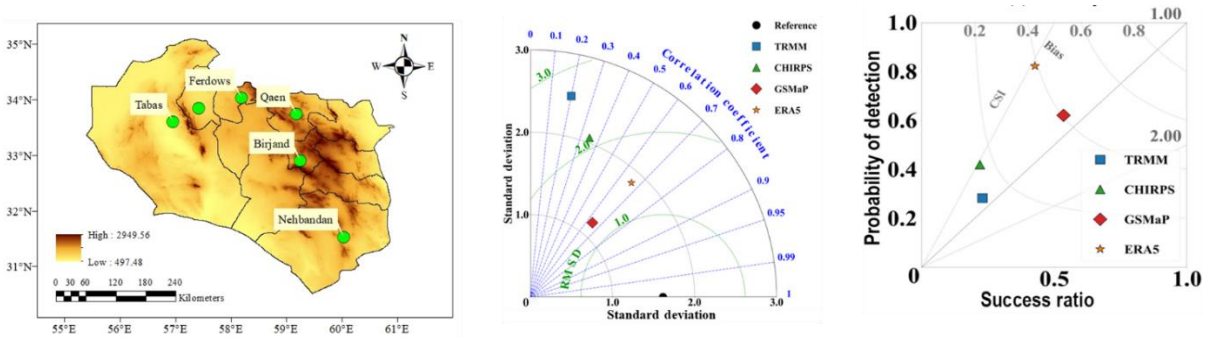
Journal of Applied Research in Water and Wastewater
E-ISSN: 2476-6283
Journal homepage: https://arww.razi.ac.ir

Assessing the suitability of gridded precipitation products for hydro-climatic applications in a sparsely Gauged arid basin of Iran

Ali Nasirian*, Mahsa Mardani

Department of Civil Engineering, Faculty of Engineering, University of Birjand, Birjand, Iran.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article type:
Research Article

Article history:
Received 19 August 2025
Received in revised form 13 November 2025
Accepted 20 November 2025
Available online 30 December 2025

Keywords:
Complex topography
Data-scarce region
Error characterization
Multi-scale
Seasonal performance

ABSTRACT

Hydro-climatic research and water resource management in arid, data-scarce regions depend fundamentally on precise precipitation data. This study presents the first comprehensive, multi-scale evaluation of four prominent gridded precipitation products (GPPs)—TRMM, CHIRPS, GSMaP, and ERA5—in the climatically challenging and sparsely gauged South Khorasan province of Iran (2010–2019). Using ground-based observations as a reference, GPP performance was evaluated across multiple timescales with a suite of statistical metrics. The evaluation framework leverages diagnostic visualizations, such as Taylor and performance diagrams, to provide deeper insights into error structures than can be achieved through traditional map-based assessments. The analysis revealed a clear performance ranking: the satellite-based TRMM and GSMaP consistently performed best, showing higher accuracy (median RMSE ≈ 2.91–3.05 mm/day), stronger correlation (median CC ≈ 0.63–0.65), and a more balanced detection skill (median CSI ≈ 0.43–0.45). In contrast, the ERA5 reanalysis product, despite achieving the highest probability of detection (POD ≈ 0.78), suffered from notable systematic biases and the largest random errors. Performance for all products degraded during the arid summer, and estimation errors systematically increased in wetter regions. We conclude that the gauge-adjusted satellite products, GSMaP and TRMM, provide the most dependable precipitation estimates for the study area. These findings offer a critical, evidence-based guide for selecting appropriate GPPs in this vulnerable environment and provide insights for future algorithm development.

1. Introduction

Precipitation serves as a crucial element in the Earth's water and energy cycles and is an essential input variable for various applications, including hydrological modeling, agricultural planning, and the management of water-related natural hazards such as floods and droughts (Shirmohammadi–Aliakbarkhani *et al.*, 2025; Wei *et al.*, 2025). High-quality precipitation data with precise spatiotemporal resolution are essential for scientific research and operational decision-making, especially in arid and semi-arid regions facing considerable water resource challenges (Tosan *et al.*, 2025a). In these water-limited environments, inaccurate precipitation estimates can lead to inefficient reservoir operations, failed agricultural outputs, and inadequate preparedness for flash floods, thereby directly threatening both food security and public safety (En-nagre *et al.*, 2025). While ground-based rain gauges provide the most direct measurements of precipitation, their primary limitation stems not from the measurement itself, but from the network's spatial coverage. Most gauge networks are sparse and unevenly distributed (Tosan *et al.*, 2025b). This challenge is exacerbated in areas characterized by complex topography or remote access, where station density is often insufficient to resolve critical spatial variations in rainfall (Li *et al.*, 2025; Singh, Thakur and Mohanty, 2025). Consequently, these data gaps create a significant bottleneck, impeding progress in regional climate studies and hydrological modeling, and they complicate the design of effective water policies (Ganiyu *et al.*, 2025). To overcome these limitations, gridded precipitation products (GPPs) are widely utilized. These datasets are derived from either satellite remote sensing or atmospheric reanalysis models. Satellite-based products offer quasi-global coverage with increasingly fine spatiotemporal resolutions. Two well-known examples are the tropical rainfall measuring mission (TRMM) and the Global Satellite Mapping of Precipitation (GSMaP) (Huang *et al.*, 2025). Reanalysis products are another important source. They provide datasets that are spatiotemporally consistent. A primary example is the

*Corresponding author Email: a.nasirian@birjand.ac.ir

Cite this article: A. Nasirian, M. Mardani, Assessing the suitability of gridded precipitation products for hydro-climatic applications in a sparsely Gauged arid basin of Iran, *Journal of Applied Research in Water and Wastewater*, 12 (2), 2025, 167-177. doi: https://doi.org/10.22126/arww.2025.12654.1400

fifth-generation ECMWF reanalysis (ERA5) (Majidi *et al.*, 2025). However, they are subject to inherent uncertainties and systematic biases. These issues arise from retrieval algorithms, sensor limitations, and model parameterizations (Ganiyu *et al.*, 2025). Consequently, a crucial prerequisite for their reliable application is a rigorous regional-scale validation against ground-based observations. The application of GPPs in any hydro-climatic study is therefore contingent upon this foundational validation step Abbas *et al.*, 2024; Bekić and Leskovar, 2025). While numerous GPP evaluations have been conducted worldwide, including in other parts of Iran (e.g., the neighboring Khorasan Razavi province), their findings cannot be reliably extrapolated to the unique environment of South Khorasan. This province is characterized by a particularly challenging combination of extreme aridity, complex mountainous terrain, and distinct local climatic influences, all of which are known to amplify GPP uncertainties. Therefore, the existing literature suffers from a critical research gap: the absence of a dedicated and comprehensive GPP assessment tailored specifically to the hydro-climatic conditions of South Khorasan. Such a localized evaluation is not merely an incremental addition but an essential prerequisite for any reliable application of GPPs in this vulnerable, water-scarce region.

Therefore, this study aims to fill this critical research gap by conducting a rigorous and multi-faceted evaluation of four prominent GPPs: TRMM, CHIRPS, GSMaP, and ERA5. The performance of these products is evaluated against data from available ground-based synoptic stations in South Khorasan province, Iran, for the period spanning 2010 to 2019. The novelty of this work is threefold: (1) it provides the first dedicated GPP intercomparison for the data-scarce and climatically challenging South Khorasan province; (2) it assesses a unique combination of legacy (TRMM) and modern (GSMaP, ERA5) products across multiple temporal scales; and (3) it establishes a complete analytical workflow using statistical and diagnostic visualizations to evaluate temporal and error structures, offering a valuable alternative to traditional map-based assessments. The findings are intended to provide an essential, evidence-based guide for the selection and application of GPPs for critical water resource management in this vulnerable region and to contribute to the broader understanding of GPP performance in arid, topographically complex environments.

2. Materials and methods

2.1. Study area

This study was conducted in South Khorasan province, a data-scarce region of approximately 151,000 km² located in eastern Iran. The province spans from approximately 30° 31' to 34° 53' N latitude and 57° 3' to 61° 0' E longitude. Fig. 1 displays the province's location, its digital elevation model (DEM), and the spatial layout of the synoptic stations utilized in this study.

Table 1 lists the specific details of these stations. The region is characterized by a complex topography, with extensive mountain ranges that generally exhibit a north-south orientation (Tosan *et al.*, 2015). Elevations range from approximately 650 m in the vast plains and the Dasht-e Lut desert basin to a maximum of 3615 m in the mountainous areas. This varied terrain directly results in significant spatial variability in precipitation (Rezvani Moghaddam *et al.*, 2016).
The regional climate is predominantly arid to semi-arid, with typically arid, desert-like conditions in the lowlands and semi-arid conditions in mountainous areas. The province's proximity to major deserts intensifies these dry conditions (Feizi and Tosan, 2017). Synoptic-scale winds are also a significant influencing factor, particularly the well-known 120-day Sistan wind. Consequently, the region is characterized by a challenging climate and a limited number of long-term synoptic stations. This combination of factors renders the accurate estimation of precipitation a critical task, as such estimates are essential for both water resource management and agricultural planning in the province.

