Quarterly Journal of the Royal Meteorological Society | RMetS

A comparison of two statistical postprocessing methods for heavy-precipitation forecasts over India during the summer monsoon

Michael Angus[1] | **Martin Widmann**[1] | **Andrew Orr**[2] |
Raghavendra Ashrit[3] | **Gregor C. Leckebusch**[1,4] | **Ashis Mitra**[3,5]

[1]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

[2]British Antarctic Survey, Cambridge, UK

[3]National Centre for Medium Range Weather Forecasting, Ministry of Earth Sciences Government of India, New Delhi, India

[4]Institute for Meteorology, FU Berlin, Berlin, Germany

[5]TERI School of Advanced Studies, New Delhi, India

Correspondence
Martin Widmann, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.
Email: m.widmann@bham.ac.uk

Present address
Michael Angus, Met Office, Exeter, UK

Abstract

Accurate ensemble forecasts of heavy precipitation in India are vital for many applications and essential for early warning of damaging flood events, especially during the monsoon season. In this study we investigate to what extent Quantile Mapping (QM) and Ensemble Model Output Statistics (EMOS) statistical postprocessing reduce errors in precipitation ensemble forecasts over India, in particular for heavy precipitation. Both methods are applied to day-1 forecasts at 12-km resolution from the 23-member National Centre for Medium Range Weather Forecasting (NCMRWF) global ensemble prediction system (NEPS-G). By construction, QM leads to distributions close to the observed ones, while EMOS optimizes the ensemble spread, and it is not a priori clear which is better suited for practical applications. The methods are therefore compared with respect to several key aspects of the forecasts: local distributions, ensemble spread, and skill for forecasting precipitation amounts and the exceedance of heavy precipitation thresholds. The evaluation includes rank histograms, Continuous Ranked Probability Skill Scores (CRPSS), Brier Skill Scores (BSS), reliability diagrams, and receiver operating characteristic. EMOS performs best not only with respect to correcting under- or overdispersive ensembles, but also in terms of forecast skill for precipitation amounts and heavy precipitation events, with positive CRPSS and BSS in most regions (both up to about 0.4 in some areas), while QM in many regions performs worse than the raw forecast. QM performs best with respect to the overall local precipitation distributions. Which aspects of the forecasts are most relevant depends to some extent on how the forecasts are used. If the main criteria are the correction of under- or overdispersion, forecast reliability, match between the forecasted distribution for individual days and observations (CRPSS), and the skill in forecasting heavy-precipitation events (BSS), then EMOS is the better choice for postprocessing NEPS-G forecasts for short lead times.

KEYWORDS
EMOS, heavy precipitation, India, quantile mapping, statistical postprocessing, summer monsoon, weather forecasts

1 | INTRODUCTION

Millions of people across India are exposed to natural disasters such as floods and landslides triggered by heavy-precipitation events, particularly during the summer monsoon season (Ali *et al.*, 2019; Gupta & Nair, 2011; UNDRR, 2020; van Oldenborgh *et al.*, 2016; Varikoden & Reji, 2022; Wallemacq *et al.*, 2015). These events are typically low-probability and isolated, and originate from interactions between synoptic-scale disturbances on scales of 1000 km or more with mesoscale convective systems on scales of 5 to 100 km, with possible orographic enhancement (Francis & Gadgil, 2006; Mohandas *et al.*, 2020; Sillmann *et al.*, 2017; Sreenath *et al.*, 2022; Srinivas *et al.*, 2018; Varikoden & Reji, 2022; Viswanadhapalli *et al.*, 2019). Disasters associated with extreme precipitation can result in large number of deaths (Mahapatra *et al.*, 2018; Ray *et al.*, 2021; UNDRR, 2020), as well as extensive damage to property and infrastructure, loss of livestock, and destruction of crops and agricultural land (Revadekar & Preethi, 2012). Moreover, the frequency, intensity and spatial variability of extreme precipitation events over India during the monsoon season have shown significant increasing trends over recent decades, which are predicted to continue to increase throughout the 21st century (Ali *et al.*, 2019; Ghosh *et al.*, 2012; Goswami *et al.*, 2006; Mukherjee *et al.*, 2018; Pattanaik & Rajeevan, 2010; Roxy *et al.*, 2017; Singh *et al.*, 2019; Sooraj *et al.*, 2016). Timely, high-quality, and reliable predictions of the likelihood of such extreme events and their distribution over India is therefore essential to provide effective early warnings for authorities to improve the response to and preparedness for disasters (Basher, 2006; Mahanta & Das, 2017; Uccellini & Ten Hoeve, 2019).

The National Centre for Medium Range Weather Forecasting (NCRMWF) in India produces numerical weather forecasts using global and regional configurations of the NCRMWF version of the UK Met Office Unified Model. Because deterministic forecasting of precipitation, and in particular of extreme events, is challenging due to the chaotic nature of weather and the associated exponential growth of forecast errors (caused, e.g., by model limitations in the representation of moist convection and errors in the initial conditions) ensemble forecasts are the preferred approach. They provide estimates for the range of plausible future states and thus quantify uncertainties, as well as yield probabilities for the occurrence of extreme weather events (Ashrit *et al.*, 2020; Mukhopadhyay *et al.*, 2021). The NCMRWF ensemble prediction system (NEPS) currently consists of (a) global forecasts (NCMRWF global ensemble prediction system [NEPS-G]) with 23 members (one control and 22 perturbed members) with lead times of up to 10 days at 12-km resolution in which convection is

partially resolved as well as parametrized, and (b) regional forecasts (NEPS-R) with 12 members (one control and 11 perturbed members) with lead times of up to 75 hr at 4-km resolution in which convection is explicitly resolved instead of parametrized (Ashrit *et al.*, 2020; Chakraborty *et al.*, 2021; Mamgain *et al.*, 2020a, 2020b; Mukhopadhyay *et al.*, 2021; Sarkar *et al.*, 2021).

The challenges of forecasting heavy precipitation associated with the monsoon for the Indian region have been pointed out in Mitra *et al.* (2011), who investigated multi-model ensembles and found low skill for lead times longer than three days for individual ensemble members as well as for the ensemble mean. Chakraborty *et al.* (2021) showed that the NEPS-G system had relatively good skill at forecasting low to moderate daily precipitation but had difficulty accurately representing higher amounts. This is likely because the 12-km resolution NEPS-G system is unable to properly capture regional convective processes, which are highly relevant for heavy precipitation (Dirmeyer *et al.*, 2012; Konduru & Takahashi, 2020; Sillmann *et al.*, 2017; Willetts *et al.*, 2017), and also does not simulate critical interactions with small-scale orographic features (Baisya & Pattnaik, 2019; Rotunno & Houze, 2007; Webster *et al.*, 2008). Moreover, the model needs a good representation of the link between the lower and the upper troposphere in order to correctly forecast convective instability and the associated bursts and breaks in the monsoon (Befort *et al.*, 2016).

Statistical postprocessing can provide for the foreseeable future an important tool to reduce errors in ensemble forecasts because it is computationally cheap and therefore can be applied to a large number of forecasts consisting of all ensemble members and lead times. Statistical postprocessing attempts to reduce simulation errors by adjusting model output based on observations and has been widely used in climate research (Maraun & Widmann, 2018) and also in weather forecasting (Vannitsem *et al.*, 2021). However, until now no statistical postprocessing methods for precipitation have been applied to NCRMWF ensemble forecasts. Developing and applying such methods is thus a timely contribution to improve heavy precipitation forecasts over India, and to provide the basis for further integration of meteorological and hydrological predictions (Nanditha & Mishra, 2021; Widmann, Blake, *et al.*, 2019).

In atmospheric science two types of postprocessing methods are used, and often include downscaling. The first are Perfect Prog methods, where statistical links are established between observed predictors and observed predictands, which are then applied to model output. The second are Model Output Statistics (MOS) methods, where links are established between simulated predictors and observed predictands. The former methods often link

large-scale atmospheric states to local or regional meteorological variables, while the latter are mostly used to link the same physical variables on similar spatial scales, although a moderate downscaling step is still often part of the approach. In the context of weather forecasting MOS is the preferred approach. Quantile Mapping (QM) is the most commonly used MOS method for climate simulations (e.g., Maraun & Widmann, 2018; Piani *et al.*, 2010; Themeßl *et al.*, 2011) and has also been applied to weather forecasts (Scheuerer & Hamill, 2015), including for improving forecasts for heavy precipitation in mid-latitudes (Javanshiri *et al.*, 2021; Jha *et al.*, 2018). It maps the forecasted, unconditional distribution derived from all forecasts at a given location onto the observed distributions. This also affects the conditional distributions given by the ensemble forecasts at a given time, but not in an optimized way. It cannot meaningfully compensate for errors in the simulated large-scale meteorological states (e.g., Eden *et al.*, 2012; Maraun *et al.*, 2017; Maraun & Widmann, 2018), but as NEPS-G forecasts up to day 5 represent the large-scale atmospheric circulation well (Chakraborty *et al.*, 2021), it is in principle suitable for being applied to these forecasts. In contrast, the MOS method Ensemble Model Output Statistics (EMOS) has been specifically developed to transform the distributions of the ensemble members for a given forecast such that the postprocessed forecasts perform better. There are different versions of EMOS depending on the assumptions for the conditional distributions, the transformations, and the optimization criteria (Javanshiri *et al.*, 2021; Schefzik, 2017; Scheuerer, 2014; Scheuerer & Hamill, 2015; Wilks & Hamill, 2007), and the details of our choice are explained in Section 2. *Data and methods*.

In this pilot study, we examine the suitability of QM and EMOS postprocessing for improving NEPS-G forecasts over India for the monsoon seasons 2018–2022 with respect to the whole range of values and with respect to heavy-precipitation events.

2 | DATA AND METHODS

2.1 | Ensemble forecast and observed precipitation data

We apply QM and EMOS to postprocess daily precipitation forecasts for all 23 NEPS-G ensemble members for the 2018–2022 monsoon seasons (June–September). Forecasts prior to 2018 could not be used because postprocessing is model-specific and the version of NEPS-G used in this study was only implemented in 2018.

