https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Evaluation of NDVI-Based downscaled precipitation datasets in the Peruvian Andes

Jose P. Terán[1,*], Jhon Vila[2,*], Juan García-Quijano[3], Anne Gobin[2]

[1]Department of Water and Climate, Vrije Universiteit Brussel, Brussels, 1050, Belgium
[2]Department of Earth and Environmental Sciences, KU Leuven, Leuven, 3000, Belgium
[3]Hydrology and Water Management, WSP, Miraflores, Lima, Perú

* *These authors contributed equally to this work.*

Correspondence to: José P. Terán (jose.pablo.teran.orsini@vub.be) & Jhon F. Vila Solier (jvila.solier@gmail.com)

Abstract. Remotely sensed or model-based gridded precipitation products are increasingly used for hydrological assessments, especially in data scarce regions. However, their coarse spatial resolution often limits the accurate representation of daily rainfall variability and extremes, particularly in complex mountainous terrain. To address this, several methods have been developed to downscale these products and improve their spatiotemporal representation. In the Peruvian Andes, sparse and discontinuous rain gauge networks have led to a strong reliance on satellite-based precipitation products for hydrological assessments, despite persistent uncertainties in their spatio-temporal accuracy. In this study, we developed daily downscaled precipitation datasets covering the Andean region of Peru based on the Global Precipitation Measurement (GPM) IMERG product using parsimonious relationships, including exponential regression (EXP) and geographically weighted regression (GWR), incorporating environmental covariates such as NDVI, topography, and geographical location. These downscaled datasets were evaluated against in situ meteorological stations and benchmarked against regionally optimised products, such as PISCO and Rain4PE, using statistical performance indices, indicators of overall rainfall pattern representation, and extreme-value metrics. Results show that Rain4PE achieves the highest overall performance with a median Kling-Gupta Efficiency (KGE) of 0.67, whereas PISCO faces some limitations with overall performance (median KGE ≈ 0.25), but it excels at representing extreme precipitation. GPM-based datasets exhibit systematic limitations in both temporal coherence and detection of extremes. The GWR approach enhanced spatial detail while preserving the performance of the source GPM dataset (median KGE ≈ 0.28), outperforming the simpler EXP (median KGE ≈ 0.21) method without degrading temporal coherence and extreme-event representation. These findings highlight the potential of parsimonious downscaling strategies to improve precipitation datasets in complex, data-scarce mountainous regions, while underscoring the continued importance of regional gauge-informed products.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

1 Introduction

To adequately assess the hydrological processes of a watershed, the quality of the input data is essential. In the case of Peru, the continuity and quality of precipitation time series are associated with different uncertainties that compromise the quality data of conventional rainfall station data (Hunziker et al., 2017). In the highlands and rainforest regions in particular, there is a low density of pluviometry stations (SENAMHI, 2017), which could lead to an inadequate spatial representation of rainfall, considering that even within the same pixel in the same Andean area comparisons between rainfall measurements showed around 5 to 7% differences (Padrón et al., 2020). Therefore, biased or uncertain results are likely when applying this data to more specific purposes (e.g., hydrological modelling). Due to the limited availability and spatial density of climate data, satellite-based precipitation products have been produced, such as PISCO (Aybar et al., 2020; SENAMHI, 2017), Rain4PE (Fernandez-Palomino et al., 2021), NASA's Tropical Rainfall Measuring Station (TRMM), and Global Precipitation Measurement (GPM), the latter of which is the most recent and provides global coverage, including tropical and subtropical regions (Blumenfeld, 2015). GPM's Integrated Multi-satellitE Retrievals (IMERG) processes raw satellite measurements into 'Early', 'Late', and 'Final' products, each with a higher processing level (Huffman et al., 2014). The final product is agrid of half-hourly, daily, and monthly precipitation data at a 0.1° × 0.1° resolution.

Studies around the world have evaluated and assessed these types of products for use in hydrological assessments; some have found that products developed before GPM (e.g., TRMM, National Oceanic and Atmospheric Administration Satellite-Based Rainfall Estimates, PERSIANN) can adequately represent the monthly discharges of a watershed, when used as an input for rainfall in hydrological modelling. However, the daily values were usually over- or under-estimated, and usually presented a bias (Artan et al., 2007; Su et al., 2008). Furthermore, provided the grid size of the product, performance is usually better in larger catchments (Yilmaz et al., 2005).

The GPM product undergoes multiple stages of processing and is available at different temporal frequencies. Daily and sub-daily GPM rainfall estimates perform better than their predecessor, the TRMM, for hydrological modelling in China (Tang et al., 2016), which is a significant improvement for this purpose. In addition, the product was used to generate probability distribution functions, and the results were comparable to those obtained from ground stations. However, systematic bias, overestimation of peak flows, and delayed diurnal peaks were detected. The representation of dry areas was the weakest point, indicating that improvements were warranted (Tang et al., 2016). GPM has shown good performance in mid- to low-latitude mountain regions, particularly at altitudes below 3000-4000 m, as opposed to higher-range mountains (Sun et al., 2022). A study conducted in Malaysia using the SWAT model in combination with GPM to simulate a significant flood in 2014-2015, compared the different processing stages of the GPM product, and showed that the final stage (IMERG_F) of the daily precipitation estimation outperformed simulations using ground station data (Tan et al., 2018). However, GPM's resolution may be too coarse to represent basin-scale hydrology, and studies have shown that the quality of the spatial representation of rainfall significantly influences hydrological analysis, particularly in hydrological models (Immerzeel et al., 2009).

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Furthermore, hydrological assessments generally perform better when spatially distributed datasets are used instead of point data (i.e., ground-based meteorological measurements) (Guo et al., 2004; Smith et al., 2004).

In the Peruvian Andes, similar results to those from other regions of the world have been observed for the performance of satellite precipitation products (SPPs). Studies have found that the TRMM multi-satellite precipitation analysis (TMPA) can capture the most prominent seasonal patterns in the Andean region. However, it generally fails to do so on a daily basis. There is improved performance with temporal aggregation, where deficiencies in the detection of orographic rainfall are one of the main drawbacks (Ochoa et al., 2014; Scheel et al., 2011). Derin et al. (2019) evaluated GPM- and TRMM-based precipitation products across multiple mountainous regions, including the Peruvian Andes. The authors found patterns consistent with previous studies, noting the good performance of the GPM IMERG_F product over annual scales, which reduces at higher temporal resolutions. In addition, the inclusion of orographic rainfall correction methods, as occurs on the Global Satellite Mapping of Precipitation (GSMaP) product, significantly increased quantitative detection of rainfall in these regions.

Given the generally adequate accuracy but low spatial resolution of the GPM product, several methods have been proposed to downscale it to a finer spatial resolution. There is a relationship between the Normalized Difference Vegetation Index (NDVI) and precipitation at different spatial and temporal scales (Davenport & Nicholson, 1993; Immerzeel et al., 2005; Malo & Nicholson, 1990), which forms the basis of the downscaling method (Immerzeel et al., 2009). In this method, an exponential equation is used to relate higher-resolution NDVI data to annual rainfall derived from TRMM. Once the relationship has been established, the equation can be used to generate a high-resolution annual rainfall grid. Furthermore, NDVI and elevation were used with a multi-linear regression approach to down-scale the GPM instead of the TRMM (Jia et al., 2011). Many other studies have proposed downscaling algorithms based on the relationship between NDVI and rainfall as a base and elevation as a secondary variable, applying geographically weighted regression (GWR; Xu et al., 2015; Zhan et al., 2018) which allows the downscaling to be applied to a monthly time step, or a boosting decision tree-based approach, or similar machine learning techniques (Chen et al., 2020; Shen & Yong, 2021). The latter approach yields a robust method, particularly in regions with pronounced elevation gradients. However, the results are only reliable for the calibration period. In general, machine learning techniques provide better results than parametric methods (e.g., exponential regression and GWR). However, when corrected using residuals, parametric methods perform equally well and have the advantage of being grounded in physical relationships that extend beyond the calibration period of the algorithm, whereas machine learning methods do not.

While several satellite-based and gauge-corrected precipitation products exist, most studies find that performance is adequate at low spatial resolutions. Yet, it remains unclear whether simple downscaling methods can increase spatial resolution without degrading temporal performance, thereby improving the assessment of daily rainfall and extreme event representation in data-scarce regions where the low resolution of existing products is insufficient. Previous studies applied the correlation between NDVI and precipitation across various temporal and spatial scales was to derive a relationship for downscaling the GPM product. Despite the extensive development of NDVI-based and terrain-informed downscaling methods, it remains unclear whether simple, physically informed and parsimonious approaches can enhance the spatial resolution of daily satellite-based precipitation products without degrading temporal coherence, precipitation detection skill, and extreme-event representation.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

This uncertainty is particularly acute in complex, data-scarce mountainous regions such as the Peruvian Andes, where both calibration and validation are constrained by sparse and heterogeneous station networks. As a result, the practical value of applying simple downscaling strategies relative to regionally optimized, gauge-corrected products remains insufficiently tested. We addressed this gap by developing two GPM-based, downscaled daily precipitation datasets at 0.01° resolution using different parsimonious downscaling approaches, compared the results with two other gridded datasets developed for Peru (PISCO and Rain4PE), and evaluated all datasets against daily observations and indicators associated with extreme event occurrence. The analysis focused on the Peruvian Andes, where complex topography and sparse station networks pose significant challenges for accurately representing rainfall. Through this analysis, we assessed the effectiveness of parsimonious downscaling algorithms in enhancing rainfall representation in the Andes region. Moreover, we provided a comparative benchmark against existing national ground-based and satellite-derived gridded datasets.

2 Study Area

The study area was defined by selecting a total of 12 national basins, as defined by the National Service of Meteorology and Hydrology of Peru (SENAMHI). These basins are located in the Andean region of Peru (Figure 1), and cover an area of approximately 970 km^2, with elevations ranging from 170 to more than 6000 meters above sea level. The regions were selected based on an assessment of the availability of local data on daily precipitation, taking into account both temporal and spatial coverage.

