NMRF/RR/03/2023

RESEARCH REPORT

Evaluation of Quantitative Precipitation Forecast Performance of NWP models in Indian River Basins

Kondapalli Niranjan Kumar[1], Raghavendra Ashrit[1], Sushant Kumar[1], C. J. Johny[2], Ashok Das[2], Ananda Das[2], B. P. Yadav[2] and Ashis K. Mitra[1]

[1]National Centre for Medium Range Weather Forecasting, Ministry of Earth Sciences, India
[2]India Meteorological Department, Ministry of Earth Sciences, India

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201 309, INDIA

Evaluation of Quantitative Precipitation Forecast Performance of NWP models in Indian River Basins

Kondapalli Niranjan Kumar[1], Raghavendra Ashrit[1], Sushant Kumar[1], C. J. Johny[2], Ashok Das[2], Ananda Das[2], B. P. Yadav[2] and Ashis K. Mitra[1]

[1]National Centre for Medium Range Weather Forecasting, Ministry of Earth Sciences, India

[2]India Meteorological Department, Ministry of Earth Sciences, India

Document Control Data Sheet

1	Name of the Institutes	National Centre for Medium Range Weather Forecasting & India Meteorological Department
2	Document Number	NMRF/RR/03/2023
3	Date of Publication	March 2023
4	Title of the document	Evaluation of Quantitative Precipitation Forecast Performance of NWP models in Indian River Basins
5	Type of the document	Research Report
6	Number of pages, figures, and Tables	33 Pages, 15 Figures, 3 Tables
7	Authors	Kondapalli Niranjan Kumar, Raghavendra Ashrit, Sushant Kumar, C. J. Johny, Ashok Das, Ananda Das, B. P. Yadav and Ashis K. Mitra
8	Originating Unit	National Centre for Medium Range Weather Forecasting (NCMRWF)
9	Abstract (brief)	This report deals with the verification of Quantitative Precipitation Forecast (QPF) across the river basins of India from four different operational models run by the India Meteorological Department (IMD) and National Centre for Medium Range Weather Forecasting (NCMRWF), Ministry of Earth Sciences, Govt. of India (GoI). The operational models, namely, NCMRWF Unified Model (NCUM)-Global (NCUM-G), NCUM-Regional (NCUM-R), Global Forecast System (GFS), and Weather Research and Forecasting (WRF), where GFS and WRF are operational at IMD. IMD issues QPF for river Basins/ sub-Basins through the 'Flood Meteorological Offices' (FMOs), which is the main input to the flood forecasting models through which the Central Water Commission(CWC), Ministry of Jal Shakti, GoI issues flood forecast. The models' skill is verified with respect to the satellite and gauge (SAT+GAUGE) merged dataset at different spatial scales such as FMO regional scale and Basin/Sub-basin scales during the south-west monsoon period of 2021. The SAT+GAUGE dataset is significantly correlated with the ground truth gauge-only data set with a correlation coefficient of 0.9 and a mean bias of <2mm across several river basins of India. The four operational models' skill is assessed based on continuous and categorical skill scores. The comparison of root mean square error (RMSE) and mean bias between four models is more or less similar with estimated values less than 15mm, 10mm, respectively for different FMO regions. The verification is also done for some important rivers such as the Ganga, Brahmaputra, and Mahanadi basins, which are prone to floods during the monsoon period. Though the performance of different models over these basins is slightly different but relatively, NCUM-R outperforms other models for basin-wide average rainfall. All the models are reasonably good in the prediction of light to moderate rainfall categories. However, the detection of extreme rainfall events is relatively better in high-resolution models such as the NCUM-R and WRF over different basins but at the expense of a greater number of false alarms. Nevertheless, the Equitable Threat Score (ETS) indicates the global NCUM also has a better skill for different forecast lead times. Overall, the results demonstrated that forecasts predict observed rainfall reasonably good approximately 5 days in advance and therefore could prove valuable in rainfall prediction at river basin level in India.
10	References	3
11	Security classification	Unrestricted
12	Distribution	General

Table of Contents

Abstract

This report deals with the verification of Quantitative Precipitation Forecast (QPF) across the river basins of India from four different operational models run by the India Meteorological Department (IMD) and National Centre for Medium Range Weather Forecasting (NCMRWF), Ministry of Earth Sciences, Govt. of India (GoI). The operational models, namely, NCMRWF Unified Model (NCUM)-Global (NCUM-G), NCUM-Regional (NCUM-R), Global Forecast System (GFS), and Weather Research and Forecasting (WRF), where NCUM-G and GFS are global models with a spatial resolution of ~12km while the other two NCUM-R and WRF are regional models with a spatial resolution of ~3-4km. IMD issues QPF for river Basins/ sub-Basins through the 'Flood Meteorological Offices' (FMOs), which is the main input to the flood forecasting models through which the Central Water Commission(CWC), Ministry of Jal Shakti, GoI issues flood forecast.

The model's skill is verified with respect to the satellite and gauge (SAT+GAUGE) merged dataset at different spatial scales such as FMO regional scale and Basin/Sub-basin scales during the south-west monsoon period of 2021. The SAT+GAUGE dataset is significantly correlated with the ground truth gauge-only data set with a correlation coefficient of 0.9 and a mean bias of <2mm across several river basins of India. The four operational models' skill is assessed based on continuous and categorical skill scores. The comparison of root mean square error (RMSE) and mean bias between four models is more or less similar with estimated values less than 15mm, 10mm, respectively for different FMO regions. The verification is also done for some important rivers such as the Ganga, Brahmaputra, and Mahanadi basins, which are prone to floods during the monsoon period. Though the performance of different models over these basins is slightly different but relatively, NCUM-R outperforms other models for basin-wide average rainfall. All the models are reasonably good in the prediction of light to moderate rainfall categories. However, the detection of extreme rainfall events is relatively better in high-resolution models such as the NCUM-R and WRF over different basins but at the expense of a greater number of false alarms. Nevertheless, the Equitable Threat Score

(ETS) indicates the global NCUM also has a better skill for different forecast lead times. Overall, the results demonstrated that forecasts predict observed rainfall reasonably good approximately 5 days in advance and therefore could prove valuable in rainfall prediction at river basin level in India.

1. Background

River basins and their tributaries are essential for the livelihood of human beings and living things residing in these basins. The water resources in the river basins are used for various purposes such as irrigation, industry, hydropower, navigation, etc. The river basins of India are primarily classified as the Himalayan, Peninsular, coastal, and inland-drainage basin rivers (Rai et al., 2011). The main river basins of the Himalayas are the Indus, Ganga, and the Brahmaputra. These rivers are primarily snow-fed; however, high average annual rainfall amounts in these river catchments further add to their flow. On the other hand, the peninsular rivers such as the Godavari, Krishna, Cauvery, Narmada, Tapti, Mahanadi, Damodar, etc are rainfed, and hence very little flow during non-monsoon seasons, also immensely fluctuate in volume. The coastal rivers found primarily on the west coast are short consisting of small catchments and episodic due to scant rainfall in drought years. The inland-drainage river basins are very few in numbers and they flow for a very short period. For example, the Sambhar River that vanishes in the desert sands, and the Luni that drains into the Rann of Kutch.

Being blessed with many rivers, it is also quite important to accurately provide hydrological forecasts over all the river catchments specifically for drought and flood information services. Quantitative Precipitation Forecast (QPF) from the numerical weather prediction (NWP) models remains the primary source of rainfall data for input into hydrological forecasting models. However, the performance of flood forecasts from such hydrological models is highly dependent on the accuracy of the rainfall distribution and intensity. The India Meteorological Department (IMD) is the nodal agency for issuing the QPF for river Bains/sub-Basins whereas the Central Water Commission (CWC) is the nodal agency for issuing the flood forecast in India. IMD issues the QPF forecasts through their field offices called 'Flood Meteorological Offices' (FMOs) during the flood season. There are nearly 14 FMOs along with Damodar Valley Corporation (DVC) met service stations throughout the country. Through these FMOs IMD issues the sub-basin-wise QPF on an operational basis daily for the next 5 days (Yadav et al., 2022). Therefore, in this report, we evaluate QPF forecasts for different river basins/sub-basins during the 2021 southwest monsoon, from four different deterministic models running under the umbrella of the Ministry of Earth Sciences (MoES), Govt. of India.

2. Data and Methodology

Four different operational forecast models, namely, NCMRWF Unified Model (NCUM)-Regional, NCUM-Global, Global Forecast System (GFS), and Weather Research and Forecasting (WRF) model have been used in this study. Note that the GFS and WRF model forecasts are from IMD. The accuracy of the model forecast data is assessed against rainfall observations from IMD-NCMRWF (Satellite+Gauge [SAT+GAUGE])) merged dataset (Mitra et al., 2003) having a grid resolution of 0.25° in latitude-longitude direction. However, we also validated SAT+GAUGE dataset from the basin-wide average rainfall from gauge-onlymeasurements (resolution) before its evaluation against model forecasts. Table 1 below indicates the basic configuration details of four operational models evaluated in this report.

Model	**Domain**	**Resolution**	**Grid**	**Forecast Length**
NCUM-G	Global	~12km, Levels 70 (Top: 80 km)	2048 × 1536	10 days (based on 00 UTC initial conditions)
NCUM-R	Regional	~4km, Levels: 80 Top:38.5 km	1200 × 1200	3 days (based on 00 UTC initial conditions)
GFS	Global	~12km, 64 sigma-p layers (Top: 0.27hPa)	T1534	10 days (based on 00 UTC initial conditions)
WRF	Regional	3km, Levels 45 (Top: 50hPa)	1850 X 1950	3 days (based on 00 UTC initial conditions)

Table 1: *Configuration details of various operational models used in this study*

The NCMWF Unified Model (NCUM) is being used for numerical weather prediction (NWP) since 2012. The uniqueness of this modelling system is its seamless modelling approach. The same dynamical core is used for various high-resolution domains at global and regional scales. More details of different model configurations using the NCUM can be found at https://www.ncmrwf.gov.in/. Here we have used two operational models from NCUM, namely, NCUM-Global (NCUM-G) and NCUM-Regional (NCUM-R) models that are run at global and

Indian region, respectively initialized at 00UTC. The other two operational models are run by the India Meteorological Department (IMD), namely, GFS and WRF models. The IMD-GFS model is a spectral global model with a dynamical core of GFS based on Semi-Lagrangian and Semi-implicit time scheme having a reduced Gaussian linear grid. It should be noted here that the initial conditions for IMD-GFS model are generated from the four-dimensional ensemble-variational data assimilations system provided by the NCMRWF. On the other hand, IMD-WRF is run at ~3km resolution with initial and boundary conditions from IMD-GFS. The WRF model forecast data is available after the 6hrs of model run time for the next 3 days. More details about the GFS and WRF model operationalized by IMD can be found at https://mausam.imd.gov.in/

The forecast performance of different operational models at a range of scales compared to the parameter of interest which is the QPF at different FMOs and basin/sub-basin average rainfall. There are 14 different FMOs along with DVC met service stations. The boundaries defined for various FMOs throughout the country by the IMD are depicted in Figure 1. These FMOs issue QPF for different river basins/sub-basins with a lead-time of 7 days but with the validity of sub-basin-wise QPF up to 5 days. Here we evaluated the average rainfall across different scales such as FMO level, basin/sub-basin levels. Figure 1 also indicates the boundaries of different basin/sub-basins. The names of these different basins and sub-basins categorized under each FMO are given in Annexure-I. A detailed account of river basins/sub-basins in respective of FMOs/DVC can be found elsewhere in Yadav et al., 2022.

