NMRF/RR/04/2023

RESEARCH REPORT

सत्यमेव जयते

Implementation of Statistical Bias Correction Methods for NCUM-Global Operational Forecasts in Support of iFLOWS-Mumbai

Sukhwinder Kaur, Kondapalli Niranjan Kumar, Sumit Kumar, Harvir Singh, Mohana S. Thota, Raghavendra Ashrit, & V. S. Prasad

Nov 2023

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

Implementation of Statistical Bias Correction Methods for NCUM-Global Operational Forecasts in Support of iFLOWS-Mumbai

Sukhwinder Kaur, Kondapalli Niranjan Kumar, Sumit Kumar, Harvir Singh, Mohana S. Thota, Raghavendra Ashrit, & V. S. Prasad

NMRF/RR/04/2023

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector 62, NOIDA-201309, INDIA
www.ncmrwf.gov.in

Document Control Data Sheet

1	Name of the Institute	National Center for Medium range weather Forecasting
2	Document Number	NMRF/RR/04/2023
3	Date of Publication	November 2023
4	Title of the document	Implementation of Statistical Bias Correction Methods for NCUM-Global Operational Forecasts in Support of iFLOWS-Mumbai
5	Type of the document	Research Report
6	Number of pages, figures and Tables	25 pages and 11 figures
7	Authors	Sukhwinder Kaur, Kondapalli Niranjan Kumar, Sumit Kumar, Harvir Singh, Mohana S. Thota, Raghavendra Ashrit, & V. S. Prasad
8	Originating Unit	National Centre for Medium Range Weather Forecasting (NCMRWF), A-50, Sector-62, NOIDA 201 309, India
9	Abstract	Systematic biases in model forecasts, especially in crucial variables like precipitation, hamper their direct use in impact studies, consequently, several statistical bias-correction approaches have been developed to calibrate model outputs relative to the observations. In this context, this study endeavors to determine the efficiency of various Quantile Methods (QM) bias-correction techniques based on both empirical and parametric methods in improving the Global National Centre for Medium-Range Weather Forecasting (NCMRWF) Unified Model (NCUM-G) precipitation forecasts over Mumbai (BOM) region during monsoon season (June to September, JJAS) 2022. This location is noteworthy for supporting the Integrated Flood Warning System (iFLOWS), a key instigated by the Ministry of Earth Sciences, Government of India, which is an essential source of prompt notifications and guidance during catastrophes like floods. The empirical QM approach outperforms in adjusting quantiles, aligning calibrated precipitation closely with the cumulative distribution observed at the BOM location. However, parametric QM methods demonstrate potential in effectively calibrating precipitation during extreme rainfall events. Thus, this work reveals the effectiveness of QM approaches in enhancing the accuracy of precipitation forecasts, which is crucial for the advancement of urban flood models.
10	References	32
11	Security classification	Unrestricted
12	Distribution	General

Table of Contents

NMRF/RR/04/2023

Implementation of Statistical Bias Correction Methods for NCUM-Global Operational Forecasts in Support of iFLOWS-Mumbai

Sukhwinder Kaur, Kondapalli Niranjan Kumar, Sumit Kumar, Harvir Singh, Mohana S. Thota, Raghavendra Ashrit, & V. S. Prasad

सारांश

मॉडल पूर्वानुमानों में व्यवस्थित पूर्वाग्रह, विशेष रूप से वर्षा जैसे महत्वपूर्ण चर में, प्रभाव अध्ययन में उनके प्रत्यक्ष उपयोग में बाधा डालते हैं, परिणामस्वरूप, अवलोकनों के सापेक्ष मॉडल आउटपुट को कैलिब्रेट करने के लिए कई सांख्यिकीय पूर्वाग्रह-सुधार दृष्टिकोण विकसित किए गए हैं। इस संदर्भ में, यह अध्ययन वैश्विक नेशनल सेंटर फॉर मीडियम-रेंज वेदर फोरकास्टिंग (रा.म.अ.मौ.पू.कें.) यूनिफाइड मॉडल (एन.सी.यू.एम.-जी) वर्षा पूर्वानुमानों को बेहतर बनाने में अनुभवजन्य और पैरामीट्रिक दोनों तरीकों के आधार पर विभिन्न क्वांटाइल मेथड्स (क्यू.एम.) पूर्वाग्रह-सुधार तकनीकों की दक्षता निर्धारित करने का प्रयास करता है, मानसून सीजन (जून से सितंबर, जे.जे.ए.एस.) 2022 के दौरान मुंबई (बी.ओ.एम.) क्षेत्र में। यह स्थान एकीकृत बाढ़ चेतावनी प्रणाली (आई.एफ.एल.ओ.डब्ल्यू.एस.) का समर्थन करने के लिए उल्लेखनीय है, जो भारत सरकार के पृथ्वी विज्ञान मंत्रालय द्वारा शुरू की गई एक महत्वपूर्ण पहल है, जो बाढ़ जैसी आपदाओं के दौरान त्वरित सूचनाएं और मार्गदर्शन का एक आवश्यक स्रोत है। अनुभवजन्य क्यू.एम. दृष्टिकोण मात्राओं को समायोजित करने, बीओएम स्थान पर देखे गए संचयी वितरण के साथ कैलिब्रेटेड वर्षा को बारीकी से संरेखित करने में बेहतर प्रदर्शन करता है। हालाँकि, पैरामीट्रिक क्यू.एम. विधियाँ अत्यधिक वर्षा की घटनाओं के दौरान वर्षा को प्रभावी ढंग से अंशांकित करने की क्षमता प्रदर्शित करती हैं। इस प्रकार, यह कार्य वर्षा पूर्वानुमानों की सटीकता को बढ़ाने में क्यू.एम. दृष्टिकोण की प्रभावशीलता को प्रकट करता है, जो शहरी बाढ़ मॉडल की प्रगति के लिए महत्वपूर्ण है।

Abstract

Systematic biases in model forecasts, especially in crucial variables like precipitation, hamper their direct use in impact studies, consequently, several statistical bias-correction approaches have been developed to calibrate model outputs relative to the observations. In this context, this study endeavors to determine the efficiency of various Quantile Methods (QM) bias-correction techniques based on both empirical and parametric methods in improving the Global National Centre for Medium-Range Weather Forecasting (NCMRWF) Unified Model (NCUM-G) precipitation forecasts over Mumbai (BOM) region during monsoon season (June to September, JJAS) 2022. This location is noteworthy for supporting the Integrated Flood Warning System (iFLOWS), a vital initiative launched by the Ministry of Earth Sciences, Government of India, which is an essential source of prompt notifications and guidance during catastrophes like floods. The empirical QM approach outperforms in adjusting quantiles, aligning calibrated precipitation closely with the cumulative distribution observed at the BOM location. However, parametric QM methods demonstrate potential in effectively calibrating precipitation during extreme rainfall events. Thus, this work reveals the effectiveness of QM approaches in enhancing the accuracy of precipitation forecasts, which is crucial for the advancement of urban flood models.

