



Comparative evaluation of seven precipitation products in rainfall estimation over tropical Peninsular Malaysia

Cia Yik Ng¹ · Faridah Othman¹ · Wan Zurina Wan Jaafar¹ · Sai Hin Lai^{2,3} · Yiwen Mei⁴ · Liew Juneng⁵ · Eugene Zhen Xiang Soo⁶

Received: 9 September 2025 / Accepted: 28 December 2025 / Published online: 30 January 2026
© The Author(s) 2026

Abstract

Accurate quantification of rainfall is essential for climate assessments and disaster risk reduction. This study compares the performance of seven precipitation products in rainfall estimation against a dense rain gauge network over Peninsular Malaysia during 2014–2019. Assessments were conducted at daily and seasonal timescales using both continuous and categorical error metrics to examine spatiotemporal accuracy. The ability of each product to represent rainfall climatology, including the annual cycle and interannual extremes, was also investigated. In point-to-grid comparisons, MSWEP and IMERG exhibit good point-scale accuracy, though substantial spatial inconsistencies persist across all products. IMERG and CMORPH are the most skillful in replicating daily rainfall dynamics based on the modified Kling–Gupta efficiency and demonstrate more balanced categorical performance in detecting rain events. However, weak correlation remains a key limitation across all products. Spatiotemporal analysis reveals higher skills in wetter regions but lower skills in drier regions across all timescales, suggesting a magnitude-dependent error structure. Most products fail to replicate seasonal shifts in rainfall distribution, especially for lower and moderate intensity classes. While rainfall climatology is misrepresented in terms of magnitude, temporal variability is reasonably tracked. Overall, IMERG emerges as the best-performing product. The collective findings shed light on the reliability of precipitation products in estimating tropical rainfall for water resources management purposes. Future work should quantify uncertainty propagation in rainfall–runoff processes and characterize cross-correlation structures between rainfall and atmospheric drivers.

Keywords Satellite precipitation products · Reanalysis datasets · Uncertainty analysis · Kling–Gupta efficiency · Categorical metrics · Tropical monsoon cycle

✉ Faridah Othman
faridahothman@um.edu.my; faridah_ce@yahoo.com

Cia Yik Ng
ng.ciaiyik@gmail.com

Wan Zurina Wan Jaafar
wzurina@um.edu.my

Sai Hin Lai
shlai@unimas.my

Yiwen Mei
meiyw3@mail.sysu.edu.cn

Liew Juneng
juneng@ukm.edu.my

Eugene Zhen Xiang Soo
esoo@ums.edu.my

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

² Department of Civil Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

³ UNIMAS Water Centre (UWC), Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

⁴ Carbon-Water Observation and Research Station in Karst Regions of Northern Guangdong, School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

⁵ Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

⁶ Civil Engineering Programme, Faculty of Engineering, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia

1 Introduction

Accurate rainfall measurements across various spatio-temporal scales are essential for assessing the evolving patterns of climate extremes and facilitating disaster prevention. Lack of reliable, timely weather information hinders accurate prediction of floods and droughts, shortening disaster preparation. However, rainfall measurement can be extremely challenging due to its dynamic nature, as it varies in space and time. In practice, rainfall is usually estimated through direct gauge measurements, satellite retrievals and reanalysis products, each with distinct advantages and limitations. Rain gauges are direct measurements of rainfall, but they are point-scale measurements with limited space–time coverage over the globe and are subjected to instrumental under-catch, especially over mountainous and coastal regions (Kidd et al. 2017; Rana et al. 2015; Xie and Arkin 1996). Considering the high spatial heterogeneity of rainfall, measurements from point-scale gauges often fail to reflect meteorological conditions of broader areas. This situation is particularly noticeable in tropical regions, where the microclimate is highly variable due to convection-dominated processes (Meteorological Service Singapore 2015).

To overcome limitations in gauge measurement, satellite and reanalysis precipitation products have been developed, benefitting from innovations in numerical weather modeling, satellite retrieval and data merging (Gebremichael 2010; Maggioni et al. 2016; Sun et al. 2018). These precipitation products mitigate spatial coverage gaps inherent in rain gauge networks. However, they differ significantly in latency, retrieval algorithms, data sources, record length and spatiotemporal resolutions, leading to varying estimation skills (Beck et al. 2019a, 2019b). In addition, the choice of a precipitation product for a specific application largely depends on its design objectives. For example, the Integrated Multi-satellite Retrievals for the Global Precipitation Measurement Mission (IMERG) was developed as a high-resolution product that provides short-interval estimates. To achieve this, IMERG relies on a continuously changing constellation of satellite inputs to offer the best instantaneous precipitation estimation (Huffman et al. 2020). However, the changes in the constellation may cause variations in the products and introduce non-climatic shift into the time series, thus may not be suitable for climate-scale studies that are focusing on climate change signals, trends, interannual variability or extremes. In contrast, while Global Precipitation Climatology Centre dataset has a coarser resolution compared to IMERG, it emphasizes spatiotemporal homogeneity and climate-scale consistency over instantaneous accuracy, making it more suitable than IMERG for long-term analysis (Huffman et al. 2023).

Previous evaluations of precipitation products in Peninsular Malaysia have yielded mixed results. Several studies reported reasonable agreement between precipitation products and rain gauge observations, particularly during the wetter monsoon periods (Ayoub et al. 2020; Tan et al. 2015). However, substantial uncertainties remain in precipitation products even after bias adjustment (Mohd Zad et al. 2018; Soo et al. 2022), which can propagate into subsequent rainfall-runoff processes and affect hydrologic predictability (Mei et al. 2017; Seo et al. 2019). This underscores the need to improve their input data quality and retrieval algorithms. The release of the IMERG product represented a major advancement over its predecessor with improved detection of light rainfall events due to enhanced microwave imaging capabilities (Tan and Santo 2018). Yet, evidence from other tropical regions suggests that IMERG still performs poorly compared to the artificial neural network-based Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks—Climate Data Record (PERSI-ANN-CDR), particularly in sparsely gauged or data-scarce catchments (Ganiyu et al. 2025).

Despite valuable insights from past studies, most evaluations in Malaysia have been limited to satellite or gauge-based products, with little attention given to reanalysis datasets that may exhibit greater skills than satellite-based products near orography and in cold season environments (Beck et al. 2017; Prein et al. 2015). For example, the earlier-generation Princeton Global Forcings (PGFv3) reanalysis product performed poorest against the satellite-based Global Satellite Mapping of Precipitation (GSMaP) and the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) blended dataset (Ayoub et al. 2020), raising concerns about the reliability of reanalysis product in Peninsular Malaysia. While newer reanalysis products such as the European Centre for Medium-Range Weather Forecasts fifth generation of European ReAnalysis (ERA5) offer improvements over their predecessors, recent evidence suggests limitations remain when compared with gauge-based datasets for drought estimation (Du et al. 2025). However, their performance on rainfall estimation over Peninsular Malaysia remains unclear. Apart from reanalysis products, emerging high-resolution, multi-source products such as the Multi-Source Weighted-Ensemble Precipitation (MSWEP), which merges gauges, satellites and reanalysis datasets, have shown promising skill in rainfall estimation and hydrological applications at the global scale (Beck et al. 2017, 2019a). However, MSWEP has yet to be systematically assessed over Peninsular Malaysia. The absence of comprehensive validation of these emerging reanalysis and merged precipitation products against dense rain gauge networks represents a critical gap in understanding their applicability in convection-dominated tropical Peninsular Malaysia.

Given the considerable knowledge gaps, this study intends to shed light on the reliability of precipitation products over tropical Peninsular Malaysia by addressing several pertinent questions: (a) How do satellite, reanalysis and blended precipitation products compare in estimating rainfall across Peninsular Malaysia at daily and seasonal timescales? (b) How does their skills vary at different climatic regions? (c) Are they able to capture the climatology in Peninsular Malaysia, including annual cycle and inter-annual extremes? In the following sections, the evaluation is conducted by comparing seven precipitation products against gauge observations over Peninsular Malaysia at daily and seasonal timescales using both continuous and categorical error metrics. Their performances in capturing annual cycle and interannual extremes have also been further investigated.

2 Data and methods

2.1 Study area

Peninsular Malaysia forms the western portion of Malaysia, extending southward from the mainland Asia towards Singapore. It is separated from East Malaysia on the island of Borneo by the South China Sea. Geographically, Peninsular Malaysia has a tongue-shaped landmass situated between 1°N and 7°N latitude as well as 99.5°E and 104.5°E longitude (Fig. 1). In terms of topography, most part of the 132,090 km² land surface area has relatively flat terrains, except for the interior land mass with orographic features. Peninsular Malaysia experiences a tropical rainforest climate, characterized by consistently high annual rainfall, relative humidity and stable surface temperature throughout the year (Ng et al. 2024). With an average of six hours of sunshine daily, the interior land of Peninsular Malaysia is covered with thick and dense vegetation that hosts a diverse species of flora and fauna in its tropical rain forest.

The regional rainfall distribution in Peninsular Malaysia is heavily influenced by the tropical monsoon system, predominantly orchestrated by the expansion and contraction of the Siberian high-pressure system (Ding 1990; Wu and Chan 1995). The northeast monsoon (NEM) season (November to March) is divided into wet and dry phases (Tan and Santo 2018). The wet phase (November to early January) frequently brings widespread heavy rainfall to the eastern part of Peninsular Malaysia, while the dry phase (late January to March) is characterized by drier conditions, particularly in the north. Based on the temporal variations of domain-averaged daily rainfall over Peninsular Malaysia from 2014 to 2019, several rainfall peaks are clearly identifiable in the time series, notably in December 2014,

January 2017, late 2017/early 2018, late 2018/early 2019 and December 2019 (Fig. S1; Supplementary Materials). The timing of these peaks aligns with historical flood events documented by the Malaysia Department of Irrigation and Drainage. The highest domain-averaged single-day rainfall (00:00 to 23:59 UTC) was recorded during the January 2017 event (~54 mm), followed by events in December 2014 (~51 mm), December 2019 (~41 mm), 2017/2018 (~38 mm) and 2018/2019 (~31 mm). All of these severe events occurred during the wet phase of NEM season.

The southwest monsoon (SWM) season (May to September) is a drier period compared to the NEM season, characterized by relatively low rainfall totals (Table S1; Supplementary Materials). The inter-monsoon (IM) seasons, occurring in April and October, are transitional periods between the major monsoons. They are marked by weak monsoonal winds and are dominated by localized convective activity (Mohd Akhir et al. 2014), bringing short-lived but intense rainfall events across the region. NEM rainfall is frontal-dominated, while SWM and IM rainfall is mainly convective and thermodynamically driven (Jamaluddin et al. 2018; Juneng and Tangang 2010). Given that the climate regime in Peninsular Malaysia is heavily modulated by the tropical monsoon system, it is important to evaluate the performance of precipitation products across various monsoon seasons. To reflect these regional rainfall characteristics, Peninsular Malaysia is divided into five main climatic regions (Bakar et al. 2020; Lim 1976), namely the north, east, southwest, south and central regions. The east region has wetter conditions with higher mean annual rainfall (during the study period), while the southwest and south regions are drier than average with lower annual rainfall (Fig. 1).

2.2 Gauge observations and precipitation products

Daily rainfall records aggregated from 00:00 to 23:59 UTC were obtained from the Malaysia Department of Irrigation and Drainage rain gauge network, which consists of 626 stations across Peninsular Malaysia. Station density was relatively sparser in mountainous terrains (e.g., central region, eastern interior) than in coastal areas. The study period was set from 2014 to 2019 (6-year period), which offers the most complete coverage across precipitation products. Quality control followed Ng et al. (2022): (a) outliers were retained or replaced as missing based on documented historical events; (b) gauging stations with > 10% of missing data were discarded; (c) missing values in the remaining 408 stations (Fig. 1) were imputed using data from the nearest reliable stations, prioritizing similar elevations; (d) temporal homogeneity was assessed for all series to ensure consistency (Wang et al. 2007; Wang 2008).