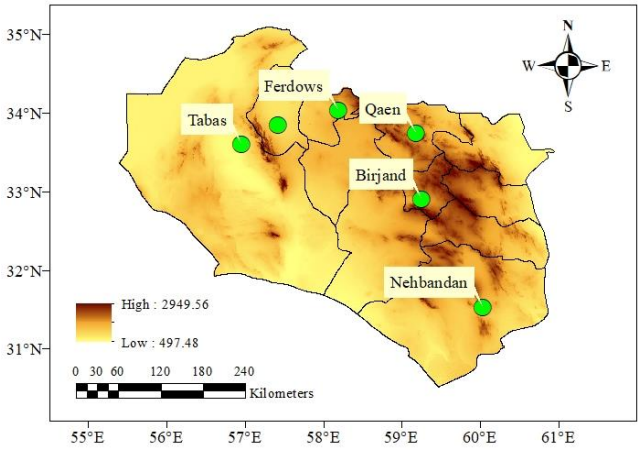


Fig. 1. The study area, South Khorasan province, and the location of the ground-based synoptic stations.

Table 1. Geographical characteristics of the synoptic stations used in the study.

Station ID	Station name	Longitude (°E)	Latitude (°N)	Elevation, m	Mean annual precipitation (MAP), mm/year	SD of annual precipitation, mm/year
40809	Birjand	59.25	32.9	1491	124.61	51.2
99407	Boshruyeh	57.42	33.85	879	72.7	37.8
40792	Ferdows	58.18	34.03	1293	118.49	49.5
40827	Nehbandan	60.03	31.53	1188	89.41	42.1
40793	Qaen	59.18	33.74	1432	150.37	58.6
40791	Tabas	56.95	33.6	711	65.79	36.3

2.2. Datasets

2.2.1. Ground-based observations

Daily precipitation records, spanning the 2010–2019 period, were obtained from the Iran meteorological organization (IRIMO). These data were sourced from a network of synoptic stations distributed throughout the study area. Quality control procedures were applied to these in-situ measurements to ensure data integrity. This quality-controlled dataset subsequently served as the reference ground truth for evaluating the GPPs.

2.2.2. GPPs

For this evaluation, four widely used GPPs were selected, comprising three satellite-based products and one reanalysis product. All datasets were acquired for the same 2010–2019 period to ensure a consistent comparative framework. The analysis utilized the TRMM multi-satellite precipitation analysis (TMPA) product, version 3B42 v7. The retrieval algorithm for this product merges microwave with infrared data from multiple satellites, and the merged data is subsequently adjusted using rain gauge measurements. This process yields precipitation estimates at a 0.25° spatial resolution with a 3-hourly temporal resolution. For almost two decades, TRMM was a key benchmark in hydro-climatic studies and has been successfully applied previously in this specific region (Zhang *et al.*, 2022). The CHIRPS v2.0 dataset was also used for this study. This dataset provides quasi-global precipitation information and is characterized by a very high spatial resolution of 0.05°. The algorithm begins with thermal infrared observations of cold cloud duration and then blends this satellite information with data from a large network of ground stations. The goal of this final blending step is to reduce bias and improve accuracy (Du *et al.*, 2024).

This study also selected the gauge-adjusted, standard near-real-time product of GSMaP, version 8. This product originates from the GPM-era of satellites. Its core data is derived from passive microwave radiometers; this information is then enhanced using infrared data (Nourani *et al.*, 2025). A final adjustment is applied using daily data from a global network of gauges, resulting in a high-quality estimate with a 0.1° spatial and an hourly temporal resolution (Taña *et al.*, 2025).
ERA5 is the fifth-generation global climate reanalysis dataset from the european centre for medium-range weather forecasts (ECMWF). It provides hourly precipitation estimates at a 0.25° spatial resolution. The dataset's creation involves assimilating a vast number of historical observations into an advanced data assimilation and modeling system. Its application in climate and hydrological research is growing. A key characteristic of this product, however, is that its performance can vary significantly by region (Keune *et al.*, 2025).

2.3. Evaluation methodology

2.3.1. Comparison approach and preprocessing

The performance of each GPP was evaluated against ground-based station data using a point-to-pixel comparison methodology. For each

station, the precipitation time series was extracted from the corresponding grid cell of each GPP dataset. Two preprocessing steps preceded the main analysis. First, GPPs with sub-daily native resolutions (TRMM, GSMaP, and ERA5) were aggregated to a consistent daily timescale. Second, to harmonize the GPP estimates with the detection capabilities of physical rain gauges and mitigate the "drizzle" effect common in reanalysis and satellite products, a daily precipitation threshold of 1.0 mm/day was applied. This value is a widely adopted standard in GPP validation studies to account for the detection limitations of standard gauges and effectively filter out artificial light precipitation in gridded products. Any GPP-estimated daily rainfall value below this threshold was set to zero. While this temporal aggregation is essential for a consistent comparison, we acknowledge that it may mask the differential performance of products in capturing sub-daily precipitation dynamics, particularly for the short-duration, high-intensity convective events common in the region's arid climate. An investigation at the native resolutions of these products would be a valuable direction for future studies.

2.3.2. Evaluation metrics

The performance of the GPPs was quantified using a comprehensive set of seven statistical metrics, which are summarized in Table 2. These metrics are divided into two categories to assess both the accuracy of the precipitation amount (continuous metrics) and the skill of event detection (categorical metrics).

Four Continuous metrics were used to evaluate the agreement and magnitude of error between GPP estimates and gauge observations (Table 2): The pearson correlation coefficient (CC) measures the linear correlation between the estimated and observed values, ranging from -1 to 1. The Root Mean Square Error (RMSE) represents the standard deviation of the prediction errors, penalizing larger errors more heavily. The mean absolute error (MAE) measures the average magnitude of the errors, providing a linear score of the error quantity.

The Mean Bias Error (MBE) indicates the average tendency of the GPP to overestimate (positive values) or underestimate (negative values) precipitation. Three categorical metrics were used to evaluate the precipitation detection skill based on a contingency table constructed from a 1.0 mm/day threshold. An event was classified as a hit (H) if both the gauge and GPP recorded precipitation ≥ 1.0 mm, a miss (M) if the gauge recorded rain but the GPP did not, and a false Alarm (F) if the GPP recorded rain but the gauge did not.

The probability of detection (POD) measures the fraction of observed rain events that were correctly detected by the GPP. The false alarm ratio (FAR) measures the fraction of GPP-detected rain events that did not actually occur. The critical success index (CSI), also known as the threat score, is a balanced metric that penalizes both misses and false alarms, providing an overall measure of detection skill.

2.3.3. Analysis framework

The evaluation was conducted across several temporal and analytical scales. The initial analysis focused on daily performance, for which the seven statistical metrics were calculated at each station and visualized using boxplots. The subsequent analysis investigated seasonal performance by aggregating the daily data into monthly sums. This seasonal assessment relied on two key diagnostic tools: Taylor diagrams and performance diagrams.

3. Results and discussion

3.1 Overall performance assessment of GPPs

This section presents a comprehensive performance assessment of four GPPs (TRMM, CHIRPS, GSMaP, and ERA5) against daily ground-based observations from synoptic stations in South Khorasan province, Iran, for the 2010–2019 period. Fig. 2 visualizes the statistical distributions of seven metrics, including both continuous and categorical types, allowing for a robust intercomparison of each product's accuracy, bias, and event detection skill. This evaluation approach aligns with standard practices in GPP validation studies (Akbas and Ozdemir, 2024; Wang, Chen and Li, 2025).

In terms of correlation with ground observations (Fig. 2a), TRMM exhibited the strongest agreement, achieving the highest median CC of approximately 0.65. Its performance was also highly consistent across the study area, as reflected by a compact interquartile range (IQR) of only 0.14 (0.58–0.72). GSMaP also performed favorably, with a median CC of 0.63 and a similar degree of consistency. These strong results for satellite products are in line with previous findings in other parts of Iran (Shirmohammadi–Aliakbarkhani *et al.* 2025).

Table 2. List of statistical metrics used for evaluating the performance of SPPs.

Category	Metric	Formula
Continuous metrics	CC	$\frac{\sum_{i=1}^{n}(O_i-\bar{O})(E_i-\bar{E})}{\sqrt{\sum_{i=1}^{n}(O_i-\bar{O})^2}\sqrt{\sum_{i=1}^{n}(E_i-\bar{E})^2}}$
	RMSE	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i-O_i)^2}$
	MAE	$\frac{1}{N}\sum_{i=1}^{N}\lvert E_i-O_i\rvert$
	MBE	$\frac{1}{N}\sum_{i=1}^{N}(E_i-O_i)$
Categorical metrics	POD	$\frac{H}{H+M}$
	FAR	$\frac{F}{H+F}$
	CSI	$\frac{\mathrm{H}}{\mathrm{H+M+F}}$

