Evaluation of five high-resolution global model rainfall forecasts over India during monsoon 2020

RAGHAVENDRA ASHRIT1,* , MOHAN S THOTA1, ANUMEHA DUBE1, KONDAPALLI NIRANJAN KUMAR1, S KARUNASAGAR2, SUSHANT KUMAR1, HARVIR SINGH1, RAJASEKHAR MEKA3, R PHANI MURALI KRISHNA4 and ASHIS K MITRA1

2 Meteorological Centre, India Meteorological Department, Ministry of Earth Sciences, Government of India, Amaravathi, India.
3 Indian Space Research Organisation, Sathish Dhawan Space Centre, Sriharikota, India.
4 Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Government of India, Pune, India.
*Corresponding author. e-mail: raghu@ncmrwf.gov.in  raghu.ashrit@nic.in  rgashrit@gmail.com

MS received 2 February 2022; revised 11 June 2022; accepted 27 June 2022

This study aims to evaluate the performance of five global medium-range operational NWP model rainfall forecasts, namely NCUM, UKMO, IMD GFS, NCEP GFS and ECMWF to provide an intercomparison of rainfall forecasts over India in terms of skill in predicting daily rainfall (24-hr accumulated rainfall). Verification and intercomparison of rainfall forecasts over India during monsoon 2020 (JJAS) are carried out using both (i) standard traditional verification methods (POD, FAR, RMSE, etc.) and (ii) advanced spatial verification methods (MODE, FSS). The evaluation also includes assessment of large-scale mean patterns, temporal evolution of spells during the season, dominant modes using spectral analysis, basin-scale rainfall time series and isolated heavy rainfall cases. Our analysis suggests that some of the key large-scale aspects of monsoon (seasonal mean, active/break spells, and northward propagation) are realistically represented in all the models, with slight discrepancies. In addition, the spectral analysis of rainfall is in association with observed rainfall in Day-1 forecast and deteriorates with lead times. Synoptic variance in NCUM on longer leading times is closer to observations. While the standard categorical verification over India as a whole (spatial averaged) suggests that ECMWF forecast skill is relatively high among the five models, the verification over the sub-regions shows mixed results with no clear unique higher performer among the models. In addition, basin-scale verification of rainfall forecasts for five rivers over the Indian subcontinent shows a fairly good amount of skill in terms of CC and RMSE up to Day-3 with comparable scores among the models. The advanced spatial verification metrics, like MODE and FSS, applied to the models show varying skills with different attributes. However, for FSS, forecast skill was high (low) for lower (higher) rainfall thresholds of 20 mm/day (100 mm/day). Though different models with different spatial resolutions show reasonable skill scores for larger regions, for high-impact heavy rainfall events, which are generally localised, the models have very comparable poor skill with no clear edge by a model among the five models.

Keywords. Monsoon; heavy rains; forecast verification; spatial verification.

Supplementary material pertaining to this article is available on the Journal of Earth System Science website (http://www.ias.ac.in/Journals/Journal_of_Earth_System_Science).

Published online: 15 December 2022
1. Introduction

Asian summer monsoon (ASM) rainfall largely influences the agriculture and economics of the south and southeast Asian countries, and the heating associated with it impacts the large-scale circulation over tropical regions (Yasunari and Seki 1992; Webster et al. 1998). In specific, within the season (i.e., June through September, JJAS), rainfall over the Indian subcontinent fluctuates between enhanced and suppressed convection known as active and break spells, respectively (Ramanurthy 1969). These spells are inherent features of the monsoon and their occurrence and duration dictate the magnitude of seasonal rainfall (Gadgil 2003; Rajeevan et al. 2006, 2010). Studies indicate that these rainfall spells are the super-position of several time-scale oscillations, ranging from synoptic (2–7 days); quasi-biweekly (10–20 days) and sub-monthly (30–50 days), also known as Madden–Julian oscillation (MJO) (Krishnamurti and Gadgil 1980; Bhat 2006; Rajeevan et al. 2010). These intraseasonal oscillations (ISOs) influence the timing of onset and withdrawal of monsoon and associated precipitation (Ajaya Mohan and Goswami 2003) and also modulate sub-seasonal variability over the Indian subcontinent. Therefore, accurate simulation of space-time characteristics of the monsoon using the forecast system is vital for the better prediction of seasonal monsoon rainfall.

In recent times, there has been a considerable improvement in the numerical models and data assimilation techniques. As a result, the accuracy of the weather forecasts is also improved. Despite continuous efforts by various numerical weather prediction (NWP) and research centers, the ability to accurately simulate the key features and skill in predicting the monsoon is still an ongoing challenge (DelSole and Shukla 2002; Turner and Annamalai 2012) and systematic errors in key atmospheric and ocean variables persist at various time scales (Sperber et al. 2013; Annamalai et al. 2017). General circulation models (GCM) have become an essential tool for dynamical prediction of monsoon rainfall and the skill in predicting the seasonal mean monsoon rainfall by GCMs remains limited (Kang et al. 2004; Kumar et al. 2011). In a study, Sperber et al. (2013) using the CMIP model analysis, found that the cause for the low skill in dynamical prediction of monsoon is the inaccurate simulation of the annual cycle in global models. One of the prominent systematic errors in all the models is the presence of dry bias over south Asia and wet bias over the equatorial Indian Ocean (EIO) region. In a subsequent study, Annamalai et al. (2017) demonstrated the importance of coupled processes over EIO as a prominent cause for the systematic errors of monsoon. Earlier Martin et al. (2010), using hindcast data, also noted these model errors in rainfall over the Indian Ocean. Few studies have attributed these biases to the mis-representation of physical processes, especially convective parameterisation in models (Martin et al. 2010; Boos and Hurley 2013; Levine et al. 2013) A number of studies in the past also demonstrated the capability of regional models embedded in GCM to simulate the Indian summer monsoon climatology and concluded that the improvement in spatiotemporal distribution is due to the increased model resolution (Dash et al. 2006). In addition, few studies also performed the sensitivity of convective parameterisation schemes on mean monsoon simulations and found that performance of simplified Arakawa–Schubert (SAS) and Grell scheme are better in simulating rainfall distributions compared to other schemes (Venkata Ratnam and Kumar 2005; Dash et al. 2006). Along with biases in the longer climate simulations mentioned above, forecasting monsoons in shorter time scales is also challenging (Gadgil and Srinivasan 2012; Ranade et al. 2014). It is noted that this bias is not only in climate simulations but also seen on NWP scales (Mitra et al. 2013; Prakash et al. 2016a, b). While these biases can be improved by combining post-processing techniques with bias correction (Mitra et al. 2011; Joshi and Kar 2016), it is very important to capture the physics and dynamics of the monsoon well for better forecasts. Evaluation of UM rainfall forecasts over India and improved skill in predicting heavy rains in recent years is discussed in Ashrit et al. (2015) and Sharma et al. (2017, 2021). While there is a general improvement in skill in predicting heavy rains, there are challenges when it comes to real-time forecasting of specific events like the heavy rainfall over Uttarakhand during 2013 (Dube et al. 2014), Kerala during August 2018 (Ashrit et al. 2020). Prediction of heavy and extreme rain is still a challenge even for high-resolution models and perhaps ensemble-based probabilistic approach is more suited, as shown in
Ashrit et al. (2020) and Mukhopadhyay et al. (2021).

