



Comparative evaluation of high-resolution rainfall products over South Peninsular India in characterising precipitation extremes

M. R. Sneha¹ · Archana Nair¹

Received: 14 October 2022 / Accepted: 21 March 2023 / Published online: 4 April 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Extreme rainfall events are becoming more frequent in South Peninsular India (SPI), which is resulting in an increase in flash floods, landslides, and damage to agriculture and infrastructure. However, because of the scarcity of rainfall data over remote areas and oceans, the reanalysis datasets are a boon for understanding various meteorological phenomenon. India's first high-resolution reanalysis dataset, Indian Monsoon Data Assimilation and Analysis (IMDAA), simulates past climate data at the regional or local levels. In this study, a comprehensive evaluation of IMDAA is carried out with respect to Indian Meteorological Department (IMD) daily gridded dataset over SPI during 2000–2020. It was found that monsoon and post-monsoon seasons demonstrated strong compatibility whereas annual and pre-monsoon seasons displayed some dissimilarity as depicted by Mahalanobis metric. Spatiotemporal analysis of IMDAA in capturing seasonal and annual climatic variations implied that the reanalysis product considerably showed a similar pattern to that of IMD neglecting some overestimations. According to the study, an analogy of 96% can be seen between IMDAA and IMD on an average scale. Results also suggest the efficiency of IMDAA model in capturing some extreme rainfall episodes better than the IMD. Consequently, the findings provide insight into the reanalysis product's ability to depict climatic variability and reliability in employing precipitation data estimated by IMDAA in modelling extreme events over SPI.

Keywords IMDAA · IMD · Rainfall extremes · South Peninsular India · Mahalanobis distance

✉ Archana Nair
archananair@kh.amrita.edu

M. R. Sneha
KH.SC.R4.MAT20001@asas.kh.amrita.edu

¹ Department of Mathematics, Amrita Vishwa Vidyapeetham, Ernakulam, Kerala 682024, India

1 Introduction

The Indian subcontinent is one of the most densely populated regions of the world. The region's socioeconomic development is significantly influenced by rainfall, and rainfed crops constitute a major part of the agriculture sector (Ashrit et al. 2020). Rainfall is an essential part of various hydrological and meteorological scenarios over the area. Spatiotemporal rainfall patterns play a crucial role in energy cycles that rely on land coupling and atmosphere (Chakraborty et al. 2015). Therefore, the availability of precise or reanalysed rainfall datasets is a crucial prerequisite for monitoring weather variabilities, modelling natural phenomena, forecasting and hydrometeorological studies.

Most accurate observations can be obtained from gauge measurements, but the scarcity of gauge stations remains a challenge. The sparsity of gauge stations over remote areas and oceans hinders in prediction and forecasting of climatic variations in those regions. Reanalysis datasets are characterised by the combined effect of empirical and satellite datasets to provide the most recent gridded atmospheric conditions for a certain time period (Blacutt et al. 2015). The precision of reanalysis datasets depends on the data assimilation scheme and the underlying model. Reanalysis products provide high-quality, long-term information about climate variability and atmospheric circulations. With the aid of a stable and individual data assimilation technique, the modelling system can combine limited and varying observations producing gridded coherent meteorological data (Rienecker et al. 2011). Over the past few decades, further developments of reanalysis datasets and their analysis have been extensively carried out (Kalnay et al. 1996; Saha et al. 2010; Dee et al. 2011; Mausson et al. 2014). The establishment of various global and regional reanalysis datasets paved the way for new high-resolution datasets that can provide insight into different atmospheric variabilities (Fortelius et al. 2002; Mesinger et al. 2006; Onogi et al. 2007; Kobayashi et al. 2015; Bollmeyer et al. 2015). Due to the requirement of enormous computational power, labour, and storage, global reanalysis products are produced with coarse grid resolutions. This moderate resolution of the global reanalysis product impedes its application in various mesoscale processes and prediction of climatic extremes which necessitates high spatial resolution. A regional reanalysis product employs a high-resolution model with restricted spatial coverage with boundary conditions and initial conditions derived from global reanalysis and can evaluate regional synoptic incidents more precisely (Dahlgren et al. 2016). National Centre for Medium Range Weather Forecasting (NCMRWF), Ministry of Earth Sciences, India, recently released two high-resolution reanalysis datasets, namely Indian Monsoon Data Assimilation and Analysis (IMDAA) and NCMRWF Global Forecast System model (NGFS). With the help of these reanalysis products, high-resolution datasets can be made available. Another positive aspect is a better understanding of extreme events and monsoon patterns at finer scales, particularly over India. IMDAA reanalysis product is developed for a limited area to produce a high-resolution (12 km) dataset rather than global reanalyses. IMDAA is better at quantifying rainfall in the past period and predicting future trends than NGFS (Mahmood et al. 2018).

Spatial aggregations of India show an adverse increase in extreme precipitation events, mainly during the monsoon season (Ghosh et al. 2016). About 80% of the overall rainfall received by India is during the monsoon season, i.e., June to September. Climatic changes including global warming, greenhouse gas emissions and land cover increase due to urbanisation inclining the chances of occurrence of extreme events as shown by the study in (Kharin et al. 2007). Every 1 °C of warming results in an increase in the water-holding capacity of the atmosphere, by 6–7% which leads to an increase in extreme rainfall

(Trenberth 2005). The urbanisation rate of India in 2020 is estimated as 34.92%, out of which the South Peninsular Region's contribution is very high. An increase in the number of extreme events during 1901–2010 in many urban cities indicated the influence of urbanisation on the occurrence of such events (Ali et al. 2014). Extreme rainfall events associated with the water vapour transport from the Arabian Sea and the Indian Ocean caused floods in Kerala in 2018 (Lyngwa and Nayak 2021). Similarly, many states of South Peninsular India are facing havoc due to extreme rainfall events. Explanations of spatiotemporal trends in extremes, mean precipitation and other hydrometeorological factors remain challenging because of the uncertainties in the multi-model observations and disagreement in data assimilation strategies in reanalysis products (Zhang et al. 2017). Henceforth, an investigation of the reanalysis products in estimating extreme precipitation scenarios and quantifying the rainfall amount received by the area is essential for societal growth and prediction of future trends in the spatial variability of precipitation and other factors. A limited number of studies have been conducted to analyse the performance of IMDAA over India; notably, no studies were conducted to study the rainfall occurrences and extreme events over South Peninsular India (SPI) using IMDAA. In (Aggarwal et al. 2022), an investigation of IMDAA's ability to characterise the Indian Summer Monsoon (ISM) was carried out over Northwest India for the years 1979–2018. The results showed the capability of IMDAA in realistically representing the ISM features and the linkage between moisture availability and convective precipitation formation. For the year 2018, a study was conducted to assess the ability of NCMRWF in predicting ISM for the months of June to September revealing the qualitative reliability of IMDAA in forecasting precipitation and zonal winds (Chakraborty et al. 2021). Precipitation forecast analysis was done with respect to Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG GPM), predictive skills were calculated based on Tropical Rainfall Measuring Mission (TRMM) observations and skills for zonal winds forecasts were computed using European Centre for Medium-Range Weather Forecasts Reanalysis (ERA-interim). A study over Western Himalaya region depicts that IMDAA is efficient in demonstrating heavy rainfalls as well as characteristics of winter precipitation at seasonal, diurnal and interannual scales associated with western disturbances. The study also implied that IMDAA is coherent in exhibiting spatial patterns comparing to gauge based and satellite products, even in higher magnitudes due to its high resolution (Sharma et al. 2022). Another intercomparison research on gridded rainfall datasets of nine global datasets over Sri Lanka implies that IMDAA well identified rainfall patterns and suitable for hydrological applications (Bandara et al. 2022). Significance of high-resolution datasets is emphasized in a study over Western Himalayas, in which, authors analysed winter seasonal precipitation over the area using different satellite (IMERG), observational (IMD) and reanalysis products (European Centre for Medium-Range Weather Forecasts (ERA5), IMDAA). Evaluation also revealed the efficiency of using IMDAA precipitation data in moisture transport, cloud cover, upper tropospheric circulations and surface temperature simulations (Punde et al. 2022).

This study focuses on the evaluation of the high-resolution reanalysis dataset, IMDAA during the years 2000–2020 in estimating precipitation characteristics and extreme events over SPI. The research study is categorised into the following details:

1. Spatiotemporal variability of daily average precipitation amounts on annual and seasonal scale is carried out using IMDAA with Indian Meteorological Department (IMD) daily gridded dataset as the benchmark.

2. Reliability of IMDAA is further emphasised by statistically analysing IMDAA in capturing precipitation using different agreement and categorical indices.
3. Spatial analysis over the grid points showing significant differences in percentile values of precipitation is done using a new approach known as Mahalanobis distance. Mahalanobis distance is preferred for multivariate comparison since a multivariate system is converted to a univariate system in the process and it exhibits sensitivity in intervariable changes in the system (Mahalanobis 1930). Also, a study conducted over San Cristobal, Venezuela concluded that Mahalanobis distance is better than Euclidean distance and Manhattan distance in detecting atypical observations in monthly precipitation series (Rivas et al. 2020).
4. Further, a statewise analysis of extreme rainfall episodes detection by IMDAA and extreme event prediction using extreme value distribution is carried out.
5. ENSO (El Niño–Southern Oscillation) climatic phenomenon is used to analyse its resonance with the extreme years using both IMDAA and IMD standardised precipitation data.

The paper is organised as follows: Sect. 2 describes the data and climatology of the site; Sect. 3 consists of the methodology of the research done; the results are discussed in Sect. 4; Summary and main conclusions of the study are presented in Sect. 5.

2 Data and climatology of the site

2.1 Data

Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis datasets of 12 km resolution are downloaded from the NCMRWF database from 2000 to 2020 (Rani et al. 2021). This high-resolution reanalysis dataset is comprehensively analysed using the Indian Meteorological Department (IMD) observations for the same period (Pai et al. 2014).

2.2 IMD precipitation data

Indian Meteorological Department (IMD) provides $0.25^\circ \times 0.25^\circ$ daily gridded precipitation datasets over India. This agency provides datasets based on ground-based gauge stations across the country. Around 6955 gauge stations are utilised to estimate the precipitation over the region. These 6955 gauge stations include observations from 547 IMD observatory stations, 494 hydro-meteorology observatories, 74 Agromet observatories, and the rest are rainfall stations maintained by the government (Rajeevan et al. 2008). Due to the topography and complex hilly terrains, there is a paucity of gauge stations in the country, mainly in the northernmost regions of the country. There is a non-uniform distribution of the gauge stations in the country, and the Inverse Distance Weighted interpolation (IDW) method was implemented to reduce this effect of non-uniformity and temporal unavailability of gauge stations (Rajeevan et al. 2006). The precipitation data from 2000 to 2020 downloaded from its official website are used as a benchmark in assessing the reanalysis dataset.

2.3 IMDAA precipitation data

IMDAA is a high-resolution reanalysis dataset implemented in the combined endeavour of the National Centre for Medium Range Weather Forecasting (NCMRWF), India and the Met Office (MO), UK, in association with the IMD as part of the National Monsoon Mission (NMM) project of the Ministry of Earth Sciences, Government of India (Shepard 1968). It is the foremost reanalysis dataset produced by NCMRWF with a horizontal resolution of 12 km spanning south Asia with neighbouring regions to capture all the necessary variables affecting the climatic variations or occurrence of the monsoon season. The database consists of hourly, three hourly and six hourly datasets of different observed variables, such as precipitation, moisture, temperature, and humidity, and non-observed variables, such as cloud cover and fluxes, for the period of 1979–2022 at present. This mesoscale version is built on a four-dimensional assimilation method and with 63 vertical levels leading to a height of 40 km. The precipitation data are downloaded from RDS NCMRWF repository.

2.4 Climatology of SPI

Assessment of reanalysis product over a region necessitates the knowledge of climatology of the study region. The performance of each precipitation product differs from area to area due to topological and climatic characteristics. This study focuses on the precipitation characteristics over the South Peninsular region of India as shown in Fig. 1 using IMDAA and IMD.

SPI is a triangle-shaped landmass surrounded by the Bay of Bengal on the east, the Arabian Sea on the west and the Indian Ocean on the south. SPI consists of five states: Kerala, Karnataka, Tamil Nadu, Andhra Pradesh, and Telangana.

Along with the sea, ocean and other water resources, the Western Ghats and the Eastern Ghats present in the borders play a significant role in the wind circulations. The circulations transport moisture content across the country, leading to monsoon and several precipitations phenomena. These moisture contents from the Arabian sea move towards the Narmada and Tapi rivers causing precipitation in areas of central India, and the combined flow of moisture content from the Indian Ocean and the Bay of Bengal leads to widespread rain in the northeast parts of India (Singh et al. 2021). Hence, an analysis related to SPI is critical in the context of climatology because of the area's topography.

