

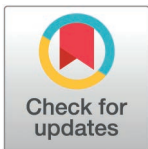
RESEARCH ARTICLE

Toward reliable water resource assessment: Correcting bias in satellite-based rainfall and model-derived discharge in Kenya's Central Highlands

Felicia Yeboah^{1*}, Afua Owusu², Mansoor Leh³, Komlavi Akpoti², Kirubel Mekonnen⁴, Elias Muluken Adamseged⁴, Emma Odera⁵, Naga Velpuri³

1 Department of Earth Science, University of Ghana, Legon, Accra, Ghana, **2** International Water Management Institute (IWMI), Accra, Ghana, **3** International Water Management Institute (IWMI), Colombo, Sri Lanka, **4** International Water Management Institute (IWMI), Addis Ababa, Ethiopia, **5** State Department for Irrigation, Nairobi, Kenya

* fyeboah011@st.ug.edu.gh



OPEN ACCESS

Citation: Yeboah F, Owusu A, Leh M, Akpoti K, Mekonnen K, Adamseged EM, et al. (2026) Toward reliable water resource assessment: Correcting bias in satellite-based rainfall and model-derived discharge in Kenya's Central Highlands. PLOS Water 5(9): e0000623. <https://doi.org/10.1371/journal.pwat.0000623>

Editor: Christopher P Konrad, US Geological Survey, UNITED STATES OF AMERICA

Received: April 7, 2026

Accepted: August 28, 2026

Published: September 25, 2026

Copyright: © 2026 Yeboah et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The observed rainfall data is available on the CETRAD website via this link: <https://wlrc-ken.org/data/timeseries/home> Shapefiles of Administrative boundaries were obtained from the geoBoundaries (<https://www.geoboundaries.org/>) The

Abstract

Reliable precipitation and streamflow data are essential for hydrological modelling and water resources management, yet observational networks in many African regions remain sparse. This study evaluated five satellite precipitation products (SPPs) CHIRPS, PERSIANN-CCS, PERSIANN-PDIR, TAMSAT, and MSWEP and modelled stream-discharge data from VegET–mizuRoute model against in situ observations across the Central Highlands of Kenya for 2012–2021. Performance was assessed at eight precipitation and six discharge stations using pairwise statistics (RMSE, NSE, CC, RMB) and categorical metrics (POD, FAR) on a monthly scale. Four bias correction techniques (Least Squares, Quantile Mapping, Delta Method, Random Forest) were assessed to enhance the utility of the best-performing SPP and the modelled discharge data. CHIRPS and MSWEP consistently outperformed the other products: CHIRPS achieved RMSE of 34.85–134.26 mm, NSE of –0.94 to 0.44, and CC of 0.52–0.87, while MSWEP recorded RMSE of 24.17–125.66 mm, NSE of –0.91 to 0.65, and CC of 0.45–0.84. The VegET–mizuRoute model reproduced seasonal flow dynamics but showed station-dependent biases, with NSE from –2.28 to 0.23, RMB from –0.59 to 0.77, and CC from 0.46 to 0.71. Following bias correction, Random Forest yielded substantial improvements for both precipitation and discharge, underscoring the capacity of machine learning to address complex, non-linear biases. It raised precipitation NSE to a consistently positive 0.85–0.95 across all eight stations while halving RMSE at most sites (9.05–46.23 mm). For discharge, Random Forest improved NSE at most stations to 0.80–0.88, with CC rising to 0.92–0.95, except at one station. Our findings confirm that systematic bias remains a dominant error source in global hydrological products in this region, and that machine learning-based bias correction is essential for the operational reliability and accuracy of these datasets for water management and flood applications.

VegDischarge dataset used in this study is available upon reasonable request from the International Water Management Institute (IWMI), licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Requests for access to the dataset should be directed to Komlavi Akpoti, Researcher, Remote Sensing and Hydrologic Modeling – Water Data Science for Action (WDSA), IWMI, Accra, Ghana (k.akpoti@cgiar.org). The codes used are available on Zenodo accessible with this DOI: <https://doi.org/10.5281/zenodo.21960308> or via this link: <https://zenodo.org/records/21960308>.

Funding: The authors gratefully acknowledge the financial support of CGIAR and The Leona M. and Harry B. Helmsley Charitable Trust. The opinions expressed in this study are those of the authors and do not necessarily reflect the views of the funders.

Competing interests: The authors have declared that no competing interests exist.

1. Introduction

Water is essential for agricultural productivity and ensuring food security. In the face of growing water scarcity and declining water quality, improving water-use efficiency is crucial [1]. Thus, quantifying the movement of water and how its storage changes within the environment, water budget analysis can aid decision-making for water resources management [2]. Unfortunately, such analysis is often limited by a lack of sufficient insitu hydroclimatic data, including precipitation and discharge measurements, which are fundamental for hydrometeorological research and water management operational applications [3]. Moreover, even where in-situ observations exist, they are often spatially sparse, temporally discontinuous, or affected by measurement inconsistencies, while global hydrological models and satellite-based products, though increasingly available, can exhibit systematic biases that limit their ability to represent local hydroclimatic conditions.

Historically, rain gauges measurements have been the primary source of precipitation data, but in many countries, the network of rain gauges is inadequate and insufficient for generating accurate distributed precipitation maps [4]. The sparse distribution is confounded by temporally incomplete records, which introduce uncertainties when gauged data are used for hydrometeorological studies and decision-support tools. Similarly, accurate discharge measurements are needed for efficient allocation of water resources and flood risk management [5]. However, obtaining reliable and long-term global precipitation and discharge data with both high spatial and temporal resolution remains a significant challenge.

Recent developments have shown that the spatiotemporal distribution of precipitation over large areas can be reliably presented by blending data from gauge stations and remote sensing satellites [6–11]. Several satellite-derived precipitation datasets are now accessible, including the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) [12]; Global Satellite Mapping of Precipitation (GSMaP) for the Global Precipitation Measurement (GPM) mission [13]; Integrated Multi-satellite Retrievals for GPM (IMERG) [14]; Tropical Applications of Meteorology using Satellite data and ground-based observations (TAMSAT) [15]; Climate Prediction Centre (CPC) Morphing technique (CMORPH) [16]; Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) [17]; and its Cloud Classification System variant, PERSIANN-CCS [18]; TAMSAT African Rainfall Climatology And Time-series (TARCAT) [15]; Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) [19]; Multi-Source Weighted-Ensemble Precipitation (MSWEP) [20]; Famine Early Warning System (FEWS) Rainfall Estimate (RFE) [21]; and, African Rainfall Climatology (ARC) [22]. Many of these products merge precipitation radar, passive microwave, and infrared sensors to produce high-resolution data [12,16]. However, the performance of satellite precipitation products (SPPs) must be evaluated to establish their accuracy and quality, as satellite estimates are inherently prone to biases and uncertainties owing to their indirect sensing approach [23]. These biases vary across regions, seasons, terrain, and climatic conditions [24,25]. Therefore, evaluating their performance against ground observations is crucial before they can be reliably used in hydrological research.

Numerous studies have evaluated the performance of SPPs at global and regional scales. For example, Mekonnen et al. [25] evaluated eight gauge-corrected precipitation products across Africa (2001–2020) using in-situ observations and multiple evaluation metrics. While they identified poorer accuracy at daily timescales, they found reliable performance at monthly and annual scales, particularly for RFE v2.0, ARC v2.0, and MSWEP v2.8. However, product performance varied by region, climatic zone, and topographical profile. For example, MSWEP was most reliable in Northern Africa, RFE in Western and Southern Africa, ARC in Central Africa, and CHIRPS in Eastern Africa [25]. In East Africa, several studies have been carried out to assess the performance of several SPPs, such as CHIRPS, TMPA, TRMM, CMORPH, MSWEP, RFE, etc. [4,7–9]. For instance, Omondi et al. [9] evaluated four satellite precipitation estimates (CHIRPS, CMORPH, MSWEP, and RFE) against in-situ data from 20 stations in Kenya's Lake Victoria basin (2012–2018) for crop growth simulation. They found CMORPH and MSWEP performed best, whereas CHIRPS frequently over-detected precipitation. Macharia et al. [8] assessed CHIRPS, TRMM, one gauge-interpolated product (Global Precipitation Climatology Centre (GPCC)) and one re-analysis product (Modern-Era Retrospective Analysis for Research and Application (MERRA)) for (1998–2013) across Kenya's agro-ecological zones and found TRMM performed best in humid zones, GPCC and CHIRPS, in semi-arid areas, and MERRA, in high-elevation areas, stressing the importance of region-specific calibrations. Dinku et al. [26] evaluated four SPPs (CHIRP, CHIRPS, ARC2, and TAMSAT) across East Africa using 1200 stations and found that CHIRP and CHIRPS performed best at dekadal/monthly scales, whereas TAMSAT was superior at the daily scale. Although numerous evaluation studies have been conducted across various regions, localized assessments remain essential, as the choice of an appropriate precipitation product is context-dependent, and validation tailored to specific basins or study scopes is required to address the particular development and water management challenges of a given region.

To address systematic errors and biases in SPPs, various bias correction techniques have been employed, including Quantile Mapping [27,28]; Empirical Cumulative Distribution Function Mapping [29]; Linear Scaling (LS) [30–32]; Bayesian Method (BM) [7]; Power Transformation (PT) [33]; Local Intensity Scaling (LOCI) [34]; the Delta Method (DM) [35]; and Machine Learning (ML) methods [36]. These techniques aim to reduce systematic biases and improve the reliability for hydrological applications. In East Africa, Kimani et al. [7] applied a Bayesian approach to merge historical rain gauge data with CHIRPS v2 precipitation estimates across East Africa (1981–2013), reducing systematic errors by up to 60% and improving correlations by 21%. The approach was effective in correcting magnitude-related biases and is suitable for long-term applications. In Kenya, Vigna et al. [37] compared the Kenyan Meteorological Department (KMD), GPCC, and CHIRPS against historical observations for North Horr Sub-County in northern Kenya and found GPCC to be the most accurate. Using Quantile Mapping, they combined the reliability of GPCC with the KMD to produce a bias-corrected monthly dataset (1983–2014).

Machine learning has recently gained attention as an effective approach for bias correction of satellite precipitation products (SPPs), offering the ability to capture nonlinear relationships, handle high-dimensional datasets, conduct comprehensive analyses of large volumes of data, and identify patterns within the data [38,39]. With respect to machine learning approaches, RF and Long Short-Term Memory Networks (LSTM) have shown particularly strong potential for SPP bias correction [36,40]. Improvements in precipitation estimates alone, however, do not guarantee reliable hydrological simulations; accurate stream discharge measurements are also essential to calibrate models and validate the resulting runoff predictions. Improvements in precipitation estimates alone, however, do not guarantee reliable hydrological simulations; as the primary input to the water balance, corrected precipitation must ultimately be evaluated through its ability to reproduce observed streamflow, making accurate discharge measurements essential for model calibration and validation of runoff predictions.

Traditionally, hydrological models are calibrated using in-situ discharge measurements, but such measurements are unavailable in ungauged basins. Combining remote sensing data with hydrologic or hydrodynamic models presents new opportunities for estimating discharge in such regions, with the added advantage that satellite observations provide

spatially continuous coverage enabling discharge estimation across entire river systems rather than at discrete gauging locations. For instance, calibrating models with altimetry-derived water levels [41] or river width observations [42] has proven effective for estimating continuous daily river discharge. In addition to these approaches, several model-derived products that use large-scale hydrological or land surface models that simulate the movement of water through the terrestrial system based on climatic inputs and physical processes have been developed to provide global and regional estimates of river discharge. Examples include the Global Land Data Assimilation System (GLDAS) [43]; PCRaster Global Water Balance (PCR-GLOBWB) [44]; Water – Global Assessment and Prognosis (WaterGAP) [45]; and European ReAnalysis version 5 – Land component (ERA5-Land) [46]; and Community Water Model (CWatM) [47]. Such discharge models offer consistent, long-term, and spatially continuous discharge estimates, providing valuable alternatives where ground observations are limited or absent. Moreover, they support a wide range of hydrological applications, including water resource management, flood forecasting, and climate impact assessment.

More recently, the VegDischarge v1 dataset has emerged as a novel continental-scale product that integrates vegetation dynamics and hydrological signals to estimate natural river discharge across Africa [48]. VegDischarge is developed through a coupled modeling framework in which gridded runoff, simulated at 1-km resolution by the agro-hydrologic VegET model developed by the U.S. Geological Survey (USGS) [49], is routed through the river network using the vector-based mizuRoute routing model [50], producing discharge estimates for more than 64,000 river reaches across the continent [48]. A distinctive feature of the underlying VegET model is its reliance on remote sensing forcings such as satellite-based precipitation and NDVI-derived land surface phenology, which allows the spatial and temporal variability of hydrological processes to be represented even in regions with sparse ground observations. Covering the period 2001–2021 at monthly and annual time steps, VegDischarge provides continuous, spatially explicit estimates of naturalized flow, and its validation against 242 Global Runoff Data Centre (GRDC) stations demonstrated strong continental-scale performance ($R^2=0.95$, $NSE=0.95$, $KGE=0.84$), supporting its suitability for water resources assessment across diverse African hydrological regimes [48].

Generally, gauge-based datasets are expected to show better performance than satellite and model-based products in data-rich regions; however, satellite and model-based datasets provide near real-time estimates, making them attractive for operational applications [51]. In this context, satellite-derived precipitation and model-derived discharge estimates can effectively substitute sparse gauge data if they are properly validated and corrected. Moreover, combining well-performing satellite or model products with gauge data can yield more robust hybrid datasets [52].

Model-derived discharge remains susceptible to compounded errors arising from both biased precipitation inputs and model parameterizations [53]. As such, they often exhibit systematic biases arising from model structure, input data limitations, or parameterization errors, which affect both the magnitude and variability of simulated streamflow, limiting their reliability for local or regional applications [54]. To address this, bias correction techniques have been commonly applied to align model outputs more closely with observed data. For instance, Hales et al. [55] evaluated the GEOGloWS ECMWF Streamflow hydrologic model by comparing its simulated discharge against observations at almost 13,000 gauging stations globally. They identify that the primary model error is bias in flow magnitudes, particularly overprediction in many locations. To mitigate this, the authors applied a revised SABER (Streamflow Adequacy Bias-correction using Regression/clustering) method, which combines watershed clustering, machine learning, and geospatial/statistical techniques to generalize bias patterns from gauged to ungauged basins and validated the bias-corrected output using performance metrics such as mean error (ME), RMSE, and KGE. Their results showed that bias correction reduced mean error in about 90% of gauges, the median KGE increased to 0.47 and the median RMSE decreased to 64 m³/s, improving overall discharge estimates worldwide.

Despite these methodological advances, the application and evaluation of bias correction techniques, both statistical and machine learning-based, remain limited across many African basins, where sparse ground observations and complex topography challenge the reliability of remote sensing and model-derived datasets. To bridge this gap, the present study

applies and assesses multiple bias correction methods for improving both satellite-derived precipitation and model-simulated discharge in the central highlands of Kenya. The Central Highlands (Mount Kenya) region serves as a vital watershed and agricultural hub for Kenya, supporting biodiversity, farming, and diverse communities. Recognized as a UNESCO World Heritage Site, it is the second-highest peak in Africa. This region is also a crucial water source, supplying several rivers which are essential for agriculture, wildlife, and human needs in Kenya. Given the ecological and socio-economic importance of the region, accurate precipitation information is indispensable for water resource planning and management. Specifically, this study seeks to: (i) comparatively evaluate five SPPs; Climate Hazards Group Infrared Precipitation with Stations (CHIRPS); Multi-Source Weighted-Ensemble Precipitation (MWSEP); Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN); its Cloud Classification System variant (PERSIANN-CCS); Tropical Applications of Meteorology using Satellite data and ground-based observations (TAMSAT), against in-situ data, (ii) assess the performance of four bias correction techniques in improving the satellite precipitation estimates of the best performing SPP, and (iii) validation of discharge against in-situ observations and assess the performance of the four bias correction techniques in improving model-derived VegDischarge data. In view of evaluating the SPPs, the study aims to unify all the above results by analysing the SPPs against available daily and monthly gauged precipitation data for 10 years (2012–2021). This study makes several novel contributions:

- (i) it focuses on the Central Highlands region of Kenya, a critical yet underexplored area where reliable precipitation information is vital for agriculture, water management, and biodiversity conservation;
- (ii) unlike previous studies that emphasize only a few widely used datasets such as IMERG, CMORPH, or TRMM, this research evaluates a broader set of satellite precipitation products, including PERSIANN-CCS and PERSIANN-PDIR, which remain scarcely validated in the region;
- (iii) thirdly, it extends bias correction beyond precipitation datasets to the VegDischarge product, offering one of the first attempts to correct model-derived discharge data against observed streamflow in East Africa.

2. Study area

This study focuses on the Central Highlands region of Kenya (latitude 1.00°S - 3.00°N; longitude 35.00°-39.00°E), which encompasses an area of 52,064 km². [Fig 1](#) shows the location, relief and drainage of the Central Highlands study area in Kenya. The region's ecosystem exhibits significant diversity, with afro-alpine vegetation at higher altitudes transitioning into montane forests, which gradually give way to savannas and grasslands at lower elevations [56]. The Mount Kenya forest ecosystem functions as a vital water tower, supplying over 40% of Kenya's freshwater that supports urban consumption, agriculture, and hydroelectric power [57]. Beyond water provision, it plays a critical role in climate regulation, carbon sequestration, and water catchment functions [57]. Temperatures in the region range from around 5 °C in the upper elevations above 3,000 m asl to about 25 °C in the lower areas, with an average decline of approximately 0.6 °C for every 100 m gain in altitude [57]. The Central Highlands experiences humid to sub-humid tropical highland climate with a bimodal rainfall regime. The long rains typically occur from March to May, peaking in April, while the short rains extend from October to December, with a peak in November. Annual rainfall generally ranges from approximately 700 mm in the lower-elevation areas to over 2,000 mm in the highlands surrounding Mount Kenya, reflecting strong orographic influences. The relatively dry periods occur between January–February and June–September, although the timing and duration of the rainy seasons vary with elevation and local topography [58].

