atmosphere

MDPI

Article

Comparative Evaluation of Short-Range Extreme Rainfall Forecast by Two High-Resolution Global Models

Tanmoy Goswami [1,2,*], Seshagiri Rao Kolusu [3], Subharthi Chowdhuri [4], Malay Ganai [1] and Medha Deshpande [1]

[1] Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pashan, Pune 411008, India; malay.cat@tropmet.res.in (M.G.); medha_d@tropmet.res.in (M.D.)
[2] Savitribai Phule Pune University, Ganeshkhind Rd, Pune 411007, India
[3] Met Office, FitzRoy Road, Exeter EX1 3PB, UK; seshagirirao.kolusu@metoffice.gov.uk
[4] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland; subharthi.chowdhuri@luke.fi
* Correspondence: tanmoy.cat@tropmet.res.in

Abstract

Accurate prediction of extreme rainfall events during the Indian Summer Monsoon (ISM, June to September) is critical for disaster preparedness and mitigation. This study evaluates the performance of two operational numerical weather prediction models, a high-resolution version of Global Forecast System (GFS T1534) and the control member of the Met Office Global and Regional Ensemble Prediction System-Global (MOGREPS-G), in forecasting such events during the ISM from 2020 to 2023. The results demonstrate that, with respect to observations, both models tend to underestimate the mean and variability of rainfall; GFS-T1534 represents the mean and correlation better while MOGREPS-G represents the variability better over the Indian landmass. To assess the models' performance for extreme rainfall prediction, we fix a rainfall threshold of 50 mm day^{-1}, and the skill scores are computed including Probability of Detection, False Alarm Rate, Bias score and F1 score. Together, these scores indicate that both models show potential in short-range forecasting of extreme rainfall events, particularly within 24 h, but their skills remain limited at longer lead times. Specifically, the model biases vary over different geographical locations, often showing contrasting features. This underscores the need for model-specific post-processing and calibration techniques if these forecasts are to be used effectively for operational decision-making.

Keywords: forecast skill; extreme rainfall; global forecast system

Academic Editor: Tin Lukić

Received: 3 January 2026
Revised: 5 March 2026
Accepted: 6 March 2026
Published: 17 March 2026

1. Introduction

The rise of extreme rainfall in different parts of the globe is well documented and reported by many previous studies [1–5], which primarily attribute this as a consequence of global warming. The wet regions in the tropics are getting wetter and dry regions are getting drier [6]. The extreme events also have various social, economic and environmental impacts [7]. Additionally, extreme events demand preparedness for emergency response in different sectors including the health [8]. A recent study reported that extreme rainfall reduces worldwide macroeconomic growth rates, which, in turn, slows down the global economy [9]. Considering this rising trend of extreme rainfall events, India is not an exception. Many previous studies have reported a significant rise in extreme rainfall events over Indian landmass [10–13]. The number of wet days with rainfall over 99th percentile is

increasing in the Indian region [14]. Moreover, there is a significant rise in the intensity of rainfall over Indian landmass [15]. Another recent study highlighted that the frequency of occurrence of widespread extreme rainfall ($\geq$150 mm day^{-1}) rose to three-fold during the year 1950–2015 over the Central Indian region [16].

Keeping the socio-economic impact in mind, the accurate prediction of these events is essential, especially over the Indian landmass, where the extreme events pose constant threats to the population and infrastructure. Past studies have highlighted that extreme rainfall is generally underestimated by current-generation climate models. Specifically, they tend to underestimate both frequencies and intensities of extreme rainfall events [17,18]. To meet the operational requirements of the Indian Meteorological Department (IMD), a high-resolution Global Forecast System (GFS T1534) for short- and medium-range prediction is being implemented by the Indian Institute of Tropical Meteorology. This model shows reasonable skills to capture the broader structure of the rainfall probability distribution functions [19]. On the other hand, the UK Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) shows good skill towards predicting extreme rainfall events over the Indian region [20]. Furthermore, it is reported that the MOGREPS-G-based system NEPS (National Centre for Medium-Range Weather Forecasting (NCMRWF) ensemble prediction system) shows better skill across different rainfall thresholds when compared with NGEFS (NCMRWF's Global Ensemble Prediction System) [21]. Despite such advances, direct comparisons between MOGREPS-G and GFS T1534 in forecasting extreme precipitation events during the Indian Summer Monsoon (ISM) remain limited. In particular, their relative performance in capturing the spatial distribution, intensity, and frequency of extreme events across India's heterogeneous agro-climatic zones has not been systematically assessed.

Therefore, a critical research gap exists in evaluating the short-range (up to 72 h lead time) prediction skill of these high-resolution global models during the ISM, especially over the regions strongly influenced by orographic lifting and convective dynamics, such as the Western Ghats and the Gangetic Plains. To address these gaps, the present study evaluates and contrasts the forecast skill of the control member of MOGREPS-G and GFS T1534 in predicting extreme precipitation events during the ISM form 2020 to 2023. The central research question guiding this study is how the high-resolution global models, namely the MOGREPS-G control-only configuration and the deterministic GFS T1534, compare in their ability to forecast the spatial distribution, intensity, and frequency of extreme precipitation events during the Indian Summer Monsoon, particularly over regions characterized by complex topography and strong convective activity.

This evaluation is important to emphasize the strengths and limitations of each model in forecasting extreme rainfall events, thereby contributing to improve weather prediction and risk management. This article is organized in three different sections. In Section 2, we describe the model setups of GFS T1534 and MOGREPS-G, the observational dataset being used for the validation purposes, and the methodologies being adopted to evaluate the model skill scores. In Section 3, the results for the Indian landmass from model inter-comparisons are described, while Section 4 presents the conclusions and future outlook.

2. Model, Data, and Methodology

2.1. Model Description

The Global Forecast System (GFS) is a spectral model with 1534 triangular truncation and a horizontal resolution of around 12.5 km with 64 hybrid vertical level (top level pressure 0.27 hPa). The model is adopted from the NCEP and its current version is 14. It has a two-time-level, semi-implicit, semi-Lagrangian discretization approach-based dynamical core, while the physics are computed over a linear reduced Gaussian grid [22]. It uses the

revised simplified Arakawa–Schubert (RSAS) convective parameterization scheme and mass-flux-based SAS shallow convection scheme [23]. For microphysics, it uses Zhao–Carr microphysics parameterization scheme [24]. It has orographic gravity wave drag, mountain-drag and stationary convective gravity wave drag. For the Planetary Boundary Layer (PBL), it uses a hybrid eddy diffusion mass flux turbulence/vertical diffusion scheme. The radiation is based on Rapid Radiative Transfer Model (RRTM) [25,26]. This model is operationally used for short-range prediction in India [19]. The initial condition for the model run is generated by NCMRWF using ensemble Kalman filter (EnKF) component of hybrid global data assimilation system (GDAS) [27]. The model is initialized every day and integrated for next 10 days for the operational short and medium range forecast. The model run is carried out at the Mihir High-Performance Computing (HPC) system of NCMRWF, New Delhi under the Ministry of Earth Sciences.

The Met Office Global and Regional Ensemble Prediction System-Global MOGREPS-G) is a high-resolution (≈20 km) ensemble-based numerical weather prediction (NWP) system being developed by the UK Met Office for medium-range weather predictions [28]. It is built on the Unified Model (UM) framework, which supports a wide range of time scales, such as weather, sub-seasonal, seasonal, and climate simulations [29,30]. It is a critical component of the UK Met Office's NWP suite, providing both standalone global forecasts as well as the boundary conditions for nested regional models such as MOGREPS-UK. MOGREPS-G is coupled to a quarter-degree ocean model, enabling two-way interaction between atmosphere and ocean, which is critical for capturing tropical and extra-tropical variability. MOGREPS-G operates at a nominal horizontal resolution of ≈16 km at the equator (N640, reduced Gaussian grid configuration) with 70 vertical levels extending up to 80 km altitude. The vertical coordinate system combines height- and pressure-based levels, enabling detailed representation of atmospheric processes. The model grid uses an Arakawa A configuration with an equi-rectangular projection [31]. The MOGREPS-G uses a mass-flux convection scheme [32]. The microphysics used is a single-moment scheme [33], with extensive modifications to better represent cloud and precipitation processes [34]. MOGREPS-G provides forecasts for up to 10 days and 6 h (T+252). The ensemble consists of 18 members—one control and 17 perturbed members—and run four times daily (00, 06, 12, and 18 UTC). Perturbations are generated using the Ensemble Transform Kalman Filter (ETKF) with localization [35], and model uncertainty is addressed through stochastic physics schemes, notably Stochastic Kinetic Energy Backscatter (SKEB) and Stochastic Perturbation of Tendencies (SPPT). Since 2019, MOGREPS-G incorporates hybrid 4D-EnVar for atmospheric data assimilation and uses an ensemble of DA systems to provide flow-dependent covariances for initial-condition spread [36]. This approach enhances the representation of uncertainty in initial conditions and improves forecast reliability. In this study, we have exclusively utilized control simulation from MOGREPS-G because of the fact that GFS T1534 is a deterministic model. Model forecast outputs for the period 2020 to 2023, restricted to the months of June to September, were considered for this study.