In contrast, both CHIRPS and ERA5 showed weaker correlations (median CCs of 0.60 and 0.58, respectively). The moderate result for ERA5 is consistent with established patterns, as previous research has shown that reanalysis products often struggle in areas with complex topography (Ahmed *et al.*, 2024; Yan *et al.*, 2024). An analysis of error magnitudes (RMSE, Fig. 2b; MAE, Fig. 2c) identified TRMM and GSMaP as the top-performing products. They achieved the lowest median RMSEs (≈2.91 and 3.05 mm/day) and the lowest MAEs (≈1.1 mm/day). For GSMaP, this high level of accuracy is consistent with results from various regional evaluations (Masood *et al.*, 2023; Kumar *et al.*, 2025). Conversely, ERA5 produced the largest errors, with a median RMSE of 3.85 mm/day and the widest error distribution, which indicates a risk of large, occasional miscalculations. This finding in our mountainous study area echoes other reports where ERA5 has struggled against satellite products in similar terrains (Chen, Collet and Di Luca, 2024; Li *et al.*, 2025). The CHIRPS product exhibited intermediate performance, with a median RMSE of approximately 3.40 mm/day. The MBE (Fig. 2d) analysis provides critical insight into systematic over- or underestimation tendencies. The results indicate a distinct behavioral pattern: TRMM and CHIRPS exhibited a tendency toward underestimation, with median MBE values of approximately -0.11 mm/day and -0.09 mm/day, respectively. Conversely, GSMaP and ERA5 displayed a tendency toward overestimation, with median MBEs of +0.15 mm/day and +0.21 mm/day. The tendency of ERA5 to overestimate precipitation is a well-documented characteristic reported in other evaluations (Chang, Qi and Wang, 2024; Fatolahzadeh Gheysari *et al.*, 2024). Regarding bias, while the median value for all products is near zero, ERA5 exhibits much greater variability, with the largest IQR for MBE (ranging from -0.05 to +0.50 mm/day). This high station-to-station variability renders it less reliable than TRMM and GSMaP. In comparison, both TRMM and GSMaP demonstrated a more consistent and predictable bias. The products' ability to identify precipitation events was evaluated using the POD, FAR, and CSI, as shown in Fig. 2e, 2f, and 2g. The CSI, which provides a combined measure of skill, identified GSMaP as the most effective product overall, with the highest median score of 0.45. TRMM was a close second with a strong CSI of 0.43. This top ranking reflects a successful balance between two competing metrics. For example, ERA5 was the most sensitive in detecting rainfall events, achieving the highest POD (median ≈ 0.78), a known characteristic of some reanalysis products (Devadarshini *et al.* 2024). However, ERA5 also had the highest FAR (median ≈ 0.48), indicating that the product generated a high number of false alarms. In contrast, TRMM and CHIRPS had the lowest FAR (median ≈ 0.35), although their lower FAR was associated with lower POD scores due to missing more rainfall events. A trade-off between detection and false alarms is often reported in SPP evaluations (Singh, Thakur and Mohanty, 2025). The performance of GSMaP and TRMM was more balanced, resulting in higher reliability for identifying both rain and no-rain days. For hydro-meteorological applications, such balanced reliability is more important than a high detection rate alone (Gu *et al.* 2024).

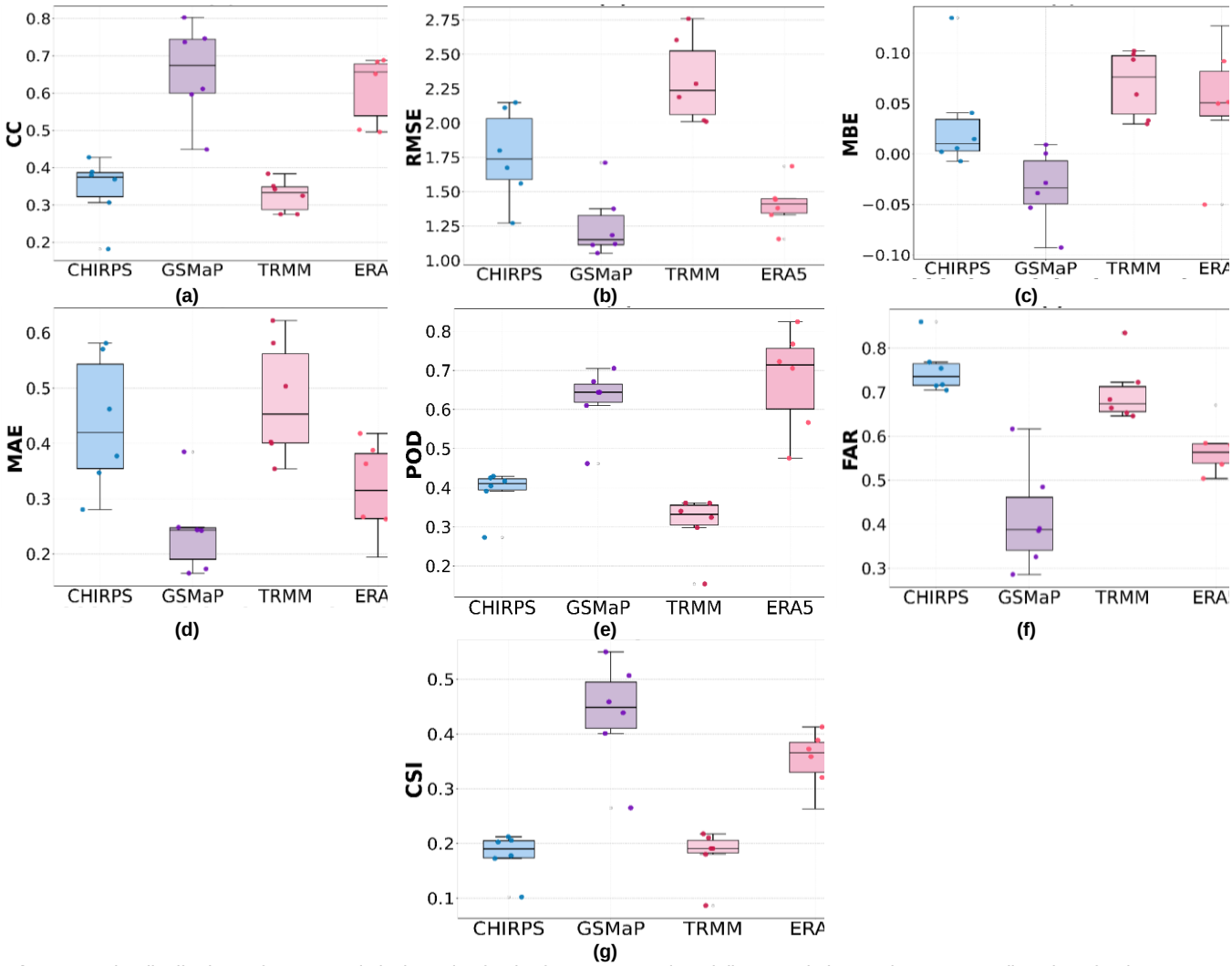


Fig. 2. Boxplot distributions of seven statistical metrics for the four GPPs against daily ground observations across all stations for the 2010–2019 period. The metrics include (a) CC, (b) RMSE, (c) MAE, (d) MBE, (e) POD, (f) FAR, and (g) CSI. The box represents the interquartile range (IQR), the horizontal line is the median, and the whiskers extend to 1.5 times the IQR.

3.2. Monthly and seasonal performance assessment using Taylor diagrams

To assess the performance of the GPPs at a coarser temporal scale, daily precipitation estimates were aggregated into monthly sums, a scale at which performance is often enhanced compared to daily evaluations (Sultan *et al.*, 2024; Ganiyu *et al.*, 2025). The statistical performance for each month was then evaluated using Taylor diagrams (Singh, Thakur and Mohanty, 2025). These diagrams offer a concise summary of three key statistics: the CC, the Centered Root Mean Square Difference (CRMSD), and the standard deviation (SDEV). As presented in Fig. 3, this approach effectively illustrates the seasonal variability in the performance of the four GPPs.

During the primary precipitation season from December to April, all GPPs demonstrated their most robust performance. In the peak wet month of February, for instance, the satellite-based products exhibited exceptional agreement with the reference observations. TRMM and GSMaP achieved the highest correlations, with CC values of 0.88 and 0.86, respectively. Furthermore, their standard deviations (SDEV of 18.1 mm/month for TRMM and 19.5 mm/month for GSMaP) closely matched the observed variability (reference SDEV = 19.9 mm/month), placing them nearest to the reference point on the diagram. CHIRPS also exhibited a high correlation (CC = 0.81) but slightly underestimated the precipitation variability (SDEV = 16.5 mm/month). While ERA5 also registered a high correlation (CC = 0.79), it markedly overestimated the monthly variability during this period, with an SDEV of 22.4 mm/month. These findings are consistent with other studies in different regions of Iran that support the general suitability of these products for seasonal analysis (Nozarpour, Mahjoobi and Golian, 2024).

The performance of all GPPs degraded substantially during the arid summer period from May to September. This is a known issue for GPPs in arid and semi-arid climates (Helmi and Abdelhamed, 2023; Salih *et al.*, 2024). This decline is largely attributable to the nature of summer precipitation in the region, which is often dominated by short-duration, high-intensity, and spatially scattered convective storms that are notoriously difficult for the relatively coarse resolution of GPPs to capture accurately. In August, the driest month, the correlation for all products weakened considerably, with CC values ranging from -0.10 (ERA5) to +0.15 (CHIRPS). The standard deviation for all products was also much lower than the observed variability. For the summer months, the points on the Taylor diagram cluster near the origin, indicating the products' inability to reliably capture either the pattern or the magnitude of infrequent summer rainfall events. Feng *et al.* (2025) have also identified this limitation. The autumn months represented a transitional period during which distinct differences emerged in the GPPs' ability to capture the onset of the precipitation season. As rainfall patterns began to re-establish in November, the performance of all products improved relative to the summer lows. TRMM and GSMaP once again demonstrated superior performance, with high correlations of 0.81 and 0.79, respectively, and the lowest CRMSD values among the products. In contrast, ERA5's performance appeared to respond more slowly to the seasonal change, exhibiting a much lower correlation of 0.65 for the same month. This may suggest a lag in the model's ability to simulate the synoptic patterns that initiate the wet season.

The overall monthly results reveal clear differences in performance among the products. TRMM and GSMaP generally yielded the most accurate and reliable monthly precipitation estimates, with particularly strong performance during the hydro-climatologically important wet season. A common characteristic was observed among all satellite products (TRMM, CHIRPS, GSMaP), as they all exhibited a tendency to underestimate the magnitude of monthly precipitation variability. Conversely, the ERA5 reanalysis product frequently overestimated this variability, particularly during wetter months. This underscores the critical importance of seasonal context in GPP evaluation, as product skill fluctuates dramatically between wet and dry periods (Hisam *et al.*, 2023; Zhu *et al.*, 2025).

