Article

Assessment of Climate Models Performance and Associated Uncertainties in Rainfall Projection from CORDEX over the Eastern Nile Basin, Ethiopia

Sadame M. Yimer [1,2,3], **Abderrazak Bouanani** [1], **Navneet Kumar** [4,*], **Bernhard Tischbein** [4] **and Christian Borgemeister** [4]

1   Laboratory 25, Department of Hydraulics, Faculty of Technology, Abou Bakr Belkaid University of Tlemcen, B.P. 119 | Pôle Chetouane, Tlemcen 13000, Algeria; e.sadam91@gmail.com (S.M.Y.); a_bouananidz@yahoo.fr (A.B.)
2   Department of Natural Resources Management, Wollo University, Dessie P.O. Box 1145, Ethiopia
3   Department of Water Engineering, Pan African University Institute of Water and Energy Sciences (Including Climate Change) (PAUWES), c/o University of Tlemcen, B.P. 119 | Pôle Chetouane, Tlemcen 13000, Algeria
4   Department of Ecology and Natural Resources Management, Center for Development Research (ZEF), University of Bonn, Genscherallee 3, 53113 Bonn, Germany; tischbein@uni-bonn.de (B.T.); cb@uni-bonn.de (C.B.)
*   Correspondence: nkumar@uni-bonn.de

Abstract: The adverse impact of climate change on different regionally important sectors such as agriculture and hydropower is a serious concern and is currently at the epicentre of global interest. Despite the extensive efforts to project the future climate and assess its potential impact, it is surrounded by uncertainties. This study aimed to assess climate models' performance and associated uncertainties in rainfall projection over the eastern Nile basin, Ethiopia. Seventeen climate models from Coordinated Regional Climate Downscaling Experiment (CORDEX) and their four ensemble models were evaluated in terms of their historical prediction performance (1986–2005) and future simulation skill (2006–2016) at rainfall station (point location), grid-scale (0.44° × 0.44°) and basin scale. Station-based and spatially interpolated observed rainfall data were used as a reference during climate model performance evaluation. In addition, CRU data was used as an alternative reference data to check the effect of the reference data source on the climate models evaluation process. As the results showed, climate models have a large discrepancy in their projected rainfall and hence prior evaluation of their performance is necessary. For instance, the bias in historical mean annual rainfall averaged over the basin ranges from +760 mm (wet bias) to −582 mm (dry bias). The spatial pattern correlation (r) of climate models output and observed rainfall ranges from −0.1 to 0.7. The ensemble formed with selected (performance-based) member models outperforms the widely used multi-model ensemble in most of the evaluation metrics. This showed the need for reconsidering the widely used multi-model approach in most climate model-based studies. The use of CRU data as a reference resulted in a change in the magnitude of climate model bias. To conclude, each climate model has a certain degree of uncertainty in the rainfall projection, which potentially affects the studies on climate change and its impact (e.g., on water resources). Therefore, climate-related studies have to consider uncertainties in climate projections, which will help end-users (decision-makers) at least to be aware of the potential range of deviation in the future projected outcomes of interest.

Keywords: climate change; climate modelling; CORDEX-RCMs; climate uncertainty; rainfall projection; climate model ensemble; Eastern Nile Basin

1. Introduction

According to the Intergovernmental Panel on Climate Change [1], there has been a sharp increase in the earth's surface temperature over the last half-century, which coincides

with the massive industrial revolution around the world. The increase in greenhouse gas (GHGs) emissions from anthropogenic activities is undoubtedly admitted as the major cause of global warming and climate change [2–5]. The complex inter-relations between the climate system and several natural processes of the earth such as the hydrological cycle, biodiversity, and health of the ecosystems [6], etc., extends the potential impacts of climate change to the existence of several endeavours of humans in a very diverse way [1]. Unfortunately, Africa is identified as one of the most vulnerable regions to impacts of climate change [1,7,8] and its terrible consequence such as flooding, prolonged drought, and heat wave phenomena have been witnessed frequently over the last few decades.

Several researchers have been working to understand the impact of climate change on different regionally important sectors such as water resources availability, economic activities, agriculture, health, social instability, etc. [9–14]. However, climate-related studies are surrounded by uncertainties in climate projections. Uncertainty can be defined as the lack of confidence about something ranging from small doubts and minor impressions to a complete lack of definite knowledge [15,16]. The concept of uncertainty has a vital role in global environmental changes research including climate change and its impact studies and leads to difficulties in decision making. Hence, quantifying the extent of uncertainties in climate model projections and understanding its impact has recently become an important component in many climate-related studies [15–20]. Policy/decision-makers, who are interested to develop a climate-resilient strategic plan, should be made aware of the future potential variation to the projected climate and its impact [21], which emphasized the need for an in-depth uncertainty study at the regional scale.

In climate projection, there are three primary sources of uncertainties: (i) the natural climate variability in the climate system without influence from an external force, (ii) the incomplete knowledge of the future trajectories variables that can potentially affect the climate system, which is mostly noted as Green House Gases (GHGs) emissions [4,15,19,20,22–24], and (iii) climate modelling uncertainties that arise from incomplete knowledge regarding how the changes in climate forcing factors translated into changes in the climate system [4,16,22,23,25]. The first two sources of uncertainty are more or less unavoidable while the latter source is the most viable to characterize and potentially can be accounted for to reduce the uncertainties [26]. Global Circulation Models (GCMs) are computer programs that mathematically represent the fundamentals of the earth's climate system. In the field of climate science and related applications, GCMs have been acknowledged worldwide as the most powerful tool for providing future climate information in response to various forcing factors up to a century ahead [4,8,27–30]. Despite successes in the climate modelling efforts, the complexity of our Earth's climate system [31] imposes limitations from being completely understandable and represented by the existing climate models [32]. In addition, the coarse resolution of GCMs restricts their capability to capture important climate phenomena such as the influence of orographic and vegetation heterogeneity on climate systems at a regional to local scale. Furthermore, due to the differences in theories, initial boundary conditions, and overall algorithms applied, different climate models provide different climate projections [6,16] which are often reported as significant sources of uncertainty in the climate projections, particularly referring to rainfall projection [33]. Hence, no single global climate model is perfect, and all climate models exhibit different levels of uncertainties accompanying their climate modelling and climate projections. Hence, to reduce the uncertainties that arise from climate models choice, researchers often use the ensemble approach where the projection result from more than one climate model is merged. The multi-model ensemble with consideration of all available climate models (regardless of their individual performance) is the widely used approach in many published research works [8,34]. However, recent studies showed that ensembles formed with selective member models have an outperforming skill compared to those formed with all the available climate models together [25,26], which is questioning the use of the blind multi-model ensemble approach. Hence, it is important to analyse individual climate models' performance before ensemble formation.

In climate model evaluation, the most frequently asked question and may be difficult to answer completely is what does the model's historical performance means for their future projection accuracy? Most of the evaluation studies focused on the comparison of climate models' historical prediction skills with less attention to validating their ability in simulating future climate [6]. However, as far as future climate is concerned, exploring the possibilities to crosscheck climate model's future performance is quite worthy and pushes one step ahead in understanding better the uncertainties caused by the model choice. This can be achieved by forcing climate models to do a projection for the period to which observed climate data is already existing and more importantly the time-dependent inputs for climate projection of this period should not be part of the initial climate model calibration process. On the other hand, given the importance of climate model evaluation and validation, one major challenge to this effort is the presence of uncertainty in observational or reference climate datasets and this often made the climate model evaluation process even further complicated [21,35,36]. In some cases, observational data uncertainty can even dominate uncertainty from a climate model [21], which points out the importance of the proper selection of reference data for reliable climate model evaluation.

The Northern Ethiopian highland is one of the East African sub-regions that is characterized by the high spatially and temporally varying climate conditions in particular rainfall [28]. Due to the importance of understanding climate characteristics and their potential impact in different sectors (e.g., agriculture, flood and drought management, hydropower, etc.), extensive research has been conducted in the region, e.g., [37,38]. Despite some efforts of evaluating different GCMs and regional climate models (RCMs) at the regional levels like the East African region or some at the country level, there were no prior studies reported in the area that explicitly assessed the potential uncertainties in the climate projections due to climate models choice. Yet there is a strong remark on the need for a spatially explicit climate model evaluation and selection of the best performing climate model to the region of interest. Hence, to fulfil this research gap, the main objective of this study is to evaluate the performance of different climate models, selected ensembles, and multi-model ensembles (all together) and analyse the associated uncertainties in rainfall projection over the Eastern Nile basin (Tekeze River basin), Ethiopia. To achieve this, more specific objectives were derived. These include: (i) evaluation of the dynamically downscaled GCMs under the CORDEX-Africa domain in predicting the historical rainfall and their future simulation skill (as a potential validation step) in the study region, (ii) development of an approach to form a climate model ensembles using selective (performance-based) member climate models and compare their performance with the widely used multi-model ensemble, and (iii) assessment of uncertainties in reference dataset for climate model evaluation process.

2. Materials and Methods

2.1. Study area Description

The study area is the Tekeze river basin, which is one of the major 12 river basins in Ethiopia. It is located in the north-western part of the country between 11°36″ to 14°18″ N and 37°30″ to 39°54″ E (Figure 1). It is a transboundary river basin that drains most of the northern Ethiopian highlands into the Atbara River in Sudan, which later joins the Nile River system as the last tributary before reaching the Mediterranean Sea. The basin has an elevation ranging from 4540 m.a.s.l. up to 825 m.a.s.l at the Embamdre hydrological station located with a GPS reading of 13°43′48″ N of Latitude and 38°12′00″ E of longitude (Figure 1). At this drainage point, the basin covers an area of around 45,589 km². The complex topographic features of the basin such as the rugged mountainous nature of the landscape and heterogeneity of the vegetation cover and its consequent interaction with large-scale climate forcing mechanisms contributed to the high spatial variability of rainfall distribution in the area [28]. The study area has a unimodal rainfall pattern (only one rainfall peak) with light rainfall from March to May followed by a long rainy season from June to September. The seasonal distribution of rainfall is strongly linked with

the north-south movement of the intertropical convergence zone (ITCZ) [28]. Given the region's physiographic pattern and distinct relation between altitude and temperature, the region's mean annual temperature ranges from less than 10 °C in the elevated mountainous area to 35 °C and above in the low land area. Rainfed agriculture (mainly wheat, maize, and sorghum) is the dominant land-use type and due to its climate sensitivity, it is the most impacted and vulnerable sector in the region due to climate change.
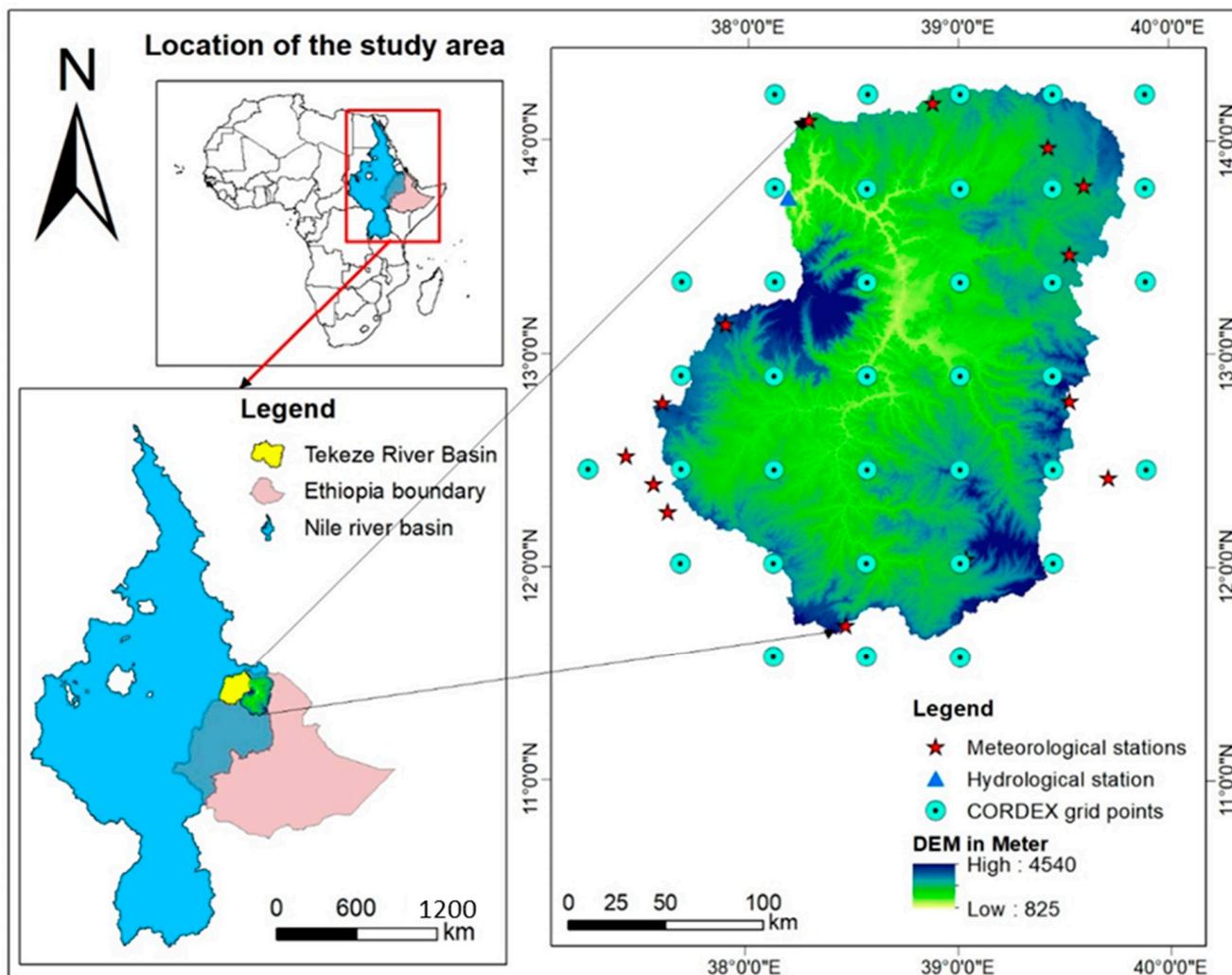

Figure 1. Location of the study area along with digital elevation model (DEM), location of hydrological and meteorological stations, and CORDEX grid points.

2.2. Observed Rainfall Data

Station-based daily observed rainfall data were collected from the National Meteorological Agency of Ethiopia (NMAE). Fourteen stations that have a relatively long record period (1986–2016) for daily rainfall were selected (Figure 1). The collected rainfall data were subjected to quality control procedures such as rainfall data gap-filling and homogenization before being used in this analysis [39].

2.3. Climate Model Data

Coordinated Regional Downscaling Experiment (CORDEX) is a climate experiment program established by the World Climate Research Program (WCRP) to provide regionally downscaled climate datasets to all parts of the world [8]. CORDEX experiment is the first effort of its kind that focuses on the whole of Africa's climate by providing high-resolution climate data processing and archiving under the CORDEX-Africa domain [8]. In this study,

daily time-step for historical (1951 to 2005) and future projected rainfall datasets (2006 to 2100) from the second phase of CORDEX output, which are the dynamically downscaled Coupled Model Inter-comparison Project 5 (CMIP5) GCMs using different RCMs (from M1 to M17 as shown in Table 1), were acquired from CORDEX-Africa domain archive. Hence, this study considered a daily time step climate (rainfall) projection from 17 climate models.

Table 1. The list of climate models that have been collected from the CORDEX-Africa domain.

Model Code	Driving GCM Name	Short GCM Name	RCM
M1	CNRM-CERFACS-CNRM-CM5	CNRM-CM5	CLMcom-CCLM4-8-17
M2	CNRM-CERFACS-CNRM-CM5	CNRM-CM5	SMHI-RCA4
M3	ICHEC-EC-EARTH	EC-EARTH	KNMI-RACMO22T
M4	ICHEC-EC-EARTH	EC-EARTH	DMI-HIRHAM5
M5	ICHEC-EC-EARTH	EC-EARTH	CLMcom-CCLM4-8-17
M6	ICHEC-EC-EARTH	EC-EARTH	MPI-CSC-REMO2009
M7	ICHEC-EC-EARTH	EC-EARTH	SMHI-RCA4
M8	MPI-M-MPI-ESM-LR	MPI-ESM-LR	CLMcom-CCLM4-8-17
M9	MPI-M-MPI-ESM-LR	MPI-ESM-LR	MPI-CSC-REMO2009
M10	MPI-M-MPI-ESM-LR	MPI-ESM-LR	SMHI-RCA4
M11	NCC-NorESM1-M	NorESM1-M	DMI-HIRHAM5
M12	CCCma-CanESM2	CanESM2	SMHI-RCA4
M13	CSIRO-QCCCE-CSIRO-Mk3-6-0	CSIRO-Mk3-6-0	SMHI-RCA4
M14	IPSL-IPSL-CM5A-MR	IPSL-CM5A-MR	SMHI-RCA4
M15	MIROC-MIROC5	MIROC5	SMHI-RCA4
M16	NCC-NorESM1-M	NorESM1-M	SMHI-RCA4
M17	NOAA-GFDL-GFDL-ESM2M	GFDL-ESM2M	SMHI-RCA4

Climate model performance varies between the regions and thus evaluation, inter-comparison, selection of the best climate model (representing the climate system) of a particular area, and ensembles of a selected climate model are required and have been performed in this study.