This region encompasses parts of eleven counties surrounding Mount Kenya: Samburu, Laikipia, Meru, Isiolo, Nakuru, Nyandarua, Nyeri, Murang'a, Embu, Kirinyaga, and Tharaka-Nithi, as shown in [Fig 1](#). Counties such as Samburu, Isiolo, and Laikipia are primarily characterized by shrubland, whereas Embu, Tharaka-Nithi, Nyandarua, Nakuru, and Meru feature extensive rainfed croplands, which play a significant role in regional food security. Moreover, Meru, Nyeri,

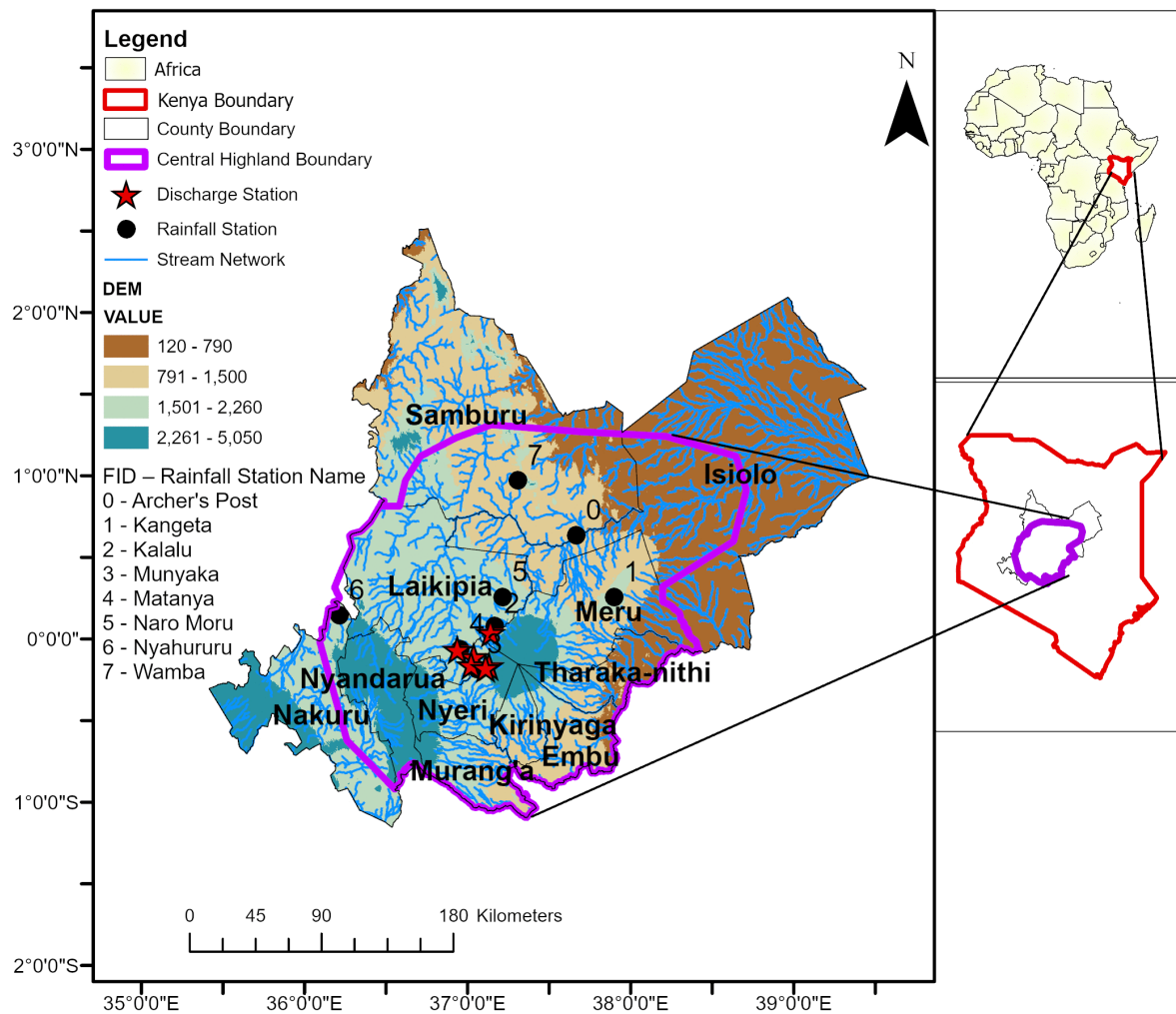


Fig 1. Central Highlands region showing the location of gauge stations and counties. Insets display the position of Kenya within Africa (top right) and the region within Kenya (bottom right). Map created using ArcGIS Pro. Base map shapefiles were obtained from Natural Earth @ <https://www.naturalearthdata.com/> (<http://www.naturalearthdata.com/about/terms-of-use/>), for continental and national boundaries, and openAFRICA (<https://open.africa/dataset/kenya-counties-shapefile>, CC BY 4.0) for Kenya county administrative boundaries.

<https://doi.org/10.1371/journal.pwat.0000623.g001>

and Tharaka-Nithi have notable irrigated agricultural areas of 32,821ha, 44,794ha and 34,201ha, respectively. Forested regions are predominantly found in Nyeri, Embu, and Kirinyaga, where both closed and open forests contribute to ecological balance [56].

3. Methodology

The methodological process undertaken in this study can be divided into five main components: data collection, data cleaning and pre-processing, validation, bias correction, and performance evaluation, as illustrated in Fig 2. Data collection involved compiling in situ, satellite, and model-derived datasets to capture precipitation and discharge variability, followed by cleaning and pre-processing to ensure consistency. Validation compared satellite/modelled and observed values, bias correction minimised systematic errors, and performance evaluation quantified overall improvements in accuracy.

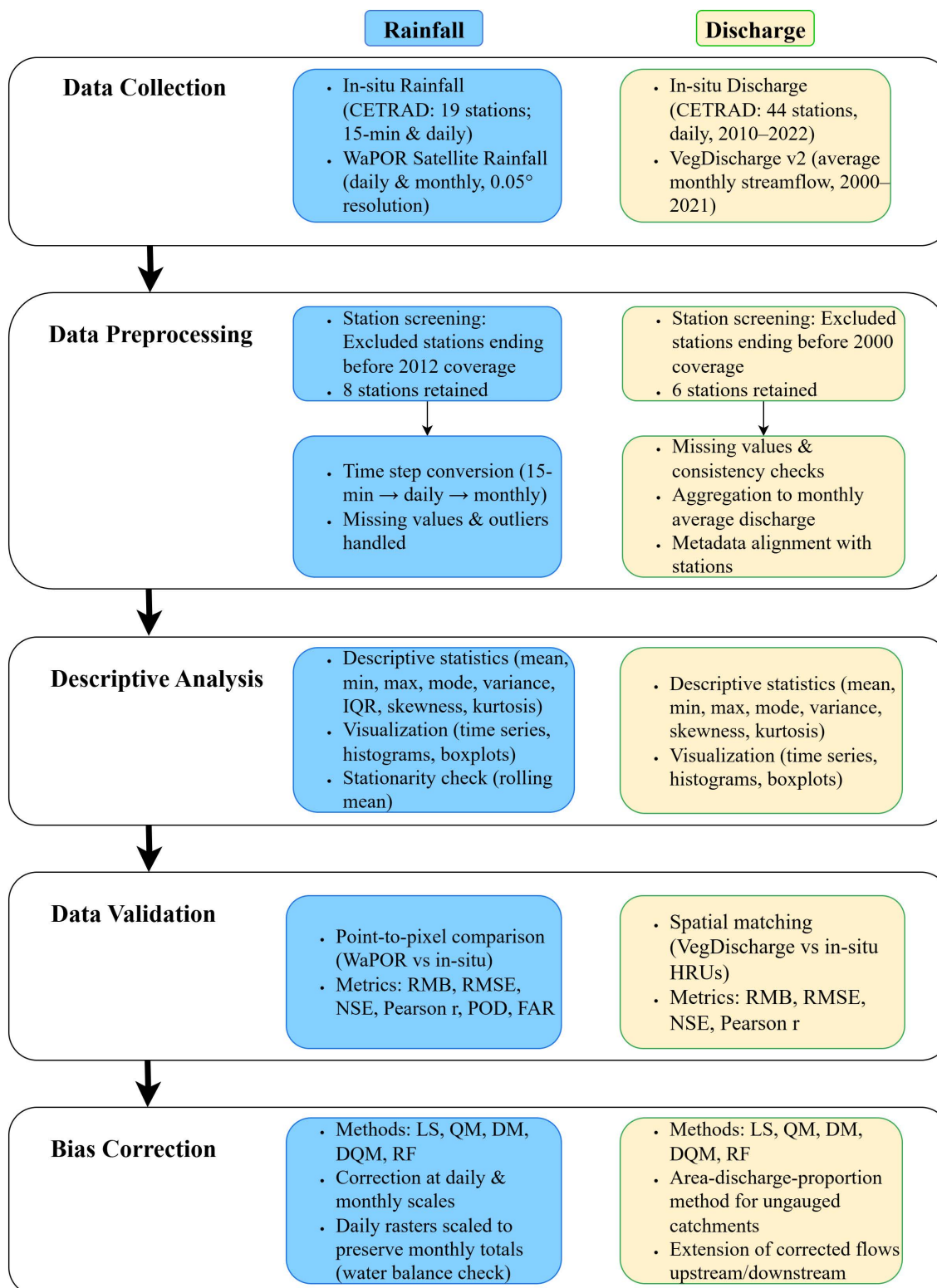


Fig 2. Methodological roadmap for the study.

<https://doi.org/10.1371/journal.pwat.0000623.g002>

3.1. Datasets used

The datasets used in this study were derived from a combination of ground-based observations, satellite-derived products, and model-generated outputs. The specific spatial and temporal characteristics of these datasets are detailed in [Table 1](#). Ground observations provided reference data, while satellite and model products offered continuous spatial representation and consistent temporal records of Precipitation and Stream discharge.

3.1.1. Precipitation data. Precipitation data and locations of in-situ gauge stations for the Central Highlands region were obtained from the Centre for Training and Integrated Research in ASAL Development (CETRAD, where ASAL stands for arid and semi-arid lands) through its Social Hydrological Information Platform (SHIP). However, only eight (8) in-situ gauge stations had historical data from 2012 to 2021. Consequently, these eight stations were selected for the bias correction and validation of the satellite precipitation products. The in-situ observations comprised measurements from both automated and manual rain gauges, recorded at daily or 15-minute intervals. The selected rainfall stations are distributed across a range of physiographic settings within and around the Central Highlands, enabling the network to capture variations in precipitation associated with differences in elevation and topography. As shown in the study area map ([Fig 1](#)), the stations extend from the high-elevation regions of Nyandarua, Nyeri, and Laikipia to the lower-elevation, semi-arid areas of Samburu and Isiolo. Despite covering different physiographic zones, the rainfall stations are relatively sparse considering the complexity of the terrain. Consequently, the existing network may not adequately capture fine-scale variations in precipitation, especially in areas where stations are widely spaced. Furthermore, with only eight stations available for analysis, some portions of the northeastern and peripheral regions remain underrepresented, potentially limiting the ability of the rainfall observations to fully characterize the spatial distribution of precipitation across the entire Central Highlands.

The various satellite precipitation data used in this study are CHIRPS v2, PERSIAN-CCS, PERSIANN-PDIR, TAMSAT, and MWSEP v2. Satellite precipitation products (SPPs) are subject to inherent uncertainties arising from their indirect rainfall approaches. SPPs infer rainfall primarily from infrared cloud-top and/or microwave measurements, introducing structural uncertainties that vary with rainfall type and algorithm [28]. Infrared-based approaches can have difficulty representing warm-cloud and non-convective precipitation, whereas microwave measurements may introduce biases in intense convective systems [28]. These limitations are often amplified in mountainous regions, where complex terrain and

Table 1. Summary of datasets used.

Dataset Type	Source	Variables	Temporal & Spatial Resolution	Reference
In-situ Precipitation	CETRAD https://wlrc-ken.org/	Precipitation (mm)	15-minutes, daily	https://wlrc-ken.org/
In-situ Discharge	CETRAD https://wlrc-ken.org/	River discharge (m ³ /s)	Daily	N/A
Station Metadata	CETRAD https://wlrc-ken.org/	Coordinates, shapefiles	N/A	N/A
Satellite Precipitation	CHIRPS v2 https://data.chc.ucsb.edu/products/CHIRPS-2.0/	Precipitation (mm)	0.05°, daily, monthly	[19]
	PERSIANN PDIR https://chrsdata.eng.uci.edu/	Precipitation (mm)	0.04°, daily, monthly	[17]
	PERSIANN CCS https://chrsdata.eng.uci.edu/	Precipitation (mm)	0.04°, daily, monthly	[17]
	TAMSAT https://data.tamsat.org.uk/	Precipitation (mm)	0.0375°, daily, monthly	[15]
	MSWEP v2 https://www.gloh2o.org/	Precipitation (mm)	0.1°, daily, monthly	[20]
Model Discharge	VegDischarge Model v2	River discharge (m ³ /s)	Monthly average	[49]

<https://doi.org/10.1371/journal.pwat.0000623.t001>

orographic processes are not fully resolved by satellite algorithms, resulting in systematic errors even after bias correction [12,17,23]. Performance also varies across temporal scales, seasons, climatic regimes, and geographic regions, meaning that no single SPP consistently provides optimal estimates under all conditions [37]. In addition, limited information on precipitation extremes can constrain the suitability of some SPPs for flood-related applications, particularly in high-rainfall and topographically complex environments [21,27]. Because precipitation estimates are subsequently used as inputs to hydrological and other environmental models, errors can propagate through simulations, and satisfactory performances for one application may not necessarily be comparable to another application [25].

3.1.2. Discharge data. In-situ discharge measurements from 44 stations were collected from CETRAD. However, data extending beyond 2010 was available for only 6 stations, so only these were used for further analysis. The in-situ discharge data are available as daily averages in cubic meters per second (m^3/s), providing the reference data against which the model-derived discharge was evaluated. The selected streamflow gauging stations are concentrated primarily within the central highland region, where the major river headwaters originate. The discharge network provides less comprehensive spatial coverage than the rainfall network, as they are clustered within the central headwater region, leaving the northern, eastern, and western portions of the study area with little or no streamflow monitoring. As a result, hydrological responses in lower-elevation catchments, semi-arid environments, and downstream river reaches remain comparatively poorly characterized because of the limited availability of direct streamflow observations. This study used VegDischarge v2, a high-resolution dataset providing naturalized monthly discharge estimates (m^3/s) for more than 650,000 river segments across Africa for the period 2000–2021, a 10-fold higher resolution from version 1 [48]. The dataset is developed through a coupled modeling framework in which runoff simulated by the agro-hydrologic VegET model is routed through the Multi-Error-Removed Improved-Terrain (MERIT) Hydro high-resolution river network [59] using the vector-based mizuRoute routing model [50]. By combining remote sensing and hydrological modeling, VegDischarge fills critical data gaps in small catchments and seasonal rivers where in-situ observations are limited [48]. A major strength of the product lies in its use of satellite-observed vegetation indices (e.g., NDVI) to dynamically constrain evapotranspiration and soil moisture, establishing a physically consistent link between land surface processes and runoff generation [50]. This vegetation-driven approach enhances hydrological realism in data-scarce regions where meteorological or ground-based inputs are sparse a persistent limitation of many global hydrological models (GHMs) and reanalysis products such as GEOGLOWS, MERRA, and GLDAS. Furthermore, its open-access, continental-scale framework supports near-real-time monitoring and independent validation against in-situ discharge observations, offering improved spatiotemporal consistency and flexibility for regional hydrological applications across Africa. Model-derived discharge data inherit and often amplify the uncertainties of their precipitation inputs, since errors in the rainfall forcing propagate through the rainfall–runoff process and can dominate overall streamflow uncertainty [53,54]. Even purpose-built high-resolution products such as VegDischarge still require substantial regionalization or bias correction before their simulated flows can be considered reliable at the local scale [48,55].

3.2. Data cleaning and preprocessing

Data cleaning was carried out to ensure accuracy, completeness, and consistency across precipitation datasets. The procedure addressed missing values, outliers, and temporal inconsistencies between automatic and manual stations, which were standardised to daily and monthly intervals using Python's Pandas library. Descriptive statistics, including measures of central tendency, dispersion, and distribution, were then applied to summarise the datasets, identify seasonality and trends, and prepare the data for subsequent statistical analysis and bias correction. In addition to numerical summaries, data visualisation was used to reveal characteristics of the datasets requiring further attention before analysis [58]. Time-series line plots (S1 Fig) illustrated the chronological progression of precipitation, enabling the identification of patterns, seasonal behaviour, and irregularities.

3.3. Data validation

3.3.1. Comparison of Satellite Precipitation Products (SPPs) with in-situ precipitation data. Satellite precipitation products (SPPs) are commonly validated against in-situ gauge data using two approaches: point-to-pixel and pixel-to-pixel comparisons [51]. The point-to-pixel method directly compares rain gauge data with satellite pixel values [6], while the pixel-to-pixel method interpolates gauge data to match the satellite resolution [60]. Given that interpolation can smooth extreme precipitation values and reduce gauge accuracy [61], this study applied the point-to-pixel approach to preserve original gauge values. The extracted satellite values were paired with the corresponding in-situ observations at each rainfall station, and all statistical metrics were computed independently for each station. Validation covered the overlapping period 2012–2021, with satellite grid values extracted at gauge locations on daily and monthly scales. Since satellite estimates reflect spatial averages, some systematic differences from point observations are expected [11].

Two statistical approaches were applied for evaluation: pairwise comparison statistics and categorical validation statistics [11]. Pairwise measures included Relative Mean Bias (RMB), which quantifies systematic over- or underestimation [60]; Root Mean Square Error (RMSE), which emphasizes large errors [51]; Nash–Sutcliffe Efficiency (NSE), which assesses skill relative to the gauge mean [11]; and the Pearson correlation coefficient (CC), which measures linear association [62]. For categorical validation, Probability of Detection (POD) and False Alarm Ratio (FAR) [63] were calculated using daily precipitation data to assess rain detection skill, specifically targeting a threshold of > 0.1 mm. Together, these six metrics (Table 2) provided a comprehensive evaluation of the precipitation products. Based on the collective performance across all metrics, the best-performing satellite precipitation product, which demonstrates the highest consistency with in-situ data, was selected for subsequent bias correction and further analyses.

3.3.2. Comparison of VegDischarge with observed data. The validation of VegDischarge data against in-situ river discharge observations involved spatially matching gauge locations to VegDischarge river network segments. Each VegDischarge river segment is associated with a Hydrological Response Unit (HRU) identified by a unique segment ID. The coordinates of in-situ gauging stations were overlain on the VegDischarge dataset to determine the HRUs in which each station is located. The corresponding HRU ID was then linked to the appropriate segment ID, and the discharge time series of that segment was extracted for the station location. Validation was carried out over the overlapping period of 2000 – 2021, using both statistical performance metrics (Table 2), which together quantified accuracy, bias, error magnitude, predictive skill, and linear association between simulated and observed discharges, as well as visual

Table 2. Performance Indicators for validation.

Name	Parameter Evaluated	Formula	Perfect Score
Relative Mean Bias	Rainfall Discharge	$RMB = \frac{S - \bar{G}}{\bar{G}}$	0
Root Mean Square Error (RMSE)	Rainfall Discharge	$RMSE = \sqrt{\frac{\sum_{i=1}^N (S_i - G_i)^2}{N}}$	0
Nash–Sutcliffe Efficiency coefficient (NSE)	Rainfall Discharge	$Eff = 1 - \frac{\sum (S - G)^2}{\sum (\bar{G} - \bar{G})^2}$	1
Pearson correlation coefficient (CC)	Rainfall Discharge	$r = \frac{\sum (\bar{G} - \bar{G})(\bar{S} - \bar{S})}{\sqrt{\sum (\bar{G} - \bar{G})^2} \sqrt{\sum (\bar{S} - \bar{S})^2}}$	1
Probability of Detection (POD)	Rainfall	$POD = \frac{H}{H+M}$	1
False Alarm Ratio (FAR)	Rainfall	$FAR = \frac{F}{H+F}$	0

^aFor continuous statistics; $\bar{G}$ = in situ measurement, G_i = in situ measurement value at time step i , $\bar{G}$ = average in situ measurement, S = satellite estimate, S_i = satellite estimate value at time step i , $\bar{S}$ = average satellite estimate, and N = number of data pairs.

^bFor categorical statistics; H = number of hits, F = number of false alarms, and M = number of misses.

^cThese statistical metrics were estimated using the numpy [64], scipy.stats [65], sklearn.metrics [66], and hydroerr [67] packages in python programming.