1. Introduction

India has experienced a significant rise/increase in the frequency, variability, and intensity of extreme rainfall events in recent decades, which have been accompanied by widespread flood-like conditions and catastrophic losses of life (Rajeevan et al., 2008, Kulkarni et al. 2020, Krishnan et al. 2020). For instance, the extreme rainfall events that occurred in December 2015 over Chennai (Boyaj et al., 2018) and in July 2005 over Mumbai (Jenamani, et al., 2006; Kumar et al., 2008) were disastrous and unprecedented. In addition, Kerala state experienced flood in August 2018 that was unprecedented in the record of the past 66 years (Mishra and Shah, 2018; Mishra et al., 2018; Lyngwa and Nayak, 2021). Indian Summer Monsoon Rainfall (ISMR) during June–September (JJAS) accounts for 70% of the total rainfall, and the large variability in ISMR at multiple time-scales significantly affects the lives of numerous people and the agriculture sector over the subcontinent (Webster 1998; Gadgil 2003). Among the rainfall coherent regions, India's western coast experiences heavy rainfall. Mumbai, which is located on the west coast, has a history of flooding during extreme rainfall events. The city experienced massive floods in July 2005, 2017, and most recently in 2022. Thus, the prior information on urban floods caused by extreme rainfall is essential to prevent associated social and economic risks in this climate sensitive location.

To that purpose, the Ministry of Earth Sciences (MoES), the Government of India (GoI), in collaboration with the Municipal Corporation of Greater Mumbai (MCGM), has introduced an integrated flood warning system, iFLOWS. Mumbai is the second Indian city to have such a system after Chennai. The iFLOWS system consists of seven modules, including a decision support system, vulnerability, risk, flood, inundation, dissemination, and data assimilation. The Geographic Information System (GIS) is used as the flood warning system. This system comprises numerical weather prediction (NWP) models from the National Centre for Medium

Range Weather Forecasting (NCMRWF), the India Meteorological Department (IMD), and observations from the rain gauge network stations.

With the advancement of technology, NWP models have undergone significant upgrades during the last few decades, enabling high-resolution weather forecasts However, the model's forecasts often reveal systematic errors in comparison to observations, particularly for essential variables such as precipitation, which have a significant impact on society (Cannon et al., 2015; Munday and Washington, 2018). Thus, correcting the model errors is crucial, in order to utilize the NWP forecasts effectively for decision-making applications pertaining to flood risk management.

The aim of this report is to assess various quantile mapping bias correction (QMBC) methods that are widely acknowledged for outperforming other approaches (Teutschbein and Seibert, 2012; Maraun and Widmann, 2018; Kim et al., 2019), in improving rainfall forecasts of the Global National Centre for Medium Range Weather Forecasts (NCMRWF) Unified Model (NCUM-G) operational model over the Mumbai region supporting iFLOWS program of MoES.

2. Study area and Data

2.1. *Study area*

In this study, we considered the urban city -- Mumbai (BOM), which is the country's most populous coastal metropolitan city located on India's western coast (see Fig.1). The motivation behind choosing Mumbai city is that this region witnessed several extreme rainfall events in recent decades (see Fig.2). In general, BOM has a humid tropical monsoon climate with a significant impact from southwest monsoon. During the southwest monsoon, Mumbai, which is on the windward side of the Western Ghats of India, experiences heavy rainfall due to the orographic effect. Additionally, several studies suggested that offshore vortices, depressions in the Arabian Sea, and mid-tropospheric cyclones (MTC) also play a vital role in the production/generation of heavy rainfall over Mumbai region (Rao 1976, Miller and Keshvamurthy 1968, Krishnamurti and Hawkins 1970, Ayantika et al. 2018). Thus, accurate precipitation predictions over the BOM are critical to supporting decision-making applications for flood risk management.

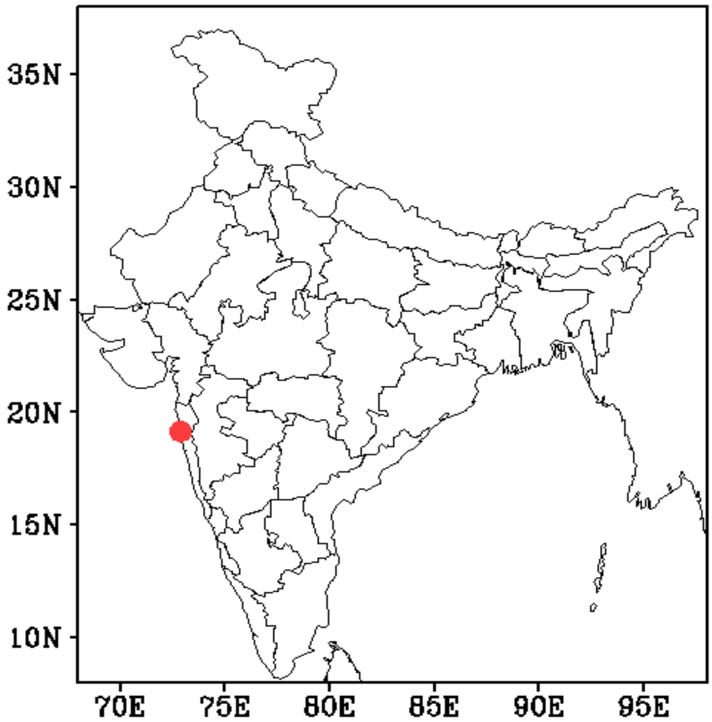

Fig.1. Map showing the location examined in the study.

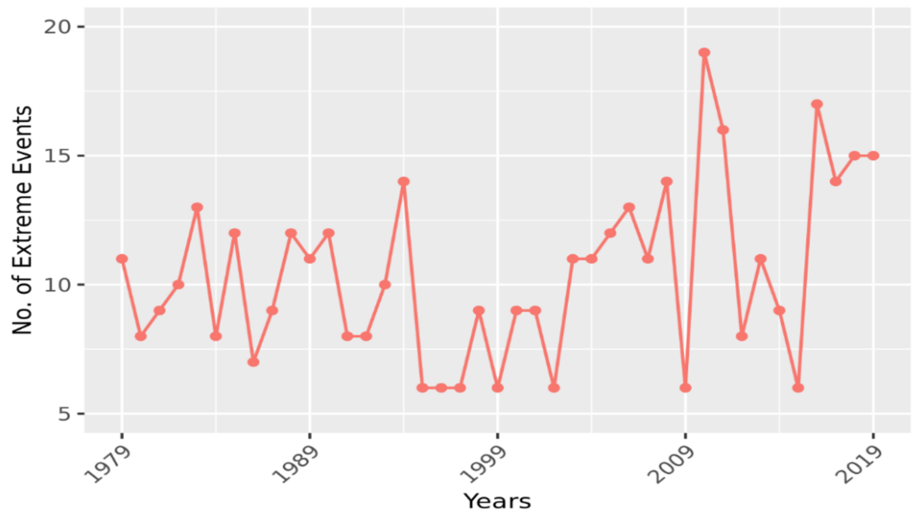

Fig. 2. Frequency of the extreme annual rainfall (>64.5mm/day) over the Mumbai, Santacruz location based on the IMD station data during 1979-2019.

2.2. *Data*

In this study, a high-resolution Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis for the period 1979-2019 (Rani et al., 2020, 2021), IMD observed station-based dataset from 1979 -2019 (Jenamani, et al., 2006), and the NCUM-G operational forecast dataset for 2022 (Sumit Kumar et al., 2020) are used. These datasets were utilized to create the BC model using various statistical methods, as well as for the model's training and testing phase (for details see Niranjan Kumar et al., 2022). The NWP dynamical core parameterization techniques used to produce the IMDAA reanalysis dataset and the NCUM-G are relatively similar. As they use the same model physics, IMDAA can be used to efficiently correct NCUM-G real-time forecasts. It is also worth noting that the calibration procedures correct the model's systematic biases, which are equivalent to several years of improvement to the basic model. Further, to evaluate the real-time application of the calibration methods, the BC model is applied to the NCUM-G operational forecast for improving/enhancing the rainfall forecast at local scale, particularly, at Mumbai for JJAS season.

