## RESEARCH

**Open Access** 

# Evaluation of CORDEX Africa regional climate models performance in simulating climatology of Zarima sub-basin northwestern Ethiopia

Meaza Kassahun<sup>1,2\*</sup>, Kassahun Ture<sup>2</sup> and Dessie Nedaw<sup>2,3</sup>

## Abstract

Climate models are basic tools to obtain reliable estimates of future climate change and its effects on the water resources and agriculture in given basin. However, all climate models are not equally valuable for all areas. Therefore, determining the most appropriate climate model for a specific study area is essential. This study examines the performance of 10 CORDEX-AFRICA-22<sup>0</sup> Regional Climate Models (RCMs), three downscaling institutional based ensembles mean (Reg ensemble, CCLM ensemble and REMOO ensemble) and the multi-model ensemble mean. The models were evaluated based on their ability in replicating the seasonal and annual rainfall, minimum and maximum temperature and inter-annual variability for the period of 1986–2005 using statistical metrics such as BIAS, Root Mean Square Error (RMSE), Pearson correlation coefficient (r), coefficient of variation (CV), Kling Gupta Efficiency (KGE) and Taylor diagram. The findings indicated that HadREMOO, MPI-Reg4-7, HadReg4-7, Reg ensemble, and multi-model ensemble mean performed relatively better in representing the mean annual observed rainfall at the Adiramets, Debarik Ketema, Niguse Maystebri, and Zarima stations, respectively. Whereas, NorESM-CCLM, MPI-CCLM, NorESM-Reg4-7, and NorESM-REMOO exhibited a weak performance in reproducing the observed mean annual rainfall at the Adiramets, Debarik Ketema Niguse, Maystebri, and Zarima stations, respectively. Similarly, RCMs generally capture the mean annual maximum temperature of climatic stations of Zarima subbasin well. Specifically, the MPI-Req4-7 simulation performs well in representing the mean annual observed maximum temperature at Adiramets and Maytsebri stations, while the Debarik and Ketema Niguse stations exhibit superior performance in the HadReg4-7 simulation and the Zarima station shows better representation in the CCLM ensemble simulations. The majority of the model simulations exhibit good representation of mean annual minimum temperature at Adiramets, Debarik, and Zarima stations. Specifically, CanESM-RCM, HadReg4-7, REMOOensemble, multi-model ensemble, and Regensemble simulations perform better at Adiramets, Debarik, Ketema niguse, Maystebri and Zarima stations respectively. This suggests that these models may have biases or shortcomings in capturing the temperature values in the subbasin. Furthermore, NorESM-CCLM at Adiramets, Ketema niguse, and Zarima stations, NorESM-REMOO at Debarik station, and HadReg4-7 at Maystebri station demonstrate poor performance in representing the observed mean minimum temprature. Majority of the RCMs, all institutional based ensemble means and the multi-model ensemble mean simulations overestimate the observed mean annual rainfall of the Zarima subbasin with minimum bias of 0.02 mm at Ketema niguse HadReg4-7and maximum bias of 2.81 mm at Maytsebri MPI-CCLM simulation. Similarly, HadReg4-7 simulation of Ketama Niguse MPI-CCLM showed a minimum 0.02 mm and Maytsebri simulation kiremit season mean rainfall showed a maximum bias of and 2.99 mm. Regarding mean annual and kiremit season

\*Correspondence: Meaza Kassahun

meaza.kassahun@aau.edu.et

Full list of author information is available at the end of the article



© The Author(s) 2023. **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/.

maximum and minimum temperature of the Zarima subbasin were overestimated by majority of the simulation and the ensemble means. The correlation (r) of observed and model simulated mean annual and kiremit season rainfall was strong (0.60–0.79) and very strong (0.80–0.99) in the majority of the simulations except Ketema niguse station mean annual and kiremit season rainfall simulations of MPI-REMOO, NorESM-Reg4-7; Debarik station kiremit season rainfall of NorESM-CCLM and NorESM-REMOO, MPI-Reg4-7 and MPI-REMOO, which showed moderate correlation. The performance of the RCMs, institutional based ensemble means and multi-model ensemble mean were different in statistical metrics (BIAS, RMSE, r, CV and KGE) and Taylor diagram. Among the simulations and ensemble means, the multi-model ensemble mean was superiors in two or more of statistical metrics at each station of the Zarima subbasin except Maytsebri station kiremit season rainfall, where the CCLM ensemble was better. Consistently, the Taylor diagram showed that the multi-model ensemble was better in the replication of the areal annual and kiremit season rainfall, maximum and minimum temperature of the subbasin. This finding evidenced that selecting the best RCMs and ensemble mean is necessary for climate projection and climate change impact assessment study.

Keywords Climate models, CORDEX-Africa, Performance, Annual and seasonal

## Introduction

Climate models are fundamental tools used in projecting future climate conditions and developing strategies for adaptation and mitigation (Giorgi et al. 2009; Endris et al. 2017; Luhunga et al. 2018). Global Climate Models (GCMs) are widely used for climate projections; however, their low spatial resolution can limit their effectiveness in capturing small-scale climate variations influenced by topography and land surface heterogeneity (Dosio and Panitz 2016; Yimer et al. 2022).

Regional Climate Models (RCMs) are developed with higher spatial resolution compared to GCMs, making them better suited for simulating climate in coastal, mountainous regions, and at regional scales. RCMs are capable of providing more detailed and accurate information about local climate conditions (Giorgi et al 2009; Endris et al. 2013; Dosio and Panitz 2016).

Studies have shown that RCMs outperform GCMs in simulating climate in coastal and mountainous areas, as well as at regional levels. This is due to their ability to capture localized features such as land-sea interactions, complex terrain, and mesoscale atmospheric processes (Endris et al. 2013; ARSET 2022). The higher spatial resolution of RCMs makes them more suitable for endusers, such as water resource managers and agricultural planners, who require predictions of future regional and local climate changes. These predictions are crucial for conducting climate change impact assessments on water resources and agriculture at regional and local scales (Endris et al. 2013; Kassie et al. 2014; Matiu et al. 2020). By providing more detailed and localized information, RCMs can improve the accuracy of climate change impact assessments and aid in the development of targeted adaptation and mitigation strategies (Ilori and Balogun 2022; Demessie et al. 2023).

Currently, the Coordinated Regional Climate Downscaling Experiment (CORDEX) program sponsored by the World Climate Research Programs aims to fill this gap by coordinating international efforts towards regional climate downscaling (Giorgi et al. 2009; Gutowski and Solman 2019). The CORDEX provides dynamically downscaled regional climate information for the climate modeling community and climate information to end users (Hernández-Díaz et al. 2013; Nikulin et al. 2012).

The CORDEX Africa is a CORDEX domain, experimentally developed specifically for climate impact studies in Africa, which is especially vulnerable to climate change due to limited adaptive capacity (Giorgi et al. 2009). Phase two CORDEX-AFRICA 0.22 has 10 Regional Climate Models (RCMs) provides valuable climate model data in countries like Ethiopia, where data scarcity exists. The CORDEX- AFRICA RCMs are particularly useful for climate projections and impact assessments due to their ability to generate downscaled climate information at regional scales. To gain confidence in evaluating how climate change affects water resources and agriculture in regions with complex topography like Northwestern part of Ethiopia best RCMs should be selected (Yimer et al. 2022). However, before selecting RCMs for climate change projections and impact assessments it is essential to evaluate the RCMs and the ensemble mean performance in replicating the historical climate of area is important (Endris et al. 2013; Worku et al. 2018; Kamworapan and Surussavadee 2019; Demissie and Sime 2021).

The Zarima subbasin in Northwestern part Ethiopia is characterized by its complex topography and is home to the Zarima river, which serves as a crucial water source for water supply and traditional irrigation systems in the region. This river, along with its tributaries, plays a vital role in supporting agricultural activities and meeting the water needs of the local communities. However, the subbasin faces challenges related to water management and competition for irrigation water. Without proper water management practices in place, there may be conflicts and unsustainable use of water resources. The lack of coordinated water management can lead to inefficient allocation and utilization of water, potentially impacting agricultural productivity and livelihoods (Lemma et al. 2010).

Additionally, there is construction of an irrigation project covering approximately 9500 hectares of land in the Zarima subbasin is a significant development. The dam can store a total of 3.6 billion cubic meters of water, the project aims to support irrigation for a sugarcane plantation spanning 50,000 hectares (Weldesadik 2021). This initiative has the potential to enhance local food security, contribute to the country's economy by creating employment opportunities, and facilitate the export of industrial products, particularly from the sugarcane industry. To sustain this benefit addressing the water management challenges such as land use land cover and climate change in the Zarima subbasin is crucial for sustaining the availability of water resources, supporting agriculture, and promoting the well-being of the local communities that rely on the Zarima river and its tributaries. Thus, evaluating the performance of CORDEX-Africa RCMs, as well as the ensemble mean of these models, in replicating historical rainfall and temperature patterns in the Zarima subbasin is a crucial step. By assessing how well the RCMs and the ensemble mean replicate historical climate data, researchers can gain insights into the models' accuracy and reliability for future climate projections. This evaluation provides valuable information for selecting appropriate RCMs to inform water resource management and climate change adaptation strategies in the Zarima subbasin. The main objective of the study is to evaluate the performance of available CORDEX-Africa RCMs and the ensemble means in replicating historical rainfall and temperature patterns in the Zarima subbasin. This evaluation is an important step in ensuring the reliability and applicability of the selected RCMs for climate change projection and impact assessment studies in the region.

## Materials and method

## Description of the study area

The study was carried out at Zarima subbasin, which is part of the Tekeze river basin and where Zarima river flows. According to information obtained from the Ministry of Water and Energy (MoWE), the subbasin is located in the northwest highlands of Ethiopia, 804.9 km from Addis Ababa, and has a surface area of 663,157.58 ha, Geographically, it is located between latitudes 37<sup>0</sup>30'-38<sup>0</sup>27'E and 13<sup>0</sup>15'-14<sup>0</sup>13'N latitude, with an elevation between 736 and 4399 m a.s.l (Fig. 1). The sub basin's topography is characterized by steep, undulating hills and a narrow gorge that includes a portion of Semen Mountain (Weldesadik 2021).



Fig. 1 Location map of Zarima subbasin

The subbasin is characterized by four agroclimatic zones; 74.83% of the subbasin is covered by the lowland agroclimatic zone, 18.96% the midland, 4.09% highland, and 2.12% and upper highland agroclimatic zone (Fig. 2a) (Zegeye et al. 2022).

The major land use types of the subbasin are agricultural land, forestland, shrubland, grassland, bare land and water bodies. Most of the lower portions are lowland areas, owing to the increased demand for agricultural, while forests and shrubland are found in the upper portions of the basin. 57.81% of the subbasin is covered by agricultural land, 23.81% of shrubland and 16.26% covered by forest land (Fig. 2b).

Agriculture is one of the most basic and important economic activities that sustains the subbasin's people's livelihood. Irrigated agriculture, commercial farming, agropastoral, gum/incense collection methods, and cereal cropping are examples. Mixed agricultural techniques, which involve crop and livestock production, are widespread in the subbasin. Smallholder farmers harvest grain for sustenance with a traditional ox-drawn rainfed plough (Mowie 2009).

