Scientific Research Publishing

Atmospheric and Climate Sciences, 2025, 15(2), 495-531
https://www.scirp.org/journal/acs
ISSN Online: 2160-0422
ISSN Print: 2160-0414

Application of Satellite Rainfall Estimates in Quantitative Forecasting of Monthly Rainfall Using a Multi-Model Ensemble Approach, Kafue River Basin

Miyanda M. Syabwengo, Progress Nyanga, Moses Chisola

Department of Geography and Environmental Studies, School of Natural Sciences, University of Zambia, Lusaka, Zambia
Email: miyamala@gmail.com

How to cite this paper: Syabwengo, M.M., Nyanga, P. and Chisola, M. (2025) Application of Satellite Rainfall Estimates in Quantitative Forecasting of Monthly Rainfall Using a Multi-Model Ensemble Approach, Kafue River Basin. *Atmospheric and Climate Sciences*, **15**, 495-531.
https://doi.org/10.4236/acs.2025.152026

Received: October 28, 2024
Accepted: April 22, 2025
Published: April 25, 2025

Abstract

This study evaluates the reliability of North American Multi-Model Ensemble (NMME) models in forecasting monthly rainfall over the Kafue River Basin using a well-selected multi-model ensemble approach. Gridded monthly rainfall forecasts were derived from global NMME models and validated against satellite-based rainfall products (SRPs) over the basin. To establish a reliable gridded rainfall dataset, three SRPs—TAMSAT, CHIRPS, and ARC2—were assessed against observed station data. Historical data were divided into a calibration period (1983-2003) at the station level and a validation period (2004-2022) using gridded datasets. The NMME models—CMC2 CANSIPSv2, NASA-GEOSS2S, CANCM4i, GFDL-CM2p1, GFDL-CM2p5-FLOR-B01, GFDL-CM2p5, NCEP-CFSv2, and COLA-RSMAS-CCSM—were downscaled using the Canonical Correlation Analysis (CCA) algorithm and evaluated using Spearman's correlation coefficient, mean bias, and root mean square error (RMSE). The Anomaly Correlation Coefficient (ACC) was used to assess forecast reliability. Results show that CHIRPS outperformed TAMSAT and ARC2 in representing observed rainfall and was used to generate a gridded time-series dataset. NMME model performance improved when validated against gridded datasets rather than station-based point data. The ensemble forecasting approach demonstrated reliable monthly rainfall predictions for December, January, and March (2004-2022). However, caution is advised when using NMME models for October and February, as these months exhibited negative ACC values (−1) over much of the basin. The study highlights spatial and temporal variability in the reliability of individual NMME models, emphasizing the importance of understanding model strengths and limitations for effective climate adaptation and water resource management.

Keywords

Satellite Rainfall Estimates, Quantitative Forecasting, Monthly Rainfall, Multi-Model Ensemble, Kafue River Basin

1. Introduction

Weather forecasting is one of the most important scenarios of scientific computing, which offers the ability to predict future weather changes [1]. Most forecasting techniques collect data about the current state of the atmosphere at specific observation stations and use the understanding of atmospheric processes to determine how the atmosphere evolves in the future. In weather forecasting, one needs to know what happened in the past and what is happening now for initial conditions to be used in predicting the future. Weather forecasts are categorized as short, medium and long-range forecasts based on the validity time. Seasonal rainfall forecast, which includes monthly forecasts, falls within the long-range type of forecast.

Conversely, weather forecasting systems can be classified into two basic types: deterministic and probabilistic, based on the methodology used for forecasting. Deterministic methods provide accurate values of weather forecasts for a specific location, whereas probabilistic methods recommend probabilities of weather events [2]. Broadly, there are three kinds of methods one can use to make seasonal forecasts: empirical (or statistical), dynamical, and hybrid [3]. Empirical methods use statistical relationships between the predictor(s) and the variable(s) used to make the forecast. Statistical methods require low computing resources and are easy to implement operationally. According to [3], dynamical seasonal forecast methods, which are carried out using global climate models, are being used with increasing frequency, though they are very resource-intensive and require access to an extensive computing infrastructure both for predictions and analysis to generate initial conditions. Global Climate Models (GCMs) and Regional Climate Models (RCMs) fall under model-based NWP which are based on highly complex mathematical representations of atmospheric, oceanic, and continental processes [3]. Hybrid approaches make use of a blend of these two methods, taking advantage of both; they use a physics-based model output to represent the different processes in the climate system and statistical models to bias-correct and calibrate. According to the IPCC technical paper on data for adaptation at different temporal and spatial scales in [4], weather forecasts are particularly relevant in the case of extreme weather that requires the implementation of an evacuation plan or other safety measures and therefore provide the basis for issuing alerts.

[5] indicated that the use of climate data in Africa for research and applications has been restricted mainly due to limited availability and access to quality climate time series data. Precise areal precipitation measurements are always a challenge because of the large spatiotemporal variability and inherent errors associated with

the various measuring instrument and uncertainties associated with weather-related elements [6]. Mostly, rain gauge observations give relatively accurate point measurements but errors may be introduced when representing areal rainfall due to limited availability and distribution of gauge stations. Satellite observations can be used to derive rainfall estimates at a resolution of about 4 km * 4 km over much of the globe but are associated with non-negligible bias and random errors due to inadequate sampling, algorithm errors and the indirect nature of the physical relationship between precipitation and observations. These are the common challenges one has to face in dealing with rainfall data-related analysis. To overcome these challenges, traditional rain gauges with their advantage of providing direct rainfall measurements, often serve as a reference for the validation of radar and satellite-based precipitation products [6]. Merging precipitation observations from diverse sources is a powerful means for improving the overall measurement quality of area rainfall.

The reality of Climate Change as established by the Intergovernmental Panel on Climate Change (IPCC) [7] can be seen from the increased frequency and intensity of extreme events such as heatwaves, drought and floods occurring the world over Zambia inclusive. The Kafue River Basin, like any other part of the world, is subjected to the effects of Climate Change/Variability impacting the rain-fed economic sectors and therefore driving the need to adequately plan and adapt to the anticipated effects. The three-month averaged probabilistic approach in seasonal rainfall forecasting at the Zambia Meteorological Department (ZMD) is relevant, highly recognized and used by different sectors. However, monthly precipitation forecasts at a local scale are valuable and reliable monthly precipitation forecasts for several months ahead can apply to various sectors [8] as input into further sector-specific planning and decision-making. Today's planning and decision-making approaches, such as modeling in sector-specific applications (Agriculture, Hydrology, and Energy), are becoming more and more technically complex, demanding accurate numeric input of weather elements such as monthly rainfall forecast at specific points. ZMD has a sparse observation network with only eight stations lying within the Kafue River Basin being integrated into generating the seasonal rainfall forecast. The aforementioned, coupled with the three-month averaged probabilistic, is thereby exposing seasonal rainfall forecasts generated by ZMD to more complex constraints. The study focused on assessing the reliability of NMME models in quantitative forecasting of monthly rainfall over the Kafue River Basin. The study employed a well-selected multi-model ensemble approach to infer gridded monthly rainfall forecast from global NMME models using validated gridded SRPs over the Kafue River Basin.

The Kafue River Basin plays an important role in Zambia's economy as most of the commercial and domestic water use activities lay within this catchment. Other specific water demands and uses in the basin include industrial, mining, hydropower, agriculture (irrigation and aquaculture), environmental, and recreation. The installed electricity power generation capacity in Zambia is 3356.6 MW, of which 83 percent is hydroelectric power. The Kafue River Basin houses the Kafue Gorge lower and upper hydropower stations and the Itezhi-Tezhi Dam, contributing more than 50% of the installed electricity power generation capacity. The basin also contributes to major industrial, socio-economic, and ecological significance for Zambia, with its water use of about 50% representing Hydropower production, 20% of the National herd, 7% of National fisheries, and 25% of Maize production [9]. The Kafue River Basin has a heavy dependence on rain-fed agriculture (including livestock production), irrigation schemes, hydropower generation, and tourism, among others. This means that the communities in the basins are highly exposed to climate change and its impacts and, at the same time, have the least capacity to adequately adapt. Observational data supports all stages of the adaptation process and several studies have shown that scientific data provide an important starting point for effective adaptation policies in response to climate change.

The generated reliable and gridded monthly rainfall forecasts were envisioned to help sector-specific applications in the scientific modeling of different applications. The validated satellite data also endeavored to provide reliable monthly rainfall gridded time series data, which improved from the sparse meteorological station point data. In the research community, the study may be useful to those focusing on finding the appropriate validated satellite rainfall product and/or a superlative NMME product that can capture the spatial and temporal variations of monthly rainfall over the Kafue River Basin. Consequently, the study will not only fill in the gaps by providing time series gridded rainfall data but also generate reliable and accurate numeric monthly rainfall forecast, providing a monthly value at each grid point. The quantitative monthly forecast can be used as input to scientific meteorological data-based planning and adaptation processes and empower decision-makers in formulating policies for planning and early warning to mitigate the impacts of Climate Change/Variability and enhance economic development in the basin.

2. Data and Methods

2.1. Data Sources

The data sets used for this study were all secondary archived and derived data sets. Daily archival time-series rainfall data were obtained from the Zambia Meteorological Department records for the period January 1983 to December 2022. The other derived datasets used were satellite rainfall products and NMME models downloaded online, as described in detail in the subsequent sections.

2.1.1. Observed Meteorological Station Data

Observed daily rainfall data measured using a standard rain gauge is entered and archived in the ZMD database for supply to the users. The study used data from eight manual weather stations lying within the Kafue River Basin. **Table 1** gives brief attributes of the meteorological stations used in this study.

Table 1. Manual meteorological stations within the Kafue River Basin.

STN ID	STATION	LONGITUDE	LATITUDE	ELEVATION
67563	Kafironda	28.14758	−12.614130	1220
67659	Kafue Polder	27.92140	−15.777080	0976
67541	Kasempa	26.00003	−13.456900	1334
67651	Magoye	27.61725	−15.998260	1025
67667	Mt Makulu	28.24822	−15.548233	1221
67655	Mumbwa	27.18938	−15.078080	1209
67561	Ndola	28.65863	−12.994180	1259
67551	Solwezi	26.36678	−12.170660	1333

2.1.2. Observed Gridded Satellite Rainfall Products

The three satellite rainfall products whose accuracies were compared and evaluated against observed rain gauge data are the Climate Hazards Group Infra-Red Precipitation (CHIRPS), Tropical Application of Meteorology Using Satellite Data and Ground-Based Observations (TAMASAT) and African Rainfall Climatology version 2 (ARC2). Daily images extending from 11˚S to 17˚S and 25˚E and 29˚E from 1983 to 2022 were downloaded online. The spatial resolution and download link details of the individual SRPs are indicated in Table 2.

Table 2. Gridded satellite rainfall products.

SRP	SPATIAL RESOLUTION	DATA SOURCE
CHIRP	0.05˚ × 0.05˚	https://data.chc.ucsb.edu/products/CHIRPS-2.0/
TAMSAT	0.037˚ × 0.037˚	http://www.tamsat.org.uk/data-subset/index.html
ARC2	0.1˚ × 0.1˚	https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/.Africa/.DAILY/.ARC2/.daily/?Set-Language=en

2.1.3. Hindcast and Forecast North America Multi-Model Ensemble Models

The North American Multi-Model Ensemble (NMME) is a seasonal prediction system that combines forecasts from the leading North American climate models. It constitutes a multi-agency and multi-institutional research-to-operations (R2O) effort jointly led by the NOAA Climate Program Office (CPO) Modelling, Analysis, Predictions, and Projections (MAPP) program and the NOAA Climate Test Bed (CTB). The NMME's increased ensemble size compared to NCEP's operational Climate Forecast System (CFSv2) and its diversity of models contribute to superior seasonal forecast skill, on average, relative to other seasonal prediction systems. NMME predictions and hindcasts provide a unique research platform for predictability and prediction research. The different ensembles from different institutions used for this study had a spatial resolution of 1˚ × 1˚ covering 11˚S to

17˚S and 25˚E and 29˚E from 1983 to 2022 and lead time ranging from 0.5 to 11.5 months with more information indicated in **Table 3**, downloaded within the Climate Prediction Tool.

Table 3. North America Multi-Model Ensemble summary details.

North America Multi-Model Ensembles (NMME) Summary			
MODEL	NUMBER OF ENSEMBLE MEMBERS (M)	INSTITUTE	COUNTRY
CMC2 CanSIPSv2	6	Canadian Meteorological Centre	Canada
CANCM4i	10	National Centers for Environmental Prediction	Canada
GFDL-CM2p1	10	National Oceanic and Atmospheric Administration	North America
GFDL-CM2p5-FLOR-B01	10	National Oceanic and Atmospheric Administration	North America
GFDL-CM2p5	10	National Oceanic and Atmospheric Administration	North America
NASAGEOSS2	11	National Aeronautics and Space Administration	North America
NCEPCFSv2	24/28	National Centers for Environmental Prediction	North America
COLA RSMASCCSM4	10	National Centers for Environmental Prediction	North America

2.2. Monthly Weather Forecasting Concept

According to [10], the concept of weather forecasting is based on the principle that current and past knowledge can be used to make predictions. For time series data in particular, there is the belief that it is possible to identify patterns such as trends in historical values, and positively implement them in the process of forecasting future values. However, forecasting theories are developed as forecasting methods and models. In this case, methods are defined to be a set of predetermined sequences of steps that produce forecasts for future periods [10]. The main steps in weather forecasting are data acquisition, data pre-processing, model training, performance validation, and finally visualization [2] as integrated into the study are indicated in **Figure 1**.

