https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

Probabilistic seasonal outlook for the rainy season over India by monitoring the onset dates using GPM IMERG satellite-based precipitation

Jayasankar Chempampadam Balasubramannian[1, 2], and Vasubandhu Misra[1, 2, 3]

[1]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, USA.

[2]Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida, USA.

[3]Florida Climate Institute, Florida State University, Tallahassee, Florida, USA.

Corresponding Author (Jayasankar Chempampadam Balasubramannian, cbjayasankar@gmail.com)

Abstract

We utilized the Integrated Multi-Satellite Retrievals for Global Precipitation Mission version 6 (IMERG) rainfall observation (available in real time) over India to determine the onset and demise of the rainy season. The annual mean climatology derived from IMERG observations over India aligned closely with the rain gauge-based India Meteorological Department observation. The IMERG rainfall time series was randomly perturbed to generate 101 ensemble members at every grid point of the rainfall analysis to obtain a corresponding ensemble of the onset and demise dates of the rainy season. The perturbations were designed to sample the uncertainty due to random synoptic or mesoscale rain events influencing the diagnosis of the onset/demise dates at the granularity of the IMERG observations (at 10 km grid). Following earlier studies, we find from the IMERG dataset that seasons with an earlier onset date are strongly related to a lengthier and wetter season, whereas seasons with a later onset date correspond to a shorter and drier season. In contrast, the connections between the onset, demise, seasonal length, and rainfall with ENSO and IOD were comparatively weaker over most of India. The generation of ensembles in this study underscores the potential for real-time application of generating reliable, probabilistic seasonal outlooks of the rainy season over India by leveraging the local links amongst onset date, seasonal length, and seasonal rainfall anomalies. This potential is further confirmed by the high probabilistic skill scores of the seasonal outlooks using the area under the relative operating characteristic curve method.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

1. Introduction

India's water and agriculture sectors rely heavily on copious summer monsoon rainfall, emphasizing its critical role in defining the country's agrarian economy and water resource management (Gadgil and Gadgil 2006). The seasonal progression of the Indian Summer Monsoon (ISM) season has an impact not only on agricultural outputs but also on other sectors such as economy and ecology (Lal 2000). Based on rainfall patterns and wind conditions, the India Meteorological Department (IMD) officially declares the date of monsoon onset over Kerala (MoK) each year. In a normal year, the ISM season begins in southern peninsular India on June 1st (with MoK), spreads throughout the country by mid-July, and then begins to retreat on September 1st (Pai et al. 2020).

There are many different definitions for the monsoon onset; however, it is typically considered as a quick, substantial, and prolonged increase in rainfall after May 10th (Pai and Rajeevan 2009). In addition to the rainfall-based definitions (e.g., Ananthakrishnan and Soman 1988; Noska and Misra 2016; Misra et al. 2017b), several other atmospheric dynamical and thermodynamical variables (winds, outgoing longwave radiation, temperature, moisture flux convergence, and precipitable water) were utilized to identify the onset of summer monsoon rainfall (Fasullo and Webster 2003; Zeng and Lu 2004; Prasad and Hayashi 2005; Joseph et al. 2006; Wang et al. 2009; Walker and Bordoni 2016; Stolbova et al. 2016). These variables provide a comprehensive understanding of the atmospheric conditions and dynamics that influence the onset and seasonal total rainfall over India. All these approaches produce somewhat identical onset dates, but their connections with the local-scale onset of the rainfall during the ISM differ (Moron and Robertson 2014; Noska and Misra 2016).

Owing to variations in the onset and demise date of the monsoon season there are considerable spatial and temporal variations in the availability of the rainwater and length of the rainy season over India (Joseph et al. 1994; Lal 2000; Wang et al. 2002; Misra et al. 2017a; Misra et al. 2018). Preenu et al. (2017) indicate from IMD archive that the earliest start of the ISM was May 11, 1918, and the most delayed onset was Jun 18, 1972. Several studies discuss the interannual variability of onset dates in India and its teleconnection with El Niño and the Southern Oscillation (ENSO) and

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

the Indian Ocean Dipole (IOD, Xavier et al. 2007; Sankar et al. 2010; Misra et al. 2017a; Pradhan et al. 2017; Choudhury et al. 2021). In ENSO and IOD years, the monsoon onset is altered by the modulation of the teleconnections with SST anomalies in the tropical oceans, which affect the Walker and Hadley circulations, respectively (Pradhan et al. 2017). Significant changes in the large-scale atmospheric patterns over the monsoon areas are identified during the onset of the monsoon over India (Joseph et al. 1994). According to Wang et al. (2013), the Interannual variations in ocean-atmosphere interaction processes significantly influence the association between the monsoon onset and ENSO, which is driven by the meridional SST gradient across the Indian Ocean. Noska and Misra (2016) show that the variations in the onset and demise date of the rainy season over India that largely overlaps with ISM are linked to the variability of the cross-equatorial upper-ocean heat transport in the Indian Ocean, variations in large-scale atmospheric and oceanic circulations, and regional ocean-atmosphere thermal gradients.

It may be noted that we make a subtle distinction between the ISM and rainy seasons, with the latter solely determined by rain rates (e. g., Noska and Misra 2016) while the former is additionally influenced by circulation features and other thermodynamic factors. Figure S1 highlights the difference between the variable season and fixed southwest monsoon season (June to September), which makes it evident that the variable season accumulates more rainfall in southern peninsular India and in regions east of the central monsoon region. Similarly, when compared to the fixed season, the variable season has higher total rainfall over both western and northern India. Notably, pre-monsoon rainfall and rainfall from pre-monsoon and post-monsoon tropical cyclones are also included in the rainy season defined here, and they significantly contribute to the annual rainfall. Pre- and post-monsoon tropical cyclones can contribute up to 25% of the yearly rainfall in some parts of India, even though the southwest monsoon is the main rainy season (Khouakhi et al. 2017).