The spatial resolution of selected CORDEX data is 0.44° (nearly 50 km) and temporal resolution is daily time steps. The link to download the CORDEX data: https://climate4impact.eu/impactportal/general/index.jsp (accessed on 10 February 2018). Climate data from CORDEX are quality controlled and can be used according to the terms of use (http://wcrp-cordex.ipsl.jussieu.fr/, accessed on 1 May 2022).

2.4. Ensemble Formation

Ensemble in climate modelling is a means of merging more than one climate model based on different statistical values such as ensemble variance, ensemble mean, ensemble sum, etc. In many applications of climate models, the use of ensemble mean (Ensmean) is the most widely recommended approach, where each of the candidate models is equally weighted (as model democracy) and hence, this study also considered the Ensmean approach. In this study, two categories of the ensemble were considered. The first one is the multi-model ensemble (M18) where all the 17 climate model projections from CORDEX were included, which is like a "blind ensemble". In the second category, three selected climate model ensembles were formed with different numbers of selected member climate models. Here, the member climate models were evaluated and selected based on their performance, and then ensembles were formed between the selected member models. These selected model ensembles (coded as M19, M20, and M21) were formed based on the recommendation from IPCC guidelines on the use of systematically selected candidate models for ensemble formation. Some researchers have also strongly advocated forming the ensemble from a small subset of climate models that have good performance in their past predictions rather than using the blind multi-model ensemble technique [40,41]. In this study, the selected climate model ensembles are formed after evaluating the entire individual models at station-scale, grid-scale, and basin-scale and the primary evaluation

results have been used as a basis to select individual member climate models for each of these ensemble models. Six criteria were set to filter candidate models for ensemble formation: (i) bias in basin average rainfall, (ii) bias in annual average rainfall, (iii) bias in monthly average rainfall, (iv) standard error (StE), (v) root mean square error (RMSE) and (vi) spatial pattern correlation (*r*). Equations are given from Equations (1)–(6) respectively. In the bias calculation, the absolute value of the bias in each year and/or month was considered to avoid the offsetting effect from opposite sign bias in the given model, which otherwise can hide the actual bias of the model. In each of these criteria, a candidate model was selected. In the first three criteria, a bias percentage of 20% was used as a cutting edge or threshold to decide whether the given model can be a candidate or not. The basis for this threshold is adapted from the WMO recommendation for missing climate data [42]. For StE (standard error) and RMSE (root mean square error), models with less than or equal to the average StE and average RMSE value respectively from all individual models were considered as a candidate. Regarding the spatial pattern correlation (*r*), models with $r \geq 0.5$ were given a candidacy flag.

$$Bias\ in\ basin\ average\ Rf = \left| \left(\frac{Rf_p - Rf_o}{Rf_o} \right) \right| \tag{1}$$

$$Bias\ in\ annual\ average\ Rf = \frac{\Sigma \left(\left| \frac{Rf_p - Rf_o}{Rf_o} \right| \right)}{N} \tag{2}$$

$$Bias\ in\ monthly\ average\ Rf = \frac{\Sigma \left(\left| \frac{Rf_p - Rf_o}{Rf_o} \right| \right)}{12} \tag{3}$$

$$StE = \frac{\sigma}{\sqrt{n}} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Rf_p - Rf_o \right)^2} \tag{5}$$

$$r = \frac{n \sum Rf_o Rf_p - \sum Rf_o \sum Rf_p}{\sqrt{\left[n \sum Rf_o^2 - (\sum Rf_o)^2 \right] \left[n \sum Rf_p^2 - (\sum Rf_p)^2 \right]}} \tag{6}$$

where Rf_p is the predicted basin or annual or monthly rainfall from each climate model, Rf_o is the observed basin or annual or monthly rainfall data, σ is the standard deviation of the time series data, "*n*" is the number of time-series data while "*N*" is the number of years considered.

Following the model screening process, three ensembles are formed. The first ensemble (M19) is formed using models that have been selected in at least three of the above six criteria while the second ensemble (M20) includes those selected in at least five criteria and the third ensemble (M21) is formed using the best models across all the six criteria mentioned above. Such ensemble formation with a different hierarchy of member models' performance allows for explicitly evaluating the importance of systematic model selection prior to ensemble formation. Therefore, in total, 21 climate models (17 individual and 4 ensemble models) are considered in this study. The overall methodological framework applied in this research is summarized in Figure 2.
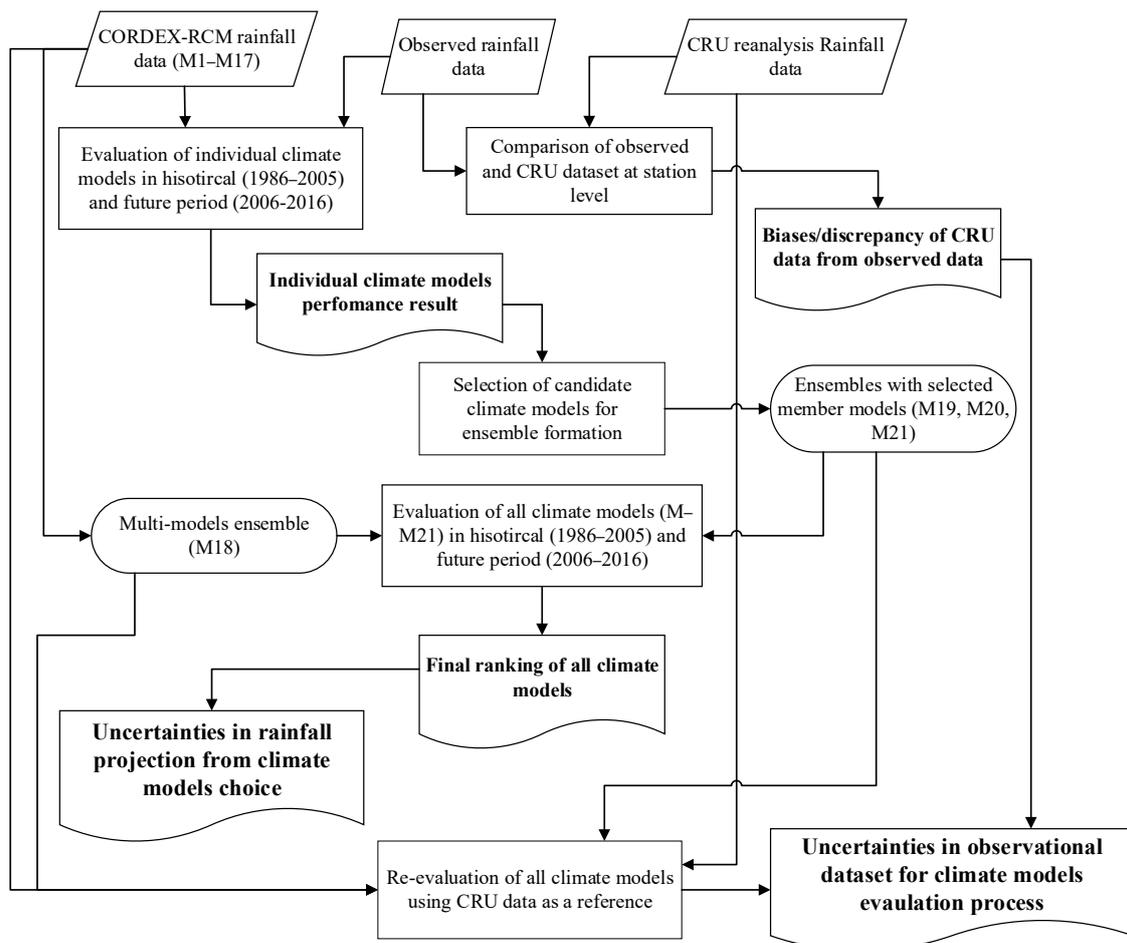
Figure 2. Flow diagram showing the overall methodological frameworks applied in this research (Part of the figure components that are written in bold are the results from the previous analysis in the work flow).

2.5. Evaluation of Individual Climate Models and Climate Model Ensembles for Rainfall Projection and Quantify the Associated Uncertainties

2.5.1. Climate Models Evaluation Periods

In this study, the comparison between individual climate models (dynamically down-scaled GCMs from the CORDEX Africa domain), multi-model ensembles, and selected ensembles was performed in two steps. First, climate models were evaluated in their historical prediction performance using their historically generated rainfall data and the corresponding observed data at three spatial scales i.e., at the station (point location), grid ($0.44° \times 0.44°$), and basin level. Due to the limitation of available observed data, the historical comparison was restricted from 1986 to 2005 only, which is 20 years and fair enough to capture the individual's model performance [6]. In the CMIP5 phase, GCMs' future simulations started from 2006 until the end of the 21st century. Standing on the present, the starting period of future climate simulation by CMIP5 GCMs [43], which is 2006, gives the advantage to have an intersection period between future simulated climate models data and actual observed data on the ground. This provided an opportunity for potential validation of the climate model's performance in simulating the future climate.

Nevertheless, unlike the historical evaluation process, validation of climate model performance using their future simulation dataset is not a straightforward process. Because, in future climate projection, there are four alternate experiments for each of the Representative Concentration Pathway (RCP) emission scenarios (RCP2.6, RCP4.5, RCP6.0, and RCP8.5) given by IPCC. However, in this model validation step, the rainfall projection

from the RCP8.5 scenario was used. The same choice has been used by [6]. This choice is due to the RCP8.5 scenario assumption of a continuing trend of GHGs emissions observed in the early 21st century (business as usual scenario), which is assumed to be similar to the actual emission situations on the ground so far. This case is based on the assumption that there was no significant intervention globally towards the reduction of GHGs emissions to the extent that can disprove the unlikeness of the RCP8.5 scenario over the last decade. In addition, today's climate is affected not only by current forcing but also by historical forcing factors in particular past emission conditions [32], which strengthens the above assumption.

Hence in this study, due to overlapped time advantage coupled with the above assumptions, the remaining observed rainfall datasets from 2006 to 2016 were used as a benchmark to which the corresponding future simulated climate models data from the RCP8.5 scenario were compared as a validation step. Although the duration of rainfall data seems short (only 11 years), it is assumed to be informative enough to observe systematic biases among the candidate climate models. According to [44], the difference between model values predominantly stems from their respective simulation techniques rather than long-term climate variability.

2.5.2. Climate Models Performance Evaluation Techniques

The evaluation process was carried out at three spatial scales: (i) at the rainfall station (point location), (ii) grid-scale ($0.44° \times 0.44°$), and (iii) basin scale. The station-level comparison was performed in both daily and monthly time steps, the grid-level comparison was using mean annual rainfall while the basin-level comparison considered mean monthly and mean annual rainfall.

(i) At rainfall station level (point) comparison: observed rainfall data from the rainfall station and the corresponding grid data from the climate model projected rainfall were compared. The station-level comparison is targeted to understand the skill of the climate model's performance in capturing the temporal variability of rainfall at each station. Four commonly used statistical evaluation techniques which are Root Mean Square Error (RMSE), correlation coefficient (r), Mean Absolute Error (MAE), and Percent Bias (PBIAS) were used [45]. The formula for the first two techniques is already given in Equations (5) and (6), respectively, while the formula for the latter two are provided in Equations (7) and (8) respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| Rf_p - Rf_o \right| \tag{7}$$

$$PBIAS = \left[\frac{\sum_{i=1}^{n} \left(Rf_o - Rf_p \right) * (100)}{\sum_{i=1}^{n} Rf_o} \right] \tag{8}$$

Since the comparison is carried out at each station separately and using more than one metric, different climate models can have different performance results at different stations and even across the evaluation techniques considered. Therefore, to rank the candidate climate models based on their overall performance in all the 14 rainfall stations and metrics used, each climate model was given a rank for each of the evaluation metrics in every station separately. This means, depending on each of the climate model's performance in each of the metrics at the given station, they were rated from 1 (best performing) to 17 (poorest performing). Here, 1 to 17 is the total number of individual climate models that have been evaluated initially. Then the average score of the given climate model from all metrics at the given station is calculated using Equation (9).

$$A_{si} = \frac{S_1 + S_2 + S_3 + S_4}{4} \tag{9}$$

where, A_{si} is the average score of the given climate model in station "*i*" from the four evaluation metrics, S_1, S_2, S_3 and S_4 are the score/rank of the given climate model in evaluation metrics 1, 2, 3 and 4 respectively.

Then, the score range in one station runs from 1 (which is $1 \times 4/4$) to 17 (which is $17 \times 4/4$). Finally, the overall performance of each climate model considering all rainfall stations together was computed using Equation (10). Similarly, the best performing climate model under all four evaluation metrics and fourteen rainfall stations should have a score of 1 (which is $1 \times 14/14$) while the poorly performing gets 17 (which is $17 \times 14/14$). A similar comparison approach has been applied in other research such as [46].

$$A_s = \frac{\sum_{i=1}^{14} A_{si}}{14} \tag{10}$$

where, A_s is the average score of the given climate model to all stations and from the four evaluation metrics.

It should be noted that when the blind ensemble (M18) and newly formed three ensembles (M19, M20, M21) joins the comparison later, the same evaluation and overall scoring procedure will be applied; yet, the score range becomes from 1 to 21.

(ii) In grid-level comparison: First the 14 station-based observed rainfall data were subjected to three spatial interpolation techniques (Inverse Distance Weighting method (IDW), Ordinary Kriging (OK), and Empirical Bayesian Kriging (EBK). Then, their performance was evaluated using three metrics given in Equations (5), (7) and (8). Five spatially well-distributed rainfall station points were used for this evaluation process. Finally, the spatially interpolated observed data from the best performing method was chosen to be used as reference data after being resampled into a similar grid size (0.44°) as that of CORDEX data using a bilinear interpolation technique. The details of the theoretical background and equations of these spatial interpolation techniques can be found in [47].

The grid values of mean annual rainfall were extracted from all climate models and compared with the corresponding grid values from the spatially interpolated and equally gridded observed rainfall dataset. In this comparison phase, the Taylor diagram which is the widely used model evaluation technique has been applied [48]. This comparison is intended to understand the skill of climate models in capturing the spatial variability of rainfall in the region.

(iii) The basin-scale comparison: Here the mean annual and mean monthly rainfall were used to evaluate the climate model's skill in capturing the inter and intra-annual variability and mean annual rainfall in the basin. The applied methodologies mentioned above were the same for historical (1986–2005) and future validation (2006–2016) steps.

2.5.3. Effect of the Reference Data Source on the Climate Models Evaluation Process

The choice of reference dataset during climate model evaluation determines the magnitude of climate model bias or level of agreement. Whenever available, observed rainfall data from the ground station is considered the most reliable reference dataset. However, in many cases where observed data is limited or scarce, different reanalysis gridded datasets are commonly used in climate model performance evaluation. In this regard, the Climate Research Unit (CRU) data is one of the widely used rainfall datasets in particular for climate model evaluation, e.g., [49,50]. Thus, to analyse the potential discrepancy in model performance due to the choice of different reference datasets, the Climate Research Unit (CRU) data has been considered in this study as one alternative reference rainfall dataset. CRU data was obtained from the University of East Anglia's Climate Research Unit (CRU TS4.03) [51]. The CRU TS4.03 data version covers a period from 1901 to 2018 and has a spatial resolution of 0.5°. Its reliable accuracy has been evidenced by several researchers such as [52,53]. To match the same spatial resolution with climate models from CORDEX, the CRU dataset was re-gridded into a resolution of 0.44° by 0.44° using a bilinear interpolation method, which is the widely used resampling technique e.g., [8,50].