State-of-the-art high-resolution NWP models allow for more realistic structure and variability in the rainfall patterns over the tropics and include better representation of topographically influenced rainfall associated with mountains and coastlines. Quantifying the accuracy of high-resolution quantitative precipitation forecasts (QPFs) can be tricky, as traditional verification statistics based on point observations from gauges and/or radar data severely penalise fine-scale differences of a severe weather event. This results in the problem of double penalty, i.e., the high-resolution models get penalised twice (once for not occurring where it should be (missed event) and second for occurring where it should not be (false alarm)) (Roebber et al. 2004; Rossa et al. 2008; Zingerle and Nurmi 2008). These false alarms and missed events dominate the 2 × 2 contingency table and result in poor traditional verification scores. In addition, these poorer verification scores also tend to undermine the perceived added benefit of the high-resolution modelling, in terms of forecast realism.

In recent years, many new spatial verification approaches have been proposed which are able to reflect more adequately various advantages of high-resolution forecasts (Casati et al. 2008; Gilliland et al. 2009). In a study, Gallus et al. (2010) note that spatial verification methods provide more informative measures of forecast performance, which better reflect the quality of high-resolution forecasts. These new methods do not rely on exact grid-to-grid match but compare the properties of matched pairs of forecast and observed objects. These are known as object-based techniques, which verify the location, size, shape, intensity, and other attributes of the object, and are therefore very intuitive in their interpretation (Ebert and Gallus 2009). Method of object-based diagnostics evaluation (MODE) (Davis et al. 2006a, b) is an object/feature-based technique that quantifies forecast objects in terms of various object parameters (attributes). A brief summary of MODE is presented in section 2.2 with a detailed description in Appendix A1. The results of MODE analysis are discussed in section 4.1. Neighbourhood, also known as fuzzy, methods measure the accuracy of the forecasts within space-time neighbourhoods. All grid-scale values within a spatial and/or temporal neighbourhood of the observation are considered to be equally likely estimates of the true value. In ‘neighbourhood verification’, scales for which the forecasts have useful skills are determined by varying the sizes of the neighbourhoods and performing the verifications at multiple scales and for multiple intensity thresholds. Fractions skill score (FSS; Roberts and Lean 2008) is an example, which is used in this study and described briefly in section 2 (details in Appendix A2).

In the present study, detailed verification of rainfall forecasts from five operational high-resolution global models is carried out over India during JJAS 2020. The five operational models are from (i) National Centre for Medium-Range Weather Forecasting Unified Model (NCUM), (ii) Met Office UK’s Model (UKMO), (iii) India Meteorological Department’s Global Forecasting System (IMD GFS), (iv) National Centre for Environmental Prediction’s Global Forecasting System (NCEP GFS) and (v) European Centre for Medium-Range Weather Forecasting (ECMWF). Additional details of the modelling systems are given briefly in table 1. NCUM model has Advanced ‘ENDGame’ (Even Newer Dynamics for General atmospheric modelling of the environment) dynamical core of Met Office UK. It employs the ‘Hybrid 4-D var DA system, which uses the ensemble forecasts of ‘NCMRWF Ensemble Prediction System (NEPS)’. The NCEP GFS has a dynamical core of finite volume cubed sphere (FV3) with hybrid 4-D var data assimilation (DA) system. The dynamical core in the Met Office model uses grid point discretisation on a latitude–longitude system with a rotated pole and it employs the hybrid incremental 4-D var DA system similar to that of ECMWF. The IMD GFS model has 64 hybrid sigma pressure levels with a dynamical core based on semi-lagrangian spectral global model and GDAS EnKF–GSI hybrid data assimilation system (Mukhopadhyay et al. 2019). Utilising the suite of operational models, this study aims to provide a detailed intercomparison of rainfall forecasts over India in terms of skill in predicting rainfall. The evaluation presented in the study is quite exhaustive. It is based on standard traditional verification methods using 2 × 2 contingency table and also uses advanced spatial verification methods. The evaluation includes assessment of large-scale mean patterns, temporal evolution of rainfall spells during the season, dominant modes using spectral analysis, basin-scale rainfall time series and isolated heavy rainfall cases.
Table 2. Cases of heavy rainfall events over different sub-regions over India for which fractions skill scores (FSS) have been computed.

<table>
<thead>
<tr>
<th>Sub-regions</th>
<th>Domain</th>
<th># Cases</th>
<th>Heavy rain cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Coast South</td>
<td>WCS</td>
<td>14</td>
<td>Jun(1-4, 17-18), Jul(3, 8, 16), Aug(4-5), Sept(7, 22)</td>
</tr>
<tr>
<td>West Coast Central</td>
<td>WCC</td>
<td>13</td>
<td>Jun(2, 3, 30), Jul(17, 24, 29, 30), Aug(2, 8-10), Sept(3, 7, 20)</td>
</tr>
<tr>
<td>Mumbai Region</td>
<td>MUM</td>
<td>13</td>
<td>Jun(13), Jul(4-6, 15-17, 27), Aug(4, 6, 29), Sept(23-24)</td>
</tr>
<tr>
<td>Northwest India</td>
<td>NW</td>
<td>13</td>
<td>Jul(5-8), Aug(8, 14, 22-25, 30-31)</td>
</tr>
<tr>
<td>Central India</td>
<td>CI</td>
<td>12</td>
<td>Jun(16), Jul(7, 10, 23, 29), Aug(5, 13, 19, 20, 26-28)</td>
</tr>
<tr>
<td>Northeast India</td>
<td>NE</td>
<td>15</td>
<td>Jun(18, 26-27), Jul(2, 9-12, 19, 22), Aug(4, 18), Sep(1, 22-23)</td>
</tr>
</tbody>
</table>

2. Verification strategy and data used

2.1 Observed and forecast rainfall data

The forecasts used from five models feature slightly varying grid resolutions, which can all be considered high resolution (table 1). NCUM, UKMO, IMD GFS (GFS hereafter), NCEP GFS (NCEP hereafter) and ECMWF are the five models. As can be understood from table 1, all five models have fine grid resolutions (~10–20 km grid). Models with such high resolution are expected to capture high rainfall amounts typically associated with steep orographic and monsoon low-pressure systems.

