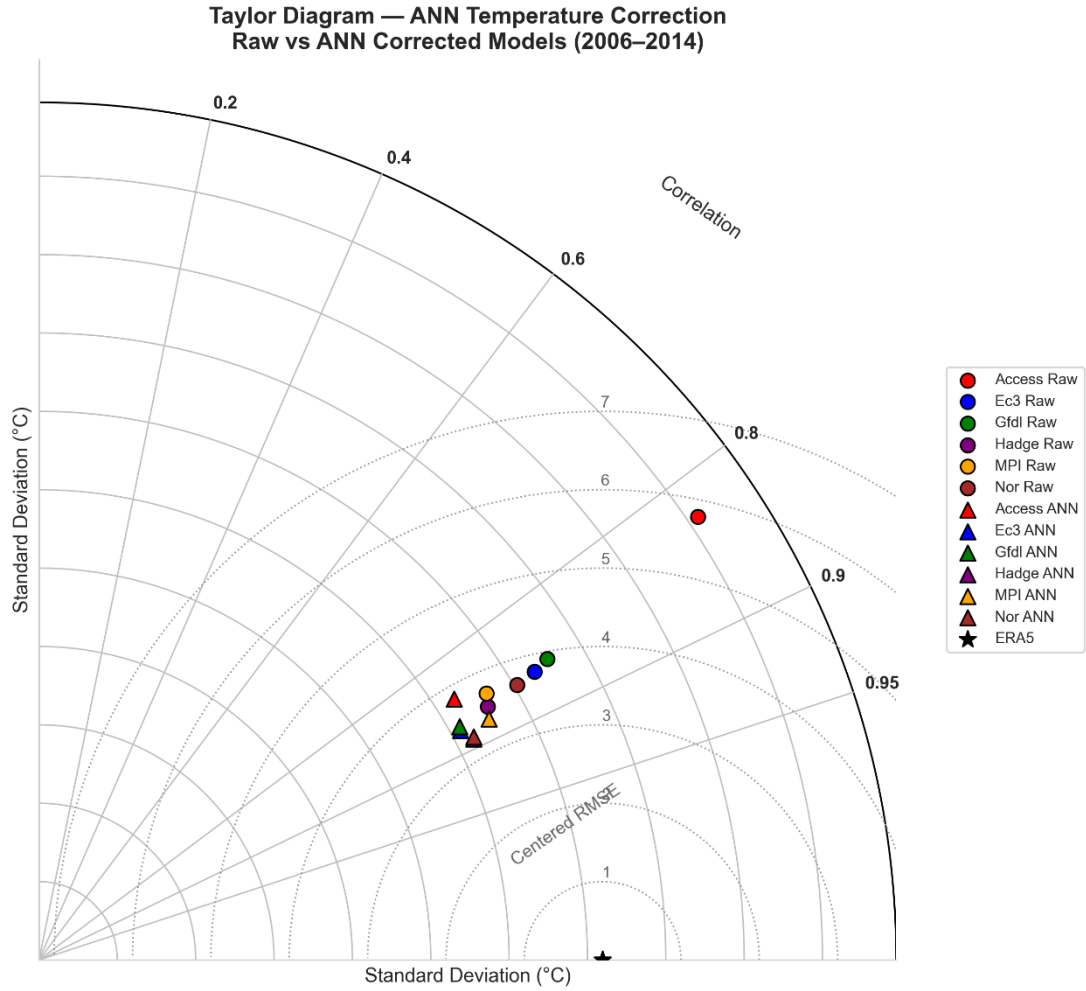


Figure 9: Taylor diagram comparing raw and ANN-corrected temperature for ACCESS-CM2 against ERA5 observations (2006–2014).

Figure 9 confirms the improvement brought by ANN correction across all six models. The raw model points show a wider spread in standard deviation and correlation, with ACCESS-CM2 deviating the farthest from the ERA5 reference point. After ANN correction, all models approach the ERA5 reference, clustering at correlation values of 0.9 and above. The spread of ANN-corrected models is also tighter than that of the raw outputs, implying that ANN correction reduces inter-model spread in addition to reducing bias in individual model runs.

Studies on ANN-based temperature bias correction have found that neural networks generally outperform traditional statistical methods in reducing mean bias, with deep learning approaches showing competitive performance in temperature downscaling (Baño-Medina et al., 2020; Vandal et al., 2017). However, their added

value is largely restricted to mean correction, with limited improvement for extremes (Baño-Medina et al., 2020). This aligns with the regression to the mean behavior observed in this study, where overall RMSE was reduced but tail performance degraded. Vandal et al. (2017) reported similar findings, noting that neural network frameworks capture large-scale climate signals but struggle to reproduce tail distributions.

5.4.2. Precipitation

Table 16 presents the ANN bias correction results for daily precipitation across all six CMIP6 models, comparing raw and corrected performance metrics for the testing period 2006–2014.

Table 16: ANN bias correction performance metrics for daily precipitation across six CMIP6 models (2006–2014).

Model	RMSE		Bias		Correlation	
	Raw	ANN	Raw	ANN	Raw	ANN
MPI-ESM1-2-LR	5.281	4.048	0.250	0.102	0.107	0.099
NorESM2-MM	5.219	4.071	-0.230	0.176	0.017	0.028
HadGEM3-GC31-LL	6.333	4.049	0.170	0.160	0.057	0.078
GFDL-ESM4	5.345	4.063	0.185	0.114	0.062	0.062
EC-Earth3-CC	5.341	4.051	-0.130	0.102	0.070	0.096
ACCESS-CM2	6.087	4.072	0.085	0.108	-0.006	-0.003

ANN bias correction reduced RMSE across all six models, with corrected values converging narrowly between 4.048 and 4.072 mm/day, compared to a raw spread of 5.219 to 6.333 mm/day. However, correlation coefficients remained very low after correction for all models, with the highest being 0.099 for MPI-ESM1-2-LR and the lowest -0.003 for ACCESS-CM2. Most models showed reduced bias, though corrections were of both signs. NorESM shifted from a raw bias of -0.230 mm/day to a corrected bias of 0.176 mm/day. Despite the homogenization of RMSE across the ensemble, the near-zero correlations suggest that the ANN struggles to capture the temporal structure of daily precipitation, a known challenge in point-scale precipitation modelling.

Table 17 presents the ANN bias correction performance for precipitation extremes across all six CMIP6 models, reporting extreme RMSE and extreme bias for both raw and ANN-corrected outputs (2006–2014).

Table 17: ANN bias correction performance for precipitation extremes across six CMIP6 models (2006–2014).

Model	Extreme Bias		Extreme RMSE	
	Raw	ANN	Raw	ANN
MPI-ESM1-2-LR	-12.387	-13.930	15.306	15.530
NorESM2-MM	-14.447	-14.050	16.361	15.639
HadGEM3-GC31-LL	-13.559	-13.965	15.938	15.567
GFDL-ESM4	-13.101	-14.062	16.085	15.650
EC-Earth3-CC	-13.462	-14.066	15.824	15.653
ACCESS-CM2	-14.375	-14.067	16.656	15.645

For precipitation extremes, ANN correction yielded only limited improvement, with small reductions in extreme RMSE observed across all models except MPI-ESM1-2-LR, where extreme RMSE marginally increased from 15.306 to 15.530 mm/day. The largest reduction was found for NorESM, from 16.361 to 15.639 mm/day. Extreme bias worsened after correction for all models, with ACCESS-CM2 reaching 14.067 mm/day. These results indicate that while ANN correction modestly reduces extreme RMSE, it consistently amplifies the underestimation of precipitation extremes, further confirming the regression to the mean behavior noted earlier. Figure 10 presents a representative time series of raw and ANN-corrected daily precipitation for ACCESS-CM2 against ERA5 observations during a high rainfall period (December 2009 – March 2010).

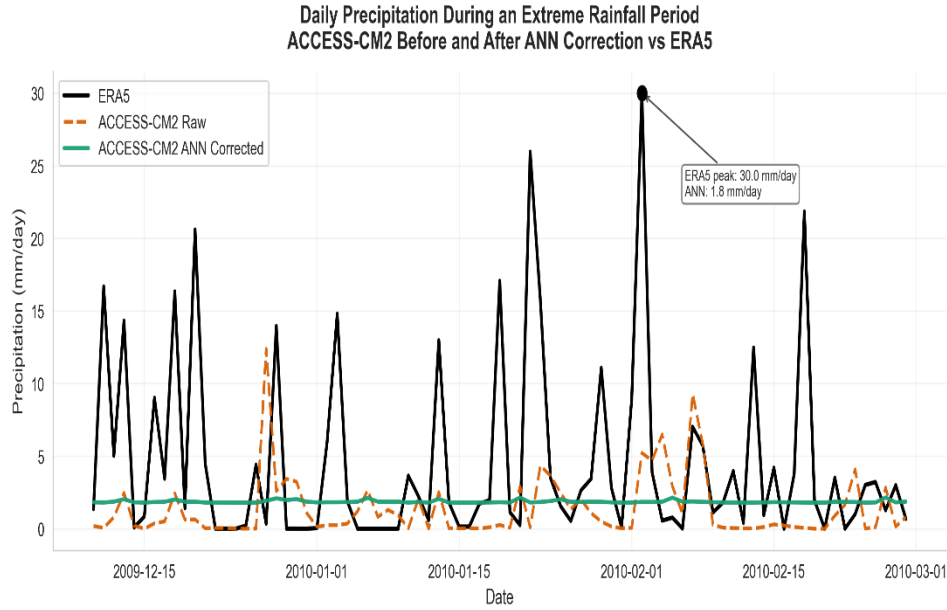


Figure 10: Representative time series of raw and ANN-corrected daily precipitation for ACCESS-CM2 against ERA5 observations during a high rainfall period (December 2009 – March 2010).