The information about the arrival of the monsoon rainfall is imperative for farmers to plan their strategy for the upcoming season. There are many attempts to predict the MoK (Kung and Shariff 1980; Rajeevan and Dubey 1995; Pai and Rajeevan 2007; Pradhan et al. 2017). But many studies have shown that variations of the MoK have very little influence on the seasonal rainfall anomalies of the ISM (Dhar et al. 1980; Mooley and Parthasarthy 1984; Mooley and Shukla 1987; Misra and DiNapoli 2014). Furthermore, MoK is known to have insignificant relationship with subsequent

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

progression of the onset isochrones of the ISM (Bansod et al. 1991; Fasullo and Webster 2003; Pai and Rajeevan 2007). Noska and Misra (2016) proposed an objective method to define the onset and demise of the rainy season based on area averaged all India daily rainfall, and they found that seasonal mean anomalies are closely linked with variations in the onset and demise dates. In a following study, Misra et al. (2018) obtained onset (demise) dates of the Indian rainy season at the granularity of the rainfall analysis and showed that they are negatively (positively) correlated with seasonal rainfall anomalies across the Indian region. These studies suggest that monitoring the onset date provides a good indication of the evolution of the seasonal length and rainfall of the forthcoming rainy season. Bhardwaj and Misra (2019) found that Remotely Sensed Rainfall Products such as Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) produce onset and demise dates similar to IMD observations. This simple rainfall-based definition was found useful for monitoring the onset in real-time and issuing the seasonal outlook over various regions such as Florida (Misra et al. 2022), and Central America (Rodgers et al. 2024).

Many studies have used the IMERG data set as one of their primary dataset for the analysis of the ISM (Bushair et al. 2019; Thakur et al. 2020; Saikrishna et al. 2021; Phadtare et al. 2023). In this study, we employ a rainfall-based objective method to define the onset and demise of the rainy season, leveraging the high-resolution IMERG dataset available since January 2001. Utilizing its 12-hour latency product, we demonstrate the potential for real-time monitoring of the onset of the rainy season, which aids in anticipating the anomalies of the seasonal length and seasonal rainfall across India. Motivated from the grid-point level onset and demise dates established by Misra et al. (2018) and Bhardwaj and Misra (2019), this study advances this framework by how the IMERG 12-hour latency product can be effectively used to monitor these parameters at fine spatial scales, enabling better anticipation of anomalies of the seasonal length and seasonal rainfall. In addition, the novelty of this work lies in the use of a perturbation technique applied to rainy season following Misra et al. (2023) and Rodgers et al. (2024) to generate 101 ensembles of daily rainfall data. This approach is the first of its kind over the Indian region and allows us to define the onset and demise dates, providing a probabilistic estimate of the evolution of the rainy season that accounts for observational and analysis uncertainties.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

Additionally, this study explores the interannual variation of the evolution of the rainy season and its link to large-scale forcing such as ENSO and IOD and compares their viability as predictors of the rainy seasons to the local links we establish with the variations of the onset dates. We believe that the insights from this study will have significant potential applications in agricultural planning and water resource management.

2. Data and Methodology

2.1 Data

The analysis performed in this study utilized daily precipitation data from the IMERG version 6 (Huffman et al. 2019). This dataset is part of the Global Precipitation Measurement (GPM) mission, which was launched in 2014 and is co-operated by NASA and the Japan Aerospace Exploration Agency (JAXA). The 0.1° grid resolution (~10 km) IMERG data are available at half-hourly intervals from June 2000 to the present. The dataset comprises early, late, and final run products, which have latencies of approximately 4 hours (Early), 12 hours (Late), and 3.5 months (Final). The IMERG late run incorporates data from multiple sources, such as satellite microwave and infrared estimations, precipitation gauge analysis, and other potential precipitation estimators. This dataset provides detailed temporal and spatial coverage for both the TRMM and GPM eras worldwide. However, in this study, we used the daily 12-hour latency product (daily averages were estimated from half-hourly products), which has great potential for monitoring the rainy season in real-time. To assess the fidelity of the IMERG product over India we have used rain gauge-based IMD gridded rainfall data for the period 2001 - 2023 at 0.25° × 0.25° resolution (Pai et al. 2014) as a validation dataset. Additionally, we defined the onset and demise dates and compared them with the IMERG using the daily NOAA Climate Prediction Center (CPC) precipitation data set (Xie et al. 2007).

2.2 Methodology

2.2.1 Onset and Retreat

This study uses a simple objective definition to identify the onset and cessation dates of the wet season over India which are determined by finding the minima and maxima of the cumulative rainfall anomaly curve (Liebmann and Marengo 2001). For a year i, the cumulative anomaly of the daily rainfall at a day j ($i.e.,\ C_i(j)$) at each grid point is estimated as;

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

$$C_i(j) = \sum_{j=1}^{J}(R_i(j) - \bar{R}) \tag{1}$$

where $R_i(j)$ is the rainfall for the day j, $\bar{R}$ is the climatology of the annual mean rainfall. The $C_i(j)$ for 365/366 days in a year represents the cumulative rainfall anomaly curve. After the onset and demise dates of the season are determined, we estimate the number of days between them to define the seasonal length. Seasonal rainfall refers to the daily rainfall that accumulates from the day of onset date to the demise date.

2.2.2 Perturbation

The motivation to perturb the timeseries is to account for the uncertainty of random synoptic or mesoscale events that is potentially unrelated to the seasonal cycle, which could affect the diagnosis of the onset/demise date of the rainy season. The threat of false diagnosis of onset/demise dates of the rainy season is acute from the proposed methodology since this diagnosis is computed at the granularity of the rainfall analysis. Therefore, generating an ensemble of onset/demise dates of the rainy season that accounts for this uncertainty is essential. The perturbations are generated randomly by replacing the rainfall of each day in the original timeseries for a given grid point by rainfall in the range of $\pm$ 3 days. The range of $\pm$ 3 days used in the generation of the perturbed timeseries covers the uncertainty in the occurrence of the synoptic (1 to 7 days) to meso-scale (1 to 3 days) rain features of the rainy season. In this manner, an ensemble of 101 (100 perturbations + 1 original) timeseries are generated. The diagnosis of the onset/demise date of the rainy season will converge or diverge amongst the ensemble members if it is found to be insensitive or sensitive to these random rain events, respectively.

2.2.3 Signal-to-noise Ratio

We estimated the signal-to-noise ratio of the four quantities by utilizing the 101 ensemble members.

$$S_{noise}^2 = \frac{1}{A(B-1)}\sum_{a=1}^{A}\sum_{b=1}^{B}(X_{ab} - \bar{X}_b)^2 \tag{2}$$

$$S_{signal}^2 = S_{em}^2 - \frac{1}{b}S_{noise}^2 \tag{3}$$

Where, X belongs to any of the four quantities such as onset date, demise date, seasonal length, and seasonal rainfall anomalies, a and b are indices for A years and B ensemble members, respectively.

$$S_{em}^2 = \frac{1}{(A-1)}\sum_{a=1}^{A}\left(\bar{X}_a - \bar{\bar{X}}\right)^2,$$

And $\bar{X}_a = \frac{1}{(B-1)}\sum_{b=1}^{B} X_{ab}$, $\bar{\bar{X}} = \frac{1}{(A-1)}\sum_{a=1}^{A} \bar{X}_a$,

The signal-to-noise ratio is then given by:

$$ratio = \frac{S_{signal}^2}{S_{noise}^2}$$

When $ratio < 1$ then it indicates that noise (or chaotic variations) is dominant, and the signal dominates when $ratio > 1$. A strong signal indicates that perturbing the rainfall time series will not substantially influence the diagnostics of the onset date, demise date, seasonal length, and seasonal rainfall anomalies.

