# Evaluating Statistical Bias Correction Methods for Improving NCMRWF Unified Model Operational Forecasts in Support of iFLOWS-Mumbai

**Sukhwinder Kaur*, Kondapalli Niranjan Kumar*, Mohana Satyanarayana Thota & Harvir Singh**

National Centre for Medium Range Weather Forecasting, Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201 309, INDIA
Email: sukhwinder@ncmrwf.gov.in, niranjan.kondapalli@gov.in

***ABSTRACT***

*The study addresses systematic errors in the forecasts of a numerical weather prediction (NWP) model, particularly in critical variables such as precipitation that have significant societal implications. Correcting these errors is imperative for enhancing the accuracy of the NWP model in flood risk management decisions. This research evaluates various quantile mapping (QM) bias correction approaches, employing empirical and parametric methods, to rectify precipitation forecasts generated by the National Centre for Medium-Range Weather Forecasting (NCMRWF) Unified Model, with a focus on Mumbai during the southwest monsoon season. The chosen location is important to the Integrated Flood Warning System (IFLOWS), a key program under the Ministry of Earth Sciences, Government of India, providing early warnings and decision support during flooding. The precipitation forecasts, calibrated using various QM techniques over the Mumbai region, demonstrated significant improvements compared to the raw forecasts, especially for higher thresholds. Particularly noteworthy is the better performance of parametric methods, specifically the Generalized Pareto parametric QM, in surpassing raw forecasts, establishing greater effectiveness for regional-scale flood warning applications during extreme rainfall events. This study highlights the efficacy of QM methodologies in treating precipitation forecasts, contributing valuable insights to the advancement of urban flood modelling, and associated decision-making processes.*

***Keywords:*** *Quantile Mapping (QM), Integrated Flood Warning System (iFLOWS), empirical QM method, parametric QM methods*

## 1. Introduction

Extreme rainfall events have occurred more frequently in recent years across the country, with a profound impact on society and ecosystems (Rajeevan et al., 2008, Goswami et al., 2006; Roxy et al., 2017; Kulkarni et al., 2020, Krishnan et al., 2020). These extreme events draw considerable attention every year during the monsoon season that lasts from June to September (JJAS) in India. Monsoon rainfall, also referred to as Indian Summer Monsoon Rainfall, contributes approximately 80% to the total annual precipitation amount in the Indian subcontinent (Sahai et al., 2003; Kumar et al. 2010) with large spatial-temporal variability. These variations have far-reaching impacts on the lives of numerous people, agriculture, and the GDP of the country (Gadgil 2003; Gadgil & Gadgil, 2006). The monsoon season brings heavy rainfall to India's western coast, with some parts experiencing an average rainfall exceeding 250 cm, primarily attributed to the orographic effect (Rao 1976; Francis and Gadgil 2006). Mumbai, situated on India's western coast, frequently experiences flooding during intense precipitation events. Apart from the orographic effect, offshore vortices, troughs, depressions in the Arabian Sea, and mid-tropospheric cyclones (MTCs) contribute significantly to hazardous rainfall in Mumbai during this period (Rao 1976; Miller and Keshvamurthy 1968; Krishnamurti and Hawkins 1970; Ayantika et al. 2018). Remarkably massive floods hit the city in July 2005, 2017, and, most recently, 2022 (Jenamani, et al., 2006; Kumar et al., 2008). Therefore, having prior knowledge of urban floods triggered by extreme rainfall events is vital for mitigating associated social and economic risks in such climate-sensitive locations.

Consequently, the Ministry of Earth Sciences of the Government of India, in partnership with the Disaster Management Department, Municipal Corporation of Greater Mumbai (MCGM), Government of Maharashtra, has implemented an

integrated flood warning system on 12th Jun 2020 for Mumbai called iFLOWS-Mumbai. Mumbai is India's second city to adopt such a system, following Chennai. The iFLOWS system comprises seven modules: a decision support system, vulnerability, risk, flood, inundation, dissemination, and data integration. The flood warning system makes use of a Geographic Information System. This system consists of NWP models from the National Centre for Medium-Range Weather Forecasting (NCMRWF) and the India Meteorological Department (IMD), along with field data collected from the rain gauge network stations established by the Indian Institute of Tropical Meteorology, MCGM, and IMD. Additionally, thematic layers on land use, infrastructure, and other relevant factors are provided by MCGM.

Over the past few decades, there have been significant improvements in global and regional models. But still, the general circulation models are usually inadequate to address the local and regional aspects due to their low resolution. This problem is better solved through downscaling techniques (Themeßl et al., 2011). Dynamical downscaling improves the representation of regional features in NWP forecasts but often suffers from important local inaccuracies owing to insufficient resolution and the uncertainties in the representation of small-scale processes such as clouds, convection, boundary-layer radiative transfer, and internal variability due to their chaotic nature. Consequently, the direct applications of regional models in impact studies are hampered by model biases, especially for essential variables such as precipitation, which have significant societal implications (Cannon et al., 2015; Eden et al., 2012; Munday and Washington, 2018). Therefore, it is imperative to correct the model inaccuracies for effective utilization of the NWP forecasts in decision-making applications related to flood risk management.

In this context, several bias-correction techniques have been adopted to adjust the simulated statistics (i.e., the mean, variance, higher moments, or correlation properties such as lag-one correlation), to those of observations (Déqué, 2007; Piani et al., 2010; Amengual et al., 2012; Berg et al., 2012; Gudmundsson et al., 2012; Wetterhall et al., 2012; Teutschbein & Seibert, 2012; Gutjahr & Heinemann, 2013; Maraun & Widmann, 2015; Macias et al., 2018; Kim et al., 2019; Yoshikane & Yoshimura, 2022). The “delta change method”, also known as “delta change scaling method” is the simplest bias-correction method, that utilizes scaling factors to adjust model simulations based on an observed baseline. This method adjusts the observed time series by incorporating the difference between future and control simulations (see Figure S1).

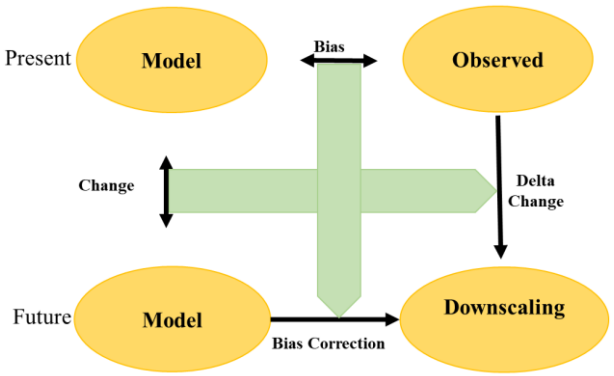


**Figure S1: Visual representation of delta change method.**

Indeed, this method is more commonly employed in climate simulations rather than NWP models. Moreover, it's important to note that the delta change method utilizes changes that model simulates, but the higher-order statistical moments in the observed time series remain unaltered. On the other hand, Widmann & Bretherton (2000) proposed a scaling factor defined as the ratio of present-day mean simulated precipitation to the mean observed precipitation for correcting forecasted precipitation. This method uses a constant coefficient of variation (COV) approach, resulting in both mean and variance being rescaled by the same factor, while higher-order moments in the calibrated data stay unchanged. Accurately representing these higher-order statistical moments, such as skewness and kurtosis, holds significant importance in flood forecasting. These moments influence the shape and tail of the probability distribution of precipitation or runoff, thereby directly impacting flood occurrence. Misrepresentations could lead to inadequate preparedness or unnecessary alarm, particularly in flood-prone regions, where the consequences of

**Table S1. Classification of Rainfall as per IMD**

| Category | 24 hour rainfall over a station |
|---|---|
| Very light rain | Trace - 2.4 mm |
| Light rain | 2.5 – 15.5 mm |
| Moderate rain | 15.6 – 64.4 mm |
| Heavy rain | 64.5 – 115.5 mm |
| Very heavy rain | 115.6 – 204.4 mm |
| Extremely heavy rain | >204.4 mm |