Figure 10 illustrates a clear limitation of ANN bias correction for precipitation. The raw ACCESS-CM2 output closely follows the temporal structure of ERA5, while after ANN correction the output remains nearly constant throughout the period, failing to capture any of the observed peaks. At the ERA5 peak of approximately 30.0 mm/day in early February 2010, the ANN-corrected output was only 1.9 mm/day, a reduction of over 93%. This is a direct consequence of regression to the mean, where the network predicts values near the mean rather than reproducing the intermittent and highly variable nature of daily precipitation.

The difficulties of ANN-based bias correction methods in reproducing precipitation extremes are consistent with findings in the broader statistical downscaling literature. Baño-Medina et al. (2020) showed that there are significant advantages of deep learning over classical methods for correcting mean precipitation, but deep learning models tend to underestimate extremes. Bürger et al. (2012) showed that precipitation-related extreme indices were much harder to reproduce than temperature indices for a range of statistical downscaling methods. This suggests that difficulties in reproducing extreme values may be common among regression-based approaches that are primarily optimized using average error measures.

5.5. Comparison of Bias-Correction Methods

5.5.1. Temperature

Table 18, 19, and 20 present a direct comparison of RMSE, bias, and correlation across all bias correction methods for daily temperature during the testing period 2006–2014.

Table 18: RMSE comparison of raw and bias-corrected daily temperature across all methods (2006–2014).

Model	RMSE			
	Raw	Linear	QM	ANN
ACCESS-CM2	6.556	5.791	4.186	3.851
EC-Earth3-CC	3.852	3.780	3.634	3.465
GFDL-ESM4	4.030	3.913	3.748	3.523
HadGEM3-GC31-LL	3.592	3.552	3.525	3.266
MPI-ESM1-2-LR	3.708	3.714	3.768	3.433
NorESM2-MM	3.721	3.695	3.696	3.350

Table 19: Bias comparison of raw and bias-corrected daily temperature across all methods (2006–2014).

Model	Bias			
	Raw	Linear	QM	ANN
ACCESS-CM2	-3.091	-0.328	-0.473	-0.525
EC-Earth3-CC	0.740	-0.011	-0.089	-0.350
GFDL-ESM4	-0.986	-0.223	-0.301	-0.508
HadGEM3-GC31-LL	-0.599	0.272	0.254	-0.138
MPI-ESM1-2-LR	0.049	-0.224	-0.189	-0.451
NorESM2-MM	-0.583	-0.379	-0.401	-0.611

Table 20: Correlation comparison of raw and bias-corrected daily temperature across all methods (2006–2014).

Model	Correlation			
	Raw	Linear	QM	ANN
ACCESS-CM2	0.830	0.830	0.838	0.848
EC-Earth3-CC	0.864	0.864	0.867	0.878
GFDL-ESM4	0.860	0.860	0.861	0.875
HadGEM3-GC31-LL	0.871	0.871	0.875	0.891
MPI-ESM1-2-LR	0.859	0.859	0.864	0.881
NorESM2-MM	0.867	0.867	0.870	0.889

Comparison of the correction methods for daily temperature is shown in Tables 18, 19 and 20. In general, the ANN achieved the lowest RMSE values. The RMSE for ACCESS-CM2 was reduced from 6.56°C to 3.85°C and for GFDL from 4.03°C to 3.52°C. Both linear scaling and quantile mapping were much less effective than ANN. The correlation coefficients improved the most with ANN. However, the ANN corrections introduced larger negative biases. The largest were for NorESM and ACCESS, with -0.611°C and -0.525°C respectively, indicating overcorrection in these two models. LS produced the smallest improvement in RMSE and correlation coefficient, while QM produced intermediate performance. ANN achieved the strongest overall temperature correction but introduced systematic negative biases in some models. Figure 11 shows Taylor diagram for ACCESS-CM2 temperature, comparing raw and all three bias-corrected outputs against ERA5 observations.

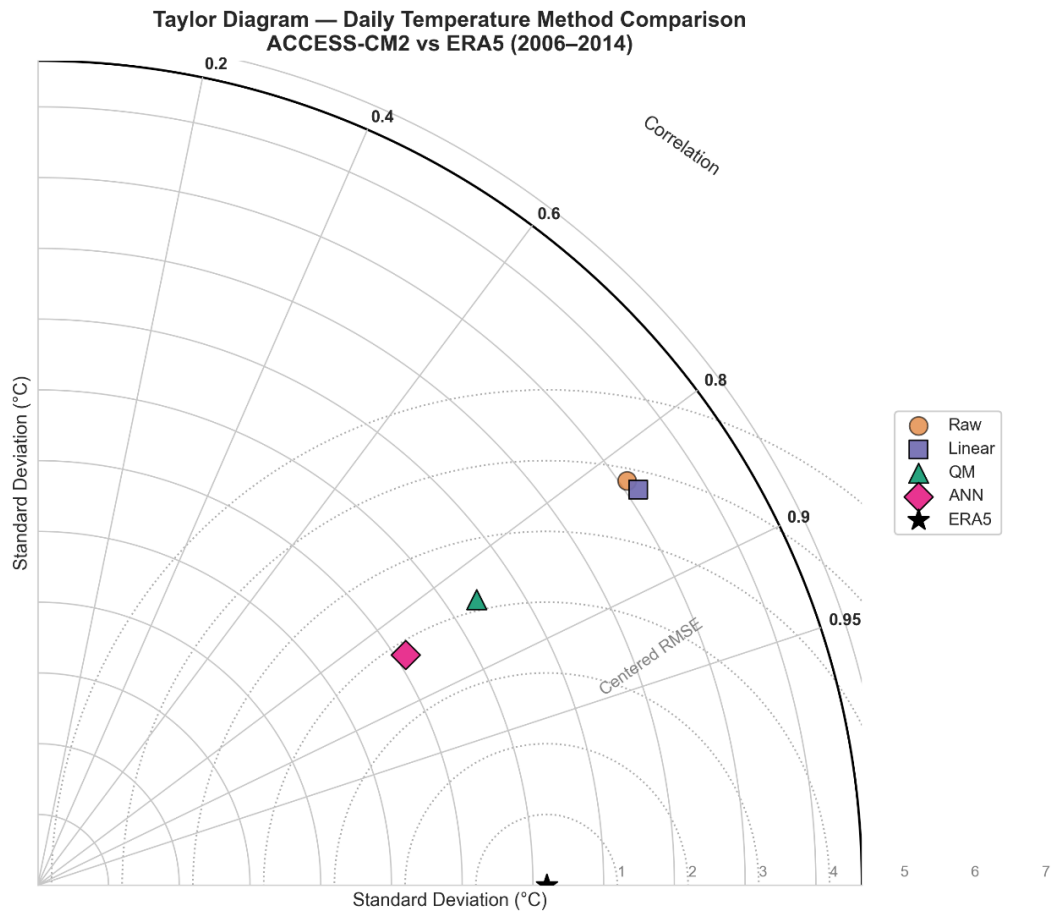


Figure 11: Taylor diagram comparing bias correction methods for ACCESS-CM2 daily temperature (2006–2014).

The Taylor diagram for ACCESS-CM2 (Figure 11) shows the same ordering of the methods as the metrics tables. The ANN corrected output is closest to the ERA5 reference point, achieving the highest correlation and lowest centered RMSE among all bias-corrected outputs. QM shows a marked improvement compared to the raw simulation in terms of both correlation and standard deviation, while linear scaling shows only marginal improvement, with the standard deviation and correlation remaining close to the raw simulation. Overall, ANN performs best for temperature correction, followed by QM, while linear scaling only achieves small improvements beyond mean bias removal.

The better performance of ANN temperature correction is consistent with Baño-Medina et al. (2020), who found that deep learning methods outperformed classical statistical ones in climate downscaling. The poor performance of linear scaling is also consistent with Teutschbein and Seibert (2012), who found that linear scaling ranked lowest among bias correction methods due to its inability to correct statistics beyond the mean. The intermediate performance of QM is consistent with Tong et al. (2021), who found that QM effectively removed systematic temperature biases during the validation period of regional climate model simulations.

5.5.2. Precipitation

Table 21, 22, and 23 present a direct comparison of RMSE, bias, and correlation across all bias correction methods for daily precipitation during the testing period 2006–2014.