The uncertainty in the model evaluation due to the choice of reference data was evaluated in two ways. First, observed monthly rainfall from stations was compared with their corresponding value from CRU data over the period of 1986–2005 using the error

metrics given in Equations (5), (7) and (8). Secondly, climate model performance was assessed at the grid and basin level using CRU data as a reference dataset.

3. Results and Discussion

3.1. Evaluation of Climate Models Performance and Associated Uncertainties in Rainfall Projection

The results of the evaluation of the climate model's performance and associated uncertainties in rainfall projection are presented in two sections. The first analysis (Section 3.1) considered all individual climate models (M1 to M17) while the latter (Section 3.3) includes all the aforementioned models, the multi-model ensemble (M18), and the three selected ensemble models (M19, M20, and M21) that are formed following the first evaluation results (see Section 3.2).

3.1.1. Climate Models Performance over the Historical Period (1986–2005)
At Station Level

The results of the historical time series comparison (averaged related to daily and monthly time-step) for all the 14 stations and 4 metrics considered are given in Table 2 (in the 2nd and 3rd columns). As the results indicate, the first six best-performed climate models in daily basis comparison (with their score in bracket) were M9 (6.3), M12 (6.4), M16 (6.6), M2 (7.3), M6 (7.4), and M13 (7.9). On the other hand, the least performed climate model was M1 (12.8) followed by M4 (12.2), M11 (12.0), and M5 (11.5). In a monthly time-step analysis, M9 was again the best performing individual model with a score of 3.4 followed by M8 (6.4), M6 (7.1), and M10 (7.6), while M1 again showed the least performance followed by M14, M4, M11, and M5. It is important to note that M9, M8, and M10 ranked first, second, and fourth respectively at the monthly time step have the same driving GCMs (MPI-ESM-LR). This result indicates that not only RCMs or downscaling methods that matter the most. The driving GCMs has also an equally important influence on the quality of projected climate data from CORDEX or other climate downscaling experiments in general. Despite a little difference in the climate model's rank under daily and monthly time steps, M9 and M1 were found as the best and least performing models respectively in both time steps. The average RMSE value from all stations to each climate model is given in Table 2 (in the 4th and 5th columns). As the results reveal, the RMSE value ranges from 7.7 mm (M12) to 13.6 mm (M4 and M11) daily and from 73.6 mm (M9) to 199.4 mm (M4) in monthly time-step. It is interesting to note that climate models have shown different rankings under different metrics considered (see Table S1), which is in line with other researchers finding such as [46,54].

At Grid Level

In the grid-based comparison, station-based observed rainfall data were subjected to three different spatially interpolation techniques (IDW, OK, and EBK). Among them, EBK showed a relatively good interpolation skill to map the spatial distribution of rainfall in the study area (Table S2) and thus its result has been considered as a reference data for spatial rainfall analysis and comparison of the climate model rainfall projection performance. The grid-based comparison results depicted in Figure 3a show the individual climate model's performance in capturing the spatial variability of rainfall in the region. The results indicated a considerable difference in climate model performance. Among all candidates, M8 was the best-performing individual climate model with a value of 0.7 and 225 mm, for pattern correlation coefficient and RMSE respectively, and with the closest standard deviation value (285 mm) compared to observed data (nearly 175 mm). The next best-performing climate model was M9, followed by M1, M6, M3, and M5. M4, M11, M10, M7, M13, M15, M2, and M17 showed the poorest skill in capturing the spatial pattern. Some climate models such as M2 have a negative spatial pattern correlation (nearly −0.1) with observed data. Whereas M4 provided an exaggerated spatial variability with a value of 1450 mm compared to the observed value of 175 mm.

Table 2. Average overall score and RMSE value of individual climate models from all rainfall stations in daily and monthly time steps over the historical period (1986–2005).

Climate Models	Average Score over the Historical Period		Average RMSE from All Stations	
	Daily Time Step	Daily Time Step (mm)	Daily Time Step (mm)	Monthly Time Step
M1	12.8	10.6	10.6	123.8
M2	7.3	9.3	9.3	124.2
M3	9.4	9.4	9.4	105.5
M4	12.2	13.6	13.6	199.4
M5	11.5	10.4	10.4	115.8
M6	7.4	8.6	8.6	104.1
M7	9.8	10.9	10.9	162.6
M8	9.0	9.9	9.9	91.7
M9	6.3	8.0	8.0	73.7
M10	9.4	10.9	10.9	156.3
M11	12.0	13.6	13.6	181.4
M12	6.4	7.7	7.7	86.4
M13	7.9	9.9	9.9	98.8
M14	8.2	8.2	8.2	105.4
M15	8.1	9.9	9.9	91.6
M16	6.6	9.3	9.3	91.2
M17	8.9	9.8	9.8	134.4

N.B. The colour scale is for each column and ranges from "*green*" for best to "*red*" for poor skill. The 2nd and 3rd columns show the average score from the comparison result between the 17 individual climate models while the 4th and 5th columns show the RMSE value averaged from all stations.

Most of the poorly performed climate models in the grid level comparison also showed a similarly poor performance in station level comparison. However, M1 which had the least performance at the station level has shown a U-turn with a good performance in the grid-based comparison (Figure 3a). M9, which was the best individual climate model at station level comparison, repeated its good performance by capturing the spatial variability with a value of 0.6, 270 mm, and 300 mm respectively for pattern correlation, RMSE, and standard deviation.

Basin Level Evaluation (Mean Annual Rainfall, Inter and Intra-Annual Rainfall Variability)

In this section, basin-level comparison results that include basin average mean annual rainfall, inter, and intra-annual rainfall variability are presented, and the results are depicted in Figure 4a–c respectively. As the results in Figure 4a show, the majority of climate models overestimated the basin average mean annual rainfall while others presented an underestimated amount. M4 (760 mm), M11 (679 mm), and M1 (329 mm) have the highest overestimated rainfall while M14 and M12 provide the highest underestimated rainfall with a bias of roughly −582 mm rainfall each. The discrepancy between models' results indicates the magnitude of uncertainties in rainfall projection from CORDEX climate models. For instance, considering the bias from the two extreme models (M4 and M12), leads to a difference of around 1342 mm rainfall amount per year. This discrepancy in annual rainfall is even higher than the maximum observed rainfall amount in the region. Among all the climate models, M6 has the least biases with only −1.2 mm followed by M15 (29.5 mm), M9 (−55.5 mm), and M8 (94.3 mm). Reports from [8,28] also showed a large positive bias by RCMs in the East Africa region and Ethiopian highlands respectively, which supported the finding of this study. Ref. [55] compared two GCMs and their dynamically downscaled version from CORDEX over Africa and reported wet bias in the Ethiopian highlands and elevated terrains of Sudan. Other studies by [50,56] conducted CMIP5 GCMs performance evaluation over East Africa and the Great Horn of Africa respectively and reported overestimation of basin average rainfall amount by most of the CMIP5 GCMs and their ensembles.
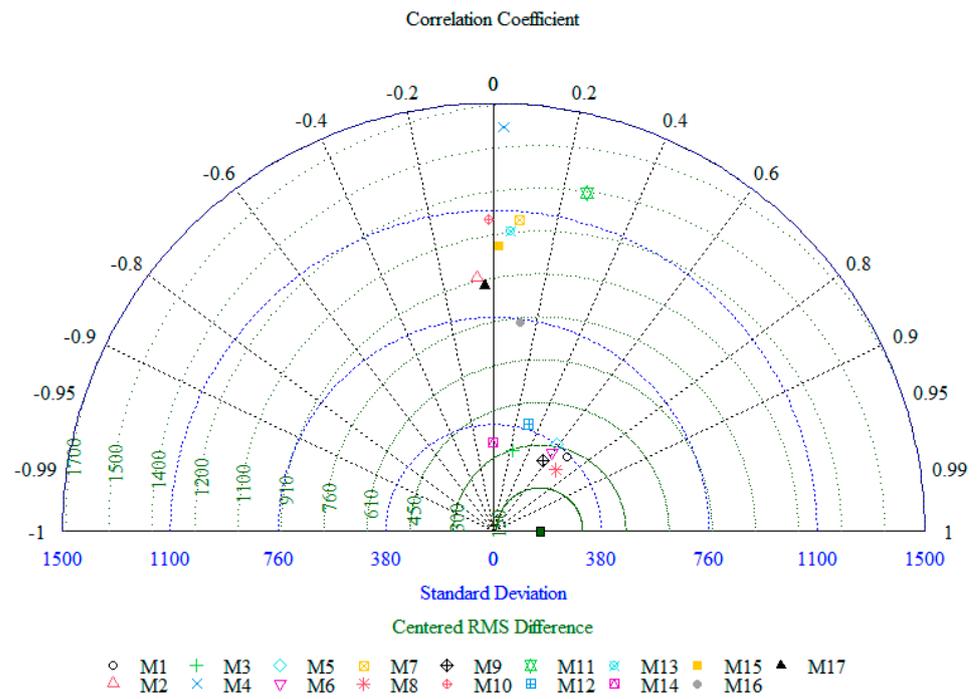
(**a**)
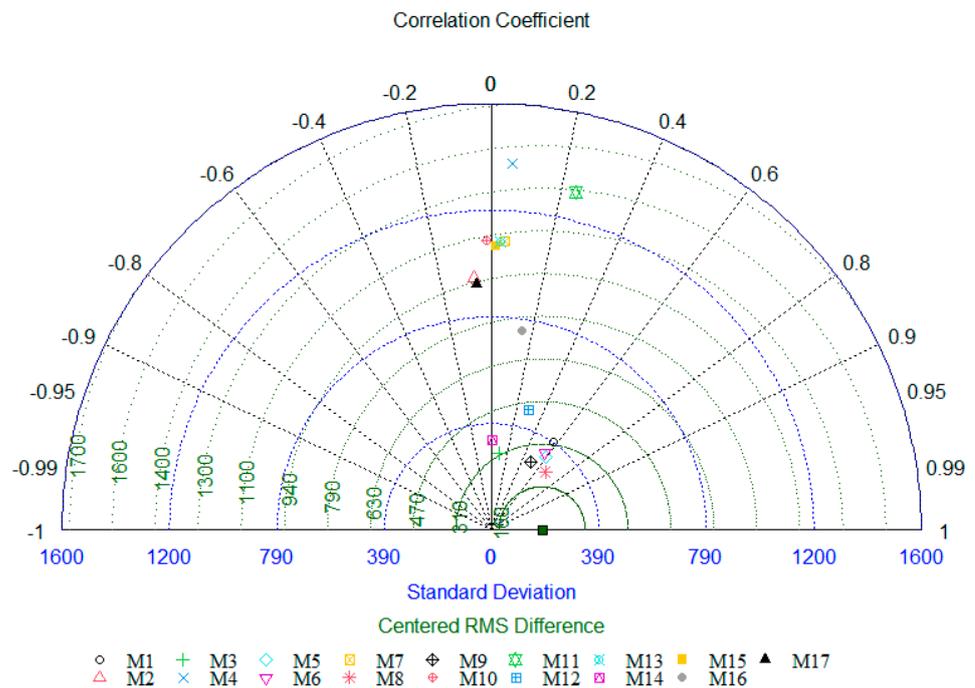
(**b**)

Figure 3. Grid-based comparison results for individual climate models using observed data as a reference and for the historical period (1986–2016) in (**a**) and the validation period (2006–2016) in (**b**).
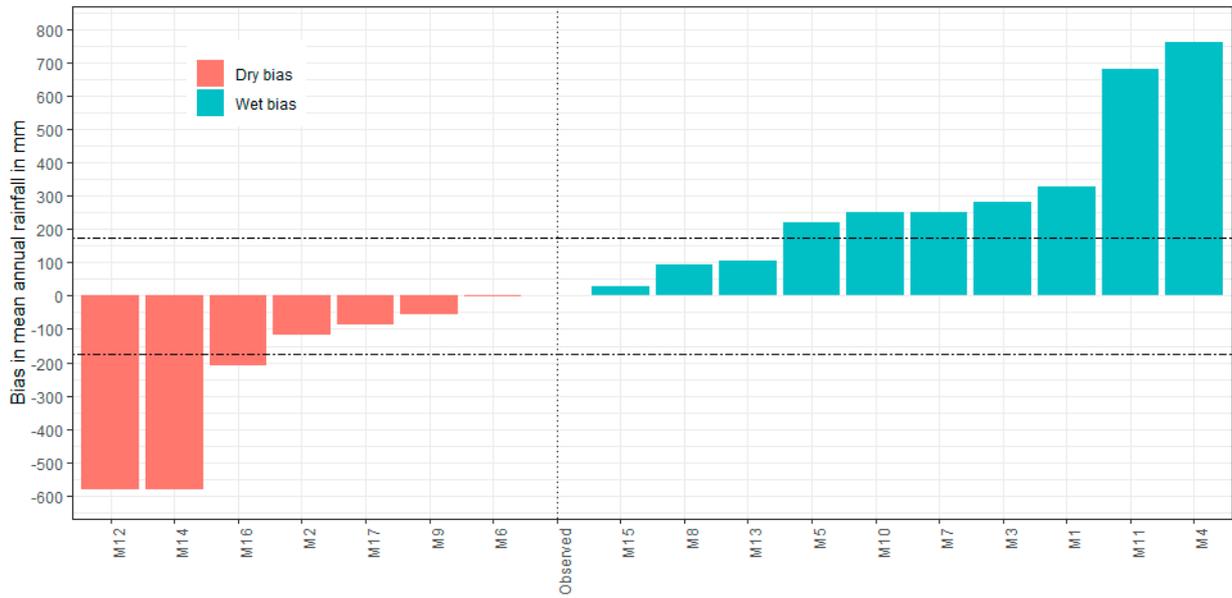
(**a**)
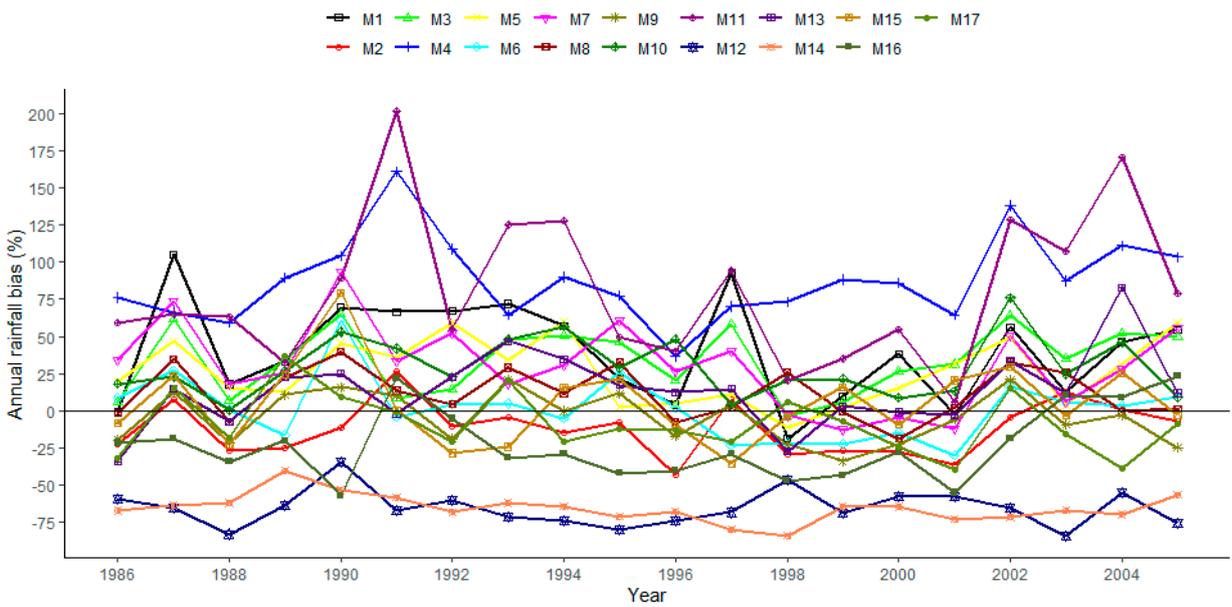
(**b**)