2.2. Data and Methodologies

To evaluate the model rainfall forecast, we have used a high-resolution gridded data set from IMD-GPM, which was created by merging Global Precipitation Measurement (GPM) data with available rain gauge observations from Indian Meteorological Department. This dataset is available at $0.25^\circ \times 0.25^\circ$ [37] and the period being considered in this study is from 2020 to 2023. For purpose of comparison, the model outputs during the same time period were re-gridded to the same resolution as the observation.

In order to see how well the model forecasts agree with observations during Indian Summer Monsoon months, June to September (JJAS), we have used statistical techniques

such as mean, bias, standard deviation, and percentile analysis. Additionally, threshold based precipitation index such as, R50 days were calculated by counting the number of days with rainfall greater than or equal to 50 mm day^{-1}. Similarly, R150 days were also computed. The R50 and R150 provide crucial information about the model's ability to simulate extreme rainfall. Furthermore, to assess the forecast skill of the models, standard verification metrics including the Probability of Detection (POD), False Alarm Rate (FAR), and Bias Score were first computed [38]. These skill scores are derived from a 2 × 2 contingency table constructed using observed and forecasted events, categorized as hits (a), false alarms (b), misses (c), and correct negatives (d). The POD represents the proportion of correctly forecasted 'yes' events relative to the total number of observed 'yes' events. The POD can range from 0 to 1 with 1 being the perfect score. It is calculated as

$$\mathrm{POD} = \frac{a}{a+c}. \tag{1}$$

The FAR shows the ratio of forecasted 'yes' events that did not occur. It also ranges from 0 to 1 with 0 being the ideal score and is calculated as follows:

$$\mathrm{FAR} = \frac{b}{a+b}. \tag{2}$$

Subsequently, the Bias Score is a metric that calculates the ratio of forecasted 'yes' to the observed 'yes' occurrences. The perfect score is 1 while a score of <1 (>1) indicates under forecasting (over forecasting). It is calculated as

$$\mathrm{BIAS\ \ score} = \frac{a+b}{a+c}. \tag{3}$$

For rare events, the traditional score can be misleading and a more comprehensive score that takes care of the false alarm is F1 score [39]. It ranges between 0 to 1, where 1 is the perfect score. A higher F1 score implies better ability to detect while minimizing false alarm. It is calculated as

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \tag{4}$$

where Precision $= a/a+b$ and Recall $= a/a+c$.

3. Results and Discussion

It is important to recognize that extreme rainfall events are characterized by anomalously large deviations from the climatological mean rainfall. Therefore, understanding the ability of numerical weather prediction models to realistically simulate the mean rainfall distribution is a critical step before evaluating their skill in forecasting extremes. An accurate representation of the seasonal mean rainfall is essential, as systematic biases in the mean field can influence the reliability of extreme event simulations and forecasts.

3.1. Mean and Standard Deviations

Figure 1 illustrates the spatial distribution of model bias (model minus observation) of the mean Indian Summer Monsoon (JJAS) rainfall at lead times of up to three days. It can be seen that both models tend to overestimate the mean rainfall over various parts of the Indian region. The overestimation is higher in MOGREPS-G, particularly over the north and northwestern part of India at all lead times. On the contrary, GFS T1534 shows underestimation over the same region. It is also to be noted that the spatial variation of bias is smoother in MOGREPs-G as compared to GFS T1534. Furthermore, both models underestimate the mean rainfall over the Bay of Bengal, which is one of the major convective

centers during ISM. This underestimation is lower in MOGREPS-G compared to GFS T1534. Additionally, MOGREPS-G overestimates the mean rainfall more over the Western Ghats and the Himalayan foothills.

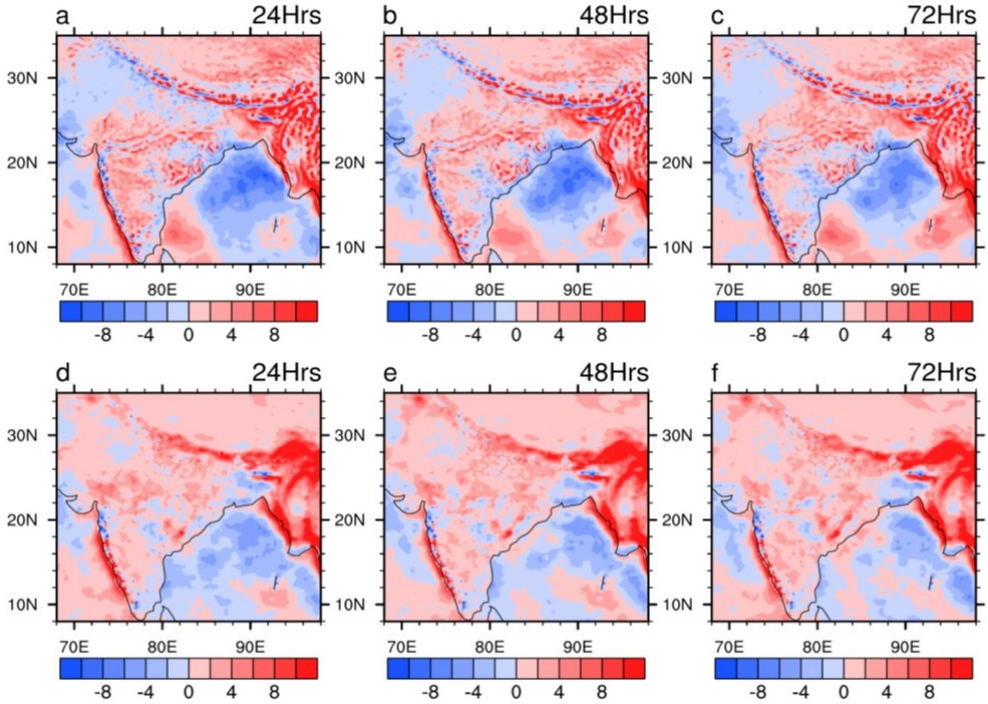

Figure 1. Spatial pattern of model bias (model minus observation) of mean JJAS rainfall (mm day^{-1}) from (**a**–**c**) GFS T1534, (**d**–**f**) MOGREPS-G. (**a**,**d**), (**b**,**e**) and (**d**,**f**) represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 55 mm day^{-1} in the interval of 5 mm day^{-1}.

To quantify the models' performance in simulating mean rainfall patterns, correlation coefficient was computed between the model-simulated and observed rainfall fields, as illustrated in Figure 2. The spatial map of the correlation coefficient provides an overall measure of agreement in the temporal variability and is a key diagnostic in assessing the model fidelity. To test its statistical significance, a Student's t-test was performed and the values in Figure 2 are only shown at a 95% significance level. The results indicate that the GFS T1534 exhibits positive spatial correlation with observations, with values reaching up to 0.5 at 24 h lead. However, the correlation values reduce with increasing lead time. On the other hand, the correlation coefficient between observations and MOGREPS-G is similar to the correlation with GFS T1534 at 24 h lead time but reduces drastically with increase in lead time and loses its statistical significance over a majority of the Indian landmass.

Relevant to the assessment of extreme rainfall, another important statistical measure is the standard deviation of rainfall. The standard deviation of JJAS rainfall over the Indian region is illustrated in Figure 3. One can clearly see that the standard deviation is underestimated by both models, implying a limited ability to reproduce the observed rainfall variability. The magnitude of this underestimation is higher in GFS T1534, especially over Central India, the Bay of Bengal, and the northeast region. It should also be noted that the MOGREPS-G overestimates the standard deviation in some parts of Northeast India.