2.2.4 Significance test

In this study, we first estimate the two-tailed p-values of the correlation coefficients using the t-statistic. Further, following Benjamini and Hochberg (1995), we conduct a test for false rejection on all temporal correlations to assess its robustness, which is necessary because of the large number of simultaneous statistical significance tests conducted across all grid points of the domain. This process involves adjusting the p-value (Benjamini and Hochberg 1995) to control the false rejection of the null hypothesis, also known as the false discovery rate. By controlling the false discovery rate, we enhance the confidence in the significance of the findings.

3. Results

The seasonal mean rainfall over India observed from IMERG is found to be comparable with IMD observation and superior to other satellite products over the region (Saikrishna et al., 2021). This is further confirmed in Figs. 1a-c which compares the 23-year annual mean rainfall climatology between IMERG observations and the corresponding IMD rainfall dataset. It may be noted that in Equation (1), the annual mean rainfall climatology ($\bar{R}$) at each grid point is used to determine onset dates and demise dates of the rainy season over India (e.g., Misra et al., 2018). Overall, the

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

annual mean rainfall climatology from IMERG compares well with the IMD rain gauge dataset. However, the overestimation of rainfall in the Indo-Gangetic plains and parts of northeast India and underestimation of rainfall over the northern Western Ghats, northern sections of northeast India, and the bulk of Jammu and Kashmir by IMERG relative to IMD dataset is apparent (Fig. 1). Bushair et al., (2019) noted that IMERG underestimates the rainfall over high-altitude regions compared to IMD observations, which are rain gauge based. Further, Fig. 1d shows the cumulative rainfall anomaly curve generated from the 23-year (2001 - 2023) daily rainfall climatology area averaged over all of India for IMERG (blue) and IMD (red) observations. The onset (inflection at the nadir of the cumulative anomaly curve) and demise (inflection at the zenith of the cumulative anomaly curve) dates estimated from IMERG and IMD closely match each other, with the onset dates being 31st May and 3rd June, and the demise dates being 7th October and 8th October, respectively.

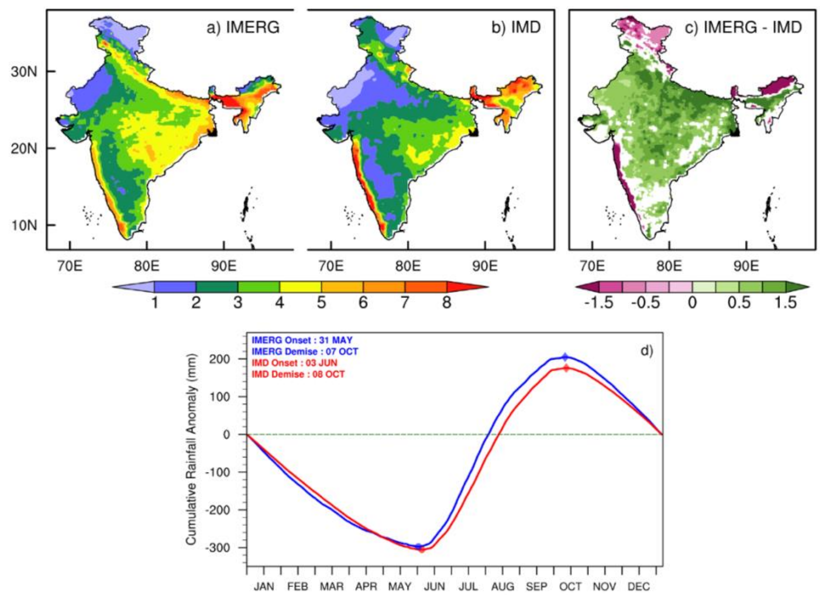

Figure 1: *Spatial pattern of the 23-year climatology of the mean annual rainfall during the period 2001-2023 from a) IMERG observation, b) IMD observation, and c) the difference between the IMERG and IMD (only differences significant at 5% percentile level on t-test is shaded). d) The cumulative rainfall anomaly curve generated from the 23-year daily rainfall climatology of*

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

IMERG (blue) and IMD (red) observations. The onset dates (filled circles) and demise dates (filled diamonds) obtained from both datasets are marked.

Figures 2a-d depict the 23-year local climatological onset date, demise date, seasonal length, and seasonal accumulated rainfall from IMERG, with the associated standard deviation displayed in Figs. 2e-h. The spatial distribution of the climatology of the onset dates (Fig. 2a) shows the earliest onset occurs over northeast regions of India followed by southern Kerala, and then it gradually advances to the other parts of the country and looks like the typical isochrone evolution of the ISM (Ramage 1971; Rao 1976; Janowiak and Xie 2003). The delayed onset occurs over the east coast of Tamil Nadu followed by Jammu and Kashmir. The onset dates over Kerala and adjacent regions mostly start from early May and do not coincide with the MoK, and this is because this methodology probably detects the onset dates early due to strong and continuous spells of pre-monsoon rainfall. Similarly, the earliest withdrawal (Fig. 2b) of the ISM occurs over northwestern India such as Rajasthan, Punjab, Haryana, and Himachal Pradesh. The demise date is delayed over southern peninsular India, with the most delayed demise occurring over the southeast coast of Tamil Nadu. The seasonal length (Fig. 2c) is shorter in the west and northwest regions of India, with the shortest season in Jammu and Kashmir. The longer season over the peninsular India, as well as the coastal regions of Odisha and West Bengal, is linked to the season's earlier start date. The spatial distribution of the seasonal accumulated rainfall (Fig. 2d) closely follows the typical seasonal mean ISM distribution. The standard deviation patterns of the onset dates show central parts of India have less variability than other regions, such as southern peninsular India, northeast, and northwest regions. Similar patterns of variability are observed in the demise dates and seasonal length with the largest variations found over southern peninsular India and northwest India and least over central India. On the other hand, the seasonal rainfall exhibits a significant range of variability across most of India, whereas comparatively less variability is observed over Rajasthan and Jammu and Kashmir.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