<https://doi.org/10.1371/journal.pwat.0000623.t002>

comparisons. Time series plots were generated to compare the temporal patterns of VegDischarge estimates with in-situ discharge, allowing for assessment of seasonal dynamics, anomaly capture, and potential systematic deviations.

3.4. Bias correction of precipitation

Four bias-correction (BC) methods (Table 3) were evaluated: three conventional statistical bias correction approaches, namely Linear Scaling (LS), Quantile Mapping (QM), and the Delta Method (DM), and one machine-learning approach, Random Forest (RF).

These methods were selected because they represent a diverse set of widely used bias-correction approaches, encompassing simple scaling, distribution-based adjustment, change-preserving techniques, and machine-learning-based corrections. The bias correction (BC) methods were applied station-wise to correct satellite rainfall data at both daily and monthly temporal scales for the period 2012–2021.

At the daily time scale, the corrected daily precipitation rasters were scaled against the corresponding bias-corrected monthly rasters to ensure consistency between temporal resolutions, such that the sum of the daily values matched the corrected monthly totals. The daily bias-corrected rasters were summed up pixel-wise for each month, and a scaling factor was computed at the pixel level as:

$$F_s = \frac{\sum P^{corr}(m)}{\sum P^{corr}(d)}, \quad (5)$$

Where, F_s is the scaling factor, $P^{corr}_{(m)}$ is the corrected monthly precipitation, and $P^{corr}_{(d)}$ is the corrected daily precipitation. Finally, each corrected daily precipitation was adjusted by multiplying it by the scaling factor. This step ensured the preservation of the monthly water balance and prevented the propagation of cumulative errors.

Table 3. Bias correction techniques used.

Method	Formula	Reference
Linear Scaling (LS)	$P^{corr}_{(t)} = P^{sat}_{(t)} \times \frac{\mu_m(P^{obs}_{(t)})}{\mu_m(P^{sat}_{(t)})} \quad (1)$	[29]
Quantile Mapping (QM)	$P^{corr}_{(t)} = F^{-1}_{obs} (F_{sat} (P^{sat}_{(t)})) \quad (2)$	[30]
Delta Method (DM)	$P^{corr}_{(t)} = P^{sat}_{(t)} \times \frac{\mu_{obs}}{\mu_{sat}} \quad (3)$	[34]
Random Forest (RF)	$f_{RF}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (4)$	[35]

Where:

F^{-1}_{obs} : Inverse of empirical cumulative distribution functions of observed precipitation

F_{sat} : Cumulative distribution functions of satellite-based precipitation

$P^{corr}_{(t)}$: Bias-corrected precipitation at time t

$P^{sat}_{(t)}$: Uncorrected (satellite) precipitation at time t

$P^{obs}_{(t)}$: Observed precipitation at time t

μ_{obs} : Mean of observed precipitation

μ_{sat} : Mean of modelled precipitation

μ_m : The long-term monthly mean of precipitation of that corresponding month

$f_{RF}(x)$: The final corrected value for a specific input x

x: Input vector/predictor

B: Number of trees

b: A counter representing each individual tree, from the 1st to the B-th

$T_b(x)$: The prediction made by a single decision tree b based on a random subset of the data

<https://doi.org/10.1371/journal.pwat.0000623.t003>

3.5. Bias correction of model-derived discharge

The same four bias-correction techniques were used to correct the VegDischarge data. However, the discharge-area proportion method was used to estimate monthly discharges for ungauged upstream and downstream segments [68]. For each observed station, its corresponding upstream catchment area and downstream catchment area were identified, and the discharge for each monthly column was scaled using the area ratio:

$$Qu_t = \frac{Au}{Ag} Qg_t, \quad (6)$$

where Qu_t is the streamflow on day t at the ungauged catchment, Qg_t is the streamflow on day t at the reference stream gauge, Au is the drainage area of the ungauged catchment, and Ag is the drainage area to the reference stream gauge. This procedure combines the bias correction methods with a physically based area-proportion method for spatial extrapolation. The area-based estimation extends the utility of the corrected data to ungauged segments.

3.6. Evaluation of bias-corrected datasets

The Statistical Evaluation metrics described in Table 2 were used for evaluating the overall performance of the bias correction methods, as they measure the deviation between the bias-corrected and reference data. These statistical metrics have been widely used, often together, to evaluate the degree of agreement between observed and bias-corrected data. We also used a Taylor diagram to compare datasets produced by different bias correction methods, providing a concise summary of how well the corrected precipitation aligns with gauge observations in terms of spatial pattern, variability, and the magnitude of that variability.

As an assessment of the physical consistency of the bias-corrected datasets, an annual water-balance check was conducted using the bias corrected precipitation (P), and discharge (Q), together with the remaining hydrological fluxes and storage-related components of the system. The resulting annual change in storage was positive for all evaluated years, with values of 2.8, 4.8, 8.3, and 8.1 km³ yr⁻¹ for 2018, 2019, 2020, and 2021, respectively. These results indicate a net annual water-balance residual across the study period. The magnitude of these residuals was assessed relative to the corresponding annual water fluxes to evaluate the overall closure of the water balance. This provides an assessment of the physical plausibility of the bias-corrected datasets, although it does not necessarily rule out imbalances at shorter temporal scales or under extreme hydrological conditions.

4. Results

4.1. Preprocessing and descriptive analysis

Fig 3 and Fig 4 provide a summary of the basic descriptive statistics for the monthly precipitation in millimetres (mm) and average monthly discharge measurements in cubic meters per second (m³/s), respectively, before bias correction. Overall, the boxplots reveal strong seasonality, substantial interannual variability, and marked spatial heterogeneity in rainfall regimes. Across most stations, rainfall exhibits a bimodal to trimodal seasonal structure, with pronounced peaks during the long rains (March–May) and a secondary increase during the short rains (October–November). Dry conditions dominate January–February and July–September, although the magnitude and persistence of dryness vary by location (S1 Fig). Archer's Post and Wamba are predominantly dry with occasional heavy bursts; Kalalu, Matanya, Munyaka, and Nyahururu are moderately wet; Kangeta is generally wet with substantial variability and large-event potential; and Naro-Moru is dominated by rare, extreme outliers that overwhelm central tendencies.

Discharge follows a distinct seasonal cycle that closely mirrors the rainfall regime (S2 Fig). Low flows dominate the dry months (January–March and July–September), while marked increases occur during the long rains (April–May) and the short rains (October–November). Peak flows are consistently observed in April–May, with a secondary peak in November.

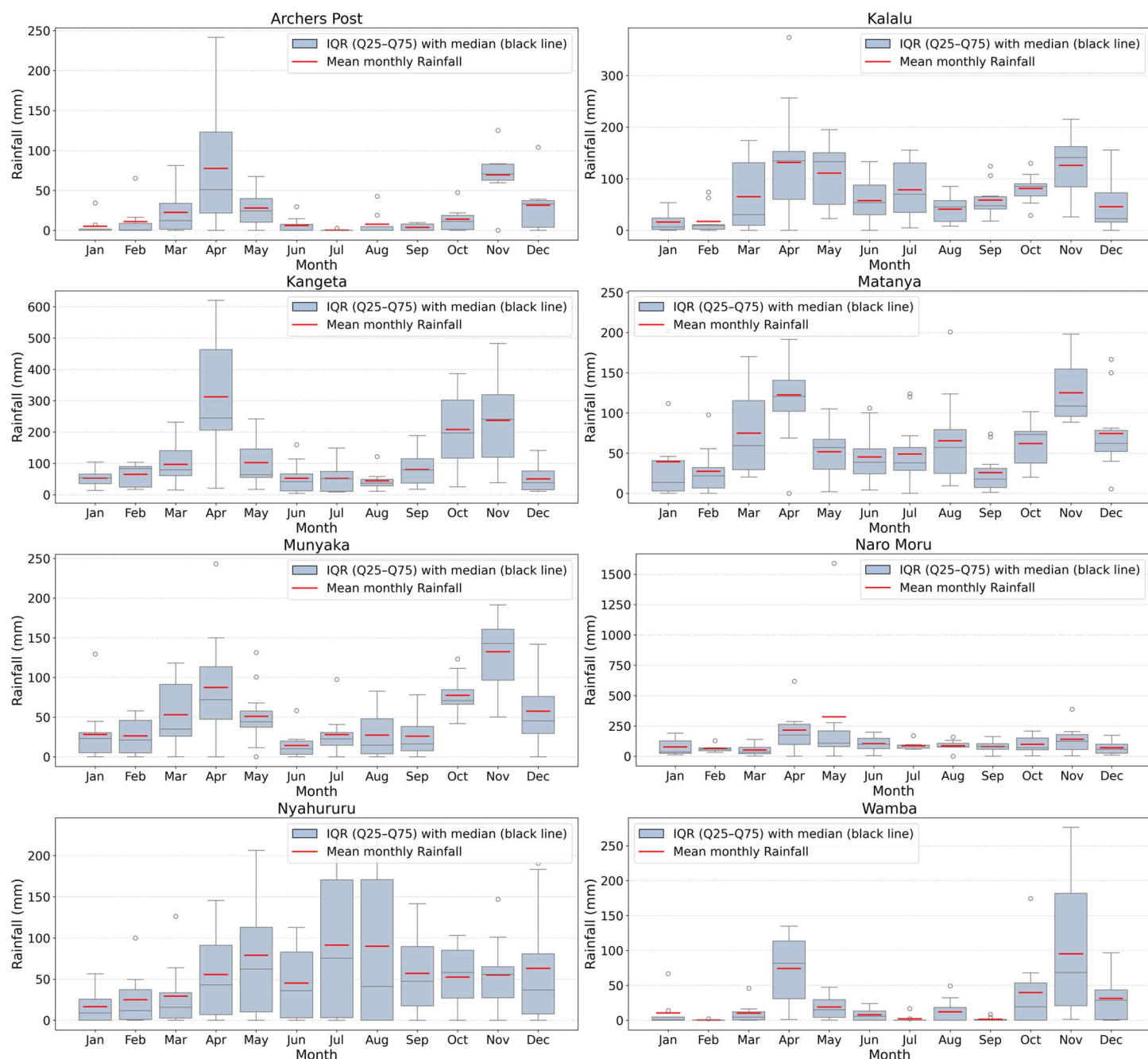


Fig 3. Monthly boxplots of rainfall representing the monthly mean rainfall computed across all years.

<https://doi.org/10.1371/journal.pwat.0000623.g003>

A3 Naro Moru shows moderate baseflow with sharp increases and multiple high-flow outliers during the long rains and a secondary peak in the short rains, indicating episodic runoff generation. A4 Naro Moru is dominated by low flows for most of the year, yet occasional pronounced outliers during wet months reveal threshold-driven responses to intense storms. A5 Naro Moru displays consistently higher discharge and variability, with frequent and very large outliers during both rainy seasons, reflecting a highly responsive catchment prone to flood events. A6 Naro Moru is characterized by low median

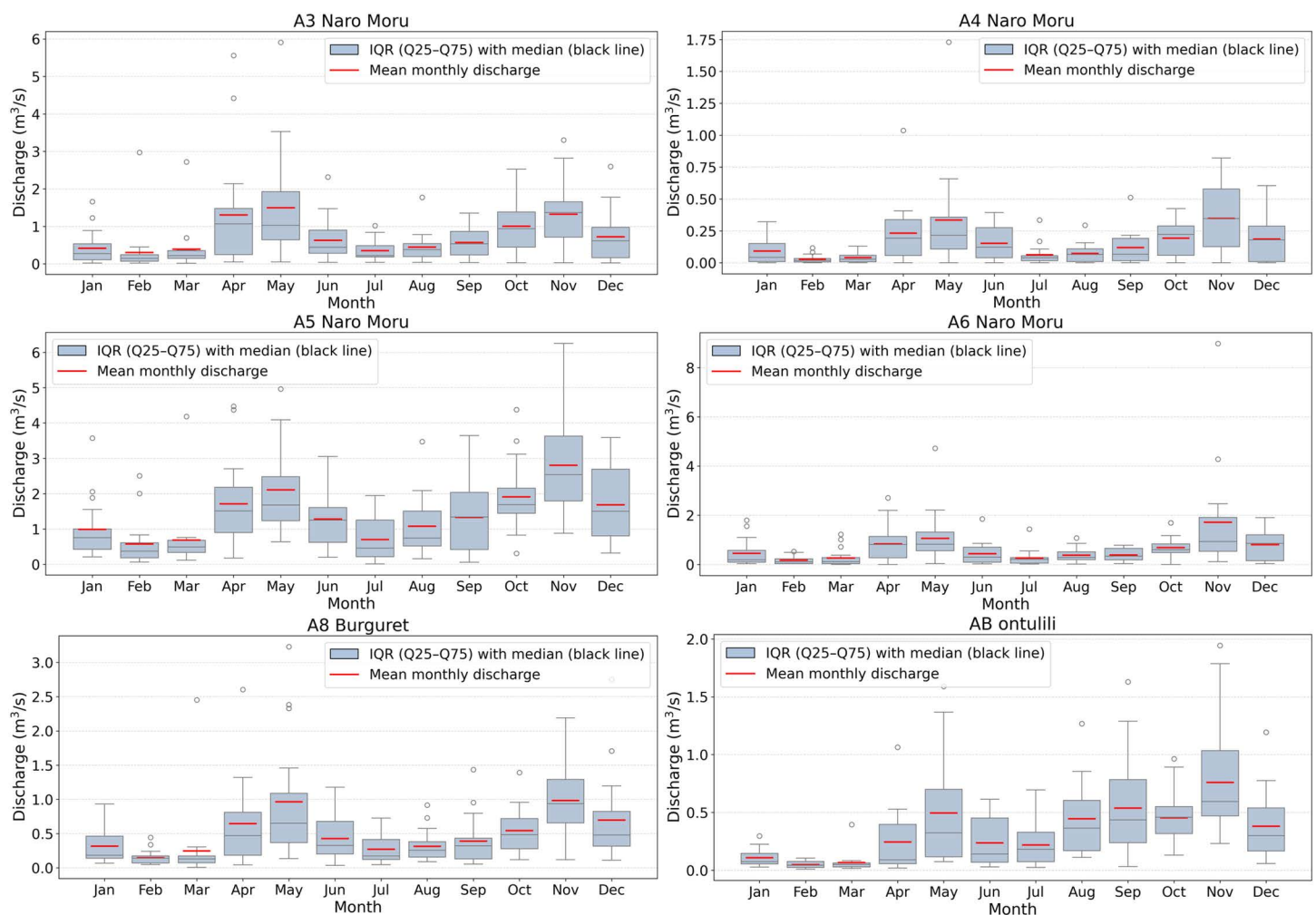


Fig 4. Monthly boxplots of stream discharge representing monthly mean discharge computed across all years.

<https://doi.org/10.1371/journal.pwat.0000623.g004>

flows but exceptionally large outliers, particularly during peak rainfall months, indicating rare but severe flood pulses that strongly skew monthly means. A8 Burguret shows moderate seasonal increases with occasional high-end outliers during wet months, suggesting episodic but less extreme runoff events. AB Ontulili exhibits sustained discharge increases and repeated outliers from the long into the short rains, indicating prolonged wet-season influence combined with event-scale flow extremes.

4.2. Performance of satellite precipitation products against in-situ data

The accuracy of precipitation magnitudes and the representation of extreme events varied across products and stations. The evaluation of the SPPS was done at daily, monthly, annual and seasonal scales. The seasonal scale comprised the major rainy season spanning March, April and May (MAM), and the minor rainy season spanning October, November and December (OND). Fig 5 shows the comparison of satellite-derived rainfall products with in situ observations using scatterplots at each station. Among the evaluated products, CHIRPS and MSWEP demonstrate the closest agreement with in-situ data, exhibiting lower scatter and reduced deviation from the 1:1 line across multiple stations (e.g., Kangeta,

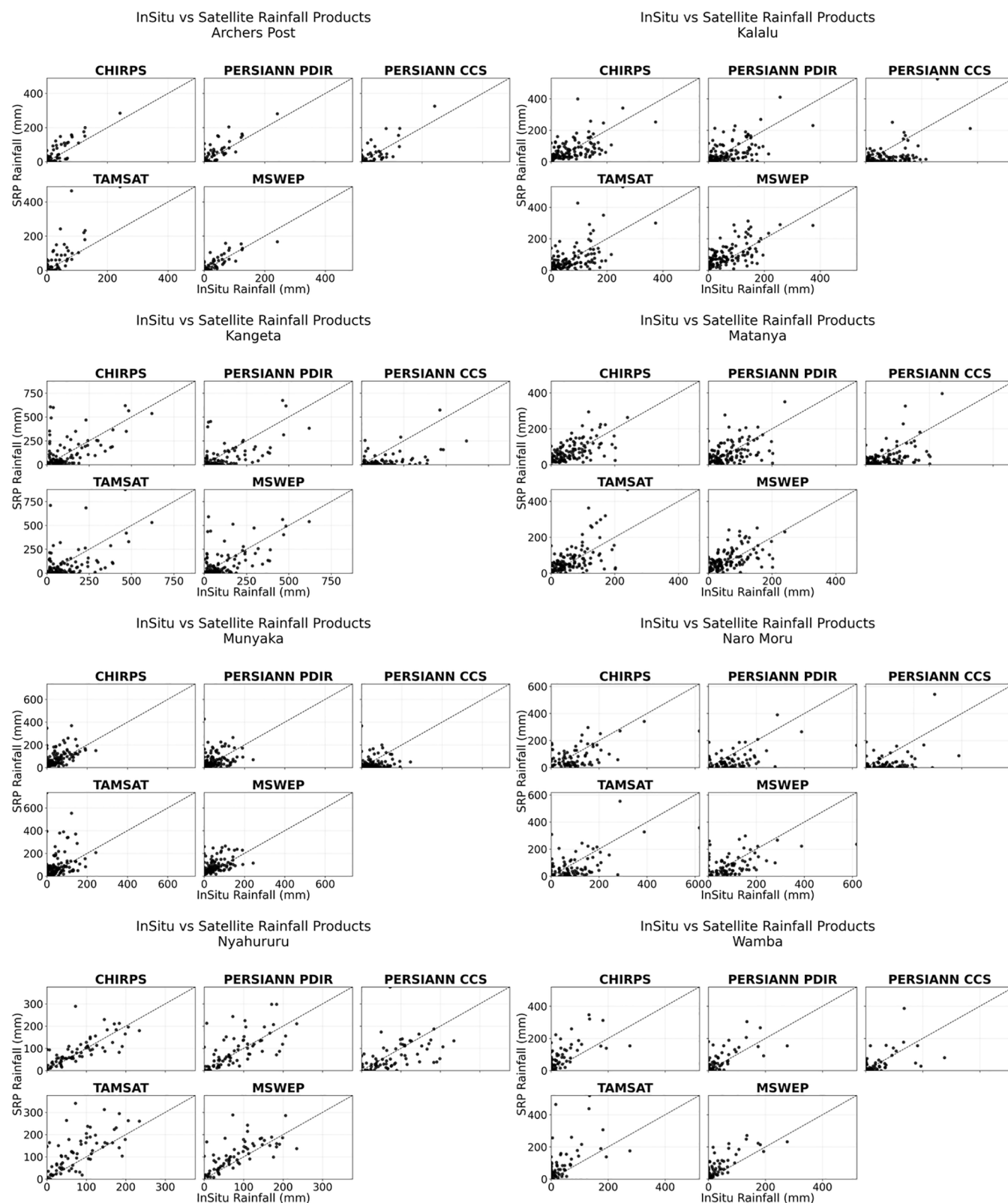


Fig 5. Scatterplots of in situ versus satellite rainfall products, showing mean monthly rainfall values for all stations.