2.3. Data Analysis in Generating Monthly Rainfall Forecast

Observational records are affected by a large number of data quality issues and therefore data quality assessment and control of such data sets is an important but mostly overlooked aspect in weather and climate analysis [11]. Other than the data assessment and quality control carried out at the station level, the study used the Climate Data Tool's data quality function to assess its availability, spatial check,

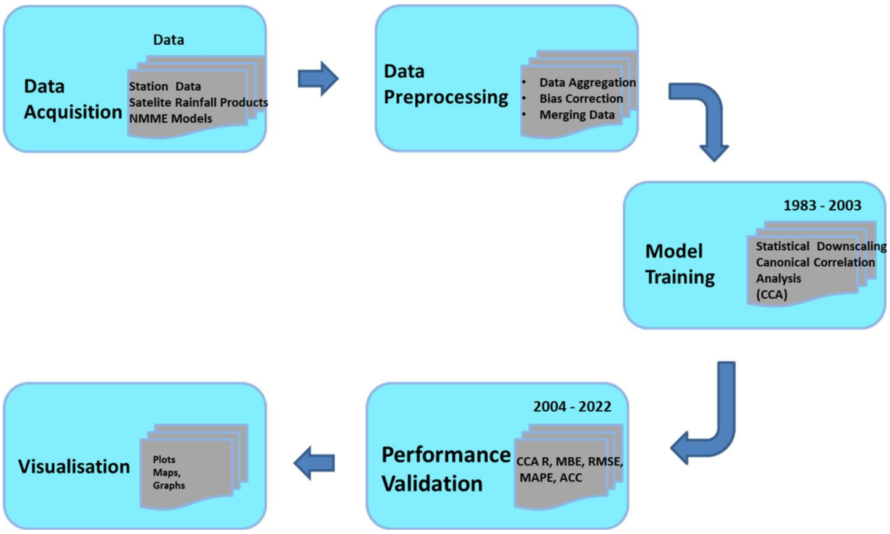


Figure 1. Basic representation of steps in weather forecasting adopted from Jaseena & Kovoor, 2020.

and outliers. To generate a time series of monthly gridded observational data sets, the study validated three satellite rainfall estimate products: Chirps, ARC 2, and TAMSAT against station point data. The disaggregate function within the CDT was used to enhance the spatial resolution of Chirps and ARC 2 satellite images to that of (0.037˚) for easy comparison. The results from the validation were used to select the best-performing satellite and the bias correction coefficients were computed using the quantile mapping with empirical distribution algorithm. The bias corrections were applied before merging the satellite images with station data to generate a time series gridded dataset from 1983 to 2022.

To post-process and recalibrate raw GCMs, one of the two downscaling techniques can be used: statistical downscaling or dynamical downscaling which is based on the conceptual and mathematical frameworks [12]. Statistical downscaling methods endeavor to link the large-scale atmospheric predictor variables to local- or regional-scale meteorological time series data. It is capable of extracting point-scale climate information from a GCM simulation on the assumption that statistical relationships remain unchanged in a changed future climate. On the other hand, dynamic downscaling approaches use finer-gridded models, technically referred to as regional climate models (RCMs), to redefine the climate change predictions at a regional scale from the coarse projection of the GCMs [13]. Dynamical downscaling requires huge computing resources and highly skilled human resources [14]. Statistical downscaling models are based on statistical relationships and hence require less computational time and costs. A combination of the two methods creates a hybrid and can be complemented with an ensemble approach. Statistical downscaling applied individually to individual ensemble forecasts looks into the benefits of multi-model forecasting as a post-processing measure [15]. The study inferred local monthly rainfall from global NMME models using the Canonical Correlation Analysis algorithm within the Climate Prediction Tool (CPT). Time series data was divided into the calibration period (1983-2003) and the validation period (2004-2022), and all were compared against the extended climate normal period (1983-2020). To validate the performance of the NMME model against station data, quantitative validation metrics, Spearman's correlation coefficient, mean bias, and root mean square error obtained from the canonical correlation analysis were used to evaluate individual NMME model performance. The best-performing NMME models were selected for each month and then validated again using the gridded time series data within the CPT environment. Validation metrics based on the generated gridded monthly forecast were also assessed before being adopted for reliability test using the Anomaly Correlation Coefficient (ACC). QGIS's raster calculator was used to assess selected individual models' reliability in forecasting monthly using the Anomaly Correlation Coefficient formula. The monthly rainfall ensemble forecast was generated by averaging the selected models for each month and again assessing their reliability using the Anomaly Correlation Coefficient to assess their reliability in ensemble forecasting.

2.4. Mathematical Equations for the Analysis

2.4.1. Generating Gridded Monthly Rainfall

The best-performing SRP was adopted and merged with time series meteorological station data. Before merging with station data, an empirical Quantile Mapping (QM) bias correction algorithm was used to correct systematic distributional biases in satellite rainfall products in the CDT software environment. Empirical QM is a nonparametric method that relaxes the normal distribution assumption [16]. The guiding equation for quantile mapping is indicated in Equation (1).

Given:

- F_0^{-1} = Inverse cumulative distribution function (CDF) of the observed data
- F_m = CDF of the modeled data
- X_m = Modeled data value

The quantile mapping correction can be written as:

$$X_0 = F_0^{-1}\left(F_m\left(X_m\right)\right) \tag{1}$$

where:

- X_0 is the corrected (bias-adjusted) modeled value.
- X_m is the original modeled value.
- $F_m(X_m)$ is the quantile of the modeled value X_m.
- F_0^{-1} maps this quantile to the corresponding value in the observed distribution.

Precipitations derived from satellites have proven to be a good alternative to ground-based precipitation observations [17]. The main advantages of satellite rainfall products over ground-based observation are their larger coverage area and high spatio-temporal resolution. However, all satellite precipitation products provide indirect estimates of precipitation and therefore drive the need to assess the products by comparing their data with ground observations.

Evaluation of the satellite rainfall products was based on cross validating satellite rainfall products against ground observed data with a 95% confidence level extracted at the same location for the eight meteorological stations in the Kafue River Basin from 1983 to 2022. The validation metrics as described and recommended by Joint Working Group on Forecast Verification Research [18] used to compare the three SRPs and the ground observations were Correlation Coefficient, Bias, Maen Absolute Error and the Root Mean Square Error and Equations (2), (3), (4) and (5) respectively were used.

where:

(S_i) denotes the rank of the ith observation among n observations.

(O_i) denotes the rank of the ith value among n predicted values.

(R) denotes the difference between the ranks of the corresponding values for each pair of observations and predicted values.

Correlation Coefficient (r)

$$\text{r} = \frac{\sum\left(S-\bar{S}\right)\left(O-\bar{O}\right)}{\sqrt{\sum\left(S-\bar{S}\right)^2}\sqrt{\sum\left(O-\bar{O}\right)^2}} \tag{2}$$

$$\text{Bias} = \frac{1}{N}\sum\nolimits_{i=1}^{n}\left(S_i - O_i\right) \tag{3}$$

$$\text{MAE} = \frac{1}{N}\sum\nolimits_{i=1}^{n}\left|S_i - O_i\right| \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{n}\left(S-\bar{S}\right)^2} \tag{5}$$

2.4.2. Evaluating the Validity of North America Multi-Model Ensemble Performance in Forecasting Monthly Rainfall

The first approach for comparison in the forecasts quality of the NMME forecast is based on the deterministic forecasts using the ensemble mean as a deterministic representative for the ensemble forecast and applying standard verification scores. The verification of the deterministic forecast was mainly based on the measure of the difference between forecasts and observations. Standard methods for quantitative forecast verification recommended by the World Meteorological Organization [19] used were Spearman's correlation coefficient, mean bias, root mean square error, and mean absolute percentage error.

1) Spearman's Rank Correlation Coefficient (rs)

This statistic measures the dependence between observations and predicted values using some monotonic function [20] given by the equation,

$$\text{rs} = \frac{\sum\nolimits_{i=1}^{n}\left(R\left(S_i\right)-\overline{R\left(S\right)}\left(R\left(O_i\right)\overline{R\left(O\right)}\right)\right)}{\sqrt{\sum\left(R\left(S\right)-\overline{R\left(S\right)}\right)^2}\sqrt{\sum\left(R\left(O\right)-\overline{R\left(O\right)}\right)^2}} \tag{6}$$

2) Mean Absolute Error (MAE)

Basically, identifies the systematic biases in the forecasts or estimates as indicated in Equation (4).

3) Root Mean Square Error (RMSE)

The RMSE corresponds to the average squared difference between the forecast and observation pairs, as shown in Equation (5). The RMSE has been used as a standard statistical metric to quantify model performance in meteorology, air quality, and climate research studies [21].

4) Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) is the mean of all absolute percentage errors between the predicted and actual values. The equation for calculating the MAPE is given by:

$$\text{MAPE} = \frac{1}{n}\sum\left|A_t - F_t\right| \times 100 \tag{7}$$

where: n is the size of the sample, A_t is the value predicted by the model for time point t, and F is the value observed at time point t.

In order to calculate MAPE value, we need to divide the difference between the forecast and observation by the observed value and thus if you have actual values close to or equal to 0 then your MAPE score will either receive a division by 0 error, or be extremely large. Therefore, it is advised not to use MAPE when you have actual values close to 0 [22]. All values closer to or equal to zero were excluded from the analysis especially for the month of October.

5) Anomaly Correlation Coefficient (ACC)

ACC is the correlation between anomalies of forecasts and those of verifying values with the reference values, such as climatological values. Quantum GIS's raster calculator was used to generate the ACC values. ACC was given by

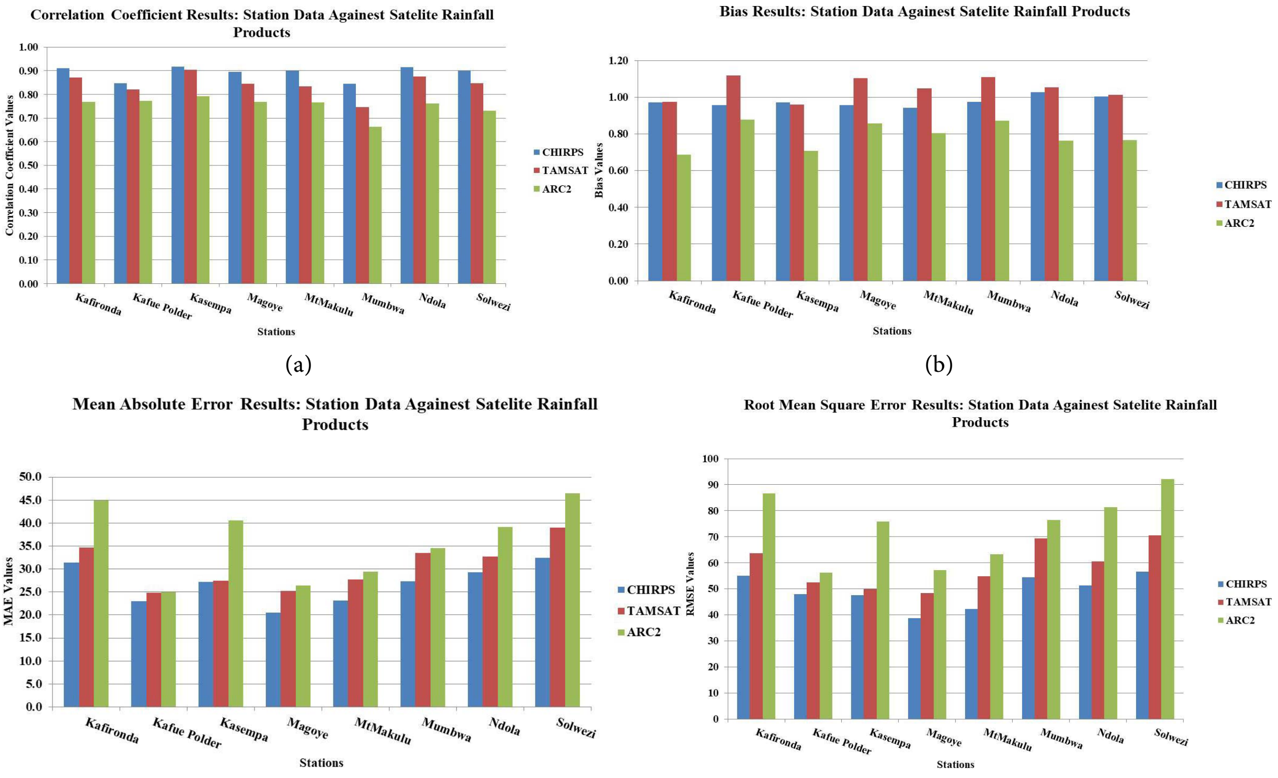
$$\text{ACC} = \frac{\sum i\left(O_i - \bar{O}\right)\sum i\left(S_i - \bar{S}\right)}{\sqrt{\sum i\left(O_i - \bar{O}\right)^2}\sqrt{\sum i\left(S_i - \bar{S}\right)^2}} \tag{8}$$

3. Results and Discussion

3.1. Results

3.1.1. Determining the Best Satellite Rainfall Product over the Kafue River Basin

At a 95% confidence level, CHIRPS SRP had the best correlation with the observed station rainfall data; the strongest correlation was at Kafironda, Kasempa, and Ndola ($r > 0.9$) and the least correlation was at Kafue Polder and Mumbwa ($r = 0.85$). TAMSAT was the second-best SRP with the strongest correlation over Kasempa ($r > 0.9$), and the least correlation over Mumbwa ($r = 0.75$). ARC SRP was least correlated with station data ($r < 0.8$) for all stations. In terms of bias, values equal to zero bias are called unbiased, and the more the value moves away from zero, the greater the bias. Comparing the three SRPs, ARC2 had the least mean bias with values ranging from 0.69 over Kafironda to 0.88 over Kafue Polder, followed by CHIRPSS with 0.94 (Mt Makulu) to 1.03 (Ndola) and TAMSAT 0.96 (Kasempa to 1.11, Mumbwa). CHIRPSS performed better than TAMSAT and ARC2 for the remaining two validation metrics, as it had values that were lower than the two other SRPs for all the stations in the basin. The CHIRPS satellite product was merged with station observed data to generate a time series gridded data set for the Kafue River Basin for the period 1983 to 2022 which was used as an input in the quantitative monthly forecasting of the Kafue River Basin. Results from the various validation metrics are indicated in Figure 2.



(a) (b)

(c) (d)

Figure 2. Station monthly rainfall results: correlation coefficient, bias, mean absolute error, and root mean square error.

