

Climatological evaluation of satellite and reanalysis precipitation products across temporal scales and rainfall extremes over the Guder Sub-basin, Ethiopia

Rahel Gezahegn Abate

rahelgeza@gmail.com

School of Food and Agriculture Engineering, Department of Agricultural Engineering,

Gebrekidan Worku Tefera

Prairie View A&M University

Temesgen Gashaw Tarkegn

Prairie View A&M University

Negusu Tarekegn Endalew

Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan.

Filagot Mengistu Walle

Institute of Geophysics Space Science and Astronomy, Addis Ababa University, Addis Ababa, Ethiopia

Tena Alamirew Agumassie

Water and Land Resource Centre

Research Article

Keywords:

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10870639/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Climatological evaluation of satellite and reanalysis precipitation products across temporal scales and rainfall extremes over the Guder Sub-basin, Ethiopia

Rahel Gezahegn Abate¹, Gebrekidan Worku Tefera², Temesgen Gashaw Tarkegn³, Negusu Tarekegn Endalew³, Filagot Mengistu Walle⁴, Tena Alamirew Agumassie⁵

¹School of Food and Agriculture Engineering, Department of Agricultural Engineering, Ambo University, Ambo, Ethiopia.

²College of Agriculture, Food and Natural Resources, Prairie View A&M University, Prairie View, Texas, USA.

³Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan.

⁴Institute of Geophysics, Space Science and Astronomy, Addis Ababa University, Addis Ababa, Ethiopia.

⁵Water and Land Resource Center, Addis Ababa University, Addis Ababa, Ethiopia.

Abstract

Precipitation data are essential for exploring climate variability, eco-hydrological modeling, agricultural planning and drought and flood monitoring, and water-resource management. However, sparse and unevenly distributed rain-gauge networks limit the availability of reliable rainfall observations in the Upper Blue Nile Basin. This study evaluates the performance of six widely used satellite and reanalysis precipitation products CHIRPS v2.0 (CHIRPS), IMERG v6.0 (IMERG), TAMSAT v3.1 (TAMSAT), MSWEP v2.8 (MSWEP), PERSIANN-CDR, and ERA5, against gauge observations from seven meteorological stations in the Guder sub-basin during 2001–2015. Performance was assessed at daily, monthly, and seasonal scales using categorical indices (POD, FAR, FBI, CSI), continuous statistics (Bias, R, RMSE, KGE), and ETCCDI-based extreme precipitation indices. For precipitation occurrence, MSWEP and IMERG showed superior detection skill, with higher POD and CSI and lower FAR, whereas ERA5 consistently performed worst. For precipitation totals, clear scale dependency was evident. MSWEP performed best at the daily scale ($R = 0.73$, $KGE = 0.71$, $RMSE = 3.29$ mm), followed by TAMSAT and IMERG. At the monthly scale, performance improved for all products, with

CHIRPS and IMERG emerging as the most reliable datasets, both achieving correlations above 0.97 and strong KGE values. Seasonally, CHIRPS and MSWEP performed better during the small rainy (February–March) and dry (October–January) seasons, while all products showed reduced skill during the main rainy season (June–September) due to intense monsoonal variability. For precipitation extremes, no single dataset consistently outperformed others. MSWEP showed balanced performance for wet/dry spell indices and moderate rainfall, CHIRPS performed better for annual totals and moderate extremes, IMERG captured heavy and very heavy rainfall events (R95p, R20mm) more effectively, and PERSIANN-CDR performed relatively better in rainfall intensity (SDII). ERA5 remained the weakest across most extreme indices. CHIRPS, MSWEP, and IMERG were identified as the most reliable precipitation datasets for hydrological and climate-related applications in the Guder Sub-basin.

1. Introduction

Reliable precipitation information is an essential component of eco-hydrological modeling and water resources management. Precipitation data are widely used in agricultural planning, drought and flood assessment, climate risk analysis, and ecosystem studies. In data-scarce regions, uncertainty in precipitation inputs directly affects streamflow simulation, water balance estimation, and climate impact assessment (Peng et al., 2021; Xiang et al., 2021). This issue is critical in Ethiopia, where rainfall variability strongly influences rain-fed agriculture, food security, water availability, soil erosion, and hydropower generation (Bekele & Gemi, 2020; Hilemelekot et al., 2021; Van der Zwaan et al., 2017). However, obtaining continuous and spatially representative rainfall observations from conventional rain-gauge networks remains a major challenge in Ethiopia. Gauge stations are sparse and unevenly distributed, and many records are affected by missing data, inconsistent temporal coverage, and quality-control limitations (Ayehu et al., 2018; Lakew et al., 2017). These limitations are particularly severe in mountainous basins, where rainfall can vary considerably over short distances because of complex topographic effects (Ayehu et al., 2018; Dinku et al., 2018).

To address these limitations, satellite and reanalysis precipitation products (SREs) have emerged as valuable alternative and complementary sources in data-limited regions. These products integrate satellite observations, rain-gauge measurements, reanalysis outputs, and multi-source merging techniques to generate gridded rainfall estimates (Beck et al., 2019; Funk et al., 2015; Huffman 2019). Products such as CHIRPS, IMERG, TAMSAT, MSWEP, PERSIANN-CDR, and ERA5 are now widely used in hydrological, climatological, agricultural, and environmental studies (Ashouri et al., 2015; Hersbach et al., 2020; Maidment et al., 2017). Nevertheless, their performance is not uniform and varies according to rainfall regime, topographic complexity, spatial and temporal scale, retrieval algorithm, gauge-adjustment procedures, and rainfall intensity (Xiang et al., 2021). Therefore, SREs should be evaluated for specific regional applications to ensure their suitability and reliability.

Several studies have evaluated satellite and gridded rainfall products in different parts of Africa, including those by Bagiliko et al. (2025), Houngnibo et al. (2023), and Waal et al. (2025). These studies reported relatively good performance of products such as TAMSAT, IMERG, RFE2, ARC2, MSWEP, ENACTS, and CHIRPS. Moreover, studies conducted in East Africa by Sahlu

et al. (2017), Dinku et al. (2018), Ageet et al. (2021), Omay et al. (2025), Muhammad et al. (2026), and Ogembo et al. (2026) indicated that CHIRPS2, CMORPH, CHIRPS, CPC-RFEv2, GPM-IMERG, GPCC, and MSWEP generally show better performance across different climatic conditions and temporal scales. In Ethiopia, Ewnetu et al. (2025) identified MSWEP, CHIRPS, and IMERG as the most reliable products across different metrics, rainfall intensities, and terrain zones. Alemu et al. (2025) reported that IMERG performed better at the daily scale, whereas CHIRPS showed superior performance at the monthly scale. Similarly, Addisu et al. (2024) found that CHIRPS and MSWEP better captured rainfall intensity classes in tropical agroecological zones, while TAMSAT performed better in temperate and alpine zones. In the Upper Blue Nile Basin, CHIRPS has generally shown better performance at dekadal and monthly scales compared to TAMSAT and ARC2 (Ayehu et al., 2018). At the same time, hydrological studies suggest that these datasets can still be very useful for water resource applications, provided that they are carefully evaluated and adjusted for local conditions (Lakew et al., 2017). Overall, existing evidence suggests that no single precipitation product consistently outperforms others across different climatic and physiographic settings.

More recently, increasing attention has been directed toward evaluating the ability of satellite and reanalysis precipitation products (SREs) to represent precipitation extremes, including prolonged dry spells, wet spells, heavy rainfall events, and extreme precipitation intensities. Accurate characterization of these events is critical because droughts and floods are among the most significant climate-related hazards affecting agriculture, water resources, infrastructure, and ecosystem sustainability. Previous studies have shown that product performance varies depending on the type and magnitude of the event being analyzed.

CHIRPS has been reported as effective for drought monitoring because of its ability to capture low-rainfall conditions and long-term rainfall variability (Degefu & Bewket, 2023; Lemma et al., 2025), while ARCV2.0 and CHIRPS have demonstrated good skill in representing dry conditions and drought-year rainfall patterns (Wodebo et al., 2025). In contrast, IMERG-Fv6B and TAMSAT have shown greater capability in detecting heavy rainfall events and flood-related extremes (Wodebo et al., 2025). Other studies have highlighted the benefits of MSWEP, CHIRPS, and IMERG for flood forecasting, drought assessment, and hydrological applications across diverse climatic and topographic conditions (Ewnetu et al., 2025; Tsegaye et al., 2025;

Zargar et al., 2025). These findings suggest that evaluations based solely on conventional statistical metrics may not adequately reflect the overall suitability of precipitation datasets. Therefore, assessing both conventional rainfall statistics and the ability of products to capture precipitation extremes is essential for determining their suitability for hydrological and climate-risk applications.

Despite recent advances in the evaluation of SREs, most studies have focused on large river basins and regional-scale domains. Products such as MSWEP, CHIRPS, and IMERG have been reported as reliable for hydrological modeling and water-resource management in the Upper Blue Nile Basin (Ewnetu et al., 2025; Gado et al., 2024). However, evaluations at smaller hydrological units, such as the Guder Sub-basin, remain limited, despite the fact that uncertainties in precipitation estimates can substantially affect hydrological simulations, agricultural water planning, and climate-risk assessments at local scales. Furthermore, precipitation-product performance is not spatially uniform; datasets that perform well at regional scales may exhibit different levels of accuracy in areas characterized by complex terrain, high rainfall variability, and sparse rain-gauge networks (Dinku et al., 2018; Sahlu et al., 2017; Xiang et al., 2021).

The Guder Sub-basin, located within the Upper Blue Nile Basin, is characterized by rugged terrain, pronounced rainfall gradients, and distinct seasonal precipitation patterns. As one of the major tributaries of the Upper Blue Nile Basin, it supports rain-fed agriculture, water-resource development, and ecosystem services. Accurate rainfall information is therefore essential for sustainable watershed management and climate-resilience planning in the region. Despite numerous precipitation-product evaluation studies conducted in Ethiopia and the Upper Blue Nile Basin, significant knowledge gaps remain for the Guder Sub-basin. Most previous studies have focused on large regional domains, a limited number of products, or a single temporal scale. Furthermore, rainfall-event detection capability, precipitation-intensity representation, and extreme-rainfall characteristics have rarely been evaluated simultaneously. Consequently, the relative suitability of different precipitation products for hydrological and climate-related applications in the Guder Sub-basin remains uncertain. To address this gap, the present study evaluates six gridded precipitation datasets, CHIRPS, IMERG, TAMSAT, MSWEP,

PERSIANN-CDR, and ERA5 against in situ rainfall observations in the Guder Sub-basin. The evaluation was conducted at daily, monthly, and seasonal timescales using statistical metrics, rainfall-event detection indices, precipitation-intensity analyses, and extreme-rainfall assessments. The findings are expected to provide guidance for selecting appropriate precipitation products for hydrological, agricultural, and climate-related applications in data-scarce regions.

