Theoretical and Applied Climatology (2026) 157:121
https://doi.org/10.1007/s00704-026-06066-5

RESEARCH

Evaluating the performance of selected CMIP6 GCMs for simulations of historical temperature over Ethiopia

Alehegn Belay Tagele[1] · Temesgen Gashaw Tarkegn[1,2,3] · Pieter R. van Oel[3] · Ram Lakhan Ray[2] · Gebrekidan Worku Tefera[2] · Sintayehu Fetene Demessie[4,12] · Alemayehu Abate Shawul[5] · Othoo Calvince Ouko[6] · Daniel Berhanu[7] · Amare Haileslassie[8] · Abeyou W. Worqlul[9] · Amare Bantider[7] · Yihun T. Dile[10] · Abebe D. Chukalla[11] · Tazebew Abera Tadesse[1] · Enyew Adgo[1]

Received: 7 August 2025 / Accepted: 21 January 2026 / Published online: 5 February 2026
© The Author(s) 2026

Abstract
This study evaluates the performance of seven Global Climate Models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) and their Ensemble mean in reproducing historical maximum (Tmax) and minimum (Tmin) temperatures across five agro-ecological zones (AEZs) of Ethiopia during 1995–2014. The assessment covers daily to annual time scales using observational and GCM datasets. Model performance was evaluated using Percent Bias (PBIAS), Root Mean Square Error (RMSE), and correlation coefficient (r), while the Comprehensive Rating Index (CRI) was applied for ranking. Results show substantial spatial and temporal variability in model performance. For Tmax, the Ensemble mean and EC-Earth3-veg performed best at the daily scale, while EC-Earth3-veg, MPI-ESM1-2-LR and BCC-CSM2-MR excelled at the monthly scale across most AEZs. Seasonally, top-performing models included Ensemble mean, MPI-ESM1-2-LR and MRI-ESM2-0 during *Bega* (October-January), EC-Earth3-veg, CNRM-CM6-1 and Ensemble mean during *Belg* (February-May) and CNRM-CM6-1 and MPI-ESM1-2-LR during *Kiremt* (June-September). For Tmin, CNRM-CM6-1, BCC-CSM2-MR and Ensemble mean ranked highest at both daily and monthly scales in most AEZs. At the annual scale, MRI-ESM2-0, Ensemble mean, MPI-ESM1-2-LR and CNRM-CM6-1 excel for Tmax, while EC-Earth3-veg, BCC-CSM2-MR and Ensemble mean lead for Tmin across most AEZs. MIROC6 consistently exhibited the weakest performance for both Tmax and Tmin across most AEZs and time periods. Thus, the Ensemble mean of all evaluated models does not consistently rank among the top three performers across all AEZs and time scales. The identified best-performing CMIP6 models provide valuable tools for assessing climate change impacts and developing region-and time-specific adaptation strategies in Ethiopia. Given the variability in GCM performance across AEZs and time scales, national or large-scale studies would benefit from using the Ensemble mean derived from the best-performing models. This study provides crucial insights into the strengths and weaknesses of different GCMs, supporting evidence-based decision-making and enhancing efforts to build climate resilience and adaptation strategies in the region.

Keywords Agro-ecological zones · Climate models · CMIP6 · GCMs · Ethiopia · Temperature

1 Introduction

Global Climate Models (GCMs) are advanced mathematical representations of the climate system, which includes the atmosphere, ocean, land surface, and cryosphere (Eyring et al. 2016; Randall et al. 2007). These models are invaluable tools for studying the global climate system across historical and future timelines (Gebrechorkos et al. 2023a; Rettie et al. 2023). Since 1995, researchers from various global institutions have been refining GCMs to enhance their efficiency and reliability in simulating the climate system (Edwards 2011). Outputs from these models are aggregated by the World Climate Research Program (WCRP) Working Group on Coupled Modeling (WGCM) to facilitate climate model experiments within the framework of the Coupled Model Intercomparison Project (CMIP) (Eyring et al. 2016). The

Extended author information available on the last page of the article latest iteration, the CMIP Phase 6 (CMIP6) GCMs, represents a significant advancement over its predecessors in many parts of the globe (Cruz-González et al. 2025; Gulakhmadov et al. 2025; Tanimu et al. 2025; Wang et al. 2025; Bağçaci et al. 2021; Kamruzzaman et al. 2021). Thus, analysis of CMIP6 data enables researchers to gain deeper insights into climate variability, change, and its impacts on society and the environment (Kawai et al. 2019).

Nevertheless, most climate model evaluation studies conducted in Ethiopia have primarily focused on climate models predating CMIP6 (Demessie et al. 2023; Worku et al. 2018). In addition, the majority of previous studies in Ethiopia have concentrated on evaluating the performance of climate models in simulating rainfall (Ersado and Awoke 2024; Demessie et al. 2023; Worku et al. 2018). This focus on rainfall may stem from the common assumption that climate models can more accurately simulate temperature, whereas capturing rainfall patterns is considerably more challenging (Peng et al. 2023). Nevertheless, recent research conducted in Ethiopia (Gashaw et al. 2024; Lebeza et al. 2024; Demessie et al. 2023; Alaminie et al. 2021) and other regions (Belazreg et al. 2022) suggests that models capable of accurately simulating observed rainfall may not necessarily perform well in simulating temperature, highlighting the complexity and variability of climate model performance across different variables.

Previous evaluations of climate models in Ethiopia have also primarily been conducted using areal averages comparisons (Gashaw et al. 2024), with most studies focusing on sub-basin or basin scales (Alaminie et al. 2021; Worku et al. 2018). Consequently, there is a noticeable gap in comprehensive national studies that assess the performance of CMIP6 models in simulating temperature across the country's diverse agro-ecological zones (AEZs). In contrast, a significantly larger body of research exists globally on the performance evaluation of climate models for temperature simulation (Belazreg et al. 2022; Hamed et al. 2022; Lovino et al. 2021; Yang et al. 2021; Jia et al. 2019). In addition, most of the GCMs evaluation studies in different watersheds of the country for simulating Tmax and Tmin did not evaluate the Ensemble mean of their chosen GCMs. This disparity highlights the urgent need for studies in Ethiopia to identify the most effective climate models for various AEZs. Such efforts are critical for informing and enhancing climate adaptation strategies tailored to the country's unique ecological and climatic conditions.

This study evaluates the ability of seven GCMs from the CMIP6 and the Ensemble mean of those GCMs in simulating historical maximum temperature (Tmax) and minimum temperature (Tmin) across daily to annual temporal scales for all five AEZs of Ethiopia. The study leverages the advancements of CMIP6 and addresses a critical gap, as there has been no comprehensive national-level evaluation of climate models for Tmax and Tmin using CMIP6 data. Identifying the best-performing GCMs is essential for enhancing our understanding of climate change and its future characteristics (Cruz-González et al. 2025; Meliho et al. 2025; Shah and Sharifi 2025; Song et al. 2025; Gebrechorkos et al. 2023a). It also provides a foundation for assessing climate change impacts on water resources (Waheed et al. 2024; Getachew and Manjunatha 2022; Worku et al. 2021), and drought propagation (Tadase and Tekile 2025). Given Ethiopia's strong dependence on agriculture, identifying the best-performing climate models is vital for evaluating climate change impacts on crop production (Chen et al. 2019), planning adaptation measures based on future climate projections, determining future growing season lengths, and selecting suitable crop types for specific AEZs. Overall, this study provides a crucial foundation for climate-resilient modeling aimed at achieving water and food security (Zhang et al. 2023) by utilizing the most reliable GCMs tailored to each AEZ.

The novelty of this study lies in the following key aspects: (1) Evaluation of seven CMIP6 GCMs and their Ensemble mean across a full range of temporal scales, from daily to annual, 2) Performance assessments of CMIP6 GCMs across distinct AEZs, 3) Independent evaluations of GCMs for Tmax and Tmin, and 4) Application of the Comprehensive Rating Index (CRI) to rank GCM performance. The best-performing GCMs for Tmax and Tmin in each AEZ can be utilized to understand future temperature patterns under different climate change scenarios and to evaluate the impacts of climate change on various environmental components. Additionally, the outputs of these models can serve as critical inputs for climate-resilient modeling efforts in Ethiopia, supporting sustainable adaptation and mitigation strategies.

2 Materials and methods

2.1 Description of the study area

This research was conducted within the geographical confines of Ethiopia situated between 3°00′ N to 15°00′ N latitudes, and 33°00′ E to 48°00′ E longitudes in the eastern segment of Africa (Fig. 1). The nation's climate is predominantly tropical, characterized by significant spatial and temporal variability in rainfall patterns. The Ethiopian highlands, areas located above 1500 m above sea level (m a.s.l), experience relatively stable temperatures, whereas the lowlands (< 1500 m a.s.l), exhibit greater temperature fluctuations. A diverse landscape that includes mountains, expansive plateaus, profound gorges, extensive river

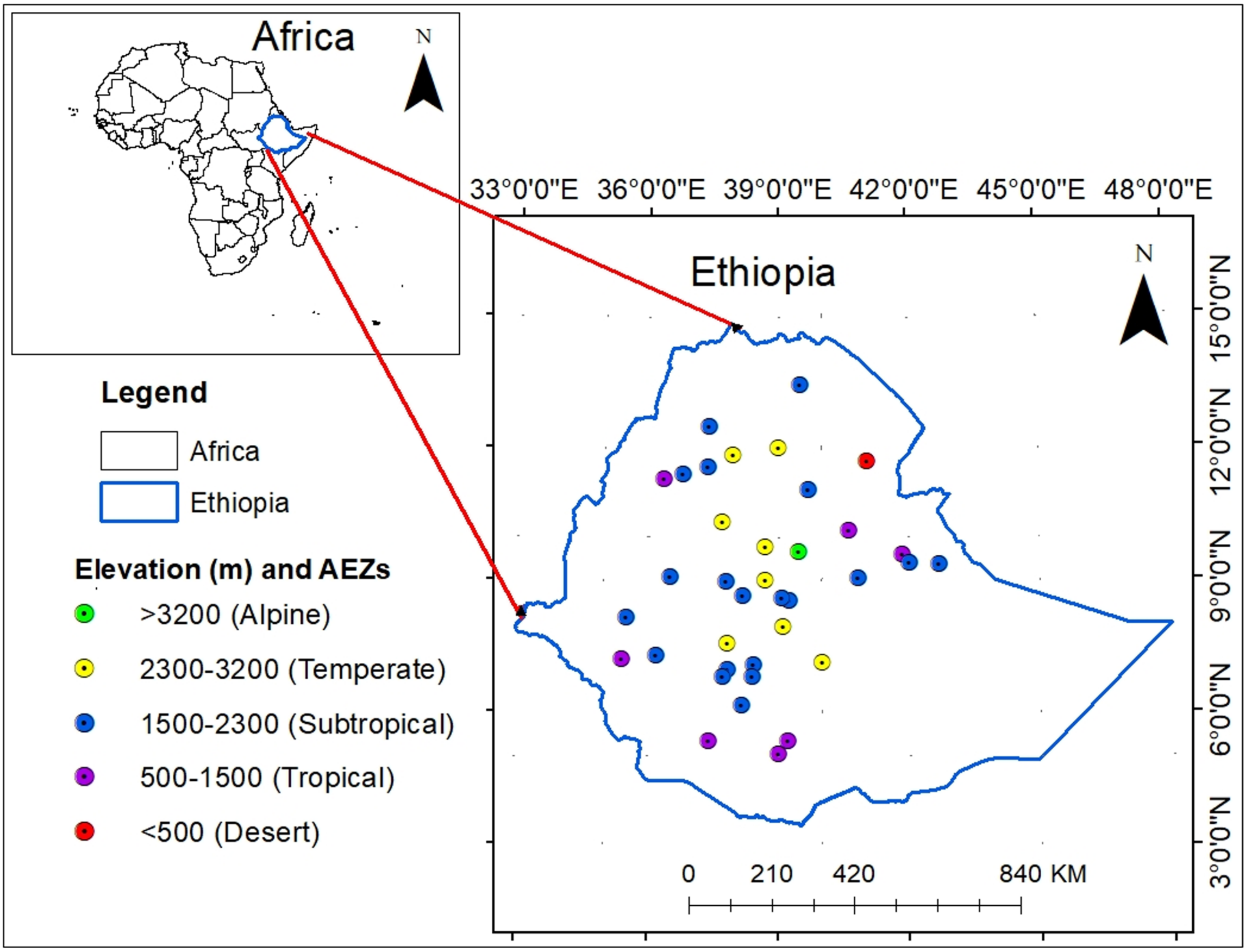


Fig. 1 The location of Ethiopia and its different AEZs

valleys, and plains distinguishing it from other African nations, marks Ethiopia's topography. It shares borders with Somalia and Djibouti to the east, Kenya to the south, Eritrea to the north and northeast, and Sudan to the west, with South Sudan lying to the west and southwest.