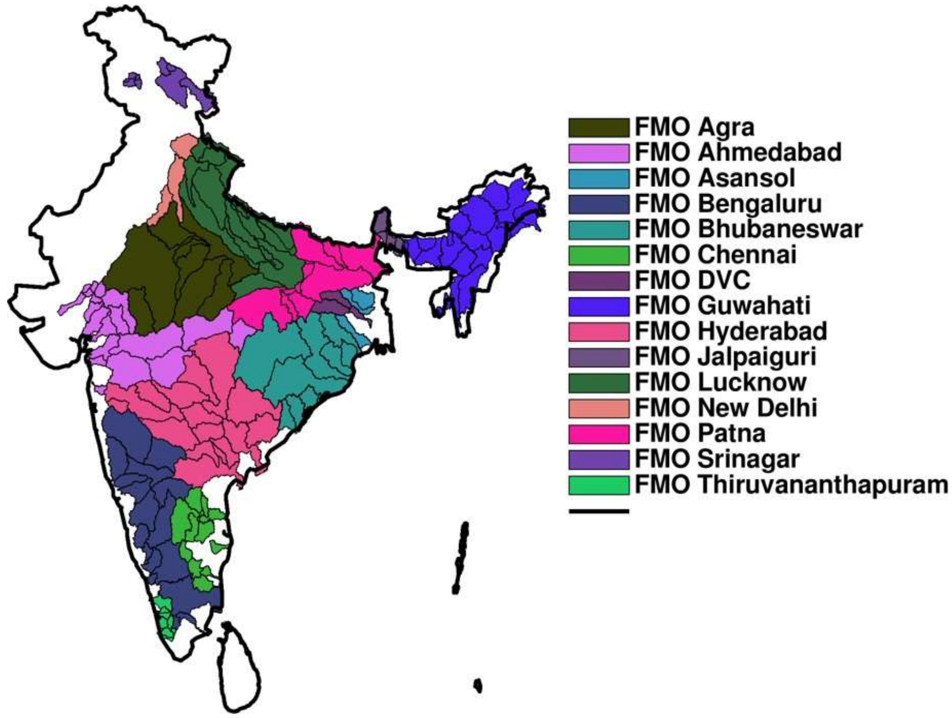

Figure 1*: Map of Flood Meteorological Offices (FMOs) along with different sub-basins based on the data obtained from the India Meteorological Department (IMD) 2021*

2.1 Accuracy Assessment

The forecast performance at a different lead-time compared against the observed rainfall for various spatial means within each FMO. For instance, Figure 2 below indicates the grid locations of SAT+GAUGE and NCUM-G model located in FMO AGRA. The accuracy assessment in this report is done based on the areal average rainfall for the grids within the polygon of different spatial scales (shown as an example for FMO AGRA and other basin/sub-basins in Figure 2). Here it is also important to note that we did not gridded the model forecast data to match the grid locations of SAT+GAUGE grids. Therefore, the accuracy assessment is done based on grid average rainfall under different polygons of FMO-scale, basin/sub-basins. The forecast outputs are assessed based on different accuracy measures including continuous and categorical (dichotomous) verification scores.

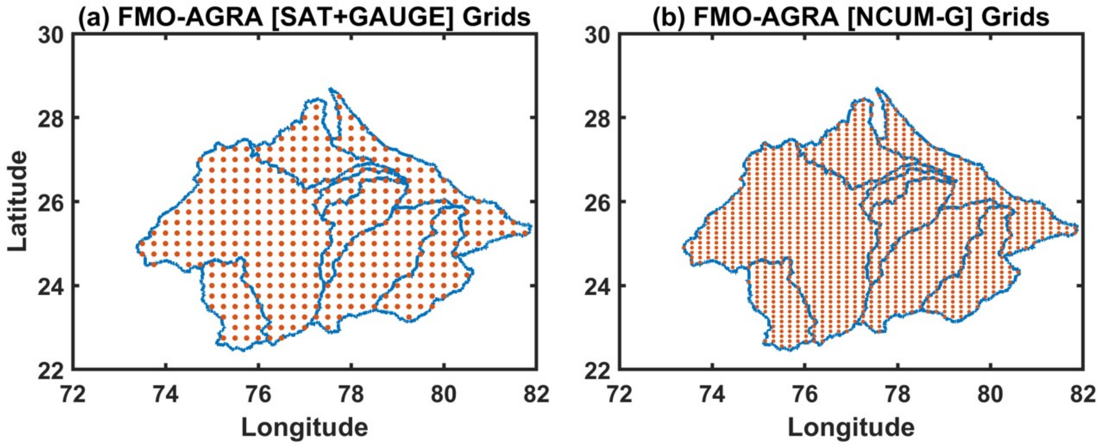

Figure 2: *(a) Grid locations of SAT+GAUGE rainfall data under different basins/sub-basins for FMO AGRA. (b) same as (a) but for the NCUM-G model grids*

2.1.1 Continuous statistics

Four different continuous accuracy metrics namely, Bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson's correlation were used for accuracy assessment. Bias is defined as the average error which can be estimated from simple statistic given in Eqn. 1. Bias, however, is an inadequate measure of skill since negative errors can compensate for positive errors. A simple way to avoid the compensation of positive and negative forecast errors is to consider the MAE defined in Eqn. 2. Another widely used forecast score is the RMSE also defined based on simple statistic defined in Eqn. 3. Another important score in weather and climate forecasting is the Pearson's correlation coefficient (ρ) because of its invariance properties and is defined in Eqn. 4.

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i) \qquad Eqn. 1$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_i - x_i| \qquad Eqn. 2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2} \qquad Eqn. 3$$

$$\rho = cor(x, \hat{x}) = \frac{cov(x,\hat{x})}{\sqrt{var(x)var(\hat{x})}} \qquad Eqn. 4$$

Where $\hat{x}_i$ is the forecast value and x_i is the observed value at time i. $cov(x, \hat{x})$ is the covariance between the observations and forecasts and $\boldsymbol{var}(\boldsymbol{x})$ = $\boldsymbol{cov}(\boldsymbol{x},\boldsymbol{x})$ and $\boldsymbol{var}(\widehat{\boldsymbol{x}})$= $\boldsymbol{cov}(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{x}})$ are the variances of the observations and forecasts, respectively. The covariance can be estimated from the sample of past forecasts and observations, e.g., $\boldsymbol{cov}(\boldsymbol{x},\widehat{\boldsymbol{x}})$ defined by

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_i)(\widehat{x_i} - \bar{\widehat{x_i}})$$

2.1.2 Categorical statistics

Deterministic forecasts are often interpreted as probabilistic forecasts having only probabilities of 0 and 1 (i.e., no uncertainty), yet it is more realistic to interpret them as probabilistic forecasts in which the uncertainty is not provided. The verification statistics for different rainfall thresholds are formulated using a contingency table shown in Table 2.

		Event Observed	
Event Forecast		Yes	No
	Yes	A (hits)	B (false alarms)
	No	C (misses)	D (correct rejections)

Table 2: *Example of a 2 × 2 contingency table (showing the number of times that the event occurred in each category)*

Many types of skill scores can be devised to adequately describe the forecast performance. Three different important categorical scores such as Probability of Detection (POD), False-Alarm Rate (FAR), and Equitable Threat Score (ETS) are evaluated in this study for assessing the forecast quality of different operational models. POD is the proportion of events when a forecast for a threshold exceedance was correct which is defined in Eqn. 5. On the otherhand, FAR is defined as the proportion of forecasts when a forecast for a threshold exceedance was incorrect (Eqn. 6). The other important skill score is the ETS (Eqn. 7) commonly used for the verification of deterministic forecasts of rare events (e.g., rainfall amounts above a threshold). This ETS is defined as the number of hits that would have been obtained purely by chance. It is also sometimes referred to as Gillbert's skill score.

$$POD = \frac{A}{A + C} \qquad Eqn. 5$$

$$FAR = \frac{B}{B + D} \qquad Eqn. 6$$

$$ETS = \frac{A - A_R}{A - A_R + B + C} \qquad Eqn. 7$$

where $A_R = (A + B)(A + C)/N$ is the number of hits expected for forecasts independent of observations (pure chance) and $N = A + B + C + D$.

The above categorical scores are evaluated at different rainfall thresholds as per the operational QPF issued at sub-basin wise as an average areal precipitation forecast by the FMOs daily in the following categories (Yadav et al., 2022)

i. **0 (No Rain)**

ii. **0.1 – 10mm**

iii. **11 – 25mm**

iv. **26 – 50mm**

v. **51-100mm**

vi. **>100mm**

Hence, for all the above rainfall thresholds using the continency table defined in Table 2, we have estimated the different skill scores for different lead times. The sub-basin-wise areal rainfall has been computed within the polygons (boundaries) at different basin/sub-basins (Figure 1) from the SAT+GAUGE and operational model forecasts defined in Table 1.

3. Results and Discussions

3.1 Evaluation of SAT+GAUGE merged rainfall product

In this section, we validate the merged (gridded) rainfall product (SAT+GAUGE) with rain gauge measurements. The gauge-only rainfall data obtained as sub-basin-wise areal rainfall computed from the daily station-wise rainfall data by using the isohyetal technique (Yadav et al., 2022), hereafter, referred to as AAP (Areal Average Product). Nevertheless, we primarily used the merged rainfall product for the assessment of rainfall in different basins. However, before the verification of model forecasts, we discuss the differences in the AAP and merged rainfall product. The comparison between these two rainfall products is done on the areal average of basin/sub-basin-wide daily rainfall under different FMOs discussed in section 2.1. Figure 3(a) shows the box plot of daily mean rainfall over different FMOs from AAP and SAT+GAUGE rainfall products during JJAS2021. On each box plot, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The outliers are plotted individually using the '+' symbol. The outliers are identified based on the Inter Quantile Range (IQR) threshold, where IQR is defined by the difference between the 75th percentile (Q75) and 25th percentile (Q25). The outliers are then defined by greater than Q75+1.5*IQR and less than Q25-1.5*IQR. It is interesting to note that despite the low resolution of SAT+GAUGE data, the distribution of rainfall across different basins is well compared with the AAP data. However, few extreme rainfall events are missed in some basins when comparing both datasets. For instance, in the case of FMO Ahmedabad and FMO DVC, extreme rainfall events with a magnitude more than 200 mm/day are noted in AAP data which are missed in the SAT+GAUGE data. The reason is obvious as the extremerainfall events sometimes smoothed out during the merging process relative to gauge-only measurements. However, it can also be noted from Figure 3a that some extreme rainfall events of more than 150 mm/day were also missed in AAP data but are observed by SAT+GAUGE data. This is probably related to the low density of stations across the river basins which is also one of the drawbacks of the gauge-only observations.