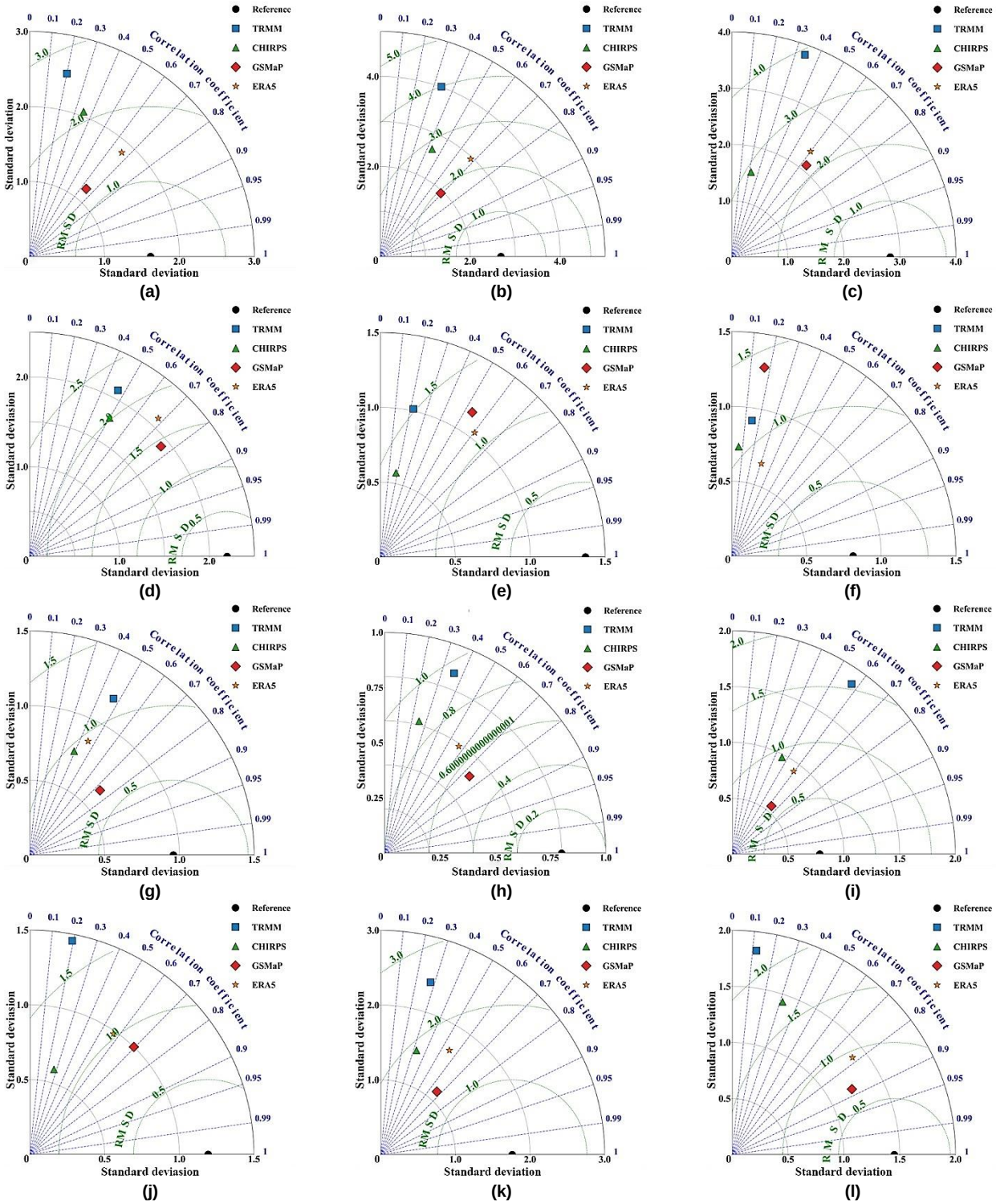


Fig. 3. Monthly Taylor diagrams comparing the four GPPs against ground observations for the 2010–2019 period. Each diagram corresponds to a specific month from (a) January to (l) December. The plots summarize the CC (indicated by the azimuthal angle), the Centered Root Mean Square Difference (gray semicircles centered on the reference point), and the standard deviation (radial distance from the origin). The black circle on the x-axis represents the reference observation.

3.3. Monthly precipitation detection skill assessment

To complement the previous analysis, monthly precipitation detection skill was evaluated using performance diagrams (Fig. 4). These diagrams simultaneously visualize three key metrics—POD, Success Ratio (SR = 1 - FAR), and CSI—providing a comprehensive view of each product's detection capabilities and its inherent trade-offs.

During the main wet season, distinct performance differences emerged among the products. In February, for example, ERA5 exhibited the highest detection sensitivity with a POD of 0.84. However, this high sensitivity came at the cost of a high number of false alarms, resulting in a Success Ratio of only 0.63 (FAR = 0.37). In contrast, GSMaP and TRMM offered a more balanced performance profile. GSMaP achieved a POD of 0.78 and a higher success ratio of 0.71, yielding the month's highest CSI score of 0.59. TRMM also registered a high CSI of 0.55. The results for CHIRPS revealed a different pattern, as it had the highest Success Ratio (0.75) at the expense of the lowest POD (0.65). This result is typical for infrared-based algorithms, which often prioritize precision over detection rates (Prakash and Bhan, 2023).

The detection skill for all products deteriorated during the arid summer months. In July, the CSI for all GPPs fell below 0.25, highlighting the difficulty of accurately identifying sparse precipitation (Salih *et al.*, 2024). During this period, ERA5 was particularly prone to generating false alarms; its success ratio dropped to 0.45 (FAR = 0.55), meaning that more than half of its rain predictions were incorrect. The performance of other products like GSMaP and TRMM also declined. However, their slightly higher CSI scores indicated a somewhat more

reliable detection capability during these challenging months, even though their overall performance was also weak. A trade-off between detection sensitivity and precision was evident across the entire annual cycle. ERA5 consistently functioned as a high-sensitivity product, maximizing the POD but at the cost of a consistently high FAR. In contrast, CHIRPS operated as a high-precision, low-sensitivity model (Singh, Thakur and Mohanty, 2025). The CSI metric is designed to balance these competing characteristics. The CSI analysis revealed that satellite-native products exhibited superior overall performance. According to this metric, GSMaP emerged as the top-performing product, achieving the highest CSI in 8 out of the 12 months. This result is consistent with other studies that also rank it highly among GPM-era products (Gao *et al.* 2024; Kumar *et al.*, 2025). TRMM's performance was a close second. Therefore, GSMaP and TRMM represent the most suitable choices for applications in the study region that require a balanced approach to precipitation detection.