Ethiopia experiences three distinct climatic seasons: the primary rainy season from June to September (JJAS), the dry season from October to January (ONDJ), and the secondary rainy season from February/March to May (FMAM), locally known as *Kiremt*, *Bega*, and *Belg*, respectively (Ariley et al. 2023; Gashaw et al. 2023; MoA 1998). The climate varies from semi-arid deserts to humid and temperate zones (Gashaw et al. 2023). The long-term mean annual rainfall (1981–2016) of the country is 620 mm (Gebrechorkos et al. 2018). Mean annual temperatures range from 15 °C in the highlands to 25 °C in the lowlands (NMSA, 2001). Ethiopia's rich agroecological diversity supports a wide array of crops and livestock. The country is categorized into five AEZs based on altitude and climate: desert (*Bereha*), tropical (*Kolla*), sub-tropical (*Weyna Dega*), temperate (*Dega*), and alpine (*Wurch*) AEZs (MoA 1998) (Fig. 1).

2.2 Data source

2.2.1 Observational data

Temperature data spanning two decades (1995–2014) was sourced from the Ethiopian Meteorology Institute (EMI). This dataset encompasses daily recorded temperatures from 37 meteorological stations across Ethiopia and contains less than 25% missing values. Like our study, due to the sparse distributions of meteorological studies, preceding studies in Ethiopia also used stations that contain missing values up to 30% (Gashaw et al. 2023, 2024; Mohammed et al. 2022). The distributions of stations were well-balanced throughout Ethiopia, with a denser presence in the central, southern, northern, and eastern regions, and fewer in the southeastern and western regions (Fig. 1).

To fill the gaps in the observed climate data, the Multivariate Imputation by Chained Equations (MICE) algorithm was employed (Buuren et al. 2022), utilizing the R statistical software. The MICE algorithm outperforms other methods like inverse distance weighting (IDW) and multiple linear regressions (MLR) by leveraging the complete observed values from all stations as predictors for imputing missing values at a single station (Worku et al. 2018). It generates multiple imputations for each missing value, incorporates uncertainties, and calculates standard errors (Buuren et al. 2022). Outliers in daily Tmax and Tmin were identified through the quality control procedure using RClimDex 1.1 (Zhang and Yang 2004). Detailed information regarding the selected meteorological stations is available in Supplementary Table 1.

2.2.2 GCM's data

This study evaluates the performance of seven GCMs from CMIP6 and the Ensemble mean of those GCMs. Table 1 provides detailed information about these GCMs, whose historical simulation outputs from 1995 to 2014 were used for this analysis. The corresponding data are archived and can be accessed using the following criteria: the Source ID, which corresponds to the GCM names listed in Table 1; the variables Tmax and Tmin; Frequency set to daily; and Experimental ID specified as amip-historical. Additionally, each model's variant label, which uniquely identifies ensemble members based on realizations, is critical for locating and distinguishing the data (Papalexiou et al. 2020).

The selection of GCMs for this study was based on the proven performance of their CMIP6 phase in simulating key climate variables (temperature, precipitation, and large-scale circulation) in different watersheds or basins of Ethiopia (Gashaw et al. 2024; Lebeza et al. 2024; Alaminie et al. 2021), the broader East African region (Gebrechorkos et al. 2023b), and other global locations (Zabihi and Ahmadi 2024; Shiru and Chung 2021; Karim et al. 2020). For instance, EC-Earth3-veg, MPI-ESM1-2-LR, and CNRM-CM6-1 were selected due to their superior ability to simulate temperature in the Bale Eco-Region (Gashaw et al. 2024). Similarly, MPI-ESM1-2-LR, BCC-CSM-2-MR, and MRI-ESM2-0 demonstrated strong performance in the Lake Tana sub-basin (Lebeza et al. 2024), while MRI-ESM2-0 excelled in the Upper Blue Nile Basin (Alaminie et al. 2021). In addition, the selected models have also been widely applied in previous studies within Ethiopia under CMIP5 (Demessie et al. 2023; Worku et al. 2018). Furthermore, these models were chosen due to their use in generating high-resolution climate data for Ethiopia (Gebrechorkos et al. 2023b; Rettie et al. 2023), ensuring their relevance and applicability to this study's objectives.

Although the performance of the selected climate models has been evaluated in previous studies (Woldemariam et al. 2025; Gashaw et al. 2024; Lebeza et al. 2024; Alaminie et al. 2021), this research is essential because earlier assessments were primarily conducted at watershed or basin scales within Ethiopia, and thus did not provide a comprehensive understanding of GCM performance at the national level. Furthermore, studies conducted in other global regions such as in Algeria (Belazreg et al. 2022), China (Wang et al. 2025), and Thailand and nearby areas (Kamworapan et al. 2021) may not accurately represent climatic conditions in Ethiopia. In addition, previous evaluations were not performed at the AEZ scale (Gashaw et al. 2024; Lebeza et al. 2024), focusing instead on areal averages of the study domains. Given that the performance of climate models can vary across regions, evaluating these models at different AEZs is therefore essential to capture spatial variability in model performance.

2.3 Evaluation methods

The historical outputs of seven CMIP6 GCMs and their Ensemble mean were assessed for their ability to simulate Tmax and Tmin across daily to annual temporal scales over Ethiopia's five AEZs: desert, tropical, sub-tropical, temperate, and alpine AEZs. Observational data from 37 meteorological stations were used for this evaluation (Fig. 1). The GCM data for Tmax and Tmin were extracted for each

Table 1 Basic information of the studied CMIP6 GCMs

CMIP6 models	Institute	Country	Resolution (Lon. by lat.) in degrees	Variant label
ACCESS-ESM1-5	Australian Community Climate and Earth System Simulator	Australia	1.9°×1.2°	r1i1p1f1
BCC-CSM2-MR	Beijing Climate Center (BCC)	China	1.1°×1.1°	r1i1p1f1
CNRM-CM6-1	Recherches Météorologiques	France	1.4 × 1.4◦	r1i1p1f2
EC-Earth3-veg	EC-Earth-Consortium	Sweden	0.7° × 0.7°	r1i1p1f1
MIROC6	Japan Agency for Marine-Earth Science and Technology (JAMSTEC)	Japan	1.4°×1.4°	r1i1p1f1
MPI-ESM1-2-LR	Max-Planck-Institute für Meteorology	Germany	1.9°×1.9°	r1i1p1f1
MRI-ESM2-0	Meteorological Research Institute	Japan	1.1°×1.1°	r1i1p1f1

meteorological station using the CMhyd tool (Rathjens et al. 2016), which is specifically designed for extracting and bias-correcting GCM data. In this study, CMhyd was employed only to extract GCM data corresponding to each observed climate station. Of the 37 meteorological stations, 20, 7, and 8 are located in the sub-tropical, tropical, and temperate AEZs, respectively. The alpine and desert AEZs are each represented by a single meteorological station due to the unavailability of other stations with long-term data and minimal missing values in these regions. To represent each AEZ, the average values of the meteorological stations within the same AEZ were calculated. Therefore, a point-to-pixel approach (Gashaw et al. 2024; Lebeza et al. 2024) was applied to evaluate the performance of the GCMs, ensuring alignment between station-based observations and the GCM grid outputs.

This study utilized three statistical metrics to evaluate the performance of the GCMs: Percent Bias (PBIAS), Root Mean Square Error (RMSE), and correlation coefficient (r). The selection of these performance measures is supported by extensive literature (Cruz-González et al. 2025; Hoseini et al. 2025; Tadase and Tekile 2025; Abdulsahib et al. 2024; Hassan et al. 2023). For example, Tadase and Tekile (2025) employed PBIAS, RMSE, and r exclusively to assess the performance of GCMs. The computations of these metrics were performed using the "hydroGOF" package (Zambrano-Bigiarini et al. 2017) in R software, a widely utilized tool for performance assessments in many studies (Woldemariam et al. 2025; Gashaw et al. 2024; Lebeza et al. 2024).

PBIAS (Eq. 1) evaluates the tendency of a model to under- or overestimate observed data, with optimal performance indicated by values near zero. A negative PBIAS value suggests overestimation, while a positive value indicates underestimation. RMSE (Eq. 2) measures the magnitude of error between the GCM outputs and observed data, providing a summary of errors in terms of magnitude. Lower RMSE values denote a better simulation with reduced error (Gashaw et al. 2024). The correlation coefficient (r) (Eq. 3) quantifies the degree of relationship or similarity between the GCM and observed data. A correlation coefficient of −1 indicates a perfect negative correlation, while +1 signifies a perfect positive correlation between the model outputs and the observed temperature data. These metrics collectively provide a robust framework for assessing the performance of GCMs across spatial and temporal scales.

$$PBIAS = \frac{\sum (G - S)}{\sum (G)} \times 100 \tag{1}$$

$$RMSE = \sqrt{\frac{\sum (G - S)^2}{N}} \tag{2}$$

$$r = \frac{\sum (G - GG)\ (S - SS)}{\sqrt{\sum (G - GG)^2}\ \sqrt{\sum (S - SS)^2}} \tag{3}$$

Where G is the observed temperature, GG is the average observed temperature, S is the GCM's temperature estimate, SS *is* the average of GCM's temperature estimate, and N is the number of data pairs compared.

The GCMs were subsequently ranked based on their ability to simulate Tmax and Tmin for each of the studied AEZs and temporal scales. This ranking was conducted using the three previously mentioned metrics (PBIAS, RMSE, and r) and applying the Comprehensive Rating Index (CRI), following methodologies outlined in prior studies (Baig et al. 2025; Woldemariam et al. 2025; Gashaw et al. 2024; Lebeza et al. 2024; Rivera and Arnould 2020). CRI (Eq. 4) is a robust and widely adopted ranking method that provides a comprehensive framework for evaluating and comparing the performance of GCMs across multiple metrics and conditions.

$$CRI = 1 - \frac{1}{nm} \sum {}_{i=1}^{n} RANK_i \tag{4}$$

Where n is the number of indices used to evaluate model performance, m is the total number of GCMs. The rank value of the model with the best performance is assigned as 1. Therefore, the closer to 1 the value of CRI is the better the model performs. Figure 2 demonstrates the methodological flows used in this study.

3 Results and discussions

3.1 Performance of GCMs for simulating Tmax

The performance of the seven GCMs and the Ensemble mean of those GCMs for simulating Tmax from daily to annual temporal scales during the 1995–2014 period in the five AEZs of Ethiopia, namely: desert, tropical, sub-tropical, temperate and alpine AEZs are presented in Figs. 3 and 4 and Supplementary Tables 2–4. The results indicated that the performances of the studied GCMs vary with AEZs and temporal scales. The top-ranked GCMs for simulating the daily Tmax over desert AEZ are Ensemble mean, EC-Earth3-veg and MRI-ESM2-0 while Ensemble mean, MPI-ESM1-2-LR and EC-Earth3-veg ranked highest over the tropical AEZ (Fig. 3). Ensemble mean, ACCESS-ESM-1–5 and MPI-ESM1-2-LR in the sub-tropical AEZ, and EC-Earth3-veg, Ensemble mean and MRI-ESM2-0 over temperate AEZ are the best-performing GCMs for simulating daily Tmax. The best-performing models over the alpine