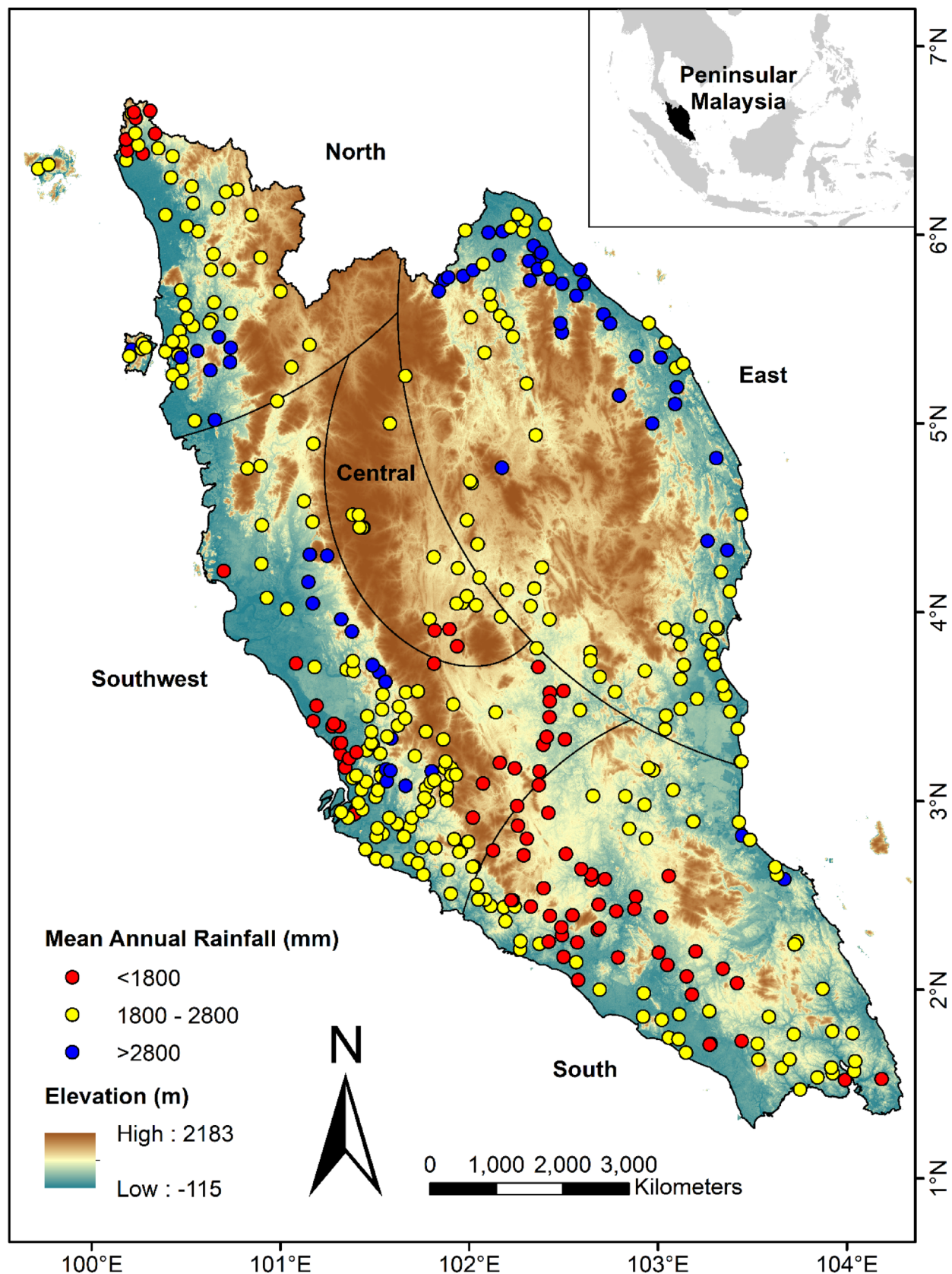


Fig. 1 Distribution of rain gauges overlaying on the elevation map of Peninsular Malaysia. Solid lines represent the boundaries of the five climatic regions

Table 1 List of gauge observations and seven precipitation products evaluated in this study

Abbreviation	Name	Data Sources	Spatial resolution	Temporal resolution	Reference
Observations	Ground rain gauge network	Gauge	Point	Daily	Malaysia Department of Irrigation and Drainage
CMORPH	Climate Prediction Center morphing technique (CMORPH) bias corrected (CRT) V1.0	Gauge, Satellite	8 km (~0.07°)	30 min	Xie et al. (2017)
GSMaP	Global Satellite Mapping of Precipitation gauge-adjusted (GSMaP_Gauge) standard product V7	Gauge, Satellite	0.10°	Daily	Kubota et al. (2020)
IMERG	Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG) Final Run V06	Gauge, Satellite	0.10°	Daily	Huffman et al. (2020)
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks—Climate Data Record (PERSIANN-CDR)	Gauge, Satellite	0.25°	Daily	Ashouri et al. (2015)
ERA5-Land	European Centre for Medium-Range Weather Forecasts (ECMWF) fifth generation of European ReAnalysis (ERA5) high resolution realisation (HRES) land component	Reanalysis	9 km (~0.08°)	Hourly	Muñoz-Sabater et al. (2021)
CHIRPS	Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) V2.0	Gauge, Satellite, Reanalysis	0.05°	Daily	Funk et al. (2015)
MSWEP	Multi-Source Weighted-Ensemble Precipitation (MSWEP) V2.8	Gauge, Satellite, Reanalysis	0.10°	Daily	Beck et al. (2019b)

Table 2 Daily rainfall intensity classes

Category	Rainfall (mm/day)
Negligible rain	0–1
Light rain	1–2
Low moderate rain	2–5
High moderate rain	5–10
Low heavy rain	10–20
High heavy rain	20–50
Very heavy rain	> 50

Seven state-of-the-art precipitation products were evaluated (Table 1). These precipitation products feature global or quasi-global coverage, high spatial resolutions and considerable latency (Ashouri et al. 2015; Beck et al. 2019b; Funk et al. 2015; Huffman et al. 2020; Kubota et al. 2020; Muñoz-Sabater et al. 2021; Xie et al. 2017). They may be categorized into three groups according to the data sources: (a) gauge-adjusted satellite precipitation products: (CMORPH, GSMaP, IMERG and PERSIANN-CDR); (b) precipitation

reanalysis (ERA5-Land); (c) merged precipitation products (CHIRPS and MSWEP). All products were aggregated to daily resolution (00:00 to 23:59 UTC). The daily rainfall intensity was categorized into seven classes for the probability distribution functions (PDFs) analysis (Table 2) (Tan et al. 2015).

2.3 Error metrics

The evaluation of precipitation products against the gauge observations was conducted at daily scale. Native-resolution grids of the products collocated with gauges were extracted for comparison using continuous and categorical error metrics. For continuous error metrics, the modified Kling–Gupta efficiency (KGE) was applied to quantify the uncertainty of these precipitation products (Kling et al. 2012). KGE integrates the correlation coefficient (reflecting temporal error), bias ratio (reflecting systematic error in distribution mean) and variability ratio (reflecting error

in distribution of rainfall intensity) without interdependence among these components:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \quad (1)$$

$$\beta = \frac{\mu_s}{\mu_o}, \quad (2)$$

$$\gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}, \quad (3)$$

where r is the linear correlation coefficient between the rainfall estimates (s) and the gauge observations (o); β is the bias ratio of the rainfall means, μ ; γ is the ratio between the estimated and the observed rainfall variability; σ represents the standard deviation. KGE and its three error components (r , β and γ) are dimensionless, with an optimal value of 1. Additionally, the normalized root mean square error (NRMSE) was used to quantify the residuals between observed and estimated rainfall series. It is computed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (S_i - O_i)^2}{N}}, \quad (4)$$

$$NRMSE = \frac{RMSE}{\bar{O}}, \quad (5)$$

where N is the total number of data points; O_i and S_i are the i -th observed and estimated daily rainfall, respectively; $\bar{O}$ is the mean observed daily rainfall. A lower NRMSE indicates a better fit between the precipitation products and gauge observations, with 0 being the optimum value.

Apart from the continuous error metrics, five detection-based categorical metrics, namely accuracy (ACC), probability of detection (POD), false alarm ratio (FAR), critical success index (CSI) and Heidke skill score (HSS), were applied to measure the detection error of rainfall event (Mei et al. 2020; Soo et al. 2018). These categorical metrics are commonly applied as diagnostic verification measures for assessing the quality of non-probabilistic forecasts or estimations (Wilks 2011). They were computed based on four possible outcomes of forecast–event combinations. Among these categorical metrics, ACC reflects level of correspondence between estimated rainfall event and observations; POD measures the hit rate of correct detections in relation to the total actual occurrences of wet days; FAR represents the proportion of the estimated rainfall event associated with non-occurrences; CSI calculates the critical success rate that integrates both POD and FAR without considering correct non-occurrence of rainfall event (correct non-alarm); HSS

provides the measure of overall capability of precipitation products in event detection without factoring in the probability of correct detection by chance. These categorical metrics are derived as:

$$ACC = \frac{A + D}{n}, \quad (6)$$

$$POD = \frac{A}{A + C}, \quad (7)$$

$$FAR = \frac{B}{A + B}, \quad (8)$$

$$CSI = \frac{A}{A + B + C}, \quad (9)$$

$$HSS = \frac{2(A \cdot D - B \cdot C)}{(A + C) \cdot (C + D) + (A + B) \cdot (B + D)} \quad (10)$$

where A , B , C and D are the number of hit, false alarm, missing and correct non-alarm of rainfall event conditioned on the 1 mm/day threshold; n is the total number of possible event combinations ($A+B+C+D$). ACC, POD, FAR and CSI have their values range from 0 to 1, with 1 being the optimum for ACC, POD and CSI and 0 being optimal for FAR. The value of HSS ranges from $-\infty$ to 1, with 1 being the optimum value, zero HSS value represents equivalent performance with observations and negative HSS value indicates that the estimation is worse than the observations.

2.4 Evaluation of climatology and extreme indices

To further explore the performance of precipitation products in estimating the climatology in Peninsular Malaysia, the annual cycles of three climate indicators, namely wet day frequency (FREQ), rainfall intensity during wet day (SDII) and 90th percentile of rainfall during wet day (Q90) (Table 3), were computed using time series of both gauge observations and precipitation products following Ngai et al. (2020). These three indicators can provide insights on the suitability of each precipitation product to be applied in future climate studies. Additionally, the potential application of precipitation products in estimating interannual precipitation extremes was also investigated by computing four extreme indices that reflect the frequency and intensity of extreme rainfall (World Meteorological Organization 2009) (Table 3). These extreme indices, recommended by the former World Meteorological Organization's Expert Team on Climate Change Detection and Indices, are widely used as quantitative indicators in decision-making processes due to their simplicity and clarity, which facilitate the monitoring

Table 3 Definitions of selected indicators for assessing annual cycle and interannual precipitation extremes. A wet/dry day is defined when daily rainfall $\geq / < 1$ mm

Index	Definitions	Units
<i>Annual cycle</i>		
FREQ	Monthly count of wet days	days
SDII	Intensity of rainfall – the ratio of monthly total rainfall to the number of wet days	mm/day
Q90	90th percentile of rainfall in wet days	mm
<i>Interannual extremes</i>		
PRCPTOT	Annual total rainfall from wet days	mm
Rx5day	Annual maximum consecutive 5-day rainfall	mm
R95p	Annual total rainfall from days > 95th percentile	mm
R20mm	Annual count of days when rainfall ≥ 20 mm	days

of trends in extreme frequency and intensity (IPCC 2021; Zhang et al. 2011). The selected indices represent moderate extremes that typically occur several times per year (Dunn et al. 2022; Tefera et al. 2024; Zhang et al. 2011), rather than rare, high-impact events with decadal return periods (Alexander et al. 2025). The four indices used in this study include three intensity-based indices (PRCPTOT, Rx5day and R95p) and one frequency-based index (R20mm) (Ng et al. 2022).

