Theoretical and Applied Climatology (2026) 157:204
https://doi.org/10.1007/s00704-026-06091-4

RESEARCH

Enhancing flood simulation in data-sparse Niger central hydrological area river basin in Nigeria using machine learning-based data fusion

Habeeb Oladimeji Ganiyu[1,2] · Wan Zurina Wan Jaafar[1] · Faridah Othman[1] · Cia Yik Ng[1]

Received: 19 August 2025 / Accepted: 27 January 2026
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2026

Abstract

Flooding remains a persistent and devastating natural hazard in Nigeria. Accurate precipitation data are critical for hydrological simulation of floods. However, many river basins in Nigeria lack adequate ground-based meteorological data. This has created a need for alternative rainfall data sources, with satellite-based precipitation products (SPPs) offering promising solutions. However, SPP estimates often exhibit biases and uncertainties, particularly in complex tropical environments. Therefore, enhancing SPP accuracy through data fusion is essential for reliable hydrological modelling. This study enhances flood event simulation in a data-sparse Niger Central Hydrological Area River Basin in Nigeria by employing Extreme Gradient Boosting (XGB), Random Forest (RF), and Long Short-Term Memory (LSTM) models to fuse daily downscaled PERSIANN-CDR with observed rainfall data from 2013 to 2022. The RF-fused PERSIANN-CDR exhibited higher accuracy than the XGB and LSTM models. It improved the downscaled PERSIANN-CDR mean correlation coefficient (R) by 113.48% and reduced the mean MAE, RMSE, and Bias by 29.90%, 26.76%, and 75.71%, respectively. The reliability of the RF-fused PERSIANN-CDR was validated hydrologically using the HEC-HMS model. The simulated flow relative to the observed flow indicated improved NSE and R^2 values, alongside reduced RSR values during the calibration and validation periods, compared to the individual datasets. The robustness of the RF-fused data improvement was further validated using bootstrapped 95% confidence intervals and paired Wilcoxon tests ($p<0.05$). This study presents a practical framework that demonstrates the potential of data fusion to enhance satellite precipitation estimates for flood simulations, offering a viable approach for improving flood preparedness in data-sparse regions.

1 Introduction

Floods are multifaceted phenomenon that fascinates researchers worldwide, as they strive to comprehend and investigate prevention and management strategies (Tadesse et al. 2022). According to the World Bank (2019), climate change is potentially responsible for at least 150,000 fatalities annually. This number is projected to double by 2030, with connections to floods and water-borne diseases (WHO 2021). Historical flood events in Nigeria have been extensively discussed in the literature (Magami et al. 2014; Nkwunonwo 2016; Umar and Gray 2022). The most recent, prominent, and catastrophic flood event occurred in 2022. This adversely affected 95% of Nigeria's states, resulting in an unprecedented displacement of over 1.4 million people, 662 deaths, 2,400 injuries, 82,035 houses damaged, 332,327 hectares of farmland destroyed, and estimated losses and damages worth $6.68 billion (NEMA 2022). These consequences from floods over the years have retarded Nigeria's progress in achieving the Sustainable Development Goals (SDGs). This includes maintaining sustainable cities and communities (SDG 11), climate action (SDG 13), and clean water and sanitation (SDG 6). The Niger Central Hydrological Area (NCHA), a basin encompassing the major river in Nigeria (River Niger), has experienced multiple floods annually. Flood management employs two approaches, which

✉ Wan Zurina Wan Jaafar
wzurina@um.edu.my

✉ Faridah Othman
faridahothman@um.edu.my

Habeeb Oladimeji Ganiyu
oladimejihabeeb88@gmail.com

Cia Yik Ng
ng.ciayik@gmail.com

[1] Department of Civil Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia

[2] Department of Civil Engineering, Faculty of Engineering and Technology, Kwara State University, Malete 241104, Nigeria

Published online: 12 March 2026

are structural and nonstructural. The former is designed to mitigate flood risks by regulating water flow outside and within urban areas. However, these structures may be overwhelmed by incidents that surpass their intended capacities, and they are expensive to construct. Conversely, nonstructural approaches focus on protecting individuals from flooding through enhanced urban planning and management (Jha et al. 2012). These measures are generally more economical to implement (Haque et al. 2022).

To develop effective non-structural measures for flood hazard management, it is crucial to obtain accurate rainfall measurements, which enhances the reliability of hydrological modelling (Yuan et al. 2017). This task is particularly challenging in regions with limited data availability (Kittel et al. 2020), particularly in Nigeria, which has the problem of data sparsity (Salami et al. 2025). Nigeria operates 87 rain gauge stations across its approximately 923,300 km² land area, equating to one gauge station per 10,600 km² (Ngene and Obianigwe 2018). To achieve optimal coverage and comply with the World Meteorological Organization's standards for Tropical, Mediterranean, and Temperate regions, a minimum of 1,057 rain gauges are required (Ngene and Obianigwe 2018). Therefore, SPPs offer an alternative source for rainfall estimates at various spatio-temporal resolutions for hydrological simulations (Goshime et al. 2019; Min et al. 2020; Worqlul et al. 2017). Recent global studies have evaluated the effectiveness of SPPs by comparing them with observed rainfall data across various hydrological models in data-scarce regions. These include the Porijogi catchment in Estonia, Northern Europe (Moges et al. 2022), Huai River Basin in China (Khan et al. 2022), Bosten Lake Basin in Northwest China (Peng et al. 2021), and Chitral River Basin in Pakistan (Usman et al. 2022). In Africa, case studies encompass the Dhidhessa River Basin in Ethiopia (Wedajo et al. 2021), Upper Nan River Basin in Northern Thailand (Gunathilake et al. 2020), Souss River Basin in Central-Western Morocco (Bouizrou et al. 2023), Ouémé River Basin in Benin, West Africa (Bodjrènou et al. 2025), selected basins from West and Northern Central Africa (Kouakou et al. 2023), Nasia River Basin in Ghana (Darko et al. 2021), Mono River catchment in Togo and Benin (Houngu' e et al. 2021), Gambia River Basin in West Africa (Faty et al. 2023), and Niger Basin in Nigeria (Oyerinde et al. 2017). The use of SPPs for hydrological modelling in Nigeria remains limited. Overall, the reviewed studies indicate that simulated discharge from SPPs often fails to accurately reflect observed discharge when compared to observed rainfall data, particularly on a daily timescale, resulting in lower NSE values. The low accuracy of SPPs in the direct application of hydrological modelling is subject to inherent errors and uncertainties due to indirect estimation, sampling uncertainty, and retrieval algorithms (Beck et al. 2019; Gebregiorgis et al. 2018).

The enhancement of SPPs for hydrological applications involves integrating (fusing) SPPs with observed rainfall data, which serves as a promising approach by leveraging the strengths and weaknesses of both datasets. Previous studies have increasingly employed machine learning (ML) models to fuse SPPs with observed rainfall data. This approach leverages their capacity to exploit nonlinear relationships from different data sources, thereby enhancing the accuracy of satellite precipitation estimates. The Random Forest (RF) algorithm has been applied to fuse multiple sources of satellite precipitation and auxiliary data with observed rainfall data in South Korea (Nguyen et al. 2021), Saudi Arabia (Assiri and Qureshi 2022), Jiangsu, Zhejiang, and Shanghai in China (Fan et al. 2021), Sichuan province in China (Li et al. 2023), and the Melkakuntire catchment in Ethiopia (Mekonnen et al. 2022). Long Short-Term Memory (LSTM) has been employed to integrate IMERG with observed rainfall on the east coast of Peninsular Malaysia (Toh et al. 2023). Furthermore, Wu et al. (2020) proposed CNN-LSTM data fusion model using TRMM 3B42 V7 and thermal infrared images, combined with rain gauge data in China. Sadeghi et al. (2020) applied U-Net convolutional neural network to fuse PERSIANN-CCS and geographical information. RF, LightGBM, Extreme Gradient Boosting (XGB), Multiple Linear Regression (MLR), and cokriging have been applied to fuse multi-source precipitation and auxiliary data with observed rainfall data in the Northern Guangdong Province of China (Zhang et al. 2022), while RF, GBM, and XGB have been utilized in the contiguous United States (Papacharalampous et al. 2023), and gradient boosting decision tree (GBDT), XGBoost, and RF were used in China (Lei et al. 2022).

Notably, the RF, XGB, and LSTM models have demonstrated strong performance in data-scarce regions (Fan et al. 2021; Lei et al. 2022; Toh et al. 2023). The high skill of RF can be attributed to its robustness with small sample sizes and noisy, heterogeneous datasets through ensemble averaging, which reduces overfitting (Fan et al. 2021; Tyralis et al. 2019). XGB achieves high predictive accuracy and efficiently handles sparse data through gradient boosting and regularization (Lei et al. 2022; Zhang et al. 2022). Meanwhile, LSTM effectively captures temporal dependencies and preserves important nonlinear hydrological signals even when observations are limited (Toh et al. 2023). These characteristics make them more suitable than other ML models in regions with limited observational data, such as Africa and other developing countries. Although these three models show high reliability in enhancing SPPs, few studies have evaluated the effectiveness of ML-fused SPPs for hydrological modelling of flood events in Africa. In Ethiopia, Mekonnen et al. (2022)

applied RF-fused SPPs at a daily scale to improve flow simulation using the HBV-light model in the Melkakuntire catchment. Awad et al. (2025) enhanced the accuracy of discharge simulations using improved four PERSIANN family products in the Upper Blue Nile Basin (UBNB) by integrating three machine learning models (Random Forest Regression, Boosted Regression Forest, and CatBoost Regression), with Variational Mode Decomposition (VMD). Mulu et al. (2025) improved daily streamflow predictions in the Bilate River watershed by integrating simulated outputs of the GR4J conceptual hydrological model with climate input features using machine learning techniques. However, conceptual and process-based models are limited in their ability to capture complex nonlinear dynamics.

Therefore, this study addresses this knowledge gap by evaluating RF-, XGB-, and LSTM-based fusion of daily downscaled SPP (PERSIANN-CDR) with observed rainfall and assessing their efficacy in improving flood simulations using the HEC-HMS model in a data-sparse river basin in Nigeria. This demonstrates how enhanced SPPs influence extreme flood simulations, providing a robust foundation for future spatial regionalization and effective flood risk mitigation. The downscaled PERSIANN-CDR was selected based on its statistical evaluation performance compared to other SPPs considered in our previous study (Ganiyu et al. 2025). Salami et al. (2024) emphasised the selection of the most suitable SPPs for enhanced disaster preparedness and resource management across Nigerian regions. The objectives of this study were to (1) evaluate the fusion of the daily downscaled PERSIANN-CDR and observed rainfall data using three machine learning models (RF, XGB, and LSTM) across five gauging stations between 2013 and 2022 and (2) validate the effectiveness of the enhanced SPP by applying the fused PERSIANN-CDR from the best-performing fusion model and the individual datasets as model inputs in the HEC-HMS model for the simulations of three extreme flood events.

2 Data and methods

2.1 Study area

The study area is situated between 7°39'E and 9°52'E longitude and 4°36'N and 7°35'N latitude, located within a transitional climatic zone between the northern semi-arid and southern sub-humid climates. The weather patterns in this region are influenced by the seasonal movement of the Intertropical Convergence Zone, resulting in distinct rainy and dry periods. The rainy season occurs from April to October, and the dry season extends from November to March (Otache et al. 2016). The geology of the area comprises approximately 20% sedimentary rocks and 80% basement complex rocks (NIHSA 2022). The soils near the river valleys are sandy, whereas the nearby lateritic soils are heavily weathered and exhibit grey or red hues. Generally, the soil has a sandy top layer with a weakly formed clay layer beneath it (Matazu and Chioma 2014). This study covered a total area of 47,310 km², which is approximately 25% of the total Niger Central Hydrological Area (NCHA). It encompasses Lokoja and Kontokari in Kogi, Pategi in Kwara, and Agele, Bida, Gbako, Katcha, Lapai, and Lavun in Niger State. Figure 1 shows a map of the study area, with Nigeria as an inset.

2.2 Data collection and quality check

The model input data used in this study included observed rainfall data, satellite precipitation product, hydrological data, digital elevation model (DEM), soil map, and land use map. Figure 2 shows the maps of spatial data used for the hydrological modelling.

2.2.1 Observed rainfall data

The study utilized 10 years of daily observed rainfall data from 2013 to 2022, obtained from the Nigerian Meteorological Agency (NIMET), located at Oshodi, Lagos State, Nigeria. The study area included five gauging stations: Abuja, Baro, Bida, Lokoja, and Minna. All of the stations are traditional rain gauges, except for Baro, which is an Automated Weather Station (AWOS).

2.2.2 Satellite precipitation product

The Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California developed the PERSIANN-CDR, a satellite-based precipitation dataset. This dataset provides reliable, long-term, and high-resolution global rainfall information, which is essential for climate and hydrological studies, particularly those focusing on extreme-precipitation events. PERSIANN-CDR was derived from the PERSIANN algorithm utilizing GridSat-B1 infrared data. To reduce biases in the estimated precipitation, the model outputs were calibrated using the monthly Global Precipitation Climatology Project (GPCP) product (Sun et al. 2018). The daily PERSIANN-CDR data is available from https://www.ncei.noaa.gov/data/precipitation-persiann/access/, at 0.25° spatial resolution.

2.2.3 Hydrological data

Hydrological data for the Lokoja gauge station, including discharge and water levels, were obtained from the Nigerian

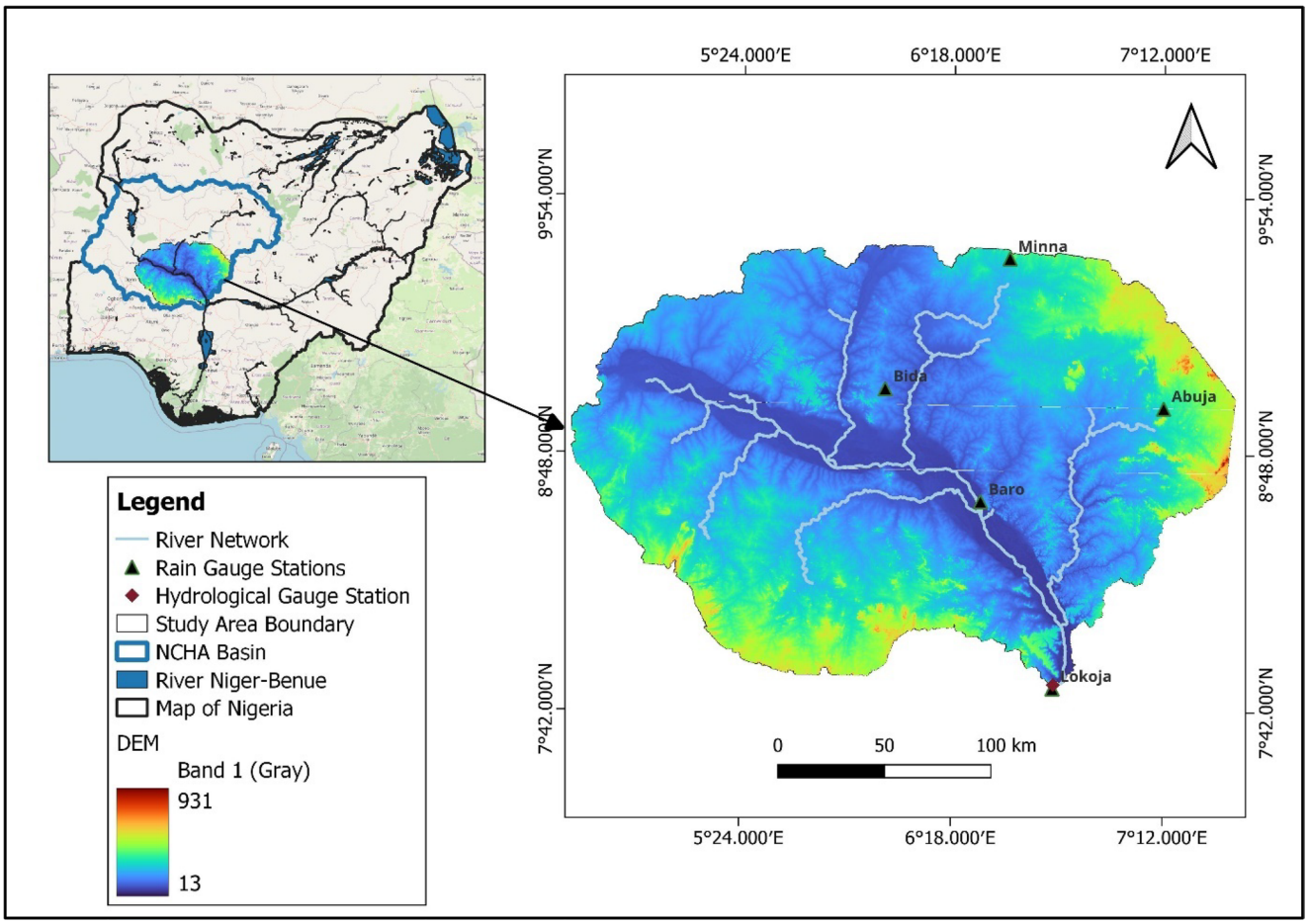


Fig. 1 Map of the study area with Nigeria as an inset

Inland Waterways (NIWA), situated at Lokoja, Kogi State. The collected daily data spanned the period from January 2013 to December 2022. This was necessary to capture the most recent extreme flood events.

