**Quantile mapping bias correction methods to IMDAA reanalysis for calibrating NCMRWF unified model operational forecasts**

Kondapalli Niranjan Kumar1\*, Mohana Satyanarayana Thota1, Raghavendra Ashrit1, Ashis K. Mitra1, Madhavan Nair Rajeevan2

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

2Ministry of Earth Sciences, India.

**Chart, line chart

Description automatically generated**

**Figure S1.** *The cumulative distribution from daily precipitation over CHN (left panel) and BOM (right panel) locations over the validation period (2009-2018). The observed (GHCN) and raw IMDAA (IMDAAnocal) are indicated by “black” and “red” lines. While the bias-corrected precipitation distribution from empirical quantile methods (EQM[IMDAAeqm] and EQM-MV[(IMDAAeqmv]) and parametric quantile methods (PQM[IMDAApqm] and GPQM[IMDAAgpqm]) are indicated in different colors*

The cumulative distribution in Figure S1 is plotted for the post-monsoon and summer monsoon seasons over the Chennai(CHN) and Mumbai(BOM) locations over the validation period between 2009-2018. We showed the observed (GHCN), raw IMDAA precipitation (IMDAAnocal), and calibrated precipitation from different quantile mapping methods described in section 3. The IMDAAnocal distribution is overestimating the weday frequency with respect to observed precipitation in the very light rainfall regime (<1mm/day). However, even the distribution in the light and moderate rainfall regimes does not follow the observed distribution. The different calibrated methods used in this study clearly improves the distribution close to the observed GHCN data. For instance, the empirical distribution methods are quite impressive in correcting adjusting the IMDAAnocal distribution very close to GHCN data. Nonethless, the parametric method based on gamma (PQM) also well fixed the raw precipitation. However, the PQM based on gamma and Generalised Pareto Distribution (GPQM) in the moderate rainfall regime deviates from observed distribution. Nevertheless, it quite interesting to see all the methods have significantly corrected the raw distribution.

The quantile-quantile plot (QQ-plot) of observed (OBS;”black line”), raw IMDAA reanalysis (IMDAAnocal; ” solid blue circles”), and calibrated IMDAA precipitation (IMDAAcal; “solid red circles) over the CHN location for the validation period of 2009-2018 is shown in Figure S2 shown below.

Chart, scatter chart

Description automatically generated

**Figure S2.** *The quantile-quantile plot between Observed (GHCN) and IMDAA precipitation at CHN location over a period 2009-2018. The “blue” solid circles indicate the raw IMDAA precipitation (IMDAAnocal) before bias correction, while the “red” solid circles indicates the bias corrected or calibrated precipitation (IMDAAcal) using (a) EQM (b) PQM (c) GPQM and (d) EQM-MV quantile methods.*

A clear underestimation of moderately heavy rainfall (approx. 5-100mm/day) and heavy to very heavy rainfall regime (>100mm/day) can be seen from the IMDAA dataset (Figure S2; “solid blue circles”) relative to observed station dataset (Figure S2; “black line”). Here, we used empirical (EQM) and parametric (PQM and GPQM) quantile methods to correct these biases. For instance, Figure S2a shows the calibrated precipitation (“solid red circles”) based on the EQM method. It is noted from the QQ-plot that the EQM calibrated precipitation is very well close to the observed distribution until low- to moderate- rainfall, while slight overestimation can be noted at heavy- to very heavy rainfall amounts. Similarly, the PQM calibrated precipitation indicates that the distribution is relatively well-matched, except that the overestimation is higher in heavy rainfall amounts relative to the EQM method. The GPQM further overestimates even in the low precipitation regime along with extreme rainfall amounts (Figure S2c). Therefore, both the gamma distribution and gamma + GPD parametric distributions overestimate the extreme rainfalls. Noting that the EQM performance better relative to the other parametric methods, we further applied the moving window technique that depends on the EQM. The EQM method in Figure S2a uses the whole training period; however, it is also interesting to see the correction of each day or a group of days independently by using a moving window (EQM-MV) technique. Hence, instead of the same correction function for the whole group, each day of the year centred in a moving window for precipitation correction. Here, we used a correction window of 30 days after several trials and errors, and the result of calibrated precipitation is shown in Figure S2d. It is interesting to note that both EQM and EQM-MV indicate similar calibrated precipitation distribution. However, the very heavy rainfall distribution is now slightly better in the EQM-MV method. Similarly, over the BOM location (Figure S3), we can see the IMDAA precipitation is underestimated relative to observed station data. It is interesting to note that both empirical and parametric quantile methods are performing well for BOM locations. Specifically, the light, moderate, and heavy rainfall amounts are well-matched with the observed precipitation distribution. However, the very heavy rainfall distribution still has slightly deviated from the observed quantiles. Overall, at both locations, the performance of the empirical method is relatively better.

Chart, line chart

Description automatically generated

**Figure S3.** *Same as Figure S2 except for “BOM” location*

Further, the annual cycle of monthly accumulated precipitation for the validation period (2009-2018) is shown for CHN and BOM locations in Figure S4(left panel) and Figure S4(right panel), respectively. Figure S4a and S4b show the empirical (EQM and EQM-MV) and parametric quantile (PQM and GPQM) methods over CHN location along with observed (GHCN) and raw IMDAAnocal precipitation. As we can see, the IMDAAnocal is underestimated relative to observed precipitation at both the CHN (also refer to Figure S2 and Figure S3), specifically, during the post-monsoon. While the calibrated IMDAA precipitation (IMDAAeqm and IMDAAeqmv) significantly improves the monthly totals in the inter-quantile range (IQR)along with extreme rainfall amounts. The EQM and EQM-MV approaches show similar precipitation improvement relative to observation (GHCN).

A picture containing text, writing implement, stationary, pencil

Description automatically generated

**Figure S4.** *The box plot indicating annual cycle of monthly accumulated precipitation over CHN (left panel) and BOM (right panel) locations over the validation period (2009-2018). The top panel indicates the bias corrected precipitation from empirical quantile methods (EQM[IMDAAeqm] and EQM-MV[(IMDAAeqmv]) while bottom panel indicates the parametric quantile methods (PQM[IMDAApqm] and GPQM[IMDAAgpqm]). In both panels also shown are observed (GHCN) and raw IMDAA (IMDAAnocal) precipitation at the respective locations. On each box, the central line is the median, and the edges of the box are 25th and 75th percentiles. The outliers are shown ‘o’ symbol.*

On the other hand, the parametric quantile methods (PQM and GPQM) overestimate the monthly rainfall amounts relative to empirical methods, primarily through the GPQM approach. Similar to Figures S4a and S4b, Figures S4c and S4d are at the BOM location. As noted in CHN location, the IMDAAnocal underestimates the monthly rainfall amounts relative to observed precipitation over the BOM. While the empirical methods (Figure S4c) again significantly improve the rainfall amounts during the summer monsoon season. The parametric methods (PQM and GPQM) also adjust the monthly totals relative to observed rainfall; however, the overestimation is still discernible, especially during June and July. It is worth noting that the calibrated precipitation shows clear improvement relative to observed GHCN precipitation, specifically in IQR. However, it is also important to note some overestimations compared to observed precipitation totals from the PQM and GPQM methods, resulting from overestimating extreme rainfalls. Nevertheless, the downscaled IMDAA rainfall has a better annual cycle of monthly accumulated precipitation. The critical point to be noted here that we have chosen annual quantiles irrespective of the season for bias correcting. However, the rainfall occurrence over these two stations is highly seasonal. Hence, a seasonal quantile correction would be more appropriate for better results (*refer main manuscript for more details*).