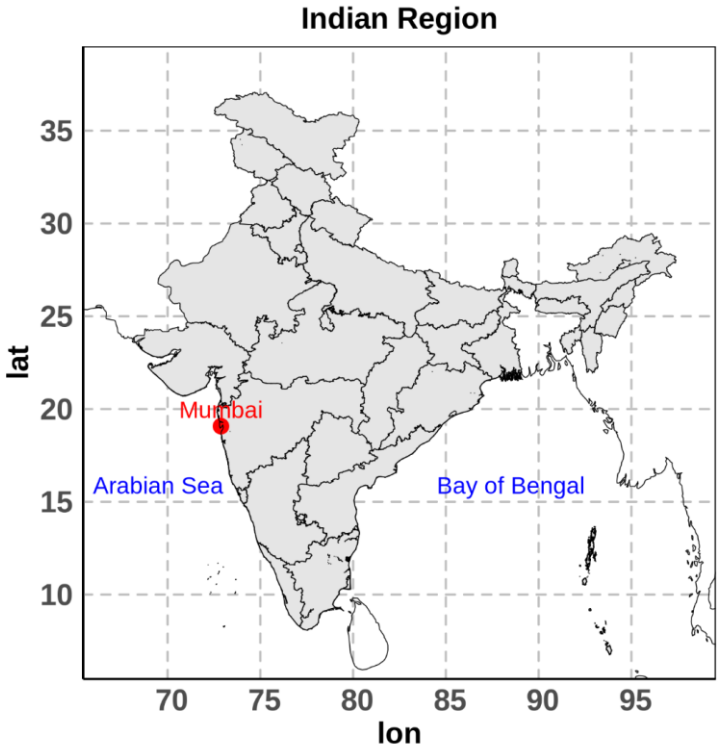


**Figure 1: Map indicating the study area's location.**

inaccurate forecasts can be severe. Therefore, precise calibration and validation of models to capture these aspects is essential for effective flood forecasting and risk mitigation. As a result, recent studies emphasize quantile mapping bias correction (QMBC) methods that adjust distinct quantiles individually to correct the higher-order statistical moments of the forecasted precipitation time series (Gudmundsson et al., 2012; Themeßl et al., 2012; Cannon et al., 2015; Kim et al. 2019). Furthermore, these techniques outperform other methods, particularly when it comes to precipitation data (Teutschbein and Seibert 2012, Themeßl et al., 2012; Maraun & Widmann, 2015; Fang et al., 2015; Kim et al. 2019). Therefore, this study assesses the performance of various QMBC methods in improving rainfall forecasts of the global NCMRWF Unified Model (NCUM-G) operational model over Mumbai to support the MoES iFLOWS program.

The article is organized as follows. An overview of the study area and data used is provided in section 2. Section 3 describes the various statistical methods, including empirical and parametric quantile mapping, utilized in this study. The results and subsequent discussion are detailed in Section 4, followed by a summary and conclusion in Section 5.

## 2. Study area and Data

### 2.1 Study area

This study focuses on the Santacruz station ( 72.85 0E, 19.117 0N) in Mumbai (See Fig. 1), situated on the west coast of India. Mumbai's rapid urbanization and infrastructure expansion render it

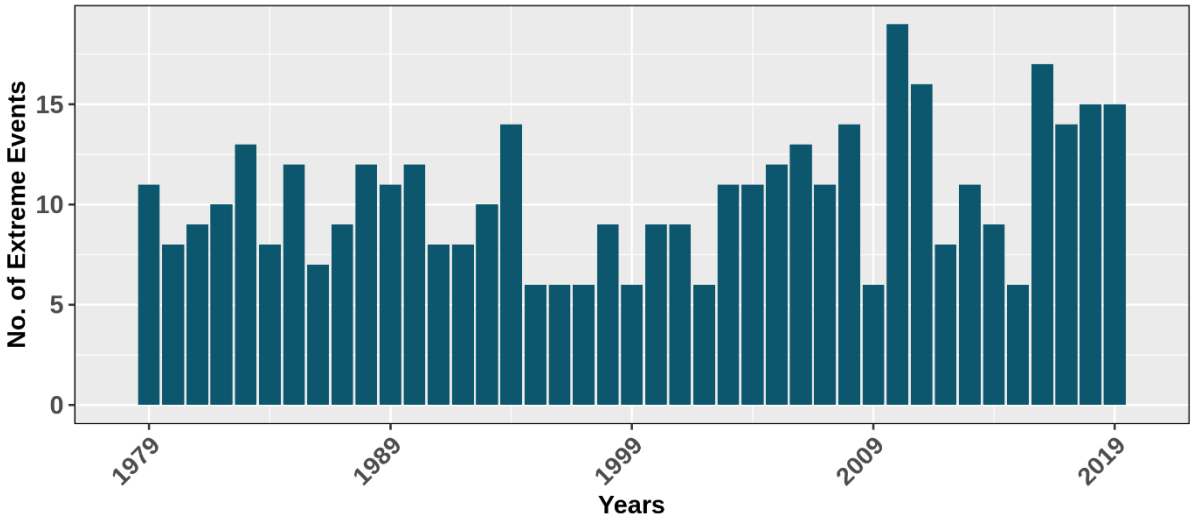


**Figure 2: Frequency of the extreme annual rainfall events (>64.5mm/day) at Santacruz station in Mumbai based on the IMD station data for the period 1979-2019.**

highly vulnerable to both natural and man-made disasters. Being situated on the windward side of the Western Ghats of India, Mumbai receives high-magnitude and intense rainfall during the southwest monsoon season owing to the orographic effect (Francis and Gadgil 2006, Singh et al. 2017). The selection of Mumbai as the study area is motivated by the fact that the region has experienced the most severe extreme precipitation events (defined as exceeding 64.5 mm/day according to IMD criteria, as detailed in supplementary Table S1) in recent decades (see Fig. 2). Therefore, accurate precipitation predictions over Mumbai are crucial to facilitate decision-making applications related to flood risk management.

### 2.2 Data

This study employs a high-resolution Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis, covering the period from 1979 to 2019 (Ashrit et al., 2020; Rani et al., 2021). Additionally, an IMD observed station dataset spanning the same timeframe (1979 to 2019) is utilized (Jenamani et al., 2006), along with the NCUM-G operational forecast dataset for the southwest monsoon season from 2020-2023 (Sumit Kumar et al., 2020) over the Santacruz station in Mumbai. These datasets form the basis for constructing a bias correction (BC) model using various statistical methods, as detailed in Niranjan Kumar et al. (2022), and subsequently for training and testing the model.

Furthermore, the bias correction model is applied to rectify the NCUM-G operational forecast data, evaluating the effectiveness of the calibration methods for flood applications. It is essential to note that the NWP dynamical core parameterization techniques employed for generating both the IMDAA reanalysis dataset and the NCUM-G are relatively similar. Given their shared model physics, IMDAA proves effective in correcting real-time forecasts from NCUM-G.

## 3. Methodology

### 3.1 Statistical Bias Correction (BC) approaches

Bias Correction (BC) methods function as post-processing tools in numerical modeling, aiming to refine the agreement between the model and observations. Statistical BC approaches develop a functional relationship between observed and simulated variables over a historical period, subsequently utilizing this relationship for model forecasts. The following subsections provide a concise overview of the different quantile methods employed in this study.

#### 3.1.1 Empirical Quantile Method (EQM)

The Empirical Quantile Method (EQM) is a non-parametric statistical BC approach that adjusts mean, standard deviation (variability), and shape errors by computing quantile-by-quantile changes in the simulated cumulative distribution function (CDF) (Boè et al., 2007; Déqué, 2007; Amengual et

al., 2012; Gudmundsson et al., 2012; Niranjan Kumar et al., 2022).

Let "CAL" be the calibrated precipitation acquired after bias correction and "Obs" denote the observed precipitation. Then, for i^th value of CDF, the calibrated precipitation is given as:

$$CAL_i = Obs_i + \bar{\Delta} + \Delta_i' \quad (1)$$

where $\Delta_i$ defines the difference between control ($ctl_i$) and future ($fut_i$) raw simulated precipitation, i.e., $\Delta_i = (fut_i - ctl_i)$. Then, the mean regime shift ($\bar{\Delta}$) can be represented as $\bar{\Delta} = \frac{\sum_{i=1}^{N} \Delta_i}{N}$ and the corresponding deviation from this shift, denoted as $\Delta_i'$, is given by $\Delta_i' = \Delta_i - \bar{\Delta}$.

This method, besides its simplicity and non-parametric nature, has demonstrated effectiveness compared to other distribution and scaling BC methods (Boé et al., 2007; Gudmundsson et al., 2012). However, instabilities emerge at higher quantiles (Gudmundsson et al., 2012; Gutjahr and Heinemann, 2013). Additionally, this method depends on numerous degrees of freedom, leading to non-stationarity for future periods (Piani et al. 2010b, Gutjahr and Heinemann 2013). Therefore, we are also exploring parametric BC approaches discussed below, which rely on a lower degree of freedom.

