



Application of Boruta algorithms as a robust methodology for performance evaluation of CMIP6 general circulation models for hydro-climatic studies

I. M. Lawal^{1,2} · D. Bertram¹ · C. J. White¹ · S. R. M. Kutty³ · I. Hassan² · A. H. Jagaba^{2,3}

Received: 18 October 2022 / Accepted: 17 April 2023
© The Author(s) 2023

Abstract

Regional climate models are essential for climate change projections and hydrologic modelling studies, especially in watersheds that are overly sensitive to changes in climate. Accurate hydrologic model development is a daunting task in data-sparse regions where climate change's impact on hydrologic and water quality processes is necessary for a well-informed policy decision on adaptation and hazard mitigation strategies. Novel approaches have been evolving that evaluated GCMs with the objective of improved parameterization to limit uncertainty and improve hydrologic model development. However, conclusions drawn should be purpose-driven based on intended usage. This study provides an overview of the state-of-the-art Boruta random forest as a robust methodology in the performance evaluation of GCMs models for hydroclimatic study. Highlights from the assessment indicate that (1) there is consistency in replicating the three observed climate variables of daily precipitation, maximum and minimum temperature respectively, (2) better temporal correlation ($R^2 = 0.95$) in annual precipitation with a mean bias of 0.638mm/year, when compared to symmetrical uncertainty (SU) ($R^2 = 0.82$), and all models ensembles (AME) ($R^2 = 0.88$) with associated biases of 68.19mm/year and 10.57mm/year, respectively. Evaluation of the multi-year climate extreme indices, trends and magnitude reveal that there is a fair representation of basin-scale observed climate extreme events. However, the Boruta random forest approach exhibited a better statistical trend and magnitude of the extreme event in the basin. The findings of the study revealed enhanced GCM dataset evaluation and present a simple and efficient methodology to examine the limitations associated with the selected GCM ensemble for impact study in hydrology.

Abbreviations

AIS	Aggregate importance score	MBE	Mean bias error
AME	All model ensemble	PGF	Princeton University Global Meteorological Forcing
BRF	Boruta-random forest optimizer algorithms	QM	Quantile mapping
CMIP6	Coupled model inter-comparison project phase 6	SC	Similarity coefficient
CPC	Climate prediction centre	SPEI	Standardized precipitation evapotranspiration index
EQM	Empirical quantile mapping	SU	Symmetrical uncertainty
GCM	Global climate models or general circulation models		
IS	Importance score		

✉ I. M. Lawal
ibrahim.lawal@strath.ac.uk

¹ Department of Civil and Environmental Engineering, University of Strathclyde, James Weir Building, 75 Montrose Street, G1 1XJ, Glasgow, UK

² Department of Civil Engineering, Abubakar Tafawa Balewa University, Bauchi, Nigeria

³ Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia

1 Introduction

To reduce observational uncertainty and the impact of projected changes in climate and catchment hydrologic variables on the availability of water resources for an accurate assessment of sustainability concerns at local, regional, and global scales, a successful hydro-climatic study requires accurate data with high temporal and spatial resolution. Therefore, future variations of global surface air temperature and precipitation are deemed important for climate change policy formulation, and hazard assessment which

may affect the human livelihood and regional economy, especially in developing economies (Akhter et al., 2017) and ensure proper mitigation and adaptation strategies are adopted for integrated water resources management (Ahmed, Sachindra, et al. 2019a).

Global energy balance has undergone periodic changes due to increased emission of atmospheric greenhouse gas owing to widespread fossil fuels usage and industrial activities across the globe (Chu et al., 2010; Huang et al., 2011; Salman et al., 2018). The rise in the concentration of greenhouse gases has been noted to increase the Earth's temperature at a rate of 0.15 °C/decade (IPCC, 2013), and has potentially a severe effect on the earth's ecosystem and most notably in the tropical regions (Wang et al. 2014; Mohsenipour et al. 2018; Khan et al. 2019). (Mishra and Liu, 2014) observed that the impact of the variations of rainfall-induced by projected climate change would be more intense in the tropical regions of the world. This phenomenon has become a significant socio-economic and political issue (Alamgir et al., 2019).

General circulation models are the main tool for predicting future changes in the global climate (Maraun et al., 2010) because they possessed the potential in replicating historical climatic changes as well as future changes considering the greenhouse gas concentration (Goyal et al., 2012) and other shared socio-economic pathways integrated into the recently developed coupled model inter-comparison project phase six models. Evaluating the performance of general circulation models (GCM) outputs is instrumental for simulating the historical and future basin scale hydrologic cycles using hydrological models. The accuracy of the precipitation and temperature outputs from GCMs, when used as an input for hydrological modelling, affects the reliability of hydrologic variables prediction, and therefore the spatial and temporal performances as well as their seasonal variations need to be evaluated arising from natural and climate-induced radiative forcing in a multi-model context (Eyring et al. 2016; Gusain, Ghosh, and Karmakar 2020; Wang et al. 2021) to reduce uncertainties and enhance prediction accuracy for reliable policy planning of basin scale hydrology.

GCMs are developed to produce projections at a coarse spatial scale and could not resolve finer scale features such as clouds and land use change. Studies at the regional scale require that the output be downscaled to finer a resolution. The mismatch between the hydro-climatic information required and the GCM output is a major cause for concern or obstacle in hydrologic impact studies (Willems and Vrac, 2011). Careful assessment of various downscaling techniques is important to enable accurate prediction and a solution to the mismatch between regional hydro-climatic information and GCM outputs at a spatial scale of between 5 and 50 km (Yang and Delsole, 2012). However, McSweeney et al., (2015) argued that downscaling the

complete ensemble may not be desirable or necessary to produce a meaningful range of future climate conditions relevant to evaluate hazards associated with future climate change because high resolution downscaling is labour and computing resource intensive and therefore, various strategies need to be explored to sample from the available CMIP6 GCMs and their shared socio-economic pathways (SSPs) scenarios to generate projections relevant for water policies at catchment scale. The assessment process is to identify the challenges and opportunity to exclude models deemed unsatisfactory in the representation of key climate features.

These models have been utilized to simulate the historical and projected changes in climate at global (Sachindra et al., 2014; Wright et al., 2015), regional (Salman et al. 2018; Ahmed, Sachindra, et al. 2019a; Abbasian, Moghim, and Abrishamchi 2019) and local basin scale (Akhter et al., 2017; Hassan et al., 2020). However, over the years, six phases of the Coupled Model Inter-comparison Projects (CMIPs) have been developed by various modelling centres for climate change studies. Previous studies utilizing the fifth phase (CMIP5) GCMs indicated some significant improvements in simulated climate variables relative to the latter generation (CMIP3) models (Wang et al. 2016), due to improved knowledge of climate science. Studies that utilized CMIP5 GCMs have identified the Sahel region, tropical West Africa and Southern part of Africa as hotspot for severe impact of regional climate change (Differbaugh and Giorgi 2012; Niang, Ruppel, Abdrabo, Essel, Lennard, Padgham 2014). Although studies are evolving utilizing CMIP6 models recently to understand their ability in replicating historical and future climate at the regional and global scale. There are notable differences between the CMIP6 GCMs and the earlier phases which integrate new specification for greenhouse gas concentration and emission scenarios, as well as land use scenarios (Gidden et al., 2019). A limited number of studies utilizing the CMIP6 GCMs indicated improvement and its robustness over earlier phases in some regions, for example Australia (Grose et al., 2020); Africa (Almazroui et al., 2020; Ayugi et al., 2021; Sian et al., 2021); Nigeria (Shiru and Chung, 2021); and South Korea (Song et al., 2020). The performances may vary from one region to another and therefore, necessitate the evaluation of their performance at any region of interest, especially in data sparse catchments before adoption for proper representation of the climatic feature to avoid projections with large uncertainty range which may result to over confidence and poor adaptation (McSweeney et al., 2015).

Several techniques have been used to assess the performance of climate models such as ensemble averaging (Giorgi and Mearns, 2002), combined statistical measures like root mean square error, mean bias error, mean absolute error, correlation coefficient into one performance index (Gu et al. 2015; Wu et al. 2017), relative entropy (Shukla

et al., 2006), probability density functions (PDF) (Perkins et al., 2007), Bayesian regression (Chandler, 2013), clustering (Knutti et al., 2013), correlation (Li et al., 2016; Xuan et al., 2017), and symmetrical uncertainty (Salman et al., 2018) and often times the methods are quite cumbersome or produce inconsistent performance across climate variables that will require a multi-criteria decision tool for selection of appropriate GCMs for hydrologic impact studies.

However, based on previous studies, no consensus was made regarding the choice of GCM selection approach and McMahan et al., (2015) posited that no single criterion is accepted universally for GCM performance, although assessment at multiple time scales and the ability of a GCM capturing the spatial structure of a catchment's key climate feature may give crucial information for water resource management accurately, especially at basins with high climate variability (Gleckler, Taylor, and Doutriaux 2008; Ahmed, Sachindra, et al. 2019a). The downside of the statistical metric is that they only assessed certain features of the time series data when compared to the observed data (McSweeney et al., 2015; Weigel et al., 2010), and often provide contradictory results across different metrics (Nashwan and Shahid, 2019; Raju et al., 2017).

Studies on GCM selection are grouped into the past performance method (Biemans et al., 2013), where the models are selected based on their capability to replicate the historical climate and the envelope method (Warszawski et al., 2014) where the ensemble of GCMs are selected based on the ability to encompass the whole range of future projections. However, both of these approaches have weaknesses; for example the former method neglects agreement between GCMs to simulate projected future climate, while the latter method ignores the ability of GCMs to replicate the historical observed climate (Ahmed et al., 2019b).

Machine learning algorithms are gaining a lot of attention, especially filters and wrappers such as clustering (Knutti et al., 2013; Raju and Kumar, 2016), Bayesian weighting (Min and Hense, 2006), and weighted skill score (Maxino et al., 2008; Perkins et al., 2007). Furthermore, these techniques are used to evaluate and select the GCMs using a single performance index (Ahmadalipour et al., 2017; Fu et al., 2013; Raju and Kumar, 2016; Wójcik et al., 2014). However, the techniques are found to have some inherent weaknesses such as the inability to capture the temporal variability of climate, and, the variation of frequency of climate extreme which is found to be an important factor in the assessment of model performance (Salman et al., 2018).

The major advantages of machine learning techniques are that, in feature selection where the dependent variables are analysed and ranked based on their importance or impact on the independent variable, and where features that are likely to decrease the efficiency of a model are screened out to reduce uncertainty in model development. It is important in

machine learning application to have the observed data well represented and the information within the data series possessed the ability to be learnable for the modelling process, as this is critical for the final performance (Kursa, 2016).

In GCM selection for hydro-climatic study, the information entropy based filter referred to as symmetrical uncertainty (SU) (Witten et al., 2005) has gained the attention of researchers due to its ability to select variable without bias and reliably. The technique was used to rank GCMs according to their degree of similarity or otherwise with the observations for the entire time series data and it has the advantage of giving a universal metric for the relationship between dependent and independent features irrespective of the shape of the underlying distributions (Wu and Zhang 2004) and has been used in various studies (Ahmed et al., 2019b; Nashwan and Shahid, 2019; Pour et al., 2018; Salman et al., 2018).

The technique has shown to be promising and possessed a similar or better skill for performance evaluation and selection of appropriate GCM dataset for hydro-climatic study as compared to other available methods such as compromise programming, wavelet-based skill score (WSS), statistical metrics which sometimes exhibits contradictory results and makes decisions on performance and selection difficult (Ahmed, Shahid, et al. 2019b), although most filter based classifiers are known to use single algorithms to integrate variable selection and modelling and often evaluate univariate or very simple interactions between attributes and decisions, which affects the outcomes or performances of features (Kursa, 2016).

However, machine learning approaches are evolving, for example, the wrapper-based Boruta algorithms for feature selection in model building and has been applied in other disciplines such as, Kursa (2016) showed the relative skill or effectiveness of the methodology in feature selection of random ferns classifier, Ahmed et al. (2021) applied the technique for soil moisture estimation under global warming scenarios. Advancement in computational capabilities has ensured that machine and deep learning techniques are useful for accurate variable prediction due to their ability to extract, process and handle relatively large amount of complex data with high degree of variable mapping skills and efficiency (Gong et al., 2019).

Numerous studies have successfully implemented different feature selection algorithms such as information entropy (Shukla et al., 2006), Bayesian weighting (Min and Hense, 2006), elastic net and ridge regression (Hammami et al., 2012), artificial neural networks (Hajnayeb et al., 2011), support vector machine (Maldonado and Weber, 2009), random forest (Genuer et al., 2010), neighbourhood component analysis for regression (Ghimire et al. 2019a), and iterative input selection algorithm (Prasad et al., 2017) for hydro-climatic studies. Among the techniques, symmetrical

uncertainty (SU) that is based on information entropy (William et al., 1996) is popular and has gain the attention of researchers (Kannan and Ramaraj 2010; Nashwan and Shahid 2019; Ahmed, Shahid, et al. 2019b; Sa'adi et al. 2020). However, there are identified shortcomings of the various method such as, inter-dependencies among the models are ignored for a known feature which may result in the selection of inappropriate GCMs due to overfitting and some performance indices are based on the state of the climate as a whole and temporal variability is not considered which is critical in model performance assessment (Reichler and Kim 2008; Ahmed, Shahid, et al. 2019b; Hassan et al. 2016).

This study employed and pilot the use of Boruta-random forest algorithms (BRF) developed by Kurasa et al., (2010) for performance evaluation and selection of appropriate CMIP6 GCM models to examined their individual capability to accurately simulate observed daily historical climate variables in the basins due to its high sensitivity to climate change, and in this case, Lake Chad basin was adopted to examined the robustness of the methodology due in part to its highly skilled variable mapping as a requirement for input parameters required in hydrologic study, uniqueness and high climate variability across the basin of interest and this technique has been used and recommended based on previous study for example, Prasad et al. (2019) utilised BRF to predict soil moisture, Christ et al. (2016) applied BRF for industrial big data application in distributed and parallel time series feature extraction, Leutner et al. (2012) predict forest biodiversity and Lyu et al. (2017) applied the concept to forecast air quality, where these algorithms were used to define significant input parameters by comparing the real features or variables to those of random probes and all the studies have suggested a convincing outcome of model accuracy. However, the authors acknowledged that, this is first attempt to have applied the proposed technique for GCM evaluation and selection and to validate the methodology, an information aggregation approach was adopted to combine the ranks of the GCMs across the grid points in the entire basin to identify the best ensemble of the CMIP6 GCM for simulation of the above variables. The result of this approach will be compared with the well-established symmetrical uncertainty technique to understand the efficacy, applicability and the inherent uncertainties of the model evaluation and selection approach.