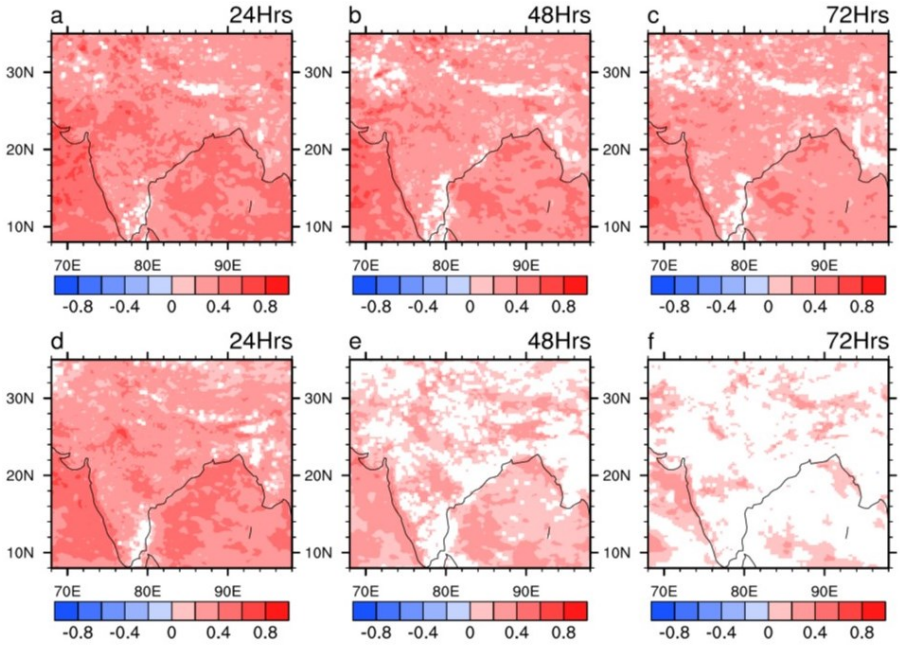

Figure 2. Spatial pattern of correlation between observation and models. (**a**–**c**) GFS T1534 and (**d**–**f**) MOGREPS-G. First, second and third column represents 24 h, 48 h and 72 h forecast respectively. The correlation values exceeding the 95% significance level are only shown. The color bar ranges from −1 to 1 in the interval of 0.2.

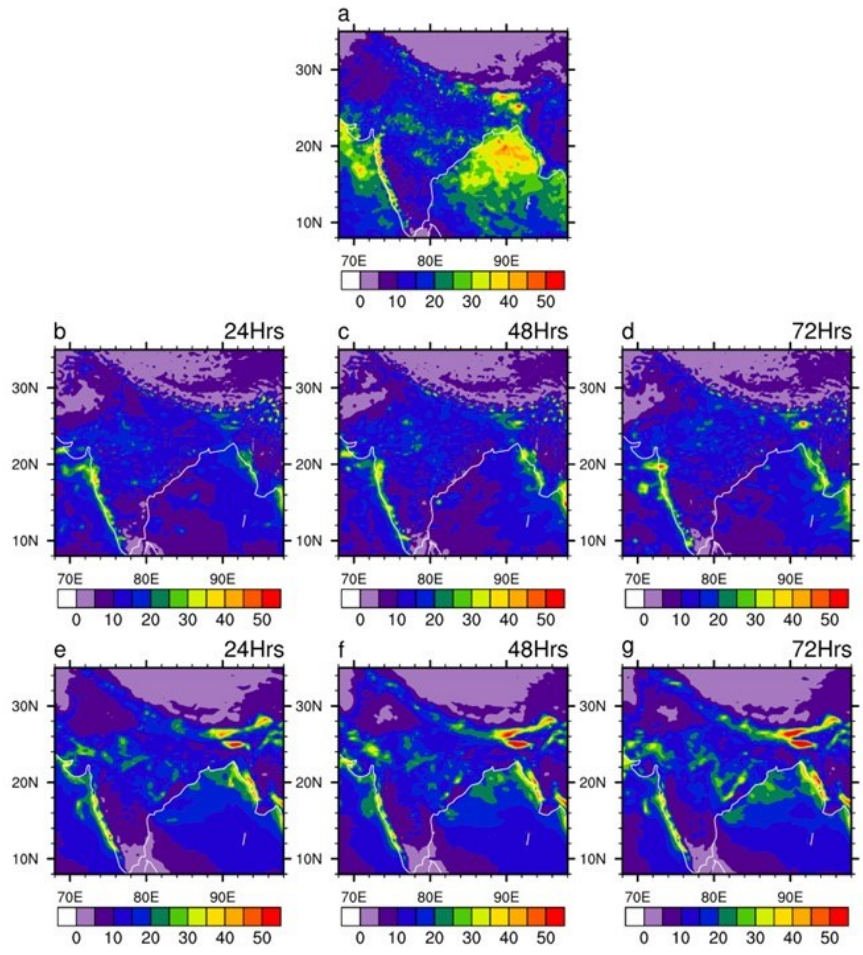

Figure 3. Spatial pattern of standard deviation (mm day^{-1}) of JJAS rainfall derived from (**a**) IMD-GPM, (**b**–**d**) GFS T1534, (**e**–**g**) MOGREPS-G. (**b**,**e**), (**c**,**f**) and (**d**,**g**) represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 55 mm day^{-1} in the interval of 5 mm day^{-1}.

3.2. Extreme Rainfall Events

To investigate the capabilities of prediction of extreme rainfall events, we statistically fix the threshold of extreme rainfall by using percentile values in accordance with previous studies [40,41]. Figure 4 shows the spatial distribution of 95th percentile values from model and observation over the Indian region. Figure 4 demonstrates that both models underestimate the 95th percentile threshold over a large part of the Indian landmass and the Bay of Bengal. This underestimation increases with increase in lead times. Although both models underestimate the 95th percentile threshold, MOGREPs-G is able to simulate the spatial structure of the extreme rainfall events better when compared to observations. Specifically, the extreme rainfall events simulated by MOGREPS-G over Western Ghats and Northeast India are, in fact, in close agreement with the observations.

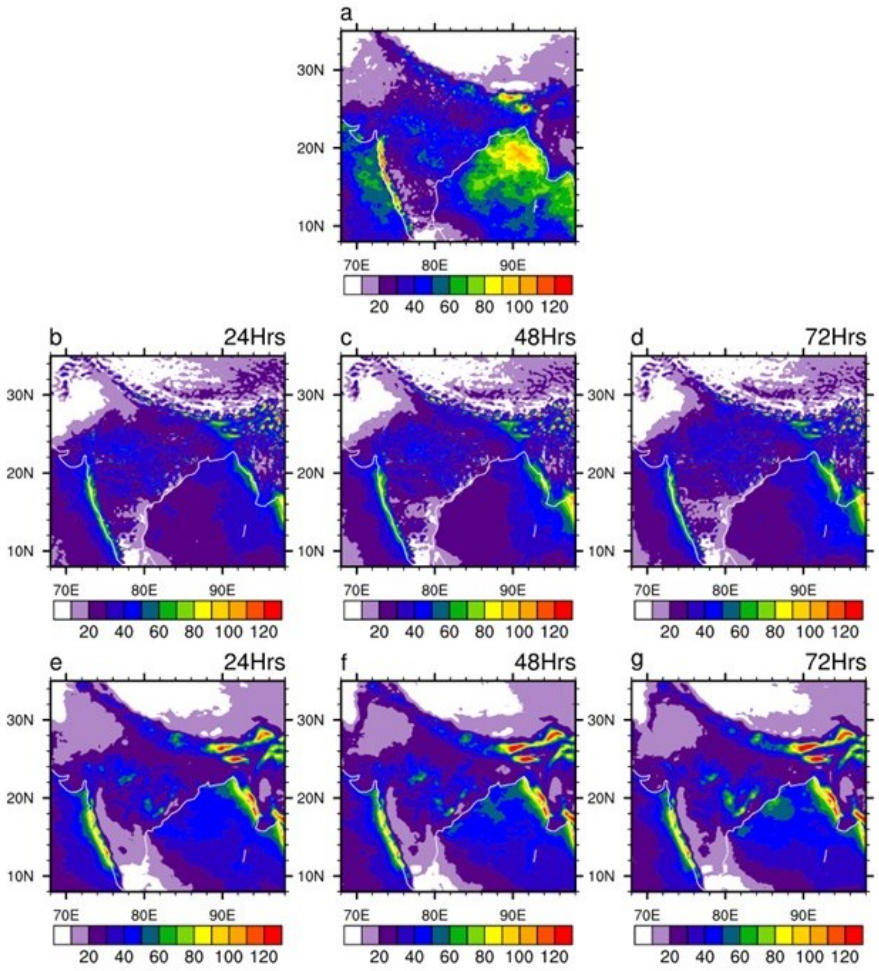

Figure 4. Spatial pattern of 95th percentile threshold (mm day^{-1}) of JJAS rainfall derived from (**a**) IMD-GPM, (**b**–**d**) GFS T1534, (**e**–**g**) MOGREPS-G. (**b**,**e**), (**c**,**f**) and (**d**,**g**) represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 130 mm day^{-1} in the interval of 10 mm day^{-1}.