3.1.2. North America Multi-Model Ensemble Validity Performance in Quantitative Forecasting of Monthly Rainfall

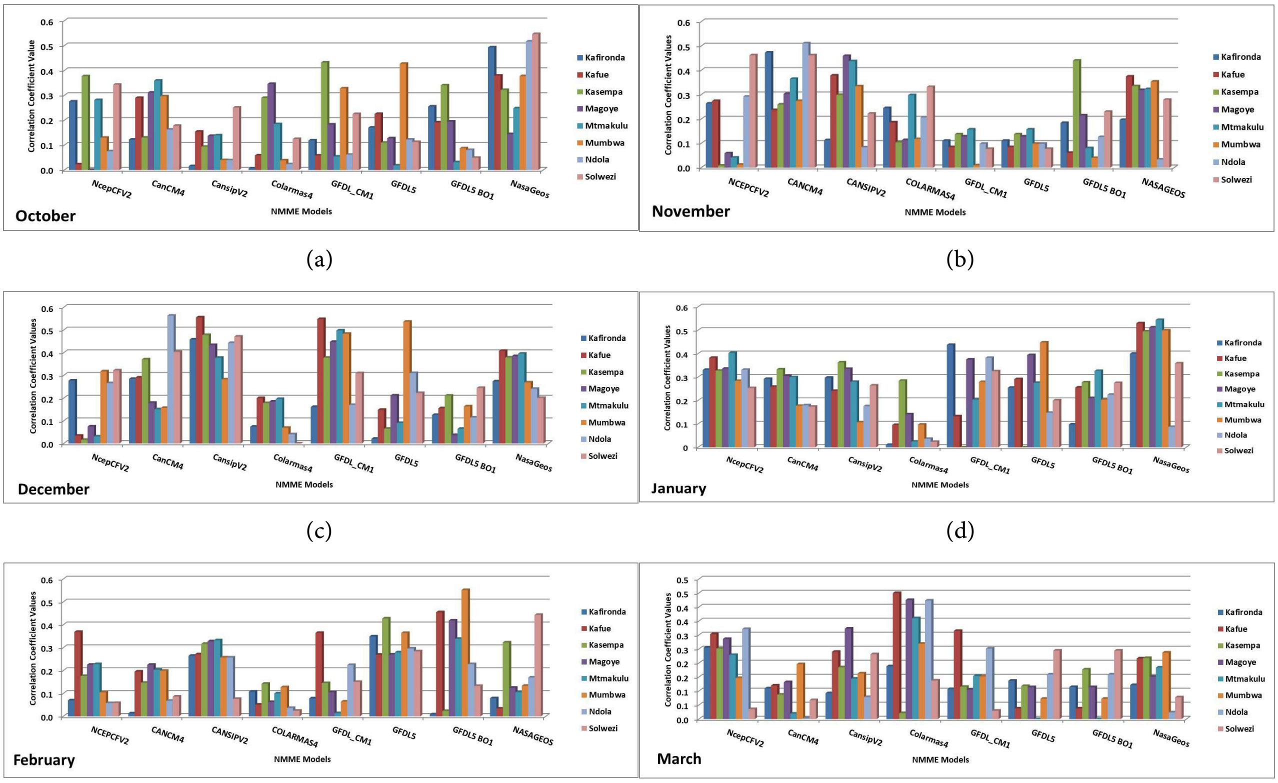
Based on the results obtained from the CCA, the best-performing models were selected for each month centered on the performance of the described quantitative forecasting validation metrics. Table 4 shows the assumptions considered in selecting the best-performing models and generating the quantitative ensemble monthly forecast for the Kafue River Basin.

Analysis results based on monthly results indicated that October had only three models that had four or more stations having correlation coefficient values greater than 0.30; NCEPCFV2, CANCM4i, and NASAGEOS. In November, NCEPCFV2, CANCM4i, CANSIPV2, and NASAGEOS were adopted with CANCM4i, CANSIPV2, and GFDL_CM1 for December based on association in response to Spearman's correlation coefficient values. In terms of strength of association (correlation) between the NMME models and the observed monthly rainfall for the month of January, only the COLARMAS4, GFDL5BO1, and the GFDL_CM1 model had less than four stations with coefficient values less than 0.30, the rest were adopted.

Table 4. NMME model selection and forecasting assumptions.

ASSUMPTION	PERFECT SCORE	SUPPORTING LITERATURE
Spearman's Correlation (≥0.30)	**1** (range: −1 to 1)	[23]
Bias (Mean Error) > 1	**0** (range: 0 to ∞)	Evaluated within the data set no general rule
Normalized Root Mean Square Error < 0.29	**0** (range: 0 to ∞)	Evaluated within the data set no general rule but normalised for easy comparison [24]
Mean Absolute Percentage Error (1 to 50)	**1** (1 to 100)	1 to 50 lies with highly accurate to Reasonably accurate forecasting [25]
Anomaly Correlation Coefficient (1)	**1** (Range −1 to 1)	[26]
Extended Climatological Normal	More than the standard 30-year period	[27]
Simple Composite Method (SCM); for Ensemble Generation		[28]

The CANSIPV2, GFDL5, and GFDL5BO1 were adopted for the month of February with March having NCEPCFV2 and COLARMAS4. The general performance for all the models at the station level for October, November, December, January, and February up to March shows that the least correlation coefficient value was zero indicating no correlation and the strongest was 0.60. **Figure 3** shows the correlation values for the months of October, November, December, January, February, and March for each model.



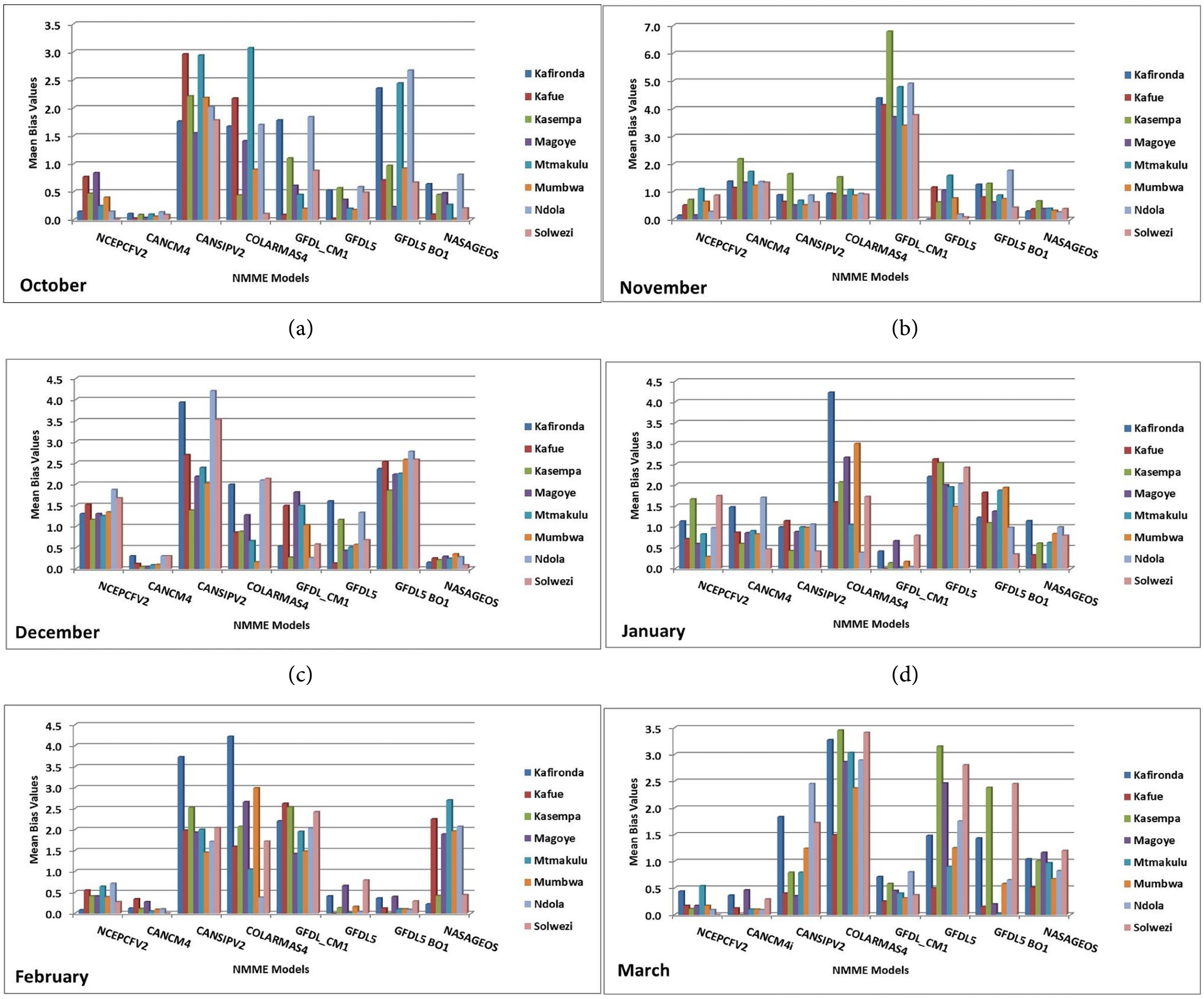
(a) (b)

(c) (d)

(e) (f)

Figure 3. NMME models against station monthly rainfall; correlation coefficient results.

The selected models based on the mean bias for the month of October were NCEPCFV2, CANCM4i, and NASAGEOS; November had NCEPCFV2, CANSIPV2, and NASAGEOS. For the months of December and January, NASAGEOS and CANCMi4; and CANCMi4, CANSIPV2, GFDL_CM1, and NASAGEOS were adopted respectively. The NCEPCFV2, CANCMi4, GFDL 5, and GFDL5 BO1 were selected for the month of February. Finally, NCEPCFV2, CANCMi4, GFDL_CM1, GFDL5 BO1, and NASAGEOS were adopted for the month of March. See Figure 4 (October, November, December, January, February, and March) with at least four stations with mean bias values equal to or less than 1.



(a) (b)

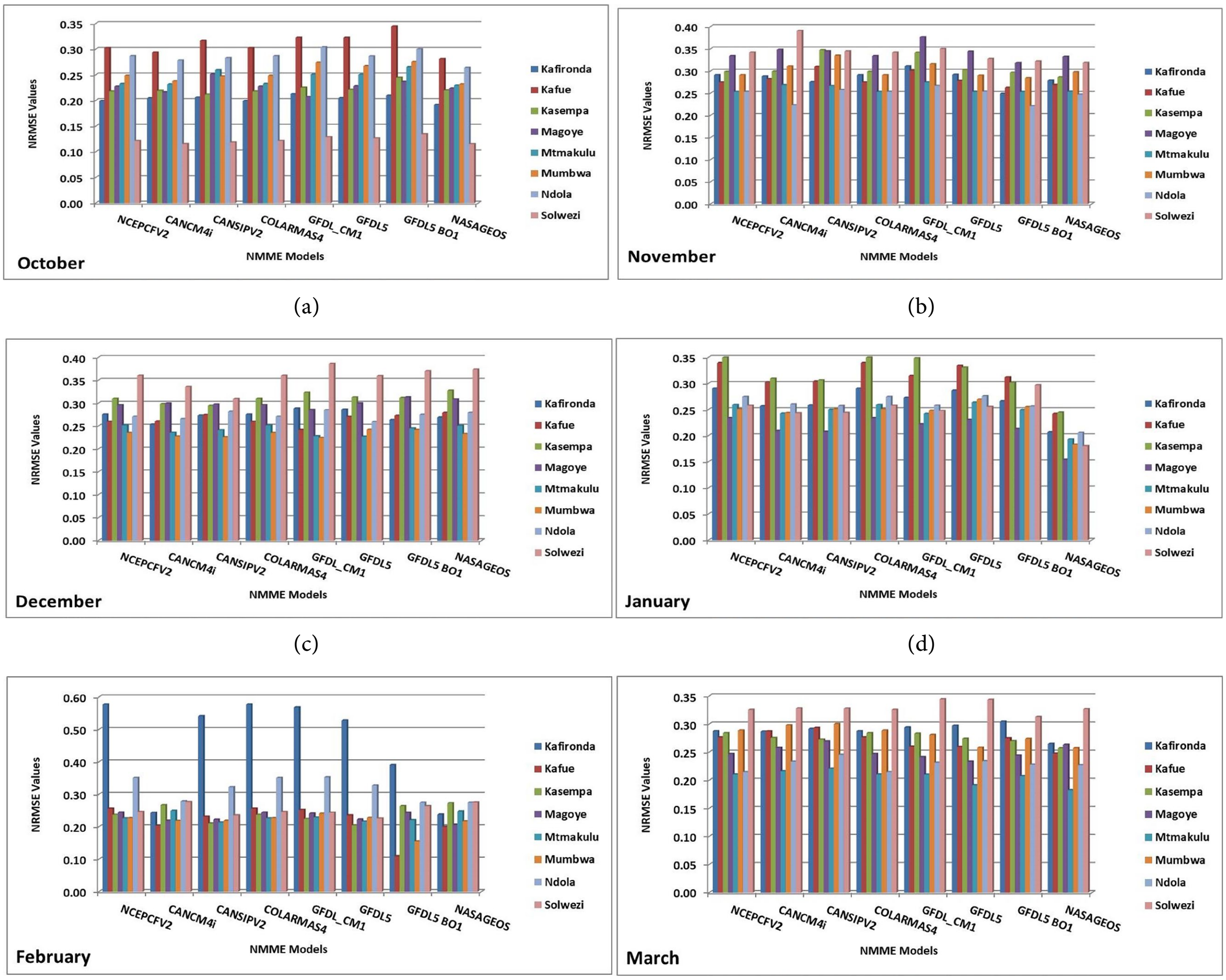
(c) (d)

(e) (f)

Figure 4. NMME models against station monthly rainfall; mean bias results.

Models that had at least four stations with NRMSE between 0 and 0.29 were considered and adopted for each month. The monthly results showed that all eight models were considered and adopted based on the test for model sharpness, with the NRMSE values ranging from 0.12 to 0.34 in October. In November, the NASA-

GEOS and GFDL5 BO1 models had five stations as the highest number with NRMSE values less than 0.30; while CANCMi4 and GFDL5 had four stations. For the remaining four months, all the models were considered, although NRMSE values greater than 0.50 were observed over Kafironda station from the NCEPCFV2, CANSIPV2, GFDL_CM1 GFDL 5, and COLARMAS4 models in the month of February. See **Figure 5** for NRMSE values for October, November, December, January, February, and March.



(a) (b)

(c) (d)

(e) (f)

Figure 5. NMME Models against station monthly rainfall; normalised root mean square error (NRMSE).