### 3.1.2 Parametric Quantile Method (PQM)

The Parametric Quantile Method (PQM) is a parametric BC method that employs theoretical distribution rather than empirical distribution. This method adjusts the CDF of the simulated output to the corresponding observed distribution via a transfer function (i.e., two-parameter gamma distributions) (Piani et al., 2010; Niranjan Kumar et al., 2022). The gamma distribution's probability density function (pdf; *y*) is:

$$y = f(x|\xi,\sigma) = \frac{1}{\sigma^{\xi}\Gamma(\xi)} x^{\xi-1} e^{\frac{-x}{\sigma}} \quad (2)$$

Where Γ(.) represent the gamma function. ξ and σ are the shape and scale parameters of the gamma distribution, respectively. x denotes the normalized daily precipitation. The drawback of the PQM method is that there is no restriction on the upper limit, which can lead to false alarms for extreme rainfall events (Gutjahr and Heinemann, 2013). So, in this study, we adopted a new approach that combines the gamma distribution with the Generalized Pareto Distribution (GPD), which we further elaborate on below.

### 3.1.3 Generalized Parametric Quantile Method (GPQM)

The Generalized Parametric Quantile Method (GPQM) is also a parametric BC method that combines gamma and Generalized Pareto Distribution (GPD) (Niranjan Kumar et al., 2022). The pdf of GPD is:

$$y = f(x|\xi,\sigma,\theta) = \left(\frac{1}{\sigma}\right)\left(1 + \frac{\xi(x-\theta)}{\sigma}\right)^{-1-\frac{1}{\xi}} \quad (3)$$
$$for\ \theta < x, when\ \xi > 0, or\ for\ \theta < x < \theta - \frac{\sigma}{\xi}\ when\ \xi < 0$$

where, ξ≠0,σ,and θ are the shape, scale, and threshold parameters of GPD, respectively. Here, the gamma distribution is applied to values below 95th percentile (Yang et al. 2010) and a GPD to values above 95th percentile (Coles 2001). Thus, the GPQM method can be formulated as

$$\begin{cases} f_{obs,gamma}^{-1}\left(f_{IMDAA,gamma}\right), & if\ x < 95th\ percentile \\ f_{obs,GPD}^{-1}\left(f_{IMDAA,GPD}\right), & if\ x > 95th\ percentile \end{cases} \quad (4)$$

The methodology involves the calibration and testing of various quantile methods, as outlined in the flow diagram (Refer to Fig. 3). IMDAA reanalysis data from the grid nearest to the Santacruz station's location and observed station data from IMD, spanning a common period of 41 years (i.e., 1979-2019), are utilized for assessment. The entire 41-year dataset is considered for training and testing the model from 2020-2023 (For more details, readers can refer to Niranjan Kumar et al., 2022). Additionally, the calibration methods are applied to bias-correct the real-time operational NCUM-G forecasts.

### 3.2 Categorical verification scores

The performance of different quantile bias correction methods is evaluated through categorical verification scores, which include metrics such as the probability of detection or hit rate (POD), false alarm ratio (FAR), and Equitable Threat Score (ETS). The categorical approach of

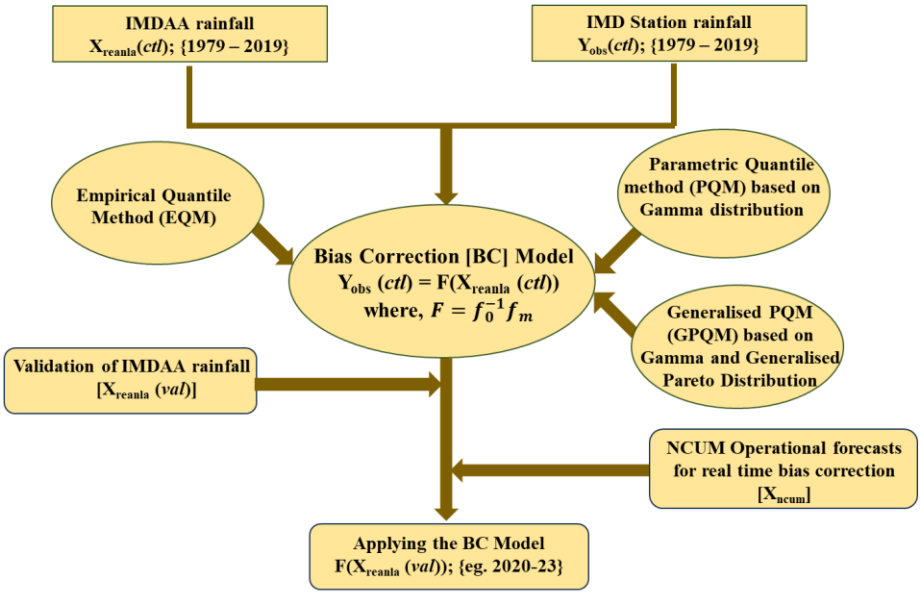


**Figure 3: Flow diagram depicting the methodology adopted in this study.**

**Table 1. Contingency table**

| | | Observed | | Total |
|---|---|---|---|---|
| | | **Yes** | **No** | |
| **Forecast** | **Yes** | Hits | False alarms | Forecast Yes |
| | **No** | Miss | Correct Negative | Forecast No |
| **Total** | | Observed Yes | Observed No | Total |

verifying quantitative precipitation forecast is based on the 2×2 contingency table given below (see Table 1).

In this context, we define a "hit" as an event where the prediction matches the observation on a grid point. Conversely, if an event on a grid point is predicted but not observed, it is termed a "false alarm". A "miss" happens when an event is not predicted but is observed. Lastly, a "correct negative" occurs when an event does not happen and the model does not predict it. Utilizing these components of the contingency table, categorical skill scores are calculated for different rainfall thresholds. The following subsections provide a concise overview of the various categorical verification scores employed in this study.

### 3.2.1 Probability of Detection (POD)

The Probability of Detection (POD) or hit rate measures the model's ability to correctly detect the occurrence of specific weather events. It is calculated as the number of hits divided by the total number of event observations and expressed as:

$$POD = \frac{hits}{hits + misses} \quad (5)$$

Its value ranges from 0 to 1, with 1 indicating a perfect score.

### 3.2.2 False Alarm Ratio (FAR)

The False Alarm Ratio (FAR) represents the proportion of forecasted events that were incorrect. It is computed by dividing the number of false alarms by the total number of forecasted events. Mathematically, it can be expressed as:

$$FAR = \frac{false\ alarms}{hits + false\ alarms} \quad (6)$$

This score ranges from 0 to 1, with 0 indicating a perfect score.

### 3.2.3 Equitable Threat Score (ETS)

The Equitable Threat Score (ETS) is especially useful for assessing deterministic forecasts and is commonly employed to verify rainfall in numerical weather prediction models as it penalizes constant and purely random forecasts heavily (Gandin and

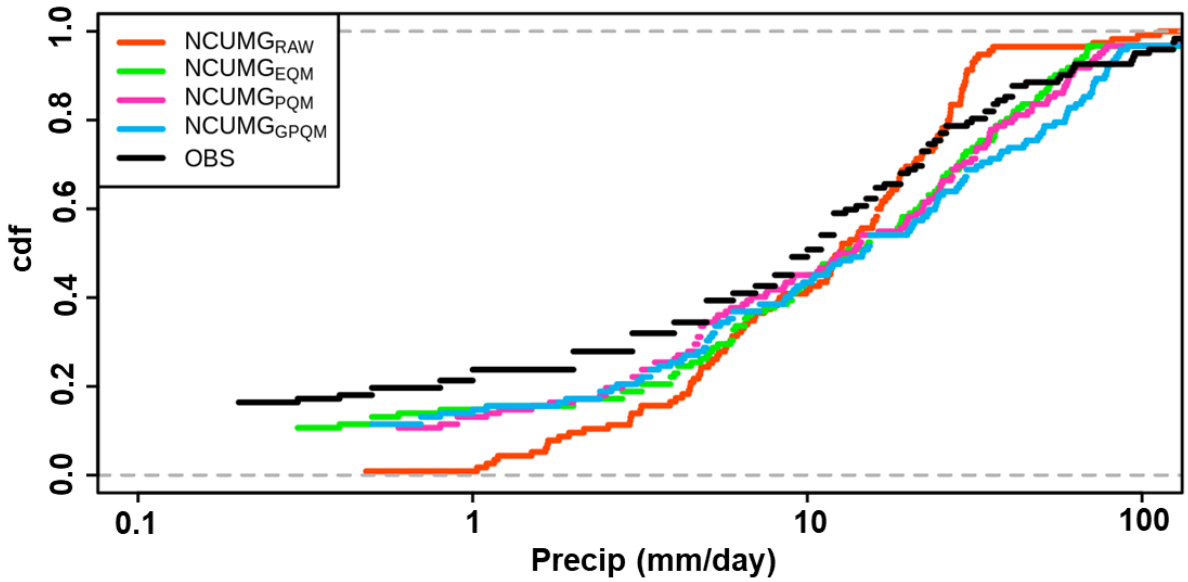


**Figure 4: The CDF of daily precipitation obtained from the IMD station dataset (OBS), NCUM-G Day-01 raw forecast precipitation (NCUMG$_{RAW}$), and bias-corrected Day-01 forecast precipitation based on EQM (NCUMG$_{EQM}$), PQM (NCUMG$_{PQM}$), and GPQM (NCUMG$_{GPQM}$) during the southwest monsoon (JJAS, 2020-23) over Santacruz station .**

