1. **Introduction**

We live in an Era where severe weather events are becoming common, and precise and high-resolution forecasting is not just a scientific challenge but a societal need. Coarse-resolution data sets play the most essential role in predicting the future; however, scale discrepancy limits the coarse-resolution datasets from being directly used for impact assessments and decision-making (WANG et al. 2004), nevertheless, in this age where accurate predictions are imperative rather than optional. The need for finer-resolution data sets is ever-increasing. When going through the Intergovernmental Panel on Climate Change (IPCC) assessment reports, the fifth report highlights the importance of regional-scale climate change (IPCC, 2013). To get over it, we use techniques to infer high-resolution data from these low-resolution coarse data, i.e. GCM-simulated predictors, to obtain locally downscaled values. The method we use to tackle this issue is called downscaling.

* 1. **What is Downscaling**

Downscaling just as it says, is the process of reduction in size, scale, or extent (Merriam-Webster); As such, it is the method of decreasing the pixel size of remotely sensed images (Atkinson 2013) as a means to translate info from global and other large climate models to the scale of local/regional models. Then, why not just use regional models? The issue with regional models is that the computational cost is high; not only that, even if you have that computational power, the regional model is very boundary condition dependent as it derives from a Global Climate Model (GCM). Furthermore, if the GCM is biased, this model inherits and amplifies those biases, as it works in a “garbage in, garbage out” manner (Rummukainen 2010). Also, the initiation process would need localized observations, which are hard to come by, and running it requires expertise from different fields. These are but a few of the many problems faced while running an RCM (Regional Climate Model). Downscaling holds a strong hand in the production of regional-scale predictions and circulation models GCMs e.g., (Tapiador et al. 2020). The process of Downscaling is commonly done through 2 major methods: Statistical Downscaling and Dynamic Downscaling.

* 1. **Dynamic downscaling**

Dynamical downscaling relies on a regional climate or numerical weather model to provide high-resolution climate factors by simulating the physical processes of the coupled land-atmosphere system (Rummukainen 2010). A high-resolution grid is nested on the coordinates required for the high-resolution global model to infer high-resolution data from a low-resolution global model. Even so, it has its limitations, such as requiring a large amount of processing power to produce a reasonable output while being prone to model errors and not being able to be applied to another domain and deliver results with the same impact. There is also the fact that most of the nested models are optimised for the northern hemisphere, even when many countries in the southern hemisphere encounter extreme weather conditions (Hewitson and Crane 1996), a topic of discussion for another day. Then there is the even more significant fact that it is still challenging for the current downscaling approaches to adequately represent spatial and temporal variability of the local-scale climate and capture local-, small-scale features such as local extreme events (Hertig et al. 2019).

* 1. **Statistical Downscaling**

Statistical Downscaling is the basis for developing an empirical relationship between local high-resolution and low-resolution data, either through linear regression or analog techniques and, in some cases, even both (Suarez-Gutierrez, Milinski, and Maher 2021). This gets more accurate as the quality and quantity of these data increase. This relationship or empirical formula is derived from comparing the local model data to observed and high-resolution data using mathematical or statistical equations. Although the Statistical downscaling method is more computationally efficient than nested models, it also has drawbacks, such as a lack of universality; if the relation is derived for an area, it cannot be used for another area, as the forcing functions may differ. Thus, finding a method to obtain accuracy and universality is necessary.

In this paper, we will evaluate the accuracy of the NCUM operational forecast. Then, use statistical Bias correction methods by using historical observational data to correct systematic errors in model outputs. By adjusting the outputs to take into account the bias, we can bring the simulations closer to observational data. The evaluation of the accuracy of NCUM-R G (Global NCMRWF Unified Model - Regional, NCUM-G) done by comparing NCUM-G (Global NCMRWF Unified Model - Global, NCUM-G). Later, we will see how the Bias correction improves the NCUM forecast by comparing the model product and the output given out by various statistical downscaling methods.

In the second section of our project, we will discuss the different aspects of the climate models referred to here in greater detail. In later sections, we will verify the downscaling performance of the Dynamic core, and explore different statistical downscaling methods.

1. **Modelling** 
   1. **Introduction to NCUM**

The NCUM or NCMRWF unified model is a numerical weather prediction model operated by Partnerships' Unified Model (UM) since 2012. It works using a Unified Model (UM) developed by the UK Met Office. NCUM is used for both global (NCUM-G) and regional (NCUM-R) weather prediction.

* + 1. **NCUM-G**

From the latest model verification report, ‘NCUM Global Model Verification: Pre-monsoon (MAM) 2024,’(Kumar et al. 2024) we can see that the global model runs with a resolution of 25 km. Moreover, the latest iteration of the model, NCUM-G V7, has a horizontal grid resolution of 12 km with 70 vertical levels in the atmosphere reaching 80 km. This model is adapted from Unified Model version 11.2 (UM11.2), part of the latest “Operational Global Suite” (PS43) of UK Met Office runs based on an advanced “ENDGame” (Even Newer Dynamics for General atmospheric modelling of the environment) dynamical core. This is adept at solving compressible non-hydrostatic equations of motion with semi-Lagrangian advection and semi-implicit time stepping. The model runs in 5-minute time steps, providing a 10-day forecast. Similar to every NWP (Numerical Weather Prediction) model, this model also needs Quality control that has the least amount of error, and it is provided with a Hybrid 4D Var data assimilation process, where a hybrid of climatology and forecast ensemble is used; where the sub-grid process undergoes physical parameterisation (GA7.2).

* + 1. **NCUM-R**

Among them, NCUM-R is a high-resolution domain for the Indian region and is configured with the UM seamless prediction for various applications. The model, which ran on Singapore version 2 (SINGV2) from 2016, now uses the UK Met Office’s “Regional Atmosphere and Land version 3 (RAL3) to run from October of 2022. This collective of the high-resolution models includes 3 models, each with different resolutions: the 4km - All India Model – which is of most interest to us, a 1.5km - Delhi and neighbourhood model and finally a 330m - Delhi city model. NCUM-R has a rotated (slightly tilted) latitude-longitude horizontal grid (to make the ‘Equator’ on top of the area of interest) with Arakawa-C staggering and a terrain-following hybrid vertical coordinate with Charney-Philips staggering. The 4km domain model, which is run all year round, covers the area of 62⁰E-106 ⁰E; 6 ⁰S 41 ⁰N respectively with a time step of two minutes. The horizontal and vertical extent of the model are 1200x1200 grid points horizontally and 90 hybrid levels vertically, with a top at 40km. With initial conditions set at 00UTC and 12UTC. These high-resolution initial conditions are set with the “4D-Var” data assimilation system, which is used in NCUM-G. Even then, NCUM-R updates its lateral boundary conditions from operational global model forecasts (NCUM-G) every hour. It is to be noted that even when the different parametrization exists in the model, convection is explicit, and the subgrid scale deep convection parametrization is absent.

1. **Evaluation of NCUM-R**

Dynamical downscaling improves regional-scale weather and climate predictions by refining coarse-resolution global model outputs. However, the effectiveness of the dynamical downscaling process is heavily dependent on the Global climate models (GCMs) and parameterisation schemes used. In this study, we will compare the effectiveness of this approach to see how much this method improves upon the parent model and how accurately it represents observed climate patterns. To assess the performance of the NCUM-R model**,** we compare its simulated rainfall against observations and the coarse NCUM-G dataset. The verification focuses on mean rainfall, bias, and variance for the JJAS monsoon seasons from 2021 to 2024. Additionally, case studies of major extreme rainfall events in the same period are also included.

We analyse the biases in NCUM-R to determine whether it improves upon NCUM-G in capturing spatial and temporal rainfall patterns. Extreme rainfall events are also examined to evaluate the model’s ability to capture heavy precipitation events using intensity and frequency distributions.

So, for the verification of the dynamical downscaling method, we take data from 3 major data sets

1. **NCUM-R** (Regional Model) – Daily surface rainfall (APCP\_Surface) output from the high-resolution regional model.
2. **NCUM-G** (Global Model) – Daily surface rainfall (APCP\_Surface) output from the lower-resolution global model.
3. **IMD-MSG** (Observed Dataset) – The observed rainfall(rf) dataset is used as a reference.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Resolution** | **Temporal Coverage** | **Variable Used** |
| NCUM-G | ~12 km | Multiple forecast lead times (Day 1, 3, 5) | Daily precipitation |
| NCUM-R | ~4 km | Multiple forecast lead times (Day 1, 2, 3) | Daily precipitation |
| IMD-MSG | ~10 km | Observed rainfall data | Daily precipitation |

Table 1: table showing different datasets used in this study and their resolution and other details

* 1. **Primary Analysis and Results**

For this analysis, we have taken the daily rainfall 2021–2024 for the JJAS (June–September) monsoon period in the spatial range of 65⁰E-106 ⁰E, 6 ⁰S 41 ⁰N. This approximately matches the domain size of the NCUM-R. (The western boundary 65⁰E is taken into account to compensate for the errors in the IMD-MSG observation dataset, which ensures consistent comparison across all datasets.) The dataset is masked using India's shapefile so that only data over the country is considered before further analysis. This is done for all the 3 datasets before doing any further analysis. We analyse the Mean Rainfall for the period, Variance, Biases, and some forecast verification of the 2 models. Along with some model comparisons on some extreme weather events in the period.

* + 1. **Mean rainfall**

The spatial distribution of mean rainfall is plotted for each dataset to highlight discrepancies and improvements in NCUM-R over NCUM-G. To evaluate how well NCUM-R captures regional rainfall patterns compared to the global NCUM-G.

For this, first, all the datasets need to be gridded to the same resolution of the IMD-MSG observation dataset. This is done by the ‘nearest-neighbour interpolation’ method, where each point in the new grid is assigned the value of the nearest data point in the original grid. It does not perform any averaging or weighting of surrounding points. It is a simple and fast method but may not be as smooth as other interpolation methods like linear or cubic interpolation (xarray Documentation).

After this step, we calculate the mean rainfall over the selected period. Mathematically, the mean rainfall (i,j) for grid point (i,j) is calculated by;

Where Is the rainfall at grid point (i,j) at time t, and T is the total number of time steps. Using that, we find the mean precipitation for the observation dataset and Day-1, Day-2, and Day-3 of NCUM-G and NCUM-R, respectively. For ease of comparison, we make a figure of both separately but with consistent colour scales.

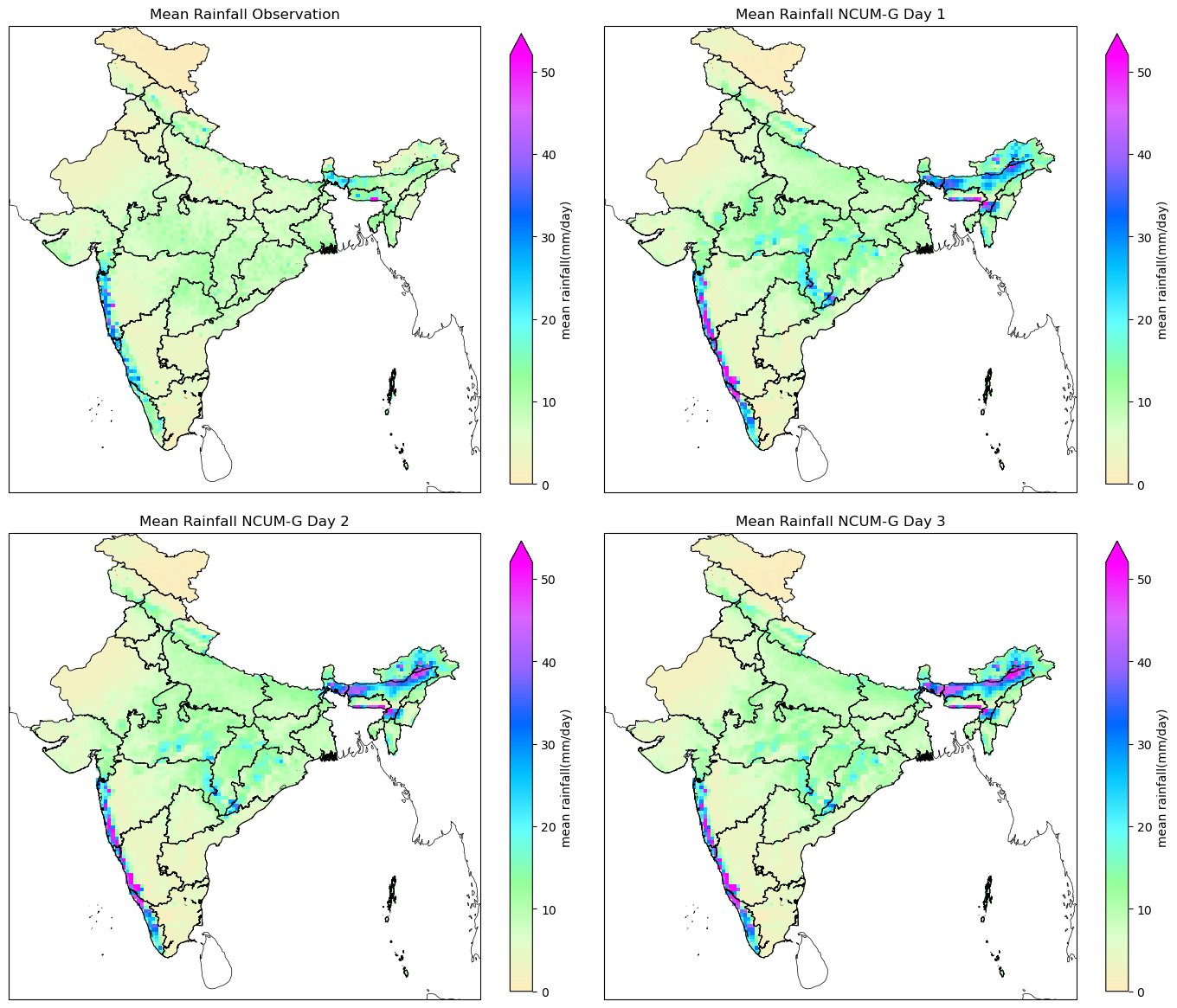


Figure 1 : Mean daily rainfall (mm/day) over India from observations and NCUM-G forecasts. (Top-left) Mean observed rainfall (IMD-MSG). (Top-right) NCUM-G Day 1 forecast. (Bottom-left) NCUM-G Day 2 forecast. (Bottom-right) NCUM-G Day 3 forecast. The colour scale represents rainfall intensity, ranging from 0 mm/day (beige) to over 50 mm/day (purple)

Looking at figures 1 and 2, we can see that the first plots on both the figures depict the observation mean rainfall. In it, high rainfall is concentrated in the Western Ghats (Kerala, Karnataka, Goa, Maharashtra) and Northeast India (Assam, Meghalaya) (~30 mm/day, and up to 50+ mm/day). This serves as a benchmark for evaluating both models.