3. Methodology

Statistical BC approaches establish a functional relationship between simulated and observed variables over some historical period, which is subsequently used for the model projections (Niranjan Kumar et al., 2022). A brief description of the various quantile approaches used in this study is given below.

3.1. *Empirical Quantile Method (EQM)*

The EQM is a non-parametric BC method that corrects the mean, standard deviation (variability), and shape errors by mapping the quantiles of the simulated cumulative distribution function (CDF) to those of the observations through an empirical transfer function (Boè et al. 2007; Déqué, 2007; Amengual et al., 2012, Gudmundsson et al. 2012; Niranjan Kumar et al., 2022).

Let cal represents the calibrated precipitation obtained after the bias correction and obs be the observed precipitation. Then the calibrated precipitation for the CDF's i^{th} ranked value is expressed as

$$cal_i = obs_i + \bar{\delta} + {\delta_i}' \qquad (1)$$

where δ_i is defined as 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 expressed as $\bar{\delta} = \frac{\sum_{i=1}^{N} \delta_i}{N}$ and the corresponding deviation from this shift i.e., ${\delta_i}' = \delta_i - \bar{\delta}$.

This method, apart from being simple and non-parametric, has also proved to be effective in comparison to other distribution and scaling BC approaches (Boé et al. 2007, Gudmundsson et al. 2012). However, instabilities arise at higher quantiles (Gudmundsson et al., 2012; Gutjahr and Heinemann, 2013). In addition, this method also relies on many degrees of freedom, which results in non-stationary for future time periods (Gutjahr and Heinemann, 2013). Thus, we are also considering parametric BC methods as discussed below, which rely on a lesser degree of freedom.

3.2. *Parametric Quantile Method (PQM)*

The PQM is a parametric BC method that uses theoretical distribution instead of empirical distribution. This method adjusts the theoretical CDF of the simulated output onto the corresponding observed distribution through a parametric transfer function (i.e., two-parameter gamma distributions) (Niranjan Kumar et al., 2022; Piani et al., 2010). The probability density function (y) of the gamma distribution is:

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

Where $\Gamma(.)$ denotes the gamma function. ξ, σ are the shape and scale parameters of the gamma distribution, respectively. x is the normalized daily precipitation.

The limitation of PQM method is that there is no restriction on the upper limit resulting in possible false alarms for the extreme rainfall events (Gutjahr and Heinemann, 2013). Therefore, in this study, we used a new method based on the gamma distribution combined with Generalized Pareto Distribution (GPD), which is further discussed below.

3.3. *PQM based on gamma and Generalised Pareto Distribution (GPQM)*

The GPQM is also a parametric BC method based on the combination of gamma and Generalized Pareto Distribution (GPD) (Niranjan Kumar et al., 2022). The probability density function (pdf) of GPD is expressed as (Coles 2001)

$$y = f(x|\xi, \sigma, \theta) = \left(\frac{1}{\sigma}\right)\left(1 + \frac{\xi(x-\theta)}{\sigma}\right)^{-1-\frac{1}{\xi}} \qquad (3)$$

$$for\ \theta < x, when\ \xi > 0, or\ for\ \theta < x < \theta - \frac{\sigma}{\xi}\ when\ \xi < 0$$

where, $\xi \neq 0, \sigma, and\ \theta$ represents the shape, scale, and threshold parameter of GPD, respectively. Here, the values lesser than 95th percentile are assumed to follow gamma distribution, while GPD for values above the 95th percentile. Therefore, the GPQM method can be constructed as

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

The methodology adopted here comprises calibrating and testing various quantile methods described in the flow diagram (see Fig. 3). The observed station data from IMD and IMDAA reanalysis data of the closest grid to BOM location has been considered for the assessment. Overall, we have 41-years (i.e., 1979-2019) of daily precipitation data in common from observations and reanalysis. Hence, we have used the whole 41-years of the data for training and tested the model for the year 2022. A more detailed discussion can also be found from Niranjan Kumar et al., 2022. Furthermore, the calibration methods are utilized to bias-correct the real-time operational NCUM-G forecasts.

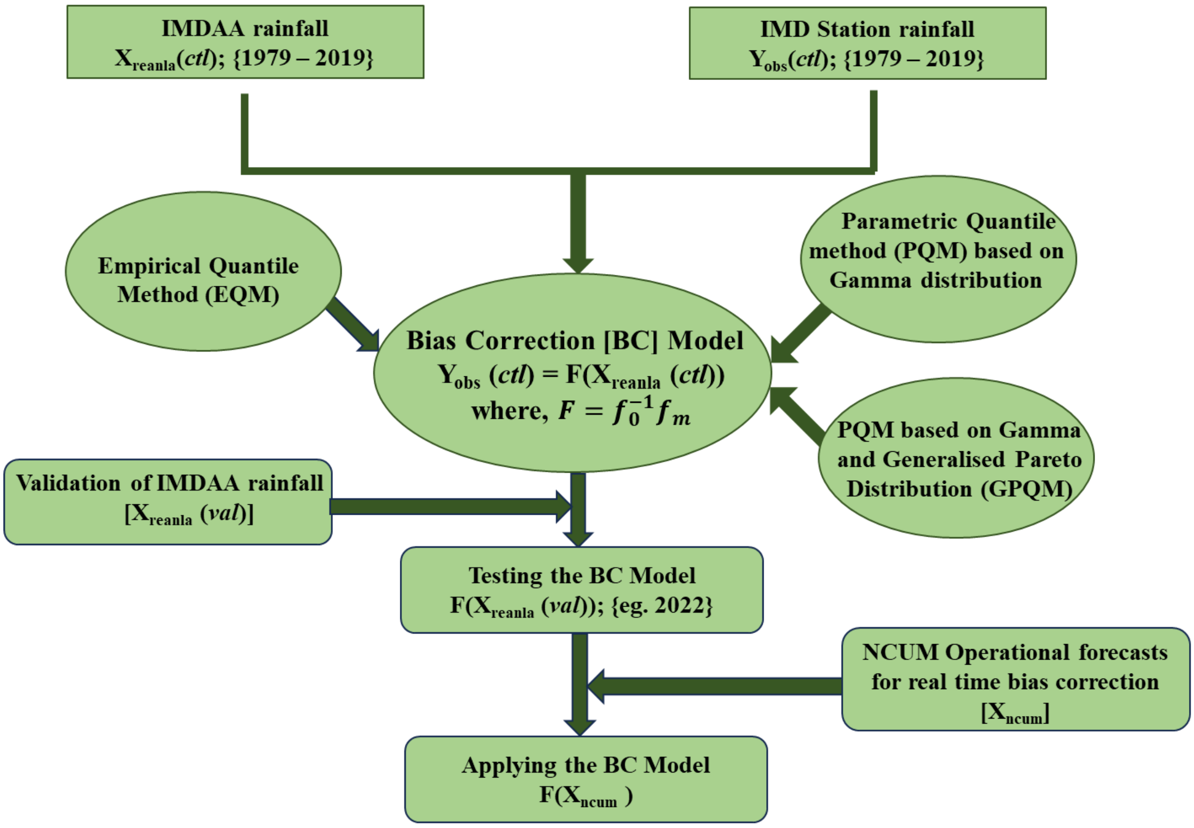

Fig.3. Flow diagram of the methodology adopted in this study.