3 Results

3.1 Daily and seasonal performance

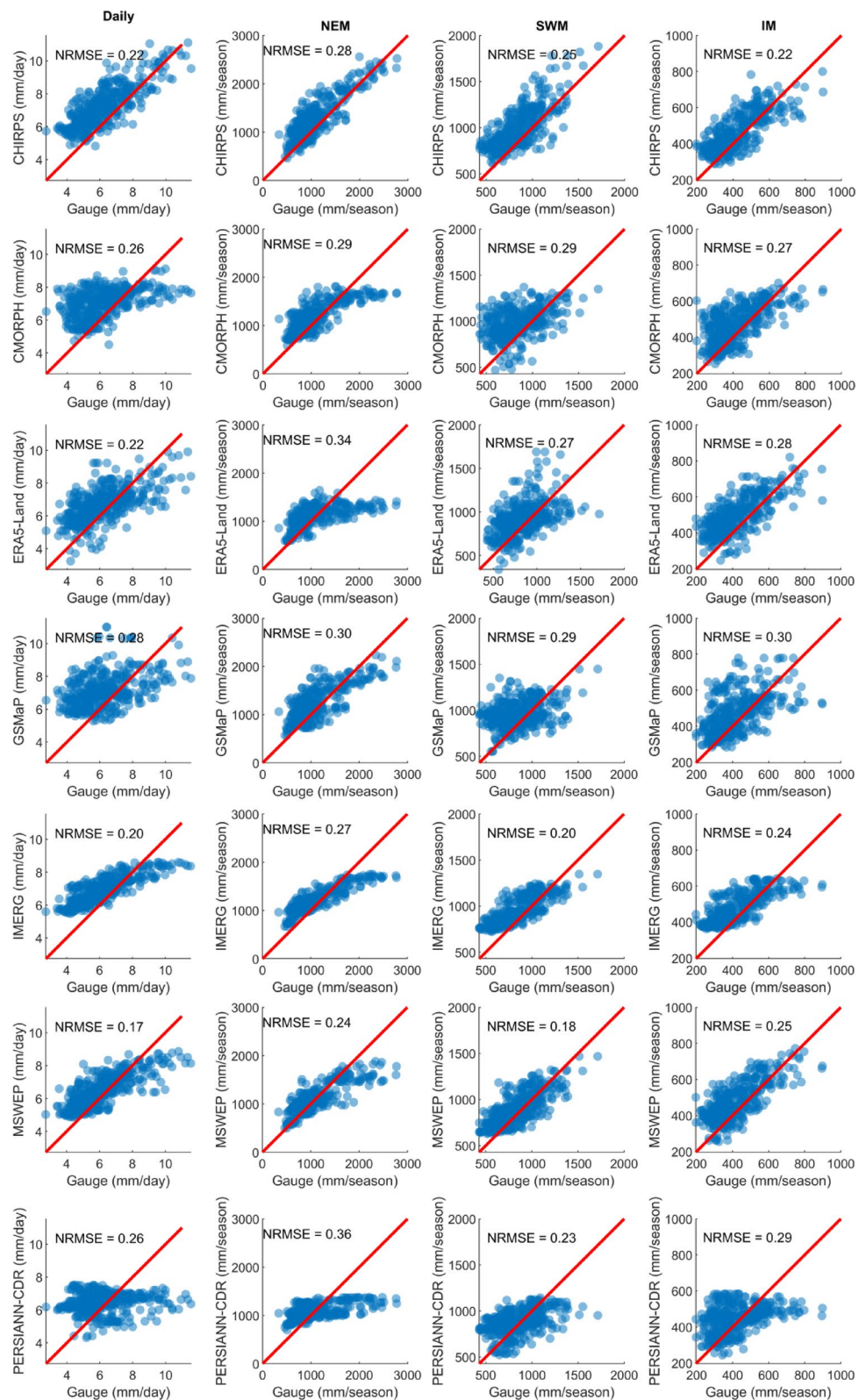
Figure 2 compares precipitation products in capturing point-scale rainfall. MSWEP exhibits the smallest residual errors with the lowest NRMSE (0.17), followed by IMERG (0.20), CHIRPS (0.22), ERA5-Land (0.22), CMORPH (0.26), PERSIANN-CDR (0.26) and GSMaP (0.28). The spread patterns show that all products underestimate rainfall at wetter stations and overestimate at drier ones, producing a flattened distribution relative to the 1:1 line. MSWEP and IMERG have tighter clustering around the 1:1 line, reflecting higher accuracy and spatial consistency. CHIRPS performs best at wet stations, likely due to its finer resolution ($0.05^\circ \times 0.05^\circ$) with better representation of localized rainfall. In contrast, CMORPH, ERA5-Land, GSMaP and PERSIANN-CDR display substantial scatter and poor alignment with the 1:1 line, indicating inaccuracies across most of the stations, which is an important consideration when applying these datasets in regional or local-scale studies. These results highlight the importance of spatial analysis, as the performance of precipitation products is highly location dependent in Peninsular Malaysia.

For seasonal analysis, precipitation products exhibit the highest accuracy during the SWM season, except for CHIRPS and CMORPH. However, scatter patterns reveal poor spatial consistency during the SWM and IM seasons. In contrast, data points during the NEM season are more tightly clustered, despite obvious underestimation of rainfall at wet stations. MSWEP achieves the highest accuracy during the NEM and SWM seasons, while CHIRPS is the most accurate during the IM season. IMERG ranks as the second-best product across all seasons, with NRMSE ranges from 0.20 to 0.27. Despite performing well during the SWM season, ERA5-Land and PERSIANN-CDR show poor accuracy during the NEM season (NRMSE > 0.30). The results highlight varying accuracy of precipitation products under different meteorological conditions.

The skills of precipitation products in estimating the rainfall dynamics, such as temporal variations, systematic biases and rainfall intensity distribution, are further examined using KGE (Fig. 3). The median KGE scores at the daily scale range from 0.12 to 0.45, with IMERG showing the highest score, followed by CMORPH, MSWEP and GSMaP. In contrast, CHIRPS, ERA5-Land and PERSIANN-CDR show lower KGE. IMERG's violin plot distribution is more tightly clustered and skewed toward higher values, indicating better consistency and reduced spatial variability. GSMaP, however, displays the widest spread of violin plot, suggesting higher spatial inconsistency.

To better understand performance across different error dimensions, KGE can be decomposed into three components: correlation, bias and variability ratio. With largest deviation from the optimum, correlation is the weakest component across all products (Table S2; Supplementary Materials). KGE rankings closely mirror those of correlation, suggesting that low daily KGE are largely attributed to temporal mismatches between product estimates and gauge observations. Only IMERG and MSWEP achieve median correlation values > 0.50 (0.55 and 0.54, respectively), while others range from 0.32 to 0.47. MSWEP's relatively high correlation is likely due to its adjustment for gauge reporting times, which reduces temporal misalignment (Beck et al. 2019a). However, the overall poor correlations imply that, despite recent advances, current generation of precipitation products still struggle to capture day-to-day rainfall variations in tropical region. Therefore, improving temporal estimation is crucial to enhance the KGE scores. For bias, all products show reasonable performance, with median bias ratios ranging from 1.08 to 1.17, within ± 0.2 of the optimum. Small biases are expected, as most products are bias-corrected using global gauge data. Notably, the unadjusted ERA5-Land performs on par with other bias-corrected datasets. All precipitation products show greater variability errors compared to bias, with values ranging from 0.52 to

Fig. 2 Scatter plot of mean rainfall at each station for daily and seasonal timescales. Red line represents the 1:1 line of perfect agreement



0.93. They underestimate the rainfall variability ratio, with CMORPH showing the best performance (0.93), followed by IMERG (0.77) and GSMaP (0.74). ERA5-Land, PERSIANN-CDR, CHIRPS and MSWEP exhibit greater errors

in variability (0.52–0.60), likely due to overestimation of rainfall frequency (Beck et al. 2017, 2019a).

At the seasonal scale, all products show higher KGE during the NEM season than daily scale with similar

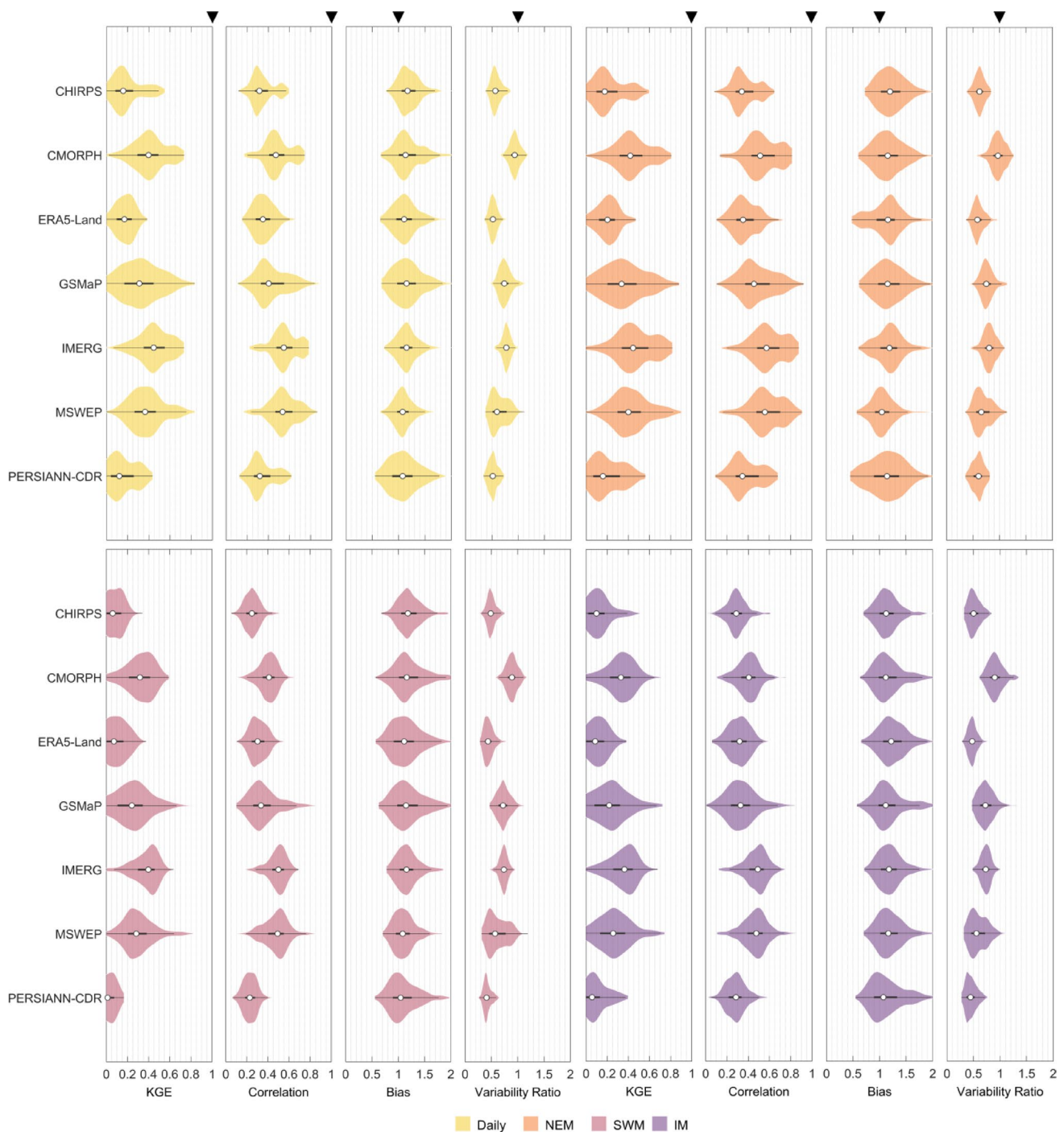


Fig. 3 Violin plots of KGE scores and their error components (correlation, bias and variability ratio) at daily and seasonal timescales over Peninsular Malaysia during 2014–2019. Inverted black triangles at the top indicate the optimum value for each metric

performance rankings, whereby MSWEP, CMORPH and IMERG perform better than other products. These gains are driven by increased correlation (up to +0.06) and variability ratio (up to +0.09) scores, suggesting enhanced skill in capturing temporal dynamics and intensity distribution during the wet season. However, the violin plot distributions of KGE are flatter during this period, indicating greater

spatial variability in KGE scores. In contrast, KGE scores decline during the SWM and IM seasons, particularly in the dry SWM season, suggesting that low daily KGE values are caused by poor performance during dry periods. All error components show lower median values in these seasons, reflecting weaker temporal agreement, greater bias and reduced variability skill. The results indicate that

although precipitation products achieve higher accuracy in estimating seasonal totals during the SWM season (Fig. 2), the temporal correlations, systematic biases and rainfall intensity distribution are more poorly captured than in the NEM season (Fig. 3), revealing clear seasonal differences in performance.

Figure 4 presents the categorical performance of the precipitation products in detecting rain events. Large discrepancies in performance are observed among the products for POD, FAR and HSS, whereas their ACC and CSI scores are relatively consistent (Table S3; Supplementary Materials). For ACC, all products demonstrate reasonably good accuracy (~ 0.60), comparable to the gauge-based Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation (APHRODITE) dataset reported in a previous study over Peninsular Malaysia (Tan et al. 2015). The comparable ACC values suggest that all products have

a similar capability in detecting both wet and dry days. In terms of POD, all products perform well with values exceeding 0.70 (Tan et al. 2015), though substantial differences are evident between the best and worst performers. ERA5-Land achieves the highest POD score (0.95), likely due to its dependency on input components that frequently yield nonzero estimates (Hersbach et al. 2020), thereby reducing missed events (Ma et al. 2022). Its POD is slightly higher than that of MSWEP, PERSIANN-CDR, IMERG and GSMaP, while CHIRPS and CMORPH exhibit lower POD values (< 0.80). However, ERA5-Land's tendency to produce nonzero precipitation estimates also leads to rounding bias and an increased risk of false alarms. As a result, it ranks lowest in FAR, along with PERSIANN-CDR (both at 0.52). CMORPH achieves the best FAR score (0.38), while the remaining products fall between 0.40 and 0.50.