<https://doi.org/10.1371/journal.pwat.0000623.g005>

Nyahururu, and Wamba). In contrast, PERSIANN-CCS shows the widest dispersion and most pronounced underestimation at higher rainfall regimes, while PERSIANN-PDIR and TAMSAT perform intermediately. The comparison also shows that station-specific characteristics shaped performance. For example, in the high-precipitation and high-elevation stations such as Naro Moru and Nyahururu, all products performed worse, as seen in the metrics (low or negative NSE values and reduced CC). Conversely, in drier or semi-arid stations such as Archer's Post and Wamba, products aligned better with in-situ data, which is reflected in higher CC values and more moderate bias values.

At the monthly evaluation scale, CHIRPS performs reliably overall, with moderate-to-high correlation (mean CC = 0.64), modest RMB (mean = 0.28), and low-to-moderate RMSE (mean = 65.68 mm/year). It shows positive NSE at some stations (e.g., Archer's Post, Nyahururu) but negative NSE at more variable sites (e.g., Kangeta, Wamba), where NSE values are negative (Table 4), indicating poor representation of variability. The positive RMB at most stations (e.g., 0.47 at Archer's Post, 0.43 at Munyaka) confirms that CHIRPS slightly overestimates precipitation totals, although underestimation occurs

Table 4. Results of monthly satellite precipitation validation.

Satellite product	Error Metric	Archers Post	Kalalu	Kangeta	Matanya	Munyaka	Naro Moru	Nyahururu	Wamba
CHIRPS	RMSE	34.85	60.04	134.26	50.68	55.27	84.4	43.57	62.33
	NSE	0.3	0.1	-0.24	0.07	-0.94	0.17	0.44	-0.58
	CC	0.87	0.55	0.52	0.59	0.61	0.54	0.71	0.72
	RMB	0.47	0.05	0.09	0.17	0.43	-0.21	0.10	1.12
	POD	0.17	0.23	0.24	0.26	0.22	0.22	0.37	0.42
	FAR	0.80	0.67	0.64	0.65	0.77	0.61	0.61	0.71
PERSIANN-CCS	RMSE	34.34	75.5	129.82	61.28	65.43	204.09	78.28	46.4
	NSE	0.21	-0.43	-0.16	-0.36	-0.86	-0.21	-0.72	0.04
	CC	0.75	0.45	0.51	0.47	0.18	0.18	0.21	0.58
	RMB	0.16	-0.56	-0.68	-0.40	-0.39	-0.70	0.18	0.01
	POD	0.25	0.17	0.19	0.23	0.20	0.13	0.44	0.40
	FAR	0.82	0.58	0.65	0.59	0.69	0.69	0.67	0.66
TAMSAT	RMSE	66.55	69.38	137.46	63.5	113.06	186.89	74.84	86.12
	NSE	-1.69	-0.2	-0.3	-0.46	-4.56	-0.01	-0.66	-2.3
	CC	0.81	0.56	0.5	0.56	0.28	0.29	0.6	0.62
	RMB	0.85	0.00	-0.19	0.16	0.79	-0.36	0.52	1.53
	POD	0.45	0.35	0.28	0.41	0.35	0.25	0.65	0.57
	FAR	0.81	0.63	0.70	0.63	0.74	0.72	0.70	0.75
PERSIANN-PDIR	RMSE	35.98	63.12	121.83	57.63	72.78	185.28	85.04	45.02
	NSE	0.13	0.02	0.01	-0.2	-1.26	0.01	-1.09	0.14
	CC	0.78	0.54	0.57	0.47	0.25	0.35	0.34	0.73
	RMB	0.59	-0.16	-0.33	-0.01	0.25	-0.54	0.63	0.56
	POD	0.29	0.29	0.29	0.35	0.31	0.21	0.59	0.52
	FAR	0.88	0.63	0.72	0.69	0.78	0.77	0.72	0.76
MWSEP	RMSE	24.17	63.54	125.66	53.52	66.88	90.09	54.60	48.55
	NSE	0.65	0.01	-0.06	-0.03	-0.91	0.05	0.14	0.00
	CC	0.84	0.63	0.56	0.54	0.47	0.45	0.71	0.81
	RMB	0.17	0.38	-0.05	0.21	0.72	-0.17	0.32	1.04
	POD	0.59	0.73	0.29	0.67	0.73	0.51	0.84	0.96
	FAR	0.74	0.68	0.83	0.65	0.78	0.62	0.65	0.72

<https://doi.org/10.1371/journal.pwat.0000623.t004>

at some sites (e.g., Kalalu, RMB=0.05). Despite these limitations, the relatively low RMSE at several stations and good CC suggest that CHIRPS is one of the more reliable products in this comparison. Looking at the median values across all stations (Table 5), CHIRPS recorded a median RMSE of 57.5 mm/year, a mean CC of 0.60, and an overall RMB of 0.14, reflecting moderate magnitude error, good temporal agreement, and generally modest bias. The median NSE was 0.08, indicating that although CHIRPS captures variability well, magnitude errors remain at some locations. Similarly, MSWEP consistently performs reasonably well. This is reflected in generally low RMSE (e.g., 24.17 mm/year at Archers Post) and relatively high NSE (0.65 at Archers Post, 0.14 at Nyahururu), with many stations outperforming the other products. The CC are also relatively high, especially at Wamba (0.81) and Archer's Post (0.84). RMB values are mostly modest, though some overestimation appears at Wamba (1.04) and Munyaka (0.72). MSWEP also demonstrates good rainfall-detection skill, as indicated by higher POD values (e.g., 0.96 at Wamba, 0.84 at Nyahururu). This is countered by a relatively high False Alarm Ratio (FAR), which ranges from 0.62 at Naro Moru to 0.83 at Kangeta, suggesting the product frequently overestimates the occurrence of precipitation events. Across all stations, MSWEP achieved a median RMSE of 59.07 mm/year, a mean CC of 0.60, an overall RMB of 0.27, and a median NSE of 0.01.

PERSIANN-CCS shows inconsistent performance with frequent underestimation at some stations (e.g., RMB -0.56 at Kalalu, -0.70 at Naro Moru) and overestimation at others (e.g., 0.16 at Archer's Post, 0.18 at Nyahururu), high RMSE (e.g., 204.09 mm/year at Naro Moru), weak-to-moderate correlation (often CC<0.5), and mostly negative NSE, reflecting limited skill in capturing variability. TAMSAT consistently overestimates totals across most stations (positive RMB, e.g., 0.85 at Archer's Post, 0.79 at Munyaka, 1.53 at Wamba), resulting in large RMSE (e.g., 113.06 mm/year at Munyaka, 186.89 mm/year at Naro Moru) and very poor NSE (often strongly negative, down to -4.56 at Munyaka), indicating failure to reproduce observed dynamics despite occasional moderate correlation mainly in timing. PERSIANN-PDIR performs noticeably better than PERSIANN-CCS, with moderate correlations (CC 0.47–0.78), mixed bias (slight over- or underestimation depending on the site), near-zero or negative NSE, and lower RMSE than TAMSAT at several stations (e.g., 57.63 mm at Matanya), though it remains limited by frequent false alarms and moderate FAR; this relative advantage over PERSIANN-CCS is consistent with findings from Saemian et al. [51] in the Middle East. The evidence suggests that PERSIANN-PDIR is an intermediate performer, better than PERSIANN-CCS and TAMSAT in many contexts, but less consistent than CHIRPS and MSWEP.

The validation of the SPPs during the March–April–May (MAM) rainy season reveals substantial differences in their ability to reproduce observed rainfall across the eight stations (S1 Table). The products generally perform better at Archers Post and Naro Moru than at stations such as Munyaka, Nyahururu, and Wamba, where all products experience marked reductions in skill. CHIRPS shows good agreement with observations in terms of correlation, with high coefficients at Archers Post (CC=0.94), Naro Moru (CC=0.83), and Wamba (CC=0.75), indicating that it captures the temporal variability of MAM rainfall. Its best performance is at Archers Post (RMSE=97.11 mm; NSE=0.09). Naro Moru also has a positive NSE (0.20), although with a high RMSE (516.78 mm). In contrast, CHIRPS performs poorly at Munyaka (NSE=-14.89), Nyahururu (NSE=-5.41), and Wamba (NSE=-17.37), indicating that the magnitude of seasonal rainfall is

Table 5. Results of satellite precipitation validation showing median values for error metrics.

SPP	Performance Metric (Median)					
	RMSE (mm/year)	NSE	CC	RMB	POD	FAR
CHIRPS	57.6	0.08	0.60	0.14	0.24	0.66
MSWEP	59.07	0.01	0.60	0.27	0.70	0.70
PERSIANN-CCS	70.47	-0.29	0.46	-0.40	0.22	0.67
PERSIANN-PDIR	67.95	0.01	0.51	0.12	0.30	0.74
TAMSAT	80.48	-0.56	0.56	0.34	0.38	0.71

<https://doi.org/10.1371/journal.pwat.0000623.t005>

not well represented. MSWEP shows the lowest RMSE at Archers Post (50.17 mm), Kalalu (151.07 mm), and Munyaka (264.29 mm), and achieves the highest CC (0.96) and NSE (0.76) at Archers Post. At Naro Moru, MSWEP performs well, with NSE = 0.21 and CC = 0.84, but its performance deteriorates at Kangeta, Nyahururu, and Wamba, where NSE values become strongly negative and the correlation at Kangeta is even negative (CC = -0.61). Despite these localized deficiencies, MSWEP maintains relatively low rainfall biases (RMB 0.65–2.75). PERSIANN CCS shows mixed performance across the study area. It performs best at Naro Moru, with the highest CC (0.95), highest NSE (0.33), and lowest RMSE (470.14 mm), indicating excellent agreement with observed rainfall. Correlations remain strong at Archers Post (0.91) and Wamba (0.78). In contrast, performance deteriorates at Kalalu, Kangeta, Matanya, Munyaka, and Nyahururu, where NSE values are negative despite moderate to high correlations. Rainfall bias is moderate overall, with RMB values from 0.49 to 1.97, showing that while rainfall patterns are often captured, rainfall magnitude is frequently misestimated. PERSIANN PDIR shows consistently competitive performance and has the lowest RMSE at Kangeta (250.39 mm), Matanya (113.78 mm), Nyahururu (225.80 mm), and Wamba (171.27 mm), and the highest or least negative NSE at these sites. Positive NSE values at Archers Post (0.31), Kalalu (0.05), Kangeta (0.13), and Naro Moru (0.19) indicate good predictive skill across diverse climatic conditions. Correlation coefficients are high, up to 0.94 at Archers Post and 0.85 at Naro Moru, although it overestimates rainfall at all stations (RMB = 0.49–2.20). TAMSAT shows the weakest performance across the study area, with the highest RMSE at most stations: Archers Post (278.01 mm), Kalalu (226.47 mm), Matanya (238.40 mm), Munyaka (519.48 mm), Nyahururu (462.63 mm), and Wamba (430.71 mm). It also yields strongly negative NSE values, reaching -45.74 at Munyaka and -39.63 at Wamba. Although correlation coefficients are relatively high at Archers Post (0.93) and Naro Moru (0.91), large magnitude errors prevent accurate rainfall estimation. At the MAM rainfall season, MSWEP is the most reliable SPP, owing to its consistently low RMSE values, high correlation coefficients, and superior NSE performance across multiple stations.

The validation of the SPPs during the October–November–December (OND) minor rainy season indicates most products continue to exhibit predominantly negative Nash–Sutcliffe Efficiency (NSE) values ([S2 Table](#)). CHIRPS demonstrates moderate OND-season performance, capturing rainfall variability at several stations but showing inconsistent model efficiency. The product records best performance at Naro Moru, achieving a positive NSE of 0.42, a high correlation coefficient (CC = 0.89), and an RMSE of 173.82 mm, indicating agreement with observed rainfall. CHIRPS also produces low RMSE values at Kalalu (111.11 mm) and Matanya (105.57 mm) and moderate correlations at Archers Post (CC = 0.52), Matanya (CC = 0.56), and Wamba (CC = 0.71). Rainfall bias is consistently positive (RMB = 1.10–1.99), suggesting rainfall overestimation throughout the study area. MSWEP exhibits good performance in OND. It records the lowest RMSE at Archers Post (69.72 mm), Matanya (96.44 mm), Munyaka (129.13 mm), and Wamba (118.35 mm). Positive NSE values occur at Naro Moru (0.41) and Wamba (0.13), while the NSE at Archers Post (-0.28) is among the least negative across products. Correlations are moderate to high, reaching 0.74 at Naro Moru and 0.87 at Wamba, although they are negative at Kangeta (CC = -0.43). Rainfall bias remains moderate (RMB = 0.78–1.60), indicating limited systematic error. PERSIANN CCS records the largest RMSE at Kalalu (231.05 mm), Kangeta (519.15 mm), Matanya (223.24 mm), Munyaka (234.21 mm), and Naro Moru (318.83 mm), negative NSE across all stations, reaching -9.43 at Kalalu and -8.52 at Munyaka. Correlation coefficients are lowest among products, rarely exceeding 0.34 and becoming negative at Kangeta and Nyahururu. Although PERSIANN CCS exhibits the smallest rainfall bias (RMB = 0.15–0.83), indicating limited overestimation, this bias does not translate into predictive skill because the product fails to reproduce rainfall variability. PERSIANN CCS demonstrates the weakest performance during the OND season. TAMSAT demonstrates intermediate performance during the OND rainy season. It records the lowest RMSE at Kalalu (128.74 mm) and Naro Moru (114.18 mm) and the highest NSE (0.75) at Naro Moru, indicating excellent agreement with observations there. Correlation coefficients are also high at Naro Moru (CC = 0.87) and Wamba (CC = 0.79). However, TAMSAT performs poorly across the remaining stations, with negative NSE values from -4.25 at Munyaka to -0.08 at Wamba. Rainfall bias is generally higher than PERSIANN PDIR and MSWEP, with RMB values between 0.61 and 1.94. Its inconsistent

behaviour across the remaining stations thus limits overall applicability. PERSIANN PDIR provides the best performance among satellite products during OND rainy season. It records the lowest RMSE at Kangeta (409.59 mm), Naro Moru (144.70 mm), Nyahururu (149.67 mm), and Wamba (90.74 mm), and the highest CC (0.90) at Naro Moru. PERSIANN PDIR achieves positive NSE values at Naro Moru (0.60) and Wamba (0.49), representing the strongest efficiencies among products there. Although NSE values remain negative elsewhere, they are less negative than those of competing products. Rainfall bias is modest (RMB = 0.69–1.31), reflecting balanced estimation. PERSIANN PDIR emerges as the best-performing satellite rainfall product during the OND season.

The annual validation of the five satellite precipitation products (SPPs) reveals generally poorer performance than that observed during the MAM season (S3 Table). CHIRPS shows relatively stable performance across most stations, with moderate errors and strong correlations with observed rainfall. Its best performance is at Kalalu (NSE = 0.14), the only case where annual rainfall variability is satisfactorily reproduced. CHIRPS also attains a positive NSE at Naro Moru (0.27) and maintains strong correlations at Archers Post (CC = 0.74) and Naro Moru (CC = 0.88). Although RMSE remains high at wetter stations such as Naro Moru (715.60 mm) and Nyahururu (717.04 mm). Rainfall bias is moderate (RMB = 0.75–2.26), indicating overestimation across the study area, yet despite negative NSE values at several stations, CHIRPS still exhibits relatively consistent skill and compares favourably with the other satellite products. MSWEP shows mixed performance at the annual scale. It achieves the lowest RMSE at Archers Post (114.72 mm), indicating good agreement with observations at this station. It also attains a positive NSE at Naro Moru (0.24) and relatively high correlation coefficients at Kalalu (CC = 0.65) and Wamba (CC = 0.74). However, MSWEP performs poorly at the other stations, with negative NSE values everywhere else and negative correlations at Kangeta (-0.22), Munyaka (-0.21), and Nyahururu (-0.24). Although rainfall bias is relatively modest (RMB = 0.77–2.03), the large RMSE values at several stations, particularly Nyahururu (812.31 mm) and Naro Moru (730.05 mm), indicate reduced accuracy in estimating annual rainfall totals. PERSIANN CCS shows the weakest consistency among the PERSIANN products. While rainfall bias is small, with RMB values ranging from 0.31 to 1.23, the product records negative NSE values across all stations, indicating poor skill. RMSE values are among the highest products. Correlation coefficients remain moderate at stations including Naro Moru (CC = 0.71) and Kangeta (CC = 0.50), suggesting the product captures rainfall seasonality but fails to reproduce rainfall magnitude. PERSIANN PDIR records lowest RMSE at Kalalu (295.28 mm), Matanya (236.01 mm), Nyahururu (666.59 mm), and Wamba (219.55 mm), and highest correlation coefficient at Naro Moru (CC = 0.90). Importantly, PERSIANN PDIR is the only product besides CHIRPS to achieve a positive NSE at Kalalu (0.01), while maintaining less negative NSE values than competing products at several stations. Rainfall bias remains moderate (RMB = 0.48–1.56), reflecting limited systematic overestimation. Although the product still struggles at stations Munyaka and Archers Post. TAMSAT records the highest RMSE values at most stations, including Archers Post (365.28 mm), Matanya (274.90 mm), Munyaka (655.34 mm), Nyahururu (997.83 mm), and Wamba (542.23 mm), with negative NSE values at nearly all stations, reaching -14.56 at Archers Post and -13.57 at Munyaka. Although TAMSAT produces a positive NSE at Naro Moru (0.33) and high correlation coefficients at Naro Moru (CC = 0.85) and Wamba (CC = 0.50), these strengths are outweighed by large errors and rainfall overestimation, with RMB values from 0.69 to 2.41, making it the least reliable for among the other products. CHIRPS emerge as the most reliable product at the annual scale, followed by PERSIANN PDIR.

