Spatiotemporal performance evaluation of high-resolution multiple satellite and reanalysis precipitation products over the semiarid region of India

Elangovan Devadarshini[1], Kulanthaivelu Bhuvaneswari[2], Shanmugam Mohan Kumar[1], Vellingiri Geethalakshmi[3*], Manickam Dhasarathan[1], Alagarsamy Senthil[4], Kandasamy Senthilraja[5], Shahbaz Mushtaq[6], Thong Nguyen-Huy[6] and Thanh Mai[6]

[1]Agro Climate Research Centre, Tamil Nadu Agricultural University, Coimbatore, India

[2]Centre for Rural Development Studies, Tamil Nadu Agricultural University, Coimbatore, India

[3]Office of the Vice-Chancellor, Tamil Nadu Agricultural University, Coimbatore, India

[4]Department of Crop Physiology, Tamil Nadu Agricultural University, Coimbatore, India

[5]Directorate of Research, Tamil Nadu Agricultural University, Coimbatore, India

[6]University of Southern Queensland, Australia

*Corresponding author email: geetha@tnau.ac.in (ORCID ID: 0000-0003-1631-121X)

Abstract

The present investigation evaluates three satellite precipitation products (SPPs), namely, Multi-Source Weighted-Ensemble Precipitation (MSWEP), Global Precipitation Climatology Centre (GPCC), Climate Hazard Infrared Precipitation with Station Data (CHIRPS) and two reanalysis datasets, namely, the ERA5 atmosphere reanalysis dataset (ERA5) and Indian Monsoon Data Assimilation and Analysis (IMDAA), against the good quality gridded reference dataset (1991-2022) developed by the India Meteorological Department (IMD). The evaluation was carried out in terms of the rainfall detection ability and estimation accuracy of the products using metrics such as the false alarm ratio (FAR), probability of detection (POD), misses, root mean square error (RMSE), and percent bias (PBIAS). Among all the rainfall products, ERA5 had the best ability to capture rainfall events with a higher POD, followed by MSWEP. Both MSWEP and ERA5 had PODs of 70-100% in more than 90% of the grids and less than 35% of missing rainfall events in the entire Tamil Nadu. In the case of the rainfall estimation accuracy evaluation, the MSWEP exhibited superior performance, with lower RMSEs and biases ranging from -25 to 25% at the annual and seasonal scales. In NEM, CHIRPS demonstrated a comparable performance to that of MSWEP in terms of the RMSE and PBIAS. These findings will help product users select the best reliable rainfall dataset for improved research, diversified applications in various sectors and policy-making decisions.

Keywords: Tamil Nadu, satellite precipitation products, MSWEP, GPCC, CHIRPS, ERA5, IMDAA

Introduction

The global food insecurity crisis is a significant problem exacerbated by the detrimental impacts of climate variability and change. Extreme weather events such as floods, droughts, heatwaves, and cold waves are causing extensive agricultural and socioeconomic losses. The frequency and intensity of these extreme weather events are increasing and projected to increase further in the future (Kalyan et al., 2021; Fowler et al., 2021). The Intergovernmental Panel on Climate Change (IPCC) has highlighted the major impacts experienced by vulnerable regions such as South Asia, including India, due to their geography and rising temperatures. In Tamil Nadu, the availability of water for crop cultivation is uncertain due to erratic and rainfall. This uncertainty is a significant concern because agriculture is the primary livelihood for many people in the region.

Investigating the spatial-temporal dynamics of hydrometeorological variables in the context of climate change, particularly in countries with rainfed agriculture, is important for assessing climate-driven variability and suggesting adaptation strategies (Asfaw et al., 2018). Reliable climate data are crucial in multiple sectors, especially agriculture, where climate strongly impacts crop growth. Precipitation and temperature are the most significant meteorological variables for studying regional climate variability, extreme weather events, and their influence on crop yield and food security (Yuvaraj et al., 2016). Geethalakshmi et al. (2008) reported that changes in these variables significantly affect crop production, food availability, and food prices. Precipitation data are also essential for flood prediction, water balance determination, and other practical applications. Spatiotemporal precipitation data has been extensively used in many pivotal spheres including agriculture, natural disaster assessment, water resources assessment and management (Collier, 2007; Behrangi et al., 2011; Zeng et al., 2012; Shah & Mishra, 2016; Peng et al., 2020).

Water usage has increased globally in the last century (Kummu et al., 2016), and the current climate crisis is expected to further increase the water requirements for crops and irrigation while reducing water availability due to global warming (Rockström et al., 2012). High-resolution precipitation data can help stakeholders devise effective water management strategies, improve intervention capacities for water conservation, and reduce water usage. Spatially detailed rainfall measurements can enhance the performance and accuracy of hydrological models (Silberstein, 2006; Merlin et al., 2008; Ragettli et al., 2014). High-quality precipitation data are crucial for crop insurance companies to develop appropriate index-based crop insurance products, mitigating the financial loss faced by farmers in the event of destructive weather events (Black et al., 2016; Enenkel et al., 2019).

Precipitation is one of the key climate variables that needs to be thoroughly analysed in terms of spatial and temporal distribution, variability, trends and precipitation extremes with a high degree of precision to assess its risks related to crop production and designing mitigation and evolving potential adaptation strategies. Despite the huge requirement for high-resolution rainfall data insufficient and unequally distributed raingauge networks make data too scanty to describe rainfall characteristics and pattern capturing the high spatial variability mainly in developing countries (Dinku, 2019). Although ground-based observations are important in understanding meteorological parameters, they can be limited in geographic scope due to factors like high altitude and complicated topography (Tapiador et al., 2012; Dinku et al., 2018). To conduct regional and global meteorological research, it is necessary to have a high-resolution database capable of capturing spatial-temporal changes in climate (Malvern & Maurice 2018; Gleixner et al., 2020). Mohan Kumar et al. (2022) highlighted that a dense network of stations is essential to provide comprehensive and accurate climate information across regions. In addition to ground-based observations, satellite retrievals and reanalysis products are valuable tools for climate monitoring.

In areas with limited rain gauges, satellite-based precipitation products (SPPs) offer a useful alternative due to their worldwide availability and high spatiotemporal resolution, providing more detailed weather information (Alijanian et al., 2019). The accuracy of SPPs varies depending on the region and precipitation type (Alijanian et al., 2019). Satellites equipped with remote sensing instruments can collect data over large geographic areas, offering a broader perspective on climate patterns. Reanalysis datasets combine global and regional weather models with observations to provide reliable historical data at different spatiotemporal resolutions. However, complex terrain and limited observations can introduce biases (Luo et al., 2019).

High-resolution precipitation and temperature products, such as the fifth-generation ECMWF global reanalysis (ERA-5) (Hersbach et al., 2020), Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis dataset (Rani et al., 2021), Multisource Weighted Ensemble Precipitation (MSWEP) and Multi Source Weather (MSWX) (Beck et al., 2017; 2019), Climate Hazards Group (CHG) InfraRed Precipitation with Stations data (CHIRPS) (Funk et al., 2015), and Global Precipitation Climatology Project (GPCP) (Huffman et al., 1997; 2001), have become available. These products merge satellite data, reanalysis products, and in situ measurements. These products can be used to identify climatic trends and patterns, help researchers and policymakers understand the impact of climate variability on food security and develop mitigation strategies.

In India, ERA-5 performed better at estimating daily rainfall than did CHIRPS across various climatic basins (Kolluru et al., 2020). The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis

(TMPA-3B42) measured rainfall closely matched ground-truth rainfall observations (Prakash, 2014). Several multisatellite high-resolution precipitation products (HRPPs), including Climate Prediction Center Morphing (CMORPH) version 1.0, TMPA-3B42, Naval Research Laboratory (NRL)–blended, and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), demonstrated high POD (probability of detection) and low FAR (false alarm rate) in most regions of India. However, these four HRPPs struggled to detect rainfall events in the rain shadow region of southeast peninsular India (where Tamil Nadu is located), semiarid parts of northwest India, and hilly parts of northern India (Prakash, 2014). CMORPH, PERSIANN, TMPA-3B42, and the Global Precipitation Climatology Project (GPCP) were found to be better at capturing rainfall events with high POD and fewer missing values (Sunilkumar et al., 2015). Comparatively, TRMM showed a better ability than CHIRPS for rainfall estimates in the catchment of the Gurupura River in India (Sharannya et al., 2020).

Despite significant developments in satellite-based precipitation data, which are more accurate and reliable (Michaelides et al., 2009; Levizzani et al., 2019), several studies have shown that the performance of satellite-based precipitation data can vary from one region to another (Gebremichael et al., 2014; Bharti & Singh, 2015). Therefore, it is necessary to comprehensively evaluate satellite products in various geographic regions to enhance the usability of satellite-based weather products. The performance of satellite-based weather products also varies depending on the climatic area. Kolluru et al. (2020) observed a greater accuracy of satellite precipitation products (SPPs) in tropical and humid regions than in arid and semiarid regions in India. Conducting evaluations of SPPs in specific areas of interest is crucial to consider the local topography and climatic conditions that could impact the accuracy of SPP estimation (Bitew & Gebremichael, 2011). Therefore, rigorous evaluation and site-specific validation of SPPs are necessary before utilizing them for various applications (Belay et al., 2019).

Tamil Nadu has diverse topography, ranging from coastlines to high-elevation hilly regions. The terrain features, including the elevated western Ghats and coastal areas along the Bay of Bengal, experience rainfall variability and extreme rainfall patterns that significantly impact the agricultural sector. To overcome the limited availability of gauge-based rainfall observations in these areas, remote sensing measurements of meteorological variables are desirable. Such measurements can provide rainfall data at high spatial and temporal resolutions. Therefore, it is important to evaluate different SPPs to understand their performance in Tamil Nadu. This evaluation will help describe long-term rainfall variability and analyse climate change-induced changes in rainfall characteristics. Reliable rainfall information derived from the best SPP can enable informed decision-making in sectors dependent on rainfall. However, there have been limited studies evaluating multiple satellite products for Tamil Nadu, and no studies have focused on the grid wise comprehensive evaluation and comparison of MSWEP, GPCC, CHIRPS, ERA5, and IMDAA products at different temporal scales. This present research is intended to provide a deep and inclusive understanding of SPPs as reliable sources of rainfall data for various applications in regions with limited data availability. This study aimed to investigate the ability of SPPs (MSWEP, GPCC, CHIRPS, ERA5, and IMDAA) to detect and estimate rainfall and identify the most suitable SPPs to enhance future research in climate system, effective water resource management, weather and water smart agricultural planning, designing weather-based crop insurance products, improved weather forecasting and developing climate change adaptation strategies.

Materials and methods

Description of the study area

Tamil Nadu is a state in southern India located on the southeastern coast. It covers an area of 130,058 km^2, with latitudes ranging from 8°25' to 13°5' N and longitudes ranging from 76°5' to 80°25' E. The state is divided into 38 districts, which are classified into seven agroclimatic zones: the western zone (WZ), southern zone (SZ), north-western zone (NWZ), north-eastern zone (NEZ), high rainfall zone (HRZ), high altitude and hilly zone (HAHZ), and Cauvery delta zone (CDZ), as shown in Figure 1. Tamil Nadu has a tropical climate influenced by the Bay of Bengal, Western Ghats, Northeast, and Southwest monsoons. The average annual rainfall is 945 mm, with the majority occurring during the Northeast monsoon (48 percent) and the Southwest monsoon (36 percent). The temperature ranges from 23.4°C to 33.8°C, with mean minimum and maximum temperatures, respectively (Prajesh et al., 2019). The state experiences distinct wet and dry seasons, with monsoonal rains bringing relief from extreme heat. The monsoon greatly impacts agriculture, affecting the onset and cessation of monsoons, prolonged droughts, quantity and distribution of rainfall, duration and frequency of dry and wet spells, and extreme rainfall events. Agriculture in Tamil Nadu is mainly rainfed, making monsoonal rainfall a crucial factor for crop cultivation. Rice cultivation is common in coastal areas, while cotton, pulses, and oilseeds are grown in dryland regions. Hilly areas are suitable for tea, coffee, and spices. In summary, Tamil Nadu has diverse geographical features and a unique tropical climate that significantly influences its agricultural sector.