Table 21: RMSE comparison of raw and bias-corrected daily precipitation across all methods (2006–2014).

Model	RMSE				
	Raw	Linear	QM	Wet-QM	ANN
ACCESS-CM2	6.087	5.887	5.963	6.183	4.072
EC-Earth3-CC	5.341	5.691	5.748	5.601	4.051
GFDL-ESM4	5.345	4.986	5.307	5.715	4.063
HadGEM3-GC31-LL	6.333	6.229	5.819	5.902	4.049
MPI-ESM1-2-LR	5.281	5.233	5.782	5.740	4.048
NorESM2-MM	5.219	5.716	5.718	5.489	4.071

Table 22: Bias comparison of raw and bias-corrected daily precipitation across all methods (2006–2014).

Model	Bias				
	Raw	Linear	QM	Wet-QM	ANN
ACCESS-CM2	0.085	-0.026	-0.019	0.255	0.108
EC-Earth3-CC	-0.130	0.093	0.095	0.066	0.102
GFDL-ESM4	0.185	-0.116	-0.136	0.471	0.114
HadGEM3-GC31-LL	0.170	0.119	0.101	0.136	0.160
MPI-ESM1-2-LR	0.250	0.212	0.234	0.383	0.102
NorESM2-MM	-0.230	0.113	0.103	-0.105	0.176

Table 23: Correlation comparison of raw and bias-corrected daily precipitation across all methods (2006–2014).

Model	Correlation				
	Raw	Linear	QM	Wet-QM	ANN
ACCESS-CM2	-0.006	-0.006	-0.007	-0.003	-0.003
EC-Earth3-CC	0.070	0.070	0.068	0.070	0.096
GFDL-ESM4	0.062	0.062	0.058	0.059	0.062
HadGEM3-GC31-LL	0.057	0.057	0.059	0.060	0.078
MPI-ESM1-2-LR	0.107	0.107	0.102	0.107	0.099
NorESM2-MM	0.017	0.017	0.016	0.017	0.028

For daily precipitation, ANN achieved the lowest RMSE across all six models, with corrected values ranging from 4.048 to 4.072 mm/day, considerably outperforming the other methods for this metric. For linear scaling, QM, and wet-day QM, RMSE improvement was inconsistent, with some models showing slight increases compared to raw outputs. Correlation coefficients remained near zero for all models and all correction methods, confirming that no method improved the temporal representation of daily precipitation events. All correction methods were somewhat effective in reducing bias, though wet-day QM introduced larger positive biases for several models, particularly GFDL at 0.471 mm/day and MPI-ESM1-2-LR at 0.383 mm/day. ANN showed the least variance in bias across models. However, RMSE alone may not fully capture method performance for precipitation, as a method predicting values close to the mean will achieve low RMSE while failing to reproduce the actual distribution, which provides a more complete assessment of distributional performance (Figure 12).

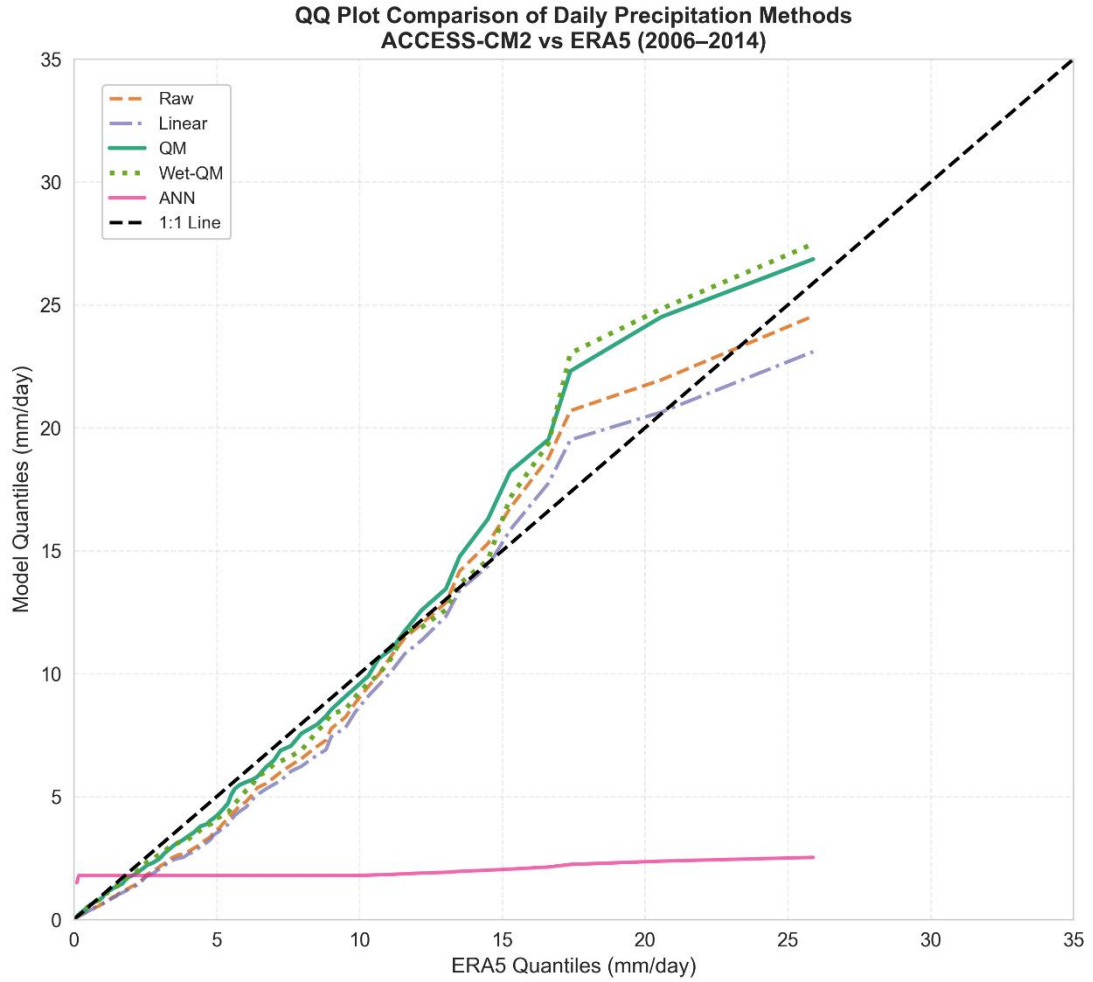


Figure 12: Quantile-quantile plot comparing all bias correction methods for ACCESS-CM2 daily precipitation against ERA5 observations (2006–2014).

The distributional performance of all correction methods is illustrated in Figure 12. In the lower to mid quantiles, both QM and wet-day QM approximately follow the 1:1 reference line. Linear scaling shows improvement over the raw simulation but remains below the 1:1 line for most quantiles. The ANN correction produces a near-zero line across all quantiles. While the RMSE is small, the correction simply rescales the distribution toward the mean, which is even more pronounced for moderate and heavy precipitation, both of which are completely under-predicted. These results suggest that wet-day QM provides the best balance between mean bias correction and reproduction of the precipitation distribution, while ANN's low RMSE reflects regression to the mean rather than genuine distributional bias correction.

The regression to the mean behavior of the ANN precipitation correction is in line with the results from Baño-Medina et al. (2020), who found that deep learning models tend to underestimate precipitation extremes even when performing well overall, as well as with Gudmundsson et al. (2012), who found that the distributional performance of quantile mapping is superior to that of linear scaling, with nonparametric QM achieving the highest skill at reducing precipitation biases. The importance of correcting not only precipitation amount but also the occurrence of precipitation, as achieved by wet-day QM, is supported by Vrac et al. (2016), who found that correcting wet-day occurrence and precipitation amounts simultaneously produces greater total bias correction skill than correcting precipitation amounts alone. Lafon et al. (2013) argued that precipitation bias correction evaluated solely on RMSE is insufficient, and that higher order moments are more sensitive to bias correction method choice than the mean.

5.6. Comparison with NASA NEX-GDDP Data

To be contextualized using NASA NEX-GDDP-CMIP6 data as an external benchmark, ANN-corrected output was chosen as the representative corrected product for temperature, and the wet-day QM corrected output was used for precipitation.

5.6.1. Temperature

Tables 24 and 25 present the standard performance metrics and extreme temperature indices (respectfully) for NASA NEX-GDDP compared against ERA5 observations for the testing period 2006–2014.

