Journal of
Water & Climate Change

Evaluation of four bias correction methods and random forest model for climate change projection in the Mara River Basin, East Africa

Priyanko das, Zhenke Zhang* and Hang Ren

Center of African Studies, School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210046, China
*Corresponding author. E-mail: zhangzk@nju.edu.cn

ABSTRACT

This study evaluates the performance of four bias correction methods based on CORDEX (coordinated regional climate downscaling experiment) domain six regional climate models (RCMs) at the Mara River Basin. A suitable bias correction method was considered to develop the future climate scenario. The performance of bias correction methods was evaluated by various statistical metrics based on the historical period and revealed that the distribution mapping (DM) techniques have strong performance under the different climatic conditions. The effectiveness of the DM method is found to be better at capturing the coefficient of variation and standard deviation of observed rainfall and temperature. Therefore, this study considers the future climate (2026–2095) from bias-corrected RCMs output using DM techniques. The results from bias-adjusted RCMs show an increase of rainfall (+118.3%) and temperature (+2.91) in the future climate under Representative Concentration Pathways (RCPs) 2.6, 4.5 and 8.5. In addition, this study tested the random forest (RF) method to determine the capacity of each bias-corrected RCMs for reproducing the future rainfall and temperature under the RCP 4.5 and RCP 8.5 scenario. The results demonstrate that the RF can reproduce the climate variable with its average correlation (R2) of 0.93 for rainfall and 0.95 for temperature.

Key words: BIAS correction, climate change, CORDEX Africa, RCMs, RF

HIGHLIGHTS

- The performance of four bias correction methods and climate simulation of six CORDEX Africa regional climate models (RCMs) was evaluated.
- The statistical matrix showed that the distribution mapping (DM) method results were better when compared to the other methods.
- The random forest (RF) model was performed for reproducing the climate scenario from the RCMs.
- It was found that the ensemble of all RCMs showed better results for the future climate scenario.

GRAPHICAL ABSTRACT
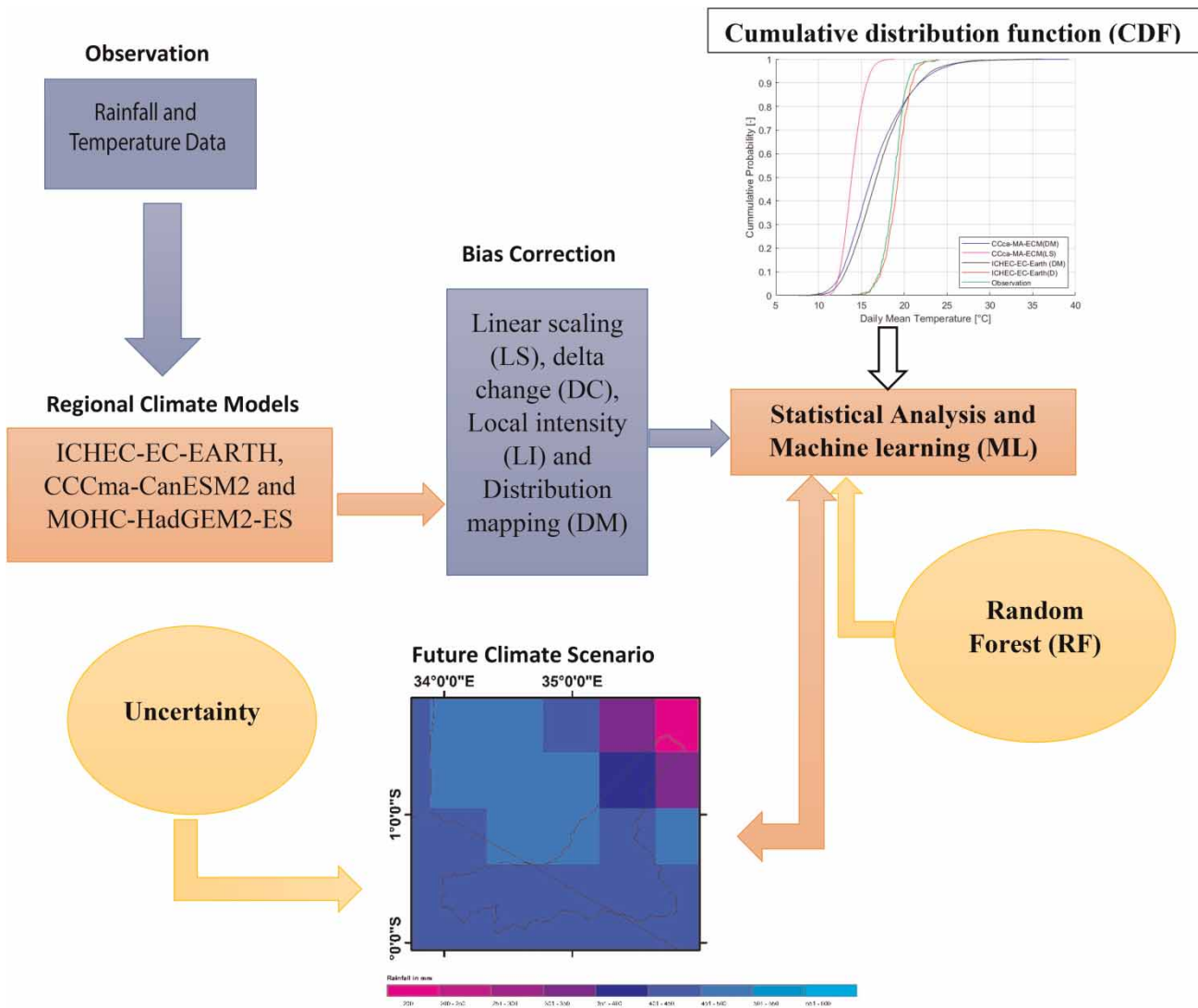
1. INTRODUCTION

The high-resolution global climate models (GCMs) and regional climate models (RCMs) are essential for understanding the global climate change variability and mitigation (Rauscher *et al.* 2010; Oo *et al.* 2020). As a major vulnerable region of climate variability, Africa is the first continent selected by the World Climate Research Program (WCRP) CORDEX (coordinated regional climate downscaling experiment) to generate an ensemble of high-resolution climate projections. The output of CORDEX evaluates the GCMs to understand the local climate variability and changes through the downscaling method. In general, it establishes coordination among regional climate downscaling for producing the GCMs (Dosio & Panitz 2016). Thus, downscaling the various GCMs output simulation is important for local climate change impact studies (Worku *et al.* 2020). Also, it is vital to select a better downscaling approach for GCMs simulation at the local scale to reduce the uncertainty.

Simulating the African climate models using the GCM is challenging due to complex and diverse processes, and the models have difficulties simulating the current climate (Laprise *et al.* 2013). Another factor is scale issues in the simulation process due to low spatial resolution 250 km × 250 km, respectively (Alemseged & Tom 2015), although GCM grid resolution (2°) is not a suitable approach for representing climate variables such as precipitation and temperature at regional climate (Dobler

& Ahrens 2008; Urrutia & Vuille 2009; Lafon *et al.* 2013; Fiseha *et al.* 2014; Teng *et al.* 2015; Roy *et al.* 2018). Through the better representation of regional climate, regional climate models (RCMs) provide high-resolution (50 km × 50 km) climate simulation and variability (Urrutia & Vuille 2009). Most of the previous studies proved that the output of CORDEX domain RCMs could reproduce the feature of African climate (Kim *et al.* 2014; Dosio & Panitz 2016; Favre *et al.* 2016; Pinto *et al.* 2016). In general, very few studies simulated the CORDEX domain RCMs at a local scale in the African continent (Hernández-Díaz *et al.* 2013; Laprise *et al.* 2013; Alemseged & Tom 2015; Dosio & Panitz 2016). In addition, the evaluation of CORDEX shows a reasonable standard and is used in various climate change impact studies of Africa (Worku *et al.* 2020). However, the downscaled RCMs cannot be used directly without bias correction for climate change impact studies due to systematic error and the reliability of the simulation.

Bias correction is a scaling method of climate model variables with observation values (Gudmundsson *et al.* 2012; Lafon *et al.* 2013). Although various bias correction methods have been developed and used in numerous studies to reduce the RCMs simulation biases, in general, most of the existing studies only focus on projecting the future climate variable but not for bias correction methods (Kim *et al.* 2019). Tschöke *et al.* (2017) evaluate two bias correction methods – gamma distribution and power transformation (PT) for precipitation simulation; they found that gamma distribution methods are more efficient in reducing the error from RCMs. Worku *et al.* (2020) evaluate four bias correction methods such as distribution mapping (DM), linear scaling (LS), variance scaling and PT for adjusting the RCMs temperature and precipitation with observation and developed future climate scenarios. The study proved that the DM correction method is more effective in adjusting the frequency of temperature and rainfall. Moreover, very few studies utilize the statistical metrics (e.g. Räty *et al.* 2014; Worku *et al.* 2020) to find the suitable bias correction method that has not yet been tested in the Mara River Basin (MRB). In addition, there are some limitations on various bias correction methods due to its strong dependency on observation data (Teng *et al.* 2015). However, bias correction methods factor in various climatic studies for reducing the error from climate models. Therefore, it is important to investigate the strongest bias correction method from RCMs output before developing the future climate change scenario.

The bias correction approach cannot correct the non-stationary RCMs biases and uncertainties (Wang *et al.* 2018). The machine learning (ML) approach for an ensemble of various GCMs/RCMs is selected in numerous studies to reduce these uncertainties (Tripathi Srinivas & Nanjundiah 2006). Campozano *et al.* (2016) compared two ML approaches – artificial neural network (ANN) and least square support vector machines (LS-SVM) – and both methods performed equally for downscaling the monthly precipitation. Crawford *et al.* (2019) used six ML approaches, including random forest (RF) and SVM and linear regression (LR), to evaluate the performance of ensemble coupled model intercomparison project phase 5 (CMIP5); they found that the RF method performed better than other approaches to simulate temperature. In a similar context, Wang *et al.* (2018) compared four ML methods (RF, Bayesian model averaging, the arithmetic ensemble means (EM) and SVM) for reproducing the observed monthly temperature and rainfall. The output of this study indicates that the RF ensemble method performs well and can reproduce monthly climatic datasets. However, no published work is found for the application of ML to reproduce the future rainfall and temperature datasets from the bias-corrected multiple RCMs ensemble. This study considers an RF ML approach to reproduce the future rainfall and temperature.

The MRB hydrologic system has been selected to evaluate the various bias corrections and an ML method using RCMs from CORDEX. Dessu & Melesse (2012) evaluate the five GCMs for climate change assessment at the MRB using a single bias correction method. Similarly, Roy *et al.* (2018) used the CIMP5 GCM to develop the future climate change scenario on the hydrological application of the MRB. However, most studies were based on the GCM instead of the RCM, and very few used RCMs at the basin scale, although dissimilar topography and various extra-terrestrial climatic phenomena impact the MRB rainfall and temperature pattern. For example, the rainfall of MRB is strongly influenced by the intertropical convergence zone (ITCZ) and Indian ocean surface temperature. In addition, these rainfall variabilities are also controlled by Eastern Pacific Ocean surface temperature, ENSO (El Nino Southern Oscillation) and La-Nina conditions (USAID 2019). Furthermore, it is important to understand the complex interaction of the climate component of the region (Laprise *et al.* 2013). Therefore, high-resolution simulation from RCMs is needed to improve climate change assessment representation on water resources and agriculture at the MRB.