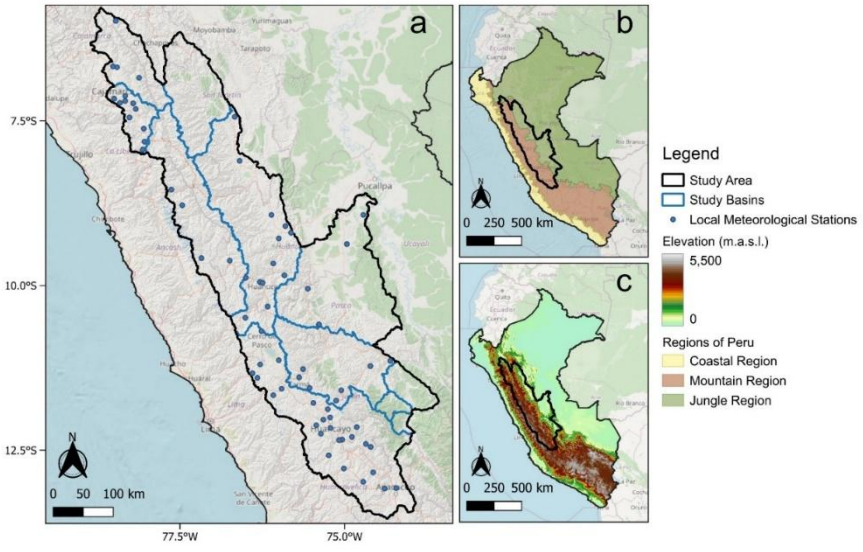

Figure 1: Region of Analysis, national basins and local meteorological stations (a), environmental regions of Peru (b), and topography (c). Basemap from OpenStreetMap contributors, Open Database License (ODbL).

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

The study area falls within two recognized environmental regions of the country, the Mountainous Region and the Jungle. However, the section that falls within the Jungle area represents the transition between the mountainous ecosystem and the Amazon forest. In this sense, the area encompasses a variety of multiple environmental and topographic conditions and settings found in the Peruvian Andes.

3 Materials and methods

3.1 Datasets

3.1.1 Precipitation

Multiple datasets (Table 1) were used in this study for daily precipitation, based on satellite-based precipitation products, meteorological station corrected gridded datasets, and in situ observations. Among them, PISCOp v2.1 is a gridded rainfall product developed by SENAMHI for Peru, providing daily and monthly precipitation fields at 0.1° spatial resolution. It blends satellite-based products with rain-gauge observations and applies bias corrections to improve consistency with ground data, including adjustments to enhance daily estimates (Aybar et al., 2020). Compared with earlier versions, PISCOp v2.1 incorporates more stations and refined corrections, resulting in more reliable rainfall patterns and magnitudes, and is therefore widely used for hydrological and climate applications in the region.

Table 1: Precipitation data utilized in the study: Entirely satellite-based (S) precipitation data, completely/partially corrected datasets using gauge (G) data.

Dataset (version)	Type	Spatial resolution	Spatial coverage	Temporal resolution	Temporal coverage	Source
Gauge-corrected datasets						
PISCOp (v2.1)	S, G	0.1°	Peru	Daily	1981-2016	Aybar et al. (2020)
Rain4PE	S, G	0.1°	Peru-Ecuador	Daily	1981-2015	Palomino et al. (2021)
Non-gauge corrected datasets						
GPM IMERGF (v6)	S	0.1°	Global	Daily	2000-present	
Meteorological Stations						
Local Stations	-	-	Peru	Daily	2000-2018	Autoridad Nacional del Agua (ANA)

Rain4PE is a daily gridded precipitation dataset for Peru and Ecuador (1981–2015) developed using a combination of satellite-based products, reanalysis products, and ground observations merged through machine learning (Fernandez-Palomino et al., 2021). Subsequent hydrological corrections were applied to catchments where rainfall had been underestimated, improving

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

the consistency of the water balance. This approach yields a product that is spatially robust that more accurately represents rainfall dynamics across the Andes and Amazon than commonly used datasets, thereby supporting hydrological and climate applications in the region.

The GPM IMERGF dataset is the successor to TRMM and forms part of NASA's Global Precipitation Measurement mission, which provides high-resolution estimates of global rainfall and snowfall. Building on TRMM's foundation, GPM extends spatial coverage beyond the tropics and improves the detection of light rain, snow, and microphysical precipitation features (Gabella et al., 2017; Sun et al., 2022). IMERG products offer a spatial resolution of 0.1° and temporal coverage ranging from sub-hourly to 3-hourly, with Early, Late, and Final versions supporting near-real-time monitoring and post-processed analyses. This makes IMERG a key global dataset for hydrological, meteorological, and climatological applications.

Finally, ground-based meteorological stations from the meteorological monitoring network in the study area were utilised, operated by the National Service of Meteorology and Hydrology of Peru (SENAMHI). A total of 159 conventional meteorological stations are present in the region, which measure different variables daily, depending on their configuration. However, of these stations, only 70 have reliable data that overlap with the periods covered by the other datasets (PISCO, RAIN4PE, GPM IMERGF). Unfortunately, most of these stations are no longer operational, resulting in a lack of information on the area's climatology for future studies.

3.1.2 Topography and Vegetation

For all analyses, a 12.5 m-resolution Digital Elevation Model (DEM) was used. This DEM is the ALOS PALSAR-RTC product (ASF DAAC, 2014). Additionally, the 16-day average Normalized Difference Vegetation Index (NDVI) product from MODIS TERRA (Didan, 2021) was obtained and processed to yield a yearly mean for the period 2000-2018, calculated using Google Earth Engine (GEE) across Peru.

3.2 Downscaling procedures

Multiple downscaling procedures were applied tothe GPM product. All of these methods used NDVI as the primary explanatory variable to derive precipitation at the resolution of the NDVI product (0.01°). Two different regression methods were used: Exponential Regression (EXP) and Geographically Weighted Regression (GWR). The downscaling procedures were based on the methods developed by Immerzeel et al. (2009) for EXP, and Oshan et al. (2019), Zhan et al. (2018) and Foody (2003) for GWR.

All applied methods assumed that NDVI correlates with precipitation across temporal and spatial scales. A relationship between the two variables can be derived, either exponential or multilinear, by incorporating additional explanatory variables to account for the spatial variability of this relationship (i.e., the grid cell's geographical location).

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

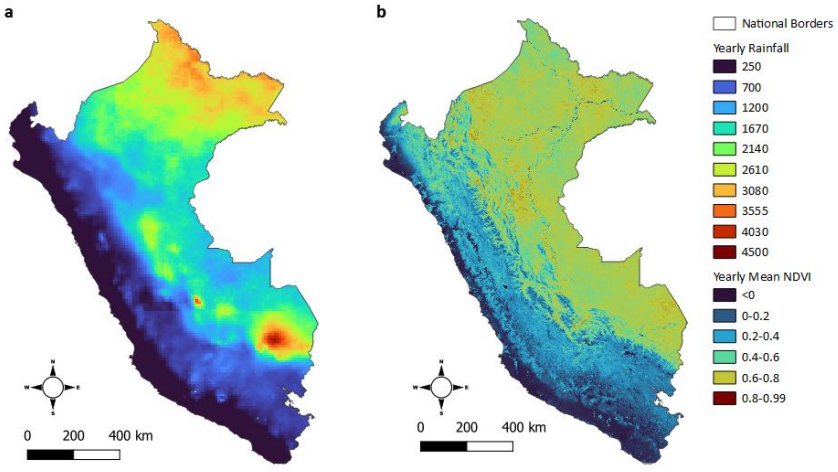

Figure 2: Peru's Yearly Precipitation based on GPM IMERGF v06 product for 2020 (a). Peru's Yearly Mean NDVI for 2020 (b).

3.2.1 Exponential Regression (EXP)

First, an annual downscaling was performed. The annual mean NDVI was compared with the annual rainfall of the GPM for the period 2001-2020 for different resolutions (0.10°, 0.25°, 0.50°, 0.75°, 1.00°, 1.25°) to identify the scale at which the NDVI–precipitation relationship was strongest. The original resolutions of the datasets are 0.01° and 0.10°, respectively; grid-cell averaging was applied for rescaling. For each resolution, an exponential regression of NDVI on GPM precipitation was fitted, and outliers associated with non-precipitation-limited vegetation, i.e., water bodies, urban areas, irrigated agriculture, or deserts, were removed. The regression with the best fit (i.e., the highest R^2) occurred at 0.75° and was used as the basis for downscaling.

At the native low GPM resolution of 0.10° (LR), exponential regression was applied to estimate precipitation from NDVI (Eq. 1).

$$P_e(NDVI_{LR}) = a \cdot e^{b \cdot NDVI_{LR}} \quad (1)$$

where $P_e(NDVI_{LR})$ is the NDVI-based yearly rainfall (mm year^{-1}) for low resolution (LR), a and b are the exponential regression coefficients, and $NDVI_{LR}$ is the yearly mean NDVI for LR.

The following step was the estimation of a residual for LR (Eq. 2). The residual between observed GPM precipitation and NDVI-based estimates was interpreted as the portion of rainfall not explained by vegetation patterns.

$$\Delta GPM_{LR} = P_{GPM} - P_e(NDVI_{LR}) \quad (2)$$

where ΔGPM_{LR} is the LR residual, interpreted as the amount of rainfall that the regression function cannot explain. The residual for high resolution (HR), ΔGPM_{HR}, with a resolution of 0.01°, was obtained by resampling ΔGPM_{LR} with a cubic spline interpolation algorithm. This HR residual was added to the high-resolution NDVI-derived precipitation estimate (Eq. 3) to ensure consistency with the original GPM totals.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

The next step was to estimate the NDVI-based precipitation at high resolution (HR) (Eq. 3).

$$P_e(NDVI_{HR}) = a \cdot e^{b \cdot NDVI_{HR}} \quad (3)$$

The final downscaled yearly rainfall was $P_e(NDVI_{HR})$ corrected by the HR residual (Eq. 4).

$$Py_{ds} = P_e(NDVI_{HR}) + \Delta GPM_{HR} \quad (4)$$

A parsimonious approach was used to temporally downscale annual to daily precipitation, employing a scaling coefficient to derive daily precipitation from annual precipitation.

First, the coefficient was found using the LR GPM data (Eq. 5).

$$CdLR_{i,j} = \frac{PdLR_{i,j}}{PyLR_j} \quad (5)$$

Where $CdLR_{i,j}$ is the daily coefficient for the day i of the year j, $PdLR_{i,j}$ is the GPM's LR daily precipitation for the day i of the year j, and $PyLR_j$ is the GPM's LR yearly precipitation for the year j. This coefficient was derived from LR data at a resolution of 0.10° and subsequently interpolated to 0.01° to incorporate spatial variability at higher resolutions. The HR daily coefficient $CdHR_{i,j}$ was derived by resampling the LR coefficient with a cubic spline interpolation algorithm. Finally, the daily downscaled rainfall was derived using Eq. 6.

$$Pd_{ds_{i,j}} = Py_{ds_j} \cdot CdHR_{i,j} \quad (6)$$

Where $Pd_{er_{i,j}}$ is the downscaled, daily precipitation for the day i of the year j and Py_{ds_j} is the downscaled yearly precipitation for the year j. The resulting dataset is referred to as GPM-EXP.