Based on the rainfall amount Ethiopia has three local seasons namely Bega, Belg and Kiremit. Bega is the dry season that covers period from October to January. The Bega season is characterized by hot, dry days and cool nights. Frosty in early mornings in the majority of the highland areas (NMA 2015). Belg is the small rainy season in Ethiopia except southern and southeastern lowlands areas. It covers the period from February to May. High variability of rainfall in time and space and high maximum temperature are common characters of Belg season (Ven Chow et al. 1988). It's the warmest season as March, April and May months are the warmest months of the year (NMA 2015). Kiremt is the main rainy season which contributed 85-95% of the annual rainfall and food crop production of the country (Ven Chow 1988). It spans from June to September with frequent rains and homogeneous temperatures in July and August (NMA 2015).

The subbasin received annual rainfall ranging from 1051 mm at the low land areas to 1863.72 mm at the high land area from 1984 to 2018. The average annual temperature was 13.7 °C in the highlands and 20.4 °C around the lowlands (Zegeye et al. 2022). The Sub-basin



Fig. 2 Agroclimatic zone (a) and land use land cover (b) of Zarima subbasin



from the Ethiopian Meteorological Institute for the period of 1984 to 2018

receives approximately 80 to 92% of its annual rainfall during the rainy season (Kiremit season) Fig. 3 (Daba 2014).

## Source of data

## Observed climate data

The gridded daily observational rainfall and temperature data sets for the Zarima subbasin was obtained for the period of 1984–2018 were obtained from the Ethiopian Meteorological Institute (EMI). To compare the observed and model simulated climate data for the Maytsebri, Adiramets, Debarik, Ketema Niguse, and Zarima stations of the Zarima subbasin for the time period of 1986–2005 the station data were extracted using the R statistical package Climate Data Tools (CDT).

## Model climate data

The CORDEX Africa 0.22<sup>0</sup> (CORDEX -CORE) data consists of 10 RCMs (Table1). The historical simulations rainfall and temperature data of 10 RCMs for the time period 1986–2005 were downloaded freely from Earth System Grid Federation (ESGF) (https://esg-dn1.

#### Table 1 List of stations and their location

nsc.liu.se/search/cordex). The downloaded CORDEX Africa was extracted using the R statistical software Climate data tool (CDT) package for the Maytsebri, Adiramets, Debarik, Ketema niguse and Zarima stations of Zarima subbasin.

#### Data analysis

There is no single criterion used to identify the best RCMs. In this study, combined performance metrics Bias (BIAS)(Eq. 2), RMSE (Eq. 2) and Pearson's correlation (r) (Eq. 3), coefficient of variation (CV) (Eq. 4), Kling-Gupta Efficiency (KGE) (Eq. 5) and Taylor diagram were used to evaluate the performance of the CORDEX-AFRICA-RCMs, the institutional based ensembles mean (Reg ensemble, CCLM ensemble and REMOO ensemble) and the multi-model ensemble mean (grand ensemble) in replicating kiremit season and annual climate of the subbasin. BIAS, RMSE and r are commonly used in multiple studies to evaluate the performance of CORDEX-AFRICA- RCMs (Dibaba et al. 2019; Ayugi et al. 2020; Mendez et al. 2020).

The BIAS measures the systematic error between the observed and simulated climate variables and zero indicates good performance, while values away from zero show the deviations from observed data. RMSE measures how accurately climate models simulate climate variables. There is no common acceptable value for BIAS and RMSE. Smaller values of RMSE close to zero had good model performance and vice versa (Gleckler et al. 2008; Chai and Draxler 2014). On the other hand, the correlation coefficient (r) values showed the linear relationship between observed and simulated by RCMs. Correlation coefficient values can range from -1 for a perfect negative correlation to 1 for a perfect positive correlation between the modeled (RCMs) and the observed climate variables (Schober et al. 2018). Furtherer more, the correlation strength was evaluated using Evans (1996) suggests that the value of r (0.00-0.19=very weak), (0.20 - 0.39 = weak),(0.40 - 0.59 = moderate),(0.60 -0.79 = strong), and (0.80 - 1.00 = very strong).

No	Stations	Latitude	Longitude	Elevation	Data availability	Data type
						Rainfall & temperature
1	Maytsebri	15.056	39.51691	1689	1984–2018	Yes
2	Adiramets	13.7439	37.32444	2087	1984-2018	Yes
3	Debarik	13.1585	37.89904	2854	1984-2018	Yes
4	Ketema niguse	13.3929	37.40125	2844	1984-2018	Yes
5	Zarima	13.3408	37.88243	1295	1984-2018	Yes

$$BIAS = \frac{1}{n} \sum_{i=1}^{n} (S_i - O_i)$$
(1)

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

$$r = \frac{\sum_{i=1}^{n} (S_i - S_m) (O_i - O_m)}{\sqrt{\sum_{i=1}^{n} (S_i - S_m)^2} \sqrt{\sum_{i=1}^{n} (O_i - O_m)^2}}$$
(3)

$$CV = \frac{SD}{\overline{X}}$$
(4)

where,  $S_i$  is simulated rainfall or temperature,  $O_i$  is observed rainfall or temperature,  $O_m$ . is observed mean of rainfall or temperature,  $S_m$  = simulated value.

In this study, CV is used to classify the degree of variability as less (CV < 20%), moderate (20 < CV < 30%), high (CV > 30%), very high (CV > 40%) and CV > 70% indicate extremely high inter annual variability of the rainfall and temperature data which was applied by Eshetu, (2020).

Kling-Gupta efficiency (KGE) is a performance metric which was developed as an upgrade to the commonly used Nash–Sutcliffe efficiency, taking into account multiple types of model errors, namely mean, variability, and correlation. It was introduced by Gupta et al. (2009) and modified by Kling et al. (2012) and is defined as Eq. 5.

There are three major components involved in the calculation of this KGE index:

- 1. The Pearson product-moment correlation coefficient, denoted as r. The ideal value is r = 1.
- 2. Beta ( $\beta$ ): the ratio of the mean simulated ( $\mu_s$ ) values to the mean observed values ( $\mu_o$ ) Beta = 1 is the ideal value.
- 3. Alpha ( $\alpha$ ): variability ratio, which is coefficient of variation of simulated ( $CV_S$ ) and observed ( $CV_O$ )value

$$KGE = \dot{I} - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (r - 1)^2}$$
(5)  
where  $\beta = \frac{\mu_s}{\mu_o}$  and  $\alpha = \frac{CV_s}{CV_o}$ .

KGE value ranges between negative infinity and one. Model performance to be poor for 0.5 > KGE > 0 (Vrugt and de Oliveira 2022).

In addition to statistical metrics, monthly and annual plots of observed versus simulated rainfall, maximum and minimum temperature were used.

Cumulative distribution function (CDF) was used to evaluate the areal rainfall distribution. Number

of days for dry (RR=0), light (RR<10 mm), heavy (RR > = 10 mm) and very heavy rainfall (RR > = 20 mm)were used to compare the observed with RCMs, institutional based ensemble means and over all ensemble mean as applied by Demissie and Sime 2021; Demessie et al. 2023. Further the Taylor diagram was used to evaluate the areal kiremit season and annual rainfall of the subbasin. Taylor diagram is a diagrammatical representation of statistical methods: Pearson correlation coefficients (r), standard deviation (SD), and centered RMS difference (Taylor 2001). It was used to evaluate the performance and consistency of each CORDEX-AFRICA-RCMs, downscaling institutional based ensembles mean and multi-model ensemble mean in replicating the mean annual rainfall of each station in the Zarima subbasin. It is used by Ilori and Balogun, (2022) and Demessie et al., (2023), for the graphical synthesis of the extent of the agreement between simulation and observation data.

### **Results and discussion**

## Performance of RCMs in simulating rainfall in different temporal scale

Figures 4a-e depict the long-term (1986-2005) mean monthly rainfall cycle of the observed, RCMs, the institutional based ensembles mean (Reg ensemble, CCLM ensemble and REMOO ensemble) and the multi-model ensemble mean (ensemble in figures) each station of Zarima subbasin. The simulation of a monomodal rainfall from May to September. Some models overestimated rainfall in dry months while underestimating it in wet months. The multi-model ensemble mean and CanESM-RCM monthly rainfall simulations performed well over most subbasins, except for the Ketema niguse station. At the Ketema niguse station, the REMOO ensemble and Reg ensemble showed better performance. Additionally, at the Debarik and Maytsebri stations, the NorESM-Reg4-7 simulation was good. At the Adiramets station, the CCLM ensemble performed well. Generally, RCMs, institutional based ensembles mean and the multi-model ensemble mean showed different pattern at each station of the subbasin but the multi-model ensemble was better in replicating the observed rainfall of the subbasin (Figs. 4a-e). This finding implied that different simulations or ensembles performed better at different stations, highlighting the spatial variability in their performance. Similarly, Some RCMs models overestimated the rainfall in the dry months and underestimated it in the wet months in south West Ethiopia (Demissie and Sime 2021).

The long-term (1986–2005) mean annual of the observed, RCMs, the institutional based ensembles mean



Mean 1 Mean 1 Jun Months Jan Feb Mar Apı Ma Jul Aug Sep Oct Not Dec ---- Observed HadReg4-7 -HadCCLM HadREMOO MPI-Reg4-7 MPI-REMOO NorESM-Reg4-7 ----NorESM-CCLM -CanESM-RCM - Ensemble

Fig. 4 Monthly plot of mean annual rainfall Maystebri (a), Adiramets (b), Debarik (c), Ketema Niguse (d) and Zarima (e) stations



Table 2	The list of	CORDEX-Africa	RCMs and	their	driving	GCMs

Institutes	Driven GCM Model	RCMs	RCM abbreviation	Abbrevations used in the study
Met Office Hadley center	MOHC-HadGEM2	CLMcom-KIT-CCLM5	HadGEM2-CCLM5-0-15	HadCCLM
	MPI-M-MPI-ESM-LR	CLMcom-KIT-CCLM5	MPI-ESM-LR-CCLM5-0-15	MPI-CCLM
	NCC-NorESM1-M	CLMcom-KIT-CCLM5	NorESM1-M -CCLM5-0-15	NorCCLM
Max Planck Institute for meteorology (Germany)	MOHC-HadGEM2	GERICS-REMO2015	HadGEM2-REMO2015	HadREMOO
	MPI-M-MPI-ESM-LR	GERICS-REMO2015	MPI-ESM-LR-REMO2015	MPI-REMOO
	NCC-NorESM1-M	GERICS-REMO2015	NorESM1-M-REMO-2015	NorREMOO
Norwegian Climate Center (Norway)	MOHC-HadGEM2	ICTP-RegCM4-7	NorESM1-M-REGCM4-7	HadReg4-7
	MPI-M-MPI-ESM-LR	ICTP-RegCM4-7	MPI-ESM-LR-REGCM4-7	MPI-Reg4-7
	NCC-NorESM1-M	ICTP-RegCM4-7	NorESM1-M-REGCM4-7	NorReg4-7
Canadian Centre for Climate Modelling and Analy- sis (Canada)	Cccma-CanESM2	CCCMA-CanESM-RCM4	CanESM2-CanESM-RCM4	CanESM-RCM

Source: https://esg-dn1.nsc.liu.se/search/cordex

(Reg ensemble, CCLM ensemble and REMOO ensemble) and the multi-model ensemble mean of Zarima subbasin is presented in Fig. 5. The findings indicated that HadREMOO, MPI-Reg4-7, HadReg4-7, Reg ensemble, and multi-model ensemble showed a stronger capability in capturing the observed mean annual rainfall at the Adiramets, Debarik, Ketema Niguse, Maystebri, and Zarima stations respectively. On the contrary, the RCMs NorESM-RCLM, MPI-CCLM, NorESM-Reg4-7, and NorESM-REMOO exhibited a weak performance in reproducing the observed mean annual rainfall at the Adiramets, Debarik, Ketema Niguse, Maystebri and Zarima stations respectively. These findings implied that certain RCMs perform well in representing mean annual

rainfall at specific stations while exhibiting poorer performance at others (Table 2).