2. Study area and datasets

2.1. Site Description

The Guder Sub-basin, located in the Upper Blue Nile Basin in Ethiopia, is an important tributary of the Blue Nile River (Figure 1). Covering approximately 7,011 square kilometers, the sub-basin features diverse topography, including high mountains and low valleys, and is situated between 7°30' to 9°30' N latitude and 37°00' to 39°00' E longitude. The Guder River originates in the central highlands of Ethiopia and flows into the Blue Nile. The region has a tropical highland climate with a distinct wet season from June to September, driven by the Inter-Tropical Convergence Zone (ITCZ). Annual rainfall varies from 1,200 mm in lower areas to over 1,800 mm in higher elevations. The Guder Sub-basin's hydrology is influenced by its varied topography and land use, which includes agriculture, forests, and grasslands. The area is prone to soil erosion due to its steep slopes and intensive farming. The sub-basin plays a crucial role in the hydrology of the Upper Blue Nile Basin, contributing significantly to water flow and sediment load. Despite its location near the equator, the Guder Sub-basin has a relatively mild climate due to its high elevation, ranging from 1,500 to 3,000 meters above sea level.

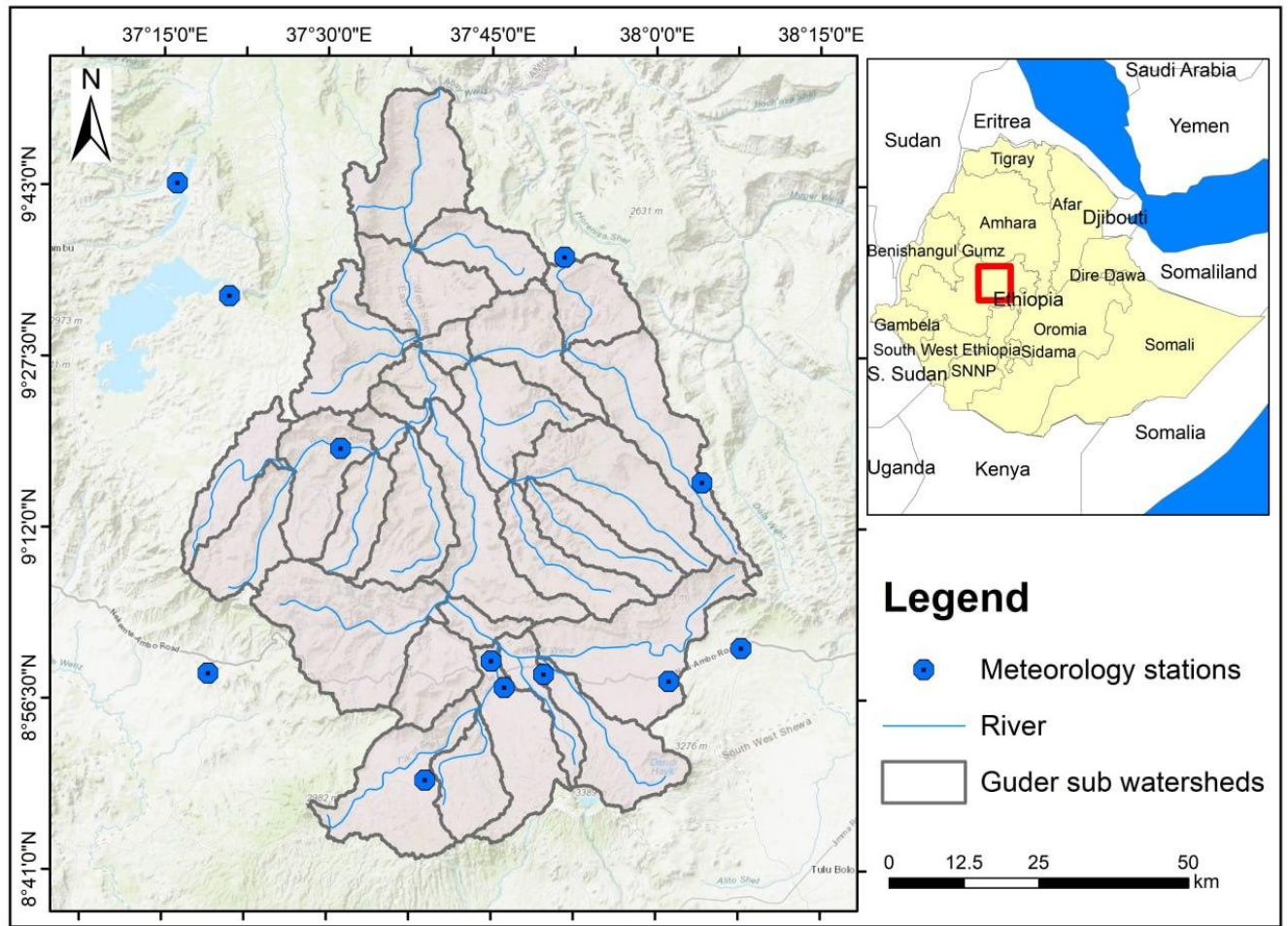


Figure 1. Location map of the Guder sub basin

2.2 Datasets

2.2.1 Observed climate data

In this study, long-term observed daily precipitation data were used to evaluate six satellite rainfall estimates (SREs) for the period 2001–2015. Data from ten rain gauge stations were obtained from the National Meteorological Agency (NMA) of Ethiopia. Of these, seven stations were selected based on data consistency and completeness, with less than 10% missing records as presented on Table 1. Data quality was assessed using the double mass curve method (Searcy and Hardison, 1960), while continuity was evaluated based on the proportion of missing values. Stations with inconsistent or incomplete records were excluded from the analysis. The study period (2001–2015) was chosen because it provided the most reliable and complete gauge

observations. During this period, missing data remained below 10% across the selected stations, ensuring sufficient data quality for validation. Years beyond 2015 were excluded due to increased data gaps and inconsistencies, which could introduce additional uncertainty into the evaluation.

Table 1. Geographical coordinates, elevation, percentage of missing daily rainfall data, and mean annual rainfall of the selected stations (2001–2015)

Station name	Longitude(°E)	Latitude (°N)	Altitude(m)	Missed data (%)
Ambo Met	37.83	8.98	2130	0.33
Fincha	37.35	9.55	2248	1.75
Gedo	37.75	9.00	2520	0.09
Kachise(Rs)	37.86	9.61	2520	1.80
Neshi	37.27	9.72	2060	6.40
Guder	37.77	8.96	2053	0.00
Abebe Keranso	38.02	8.97	2475	6.78

2.2.2 Gridded Precipitation Products

As shown on Table 2, six widely used gridded precipitation datasets, namely CHIRPS, IMERG, TAMSAT, PERSIANN-CDR, MSWEP, and ERA5, were evaluated. These datasets represent different approaches to precipitation estimation. ERA5 is a reanalysis product derived from numerical weather prediction models and data assimilation techniques. In contrast, CHIRPS and MSWEP integrate satellite observations with in-situ gauge measurements and/or reanalysis data to enhance accuracy. Meanwhile, IMERG, TAMSAT, and PERSIANN-CDR are primarily satellite-based products supplemented with gauge-based calibration. The selection of these datasets was based on their relatively high spatial and temporal resolution, broad spatial coverage, and demonstrated performance in regions with complex topography. Previous studies have shown that MSWEP, CHIRPS, and IMERG consistently provide balanced and reliable performance across multiple evaluation metrics and rainfall intensities (Taye et al., 2022; Ahmed et al., 2024; Assamnew et al., 2025; Wodebo et al., 2025; Ewnetu et al., 2025). Moreover, these datasets have been extensively evaluated at global and regional scales and generally show good agreement with ground-based observations (Shen et al., 2020; Dinku et al., 2018; Ageet et al., 2022; Omay et al., 2025; Muhammad et al., 2026). In addition, they have been widely applied in hydro-climatic, agro-climatic, and ecological studies (Satgé et al., 2021; Gebremicael et al., 2022; Wodebo et al., 2025; Mekonnen et al., 2026). However, their performance may vary across different climatic and topographic settings. Therefore, evaluating these products over a medium-sized watershed such as the Guder Sub-basin is essential to support their application in water resources management, agro-hydrological modeling, and ecosystem assessment. Although newer versions of some datasets (e.g., IMERG v07, MSWEP v3, and TAMSAT v4) are available, the versions used in this study were selected to maintain consistency across all datasets and to ensure comparability with previous studies conducted in Ethiopia. Detailed descriptions of each dataset are provided in the following subsections.

CHIRPS (version 2.0)

CHIRPS (version 2.0) (Climate Hazards Group InfraRed Precipitation with Station data) is a widely used satellite-based precipitation dataset that covers land areas between 50°S and 50°N from 1981 to near-present (Funk et al., 2015). It provides high spatial resolution (0.05°) rainfall estimates by blending satellite infrared data with a large network of rain gauge observations and

advanced interpolation techniques. CHIRPS offer precipitation products at daily, 5-day, and monthly precipitation products. This makes it valuable for monitoring drought, analyzing climate variability and change, hydrological modeling, agricultural planning, and food security early warning systems.

IMERG (version 06)

IMERG (version 06) (Integrated Multi-satellitE Retrievals for Global Precipitation Measurement) is a satellite-based precipitation product developed under the GPM mission, combining data from multiple satellite sensors to generate global rainfall estimates (Huffman 2019). It provides half-hourly precipitation data at a spatial resolution of 0.1° . IMERG incorporates calibration using observations from the GPM Core Observatory, significantly improving rainfall estimation accuracy across different climatic regions. The dataset is available in near-real-time and research-quality formats, supporting hydrological and climate studies.

TAMSAT v3.1

TAMSAT v3.1 (Tropical Applications of Meteorology using Satellite data) is a satellite-derived rainfall dataset primarily developed for Africa. It utilizes thermal infrared (TIR) imagery from Meteosat satellites and ground-based rain gauge data for calibration (Maidment et al., 2017). The dataset provides precipitation estimates at a spatial resolution of 0.0375° (~ 4 km). TAMSAT offers multiple temporal scales, including daily, pentadal, dekadal, monthly, and seasonal data, with coverage from 1983 to the present, making it valuable for climate monitoring and agricultural applications.

PERSIANN-CDR

PERSIANN-CDR (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Climate Data Record) provides long-term daily precipitation estimates at a spatial resolution of $0.25^\circ \times 0.25^\circ$ (Ashouri et al., 2015). The dataset covers the latitude band 60°S to 60°N from 1983 to near-present. It is derived from infrared satellite data and bias-corrected using Global Precipitation Climatology Project (GPCP) data. Due to its long-term consistency, PERSIANN-CDR is widely used for climate trend analysis, although its performance may vary in capturing localized rainfall events.

MSWEP v2.8

MSWEP v2.8 (Multi-Source Weighted-Ensemble Precipitation) is a high-resolution global precipitation dataset that integrates gauge observations, satellite estimates, and reanalysis data (Beck et al., 2019). It provides precipitation data at 0.1° spatial resolution and 3-hourly temporal resolution from 1979 to near-present. Advanced bias correction and merging techniques enhance its reliability, making MSWEP particularly suitable for hydrological modeling and climate studies, especially in data-scarce regions.