Furthermore, earlier studies that utilized simulation based studies to investigate the impact of climate change on freshwater hydrology required data in the form of daily precipitation and temperature series to drive various hydrological watershed models; however, the performance of the model is dependent upon the driving general circulation models, internal parameterizations and model domain configuration (Déqué, 2007).

Therefore, it is important to note that conclusions drawn from studies that evaluated and select GCM performance at monthly time scale (Salman et al. 2018; Abbasian, Moghim, and Abrishamchi 2019; Ahmed, Shahid, et al. 2019b) may well not be relied upon for studies that requires climate data at daily time step for hydrologic modelling processes, in order to reduce biases and observational uncertainties for realistic predictions of climate change impact, adaptation and resilience in hazard assessment. The performance of GCMs from BRF, SU and an all-ensemble average approach adopted in this study will be evaluated for a realistic assessment of basin scale features and climate variable dynamics.

Finally, this study intends to propose and provide a robust approach that enhance and preserve climate signals' internal parameterization exerted by re-gridding, downscaling and bias correction to realistically utilize the optimal amount of GCM dataset capable of assessing the complex interactions within the Earth system (hydrologic) models which are essential for accurate understanding and forecasting of long-term changes of basin hydrology resulting from external forcing which are not adequately addressed in previous research especially in data-sparse regions within an acceptable level of uncertainty.

2 Case study area and data

2.1 Case study area

The Lake Chad Basin is one of the world's largest endorheic basins, with an estimated area of $\sim 2,500,000$ km², (Coe and Foley, 2001; Gao et al., 2011). It is situated at latitudes of 5.2° – 25.3° N and longitudes of 6.9° – 24.5° E, in the transition zone between the Sahara and the Sudano-Sahelian regions of West Africa (Ndehedehe et al., 2018), (Fig. 1). The basin provides the primary source of freshwater for livestock grazing, agricultural production, and fish farming (Buma et al., 2016). The basin is divided into four climatic zones namely the Saharan, Sahelo-Saharan, Sahelo-Sudanian, and the Sudano-Guinean zone, with an annual precipitation range of < 100–1500 mm respectively, and has an average annual temperature of between 35 and 40 °C in the northern part to as low as 26.5 °C in the southern part of the basin (Nkiaka, Nawaz, and Lovett 2018a) characterised by hot, wet and dry weather condition from March to June, from June to October, and from November to February, respectively (Mahmood et al., 2019). The basin is situated in an area with minimal relief, no surface outflow, and a spatial extent that is highly vulnerable to climate change, with elevations ranging from –330 to 3446 m (see Fig. 1), However, according to (Coz et al. 2009; Nkiaka, Nawaz, and Lovett 2018b), with the exception of few isolated hills, plateaus, and mountains

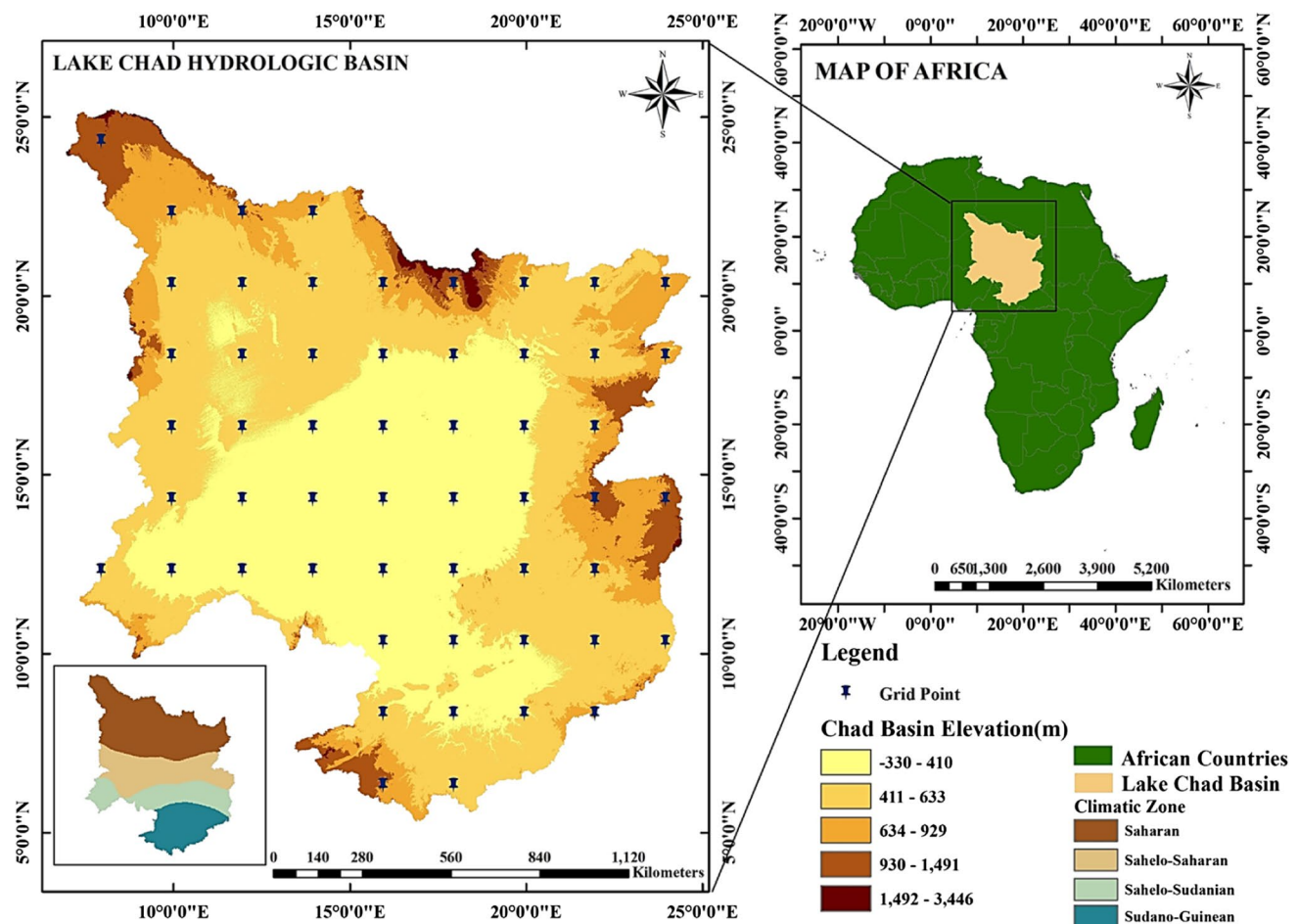


Fig. 1 Lake Chad basin showing elevation and climate stations for the proposed study

in the southern and northern regions, the basin is generally flat ground with an average slope of $< 1.3\%$.

2.2 Gridded data and sources

Several gridded climate data sources were explored and analysed for suitable application in climate studies in the Chad Basin due to inadequate gauged based meteorological data in most part of the world Sub-Saharan Africa and the Mediterranean, in particular. This study utilized the daily gridded precipitation data of the US Climate Prediction Centre (CPC), optimal interpolation of station or gauged based records of GTS (Xie et al., 2007) and daily maximum and minimum temperature data of the Princeton University Global Meteorological Forcing PGF v.2 from forcing's of NCEP-NCAR reanalysis and other global data by bilinear interpolation (Sheffield et al., 2006), from 1979 to 2012 and available at <https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globalprecip.html> and <http://hydrology.princeton.edu/data.pgf.php> respectively, as recommended by Lawal

et al., (2021) to represent adequately the true climatic features of the selected basin.

2.3 General circulation model data and sources

In this study, 16 coupled inter-comparison project phase 6 (CMIP6) general circulation models at daily time resolution were considered. These are available at <https://esgf-node.llnl.gov/projects/esgf-llnl/CMIP6> for the period 1979–2012 and consistent with the gridded data time scale for realistic performance evaluation. Details of the GCMs name, spatial resolutions, models' development centres and country of origin are provided in Table 1. The first ensemble members of the GCM were adopted for fair assessment. The models were chosen because they have the requisite historical and future climate change emissions scenario data (SSP1–SSP5) at a daily timestep as an important requirement for a hydrologic modelling study. This is essential for understanding watershed projected hazards due to climate change and provide valuable information to policy makers for informed decision on integrated water resource management.

Table 1 Summary of CMIP6 models considered in this study

Country	Modelling centre	Model name	Resolution in arc Deg. (Lon × Lat)	Model number
Australia	Commonwealth Scientific and Industrial Research Organisation; ARCCSS (Australian Research Council Centre of Excellence for Climate System Science)	ACCESS CM2	1.875×1.25	M1
Australia	Commonwealth Scientific and Industrial Research Organisation; ARCCSS (Australian Research Council Centre of Excellence for Climate System Science)	ACCESS ESM1-5	1.875×1.25	M2
China	Beijing Climate Centre	BCC CSM2-MR	1.10×1.10	M3
Canada	Canadian Centre for Climate Modelling and Analysis	CanESM5	2.81×2.81	M4
Sweden	EC-Earth consortium	EC-Earth3	0.70×0.70	M5
Sweden	EC-Earth consortium	EC-Earth3-Veg	0.70×0.70	M6
China	Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) modelling (Chinese Academy of Sciences)	F-Goals-g3	2.81×2.78	M7
China	Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) modelling (Chinese Academy of Sciences)	GFDL-ESM4	1.25×1.00	M8
Russia	Institute of Numerical Mathematics of the Russian Academy of Sciences	INM-CM4.8	1.12×1.12	M9
Russia	Institute of Numerical Mathematics of the Russian Academy of Sciences	INM-CM5.0	1.87×1.25	M10
France	Institut Pierre-Simon Laplace	IPSL-CM6A-LR	2.50×1.27	M11
Japan	Centre for Climate System Research; Japan Agency for Marine-Earth Science and Technology; National Institute for Environmental Studies	MIROC6	1.40×1.40	M12
Germany	Max-Planck-Institut für Meteorologie	MPI-ESM1.2-LR	1.875×1.875	M13
Japan	Meteorological Research Institute (MRI)	MRI-ESM2.0	1.125×1.125	M14
Norway	Norwegian Climate Consortium	NorESM2-LM	2.50×1.875	M15
Norway	Norwegian Climate Consortium	NorESM2-MM	1.25×0.9375	M16

(Source: <https://esgf-node.llnl.gov/projects/esgf-llnl/CMIP6>)

3 Research methodology

3.1 Methodology flow chart

The general procedure used for the evaluation of GCMs using gridded CPC and PGF climate data as surrogate for observed, selection, and further analysis of GCM ensemble is outlined in the methodology flow chart as shown in Fig. 2.

3.2 Statistical downscaling and bias correction of GCM

A multi-point statistical downscaling and bias correction approach was considered. The CMIP6 daily precipitation, maximum temperature and minimum temperature GCM models were interpolated to a common $2^\circ \times 2^\circ$ grid and spatially downscaled using bilinear interpolation approach for smooth transformation of coarse to fine resolution as recommended by Fischer et al., (2014) resulting into 54 grid point (Fig. 1), that covered the entire study area.

The deviation of the interpolated climate data was corrected to improve the GCM models agreement with the

observed data (CPC precipitation and PGF temperature), and to provide enhancement in model output because direct GCM output cannot be relied upon for accurate assessment of watershed climate features at local and regional scale required for impact studies due to their coarse resolution.

This is achieved by combining features of local observation and simulations resulting in insignificant biases and higher resolution climate projections using three well-known bias correction approaches and the detailed methodologies can be found in the cited references namely delta change (Navarro-Racines et al., 2020), quantile mapping (Cannon et al., 2015) and empirical quantile mapping technique (U. Ghimire et al., 2019b).

The grid based bias correction performance was evaluated using three statistical metrics e.g. correlation coefficient, mean bias error and index of agreement to understand the limitations of the various techniques to prevent misuse in selecting the best suited output relative to other method for further analysis and their detailed methodology can be found in (Taylor, 1997; Willmott, 1981; Willmott and Matsuura, 2005).

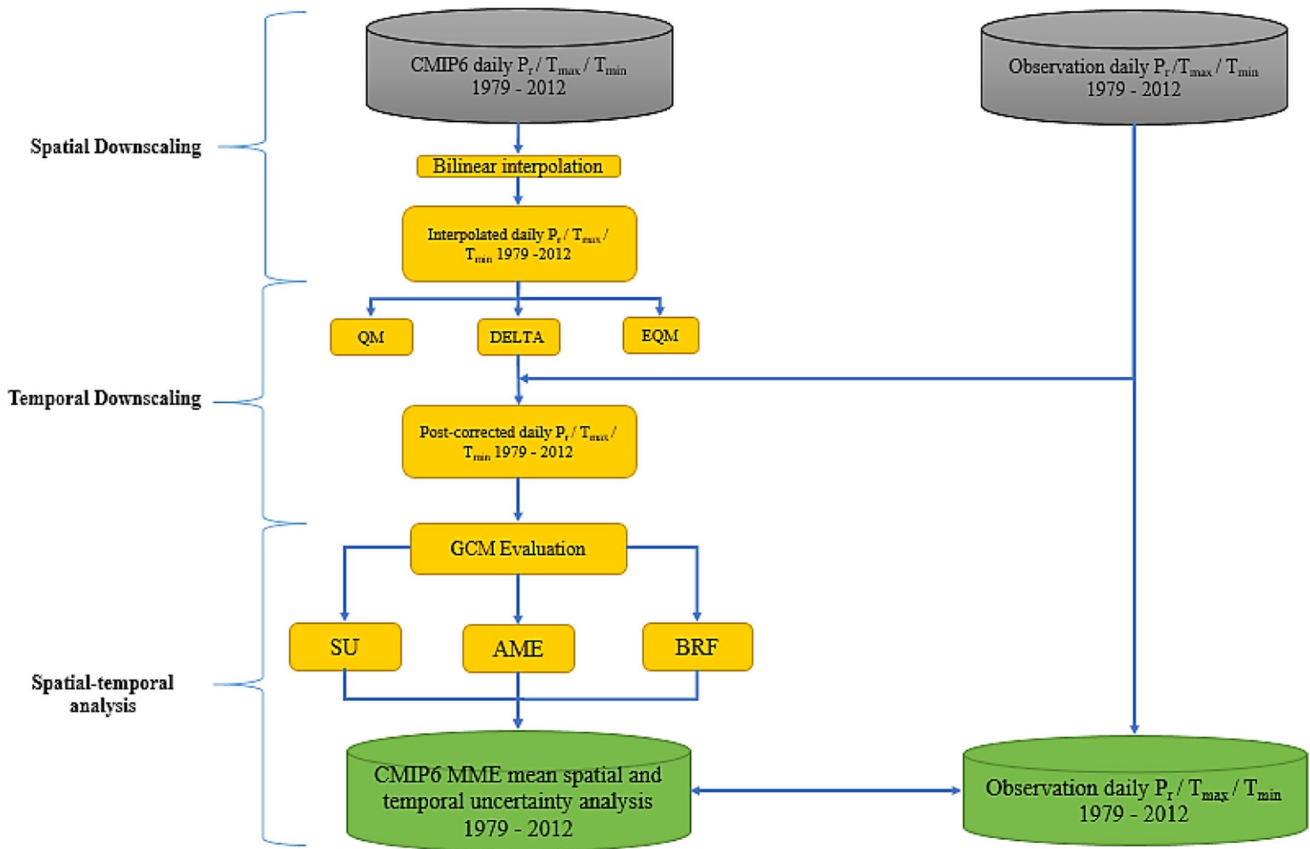


Fig. 2 GCM evaluation and selection for hydro-climatic study

3.3 Performance evaluation and selection of appropriate GCM output by machine learning technique

The daily bias corrected CMIP6 GCM outputs adopted were evaluated by two machine learning approaches to discover the significant features of the GCMs relative to the observation data. This section gives a brief overview of the machine learning based symmetric uncertainty and Boruta random forest algorithms for performance assessment and ensemble projections.