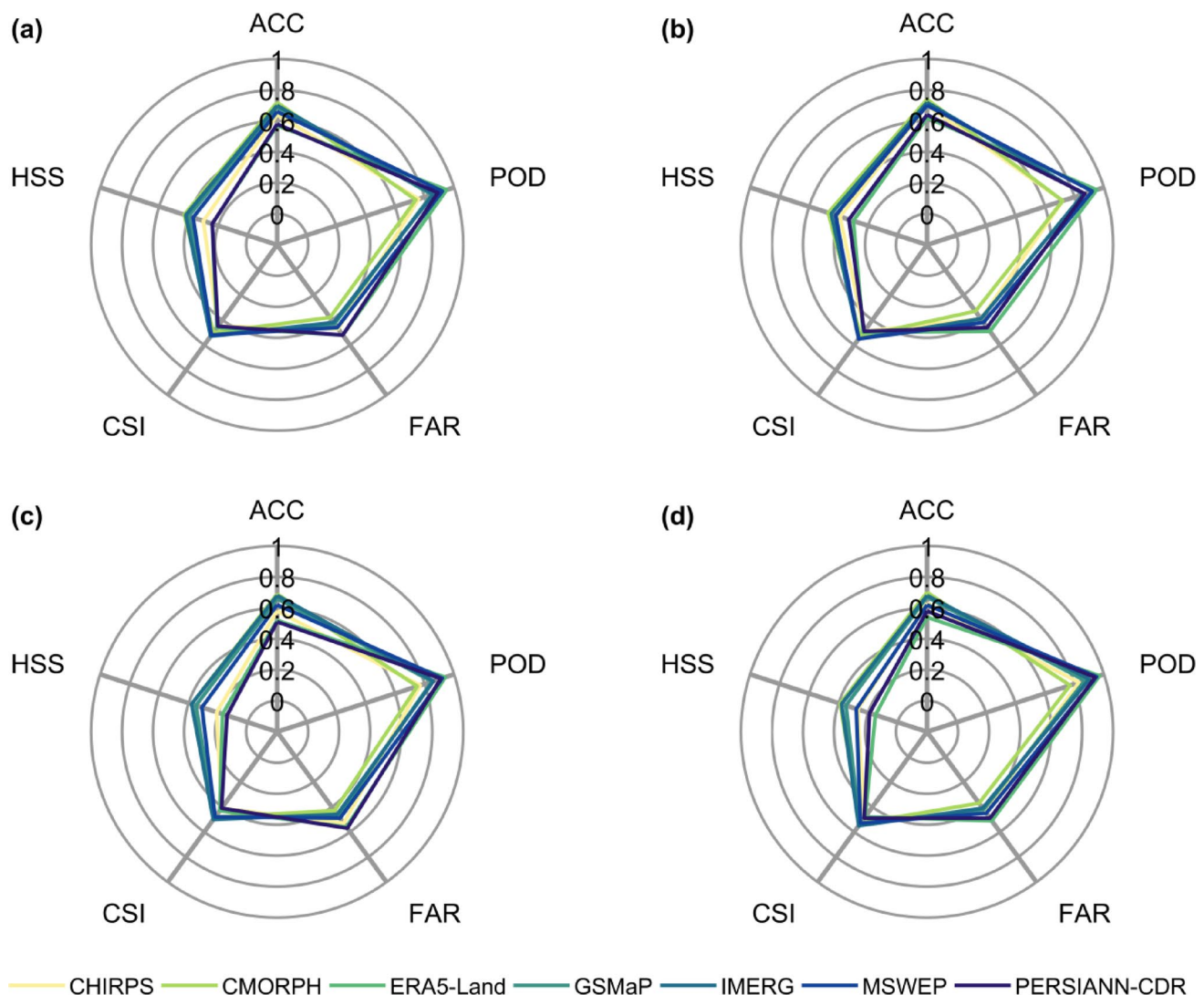


Fig. 4 Radar chart of five categorical metrics (ACC, POD, FAR, CSI and HSS) that measure the capability of the seven precipitation products in rain event detection for **a** daily timescale, **b** NEM season, **c** SWM season and **d** IM season

Given the inconsistencies observed in ACC, POD and FAR, the CSI is applied as a more comprehensive metric that integrates both POD and FAR to better evaluate the performance of precipitation products in detecting rain events. CSI accounts for both missed events and false alarms while excluding correct non-events, offering measure of detection capability without being dominated by correctly predicted non-rain days. The radar chart shows that all precipitation products perform similarly poor in terms of CSI, with values ranging from 0.45 to 0.53. This reflects inadequate skill in detecting rainfall events relative to the frequency of detection errors. Another composite metric used is the HSS, which is widely used to measure the accuracy of a forecast compared to the accuracy expected from random chance. All precipitation products demonstrate weak performance in HSS, with values ranging from 0.24 to 0.43, indicating only modest improvement over random estimations. Among the products, CMORPH, IMERG, GSMaP and MSWEP achieve relatively higher HSS scores than CHIRPS, PERSIANN-CDR and ERA5-Land, suggesting better overall performance in rain event detection.

For seasonal categorical performance, all precipitation products show better ACC, FAR, CSI and HSS during the NEM season, whereas poorer performance is observed during the SWM and IM seasons. These results indicate that seasonal variations in categorical performance are consistent with continuous error metrics, where precipitation

products generally perform better during the wet season and worse during the dry season. However, the seasonal pattern of POD differs from the other four categorical metrics. Specifically, POD scores tend to improve during the SWM and IM seasons but decrease during the NEM season. This may be attributed to the increased likelihood of missed events during the wet season, when there are more rainy days and a reduced likelihood during the dry season, when wet days are less frequent. Despite accounting for missed events, both CSI and HSS incorporate a broader combination of event pairs, thereby offering a more balanced and comprehensive assessment of categorical performance that is not overly influenced by a large number of missed events. These findings highlight that a lower score in a simple categorical metric, such as POD, does not necessarily imply poor overall performance. Thus, the result needs to be interpreted cautiously.

3.2 Geospatial analysis

To assess the performance of precipitation products across different climatic regions, regional mean KGE scores were computed for the north, east, southwest, south and central regions at daily and seasonal timescales (Fig. 5). Among all products, IMERG exhibits the highest KGE scores in the east, southwest, south and central regions, while MSWEP performs best in the north (Fig. S2 and Table

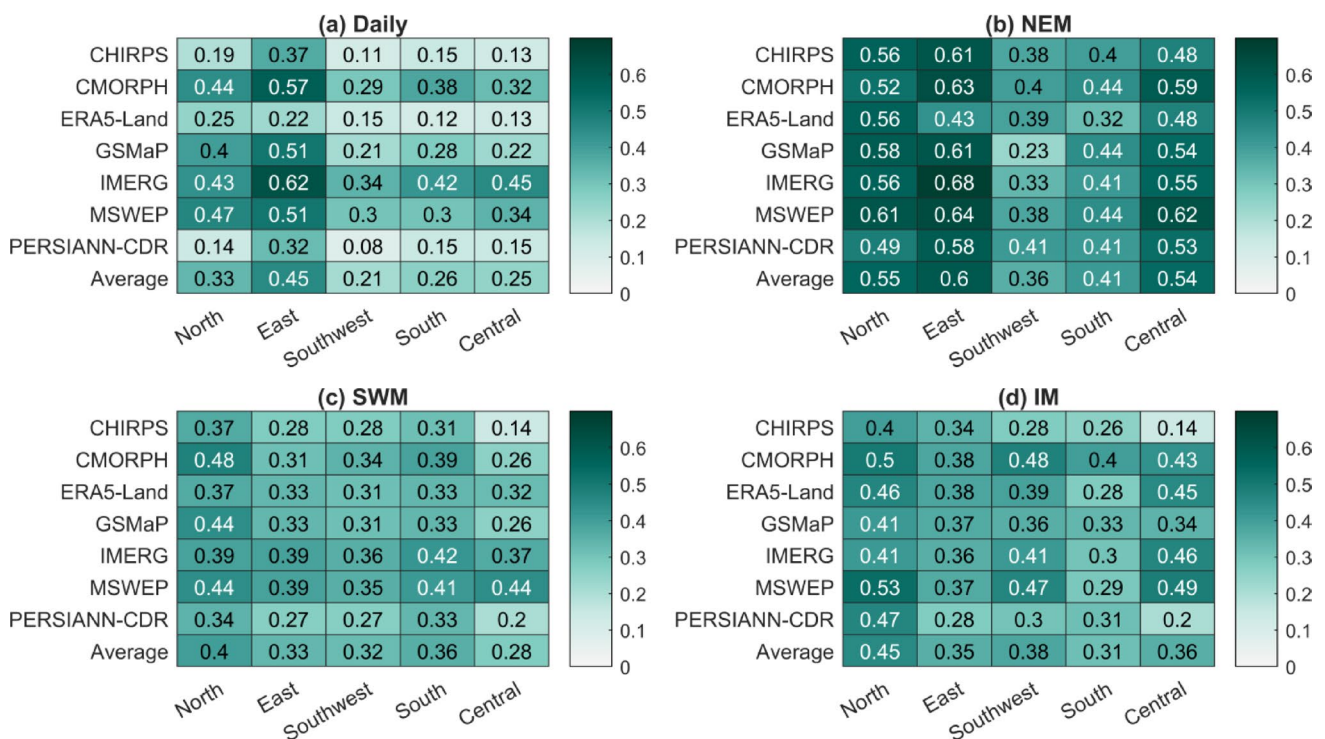


Fig. 5 Heat maps of regional mean KGE scores for the north, east, southwest, south and central regions at **a** daily timescale, **b** NEM season, **c** SWM season and **d** IM season

S4; Supplementary Materials). Spatial distribution shows $KGE > 0.6$ are mainly concentrated along the wetter east coast, whereas $KGE < 0.4$ are clustered in the drier southwest and southern regions (Fig. 6). Four out of seven products (IMERG, CMORPH, MSWEP and GSMaP) achieve regional mean $KGE > 0.50$ in the east, while none exceeds 0.40 in the southwest. This indicates a magnitude-dependent error structure in Peninsular Malaysia, whereby estimates for higher rainfall magnitudes tend to have lower errors and vice versa (Mei et al. 2014).

In the seasonal analysis, the north, east, south and central regions record their highest average KGE scores during the NEM season, while the performance in southwest peaks during the IM season (Fig. 5). Conversely, the poorest KGE scores are observed in the SWM season for the north, east, southwest and central regions and during the IM season for the south. Average KGE scores improve across all regions during the NEM season, with high values (> 0.6) predominantly clustered in the east, north and central regions (Fig. 6). During the SWM season, higher KGE scores are concentrated in the north and south, whereas in the IM season, high scores are scattered across Peninsular Malaysia without a clear pattern. These spatial patterns in seasonal performance correspond with the wettest regions identified by Ng et al. (2022) during each monsoon season, reinforcing the presence of a magnitude-dependent error structure.

Figure 7 presents a probabilistic evaluation of precipitation estimates across rainfall intensity classes using PDFs derived from regional-averaged time series. At the daily scale, all climatic regions exhibit high frequencies in three intensity categories, namely negligible, low moderate and high moderate rain. Conversely, high heavy and very heavy rain collectively represent only about 5% of total rainfall occurrences. However, regional disparities in rainfall structure are apparent. The north and east regions exhibit a higher proportion of high moderate rain ($\sim 22\%$), whereas the southwest and south regions show left-skewed distributions with a predominance of low moderate rain ($\sim 26\%$). In comparison, the central region demonstrates a flattened distribution across moderate classes but a relatively high frequency of negligible rain ($\sim 35\%$).

Precipitation products generally show consistent skill in capturing light, high heavy and very heavy rainfall. However, notable deviations arise in the negligible, low and high moderate and low heavy rain categories. CMORPH and GSMaP tend to overestimate negligible rainfall and underestimate moderate intensities in all regions, except for the central region. In contrast, ERA5-Land, PERSIANN-CDR and CHIRPS display the opposite pattern in those rain categories. IMERG has the highest skills across intensity classes in the north, east, southwest and south regions. In the central region, all precipitation products except CMORPH

clearly underestimate the negligible rainfall, while CHIRPS, MSWEP, ERA5-Land and PERSIANN-CDR substantially overestimate the high moderate rainfall.

Seasonal analysis reveals greater divergence in rainfall structure across regions, primarily driven by monsoonal influences. During the NEM season, all regions exhibit an increase in the frequency of negligible rainfall events by approximately 1–12%. Flatter probability distribution curves are also observed, indicating reduced frequencies of light to low heavy rainfall. In the east region, the frequency of high heavy and very heavy rainfall events increases by 3–4% due to direct exposure to monsoonal cold surges (Ng et al. 2022). The southwest and south regions also display slight increases in heavy rainfall during the NEM season. Among the precipitation products, CHIRPS, CMORPH and GSMaP generally underestimate moderate rainfall in all regions, while ERA5-Land tends to overestimate moderate rainfall in the east and central regions. Additionally, heavy rainfall is consistently overestimated by CHIRPS, ERA5-Land, MSWEP and PERSIANN-CDR in all regions. IMERG demonstrates the highest skills during the NEM season across all regions.