3.2.2 Geographically Weighted Regression (GWR)

The GWR approach extends the EXP method by explicitly accounting for spatial heterogeneity in the precipitation–NDVI relationship. The GWR downscaling method first resampled annual-averaged coarse-resolution NDVI and Digital Elevation Model (DEM) data from their native high-resolution resolutions (0.01° and 30 m, respectively) to the geographical grid of the original low-resolution GPM dataset at 1.00°.

An optimization routine was implemented (Eq. 7), utilizing the DEM, NDVI, and geographical location as explanatory variables for GPM precipitation at LR:

$$P_{GPM} = \varphi_{NDVI} \cdot NDVI_{LR} \cdot \delta + \varphi_{DEM} \cdot DEM_{LR} \cdot \delta + \varphi O \cdot \delta + \varepsilon_{LR} \quad (7)$$

Where P_{GPM} is the observed yearly precipitation, O is the geographical location variable, ε_{LR} is a residual at LR, with φ_{NDVI}, φ_{DEM}, δ as coefficients for optimization, which vary spatially and per year.

The LR residual was used to derive a HR residual (ε_{HR}) by applying a spline interpolation from 1.0° to 0.01°. Finally, the yearly HR precipitation was calculated based on the previous steps (Eq. 8), for each year of the studied period.

$$Py_{ds} = \varphi_{NDVI} \cdot NDVI_{HR} \cdot \delta + \varphi_{DEM} \cdot DEM_{HR} \cdot \delta + \varphi O \cdot \delta + \varepsilon_{HR} \quad (8)$$

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

The final step was to downscale the annual HR to daily precipitation. As with the EXP method, daily precipitation was derived by applying interpolated daily scaling coefficients from the GPM dataset to the downscaled annual precipitation (Equations 5 and 6), yielding a daily HR GPM-GWR precipitation dataset at 0.01° resolution.

3.3 Evaluation Metrics for Gridded Datasets

The downscaled datasets and reference products were evaluated against meteorological station observations at a daily time step using various performance metrics (Table 2). The period of evaluation matches the availability of local stations (2000 – 2018) as described on Table 1. A 'point-to-pixel' comparison was conducted, whereby the gridded dataset was sampling at the meteorological station locations. The Percent Bias (PBIAS), Root Mean Square Error (RMSE), and the Pearson Coefficient of Correlation (r) were used to assess the statistical performance of the datasets in representing observations. In addition, precipitation detection skill was evaluated using commonly applied forecast performance indices : the Critical Success Index (CSI), Frequency Bias Index (FBI), False Alarm Ratio (FAR), and Probability of Detection (POD) (Su et al., 2008; Tang et al., 2016; Xu et al., 2017; Yu et al., 2022).

Table 2: Summary of Evaluation Metrics for Gridded Datasets

Type	Indicator	Range	Ideal value
Statistical Performance Index	Percent Bias (PBIAS)	-α to +α	0
	Root-Mean-Square-Error (RMSE),	0 to α	0
	Coefficient of Correlation (r)	0 to 1	1
Forecasting Index	Critical Success Index (CSI)	0 to 1	1
	Frequency Bias Index (FBI)	-α to +α	1
	False Alarm Ratio (FAR)	0 to 1	0
	Probability of Detection (POD)	0 to 1	1

Performance indices typically use the concept of *hits, misses,* or *false alarms*. A hit represents a correct prediction of the occurrence of precipitation, while a miss is a non-prediction, and a false alarm is an incorrect prediction of the occurrence. In that sense, the CSI, sometimes referred to as Threat Score, represents the ratio of hits to misses and false alarms; FAR measures the proportion of false alarms to hits; and POD represents the proportion of hits to false alarms (Wilks, 2006). Moreover, the FBI measures the tendency of the dataset to under- or overforecast (Bougeault, 2003), defined as the ratio of the sum of forecasts (i.e., hits and false alarms) to occurrences (i.e., hits and misses).

An important aspect in calculating the CSI, FBI, FAR, and POD is the threshold for daily precipitation, which determines whether the measurements detect a real rainfall event. This is typically done to filter out noise in the GPM timeseries caused by the satellite's sensor. Any day in the time series with precipitation exceeding this threshold was classified as a rainfall event;

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

days with precipitation below this threshold were classified as without rainfall. Rainfall occurrence was defined using a daily threshold of 2.5 mm day-1, based on the World Meteorological Organization's classification categories for rainfall (World Meteorological Organization, 2018), which define an event with an intensity of less than 2.5 mm hour^{-1} as light rain. Because the comparison was based on cumulative daily values, a cumulative daily value less than 2.5 mm is likely to correspond to an event or events with an intensity of 2.5 mm hr^{-1} or less.

3.4 Indicators for Extreme Values

The capabilities of the downscaled datasets and the PISCO and Rain4PE products for detecting occurrences of extreme events were evaluated for the same period using six precipitation indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI; Peterson, 2005) (Table 3). The indices both frequency-based and amount-based extremes, including consecutive dry and wet days, heavy and very heavy precipitation days, and high-percentile rainfall totals. Indices were calculated for both the gridded datasets and the meteorological station observations. The gridded dataset-derived indicators were validated against the observed data-derived indicators using the coefficient of correlation (r). This approach emphasizes the temporal consistency of extreme-event representation rather than absolute magnitude alone.

Table 3: Summary of ETCCDI indicators to evaluate extreme events

Type	Name	Definition
Frequency	Consecutive Dry Days (CDD)	Number of consecutive days with precipitation < 1mm in a year.
	Consecutive Wet Days (CWD)	Number of consecutive days with precipitation > 1mm in a year.
	Number of heavy precipitation days (R10)	Number of days with precipitation > 10mm in a year.
	Number of very heavy precipitation days (R20)	Number of days with precipitation > 20mm in a year.
Amount	Very wet days (R95p)	Annual total precipitation when daily precipitation > 95th percentile.
	Extremely wet days (R99p)	Annual total precipitation when daily precipitation > 99th percentile.

4 Results

4.1 General performance of precipitation datasets

The distribution of daily precipitation values revealed systematic, dataset-specific differences relative to ground observations (Figure 3). Among all products, Rain4PE and PISCO showed the closest agreement with the observed distributions,

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

successfully reproducing both the central tendency and the overall spread of daily rainfall. Rain4PE and PISCO show a generally higher median as well as a more extended interquartile range. This suggests that these datasets tend to estimate small amounts of rainfall in situations where gauge stations register no rainfall, and they tend to represent more moderate-to-high rainfall occurrences, leading to a higher third-quartile value of approximately 4-5 mm per day compared with observations, which remain at 2.5 mm. The GPM and GPM-GWR products show a similar distribution, with the median closer to observations (close to zero) and a similar interquartile range. However, they still exhibit a slightly different distribution, with more data points concentrated near 2 mm per day, similar to Rain4PE and PISCO. Finally, the GPM-EXP dataset has a median and quartiles near the observations; the distribution differs from those of all datasets, with a generally wider distribution for moderate-to-low precipitation values, which may indicate an even stronger tendency to overforecast low precipitation values in no-rainfall days.

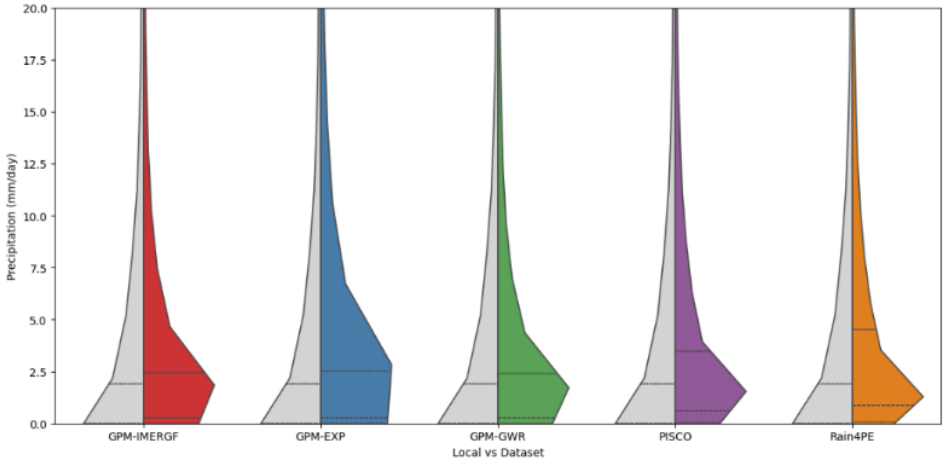

Figure 3: Distribution of daily precipitation values across all datasets (right) compared to local observations (left). The dotted line represents the median, and the dashed lines represent the first and third quartiles.

The statistical performance metrics further emphasized these contrasts (Figure 4). Rain4PE achieved the highest correlation coefficients, indicating strong temporal agreement with observations, while PISCO exhibited the lowest correlations despite its good representation of rainfall magnitudes, suggesting weaker day-to-day variability. The downscaled GPM products (EXP and GWR) showed intermediate correlation values, with GPM-GWR outperforming GPM-EXP. Root-mean-square error (RMSE) was lowest for Rain4PE and PISCO, confirming their superior accuracy in reproducing daily rainfall magnitudes, whereas the original GPM products exhibited the largest deviations. Percent bias (PBIAS) revealed that Rain4PE and PISCO maintained near-zero median bias, while GPM-IMERGF strongly underestimated rainfall and GPM-EXP tended to overestimate it. Kling–Gupta efficiency (KGE) was highest for Rain4PE, although it showed substantial variability across stations. GPM-GWR showed moderate efficiency, PISCO performed intermediate, and both GPM-EXP and GPM-IMERGF exhibited low and unstable KGE values.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

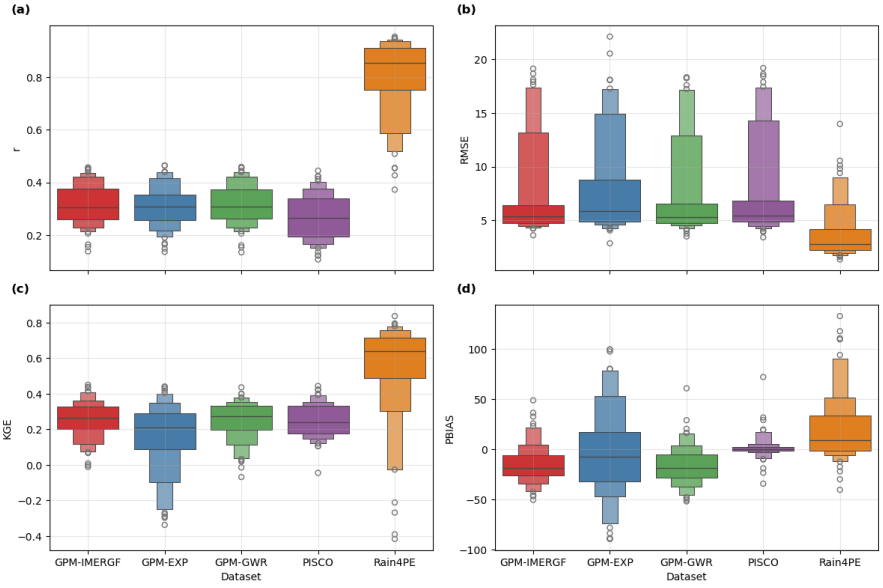

Figure 4: Distribution of performance daily statistics across all datasets, considering r (a), RMSE (b), KGE (c) and PBIAS (d). For each box plot, the central line represents the median (50th percentile). Successive nested boxes denote progressively wider quantile ranges, starting from the interquartile range (25th–75th percentiles) and extending outward by repeatedly halving the remaining data (e.g., 12.5th–87.5th, 6.25th–93.75th percentiles). Box widths reflect the density within each quantile band.