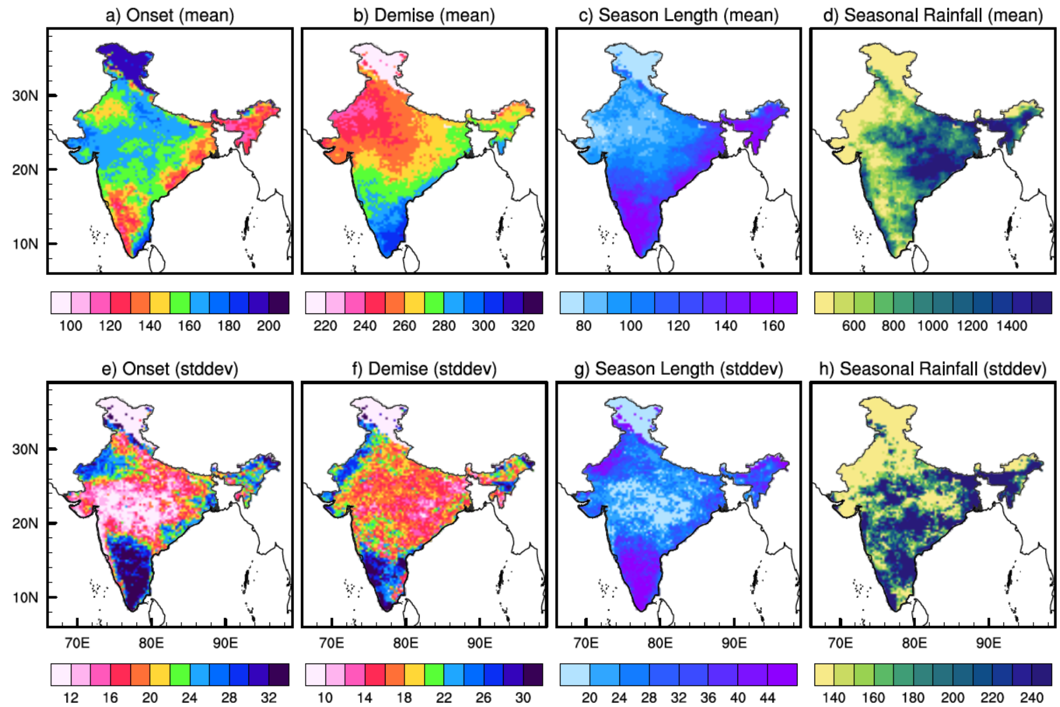

***Figure 2**: The climatological a) onset date (Julian day), b) demise date (Julian day), c) seasonal length, and d) seasonal accumulated rainfall (mm) of the rainy season from IMERG. The corresponding standard deviation of e) onset date (days), b) demise date (days), c) seasonal length (days), and d) seasonal accumulated rainfall (mm).*

The IMERG dataset, with its 23-year record, was compared to the IMD and CPC datasets spanning 1979–2023 (limited to this period in order for CPC and IMD to have a common period) to evaluate onset and demise dates, seasonal length, and seasonal accumulated rainfall of the rainy season over India (Fig. S2). The motivation to carry out this comparison is to establish that a 23-year record of IMERG is comparable to a longer record of dataset available from the other two sources besides examining its fidelity.

It is apparent from Fig. S2 that the spatial patterns of the climatological onset and demise dates, seasonal length and rainfall in IMERG is comparable to corresponding climatologies from IMD and CPC datasets. There is however a tendency for IMERG to have a bias of a slightly earlier onset of the rainy season compared to IMD and CPC, particularly over southern peninsular India,

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

including Kerala and Tamil Nadu (Fig. S2a-c). Similar inference is observed in Rajasthan, costal Odisha and West Bengal, and few isolated grid points over other regions. In contrast, over Jammu and Kashmir, IMERG exhibits a noticeably delayed onset, which makes us unsure about the results over this area. Similarly, IMERG display a systematic bias of slightly later demise dates along the Western Ghats, but overall, the rainy season demise dates are consistent across datasets (Fig. S2d-f). As a result, IMERG's rainy season seems to last longer than the other two datasets in some of these locations (Fig. S2g-i). Again, in terms of seasonal accumulated rainfall (Fig. S2j-l), IMERG overestimates precipitation compared to IMD and CPC in some of these locations. However, rainfall over the Western Ghats exhibits stronger gradients and is underestimated in IMERG while it exhibits more rainfall over western central India and Tamil Nadu than the other two datasets.

The interannual variability of these variables in IMERG also appear to be comparable to the other two datasets, especially in their spatial gradients (Fig. S3). There are however instances of some differences in the datasets. For example, over central India, the standard deviation of seasonal rainfall is significantly overestimated in IMERG compared to IMD (Fig. S3j-l), even though the demise dates (Fig. S3d-f) and seasonal length (Fig. S3g-i) exhibit less variability in IMERG. The discrepancies observed in IMERG can likely be attributed to the choice of the 12-hour latency product, which is an early version with limited calibration against ground-based observations, its shorter temporal coverage of 23 years, compared to the 45 years of data used for IMD and CPC. Despite these limitations, which seem to be tolerable for the applications of this study (compare Figs. 1, S2, and S3) we are inclined to used IMERG because of the potential for real-time applications.

The primary motivation to monitor the local onset and demise of the ISM is its significant interannual variations and spatial variability. Both the variation in seasonal length and seasonal rainfall are crucial in determining the nature of the ISM (Xavier et al. 2007; Sperber and Annamalai 2014). The correlation of the onset date with the seasonal length shows a significant negative correlation across India except in some parts of central India and Jammu and Kashmir (Fig. 3a). This negative correlation suggests that an early or later onset date is likely to be associated with a longer or a shorter rainy season, respectively. Only the correlations significant at a 5% level are shaded in Fig. 3. Similarly, early onset is associated with a wetter season, while

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

a delayed onset is linked with a drier season (Fig. 3b). However, the grid points showing significant correlation are fewer in Fig. 3b compared to Fig. 3a. The correlations of the demise date with seasonal length (Fig. 3c) and seasonal rainfall anomalies (Fig. 3d) show that later or early demise of the wet season corresponds to a longer and wetter season, whereas an earlier demise corresponds to shorter and drier seasons. However, the demise date may not be a useful predictor for the rainy season because it occurs when the rainy season ends. Figure 3e shows that onset and demise dates are largely uncorrelated, implying that their variations are independent of each other. These outcomes imply that the onset date alone can be a useful predictor for the outlook of the seasonal length and the seasonal rainfall anomaly of the forthcoming season.