NCUM-G Model mean rainfall for Day 1, Day 2 and Day 3 are seen in the latter plots of figure 1. Looking at it, the Western Ghat rainfall has more localized high-intensity rainfall (up to 50+ mm/day) and more scattered patches of heavy rainfall compared to observations, and some overestimation can also be seen on the west coast.

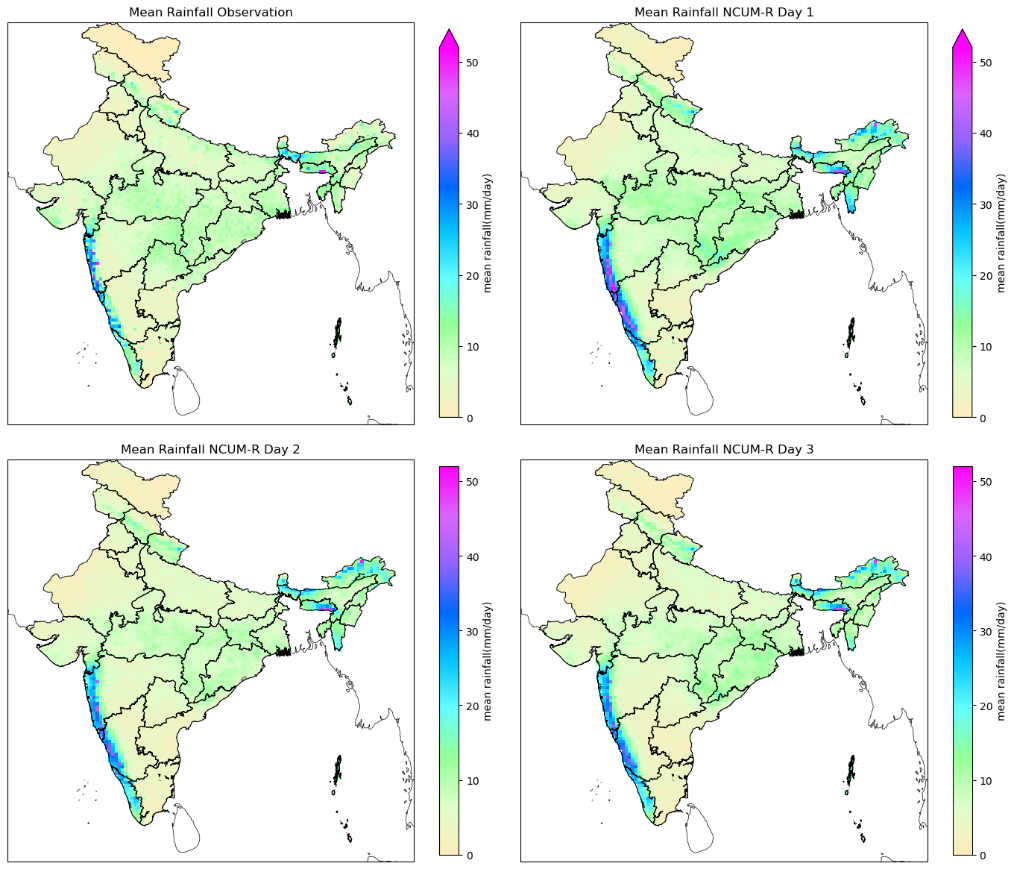


Figure 2 : Mean daily rainfall (mm/day) over India from observations and NCUM-R forecasts. (Top-left) Mean observed rainfall (IMD-MSG). (Top-right) NCUM-R Day 1 forecast. (Bottom-left) NCUM-R Day 2 forecast. (Bottom-right) NCUM-R Day 3 forecast. The colour scale represents rainfall intensity, ranging from 0 mm/day (beige) to over 50 mm/day (purple)

For the Northeast India Rainfall, the rainfall is captured well, but it shows a misaligned peak rainfall and severe overestimation. It also shows rainfall over Parts of Eastern India (Odisha, West Bengal). Day 1 forecast is closest to observations but with overestimated intensities**.** Days 2 & 3: Rainfall decreases, but the model smoothens with lead time.

When looking at the latter plots of figure 2, which shows the NCUM-R model statistics for the Western Ghats rainfall, a more continuous and uniform pattern along the coast is seen. NCUM-R captures the widespread nature of coastal rainfall better than NCUM-G. However, it overestimates rainfall intensity in regions with high orography. And as for Northeast India Rainfall, it aligns slightly better with observations, but there are some displacement issues for peak rainfall in some areas. The Day 1 forecast is more realistic compared to that of NCUM-G. As for Days 2 and 3, the rainfall prediction decreases compared to Day 1.

* + 1. **Variance**

Similar to the Mean rainfall distributions, the variance distribution also uses the daily rainfall data re-gridded to match the resolution of the observation Dataset. Then, the variance (σ2) of forecasted and observed values is calculated by the equation:

σ2

Where:

* represents individual forecast or observed values,
* Is the mean of the dataset,
* N is the total number of data points.

The variance plot is generated to visualise the dispersion of forecasted values relative to observations. Variance is generated for both NCUM-R and NCUM-G for their Day-1, Day-2, and Day-3 forecasts. These dispersion plots help in understanding the spread of model outputs by showing whether the models overestimate or underestimate variability to the observed data in various circumstances during the period of the study.

Fluctuations in variance values correspond to weather events in the forecast, where high variance shows high rainfall events. The difference between the model’s variance and the variance of observation indicates whether the model over-predicts or under-predicts, depending on the sign of the difference.

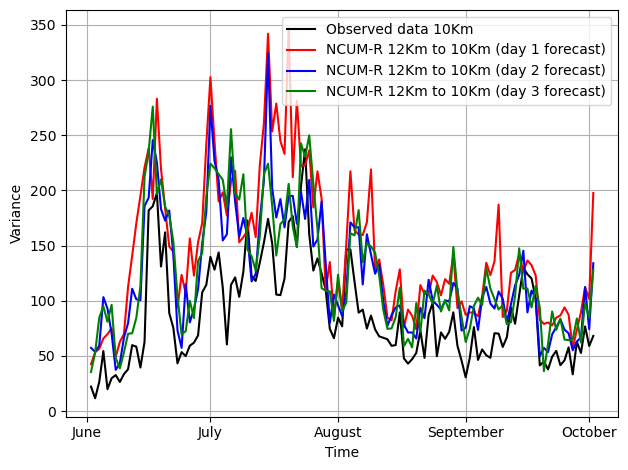
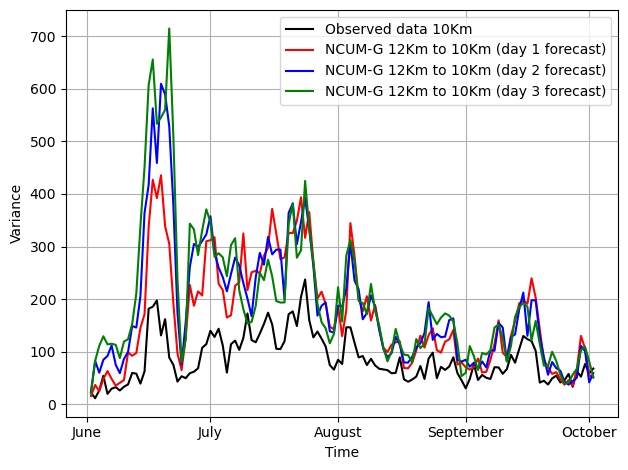


Figure 3 and 4:Temporal variance of daily rainfall over a JJAS months. The black line represents observed data, while the red, blue, and green lines represent 1. NCUM-G, 2. NCUM-R forecasts for Day 1, Day 2, and Day 3, respectively.

In the NCUM-G model (Figure 3), we can see that, although the model captures the general trend of observed values, it shows higher variance values (up to 720) during the peak monsoon months of June and July, reducing thereafter. This means the model captures the extreme events but over-amplifies variability, causing it to predict too much precipitation compared to observation in high-impact weather scenarios. The variance is seen to be greatest for the Day 1 forecast (red line) and reduces with lead time, indicating that predictions smooth out with lead time. Increased variability suggests growing uncertainty with longer lead times. This may be due to the excessive sensitivity of the model to initial conditions taken or the underlying model physics.

Now, if we take a look at the NCUM-R model, as shown in figure 4, we can see that the model variance follows the trend of observations variance very well. The values are seen to be closer to the observation compared to the NCUM-G variance (up to 350). This means that NCUM-R offers a better representation of local-scale processes compared to NCUM-G. Even then, the Day 1 forecast (red line) slightly overshoots on days of high rainfall. Similarly, Day 2 and Day 3 forecasts (green and blue line) also tend to stabilize variance better compared to NCUM-G. It also appears to underestimate variance in some cases, especially in the late monsoon season (months of September–October). This may be due to a damping effect in long-range forecasts of the NCUM-R model.

From all this, we can infer that the NCUM-G model by itself is not enough to meet our needs because the global model generally shows a higher variance than observations, meaning that it overreacts to changes in rainfall, leading to potentially exaggerated forecasts and Bias correction of this model is necessary for getting a much better data product. As seen, while NCUM-R provides more controlled variance, it still tends to both underestimate variance in some cases and overestimate it in others, which makes it imperfect.

* + 1. **Bias**

The Bias quantifies the difference between forecasted and observed rainfall. Studying bias will help identify systematic overestimation/underestimation trends in the models. The same dataset used before is also used for this study. The bias is calculated by the equation:

Bias = Forecast Rainfall − Observed Rainfall

Where the Forecast Rainfall (F) refers to the Rainfall predicted by NCUM-G or NCUM-R. And Observed Rainfall(O), which is our IMD gridded rainfall data. And this is done for all Day 1, Day 2, Day 3 forecasts of both the models. A positive bias means that the model overpredicts rainfall, and a negative bias indicates that the model underpredicts.

Since our dataset is four-dimensional with Rainfall, Latitude, Longitude and Time. We do Bias analysis in 2 ways: 1. Spatial Bias (spatial) and 2. Time Series-Bias (temporal).

* + - 1. **Spatial Rainfall Bias**

For each grid point (i,j):

Bias(i,j) = F(i,j) − O(i,j)

Then, the bias is separately found for NCUM-G and NCUM-R for multiple forecast lead times (Day 1, 2, 3). The bias is averaged over time to create mean bias maps. The spatial bias is visualized using plots red for underestimation and blue for overestimation, as seen in Figure 5.

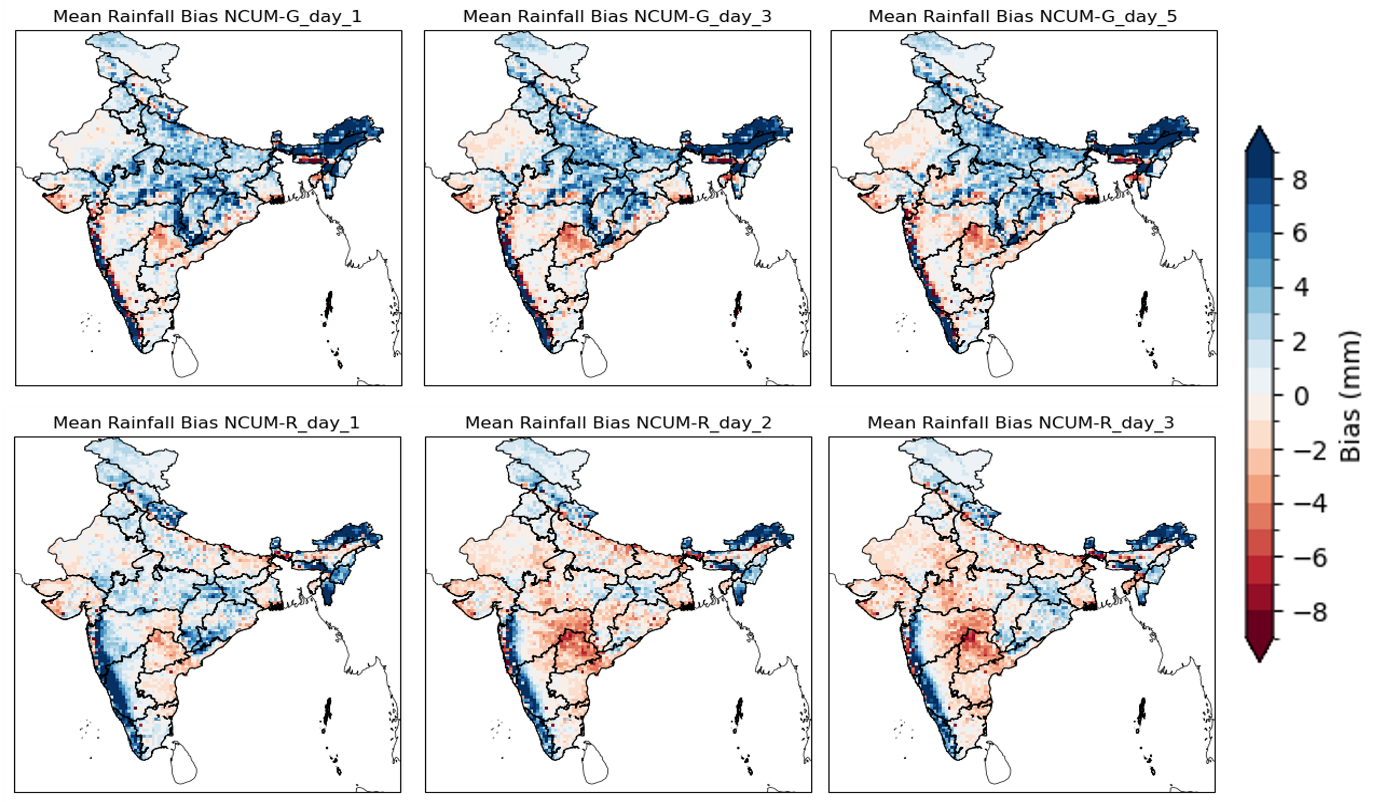


Figure 5 :Mean daily rainfall bias (mm) over India for NCUM-G and NCUM-R forecasts compared to observations. (Top row) NCUM-G biases for Day 1, Day 3, and Day 5 forecasts. (Bottom row) NCUM-R biases for Day 1, Day 2, and Day 3 forecasts. Positive biases (blue) indicate overestimation, while negative biases (red) indicate underestimation.