The main objective of this study is to evaluate the performance of four bias correction methods in adjusting the CORDEX domain RCMs at the regional scale of Africa. This evaluation mainly finds the most effective bias correction method using statistical metrics and cumulative distribution function (CDF) derived from the historical RCMs. It assumes that the most effective bias correction method from historical RCMs performed better and developed the future climate change assessment.

Therefore, the best bias-corrected ensemble of RCMs was used to generate future climatic scenarios at the MRB. In addition, this study tested an ML (RF) approach based on the bias-corrected RCMs output to reproduce the future rainfall and temperature at a basin scale. Furthermore, the RF model determines the most important variable from each RCM. Moreover, to evaluate the performance of the bias correction method, the RF model was validated using various statistical metrics.

2. MATERIALS AND METHODS

2.1. The Mara River Basin

The MRB shares the transboundary between two African countries, Kenya and Tanzania, with latitudes −0.331573°S to 1.975056°S and longitudes 33.88372°W–35.907682°W (Figure 1). 65% of the upper MRB portion covers Kenya, and the remaining 35%, the lower part, is located in Tanzania (Mati *et al.* 2008). The Mara river (395 km) originates from the Mau Forest in Kenya at 3,000 m altitude and has fed into the Victoria Lake in Tanzania (Mango *et al.* 2011). The main tributaries of the MR are Nyangores and Amala, and other sub-tributaries are Engare Egito, Talek, Kenyo and Tambura. The river basin is affected by agriculture activity, deforestation, mining and other human activity.

The seasons of MRB are influenced by the movement of ITCZ and dry north-eastern wind, which comes from the Sahara Desert. The annual rainfall pattern of the MRB varies from 1,000 to 1,750 mm at the upper catchment and 700–850 mm at the lower catchment (around Musoma), which falls between April–September and November–December (Mati *et al.* 2008). The rainfall pattern in March to June is influenced by southeast trade winds from the Indian Ocean (Dutton *et al.* 2018). The average temperature of MRB is 25°C, where the maximum temperature recorded is 28 °C in the daytime and a minimum 10 °C at
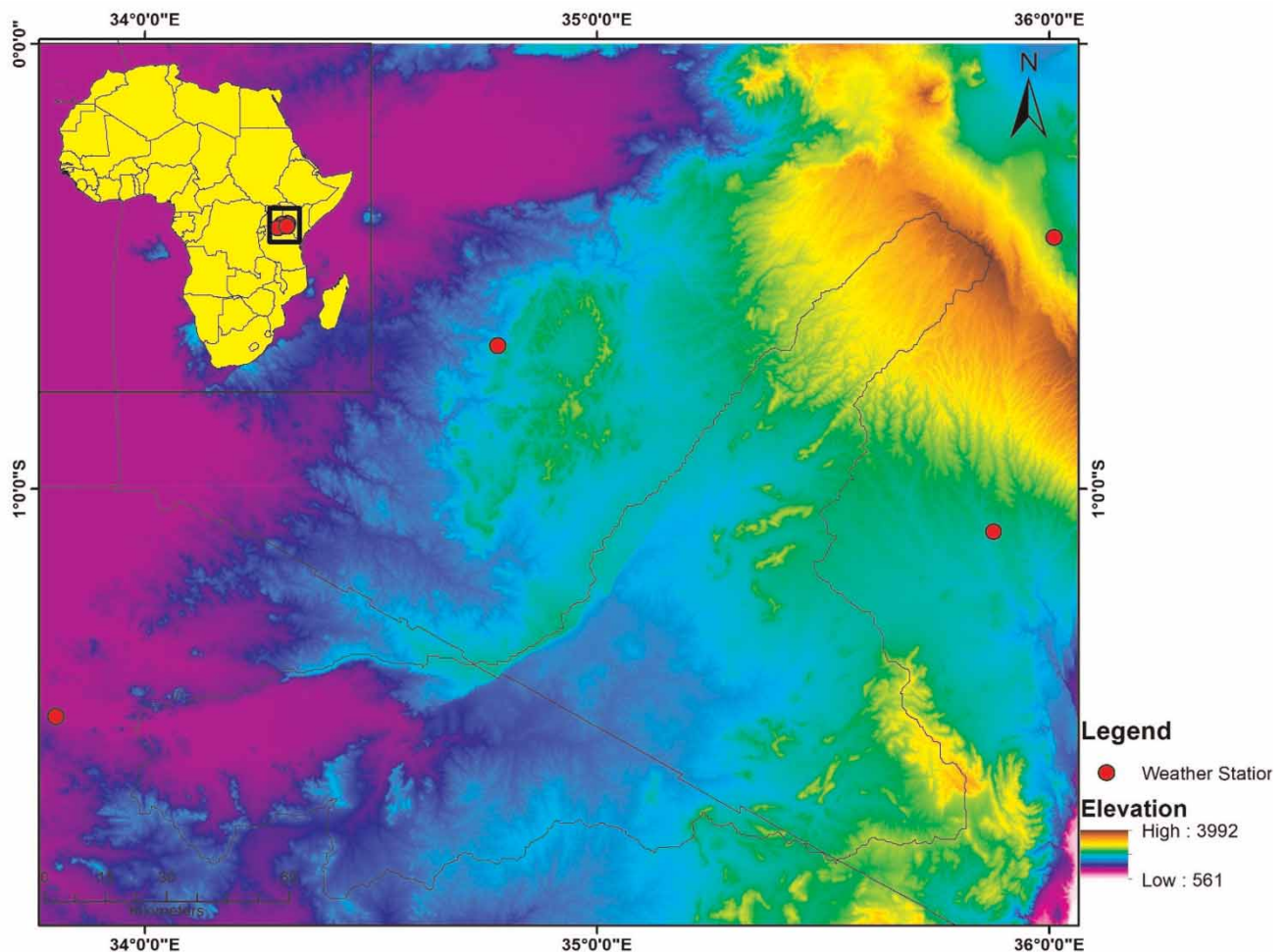

Figure 1 | Map of the MRB and the meteorological station with elevation data, obtained from the Shuttle Radar Topographic Mission (SRTM) (https://earthexplorer.usgs.gov/).

night. The climate is warm in the daytime and cool at night, which also depends on elevation. The temperature increased from October to February and slightly decreased from June to August.

2.2. Datasets

2.2.1. Meteorological station data

The historical observation of daily rainfall and temperature data was obtained from four meteorological stations in the MRB, and it is available at https://geographic.org/global_weather/. Thus, the website creates a database to store historical weather station datasets for the world based on the National Centres for Environmental Information (NCEI) (https://www.ncdc.noaa.gov/). One major problem in simulating RCMs in the African region is the lack of observation data availability, especially for the regional level, such as the river basin. It is tough to obtain African historical data from 1900; however, the historical datasets are available in the studied region from 1980. Therefore, this study considers historical observation datasets during 1986–2005 (20 years) to evaluate the various bias correction methods and develop future climate projections.

2.2.2. Regional climate models

This study obtained six historical and future RCMs from CORDEX projected to evaluate the bias correction method followed by Favre et al. (2016). The RCMs datasets are obtained from the Lawrence Livermore National Laboratory (LLNL) web portal powered by the Earth System Grid Federation (ESGF) (https://esgf-node.llnl.gov/search/esgf-llnl/). In this study, historical (1986–2005) and future (2026–2095) simulation of daily rainfall, maximum temperature (TMAX) and minimum temperature (TMIN) of ICHEC-EC-EARTH, CCCma-CanESM2 and MOHC-HadGEM2-ES models driven by six CORDEX RCMs (Table 1) were considered for bias correction and the RF model. Moreover, simulated and observation data vary in elevation and need bias adjustment in various locations (Worku et al. 2020). After retrieved the observation and RCMs rainfall and temperature, the data proceeded with a grid resolution of 0.44°×0.44° using the bilinear interpolation method to match the grid resolution of each RCMs. In addition, RCP 4.5, RCP 8.5 and RCP 2.6 emission scenarios were used to represent the radiative forcing level (Javadinejad et al. 2020; Oo et al. 2020). RCP 4.5 shows the intermediate emission levels of 4.5 W/m^2, while RCP 8.5 W/m^2 demonstrates the high radiative forcing with the emission level of 8.5 W/m^2, which could be stabilized after 2100 (Worku et al. 2020). Furthermore, RCP 2.6 is selected for considering the Paris agreement and sustainability for future climate change.

2.3. Evaluation of bias correction methods

2.3.1. Linear scaling method

The simulated average precipitation and temperature are scaled by the ratio of long-term observation and RCMs. The ratio is used to multiply the daily precipitation and temperature for each year (Lenderink et al. 2007), and this function is simplified

Table 1 | Description of RCMs used in this study

CORDEX Africa	Research Centre	RCMs	Reference
(ICHEC: Consortium of European research institute and researchers) KNMI – RACMO	Koninklijik Nederlands Meteorological Institute (Nederland)	RAC (EC-Earth)	Alemseged & Tom (2015)
(ICHEC) DMI-HIRHAM	Denmark's Meteorological institute (Denmark)	HIRHAM5 (EC-Earth)	Schmidli et al. (2007); Kim et al. (2014); Favre et al. (2016)
(ICHEC) MPI-ESM-LR	Max Planck Institute (Germany)	REMO (EC-Earth)	Navarro-Racines et al. (2020); Kim et al. (2014); Favre et al. (2016)
(ICHEC) CLMcom	CLM Community (www.clm-community.eu)	CCLM (EC-Earth)	Kim et al. (2014); Favre et al. (2016)
CCCma	Canadian Centre for Climate Modelling and Analysis (Canada)	CanESM2	Alemseged & Tom (2015)
MOHC	Met Office Hadley Centre (UK)	HadGEM2-ES	Navarro-Racines et al. (2020)

by the following equations (Zhang *et al.* 2018):

$$\rho_{con,cor\ (f)}\ =\ \rho_{con(f)} \times\ (\rho_{obs(y)}/\rho_{con(y)}) \tag{1}$$

$$\rho_{sec,cor\ (f)}\ =\ \rho_{sec(f)}\ \times (\rho_{obs(y)}/\rho_{con(y)}) \tag{2}$$

$$T_{cont.cor\ (t)} =\ T_{con(t)}\ + (T_{obs(y)}\ -\ T_{con(y)}) \tag{3}$$

$$T_{sec.cor\ (t)}\ =\ T_{sec(t)}\ + (T_{obs(y)}\ -\ T_{con(y)}) \tag{4}$$

where $\rho_{con,cor\ (f)}$ and $T_{con,cor\ (t)}$ are the corrected daily precipitation and temperature during the historical corresponding year. $\rho_{con,sec\ (f)}$ and $T_{sec,cor\ (t)}$ are the corrected daily precipitation and temperature for the future climate scenario. $\rho_{con(f)}$ and $\rho_{con(f)}$ represent the historical and future daily precipitation during the corresponding year. A similar process is followed for $T_{con(t)}$ and $T_{sec(t)}$, which represent the temperature of historical and future scenario. $\rho_{obs(y)}$ and $T_{obs(y)}$ are the observation of daily precipitation and temperature corresponding to the year. $\rho_{con(y)}$ and $T_{con(y)}$ represent simulated data for the historical and future scenarios of the corresponding year.