Note that over most of the Central Indian region, the threshold of 95th percentile corresponds to rainfall values between 30–50 mm day^{-1}, while over the Western Ghats and some parts of Northeast India, it exceeds 80 mm day^{-1}. Hence, for our study, we define extreme events as rainfalls over 50 mm day^{-1}, contrary to the threshold of 150 mm day^{-1} used by previous studies [10,16]. In the context of extreme rainfall, an important metric to evaluate the model performances and observations is R days, which denotes the number of days with rainfall over a certain threshold [29,32,42]. Figure 5 shows the spatial pattern

of R50 days over the Indian region derived from observation and models. Observations indicate that over most of the Central Indian region, the magnitude of R50 days during our period of study (2020–2023) was between 20 to 40 days, whereas over Western Ghats and part of the northeast region, it exceeded beyond 70 days (Figure 5a). The number of R50 days simulated by GFS T1534 appear to be reasonable over these regions, but are significantly underestimated over the Central Indian region at all lead times (Figure 5b–d).

On the other hand, the number of R50 days simulated by MOGREPS-G is much closer to the observation except some region of overestimation over Western Ghats and parts of Northeast India. If the same pattern is compared with R150 days, it is apparent that the number of R150 days do not exceed beyond 3–4 days over major parts of the Indian landmass for both models and observations (Figure 6). This justifies our choice of defining the extreme rainfall threshold as 50 mm day^{-1} for this study. Despite the number of R150 days being only a few, it is interesting to note that while GFS T1534 underestimates this variable, MOGREPS-G is found to overestimate it over parts of Central Indian region and Himalayan foothills region (Figure 6). This contrast between the two models increases with increasing lead days.

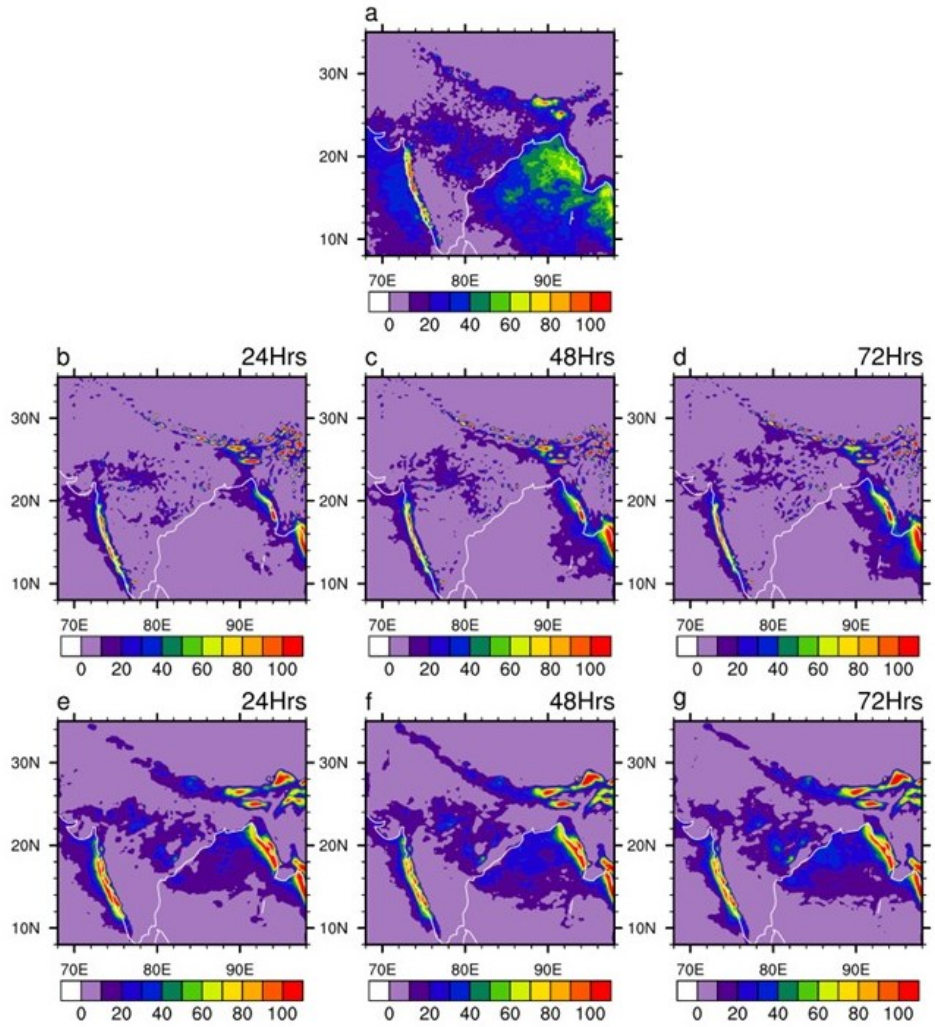

Figure 5. Spatial pattern of R50 days during JJAS derived from (**a**) IMD-GPM, (**b**–**d**) GFS T1534, (**e**–**g**) MOGREPS-G. (**b**,**e**), (**c**,**f**) and (**d**,**g**) represents 24 h, 48 h and 72 h forecast respectively.

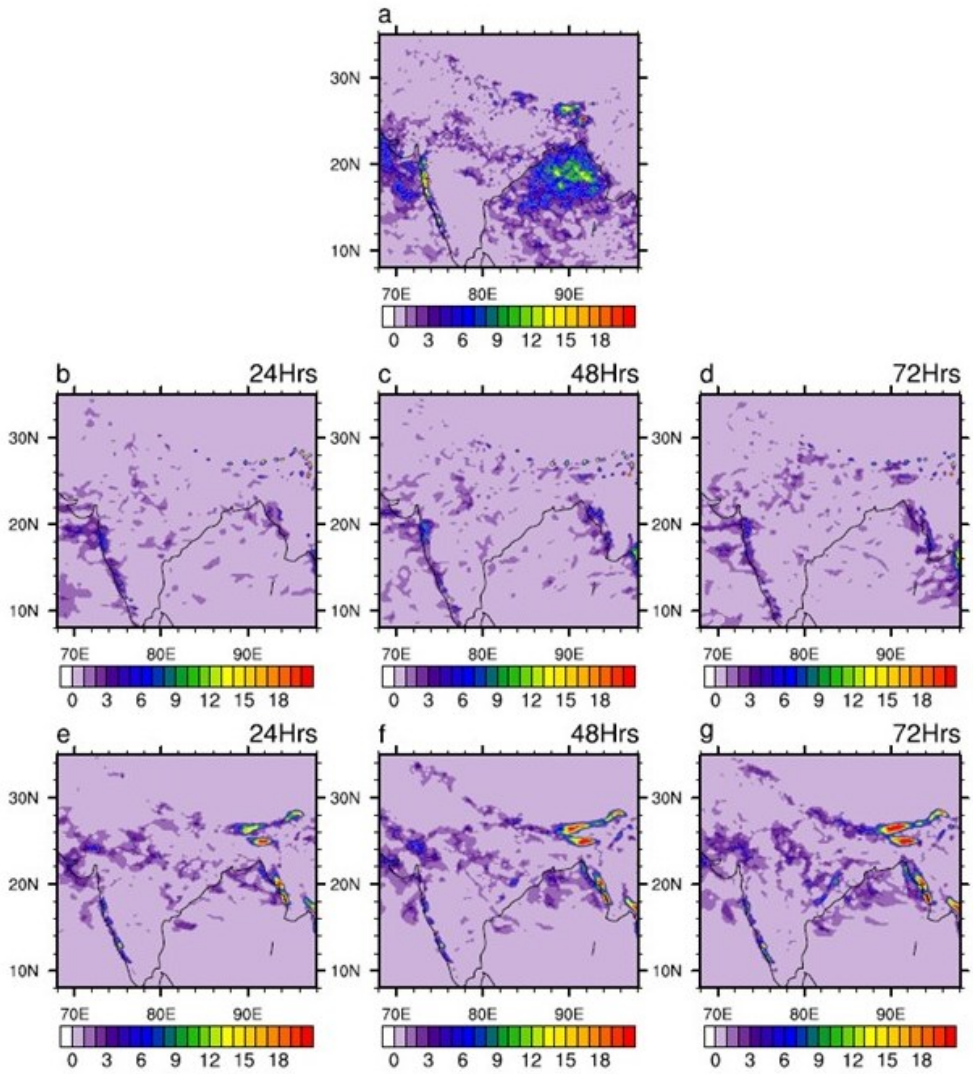

Figure 6. Spatial pattern of R150 days during JJAS derived from (**a**) IMD-GPM, (**b**–**d**) GFS T1534, (**e**–**g**) MOGREPS-G. (**b**,**e**), (**c**,**f**) and (**d**,**g**) represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 21 in the interval of 1.