**Figure 4.** *Cont.*
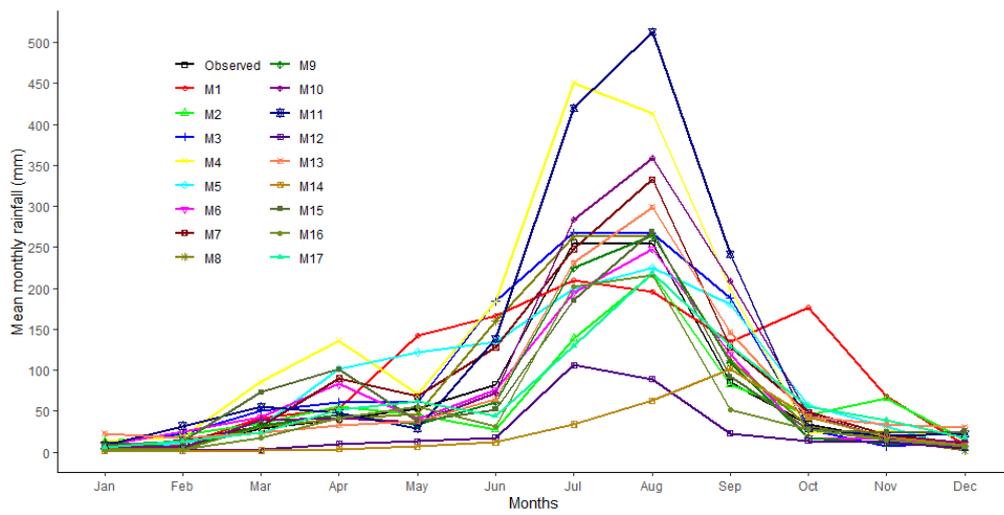
(**c**)
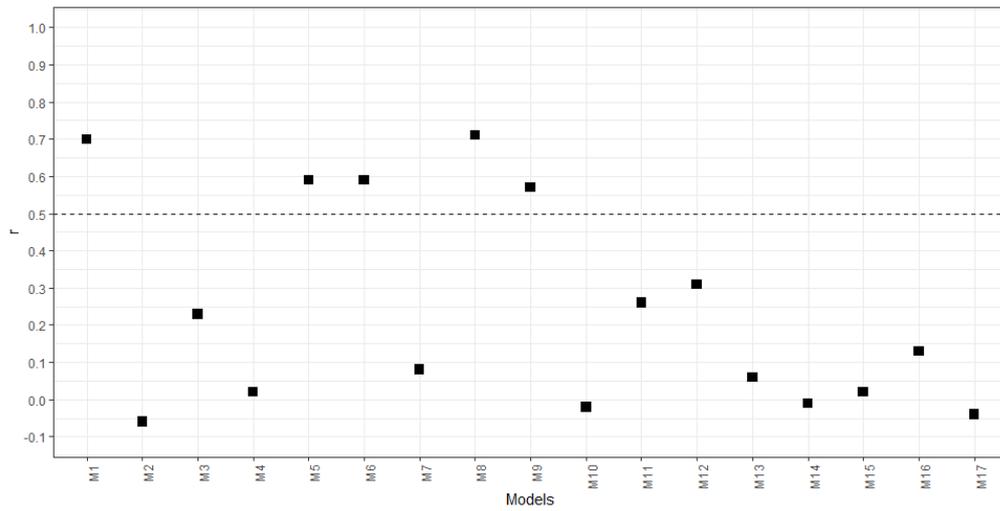
(**d**)
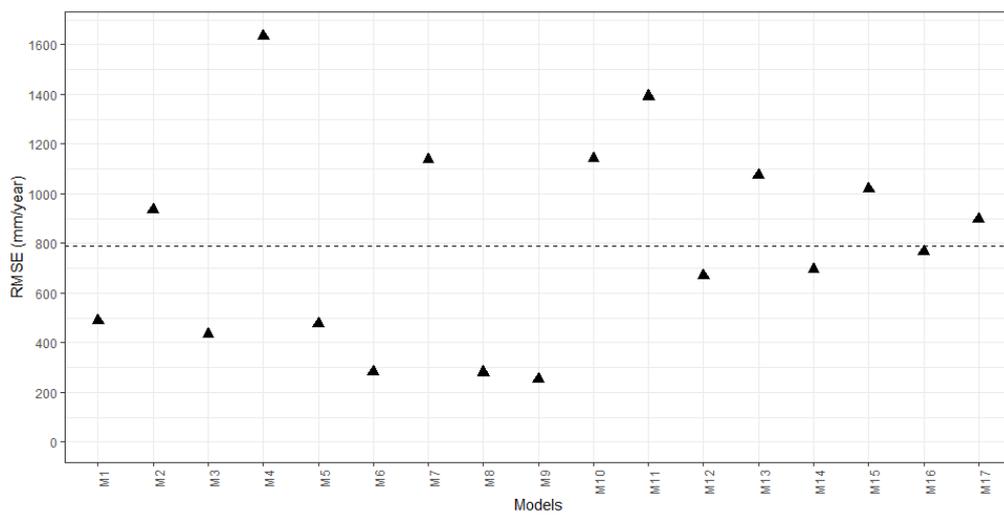
(**e**)

Figure 4. *Cont.*
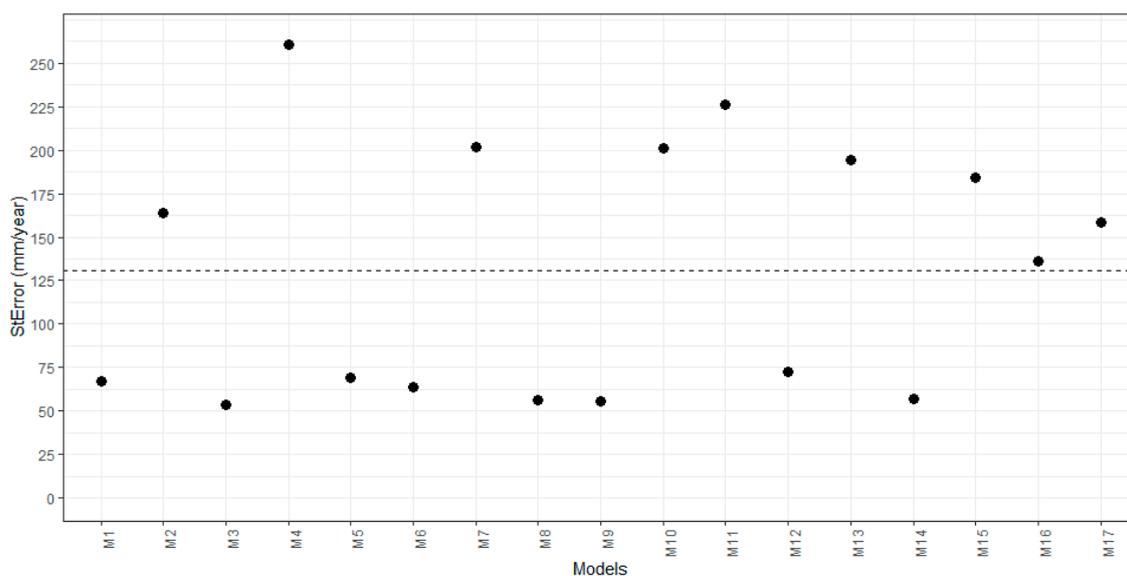
(f)

Figure 4. Comparison of results from individual CORDEX models output with respect to observed data over the period of 1986 to 2005 (**a**) Bias in basin mean annual rainfall (**b**) Bias in annual rainfall (**c**) Monthly rainfall pattern (**d**) spatial rainfall pattern correlation (r), (**e**) RMSE and (**f**) StError.

The inter-annual rainfall variability (1985–2005) from the CORDEX models is given in Figure 4b. The result showed a wide range of model performance from a continuous underestimation for instance by M12 and M14 to a continuous overestimation over the entire period by M4 and M11. Despite the difference between individuals' model bias, there is a high tendency to overestimate annual rainfall from most of the climate models. M6, M8, and M9 appeared to be relatively best-performing models with less bias in simulating the inter-annual rainfall variation though they still have sudden fluctuations in their prediction results. A similar study by [4] over the Africa domain reports the poor performance of CORDEX-RCMs in simulating the inter-annual variability of rainfall in the Ethiopian highlands, which is more or less in line with the result of this study.

The annual cycle or monthly rainfall pattern resulting from the individual climate models is given in Figure 4c. The results indicate that the majority of the climate models capture the unimodal pattern of the annual rainfall cycle in the region. Nevertheless, some models showed either a shifting of the rainy season, substantial under or overestimation, or even indicated a bimodal rainfall pattern. For instance, M14 completely underestimated and even shifted the peak rainfall period to September. M12 captured the unimodal rainfall pattern but highly underestimated the rainfall in all months including the peak value. Some climate models such as M1, M4, M6, and M15 have shown a bimodal rainfall pattern with an overestimation of the short rainy season (March to May) and underestimation of the long rain season (June to September) in the region. Among the candidates, M9 and M8 best capture the annual cycle with less bias. Similar studies by [57] in eastern Africa reported the same result where most of the CMIP3 (CMIP5) GCMs overestimate (underestimate) the short-rainy season (long-rainy season). The climate model's skill in reproducing the seasonal cycle of rainfall is an important performance indicator and major information for farmers and water resource managers [58]. Thus, climate models' inability to capture the seasonality of rainfall, in particular, underestimation of the main rainy season in the region possesses potential risks related to heavy rain such as flood hazards and damage to infrastructure, settlements, and the population.

Spatial Distribution of Rainfall

In addition to grid-based comparison, which is the numerical way of understanding climate models' skill in capturing the spatial distribution of rainfall, a visual study area map depicting the spatial distribution of mean annual rainfall (over 1986–2005) from observed data and CORDEX climate models is shown in Figure 5. As the results reveal, the majority of climate models did not capture the spatial distribution of rainfall and presented an exaggerated spatial variability of rainfall in the study region. For instance, both M4 and M11 predicted mean annual rainfall of nearly 7200 mm and 4600 mm (highly overestimated rainfall) in the west and southeast part of the region respectively, whereas the basin average and maximum observed mean annual rainfall is only 879 mm and 1125 mm respectively. Most individual models such as M2, M7, M10, M12, M13, M14, M15, M16, and M17 presented an underestimated rainfall along the gorge in the basin. The amount of simulated mean annual rainfall by the climate models in the basin ranges from nearly 0 mm to above 7000 mm, while in observed data, it ranges from 550 mm to 1125 mm (Figure 5). M12 and M14 predicted a highly underestimated rainfall amount in the basin. Among the candidate climate models, M6, M8, and M9 were the ones that showed a better skill in capturing the spatial variability of rainfall in the region. M1, M3, and M5 also featured a relatively good spatial variability but with a slightly overestimated rainfall across the basin.

The actual correlation between observed rainfall and altitude in the region is low as shown in Figure S1a. However, looking at the spatial rainfall pattern from most of the climate models (Figure 5) and the regions' topographic condition in Figure 1, the predicted rainfall from the majority of climate models has shown a relatively strong correlation pattern with altitude. As far as the result depicted in Figure S1a is concerned, the strong correlation between predicted rainfall from the climate model and altitude is most likely for the wrong reason which indicated climate models' failure to capture the topographic influence on spatial rainfall patterns in the study region. Despite the difference in magnitude of bias, it is interesting to notice that most of the climate models have shown spatial consistency in their overestimation and underestimation of annual rainfall. The west (in Ras Dashin Mountain) and southern part of the basin are highly elevated areas and the overestimated rainfall by most of the models in these areas reflects the climate model's limitation in the mountainous region. Similar studies by [59] reported that downscaled GCMs from CORDEX predicted maximum rainfall of up to 4745 mm/year and 3285 mm/year in Guinean and Cameroonian highlands respectively. Ref. [60] evaluated different climate models from the CORDEX-Africa domain over Malawi and reported that none of the individual or ensemble models has a good correlation with the observed dataset in both precipitation and temperature variable. The overestimation of rainfall in the highland area may be linked with climate model limitations to resolve topography accurately [61,62].

Ref. [61] assessed the impact of RCMs grid size and model domain on climate simulation skills. Their report showed that RCMs with 50 km resolution underestimate the actual terrain height and have difficulty in capturing the orographic influence. In this research as well, where the CORDEX data with 50 km spatial resolution is considered, the resolution is still not sufficient to capture completely the rainfall heterogeneity in the study region. For instance, from the correlation result given in Figure S1b, there are stations less than 50 km far apart but with correlation strength of as low as 0.3 in their daily rainfall amount. Yet, from the CORDEX spatial resolution (nearly 50 km), these stations are assumed to have the same rainfall amount, which may not be correct. Therefore, the climate modelling community has to continue their endeavours to find out, how they can improve climate model's skill including the representation of orographic influence to better capture spatial rainfall patterns at the local level. In addition, efforts to further downscale the climate projection from GCMs into a finer spatial resolution need to continue.
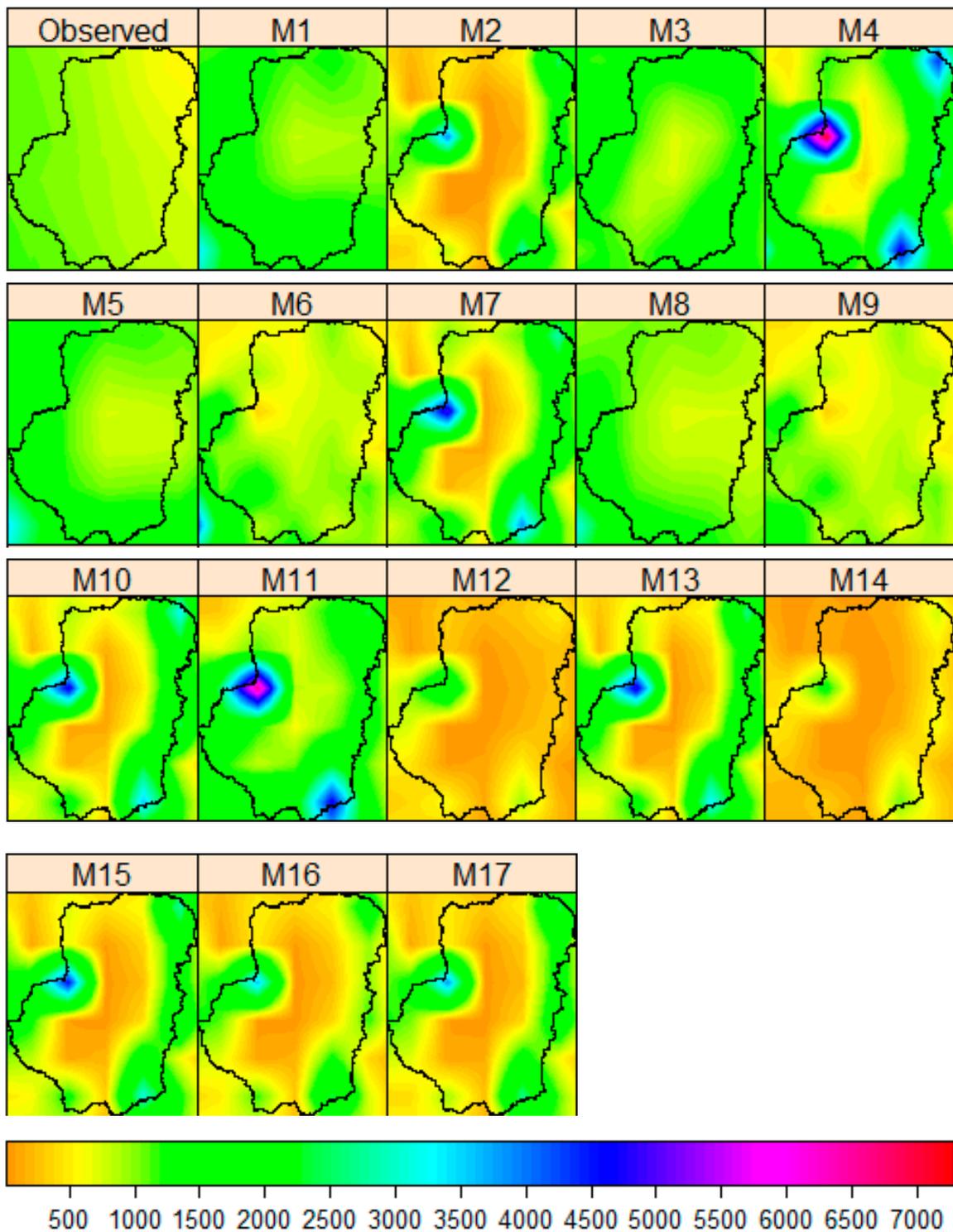
Figure 5. Spatial distribution of mean annual rainfall (in mm) from observed data, and individual climate models over the period of 1986–2005.