3.3.1 Symmetric uncertainty

Symmetric uncertainty was developed by William et al., (1996), and an entropy-based filter that assess pair-wise similarity between dependent and independent attribute irrespective of their probability distribution and interdependency (Wu and Zhang, 2004). It measures the information gain of the response random variable relative to the predictor and the lesser the entropy the greater the association of the data. The bias in the measurement is corrected by dividing the information gain with the sum of entropies of

the random features of the observation and GCMs precipitation, maximum and minimum temperature respectively. The symmetric uncertainty (SU) was estimated according to Eq. (1). Detailed methodology of the approach can be found in (Ahmed et al., 2019b; Lawal et al., 2021).

$$SU(x, y) = 2 \times \frac{MI(x, y)}{H(x) + H(y)} \tag{1}$$

In Eq. (1), $MI(x, y)$ is the mutual information gain, $H(x)$ and $H(y)$ represent the entropies of the GCM and observation respectively. Symmetric uncertainty values range from 0 to 1 indicating no similarity to perfect similarity respectively.

3.3.2 Boruta random forest algorithm

Boruta feature selection was developed based on random forest algorithm (Breiman, 2001). The approach was introduced by Stoppiglia et al. (2003) to identify significant input parameters from a host of many dependent features to match the attributes of an independent feature. The algorithm computes the Z-score of all input predictors and the distribution determines the shadow features as well as the important variables of the

predictors based on the Z-score metrics (Kursa and Rudnicki, 2010). The methodology adopted are shown below.

A duplicate or shadow random variable, q'_t for a particular vector, q_t to increase randomness and correlation between predictors and target variable (P_t), for all group of discrete inputs ($q_t \in R^n$), T and target variables ($P_t \in R$) for all inputs (n) and $t = 1, 2, 3, \dots, T$. The target variable is evaluated for association using the shadow variable q'_t and actual variable q_t .

- The variable importance measures or scores (mean decrease accuracy) for each predictor q_t and shadow q'_t attributes were computed for 500 iterations based on Eq. (2), to achieve better accuracy (Hur et al., 2017).

$$VIM = \frac{1}{m_{ree}} \sum_{m=1}^{m_{ree}} \frac{\sum_{t \in OOB} I(P_t = f(q_t)) - \sum_{t \in OOB} I(P_t = f(q'_t))}{|OOB|} \quad (2)$$

$I(\bullet)$ is an indication function, OOB is out-of-bag predicted error in the training samples; $P_t = f(q_t)$ and $P_t = f(q'_t)$ are predicted values before and after permutation.

- The standard score (Z-score) of the predictor and shadow attributes relative to the observation are computed as:

$$Z - \text{score} = \frac{VIM}{\sigma} \quad (3)$$

Where, σ is the measured standard deviation of accuracy losses, the minimum, median and maximum Z-score of the shadow features is computed and analysed relative to the predictor variable importance distribution of all dependent features. When all input features are confirmed to be important, or the iteration limit is achieved, the algorithm ends.

- The attribute is deemed important for a set iteration, if its variable importance score is higher than the maximum importance score of the shadow attributes.

3.4 Performance evaluation of selected GCM ensemble mean.

The weight or performance score of the two machine learning technique was used to rank the bias corrected precipitation, maximum and minimum temperature at individual grid point and the ranking were aggregated based on multi-criteria rating metrics using numerical averaging as recommended by Raju and Kumar (2016) for the entire study area.

Multi-model ensemble mean of the GCMs was developed by selecting four best GCM relative to the observation based on the two machine learning approaches (BRF and SU) and that of the 16 GCMs here referred as all model ensembles mean (AME) to further evaluate spatial and temporal changes in annual precipitation, maximum and minimum temperature changes, and extreme

conditions of the basin climate. Analysis of individual model bias as seen in other studies may not necessarily translate to better performance. Here, the study focused on the ensemble mean of the GCM to better understand the uncertainty (spread) of the combined model ensemble mean features, which is important and sensitive in their usage for hydrologic impact studies and reliable future projections of water resources.

3.5 Validation of evaluation approach based on return period of drought

The evaluation approaches considered, i.e., BRF, SU and AME, were further validated by examining the influence of the ensemble mean GCM precipitation, maximum and minimum temperature on the severity and returned period of drought against observation across the four climatic zone of the basin by estimating the 12-monthly standardized precipitation evapotranspiration index (SPEI) for the period of observation (1979–2012). The estimated SPEI time series were further used to assess the temporal pattern in the trend and their significance by Mann Kendall trend test and Sen’s slope estimator at 95% level of confidence.

3.6 Standardized precipitation evapotranspiration index

Standardized precipitation evapotranspiration index is a phenomenon that depicts water surplus or deficit within long climatic time scale that the difference between precipitation and potential evapotranspiration is calculated and then fitted with probability density function to estimate the return period of flood or droughts in a catchment. The severity was classified as $SPEI \geq 2.0$, 1.5 to 1.99, 1.0 to 1.49, 0.99 to -0.99 , -1.0 to -1.49 , -1.5 to -1.99 and $SPEI \leq -2.0$ to indicate extremely wet, severely wet, moderately wet, normal, moderate drought, severe drought, and extreme drought event respectively (Shekhar and Shapiro, 2019). Validation by SPEI captures the impact of temperature increase on water demand and natural variability of climate and its influence on climate change studies. Further evaluation to observe the trend and its significance was based on the methodology by (Henry, 1945; Kendall, 1948) and (Sen, 1968) given in the given equations.

The Mann-Kendall statistics is:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (4)$$

For a n number of time series, x_j and x_k are consecutive data values. The series sgn is as follows:

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{if } x_j > x_k \\ 0 & \text{if } x_j = x_k \\ -1 & \text{if } x_j < x_k \end{cases} \quad (5)$$

The mean $E(S)$, variance $V(S)$ and the Z statistics is evaluated as:

$$E(S) = 0 \quad (6)$$

$$V(S) = \frac{1}{18} \left\{ n(n-1)(2n+5) - \sum_{i=1}^p t_i(t_i-1)(2t_i+5) \right\}$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{for } S < 0 \end{cases} \quad (7)$$

Because the nonparametric Sen's slope estimator is resilient against outliers in time series analysis, it was used to calculate the size of the discovered trends in the time series data:

$$SS = \text{median} \left[\frac{x_j - x_i}{j - i} \right] \quad \text{for all } i < j \quad (8)$$

Where x_i represents the data value at a time step i and x_j represents the data value at time step j .

4 Results and discussion

4.1 Spatial and temporal downscaling and bias correction of CMIP6 GCMs

Figure 3(a–c) shows the example of post-corrected precipitation and maximum and minimum temperature data of GCM output ACCESS CM2 relative to the observations at grid point 1 (15.95°⁰, 6.31°⁰), for 1979–2014 and 1979–2012 respectively, using delta change, quantile mapping and empirical quantile mapping technique.

The evaluation results across the 54 grid point that covers the study area of the downscaled and the daily bias corrected GCM outputs of precipitation, maximum and minimum temperature indicated that delta change and empirical quantile mapping method is the most suitable for daily precipitation and maximum and minimum temperature with a recorded mean bias error, $MBE = 0$, mean correlation coefficient $R^2 = 0.8$ and modified index of agreement, $md = 0.86$ at 94% and a MBE range of $-0.01 - 0$, mean correlation coefficient $R^2 = 0.92$, and modified index of agreement $md = 0.96$ respectively, relative to other evaluation method across all the grid points. However, according to Fig. 3c, there are difficulties of downscaling models to capture peak values of minimum temperature across all evaluation methods, especially in the Sahelo-Saharan zone but this might be attributed to scale

gap between GCM outputs and observations and such variations cannot be accounted for by the GCM outputs as corroborated in Sachindra et al., (2014). Hence, GCM outputs from this evaluation technique were adopted for the next phase of analysis. The evaluation was limited to statistical downscaling because it was considered to be more flexible than dynamical technique and its projections can be downscaled to point specific locations as corroborated in Martinez-García et al., (2021).

The merit of this approach is to provide a reference for evaluating the accuracy of precipitation and temperature, which are the two critical hydrologic model input parameters, to improve the predictions of future river basin hydrologic cycles. The bias correction technique performed much better in simulating the observation in the Guinean-Sudanian zone, with a cluster of grid points with almost perfect correlation relative to the observation. However, Saharan zone exhibit some inadequacies in replicating the observed climate features which may be due in part to poor climate signals of sparse precipitation events. In general, the result indicates some improvements of the CMIP6 GCM in capturing the observation signals of the climate variables and has proven to restore the inter-station dependencies. The multi-site approach illustrated the capability in addressing the inequities of transitioning and interpolation of GCMs from coarse to finer resolution and vice versa to accurately reproduce observed multiple statistical properties of climate variable for improved hydrological climate change impact studies.

4.2 Performance evaluation and selection of CMIP6 GCMs

The individual models were coded M_1, M_2, \dots, M_{16} to represent ACCESS CM2, ACCESS ESM1-5, NorESM2-MM respectively, as appeared in Table 1, for ease of identification. The performance of the model across individual grid point was analysed based on the two machine learning approaches SU and BRF to investigate their significance in replicating the basin climatic features of precipitation and maximum and minimum temperature for the period 1979–2014 and 1979–2012 respectively, which was a compromise and limitations between CPC, PGF and historical CMIP6 GCM based on world meteorological organisation climatological standard normal period of data range and availability.

The GCMs were ranked based on their symmetric uncertainty (SU) coefficient (Section 4.2.1) and variable importance score (Section 4.2.2) at each grid point for all the climate variable to understand the degree of association or relative skill (importance) of the models with insignificant bias across the basin respectively. However, in BRF analysis, a zero score was recorded for GCMs whose variable