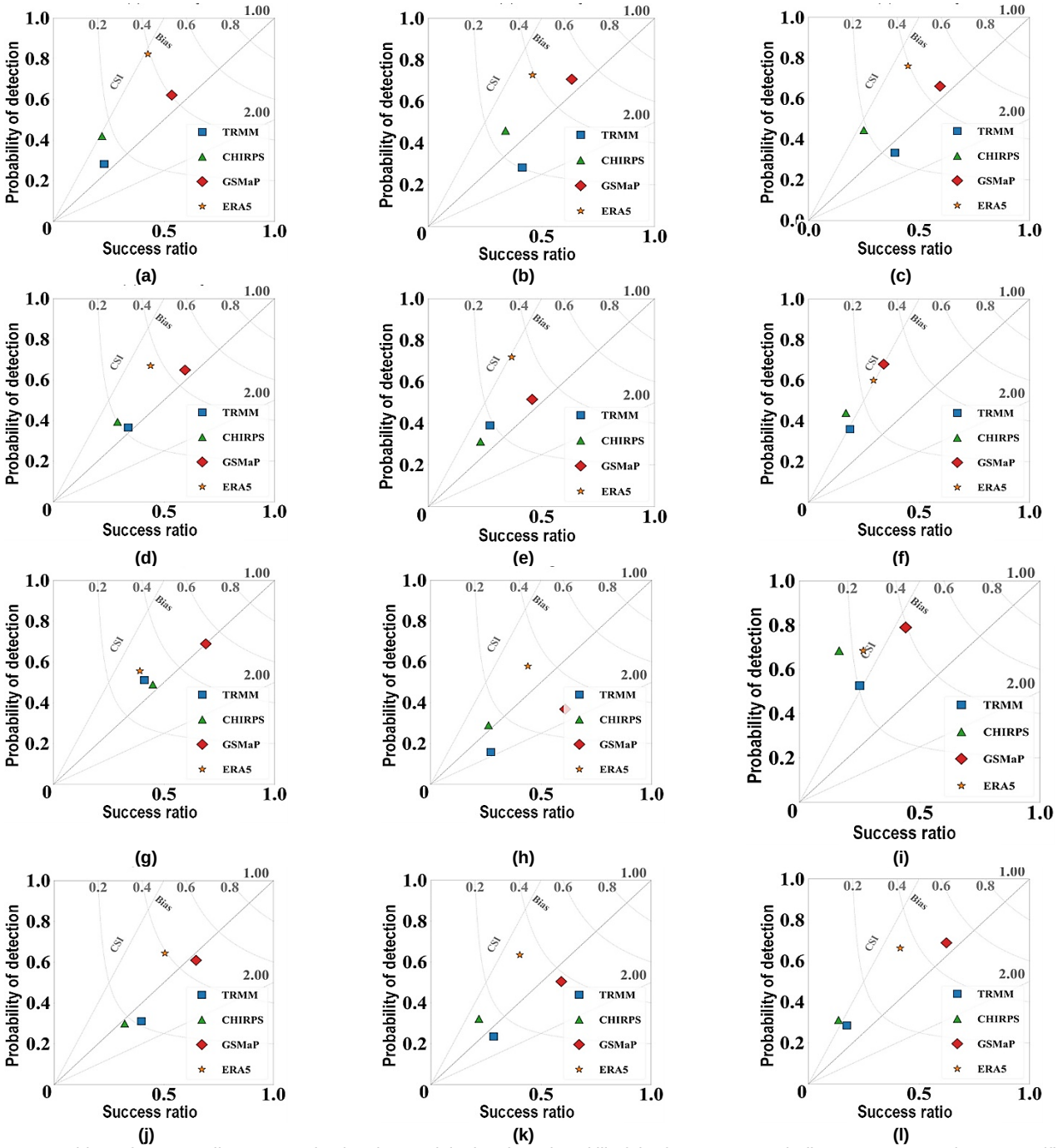


Fig. 4. Monthly performance diagrams evaluating the precipitation detection skill of the four GPPs. Each diagram corresponds to a specific month from (a) January to (l) December. The plots visualize the POD (y-axis) against the Success Ratio (SR = 1-FAR; x-axis). The dashed diagonal lines represent the Bias, and the curved dotted lines indicate the CSI.

3.4. Inter-annual performance stability

To investigate the temporal stability and potential long-term trends in GPP performance, key statistical metrics were aggregated on an annual basis for the 2010–2019 period. The resulting time series, presented in Fig. 5, reveal the year-to-year fluctuations in accuracy, bias, and detection skill. The inter-annual analysis of the CC (Fig. 5a) and RMSE (Fig. 5b) reveals a consistent performance hierarchy. Over the decade, TRMM and GSMaP consistently exhibited the strongest correlations, with annual CC values typically between 0.62 and 0.70. Their annual RMSE values were also the lowest, generally between 2.8 and 3.2 mm/day. Our analysis found no statistically significant performance trend for any product, a result that differs from some other studies that reported improvements in newer product versions (Zhu *et al.*, 2024). In this analysis, ERA5 consistently exhibited the weakest performance metrics, a finding that is consistent with results from other multi-year evaluations (Abbas *et al.,* 2024).

Fig. 5c presents the analysis of the annual MBE. The systematic biases for each product remained persistent over time. TRMM and CHIRPS consistently exhibited a slight underestimation. In contrast, ERA5 and GSMaP showed a persistent overestimation bias. The stability of this bias over time may suggest that the underlying

algorithms and calibration procedures did not undergo major revisions that would have impacted their systematic error characteristics in the study region during this period (Zhu *et al.*, 2025). The overall skill in detecting precipitation events, as measured by the CSI (Fig. 5d), also remained stable over the decade. GSMaP and TRMM consistently achieved the highest scores, with their annual CSI values generally in the 0.42 to 0.46 range. The performance difference between these top products and the others was maintained throughout the study period. In summary, the inter-annual evaluation revealed that the performance characteristics and relative ranking of the four GPPs were largely static, which provides a consistent baseline for their application in regional studies.

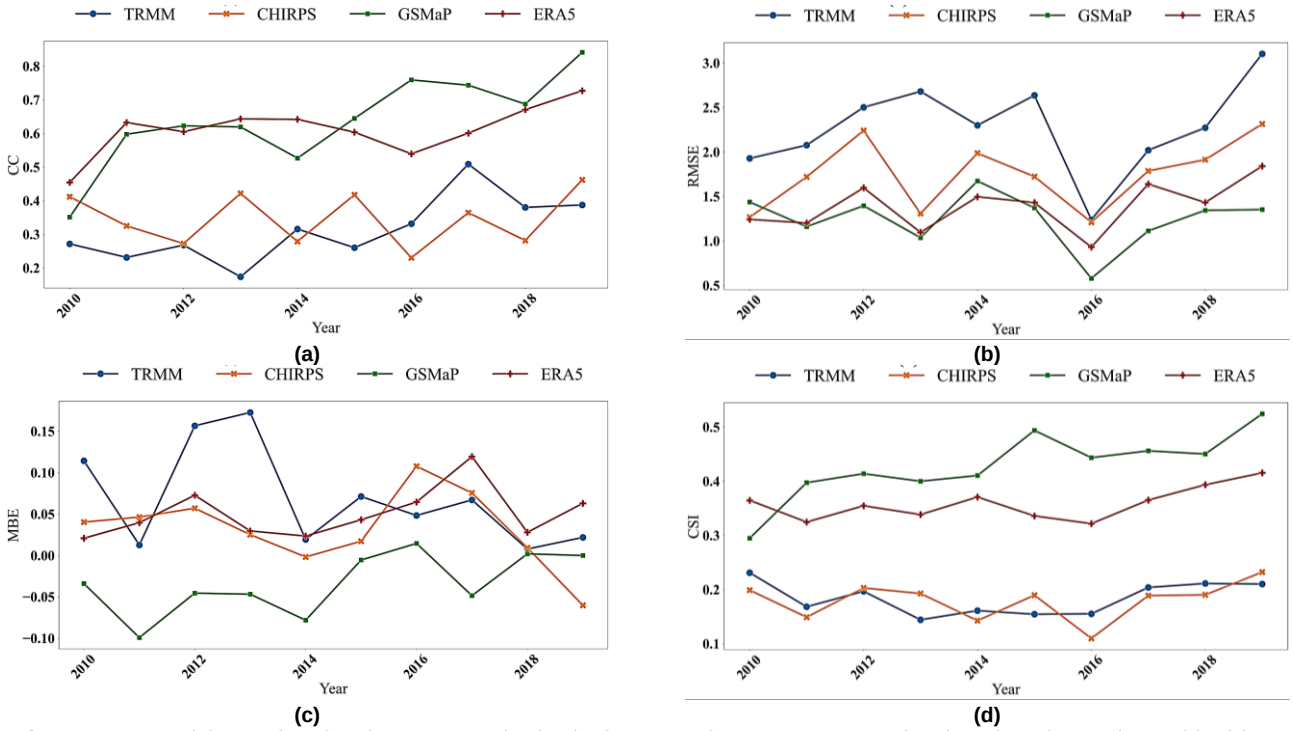
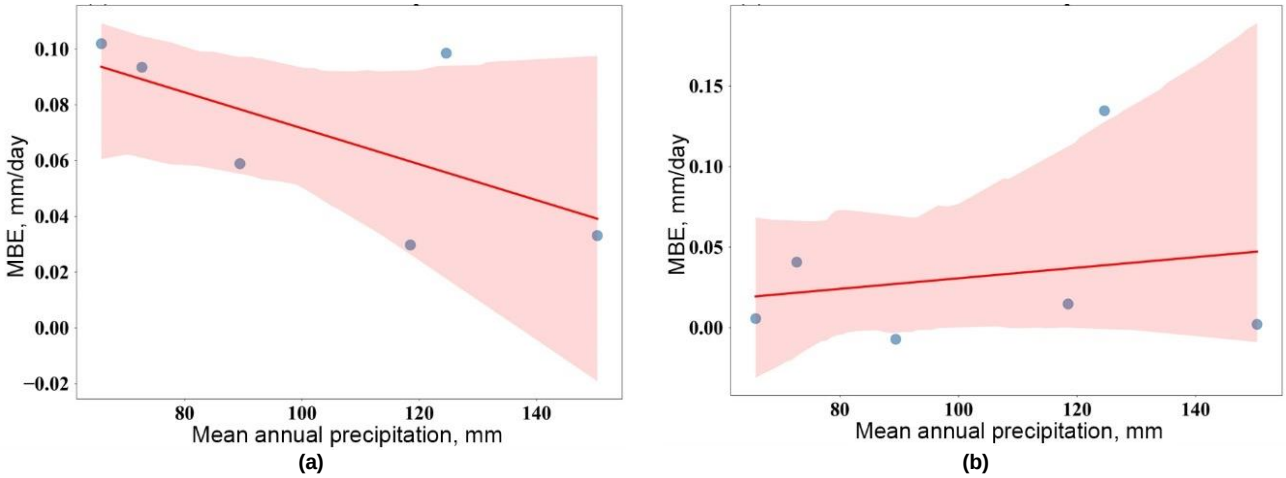


Fig. 5. Inter-annual time series of performance metrics for the four GPPs from 2010 to 2019. The plots show the yearly trend for (a) CC, (b) RMSE, (c) MBE, and (d) CSI.

3.5. Relationship between estimation errors and local climate conditions

To investigate the spatial patterns of estimation errors, the relationship between GPP performance metrics and local climate conditions was examined. This analysis serves as a proxy for understanding the spatial distribution of errors by correlating key error metrics (RMSE and MBE) with the mean annual precipitation (MAP) at each station. The results are presented in Fig. 6 and 7.