Murphy, 1992). ETS quantifies the fraction of observed and/or forecast events that were accurately predicted, considering hits associated with random chance and defined as:

$$ETS = \frac{hits - hits_{rand}}{hits + misses + false\ alarms - hits_{rand}} \quad (7)$$

where,

$$hits_{rand} = \frac{(hits + misses)(hits + false\ alarms)}{total} \quad (8)$$

The range of ETS scores is from -1/3 to 1, where 0 indicates no skill, and a perfect score is 1.

## 4. Results and Discussion

### 4.1 Assessment of QM approaches for NCUM-G forecasts

In this section, we first explore the empirical CDF relevant to the Santacruz station . Fig.4 depicts the CDF of daily precipitation sourced from the IMD station data (OBS), NCUM-G Day-01 raw forecast (NCUMGRAW), and bias-corrected Day-1 forecast precipitation acquired through EQM (NCUMGEQM), PQM (NCUMGPQM), and GPQM (NCUMGGPQM) during the summer monsoon (June–September 2020-23) over the Santacruz station In particular, for higher thresholds (> 10 mm/day), the CDF of NCUMGRAW dataset deviates from the observed distribution. Despite its high spatial resolution, significant biases in extreme rainfall intensities compared to the IMD station dataset are apparent in NCUMGRAW. These biases impede the practical utility of NCUM-G in hydrology at local and regional scales, as well as in flood-related applications. Thus, rectifying these biases is crucial before employing them for urban flood forecasting purposes. To address this, we evaluate the CDF using different BC methods applied to the NCUM-G precipitation data, following the procedures outlined in Section 3. The calibrated precipitation from various QM approaches shows a better alignment with OBS for higher thresholds (Fig. 4).

To obtain more information about the effectiveness of various QM approaches in correcting the NCUM-G precipitation forecasts, we validate the calibrated precipitation acquired through various QM methods against the IMD station dataset. For this, firstly we took the NCUM-G Day-01, Day-03, and Day-05 raw forecasts (NCUMGRAW) during the southwest monsoon season from 2020-23. Subsequently, to calibrate the NCUM-G raw forecasts for the period 2020-23, we utilized the IMDAA and station-based IMD datasets spanning from 1979 to 2019 during southwest monsoon

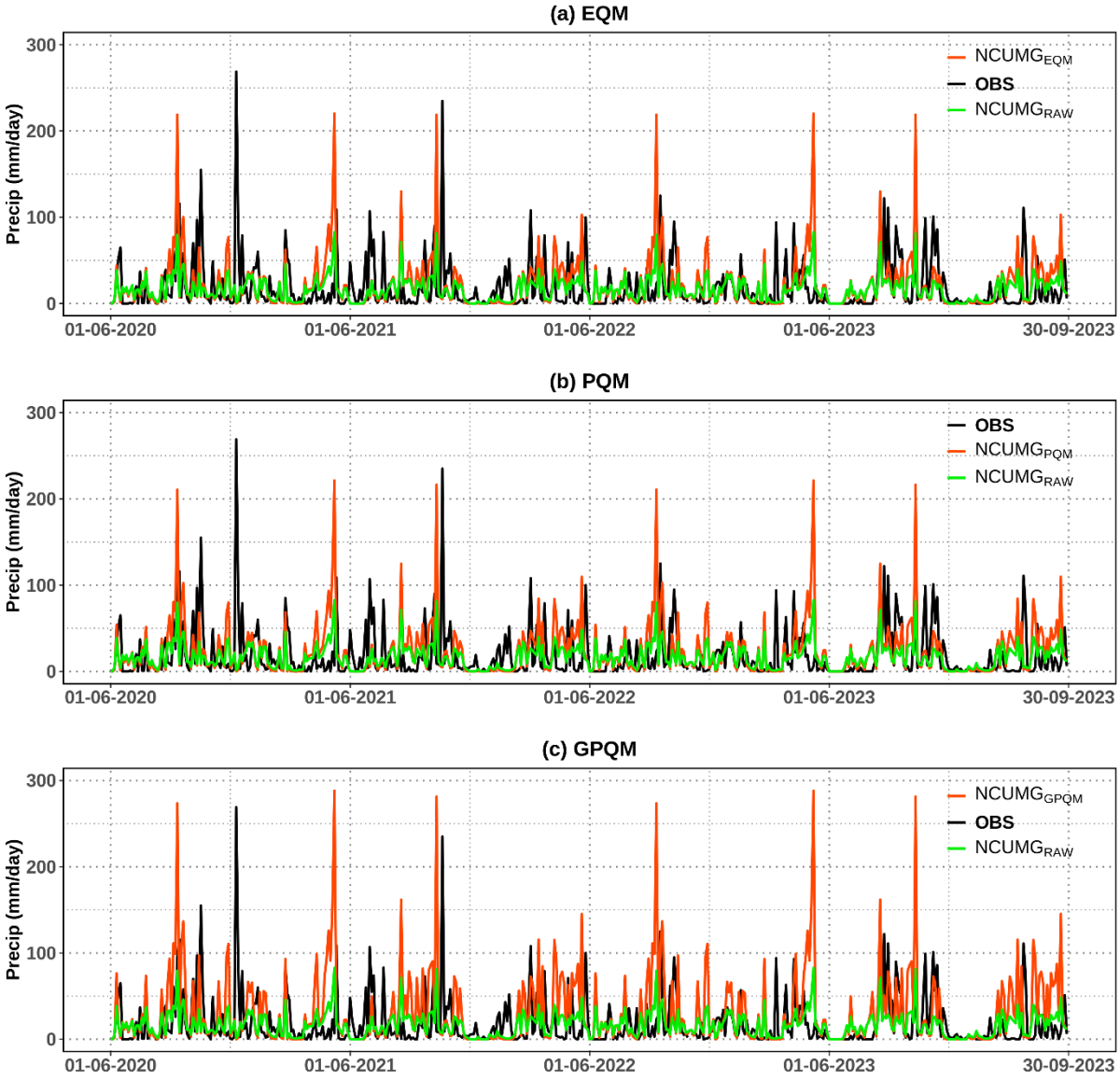


**Figure 5: Time evolution of daily precipitation obtained from IMD station dataset (OBS), NCUM-G Day-01 raw forecast ($NCUMG_{RAW}$), and bias-corrected Day-01 forecast obtained based on (a) empirical quantile methods (EQM; $NCUMG_{EQM}$), (b) parametric quantile method based on gamma distribution (PQM; $NCUMG_{PQM}$), and (c) Generalized PQM based on gamma and GPD (GPQM; $NCUMG_{GPQM}$) during the southwest monsoon (JJAS, 2020-23) over the Santacruz station.**

seasons over the Santacruz station for training purposes. This calibration is conducted based on monthly quantile technique, as elaborated earlier (refer to Methodology). Fig. 5 depicts the validation of Day-01 daily precipitation of NCUM-G raw forecasts (NCUMGRAW) against the IMD station-based dataset (OBS) at Santacruz station during the southwest monsoon season. Additionally, the calibrated precipitation obtained using the empirical method is shown in Fig. 5a (NCUMGEQM) and based on parametric methods in Fig. 5b, and 5c (NCUMGPQM, NCUMGGPQM). The Day-01 daily precipitation time evolution at the Santacruz station demonstrates that the NCUMGRAW failed to capture numerous heavy rainfall events. For instance, if we categorized 64.5 mm/day and above events as heavy to extreme rains (See supplementary Table S1), approximately 36 such events were observed at the Santacruz station during the southwest monsoon season from 2020-23. The NCUMGRAW accurately predicted only 8 events, whereas the empirical method corrected 28 events that were initially missed in the raw forecasts (Fig. 5a). However, the parametric