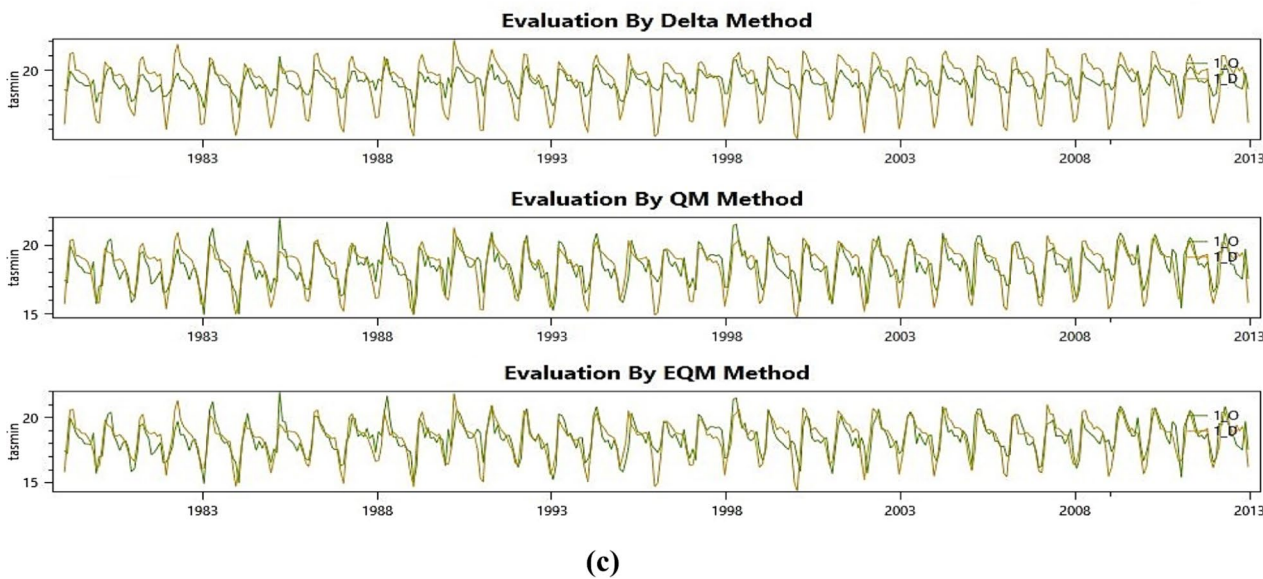
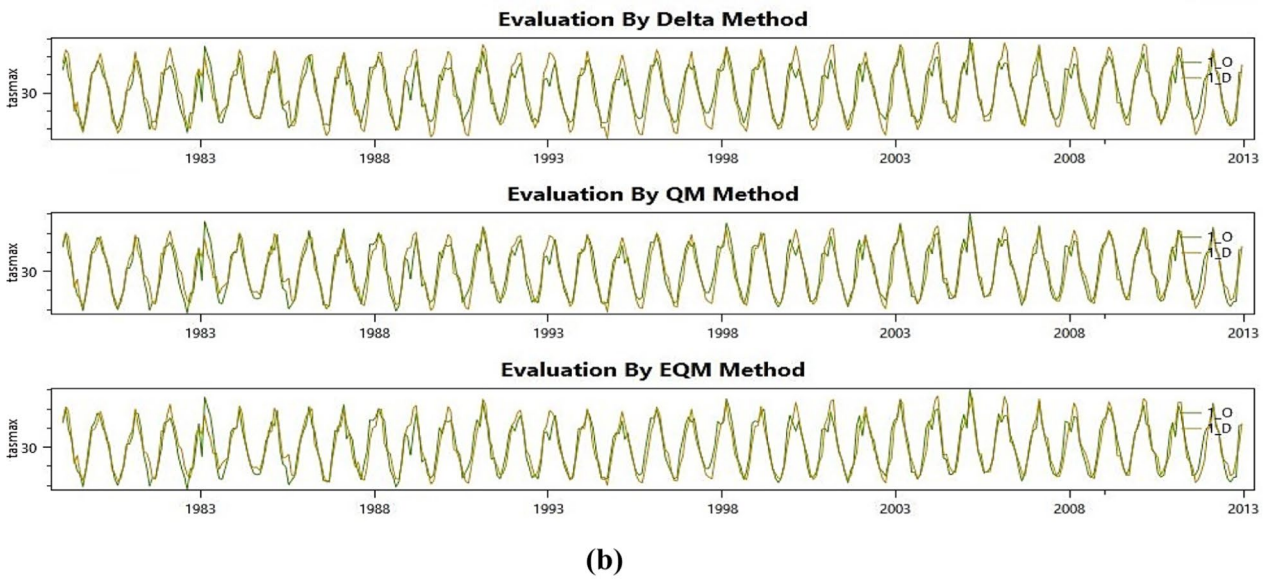
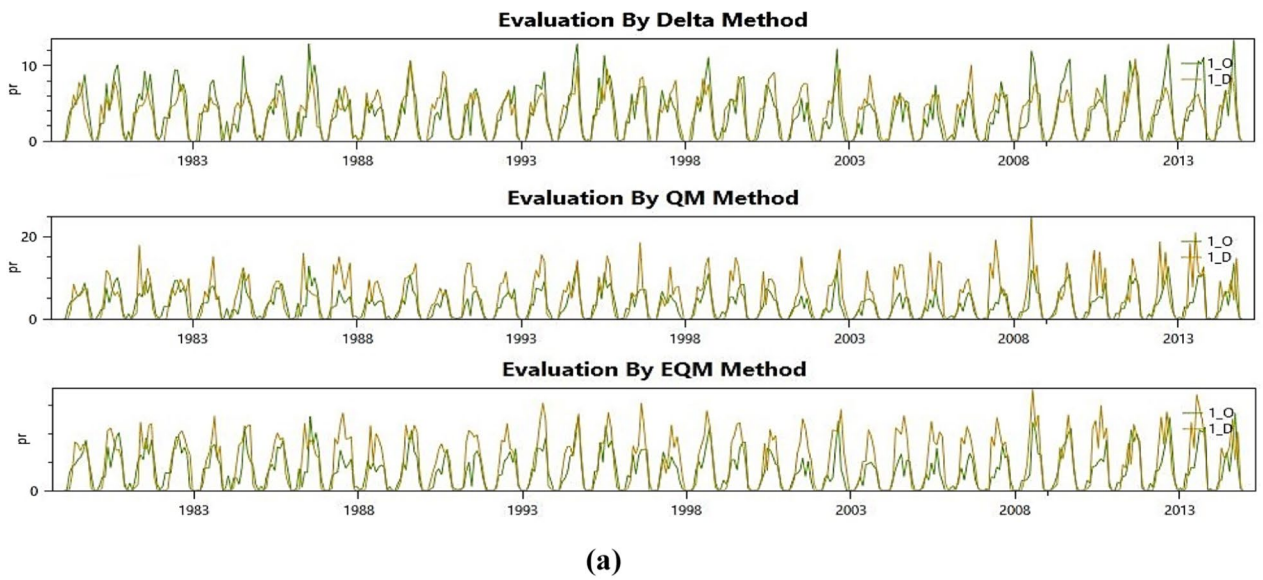


Fig. 3 Plot of bias-corrected GCMs and observed climate data using delta change, quantile mapping and empirical quantile mapping method in the Lake Chad basin. (a) Variation of GCM precipitation relative to observed CPC data. (b) Variation of GCM maximum temperature relative to observed PGF data. (c) Variation of GCM minimum temperature relative to observed PGF data

importance score is less than that of the shadow attributes at any grid before aggregation and ranking for consistency.

4.3 Evaluation based on symmetric uncertainty.

The result of symmetric uncertainty for all the climate variables were aggregated across the 54 grid points considered, which was a fair representation of the basin as shown in Table 2. The rankings based on the coefficients indicated that GFDL-ESM4, MIROC6, INM-CM4.8, ACCESS ESM1-5, and MRI-ESM2.0 exhibit a better skill for precipitation with coefficient in the range of 0.94–0.81, while MPI-ESM1.2-LR, INM-CM4.8, EC-Earth3-Veg, and EC-Earth3 are relatively better in simulating maximum temperature with coefficient in the range of 0.98–0.92 and finally, MPI-ESM1.2-LR, MIROC6, NorESM2-MM and CanESM5 are relatively better in simulating minimum temperature with a recorded coefficients in the range of 0.84–0.85. However, all

Table 2 Summary of GCM SU coefficients of precipitation, maximum and minimum temperature relative to observation data

Symmetric uncertainty					
Precipitation		Maximum temperature		Minimum temperature	
Model	SC	Model	SC	Model	SC
M8	0.94	M13	0.98	M13	0.85
M12	0.88	M9	0.94	M12	0.85
M9	0.81	M6	0.93	M16	0.84
M2	0.81	M5	0.92	M4	0.84
M14	0.81	M11	0.90	M6	0.83
M16	0.74	M16	0.90	M8	0.82
M13	0.72	M2	0.90	M5	0.81
M4	0.72	M12	0.89	M11	0.80
M3	0.71	M8	0.85	M9	0.80
M15	0.71	M4	0.83	M2	0.79
M11	0.65	M10	0.79	M15	0.77
M1	0.65	M14	0.78	M1	0.76
M10	0.65	M1	0.78	M14	0.75
M6	0.65	M15	0.76	M10	0.74
M5	0.62	M7	0.72	M7	0.73
M7	0.34	M3	0.70	M3	0.72

Note: the bold entries reflect models that best represents the climate features of the basin

the CMIP6 GCMs exhibit an above average skill except for F-Goals-g3 which was quite poor for precipitation. Varied level of GCM performance was noticed which further complicates evaluation and has been acknowledged by (McMahon et al., 2015), that GCMs have strength and weaknesses in simulating different climate variables. Simulation result of CMIP6 GCM by symmetric uncertainty exhibits some improvement in contrast to earlier phase as in (K Ahmed et al., 2019a), which may be attributed to differences in time-scale and timesteps of the chosen GCMs, improvement in model parameterizations and development and quality of observation data. This is corroborated in studies by (Ayugi et al., 2021; Grose et al., 2020; Wang et al., 2021).

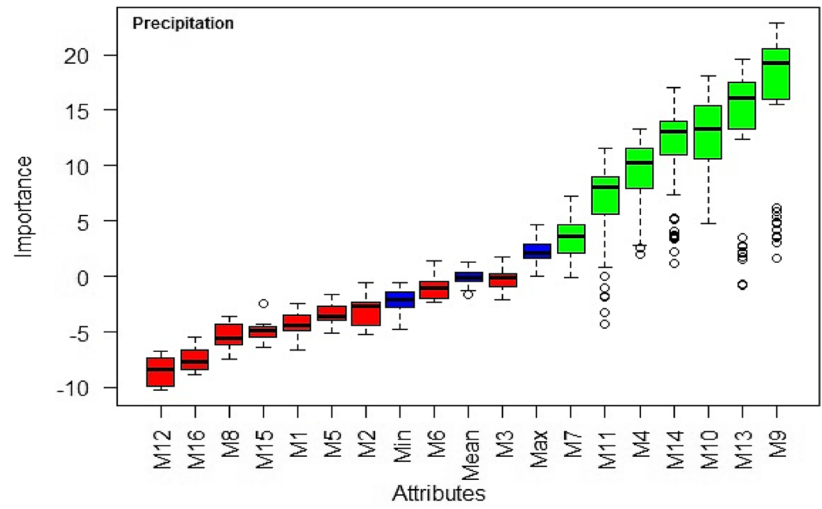
4.3.1 Evaluation based on Boruta random forest algorithm

The performance of the GCM models was assessed based on its ability to iteratively identify the importance of the original attributes (CMIP6 GCMs) with their randomised sets (shadow attributes) to truly replicate the observation data. The simulation and importance measure were generated, and the variables were ranked as shown in Fig. 4a–c for grid 1. The ranking indicates that model M9, M13, M10, M14, M4, M11 and M7 are quite important in simulating GCM daily precipitation with variable importance score in the range of 4–22, while others whose important score are below that of the maximum importance score of the shadow attributes are considered poor in simulating the observed (CPC) daily precipitation. However, all GCMs exhibit a good skill in simulating maximum and minimum temperature with a significant difference in variable importance score in the range of 18.0–54.2 and 19.1–44.6 respectively. The procedure was repeated across all grid points and attributes were filtered or rejected with a performance score below the shadow attributes. The variable importance score was aggregated across all points and were ranked based on descending order of mean importance score for all the variables as shown in Table 3.

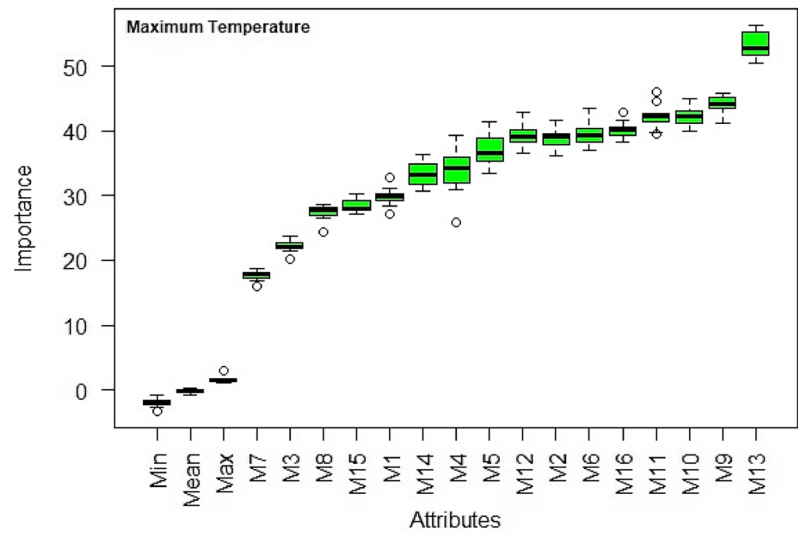
The evaluation of the climate variables by the novel Boruta random forest revealed a consistent performance of the GCMs across the grid points which makes the technique quite efficient in the ranking process and necessary for holistic assessment where the minimal optimal set of GCMs might be more useful rather than the application of the entire set of available models which may be computationally intensive, require more resources and time and decrease model proficiency.

The rankings based on Boruta algorithm indicated the difficulty of a single GCM to reliably simulate the daily observed precipitation across all grid points satisfactorily, although some GCM have a relatively better performance and are quite consistent across the grid points for maximum and minimum temperature. Figure 5a–c showed the spatial spread of the GCM performance and the aggregated importance score in Table 3, indicated that INM-CM4.8,

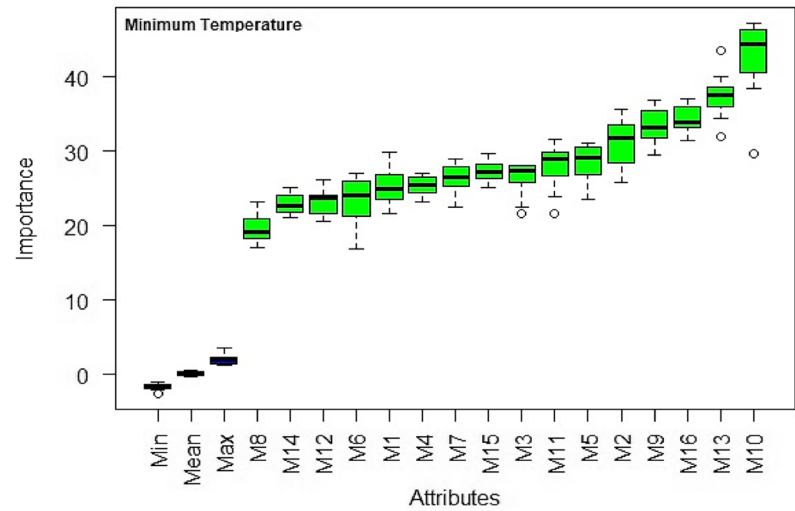
Fig. 4 Box plot of variable importance score of GCMs relative to observed climate data using BRF. **(a)** Relative importance of GCM precipitation to observed CPC data. **(b)** Relative importance of GCM maximum temperature to observed PGF data. **(c)** Relative importance of GCM minimum temperature to observed PGF data



(a)



(b)



(c)

Table 3 Summary of GCM importance score of precipitation, maximum and minimum temperature relative to observation data

Boruta algorithms					
Precipitation		Maximum temperature		Minimum temperature	
Model	IS	Model	IS	Model	IS
M9	4.68	M9	41.07	M10	32.03
M14	3.20	M13	39.44	M9	28.96
M10	2.88	M10	36.81	M13	28.86
M7	2.82	M12	34.10	M14	27.32
M4	2.68	M14	33.43	M16	27.18
M8	2.53	M2	32.25	M5	26.57
M13	2.12	M11	32.25	M7	26.50
M6	1.81	M7	31.36	M15	26.24
M12	1.33	M6	30.93	M6	26.17
M5	1.29	M16	30.87	M2	25.82
M1	1.20	M1	30.79	M1	25.52
M2	0.96	M3	29.89	M4	25.00
M11	0.75	M8	29.66	M11	24.92
M3	0.72	M5	29.62	M12	24.79
M16	0.31	M4	29.59	M8	24.53
M15	0.10	M15	27.56	M3	24.18

Note: the bold entries reflect models that best represents the climate features of the basin

MRI-ESM2.0, INM-CM5.0 and F-Goals-g3 exhibit a relatively better skill in replicating observed precipitation, while INM-CM4.8, MPI-ESM1.2-LR, INM-CM5.0 and MIROC6 for maximum temperature and INM-CM5.0, INM-CM4.8, MPI-ESM1.2-LR and MRI-ESM2.0 for minimum temperature respectively.