3 Methodology

Precipitation datasets for 21 years from 2000 to 2020 are utilised in this study. Total hourly precipitation estimated by IMDAA during the analysis time is extracted from NCMRWF official web page (Tan and Santo 2018). Daily precipitation data over India for the study period are downloaded from IMD website. The resolutions of the two datasets used here are different. So, the first step was to regrid the reanalysis dataset to a horizontal resolution of 0.25° . After interpolating the datasets to the same resolution, precipitation over SPI is obtained. A comparison of the estimation capacity of the reanalysis dataset with the observation dataset is done by examining seasonal and interannual variability. Based on the climatology of SPI, the data are classified into four seasons: (1) Winter (January–February),

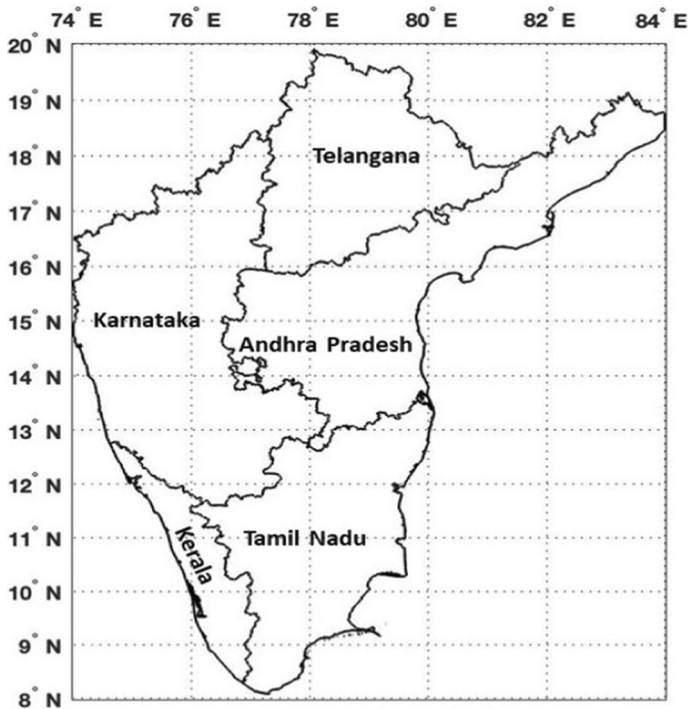


Fig. 1 Map of South Peninsular India

(2) Pre-monsoon (March–May), (3) Monsoon (June–September), and (4) Post-monsoon (October–December). For the comprehensive assessment of the reanalysis product and the observation datasets, different statistical indices are made use of.

3.1 Assessment indices

Statistical indices are calculated over 810 grid points over the land within the geographical boundaries of the states representing SPI spanning 7° N to 22° N to 73.5° E to 84.25° E for the observation period. The performance of IMDAA is analysed at annual, monthly, and seasonal scales. Here, O_i and S_i is the average precipitation obtained by IMD and IMDAA, respectively, over the area, \bar{O} and \bar{S} is the mean of O_i and S_i , respectively, n is the number of samples corresponding to space and time rd and cd are the standard deviations of IMD and IMDAA values, respectively. A detailed description of the statistical indices used is given in Table 1.

Rather than the above-mentioned metrics, Kling Gupta Efficiency (KGE) takes into account the pattern similarity, variance and error, while Nash Sutcliffe Efficiency Coefficient (NSE) gives the index value totally based on residuals. Hence, these two indexes altogether give different scores and are important to make any decision regarding the best model selection. The details of these two metrics are given as follows:

Kling Gupta Efficiency (KGE): A measure to show the degree of fit and helps to understand the relationship between variability, correlation, standard deviation, bias, etc.

Table 1 Description of statistical indices

Statistical Indices	Equations	Range of indices	References
Correlation Coefficient(CC)	$CC = \frac{\sum_{i=1}^n (o_i - \bar{o})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (o_i - \bar{o})^2} \sqrt{\sum_{i=1}^n (s_i - \bar{s})^2}}$	- 1 to 1	Chen et al. (2020)
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^n (s_i - o_i)^2}{n}}$	0 to ∞	Willmott and Matsuura (2005)
Relative Bias (RB)	$RB = \frac{\sum_{i=1}^n (s_i - o_i)}{\sum_{i=1}^n o_i}$	$-\infty$ to ∞	Moazami et al. (2016)
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^n s_i - o_i }{n}$	0 to ∞	Tait et al. (2012)
Index of Agreement (d)	$d = 1 - \frac{\sum_{i=1}^n (o_i - s_i)^2}{\sum_{i=1}^n (s_i - \bar{o} + o_i - \bar{o})^2}$	0 to 1	Thameemul Hajaj et al. (2019)

$$KGE = 1 - \sqrt{(CC - 1)^2 + \left(\frac{cd}{rd} - 1\right)^2 + \left(\frac{\bar{S}}{\bar{O}} - 1\right)^2} \tag{1}$$

Nash Sutcliffe Efficiency Coefficient (NSE): A statistic that establishes the magnitude of variability between the variances in observed and reanalysis data. Its value shows how strongly a 1:1 line fits in the observed versus reanalysis graph. The optimal value of NSE is 1 which implies an ideal model.

$$NSE = 1 - \frac{\sum_{i=1}^n (o_i - s_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \tag{2}$$

Moreover, seasonal and interannual variability of data is compared using spatial distributions and the significant difference between average precipitation during different seasons and years is analysed using the Student’s t test with the null hypothesis that there is no difference in the mean precipitation of IMD and IMDAA estimation.

Student’s t test is a statistical method of testing how significant the differences can be between two groups which are drawn from a normally distributed population where the standard deviation is unknown.

$$\text{Test statistic, } t = \frac{|O_i - S_i|}{D \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3}$$

where n_1 and n_2 are sample sizes of IMD and IMDAA, but here n_1 and n_2 are same, D is the pooled standard deviation of IMD and IMDAA.

3.2 Mahalanobis distance

Other than the usual distance metric calculated for single mean variables, a new approach is utilised which makes use of a multivariate distance metric known as Mahalanobis distance to find out if there exists any significant distance between IMD observations and reanalysis

product datasets across seasonal and annual percentiles of precipitation of each year corresponding to each grid point.

Mahalanobis distance (MD) is an approach which provides a covariance distance between data that helps to detect anomalies and similarities of unidentified samples effectively (Mimmack et al. 2001). It gives distance between centroids of two or more groups to get an idea about the divergence of characteristics associated with the groups (Matcharashvili et al. 2017). Rather than Euclidean distance, it addresses different groups based on the correlation of data points and all variables are equally characterised by giving less weightage to highly correlated variables (Kishore et al. 2016). Other important properties of Mahalanobis distances include their relationship to multidimensional scaling and the log-likelihood of multivariate normal distributions. Mahalanobis distance is used to produce univariate distance for multivariate groups with various parameters. Mahalanobis distance is given by the formula:

$$MD^2 = (x - y)^T P^{-1} (\bar{x} - \bar{y}) \quad (4)$$

where \bar{x} and \bar{y} represent sample means with sizes m and n , respectively, T is the transpose operation and P is the covariance matrix:

$$P = \frac{((m - 1)P_1 + (n - 1)P_2)}{m + n - 2} \quad (5)$$

where P_i 's denote the covariance matrices of the corresponding groups.

As the MD value gets smaller, the variables tend to fall into the same group or it implies that the groups possess similar characteristics. After the calculation of MD values, Hotelling's T^2 test is used to calculate the significant difference in the precipitation values. The T statistic is given by:

$$F = \frac{m + n - p - 1}{(m + n - 2)p} MD^2 \sim F_{p, m+n-p-1} \quad (6)$$

where p is the number of variables used. The significance test is undergone with the assumption that there is no discrepancy in the performance parameters of the groups as the null hypothesis. If $F > F_{p, m+n-p-1}$, then the null hypothesis is rejected implying the existence of a significant difference between the group parameters.

3.3 Performance diagram

Performance diagram is used to analyse the detection capability of IMDAA in seasonal, annual and percentile scales. The performance diagram plots four forecast statistics such as Probability of Detection (POD), Success Ratio (SR), Critical Success Index (CSI) and bias simultaneously. Points close to the upper right corner of the graph show the best efficiency in forecasting. POD is the probability with which a product will detect the rainfall correctly (Haile et al. 2016). CSI gives the capability of the product in estimating precipitation and SR is the inverse of the False Alarm Ratio (FAR) which gives the ratio of wrongly detected occurrences.

$$POD = \frac{H}{H + M} \quad (7)$$

$$FAR = \frac{F}{F + M} \tag{8}$$

$$CSI = \frac{H}{H + M + F} = \frac{1}{\frac{1}{(1-FAR)} + \frac{1}{POD} - 1} \tag{9}$$

where H is the number of hits, M is the number of misses and F is the number of false alarms (Tang et al. 2020).

3.4 Extreme value distribution

The occurrence of extreme precipitation events over SPI has increased during the past few decades. The geographical conditions of SPI with the ocean surrounding the land-mass on three sides lead to the development of thermal and pressure gradients increasing the chances of formation of hydrometeorological extremes more often (Baki et al. 2021). Extreme flood events are causative of these extreme precipitations. During the study period of 2000–2020, the states of SPI were prone to severe flooding events that caused catastrophic destructions of agricultural, economic, and societal aspects. The prediction of these disastrous extreme events attributed to variations in several dynamical parameters is a need. Spatial distributions of extreme rainfall episodes of each state detected by IMDAA are analysed and compared against the IMD observation dataset. Not only the detection or estimation of precipitation is needed, but a reanalysis dataset must also be capable of detailing small variations in precipitation amounts in calibrating future extremes. An extreme value distribution was used to give an insight into the prediction capability of IMDAA. Extreme value Type 1 distribution was used for finding the large extremes of the data (Phien 1987).

General representation of the probability density function of the Gumbel (maximum) distribution is given by:

$$f(x) = \frac{1}{\beta} e^{-\frac{x-\mu}{\beta}} e^{e^{-\frac{x-\mu}{\beta}}} \tag{10}$$

where x is the precipitation data, μ is the location parameter and β is the scale parameter. For the standard Gumbel distribution $\mu = 0$ and $\beta = 1$, lead to the following formula:

$$f(x) = e^{-x} e^{-e^{-x}} \tag{11}$$

The spatial distribution of grid points showing extreme precipitation amounts during the observed extreme rainfall episodes was compared based on the IMD observations.

Gumbel distribution is widely used to predict extreme values of different phenomena such as wind speeds, flood frequency, and extreme event prediction. Extreme value prediction can be done using different distribution such as Normal, Log-Normal, Weibull, Gamma, and Gumbel. The best fit distribution differs from data to data. Hence, goodness of fit measures such as the Kolmogorov–Smirnov (K–S) Test, Anderson–Darling (A–D) Test and Root Mean Square Error (RMSE) were used to determine the best fit distribution. Gumbel distribution resulted in the best fit for the data and in predicting the extremes. Also Gumbel distribution seems to be better in predicting extremes using IMD rainfall data as seen in studies related to rainfall intensity duration frequency curve over Bangladesh (Rasel and Islam 2015), intensity duration frequency duration with maximum rainfall (Barrie and

Scott 2021) and the distribution shows reliable results in flood forecasting, extreme events prediction, estimate river discharge or flow (Solomon and Prince 2013; Meeyaem and Polpinit 2014; Mamman et al. 2017; Bhagat 2017; Naz et al. 2019).

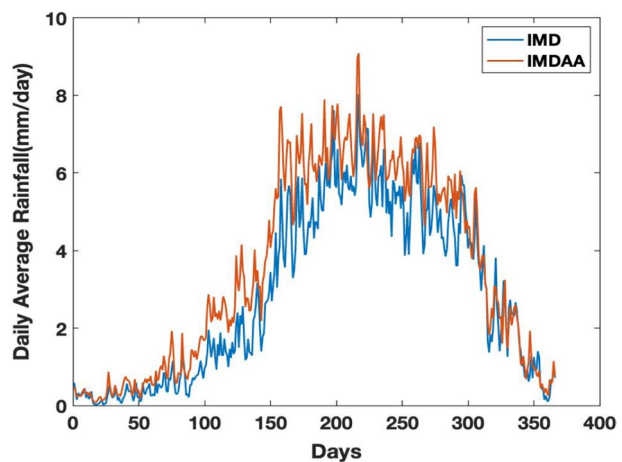
4 Results and discussion

As the primary analysis, the annual cycle of IMDAA and IMD is shown in Fig. 2. It can be noticed that daily average rainfall given by IMDAA is in close correlation with IMD rainfall data. However, a little exaggeration of the data is seen for IMDAA. A broad overview of the similar behaviour of average daily data can be seen from the figure. It is clear that IMDAA and IMD data share same patterns of crest and trough throughout the daily average of the 21 years of study. However, to guarantee the precision of IMDAA data in all respects, particularly in grid-based and extreme event prediction, an analysis in higher level skills is important. The following subsections are devoted to spatiotemporal analysis, variability assessment using different metrics and research on reliability of the data in predicting extremes.