The statistical evaluations of the mean annual rainfall are shown in Table 3. The result revealed that the majority of the RCMs, all institutional based ensembles mean and the multi-model ensemble mean simulations overestimate the observed mean annual rainfall, which ranged between 0.02 and 2.81 mm. The lowest and highest overestimation were observed in HadReg4-7 and MPI-CCLM simulations at Ketema niguse and Maytsebri stations respectively (Table3). While an underestimation of mean annual rainfall was observed in NorESM-Reg4-7 and HadReg4-7 simulations at Ketema niguse and Zarima stations respectively (Table 3). The underestimation value was ranged range of 0.02–17.93 mm. Similarly,

able 3	Statistica	I evaluation	or the mear	i annual ar	id kiremit se		iali Ui Zari	ma subbasir	i (the statis	stical evalu	ation oi ki		on mean r	aintail is W	ritten in pr	ackets
tations	Stastical metrics	Observed	HadReg4-7	HadCCLM	HadREMOO	MPI- Reg4-7	MPI- CCLM	MPI- REMOO	NorESM- Reg4-7	NorESM- CCLM	NorESM- REMOO	CanESM- RCM	Reg ensemble	CCLM ensemble	REMOO ensemble	Ensemble
diramets	BIAS		0.53 (0.22)	- 0.51 (- 0.94)	- 1.28 (- 1.49)	— 0.45 (— 0.71)	- 1.38 (- 2.00)	- 1.86 (- 2.04)	- 1.5 (- 1.93)	— 1.08 (— 1.46)	— 2.01 (— 2.12)	<b>-</b> 1.90 ( <b>-</b> 2.11)	- 0.47 (- 0.81)	- 1.15 (- 1.48)	<b>-</b> 1.25 ( <b>-</b> 1.65)	— 1.14 (— 1.46)
	RMSE		247.54 (238.61)	263.54 (245.12)	291.39 (292.22)	332.77 (322.75)	261.17 (309.86)	386.65 (398.50)	368.39 (376.85)	310.37 (309.20)	426.25 (432.90)	322.58 (350.77)	180.88 <b>(194.71)</b>	197.14 (215.08)	208.82 (226.11)	<b>180.16</b> (215.77)
	-		0.76 (0.77)	0.68 (0.75)	0.72 (0.73)	0.61 (0.65)	0.82 (0.78)	0.73 (0.70)	0.76 (0.74)	0.64 (0.67)	0.71 (0.72)	0.78 (0.73)	0.87 (0.86)	0.88 (0.87)	0.87 (0.87)	0.89 (0.88)
	S	16.74 (17.84)	17.22 (18.09)	14.93 (15.83)	18.82 (20.42)	19.85 (21.62)	20.36 (22.67)	26.94 (28.21)	27.24 (28.39)	18.12 (19.39)	29.03 (31.87)	22.61 (24.74)	17.47 (18.79)	15.67 (16.99)	16.10 (17.50)	9.14 (9.14)
	KGE		0.76 (0.77)	0.66 (0.72)	0.69 (0.69)	0.57 (0.59)	0.71 (0.64)	0.33 (0.34)	0.33 (0.35)	0.63 (0.66)	0.21 (0.16)	0.57 (0.52)	0.86 <b>(0.85)</b>	0.85 <b>(0.85)</b>	0.85 (0.85)	0.87 (0.85)
Debarik	BIAS		0.38 (- 0.15)	0.45 (- 0.47)	<b>-</b> 0.12 (0.08)	0.07 (0.03)	1.22 (0.60)	<b>-</b> 0.54 (-0.99)	— 0.10 (-0.21)	0.4 (- 0.21)	<b>-</b> 0.48 ( <b>-</b> 0.61)	<b>-</b> 0.08 ( <b>-</b> 0.23)	0.12 (- 0.11)	0.69 ( <b>-</b> 0.33)	- 0.38 (- 0.51)	0.12 (- 0.22)
	RMSE		167.39 (143.57)	139.47 (101.92)	145.73 (142.63)	146.93 (129.49)	136.59 (103.52)	189.87 (168.26)	160.95 (136.74)	186.88 (167.61)	206.89 (205.75)	129.11 (127.55)	127.76 (103.34)	103.59 (83.73)	142.11 (139.50)	84.11 (77.43)
	-		(090) 06.0	0.91 (0.77)	0.87 (0.64)	0.90 (0.56)	0.95 <b>(0.78)</b>	0.86 (0.51)	0.88 (0.60)	0.89 (0.44)	0.85 (0.48)	0.93 (0.72)	0.92 ( 0.73)	0.94 (0.84)	0.91 (0.63)	0.96 (0.87)
	S	12.02 (12.23)	<b>11.93</b> (13.63)	10.24 (10.99)	13.58 (14.33)	9.10 (9.56)	11.95 (10.89)	13.07 (14.10)	12.80 (12.49)	14.21 (13.44)	17.19 (18.73)	13.49 (14.68)	9.27 (9.57)	9.35 (9.11)	12.17 (13.48)	8.56 (13.44)
	KGE		0.89 (13.58)	0.83 (0.75)	0.81 (0.60)	0.74 (0.51)	0.92 (0.71)	0.83 (0.48)	0.86 (0.60)	0.79 (0.44)	0.54 (0.26)	0.86 (0.66)	0.76 (0.66)	0.77 (0.70)	<b>0.91</b> (0.61)	0.71 <b>(0.72)</b>
Ketema	BIAS		0.02 (0.02)	0.18 (1.57)	0.17 ( <b>—</b> 2.02)	0.20 (1.01)	0.20 (1.56)	0.11 ( <b>—</b> 2.24)	-0.02 ( <b>-</b> 2.53)	0.17 (0.92)	0.04 (- 0.27)	0.09 ( <b>-</b> 0.62)	0.07 (- 0.50)	0.18 (1.35)	0.11 (- 1.60)	0.12 (— 0.26)
niguse	RMSE		156.75 (135.76)	202.90 (180.83)	224.79 (201.65)	180.50 (159.84)	238.43 (149.41)	281.55 (256.70)	282.94 (259.73)	199.98 (145.27)	145.69 (132.76)	113.84 (119.35)	138.47 (118.40)	176.39 (114.90)	175.21 (152.87)	98.15 (94.13)
	-		0.69 (0.71)	0.73 (0.62)	0.60(0.60)	0.68 (0.66)	0.73 (0.76)	0.41 (0.45)	0.48 (0.46)	0.72 (0.74)	0.72 (0.73)	0.84 (0.78)	0.75 (0.77)	0.84 <b>(0.86)</b>	0.71 (0.73)	0.88 (0.86)
	S	16.11 (16.34)	16.13 (16.18)	16.19 (16.18)	21.20 (20.56)	16.01 (15.99)	18.36 (15.87)	22.77 (24.73)	23.69 (24.78)	18.50 (17.18)	14.39 (16.18)	15.33 (15.01)	14.89 (14.43)	15.39 (13.49)	<b>16.12</b> (16.34)	13.51 (13.44)
	KGE		0.69 (0.71)	0.71 (0.61)	0.48 (0.52)	0.68 (0.65)	0.67 (0.75)	0.27 (0.24)	0.29 (0.25)	0.68 (0.73)	0.72 (0.72)	0.83 (0.76)	0.74 (0.74)	0.80 <b>(0.77)</b>	0.70 (0.72)	0.80 (0.77)
Maytsebri	BIAS		0.04 (0.04)	2.28 (2.29)	<b>—</b> 1.73 (2.47)	1.34 (2.91)	2.81 (2.99)	<b>-</b> 2.30(1.29)	-2.63 ( <b>-</b> 0.79)	1.78 (2.25)	<b>-</b> 0.56 (0.21)	<b>-</b> 0.44 (0.83)	<b>—</b> 0.42 (0.72)	2.29 (2.51)	<b>-</b> 1.64 (1.32)	0.06 (1.46)
	RMSE		246.86 (213.73)	290.77 (243.59)	328.52 (303.37)	271.66 (231.16)	260.65 (248.68)	320.84 (280.46)	153.11 (160.84)	229.88 (191.54)	202.34 (181.56)	184.67 (166.26)	176.86 (158.97)	218.55 (185.55)	251.72 (220.37)	181.65 (154.47)
	L		0.73 (0.70)	0.64 (0.62)	0.71 (0.63)	0.70 (0.71)	0.75 (0.72)	0.70 (0.65)	0.70 (0.63)	0.68 (0.65)	0.71 (0.63)	0.72 (0.67)	0.76 (0.730	0.84 (0.81)	0.76 (0.69)	0.83 (0.80)
	S	14.82 (14.12)	32.66 (29.87)	23.86 (22.62)	30.27 (30.30)	19.63 (18.69)	19.89 (21.33)	34.12 (32.90)	21.88 (21.10)	16.07 (14.39)	26.49 (24.46)	21.08 (21.02)	23.02 (22.54)	16.27 (16.69)	28.32 (26.94)	20.72 (20.22)
	KGE		<b>-</b> 0.23 ( <b>-</b> 0.16)	0.27 (0.27)	- 0.1 (- 0.21)	0.52 (0.53)	0.53 (0.39)	- 0.34 (- 0.38)	0.44 (0.38)	0.63 (0.63)	0.16 (0.18)	0.49 (0.41)	0.39 (0.34)	0.74 (0.70)	0.05 <b>(0.04)</b>	0.55 (0.51)
Zarima	BIAS		<b>-</b> 0.08 ( <b>-</b> 0.45)	- 0.96 (1.75)	- 16.27 (- 0.41)	— 6.97 (— 0.62)	0.16 (- 0.51)	- 16.01 (0.13)	" <b>—</b> 14.87 (0.95)	- 3.08 (-1.54)	<b>-</b> 17.93 (0.44)	1.30 (1.08)	0.25 ( <b>- 0.04)</b>	- 0.92 (- 1.27)	0.25 (0.05)	<b>0.00</b> (- 0.27)

	Stastical metrics	Observed	HadReg4-7	HadCCLM	HadREMOO	MPI- Reg4-7	MPI- CCLM	MPI- REMOO	NorESM- Reg4-7	NorESM- CCLM	NorESM- REMOO	CanESM- RCM	Reg ensemble	CCLM ensemble	REMOO ensemble	Ensemble
	RMSE		0.68 (258.79)	<b>-</b> 0.27 (248.25)	-0.15 (254.44)	<b>-</b> 0.23 (280.36)	<b>0.00</b> (252.91)	- 0.03 (312.57)	— 0.09 (313.16)	— 0.18 (263.13)	<b>-</b> 0.06 (249.70)	0.80 (320.82)	0.86 (177.00)	0.93 (172.33)	0.84 (184.37)	0.95 (102.05)
	L		0.68 (0.72)	0.69 (0.73)	0.69 (0.71)	0.79 (0.78)	0.74 (0.75)	0.76 (0.76)	0.72 (0.72)	0.77 (0.78)	0.67 (0.67)	0.80 (0.76)	0.86 (0.87)	0.93 (0.95)	0.84 (0.85)	0.95 (0.96)
-	C	16.74 (17.84)	18.79 <b>(17.75)</b>	14.74 (16.24)	55.16 (16.82)	16.37 (23.55)	14.93 (18.73)	21.36 (25.04	48.91 (21.29)	14.75 (18.43)	89.76 (18.73)	22.17 (23.13)	17.80 (17.82)	13.87 (14.17)	<b>17.03</b> (17.03)	15.26 (15.35)
	KGE		0.66 (0.72)	0.66 (0.71)	<b>-</b> 1.43 (0.71)	0.62 (0.61)	0.72 (0.74)	0.18 (0.53)	— 1.06 (0.65)	0.71 (0.77)	<b>-</b> 3.45 (0.66)	0.61 (0.61)	0.84 (0.87)	0.81 (0.78)	0.84 (0.84)	0.90 (0.86)

Hernández-Díaz et al. (2013), reported that a dynamically downscaled version of CORDEX-AFRICA showed over estimation of the annual rainfall in the Ethiopian highlands and elevated area of Sudan; Ayugi et al. (2020) evaluated RCMs in East Africa and found that most of the models and ensembles overestimated the basin-average annual rainfall amount. In another study by Otieno and Anyah (2013) focused on the Great Horn of Africa, similar overestimations of basin-average annual rainfall were observed by the RCMs. Yimer et al. (2022) conducted an assessment of RCMs in the East Africa region and the Ethiopian highlands area. Their findings also indicated that the RCMs tended to overestimate rainfall in these regions. Additionally, Demessie et al. (2023) specifically examined the Guder subbasin of the upper Blue Nile basin. Their study revealed that the mean annual rainfall in this subbasin was overestimated by the CORDEX-AFRICA-RCMs.