2.3.2. Delta change method

The delta change (DC) method creates a scenario based on adding anomalies for the future climate obtained from the RCMs simulation. It assumes that the regional level bias is constant over time (Beyer *et al.* 2019; Mendez *et al.* 2020). Although the evaluation of the DC method is expressed by many researchers (Räty *et al.* 2014; Sarr *et al.* 2015; Navarro-Racines *et al.* 2020), the function is simplified by the following equations (Mendez *et al.* 2020) for precipitation and (Beyer *et al.* 2019) for temperature.

$$\rho_{Contr}^{DC}(t)\ =\ \rho_{obs}(t) \tag{5}$$

$$\rho_{frc}^{DC}(t)\ =\ \rho_{obs}(t).\left[\frac{\mu m \rho frc(t)}{\mu m \rho contr(t)}\right] \tag{6}$$

$$T_{sim}(t)\ =\ T_{obs}(0) + (T_{sim}(t) -\ T_{sim}(0)) \tag{7}$$

where ρ is the daily precipitation and 'contr' is the simulated RCMs during the period, 'obs' represents the observational time series and 'frc' is the future RCMs scenario during the period, and 'DC' is the simulated RCMs. 't' shows the time step and 'μm' is the long-term yearly average. The expression of temperature '$T_{sim}(t)$' represents the simulated anomalies based on present observation. 'T'_{obs} is the observed daily temperature and 'T'_{sim} is the simulated daily temperature during the year. 't' is the obtaining time after bias term simulated temperature.

2.3.3. Local intensity method

The local intensity (LI) method widely reported that GCMs/RCMs precipitation works on the daily precipitation. This simulated precipitation from the GCM/RCM is corrected based on frequencies and the intensity of the wet and drizzle days (Dobler & Ahrens 2008). Although this method reported the daily precipitation of RCMs simulation is larger than observation, the threshold values of the wet days are adjusted for future wet-day frequencies (Olsson *et al.* 2015; Zhang *et al.* 2018). However, many researchers studied the evaluation of the LI method (Schmidli *et al.* 2007; Dobler & Ahrens 2008; Olsson *et al.* 2015), but the function was simplified by the following equations (Zhang *et al.* 2018):

$$P_{contr}^{1}(d) = \begin{cases} 0, & \text{if}\quad P_{con(d)}\ <\ P_{thres} \\ P_{con}(d), & \text{otherwise} \end{cases} \tag{8}$$

$$P_{sce}^{1}(d) = \begin{cases} 0, & \text{if}\quad P_{sec(d)}\ <\ P_{thres} \\ P_{sec}(d), & \text{otherwise} \end{cases} \tag{9}$$

where $P_{contr}^{1}(d)$ and $P_{sce}^{1}(d)$ are the daily precipitation of the historical and future climate, $P_{con}(d)$ is the observation daily precipitation and P_{sec} is the future daily precipitation for the control period and P_{thres} is the threshold value.

In the next step, the scaling factor is calculated based on the long-term mean wet-day frequencies. The intensity scaling is estimated from the S ratio:

$$S = \mu(P_{obs,m,d}/P_{obs,m,d} > 0) - 0/\mu(P_{con,m,d}/P_{con,m,d} > P_{thres}) - P_{thres} \tag{10}$$

In the last step, the bias-corrected daily precipitation was estimated using scaling factor and daily precipitation

$$P_{con,cor}(d) = s \times P^1_{con,}(d) \tag{11}$$

$$P_{sec,cor}(d) = s \times P^1_{sec,}(d) \tag{12}$$

where $P_{con,cor}(d)$ and $P_{sec,cor}(d)$ are represented for the bias-corrected historical and future daily precipitations and P^1_{sec} is the same as shown in Equations (8–10).

2.3.4. Distribution mapping method

The DM technique is based on different approaches such as probability mapping (PMAP), empirical CDF mapping, quantile mapping (QMAP) and Kernel density distribution (KDDM) (McGinnis Nychka & Mearns 2015) and is used in various climate simulation studies (Lafon et al. 2013; Olsson et al. 2015; Switanek et al. 2017). The main principle of this method is to fit the transfer function of RCMs simulation with an observational mean value (Zhang et al. 2018). Mainly there are two functions suitable for precipitation and temperature – gamma and Gaussian distribution (Worku et al. 2020) – although the histogram of bias-corrected RCMs and observation values has properties of positive skewness (Tschöke et al. 2017) to adjust the mean value, extreme, standard deviation (SD) and distribution of precipitation.

2.4. RF model

RF is a classification approach used in various predictive climate change assessment studies. This method mainly consists of a decision tree approach where each classifier is developed from the bootstrap sample (Gaál 2012). The bootstrap sampling makes RF less susceptible to overfitting than a decision tree (Wang et al. 2018), although the decision tree methods are developed from partitioning the training data (y_i, x_i), $i = 1, \ldots, n$, which is based on the independent variable (Crawford et al. 2019). However, the RF model contains a regression tree, which consists of many regression trees like a forest and obtained the prediction model based on the two parameters: (1) the n number of regression tree (n_{tree}) and (2) the selected evidential feature at each node (m_{try}) (Wang et al. 2018). Breiman (2001) and Segal (2003) have more details of the RF computation process.

In this study, we set the number of n_{tree} ranges from 500 to 100 and selected evidential feature m_{try} ranges from 5 to 20 for reproducing the future rainfall and temperature followed by Wang et al. (2018). The six bias-corrected RCMs output are considered a predictor variable in RF and used to determine the rank of relative impotence of each predictor, which may suggest the most important RCMs for contributing to reproducing the future climate scenario.

2.5. Performance evaluation

The CMhyd tool was used to process the four bias corrections methods (LS, DC, LI and DM) and retrieved the rainfall and temperature datasets from the RCMs (Zhang et al. 2018). However, the CDF is used to adjust the simulated and non-simulated rainfall and temperature datasets, then compared to the observational values (Figure 2) (Maurer & Pierce 2014). In addition, four statistical matrices (mean, SD, coefficient of variation (CV) and 90th percentile) are selected for evaluating the bias correction methods from historical RCMs and observation data. This evaluation is used to obtain the suitable bias correction method to obtain the future climate change scenario. The strongest bias correction method was applied for developing the near-term (2026–2045) and long-term (2076–2095) climate scenario at the MRB. Furthermore, the performance of the various bias correction methods and the RF model was validated using four evaluation metrics such as (a) the coefficient of determination (R^2), (b) mean square error (MSE), (c) mean absolute error (MAE) and (d) root mean square error (RMSE).
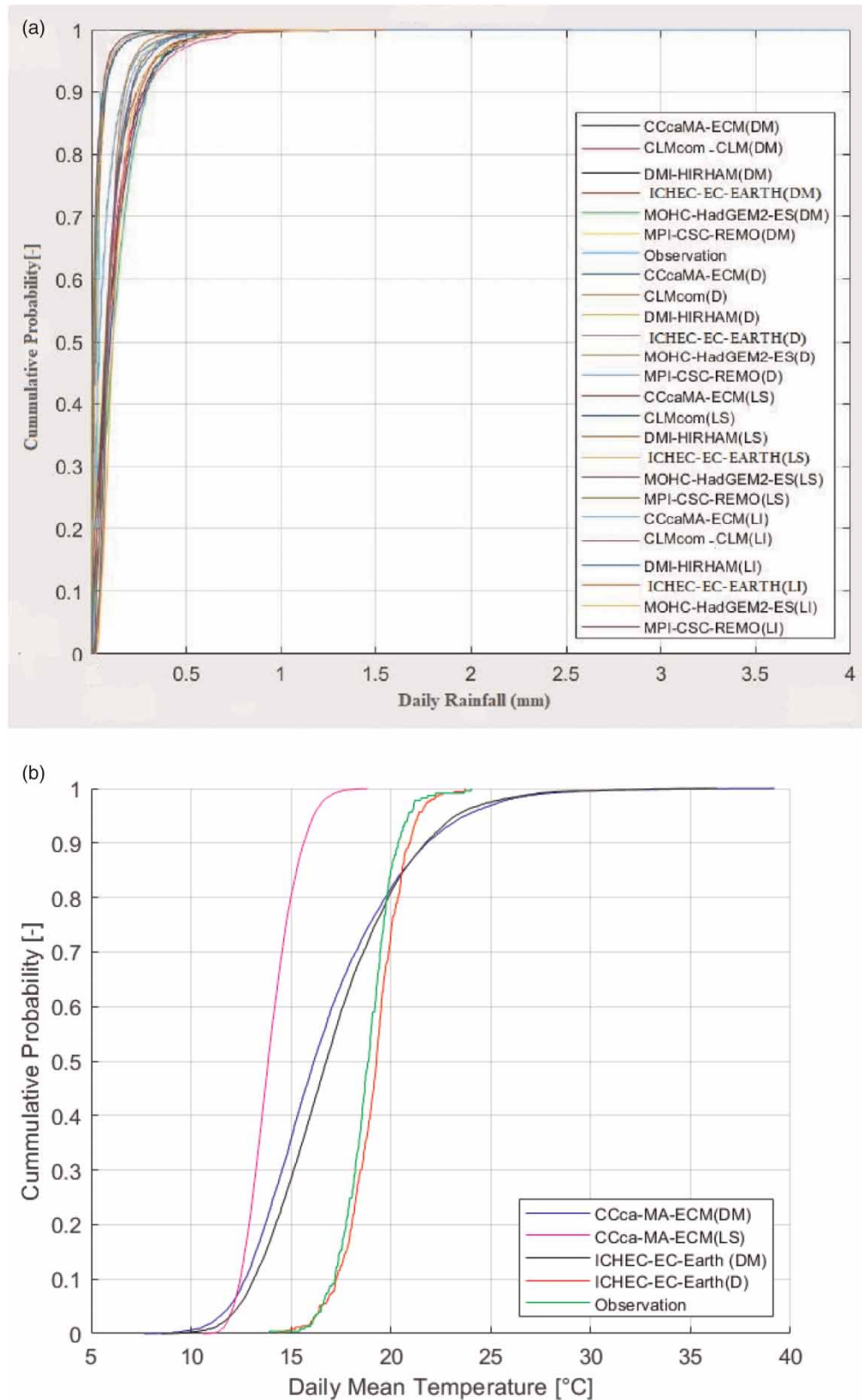
Figure 2 | CDF of simulated RCMs and observed (a) and daily rainfall (b) temperature in the MRB. DM, distribution mapping; DC, delta change; LS, linear scaling; LS, local intensity.