4.1 Interannual variability of precipitation between IMDAA and IMD

The behavioural patterns of the reanalysis product throughout the years in comparison to the IMD observations over SPI are studied. Here, the findings reveal that during monsoon season, the estimated rainfall data are more spread out which can be observed from the standard deviation (SD) values of IMD (SD=4.49 mm/day) and IMDAA (SD=4.72 mm/day), whereas the data are clustered around the average values during pre-monsoon (SD of IMD=0.8 mm/day, SD of IMDAA=1.28 mm/day) and winter seasons (SD of IMD 0.2 mm/day, SD of IMDAA=0.31 mm/day). Figure 3 shows interannual variability of average precipitation estimated by IMDAA and IMD over the years 2000–2020 during different seasons and total yearly precipitation. Annual precipitation of IMDAA is comparable with IMD but with an overestimation by IMDAA. Figure 3a, d shows a common pattern indicating average rainfall estimation of IMDAA and IMD is closely related during monsoon season and overall annual precipitation values. As depicted in Fig. 3b, e, Winter

Fig. 2 Daily average precipitation estimated by IMDAA and IMD



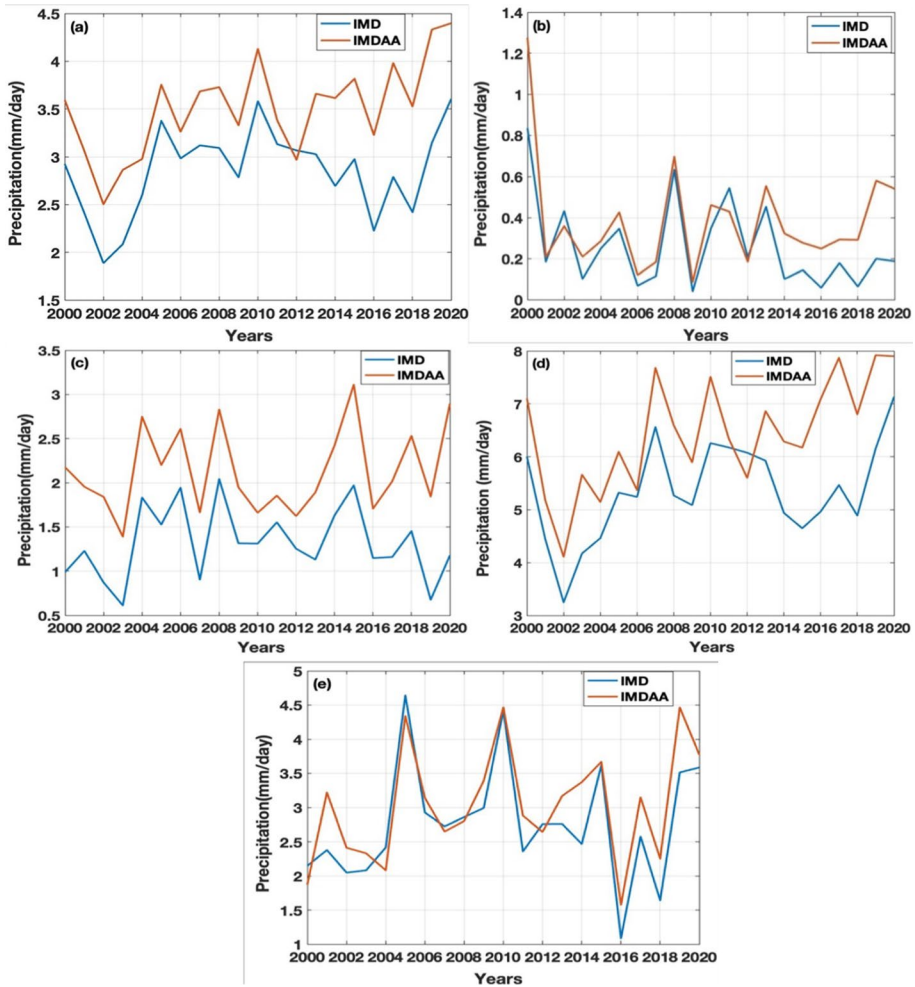


Fig. 3 Year to year variation of average daily rainfall on annual and seasonal scale, **a** Annual, **b** Winter, **c** Pre-monsoon, **d** Monsoon and **e** Post-monsoon

and post-monsoon season show similar characteristics implying that IMDAA has well-captured the rainfall during these seasons on overall seasonal scale. Out of these seasonal and annual estimations, Fig. 3c exhibits the highest deviation of precipitation amount obtained by IMDAA when compared to IMD. That is, during the pre-monsoon period, a high level of overestimation by IMDAA occurred during these 21 years.

4.2 Daily climatology comparison using scatter plots

Scatter plots define the correlation between two quantities with respect to the direction in which the points are scattered and the closeness of the points. If the data points are closely correlated, the scatter plot will show a straight line fit, and the direction of correlation depends on whether the line possesses a positive or negative slope.

Here, Fig. 4 presents scatter plots of average daily precipitation amount over SPI for 21 years during different seasons and the whole year. For a comprehensive analysis of the correlation, daily rainfall data of all grid points of SPI from both IMD and IMDAA for annual and respective seasons are pooled. All the seasonal and annual scatter plots show a positive correlation between IMDAA and IMD. In the scatter plots of annual (CC=0.8018) and monsoon (CC=0.8196) seasons, IMDAA displays a high correlation

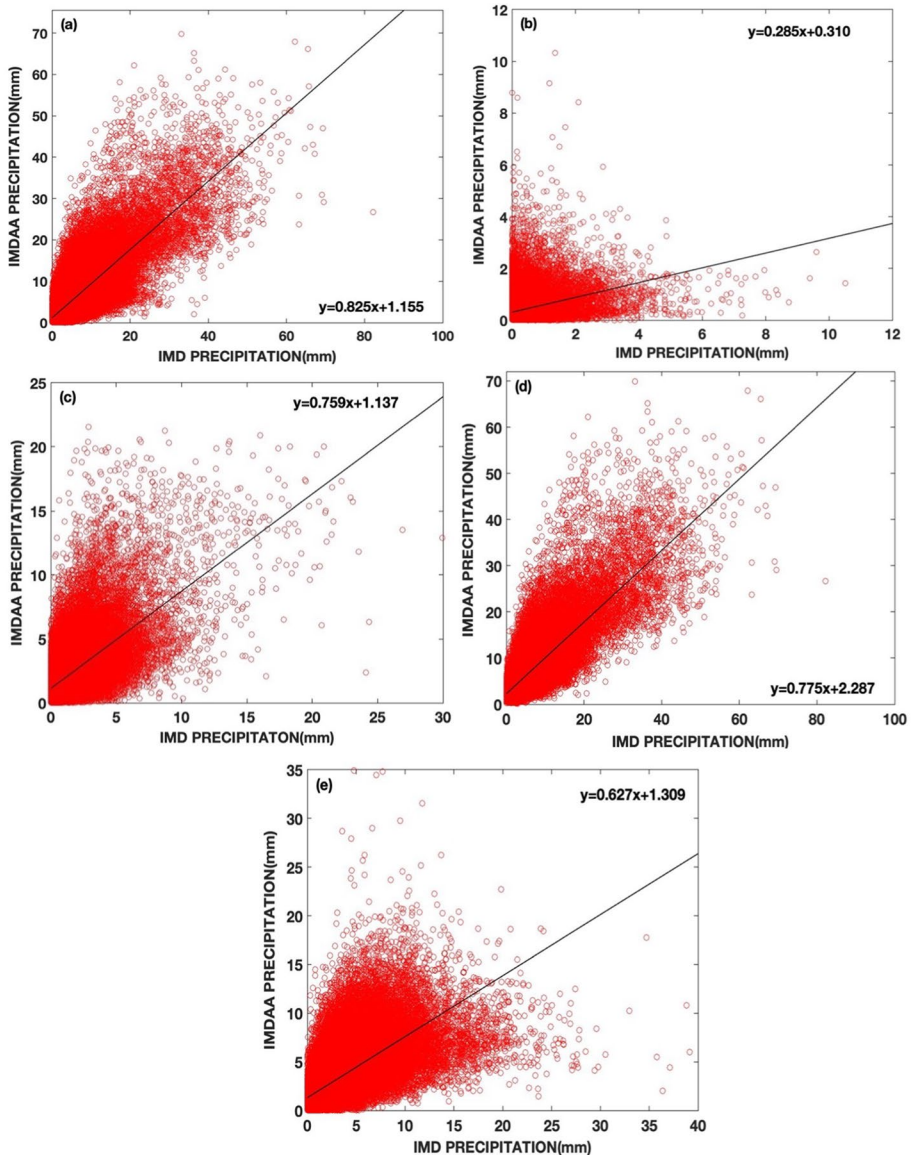


Fig. 4 Scatter plot of average daily precipitation estimated by IMD versus IMDAA during **a** annual, **b** winter, **c** pre-monsoon, **d** monsoon and **e** post-monsoon

with the IMD observations. The winter and post-monsoon seasonal rainfall are dominant over some parts of SPI with significant interannual variability. The dynamical variation in the rainfall amounts during these seasons is causing anomalies in the correlation values. As inferred from Fig. 4b, e, the correlation values are comparatively less in the case of winter ($CC=0.3294$) and post-monsoon ($CC=0.69315$) seasons. A correlation coefficient of 0.7315 indicates that pre-monsoon seasonal IMDAA data shows a moderate level of correlation with the IMD daily grid wise precipitation data. An additional information about the correlational relationship can be derived from the best fit trend line and the corresponding equation of the line which is included in the figure.

4.3 Spatial variability of IMDAA and IMD on annual and seasonal range

Spatiotemporal patterns of precipitation estimation resulting from a reanalysis product are an important analysis technique in establishing many hydrometeorological parameters.

Figure 5 depicts spatial distributions of grid points showing significant difference in annual and seasonal precipitation between IMDAA and IMD across the landmass during 2000–2020. Different colours in the distribution represent grid points which are manifesting notable positive or negative changes in mean precipitation amount when compared to the observation dataset.

Student's t test was used to analyse the extent to which there is a significant difference between IMDAA precipitation values and the observation dataset. For the 810 data points, a t test was done between 21 years' annual and seasonal precipitation values. The test resulted in zero for the grid points showing the same rainfall value as that of IMD during the years 2000–2020 period and the statistic value is one for grid points exhibiting a difference in precipitation.

In Fig. 5a, the spatial distribution shows that over the Telangana and most of the areas of Andhra Pradesh, there is a significant positive difference of 0.5–1 mm/day. It is evident that there is an overestimation by IMDAA over some grids of Karnataka over which an annual rainfall of 3 mm/day is received. Along the coastal sides of the peninsular region, IMDAA shows good agreement with the IMD observations and is having average precipitation of 10–12 mm/day. Additionally, along the coastal sides of SPI, an underestimation of IMDAA annual rainfall can be seen. Furthermore, IMDAA overestimated the amount of rainfall in most areas of SPI during the winter by 0–1 mm/day.

Pre-monsoon rainfall predicted by IMDAA is analogous to IMD as the number grids indicating significant difference is very less and an overestimation is only seen in relatively few number of grids along the coastal regions Andhra Pradesh, Karnataka and Kerala as shown in Fig. 5c. Precipitation during monsoon season estimated by IMDAA demonstrate a great homogeneity over all the parts of the landmass as given in Fig. 5d. The area received mean rainfall of 5.35 mm/day according to IMD and 6.43 mm/day based on IMDAA during the monsoon period from 2000 to 2020. A mixture of over and under estimations of IMDAA rainfall can be seen with a concentration of significantly different grids over Telangana. For the post-monsoon season, IMD and IMDAA estimated 2–3 mm/day along the northwest sides of Karnataka and over Telangana, 7–8 mm/day over Kerala and some regions of Fig. 5e indicate that IMDAA rainfall during post-monsoon season shows significant difference in the northern states of the region, especially overestimation in the range of 0–0.5 mm/day over Telangana and Karnataka. Underestimation of the rainfall amount predicted by IMDAA can be visualised through the mean difference value which is in the range of 0–1.5 mm/day.