Spatial patterns of performance further emphasize these contrasts (Figure 5). Rain4PE showed the strongest overall spatial performance, with consistently high correlation and very low bias across most stations, indicating strong agreement with both daily variability and rainfall totals. GPM-GWR performed reasonably well, with greater spatial variability, showing good results in several areas while exhibiting moderate bias and localized higher errors. PISCO performed well with respect to rainfall magnitude and bias, maintaining minor deviations across most stations. However, its spatial correlation was weak across much of the region, particularly in the southern and eastern transitional zones, i.e. the Andes-and -Amazon basin.

In contrast, both GPM-EXP and GPM-IMERGF exhibited widespread weak spatial agreement, with many stations failing to meet acceptable correlation thresholds, particularly in the northern and central sectors. Their bias patterns were also large and spatially inconsistent, with frequent extreme values, particularly in the eastern and southern areas. Overall, Rain4PE clearly outperforms the others; GPM-GWR is intermediate; PISCO is reliable in magnitude but weak in temporal agreement; and GPM-EXP and GPM-IMERGF perform the worst.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

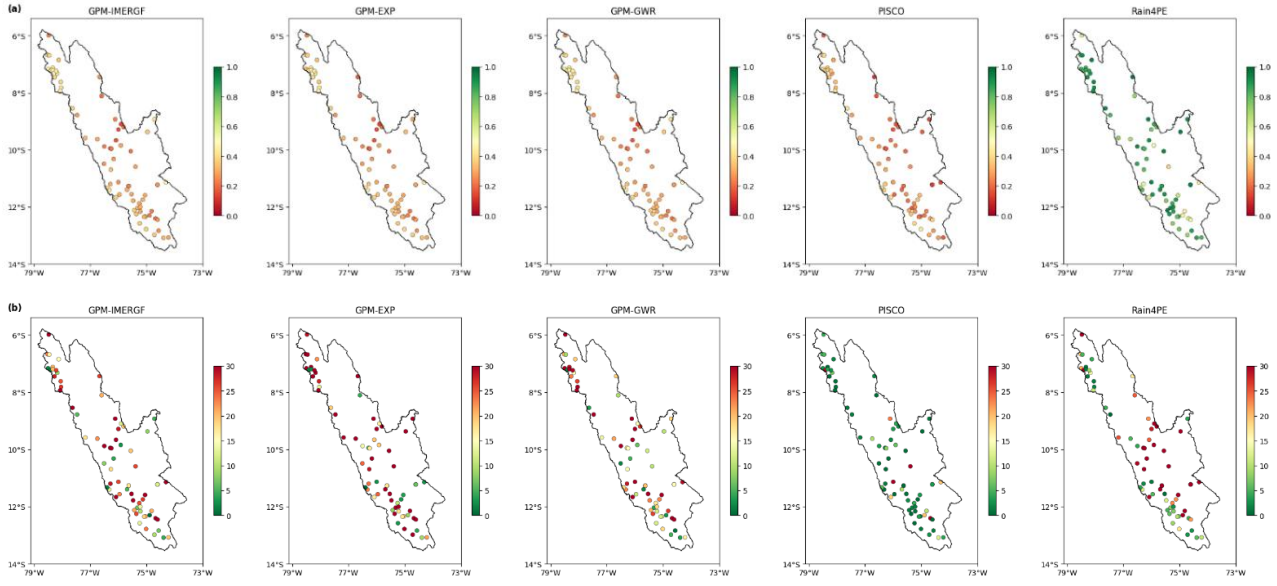

Figure 5: Spatial distribution of performance statistic r (a) and |PBIAS| (b) across the study area.

4.2 Forecasting indices and detection trade-offs across datasets

The trade-off between detection skill and false alarms, shown by CSI and FAR in Figure 6, highlights the apparent differences between the datasets. Rain4PE achieved the highest median CSI with minimal variability, indicating strong and consistent detection performance across stations, whereas PISCO recorded the lowest CSI, reflecting weaker overall detection skill. The GPM products (GPM-GWR, GPM-EXP, and GPM-IMERGF) occupied intermediate positions with overlapping distributions and generally lower medians. Consistent with the CSI results, the FAR distributions showed that Rain4PE had the lowest false-alarm rate. The three GPM products exhibited intermediate behavior with broader ranges, while PISCO had the highest FAR, indicating a greater tendency towards over-detection.

The FBI and POD distributions provided additional insight into detection characteristics (Figure 6). Rain4PE exhibited the highest median POD, confirming its superior ability to detect precipitation events. However, it also showed elevated FBI, indicating a tendency to overpredict event frequency. The remaining datasets cluster at lower FBI and POD values, particularly PISCO and GPM-GWR, which concentrate near a FBI value of unity, suggesting more conservative detection with reduced sensitivity. Overall, Rain4PE combined strong detection performance with lower false-alarm rates, albeit with a higher frequency bias, whereas the other products favored conservative behavior at the expense of sensitivity.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

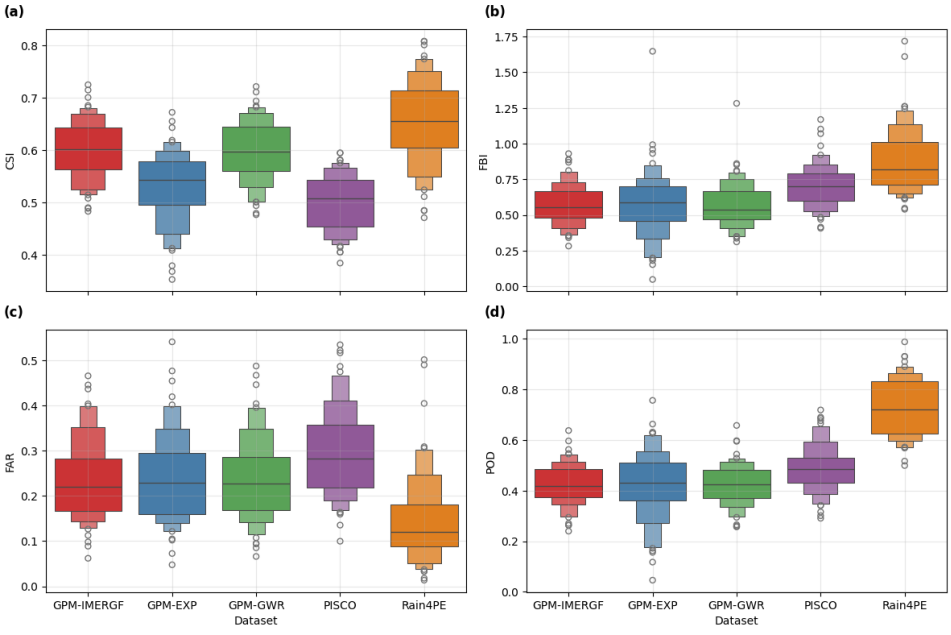

Figure 6: Distribution of forecasting indices across all datasets, considering CSI (a), FBI (b), FAR (c), and POD (d). For each box plot, the central line represents the median (50th percentile). Successive nested boxes denote progressively wider quantile ranges, starting from the interquartile range (25th–75th percentiles) and extending outward by repeatedly halving the remaining data (e.g., 12.5th–87.5th, 6.25th–93.75th percentiles). Box widths reflect the density within each quantile band.

The spatial evaluation highlighted clear contrasts in performance across the study area (Figure 7), with Rain4PE and PISCO emerging as the most reliable datasets. Both datasets showed consistently high CSI values across much of the region and balanced FBI close to unity, indicating strong detection capability with limited systematic bias. The downscaled products introduced finer spatial detail, albeit with more heterogeneous behavior. GPM-GWR showed localized improvement in CSI, particularly in regions with strong elevation gradients, while also moderating FBI towards unity. This reflected the advantage of incorporating terrain information. In contrast, GPM-EXP presented highly uneven CSI patterns, featuring isolated areas of good performance surrounded by broad regions of low skill, and widespread FBI values well above unity, indicating persistent overestimation. Despite the increased spatial detail, the downscaled products did not consistently outperform Rain4PE or PISCO in rainfall detection at the station scale. While acknowledging that large areas without observational coverage could not be evaluated, Rain4PE and PISCO proved to be the most spatially robust datasets for daily precipitation within the limits of this station-based evaluation. .