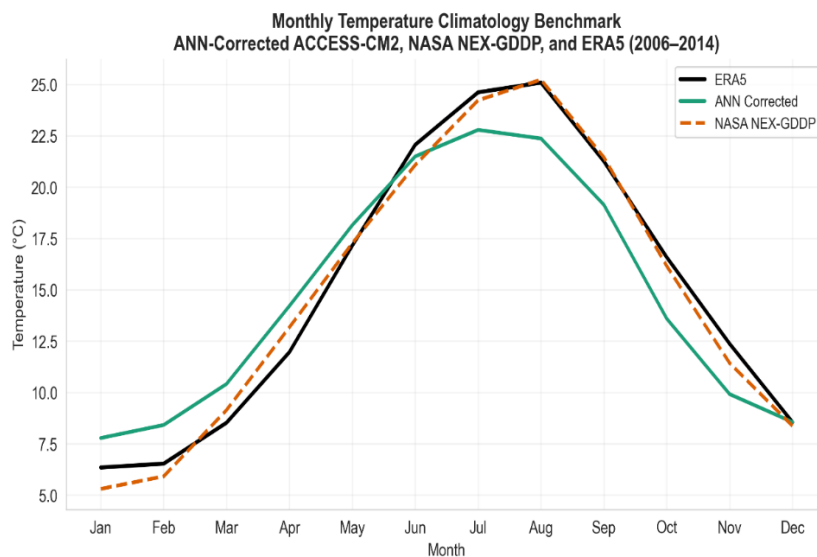
Table 24: NASA NEX-GDDP temperature performance metrics compared to ERA5 (2006–2014).

Model	Temperature		
	RMSE	Bias	Corr
ACCESS-CM2	4.146	0.147	0.835
EC-Earth3	4.197	0.695	0.838
GFDL-ESM4	4.279	0.119	0.830
HadGEM3-GC31-LL	4.136	0.505	0.843
MPI-ESM1-2-LR	4.359	0.372	0.825
NorESM2-MM	4.297	0.378	0.831

Table 25: NASA NEX-GDDP extreme temperature metrics compared to ERA5 (2006–2014).

Model	Temperature Extremes			
	Cold Bias	Hot Bias	Cold RMSE	Hot RMSE
ACCESS-CM2	5.008	-1.670	6.407	3.161
EC-Earth3	5.210	-0.825	6.688	2.602
GFDL-ESM4	4.587	-1.624	6.274	3.417
HadGEM3-GC31-LL	4.896	-0.951	6.297	2.854
MPI-ESM1-2-LR	5.107	-1.623	6.930	3.121
NorESM2-MM	4.930	-0.983	6.274	3.294

Despite model differences, NASA NEX-GDDP temperature metrics were similar with RMSE values ranging from 4.136 to 4.359°C. The mean bias ranged from 0.119 to 0.695°C, indicating that NASA systematically overestimated temperature. Correlation coefficients ranged from 0.825 to 0.843, indicating strong temporal agreement with ERA5 observations. All models showed largely positive cold biases for cold extremes, ranging from 4.587 to 5.210°C. Hot biases were negative across all models, ranging from -0.825 to -1.670°C, indicating underestimation of warm extremes. Cold RMSE was considerably larger than hot RMSE across all models, indicating that NASA performs better for warm conditions than cold conditions. Figure 13 shows the monthly temperature climatology for ANN-corrected ACCESS-CM2 and NASA NEX-GDDP compared to ERA5 observations.

**Figure 13:** Monthly temperature climatology benchmark comparing ANN-corrected ACCESS-CM2 and NASA NEX-GDDP against ERA5 observations (2006–2014).

Both datasets show a similar seasonal temperature pattern throughout the year. The highest temperature occurs in August and the lowest in January and February, representative of the Mediterranean climate of the Florya region. Although NASA follows the ERA5 seasonal cycle more closely throughout the year, with the ANN-corrected output diverging during winter and summer months, both products capture the overall seasonal temperature structure, suggesting that the locally developed ANN correction produces climatologically plausible results broadly consistent with the globally downscaled NASA benchmark (Thrasher et al., 2022).

5.6.2. Precipitation

Tables 26 and 27 present the standard performance metrics and extreme precipitation indices (respectfully) for NASA NEX-GDDP compared against ERA5 observations for the testing period 2006–2014.

Table 26: NASA NEX-GDDP precipitation performance metrics compared to ERA5 (2006–2014).

Model	Precipitation		
	RMSE	Bias	Corr
ACCESS-CM2	5.717	0.268	0.028
EC-Earth3	6.235	0.410	0.030
GFDL-ESM4	5.841	0.334	0.051
HadGEM3-GC31-LL	5.700	0.285	0.045
MPI-ESM1-2-LR	5.519	0.300	0.059
NorESM2-MM	5.984	0.256	0.029

Table 27: NASA NEX-GDDP extreme precipitation metrics compared to ERA5 (2006–2014).

Model	Precipitation Extremes	
	Extreme Bias	Extreme RMSE
ACCESS-CM2	-13.995	16.440
EC-Earth3	-13.827	16.334
GFDL-ESM4	-13.426	16.161
HadGEM3-GC31-LL	-14.085	16.175
MPI-ESM1-2-LR	-13.725	16.110
NorESM2-MM	-13.898	16.533

All six NASA NEX-GDDP models show comparable performance. RMSE values ranged from 5.519 to 6.235 mm/day, with MPI-ESM1-2-LR showing the lowest error and EC-Earth the highest. All models had positive biases ranging from 0.256 to

0.410 mm/day, while correlation coefficients were near zero across all models. Consistent with all bias correction methods evaluated in this study, this indicates that the timing of daily precipitation events at local scales was difficult to reproduce. Extreme precipitation bias was large and negative across all models, ranging from -13.426 to -14.085 mm/day, indicating consistent underestimation of extreme precipitation events relative to ERA5 observations. Extreme RMSE values were also high across all models, ranging from 16.110 to 16.533 mm/day. Figure 14 presents the wet-day precipitation cumulative distribution function for wet-QM corrected ACCESS-CM2 and NASA compared against ERA5 observations.

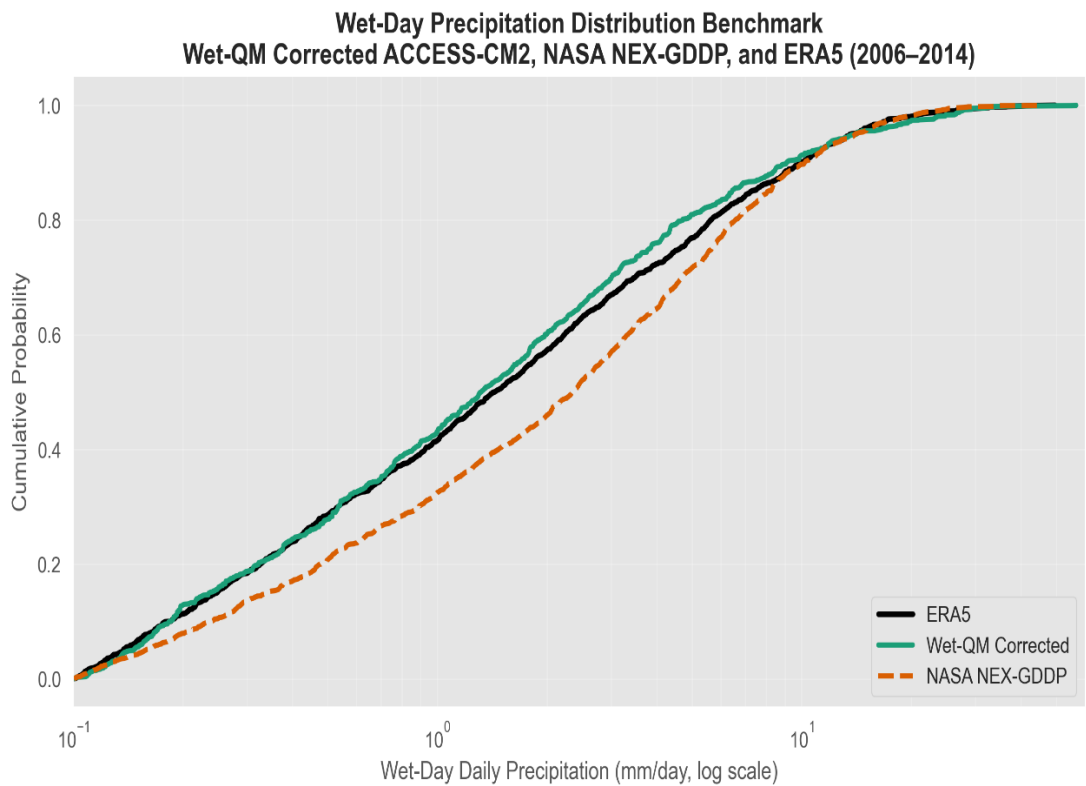


Figure 14: Wet-day precipitation distribution benchmark comparing wet-QM corrected ACCESS-CM2 and NASA against ERA5 observations (2006–2014).

As can be seen in Figure 14, the wet-QM corrected output closely follows the ERA5 distribution over the majority of the wet-day precipitation range. The NASA output shows large underestimation of wet-day precipitation throughout the lower to middle of the distribution. Both products converge on ERA5 at the upper tail, corresponding to heavy precipitation events. This result indicates that the locally

developed wet-day QM correction reproduces the wet-day precipitation distribution more accurately than the globally applied NASA product at this location, consistent with findings suggesting that global bias-corrected products may not optimally reproduce local precipitation characteristics (Thrasher et al., 2022).