The observed gridded rainfall analysis used for verification of model forecast rainfall is the IMD–NCMRWF merged (Satellite + Gauge) data for the entire season June–September 2020, which is available at 0.25° × 0.25° grid resolution (Mitra et al. 2009; Prakash et al. 2016a, b). Additionally, for verification of high-impact heavy rainfall events (select cases; table 2) over smaller regions, GPM data (Huffman et al. 2014) at a grid resolution of 0.1° × 0.1° is used.

Different verification approaches are used to bring out the skill in different aspects of the forecasts. For users, it is important to focus on metrics that are most relevant to the sectorial application. From the user’s perspective, it is important not to focus on one popular model or based on performance based on a single case or small data sample. Even the models with a very high reputation or good overall skill may not be the best for every application. Thus, the performance of these five models in forecasting rainfall over India is evaluated and presented in the following four distinct aspects in four steps:

(i) Comparison of observed and forecast mean rainfall averaged over the entire season (JJAS 2020), a comparison of observed and forecast rainfall frequencies, comparison of spatial mean time series is presented at a grid resolution of 0.25° × 0.25°. Additionally, spectral analysis of observed and forecast rainfall time series is presented to evaluate dominant modes of variability.

(ii) Quantitative verification over the Indian land region is carried out using standard traditional categorical verification based on a 2×2 contingency table. The traditional verification metrics (probability of detection (POD), false
alarm ratio (FAR), frequency bias/bias score (BIAS), Peirce’s skill score (PSS) and symmetric extremal dependence index (SEDI) are computed at a grid resolution of $0.25^\circ \times 0.25^\circ$.

(iii) Verification is also presented over smaller geographical regions of India in terms of root mean squared errors (RMSE) and correlation coefficients (CC) computed for spatial mean rainfall over six regions indicated in figure 1, described in table 2. Verification over river basins (not shown; Supplementary material in section S1) is also presented using basin averaged rainfall series in terms of RMSE and CC. These results correspond to verification at a grid resolution of $0.1^\circ \times 0.1^\circ$.

(iv) Additionally, spatial verification of rainfall forecasts is presented using the method of object-based diagnostics evaluation (MODE) for the entire season of JJAS 2020 and using fractions skill score (FSS) for high-impact heavy rainfall events over sub-regions over India indicated in figure 1, described in table 2.

A brief description of the two spatial verification methods is presented here with a detailed description in the Appendix.

![Figure 1. Six sub-domains over India chosen for regional verification of rainfall forecasts over India during monsoon (June–Sept) 2020.](image)

2.2 Spatial verification of heavy rainfall events

2.2.1 MODE

Method of object-based diagnostics evaluation (MODE) is a spatial verification method that provides an object-based verification of gridded forecasts and observations (Davis et al. 2006a, b, 2009; Bullock et al. 2016). In this method, the degree of similarity between the two object pairs (forecast and observed objects) is used to quantify the forecast quality. This procedure mimics what a human expert would do to find features and decide whether a given feature in a forecast represents an analogous feature in the observations (Davis et al. 2009).

See Appendix A1 for more details. Based on the user-defined rainfall thresholds (e.g., 20, 40, 60 mm), observed and forecast object pairs are constructed, and for each pair, differences between their attributes are computed. The paired attributes used here are: symmetric difference, centroid distance, area ratio, angle difference, and intensity ratio. Further, based on user-defined weights and fuzzy logic, the ‘total interest’ value, an integrated measure of forecast performance among all the attributes, is also computed. In the next step, based on the computed interest values, objects are matched across fields and merged within the same field. Finally, statistics are written out summarising the characteristics of the (a) single objects, (b) pairs of objects, and (c) matched/merged objects. Thus, MODE provides information that is not possible to obtain using traditional grid-point-based verification methods. In this study, thresholds of daily accumulated rainfall amounts exceeding 20, 40, 60, 80, 100, and 120 mm are chosen for MODE analysis. Object pair attributes include (a) centroid distance (small is good), (b) axis angle difference (small is good), (c) area ratio (close to 1 is good), (d) symmetric difference (small is good) and (e) total interest (close to 1 is good).

2.2.2 Fractions skill score (FSS)

Fractions skill score (FSS) belongs to the category of fuzzy or neighbourhood spatial verification methods. Fuzzy methods also do not look for a grid-to-grid match but relax the criteria by verifying forecasts in the local neighbourhood of the observations. FSS is a scale-selective method that can provide a measure of how the forecast skill
varies with the chosen spatial scale; in other words, it can help in determining the scales at which the forecasts become useful. It is usually used for the verification of high-resolution gridded rainfall forecasts. Roberts and Lean (2008) introduce and describe the FSS in detail. Further, spatial and temporal variation of skill using FSS is presented in Roberts (2008). A brief description of the method is presented in Appendix A2. The FSS values range from 0 (for the worst forecast) to 1 for the perfect forecast. FSS has the lowest value when forecasts are verified at the grid-scale only, i.e., the neighbourhood has only a single point. As the size of the neighbourhood increases, the skill also increases until it assumes an asymptotic FSS value known as AFSS.

3. Rainfall forecast verification

3.1 Mean and mean error (ME)

The observed and forecast (Day-3) mean monsoon rainfall during JJAS 2020 is shown in figure 2. Some of the key features are: (i) Observations indicate the highest mean rainfall exceeding 15 mm/day over the Western Ghats, the Arrakkan coast and parts of northeastern (NE) India adjoining Bangladesh. (ii) Another region of high rainfall >10 mm/day is prominent over eastern India (core monsoon region) and along the foothills of the Himalayas. (iii) During the active spells in the summer monsoon season, low-pressure systems form over the Bay of Bengal and move northwesterly across northern parts of India, bringing widespread rainfall over the core monsoon region. (iv) Northwestern parts of India and eastern parts of the southern peninsula show reduced rainfall amounts (<6 mm/day). The panels in figure 2(b–f) show Day-3 forecast rainfall averaged during the same period for NCUM, UKMO, GFS, NCEP and ECMWF, respectively. The observed peak rainfall amounts (>15 mm/day) along the Western Ghats, the Arrakkan coast and parts of northeastern India are predicted by all five models. However, it is found that both the unified models (NCUM and UKMO) forecast show higher rainfall amounts all over the west coast and over the Himalayas. The overestimation of rainfall over the NE region is noticed in all model forecasts. The forecasts overestimate the isolated high rainfall amounts (>10 mm/day) over the core monsoon region in NCUM, UKMO and GFS. The reduced rainfall amounts (<6 mm/day) over the eastern parts of the peninsula

Figure 2. Mean (JJAS 2020) observed and Day-3 forecast rainfall (mm/day) over India in the five high-resolution global models.
and northwest India are predicted fairly well in the models, particularly the very light rain over the southern tip of India.