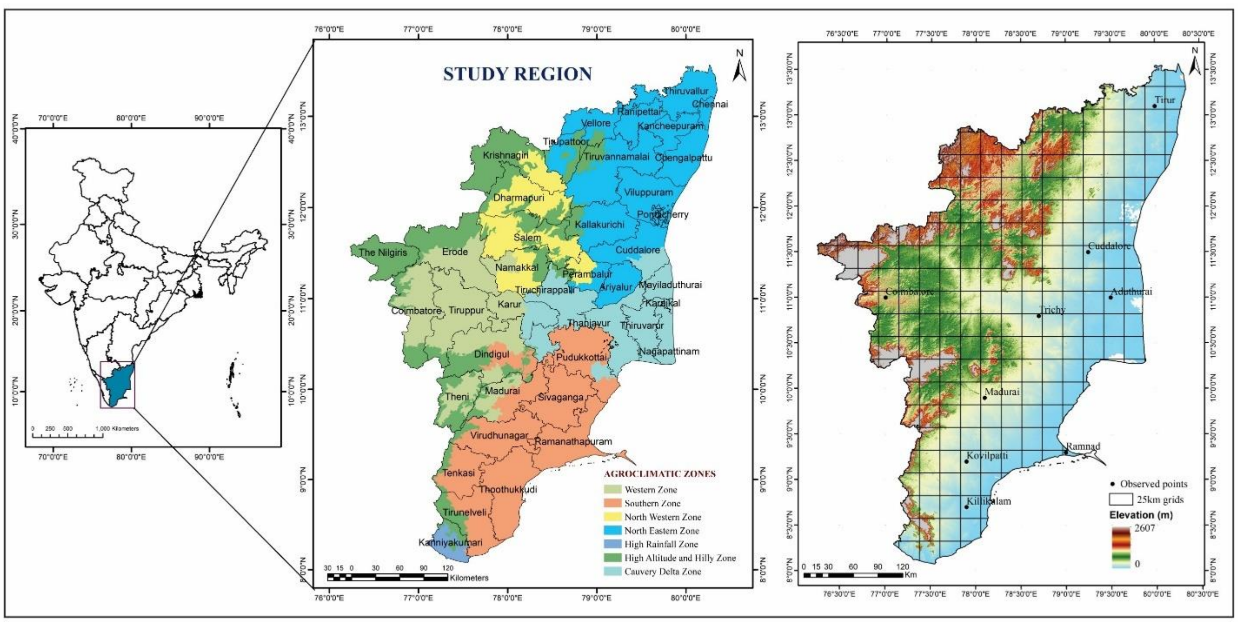

Fig. 1 The study region map in the left panel depicts the location and agroclimatic zones of the study area, and the map in the right panel shows the grids at a 25 km spatial resolution with the elevation range and distribution of ground station observation points in Tamil Nadu.

Satellite and reanalysis precipitation data

SPPs developed at global and regional scale are available in different file formats and at different grid resolution that do not overlap with station- or gauge-based gridded data (IMD gridded data). To eliminate the complexities in data extraction and for obtaining data for the same grid as the IMD, regridding was employed in the current study. All the downloaded satellite precipitation products are regridded to 0.25° spatial resolution to make it consistent and comparable with IMD. The Climate Data Operator (CDO) was used to regrid the data and commonly used tool to manipulate and analyse gridded data (Schulzweida, 2019). The nearest-neighbor method is frequently employed in precipitation analysis (Booth et al., 2018). This approach entails selecting the grid that is closest to the target grid, resulting in a mere shifting of the grid to align with the corresponding precipitation time series. The extracted satellite datasets are compared with IMD gridded data using various statistical metrics to test the accuracies and performances of these SPPs at daily, seasonal and annual timescales covering 18 x 22 grid points in different climatic zones of Tamil Nadu. The list of satellite precipitation products and their details are given in Table 1.

Table 1. Description of the precipitation products used in the study

Dataset	Reference	Spatial resolution	Institute
IMD	Pai et al., 2014	0.25° x 0.25°	India Meteorological Department
ERA5	Hersbach et al., 2020	0.1° x 0.1°	European Center for Medium Range Weather Forecasting
IMDAA	Ashrit et al., 2020	0.12° x 0.12°	National Center for Medium Range Weather Forecasting
MSWEP	Beck et al., 2017	0.1° x 0.1°	GLoH$_2$O
CHIRPS	Funk et al., 2015	0.05° x 0.05°	Climate Hazard Center
GPCC	Huffman et al., 2001	1° x 1°	Deutscher Wetterdienst

Ground-based rainfall observations

The ground-based observation data from 1991 to 2022 were collected from nine research stations across Tamil Nadu, covering Coimbatore, Cuddalore, Madurai, Ramnad, Thanjavur (Aduthurai), Tiruvallur (Tirur), Trichy, and Tuticorin (Killikulam and Kovilpatti) districts, which were used for comparison with the satellite, reanalysis datasets and IMD gridded datasets.

Satellite and reanalysis precipitation dataset evaluation using statistical metrics

Detection metrics (categorical statistics) and accuracy metrics (continuous statistics) that describe the detection capabilities and error characteristics of SPPs as well as reanalysis products are applied in the present investigation to perform statistical analysis between various precipitation products (MSWEP, GPCC, CHIRPS, ERA5, IMDAA) and reference datasets (IMD gridded dataset and gauge-based observation).

Detection metrics

Several detection metrics can be used to evaluate the rain detection capabilities of satellite products, namely, the false alarm ratio (FAR), probability of detection (POD) and misses. FAR indicates the satellite estimated rainfall where there is no ground observation. POD measures correctly detected rainfall in both the satellite and ground estimates. Misses determine rainfall not recorded by satellites but present in ground observations (Gosset et al., 2013, Sunilkumar et al., 2015).

$$\text{FAR}\ (\%) = \frac{\text{Number of days}\ \{(\mathrm{S_i} \geq 0.5)\}\ \&\ \text{Number of days}\ \{(\mathrm{R_i} < 0.5)\}}{\text{Number of days}\ \{(\mathrm{R_i} < 0.5)\}} \mathrm{X100}$$

$$\text{POD}\ (\%) = \frac{\text{Number of days}\ \{(\mathrm{S_i} \geq 0.5)\}\ \&\ \text{Number of days}\ \{(\mathrm{R_i} \geq 0.5)\}}{\text{Number of days}\ \{(\mathrm{R_i} \geq 0.5)\}} \mathrm{X100}$$

$$\text{Misses}\ (\%) = \frac{\text{Number of days}\ \{(\mathrm{S_i} < 0.5)\}\ \&\ \text{Number of days}\ \{(\mathrm{R_i} \geq 0.5)\}}{\text{Number of days}\ \{(\mathrm{R_i} \geq 0.5)\}} \mathrm{X100}$$

where S_i and R_i are the satellite- and rain gauge-based rainfall estimates, respectively, with a chosen threshold of 0.5 mm/day. A threshold of 0.5 mm d^{-1} was maintained for daily rainfall to eliminate the lower rainfall values that resulted from the interpolation of rainfall during the gridding process.

Accuracy metrics

Accuracy metrics, which include the percent bias (PBIAS), root mean square error (RMSE), index of agreement (d), coefficient of determination (R^2) and correlation coefficient (r), to determine the performance accuracy of the satellite products in estimating the precipitation amount and its variability from the reference or observed data. The RMSE has been used as a standard statistical metric to measure model performance in meteorology and climate research studies (Hodson et al., 2022). The data accuracy or the average error magnitude between the gauge and satellite products is indexed by the RMSE. It is a measure derived from the whole square root of the sum of squares of the differences between satellite and observed data divided by the number of total observations. Range: 0 to infinity, and perfect score: 0.

$$\text{RMSE} = \sqrt{\frac{\sum(\mathrm{S_i} - \mathrm{R_i})^2}{\mathrm{N}}}$$

PBIAS measures the average tendency of the satellite rainfall estimation values to be higher or lower than the observed rainfall. PBIAS values with a lesser magnitude are desired. A positive PBIAS indicates overestimation of satellite rainfall, while a negative PBIAS indicates underestimation of satellite rainfall products (Gupta and Nagar, 1999).

$$\text{BIAS}\ (\%) = \frac{\sum(\mathrm{S_i} - \mathrm{R_i})}{\sum \mathrm{R_i}}\ \mathrm{X}\ 100$$

The correlation coefficient (CC) denoted by r is used to measure the strength of the linear relationship between two variables (observed and predicted). Its value ranges from -1 to +1, where – 1 indicates a perfect negative relationship, +1 indicates a perfect positive relationship and 0 indicates no linear relationship (Ratner et al., 2009).

$$\mathbf{r} = \frac{\mathrm{n}\sum \mathrm{xy} - (\sum \mathrm{x})(\sum \mathrm{y})}{\sqrt{(\mathrm{n}\sum \mathrm{x}^2 - (\sum \mathrm{x})^2)(\mathrm{n}\sum \mathrm{y}^2 - (\sum \mathrm{y})^2))}}$$

The coefficient of determination (R^2) can be interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from −∞ to 1, indicating +1 as the best value.

$$R^2 = 1 - \frac{\sum(X - Y)^2}{\sum(\bar{Y} - Y)^2}$$

A value of 1 for the index of agreement (*d*) indicates good agreement between the simulated and observed data, while values closer to 1 indicate better predictions. A "*d*" value of zero indicates no predictability.

$$d = 1 - \frac{\sum_{i=1}^{n}(S_i - R_i)^2}{\sum_{i=1}^{n}(|S_i - \overline{R}| + |R_i - \overline{R}|)^2}$$

where $\overline{R}$ is the rain gauge observed mean value, S_i is the satellite estimated value, and R is the rain gauge observed value.

Results and discussion

Spatial distribution of rainfall by satellite and reanalysis products at annual and monsoonal scales

Spatial annual rainfall pattern

The spatial patterns of annual and predominant monsoons, *viz.*, southwest (June-September) and northeast (October-December) rainfall, were evaluated by performing gridded analysis and visual observations. The results revealed that approximately 988 mm of average annual rainfall was recorded by the IMD over Tamil Nadu; on the other hand, the SPPs, *viz.*, MSWEP, GPCC, and CHIRPS, estimated the annual rainfall to be approximately 1059, 1038, and 1186 mm, respectively. The reanalysis dataset ERA5 showed an annual rainfall of 1113 mm, and it was 1241 mm with the IMDAA. The IMDAA estimated the maximum annual average rainfall compared to the other products and showed greater deviation from the IMD than other satellite and reanalysis precipitation products.

The spatial distributions of the mean annual rainfall estimated from the MSWEP, GPCC, CHIRPS, ERA5, IMDAA and IMD datasets are presented in Fig. 2. According to the IMD, the average annual rainfall varies from 500 to 2250 mm in Tamil Nadu. Fig. 1 shows that the northeastern and Cauvery delta regions, which are situated near the east coastline of Tamil Nadu, receive high amounts of rainfall ranging from 1000 to 1500 mm. The hilly areas adjacent to the Western Ghats also exhibit high rainfall ranging from 1250 to 2000 mm, except for very few pockets with rainfall above 2000 mm. The spatial distributions of the annual mean rainfall derived from ERA5, MSWEP, and GPCC exhibited similar patterns to those of the IMD, with slight underestimations of rainfall over some portions of the Cauvery delta and eastern zones and overestimations over the hilly regions. IMDAA overestimated the rainfall compared to all products over the major area in Tamil Nadu.