Figure 5 shows the spatial bias maps of NCUM-G for forecasts of lead times Day 1, Day 2, and Day 3. We can see that both overestimation and underestimation are seen spread out on the whole of India region. Especially in central India and the interior peninsular regions, underprediction (red regions) is dominant. It is also to be noted that the bias intensity slightly reduces from Day 1 to Day 3, which indicates error dampening over longer lead times.

Similarly, also from Figure 5, we can see the spatial Bias maps of NCUM-R for forecasts of lead times Day 1, Day 2, and Day 3. We can infer that NCUM-R has strong overprediction tendencies. In the northeastern region and coastal areas (especially along the western coast), the overprediction is particularly high, even higher than in NCUM-G. Underprediction, although mild, is more widespread in the central regions of India. But the key thing to note is that, unlike the NCUM-G, the bias structure remains almost the same in all lead times, which means that it only has somewhat of a systematic model bias. This model seems to have stronger positive biases over high-rainfall regions (Western Ghats, Northeast) but underestimates rainfall over dry regions, this may very well be an effect due to the model not capturing the variation by the virtue of orography.

From all this, the key takeaway is the fact that NCUM-G bias shows less variability across different regions compared to NCUM-R; at the same time, it shows NCUM-G bias changes slightly over time, whereas NCUM-R shows a more consistent bias. NCUM-G has a strong positive bias, especially in the northeastern and western ghats parts of India, and this bias has been increasing noticeably in the years. Compared to that, the NCUM-R has a more balanced bias. But it also shows some underestimation in different parts of India.

* + 1. **Forecast verification**

Forecast verification is about checking how well weather forecasts match observed data. It assesses how accurate and how reliable the model skill is. By using verification metrics, we can quantify forecast performance and identify model strengths and weaknesses. WE will verify both NCUM\_G and NCUM\_R over different lead times. The key metrics used are Probability of Detection (POD), False Alarm Ratio (FAR), Bias, and Equitable Threat Score (ETS). We will also evaluate them over 1 to 3-day lead times to see how forecast skill changes as lead time increases. To do this, we will take the dataset and regird it concerning the observation dataset as usual for proper comparison, and we will set a threshold of 0-30 to get the general trend.

The following metrics are used to evaluate the forecast performance:

* Probability of Detection (POD):

POD measures the fraction of observed events that were correctly forecasted. A POD value of 1 indicates that all observed events were correctly forecasted, while a value of 0 indicates none were detected. Mathematically, it can be described by,

POD =

* False Alarm Ratio (FAR):

FAR measures the fraction of forecasted events that did not occur. A FAR value of 0 indicates no false alarms, while a value of 1 indicates that all forecasts were false alarms.

FAR =

* Bias:

Bias indicates whether the model tends to over-predict or under-predict events. A Bias value of 1 indicates perfect balance, while values greater than 1 indicate over-prediction, and values less than 1 indicate under-prediction. Mathematically, it is represented by:

Bias =

* Equitable Threat Score (ETS):

ETS measures the accuracy of the forecast relative to random chance, accounting for hits, false alarms, and misses. ETS ranges from -1/3 to 1, where 1 indicates a perfect forecast and 0 indicates no skill. Mathematically, it is represented by:

ETS =

Where Expected hits = (Hits + Misses) x (Hits + False Alarms) / (Hits + Misses + False alarms + Correct Rejection)

Analysis

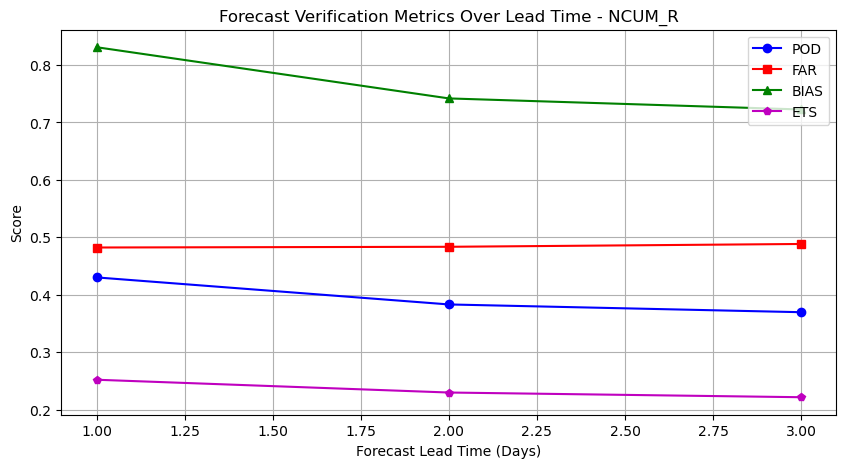
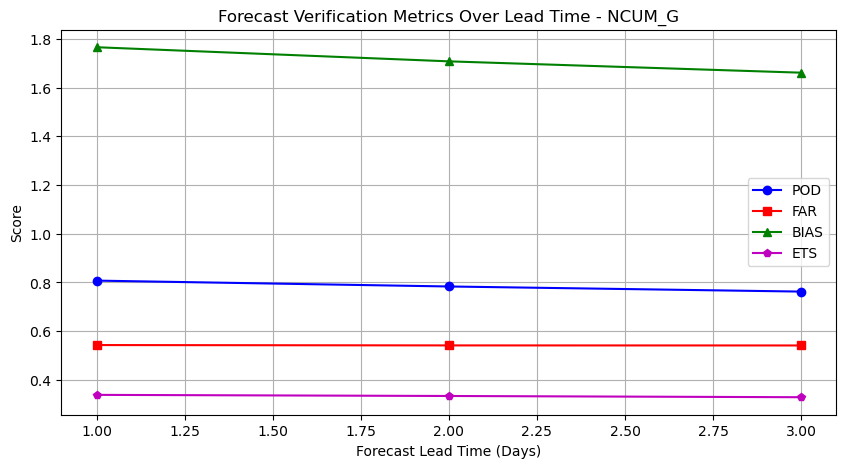


Figure 6 : Forecast verification metrics (POD, FAR, BIAS, and ETS) over different lead times for NCUM-G. The performance remains relatively stable, with minor variations in BIAS and POD as lead time increases. Figure Y: Forecast verification metrics (POD, FAR, BIAS, and ETS) over different lead times for NCUM-R. The skill scores show a gradual decline with increasing lead time, particularly for POD and ETS, indicating a reduction in forecast accuracy.

* + - 1. **NCUM\_G Plot Analysis**

POD is the measure of actual events that are correctly forecasted. As we see in Figure 6, in the case of NCUM\_G, POD starts high (around 0.8) and decreases slightly as the lead time increases, this means the ability to detect the events of rainfall is decaying over time, that is Day 1 predictions are the most reliable compared to Day 2 and Day 3 forecasts.

FAR is the fraction of forecasted events that did not occur. FAR stays the same with increasing lead with a value little above 0.5, suggesting that the model tends to over-predict events.

Bias indicates whether the model tends to over-predict (Bias > 1) or under-predict (Bias < 1) events. The Bias is quite high, around 1.8 for Day 1 forecasts, and it reduces with lead time. This means that NCUM-G tends to overpredict, and it smoothens out with lead time much like our previous analysis.

ETS measures the accuracy of the forecast relative to random chance. ETS stays relatively stable with lead time, indicating that the forecast skill remains the same, about 0.3, as the lead time increases.

* + - 1. **NCUM-R Plot Analysis**

POD in NCUM\_R starts lower (around 0.4) and decreases slightly over lead time, then remains the same value after Day 2, detection rate of NCUM\_G is better than NCUM-R.

Although POD is lower than NCUM-G, the value of FAR in NCUM\_R is less than 0.5, which means that although it struggles hard to detect the events, the accuracy of the forecast is a bit more than NCUM-G. Similar to NCUM\_G, but it starts from a lower baseline, suggesting fewer false alarms initially.

Although bias in NCUM\_R starts around 0.8, it decreases with an increase in lead time. This means that the NCUM-R tends to give predictions closer to the actual Precipitation, although it shows a bit of underestimation.

ETS in NCUM\_R decreases a bit with lead time and then remains constant. The value is somewhat similar to NCUM\_G, but starts from a lower value, indicating lower initial forecast skill compared to NCUM\_G.

* + 1. **QQ plot of Rainfall**

A quantile–quantile plot or Q-Q plot indicates bias in is made to understand the reliability of the NCUM-R plot compared to the NCUM-G model in predicting daily surface rainfall. The data set used is the same dataset we have been using. This plot is made by comparing the empirical quantiles of the observed rainfall and model predictions of both NCUM-R and NCUM-G. As the data is spatial data, the first step is to flatten the data to a long 1-dimensional data. After this, Sorting is done where both observed and predicted values are sorted in ascending order. These are done for Day 1, Day 2, and Day 3 of the Model and observation dataset. Then, quantile matching is done; it is the process of plotting corresponding quantiles of observed and predicted distributions against each other. We also add a theoretical 1:1 line (red dashed) to show the perfect agreement for reference. Separate Q-Q plots are generated for NCUM-G and NCUM-R forecasts at lead times of Day 1, Day 2, and Day 3, in python using libraries like NumPy and matplotlib.

The figure 7 corresponds to NCUM-G and NCUM-R Day 1, Day 2, Day 3 forecasts. Comparing these plots to each other will reveal which model performs better. While highlighting the biases the model has, it will shed light on how and where to correct those biases.

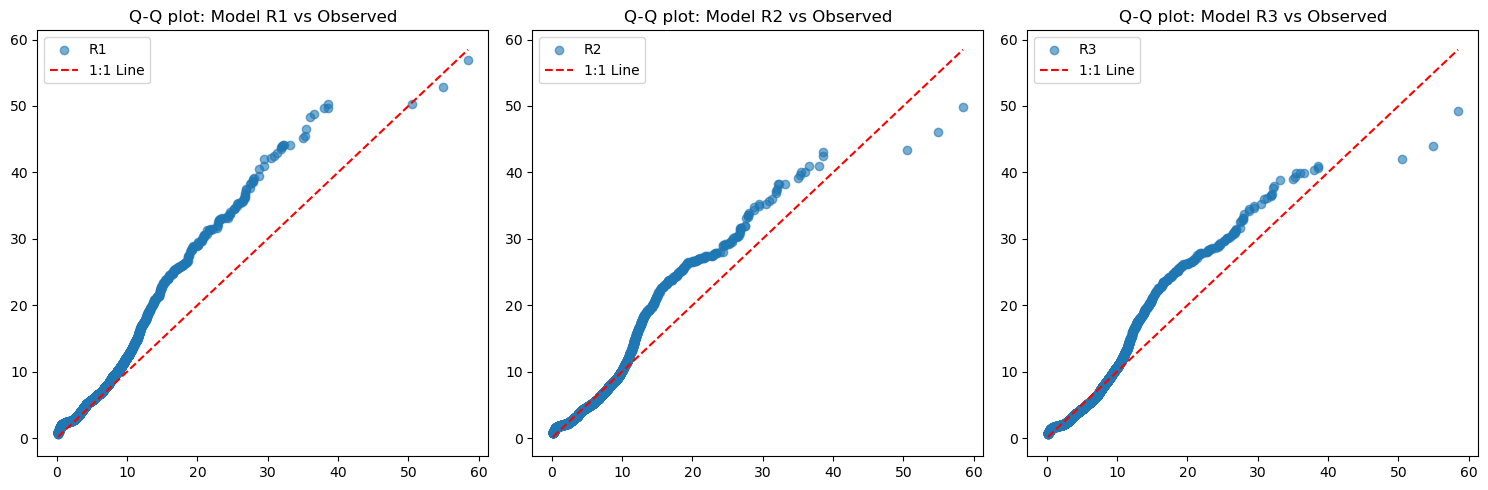
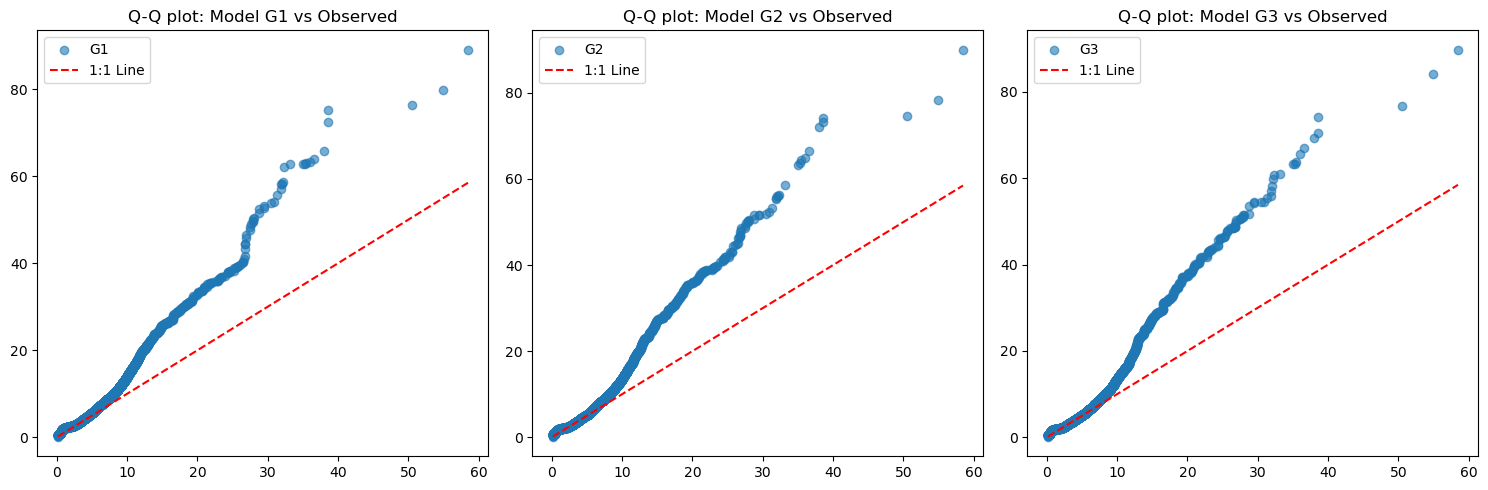


Figure 7 : Quantile-Quantile (Q-Q) plots comparing the distribution of rainfall from NCUM-G (G1, G2, G3) with observed data. Deviations from the 1:1 line indicate biases in the model, particularly for higher rainfall values. Figure : Quantile-Quantile (Q-Q) plots comparing the distribution of rainfall from NCUM-R (R1, R2, R3) with observed data. The models show better agreement with observations at lower values but exhibit biases at higher extremes.