The RCMs, institutional based ensembles mean and multi-model ensemble mean (ensemble in tables) of kiremit season rainfall of Zarima subbasin showed both over estimation and under estimation. The overestimation magnitude deviated from 0.02 to 2.99 mm. The lowest and highest over estimation was seen in HadReg4-7 and MPI-CCLM simulations at Ketema niguse and Maytsebri stations respectively. whereas, the underestimation was ranged from 0.15 to 2.53 mm, which was observed at HadReg4-7 simulation of at Debarik station and NorESM-Reg4-7 at Ketema niguse stations respectively (Table 3). In terms of bias HadReg4-7 was better at Adiramets, Ketema niguse and Maytseri stations; MPI-CCLM and CCLM ensemble were better at Debarik station and the multi-model ensemble mean at Zarima station (Table3).

The RMSE value indicated the error occurred between the observed and the model simulated mean annual rainfall reached up to 426.25 mm (Table 3). The lowest value was almost zero, which was observed at Zarima station MPI-CCLM, while the highest RMSE was 426.25 mm, which was seen at Ketema Niguse station NorESM-REMOO simulation. The multi-model ensemble mean showed lower RMSE value than the individual RCMs in Debarik and Ketema Niguse, whereas Reg ensemble, NorESMReg4-7 and MPI-CLLM showed lower RMSE values at Adiramets, Maytsebri and Zarima stations respectively. The kiremit season rainfall RMSE value ranged between 77.43 mm and 432.90 mm (Table 3). The lowest and highest values were observed at Debarik station multi-model ensemble mean simulation and Adiramets station NorESM-REMOO simulation respectively. In terms of RMSE, the multi-model ensemble mean was better in all of the stations of the subbasin except the Adiramets station where Reg ensemble mean was better (Table 3). This finding is in line with a study conducted in Eastern Africa where, the ensemble mean showed lower RMSE than other RCMs (Endris et al. 2013). The ensemble mean showed less RMSE values than the individual RCMs in most stations in southwest Ethiopia (Demissie and Sime 2021).

The correlation (r) of observed and model simulated mean annual and kiremit season rainfall was strong (0.60-0.79) and very strong (0.80-0.99) in the majority of the simulations except Ketema Niguse station mean annual and kiremit season rainfall simulations of MPI-REMOO, NorESM-Reg4-7; Debarik station kiremit season rainfall of NorESM-CCLM and NorESM-REMOO, MPI-Reg4-7 and MPI-REMOO, which showed moderate correlation (Table 3). This result coincides with a study conducted in Western Africa where the correlation coefficient value indicated that RCMs simulations of the annual rainfall well matched with the observed mean annual rainfall (Ilori and Balogun 2022). The multi-model ensemble mean had superior correlation coefficient values than the individual RCMs and institutional based ensembles mean throughout the subbasin at seasonal and the annual time scale except the Maytsebri station annual rainfall where MPI-CCLM was better (Table 3). Similarly, a study conducted in Eastern Africa RCMs (Endris et al. 2013) and Jemma subbasin (Worku et al. 2018), found out that the multi-model ensemble mean had relatively higher correlation than other RCMs.

The observed and mean annual and kiremit season rainfall showed less inter annual variability (CV<20%) over the subbasin. The RCMs, the institutional based ensembles mean and the multi-model ensemble mean annual rainfall showed low (CV=less than 20%) to extremely high (CV > 70%) variability (Table 3). The mean annual rainfall variability of the REMOO ensemble mean was closer to observed rainfall at Adiramets, Debarik and Ketema niguse stations, whereas at Maystebri station the CCLM ensemble and Zarima station Reg ensemble simulations showed closer variability than others. The kiremit season rainfall variability of REMOO ensemble, NorReg4-7, NorESM-CCLM and HadReg4-7 simulations showed closer seasonal rainfall variability at Adiramets and Ketema Niguse stations, Debarik, Maytsebri and Zarima stations respectively (Table 3). Generally, the ensemble means showed better results than the individual RCMs at annual time scale. Similarly, the ensemble was better in simulating the annual rainfall variability (Endris et al. 2013; Alemseged and Tom 2015; Worku et al. 2018; Dibaba et al. 2019; Demissie and Sime 2021).

Kling-Gupta efficiency (KGE) of the mean annual (Table 3) and mean kiremit season (Table 4) rainfall results revealed that the majority of the simulations had good performance (KGE>0.50). The multi-model

<b>ble 4</b> ason st	Statistical e atical evalu	evaluation - lation are v	of RCMs, ins vritten in bra	stitutional <mark>k</mark> acket)	based ensem	sam saldr	an and m	ulti-model	ensemble	e mean of	the annua	l and kirer	mit season	maximum	temperat	ure (kiremit
tions	Stastical metrics	Observed	HadReg4-7	HadCCLM	HadREMOO	MPI- Reg4-7	MPI- CCLM	MPI- REMOO	NorESM- Reg4-7	NorESM- CCLM	NorESM- REMOO	CanESM- RCM	Reg ensemble	CCLM ensemble	REMOO ensemble	Ensemble
ramets	BIAS		2.16 (0.27)	0.85 ( <b>- 0.02)</b>	0.71 (0.64)	<b>- 0.10</b> (0.12)	0.14 (0.12)	1.09 (0.34)	1.41 (0.27)	0.24 (0.20)	1.80 (0.25)	1.37 (0.31)	0.14 (0.22)	0.44 (0.10)	1.20 (0.41)	0.13 (0.25)
	RMSE		2.77 (0.49)	1.16 (0.88)	0.94 (1.10)	0.29 (1.55)	0.39 (0.49)	1.50 (0.57)	1.88 (0.51)	0.43 (0.37)	2.50 (0.39)	3.29 (0.46)	0.33 (0.62)	0.64 (0.37)	1.57 (0.61)	0.30 <b>(0.36)</b>
	-		0.94 (0.78)	0.92 (0.83)	0.93 (0.74)	<b>0.7</b> 9 (0.79)	0.92 (0.71)	0.87 (0.74)	<b>0.96</b> (0.81)	0.94 <b>(0.89)</b>	0.88 (0.72)	<b>0.96</b> (0.84)	0.95 (0.86)	0.94 (0.88)	0.94 (0.72)	<b>0.96</b> (0.55)
	S	1.98 (3.93)	2.09 (4.29)	2.76 (6.23)	2.35 (4.97)	4.77 (8.47)	2.36 (4.99)	2.87 (4.23)	2.87 (4.21)	2.28 (3.75)	2.19 (4.51)	2.47 (3.81)	2.99 (5.28)	2.37 (4.76)	2.34 (4.20)	2.08 (3.96)
	KGE		<b>0.94</b> (0.90)	0.60(0.40)	0.70 (0.69)	- 0.45 (- 0.17)	0.74 (0.73)	0.55 (0.91)	0.49 (0.91)	0.80 (0.94)	0.87 (0.85)	0.75 (0.95)	0.48 (0.65)	0.80 (0.79)	0.81 (0.92)	0.94 (0.98)
ebarik	BIAS		0.28 <b>(0.13)</b>	<b>0.16</b> (0.45)	0.33 (0.38)	0.03 (0.35)	0.18 (0.37)	0.77 (0.69)	0.41 (0.48)	0.32 (0.24)	0.28 (0.62)	0.34 (0.53)	0.24 (0.32)	0.22 (0.35)	0.46 (0.56)	<b>0.30</b> (0.42)
	RMSE		0.26 (0.54)	0.47 (0.61)	0.50 (0.71)	0.51 (0.63)	0.41 (0.57)	1.01 (0.85)	0.48 (0.67)	0.53 (0.74)	0.66 (0.95)	0.43 (0.73)	0.36 <b>(0.47</b> )	0.43 (0.49)	0.68 (0.66)	0.46 <b>(0.47)</b>
	_		<b>0.97</b> (0.95)	0.93 (0.89)	0.76 (0.83)	0.68 (0.82)	0.83 (0.95)	0.95 (0.95)	0.77 (0.94)	0.88 (0.97)	0.93 (0.99)	<b>0.97</b> (0.97)	0.90 (0.94)	<b>0.97</b> (0.97)	0.96 (0.95)	0.97 (0.99)
	S	2.33 (4.53)	2.38 (4.52)	2.55 (4.45)	2.22 (4.39)	2.81 (5.20)	2.56 (5.27)	2.96 (4.87)	2.89 (5.07)	2.86 (5.90)	2.90 (5.60)	2.42 (4.26)	2.57 (4.64)	2.51 (4.88)	2.47( 4.40)	2.43 (4.41)
	KGE		0.92 (0.85)	0.79 (0.90)	0.84 (0.80)	0.74 (0.81)	0.83 (0.82)	0.67 (0.86)	0.65 (0.84)	0.65 (0.66)	0.72 (0.70)	0.83 (0.84)	0.82 (0.93)	0.90 (0.91)	0.91 (0.93)	0.94 (0.96)
tema	BIAS		0.14 (0.11)	0.39 (0.32)	0.42 (0.30)	0.42 (0.40)	0.34 (0.35)	0.89 (0.45)	0.34 (0.42)	0.41 (0.22)	0.54 (0.38)	0.34 (0.36)	0.30 (0.31)	0.38 (0.30)	0.62 (0.38)	0.43 (0.33)
guse	RMSE		0.24 (0.61)	0.42 (0.69)	0.42 (0.62)	0.48 (0.70)	0.45 (0.64)	0.54 (0.87)	0.50 (0.76)	0.47 (0.68)	0.58 (0.77)	0.51 (0.84)	0.34 (0.52)	0.39 (0.48)	0.47 (0.59)	0.18 ( 0.47)
	2		0.92 (0.85)	0.81 (0.91)	0.85 (0.80)	0.84 (0.89)	0.86 (0.93)	0.81 (0.88)	0.74 (0.90)	0.73 (0.84)	0.87 (0.81)	0.84 (0.85)	0.85 (0.94)	0.93 (0.95)	0.94 (0.94)	0.97 (0.98)
	S	2.21 (4.40)	2.54 (5.16)	2.58 (4.90)	1.99 (4.10)	2.19 (3.92)	1.98 (4.11)	<b>2.20</b> (4.78)	2.28 (4.64)	2.40 (5.23)	2.67 (5.33)	2.61 (5.58)	2.23 (4.27)	2.19 (4.42)	2.17 <b>(4.40)</b>	2.16 (4.32)
	KGE		0.84 (0.80)	0.79 (0.84)	0.80 (0.87)	0.84 (0.84)	0.80 (0.88)	0.86 (0.81)	0.89 (0.86)	0.80 (0.78)	0.75 (0.76)	0.77 (0.70)	0.85(0.94)	0.94 <b>(0.95)</b>	0.95 (0.93)	0.96 (0.96)
aytsebri	BIAS		<b>0.06</b> (2.09)	0.30 (2.04)	0.30(— 0.02)	0.34 ( <b>-</b> 0.06)	0.32 <b>(0.19)</b>	0.40 (0.80)	0.39 <b>(0.00)</b>	0.30 (0.30)	0.41 (2.02)	0.32 (0.13)	0.27 (0.68)	0.31 (0.84)	0.37 (0.93)	0.31(0.93)
	RMSE		0.40 (2.67)	0.36 (2.61)	0.56 (0.66)	0.91 (0.61)	<b>0.32</b> (0.79)	1.09 (1.39)	0.70 (0.64)	0.47 (0.56)	0.41 (3.19)	0.48 <b>(0.55)</b>	0.48 (0.97)	<b>0.32</b> (1.12)	0.64 (1.48)	0.44(1.48)
	-		0.94 (0.89)	0.89 (0.89)	0.83 (0.89)	0.84 (0.88)	0.83 (0.90)	0.86 (0.83)	0.89 (0.87)	0.82 (0.88)	0.86 (0.89)	0.86 (0.86)	0.85 (0.94)	0.94 (0.96)	0.96 (0.94)	0.97 (0.97)
	S	3.36 (3.91)	5.49 (6.36)	4.32 (6.37)	<b>3.45</b> (3.30)	3.20 (3.68)	2.98 <b>(4.09)</b>	4.69 (5.91)	5.48 (4.35)	3.56 <b>(3.59)</b>	6.57 (9.43)	2.92 (3.19)	2.92 (4.44)	3.76 (4.50)	4.66 (5.63)	12.04 (5.63)
	KGE		0.36 (0.33)	0.70 (0.25)	0.92 (0.69)	<b>0.93</b>	0.86	0.58 (0.42)	0.36 (0.78)	0.92 (0.87)	0.03 ( <u> </u>	0.86 (0.75)	0.86 (0.80)	0.87 (0.80)	0.61 (0.47)	<b>—</b> 1.60 (0.47)