2.2.4 Digital elevation model (DEM)

The DEM used to generate the elevation of the study area is the ALOS Phased Array Type L-Band Synthetic Aperture Radar (PALSAR), at a higher spatial resolution of 12.5 m. This data was downloaded from the University of Alaska at https://www.asf.alaska.edu/sardata/palsar on October 13, 2023. The elevation in the basin ranges from 13 to 931 m AMSL. The DEM was projected onto WGS 1984, Northern Hemisphere, UTM Zone 31 N, using the Quantum Geographic Information System (QGIS).

2.2.5 Soil data

Soil plays a crucial role in determining curve numbers (CN). A soil map was used to determine the soil type, texture, and hydrological soil group (HSG) of the study area. The soil types available in the study area are Lithosols, Fluvisols, Ferric Luvisols, Plinthic Luvisols, Dystric Nitosols, and Eutric Nitosols. Therefore, the most dominant soil type present in the catchment is Ferric Luvisols, characterized by sandy-clay soil texture (HSG D), covering approximately 51.93%. In contrast, sandy-clay-loam (HSG C) had the lowest percentage (8.1%). The soil map, with a resolution of 10 km developed by the Food and Agriculture Organization of the United Nations (FAO-UNESCO) at a scale of 1:5,000,000 (Fischer et al. 2008), was downloaded from https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/. The soil data integrated with LULC were further processed using the HEC-GeoHMS and ArcHydro tools, which are ArcGIS extensions, to generate a CN grid map ranging from 0 to 100.

2.2.6 Land use land cover (LULC)

LULC is the most significant factor contributing to rapid changes in surface runoff resulting from increased population growth. The LULC map provides information on the spatial extent and classification of various LULC types

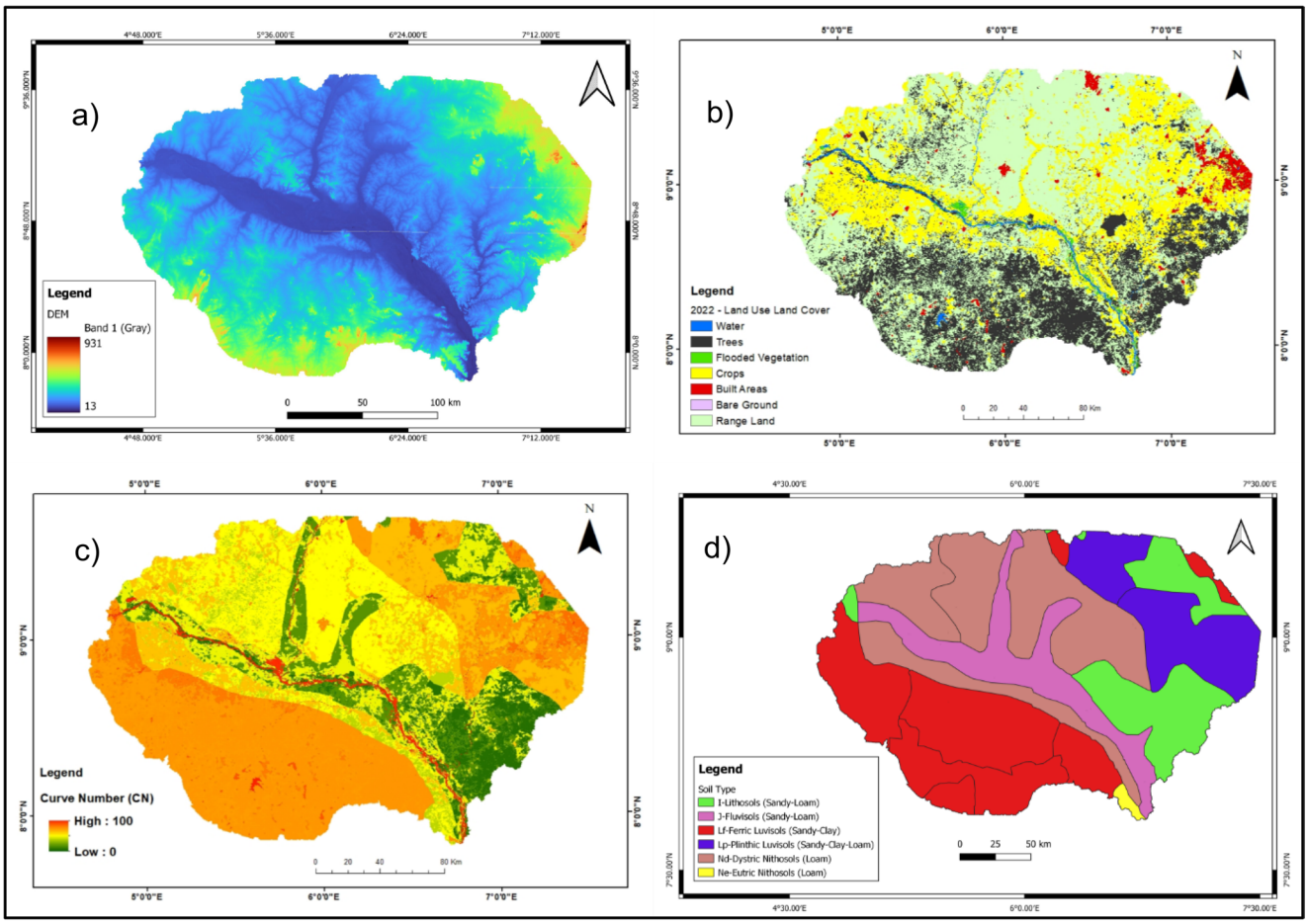


Fig. 2 Map of **a** DEM, **b** LULC, **c** CN, and **d** Dominant soil types

in the study area. The hydrological characteristics of the study area were generated by combining these data with soil data. Sentinel-2 LULC 2022 version at 10 m resolution was downloaded from https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2. LULC was categorized into water, trees, flooded vegetation, crops, built areas, bare ground, and rangeland. A larger percentage of the basin was dominated by rangelands (47.39%). This was followed by tree plantations, crop cultivation, built areas, water, flooded vegetation, and bare ground, at approximate values of 24.28%, 24.21%, 2.26%, 1.23%, 0.54%, and 0.07%, respectively.

2.2.7 Hydrological and rainfall data quality check

The data used for hydrological modelling in this study were cleaned and analyzed for reliability. The completeness of the rainfall data records was checked, and it was observed that all the in situ gauge stations had a few missing data points, except for Baro. The Abuja, Bida, Lokoja, and Minna gauge stations had missing data percentages of 0.74%, 0.33%, 0.38%, and 0.33%, respectively. However, the missing data were filled using linear interpolation in MATLAB based on data from the same season as the missing period. The homogeneity of the observed rainfall was checked using the RHtests_dlyPrcp software package, which is specifically designed to analyze the homogeneity of daily precipitation series. The software was obtained from http://etccdi.pacificclimate.org/software.shtml. RHtests_dlyPrcp offers three functionalities for detecting and rectifying artificial shifts in daily precipitation datasets without requiring reference series. The FindU, FindUD, and StepSize functions are based on the transPMFred algorithm (Wang et al. 2010). RHtests_dlyPrcp was launched at the R prompt with the input of FindU.dlyPrcp, FindUD.dlyPrcp, and StepSize.dlyPrcp for the daily precipitation data. The purpose was to detect and adjust shifts in daily precipitation series (using a threshold of 0.99 mm), as tested in these functions. However, the analysis revealed no change points at the five gauge stations in this study area. Consequently, further testing is unnecessary, and the examined time series is considered homogeneous.

For hydrological data, relatively high discontinuity and inconsistency were observed in the discharge data compared

to the water levels. Therefore, the rating curve (stage-discharge) equation developed by Ologhadien (2018) for the Lokoja gauge station was used to generate discharge data. The rating curve, as presented in Eq. 1, transforms continuous measurements of water levels into discharges. However, discontinuous missing water level data were not considered, and an interpolation method was used to fill the gaps in the missing data.

$$Q = 171.56H^2 + 500.95H - 35.04 \quad (1)$$

Where: Q = discharge in m^3/s and H = the stage or water level in m.

2.2.8 Analysis of discharge and rainfall data

The 10-year study period (2013–2022) showed a peak daily rainfall of 156.5 mm at the Bida gauge station on July 26, 2015. A high frequency of no rainfall was recorded on most days from November to March at all gauging stations. However, there was a high magnitude of rainfall between April and October, with peaks in August and September. The lowest and highest daily discharges of 1470 m^3/s and 36118.9 m^3/s were observed on 1 May 2020 and 6 October 2022, respectively, at the Lokoja gauge station. Floods, with peak discharge, are frequently recorded between September and October in Nigeria, but with less rainfall than in other wet months. This is because the other rainy months fully saturate the earth's surface, and any additional precipitation in October serves as excess rainfall, which can result in flooding. Figure 3 shows a graph of the daily observed rainfall data at the five gauging stations and the discharge data at the Lokoja gauge station between 2013 and 2022. The mean daily discharges of 19,610, 26191.4, and 36118.9 m^3/s observed in 2018, 2019, and 2022, respectively, were higher than those in the other years. Therefore, this study considered three extreme events for hydrological simulations.

2.3 Methods

This study builds upon our previous research (Ganiyu et al. 2025), in which four daily satellite precipitation products (CHIRPS2.0, CPC-CMORPH, GPM-IMERG07, and PERSIANN-CDR) were accessed at various spatial resolutions and downscaled to a uniform spatial resolution of 0.1° for a fair comparison. The downscaled datasets were comprehensively assessed with reference to observed rainfall data at five gauging stations using scatter plots, descriptive statistics, categorical statistics, and four-component error decomposition (4CED). The study demonstrated the suitability of downscaled PERSIANN-CDR in the study area. However, it recommended data fusion approaches to improve its accuracy and reduce bias before being used as model input data for hydrological modelling.

Therefore, this study presents the enhancement of satellite precipitation estimates through data fusion of daily downscaled PERSIANN-CDR and observed rainfall data using three machine learning models (RF, LSTM, and XGB), as well as the hydrological evaluation of the enhanced precipitation datasets from the best-performing machine learning model. Figure 4 illustrates the methodological framework of the model process.

2.3.1 Data fusion processes

It has been established that SPPs are prone to errors, and it is challenging to find an accurate SPP at a daily temporal resolution (Chen et al. 2019). Therefore, enhancing SPP estimates before their application in hydrological simulations is essential. In this study, three reliable data ML models (RF, XGB, and LSTM) were applied to fuse the downscaled PERSIANN-CDR and observed rainfall data at a daily temporal resolution from 2013 to 2022 in a Python environment, leveraging the strengths and mitigating the weaknesses of the two datasets. The period selected for data fusion was limited by the availability of high-quality

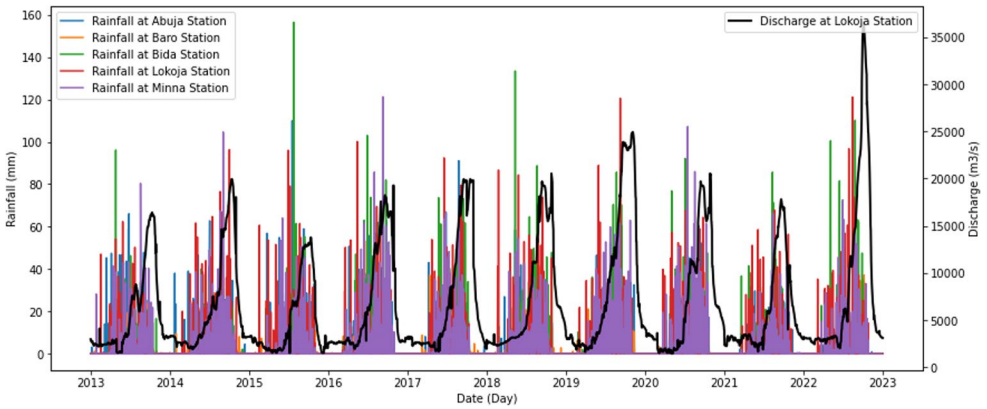


Fig. 3 Graph of daily rainfall at the five gauging stations and discharge data at Lokoja gauge station between 2013 and 2022