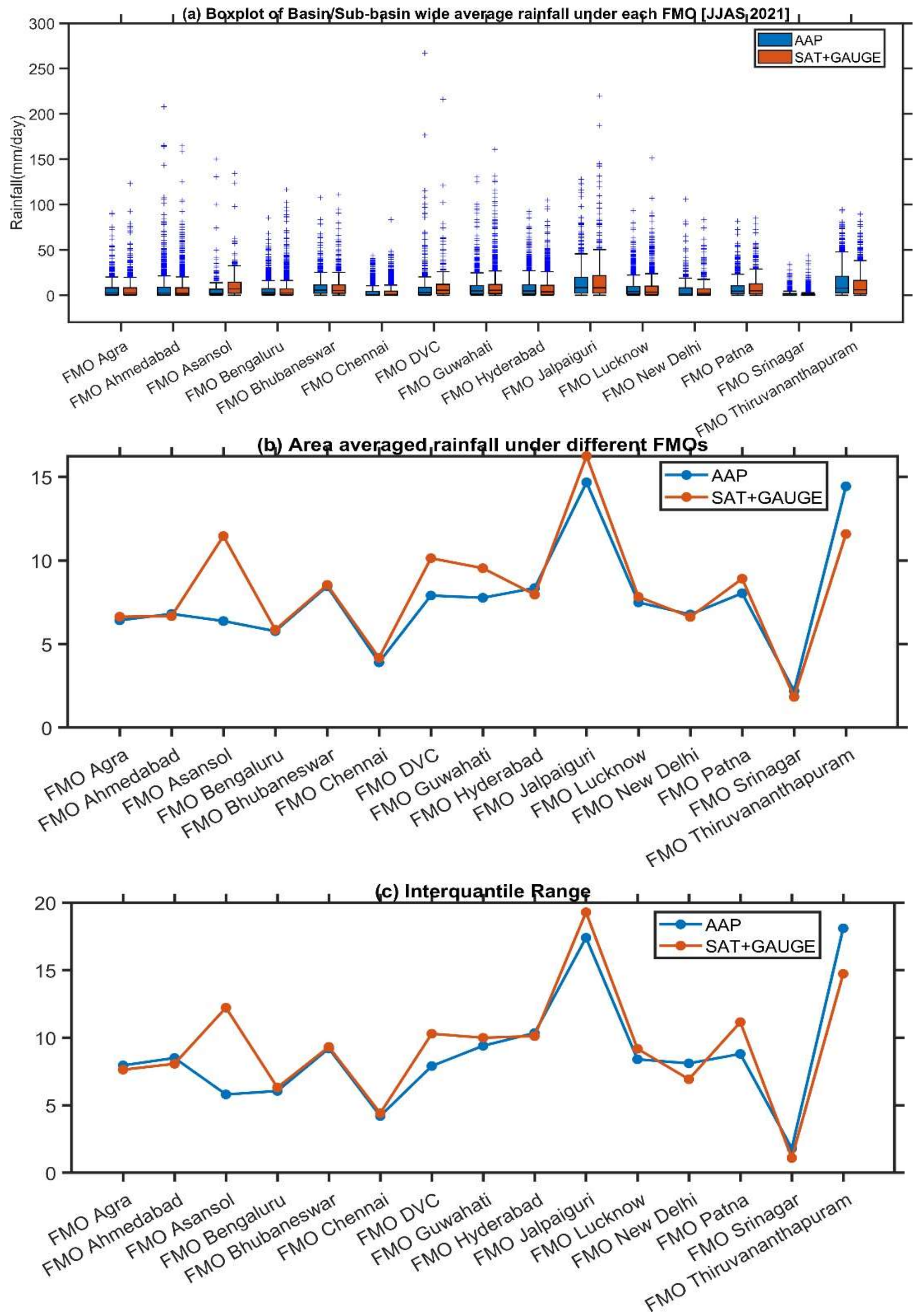

Figure 3: (a)*Box plot (refer to text for more details) of basin/sub-basin wide averaged rainfall under different FMOs (b) Area averaged rainfall under different FMOs and (c) same as (b) except the Inter quantile range of rainfall associated with basin/sub-basin wide rainfall noted over FMOs*

We also showed the area average rainfall for each FMO from both datasets in Figure 3b. The rainfall SAT+GAUGE data is quite well compared with the AAP data except for FMO Asansol (Figure 1) region where an approximately 5mm/day difference is noted. The underestimation of rainfall in AAP data is most probably due to low-density rain gauges in that basin. We also showed the IQR comparison for both these datasets in Figure 3c. Similar to Figure 3b, the IQR also indicates a good correlation between the merged and AAP datasets except for FMO Asansol. The prime reason for using the SAT+GAUGE product for the model forecast verification is due to the uncertainty of the measurements including the low gauge density network across different basins, risk of corruption, or missing values in the reporting process in AAP (gauge-only) data.

A further validation between the AAP and SAT+GAUGE datasets is shown in Figure 4, where the day-to-day variability of rainfall during JJAS 2021 for 4 different FMOs covering north, south, central, and eastern parts of India. Figure 4a shows the temporal evolution of rainfall for FMO Agra from both datasets. The rainfall variability from the two different datasets is matching well with a correlation of 0.96. The extreme rainfall event in the first week of August 2021 is also picked up in both datasets. The error quantification statistics (discussed in section 2.2) between two datasets is also indicated in Figure 4a. The rainfall evolution for the basins in FMO Bengaluru is shown in Figure 4b. The day-to-day rainfall is matching well between two datasets having a mean bias of 0.21 and an RMSE of 3.3 for JJAS 2021. Figure 4c shows the daily rainfall from AAP and SAT+GAUGE datasets for FMO Bhubaneswar. The correlation coefficient between the two datasets is 0.96 having amean bias of 0.09. Similarly, the rainfall time series for FMO Guwahati is shown in Figure 4d from both datasets. The mean bias is slightly higher for this region as the basins are located in the north- eastern part of India where the gauge density might be lower relative to other regions. Nevertheless, the day-to-day variability of basin-wide average rainfall is matching quite well with an RMSE value of ~3.5. Though the error statistics between the merged rainfall and AAP are slightly different, which is not significant, for different basins/sub-basins under different FMOs, the sub-seasonal variability is matching well between these two datasets. Hence, we use the SAT+GAUGE dataset for further verification of model forecasts in the following sections.

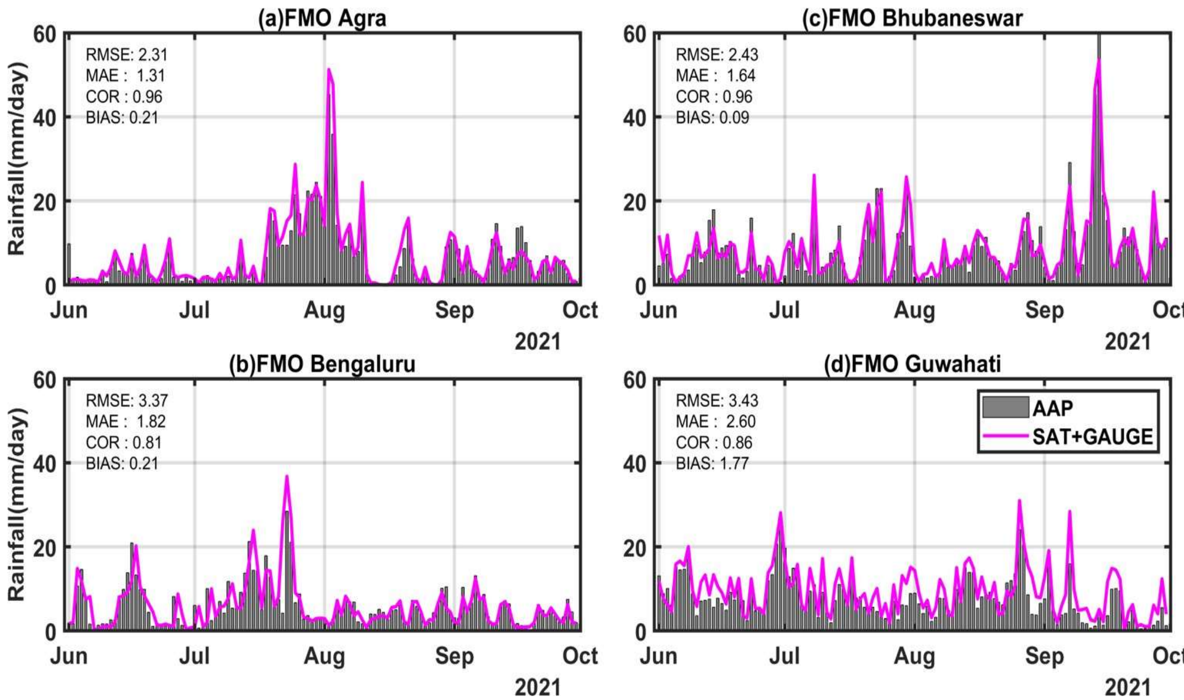

Figure 4: *Day to day variability of average rainfall during JJAS 2021 over (a) FMO Agra, (b) FMO Bengaluru, (c) FMO Bhubaneswar, and (d) FMO Guwahati*

3.2 QPF verification across different Basin/sub-basins of India

As discussed in section 2, we have verified four different operational models for which two are global (NCUM-G and GFS) and two are regional (NCUM-R and WRF) with more details listed in Table-1. We first show the temporal evolution of average rainfall across different basin/sub-basins under those FMOs (shown in Figure 4) from SAT+GAUGE data and different operational models for Day-1 to Day-3 forecast lead-times. For instance, Figure 5 below shows the time series of rainfall for FMO Agra from four operational models with the *left panel* indicating the NCUM-R and NCUM-G models and the *right panel* for GFS and WRF models. The rainfall evolution from the NCUM global and regional models is mostly similar in the Day-1 forecast. The continuous rainfall during mid-July to mid-August is nicely predicted in both models. However, as the lead-time increases, the NCUM-G indicates overestimation while the regional NCUM model indicates underestimation of heavy rainfall during the first week of August. Also, noted a few false alarms during September in the NCUM models. On the other hand, both the GFS and WRF models also forecasted the continuous rainfall from mid-July to mid-August. As the lead-time increases, the magnitudes of forecasted rainfall diminished in these two models. Further, the WRF models indicate high amount of rainfall relative to observed as well as GFS model, specifically in Day-1 and Day-2 lead-times.