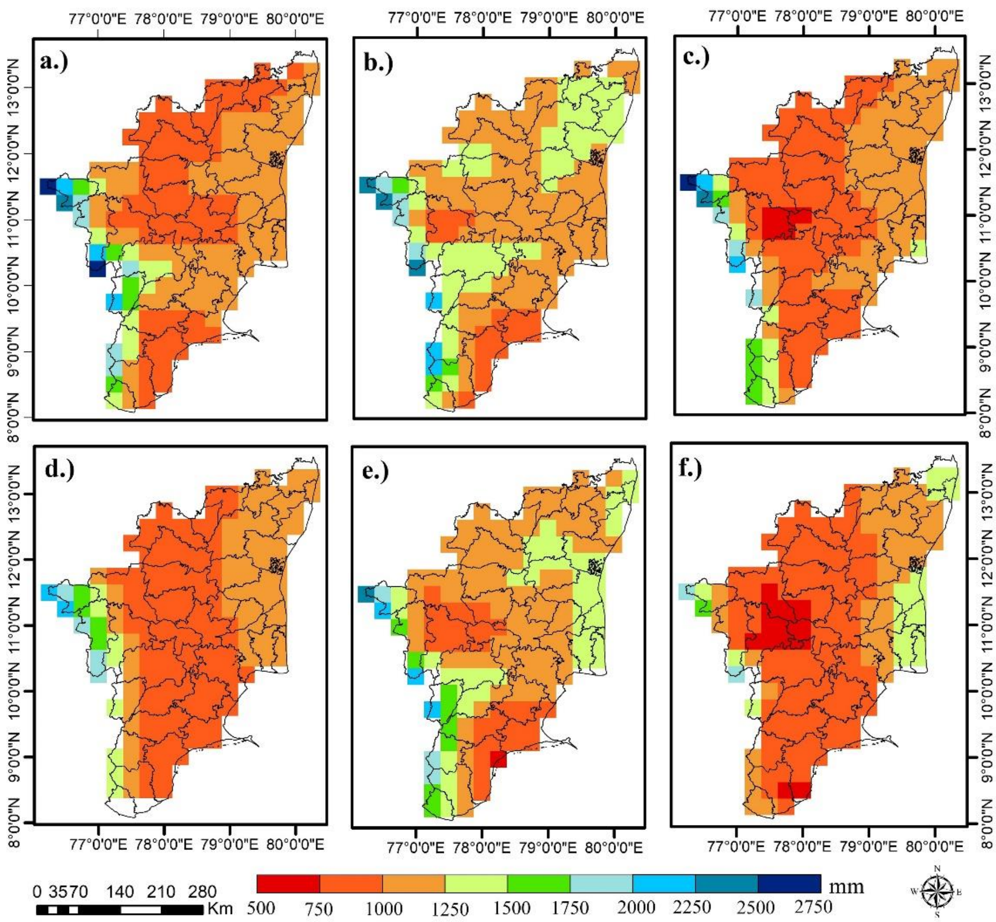

Fig. **2** Spatial distribution of annual mean rainfall in satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS, f. IMD) and heatmap showing the number of grids under each rainfall category

The heatmap shows (Fig. 3) that the total number of grid cells falls under each rainfall range with respect to each precipitation product. A lower range of 500–750 mm of rainfall is observed in minimum grids with IMD (13), MSWEP (5), and CHIRPS (1) in the western and southern parts of Tamil Nadu. The remaining products showed this low rainfall range in none of the grids. More than 80% of the grids cover average annual rainfall ranging from 750 to 1250 mm with GPCC (159 grids), MSWEP (152 grids), ERA5 (151 grids) and IMD (143 grids). IMDAA (123 grids) and CHIRPS (119 grids) also estimated similar rainfall quantity ranges in a significant number of grids distributed in 70% of the grids. Although the IMDAA (51) and CHIRPS (42) datasets exhibited 1250-1500 mm of precipitation over approximately 25% of the grids, the other four datasets, viz., the MSWEP (8), GPCC (8), ERA5 (10), and IMD (18) datasets, showed this rainfall in a very minimum area on the order of 10% of the grids. More than 1500 mm was detected in a very small number of grids comprising hilly and high-rainfall zones.

In Tamil Nadu, the average annual rainfall is observed to be the highest in the western and eastern regions, and less rainfall occurs in the central region, northwestern and southern areas. The higher rainfall pattern concentrated over the western and eastern regions of Tamil Nadu could be attributed to the orographic effect of the western Ghats Mountains near the western region and east coastline in the eastern part of Tamil Nadu, which is influenced by monsoonal winds and cyclonic activities (Phadtare, 2023). MSWEP, ERA5 and GPCC performed well in capturing the spatial rainfall variability of average annual rainfall with a close similarity to that of the IMD, whereas IMDAA overestimated the annual rainfall over a larger area.

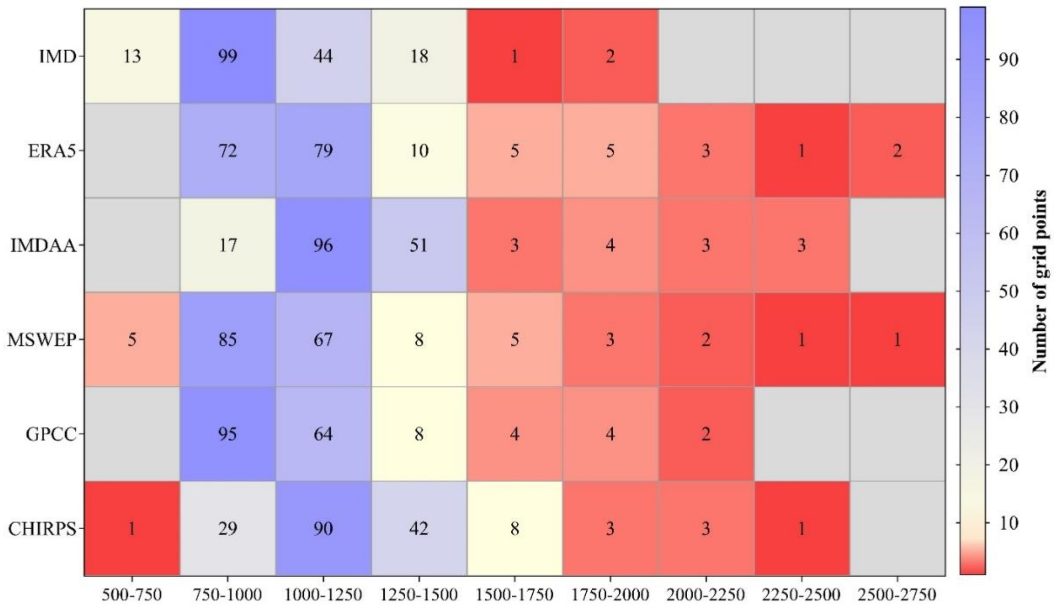

Fig. **3** Heatmap represents the number of grids in each rainfall category under various precipitation products

Spatial seasonal rainfall pattern

During the southwest monsoon, rainfall tends to increase from the southeast to the northern regions, while the hilly region in the western parts and the high-rainfall zone at the southern tip of Tamil Nadu receive more rainfall in the SWM across all the products (Fig. SI1). In the southeastern and central parts, less than 350 mm of rainfall occurs, whereas in the northern parts, the SWM rainfall ranges from 350 to 650 mm. In the hilly zone, rainfall reaches 1500 mm with the IMD. The rainfall ranging from 50 to 350 mm was observed in approximately 50% of the grids with CHRIPS and GPCC and approximately 40% of the grids with MSWEP. According to the ERA (122 grids) and IMDAA (110 grids) SWM recorded rainfall varying from 350 to 650 mm in 60 to 70% of the grids, followed by the MSWEP (98 grids: 55%) and CHIRPS (79 grids: 47%). Higher rainfall of more than 950 mm was observed in less than 10 grids in all products except for the IMDAA (12 grids) (Fig. SI2a).

The NEM, which is the chief rainfall season, receives rainfall ranging from 350 to 950 mm (Fig. SI3) in the majority of the regions. The spatial pattern of the NEM rainfall pattern manifests more rainfall over the northeast and cauvery delta regions, ranging from 650 to 950 mm for the IMD, whereas other products show rainfall ranging between 350 mm and 650 mm in the northeastern parts, except for a few patches. The NEM rainfall shows an increasing pattern from the western and southern regions towards the northeast side. More than 100 grid points in ERA5, IMDAA, MSWEP, GPCC, and CHIRPS showed rainfall amounts ranging from 350 to 650 mm, while fewer grid points in IMD (13) and MSWEP (2) predicted more than 950 mm of rainfall in the coastal regions of Tamil Nadu (Fig. SI2b).

During SWM, the spatial rainfall pattern exhibited less than 350 mm in the southeastern, southern and central regions, while 350 to 650 mm of rainfall occurred in the northwestern and northeastern zones. Higher rainfall is observed in hilly and high-rainfall zones (Fig. SI1). In NEM, as per the IMD, most of the grids have average rainfall of 350 to 950 mm covering the entire Tamil Nadu region, excluding a few coastal regions where NEM rainfall exceeds 950 mm. On the other hand, all other models showed 350 to 650 mm of rainfall in the greater portion of the area with higher rainfall of 650 to 950 mm in the coastal regions (Fig. SI3). The higher rainfall estimated by the majority of the products in coastal regions might be due to the strong influence of monsoonal behaviour in association with frequent exposure to rainfall extremes and cyclones.

All the products exhibited relatively **c**onsistent spatial SWM rainfall patterns with varying magnitudes of rainfall. MSWEP estimates of SWM rainfall ranging from 50 to 650 mm in 92% of the grids showed the best match with the SWM rainfall spatial pattern of the IMD, which exhibited a similar range in 95% of the grids. The present findings are in line with Reddy et al. (2022), who reported that MSWEP outperformed in estimating the precipitation across the Godavari River basin in India among the other precipitation products, viz., CHIRPS, TRMM, CPC, CMORPH, and PERSIANN-CDR evaluated against the IMD.

MSWEP, GPCC, CHIRPS, and IMD better captured the influence of NEM over the coastal regions of Tamil Nadu and estimated the higher rainfall over those regions compared to other regions during NEM. ERA5 and IMDAA failed to estimate higher rainfall over the coastal regions, and IMDAA exhibited a rainfall range of 350 to 650 mm in almost the entire (90% of the grids) Tamil Nadu region. The seasonal rainfall distribution shows that the study region receives the majority of the rainfall during the SWM and NEM seasons, with NEM contributing more to the rainfall in the northeastern regions and SWM contributing more to the rainfall in the western regions. This could be due to the seasonal wind swings during the monsoon period and the orography that exists due to the presence of the Western Ghats (Jegankumar et al., 2012).

Evaluation of the rain detection capabilities of satellite and reanalysis products

False Alarm Ratio

Detection metrics such as the FAR, POD and percentage of missing rainfall events were computed using daily rainfall data to understand the abilities of satellite and reanalysis products to detect daily rainfall events precisely. The spatial variation in the FAR of all five precipitation products was obtained by comparing the reference data (IMD) and is illustrated in Fig. 4. The results indicate that the FAR for the majority of the products followed a decreasing pattern from the lower portion of Tamil Nadu towards the upper regions, i.e., the satellite and reanalysis products exhibited good performance in detecting rainfall events while moving from the southern to the northern direction in Tamil Nadu. However, the spatial distribution of the FAR exhibited only two distinct patterns, indicating a lower FAR in the upper regions and a high FAR in the lower portions and not showing much variation within each portion. At the same time, CHIRPS had almost the same magnitude of FAR over Tamil Nadu. Rainfall events detected by CHIRPS matched well with the rainfall events of the IMD across Tamil Nadu, with a FAR of 10-30%. ERA5 and IMDAA attained a lower FAR over the northwestern areas, GPCC and CHIRPS achieved a lower FAR in the coastal areas, while MSWEP had an almost similar FAR over these regions. CHRIPS performed better, with a lower FAR ranging from 10-20% FAR in 13% of the grids and 20-30% FAR in 87% of the grids (154), while GPCC obtained FAR values ranging from 10 to 20% in 32% of the grids (58) and 20-30% in 47% of the grids (84). The MSWEP showed FAR values ranging between 20 and 30% for the maximum number of grids (60% of grids) and between 30 and 40% for 28% of the grids. ERA5 exhibited a lower FAR (20-30%) in fewer grids (18), a 30-40% FAR in 60% of the grids (107), and more than 40% FAR in the remaining grids. In the case of the IMDAA, all the grids displayed a FAR greater than 40% (Fig. SI4a).