4.3. Performance of VegDischarge against In-situ data

The performance of VegDischarge data compared with observed discharge data is shown in Fig 6 and Table 6. Both VegDischarge and in-situ datasets follow a similar seasonal pattern, with peaks and troughs aligned to rainfall-driven flow events. This agreement is quantitatively supported by CC, which ranges from 0.46 to 0.71 across the stations, indicating a moderate linear relationship between the modeled and observed discharge. However, in-situ discharge values are consistently higher than VegDischarge during peak flow periods at stations such as A3 Naro Moru and A5 Naro Moru, indicating that the model tends to underestimate high flows at those stations. RMB scores range from -0.59 to 0.77, with A3 Naro

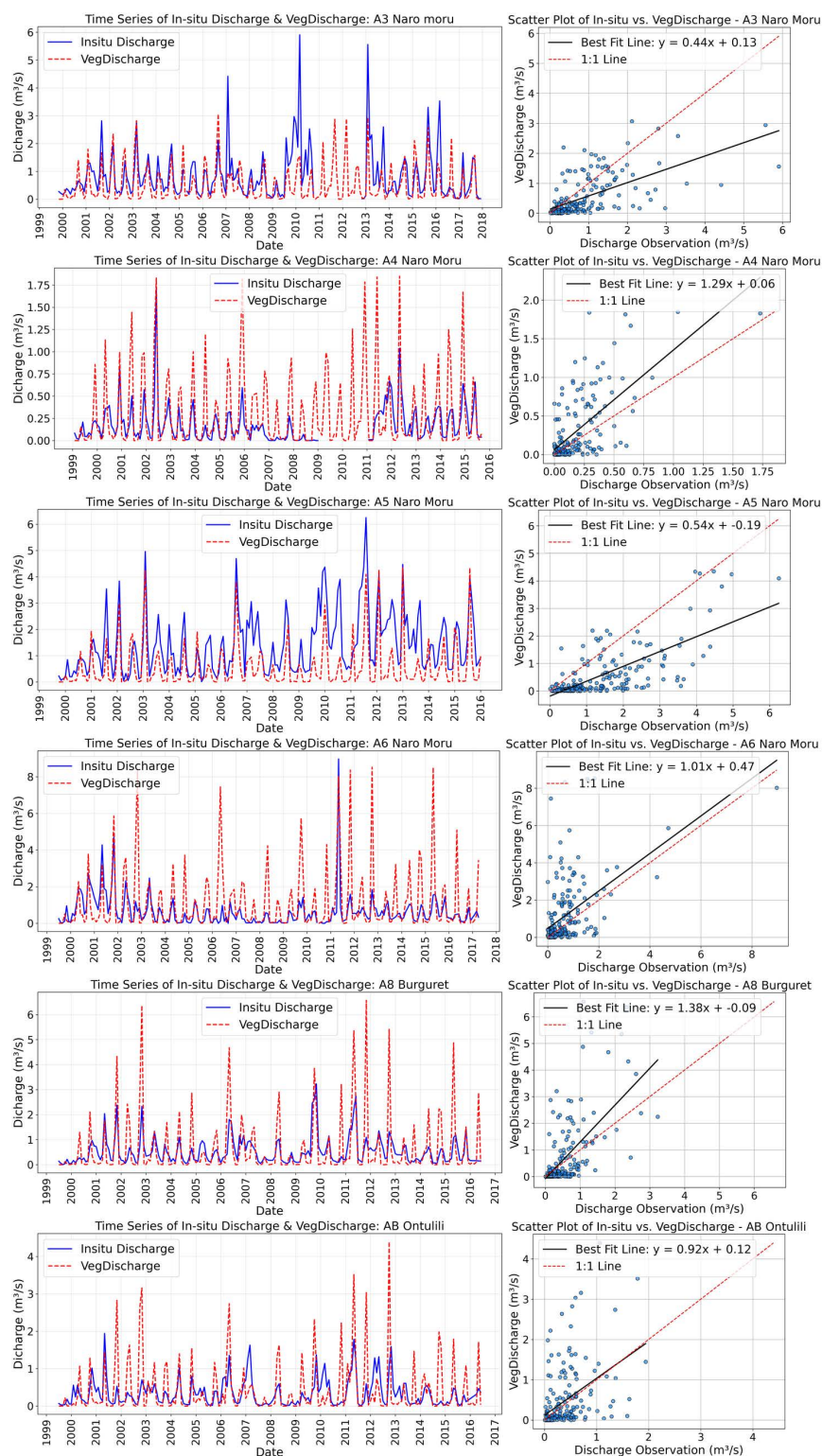


Fig 6. Time series and scatter plots comparing VegDischarge to observed discharge for all stations.

<https://doi.org/10.1371/journal.pwat.0000623.g006>

Table 6. Comparison of VegDischarge to observed discharge (m³/s).

Metric	A3 Naro Moru	A4 Naro Moru	A5 Naro Moru	A6 Naro Moru	A8 Burguret	AB Ontulili
RMB	-0.39	0.66	-0.59	0.77	0.22	0.28
RMSE	0.79	0.34	1.17	1.55	0.95	0.65
NSE	0.23	-1.54	-0.01	-2.10	-2.05	-2.28
CC	0.61	0.66	0.71	0.51	0.63	0.46

<https://doi.org/10.1371/journal.pwat.0000623.t006>

Moru and A5 Naro Moru having negative RMB scores, reinforcing the underestimation trend. For the remaining stations, VegDischarge consistently overestimates discharge magnitudes, particularly during high-flow events. A notable limitation is evident during low-flow conditions: VegDischarge systematically underestimates discharge magnitudes at most stations, often failing to reproduce observed low-flow peaks or sustained baseflow periods. This shortcoming stems primarily from uncertainties in baseflow estimation within the underlying VegET model framework, where simulated baseflow is frequently zero or near-zero [50]. These limitations highlight areas for potential model refinement to enhance reliability across the full hydrograph, particularly in water-scarce or drought-prone contexts.

4.4. Evaluation of precipitation bias correction (BC) methods

4.4.1. Daily precipitation BC assessments. Given its stronger performance, CHIRPS precipitation was selected for assessing bias correction methods. This section evaluates the daily bias correction of CHIRPS precipitation data against in-situ gauge records in the Central Highlands region from 2012 to 2021 (Fig 7). Across all eight stations, the uncorrected CHIRPS (cyan circle) generally shows a high standard deviation and a relatively low CC, placing it far from the observed reference point. The Taylor diagrams indicate that all bias correction methods significantly improved the performance of daily CHIRPS data, bringing the bias-corrected values much closer to the observed data in terms of correlation, standard deviation, and RMSD. Random Forest provided the best overall agreement with the observed rainfall across the evaluated stations, consistently exhibiting the highest correlation and lowest centered RMSD. However, Random Forest generally underestimated the observed standard deviation, indicating a tendency to dampen rainfall variability. Quantile Mapping more closely reproduced the observed standard deviation but exhibited weaker correlations and higher centered RMSD. Table 7 summarises the performance of the daily bias correction across the stations.

The azimuthal angle represents the correlation coefficient; radial distance is the standard deviation (mm/day), and the curved lines are arcs of constant RMSD (mm/day).

4.4.2. Monthly precipitation BC assessments. The evaluation of the four bias correction methods on monthly precipitation as shown in Fig 8 and Table 8, indicated that similar to the daily assessment, all methods successfully improved the raw CHIRPS precipitation estimates relative to the insitu data. However, a key finding was that the magnitude of this performance gain at the daily scale was insignificant compared with the substantial improvements observed during the monthly bias correction (judging by the performance metrics). The two simplest methods, LS and DM, demonstrated an improvement over the uncorrected data but were consistently the least effective of the applied correction methods. The RF method was the superior performer, achieving the lowest RMSE and highest NSE values across the eight locations. This demonstrates the strength of RF in capturing nonlinear relationships between satellite estimates and in situ observations.

Annual average precipitation of the Central Highlands region estimated before and after BC are presented in Fig 9. The uncorrected CHIRPS data (left panel) show a relatively smooth distribution of precipitation, with higher amounts concentrated in the central and southern parts of the basin. At the same time, the northern and eastern zones appear considerably drier, with the overall precipitation range being narrow, with maximum values not exceeding 220 mm. Following bias correction (right panel), the corrected data capture a wider range of values, extending from less than 16 mm to over 285 mm, with enhanced gradients between wetter and drier zones.

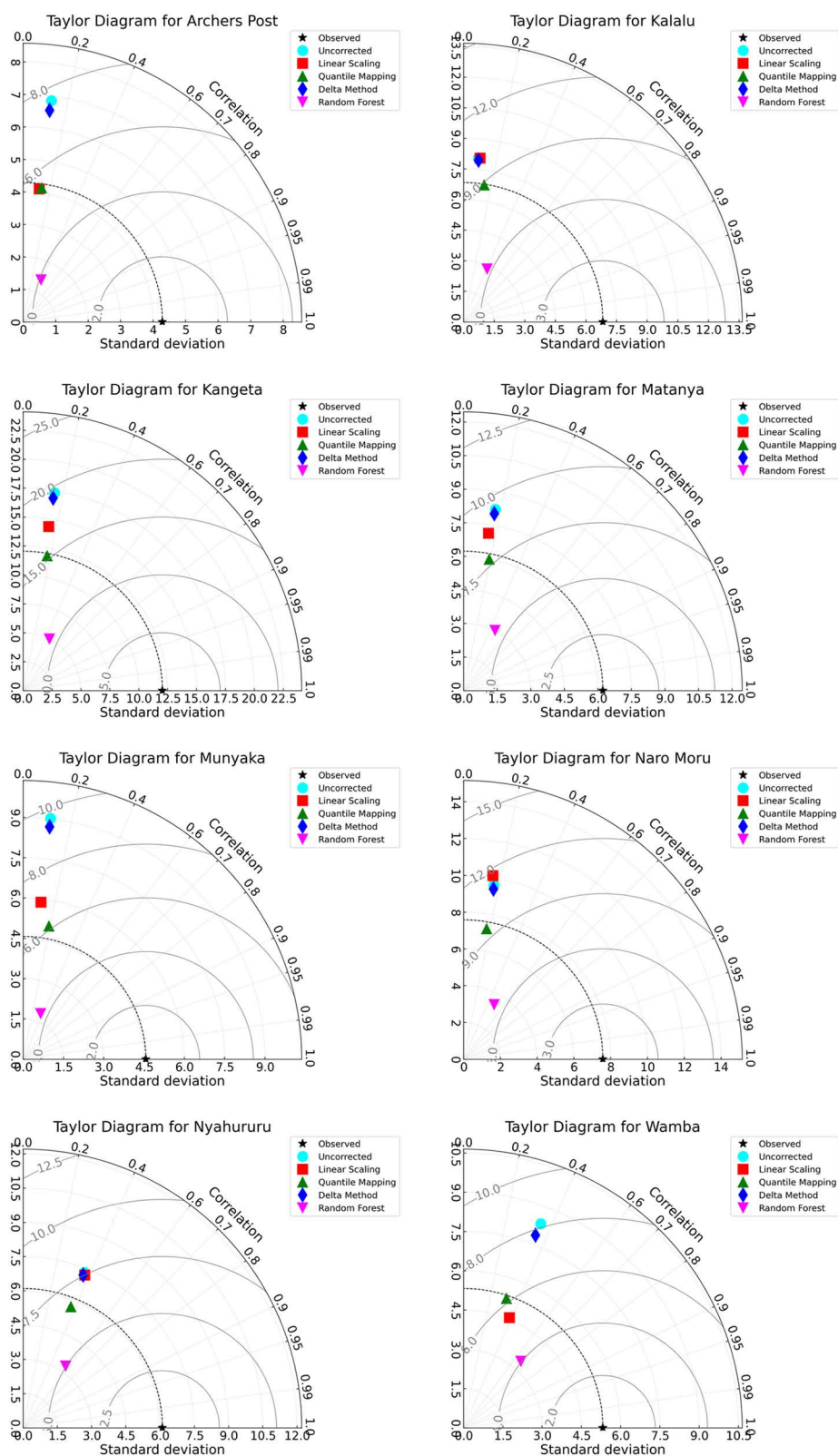


Fig 7. Taylor diagrams of uncorrected daily CHIRPS (cyan circle) and bias-corrected estimates with corresponding rain gauge data (black star) for all stations.

<https://doi.org/10.1371/journal.pwat.0000623.g007>

Table 7. Performance of Bias Correction Methods Across Stations for Daily Precipitation.

Station	Metric	Uncorrected CHIRPS	Linear Scaling	Quantile Mapping	Delta Method	Random Forest
Archer's Post	RMB	0.56	-0.53	2.60	0.64	0.01
	CC	0.13	0.44	0.26	0.26	0.58
	RMSE	7.63	3.99	5.5	5.09	3.50
	NSE	-2.17	0.13	-0.65	-0.41	0.33
Kalalu	RMB	0.05	-2.13	2.07	-0.17	0.01
	CC	0.09	0.15	0.21	0.21	0.63
	RMSE	10.07	7.11	7.59	7.47	5.34
	NSE	-1.18	-0.08	-0.23	-0.19	0.39
Kangeta	RMB	0.08	-3.36	1.49	-1.21	0.01
	CC	0.16	0.29	0.16	0.16	0.68
	RMSE	19.47	12.03	12.61	12.72	8.87
	NSE	-1.61	0	-0.1	-0.12	0.46
Matanya	RMB	0.20	-1.92	2.32	0.07	0.01
	CC	0.17	0.03	0.15	0.15	0.63
	RMSE	9.42	6.63	7.38	7.19	4.90
	NSE	-1.29	-0.14	-0.4	-0.33	0.38
Munyaka	RMB	0.46	-1.55	2.67	0.38	0.01
	CC	0.10	0.05	0.13	0.13	0.56
	RMSE	9.80	5.28	6.66	6.35	4.11
	NSE	-2.94	-0.15	-0.82	-0.65	0.31
Naro Moru	RMB	-0.12	-2.64	1.62	-0.61	0.00
	CC	0.17	0.12	0.16	0.17	0.63
	RMSE	11.18	8.01	8.33	8.29	5.91
	NSE	-1.18	-0.12	-0.21	-0.20	0.39
Nyahururu	RMB	0.08	-2.14	1.69	-0.58	0.01
	CC	0.37	0.11	0.26	0.27	0.75
	RMSE	7.62	6.73	6.48	6.38	4.10
	NSE	-0.56	-0.22	-0.13	-0.09	0.55
Wamba	RMB	1.03	-0.69	1.72	0.69	0.03
	CC	0.35	0.43	0.34	0.38	0.83
	RMSE	8.21	4.87	5.85	5.54	3.05
	NSE	-1.38	0.16	-0.21	-0.08	0.67

<https://doi.org/10.1371/journal.pwat.0000623.t007>

4.5. Assessment of discharge data bias correction

The Taylor diagrams (Fig 10), and statistical metrics (Table 9), illustrate the differences in performance among the bias correction techniques. Among the approaches, RF consistently lies closest to the observational reference across stations. Its high CCs (0.92–0.95) at 5 of the 6 stations and standard deviations comparable to observations are corroborated by very low RMSE and strongly positive NSE values. This indicates that RF not only reproduces temporal dynamics effectively but also preserves the variance structure of streamflow. An exception is A5 Naro Moru, where RF yielded negligible improvement over the raw VegDischarge (NSE = −0.01; RMB = −0.55; CC = 0.67) and Linear Scaling instead performed best (NSE = 0.45; RMB = 0.08). This station exhibits the highest discharge magnitudes and the most frequent extreme

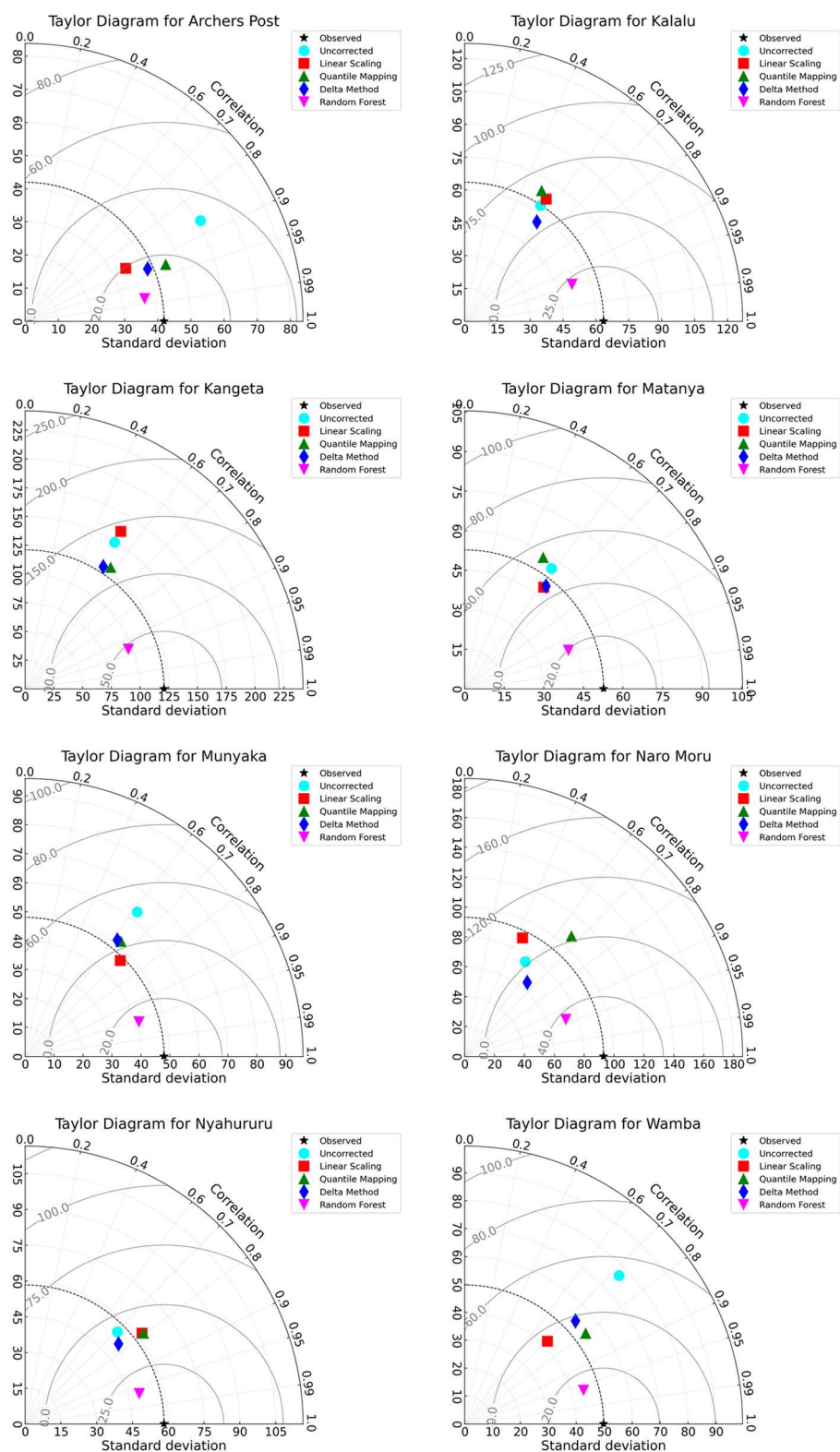


Fig 8. Taylor diagrams of uncorrected CHIRPS (cyan circle) and bias-corrected estimates on a monthly timescale, with corresponding rain gauge data (black star) for all stations. The azimuthal angle represents the correlation coefficient; radial distance is the standard deviation (mm/month), and the curved lines are arcs of constant RMSE (mm/day).

<https://doi.org/10.1371/journal.pwat.0000623.g008>

Table 8. Performance of Bias Correction Methods Across Stations for Monthly Precipitation.