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

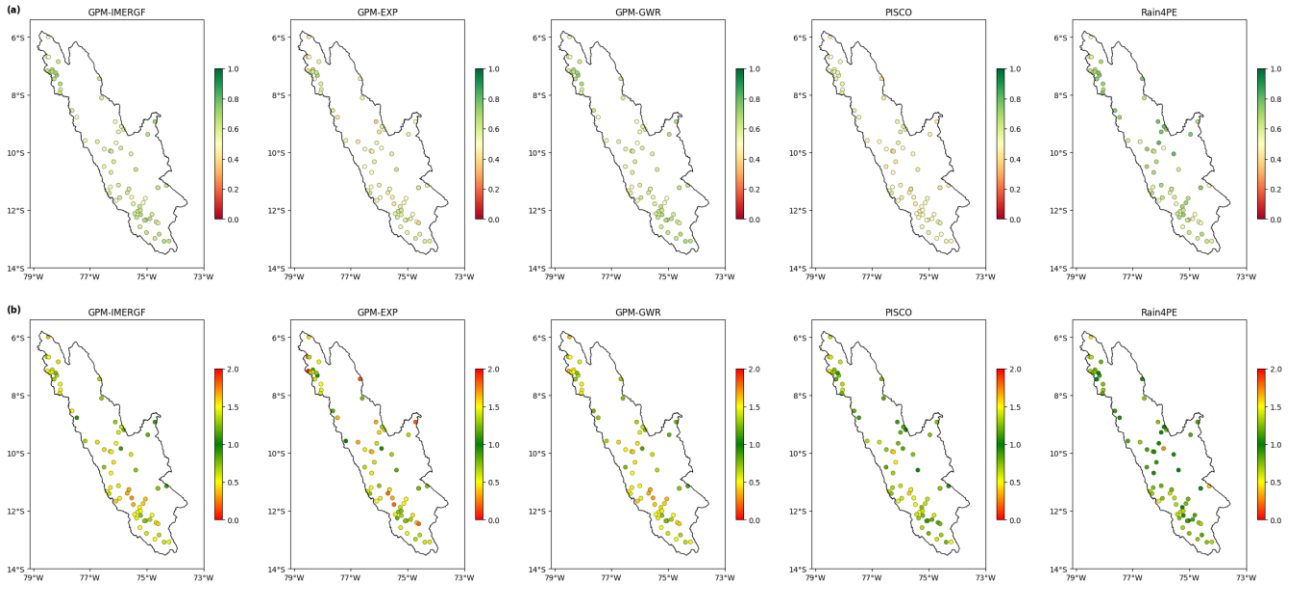

Figure 7: Spatial distribution of CSI (a) and FBI (b) across the study area.

4.3 Influence of topography on reliability

Performance metrics aggregated by elevation bands with a similar number of stations revealed a clear dependence on altitude (Figure 8). Rain4PE consistently showed the strongest overall agreement with daily observations, combining the lowest RMSE values, the highest correlations (often greater than 0.90), and the most balanced KGE. For this dataset, there is a clear and consistent reduction in PBIAS and, consequently, an increase in KGE with increasing elevation. However, the dataset exhibits the highest biases for stations below 2300m in elevation, with substantial underestimates in some cases. PISCO exhibited a moderate RMSE at stations below 2300 m and a very low RMSE at higher elevations. PISCO exhibited small biases and consistently positive KGE values, with a slight positive trend with increasing elevation. Nevertheless, it achieved the lowest overall correlation, with exceptionally low values at stations below 2300 m. GPM-IMERGF and GPM-GWR exhibited moderate overestimation and weak temporal agreement, as reflected in low correlations (0.2 to 0.4) and consistently negative PBIAS values. GPM-GWR generally improved average performance compared with GPM-IMERGF, possibly due to its elevation-aware formulation. However, this pattern was not evident across all elevations, nor was there a substantial difference in the performance index. GPM-EXP exhibited the most unstable behavior, with large variability in bias and performance across elevations. On average, its correlations were similar to those of the other two GPM datasets. Although it also tends to overestimate, like the other two datasets, for stations in the 3100-3700 elevation range, it shifted towards underestimation, showing considerable biases in some cases. In general, lowland sites performed poorly across most datasets, whereas intermediate- to high-elevation sites showed better and more stable performance.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

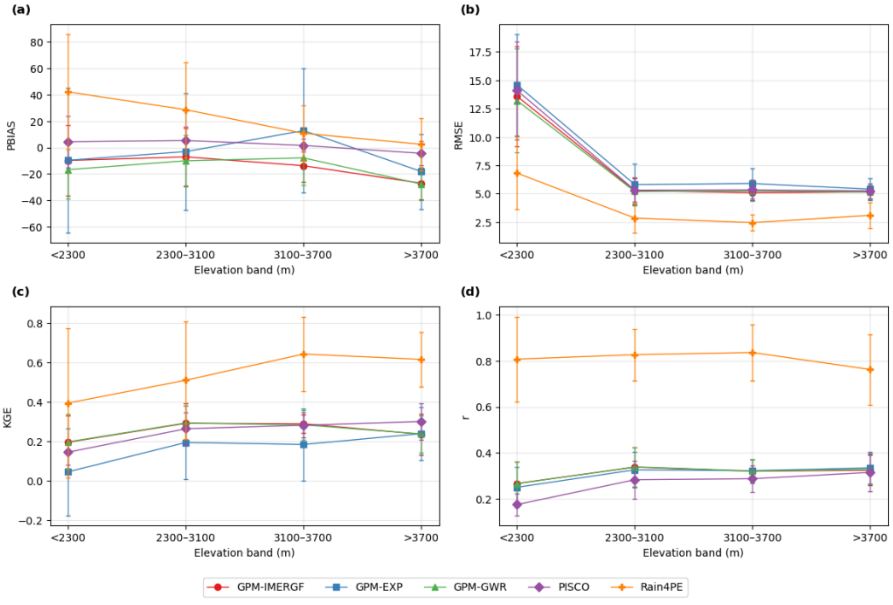

Figure 8 Relationships between elevation and daily precipitation performance metrics at stations: PBIAS (a), RMSE (b), KGE (c), and r (d).

Altitude modulated forecast skill when comparing the downscaled products with the GPM-IMERGF dataset (Figure 9).Driven solely by NDVI, GPM-EXP exhibited the highest sensitivity but it systematically overestimated, with FBI frequently exceeding 2.0 at mid-elevations. In contrast, its lower FAR largely reflected inflated event counts rather than improved reliability. Raw GPM remained closer to unity FBI with stable POD, but had a higher FAR, representing a more conservative baseline. GPM-GWR provided a better balance, moderating bias and improving detection in complex terrain, although the FAR remained variable. Trade-offs were most pronounced between 2000 and 3500 m, whereas lowland and high-mountain sites exhibit more stable behavior. Overall, NDVI-based downscaling enhanced detection in vegetation-dominated regions. However, incorporating elevation, as in GPM-GWR, resulted in more robust performance across the heterogeneous Andean landscape.

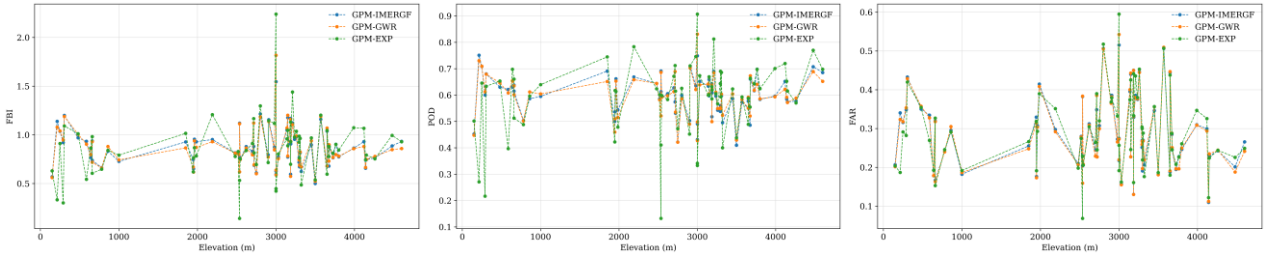

Figure 9: Altitude-sorted forecasting indices for exponential and geographically weighted regressions and the raw GPM dataset.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

4.4 Representation of extremes

There are important differences between datasets in their ability to reproduce the derived ETCCDI indices (Figure 10). The GPM-EXP product overestimated consecutive dry days (CDD) relative to the other datasets, which are very similar to local data (35 days), while Rain4PE shows a slight underestimation. Other datasets showed relatively stable but muted variability, failing to capture the full range observed at stations. GPM-EXP presented a high standard deviation of 21 days, while the rest a low deviation of 12, in contrast to the locally derived CDD, which is considerably higher than all datasets except GPM-EXP, indicating that GPM-IMERGF, GPM-GWR, Rain4PE and PISCO are stable across the region for this indicator, not necessarily capturing the variability shown by local data. In relation to consecutive wet days (CWD), all GPM datasets approximated local data, whereas Rain4PE and PISCO, on average, overestimated this indicator across the region and exhibited a high standard deviation. Rain4PE predicted more high-magnitude rainfall, with higher R10 and R20 values than those in the local data, whereas GPM products captured fewer extreme rainfall events. PISCO is the best dataset in this respect, on average, across the region. For amount-type extreme indices, on average, all products remained within an adequate order of magnitude relative to local data, with GPM-EXP showing a tendency towards overestimation that is more pronounced than the others.

Figure 10: Summary of ETCCDI indices aggregated across the region: mean value a), standard deviation b), and coefficient of correlation (r) of datasets against indices derived from local data c).

Overall, PISCO shows a better temporal representation of ETCCDI indices using local data as the reference, with an r value larger than 0.5 (Figure 10) and performs particularly well in detecting rainfall events of high magnitudes. Rain4PE follows; however, it does not perform as well for CWD and R20, showing both a tendency of overestimating rainfall events and, at times, underestimating high magnitudes or their time of occurrence. All GPM products performed poorly, with r values below 0.3 across almost all indices, particularly for detecting high-intensity and high-magnitude rainfall (i.e., R20 and R99p). GPM-GWR performs slightly better than GPM-IMERGF, with CDD, CWD, and R10 an r larger than 0.2 showed a weak but present correlation, while GPM-EXP is the worst of the three, with r higher than 0.2 only for CWD.