For fitting and validating the QM and EMOS postprocessing models, daily precipitation observations for the monsoon season (June–September) are taken from the 0.25° × 0.25° Indian Meteorological Division (IMD) gridded dataset V6.9 for 1980–2022 (Pai *et al.*, 2014). This is based on distance-weighted interpolation of quality-controlled gauge measurements from a high-density network of rain gauge stations (Pai *et al.*, 2014), and is considered to be the best available dataset to accurately capture the frequency, distribution, and intensity of rainfall extremes over India (Falga & Wang, 2022; Gupta & Takahashi, 2022).

2.2 | Setups

The QM and EMOS postprocessing models are used to correct day-1 forecasts only as this ensures that the methods and their evaluation are not adversely affected by either potential spin-up problems in the day-0 analysis or decreased skill in predicting the synoptic- and mesoscale circulation for longer lead times (Chakraborty *et al.*, 2021). We thus follow a conceptually conservative approach in which the postprocessing models are used to correct precipitation biases that mainly stem from unresolved and parametrized local (at each grid cell) processes such as convection and orographic effects, but not from biases in the atmospheric states on larger spatial scales (Eden *et al.*, 2012). We note that the performance of QM and EMOS is likely to depend in different ways on the skill of the raw forecast, and additional analysis would be needed to assess the suitability of both methods for longer forecast lead times.

To avoid conflating bias correction with addressing spatial-scale differences between the forecasts and observations (Maraun & Widmann, 2018; Voulosciuk *et al.*, 2017), the 12-km resolution NEPS-G forecasts were re-gridded to the coarser 0.25° × 0.25° resolution IMD grid for the postprocessing. The original forecast resolution can be recovered in a straightforward way through spatial disaggregation of the postprocessed values using the ratios of the full-resolution forecast and the uncorrected spatial mean on the coarser grid. Postprocessing is then applied individually for each grid cell, with the QM distributions and EMOS parameters calculated separately at each location. For both QM and EMOS we apply a five-fold cross-validation. We use four monsoon seasons to fit the EMOS model or to calculate the forecast distribution for QM, and apply the EMOS or QM transformations to the fifth year of independent data. This is repeated using different independent years and the evaluation is then based on the concatenated five independently postprocessed forecasts.

We train QM and EMOS only on observed wet days at a given grid cell (defined as daily precipitation greater than 0 mm·day^{-1}). The reason is that we aim at improving

the forecasts specifically for heavy rainfall and excluding observed dry days from EMOS fitting avoids that the EMOS transformations are affected by potential systematic deficiencies of the NEPS-G model in forecasting dry days. Because observed dry days are excluded from training, the postprocessing is designed to improve the forecasts when precipitation is observed but cannot correct the probabilities for precipitation occurrence. The size of the training dataset varies between locations and between different four-year training periods. There are large regional differences with an average of 258 wet days per grid cell within the 488 days of the four-year fitting periods (58%).

2.3 | Postprocessing methods

2.3.1 | Quantile mapping

In the QM approach (e.g., Hay & Clark, 2003; Maraun & Widmann, 2018; Piani et al., 2010), for each forecast value of precipitation x_{sim} the quantile in the forecast probability distribution function (PDF) is determined and the forecast value is replaced with the value for this quantile in the observed PDF, that is, the corrected values $x_{\mathrm{sim}}^{\mathrm{corr}}$ are given by:

$$x_{\mathrm{sim}}^{\mathrm{corr}} = F_{\mathrm{obs}}\left(F_{\mathrm{sim}}^{-1}(x_{\mathrm{sim}})\right) \qquad (1)$$

where F_{obs} denotes the cumulative density function (CDF) of the observations and F_{sim}^{-1} the inverse CDF of the ensemble forecast values. By construction, this approach removes biases in the ensemble forecast distributions.

Given our focus on heavy precipitation we do not use empirical PDFs, because these are derived directly from the observed or simulated data, and the low number of heavy precipitation events in the 2018 to 2022 monsoon seasons leads to high sampling variability of the empirical PDFs. Instead, we use a parametric approach by fitting suitable distributions to the ensemble forecast members and observed local precipitation values. The former is calculated for all wet days in all 23 day-1 forecast members during the four-year June–September training periods, while the latter is calculated from IMD observations for all wet days during all monsoon seasons (June–September) from 1980 to 2022. Although the distributions depend to some extent on the reference period due to internal variability, we decided to use all available observations to obtain a more robust estimate for the distributions. We tested different types of distributions including Gamma, Generalized Extreme Value, and mixed distributions, where different distributions are fitted to precipitation below and above the local 90th percentile (e.g., Pastén-Zapata et al., 2020). Distributions that include the Generalized Extreme Value distribution

are dominated by the heaviest precipitation, resulting in unrealistic distributions (not shown). Consistent with Pastén-Zapata et al. (2020), the best fits are obtained with separate Gamma distributions for values below and above the 90th percentile threshold (not shown), and therefore we use the Double Gamma distribution for QM.

2.3.2 | Ensemble Model Output Statistics (EMOS)

EMOS considers the conditional PDFs of the forecast ensemble members at a given time and determines a transformation of these PDFs such that the postprocessed PDFs optimally fit the observations. It is applied locally, and in our implementation the means and variances of the transformed PDFs are linear functions of the ensemble forecast means and variances, using the same regression coefficients for every timestep (e.g., Gneiting et al., 2005; Schefzik, 2017; Wilks, 2006). More general transformations, for instance based on non-homogeneous regression (Scheuerer & Hamill, 2015), are in principle possible. In contrast to the QM approach, EMOS does not guarantee that the marginal distribution of all postprocessed values at a given location matches an observed target PDF.

The type of PDF that is used to model the postprocessed conditional PDFs can be adapted to the modelled variable. For temperature a Gaussian PDF is suitable, while we represent the precipitation PDF at each forecast time by a shifted Gamma distribution left-censored at zero, following Scheuerer and Hamill (2015) and Baran and Nemoda (2016). The Gamma distribution has shape (k) and scale (θ) parameters:

$$k = \frac{\mu^2}{\sigma^2} \text{ and } \theta = \frac{\sigma^2}{\mu}, \qquad (2)$$

where μ and σ are the mean and standard deviation of the distribution. In the most general linear version of EMOS the forecast values f_i of the ith ensemble members for a given location and time (where M is the ensemble size) are linked to the mean and variance of the postprocessed Gamma distribution through:

$$\mu = a + b_1 f_1 + b_2 f_2 + \ldots + b_M f_M \text{ and } \sigma^2 = c + ds^2, \quad (3)$$

where $a, b_1, b_2, \ldots, b_M, c$ and d are the regression coefficients that have to be estimated and s^2 is the variance of the raw forecast ensemble. However, following common practice we treat the members of an ensemble forecast as statistically indistinguishable, and therefore add the constraint that all regression coefficients $b_1, b_2, \ldots, b_M$ are identical, which simplifies the equation for the mean of the

postprocessed PDF to:

$$\mu = a + b\bar{f},\qquad(4)$$

where $\bar{f}$ is the ensemble mean for a given time:

$$\bar{f} = \frac{1}{M}\sum_{i=1}^{M} f_i,\qquad(5)$$

and $b/M = b_1 = b_2 \ldots = b_M$. Baran and Nemoda (2016) find in practical tests that this model performs best from a range of possible relations between the original ensemble and the postprocessed PDF. Following Schefzik (2017), the coefficients are estimated by optimizing the Continuous Ranked Probability Score (CRPS, see Equation 7) using the Broyden–Fletcher–Goldfarb–Shannon algorithm.

If the sampling from the postprocessed PDF is done such that realizations are taken at the same quantiles that are associated with the ensemble forecast members in the raw distribution, the following relation between an original ensemble member (f_i) and the postprocessed ensemble member $(\hat{f}_i)$ holds:

$$\hat{f}_i = a + b\bar{f} + \frac{\sqrt{c + ds^2}}{s}\left(f_i - \bar{f}\right) - q,\qquad(6)$$

where q is the shift parameter from the left-censored Gamma distribution. For further details on this, please see Baran and Nemoda (2016) and Schefzik (2017).

EMOS is implemented here in the R package ensemble *MOS* (Yuen *et al.*, 2013). As mentioned in Section 2.2. *Setups* only observed wet days at a given grid cell and the corresponding ensemble forecasts are used for model fitting and cross-validation is applied.

2.4 | Validation methods

Our evaluation is done for the entire Indian region as well as for nine grid-point test locations representing western (Mumbai and Rajasthan), southern peninsula (Kerala), northwest (Shimla and Delhi), central (Hyderabad) and northeastern (Patna, Bhubaneswar, and Meghalaya) regions (Figure 1). It involves (a) comparing the raw, post-processed, and observed precipitation distributions and ensemble spread at the nine test locations, and (b) using standard forecast validation metrics to assess local forecast quality for both the whole range of precipitation values, as well as for heavy precipitation. The standard validation metrics used are the Continuous Ranked Probability Skill Score (CRPSS) for the whole range of precipitation values, and the Brier Skill Score (BSS), reliability diagrams, Receiver Operating Characteristic (ROC) curves, and ROC Skill Score (RSS) for heavy-precipitation events, which we define as either the 90th or 75th percentile of the local observed daily precipitation. The skill scores use the raw ensemble as the reference forecast. In order to assess the quality of a climatological forecast relative to
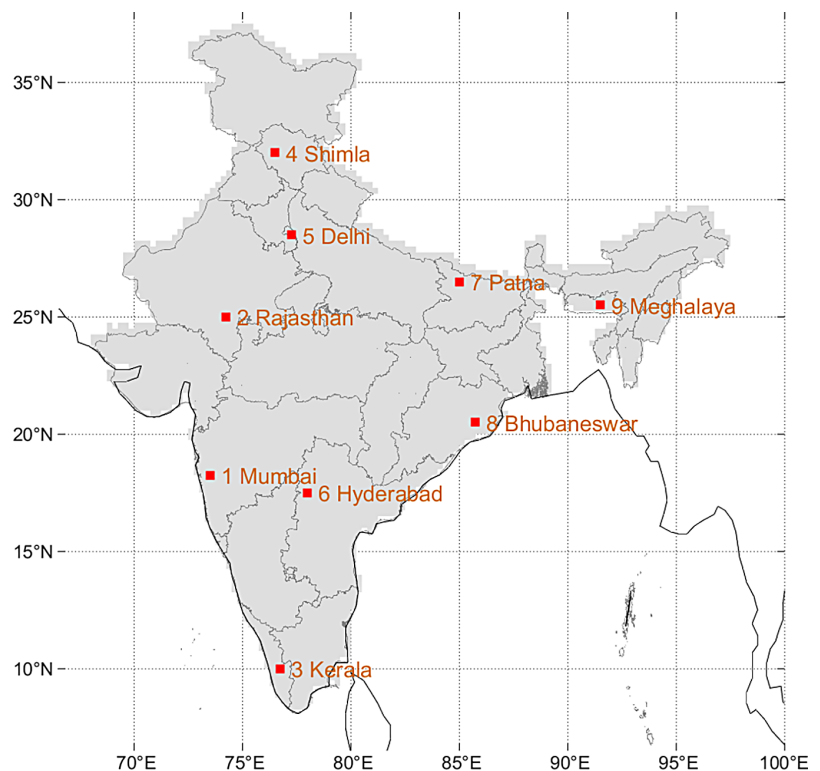