tions	Stastical metrics	Observed	HadReg4-7	HadCCLM	HadREMOO	MPI- Reg4-7	MPI- CCLM	MPI- REMOO	NorESM- Reg4-7	NorESM- CCLM	NorESM- REMOO	CanESM- RCM	Reg ensemble	CCLM ensemble	REMOO ensemble	Ensemble
ima	BIAS		0.26 (0.50)	0.18 (0.47)	5.99 (0.65)	1.2 (0.73)	0.28 ( <b>-</b> 2.75)	5.77 (6.12)	8.19 (0.52)	1.02 <b>(0.27)</b>	7.65 (0.40)	0.57 (0.80)	0.39 (0.58)	<b>0.10</b> (- 0.67)	0.99(2.39)	0.50(0.77)
	RMSE		0.37 (0.76)	0.60 (0.75)	6.79 (0.83)	1.62 (0.90)	0.66 (2.90)	6.44 (6.31)	9.28 (0.78)	1.78 (0.80)	8.64 <b>(0.69)</b>	0.68 (1.03)	0.45 (0.70)	<b>0.21</b> (0.79)	1.1 (2.46)	0.56 (0.83)
	-		0.95 (0.90)	0.24 (0.87)	0.32 (0.90)	0.23 (0.88)	0.29 (0.77)	0.73 (0.81)	0.75 (0.91)	0.11 (0.87)	0.17 (0.86)	0.12 (0.92)	0.94 <b>(0.96)</b>	0.95 (0.91)	0.96 (0.91)	<b>0.99</b> (0.91)
	S	2.39 (4.82)	2.85 (5.64	1.13 (5.03)	5.04 (4.96)	2.56 (4.28)	2.59 (3.45)	4.10 (3.82)	6.41 (5.85)	4.96 (6.54)	5.75 <b>(4.73)</b>	2.74 (6.26)	2.47 (4.99)	2.40 (4.66)	2.26 (3.93)	<b>2.37</b> (4.55)
	KGE		0.80 (0.80)	0.08 (0.86)	<b>-</b> 0.33 (0.89)	0.22 (0.83)	0.28 (0.61)	0.19 (0.61)	— 0.74 (0.76)	— 0.40 (0.62)	— 0.67 (0.86)	0.11 (0.69)	0.93 <b>(0.94)</b>	0.95 (0.90)	0.92 (0.77)	<b>0.97</b> (0.89)

ole 4 (continu	ions Stastica metrics	ma RIAS
Tabl	itatio	arim

ensemble had relatively better performance than the others in simulating the annual and kiremit season rainfall over the subbasin, except some simulations such as MPI-CCLM and CCLM ensembles mean annual rainfall of Debarik and Ketema Niguse stations respectively (Table 3). Maytsebri station CCLM ensemble mean simulation; all the institutional based ensembles simulations of Ketema Niguse and Adiramets stations had better performance in replicating the kiremit season mean rainfall (Table 4).

Figure 6a, b illustrates the taylor diagram of the long term (1986-2005) areal annual and kiremit season rainfall between observational rainfall values with RCMs, the institutional based ensemble mean and the multi-model ensemble mean. Relatively, the best models are those whose simulated patterns correspond with the observed ones and are located near the specified point on the x-axis. Accordingly, the simulation of the areal annual and kiremit season rainfall of the multi-model ensemble mean demonstrates best matches with the observed values throughout the subbasin. The overlapped models (multi-model ensemble mean) agree best with the observed rainfall, had the smallest and equivalent values of RMSE, a closer value of the standard deviation and strong correlation. The poorly performed models MPI-RECMOO and MPI-Reg4-7 showed relatively lower correlation; high RMSE and standard deviation (SD) as compared to the best fitted multi-model ensemble mean in models simulating the mean annual rainfall and kiremit season rainfall respectively (Fig. 6a, b).

Statistical metrics (bias, RMSE, r, CV and KGE) evaluation of the mean annual and kiremit season rainfall showed that the multi-model ensemble was better in two or more statistical metrics at each station except Maytsebri station kiremit season rainfall, where the CCLM ensemble was better. Similarly, Taylor diagram showed that the multi-model ensemble was better in the replication of the areal annual and kiremit season rainfall of the subbasin. This finding, in line with other studies (Nikulin et al. 2012; Endris et al. 2013; Dibaba et al. 2019; Demissie and Sime 2021; Mengistu et al. 2021) reported that ensemble mean showed relatively better performance than individual RCMs. The grand ensemble was better in replicating the mean annual and seasonal rainfall in western Africa (Ilori and Balogun 2022).

The cumulative distribution function of the multimodel ensemble mean daily rainfall of the subbasin showed that the RCMs, institutional based ensembles mean and multi-model ensemble mean were good in replicating the CDF of the daily rainfall of the observed data (Fig. 7). Similarly, the occurrence of the number of rainy days showed the same pattern as the number of daily rainy days occurrence decreases from dry to very heavy rainfall both in the observed and model simulated daily rainfall. (Fig. 8). There was a large (4500 days) number of dry days (RR=0) occurrence both observed and model simulated rainfall. On rainy days, the light rainy days (RR  $\leq$  10 mm) occurrence was larger than heavy  $(RR \ge 10 \text{ mm})$  and very heavy  $(RR \ge 20 \text{ mm})$  rainy days (Fig. 8). Additionally, the average light rainfall days were over estimated by the RCMs, institutional based ensembles mean and multi-model ensemble mean simulation, while the rate of heavy and very heavy rainy days were higher in the observed rainfall than in the model simulation except MPI-Reg4-7 and Reg ensemble mean heavy rainfall days. The result revealed that the MPI-REMOO, NorESM-REMOO, Reg ensemble and MPI-CCLM were better in replicating the number of days of dry, light,



Fig. 6 Taylor diagram of areal annual rainfall (a), areal kiremit season rainfall (b)



Fig. 7 Cumulative Distribution Function (CDF) of the multi-model observed and model simulated daily rainfall Zarima subbasin



Fig. 8 The long-term (1986–2005) daily areal observed and RCMs, institutional based ensembles mean and multi-model ensemble mean rainfall



Fig. 9 Monthly maximum temperature plots a Maytsebri, b Adiramets, c Debarik d Ketema Niguse and e Zarima

heavy and very heavy rainfall. The REMOO simulations were superior in replicating the number of dry and light rainy days of the average daily rainfall of the Zarima subbasin. This finding is in line with studies conducted in the Blue Nile basin (Worku et al. 2018; Demessie et al. 2023).

## **Performance of RCMs in simulating the maximum and minimum temperature in different temporal scale** Figure 9a–e showed the monthly pattern of the RCMs,

institutional based ensemble means and multi-model ensemble mean monthly maximum and minimum



Fig. 10 Monthly minimum temperature plots a Maytsebri, b Adiramets, c Debarik d Ketema Niguse and e Zarima

temperature the result revealed that the individual RCMs, institutional based ensemble and multi-model ensemble were good in representing the monthly pattern of the observed monthly maximum and minimum temperature even if some of the simulations were closer to the observed temperature and others were partly different in replicating the monthly maximum and minimum temperature. MPI-CCLM, multi-model ensemble mean, REMOO ensemble mean HadREMOO and Had-Reg4-7 were relatively better than the others in replicating the monthly pattern of the maximum temperature at Maytsebri, Adiramets, Debarik Ketema Niguse and Zarima stations respectively, while ensemble HadC-CLM, CCLM ensemble, MPI-CCLM, MPI-Reg4-7 were relatively better in replicating the monthly minimum





Fig. 12 The observed and model simulated mean annual minimum temperature of the Zarima subbasin

temperature at Maytsebri, Adiramets, Debarik Ketema Niguse and Zarima stations respectively Fig. 10a–e.

The observed, RCMs simulated, an institutional based ensembles mean and multi-model ensembles mean of the mean annual maximum temperature are presented in Fig. 11. The result showed that the evaluated RCMs generally good in capturing the mean annual maximum temperature of climatic stations of Zarima subbasin well. Specifically, the MPI-Reg4-7 simulation performs well in representing the mean annual observed maximum temperature at Adiramets and Maytsebri stations, while the Debarik and Ketema Niguse stations exhibit superior performance in the HadReg4-7 simulation and the Zarima station shows better representation in the CCLM ensemble simulations. However, there are some RCMs that demonstrate poor performance in capturing the mean annual maximum temperature at specific stations. The HadReg4-7 simulations for the Adiramets station, MPI-REMOO simulations for the Debarik and Maytsebri stations, Ketema Niguse station's NorESM-REMOO simulation, and Zarima station's NorESM-Reg4-7 simulations show limitations in representing the mean annual maximum temperature.

Similarly, Table 4 showed that the mean annual maximum temperature of the subbasin was over estimated by all the considered RCMs, institutional based ensembles mean and the multi-model ensemble mean simulations except Adiramets station MPI-Reg4-7 simulation, which under estimated by the magnitude of  $0.10^{\circ}$ C. The overestimating magnitude ranged between 0.06 and 8.9 °C. The lowest and highest over estimation was observed at Maytsebri HadReg4-7 and Zarima NorESM-Reg4-7 simulations (Table 4). Similar result was found in South west Ethiopia where, all evaluated CORDEX-AFRICA-RCMs overestimate the maximum temperature (Demissie and Sime 2021).