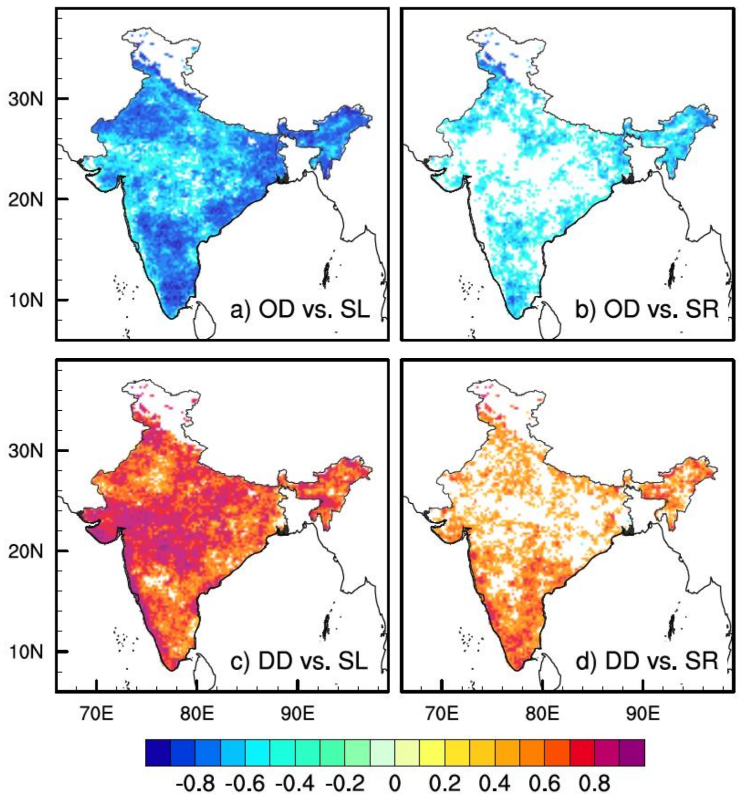

Figure 3*: The correlation coefficients of the onset date (OD) with a) seasonal length (SL), and b) seasonal rain (SR). Similarly, the correlations of DD with c) SL and d) SR. The shading indicates statistical significance at a 5% significance level according to the t-statistic, following false discovery rate testing as described by Benjamini and Hochberg (1995).*

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

We estimated the signal-to-noise ratio on the four variables of onset and demise dates, seasonal length, and rainfall by utilizing the 101-member ensemble of time-series data. This analysis examines the spread across ensemble members and informs on the uncertainty of the diagnosis to random rain events. The signal-to-noise ratio for onset dates (Fig. 4a) shows that noise dominates in many places, particularly in peninsular India and the eastern half of central India. These low signal-to-noise ratio regions suggests that onset dates of the rainy season are not very strongly tied to the seasonal cycle and predicting the onset dates in these places becomes challenging and less reliable. In contrast, the signal-to-noise ratios for demise dates (Fig. 4b) are typically more than one over most grid points. This indicates greater certainty for demise dates, particularly in areas like Kerala, coastal Karnataka, and Gujarat, where the ratios are comparatively higher than one. In the case of the seasonal length, signal dominates over noise with a signal-to-noise ratio above 1 over most areas (Fig. 4c). Compared to the other variables, seasonal rainfall generally shows much higher signal-to-noise ratios (> 1) across most of the grid points. This result is obvious given the fact that seasonal rainfall is an aggregate of the daily rainfall over the entire rainy season, while the onset and demise dates are single days of the season when they are diagnosed. Therefore, there is a tendency for the cancellation of the noise in the aggregation of the daily rainfall to seasonal rainfall, which yields a higher signal-to-noise ratio.

© Author(s) 2025. CC BY 4.0 License.

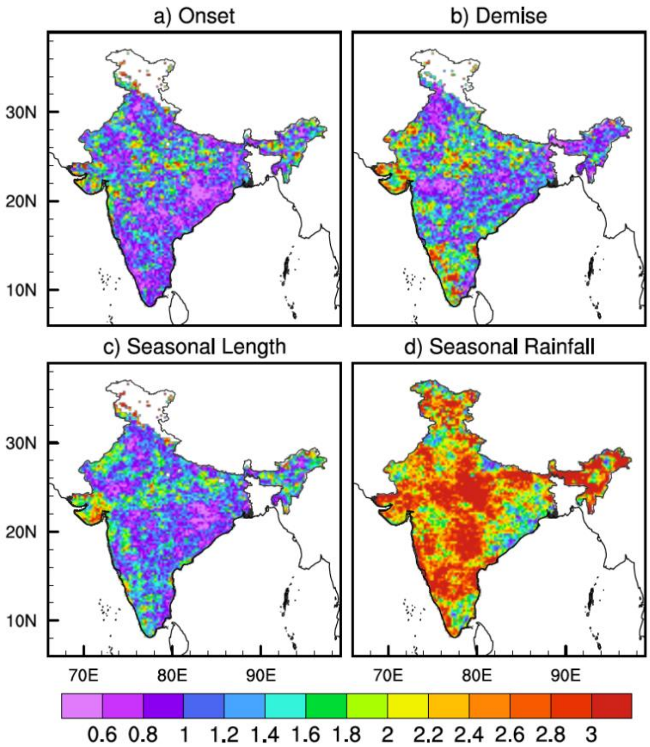

Figure 4: *The 23-year climatology of the signal-to-noise ratio of a) onset date, b) demise date, c) seasonal length, and d) seasonal rainfall estimated based on the 101 ensemble members.*

Further, we examined the probabilistic skill of the seasonal outlook by using the Area Under the relative operating characteristic (ROC) Curve (AUC) method. The AUC method is widely used to assess the skill of seasonal climate predictions (Mason and Graham 2002; Misra 2004; Narotsky and Misra 2021). Previous studies have shown that this method is useful for evaluating the probabilistic skill of predicting anomalous seasons based on anomalous onset dates of the rainy season (e.g. Rodgers et al 2024). We categorized the onset date, seasonal length, and seasonal rainfall into terciles (23-years divided into three groups): the lower tercile represents an early onset, shorter season, and drier conditions; the upper tercile signifies a delayed onset, longer season, and wetter conditions; and the middle tercile indicates normal onset, length, and seasonal rainfall. We then created a contingency table (Table S1) to assess the probability of categorical forecasts. This table evaluates how often an early or late onset is linked with a shorter or longer season and with

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

drier or wetter conditions, respectively. In addition, we also consider the effects of normal onset on normal length and seasonal rainfall.