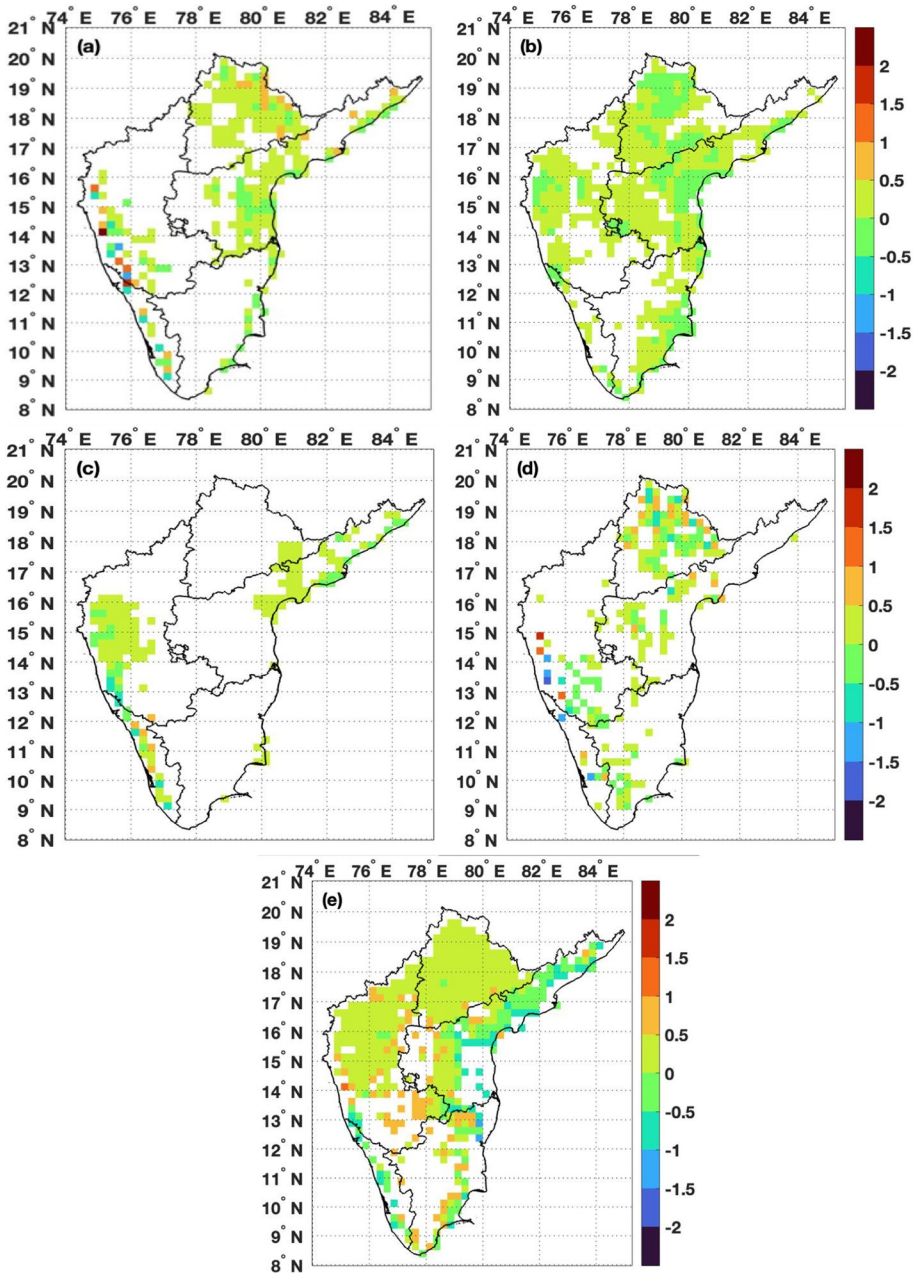


Fig. 5 Spatial distribution of significantly different grid points of seasonal and annual daily average precipitation, **a** annual, **b** winter, **c** pre-monsoon, **d** monsoon, **e** post-monsoon

Table 2 Statistical indices results for monthly analysis

Months	INDICES						
	CC	RMSE (mm/month)	RB (mm/month)	MAE (mm/month)	<i>d</i>	KGE	NSE
January	0.81	3.83	27.39	3.14	0.99	0.65	0.41
February	0.91	8.73	59.12	5.75	0.99	0.35	0.43
March	0.96	14.38	67.06	12.40	0.98	0.32	0.67
April	0.89	27.93	70.17	26.17	0.98	0.28	−0.78
May	0.83	40.95	56.33	37.11	0.98	0.39	−1.28
June	0.75	59.17	36.76	54.23	0.97	0.46	−2.21
July	0.80	43.74	17.59	36.74	0.98	0.71	0.09
August	0.93	30.67	12.69	24.90	0.99	0.83	0.60
September	0.92	33.89	18.30	28.55	0.99	0.79	0.44
October	0.91	28.26	14.19	24.16	0.99	0.83	0.62
November	0.93	19.66	5.54	16.49	0.99	0.75	0.85
December	0.83	10.65	7.78	7.59	0.99	0.80	0.61

Table 3 Statistical indices results for seasonal and annual analysis

Indices	Seasonal				Annual
	Winter	Pre-monsoon	Monsoon	Post-monsoon	
CC	0.84	0.75	0.76	0.88	0.77
RMSE (mm/season)	10.92	80.91	155.75	44.15	268.34
RB (mm/season)	45.2	62.19	20.37	10.69	23.27
MAE (mm/season)	8.34	75.44	138.34	36.22	245.67
<i>d</i>	0.98	0.97	0.97	0.99	0.97
KGE	0.48	0.31	0.66	0.84	0.66
NSE	0.23	−3.88	−1.06	0.65	−1.72

Quantitatively, the estimations of IMDAA are comparable with IMD observations excluding overestimations in the course of some seasons over particular areas of the land-mass. Hence, during 2000–2020, IMDAA well-captured the precipitation for all the seasons and overall times of the years.

The performance of IMDAA in correctly extracting precipitation amounts for each grid point in different years shows differences in several grid points during winter and post-monsoon seasons. But for the monsoon and pre-monsoon seasons, the spatial distribution follows a pattern in which there is a larger number of white grids than shaded grids indicating the efficiency of IMDAA.

4.4 Comparison using statistical metrics

Apart from the visual representation of the homogeneity of the two products, a statistical evaluation gives a better interpretation of the performance of the reanalysis product.

Tables 2 and 3 illustrate the values of various statistical measures to predict the reliability of the reanalysis product.

Monthly, seasonal, and annual variations of statistical indices are observed for IMDAA with an observation dataset from IMD across the SPI region during the span of 2000–2020. CC values are high for all the months indicating a good linear relationship between reanalysis and observation. RMSE and MAE show a similar trend moving from January to December. RMSE and MAE values are lowest for the month of January (~3–4 mm/month), from February to June there is a linear increase in RMSE and MAE values i.e., ~5.5–60 mm/month and a sudden trend of decreasing RMSE and MAE values can be seen from July to December with approximately in the range of 35–7 mm/month. Relative Bias is also showing a similar pattern to that of RMSE and MAE values. Index of Agreement values is close to one, indicating a good agreement over all twelve months even though there are some differences in the performance of IMDAA in each month as indicated by other statistics. This may be because of the sensitivity of the index due to the presence of extreme values leading to varying squared differences.

Table 3 gives statistical values for seasonal and annual scales. A similar pattern is shown by the seasonal values also. That is, the NSE, KGE and CC values are the smallest and have a high relative bias for the pre-monsoon (March–May) season. Monsoon and annual periods have a negative NSE but a lower relative bias indicating the efficiency of IMDAA. The RMSE and MAE values are very high in the case of monsoon and annual since these are calculated on a seasonal (mm/season) and annual (mm/year) basis in which the total precipitation amount is very high. Overall results show that IMDAA estimations are in close relation to IMD based on all the statistical indices even though some months are showing some discrepancies.

4.5 Geometrical evaluation using performance diagrams

Evaluation of performance of IMDAA is done using a geometrical relationship representation approach known as performance diagram. Performance diagram can depict the relationship between four important forecast metrics such as Probability of Detection (POD), Bias, Success Ratio (SR) and Critical Success Index (CSI). Percentiles of data are normalised to find out the hits and misses by fixing a threshold of zero. Hit occurs when both observation and reanalysis data behave analogously, that is, when observation is greater than zero, reanalysis is also greater than zero or vice versa. Misses are defined as a condition where observation and predicted data behaves oppositely.

Figure 6a indicates performance diagram of IMDAA during seasonal and annual scales where the data are taken yearly over areal average of 810 grid points and temporal average corresponding to seasonal or annual basis. In all the seasonal and annual data, IMDAA is showing good performance. IMDAA performs best during pre-monsoon season where the POD and SR values are close to one, winter and post-monsoon values are coinciding, annual and monsoon season plots are varying similarly. Figure 6b shows performance of IMDAA in the case of daily data for annual and seasonal scale. Prediction during annual (POD=0.98 and SR=0.98) and post-monsoon (POD=0.98 and SR=0.99) season is relatively satisfactory with respect to other seasons in daily case. POD and CSI of winter (POD=0.87) and monsoon (POD=0.84) are less than other points. Spatial scale comparison of annual and seasonal data is shown in Fig. 6c. IMDAA works better in monsoon season, which is having high POD (0.95), CSI, and SR (0.97) values. But annual, pre-monsoon and post-monsoon seasons show similar properties. Out of the five forecast values,

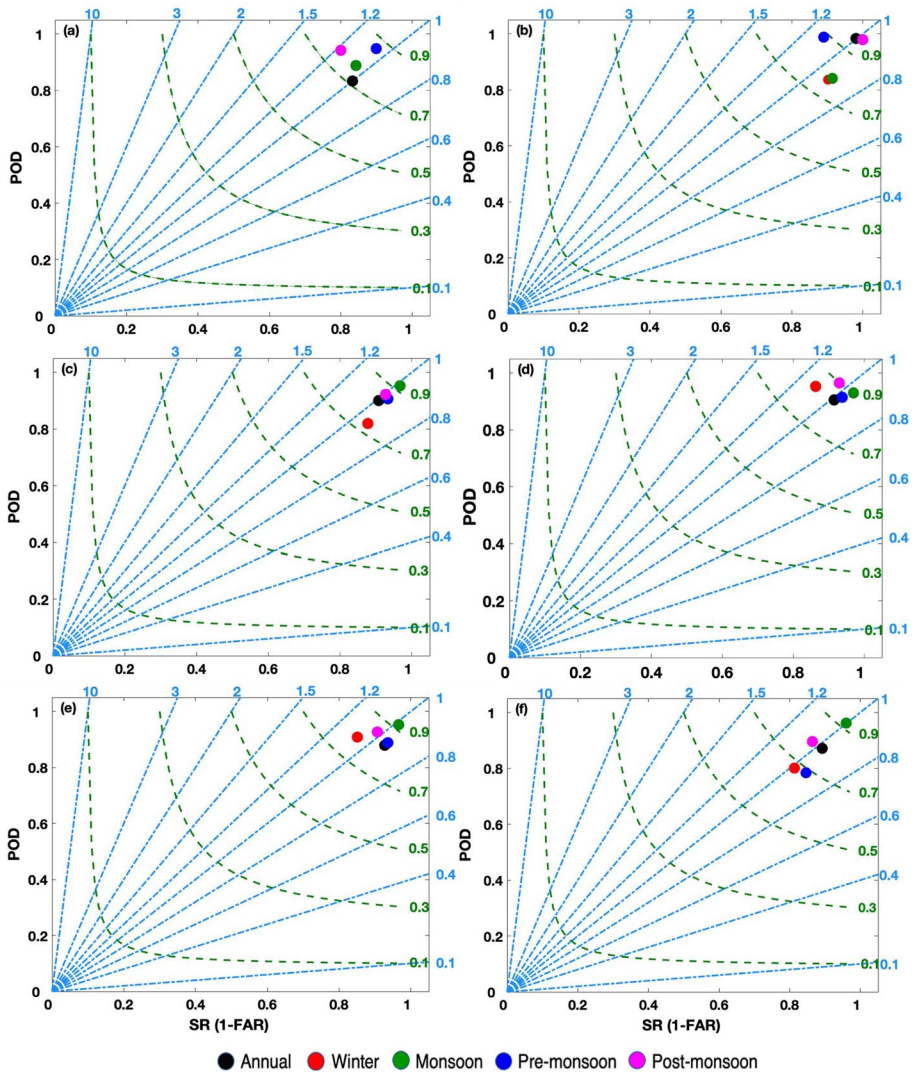


Fig. 6 Performance diagrams of IMDAA with respect to **a** annual variation, **b** daily variation, **c** spatial variation, **d** 90th percentile variation, **e** 95th percentile variation, **f** 99th percentile variation

IMDAA during winter season (POD=0.82 and SR=0.88) shows relatively worst performance. Moreover, the mean values over yearly, spatial, and daily scale data, an analysis was carried out to evaluate the reliability using percentile values. Performance diagram of 90th percentile (Fig. 6d), 95th percentile (Fig. 6e) and 99th percentile (Fig. 6f) resulted in the excellence of detection of precipitation by IMDAA during monsoon season with large POD and SR values. 90th and 95th percentile performance plots exhibit similar characteristics with annual(POD for 90th=0.90, SR for 90th=0.91 and POD for 95th=0.88, SR for 95th=0.93), pre-monsoon (POD for 90th=0.91,SR for 90th=0.94 and POD for 95th=0.89, SR for 95th=0.93), monsoon (POD for 90th=0.93, SR for 90th 0.97 and

POD for 95th=0.95, SR for 95th=0.97) and post-monsoon(POD for 90th=0.96, SR for 90th=0.93, POD for 95th=0.92, SR for 95th=0.91) seasons. Working of IMDAA during the winter season is relatively bad when compared to other seasons as seen in all the three percentile diagrams. Analysing all the performance diagrams, it can be concluded that prediction by IMDAA is close enough in all aspects, whether it is annual, daily, spatial or percentiles indicating the robustness of IMDAA in capturing precipitation occurrences.