For NCUM-G, although at lower values (near zero), the predictions are relatively close, and it shows a significant deviation from the 1:1 line, particularly in the upper quantiles. It is seen that as the values increase, the deviation from the 1:1 line becomes stronger; this shows the systematic underestimation of higher values by the model. As lead time increases, the model doesn’t have any apparent changes, suggesting that the model is consistent with lead times.

Compared with that of NCUM-R, it shows a much closer alignment to the 1:1 line. And stay close to it for bigger quantiles compared to that of NCUM-G. It is also to be noted that, while there are deviations in the upper quantiles, these discrepancies are noticeably smaller in comparison to the NCUM-G model. Which implies the NCUM-R model is less biased and is better at capturing a bigger range of observed values.

|  |  |  |
| --- | --- | --- |
| **Feature** | **NCUM G** | **NCUM R** |
| **General Fit** | Curve follows 1:1 line at lower quantiles, and diverges sharply. | Curve follows 1:1 line at lower quantiles, and diverges but not as sharp as NCUM-G. |
| **High Quantile Behaviour** | Largely overestimates large values. | Overestimates large values. |
| **Consistency** | Reasonably accurate at lower rainfall & bias increase sharply at high quantiles. | Reasonably accurate at lower rainfall & bias increase with high quantiles. |
| **Reliability** | Very high tendency to over-predict extremes. | Tendency to over-predict extremes. |

Table 2: This table shows Comparison of NCUM-G and NCUM-R models in terms of general fit, high quantile behaviour, consistency, and reliability

NCUM-G has a more pronounced bias, especially for larger values. These values deviate significantly upwards in higher quantiles, meaning that predicted values tend to be much larger than observed values, confirming overprediction. The curvature of the Q-Q plot is much predominant on NCUM-G; this proves the earlier statement.

NCUM-R is a superior model compared to NCUM-G in terms of distributional fit. The poor fit of NCUM-G suggests that the model is not well-calibrated for the observed data, possibly due to incorrect assumptions, model limitations, or inadequate parameter tuning. This may be due to a myriad of different reasons like lack of flexibility in model structure, improper training or tuning of parameters, Bias in training data, or Transformation issues.

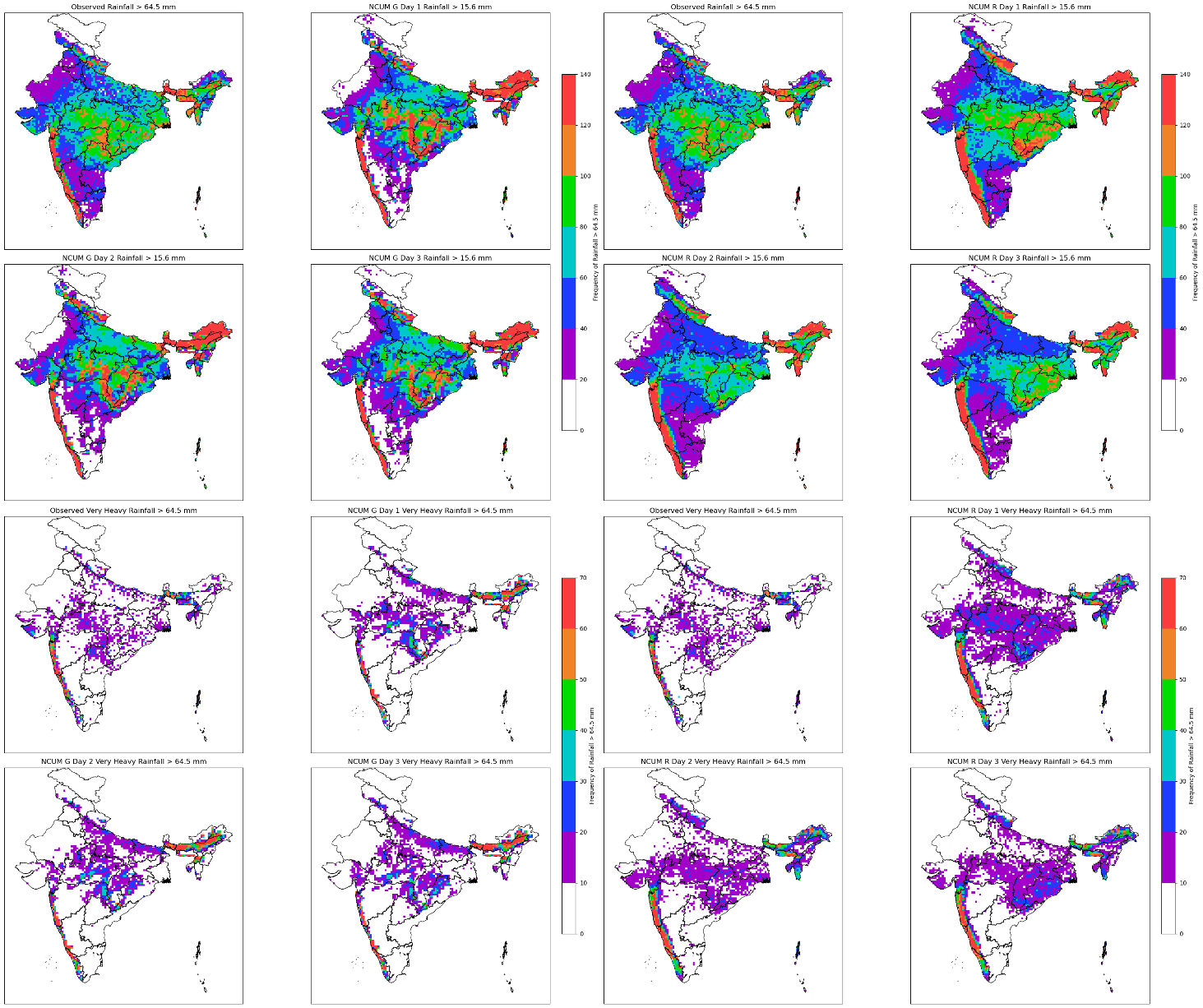
Refining the NCUM-G model by introducing more flexible parameterization, adjusting training strategies, or adding transformations to better capture extreme values. , investigating residual errors, could confirm and identify certain systematic patterns that exist in NCUM-G. The made bias was corrected to make it an ideal data for modern-day forecasts.

* + 1. **Rainfall Frequency**

So, we have understood the importance of doing bias correction on the model data to give a better data product, which will yield a better forecast. Now, the next question that remains is the source of these biases. So, for that, we have to analyse the Frequency plot of Intense rainfall. These are done primarily for two thresholds: 1. moderate rainy days (“15.6 mm/day”) and 2. heavy rainy days (“64.5 mm/day”).

In figures M and N, the model is the Frequency of moderate rainfall, i.e. events exceeding 15.6 mm across India. The analysis is conducted over three forecast lead times to assess how well each model captures the rainfall events. The colour scale represents the frequency of rainfall occurrences, ranging from low (blue) to high (red).

In figure 8, The first panel of both the plots shows the observed dataset, in it the region which receives the most frequent moderate rainfall is Northeastern India (Assam, Arunachal Pradesh, Nagaland, and Meghalaya), Western coastal regions (Kerala, Karnataka, Konkan region in Maharashtra, and Goa), Central India (Madhya Pradesh, Chhattisgarh, and Odisha), Northern states like Jammu & Kashmir, Himachal Pradesh, and Uttarakhand and the Northwest India (Rajasthan, Gujarat, and parts of Punjab and Haryana) which have lower rainfall frequencies are identified as Drier regions.

Figure 8 **:** Spatial distribution of observed and modelled rainfall over India for different lead times. The top two rows compare total rainfall from NCUM-G and NCUM-R models against observations during moderate rainfall (≥ 15.6mm), while the bottom two rows highlight areas experiencing very heavy rainfall (≥ 64.5 mm). Differences in spatial patterns indicate model biases in predicting extreme rainfall events.

When looking at the NCUM-G model, it is seen to capture the general rainfall pattern well, especially in the Northeast and Western Ghats. But it is seen to overestimate rainfall in parts of North India (Punjab, Haryana, western UP) and underpredict in Central India, particularly in Odisha and parts of Maharashtra. With an increase in lead time, the model degrades while the overestimations of wet regions persist, the model gets drier, and the drier biases regions get more apparent. This inconsistency shows the decay of forecast skill with lead time.

Comparing it to the NCUM-R model, the NCUM-R has more accurate spatial alignment and matching frequency with observed rainfall. The Western Ghats and Northeast India are well captured, and it has fewer overpredictions compared to NCUM-G. But even then, there are some underestimations still present in Odisha and parts of Maharashtra. The lower overprediction reflects thelow false alarm rate of NCUM-R. It is also noted that as the lead time of forecast increases, the model doesn’t decay much; instead, it smoothens out. NCUM-R is a better spatial match with observations. With lead time, it retains reasonable accuracy and shows less overprediction. In the Western Ghats, Northeast India, an accurate representation with realistic intensity and distribution is displayed.

NCUM-R or dynamical downscaling in general is a sure-fire way to a give a consistent way to refine global model outputs, but it has its challenges. Systematic biases creep in due to model parameterizations, boundary conditions, and the limitations of the driving model itself. This leads to noticeable discrepancies, especially for parameters like rainfall, which is highly sensitive to model physics, terrain influences and parametrisation schemes. If left uncorrected, they can reduce the reliability of high-resolution predictions.

This is where statistical bias correction comes in. Instead of relying on model physics alone, statistical methods use historical observational data to correct systematic errors in model outputs. By adjusting the outputs to take into account the bias, we can bring the simulations closer to reality, making them more useful for modern - day applications. In the next section, let’s explore these techniques in detail - how they work, their strengths, and how they improve the reliability of downscaled data.

1. **Bias Correction of NCUM-G using Statistical Methods**

So far, we have looked at and compared the effectiveness of dynamical downscaling in producing accurate and detailed downscaled datasets. However, the results have not been entirely promising, making it clear that bias correction is essential for improving model outputs. These models may not be reliable for generating accurate predictions and forecast products without proper correction. As such, in this section, we will look into various bias correction methods to make the global model output reflect the features of the observation product, to make it a more accurate and robust product.

Biases in numerical weather prediction models are due to factors such as 1) Errors in convective parameterizations, where the model struggles to include/resolve sub-grid-scale convection processes. 2) Unresolved sub-grid-scale orography, causing small-scale terrain features that impact precipitation patterns, are not fully captured by the model’s resolution. 3) Systematic model drift, due to long-term integration issues, model physics imbalances, which cause the model to consistently overestimate/underestimate rainfall in certain regions. To address these challenges, we need to look at different methods to do bias correction; this is where statistical bias corrections come into play. Statistical bias correction techniques have been shown to effectively address errors arising from convective parameterizations and unresolved sub-grid-scale orography—two major factors contributing to biases in model outputs. (Ehret et al. 2012; Niranjan Kumar et al. 2022; Velasquez, Messmer, and Raible 2019). These techniques transform model forecasts to make their statistical properties better match those of observations. Among the different statistical bias correction methods, Quantile Mapping (QM) is one of the most effective approaches for correcting systematic errors in precipitation forecasts. It directly adjusts the entire distribution of rainfall values, rather than just correcting the mean or variance. This makes it particularly useful for capturing both moderate and extreme rainfall events.

The key reason for choosing Quantile Mapping is that it captures the full distribution, unlike mean or linear bias correction methods, and it also handles the extreme/tail values better, which is ideal for usually skewed data like rainfall data. It is also noteworthy that this method can handle nonlinear biases, which models tend to have. The only downside of choosing such a method is that it can be needy of a solid reference dataset.

In this section, we are going to take the high-resolution rainfall data from Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis (Ashrit et al. 2020; Rani et al. 2021) and IMD Gauge datasets in the time frame of 1989 to 2019 as the aforementioned solid reference dataset. after that we will do the Bias correction for the data on the testing dataset NCUM-G operational forecast dataset for 2020 to 2024 JJAS months (Sumit Kumar et al., 2020). These datasets form the basis for constructing bias correction (BC) models which follows EQM (Empirical Quantile Mapping), PQM (Parametric Quantile Mapping), GPQM(Gamma-Generalised Pareto Quantile Mapping), as detailed in Niranjan Kumar et al. (Niranjan Kumar et al. 2022), and subsequently for training and testing the model. And finally compare these to the Raw forecast to see how it fares.

* 1. **Study area**

The study area of this section is the same used in the previous section, the Indian subcontinent, specifically focusing on the region within the Indian national boundaries. This area is confined to India, bounded by bounding with the shapefile used for clipping and analysis. The time frame taken is JJAS months of 1989 to 2019 to serve as reference and 2020 – 2024 to test the bias correction model.

* 1. **Statistical Bias correction**

Statistical bias correction techniques via Quantile mapping come in many flavours like Empirical Quantile mapping, Parametric Quantile Mapping, etc., but similar to other statistical bias correction methods, they aim to establish a functional relationship between model-simulated outputs and actual observed values, based on well-defined historical reference datasets. Once this relationship is calibrated, it can be applied to adjust future model outputs to reduce systematic errors. The following sections outline the quantile-based correction methods applied in this study.

* + 1. **Empirical Quantile Mapping (EQM)**

The Empirical Quantile Mapping technique (EQM) is the first bias correction approach we used; it operates without assuming a specific distribution form, making it well-suited for adjusting both the average behaviour and the distributional shape of model outputs. The working principle of EQM is that, instead of correcting the mean alone, like traditional methods, this addresses the systematic deviations across the entire range of values by matching the quantiles of simulated data with those of observed records.

To do this, first, we start constructing the empirical cumulative distribution functions (CDFs) for both the model-simulated and observed datasets during a historical reference period. At each quantile, a correction term is computed and applied to the simulated data. These adjustments are then interpolated across quantiles.

Mathematically, the adjusted (or calibrated) precipitation for the ith quantile can be described as:

Pical=Piobs++δi'

Here:

* Piobs ​: Observed precipitation at the ith quantile
* Mean correction component
* δi' = Pifut − Pictl: Quantile-specific deviation, where Pifut ​ is the future model output and Pictl​ is the historical simulation

This method has been applied in several past studies, providing a robust framework for adjusting modelled precipitation to better align with observations (e.g., Boé et al., 2007; Déqué, 2007; Amengual et al., 2012). Although simple, the result shows a great improvement, but it struggles with extreme values.