A distinct positive relationship was identified between the RMSE and MAP for all four GPPs (Fig. 6). This indicates that the magnitude of estimation error for all products systematically increases in wetter station locations, a common characteristic of GPPs evaluated in areas with diverse topography and climate (Abbas *et al.*, 2024; Ahmed *et al.* 2024). The dependency of RMSE on local climate was most pronounced for ERA5 and CHIRPS. For instance, the RMSE for ERA5 varied significantly with the local climate, increasing from an average of approximately 3.2 mm/day in drier stations (MAP < 250 mm/year) to over 4.5 mm/day in the wettest locations (MAP > 400 mm/year). In contrast, the regression slopes for TRMM and GSMaP were comparatively flatter, indicating that their accuracy is less sensitive to variations in local climate and signifying a more robust performance across the region's diverse precipitation regimes. The analysis of MBE as a function of MAP (Fig. 7) revealed complex and divergent behaviors among the products. For TRMM, its underestimation bias was more pronounced in wetter stations, which may suggest the algorithm struggles to capture the full magnitude of large precipitation events—a known limitation for satellite products when dealing with heavy rainfall (Najafi Tireh Shabankareh, Ziaee and Abedini, 2024).



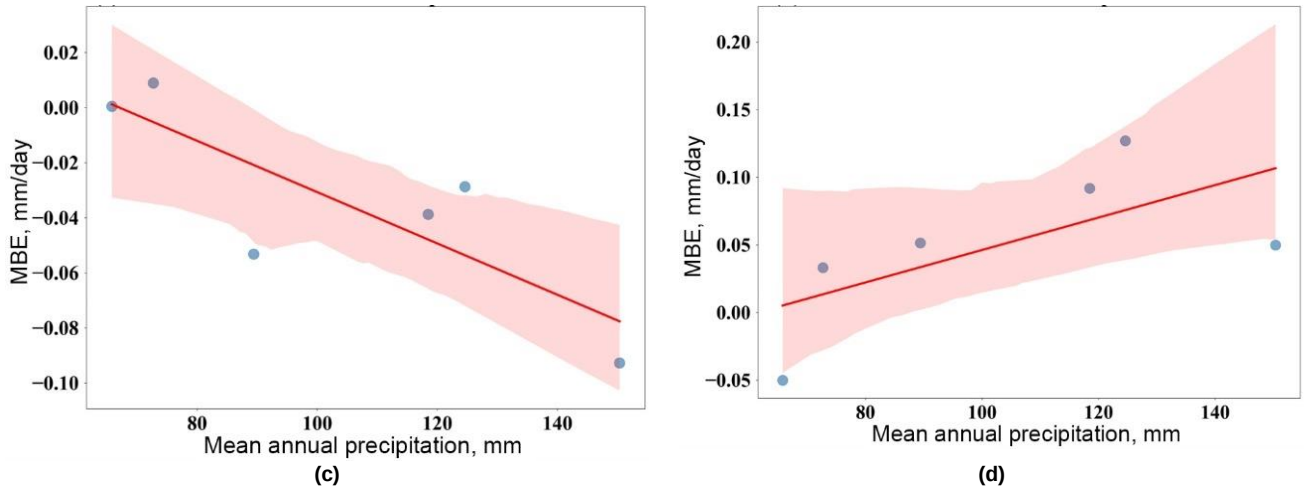


Fig. 6. Relationship between Root Mean Square Error (RMSE) and MAP for each GPP across all stations. Each panel represents a different product: (a) TRMM, (b) CHIRPS, (c) GSMaP, and (d) ERA5. The red line indicates the linear regression fit.

In contrast, ERA5 exhibited a strong positive trend, with its tendency to overestimate becoming much greater in wetter locations. GSMaP demonstrated the most stable performance; its regression line was nearly horizontal, indicating that its slight overestimation tendency was largely independent of the local precipitation climatology. This characteristic makes GSMaP the most consistent product in terms of systematic bias across the diverse climatic gradients of the study area. In summary, this analysis confirms that GPP errors are not spatially random but are systematically linked to the underlying precipitation climatolog

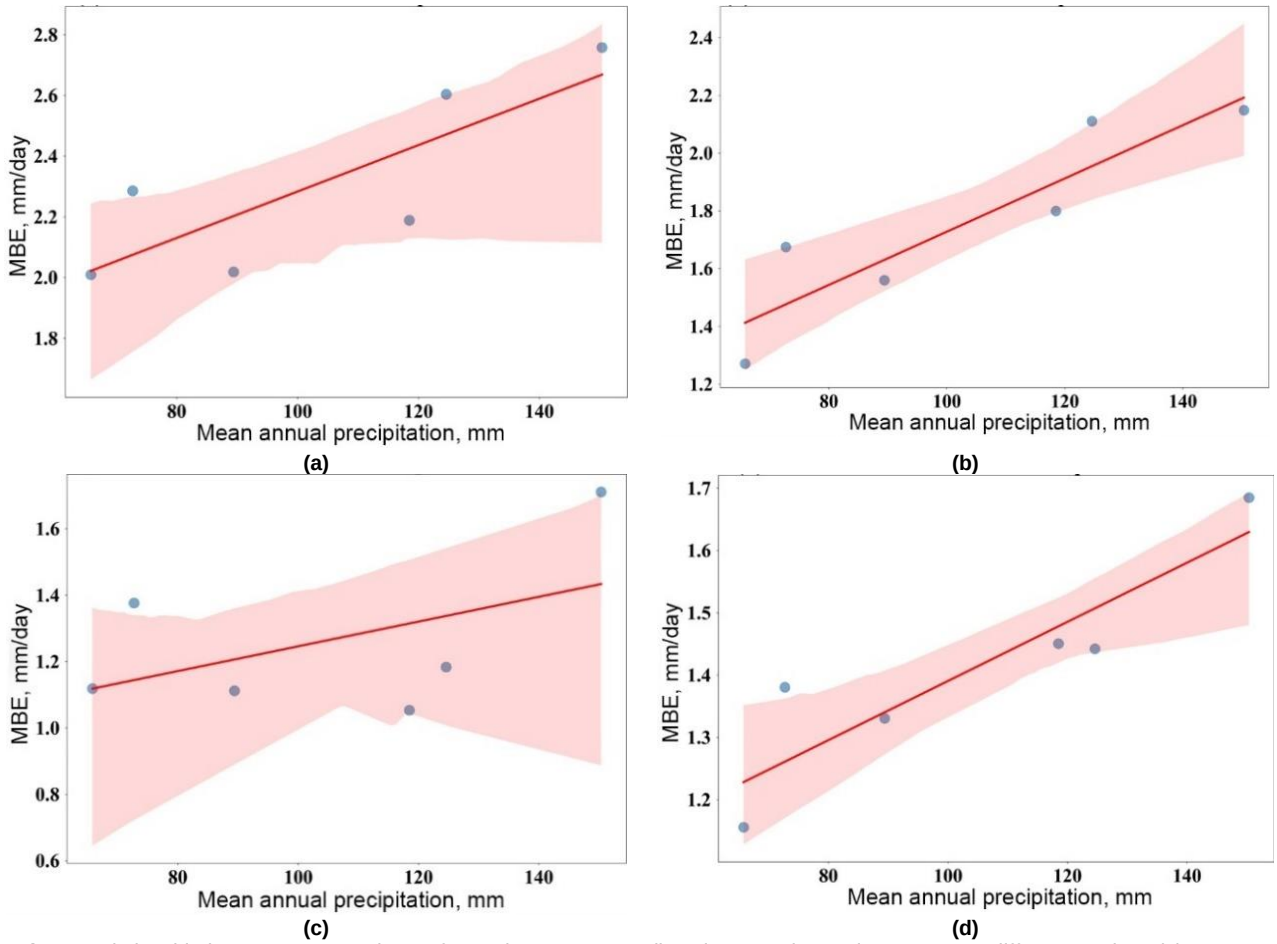


Fig. 7. Relationship between MBE and MAP for each GPP across all stations. Each panel represents a different product: (a) TRMM, (b) CHIRPS, (c) GSMaP, and (d) ERA5. The red line indicates the linear regression fit.

3.6. Overall agreement analysis using density scatter plots

To visually assess the agreement between daily GPP estimates and ground observations, two-dimensional density scatter plots were generated for each product (Fig. 8). These plots illustrate the concentration of data points around the 1:1 identity line, providing a holistic overview of each product's performance.

The results reveal distinct patterns that support the findings from the statistical metrics. For TRMM and GSMaP, the data points are highly concentrated around the 1:1 line, particularly for light to moderate rainfall events (< 20 mm/day). Both products exhibit some data dispersion for more intense events; however, their overall pattern suggests a well-calibrated estimation process, which is consistent with other evaluations of GPM-era products (Gao *et al.*, 2024). For CHIRPS, while data points gathered with reasonable density near the identity line, the spread was visibly wider, allowing a higher degree of random error to be inferred. The product also exhibited a tendency toward underestimating moderate precipitation—a behavior that echoes findings from other semi-arid regions (Al-Shamayleh *et al.*, 2024). The data points for the ERA5 reanalysis product were the most dispersed. A distinct pattern was observed: overestimation occurred for light precipitation (1–10 mm/day), while underestimation was apparent for

high-intensity events (> 30 mm/day). Such behavior is indicative of a known limitation in global reanalysis models, which have difficulty resolving the sub-grid scale processes that control convective rainfall (Gomis-Cebolla *et al.*, 2023; Chen, Collet and Di Luca, 2024; Yan *et al.*, 2024). In summary, the density scatter plots visually confirm these performance differences. The gauge-adjusted satellite products (TRMM and GSMaP) more accurately reproduced the magnitude of daily precipitation events, which underscores the reliability of their estimates. Conversely, the reanalysis data from ERA5 should be interpreted with caution, particularly for applications that are sensitive to the precise magnitude of daily rainfall.