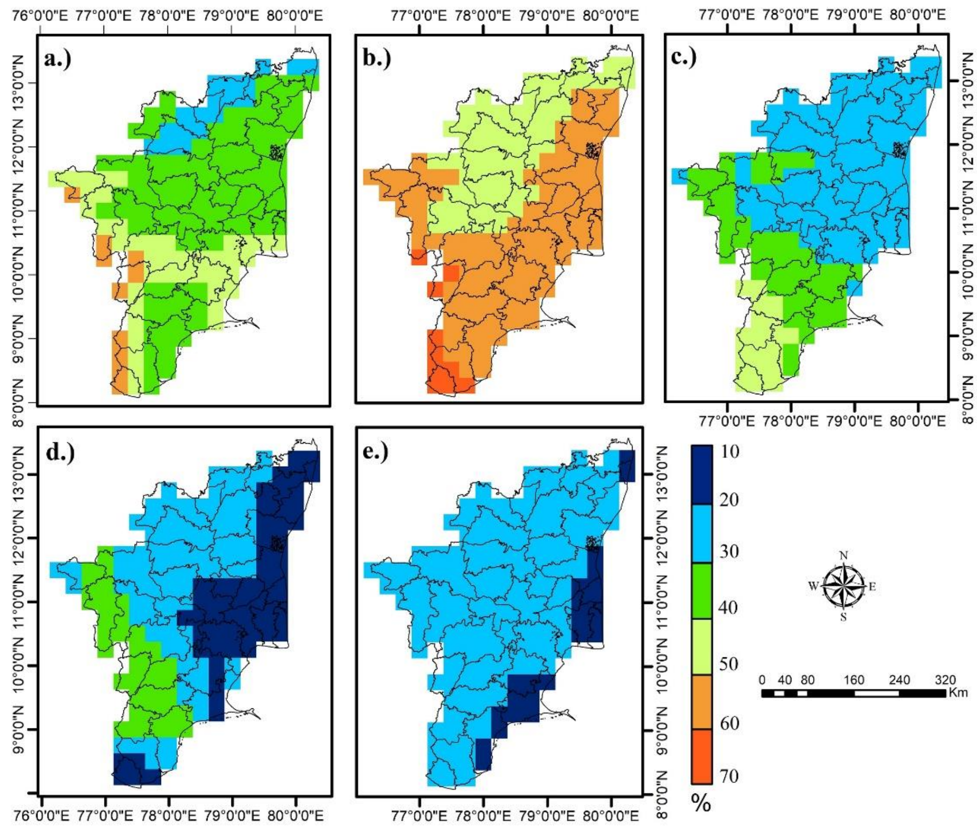

Fig. 4. Spatial distribution of the false alarm ratio (FAR) in the satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS).

Probability of detection (POD)

The spatial pattern of the probability of detection and heatmap showing the number of grid points under each specified POD range are given in Fig. 5. The spatial pattern of POD indicated the best detectability of ERA5 and MSWEP, with more than 70% POD covering 90% of Tamil Nadu; however, not much heterogeneity was found in POD between the regions. CHIRPS also displayed almost evenly distributed spatial POD values within the range of 50 to 60% in a larger area (76% grids). Similarly, a uniform distribution of PODs ranging from 70 to 80% was observed with the IMDAA in 74% of the grids, and the rest of the area had 60 to 70% PODs. The GPCC achieved 60 to 70% POD in 79 of the 177 grids (Fig. SI4b). The majority of the products performed better in identifying rainfall events in hilly rainfall zones than in identifying rainfall events in other zones.

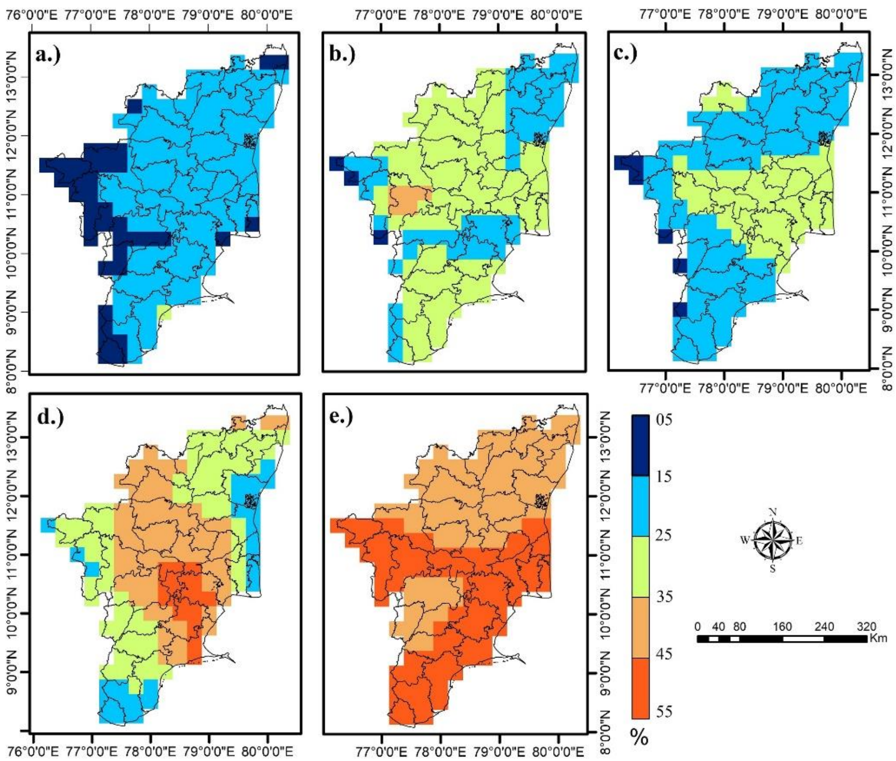

Fig. 5. Spatial distribution of the Probability of detection (POD) in the satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS)

Misses (%)

The spatial pattern of misses is depicted in Figure 6. ERA-5 exhibited good performance, with a miss percentage below 25% over Tamil Nadu, followed by the MSWEP, with a lower miss rate of 15 to 25% in the majority of the areas, and it reached 35% in some regions, including the Cauvery delta and parts of the southern zone. The number of missed rainfall events was greater in the CHIRPS, with 45 to 55% of events occurring in the lower half of Tamil Nadu from the central region, while in the upper half, the number of missed events decreased to 35%. The results showed that the spatial variation in the percentage of misses ranged from 5 to 55% in all the products, with a lower miss percentage of 5 to 15 in ERA5 over 38 grids in the western and southwestern regions of Tamil Nadu. Both ERA5 and MSWEP had percentage of misses within 35% of the range. The percentages of missing in GPCC were 25 to 35% and 35-45%, respectively, in an equal number of grids (68). The miss percentage shows a decreasing trend towards the western side, excluding CHIRPS, which had miss percentages of 35–45% in 88 grids and 45–55% in 89 grids over the Tamil Nadu region (Fig. SI4c).

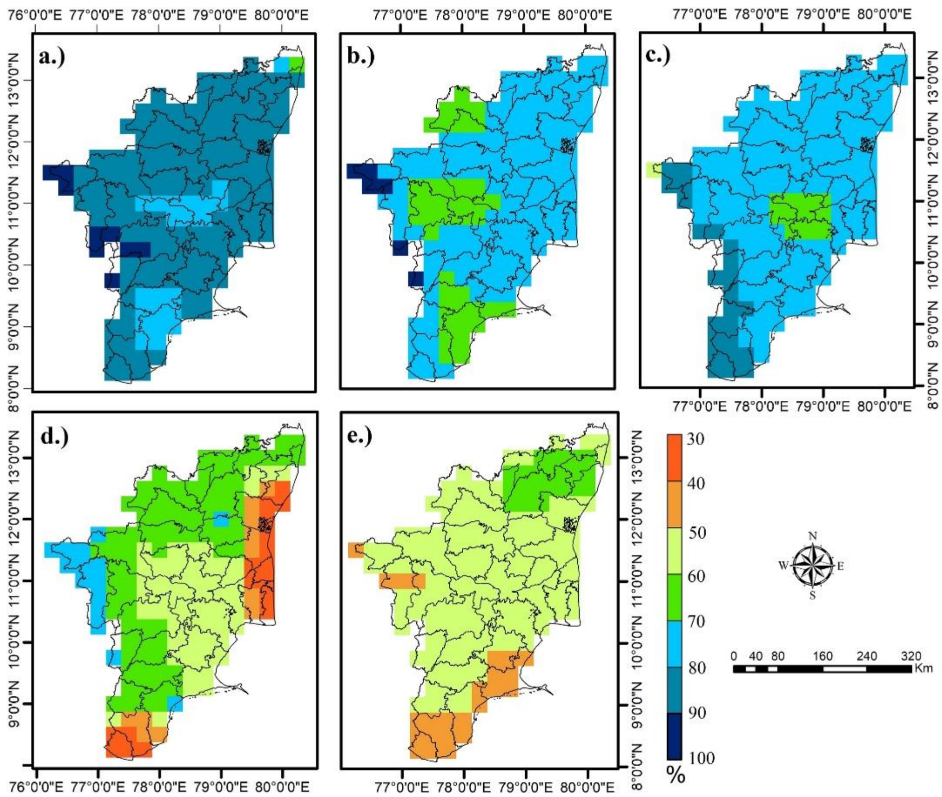

Fig. 6. Spatial distribution of the misses in the satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS)

The results of the detection metrics indicated that the POD was greater in ERA5 and MSWEP, revealing better agreement with the reference dataset, whereas the POD was lower in CHIRPS, indicating poor detection capacity. The percentage of missing rainfall events was lower in the ERA5 and MSWEP datasets and higher in the CHIRPS dataset. ERA5 and MSWEP exhibited the best performance in detecting rainfall events as a result of their high POD and low number of missing rainfall events. It is evident from the results that the products with fewer missing rainfall events had a higher POD. The CHIRPS, GPCC and MSWEP products had FAR values less than 40% at more than 90% of the grid points under study and performed better than the other products. An evaluation study of SPPs conducted by Reddy and Saravanan (2023) in India also reported that MSWEP and CHIRPS had lower false ratios than the other precipitation products. Additionally, it is important to note that high spatial resolutions allow satellites to identify atmospheric processes more precisely, which would improve rainfall estimation (Alfieri et al., 2022). The spatial distribution of the probability of detection (POD) indicated that both MSWEP and ERA5 had PODs of detecting 70-100% of the rainfall at more than 90% of the grid points. Kolluru et al. (2020) reported that ERA-5 had a high POD and poor performance of CHIRPS in five diverse Indian climatic zones. Similar results could be observed from the study conducted by Taye et al. (2023) which showed the lowest detection capacity of CHIRPSv2. The variation in the performance of different products could be linked to several factors, such as the input data used and assimilation methods applied for developing reanalysis datasets, sensor type and accuracy, data retrieval algorithms used for satellite-based products, different schemes adopted and the structure of the numerical models (Mulungu & Mukama, 2023).

Evaluation of the rain estimation capabilities of satellite and reanalysis products

Root mean square error (RMSE)

The spatial map of the RMSE for daily rainfall, along with the heatmap showing the number of grids of the RMSE presented in Fig. 7, indicates that all products had higher RMSEs in the areas receiving more rainfall, i.e., in the mountainous and coastal regions and high-rainfall zone. The results revealed a lower RMSE ranging from 7 to 11 mm in the entire middle portion, covering the region from north to south of Tamil Nadu, while in the mountainous and coastal regions and high-rainfall zone, the RMSE ranged from 11 to 17 mm across the products. ERA5 and MSWEP were more effective at estimating the daily rainfall than the other products, with 7-11 mm in more than 70% of the grids. The RMSEs of GPCC (80 grids) and CHIRPS (78 grids) were 9-11 mm for the maximum number of grids. All the products had errors in the RMSE range of 9–11 mm for more than 50 grid points, and the IMDAA showed this RMSE range for the highest number of grids (102).