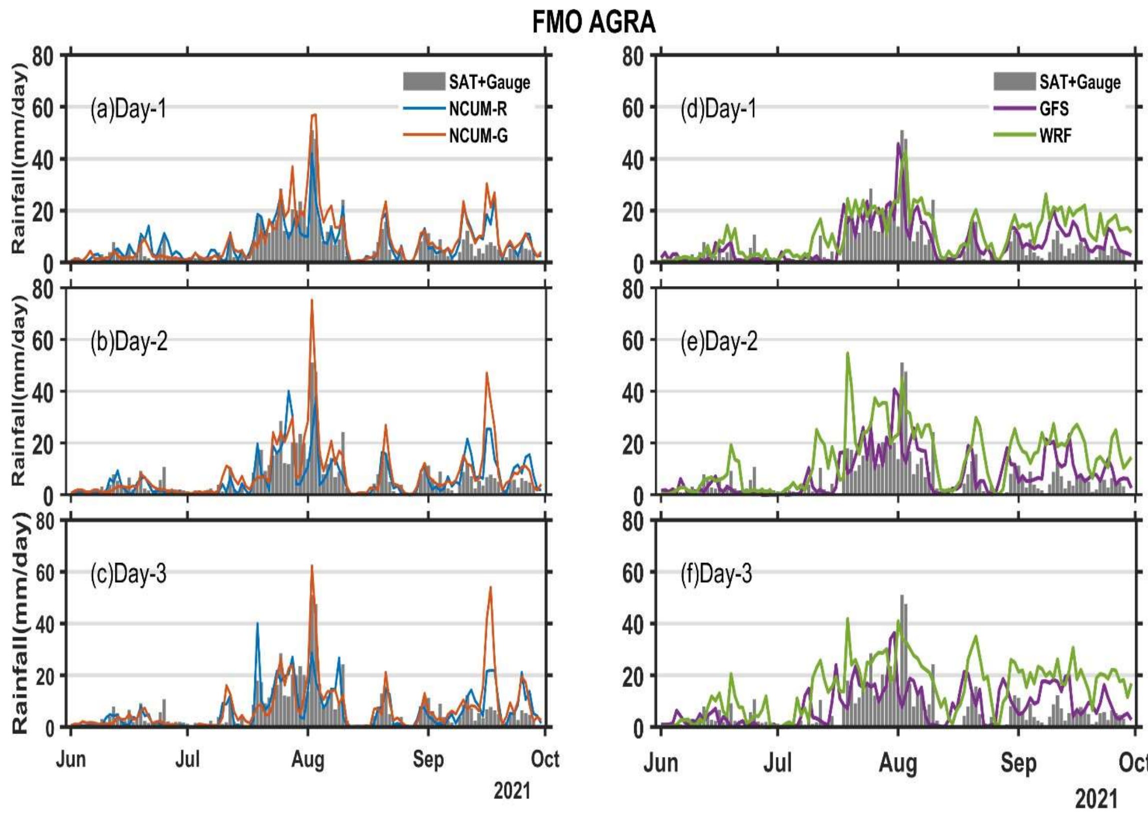

Figure 5 ***(a-c)***: *The daily mean time series of rainfall over FMO AGRA from observed and model forecasts (NCUM-Global and NCUM-Regional) during JJAS 2021. (**d-f**: right panel) Same as (a-c) but from GFS and WRF models.*

The time series of rainfall for the FMO Bengaluru region is shown in Figure 6. Similar to Figure 5, the *left panel* indicates the rainfall forecast from NCUM-G and NCUM-R models and the *right panel* indicates from GFS and WRF models along with SAT+GAUGE observed rainfall data. It can be noticed from Figure 6 that all models in all forecast lead times show lesser rainfall relative to

SAT+GAUGE data, except the NCUM-R model. However, NCUM-R is able to predict the high rainfall amounts but at the same time, the false alarms are also higher.

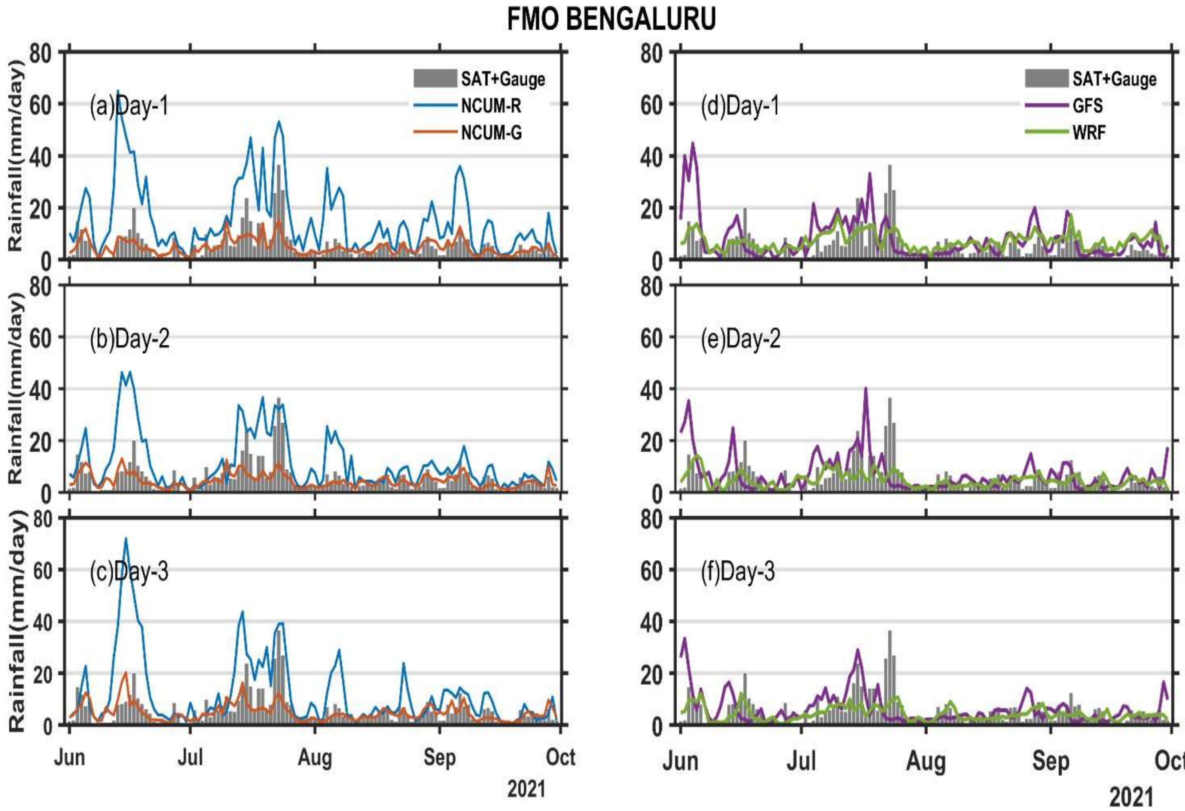

Figure 6: *Same as Figure 5 but for FMO Bengaluru.*

Similarly, the temporal evolution of mean rainfall across several basins under the FMO Bhubaneswar is shown in Figure 7. The day-to-day variability of rainfall evolution is predicted well in both NCUM-G and NCUM-R, except the magnitudes are higher relative to observed rainfall. The heavy rainfall events in the mid-September in all forecast-lead times were predicted in both NCUM models (Figure 7, *left panel*). However, this event is missed both GFS and WRF models in Day-2 and Day-3 forecasts. It is also important to note the overestimation in WRF model forecasts when the rainfall amounts are less the 20mm in Day-2 and Day-3 forecasts. At the same time, Day-1 forecasts from both GFS and WRF models show better predictability of rainfall evolution over FMO Bhubaneswar.

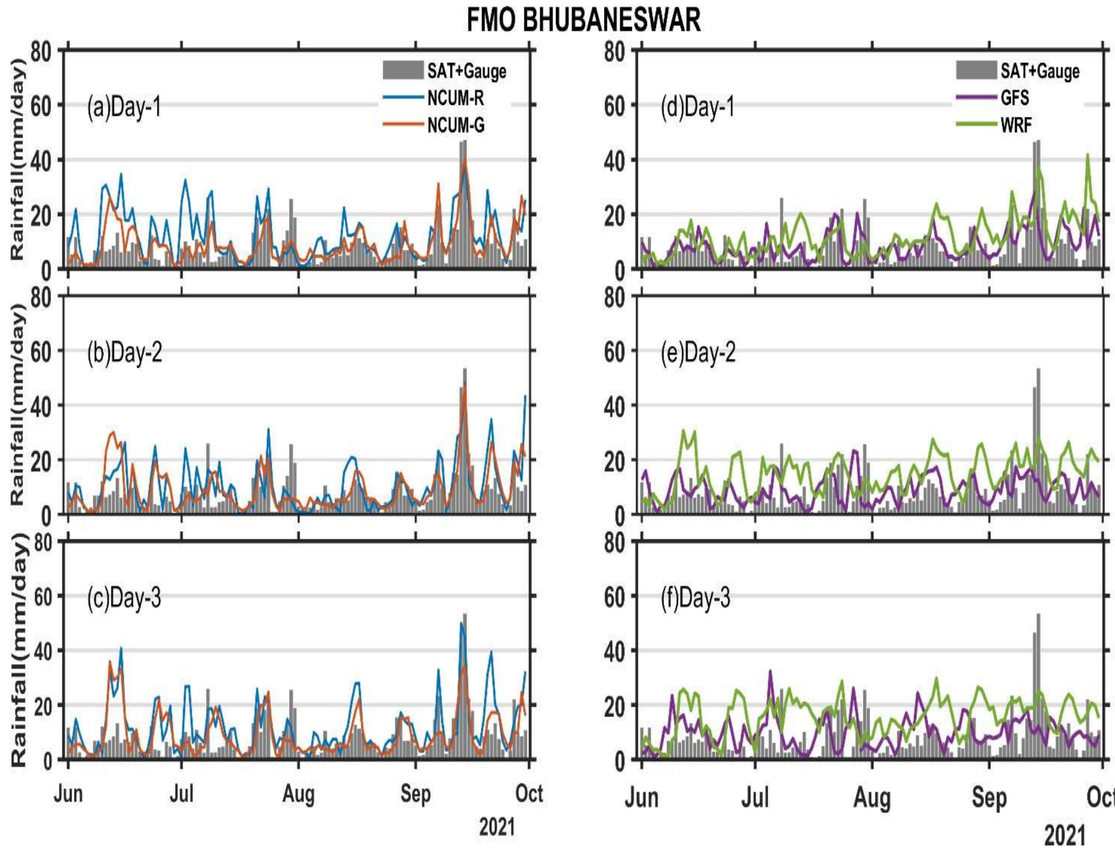

Figure 7: *Same as Figure 4 but for FMO Bhubaneswar*.