3.3. Model Skill Scores

Hitherto, the comparison has been mostly qualitative. To quantify the capability of the models in forecasting extreme rainfall events, skill scores such as the Probability of Detection (POD), False Alarm Rate (FAR), Bias Score, and F1 score were computed over Indian land mass and are listed in Table 1. These categorical verification metrics are essential in quantifying the ability of a model to correctly identify extreme events, while also accounting for false alarms and systemic biases.

Table 1 depicts that both models display quite low values of POD at all lead times. Specifically, with increase in lead times, POD values drop particularly in MOGREPS-G. This highlights the models' inefficiency to capture events with rainfall more than 50 mm day^{-1}. When it comes to FAR, both models show higher values than POD. While GFS T1534 has similar FAR values for all lead times, the FAR values for MOGREPS-G decrease with increase in lead time. In fact, compared to GFS T1534, MOGREPS-G has higher FAR at a 24 h lead time and lower FAR at a 72 h lead time. From Table 1, it is evident that GFS T1534 has a higher BIAS score of 1.14 as compared to 1.07 of MOGREPS-G at a 24 h lead time. This suggests that both models tend to over predict rainfall more than 50 mm day^{-1}.

Upon increasing the lead time, the BIAS Score of GFS T1534 is reduced to 1.09, whereas for MOGREPS-G, it increased to 1.26 at a 72 h lead time.

Table 1. Spatially-aggregated model simulated skill scores for rainfall greater than or equal to 50 mm day^{-1} over Indian landmass during JJAS of 2020 to 2023.

Model	Skill Score	Forecast Hour		
		24 h	48 h	72 h
GFS T1534	POD	0.12	0.10	0.09
	FAR	0.36	0.36	0.34
	BIAS	1.14	1.14	1.09
	F1	0.12	0.09	0.08
MOGREPS-G	POD	0.12	0.07	0.05
	FAR	0.40	0.33	0.29
	BIAS	1.07	1.20	1.26
	F1	0.11	0.06	0.04

Although these metrics provide valuable insights, POD, FAR, BIAS scores may sometimes give wrong impression for rare high threshold events. A more suitable score is F1 score, which penalizes both misses and false alarms. Table 1 shows that GFS T1534 has a higher F1 score than MOGREPS-G at all lead times. This suggests that GFS T1534 has a better capability to predict rainfall of more than 50 mm day^{-1} after taking into account the False Alarm Rate. While Table 1 give a comparative view of all scores, these are spatially-aggregated metrics and therefore contain no information about their spatial variability. To understand the models' performance at different geographical locations, we present the spatial maps of these scores.

Figure 7 shows the spatial distribution of POD across the Indian region for both models at various lead times. The POD values for the GFS T1534 model at 24 h lead time range from 0.3 to 0.7 across several key monsoonal zones, including Central India, Northwest India, Northeast India, and the Western Ghats. Among these, the Western Ghats exhibit the highest POD values, indicating a relatively higher model skill in capturing extreme events over this region. However, a consistent decline in POD values is observed with increasing forecast lead time, reflecting a typical reduction in predictive skills at longer leads. On the other hand, MOGREPS-G demonstrates higher POD values than GFS T1534 at 24 h lead time over South East part of Central India and the Himalayan foothills region, suggesting better predictability. Nevertheless, it's skill markedly deteriorates with increasing lead times. At 48 and 72 h, the POD values of MOGREPS-G decline substantially, with values dropping below 0.2 over large parts of the Central Indian region. While POD assesses the models' ability to correctly forecast the occurrence of extreme rainfall events, FAR complements this information by quantifying the frequency of incorrect or false forecasts.

Figure 8 illustrates the spatial distribution of FAR for both GFS T1534 and MOGREPS-G at different lead times. At a 24 h lead time, GFS T1534 exhibits relatively high FAR values across much of the Indian region, with most areas displaying FAR in the range of 0.6–0.8 (Figure 8a). This implies that a large fraction of the model's extreme rainfall forecasts is absent in observations, thereby reducing one's confidence in these predictions. Furthermore, as seen in Figure 8b,c, the FAR values increase with longer lead times, reaching values above 0.8 in several regions by the 72 h lead. This steady increase in FAR with forecast horizon suggests a degradation in the reliability of the GFS T1534 model in simulating extreme rainfall as the lead time extends. The analysis of FAR for MOGREPS-G, presented in Figure 8d–f, reveals even higher values across the Indian landmass compared to GFS T1534. Despite MOGREPS-G showing relatively higher POD values at the 24 h lead

(Figure 7d) over some regions, it simultaneously exhibits higher False Alarm Rates over majority of the Indian landmass. This indicates that while the model has good capabilities in capturing the occurrence of extreme events, it is also more prone to over predicting them. Interestingly, regions with higher POD values in MOGREPS-G tend to coincide with regions exhibiting higher FAR values. This spatial overlap suggests that increased detection may increase the false alarms, raising concerns about the trade-off between sensitivity and precision in the model's forecasts.

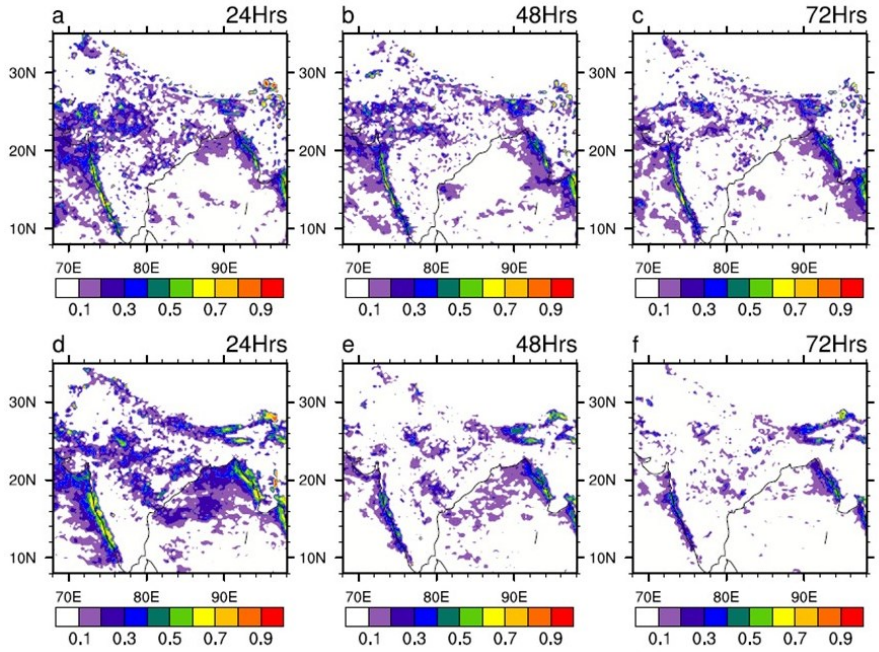

Figure 7. Spatial distribution of Probability of Detection (POD) for rainfall over 50 mm day^{-1} obtained from (**a**–**c**) GFS T1534 and (**d**–**f**) MOGREPS-G. First, second and third column represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 1 in the interval of 0.1.