* + 1. **Parametric Quantile Mapping (PQM)**

Unlike EQM, which works directly with observed distributions, the Parametric Quantile Mapping (PQM) method takes a slightly different approach by fitting probability distributions to the data. It assumes that both observed and modelled precipitation follow a known theoretical distribution — in our case, the two-parameter gamma distribution, due to its ability to flexibly capture rainfall characteristics.

In this approach, the gamma distribution parameters (shape(k) and scale(σ)) are estimated separately for the observed and simulated datasets during the calibration period. Once the distribution parameters are obtained, quantiles can be mapped from the model's fitted distribution to the observed one through their respective cumulative distribution functions (CDFs). This creates a smooth and continuous correction function, particularly useful for handling data-sparse or extreme values.

The probability density function (PDF) of the gamma distribution used is given by:

Here, x is the precipitation value, k is the shape parameter, σ is the scale parameter, and Γ(k) is the gamma function. This parametric formulation allows extrapolation beyond the calibration range, making it a strong choice for correcting extremes, as supported by Piani et al. (2010a) and others. Although as said above this method is strong with extrapolation its only valid in light to moderate rainfall where it’s smooth, continuous, and doesn't need observed values at every possible quantile; but when it comes to extreme values, (95th, 99th percentiles and beyond), gamma struggles because it's a light-tailed distribution. But the distribution sometimes results in higher than observed extremes, as there is no restriction on the upper limit, but these have a high chance of being false alarms for extreme rainfall events (Kondapalli et al. 2022).

* + 1. **Gamma-Generalised Pareto Quantile Mapping (GPQM)**

All the methods we have seen so far have struggled with extreme values, one of the major reasons that they have been sticking to a single distribution. So, to overcome the limitations of just using a single distribution for rainfall bias correction, we turn to a hybrid approach combining the gamma distribution with the Generalised Pareto Distribution (GPD). This method is known as Gamma-Generalised Pareto Quantile Mapping (GPQM), it retains the strengths of the gamma distribution for light to moderate precipitation while leveraging the GPD’s strength of being an ‘extreme value theory’ to better represent the tail-end behaviour of extreme rainfall. In this setup, the gamma distribution is used to model values below a specified threshold. GPD takes over beyond this threshold to more accurately capture the distribution’s heavy tail. We need to look at the probability density function of the GPD first, which is:

Where:

* k is the shape parameter (tail heaviness),
* σ is the scale parameter,
* θ is the threshold, and
* x > θ

In practice, a smooth transition will be implemented at the threshold to ensure continuity between the gamma and GPD portions of the distribution. This method allows for more robust handling of extremes, which are often underestimated or overestimated by lighter-tailed distributions.

What we have done in this section is that we have compared the performance of GPQM and other quantile-based correction methods and raw NCUM-G to observation data, for the evaluations we have done.

* Comparing the Mean rainfall for each of the JJAS Months of 2020-2024
* Comparing the spatial bias of rainfall in the JJAS Months of 2020-2024
* Variation in the Cumulative Distribution function (CDF) for NCUM-G and bias corrections comparison to Observations in the JJAS Months of 2021-2024
* Evaluated using standard verification metrics:
  + Probability of Detection (POD) – measures how effectively true events are captured.
  + False Alarm Ratio (FAR) – evaluates the proportion of incorrectly forecasted events.
  + Equitable Threat Score (ETS) – provides a balanced assessment of predictive skill by accounting for random hits.
* Symmetric Extreme Dependency Index (SEDI) plot for extreme rainfall at the threshold of 64.5 mm to evaluate the skill of forecasts for extreme rainfall events, focusing on how well extreme hits and false alarms are balanced.

After this blend of rigorous evaluation, we can get a robust idea of how much each bias correction method fares in terms of improving rainfall forecasts, particularly where extremes are critical.

* 1. **Data and pre-processing**

As we know quantile mapping is the matching the quantiles of the model's distribution to observed distribution. For this we find model outputs percentile in the model's CDF, then taking the observed value at that same percentile from the observed CDF. It is to be noted that the empirical methods (EQM) fit method directly implements empirical distributions. Whereas, the parametric methods (PQM, GPQM) fit specific distribution functions to the data.

Figure 9: Flow diagram of the methodology adopted in this study. The datasets used (Blue) and various quantile mapping methods used (Grey) to bias-correct the model are shown accordingly.

For fulfilling this demand we have sourced the high-resolution rainfall dataset, IMDAA Reanalysis Rainfall for the baseline period (1989 to 2019), along with Rain gauge data of the same period to serve as ground truth, to make a bias correction model which is developed by mapping the empirical or parametric cumulative distribution function (CDF) of model data to that of the observed data:

where:

* fm​: CDF of the model/reanalysis data,
* fo−1​: Inverse CDF of the observed rainfall data.

And after that the four bias correction techniques are tried, which are

* **Empirical Quantile Mapping (EQM)** a non-parametric approach using empirical CDFs.
* **Parametric Quantile Mapping (PQM)** uses gamma distributions fitted to data.
* **Generalised Pareto Quantile Mapping (GPQM)** a method similar to that of PQM, capturing extreme events using the Generalised Pareto distribution.

These derived bias correction functions are applied to IMDAA reanalysis data for the validation:

where F is the derived bias correction function

This helps to see the performance of bias correction under independent historical conditions.

And finally, the bias correction model was applied to NCUM operational forecasts for real-time correction:

where similar to before F is the derived bias correction function

The makes the corrected model forecasts which follows the statistical characteristics of past observed rainfall closer to observations, making them more reliable for further analysis and practical applications.

* 1. **Results and Discussions**

So, to see how well the bias correction methods (EQM, PQM, GPQM) worked, we have done some comparisons between the raw NCUM-G data and the observed IMD-MSG rainfall over India for the monsoon season (JJAS) from 2020 to 2024, as discussed before, we looked at things like monthly averages, how much the model was off in different regions (bias), and also checked how well the rainfall distributions matched using CDFs. On top of that, also calculated a few verification scores to properly measure the improvements.

The results below break it all down — what changed after applying each method, and which one fixed the model's errors best and is the best bias correction method for rainfall.

* + 1. **Spatial Distribution of Mean Rainfall (JJAS 2020–2024)**

In earlier sections, we have seen details on how well the NCUM-G model forecasts rainfall over India during the monsoon season (JJAS) for the years 2020–2024. Similarly, we have taken a look at three different lead times of NCUM-G (Day 1, Day 3, and Day 5 ) and compared the raw model outputs with observed rainfall data to understand where and how the model goes off. Then, we apply three bias correction techniques which are EQM, PQM, and GPQM, to see how much they can fix the overestimations and misplacements, giving special consideration to regions like the Western Ghats, Odisha, and most notably, Northeast India, where the bias was seen to be quite high when comparing NCUM-G to observation data. First, we look at the all-India pattern, then zoom in on NE India to get a better sense of how each correction method performs in that tricky region.

As seen in Figure 10, the raw NCUM-G forecasts tend to consistently overestimate rainfall, especially in regions like the Western Ghats, Odisha, and especially in Northeast India. This bias becomes more pronounced with increasing forecast lead time (Day 5 > Day 3 > Day 1). The NCUM-G model captures the general trend of rainfall pattern well, and over central and

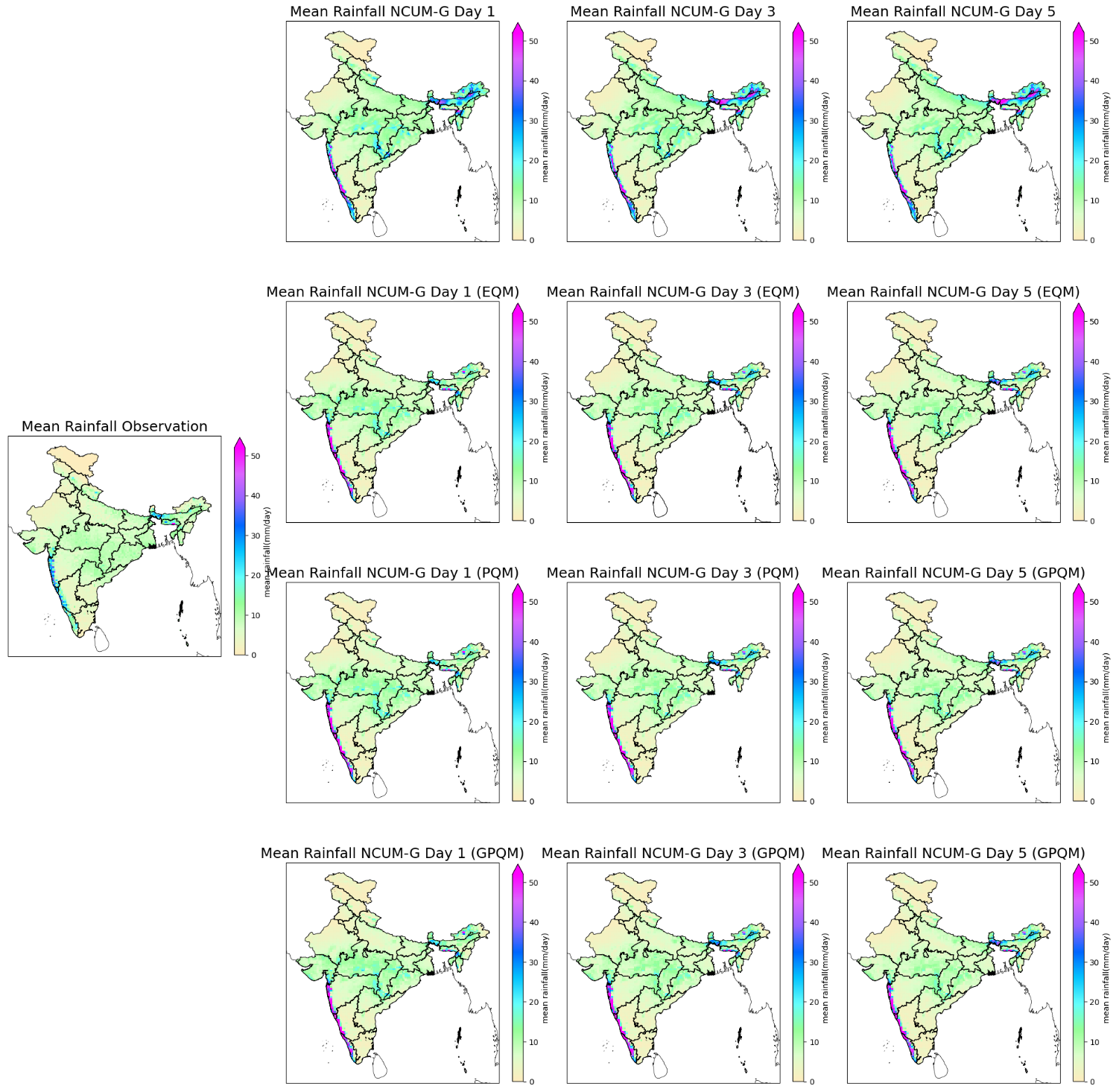


Figure 10: presents the spatial distribution of daily mean rainfall of India during the monsoon season (JJAS) averaged over 2020–2024. The maps compare NCUM-G forecasts for Days 1, 3, and 5 with observations, as well as with outputs corrected using three statistical bias correction methods: EQM (Empirical Quantile Mapping), PQM (Parametric Quantile Mapping), and GPQM (Generalised Pareto-based Quantile Mapping).

southern India, the forecasts show really good agreement, but even there, some isolated rainfall peaks that are not seen in observations persist. When we apply EQM, it provides modest correction by reducing overall rainfall magnitude, particularly in areas with widespread overestimation. However, it does not fully address extreme rainfall peaks, leading to persistent bias in high-intensity zones. Also, there are noticeable isolated peaks, although fewer compared to raw NCUM-G in the mean rainfall as well. PQM does a better job in that department. Not only does it reduce the magnitude of the bias, but it also tweaks the shape of the rainfall distribution itself, by modifying the higher quantiles, which means it leads to a more realistic depiction of the intense rainfall pattern compared to EQM. PQM patterns align more closely with observed patterns, especially in the Day 1 and Day 3 forecasts. However, some pockets of high rainfall persist unrealistically in some parts of the region.

Compared to other methods, however, GPQM stands out. It offers the cleanest, most consistent spatial pattern of rainfall, cutting down unrealistic heavy rain areas while still preserving the broader monsoon structure. Even for Day 5 forecasts, where the raw model is at its worst, GPQM holds up well, showing that it’s easily the most robust and effective of the three correction techniques. It also captures the spatial pattern of rainfall in North-East India with the highest fidelity. It is to be noted that all this smoothing out is done without removing genuine rainfall peaks.

The Indian Northeast is highly landslide/flood-prone, due to the multitude of hilly/valley regions. So even small errors in rainfall prediction can have major socio-economic impacts in this region. Thus, reducing bias in this region is critical for the country. As such, we have given special care to the northeastern region where the Bias is also seen to be consistently high.