MAPE can be understood as the inverse of model accuracy, but more specifically as the average percentage difference between forecasts and their intended targets in the dataset. In this study, the interpretation of the MAPE results was used to assess the accuracy of the predictions. A MAPE value less than 10 was considered a highly accurate forecast, and that which is greater than 50 was an inaccurate forecast adopted by Moreno *et al.*, 2013 with other interpretations shown in **Table 5**.

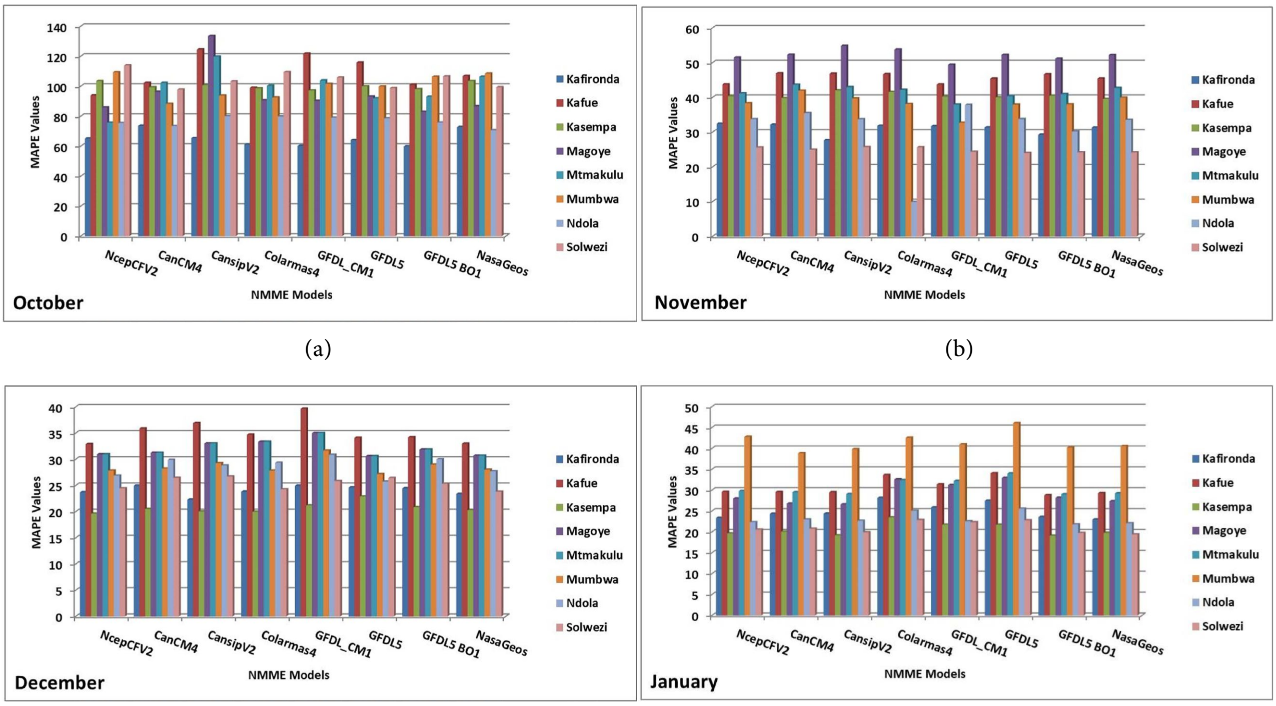
Table 5. Mean Absolute Percentage Error (MAPE) interpretation (adopted from Moreno *et al.*, 2013).

Interpretation of Typical MAPE Results	
MAPE	**INTERPRETATION**
<10	Highly Accurate Forecasting
10 to 20	Good Forecasting
20 to 50	Reasonable Forecasting
>50	Inaccurate Forecasting

MAPE values for almost all the stations in October had values ranging between 60 and 120 except for Kafironda which indicated MAPE values of below 50 for all the models and Mt Makulu which also had MAPE values less than 50 for the GFDL5 and NASAGEOS model. For the month of November, the MAPE results showed some improvement with all models having MAPE values between 20 and 50 except for Magoye which had values greater than 50 for almost all the models. MAPE values for the months of December, January and February had all stations having values between 50 and 10 for almost all the models. However, Mt. Makulu had MAPE values greater than 50 for the NCEPCFV2, CANCM4i, CANSIPV2 and GFDL5_BO1, see **Figure 6** for the individual NMME model's performance.

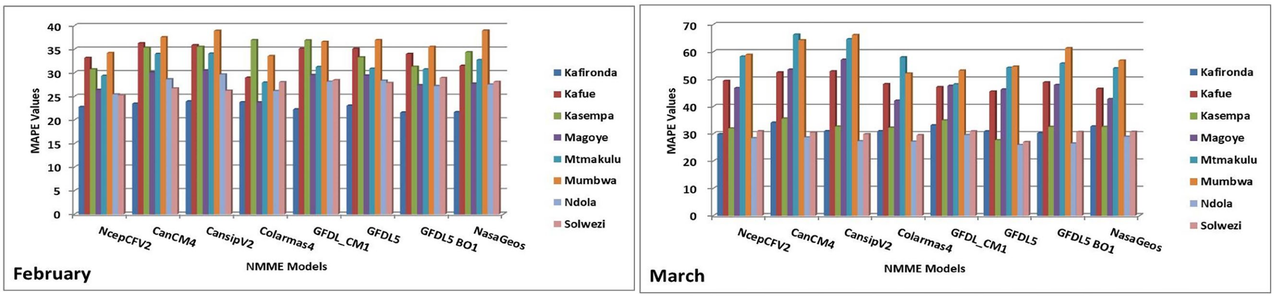
3.1.3. Selection of North America Multi Model Ensemble Models for Monthly Rainfall Forecasting

The selection of NMME models used in the quantitative forecasting of monthly rainfall was based on the four adopted quantitative validation metrics; Spearman's



(a) (b)

(c) (d)



(e) (f)

Figure 6. NMME models against station monthly rainfall; Mean Absolute Percentage Error (MAPE).

correlation coefficient, mean error, normalised root mean square error and mean absolute percentage error. Models which had at least three out of four of the validation metrics were considered in the final ensemble model. Based on the performance of individual models on a monthly basis for the model calibration period 1983 to 2003, the selected models for the months October, November, December, January, February and March are indicated in Table 6 and Table 7 (arranged according to performance for each standard validation metric).

Table 6. Selection of ensemble forecast models based on the four-validation metrics (October, November and December).

	Spearman's Correlation Coefficient	Mean Error	Normalised Root Mean Square Error	Mean Absolute Percentage Error	Ensemble Forecast Models
OCTOBER	NASAGEOS CANCM4i NCEPCFV2	CANCM4i NASAGEOS GFDL5 NCEPCFV2 GFDL_CM1 GFDL5 BO1	NASAGEOS CANCM4i NCEPCFV2 COLARMAS4 CANSIPV2 GFDL_CM1 GFDL5 GFDL5 BO1	NONE	NASAGEOS CANCM4i NCEPCFV2
NOVEMBER	CANCM4i NASAGEOS CANSIPV2 NCEPCFV2	NASAGEOS NCEPCFV2 CANSIPV2 GFDL5 BO1 COLARMAS4 GFDL5	NASAGEOS GFDL5 BO1 GFDL5 CANCM4i	NASAGEOS CANCM4i NCEPCFV2 COLARMAS4 CANSIPV2 GFDL_CM1 GFDL5 GFDL5 BO1	NASAGEOS CANCM4i NCEPCFV2 CANSIPV2 GFDL5
DECEMBER	CANSIPV2 NASAGEOS GFDL_CM1	CANCM4i NASAGEOS GFDL_CM1	GFDL_CM1 CANSIPV2 CANCM4i	GFDL5 CANCM4i COLARMAS4	NASAGEOS CANCM4i CANSIPV2

Continued

	CANCM4i NCEPCFV2	GFDL5 COLARMAS4	NCEPCFV2 COLARMAS4 GFDL5 GFDL5 BO1 NASAGEOS	GFDL_CM1 NASAGEOS GFDL5 BO1 NCEPCFV2 CANSIPV2	GFDL_CM1 NCEPCFV2

Table 7. Selection of Ensemble Forecast Models based on the four-validation metrics (January, February and March).

	Spearman's Correlation Coefficient	Mean Error	Normalised Root Mean Square Error	Mean Absolute Percentage Error	Ensemble Forecast Models
JANUARY	NCEPCFV2 NASAGEOS CANSIPV2 GFDL_CM1 GFDL5 GFDL5 BO1	GFDL_CM1 NASAGEOS CANCM4i CANSIPV2 NCEPCFV2	NASAGEOS CANCM4i CANSIPV2 GFDL_CM1 GFDL5 GFDL5 BO1 NCEPCFV2 COLARMAS4	NASAGEOS NCEPCFV2 GFDL5 CANCM4i GFDL5 BO1 GFDL_CM1 CANSIPV2 COLARMAS4	NASAGEOS CANCM4i CANSIPV2 NCEPCFV2 GFDL_CM1 GFDL5 GFDL5 BO1
FEBRUARY	CANSIPV2 GFDL5 GFDL5 BO1	NCEPCFV2 CANCM4i GFDL5 GFDL5 BO1	NASAGEOS CANCM4i GFDL5 BO1 GFDL5 NCEPCFV2 CANSIPV2 COLARMAS4 GFDL_CM1	NASAGEOS GFDL5 BO1 COLARMAS4 CANCM4i NCEPCFV2 GFDL_CM1 GFDL5 CANSIPV2	NCEPCFV2 CANSIPV2 CANCM4i GFDL5 GFDL5 BO1
MARCH	COLARMAS4 NCEPCFV2	NCEPCFV2 CANCM4i GFDL_CM1 GFDL5 BO1 NASAGEOS	NASAGEOS NCEPCFV2 COLARMAS4 CANCM4i GFDL_CM1 GFDL5 GFDL5 BO1 CANSIPV2	NASAGEOS COLARMAS4 GFDL5 GFDL_CM1 NCEPCFV2 GFDL5 BO1 CANCM4i CANSIPV2	NCEPCFV2 COLARMAS4 NASAGEOS CANCM4i GFDL_CM1 GFDL5 BO1

3.2. Quantitative Monthly Rainfall Forecasting for the Kafue River Basin Using Gridded Data

For the validation period (2004 to 2022), the gridded data set was used to generate the forecast for each of the selected models for the six months, October to March.

Gridded validation metrics were analysed for the selected individual NMME models to strengthen their reliability and also to understand their spatial and temporal variation before generating the ensemble forecast.

In terms of association, the NCEPCFV2 model showed good correlation skills (r ≥ 0.6), especially over the western half of the basin, as compared to the NASAGEOS model in October. Generally, over the whole basin, Spearman's correlation values ranged between ±0.75 to ±0.15 with lower values in the eastern part of the basin (NCEPCFV2 model) and ranging between 0.3 to −0.3 with the NASAGEOS model. The NASAGEOS model had lower RSME values (5 to 10) over much of the basin as compared to the NCEPCFV2 model (15 to 20 though both models indicated RMSE values greater than 30 over the southeastern parts of the basin for the month of October (**Figure A1**). The mean absolute percentage error (MAPE) results for all two models showed values greater than 50 over much of the Kafue River Basin, apart from small patches over the southern parts (see **Figure A2**).

In November, the NASAGEOS model had Spearman's correlation values r = 0.9 in some areas over the northern parts of the basin. The NCEPCFV2 model had a good correlation to the southern part of the basin, tending to be weaker over the northern half. The NCEPCFV2 and NASAGEOS models had closer to zero mean bias values over the central into the eastern parts of the basin, and negative values went to as low as −85, especially over the central and western parts of the basin (**Figure A3**). In terms of accuracy using the MAPE test, the NASAGEOS model generally indicated values between 10 and 20 over most parts of the basin with a few areas over the south and northeastern parts which had values between 20 and 50. Conversely, the NCEPCFV2 models had most parts with MAPE values between 20 and 50 with only a few areas over the central parts having values between 1 and 20 (**Figure A4**).

Generally, negative correlation values were observed for the month of December, ranging between 0 to slightly below −0.75, with the NCEPCFV2 performing better than the NASAGEOS model in terms of association or strength. Mean bias values were between 0 and −95 for both models over much of the basin. The RMSE values ranged between 12 and 100 for both models, with higher values observed in the central parts of the basin (see **Figure A5**). The NCEPCFV2 model performed better than the NASAGEOS model, with MAPE values ranging between 1 and 50 over much of central and northern parts, though the southeastern parts had values greater than 50 (see **Figure A6**).

In January, the NASAGEOS model indicated a good positive correlation with values r = 0.9 especially over the northern parts of the basin. The CANSIPV2 model had a mixture of positive correlation to the north and the southeastern parts and negative correlation to the western and the central parts of the basin (values between −0.45 and 0.60). For both models, the mean bias had positive values as high as 42.5 in the northwestern parts and as low as −52.5 over the southwestern parts of the basin (**Figure A7**). An overview of the MAPE results in January showed MAPE values between 1 and 20 over the northern half and between 20 and 50 over the southern half. The NASAGEOS model performed better than the CANSIPV2 model in terms of spatial coverage of the area with good results (below 20) (**Figure A8**).

For the month of February, all the models that were selected had a −1 ACC value over the whole basin, but the study also evaluated the spatial variation of the performance validation metrics. Weak correlation between the observed and the NASAGEOS model was shown especially over much of central and southeastern parts r = 0 to −0.15. The CANSIPV2 model was slightly better even though the general performance was still weak, with Spearman's correlation coefficient values between −0.3 and +0.3. Negative mean bias values were shown by both models over much of the basin, getting to as low as −57.5 and −37.5 for the CANSIPV2 and NASAGEOS models respectively. The RMSE values were between 30 and 110 for the CANSIPV2 and between 27.5 and 102.5 for the NASAGEOS model (see **Figure A9**). The general performance in terms of the MAPE had values between 20 and 50 in the extreme northern parts and greater than 50 over the rest of the basin for both models (**Figure A10**).