The skill of different quantile BC methods is assessed by categorical verification scores such as the probability of detection or hit rate (POD), false alarm ratio (FAR) and Equitable Threat Score (ETS). The POD quantifies the fraction of actual occurrences of an event that were successfully forecasted with a score of one indicating a perfect hit rate. While the FAR quantifies the portion of occurrence of an event in forecasts that were indeed wrong. The ETS is particularly valuable to evaluate deterministic forecasts and often used to verify the rainfall in numerical weather prediction models, since it heavily penalises the constant and purely random forecasts (Gandin and Murphy, 1992). ETS measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance.

4. Results and Discussions

4.1. *Assessment of QM approaches for NCUM-G forecasts*

In this section, we first discussed the empirical cumulative distribution function (*cdf*) pertaining to the BOM location. Fig.4 illustrates the JJAS seasonal precipitation *cdf* from the IMD station data (OBS), NCUM-G Day-01 raw forecast ($NCUM_{raw}$), and bias-corrected Day-1 forecast precipitation obtained using EQM ($NCUM_{eqm}$), PQM ($NCUM_{pqm}$), and GPQM ($NCUM_{gpqm}$) for the year 2022. The precipitation amount above 10 mm/day is higher in $NCUM_{raw}$ dataset. Despite having a high spatial resolution, the significant biases in extreme rainfall intensities of $NCUM_{raw}$ compared to the IMD station dataset are evident, which hamper the practical applicability of NCUM-G in local and regional-scale hydrology, and flood-related applications. Therefore, correcting/rectifying these biases is essential prior to utilizing them for urban flood forecasting applications. For this, we assess the cdf using various bias correction methods applied to the NCUM-G precipitation data following the procedures discussed in Section 3. The calibrated precipitation from different QM approaches matches well with the OBS for the higher thresholds (Fig. 4).

To get further insight into the effectiveness of different QM methods in bias-correcting the NCUM-G precipitation forecasts, we validate the calibrated precipitation obtained using different QM approaches with the IMD station dataset. For this, we firstly took the NCUM-G Day-01, Day-03, and Day-05 precipitation forecast for the year 2022 during the southwest monsoon season. Further, to calibrate the NCUM-G precipitation forecasts for the year 2022, we used the IMDAA and station based IMD reanalysis datasets from 1979 to 2019 during southwest monsoon seasons (June-September for BOM region) based on monthly quantile approach as discussed earlier (see Methodology). Fig. 5 shows the validation of Day-01 daily precipitation from NCUM-G raw forecasts ($NCUM_{raw}$) with IMD station-based dataset (OBS) over BOM location for the summer monsoon season. In addition, the calibrated precipitation from different QM approaches based on empirical methods is also presented in Fig. 5a ($NCUM_{eqm}$), and parametric methods in Fig. 5b, c ($NCUM_{pqm}$, $NCUM_{gpqm}$). The time series of Day-01 daily precipitation at the BOM location reveals that the $NCUM_{raw}$ is missing numerous heavy rainfall events. For example, if we assume 50mm/day and above events that come under heavy to extreme rains, there are nearly 14 events observed over BOM location during the southwest monsoon season. The $NCUM_{raw}$ correctly predicted 4 events while the empirical method is able to correct 5 events that were missed in raw forecasts (Fig. 5a). However, the parametric methods, specifically, the GPQM is able to predict all the ten events that were missed in raw forecasts (Fig. 5c). Similarly, with forecast lead times, the daily rainfall time series shows that the raw forecasts underestimate the actual observed rainfall, while the calibrated methods rectify those underestimations relative to the $NCUM_{raw}$ during heavy rainfall events (see Fig. 6 and 7).

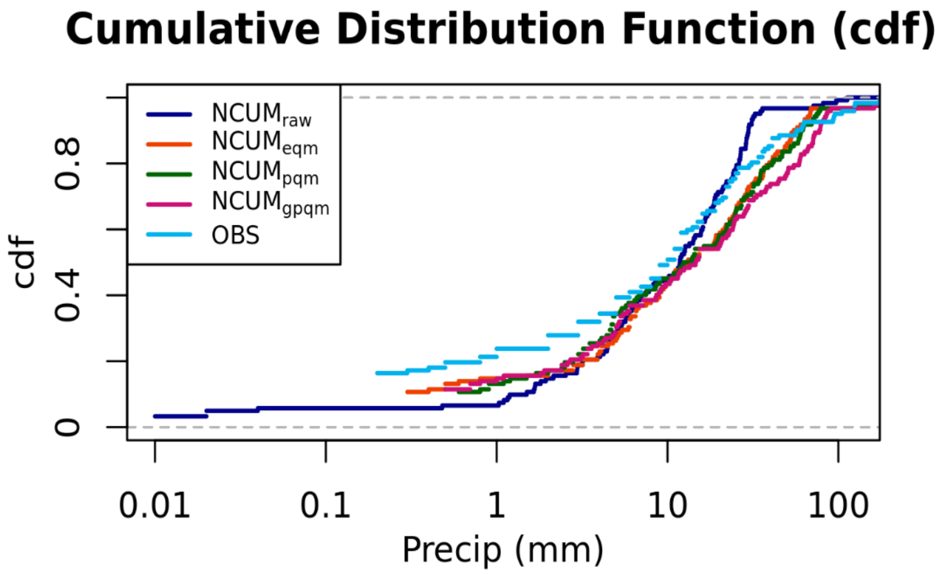

Fig.4. The cdf of daily precipitation from the IMD station dataset (OBS), raw NCUM-G Day-01 forecast precipitation ($NCUM_{raw}$), bias-corrected Day-01 forecast precipitation obtained using EQM ($NCUM_{eqm}$), PQM ($NCUM_{pqm}$), and GPQM ($NCUM_{gpqm}$) during the summer monsoon (June–September, 2022) over BOM.