5.7. Monthly Climate Analysis

For ACCESS-CM2 and GFDL-ESM4, a monthly evaluation was carried out to determine whether the bias correction techniques maintained realistic seasonal cycles throughout the entire year. Tables 28, 29, and 30 present the monthly temperature evaluation metrics (RMSE, bias, and correlation coefficient respectively) for ACCESS-CM2 and GFDL across all bias correction methods.

Table 28: Monthly temperature RMSE for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).

Model	RMSE			
	Raw	Linear	QM	ANN
GFDL-ESM4	2.126	1.897	1.861	1.962
ACCESS-CM2	5.089	4.046	2.559	2.435

Table 29: Monthly temperature bias for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).

Model	Bias			
	Raw	Linear	QM	ANN
GFDL-ESM4	-0.984	-0.221	-0.251	-0.410
ACCESS-CM2	-3.106	-0.341	-0.414	-0.473

Table 30: Monthly temperature correlation coefficient for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).

Model	Correlation			
	Raw	Linear	QM	ANN
GFDL-ESM4	0.963	0.963	0.963	0.961
ACCESS-CM2	0.922	0.922	0.933	0.936

Both models showed improved monthly temperature RMSE after bias correction, with GFDL-ESM4 reducing from 2.126°C to 1.861°C and ACCESS-CM2 from 5.089°C to 2.559°C under QM correction. QM and ANN produced the smallest residual biases across both models. Monthly correlation coefficients were already high

prior to correction, remaining above 0.92 for all methods, indicating that the seasonal temperature cycle was well preserved regardless of the correction approach applied. Tables 31, 32, and 33 present the monthly precipitation evaluation metrics (RMSE, bias, and correlation coefficient respectively) for ACCESS-CM2 and GFDL across all bias correction methods.

Table 31: Monthly total precipitation RMSE for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).

Model	RMSE			
	Raw	Linear	QM	ANN
GFDL-ESM4	41.9760	39.2370	39.9450	36.8010
ACCESS-CM2	53.5190	51.7400	51.3700	38.6050

Table 32: Monthly mean precipitation bias for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).

Model	Bias			
	Raw	Linear	QM	ANN
GFDL-ESM4	5.6340	-3.5240	-4.4350	3.3060
ACCESS-CM2	2.5920	-0.7780	-0.8000	1.5450

Table 33: Monthly precipitation correlation coefficient for ACCESS-CM2 and GFDL-ESM4 across bias correction methods (2006–2014).

Model	Correlation			
	Raw	Linear	QM	ANN
GFDL-ESM4	0.3860	0.3860	0.4030	0.3830
ACCESS-CM2	0.1300	0.1300	0.1310	0.1940

Monthly precipitation RMSE is considerably higher than daily RMSE because errors accumulate when aggregating daily projections to the monthly scale, rather than reflecting any loss of correction ability. ANN achieved the greatest RMSE reduction for both models, bringing GFDL-ESM4 from 41.976 to 36.801 mm/month and ACCESS-CM2 from 53.519 to 38.605 mm/month. Bias correction direction varied across methods for GFDL-ESM4 (Table 32), where linear scaling and QM produced negative corrected biases despite a positive raw bias, while ANN retained a small positive corrected bias of 3.306 mm/month. Low monthly correlation coefficients were likewise observed across all methods for both models, consistent with the near-zero daily correlations reported earlier, indicating poor representation of the temporal structure of precipitation at the monthly timescale. Figure 15 presents the monthly

ACCESS-CM2 evaluation before and after ANN correction for both temperature and precipitation against ERA5 observations over the testing period 2006–2014.

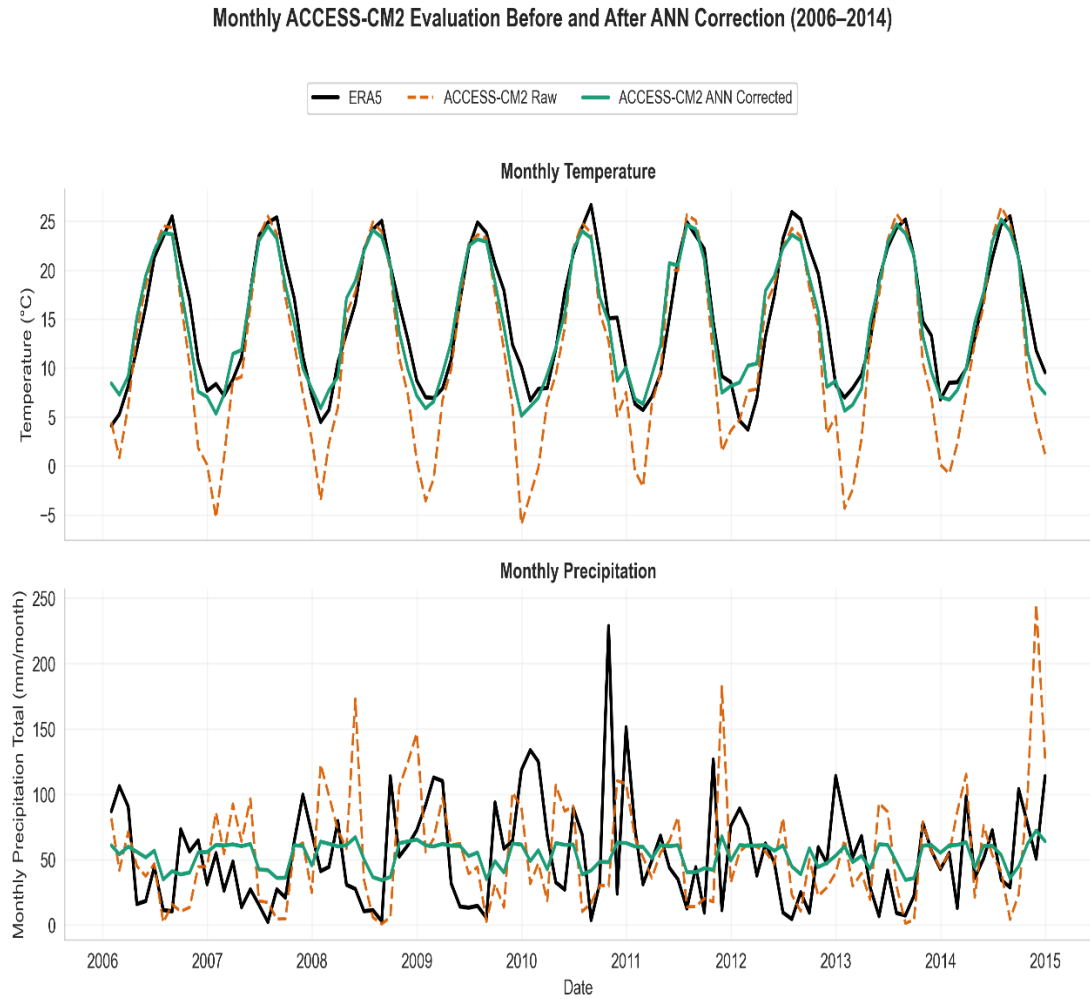


Figure 15: Monthly ACCESS-CM2 temperature and precipitation before and after ANN correction compared to ERA5 observations (2006–2014).

Monthly evaluation of bias correction methods has been widely used as a supplementary diagnostic to assess whether corrections preserve realistic seasonal cycles beyond mean error reduction. Teutschbein & Seibert (2012) found good performance for monthly temperature cycles across bias correction methods, with larger seasonal variation observed for precipitation. Gudmundsson et al. (2012) similarly showed that QM produces realistic monthly precipitation distributions, though residual biases can persist in individual months. Huang et al. (2023), evaluating deep learning and traditional bias correction approaches for CMIP6 precipitation across global continents, found that machine learning methods effectively reduced

ensemble mean bias, but limitations remained for extreme precipitation and seasonal variability, consistent with the low monthly correlations observed in this study.

5.8. Future Projections

5.8.1. Temperature Projections

Table 34 presents the ensemble mean temperature projections for the three future periods under SSP2-4.5 and SSP5-8.5.

Table 34: Ensemble mean and extreme temperature projections across three future periods under SSP2-4.5 and SSP5-8.5 (2021–2100).

Scenario	Period	Temperature		
		Mean	P95	P99
SSP245	2021-2040	15.664	26.020	27.640
SSP245	2041-2070	16.164	26.725	28.857
SSP245	2071-2100	16.741	27.445	29.885
SSP585	2021-2040	15.617	26.069	27.787
SSP585	2041-2070	16.692	27.480	29.908
SSP585	2071-2100	18.248	29.697	32.225

In both scenarios, warming occurs across all three time periods. For SSP2-4.5, the ensemble mean temperature rises from 15.664°C in the near-term to 16.741°C by 2071-2100. For SSP5-8.5, the near-future temperature of 15.617°C rises to 18.248°C in 2071-2100. For the P99 percentile, the temperature increases from 29.885°C in the near future to 32.225°C in 2071-2100. Another finding is that the largest difference is in 2071-2100, where SSP5-8.5 is 1.5°C higher than SSP2-4.5 in the late-century period due to stronger radiative forcing. Below, figure 16 shows the projection of ensemble annual mean temperature by emission scenario and figure 17 shows a 10-year moving average of the 95th percentile (P95) temperature from 2015 to 2100.