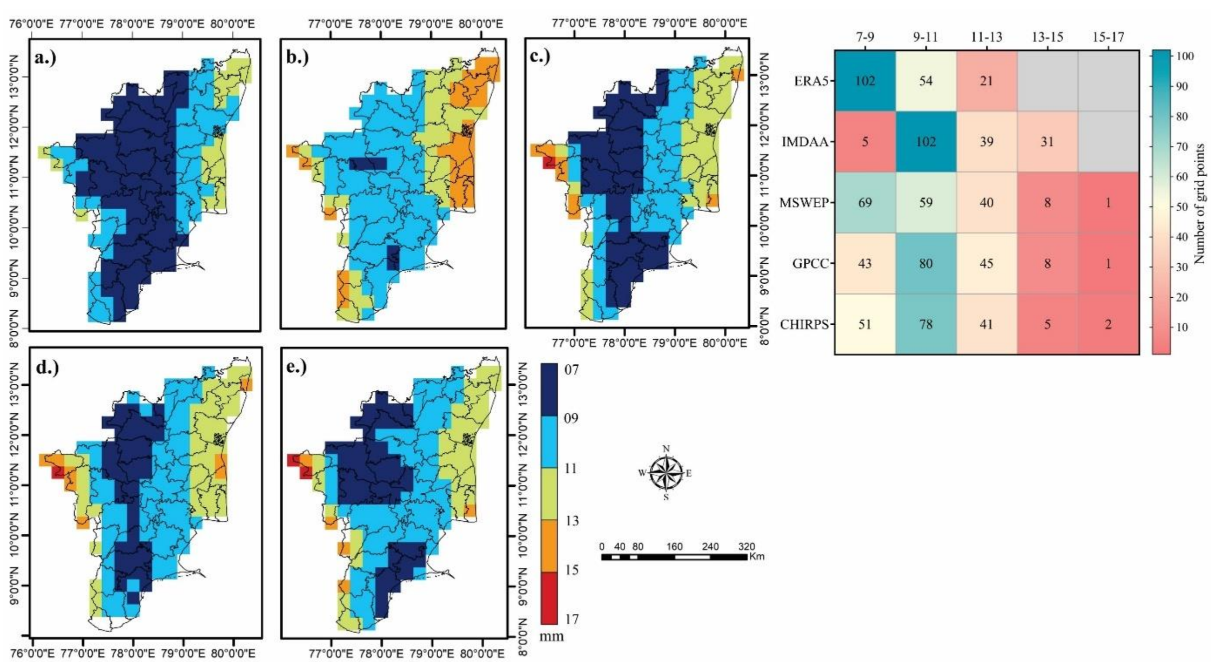

Fig. 7 Spatial distribution of the RMSE of daily rainfall in the satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS) and heatmap showing the number of grids in each RMSE category

The spatial pattern of the RMSE of the annual rainfall (Fig. 8) revealed that the western Ghats and the hills that received a good amount of rainfall from the SWM had higher RMSEs compared to other regions that received less rainfall. The RMSEs in the northwestern and northeastern parts of Tamil Nadu were low, and the RMSEs in the southern parts increased. The spatial RMSE distribution for the SWM was also consistent with the RMSE pattern of the annual rainfall (Fig. SI5). On the other hand, in NEM, the spatial RMSE distribution pattern showed a different pattern, exhibiting high RMSEs over the northeast and delta zones, which have close proximity to the coastline and receive more rainfall during NEM (Fig. SI6). However, in the remaining region, the RMSE did not vary much spatially, with the exception of a smaller area over the central part of Tamil Nadu during NEM. The lowest RMSE range of 180–380 was found to be the maximum in the MSWEP (110), followed by the GPCC (103) and ERA (82), while the IMDAA showed the highest RMSE range of 380–580 mm at the maximum grid points, and further the IMDAA obtained a high RMSE value ranging from 980 to 1180 mm across the six grids, accounting for 3.4% of the total grids in the western and southwestern regions.

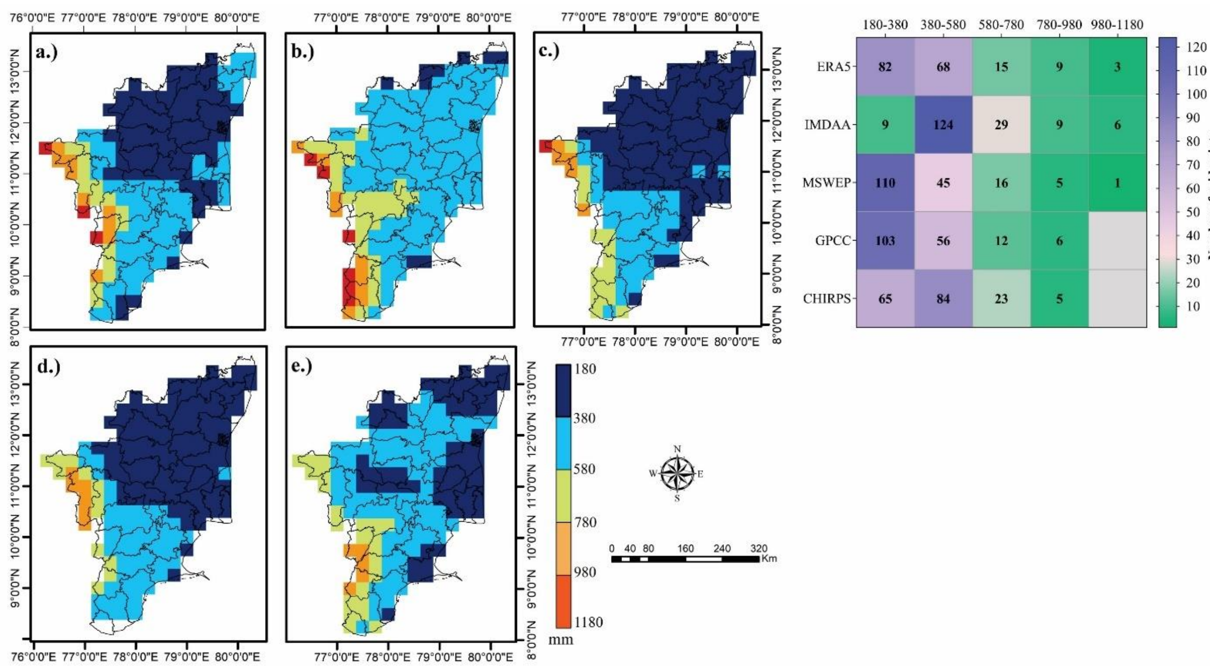

Fig. 8 Spatial distributions of the RMSEs of annual rainfall in satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS) and heatmaps showing the number of grids in each RMSE category

During SWM rainfall, the RMSE of the MSWEP, ERA5, and GPCC models exhibited similar patterns at the spatial scale, with lower values ranging from 100–200 mm on the northwestern northeastern side and higher RMSE values ranging from 700–800 mm in the western and southwestern areas (Fig. SI5). The maximum (more than 600 mm) error in rainfall was noted over the western part with all satellite products. The number of grid points with a lower RMSE range of 100–200 mm is observed in MSWEP (94), followed by ERA5 (85), GPCC (83) and CHIRPS (61) out of 177 grid points (Fig. 9a). The IMDAA had an error of 200–300 mm in the maximum number of (103 grid) points, followed by CHIRPS in 71 grid points. In NEM, MSWEP, CHRIPS and GPCC exhibited minimum RMSEs between 100 and 200 mm in 48, 66 and 52 grids, respectively, demonstrating around 30% of the grids, while these products had RMSEs of 200-300 mm in about 55% of the grids. All products had RMSEs of 200-300 mm for the maximum number of grids during NEM (Fig. 9b).

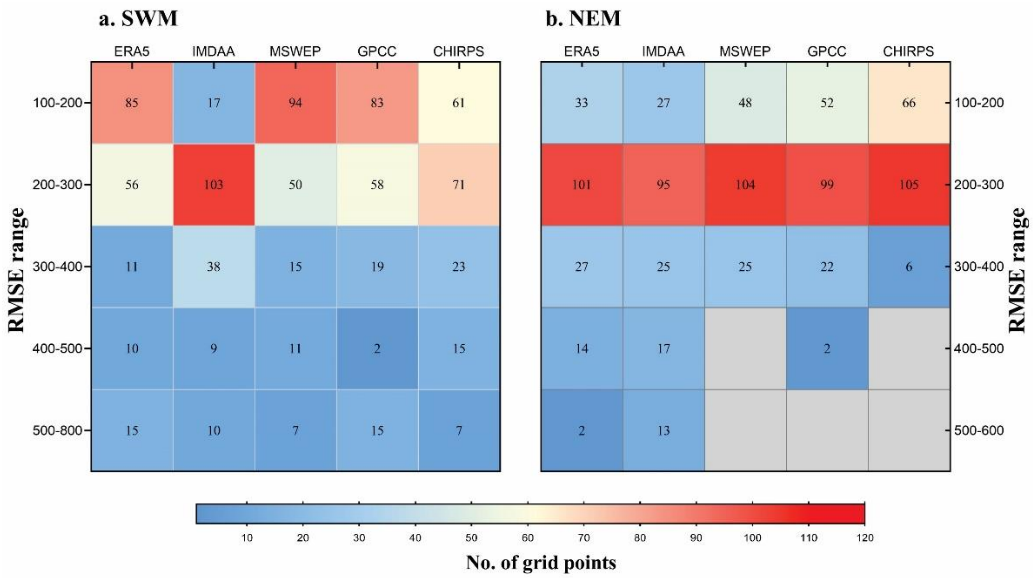

Fig. **9** Heatmap represents the number of grids in each RMSE category under various precipitation products (a. SWM and b. NEM)

The root mean square error (RMSE) serves as a straightforward metric to assess the average magnitude of errors in predictions, irrespective of their direction. In general, high values of root mean square error (RMSE) indicate inadequate predictive ability of models or satellite products. It is important to address these issues to enhance the quality of prediction in satellite products.

RMSEs in the range of 11 to 17 mm for daily rainfall were observed over the hilly and coastal regions, whereas inland regions exhibited RMSEs ranging between 7 and 11 mm across the products. The RMSE is generally larger for all products at all daily, annual and seasonal scales across the contiguous hilly regions of Western Ghats and coastal regions of Tamil Nadu. In particular, the results clearly manifested the higher RMSE values in the hilly regions adjacent to the western Ghats during the SWM while in the coastal regions during the NEM. The study of Willmott and Matsuura (2006) revealed that the RMSE is affected by geographic region, time period, and outliers in the data. Among the products, ERA5 performs better, with a lower RMSE in the daily time step, followed by MSWEP. However, the MSWEP could estimate rainfall better than the ERA at annual and seasonal time scales. The continuous statistical results confirmed that ERA-5 is a good dataset for daily time steps but is not effective at the monthly scale (Kolluru et al., 2020).

Percent Bias

The spatial patterns of the bias percentages of the annual and seasonal rainfall for all satellite products are shown in Fig. 10. The results revealed that the SPPs and reanalysis products underestimated the annual rainfall in the northeast region, especially in the coastal regions, with a bias percentage of 0 to -25%, while the annual rainfall was overestimated (>50% bias) in certain areas in the western and southwestern regions of the Western Ghats side of Tamil Nadu. In SWM, PBIAS showed an analogous spatial pattern to annual rainfall in the hilly regions of the western and southwestern parts of Tamil Nadu (Fig. SI7), and during NEM, the products underestimated the rainfall in the entire

Tamil Nadu region except for a few southern pockets. The products estimated that NEM rainfall exhibited a larger negative deviation from the IMD over some parts of the northwestern and northeastern regions (Fig. SI8).