The CAMCN4i model performed better than the NCEPCFV2 model in terms of association for the month of March. Spearman's correlation coefficient values of more than 0.9 were observed over the extreme northern parts, with all the models having a positive correlation with much of the basin. Both models indicated high RMSE values greater than 132 over the extreme southeastern parts of the basin (**Figure A11**). Most parts of the basin had MAPE values between 1 and 50, with only small pockets that had values greater than 50 for the two selected models. (**Figure A12**).

3.3. Ensemble Model Monthly Rainfall Forecasting over the Kafue River Basin

The selected NMME model's reliability in predicting gridded monthly rainfall was tested using the Anomaly Correlation Coefficient (ACC). ACC reflects the spatial consistency of the model forecast and observations. The ACC values range from −1 to 1. The larger the value of the ACC, the more consistent the spatial pattern between the forecast and the observations [26]. The interpretations of the values based on this study are,

- Approach +1: there is good agreement and the forecast is reliable.
- Lie around 0: there is poor agreement and the forecast has had no value.
- Approach −1: the agreement is in anti-phase and the forecast has been very unreliable.

3.3.1. Ensemble Model Sensitivity Analysis

The study assessed the effect of using an ensemble or single model approach in months that had more than one NMME model showing reliability using the ACC. The reliability of the selected individual models was assessed on a monthly basis before generating the ensemble forecast. Results show that the two models NCEPCFV2 and NASAGEOS selected for the month of October had ACC values equal to −1 for much of the basin except for a small area around Magoye, which had values equal to 1 (see **Figure A13**).

In November, the mean ACC values for the two selected models were the opposite of the other. The NASAGEOS model had ACC values equal to 1 over much of the basin apart from areas around Mumbwa moving northwards towards Kafironda. On the other hand, the NCEPCFV2 model had −1 ACC values over most parts of the basin apart from the few areas that had positive ACC values results with the NASAGEOS model as indicated in **Figure A14**.

For the month of December, both the NASAGEOS and the NCEPCFV2 models ACC values of 1 over much of the basin except for a few areas around Kafironda over the extreme northeastern parts which had negative values (see **Figure A15**).

Four models, CANCM4i, CANSIPV2, NCEPCFV2 and NASAGEOS were adopted for the month of January. The maps under **Figure A16** show that the NCEPCFV2 and CAMCN4i models generally had ACC values of −1 over much of the basin with only a few patches in the northern part which had ACC values equal to 1. The other two models gave opposite results, with much of the basin having ACC values equal to 1 and a few areas to the north equal to −1.

In February, the two models selected indicated anti-phase agreement with ACC values equal to −1 (see **Figure A17**).

For the month of March, four models were selected, CANCM4i, COLARMAS4, NCEPCFV2 and NASAGEOS. The NCEPCFV2 and CAMCN4i models had most parts of the basin having ACC values equal to 1 as compared to the other two models which had most parts of the basin having ACC values equal to −1 as seen in **Figure A18**.

All the models that had ACC values equal to 1 over much of the basin were selected to be part of the ensemble forecast. For the month of November, only the NASAGEOS model was adopted since it had most parts of the basin indicating reliability as compared to the NCEPCFV2 model. An ensemble of the two reduced the spatial coverage in terms of reliability.

3.3.2. Ensemble Model Monthly Rainfall Forecasting over the Kafue River Basin

The monthly ensemble forecasts were generated by calculating the average of all the selected models for each month and comparing the ensemble forecast with the observations to check for the reliability of the Ensemble forecast using the Anomaly Correlation Coefficient (ACC). Ensemble forecasts from the selected models point to ACC values equal to 1 over much of the basin for the months of December, January and March. For the month of November, a single model (NASAGEOS) which had ACC values equal to 1 over much of the basin, was used since an ensemble with the NCEPCFSV2 model reduced the spatial coverage of the area, indicating reliability compared to a single model. For the months of October and February, all the models failed the reliability test, generating an ensemble that generally had −1 ACC values. (See **Figure 7** for October and November; **Figure 8** for

December and January; and **Figure 9** for February and March)

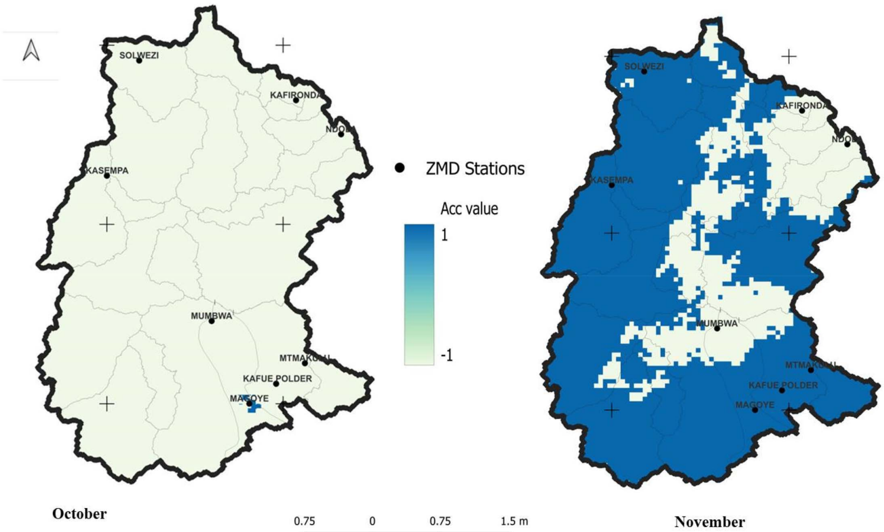


Figure 7. Ensemble Forecast Mean ACC values (October & November 2004 to 2022).

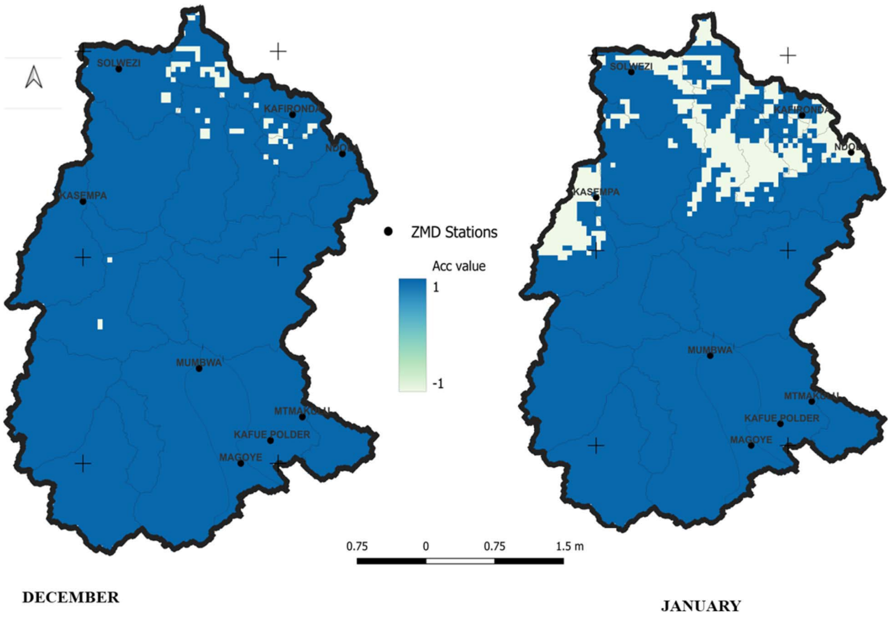


Figure 8. Ensemble Forecast Mean ACC values (December & January 2004 to 2022).

3.4. Discussion

According to [29], quality control ensures the data measured and used for weather and climate analysis has not been contaminated by unrelated factors but is relatively accurate and consistent for dependable decisions to be made. Daily rainfall data collected from Kasempa, Mumbwa and Mt Makulul for the period 1983 to 2022 had gaps of more than 10%, defying the recommendation [29]. For Mumbwa, Kafue Polder and Kasempa, which had availability below 90%, station means were used to fill in the gaps, as the stations could be completely thrown away since

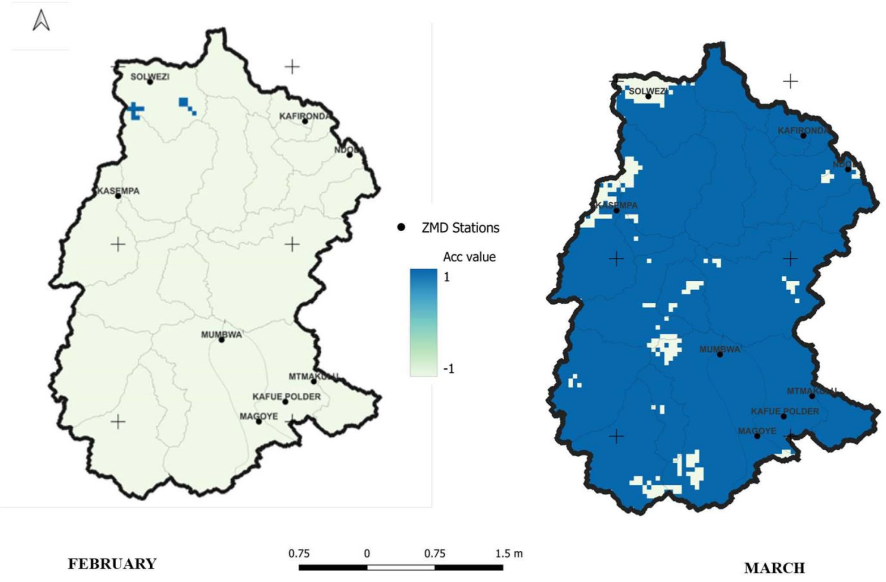


Figure 9. Ensemble Forecast Mean ACC values (February & March 2004 to 2022).

the basin already has a limited number of manual meteorological stations with data of at least 30 years or more. In general, the data was of good quality and in an acceptable form to proceed with the analysis. Satellite-based rainfall products had all the data available for the whole period (1983 to 2022).

Chirps emerged to be the best performing SRP compared to the other two (TAMSAT and ARC2) with correlation coefficient values greater than 0.9 in six out of eight stations, leaving out Mumbwa and Kafue Polder. ARC2 had the least bias results but the highest values in the other validation metrics, failing out of favour. The study merged the point station and the gridded observational data, thereby maintaining the gauge rainfall observed values at all the ZMD stations in the basin. Acquiring high quality rainfall data is very important and this data can be obtained by merging Satellite Rainfall Estimates [30] and a lot of work in validating different SRP data has been done in Zambia as well, notably [31] and [32]. However, analysis of skills of satellite rainfall estimates varies greatly with climate, topography and seasonal rainfall patterns and other factors which may include the spatial and temporal scales. Therefore, it was important for the study to validate the three SRPs before generating time series data specifically for the Kafue River Basin. The observation data can confidently be integrated into other research studies or any weather and climate related analysis.

Rainfall forecasting is one of the most challenging tasks in this context because of the existence of three patterns: the temporal, spatial and non-linear, occurring simultaneously [33]. Therefore, the performance of NMME models in quantitative forecasting of monthly rainfall cannot be evaluated based on only one specific validation metric. This study adopted the [18] description of a good forecast based on quality, the degree to which the forecast corresponds to the actual weather. The attributes of quality are association, mean bias, sharpness and accuracy were linked to the following validation metrics: Spearman's correlation coefficient, mean error, RMSE and MAPE respectively. In terms of association, the strength of correlation in this study was not dependent on the direction or sign and therefore, absolute values of Spearman's correlation were used to assess the association of NMME models with the observed gridded rainfall. [23] considered variables with correlation values of 0.30 (9% of the variance) and above to be interpreted as part of the variance, and variables with loadings below 0.30 to be not. However, [34] indicated that most researchers would probably agree that a coefficient of <0.1 indicates a negligible correlation and that which is >0.9 is a very strong relationship. Values in between are disputable. In this study, a correlation value of absolute 0.30 was considered good enough.

Generally, the correlation results indicated that each station responded differently to each model in different months, although similarities were observed, especially for stations lying within the same ecological regions. Notably, for the month of October, the CANCM4i model indicated poor correlation skills for stations under Region Two ($r \leq 0.20$) and moderate strength for stations under Region Three ($r \geq 0.30$). In November, the NASAGEOS model had poor correlation strength with values ≥ 0.20 in Ndola and Kafironda, all falling under Region Three and falling under Copperbelt Province. In December, the CANCM4i model had Mt Makulu, Mumbwa and Mt Makulu both sit under region two, having correlation values ≤ 0.20 and moderate strength with correlation values ≥ 0.30 over all stations falling under region three.

Results from this study also showed that the association between NMME models and station observed data was within 0, indicating zero association and 0.60 for moderate association for the whole period, as well as all the downscaled models under study. Conversely, the validation of NMME models with gridded data had values ranging from 0 to as high as 0.9, indicating a very strong relationship for the NASAGEOS model in November, December and January; NCEPCFV2 model in November and March; and the CANSPV2 model in March.

The performance of the simulated results produced by each model for each month was also analyzed based on the mean error (bias) values. The results indicated that the CCA predictor models were generally good and closer to the observed historical data for most stations with values equal to or less than 1. The ME values were as low as 0 (indicating no difference between forecast and observed data) and as high as 6.79; the lower the value, the closer the models were to the station data. Positive mean bias values were exhibited by all the models for all the months indicating over estimation by the models at station level. However, the gridded forecast output showed negative values or under-estimation of the forecast models in all the months and all the models in some parts of the basin. In this aspect, the study was able to resolve and capture trends that were not shown in the forecasts generated with station data, considering the large distance between stations which are sparsely distributed in the basin.

The study employed NRMSE to allow easy comparison of all the models for all the months despite having varying rainfall amounts. All the models were able to meet the criteria of having at least four stations attaining the NRMSE assumptions in **Table 5**. NMME model performance against station data was very poor in the month of October, with all the stations having MAPE values greater than 50, indicating inaccurate forecasting. The remaining months, November to March, had MAPE values ranging from 50 to 10, which was reasonable and had good forecasting. The MAPE values had similarities with correlation coefficient results, which showed good skills with gridded data compared to point station data. In some months, the NASAGEOS, NCEPCFV2 and the CANSIPV2 models had MAPE values less than 10 indicating accurate forecasting over much of the Kafue River Basin when validated with gridded data but was only ending at MAPE values between 10 and 20 indicating good forecast with station observed data.

The study has established that NMME models performed better when evaluated against bias-corrected gridded observation data than point station data. A monthly assessment of the performance of NMME models in forecasting rainfall on a monthly basis using gridded data improved the overall skills of the validation metrics for the forecast models.