ERA5

ERA5 is the fifth-generation atmospheric reanalysis dataset developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al., 2020). It provides hourly precipitation data at a spatial resolution of approximately 0.25° (~31 km) with global coverage from 1950 to the present. ERA5 combines numerical weather prediction models with a wide range of observational data through data assimilation techniques. This enables it to have consistent and physically coherent estimates for climate and hydrological analyses.

Table 2. Summary of satellite and reanalysis rainfall products used in this study

Datasets	Data source	Resolution	Frequency	Time Latency	Source
CHIRPS v2.0	SRG	0.05°	Daily	2 weeks	https://data.chc.ucsb.edu/products/CHIRPS-2.0/
IMERG v6.0	SG	0.1°	Half-hourly	12 hours	https://gpm.nasa.gov/data
TAMSAT v3.1	SG	0.0375°	Daily		https://www.tamsat.org.uk
MSWEP v2.8	SRG	0.1°	3-hourly	3 hours	http://www.gloh2o.org/
ERA5	R	0.25°	Daily	6 days	https://cds.climate.copernicus.eu/
PERSIANN-CDR	SG	0.25°	Daily	≈30–90 days	https://www.ncei.noaa.gov/data/precipitation/persiann/access/

R= Reanalysis; SG =Satellite Ground; SRG Satellite reanalysis ground

3. Research Methodology

Most of the SRE products, including CHIRPS, IMERG, ERA5, and PERSIANN-CDR, were processed and downloaded from Google Earth Engine (GEE) in CSV format. Precipitation data were extracted from the grid cell (pixel) corresponding to each rain gauge station. The remaining datasets, MSWEP and TAMSAT, were obtained in NetCDF format from their respective official sources (Table 2). Temporally, the datasets were aligned with the study period. Sub-daily datasets, including half-hourly IMERG and 3-hourly MSWEP, were aggregated to daily precipitation totals (mm) using Python for subsequent analysis and application of evaluation metrics. The evaluation was conducted at multiple temporal scales, including daily, monthly, and seasonal timescales (FMAM, JJAS, and ONDJ). A point-to-pixel comparison approach was adopted due to the sparse and uneven distribution of rain gauge stations within the Guder sub-basin. Such an approach is widely used in the evaluation of gridded precipitation products in Ethiopia and other data-scarce regions, where gauge density is insufficient to support reliable spatial interpolation (Addisu et al., 2024; Alemu et al., 2025).

3.1 Precipitation Totals

The ability of the datasets to reproduce observed precipitation amounts was evaluated using the correlation coefficient (R), root mean square error (RMSE), bias, and Kling–Gupta efficiency (KGE). These metrics quantify the agreement between estimated and observed precipitation in terms of temporal variability, magnitude, systematic error, and overall performance (Pearson, 1909; Gupta et al., 2009; Billah et al., 2006). The evaluation was conducted at daily, monthly, and seasonal timescales. Seasonal performance was assessed for the major climatic seasons of the study area, namely FMAM (February–May), JJAS (June–September), and ONDJ (October–January). The mathematical expressions, ideal values, and performance ranges of all evaluation metrics are summarized in Table 3 & 4.

Table 3. Descriptions of evaluation metrics used in the article. A, B and C represent hits, false alarms and misses respectively

Name	Formula	Range	Perfect score
Bias	$BIAS = \frac{\sum_{i=1}^N (O_i - P_i)}{N}$	0 to ∞	1
Correlation	$r = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}}$	-1 to 1	1
Root Mean Square Error (mm)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}$	0 to $+\infty$	0
Kling Gupta Efficiency	$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$	$-\infty$ to 1	0
Probability of Detection	$POD = \frac{A}{A + C}$	0 to 1	1
False Alarm Ratio	$FAR = \frac{B}{A + B}$	0 to 1	0
Critical Success Index	$CSI = \frac{A}{A + B + C}$	0 to 1	1
Frequency of Bias Index	$FBI = \frac{A + B}{A + C}$	0 to $+\infty$	1

3.2 Precipitation Occurrences

The categorical performance of the evaluated rainfall products was assessed using probability of detection (POD), false alarm ratio (FAR), frequency bias index (FBI), and critical success index (CSI). These metrics evaluate the ability of the products to detect precipitation events, minimize false alarms, and represent event frequency (Dinku et al., 2010; Wilks, 2011; Ferro and Stephenson, 2012). The contingency table used for analyzing the detection ability of the SREs is presented in Table 4. The mathematical expressions, ideal values, and possible ranges of these metrics are presented in Table 3. These categorical metrics were assessed based on precipitation intensity by classifying rainfall events into five categories, following established approaches in precipitation verification studies (Wilks, 2011; Dinku et al., 2018). The intensity thresholds were defined based on commonly used classifications in the literature and are consistent with general meteorological guidelines (WMO; Wodebo et al., 2025), including no rain (< 1 mm/day), very

light rain (1–5 mm/day), light rain (5–10 mm/day), moderate rain (10–30 mm/day), and heavy rain (> 30 mm/day).

Table 3. Contingency table used for analyzing the detection ability of the SREs

SREs	Gauge measurement		
	Yes	No	Total
Yes	Hit(A)	False Alarm(B)	A+B
No	Miss(C)	Correct Negative (D)	C+D
Total	A+C	B+D	A+B+C+D

3.3 Precipitation Extremes

Extreme precipitation characteristics were evaluated using indices recommended by the ETCCDI Alexander et al., (2006). These indices have been widely used for characterizing rainfall extremes and have been applied in similar climatic regions in previous studies (Mekonnen et al., 2020; Tsegaye et al., 2025; Muhammad et al., 2026; Tarkegn et al., 2026). Prior to analysis, negative daily precipitation values from both observed and SRE datasets were removed. These indices were derived in Python following the definitions recommended by the ETCCDI. The analysis framework was implemented in Python using the pandas, NumPy, Matplotlib, and Seaborn libraries for data processing, statistical evaluation, and visualization. ETCCDI indices were computed from daily precipitation time series, and evaluation metrics were subsequently derived to assess the agreement between satellite/reanalysis products and observed precipitation records. Continuous statistical metrics, including bias, R, RMSE and KGE were calculated again for each ETCCDI precipitation index to evaluate the capability of the satellite and reanalysis products from representing extreme precipitation events perspective. The selected extreme precipitation indices included consecutive dry days (CDD) and consecutive wet days (CWD), moderate-to-heavy rainfall frequency indices (R10mm and R20mm), very heavy and extreme precipitation indices (R95p and R99p), as well as annual total wet-day precipitation (PRCPTOT) and the simple daily intensity index (SDII) (Table 5). These metrics were calculated for each ETCCDI precipitation index to evaluate the capability of the products in representing extreme

precipitation characteristics. Consistent data screening procedures were applied throughout the analysis to exclude missing or invalid values before metric computation.

Table 4. ETCCDI-based extreme precipitation indices used in this study

Index	Definition (ETCCDI standard)	Unit	Hydrological interpretation
CDD	Maximum number of consecutive days with precipitation < 1 mm	days	Duration of dry spells / drought persistence
CWD	Maximum number of consecutive days with precipitation ≥ 1 mm	days	Persistence of wet spells
R10mm	Number of days with precipitation ≥ 10 mm	days	Frequency of moderate heavy rainfall events
R20mm	Number of days with precipitation ≥ 20 mm	days	Frequency of intense rainfall events
R95p	Annual total precipitation from days above 95th percentile	mm	Contribution of very heavy rainfall to total rainfall
R99p	Annual total precipitation from days above 99th percentile	mm	Contribution of extreme rainfall events
PRCPTOT	Annual total precipitation from wet days (≥ 1 mm)	mm	Overall rainfall availability
SDII	PRCPTOT divided by number of wet days	mm day ⁻¹	Mean precipitation intensity

4. Results and Discussion

4.1. Performances for Capturing Precipitation Occurrences

As shown in Figure 2, the performance of most SREs in categorical metrics depends on the rainfall threshold; their skill decreases as rainfall intensity increases. At low rainfall thresholds (1–5 mm/day), ERA5, CHIRPS, and MSWEP demonstrated relatively high detection ability. This is due to the limitations of indirect retrieval algorithms used in SREs, which tend to introduce both systematic and random errors especially for higher rainfall events. POD and CSI

decline as threshold values increase, indicating reduced effectiveness in capturing higher rainfall intensities.

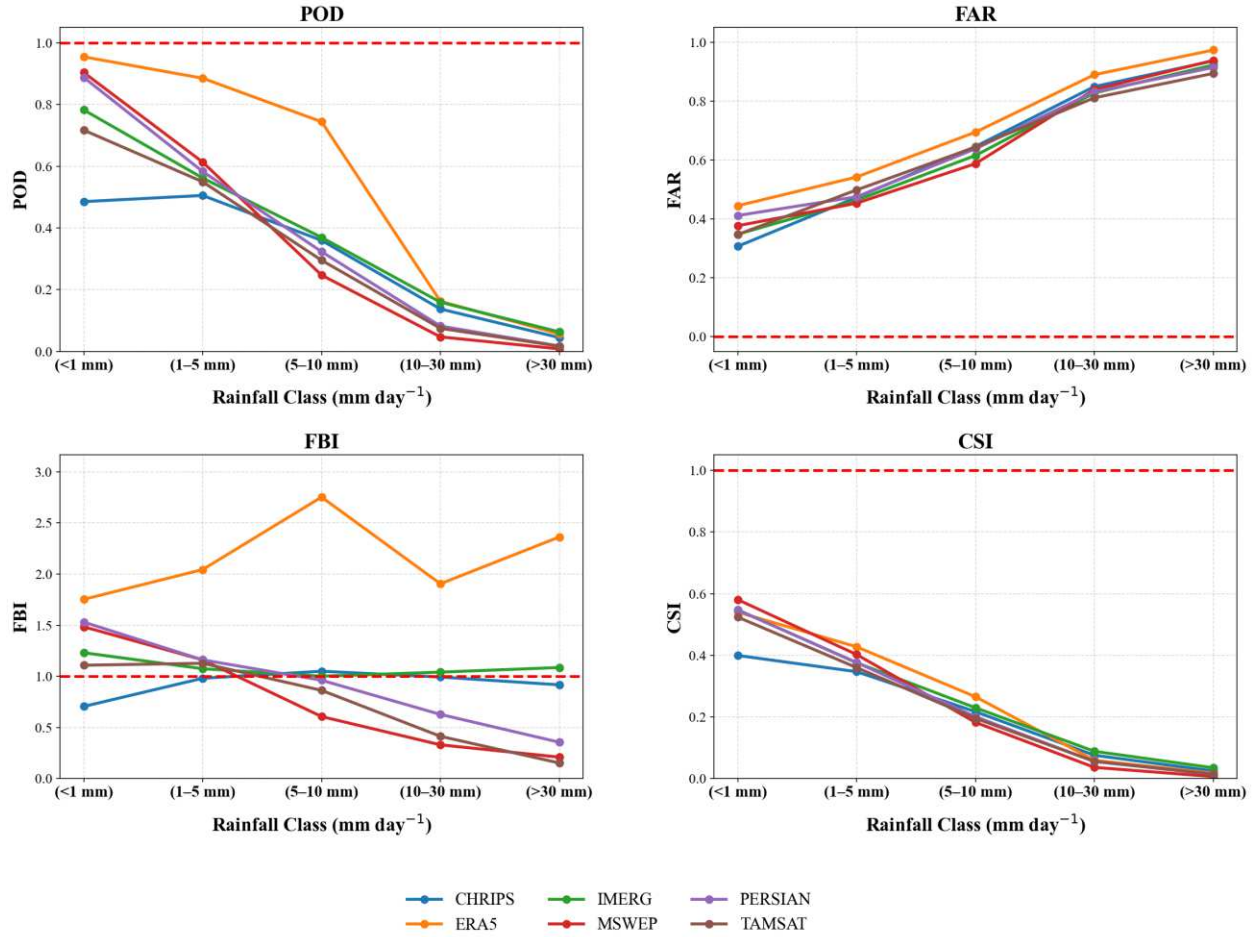


Figure 2. Daily rainfall detection ability of SREs using four categorical metrics. A red dashed line represents the optimal value of each categorical performance metric.