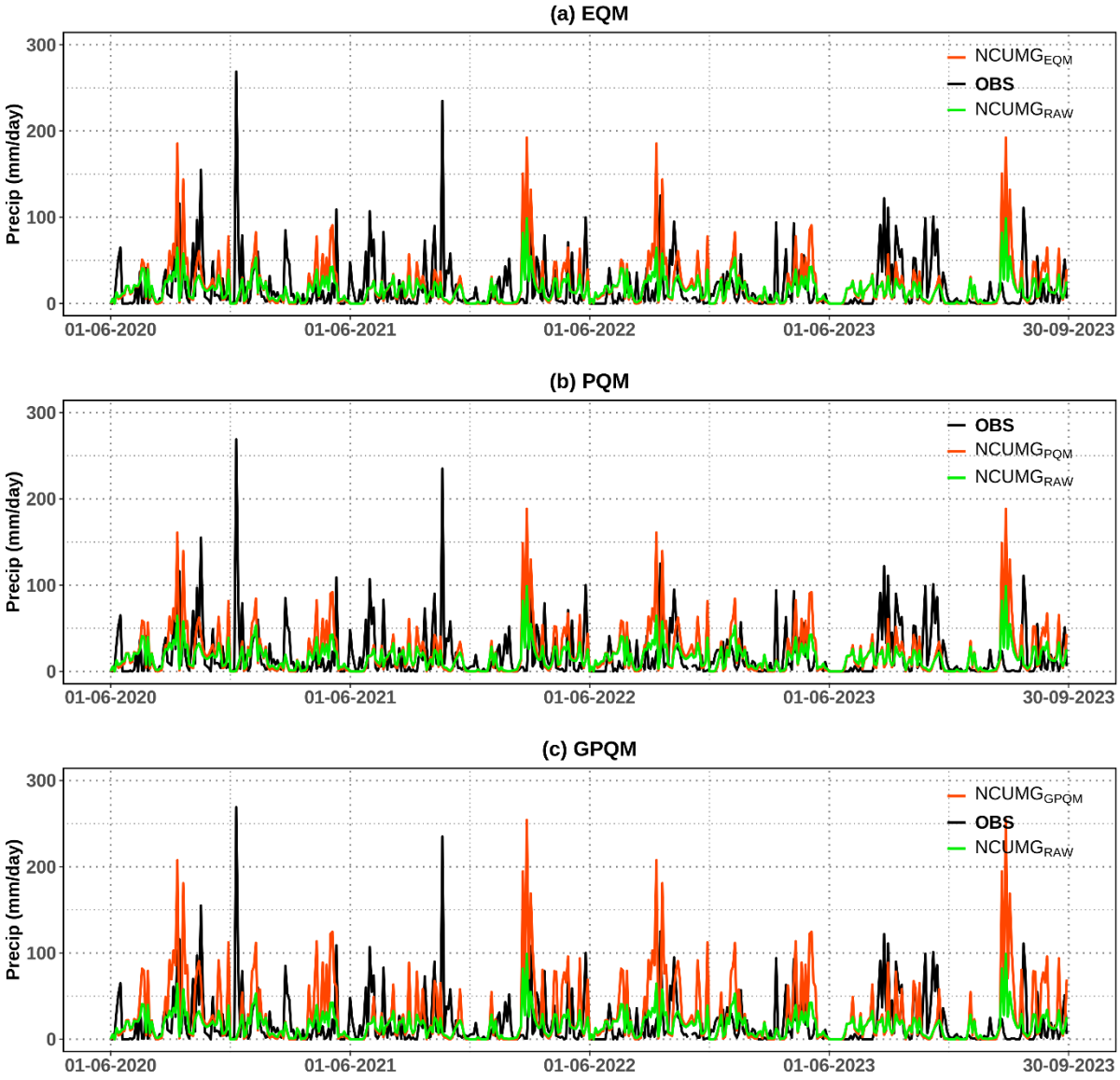


**Figure 6: Same as Fig. 5, but for the Day-03 forecast.**

methods, specifically GPQM, successfully predicted all the events that the raw forecasts missed (Fig. 5c). Similarly, as the forecast lead times progress, the time evolution of daily rainfall indicates that the NCUMG raw (NCUMGRAW) forecasts consistently underestimate the observed rainfall. However, the calibrated methods effectively address these underestimations compared to the NCUMGRAW, particularly for the heavy rainfall events (Fig. 6 and 7).

To further understand how well various QM approaches perform in correcting the biases of raw forecasts across different lead times, the skill of various methods is evaluated using categorical verification scores such as POD, FAR, and ETS. Verification scores for rainfall thresholds of up to 75 mm/day across different forecast lead times are depicted in Figure 8 for the Day-01 forecast (For Day-03 (Fig. S2) and Day-05 (Fig.S3) forecasts, please refer to Supplementary material). The calibrated precipitation exhibits an improvement in POD and ETS compared to the NCUMGRAW. Specifically, the POD and ETS of the calibrated precipitation exhibit higher values for moderate to heavy rainfall events (see Table S1) across all forecast lead times at the Santacruz station (see Supplementary Fig. S2 and S3). Moreover, with increasing forecast lead time, the calibrated methods demonstrate better performance compared to the raw forecasts (Fig. S2 and S3). However, a drawback of the POD score is its susceptibility to