3.1.2. Climate Models Skill in Simulating the Future Climate (Validation Period)

Besides the historical comparison analysis, model performance in simulating the future climate has been carried out as a potential validation step over the period of 2006 to 2016. This evaluation has been done at the grid level (using mean annual rainfall). As depicted in Figure 3b, almost all the individual climate models have kept their historical skill order in the validation period too. M8, M9, and M6 were in the top rank while

M4, M11, M10, and M7 showed poor skill which is the same as during the historical comparison. The validation result in this study supported the assumption that climate models' historical performance is a good indicator of their future simulation skill and doing so gives reassurance to the climate models' ability in the region. In addition, the similar performance result of climate models over the historical and future period (under the RCP8.5 scenario) could indicate the little influence of emission scenarios on rainfall projection uncertainty as several researchers reported that emission scenario has less contribution to rainfall projection uncertainty [17,63–65].

3.2. Selection of Candidate Models for New Ensemble Model Creation

The six criteria that have been used to select candidate climate models for new ensemble formation were the bias in basin average rainfall, bias in annual average rainfall, bias in monthly average rainfall, the correlation coefficient (r), RMSE, and StE in the spatial rainfall pattern. In climate modelling, in particular rainfall projection, the two fundamental aspects that every climate model is expected to address are: when the rain comes (the time) and where the rain falls (the place). Thus, the intra-annual (seasonal rainfall pattern) and the inter-annual rainfall result account for model skill in capturing the temporal variability of rainfall. The result of each climate model in each of these metrics is given in Figure 4d–f respectively and the summary numeric result derived from these figures is provided in Table 3. The last three metrics were derived from grid-series data, and they reflect the model's skill in capturing the spatial distribution of rainfall.

Table 3. Selection of candidate models to form an ensemble under a different range of model performance over the study area.

	The Absolute Bias of Rainfall in						Ensemble Models			
Climate Models	Basin Average	Inter-Annual	Monthly Average	≤Ave StE	≤Ave RMSE	≥0.5 r	M18	M19	M20	M21
M1	37.4	25.8	63.5	**67.2**	**489.6**	**0.70**	*	**		
M2	**13.5**	**12.2**	37.6	164.0	934.3	−0.06	*			
M3	31.9	20.8	36.4	**53.9**	**432.3**	0.23	*			
M4	86.4	55.7	90.1	260.6	1635.1	0.02	*			
M5	25.3	**17.1**	47.9	**69.1**	**475.7**	**0.59**	*	**		
M6	**0.1**	**9.5**	24.9	**63.5**	**283.9**	**0.59**	*	**	***	
M7	28.4	20.7	31.6	201.6	1136.2	0.08	*			
M8	**10.7**	**9.6**	**16.9**	**56.2**	**280.2**	**0.71**	*	**	***	****
M9	**6.3**	**10.1**	**15.8**	**55.4**	**254.5**	**0.57**	*	**	***	****
M10	28.3	**18.2**	35.9	201.2	1140.8	−0.02	*			
M11	77.2	49.8	85.1	226.4	1391.3	0.26	*			
M12	66.2	42.7	64.9	**72.7**	**669.4**	0.31	*			
M13	**11.8**	**13.2**	28.5	194.2	1075.1	0.06	*			
M14	66.2	42.7	71.9	**57.2**	**693.2**	−0.01	*			
M15	**3.4**	**12.4**	32.4	184.1	1020.7	0.02	*			
M16	23.9	**19.4**	23.4	135.8	767.2	0.13	*			
M17	**9.8**	**12.2**	37.6	158.5	895.9	−0.04	*			
Threshold values	≤20%	≤20%	≤20%	≤130.7 mm/year	≤798.6 mm/year	≥0.5				

Values highlighted in bold show the model that meets the criteria for candidacy under the given evaluation metrics. (*), (**), (***), and (****) indicate the CORDEX model that has been selected to form the ensemble models M18, M19, M20, and M21 respectively.

As the results show, seven (nine) climate models have ≤20% bias in basin average (annual average) rainfall, whereas, in monthly average rainfall, there were only two models (M8 and M9) that meet the criteria for candidacy. Regarding the StE and RMSE, there were eight and nine individual models respectively whose results fall below the average value from all candidate climate models. In the spatial pattern correlation result, there are five models with an r value ≥ 0.5. As the summary results given in Table 3 revealed, there are only two models (M8 and M9) that have been given a candidacy flag across all the six criteria applied while M6 was selected in five criteria except the bias in monthly average rainfall. As can be seen in Figure 4c, M6 presented a bimodal rainfall pattern in the region

which resulted in a bias (24.9%) exceeding the given threshold of 20%. M1 and M5 were the other candidate models that have been selected in three of the six criteria. Their candidacy comes from the spatial analysis result which shows that they better capture the spatial variability but not the temporal dynamics of rainfall. The good performance of M8 and M9 in all criteria indicates their skill in capturing both the spatial and temporal variability of rainfall in the region.

Following the selection of the candidate models under the given criteria, three different new ensembles were formed. Table 3 showed that there are five models (M1, M5, M6, M8, and M9) that have been selected in at least three criteria and these models were used to form the new ensemble model M19.

Similarly, ensemble M20 was created using M6, M8, and M9 which have a candidacy flag in at least five criteria. Then the final ensemble M21 was formed using only M8 and M9 which are the only individual model that has been selected in all the selected criteria. Unlike these three ensemble models, M18 is the muti-model ensemble created using all the available 17 CORDEX climate models regardless of the individual climate model performance. After creating these ensembles their performance was evaluated together with all the individual models and the multimodal ensemble M18 and the result is presented in the following section.

3.3. Performance of All the Individual and Ensemble Climate Models

Similar to the previous model evaluation process, evaluation of the multi-model ensemble (M18), the three newly formed ensembles (M19, M20, and M21) along with the previous individual models has been carried out over a historical period and at the station, grid, and basin scale. As the station-based evaluation results given in Table 4 show, M18 was the best performing model in the daily time-step with a score of 5 and followed by M20 (5.8), M19 (6.6), and M21 (7.8). In the monthly time step, the best model was M20 with a score of 4.8 and closely followed by M21 (4.9), then M18 (5.6), and M9 (6.5). The grid-based comparison results (in historical and validation period) are given in Figure 6a,b respectively while the spatial distribution of rainfall from observed data, all climate models, and CRU data as an alternative reference is depicted in both Figure 7 (rainfall density plot) and Figure 8 (map of the study area with rainfall amount). As these results indicate, the multi-model ensemble (M18) did not perform well and some of its member models such as M8, M9, M6, and M1 showed a better skill in this regard. M18 has a spatial pattern correlation coefficient of nearly 0.2, a standard deviation close to 620 mm, and an RMSE of 610 mm (Figure 6a,b). The model that performed well in capturing the spatial distribution of rainfall was M21 which was formed by just two models. Even the other two ensembles (M19 and M20) have shown more or less the same good skill in representing the spatial distribution of rainfall. Given the spatial dynamic nature of rainfall, the climate model's skill in capturing its spatial pattern is a major indicator of model performance. Thus, the ensembles formed with systematically selected member models deserve more credit over the multi-model ensemble M18 due to their outperforming skill in capturing the spatial rainfall pattern.

The result of bias in basin average annual rainfall is given in Figure 9a for all climate models including the ensembles. As the results show, the predicted rainfall amount by M20 (891 mm) and M21 (897 mm) have only 1.3% and 2% bias respectively from the observed value (880 mm). M6 was the best model in this regard with a bias of only 0.14%. Comparing the result with M20 and M21, the multi-model ensemble (M18) has a considerable bias in basin average rainfall (960 mm) which is a 9.1% bias. In addition, M19 has shown a substantial bias with a value of 13.3% which is slightly higher than the M18 (9.1%). The higher bias in M19 (13.3%) is attributed to the significant bias that exists in its two-member models M1(37.5% bias) and M5(25.3%). The outperforming skill of the ensemble models to most of their member models indicated the offsetting of biases between member models during the ensemble which underlines the importance of considering the individual model's bias prior to forming an ensemble.

Table 4. Average overall score and RMSE value of all climate models from all rainfall stations in daily and monthly time steps over the historical period (1986–2005).

Climate Models	Average RMSE from all Stations and Climate Models		Average Score over the Historical Period	
	Daily Time Step (mm)	Monthly Time Step (mm)	Daily Time-Step	Monthly Time-Step
M1	10.6	123.8	16.6	16.8
M2	9.3	124.2	10.0	13.2
M3	9.4	105.5	12.9	12.4
M4	13.6	199.4	15.9	15.0
M5	10.4	115.8	15.2	15.2
M6	8.6	104.1	10.4	11.3
M7	10.9	162.6	12.8	14.5
M8	9.9	91.7	12.4	9.8
M9	8.0	73.7	8.9	6.5
M10	10.9	156.3	12.4	11.3
M11	13.6	181.4	15.6	15.3
M12	7.7	86.4	8.7	14.7
M13	9.9	98.8	10.9	11.7
M14	8.2	105.4	10.6	11.3
M15	9.9	91.6	11.1	11.7
M16	9.3	91.2	9.3	6.9
M17	9.8	134.4	11.9	11.2
M18	7.0	132.2	5.0	5.6
M19	7.2	77.6	6.6	7.2
M20	7.4	74.0	5.8	4.8
M21	7.9	70.7	7.8	4.9

N.B. The colour scale is for each column and ranges from "*green*" best to "*red*" poor skill.

Regarding the inter-annual variation of rainfall, both M20 and M21 have shown good performance with better temporal stability compared with that of observed rainfall variation and then followed by M6, M8, and M9, which are the member models during their formation (Figure 9b). The result in Figure 9b,c is depicted only for selected models to ease visualization and results for the remaining models are already given in Figure 4b,c. The annual rainfall bias from all models ranges from −85% in M12 to 202% in M11, while in M21, M20, M19 and M18, it ranges from −21% to 27%, −19% to 39%, −10% to 46%, and −10.4% to 34.8% respectively. From the mean monthly rainfall pattern result as well, M21 has shown an outstanding skill with the least bias and followed by M20, M9, and M8 (Figure 9c). From the spatial rainfall distribution result given in Figure 8, the multi-model ensemble M18 featured a slight underestimation (overestimation) in the gorge (highland) part of the basin. Despite the better skill of M18 compared to some of the individual models, there are some models such as M6, M8, and M9 that outperform the M18 in capturing the spatial variability of rainfall, which indicates that multi-model ensemble is not always the best performing model.

As the overall comparison results (historical and validation period) indicate that ensembles formed with highly performing individual models (M20 and M21) have an outstanding skill in simulating the spatial distribution and temporal variation of rainfall in the study region. Even the multi-model ensemble (M18) with 17 member models was surpassed by these two models M21 (M8 and M9) and M20 (M6, M8, and M9) in most of the evaluation metrics considered. Hence, there are two important points that should be highlighted from the results found here. (i) Ensembles have the potential to provide a better climate projection result than most of the single models (if not all) since some individual models such as M8 and M9 outperformed the multi-model ensemble (M18) in some of the metrics used. This improvement in the ensemble model is attributed to the offsetting of the individual model's bias, as has been acknowledged by [1,8] (ii) Even though ensemble models are good, the result in this study explicitly showed that the skill of ensemble models is not really a function of how many individual models were involved while creating the

ensemble. The major determining factor in the performance of the ensemble is rather the skill of the individual member models considered in the ensemble. Other researchers such as [66] also pointed out the fact that the reliability of climate projections might be improved if only subsets of climate models are considered. Outperforming skills from systematically formed ensembles have been reported in other studies too like [25] over the Arabian Peninsula and [26] in central Africa.
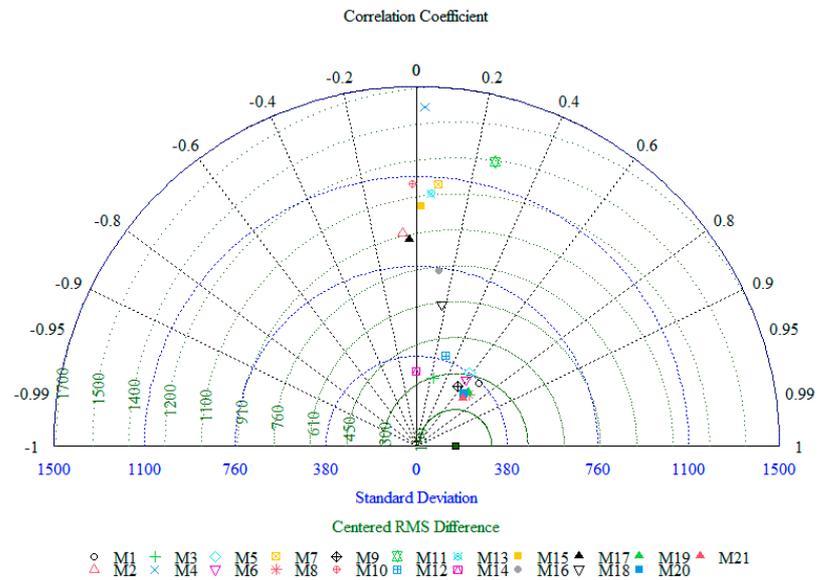
(**a**)
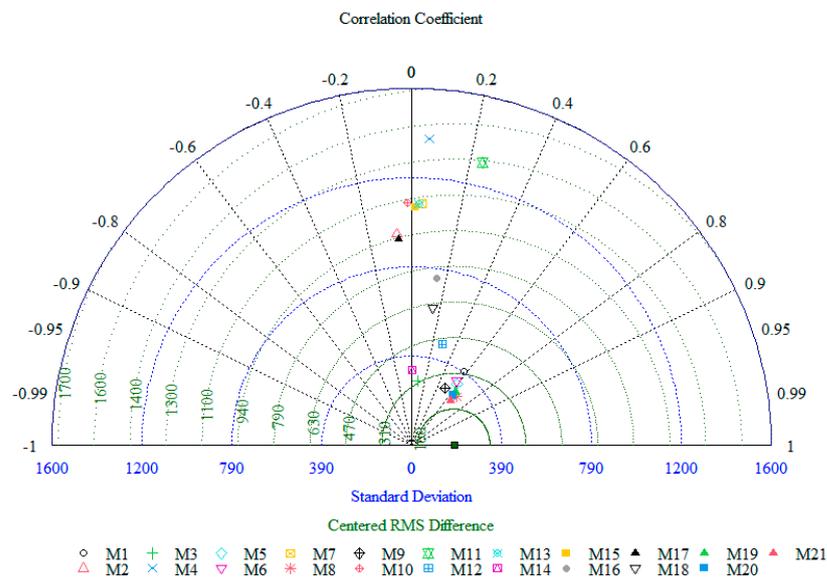
(**b**)

Figure 6. Grid-based comparison result for all climate models using observed data as a reference and for the historical period (1986–2016) in (**a**) and validation period (2006–2016) in (**b**).
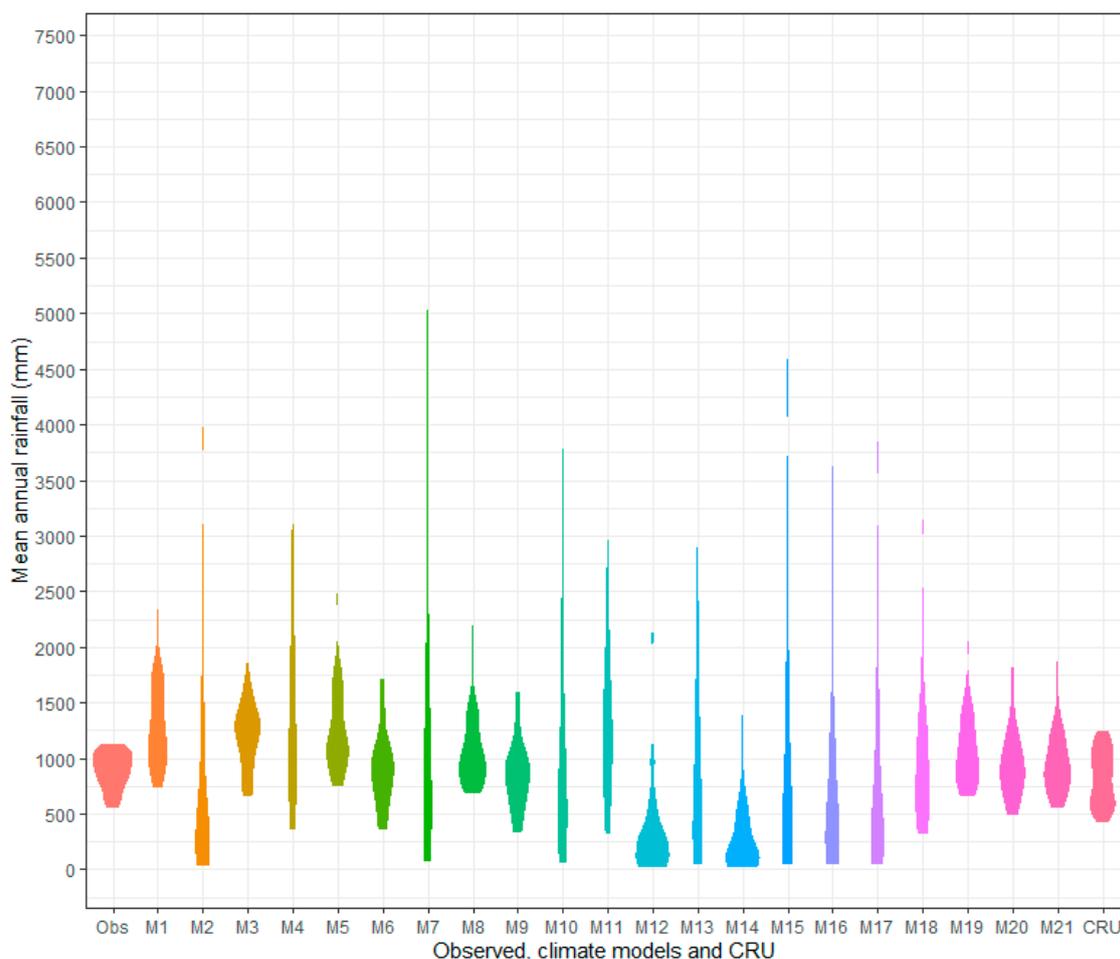
Figure 7. Density plot that shows the spatial variability of rainfall from observed data (Obs), CORDEX climate models (M1–M17), their ensembles (M18–M21), and from Climate Research Unit (CRU) reanalysis rainfall data as an alternative reference.