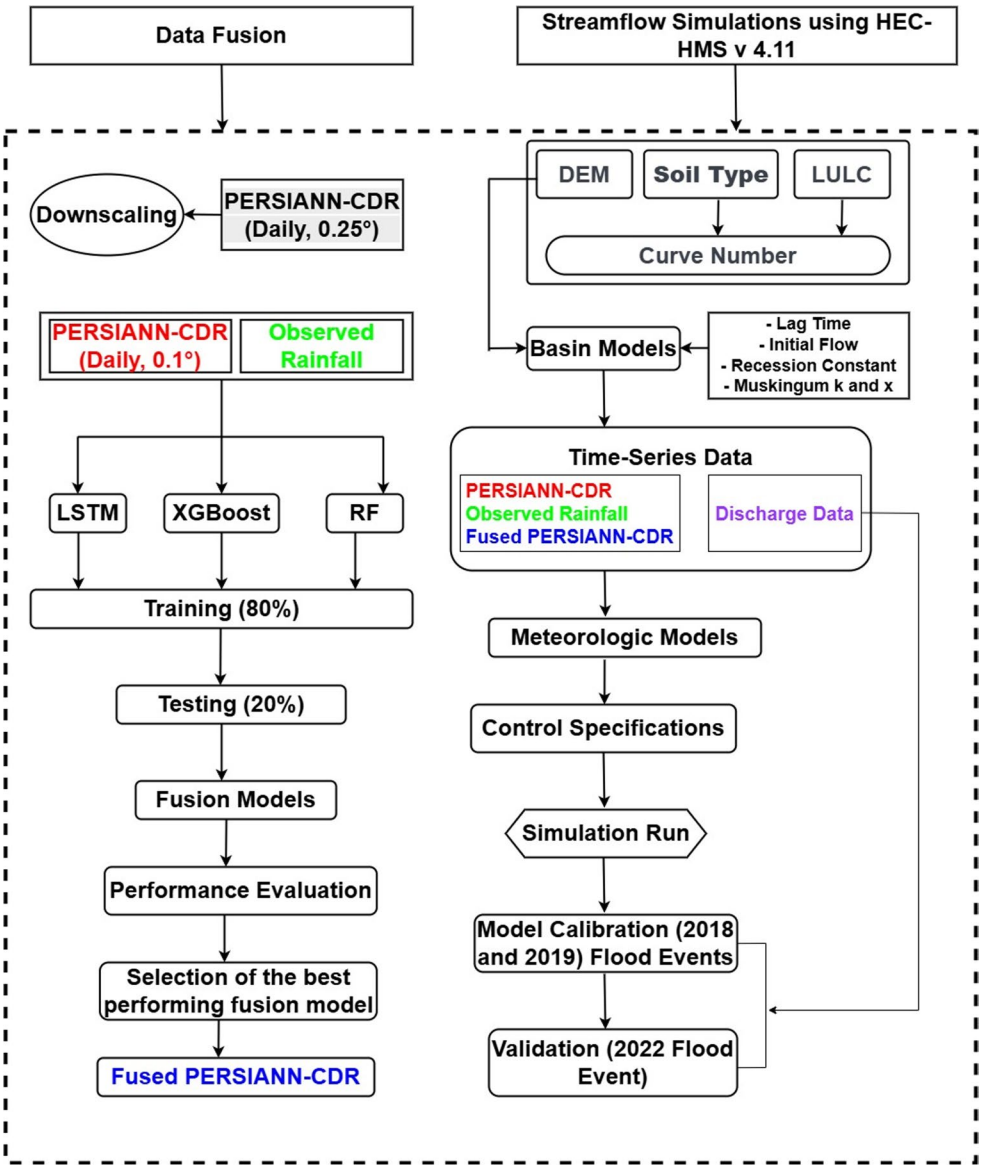


Fig. 4 Methodological framework of the study

satellite and gauge datasets. Although this timeframe is shorter, it encompasses a representative range of hydrological conditions, including both wet and dry seasons, offering a sufficient foundation for hydrological modelling and performance evaluations. The ML models were trained using downscaled PERSIANN-CDR data as the input and observed rainfall as the output. This is consistent with the findings of Shen et al. (2022) and Toh et al. (2023). A consistent modelling framework was applied across the five gauging stations, with each fusion model trained using 80% of the 10 years of data for training and the remaining 20% for testing. Furthermore, the trained model was applied to the entire data series to generate fused precipitation data for the full series, following Singh et al. (2022). The full series of fused PERSIANN-CDR was compared to the downscaled PERSIANN-CDR in reference to the observed rainfall to determine the percentage improvement at the existing gauge stations. Several trials were conducted, adjusting the percentage of data samples allocated for training and testing until the optimal allocation was achieved. For the RF model, a uniform architecture and set of hyperparameters were used across all five gauging stations to ensure comparability and consistency. The same approach was applied to the XGB and LSTM models, although each had distinct architectures and hyperparameter settings optimized for their specific algorithmic needs. This strategy enabled a fair comparison of model performance, aiming to identify the performance of the fusion models in evaluating the fused product against observed rainfall data at the existing gauge stations for subsequent hydrological assessment. The fused precipitation product (dependent variable) was generated using the downscaled PERSIANN-CDR as an independent variable

in the model. The estimated amount of precipitation based on data fusion is presented in Eq. 2.

$$P_{Fused} = f_{\theta}\ (P_{Satellite}) \quad (2)$$

Where:

P_{Fused} is the final, fused PERSIANN-CDR.
$P_{Satellite}$ represents the input from the downscaled PERSIANN-CDR.
f_{θ} represents the machine learning models (RF, XGB, and LSTM), which are functions that learn complex, nonlinear relationships between the inputs and the output.

The architectures of the machine learning models are described as follows:

1. Random forest model
The RF algorithm serves regression and classification purposes based on the decision tree model proposed by Breiman (2001). It extends the Classification and Regression Tree (CART) model, enhancing the precision and robustness of regression models. This is accomplished by constructing multiple classification regression trees and combining their forecasts to produce final predictions. RF model is effective for both classification and regression, significantly reducing overfitting (Abdo et al. 2025; Ding et al. 2025; Mao et al. 2025). Integrating the results of multiple classification regression trees using the bagging method cancels out some random errors and demonstrates good tolerance to random noise and outliers (Li et al. 2023). In this study, the regression method was applied to fuse the downscaled PERSIANN-CDR with observed rainfall data by learning the nonlinear pattern between them, thus compensating for biases in the downscaled PERSIANN-CDR. The RF model constructs an ensemble of decision trees, with each tree trained on a bootstrapped sample of the data. A random subset of features is tested, and the final prediction is the average output of all trees at each node split. The model performed fairly well at all the gauging stations with the default hyperparameters, except for the number of trees, which was adjusted to 50 for optimal results after multiple tunings. The RF algorithm can achieve optimal performance with little or no hyperparameter tuning (Tyralis et al. 2019). Equation 3 represents the estimated final prediction (Nguyen et al. 2021), and Fig. 5 illustrates the structure of the RF model.

$$y = \frac{1}{C}\sum_{c=1}^{C} T_c^*(x) \quad (3)$$

Where: y = final predicted result from the RF model, c = individual bootstrap sample, C = total number of predictions, T_n^* = individual prediction, and x = input vector for the test.

2. Extreme gradient boosting model
It is an advanced tree-based ensemble algorithm that enhances and refines the Gradient Boosting Decision Tree (GBDT) method (Niazkar et al. 2024). This approach constructs trees incrementally through feature splitting, with each newly created tree acting as a new function to address the residuals of the preceding tree.

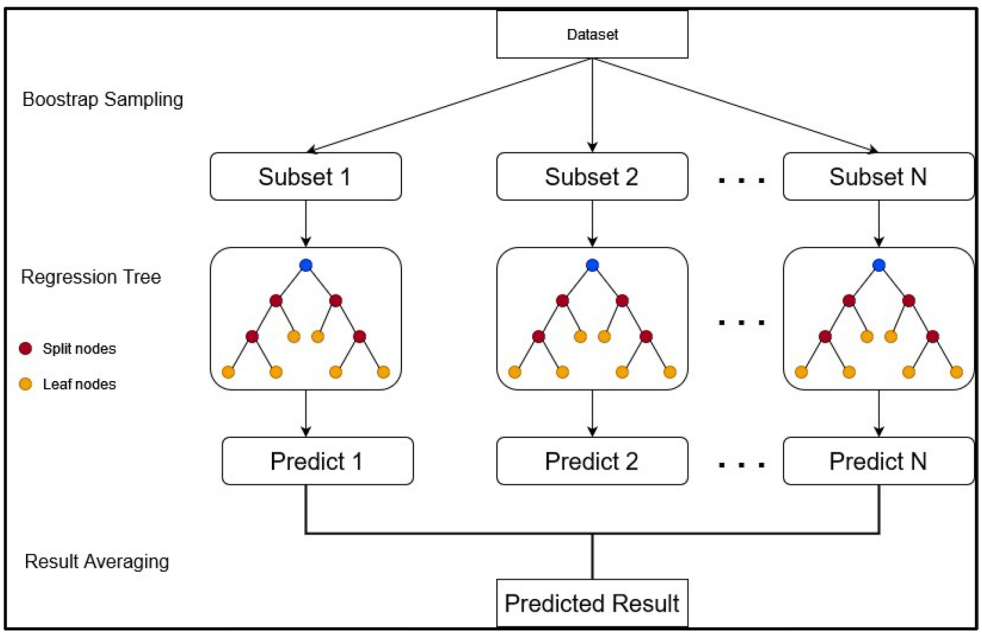


Fig. 5 Schematic diagram of RF structure. (Adapted from Yan et al. 2021)

The mathematical representation of the XGBoost model is provided in Eqs. 4–6 (Zhang et al. 2022). Figure 6 illustrates the schematic diagram of XGBoost. The optimal result was achieved with a learning rate, maximum tree depth, and number of estimators set at 0.1, 20, and 300, respectively.

For a given input x, the prediction is calculated as the sum of the outputs from multiple decision trees:

$$\widehat{y}_i = \sum_{K=1}^{K} f_k(X_i), \quad f_k \in F \tag{4}$$

Where:

$\widehat{y}_i$ = the predicted value for the i-th sample, K = the total number of trees, f_k = the prediction from the k-th decision tree, and $\mathcal{F}$ = space of all possible regression trees.

XGBoost minimizes the following regularized objective function:

$$\mathcal{L}(\varnothing) = \sum_{i=1}^{n} l(y_i, \widehat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \| w \|^2 \tag{6}$$

Where: $l(y_i, \widehat{y}_i)$ = the training error between the predicted value $\widehat{y}$ and the actual target value y_i. Ω = the regularization term, which penalizes model complexity to smooth the final learned weight and prevent overfitting, T = number of leaves in the tree, w = weights of the leaves, and γ and λ = the penalty coefficients.

3. Long Short-term memory model

It is an advanced deep learning technique that excels at extracting insights from sequential data by selecting relevant features and accounting for long-term dependencies. Its ability to model complex data distributions is facilitated by memory units that include the forget, input, and output gates. Integrating a memory unit within the hidden layer enhances the LSTM's capability to handle temporal information in long-sequence data, allowing it to identify the state characteristics of long time series data. This innovation effectively mitigates the well-known issue of exploding or vanishing gradients in recurrent neural networks (Tang et al. 2021). The LSTM was optimized using the ADAM algorithm and trained against the mean absolute error (MAE) loss, a method favored for its efficiency and rapid convergence (Toh et al. 2023). During fusion

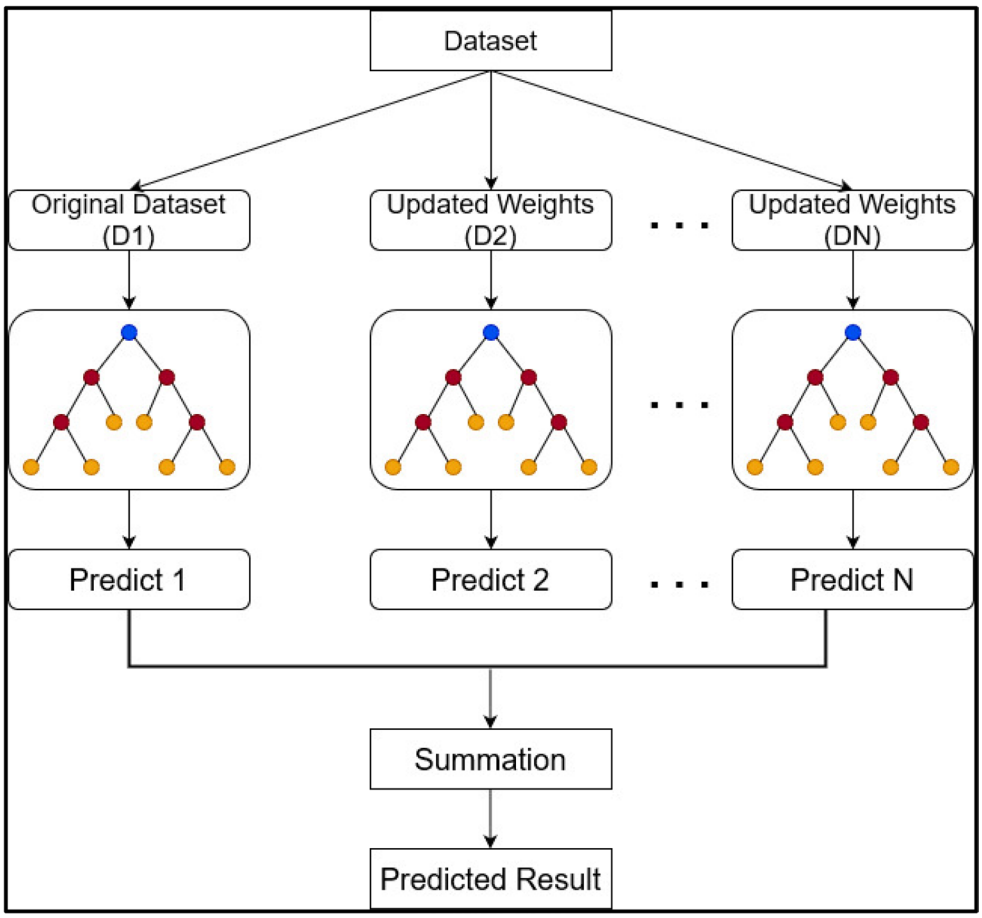


Fig. 6 Schematic diagram of XGB structure

training, the LSTM received the downscaled PERSIANN-CDR, compared its output to the observed rainfall, adjusted its weights to minimize error, and learned the relationship between the two datasets. For the fusion model, the optimal hyperparameters were determined to be one hidden layer with 47 neurons, a learning rate of 0.01, a batch size of 100, and 500 epochs after several tuning. The structure of the LSTM is depicted in Fig. 7, and the fundamental equations governing the operation of the LSTM network are presented as Eqs. 7–11 (Shen et al. 2022; Toh et al. 2023; Weber de Melo et al. 2025):

$$i_t = \sigma\left(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci} \otimes c_{t-1} + b_i\right) \quad (7)$$

$$f_t = \sigma\left(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf} \otimes c_{t-1} + b_f\right) \quad (8)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh\left(w_{xc}x_t + w_{hc}h_{t-1} + b_c\right) \quad (9)$$

$$o_t = \sigma\left(w_{xo}x_t + w_{ho}h_{t-1} + w_{co} \otimes c_{t-1} + b_o\right) \quad (10)$$

$$h_t = o_t \otimes \tanh\left(c_t\right) \quad (11)$$

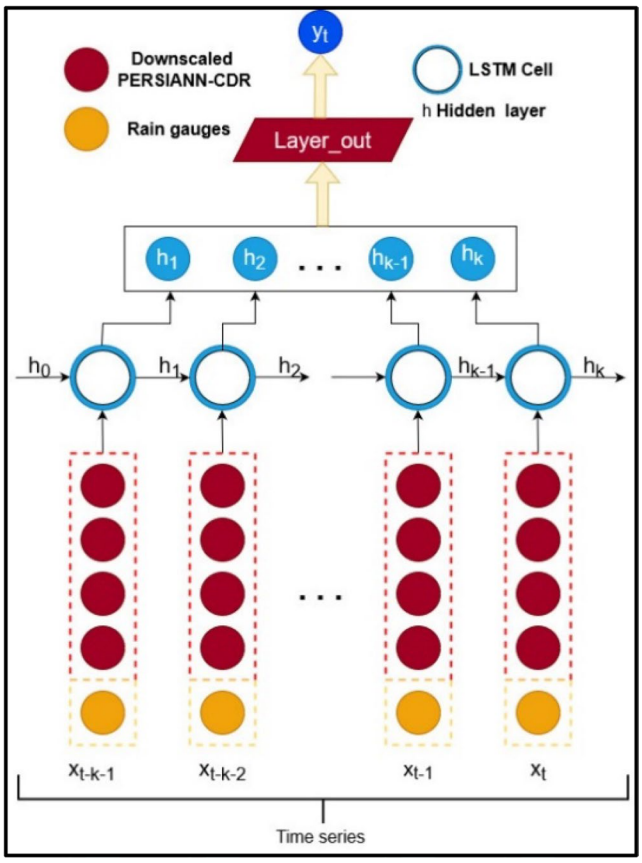


Fig. 7 Schematic diagram of LSTM structure for data fusion. (Adapted from Shen et al. 2022)

Where: i, f, o, and c = the input gate, forget gate, output gate, and memory cell, respectively, ⊗ represent element-wise multiplication, x = the input vector, w = the network's adjustable weight, b = the adjustable bias vector, h = the internal hidden state, and σ = the sigmoid activation function. Subscripts i, f, o, and c = variables corresponding to the input gate, forget gate, output gate, and memory cell, respectively, and subscript t = the time of a given variable.