Unlike the NEM season, rainfall distributions appear more sharply peaked during the SWM and IM periods, with increased moderate and low heavy rainfall and reduced negligible and very heavy rain, except for the southwest region during SWM season. This shift indicates a transition toward moderate-intensity rainfall during these seasons. However, these seasonal structural changes are not well represented by precipitation products, with relatively large deviations observed for the negligible, moderate and heavy rainfall. CHIRPS, ERA5-Land, MSWEP and PERSIANN-CDR consistently underestimate negligible rainfall and overestimate moderate and low heavy rainfall, while CMORPH and GSMaP show the opposing patterns. IMERG maintains relatively high skills across most intensity classes, though it underestimates negligible rainfall in the south and central regions. The results reveal pronounced spatiotemporal variability in the skill of precipitation products to characterize daily and seasonal rainfall structure in Peninsular Malaysia.

3.3 Annual cycle and interannual extremes

Figure 8 presents the annual cycles of domain-averaged FREQ, SDII and Q90 across Peninsular Malaysia. Based on gauge observations, the highest number of wet days occurs in November, while February records the lowest. Most precipitation products exhibit better agreement with observations during the NEM season, whereas larger biases and wider gaps are observed during the SWM and IM seasons. CMORPH shows the highest skills among the products, while ERA5-Land, MSWEP and PERSIANN-CDR

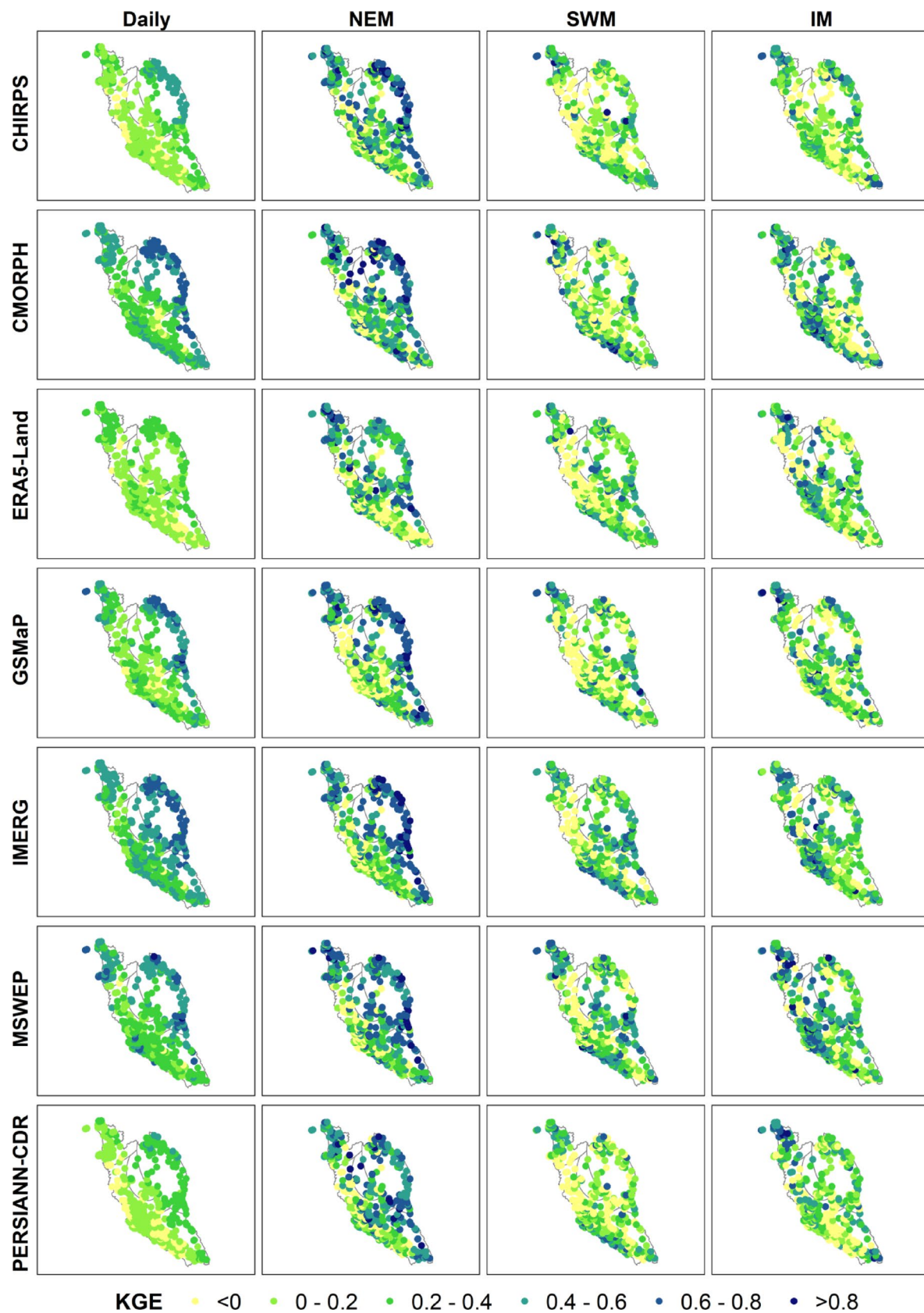


Fig. 6 Spatial distribution of KGE scores for daily and seasonal timescales

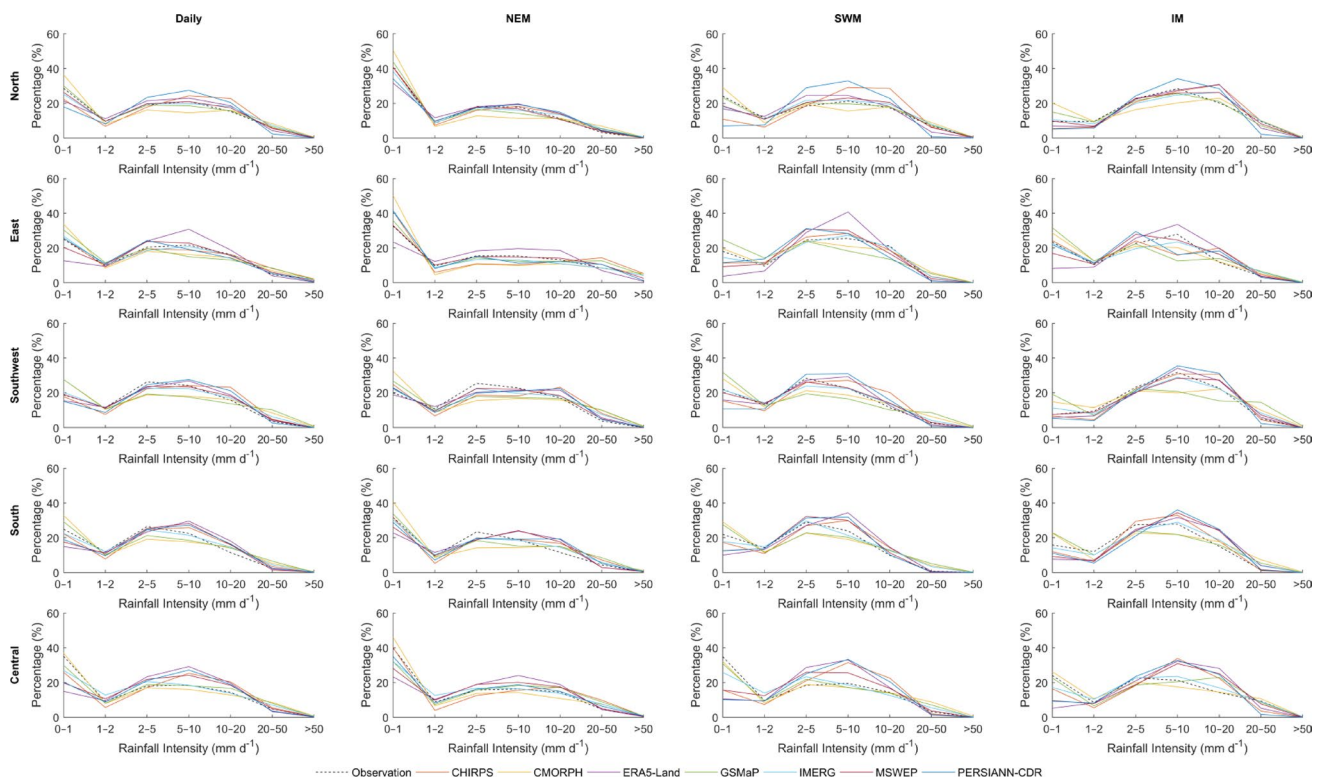


Fig. 7 Probability distribution functions of rainfall intensity across daily and seasonal timescales for the north, east, southwest, south and central regions

consistently overestimate wet-day frequencies by >5 days per month. CHIRPS performs well during the NEM season but shows reduced accuracy during the SWM season. The annual cycle of *FREQ* is further normalized (Fig. 8d) to evaluate the month-to-month structure of rainfall frequency without biases. All products correctly depict the bimodal peaks in April–May and October–November, consistent with previous findings by Ngai et al. (2020). This suggests that the discrepancies in Fig. 8a are primarily due to magnitude biases. However, all products tend to underestimate the normalized *FREQ* ratios during the early NEM months (November and December).

For *SDII*, most precipitation products show substantial underestimations, particularly from March to November, except for CHIRPS and CMORPH during the NEM season (Fig. 8b). This underestimation is likely due to the concurrent overestimation of rainfall frequency, which skews the magnitude–frequency ratio. Despite a slight overestimation during the NEM season, CMORPH remains the best performer, whereas ERA5-Land, MSWEP and PERSIANN-CDR continue to show poor skill which is consistent with their *FREQ* performance. The normalized *SDII* ratios indicate that most products capture the month-to-month intensity pattern, except for CHIRPS which overestimates *SDII* ratio in January–February and underestimates it from April to October.

In contrast to frequency and intensity, all precipitation products show good agreement with the observed *Q90* distribution, indicating reliable skill in reproducing the monthly extremes, particularly in early months of the year. However, deviations increase from April onward, likely due to greater rainfall variability associated with monsoonal transitions. During this period, most products underestimate *Q90*, except for CMORPH which slightly overestimates it. CMORPH, GSMaP and IMERG consistently outperform others throughout the annual cycle. From the normalized *Q90* ratios, it is evident that precipitation products better capture monthly extremes than frequency or intensity after removing biases. These results suggest that most products are more reliable in characterizing month-to-month extreme rainfalls than frequency or intensity.

Figure 9 presents the evaluation of precipitation product in estimating year-to-year extreme rainfall. Overall, all products demonstrate poor performance in estimating the magnitude of interannual extreme indices with evident biases. However, after normalization, most products capture the general curves of the extreme indices reasonably well. This indicates that while precipitation products are capable of reproducing interannual variability in pattern, they remain highly susceptible to biases in terms of magnitude. Among the extreme indices, precipitation products show relatively better skill in estimating the *PRCPTOT* series. All