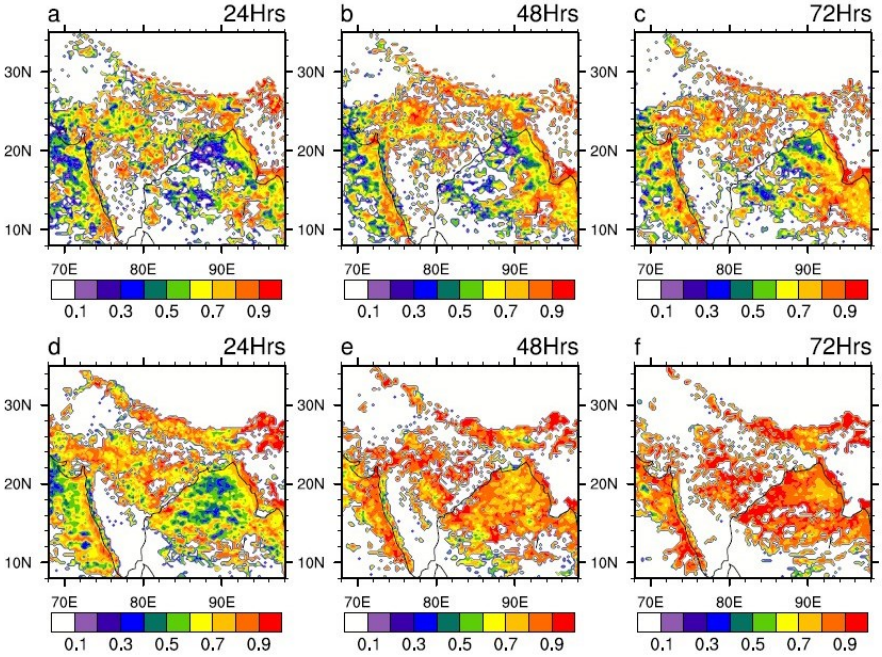

Figure 8. Spatial distribution of False Alarm Rate (FAR) for rainfall over 50 mm day^{-1} obtained from (**a**–**c**) GFS T1534 and (**d**–**f**) MOGREPS-G. First, second and third column represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 1 in the interval of 0.1.

To evaluate the performance of the model in forecasting extreme rainfall events, the Bias Score is computed for both models and shown in Figure 9. The results obtained from GFS T1534 show that this model has a tendency to under-forecast extreme rainfall events across most parts of the Indian region at all lead times (Figure 9a–c). However, there are exceptions to this general behavior. The Western Ghats, Northeast India, and isolated pockets of Central India exhibit Bias Scores greater than 1.0, indicating localized over forecasting. With increasing lead time, the model begins to exhibit over forecasting tendencies in additional small patches across Central India, suggesting a lead time-dependent shift in rainfall signal that could possibly be due to growing model errors in convective parameterizations.

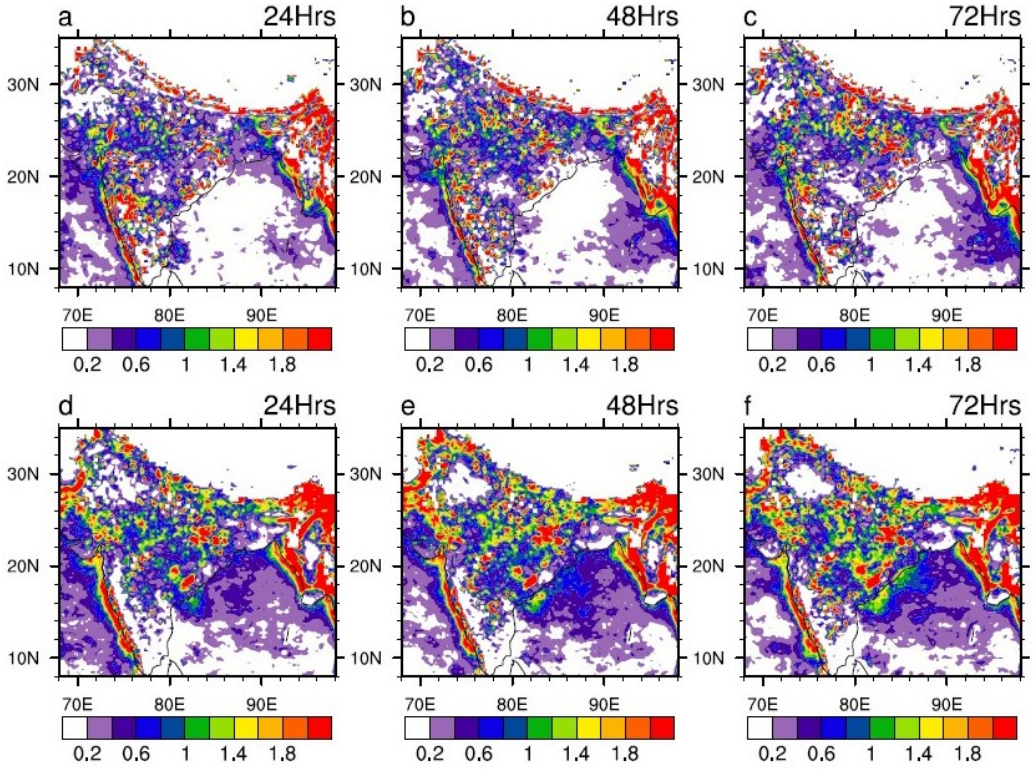

Figure 9. Spatial distribution of forecast Bias Score for rainfall over 50 mm day^{-1} computed from equation 3 for (**a**–**c**) GFS T1534 and (**d**–**f**) MOGREPS-G. First, second and third column represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 2.2 in the interval of 0.2.

In contrast to GFS T1534, the patches of under forecast are less in case of MOGREPS-G over the Indian region at all lead times (Figure 9d–f). MOGREPS-G displays widespread over-forecasting behavior across the Indian landmass, with a Bias Score exceeding 1 in most regions even at the shortest lead time. This tendency to over forecast becomes more pronounced as the lead time increases, indicating a systematic bias exists in the ensemble forecasts toward generating more extreme rainfall events than observed. These results are consistent with the previously discussed spatial distribution of False Alarm Rate (Figure 8), where MOGREPS-G was found to have high FAR values across large parts of the country. The concurrence of a high FAR and a high Bias Score suggests that the over-forecasting tendency in MOGREPS-G contributes directly to the generation of numerous false alarms. Although a higher Probability of Detection (POD) was noted in MOGREPS-G at shorter lead times, the accompanying high Bias and FAR values indicate that this detection comes at the expense of forecast precision and reliability.

Last but not the least, the F1 score gives a more comprehensive picture about the forecast accuracy (Figure 10). From Figure 10, one could see that both models perform similarly over the upper part of Central India, i.e., in the states of Rajasthan, Madhya

Pradesh, and Uttar Pradesh, at a 24 h lead time. Although the performance at the lower half of Central India varies, GFS T1534 performs better over the states like Odisha and Andhra Pradesh while MOGREPS-G demonstrates better performance over Madhya Pradesh and parts of Maharashtra at 24 h lead time. In particular, GFS T1534 performs better at higher lead times than MOGREPS-G. Finally, while GFS T1534 exhibits a general tendency to under predict extreme rainfall in few geographical locations at 24 h lead, MOGREPS-G shows a systematic overestimation, particularly at longer lead times. These contrasting biases underscore the need for model-specific post-processing and calibration techniques if these forecasts are to be used effectively for operational decision-making, especially in applications sensitive to both, missed events and false alarms.