Station-level time series from two stations illustrate these patterns (Figure 11). At Tananta station (Figure 11, Panel a) had one of the highest R20 values, while Sondor Matara (Figure 11, Panel b) had one of the highest CDD values.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

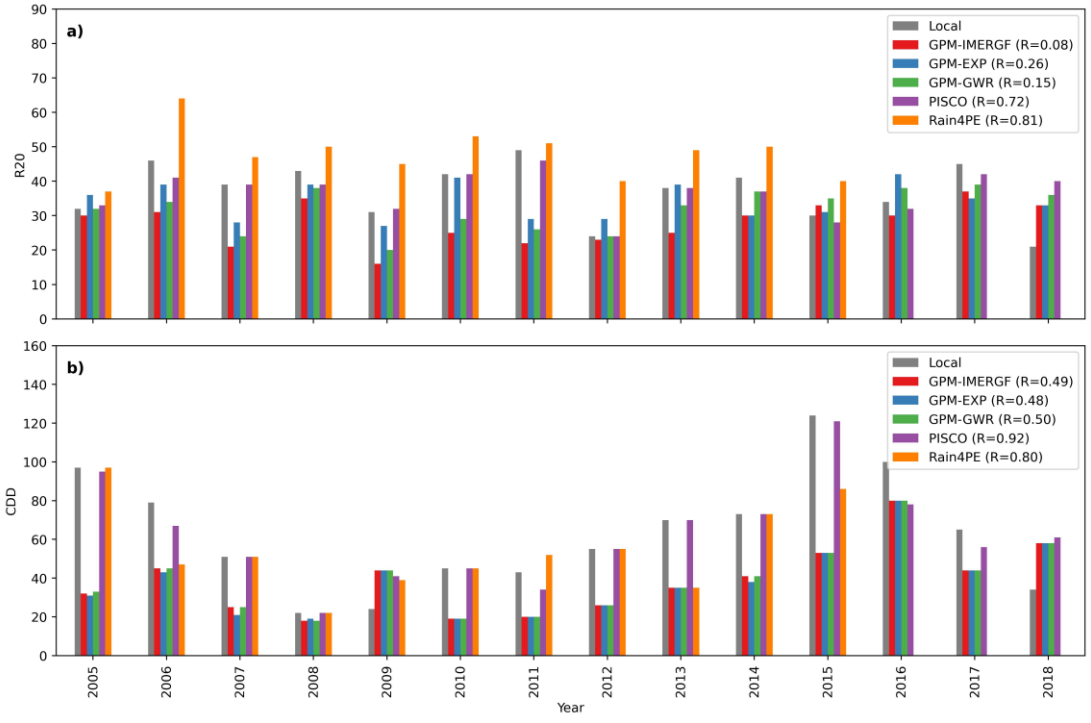

Figure 11: Yearly time series of R20 for station Tananta (a) and CDD for station Sondor Matara (b).

PISCO clearly captured temporal dynamics well for CDD at Sondor Matara, across the selected stations, and compared well with Rain4PE for R20 at Tananta. Despite adequately capturing most years, Rain4PE tended to overestimate R20 and underestimate CDD in some years, which negatively impacted its performance, as was observed in an aggregated manner for the region (Figure 10). Conversely, in Tananta, GPM-EXP more closely approximated the locally based R20 for most years than other GPM datasets, albeit with some variability. GPM-GWR and GPM-IMERGF systematically underestimated R20 in almost all years, particularly during 2006-2014. All GPM datasets systematically underestimated CDD at Sondor Matara, with no apparent difference between them. This shows that the aggregated assessment (Figure 10) might not present the full picture in absolute terms, consistent with the region's overall poor performance on this indicator.

The elevation and slope of the terrain did not show a statistically significant relationship with the performance of datasets representing extreme value indicators. However, if the elevation and slope were aggregated in ranges or bands, dividing the number of stations equally between bands(Figure 12), PISCO showed an increasing trend in performance with elevation up to 3700 m. For stations above this elevation, the mean coefficient of correlation declined, although it remained above 0.5. A similar pattern was observed for the slope, with performance declining for slopes above 16%. Rain4PE did not exhibit a clear relationship between performance and elevation; however, there was a consistent decrease as the slope increased, with an average R-value below 0.5 for slopes above 16%.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

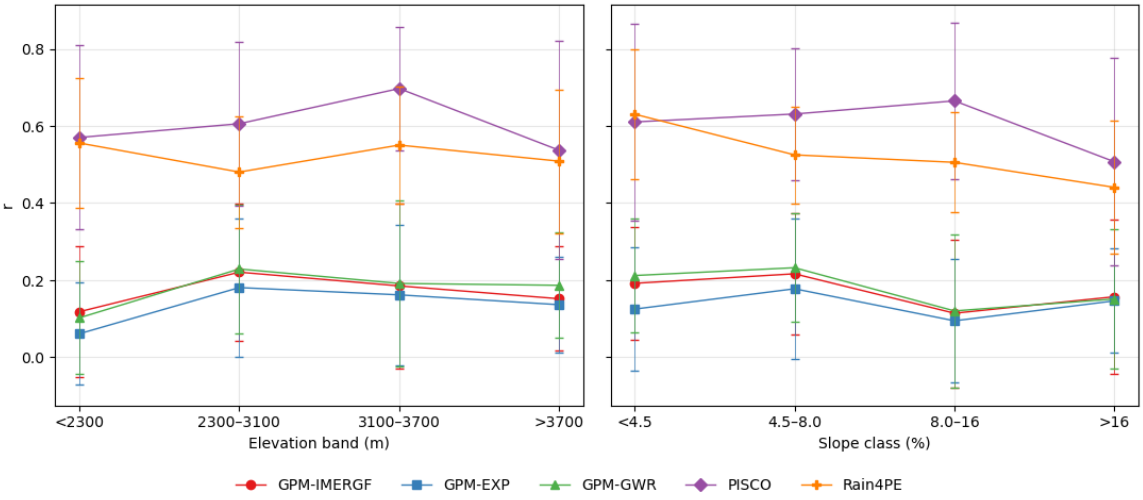

Figure 12: Mean coefficient of correlation (r) ± standard deviation among all ETCCDI indicators, per dataset, against elevation and slope ranges.

The GPM products exhibited similar patterns of performance with respect to elevation and slope, but differed somewhat from Rain4PE and PISCO. For elevations below 2300 m, performance is considerably worse across all GPM datasets. In contrast, GPM-GWR remains stable at higher elevations, whereas the other two tend to decrease uniformly. Conversely, performance decreased significantly for stations located on slopes between intermediate and high (8-16%) relative to other bands, whereas performance was highest on slopes below 8%. Nonetheless, as previously assessed, the correlation coefficient for the GPM datasets was relatively low. This analysis indicates that, for the GPM-GWR and GPM-IMERGF, only stations in an elevation range of 2300 and 3100 m, and a slope band below 8%, presented an r value higher than 0.2, which represented at least a weak correlation compared to local data. In contrast, PISCO showed a strong correlation (R>0.6) and Rain4PE at least a moderate correlation (R>0.4) in all elevations and slopes.

5 Discussion

Rain4PE delivered the strongest overall performance consistently across distributional, error-based, and detection metrics: high correlations, low RMSE, balanced KGE/PBIAS, and strong CSI/POD behavior. PISCO followed closely behind, and showed a similar quantitative performance to that of GPM and the downscaled products. However, it achieved the lowest PBIAS and performed best in terms of extreme-index representation. As expected, the GPM-IMERGF product performed substantially worse than PISCO and Rain4PE, and GPM-EXP performed at least as poorly as, and often worse than, GPM-IMERGF in terms of forecasting skill. A positive aspect was that, unlike GPM-EXP, the GPM-GWR product generally preserved or improved the representation of rainfall patterns achieved by GPM IMERG, at a much higher spatial resolution.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

5.1 Quantitative performance and spatial realism

The results confirmed that the GPM IMERG (IMERG-F) exhibits systematic weaknesses in the Andean region, particularly at the daily timescale (Derin et al., 2019). There is also an underestimation of intensities, lower temporal correlation, and reduced detection sensitivity in orographically complex environments (Bulovic et al., 2020; Tang et al., 2016). This pattern is not unique to the Andes, as comparable elevation-dependent reductions in IMERGF skill have been observed in other mountain systems such as the Himalayas, the Qilian Mountains, and the Colombian Andes and other SPPs (Kumar et al., 2022; Ochoa et al., 2014; Scheel et al., 2011; Wang et al., 2019). The performance of GPM-IMERGF and downscaled products varies with the climatic context, being better in Amazonian rainforest climates and weaker on Pacific-facing slopes, underscoring the importance of regional evaluation rather than relying on global performance summaries (Chen et al., 2022). This highlights the value of incorporating local rain-gauge data into the development of regionally optimised datasets such as Rain4PE and PISCO (Aybar et al., 2020), and of implementing machine learning algorithms alongside other explanatory variables and data sources (Fernandez-Palomino et al., 2021). These methodologies naturally enable better capture of rainfall from local or complex processes that SPPs alone often fail to capture.

Downscaling improves spatial detail while generally preserving the performance of the source dataset. GWR downscaling (GPM-GWR) produced the most useful compromise: improved spatial coherence, a closer match to orographic rainfall cores (R10/R20), and higher CSI. However, GWR tends to be more conservative (lower POD than PISCO), i.e., it generally misses some small events but produces fewer false alarms (Wang et al., 2022). GPM-EXP increases sensitivity (high POD) but at the cost of excessive false alarms and systematic overestimation. This behavior illustrates how downscaling methods that rely on a single explanatory variable can propagate vegetation-related signals into precipitation fields without adequately constraining physical rainfall processes. (Immerzeel et al., 2009; Jia et al., 2011). These results align with broader reviews showing that downscaling generally improves representation but remains method and site-dependent (Kofidou et al., 2023). While GWR preserves the temporal structure of the source dataset, it does not introduce artificial high-frequency variability, unlike the EXP approach, which inflates low-intensity event counts.

Elevation up to a certain level modulates rainfall representation. There is a much higher density of stations above an elevation of 2300 m, and these areas generally show the best match across products, while low-density areas at very high elevations or below 2300 m display larger uncertainty (Bulovic et al., 2020; Chen et al., 2022), which may explain the observed positive trend in average performance with increasing elevation, rather than the nature of the datasets. The GPM-GWR dataset, which is terrain-sensitive by nature, shows differences to its counterpart, GPM-EXP, showing a reduced and less variable PBIAS, by explicitly letting regression coefficients vary spatially and be adjusted by elevation (Foody, 2003; Wang et al., 2022). Moreover, on the Amazon-facing slopes of the study area, land–atmosphere interactions and evapotranspiration (ET) recycling from the Amazon basin are plausible modulators of precipitation occurrence and persistence (Claros et al., 2025; Dominguez et al., 2022). Because these processes may not be fully captured in IMERG retrievals and are not accounted for by our simple

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

downscaling schemes, they likely contribute to important biases that persist even after downscaling. These biases might be better captured by Rain4PE and PISCO, which incorporate local data.

5.2 Implications for extreme event analysis

Substantial differences exist among datasets in their ability to reproduce ETCCDI indices relative to local observations. PISCO consistently exhibits the strongest temporal agreement and performs particularly well for high-magnitude rainfall indicators (R20, R95p, R90p). Rain4PE ranks second, capturing temporal variability reasonably well, although it shows some disagreement with observations, particularly overestimating rainfall frequency and magnitude for R10 and R20, and performing more poorly for CWD. In contrast, all GPM-based datasets perform weakly, with correlation coefficients generally below 0.3, particularly for high-intensity and amount-based extremes. Among them, GPM-GWR moderately improves upon GPM-IMERGF, suggesting benefits from the GWR downscaling method, although correlations remain generally low.