2.3.2 Hydrological model setup and model calibration and validation

The United States Army Corps of Engineers developed HEC-HMS, a mathematical computer program available for download from the HEC website (http://www.hec.usace.army.mil/software/hec-hms/). HEC-HMS employs multiple techniques to model the behavior of a catchment (basin), incorporating both hydrological and hydraulic elements (Fleming 2013). HEC-HMS comprises four primary components: a basin model, a meteorologic model, control specifications, and rainfall data input, which can be in time series, paired, or gridded formats. The three main elements of the basin model are the loss, transform, and baseflow models. Within a subbasin (a portion of the catchment), each model component performs different functions related to the rainfall-runoff process. Consequently, the principal outcome of hydrological modelling is the calculation of streamflow hydrographs at the basin outlet (Jabbar et al. 2021). The HEC-HMS model has several advantages over other hydrological models. This includes the computation of various hydrological responses for watershed development (Ismail et al. 2022) and the development of a reliable model for both event-based and continuous simulations, which is also applicable in both wet and dry regions (Min et al. 2020). Therefore, this study applied HEC-HMS version 4.11 to simulate the event-based rainfall-runoff process in the basin. Moreover, this version was coupled with a GIS interface to develop basin models.

The basin was delineated into sub-basins, and its characteristics, such as the stream network, reaches, and junctions, were identified using the DEM and GIS tools embedded in the HEC-HMS model. The drainage basin was initially delineated into 22 subbasins; however, two subbasins were merged to form 21 subbasins because of the low estimated area. The 21 hydrological sub-basins have varying estimated areas, ranging from 424.24 to 6908 km², with a total area of 47,309.51 km². Hydrological components were integrated into the basin network to model the rainfall and flow correlations. Figure 8 illustrates the basin model of the study area. The HEC-HMS user manual USACE

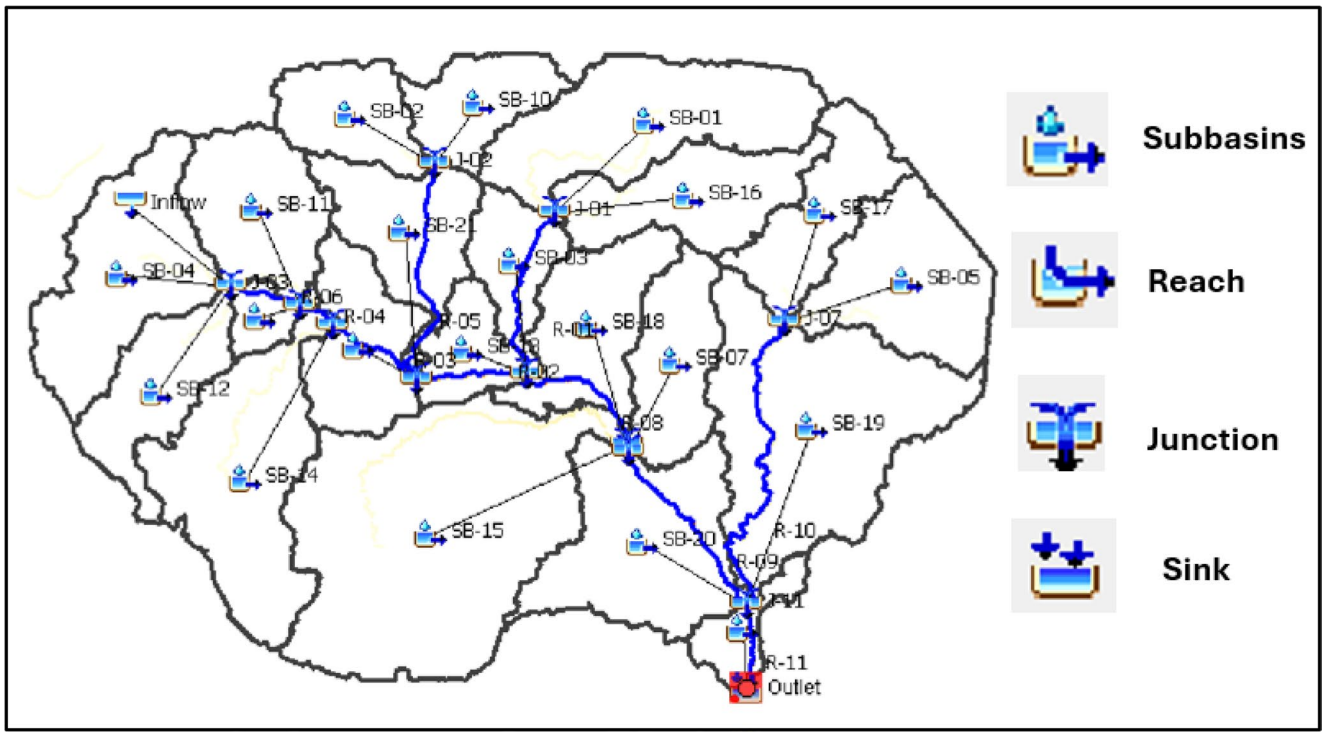


Fig. 8 Basin model of the study area

(2022), outlines the array of hydrological modelling options available in the basin model, encompassing twelve precipitation loss types, eight transform model varieties, five baseflow methodologies, and eight routing techniques. Therefore, this study employed the Recession, SCS curve number, SCS unit hydrograph, and Muskingum methods to simulate streamflow. These methods were selected based on recommendations for event-based simulations in various hydrological studies (Gilewski and Nawalany 2018; Natarajan and Radhakrishnan 2021; Tangam et al. 2024). The parameters of the basin model used in this study are presented in Table 1.

The SCS CN method was used to analyze the LULC changes and excess rainfall, as shown in Eqs. 12 and 13. The CN is a physical parameter determined for each subbasin, generated from the combination of a hydrologic soil group and the LULC of the basin, using the CN Grid tool of the HEC-GeoHMS coupled in ArcGIS.

$$R_e = \frac{(R - I_a)^2}{R - I_a + S}, \quad I_a = 0.2S \tag{12}$$

$$S = 254\left(\frac{100 - CN}{CN}\right) \tag{13}$$

Where; R_e is the excess rainfall (mm); R is the rainfall (mm); I_a is the initial abstraction (mm); S is the potential maximum retention (mm), a measurement of the capacity of a catchment to abstract and retain storm precipitation, and CN is the curve number.

The SCS-unit hydrograph method was used to transform excess rainfall into direct runoff. This method mainly depends on Tlag, with the concentration time (Tc, Eq. 14) being a factor. The initial Tlag value for each subbasin was estimated using Eq. 15.

$$Tc = \frac{l^{0.8}(S + 1)^{0.7}}{1140Y^{0.5}} \tag{14}$$

$$Tlag = 0.6Tc \tag{15}$$

Where Tc is the concentration time (hr), l is the longest flow path (ft), S = (1000/CN) −10 (in) is the potential maximum retention (in), and Y is the average slope (%).

The recession method yields exponentially decreasing base flow from single or multiple sequential events. The parameters required for generating baseflow using the recession method are the initial discharge, recession constant, and ratio to peak. The initial discharge is defined as discharge (m^3/s) or discharge per area ($m^3/s/m^2$). The discharge per unit area was selected as the initial discharge in this study. The recession constant indicates the rate of baseflow decline between storm events, and the ratio to the peak was selected as the threshold. However, all initial values of these parameters were adjusted during the calibration process. The Muskingum technique employs a simple

Table 1 Parameters generated for sub-basins in the NCHA basin

S/*N*	Subbasin	Longest Flow path (km)	CN	Slope (%)	Tc (hr)	Lag (hr)	Ia (mm)	Recession Constant
1	SB-01	187.18	72.93	6.19	44.58	26.75	18.86	0
2	SB-02	78.98	69.88	5.62	25.51	15.30	21.89	0
3	SB-03	95.73	63.24	5.19	36.84	22.11	29.53	0
4	SB-04	144.18	77.70	5.85	32.41	19.44	14.58	0
5	SB-05	119.72	80.46	9.46	20.18	12.11	12.34	0
6	SB-06	85.01	71.48	6.14	24.79	14.87	20.27	0
7	SB-07	99.62	63.59	7.01	32.45	19.47	29.09	0
8	SB-09	68.35	60.40	13.18	18.99	11.39	33.31	0
9	SB-10	87.97	69.46	5.63	28.09	16.85	22.34	0
10	SB-11	118.67	70.09	5.78	34.64	20.79	21.68	0
11	SB-12	138.15	82.72	6.60	25.20	15.12	10.61	0
12	SB-13	65.58	63.03	5.48	26.65	15.99	29.80	0
13	SB-14	203.43	83.42	7.92	30.62	18.37	10.10	0
14	SB-15	274.14	82.85	7.67	40.26	24.15	10.52	0
15	SB-16	152.22	72.52	6.17	38.27	22.96	19.24	0
16	SB-17	127.61	75.61	7.42	27.77	16.66	16.39	0
17	SB-18	144.73	67.82	5.74	43.28	25.97	24.11	0
18	SB-19	206.73	59.16	11.23	51.46	30.88	35.07	0
19	SB-20	135.75	74.32	9.63	26.58	15.95	17.55	0
20	SB-21	131.98	64.06	6.85	40.60	24.36	28.50	0
21	SB-22	52.24	67.52	4.81	21.08	12.65	24.44	0

mass-conservation approach for channel flow through a stream reach. Unlike other methods, this method does not assume a specific water surface level. This approach was applied to the route flow in the downstream reaches of the Niger River. The Muskingum method utilizes two parameters: travel time, k (measured in hours), and dimensionless weight factor, x. The former represents the duration required for water to move from the upstream to downstream end of the reach. In contrast, the latter determines the comparative significance of inflow and outflow in the storage equations.

The meteorologic model component of the HEC-HMS model encompasses precipitation data for all sub-basins. A meteorologic model was developed for the three flood events assessed in this study. The model input data for simulating streamflow included observed rainfall, downscaled PERSIANN-CDR, and fused PERSIANN-CDR data at each of the five gauging stations, serving as the primary rainfall input for the HEC-HMS model. To spatially distribute these point-based precipitation data within each subbasin in HEC-HMS, the "Gage Weight" method was used. This method assigns a specific weight to each rain gauge within a subbasin. Thiessen polygons were constructed in QGIS to determine these weights. The Thiessen polygon approach divides the study area into polygons, ensuring that every point within a given polygon is closer to its associated gauge than to any other gauge point. For each subbasin, the weight of each contributing gauge was calculated as the proportion of its Thiessen polygon area within that subbasin's boundary. These calculated weights were then incorporated into the meteorologic model component of HEC-HMS for each respective subbasin. Figure 9 shows the spatial distribution of the gauge weight for each subbasin using the Thiessen Polygon.

The control specifications were used to set the start and end dates of the simulation period and specify the required time interval (in seconds, minutes, hours, or days). A daily time step was selected for this study, which aligned with the precipitation time interval for event-based simulations. The HEC-HMS model features sixteen distinct types of time series, depending on the available data for the simulation. This study focused only on precipitation gauges for model simulation and discharge gauges for calibration and validation. The precipitation gauges included time series data from Abuja, Baro, Bida, Lokoja, and Minna gauge stations, while the discharge gauge was located at the Lokoja gauge station.

The model was manually calibrated by adjusting the hydrological parameters to improve calibration results. The most sensitive parameters, including CN in the SCS CN method, lag time (Tlag) in the SCS unit hydrograph, and Muskingum K and x, were adjusted to match the simulated and observed flows in terms of the peak discharge value and shape of the hydrograph. The study employed a split-sample methodology, utilizing two flood events (6th–20th September 2018 and 29th October–12th November 2019) for model calibration, while the third event, which was more intense (27th September–11th October 2022), was used for validation for each dataset. The average of the optimal parameters derived from the two calibration events was used for validation. This approach enhances model accuracy and reduces

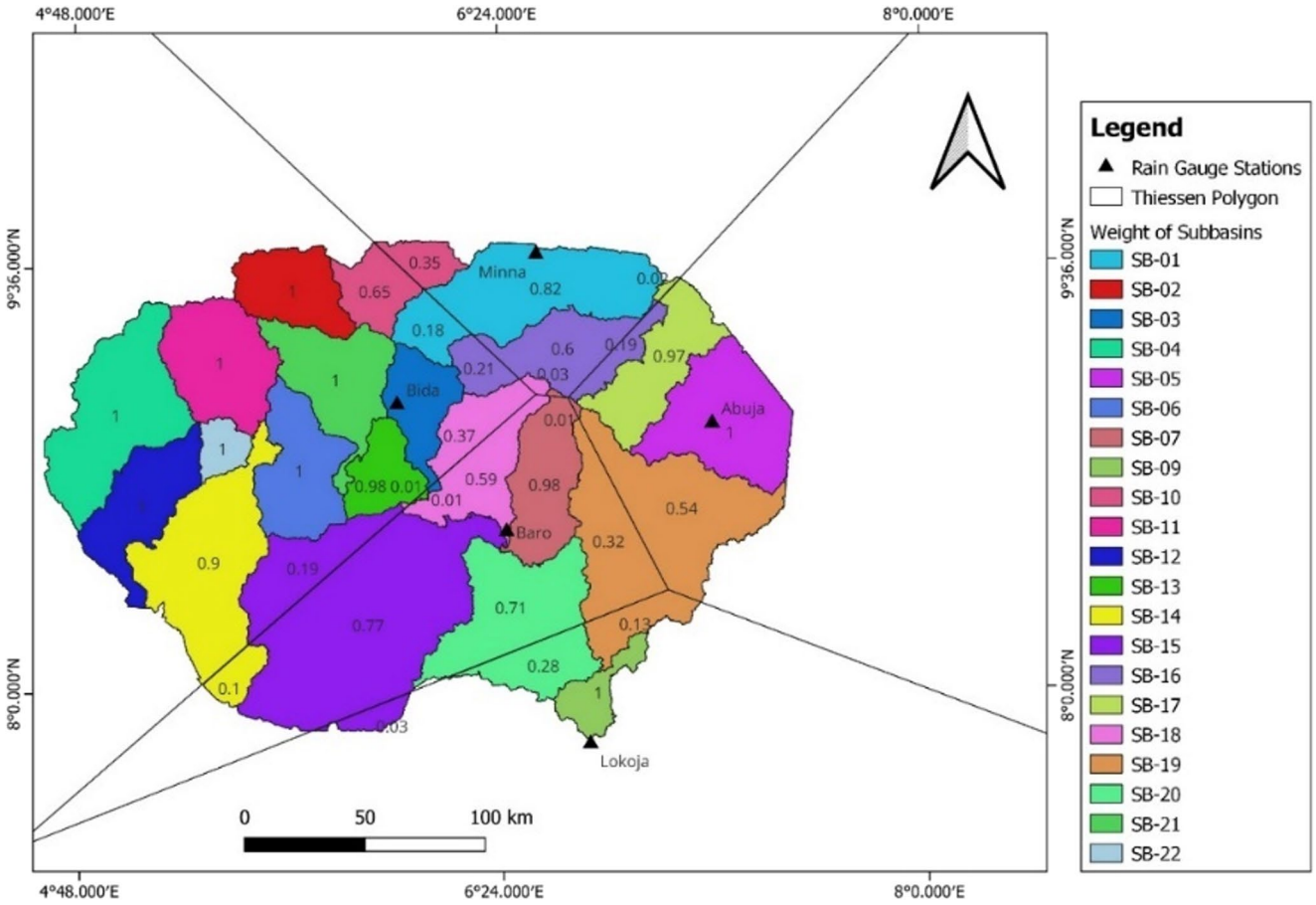


Fig. 9 Spatial distribution of gauge weight for each of the subbasins

overfitting (Reynolds et al. 2020), and it is consistent with previous studies (Dhanapala et al. 2022; Tangam et al. 2024). Table 2 presents the calibration parameters and value ranges for the HEC-HMS model.