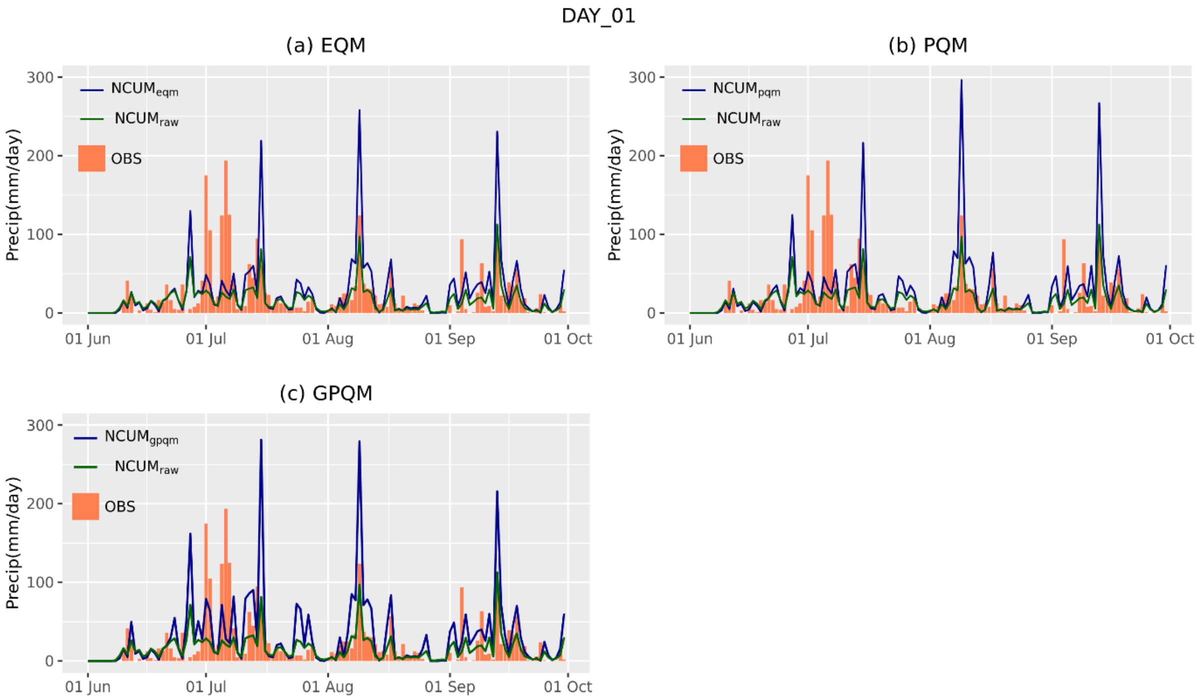

Fig. 5. Time series of daily precipitation from IMD station dataset (OBS), raw NCUM-G Day-01 forecast ($NCUM_{raw}$), and bias-corrected Day-01 forecast obtained using (a) empirical quantile methods (EQM, [$NCUM_{eqm}$]), (b) parametric quantile method based on gamma distribution (PQM, [$NCUM_{pqm}$]), and (c) PQM based on gamma and GPD (GPQM, [$NCUM_{gpqm}$]) during the summer monsoon (June–September, 2022) over BOM.

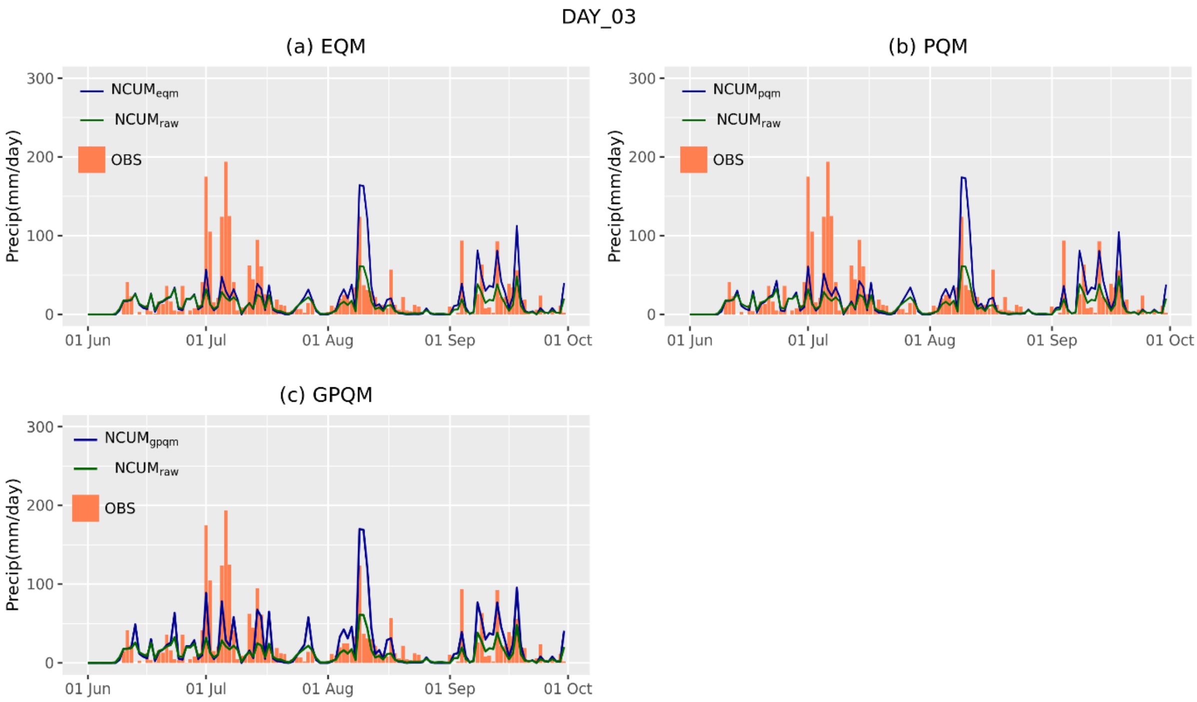

Fig. 6. Same as Figure 5, but for Day-03 forecast.

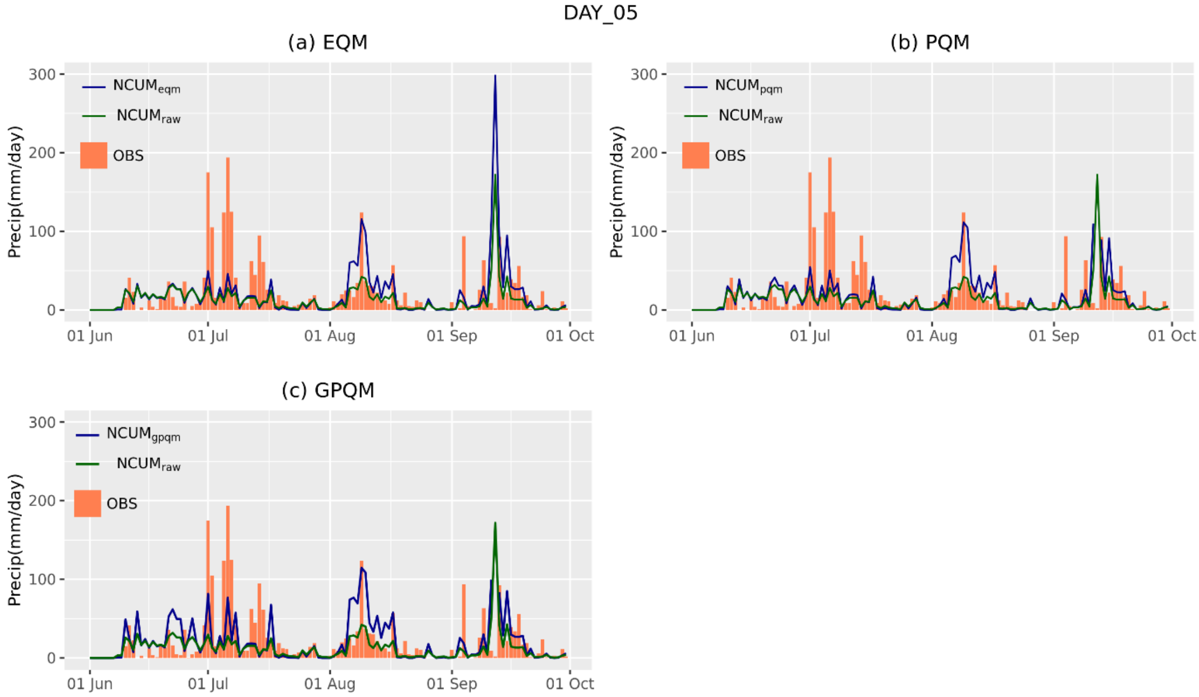

Fig. 7. Same as Fig. 5, but for Day-05 forecast.