Time-series analysis at the station level confirms these regional findings: PISCO reproduces the temporal dynamics of both CDD and R20 most consistently; Rain4PE shows mixed performance depending on the indicator; and GPM products systematically underestimate extremes, particularly dry-spell duration (CDD). Stratification by elevation and slope reveals that PISCO and Rain4PE retain moderate-to-strong performance across most terrain classes. In contrast, the GPM datasets exhibit their weakest performance at low elevations and on steeper slopes, with only limited improvement at mid-elevations and in gentle terrain. However, low station density within specific elevation ranges increases the uncertainty of these dynamics.

A combination of data sources, methodological approaches per dataset, and system complexity in the Andes can explain the observed performance contrasts. PISCO's superior performance in this aspect is expected, as it explicitly incorporates filtered and gap-filled in situ data (Aybar et al., 2020). Therefore, coherence with the observed data used in this study is expected. Rain4PE, while also gauge-adjusted, relies on blending strategies and a machine-learning algorithm that can produce smooth precipitation fields and thus overestimate rainfall persistence, as evidenced by the high CWD and R20 values. In contrast, GPM-based datasets struggle in this component, consistent with their previously observed weaker quantitative performance, due to the limited capacity of IMERG retrievals to represent complex local processes in the Andes (Derin et al., 2019). Moreover, the study reveals that GPM datasets also struggle to represent low-intensity, persistent events or dry spells temporally. The elevation-dependent behavior observed here is consistent with previous studies showing that satellite precipitation products perform best at mid-elevations with moderate slopes, where precipitation systems have a stronger vertical development and are less scattered, while skill degrades in lowland, foothill transition zones influenced by Amazonian moisture recycling and land–atmosphere coupling (Dominguez et al., 2022; Espinoza et al., 2015).

5.3 Practical implications for hydrological applications

From an operational perspective, the choice of precipitation dataset should be guided by application scale and objective. Our findings indicate that for event-based analyses and hydrological modelling, where accurate totals and event timing are critical, Rain4PE is the preferred and most reliable baseline. In contrast, PISCO is preferable for long-term climatological consistency

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

or drought persistence metrics are the focus, as it offers superior performance in capturing extremes and long-term variability. When higher spatial resolution is required, particularly in data-scarce regions, our study shows that the GWR downscaling method, using NDVI and elevation as main explanatory variables, can refine the product spatially without degrading the performance of the original dataset. The 0.01° GPM-GWR better resolves orographic precipitation structures and exhibits improved detection skill relative to raw GPM-IMERGF, although it still underestimates magnitudes in some areas. Even high-performing products such as Rain4PE may require additional processing in hydrological modelling assessments (Fernandez-Palomino et al. (2021). Additional bias correction or reverse modelling approaches may be required to achieve satisfactory performance in representing hydrological variables such as streamflow. Moreover, particularly in this region, fog capture in the paramo ecosystem can constitute a significant component of the water balance (Cárdenas et al., 2017). However, these fluxes are not measured by rain gauges, not captured by satellite observations and therefore not accounted for in gridded datasets. This phenomenon should be considered in water-resources assessments and hydrological modelling applications. These findings suggest that parsimonious, physically informed downscaling can add value in regions where gauge-based products are unavailable or outdated, but it should not be viewed as a substitute for high-quality, regionally calibrated datasets.

5.4 Uncertainty in validation and station representativeness

A persistent constraint in mountainous validation is not only the sparse distribution of rain gauges, but also their limited representativeness of local precipitation variability. This dual challenge is evident in altitude performance analyses and spatial error-metric distributions, where regions with many closely spaced stations sometimes still exhibit high RMSE, wide scatter in correlation (r), and inconsistent bias patterns across products. Importantly, as Bulovic et al. (2020) emphasize, the presence of multiple rain gauges within a satellite pixel or grid cell should not be interpreted as a definitive “ground truth.” In the tropical Andes, they found that even when several gauges fall within a single satellite footprint, native GPM retrievals can deviate markedly from the gauge aggregate due to both retrieval limitations and station measurement uncertainties such as wind effects, exposure, wet-bias, or under-catch. This means that point-to-pixel validation can overstate or misrepresent satellite error if gauge quality and representativeness are not critically assessed.

In our study, this manifests as clusters of stations at mid-elevations producing performance statistics that appear stronger on average. However, the behaviour of individual stations remains highly variable, particularly for GPM-based datasets. This variability underscores the importance of explicitly acknowledging validation uncertainty, both in areas with few stations and in areas with multiple stations that may not adequately sample local microclimates. Accordingly, visualizing station density and confidence metrics alongside performance maps can provide a more balanced assessment of dataset reliability, help avoid overinterpretation of localized discrepancies and help interpret product differences in view of both gauge sparsity and representativeness (Bulovic et al., 2020; Hunziker et al., 2017).

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

5.5 Future directions

Our results confirm the effectiveness of products such as Rain4PE and PISCO in the region. However, they still remain relatively imprecise for assessing potential requirements in hydroclimatic processes. The application of a simple, parsimonious downscaling method, such as the GWR approach, proved particularly effective in preserving or even enhancing performance and increasing spatial detail. Future work will therefore focus on integrating this downscaling framework with Rain4PE and/or PISCO, combining their respective strengths to improve spatial representation. In addition, the downscaled products themselves could be refined by applying scaling procedures and correcting residual biases using multivariate approaches or relationships with complementary datasets. Finally, the suitability of the resulting downscaled products will need to be evaluated in hydrological modelling applications, where improved spatial and temporal representation of precipitation could be key to achieving more realistic simulations of streamflow and the water balance.

6 Conclusions

This study evaluated the performance of satellite-based, regionally optimized and downscaled precipitation datasets in representing daily rainfall characteristics and extremes across the complex terrain of the Peruvian Andes, using in situ rain-gauge observations as a reference. The GPM-IMERGF SPP product and two parsimoniously downscaled variants, using an NDVI-based exponential approach (GPM-EXP) and a geographically weighted regression approach (GPM-GWR), were compared with two regionally optimized gauge-informed datasets: Rain4PE and PISCO. The study focused on their reliability for spatio-temporal representation of daily precipitation, precipitation detection, and extreme-event applications. Overall, our findings support that downscaling adds spatial realism, but only gauge-informed products fundamentally improve temporal and extreme-event skill.

The specific conclusions are that:

- Rain4PE provides the strongest overall performance across quantitative, detection, and temporal metrics, making it the most suitable baseline product for daily-scale applications.
- PISCO excels in representing precipitation extremes and persistence-related indices, benefiting from its strong grounding in local observations and long-term climatological consistency.
- GPM-IMERGF shows systematic limitations in complex Andean terrain, particularly for daily extremes and persistence metrics, although these weaknesses are consistent with findings from other mountainous regions.
- Downscaling enhances spatial realism without necessarily degrading performance, provided that the method accounts for spatial heterogeneity. Among the tested approaches, GWR-based downscaling stands out as it preserves the temporal characteristics of the original dataset while substantially increasing spatial detail, outperforming the simpler NDVI-only exponential method, which tends to amplify biases and false detections.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

- Parsimonious, physically informed downscaling, such as the GWR approach, offers a robust pathway for refining coarse precipitation datasets, particularly in data-scarce regions, as it avoids excessive amplification of biases while allowing relationships between precipitation and terrain-related predictors to vary spatially. However, it should be viewed as complementary to, rather than a substitute for, regionally calibrated, gauge-informed datasets.

Overall, the results highlight the value of regionally optimized products and the potential of physically-informed downscaling methodologies. Overall, the analyses demonstrate that meaningful gains in daily precipitation representation in complex mountainous environments are primarily achieved through the integration of high-quality local observations, while terrain-aware downscaling can add spatial value where such observations are unavailable. The demonstrated strengths of the GWR method support its computationally efficient foundation for integration into high-quality spatio-temporal precipitation products, as well as its application in hydrological modelling, with great potential to enhance water resource assessments in the Andes and similar data-limited mountainous regions.

Code, data, or code and data availability

Processing and analysis scripts, as well as processed data for the selected stations of study are available at our GitHub repository (https://github.com/Jvi9/Precipitation-performance). The gridded precipitation datasets used in this study are publicly available open-source products and can be accessed from their respective official repositories.

Author contributions

JT and JV contributed in the conceptualization, data curation, formal analysis, investigation, methodology, resources, software, visualization, and writing of the original draft and edition of the manuscript. JG-Q contributed in conceptualization and review of the manuscript. AG contributed in conceptualization, funding acquisition, supervision and the review and editing of the manuscript.

Competing interests

The authors declare that they have no known competing financial or personal interests that could have influenced the work reported in this paper.

Acknowledgements

Authors wish to thank KU Leuven internal funding STG/21/027 and acknowledge VLIR-UOS ICP Master's scholarships which supported the studies of JT and JV; this work is part of their master's theses at the Interuniversity Programme of Water Resources Engineering organized by KU Leuven and VUB.

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

JT also acknowledges the Research Foundation – Flanders (FWO) funded International Coordination Action (ICA) “Open Water Network: impacts of global change on water quality" (project code G0ADS24N).

References

Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., & Verdin, J. P. (2007). Adequacy of satellite derived rainfall data for stream flow modeling. *Natural Hazards*, *43*(2), 167–185. https://doi.org/10.1007/s11069-007-9121-6

ASF DAAC. (2014). *PALSAR_Radiometric_Terrain_Corrected_high_res* [Dataset]. [object Object]. https://doi.org/10.5067/Z97HFCNKR6VA

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., & Felipe-Obando, O. (2020). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, *65*(5), 770–785. https://doi.org/10.1080/02626667.2019.1649411

Blumenfeld, J. (2015). *From TRMM to GPM: The Evolution of NASA Precipitation Data | Earthdata*. https://www.earthdata.nasa.gov/learn/articles/trmm-to-gpm

Bougeault, P. (2003). *The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events*.