3. RESULTS

3.1. Evaluation of bias correction methods

The African continent is not suitable for a large global and regional model of precipitation change with different aspects (Déqué *et al.* 2017). In this study, results suggest that the four bias correction methods, LS, DC, LI and DM, are effective for adjusting the mean annual RCM simulation of rainfall. However, a significant difference was found between bias-corrected RCM and raw RCM for the spatial distribution of rainfall in the MRB (Figure 3). Also, the CDF results indicate the overestimation and underestimation of the annual rainfall were adjusted sufficiently in the entire sub-basin using LS, DC, LI and DM methods (Figure 2) (Worku *et al.* 2020). These four methods have comparable performance in adjusting the mean annual observed rainfall and RCMs output of the MRB (Figure 3). But the significant distribution of the total amount of rainfall was overestimated and underestimated depending on the changes in elevation (Gudmundsson *et al.* 2012).

The seasonal variation of precipitation derived from the bias-corrected RCMs shows comparable performance at a regional scale (Urrutia & Vuille 2009). In general, the mean monthly rainfall of simulated RCMs is adjusted through observed precipitation. The seasonal distribution of rainfall from bias-corrected RCMs and observation output provides the maximum peak of rainfall in the summer month of MJJ (May, June and July) and winter months of NDJ (November, December and January) (Figure 4(a)). Similarly, the LS and DC methods were found inadequate to adjust the CV of rainfall with an average value of 1.5, which is overestimated compared to the observation (Figure 4(d)). In the same contrast, LI and DM show good strength for CV with a value of 0.8 to adjust the rainfall from individual simulated RCMs. The rainfall pattern concerns the SD from various computed RCMs simulation (Figure 4(b)) at individual grid (Meehl Arblaster & Tebaldi 2005), although simulated CCCma-CanESM2 RCMs shows the maximum peak in SD compared to other simulated RCMs for rainfall and

Figure 3 | Annual mean rainfall (mm) (a), observation (b), KNMI (EC-EARTH) using raw simulation (c), KNMI (EC-EARTH) using DM (d), KNMI (EC-EARTH) using linear scaling (LS) (e), KNMI (EC-EARTH) using LI (f), HadGEM2 DM (g), HadGEM2-ES (MOHC) using DC (h), REMO (EC-EARTH) using linear scaling (LS) (i) and REMO (EC-EARTH) using LI (j) CanESM (CCCma) using LI bias-corrected output from the period (1986–2005).
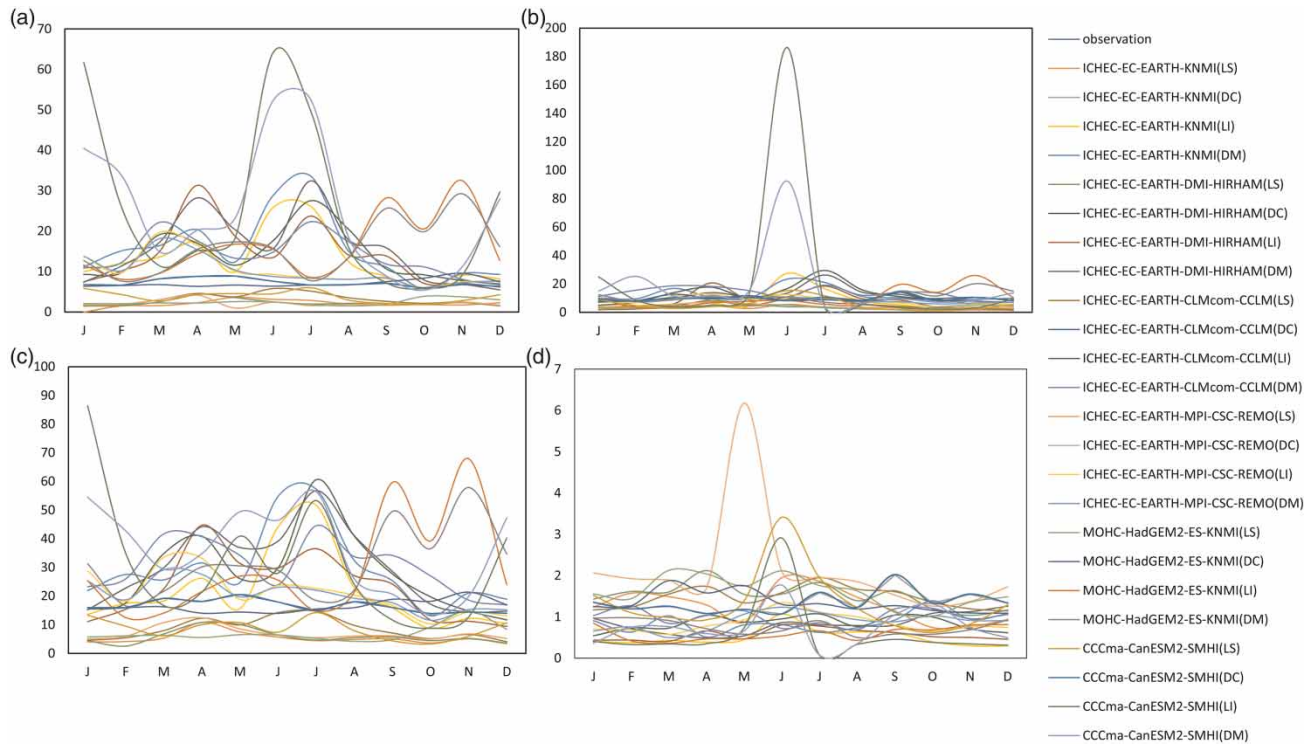
Figure 4 | The ability of LS, DC, LI and DM to adjust daily rainfall measured by mean (a), SD (b), 90th percentile (c) and CV (d) in the MRB.

shows the poor performance with a MSE value of 204.33 mm (Table 2). In addition, the 90th percentile of precipitation indicates a significant result between the DM-corrected RCM and the observation value (Figure 4(c)). However, the LI and DM methods show similar results in adjusting the annual mean rainfall corresponding to observation values.

Other studies also suggest a comparable performance among various bias correction methods in adjusting the rainfall and temperature values (Liang 2004; Räty et al. 2014; Tschöke et al. 2017; Wang et al. 2018). In addition, there is a limitation in various bias correction processes at individual RCMs, which is mainly based on resolution (Rauscher et al. 2010). However, LS, DC and LI methods shows overestimated or underestimated rainfall variation in the statistical framework. Therefore, the DM bias correction method was selected to compare the performance of individual RCMs output against observation output, followed by Worku et al. (2020).

Four evaluation matrices (R^2, RMSE, MAE and MSE) were selected to evaluate the performance of bias-corrected individual RCM and the EM of six RCMs. The bias-corrected DMI-HIRHAM (R^2=0.35) and HadGEM2-ES (R^2=0.26) RCMs showed excellent performance to reproduce the rainfall and temperature compared to observation values (Table 2). In addition, RMSE error values of these two RCMs were 12.94 mm (DMI-HIRHAM) and 11.84 mm (HadGEM2- ES), respectively, although the bias-corrected KNMI (ICHEC-EC-Earth) RCMs output showed poor performance compared to observation values with the RMSE of 10.54 mm and the MSE of 111.2 mm. However, the ensemble of all bias-corrected

Table 2 | R^2, RMSE, MAE and MSE between observation and simulated RCMs

Error	KNMI (EC-EARTH)	DMI- HIRHAM (EC-EARTH)	CLMcom – CCLM (EC-EARTH)	MPI-CSC-REMO (EC-EARTH)	KNMI-HadGEM2-ES (MOHC)	CanESM (CCCma)	Ensemble
R^2	0.11	0.35	0.12	0.30	0.26	0.21	0.45
RMSE	10.54	12.94	8.12	3.53	11.84	14.29	3.433
MAE	6.74	9.75	5.55	2.66	8.88	9.47	2.32
MSE	111.2	167.55	65.92	12.46	140.04	204.33	11.78

RCMs showed excellent performance with a correlation coefficient (R^2) of 0.4 and a minimum value for RMSE, MAE and MSE (3.43, 2.32 and 11.78). Therefore, this study selected DM-corrected ensemble of six RCMs output to develop the short-term (2026–2045) and long-term (2076–2095) future climate under the RCP 4.5, RCP 8.5 and RCP 2.6 scenario.

3.2. Future rainfall and temperature scenario in the MRB

The future rainfall and temperature pattern in the MRB were projected using the ensemble RCMs output (Figure 5). The individual RCMs of rainfall results were compared to historical simulated RCMs (Table 3). However, the future mean annual rainfall and temperature were categorized into short-term (2026–2045) and long-term (2076–2095) based on the emission scenarios of RCP 2.6, RCP 4.5 and RCP 8.5. The spatial variation of rainfall for the near-term (2026–2045) and long-term (2076–2095) is shown in Figure 6. The results from the ensemble RCMs demonstrate the increasing trend of future rainfall variation under various RCP scenarios such as RCP 4.5 (+110.45%), RCP 8.5 (+ 108.87%) and RCP 2.6 (+134.28%) (Table 3) in the near term (2026–2045). Similarly, the long-term (2076–2095) climate of rainfall trend increased +114.32, +123.305 and +121.57% under the emission scenario of RCP 4.5 and RCP 8.5. But the highest increased rainfall recorded based on the ensemble RCM output in the near-term future (2026–2045) with the amount of +123, +117 and +115% under the emission scenarios of RCP 4.5, RCP 8.5 and RCP 2.6. In addition, a slight increasing rainfall trend was found (13–23%) for REMO (MPI-ESM-LR) RCMs in the near-term and long-term scenario. Moreover, all the RCMs output and EM suggest the increasing rainfall trends for short-term (2026–2045) and long-term (2076–2095) climate in the MRB.