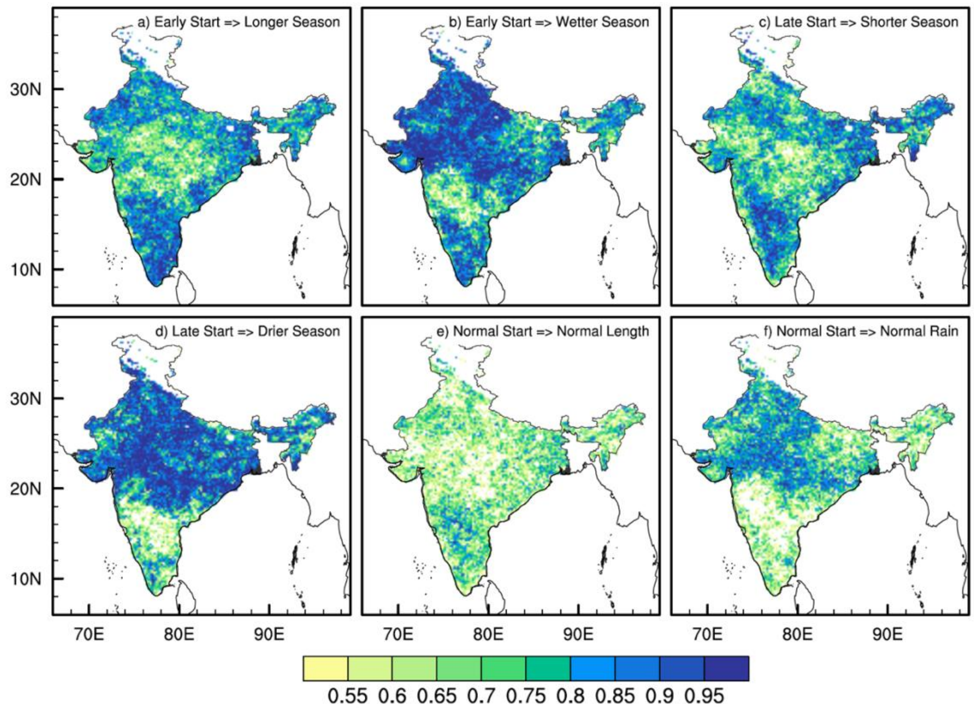

***Figure 5**: The probabilistic skill score as measured by the Area Under the relative operating characteristic Curve (AUC) for early start (lowest tercile) season associated with a) longer (highest tercile) and b) wetter (highest tercile) season, late start (highest tercile) season associated with c) shorter (lowest tercile) and d) drier (lowest tercile) season, and normal (middle tercile) start season associated with e) normal (middle tercile) seasonal length and f) normal (middle tercile) seasonal rain. The AUC above 0.5 is shaded.*

The probabilistic skill scores derived from the AUC method are shown in Fig. 5. Here, only grid points with AUC values ≥ 0.5 are shaded, as these points indicate skillful seasonal outlooks that outperform random predictions (Mason and Graham, 2002). The seasons with an early start and longer season (Fig. 5a) show the most skill (>0.9) over peninsular India, northwest India, and some parts of northeast India. Similarly, the early start of the season, along with the wetter season (Fig. 5b), results in high skill levels across most parts of India except the majority of Maharashtra and

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

north interior Karnataka. The late start and associated shorter seasons shown in Fig. 5c exhibit the highest skill score over some parts of peninsular India. The late start with the drier season also resulted in high skill scores across most of the areas except Kerala, Tamil Nadu, Karnataka and Andhra Pradesh. The anomalous seasons (Fig. 5a-d) demonstrate higher skill levels across India compared to the normal seasons (Fig. 5e-f). This is due to the leveraging of the linear relationships of the onset date variations with the rainy season variations (Figs. 3a and b). In contrast, seasons with a normal start of the rainy season exhibit lower AUC skill scores (Figs. 5e and f) across most grid points relative to the anomalous start of the rainy seasons (Figs. 5a-d). However, the skill scores for a normal start with normal rainfall (Fig. 5f) are slightly better than those for a normal start with normal length (Fig. 5e). In summary, the behavior of the forthcoming season is more predictable if the onset date is early or delayed than the normal. It is easy to adopt this methodology for real-time applications. The evolution of the daily cumulative anomaly curve of rainfall could be monitored in real-time to find the minima in the curve as the onset date. However, to avoid misdiagnosing the onset date of the rainy season, one could wait for a period of time after this diagnosis (typically a week) to confirm that the minima were indeed reached to declare the onset date. Once the onset date is diagnosed then one could use the linear relationships shown in Fig. 3 to develop a seasonal outlook for the rainy season. This is done routinely in Florida (Misra et al. 2022). With the availability of the merged IMERG estimates with the rain-gauge based IMD rainfall following Mitra et al., (2009) our proposed methodology could be adopted for real-time applications over India.

Further, we investigated the interannual variability of the onset dates over India and its association with ENSO (Figure S4) and the Indian Ocean Dipole (IOD) variations (Figure S5). In comparison to Figure 3, Figures S4 and S5 failed the BH test suggesting the lack of robustness in the teleconnections of the rainy season variations with either ENSO or IOD. In light of these results, the significance of the reliability of the seasonal outlook shown in Figure 5 assumes greater significance. Furthermore, the monitoring of the observations of the evolution of the rainy season to diagnose the onset dates seems an attractive approach with the relatively lower signal-to-noise ratio of the onset date of the rainy season shown in Figure 4a relative to the seasonal length in Figure 4c or seasonal rainfall in Figure 4d. We are then able to leverage a low signal-to-noise ratio quantity like the onset date of the rainy season to provide a seasonal outlook of relatively higher signal-to-noise ratio quantities of seasonal length and rainfall.

4. Summary and Concluding Remarks

The seasonal prediction of the Indian rainy season is a considerable challenge given its complex spatio-temporal variations. In this study, we offer a simple and reliable technique for seasonal outlook of the rainy season, which is viable for real-time applications as well. Using precipitation estimates of IMERG version 6 over India at 10km grid, we generate an ensemble of 101 members from randomly perturbing the series to assess the uncertainty of the diagnosed onset/demise dates of the rainy season to random synoptic-meso scale rain events unconnected to the seasonal cycle.