FIGURE 1 Locations of the nine test sites within India. Red squares represent chosen Indian Meteorological Division (IMD) grid cells, with labels providing nearest major city or region for reference.

the raw forecast we also calculate the CRPSS and BSS for a climatological forecast given by the distribution of daily precipitation on wet days during the 1980–2022 monsoon seasons at a given grid cell. The climatological probability of exceeding the 90th percentile of the precipitation distribution on wet days is then given by 0.1 scaled with the fraction of wet days in all days.

We calculate the raw and postprocessed distributions and quantile–quantile plots, show the ensemble spread as a function of the raw ensemble mean forecast, and provide rank histograms (Hamill, 2001) in which the observations are ranked relative to the ensemble members. If the ensemble has too little spread to be consistent with the forecast errors (underdispersive, overconfident ensemble), the observations will frequently fall outside of the ensemble range and have the lowest or highest rank, resulting in a U-shaped rank histogram. If the ensemble has too much spread (overdispersive, underconfident ensemble), the observations will frequently be ranked near the centre of the range resulting in a dome-shaped (or inverse U-shaped) histogram.

The CRPS quantifies the mean across all forecasts of the integral of the squared difference between the forecasted cumulative distributions and the step function distributions associated with the observations (Hersbach, 2000). This is defined as:

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left\{F_i(y) - I_{y\geq x_i}\right\}^2 dy, \qquad (7)$$

where N is the total number of forecast days available, F_i are the ensemble forecast CDFs for days i, x_i are the observations for days i, and $I_{y\geq x_i}$ is the Heaviside step function that changes from 0 to 1 when the integration variable y equals the observation x_i. Smaller CRPS values indicate better performance, with a perfect forecast where all ensemble members predict the observed value having a CRPS of 0. Using this, the CRPSS is then calculated as:

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}}, \qquad (8)$$

where $CRPS_{ref}$ is the CRPS of a reference forecast, which in our case is the raw forecast. Positive (negative) CRPSS values therefore indicate better (worse) performance than the raw forecast, with a value of 1 indicating a perfect forecast.

The Brier Score (BS) measures how well a probabilistic forecast predicts binary events by calculating the average squared difference between the forecast probabilities for the event occurring and the observations (Brier, 1950; Javanshiri et al., 2021; Wilks, 2011). This is

defined as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(y_i - o_i)^2, \qquad (9)$$

where y_i is the forecasted probability of occurrence of an event for day i (with values ranging from 0 to 1), o_i is the observation for day i with binary values (0 for event not observed, 1 for observed). The lower the BS the better the forecast performance, with a value of 0 being associated with a perfect forecast where all ensemble members correctly predict the outcome. For the BS we use the local 90th percentile of observed precipitation as the event threshold. Using this, the BSS is then defined as:

$$BSS = 1 - \frac{BS}{BS_{ref}}, \qquad (10)$$

where BS_{ref} is the BS of the reference (raw) forecast. Positive (negative) BSS values therefore indicate better (worse) performance than the reference forecast, with a value of 1 obtained for a perfect forecast.

The ROC curves assess the ability of probabilistic forecasts to discriminate an event from a non-event by plotting the hit rate (number of correctly forecasted events/total number of events) against the false alarm rate (number of wrongly forecasted events/total number of non-events) for different thresholds of the forecasted probability that are used to transform the probability into a binary decision on the event occurrence (Javanshiri et al., 2021; Wilks, 2011). Consistent with the BSS calculations we use the local 90th percentile of observed precipitation as the event threshold. In good discrimination forecasts, the ROC curve approaches the top-left corner (high hit rate and low false alarm rate for most probability thresholds), while in poor discrimination forecasts these curves are close to the diagonal line (Mason & Graham, 1999). Using this, the RSS is defined as:

$$RSS = \frac{A - A_{ref}}{1 - A_{ref}}, \qquad (11)$$

where A is the area under the ROC curve of the postprocessed forecast, and A_{ref} is the area under the curve of the reference (raw) forecast. Positive (negative) RSS values indicate better (worse) discrimination than the reference forecast, with a value of 1 obtained for perfect discrimination for all probability thresholds.

Finally, an important criterion for the quality of probabilistic binary forecasts is how close the observed relative frequency of an event is to the forecast probability. This is analyzed with reliability diagrams (Hamill, 1997; Wilks 2011), which plot the observed event frequency against the forecast probability for a particular threshold value. A perfectly reliable forecast for which the frequency of occurrence is equal to the forecast probability leads to a diagonal
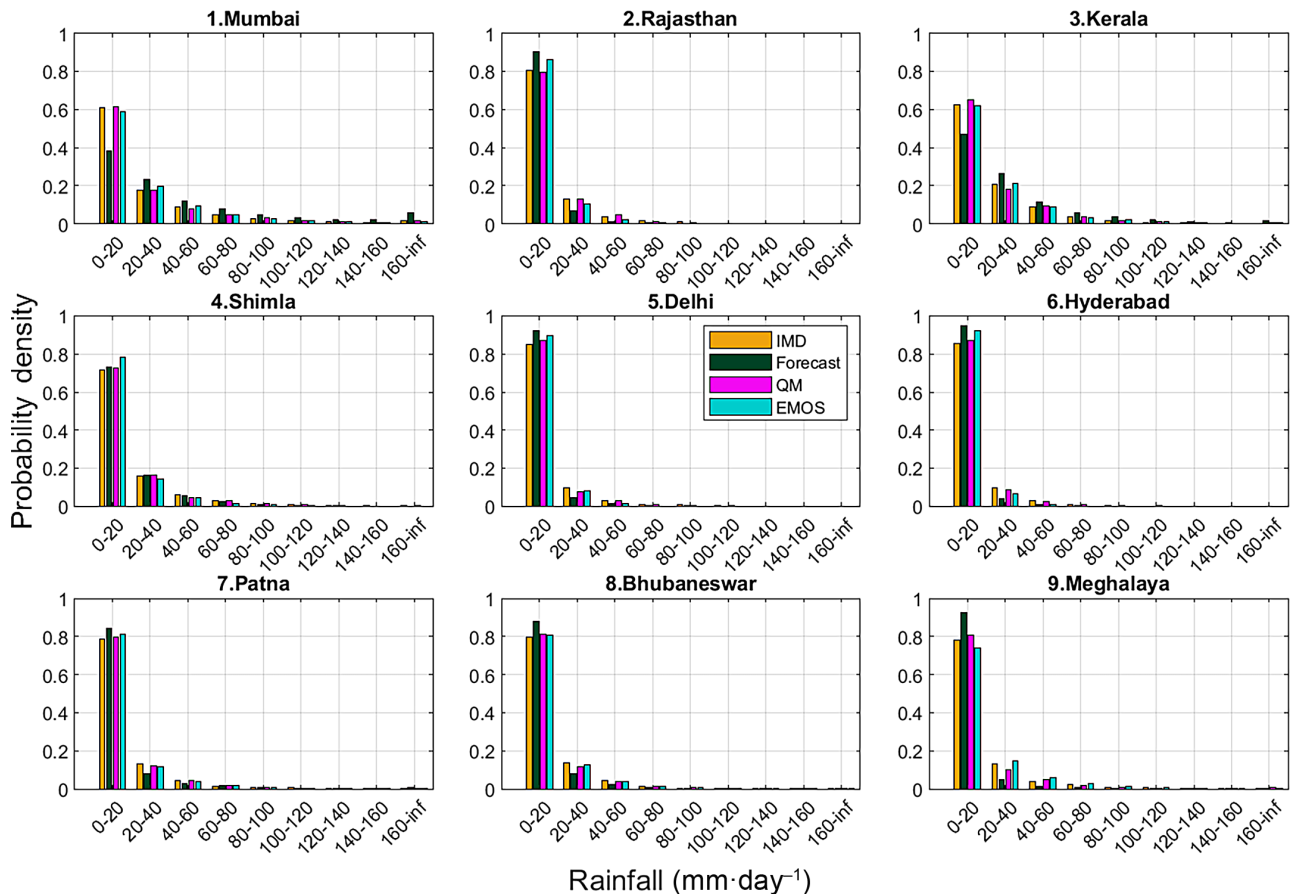
FIGURE 2 Relative frequency of observed (Indian Meteorological Division [IMD]) daily precipitation, and of raw, Quantile Mapping (QM)-postprocessed, and Ensemble Model Output Statistics (EMOS)-postprocessed forecasts for observed wet days (precipitation > 0 mm·day^{-1}) in the 2018–2022 monsoon seasons (June–September) at the nine test locations.

line in the reliability diagram. It would be preferable to use again an event threshold defined as the observed local 90th percentile but there are too few cases to robustly estimate the observed relative frequencies for a given forecast probability, resulting in noisy and non-informative reliability diagrams. We therefore use the local 75th percentile as the event threshold for reliability diagrams.