Most of the individual RCMs showed an increasing trend in annual rainfall (Table 3), except simulated REMO output (Worku *et al.* 2020). Other studies show similar rainfall trends for the future scenario in the MRB. For example, Dessu & Melesse (2012) found a 30–50% increase in the rainfall trend during the wet season (MAM) using five GCMs output for the year 2050. Also, this study's results suggest a significant increase in the water flow of MRB from 2046 to 2065 and 2081 to 2100. Roy *et al.* (2018) projected 31 CMIP5 RCMs output to evaluate the daily precipitation and found an increased rainfall trend in the MRB for the short-term future scenario (2020–2050). Similar studies carried out by Mellander *et al.* (2013)
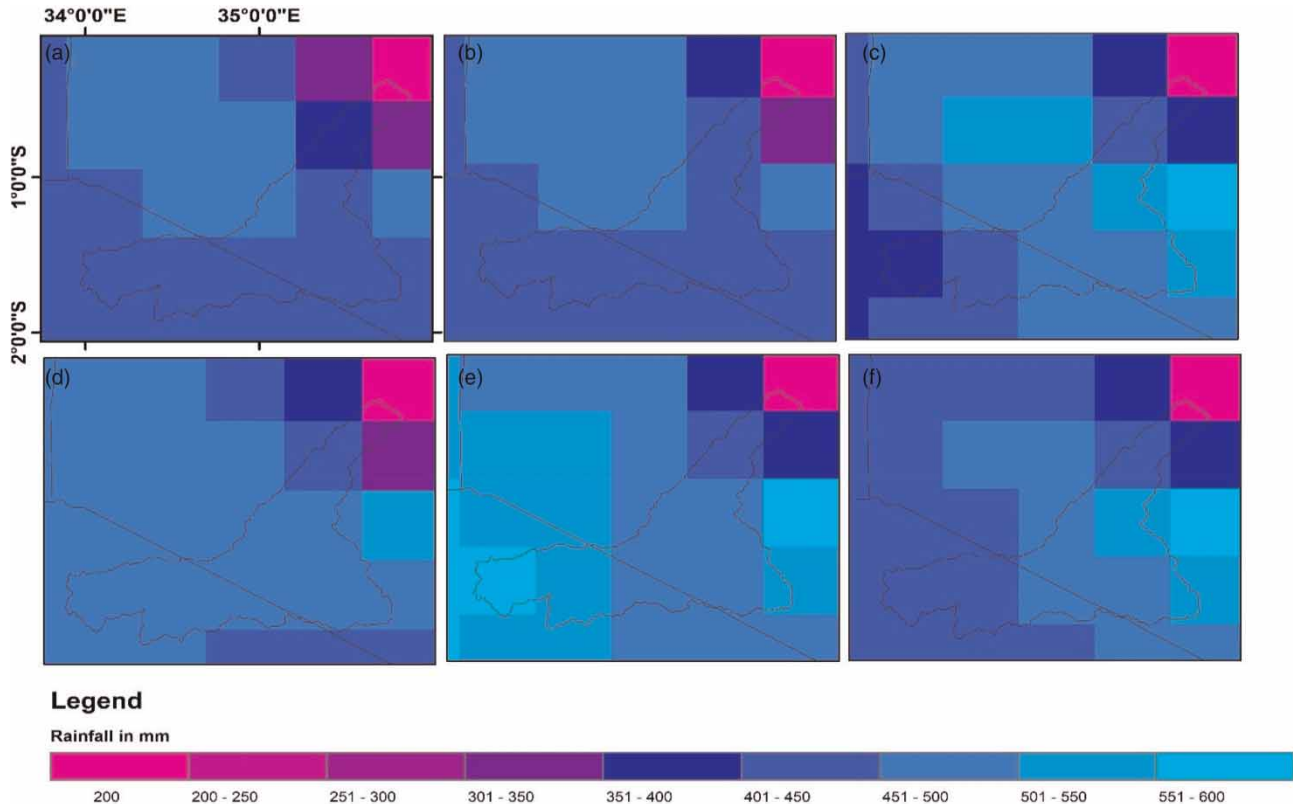

Figure 5 | Future RCMs ensemble of rainfall (a) EM of RCP 4.5 (2026–2045), (b) EM of RCP 8.5 (2026–2045), (c) EM of RCP 2.6 (2026–2045), (d) EM of RCP 4.5 (2076–2095), (e) EM of RCP 8.5 (2076–2095) and (f) EM of RCP 2.6 (2076–2095).

Table 3 | Future annual mean rainfall for the near-term (2026–2045) and long-term (2076–2095) compared to the historical rainfall (1986–2005)

Scenario	RCMs	2026–2045	Changes in %	2076–2095	Changes in %
	Observation (1986–2005)	148.16			
RCP 4.5	KNMI (ICHEC-EC-EARTH)	250.80	+64.8	252.95	+70.72
	CLMcom – CCLM (ICHEC-EC-EARTH)	262.53	+77.19	281.93	+90.28
	DMI-HIRHAM (ICHEC-EC-EARTH)	262.98	+77.49	278.93	+88.26
	MPI-CSC-REMO (ICHEC-EC-EARTH)	180.23	+21.64	172.49	+16.42
	SMHI (CCCma-CanESM)	294.05	+98.46	374.32	+152.64
	KNMI (MOHC-HadGEM2-ES)	620.45	+318.77	544.63	+267.59
	EM	311.84	+110.45	317.54	+114.32
RCP 8.5	KNMI (ICHEC-EC-EARTH)	263.39	+77.77	267.30	+80.41
	CLMcom – CCLM (ICHEC-EC-EARTH)	246.99	+66.70	266.83	+80.09
	DMI-HIRHAM (ICHEC-EC-EARTH)	266.48	+79.85	308.52	+108.23
	MPI-CSC-REMO (ICHEC-EC-EARTH)	183.09	+23.57	167.89	+13.31
	SMHI (CCCma-CanESM)	291.79	+96.94	420.32	+183.69
	KNMI (MOHC-HadGEM2-ES)	605.11	+308.41	554.24	+274.08
	EM	309.475	+108.87	330.85	+123.305
RCP 2.6	KNMI (ICHEC-EC-EARTH)	260.07	+75.53	263.24	+77.67
	MPI-CSC-REMO (ICHEC-EC-EARTH)	183.20	+23.65	180.99	+22.16
	KNMI (MOHC-HadGEM2-ES)	598.05	+303.65	540.65	+264.90
	EM	347.11	+134.28	328.29	+121.57

showed the +6% increased rainfall pattern in the upper Blue Nile basin based on the downscale of ECHAM5/MP1-OM during 2050–2100. In addition, Fenta Mekonnen & Disse (2018) reported the downscaled annual mean precipitation increased from 2.1 to 43.8% under A1B, A2, and B1 scenarios for 2006–2100 using the LARS-WG model.

Similarly, maximum temperature (TMAX) and minimum temperature (TMIN) showed an increasing trend in the near-term (2026–2045) and long-term (2076–2095) climate under the emission scenarios of 2.6, 4.5 and 8.5 (Figures 7 and 8). However, the ensemble of temperature output have shown the rise of TMAX 1.31 °C and TMIN 2.66 °C under the RCP 4.5 scenario and an increasing trend of TMAX 1.85 °C and TMIN 3.6 °C under the RCP 8.5 scenario during the near-term period (2026–2045). The maximum increase trend from the ensemble of TMAX (11.81 °C) and TMIN (15.12 °C) has been recorded in the long-term (2076–2095) scenario based on the RCP 8.5. The lower emission of increased TMAX and TMIN was projected under the RCP 2.6 scenario (Table 4). The spatial pattern of future TMAX and TMIN is shown in Figures 7 and 8 based on ensemble RCMs and observed values. IPCC projected a similar condition of an increased trend in East Africa (IPCC 5th assessment), which is estimated to change from 0 to 4 °C warmer by 2050 under the RCP 4.5 scenario (Solomon et al. 2007). The other studies also found a similar increasing trend in the MRB. For example, USAID (2019) reported the average temperature of MRB would increase by 1.97 °C (2030) and 2.71 °C (2050), which is a similar output of the ensemble near-term (2026–2045) and long-term future (2076–2095) under the RCP 4.5 scenario (Table 4). Therefore, the emission scenario of RCP 4.5 and RCP 8.5 was selected for the RF model to reproduce the rainfall and temperature.

3.3. RF model prediction

The study used EM of six observed RCMs output in the RF model to reproduce future climatic datasets. For developing the RF-based future climate scenario (precipitation and temperature), RCP 4.5 and RCP 8.5 were selected during the period (2006–2100). Most previous studies used 70% of datasets as training and the remaining 30% for testing in various ML models (Tripathi et al. 2006; Anandhi et al. 2012; Gaál 2012). Therefore, 70% of datasets (2006–2069) are considered training
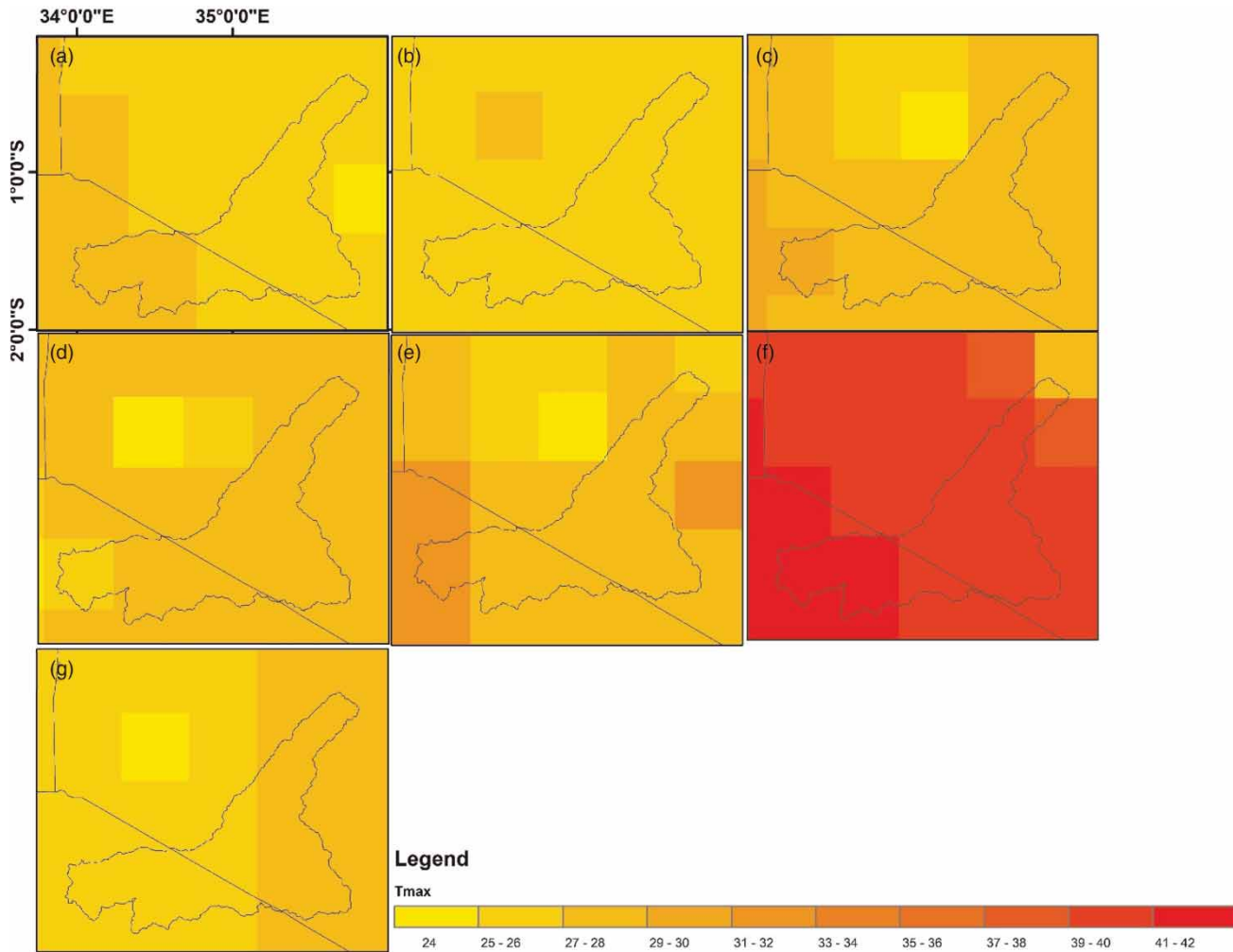
Figure 6 | Future RCMs ensemble of Tmax (a), observation (b) EM RCP 4.5 (2026–2045), (c) EM RCP 8.5 (2026–2045), (d) EM RCP 2.6 (2026–2045), (e) EM RCP 4.5 (2076–2095), (f) EM RCP 8.5 (2076–2095) and (g) EM RCP 2.6 (2076–2095).

data to establish the relationship between observed EM and predicted EM and the rest of the 30% datasets (2070–2100) are used for testing. Four evaluation parameters (R^2, RMSE, MSE and MAE) were used to compare the performance of RF-based predicted EM against observed EM (Table 5). The results indicate that the RF model can predict and reproduce the future precipitation and temperature from each RCM. All the RF model-based EMs provide best agreement with observed EM, and each RCP shows an R^2 value of more than 0.90, which indicates excellent performance. The results also demonstrate the small error with the RMSE of 0.26 °C on average temperature and 0.45 mm on average rainfall.