Nevertheless, most of the climate change and impact studies often use the multi-model ensemble that considers all the available individual models regardless of the individual model's performance [13,67]. Therefore, as far as reliable climate study is concerned, end-users of the climate model output need to be cautious about how to form their ensemble model. For instance, from the result found here, both M20 and M21 ensemble models are the best and most reasonable alternative ensemble climate models to be used compared to other models including the multi-model ensemble M18. Thus, previous climate change and its impact studies that use the multi-model ensemble without prior model evaluation are more likely to be biased in their result at least in this study area as indicated by the findings of our study. Similar to the previous remarks given by [8,50,68] about CMIP5 GCMs and CORDEX-RCMs biases in the East Africa region, the result of this study again reinforces the need for caution while using downscaled climate model outputs from CORDEX archives. Policymakers are among the end-users of climate research outputs or reports and any possible uncertainties in the given climate study reports will continue to propagate along with the entire decision-making process. This will eventually cause a huge socio-economic burden as far as the implementation of climate change adaption and mitigation strategies with a lot of financial investment is concerned. Therefore, the primary evaluation of climate model performance should not be overlooked while working on climate change projection and its associated impact in different sectors.
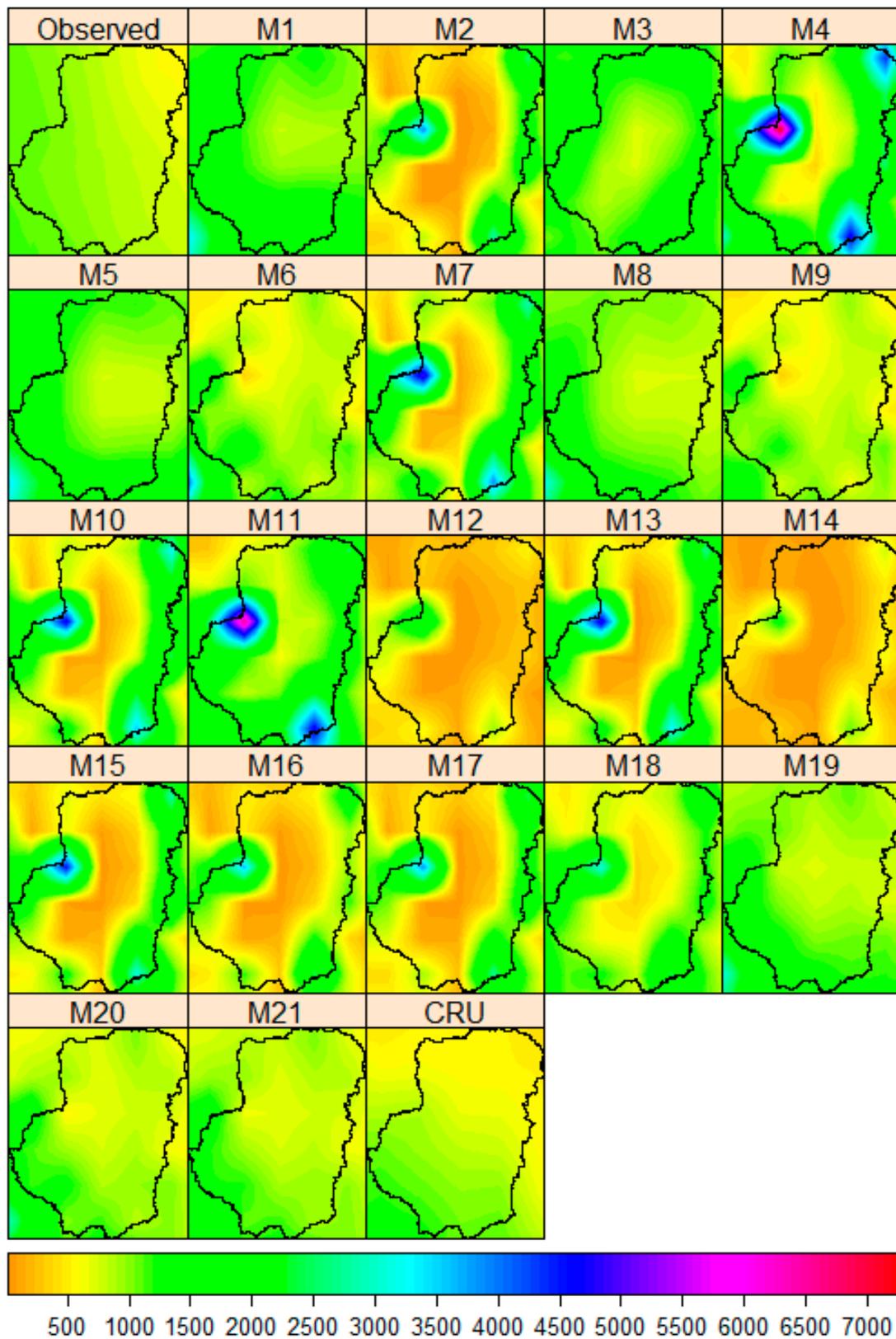
Figure 8. Spatial distribution of mean annual rainfall (in mm) from observed data, all climate models, their ensembles, and CRU data over the period of 1986–2005.
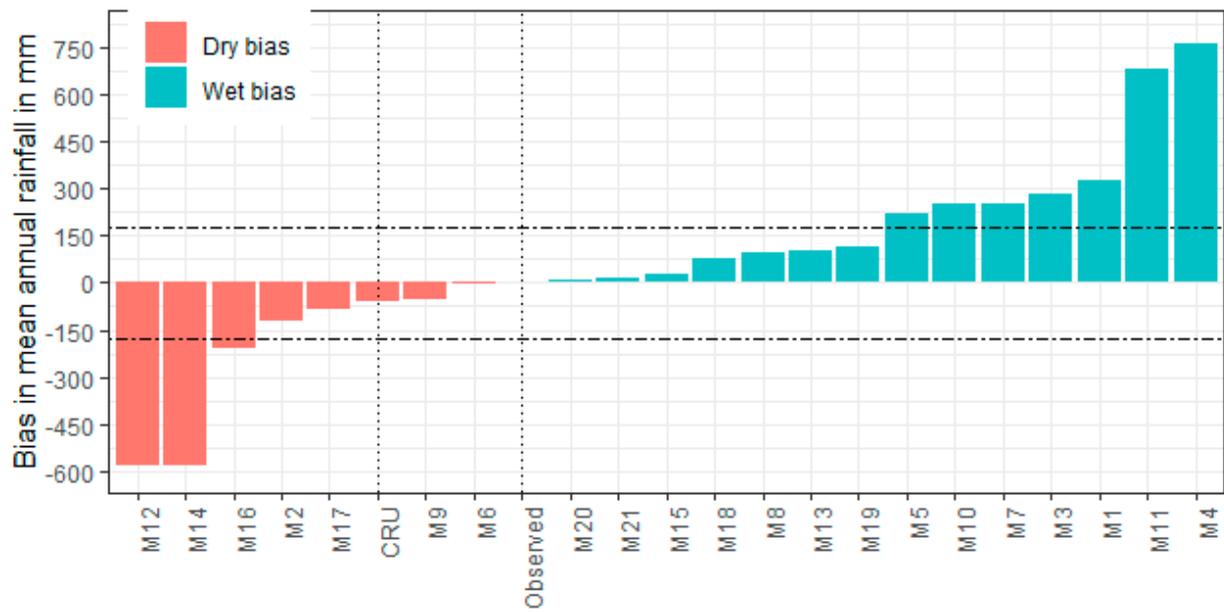
(**a**)
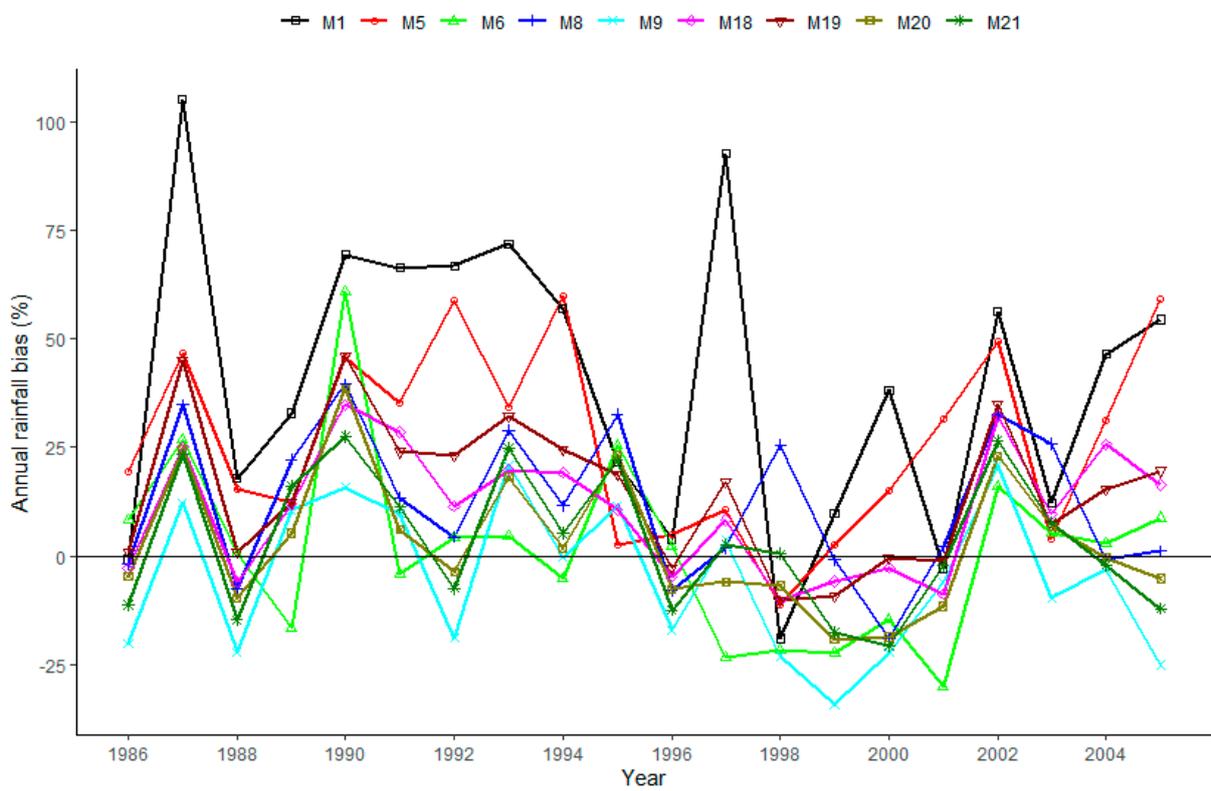
(**b**)

Figure 9. *Cont.*
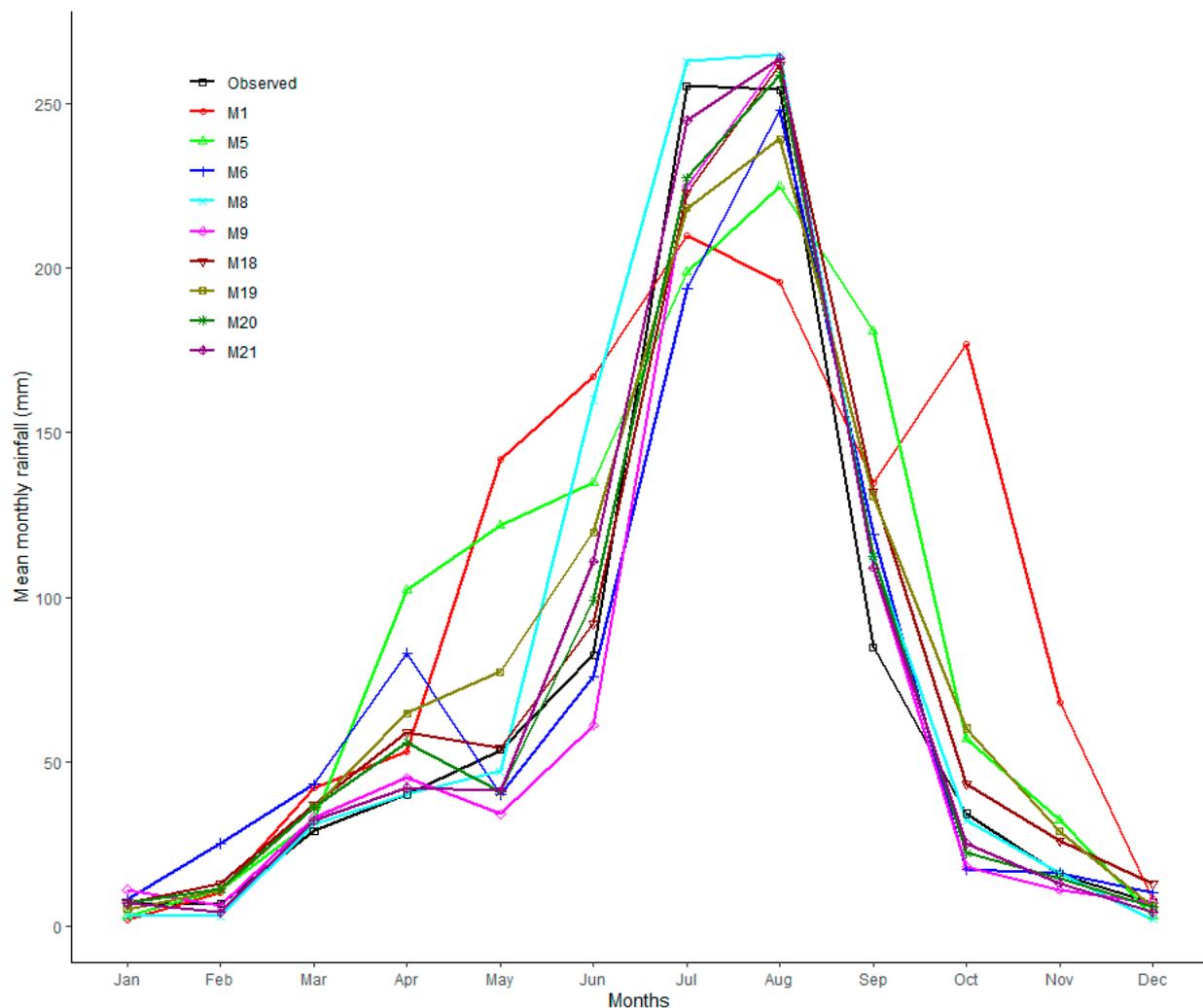
(**c**)

Figure 9. Climate models (including ensembles) comparison result using observed data as a reference over the historical period (1986–2005) (**a**) Bias in mean annual rainfall from all models, (**b**) Bias in annual rainfall in selected models (**c**) Annual rainfall pattern in selected models.