Regarding the bias of the mean annual maximum temperature of Adiramets station MPI-Reg4-7, the Debarik station multi-model ensemble mean, Ketema niguse and Maytsebri stations HadReg4-7 and Zarima stations the CCLM ensemble mean simulations demonstrated superior performance than the others.

Based on analysis of the observed and model simulated mean annual minimum temperature majority of the model simulations exhibit good representation at Adiramets, Debarik, and Zarima stations (Fig. 12). Specifically, CanESM-RCM, HadReg4-7, REMOOensemble, multi-model ensemble, and Regensemble simulations perform better at Adiramets, Debarik, Ketema Niguse Maystebri, and Zarima stations, respectively. On the other hand, the mean annual minimum temperature simulations of NorESM-CCLM at Adiramets, Ketema Niguse, and Zarima stations, NorESM-REMOO at Debarik station, and HadReg4-7 at Maystebri station demonstrate poor performance in representing the observed values.

The statistical evaluation of the annual minimum temperature is tabulated in Table 5. The result showed that similar to the annual maximum temperature majority of the simulation overestimated the mean annual minimum temperature. The annual minimum temperature bias values range was from 0.43 to 3.57°C. The lowest and highest bias values are observed HadReg4-7 and MPI-CCLM simulation at Ketema niguse and Zarima station respectively. Whereas the underestimation values were ranged from 0.09 to 5.07 °C which was observed at REMOO ensemble simulation of Zarima station and HadReg4-7 simulation of the Adiramets station (Table 5). The over estimation was high in highland areas and low in lowland area of the subbasin. Similarly, the study conducted in Tanzania the bias was relatively high in highly elevated area and low in low elevated area (Luhunga et al. 2016). The RCMs performance varies based on locations and topography in upper Blue Nile basin (Dibaba et al. 2019). North western Ethiopia climate simulations CORDEX-AFRICA-RCMs are sensitive to elevation, resulting higher biases for higher elevation (Van Vooren et al. 2019).

The simulations of the mean annual minimum temperature at Adiramets station the multi-model ensemble mean, HadReg4-7at Debarik station, REMOO ensemble mean at Ketema niguse and Zarima stations, and CanESM-RCM ensemble at Maystebri stations outperformed than other simulations in terms of bias.

The kiremit season mean maximum temperature was over estimated by majority of RCMs, institutional based ensembles mean and multi-model ensemble mean except HadCCLM at Adiramets station, HadREMOO and MPI-Reg4-7 at Maytsebri station and MPI-CCLM, CCLM ensemble mean at Zarima station which underestimated the mean kiremit season maximum temperature with magnitude ranged between 0.02 and 2.75 °C. The lowest and highest underestimations were observed at Adiramets station HadCCLM simulation and Zarima station CCLM ensemble simulation. On the other hand, the overestimation magnitude was ranged from 0.1 to 6.12 °C. The lowest and highest over estimation were observed at Adiramets CCLM ensemble mean simulation and Zarima station MPI-REMOO simulation respectively (Table 4).

Similar to the kiremit season mean maximum temperature mean minimum temperature of the kiremit season was overestimated by the majority of the simulation with values ranged from 0.01 to 6.03 °C whereas the underestimation value was ranged between 0.32 and 7.01 °C. The highest and lowest overestimation was observed in HadREMOO and MPI-REMOO simulations at Debarik and Adiramets stations respectively while the underestimation highest and lowest value was observed at Ketema niguse REMOO simulation and Adiramets station HadReg4-7 simulation respectively (Table 5). Similarly, almost 50% of the CORDEX-RCMs simulations showed overestimation of minimum temperature during the kiremit season (JJAS) (Tumsa 2022).

The simulation of the HadCCLM at Adiramets station HadReg4-7 at Debarik and Ketema niguse; NorESM-CCLM at Maystebri and Zarima stations kiremit season mean maximum temperature showed superior performance than the other RCMs. HadReg4-7 outperformed kiremit season mean minimum temperature over the subbasin except Ketema niguse station where REMOO ensemble was better.

The RMSE values between the observed and the RCMs simulated, the institutional based ensembles mean and the multi-model ensemble mean of the annual maximum temperature was low (0.24 °C) at Ketema niguse station HadReg4-7 simulation and high (9.28 °C) at Zarima station NorESM-Reg4-7 simulation (Table 5). Whereas, the NorESM-REMOO and MPI-REMOO simulations

bracket)																
Stations	Stastical metrics	Observed	HadReg4-7	HadCCLM	HadREMOO	MPI- Reg4-7	MPI- CCLM	MPI- REMOO	NorESM- Reg4-7	NorESM- CCLM	NorESM- REMOO	CanESM- RCM	Reg ensemble	CCLM ensemble	REMOO ensemble	Ensemble
Adiramets	BIAS		- 5.07 (- 7.01)	2.26 (0.52)	- 3.84 (0.01)	- 1.73 (0.32)	1.02 (0.53)	- 0.90 (0.54)	3.79 (0.56)	2.44 (0.64)	<b>-</b> 2.74 (0.38)	0.47 (0.56)	- 1.31 (-2.04)	0.71 (0.56)	0.43 (0.31)	0.00(- 0.30)
	RMSE		4.08 (5.55)	1.88 (0.52)	3.20 (0.69)	3.53 (0.41)	1.98 (0.56)	4.51 (0.51)	3.17 (0.59)	2.04 (0.61)	2.29 (0.42)	0.42 (0.55)	1.07 (1.63)	0.59 (0.49)	0.41 (0.34)	0.09 (0.27)
	L		0.90 (0.84)	0.44 (0.86)	<b>-</b> 0.19 (0.86)	0.12 (0.88)	0.05 (0.86)	0.20 (0.89)	0.25 (0.88)	0.59 (0.86)	0.19(0.88)	0.92 (0.85)	0.96 (0.94)	0.93 (0.95)	0.94 (0.94)	(86.0) 66.0
	S	2.69 (3.68)	4.23 (9.10)	2.73 (3.17)	4.48 (6.76)	20.54 (3.97)	9.78 (4.13)	27.30 (3.22)	4.48 (4.73)	3.63 (4.05)	3.62 (3.27)	2.77 (3.41)	3.30 (5.10)	<b>2.65</b> (3.43)	3.34 (4.07)	2.93 (3.91)
	KGE		0.38 (- 0.51)	0.43 (0.80)	<b>-</b> 0.38 (0.15)	<b>-</b> 5.69 (0.85)	<b>-</b> 1.78 (0.81)	<b>-</b> 8.18 (0.83)	— 0.015 (0.69)	0.45 (0.83)	0.11 (0.84)	0.92 (0.83)	0.76 (0.60)	<b>0.92</b> (0.91)	0.75 (0.88)	0.91 <b>(0.93)</b>
Debarik	BIAS		0.46 (0.46)	1.52 (0.93)	1.31 (1.28)	1.42 (1.23)	1.51 (1.38)	4.05 (6.03)	1.01 (0.62)	1.84 (1.65)	5.06 (1.25)	1.26 (0.97)	0.96 (0.77)	1.62 (1.32)	3.47 (2.85)	1.94 (1.58)
	RMSE		<b>0.41</b> (0.54)	0.87 (0.69)	0.76 (0.88)	0.83 (0.81)	0.87 (0.88)	2.57 (4.21)	0.63 (0.56)	1.06 (1.04)	2.82 (0.82)	0.76 (0.71)	0.56 <b>(0.51)</b>	0.92 (0.84)	1.96 (1.79)	1.08 (0.93)
	L		0.79 (0.76)	0.90 (0.73)	0.89 (0.77)	0.83 (0.78)	0.88(0.81)	0.77 (0.72)	0.82 (0.76)	0.82 (0.79)	0.84 (0.79)	0.80 (0.72)	0.95 (0.92)	0.94 (0.86)	0.90 (0.86)	0.96 (0.95)
	S	4.25 (5.04)	4.21 (5.58)	3.80 (3.45)	3.24 (5.63)	3.36 (4.04)	3.31 (3.48)	11.53 (17.50)	3.79 (4.63)	3.05 (2.71)	4.57 (4.17)	2.93 (4.06)	3.21 (3.92)	3.11 (2.89)	5.85 (8.26)	3.81 <b>(4.55)</b>
	KGE		0.79 (0.74)	0.84 (0.58)	0.73 (0.74)	0.72 (0.70)	0.74 (0.63)	<b>-</b> 0.74 (-1.51)	0.78 (0.74)	0.65 (0.49)	0.71 (0.72)	0.62 (0.66)	0.75 (0.76)	0.72 (0.55)	0.58 (0.33)	0.86 (0.87)
Ketema	BIAS		<b>-</b> 4.04 ( <b>-</b> 3.31)	1.77 (3.31)	0.48 (0.38)	<b>-</b> 0.17 ( <b>-</b> 0.47)	3.56 (4.60)	- 1.07 (- 1.72)	3.11 (5.50)	3.52 (4.99)	0.84 (0.38)	0.78 (0.81)	— 0.37 (0.57)	2.95 (4.30)	0.08 (- 0.32)	0.88 (1.45)
niguse	RMSE		3.14 (2.56)	1.45 (2.70)	0.59 (0.90)	0.91 (1.55)	2.76 (3.50)	0.98 (1.36)	2.46 (4.19)	2.76 (3.81)	0.75 (0.54)	0.67 (0.75)	1.13 (0.74)	1.60 (3.27)	0.58 (0.45)	0.29 (1.11)
	L		0.85 (0.78)	0.83 (0.79)	0.86 (0.73)	0.89 (0.73)	0.87 (0.85)	0.82 (0.82)	0.75 (0.74)	0.74 (0.72)	0.77 (0.72)	0.86 (0.75)	0.93 (0.83)	0.91 (0.89)	0.90 (0.83)	0.96 (0.95)
	C	2.85 (4.04)	3.75 (6.10)	4.33 (7.72)	4.65 (7.26)	7.85 (12.08)	2.23 (2.23)	5.22 (3.94)	3.99 (3.32)	3.51 (1.40)	3.26 (2.64)	3.24 (3.06)	3.30 (6.15)	2.65 (3.31)	3.34 <b>(4.16)</b>	<b>2.93</b> (3.85)
	KGE		0.60 (0.43)	0.45 (0.05)	0.35 (0.16)	<b>-</b> 0.76 ( <b>-</b> 1.01)	0.70 (0.48)	0.15 (0.80)	0.51 (0.59)	0.62 (0.25)	0.72 (0.55)	0.80 (0.65)	0.33 (0.45)	0.83 (0.71)	0.59 (0.83)	0.74 (0.90)
Maytsebri	BIAS		0.72 (0.53)	0.72 (0.66)	0.88 (0.76)	0.75(0.66)	0.72(0.56)	0.66 (0.59)	0.83 (0.58)	0.92 (0.72)	0.78 (0.71)	<b>0.59</b> (0.65)	0.76 (0.59)	0.78 (0.65)	0.78 (0.69)	0.76 (0.64)
	RMSE		0.49 (0.42)	0.51 (0.50)	0.60 (0.58)	0.52 (0.52)	0.57 (0.48)	0.49 (0.46)	0.56 (0.46)	0.61 (0.55)	0.53 (0.54)	0.48 (0.58)	0.50 (0.41)	0.53 (0.46)	0.51 (0.48)	0.47 (0.41)
	L		0.87 (0.89)	0.87 (0.86)	0.88 (0.82)	0.84 (0.85)	0.79 (0.82)	0.79 (0.83)	0.88 (0.86)	0.86 (0.84)	0.90 (0.80)	0.84 (0.74)	0.94 (0.94)	0.89 (0.93)	<b>0.95</b> (0.91)	0.95 (0.96)
	C	2.90 (3.32)	2.67 (3.80)	3.23 (3.47)	3.50 (3.84)	2.59 (3.81)	4.02 (3.85)	2.40 (3.18)	3.05 (3.69)	2.81 (3.71)	<b>2.96</b> (2.80)	3.95 (4.21)	3.21 (3.45)	3.11 (3.35)	5.85 (2.95)	3.81 (3.15)
	KGE		0.85 (0.82)	0.82 (0.85)	0.76 (0.76)	0.81 (0.79)	0.56 (0.76)	0.72 (0.82)	0.87 (0.82)	0.86 (0.80)	0.90 (0.75)	0.61(0.62)	0.86 <b>(0.93)</b>	0.86 (0.92)	0.90 (0.85)	0.93 (0.93)
Zarima	BIAS		0.43 <b>(0.31)</b>	1.78 (1.46)	0.19 (0.70)	0.80 (0.79)	2.28 (1.98)	- 1.12 (0.82)	1.50 (1.80)	2.44 (2.35)	0.67 (2.29)	1.40 (1.54)	0.91 (1.00)	2.17 (1.93)	<b>- 0.09</b> (1.27)	1.04 (1.42)
	RMSE		0.39 (0.68)	1.14 (1.02)	0.33 (0.64)	0.58 (0.74)	1.45 (1.30)	0.78 (0.66)	1.19 (2.21)	1.55 (1.57)	0.51 (1.53)	0.90 (1.10)	0.64 (0.90)	1.37 (1.28)	0.23 (0.88)	0.66 (0.93)
	L		0.86 (0.72)	0.89 (0.72)	0.78 (0.71)	0.82 (0.75)	0.91 (0.84)	0.77 (0.75)	0.70 (0.69)	0.87 (0.70)	0.82 (0.68)	0.89 (0.65)	0.88 (0.85)	0.95 (0.63)	0.89 (0.60)	0.96 (0.73)
	S	3.62 (4.60)	3.99 (6.70)	3.14 (3.10)	3.10 (2.43)	3.47 <b>(4.74)</b>	2.79 ( 3.05)	3.74 (3.39)	6.80 (15.37)	3.17 (1.82)	3.21 (3.45)	2.89 (4.08)	4.73 (7.50)	2.97 (2.31)	3.98 (2.53)	<b>3.58</b> (3.65)
	KGE		0.83 (0.46)	0.81 (0.57)	0.74 (0.44)	0.81 <b>(0.74)</b>	0.73 (0.62)	0.77 (0.64)	0.07 (-1.36)	0.79 (0.32)	0.79 (0.58)	0.76 (0.63)	0.79 (0.35)	0.75 (0.37)	0.79 (0.40)	<b>0.84</b> (0.65)