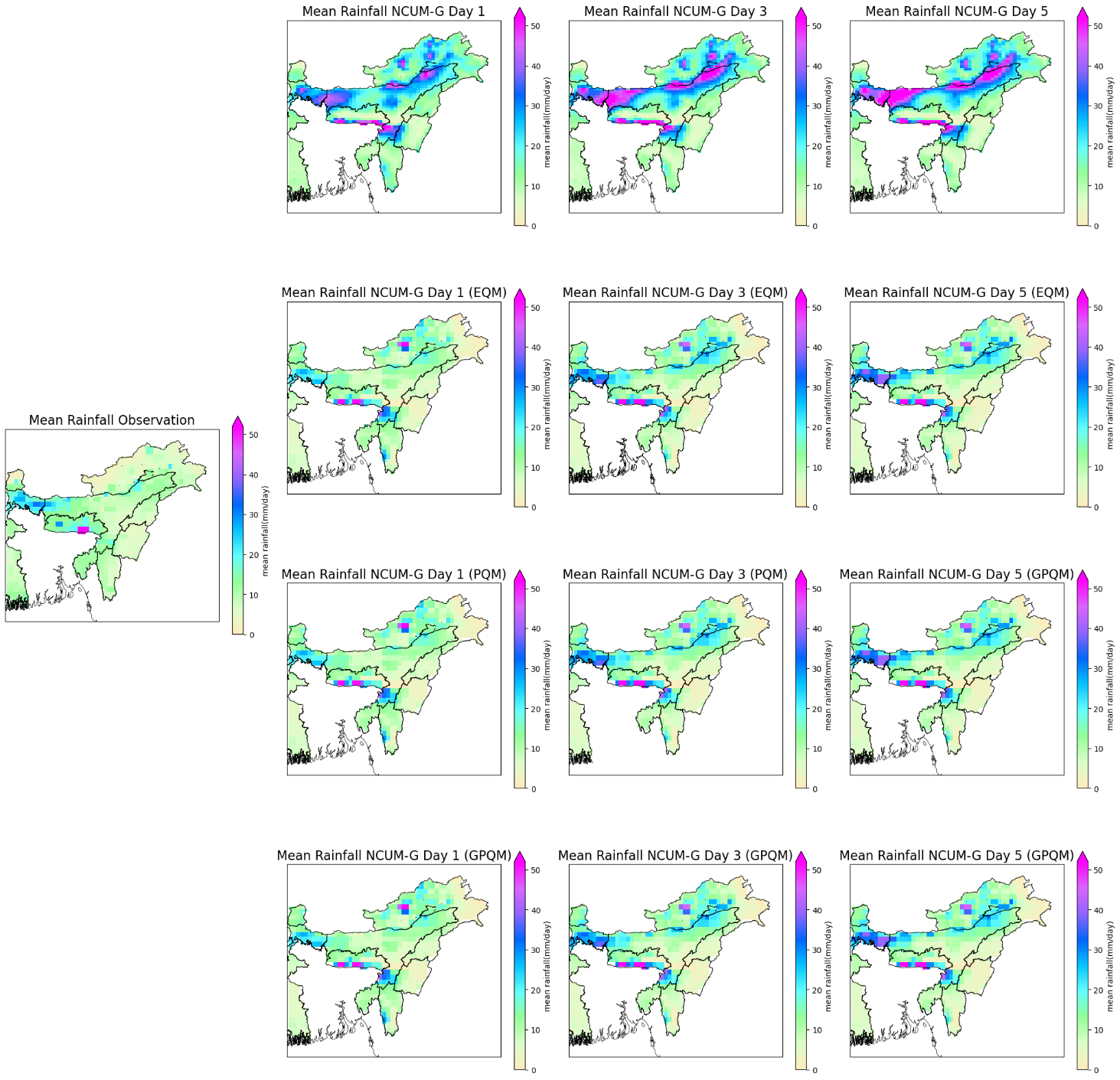


Figure 11 presents the spatial distribution of daily mean rainfall of North India Region during the monsoon season (JJAS), averaged over 2020–2024. The maps compare NCUM-G forecasts for Days 1, 3, and 5 with observations, as well as with outputs corrected using three statistical bias correction methods: EQM (Empirical Quantile Mapping), PQM (Parametric Quantile Mapping), and GPQM (Generalised Pareto-based Quantile Mapping).

As discussed before, Northeast India (NEI) is one of the most sensitive regions in the monsoon system and also one of the most challenging to model accurately due to its complex terrain and orographic effects. This region is consistently identified as a high-bias zone across all lead times in the raw NCUM-G outputs. The raw forecasts tend to grossly overestimate rainfall in this region. Exceeding observed values can be seenfar and wide in regions like Assam, Meghalaya, Arunachal Pradesh, and adjoining states. Another thing to note is that it’s not just the magnitude that has errors, but also manifests in rain bands, where they are often misplaced.

When comparing EQM, the correction somewhat reduces the magnitude of overestimations but fails to correct the spatial placement and distribution of intense rainfall zones (misplaced rain bands). The biases, particularly in foothill regions and high rainfall corridors like the Meghalaya Plateau, remain noticeable.

PQM improves this further by modifying the higher quantiles, leading to a more realistic depiction of intense rainfall patterns. As such, the patterns align more closely with observed patterns, especially in the Day 1 and Day 3 forecasts. However, pockets of high rainfall persist unrealistically in some parts of the region.

GPQM shows a remarkable improvement over both EQM and PQM. Areas of consistently overestimated rainfall in raw NCUM-G are effectively handled without removing genuine rainfall peaks. Even in Day 5 forecasts, GPQM maintains a steady rainfall pattern structure, indicating robustness across lead times. So, we can say that it can capture the spatial pattern of rainfall in NEI with the highest fidelity.

GPQM's ability to tackle rainfall magnitude and spatial misalignment makes it an ideal method for operational bias correction in data-scarce, high-impact regions like NEI.

* + 1. **Spatial Bias**

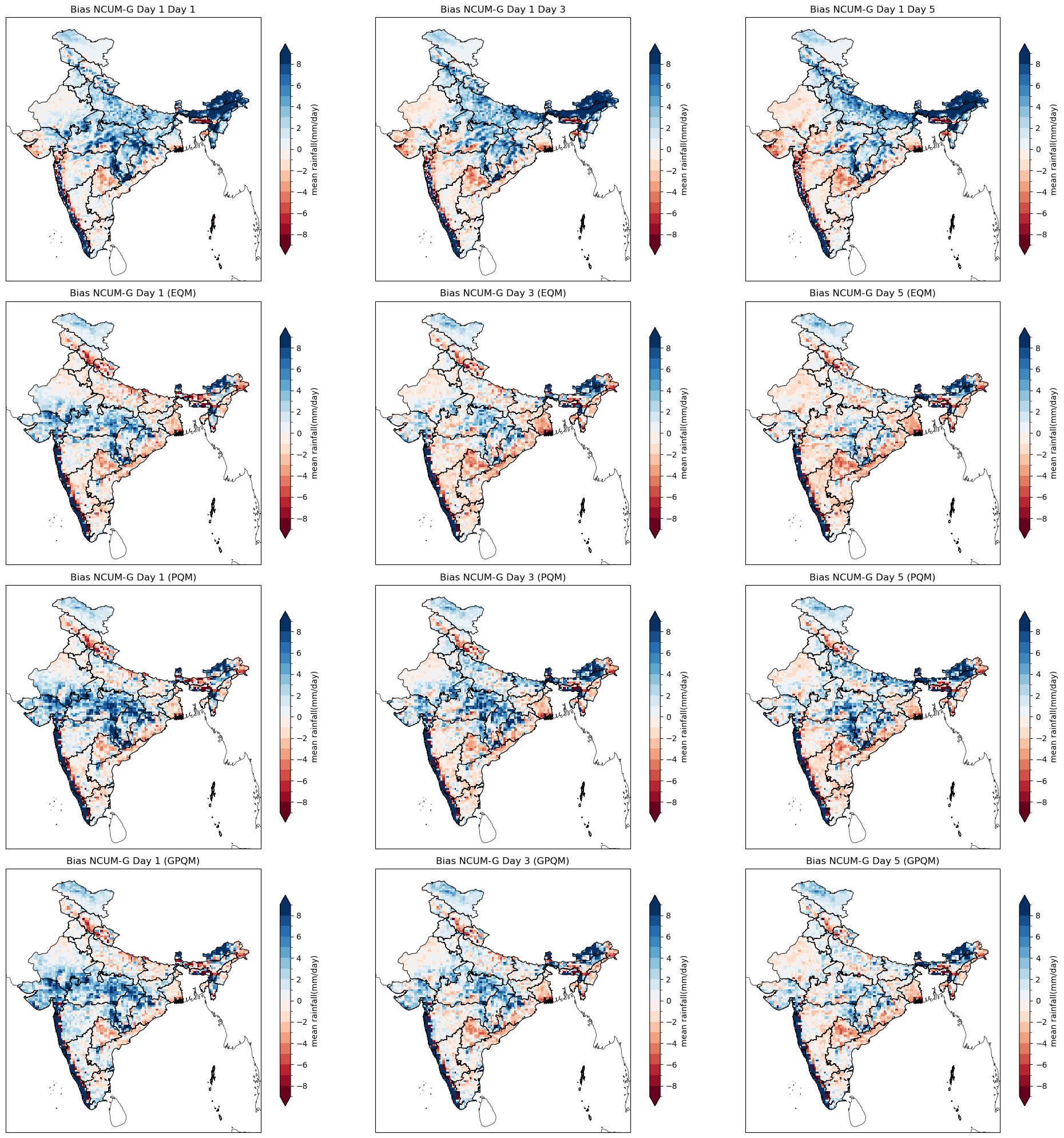


Fig. 12: Spatial distribution of mean daily rainfall bias (Forecast – IMD-MSG observations) during JJAS 2020–2024 over India for NCUM-G forecasts at Day 1, Day 3, and Day 5 lead times, both before (raw) and after bias correction using EQM, PQM, and GPQM methods. Blue shades indicate overestimation, and red shades indicate underestimation.

To see how much the bias correction methods perform and how they vary across different regions, we looked at the spatial rainfall bias across India during the monsoon months (June to September) for the years 2020 to 2024.

From Figure 12, we can see that the raw NCUM-G forecasts show a strong tendency to overestimate rainfall over orographically influenced areas like the Western Ghats and especially Northeast India. It's also noted that the magnitude of this overestimation increases with forecast lead time, with Day 5 showing the worst (highest and lowest) bias. While in the central and southern parts of the country, it shows more moderate biases, though there are still pockets of unrealistically high rainfall forecasts that aren't seen in the observational datasets.

Similarly, the EQM Bias correction brings a noticeable decrease in widespread overestimation, especially across the Indo-Gangetic plains and peninsular India. However, it's not efficient in addressing extreme rainfall outliers, and NEI regions show persistent biases.

On the other hand, in the case of PQM. It matches observations to a degree as it not only reduces overall bias but also does a better job at reshaping the rainfall distribution, all the while helping to fix those intense rainfall anomalies. Spatial patterns become more realistic and aligned with observations, particularly for Day 3 and Day 5 forecasts. Even then, some overestimations are left, especially in the northeastern hill/valley terrain.

Now, taking a look at GPQM. It gives a cleaner spatial correction across most regions, with solid magnitude and pattern. Biases are dramatically reduced even in NE India, though not fully eliminated, and GPQM maintains realistic gradients without introducing weird artefacts. It is the most robust method across all lead times, especially for longer-range forecasts where biases tend to be worse.

NEI Region stands out as a particularly challenging region. The raw NCUM-G forecasts consistently show heavy overestimation here, and bias intensifying further at longer lead times. This is maybe due to the region’s unique geography and strong orographic influence, which the model struggles to handle realistically.

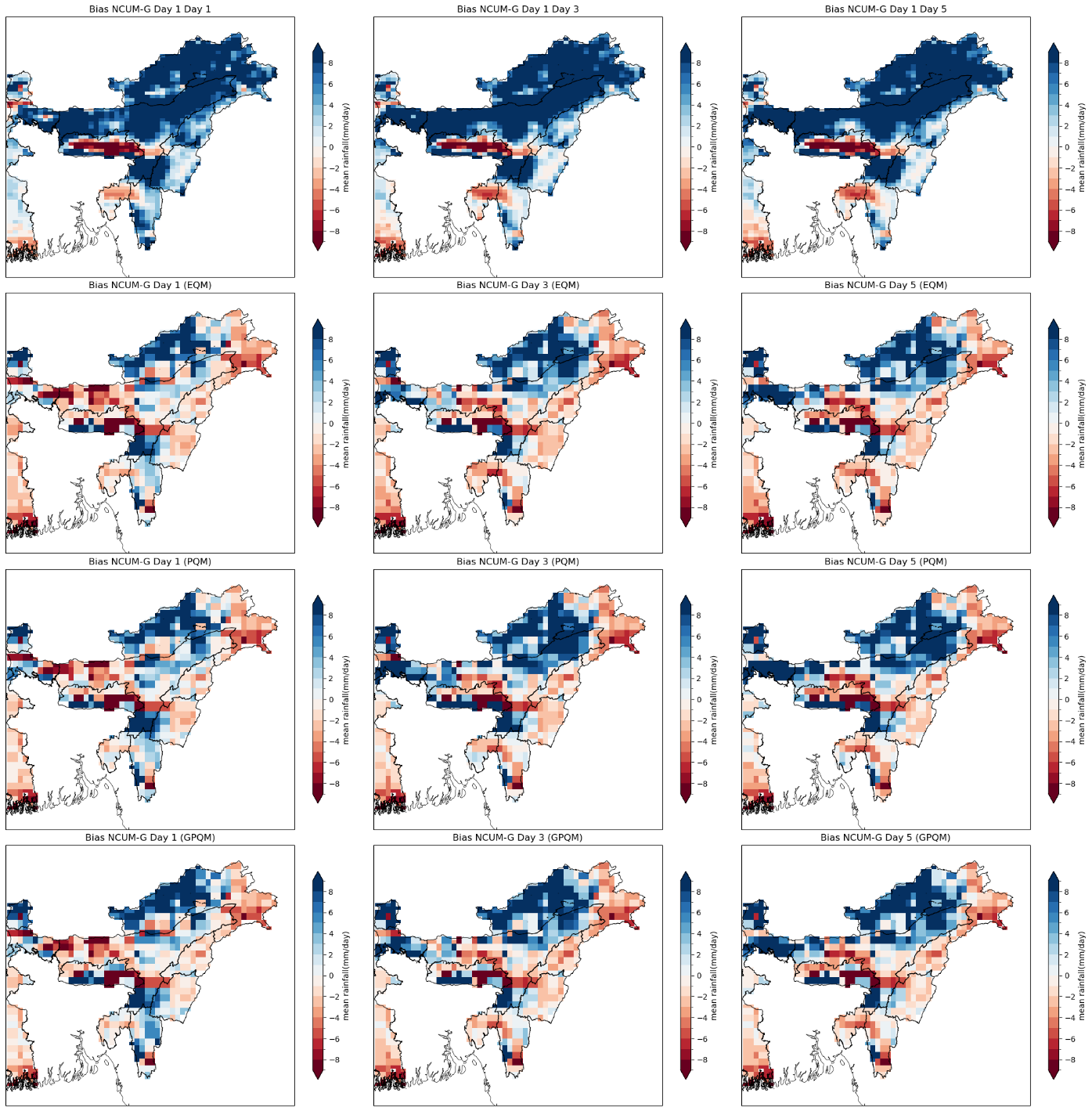


Figure 13: Spatial distribution of mean daily rainfall bias (Forecast – IMD-MSG observations) during JJAS 2020–2024 over Northeast India for NCUM-G forecasts at Day 1, Day 3, and Day 5 lead times, both before (raw) and after bias correction using EQM, PQM, and GPQM. Blue shades indicate overestimation, while red shades denote underestimation of rainfall.

The EQM method brings down the bias in bulk rainfall, but it struggles with the extreme events. Although the bias is slightly reduced, large patches of local overestimation remain untouched, like in hilly zones like Arunachal and parts of Assam. PQM does a better job, reshaping the distribution and reducing the intensity of those inflated rainfall signals. Still, some spatial noise lingers, particularly on Day 5 forecasts.