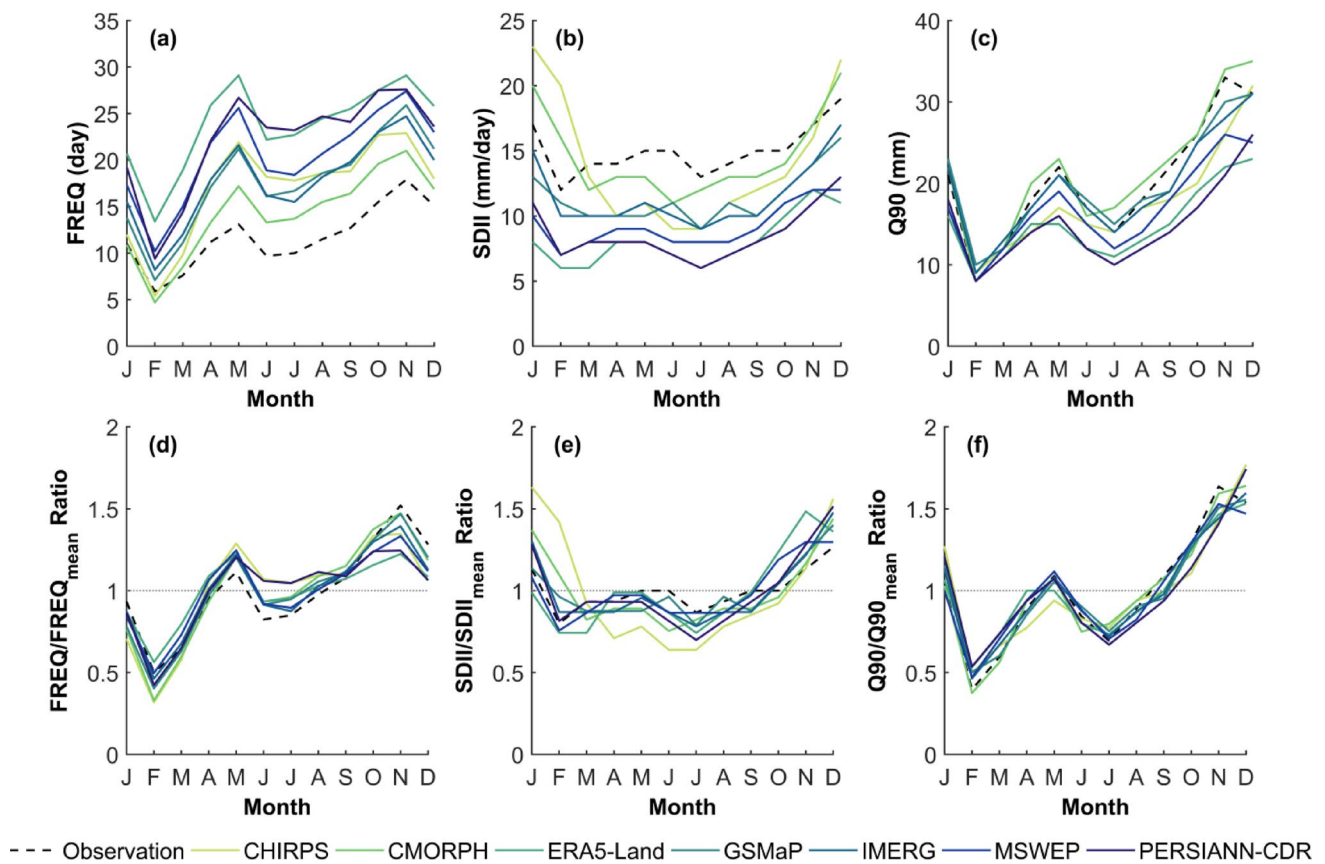


Fig. 8 Annual cycle for **a** FREQ, **b** SDII, **c** Q90 and their respective normalized ratios **d** FREQ/FREQ_{mean}, **e** SDII/SDII_{mean} and **f** Q90/Q90_{mean} over Peninsular Malaysia during 2014–2019

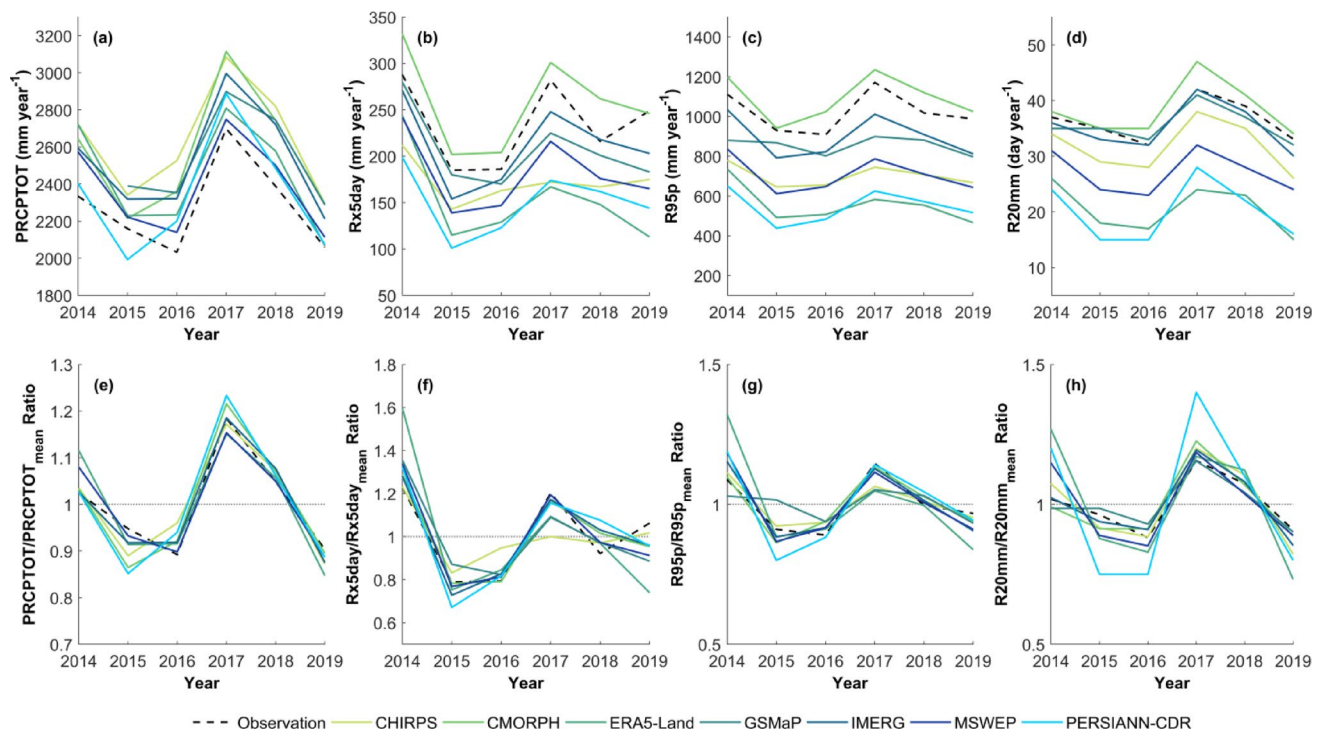


Fig. 9 Domain-averaged time series for **a** PRCPTOT, **b** Rx5day, **c** R95p, **d** R20mm and their respective normalized ratios **e** PRCPTOT/PRCPTOT_{mean}, **f** Rx5day/Rx5day_{mean}, **g** R95p/R95p_{mean} and **h** R20mm/R20mm_{mean} over Peninsular Malaysia during 2014–2019

products overestimate PRCPTOT, except for PERSIANN-CDR which underestimates it in 2015. When considering the normalized PRCPTOT ratios, all products align well with gauge observations across most years, except for 2015, 2016 and 2017. Previous studies have identified 2015/2016 as among the driest years in the past decade and 2017 as the wettest year in the last 30 years, largely due to El Niño–Southern Oscillation (ENSO) influences (Ng et al. 2022). While the causal linkage between ENSO events and precipitation product performance remains unclear, the current findings suggest that these products may struggle to accurately reflect regional extreme rainfall variations driven by ENSO. Further investigation using long-term rainfall records is necessary to confirm this relationship with stronger evidence.

For Rx5day and R95p, similar performance is observed from all precipitation products, whereby most products underestimate the magnitude of both indices, except for CMORPH which overestimates them. CHIRPS, ERA5-Land, GSMaP and PERSIANN-CDR exhibit poor accuracy in estimating normalized Rx5day and R95p ratios, while CMORPH, IMERG and MSWEP show relatively better agreement with observations. Performance for the R20mm index is generally poor across most products, with substantial deviations from observed series. However, GSMaP and IMERG demonstrate noticeably better accuracy in capturing both magnitude and normalized curves. After normalization, most precipitation products show consistent performance across the time series, although PERSIANN-CDR continues to underperform, especially in 2015, 2016 and 2017.

Figure 10 presents the spatial distribution of mean extreme indices values across Peninsular Malaysia. Based on gauge observations, high values of PRCPTOT and R20mm are primarily concentrated in the north, east and southwest regions, whereas high mean values of Rx5day and R95p are spatially clustered in the north and along the eastern coastline. The inland areas of the southwest and south regions represent drier zones with lower magnitudes of extremes, consistent with previous findings by Ng et al. (2022). For PRCPTOT, all precipitation products tend to overestimate the mean values at dry stations in the southwest and south regions, while underestimating them at wet stations in the east. Large deviations from observations (>1000 mm/year) are observed in CMORPH, GSMaP and PERSIANN-CDR, whereas CHIRPS, ERA5-Land, IMERG and MSWEP exhibit smaller discrepancies. For Rx5day, all precipitation products show a consistent pattern of large underestimation in the east region. This suggests that the deviations observed in Fig. 9b primarily originate from the wet stations located in the east. In contrast, other regions show minimal differences from observations (<100 mm/year).

For R95p and R20mm, the products exhibit mixed performance in different regions. ERA5-Land, MSWEP and PERSIANN-CDR strongly underestimate both indices across most of Peninsular Malaysia, while IMERG underestimates them at several stations in the east. CHIRPS underestimates R95p consistently across the region, but shows underestimation of R20mm mainly in the north, southwest and south. GSMaP underestimates R95p across most regions, but underestimates R20mm only in the north, east and southwest. In contrast, CMORPH demonstrates a consistent spatial pattern for both indices, underestimating values in the east while overestimating them in inland areas between the east, southwest and south regions. These results highlight that the ability of precipitation products to estimate both the intensity and frequency of wet extremes is highly region-specific, particularly for R95p and R20mm.

4 Discussions

Based on the results, these rainfall estimates show a generally low level of agreement with gauge observations at both daily and seasonal scales, consistent with recent findings by Tan et al. (2015) and Tan and Santo (2018). They generally exhibit substantial random errors, primarily attributed to temporal sampling limitations (Tapiador et al. 2017) and mismatches in gauge reporting times (Beck et al. 2019b). Additionally, they inherit systematic regional biases, particularly in convection-dominated dry regions such as the southwest and southern parts of Peninsular Malaysia (Chang et al. 2005; Tangang et al. 2012). The poor skills observed in these dry regions may stem from the inability of precipitation products to correctly capture small-scale convective mass flux when multiple gauge stations fall within a single grid cell. In such cases, these products may interpret localized convective events as grid-scale rainfall, leading to overestimations especially when only a few gauges record rain. Conversely, when several gauges within a grid cell record rainfall, those with higher rainfall amounts tend to be underestimated due to aggregation across the entire grid. This effect is clearly illustrated in the scatter plot analysis, where all products tend to overestimate low rainfall and vice versa.

Rainfall estimates show higher skill during the NEM season compared to the SWM and IM seasons. This improvement can be attributed to the dominance of large-scale synoptic frontal systems, particularly the northeasterly cold surge along the east coast (Juneng and Tangang 2010). This phenomenon is triggered by the southward expansion of the Siberian High (Ding 1990; Wu and Chan 1995), creating strong low-level convergence and high vorticity conducive to rainfall formation (Chang et al. 2005; Ooi et al. 2011).

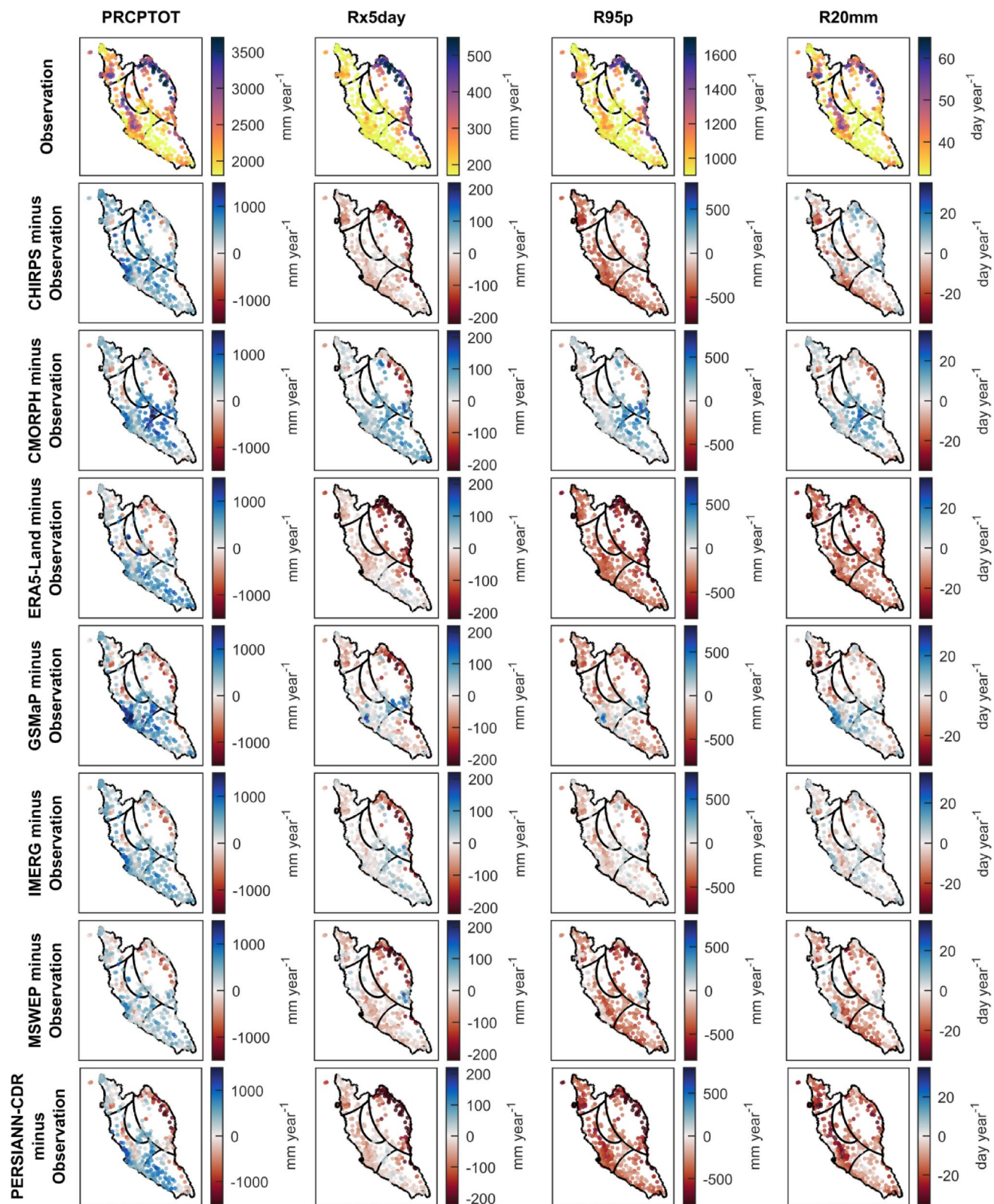


Fig. 10 Spatial distribution of mean PRCPTOT, RX5day, R95p and R20mm values from observations and their differences compared to precipitation products across Peninsular Malaysia

Compared to convective storms, these frontal-dominated systems are more predictable and spatially coherent (Prein et al. 2015), contributing to improved rainfall detection by most products during this season. The eastern coastline, fully exposed to these cold surges, naturally records the highest annual rainfall and extremes.