3.4. Climate Models' Potential Contribution to Future Rainfall Projection Uncertainty

The ultimate goal of the climate modelling community is to project the future climate with the maximum possible accuracy although the future is always uncertain by nature. Policymakers on the other hand are interested to know the future impact of climate change in which the future projected climate data is a major input. Thus, given the importance of the future climate projection and the considerable differences that have been found between the climate model's performance, this study considered assessing the significance of differences in the future projected rainfall from the multi-model ensemble (M18) and the relatively best model (M21). In this regard, the F-test two-sampled for variance was applied to the future project rainfall from these two models (M18 and M21). The projected rainfall under RCP8.5 and RCP4.5 emission scenarios and over the two-time horizons (2020–2049 and 2070–2099) were considered. As the results reveal, the future projected rainfall from the two models (M18 and M21) using the RCP8.5 scenario has shown a significant difference at a 5% significance level in both time horizons. The projected result under the RCP4.5 scenario as well has shown a significant difference at a 5% significance level but only in the near-term period. The cumulative distribution function (CDF) of these projected rainfalls is depicted in Figure 10 and as the results indicate, the two models have a considerable

difference in their projection result in both scenarios and time-horizon. The significant variations detected between the projected rainfall under the most widely used ensemble M18 and the best performing ensemble model M21 in this study clearly showed how far biased the future climate projection and its impact studies could be in case proper model evaluation and selection were not in place. This again emphasized the importance of model selection with a thorough evaluation procedure right to the area of interest.
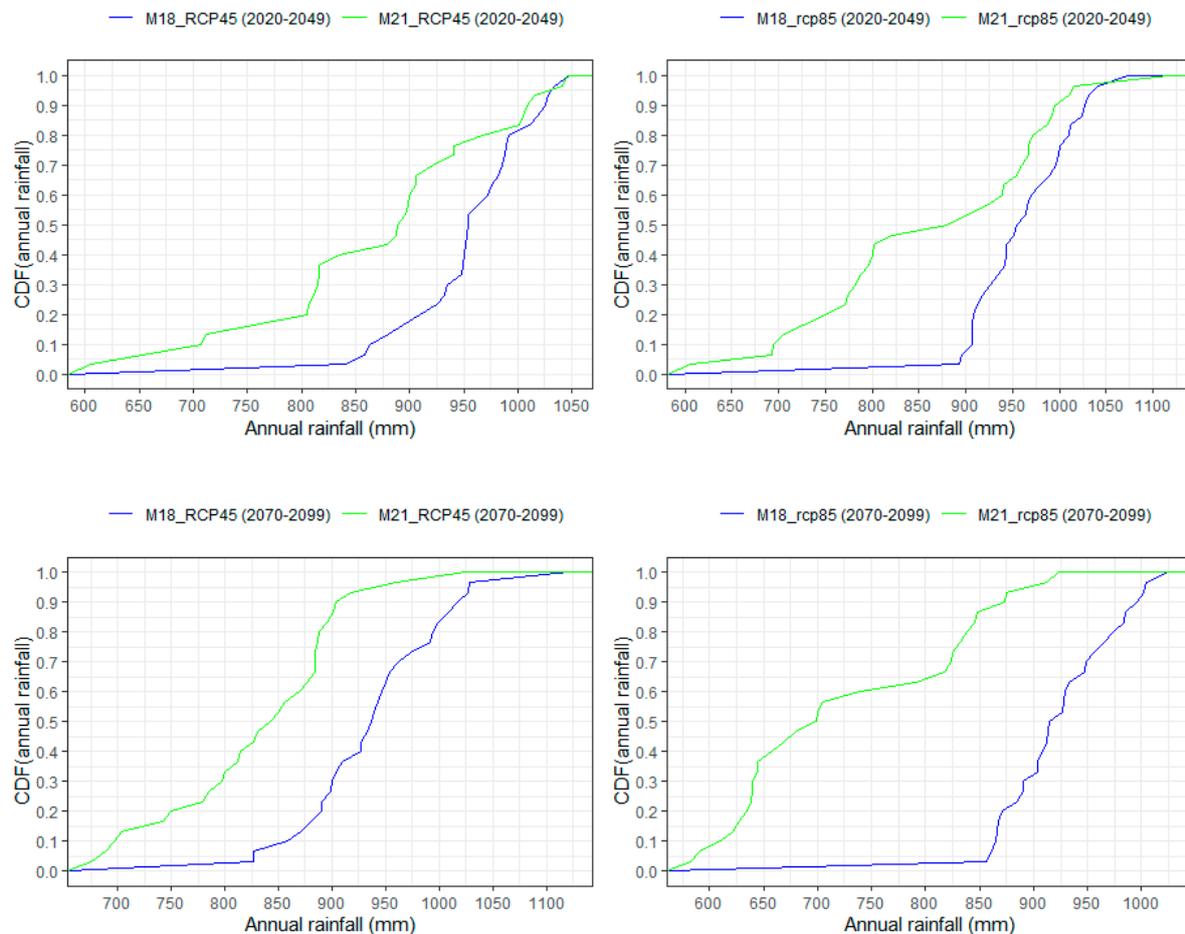
Figure 10. The cumulative distribution function (CDF) of future projected annual rainfall under RCP4.5 and RCP8.5 emission scenarios from the multi-model ensemble (M18) and best performing model (M21) over the two-time horizons (2030s and 2080s).

3.5. Uncertainty in Observational (Reference) Dataset for Climate Model Evaluation Process

In climate model evaluation, the reliability of the reference climate data has an important influence in determining the magnitude of the candidate model bias or agreement which will eventually determine the selection of the relatively best model in the study area. Thus, this study considered quantifying uncertainties in climate model evaluation due to the choice of reference data. In addition to the observed rainfall data, CRU rainfall data has been used as alternate reference data. As the results of the station-based comparison between the observed rainfall data and CRU data showed (Table S3), there is a considerable discrepancy between the two reference datasets. The value of RMSE, MAE, PBIAS, and r in all stations ranges from (32.7 mm to 69.4 mm), (17.9 mm to 39.8 mm), (-43.9% to 4.9%), and (0.75 to 0.92) respectively. As far as their difference are concerned, it is obvious to expect uncertainty in model evaluation results under these two reference datasets.

Nevertheless, to explore the actual effect of reference data on the model selection process, all the previously mentioned climate models were re-evaluated using CRU rainfall data as an alternative reference. The grid-based comparison results of this evaluation for

the historical and validation period are given in Figure 11a,b respectively. As the results show, there is a slight shift in best-performing climate models in at least one of the three measures considered in the Taylor diagram, if not in all circumstances. For example, in Figure 11a, M1 became the best spatially correlated data series with a value of nearly 0.8 while M21 has a value close to 0.7. Despite the change in model bias under some of the metrics, M21 remained to be the best performing model in overall comparison. The spatial distribution (density) of rainfall from CRU data is also given in Figure 8 (Figure 7) together with all climate models. As the results show, the rainfall from CRU data has more or less the same spatial pattern as that derived from observed data. But, from Figure 7 result, the CRU data has shown a slightly higher range of variability compared to observed data.
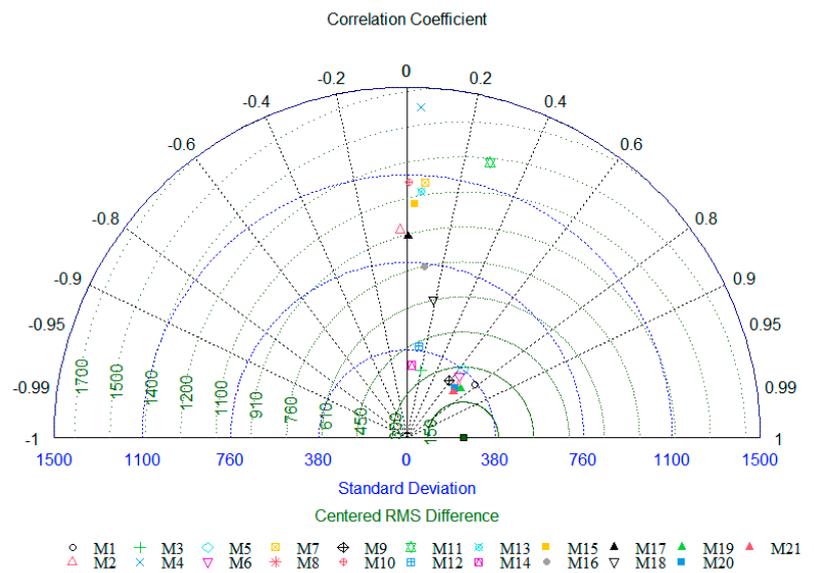
(**a**)
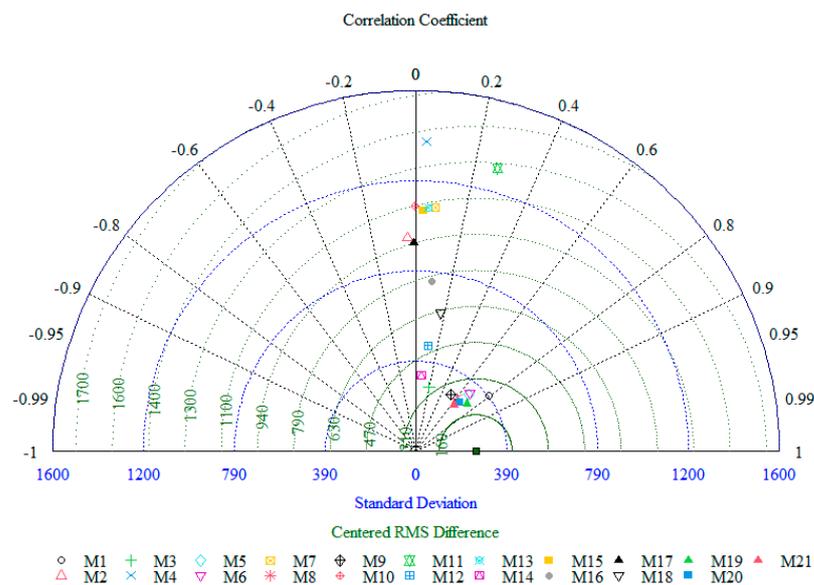
(**b**)

Figure 11. Grid-based comparison result for all climate models (including ensembles) using CRU reanalysis rainfall data as reference and for the historical period (1986–2016) in (**a**) and validation period (2006–2016) in (**b**).

Ref. [18] reported that climate models sometimes could provide a more faithful result to some climate quantities better than the reanalysis dataset, which indicates the improvement in modern climate models. Findings from [8] also show that rainfall biases of reanalysis can be as large as the bias from RCMs. In the interest of seeing such a possible scenario, the current study attempted to evaluate the performance of CRU rainfall data together with all the climate models at the basin level and the result of bias in basin average rainfall is given in Figure 9a. As the result revealed, CRU data underestimated the mean annual rainfall by nearly 56 mm. From this result, there are six candidate climate models (M6, M20, M21, and M15) that have shown better agreement with observed data than CRU data, yet this dataset has been used as a reference in several climate model evaluations research works. For instance, if the best model were to be chosen using the bias in basin average rainfall and CRU data as a reference, then M9 would have been the best model with a little magnitude of bias compared to CRU and probably followed by M17. This result clearly shows how uncertain the model evaluation could be if appropriate reference datasets were not used. [21] also reported a shift in the rank of candidate models when evaluated under different reference datasets which are in support of the finding in this research. The outperforming result of some climate models over the CRU reanalysis dataset in some stations is also in line with the report of [8] where RCMs were found to have a better simulation result than the ERA-Interim reanalysis dataset in Africa's climate.

The results of the overall re-evaluation led to the important conclusion that the use of different data sources as a reference may or may not change which model to choose, but it certainly hides the actual bias of candidate models. However, the magnitude of climate model bias has particular importance to climate model developers who are more concerned to know the level of bias/discrepancy in their model projection which gives a foundation to their efforts in further improvement. One of the big challenges in particular to the African climate is the lack of ground-measured datasets and the use of satellite reanalysis climate products is often the only option. Despite that, previous studies in the East Africa region reported a large discrepancy between the observational data set [8,69], which urges the selection of appropriate reference data sources. Ref. [70] also mentioned that understanding the possible reasons for both common and model-specific problems require a deep and dedicated analysis with explicit consideration of uncertainties in the observational references. Thus, as far as reliable model selection and, in turn, climate projection is concerned, an appropriate reference dataset has to be selected.

4. Conclusions

The issue of climate change is currently at the epicentre of global interest and a wide range of studies has been carried out to investigate the potential impact of climate change in different regionally important sectors. Nevertheless, climate change and its impact studies are accompanied by uncertainties and climate models are often reported to have a considerable contribution in this regard and are particularly related to rainfall projection. Thus, this study was intended to explicitly assess the performance of climate models and associated uncertainties in rainfall projections from the CORDEX data archive over the Eastern Nile basin (Tekeze river basin) in Ethiopia. The study also proposed and demonstrated an approach for the selection of a candidate climate model that can be used to form an ensemble which has the potential towards reaching a better performance than the widely used multi-model ensemble.

The overall evaluation of climate models results in a large discrepancy in the projected rainfall. Most of the climate models resulted in an overestimated rainfall amount over the basin whereas some others underestimated substantially. For instance, one of the poorly performing models (M4) and the widely used multi-model ensemble (M18) presented the maximum mean annual rainfall of nearly 7200 mm and 3250 mm respectively, yet the maximum observed amount in the basin is not more than 1125 mm. Interestingly, the majority of the models showed a spatial consistency in underestimation (in the river gorge) and overestimation (highland area) of mean annual rainfall, which may indicate

the models' difficulties to represent topographic influence in the rainfall distribution The overall deviation of climate models output from observed rainfall data clearly showed the need for a thorough model evaluation to be able to select and use at least the relatively best-fitting model output. The difference in projection output between models is certainly attributed to their difference in understanding and representing the different processes in the climate system. Thus, it is important for climate modellers to get together and share expertise towards the success of climate modelling with better accuracy.

The role of ensembles in improving the quality of individual climate models has been acknowledged by several researchers that worked mainly with model evaluation in different parts of the world [8,40,59,71]. The results of this study also confirmed the power of the ensemble approach in improving climate data quality. However, as some researchers already unveiled and the result in this study showed, the multi-model ensemble (considering all the available climate models) may not always have an outperforming skill. After the initial model evaluation results, three different ensemble models were formed using only a subset of the available models and these models have shown a promising prediction performance compared to the multi-model ensemble (M18). For instance, the multi-model ensemble (M18) with 17 member models had a poor skill in capturing the spatial variability of rainfall and basin average rainfall amount. However, M20 and M21, which only had three and two selected member models respectively outperformed all the candidate models including the widely used multi-model ensemble M18, in most of the evaluation metrics considered. This indicated that the power of ensemble in improving climate models' output is not a function of how many individual climate models were involved during its formation. Rather a high-quality climate projection can be achieved using small but selective member climate models. Therefore, to harness the power of an ensemble, it rather has to be formed with a systematic selection of its members than just considering all the available climate models together.

In conclusion, users of climate model output such as water resources managers, water infrastructure designers, and policymakers working out adaptation and mitigation strategies for the future potential climate threats need to carefully consider the reliability of the data on which their decisions are based. As the modellers keep their effort toward model improvement, the climate user community has to do a proper evaluation and selection of the best climate model that represents the climate in the area of interest. The candidate model selection procedure used in this study can be applied in other study regions as well to form an ensemble model that best projects the climate in their area of interest.