The basic idea of a multi-model concept is to account for these inherent model errors by using a number of independent and skillful models in the hope of better coverage of the whole possible climate phase [35]. Even if Multi-Model Ensembles (MMEs), are often used to reduce the uncertainties related to GCM projections [36], the study showed that it is important to individually assess the capability of each model before generating an ensemble forecast. This was evident in the ensemble model sensitivity analysis the study employed to assess all the models selected for each month. In the month of November, combining the NASAGEOS and NCEPCFV2 models reduced the spatial extent with ACC values equal to 1 indicating reliability. The NASAGEOS model individual had a large spatial coverage in terms of reliability as compared to generating an ensemble with the NCEPCFV2 model.

In evaluating NMME model's reliability in forecasting monthly rainfall over the Kafue River Basin, the ACC values showed that the NASAGEOS model is reliable in predicting rainfall over much of the basin in November, December, January and March. The NCEPCFV2 model had ACC values equal to 1 over most parts of the basin in December with the CANSIPV2 and the CANCM4i showing reliability in January and March respectively. The eight NMME models assessed in this study are not reliable in forecasting monthly rainfall for the months of October and February with ACC values equal to −1 over much of the Kafue River Basin implying the agreement between the NMME models and the gridded observed data is in ant-phase and the forecast has been very unreliable. In October, as the rain season sets in, the oscillation of the Congo Air Boundary and other local phenomena, such as westerly waves and the Angola Low-Pressure System, play an important role in contributing to the total monthly rainfall. Global models generated through Numeric or Prediction may not be able to resolve such variations which are probably considered at a more local scale.

All the models under the Geophysical Fluid Dynamics Laboratory Earth System

Model with Modular Ocean Model were excluded in the final analysis despite having performed well at evaluation stage because updates were not provided for the year 2021 and 2022.

4. Conclusions

The primary goal of this study was to assess the reliability of NMME models in quantitative forecasting of monthly rainfall over the Kafue River Basin for the six months of the rainy season in Zambia: October, November, December, January, February and March.

The first step in assessing the reliability of NMME models in the quantitative prediction of monthly rainfall over the Kafue River Basin was the generating of a time series gridded dataset using a merge of station gauge data and the CHIRPS satellite rainfall product using quantile mapping empirical distribution algorithm.

The second objective was achieved by generating a time series of monthly forecasts for the Kafue River Basin using canonical correlation analysis. Time series monthly station data lying within the basin was used to calibrate a collection of eight state-of-the-art GCMs from the NMME project using the Canonical Correlation Analysis for the training period 1983 to 2003. Four quantitative validation metrics recommended by the WMO: Spearman's correlation, mean error, root mean square error and mean absolute percentage error were used to assess the skill of each model for each individual month. Based on the results obtained, the NMME models were tabulated according to performance for each month and the gridded time series was used to generate the gridded monthly forecast for the validation period (2004 to 2022). The deterministic skill validation metrics generally indicated that NMME models performed well in January with strong correlation coefficient values (r = 0.9) and MAPE values falling within the accurate forecasting range (1 to 10), especially over the northern half of the basin. This could be attributed to the strong association of the total monthly rainfall to the global phenomenon (Inter Tropical Convergence Zone), which usually reached its peak in Zambia during that period.

In conclusion, the study showed that there is a spatial and temporal consistency in the reliability of individual NMME models in forecasting monthly rainfall over the basin. Hence, understanding the strengths and limitations of the models used in various months is critical to properly using and interpreting the forecast information.

The following recommendations were drawn from the study,

1) The study recommends the NASAGEOS to be the best model for quantitative forecasting of monthly rainfall in (November, December and January), NCEPCFV2 (November, December and March), CANSIPV2 (January) and the CANCMI4 model for (March).

2) Due to the increasing number of Automatic Weather Stations, future researchers should integrate them into the analysis to improve the accuracy after data accumulation of 30 years and above.

3) The study would also like to encourage professionals within the Meteorology fraternity to consider integrating regional/local effects, especially in forecasting monthly rainfall at the beginning and towards the end of the rainy season, in the bid to improve the accuracy of the Global NMME models.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bi, K.F., Xie, L.X., Zhang, H.H., Cheng, X., Gu, X.T. and Tian, Q. (2022) Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast. arXiv: 2211.02556. https://doi.org/10.48550/arXiv.2211.02556

[2] Jaseena, K.U. and Kovoor, B.C. (2022) Deterministic Weather Forecasting Models Based on Intelligent Predictors: A Survey. *Journal of King Saud University—Computer and Information Sciences*, **34**, 3393-3412. https://doi.org/10.1016/j.jksuci.2020.09.009

[3] WMO-No 1246 (2020) Guidance on Operational Practices for Objective Seasonal Forecasting. WMO Secretariat.

[4] IPCC Report (2022) Climate Change 2022: Impacts, Adaptation and Vulnerability. IPCC.

[5] Dinku, T., Faniriantsoa, R., Islam, S., Nsengiyumva, G. and Grossi, A. (2022) The Climate Data Tool: Enhancing Climate Services across Africa. *Frontiers in Climate*, **3**, Article 787519. https://doi.org/10.3389/fclim.2021.787519

[6] Wang, Q., Xia, J., She, D., Zhang, X., Liu, J. and Zhang, Y. (2021) Assessment of Four Latest Long-Term Satellite-Based Precipitation Products in Capturing the Extreme Precipitation and Streamflow across a Humid Region of Southern China. *Atmospheric Research*, **257**, Article ID: 105554. https://doi.org/10.1016/j.atmosres.2021.105554

[7] IPCC (2021) Climate Change Widespread, Rapid, and Intensifying. IPCC.

[8] Yuan, F., Zhang, L., Soe, K.M.W., Ren, L., Zhao, C., Zhu, Y., *et al.* (2019) Applications of TRMM- and GPM-Era Multiple-Satellite Precipitation Products for Flood Simulations at Sub-Daily Scales in a Sparsely Gauged Watershed in Myanmar. *Remote Sensing*, **11**, Article 140. https://doi.org/10.3390/rs11020140

[9] Water Resource Management Authority (2022) WARMA Strategic Plan 2022 to 2026. WARMA.

[10] Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Ben Taieb, S., *et al.* (2022) Forecasting: Theory and Practice. *International Journal of Forecasting*, **38**, 705-871. https://doi.org/10.1016/j.ijforecast.2021.11.001

[11] Hunziker, S., Brönnimann, S., Calle, J., Moreno, I., Andrade, M., Ticona, L., *et al.* (2018) Effects of Undetected Data Quality Issues on Climatological Analyses. *Climate of the Past*, **14**, 1-20. https://doi.org/10.5194/cp-14-1-2018

[12] Faiz, M.A., Liu, D., Fu, Q., Li, M., Baig, F., Tahir, A.A., *et al.* (2018) Performance Evaluation of Hydrological Models Using Ensemble of General Circulation Models in the Northeastern China. *Journal of Hydrology*, **565**, 599-613. https://doi.org/10.1016/j.jhydrol.2018.08.057

[13] Le Roy, B., Lemonsu, A. and Schoetter, R. (2021) A Statistical-Dynamical Downscaling Methodology for the Urban Heat Island Applied to the EURO-CORDEX Ensem-

ble. *Climate Dynamics*, **56**, 2487-2508. https://doi.org/10.1007/s00382-020-05600-z

[14] Trzaska, S. and Schnarr, E. (2014) A Review of Downscaling Methods for Climate Change. United States Agency for International Development by Tetra Tech ARD, 1-42.

[15] Qu, B., Fang, Y. and Li, L. (2022) Analysis of Statistical Post-Processing Methods for Multi-Model Ensemble Runoff Forecasts in Flood Seasons. *IOP Conference Series: Earth and Environmental Science*, **1087**, Article ID: 012052. https://doi.org/10.1088/1755-1315/1087/1/012052

[16] Panofsky, H.A. and Brier, G.W. (1958) Some Applications of Statistics to Meteorology. Mineral Industries Extension Services, Penn State University.

[17] Yan, G., Liu, Y. and Chen, X. (2018) Evaluating Satellite-Based Precipitation Products in Monitoring Drought Events in Southwest China. *International Journal of Remote Sensing*, **39**, 3186-3214. https://doi.org/10.1080/01431161.2018.1433892

[18] JWGFVR Joint Working Group on Forecast Verification Research (2017) Types of Forecasts and Verifications. CAWCR.

[19] WMO (2018) WMO Guidance on Verification of Operational Seasonal Climate Forecasts. WMO Secretariat.

[20] Gutierrez, J.M., San-Martin, D., Cofino, A.S., Herrera, S., Manzanas, R.G. and Frias, M.D. (2011) User Guide of the ENSEMBLES Downscaling Portal. Santander Meteorology Group.

[21] Hodson, T. (2022) Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not 1U.S. Geological Survey Central Midwest Water Science Center.

[22] Kim, S. and Kim, H. (2016) A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts. *International Journal of Forecasting*, **32**, 669-679.

[23] Meloun, M. and Militký, J. (2011) Statistical Analysis of Multivariate Data. In: Meloun, M. and Militký, J., Eds., *Statistical Data Analysis*, Elsevier, 151-403. https://doi.org/10.1533/9780857097200.151

[24] Cannizzaro, D., Aliberti, A., Bottaccioli, L., Macii, E., Acquaviva, A. and Patti, E. (2021) Solar Radiation Forecasting Based on Convolutional Neural Network and Ensemble Learning. *Expert Systems with Applications*, **181**, Article ID: 115167. https://doi.org/10.1016/j.eswa.2021.115167

[25] Montaño Moreno, J., Palmer Pol, A., Sesé Abad, A. and Cajal Blasco, B. (2013) Using the R-MAPE Index as a Resistant Measure of Forecast Accuracy. *Psicothema*, **4**, 500-506. https://doi.org/10.7334/psicothema2013.23

[26] Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., *et al.* (2021) A Deep Learning Method for Bias Correction of ECMWF 24-240 H Forecasts. *Advances in Atmospheric Sciences*, **38**, 1444-1459. https://doi.org/10.1007/s00376-021-0215-y

[27] WMO No 1203 (2017) WMO Guidelines on the Calculation of Climate Normals. WMO Secretariat.

[28] Wang, B., Zheng, L., Liu, D.L., Ji, F., Clark, A. and Yu, Q. (2018) Using Multi-Model Ensembles of CMIP5 Global Climate Models to Reproduce Observed Monthly Rainfall and Temperature with Machine Learning Methods in Australia. *International Journal of Climatology*, **38**, 4891-4902. https://doi.org/10.1002/joc.5705

[29] WMO (2021) WMO Guidelines on the Quality Control of Surface Climatological Data. WMO Secretariat.

[30] Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-

Osorio, J.D. and Thinh, N.X. (2018) Temporal and Spatial Evaluation of Satellite Rainfall Estimates over Different Regions in Latin-America. *Atmospheric Research*, **213**, 34-50. https://doi.org/10.1016/j.atmosres.2018.05.011

[31] Libanda, B., Mie, Z., Nyasa, L. and Chilekana, N. (2020) Deciphering the Performance of Satellite-Based Daily Rainfall Products over Zambia. *Acta Geophysica*, **68**, 903-919. https://doi.org/10.1007/s11600-020-00429-w

[32] Omondi, C.K. (2017) Assessment of Bias Corrected Satellite Rainfall Products for Stream Flow Simulation: A Top Model Application in the Kabompo River Basin Zambia. University of Twente.

[33] Saha, U., Singh, T., Sharma, P., Gupta, M.D. and Prasad, V.S. (2020) Deciphering the Extreme Rainfall Scenario over Indian Landmass Using Satellite Observations, Reanalysis and Model Forecast: Case Studies. *Atmospheric Research*, **240**, Article ID: 104943. https://doi.org/10.1016/j.atmosres.2020.104943

[34] Schober, P., Boer, C. and Schwarte, L.A. (2018) Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, **126**, 1763-1768. https://doi.org/10.1213/ane.0000000000002864

[35] Hagedorn, R., Doblas-Reyes, F.J. and Palmer, T.N. (2005) The Rationale behind the Success of Multi-Model Ensembles in Seasonal Forecasting—I. Basic Concept. *Tellus A*, **57**, 219-233. https://doi.org/10.1111/j.1600-0870.2005.00103.x

[36] Ahmed, K., Sachindra, D.A., Shahid, S., Iqbal, Z., Nawaz, N. and Khan, N. (2020) Multi-Model Ensemble Predictions of Precipitation and Temperature Using Machine Learning Algorithms. *Atmospheric Research*, **236**, Article ID: 104806. https://doi.org/10.1016/j.atmosres.2019.104806

Appendices

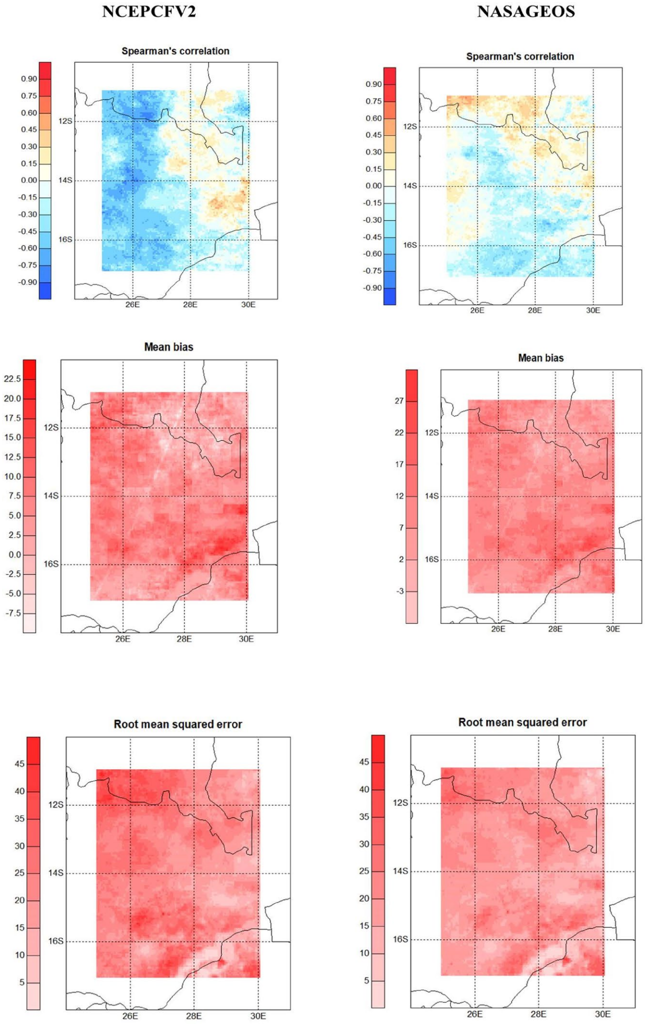


Figure A1. Gridded retroactive validation metrics for the months of October 2004 to 2022.