Station	Metric	Uncorrected CHIRPS	Linear Scaling	Quantile Mapping	Delta Method	Random Forest
Archer's Post	RMB	0.47	-0.14	0.11	0.04	-0.05
	CC	0.87	0.89	0.93	0.92	0.98
	RMSE	34.85	19.97	17.28	16.46	9.05
	NSE	0.3	0.77	0.83	0.84	0.95
Kalalu	RMB	0.05	0.07	0.04	0	-0.01
	CC	0.55	0.56	0.51	0.59	0.95
	RMSE	60.04	61.59	65.79	54.46	22.14
	NSE	0.1	0.05	-0.08	0.26	0.88
Kangeta	RMB	0.09	0.17	0.03	0.02	-0.02
	CC	0.52	0.52	0.58	0.54	0.93
	RMSE	134.26	142.46	114.88	118.17	46.23
	NSE	-0.25	-0.4	0.09	0.04	0.85
Matanya	RMB	0.17	-0.03	0.05	0	-0.01
	CC	0.59	0.61	0.51	0.62	0.94
	RMSE	50.68	44.62	54.63	44.46	19.72
	NSE	0.07	0.28	-0.08	0.28	0.86
Munyaka	RMB	0.43	-0.05	0.05	0.01	0
	CC	0.61	0.7	0.64	0.62	0.96
	RMSE	55.27	36.32	42.29	43.11	14.71
	NSE	-0.34	0.42	0.22	0.19	0.91
Nyahururu	RMB	0.1	0.1	0.02	0	-0.03
	CC	0.71	0.79	0.79	0.76	0.97
	RMSE	43.8	39.65	38.87	38.42	16.52
	NSE	0.43	0.53	0.55	0.56	0.92
Naro Moru	RMB	-0.21	0.05	0.06	0	0.01
	CC	0.54	0.44	0.66	0.65	0.94
	RMSE	84.4	95.7	83.07	70.66	35.19
	NSE	0.17	-0.07	0.19	0.42	0.86
Wamba	RMB	1.12	-0.04	0.1	0.11	-0.03
	CC	0.72	0.71	0.8	0.73	0.96
	RMSE	62.33	35.64	33.04	38.16	14.01
	NSE	-0.58	0.48	0.55	0.41	0.92

<https://doi.org/10.1371/journal.pwat.0000623.t008>

outliers in the network (Section 4.1), suggesting that flow magnitudes beyond the range represented in the training data constrained the RF correction. Overfitting was mitigated through seasonal stratification, data filtering, and ensemble averaging inherent to Random Forests, which employs bootstrapping to train multiple decision trees on different resampled subsets of the training data. By drawing samples with replacement, individual trees are exposed to varying data realisations, reducing sensitivity to noise and preventing any single subset from dominating model behaviour. The aggregation of predictions across independently trained trees through averaging enhances model stability and robustness, thereby limiting overfitting and improving the reliability of bias correction, and performance metrics therefore reflect correction skill rather than out-of-sample generalization. The traditional statistical techniques exhibit moderate performance. LS and QM cluster at intermediate distances from the observation, generally maintaining reasonable CCs (0.60–0.73) but showing some deviations in standard deviation. Their RMSE values remain higher than RF, and NSE values are often marginal or even negative at some stations. DM tends to exaggerate variability, appearing displaced away from the reference along

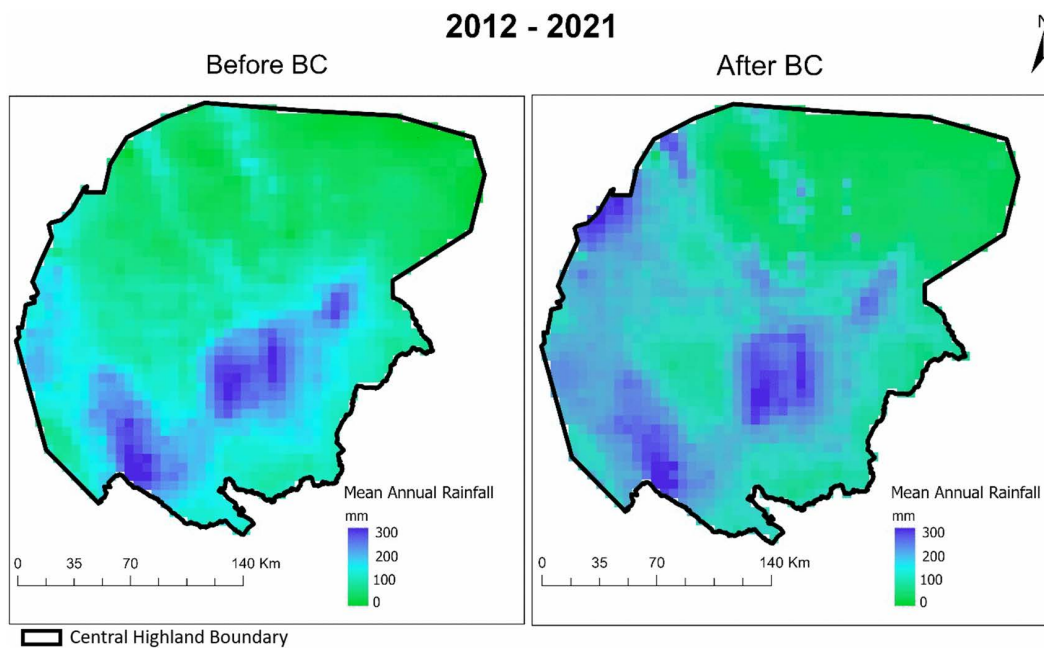


Fig 9. Annual average precipitation of CHIRPS before and after bias-correction (2012-2021). Maps were generated using ArcGIS Pro. The study area boundary is identical to that shown in Fig 1, with base county shapefiles sourced from openAFRICA (<https://open.africa/dataset/kenya-counties-shapefile>, CC BY 4.0).

<https://doi.org/10.1371/journal.pwat.0000623.g009>

the standard deviation axis. Although it achieves CC similar to LS, its higher RMSE and low NSE reflect a weaker correction of errors.

Fig 11 compares mean monthly discharge time series from in-situ observations, VegDischarge, and the four bias-correction methods (LS, QM, DM, and RF) across all stations. LS and the DM produce conservative corrections characterised by smooth hydrographs and systematically dampened peak flows, often closely overlapping, indicating a primary adjustment of mean discharge rather than flow variability. QM introduces greater variability and sharper peaks, particularly during high-flow periods, enhancing responsiveness to extreme events but occasionally amplifying peak magnitudes. In contrast, the RF approach demonstrates a more balanced behaviour, preserving temporal variability and peak-flow dynamics without the pronounced smoothing observed in the statistical methods or the occasional peak exaggeration associated with QM.

5. Discussion

Reliable estimation of precipitation and river discharge is essential for understanding hydrological processes and managing water resources effectively, particularly in data-scarce regions where in-situ gauge networks are limited and unevenly distributed. We conducted a comparative study of the precipitation time series at eight stations from 2012 to 2021 to evaluate the performance of five SPPs: CHIRPS, PERSIANN-CCS, PERSIANN-PDIR, TAMSAT, and MSWEP. Four pairwise comparison statistics (RMSE, NSE, RMB and CC) and two categorical validation statistics (POD and FAR) were used for assessment. Our findings show that all the SPPs significantly overestimate rainfall at most stations except for PERSIANN-CCS, a pattern documented in complex mountainous terrain where orographic processes contribute to inaccuracies in satellite-based precipitation estimates [37]. Even so, CHIRPS and MSWEP performed most reliably, with each excelling in different aspects of rainfall representation. This duality mirrors findings from several other regional studies:

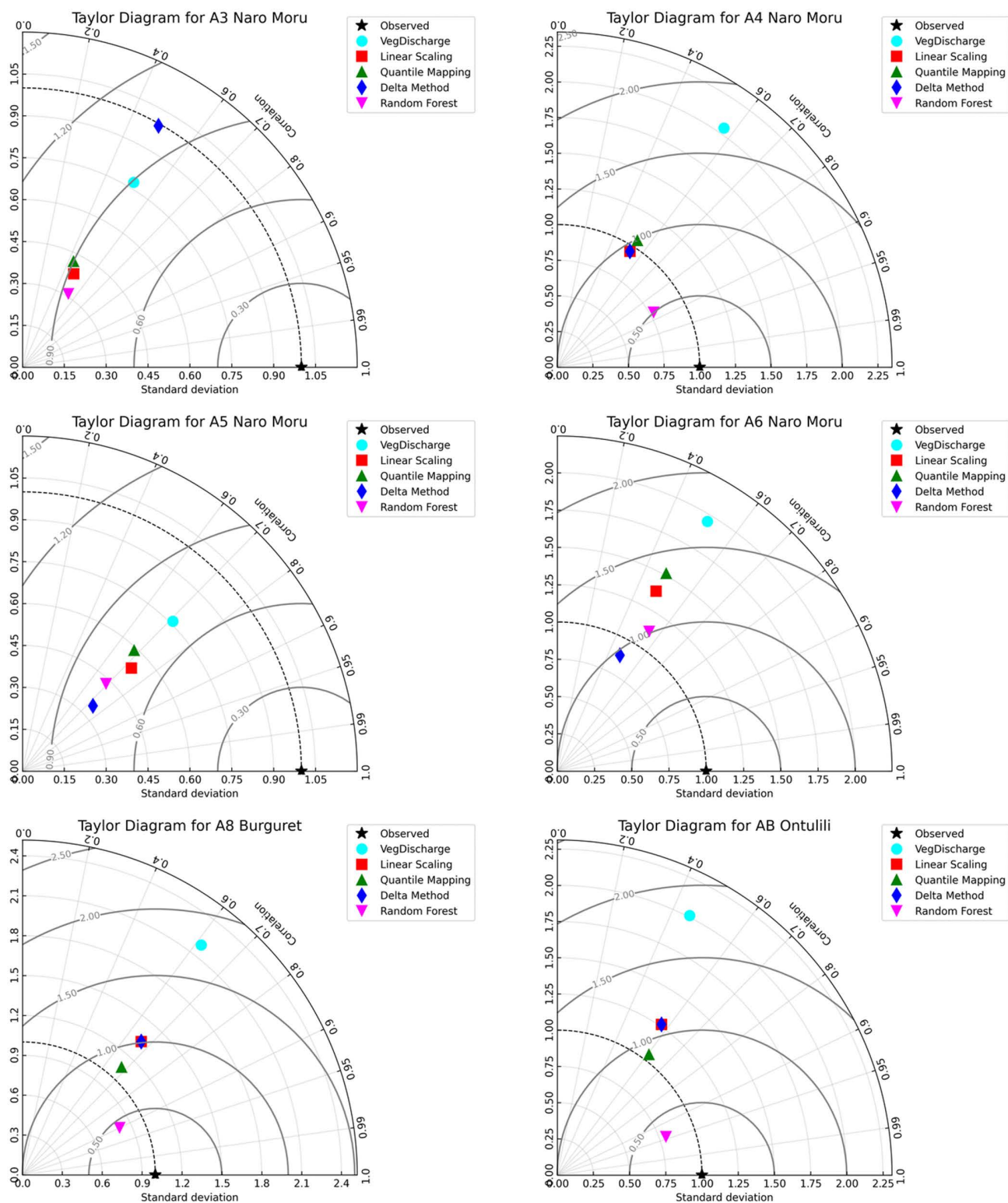


Fig 10. Taylor diagrams of uncorrected vegDischarge (cyan circle) and bias-corrected estimates, with corresponding observed data (black star) for all stations. The azimuthal angle represents the correlation coefficient; radial distance is the standard deviation (mm/month), and the curved lines are arcs of constant RMSE (m^3/s).

<https://doi.org/10.1371/journal.pwat.0000623.g010>

Table 9. Performance of Bias Correction Methods Across Discharge Stations.

Station	Metric	VegDischarge	Linear Scaling	Quantile Mapping	Delta Method	Random Forest
A3 Naro Moru	RMB	-0.39	0.00	-0.02	-0.01	-0.02
	RMSE	0.79	0.66	0.76	0.79	0.32
	NSE	0.23	0.46	0.28	0.24	0.88
	CC	0.61	0.68	0.64	0.59	0.95
A4 Naro Moru	RMB	0.66	0.13	-0.08	-0.07	0.05
	RMSE	0.34	0.15	0.19	0.18	0.09
	NSE	-1.54	0.48	0.18	0.32	0.83
	CC	0.66	0.70	0.60	0.63	0.93
A5 Naro Moru	RMB	-0.59	0.08	-0.56	-0.55	-0.55
	RMSE	1.17	0.27	1.20	1.11	1.17
	NSE	-0.01	0.45	-0.05	0.09	-0.01
	CC	0.71	0.68	0.64	0.73	0.67
A6 Naro Moru	RMB	0.77	0.12	-0.01	-0.01	-0.01
	RMSE	1.55	0.74	0.92	0.85	0.40
	NSE	-2.1	0.30	-0.10	0.06	0.80
	CC	0.51	0.55	0.46	0.47	0.92
A8 Burguret	RMB	0.22	-0.02	-0.12	-0.11	-0.02
	RMSE	0.95	0.35	0.42	0.51	0.21
	NSE	-2.05	0.57	0.40	0.13	0.86
	CC	0.63	0.76	0.73	0.71	0.93
AB Ontulili	RMB	0.28	0.08	-0.08	-0.07	0.01
	RMSE	0.65	0.27	0.33	0.39	0.13
	NSE	-2.28	0.45	0.17	-0.16	0.87
	CC	0.46	0.68	0.61	0.57	0.94

<https://doi.org/10.1371/journal.pwat.0000623.t009>

CHIRPS consistently shows a stronger ability to capture temporal variability across East Africa [25,26,69], while MSWEP is frequently noted for its accuracy in reproducing precipitation magnitudes across diverse climates [70,71]. The good performance of CHIRPS and MSWEP in our study, as compared to the other SPPs, aligns with regional patterns documented in the literature, suggesting that both products represent robust options for hydrological applications in equatorial East Africa [71–74].

In contrast, TAMSAT, PERSIANN-PDIR, and PERSIANN-CCS performed relatively poorly, characterised by large RMSEs, strongly negative NSEs, and weak CC, suggesting limited ability to replicate observed precipitation patterns. The poorer performance in TAMSAT and the PERSIANN variants aligns with earlier work suggesting that finer spatial resolution does not necessarily translate to better accuracy in complex terrain [37,74].

MSWEP showed clear advantages in terms of RMSE and NSE at Archers Post, Kangeta, Wamba, and Munyaka, indicating lower magnitude error at those stations, while CHIRPS tended to achieve stronger CC at most stations (Archers Post, Matanya, Naro Moru, Nyahururu, Kangeta), suggesting a better ability to capture temporal variability. RMB was mixed and site-dependent, with neither CHIRPS nor MSWEP being uniformly less biased; each was closer to unity at four stations. These findings suggest that while MSWEP may reduce magnitude errors at several stations, CHIRPS provides stronger temporal agreement and enhanced performance in mountainous terrain, making it particularly valuable for applications in heterogeneous landscapes [26,70]. The findings in the performance of SPPs in the Central highlands are in line with Kimani et al. [75], who also found that CHIRPS, as well as CMORPH and TRMM, consistently outperformed

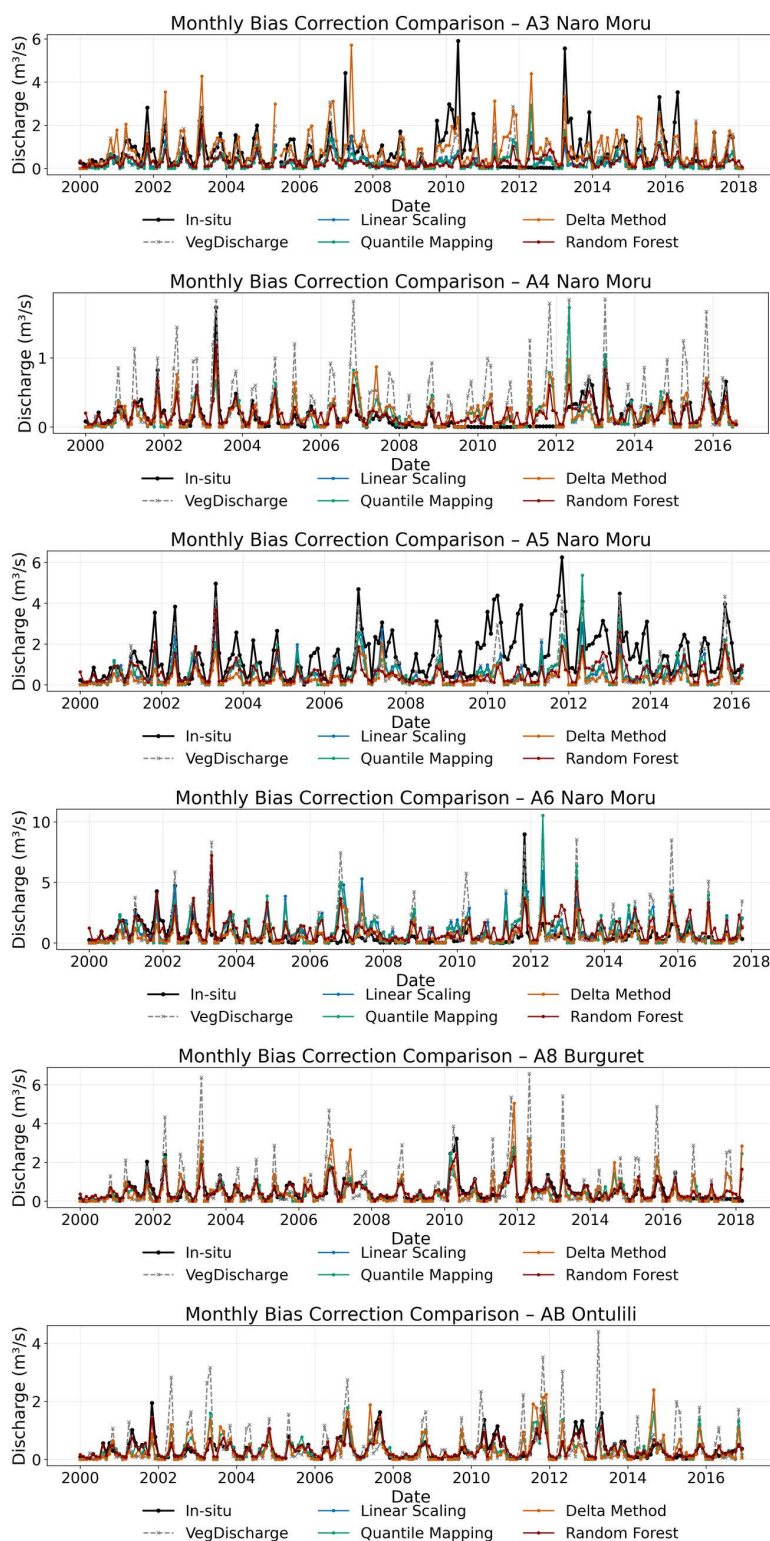


Fig 11. Monthly discharge time series comparing in-situ observations and bias-corrected VegDischarge estimates for all stations.

<https://doi.org/10.1371/journal.pwat.0000623.g011>

other products such as TAMSAT, TARGAT, PERSIANN-CDR, CMAP, and GPCP when they evaluated their performance against high-resolution gridded gauge data across East Africa (1998–2012). Similarly, in another study by Ageet et al. [72], MSWEP and CHIRPS performed best at the monthly and annual time steps, respectively, compared to IMERG and TMPA over equatorial East Africa for the period 2001–2018. Nkiaka and Tume [74] also found CHIRPS and MSWEPv2 as the best performing SPPs compared to ARCV2, GPM, RFEv2, TAMSAT, ERA5, AgERA5 and TerraClimate (TC) in the Mount Fanko region in Central-West Africa using gauge data from six locations.

POD and FAR were calculated using daily precipitation data to assess rain detection skill, and the results revealed that most SPPs exhibited relatively high false alarm ratios ($FAR > 0.6$) and low to moderate rainfall detection skill ($POD \leq 0.7$). Among the products, CHIRPS and MSWEP performed better than the others, though their skill remained limited.