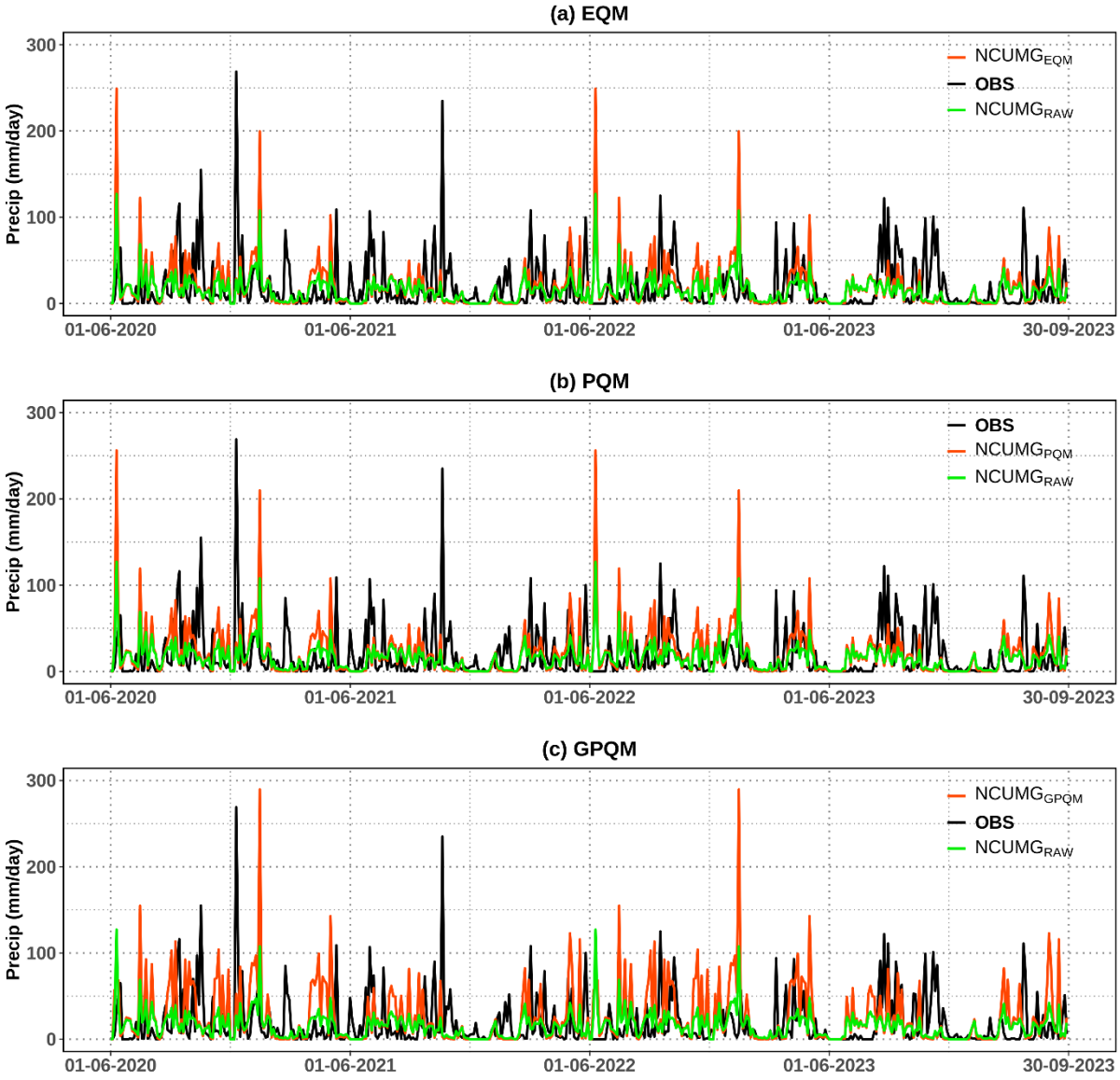


**Figure 7: Same as Fig. 5, but for the Day-05 forecast.**

the number of hits, disregarding false alarms. Therefore, we have also included the FAR score, as depicted in Figures 8(b) for Day-01 forecast (For Day-03 in Fig. S2(b) and Day-05 in Fig.S3(b)). The FAR measures the fraction of events falsely predicted to occur, with a perfect FAR score being zero. In Day-01 to Day-05 forecasts, various QM approaches tend to overestimate false alarms for moderate to heavy rainfall events. However, the overestimation of such events in calibrated precipitation slightly diminishes with forecast lead time, as evidenced by the reduction in FAR score (Figs. 8(b), S2(b) and S3(b)). Therefore, the verification analysis indicates that the QM methods exhibit a considerable improvement in the detection of local extreme rainfall events.

### 4.2 Real time implementation of QM techniques at Mumbai

The Operational implementation of different QM methods to NCUM-G precipitation forecasts has been conducted. For verification, we have selected two heavy rainfall events reported by IMD. Specifically, we will verify the heavy rainfall events that happened on 21 July 2023 and 26 July 2023, to assess the improvements in bias-corrected rainfall obtained using various QM methods for such events over the Santacruz station. Fig. 9 depicts the rainfall event which occurred on 21 July 2023, across different forecast lead times (highlighted by the black ellipse). On 21 July 2023, the observed rainfall exceeded 100mm/day,

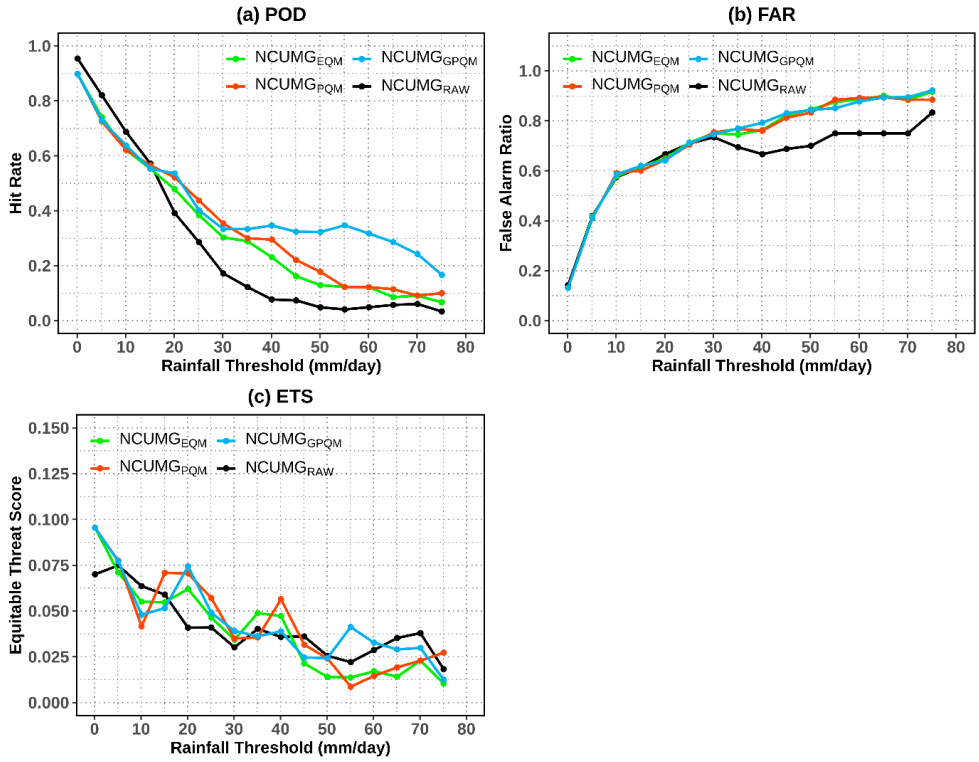


**Figure 8: Categorical rainfall scores (a) POD (top left), (b) FAR (top right), and (c) ETS (bottom left) at different thresholds for raw (NCUMG$_{RAW}$) and calibrated rainfall obtained using different QM approaches at the Santacruz station for Day-1 forecasts.**

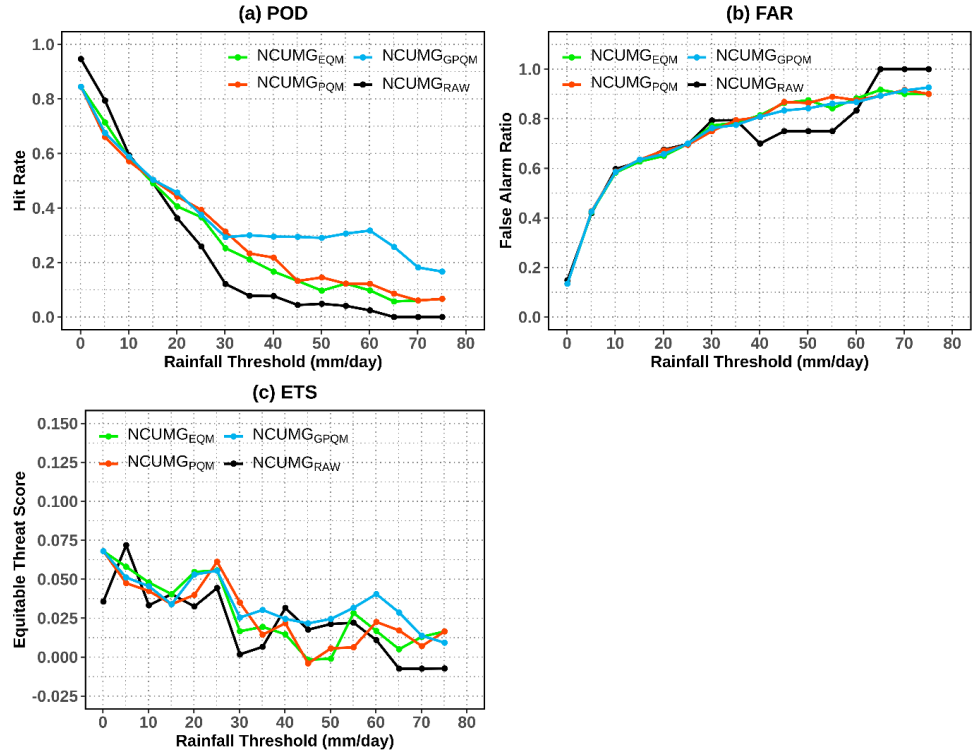


**Figure S2: Categorical rainfall scores (a) POD (top left), (b) FAR (top right), and (c) ETS (bottom left) at different thresholds for raw (NCUMG$_{RAW}$) and calibrated rainfall obtained using different QM approaches at the Santacruz station for Day-03 forecasts.**