In order to assess how well the QM approaches bias-correct the raw forecasts in different lead days, the skill of different quantile BC methods is assessed by categorical verification scores (i.e., POD, FAR, and ETS) and are displayed in Figures 8 and 9. The POD, FAR, and ETS scores are computed for moderate (i.e., at 15.6mm/day threshold) and heavy (i.e., at 64.5mm/day threshold) rainfall events at different lead times over BOM. The POD and ETS from calibrated precipitation clearly show an improvement relative to the $NCUM_{raw}$. For instance, the magnitude of POD and ETS scores of bias-corrected forecast is higher for moderate and heavy rainfall events (i.e., >15.4 mm/days and >64.5mm/day) in all forecast lead times over the BOM region (Fig. 8 and 9). Furthermore, as the lead time increases the performance of the calibrated methods is relatively better than the raw forecasts (Fig. 8 and 9). However, one of the limitations of the POD scores is that it is susceptible to the number of hits ignoring the false alarms. Thus, we have also shown the FAR score in Figures 8(b) and 9(b). The FAR is defined as the occurrence fraction of events for which the event did not occur, and hence a perfect FAR score is zero. In Day-01 to Day-04 forecasts, there is an overestimation of false alarms by different QM approaches in case of moderate to heavy rainfall events. However, the overestimation of moderate to heavy rainfall events in raw NCUM-G precipitation ($NCUM_{raw}$) is slightly improved with forecast lead time, as we can see there is a reduction in FAR (Figs. 8(b) and 9(b). Hence, the verification analysis suggests that the quantile methods have significantly improved the detection of the local extreme rainfall events.

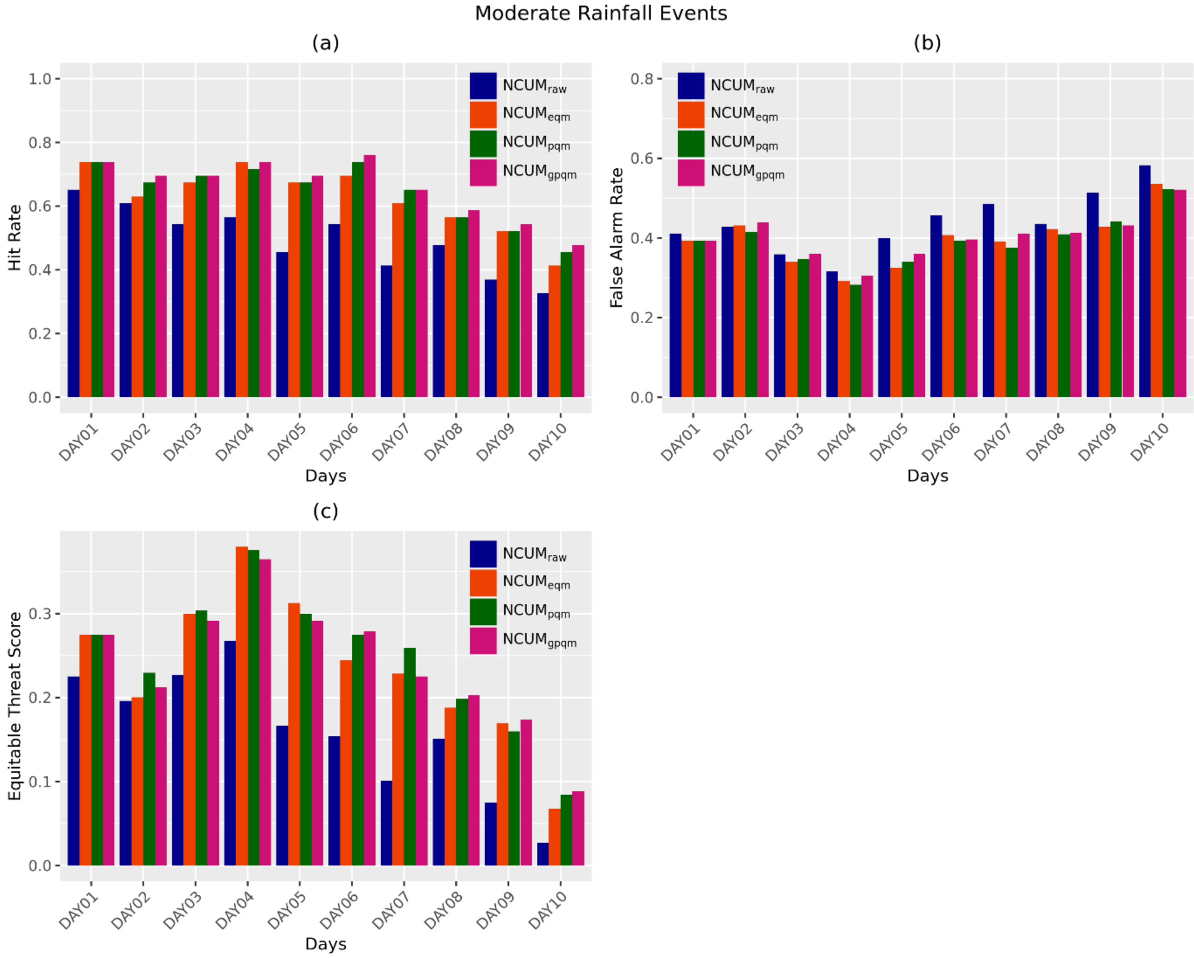

Fig.8. The categorical verification scores (a) the probability of detection (POD), (b) false alarm ratio (FAR), and (c) Equitable Threat Score (ETS) from the raw ($NCUM_{raw}$) and calibrated rainfall using different QM approaches at the BOM location for moderate rainfall events.