The time series of rainfall evolution over FMO Guwahati is shown in Figure 8, similar to Figure 5. Here, both the NCUM models (Figure 8, left panel) indicate an overestimation of rainfall, specifically in Day-2 and Day-3 lead-times, however, in Day-1 forecasts rainfall is predicted well. In the case of GFS and WRF model predictions, the day-to-day rainfall variability is relatively better than NCUM for the FMO Guwahati. We can notice, the overestimation in all the models, particularly in NCUM models than GFS and WRF at longer lead times. Nevertheless, the model predictions are quite different in different regions and lead times. It is also quite hard to say which model's performance is better in which region. Hence, we quantified these errors at the FMO scale, i.e., mean rainfall of basin/sub-basins under each FMO, using the continuous verification scores discussed in Section 2.2

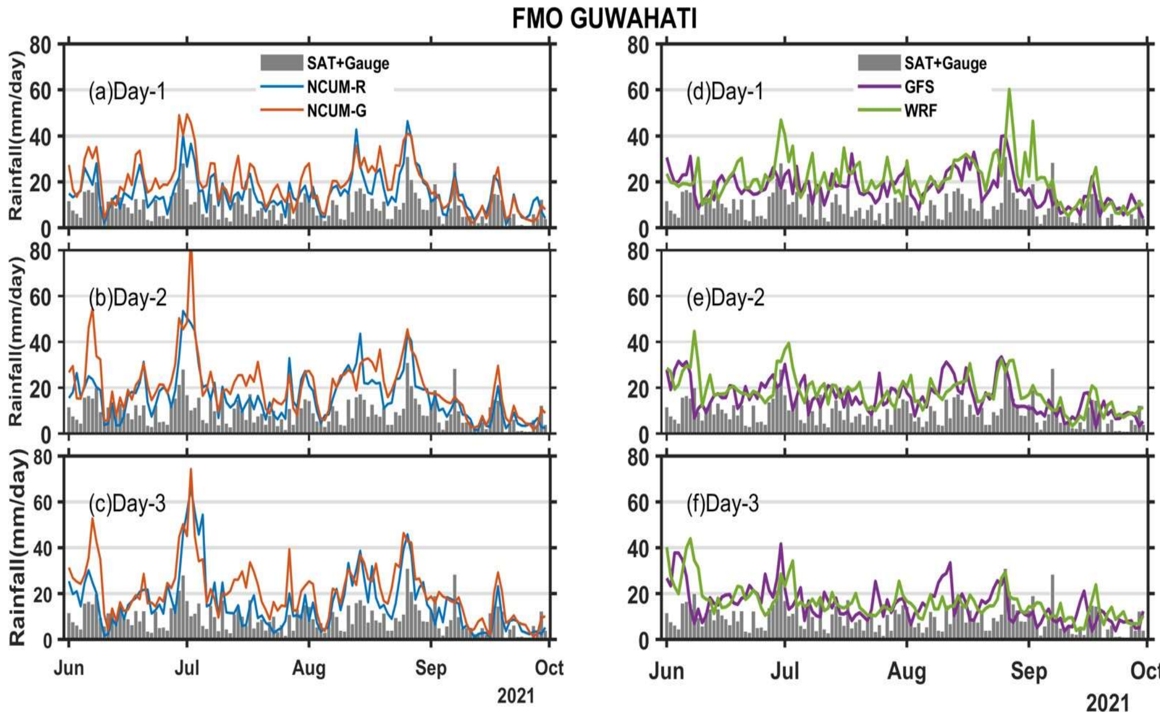

Figure 8: *Same as Figure 4 but for FMO Guwahati.*

Figure 9 shows the two verification scores, namely, RMSE (*left panel*) and MAE (*right panel*) defined in section 2.2, estimated for different forecast-lead times from the four operational models with reference to SAT+GAUGE data. It is interesting to note that the variability of the magnitudes of RMSE at most of the FMOs are similar, except at FMO Asansol where the RMSE is higher in NCUM-R relative to other models. The magnitude of RMSE is below 15mm/day at many FMO with magnitudes increasing from Day-1 to Day-3 forecasts. Overall, the RMSE values are higher in WRF operational model at many FMOs relative to other models at different forecast-lead times. Similarly, the magnitude of MAE is also higher in the WRF model in most of the FMOs relative to other models. Figure 10 shows the two other verification scores such as the correlation coefficient and mean bias estimated based on the definitions discussed in section 2.-2.

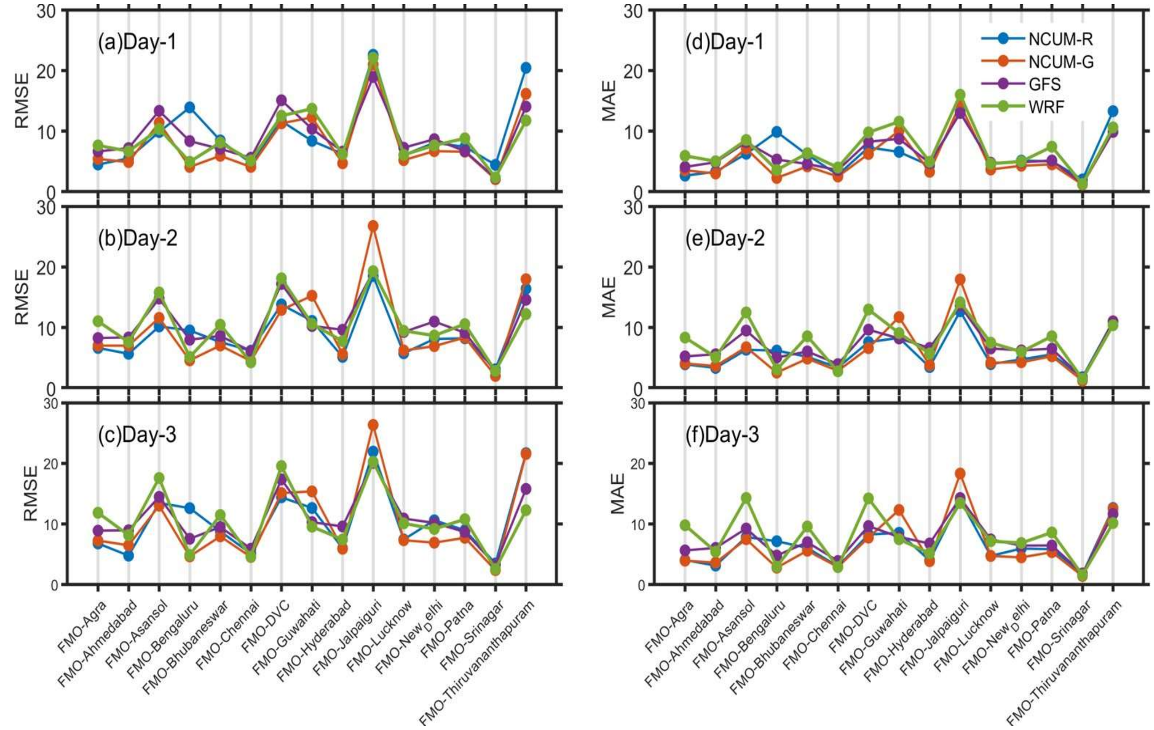

Figure 9: ***(left panel)*** *Root Mean Square Error (RMSE) estimated from model forecasts with respect to SAT+GAUGE data for (a) Day-1, (b) Day-2, and (c) Day-3 lead times.* ***(right panel)*** *Same as (a-c) but for Mean Absolute Error (MAE) for all the FMOs shown in Figure 1.*

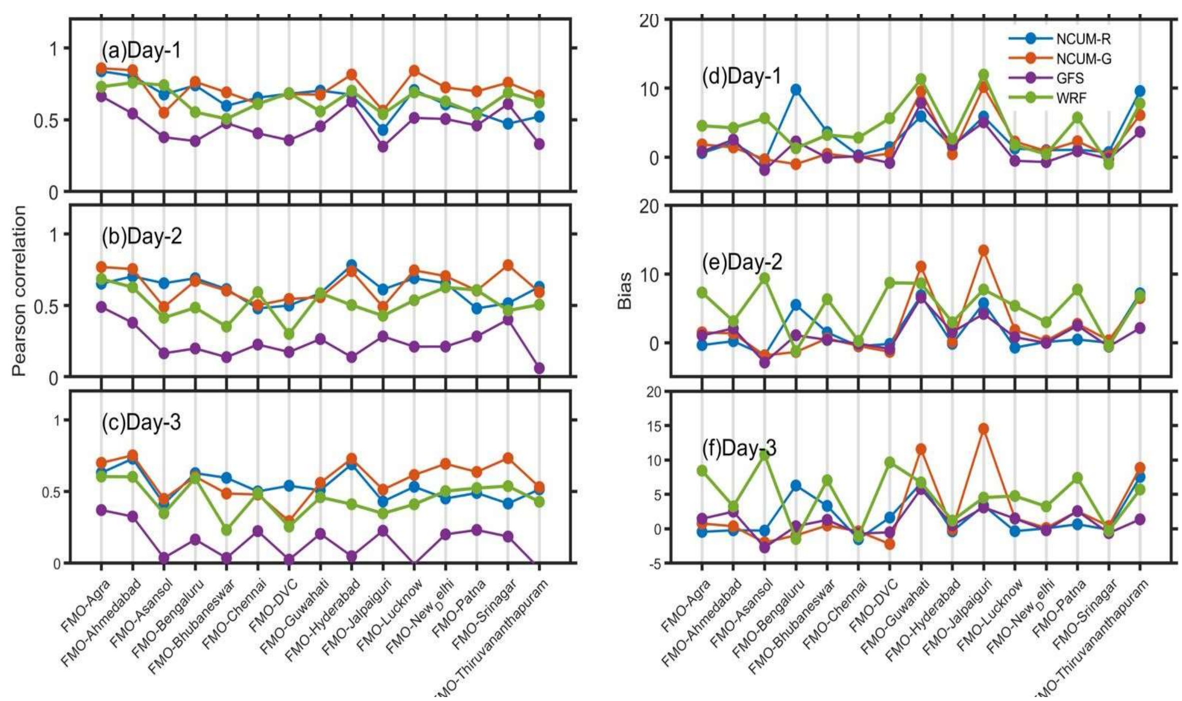

Figure 10: ***(left panel)*** *Correlation Coefficient estimated between model forecasts and SAT+GAUGE data for (a) Day-1, (b) Day-2, and (c) Day-3 lead times.* ***(right panel)*** *Same as (a-c) but for Bias for all the FMOs shown in Figure 1.*

The correlation coefficient (Figure 10, *left panel*) is higher in Day-1 forecasts and decreases as the lead-time increases in all the models. Also, it is important to note the GFS model indicates the lower correlation relative to all other models at all lead times. Both the NCUM-G and NCUM-R indicate a relatively better correlation of rainfall across different FMOs than GFS and WRF models. Further, the bias score is shown in Figure 10, *right panel* for Day-1 to Day-3 (top to bottom) forecasts lead times. Again, the WRF model indicates a higher bias across many FMOs relative to other operational models. The mean bias is relatively better in NCUM-R and GFS models along with NCUM-G except for FMO Guwahati and FMO Jalpaiguri where the bias in NCUM-G is relatively higher. The performance of all the operational models in FMO Jalpaiguri is relatively poor at all forecast lead times compared to all other FMOs. Overall, the performance of NCUM operational models is relatively better when we look at various scores across various FMOs in Figure 9 and Figure 10.