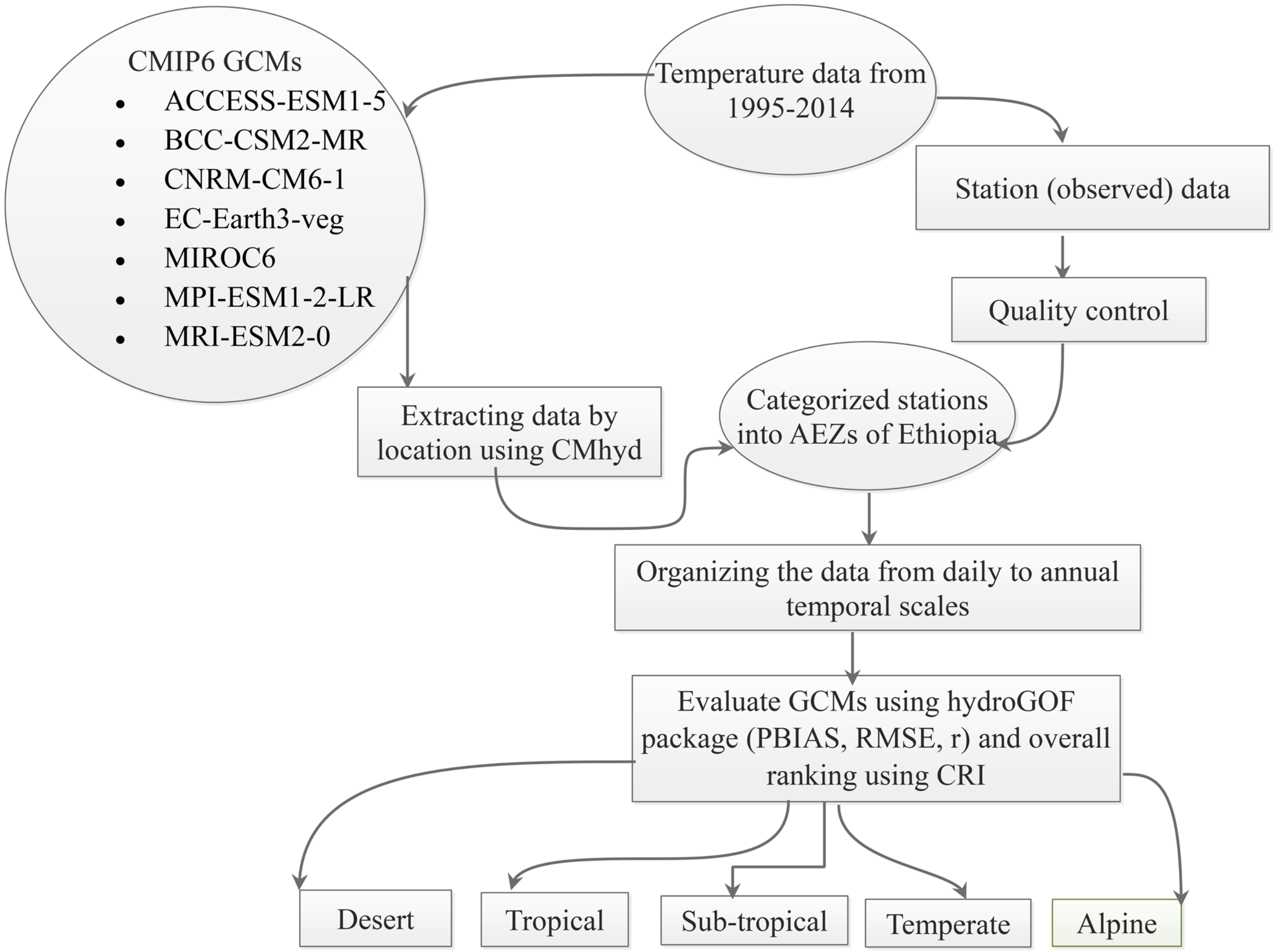


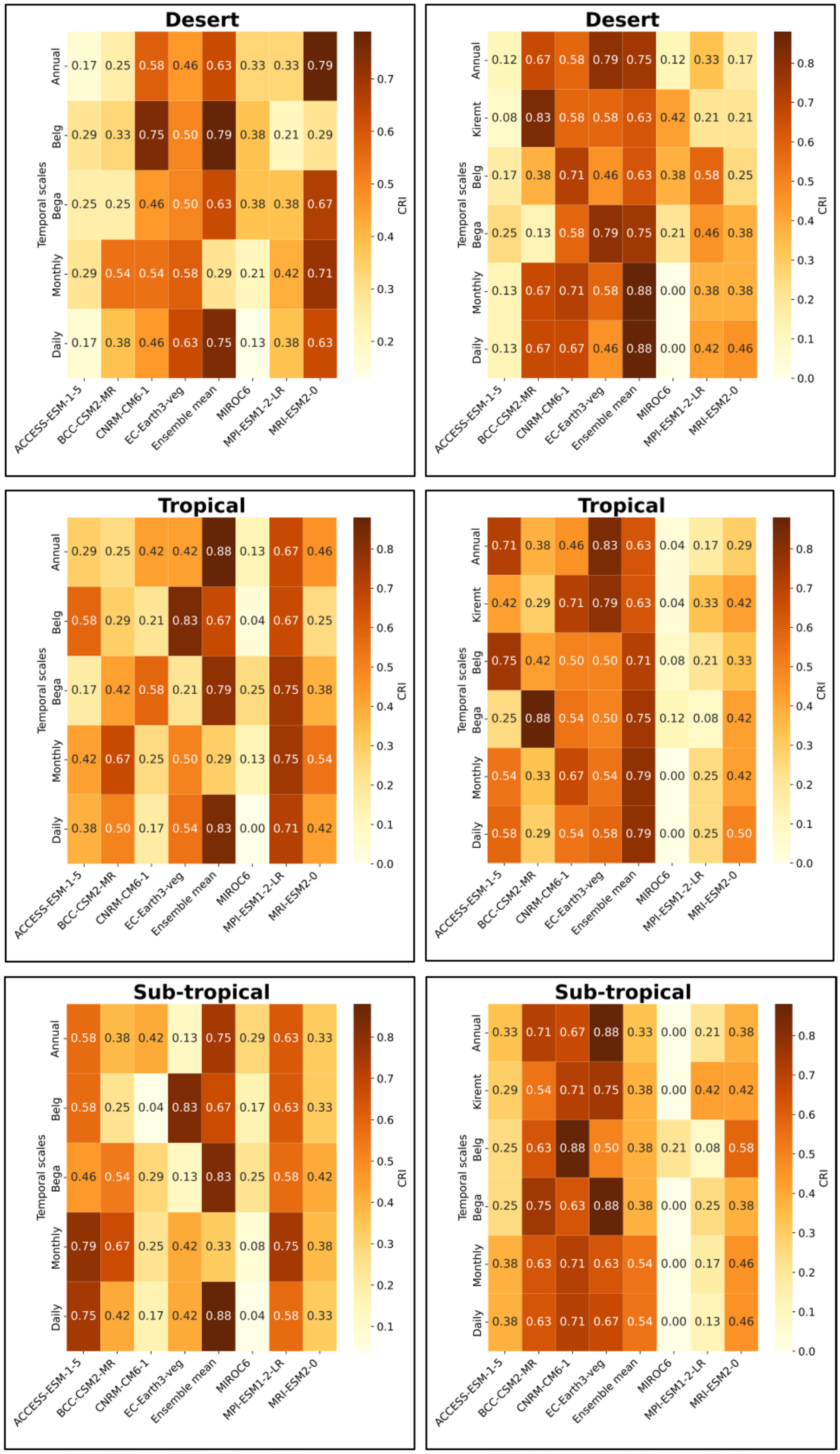
Fig. 2 The methodological flows implemented in our study

AEZ are EC-Earth3-veg, CNRM-CM6-1 and Ensemble mean. Thus, the Ensemble mean of the studied GCMs is among the best three datasets for estimating daily Tmax over the five AEZs of the country. EC-Earth3-veg over the four AEZs is among the three best GCMs for simulating daily Tmax. The poorest dataset for estimating daily Tmax across all AEZs is MIROC6.

Consistent with the findings of this study, EC-Earth3-veg has demonstrated strong performance in simulating daily Tmax over the Bale Eco-Region in southern Ethiopia (Gashaw et al. 2024). In our study, MPI-ESM1-2-LR is not among the three best-performing GCMs across most AEZs. However, this model has shown superior performance over the Lake Tana sub-basin (Lebeza et al. 2024) and the Bale Eco-Region (Gashaw et al. 2024). The apparent discrepancy between our findings and previous studies highlights that the performance of MPI-ESM1-2-LR is likely region-dependent, shaped by the interaction of local climatic conditions, topography, and atmospheric processes. On the other hand, MIROC6 consistently exhibited the poorest performance across all five AEZs in this study. This result is supported by previous findings, where MIROC6 was among the weakest GCMs for simulating daily Tmax in the Lake Tana sub-basin (Lebeza et al. 2024) and performed poorly across all scenarios (land-only, sea-only, and combined land-sea cases) in Thailand (Shiru and Chung 2021).

The best-performing models for simulating Tmax at the monthly temporal scale include MRI-ESM2-0, EC-Earth3-veg, BCC-CSM-2-MR and CNRM-CM6-1 over the desert AEZ, MPI-ESM1-2-LR, BCC-CSM-2-MR and MRI-ESM2-0 over tropical AEZ, and ACCESS-ESM-1–5, MPI-ESM1-2-LR and BCC-CSM-2-MR over the sub-tropical AEZ (Fig. 3). The GCMs that are defined as best performing for simulating monthly Tmax over the temperate AEZ include EC-Earth3-veg, CNRM-CM6-1, MPI-ESM1-2-LR and MRI-ESM2-0 while EC-Earth3-veg, Ensemble mean and CNRM-CM6-1are effective GCMs over alpine AEZ. Thus, EC-Earth3-veg, MPI-ESM1-2-LR and BCC-CSM-2-MR are among the three best-performing models in the three AEZs. Unlike the daily time scale, the Ensemble

Fig. 3 The overall ranking of the studied climate models at different temporal scales for simulating Tmax (left) and Tmin (right) across different climate regions of Ethiopia. Note: Darker colors represent the best ranking



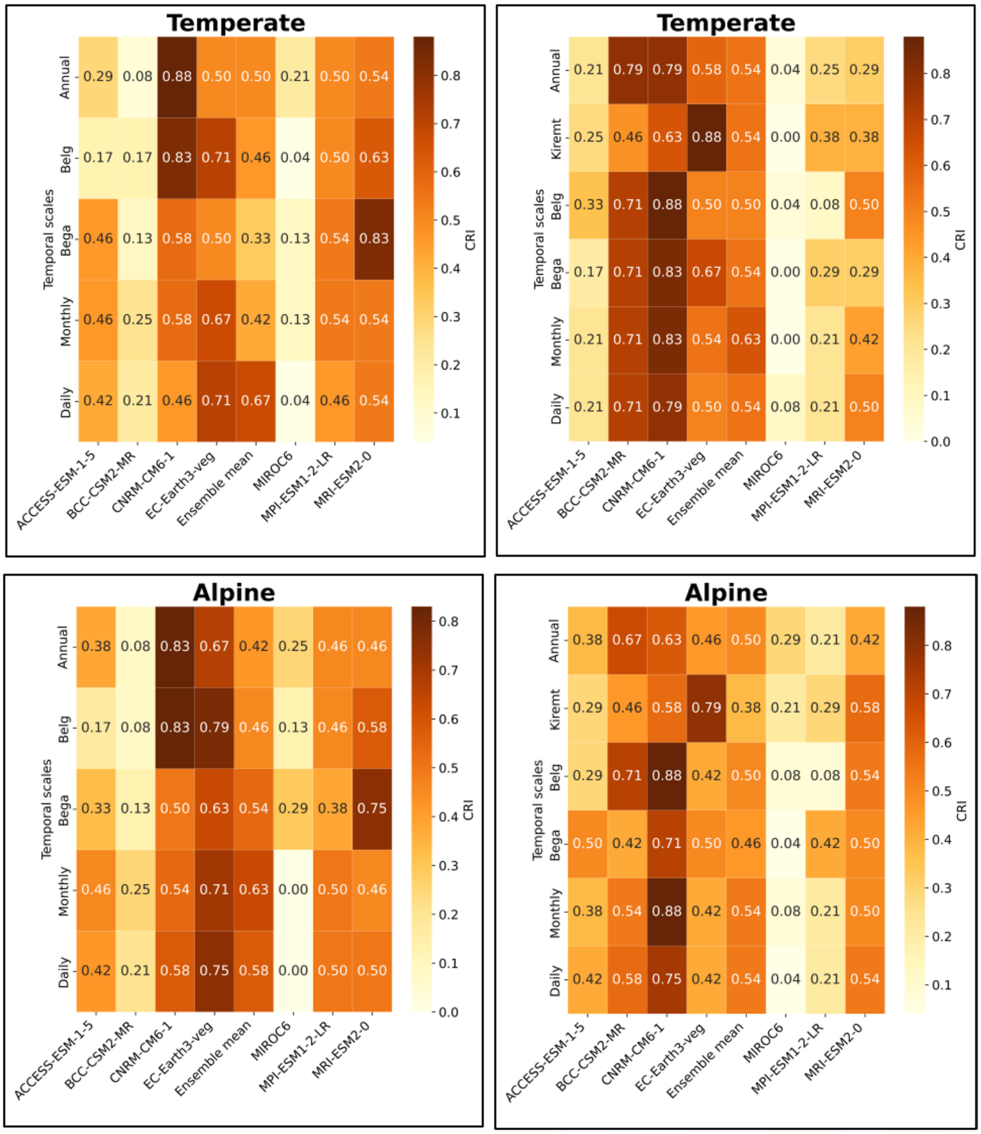


Fig. 3 (continued)

mean is not among the three best datasets in the four AEZs. Similar to its performance on the daily time scale, MIROC6 showed the weakest simulation accuracy on the monthly time scale across all AEZs.