Quantification of rainfall bias in terms of models means error (ME) in Day-3 forecasts is presented in figure 3. The positive (negative) values indicating wet (dry) bias are shown in blue (red). (i) Over the north and northeastern parts of India, models have a wet bias. Similarly, the wet bias is evident along the west coast of India over the Sea in NCUM, UKMO and GFS and relatively less evident in NCEP and ECMWF. (ii) Dry bias over western India and adjoining a large part of the Arabian Sea and head Bay of Bengal (BoB) is evident in all the models. Widespread dry bias over BoB extending to parts of the eastern peninsula is also prominent in all five models. It is found that the magnitude of biases grows with lead time (not shown). It is remarkable to note that all five models exhibit very similar biases in magnitude and location. Large parts of the Bay of Bengal region and the Arabian Sea near the Indian west coast prominently feature dry bias in all five models. Similarly, over the west coast of India, central and eastern India (core monsoon zone) and NE India show prominent wet bias.

3.2 Prediction of rainy days, moderate and heavy rain days

To further assess the wet bias over the land regions, observed and forecast counts of rainfall exceeding different thresholds are examined in this section. Results are discussed for rainy days (>1 mm/day), moderate rain (>15.6 mm/day) and heavy rain (>64.5 mm/day) days, as per the IMD definitions. The observed rainfall counts and POD (FAR) in the Day-3 forecast are shown in figure 4 (figure 5) for the above three categories. Observations show (a) 50–75 rainy days (yellow shaded) over a large area except over NW India, the J & K region and parts of the southern tip of the peninsula. (b) More than 75 rainy days (orange shade) can be seen over some parts of the core monsoon region, the west coast of India and parts of NE India. Over the dry regions of NW India and the eastern peninsula, the number of rainy days is lower, in the range of 25–50 days (green shade). Except for some isolated locations over the west coast of India, observations do not indicate counts >100 anywhere over the Indian land region. However, the model forecasts over most parts of India show an exceedingly high number of rainy days.
day counts (not shown). High values of POD (>0.9) over eastern India, NE India and the west coast are prominent in NCUM and ECMWF (figure 4b–f). High POD (>0.9) in UKMO, GFS and NCEP is not as widespread as in the other two model forecasts. Further, the forecasts feature a reduction in the number of rainy days from Day-1 to Day-5 (not shown). On the other hand, FAR in figure 5(b–f) show widespread FAR values in the range of 0.3–0.6, indicating false alarms.

The moderate rainfall day counts in the observations (figure 4g) is at least >=1 (purple) over NW India and the eastern peninsula. Nearly most of the Indian land region shows counts of >1. Over most parts of northern India (except NW India) and the peninsula, moderate rain day count >5 (blue) is prominent. Large parts of eastern India (core monsoon region) show counts of 10–25 (cyan) in the observations, with isolated regions showing 25–50 (green). The forecasts tend to show a rather moderate (0.3–0.6) magnitude of POD (figure 4h–l) over most parts of India in predicting the moderate rain category. NCUM, UKMO and ECMWF forecasts indicate slightly higher (>0.6) POD along the west coast and NE India. On the other hand, the forecasts feature higher magnitudes of FAR (>0.6) as can be seen in figure 5(h–l).

For the heavy rain category (bottom panels; figure 4m–r), observations indicate counts >10 over the west coast (cyan) and >5 or so in isolated locations over central India (blue). However, large areas over central India indicate counts of at least 1 (purple). The POD values in the model forecasts are rather too low in all the models. However, over the west coast and isolated locations widespread over central India, POD values >0.3 are prominent. The FAR values for heavy rains (figure 5m–r) show very high values (>0.9) widespread over central and eastern India. The spatial distribution of observed heavy rain counts (figure 5m) matches very well with the spatial distribution of FAR (figure 5n–r). This suggests that models successfully capture the heavy rainfall climatology. It is due to spatial and temporal miss-match in forecast of heavy rains that lead to poor POD and other metrics. These forecasts could be very useful to develop probabilistic products for heavy rainfall warnings based on neighborhood processing (Schwartz and Sobash 2017) using very high-resolution forecasts.

Thus the wet bias over central and eastern India discussed in figure 3 can be attributed to high POD values of rainy day counts and even higher FAR values of the moderate and heavy rain day count.

Figure 4. Observed counts of rainy days and probability of detection (POD) in the Day-3 forecasts (a–f). Panels (g–l) correspond to moderate rain days and panels (m–r) correspond to heavy rain days over India during monsoon (JJAS) 2020. (Rainy, moderate rain and heavy rain days are defined by 24-hour accumulated rainfall > 1, 15.5 and 64.5 mm/day thresholds, respectively, as per the IMD defined rainfall categories).
All five models consistently overestimate rainy days and heavy rain counts, which is reflected in higher false alarms.

3.3 Active/break spells and northward propagation

During monsoon season, a major component of variability exists from the intraseasonal oscillation of active (enhanced convection) and break (subdued or no convection) spells over the Indian subcontinent. Active and break spells over the Indian subcontinent during monsoon season JJAS 2020 were identified using the standardised rainfall anomalies from merged (IMD–NCMRWF) gridded daily rainfall data over central India (15°–25°N, 72°–85°E; or core monsoon zone, CMZ). If the standardised rainfall anomaly is more than +1 (−1) (standard deviation) SD for three continuous days, it is considered an active (break) spell (Rajeevan et al. 2010). Based on the criterion, five active spells and one break event were identified. The occurrence time and the duration of active and break spells are shown in table 3. The spells are short and occurred mostly in the second half (i.e., during August and September) of the monsoon 2020. The active spell during the first week of June is due to the severe cyclonic storm (SCS) ‘Nisarga’ that passed over the Indian subcontinent.

Observed and model forecast rainfall averaged over the CMZ is shown in figure 6. From the figure, a clear one-to-one comparison between observations and rainfall forecast is seen. Peaks and magnitudes of the rainfall forecasts are well predicted over CMZ. Specifically, suppressed and enhanced rainfall phases during the months of July and August are well predicted in all the models on all the forecast days. Despite having a very good comparison between modelled precipitation and observations, a close examination reveals that with lead time, i.e., in Day-3 and Day-5 forecasts exhibit slight discrepancies in rainfall magnitudes. For instance, Day-5 forecasts from all the models show significant differences from the observations. Also, NCUM shows a clear overestimation, especially during the first week of August.