We verified the IMERG rainfall observations against gauge-based IMD observations. The mean annual rainfall climatology of IMERG closely matches that of the IMD. However, IMERG tends to underestimate mean annual rainfall in high-altitude regions and overestimate it in the Indo-Gangetic plains. Further, we examined the relationships between the onset and demise dates with seasonal length and rainfall and discovered that variations in onset and demise dates of the rainy season across India have a significant impact on the seasonal length and the seasonal rainfall variations of the rainy season. It is found that an earlier onset date of the rainy season is strongly related to a longer and wetter season, whereas a later onset date corresponds to a shorter and drier season. However, the relationship between the onset, demise, seasonal length, and rainfall with large-scale climate drivers such as ENSO and IOD is comparatively weaker in major parts of India. This study shows that by estimating the onset date of the rainy season alone we can effectively provide a reliable seasonal outlook for both the seasonal length and total rainfall of the upcoming season by exploiting the existing local relationships. The probabilistic skill scores presented in this study also demonstrate that this method has a high potential for providing seasonal outlooks for the forthcoming season. These seasonal outlooks have numerous potential applications, and many local communities could greatly benefit from them. The proposed methodology for the seasonal outlook of the rainy season over India could be easily adapted for real-time applications with the current availability of IMERG rainfall products in real-time.

Acknowledgments

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

We acknowledge the support from NASA grant 80NSSC22K0595.

Code/Data Availability

The IMERG rainfall dataset is available from the NASA GES DISC (https://gpm1.gesdisc.eosdis.nasa.gov/data/GPM_L3/GPM_3IMERGDL.06/). The India Meteorological Department rainfall data set used for the analysis in this study is available at https://www.imdpune.gov.in/cmpg/Griddata/Rainfall_25_NetCDF.html. The daily CPC data is obtained from https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html. The Python and NCL codes used for the analysis can be provided by the corresponding authors upon request.

Author contribution

The study was conceptualized by CBJ and VM. CBJ carried out the analysis and validation and prepared the original manuscript. VM supervised the work, acquired funding, and reviewed and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Reference:

1. Ananthakrishnan, R., and M. K. Soman, 1988: The onset of the southwest monsoon over Kerala: 1901–1980. J. Climatol., **8**, 283–296, doi:10.1002/joc.3370080305.
2. Bansod, S. D., S. V. Singh, and R. H. Kripalani, 1991: The relationship of monsoon onset with subsequent rainfall over India. *Int. J. Climatol.*, **11**, 809–817, doi:10.1002/joc.3370110707.
3. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
4. Bhardwaj, A., & Misra, V. (2019). Monitoring the Indian summer monsoon evolution at the granularity of the Indian meteorological sub-divisions using remotely sensed rainfall products. Remote Sensing, 11(9), 1080.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

5. Bushair, M. T., Kumar, P., & Gairola, R. M. (2019). Evaluation and assimilation of various satellite-derived rainfall products over India. International Journal of Remote Sensing, 1–24.
6. Choudhury, D., Nath, D., & Chen, W. (2021). The modulation of Indian summer monsoon onset processes during ENSO through equatorward migration of the subtropical jet stream. Climate Dynamics, 57(1), 141-152.
7. Dhar, O. N., P. R. Rakhecha, and B. N. Mandal (1980), Does the early or late onset of monsoon provide any clue to subsequent rainfall during the monsoon season, Mon. Weather Rev., 108, 1069–1072, doi:10.1175/1520-0493(1980)1082.0.CO;2
8. Fasullo J, Webster PJ (2003) A hydrological definition of Indian monsoon onset and withdrawal. J Climate 16:3200–3211
9. Gadgil, S.; Gadgil, S. The Indian Monsoon, GDP and Agriculture. Econ. Polit. Wkly. 2006, 41, 4887–4895
10. Huffman, G. J., Adler, R. F., Bolvin, D. T., Hsu, K., Kidd, C., Nelkin, E. J., et al. (2019). Algorithm Theoretical Basis Document (ATBD) for Global Precipitation Climatology Project Version 3.0. Precipitation Data. Greenbelt, MD: MEaSUREs project.
11. Janowiak, J. E., and P. Xie (2003), A global-scale examination of monsoon-related precipitation, J. Clim., 16, 4121–4133.
12. Joseph PV, Eishcheid JK, Pyle RJ. 1994. Interannual variability of the onset of the Indian summer monsoon and its association with atmospheric features, El Nino, and sea surface temperature anomalies. Journal of Climate 7: 81–105.
13. Joseph PV, Sooraj KP, Rajan CK (2006) The summer monsoon onset process over South Asia and an objective method for the date of monsoon onset over Kerala. Int J Climatol 26:1871–1893
14. Joseph, P.V.; Eischeid, J.K.; Pyle, R.J. Interannual Variability of the Onset of the Indian Summer Monsoon and Its Association with Atmospheric Features, El Niño, and Sea Surface Temperature Anomalies. J. Clim. 1994, 7, 81–105.
15. Khouakhi, A., Villarini, G., & Vecchi, G. A. (2017). Contribution of tropical cyclones to rainfall at the global scale. Journal of Climate, 30(1), 359-372.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

16. Kung, E. C., and Sharif, T. A., 1980, Regression forecasting of the onset of the India summer monsoon with antecedent upper air conditions, J. Appl. Meteorol., V19, pp370—380.
17. Lal, M. (2000). Climatic change-implications for India's water resources. Journal of Social and Economic Development, 3, 57-87.
18. Liebmann, B., and J. Marengo, 2001: Interannual variability of the rainy season and rainfall in the Brazilian Amazon basin. J. Climate, 14, 4308–4318, https://doi.org/10.1175/1520-457 0442(2001) 014,4308:IVOTRS.2.0.CO;2.
19. Mason, S. J. and N. E. Graham, (2002): Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. Quar. Roy. Soc., 128, 2145-2166.
20. Misra, V. (2004) An evaluation of the predictability of austral summer season precipitation over South America. Journal of Climate, **17**, 1161–1175.
21. Misra, V., Dixit, S., & Jayasankar, C. B. (2023). The regional diagnosis of onset and demise of the rainy season over tropical and subtropical Australia. Earth Interactions, 27(1), 220026.
22. Misra, V., Jayasankar, C. B., Beasley, P., & Bhardwaj, A. (2022). Operational monitoring of the evolution of rainy season over florida. Frontiers in Climate, 4, 793959.
23. Misra, V.; Bhardwaj, A.; Mishra, A. (2018) Local onset and demise of the Indian summer monsoon. Clim. Dyn., 51, 1609–1622.
24. Misra, V.; Bhardwaj, A.; Noska, R. (2017) Understanding the Variations of the Length and the Seasonal Rainfall Anomalies of the Indian Summer Monsoon. J. Clim., 30, 1753–1763.
25. Misra V, DiNapoli S (2014) The variability of the southeast Asian summer monsoon. Int J Climatol 34:893–901. https://doi.org/10.1002/joc.3735
26. Mitra, A. K., Bohra, A. K., Rajeevan, M. N., & Krishnamurti, T. N. (2009). Daily Indian precipitation analysis formed from a merge of rain-gauge data with the TRMM TMPA satellite-derived rainfall estimates. 87A, 265-279.
27. Mooley, D. A., and B. Parthasarthy (1984), Fluctuations in all-India summer monsoon rainfall during 1871–1978, Clim. Change, 6(3), 287–301, doi:10.1007/BF00142477.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