Previously, we have verified model forecasts at a spatial scale of FMO (Figure 1) by averaging rainfall over various basins/sub-basins with each FMO. Here, we also evaluated model forecasts on a basin scale by selecting some important large basins that are prone to floods during the summer monsoon. Figure 11A shows the map of the Ganga basin located under FMO Lucknow. There were three sub-basins within the large Ganga basin shown in different colours (Figure 11A). The time series of the average rainfall of all the three sub-basins of Ganga is shown in Figure 11B at different forecast lead times. Figure 11B (left panel) indicates the rainfall from NCUM model forecasts along with observed data. The observed rainfall shows several extreme rainfall episodes of more than 20mm/day basin-wide average rainfall. The NCUM models predicted these events very well, except that the rainfall intensity is higher than the observed rainfall, particularly in NCUM-G. Also, it is important to note that the false alarms are also higher over the Ganga basin in the NCUM-R model, especially, in the Day-3 forecast.

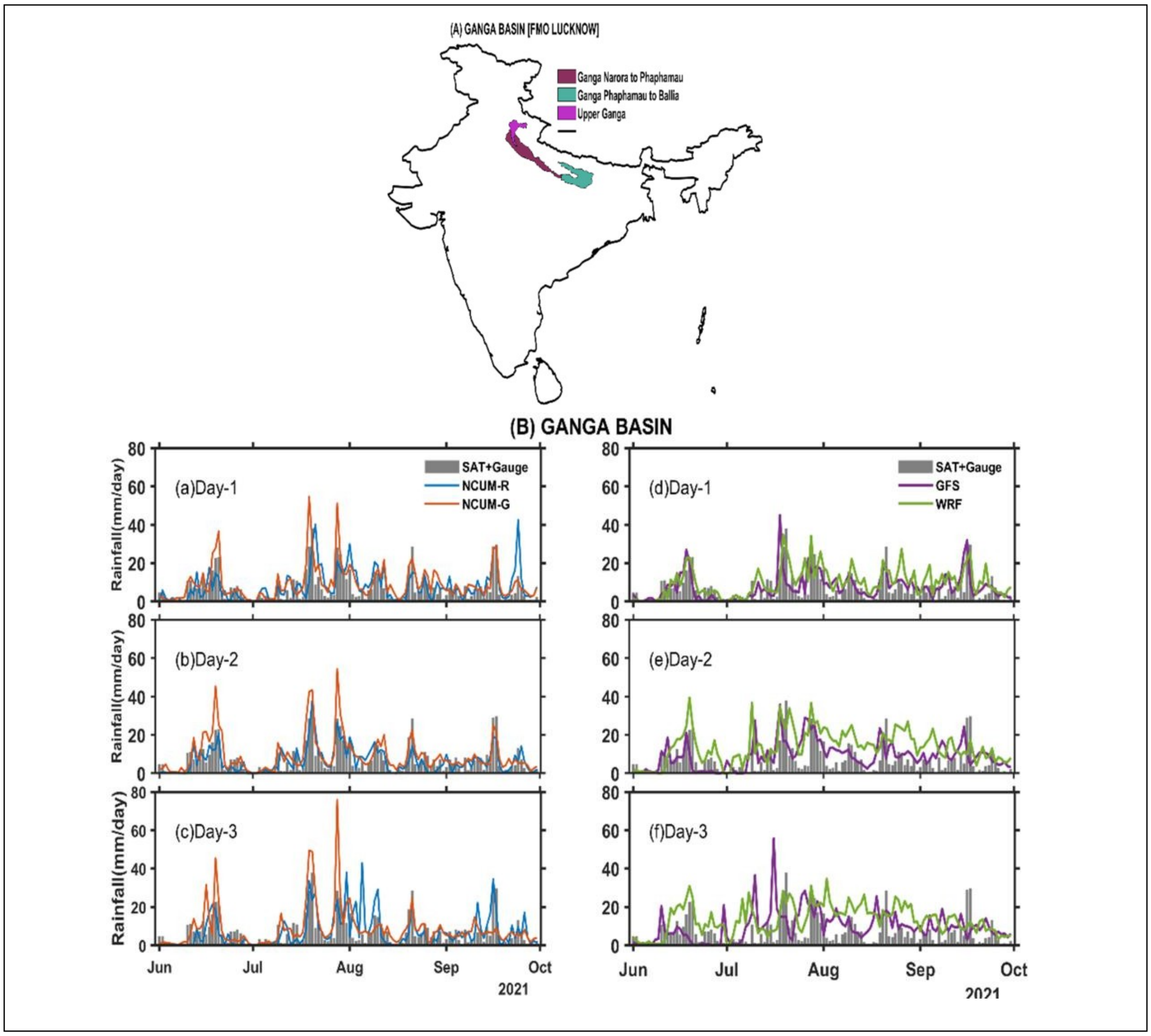

Figure 11*: (A) Map of different sub-basins of Ganga River under FMO Lucknow. **(B) (a-c)** temporal evolution of rainfall from SAT+GAUGE data and NCUM-G and NCUM-R model forecasts for Day-1 to Day-3 lead times. (d-f) same as (a-c) but from GFS and WRF models*

The rainfall time-series from GFS and WRF models forecasts are shown in Figure 11B (*right panel*). Both the GFS and WRF models also predicted rainfall events of more than 200mm/day, however, the WRF model shows many false alarms than GFS and NCUM, particularly, at longer lead times. Further, we showed the continuous verification scores discussed previously to quantify the uncertainty in the model forecasts over the Ganga basin in Figure 12. The RMSE is shown in Figure 12a with different lines indicating four operational models. It is obvious to see that the RMSE is increasing with increasing lead times in all the models, except in the NCUM-R model where Day-2 forecasts are relatively better than Day-1. Over the Ganga basin, the RMSE shows that the NCUM-G model performance is relatively better than the other models at all forecast lead times. Similar to RMSE, the MAE (Figure 12b) also indicates that the NCUM-G model performs well as the magnitudes of MAE are less relative to all other models at various lead times. Further, the bias score (Figure 12c) shows that the NCUM-R has a lesser mean bias than other models, while NCUM-G shows almost constant bias in all the lead times. The GFS and WRF models show that biases are larger at longer lead times compared to NCUM models. The correlation coefficient is shown in Figure 12d, which indicates the higher correlation of the NCUM-G model at all lead times relative to other models followed by NCUM-R. The GFS model particularly shows a drastic reduction of correlation from Day-3 forecasts. Nevertheless, all models indicate a decreasing correlation with increasing lead times. Therefore, Figure 12 indicates, the performance, of NCUM-G is better compared to all other models based on the continuous verification scores.

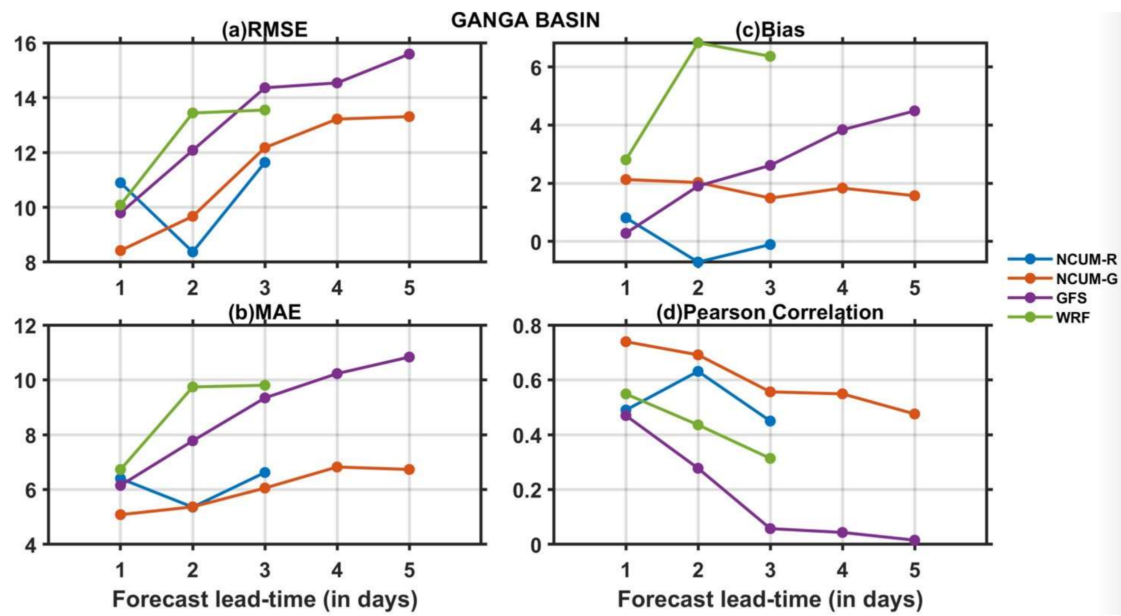

Figure 12: *(a) RMSE, (b)MAE, (c)Bias, and (d) Correlation Coefficient estimated for 4 different models with respect to SAT+GAUGE data for different forecast lead-timings over the Ganga Basin.*

Further, Figure 13A shows the map of different sub-basins in the Brahmaputra basin under the FMO Guwahati. There were 6 different sub-basins in the Brahmaputra basin shown in different colours in Figure 13A. Of all the rivers in India, the Brahmaputra has the greatest volume of water because of heavy annual rainfall in its basin. At Dibrugarh, the annual rainfall averages 2,800 mm, and at Shillong, it averages 2,430 mm. The verification scores are shown in Figure 13B. In contrast to the Ganga basin, the NCUM-G model has relatively higher RMSE, MAE and mean bias over the Brahmaputra basin at all forecast lead times. Interestingly, the co-variability is quite good in the NCUM-G model relative to other models (Figure 13B.d). Overall, the NCUM-R and GFS are shown good performance over the Brahmaputra basin as revealed from Figure 13B.