2.3.3 Evaluation approaches

Four descriptive statistical metrics were applied to evaluate the daily downscaled PERSIANN-CDR and the fused PERSIANN-CDR against the observed rainfall at each gauge station. In this study, the metrics employed included Pearson's product-moment correlation coefficient (R) to assess the linear relationship, mean absolute error (MAE) to evaluate the average magnitude of errors without regard to direction, root mean square error (RMSE) to measure the average magnitude of error, and bias to measure the average difference between the SPPs and observed rainfall. Lower MAE, RMSE, and bias values indicate improved accuracy, whereas higher R values indicate stronger agreement. The four predictive statistics used in this study are presented in the Eqs. 16–19.

Table 2 Calibration parameters and the value range for the HEC-HMS model

Model Component	Parameter (Unit)	Values Range
Loss Model (SCS)	Initial Abstraction (mm)	0–35.07
	Curve Number, CN	47.33–95.28
	Recession Constant	1.0
Transform Model (SCS-UH)	Lag Time, Tlag (min)	341.83–2593.50
	Percent Impervious (%)	20–50
Muskingum Routing Model	Travel Time, k (hours)	15–60
	Attenuation Coefficient, x	0.1 −0.5

$$\mathrm{R} = \frac{\sum_{i=1}^{n}\left(P_{o,i} - \bar{P}_g\right)\left(P_{s,i} - \bar{P}_s\right)}{\sqrt{\left(P_{o,i} - \bar{P}_o\right)^2}\sqrt{\left(P_{s,i} - \bar{P}_s\right)^2}} \tag{16}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|P_{s,i} - P_{o,i}\right| \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{s,i} - P_{o,i})^2} \quad (18)$$

$$\mathrm{BIAS} = \sum_{i=1}^{n}(P_{s,i} - P_{o,i}) \quad (19)$$

Where: P=Precipitation, $\bar{P}$ = Average precipitation, s = satellite precipitation, o = observed rainfall, n = number of precipitation events.

The Nash–Sutcliffe efficiency (NSE), coefficient of determination (R^2), ratio of root mean square error (RMSE) to standard deviation ratio (RSR), and percent bias (PBIAS) were used to evaluate the streamflow simulation process output. The percent error in peak discharge (PEP) is a metric used to assess the accuracy of peak discharge predictions in hydrological models. It quantifies the deviation between the observed and simulated peak discharges. The percent error in the runoff volume (PEV) assesses the accuracy of the simulated peak volume compared to the observed value. The same calculation formula as PEP, but on the peak volume of the event. Equations 20–25 present the formulas for NSE, R^2, RSR, PBIAS, PEP, and PEV, respectively. Table 3 lists the performance ratings for the evaluation criteria of the HEC-HMS model.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{o,i} - Q_{s,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{o,i} - \bar{Q}_{o,i}\right)^2} \quad (20)$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(Q_{o,i} - \bar{Q}_{o,i}\right)\left(Q_{s,i} - \bar{Q}_{s,i}\right)\right]^2}{\sum_{i=1}^{n}\left(Q_{o,i} - \bar{Q}_{o,i}\right)^2 \sum_{i=1}^{n}\left(Q_{s,i} - \bar{Q}_{s,i}\right)^2} \quad (21)$$

$$\mathrm{RSR} = \frac{\sqrt{\sum_{i=1}^{n}\left(Q_{o,i} - Q_{s,i}\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(Q_{o,i} - \bar{Q}_{o,i}\right)^2}} \quad (22)$$

$$\mathrm{PBIAS} = \frac{\sum_{i=1}^{n}\left(Q_{o,i} - Q_{s,i}\right)}{\sum_{i=1}^{n}\left(Q_{o,i}\right)} \quad (23)$$

$$\mathrm{PEP} = \frac{Q_{s,i}^{peak} - Q_{o,i}^{peak}}{Q_{o,i}^{peak}} \times 100 \quad (24)$$

$$\mathrm{PEV} = \frac{V_{s,i} - V_{o,i}}{V_{o,i}} \times 100 \quad (25)$$

Where: $Q_{o,i}$ and $Q_{s,i}$ are the observed and simulated daily discharges at time step i, respectively; $\bar{Q}_{o,i}$ and $\bar{Q}_{s,i}$ are the average observed and simulated daily discharges at time step i, respectively; n is the number of observations; $Q_{o,i}^{peak}$ and $Q_{s,i}^{peak}$ are the observed and simulated peak discharges for a given flood event i, respectively; and $V_{o,i}$ and $V_{s,i}$ are the observed and simulated flood event runoff volumes for a given event i, respectively.

This study further estimated uncertainty and tested statistical significance using resampling methods applied to the existing paired observed and simulated hydrographs of the validation period for the three model inputs. This was intended to provide more confidence in the model output, supporting the superiority of RF-fused PERSIANN-CDR over downscaled PERSIANN-CDR and observed rainfall. The 95% confidence intervals for R^2, NSE, PBIAS, and RSR were obtained using block bootstrap resampling of the hydrographs with 1,000 iterations to preserve temporal dependence. The block length was chosen based on the hydrograph autocorrelation timescale. Paired comparisons between forcings (RF-fused versus downscaled PERSIANN-CDR and RF-fused versus observed rainfall) used the nonparametric Wilcoxon signed rank test and a paired bootstrap test of differences in medians with 1,000 iterations to produce p-values and 95% bootstrap confidence intervals. All model performance evaluations and result visualizations were performed in the RStudio environment using the R programming language.

3 Results

3.1 Comparison of the data fusion methods and downscaled PERSIANN-CDR

Descriptive statistical metrics (R, MAE, RMSE, and Bias) were used to evaluate the accuracy of daily downscaled

Table 3 General performance ratings for evaluation criteria (Moriasi et al. 2007)

S/*N*	Performance						
	Ratings	NSE	PBIAS	RSR	R^2	PEP (%)	PEV (%)
1	Very Good	0.75–1.0	< ± 10	0–0.5	0.75–1.0	< ± 15	< ± 10
2	Good	0.65–0.75	± 10 - ± 15	0.5–0.6	0.65–0.75	15–30	± 10 - ± 15
3	Satisfactory	0.50–0.65	± 15 - ± 25	0.6–0.7	0.50–0.65	30–40	± 15 - ± 25
4	Unsatisfactory	< 0.50	> ± 25	> 0.70	< 0.50	> 40	> ± 25

PERSIANN-CDR (Ganiyu et al. 2025) and RF, XGB, and LSTM fusion models in reference to observed rainfall data at each gauge station. Gauge-level evaluation is essential for capturing the spatial variability of basin precipitation (Mohammed et al. 2022). Across the rain gauge stations, a significant improvement was observed in all the metrics when comparing the fusion models with the downscaled PERSIANN-CDR, except MAE at Bida for XGB, R at Bida, Lokoja, and Minna, RMSE at Bida, Lokoja, and Minna, and Bias at Abuja, Bida, Lokoja, and Minna gauge stations for LSTM. This result is consistent with previous studies in data-sparse regions (Fan et al. 2021; Nguyen et al. 2021; Zhang et al. 2021). Conversely, the poor performance of the daily downscaled PERSIANN-CDR can be attributed to several factors, including the rainfall retrieval techniques and the sensors used to develop the satellite data (Hu et al. 2019; Sun et al. 2018). Further explanation of the weak performance of SPP at daily temporal resolution could be found in the previous studies by Gebremedhin et al. (2021) and Rahmawati and Lubczynski (2018). Table 4 presents the performances of the downscaled PERSIANN-CDR and fused PERSIANN-CDR in reference to the gauge observations.

The RF model outperformed other fusion models, achieving correlation values above 0.5 at all five stations, which demonstrates the model's high generalizability. The RF model performed best in three of four metrics across gauge stations, except for Baro, which performed well only in R and RMSE. However, LSTM recorded the lowest MAE values of 2.93 mm, 1.62 mm, and 3.56 mm at Abuja, Baro, and Lokoja stations, respectively. Figure 10 shows box-and-whisker plots of model performance, where box borders represent 75th and 25th percentiles, with the median line inside. The LSTM model had the lowest performance with a median R of 0.31, the highest RMSE of 9.97 mm, and the highest bias of −2.9. The RF model achieved the highest median $R = 0.66$ and the lowest median values of MAE = 3.11 mm, RMSE = 7.05 mm, and bias = 0.04, making it the best-performing fusion model. Papacharalampous et al. (2023) found XGB superior to RF in merging PERSIANN and IMERG data, likely due to the high density of gauge stations in the United States. XGB is recommended for large-scale merging due to its computational efficiency (Lei et al. 2022), whereas RF performs well in limited data regions (Biau and Scornet 2016). Nguyen et al. (2021) showed RF's accuracy in data-sparse regions, with median R, MAE, and RMSE of 0.95, 1.09, and 4.44 mm, respectively. Mohammed et al. (2022) confirmed the reliability of the RF model in enhancing satellite precipitation estimates in data-scarce basins.

Figure 11 illustrates the average percentage enhancement in R, MAE, RMSE, and Bias for the RF fusion model over downscaled PERSIANN-CDR across the gauge stations. The RF model enhanced the accuracy of downscaled PERSIANN-CDR by 89.71–146.25.71.25% and decreased errors (MAE, RMSE, and Bias) by 21.67–45.38%, 19.23–38.21%, and 56.21–88.05% across the five gauge stations. On average, the RF model increased accuracy by 113.48% and reduced MAE, RMSE, and Bias by 29.9%, 26.76%, and 75.71%, respectively, across the entire basin.

Table 4 Performance evaluation of downscaled PERSIANN-CDR and fused PERSIANN-CDR models in reference to the gauge observations at daily Temporal resolution from 2013 to 2022

Models	Metrics	Gauge Stations				
		Abuja	Baro	Bida	Lokoja	Minna
Downscaled PERSIANN-CDR**	R	0.36	0.30	0.35	0.24	0.33
	MAE	4.26	3.14	4.18	4.90	4.38
	RMSE	8.53	6.05	10.30	11.00	9.66
	BIAS	−0.42	−1.33	0.29	0.26	−0.13
RF	R	**0.72**	**0.66**	**0.66**	**0.59**	**0.70**
	MAE	2.99	1.72	**3.11**	3.84	**3.20**
	RMSE	**6.11**	**3.74**	**8.12**	**8.91**	**7.05**
	BIAS	**0.08**	0.16	−0.13	**−0.04**	**0.04**
XGB	R	0.49	0.43	0.45	0.37	0.45
	MAE	3.99	2.21	4.19	4.89	4.24
	RMSE	7.68	4.47	9.72	10.30	8.84
	BIAS	0.16	**0.13**	**−0.09**	−0.05	0.06
LSTM	R	0.42	0.40	0.31	0.17	0.27
	MAE	**2.93**	**1.62**	3.23	**3.56**	3.56
	RMSE	8.42	4.73	10.80	11.60	9.97
	BIAS	−2.46	−1.23	−2.98	−3.56	−2.90

*The best-performing values for each metric are highlighted in bold

**The results for Downscaled PERSIANN-CDR is available from Ganiyu et al. (2025)

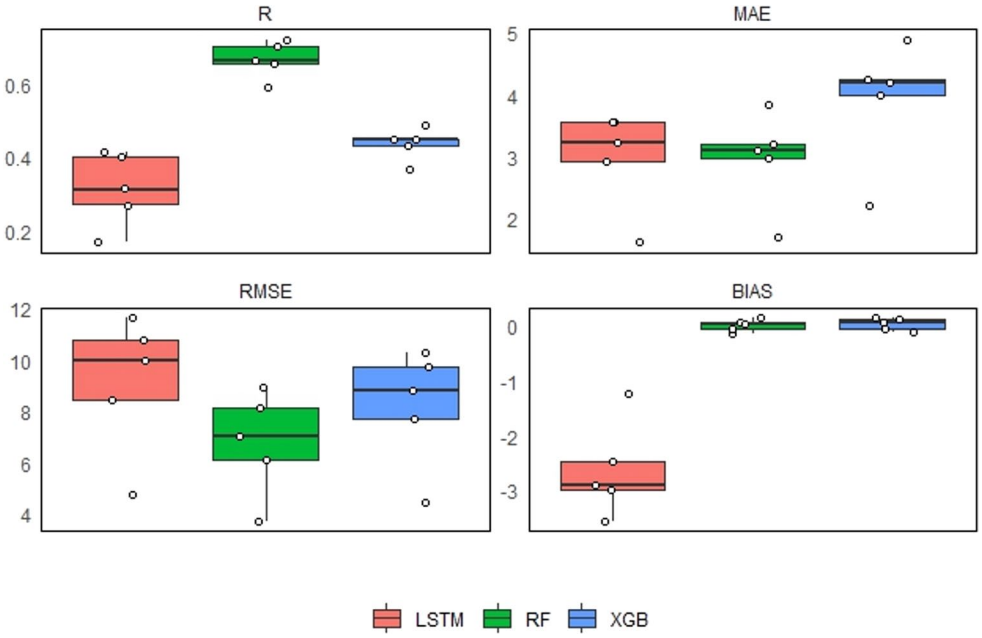
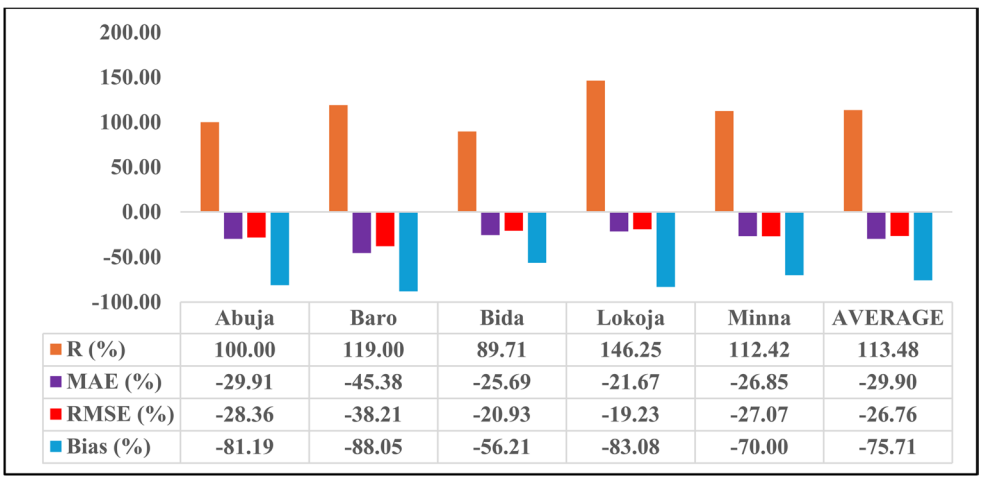