Similar findings have been reported in previous studies (Mekonnen et al., 2020; Bagiliko et al., 2025; Wodebo et al., 2025; Jhonny et al 2026). Bagiliko et al. (2025) showed that detection skill is high for dry days (POD ~70–100%), moderate for light rainfall (<50%), and very low for heavy rainfall events. Wodebo et al. (2025) also reported a decline in SRE performance with increasing rainfall intensity indicating weak skill to capture moderate to heavy rainfall events. Meanwhile, the false alarm ratio (FAR) increases, for ERA5, indicating a tendency to overestimate rainfall, which is also supported by their frequency bias index (FBI > 1). Ewnetu et al. (2025) also reported high FBI values for ERA5, indicating an overestimation of observed rainfall.

Table 5. The continuous statistical evaluation of daily and monthly gridded precipitation datasets analyzed in the Guder sub-basin from 2001 to 2015

	Daily				Monthly			
	Bias	R	KGE	RMSE	Bias	R	KGE	RMSE
CHIRPS	0.88	0.59	0.45	4.82	0.88	0.98	0.88	26.7
ERA5	2.03	0.52	-0.13	9.42	2.03	0.84	-0.05	165.5
IMERG	1.01	0.64	0.59	4.53	1.01	0.97	0.96	24.5
MSWEP	0.90	0.73	0.71	3.29	0.90	0.97	0.88	30.4
PERSIANN-CDR	0.92	0.64	0.59	4.23	0.91	0.94	0.88	37.0
TAMSAT	0.86	0.67	0.63	3.81	0.86	0.95	0.85	37.7

4.2 Performances for Estimating Total Precipitation

4.2.1 Daily Precipitation

As shown in Table 6, MSWEP showed the best performance in representing daily rainfall variability among the evaluated SREs over the Guder Sub-basin. It achieved highest correlation of 0.73, lowest RMSE of 3.29 mm, and the strongest KGE of 0.71. Similarly, at the monthly scale, MSWEP also demonstrated strong performance with high correlation ($R = 0.97$), low RMSE (30.4 mm), and KGE of 0.88. These results reflect the strength of MSWEP's multi-source integration and bias correction strategies, which minimize regional errors. Similarly, Ewnetu et al. (2025) reported that MSWEP outperformed other satellite rainfall products, with correlation coefficients ranging from 0.50 to 0.64 across both highland and lowland regions. These findings suggest that MSWEP provides superior performance and can be used as an alternative data source for various water resources applications, as also reported by Lakew et al. (2020) and Aniley et al. (2023).

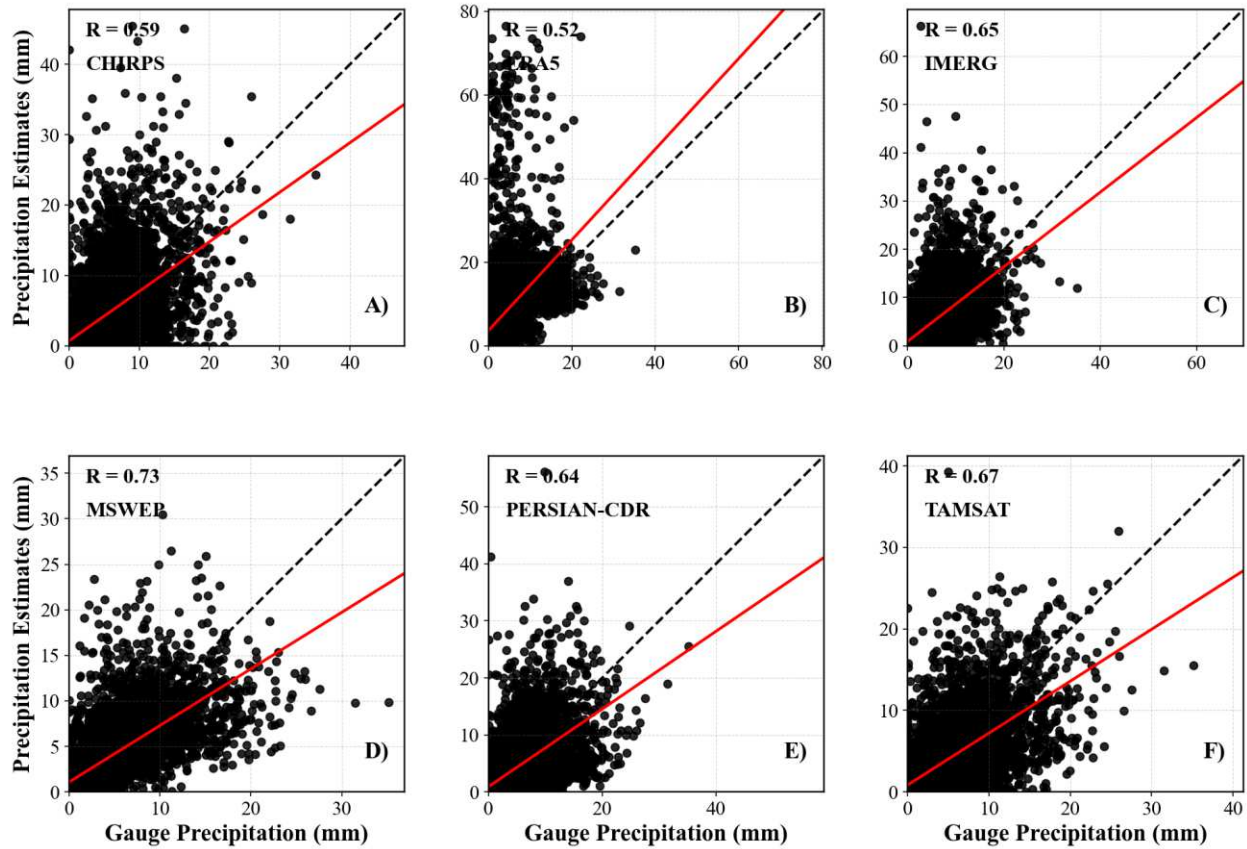


Figure 3. Scatter plots of daily average rainfall across the basin from satellite rainfall estimates against gauge observations for the period 2001–2015.

TAMSAT and PERSIANN-CDR also showed good performance, with moderate correlations, low RMSE values and acceptable KGE at both daily and monthly scales. At the monthly scale, TAMSAT and PERSIANN-CDR achieved correlations of 0.95 and 0.94 respectively, with KGE values of 0.85 and 0.88, indicating strong capability in reproducing total precipitation. IMERG and CHIRPS demonstrated moderate performance in estimating total precipitation. Both products exhibited strong monthly-scale correlations (0.97 and 0.98, respectively), with KGE values of 0.96 and 0.88, and relatively low RMSE values (24.5 mm and 26.7 mm). Their biases were close to unity which shows only slight deviations from observed rainfall totals. ERA5 performed the weakest among all evaluated products. It showed lower correlation at the daily scale ($R = 0.52$), a negative KGE (-0.13), and the highest RMSE (9.42 mm). At the monthly scale, ERA5 still showed relatively lower skill with $R = 0.84$, negative KGE (-0.05), and very high RMSE (165.5 mm), indicating overestimation of total precipitation.

Moreover, Figure 3 scatter plots compare precipitation estimates against ground gauge precipitation measurements in millimeters. Among the products evaluated, MSWEP demonstrates the highest linear correlation with ground measurements ($R = 0.73$), whereas ERA5 exhibits the lowest correlation ($R = 0.52$).

4.2.2. Monthly Precipitation

The monthly precipitation evaluation as shown on Figure 4 and Table 6 reveals different results compared to the daily scale. All products showed improvement at the monthly scale, with correlations ranging from 0.84 to 0.98 and KGE values ranging from 0.85 to 0.96 except for ERA5 with poor KGE value. Figure 4 shows that the scatter plots indicate all SRE products exhibit strong correlation with the observed data.

CHIRPS and IMERG demonstrated the best monthly performance. CHIRPS achieved the highest correlation value of 0.98 and a strong KGE of 0.88, while IMERG achieved the highest KGE of 0.96 with a correlation of 0.97. Both products showed low RMSE values of 26.7 mm for CHIRPS, 24.5 mm for IMERG and near-unity biases (0.88 and 1.01, respectively). Previous studies (Dinku et al., 2018; Hordofa et al., 2021; Ahmed et al. 2024; Gado et al., 2024; Zargar et al., 2025). Similarly, Hordofa et al., 2021 and [Alemu et al., 2025](#) underpinned that both GPM-IMERG and CHIRPS performed best at the monthly scale, where the CHIRPS products slightly outperformed the GPM-IMERG products. Moreover, reported that the strong capability of CHIRPS to reproduce monthly observed rainfall patterns with high correlation and low bias. Wedajo et al. (2021) reported that CHIRPS showed superior performance in the Dhidhessa River Basin, exhibiting higher r and E values and lower bias, ME, and RMSE at both annual and monthly scales. MSWEP closely follows CHIRPS and IMERG with high accuracy of correlation (0.97), KGE (0.88), and a moderate RMSE of 30.4 mm. Almeida et al. (2025) also reported that MSWEP and IMER show reasonable skill, with KGE values up to 0.4, and generally perform better than the other datasets across all timescales.