4.6 Mahalanobis distance metric for percentiles

The intensity of rainfall events can adversely affect agriculture, biodiversity, lives of people, constructions, etc. Extreme rainfall can cause flash floods leading to catastrophes. The frequency of such extreme events increases the chance of distorting the balance of the climatic conditions. Hence a study verifying the changes in percentiles of precipitation over the years 2000–2020 of each grid point rather than mean rainfall is also important. Rather than the usual Euclidean metric, a new technique of finding multivariate distance is used to evaluate the variations in percentiles of IMDAA and IMD. Corresponding to each grid point, the 90th, 95th, and 99th percentile of precipitation for 21 years are calculated. Two groups are formed relative to IMDAA and IMD datasets using these percentiles as the three variables. Mahalanobis distance was calculated between these groups in annual and seasonal cases.

Mahalanobis distance considers correlation between variables explaining the variability between groups by measuring distance between centroids of clusters of different groups. Average MD values over the area are smaller for winter (1.29), monsoon (3.09) and post-monsoon (2.64) seasons while annual (8.38) and pre-monsoon (6.45) shows distinguishably high MD values. With reference to the IMD, IMDAA shows distinct percentile characteristics during different seasons as depicted by the MD values.

After calculating the distance, Hotelling's T^2 test was utilised to find the grid points with a significant difference in percentile values or which have high Mahalanobis distance. Spatial plots showing the grid points of significant variations are shown in Fig. 7. At 5% significance level, the annual percentile groups show the highest number of significantly different grids with account of 810 grids. Grid points spread out in the central regions of SPI show distinguishable variation in percentiles considering the yearly data (Fig. 7a). Telangana and different parts of Andhra Pradesh also contain grid points with percentile differences during the pre-monsoon season, as seen in Fig. 7c. 188 out of 810 grid points, mostly lying in Telangana and coastal regions of Andhra Pradesh shows significant difference during pre-monsoon season. Apart from the lower MD values, the similarity between IMDAA and IMD percentile groups for winter (1 grid), monsoon (19 grids) and post-monsoon (5 grids) seasons is again shown by the smaller number of significantly different grids. Indifferent behavioural patterns in percentiles are not displayed by IMDAA during monsoon and post-monsoon seasons as presented in Fig. 7d, e. In a nutshell, a comparison of IMDAA and IMD in terms of distance between groups dependent on percentiles resulted in the depiction of proficiency of reanalysis product in capturing precipitation to a great extent.

4.7 Spatial distribution of extreme rainfall episodes over SPI

The agricultural and economic growth of India is largely dependent on rainfall received by the country. The occurrence of extreme rainfall events over the country is exceeding the

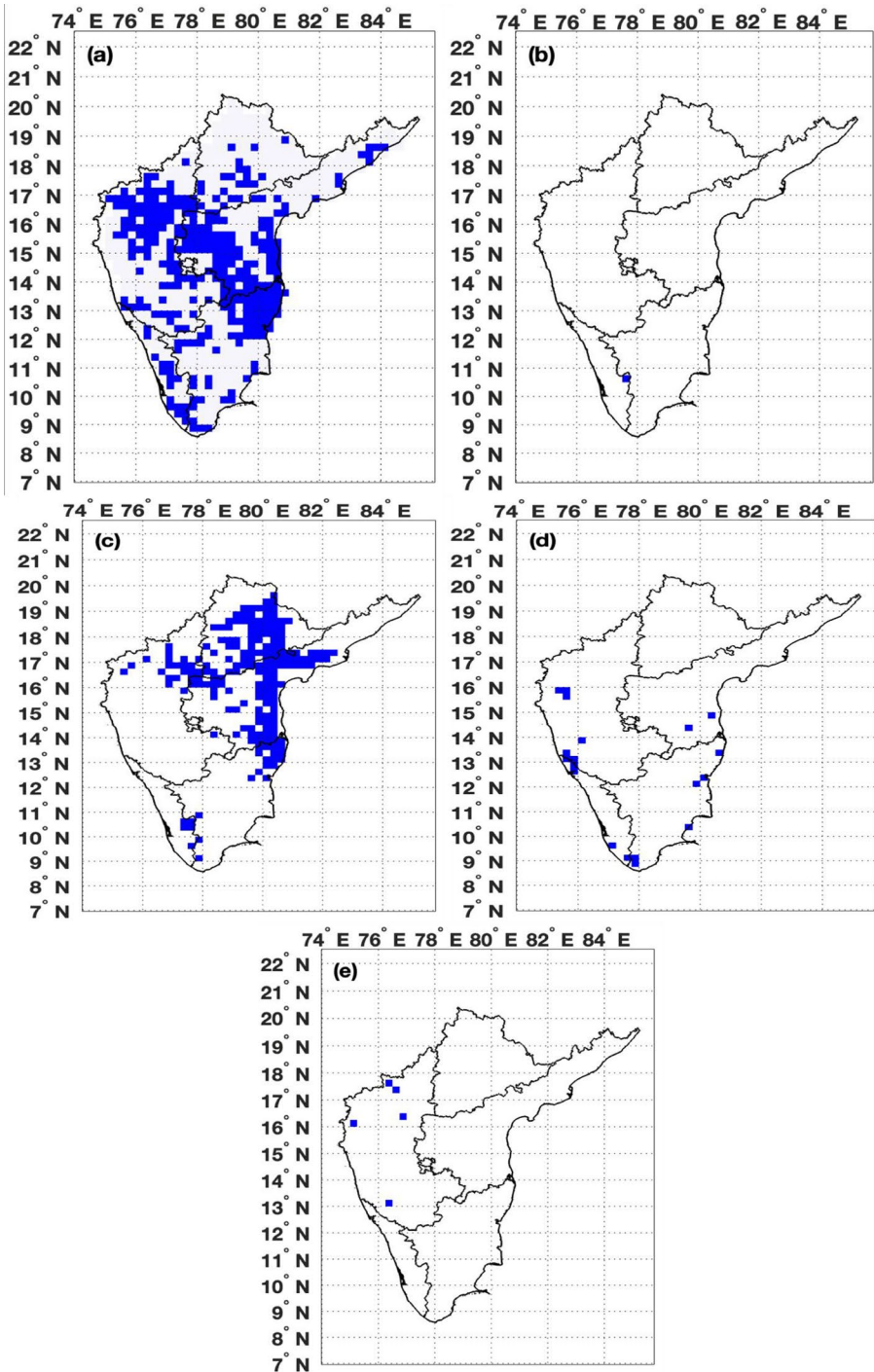


Fig. 7 Spatial plot of grid points showing significant Mahalanobis distance during a annual, b winter, c pre-monsoon, d monsoon and e post-monsoon

normal frequency for the past few years. Regional variations in land cover, urbanization, global warming, deforestation, etc. affect the amount of rainfall received by the country. Low-pressure systems over the Bay of Bengal and the increase in cyclonic activity over the Arabian sea are also causing extreme events in many parts of the country (Chaluvadi et al. 2021). To avert the significant economic damage caused by such catastrophic events, an effective reanalysis dataset is needed.

This section examines the statewide assessment of severe rainfall incidents that occurred between 2000 and 2020. Analyses are done to see how well IMDAA and IMD can estimate supposedly severe occurrences that have already taken place.

Between the Arabian Sea to the west and the Western Ghats to the east sits Kerala, which covers 38,863 km² of the total area of the mainland. Kerala has a humid marine climate with the monsoon season from June to August receiving an average of 3100–7030 mm annual rainfall. The state consists of land imposed by the Western Ghats composed of mountains, valleys, etc. About 44 rivers are present in Kerala, out of which 40 are originating in the Western Ghats. For the past few years, the frequency of flooding has increased in the state (Lal et al. 2020). The spatial distribution of average rainfall obtained by the area during extreme rainfall events from 6 to 18 August 2018 estimated by IMDAA and IMD is displayed in Fig. 8a, b. This extreme rainfall event caused havoc in the state with huge economic loss. Spatial distributions are plotted with the original resolutions of both datasets. That is, IMDAA is having 12 km resolution and IMD with 25 km resolution. IMDAA and IMD are exhibiting comparable spatial patterns from 6 August to 18 August 2018 with heavy rains in several areas of Kerala.

Tamil Nadu is one of the largest states in India, having a total area of 130,058 km². The Indian Ocean borders it on the south, the Bay of Bengal to the east, Kerala to the west, Karnataka to the northwest, and Andhra Pradesh to the north. Much of the rain-bearing clouds are obstructed by the Western Ghats, leading to variations in monsoon over Tamil Nadu, effectively causing droughts in some regions. But a distinct sudden rainfall is causing severe floods in Tamil Nadu. An extreme rainfall scenario is observed from 26th November to 2nd December 2015 (Singh et al. 2018). Figure 8c, d depicts spatial analysis on these rainy days. Results imply that IMDAA underestimated rain in some regions of Tamil Nadu as seen in the distribution image. Discrepancies between the IMDAA and IMD outcomes are seen in the figures where IMD showed heavy rainfall in coastal lines but IMDAA showed heavy rainfall in some grid points other than one shown by IMD.

Andhra Pradesh is located between latitudes 12° 41' N and 19.07°N and longitudes 77° E and 84° 40' E. Andhra Pradesh has the second longest coastal line in the country and consists of major parts of the Eastern Ghats and Deccan Plateau. A flood occurred from 29th September to 2nd October 2009 due to heavy precipitation in the cause of a mesoscale system (Kumar et al. 2012). Figure 8e, f represents these rainfall episodes and shows how well IMDAA and IMD captured this event. IMD fails in capturing these events properly, whereas IMDAA shows higher precipitation values in Andhra Pradesh during these days.

Telangana is a semi-arid state located on Deccan Plateau. Godavari and Krishna are the two major rivers in the state. A yearly rainfall of 700–900 mm is received by the southern region and 900–1500 mm by the northern region of the state, with major influence by the summer monsoon season. About 250 mm of rainfall is received by the north-western parts of the state from 23–24 September 2016 (Boyaj et al. 2020). On 24 September 2016, a rainfall of 390 mm was reported at the station Armoor (18.9° N, 78.29° E). These rainfall episodes are accurately estimated by the two products, as seen in Fig. 8g, h.

Karnataka is in the plateau of convergence of the Western Ghats and the Eastern Ghats. The topographical features of Karnataka are diverse, with forests, mountains, coastal

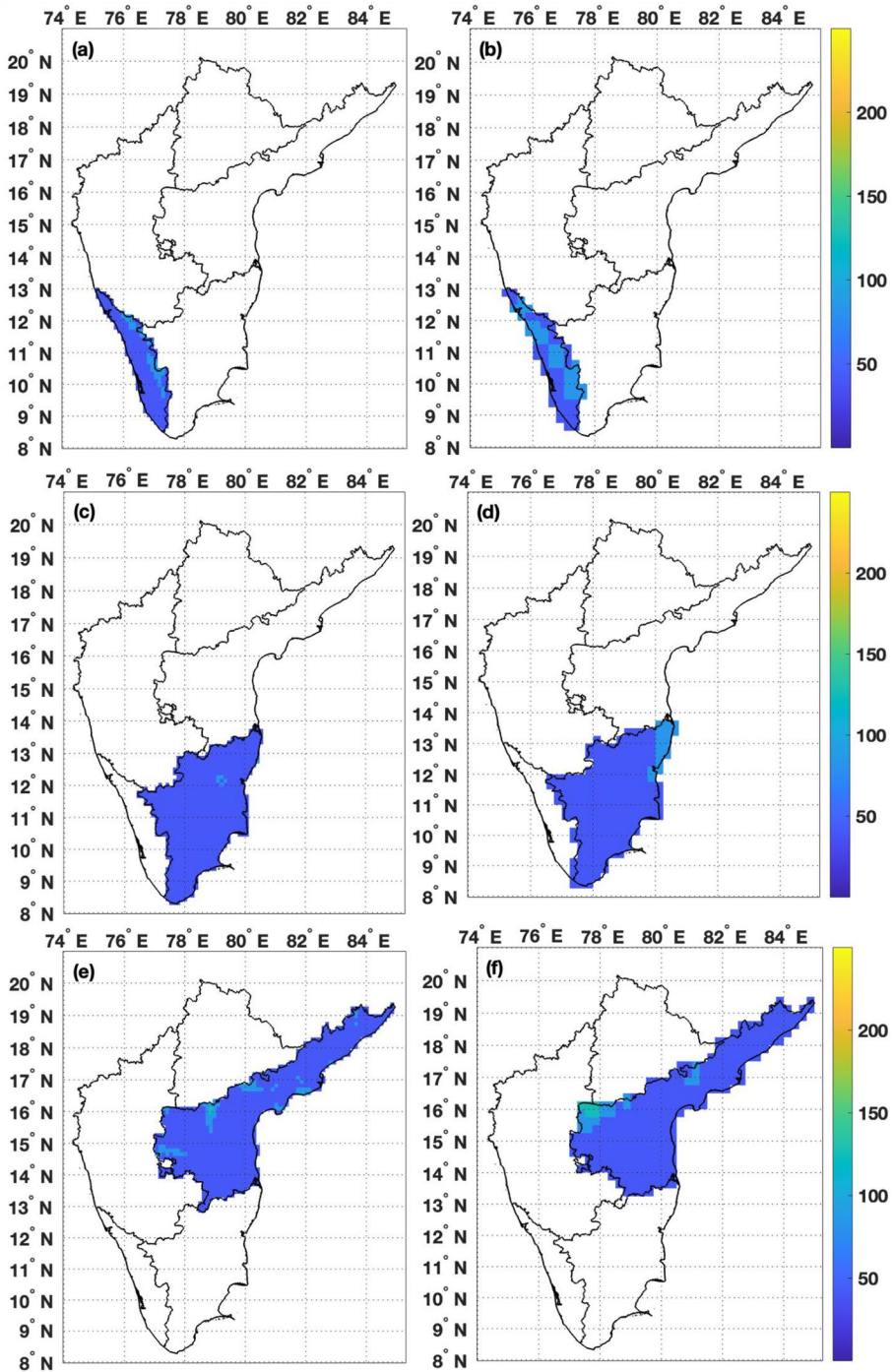


Fig. 8 Spatial distribution of extreme rainfall episodes estimated by IMDAA over **a** Kerala, **c** Tamil Nadu, **e** Andhra Pradesh, **g** Telangana, **i** Karnataka, and IMD over **b** Kerala, **d** Tamil Nadu, **f** Andhra Pradesh, **h** Telangana, **j** Karnataka