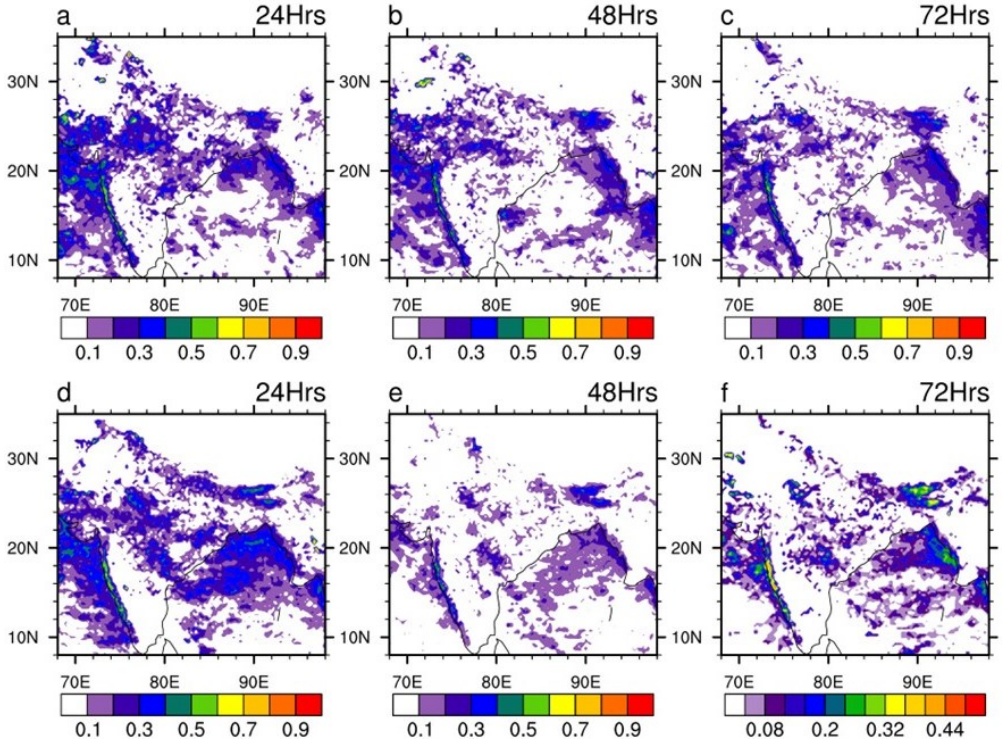

Figure 10. Spatial distribution of F1 score computed from equation 4 for (**a**–**c**) GFS T1534 and (**d**–**f**) MOGREPS-G. First, second and third column represents 24 h, 48 h and 72 h forecast respectively. The color bar ranges from 0 to 1 in the interval of 0.1.

4. Conclusions

This study presents a comprehensive evaluation of the forecast skill of GFS T1534 and MOGREPS-G models in predicting extreme rainfall events over the Indian region during the Indian Summer Monsoon (JJAS) for the period 2020–2023. Both models exhibit substantial regional biases in simulating the seasonal mean rainfall, characterized by overestimation in some regions and underestimation in others. While GFS T1534 represents the spatial distribution of mean rainfall and its correlation with observations more realistically over the Indian landmass, MOGREPS-G shows relatively lesser underestimation over the Bay of Bengal but still fails to fully capture the observed rainfall magnitude. The spatial correlation between the model forecasts and observations decreases with increasing lead time, with a more rapid degradation observed in MOGREPS-G beyond the 24 h lead, indicating reduced spatial coherence at longer forecast horizons. Analysis of rainfall variability reveals that both models underestimate the observed standard deviation. The underestimation is more pronounced in GFS T1534, while MOGREPS-G occasionally overestimates the variability in certain regions.

The number of R50 days were assessed and the results revealed that the MOGREPS-G simulates the number relatively closer to the observation over the Central Indian region compared to GFS T1534. Despite this, the POD values suggest that both models have poor skill in detecting extreme rainfall events at 1 to 3 day lead times. At a 24 h lead, both models have similar POD values over the northwestern part of the Central Indian region, but over the southeastern part of Central India, MOGREPS-G slightly outperforms GFS T1534. However, in terms of POD values, beyond the 24 h lead, GFS T1534 performs better than MOGREPS-G. The FAR and Bias Score analyses indicate that MOGREPS-G suffers from over forecasting and higher False Alarm Rates. On the other hand, GFS T1534 is somewhat more conservative in its prediction of extreme rainfall events. The F1 score provided a clearer picture of their relative performance by accounting for the over prediction. Both models display similar F1 score at 24 h lead time over the upper part of Central India in the states of Rajasthan, Madhya Pradesh and Uttar Pradesh. Although the models' performance at the lower half of Central India varies, at 24 h lead time, GFS T1534 performs better over states like Odisha and Andhra Pradesh while MOGREPS-G demonstrates better performance over Madhya Pradesh and parts of Maharashtra. At higher lead times GFS T1534 performs better over the land region. Furthermore, it was observed that, MOGREPS-G outperforms GFS T1534 over the Bay of Bengal region at all lead times.

Overall, while both models demonstrate potential in short-range forecasting, particularly within 24 h, at longer lead times, their performance deteriorates, with GFS T1534 performing slightly better. Such variations in models' performance may result from differences in their convective parameterization, cloud microphysics schemes, and data assimilation techniques. However, a detailed investigation on these aspects is out of the scope of the present study and will be reported in a future work. In terms of practical applications, the results reported here underscore the need for further model developments and regional post-processing techniques. These insights are crucial for improving early warning systems and disaster risk reduction strategies related to extreme rainfall events over India.

Author Contributions: Conceptualization: T.G. and S.R.K.; Methodology: T.G.; Formal analysis: T.G. and S.R.K.; Data curation: M.G. and S.R.K.;Writing original draft: T.G.; Review and editing: M.D., S.R.K. and S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The IMD-GPM data is available at IMD Pune's website (https://imdpune.gov.in/lrfindex.php accessed on 12 February 2026). The GFS T1534 data is available by upon request via email to Director, IITM (director@tropmet.res.in). The MOGREPS-G data is available upon request via email to Dr. Seshagirirao Kolusu (seshagirirao.kolusu@metoffice.gov.uk).

Acknowledgments: The Indian Institute of Tropical Meteorology (Pune, India) is fully funded by the Ministry of Earth Sciences, Government of India, New Delhi. We thank IMD for providing the IMD-GPM merged data. Authors thank Director, IITM, Pune for motivation and encouragement in the study. GFS T1534 runs were carried out by using Pratyush HPC facility located at IITM, Pune. The authors thank the IITM Pratyush HPC and NCMRWF Mihir HPC support team for all the necessary help. The contribution to this work by SK was funded by the Met Office Weather and Climate Science for Service Partnership (WCSSP) India project under the International Science Partnership Fund (ISPF). WCSSP India is a collaborative initiative between the Met Office and the Indian Ministry of Earth Sciences (MoES).

Conflicts of Interest: The author S.C. is the Guest Editor of a Special Issue, entitled "Novel Approaches to Predict Extreme Events in Atmospheric Flows: From Turbulence to Climate".