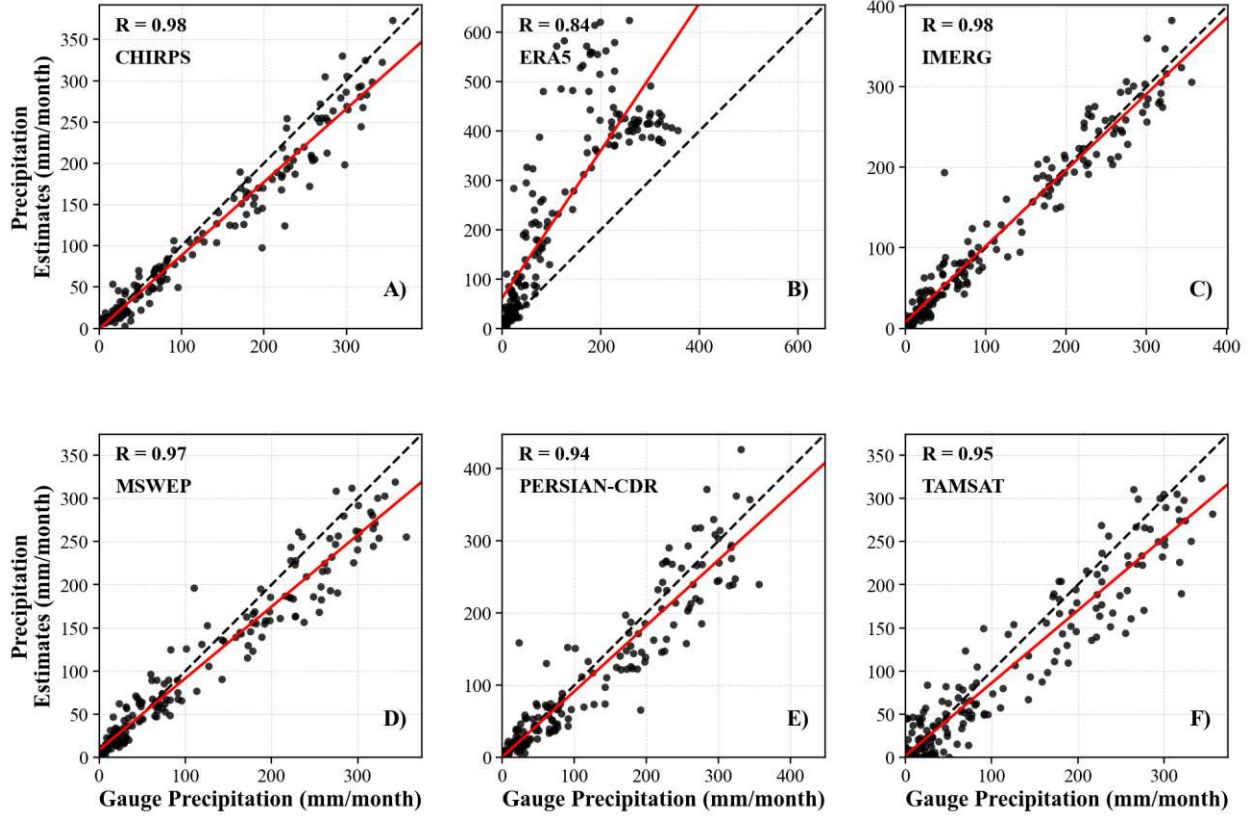


Figure 4. Scatter plots of monthly average rainfall across the basin from satellite rainfall estimates against gauge observations for the period 2001–2015

The strong performance of MSWEP can be attributed to its multi-source merging approach and bias correction procedures. Thus, their high KGE and low bias at monthly scales make them suitable for water balance studies and hydrological modeling in data-sparse regions. PERSIANN-CDR and TAMSAT show moderate performance, with correlation coefficients of 0.94 and 0.95, respectively. However, their slightly higher RMSE values near 37mm suggest minor deviations from observed rainfall. In contrast, ERA5 remained the poorest performer at the monthly scale, with a negative KGE (−0.05), a large bias of 2.03, and the highest RMSE (165.5 mm), which reflect substantial overestimation and unreliability in complex terrain. Likewise, poor performance of ERA5 has also been reported in studies (Almeida et al., 2025; Wodebo et al., 2025) in regions with complex topography.

4.3.3. Seasonal Precipitation

Figure 6 shows the seasonal evaluation of SREs over the Guder sub-basin across the three main rainfall regimes. During (FMAM), strong performance is observed for MSWEP and CHIRPS. In

addition Figs. 5 and 7 present the scatter plot and cumulative distribution function (CDF) respectively. MSWEP demonstrates high agreement with observations with KGE of 0.87, R of 0.88, bias of 0.96, while CHIRPS also performs robustly with KGE of 0.82 and R of 0.92. Wedajo et al. (2021) and Taye et al. (2023) suggest that CHIRPS has better performance in capturing the small rainy season precipitation in the Dhidhessa River Basin (DRB) and the Central Highlands of the Abbay Basin, Ethiopia, respectively. Moderate performance is shown by TAMSAT, which captures variability reasonably well with R of 0.69 but exhibits a dry bias of 0.74. PERSIANN-CDR and IMERG fall into the low-to-moderate performance category, with KGE values ranging from 0.47 to 0.53 and correlations between 0.49 and 0.55. In contrast, ERA5 shows poor performance, characterized by a large positive bias (2.56) and negative KGE (−0.60) that shows its poor reliability during this season.

JJAS (Kiremt) is the dominant rainfall season in the basin and presents the greatest challenge for all precipitation products. Overall, correlation analysis shows that most products exhibit good agreement with observations, except ERA5 reanalysis and MSWEP, which show poor correlations. However, performance based on KGE indicates generally poor skill for most products, with the exception of IMERG and TAMSAT, which demonstrate better efficiency compared to the other datasets. Regarding bias, all products show good performance, with values ranging from 0.87 to 0.99, whereas ERA5 reanalysis exhibits a higher bias of 1.77. In terms of RMSE, all products except IMERG show a high error magnitude which indicates reduced accuracy to capture peak monsoonal rainfall. Overall, although most products show acceptable correlation during JJAS, their hydrological performance remains limited due to poor KGE and high error levels, particularly under intense rainfall conditions (Figure 5). These limitations are consistent with challenges in representing mesoscale convective systems, as reported in previous studies (Taye et al., 2022; Aniley et al., 2023; Fraga et al., 2024). For example, Xin et al., 2022 found that during JJAS, all precipitation products showed stronger agreement with station observations, although RMSE remained high, reflecting its sensitivity to precipitation magnitude. Across all products, JJAS remains the most challenging season, with KGE values generally low and not exceeding 0.54. This confirms that the main rainy season is consistently the most difficult period for satellite and reanalysis precipitation products to accurately capture. Several studies have reported similar findings across different regions (Gashaw et al., 2026; Ng et al., 2026; Wodebo et al., 2026). For instance, Gashaw et al. (2026) found that SREs performance

deteriorates during JJAS, while they reproduce rainfall more reliably during MAM across the South and North Gojjam sub-basins. Similarly, Ng et al. (2026) reported that intense seasonal rainfall such as JJAS reduces product reliability in tropical regions like Peninsular Malaysia. Therefore, for water resource management and hydrological applications, these products require regional calibration or advanced multi-source data fusion approaches to improve their reliability.

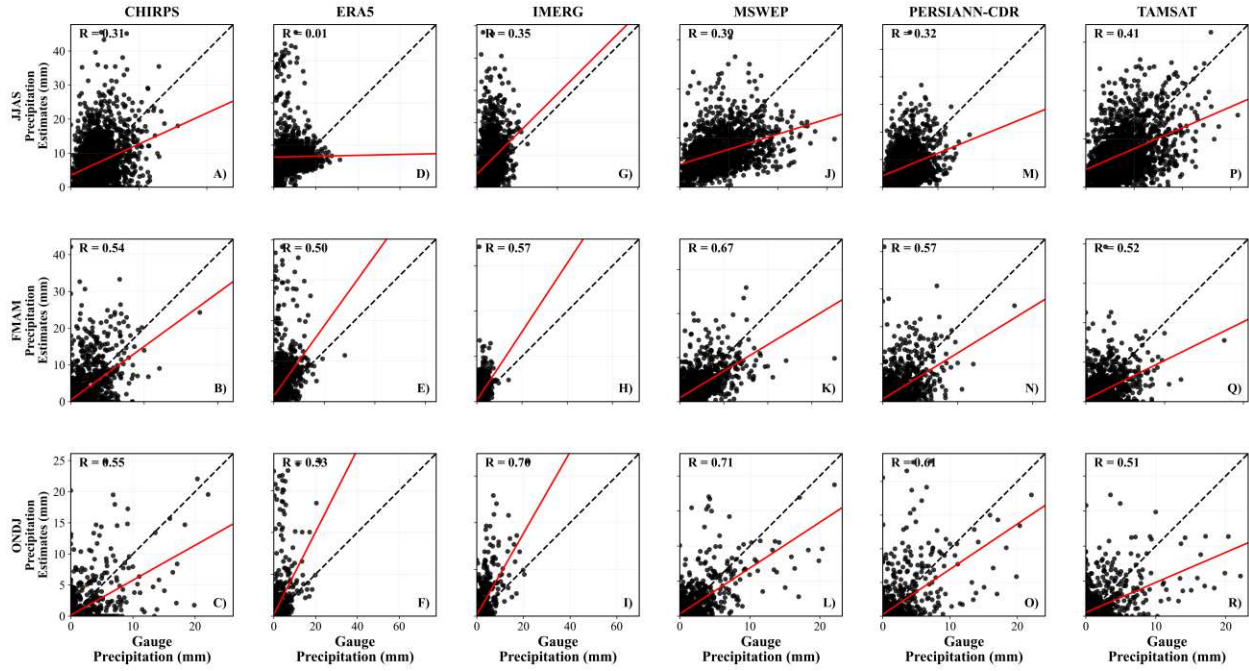


Figure 5. Scatter plots of seasonal average rainfall across the basin from satellite rainfall estimates against gauge observations for the period 2001–2015

During ONDJ (Bega), performance improves for most precipitation products due to lower rainfall intensity and frequency. CHIRPS and MSWEP exhibit strong performance, with high correlation values of 0.90 for both and good KGE values of 0.69 and 0.64, respectively, along with low RMSE values of 33.1 mm and 23.2 mm. PERSIANN-CDR also shows moderate to strong performance, with R and KGE value of 0.74 and 0.63 respectively. TAMSAT demonstrates low to moderate skill, with R = 0.39 and KGE = 0.29 which indicates reduced capability during the post-rainy season. TAMSAT demonstrates low to moderate skill, with R = 0.39 and KGE = 0.29 which indicate reduced performance during the post-rainy season. This may be due to limitations of thermal infrared (TIR) retrieval methods under low-rainfall conditions, where light and scattered precipitation is difficult to detect using satellite-based thermal infrared techniques. ERA5 reanalysis and IMERG show contrasting behavior in ONDJ.

IMERG performs well in terms of correlation ($R = 0.93$) and KGE (0.78), although it still exhibits bias. ERA5 remains the weakest performer, with low efficiency (KGE = -0.95) and substantial bias (2.92), indicating persistent overestimation.