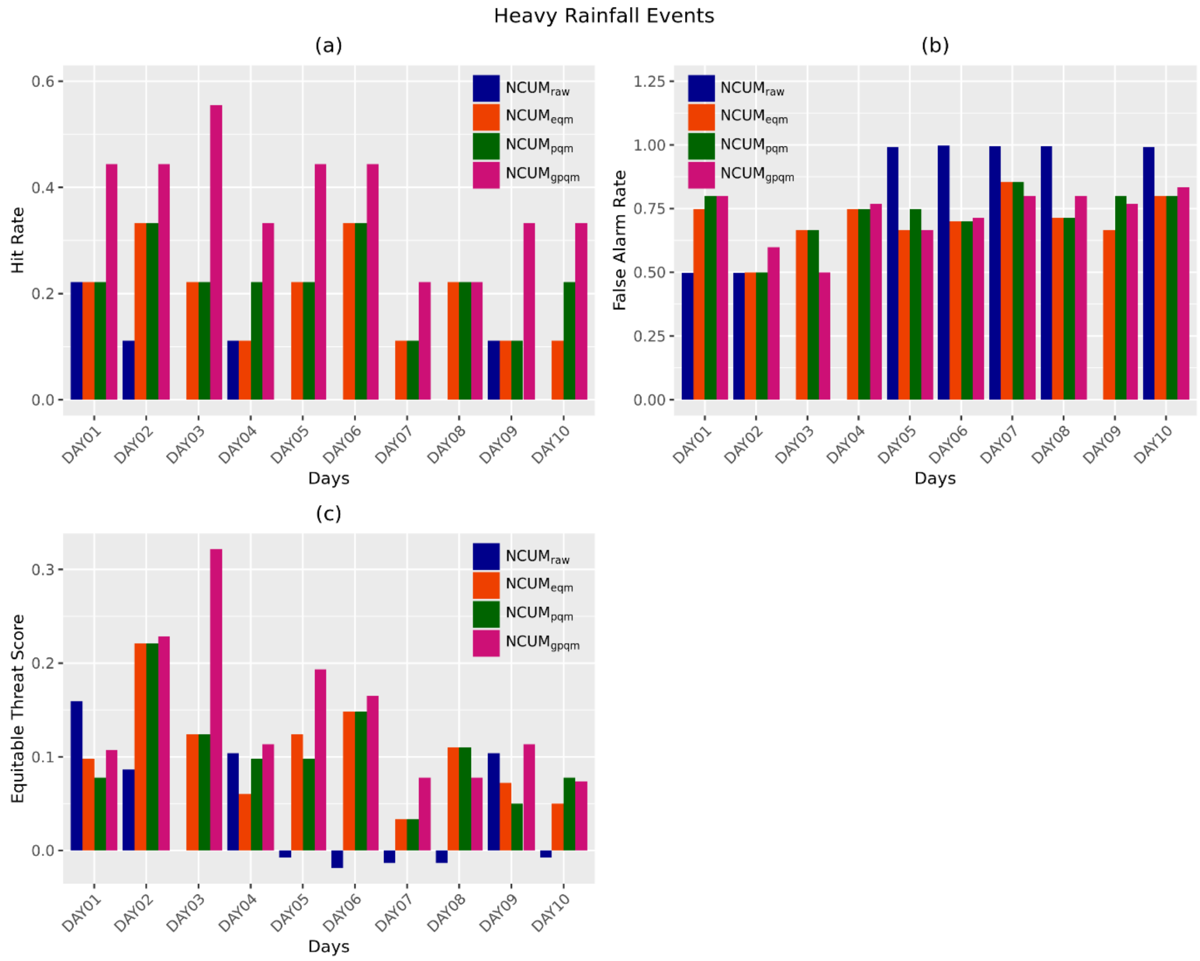

Fig.9. The categorical verification scores (a) the probability of detection (POD), (b) false alarm ratio (FAR), and (c) Equitable Threat Score (ETS) from the raw ($NCUM_{raw}$) and calibrated rainfall using different QM approaches at the BOM location for heavy rainfall events.

4.2. *Real time implementation of BC approaches for Mumbai during July 2023*

The Operational implementation of the various quantile mapping bias correction (QMBC) approaches to NCUM-G precipitation forecasts has been made. We have chosen two heavy rainfall cases as reported by IMD for this verification analysis. Here, we verify the heavy rainfall events that occurred on 21st July 2023 and 26th July 2023, to investigate how the bias-corrected rainfall derived using different QM methods improved such events at the BOM location. Fig. 10 illustrates the 21st July 2023 rainfall event in different lead forecast times. As we can see, on 21st July 2023, the observed rainfall exceeded 100mm/day, while $NCUM_{raw}$

exhibited a substantial underestimation of this event. The calibrated precipitation from various QM approaches, specifically, the GPQM demonstrates significant enhancement, ranging from 59 to 96.3 mm/day and maximum magnitude in Day-03 forecasts. Similarly, Fig. 11 shows another heavy rainfall event that occurred on 26^{th} July 2023 over BOM. Mumbai experienced heavy rainfall surpassing 100mm/day on 26th July 2023 as evident from the observation. The $NCUM_{raw}$ underestimated this event with the magnitude ranging from 8.49 to 31.48mm/day in Day-05 to Day-01 forecasts, respectively. In contrast, all the parametric methods show an improvement in forecasting this event in Day-01 to Day-03 forecasts, with magnitudes closely matching the observed rainfall in Day-01 forecast. In the Day-05 forecast, this event is missing in $NCUM_{raw}$. One possible reason for this absence could be the model's spatial dislocation of this synoptic event in its forecasts. Moreover, the calibrated rainfall from different QM techniques does not even predict the rainfall. It's worth highlighting that the statistical adjustment via QM relies on historical quantiles, making it challenging to bias-correct highly specific mesoscale events associated with substantial rainfall. Nevertheless, it remains crucial to correct biases in such specific events and an alternative for accurately identifying such forecasts is through the use of BC techniques based on synoptic events.

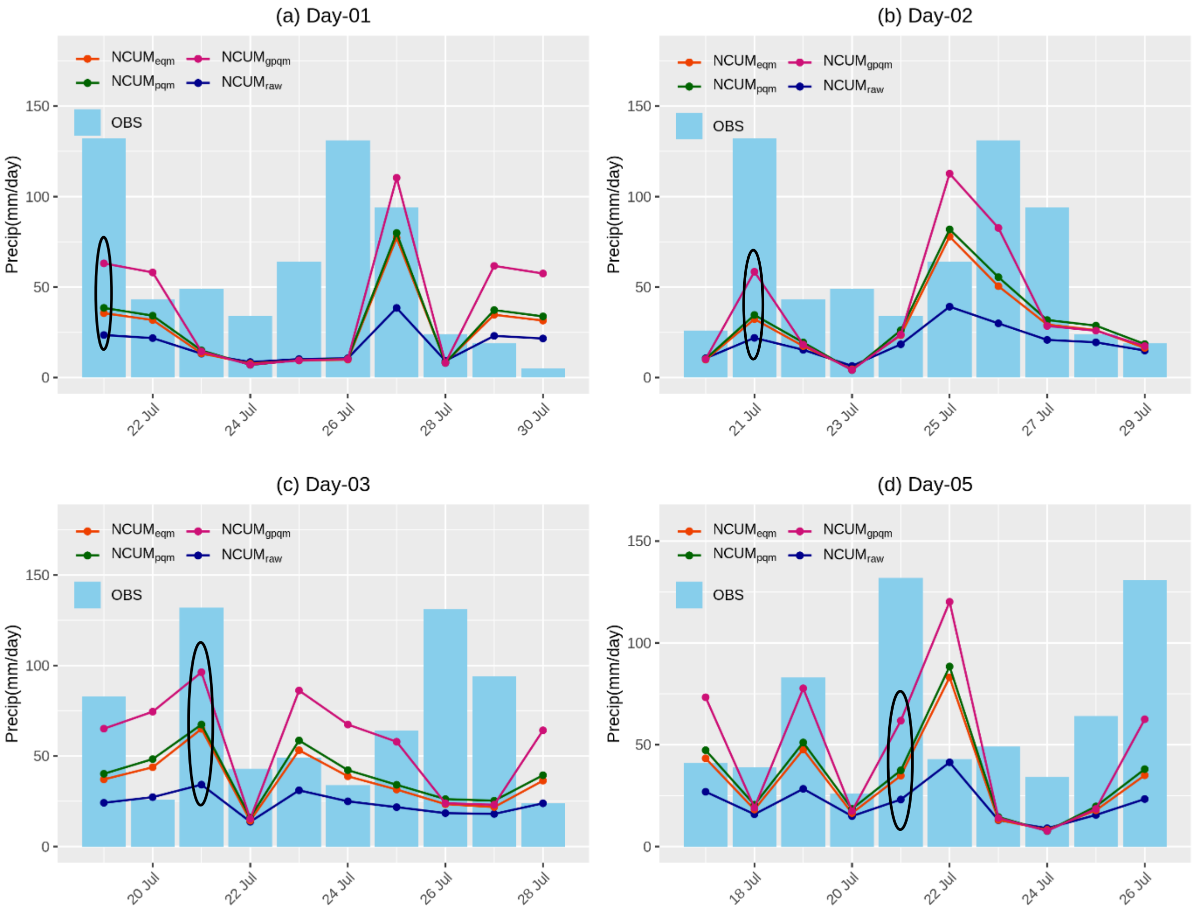

Fig.10. Time series of daily precipitation from NCUM-G raw forecast precipitation (NCUM$_{raw}$), bias-corrected precipitation forecast obtained using EQM (NCUM$_{eqm}$), PQM (NCUM$_{pqm}$), and GPQM (NCUM$_{gpqm}$) on 21st Jul 2023 over BOM.