The advance of the monsoon over Indian land regions is manifested by the northward propagation of the rainfall bands. The Hovmoller plots of observed and model forecast rainfall averaged over longitudes 70°–85°E for Day-3 are shown during JJAS 2020 in figure 7(a–f). The x-axis shows the days within the season and the y-axis in each panel show latitudes. The northward propagation of

![Figure 5. Observed counts of rainy days and false alarm ratio (FAR) in the Day-3 forecasts (a–f). Panels (g–l) correspond to moderate rain days and panels (m–r) correspond to heavy rain days over India during monsoon (JJAS) 2020. (Rainy, moderate rain and heavy rain days are defined by 24-hour accumulated rainfall > 1, 15.5 and 64.5 mm/day thresholds, respectively, as per the IMD-defined rainfall categories).](image-url)
rainfall is depicted by rainfall averaged in the longitudes 70°–85°E. From the observation (figure 7a), it is clear that there are few active spells during JJAS 2020. All the five model forecasts indicate a realistic representation of the active/break spells of rainfall activity during the entire season along with northward propagation of rainfall with time. The rainfall biases discussed in earlier sections are again reflected in terms of the intensity of rainfall activity within active monsoon periods. NCUM and UKMO (figure 7b and c) seem realistic in predicting long-lived intense spells of rainfall, while the other three models predict short-lived spells of rainfall with reduced intensity. A prolonged dry spell during the 2nd half of August, especially at lower latitudes, is realistic in all the models.

3.4 Spectral analysis

The rainfall over India exhibits high and low-frequency variations along with synoptic scale variations during the summer monsoon months. The three significant modes of mean monsoon rainfall variability are broadly categorised into (i) synoptic scale (<7 days); (ii) quasi-biweekly scale (QBWO; 10–20 days); (iii) low-frequency intra-seasonal oscillation (ISO; 30–60 days) (Krishnamurti and Bhalme 1976). Synoptic variability includes the monsoon troughs, low-pressure systems, monsoon

<table>
<thead>
<tr>
<th>Sl. no.</th>
<th>Active</th>
<th>Break</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1–5 June 2020</td>
<td>26–28 July 2020</td>
</tr>
<tr>
<td>2</td>
<td>13–18 August 2020</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20–22 August 2020</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>27–30 August 2020</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>21–23 September 2020</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Time series of area-averaged rainfall over core monsoon zone (72°–85°E, 15°–25°N, CMZ) from observations and (a) Day-1, (b) Day-3, and (c) Day-5 forecasts from various models during monsoon 2020. In the figure, the rectangular orange and blue shaded regions indicate break and active spells, respectively. Also note that despite having relatively large peaks in the rainfall forecast time series, these are not highlighted since they did not meet the 1.0 SD level criterion.
depressions, cyclones, offshore vortices, etc. At the same time, the QBWO mode is a westward propagating mode along the monsoon trough with a dominant wavelength of 6000 km and speed of 4.5–6 m s⁻¹ (Krishnamurti and Ardanuy 1980). This mode is also one of the main controls on the active/break cycles of the Indian monsoon system, along with ISO mode. On the other hand, the ISO is a prominent northward propagating mode providing significant rainfall over the south Asian monsoon region. Hence, it is an interesting note about these modes and their amplitude in the model forecasts. Therefore, we showed in figure 8 the periodogram analysis of rainfall over the Indian land region during JJAS 2020 from the observations and different model predictions with 1, 3, and 5 leading days. The observed spectrum from IMD in figure 8 depicts (with x-axis representing the period and y-axis indicating the normalised power associated with each spectral mode) two low-frequency modes (QBWO and ISO) above 90% confidence level.

The synoptic variability, i.e., the peak in 6–10 day spectral band is also clearly seen in the observed spectrum; however, it is not significant at 90% confidence level. Nevertheless, it is important...
to note that in Day-1 forecasts (figure 8a), peak magnitudes are underestimated compared to the observed spectral amplitudes. However, as the forecast lead-time increases, i.e., in Day-3 (figure 8b) and Day-5 (figure 8c) forecasts, the amplitudes of synoptic variability are larger than the observed spectral amplitudes in UKMO and NCUM models with not much change in the amplitudes in other models. Therefore, the predicted amplitude of the synoptic variability is different for different models, except in Day-1 forecasts, where the amplitude is underestimated by all models relative to the observed amplitude. Further, the NCUM model on longer leading times is closer to observed, while the UKMO overestimates the synoptic variance.

On the other hand, the QBWO (peak in 10–20 day spectral band in figure 8) is also clearly noticed in the observed spectrum, which is significant at 90% confidence level. It can be noticed that the QBWO spectral peak is clearly distinguishable in all the models. However, relative to the observed spectrum, the amplitudes in model forecasts are slightly underestimated. Also, the amplitude of QBWO does not change much with forecast-lead times in ECMWF, UKMO, and NCUM models. However, in the GFS and NCEP models, it can be noticed that the amplitude of QBWO decreases as the forecast-lead time increases.

At the ISO time scales (i.e., 30–60 day spectral band in figure 8), the NCUM and UKMO models are close to the observed amplitudes, while all the other models underestimated the ISO spectral amplitudes in Day-1 forecasts. However, at the longer leading times, the NCUM and UKMO models severely underestimate the ISO amplitudes relative to other models. These differences in ISO time scales may also cause significant over- and under-estimations of mean monsoon rainfall as this is the primary mode of Indian summer monsoon variability. One probable reason for this drastic reduction in the spectral amplitudes at the ISO time scale is associated with model initialisation. The spectral analysis applied to the time series constructed based on the Day-1, Day-3 and Day-5 forecasts, where the initialisation is different for every day. Hence, when the initialisation is not at the right phase and amplitude of the ISO, the forecasts also get affected. The spectral analysis here only indicates the mean amplitude of major modes of intraseasonal variability, but a more detailed analysis is required for understanding the differences in the spatial and phase propagation characteristics of these low-frequency modes, which will give a deeper understanding of the biases related to each model.

3.5 Categorical verification

Categorical verification of rainfall forecasts is carried out using the standard approach by constructing a 2×2 contingency table of the counts of

![Figure 8. The spectral analysis of JJAS 2020 All India daily mean rainfall from observations (IMD) and various model forecasts for (a) Day-1, (b) Day-3, and (c) Day-5. The dashed line indicates the 90% confidence level.](image-url)
hits, misses, false alarms and correct negatives. The categorical scores are computed over India as a whole for various thresholds up to 30 mm/day (figure 9). The top panels in figure 9 suggest low values of POD (high FAR) for higher rainfall amounts. ECMWF forecasts indicate the highest POD at all thresholds up to about 15 mm/day and UKMO shows least with all models in between. FAR values for all the models seem very close, with the highest FAR evident in GFS forecasts. The BIAS score (frequency bias) clearly indicates an overestimation of rainfall frequency at all thresholds up to 20 mm/day and underestimation at higher thresholds. NCUM and UKMO indicate marginal underestimation, while NCEP and ECMWF show significant underestimation. ETS, PSS and SEDI metrics indicate relatively higher skill in the ECMWF forecasts, with comparable skill among all the other models.