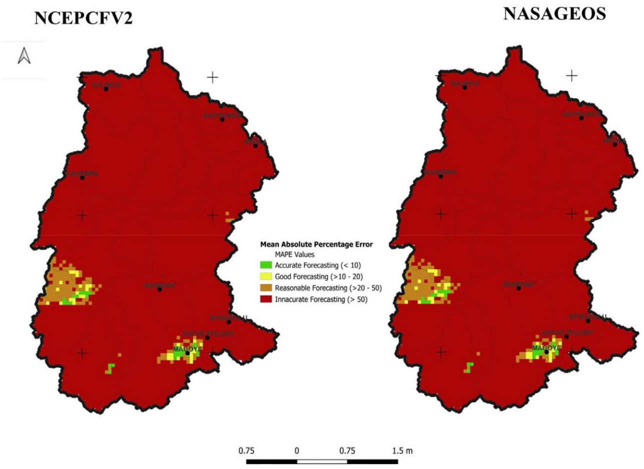


Figure A2. Mean absolute percentage error results for October 2004 to 2022.

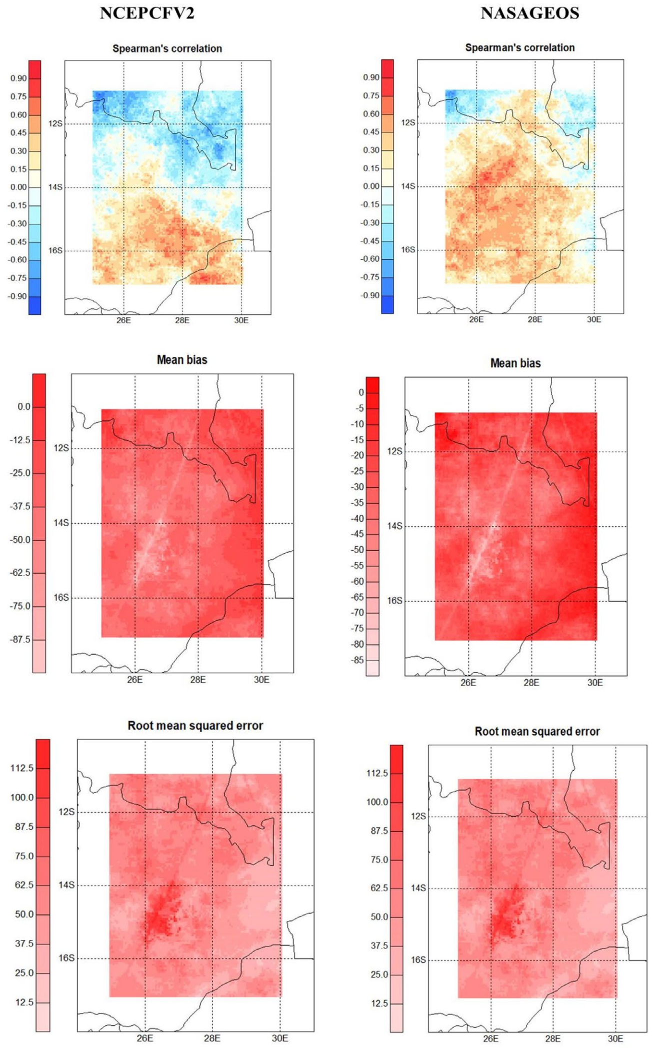


Figure A3. Gridded retroactive validation metrics for the months of November 2004 to 2022.

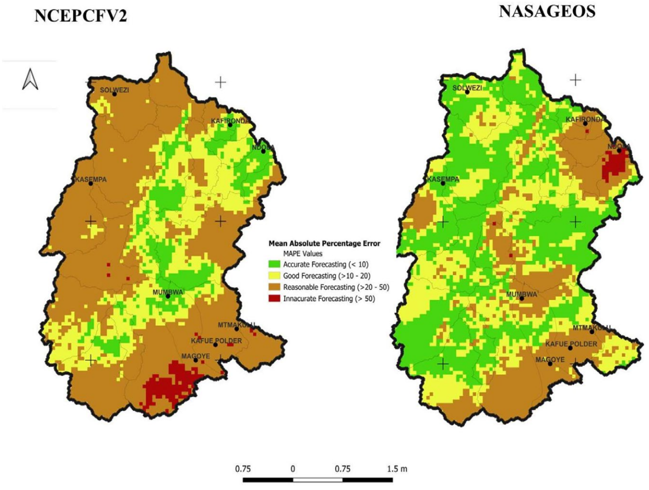


Figure A4. Mean absolute percentage error results for November 2004 to 2022.

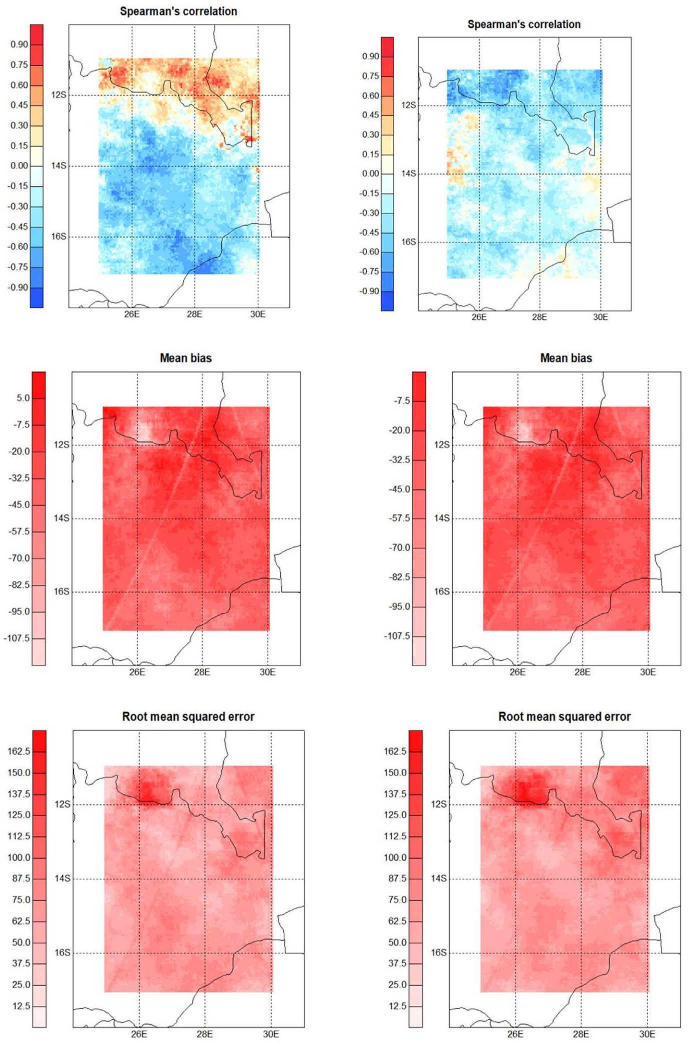


Figure A5. Gridded retroactive validation metrics for the months of December 2004 to 2022.

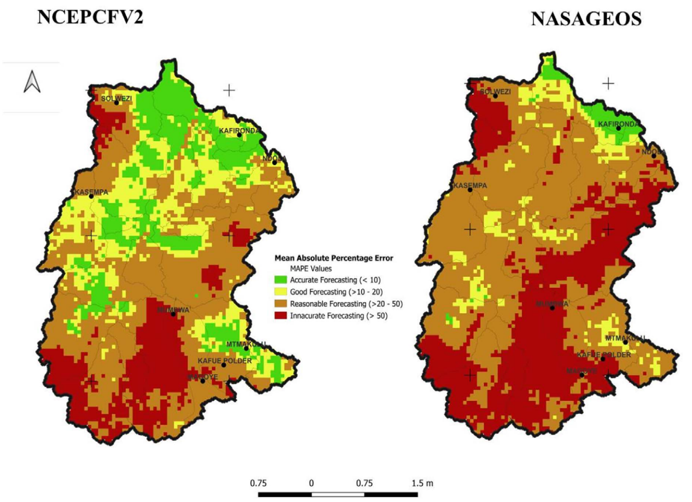


Figure A6. Mean absolute percentage error results for December 2004 to 2022.

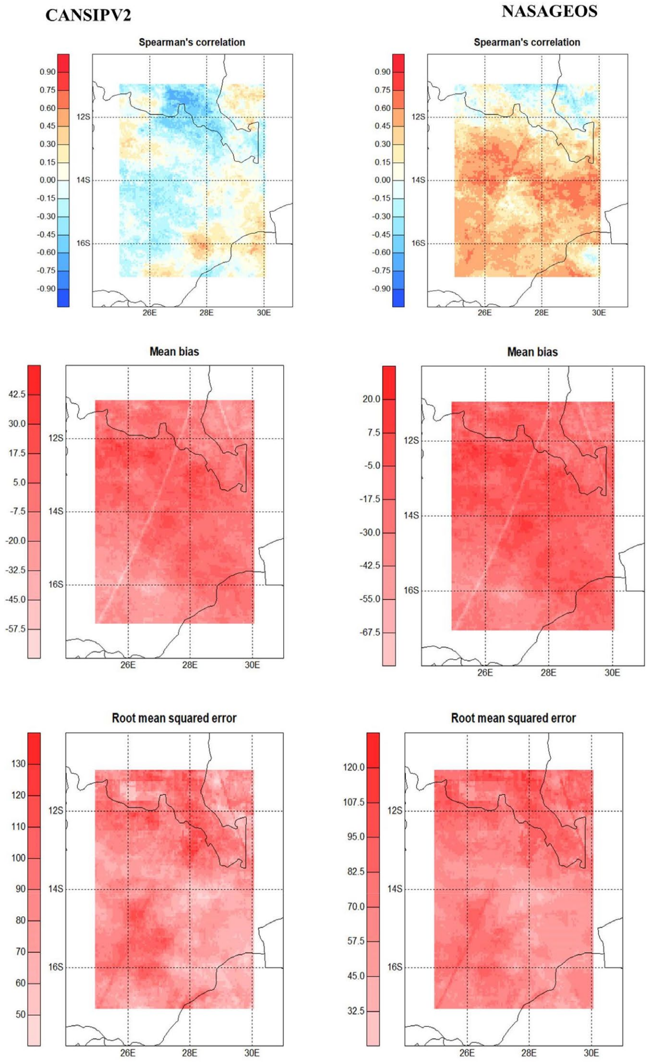


Figure A7. Gridded retroactive validation metrics for the months of January 2004 to 2022.

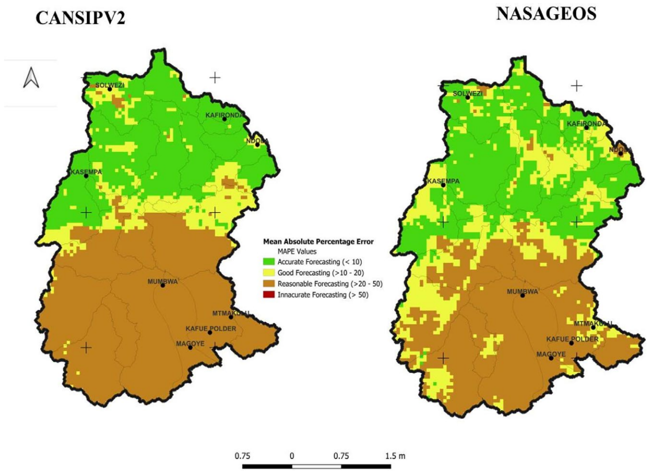


Figure A8. Mean absolute percentage error results for January 2004 to 2022.

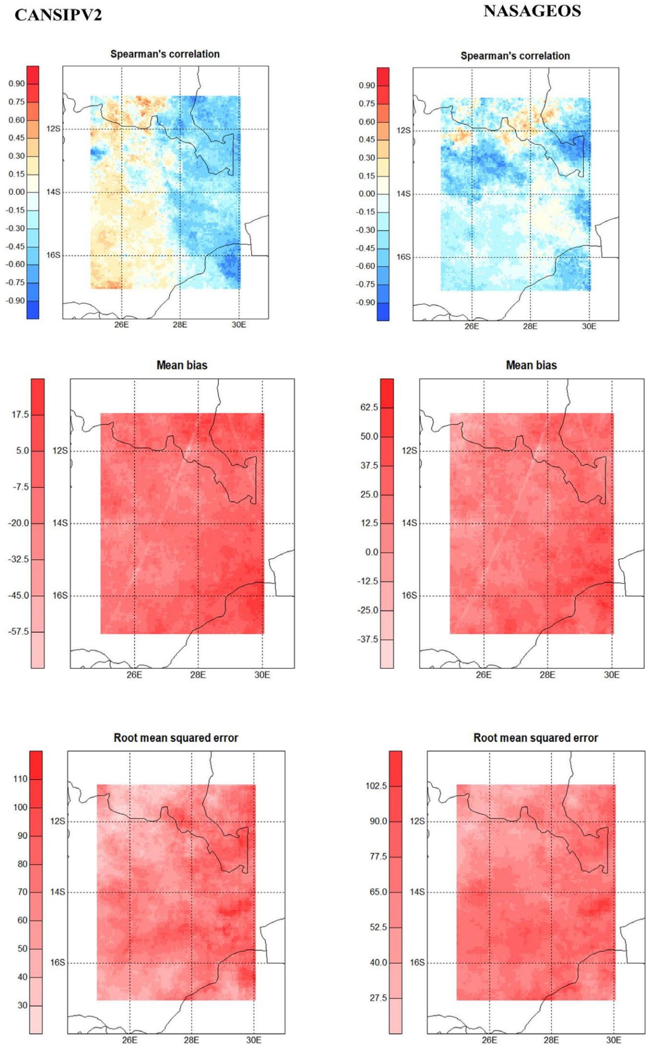


Figure A9. Gridded retroactive validation metrics for the months of February 2004 to 2022.