Fig. 10 Evaluation of the performance of LSTM, RF, and XGB fusion models using R, MAE, RMSE, and Bias at daily temporal resolution



	Abuja	Baro	Bida	Lokoja	Minna	AVERAGE
R (%)	100.00	119.00	89.71	146.25	112.42	113.48
MAE (%)	-29.91	-45.38	-25.69	-21.67	-26.85	-29.90
RMSE (%)	-28.36	-38.21	-20.93	-19.23	-27.07	-26.76
Bias (%)	-81.19	-88.05	-56.21	-83.08	-70.00	-75.71

Fig. 11 Average percentage enhancement of RF Fusion Model over downscaled PERSIANN-CDR

3.2 Performance of fused precipitation in hydrological simulation

The performance of the fused PERSIANN-CDR was further assessed through hydrological simulation of extreme flood events using HEC-HMS. The downscaled PERSIANN-CDR, observed rainfall, and fused PERSIANN-CDR data, at a daily temporal resolution, served as the model input data during the calibration and validation process. The calibration and validation results are presented graphically (Fig. 12). The results demonstrated that the performance of the fused PERSIANN-CDR substantially improved compared to the downscaled PERSIANN-CDR and observed rainfall as model input data. The performance of the simulated flows based on the model input data was evaluated against the observed discharge at the basin outlet using several statistical indices, as recommended by Moriasi et al. (2007). The statistical index during the 2018 and 2019 flood events (calibration periods) indicated that simulated flow by the fused PERSIANN-CDR best agreed with the observed discharge data, followed by observed rainfall. The fused PERSIANN-CDR had the highest values of R^2 (0.87 and 0.64) and NSE (0.75 and 0.60), whereas the downscaled PERSIANN-CDR had the lowest NSE values (0.21 and 0.22). For the lowest value of R^2, the downscaled PERSIANN-CDR had 0.63, while the observed rainfall had 0.47 for the 2018 and 2019 flood events, respectively. According to Moriasi et al. (2007), fused PERSIANN-CDR demonstrated outstanding performance in reproducing the observed pattern of the flow hydrograph for R^2 and NSE in 2018, but was satisfactory in the 2019 flood event. The observed rainfall data performed well and satisfactorily in

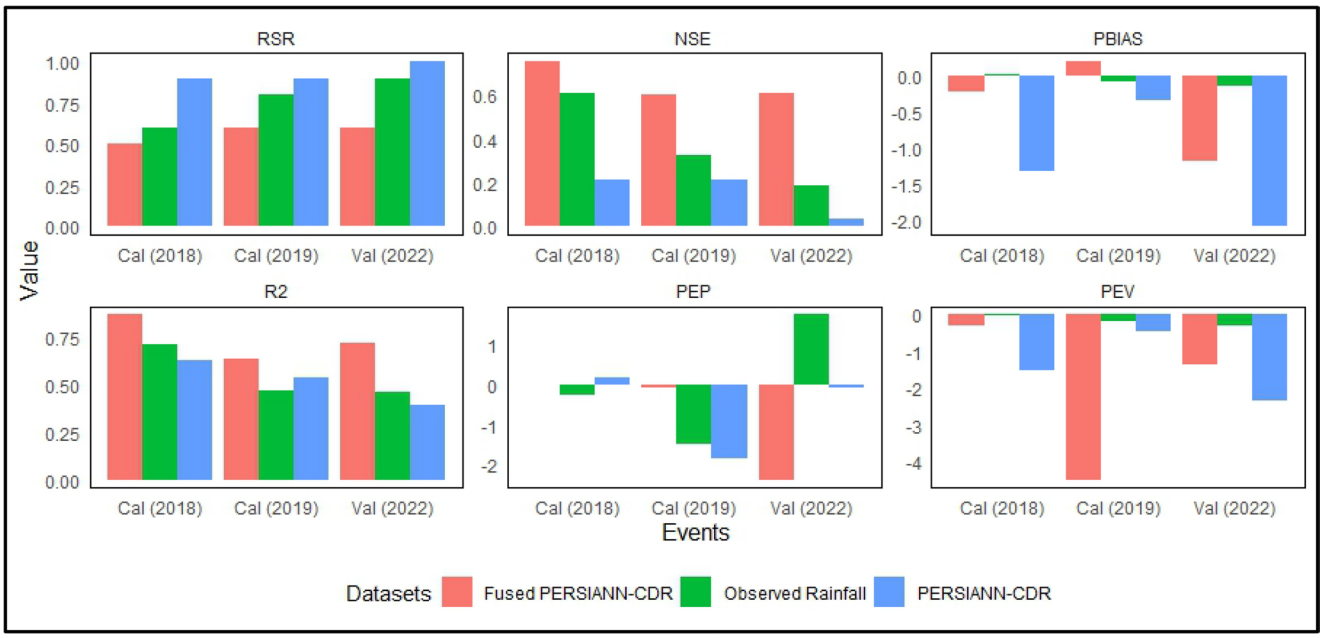


Fig. 12 Computed daily RSR, NSE, PBIAS, R2, PEP, and PEV for fused PERSIANN-CDR, observed rainfall, and downscaled PERSIANN-CDR during calibration (2018 and 2019) and validation (2022) periods

terms of R^2 and NSE in the 2018 flood event. However, downscaled PERSIANN-CDR performed satisfactorily in terms of R^2 in the 2018 and 2019 flood events. In the case of PBIAS, the observed rainfall had the lowest values of 0.01% and −0.08%, whereas the downscaled PERSIANN-CDR had the highest values of −1.33% and −0.34% for the 2018 and 2019 simulation periods, respectively. Based on these results, the simulated flows from the three model inputs performed very well, with a PBIAS of less than 10%. For RSR, fused PERSIANN-CDR had the lowest values of 0.5 and 0.6, while downscaled PERSIANN-CDR had the highest values of 0.9 for both calibration periods. This demonstrated the high performance of the fused PERSIANN-CDR in 2018 and a good performance in 2019. Whereas the simulation obtained from observed rainfall also indicated good performance in the 2018 flood event.

Considering the validation periods in 2022, the simulated flow hydrographs from the fused PERSIANN-CDR showed a strong agreement with the observed discharge, with an R^2 value of 0.72 and a satisfactory performance, as indicated by an NSE value of 0.61, compared to the downscaled PERSIANN-CDR and observed rainfall. The simulated flows from the three datasets performed very well in terms of PBIAS, with values less than 10%. Whereas, only fused PERSIANN-CDR performed satisfactorily, with the lowest RSR value of 0.6. The simulated flows from the three datasets demonstrated outstanding performance in accurately recording the peak flow. However, they underestimated the observed flow, except downscaled PERSIANN-CDR and observed rainfall, with values of 0.17% and 1.74% in 2018 and 2022, respectively. The results revealed that the simulated peak flow by fused PERSIANN-CDR showed better agreement with the observed peak flow, with the lowest PEP values of 0.01% and 0.10% during the calibration periods of 2018 and 2019, respectively. However, the simulated peak flow from the observed rainfall agreed better than the other datasets, with a PEP value of 1.74% during the validation period. Similarly, the three datasets performed very well in terms of the PEV, with values of less than 10% for the simulation periods. The simulated peak volumes from the three datasets underestimated the observed peak volume, with PEV ranging from 0.01% to 4.49% for both the calibration and validation periods. Unlike the PEP, the simulated peak volume from the observed rainfall agreed more with the observed peak volume than the other datasets for the calibration and validation periods. Figure 13 shows the hydrograph patterns of the daily simulated and observed flows for the calibration and validation periods using the three datasets. The simulated flow hydrograph from the fused PERSIANN-CDR dataset matched the observed discharge hydrograph better than the downscaled PERSIANN-CDR and observed rainfall data across the three flood events.

4 Discussion

The RF model demonstrates substantial improvement even with minimal hyperparameter tuning. Its performance may be attributed to its ability to handle small sample sizes (Tyralis et al. 2019) and independent subsets that facilitate

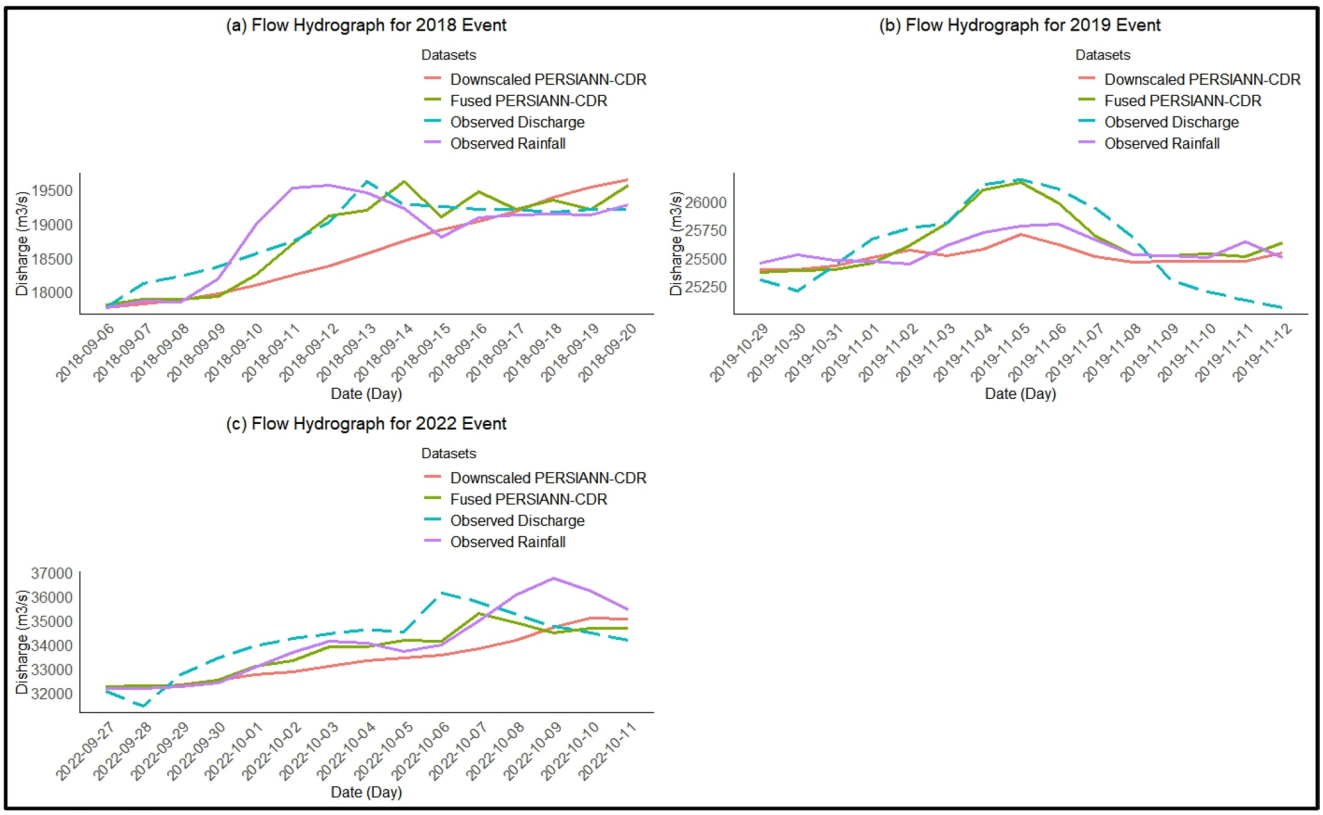


Fig. 13 Hydrographs of daily simulated (with downscaled PERSIANN-CDR, observed rainfall, and fused PERSIANN-CDR) and observed flows for the calibration **a** 2018 Event: 06–20 Sep **b** 2019 Event: 29 Oct–12 Nov, and validation **c** 2022 Event: 27 Sept–11 Oct

parallel computing (Zhang et al. 2022). Similar enhancements were observed by Li et al. (2023) in Sichuan Province, China, with a 407% increase in accuracy and a 24% reduction in RMSE, likely due to the integration of satellite, reanalysis, and auxiliary data in a densely gauged area. Mohammed et al. (2022) reported improved fused data by integrating multi-satellite precipitation with geographic information in the Upper Blue Nile River basin. However, integrating geographic information with downscaled PERSIANN-CDR had no significant impact, leading to its exclusion from fusion models in this study. This is consistent with previous studies (Baez-Villanueva et al. 2020; Bhuiyan et al. 2018; Zhang et al. 2021). This novel study on data fusion is unprecedented in Nigerian research. While applying consistent RF architecture and hyperparameters across stations enhanced model reliability, the model's localized nature limits its applicability in ungauged regions. Despite this limitation, the evaluation of data fusion at existing gauge locations offers a solid foundation for hydrological studies in Nigeria, justifying continued research into developing more robust and regionally applicable precipitation inputs for hydrological applications in data-sparse regions, such as exploring regionalization techniques to enhance spatial generalizability.

The low performance of PERSIANN-CDR in this study conforms with the results obtained even when forced in another hydrological model (SWAT), irrespective of either event- or continuous-based simulation at daily temporal resolution in the data-scarce regions using SPPs as alternative model input data (Vu et al. 2018; Musie et al. 2019; Le et al. 2020; Saddique et al. 2022; Eini et al. 2022). This shows that the inherent bias in the SPP is reflected in the hydrological simulation, resulting in low performance compared to the other model inputs. The RF model's fused precipitation data significantly enhanced the hydrological simulation of extreme flood events. The improved simulated flow shows that the SPP benefits from the RF's ability to leverage the strengths and reduce the weaknesses of both the SPP and observed rainfall data. Therefore, using fused precipitation data will increase the accuracy of flood forecasts, strengthen early warning systems, support emergency preparedness, and help policymakers develop effective flood mitigation strategies.Tang et al. (2021) findings revealed the high performance of data fusion (IMERG and observed rainfall data) compared to individual datasets during the hydrological simulation of flash flood events in the data-sparse river basin of China using the HEC-HMS model.