3.6 Verification over sub-regions

Further, the verification is carried out over six sub-regions (see section 2, table 2 for more details), as shown in figure 1(a). The time series of observed and forecast rainfall averaged over the sub-regions are presented in figure 10. The time series of all model forecasts are in general agreement with observation over MUM, WCC, and NW regions, whereas overestimation of rainfall is noticed over NE, CI and WCS regions. In general, among the six sub-regions, the west coastal areas (MUM, WCS and WCC) and the NE regions experience intense

![Figure 9. Categorical verification of Day-3 rainfall forecasts over India during monsoon (JJAS) 2020.](image-url)
rainfall activity due to the interaction of monsoon winds with orography, whereas in the CI and NW regions, which consist of low orography, rainfall is usually associated with movement of monsoon troughs and low-pressure systems. A close examination indicates that, during June, over the WCS region (figure 10f), NCUM and UKMO models show excessive rainfall amounts. Similarly, over the NE region (figure 10c), almost all the rainfall peaks are overestimated in all five models. For completeness, correlation coefficient (CC) and RMSE have been computed for each sub-regions (figures 11 and 12). The lowest RMSE and highest CC are evident in ECMWF forecasts over WCS, CI and NE, whereas other models show overestimation (figure 11). CC values from all the models are comparable at different lead times, except over WCS (figure 11f), where UKMO, NCEP and ECMWF show higher values after Day-1. Similar conclusions can be drawn from RMSE plots (figure 12). RMSE in ECMWF (NCEP) forecasts is lowest at all lead times over NE (WCS) regions. Thus, while the categorical verification scores discussed in section 3.5 indicate relatively higher forecast skill in ECMWF over the smaller sub-regions, the RMSE and CC analysis do not indicate any clear winner among the models. This clearly suggests challenges in forecasting events over smaller regions. Along with the rainfall verification over sub-regions, we also evaluated rainfall forecasts for hydrological modelling over five river basins (not shown) for flood and water management. We believe this exercise will allow us to assess the performance of NWP model forecasts at a basin scale (refer section S1 of the Supplementary material).

Figure 10. Observed and Day-3 forecast rainfall averaged over each of the six sub-domains over India.
4. Spatial verification of rainfall forecasts

In section 3, comprehensive rainfall verification over the Indian region is presented using conventional techniques like categorical scores and time series analysis. In the following section, we extend our analysis to further highlight the pros and cons of the models used by utilising the spatial verification methods (section 2.2), which will be discussed in the following sections in brief.
4.1 MODE analysis

In the present study, MODE analysis is carried out using a convolution radius (CR) of 2 grid squares (1 grid size $\sim 25$ km). Further, convolution threshold (CT) values of 20 and 100 mm/day are chosen and the results of an inter-comparison of five paired attributes are presented in figure 13. Details of the paired attributes can be found in Appendix A1. Figure 13(a, f) shows symmetric differences for CT of 20 and 100 mm/day, respectively. For 20 mm/day CT, all five models have nearly the same median values, while the 75 centile values indicate lower values of NCEP and ECMWF. However, for a higher CT of 100 mm/day, a difference among the models is evident in median values and 75th centile. Both are lowest in NCEP and highest in NCUM. On a similar note, the centroid distance is comparable among models for lower CT value (figure 13b), but the performance of ECMWF (GFS) models shows a relatively larger (smaller) for 100 mm/day CT value (figure 13g). For area ratio (figure 13c) for 20 mm/day CT, there is little difference among the five models in terms of median as well as 75th centile. However, for 100 mm/day CT, the intermodal differences are quite clear. ECMWF features highest value in area ratio, with GFS showing the lowest median. The angle difference is lowest in NCUM and highest in ECMWF (figure 13i) for 100 mm/day CT.

When it comes to the total interest, all five models have comparable values for lower CT values and for higher CT, UKMO shows relatively large value, followed by NCUM and GFS, and ECMWF shows the lowest median value. Thus it can be understood from MODE analysis that the forecast performance by models is varying with different attributes of the forecast. Based on the total interest (TI), which is an integrated quantitative measure based on all the attributes, UKMO performance is the best among the five models under consideration.

4.2 Fractions skill score (FSS)

As noted in section 2.2.2, FSS is a scale-selective method that can provide a measure of how the forecast skill varies with the chosen spatial scale. As an example, 6th August 2020 extremely heavy rainfall event (>204.5 mm/day) over Mumbai is considered (please see Supplementary section S2 and figure S2a). Spatial maps of observed and Day-1 forecast rainfall over the MUM region are shown in figure 14(a–f). The shaded panels show rainfall excess of 50, 100 and 200 mm/day. As can be seen, like in the observations, forecasts also show large regions of MUM domain covered with rainfall excess of 200 mm/day. However, while NCUM and UKMO have very similar patterns, other three models distinctly feature differing spatial

![Figure 13](image-url) Intercomparison of MODE verification results in the Day-1 rainfall forecasts over India during JJAS 2020. The panels show (a) symmetric difference, (b) centroid distance, (c) area ratio, (d) angle difference, (e) total interest computed for 20 mm/day rainfall objects during JJAS 2020 season. Similarly, bottom panels (f–j) correspond to 100 mm/day objects during JJAS 2020.
FSS computed for this typical case of extremely heavy rain events is shown in figure 15(a–d) for six different rainfall thresholds. The FSS computation results are presented for six spatial scales of 0, 50, 100, 150, 200 and 250 km. The verification is carried out at $0.1^\circ \times 0.1^\circ$ grid resolution using the GPM rainfall data (Huffman et al. 2014). The FSS for the scale of 0 indicates actual grid-to-grid matching without considering any neighbourhood points. It can be seen that for rainfall thresholds (5, 10, 20 and 50 mm/day), the FSS values are >0.5 indicating skillful forecasts event at finest resolution in all models. The magnitude of FSS shows a general decrease with increasing rainfall threshold at all scales. Even for higher threshold of 100 mm/day, except NCEP, all models indicate skilful forecast even at highest resolution. NCEP forecasts show skill at 50 km resolution or so. For 200 mm/day threshold, ECMWF and NCEP show sharp decrease in FSS value. Displaced peak rainfall amounts to the south in NCEP and northeast in ECMWF have resulted in reduced FSS. Similar decay of FSS with increasing rainfall threshold is also evident in Day-3 forecast (please see Supplementary discussion S2 and figure S2b and c).