Kassahun et al. Environmental Systems Research

showed the lowest (0.36 °C) and highest (6.31 °C) RMSE values at Adiramets and Maytsebri stations of kiremit season mean maximum temperature respectively (Table 4). Generally, in terms of RMSE the multi-model ensemble mean was better at Adiramets and Ketema niguse station, HadReg4-7 at Debarik station and CCLM ensemble at Maystebri and Zarima stations. The kiremit season RMSE value showed that the multi-model ensemble mean was superior in simulating the observed kiremit season mean maximum temperature at adirmets, Debarik and Ketema niguse stations, whereas at Maytsebri and Zarima stations CanESM-RCM and NorESM-REMOO was better than the other respectively.

Unlike the maximum temperature the multi-model ensemble mean and HadReg4-7 simulated minimum temperature showed low RMSE values in both annual and kiremit season time scale at Adiramets station. The RMSE value observed at annual and kiremit season time scale ranged between 0.09-4.08 °C and 0.27-5.55 °C respectively. Regarding RMSE value the multi-model ensemble was superior in simulating the annual and kiremit season mean minimum temperature over the subbasin except Debarik and Zarima mean annual minimum temperature and Ketema niguse kiremit season mean minimum temperature where HadReg4-7, REMOO ensemble mean and Reg ensemble mean was better.

The majority of the annul and kiremit season maximum temperature simulations showed strong correlation with the observed mean annual and seasonal maximum temperature except Zarima station HadCCLM, HadREMOO, MPI-CCLM, NorESM-CCLM, NorESM-REMOO mean annual maximum temperature simulations which showed weak correlation. The multi-model ensemble showed superior correlation performance than the individual RCMs, institutional based ensembles mean in majority of the subbasin areas at annual and kiremit season time scales (Tables 4 and 5). Further CanESM-RCM showed equal performance with the multi-model ensemble mean at Adiramets and Debarik station and HadReg4-7 and CCLM ensemble at Debarik station in simulating the mean annual maximum temperature (Table 4). Similarly, the multi-model ensemble mean showed better correlation performance than the other simulations at annual and kiremit season time scale minimum temperature except kiremit season REMOO ensemble mean simulation at a Zarima station (Table 5). Contrary, negative correlation was observed at Adiramets station HadREMOO mean annual minimum temperature simulation (Table 5).

The KGE value of the annual and kiremit season mean maximum and minimum temperatures demonstrated good performance (KGE > 50) over the subbasin area, with the exception of few areas where the other RCMs were better. The mean annual maximum temperature of MPI-Reg4-7at Maystebri station and CCLM ensemble simulation of Adiramets and Ketema niguse mean minimum temperature performed better. Seasonally, the CCLM ensemble simulation of the Maystebri station kiremit season mean maximum temperature and Reg ensemble mean and MPI-Reg 4–7 simulation of the Zarima station mean maximum and minimum temperature respectively had superior performance than the other RCMs.

The RCMs, institutional based ensemble means and multi-models ensemble mean of annual and kiremit season mean maximum and minimum showed low inter annual and seasonal variability (CV < 20) throughout the subbasin. In terms of CV, the mean annual maximum temperature simulation was better in HadReg4-7 at Adiramets and Debarik stations; MPI-REMOO at Ketema Niguse and MPI-Reg4-7 at Maytsebri and Zarima stations (Table 4) whereas, Adiramets station NorESM-REMOO; Debarik station Reg ensemble mean; Ketema niguse REMOO ensemble mean; Maytsebri station MPI-CCLM and Zarima station HadREMOO simulations showed closer simulation to kiremit season mean maximum temperature (Table 4). This finding is in line with study conducted in southwestern Ethiopia where the CORDEX-AFRICA-RCMs showed low variability of the maximum and minimum temperatures on annual and seasonal time scales (Demissie and Sime 2021).

Among the evaluated simulations of the mean annual minimum temperature closer CV value was shown in the multi-model ensemble mean simulation of Ketema Niguse and Zarima stations; CCLM ensemble of Adiramets stations; HadReg4-7 of Debarik and NorESM-REMOO of Zarima station (Table 5). The multi-model ensemble mean, REMOO ensemble, CCLM ensembles and MPI-Reg4-7 mean kiremit season minimum temperature showed closer CV value with the observed minimum temperature at Adiramets and Debarik, Ketema Niguse, maytsebri and Zarima stations respectively (Table 5).

The Taylor diagram of the mean annual and kiremit season maximum and minimum temperature showed that the multi-model ensemble mean was better in replicating the subbasin areal mean annual and kiremit season maximum and minimum temperature (Fig. 13a–d). Like the rainfall, the mean kiremit and annual multi-model ensemble mean maximum and minimum temperature was better in two to four of the evaluation metrics thus the multi-model ensemble was better in replicating the maximum and minimum temperature of the subbasin.



Fig. 13 Taylor diagram of a areal annual maximum temperature, b the areal kiremit season maximum temperature, c areal annual minimum temperature and d the areal kiremit season minimum temperature

## Conclusion

This study evaluated the performance of historical simulation of 10 CMIP5 CORDEX-AFRICA-22<sup>0</sup> RCMs, three institutional based ensembles mean (Reg ensemble, CCLM ensemble REMOO ensemble), and one over all ensemble mean in replicating the observed mean annual and kiremit season rainfall, maximum and minimum over Zarima subbasin between 1986 and 2005 using statistical metrics, including bias, RMSE (Root Mean Square Error), r (Pearson correlation coefficient), CV (coefficient of variation), KGE (Kling-Gupta Efficiency)and Taylor diagram which shows the r, standard deviation, and centered root mean square difference graphically. The result indicated that monthly rainfall pattern of the multimodel ensemble and CanESM-RCM simulations exhibited better performance across the subbasin, except for the Ketema niguse station where, the REMOO ensemble and Reg ensemble showed superior performance in capturing the monthly rainfall patterns. Additionally, the NorESM-Reg4-7 simulation demonstrated good performance at the Debarik and Maytsebri stations, while the CCLM ensemble performed well at the Adiramets station in terms of capturing the monthly rainfall variations.

The monthly pattern of monthly maximum and minimum temperatures in the RCMs, institutional based ensemble means, and multi-model ensemble mean revealed that they generally performed well in representing the observed patterns. While some simulations were closer to the observed maximum and minimum temperatures, others showed some differences in replicating the monthly maximum and minimum temperatures. Specifically, the MPI-CCLM simulation, multi-model ensemble mean, REMOO ensemble mean, HadREMOO, and HadReg4-7 performed relatively better than other simulations in replicating the monthly pattern of maximum temperature at specific stations. Specifically, these simulations showed good agreement with the observed monthly patterns of maximum temperature at the Maytsebri, Adiramets, and Debarik stations. On the other hand, for the monthly minimum temperature, the multimodel ensemble mean, HadCCLM, CCLM ensemble mean, MPI-CCLM, MPI-Reg4-7 were relatively better compared to other simulations in replicating the monthly pattern. These simulations exhibited closer agreement

with the observed monthly patterns of minimum temperature at the Maytsebri, Adiramets, Debarik, Ketema niguse, and Zarima stations.

The mean annual rainfall analysis showed that Had-REMOO, MPI-Reg4-7, HadReg4-7, Reg ensemble, and multi-model ensemble mean performed relatively better in representing the mean annual observed rainfall at the Adiramets, Debarik Ketema, Niguse Maystebri, and Zarima stations, respectively. Whereas, NorESM-CCLM, MPI-CCLM, NorESM-Reg4-7, and NorESM-REMOO exhibited a weak performance in reproducing the observed mean annual rainfall at the Adiramets, Debarik Ketema niguse, Maystebri, and Zarima stations, respectively. Similarly, RCMs generally capture the mean annual maximum temperature of climatic stationsof Zarima subbasin well. Specifically, the MPI-Reg4-7 simulation performs well in representing the mean annual observed maximum temperature at Adiramets and Maytsebri stations, while the Debarik and Ketema niguse stations exhibit superior performance in the HadReg4-7 simulation and the Zarima station shows better representation in the CCLM ensemble simulations. The majority of the model simulations exhibit good representation of mean annual minimum temperature at Adiramets, Debarik, and Zarima stations. Specifically, CanESM-RCM, Had-Reg4-7, REMOOensemble, multi-model ensemble, and Regensemble simulations perform better at Adiramets, Debarik, Ketema niguse, Maystebri and Zarima stations respectively.

The majority of RCMs, as well as the ensemble mean simulations from various institutions and the multi-model ensemble were found to overestimate the observed mean annual and kiremit season rainfall. Specifically, the annual HadReg4-7 simulation showed the highest overestimation at the Maytsebri station, while the MPI-CCLM simulation had the lowest overestimation at the Ketema niguse station. On the other hand, the NorESM-Reg4-7 at Ketema niguse station and HadReg4-7 simulations at the Zarima station were observed to underestimate the mean annual rainfall. Seasonally, the minimum over estimation was seen in HadReg4-7 simulation of Ketema niguse station and maximum overestimation was seen MPI-CCLM simulations of Maytsebri stations. The RCMs are sensitive to elevation as the finding indicated that highest BIAS values were observed at highly elevated area of subbasin Ketema niguse and lowest BIAS values was observed at low land area of the Maytsebri and Zarima stations.