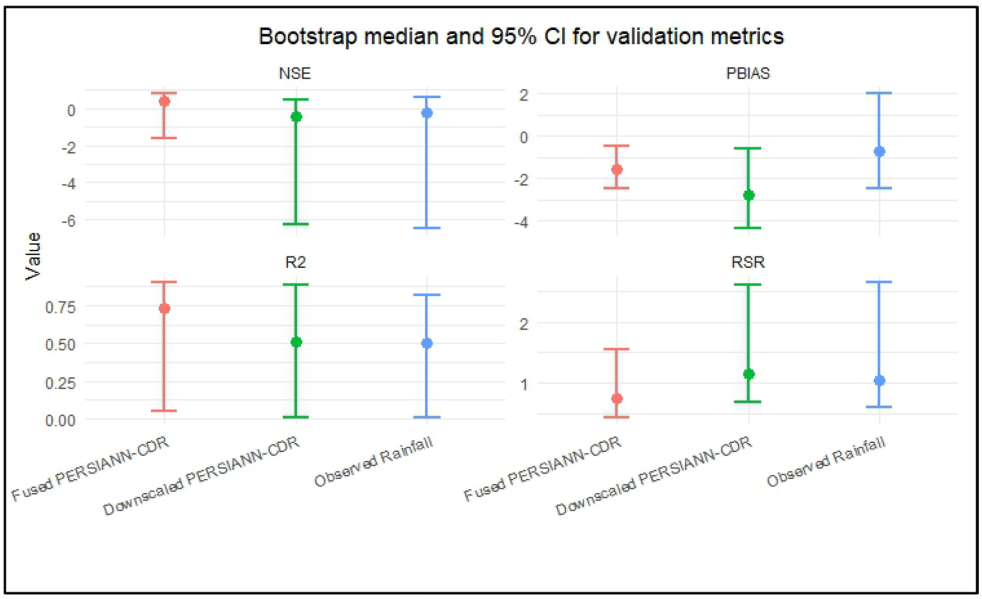


Fig. 14 Bootstrap medians and error bars of 95% percentile CIs for validation metrics

In Africa, a previous study enhanced the accuracy of streamflow predictions in the Bilate River watershed by integrating the GR4J model with machine learning techniques, including RF, ELM, XGB, and LSTM (Mulu et al. 2025). The study concluded that this hybrid model outperformed the GR4J conceptual hydrological model, which is crucial for flood and drought monitoring applications. Another study in Ethiopia improved the hydrological simulation of daily discharge in the Upper Blue Nile Basin by reducing biases in PERSIANN family products through the integration of VMD and machine learning techniques. The study revealed that PERSIANN-CDR showed the most improved discharge outcomes (Ali et al. 2025). Mekonnen et al. (2022) demonstrated the effectiveness of the RF-fused model in enhancing the hydrological simulation of daily, seasonal, and high flows in the data-sparse MelkaKuntire catchment of Ethiopia using HBV-Light. Their findings indicated that fused downscaled SPPs performed better than both the fused undownscaled and the original SPPs. Consequently, our findings are consistent with previous studies that demonstrated the importance of reducing bias in SPPs using machine learning-based data fusion for improved hydrological applications.

To further support the evidence that the RF-fused SPP produces superior flood simulations in our study, we employed block bootstrap confidence intervals and paired tests in the validation period. Figure 14 shows the bootstrapped medians and error bars of 95% percentile confidence intervals, and Table 5 presents the summary of paired Wilcoxon test results using the median bootstrap of validation metrics. The error bar plots illustrated in Fig. 14 demonstrate that the median NSE, R^2, and RSR for RF-fused SPP performed better than the individual datasets. Additionally, these plots highlight the variability in model performance, with the fused PERSIANN-CDR exhibiting narrower error bars across the metrics, indicating reduced uncertainty. This confirms that the fused precipitation data consistently outperforms both satellite-only and gauge-only inputs within acceptable uncertainty bounds. The paired Wilcoxon tests, which produced p-values approximately 0 ($p < 0.05$) for differences in metrics, indicated statistical significance at the 95% level.

Table 5 Summary of paired Wilcoxon test results using median bootstrap of validation metrics

Metrics	Comparison	Approx. Wilcoxon_p	Bootstrap median_diff	Bootstrap CI lower	Bootstrap CI upper	Remarks
NSE	RF-Fused vs. Downscaled PERSIANN-CDR	0.0000	0.8267	0.7829	0.8995	Significant
	RF-Fused vs. Observed Rainfall	0.0000	0.4665	0.4288	0.52	Significant
R^2	RF-Fused vs. Downscaled PERSIANN-CDR	0.0000	0.2078	0.192	0.2197	Significant
	RF-Fused vs. Observed Rainfall	0.0000	0.2162	0.2023	0.2264	Significant
PBIAS	RF-Fused vs. Downscaled PERSIANN-CDR	0.0000	1.163	1.118	1.195	Significant
	RF-Fused vs. Observed Rainfall	0.0000	−0.8605	−0.9148	−0.829	Significant
RSR	RF-Fused vs. Downscaled PERSIANN-CDR	0.0000	−0.4258	−0.4423	−0.413	Significant
	RF-Fused vs. Observed Rainfall	0.0000	−0.2612	−0.2784	−0.2419	Significant

Despite the proposed RF fusion model's better performance in enhancing the extreme flood events in the data-sparse region of the Niger Central Hydrological Area in Nigeria, certain limitations must be acknowledged. The performance of the fusion model may vary under different catchment characteristics or climatic conditions. Manual calibration of the HEC-HMS model was performed by iteratively adjusting the hydrological parameters, including CN, lag time, initial abstraction, recession constant, travel time, percent imperviousness, and attenuation coefficient. The calibrated parameter ranges presented in Table 2 reflect the variability explored during the model tuning. The calibration process inherently involves assessing the influence of individual parameters on the model performance. Future studies should evaluate the performance of the fusion model approach across diverse basins in data-sparse regions, employ ensemble models for data fusion, and further extend the uncertainty quantification analysis by exploring additional approaches, such as probabilistic frameworks, for a more robust assessment of confidence in hydrological predictions.

5 Conclusion

This study enhances the hydrological simulation of extreme flood events in a data-sparse river basin in Nigeria by employing Extreme Gradient Boosting (XGB), Random Forest (RF), and Long Short-Term Memory (LSTM) models to fuse daily downscaled PERSIANN-CDR satellite precipitation with observed rainfall data at the existing gauge locations from 2013 to 2022. The RF-fused PERSIANN-CDR demonstrated higher accuracy than XGBoost and LSTM, maintaining consistent model architecture and hyperparameters across all gauging stations. It enhanced the mean correlation coefficient of the downscaled PERSIANN-CDR by 113.48% and reduced the mean MAE, RMSE, and Bias by 29.90%, 26.76%, and 75.71%, respectively. The reliability of the RF-fused PERSIANN-CDR was hydrologically validated using the HEC-HMS model to simulate 2018, 2019, and 2022 extreme flood events. The simulated flow with the fused data exhibited higher NSE and R^2 values and lower RSR values compared to individual datasets during both calibration and validation periods. To enhance the reliability of the model evaluation, uncertainty quantification was integrated using bootstrap median differences and 95% CIs and Wilcoxon tests for the validation metrics. The resulting confidence intervals and p-values offered a measure of the statistical robustness of the observed improvements across all performance metrics. Additionally, the error-bar plots illustrated the variability in model performance, confirming the improvements of fused precipitation data compared to the individual datasets within acceptable uncertainty bounds. These analyses provide greater confidence in the robustness and reproducibility of the fusion model's results. Therefore, this study offers a practical and scalable framework that highlights the potential of data fusion in enhancing satellite precipitation estimates at existing gauge locations, thereby improving extreme flood simulation in the data-sparse river basin. However, the gauge station-based evaluations may limit data transferability to ungauged regions, necessitating retraining or additional calibration. Although the fusion model demonstrates promising results, its performance may vary under different catchment characteristics or climatic conditions. Therefore, future studies should apply fused precipitation data for flood forecasting, assess the fusion model across different basins in data-sparse regions, and evaluate its applicability in ungauged basins using regionalization techniques such as hydrological similarity and K-nearest neighbour (KNN) donor basin transfer. Transfer learning, particularly fine-tuning pretrained models on physiographically similar basins and applying domain adaptation methods, should be explored to improve the model's scalability and generalization. Additionally, employing ensemble models for data fusion and using probabilistic frameworks for uncertainty quantification are recommended. These comprehensive approaches will enhance the evaluation of confidence in both the fusion model and the resulting hydrological predictions.

Acknowledgements This work is part of research funded by the Tertiary Education Trust Fund (TETFUND), Nigeria, under the Overseas PhD Scholarship Scheme. The authors also appreciate the financial support from the Ministry of Higher Education via the FRGS grant (Grant number: FRGS/1/2020/TK0/UM/02/19). We thank the Nigerian Meteorological Agency (NiMET) and the National Inland Waterways Authority (NIWA) for making the observed rainfall and hydrological data available.

Authors' contributions All authors contributed to the study's conception and design. Supervision and the study's framework were done by Faridah Othman and Wan Zurina Wan Jaafar. Material preparation, data collection and analysis were performed by Habeeb Oladimeji Ganiyu. The first draft of the manuscript was written by Habeeb Oladimeji Ganiyu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work is part of research funded by the Tertiary Education Trust Fund (TETFUND), Nigeria, under the Overseas PhD Scholarship Scheme. The authors also appreciate the financial support from the Ministry of Higher Education via the FRGS grant (Grant number: FRGS/1/2020/TK0/UM/02/19).

Data availability The observed rainfall and discharge data used for this study are available from the first author upon reasonable request. PERSIANN-CDR is available at [https://www.ncei.noaa.gov/data/precipitation-persiann/access/](https://www.ncei.noaa.gov/data/precipitation-persiann/access).

Declarations

Competing interests The authors declare no competing interests.

References

Abdo HG, Richi SM, Alqadhi S, Zeng T, Prasad P, Kotaridis I, Alharbi MM, Khaddour LA, Mallick J (2025) A hybrid machine learning modelling for optimization of flood susceptibility mapping in the Eastern Mediterranean. Nat Hazards 121(6):7199–7228. https://doi.org/10.1007/s11069-024-07081-3

Ali AM, Abdallah M, Mohammadi B, Elzain HE (2025) Three-stage hybrid modeling for real-time streamflow prediction in data-scarce regions. J Hydrol Reg Stud. https://doi.org/10.1016/j.ejrh.2025.102337

Assiri ME, Qureshi S (2022) A multi-source data fusion method to improve the accuracy of precipitation products: a machine learning algorithm. Remote Sens. https://doi.org/10.3390/rs14246389

Awad MA, Mohammed A, Babak M, Hussam EE (2025) Three-stage hybrid modeling forreal-time streamflow prediction in data-scarce regions. Journal of Hydrology: Regional Studies 59:102337. https://doi.org/10.1016/j.ejrh.2025.102337

Baez-Villanueva OM, Zambrano-Bigiarini M, Beck HE, McNamara I, Ribbe L, Nauditt A, Birkel C, Verbist K, Giraldo-Osorio JD, Xuan Thinh N (2020) RF-MEP: a novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens Environ 239:111606. https://doi.org/10.1016/j.rse.2019.111606

Beck HE, Pan M, Roy T, Weedon GP, Pappenberger F, Van Dijk AIJM, Huffman GJ, Adler RF, Wood EF (2019) Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol Earth Syst Sci 23(1):207–224. https://doi.org/10.5194/hess-23-207-2019

Bhuiyan MAE, Nikolopoulos EI, Anagnostou EN, Quintana-Seguí P, Barella-Ortiz A (2018) A nonparametric statistical technique for combining global precipitation datasets: development and hydrological evaluation over the Iberian Peninsula. Hydrol Earth Syst Sci 22(2):1371–1389. https://doi.org/10.5194/hess-22-1371-2018

Biau G, Scornet E (2016) A random forest guided tour. Test 25(2):197–227. https://doi.org/10.1007/s11749-016-0481-7

Bodjrènou M, Peng K, Affèwé DJ, Hounkpè J, Donnou HEV, Adounkpè J, Akpo AB (2025) Assessment of satellite-based rainfall products for flood modeling in the Ouémé River Basin in Benin (West Africa). Hydrology. https://doi.org/10.3390/hydrology12040071

Bouizrou I, Bouadila A, Aqnouy M, Gourfi A (2023) Assessment of remotely sensed precipitation products for climatic and hydrological studies in arid to semi-arid data-scarce region, central-western Morocco. Remote Sens Appl Soc Environ. https://doi.org/10.1016/j.rsase.2023.100976

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010950718922

Chen F, Gao Y, Wang Y, Qin F, Li X (2019) Downscaling satellite-derived daily precipitation products with an integrated framework. Int J Climatol 39(3):1287–1304. https://doi.org/10.1002/joc.5879

Darko S, Adjei KA, Gyamfi C, Odai SN, Osei-Wusuansa H (2021) Evaluation of RFE satellite precipitation and its use in streamflow simulation in poorly gauged basins. Environ Processes 8(2):691–712. https://doi.org/10.1007/s40710-021-00495-2

Dhanapala L, Gunarathna MHJP, Kumari MKN, Ranagalage M, Sakai K, Meegastenna TJ (2022) Towards coupling of 1D and 2D models for flood simulation-a case study of Nilwala River Basin, Sri Lanka. Hydrology 9:17. https://doi.org/10.3390/hydrology9020017

Ding J, Wang X, Zhou S, Lei S (2025) Holistic risk assessment using a hybrid approach in a flash flood disaster-prone area of the Poyang Lake basin. Nat Hazards. https://doi.org/10.1007/s11069-025-07306-z

Eini MR, Rahmati A, Piniewski M (2022) Hydrological application and accuracy evaluation of PERSIANN satellite-based precipitation estimates over a humid continental climate catchment. J Hydrol Reg Stud. https://doi.org/10.1016/j.ejrh.2022.101109

Fan Z, Li W, Jiang Q, Sun W, Wen J, Gao J (2021) A comparative study of four merging approaches for regional precipitation estimation. IEEE Access 9:33625–33637. https://doi.org/10.1109/ACCESS.2021.3057057

Faty B, Sterk G, Ali A, Sy S, Dacosta H, Diop S, Descroix L (2023) Satellite-based rainfall estimates to simulate daily streamflow using a hydrological model over Gambia watershed. Water Sci 37(1):1–18. https://doi.org/10.1080/23570008.2023.2225898

Fischer GF, Nachtergaele S, Prieler HT, van Velthuizen L, Verelst D, Wiberg (2008) Global Agro-ecological zones assessment for agriculture (GAEZ). IIASA, Laxenburg, Austria and FAO, Rome, Italy

Fleming MJ (2013) Hydrologic Modeling System HEC-HMS: Quick Start Guide. US Army Corps of Engineers, Hydrologic Engineering Center HEC, Davis, CA

Ganiyu HO, Othman F, Jaafar WZW, Ng CY (2025) Comprehensive evaluation of satellite precipitation products over sparsely gauged river basin in Nigeria. Theor Appl Climatol 156(3):156–169. https://doi.org/10.1007/s00704-025-05388-0

Gebregiorgis AS, Kirstetter PE, Hong YE, Gourley JJ, Huffman GJ, Petersen WA, Xue X, Schwaller MR (2018) To what extent is the day 1 GPM IMERG satellite precipitation estimate improved as compared to TRMM TMPA-RT? J Geophys Research: Atmos 123(3):1694–1707. https://doi.org/10.1002/2017JD027606

Gebremedhin MA, Lubczynski MW, Maathuis BHP, Teka D (2021) Novel approach to integrate daily satellite rainfall with in-situ rainfall, Upper Tekeze Basin, Ethiopia. Atmos Res. https://doi.org/10.1016/j.atmosres.2020.105135

Gilewski P, Nawalany M (2018) Inter-comparison of rain-gauge, radar, and satellite (IMERG GPM) precipitation estimates performance for rainfall-runoff modeling in a mountainous catchment in Poland. Water. https://doi.org/10.3390/w10111665

Goshime DW, Absi R, Ledésert B (2019) Evaluation and bias correction of CHIRP rainfall estimate for rainfall-runoff simulation over Lake Ziway Watershed, Ethiopia. Hydrology. https://doi.org/10.3390/hydrology6030068