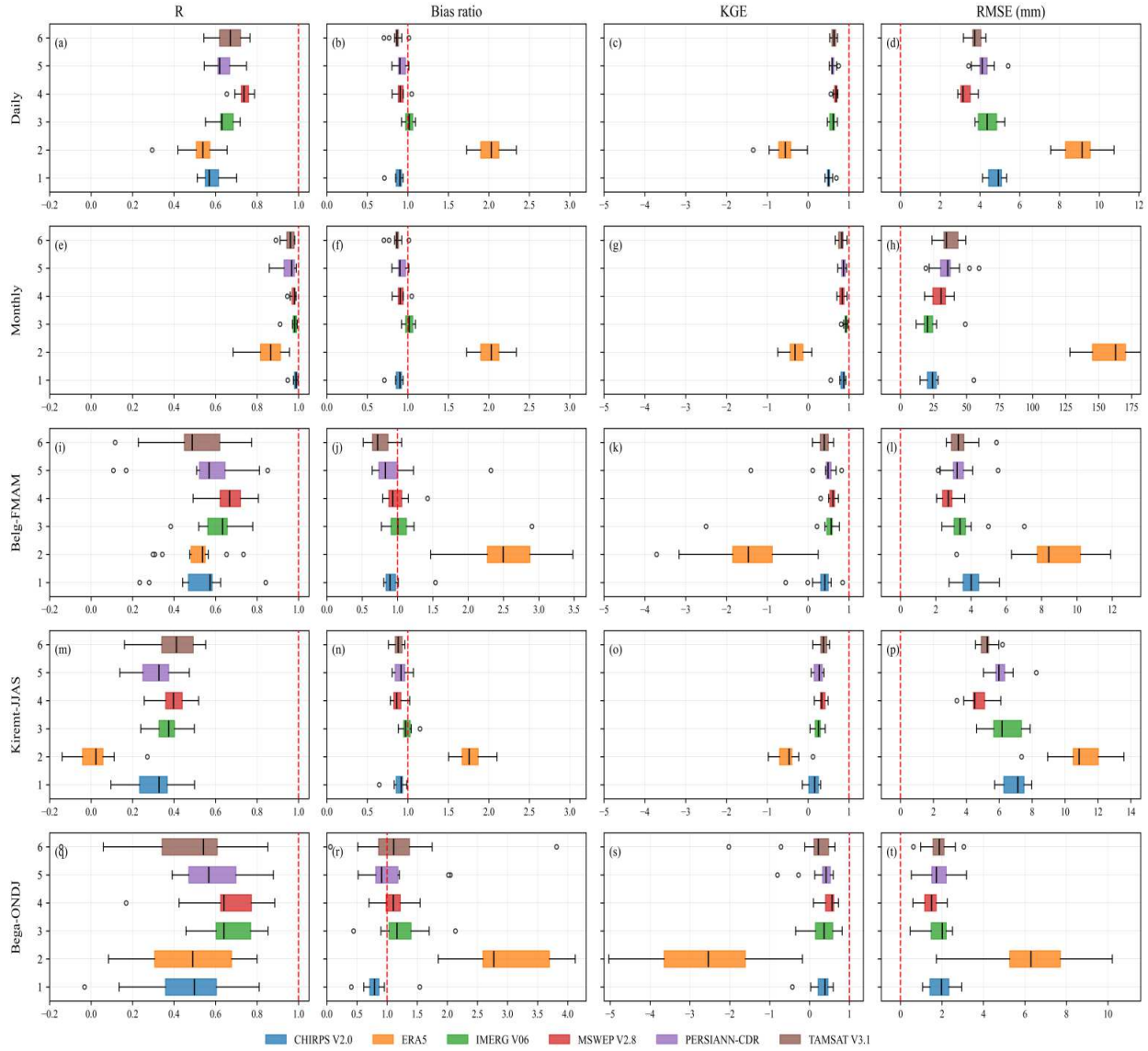


Figure 6. Comparative performance of SREs over Guder Sub basin across daily, monthly, and seasonal timescales based on continuous evaluation metrics. The red dashed line indicates the optimal value of each metric, while boxes, whiskers, and circles represent the interquartile range, data spread, and outliers, respectively.

As shown on Figure 7, MSWEP and TAMSAT show strong agreement with the observed cumulative distribution function (CDF) during the small rainy season. They effectively capture both the frequency and intensity of rainfall events across different magnitudes. These shows that their enhanced capability to capture both rainfall occurrence and intensity characteristics in the study area. For JJAS, PERSIANN-CDR and TAMSAT align closely with observations, while CHIRPS, MSWEP, and TAMSAT demonstrate similar consistency during ONDJ. PERSIANN-CDR and TAMSAT show reasonable agreement at lower to moderate rainfall intensities but tend to underestimate the upper tail of the distribution, reflecting limitations in detecting heavy precipitation events.

IMERG provides a realistic representation of the observed precipitation distribution across the 3 seasons (Figure 7). The close correspondence between the IMERG and gauge-based CDFs suggests that the product accurately captures the frequency and magnitude of daily rainfall events. Minor rightward deviations observed at DNDJ imply a slight overestimation of moderate rainfall intensities; however, these discrepancies are substantially smaller than those observed for ERA5. Across all three seasons, ERA5 demonstrates a rightward shift relative to the observed CDFs. This indicates systematic overestimation of rainfall intensity which affects its suitability for hydrological and extreme-event applications without bias correction. These deviations are consistent with their large positive biases and poor error metrics observed in the statistical analyses. In general, the CDF analysis demonstrates that CHIRPS (excluding FMAM), MSWEP, PERSIANN-CDR (excluding ONDJ), and TAMSAT provide a more accurate representation of precipitation distribution across the Guder watershed.

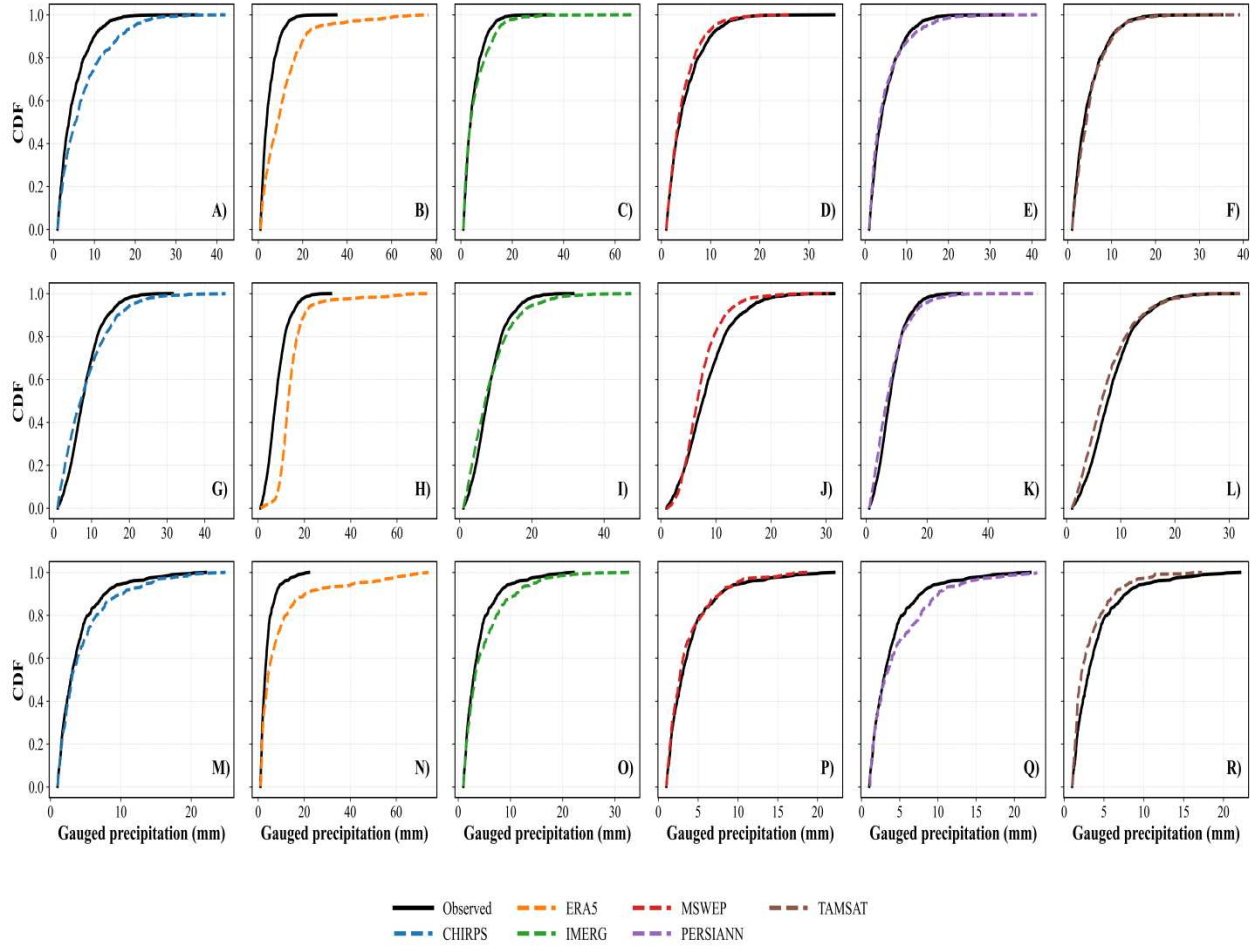


Figure 7. Cumulative distribution functions of precipitation of observed and SREs.

4.3. Performances of Capturing Precipitation Extremes

4.3.1. Performance on Consecutive Dry and Wet Days

The performance of the satellite and reanalysis precipitation products for (CDD) and (CWD) varied considerably (Figure 8). For CDD, PERSIANN-CDR showed the best overall performance, with the highest correlation of 0.83, highest KGE of 0.69, and the lowest RMSE of 14.43. Although it exhibited a moderate negative bias of -10.20, its overall agreement with observations was superior to the other products. MSWEP also performed well, with a strong correlation of 0.72, KGE of 0.61, and low RMSE of 15.22.

These results suggest improved representation of prolonged dry spell conditions. ERA5 and IMERG showed moderate performance, whereas TAMSAT displayed weak correlation of 0.13 and low KGE of 0.13. CHIRPS also showed relatively weak performance, with low correlation of 0.42 despite a moderate KGE of 0.25. In contrast, all products showed weaker skill in capturing CWD across most evaluation metrics. MSWEP again emerged as the best-performing dataset, with the highest correlation of 0.47, highest KGE of 0.45, and the lowest RMSE of 20.79. Improved representation of CWD by MSWEP indicates greater reliability for applications involving prolonged wet spells and associated hydrological responses. ERA5 ranked second, recording a correlation of 0.37 and KGE of 0.33, although it had a positive bias of 18.20. CHIRPS, IMERG, PERSIANN-CDR, and TAMSAT exhibited poor performance, characterized by low correlations of $R \leq 0.09$ for some datasets and strongly negative KGE values. IMERG and PERSIANN-CDR showed very poor KGE values of -1.21 and -0.76, respectively. Those results infer their weak performance to reproduce wet spell persistence. These findings are consistent with previous studies such as Xiao et al. (2020), which reported poor performance of most satellite precipitation products in representing CWD. This limitation may be due to the high spatial and temporal variability of prolonged rainfall events.