GCM models evaluated across the grid points indicated that a few are quite consistent in replicating the observation of the three climate variables; for example, INM-CM4.8 and INM-CM5.0 have shown to exhibit a better skill in simulating the observed daily precipitation at 24.1% and 18.52% of the grid points, while INM-CM4.8 and MPI-ESM1.2-LR at 35.19% and 16.67% for daily maximum temperature and finally INM-CM5.0 and MPI-ESM1.2-LR at 53.70% and 25.93% for daily minimum temperature respectively.

4.4 Identification and evaluation of multi-model ensemble mean of GCMs

The GCMs evaluated by the approaches considered were a precursor in understanding their skills necessary to match observations. However previous studies, for example Kim et al., (2016) have shown that uncertainties in climate projection can be reduced by identifying and adopting GCMs with better performance for impact assessment

studies. Earlier literature such as Weigel et al., (2010) and Miao et al., (2012) recommended the use of a collection of GCMs ensemble mean to optimize reliability in prediction and minimize uncertainty in climate variable assessment. In this study, four best GCM were selected after re-aggregation (Table 4), due in part to the significant difference observed in the aggregate importance score value between model M14 and M12 in Boruta random forest evaluation to form the multi-model ensemble mean herein referred to as SU and BRF and a combination of the 16 GCM model referred as AME and were further analysed and validated for spatial pattern of precipitation and temperature and return period of drought for the study period and their implication for hydrologic modelling.

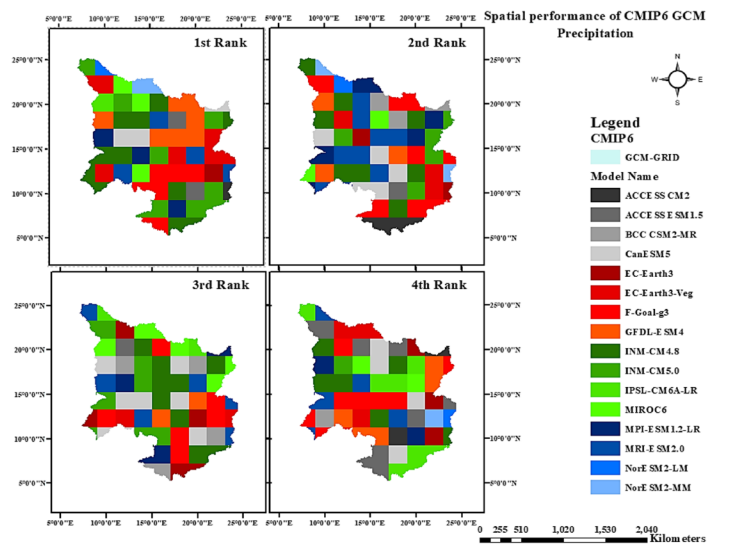
The result of the overall ranking indicated that MPI-ESM1.2-LR, MIROC6, INM-CM4.8 and NorESM2-MM are quite suitable from symmetric uncertainty approach, while INM-CM4.8, INM-CM5.0, MPI-ESM1.2-LR and MRI-ESM2.0 for Boruta random forest approach and are limited to four GCMs to others, due in part to the significant difference in their skills from the rating metrics scores in Table 4.

4.4.1 Spatial and temporal pattern of precipitation and temperature of GCM ensemble mean

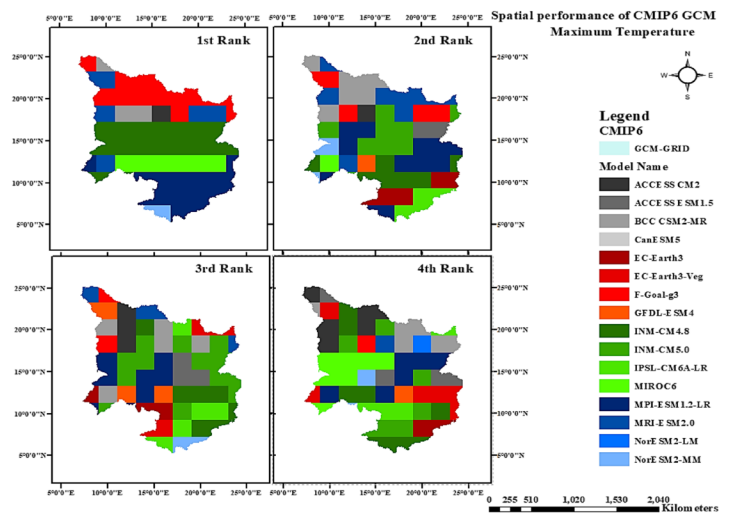
The spatial correlation and pattern of the GCM ensemble mean annual precipitation and temperature was used to validate and measure uncertainty range of the three different approaches relative to the observation as in Fig. 6a–b and Fig. 7a–b for the period 1979–2012. The result of the spatial correlation between the mean annual precipitation and temperature indicated that BRF is consistent with the observation having a correlation value in the range of 0.641–0.9995 (0.9991) and 0.4423–0.8345 (0.7136) respectively. Evaluation from SU and AME is quite satisfactory; however, there are mismatches or poor correlation observed, especially in the Sahelo-Sudanian zone with a recorded spatial correlation as low as 0.15 and 0.19 for SU and 0.01 and 0.36 for AME relative to the observations for annual mean precipitation and temperature respectively.

The results obtained in the temporal assessment of the evaluated approaches have shown that the correlation between the ensemble mean temperature are similar and quite skilful with $R^2 = 0.984$ and a mean bias of 0.49 °C, 0.49 °C and 0.50 °C for BRF, SU and AME respectively, however, the BRF approach indicated a better temporal correlation of 0.95 and an annual mean bias of 0.638 mm/year as compared to SU and AME with spatial correlation of 0.82 and 0.88 and annual mean bias precipitation of 68.19 mm/year and 10.57 mm/year respectively. The biases are found to be significant and visible in the south-western part of the basin as seen in the grid-based analysis of the annual

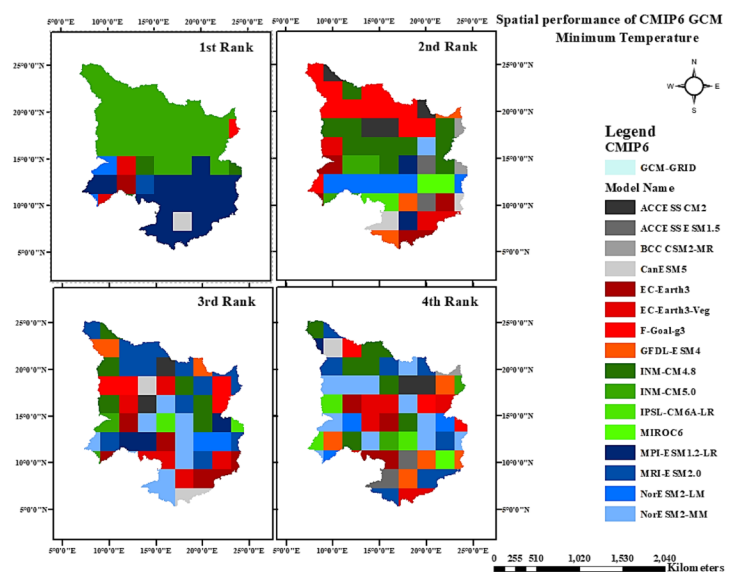
Fig. 5 Spatial spread of GCMs performance relative to observed climate data using BRF. **(a)** Ranking of spatial spread GCM precipitation relative to observed CPC data. **(b)** Ranking of spatial spread GCM maximum temperature relative to observed PGF data. **(c)** Ranking of spatial spread GCM minimum temperature relative to observed PGF data



(a)



(b)



(c)

Table 4 Summary of GCM overall ranking based on aggregated SU and IS of climate variables relative to observation data

Symmetric uncertainty		Boruta algorithms	
Model	SC	Model	AIS
M13	0.89	M9	60.22
M12	0.87	M10	45.13
M9	0.85	M13	29.54
M16	0.83	M14	15.11
M6	0.80	M12	10.43
M8	0.80	M16	8.54
M4	0.80	M7	8.49
M5	0.78	M2	8.14
M11	0.78	M5	6.75
M2	0.78	M6	6.52
M14	0.78	M11	6.49
M15	0.75	M1	5.18
M1	0.73	M15	5.15
M10	0.73	M4	4.52
M3	0.71	M8	4.36
M7	0.60	M3	3.97

Note: the bold entries reflect models that best represents the climate features of the basin

precipitation in Fig. 8 shown below, where significant deviations are noticed between grid point 1 to 10 for SU and 1 to 26 for AME, while BRF showed almost a perfect match relative to the observation across the grid points.

4.4.2 Assessment of climate extremes events of GCM ensemble mean

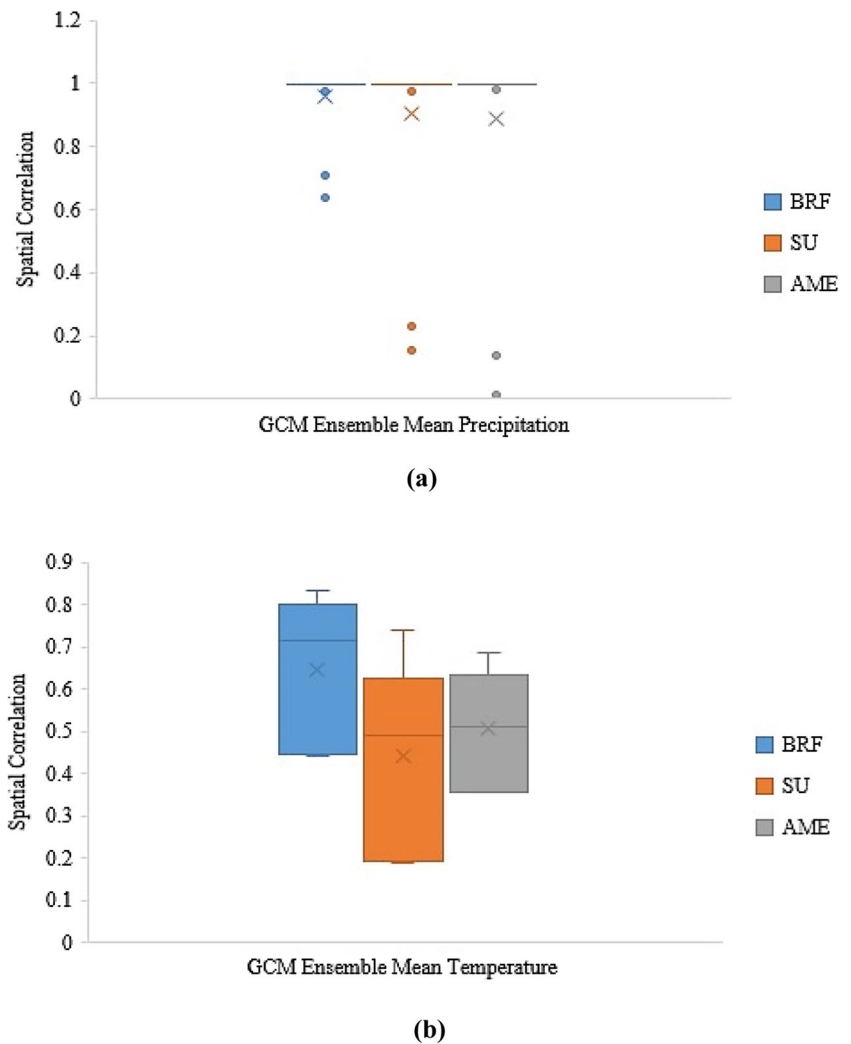
The temporal analysis of climate extreme indices was assessed at 12-month time step by the BRF, SU and AME and compared to the observation at the four climatic zone of the basin to understand the relative skills in predicting the pattern and frequency of extreme event and shift in trend within the study period in the four climatic zones. The result indicates an inherent shift (Fig. 9a), relatively in wet climate (1980–1998) to a transition from moderate to extreme droughts (1999–2012) in the Saharan zone with the frequency of SPEI values of the BRF, SU and AME approach consistent with the observation at 53.3%, 45.5% and 48.9% of the time respectively. Results of the SPEI values of the approaches relative to the observation (Figure 9b–d), in the Sahelo-Saharan zone is 48.99%, 53.28% and 47.22%, Sahelo-Sudanian zone is 45.2%, 45.96%, and 43.43% and Sudano-Guinean zones is 48.23%, 50.25% and 39.14% respectively. However, the statistical trends based on multi-year SPEI indices

for the period 1980–2012 in the four climatic zones of the basin indicated that the BRF approach captured the extreme event direction quite accurately relative to the observation as seen in the z-statistic values within the same trend envelope Table 5, although all the approaches showed a satisfactory result in predicting the trend direction but there are under-estimation in the magnitude of the extreme event by the SU (− 0.0013) and AME (− 0.0033) approach in the Saharan zone which indicate an insignificant shift from wet to drought events (drying trend) as against the trend magnitude exhibited by the observation (− 0.0057) which is consistent with the BRF (− 0.0058) approach. The approaches showed difficulty in predicting magnitude of the trend in the Sahelo-Sudanian zone which indicated an over-estimation relative to the observation with statistically significant wetting trend and the magnitude in the order of 0.0031, 0.0038 and 0.0057 for BRF, SU and AME respectively, as against observation (0.001). However, the deviation is more pronounce in the AME approach and it is observed that there is a frequent and consistent shift in trend from wetting to drying period across the climatic zone and is consistently captured by the BRF approach relative to the observation. Bold values in Table 5 indicate consistent agreement between simulated GCM ensemble mean and observation that better represent the basin climatic features and the outcome will be suited to limit the magnitude of uncertainties and accurate hydroclimatic hazard representation in impact studies.