Aligned with the findings of this study, EC-Earth3-veg and MPI-ESM1-2-LR are among the best-performing GCMs for simulating monthly Tmax over the Bale Eco-Region (Gashaw et al. 2024). Additionally, MPI-ESM1-2-LR has demonstrated strong performance over the Lake Tana sub-basin (Lebeza et al. 2024). Similarly, BCC-CSM-2-MR has been found to perform exceptionally well in the Rift Valley Basin (Woldemariam et al. 2025). In contrast, MIROC6, which exhibited poor performance on a daily temporal scale, also ranks as the weakest GCM for simulating monthly Tmax across the five AEZs in this study (Fig. 3). This finding is consistent with previous studies in Ethiopia (Lebeza et al. 2024) and other regions in the world (Belazreg et al. 2022), further underscoring its limitations in temperature simulation. Although the Ensemble mean was not among the top three models for simulating monthly Tmax in this study in most AEZs, the Ensemble of 27 models was ranked among the best-performing models in the Rift Valley Basin (Woldemariam et al. 2025), underscoring that the performance of the Ensemble mean varies depending on the set of models included and the region considered.

Regarding the estimation bias of the studied GCMs across AEZs, a greater number of GCMs have an underestimation bias for simulating daily and monthly Tmax over the desert, tropical, and sub-tropical AEZs (Supplementary Table 2). In contrast, the entire GCMs in alpine AEZ and the majority/entire GCMs in the temperate AEZ have overestimation bias for estimating daily and monthly Tmax. However, a study in

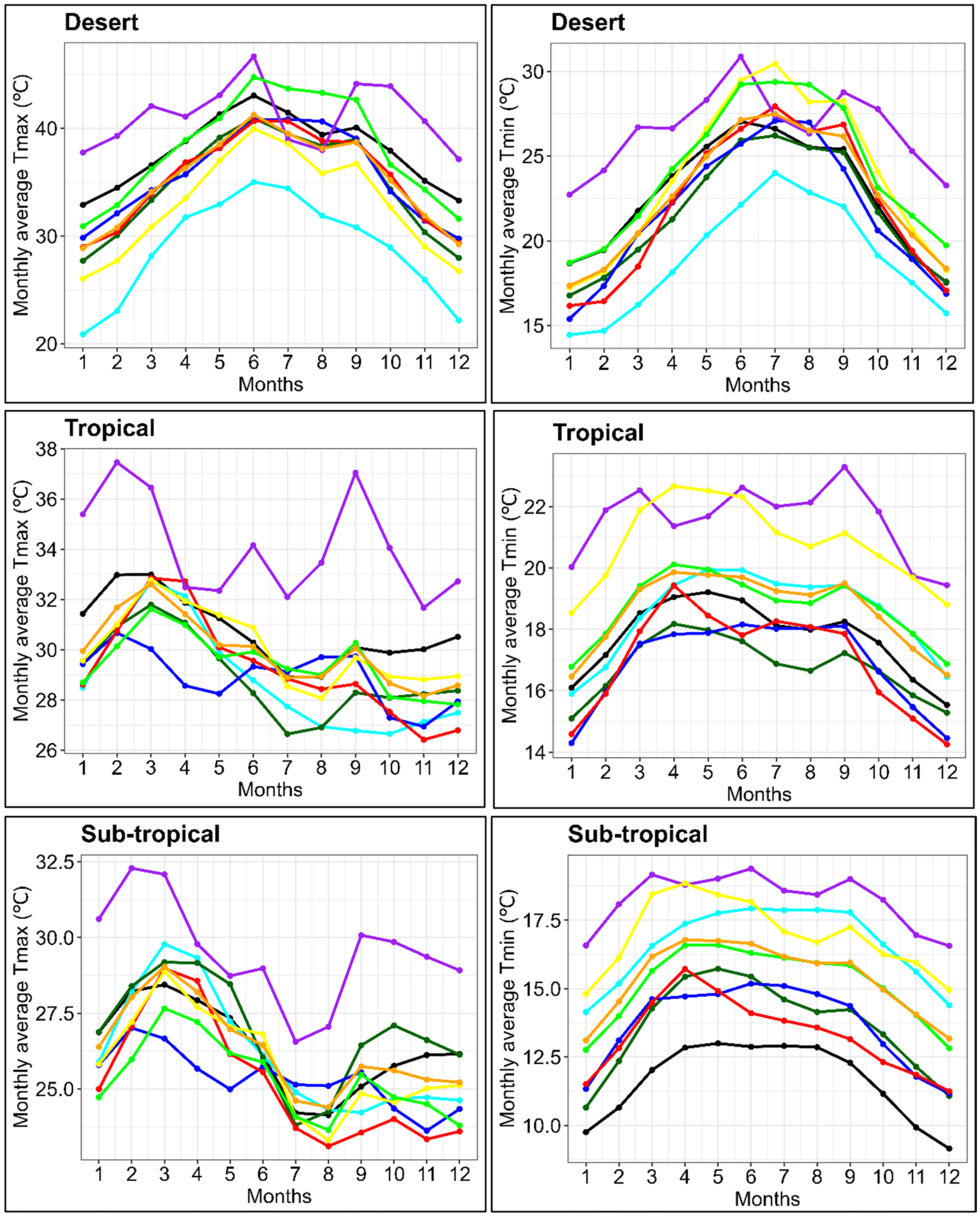


Fig. 4 The long-term monthly average (1995–2014) Tmax (left) and Tmin (right) values of the observed and studied GCMs over the desert, tropical, sub-tropical, temperate, and alpine AEZs of Ethiopia

the Bale Eco-Region indicated that six out of the ten studied GCMs exhibited overestimation bias during the daily and monthly Tmax estimates (Gashaw et al. 2024). At both daily and monthly temporal scales, the alpine AEZ exhibits the highest extent of biases in simulating Tmax compared to the other four AEZs (Supplementary Table 2).

The long-term monthly average (1995–2014) Tmax values of the observed and studied GCMs over the five AEZs of Ethiopia are demonstrated in Fig. 4. The results indicated that ACCESS-ESM-1–5 over the desert AEZ and MIROC6 over the remaining four AEZs displayed the highest deviation from the observed Tmax (Fig. 4). Most of the GCMs underestimated the long-term monthly average Tmax over the desert and tropical AEZs. In contrast, all GCMs over the alpine and the majority of GCMs over the temperate AEZs overestimated the observed Tmax (Fig. 4).

During *Bega* season, MRI-ESM-2-0, Ensemble mean and Ec-Earth3-veg over desert AEZ, Ensemble mean, MPI-ESM1-2-LR and CNRM-CM6-1over tropical AEZ, and Ensemble mean, MPI-ESM1-2-LR and BCC-CSM2-MR over the sub-tropical AEZs are the best models for simulating Tmax (Fig. 3). In temperate AEZ, MRI-ESM-2-0, CNRM-CM6-1 and MPI-ESM1-2-LR ranked from one to three, respectively. The best-performing models over the alpine AEZ are MRI-ESM-2-0, Ec-Earth3-veg and Ensemble mean. Thus, the Ensemble mean over the four AEZs, and MPI-ESM1-2-LR and MRI-ESM-2-0 over the three AEZs are among the three best-performing climate models for

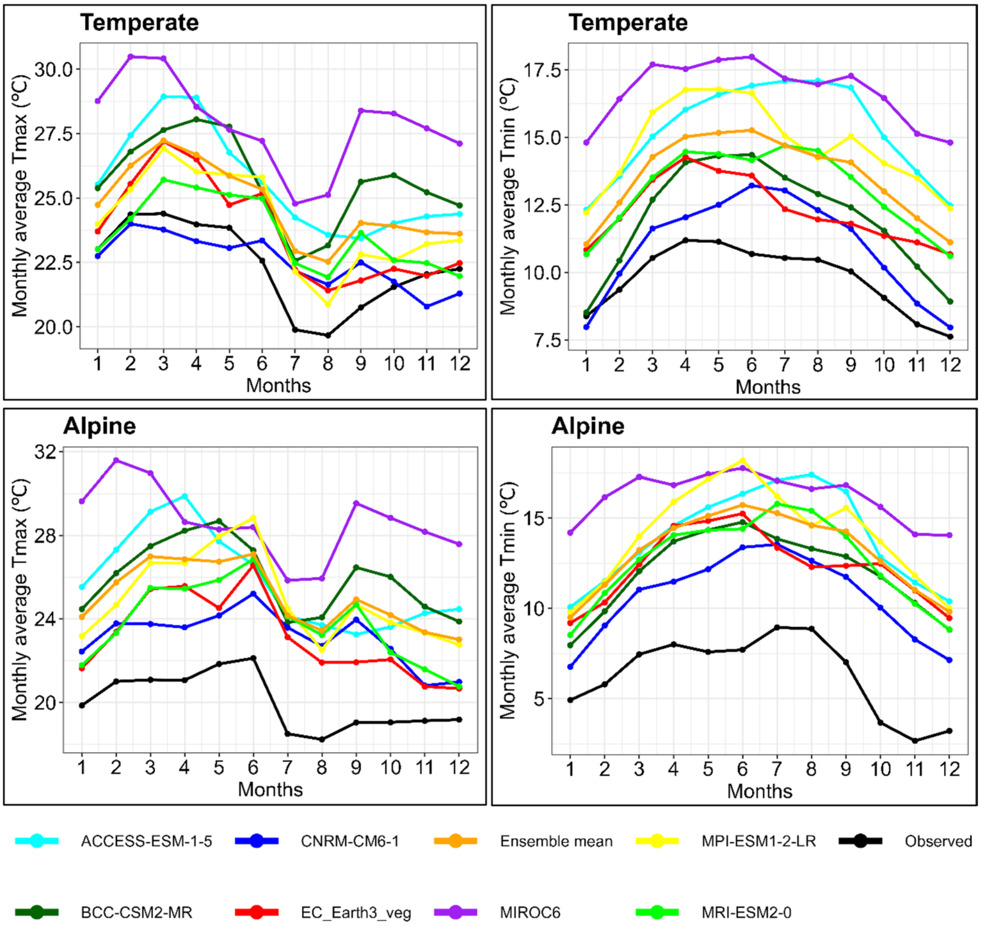


Fig. 4 (continued)

simulating Tmax during *Bega* season. The poor-performing GCMs include ACCESS-ESM-1–5 and BCC-CSM2-MR over the desert AEZ, ACCESS-ESM-1–5 and Ec-Earth3-veg over tropical, Ec-Earth3-veg and MIROC6 over sub-tropical, and BCC-CSM2-MR and MIROC6 over the temperate and alpine AEZs.

During *Belg* season, Ensemble mean, CNRM-CM6-1 and Ec-Earth3-veg in desert AEZ, and Ec-Earth3-veg, MPI-ESM1-2-LR and Ensemble mean in tropical and sub-tropical AEZs are the best performing three models (Fig. 3). CNRM-CM6-1, Ec-Earth3-veg and MRI-ESM-2-0 over temperate and alpine AEZs showed superior performance. Therefore, Ec-Earth3-veg over the five AEZs, and CNRM-CM6-1 and Ensemble mean over the three AEZs are among the three best climate datasets. The poor-performing GCMs include MPI-ESM1-2-LR, MRI-ESM-2-0 and ACCESS-ESM-1–5 in desert AEZ, MIROC6 and CNRM-CM6-1 in tropical and sub-tropical AEZs, MIROC6, ACCESS-ESM-1–5 and BCC-CSM2-MR in temperate AEZ, and BCC-CSM2-MR and MIROC6 in alpine AEZ (Fig. 3).

During the *Kiremt* season, GCMs such as CNRM-CM6-1, MIROC6 and Ensemble mean perform best for simulating Tmax during the *Kiremt* season in desert regions (Fig. 3). In tropical AEZ, the better models are Ensemble mean, MPI-ESM1-2-LR and MRI-ESM-2-0. For sub-tropical AEZ, the best models include MPI-ESM1-2-LR, ACCESS-ESM-1–5, and BCC-CSM2-MR. CNRM-CM6-1, MPI-ESM1-2-LR and Ec-Earth3-veg in the temperate AEZ, and Ec-Earth3-veg, ACCESS-ESM-1–5, and CNRM-CM6-1 in alpine AEZ stood as the leading best GCMs for simulating *Kiremt* season Tmax. The GCMs that exhibited poor performance for simulating Tmax during *Kiremt* season include ACCESS-ESM-1–5 and MRI-ESM-2-0 over the desert AEZ, MIROC6 and ACCESS-ESM-1–5 over the tropical and temperate AEZs, MIROC6 and Ec-Earth3-veg over sub-tropical AEZ, and MIROC6 and BCC-CSM2-MR over alpine AEZ (Fig. 3). Hence, CNRM-CM6-1 and MPI-ESM1-2-LR are the best models for estimating *Kiremt* season Tmax over the three AEZs. MIROC6 over the four AEZs and ACCESS-ESM-1–5 over the three AEZs are the poorest climate datasets.