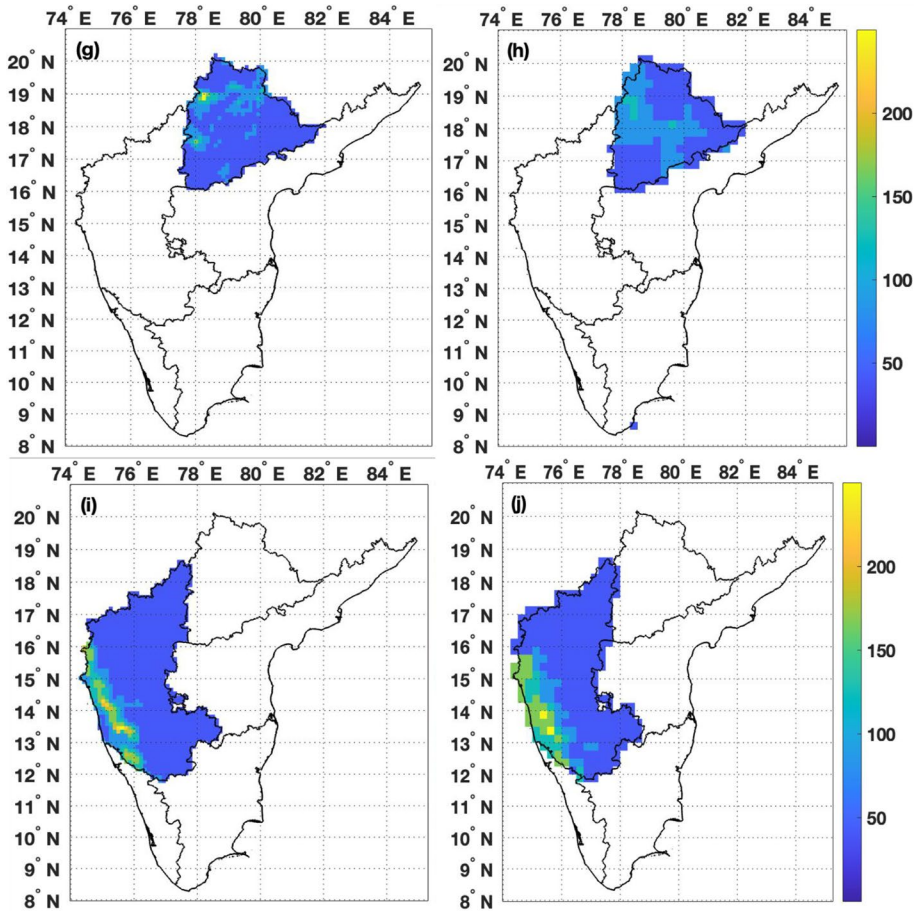


Fig. 8 (continued)

plains, and remnant hills. About 80% of the rainfall received by the state is during the southwest monsoon season. Depression in the Bay of Bengal caused heavy rainfall of around 224 mm in the state on 6–10 August 2019. It is evident from Fig. 8i, j that IMDAA and IMD well-captured the precipitation amount.

4.8 Extreme event prediction—utilisation of extreme value distribution

The accuracy of IMDAA in estimating the level of severe rainfall occurrences was examined in the preceding section. In this section, an extreme value distribution is utilised to predict extremes when given precipitation data. Extremes are predicted using precipitation data during each of the rainfall episodes mentioned in Sect. 4.7, so that the effectiveness of the data may be evaluated. Corresponding to each state's extreme rainfall days, extreme value distribution was used to generate grid points illustrating the extremes on the specified days. Gumbel distribution was used to fit the data and examine the extreme grid points. On each rainfall episodes, distribution was utilised to find out the grid points showing extremes. For the five states, precipitation data predicted by IMDAA and IMD

during extreme rainfall episodes happened in each state were observed. The observed rainfall during these episodes predicted by IMDAA and IMD were used to determine extremes, which in hand presented the difference between the reliability of IMDAA and IMD data in predicting extremes. The extreme grid points predicted by IMDAA and IMD are shown in Fig. 9.

In Fig. 9a, b, IMDAA is showing overall overestimation when compared to IMD, but over Kerala, IMDAA shows more grid points with extremes. Extreme valued grid points are similar for both IMDAA and IMD in the case of Tamil Nadu rainfall episodes (Fig. 9c, d). IMDAA and IMD are forecasting the same grid points with extreme value during the Andhra Pradesh extremes, with the exception of two or three grid points (Fig. 9e, f). IMDAA overestimated the number of extreme grid points but showed compatible results over Telangana (Fig. 9g, h).

A record amount of rainfall was received by Karnataka during 6–10 August 2019. Still, the extreme grid points detected using IMD data are unsatisfactory since no grid points exhibit extremes in Karnataka. In contrast, IMDAA spatial plot reveals most of the grid points which had extreme precipitation, especially over Karnataka. IMDAA precipitation data, as opposed to IMD precipitation data, accurately represents extremes, according to the findings. Also, IMDAA being 12 km resolution data available even on an hourly scale, utilisation of it for predicting extreme events stands to reason. The unreliable behaviour of IMD, since it failed to determine extreme grid points which was correctly predicted by IMDAA implies that usage of IMDAA helps in accurately predicting extremes which in turn helps assessing the spatiotemporal variability much better.

4.9 Teleconnection of ENSO with SPI rainfall

The interannual variability of tropical climate system is influenced by a large-scale phenomenon known as ENSO (El Niño-Southern Oscillation). Research suggests that an inverse relationship exists between ENSO and Indian summer monsoon rainfall. El Niño is referred to as a situation where five consecutive three month moving average Oceanic Niño Index Average (ONIA) values exceed $0.5\text{ }^{\circ}\text{C}$ (Azad and Rajeevan 2016).

An attempt is made to investigate the role of ENSO in the interannual variation of rainfall over SPI by determining how resonant the extreme years predicted using IMDAA with the ENSO signal. In general, El Niño events are associated with dry weather and La Niña occurrences are related to wet conditions. Figure 10 represents the interannual variability of standardised rainfall time series with depiction of El Niño and La Niña year's corresponding to IMDAA and IMD from 2000–2020.

El Niño years are shown in red and La Niña years are shown in blue colour. Fig. 10a, b implies that IMDAA and IMD behaves similarly considering the variability with respect to ENSO. El Niño and La Niña are the warm and cold phases of the climate patterns, respectively. In most of the El Niño years, normalised precipitation has got a negative value and for the La Niña years, a positive value can be seen which is in tune with the characteristics of El Niño and La Niña events. The La Niña years of 2000, 2007, 2008, 2010 and 2020 imply a positive normalised precipitation with consistent behaviour in both IMDAA and IMD. The wet year 2014 is not associated with La Niña episodes. Over the SPI, the relatively dry year of 2002 is not related to El Niño in case of IMD data, but it is related when normalised precipitation is from IMDAA. The wet years of 2006, 2015, 2016 and 2020 is not completely account for La Niña episodes. During the year of 2018 and 2019, the southern parts of SPI has experienced flood, though there was El Niño episodes. Even though

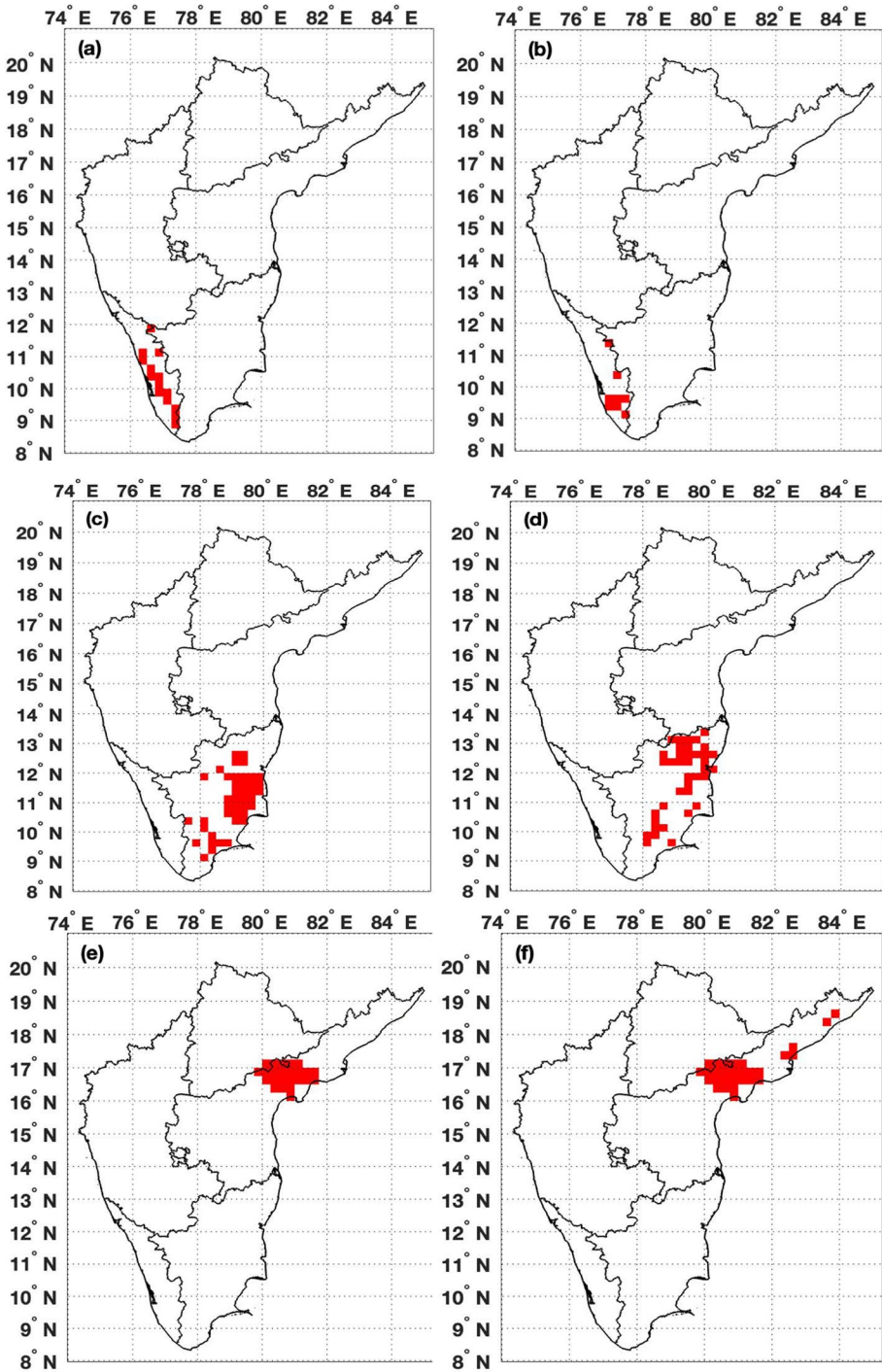


Fig. 9 Spatial plots of grid points with extreme rainfall events estimated using IMDAA over **a** Kerala, **c** Tamil Nadu, **e** Andhra Pradesh, **g** Telangana, **i** Karnataka, and IMD over **b** Kerala, **d** Tamil Nadu, **f** Andhra Pradesh, **h** Telangana, **j** Karnataka