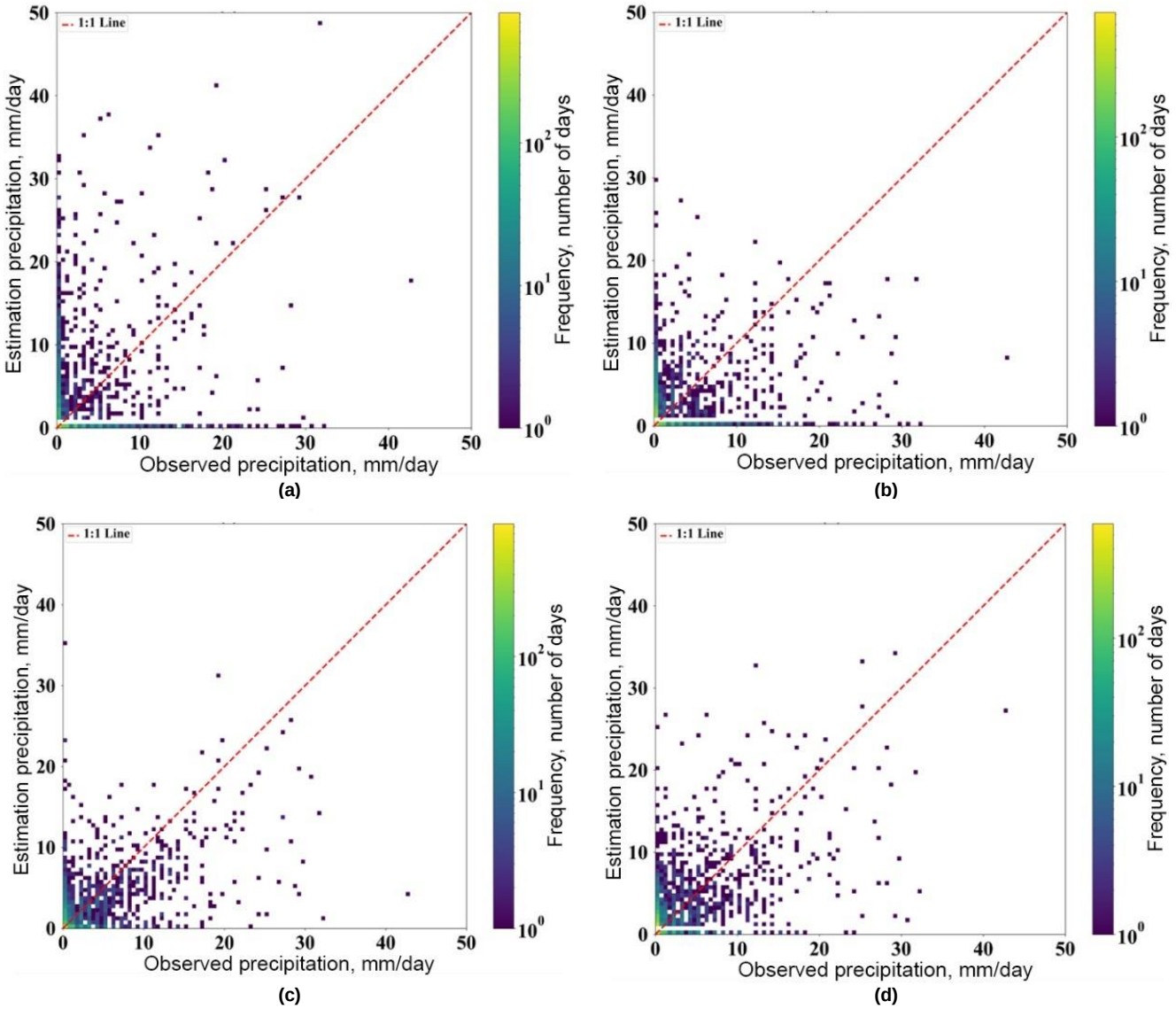


Fig. 8. Two-dimensional density scatter plots comparing daily GPP estimates against ground observations for the 2010–2019 period. Each panel corresponds to a GPP: (a) TRMM, (b) CHIRPS, (c) GSMaP, and (d) ERA5. The color bar indicates the frequency of data points (number of days) on a logarithmic scale. The dashed red line represents the 1:1 identity line.

4. Conclusions

This study evaluated four major GPPs (TRMM, CHIRPS, GSMaP, and ERA5) against ground-based station data in the complex, data-scarce terrain of South Khorasan province, Iran (2010–2019). By employing a variety of metrics across multiple timescales, the analysis aimed to comprehensively identify the strengths and weaknesses of each dataset for hydro-climatic applications. It is also important to acknowledge that the gauge-adjusted nature of the satellite products (TRMM, GSMaP, and CHIRPS) likely gives them an inherent advantage over the unadjusted ERA5 reanalysis product within this specific point-to-pixel validation framework. Our analysis revealed a clear and consistent performance ranking. The satellite-based products, especially the gauge-adjusted TRMM and GSMaP, emerged as the top performers. They exhibited the highest correlation and lowest errors (RMSE and MAE). Their skill in detecting rain events, measured by CSI, was also the most balanced. The performance profile of the ERA5 reanalysis product was markedly different; although it demonstrated the highest sensitivity in detecting the presence of rain, it was characterized by large systematic biases, the greatest random errors, and a high rate of false alarms, rendering it less suitable when precise rainfall amounts are important. The primary conclusion of this study is that the optimal GPP for this semi-arid region depends on the specific application. For tasks requiring accurate daily and monthly rain data, such as hydrological modeling, GSMaP and TRMM are the strongest options. GSMaP, in particular, delivered the best overall results, combining high accuracy with a systematic bias that remained stable across different local climates. The analysis also confirmed that performance is seasonally dependent, with all products functioning less effectively during the dry summer months. These results are practical for researchers and water managers in the region. However, it is crucial to recognize that the sparse gauge network itself imposes a significant limitation on the spatial generalizability of our findings**.** While the point-to-pixel evaluation is sound for the available stations, the performance of GPPs across the vast, ungauged portions of the basin may differ, particularly in areas of complex topography. Therefore, this local validation is essential before these products are used operationally. For future work, we suggest creating multi-product ensembles, possibly with machine learning, that are specifically designed to leverage the complementary strengths identified in this study. For instance, such a framework could combine the high detection sensitivity (POD) of ERA5 with the superior quantitative accuracy and stable bias of GSMaP. The goal is to create a better, high-resolution precipitation dataset for northeastern Iran to improve local climate studies and hydrological forecasting.

Author Contributions

Ali Nasirian: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing Review & Editing, Visualization, Supervision, Project administration.
Mahsa Mardani: Conceptualization, Resources, Writing Draft.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data Availability Statement

The data presented in this study is available on request from the corresponding author.

Acknowledgment

The authors would like to thank the Iran Meteorological Organization for providing the data for this research.

References

Abbas, H. *et al.* (2024) 'Validation of CRU TS v4.08, ERA5-Land, IMERG v07B, and MSWEP v2.8 precipitation estimates against observed values over Pakistan', *Remote Sensing*, 16(24), p. 4803. doi: https://doi.org/10.3390/rs16244803

Ahmed, J. S. *et al.* (2024) 'Evaluation of ERA5 and CHIRPS rainfall estimates against observations across Ethiopia', *Meteorology and Atmospheric Physics*, 136(3), p. 17. doi: https://doi.org/10.1007/s00703-024-01008-0

Akbas, A. and Ozdemir, H. (2024) 'Comparing satellite, reanalysis, fused and gridded (In Situ) precipitation products over Türkiye', *International Journal of Climatology*, 44(16), pp. 5873-5889. doi: https://doi.org/10.1002/joc.8643

Al-Shamayleh, S. *et al.* (2024) 'Performance of CHIRPS for estimating precipitation extremes in the Wala Basin, Jordan', *Journal of Water and Climate Change*, 15(3), pp. 1349-1363. doi: https://doi.org/10.2166/wcc.2024.015

Bekić, D. and Leskovar, K. (2025) 'Evaluating CHIRPS and ERA5 for long-term runoff modelling with SWAT in Alpine headwaters', *Water*, 17(14), p. 2116. doi: https://doi.org/10.3390/w17142116

Chang, Y., Qi, Y. and Wang, Z. (2024) 'Comprehensive evaluation of IMERG, ERA5-Land and their fusion products in the hydrological simulation of three karst catchments in Southwest China', *Journal of Hydrology*: Regional Studies, 52, p. 101671. doi: https://doi.org/10.1016/j.ejrh.2024.101671

Chen, T. C., Collet, F. and Di Luca, A. (2024) 'Evaluation of ERA5 precipitation and 10-m wind speed associated with extratropical cyclones using station data over North America', *International Journal of Climatology*, 44(3), pp. 729-747. doi: https://doi.org/10.1002/joc.8329

Devadarshini, E. *et al.* (2024) 'Spatiotemporal performance evaluation of high-resolution multiple satellite and reanalysis precipitation products over the semiarid region of India', *Environmental Monitoring and Assessment*, 196(11), p. 1006. doi: https://doi.org/10.1007/s10661-024-13152-6

Du, H. *et al.* (2024) 'Evaluating the effectiveness of CHIRPS data for hydroclimatic studies', *Theoretical and Applied Climatology*, 155(3), pp. 1519-1539. doi: https://doi.org/10.1007/s00704-023-04721-9

En-nagre, K. *et al.* (2025) 'Assessment of three satellite precipitation products for hydrological studies in a data-scarce context: Ouarzazate Basin, Southern Morocco', *Natural Hazards Research*, In press, doi: https://doi.org/10.1016/j.nhres.2025.02.008