The median performance metrics across stations, show that CHIRPS performs better overall, with lower mean RMSE and RMB, and stronger correlation, while MSWEP performs better in NSE. In terms of daily rainfall detection skill, CHIRPS performs better in FAR, while MSWEP performs better in POD. This complementary behaviour aligns with findings from Omondi et al. [9], who similarly observed that MSWEP demonstrated good rainfall detection skill, with the highest median POD across crop stages, while CHIRPS exhibited the best results in terms of falsely detected rainfall, with the lowest FAR values in Kenya. Furthermore, CHIRPS has been reported to perform better at coarser temporal scales [70,72,75,76]. Beyond the monthly scale, the seasonal and annual validations revealed scale-dependent product rankings. MSWEP was the most reliable product during the long rains (MAM; S1 Table), PERSIANN-PDIR emerged as the strongest performer during the short rains (OND; S2 Table), while CHIRPS proved the most consistent at the annual scale (S3 Table). The predominantly negative NSE values across all products at seasonal and annual scales particularly at the wetter, high-elevation stations indicate that no single product reliably reproduces rainfall magnitudes across all timescales, reinforcing that product selection should be matched to the temporal scale, season, and location of the intended application. During the course of this research, CHIRPS version 3 was released, incorporating methodological and input-data improvements relative to earlier versions. It is therefore recommended that future studies evaluate the performance of CHIRPS v3 in similar data-scarce environments and assess its response to bias correction approaches, particularly in comparison with CHIRPS v2.

This study further applied four bias correction techniques: LS, QM, DM, and RF to bias-correct CHIRPS data at both daily and monthly scales, to enhance its accuracy and reduce systematic biases. At the monthly scale, RF achieved the most substantial improvement in CHIRPS precipitation estimates, yielding the lowest RMSE values (ranging from 9.05 mm at Archer's Post to 46.23 mm at Kangeta), the highest correlation coefficients ($CC = 0.93$ – 0.98), and NSE values between 0.85 and 0.95 across all stations. This strong performance is consistent with similar findings in the literature; for example, Iqbal et al. [36] evaluated RF and K-Nearest Neighbors (KNN) algorithms for the bias correction of multiple Satellite Precipitation Products (SPPs) over Peninsular Malaysia. They found that RF consistently outperformed KNN and conventional statistical methods, achieving higher classification skill scores and lower RMSE than with uncorrected satellite data. In a related effort, Tanimu et al. [77] assessed thirteen bias correction methods (both traditional and machine learning-based) for correcting and downscaling four CMIP6 GCMs over Nigeria. They concluded that RF better corrects the bias of all three climate variables: rainfall, maximum temperature, and minimum temperature. This demonstrates the capability of machine learning approaches to capture complex, nonlinear relationships between satellite estimates and ground observations that traditional statistical methods may overlook.

The DM bias correction method ranked second, producing moderate improvements in the metric scores, suggesting its effectiveness in correcting mean biases, though less adept at capturing nonlinear variability. QM showed mixed performance, with improved distributional matching at some locations but occasional additional bias under highly variable conditions. Similar behaviour was reported by Beyer et al. [78] who found that DM performs slightly better, albeit not always to a statistically significant degree, at minimising the median absolute bias between empirical data and debiased simulations for both temperature and precipitation than Generalised additive models (GAMs) and QM. They also reported

considerable spatial and temporal variability in method performance, reinforcing the pattern observed in this study: DM provides consistent but limited improvements, while QM can enhance distributional structure but remains sensitive to spatial and climatological variability. LS also improved the CHIRPS data, though it was the least effective. This can be attributed to its uniform correction approach, which limits its responsiveness to spatial and temporal precipitation variations. LS was also found to perform the least compared to other BC methods, such as power transformation of precipitation (PTR), empirical quantile mapping (EQM), gamma quantile mapping (GQM), and gamma-pareto quantile mapping (GPQM), and DM over Costa Rica for monthly precipitation outputs from five General Circulation Models (GCMs) [79]. Similarly, Bhatti et al. [24] observed that simple scaling approaches, such as LS, underperform in capturing high temporal variability and fail to accurately simulate heavy rainfall intensities. Collectively, these studies highlight that while LS can correct the mean, its inability to adjust the frequency and distribution of rainfall makes it the least robust method for capturing the complex precipitation patterns required for hydrological applications. The markedly larger correction gains at the monthly scale relative to the daily scale reflect the nature of the underlying errors: daily satellite–gauge comparisons are dominated by stochastic errors and event-timing mismatches that no deterministic mapping can fully remove, whereas temporal aggregation cancels much of this random component, leaving a predominantly systematic bias that is far more amenable to correction. The contrasting performance of methods at different temporal scales underscores the importance of context-specific correction strategies.

The study further examined the performance of the VegDischarge dataset relative to observed streamflow data. VegDischarge was selected because VegET generates runoff at approximately 1 km resolution, which is subsequently routed through mizuRoute, producing mean monthly discharge. This provides finer spatial representation of runoff generation than PCR-GLOBWB (5 arc-minutes; ~10 km), WaterGAP (0.5°; ~55 km), and the standard global CWatM configuration (0.5°; ~55 km) [43,44,46]. The VegDischarge dataset successfully captured seasonal flow dynamics, with peaks and troughs broadly aligning with rainfall-driven events. However, quantitative validation revealed station-dependent biases: peak flows were underestimated at the higher-yielding Naro Moru stations (A3 and A5), discharge was overestimated at the remaining stations, and low flows were systematically underestimated across nearly all sites, a shortcoming linked to the near-zero baseflow frequently simulated by the underlying VegET model, and to its documented limitations in partitioning surface runoff and deep drainage [50]. Consequently, initial model outputs showed poor statistical agreement, often resulting in negative NSE values at most stations, and scatterplots clearly showed the majority of simulated points falling below the 1:1 line (Fig 6), underscoring the necessity of bias correction before operational use. A similar challenge was reported by Siderius et al. [80] when evaluating the LPJmL (Lund-Potsdam-Jena managed Land) global hydrological model in the East African Rufiji Basin, they found that the model struggled to reproduce both the magnitude and timing of flows and emphasized that for basin-scale water-management or scenario-analysis applications, additional calibration and validation against observed discharge are required. This parallel reinforces the observation that even well-established global hydrological models often require local adjustment when applied in local basins.

Following bias correction, RF again delivered the most significant improvements, achieving correlations above 0.9 and NSE values exceeding 0.8 at multiple stations. It also preserved variance structures more effectively than the statistical methods, suggesting that RF not only corrects mean bias but also accounts for variability in discharge magnitudes. A similar result was found in a previous study, Shen et al. [81], which utilised RF to bias-correct PCR-GLOBWB streamflow predictions, employing both meteorological and hydrological variables as predictors, and found that the RF-based error-correction model significantly improved the PCR-GLOBWB predictions. The findings demonstrated the RF method's ability to estimate prediction errors. Further, they indicated that considerable improvement in predictive accuracy could be obtained by using the bias correction method, without the cumbersome work of calibrating the PCR-GLOBWB model. Among the statistical bias correction methods, LS and QM provided moderate improvements but were less effective in addressing extreme flows. Nevertheless, the weak RF performance at A5 Naro Moru illustrates a structural limitation of tree-based ensemble methods, because Random Forest predictions are bounded by the range of the training data, the

method cannot extrapolate to flow magnitudes unrepresented during training, making it least reliable precisely at the extreme high flows that matter most for flood-related applications. This limitation strengthens the case for the physics-informed and hybrid correction approaches identified below as future research needs. This limitation was also noted by Hagemann et al. [82], who showed that conventional single-factor correction approaches struggle with biases that differ across low, medium, and high-flow regimes, prompting the development of a three-quantile method that more effectively corrected discharge extremes, improving KGE from as low as -0.79 to as high as 0.90 after adjustment. The DM tended to exaggerate variability, resulting in poorer performance.

Comparable efforts to improve discharge realism at larger spatial scales include Liu et al. [83] who evaluated a global discharge reanalysis (the GloFAS / ERA5-based global river discharge product and related reanalysis outputs) against observed streamflow records at many gauged catchments, and reported that the uncorrected global reanalysis exhibited large, location-dependent biases and variable skill, with some humid/seasonal catchments showing reasonable skill, while many arid or complex basins exhibit poor performance. To improve the raw reanalysis, the authors employed a Piecewise Random Forest (P-RF) bias-correction system (a machine-learning post-processor designed to treat different flow regimes separately) and documented substantial improvements across performance measures. Similarly, Sanchez-Lozano et al. [84] demonstrated that locally informed, seasonally adjusted bias correction significantly improves the performance of the GEOGloWS ECMWF Streamflow Services (GESS) across diverse hydroclimatic regions (Australia, Brazil, Colombia, Dominican Republic, Peru). Initial simulations showed a poor fit, with KGE values below 0.4 in 2012 across Latin American basins, indicating systematic errors in flow magnitude and variability. Applying the seasonal bias correction substantially enhanced model reliability, particularly in Brazil, Colombia, and Peru, where post-correction KGE values often exceeded 0.6 – 0.7 , and mean errors (ME) approached zero. Although their method differs conceptually from the machine-learning framework applied in the present study, both approaches highlight the importance of flexible correction strategies capable of capturing non-uniform bias behaviour.

Despite these advancements, several limitations must be considered. Firstly, the spatial distribution and density of observational networks remain a constraint. While the rainfall stations cover a range of physiographic settings, the network is relatively sparse given the complex terrain of the Central Highlands. With only eight rainfall stations available for analysis, some northeastern and peripheral areas are underrepresented, limiting the observations' ability to fully characterize fine-scale spatial variations in precipitation. The streamflow gauging network offers even less comprehensive spatial coverage, with stations primarily concentrated in the central headwater region. Consequently, hydrological responses in lower-elevation catchments, semi-arid environments, and downstream river reaches cannot be characterized with the same level of confidence due to the limited availability of direct streamflow observations. This uneven observational coverage may also constrain the spatial generalizability of the bias-correction models, particularly in locations poorly represented in the training data. Future studies should incorporate additional observations and independent datasets from underrepresented areas to enhance spatial validation and improve confidence in the transferability of the corrected products. Secondly, although the RF approach demonstrated strong overall performance, its effectiveness varied across stations, emphasizing the importance of balancing model complexity with spatial generalizability. The transferability of the bias-correction models across different regions and time periods remains uncertain. A model that performs well under the hydroclimatic conditions represented in the available training and validation data may not necessarily maintain the same performance under substantially different conditions. Future research should evaluate the transferability of the correction models across independent regions, longer time periods, and contrasting hydroclimatic regimes, including conditions not represented in the current training data. The annual water-balance check provides partial reassurance on this point: the positive storage changes of 2.8 – $8.1 \text{ km}^3 \text{ yr}^{-1}$ (2018–2021) are physically plausible for a recharge-dominated water-tower system. However, because part of this residual may absorb remaining biases in the corrected fluxes, closure at the annual scale should be interpreted as a necessary but not sufficient indicator of physical consistency and does not preclude imbalances at monthly or event scales. Thirdly, the present study primarily evaluates model performance using deterministic statistical metrics and does not

formally quantify the uncertainties associated with the input satellite precipitation products, embedded atmospheric corrections, or the structural uncertainty of the VegDischarge model. Consequently, uncertainty associated with the corrected datasets remains insufficiently characterized. Future research should incorporate formal uncertainty quantification through approaches such as ensemble-based product comparison, probabilistic bias-correction frameworks, and structural uncertainty analysis to provide uncertainty bounds alongside the corrected estimates.

6. Conclusion

Accurate and reliable precipitation estimates are essential for effective water resource management, particularly in data-scarce regions such as the Central Highlands of Kenya. This study evaluated the performance of five SPPs, CHIRPS, PERSIANN-CCS, PERSIANN-PDIR, TAMSAT, and MSWEP, against in-situ rainfall observations. Although all products showed some degree of correlation with gauge data, their performance varied considerably across stations and metrics. CHIRPS and MSWEP consistently emerged as the most reliable products, each excelling in complementary aspects of performance. Station-level analyses showed that both products performed equally well in terms of RMSE and NSE, with each outperforming the other at four stations, while bias performance was evenly distributed between them. Notably, CHIRPS demonstrated superior temporal correlation at five stations, indicating a stronger ability to capture rainfall timing and variability. These results underscore the value of satellite-based precipitation products in providing spatially and temporally continuous rainfall information in regions where gauge networks are sparse or unevenly distributed.

The application of bias correction techniques to the CHIRPS product led to substantial improvements in precipitation estimation accuracy. Among the four methods evaluated, LS, QM, DM, and RF, the RF approach delivered the most pronounced improvements, particularly at the monthly scale, achieving high correlation coefficients (0.93–0.98) and NSE values ranging from 0.85 to 0.95. While the statistical bias correction methods reduced mean bias to varying degrees, they were generally less effective in correcting variability at stations characterized by highly dynamic rainfall regimes. In contrast, the Random Forest approach was better able to capture nonlinear error structures, although its performance varied across stations, highlighting the need to balance model complexity with spatial generalizability.

Evaluation of the model-derived VegDischarge dataset against observed streamflow revealed that while the dataset successfully captured seasonal flow dynamics, it consistently underestimated discharge magnitudes, particularly during peak-flow and low-flow periods. This systematic underestimation resulted in weak statistical agreement, with negative NSE values at most stations. Despite these limitations, the results emphasize the role of model-derived discharge datasets in providing spatially and temporally continuous streamflow information in regions where monitoring networks are sparse or incomplete, such datasets may represent the only available source of hydrological information. Applying the same bias correction techniques to the VegDischarge dataset resulted in marked performance improvements. As observed for precipitation, the Random Forest approach consistently outperformed the statistical methods, achieving correlation coefficients above 0.9 and NSE values exceeding 0.8 at multiple stations. These improvements confirm that bias correction is critical for enhancing the usability of model-derived discharge products and for improving their representation of both flow magnitude and variability. The consistent superiority of the Random Forest approach for both precipitation and discharge correction aligns with recent advances highlighting the effectiveness of machine-learning-based bias correction methods. The results also carry broader implications for hydrological modelling and water resource planning in Kenya and similar data-scarce regions. Corrected satellite and model-derived datasets can enhance water allocation planning by providing more reliable precipitation and discharge inputs. Such integrated and corrected datasets are likely to be more credible for informing water allocation decisions in catchments that experience recurring water resource conflicts among multiple users.

Crucially, for the Central Highlands of Kenya, a region vital to the nation's water tower, agricultural production, and unique biodiversity, the high-accuracy, bias-corrected CHIRPS and VegDischarge datasets provide the precise inputs needed to transition toward informed, evidence-based water allocation strategies and environmental conservation efforts. More broadly, the study highlights the importance of blended data approaches in supporting transparent, data-driven, and defensible water

resources management frameworks. Furthermore, the demonstrated success of a machine-learning-based bias correction method highlights opportunities to integrate advanced data-driven methods into hydrological workflows. However, the transferability of such models across regions and time periods remains uncertain and warrants further research.

Supporting information

S1 Fig. Monthly rainfall time series plots for the in-situ rain gauge stations.
(TIF)

S2 Fig. Mean monthly discharge time series plots for observed discharge stations.
(TIF)

S1 Table. Validation statistics of the SPPS during the March–April–May (MAM) rainy season.
(DOCX)

S2 Table. Validation statistics of the SPPS during the October–November–December (OND) rainy season.
(DOCX)

S3 Table. Annual validation statistics of the SPPs.
(DOCX)

Acknowledgments

This research was conducted as part of CGIAR Sustainable Farming Science Program and the International Water Management Institute's Digital Innovations for Water Secure Africa (DIWASA). The authors would like to express their gratitude to the Centre for Training and Integrated Research in ASAL Development (CETRAD), The Nature Conservancy (TNC), Kenya and the National Irrigation Authority for their support in this work.

Author contributions

Conceptualization: Felicia Yeboah, Afua Owusu, Naga Velpuri.

Data curation: Afua Owusu, Komlavi Akpoti, Elias Muluken Adamseged, Emma Odera.

Formal analysis: Felicia Yeboah.

Methodology: Felicia Yeboah.

Resources: Komlavi Akpoti, Elias Muluken Adamseged.

Supervision: Afua Owusu, Naga Velpuri.

Validation: Felicia Yeboah.

Visualization: Felicia Yeboah, Afua Owusu, Kirubel Mekonnen.

Writing – original draft: Felicia Yeboah.

Writing – review & editing: Felicia Yeboah, Afua Owusu, Mansoor Leh, Komlavi Akpoti, Kirubel Mekonnen, Elias Muluken Adamseged, Emma Odera, Naga Velpuri.

References

1. Bughici T, Lazarovitch N, Fredj E, Tas E. Evaluation and Bias Correction in WRF Model Forecasting of Precipitation and Potential Evapotranspiration. *Journal of Hydrometeorology*. 2019;20(5):965–83. <https://doi.org/10.1175/jhm-d-18-0160.1>
2. Lazin R, Shen X, Koukoulou M, Anagnostou E. Evaluation of the Hyper-Resolution Model-Derived Water Cycle Components Over the Upper Blue Nile Basin. *Journal of Hydrology*. 2020;590:125231. <https://doi.org/10.1016/j.jhydrol.2020.125231>