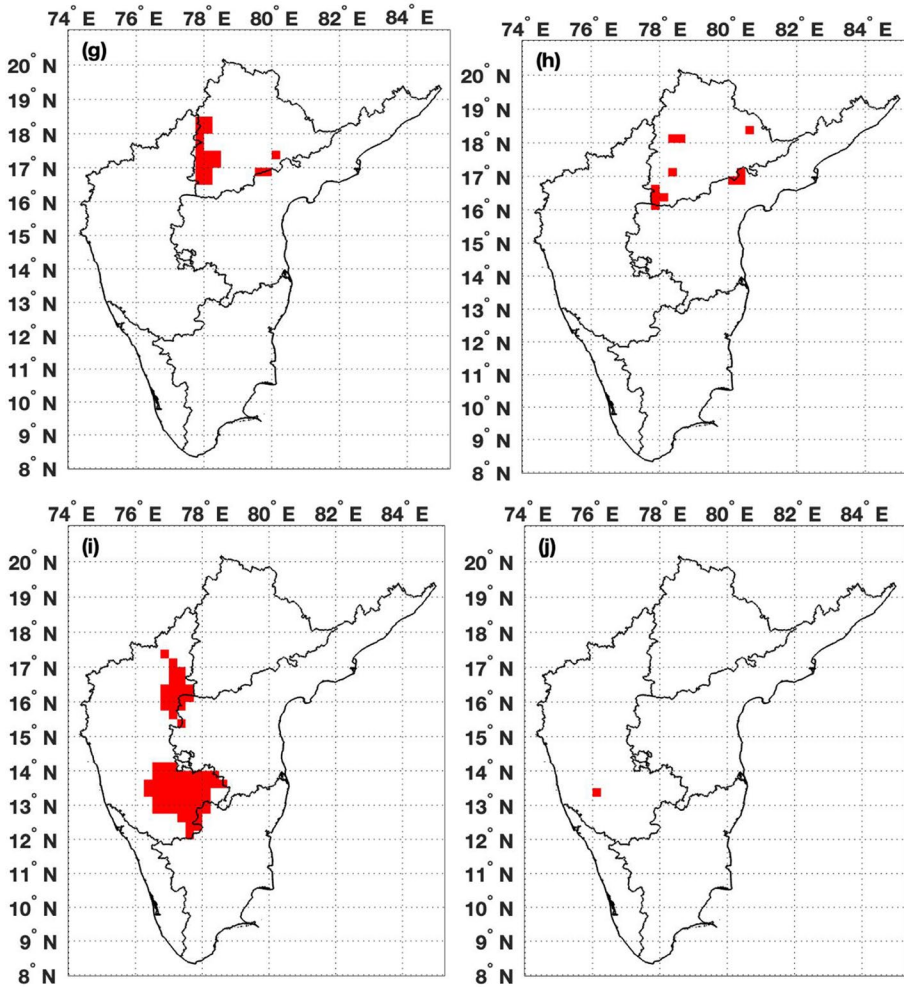


Fig. 9 (continued)

IMD normalised precipitation is positive, it is far less than IMDAA, which in turn represent the wet year more efficiently. Conclusively, it is obvious that the interannual variability is dependent on the ENSO events but cannot completely rely on the El Niño and La Niña episodes to predict extreme years.

5 Summary and conclusions

For the past few decades, South Peninsular India has been facing extreme rainfall events causing severe meteorological impacts, including agricultural and infrastructural losses. A robust reanalysis product with high resolution is required for predicting future forecasts to take adequate measurements and reduce risks of extreme events such as flash floods and landslides.

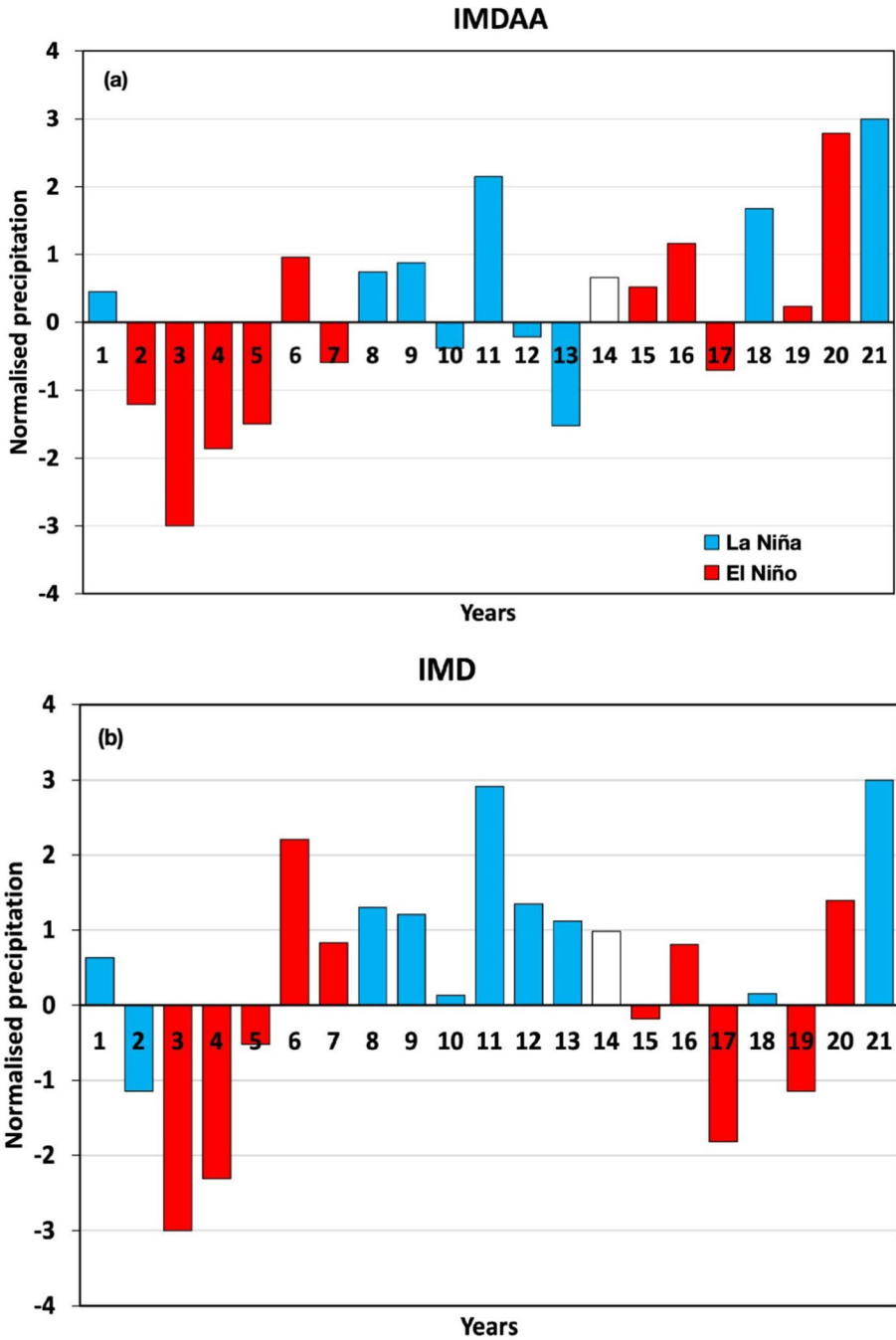


Fig. 10 Interannual variability of standardised rainfall time series with El Niño and La Niña year's corresponding to a IMDAA and b IMD from 2000 to 2020

A high-resolution 12 km gridded rainfall dataset produced by National Centre for Medium Range Weather Forecasting (NCMRWF), namely IMDAA, is analysed to estimate rainfall characteristics and extreme events over South Peninsular India during 2000–2020. IMD daily gridded rainfall dataset is explored for the benchmark of the study. The datasets considered are acquired using different assimilation techniques and can be utilised for comparison without any overlapping. Proper assessment of monthly, seasonal and annual scale precipitation estimation by IMDAA is done with IMD observations. Investigation of spatial trends of mean precipitation during seasonal and annual scale, statewide rainfall episodes, identification of significant differences in precipitation using distance metric known as Mahalanobis distance, and prediction of extreme events using extreme value distribution is done. Several visual and statistical assessment is also implemented, such as calculating several statistical measures and graphical representation of detection capability using a Performance diagram. The major conclusions of the research can be summed up as follows:

1. IMDAA slightly overestimated average precipitation over SPI throughout all the years under investigation. Seasonal scale measurements reveal that the IMDAA's assessment varies from season to season. Daily average precipitation obtained by IMDAA over SPI during 2000–2020 represents a close correlation with IMD observations.
2. Utilisation of Student's *t* test to find the grid points showing significant variations in average precipitation over the years implied that the number of grid points indicating substantial difference is more prominent in winter and post-monsoon season compared to monsoon and pre-monsoon season, according to a spatial plot displaying grid points exhibiting significant differences throughout each season and yearly.
3. Correlation values are pretty high for all monthly, seasonal, and annual cases, but RMSE, RB, and MAE values vary in a similar pattern. Both seasonal and monthly statistics show identical characteristics, such as RMSE, RB, and MAE being smaller for the winter season, with much higher values till starting of the monsoon season and decreasing from July to December. Index of Agreement values is close to one in all cases. NSE and KGE values are also changing concerning the months or seasons of study. It is clear that, even though the statistical metrics show different values based on the seasons, all metrics are close to optimal values.
4. Graphical representation of precipitation detection capacity of IMDAA during daily, annual, spatial and percentile using Performance diagram indicates that POD and SR are close to one in most cases. For the annual scale, yearly and seasonal plots, IMDAA exhibits good performance. Daily scale performance of IMDAA is better in annual and post-monsoon seasons, whereas over the spatial scale, it performs best during monsoon season. 90th and 95th percentile performance diagrams revealed comparable characteristics.
5. Mahalanobis distance for percentiles implied that monsoon and post-monsoon seasons demonstrated strong compatibility between IMDAA and IMD; however, annual and pre-monsoon seasons displayed some considerably dissimilar grid points in terms of distance.
6. The five states constituting SPI are diverse in their climatology. Each state is impacted by rainfall differently due to its unique geographic and climatic characteristics. Spatial distributions of rainfall episodes of each state acted differently. For the state of Kerala, IMDAA and IMD exhibited similar distributions. A dry bias was shown by IMDAA over Tamil Nadu for precipitation from 26th November to 2nd December 2015. On the

- other hand, IMD could not effectively capture the rainfall episodes from 29th September to 2nd October 2009 over Andhra Pradesh, while IMDAA successfully captured the rainfall. IMDAA indicated the reported rainfall exactly over Telangana and Karnataka.
7. The extreme value distribution overestimated extreme grid points by IMDAA data in the Telangana and Kerala rainfall episodes. Extreme grid points were satisfactorily the same for IMDAA and IMD in the case of Tamil Nadu and Telangana. However, IMDAA correctly detected extreme grid points in the Karnataka region while IMD failed to do so.
 8. ENSO phenomenon impacts the tropical climate including Indian monsoon. Interannual variability of standardised rainfall time series of IMDAA and IMD with respect to El Niño and La Niña years indicated that extreme years are resonant with the ENSO signal over the region but cannot completely rely on ENSO to predict extremes.

The results suggest that IMDAA reanalysis dataset is better at estimating accurate precipitation over SPI. The research study also gives insight into the certainty that IMDAA data may be used to forecast extreme rainfall events. The overestimation of rainfall amount estimated by IMDAA in some seasons and regions can be corrected using bias correction. The future study will include the bias correction of the reanalysis product and consider other factors responsible for the extreme events.

Acknowledgements Authors gratefully acknowledge NCMRWF, Ministry of Earth Sciences, Government of India, for IMDAA reanalysis. IMDAA reanalysis was produced under the collaboration between UK Met Office, NCMRWF, and IMD with financial support from the Ministry of Earth Sciences, under the National Monsoon Mission programme. Authors also thankfully acknowledge Indian Meteorological Department (IMD) for providing daily gridded precipitation datasets. The authors are thankful to Guoqiang Tang, University of Saskatchewan Coldwater Lab, Canada, for suggesting ideas in implementing some codes.

Author contributions Both the authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Sneha M R. The first draft of the manuscript was written by Sneha M R and Archana Nair commented on previous versions of the manuscript. Archana Nair read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Availability of data and materials The datasets generated during and/or analysed during the current study are available in the RDS NCMRWF repository, (<https://rds.ncmrwf.gov.in/dashboard/download>) and IMD repository (<https://mausam.imd.gov.in>).

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval The submitted manuscript is original and has not been submitted or published in any other language or form. The results are not manipulated, and the authors did not use any fabricated data.