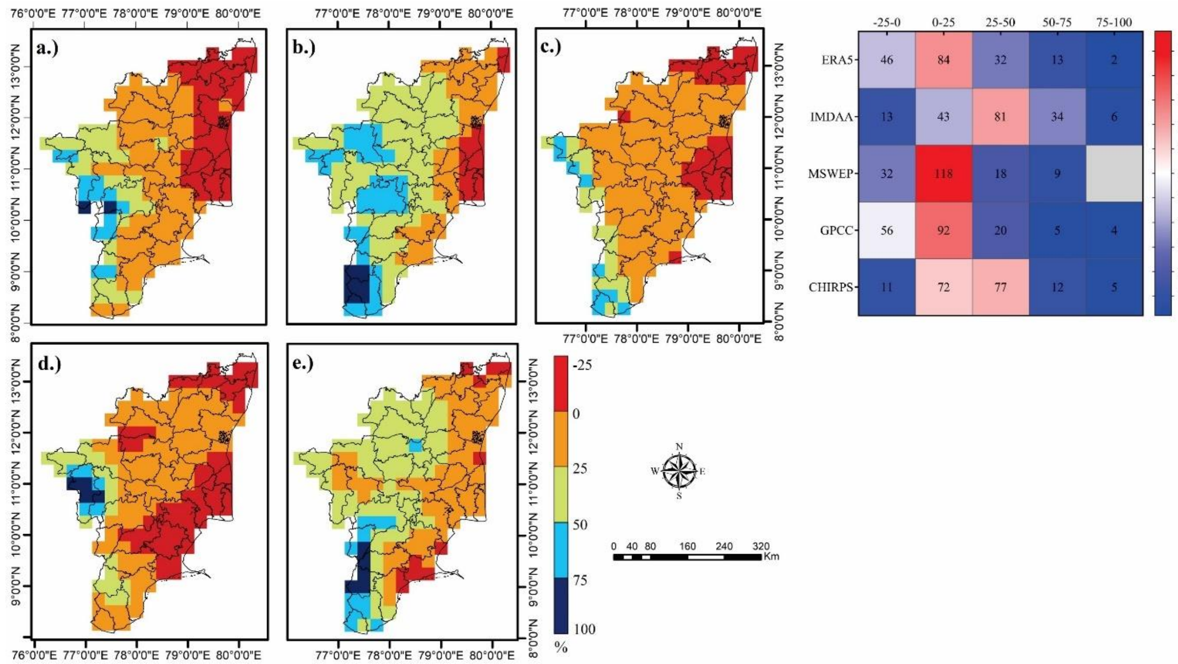

Fig. 10 Spatial distribution of the percent bias of annual rainfall in satellite and reanalysis products (a. ERA5, b. IMDAA, c. MSWEP, d. GPCC, e. CHIRPS) and heatmap showing the number of grids in each PBIAS category

For annual rainfall, a lower bias percentage of 0 to 25 was noted at the maximum grid points of MSWEP (118), followed by GPCC (92), ERA5 (84) and CHIRPS (72 grid points) (Fig. 10). Among the products, MSWEP had a lower bias (-25 to +25%) over the maximum grids of 150, covering 84.7% of the region. In the SWM, the MSWEP performed better, with a percent minimum bias (0 to 25%) in 98 grids, followed by the ERA5 (77) and GPCC (71 grid points). Both IMDAA and CHIRPS had biases of more than 25 percent in the maximum number of grid points compared to the other products (Fig 11a). In NEM, a smaller underestimated bias by CHRIPS demonstrated a better performance with a lower bias range of 0 to 20 PBIAS in 126 grids, covering 71.2% of the region, and a parallel performance was observed in MSWEP (111). GPCC showed a negative bias within the lower PBIAS range from 0 to 20% in 107 grids (Fig 11b).

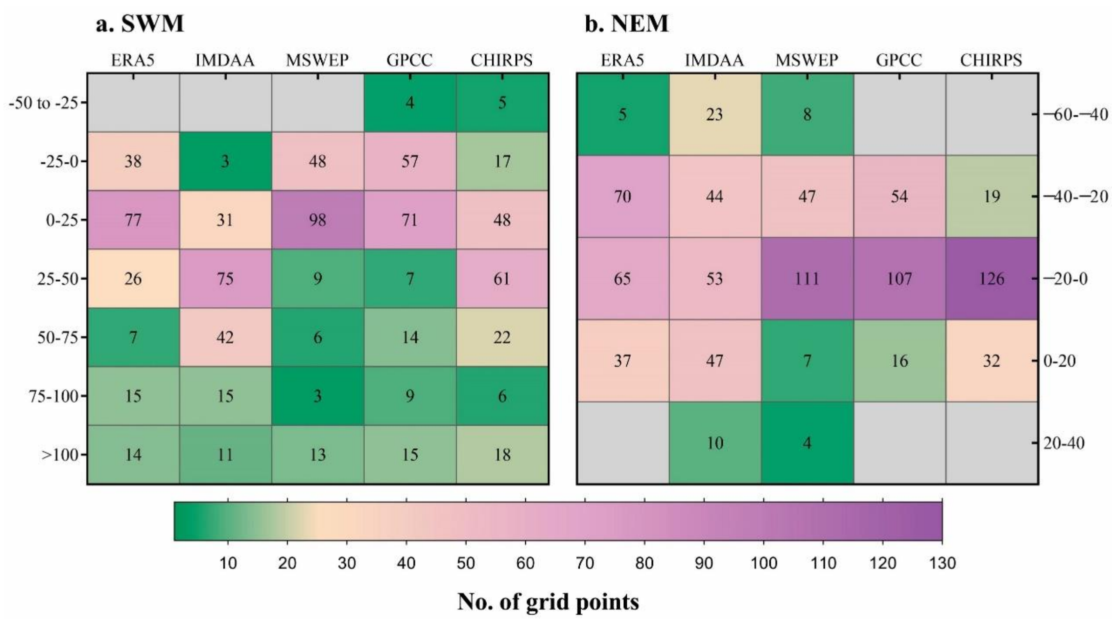

Fig. **11** Heatmap represents the number of grids in each percent bias category under various precipitation products (a. SWM and b. NEM)

The rainfall estimation accuracy metrics (continuous metrics) exhibited variations in the seasonal performances of the products, indicating better estimates of CHRIPS in NEM than in SWM. The SPPs and reanalysis data exhibited weak performance with an overestimation of rainfall over hilly and highly elevated regions of the Western Ghats. The rainfall detection and estimation ability of infrared and microwave sensors and retrieval algorithms employed in SPPs might have impacted the accuracy of SPPs in mountainous regions. Many studies have shown that SPPs do not perform efficiently in high-elevation regions (Yin et al., 2008, Ngo-Duc et al., 2013, Toté et al., 2015, Hobouchian et al. 2017). Reanalysis datasets blend global and regional weather models with observations to create reliable historical datasets at various spatiotemporal resolutions. A larger bias in the reanalysis rainfall datasets over mountainous terrain than over other regions might be associated with sparse rain gauge observations and complicated terrain. Although reanalysis data combine rain gauge-based observations with weather models to generate reanalysis data with increased accuracy, factors such as high elevation, complex topography and insufficient ground observations can lead to potential errors and limit the production of high-quality rainfall datasets. These findings are in conformity with Luo et al. (2019), who indicated that complex topography and inadequate observations can tend to increase the biases in rainfall estimation.

Validation of the precipitation products with ground station observations

Continuous statistical metrics such as the correlation coefficient (r), coefficient of determination (R^2), index of agreement (*d*), RMSE and PBIAS were employed to evaluate the performances of the satellite and reanalysis precipitation products against the nine ground-based observations made over Tamil Nadu. The precipitation products were evaluated for both the seasonal (i.e., SWM and NEM) and annual time scales.

The correlation analysis between the ground station observations and various precipitation products indicated a positive correlation for all the SPPs. Among the products, the MSWEP demonstrated a highly positive correlation followed by IMD at all annual and during the two monsoon periods (Table 2). The annual rainfall of the MSWEP at all rain gauge stations was strongly positively correlated (0.61 to 0.88) with the ground station observations, except for two-gauge stations (0.49 and 0.52) which had significantly weak positive correlation. Similarly, seven stations had significantly strong positive correlations (0.74 to 0.91) in NEM and SWM (0.53 to 0.87) for MSWEP. In the case of the IMD, one stations only showed a significantly strong positive correlation for annual rainfall (0.6) as well as in the SWM (0.7) and seven stations (0.61 to 0.87) in the NEM. Among all the products, the IMDAA was weakly correlated with the ground station observations.

For MSWEP, the positive correlation with ground observations was significant at all locations annually and NEM whereas positive correlation was nonsignificant at two locations in SWM. IMD had a significantly positive correlation for all stations only in NEM, five stations at annual scale and four locations in SWM. The correlation between the MSWEP and gauge data was better than the correlation between the IMD and gauge data at all-time scales at all locations. Among all precipitation products MSWEP had a lower RMSE followed by IMD (Fig. 12) and PBIAS values was lower with IMD followed by MSWEP (Fig. 13). Comparing the seasons, the higher correlation coefficient, lower PBIAS and RMSE were higher with NEM than SWM. MSWEP had a good agreement with gauge data than IMD with gauge data at all the time scales in all locations.

Table 2. Correlation statistics between precipitation products and ground observations (correlation significance was tested statistically only for the top performing precipitation product with higher r values (MSWEP) and IMD)

Locations	IMD	ERA5	IMDAA	MSWEP	GPCC	CHIRPS	IMD	ERA5	IMDAA	MSWEP	GPCC	CHIRPS	IMD	ERA5	IMDAA	MSWEP	GPCC	CHIRPS
	Annual						SWM						NEM					
Aduthurai	0.58**	0.78	0.48	0.78**	0.75	0.56	0.4*	0.64	0.46	0.69**	0.66	0.63	0.74**	0.72	0.68	0.76**	0.74	0.58
Coimbatore	0.42*	0.59	0.39	0.82**	0.85	0.63	0.03	0.45	0.71	0.76**	0.67	0.63	0.81**	0.65	0.35	0.88**	0.85	0.7
Cuddalore	0.6**	0.65	0.42	0.74**	0.74	0.72	0.4*	0.54	0.55	0.78**	0.7	0.69	0.54**	0.67	0.51	0.78**	0.74	0.73
Killikulam	0.27	0.27	0.35	0.49**	0.74	0.46	-0.3	0.06	0.02	0.32	0.36	-0.01	0.61**	0.65	0.61	0.53**	0.54	0.65
Kovilpatti	0.14	0.55	0.51	0.65**	0.79	0.58	-0.26	0.53	0.56	0.36	0.44	0.37	0.47*	0.45	0.41	0.46*	0.52	0.63
Madurai	0.41*	0.49	0.39	0.73**	0.6	0.51	0.27	0.31	0.51	0.53**	0.34	0.32	0.87**	0.39	0.44	0.74**	0.77	0.69
Ramnad	0.04	0.75	0.45	0.88**	0.79	0.62	0.03	0.58	0.41	0.87**	0.82	0.64	0.66**	0.76	0.61	0.89**	0.74	0.73
Tirur	0.27	0.67	0.43	0.52**	0.56	0.36	0.7**	0.7	0.67	0.76**	0.71	0.52	0.82**	0.78	0.56	0.91**	0.91	0.82
Trichy	0.46*	0.32	0.19	0.61**	0.46	0.3	0.41*	0.39	0.28	0.59**	0.43	0.33	0.81**	0.53	0.49	0.8**	0.78	0.61

*. Correlation is significant at the 0.05 level (2-tailed), **. Correlation is significant at the 0.01 level (2-tailed)

Categorisation of the strength and nature (positive/negative) of correlation: .00-.19 “very weak”, .20-.39 “weak” • .40-.59 “moderate” • .60-.79 “strong” • .80-1.0 “very strong” as per Evans (1996).