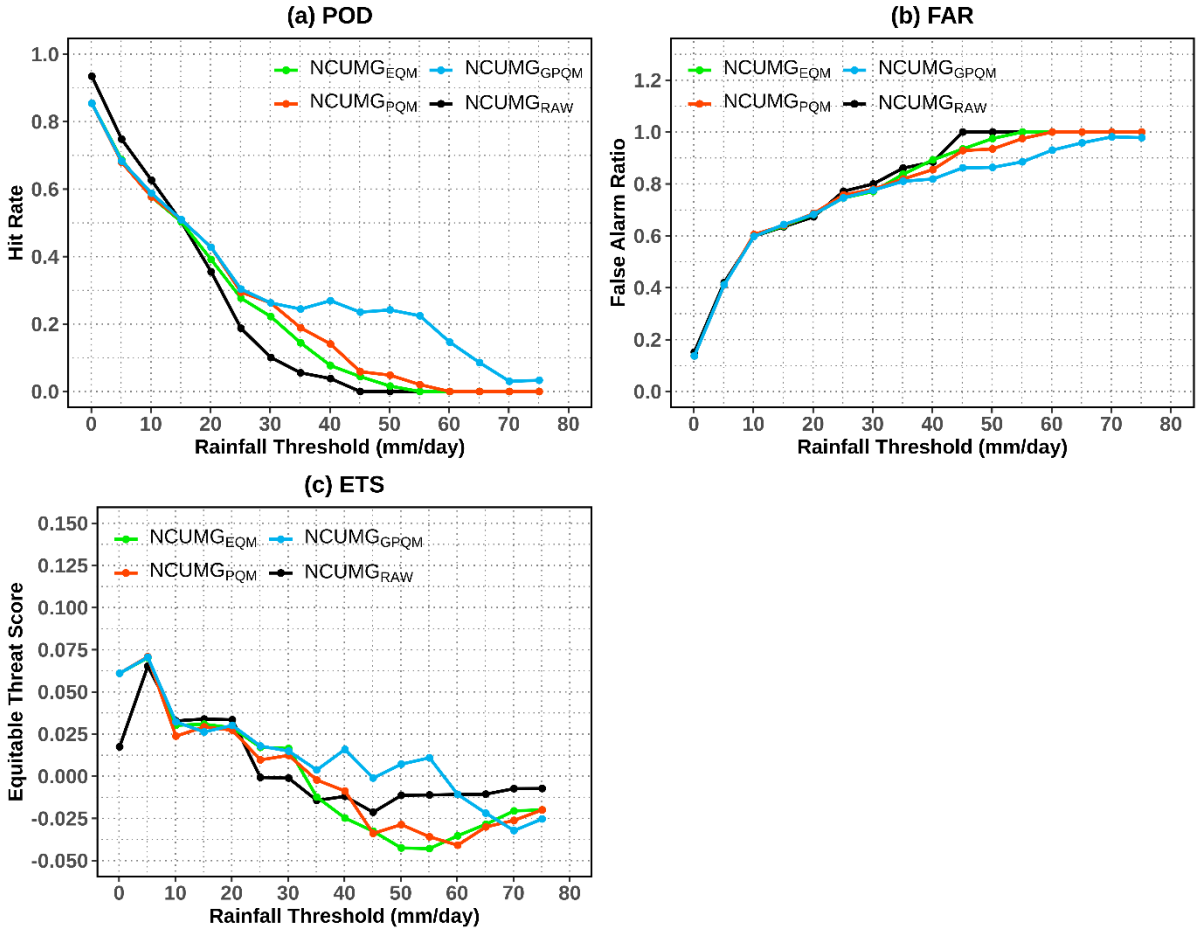


**Figure S3: Categorical rainfall scores (a) POD (top left), (b) FAR (top right), and (c) ETS (bottom left) at different thresholds for raw ($NCUMG_{RAW}$) and calibrated rainfall obtained using different QM approaches at the Santacruz station for Day-05 forecasts.**

whereas NCUMGRAW significantly underestimated this event. The calibrated precipitation obtained from different QM methods, particularly GPQM, shows substantial improvement, ranging from 59 to 96.3 mm/day. The most notable the highest magnitude observed in the Day-03 forecasts, attributed to the model spin-up issue, resulting in better forecasts with increased forecast lead times.

Similarly, in Fig. 10, another heavy rainfall event on 26 July 2023, is depicted over the Santacruz station . Mumbai experienced heavy rainfall exceeding 100mm/day on 26 July 2023, as per IMD. The NCUMGRAW underestimated this event, with magnitudes ranging from 8.5 to 31.5 mm/day in Day-05 to Day-01 forecasts. Whereas, all parametric methods exhibit improved forecasting of this event in Day-01 to Day-03 forecasts, with magnitudes closely aligning with the observed rainfall. Notably, this event is missing in the Day-05 forecast of NCUMGRAW. One potential reason for this absence might be the spatial dislocation of this synoptic event in the model's forecasts. Furthermore, the calibrated rainfall obtained from various QM methods does not anticipate this rainfall at all. It is important to note that QM's statistical adjustments rely on historical quantiles, posing challenges in bias-correcting highly localized mesoscale events with significant rainfall. Nevertheless, it remains imperative to address biases in such specific events, and an alternative approach for accurately identifying such forecasts is essential through the utilization of BC techniques based on synoptic events.

## 5. Summary and Conclusion

This study systematically assesses the efficacy of three QM bias correction techniques, incorporating

both empirical and parametric methods, to improve NCMRWF precipitation forecasts for Santacruz station ( 72.85 0E, 19.117 0N) in Mumbai during the southwest monsoon season of 2020-23. The CDF of calibrated precipitation, derived from different QM approaches, demonstrates a close alignment with the IMD station dataset, particularly for higher thresholds. Furthermore, the temporal evolution of daily rainfall reveals that the GPQM is the best method for predicting most of the events across various forecast lead times, improving from raw forecasts. Additionally, the skill of calibrated rainfall is evaluated for moderate to heavy rainfall occurrences, with all QM approaches showing higher POD and ETS compared to raw forecasts. Notably, in Day-01 to Day-05 forecasts, different QM techniques exhibit a tendency to overestimate false alarms for moderate to heavy rainfall events. However, as the forecast lead time increases, there is a discernible improvement in mitigating the overestimation of calibrated precipitation during such events compared to raw forecasts. Consequently, the findings of this study strongly suggest the advantage of calibrated rainfall using various QM techniques, especially GPQM, over raw forecasts. This calibrated approach proves to be more suitable for hydrological applications like regional-scale flood forecasting.

This study contributes valuable insights into the effectiveness of three QM techniques in enhancing precipitation forecast accuracy, a critical factor in advancing flood prediction. But still, this study has several caveats, including the spatial dislocation of synoptic events in the model's forecasts, higher false alarm rates resulting from smaller sample sizes, etc. Additionally, the fixed climatological distribution used for bias correction may not account for changes in climate, potentially hindering improvements in rainfall prediction. These issues lead to discrepancies between observed and forecasted physics, impacting calibration as the statistical adjustment via quantitative methods relies on historical quantiles. Consequently, further analysis is warranted based on improving calibration techniques to enhance flood prediction capabilities by developing more dynamic approaches that can adapt to spatial and temporal variations in extreme weather events and evolving climatic conditions.

**Acknowledgements**

The IMDAA regional reanalysis dataset is the result of the National Monsoon Mission Programme of the Ministry of Earth Sciences, Government of India; the Met Office, UK; and many individuals who have contributed directly or indirectly to setting up and running this complex reanalysis system. This work is supported by the Ministry of Earth Sciences, Government of India.

**References**

Amengual, A., Homar, V., Romero, R., Alonso, S., & Ramis, C., 2012. A Statistical Adjustment of Regional Climate Model Outputs to Local Scales: Application to Platja de Palma, Spain, Journal of Climate, 25(3), 939-957. https://doi.org/10.1175/JCLI-D-10-05024.1

Raghavendra Ashrit, S. Indira Rani, Sushant Kumar, S. Karunasagar, T. Arulalan, Timmy Francis, Ashish Routray, S. I. Laskar, Sana Mahmood, Peter Jermey, Adam Maycock, Richard Renshaw, John P. George, and E. N. Rajagopal (2020), "IMDAA Regional Reanalysis: Performance Evaluation During Indian Summer Monsoon Season", JGR Atmospheres, Vol.125, Issue2, AGU Publications.doi:10.1029/2019JD030973

Ayantika, D.C., Krishnan, R., Ramarao, M.V.S., Vellore, R., Singh, M. and Mapes, B., 2018. Phenomenological paradigm for midtropospheric cyclogenesis in the Indian summer monsoon. J. Atmos. Sci., 75, 2931–2954, https://doi.org/10.1175/JAS-D-17-0356.1.

Berg, P., Feldmann, H. and Panitz, H.J., 2012. Bias correction of high resolution regional climate model data. Journal of Hydrology, 448, pp.80-92.