Figure 8(a) and 8(b) shows the functional relationship between observed EM and predicted EM using the RF model during the testing period (2070–2100). RF-based predicted EM RCP 4.5 is similar to the observed EM for precipitation (Figure 8(a)), Tmax and Tmin (Figure 8(b)). The maximum difference between predicted EM and observed EM datasets is larger during the RCP 8.5 scenario for temperature and precipitation. Also, a precise degree of scattering was observed between estimated EM and predicted EM (Figure 8(c) and 8(d)) during the testing period, which indicates that the RF model can reproduce future rainfall and temperature datasets. The performance of each RCM was evaluated based on their relative importance shown in Figure 8(e) and 8(f). The relative performance of each RCM is automatically assessed in the RF model based on their variance of importance (Wang *et al.* 2018). The results suggest that the HadGEM2_ES RCM had more important variables in the RF model for production precipitation and CLMcom_CCLM for temperature, including RCP 4.5 and RCP 8.5 scenarios.
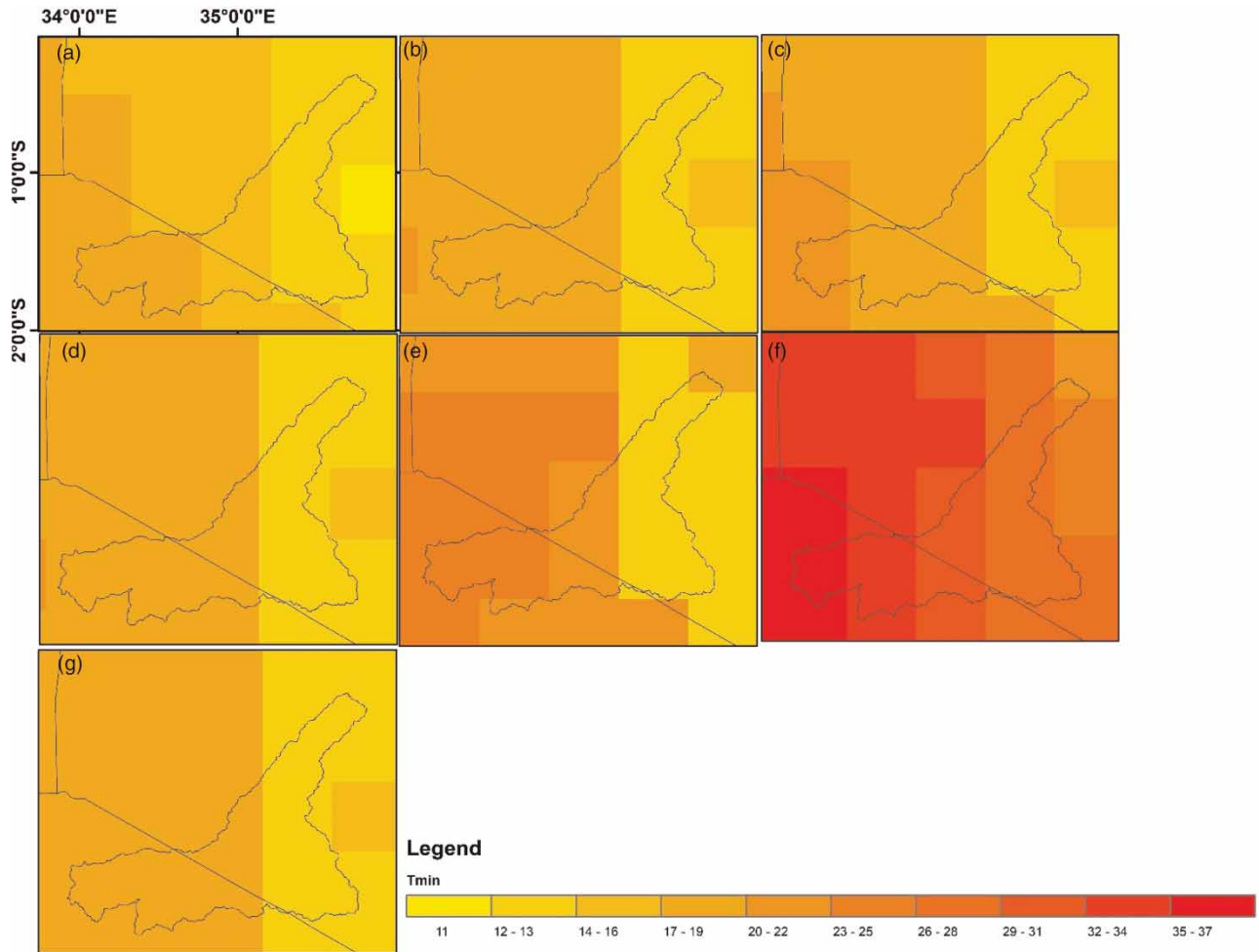
Figure 7 | Future RCMs ensemble of Tmin (a), observation (b) EM RCP 4.5 (2026–2045), (c) EM RCP 8.5 (2026–2045), (d) EM RCP 2.6 (2026–2045), (e) EM RCP 4.5 (2076–2095), (f) EM RCP 8.5 (2076–2095) and (g) EM RCP 2.6 (2076–2095).

4. UNCERTAINTY

The study evaluates various bias correction methods from six RCMs (KNMI, CLMcom – CCLM, CLMcom – CCLM, REMO, CanESM2 and HadGEM2- ES), obtained from three CORDEX (EC – Earth, CCCma and MOHC). The uncertainty of the observation and individual bias-corrected RCMs pattern was captured from monthly mean rainfall and temperature (Figure 9). This study did not trigger the climate change of future rainfall (Räty *et al.* 2014). The bias-corrected monthly rainfall from RCMs showed a comparable increase trend in MJJ and NDJ. The EMs of RCMs and individual RCMs showed uniformity in future climate change scenarios with different magnitude. But strong similarities were found in observation and EMs rainfall. Furthermore, HadGEM2 RCMs projected the highest increasing rainfall trend under every emission scenario. The individual RCMs showed various rainfall patterns in the sub-grid process, which depends upon the grid resolution.

In this study, the results demonstrate the rising rainfall pattern in the near-term and long-term climate under the emission scenario of RCP 4.5 and RCP 8.5. However, this emission scenario described the changes in rainfall that may trigger the convective activity and moisture of the atmosphere over the MRB, although various regional climatic phenomena and local forcing are also the major driving factors for rainfall in the MRB, East Africa. Also, the changes in ITCZ could impact the rainfall variation in the near-term and long-term future. For considering the Paris agreement, the future temperature and rainfall are also investigated under the RCP 2.6 scenario. The RCP 2.6 scenario showed a lower percentage of rainfall increase in the MRB. KNMI (EC- EARTH) and CCLM (EC-EARTH) showed the lower increased rainfall rate under the RCP 4.5 and RCP 8.5 scenario. But the REMO (EC-EARTH) showed a slight increase in the rainfall pattern under the RCP 4.5, RCP
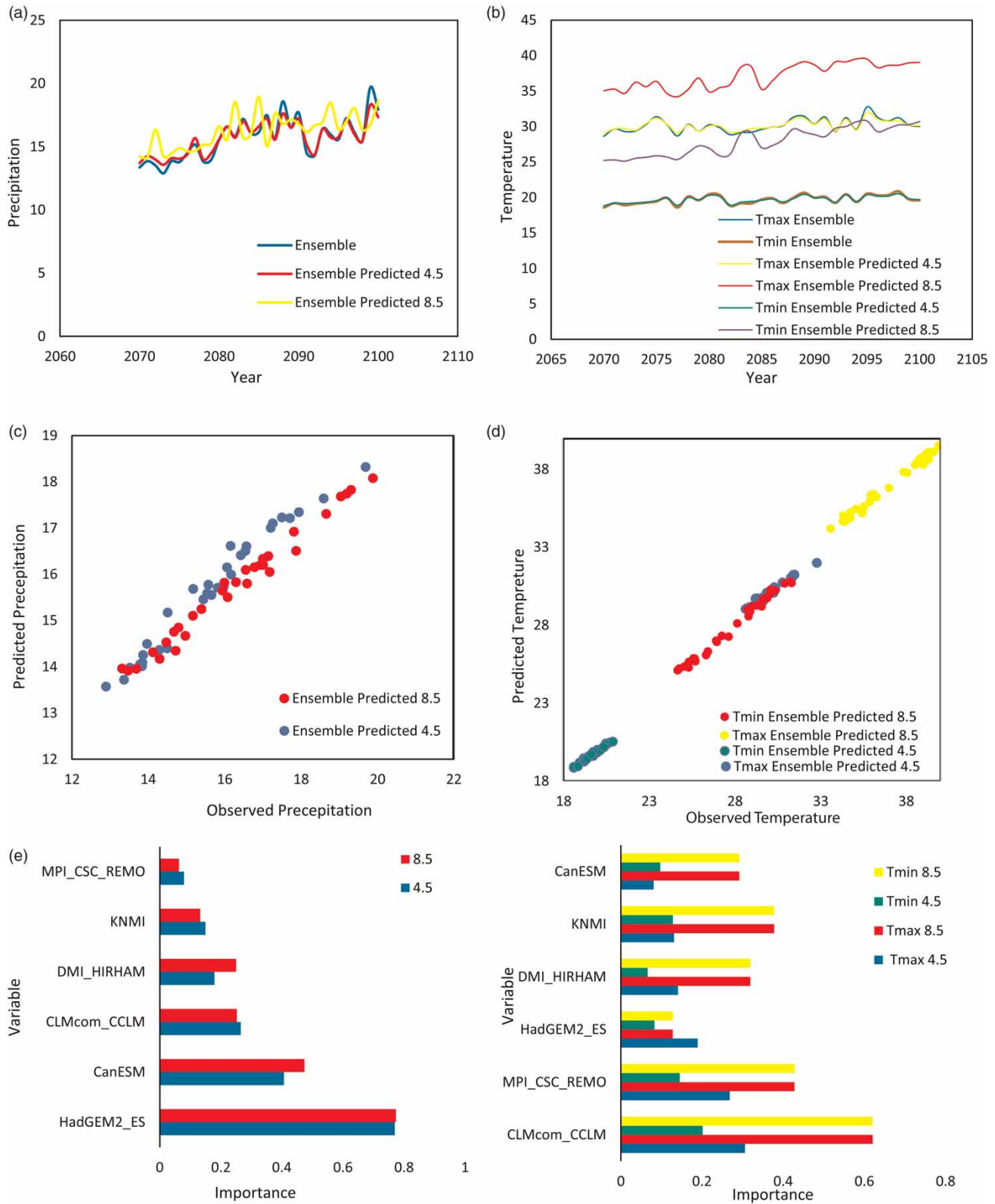
Figure 8 | Time series of observation and predicted ensemble based on RF model (a) precipitation (b) temperature, scatter plot of observed and predicted ensemble for (c) precipitation and (d) temperature, (e) (precipitation) and f (temperature) showing the importance of each RCMs to predict and reproduced the datasets during the testing period (2070–2100).