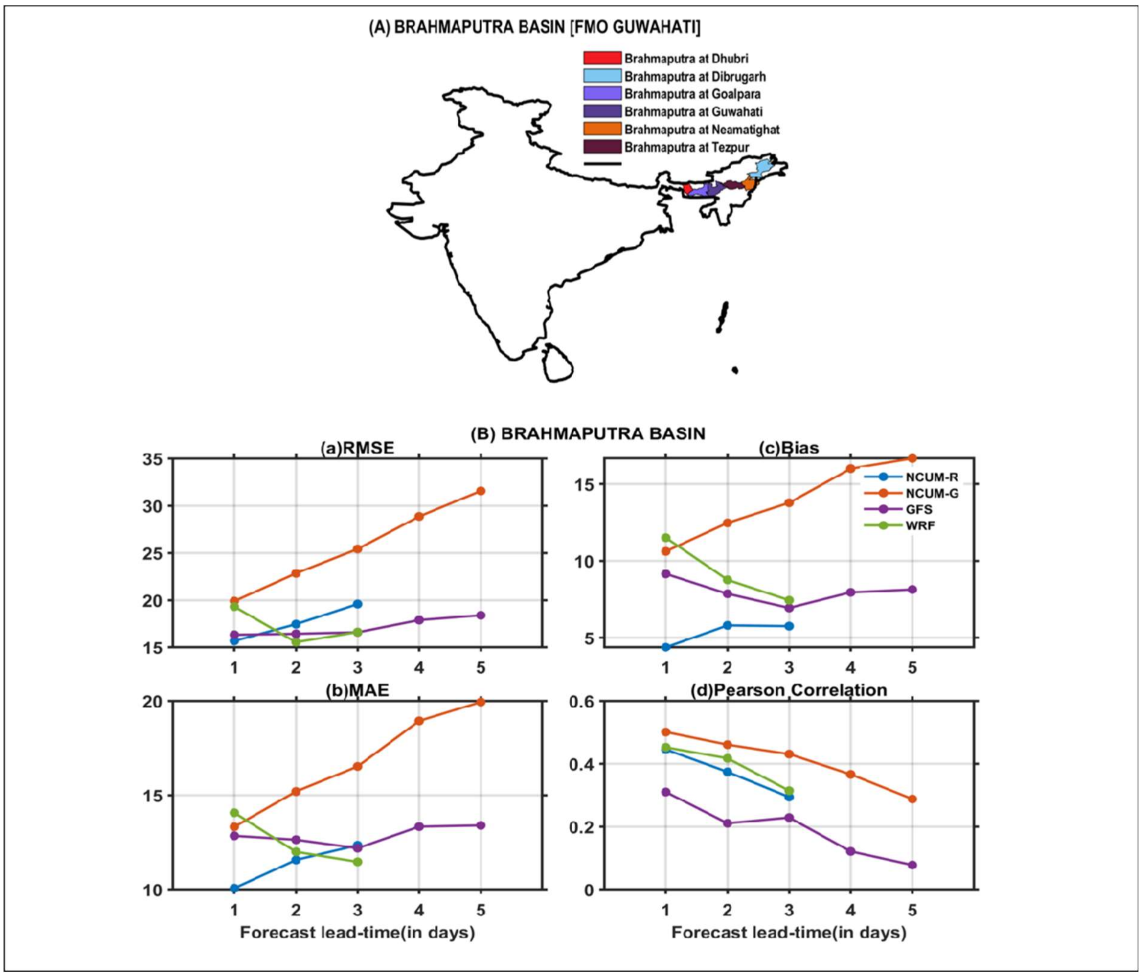

Figure 13: ***(A)*** *Map of different sub-basins of Brahmaputra River under FMO Guwahati.* ***(B)*** *(a) RMSE, (b)MAE, (c)Bias, and (d) Correlation Coefficient estimated for 4 different models with respect to SAT+GAUGE data for different forecast lead-timings.*

We have also shown the models' performance in the Mahanadi basin (Figure 14B) with the map of different sub-basins in Figure 14A. There were two sub-basins within the Mahanadi basin shown in different colours. The RMSE, MAE, bias, and correlation plots shown in Figure 14B indicates that the NCUM-R model is performing very well relative to other models with the scores increasing with increasing lead days, except the correlation where it decreases with increasing lead days. On the other hand, the WRF model shows higher RMSE, MAE, and bias than other operational models. Therefore, the performance of different models at different basins is slightly different; however, the performance of high-resolution model such as NCUM-R is relatively better on basin-wide average rainfall.

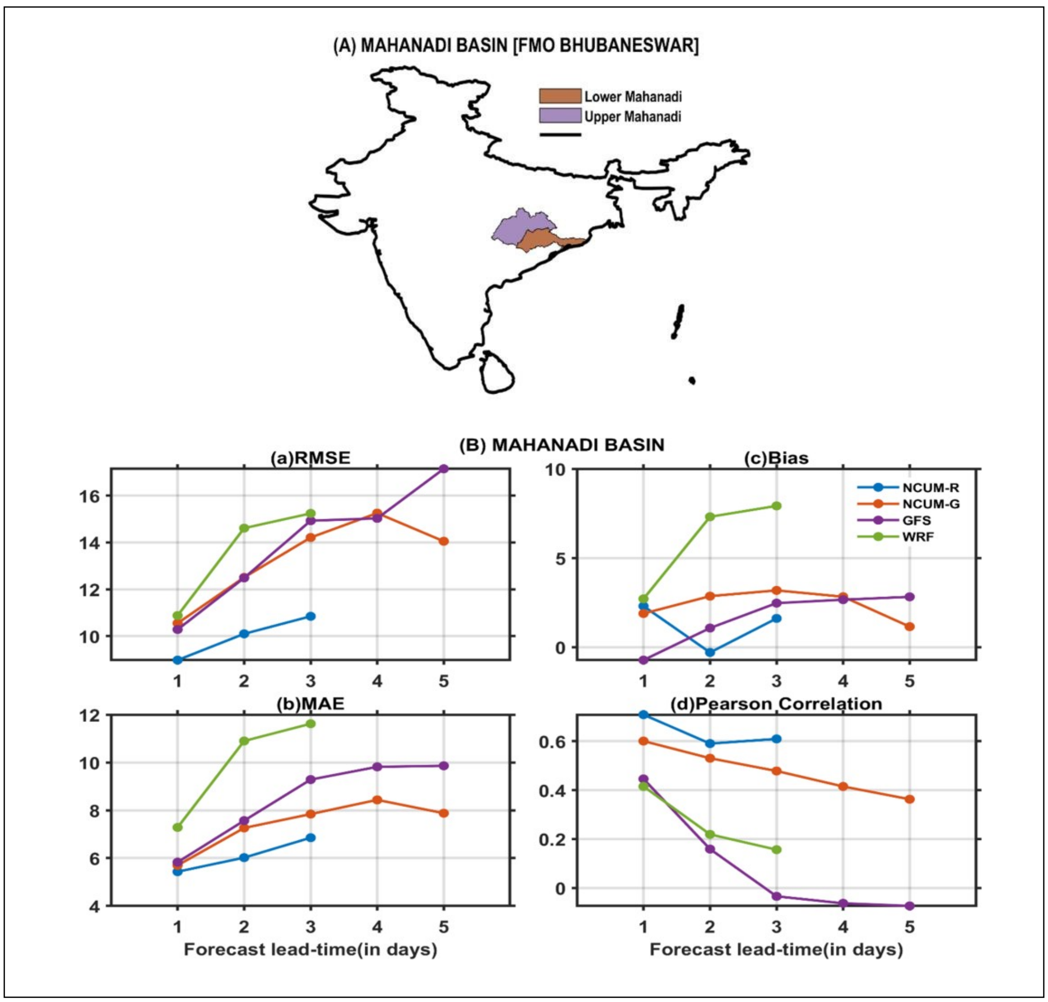

Figure 14: ***(A)*** *Map of different sub-basins of Mahanadi Basin under FMO Bhubaneswar.* ***(B)*** *(a) RMSE, (b)MAE, (c)Bias, and (d) Correlation Coefficient estimated for 4 different models with respect to SAT+GAUGE data for different forecast lead-timings*.

In all the previous discussions, we discussed models' skills based on the continuous verification scores. However, one of the disadvantages of these scores is that it is average performance within the period we have chosen across all the rainfall categories. However, for the case of river basins, it is also important to understand the verification based on the different rainfall categories as extreme rainfall amounts cause the flooding. IMD operational QPF is issued sub- basin-wise as an average areal precipitation forecast by the FMOs daily during the season in different rainfall categories discussed in section 2.3. Hence, we also verified the models' performance in those categories.

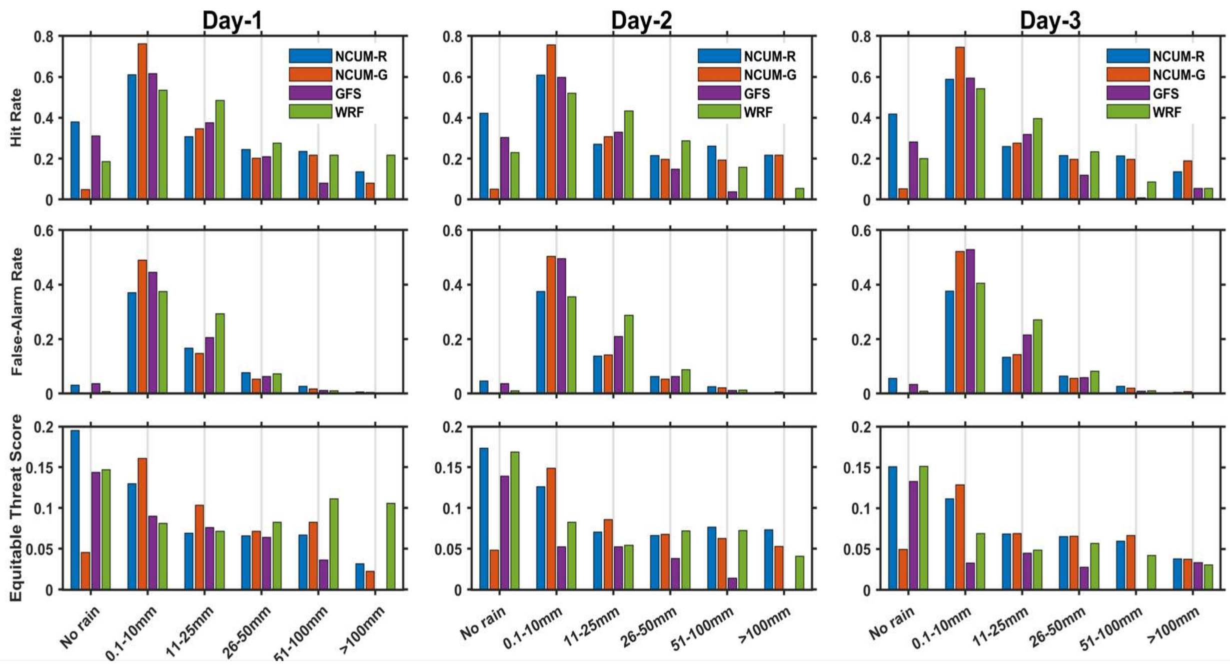

Figure 15: (left panel) *Categorical skill scores Hit rate, False-Alarm Rate and Equitable Threat Score estimated from different models with respect to SAT+GAUGE data for Day-1 forecast.* ***(middle panel)*** *Day-2 and (**right panel) Day-3** forecasts for all the FMO regions shown in Figure 1*

Figure 15 above indicates the four operational models' skill for different forecast lead times (*left to right*). Three different scores defined in section 2.3, such as Hit rate, FAR, and ETS are shown in Figure 15 (*top to bottom*), respectively. Figure 15 (*left panel*) indicates the model skill for Day-1 forecasts indicating the Hit rate between 0.1-10mm (light rainfall category) is higher in all the model forecasts following the 11-25mm category (Figure 15, *top panel*). At the same time, the number of false alarms is also higher in those categories (Figure 15, *middle panel*). In the heavy to very heavy rainfall category (e.g., >50mm) skill is good in both the NCUM and WRF model forecasts. The models' performance is similar for the Day-2 and Day-3 forecasts, except that the skill is better in NCUM models relative to GFS and WRF models in heavy (70-110mm) to very heavy rainfall (120-200m) categories that are defined based on IMD (Yadav et al., 2022). Further, Figure 15 (*bottom panel*) shows the ETS which measures the fraction of observed and/or forecast events that were correctly predicted, and not by any random chance in the case of a number of hits. The model skill is perfect when the ETS value is 1. In the case of light to moderate rainfall categories, the performance of the NCUM is model is better than other models. However, in the heavy to very heavy rainfall categories, the WRF model has also a better skill along with NCUM models. The skill of the model is deteriorating as the lead time increases. Figure 15 provides the model skill across all the FMOs defined in Figure 1.