In terms of point-to-grid comparisons, higher spatial resolution products are often more accurate in representing rainfall patterns than those with coarser resolutions (Kidd and Huffman 2011). This is evident in the $0.10^\circ \times 0.10^\circ$ IMERG, which consistently shows the highest skill, whereas the $0.25^\circ \times 0.25^\circ$ PERSIANN-CDR is among the worst. These results are reasonable given that gridded datasets represent spatially averaged rainfall within each grid cell, and errors arising from scale mismatch in point-to-grid comparisons tend to diminish with finer spatial resolution. However, their performance rankings may differ under grid-to-grid evaluations or when areal reduction factors are applied. Nonetheless, our findings demonstrate that spatial resolution alone does not guarantee performance, as algorithms, interpolation techniques and input data sources also play significant roles. For example, CHIRPS, despite having the finest resolution ($0.05^\circ \times 0.05^\circ$), does not outperform other coarser-resolution products in most aspects. In fact, it shows low agreement with gauge observations in spatio-temporal analysis, rainfall climatology and event detection. This may be due to its reliance on thermal infrared-based (IR) Cold Cloud Duration estimates, which exclude passive microwave-derived (PMW) estimates that is known for more skillful rainfall sensing (Ebert et al. 1996; Funk et al. 2015). Similarly, PERSIANN-CDR uses an IR-based estimation algorithm, even though its artificial neural network model is trained with PMW-derived estimates (Ashouri et al. 2015). The combination of coarse resolution and IR-only estimates limits its performance in tropical regions. While IR sensors provide higher sampling frequency, they rely on cloud-top brightness temperature and often misinterpret non-precipitating clouds as rainfall.

In contrast, PMW sensors measure microwave energy from hydrometeors in the atmospheric column, providing more direct and reliable estimates (Tapiador et al. 2017; Wilheit et al. 1977). As such, some of the best-performing satellite products, namely IMERG, CMORPH and GSMaP, employ IR–PMW combined techniques, integrating the frequent coverage of IR estimates with the physical robustness of PMW retrievals (Huffman et al. 2020; Kubota et al. 2020; Xie et al. 2017). These satellite precipitation products with IR–PMW combined methods display clear advantages over other products in estimating tropical rainfall over Peninsular Malaysia, as demonstrated in the results. Among them, IMERG and CMORPH outperform GSMaP, likely due to additional quality control mechanisms in their advection

morphed schemes, which enhance spatial consistency (Tapiador et al. 2017). IMERG is the best-performing product overall, due to its use of the advanced GPM dual-frequency precipitation radar, which offers higher sensitivity to light rainfall and allows for detailed characterization of microphysical processes (Hou et al. 2014). Its advantages over CMORPH are evident in Peninsular Malaysia, where CMORPH tends to overestimate low rainfall.

In addition to satellite-based datasets, ERA5-Land, a state-of-the-art reanalysis product, is increasingly popular for climate studies (Harrigan et al. 2020; Ng et al. 2024), due to improvements in model physics and data assimilation (Gleixner et al. 2020; Sheridan et al. 2020). It relies on in situ gauge and satellite observations assimilated into the numerical model throughout the reanalysis period to produce brief, short-range weather forecasts (Hersbach et al. 2020; Muñoz-Sabater et al. 2021; Simmons et al. 2010). However, despite recent advancements, ERA5-Land generally performs poorly in this study, especially in capturing convective rainfall and climatology. This aligns with earlier studies showing that numerical model-based reanalyses struggle in tropical regions due to limitations in convective parameterization (Ebert et al. 2007). Moreover, ERA5-Land's performance may be compromised by inhomogeneities arising from changes in the global observation network (Zhang et al. 2011). While ERA5-Land remains useful for estimating mean rainfall, it still lacks sufficient skills to estimate extremes and convective events in tropics compared to the satellite technique. This highlights the need for further improvements in simulating tropical convection in reanalysis models.

In contrast, MSWEP exhibits relatively good performance compared to ERA5-Land, as it blends gauge, satellite and reanalysis precipitation estimates to improve skills in densely gauged regions (Beck et al. 2019b). It shows higher correlation with daily observations as it adjusts for gauge reporting times to minimize temporal mismatches. However, MSWEP demonstrates lower agreement with the gauge observations in variability metrics and rainfall PDFs. This could be due to its use of ERA5 climatology to adjust merged satellite-reanalysis estimates, which are known to produce flatter and less peaky rainfall distributions (GloH2O 2022). Additionally, MSWEP underperforms in capturing the annual cycle and interannual extremes of Peninsular Malaysia compared to other products, possibly due to inhomogeneities from constant changes of data sources. The results reflect the necessity to enhance the representation of the rainfall PDFs and climatology to achieve comparable skills with IMERG in Peninsular Malaysia.

5 Conclusion

This study focused on the evaluation of seven state-of-the-art precipitation products against gauge observations in Peninsular Malaysia from 2014 to 2019. The assessment was conducted at daily and seasonal timescales using both continuous and categorical error metrics to quantify their spatiotemporal performance. The ability of these products to estimate the regional climatology, including the annual cycle and interannual extremes, was also examined. The major findings are summarized as follows.

In point-to-grid comparisons, MSWEP shows highest point-scale accuracy, although all products exhibit considerable spatial inconsistencies. Based on KGE scores, IMERG and CMORPH are the most skillful in representing daily rainfall dynamics. However, poor correlation remains a key limitation across all products. For rain event detection, all products performed reasonably well in terms of ACC, POD and FAR, but poorly in CSI and HSS. CMORPH and IMERG exhibit more balanced categorical performance. Region-specific analysis reveals higher skills along the wetter east coast and lower skills in the drier southwest and south regions, reflecting a magnitude-dependent error structure. IMERG performs best over the east, southwest, south and central regions, whilst MSWEP outperforms the rest in the north. These regional discrepancies become more evident in seasonal analysis, with product performance improving during the wetter NEM season but declining during the drier SWM and IM seasons. The performance ranking of precipitation products in each region also varies seasonally. Most products fail to replicate seasonal shifts in rainfall structure, particularly for negligible and moderate intensity classes. In terms of climatology, all products misrepresent the rainfall frequency and intensity in the annual cycle due to substantial biases, although extreme quantiles are better captured. Interannual extremes are generally underestimated in magnitude but reasonably tracked temporally, except during ENSO years when products struggled to replicate the observed variability.

Several issues related to algorithm design, input data sources and spatial resolution were highlighted in the discussions as key factors affecting product performance. The overall findings emphasize the importance of region-specific validation in complex tropical climates before applying precipitation products to operational hydrology (e.g., flood forecasting, drought monitoring) and coupled land-atmosphere modeling. These findings also underscore the need for further algorithmic refinements and bias correction techniques to enhance rainfall estimation in convection-dominated tropical environments. Future studies should examine the capability of precipitation products to simulate the diurnal cycle in tropics and explore the propagation of

uncertainty in rainfall-runoff processes for flood prediction and disaster response. Additional quantitative estimation of errors due to spatial mismatch and areal reduction factors under different storm types, durations and elevation is recommended in future studies to better assess spatial representativeness.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00477-025-03166-y>.

Acknowledgements The authors highly appreciate the financial support provided by the Ministry of Higher Education, Malaysia via the Fundamental Research Grant Scheme (Grant number: FRGS/1/2020/TK0/UM/02/19). We acknowledge the Malaysia Department of Irrigation and Drainage for making the daily rainfall observed data available in this study. We also thank the providers/developers of precipitation products for keeping the datasets freely available online. Yiwen Mei is supported by the National Key R&D Program of China (2024YFC3013302).

Author contributions All authors contributed to the study conception and design. Data collection and analysis were performed by Cia Yik Ng. Project administration and funding acquisition were performed by Faridah Othman. The first draft of the manuscript was written by Cia Yik Ng. All authors reviewed and approved the final manuscript.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Malaya. This work was funded by the Ministry of Higher Education, Malaysia (Grant number: FRGS/1/2020/TK0/UM/02/19).

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical approval The authors confirm that this research adheres to ethical guidelines.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Alexander AA, Rasool T, Kumar C, Sahoo S, Bhowmik RD, Kumar DN (2025) Unprecedented rainfall events increase the magnitude of design storms. *Environ Res Lett* 20:064011. <https://doi.org/10.1088/1748-9326/add175>
- Ashouri H, Hsu K-L, Sorooshian S, Braithwaite DK, Knapp KR, Cecil LD, Nelson BR, Prat OP (2015) PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull Am Meteorol Soc* 96:69–83. <https://doi.org/10.1175/BAMS-D-13-00068.1>
- Ayoub AB, Tangang F, Juneng L, Tan ML, Chung JX (2020) Evaluation of gridded precipitation datasets in Malaysia. *Remote Sens* 12:613. <https://doi.org/10.3390/rs12040613>
- Bakar MAA, Ariff NM, Jemain AA, Nadzir MSM (2020) Cluster analysis of hourly rainfalls using storm indices in Peninsular Malaysia. *J Hydrol Eng* 25:05020011. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001942](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001942)
- Beck HE, Vergopolan N, Pan M, Levizzani V, van Dijk AIJM, Weedon GP, Brocca L, Pappenberger F, Huffman GJ, Wood EF (2017) Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol Earth Syst Sci* 21:6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>
- Beck HE, Pan M, Roy T, Weedon GP, Pappenberger F, van Dijk AIJM, Huffman GJ, Adler RF, Wood EF (2019a) Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol Earth Syst Sci* 23:207–224. <https://doi.org/10.5194/hess-23-207-2019>
- Beck HE, Wood EF, Pan M, Fisher CK, Miralles DG, van Dijk AIJM, McVicar TR, Adler RF (2019b) MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bull Am Meteorol Soc* 100:473–500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
- Chang C-P, Wang Z, McBride J, Liu C-H (2005) Annual cycle of Southeast Asia—Maritime Continent rainfall and the asymmetric monsoon transition. *J Clim* 18:287–301. <https://doi.org/10.1175/jcli-3257.1>
- Ding Y (1990) Build-up, air mass transformation and propagation of Siberian high and its relations to cold surge in East Asia. *Meteorol Atmos Phys* 44:281–292. <https://doi.org/10.1007/BF01026822>
- Du H, Tan ML, Chun KP, Zhang F (2025) Evaluation of four gridded climate products for streamflow and drought simulations in the Kelantan River Basin, Malaysia. *Geocarto Int* 40:2453615. <https://doi.org/10.1080/10106049.2025.2453615>
- Dunn RJH, Donat MG, Alexander LV (2022) Comparing extremes indices in recent observational and reanalysis products. *Front Clim*. <https://doi.org/10.3389/fclim.2022.989505>
- Ebert EE, Manton MJ, Arkin PA, Allam RJ, Holpin GE, Gruber A (1996) Results from the GPCP algorithm intercomparison programme. *Bull Am Meteorol Soc* 77:2875–2888. [https://doi.org/10.1175/1520-0477\(1996\)077%3c2875:RFTGAI%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077%3c2875:RFTGAI%3e2.0.CO;2)
- Ebert EE, Janowiak JE, Kidd C (2007) Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull Am Meteorol Soc* 88:47–64. <https://doi.org/10.1175/BAMS-88-1-47>
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data* 2:150066. <https://doi.org/10.1038/sdata.2015.66>
- Ganiyu HO, Othman F, Jaafar WZW, Ng CY (2025) Comprehensive evaluation of satellite precipitation products over sparsely gauged river basin in Nigeria. *Theor Appl Climatol* 156:169. <https://doi.org/10.1007/s00704-025-05388-0>
- Gebremichael M (2010) Framework for satellite rainfall product evaluation. Rainfall: state of the science. American Geophysical Union, Washington DC, pp 265–275
- Gleixner S, Demissie T, Diro GT (2020) Did ERA5 improve temperature and precipitation reanalysis over East Africa? *Atmosphere* 11:996. <https://doi.org/10.3390/atmos11090996>
- GloH2O (2022) MSWEP V2.8 technical documentation. <https://www.gloh2o.org/mswep>. Accessed 9 Oct 2022
- Harrigan S, Zsoter E, Alfieri L, Prudhomme C, Salamon P, Wetterhall F, Barnard C, Cloke H, Pappenberger F (2020) GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst Sci Data* 12:2043–2060. <https://doi.org/10.5194/essd-12-2043-2020>
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J-N (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, Oki R, Nakamura K, Iguchi T (2014) The global precipitation measurement mission. *Bull Am Meteorol Soc* 95:701–722. <https://doi.org/10.1175/bams-d-13-00164.1>
- Huffman GJ, Bolvin DT, Braithwaite D, Hsu K-L, Joyce RJ, Kidd C, Nelkin EJ, Sorooshian S, Stocker EF, Tan J, Wolff DB, Xie P (2020) Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG). Satellite precipitation measurement, vol 1. Springer, Cham, pp 343–353
- Huffman GJ, Adler RF, Behrangi A, Bolvin DT, Nelkin EJ, Gu G, Ehsani MR (2023) The new version 3.2 Global Precipitation Climatology Project (GPCP) monthly and daily precipitation products. *J Clim* 36:7635–7655. <https://doi.org/10.1175/JCLI-D-23-0123.1>
- IPCC (2021) Weather and climate extreme events in a changing climate. Climate change 2021: the physical science basis Contribution of Working Group I to the sixth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 1513–1766
- Jamaluddin AF, Tangang FT, Chung JX, Juneng L, Sasaki H, Takayabu I (2018) Investigating the mechanisms of diurnal rainfall variability over Peninsular Malaysia using the non-hydrostatic regional climate model. *Meteorol Atmos Phys* 130:611–633. <https://doi.org/10.1007/s00703-017-0541-x>
- Juneng L, Tangang FT (2010) Long-term trends of winter monsoon synoptic circulations over the maritime continent: 1962–2007. *Atmos Sci Lett* 11:199–203. <https://doi.org/10.1002/asl.272>
- Kidd C, Huffman G (2011) Global precipitation measurement. *Meteorol Appl* 18:334–353. <https://doi.org/10.1002/met.284>
- Kidd C, Becker A, Huffman GJ, Muller CL, Joe P, Skofronick-Jackson G, Kirschbaum DB (2017) So, how much of the Earth's surface is covered by rain gauges? *Bull Am Meteorol Soc* 98:69–78. <https://doi.org/10.1175/bams-d-14-00283.1>
- Kling H, Fuchs M, Paulin M (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J Hydrol* 424:264–277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Kubota T, Aonashi K, Ushio T, Shige S, Takayabu YN, Kachi M, Arai Y, Tashima T, Masaki T, Kawamoto N, Mega T, Yamamoto MK, Hamada A, Yamaji M, Liu G, Oki R (2020) Global satellite mapping of precipitation (GSMaP) products in the GPM era. Satellite precipitation measurement, vol 1. Springer, Cham, pp 355–373