A trade off was created in the evaluation process to justify the selected model performance using the two techniques irrespective of climate variable of interest and a varied level of performance was noticed from one grid point to another. The ensemble mean approach was quite essential because it led to reduced biases and their combinations emphasizing on few models with good performance are required and this is particularly important in watersheds with sparse observed climate data and high climate variability.

The BRF approach has shown to be promising in the evaluation with a recorded lower bias in the temperature and precipitation and a more accurate representation of the magnitude, pattern, and trends of extreme event. Overestimation or considerable bias was observed by the SU and AME approaches in the southwestern part of the basin. The associated uncertainties can be evaluated further by considering the sensitivities of ensemble with alternative metrics or combination of approaches. However, the BRF approach has shown to be quite robust in evaluating the integrity of the regionalisation of GCMs over different timescale that exhibit good performance during the baseline period and their combination may

Fig. 6 Comparison of spatial correlation of GCM ensemble mean performance relative to observed climate data. **(a)** Annual precipitation. **(b)** Annual temperature



likely be valid to represent the future period under climate change scenarios with certainty. The weighing scheme developed in this study is an exploratory framework that can be tested in various watersheds of interest for better water policy planning.

5 Conclusion

General circulation models are important and provides a pathway for simulation and assessment of the perceived impact of climate change on local, regional, and global hydrology. However, the choice of GCM input data, interpolation and downscaling method, timescale and timesteps are essential and critical for effective and accurate representation of the past, present, and future basin hydrologic

process for fair and equitable river basin management and policy planning for sustainability.

Earlier studies suggest that most evaluations ignored the inter-dependencies among models of a known variable and could create over-fitting problems. This study is based on an ensemble of 16 CMIP6 GCMs at daily timestep evaluating the efficacy and robustness of the state of the art Boruta random forest algorithm technique has shown this to be a viable tool for selection of relevant models to reduce redundancy, complexity and over-fitting problems associated with climate models to ensure sufficient overlap of chosen models ensemble mean with observations, seeking to limit the drawbacks encountered from existing techniques such as but not limited to inability to analyse complicated inputs, stochastic aspects, and climatic and hydrologic properties that are intricately inter-related, reduce the transfer of uncertainties into hydrologic

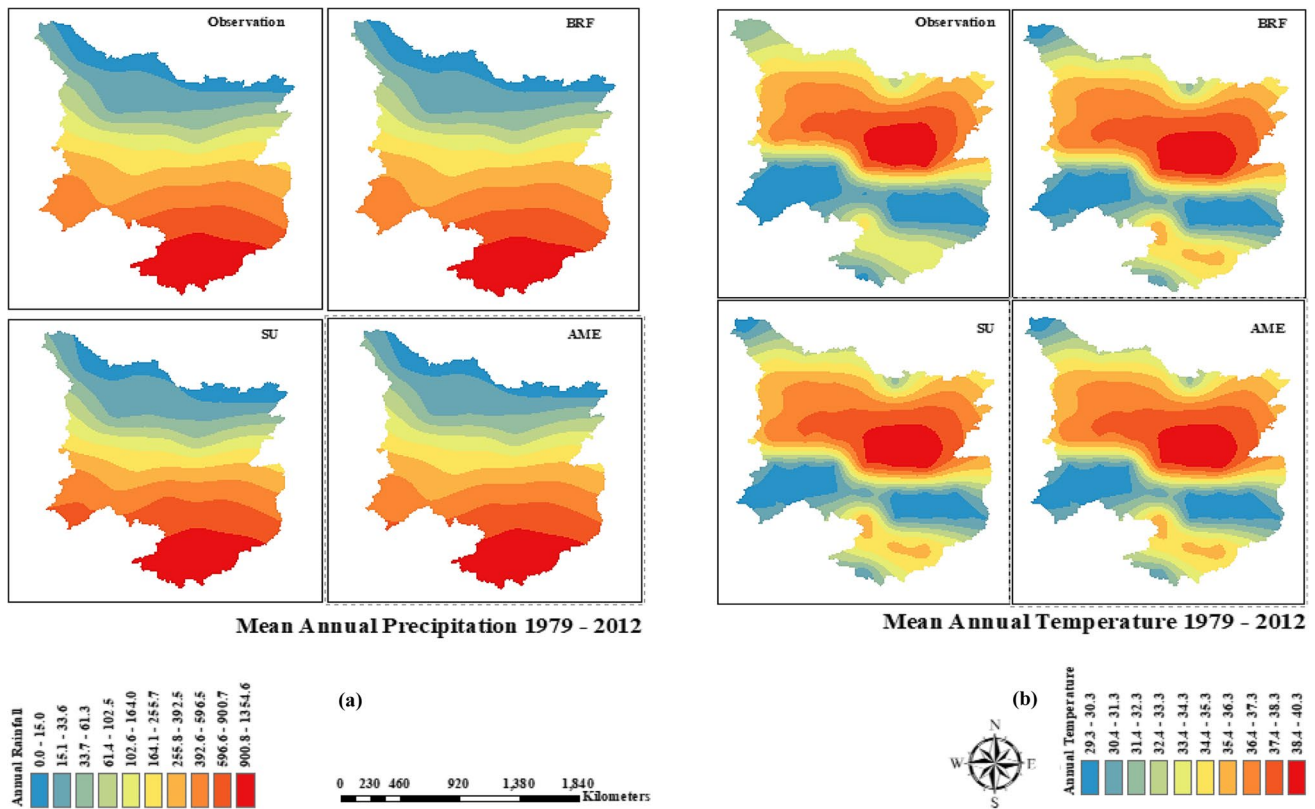


Fig. 7 Comparison of the spatial pattern of GCM ensemble performance relative to observed climate data. (a) Annual precipitation. (b) Annual temperature

processes and address critical temporal and spatial behaviour of the climate variables as a precursor for reliable and accurate predictions.

Highlights from the study revealed that there are inherent weaknesses associated temporal and spatial downscaling techniques and multiple techniques should be tried and examined to limit uncertainty range and inadequacies of GCMs, because exploring different multi-site

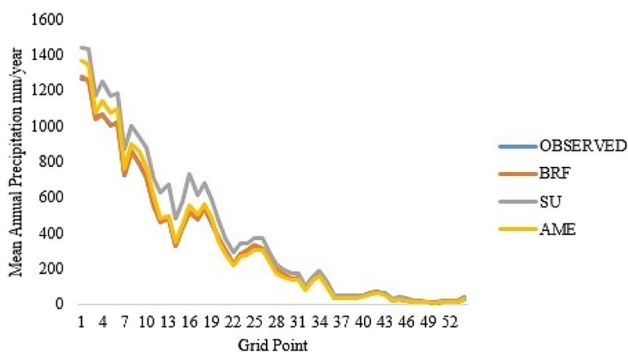
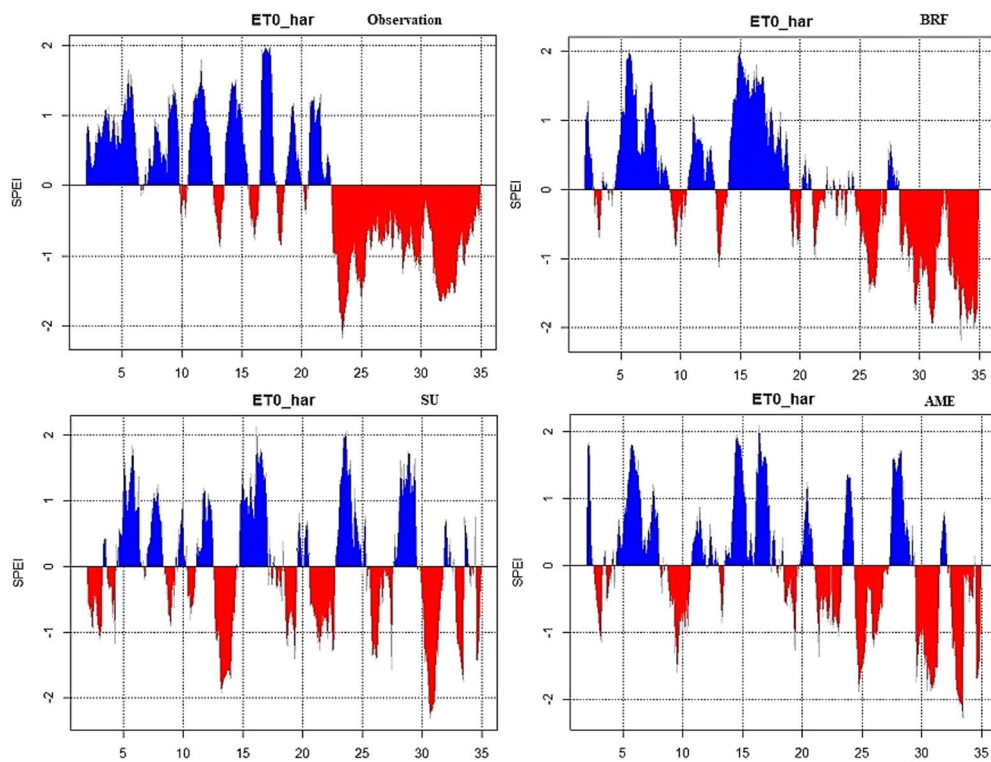


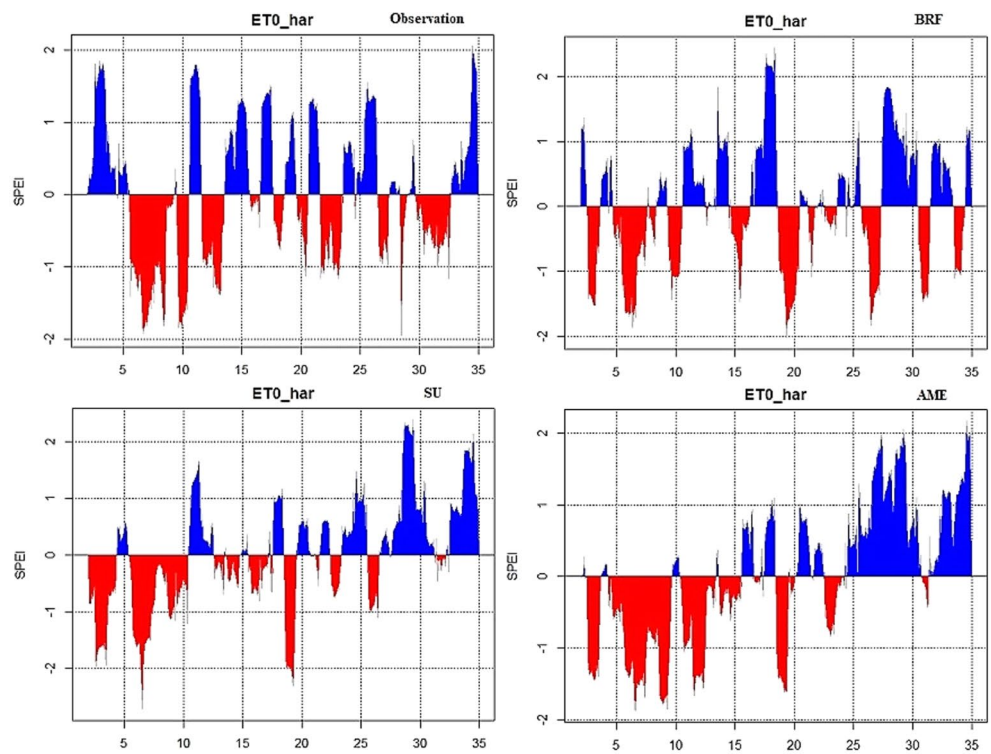
Fig. 8 Variation of the temporal pattern of mean annual precipitation across grid points for 1979-2012

downscaling techniques is very important in increasing the effectiveness of GCMs performance, combination of appropriate GCMs can enhance spatial and temporal variability to accurately reproduce observed multiple statistical properties of climate variables for improved output and reduced uncertainty in hydrologic modelling at regional and local basin scale. The selected models from Boruta random forest technique adequately have the capability in reducing biases in precipitation compared to the other approaches, although similar performances were observed in terms temperature and can capture the trends, patterns, and magnitude of extreme events within the accepted confidence limit.

The findings associated with this study are generally not meant to be a process to identify viable GCM dataset suitable for hydroclimatic study, but also to present a simple and efficient methodology to examine the limitations associated with the selected GCM ensemble for impact study. Therefore, the methodology proposed is not unique and therefore be explored to other basins for reliable representation of catchment climatology, representing a key step forward in GCM ensemble impact research.