Table 4 | Future annual mean of TMAX and TMIN for future climate scenarios compared to observation

Scenario	RCMs	TMAX (2026–2045)	Changes in °C	TMAX (2076–2095)	Changes in °C	TMIN (2026–2045)	Changes in °C	TMIN (2076–2095)	Changes in °C
	Observation (1986–2005)	25.47		25.47		13.03		13.03	
RCP 4.5	KNMI (ICHEC-EC-EARTH)	25.67	+0.20	28.39	+2.92	15.47	+2.44	19.31	+6.28
	CLMcom – CCLM (ICHEC-EC-EARTH)	25.50	+0.03	29.03	+3.56	14.72	+1.69	18.81	+5.78
	DMI-HIRHAM (ICHEC-EC-EARTH)	25.52	+0.05	28.27	+2.80	15.51	+2.48	18.54	+5.51
	MPI-CSC-REMO (ICHEC-EC-EARTH)	24.11	−1.47	27.85	+2.38	15.02	+1.99	19.95	+6.92
	SMHI (CCCma-CanESM)	29.91	+4.44	35.51	+10.04	16.23	+3.2	19.93	+6.9
	KNMI (MOHC-HadGEM2-ES)	30.38	+4.91	31.89	+6.42	17.18	+4.15	22.14	+9.11
	EM	26.78	+1.31	30.16	+4.69	15.69	+2.66	19.78	+6.75
RCP 8.5	KNMI (ICHEC-EC-EARTH)	26.199	+0.73	33.82	+8.35	16.53	+3.5	28.1	+15.07
	CLMcom – CCLM (ICHEC-EC-EARTH)	26.42	+0.95	37.44	+11.97	15.94	+2.91	27.15	+14.12
	DMI-HIRHAM (ICHEC-EC-EARTH)	25.76	+0.29	35.33	+9.86	15.6	+2.57	26.69	+13.66
	MPI-CSC-REMO (ICHEC-EC-EARTH)	25.78	+0.31	30.69	+5.22	16.27	+3.24	24.89	+11.86
	SMHI (CCCma-CanESM)	30.55	+5.08	47.77	+22.30	16.84	+3.81	29.27	+16.24
	KNMI (MOHC-HadGEM2-ES)	29.22	+3.75	38.65	+13.18	18.59	+5.56	32.81	+19.78
	EM	27.32	+1.85	37.28	+11.81	16.63	+3.6	28.15	+15.12
RCP 2.6	KNMI (ICHEC-EC-EARTH)	25.21	+0.26	24.85	−0.62	14.89	+1.86	14.73	+1.7
	MPI-CSC-REMO (ICHEC-EC-EARTH)	23.588	−1.882	24.37	−1.1	14.63	+1.6	15.06	+2.03
	KNMI (MOHC-HadGEM2-ES)	28.81	+3.34	28.21	+2.74	16.76	+3.73	17.01	+3.98
	EM	25.87	+0.40	25.81	+0.35	15.43	+2.4	15.59	+2.56

Table 5 | Evaluation matrices of the emission scenarios of RCP 4.5 and RCP 8.5 using the RF model during the testing period (2070–2100)

	PRCP 4.5	PRCP 8.5	Tmax 4.5	Tmax 8.5	Tmin 4.5	Tmin 8.5
R^2	0.93	0.94	0.92	0.97	0.93	0.98
RMSE	0.44	0.46	0.27	0.35	0.17	0.25
MSE	0.194	0.209	0.0714	0.119	0.027	0.063
MAE	0.323	0.376	0.204	0.279	0.143	0.200

8.5 and RCP 2.6 scenario. The importance of large uncertainty was found in various RCMs, which is comparable to a different scenario. However, it found that the global emission will not be reduced by 2030 (10%), which could not be achieved according to the Paris climate agreement (Déqué *et al*. 2017). `

In a similar contrast, the TMAX and TMIN projected high emissions in the near-term and long-term future. The long-term future (2076–2095) scenario showed the high emission of greenhouse gases, increasing the temperature under the RCP 4.5 and RCP 8.5 scenario. But the lower increase rate is observed under RCP 2.6. The IPCC report also contributed that the emission scenario is a major driver of global temperature attributed to global climate change.

In addition, the RF method reduced the uncertainty of the bias-corrected ensemble RCM output and reproduced the future climate variable. The relative importance of each RCM for developing the future rainfall and temperature is beneficial for understanding the contribution of each model adding the value to the ensemble (Wang *et al*. 2018).
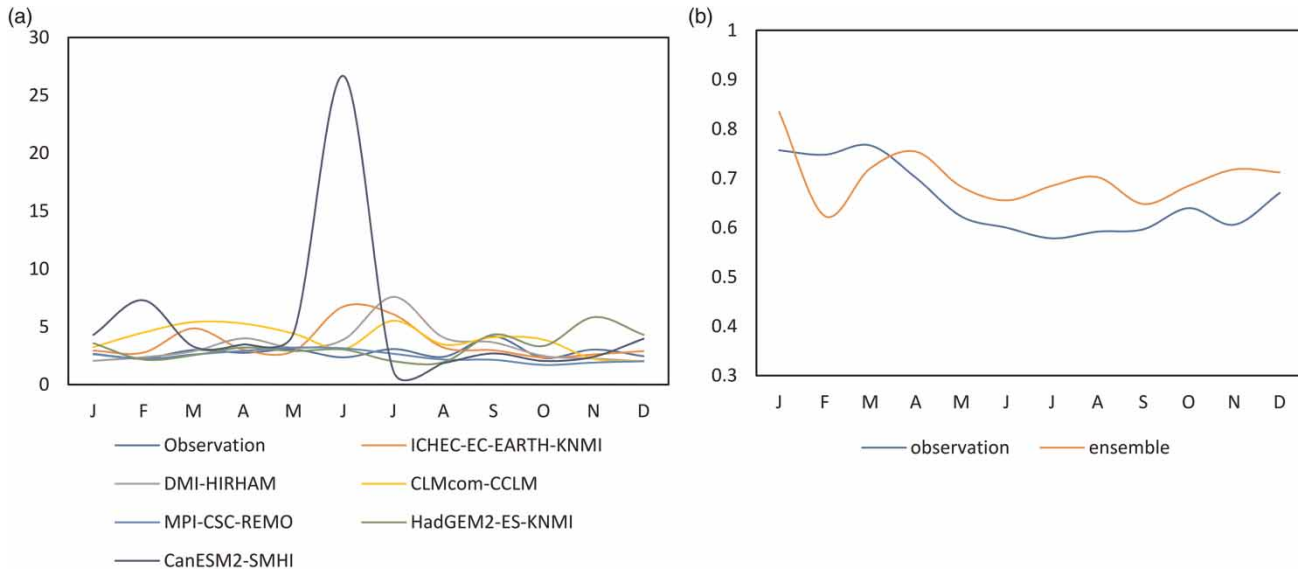
Figure 9 | Uncertainty among the observation and RCMs simulation (a) rainfall and (b) temperature. The comparison with observation and historical bias-corrected period (1986–2005).

5. CONCLUSION

The bias correction method is a major pre-processing step to investigate the climate variable at a regional scale from RCMs results. This study tested four bias correction methods (LS, DC, LI and DM) to adjust the rainfall and temperature retrieved from six historical RCMs. These methods were evaluated in a statistical metrics framework to project future rainfall and temperature at the basin scale. Our results suggest that the DM method showed the best performance in adjusting the rainfall and temperature based on the historical period. Therefore, the DM method was applied to generate future climate scenarios at the MRB. The future climate scenario from the individual RCMs and all EM of RCMs revealed that wetter conditions and high temperature characterize the future climate of the MRB. In a similar context, the DM method can analyze the climatic extremes such as flood and drought and design new policies to reduce the impact of extreme events. In addition, to reduce the uncertainties of the EM of RCMs, we used an ML approach, namely RF. Thus, the method was applied to develop the most robust ensemble-based future climate scenario. The RF ensemble provides comparable results based on the evaluation matrices and perfectly reproduces future rainfall and temperature, although the relative importance was assigned based on the performance of each RCM in RF ensemble simulation. However, this study used DC, LS and LI methods to adjust the RCMs simulation and develop future climate scenarios. Moreover, it will assist in future water resource management and planning in the MRB.

ETHICAL APPROVAL

All the work is in compliance with ethical standards.

COMPETING INTERESTS

None declared.

AVAILABILITY OF DATA AND MATERIALS

Publicly available datasets were analyzed in this study. This data can be found here: https://esgf-node.llnl.gov/search/esgf-llnl/ and https://www.ncdc.noaa.gov/.

REFERENCES

Alemseged, T. H. & Tom, R. 2015 Evaluation of regional climate model simulations of rainfall over the Upper Blue Nile basin. *Atmospheric Research* **161–162**, 57–64. doi:10.1016/j.atmosres.2015.03.013.

Anandhi, A., Srinivas, V. V., Kumar, D. N. & Nanjundiah, R. S. 2012 Daily relative humidity projections in an Indian river basin for IPCC SRES scenarios. *Theoretical and Applied Climatology* **108** (1–2), 85–104. doi:10.1007/s00704-011-0511-z.

Beyer, R., Krapp, M. & Manica, A. 2019 A systematic comparison of bias correction methods for paleoclimate simulations [Preprint]. *Climate of the Past Discussions*. doi:10.5194/cp-2019-11.

Breiman, L. 2001 Random forest. *Machine Learning* **45**, 5–32.

Campozano, L., Tenelanda, D., Sanchez, E., Samaniego, E. & Feyen, J. 2016 Comparison of statistical downscaling methods for monthly total precipitation: case study for the Paute River Basin in Southern Ecuador. *Advances in Meteorology* **2016**, 1–13. doi:10.1155/2016/6526341.

Crawford, J., Venkataraman, K. & Booth, J. 2019 Developing climate model ensembles: a comparative case study. *Journal of Hydrology* **568**, 160–173. doi:10.1016/j.jhydrol.2018.10.054.

Déqué, M., Calmanti, S., Christensen, O. B., Dell Aquila, A., Maule, C. F., Haensler, A. & Teichmann, C. 2017 A multi-model climate response over tropical Africa at +2 °C. *Climate Services* **7**, 87–95. doi:10.1016/j.cliser.2016.06.002.

Dessu, S. B. & Melesse, A. M. 2012 Impact and uncertainties of climate change on the hydrology of the Mara River basin, Kenya/Tanzania: Mara River Basin: climate change and hydrology. *Hydrological Processes*. doi:10.1002/hyp.9434.