Gunathilake MB, Amaratunga YV, Perera A, Karunanayake C, Gunathilake AS, Rathnayake U (2020) Statistical evaluation and hydrologic simulation capacity of different satellite-based precipitation products (SbPPs) in the Upper Nan River Basin, Northern Thailand. J Hydrol Reg Stud. https://doi.org/10.1016/j.ejrh.2020.100743

Haque A, Saha OR, Aziz MA (2022) Flood hazard mapping of the floodplain of the Jamuna river. Comput Eng Phys Model 5(1):89–101. https://doi.org/10.22115/cepm.2022.327103.1201

Hounguè NR, Ogbu KN, Almoradie ADS, Evers M (2021) Evaluation of the performance of remotely sensed rainfall datasets for flood simulation in the transboundary Mono River catchment, Togo and Benin. J Hydrol Reg Stud. https://doi.org/10.1016/j.ejrh.2021.100875

Hu Q, Li Z, Wang L, Huang Y, Wang Y, Li L (2019) Rainfall spatial estimations: a review from spatial interpolation to multi-source data merging. Water. https://doi.org/10.3390/w11030579

Ismail H, Kamal MR, Mojid MA, Abdullah AFB, Hin LS (2022) Loss methods in HEC-HMS model for streamflow projection under

climate change: a review. Int J Hydrology Sci Technol 13(1):23–42. https://doi.org/10.1504/ijhst.2022.119234
Jabbar LA, Khalil IA, Sidek LM (2021) HEC-HMS hydrological modelling for runoff Estimation in Cameron Highlands, Malaysia. Int J Civil Eng Technol (IJCIET) 12(9). https://doi.org/10.34218/ijciet.12.9.2021.004
Jha A, Lamond J, Proverbs D,Bhattacharya-Mis N, Barker R (2012) Cities and Flooding: A Guide to IntegratedUrban Flood Risk Management for the 21st Century; World Bank Publications:Washington, DC, USA
Khan S, Khan F, Guan Y (2022) Assessment of gridded precipitation products in the hydrological modeling of a flood-prone mesoscale basin. Hydrol Res 53(1):85–106. https://doi.org/10.2166/NH.2021.073
Kittel CMM, Arildsen AL, Dybkjær S, Hansen ER, Linde I, Slott E, Tøttrup C, Bauer-Gottwein P (2020) Informing hydrological models of poorly gauged river catchments – a parameter regionalization and calibration approach. J Hydrol. https://doi.org/10.1016/j.jhydrol.2020.124999
Kouakou C, Paturel JE, Satgé F, Tramblay Y, Defrance D, Rouché N (2023) Comparison of gridded precipitation estimates for regional hydrological modeling in West and Central Africa. J Hydrol Reg Stud. https://doi.org/10.1016/j.ejrh.2023.101409
Lei H, Zhao H, Ao T (2022) A two-step merging strategy for incorporating multi-source precipitation products and gauge observations using machine learning classification and regression over China. Hydrol Earth Syst Sci 26(11):2969–2995. https://doi.org/10.5194/hess-26-2969-2022
Le MH, Lakshmi V, Bolten J, Bui DD (2020) Adequacy of satellite-derived precipitation estimate for hydrological modeling in Vietnam basins. J Hydrol. https://doi.org/10.1016/j.jhydrol.2020.124820
Li Z, Wang H, Zhang T, Zeng Q, Xiang J, Liu Z, Yang R (2023) Multi-source precipitation data merging for high-resolution daily rainfall in complex terrain. Remote Sens. https://doi.org/10.3390/rs15174345
Magami IM, Yahaya S, Mohammed K, Magami IM, Mohammed K (2014) Causes and consequences of flooding in nigeria: a review. In Biol Environ Sci J Tropics 11(2). https://www.researchgate.net/publication/262562763
Mao Y, Qin H, Yaojun S, Zilong H, Zhaohui G, Decheng M, Kouhdaragh M (2025) Implementing an explored advanced and integrated deep random forest learning-based model to monitor the enhanced landslide susceptibility mapping. Nat Hazards. https://doi.org/10.1007/s11069-025-07415-9
Matazu MB, Chioma M (2014) The application of Geospatial techniques in flood risk and vulnerability mapping for disaster management at Lokoja, Kogi State, Nigeria. J Environ Earth Sci 4(5). https://www.iiste.org/Journals/index.php/JEES/article/download/11627/11970
Mekonnen K, Melesse AM, Woldesenbet TA (2022) Merging satellite rainfall estimates and daily rain gauge observations for improved flood simulation in MelkaKuntire catchment, upper Awash Basin, Ethiopia. Remote Sensing Applications: Society and Environment. https://doi.org/10.1016/j.rsase.2022.100701
Min X, Yang C, Dong N (2020) Merging satellite and gauge rainfalls for flood forecasting of two catchments under different climate conditions. Water. https://doi.org/10.3390/w12030802
Moges DM, Kmoch A, Uuemaa E (2022) Application of satellite and reanalysis precipitation products for hydrological modeling in the data-scarce Porijõgi catchment, Estonia. Journal of Hydrology: Regional Studies. https://doi.org/10.1016/j.ejrh.2022.101070
Mohammed J, Mengiste Y, Singh VP (2022) Improving spatio-temporal precipitation estimates in data scarce river basins: an application of machine learning-based multi-source data merging. Stoch Env Res Risk Assess 37:1353–1369. https://doi.org/10.1007/s00477-022-02346-4
Moriasi D, Arnold J, Van Liew M, Bingner R, Harmel R, Veith T (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50:885–900. https://doi.org/10.13031/2013.23153
Mulu BA, Zimale FA, Kebede MG (2025) Integration of conceptual hydrological and machine learning models via output augmentation for enhanced streamflow prediction. Water Soil Manage Model 5(4):95–115. https://doi.org/10.22098/mmws.2025.18128.1648
Musie M, Sen S, Srivastava P (2019) Comparison and evaluation of gridded precipitation datasets for streamflow simulation in data scarce watersheds of Ethiopia. J Hydrol. https://doi.org/10.1016/j.jhydrol.2019.124168
Natarajan S, Radhakrishnan N (2021) Simulation of rainfall-runoff process for an ungauged catchment using an event-based hydrologic model: a case study of Koraiyar basin in Tiruchirappalli city, India. J Earth Syst Sci 130:30. https://doi.org/10.1007/s12040-020-01532-8
National Emergency Management Agency, NEMA (2022) Report on Flood Disasters in Nigeria. Government Press, Abuja
Ngene BU, Obianigwe N (2018) Nigerian rain gauge station optimization and national development: the importance of headcount. IOP Conference Series: Materials Science and Engineering 413 012022. pdf (iop.org)
Nguyen GV, Le XH, Van LN, Jung S, Yeon M, Lee G (2021) Application of random forest algorithm for merging multiple satellite precipitation products across South Korea. Remote Sens. https://doi.org/10.3390/rs13204033
Niazkar M, Menapace A, Brentan B, Piraei R, Jimenez D, Dhawan P, Righetti M (2024) Applications of XGBoost in water resources engineering: a systematic literature review (Dec 2018–May 2023). Environ Model Softw. https://doi.org/10.1016/j.envsoft.2024.105971
NIHSA (2022) Nigerian Hydrological Services Agency. Annual flood outlook publication. https://nihsa.gov.ng/publications/. Accessed 23 Feb 2023
Nkwunonwo UC (2016) A review of flooding and flood risk reduction in Nigeria. Global J Human-Social Sci (B) 16(2). https://globaljournals.org/
OloghadienI (2018) CODEN(USA): JSERBR development of updated rating curves/equations forNiger and Benue River Systems. Journal of Scientific and Engineering Research5(11): 90–96.
Otache MY, Tyabo MA, Animashaun IM, Ezekiel LP (2016) Application of parametric-based framework for regionalisation of flow duration curves. J Geoscience Environ Prot 04(05):89–99. https://doi.org/10.4236/gep.2016.45009
Oyerinde GT, Fademi IO, Denton OA (2017) Modeling runoff with satellite-based rainfall estimates in the Niger basin. Cogent Food Agric. https://doi.org/10.1080/23311932.2017.1363340
Papacharalampous G, Tyralis H, Doulamis A, Doulamis N (2023) Comparison of tree-based ensemble algorithms for merging satellite and earth-observed precipitation data at the daily time scale. Hydrology. https://doi.org/10.3390/hydrology10020050
Peng J, Liu T, Huang Y, Ling Y, Li Z, Bao A, Chen X, Kurban A, De Maeyer P (2021) Satellite-based precipitation datasets evaluation using gauge observation and hydrological modeling in a typical arid land watershed of central Asia. Remote Sens 13(2):1–26. https://doi.org/10.3390/rs13020221
Rahmawati N, Lubczynski MW (2018) Validation of satellite daily rainfall estimates in complex terrain of Bali Island, Indonesia. Theor Appl Climatol 134(1–2):513–532. https://doi.org/10.1007/s00704-017-2290-7
Reynolds JE, Halldin S, Seibert J, Xu CY, Grabs T (2020) Robustness of flood-model calibration using single and multiple events. Hydrol Sci J 65(5):842–853. https://doi.org/10.1080/02626667.2019.1609682

Saddique N, Muzammil M, Jahangir I, Sarwar A, Ahmed E, Aslam RA, Bernhofer C (2022) Hydrological evaluation of 14 satellite-based, gauge-based and reanalysis precipitation products in a data-scarce mountainous catchment. Hydrol Sci J 67(3):436–450. https://doi.org/10.1080/02626667.2021.2022152

Sadeghi M, Nguyen P, Hsu K, Sorooshian S (2020) Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. Environ Model Softw. https://doi.org/10.1016/j.envsoft.2020.104856

Salami AA, Olanrewaju RM, Olorunfemi JF (2024) Performance assessment of satellite rainfall estimates in rain detection capabilities at different thresholds over Nigeria. Hydrol Sci J 69(14):1997–2009. https://doi.org/10.1080/02626667.2024.2395462

Salami AA, Olorunfemi JF, Eludoyin AO (2025) Multi-criteria assessment of gridded rainfall data accuracy in Nigeria using compromise programming. Earth Syst Environ. https://doi.org/10.1007/s41748-025-00601-6

Shen J, Liu P, Xia J, Zhao Y, Dong Y (2022) Merging multisatellite and gauge precipitation based on geographically weighted regression and long short-term memory network. Remote Sens. https://doi.org/10.3390/rs14163939

Singh L, Mishrab PK, Pingale SM, Kharea D, Thakur HP (2022) Streamflow regionalisation of an ungauged catchment with machine learning approaches. Hydrol Sci J 67(6):886–897

Sun R, Yuan H, Yang Y (2018) Using multiple satellite-gauge merged precipitation products ensemble for hydrologic uncertainty analysis over the Huaihe Riverbasin. Journal of Hydrology 566:406–420. https://doi.org/10.1016/j.jhydrol.2018.09.024

Tadesse D, Suryabhagavan KV, Nedaw D, Hailu BT (2022) A model-based flood hazard mapping in Itang District of the Gambella region, Ethiopia. Geology, Ecology, and Landscapes 8(1):8–25. https://doi.org/10.1080/24749508.2021.2022833

Tangam IS, Yonaba R, Niang D, Adamou MM, Keïta A, Karambiri H (2024) Daily simulation of the rainfall-runoff relationship in the Sirba River basin in West Africa: insights from the HEC-HMS model. Hydrology. https://doi.org/10.3390/hydrology11030034

Tang X, Yin Z, Qin G, Guo L, Li H (2021) Integration of satellite precipitation data and deep learning for improving flash flood simulation in a poor-gauged mountainous catchment. Remote Sens. https://doi.org/10.3390/rs13245083

Toh SC, Lai SH, Mirzaei M, Soo EZX, Teo FY (2023) Sequential data processing for IMERG satellite rainfall comparison and improvement using LSTM and ADAM optimizer. Appl Sci. https://doi.org/10.3390/app13127237

Tyralis H, Papacharalampous G, Langousis A (2019) A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water. https://doi.org/10.3390/w11050910

Umar N, Gray A (2022) Flooding in Nigeria: a review of its occurrence and impacts and approaches to modelling flood data. Int J Environ Stud 80(3):540–561. https://doi.org/10.1080/00207233.2022.2081471

USACE (2022) Hydrologic Modeling System HEC-HMS User's Manual. Hydrologic Engineering Center, Davis

Usman M, Ndehedehe CE, Ahmad B, Manzanas R, Adeyeri OE (2022) Modeling streamflow using multiple precipitation products in a topographically complex catchment. Model Earth Syst Environ 8(2):1875–1885. https://doi.org/10.1007/s40808-021-01198-1

Vu TT, Li L, Jun KS (2018) Evaluation of multi-satellite precipitation products for streamflow simulations: a case study for the Han River Basin in the Korean Peninsula, East Asia. Water. https://doi.org/10.3390/w10050642

WangXL, Chen H, Wu Y, Feng Y, Pu Q (2010) New techniques for the detection andadjustment of shifts in daily precipitation data series. Journal of AppliedMeteorology and Climatology 49(12):2416–2436. https://doi.org/10.1175/2010JAMC2376.1

Weber de Melo W, Iglesias I, Pinho J (2025) Early warning system for floods at estuarine areas: combining artificial intelligence with process-based models. Nat Hazards 121(4):4615–4638. https://doi.org/10.1007/s11069-024-06957-8

Wedajo GK, Muleta MK, Awoke BG (2021) Performance evaluation of multiple satellite rainfall products for Dhidhessa river basin (DRB), Ethiopia. Atmos Meas Tech 14(3):2299–2316. https://doi.org/10.5194/amt-14-2299-2021

World Bank Report (2019) Building Climate Resilience: Experience from Nigeria. https://www.worldbank.org/en/results/2019/04/18/building-climate-resilience-experiencefrom-nigeria. Accessed 23 May 2023

World Health Organization, WHO (2021) Climate Change. https://www.who.int/heli/risks/climate/climatechange/en/on. Accessed 20 Apr 2023

Worqlul AW, Yen H, Collick AS, Tilahun SA, Langan S, Steenhuis TS (2017) Evaluation of CFSR, TMPA 3B42 and ground-based rainfall data as input for hydrological models, in data-scarce regions: the upper Blue Nile Basin, Ethiopia. CATENA 152:242–251. https://doi.org/10.1016/j.catena.2017.01.019

Wu H, Yang Q, Liu J, Wang G (2020) A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. J Hydrol. https://doi.org/10.1016/j.jhydrol.2020.124664

Yan X, Chen H, Tian B, Sheng S, Wang J, Kim JS (2021) A downscaling–merging scheme for improving daily spatial precipitation estimates based on random forest and cokriging. Remote Sens. https://doi.org/10.3390/rs13112040

Yuan F, Zhang L, Wah Win KW, Ren L, Zhao C, Zhu Y, Jiang S, Liu Y (2017) Assessment of GPM and TRMM multi-satellite precipitation products in streamflow simulations in a data-sparse mountainous watershed in Myanmar. Remote Sens 9(3). https://doi.org/10.3390/rs9030302

Zhang J, Xu J, Dai X, Ruan H, Liu X, Jing W (2022) Multi-source precipitation data merging for heavy rainfall events based on cokriging and machine learning methods. Remote Sens. https://doi.org/10.3390/rs14071750

Zhang L, Li X, Zheng D, Zhang K, Ma Q, Zhao Y, Ge Y (2021) Merging multiple satellite-basedprecipitation products and gauge observations using a novel double machinelearning approach. Journal of Hydrology, 594. https://doi.org/10.1016/j.jhydrol.2021.125969

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.