4.3.2. Performance on Moderate and Heavy Rainfall Days

The evaluated datasets showed mixed performance for moderate rainfall events (R10mm) and heavy rainfall events (R20mm) (Figure 8). For R10mm, ERA5 and CHIRPS recorded the highest correlations of 0.72. Moreover, CHIRPS achieved the best overall performance with positive KGE of 0.53, relatively small bias of 1.73, and low RMSE of 6.26. This is consistent with the findings of Tsegaye et al. (2025), who reported that CHIRPS outperformed other products in capturing moderate and heavy rainfall events across subtropical agro-ecological zones. They also suggest CHIRPS as a suitable product for flood and drought forecasting, water management, and agricultural planning. IMERG also performed well with a correlation of 0.68 and KGE of 0.66.

ERA5 exhibited a very large positive bias, high RMSE, and strongly negative KGE, indicating systematic overestimation despite its high correlation. MSWEP and PERSIANN-CDR demonstrated moderate skill, while TAMSAT showed the weakest performance with low

correlation and negative KGE. These findings suggest that CHIRPS and IMERG are relatively more capable of detecting moderate rainfall events. For R20mm, performance declined across all products. IMERG achieved the highest correlation and a positive KGE, together with a relatively small bias and RMSE value of 0.55, 0.12, 5.20 and 5.92 respectively. Similarly, Mekonnen et al. (2020) underlined that IMERG demonstrated the best performance for the R10mm index and also performed well for the R20mm index in both highland and lowland regions.

CHIRPS also showed moderate skill with a correlation of 0.49, although its KGE was strongly negative. MSWEP recorded the lowest RMSE value of 2.81 and smallest bias of -1.20, but its correlation remained weak value 0.12. ERA5 performed very poorly, with negative correlation, extremely large bias, and highly negative. TAMSAT and PERSIANN-CDR also exhibited weak performance with negative KGE values. The generally poor performance for R20mm may be attributed to the highly localized and convective nature of intense rainfall events, which are difficult to capture accurately in regions with complex topography. Similar findings were reported by Ahmed et al. (2024) and Ewnetu et al. (2025) very heavy rainfall occurrences linked to flood generation and erosive hydrological events.

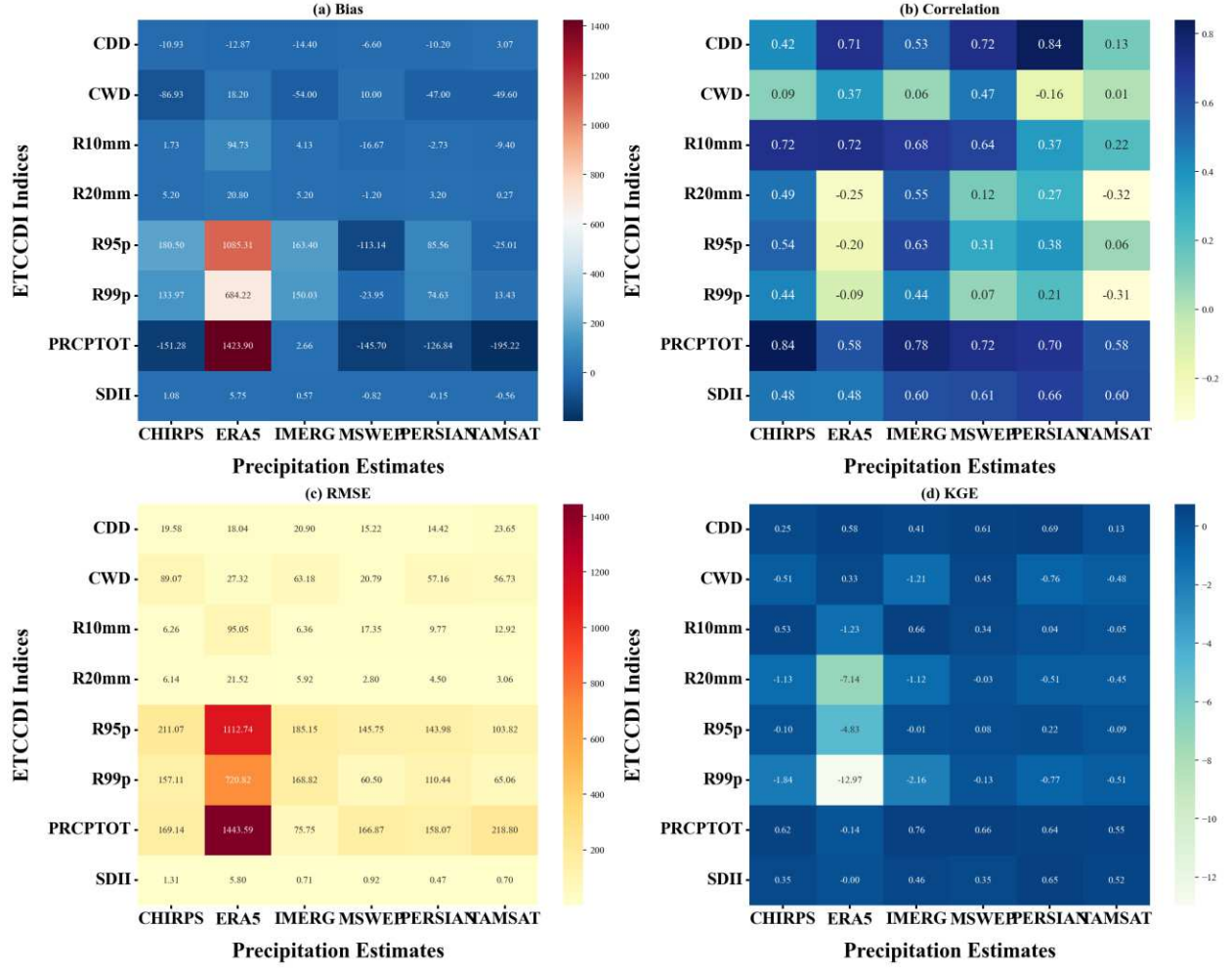


Figure 8. Heatmap comparison of precipitation datasets based on (a) Bias, (b) Correlation, (c) RMSE, and (d) KGE for selected extreme precipitation indices.

4.3.3. Performance on Very Heavy and Extreme Rainfall Indices

The performance of the datasets further deteriorated for very heavy and extreme precipitation indices (R95p and R99p). For R95p, IMERG achieved the highest correlation and relatively strong KGE value of 0.68 and 0.53 respectively. Thus, it is suitable for analyzing extreme rainfall and assessing climate-related risks. Previous studies by Ageet et al. (2022) and Zhang et al. (2023) also identify IMERG as a promising dataset, highlighting its strong capability in detecting extreme rainfall events and its overall accuracy.

However, it exhibited a large positive bias and high RMSE, which indicate a pronounced overestimation of rainfall magnitude. This may be due to the limited effectiveness of gauge correction caused by the sparse and uneven distribution of stations in Ethiopia which reduces the accuracy of bias adjustment (Xu et al 2017; Xie et al., 2022). CHIRPS also showed moderate performance with a correlation of 0.50 and KGE of 0.50. MSWEP and PERSIANN-CDR demonstrated moderate skill, whereas ERA5 performed poorly with negative correlation (-0.08), large positive bias (329.86), very high RMSE (403.05), and strongly negative KGE (-1.04). TAMSAT displayed weak performance with low correlation (0.17) and near-zero KGE (-0.01).

Similarly, for R99p, all products showed limited skill in capturing extreme rainfall variability. PERSIANN-CDR achieved the best overall performance with the highest correlation (0.37) and highest KGE (0.28), together with relatively low RMSE (56.51) and moderate bias (7.75). MSWEP and IMERG showed moderate performance, while CHIRPS and TAMSAT exhibited weak correlations and near-zero or negative KGE values. ERA5 again showed the poorest performance, with negative correlation (-0.05), extremely large bias (101.02), high RMSE (156.90), and negative KGE (-1.33). The reduced skill of all products in representing R95p and R99p may result from the localized nature of convective storms, coarse spatial resolution of satellite products, and smoothing effects in reanalysis datasets. Similar findings were reported by Mekonnen et al. (2020), Bhattacharyya et al. (2022), and Fraga et al. (2024), who highlighted the challenges of capturing extreme rainfall events in mountainous and topographically complex regions. Therefore, caution is required when applying these datasets for flood risk assessment, climate extremes analysis, and hydrological hazard studies.

4.3.4. Performance on Annual Total and Intensity Indices

For annual total wet-day precipitation (PRCPTOT), CHIRPS demonstrated the best overall performance with highest correlation (0.84), positive KGE (0.61), relatively small RMSE (169.14), and lower negative bias (-151.28). IMERG and MSWEP also showed good skill, with correlations of 0.78 and 0.72, and positive KGE values of 0.76 and 0.66, respectively. PERSIANN-CDR performed reasonably well with a KGE of 0.64 and comparatively low RMSE (158.07). In contrast, ERA5 performed very poorly despite a moderate correlation (0.58), due to

its extremely large positive bias (1423.90), very high RMSE (1443.59), and negative KGE (-0.14). TAMSAT also showed weak performance with lower correlation and high RMSE.

For the simple daily intensity index (SDII), PERSIANN-CDR showed the best overall performance by achieving the highest KGE (0.65) and correlation (0.66), lowest RMSE (0.47), and a very small negative bias (-0.15). TAMSAT also performed well, with KGE of 0.52 and low RMSE (0.70). IMERG and MSWEP showed moderate skill with positive KGE values of 0.46 and 0.35, respectively. CHIRPS displayed acceptable performance, whereas ERA5 again showed weaker performance due to its large positive bias and comparatively high RMSE. These findings agree with previous studies such as Tsegaye et al. (2025), who reported that CHIRPS and PERSIANN-CDR effectively capture annual precipitation totals and rainfall intensity indices across Ethiopian highland basins.

5. Conclusion

This study evaluated the performance of six satellite and reanalysis precipitation products, namely CHIRPS, IMERG, TAMSAT, MSWEP, PERSIANN-CDR, and ERA5, against observed rainfall data in the Guder Sub-basin, Upper Blue Nile Basin, Ethiopia, during the period 2001–2015. The assessment was conducted at daily, monthly, and seasonal timescales using continuous statistical metrics, categorical rainfall detection indices, and ETCCDI-based extreme precipitation indices. The results demonstrated considerable variability in the performance of the evaluated datasets across temporal scales, rainfall regimes, and precipitation intensity classes. At the daily scale, MSWEP showed the best overall performance, with the highest correlation, strongest KGE, and lowest RMSE, indicating superior capability in reproducing daily rainfall variability. TAMSAT and IMERG also demonstrated acceptable performance, whereas ERA5 consistently showed poor reliability because of substantial overestimation and weak efficiency. At the monthly scale, all products improved considerably, with CHIRPS and IMERG emerging as the most reliable datasets due to their high correlation, low RMSE, and strong KGE values. Seasonal analysis revealed that product performance strongly depends on rainfall regime. CHIRPS and MSWEP performed better during the FMAM and ONDJ seasons, while all products exhibited reduced skill during the JJAS season because of difficulties in capturing intense monsoonal rainfall and convective precipitation systems. ERA5 remained the weakest performer across all seasons. The evaluation of extreme precipitation indices further indicated that no single product consistently outperformed others for all indices. MSWEP demonstrated the most balanced and reliable performance for consecutive dry and wet days and moderate rainfall indices. CHIRPS showed strong skill for PRCPTOT and moderate rainfall events, while IMERG performed relatively better in detecting heavy and extreme rainfall events such as R20mm and R95p.