Concerning the estimation bias of the GCMs, most GCMs in desert, tropical and sub-tropical AEZs underestimated *Bega* season Tmax (Supplementary Table 3). In contrast, most GCMs in the temperate AEZ and all GCMs in alpine AEZ overestimated Tmax during *Bega* season. During *Belg* season, the majority of GCMs in desert and tropical AEZs underestimated Tmax, but most GCMs over sub-tropical and temperate AEZs overestimated Tmax. In alpine AEZ, all GCMs overestimated the observed Tmax during *Belg* season (Supplementary Table 3). During *Kiremt*, all the studied GCMs overestimated Tmax over temperate and alpine AEZ (Supplementary Table 4). In addition, most GCMs over desert AEZ overestimated Tmax, but most GCMs underestimated the *Kiremt* season Tmax in tropical and sub-tropical AEZs. In general, the findings indicated a higher estimation bias in climate models over the alpine AEZ during the *Bega*, *Belg*, and *Kiremt* seasons (Supplementary Tables 3–4).

At annual time scale, the leading best-performing GCMs for simulating Tmax are MRI-ESM-2-0, Ensemble mean and CNRM-CM6-1 over desert AEZ, Ensemble mean, MPI-ESM1-2-LR and MRI-ESM-2-0 in tropical AEZ, Ensemble mean, MPI-ESM1-2-LR and ACCESS-ESM-1–5 over sub-tropical AEZ (Fig. 3). CNRM-CM6-1, MRI-ESM-2-0, Ec-Earth3-veg, MPI-ESM1-2-LR and Ensemble mean in temperate AEZ, and CNRM-CM6-1, Ec-Earth3-veg, MPI-ESM1-2-LR, and MRI-ESM-2-0 over the alpine AEZ are the other best-performing climate models. The GCMs with poor simulation performance include ACCESS-ESM-1–5 and BCC-CSM2-MR in the desert, Ec-Earth3-veg and MIROC6 in sub-tropical AEZ, and BCC-CSM2-MR and MIROC6 over tropical, temperate and alpine AEZs (Fig. 3). In general, MRI-ESM-2-0, Ensemble mean and MPI-ESM1-2-LR over four AEZs, and CNRM-CM6-1 over three AEZs are among the best three GCMs for simulating annual Tmax. On the other hand, BCC-CSM2-MR and MIROC6 are the poor-performing GCMs in the four AEZs of the country. Related to the findings of our study, the better performance of CNRM-CM6-1 and the poor performance of MIROC6 were reported over Thailand (Kamworapan et al. 2021). Likewise, the better performance of the Ensemble mean of 18 GCMs for estimating annual Tmax was also reported in Bangladesh (Kamruzzaman et al. 2023). The poor performance of MIROC6 was also stated over the Lake Tana sub-basin in Ethiopia (Lebeza et al. 2024).

Concerning the estimation bias of the studied GCMs for simulating annual Tmax, the finding revealed that the majority of the GCMs over the desert, tropical, and sub-tropical AEZs exhibited an underestimation bias (Supplementary Table 4). In contrast, the analyzed GCMs displayed an overestimation bias in temperate and alpine AEZs. A study in the Bale Eco-Region, which took the areal average values of the GCMs and observed data, indicated that most of the studied GCMs overestimated annual Tmax (Gashaw et al. 2024). Thus, the findings from this study clearly indicated that performance evaluations of GCMs over the different AEZs are imperative to identify the direction of the estimation biases of the GCMs, which will provide inputs for future model development. Like the daily, monthly and seasonal temporal scales, the greatest estimation bias of the GCMs is obtained over the alpine AEZ (Supplementary Table 4).

The finding also indicated that there are significant differences between the models, and no single model could adequately represent every characteristic of the observational datasets. For example, CNRM-CM6-1, which ranked first for *Belg* season Tmax over alpine and temperate, showed lowest ranking over the sub-tropical AEZ (Fig. 3). Similarly, ACCESS-ESM-1–5, which is a leading GCM for simulating monthly Tmax over the sub-tropical AEZ of Ethiopia, is among the lowest performing GCMs over the desert AEZ (Fig. 3). Moreover, MIROC6, which is one the poorest performing models for simulating daily and monthly Tmax in the entire AEZs of the country, is among the best three models for simulating *Kiremt* season Tmax over the desert AEZ. Furthermore, EC-Earth3-veg is one of the top three performing GCMs for simulating daily Tmax across the alpine, temperate, tropical and desert AEZs and monthly Tmax in the three AEZs. However, it performs poorly in simulating the *Bega* season in both tropical and sub-tropical AEZs. Our results are in agreement with the findings reported for the Middle Awash sub-basin (Tesfaye et al. 2023), which indicated that no single GCM could accurately reproduce the spatial pattern in different temporal scales.

Therefore, it is essential to evaluate climate models to identify the most accurate ones tailored to specific AEZs and temporal scales. The inconsistency of GCMs in simulating Tmax across different AEZs can be attributed to various factors. One potential explanation is the quality of the observed data, which may be compromised by inhomogeneities caused by station malfunctions, changes in measurement techniques, station relocations, or evolving environmental conditions over time (Ducré-Robitaille et al. 2003). Another critical factor is the influence of topography and climatic variability. Research conducted in the Tibetan Plateau (Lun et al. 2021) highlighted significant differences in how GCMs simulated various climatic variables, underscoring the impact of complex terrain. Similarly, Ethiopia's rugged topography and substantial altitudinal variations are likely to influence the performance of GCMs across different AEZs, further emphasizing the need for region-specific evaluations.

3.2 Performance of GCMs for simulating Tmin

The performances of the studied GCMs for simulating Tmin at different temporal scales over the five AEZs of the country are shown in Figs. 3 and 4 and Supplementary Tables 5–7. The result indicated that Ensemble mean, BCC-CSM-2-MR and CNRM-CM6-1 are the top-ranking GCMs for daily Tmin across the desert and temperate AEZs while Ensemble mean, ACCESS-ESM-1–5 and EC-Earth3-veg record the highest over the tropical AEZ (Fig. 3). CNRM-CM6-1, EC-Earth3-veg, and BCC-CSM-2-MR in sub-tropical AEZ are the top-performing GCMs for simulating daily Tmin. The best-performing GCMs for simulating daily Tmin over alpine AEZ include CNRM-CM6-1, BCC-CSM-2-MR, MRI-ESM2-0 and Ensemble mean. Therefore, CNRM-CM6-1 and BCC-CSM-2-MR over four AEZs, and Ensemble mean over three AEZs are among the three top performing datasets for estimating daily Tmin. Like Tmax, the poorest performing GCM for simulating daily Tmin across all AEZs is MIROC6 (Fig. 3).

The top-ranking GCMs for monthly Tmin in the alpine, temperate and desert AEZs include CNRM-CM6-1, BCC-CSM-2-MR, and Ensemble mean, while the GCMs with the greatest rankings across tropical AEZ are Ensemble mean, CNRM-CM6-1, ACCESS-ESM-1–5 and EC-Earth3-veg (Fig. 3). In the sub-tropical AEZ, CNRM-CM6-1, BCC-CSM-2-MR and EC-Earth3-veg exhibited the highest performance for simulating monthly Tmin. In general, CNRM-CM6-1 is one of the three best-performing GCMs for simulating monthly Tmin over the five AEZs while BCC-CSM-2-MR and Ensemble mean are among the top three outclassing GCMs over the four AEZs (Fig. 3). MIROC6 is the poorest performing GCM for monthly Tmin across all AEZs.

Consistent with the findings of this study, CNRM-CM6-1 ranks among the top-performing GCMs for simulating both daily and monthly Tmin over the Bale Eco-Region in Ethiopia (Gashaw et al. 2024). Similarly, BCC-CSM-2-MR and CNRM-CM6-1 are identified as two of the four best-performing GCMs for estimating monthly Tmin in the Rift Valley Lakes Basin (Woldemariam et al. 2025). Furthermore, CNRM-CM6-1 has been reported as the highest-performing GCM for simulating daily Tmin over Thailand (Shiru and Chung 2021). Similarly, the Ensemble mean of 27 GCMs is among the top performers for simulating monthly Tmin in the Central Rift Valley region of Ethiopia (Woldemariam et al. 2025). Although MRI-ESM2-0 is not among the three best climate models in any of the AEZs for estimating monthly Tmin in our study, it is recognized as one of the top-performing GCMs in Ethiopia's Upper Blue Nile Basin (Alaminie et al. 2021). Conversely, the poor performance of MIROC6 in simulating monthly Tmin in this study aligns with findings from the Lake Tana sub-basin, where it was also identified as a weak performer (Lebeza et al. 2024).

Most studied GCMs underestimated Tmin over the desert AEZ, while many GCMs overestimated it over the tropical AEZ on daily and monthly temporal scales (Supplementary Table 5). In sub-tropical, temperate, and alpine AEZs, the entire GCMs illustrated an overestimation bias when simulating daily and monthly Tmin. In alignment with the findings of this study, an evaluation of 13 CMIP6 GCMs conducted across East Africa-including Kenya, Rwanda, Tanzania, Uganda, and Burundi-indicated that most GCMs tended to overestimate daily Tmin, with only a few instances of underestimation (Ayugi et al. 2021). Similarly, a study conducted in the Bale Eco-Region of Ethiopia found that the majority of the evaluated GCMs overestimated both daily and monthly Tmin during the 1995–2014 period (Gashaw et al. 2024). Compared to the remaining four AEZs, the greatest estimation bias of the GCMs for simulating daily and monthly Tmin in our study was attained over the alpine AEZ (Supplementary Table 5).

The long-term monthly average Tmin (1995–2014) values of the observed dataset and studied GCMs are depicted in Fig. 4. As illustrated in the graph, CNRM-CM6-1, EC-Earth3-veg, the Ensemble mean and BCC-CSM-2-MR closely align with the observed monthly average Tmin over the desert AEZ during the June-September period. In the sub-tropical and temperate AEZs, CNRM-CM6-1, EC-Earth3-veg, and BCC-CSM-2-MR show closer agreement with observations across most months. Similarly, over the alpine AEZ, these three models demonstrate better correspondence with observed Tmin during June-September. Except for CNRM-CM6-1 for temperate AEZ in January, all GCMs overestimate the monthly average Tmin for all months over sub-tropical, temperate, and alpine AEZs. Across the desert and tropical AEZs, some GCMs exhibited underestimation biases, whereas the remaining models showed overestimation biases. Overall AEZs, MIROC6 is far from the observed line graph, hence not recommended for all months.

During *Bega* season, Ec-Earth3-veg, Ensemble mean and CNRM-CM6-1 over desert, BCC-CSM2-MR, Ensemble mean and CNRM-CM6-1 over tropical, and Ec-Earth3-veg, BCC-CSM2-MR and CNRM-CM6-1 over sub-tropical and temperate AEZs are the best-performing GCMs for simulating Tmin (Fig. 3). The climate models that showed leading performance for simulating Tmin during *Bega* season over the alpine AEZ are CNRM-CM6-1, Ec-Earth3-veg, MRI-ESM2-0 and ACCESS-ESM-1–5. Therefore, CNRM-CM6-1, Ec-Earth3-veg and BCC-CSM2-MR over the five, four and three AEZs, respectively, are among the three best-performing GCMs for simulating *Bega* season Tmin. Similar to our study, CNRM-CM6-1 and Ec-Earth3-veg are among the best-performing GCMs in the Bale Eco-Region (Gashaw et al. 2024). BCC-CSM2-MR and MIROC6 over desert AEZ, MPI-ESM1-2-LR and MIROC6 over tropical, MIROC6, MPI-ESM1-2-LR and ACCESS-ESM-1–5 over sub-tropical, MIROC6 and ACCESS-ESM-1–5 over temperate, MIROC6, MPI-ESM1-2-LR and BCC-CSM2-MR over alpine AEZs are the poor-performing GCMs (Fig. 3). Thus, MIROC6 over the five AEZs and MPI-ESM1-2-LR over the three AEZs are among the poorly performing GCMs for simulating Tmin during *Bega* season. Comparable to our finding, the poor-performance of MIROC6 for simulating dry season Tmin was reported over the Lake Tana sub-basin (Lebeza et al. 2024).