3 | RESULTS: DISTRIBUTIONS AND ENSEMBLE SPREAD

Figures 2–5 compare the raw, postprocessed, and observed precipitation distributions and ensemble spread at the nine test locations. The figures are based on only the observed wet days (precipitation >0 mm·day^{-1}) during the 2018–2022 monsoon seasons and the corresponding raw and postprocessed forecasts. We take this approach to be consistent with the model-fitting setup, and because assessing the performance of the postprocessing on observed dry days is not an objective of this study.

Prior to evaluation we have calculated maps for the coefficients in Equation (6) that specify the EMOS transformation (not shown). All coefficients are constrained to be non-negative. The overall shift in the mean (coefficient a) is over most of India between zero and about 50% of the mean. Over dry areas it can reach up to 200% but this is still only a small absolute shift. The proportionality factors between the mean of the postprocessed distribution and the simulated ensemble mean (coefficient b) lie mainly between 0.4 and 0.8 but there are also lower values and high values reaching up to about 1.5. The scaling of the ensemble spread is more difficult to assess because the combined effect of the two coefficients c and d depends on the standard deviation of the ensemble spread. For the average standard deviation there are a few locations with a reduction in the ensemble spread and large areas with a small or moderate (up to a factor of 3) and also areas with a large (up to a factor of 5) increase in the ensemble spread. In a few isolated locations the increase is even higher. However, it will be shown later that in heavy precipitation situations EMOS reduces the spread at the majority of the test locations.
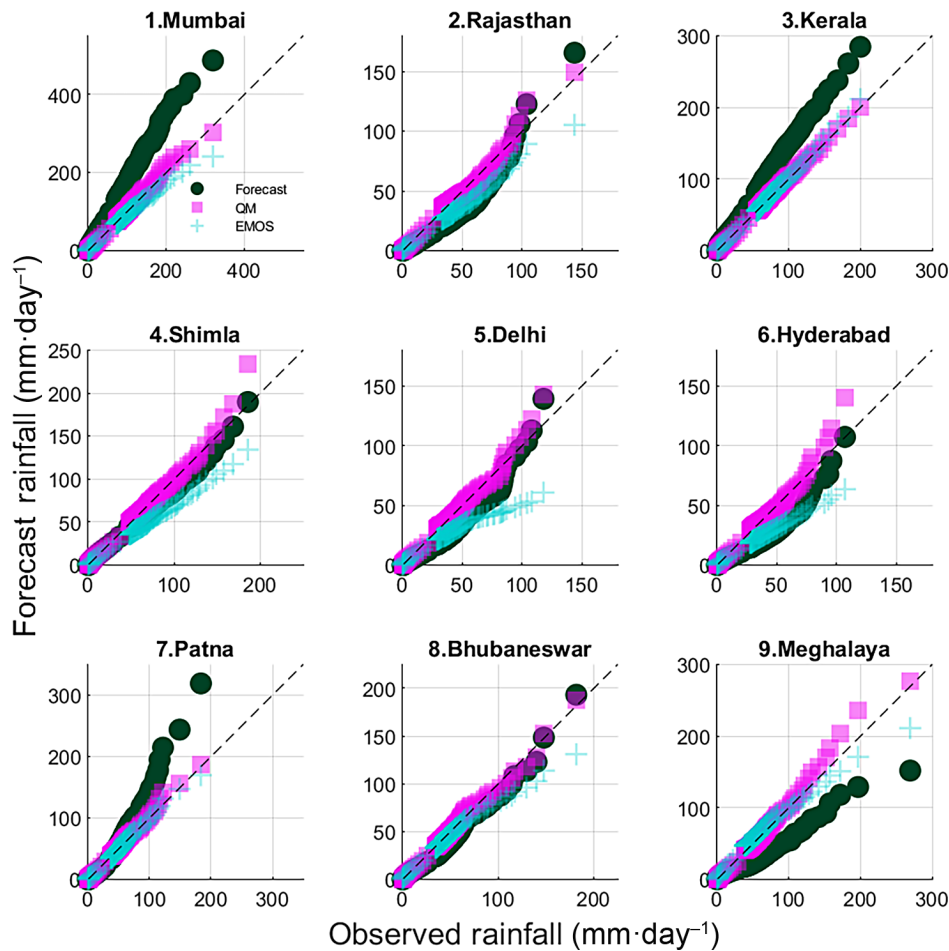
FIGURE 3 Quantile-quantile plots for raw, Quantile Mapping (QM)-postprocessed, and Ensemble Model Output Statistics (EMOS)-postprocessed forecasts of daily precipitation against observations for observed wet days (precipitation > 0 mm·day^{-1}) during the 2018–2022 monsoon seasons (June–September) at the nine test locations. Quantiles shown from 1st to 90th are at intervals of 5 percentile points, and 0.1 percentile points from 90th to 100th.

Frequency histograms of the raw forecast, QM- and EMOS-postprocessed forecasts, and observed daily precipitation at the test locations for observed wet days during the 2018–2022 monsoon seasons are shown in Figure 2. It is noteworthy that based on the observations the 0–20 mm·day^{-1} bin makes up the largest fraction of daily precipitation for most of the locations (Rajasthan, Shimla, Delhi, Hyderabad, Patna, Bhubaneswar, Meghalaya), with frequency values of around 0.8 and over, that is, the occurrence of relatively low rainfall amounts is high and the precipitation distributions are severely skewed. At these locations the raw forecast overestimates the fraction of precipitation in this bin. Exceptions are the wetter locations of Mumbai and Kerala, which have reduced frequency values for the 0–20 mm·day^{-1} bin (around 60%) but higher frequency values for the higher rainfall bins, e.g., 20% for the 20–40 mm·day^{-1} bin and 10% for the 40–60 mm·day^{-1} bin. At these two locations the raw forecast underestimates the fraction of precipitation in the 0–20 mm·day^{-1} bin but overestimates the fraction for the higher rainfall bins.

As expected, Figure 2 shows that the QM-postprocessed distributions are generally close to the observed ones regardless of whether the raw forecast overestimates or underestimates the number of days in a given precipitation bin. For example, for Kerala QM corrects the underestimation of the frequency of precipitation values in the 0–20 mm·day^{-1} bin, as well as the overestimation of the frequencies in the 20–40, 40–60, and 60–80 mm·day^{-1} bins. Although the purpose of QM is to bring the postprocessed distribution close to the observed one, the QM and observed distributions are not identical because of the cross-validation and because the target distribution in QM is the Double Gamma distribution fitted to the IMD observations for 1980–2022 rather than the empirical distribution for the 2018–2022 monsoon seasons. It is also noteworthy that at some locations the raw forecasts are in relatively good agreement with the observations (e.g., Shimla and Patna).

By contrast, as EMOS optimizes the ensemble spread it does not guarantee that the postprocessed local distributions are close to the observed ones. In many locations the QM distributions are therefore closer to the observed ones than the EMOS distributions. This is especially apparent for the 0–20 mm·day^{-1} bins, which dominate the distribution. At some locations the EMOS distributions are similar to the raw forecast but substantially different from the observed (e.g., Rajasthan and Hyderabad).

FIGURE 4 Observations and ensemble mean and spread of daily precipitation forecasts against the ensemble mean of the raw forecast for the raw forecast (green), Quantile Mapping (QM)-postprocessed (pink), Ensemble Model Output Statistics (EMOS)-postprocessed (blue), and Indian Meteorological Division (IMD) observations (gold) at the nine test locations. Only values with a raw forecast mean over the 90th percentile of observed wet days during the 2018–2022 monsoon seasons (June–September) are shown. The ensemble mean is shown by the respective symbol and the spread by the bars which show ±2 standard deviations.
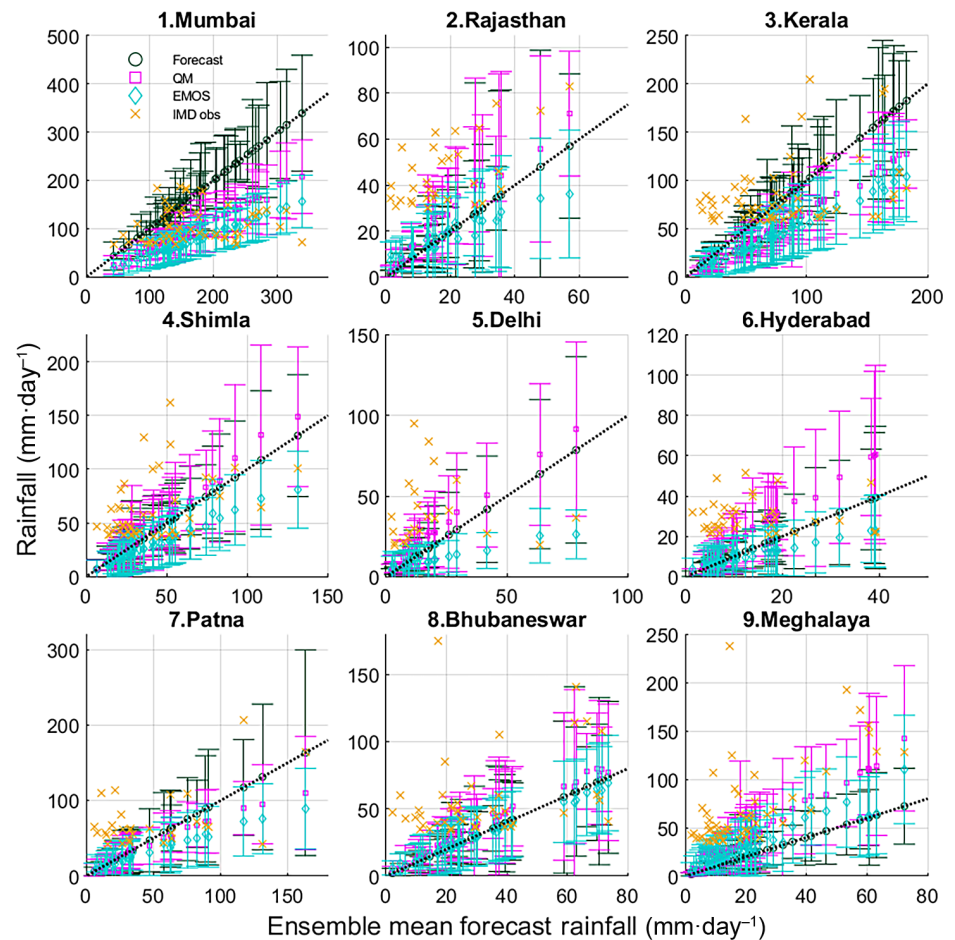
However, EMOS outperforms QM in Kerala, especially in the 20–40 mm·day^{-1} bin where the raw forecast is wetter than the observations.