PERSIANN-CDR demonstrated comparatively better capability in representing rainfall intensity (SDII). However, most datasets showed limited skill in capturing very extreme precipitation indices, highlighting the challenges of representing localized convective rainfall in complex topographic regions. Overall, CHIRPS, MSWEP, and IMERG were identified as the most suitable datasets for hydrological modeling, drought and flood assessment, agricultural planning, and water-resource management in the Guder Sub-basin. The findings emphasize that

precipitation product performance is highly region- and application-dependent; therefore, local evaluation and calibration are essential before operational application. Future studies should focus on bias-correction techniques, multi-source data fusion approaches, and the integration of additional ground observations to improve precipitation estimation in data-scarce mountainous regions. The findings underline the importance of regional evaluation and product-specific selection before applying satellite rainfall estimates in complex topographic and data-scarce regions.

Funding

This research received no specific funding from any public, commercial, or not-for-profit organization.

References

- Addisu, S., Aniley, E., Gashaw, T., Kelemu, S., & Fetene Demessie, S. (2024). Evaluating the performances of gridded satellite products in simulating the rainfall characteristics of Abay Basin, Ethiopia. *Sustainable Environment*, 10(1), 2381349.
- Ageet, S., Fink, A. H., Maranan, M., Diem, J. E., Hartter, J., Ssali, A. L., & Ayabagabo, P. (2022). Validation of satellite rainfall estimates over equatorial East Africa. *Journal of Hydrometeorology*, 23(2), 129–151.
- Ahmed, J. S., Buizza, R., Dell’Acqua, M., Demissie, T., & Pè, M. E. (2024). Evaluation of ERA5 and CHIRPS rainfall estimates against observations across Ethiopia. *Meteorology and Atmospheric Physics*, 136(3), 17.
- Alemu, M. M., Zaitchik, B. F., Enku, T., Worqlul, A. W., Yimer, E. A., & van Griensven, A. (2025). Statistical and hydrological evaluation of remotely sensed rainfall products in the Upper Blue Nile basin, Ethiopia. *Modeling Earth Systems and Environment*, 11(2), 99.
- Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., ... & Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111(D5).
- Aniley, E., Gashaw, T., Abraham, T., Demessie, S. F., Bayabil, H. K., Worqlul, A. W., ... & Wubaye, G. B. (2023). Evaluating the performances of gridded satellite/reanalysis products in representing the rainfall climatology of Ethiopia. *Geocarto International*, 38(1), 2278329.
- Ashouri, H., Hsu, K. L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., & Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1), 69–83.
- Assamnew, A. D., Gobie, B. G., & Habtemicheal, B. A. (2025). Evaluating the accuracy of gridded precipitation datasets and CMIP6 GCM ensemble means against gauge observations in South Wollo, Ethiopia. *Scientific Reports*, 15(1), 34197.
- Bagiliko, J., Stern, D., Ndanguza, D., & Torgbor, F. F. (2025). Validation of satellite and reanalysis rainfall products against rain gauge observations in Ghana and Zambia. *Theoretical and Applied Climatology*, 156(5), 241.
- Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., Huffman, G. J., Adler, R. F., & Wood, E. F. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data. *Hydrology and Earth System Sciences*, 23(1), 207–224.

- Bhattacharyya, S., Sreekesh, S., & King, A. (2022). Characteristics of extreme rainfall in different gridded datasets over India. *Atmospheric Research*, 267, 105930.
- Billah, B., Hyndman, R. J., & Koehler, A. B. (2005). Empirical information criteria for time series forecasting model selection. *Journal of Statistical Computation and Simulation*, 75(10), 831–840.
- Dinku, T., Ruiz, F., Connor, S. J., & Ceccato, P. (2010). Validation and intercomparison of satellite rainfall estimates over Colombia. *Journal of Applied Meteorology and Climatology*, 49(5), 1004–1014.
- de la Fraga, P., Del-Toro-Guerrero, F. J., Vivoni, E. R., Cavazos, T., & Kretzschmar, T. (2024). Evaluation of gridded precipitation datasets in mountainous terrains of Northwestern Mexico. *Journal of Hydrology: Regional Studies*, 56, 102019.
- Ewnetu, S. S., Dessie, M., Belete, M. A., van Griensven, A., Walraevens, K., Frankl, A., ... & Verhoest, N. E. (2025). Spatial and temporal evaluation of gridded precipitation products over Lake Tana Basin, Ethiopia. *Water*, 17(24), 3536.
- Ferro, C. A., & Stephenson, D. B. (2012). Deterministic forecasts of extreme events and warnings. In *Forecast verification: A practitioner's guide in atmospheric science* (pp. 185–200).
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations (CHIRPS). *Scientific Data*, 2(1), 150066.
- Gashaw, T. T., Melesse, A. M., & Abate, B. (2025). Performance assessment of satellite-based rainfall products in the Abbay Basin, Ethiopia. *Remote Sensing*, 18(1), 2.
- Gebremicael, T. G., Deitch, M. J., Gancel, H. N., Croteau, A. C., Haile, G. G., Beyene, A. N., & Kumar, L. (2022). Satellite-based rainfall estimates evaluation using a parsimonious hydrological model. *Atmospheric Research*, 266, 105939.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria. *Journal of Hydrology*, 377(1–2), 80–91.
- Han, J., Miao, C., Gou, J., Zheng, H., Zhang, Q., & Guo, X. (2023). A new daily gridded precipitation dataset for the Chinese mainland. *Earth System Science Data*, 15(7), 3147–3161.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Simmons, A. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.
- Hordofa, A. T., Leta, O. T., Alamirew, T., Kawo, N. S., & Chukalla, A. D. (2021). Performance evaluation of satellite-derived rainfall datasets over the Ziway lake basin. *Climate*, 9(7), 113.

- Houngnibo, M. C., Minoungou, B., Traore, S. B., Maidment, R. I., Alhassane, A., & Ali, A. (2023). Validation of high-resolution satellite precipitation products over West Africa. *Frontiers in Climate*, 5, 1185754.
- Huffman, G. (2019). IMERGquality index. NASA GPM IMERG documentation.
- Lakew, H. B., Moges, S. A., & Asfaw, D. H. (2020). Hydrological performance evaluation of satellite precipitation products in the Upper Blue Nile basin. *Journal of Hydrology: Regional Studies*, 27, 100664.
- Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., ... & Alcántara, E. M. U. (2017). A new long-term daily satellite-based rainfall dataset for Africa. *Scientific Data*, 4(1), 170063.
- Mekonnen, K., Velpuri, N. M., Leh, M., Akpoti, K., Owusu, A., Tinonetsana, P., ... & Munzimi, Y. (2023). Accuracy of satellite and reanalysis rainfall estimates over Africa. *Journal of Hydrology: Regional Studies*, 49, 101514.
- Montiel, J. I. P., Epiayu, A. M. C., & Moscote, C. D. (2026). Evaluation of satellite precipitation products across climatic and topographic gradients. *Environmental Challenges*, 101426.
- Muhammad, W. I., Wu, Z., Gebresellase, S. H., Ibrahim, A., & Badji, O. (2026). Multi-criteria framework for evaluating gridded precipitation products in the Sahel region. *Journal of Hydrology: Regional Studies*, 64, 103240.
- Ng, C. Y., Othman, F., Wan Jaafar, W. Z., Lai, S. H., Mei, Y., Juneng, L., & Soo, E. Z. X. (2026). Comparative evaluation of seven precipitation products in tropical Malaysia. *Stochastic Environmental Research and Risk Assessment*, 40(2), 45.
- Ogembo, V., Mukama, E. B., Ronoh, E. K., & Akinyi, G. (2026). Assessment of remote sensing precipitation products for drought monitoring in Tanzania. *Climate*, 14(2), 36.
- Omay, P. O., Muthama, N. J., Oludhe, C., Kinama, J. M., Artan, G., & Atheru, Z. (2025). Evaluation of satellite-based rainfall estimates over the IGAD region. *Meteorology and Atmospheric Physics*, 137(2), 22.
- Pearson, K. (1909). Determination of the coefficient of correlation. *Science*, 30(757), 23–25.
- Satgé, F., Pillot, B., Roig, H., & Bonnet, M. P. (2021). Are gridded precipitation datasets a good option for streamflow simulation across the Juruá River basin? *Journal of Hydrology*, 602, 126773.
- Searcy, J. K., & Hardison, C. H. (1960). Double-mass curves (No. 1541). US Government Printing Office.

- Tarkegn, T. G., Ray, R. L., Tefera, G. W., & Wubetu, S. T. (2026). Performance of gridded temperature datasets in the Brazos River Basin. *Frontiers in Climate*, 8, 1731069.
- Taye, M., Mengistu, D., & Sahlu, D. (2023). Performance evaluation of satellite rainfall datasets in Abbay Basin. *European Journal of Remote Sensing*, 56(1), 2233686.
- Tsegaye, S., Mohammed, J. A., Tarkegn, T. G., Ray, R. L., & Tefera, G. W. (2025). Evaluation of rainfall products for rainfall extremes in Jemma sub-basin. *Physical Geography*, 1–29.
- Wedajo, G. K., Muleta, M. K., & Awoke, B. G. (2021). Performance evaluation of satellite rainfall products for Dhidhessa River Basin. *Atmospheric Measurement Techniques*, 14(3), 2299–2316.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic Press.
- Wodebo, D. Y., Melesse, A. M., Woldesenbet, T. A., Mekonnen, K., Amdihun, A., Korecha, D., & Teshome, A. (2025). Comprehensive performance evaluation of satellite-based rainfall products in Ethiopia. *Journal of Hydrology: Regional Studies*, 57, 102150.
- Xiao, S., Xia, J., & Zou, L. (2020). Evaluation of multi-satellite precipitation products over the Yangtze River Basin. *Water*, 12(4), 1179.
- Xie, W., Yi, S., & Leng, C. (2022). Impacts of gauge data bias on satellite precipitation evaluation. *Water*, 14(12), 1860.
- Xu, J., Ma, Z., Yan, S., & Peng, J. (2022). Do ERA5 and ERA5-Land outperform satellite precipitation products? *Journal of Hydrology*, 605, 127353.
- Zargar, M., Bronstert, A., Francke, T., Zimale, F. A., Worku, K. B., Wiegels, R., & Kunstmann, H. (2025). Comparison and hydrological evaluation of precipitation data for the Blue Nile Basin. *Frontiers in Water*, 7, 1536881.
- Zhang, L., Xin, Z., Zhang, C., Song, C., & Zhou, H. (2022). Bias-corrected satellite precipitation for streamflow simulation. *Journal of Hydrology: Regional Studies*, 43, 101192.