On the other hand, the best performing GCMs for simulating *Belg* season Tmin include CNRM-CM6-1, Ensemble mean and MPI-ESM1-2-LR over desert AEZ, ACCESS-ESM-1–5, Ensemble mean, CNRM-CM6-1 and Ec-Earth3-veg over tropical AEZ, and CNRM-CM6-1, BCC-CSM2-MR, and MRI-ESM-2-0 over sub-tropical and alpine AEZs (Fig. 3). The GCMs that displayed better performance for estimating Tmin over the temperate AEZ are CNRM-CM6-1, BCC-CSM2-MR, Ec-Earth3-veg, MRI-ESM-2-0 and Ensemble mean. Thus, CNRM-CM6-1 over the five AEZs, and Ensemble mean, BCC-CSM2-MR and MRI-ESM-2-0 over the three AEZs are among the three best preforming models for simulating Tmin during *Belg* season. The least performing GCMs during *Belg* season over the tropical, sub-tropical, temperate, and alpine AEZs include MIROC6 and MPI-ESM1-2-LR. In addition, ACCESS-ESM-1–5 and MRI-ESM-2-0 over the desert AEZ are also the poor-performing GCMs during this season (Fig. 3).

The best-performing GCMs for simulating Tmin during *Kiremt* season over desert AEZ are BCC-CSM2-MR, Ensemble mean, CNRM-CM6-1 and Ec-Earth3-veg while Ec-Earth3-veg, CNRM-CM6-1 and Ensemble mean are leading performing models in tropical and temperate AEZs (Fig. 3). The other leading climate models include Ec-Earth3-veg, CNRM-CM6-1 and BCC-CSM2-MR over sub-tropical AEZ, and Ec-Earth3-veg, CNRM-CM6-1 and MRI-ESM-2-0 over alpine AEZ (Fig. 3). Therefore, Ec-Earth3-veg and CNRM-CM6-1 over the five AEZs, and the Ensemble mean over the three AEZs are among the three best datasets for simulating *Kiremt* season Tmin. Similarly, Ec-Earth3-veg and CNRM-CM6-1are also among the best-performing GCMs for simulating wet rainy season Tmin over the Bale Eco-Region, Ethiopia (Gashaw et al. 2024). In contrast, MIROC6 over the tropical, sub-tropical, temperate, and alpine AEZs and ACCESS-ESM-1–5 in the desert AEZ are the poorest-performing GCMs for simulating Tmin during *Kiremt* season (Fig. 3). Likewise, the poor performance of MIROC6 for simulating Tmin during the main rainy season was also reported over the Lake Tana sub-basin (Lebeza et al. 2024).

Concerning the estimation bias of the GCMs for estimating Tmin during the three seasons, all GCMs, except CNRM-CM6-1 in sub-tropical and BCC-CSM2-MR in alpine AEZs during *Bega* season, overestimated *Bega*, *Belg* and *Kiremt* seasons Tmin over the sub-tropical, temperate and alpine AEZs (Supplementary Tables 6–7). Over the desert and tropical AEZs, half of the GCMs overestimated while the remaining half of them had underestimation bias during *Bega* season, but the majority of GCMs overestimated Tmin during *Kiremt* season (Supplementary Tables 6–7). For *Belg* season, most GCMs underestimated Tmin over the desert AEZ, while a large number of GCMs overestimated Tmin over the tropical AEZ (Supplementary Table 6). Our finding is related to the results reported for the Bale Eco-Region, which indicated an overestimation of most GCMs for simulating seasonal Tmin (Gashaw et al. 2024). In all three seasons, the highest number of estimation bias was recorded over the alpine AEZ. In addition, the estimation bias of the GCMs is greater in the temperate AEZ during *Kiremt* season compared to the desert, tropical, and sub-tropical AEZs.

For simulating annual Tmin, Ec-Earth3-veg, Ensemble mean, BCC-CSM2-MR over desert, Ec-Earth3-veg, ACCESS-ESM-1–5 and Ensemble mean over tropical, Ec-Earth3-veg, BCC-CSM2-MR and CNRM-CM6-1 over sub-tropical and temperate, and BCC-CSM2-MR, CNRM-CM6-1 and Ensemble mean over alpine AEZs are the top performing GCMs (Fig. 3). Therefore, Ec-Earth3-veg and BCC-CSM2-MR over four, and Ensemble mean over the three AEZs are among the three best performing datasets for estimating annual Tmin. Aligned with the findings of our study, Ec-Earth3-veg has also shown superior performance over the Bale Eco-Region (Gashaw et al. 2024). The poor-performing GCMs are MIROC6 and MPI-ESM1-2-LR over the tropical, sub-tropical and alpine AEZs, and MIROC6 and ACCESS-ESM-1–5 over the desert and temperate AEZs. Likewise, ACCESS-ESM-1–5 is also among the poorest performing GCMs over the Bale Eco-Region (Gashaw et al. 2024). Similarly, the superior performance of the Ensemble mean of 18 GCMs in estimating annual Tmax was also reported in Bangladesh (Kamruzzaman et al. 2023). In addition, MIROC6 is among the poor-performing GCMs for simulating Tmin over the Lake Tana sub-basin at an annual temporal scale (Lebeza et al. 2024).

Regarding the estimation bias of the considered GCMs for simulating annual Tmin, all GCMs over the sub-tropical, temperate, and alpine AEZs and most GCMs over the tropical AEZs have overestimated annual Tmin (Supplementary Table 7). In the desert AEZ, half of them exhibited overestimation bias while the remaining half of the GCMs demonstrated underestimation bias. Likewise, the overestimation of most GCMs for simulating annual Tmin was also reported over the Bale Eco-Region (Gashaw et al. 2024). Compared to the remaining AEZs, the estimation bias is generally higher in the alpine AEZ. The result also indicated that, except for CNRM-CM6-1 over temperate AEZ, the estimation bias is greater in temperate AEZ compared to desert, tropical and sub-tropical AEZs (Supplementary

Table 7). Therefore, the result indicated the presence of greater estimation bias in high altitude areas.

3.3 Implications of the findings

Our study found that no single model consistently performed well in simulating both Tmax and Tmin across the studied spatial and temporal scales, a result consistent with previous studies conducted in Ethiopia (Legass et al. 2025; Gashaw et al. 2024; Lebeza et al. 2024; Demessie et al. 2023) and other regions globally (Cruz-González et al. 2025). Consequently, Gashaw et al. (2024) identified the best-performing climate models for Tmax and Tmin separately for their future climate change analysis of the Bale Eco-Region, while Legass et al. (2025) adopted a similar approach for the Awash Basin. The primary reason a climate model may perform well in estimating Tmax but poorly for Tmin, or vice versa, may be attributed to its representation of cloud processes (Chen et al. 2022; Lobell et al. 2007). These findings underscore the importance of using the best-performing models for Tmax and Tmin in analyzing a future climate. Employing the same GCM for both variables without performance assessment could introduce significant uncertainties in future climate projections. This, in turn, could affect the reliability of studies on climate change impacts on critical sectors such as water resources, agriculture, sediment transport, and ecosystem services, as well as efforts to climate-resilient modeling.

The Ensemble mean of the evaluated GCMs failed to demonstrate superior performance across some temporal scales and AEZs. Specifically, the Ensemble mean is not among the three best climate models for estimating Tmax across most AEZs for monthly and *Kiremt* temporal scales, and for simulating Tmin during *Bega* season. This may be due to the fact that the Ensemble mean contains all the 7 climate models studied, including the lowest performing GCMs such as MIROC6, for Tmax and Tmin across most AEZs and temporal scales. Therefore, using the Ensemble mean of all climate models for climate change analysis without prior performance evaluation could introduce greater uncertainties compared to employing the Ensemble mean of the best-performing models. In this context, Woldemariam et al. (2025) found that, in the Rift Valley Lake Basin of Ethiopia, the Ensemble mean of 27 CMIP6 GCMs underperformed compared to the Ensemble mean of the four best-performing models for estimating Tmax and Tmin. In addition, Berhanu et al. (2023) reported that, for simulating rainfall over Ethiopia, the Ensemble mean of four selected models outperformed the Ensemble mean of all models. Moreover, Yang et al. (2023) reported that the Ensemble mean of all models for future climate change analysis should be used with caution, as it may incorporate information from poorly performing models. This highlights the importance of selecting and prioritizing well-performing models to ensure more reliable and accurate climate change projections.

Given the variability in GCM performance across AEZs, using the same set of models for future climate analyses in studies that rely on point-based data, such as most hydrological models, across an entire region may introduce substantial uncertainty. Instead, selecting the best-performing climate models for each specific region would provide more reliable projections. For grid-based studies such as country-level studies, it would be more appropriate to use the Ensemble mean of the best-performing models from different regions, ensuring a more nuanced and accurate representation. Similarly, as the performance of climate models varies across temporal scales, employing a single model for an entire temporal scale without prior performance assessment can introduce significant uncertainties. Selecting the best-performing models for each temporal scale, depending on the study's objectives, is a more robust approach. For studies spanning multiple temporal scales or the entire temporal range, using the Ensemble mean of the best-performing models for each temporal scale is advisable (Gashaw et al. 2024). This approach ensures greater accuracy and reduces uncertainties in climate projections.

3.4 Limitations and contributions of the study

A key limitation of this study is the reliance on a relatively small number of meteorological stations to represent a country spanning ~ 1.12 million km². This limitation arises primarily from the criterion of permitting no more than 25% missing data. Furthermore, the selection of stations was influenced by the corresponding authors' familiarity with specific locations. Similar challenges have been reported in many temperature studies conducted in Ethiopia, where reliance on a relatively small number of meteorological stations has been common due to data availability constraints (Tadesse et al. 2025; Gashaw et al. 2023; Tesfaye et al. 2017). For example, Gashaw et al. (2023) and Tadesse et al. (2025) used the same number of meteorological stations for their study in Ethiopia. In addition, Tesfaye et al. (2017) also used data from 15 meteorological stations to evaluate gridded temperature products in Ethiopia. However, if there are many stations that contain less than 25% missing values and the selection of stations is not influenced by the corresponding authors' familiarity with specific locations, incorporating additional stations, selected to represent the spatial coverage of the country, can improve the strength of the findings when evaluating the performance of CMIP6 GCMs. The evaluation of the climate models merely using station data is also another limitation of the study.

Another limitation relates to the number of GCMs employed in this study. It is noted that our evaluation assessed the performance of these models across six temporal scales (daily, monthly, *Bega*, *Belg*, *Kiremt*, and annual), five AEZs (desert, tropical, sub-tropical, temperate, and alpine), and for Tmax and Tmin separately. Given this multidimensional evaluation framework, the decision to focus on a smaller set of models was deliberate, as expanding the number of GCMs would considerably increase analytical complexity. Unlike many previous studies that were conducted without considering AEZs scale (Gashaw et al. 2024; Lebeza et al. 2024; Alaminie et al. 2021) and focused on a few of the temporal scales, our study provides a more comprehensive assessment by covering the full temporal spectrum, multiple AEZs, and separate analyses of Tmax and Tmin. However, considering that more than 40 CMIP6 GCMs are available, evaluating the full suite of models, or at least more than half of them, would further enhance the robustness of the findings. In addition, this study did not assess the performance of CMIP6 GCMs for rainfall, which represents another limitation. Conducting evaluations for Tmax and Tmin alongside precipitation would provide a more complete assessment, especially given that models performing well for temperature variables may not necessarily exhibit similar skill in simulating precipitation.

Another limitation of this study is that we did not separately evaluate the performance of the best-performing GCMs. Conducting evaluations of the raw GCMs, the Ensemble mean of all models, and the Ensemble mean derived only from the best-performing GCMs together would provide quantitative evidence on whether the Ensemble mean of the best-performing models typically outperforms or underperforms the Ensemble mean of all models. Further, the lack of exploration for Regional Climate Models (RCM), mainly the Coordinated Regional Climate Downscaling Experiment (CORDEX)-Africa runs, which were downscaling CMIP6 models, is another limitation of the study. However, because RCMs better capture local climatic processes than GCMs, evaluating the downscaled GCMs used to drive the CORDEX RCMs would offer more meaningful guidance for future climate studies than assessments based solely on the raw CMIP6 GCMs.

Despite these limitations, this study makes important contributions. It demonstrates the performance of multiple datasets across different temporal scales and evaluates climate models within distinct AEZs, highlighting performance variations across regions and temporal scales. The findings underscore the importance of carefully assessing climate models for specific temporal scales and locations. Furthermore, the study emphasizes the need to evaluate climate models separately for Tmax and Tmin, offering valuable insights for future climate research and applications.