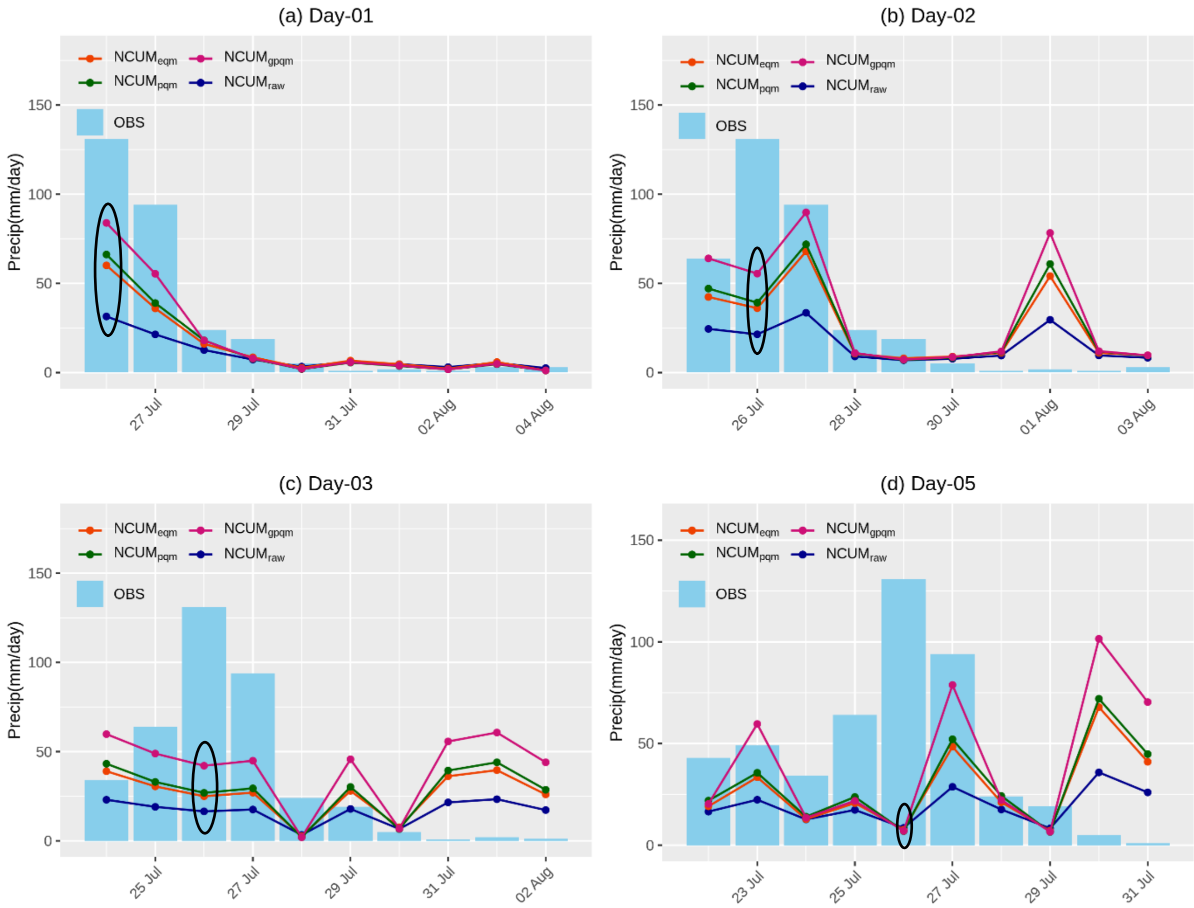

Fig. 11. Time series of daily precipitation from NCUM-G raw forecast precipitation (NCUM$_{raw}$), bias-corrected precipitation forecast obtained using EQM (NCUM$_{eqm}$), PQM (NCUM$_{pqm}$), and GPQM (NCUM$_{gpqm}$) on 26th Jul 2023 over BOM.

5. Summary and Conclusions

This report evaluates the quantile mapping bias correction (QMBC) approaches based on both empirical and parametric methods (i.e., EQM, PQM, and GPQM) to bias-correct the NCMRWF precipitation forecasts over Mumbai during the southwest monsoon of 2022. The IMDAA and station based IMD reanalysis datasets from 1979 to 2019 during JJAS is utilized for training purposes based on monthly quantile approach as discussed in methodology (Section 3).

Overall, the cdf of bias-corrected precipitation obtained from various QM approaches exhibits a substantial alignment with the IMD station dataset, specifically for the higher thresholds. Moreover, the daily rainfall time series shows that the GPQM method successfully predicts all the events while the raw forecasts fail to capture in all forecast lead times. The skill of calibrated rainfall is also verified for both moderate and heavy rainfall events. In moderate to heavy rainfall cases, all QM approaches have demonstrated high POD and ETS scores compared to $NCUM_{raw}$. Further, the FAR is included in the POD skill score to account for its inclination toward the frequency of correct predictions. In Day-01 to Day-04 forecasts, various Quantile Mapping (QM) approaches tend to overestimate false alarms for moderate to heavy rainfall events. However, the overestimation of moderate to heavy rainfall events in $NCUM_{raw}$ is slightly improved with forecast lead time. Thus, the results clearly indicate that the rainfall data calibrated using QM methods outperforms raw forecasts and is better suited for regional-scale flood warning applications.

Although this study provides valuable insight into the effectiveness of various QM approaches in improving the accuracy of precipitation forecasts, a critical aspect for the progression of flood models, however, it should be noted that the statistical adjustment via QM relies on historical quantiles, making it challenging to bias-correct substantial rainfall specifically associated with mesoscale events. Therefore, further analysis is warranted based on synoptical scale BC techniques to better understand and bias correct the extreme event forecasts at local scale.

References

1. 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
2. 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
3. Boyaj, A., Ashok, K., Ghosh, S., et al. 2018. The Chennai extreme rainfall event in 2015: The Bay of Bengal connection. *Clim. Dyn.* 50, 2867–2879. https://doi.org/10.1007/s00382-017-3778-7
4. 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
5. 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.
6. Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. *Springer*, 208 pp
7. 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

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

9. Gandin, L. S., & Murphy, A., 1992. Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*,120, 361–370

10. 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

11. 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

12. 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.

13. 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

14. 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

15. 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.

16. 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.

17. 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

18. Lyngwa, R. V., & Nayak, M. A., 2021. Atmospheric river linked to extreme rainfall events over Kerala in August 2018. *Atmos. Res.*, 253, https://doi.org/10.1016/j.atmosres.2021.105488

19. 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

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

21. Mishra, V., & Shah, H. L., 2018. Hydroclimatological Perspective of the Kerala Flood of 2018. *J. Geol. Soc. India* 92, 645–650. https://doi.org/10.1007/s12594-018-1079-3

22. Mishra, V., Aaadhar, S., Shah, H., Kumar, R., Pattanaik, D. R., & Tiwari, A. D., 2018. The Kerala flood of 2018: combined impact of extreme rainfall and reservoir storage. Hydrol. *Earth Syst. Sci. Discuss*. [preprint]. https://doi.org/10.5194/hess-2018-480

23. 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

24. 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

25. 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 (2010). https://doi.org/10.1007/s00704-009-0134-9

26. 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

27. 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

28. 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

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

30. 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.

31. 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

32. Webster, P. J., Magna, V. O., Palmer, T. N., Shukla, J., & Tomas, R. A., 1998. Monsoons: processes, predictability and the prospects for prediction, *J. Geophys. Res.,* 103, 4451-4510.