Dobler, A. & Ahrens, B. 2008 Precipitation by a regional climate model and bias correction in Europe and South Asia. *Meteorologische Zeitschrift* **17** (4), 499–509. doi:10.1127/0941-2948/2008/0306.

Dosio, A. & Panitz, H.-J. 2016 Climate change projections for CORDEX-Africa with COSMO-CLM regional climate model and differences with the driving global climate models. *Climate Dynamics* **46** (5–6), 1599–1625. doi:10.1007/s00382-015-2664-4.

Dutton, C. L., Subalusky, A. L., Anisfeld, S. C., Njoroge, L., Rosi, E. J. & Post, D. M. 2018 The influence of a semi-arid sub-catchment on suspended sediments in the Mara River, Kenya. *PLoS ONE* **13**, e0192828. https://doi.org/10.1371/journal.pone.0192828.

Favre, A., Philippon, N., Pohl, B., Kalognomou, E.-A., Lennard, C., Hewitson, B. & Cerezo-Mota, R. 2016 Spatial distribution of precipitation annual cycles over South Africa in 10 CORDEX regional climate model present-day simulations. *Climate Dynamics* **46** (5–6), 1799–1818. doi:10.1007/s00382-015-2677-z.

Fenta Mekonnen, D. & Disse, M. 2018 Analyzing the future climate change of Upper Blue Nile River basin using statistical downscaling techniques. *Hydrology and Earth System Sciences* **22** (4), 2391–2408. doi:10.5194/hess-22-2391-2018.

Fiseha, B. M., Setegn, S. G., Melesse, A. M., Volpi, E. & Fiori, A. 2014 Impact of climate change on the hydrology of Upper Tiber River basin using bias corrected regional climate model. *Water Resources Management* **28** (5), 1327–1343. doi:10.1007/s11269-014-0546-x.

Gaál, M. 2012 Modelling the impact of climate change on the Hungarian wine regions using random forest. *Applied Ecology and Environmental Research* **10** (2), 121–140. doi:10.15666/aeer/1002_121140.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. & Engen-Skaugen, T. 2012 Technical note: downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System Sciences* **16** (9), 3383–3390. doi:10.5194/hess-16-3383-2012.

Hernández-Díaz, L., Laprise, R., Sushama, L., Martynov, A., Winger, K. & Dugas, B. 2013 Climate simulation over CORDEX Africa domain using the fifth-generation Canadian Regional Climate Model (CRCM5). *Climate Dynamics* **40** (5–6), 1415–1433. doi:10.1007/s00382-012-1387-z.

Javadinejad, S., Dara, R. & Jafary, F. 2020 Climate change scenarios and effects on snow-melt runoff. *Civil Engineering Journal* **6** (9), 1715–1725. doi:10.28991/cej-2020-03091577.

Kim, J., Waliser, D. E., Mattmann, C. A., Goodale, C. E., Hart, A. F., Zimdars, P. A., Crichton, D. J., Jones, C., Nikulin, G., Hewitson, B., Jack, C., Lennard, C. & Favre, A. 2014 Evaluation of the CORDEX-Africa multi-RCM hindcast: systematic model errors. *Climate Dynamics* **42** (5–6), 1189–1202. doi:10.1007/s00382-013-1751-7.

Kim, D.-I., Kwon, H.-H. & Han, D. 2019 Bias correction of daily precipitation over South Korea from the long-term reanalysis using a composite Gamma-Pareto distribution approach. *Hydrology Research* **50**, 1138–1161. doi:10.2166/nh.2019.127.

Lafon, T., Dadson, S., Buys, G. & Prudhomme, C. 2013 Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods: bias correction of daily precipitation simulated by a regional climate model. *International Journal of Climatology* **33** (6), 1367–1381. doi: 10.1002/joc.3518.

Laprise, R., Hernández-Díaz, L., Tete, K., Sushama, L., Šeparović, L., Martynov, A., Winger, K. & Valin, M. 2013 Climate projections over CORDEX Africa domain using the fifth-generation Canadian Regional Climate Model (CRCM5). *Climate Dynamics* **41** (11–12), 3219–3246. doi:10.1007/s00382-012-1651-2.

Lenderink, G., Buishand, A. & Van Deursen, W. 2007 *Estimates of Future Discharges of the River Rhine Using two Scenario Methodologies: Direct Versus Delta Approach*.

Liang, X.-Z. 2004 Regional climate model simulation of summer precipitation diurnal cycle over the United States. *Geophysical Research Letters* **31** (24), L24208. doi:10.1029/2004GL021054.

Mango, L. M., Melesse, A. M., McClain, M. E., Gann, D. & Setegn, S. G. 2011 Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: results of a modeling study to support better resource management. *Hydrology and Earth System Sciences* **15**, 2245–2258. https://doi.org/10.5194/hess-15-2245-2011.

Mati, B. M., Mutie, S., Gadain, H., Home, P. & Mtalo, F. 2008 Impacts of land-use/cover changes on the hydrology of the transboundary Mara River, Kenya/Tanzania: land use change, hydrology, and the Mara basin. *Lakes & Reservoirs: Research and Management* **13**, 169–177. https://doi.org/10.1111/j.1440-1770.2008.00367.x.

Maurer, E. P. & Pierce, D. W. 2014 Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences* **18** (3), 915–925. doi:10.5194/hess-18-915-2014.

McGinnis, S., Nychka, D. & Mearns, L. O. 2015 A new distribution mapping technique for climate model bias correction. In: *Machine Learning and Data Mining Approaches to Climate Science* (Lakshmanan, V., Gilleland, E., McGovern, A. & Tingley, M., eds). Springer International Publishing, Cham, pp. 91–99. doi:10.1007/978-3-319-17220-0_9.

Meehl, G. A., Arblaster, J. M. & Tebaldi, C. 2005 Understanding future patterns of increased precipitation intensity in climate model simulations: patterns of precipitation extremes. *Geophysical Research Letters* **32** (18). doi:10.1029/2005GL023680.

Mellander, P.-E., Gebrehiwot, S. G., Gärdenäs, A. I., Bewket, W. & Bishop, K. 2013 Summer rains and dry seasons in the Upper Blue Nile Basin: the predictability of half a century of past and future spatiotemporal patterns. *PLoS One* **8**, e68461.

Mendez, M., Maathuis, B., Hein-Griggs, D. & Alvarado-Gamboa, L.-F. 2020 Performance evaluation of bias correction methods for climate change monthly precipitation projections over Costa Rica. *Water* **12** (2), 482. doi:10.3390/w12020482.

Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A. & Ramirez-Villegas, J. 2020 High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Scientific Data* **7** (1), 7. doi:10.1038/s41597-019-0343-8.

Olsson, T., Jakkila, J., Veijalainen, N., Backman, L., Kaurola, J. & Vehviläinen, B. 2015 Impacts of climate change on temperature, precipitation and hydrology in Finland – studies using bias corrected Regional Climate Model data. *Hydrology and Earth System Sciences* **19** (7), 3217–3238. doi:10.5194/hess-19-3217-2015.

Oo, H. T., Zin, W. W. & Thin Kyi, C. C. 2020 Analysis of streamflow response to changing climate conditions using SWAT model. *Civil Engineering Journal* **6** (2), 194–209. doi:10.28991/cej-2020-03091464.

Pinto, I., Lennard, C., Tadross, M., Hewitson, B., Dosio, A., Nikulin, G., Panitz, H.-J. & Shongwe, M. E. 2016 Evaluation and projections of extreme precipitation over Southern Africa from two CORDEX models. *Climatic Change* **135** (3–4), 655–668. doi:10.1007/s10584-015-1573-1.

Räty, O., Räisänen, J. & Ylhäisi, J. S. 2014 Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations. *Climate Dynamics* **42**, 2287–2303.

Rauscher, S. A., Coppola, E., Piani, C. & Giorgi, F. 2010 Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Climate Dynamics* **35** (4), 685–711. doi:10.1007/s00382-009-0607-7.

Roy, T., Valdés, J. B., Lyon, B., Demaria, E. M. C., Serrat-Capdevila, A., Gupta, H. V., Valdés-Pineda, R. & Durcik, M. 2018 Assessing hydrological impacts of short-term climate change in the Mara River basin of East Africa. *Journal of Hydrology* **566**, 818–829. doi:10.1016/j.jhydrol.2018.08.051.

Sarr, M. A., Seidou, O., Tramblay, Y. & El Adlouni, S. 2015 Comparison of downscaling methods for mean and extreme precipitation in Senegal. *Journal of Hydrology: Regional Studies* **4**, 369–385. doi:10.1016/j.ejrh.2015.06.005.

Schmidli, J., Goodess, C. M., Frei, C., Haylock, M. R., Hundecha, Y., Ribalaygua, J. & Schmith, T. 2007 Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research* **112** (D4), D04105. doi:10.1029/2005JD007026.

Segal, M. R. 2003 Machine Learning Benchmarks and Random Forest Regression. 14.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L. 2007 Climate change 2007: The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change (S. D. Solomon, ed.). IPCC, Geneva, Switzerland.

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R. & Demaria, E. M. C. 2017 Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences* **21** (6), 2649–2666. doi:10.5194/hess-21-2649-2017.

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J. & Evans, J. P. 2015 How does bias correction of regional climate model precipitation affect modelled runoff? *Hydrology and Earth System Sciences* **19** (2), 711–728. doi:10.5194/hess-19-711-2015.

Tripathi, S., Srinivas, V. V. & Nanjundiah, R. S. 2006 Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330** (3–4), 621–640. doi:10.1016/j.jhydrol.2006.04.030.

Tschöke, G. V., Kruk, N. S., de Queiroz, P. I. B., Chou, S. C. & de Sousa Junior, W. C. 2017 Comparison of two bias correction methods for precipitation simulated with a regional climate model. *Theoretical and Applied Climatology* **127** (3–4), 841–852. doi:10.1007/s00704-015-1671-z.

United States Agency for International Development (USAID) 2019 *Vulnerability and Adaptation in the Mara River Basin*. Chemonics International Inc., Washington, DC, p. 103.

Urrutia, R. & Vuille, M. 2009 Climate change projections for the tropical Andes using a regional climate model: temperature and precipitation simulations for the end of the 21st century. *Journal of Geophysical Research* **114** (D2), D02108. doi:10.1029/2008JD011021.

Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A. & Yu, Q. 2018 Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *International Journal of Climatology* **38** (13), 4891–4902. doi:10.1002/joc.5705.

Worku, G., Teferi, E., Bantider, A. & Dile, Y. T. 2020 Statistical bias correction of regional climate model simulations for climate change projection in the Jemma sub-basin, upper Blue Nile Basin of Ethiopia. *Theoretical and Applied Climatology* **139** (3–4), 1569–1588. doi:10.1007/s00704-019-03053-x.

Zhang, B., Shrestha, N., Daggupati, P., Rudra, R., Shukla, R., Kaur, B. & Hou, J. 2018 Quantifying the impacts of climate change on streamflow dynamics of two Major Rivers of the Northern Lake Erie Basin in Canada. *Sustainability* **10** (8), 2897. doi:10.3390/su10082897.