A comparison of the raw and postprocessed forecast precipitation distributions and the observed distributions is also shown as quantile–quantile plots in Figure 3. They show that for the lower quantiles at some locations the raw forecasts tend to agree relatively well with the observations (Shimla, Rajasthan, Delhi, Bhubaneswar), while at other locations the raw forecast distributions can be either skewed to the left (Mumbai, Kerala, Patna) or to the right (Meghalaya) compared to the observation distribution. At the majority of locations, the QM distributions are closer than the EMOS distributions to the observed distributions, while at some locations the QM and EMOS distributions are very similar (Kerala, Patna, Bhubaneswar). The EMOS distributions are sometimes closer to the observed ones than those from the raw forecasts (especially at Mumbai, Kerala, Meghalaya). At all locations the largest disagreements between the raw forecast and observed values occur for the upper quantiles, which are corrected much better using the QM method compared to the EMOS method (especially at Delhi and Meghalaya). Note that by construction, the upper quantiles in the QM method are

always shifted towards the observed values, while in some cases (Mumbai, Shimla, Delhi) the EMOS method results in even larger differences to the observations.

In Figure 4 we investigate how the postprocessing affects the ensemble mean and spread. For QM the change in spread is strongly related to how the method affects individual values and the ensemble mean. When the raw forecast underestimates (overestimates) the observed rainfall, the increase (decrease) in values through QM is similar to a scaling with a factor larger (smaller) than 1, and thus an increase (decrease) in the ensemble mean is associated with an increase (decrease) in the ensemble spread. An example for underestimation is Meghalaya (cf. Figures 2 and 3), where QM has a larger ensemble spread than the raw forecast, while Kerala is an example for overestimation which leads to a reduction of ensemble spread by QM. For other locations the link is less clear because over- and underestimation vary between quantiles and QM is based on fitting separate distributions for the lower and higher values.

In EMOS the postprocessed ensemble mean and spread are determined through optimizing the CRPS. The change in the ensemble spread is independent of the change of the ensemble mean, and the spread is therefore not directly
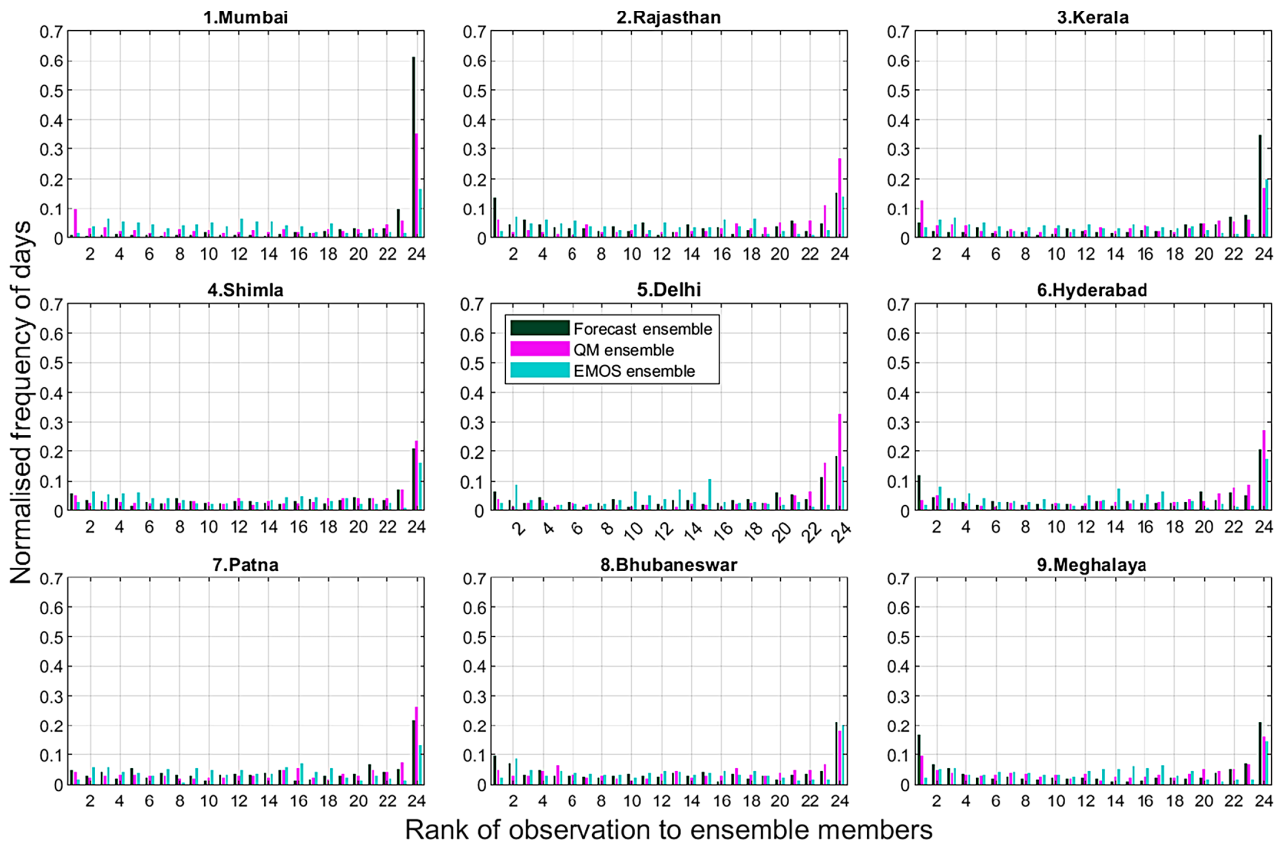
FIGURE 5 Rank histogram of observed daily precipitation at the nine test locations with respect to the 23 ensemble members of the raw, Quantile Mapping (QM)-postprocessed, and Ensemble Model Output Statistics (EMOS)-postprocessed forecasts for observed wet days (precipitation >0 mm·day^{-1}) for the 2018–2022 monsoon seasons (June–September). A rank of 1 (24) means the observed precipitation is higher (lower) than all 23 ensemble members.

related to over- or underestimation. For example, in Kerala the EMOS ensemble mean decreases and the spread is similar to the raw forecast, while in other locations with a decrease in the mean there is a reduction of the spread (Mumbai, Shimla, Delhi, Patna). In Bhubaneswar EMOS reduces the ensemble spread while there is no substantial change in the mean. At the majority of the test locations the ensemble spread is reduced by EMOS, while in Meghalaya there is a tendency for increased spread.

The rank histograms (Figure 5) for the raw forecast are at some test locations U-shaped, which indicates an underdispersive ensemble, and in other locations only show a high fraction of observations in the lowest rank but not in the high ranks, which means all forecast ensemble members are often too wet while not also often being too dry. The EMOS approach leads at all test locations to some extent to more even, but not in all cases flat, distributions of the observation ranks. In contrast, QM does not improve the rank histograms and in several cases exacerbates the problems. For example, the raw forecasts for Rajasthan are too dry for most quantiles with the exception of very high ones (Figure 3) but there are also too many individual forecasts that are wetter than the observations

(high number of rank 24 in Figure 5). As QM increases most of the values to correct the low bias, the fraction of individual forecasts that are wetter than the observations increases even more.

The link between how EMOS affects the ensemble spread in Figure 4 and how it modifies the rank histograms is not straightforward because Figure 4 only shows days when the ensemble mean forecast was above the local 90th percentile. Nevertheless, the results are consistent because the locations with a strong reduction of ensemble spread in Figure 4 (Mumbai, Shimla, Delhi, Patna, Bhubaneswar), which indicates an overdispersive ensemble, are different from the locations for which the rank histograms in Figure 5 indicate an underdispersive ensemble (Rajasthan, Hyderabad, Meghalaya).