For the six sub-regions discussed in section 3.6 (and table 2), FSS is computed for all cases of heavy rainfall events listed in table 2. Figure 16 shows the FSS values computed for Day-1 forecasts in all six regions and gives a comparison of FSS among the five models. As can be expected, it is seen from figure 16 that the FSS values for lower rainfall thresholds 20 mm/day (top panels; figure 16a–c) are always higher than 0.5 (skillful forecast) at all spatial levels over WCS, WCC and MUM regions. However, over NW, CI and NE (figure 16d–f), the FSS values are lower than 0.5 at scales near zero (i.e., native grid). For higher rainfall thresholds of 100 mm/day (bottom/panels;
Figure 16. The models mostly show low values (poor skill) of FSS at all spatial scales up to 50 or 100 km. Over MUM, NW and NE (Figure 16i, j and l), the FSS values for NCUM and UKMO forecasts are higher than 0.5, suggesting skilful forecasts over spatial scales of >50 km or so. The performance of other models is marginally lower at all scales. Figure 17 shows the FSS computed for Day-2 forecasts from the different models considered in this study. It can be noted that there is a general drop in the FSS values over all the domains for both 20 and 100 mm/day thresholds. For the 20 mm/day threshold, the FSS values are below 0.5 for all regions and horizontal scales except for WCC, NE and the MUM regions, indicating a better performance of the models over these areas in Day-2. For 100 mm/day threshold, drop in skill in Day-2 compared to Day-1 is remarkable in WCS, WCC, CI and NE.

It is evident from both the spatial verification methods, MODE and FSS, (a) all five models have very comparable skills at lower rainfall thresholds of 20 mm/day. The skill is very low for a higher threshold of 100 mm/day. (b) Among the five models, there is no clear winner among the models when it comes to forecasting heavy rainfall over India during the monsoon season.

5. Summary and conclusions

This study presents verification and intercomparison of rainfall forecasts over India during JJAS 2020 by five operational models, namely NCUM, UKMO, IMD GFS, NCEP GFS and ECMWF. The evaluation is focused on rainfall forecasts over India in terms of skill in predicting rainfall. The exhaustive evaluation presented in the study uses (a) standard traditional verification methods and (b) advanced spatial verification methods. The evaluation includes assessment of large-scale mean patterns, temporal evolution of spells during the season, dominant modes using spectral analysis,
Figure 16. Fractions skill score (FSS) computed for Day-1 forecast rainfall amounts exceeding 20 mm/day (a-f; top panels) and 100 mm/day (g-l; bottom panels) in the five models for each of the six sub-regions over India.
Figure 17: Fractions skill score (FSS) computed for Day-2 forecast rainfall amounts exceeding 20 mm/day (a–f; top panels) and 100 mm/day (g–l; bottom panels) in the Be models for each of the six sub-regions over India.
basin-scale rainfall time series and isolated heavy rainfall cases.

An assessment of mean rainfall patterns and biases (mean error), suggests remarkable similarity in the biases (magnitude and location) among all five models. Large parts of the Bay of Bengal and the Arabian Sea near the Indian west coast prominently feature dry bias in all models. Similarly, over the west coast of India, central and eastern India (core monsoon zone) and NE India show prominent wet bias. All the models consistently overestimate rainy days and heavy rain counts, which is reflected in higher false alarms. Some of the key points are summarised below:

- All the model forecasts indicate a realistic representation of the active/break spells of rainfall activity during the JJAS season along with northward propagation of rainfall. In specific, NCUM and UKMO seem realistic in predicting long-lived intense spells of rainfall, while other models predict short-lived spells of rainfall with reduced intensity. A prolonged dry spell during the 2nd half of August, especially at lower latitudes, is realistic in all the models.

- The amplitude of the synoptic variability is different for different models, except for the Day-1 forecasts, where the amplitude is underestimated by all models relative to the observed amplitude. The NCUM model on longer leading times is closer to observed, while the UKMO overestimates the synoptic variance. Hence, the multi-model assessment of sub-seasonal variability indicates differences in model predictions. These differences may also cause significant over- and under-estimations of mean monsoon rainfall over the Indian region.

- Further, the categorical verification metrics indicate that ECMWF forecasts have the highest POD at all thresholds up to about 15 mm/day and UKMO shows the least with other (NCUM, IMD GFS and NCEP GFS) models in between. The BIAS score (frequency bias) clearly indicates an overestimation of rainfall frequency at all thresholds up to 20 mm/day and underestimation at higher thresholds. ETS, PSS and SEDI metrics indicate relatively improved skill in the ECMWF forecasts, with very comparable skill among all the other models.

However, while the categorical verification scores indicate relatively higher forecast skill in ECMWF, the verification over sub-region does not indicate any clear winner among the models. The time series of all model forecasts are in general agreement with observation over MUM, WCC, NW regions, whereas overestimation of rainfall is noticed over NE, CI and WCS regions. This clearly suggests challenges in forecasting events over smaller regions.

In addition, verification is also presented over five river basins (please refer to Supplementary section S1) to assess the skill at a basin scale. The five river basins are Godavari, Mahanadi, Krishna, Cauvery and Narmada. The Day-3 forecasts seem to successfully capture the events with biases in magnitude. In general, all the models show a good CC (>0.7) in Day-1 except GFS over Cauvery. Among all the models studied, ECMWF shows the highest CC on Day-1 to Day-3 for all the river basins except for Mahanadi in Day-1, where all the models perform equally well and for Narmada on Day-3 where UKMO has outperformed. In higher lead times (Day-4 and Day-5), models have shown mixed behaviour.

- Method of object-based diagnostics evaluation (MODE) analysis, a spatial verification technique, allows verification of different attributes of the rainfall forecasts. MODE analysis shows that the forecast performance by models is varying with different attributes of the forecast. Based on the total interest (TI), which is an integrated quantitative measure based on all the attributes, UKMO performs the best among the five models studied.

- Fractions skill score (FSS), another spatial verification metric, is computed for select cases of heavy rainfall events over the sub-regions, the west coastal areas (MUM, WCS and WCC), and the NE regions. For lower rainfall thresholds 20 mm/day FSS suggests skillful forecasts at all spatial scales. For higher rainfall thresholds of 100 mm/day the models mostly show poor skill for spatial scales up to 50 or 100 km.