Boé, J., Terray, L., Habets, F., & Martin, E., 2007. Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. Int. J. Climatol., 27: 1643-1655. https://doi.org/10.1002/joc.1602

Cannon, A. J., Sobie, S. R., & Murdock, T. Q., 2015. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, Journal of Climate, 28(17), 6938-6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, 208 pp

De´que´, M., 2007. Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. Global Planet. Change, 57, 16–26. https://doi.org/10.1016/j.gloplacha.2006.11.030

Eden, J.M., et al., 2012. Skill, correction, and downscaling of GCM-simulated precipitation. Journal of Climate, 25 (11), 3970–3984. doi:10.1175/JCLI-D-11-00254.1

Fang, G.H., Yang, J., Chen, Y.N. and Zammit, C., 2015. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. Hydrology and Earth System Sciences, 19(6), pp.2547-2559.

Francis, P.A. and Gadgil, S., 2006. Intense rainfall events over the west coast of India. Meteorology and Atmospheric Physics, 94(1-4), pp.27-42.

Gadgil, S., 2003. The Indian monsoon and its variability. Annu. Rev. Earth Planet. Sci., 31, 429-467.

Gadgil, S. and Gadgil, S., 2006. The Indian monsoon, GDP and agriculture. Economic and political weekly, pp.4887-4895.

Goswami, B.N., Venugopal, V., Sengupta, D., Madhusoodanan, M.S. and Xavier, P.K., 2006. Increasing trend of extreme rain events over India in a warming environment. Science, 314(5804), pp.1442-1445.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T., 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012

Gudmundsson, L., Bremnes, J.B., Haugen, J.E. and Engen-Skaugen, T., 2012. Downscaling RCM precipitation to the station scale using statistical transformations–a comparison of methods. Hydrology and Earth System Sciences, 16(9), pp.3383-3390.

Gutjahr, O., & Heinemann, G., 2013. Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM, Theor. Appl. Climatol., 114, 511–529, https://doi.org.10.1007/s00704-013-0834-z

Jenamani, R. K., Bhan, S. C., & Kalsi, S. R., 2006. Observational/forecasting aspects of the meteorological event that caused a record highest rainfall in Mumbai. Current Science, 1344-1362.

Kim, D.I., Kwon, H.H., & Han, D., 2019. Bias correction of daily precipitation over South Korea from the long-term reanalysis using a composite Gamma-Pareto distribution approach. Hydrol. Res., 50, 1138–1161. https://doi.org/10.2166/nh.2019.127

Krishnamurti, T. N., & Hawkins, R. S., 1970. Mid-Tropospheric Cyclones of the Southwest Monsoon, Journal of Applied Meteorology and Climatology, 9(3), 442-458. https://doi.org/10.1175/1520-0450(1970)009<0442:MTCOTS>2.0.CO;2

Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A. & Chakraborty, S., 2020. Assessment of climate change over the Indian region: a report of the Ministry of Earth Sciences (MOES), government of India (p. 226). Springer Nature.

Kulkarni, A., Sabin, T.P., Chowdary, J.S., Rao, K.K., Priya, P., Gandhi, N., Bhaskar, P., Buri, V.K., Sabade, S.S., Pai, D.S. & Ashok, K., 2020. Precipitation changes in India. Assessment of climate change over the Indian region: a report of the Ministry of Earth Sciences (MoES), Government of India, pp.47-72.

Kumar, A., Dudhia, J., Rotunno, R., Niyogi, D., & Mohanty, U.C., 2008. Analysis of the 26 July 2005 heavy rain event over Mumbai, India using the

Weather Research and Forecasting (WRF) model. Q.J.R. Meteorol. Soc., 134: 1897-1910. https://doi.org/10.1002/qj.325

Kumar, V., Jain, S.K. and Singh, Y., 2010. Analysis of long-term rainfall trends in India. Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55(4), pp.484-496.

Macias, D., Garcia-Gorriz, E., Dosio, A., Stips, A. and Keuler, K., 2018. Obtaining the correct sea surface temperature: bias correction of regional climate model data for the Mediterranean Sea. Climate dynamics, 51, pp.1095-1117.

Maraun, D., & Widmann, M., 2015. The representation of location by regional climate models in complex terrain. Hydrol. Earth Syst. Sci. Discuss., 12, 3011–3028, https://doi.org.10.5194/hessd-12-3011-2015

Miller, F. R., & Keshvamurthy, R. N., 1968. Structure of Arabian Sea monsoon system. I. I. O. E Meteorol Monograph, I, 94

Munday, C., & Washington, R., 2018. Systematic Climate Model Rainfall Biases over Southern Africa: Links to Moisture Circulation and Topography. Journal of Climate. 31(18), 7533-7548. https://doi.org/10.1175/JCLI-D-18-0008.1

Niranjan Kumar, et al., 2022. Quantile mapping bias correction methods to IMDAA reanalysis for calibrating NCMRWF unified model operational forecasts, Hydrological Sciences Journal, 67:6, 870-885.

https://doi.org/10.1080/02626667.2022.2049272

Phatak, J. & Hukku, A., 2015. Case of Greater Mumbai, In: Monographs of the International Conference on Water Megacities and Global Change; 1–4 December 2015 at UNESCO HQ, Paris, France, UNESCO, Paris, pp. 1–32.

Piani, C., Haerter, J.O., & Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theor. Appl. Climatol., 99, 187–192. https://doi.org/10.1007/s00704-009-0134-9

Rajeevan, M., Bhate, J., & Jaswal, A. K., 2008. Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. Geophys. Res. Lett., 35, L18707, https://doi.org/10.1029/2008GL035143

Rani, S. I, George, J. P., Arulalan, T., Sumit Kumar, Das Gupta, M., Rajagopal, E. N., & Renshaw, R., 2020. "Evaluation of High Resolution IMDAA Regional Reanalysis Precipitation over India during Summer Monsoon Season", CLIVAR Exchanges - Special Issue: India's Monsoon Mission, No.79, Pages 31-33, November 2020 DOI: https://doi.org/10.36071/clivar.79.2020

Rani, S. I., T, A., George, J. P., Rajagopal, E. N., Renshaw, R., Maycock, A., Barker, D. M., & Rajeevan, M., 2021. IMDAA: High Resolution Satellite-era Reanalysis for the Indian Monsoon Region, Journal of Climate, 1-78. Doi: https://doi.org/10.1175/JCLI-D-20-0412.1

Rao, Y. P., 1976. Southwest monsoon. Meteorological Monograph No 1, India Meteorological Department, New Delhi.

Roxy, M.K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P. and Rajeevan, M., 2017. A threefold rise in widespread extreme rain events over central India. Nature communications, 8(1), p.708.

Sahai, A.K., Grimm, A.M., Satyan, V. and Pant, G.B., 2003. Long-lead prediction of Indian summer monsoon rainfall from global SST evolution. Climate Dynamics, 20, pp.855-863.

Singh, J., Sekharan, S., Karmakar, S., Ghosh, S., Zope, P.E. and Eldho, T.I., 2017. Spatio-temporal analysis of sub-hourly rainfall over Mumbai, India: Is statistical forecasting futile?. Journal of Earth System Science, 126, pp.1-15.

Sumit Kumar, Bushair, M. T., Buddhi Prakash J., Lodh, A., Priti Sharma, Gibies George, Indira Rani, S., George, J. P., Jayakumar, A., Saji Mohandas, Sushant Kumar, Kuldeep Sharma, Karunasagar, S., & Rajagopal, E. N., 2020. NCUM Global NWP System: Version 6 (NCUM-G:V6), NCMRWF Technical Report, NMRF/TR/06/2020.

Teutschbein, C., & Seibert, J., 2012. Bias correction of regional climate model simulations for

hydrological climate-change impact studies: Review and evaluation of different methods. J. Hydrol., 456, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Themeßl, M.J., Gobiet, A., and Leuprecht, A., 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. International Journal of Climatology, 31, 1530–1544. doi:10.1002/joc.2168

Themeßl, M.J., Gobiet, A. and Heinrich, G., 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Climatic Change, 112, pp.449-468.

Wetterhall, F., Pappenberger, F., He, Y., Freer, J. and Cloke, H.L., 2012. Conditioning model output statistics of regional climate model precipitation on circulation patterns. Nonlinear Processes in Geophysics, 19(6), pp.623-633.

Widmann, M. and Bretherton, C.S., 2000. Validation of mesoscale precipitation in the NCEP reanalysis using a new gridcell dataset for the northwestern United States. Journal of Climate, 13(11), pp.1936-1950.

Yoshikane, T. and Yoshimura, K., 2022. A bias correction method for precipitation through recognizing mesoscale precipitation systems corresponding to weather conditions. PLoS Water, 1(5), p.e0000016.