4 | RESULTS: VALIDATION METRICS

As in Section 3 the evaluation in this section is also based on only the observed wet days during the 2018–2022 monsoon periods and the corresponding raw and
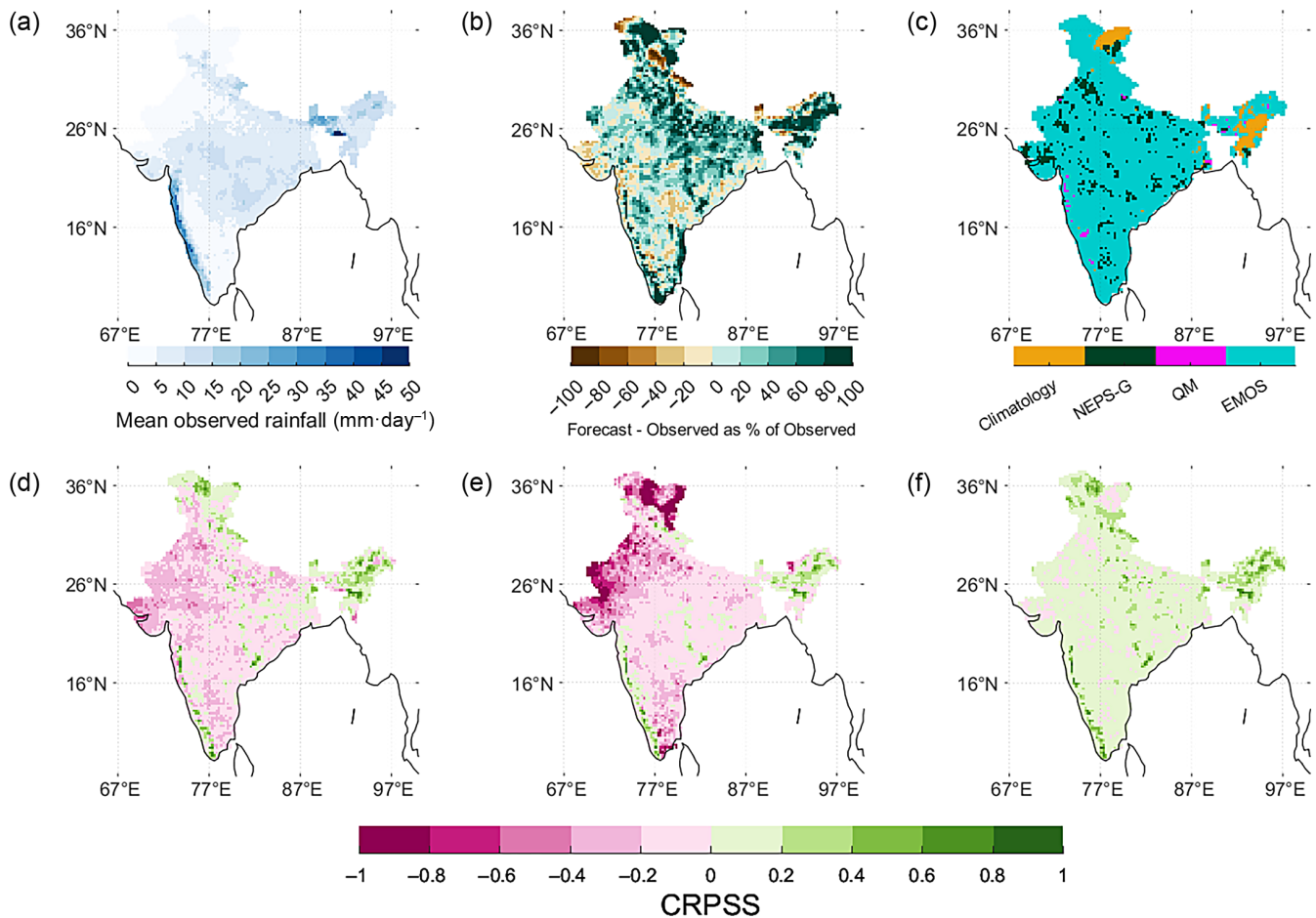
FIGURE 6   (a) Observed (Indian Meteorological Division [IMD]) mean daily precipitation for the 1980–2022 monsoon seasons (June–September); (b) difference between precipitation means for the 2018–2022 monsoon seasons in the raw forecast and in the IMD observations divided by the observed mean (relative bias); (c) method that yields the best Continuous Ranked Probability Skill Scores (CRPSS) at each location. The CRPSS for forecasts for daily precipitation on observed wet days for the monsoon seasons 2018–2022 are shown in (d–f) with (d) for the climatological forecast, (e) for the Quantile Mapping (QM)-postprocessed forecasts and (f) for the Ensemble Model Output Statistics (EMOS)-postprocessed forecasts. Positive (negative) CRPSS indicates better (worse) performance than the raw forecast. NEPS-G, National Centre for Medium Range Weather Forecasting (NCMRWF) global ensemble prediction system.

postprocessed forecasts. For most parts of India, the CRPSS for daily precipitation amounts for the climatological forecast (Figure 6d) and for the QM-postprocessed forecast (Figure 6e) is negative, with values for QM lower than −0.8 over the northwest region. This means these forecasts are considerably worse than the raw forecasts. Exceptions with positive CRPSS for the climatological and QM-postprocessed forecasts include the western coast of India and some localized central, northern, and northeastern regions, which generally have a higher average precipitation (Figure 6a) and are influenced by mountain ranges such as the Western Ghats along the western coast and the Himalayas to the north. The CRPSS for the climatological forecast is also positive in the dry, mostly northern areas, which are also strongly influenced by topography. There is no systematic relationship between

the bias of the raw forecast (Figure 1b) and the CRPSS of the climatological or QM-postprocessed forecast.

The CRPSS for EMOS is positive in most regions across India, with values reaching 0.4 and in a few locations up to 0.8 over the western coast and some central and eastern regions (Figure 6f). The exceptions to this are the relatively dry areas in the northwest (Figure 6a) where the CRPSS is slightly negative, but not as negative as for QM. Figure 6c shows for each location the forecast type with the highest CRPSS and EMOS performs best over most of India. Over the regions where the EMOS CRPSS is negative the raw forecast is slightly better, while there are only a few isolated areas where QM would be the best choice. In the dry northern and northwestern regions the climatological forecast outperforms both the raw and postprocessed forecasts, which indicates that in these regions
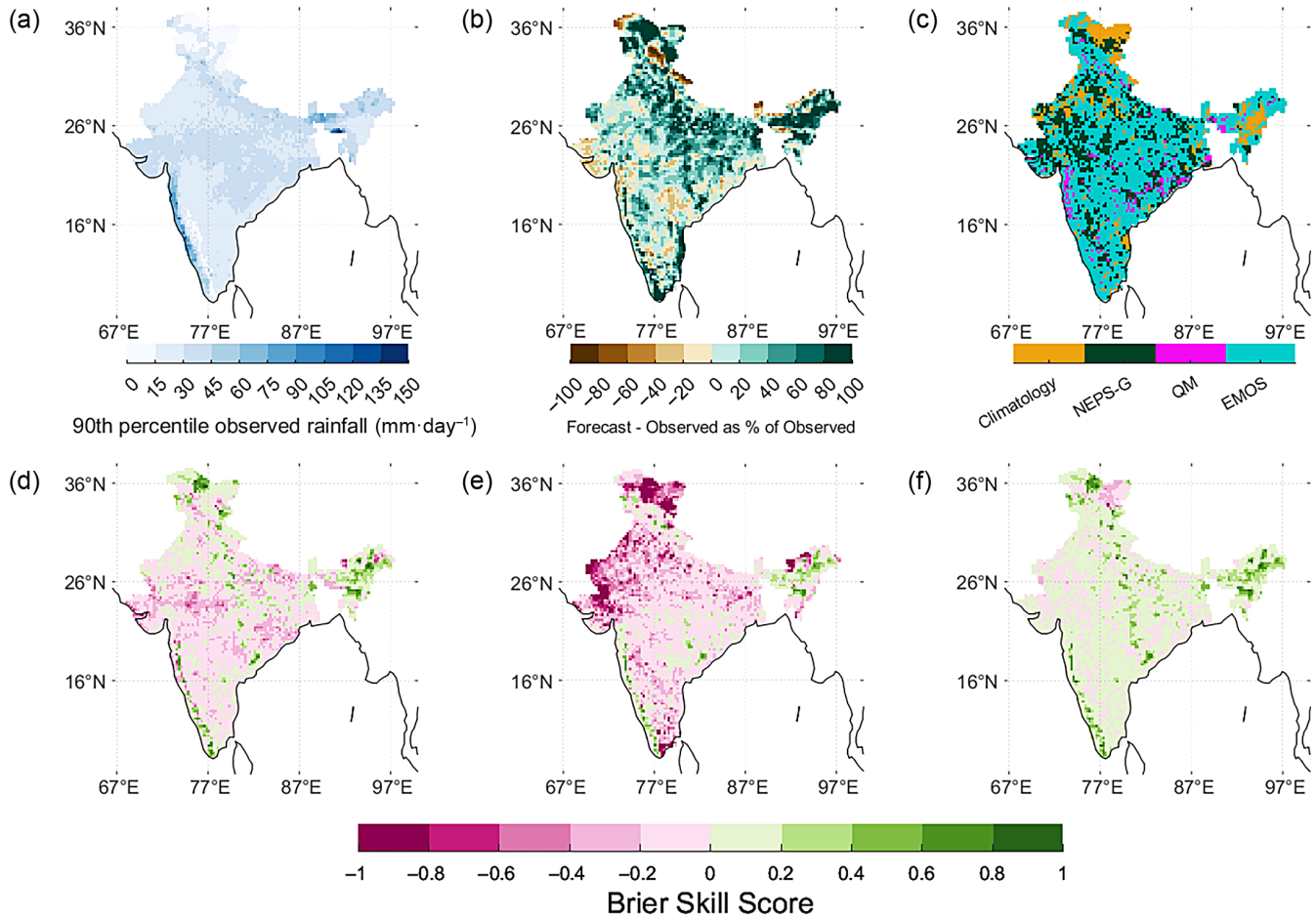
FIGURE 7 (a) Observed (Indian Meteorological Division [IMD]) 90th percentile of daily precipitation on wet days for the 1980–2022 monsoon seasons (June–September); (b) difference between precipitation means for the 2018–2022 monsoon seasons in the raw forecast and in the IMD observations divided by the observed mean (relative bias); (c) method that yields the best Brier Skill Scores (BSS) at each location. The BSS for forecasts for the probability of exceeding the 90th percentile (defined as in panel a) on observed wet days for the monsoon seasons 2018–2022 is shown in (d–f) with (d) for the climatological forecast, (e) for the Quantile Mapping (QM)-postprocessed forecasts and (f) for the Ensemble Model Output Statistics (EMOS)-postprocessed forecasts. Positive (negative) BSS indicates better (worse) performance than the raw forecast. NEPS-G, National Centre for Medium Range Weather Forecasting (NCMRWF) global ensemble prediction system.

the raw forecast contains not much information about the daily precipitation variability that can be exploited through postprocessing.

The comparison of the BSS for the exceedance of the local 90th percentile of precipitation (Figure 7) yields results that are broadly similar to those for the CRPSS (Figure 6). The BSS for the climatological forecast (Figure 7d) and for QM (Figure 7e) is negative almost everywhere with values lower than −0.8 in the same regions in which the CRPSS has the lowest values. The exception are slightly positive BSS values in the same western, central, northern and northeastern regions for which the CRPSS for the climatological and QM-postprocessed forecasts is positive. This means both forecasts are considerably worse than the raw forecast for both the daily mean and the exceedance probability of the 90th percentile. The

BSS for EMOS (Figure 7f) is positive almost everywhere, with values reaching 0.4 and in a few locations up to 0.8. Again, EMOS generally performs best over most of India (Figure 7c), but in comparison with the CRPSS results the BSS maps show more small-scale variability, characterized by randomly scattered locations where the raw forecast performs best, and there are also a few more locations where QM or the climatological forecast performs best. These characteristics may partly reflect the small spatial scales of heavy-precipitation events, and also their smaller sample size and associated higher sampling variability of performance measures.