28. Mooley, D. A., and J. Shukla (1987), Variability and forecasting of the summer monsoon rainfall over India, in Monsoon Meteorology, edited by C. P. Chang and T. N. Krishnamurti, pp. 26–59, Oxford Univ. Press, Oxford, U. K
29. Moron V, Robertson AW, (2014): Interannual variability of Indian summer monsoon rainfall onset date at local scale. Int J Climatol, 34(4). doi:10.1002/joc.3745
30. Noska, R.; Misra, V. (2016) Characterizing the onset and demise of the Indian summer monsoon. Geophys. Res. Lett., 43, 4547–4554.
31. Pai DS, Rajeevan M (2007) Indian Summer Monsoon Onset: variability and prediction. NCC Research Report No. 6. National Climate Centre, Indian Meteorological Department.
32. Pai, D. S., & Rajeevan, M. N. (2009). Summer monsoon onset over Kerala: New definition and prediction. Journal of Earth System Science, 118(2), 123-135.
33. Pai, D. S., Rajeevan, M., Sreejith, O. P., Mukhopadhyay, B., & Satbha, N. S. (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
34. Pai, D. S., Bandgar Arti, Devi Sunitha, Musale Madhuri, M. R. Badwaik, A. P. Kundale, Gadgil Sulochana, M. Mohapatra, and M. Rajeevan. "Normal dates of onset/progress and withdrawal of southwest monsoon over India." Mausam 71, no. 4 (2020): 553-570.
35. Phadtare, J. A., Fletcher, J. K., Ross, A. N., Turner, A. G., Schiemann, R. K., & Burns, H. L. (2023). Unravelling the mechanism of summer monsoon rainfall modes over the west coast of India using model simulations. Quarterly Journal of the Royal Meteorological Society, 149(757), 3164-3182.
36. Pradhan, M., Rao, A. S., Srivastava, A., Dakate, A., Salunke, K., & Shameera, K. S. (2017). Prediction of Indian summer-monsoon onset variability: A season in advance. Scientific reports, 7(1), 14229.
37. Prasad, V. S. & Hayashi, T. Onset and withdrawal of Indian summer monsoon. Geophys. Res. Lett. **32**, 1–5 (2005).
38. Preenu, P. N., Joseph, P. V., & Dineshkumar, P. K. (2017). Variability of the date of monsoon onset over Kerala (India) of the period 1870–2014 and its relation to sea surface temperature. Journal of Earth System Science, 126, 1-19.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

39. Rajeevan, M. and Dubey, D. P., 1995, Long range prediction of monsoon onset over Kerala, Mausam, V46, pp287-290.
40. Ramage, C. (1971), Monsoon Meteorology, Int. Geophys. Ser., vol. 15, 296 pp., Academic Press, San Diego, Calif.
41. Ramesh Kumar, M. R., 2004, Forecasting of onset of southwest monsoon over Kerala coast using satellite data, IEEE Geosci. Remote Sens. Lett., V1 (4), pp265-267
42. Rao, Y. P. (1976) Southwest monsoon. Meteorological Monograph, Synoptic Meteorology, 367.
43. Rodgers, J., Misra, V., & Jayasankar, C. B. (2024). Using the observed variations of the start date of the rainy season over Central America for its reliable seasonal outlook. Journal of Climate. DOI: https://doi.org/10.1175/JCLI-D-23-0699.1
44. Rodgers, J., Misra, V., & Jayasankar, C. B. (2024). Using the observed variations of the start date of the rainy season over Central America for its reliable seasonal outlook. Journal of Climate.
45. Saikrishna, T. S., Ramu, D. A., & Osuri, K. K. (2021). Inter-comparison of high-resolution satellite precipitation products over India during the summer monsoon season. Meteorology and Atmospheric Physics, 133, 1675-1690.
46. Sankar, S., Kumar, M. R., & Reason, C. (2011). On the relative roles of El Nino and Indian Ocean Dipole events on the Monsoon Onset over Kerala. Theoretical and applied climatology, 103, 359-374
47. Sperber, K. R., and H. Annamalai, 2014: The use of fractional accumulated precipitation for the evaluation of the annual cycle of monsoons. Climate Dyn., 43, 3219–3244, doi:10.1007/s00382-014-2099-3.
48. Stolbova, V., Surovyatkina, E., Bookhagen, B., & Kurths, J. (2016). Tipping elements of the Indian monsoon: Prediction of onset and withdrawal. Geophysical Research Letters, 43(8), 3982-3990.
49. Thakur, M. K., Kumar, T. L., Narayanan, M. S., Kundeti, K. R., & Barbosa, H. (2020). Analytical study of the performance of the IMERG over the Indian landmass. Meteorological Applications, 27(3), e1908.

https://doi.org/10.5194/egusphere-2025-273
Preprint. Discussion started: 18 February 2025
© Author(s) 2025. CC BY 4.0 License.

50. Walker, J. M., & Bordoni, S. (2016). Onset and withdrawal of the large-scale South Asian monsoon: A dynamical definition using change point detection. Geophysical Research Letters, 43(22), 11-815.
51. Wang B, Ding Q, Joseph PV (2009) Objective definition of the Indian summer monsoon onset. J Clim 22(12):3303–3316.
52. Wang B, LinHo (2002) Rainy season of the Asia-Pacific summer monsoon. J Climate 15:386–398
53. Wang, B., Ding, Q. & Joseph, P. V. Objective definition of the Indian summer monsoon onset. J. Clim. 22, 3303–3316 (2009)
54. Wang, X., Jiang, X., Yang, S., & Li, Y. (2013). Different impacts of the two types of El Niño on Asian summer monsoon onset. Environmental Research Letters, 8(4), 575–591.
55. Xavier, P. K., C. Marzin, and B. N. Goswami, 2007: An objective definition of the Indian summer monsoon season and a new perspective on the ENSO–monsoon relationship. Quart. J. Roy. Meteor. Soc., 133, 749–764, doi:10.1002/qj.45.
56. Xie, P., A. Yatagai, M. Chen, T. Hayasaka, Y. Fukushima, C. Liu, and S. Yang (2007), A gauge-based analysis of daily precipitation over East Asia, J. Hydrometeorol., 8, 607. 626.
57. Zeng X, Lu E (2004) Globally unified monsoon onset and retreat indexes. J Climate 17:2241–2248