3. Islam MdA. Statistical comparison of satellite-retrieved precipitation products with rain gauge observations over Bangladesh. *International Journal of Remote Sensing*. 2018;39(9):2906–36. <https://doi.org/10.1080/01431161.2018.1433890>
4. Muthoni FK, Odongo VO, Ochieng J, Muglavai EM, Mourice SK, Hoesche-Zeledon I, et al. Long-term spatial-temporal trends and variability of rainfall over Eastern and Southern Africa. *Theor Appl Climatol*. 2018;137(3–4):1869–82. <https://doi.org/10.1007/s00704-018-2712-1>
5. Nickles C, Beighley E, Durand M, Prata de Moraes Frasson R. Integrating Lateral Inflows Into a SWOT Mission River Discharge Algorithm. *Water Resources Research*. 2020;56(10). <https://doi.org/10.1029/2019wr026589>
6. Dembélé M, Zwart SJ. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *International Journal of Remote Sensing*. 2016;37(17):3995–4014. <https://doi.org/10.1080/01431161.2016.1207258>
7. Kimani MW, Hoedjes JCB, Su Z. Bayesian Bias Correction of Satellite Rainfall Estimates for Climate Studies. *Remote Sensing*. 2018;10(7):1074. <https://doi.org/10.3390/rs10071074>
8. Macharia JM, Ngetich FK, Shisanya CA. Comparison of satellite remote sensing derived precipitation estimates and observed data in Kenya. *Agricultural and Forest Meteorology*. 2020;284:107875. <https://doi.org/10.1016/j.agrformet.2019.107875>
9. Omondi CK, Rientjes THM, Booij MJ, Nelson AD. Satellite rainfall bias assessment for crop growth simulation – A case study of maize growth in Kenya. *Agricultural Water Management*. 2021;258:107204. <https://doi.org/10.1016/j.agwat.2021.107204>
10. Saber M, Yilmaz KK. Evaluation and Bias Correction of Satellite-Based Rainfall Estimates for Modelling Flash Floods over the Mediterranean region: Application to Karpuz River Basin, Turkey. *Water*. 2018;10(5):657. <https://doi.org/10.3390/w10050657>
11. Toté C, Patricio D, Boogaard H, Van der Wijngaart R, Tarnavsky E, Funk C. Evaluation of Satellite Rainfall Estimates for Drought and Flood Monitoring in Mozambique. *Remote Sensing*. 2015;7(2):1758–76. <https://doi.org/10.3390/rs70201758>
12. Huffman GJ. The Transition in Multi-Satellite Products from TRMM to GPM (TMPA to IMERG) The transition from the Tropical Rainfall Measuring Mission (TRMM) data products to the Global Precipitation Measurement (GPM) mission products is well underway. 2019 https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V6.pdf
13. Kubota T, Aonashi K, Ushio T, Shige S, Takayabu YN, Arai Y, et al. Recent progress in global satellite mapping of precipitation (GSMAP) product. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017. 2712–5. <https://doi.org/10.1109/igarss.2017.8127556>
14. Pan X, Wu H, Chen S, Nanding N, Huang Z, Chen W, et al. Evaluation and Applicability Analysis of GPM Satellite Precipitation over Mainland China. *Remote Sensing*. 2023;15(11):2866. <https://doi.org/10.3390/rs15112866>
15. Maidment RI, Grimes D, Allan RP, Tarnavsky E, Stringer M, Hewison T, et al. The 30 year TAMSAT African Rainfall Climatology And Time series (TARCAT) data set. *J Geophys Res Atmos*. 2014;119(18). <https://doi.org/10.1002/2014jd021927>
16. Xie P, Joyce R, Wu S, Yoo S-H, Yarosh Y, Sun F, et al. Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998. *Journal of Hydrometeorology*. 2017;18(6):1617–41. <https://doi.org/10.1175/jhm-d-16-0168.1>
17. Nguyen P, Ombadi M, Sorooshian S, Hsu K, AghaKouchak A, Braithwaite D, et al. The PERSIANN family of global satellite precipitation data: a review and evaluation of products. *Hydrol Earth Syst Sci*. 2018;22(11):5801–16. <https://doi.org/10.5194/hess-22-5801-2018>
18. Sadeghi M, Nguyen P, Naeini MR, Hsu K, Braithwaite D, Sorooshian S. PERSIANN-CCS-CDR, a 3-hourly 0.04° global precipitation climate data record for heavy precipitation studies. *Sci Data*. 2021;8(1):157. <https://doi.org/10.1038/s41597-021-00940-9> PMID: 34162874
19. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data*. 2015;2:150066. <https://doi.org/10.1038/sdata.2015.66> PMID: 26646728
20. Beck HE, Wood EF, Pan M, Fisher CK, Miralles DG, van Dijk AIJM, McVicar TRF. MSWEP V2 Global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 2019;100(3), 473–500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
21. Funk C, Shukla S, Mamadou Thiaw W, Rowland J, Hoell A, McNally A, et al. Recognizing The Famine Early Warning Systems Network Over 30 Years of Drought Early Warning Science Advances and Partnerships Promoting Global Food Security. *American Meteorological Society*, 2018;1011–27. <https://doi.org/10.1175/BAM>
22. Novella NS, Thiaw WM. African Rainfall Climatology Version 2 for Famine Early Warning Systems. *Journal of Applied Meteorology and Climatology*. 2013;52(3):588–606. <https://doi.org/10.1175/jamc-d-11-0238.1>
23. Gebregiorgis A, Hossain F. Making Satellite Precipitation Data Work for the Developing World. *IEEE Geosci Remote Sens Mag*. 2014;2(2):24–36. <https://doi.org/10.1109/mgrs.2014.2317561>
24. Bhatti HA, Rientjes T, Haile AT, Habib E, Verhoef W. Evaluation of Bias Correction Method for Satellite-Based Rainfall Data. *Sensors (Basel)*. 2016;16(6):884. <https://doi.org/10.3390/s16060884> PMID: 27314363
25. Mekonnen K, Velpuri NM, Leh M, Akpoti K, Owusu A, Tinonetsana P, et al. Accuracy of satellite and reanalysis rainfall estimates over Africa: A multi-scale assessment of eight products for continental applications. *Journal of Hydrology: Regional Studies*. 2023;49:101514. <https://doi.org/10.1016/j.ejrh.2023.101514>
26. Dinku T, Funk C, Peterson P, Maidment R, Tadesse T, Gadain H, et al. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quart J Royal Meteor Soc*. 2018;144(S1):292–312. <https://doi.org/10.1002/qj.3244>
27. Trentini L, Dal Gesso S, Venturini M, Guerrini F, Calmanti S, Petitta M. A Novel Bias Correction Method for Extreme Events. *Climate*. 2023;11(1). <https://doi.org/10.3390/cli11010003>

28. Valdés-Pineda R, Demaría EMC, Valdés JB, Wi S, Serrat-Capdevilla A. Bias correction of daily satellite-based rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa. *Hydrology and Earth System Sciences*. 2016 <https://doi.org/10.5194/hess-2016-473>
29. Abera W, Brocca L, Rigon R. Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin. *Atmospheric Research*. 2016;178–179:471–83. <https://doi.org/10.1016/j.atmosres.2016.04.017>
30. Chaudhary S, Dhanya CT. (2019). Investigating the performance of bias correction algorithms on satellite-based precipitation estimates. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI* (p. 39). SPIE. <https://doi.org/10.1117/12.2533214>
31. Choudhary A, Dimri AP. On bias correction of summer monsoon precipitation over India from CORDEX-SA simulations. *Intl Journal of Climatology*. 2018;39(3):1388–403. <https://doi.org/10.1002/joc.5889>
32. Kagone S, Velpuri NM, Khand K, Senay GB, van der Valk MR, Goode DJ, et al. Satellite precipitation bias estimation and correction using in situ observations and climatology isohyets for the MENA region. *Journal of Arid Environments*. 2023;215:105010. <https://doi.org/10.1016/j.jaridenv.2023.105010>
33. Vernimmen RRE, Hooijer A, Aldrian E, van Dijk AIJM. Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. *Hydrol Earth Syst Sci*. 2012;16(1):133–46. <https://doi.org/10.5194/hess-16-133-2012>
34. Andari R, Nurhamidah N, Daoed D, Marzuki. Evaluation of bias correction methods for multi-satellite rainfall estimation products. *IOP Conf Ser: Earth Environ Sci*. 2024;1317(1):012008. <https://doi.org/10.1088/1755-1315/1317/1/012008>
35. Nigussie AB, Tenfie HW, Zimale FA, Endalew A, Wudiye G. Evaluation of multiple bias correction methods with different satellite rainfall products in the Main Beles Watershed, Upper Blue Nile (Abbey) Basin, Ethiopia. *Journal of Water and Climate Change*. 2022;14(1):156–74. <https://doi.org/10.2166/wcc.2022.244>
36. Iqbal Z, Shahid S, Ahmed K, Wang X, Ismail T, Gabriel HF. Bias Correction Method of High-Resolution Satellite-Based Precipitation Product for Peninsular Malaysia. 2021. <https://doi.org/10.21203/rs.3.rs-1141784/v1>
37. Vigna I, Bigi V, Pezzoli A, Besana A. Comparison and Bias-Correction of Satellite-Derived Precipitation Datasets at Local Level in Northern Kenya. *Sustainability*. 2020;12(7):2896. <https://doi.org/10.3390/su12072896>
38. Elsebaie IH, Kawara AQ, Alharbi R, Alnahit AO. Bias Correction Methods Applied to Satellite Rainfall Products over the Western Part of Saudi Arabia. *Atmosphere*. 2025;16(7):772. <https://doi.org/10.3390/atmos16070772>
39. Yao N, Ye J, Wang S, Yang S, Lu Y, Zhang H, et al. Bias correction of the hourly satellite precipitation product using machine learning methods enhanced with high-resolution WRF meteorological simulations. *Atmospheric Research*. 2024;310:107637. <https://doi.org/10.1016/j.atmosres.2024.107637>
40. M R S, Nair A, Somasundaram K. Spatiotemporal bias correction of satellite precipitation products using multimodel techniques over temporally coherent clusters in South Peninsular India. *Atmospheric Research*. 2025;325:108244. <https://doi.org/10.1016/j.atmosres.2025.108244>
41. Huang Q, Long D, Du M, Han Z, Han P. Daily Continuous River Discharge Estimation for Ungauged Basins Using a Hydrologic Model Calibrated by Satellite Altimetry: Implications for the SWOT Mission. *Water Resources Research*. 2020;56(7). <https://doi.org/10.1029/2020wr027309>
42. Sun W, Fan J, Wang G, Ishidaira H, Bastola S, Yu J, et al. Calibrating a hydrological model in a regional river of the Qinghai–Tibet plateau using river water width determined from high spatial resolution satellite images. *Remote Sensing of Environment*. 2018;214:100–14. <https://doi.org/10.1016/j.rse.2018.05.020>
43. Fang H, Beaudoin HK, Rodell M, Teng WL, Vollmer BE. (2009, March 8). Global Land Data Assimilation System (GLDAS) Products, Services and Application from NASA Hydrology Data and Information Services Center (HDISC). ASPRS 2009 Annual Conference. <http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html>
44. Sutanudjaja EH, Van Beek R, Wanders N, Wada Y, Bosmans JH, Drost N, et al. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*. 2018;11(6), 2429–53. <https://doi.org/10.5194/gmd-11-2429-2018>
45. Müller Schmied H, Cáceres D, Eisner S, Flörke M, Herbert C, Niemann C, et al. The global water resources and use model WaterGAP v2.2d: model description and evaluation. *Geosci Model Dev*. 2021;14(2):1037–79. <https://doi.org/10.5194/gmd-14-1037-2021>
46. Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, et al. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst Sci Data*. 2021;13(9):4349–83. <https://doi.org/10.5194/essd-13-4349-2021>
47. Burek P, Satoh Y, Kahil T, Tang T, Greve P, Smilovic M, Wada Y. Development of the Community Water Model (CWatM v1. 04)—a high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geoscientific Model Development*. 2020;13(7):3267–98. <https://doi.org/10.5194/gmd-13-3267-2020>
48. Akpoti K, Velpuri NM, Mizukami N, Kagone S, Leh M, Mekonnen K, et al. Advancing water security in Africa with new high-resolution discharge data. *Sci Data*. 2024;11(1):1195. <https://doi.org/10.1038/s41597-024-04034-0> PMID: 39500890
49. Senay GB, Kagone S, Parrish GEL, Khand K, Boiko O, Velpuri NM. Improvements and Evaluation of the Agro-Hydrologic VegET Model for Large-Area Water Budget Analysis and Drought Monitoring. *Hydrology*. 2023;10(8). <https://doi.org/10.3390/hydrology10080168>
50. Mizukami N, Clark MP, Gharari S, Kluzek E, Pan M, Lin P, et al. A Vector-Based River Routing Model for Earth System Models: Parallelization and Global Applications. *J Adv Model Earth Syst*. 2021;13(6). <https://doi.org/10.1029/2020ms002434>
51. Saemian P, Hosseini-Mogharri S-M, Fatehi I, Shoarinezhad V, Modiri E, Tourian MJ, et al. Comprehensive evaluation of precipitation datasets over Iran. *Journal of Hydrology*. 2021;603:127054. <https://doi.org/10.1016/j.jhydrol.2021.127054>

52. Chen S, Xiong L, Ma Q, Kim J-S, Chen J, Xu C-Y. Improving daily spatial precipitation estimates by merging gauge observation with multiple satellite-based precipitation products based on the geographically weighted ridge regression method. *Journal of Hydrology*. 2020;589:125156. <https://doi.org/10.1016/j.jhydrol.2020.125156>
53. Sperna Weiland FC, Vrugt JA, van Beek R (L.) PH, Weerts AH, Bierkens MFP. Significant uncertainty in global scale hydrological modeling from precipitation data errors. *Journal of Hydrology*. 2015;529:1095–115. <https://doi.org/10.1016/j.jhydrol.2015.08.061>
54. Moges E, Demissie Y, Larsen L, Yassin F. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water*. 2020;13(1):28. <https://doi.org/10.3390/w13010028>
55. Hales RC, Williams GP, James Nelson E, Sowby RB, Ames DP, Lozano JLS. Bias correcting discharge simulations from the GEOGloWS global hydrologic model. *Journal of Hydrology*. 2023;626:130279. <https://doi.org/10.1016/j.jhydrol.2023.130279>
56. Owusu A, Matheswaran K, Velpuri NM, Magesa R, Schmitter P. Use case report on scenarios of water availability and use in the Central Highland Ecoregions Foodscapes (CHEF) of Kenya. 2024 <https://doi.org/10568/173216>
57. Otieno TA, Otieno LA, Rotich B, Löhr K, Kipkulei HK. Modeling climate change impacts and predicting future vulnerability in the Mount Kenya forest ecosystem using remote sensing and machine learning. *Environ Monit Assess*. 2025;197(6):631. <https://doi.org/10.1007/s10661-025-14089-0> PMID: 40329020
58. Downing T, Olago D, Nyumba T. Perceptions of Ecosystem Services and Climate Change in the Communities Surrounding Mt. Kenya and Mt. Elgon, Kenya. *Sustainability*. 2023;15(14):11470. <https://doi.org/10.3390/su151411470>
59. Yamazaki D, Ikeshima D, Sosa J, Bates PD, Allen GH, Pavelsky TM. MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research*. 2019;55(6):5053–73. <https://doi.org/10.1029/2019wr024873>
60. Gibert K, Sánchez-Marré M, Izquierdo J. A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. *AIC*. 2016;29(6):627–63. <https://doi.org/10.3233/aic-160710>
61. Ullah W, Wang G, Ali G, Tawia Hagan DF, Bhatti AS, Lou D. Comparing Multiple Precipitation Products against In-Situ Observations over Different Climate Regions of Pakistan. *Remote Sensing*. 2019;11(6):628. <https://doi.org/10.3390/rs11060628>
62. Nerini D, Zulkaffi Z, Wang L-P, Onof C, Buytaert W, Lavado-Casimiro W, et al. A Comparative Analysis of TRMM–Rain Gauge Data Merging Techniques at the Daily Time Scale for Distributed Rainfall–Runoff Modeling Applications. *Journal of Hydrometeorology*. 2015;16(5):2153–68. <https://doi.org/10.1175/jhm-d-14-0197.1>
63. Yonaba R, Belemtougri A, Fowé T, Mounirou LA, Nkiaka E, Dembélé M, et al. Rainfall estimation in the West African Sahel: comparison and cross-validation of top-down vs. bottom-up precipitation products in Burkina Faso. *Geocarto International*, 2024;39(1). <https://doi.org/10.1080/10106049.2024.2391956>
64. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357–62. <https://doi.org/10.1038/s41586-020-2649-2> PMID: 32939066
65. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261–72. <https://doi.org/10.1038/s41592-019-0686-2> PMID: 32015543
66. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 2011;12:2825–30.
67. Jackson EK, Roberts W, Nelsen B, Williams GP, Nelson EJ, Ames DP. Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environmental Modelling & Software*. 2019;119:32–48. <https://doi.org/10.1016/j.envsoft.2019.05.001>
68. Akpoti K, Mekonnen K, Leh M, Owusu A, Dembélé M, Tinonetsana P, et al. State of continental discharge estimation and modelling: challenges and opportunities for Africa. *Hydrological Sciences Journal*. 2024;69(15):2124–52. <https://doi.org/10.1080/02626667.2024.2402938>
69. Das P, Ghosh S, Zhang Z, Feng S. Assessment of CHIRPS and PERSIANN-CDR products for meteorological drought monitoring and future trends in East Africa. *Environ Res: Climate*. 2025;4(3):035002. <https://doi.org/10.1088/2752-5295/add28f>
70. Almeida M, Zhu S, Coelho P. Evaluation of eight satellite and reanalysis precipitation products over Angola: The value of targeted regional assessment for water management. *Journal of Hydrology: Regional Studies*. 2025;62:102955. <https://doi.org/10.1016/j.ejrh.2025.102955>
71. Sahlu D, Moges S, Nikolopoulos EI, Anagnostou EN, Hailu D. Evaluation of high-resolution multisatellite and reanalysis rainfall products over East Africa. *Advances in Meteorology*, 2017;(1):4957960. <https://doi.org/10.1155/2017/4957960>
72. Ageet S, Fink AH, Maranan M, Diem JE, Hartter J, Ssali AL, et al. Validation of Satellite Rainfall Estimates over Equatorial East Africa. *Journal of Hydrometeorology*. 2022;23(2):129–51. <https://doi.org/10.1175/jhm-d-21-0145.1>
73. Cattani E, Ferguglia O, Merino A, Levizzani V. Precipitation Products' Inter-Comparison over East and Southern Africa 1983–2017. *Remote Sensing*. 2021;13(21):4419. <https://doi.org/10.3390/rs13214419>
74. Nkiaka E, Tume SJP. Evaluating the performance of high-resolution gridded precipitation products in regions with complex topographies and high precipitation. *Hydrological Sciences Journal*. 2025;71(1):49–61. <https://doi.org/10.1080/02626667.2025.2579875>
75. Kimani M, Hoedjes J, Su Z. An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa. *Remote Sensing*. 2017;9(5):430. <https://doi.org/10.3390/rs9050430>

76. Goudiaby O, Bodian A, Dezetter A, Diouf I, Ogilvie A. Evaluation of Gridded Rainfall Products in Three West African Basins. *Hydrology*. 2024;11(6):75. <https://doi.org/10.3390/hydrology11060075>
77. Tanimu B, Bello A-AD, Abdullahi SA, Ajibike MA, Yaseen ZM, Kamruzzaman M, et al. Comparison of conventional and machine learning methods for bias correcting CMIP6 rainfall and temperature in Nigeria. *Theor Appl Climatol*. 2024;155(6):4423–52. <https://doi.org/10.1007/s00704-024-04888-9>
78. Beyer R, Krapp M, Manica A. An empirical evaluation of bias correction methods for palaeoclimate simulations. *Clim Past*. 2020;16(4):1493–508. <https://doi.org/10.5194/cp-16-1493-2020>
79. Mendez M, Maathuis B, Hein-Griggs D, Alvarado-Gamboa L-F. Performance Evaluation of Bias Correction Methods for Climate Change Monthly Precipitation Projections over Costa Rica. *Water*. 2020;12(2):482. <https://doi.org/10.3390/w12020482>
80. Siderius C, Biemans H, Kashaigili JJ, Conway D. Going local: Evaluating and regionalizing a global hydrological model's simulation of river flows in a medium-sized East African basin. *Journal of Hydrology: Regional Studies*. 2018;19:349–64. <https://doi.org/10.1016/j.ejrh.2018.10.007>
81. Shen Y, Ruijsch J, Lu M, Sutanudjaja EH, Karssenberg D. Random forests-based error-correction of streamflow from a large-scale hydrological model: Using model state variables to estimate error terms. *Computers & Geosciences*. 2022;159:105019. <https://doi.org/10.1016/j.cageo.2021.105019>
82. Hagemann S, Nguyen TT, Ho-Hagemann HTM. A three-quantile bias correction with spatial transfer for the correction of simulated European river runoff to force ocean models. *Ocean Sci*. 2024;20(6):1457–78. <https://doi.org/10.5194/os-20-1457-2024>
83. Liu L, Zhou L, Gusyev M, Ren Y. Unravelling and improving the potential of global discharge reanalysis dataset in streamflow estimation in ungauged basins. *Journal of Cleaner Production*. 2023;419:138282. <https://doi.org/10.1016/j.jclepro.2023.138282>
84. Sanchez Lozano J, Romero Bustamante G, Hales RC, Nelson EJ, Williams GP, Ames DP, et al. A Streamflow Bias Correction and Performance Evaluation Web Application for GEOGloWS ECMWF Streamflow Services. *Hydrology*. 2021;8(2):71. <https://doi.org/10.3390/hydrology8020071>