The reliability diagrams for exceedance of the 75th percentile of precipitation (Figure 8) are in general quite noisy, which is likely because of the relatively limited sample sizes. Nevertheless, they are still informative, and

FIGURE 8 Reliability diagram for the exceedance of the 75th percentile of daily precipitation for observed wet days during the 2018–2022 monsoon seasons (June–September) for raw, Quantile Mapping (QM)-postprocessed, and Ensemble Model Output Statistics (EMOS)-postprocessed forecasts at the nine test locations. The horizontal axis shows the forecasted event probability, and the vertical axis the observed relative frequency.
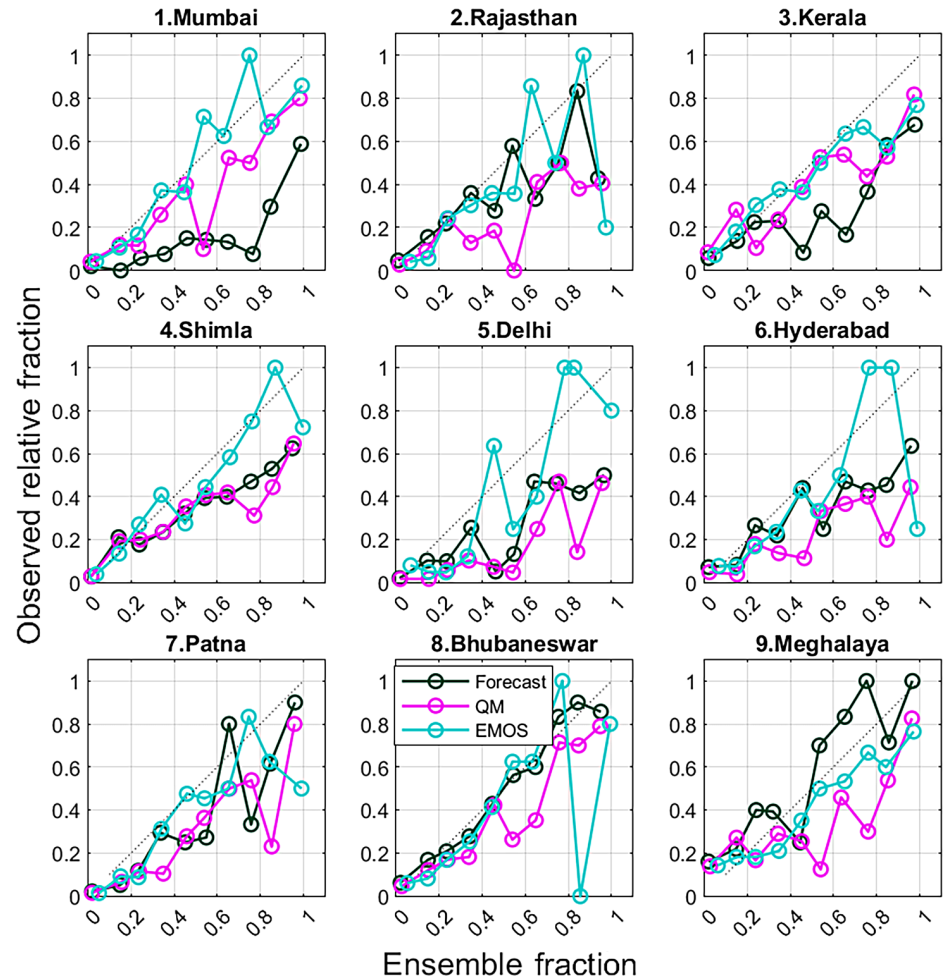
show that for the majority of locations the slope for the raw forecasts is shallower than the diagonal, which means that the forecast probability of an event is greater than the observed relative frequency. For example, when the raw forecast probability at Shimla is 80% the actual chance of observing the event is about 50%. The exceptions are Rajasthan, Bhubaneswar, and Meghalaya, where the slope is closer to the diagonal for most of the probability range. In some cases, QM improves the reliability (Mumbai, Kerala), while in others it makes it worse (Rajasthan, Delhi, Bhubaneswar, Meghalaya). By contrast, EMOS either considerably improves the reliability at the locations where the raw forecasts overestimate the event probability, especially for relatively high forecast probabilities (Mumbai, Kerala, Shimla, Delhi, Hyderabad, Patna), or does not reduce the already good reliability of the raw forecasts (Rajasthan, Bhubaneswar, Meghalaya). Overall, the reliability after applying EMOS is higher than that of both the raw forecasts and QM postprocessing.

The ROC curves (Figure 9) for the exceedance of the local 90th percentile of observed precipitation show small to moderate differences between the raw and postprocessed forecasts for the majority of locations (Mumbai, Rajasthan, Kerala, Shimla, Patna, Bhubaneswar, Meghalaya), with only Delhi and Hyderabad showing larger differences in the area under the curves. Raw and postprocessed forecasts have the largest area under the ROC curves (i.e., highest hit rates and lowest false alarm rates for all probability thresholds) at Mumbai and to a lesser extent Kerala, which are both characterized by relatively high average precipitation (Figure 3). At these locations the raw and postprocessed forecasts perform similarly. The difference between the curves for the raw and postprocessed forecasts increases at locations such as Delhi and Hyderabad, which are characterized by drier conditions, and in Delhi QM outperforms both EMOS and the raw forecast.

The RSS values for the exceedance of the local 90th percentile of observed wet-day precipitation for QM (Figure 10a) are positive in the southern, western, and northern regions, indicating improved event discrimination compared to the raw forecast, while in central and eastern regions the RSS is close to zero or negative. The RSS values for EMOS (Figure 10b) are mainly positive in the western region and mostly negative in central, northern, and eastern regions. The method with the highest RSS at
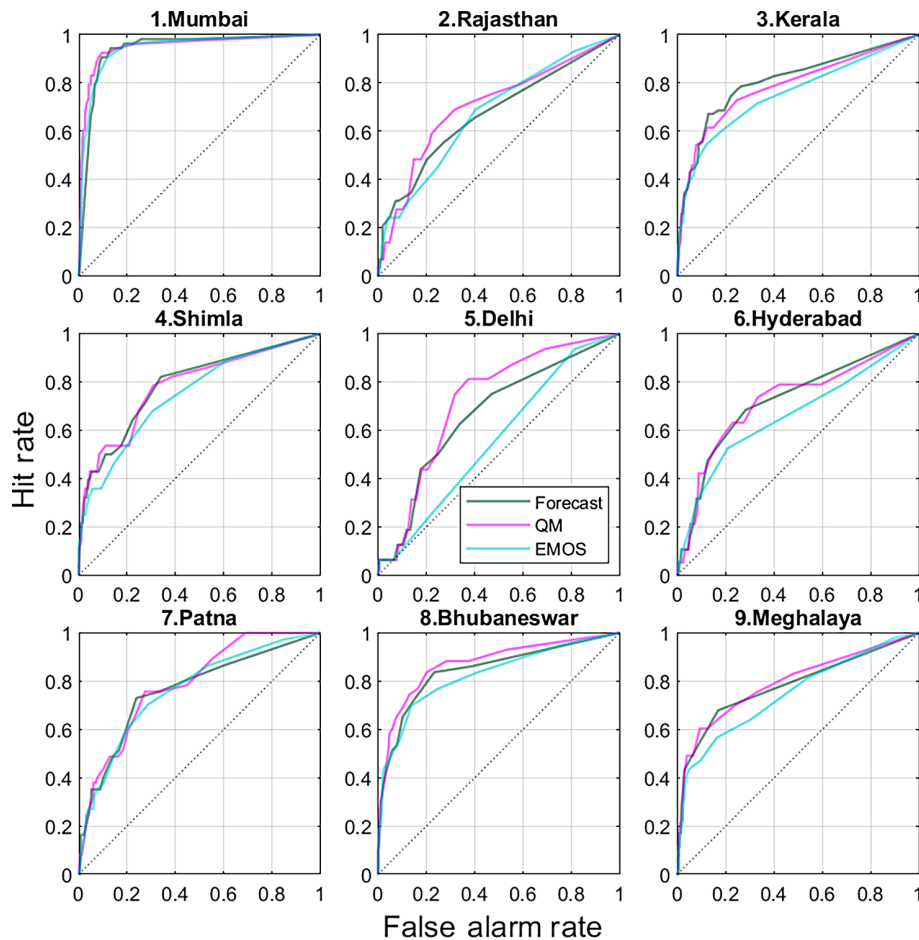
FIGURE 9 Receiver Operating Characteristic (ROC) curves for the exceedance of the 90th percentile of daily precipitation for observed wet days (precipitation >0 mm·day^{-1}) during the 2018–2022 monsoon seasons (June–September) for raw, Quantile Mapping (QM)-postprocessed, and Ensemble Model Output Statistics (EMOS)-postprocessed forecasts at the nine test locations.

a given location is shown in Figure 10c. For most of India, QM is better than EMOS (although EMOS performs best in the wet regions of the Western Ghats), with scattered locations where the raw forecasts perform best.

5 | DISCUSSION AND CONCLUSIONS

In our pilot study we have evaluated QM and EMOS post-processing of ensemble precipitation forecasts with a lead time of one day over India during five summer monsoon periods, with respect to all precipitation amounts as well as with respect to heavy precipitation events. The comparison of the raw forecasts with observations revealed substantial errors for all precipitation values and for heavy precipitation, despite the short lead time (Figures 2, 3, 5, and 8). This is likely due to limitations in the representation of multiscale interactions in the 12-km resolution NEPS-G model, ranging from synoptic-scale to local-scale processes, with the latter highly dependent on model resolution. Important small-scale processes for precipitation, and especially heavy precipitation, include the representation of convective events (Konduru & Takahashi, 2020;

Sillmann *et al.*, 2017; Willetts *et al.*, 2017), whose locations are quasi-random, and the effects of complex orography (Baisya & Pattnaik, 2019; Rotunno & Houze, 2007; Webster *et al.*, 2008).