References

- Aggarwal D, Attada R, Shukla KK, Chakraborty R, Kunchala RK (2022) Monsoon precipitation characteristics and extreme precipitation events over Northwest India using Indian high resolution regional reanalysis. *Atmospheric Res* 267:105993. <https://doi.org/10.1016/j.atmosres.2021.105993>
- Ali H, Mishra V, Pai DS (2014) Observed and projected urban extreme rainfall events in India. *J Geophys Res Atmospheres* 119(22):12–621. <https://doi.org/10.1002/2014JD022264>

- Ashrit R, Indira Rani S, Kumar S, Karunasagar S, Arulalan T, Francis T et al (2020) IMDAA regional reanalysis: performance evaluation during Indian summer monsoon season. *J Geophys Res Atmospheres* 125(2):e2019JD030973. <https://doi.org/10.1029/2019JD030973>
- Azad S, Rajeevan M (2016) Possible shift in the ENSO-Indian monsoon rainfall relationship under future global warming. *Sci Rep* 6(1):20145. <https://doi.org/10.1038/srep20145>
- Baki H, Chinta S, Balaji C, Srinivasan B (2021) Determining the sensitive parameters of WRF model for the prediction of Tropical cyclones in Bay of Bengal using Global sensitivity analysis and Machine learning. *arXiv prepr. arXiv:2107.04824*. <https://doi.org/10.5194/gmd-15-2133-2022>
- Bandara U, Agarwal A, Srinivasan G, Shanmugasundaram J, Jayawardena IS (2022) Intercomparison of gridded precipitation datasets for prospective hydrological applications in Sri Lanka. *Int J Climatol* 42(6):3378–3396. <https://doi.org/10.1002/joc.7421>
- Barrie A, Scott MB (2021) Intensity duration frequency relationship of maximum rainfall in a data scarce urbanized environment: a case study of the Guma Catchment in Sierra Leone. *Int J Water Resources Environ Eng* 13(2):123–134. <https://doi.org/10.5897/IJWREE2020.0938>
- Bhagat N (2017) Flood frequency analysis using Gumbel's distribution method: a case study of Lower Mahi Basin, India. *J Water Resources Ocean Sci* 6(4):51–54. <https://doi.org/10.11648/j.wros.20170604.11>
- Blacutt LA, Herdies DL, de Gonçalves LGG, Vila DA, Andrade M (2015) Precipitation comparison for the CFSR, MERRA, TRMM3B42 and Combined Scheme datasets in Bolivia. *Atmospheric Res* 163:117–131. <https://doi.org/10.1016/j.atmosres.2015.02.002>
- Bollmeyer C, Keller JD, Ohlwein C, Wahl S, Crewell S, Friederichs P et al (2015) Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q J R Meteorol Soc* 141(686):1–15. <https://doi.org/10.1002/qj.2486>
- Boyaj A, Dasari HP, Hoteit I, Ashok K (2020) Increasing heavy rainfall events in south India due to changing land use and land cover. *Q J R Meteorol Soc* 146(732):3064–3085. <https://doi.org/10.1002/qj.3826>
- Chakraborty S, Saha U, Maitra A (2015) Relationship of convective precipitation with atmospheric heat flux—a regression approach over an Indian tropical location. *Atmospheric Res* 161:116–124. <https://doi.org/10.1016/j.atmosres.2015.04.008>
- Chakraborty P, Sarkar A, Bhatla R, Singh R (2021) Assessing the skill of NCMRWF global ensemble prediction system in predicting Indian summer monsoon during 2018. *Atmospheric Res* 248:105255. <https://doi.org/10.1016/j.atmosres.2020.105255>
- Chaluvadi R, Varikoden H, Mujumdar M, Ingle ST, Kuttippurath J (2021) Changes in large-scale circulation over the Indo-Pacific region and its association with 2018 Kerala extreme rainfall event. *Atmospheric Res* 263:105809. <https://doi.org/10.1016/j.atmosres.2021.105809>
- Chen S, Zhang L, Zhang Y, Guo M, Liu X (2020) Evaluation of Tropical Rainfall Measuring Mission (TRMM) satellite precipitation products for drought monitoring over the middle and lower reaches of the Yangtze River Basin, China. *J Geograph Sci* 30:53–67. <https://doi.org/10.1007/s11442-020-1714-y>
- Dahlgren P, Landelius T, Källberg P, Gollvik S (2016) A high-resolution regional reanalysis for Europe. Part I: Three-dimensional reanalysis with the regional High-Resolution Limited-Area Model (HIRLAM). *Q J R Meteorol Soc* 142(698):2119–2131. <https://doi.org/10.1002/qj.2807>
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. <https://doi.org/10.1002/qj.828>
- Fortelius C, Andr e U, Forsblom M (2002) The BALTEX regional reanalysis project. *Boreal Environ Res* 7(3):193
- Ghosh S, Vittal H, Sharma T, Karmakar S, Kasiviswanathan KS, Dhanesh Y et al (2016) Indian summer monsoon rainfall: Implications of contrasting trends in the spatial variability of means and extremes. *PLoS ONE* 11(7):e0158670. <https://doi.org/10.1371/journal.pone.0158670>
- Haile AT, Tefera FT, Rientjes T (2016) Flood forecasting in Niger-Benue basin using satellite and quantitative precipitation forecast data. *Int J Appl Earth Observ Geoinf* 52:475–484. <https://doi.org/10.2166/h2oj.2021.094>
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteor Soc* 77(3):437–472. [https://doi.org/10.1175/1520-0477\(1996\)077%3c0437:TNYRP%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077%3c0437:TNYRP%3e2.0.CO;2)
- Kharin VV, Zwiers FW, Zhang X, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J Clim* 20(8):1419–1444. <https://doi.org/10.1175/JCLI4066.1>
- Kishore P, Jyothi S, Basha G, Rao SVB, Rajeevan M, Velicogna I, Sutterley TC (2016) Precipitation climatology over India: validation with observations and reanalysis datasets and spatial trends. *Clim Dyn* 46:541–556. <https://doi.org/10.1007/s00382-015-2597-y>

- Kobayashi S, Ota Y, Harada Y, Ebata A, Moriya M, Onoda H et al (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *J Meteorol Soc Jpn Ser II* 93(1):5–48. <https://doi.org/10.2151/jmsj.2015-001>
- Kumar OB, Suneetha P, Rao SR, Kumar MS (2012) Simulation of heavy rainfall events during retreat phase of summer monsoon season over parts of Andhra Pradesh. *Int J Geosci* 3(4):737. <https://doi.org/10.4236/ijg.2012.34074>
- Lal P, Prakash A, Kumar A, Srivastava PK, Saikia P, Pandey AC et al (2020) Evaluating the 2018 extreme flood hazard events in Kerala, India. *Remote Sens Lett* 11(5):436–445. <https://doi.org/10.1080/2150704X.2020.1730468>
- Lyngwa RV, Nayak MA (2021) Atmospheric river linked to extreme rainfall events over Kerala in August 2018. *Atmos Res* 253:105488. <https://doi.org/10.1016/j.atmosres.2021.105488>
- Mahalanobis PC (1930) On test and measures of group divergence: theoretical. *J as Soc Beng* 26:541–588
- Mahmood S, Davie J, Jerney P, Renshaw R, George JP, Rajagopal EN, Rani SI (2018) Indian monsoon data assimilation and analysis regional reanalysis: configuration and performance. *Atmos Sci Lett* 19(3):e808. <https://doi.org/10.1002/asl.808>
- Mamman MJ, Martins OY, Ibrahim J, Shaba MI (2017) Evaluation of best-fit probability distribution models for the prediction of inflows of Kainji Reservoir, Niger State, Nigeria. *Air Soil Water Res.* <https://doi.org/10.1177/1178622117691034>
- Matcharashvili T, Zhukova N, Chelidze T, Founda D, Gerasopoulos E (2017) Analysis of long-term variation of the annual number of warmer and colder days using Mahalanobis distance metrics—a case study for Athens. *Physica A* 487:22–31. <https://doi.org/10.1016/j.physa.2017.05.065>
- Mausson F, Scherer D, Mölg T, Collier E, Curio J, Finkelnburg R (2014) Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia Reanalysis. *J Clim* 27(5):1910–1927. <https://doi.org/10.1175/JCLI-D-13-00282.1>
- Meeyaem K, Polpinit P (2014) Mathematical model for flood forecasting of the Chi River Basin. In: International proceedings of chemical, biological and environmental engineering (IPCBBE), vol 63, pp 5–9
- Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W et al (2006) North American regional reanalysis. *Bull Am Meteor Soc* 87(3):343–360. <https://doi.org/10.1175/BAMS-87-3-343>
- Mimmack GM, Mason SJ, Galpin JS (2001) Choice of distance matrices in cluster analysis: defining regions. *J Clim* 14(12):2790–2797. <https://doi.org/10.7916/d8-7gwk-z354>
- Moazami S, Golian S, Hong Y, Sheng C, Kavianpour MR (2016) Comprehensive evaluation of four high-resolution satellite precipitation products under diverse climate conditions in Iran. *Hydrol Sci J* 61(2):420–440. <https://doi.org/10.1080/02626667.2014.987675>
- Naz S, Baig MJ, Inayatullah S, Siddiqi TA, Ahsanuddin M (2019) Flood Risk Assessment of Guddu Barrage using Gumbel's Distribution. *Int J Sci* 8(04):33–38
- Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H et al (2007) The JRA-25 reanalysis. *J Meteorol Soc Jpn Ser II* 85(3):369–432. <https://doi.org/10.2151/jmsj.85.369>
- Pai DS, Rajeevan M, Sreejith OP, Mukhopadhyay B, Satbha NS (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. <https://doi.org/10.54302/mausam.v65i1.851>
- Phien HN (1987) A review of methods of parameter estimation for the extreme value type-I distribution. *J Hydrol* 90(3–4):251–268. [https://doi.org/10.1016/0022-1694\(87\)90070-9](https://doi.org/10.1016/0022-1694(87)90070-9)
- Punde P, Attada R, Aggarwal D, Radhakrishnan C (2022) Numerical simulation of winter precipitation over the Western Himalayas using a weather research and forecasting model during 2001–2016. *Climate* 10(11):160. <https://doi.org/10.3390/cli10110160>
- Rajeevan M, Bhate J, Kale JD, Lal B (2006) High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells. *Curr Sci* 91:296–306
- Rajeevan M, Bhate J, Jaswal AK (2008) Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys Res Lett.* <https://doi.org/10.1029/2008GL035143>
- Rani SI, Arulalan T, George JP, Rajagopal EN, Renshaw R, Maycock A et al (2021) IMDAA: High-resolution satellite-era reanalysis for the Indian monsoon region. *J Clim* 34(12):5109–5133. <https://doi.org/10.1175/JCLI-D-20-0412.1>
- Rasel MM, Islam MM (2015) Generation of rainfall intensity-duration-frequency relationship for North-Western Region in Bangladesh. *IOSR J Environ Sci Toxicol Food Technol* 9(9):41–47. <https://doi.org/10.9790/2402-09914147>
- Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E et al (2011) MERRA: NASA's modern-era retrospective analysis for research and applications. *J Clim* 24(14):3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>

- Rivas DV, Pino MM, Pérez YP, Rivas SV, Torres OG, Fernández VC, Ysa RS (2020) Comparison of distance methods for detection of atypical observations in monthly precipitation series. *Revista de Climatología* 20
- Saha S, Moorthi S, Pan HL, Wu X, Wang J, Nadiga S et al (2010) The NCEP climate forecast system reanalysis. *Bull Am Meteor Soc* 91(8):1015–1058. <https://doi.org/10.1175/2010BAMS3001.1>
- Sharma N, Attada R, Hunt KM (2022) Evaluating winter precipitation over the western Himalayas in a high-resolution Indian regional reanalysis using multi-source climate datasets. *J Appl Meteorol Climatol*. <https://doi.org/10.1175/JAMC-D-21-0172.1>
- Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 23rd ACM national conference*, pp 517–524. <https://doi.org/10.1145/800186.810616>
- Singh AK, Singh V, Singh KK, Tripathi JN, Kumar A, Soni AK et al (2018) A case study: Heavy rainfall event comparison between daily satellite rainfall estimation products with IMD gridded rainfall over peninsular India during 2015 winter monsoon. *J Indian Soc Remote Sens* 46:927–935. <https://doi.org/10.1007/s12524-018-0751-9>
- Singh T, Saha U, Prasad VS, Gupta MD (2021) Assessment of newly-developed high resolution reanalyses (IMDAA, NGFS and ERA5) against rainfall observations for Indian region. *Atmospheric Res* 259:105679. <https://doi.org/10.1016/j.atmosres.2021.105679>
- Solomon O, Prince O (2013) Flood frequency analysis of Osse river using Gumbel's distribution. *Civil Environ Res* 3(10):55–59
- Tait A, Sturman J, Clark M (2012) An assessment of the accuracy of interpolated daily rainfall for New Zealand. *J Hydrol (New Zealand)* 25–44. <http://www.jstor.org/stable/43944886>
- Tan ML, Santo H (2018) Comparison of GPM IMERG, TMPA 3B42 and PERSIANN-CDR satellite precipitation products over Malaysia. *Atmospheric Res* 202:63–76. <https://doi.org/10.1016/j.atmosres.2017.11.006>
- Tang G, Clark MP, Papalexiou SM, Ma Z, Hong Y (2020) Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sens Environ* 240:111697
- Thameemul Hajaj PM, Yarrakula K, Durga Rao KHV, Singh A (2019) A semi-distributed flood forecasting model for the Nagavali River using space inputs. *J Indian Soc Remote Sens* 47:1683–1692. <https://doi.org/10.1007/s12524-019-01019-0>
- Trenberth KE (2005) The impact of climate change and variability on heavy precipitation, floods, and droughts. *Encyclopedia Hydrol Sci* 17:1–11. <https://doi.org/10.1002/0470848944.hsa211>
- Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 30(1):79–82. <https://doi.org/10.3354/cr030079>
- Zhang Q, Pan Y, Wang S, Xu J, Tang J (2017) High-resolution regional reanalysis in China: evaluation of 1 year period experiments. *J Geophys Res Atmospheres* 122(20):10–801. <https://doi.org/10.1002/2017JD027476>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.