- Lim JT (1976) Rainfall minimum in Peninsular Malaysia during the northeast monsoon. *Mon Weather Rev* 104:96–99. [https://doi.org/10.1175/1520-0493\(1976\)104%3c0096:Rmipmd%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104%3c0096:Rmipmd%3e2.0.CO;2)
- Ma Z, Xu J, Ma Y, Zhu S, He K, Zhang S, Ma W, Xu X (2022) AERA5-Asia: a long-term Asian precipitation dataset (0.1°, 1-hourly, 1951–2015, Asia) anchoring the ERA5-Land under the total volume control by APHRODITE. *Bull Am Meteorol Soc* 103:E1146–E1171. <https://doi.org/10.1175/BAMS-D-20-0328.1>
- Maggioni V, Meyers PC, Robinson MD (2016) A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *J Hydrometeorol* 17:1101–1117. <https://doi.org/10.1175/jhm-d-15-0190.1>
- Mei Y, Anagnostou EN, Nikolopoulos EI, Borga M (2014) Error analysis of satellite precipitation products in mountainous basins. *J Hydrometeorol* 15:1778–1793. <https://doi.org/10.1175/JHM-D-13-0194.1>
- Mei Y, Anagnostou EN, Shen X, Nikolopoulos EI (2017) Decomposing the satellite precipitation error propagation through the rainfall-runoff processes. *Adv Water Resour* 109:253–266. <https://doi.org/10.1016/j.advwatres.2017.09.012>
- Mei Y, Maggioni V, Houser P, Xue Y, Rouf T (2020) A nonparametric statistical technique for spatial downscaling of precipitation over high mountain Asia. *Water Resour Res* 56:e2020WR027472. <https://doi.org/10.1029/2020WR027472>
- Meteorological Service Singapore (2015) Singapore's second national climate change study – phase 1. Met Office, Exeter, UK, Centre for Climate Research Singapore, Singapore, National Oceanography Centre, Liverpool, UK, CSIRO, Australia, and Newcastle University, Newcastle, UK. <http://ccrs.weather.gov.sg/Publications-Second-National-Climate-Change-Study-Science-Reports>. Accessed 1 Oct 2023
- Mohd Akhir MF, Zakaria NZ, Tangang F (2014) Intermonsoon variation of physical characteristics and current circulation along the east coast of Peninsular Malaysia. *Int J Oceanogr* 2014:527587. <https://doi.org/10.1155/2014/527587>
- Mohd Zad SN, Zulkafli Z, Muharram FM (2018) Satellite rainfall (TRMM 3B42-V7) performance assessment and adjustment over Pahang River Basin, Malaysia. *Remote Sens* 10:388. <https://doi.org/10.3390/rs10030388>
- Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles DG, Piles M, Rodríguez-Fernández NJ, Zsoter E, Buontempo C, Thépaut JN (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst Sci Data* 13:4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Ng CY, Wan Jaafar WZ, Mei Y, Othman F, Lai SH, Liew J (2022) Assessing the changes of precipitation extremes in Peninsular Malaysia. *Int J Climatol* 42:7914–7937. <https://doi.org/10.1002/joc.7684>
- Ng CY, Wan Jaafar WZ, Othman F, Lai SH, Mei Y, Juneng L (2024) Assessment of evaporative demand drought index for drought analysis in Peninsular Malaysia. *Sci Total Environ* 917:170249. <https://doi.org/10.1016/j.scitotenv.2024.170249>
- Ngai ST, Juneng L, Tangang F, Chung JX, Salimun E, Tan ML, Amalia S (2020) Future projections of Malaysia daily precipitation characteristics using bias correction technique. *Atmos Res* 240:104926. <https://doi.org/10.1016/j.atmosres.2020.104926>
- Ooi SH, Samah AA, Braesicke P (2011) A case study of the Borneo vortex genesis and its interactions with the global circulation. *J Geophys Res Atmos*. <https://doi.org/10.1029/2011JD015991>
- Prein AF, Langhans W, Fossler G, Ferrone A, Ban N, Goergen K, Keller M, Tölle M, Gutjahr O, Feser F, Brisson E, Kollet S, Schmidli J, van Lipzig NPM, Leung R (2015) A review on regional convection-permitting climate modeling: demonstrations, prospects, and challenges. *Rev Geophys* 53:323–361. <https://doi.org/10.1002/2014RG000475>
- Rana S, McGregor J, Renwick J (2015) Precipitation seasonality over the Indian subcontinent: an evaluation of gauge, reanalyses, and satellite retrievals. *J Hydrometeorol* 16:631–651. <https://doi.org/10.1175/jhm-d-14-0106.1>
- Seo SB, Das Bhowmik R, Sankarasubramanian A, Mahinthakumar G, Kumar M (2019) The role of cross-correlation between precipitation and temperature in basin-scale simulations of hydrologic variables. *J Hydrol* 570:304–314. <https://doi.org/10.1016/j.jhydrol.2018.12.076>
- Sheridan SC, Lee CC, Smith ET (2020) A comparison between station observations and reanalysis data in the identification of extreme temperature events. *Geophys Res Lett* 47:e2020GL088120. <https://doi.org/10.1029/2020GL088120>
- Simmons AJ, Willett KM, Jones PD, Thorne PW, Dee DP (2010) Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: inferences from reanalyses and monthly gridded observational data sets. *J Geophys Res Atmos*. <https://doi.org/10.1029/2009JD012442>
- Soo EZX, Wan Jaafar WZ, Lai SH, Islam T, Srivastava P (2018) Evaluation of satellite precipitation products for extreme flood events: case study in Peninsular Malaysia. *J Water Clim Change* 10:871–892. <https://doi.org/10.2166/wcc.2018.159>
- Soo EZX, Wan Jaafar WZ, Lai SH, Othman F, Elshafie A (2022) Enhancement of satellite precipitation estimations with bias correction and data-merging schemes for flood forecasting. *J Hydrol Eng* 27:05022009. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002190](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002190)
- Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu K-L (2018) A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev Geophys* 56:79–107. <https://doi.org/10.1002/2017rg000574>
- Tan ML, Santo H (2018) Comparison of GPM IMERG, TMPA 3B42 and PERSIANN-CDR satellite precipitation products over Malaysia. *Atmos Res* 202:63–76. <https://doi.org/10.1016/j.atmosres.2017.11.006>
- Tan ML, Ibrahim AL, Duan Z, Cracknell AP, Chaplot V (2015) Evaluation of six high-resolution satellite and ground-based precipitation products over Malaysia. *Remote Sens* 7:1504–1528. <https://doi.org/10.3390/rs70201504>
- Tangang F, Juneng L, Salimun E, Sei K, Loh JL, Muhamad H (2012) Climate change and variability over Malaysia: gaps in science and research information. *Sains Malays* 41:1355–1366
- Tapiador FJ, Navarro A, Levizzani V, García-Ortega E, Huffman GJ, Kidd C, Kucera PA, Kummerow CD, Masunaga H, Petersen WA, Roca R, Sánchez JL, Tao WK, Turk FJ (2017) Global precipitation measurements for validating climate models. *Atmos Res* 197:1–20. <https://doi.org/10.1016/j.atmosres.2017.06.021>
- Tefera GW, Ray RL, Wooten AM (2024) Evaluation of statistical downscaling techniques and projection of climate extremes in central Texas, USA. *Weather Clim Extrem* 43:100637. <https://doi.org/10.1016/j.wace.2023.100637>
- Wang XL (2008) Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or f test. *J Appl Meteorol Climatol* 47:2423–2444. <https://doi.org/10.1175/2008jamc1741.1>
- Wang XL, Wen QH, Wu Y (2007) Penalized maximal t test for detecting undocumented mean change in climate data series. *J Appl Meteorol Climatol* 46:916–931. <https://doi.org/10.1175/jam2504.1>
- Wilheit TT, Chang ATC, V. Rao MS, Rodgers EB, Theon JS (1977) A satellite technique for quantitatively mapping rainfall rates over the oceans. *J Appl Meteorol* 16(5):551–560. [https://doi.org/10.1175/1520-0450\(1977\)016%3c0551:ASTFQM%3e2.0.CO;2](https://doi.org/10.1175/1520-0450(1977)016%3c0551:ASTFQM%3e2.0.CO;2)

- Wilks DS (2011) Chapter 8—Forecast verification. International geophysics. Academic Press, Oxford, pp 301–394
- World Meteorological Organization (2009) Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. <https://library.wmo.int/idurl/4/48826>. Accessed 18 Feb 2024
- Wu MC, Chan JCL (1995) Surface features of winter monsoon surges over South China. *Mon Weather Rev* 123:662–680. [https://doi.org/10.1175/1520-0493\(1995\)123%3c0662:Sfowms%3e2.0.Co;2](https://doi.org/10.1175/1520-0493(1995)123%3c0662:Sfowms%3e2.0.Co;2)
- Xie P, Arkin PA (1996) Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J Clim* 9:840–858. [https://doi.org/10.1175/1520-0442\(1996\)009%3c0840:Aogmpu%3e2.0.Co;2](https://doi.org/10.1175/1520-0442(1996)009%3c0840:Aogmpu%3e2.0.Co;2)
- Xie P, Joyce R, Wu S, Yoo S-H, Yarosh Y, Sun F, Lin R (2017) Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J Hydrometeorol* 18:1617–1641. <https://doi.org/10.1175/JHM-D-16-0168.1>
- Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wires Clim* 2:851–870. <https://doi.org/10.1002/wcc.147>

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