GPQM is a few steps ahead. It not only brings the magnitude down to more reasonable levels but also smooths out those overly sharp gradients that make the raw and EQM-corrected maps look messy. For NEI, GPQM shows the most spatially consistent correction, with reduced hotspots and better alignment to observed rainfall distribution, even at longer lead times. Given how tricky the NEI region is to forecast, GPQM comes across as the most promising bias correction approach for this region.

* + 1. **Cumulative Distribution Function**

To get a better understanding, we compare the Cumulative Distribution Function of Day 1, Day 3, and Day 5 forecasts to get a comparison of how the raw and bias-corrected NCUM forecasts align with observed rainfall.

Looking at Figure 14, we can see that across all lead times, the raw forecasts show a consistent deviation from observations, mainly in the lower and middle rainfall ranges, which means there is a general overestimation of rainfall events. This increases with lead time, and is especially noticeable in Day 5 forecasts.

Among the bias correction techniques, GPQM demonstrates the best overall alignment with the observed CDF across all lead times. It effectively narrows the gap between the model and observed distributions, especially in the critical mid-range rainfall values, where forecast errors are most sensitive. EQM and PQM offer moderate improvements but tend to underperform at the tails—either underestimating light rain or failing to fully correct for extreme values.

Notably, GPQM’s ability to maintain a close fit even on Day 5 highlights its robustness, especially under conditions of increasing uncertainty. These results further support GPQM as the most reliable method for statistical post-processing of NCUM rainfall forecasts in this study.

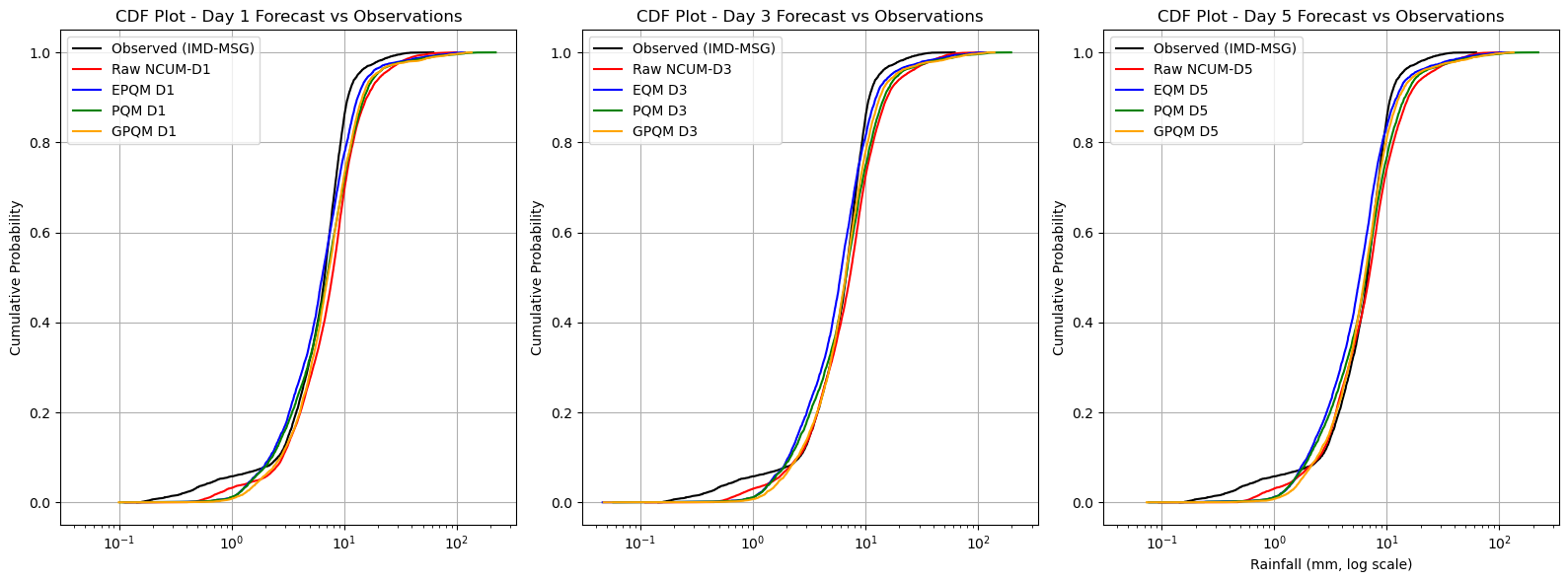


Figure 14: Cumulative Distribution Function (CDF) plots comparing observed daily rainfall (IMD-MSG) with NCUM forecasts (raw and bias-corrected using EQM, PQM, and GPQM) for Day 1, Day 3, and Day 5 lead times during the monsoon season (JJAS 2020–2024). Rainfall values are plotted on a logarithmic scale.

* + 1. **Forecast verification**

In this part we try to compare the performance of raw and bias-corrected rainfall forecasts using categorical verification metrics we discussed previous chapters like Probability of Detection (POD), False Alarm Ratio (FAR), Bias, and Equitable Threat Score (ETS). These were computed for different lead times (Day 1 to Day 5) over the whole India and the NEI region.

The raw NCUM\_G consistently showed higher POD and ETS values, especially at shorter lead times, indicating good detection of rainfall events. But this is not true at all, as this product has high Bias and FAR, meaning the model tended to overpredict rainfall consistently and produced many false alarms.

While looking at EQM, while Bias decreases compared to the raw forecast, generally underperformed compared to other methods in terms of POD and ETS, may be due to its tendency to overly suppress rainfall extremes.

One thing to note is that PQM is skilled in maintaining detection skill and reducing false alarms, as evidenced by its relatively high POD and ETS alongside lower Bias.

Even then GPQM is the most effective at reducing overestimation, yielding the lowest Bias and FAR. It is noted that even when ETS of the Bias corrected models is lower its not much less than the Raw forecast.

Forecast skill is seen to decrease with increasing lead time for all models and correction methods, as expected. However, the corrected models, especially with PQM and GPQM, maintained relatively better skill even at longer lead times, indicating that statistical post processing can effectively extend the utility of numerical forecasts.

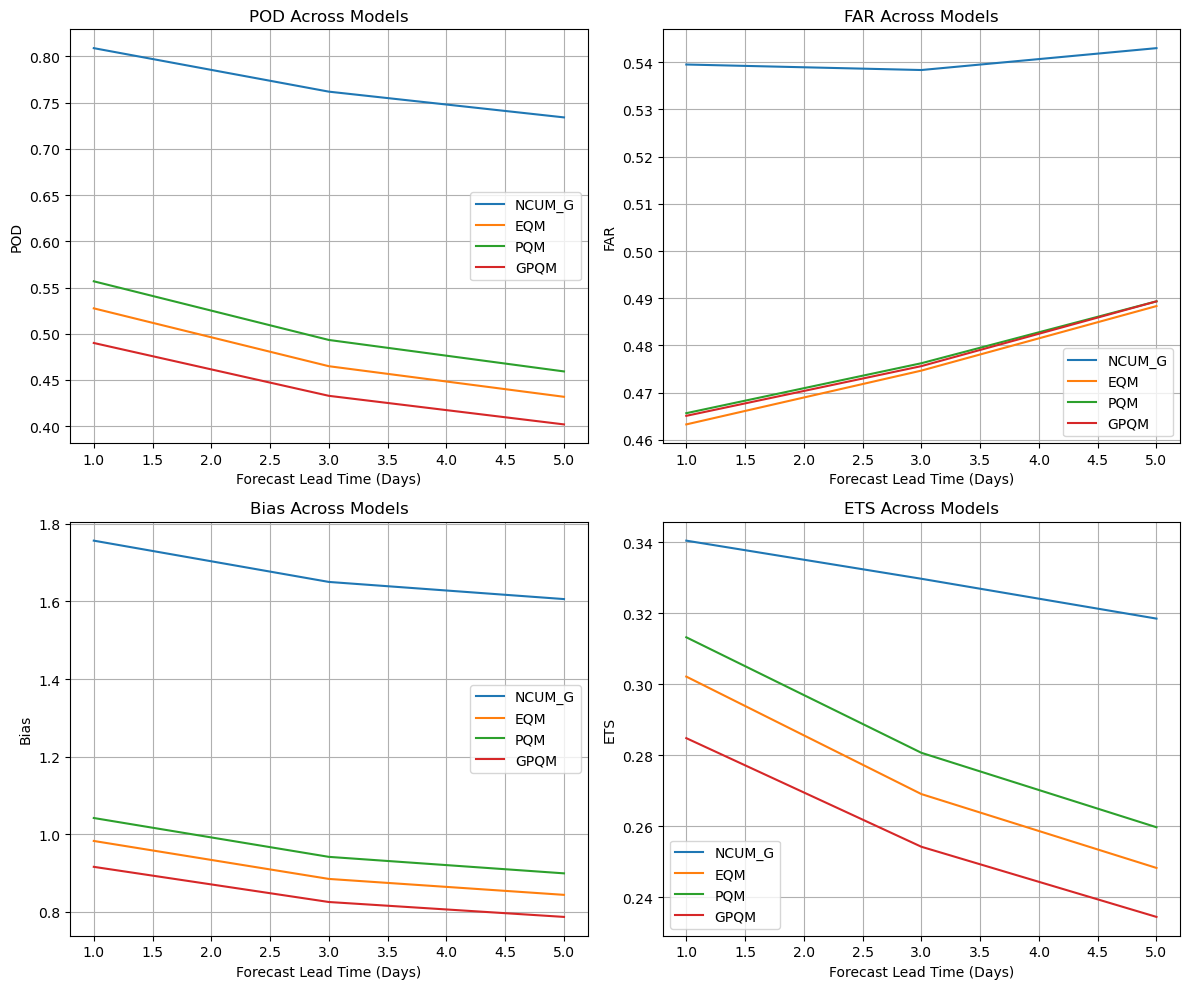


Figure 15: The spatial distribution of four categorical verification metrics—Probability of Detection (POD), False Alarm Ratio (FAR), Bias, and Equitable Threat Score (ETS)—for daily rainfall forecasts over the Indian landmass during the monsoon seasons from 2020 to 2024. Metrics are shown for Days 1, 3, and 5 lead times across four different forecast versions: raw NCUM-G output (first row) and outputs bias-corrected using Empirical Quantile Mapping (EQM), Parametric Quantile Mapping (PQM), and Generalised Pareto Quantile Mapping (GPQM).

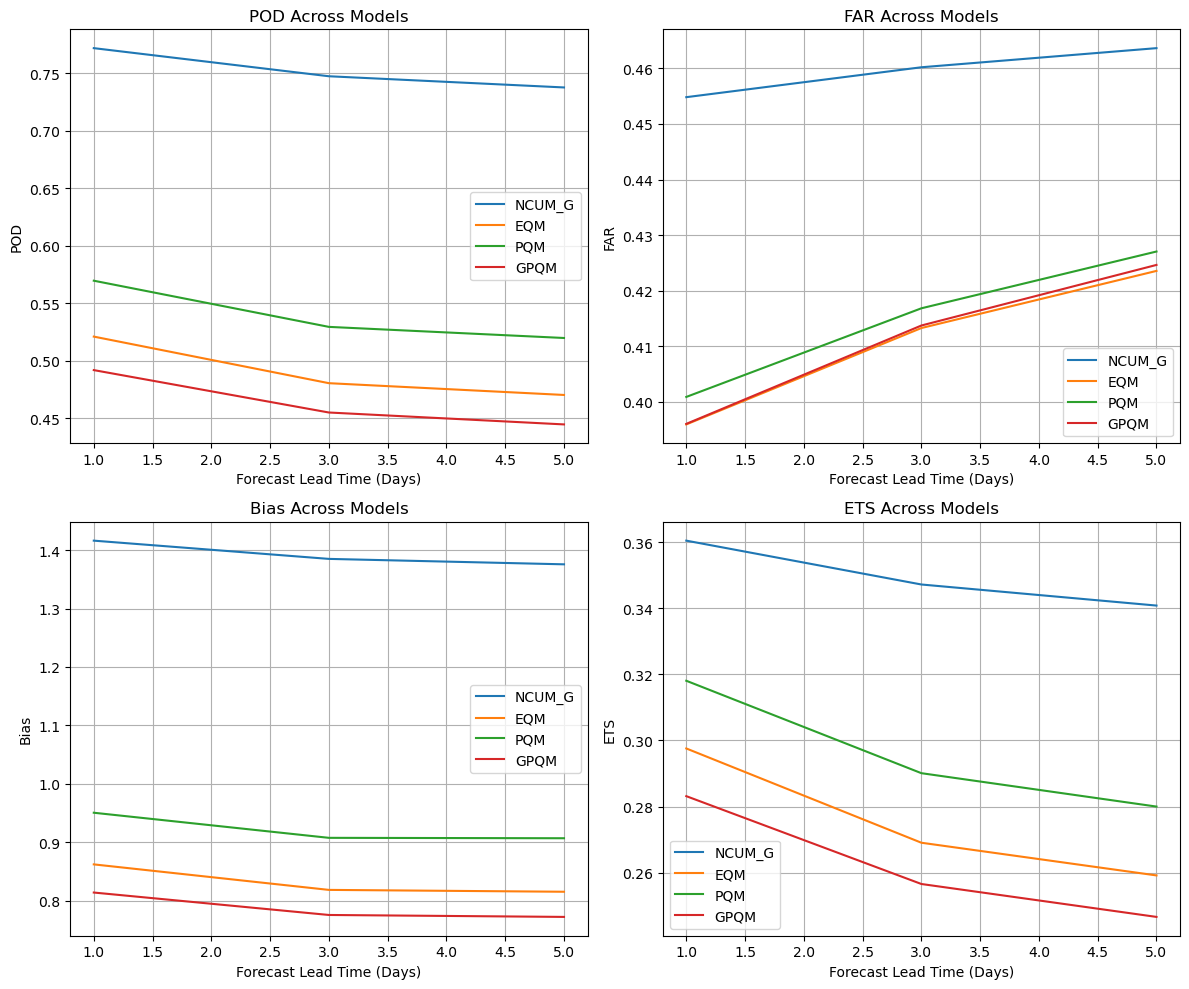


Figure 16: The spatial distribution of four categorical verification metrics—Probability of Detection (POD), False Alarm Ratio (FAR), Bias, and Equitable Threat Score (ETS)—for daily rainfall forecasts over the NEI region during the monsoon seasons from 2020 to 2024. Metrics are shown for Days 1, 3, and 5 lead times across four different forecast versions: raw NCUM-G output (first row) and outputs bias-corrected using Empirical Quantile Mapping (EQM), Parametric Quantile Mapping (PQM), and Generalised Pareto Quantile Mapping (GPQM).