References

1. Gordon, H.; Whetton, P.; Pittock, A.; Fowler, A.; Haylock, M. Simulated changes in daily rainfall intensity due to the enhanced greenhouse effect: Implications for extreme rainfall events. *Clim. Dyn.* **1992**, *8*, 83–102. [CrossRef]
2. Hennessy, K.; Gregory, J.M.; Mitchell, J. Changes in daily precipitation under enhanced greenhouse conditions. *Clim. Dyn.* **1997**, *13*, 667–680. [CrossRef]
3. Easterling, D.R.; Evans, J.L.; Groisman, P.Y.; Karl, T.R.; Kunkel, K.E.; Ambenje, P. Observed variability and trends in extreme climate events: A brief review. *Bull. Am. Meteorol. Soc.* **2000**, *81*, 417–426. [CrossRef]
4. Trenberth, K.E.; Dai, A.; Rasmussen, R.M.; Parsons, D.B. The changing character of precipitation. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 1205–1218. [CrossRef]
5. Min, S.K.; Zhang, X.; Zwiers, F.W.; Hegerl, G.C. Human contribution to more-intense precipitation extremes. *Nature* **2011**, *470*, 378–381. [CrossRef]
6. Allan, R.P.; Soden, B.J.; John, V.O.; Ingram, W.; Good, P. Current changes in tropical precipitation. *Environ. Res. Lett.* **2010**, *5*, 025205. [CrossRef]
7. Wernberg, T.; Smale, D.A.; Tuya, F.; Thomsen, M.S.; Langlois, T.J.; De Bettignies, T.; Bennett, S.; Rousseaux, C.S. An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nat. Clim. Change* **2013**, *3*, 78–82. [CrossRef]
8. Curtis, S.; Fair, A.; Wistow, J.; Val, D.V.; Oven, K. Impact of extreme weather events and climate change for health and social care systems. *Environ. Health* **2017**, *16*, 128. [CrossRef]
9. Liang, S.; Wang, D.; Ziegler, A.D.; Li, L.Z.; Zeng, Z. Madden–Julian Oscillation-induced extreme rainfalls constrained by global warming mitigation. *Npj Clim. Atmos. Sci.* **2022**, *5*, 67. [CrossRef]
10. Goswami, B.N.; Venugopal, V.; Sengupta, D.; Madhusoodanan, M.; Xavier, P.K. Increasing trend of extreme rain events over India in a warming environment. *Science* **2006**, *314*, 1442–1445. [CrossRef]
11. Rajeevan, M.; Bhate, J.; Jaswal, A.K. Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett.* **2008**, L18707.
12. Ajayamohan, R.; Merryfield, W.J.; Kharin, V.V. Increasing trend of synoptic activity and its relationship with extreme rain events over central India. *J. Clim.* **2010**, *23*, 1004–1013. [CrossRef]
13. Pattanaik, D.; Rajeevan, M. Variability of extreme rainfall events over India during southwest monsoon season. *Meteorol. Appl.* **2010**, *17*, 88–104. [CrossRef]
14. Pal, L.; Ojha, C.S.P.; Dimri, A. Characterizing rainfall occurrence in India: Natural variability and recent trends. *J. Hydrol.* **2021**, *603*, 126979. [CrossRef]
15. Singh, D.; Tsiang, M.; Rajaratnam, B.; Diffenbaugh, N.S. Observed changes in extreme wet and dry spells during the South Asian summer monsoon season. *Nat. Clim. Change* **2014**, *4*, 456–461. [CrossRef]
16. Roxy, M.K.; Ghosh, S.; Pathak, A.; Athulya, R.; Mujumdar, M.; Murtugudde, R.; Terray, P.; Rajeevan, M. A threefold rise in widespread extreme rain events over central India. *Nat. Commun.* **2017**, *8*, 708. [CrossRef] [PubMed]
17. Durman, C.; Gregory, J.M.; Hassell, D.C.; Jones, R.; Murphy, J. A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Q. J. R. Meteorol. Soc.* **2001**, *127*, 1005–1015. [CrossRef]
18. Li, F.; Collins, W.D.; Wehner, M.F.; Williamson, D.L.; Olson, J.G. Response of precipitation extremes to idealized global warming in an aqua-planet climate model: Towards a robust projection across different horizontal resolutions. *Tellus A Dyn. Meteorol. Oceanogr.* **2011**, *63*, 876–883. [CrossRef]
19. Mukhopadhyay, P.; Prasad, V.; Krishna, R.P.M.; Deshpande, M.; Ganai, M.; Tirkey, S.; Sarkar, S.; Goswami, T.; Johny, C.; Roy, K.; et al. Performance of a very high-resolution global forecast system model (GFS T1534) at 12.5 km over the Indian region during the 2016–2017 monsoon seasons. *J. Earth Syst. Sci.* **2019**, *128*, 155. [CrossRef]
20. Arora, K.; Ashrit, R.; Iyengar, G.; Rajagopal, E. MOGREPS Temperatures and Total Precipitation Climatology Over India. Technical Report, NCMRWF. 2016. Available online: https://www.researchgate.net/profile/Kopal-Arora/publication/303737832_MOGREPS_Temperatures_and_Total_Precipitation_Climatology_Over_India/links/57500a5608aeb753e7b49eda/MOGREPS-Temperatures-and-Total-Precipitation-Climatology-Over-India.pdf (accessed on 3 January 2026).
21. Dube, A.; Ashrit, R.; Singh, H.; Arora, K.; Iyengar, G.; Rajagopal, E. Evaluating the performance of two global ensemble forecasting systems in predicting rainfall over I ndia during the southwest monsoons. *Meteorol. Appl.* **2017**, *24*, 230–238. [CrossRef]
22. Sela, J.G. The Derivation of the Sigma Pressure Hybrid Coordinate Semi-Lagrangian Model Equations for the GFS. Technical Report, NOAA. 2010. Available online: https://repository.library.noaa.gov/view/noaa/6971 (accessed on 3 January 2026).
23. Han, J.; Pan, H.L. Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Weather. Forecast.* **2011**, *26*, 520–533. [CrossRef]
24. Zhao, Q.; Carr, F.H. A prognostic cloud scheme for operational NWP models. *Mon. Weather. Rev.* **1997**, *125*, 1931–1953. [CrossRef]
25. Mlawer, E.J.; Taubman, S.J.; Brown, P.D.; Iacono, M.J.; Clough, S.A. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res. Atmos.* **1997**, *102*, 16663–16682. [CrossRef]

26. Mlawer, E.; Clough, S. On the extension of rapid radiative transfer model to the shortwave region. In Proceedings of the 6th Atmospheric Radiation Measurement (ARM) Science Team Meeting, US Department of Energy, CONF-9603149, Citeseer, San Antonio, TX, USA, 4–7 March 1997; Volume 510.
27. Prasad, V.; Johny, C.; Sodhi, J.S. Impact of 3D Var GSI-ENKF hybrid data assimilation system. *J. Earth Syst. Sci.* **2016**, *125*, 1509–1521. [CrossRef]
28. Bowler, N.E.; Arribas, A.; Mylne, K.R.; Robertson, K.B.; Beare, S.E. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **2008**, *134*, 703–722. [CrossRef]
29. Kolusu, S.R.; Mittermaier, M.; Robbins, J.; Ashrit, R.; Mitra, A.K. Novel evaluation of sub-seasonal precipitation ensemble forecasts for identifying high-impact weather events associated with the Indian monsoon. *Meteorol. Appl.* **2023**, *30*, e2139.
30. Mittermaier, M.; Kolusu, S.R.; Robbins, J. Mitigating against the between-ensemble-member precipitation bias in a lagged sub-seasonal ensemble. *Meteorol. Appl.* **2024**, *31*, e2197. [CrossRef]
31. Dietrich, D.E.; Ko, D.S. A semi-collocated ocean model based on the SOMS approach. *Int. J. Numer. Methods Fluids* **1994**, *19*, 1103–1113. [CrossRef]
32. Gregory, D.; Kershaw, R.; Inness, P. Parametrization of momentum transport by convection. II: Tests in single-column and general circulation models. *Q. J. R. Meteorol. Soc.* **1997**, *123*, 1153–1183. [CrossRef]
33. Wilson, D.R.; Ballard, S.P. A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.* **1999**, *125*, 1607–1636. [CrossRef]
34. Walters, D.; Baran, A.J.; Boutle, I.; Brooks, M.; Earnshaw, P.; Edwards, J.; Furtado, K.; Hill, P.; Lock, A.; Manners, J.; et al. The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. *Geosci. Model Dev.* **2019**, *12*, 1909–1963. [CrossRef]
35. Bishop, C.H.; Etherton, B.J.; Majumdar, S.J. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Weather. Rev.* **2001**, *129*, 420–436. [CrossRef]
36. Inverarity, G.; Tennant, W.; Anton, L.; Bowler, N.; Clayton, A.; Jardak, M.; Lorenc, A.; Rawlins, F.; Thompson, S.; Thurlow, M.; et al. Met Office MOGREPS-G initialisation using an ensemble of hybrid four-dimensional ensemble variational (En-4DEnVar) data assimilations. *Q. J. R. Meteorol. Soc.* **2023**, *149*, 1138–1164. [CrossRef]
37. Mitra, A.K.; Prakash, S.; Pai, D. Operational 24-hour accumulated rainfall dataset over Indian region by merging rain gauge and multi-satellite estimate from 'Global Precipitation Measurement'mission. *J. Earth Syst. Sci.* **2025**, *134*, 1–12. [CrossRef]
38. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*; Academic Press: Cambridge, MA, USA, 2011; Volume 100.
39. Datta, A.; Si, S.; Biswas, S. Complete statistical analysis to weather forecasting. In *Computational Intelligence in Pattern Recognition: Proceedings of the CIPR 2019*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 751–763.
40. Krishnamurthy, C.K.B.; Lall, U.; Kwon, H.H. Changing frequency and intensity of rainfall extremes over India from 1951 to 2003. *J. Clim.* **2009**, *22*, 4737–4746. [CrossRef]
41. Goswami, T.; Mukhopadhyay, P.; Krishna, R.P.M.; Rajeevan, M.; Chowdhuri, S. Can a 12-km GFS Model Simulate the Observed Relationship between Cloud Optical Properties and Extreme Rainfall of Indian Summer Monsoon? *Weather. Forecast.* **2025**, *40*, 79–91. [CrossRef]
42. Zhu, H.; Wheeler, M.C.; Sobel, A.H.; Hudson, D. Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Weather. Rev.* **2014**, *142*, 1556–1569. [CrossRef]