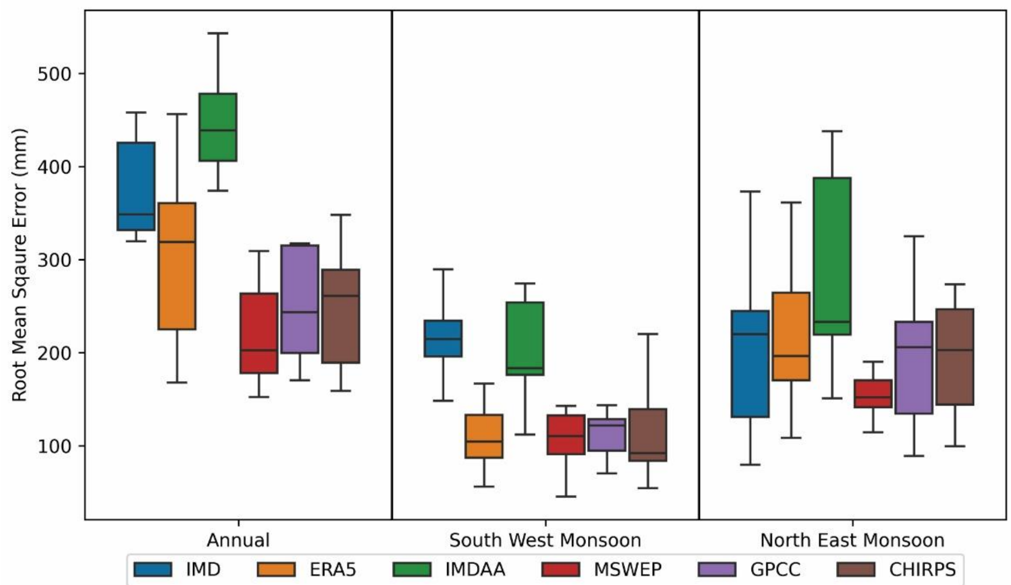

Fig. 12 Spatial (across nine-gauge stations) RMSE distribution of the annual rainfall, Southwest monsoon (SWM) and Northeast monsoon (NEM) for satellite and reanalysis products in comparison with the gauge rainfall data. In a box and whisker plot, the bottom of the box represents the 25th percentile (Q1) and top of the box represents the 75th percentile (Q3), the solid line within the box shows the median. The whiskers at the bottom and top indicate the minimum (Q1-1.5times interquartile range) and maximum (Q3+1.5times interquartile range) values respectively.

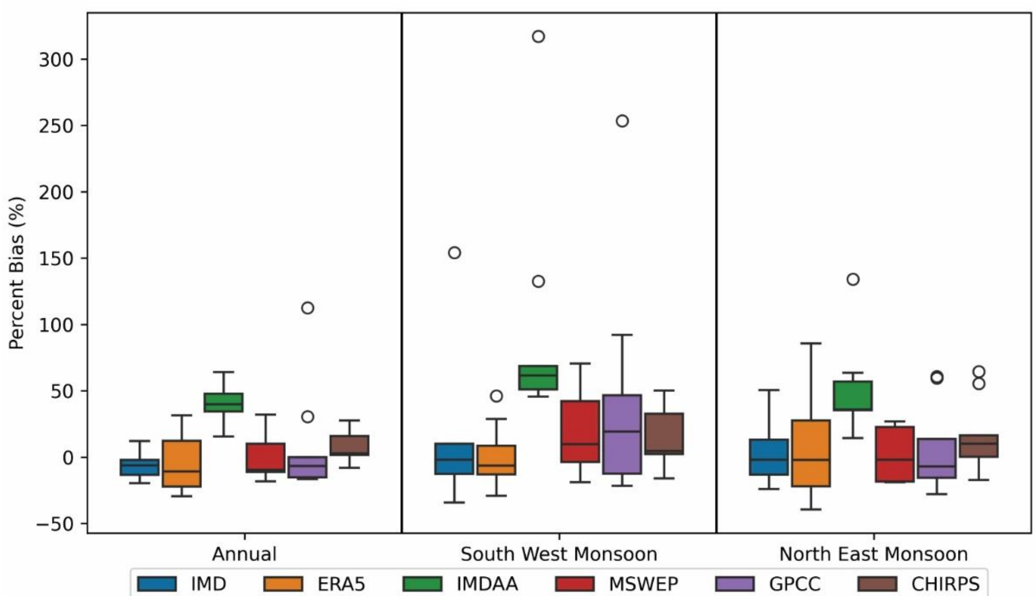

Fig. 13 Spatial (across nine-gauge stations) PBIAS distribution of the annual rainfall, Southwest monsoon (SWM) and Northeast monsoon (NEM) for satellite and reanalysis products in comparison with the gauge rainfall data. In a box and whisker plot, the bottom of the box represents the 25th percentile (Q1) and top of the box represents the 75th percentile (Q3), the solid line within the box shows the median. The whiskers at the bottom and top indicate the minimum (Q1-1.5times interquartile range) and maximum (Q3+1.5times interquartile range) values respectively. Black round points are outliers.

From the comparative analysis of precipitation products with ground station data, it was found that the MSWEP performed better than other precipitation products with a strong correlation, lower PBIAS and RMSE. The efficacy of MSWEP and IMD in representing the ground station observations was also tested using two more statistical indices (the index of agreement and the coefficient of determination). Index of agreement (d) values were higher with MSWEP than IMD for all ground stations at both annual and SWM scales. Similarly, during NEM, the d values for MSWEP were higher in the majority of stations except three stations where IMD had slightly higher d values (0.02, 0.04, 0.09) than MSWEP, which is also negligible (Table 3). The coefficient of determination (r^2) also showed a similar pattern of index of agreement (d) in the comparison of precipitation products (IMD and MSWEP) with ground stations.

Table 3. Comparison of IMD and MSWEP with ground station rainfall datasets using Index of agreement (d) and Coefficient of determination (r^2)

Stations	Annual		Southwest monsoon		Northeast monsoon		Annual		Southwest monsoon		Northeast monsoon	
	Index of agreement (d)						Coefficient of determination (r^2)					
	IMD	MSWEP	IMD	MSWEP	IMD	MSWEP	IMD	MSWEP	IMD	MSWEP	IMD	MSWEP
Aduthurai	0.71	0.87	0.54	0.82	0.82	0.86	0.34	0.61	0.16	0.48	0.55	0.58
Coimbatore	0.59	0.69	0.21	0.66	0.85	0.85	0.18	0.67	0.00	0.58	0.66	0.78
Cuddalore	0.68	0.82	0.57	0.86	0.68	0.83	0.35	0.54	0.16	0.61	0.29	0.61
Killikulam	0.50	0.58	0.20	0.52	0.63	0.66	0.07	0.24	0.09	0.10	0.37	0.28
Kovilpatti	0.37	0.67	0.14	0.55	0.65	0.63	0.02	0.42	0.07	0.13	0.22	0.22
Madurai	0.61	0.78	0.47	0.68	0.93	0.84	0.17	0.54	0.07	0.28	0.75	0.55
Ramnad	0.33	0.85	0.25	0.88	0.77	0.79	0.00	0.78	0.00	0.75	0.43	0.80
Tirur	0.46	0.47	0.78	0.81	0.89	0.91	0.07	0.27	0.48	0.57	0.64	0.82
Trichy	0.65	0.68	0.56	0.67	0.87	0.83	0.20	0.37	0.16	0.35	0.66	0.63

Conclusion

The study investigates the performance of high-resolution precipitation products (MSWEP, GPCC, CHIRPS, ERA5, IMDAA) to select the best precipitation products for Tamil Nadu by employing rainfall detection and estimation accuracy metrics at each grid. The results obtained from the spatial performance analysis of precipitation products against IMD indicated that ERA proved to be effective in detecting rainfall followed by MSWEP. Even though ERA is effective in detecting rainfall, False Alarm Ratio (FAR) is higher in ERA than MSWEP. The Rain day detection capability of most of the products is greater in hilly regions compared to other regions.

The accuracy metrics (continuous statistics) results exhibit a better performance of satellite precipitation products compared to reanalysis products. MSWEP demonstrates optimal performance in capturing rainfall events and achieving good rainfall estimates on both annual and seasonal scales. At the same time, CHIRPS is equally effective as MSWEP at estimating the rainfall during NEM. Among all precipitation products, the IMDAA shows poor performance in rainfall estimation.

IMD precipitation dataset is predominately used as a high-quality reference dataset to compare satellite and reanalysis precipitation datasets for India. Considering the wider usage of the IMD precipitation dataset, this study attempted to verify the performance of IMD dataset along with the other satellite and reanalysis precipitation datasets by comparing with the observed ground station rainfall. Comparison of precipitation products with ground station rainfall reveals a higher efficiency of MSWEP in estimating the rainfall over the other precipitation products by having a strong and significant correlation with ground station rainfall and a lower RMSE. PBIAS of MSWEP is slightly higher than that of IMD. In view of correlation (r), Index of agreement (d), Coefficient of determination (r2), RMSE values, MSWEP outperformed the IMD at both annual and seasonal scales. Hence, high resolution MSWEP data could be suitable for operational applications in various sectors.

All precipitation products, excluding IMDAA, present both an overestimation and an underestimation of rainfall across the ground stations. The results of grid analysis performed over Tamil Nadu show that the magnitude of overestimation and underestimation of rainfall by SPPs varied spatially and temporally. It clearly indicates that the accurate spatiotemporal estimation of rainfall by SPPs remains a challenge. This sort of inaccuracy and error may stem from bias correction of precipitation products with insufficient gauge distributions, limitations of remote sensing sensors used and uncertainty in SPPs estimation also attributed to the algorithm used by different SSPs to estimate rainfall.

Overall, the comparison of precipitation products against IMD gridded rainfall indicates the better performance of MSWEP rainfall dataset over other datasets. ERA5, CHIRPS and GPCC datasets can also be used for locations where gauge-based rainfall datasets are scarce and unavailable. The comparison study of IMD and MSWEP with ground station observations shows the suitability of high-resolution MSWEP rainfall datasets (0.1°×0.1°) alternative to IMD gridded (0.25°×0.25°) rainfall datasets.

The best precipitation product identified through the robust evaluation would increase the confidence of researchers and practitioners to apply the data to bolster operational meteorological research, crop planning, enhanced weather forecasting, framing suitable water management plans, crop simulation and hydrological modelling and to provide consistent and precise rainfall information needed for policy-making decisions. The spatiotemporal evaluation results of precipitation products form the basis for enhancing the accuracy of precipitation products at the regional level and increasing their practical applications.

Author's contribution

The study conception and design were performed by Elangovan Devadarshini and Kulanthaivelu Bhuvaneswari. Material preparation, data collection, and analysis were performed by Elangovan Devadarshini, Shanmugam Mohan Kumar, Manickam Dhasarathan and Kandasamy Senthilraja under supervision of Vellingiri Geethalakshmi. The first draft of the manuscript was written by Alagarsamy Senthil, Shahbaz Mushtaq, Thong Nguyen-Huy and Thanh Mai and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the Department of Science and Technology (DST), Government of India (GOI), through the project entitled “Enhancing climate change adaptation processes for farmers and agribusinesses (DST No: DST/INT/AUS/P-77/2020(G))”. The funders had no role in the work design, data collection and analysis, or decision and preparation of the manuscript.

Acknowledgments

The authors greatly acknowledged the Computational facility by the Tamil Nadu Agricultural University, Coimbatore and University of Southern Queensland, Australia for technical support.

Conflict of Interest Statement

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Data Availability Statement

Data will be made available on request.

Declarations

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Competing interests

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

References

Alfieri, L., Avanzi, F., Delogu, F., Gabellani, S., Bruno, G., Campo, L., Libertino, A., Massari, C., Tarpanelli, A., Rains, D. & Miralles, D.G. (2021). High resolution satellite products improve hydrological modelling in northern Italy. *Hydrology and Earth System Sciences Discussions*, *2021*,1-29.

Alijanian, M., Rakhshandehroo, G.R., Mishra, A., & Dehghani, M. (2019). Evaluation of remotely sensed precipitation estimates using PERSIANN-CDR and MSWEP for spatiotemporal drought assessment over Iran. *Journal of hydrology*, *579*, 124189.

Asfaw, A., Simane, B., Hassen, A., & Bantider, A. (2018). Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: A case study in woleka sub-basin. *Weather and Climate Extremes*, 19, 29–14. https://doi.org/10.1016/j.wace.2017.12.002

Ashrit, R., Indira Rani, S., Kumar, S., Karunasagar, S., Arulalan, T., Francis, T., … & Jermey, P.J.J. (2020). IMDAA regional reanalysis: Performance evaluation during Indian summer monsoon season. *Journal of Geophysical Research: Atmospheres, 125* (2):e2019JD030973.

Beck, H.E., Van Dijk, A.I., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., & Sciences, E.S. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology Earth System Sciences*, *21* (1):589-615.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I., … & Adler, R.F. (2019). MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society, 100* (3):473-500.

Behrangi, A., Khakbaz, B., Jaw, T.C., AghaKouchak, A., Hsu, K., & Sorooshian, S. (2011). Hydrologic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology*, *397*(3-4), 225-237.