4 Conclusion

This study evaluated the performance of seven CMIP6 GCMs and their Ensemble mean in simulating historical (1995–2014) Tmax and Tmin across Ethiopia's five AEZs on daily to annual temporal scales. The analysis revealed significant variations in GCM performance across different AEZs and time scales. For instance, EC-Earth3-veg emerged as one of the best-performing models for simulating daily, monthly, and *Belg* season Tmax across temperate, alpine and desert AEZs. However, it performed poorly in simulating Tmax during the *Bega* season over tropical and sub-tropical AEZs and *Kiremt* season Tmax over sub-tropical AEZ. Conversely, while MIROC6 is one of the poorest-performing GCMs for simulating daily and monthly Tmax across all AEZs, it demonstrated strong performance for simulating Tmax during *Kiremt* season in the desert AEZ.

The results of this study revealed that the Ensemble mean and EC-Earth3-veg models at the daily scale, and EC-Earth3-veg, MPI-ESM1-2-LR and BCC-CSM2-MR at the monthly scale, are among the most reliable GCMs for estimating Tmax across most AEZs. Furthermore, Ensemble mean, MPI-ESM1-2-LR and MRI-ESM2-0 during the *Bega* season, EC-Earth3-veg, CNRM-CM6-1 and Ensemble mean during the *Belg* season, CNRM-CM6-1 and MPI-ESM1-2-LR during the *Kiremt* season, and MRI-ESM2-0, Ensemble mean, MPI-ESM1-2-LR and CNRM-CM6-1 at the annual scale performed best for Tmax estimation in most AEZs. Similarly, for Tmin, the CNRM-CM6-1, BCC-CSM2-MR and Ensemble mean models demonstrated superior performance at both daily and monthly time scales across the majority AEZs. The CNRM-CM6-1, EC-Earth3-veg and BCC-CSM2-MR models during the *Bega* season, CNRM-CM6-1, Ensemble mean, BCC-CSM2-MR and MRI-ESM2-0 during the *Belg* season, EC-Earth3-veg, CNRM-CM6-1 and Ensemble mean during the *Kiremt* season, and EC-Earth3-veg, BCC-CSM2-MR and Ensemble mean at the annual scale performed best in most AEZs. In contrast, MIROC6 consistently exhibited the weakest performance in simulating both Tmax and Tmin across most AEZs and temporal scales.

The findings indicated that one of the best models for simulating Tmax in tropical and sub-tropical AEZs, the MPI-ESM1-2-LR, performed poorly when simulating Tmin over most studied temporal scales (daily to annual). This finding suggests the need to evaluate climate models for Tmax and Tmin independently. Additionally, the study found that biases in simulating both Tmax and Tmin were relatively higher in alpine AEZs compared to other regions, highlighting the challenges climate models face in accurately capturing temperature dynamics in high-elevation areas. This finding calls for further research to understand the underlying reasons for these biases in mountainous areas.

The findings indicate that the Ensemble mean of the evaluated GCMs did not consistently achieve superior performance across certain temporal scales and AEZs. Consequently, relying on the Ensemble mean of all models may introduce greater uncertainty compared to using an Ensemble mean based solely on the best-performing models. The identified top-performing CMIP6 models serve as valuable tools for assessing climate change impacts and developing region- and season-specific adaptation strategies in Ethiopia. For studies encompassing the entire country and multiple temporal scales, employing the ensemble of the best-performing GCMs is recommended, given the variability in model performance across AEZs and time periods. These results provide practical guidance for climatologists, hydrologists, and water resource managers in selecting suitable GCMs for sectoral applications and offer a useful reference for regions with hydro-meteorological conditions similar to Ethiopia. Climate models that perform well for Tmax or Tmin may not necessarily perform well for precipitation. Therefore, this study recommends further research focusing on the evaluation of CMIP6 GCMs for precipitation, Tmax, and Tmin.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00704-026-06066-5.

Acknowledgements The authors are glad to thank the Ethiopian Meteorology Institute (EMI) for providing the observed climate data. This research was undertaken during the corresponding author's stays in Bahir Dar University (Ethiopia), Wageningen University and Research (Netherlands), and Prairie View A & M University (USA), and hence the authors acknowledge these institutions for their support. The authors are also grateful to the Editor of the journal and the two anonymous reviewers for their comments, which thoroughly improved the quality of the paper.

Author contributions A.B.T: Conceptualization, data curation, methodology, software, visualization, writing-original draft, and writing-review & editing; T.G.T: Conceptualization, data curation, methodology, software, visualization, supervision, writing-original draft, and writing-review & editing; P.R.v.O, R.L.R, G.W.T, S.F.D, A.A, D.B: Conceptualization, writing-original draft, and writing-review & editing; O.C.O, A.H, A.D.C: Writing-original draft, and writing-review & editing; A.W.W, A.B, Y.T.D, E.Y: Writing-review & editing; T.A.T: Software, writing-original draft, and writing-review & editing. All authors read and approved the final version of the manuscript.

Funding This research did not obtain financial support from any institutions.

Data availability Since the observed climate data obtained from the Ethiopian Meteorology Institute (EMI) are for self-use only, the authors do not have permission to share the data. However, the studied CMIP6 GCMs are freely accessible from the Earth System Grid Federation (ESGF) database at https://aims2.llnl.gov/search.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abdulsahib SM, Zubaidi SL, Almamalachy Y, Dulaimi A (2024) Temperature and precipitation change assessment in the north of Iraq using LARS-WG and CMIP6 models. Water. https://doi.org/10.3390/w16192869

Alaminie AA, Tilahun SA, Legesse SA, Zimale FA, Tarkegn GB, Jury MR (2021) Evaluation of past and future climate trends under CMIP6 scenarios for the UBNB (Abay), Ethiopia. Water. https://doi.org/10.3390/w13152110

Aniley E, Gashaw T, Abraham T, Demessie SF, Bayabil HKW, Abeyou W, Oel PRv, Dile YT, Chukalla AD, Haileslassie A, Wubaye GB (2023) Evaluating the performances of gridded satellite/ reanalysis products in representing the rainfall climatology of Ethiopia. Geocarto Int. https://doi.org/10.1080/10106049.2023.2278329

Ayugi B, Ngoma H, Babaousmail H, Karim R, Iyakaremye V, Lim Kam Sian TC, Ongoma K, V (2021) Evaluation and projection of mean surface temperature using CMIP6 models over East afric. https://doi.org/10.20944/preprints202101.0611.v1

Bağçaci SÇ, Yucel I, Duzenli E, Yilmaz MT (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. Atmos Res. https://doi.org/10.1016/j.atmosres.2021.105576

Baig MA, Cui P, Ullah S, Bazai NA, Wei R, Ali M, Wang Y, Ullah W, Hussain I, Ramzan M, Asfandyar M, Purohit S (2025) Evaluation and projection of temperatures over pakistan: insights from the downscaled NEX-GDDP-CMIP6 models. Earth Syst Environ. https://doi.org/10.1007/s41748-024-00554-2

Belazreg NEH, Hasbaia M, Şen Z, Ferhati A (2022) Historical evaluation and future projections of monthly precipitation and temperature under CMIP6 GCMs, case of Hodna basin (central Algeria). Arab J Geosci. https://doi.org/10.1007/s12517-022-11124-z

Berhanu D, Alamirew T, Taye MT, Tibebe D, Gebrehiwot S, Zeleke G (2023) Evaluation of CMIP6 models in reproducing observed rainfall over Ethiopia. J Water Clim Change 14:2583–2605. https://doi.org/10.2166/wcc.2023.502

Buuren S, Groothuis-Oudshoorn K, Vink G, Schouten R, Robitzsch A, Rockenschaub P, Doove L, Jolani S, Moreno-Betancur M, White I, Gaffert P, Meinfelder F, Gray B, Arel-Bundock V, Cai M, Volker T, Costantini E, Lissa C, Oberman H (2022) Package 'mice' version 3.15.0; multivariate imputation by chained equations. https://doi.org/10.18637/jss.v045.i03

Chen Y, Marek GW, Marek TH, Moorhead JE, Heflin KR, Brauer DK, Gowda PH, Srinivasan R (2019) Simulating the impacts of climate change on hydrology and crop production in the Northern high plains of Texas using an improved SWAT model. Agric Water Manage 221:13–24. https://doi.org/10.1016/j.agwat.2019.04.021

Chen G, Wang WC, Bao Q, Li J (2022) Evaluation of simulated cloud diurnal variation in CMIP6 climate models. J Geophys Res Atmos. https://doi.org/10.1029/2021jd036422

Cruz-González A, Arteaga-Ramírez R, Soria-Ruiz J, Sánchez-Cohen I, Monterroso-Rivas AI, Quevedo-Nolasco A (2025) Climate change projections in temperature and precipitation using cmip6 in Central Mexico. Theor Appl Climatol. https://doi.org/10.1007/s00704-024-05345-3

Demessie SF, Dile YT, Bedadi B, Gashaw T, Tefera GW (2023) Evaluations of regional climate models for simulating precipitation and temperature over the Guder sub-basin of upper blue nile Basin, Ethiopia. Model Earth Syst Environ 9:4455–4476. https://doi.org/10.1007/s40808-023-01751-0

Ducré-Robitaille JF, Vincent LA, Boulet G (2003) Comparison of techniques for detection of discontinuities in temperature series. Int J Climatol 23:1087–1101. https://doi.org/10.1002/joc.924

Edwards PN (2011) History of climate modeling. WIREs Clim Change 2:128–139. https://doi.org/10.1002/wcc.95

Ersado DL, Awoke AG (2024) A multi-criteria decision analysis approach for ranking the performance of CMIP6 models in reproducing precipitation patterns over Abaya-Chamo sub-basin. Ethiopia Heliyon 10:e32442. https://doi.org/10.1016/j.heliyon.2024.e32442

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9:1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Federal Democratic Republic of Ethiopia Ministry of Water Resources National Meteorological Services Agency (2001) Initial National communication of Ethiopia to the united nations framework convention on climate change (UNFCCC), vol NMSA. Addis Ababa, Ethiopia

Gashaw T, Wubaye GB, Worqlul AW, Dile YT, Mohammed JA, Birhan DA, Tefera GW, van Oel PR, Haileslassie A, Chukalla AD, Taye MT, Bayabil HK, Zaitchik B, Srinivasan R, Senamaw A, Bantider A, Adgo E, Seid A (2023) Local and regional climate trends and variabilities in Ethiopia: implications for climate change adaptations. Environ Challenges. https://doi.org/10.1016/j.envc.2023.100794

Gashaw T, Worqlul AW, Taye MT, Lakew HB, Seid A, Ayele G, Haileslassie A (2024) Performance evaluations of CMIP6 model simulations and future projections of rainfall and temperature in the Bale Eco-Region, Southern Ethiopia. Theoret Appl Climatol 155:5069–5092. https://doi.org/10.1007/s00704-024-04904-y

Gebrechorkos SH, Hülsmann S, Bernhofer C (2018) Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania. Int J Climatol 39:18–30. https://doi.org/10.1002/joc.5777

Gebrechorkos SH, Taye MT, Birhanu B, Solomon D, Demissie T (2023a) Future changes in climate and hydroclimate extremes in East Africa. Earth's Future 11. https://doi.org/10.1029/2022ef003011

Gebrechorkos S, Leyland J, Slater L, Wortmann M, Ashworth PJ, Bennett GL, Boothroyd R, Cloke H, Delorme P, Griffith H, Hardy R, Hawker L, McLelland S, Neal J, Nicholas A, Tatem AJ, Vahidi E, Parsons DR, Darby SE (2023b) A high-resolution daily global dataset of statistically downscaled CMIP6 models for climate impact analyses. Sci Data 10:611. https://doi.org/10.1038/s41597-023-02528-x

Getachew B, Manjunatha BR (2022) Potential climate change impact assessment on the hydrology of the Lake Tana Basin, Upper Blue Nile River Basin, Ethiopia. Physics and Chemistry of the Earth, Parts A/B/C. https://doi.org/10.1016/j.pce.2022.103162

Gulakhmadov A, Chen X, Gulahmadov N, Rizwan M, Gulakhmadov M, Nadeem MU, Rakhimova M, Liu T (2025) Modeling of historical and future changes in temperature and precipitation in the Panj river basin in central Asia under the CMIP5 RCP and CMIP6 SSP scenarios. Sci Rep 15:3037. https://doi.org/10.1038/s41598-025-86366-4

Hamed MM, Nashwan MS, Shahid S, Ismail T, Wang X-j, Dewan A, Asaduzzaman M (2022) Inconsistency in historical simulations and future projections of temperature and rainfall: a comparison of CMIP5 and CMIP6 models over Southeast Asia. Atmos Res. https://doi.org/10.1016/j.atmosres.2021.105927