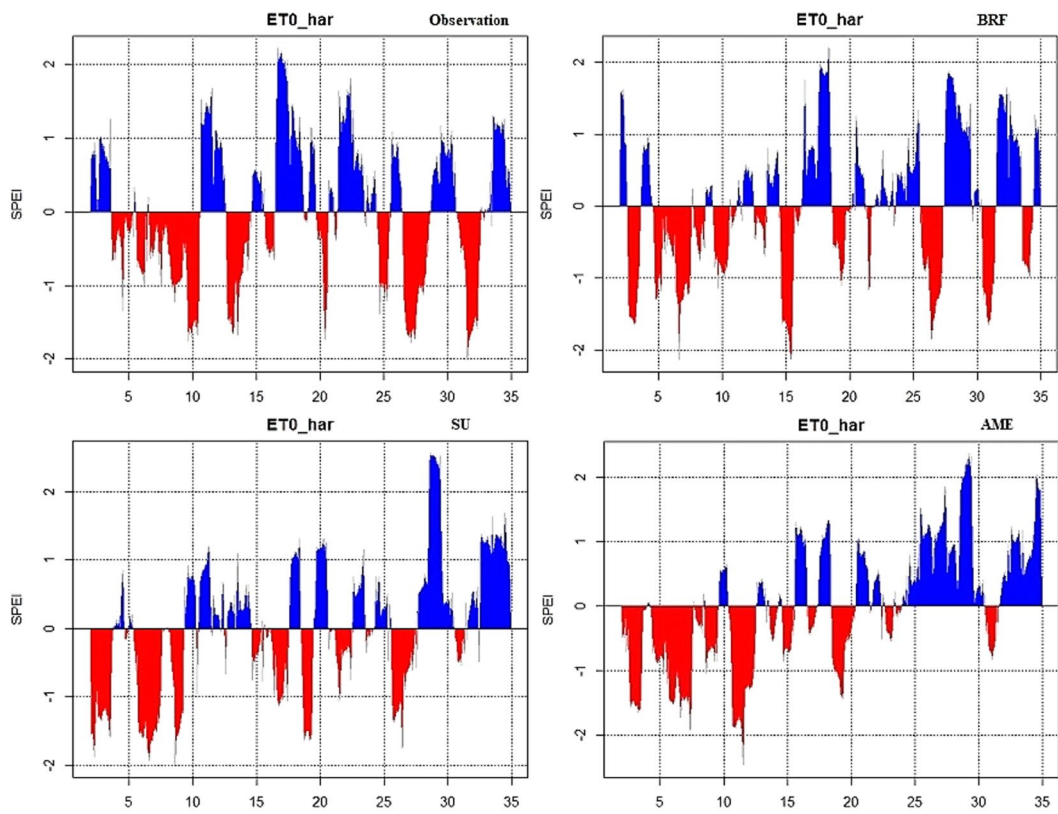


(a)

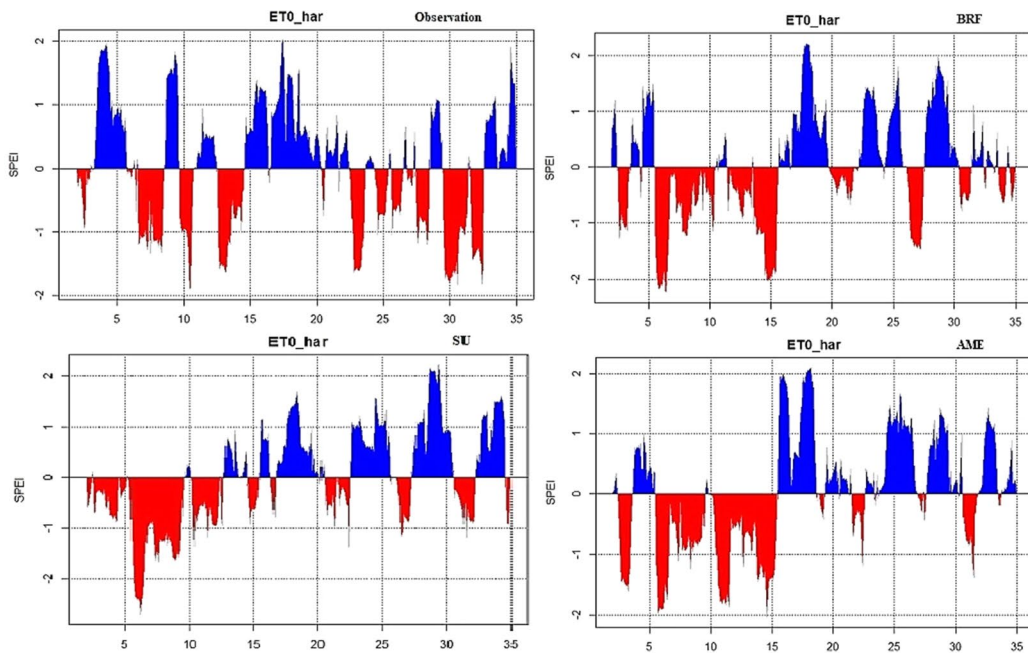


(b)

Fig. 9 Temporal variations climate extreme event of GCMs ensemble mean performance of relative to observation using BRF, SU and AME approach in the Lake Chad basin. (a) Saharan zone. (b) Sahelo-Saharan zone. (c) Sahelo-Sudanian zone. (d) Sudano-Guinean zone



(c)



(d)

Fig. 9 (continued)

Table 5 Mann Kendall Z-statistic of linear trend of extreme event for the period 1980–2012

Climatic zones	SPEI trend			
	Observation	BRF	SU	AME
Saharan	−2.44*	−2.25*	−1.04	−1.33
Sahelo-Saharan	1.39	1.04	3.25*	5.78*
Sahelo-Sudano	0.38	2.27*	3.07*	4.94*
Sudano-Guinean	0.69	0.74	2.28*	2.28*

*indicates statistically significant trend at 0.05 level of significance and bold values a match of similar trend envelop

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00704-023-04466-5>.

Acknowledgements This research work was supported by the Petroleum Technology Development Fund (Nigeria) and the Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow. The authors wish to extend their gratitude to all agencies whose data sources were used in this study for making their resources available and free for access.

Author contribution Conceptualization of research was done by, I.M.L., D.B., and C.J.; Data curation for the research was performed by, I.M.L.; Formal analysis of work by, I.M.L. I.H. and A.H.J.; Funding for APC was sought for by, D.B. and C.J.W.; Investigation, I.M.L., D.B., I.H. and S.R.M.K.; Methodology, I.M.L., and D.B.; Resources for computation, I.M.L., D.B., C.J.W., A.H.J., and S.R.M.K.; Software, I.M.L.; Supervision, D.B. and C.J.W.; Validation, D.B., C.J.W., A.H.J., I.H. and S.R.M.K.; Visualization, D.B., C.J.W., A.H.J., I.H., and S.R.M.K.; Writing—original draft, I.M.L.; Writing, Review & Editing, I.M.L., D.B., C.J.W., A.H.J., I.H., and S.R.M.K. All authors have read and agreed to the published version of the manuscript.

Funding This research was undertaken during the PhD studentship funded by the Petroleum Technology Development Fund, Nigeria, award number [PTDF/ED/OSS/PHD/IML/1537/19], and the University of Strathclyde.

Data availability All datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

Code availability Custom software source codes used in this study are available upon request from the corresponding author.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbasian M, Moghim S, Abrishamchi A (2019) Performance of the general circulation models in simulating temperature and precipitation over Iran. *Theor Appl Climatol* 135:1465–1483. <https://doi.org/10.1007/s00704-018-2456-y>
- Ahmadalipour A, Rana A, Moradkhani H, Sharma A (2017) Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis. *Theor Appl Climatol* 128:71–87. <https://doi.org/10.1007/s00704-015-1695-4>
- Ahmed K, Sachindra DA, Shahid S, Demirel MC, Chung E-S (2019a) Selection of multi-model ensemble of general circulation models for the simulation of precipitation and maximum and minimum temperature based on spatial assessment metrics. *Hydrol Earth Sys Sci* 23:4803–4824. <https://doi.org/10.5194/hess-23-4803-2019>
- Ahmed K, Shahid S, Sachindra DA, Nawaz N, Chung E (2019b) Fidelity assessment of general circulation model simulated precipitation and temperature over Pakistan using a feature selection method. *J Hydrol* 573:281–298. <https://doi.org/10.1016/j.jhydrol.2019.03.092>
- Ahmed AAM, Deo RC, Ghahramani A, Raj N, Feng Q, Yin Z, Yang L (2021) LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios. *Stoch Environ Res Risk Assess* 3:1–31. <https://doi.org/10.1007/s00477-021-01969-3>
- Akhter J, Das L, Deb A (2017) CMIP5 ensemble-based spatial rainfall projection over homogeneous zones of India. *Climate Dyn* 49:1885–1916. <https://doi.org/10.1007/s00382-016-3409-8>
- Alamgir M, Ahmed K, Homsi R, Dewan A, Wang J, Shahid S (2019) Downscaling and projection of spatiotemporal changes in temperature of Bangladesh. *Earth Syst Environ* 3:381–398. <https://doi.org/10.1007/s41748-019-00121-0>
- Almazroui M, Saeed S, Saeed F, Islam MN, Ismail M (2020) Projections of Precipitation and temperature over the South Asian Countries in CMIP6. *Earth Syst Environ* 4:297–320. <https://doi.org/10.1007/s41748-020-00157-7>
- Ayugi B, Zhihong J, Zhu H, Ngoma H, Babaousmail H, Rizwan K, Dike V (2021) Comparison of CMIP6 and CMIP5 models in simulating mean and extreme precipitation over East Africa. *Int J Climatol* 41:6474–6496. <https://doi.org/10.1002/joc.7207>
- Biemans H, Speelman LH, Ludwig F, Moors EJ, Wiltshire AJ, Kumar P, Gerten D, Kabat P (2013) Future water resources for food production in five South Asian river basins and potential for adaptation - a modeling study. *Sci Total Environ* 468–469:S117–S131. <https://doi.org/10.1016/j.scitotenv.2013.05.092>
- Breiman L (2001) Random forests. *Machine learning* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Buma WG, Lee S, Seo JY (2016) Hydrological evaluation of Lake Chad basin using space borne and hydrological model observations. *Water (Switzerland)* 8:1–15. <https://doi.org/10.3390/w8050205>
- Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of GCM precipitation by quantile mapping: how well do methods preserve

- changes in quantiles and extremes? *J Climate* 28:6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Chandler RE (2013) Exploiting strength, discounting weakness: combining information from multiple climate simulators. *Phil Trans Royal Soc A* 371:1–19. <https://doi.org/10.1098/rsta.2012.0388>
- Christ M, Kempa-Liehr AW, Feindt M (2016) Distributed and parallel time series feature extraction for industrial big data applications. Neurocomputing, Karlsruhe, Germany
- Chu JT, Xia J, Xu CY, Singh VP (2010) Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theor Appl Climatol* 99:149–161. <https://doi.org/10.1007/s00704-009-0129-6>
- Coe MT, Foley JA (2001) Human and natural impacts on the water resources of the Lake Chad basin. *J Geophys Res Atmos* 106:3349–3356. <https://doi.org/10.1029/2000JD900587>
- Coz LM, Delclaux F, Genthon P, Favreau G (2009) Assessment of Digital Elevation Model (DEM) aggregation methods for hydrological modeling: Lake Chad basin, Africa. *Comp Geosci* 35:1661–1670. <https://doi.org/10.1016/j.cageo.2008.07.009>
- Déqué M (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Glob Planet Change* 57:16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>
- Diffenbaugh NS, Giorgi F (2012) Climate change hotspots in the CMIP5 global climate model ensemble. *Climatic Change* 114:813–822. <https://doi.org/10.1007/s10584-012-0570-x>
- Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev* 9:1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Fischer R, Nowicki S, Kelley M, Schmidt GA (2014) A system of conservative regridding for ice-atmosphere coupling in a General Circulation Model (GCM). *Geosci Model Dev* 7:883–907. <https://doi.org/10.5194/gmd-7-883-2014>
- Fu G, Liu Z, Charles SP, Xu Z, Yao Z (2013) A score-based method for assessing the performance of GCMs: a case study of southeastern Australia. *J Geophys Res Atmos* 118:4154–4167. <https://doi.org/10.1002/jgrd.50269>
- Gao H, Bohn TJ, Podest E, McDonald KC, Lettenmaier DP (2011) On the causes of the shrinking of Lake Chad. *Environ Res Lett* 6:1–7. <https://doi.org/10.1088/1748-9326/6/3/034021>
- Genuer R, Poggi JM, Tuleau-Malot C (2010) Variable selection using random forests. *Pattern Recog Lett* 31:2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Ghimire S, Deo RC, Raj N, Mi J (2019a) Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction. *Energies* 12:2407. <https://doi.org/10.3390/en12122407>
- Ghimire U, Srinivasan G, Agarwal A (2019b) Assessment of rainfall bias correction techniques for improved hydrological simulation. *Int J Climatol* 39:2386–2399. <https://doi.org/10.1002/joc.5959>
- Gidden MJ, Riahi K, Smith SJ, Fujimori S, Luderer G, Kriegler E, Van Vuuren DP, Van Den Berg M, Feng L, Klein D, Calvin K, Doelman JC, Frank S, Fricko O, Harmsen M, Hasegawa T, Havlik P, Hilaire J, Hoesly R et al (2019) Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geosci Model Dev* 12:1443–1475. <https://doi.org/10.5194/gmd-12-1443-2019>
- Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “Reliability Ensemble Averaging” (REA) Method. *J Climate* 15:1141–1158. [https://doi.org/10.1175/1520-0442\(2002\)016](https://doi.org/10.1175/1520-0442(2002)016)
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res Atmos* 113:1–20. <https://doi.org/10.1029/2007JD008972>
- Gong G, An X, Mahato NK, Sun S, Chen S, Wen Y (2019) Research on short-term load prediction based on Seq2seq model. *Energies* 12:1–18. <https://doi.org/10.3390/en12163199>
- Goyal MK, Burn DH, Ojha CSP (2012) Evaluation of machine learning tools as a statistical downscaling tool: temperatures projections for multi-stations for Thames River Basin, Canada. *Theor Appl Climatol* 108:519–534. <https://doi.org/10.1007/s00704-011-0546-1>
- Grose MR, Narsey S, Delage FP, Dowdy AJ, Bador M, Boschat G, Chung C, Kajtar JB, Rauniyar S, Freund MB, Lyu K, Rashid H, Zhang X, Wales S, Trenham C, Holbrook NJ, Cowan T, Alexander L, Arblaster JM, Power S (2020) Insights from CMIP6 for Australia’s future climate. *Earth’s Future* 8:1–24. <https://doi.org/10.1029/2019EF001469>
- Gu H, Yu Z, Wang J, Wang G, Yang T, Ju Q, Yang C, Xu F, Fan C (2015) Assessing CMIP5 general circulation model simulations of precipitation and temperature over China. *Int J Climatol* 35:2431–2440. <https://doi.org/10.1002/joc.4152>
- Gusain A, Ghosh S, Karmakar S (2020) Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmos Res* 232:104680. <https://doi.org/10.1016/j.atmosres.2019.104680>
- Hajnayeb A, Ghasemloonia A, Khadem SE, Moradi MH (2011) Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis. *Exp Syst Appl* 38:10205–10209. <https://doi.org/10.1016/j.eswa.2011.02.065>
- Hammami D, Lee TS, Ouarda TBMJ, Le J (2012) Predictor selection for downscaling GCM data with LASSO. *J Geophys Res Atmos* 117:D17116. <https://doi.org/10.1029/2012JD017864>
- Hassan I, Lawal IM, Mohammed A, Abubakar S (2016) Analysis of geostatistical and deterministic techniques in the spatial variation of groundwater depth in the North-western part of Bangladesh. *Am J Eng Res (AJER)* 5:29–34
- Hassan I, Kalin RM, White CJ, Aladejana JA (2020) Selection of CMIP5 GCM ensemble for the projection of spatio-temporal changes in precipitation and temperature over the Niger Delta, Nigeria. *Water (Switzerland)* 12:1–19. <https://doi.org/10.3390/w12020385>
- Henry BM (1945) Nonparametric tests against trend. *Econometrica* 13:245–259
- Huang W, Wang S, Chan JCL (2011) Discrepancies between global reanalyses and observations in the interdecadal variations of Southeast Asian cold surge. *Int J Climatol* 31:2272–2280. <https://doi.org/10.1002/joc.2234>
- Hur J, Ihm S, Park Y (2017) A variable impacts measurement in random forest for mobile cloud computing. *Wireless Comm Mobile Comp* 2017:1–13. <https://doi.org/10.1155/2017/6817627>
- IPCC (2013) Climate change 2013 the physical science basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA. <https://doi.org/10.1017/CBO9781107415324>
- Kannan SS, Ramaraj N (2010) A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm. *Knowl-Based Syst* 23:580–585. <https://doi.org/10.1016/j.knsys.2010.03.016>
- Kendall MG (1948) Rank Correlation Methods, 4th edn. Griffin, London
- Khan N, Pour SH, Shahid S, Ismail T, Ahmed K, Chung E, Nawaz N, Wang X (2019) Spatial distribution of secular trends in rainfall