A similar analysis was conducted for the NEI region, Figure 16, which is characterised by complex terrain and intense monsoonal rainfall, often making it a challenging domain for numerical weather prediction as discussed before. The verification results for NEI were consistently better across all metrics compared to the all-India domain.

The raw NCUM\_G model in this region showed higher POD and ETS and lower FAR and Bias than for India as a whole.

GPQM again yielded the lowest FAR and Bias, showing strong capability in removing systematic model overestimations. While PQM is also great by in maintaining high POD and ETM. An in case of EQM, improved the Bias, slightly reduced POD and ETS compared to PQM can be seen.

Over NEI region, the decline in forecast skill with lead time is less pronounced. which means that bias correction methods are more effective in this region. The improved performance also reflects the suitability of GPQM for high-rainfall regions, where preserving the distributional characteristics of rainfall is essential.

* + 1. **Symmetric Extremal Dependence Index (SEDI)**

The Symmetric Extremal Dependence Index (SEDI) is a monitoring metrics for precipitation forecasts. This score is a complementary assessments of forecast performance, by quantifying general performance in the prediction of dry higher threshold events. This metric can handle the common limitations of conventional verification scores (e.g., FAR, POD, CSI) that often degrade under low base rates, a typical scenario for extreme events.

SEDI score is computed using the formula:

Where POD and FAR are derived from a contingency table based on a binary threshold of extreme rainfall. SEDI values range from -1 to +1, where values closer to +1 imply high skill in detecting extremes and values near 0 or negative indicate little to no discrimination skill.

This makes SEDI especially suitable for evaluating model performance on the tails of precipitation distributions.

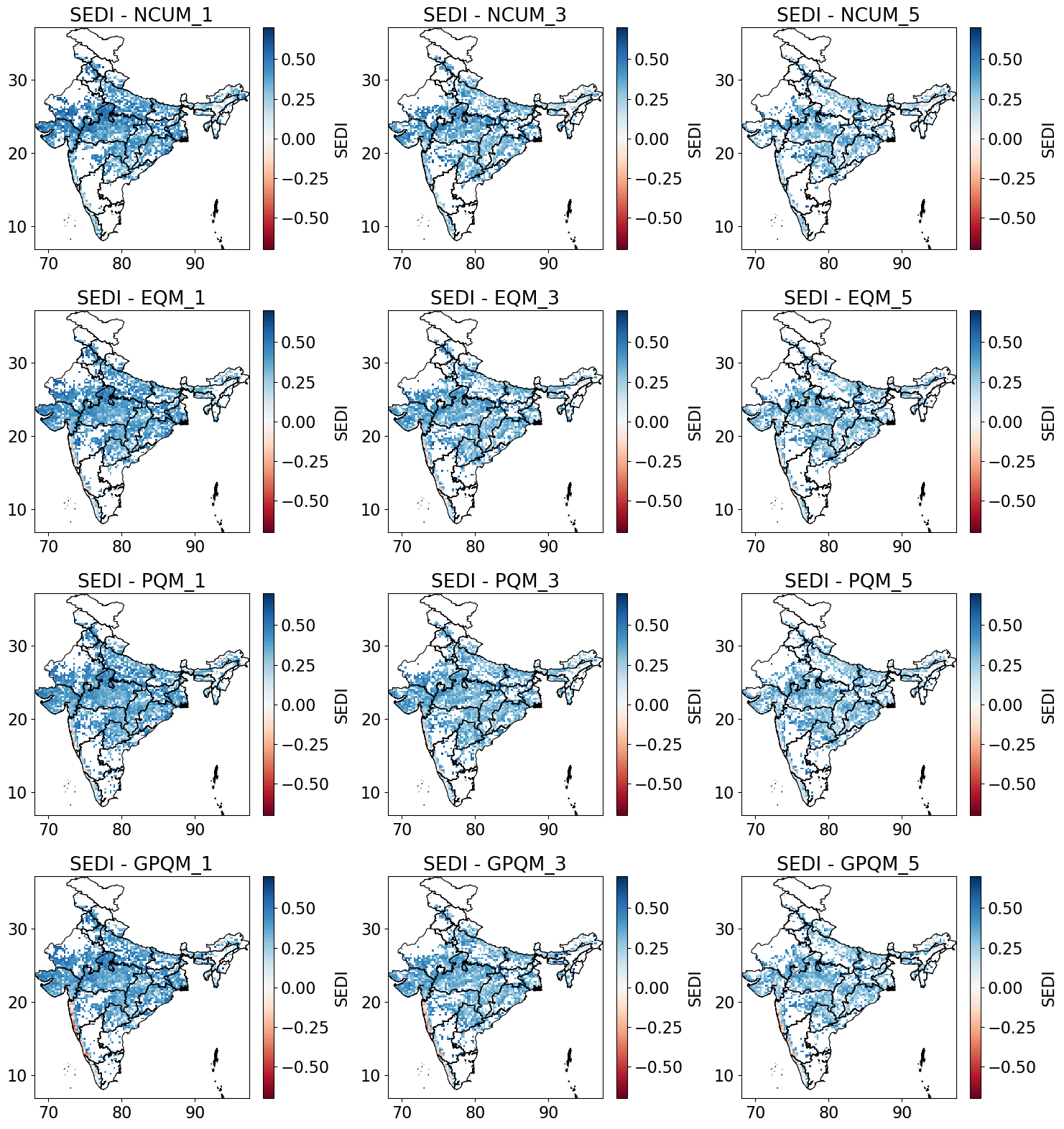


Figure 17 : Spatial distribution of Symmetric Extreme Dependency Index (SEDI) scores for heavy rainfall events (≥ 64.5 mm/day) over India during JJAS (2020–2024).  
Raw NCUM-G forecasts and their bias-corrected versions using EQM, PQM, and GPQM across lead times of Days 1, 3, and 5. Higher SEDI scores are shown as blue and negetive SEDI scores are shown as red

Utilising this Equation spatial maps of the SEDI metrics are generated for 1-day, 3-day, and 5-day accumulated rainfall forecasts with threshold 64.5 mm rain, from this skill of raw and post-processed model outputs in detecting extreme rainfall events over India.

The Raw NCUM outputs exhibit consistently lower SEDI values over central, southern, and eastern regions of India. This pattern is more prominent in lower lead times (1-day), with large spatial extents showing SEDI values near or below zero, suggesting a lack of discrimination skill in capturing rainfall extremes. This reinforces the need for statistical post-processing to improve extreme event detection.

When looking at EQM, it shows increase in SEDI values over most parts of India, particularly in central, eastern, and southern regions. Shows a slight but steady decrease in SEDI from 1-day to 5-day forecasts.

PQM meanwhile yields a smoother spatial SEDI patterns withimprovements in north-central and southern India, though slightly weaker in the northeast. Shows increasing SEDI with longer lead times, similar to EQM. It’s also to be noted that the SEDI value decreases beyond zero near western ghats region.

GPQM records the highest SEDI values across most regions compared to other methods, particularly the NEI region, and core monsoon region, which are areas of high importance as prone to frequent extremes. SEDI improves significantly with increasing lead time compared to others, making GPQM especially valuable for 3–5day forecasts relevant to flood preparedness and impact forecasting.

This analysis also shows the value of bias correction methods in improving the detection of extreme rainfall events in the NCUM model, especially in NEI region. While all three methods enhance SEDI scores over raw outputs,GPQM consistently outperforms,especially for Day 3 and Day 5 forecasts it is very powerful.

* 1. **What that means**

In this section we looked into how well different bias correction methods — specifically Empirical Quantile Mapping (EQM), Parametric Quantile Mapping (PQM), and Generalised Quantile Mapping (GPQM) — can fix errors in regional rainfall forecasts from the NCUM-G and NCUM-R models. With help of multiple verification tools like mean rainfall plots, bias maps, SEDI analysis, CDF plots and different forecast verification tools to compare how close the corrected data came to the actual observations from the IMD-MSG dataset. Out of the three methods, even though EQM managed to give a really good prediction in low and moderate intensity rainfall, **GPQM** turned out to perform the best overall — it reduced bias the most and gave rainfall distributions that matched up more closely with what actually happened. This basically shows that using statistical bias correction can really help in making model rainfall forecasts more accurate and realistic, especially during the monsoon period over India. It's a solid step towards bridging the gap between model output and real-world rainfall patterns.

1. **Conclusion**

In this study we mainly focused on understanding how well regional rainfall forecasts can be improved using both dynamical and statistical downscaling methods. In the first section, we looked into NCUM-R, which is a high-resolution regional model dynamically downscaled from the global NCUM-G. NCUM-R did a better job than NCUM-G when it came to capturing the monsoon rainfall patterns over India, but there were still some gaps, especially when comparing it with the observed IMD-MSG data. And we also understood how biased the global model NCUM-G actually is.

Later, to make the global model NCUM-G actually viable for climatology and forecast purposes, we and tried correcting it using three different statistical bias correction methods: Empirical Quantile Mapping (EQM), Parametric Quantile Mapping (PQM), and Generalised Quantile Mapping (GPQM). After applying these corrections, we used different verification tools to check how well the corrected data matched up with observations. Among the three methods, GPQM showed the most consistent improvements, especially in reducing bias and giving a more realistic distribution of rainfall even in extreme rainfall conditions.

All in all, this study shows that while dynamical downscaling (like NCUM-R) is already a big step forward, and how using bias corrections on coarse global models like NCUM-G, can bring them much closer to reality. Making them more usable for practical weather applications.

1. **Future Scope**

The next step in this research will be to apply AI-based downscaling techniques to generate high-resolution rainfall forecasts. With the availability of bias-corrected datasets from models like NCUM-G, there's now a solid dataset to train data-driven models based on Convolutional Neural Networks (CNNs). These models have shown promise in recent studies for capturing non-linear relationships in weather data and can potentially outperform traditional methods in specific cases, particularly when it comes to local rainfall variability and extreme events.

Another important direction is merging multiple data sources like satellite rainfall, rain gauges, and DEMs to create richer training datasets for these AI models. This kind of fusion can help the model learn more spatial and topographical patterns, which is especially useful for a complex region like India.

Eventually, the goal is to build a deep learning-based downscaling system that can take in bias-corrected global model forecasts and output high-resolution, more realistic rainfall predictions. Such a system could be integrated into existing forecasting pipelines and help improve impact-based forecasting, especially during the monsoon season when accurate predictions are most needed.

1. **References**

Ashrit, Raghavendra, A. K. Mitra, G. R. Iyengar, Sushant Kumar, Arun Kumar, S. K. Dutta, V. R. Durai, John P. George, and E. N. Rajagopal. 2020. “IMDAA: High-Resolution Satellite-Era Reanalysis for the Indian Monsoon Region.” *Journal of Earth System Science* 129 (1): 68–68. https://doi.org/10.1007/s12040-019-1306-3.

Atkinson, Peter M. 2013. “Downscaling in Remote Sensing.” *Spatial Statistics for Mapping the Environment* 22 (June):106–14. https://doi.org/10.1016/j.jag.2012.04.012.

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert. 2012. “HESS Opinions ‘Should We Apply Bias Correction to Global and Regional Climate Model Data?’” *Hydrology and Earth System Sciences* 16 (9): 3391–3404. https://doi.org/10.5194/hess-16-3391-2012.

Hertig, Elke, Douglas Maraun, Judit Bartholy, Rita Pongracz, Mathieu Vrac, Ileana Mares, José M. Gutiérrez, Joanna Wibig, Ana Casanueva, and Pedro M. M. Soares. 2019. “Comparison of Statistical Downscaling Methods with Respect to Extreme Events over Europe: Validation Results from the Perfect Predictor Experiment of the COST Action VALUE.” *International Journal of Climatology* 39 (9): 3846–67. https://doi.org/10.1002/joc.5469.

Hewitson, B. C., and R. G. Crane. 1996. “Climate Downscaling: Techniques and Application.” *Climate Research* 7 (2): 85–95

.

Kumar, K. Niranjan, Sukhwinder Kaur, Mohana S. Thota, M. Venkatarami Reddy, Harvir Singh, Sushant Kumar, Anumeha Dube, and Raghavendra Ashrit. 2024. “NCUM Global Model Verification: Pre-Monsoon (MAM) 2024.” Noida, India: National Centre for Medium-Range Weather Forecasting (NCMRWF). https://www.ncmrwf.gov.in/Reports-eng/NCUMG\_MAM2024.pdf.

Niranjan Kumar, Kondapalli, Mohana Satyanarayana Thota, Raghavendra Ashrit, Ashis K. Mitra, and Madhavan Nair Rajeevan. 2022. “Quantile Mapping Bias Correction Methods to IMDAA Reanalysis for Calibrating NCMRWF Unified Model Operational Forecasts.” *Hydrological Sciences Journal* 67 (6): 870–85. https://doi.org/10.1080/02626667.2022.2049272.

Rani, S. Indira, T. Arulalan, John P. George, E. N. Rajagopal, Richard Renshaw, Amanda Maycock, and M. Rajeevan. 2021. “Evaluation of IMDAA High-Resolution Reanalysis Rainfall Products over India During Monsoon Seasons.” *Journal of Geophysical Research: Atmospheres* 126 (5). https://doi.org/10.1029/2020JD033454.

Rummukainen, Markku. 2010. “State-of-the-Art with Regional Climate Models.” *WIREs Climate Change* 1 (1): 82–96. https://doi.org/10.1002/wcc.8.

Suarez-Gutierrez, Laura, Sebastian Milinski, and Nicola Maher. 2021. “Exploiting Large Ensembles for a Better yet Simpler Climate Model Evaluation.” *Climate Dynamics* 57 (9): 2557–80. https://doi.org/10.1007/s00382-021-05821-w.

Velasquez, Patricio, Marco Messmer, and Christoph C. Raible. 2019. “A New Bias-Correction Method for Precipitation over Complex Terrain Suitable for Different Climate States.” *Geoscientific Model Development* 12 (12): 5337–58. https://doi.org/10.5194/gmd-12-5337-2019.

WANG, Yuqing, L. Ruby LEUNG, John L. McGREGOR, Dong-Kyou LEE, Wei-Chyung WANG, Yihui DING, and Fujio KIMURA. 2004. “Regional Climate Modeling: Progress, Challenges, and Prospects.” *Journal of the Meteorological Society of Japan. Ser. II* 82 (6): 1599–1628. https://doi.org/10.2151/jmsj.82.1599.