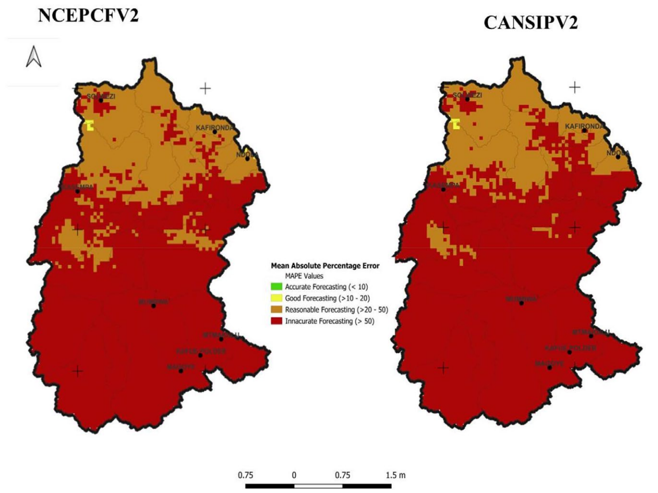


Figure A10. Mean absolute percentage error results for February 2004 to 2022.

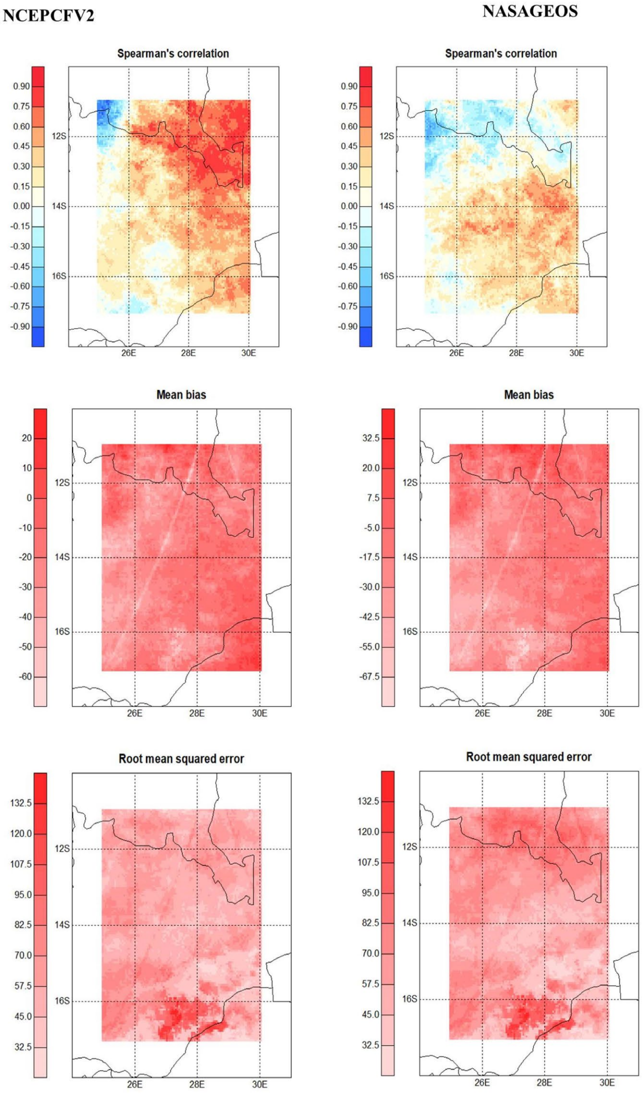


Figure A11. Gridded retroactive validation metrics for the months of March 2004 to 2022.

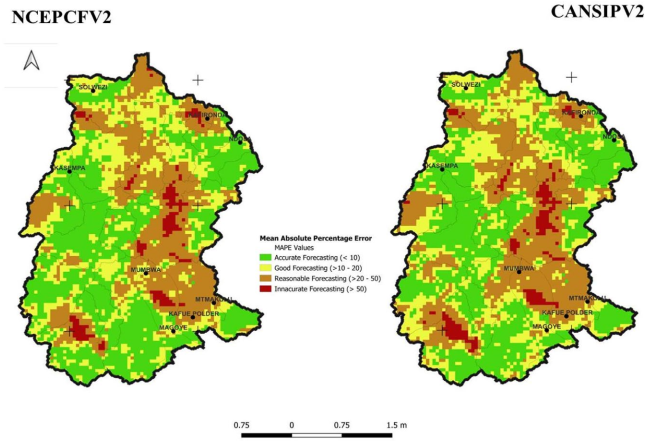


Figure A12. Mean absolute percentage error results for March 2004 to 2022.

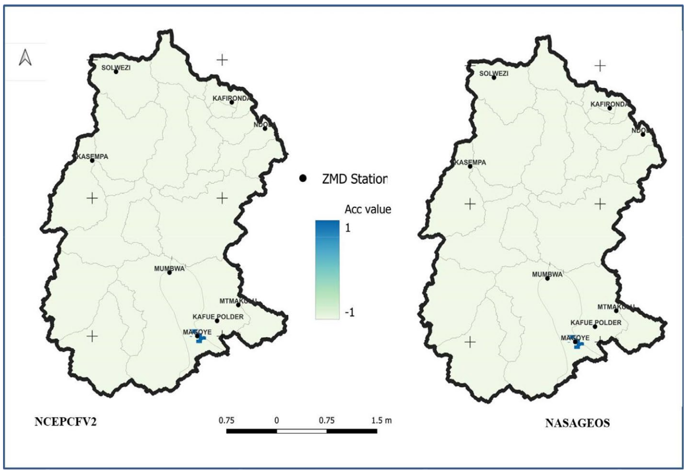


Figure A13. Individual model mean anomaly correlation coefficient values (October 2004 to 2022).

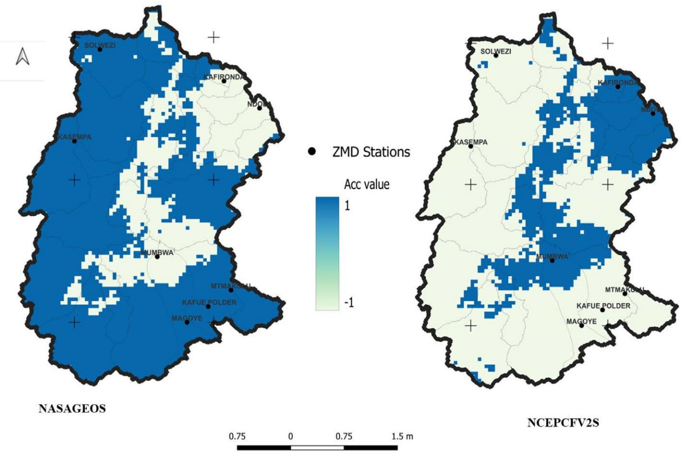


Figure A14. Individual model mean anomaly correlation coefficient values (November 2004 to 2022).

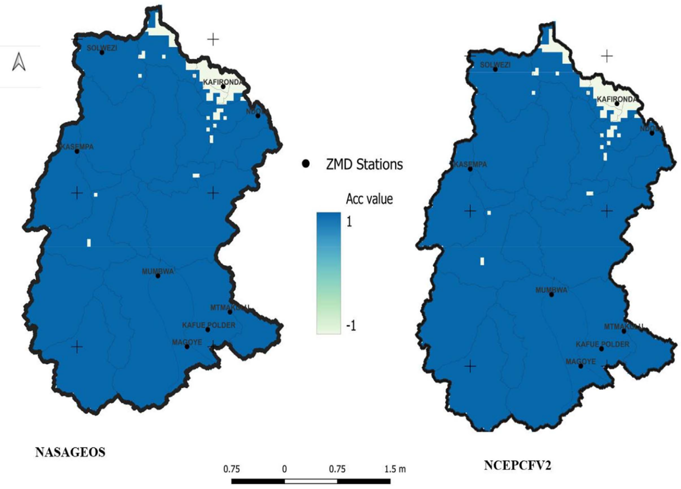
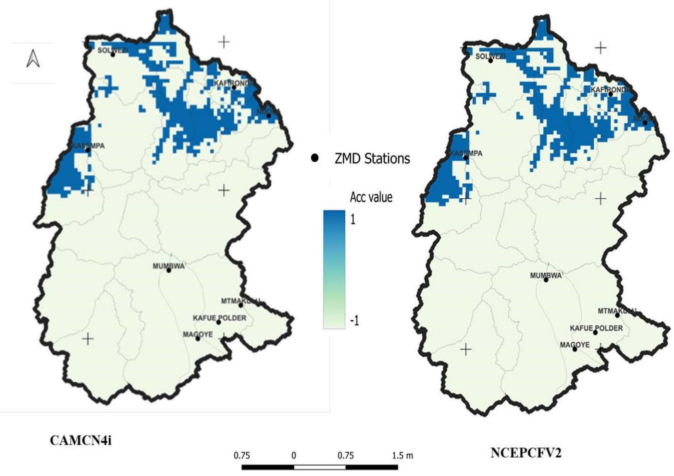


Figure A15. Individual model mean anomaly correlation coefficient values (December 2004 to 2022).



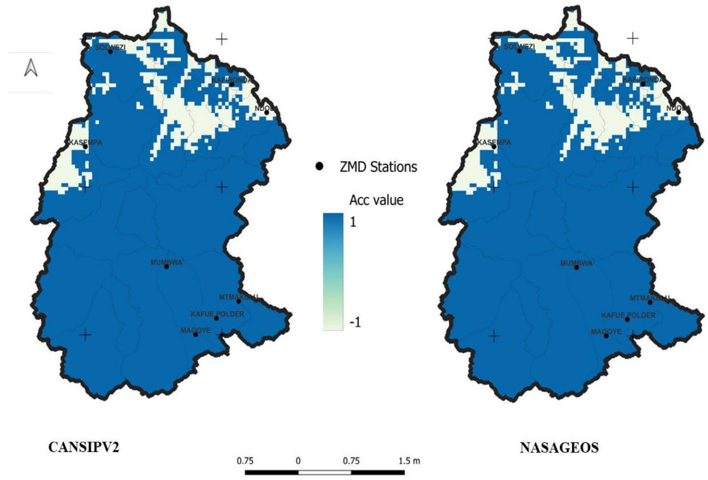


Figure A16. Individual model mean anomaly correlation coefficient values (January 2004 to 2022).

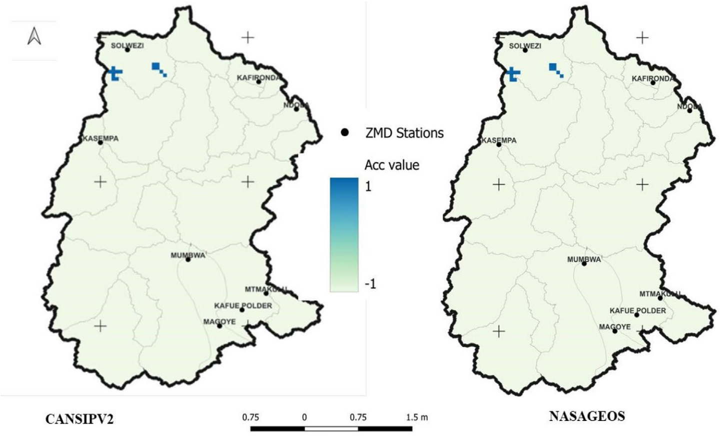


Figure A17. Individual model mean anomaly correlation coefficient values (February 2004 to 2022).

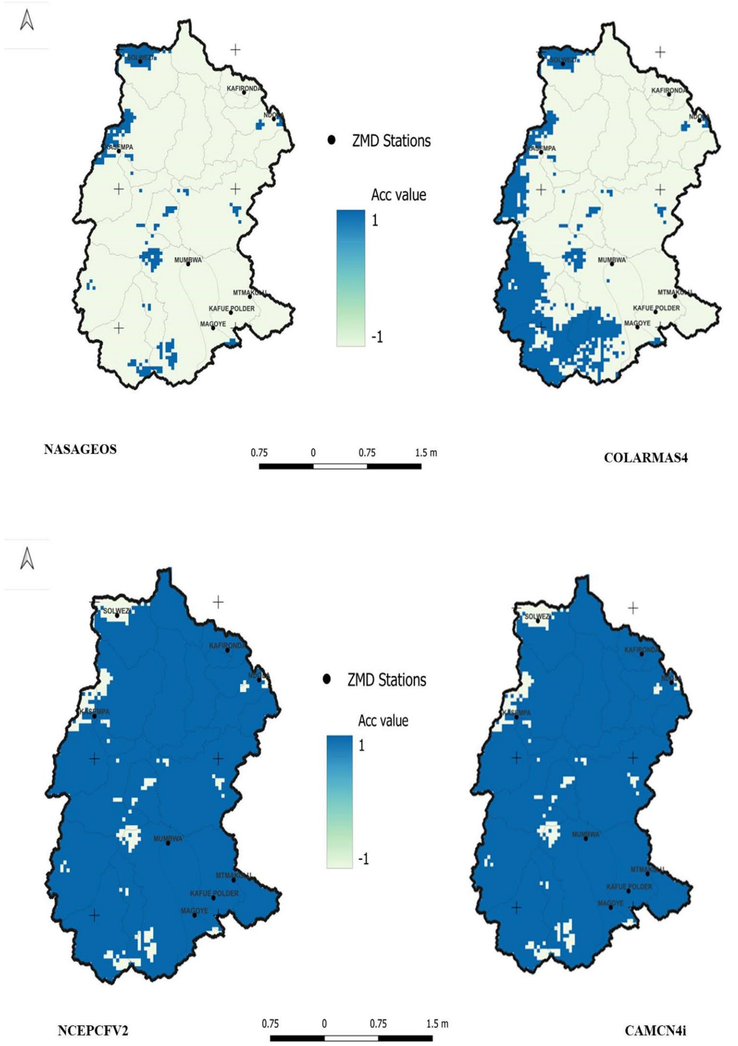


Figure A18. Individual model mean anomaly correlation coefficient values (March 2004 to 2022).