Bulovic, N., McIntyre, N., & Johnson, F. (2020). Evaluation of IMERG V05B 30-Min Rainfall Estimates over the High-Elevation Tropical Andes Mountains. *Journal of Hydrometeorology*, *21*(12), 2875–2892. https://doi.org/10.1175/JHM-D-20-0114.1

Cárdenas, M. F., Tobón, C., & Buytaert, W. (2017). Contribution of occult precipitation to the water balance of páramo ecosystems in the Colombian Andes. *Hydrological Processes*, *31*(24), 4440–4449. https://doi.org/10.1002/hyp.11374

Chen, C., Chen, Q., Qin, B., Zhao, S., & Duan, Z. (2020). Comparison of Different Methods for Spatial Downscaling of GPM IMERG V06B Satellite Precipitation Product Over a Typical Arid to Semi-Arid Area. *Frontiers in Earth Science*, *8*, 536337. https://doi.org/10.3389/feart.2020.536337

Chen, M., Huang, Y., Li, Z., Larico, A. J. M., Xue, M., Hong, Y., Hu, X.-M., Novoa, H. M., Martin, E., McPherson, R., Zhang, J., Gao, S., Wen, Y., Perez, A. V., & Morales, I. Y. (2022). Cross-Examining Precipitation Products by Rain Gauge,

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Remote Sensing, and WRF Simulations over a South American Region across the Pacific Coast and Andes. *Atmosphere*, *13*(10), 1666. https://doi.org/10.3390/atmos13101666

Claros, E., Dominguez, F., & Liu, C. (2025). Moisture Sources of Precipitation Using Convection-Permitting Simulations: A Study Over South America. *Geophysical Research Letters*, *52*(23), e2025GL118806. https://doi.org/10.1029/2025GL118806

Davenport, M. L., & Nicholson, S. E. (1993). On the relation between rainfall and the Normalized Difference Vegetation Index for diverse vegetation types in East Africa. *International Journal of Remote Sensing*, *14*(12), 2369–2389. https://doi.org/10.1080/01431169308954042

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., Chang, C.-H., Chen, H., Delrieu, G., Hsu, Y., Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E., Sahlu, D., Salerno, F., Rodríguez-Sánchez, J.-P., Vergara, H., & Yilmaz, K. (2019). Evaluation of GPM-era Global Satellite Precipitation Products over Multiple Complex Terrain Regions. *Remote Sensing*, *11*(24), 2936. https://doi.org/10.3390/rs11242936

Didan, K. (2021). *MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V061* [Dataset]. [object Object]. https://doi.org/10.5067/MODIS/MOD13A2.061

Dominguez, F., Eiras-Barca, J., Yang, Z., Bock, D., Nieto, R., & Gimeno, L. (2022). Amazonian Moisture Recycling Revisited Using WRF With Water Vapor Tracers. *Journal of Geophysical Research: Atmospheres*, *127*(4), e2021JD035259. https://doi.org/10.1029/2021JD035259

Espinoza, J. C., Chavez, S., Ronchail, J., Junquas, C., Takahashi, K., & Lavado, W. (2015). Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric circulation. *Water Resources Research*, *51*(5), 3459–3475. https://doi.org/10.1002/2014WR016273

Fernandez-Palomino, C. A., Hattermann, F., Krysanova, V., Lobanova, A., Fiorella Vega-Jácome, Waldo Lavado, William Santini, Cesar Aybar, & Axel Bronstert. (2021). A novel high-resolution gridded precipitation dataset for Peruvian and Ecuadorian watersheds – development and hydrological evaluation. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-20-0285.1

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Foody, G. M. (2003). Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship. *Remote Sensing of Environment*, *88*(3), 283–293. https://doi.org/10.1016/j.rse.2003.08.004

Guo, J., Liang, X., & Ruby Leung, L. (2004). Impacts of different precipitation data sources on water budgets. *Journal of Hydrology*, *298*(1–4), 311–334. https://doi.org/10.1016/j.jhydrol.2003.08.020

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., & Xie, O. (2014). *Integrated Multi-satellitE Retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center*. ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M., & Brönnimann, S. (2017). Identifying, attributing, and overcoming common data quality issues of manned station observations. *International Journal of Climatology*, *37*(11), 4131–4145. https://doi.org/10.1002/joc.5037

Immerzeel, W. W., Quiroz, R. A., & De Jong, S. M. (2005). Understanding precipitation patterns and land use interaction in Tibet using harmonic analysis of SPOT VGT-S10 NDVI time series. *International Journal of Remote Sensing*, *26*(11), 2281–2296. https://doi.org/10.1080/01431160512331326611

Immerzeel, W. W., Rutten, M. M., & Droogers, P. (2009). Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sensing of Environment*, *113*(2), 362–370. https://doi.org/10.1016/j.rse.2008.10.004

Jia, S., Zhu, W., Lű, A., & Yan, T. (2011). A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. *Remote Sensing of Environment*, *115*(12), 3069–3079. https://doi.org/10.1016/j.rse.2011.06.009

Kofidou, M., Stathopoulos, S., & Gemitzi, A. (2023). Review on spatial downscaling of satellite derived precipitation estimates. *Environmental Earth Sciences*, *82*(18), 424. https://doi.org/10.1007/s12665-023-11115-7

Kumar, S., Amarnath, G., Ghosh, S., Park, E., Baghel, T., Wang, J., Pramanik, M., & Belbase, D. (2022). Assessing the Performance of the Satellite-Based Precipitation Products (SPP) in the Data-Sparse Himalayan Terrain. *Remote Sensing*, *14*(19), 4810. https://doi.org/10.3390/rs14194810

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Malo, A. R., & Nicholson, S. E. (1990). A study of rainfall and vegetation dynamics in the African Sahel using normalized difference vegetation index. *Journal of Arid Environments*, *19*(1), 1–24. https://doi.org/10.1016/S0140-1963(18)30825-5

Ochoa, A., Pineda, L., Crespo, P., & Willems, P. (2014). Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific–Andean region of Ecuador and Peru. *Hydrology and Earth System Sciences*, *18*(8), 3179–3193. https://doi.org/10.5194/hess-18-3179-2014

Oshan, T., Li, Z., Kang, W., Wolf, L., & Fotheringham, A. (2019). mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *ISPRS International Journal of Geo-Information*, *8*(6), 269. https://doi.org/10.3390/ijgi8060269

Padrón, R., Feyen, J., Córdova, M., Crespo, P., & Célleri, R. (2020). Rain Gauge Inter-Comparison Quantifies Deficiencies in Precipitation Monitoring. *La Granja*, *31*(1), 7–20. https://doi.org/10.17163/lgr.n31.2020.01

Peterson, T. C. (2005). *Climate Change Indices, WMO Bulletin* (54 (2), pp. 83–86). WMO. https://etccdi.pacificclimate.org/papers/WMO.Bulletin.April.2005.indices.pdf

Scheel, M. L. M., Rohrer, M., Huggel, Ch., Santos Villar, D., Silvestre, E., & Huffman, G. J. (2011). Evaluation of TRMM Multi-satellite Precipitation Analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution. *Hydrology and Earth System Sciences*, *15*(8), 2649–2663. https://doi.org/10.5194/hess-15-2649-2011

SENAMHI. (2017). *Uso del Producto Grillado PISCO de precipitación en Estudios, Investigaciones y Sistemas Operacionales de Monitoreo y Pronóstico Hidrometeorológico*. https://www.senamhi.gob.pe/load/file/01402SENA-8.pdf

Shen, Z., & Yong, B. (2021). Downscaling the GPM-based satellite precipitation retrievals using gradient boosting decision tree approach over Mainland China. *Journal of Hydrology*, *602*, 126803. https://doi.org/10.1016/j.jhydrol.2021.126803

Smith, M. B., Koren, V. I., Zhang, Z., Reed, S. M., Pan, J.-J., & Moreda, F. (2004). Runoff response to spatial variability in precipitation: An analysis of observed data. *Journal of Hydrology*, *298*(1–4), 267–286. https://doi.org/10.1016/j.jhydrol.2004.03.039

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Su, F., Hong, Y., & Lettenmaier, D. P. (2008). Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin. *Journal of Hydrometeorology*, *9*(4), 622–640. https://doi.org/10.1175/2007JHM944.1

Sun, W., Chen, R., Wang, L., Wang, Y., Han, C., & Huai, B. (2022). How do GPM and TRMM precipitation products perform in alpine regions?: A case study in northwestern China's Qilian Mountains. *Journal of Geographical Sciences*, *32*(5), 913–931. https://doi.org/10.1007/s11442-022-1978-5

Tan, M. L., Samat, N., Chan, N. W., & Roy, R. (2018). Hydro-Meteorological Assessment of Three GPM Satellite Precipitation Products in the Kelantan River Basin, Malaysia. *Remote Sensing*, *10*(7), Article 7. https://doi.org/10.3390/rs10071011

Tang, G., Ma, Y., Long, D., Zhong, L., & Hong, Y. (2016). Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *Journal of Hydrology*, *533*, 152–167. https://doi.org/10.1016/j.jhydrol.2015.12.008

Wang, G., Zhang, X., & Zhang, S. (2019). Performance of Three Reanalysis Precipitation Datasets over the Qinling-Daba Mountains, Eastern Fringe of Tibetan Plateau, China. *Advances in Meteorology*, *2019*, 1–16. https://doi.org/10.1155/2019/7698171

Wang, H., Zang, F., Zhao, C., & Liu, C. (2022). A GWR downscaling method to reconstruct high-resolution precipitation dataset based on GSMaP-Gauge data: A case study in the Qilian Mountains, Northwest China. *Science of The Total Environment*, *810*, 152066. https://doi.org/10.1016/j.scitotenv.2021.152066

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed). Academic Press.

World Meteorological Organization. (2018). *Guide to Instruments and Methods of Observation*.

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., & Hou, A. (2017). Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on a high-density rain gauge network. *Journal of Geophysical Research: Atmospheres*, *122*(2), 910–924. https://doi.org/10.1002/2016JD025418

Xu, S., Wu, C., Wang, L., Gonsamo, A., Shen, Y., & Niu, Z. (2015). A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. *Remote Sensing of Environment*, *162*, 119–140. https://doi.org/10.1016/j.rse.2015.02.024

https://doi.org/10.5194/egusphere-2026-604
Preprint. Discussion started: 2 April 2026

Yilmaz, K. K., Hogue, T. S., Hsu, K., Sorooshian, S., Gupta, H. V., & Wagener, T. (2005). Intercomparison of Rain Gauge, Radar, and Satellite-Based Precipitation Estimates with Emphasis on Hydrologic Forecasting. *Journal of Hydrometeorology*, *6*(4), 497–517. https://doi.org/10.1175/JHM431.1

Yu, L., Leng, G., & Python, A. (2022). A comprehensive validation for GPM IMERG precipitation products to detect extremes and drought over mainland China. *Weather and Climate Extremes*, *36*, 100458. https://doi.org/10.1016/j.wace.2022.100458

Zhan, C., Han, J., Hu, S., Liu, L., & Dong, Y. (2018). Spatial Downscaling of GPM Annual and Monthly Precipitation Using Regression-Based Algorithms in a Mountainous Area. *Advances in Meteorology*, *2018*, 1–13. https://doi.org/10.1155/2018/1506017