Fatolahzadeh Gheysari, A. *et al.* (2024) 'Reliability of ERA5 and ERA5-Land reanalysis data in the Canadian Prairies', *Theoretical and Applied Climatology*, 155(4), pp. 4349-4373. doi: https://doi.org/10.1007/s00704-023-04808-3

Feizi, H. and Tosan, M. (2017) 'Saffron yield variability by climatic factors in the northeast of Iran', *Saffron Agronomy and Technology*, 5(1), pp. 3-17. doi: https://doi.org/10.22048/JSAT.2017.43324

Feng, J. *et al.* (2025) 'Multi-scale accuracy assessment of meteorological satellite precipitation products in the dry-hot valley of Jinsha River', *Journal of Hydrology: Regional Studies*, 60, p. 102535. doi: https://doi.org/10.1016/j.ejrh.2025.102535

Ganiyu, H. O. *et al.* (2025) 'Comprehensive evaluation of satellite precipitation products over sparsely gauged river basin in Nigeria', *Theoretical and Applied Climatology*, 156(3), p. 169. doi: https://doi.org/10.1007/s00704-025-05388-0

Gao, R. *et al.* (2024) 'Improvements and limitations of the latest version 8 of GSMaP compared with its former version 7 and IMERG V06 at multiple spatio-temporal scales in mainland China', *Atmospheric Research*, 308, p. 107517. doi: https://doi.org/10.1016/j.atmosres.2024.107517

Gomis-Cebolla, J. *et al.* (2023) 'Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020)', Atmospheric Research, 284, p. 106606. doi: https://doi.org/10.1016/j.atmosres.2023.106606

Gu, H. *et al.* (2024) 'Evaluation of daily and hourly performance of multi-source satellite precipitation products in China's nine water resource regions', *Remote Sensing*, 16(9), p. 1516. doi: https://doi.org/10.3390/rs16091516

Helmi, A. M. and Abdelhamed, M. S. (2023) 'Evaluation of CMORPH, PERSIANN-CDR, CHIRPS v2.0, TMPA 3B42 v7, and GPM IMERG v6 satellite precipitation datasets in Arabian arid regions', *Water*, 15(1), p. 92. doi: https://doi.org/10.3390/w15010092

Hisam, E. *et al.* (2023) 'Comprehensive evaluation of Satellite-Based and reanalysis precipitation products over the Mediterranean region in Turkey', *Advances in Space Research*, 71(7), pp. 3005-3021. doi: https://doi.org/10.1016/j.asr.2022.11.007

Huang, Z. *et al.* (2025) 'How GPM IMERG and GSMaP advance hydrological applications: A global perspective', *Journal of Hydrology*, 661, p. 133514. doi: https://doi.org/10.1016/j.jhydrol.2025.133514

Keune, J. *et al.* (2025) 'ERA5–Drought: Global drought indices based on ECMWF reanalysis', *Scientific Data*, 12(1), p. 616. doi: https://doi.org/10.1038/s41597-025-04896-y

Kumar, P. *et al.* (2025) 'Long-term evaluation of CHIRPS and GSMaP_ISRO rainfall estimates for Indian summer monsoon (2000–2022)', *Remote Sensing Applications: Society and Environment*, 39, p. 101627. doi: https://doi.org/10.1016/j.rsase.2025.101627

Li, Y. *et al.* (2025) 'Coupled statistical analysis and hydrological simulation to evaluate the performance of satellite and reanalysis precipitation products in the Qilian Mountains, Northwest China', *Climate Dynamics*, 63(2), p. 134. doi: https://doi.org/10.1007/s00382-025-07607-w

Majidi, F. *et al.* (2025) 'Evaluation of the performance of ERA5, ERA5-Land and MERRA-2 reanalysis to estimate snow depth over a mountainous semi-arid region in Iran', *Journal of Hydrology: Regional Studies*, 58, p. 102246. doi: https://doi.org/10.1016/j.ejrh.2025.102246

Masood, M. *et al.* (2023) 'Evaluation of satellite precipitation products for estimation of floods in data-scarce environment', *Advances in Meteorology*, 2023, 1685720. doi: https://doi.org/10.1155/2023/1685720

Najafi Tireh Shabankareh, R., Ziaee, P. and Abedini, M. J. (2024) 'Evaluation of IMERG precipitation product over various temporal scales in a semi-arid region of southern Iran', *Journal of Arid Environments*, 220, p. 105102. doi: https://doi.org/10.1016/j.jaridenv.2023.105102

Nozarpour, N., Mahjoobi, E. and Golian, S. (2024) 'Assessment of satellite-based precipitation products in monthly, seasonal, and annual time-scale over Iran', *International Journal of Environmental Research*, 18(5), p. 76. doi: https://doi.org/10.1007/s41742-024-00619-0

Nourani, V. *et al.* (2025) 'Advances in multi-source data fusion for precipitation estimation: remote sensing and machine learning perspectives', *Earth-Science Reviews*, 270, p. 105253. doi: https://doi.org/10.1016/j.earscirev.2025.105253

Prakash, S. and Bhan, S. C. (2023) 'How accurate are infrared-only and rain gauge-adjusted multi-satellite precipitation products in the southwest monsoon precipitation estimation across India?', *Environmental Monitoring and Assessment*, 195(4), 515. doi: https://doi.org/10.1007/s10661-023-11148-2

Rezvani Moghaddam, P. *et al.* (2016) 'Saffron agronomy and technology (book of abstracts: 2013-2016)', *Saffron Agronomy and Technology*, 4(SUPPLEMENT), pp. pp. 1-78. doi: https://doi.org/10.22048/jsat.2016.39250

Salih, W. *et al.* (2024) 'A comprehensive assessment of satellite precipitation products over a semi-arid region: focus on extreme events', *Natural Hazards*, 120(3), pp. 3037-3065. doi: https://doi.org/10.1007/s11069-023-06317-y

Shirmohammadi–Aliakbarkhani, Z. *et al.* (2025) 'Assessing the standardized precipitation index utilizing satellite-based and reanalyzed precipitation products in semi-arid region, Iran', *Journal of the Indian Society of Remote Sensing*, 53, pp. 3393–3407 doi: https://doi.org/10.1007/s12524-025-02152-9

Singh, R., Thakur, D. A. and Mohanty, M. P. (2025) 'Can satellite precipitation products comprehend rainfall extremes over disaster-sensitive mountainous basins? an exhaustive inter-comparison and assessment over Nepal', *Earth Systems and Environment*, 9(2), pp. 1371-1391. doi: https://doi.org/10.1007/s41748-024-00557-z

Sultan, U. *et al.* (2024) 'Statistical assessment of uncertainty and performance of multiple satellite-based precipitation products in capturing extreme precipitation events across Punjab Province, Pakistan', *Journal of Water and Climate Change*, 15(9), pp. 4647-4665. doi: https://doi.org/10.2166/wcc.2024.1105

Taña, E. A. S. *et al.* (2025) 'Assessing the performance of GSMaP and IMERG in representing the diurnal cycle of precipitation in the Philippines during the southwest monsoon season', *Atmospheric Research*, 317, p. 107983. doi: https://doi.org/10.1016/j.atmosres.2025.107983

Tosan, M. *et al.* (2015) 'Evaluation of yield and identifying potential regions for Saffron (Crocus sativus L.) cultivation in Khorasan Razavi province according to temperature parameters', *Saffron Agronomy and Technology*, 3(1), pp. 1-12. doi: https://doi.org/10.22048/jsat.2014.9605

Tosan, M. *et al.* (2025a) 'Enhancing evapotranspiration estimation: a bibliometric and systematic review of hybrid neural networks in water resource management', *Computer Modeling in Engineering and Sciences*, 142(2), pp. 1109-1154. doi: https://doi.org/10.32604/cmes.2025.059373

Tosan, M. *et al.* (2025b) 'Evolution of ensemble machine learning approaches in water resources management: a review', *Earth Science Informatics*, 18(2), p. 416. doi: https://doi.org/10.1007/s12145-025-01911-z

Wang, L., Chen, H. and Li, Z. (2025) 'Multiscale performance of global blended satellite precipitation products over Taiwan', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, pp. 2108-2125. doi: https://doi.org/10.1109/JSTARS.2024.3499910

Wei, L. *et al.* (2025) 'Detection of the 2022 extreme drought over the Yangtze River basin using two satellite-gauge precipitation products', *Atmospheric Research*, 315, p. 107929. doi: https://doi.org/10.1016/j.atmosres.2025.107929

Yan, X. *et al.* (2024) 'Multi-scale evaluation of ERA5 air temperature and precipitation data over the poyang lake basin of China', *Water*, 16(21), p. 3123. doi: https://doi.org/10.3390/w16213123

Zhang, Y. *et al.* (2022) 'Evaluation and comparison of precipitation estimates and hydrologic utility of CHIRPS, TRMM 3B42 V7 and PERSIANN-CDR products in various climate regimes', *Atmospheric Research*, 265, p. 105881. doi: https://doi.org/10.1016/j.atmosres.2021.105881

Zhu, S. *et al.* (2024) 'How has the latest IMERG V07 improved the precipitation estimates and hydrologic utility over CONUS against IMERG V06?', *Journal of Hydrology*, 645, p. 132257. doi: https://doi.org/10.1016/j.jhydrol.2024.132257

Zhu, S. *et al.* (2025) 'Evaluation of IMERG climate trends over land in the TRMM and GPM eras', *Environmental Research Letters*, 20(1), p. 014064. doi: https://doi.org/10.1088/1748-9326/ad984e