Hassan WH, Nile BK, Kadhim ZK, Mahdi K, Riksen M, Thiab RF (2023) Trends, forecasting and adaptation strategies of climate change in the middle and west regions of Iraq. SN Applied Sciences. https://doi.org/10.1007/s42452-023-05544-z

Hoseini SM, Soltanpour M, Zolfaghari MR (2025) Climate change impacts on temperature and precipitation over the Caspian sea. Int J Water Resour Dev 41:31–56. https://doi.org/10.1080/07900627.2024.2313050

Jia K, Ruan Y, Yang Y, You Z (2019) Assessment of CMIP5 GCM simulation performance for temperature projection in the Tibetan plateau. Earth Space Sci 6:2362–2378. https://doi.org/10.1029/2019EA000962

Kamruzzaman M, Shahid S, Islam ARMT, Hwang S, Cho J, Zaman MAU, Ahmed M, Rahman MM, Hossain MB (2021) Comparison of CMIP6 and CMIP5 model performance in simulating historical precipitation and temperature in Bangladesh: a preliminary study. Theoret Appl Climatol 145:1385–1406. https://doi.org/10.1007/s00704-021-03691-0

Kamruzzaman M, Wahid S, Shahid S, Alam E, Mainuddin M, Islam HMT, Cho J, Rahman MM, Chandra Biswas J, Thorp KR (2023) Predicted changes in future precipitation and air temperature across Bangladesh using CMIP6 GCMs. Heliyon 9:e16274. https://doi.org/10.1016/j.heliyon.2023.e16274

Kamworapan S, Bich Thao PT, Gheewala SH, Pimonsree S, Prueksakorn K (2021) Evaluation of CMIP6 GCMs for simulations of temperature over Thailand and nearby areas in the early 21st century. Heliyon 7:e08263. https://doi.org/10.1016/j.heliyon.2021.e08263

Karim R, Tan G, Ayugi B, Babaousmail H, Liu F (2020) Evaluation of historical CMIP6 model simulations of seasonal mean temperature over Pakistan during 1970–2014. Atmosphere. https://doi.org/10.3390/atmos11091005

Kawai H, Yukimoto S, Koshiro T, Oshima N, Tanaka T, Yoshimura H, Nagasawa R (2019) Significant improvement of cloud representation in the global climate model MRI-ESM2. Geosci Model Dev 12:2875–2897. https://doi.org/10.5194/gmd-12-2875-2019

Lebeza TM, Gashaw T, Bayabil HK, van Oel PR, Worqlul AW, Dile YT, Chukalla AD (2024) Performance of specific CMIP6 GCMs for simulating the historical rainfall and temperature climatology of Lake Tana sub-basin, Ethiopia. Scientific African. https://doi.org/10.1016/j.sciaf.2024.e02387

Legass AM, Alamirew T, Gebrehiwot SG, Haro-Monteagudo D, Berhanu D (2025) Spatiotemporal climate change projection and trend analysis using selected downscaled CMIP6 models for water action over awash River Basin, Ethiopia. Environmental Challenges. https://doi.org/10.1016/j.envc.2024.101059

Lobell DB, Bonfils C, Duffy PB (2007) Climate change uncertainty for daily minimum and maximum temperatures: a model intercomparison. Geophys Res Lett. https://doi.org/10.1029/2006gl028726

Lovino MA, Pierrestegui MJ, Müller OV, Berbery EH, Müller GV, Pasten M (2021) Evaluation of historical CMIP6 model simulations and future projections of temperature and precipitation in

Paraguay. Clim Change. https://doi.org/10.1007/s10584-021-03012-4
Lun Y, Liu L, Cheng L, Li X, Li H, Xu Z (2021) Assessment of GCMs simulation performance for precipitation and temperature from CMIP5 to CMIP6 over the Tibetan plateau. Int J Climatol 41:3994–4018. https://doi.org/10.1002/joc.7055
Meliho M, Braun M, Khattabi A, Orlando CA (2025) Selection of climate simulations for climate change impact studies: case study of the Souss watershed, Morocco. Theor Appl Climatol. https://doi.org/10.1007/s00704-025-05353-x
MoA (1998) Agro-ecological zones of Ethiopia. Natural resources management and regulatory department. With support of German Agency for Technical Cooperation (GTZ)
Mohammed JA, Gashaw T, Worku Tefera G, Dile YT, Worqlul AW, Addisu S (2022) Changes in observed rainfall and temperature extremes in the Upper Blue Nile Basin of Ethiopia. Weather Clim Extremes. https://doi.org/10.1016/j.wace.2022.100468
Papalexiou SM, Rajulapati CR, Clark MP, Lehner F (2020) Robustness of CMIP6 historical global mean temperature simulations: trends, long-term persistence, autocorrelation, and distributional shape. Earths Future. https://doi.org/10.1029/2020ef001667
Peng S, Wang C, Li Z, Mihara K, Kuramochi K, Toma Y, Hatano R (2023) Climate change multi-model projections in CMIP6 scenarios in Central Hokkaido, Japan. Sci Rep 13:230. https://doi.org/10.1038/s41598-022-27357-7
Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J (2007) Climate models and their evaluation. Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR). Cambridge University Press, pp 589–662
Rathjens H, Bieger K, Srinivasan R, Arnold JG (2016) CMhyd User Manual: Documentation for preparing simulated climate change data for hydrologic impact studies
Rettie FM, Gayler S, Weber TKD, Tesfaye K, Streck T (2023) High-resolution CMIP6 climate projections for Ethiopia using the gridded statistical downscaling method. Sci Data. https://doi.org/10.1038/s41597-023-02337-2
Rivera JA, Arnould G (2020) Evaluation of the ability of CMIP6 models to simulate precipitation over Southwestern South America: climatic features and long-term trends (1901–2014). Atmos Res. https://doi.org/10.1016/j.atmosres.2020.104953
Shah F, Sharifi A (2025) Climate models for predicting precipitation and temperature trends in cities: a systematic review. Sustain Cities Soc. https://doi.org/10.1016/j.scs.2025.106171
Shiru MS, Chung E-S (2021) Performance evaluation of CMIP6 global climate models for selecting models for climate projection over Nigeria. Theoret Appl Climatol 146:599–615. https://doi.org/10.1007/s00704-021-03746-2
Song YH, Chung ES, Ayugi BO (2025) CMIP6 GCMs projected future Koppen-Geiger climate zones on a global scale. Earths Future. https://doi.org/10.1029/2023ef004401
Tadase AT, Tekile AK (2025) Future projections of climate variables and meteorological drought: insight from CMIP6 models in Southeast Ethiopia. Dyn Atmos Oceans. https://doi.org/10.1016/j.dynatmoce.2025.101538
Tadesse TA, Tarkegn TG, Ray RL, Tefera GW, Demessie SF, Shawul AA, Worqlul AW, van Oel PR, Dile YT, Chukalla AD, Haileslassie A, Aniley E, Tagele AB, Bantider A (2025) Performances of reanalysis products in representing the temperature climatology of Ethiopia. Theor Appl Climatol. https://doi.org/10.1007/s00704-024-05332-8
Tanimu B, Hamed MM, Bello A-A, Abdullahi SA, Ajibike MA, Yaseen ZM, Alasow AA, Muhammad M, Shahid S (2025) CMIP5 and CMIP6: contrasting perspectives on Nigeria's climate transformation. Theor Appl Climatol. https://doi.org/10.1007/s00704-024-05272-3
Tesfaye TW, Dhanya CT, Gosain AK (2017) Spatiotemporal evaluation of reanalysis and in-situ surface air temperature over Ethiopia. Eur Water 59:215–221
Tesfaye E, Abate B, Alemayehu T, Dile Y (2023) Evaluating the skills of the CMIP5 global climate models using multicriteria decision-making analysis in Ethiopia. Heliyon. https://doi.org/10.1016/j.heliyon.2023.e20320
Waheed A, Jamal MH, Javed MF, Idlan Muhammad K (2024) A CMIP6 multi-model based analysis of potential climate change effects on watershed runoff using SWAT model: a case study of kunhar river basin, Pakistan. Heliyon 10:e28951. https://doi.org/10.1016/j.heliyon.2024.e28951
Wang L, Zhu J, Wang D (2025) Comparative analysis of high-resolution CMIP6 GCM and CMIP5 RCM: unveiling biases and advancements in simulating compound extreme events in China. Clim Dyn. https://doi.org/10.1007/s00382-024-07571-x
Woldemariam YA, Woldesenbet TA, Alamirew T (2025) Evaluation and projection of CMIP6 simulations of climate variables for the Rift Valley Lakes Basin, Ethiopia. Theor Appl Climatol. https://doi.org/10.1007/s00704-025-05356-8
Worku G, Teferi E, Bantider A, Dile YT, Taye MT (2018) Evaluation of regional climate models performance in simulating rainfall climatology of Jemma sub-basin, upper blue Nile Basin, Ethiopia. Dyn Atmos Oceans 83:53–63. https://doi.org/10.1016/j.dynatmoce.2018.06.002
Worku G, Teferi E, Bantider A, Dile YT (2021) Modelling hydrological processes under climate change scenarios in the Jemma sub-basin of upper Blue Nile Basin, Ethiopia. Clim Risk Manag. https://doi.org/10.1016/j.crm.2021.100272
Yang X, Zhou B, Xu Y, Han Z (2021) CMIP6 evaluation and projection of temperature and precipitation over China. Adv Atmos Sci 38:817–830. https://doi.org/10.1007/s00376-021-0351-4
Yang Y, Zhang Y, Gao Z, Pan Z, Zhang X (2023) Historical and projected changes in temperature extremes over China and the inconsistency between multimodel ensembles and individual models from CMIP5 and CMIP6. Earth Space Sci. https://doi.org/10.1029/2022ea002514
Zabihi O, Ahmadi A (2024) Multi-criteria evaluation of CMIP6 precipitation and temperature simulations over Iran. J Hydrol Reg Stud 52:101707. https://doi.org/10.1016/j.ejrh.2024.101707
Zambrano-Bigiarini M, Nauditt A, Birkel C, Verbist K, Ribbe L (2017) Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. Hydrol Earth Syst Sci 21:1295–1320. https://doi.org/10.5194/hess-21-1295-2017
Zhang X, Yang F (2004) RClimDex (1.0) user manual. Clim Res Branch Environ Can 22
Zhang Y, Liu H, Qi J, Feng P, Zhang X, Liu L, Marek GW, Srinivasan R, Chen Y (2023) Assessing impacts of global climate change on water and food security in the black soil region of Northeast China using an improved SWAT-CO(2) model. Sci Total Environ 857:159482. https://doi.org/10.1016/j.scitotenv.2022.159482

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Alehegn Belay Tagele[1] · Temesgen Gashaw Tarkegn[1,2,3] · Pieter R. van Oel[3] · Ram Lakhan Ray[2] · Gebrekidan Worku Tefera[2] · Sintayehu Fetene Demessie[4,12] · Alemayehu Abate Shawul[5] · Othoo Calvince Ouko[6] · Daniel Berhanu[7] · Amare Haileslassie[8] · Abeyou W. Worqlul[9] · Amare Bantider[7] · Yihun T. Dile[10] · Abebe D. Chukalla[11] · Tazebew Abera Tadesse[1] · Enyew Adgo[1]

✉ Temesgen Gashaw Tarkegn
gtemesgen114@gmail.com

1 Department of Natural Resource Management, College of Agriculture and Environmental Science, Bahir Dar University, Bahir Dar, Ethiopia

2 College of Agriculture, Food and Natural Resources, Prairie View A&M University, Prairie View, TX 77446, USA

3 Water Resources Management Group, Wageningen University & Research, P.O. Box 47, Wageningen 6700 AA, the Netherlands

4 College of Business, Technology and Vocational Education, Kotebe University of Education, P.O. Box 31248, Addis Ababa, Ethiopia

5 Department of Natural Resource Management, Madda Walabu University, Bale Robe, Ethiopia

6 Department of Community Development and Environment Management (DCEM), Co-operative University of Kenya, P.O.Box 24814 -00502, Karen, Nairobi, Kenya

7 Water and Land Resource Center, Addis Ababa University, Addis Ababa, Ethiopia

8 International Water Management Institute, Addis Ababa, Ethiopia

9 International Center for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia

10 College of Agriculture and Life Sciences, Texas A&M University, College Station, TX, USA

11 The Department of Land and Water Management, IHE Delft Institute for Water Education, Delft 2611 AX, The Netherlands

12 Africa Center of Excellence for Climate-Smart Agriculture and Biodiversity Conservation, Haramaya University, Dire Dawa, Ethiopia