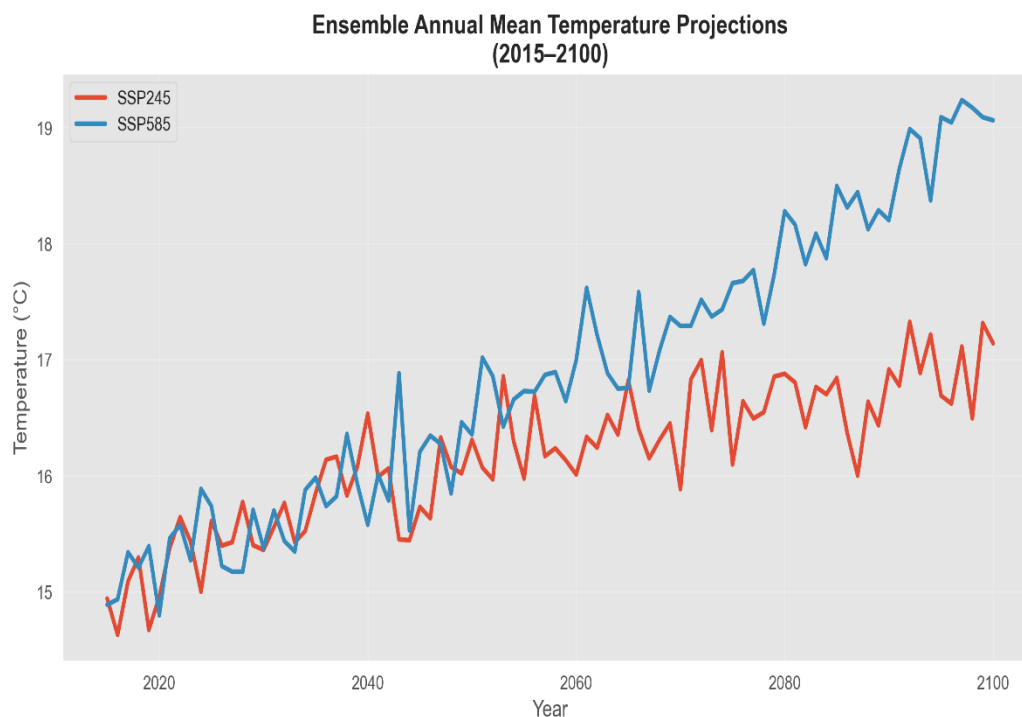


Figure 16: Ensemble annual mean temperature projections under SSP2-4.5 and SSP5-8.5 (2015–2100).

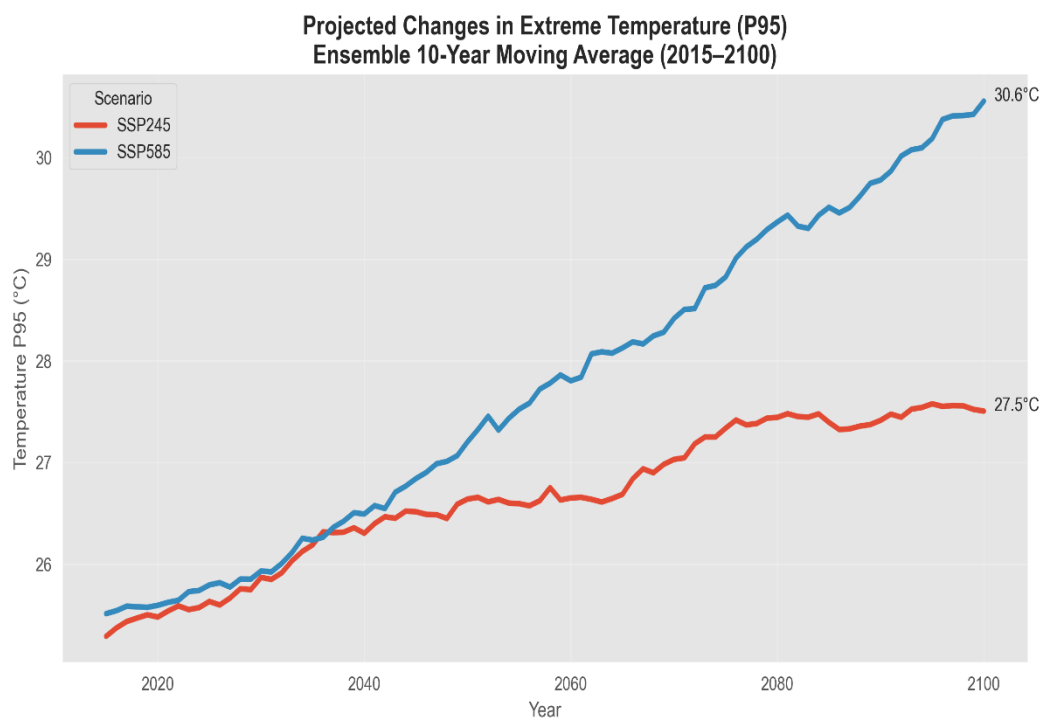


Figure 17: Projected changes in ensemble extreme temperature (P95) based on 10-year moving average under SSP2-4.5 and SSP5-8.5 (2015–2100).

Figures 16 and 17 are consistent with the projected warming in Table 34. The two scenarios are close in the near term, but diverge beginning around 2040 (and more sharply thereafter). Long-term annual averages increase to near 17°C in SSP2-4.5 by the end of the century and continue warming to 19°C in SSP5-8.5 by 2100. The upper panel of Figure 17 shows an example of these divergent climate futures for heat extremes, with P95 extreme temperatures in 2100 of 30.6°C in SSP5-8.5 and 27.5°C in SSP2-4.5, indicating more severe and frequent future heat extremes in SSP5-8.5. These results are in line with CMIP6 models, which predict considerable warming and increased temperature extremes in Turkey under high-emission climate scenarios (Bağçacı et al., 2021; Gumus et al., 2023).

5.8.2. Precipitation Projections

Table 35 presents the ensemble mean precipitation projections for the three future periods under SSP2-4.5 and SSP5-8.5.

Table 35: Ensemble mean and extreme precipitation projections across three future periods under SSP2-4.5 and SSP5-8.5 (2021–2100)

Scenario	Period	Precipitation		
		Mean	P95	P99
SSP245	2021-2040	2.021	10.736	22.336
SSP245	2041-2070	1.978	10.512	23.130
SSP245	2071-2100	1.979	10.456	23.221
SSP585	2021-2040	2.041	10.662	23.030
SSP585	2041-2070	1.913	10.136	21.882
SSP585	2071-2100	1.729	9.524	22.702

For predicted precipitation, both scenarios show declining precipitation, greater in the SSP5-8.5 scenario. The ensemble mean daily precipitation under SSP2-4.5 is predicted to be 2.021 mm/day in the near-term and 1.978 mm/day in mid-century. For SSP5-8.5, mean precipitation dropped from 2.041 mm/day to 1.729 mm/day between the near-term and 2071-2100. The extreme precipitation indices did not show a consistent trend. While P95 decreased slowly and P99 remained roughly constant, the most intense precipitation events remained relatively unchanged despite a decrease in the mean. This drying under high-emission scenarios is consistent with model projections for the Mediterranean, where warming-induced evaporation and circulation changes decrease mean precipitation while maintaining extreme events.

Figure 18 presents the projected changes in ensemble mean daily precipitation based on a 10-year moving average under both emission scenarios over the period 2015–2100.

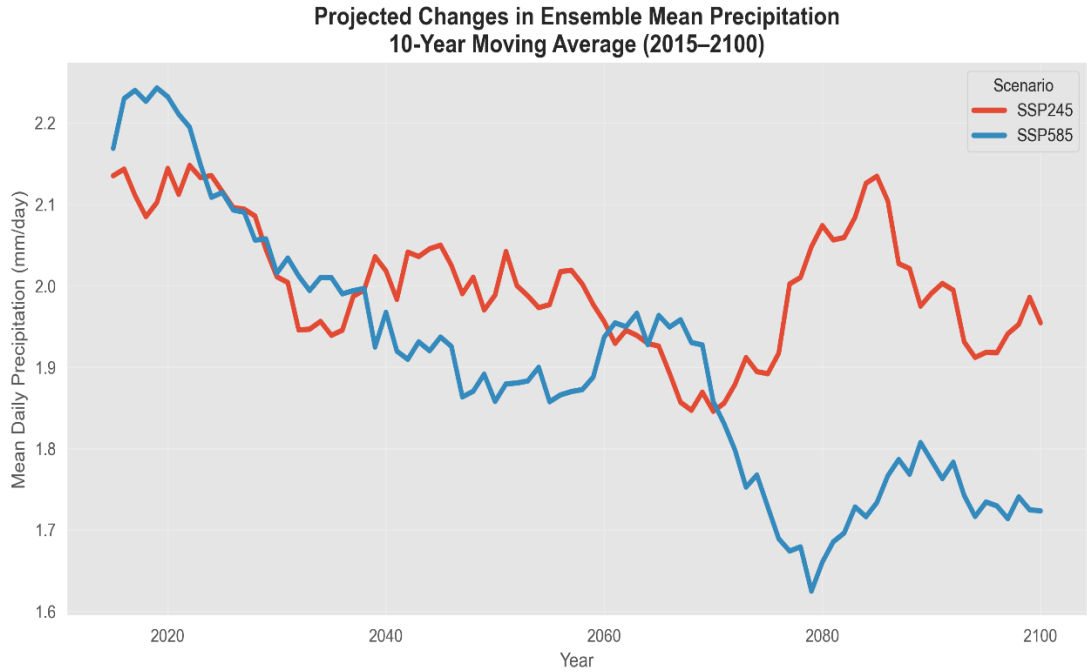


Figure 18: Projected changes in ensemble mean daily precipitation (10-year moving average) under SSP2-4.5 and SSP5-8.5 (2015–2100).