Thus it is evident from both the spatial verification methods, MODE and FSS, among the five models there is no clear winner among when it comes to forecasting heavy rainfall over India during the monsoon season. All models indicate very comparable skill for predicting heavy rains, although skill is higher in ECMWF rainfall forecasts for lower rainfall thresholds considered for large areas. This study underscores the role of different verification approaches that bring out
the skill in different aspects of the forecasts. For users, it is important to focus on metrics that are most relevant to the application. It is rather crucial for users not to get carried away based on, say one popular model or based on performance based on a single case or small data sample. Even the models with a very high reputation or good overall scores may not be the best for every application/metric.

**Acknowledgements**

NCMRWF is fully funded by Ministry of Earth Sciences (MoES), Government of India. The real-time NCUM and IMD GFS model analysis and forecast run are being carried out in the MoES High Power Computer (HPC) System at NCMRWF, Noida, India. This study benefited from the TIGGE dataset provided by European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK. The authors thank the two anonymous reviewers for their constructive suggestions and critical but insightful comments, which have helped in improving the quality and content of the manuscript. The authors also thank colleagues at NCMRWF for engaging interactive discussions and constructive feedback.

**Author statement**

Raghavendra Ashrit: Planning and execution, writing, compiling of manuscript analysis of results. Mohan S Thota: Analysis writeup and graphics for monsoon active/break cycles etc. Additional help in revision and proofreading. Anumeha Dube: Fractions skill score (FSS) computation, analysis and writing. Additional help in revision and proofreading. Kondapalli Niranjan Kumar: Computation of spectrum analysis and associated writing. Additional help in revision and proofreading. Karunasagar: Computation of verification scores, graphics and associated analysis and writing. Sushant Kumar: Use of ECMWF data, its analysis, editing of the manuscript. Ashis K Mitra: Advisory, overall planning and guidance along with manuscript editing.

**Appendix**

A1. Method of object-based diagnostics evaluation (MODE)

Object identification forms the first step in MODE analysis. In this step, objects are identified in the accumulated rainfall fields (observed and model forecast). The identification of objects is done by using a smoothing operator, ‘convolution operator’. This is governed by a convolution radius (CR) and a threshold (CT) on the intensity of the field. CR is expressed in terms of grid units and CT in terms of rainfall threshold. This step is necessary (a) to make areas more contiguous and (b) to filter out small/insignificant precipitation amounts. Thus, this step effectively selects the portion of the field that is of greatest interest to the user. There is no universally optimal choice for these parameters (CR & CT). Minimal smoothing and a very low threshold will result in a large number of objects, many of them small. Heavy smoothing and a high threshold will result in very few intense rain areas (Davis et al. 2009). In this study, R is chosen as 2 grid spacing (~20 km) as minimum spatial scale of interest. The observed and forecast rainfall data both are at high grid resolution of 0.1°×0.1°; CR = 2 is considered suitable for smoothing out tiny isolated thunderstorms. CT values indicating rainfall intensity of interest are carefully chosen at 20, 40, 60, 80, 100, and 120 mm to cover low medium and heavy rainfall events. During the monsoon, heavy rains (>80 mm/day) are common over different parts of India.

A2. Fractions skill score (FSS)

Fractions skill score (FSS) belongs to the category of fuzzy or neighborhood spatial verification methods. Like the spatial verification methods, these also do not look for a grid-to-grid match but relax the criteria by verifying forecasts in the local neighbourhood of the observations. FSS is a scale-selective method that can provide a measure of how the forecast skill varies with the chosen spatial scale; in other words, it can help in determining the scales at which the forecasts become useful. It is usually used for the verification of rainfall forecasts (Roberts and Lean 2008; Roberts 2008). A brief description of the methodology is presented here:
Step 1: The forecast probabilities and the corresponding binary observations are generated for a suitable threshold and these are used to generate fractions.

Step 2: For every grid point which has a forecast associated with it, a fraction of surrounding points is computed within a given square of length ‘n’ that has a value of 1 (i.e., where the forecasts from ensemble members have exceeded the threshold). This is presented in the equations below:

\[
O(n)(i, j) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_0 \left[ i + k - 1 - \frac{(n-1)}{2}, j + l - 1 - \frac{(n-1)}{2} \right],
\]

\[
M(n)(i, j) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_M \left[ i + k - 1 - \frac{(n-1)}{2}, j + l - 1 - \frac{(n-1)}{2} \right],
\]

\[
(I = 1, \ldots, N_x \text{ and } j = 1, \ldots, N_y), \quad (A2)
\]

where \(O(n)(i, j)\) is the resultant field of observed fractions for a square of length ‘n’ obtained from the field of binary observation \(I_0\) and \(M(n)(i, j)\) is the resultant field of forecast fractions obtained from binary field \(I_M\). \(i\) and \(j\) represent the number of columns and rows in the domain, respectively. Fractions are generated for different spatial scales by changing the value of ‘n’, which can be an odd number to a maximum of \(2N - 1\), where \(N\) is the number of points along the longest side of the domain.

Step 3: The FSS is then calculated as a variation on the Brier Score (Brier 1950):

\[
\text{FSS} = 1 - \frac{\text{FBS}}{\text{FBS}_{\text{worst}}}, \quad (A3)
\]

where FBS is the Fractions Brier score and is calculated as follows:

\[
\text{FBS} = \frac{1}{N} \sum_{j=1}^{N} (O_j - M_j)^2, \quad (A4)
\]

where \(M_j\) and \(O_j\) are the forecast and observed fractions, respectively, at each point \(j\) and have values between 0 and 1, and \(N\) is the number of pixels in the verification area.

\[
\text{FBS}_{\text{worst}} \quad (A5)
\]

It represents the largest value that can be obtained from the forecast and observed fractions when there is no closeness between the non-zero fractions.

The FSS values range from 0 (for the worst forecast) to 1 for the perfect forecast. FSS has the lowest value when forecasts are verified at the grid-scale only, i.e., the neighbourhood has only a single point. As the size of the neighbourhood increases, the skill also increases until it assumes an asymptotic value (AFSS) when the value of ‘n’ becomes its highest at \(2N - 1\).

References


Davis C A, Brown B G and Bullock R G 2006b Object-based verification of precipitation forecasts. Part II: Application
to convective rain systems; *Mon. Wea. Rev.* **134** 1785–1795.


DelSole T and Shukla J 2002 Linear prediction of Indian monsoon rainfall; *J. Clim.* **15** 3645–3658.


Ebert E E and Gallus W A 2009 Toward a better understanding of the contiguous rain area (CRA) method for spatial forecast verification; *Wea. Forecast.* **24** 1401–1415.


Ramamurthy K 1969 Monsoon of India: Some aspects of the ‘break’ in the Indian southwest monsoon during July and August; Forecasting Manual, India Meteorological Department, Pune, India, 18.3 (No. IV), 1–57.


Corresponding editor: Parthasarathi Mukhopadhyay