- indices of Peninsular Malaysia in the presence of long-term persistence. *Meteorol Appl* 26:655–670. <https://doi.org/10.1002/met.1792>
- Kim J, Ivanov VY, Fatichi S (2016) Climate change and uncertainty assessment over a hydroclimatic transect of Michigan. *Stoch Environ Res Risk Assess* 30:923–944. <https://doi.org/10.1007/s00477-015-1097-2>
- Knutti R, Masson D, Gettelman A (2013) Climate model genealogy: generation CMIP5 and how we got there. *Geophys Res Lett* 40:1194–1199. <https://doi.org/10.1002/grl.50256>
- Kursa, M.B., 2016. Embedded all relevant feature selection with random ferns.. *ArXiv abs/1604.0*, 1–13 https://doi.org/10.1007/978-3-319-60438-1_30
- Kursa MB, Rudnicki WR (2010) Feature selection with the boruta package. *J Stat Softw* 36:1–13. <https://doi.org/10.18637/jss.v036.i11>
- Kursa MB, Jankowski A, Rudnicki WR (2010) Boruta - a system for feature selection. *Fundamenta Inform* 101:271–285. <https://doi.org/10.3233/FI-2010-288>
- Lawal IM, Bertram D, White CJ, Jagaba AH, Hassan I, Shuaibu A (2021) Multi-criteria performance evaluation of gridded precipitation and temperature products in data-sparse regions. *Atmosphere* 12:1–25. <https://doi.org/10.3390/atmos12121597>
- Leutner BF, Reineking B, Müller J, Bachmann M, Beierkuhnlein C, Dech S, Wegmann M (2012) Modelling forest α -diversity and floristic composition - on the added value of LiDAR plus hyperspectral remote sensing. *Remote Sens* 4:2818–2845. <https://doi.org/10.3390/rs4092818>
- Li W, Jiang Z, Xu J, Li L (2016) Extreme precipitation indices over China in CMIP5 models. Part II: Probabilistic projection. *J Climate* 29:8989–9004. <https://doi.org/10.1175/JCLI-D-16-0377.1>
- Lyu B, Zhang Y, Hu Y (2017) Improving PM2.5 air quality model forecasts in China using a bias-correction framework. *Atmosphere* 147:1–15. <https://doi.org/10.3390/atmos8080147>
- Mahmood R, Jia S, Zhu W (2019) Analysis of climate variability, trends, and prediction in the most active parts of the Lake Chad basin, Africa. *Sci Rep* 9:1–18. <https://doi.org/10.1038/s41598-019-42811-9>
- Maldonado S, Weber R (2009) A wrapper method for feature selection using Support Vector Machines. *Inform Sci* 179:2208–2217. <https://doi.org/10.1016/j.ins.2009.02.014>
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themel M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48:1–34. <https://doi.org/10.1029/2009RG0000314>
- Martinez-García FP, Contreras-De-villar A, Muñoz-Perez JJ (2021) Review of wind models at a local scale: advantages and disadvantages. *J Mar Sci Eng* 9:318. <https://doi.org/10.3390/jmse9030318>
- Maxino CC, McAvaney BJ, Pitman AJ, Perkins SE (2008) Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *Int J Climatol* 28:1097–1112. <https://doi.org/10.1002/joc.1612>
- McMahon TA, Peel MC, Karoly DJ (2015) Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation. *Hydrol Earth Syst Sci* 19:361–377. <https://doi.org/10.5194/hess-19-361-2015>
- McSweeney CF, Jones RG, Lee RW, Rowell DP (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dyn* 44:3237–3260. <https://doi.org/10.1007/s00382-014-2418-8>
- Miao C, Duan Q, Yang L, Borthwick AGL (2012) On the applicability of temperature and precipitation data from CMIP3 for China. *PLoS One* 7:1–10. <https://doi.org/10.1371/journal.pone.0044659>
- Min S, Hense A (2006) A Bayesian assessment of climate change using multimodel ensembles. Part I: Global mean surface temperature. *J Climate* 19:3237–3256. <https://doi.org/10.1175/JCLI3784.1>
- Mishra A, Liu SC (2014) Changes in precipitation pattern and risk of drought over India in the context of global warming. *J Geophys Res*: Atmos 119:7833–7841. <https://doi.org/10.1002/2014JD021471>
- Mohsenipour M, Shahid S, Chung E, Wang X (2018) Changing pattern of droughts during cropping seasons of Bangladesh. *Water Resour Manag* 32:1555–1568. <https://doi.org/10.1007/s11269-017-1890-4>
- Nashwan MS, Shahid S (2019) Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmos Res* 230:1–10. <https://doi.org/10.1016/j.atmosres.2019.104632>
- Navarro-Racines C, Tarapues J, Thornton P, Jarvis A, Ramirez-Villegas J (2020) High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Sci Data* 7:1–14. <https://doi.org/10.1038/s41597-019-0343-8>
- Ndehedehe CE, Awange JL, Agutu NO, Okwuashi O (2018) Changes in hydro-meteorological conditions over tropical West Africa (1980–2015) and links to global climate. *Glob Planet Change* 162:321–341. <https://doi.org/10.1016/j.gloplacha.2018.01.020>
- Niang I, Ruppel OC, Abdrabo MA, Essel A, Lennard C, Padgham J, Urquhart P (2014) Africa. In: Dube P, Leary N (eds) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp 1199–1265
- Nkiaka E, Nawaz NR, Lovett JC (2018a) Effect of single and multi-site calibration techniques on hydrological model performance, parameter estimation and predictive uncertainty: a case study in the Logone catchment, Lake Chad basin. *Stoch Environ Res Risk Assess* 32:1665–1682. <https://doi.org/10.1007/s00477-017-1466-0>
- Nkiaka E, Nawaz R, Lovett JC (2018b) Assessing the reliability and uncertainties of projected changes in precipitation and temperature in Coupled Model Intercomparison Project phase 5 models over the Lake Chad basin. *Int J Climatol* 38:5136–5152. <https://doi.org/10.1002/joc.5717>
- Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J Climate* 20:4356–4376. <https://doi.org/10.1175/JCLI4253.1>
- Pour SH, Shahid S, Chung E, Wang X (2018) Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmos Res* 213:149–162. <https://doi.org/10.1016/j.atmosres.2018.06.006>
- Prasad R, Deo RC, Li Y, Maraseni T (2017) Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmos Res* 197:42–63. <https://doi.org/10.1016/j.atmosres.2017.06.014>
- Prasad R, Deo RC, Li Y, Maraseni T (2019) Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena* 177:149–166. <https://doi.org/10.1016/j.catena.2019.02.012>

- Raju SK, Kumar ND (2016) Selection of global climate models for India using cluster analysis. *J Water Climate Change* 7:764–774. <https://doi.org/10.2166/wcc.2016.112>
- Raju KS, Sonali P, Kumar ND (2017) Ranking of CMIP5-based global climate models for India using compromise programming. *Theor Appl Climatol* 128:563–574. <https://doi.org/10.1007/s00704-015-1721-6>
- Reichler T, Kim J (2008) How well do coupled models simulate today's climate? *Bull Am Meteorol Soc* 89:303–311. <https://doi.org/10.1175/BAMS-89-3-303>
- Sa'adi Z, Shahid S, Pour SH, Ahmed K, Chung E, Yaseen ZM (2020) Multi-variable model output statistics downscaling for the projection of spatio-temporal changes in rainfall of Borneo Island. *J Hydro-Environ Res* 31:62–75. <https://doi.org/10.1016/j.jher.2020.05.002>
- Sachindra DA, Huang F, Barton A, Perera BJC (2014) Statistical downscaling of general circulation model outputs to precipitation—part 1: Calibration and validation. *Int J Climatol* 34:3264–3281. <https://doi.org/10.1002/joc.3914>
- Salman SA, Shahid S, Ismail T, Ahmed K, Wang X (2018) Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. *Atmos Res* 213:509–522. <https://doi.org/10.1016/j.atmosres.2018.07.008>
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's Tau. *J Am Stat Assoc* 63:1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Sheffield J, Goteti G, Wood EF (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J Climate* 19:3088–3111. <https://doi.org/10.1175/JCLI3790.1>
- Shekhar A, Shapiro CA (2019) What do meteorological indices tell us about a long-term tillage study? *Soil Tillage Res* 193:161–170. <https://doi.org/10.1016/j.still.2019.06.004>
- Shiru MS, Chung E (2021) Performance evaluation of CMIP6 global climate models for selecting models for climate projection over Nigeria. *Theor Appl Climatol* 146:599–615. <https://doi.org/10.1007/s00704-021-03746-2>
- Shukla J, DelSole T, Fennessy M, Kinter J, Paolino D (2006) Climate model fidelity and projections of climate change. *Geophys Res Lett* 33:3–6. <https://doi.org/10.1029/2005GL025579>
- Sian KTCLK, Wang J, Ayugi BO, Nooni IK, Ongoma V (2021) Multi-decadal variability and future changes in precipitation over Southern Africa. *Atmosphere* 12:1–25. <https://doi.org/10.3390/atmos12060742>
- Song Y, Li X, Bao Y, Song Z, Wei M, Shu Q, Yang X (2020) FIO-ESM v2.0 Outputs for the CMIP6 Global Monsoons Model Intercomparison Project Experiments. *Adv Atmos Sci* 37:1045–1056. <https://doi.org/10.1007/s00376-020-9288-2>
- Stoppiglia H, Dreyfus G, Dubois R, Oussar Y (2003) Ranking a random feature for variable and feature selection. *J Mach Learn Res* 3:1399–1414. <https://doi.org/10.1162/153244303322753733>
- Taylor JR (1997) Covariance and Correlation, in: *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, Sausalito, CA, p 217
- Wang X, Piao S, Ciais P, Friedlingstein P, Myneni RB, Cox P, Heimann M, Miller J, Peng S, Wang T, Yang H, Chen A (2014) A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature* 506:212–215. <https://doi.org/10.1038/nature12915>
- Wang Y, Shi L, Zanobetti A, Schwartz JD (2016) Estimating and projecting the effect of cold waves on mortality in 209 US cities. *Environ Int* 94:141–149. <https://doi.org/10.1016/j.envint.2016.05.008>
- Wang D, Liu J, Shao W, Mei C, Su X, Wang H (2021) Comparison of CMIP5 and CMIP6 multi-model ensemble for precipitation downscaling results and observational data: the case of Hanjiang River Basin. *Atmosphere* 12:1–20. <https://doi.org/10.3390/atmos12070867>
- Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O, Schewe J (2014) The inter-sectoral impact model intercomparison project (ISI-MIP): Project framework. *Proc Nat Acad Sci United States of America*. 111:3228–3232. <https://doi.org/10.1073/pnas.1312330110>
- Weigel AP, Knutti R, Liniger MA, Appenzeller C (2010) Risks of model weighting in multimodel climate projections. *J Climate* 23:4175–4191. <https://doi.org/10.1175/2010JCLI3594.1>
- Willems P, Vrac M (2011) Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *J Hydrol* 402:193–205. <https://doi.org/10.1016/j.jhydrol.2011.02.030>
- William H, Teukolsky SA, Vetterling WT, Flannery BP (1996) *Numerical Recipes in Fortran 90: the art of parallel scientific computing*. Cambridge University Press, London, UK
- Willmott CJ (1981) On the validation of models. *Phys Geogr* 2:184–194. <https://doi.org/10.1080/02723646.1981.10642213>
- Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res* 30:79–82. <https://doi.org/10.3354/cr030079>
- Witten IH, Hall MA, Frank E (2005) *Ensemble Learning*. In: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Amsterdam, pp 351–372
- Wójcik R, Pilarski M, Miętus M (2014) Statistical downscaling of probability density function of daily precipitation on the Polish coast. *Meteorol Hydrol Water Manag* 2:27–36. <https://doi.org/10.26491/mhwm/21590>
- Wright DB, Knutson TR, Smith JA (2015) Regional climate model projections of rainfall from U.S. landfalling tropical cyclones. *Climate Dy* 45:3365–3379. <https://doi.org/10.1007/s00382-015-2544-y>
- Wu Y, Zhang A (2004) Feature selection for classifying high-dimensional numerical data. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, Washington, DC, USA, pp 251–258. <https://doi.org/10.1109/CVPR.2004.1315171>
- Wu Z, Chen X, Lu G, Xiao H, He H, Zhang J (2017) Regional response of runoff in CMIP5 multi-model climate projections of Jiangsu Province, China. *Stoch Environ Res Risk Assess* 31:2627–2643. <https://doi.org/10.1007/s00477-016-1349-9>
- Xie P, Yatagai A, Chen M, Hayasaka T, Fukushima Y, Liu C, Yang S (2007) A gauge-based analysis of daily precipitation over East Asia. *J Hydrometeorol* 8:607–626. <https://doi.org/10.1175/JHM583.1>
- Xuan W, Ma C, Kang L, Gu H, Pan S, Xu YP (2017) Evaluating historical simulations of CMIP5 GCMs for key climatic variables in Zhejiang Province, China. *Theor Appl Climatol* 128:207–222. <https://doi.org/10.1007/s00704-015-1704-7>
- Yang X, Delsole T (2012) Systematic comparison of enso teleconnection patterns between models and observations. *J Climate* 25:425–446. <https://doi.org/10.1175/JCLI-D-11-00175.1>