Overall, the models indicate that the average precipitation (Figure 18) decreases across the 21st century under both SSP scenarios, though with different rates of change. Under SSP2-4.5, mean daily precipitation remains relatively constant (at 2.0 mm/day) with small fluctuations over the course of the century. Under SSP5-8.5, the drying trend is stronger, with average rainfall decreasing from 2.2 mm/day at the start of the projection period to 1.7 mm/day towards the end of the century, with the maximum decrease occurring in the late 2060s to early 2080s, which is followed by a slight recovery in rainfall over the rest of the century. The difference between the two scenarios becomes more pronounced after mid-century, suggesting that the drying tendency projected for the study region may be considerably higher under stronger greenhouse gas forcing. This is consistent with other studies over Türkiye and the Mediterranean Basin that estimate larger reductions in precipitation under the high-emission scenario and identify the region as being particularly vulnerable to future dry conditions (Bağçacı et al., 2021; Gümüş et al., 2023).

CONCLUSION AND RECOMMENDATIONS

This study evaluated linear scaling, quantile mapping and artificial neural network bias correction techniques which were applied to the daily temperature and precipitation outputs of six CMIP6 climate models for the Florya station in Istanbul using the ERA5 reanalysis data set as an observational reference for the historical period of 1985-2014. The results of the methods were compared with the NASA's NEX-GDDP-CMIP6 climate products, which served as a benchmark. The future climate of the region was investigated under the SSP2-4.5 and SSP5-8.5 scenarios between 2015 and 2100.

The baseline evaluation found that all six raw CMIP6 models had systematic biases in temperature and precipitation compared to ERA5 data, with the largest temperature bias seen in ACCESS-CM2. The correlation coefficients for precipitation between the models and ERA5 were near zero for all models, showing that the day-to-day variability of precipitation is difficult to represent with GCMs at the point scale.

For temperature, ANN bias correction tended to perform the best, producing the lowest corrected RMSE and highest correlation coefficients across the ensemble. QM showed intermediate performance, correcting for distributional errors beyond the mean. LS had the smallest improvements, correcting only the mean error. However, ANN also had a systematic negative temperature bias in some models, overestimating cold temperatures and underestimating hot temperatures, consistent with regression to the mean.

For precipitation, no method substantially improved the temporal correlation, meaning that the timing of precipitation events cannot be reliably corrected through bias correction alone in this area. Though ANN had the lowest RMSE, the apparent improvement was largely due to regression towards the mean rather than genuine distributional correction. The corrected precipitation tended towards climatological average values and could not reproduce precipitation peaks. Wet-day quantile mapping produced the best trade-off between mean bias and distribution because it remained closest to the ERA5 wet-day distribution across the quantile range. QM and LS generally produced smaller and inconsistent precipitation improvements.

The results show that ANN correction reproduced temperature statistics similarly to the global NASA product, while the locally derived QM correction reproduced the wet-day precipitation distribution at the Florya site better than the distribution obtained when the global NASA product was applied. This indicates that locally calibrated corrections could outperform the use of globally bias-corrected products for site-specific applications.

The monthly results showed that the correction methods maintained the monthly seasonal temperature variability, with correlation coefficients greater than 0.92. The correlation coefficients for monthly precipitation were still low, similar to the daily precipitation results, showing that the correction methods did not capture seasonal precipitation variability well.

Under both scenarios, a warmer future climate is projected throughout the 21st century, with SSP2-4.5 and SSP5-8.5 remaining closely grouped during the first half of the century before beginning to diverge after 2040. Mean annual temperature under the SSP2-4.5 scenario is projected to stabilize near 17°C by the end of the century, while under the SSP5-8.5 scenario it reaches 19°C by 2100 and the P95 extreme temperature reaches 30.6°C. In both scenarios, precipitation becomes drier. The SSP5-8.5 scenario would reduce mean daily precipitation from 2.041 mm/day in the near-term period to 1.729 mm/day in the 2071-2100 period. Extreme precipitation indices showed no consistent trend, suggesting that the occurrence of intense precipitation events may continue despite the overall drying tendency.

Several limitations should be acknowledged. As this is a single grid-point study, the results may not be applicable to other regions and climate regimes. The study does not account for weather variables other than temperature and precipitation, and the ERA5 reanalysis is used as a reference instead of direct station observations. Due to the relatively short historical period used in training ANN, it may not be able to extrapolate effectively into future climate conditions.

Future work could include further development of deep learning architectures such as convolutional or recurrent networks with the objective of better reproducing the temporal structure of precipitation. In addition, including additional predictors beyond the single-point climate signal could provide further improvements to precipitation correction skill. A study extending to other locations in Istanbul and the

Marmara region would help better assess the performance of bias correction methods under a Mediterranean climate.

REFERENCES

- Azman, A. H., Tukimat, N. N. A., & Malek, M. A. (2022). Analysis of Linear Scaling Method in Downscaling Precipitation and Temperature. *Water Resources Management*, 36(1), 171–179. <https://doi.org/10.1007/s11269-021-03020-0>
- Babaousmail, H., Hou, R., Ayugi, B., Ojara, M., Ngoma, H., Karim, R., Rajasekar, A., & Ongoma, V. (2021). Evaluation of the Performance of CMIP6 Models in Reproducing Rainfall Patterns over North Africa. *Atmosphere*, 12(4), 475. <https://doi.org/10.3390/atmos12040475>
- Bağçaci, S. Ç., Yucel, I., Duzenli, E., & Yilmaz, M. T. (2021). Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey. *Atmospheric Research*, 256, 105576. <https://doi.org/10.1016/j.atmosres.2021.105576>
- Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4), 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>
- Berg, P., Feldmann, H., & Panitz, H.-J. (2012). Bias correction of high resolution regional climate model data. *Journal of Hydrology*, 448–449, 80–92. <https://doi.org/10.1016/j.jhydrol.2012.04.026>
- Bürger, G., Murdock, T., Werner, A., Sobie, S., & Cannon, A. (2012). Downscaling Extremes—An Intercomparison of Multiple Statistical Methods for Present Climate. *Journal of Climate*, 25, 4366–4388. <https://doi.org/10.1175/JCLI-D-11-00408.1>
- IPCC. (2023). Climate Change 2023: *Synthesis Report*. Intergovernmental Panel on Climate Change. <https://doi.org/10.59327/IPCC/AR6-9789291691647>
- Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, 28(17), 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Christensen, J. H., Boberg, F., Christensen, O. B., & Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, 35(20), 2008GL035694. <https://doi.org/10.1029/2008GL035694>
- Copernicus Climate Change Service. (n.d.). Coupled Model Intercomparison Project Phase 6 (CMIP6) data. European Centre for Medium-Range Weather Forecasts. <https://cds.climate.copernicus.eu>
- Climate Engine. (n.d.). *Climate Engine App*. <https://app.climateengine.com>

- Dhawan, P., Dalla Torre, D., Niazkar, M., Kaffas, K., Larcher, M., Righetti, M., & Menapace, A. (2024). A comprehensive comparison of bias correction methods in climate model simulations: Application on ERA5-Land across different temporal resolutions. *Heliyon*, 10(23), e40352. <https://doi.org/10.1016/j.heliyon.2024.e40352>
- Dutta, D., & Bhattacharjya, R. K. (2022). A statistical bias correction technique for global climate model predicted near-surface temperature in India using the generalized regression neural network. *Journal of Water and Climate Change*, 13(2), 854–871. <https://doi.org/10.2166/wcc.2022.214>
- Enayati, M., Bozorg-Haddad, O., Bazrafshan, J., Hejabi, S., & Chu, X. (2020). Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. *Journal of Water and Climate Change*, 12(2), 401–419. <https://doi.org/10.2166/wcc.2020.261>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Giorgi, F., & Mearns, L. O. (1991). Approaches to the simulation of regional climate change: A review. *Reviews of Geophysics*, 29(2), 191–216. <https://doi.org/10.1029/90RG02636>
- Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., & Francés, F. (2023). Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmospheric Research*, 284, 106606. <https://doi.org/10.1016/j.atmosres.2023.106606>
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen Skaugen, T. (2012). *Technical Note: Downscaling RCM precipitation to the station scale using quantile mapping – a comparison of methods*. Hydrometeorology/Mathematical applications. <https://doi.org/10.5194/hessd-9-6185-2012>
- Gumus, B., Oruc, S., Yucel, I., & Yilmaz, M. T. (2023). Impacts of Climate Change on Extreme Climate Indices in Türkiye Driven by High-Resolution Downscaled CMIP6 Climate Models. *Sustainability*, 15(9). <https://doi.org/10.3390/su15097202>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Huang, B., Liu, Z., Liu, S., & Duan, Q. (2023). Investigating the Performance of CMIP6 Seasonal Precipitation Predictions and a Grid Based Model