The combined analysis of statical metrics revealed that the multi-model ensemble mean, comprising multiple model simulations, outperformed the individual models in two or more statistical metrics at each station in simulating the annual and seasonal rainfall, maximum and minimum temperature, except Maytsebri station kiremit season rainfall simulation. In that case, the CCLM ensemble showed better performance. Additionally, the Taylor diagram demonstrated that the multi-model ensemble mean exhibited the best matches with the observed values for both the areal annual and kiremit season rainfall, maximum and minimum temperature across the entire subbasin. The findings implied that the multi-model ensemble mean provides a more reliable representation of the observed values and performs better across multiple evaluation criteria. This reinforces the value of using ensemble approaches to capture uncertainties and enhance the accuracy of climate model simulations.

The models had different performance in different statistical metrics at different location, and time considered (seasonal and annual) even the multi-model ensemble mean thus selecting the best representative simulation and bias correcting is important for the climate projection and climate change impact assessment study in the Zarima subbasin. This study assists water resource managers and hydrologists in selecting suitable models for their specific needs and responsibilities. Moreover, it contributes to the development of a reliable climate service assessment and facilitates decision making for climate adaptation, ultimately leading to optimal benefits in mitigating the impacts of climate change.

#### Acknowledgements

We thank the global Earth System Grid Federation for providing the CORDEX data. We would also thank the RCM data source website: https://esg-dn1.nsc.liu.se/ search/cordex/. Thank you to the Ethiopian National Meteorology Institute for providing the observed data.

#### Author contributions

Meaza Kassahun collact the data, analyzed and interpreted the data; wrote the paper. Kassahun Ture analyzed and interpret the data and edit the manuscript and Dessie Nedaw edit the manuscript.

#### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Availability of data and materials

The data that has been used is confidential. The datasets used for the study are not publicly.

#### Declarations

## Ethics approval and consent to participate

Not applicable.

#### **Consent for publication**

The manuscript has not published before.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Development and Environmental Management Studies, College of Social Sciences and Humanities, University of Gondar, Gondar, Ethiopia. <sup>2</sup>Center for Environmental Science, Addis Ababa University, Addis Ababa, Ethiopia. <sup>3</sup>School of Earth Science, Addis Ababa University, Addis Ababa, Ethiopia.

Received: 9 September 2023 Accepted: 22 November 2023 Published online: 12 December 2023

#### References

- Alemseged TH, Tom R (2015) Evaluation of regional climate model simulations of rainfall over the Upper Blue Nile basin. Atmos Res 161–162:57– 64. https://doi.org/10.1016/j.atmosres.2015.03.013
- ARSET (2022) Selecting climate change projection sets for mitigation, adaptation, and risk management applications. NASA applied remote sensing training program (ARSET). https://appliedsciences.nasa.gov/ getinvolved/training/english/arsetselectingclimate-change-proje ction-sets-mitigation-adaptation
- Ayugi B, Tan G, Gnitou GT, Ojara M, Ongoma V (2020) Historical evaluations and simulations of precipitation over East Africa from Rossby centre regional climate model. Atmos Res 232:104705. https://doi.org/10. 1016/j.atmosres.2019.104705
- Chai T, Draxler RR (2014) Root means square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. Geosci Model Dev 7:1247–1250. https://doi.org/10.5194/gmd-7-1247-2014
- Daba M (2014) Evaluating potential impacts of climate change on surface water resource availability of upper Awash Subbasin, Ethiopia rift valley basin. Paper presented at African dissemination workshop on innovative application of ICTs in addressing water-related impacts of climate change, Makerere University, Kampala, 12 Dec 2014
- 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. Model Earth Syst Environ. https://doi.org/10.1007/s40808-023-01751-0
- Demissie TA, Sime CH (2021) Assessment of the performance of CORDEX regional climate models in simulating rainfall and air temperature over southwest Ethiopia. Heliyon 7:e07791. https://doi.org/10.1016/j.heliy on.2021.e07791
- Dibaba WT, Miegel K, Demissie TA (2019) Evaluation of the CORDEX regional climate models performance in simulating climate conditions of two catchments in Upper Blue Nile Basin. Dyn Atmos Ocean 87:101104. https://doi.org/10.1016/j.dynatmoce.2019.101104
- Dosio A, Panitz HJ (2016) Climate change projections for CORDEX-Africa with COSMO-CLM regional climate model and differences with the driving global climate models. Clim Dyn 46:1599–1625. https://doi.org/10.1007/ s00382-015-2664-4
- Endris H, Misiani, G Sabiiti, MY Said, JO Ogutu, C Oludhe, Omondi P, G Galu, GO, MI (2017) Vulnerability, impacts and adaptation assessment in the east Africa region. https://www.usaid.gov/east-africa-regional
- Endris HS, Omondi P, Jain S, Lennard C, Hewitson B, Chang'a L, Awange JL, Dosio A, Ketiem P, Nikulin G, Panitz HJ, Büchner M, Stordal F, Tazalika L (2013) Assessment of the performance of CORDEX regional climate models in simulating East African rainfall. J Clim 26:8453–8475. https://doi.org/ 10.1175/JCLI-D-12-00708.1
- Eshetu M (2020) Hydro-climatic variability and trend analysis of modjo river watershed. Awash river basin of Ethiopia. Hydrol Curr Res 11:1–8
- Ethiopian National Meteorological Agency (NMA) (2015) Guide to the use of weather and climate information (WCI) for agricultural practices over selected districts of SNNP and Tigray National regional States Ethiopia, 78
- Evans JD (1996) Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co, Eagle River
- Giorgi F, Jones C, Asrar G (2009) Addressing climate information needs at the regional level: the CORDEX framework. WMO Bulletin 58:175–183
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res Atmos 113:1–20. https://doi.org/10.1029/2007J D008972

- Gutowski W, Solman S (2019) Coordinated regional downscaling experiment (CORDEX) 40th Session of the WCRP Joint Scientific Committee. https:// www.wcrp-climate.org/jsc40-about
- Hernández-Díaz L, Laprise R, Šushama L, Martynov A, Winger K, Dugas B (2013) Climate simulation over CORDEX Africa domain using the fifth-generation Canadian Regional Climate Model (CRCM5). Clim Dyn 40:1415–1433. https://doi.org/10.1007/s00382-012-1387-z
- Ilori OW, Balogun IA (2022) Evaluating the performance of new COR-DEX—Africa regional climate models in simulating West African rainfall. Model Earth Syst Environ 8:665–688. https://doi.org/10.1007/ s40808-021-01084-w
- Kamworapan S, Surussavadee C (2019) Evaluation of CMIP5 Global climate models for simulating climatological temperature and precipitation for Southeast Asia. Adv Meteorol. https://doi.org/10.1155/2019/1067365
- Kassie BT, Rötter RP, Hengsdijk H, Asseng S, Van Ittersum MK, Kahiluoto H, Van Keulen H (2014) Climate variability and change in the Central Rift Valley of Ethiopia: challenges for rainfed crop production. J Agric Sci 152:58–74. https://doi.org/10.1017/S0021859612000986
- Lemma H, Yesigat H, Jember T (2010) Analytical documentation of traditional practices and farmer innovations in agricultural water management in two traditional irrigation schemes in North-West Ethiopia, innovation and sustainable development in agriculture and food (ISDA) 2010, Montpellier, France, 10 p, hal-00530969
- Luhunga P, Botai J, Kahimba F (2016) Evaluation of the performance of COR-DEX regional climate models in simulating present climate conditions of Tanzania. J South Hemisph Earth Syst Sci. 66:32–54. https://doi.org/10. 22499/3.6601.005
- Luhunga PM, Kijazi AL, Chang'a L, Kondowe A, Ng'ongolo H, Mtongori H (2018) Climate change projections for tanzania based on high-resolution regional climate models from the coordinated regional climate downscaling experiment (CORDEX)-Africa. Front Environ Sci 6:1–20. https://doi. org/10.3389/fenvs.2018.00122
- Matiu M, Petitta M, Notarnicola C, Zebisch M (2020) Evaluating snow in EURO-CORDEX regional climate models with observations for the European Alps: biases and their relationship to orography, temperature, and precipitation mismatches. Atmosphere 2020 11(1):46. https://doi.org/10. 3390/atmos11010046
- Mendez M, Maathuis B, Hein-Griggs D, Alvarado-Gamboa LF (2020) Performance evaluation of bias correction methods for climate change monthly precipitation projections over Costa Rica. Water (Switzerland). https://doi.org/10.3390/w12020482
- Mengistu AG, Woldesenbet TA, Dile YT (2021) Evaluation of the performance of bias-corrected CORDEX regional climate models in reproducing Baro-Akobo basin climate. Theor Appl Climatol 144:751–767. https://doi.org/ 10.1007/s00704-021-03552-w
- Mowie (2009) Information and knowledge management project component 2 : strengthening water quality data generation and management. https:// www.waterethiopia.org/wp-content/uploads/2014/03/water-infor mation-and-knowledge-managementproject-component-2-strengethe ning-water-quality-data-generation-and-management-draft.pdf
- Nikulin G, Jones C, Giorgi F, Asrar G, Büchner M, Cerezo-Mota R, Christensen OB, Déqué M, Fernandez J, Hänsler A, van Meijgaard E, Samuelsson P, Sylla MB, Sushama L (2012) Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. J Clim 25:6057–6078. https:// doi.org/10.1175/JCLI-D-11-00375.1
- Schober P, Boer C, Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. Anesth Analg 126:1763–1768. https://doi.org/10. 1213/ANE.00000000002864
- Taylor KE (2001) Sumarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmospheres 106:7183–7192
- Tumsa BC (2022) Performance assessment of six bias correction methods using observed and RCM data at upper Awash basin, Oromia, Ethiopia. J Water Clim Chang 13:664–683. https://doi.org/10.2166/wcc.2021.181
- Van Vooren S, Van Schaeybroeck B, NyssenJ, Van Ginderachter M, Termonia P (2019) Evaluation of CORDEX rainfall in Northwest Ethiopia : sensitivity to the model representation of the orography. Int J Climatol 39(5):2569– 2586. https://doi.org/10.1002/joc.5971
- Ven Chow T, Maidment DR, Mays LW (1988) Applied hydrology. McgrawHill. https://doi.org/10.1061/(ASCE)1090-025X(1999)3:3(132)

- Vrugt JA, de Oliveira DY (2022) Confidence intervals of the Kling-Gupta efficiency. J Hydrol 612:127968. https://doi.org/10.1016/j.jhydrol.2022. 127968
- Weldesadik B (2021) Operational modeling for Zarima may day dam reservoir. Dissertation, Addis Ababa University
- 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 Ocean 83:53–63. https://doi.org/10.1016/j.dynatmoce.2018.06.002
- Yimer SM, Bouanani A, Kumar N, Tischbein B, Borgemeister C (2022) Assessment of climate models performance and associated uncertainties in rainfall projection from CORDEX over the Eastern Nile Basin, Ethiopia. Climate. https://doi.org/10.3390/cli10070095
- Zegeye MK, Bekitie KT, Hailu DN (2022) Spatio-temporal variability and trends of hydroclimatic variables at Zarima Sub-Basin North Western Ethiopia. Environ Syst Res. https://doi.org/10.1186/s40068-022-00273-5

#### **Publisher's Note**

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