References

1. IPCC. Climate Change 2014: Synthesis Report. In *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2014; p. 151.
2. Kondratyev, K.Y.; Varotsos, C. Atmospheric Greenhouse Effect in the Context of Global Climate Change. *Il Nuovo Cimento* **1995**, *18*, 123–151. [CrossRef]
3. IPCC. Climate Change 2007: Synthesis Report. In *Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2007; p. 104.
4. Kim, J.; Waliser, D.E.; Hart, A.F.; Nikulin, G.; Favre, A.; Mattmann, C.A.; Zimdars, P.A.; Hewitson, B.; Crichton, D.J.; Jack, C.; et al. Evaluation of the CORDEX-Africa multi-RCM hindcast: Systematic model errors. *Clim. Dyn.* **2013**, *42*, 1189–1202. [CrossRef]
5. Coninck, H.d.; Stephens, J.C.; Metz, B. Global learning on carbon capture and storage: A call for strong international cooperation on CCS demonstration. *Energy Policy* **2009**, *37*, 2161–2165. [CrossRef]
6. Yoo, C.; Cho, E. Comparison of GCM Precipitation Predictions with Their RMSEs and Pattern Correlation Coefficient. *Water* **2018**, *10*, 28. [CrossRef]
7. Boko, M.; Niang, I.; Nyong, A.; Vogel, C.; Githeko, A.; Medany, M.; Osman-Elasha, B.; Tabo, R.; Yanda, P. Africa. In *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 433–467.
8. Nikulin, G.; Jones, C.; Giorgi, F.; Asrar, G.; BüChner, M.; Cerezo-Mota, R.; Christensen, O.B.; DéQué, M.; Fernandez, J.; Nsler, A.H.; et al. Precipitation Climatology in an Ensemble of CORDEX-Africa Regional Climate Simulations. *J. Clim.* **2012**, *25*, 6057–6078. [CrossRef]
9. Schlenker, W.; Lobell, D.B. Robust negative impacts of climate change on African agriculture. *Environ. Res. Lett.* **2010**, *5*, 014010. [CrossRef]
10. Burke, M.B.; Miguel, E.; Satyanath, S.; Dykema, J.A.; Lobell, D.B. Warming increases the risk of civil war in Africa. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20670–20674. [CrossRef]
11. Dell, M.; Jones, B.F.; Olken, B.A. *Climate Change and Economic Growth: Evidence from the Last Half Century*; National Bureau of Economic Research; Massachusetts Avenue: Cambridge, MA, USA, 2008.
12. Buytaert, W.; Célleri, R.; Timbe, L. Predicting climate change impacts on water resources in the tropical Andes: Effects of GCM uncertainty. *Geophys. Res. Lett.* **2009**, 36. [CrossRef]
13. Gorguner, M.; Kavvas, M.L. Modeling impacts of future climate change on reservoir storages and irrigation water demands in a Mediterranean basin. *Sci. Total Environ.* **2020**, *748*, 141246. [CrossRef]
14. Gebrechorkos, S.H.; Bernhofer, C.; Hulsmann, S. Climate change impact assessment on the hydrology of a large river basin in Ethiopia using a local-scale climate modelling approach. *Sci. Total Environ.* **2020**, *742*, 140504. [CrossRef]
15. Kundzewicz, Z.W.; Krysanova, V.; Benestad, R.E.; Hov, Ø.; Piniewskic, M.; Otto, I.M. Uncertainty in climate change impacts on water resources. *Environ. Sci. Policy* **2018**, *79*, 1–8. [CrossRef]
16. Arora, M. Uncertainties in Climate Change Projection. *Int. J. Adv. Innov. Res.* **2019**, *6*, 1–7.
17. Gaudard, L.; Gabbi, J.; Bauder, A.; Romerio, F. Long-term Uncertainty of Hydropower Revenue Due to Climate Change and Electricity Prices. *Water Resour. Manag.* **2016**, *30*, 1325–1343. [CrossRef]
18. Reichler, T.; Kim, J. Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. *J. Geophys. Res.* **2008**, *113*, 1–13. [CrossRef]
19. Fatichi, S.; Ivanov, V.Y.; Paschalis, A.; Peleg, N.; Molnar, P.; Rimkus, S.; Kim, J.; Burlando, P.; Caporali, E. Uncertainty partition challenges the predictability of vital details of climate change. *Earths Future* **2016**, *4*, 240–251. [CrossRef]
20. Vetter, T.; Reinhardt, J.; Flörke, M.; van Griensven, A.; Hattermann, F.; Huang, S.; Koch, H.; Pechlivanidis, I.G.; Plötner, S.; Seidou, O.; et al. Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Clim. Chang.* **2016**, *141*, 419–433. [CrossRef]

21. Kotlarski, S.; Szabó, P.; Herrera, S.; Räty, O.; Keuler, K.; Soares, P.M.; Cardoso, R.M.; Bosshard, T.; Pagé, C.; Boberg, F.; et al. Observational uncertainty and regional climate model evaluation: A pan-European perspective. *Int. J. Climatol.* **2019**, *39*, 3730–3749. [CrossRef]

22. Burke, M.; Dykema, J.; Lobell, D.B.; Miguel, E.; Satyanath, S. Incorporating Climate Uncertainty in to Estimates of Climate Change Impacts. *Rev. Econ. Stat.* **2015**, *97*, 461–471. [CrossRef]

23. Maurer, E.P. Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios. *Clim. Chang.* **2007**, *82*, 309–325. [CrossRef]

24. IPCC. *IPCC Special Report Emissions Scenarios summary for Policy Maker*; A Special Report of IPCC Working Group III; IPCC: Geneva, Switzerland, 2000.

25. Almazroui, M.; Nazrul Islam, M.; Saeed, S.; Alkhalaf, A.K.; Dambul, R. Assessment of Uncertainties in Projected Temperature and Precipitation over the Arabian Peninsula Using Three Categories of Cmip5 Multimodel Ensembles. *Earth Syst. Environ.* **2017**, *1*, 23. [CrossRef]

26. Aloysius, N.R.; Sheffield, J.; Saiers, J.E.; Li, H.; Wood, E.F. Evaluation of historical and future simulations of precipitation and temperature in central Africa from CMIP5 climate models. *J. Geophys. Res. Atmos.* **2016**, *121*, 130–152. [CrossRef]

27. Maraun, D. Bias Correcting Climate Change Simulations—A Critical Review. *Adv. Model.* **2016**, *2*, 211–220. [CrossRef]

28. Endris, H.S.; Omondi, P.; Jain, S.; Lennard, C.; Hewitson, B.; Chang'a, L.; Awange, J.L.; Dosio, A.; Ketiem, P.; Nikulin, G.; et al. Assessment of the Performance of CORDEX Regional Climate Models in Simulating East African Rainfall. *J. Clim.* **2013**, *26*, 8453–8475. [CrossRef]

29. Luhunga, P.; Botai, J.; Kahimba, F. Evaluation of the performance of CORDEX regional climate models in simulating present climate conditions of Tanzania. *J. South. Hemisph. Earth Syst. Sci.* **2016**, *66*, 32–54. [CrossRef]

30. Akinsanola, A.A.; Ogunjobi, K.O.; Gbode, I.E.; Ajayi, V.O. Assessing the Capabilities of Three Regional Climate Models over CORDEX Africa in Simulating West African Summer Monsoon Precipitation. *Adv. Meteorol.* **2015**, *13*, 935431. [CrossRef]

31. Molina, M.J. Complexity in climate change science. In *Complexity and Analogy in Science: Theoretical, Methodological and Epistemological Aspects*; Pontifical Academy of Sciences: Vatican City, Italy, 2014.

32. Bader, D.; Covey, C.; Gutowski, W.; Held, I.; Kunkel, K. *Climate Models: An Assessment of Strengths and Limitations*; US Department of Energy Publications: Lincoln, NE, USA, 2008; Volume 8.

33. Pirani, A.; Meehl, G.; Bony, S. WCRP/CLIVAR working group on coupled modeling (WGCM) activity report: Overview and contribution to the WCRP crosscut on anthropogenic climate change. *Newsl. CLIVAR* **2009**, *14*, 20–25.

34. Déquéa, M.; Calmanti, S.; Christensen, O.B.; Aquila, A.D.; Maulec, C.F.; Haensler, A.; Nikulin, G.; Teichmann, C. A multi-model climate response over tropical Africa at +2 °C. *Clim. Serv.* **2017**, *7*, 87–95. [CrossRef]

35. Massoud, E.; Tian, B.; Lee, H.; Waliser, D.E.; Gibson, P.B. Climate Model Evaluation in the Presence of Observational Uncertainty: Precipitation Indices over the Contiguous United States. *J. Hydrometeorol.* **2019**, *20*, 1339–1357. [CrossRef]

36. Zumwald, M.; Knüsel, B.; Baumberger, C.; Hirsch Hadorn, G.; Bresch, D.N.; Knutti, R. Understanding and assessing uncertainty of observational climate datasets for model evaluation using ensembles. *WIREs Clim. Chang.* **2020**, *11*, e654. [CrossRef]

37. Haile, G.; Kassa, A. Investigation of Precipitation and Temperature Change Projections in Werii Watershed, Tekeze River Basin, Ethiopia; Application of Climate Downscaling Model. *J. Earth Sci. Clim. Chang.* **2015**, *6*, 300. [CrossRef]

38. Gizaw, M.S.; Biftu, G.F.; Moges, S.A.; Gan, T.Y.; Koivusalo, H. Potential impact of climate change on streamflow of major Ethiopian rivers. *Clim. Chang.* **2017**, *143*, 371–383. [CrossRef]

39. Yimer, S.M.; Kumar, N.; Bouanani, A.; Tischbein, B.; Borgemeister, C. Homogenization of daily time series climatological data in the Eastern Nile basin, Ethiopia. *J. Theor. Appl. Climatol.* **2021**, *143*, 737–760. [CrossRef]

40. Chen, J.; Brissette, F.P.; Lucas-Picher, P.; Caya, D. Impacts of weighting climate models for hydro-meteorological climate change studies. *J. Hydrol.* **2017**, *549*, 534–546. [CrossRef]

41. Suh, M.S.; Oh, S.G.; Lee, D.K.; Cha, D.H.; Choi, S.J.; Jin, C.S.; Hong, S.Y. Development of New Ensemble Methods Based on the Performance Skills of Regional Climate Models over South Korea. *J. Clim.* **2012**, *25*, 7067–7082. [CrossRef]

42. WMO. *Guidelines on the Calculation of Climate Normals*; World Meteorological Organization (WMO): Geneva, Switzerland, 2017.

43. Zong-Ci, Z.; Yong, L.; Jian-Bin, H. A Review on Evaluation Methods of Climate Modeling. *Adv. Clim. Chang. Res.* **2013**, *4*, 137–144. [CrossRef]

44. Akinsanola, A.A.; Ogunjobi, K.O.; Ajayi, V.O.; Adefisan, E.A.; Omotosho, J.A.; Sanogo, S. Comparison of five gridded precipitation products at climatological scales over West Africa. *Meteorol. Atmos. Phys.* **2016**, *129*, 669–689. [CrossRef]

45. Samadi, S.Z.; Sagareswar, G.; Tajiki, M. Comparison of General Circulation Models: Methodology for selecting the best GCM in Kermanshah Synoptic Station, Iran. *Int. J. Glob. Warm.* **2010**, *2*, 347–365. [CrossRef]

46. Gleckler, P.J.; Taylor, K.E.; Doutriaux, C. Performance metrics for climate models. *J. Geophys. Res.* **2008**, *113*, D06104. [CrossRef]

47. Pellicone, G.; Caloiero, T.; Modica, G.; Guagliardi, I. Application of several spatial interpolation techniques to monthly rainfall data in the Calabria region (southern Italy). *Int. J. Climatol.* **2018**, *38*, 3651–3666. [CrossRef]

48. Taylor, K.E.; Stouffer, R.J.; Meehl, G.A. An Overview of CMIP5 and the Experiment Design. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 485–498. [CrossRef]

49. Ayugi, B.; Tan, G.; Gnitou, G.T.; Ojara, M.; Ongoma, V. Historical evaluations and simulations of precipitation over East Africa from Rossby centre regional climate model. *Atmos. Res.* **2020**, *232*, 104705. [CrossRef]

50. Ongoma, V.; Chen, H.; Gao, C. Evaluation of CMIP5 twentieth century rainfall simulation over the equatorial East Africa. *Theor. Appl. Climatol.* **2019**, *135*, 893–910. [CrossRef]

51. Harris, I.; Jones, P.D.; Osborn, T.J.; Lister, D.H. Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset. *Int. J. Climatol.* **2014**, *34*, 623–642. [CrossRef]

52. Ongoma, V.; Chen, H. Temporal and spatial variability of temperature and precipitation over East Africa from 1951 to 2010. *Meteorol. Atmos. Phys.* **2016**, *129*, 131–144. [CrossRef]

53. Mutai, C.C.; Ward, M.N. East African Rainfall and the Tropical Circulation/Convection on Intraseasonal to Interannual Timescales. *J. Clim.* **2000**, *13*, 3915–3939. [CrossRef]

54. Mumo, L.; Yu, J. Gauging the performance of CMIP5 historical simulation in reproducing observed gauge rainfall over Kenya. *Atmos. Res.* **2020**, *236*, 104808. [CrossRef]

55. Laprise, R.; Hernández-Díaz, L.; Tete, K.; Sushama, L.; Šeparović, L.; Martynov, A.; Winger, K.; Valin, M. Climate projections over CORDEX Africa domain using the fifth-generation Canadian Regional Climate Model (CRCM5). *Clim. Dyn.* **2013**, *41*, 3219–3246. [CrossRef]

56. Otieno, V.O.; Anyah, R.O. CMIP5 simulated climate conditions of the Greater Horn of Africa (GHA). Part 1: Contemporary climate. *Clim. Dyn.* **2012**, *41*, 2081–2097. [CrossRef]

57. Yang, W.; Seager, R.; Cane, M.A.; Lyon, B. The Annual Cycle of East African Precipitation. *J. Clim.* **2015**, *28*, 2385–2404. [CrossRef]

58. Sperber, K.; Palmer, T. Interannual tropical rainfall variability in General Circulation Model Simulations Associated with the Atmospheric Model Intercomparison Project. *J. Clim.* **1996**, *9*, 2727–2750. [CrossRef]

59. Akinsanola, A.A.; Ajayi, V.O.; Adejare, A.T.; Adeyeri, O.E.; Gbode, I.E.; Ogunjobi, K.O.; Nikulin, G.; Abolude, A.T. Evaluation of rainfall simulations over West Africa in dynamically downscaled CMIP5 global circulation models. *Theor. Appl. Climatol.* **2017**, *113*, 437–450. [CrossRef]

60. Warnatzsch, E.A.; Reay, D.S. Temperature and precipitation change in Malawi: Evaluation of CORDEX-Africa climate simulations for climate change impact assessments and adaptation planning. *Sci. Total Environ.* **2019**, *654*, 378–392. [CrossRef] [PubMed]

61. Qian, J.-H.; Zubair, L. The Effect of Grid Spacing and Domain Size on the Quality of Ensemble Regional Climate Downscaling over South Asia during the Northeasterly Monsoon. *Mon. Weather. Rev.* **2010**, *138*, 2780–2802. [CrossRef]

62. Akinsanola, A.A.; Ogunjobi, K.O. Evaluation of present-day rainfall simulations over West Africa in CORDEX regional climate models. *Environ. Earth Sci.* **2017**, *76*, 366. [CrossRef]

63. Graham, L.P.; Andréasson, J.; Carlsson, B. Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods—A case study on the Lule River basin. *Clim. Chang.* **2007**, *81*, 293–307. [CrossRef]

64. Wilby, R.L.; Harris, I. A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water Resour. Res.* **2006**, *42*, W02419. [CrossRef]

65. Chen, J.; Brissette, F.P.; Poulin, A.; Leconte, R. Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resour. Res.* **2011**, *47*, W12509. [CrossRef]

66. Reto, K.; Abramowitz, G.; Collins, M.; Veronika Eyring, P.; Gleckler, J.; Hewitson, B.; Mearns, L. *IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections*; National Center for Atmospheric Research: Boulder, CO, USA, 2010; Available online: www.ipcc-wg1.unibe.ch (accessed on 12 December 2020).

67. Hagemann, S.; Chen, C.; Clark, D.B.; Folwell, S.; Gosling, S.N.; Haddeland, I.; Hanasaki, N.; Heinke, J.; Ludwig, F.; Voss, F.; et al. Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dyn.* **2013**, *4*, 129–144. [CrossRef]

68. Yang, W.; Seager, R.; Cane, M.A.; Lyon, B. The East African Long Rains in Observations and Models. *J. Clim.* **2014**, *27*, 7185–7202. [CrossRef]

69. Nikulin, G.; Asharaf, S.; Magariño, M.E.; Calmanti, S.; Cardoso, R.M.; Bhend, J.; Fernández, J.; Frías, M.D.; Fröhlich, K.; Früh, B.; et al. Dynamical and statistical downscaling of a global seasonal hindcast in eastern Africa. *Clim. Serv.* **2018**, *9*, 72–85. [CrossRef]

70. Kotlarski, S.; Keuler, K.; Christensen, O.B.; Colette, A.; Déqué, M.; Gobiet, A.; Goergen, K.; Jacob, D.; Lüthi, D.; van Meijgaard, E.; et al. Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev.* **2014**, *7*, 1297–1333. [CrossRef]

71. Wilcke, R.A.I.; Bärring, L. Selecting regional climate scenarios for impact modelling studies. *Environ. Model. Softw.* **2016**, *78*, 191–201. [CrossRef]