Belay, A. S., Fenta, A. A., Yenehun, A., Nigate, F., Tilahun, S. A., Moges, M. M., ... & Walraevens, K. (2019). Evaluation and application of multi-source satellite rainfall product CHIRPS to assess spatio-temporal rainfall variability on data-sparse western margins of Ethiopian highlands. *Remote Sensing*, *11*(22), 2688.

Bharti, V., & Singh, C. (2015). Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region. *Journal of Geophysical Research: Atmospheres*, *120*(24), 12458-12473.

Bitew, M. M., & Gebremichael, M. (2011). Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research*, *47*(6).

Black, E., Tarnavsky, E., Maidment, R., Greatrex, H., Mookerjee, A., Quaife, T., & Brown, M. (2016). The use of remotely sensed rainfall for managing drought risk: A case study of weather index insurance in Zambia. *Remote Sensing*, *8*(4), 342.

Booth, J. F., Naud, C. M., & Willison, J. (2018). Evaluation of extratropical cyclone precipitation in the North Atlantic basin: An analysis of ERA-Interim, WRF, and two CMIP5 models. *Journal of Climate, 31(6),* 2345-2360.

Collier, C.G. (2007). Flash flood forecasting: What are the limits of predictability?. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *133*(622), 3-23.

Dinku, T. (2019). Challenges with availability and quality of climate data in Africa. In *Extreme hydrology and climate variability* (pp. 71-80). Elsevier.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., & Ceccato, P. (2018). Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society*, *144*:292-312.

Enenkel, M., Osgood, D., Anderson, M., Powell, B., McCarty, J., Neigh, C., ... & Brown, M. (2019). Exploiting the convergence of evidence in satellite data for advanced weather index insurance design. *Weather, Climate, and Society*, *11*(1), 65-93.

Fowler, H.J., Ali, H., Allan, R.P., Ban, N., Barbero, R., Berg, P., … & Dale, M. (2021). Towards advancing scientific knowledge of climate change impacts on short-duration rainfall extremes. *Philosophical Transactions of the Royal Society, 379* (2195):20190542.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., … & Hoell, A.J.S. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data, 2* (1):1-21.

Gebremichael, M., Bitew, M. M., Hirpa, F. A., & Tesfay, G. N. (2014). Accuracy of satellite rainfall estimates in the B lue N ile B asin: L owland plain versus highland mountain. *Water Resources Research*, *50*(11), 8775-8790.

Geethalakshmi, V., D.J.C.C. Ga, and C. Agriculture over India, Hyderabad. 2008. *Impact of Climate Change on Agriculture over Tamil Nadu. Chap IV, Climate Change Agriculture over India*. Hyderabad, India: CRIDA.

Gleixner, S., Demissie, T., & Diro. G.T. (2020). Did ERA5 improve temperature and precipitation reanalysis over East Africa. *Atmosphere, 11* (9):996.

Gosset, M., Julien, V., Quantin, G. & Matis, A. (2013) Evaluation of several rainfall products used for hydrological applications over west Africa using two high resolution gauge networks. *Quarterly Journal of the Royal Meteorological Society*, *139*, 923–940.

Gupta, A.K., & Nagar, D.K. (1999). *Matrix Variate Distributions* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780203749289

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Nicolas, J., … & Schepers. R.M.S. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146* (730):1999-2049.

Hobouchian, M.P., Salio, P., Skabar, Y.G., Vila, D., & Garreaud, R. (2017). Assessment of satellite precipitation estimates over the slopes of the subtropical Andes. *Atmospheric Research*, *190*, 43-54.

Hodson, D. L., Bretonnière, P. A., Cassou, C., Davini, P., Klingaman, N. P., Lohmann, K., ... & Senan, R. (2022). Coupled climate response to Atlantic Multidecadal Variability in a multi-model multi-resolution ensemble. *Climate Dynamics*, *59*(3), 805-836.

Huffman, G.J., Adler, R.F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., … & Schneider. U. (1997). The global precipitation climatology project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society, 78* (1):5-20.

Huffman, G.J., Adler, R.F., Morrissey, M.M., Bolvin, D.T., Joyce, R., McGavock, B. & J. Susskind. (2001). Global precipitation at one-degree daily resolution from Mult satellite observations. *Journal of Hydrometeorology,* 2 (1):36-50.

Jegankumar, R., Nagarathinam, S., & Kannadasan, K. (2012). Spatial distribution of rainfall in Salem and Namakkal districts. *International journal of geomatics and geosciences, 2* (4):986.

Kalyan, A.V.S., Ghose, D.K., Thalagapu, R., Guntu, R.K., Agarwal, A., Kurths, J., & Rathinasamy, M. (2021). Multiscale spatiotemporal analysis of extreme events in the Gomati River Basin, India. *Atmosphere*, *12*(4), 480.

Kolluru, V., Kolluru, S., & Konkathi, P. (2020). Evaluation and integration of reanalysis rainfall products under contrasting climatic conditions in India. *Atmospheric Research*, *246*, 105121.

Kummu, M., Guillaume, J.H., de Moel, H., Eisner, S., Flörke, M., Porkka, M., ... & Ward, P.J. (2016). The world's road to water scarcity: shortage and stress in the 20th century and pathways towards sustainability. *Scientific reports*, *6*(1), 1-16.

Levizzani, V., & Cattani, E. (2019). Satellite remote sensing of precipitation and the terrestrial water cycle in a changing climate. *Remote sensing*, *11*(19), 2301.

Luo, H., Ge, F., Yang, K., Zhu, S., Peng, T., Cai, W., … & Tang. W. (2019). Assessment of ECMWF reanalysis data in complex terrain: Can the CERA-20C and ERA-Interim datasets replicate the variation in surface air temperatures over Sichuan, China? *International Journal of Climatology, 39* (15):5619-5634.

Malvern, S.F., & Maurice, C. (2018). Comparison of satellite data and groundbased weather data in Masvingo, Zimbabwe. *International Journal of Environmental Sciences & Natural Resources*, *8*:102-107.

Merlin, O., Chehbouni, A., Walker, J.P., Panciera, R., & Kerr, Y.H. (2008). A simple method to disaggregate passive microwave-based soil moisture. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(3), 786-796.

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation: Measurement, remote sensing, climatology and modeling. *Atmospheric Research*, *94*(4), 512-533.

Mohan Kumar, S., Geethalakshmi, V., Ramanathan, S., Senthil, A., Senthilraja, K., Bhuvaneswari, K., ... & Priyanka, S. (2022). Rainfall spatial-temporal variability and trends in the Thamirabharani River Basin, India: Implications for agricultural planning and water management. *Sustainability*, *14*(22), 14948

Ngo-Duc, T., Matsumoto, J., Kamimera, H., & Bui, H.H. (2013). Monthly adjustment of Global Satellite Mapping of Precipitation (GSMaP) data over the VuGia–ThuBon River Basin in Central Vietnam using an artificial neural network. *Hydrological Research Letters*, *7*(4), 85-90.

Pai, D.S., Rajeevan, M., Sreejith, O.P., Mukhopadhyay, B., Satbha, N.S. & Mukhopadhyay B. (2014). Development of a new high spatial resolution (0.25× 0.25) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam*, *65*(1), 1-18.

Peng, F., Zhao, S., Chen, C., Cong, D., Wang, Y., & Ouyang, H. (2020). Evaluation and comparison of the precipitation detection ability of multiple satellite products in a typical agriculture area of China. *Atmospheric research*, *236*, 104814.

Phadtare, J. (2023). Influence of Underlying Topography on Post-Monsoon Cyclonic Systems over the Indian Peninsula. *Meteorology*, *2*(3), 329-343.

Prajesh, P.J., Kannan, B., Pazhanivelan, S., & Ragunath, K.P. (2019). Monitoring and mapping of seasonal vegetation trend in Tamil Nadu using NDVI and NDWI imagery. *Journal of Applied and Natural Science*, *11*(1), 54-61.

Prakash, S., Sathiyamoorthy, V., Mahesh, C., & Gairola, R. M. (2014). An evaluation of high-resolution Multisatellite rainfall products over the Indian monsoon region. *International Journal of Remote Sensing*, *35*(9), 3018-3035.

Ragettli, S., Cortés, G., McPhee, J., & Pellicciotti, F. (2014). An evaluation of approaches for modelling hydrological processes in high-elevation, glacierized Andean watersheds. *Hydrological Processes*, *28*(23), 5674-5695.

Rani, S.I., Arulalan, T., George, J.P., Rajagopal, E.N., Renshaw, R., Maycock, A., ... & Rajeevan, M. (2021). IMDAA: High-resolution satellite-era reanalysis for the Indian monsoon region. *Journal of Climate*, *34*(12), 5109-5133.

Ratner, B. (2009). The correlation coefficient: Its values range between+ 1/− 1, or do they?. *Journal of targeting, measurement and analysis for marketing*, *17*(2), 139-142.

Reddy, M.V., Mitra, A.K., Momin, I.M., & Krishna, U.M. (2022). How accurately satellite precipitation products capture the tropical cyclone rainfall?. *Journal of the Indian Society of Remote Sensing*, *50*(10), 1871-1884.

Reddy, N.M., & Saravanan, S. (2023). Evaluation of the accuracy of seven gridded satellite precipitation products over the Godavari River basin, India. *International Journal of Environmental Science and Technology*, *20*(9), 10179-10204.

Rockström, J., Falkenmark, M., Lannerstad, M., & Karlberg, L. (2012). The planetary water drama: Dual task of feeding humanity and curbing climate change. *Geophysical Research Letters*, *39*(15).

Schulzweida, U. (2019). CDO user guide (version 1.9.8). *Zenodo*, pp222.

Seshapriya, V. E., & Bhaskar, B. V. Decadal Variations of the Rainfall over Tamilnadu, India.

Shah, H.L., & Mishra, V. (2016). Uncertainty and bias in satellite-based precipitation estimates over Indian subcontinental basins: Implications for real-time streamflow simulation and flood prediction. *Journal of Hydrometeorology*, *17*(2), 615-636.

Sharannya, T.M., Al-Ansari, N., Deb Barma, S., & Mahesha, A. (2020). Evaluation of satellite precipitation products in simulating streamflow in a humid tropical catchment of India using a semi-distributed hydrological model. *Water*, *12*(9), 2400.

Silberstein, R.P. (2006). Hydrological models are so good, do we still need data?. *Environmental Modelling & Software*, *21*(9), 1340-1352.

Sunilkumar, K., Narayana Rao, T., Saikranthi, K., & Purnachandra Rao, M. (2015). Comprehensive evaluation of Multisatellite precipitation estimates over India using gridded rainfall data. *Journal of Geophysical Research: Atmospheres*, *120*(17), 8987-9005.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A., ... & De Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104*, 70-97.

Taye, M., Mengistu, D., & Sahlu, D. (2023). Performance evaluation of multiple satellite rainfall data sets in central highlands of Abbay Basin, Ethiopia. *European Journal of Remote Sensing*, *56*(1), 2233686.

Toté, C., Patricio, D., Boogaard, H., Van Der Wijngaart, R., Tarnavsky, E., & Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sensing*, *7*(2), 1758-1776.

Willmott, C.J., & Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International journal of geographical information science*, *20*(1), 89-102.

Yin, Z. Y., Zhang, X., Liu, X., Colella, M., & Chen, X. (2008). An assessment of the biases of satellite rainfall estimates over the Tibetan Plateau and correction methods based on topographic analysis. *Journal of Hydrometeorology*, *9*(3), 301-326.

Yuvaraj, R. M., & Rajeswari, M. (2016). Continentality and rainfall over Cauvery delta region of Tamil Nadu. *Imperial Journal of Interdisciplinary Research*, *2*, 727-734.

Zeng, H., Li, L., & Li, J. (2012). The evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) in drought monitoring in the Lancang River Basin. *Journal of Geographical Sciences*, *22*, 273-282.

Mulungu, D. M., & Mukama, E. (2023). Evaluation and modelling of accuracy of satellite-based CHIRPS rainfall data in Ruvu subbasin, Tanzania. *Modeling Earth Systems and Environment*, *9*(1), 1287-1300.