- Heterogeneity Oriented Deep Learning Bias Correction Framework. *Journal of Geophysical Research: Atmospheres*, 128(23), e2023JD039046. <https://doi.org/10.1029/2023JD039046>
- Irwandi, H., Rosid, M. S., & Mart, T. (2023). Effects of Climate change on temperature and precipitation in the Lake Toba region, Indonesia, based on ERA5-land data with quantile mapping bias correction. *Scientific Reports*, 13(1), 2542. <https://doi.org/10.1038/s41598-023-29592-y>
- Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., Wen, J., Gao, J., & Wang, J. (2021). Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. *Journal of Hydrology*, 595, 125660. <https://doi.org/10.1016/j.jhydrol.2020.125660>
- Kara, F., Yucel, I., & Akyurek, Z. (2016). Climate change impacts on extreme precipitation of water supply area in Istanbul: Use of ensemble climate modelling and geo-statistical downscaling. *Hydrological Sciences Journal*, 61(14), 2481–2495. <https://doi.org/10.1080/02626667.2015.1133911>
- Kingma, D. P., & Ba, J. (2015). *Adam: A Method for Stochastic Optimization* (arXiv:1412.6980). arXiv. <https://doi.org/10.48550/arXiv.1412.6980>
- KLEMEŠ, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13–24. <https://doi.org/10.1080/02626668609491024>
- Lafon, T., Dadson, S., Buys, G., & Prudhomme, C. (2013). Bias correction of daily precipitation simulated by a Regional Climate Model: A comparison of methods. *International Journal of Climatology*, 33, 1367–1381. <https://doi.org/10.1002/joc.3518>
- Lavers, D. A., Simmons, A., Vamborg, F., & Rodwell, M. J. (2022). An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, 148(748), 3152–3165. <https://doi.org/10.1002/qj.4351>
- Lenderink, G., Buishand, A., & van Deursen, W. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, 11(3), 1145–1159. <https://doi.org/10.5194/hess-11-1145-2007>
- Maraun, D. (2013). Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue. *Journal of Climate*, 26(6), 2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1>
- Maraun, D. (2016). Bias Correcting Climate Change Simulations—A Critical Review. *Current Climate Change Reports*, 2(4), 211–220. <https://doi.org/10.1007/s40641-016-0050-x>

- Maraun, D., & Widmann, M. (2018). *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press. Cambridge Core. <https://doi.org/10.1017/9781107588783>
- Moghim, S., & Bras, R. L. (2017). Bias Correction of Climate Modeled Temperature and Precipitation Using Artificial Neural Networks. *Journal of Hydrometeorology*, 18(7), 1867–1884. <https://doi.org/10.1175/JHM-D-16-0247.1>
- NASA Center for Climate Simulation. (n.d.). *NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP-CMIP6)*. <https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp-cmip6>
- Ngai, S. T., Tangang, F., & Juneng, L. (2017). Bias correction of global and regional simulated daily precipitation and surface mean temperature over Southeast Asia using quantile mapping method. *Global and Planetary Change*, 149, 79–90. <https://doi.org/10.1016/j.gloplacha.2016.12.009>
- O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Oruc, S. (2022). Performance of bias corrected monthly CMIP6 climate projections with different reference period data in Turkey. *Acta Geophysica*, 70(2), 777–789. <https://doi.org/10.1007/s11600-022-00731-9>
- Panigrahi, N. (2023). Inverse Distance Weight. In B. S. Daya Sagar, Q. Cheng, J. McKinley, & F. Agterberg (Eds.), *Encyclopedia of Mathematical Geosciences*. Springer. https://doi.org/10.1007/978-3-030-85040-1_166
- 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(3–4), 199–215. <https://doi.org/10.1016/j.jhydrol.2010.10.024>
- Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolino, C., Noll, B., & Meyers, T. (2022). High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand. *Weather and Climate Extremes*, 38, 100525. <https://doi.org/10.1016/j.wace.2022.100525>
- Schmidli, J., Frei, C., & Vidale, P. L. (2006). Downscaling from GCM Precipitation: A Benchmark for Dynamical and Statistical Downscaling Methods. *International Journal of Climatology*, 26, 679–689. <https://doi.org/10.1002/joc.1287>

- Seker, M., & Gumus, V. (2022). Projection of temperature and precipitation in the Mediterranean region through multi-model ensemble from CMIP6. *Atmospheric Research*, 280, 106440. <https://doi.org/10.1016/j.atmosres.2022.106440>
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192. <https://doi.org/10.1029/2000JD900719>
- Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., Moseley, C., Pfeifer, S., Rechid, D., Remedio, A., Ries, H., Petersen, J., Preuschmann, S., Raub, T., Saeed, F., Sieck, K., & Weber, T. (2013). How Does a Regional Climate Model Modify the Projected Climate Change Signal of the Driving GCM: A Study over Different CORDEX Regions Using REMO. *Atmosphere*, 4(2), 214–236. <https://doi.org/10.3390/atmos4020214>
- Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456–457, 12–29. <https://doi.org/10.1016/j.jhydrol.2012.05.052>
- Thiemeßl, M. J., Gobiet, A., & Heinrich, G. (2012). EMPI-ESM1-2-LRrical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112(2), 449–468. <https://doi.org/10.1007/s10584-011-0224-4>
- Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T., & Nemani, R. (2022). NASA Global Daily Downscaled Projections, CMIP6. *Scientific Data*, 9(1), 262. <https://doi.org/10.1038/s41597-022-01393-4>
- Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., & Giorgi, F. (2021). Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Climate Dynamics*, 57(5), 1425–1443. <https://doi.org/10.1007/s00382-020-05447-4>
- Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. (2017). *DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution*. <https://doi.org/10.48550/arXiv.1703.03126>
- Vrac, M., Noël, T., & Vautard, R. (2016). Bias correction of precipitation through Singularity Stochastic Removal: Because occurrences matter. *Journal of Geophysical Research: Atmospheres*, 121(10), 5237–5258. <https://doi.org/10.1002/2015JD024511>
- Wang, F., & Tian, D. (2022). On deep learning-based bias correction and downscaling of multiple climate models simulations. *Climate Dynamics*, 59(11), 3451–3468. <https://doi.org/10.1007/s00382-022-06277-2>
- Weichert, A., & Bürger, G. (1998). Linear versus nonlinear techniques in downscaling. *Climate Research*, 10, 83–93. <https://doi.org/10.3354/cr010083>

- White, R. H., & Toumi, R. (2013). The limitations of bias correcting regional climate model inputs. *Geophysical Research Letters*, 40(12), 2907–2912. <https://doi.org/10.1002/grl.50612>
- Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography: Earth and Environment*, 21(4), 530–548. <https://doi.org/10.1177/030913339702100403>
- Xu, Z., Han, Y., & Yang, Z. (2019). Dynamical downscaling of regional climate: A review of methods and limitations. *Science China Earth Sciences*, 62(2), 365–375. <https://doi.org/10.1007/s11430-018-9261-5>
- Yang, X., Yang, S., Tan, M. L., Pan, H., Zhang, H., Wang, G., He, R., & Wang, Z. (2022). Correcting the bias of daily satellite precipitation estimates in tropical regions using deep neural network. *Journal of Hydrology*, 608, 127656. <https://doi.org/10.1016/j.jhydrol.2022.127656>
- Yılmaz, B., Aras, E., & Nacar, S. (2024). A CMIP6-ensemble-based evaluation of precipitation and temperature projections. *Theoretical and Applied Climatology*, 155(8), 7377–7401. <https://doi.org/10.1007/s00704-024-05066-7>
- Zhang, H., Chapman, S., Trancoso, R., Toombs, N., & Syktus, J. (2024). Assessing the impact of bias correction approaches on climate extremes and the climate change signal. *Meteorological Applications*, 31(3), e2204. <https://doi.org/10.1002/met.2204>

RESUME

Personal Information

Surname, name : Nouamane Mtarfi
Nationality : Morocco
Birth date and place : Morocco, 17/08/1998
Telephone : 05318727635
Fax :
e-mail : nouamanemt1@gmail.com

Education

Degree	Education Unit	Graduation Date
Master	Istanbul Gelişim University	2026
Bachelor	FMV Isik University	2024
High School	Abdelkarim Al Khattabi	2017

Work Experience

Year	Place	Title
------	-------	-------

Foreign Languages

English, French, Arabic

Publications

Hobbies

Reading, Football, Volleyball, Data Analysis, Programming