However, it is also interesting to see about performance at different FMOs. Hence, we have estimated the above categorical scores for Day-1 forecasts over various FMO regions and tabulated. Table-3 indicates the categorical scores of four operational models for the various FMOs. The Hit rate is particularly high for various basins in the WRF model relative to other models but at the expense of a large fraction of false alarms. On the other hand, both the NCUM models are relatively good when considering the HIR and FAR. Similarly, the ETS also indicates the performance of NCUM models, particularly, the skill of NCUM-G is relatively better across different FMOs compared to GFS and WRF models.

FMO	HIR				FAR				ETS			
	NCUM-Regional	NCUM-Global	GFS	WRF	NCUM-Regional	NCUM-Global	GFS	WRF	NCUM-Regional	NCUM-Global	GFS	WRF
Agra	0.61	0.67	0.64	0.86	0.15	0.14	0.17	0.36	0.27	0.33	0.27	0.2
Ahmedabad	0.61	0.62	0.61	0.8	0.17	0.13	0.26	0.35	0.25	0.3	0.17	0.19
Asansol	0.59	0.63	0.52	0.85	0.2	0.21	0.26	0.67	0.24	0.27	0.15	0.09
Bengaluru	0.77	0.39	0.44	0.54	0.33	0.05	0.19	0.17	0.17	0.25	0.12	0.19
Bhubaneswar	0.69	0.54	0.48	0.69	0.31	0.15	0.23	0.45	0.2	0.25	0.13	0.11
Chennai	0.4	0.24	0.28	0.65	0.1	0.07	0.1	0.23	0.17	0.11	0.1	0.17
DVC	0.62	0.6	0.52	0.77	0.29	0.22	0.27	0.67	0.18	0.23	0.14	0.04
Guwahati	0.68	0.81	0.85	0.93	0.42	0.55	0.66	0.71	0.12	0.11	0.07	0.08
Hyderabad	0.6	0.55	0.6	0.76	0.23	0.14	0.28	0.35	0.21	0.27	0.17	0.2
Jalpaiguri	0.72	0.88	0.86	0.89	0.45	0.51	0.61	0.63	0.16	0.22	0.14	0.14
Lucknow	0.56	0.68	0.47	0.61	0.23	0.17	0.16	0.27	0.18	0.32	0.18	0.17
New Delhi	0.56	0.56	0.51	0.57	0.15	0.16	0.13	0.19	0.24	0.23	0.23	0.21
Patna	0.56	0.71	0.61	0.88	0.24	0.2	0.31	0.54	0.19	0.33	0.17	0.15
Srinagar	0.33	0.45	0.22	0.08	0.05	0.03	0.03	0	0.15	0.27	0.12	0.07
Thiruvananthapuram	0.79	0.74	0.64	0.84	0.36	0.33	0.41	0.7	0.24	0.24	0.12	0.06

Table 3: *Categorical skill scores calculated between individual FMOs region averaged rainfall and corresponding SAT+GAUGE data for Day-1 forecasts of 4 different models.*

4. Summary

In this report, we have verified QPF across various river basins from four different operational models (NCUM-G, NCUM-R, GFS, WRF) for the summer monsoon season (JJAS) 2021. Floods are more frequent during the monsoon season, especially, over the basins located in the central and northern parts of India. Hence, it is very essential to see the performance of different operational models across various basins of India. IMD through its FMOs issues the QPF forecasts during the monsoon season, which is the main input in the Flood Forecasting models for issuing flood forecasts by the CWC. The highlights of this report are

- ✓ We have used the IMD-NCMRWF (referred here as SAT+GAUGE) merged rainfall data set for verifying different operational models. However, before using this product, we have verified this SAT+GAUGE product with the AAP which is based on rain gauges.

- ✓ The verification is done based on the time series analysis along with continuous and categorical skill scores to assess the accuracy of multiple operational models. Further, we have verified the model forecasts based on different spatial scales such as FMO regions (defined in Figure 1) and basin-wide average rainfall.
- ✓ The verification scores such as the RMSE, MAE and mean biasfrom four operational models indicates that the NCUM models are relatively better than the other two operational models in the 2021 monsoon season.
- ✓ The verification was also done over the scale of basin-wide average rainfall for some important rivers such as the Ganga, Brahmaputra, and Mahanadi, which are prone to floods during the monsoon period. Though the performance of different models over these basins is slightly different but overall NCUM-R outperforms other models on basin-wide average rainfall.
- ✓ The models' skill is further assessed through the categorical verification scores for various rainfall categories defined by the IMD. All the models are reasonably good in the prediction of light to moderate rainfall categories. However, during the extreme rainfall cases, the high-resolution models such as NCUM-R and WRF have better predictability at the expense of a higher number of false alarms. Nevertheless, the ETS indicates the global NCUM also have better skill for different forecast lead times.
- ✓ Among the four operational models, the NCUM models have relatively better skills. However, the high-resolution WRF model has also better skill in heavy to very heavy rainfall categories. Nevertheless, the skill of these raw operational model forecasts can be further improved across different river basins through some bias-correction procedures which will be taken up in our future works.

5. References

1. Rai R.K., Upadhyay Alka, C. S. P. Ojha, V. P. Singh, (2011), *The Yamuna River basin: water resources and environment. Water science and technology series*, vol 66, Springer, The Netherlands

2. Yadav, B. P., Ashok Kr. Das, Charu, Rahul Saxena, S. K. Manik, Asok Raja, Hemlata Bharwani, A. Sravani, H. R. Biswas, Manickam Rajavel, Manorama Mohanty, K. Santhosh, Kuldeep Srivastava, Manish Rai, N. Puviarasan, Ranendra Sarkar, Sanjay Kumar Singh, Sonam Lotus, Sourish Bondyopadhyay, Sunit Das, Uthan Kumar Ghatak (2022), *Verification of Sub-basin-wise Qunatitative Precipitation Forecast during SW Monsoon 2021*, Met. Monograph, **MoES/IMD/HS/Basin Hydrology/ 01(2022)/14**.

3. Mitra, A. K., M. D. Gupta, S. V. Singh, & T. N. Krishnamurti (2003), Daily Rainfall for the Indian Monsoon Region from Merged Satellite and Rain Gauge Values: Large-Scale Analysis from Real-Time Data, **Journal of Hydrometeorology**, 4(5), 769-781.

ANNEXURE-I

The Basin/Sub-basins categorized under different Flood Meteorological Offices (FMOs) based on the India Meteorological Department (IMD) 2021

FMO AGRA	• Betwa, Gambhir,Ken_Urmal,Kunwari,Lower_Chambal, Sind, Upper_Chambal, Yamuna_Mathura_to_Naini
FMO AHMEDABAD	• Banas, Damanganga, Lower_Tapi, Lower_Mahi, Lower_Narmada, Middle_Mahi, Middle_Narmada, Middle_Tapi, Narmada Hoshangabad to Sardar Sarovar, Sabarmati, Upper_Mahi, Upper_Narmada, Upper_Tapi
FMO ASANSOL	• Ajoy, Kangsabati, Mayurakshi
FMO BENGALURU	• Bennehalla, Ghataprabha, Hagari_or_Vedavati, Harangi, Hemavathy, Kabini, LowerCauvery, LowerVaigai, LowerBhima, Lower_Tungabhadra, Middle_Cauvery, Middle_Krishna, Middle_Tungabhadra, Uper_Vaigai, Upper_Bhima, Upper_Cauvery, Upper_Krishna, Upper_Tungabhadra
FMO BHUBANESWAR	• Baitarni, Brahmani, Burhabalang, Lower_Mahanadi, Nagavali, Rushikulya, Subarnarekha, Upper_Mahanadi, Vamsadhara
FMO CHENNAI	• Cheyyeru, Gummanur, Korttalaiyar, Kunderu, Lower_Pennar, Lower_South_Pennar, Papagni, Sagileru, Upper_Pennar, Upper_South_Pennar, Vellar
FMO DVC	• Barakar, Damodar, Damodar_Lower_Valley
FMO GUWAHATI	• Badarpurghat, Barak at Silchar, Brahmaputra at Dhubri, Brahmaputra at Dibrugar, Brahmaputra at Goalpara, Brahmaputra at Guwahati, Brahmaputra at Neamatighat, Brahmaputra at Tezpur, Buridihing_at_Khowang, Dehung at Passighat, Dhansiri at Golaghat, Dhansiri at Rly_Bridge, Gumti, Jiabharali at NT road Xin, Kapili at Kampur, Lohit at Dholla, Manas Beki at NH Xing, Manu, Sankosh, Subansiri at Badatighat
FMO HYDERABAD	• Indravati, Lower_Godavari, Lower_Krishna, Maneru, Manjra, Middle_Godavari, Munneru, Musi, Palleru, Penganga, Pravara, Purna, Sabari, Upper_Godavari, Wainganga_Pranhita, Wardha
FMO JALPAIGURI	• Jaldhaka, Lower_Teesta, Raidak, Torsa, Upper_Teesta
FMO LUCKNOW	• Alaknanda, Bhagirathi, Chhatnag_to_Mirzapur, Ganga_Narora_to_Phapha, au, Ganga_Phaphamau_to_Bal, ia, Gomti, Lower_Ghaghra, Middle_Ghaghra, Ramganga, Rapti, Sai, Sharda, Upper_Ganga, Upper_Ghaghara
FMO NEW DELHI	• Sahibi, Yamuna_upto_Hathnikund, Yamuna_upto_Mathura
FMO PATNA	• Bagmati_Adhwara, Gandak, Kanhar, Kosi__Mahananda, North_Koel, Punpun_Dhab_Nadi, Sone, Upper_Sone
FMO SRINAGAR	• Dah, Khalsi, Lidder, Lower_Jhelum, Middle_Jhelum, Nimmo, Upper_Jhelum, Upshi_Road_Bridge
FMO THIRUVANANTHAPURAM	• Achankoil, Bharathapuzha, Chalakudi, LowerPeriyar, Meenachil, Pamba, Periyar, UpperPeriyar