Overall, we find that EMOS is the best method for correcting the ensemble spread (Figure 5), and also has the best skill for forecasting the values on a given day with respect to all values, quantified by the CRPS (Figure 6), and with respect to heavy precipitation, quantified by BS (Figure 7). It also leads to the best forecast reliability (Figure 8). By contrast, QM is the best method for correcting the forecast precipitation PDF in most locations (Figures 2 and 3). With respect to the ability of the ensemble to discriminate whether heavy precipitation thresholds are exceeded, the ROC curves (Figure 9) and RSS (Figure 10) show at most locations a slight improvement by QM relative to the raw forecast and a slight reduction by EMOS.

The reason EMOS outperforms QM at most locations with respect to ensemble spread, CRPS, BS and reliability is that although QM leads by construction to a realistic overall distribution of daily precipitation values, it changes the ensemble spread in a way that at many locations reduces the forecast skill. In contrast, EMOS

FIGURE 10 Receiver Operating Characteristic (ROC) skill score (RSS) for the exceedance of the 90th percentile of daily precipitation for observed wet days (precipitation >0 mm·day^{-1}) during the 2018–2022 monsoon seasons (June–September) for (a) Quantile Mapping (QM)-postprocessing, and (b) Ensemble Model Output Statistics (EMOS) postprocessing. Positive (negative) RSS indicates better (worse) performance than the raw forecast. (c) Method that yields the best Receiver Operating Characteristic Score at a given location. (d) 90th percentile of daily precipitation for wet days for the 1980–2022 monsoon seasons. NEPS-G, National Centre for Medium Range Weather Forecasting (NCMRWF) global ensemble prediction system.
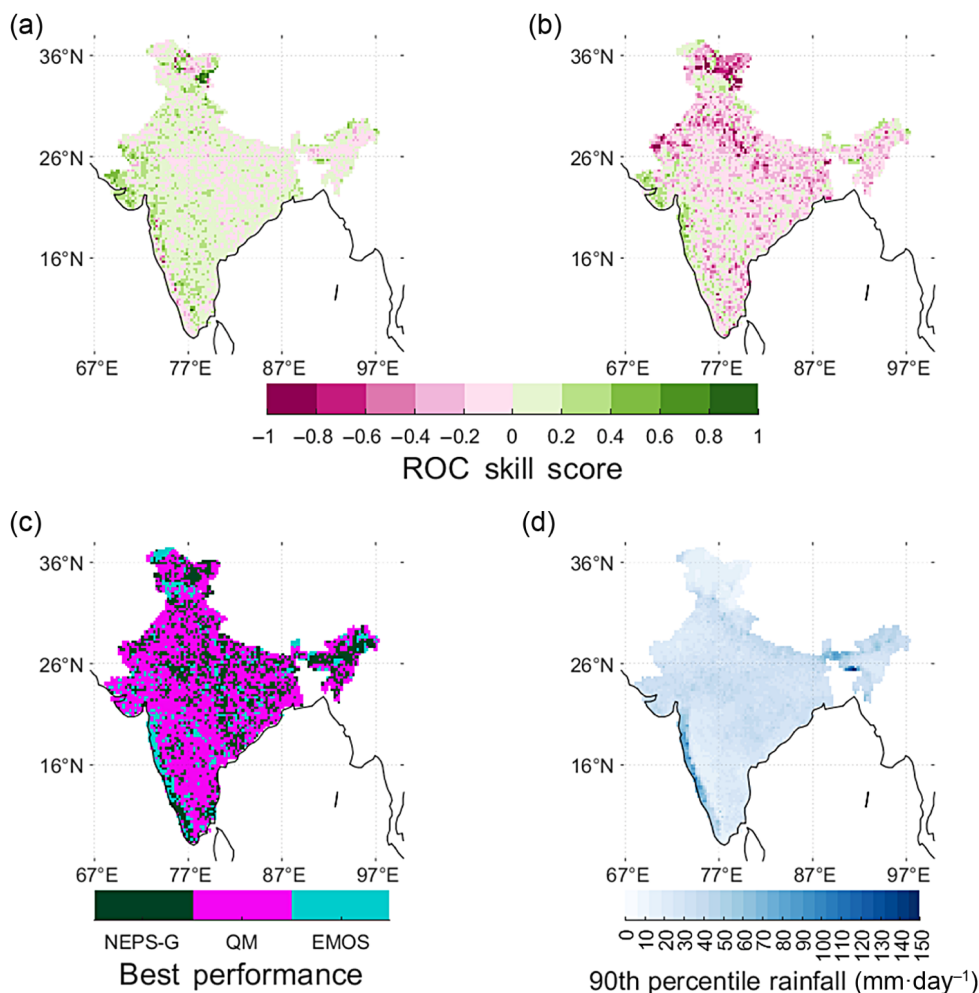
optimizes the postprocessed ensembles to fit the observations with respect to CRPS, and our results show that this also leads to good skill for forecasting heavy precipitation events (Figure 9). Which aspects of the forecasts are considered most relevant depends to some extent on how the forecasts are used. If the main criteria are the correction of under- or overdispersion, forecast reliability, match between the forecasted distribution for individual days and observations (CRPS), and the skill in forecasting heavy precipitation events (BS), EMOS is the better choice for postprocessing NEPS-G precipitation forecasts over India with one day lead time.

Both QM and EMOS allow for lead-time-dependent corrections. For QM the dependency captures potential differences in the systematic biases of the simulated marginal distribution for different lead times, and for EMOS it captures potential differences in the systematic biases of the ensemble forecasts for different lead times. While QM is only based on the simulated and observed marginal distributions, EMOS exploits the pairwise correspondence of observations and ensemble forecasts, which gives it a potential advantage and which is

likely to be a main reason for the better performance of EMOS in the day-1 forecasts in this study. However, predictability decreases with increased lead time, and EMOS may have less of an advantage. In the case of completely non-informative forecasts both QM and EMOS yield the climatological marginal distribution. A systematic comparison of QM and EMOS for longer lead times is needed to fully explore this topic.

The implementation of EMOS in an operational forecasting context is straightforward because after the fitting of the statistical postprocessing models their application for each forecast requires only minor additional computations. The postprocessing can be easily updated when new versions of forecast models or improved observations become available. As demonstrated in this study, five monsoon seasons simulated as hindcast with a new model version would allow to fit and evaluate the updated EMOS postprocessing.

As mentioned in Section 2.2. *Setups* the EMOS method presented in this study can only correct the precipitation intensity on wet days but not the precipitation probability because the fitting is based on observations and ensemble

forecasts on observed wet days. For operational implementation it needs to be decided how to determine the precipitation probability and how to avoid potential biases. If there is no systematic difference between the observed wet-day precipitation distributions during the fitting and forecasting periods there is no systematic bias of the postprocessed forecast distribution for observed wet days. If the postprocessed forecasts for the observed dry days were set to zero the distribution for all days would also not have a systematic bias. If, however, the EMOS model fitted on observed wet days was also applied to forecasts for observed dry days the output for observed dry days would include non-zero precipitation, and the overall distribution would therefore have a positive bias. In an operational context it is not possible to set the output of the postprocessing to zero when the observation is zero, because the observation for the forecast time is not known. It would be possible to mitigate against a positive bias by setting the postprocessed output for days with a substantial fraction of dry ensemble members, and thus with a high probability of an observed dry day, to zero. However, this approach lacks a sound theoretical foundation and it would be preferable to use the observed wet and dry days to fit an additional postprocessing model for precipitation probabilities, set the number of ensemble members to zero proportional to the dry-day probability, and postprocess the remaining ensemble members as described in this study.

Given the need for integrating ensemble precipitation forecasts for India with hydrological and hydrodynamic modelling (Chandra & Mujumdar, 2019; Ghosh *et al.*, 2019; Mohanty *et al.*, 2020; Nanditha & Mishra, 2021; Rupa & Mujumdar, 2019; Widmann, Blake, *et al.*, 2019), it would be important to assess to which extent the precipitation postprocessing can improve hydrological forecasting (cf. Pastén-Zapata *et al.* (2020) for a UK example). In this context the spatial structure of heavy precipitation is highly relevant, for instance the probability for joint exceedance of heavy precipitation thresholds. QM and EMOS are local postprocessing methods, which do not explicitly change the spatial structure of precipitation events. Widmann, Bedia, *et al.* (2019) have shown that the spatial structure of QM-postprocessed precipitation is mainly inherited from the raw forecast, and it can be expected that this is also the case for EMOS.

While this pilot study is limited by the relatively few monsoon seasons currently available from the NEPS-G forecast and by focussing on a lead time of one day, we demonstrated that EMOS can improve ensemble forecasts for precipitation amounts and for probabilities for exceeding the local 90th percentile over most of India with respect to multiple skill measures. Applying EMOS operationally for short lead times of a few days would thus provide an opportunity to improve precipitation forecasts at low computational costs, while the potential benefit for longer lead times needs to be explored further.

DATA AVAILABILITY STATEMENT
The 23-member ensemble precipitation forecasts for day-1 from the NEPS-G system can be downloaded from the ECMWF data portal TIGGE archive: https://apps.ecmwf .int/datasets/data/tigge/levtype=sfc/type=cf/. The IMD V6.9 data can be downloaded from: https://www.imdpune .gov.in/cmpg/Griddata/Rainfall&uscore;25&uscore;Bin .html.

ORCID
Martin Widmann https://orcid.org/0000-0001-5447 -5763
Andrew Orr https://orcid.org/0000-0001-5111-8402
Raghavendra Ashrit https://orcid.org/0000-0002-1983 -8378
Gregor C. Leckebusch https://orcid.org/0000-0001-9242 -7682
Ashis Mitra https://orcid.org/0000-0002-7882-5868

REFERENCES
Ali, H., Modi, P. & Mishra, V. (2019) Increased flood risk in Indian sub-continent under the warming climate. *Weather and Climate Extremes*, 25, 100212. Available from: https://doi.org/10.1016/j .wace.2019.100212

Ashrit, R., Sharma, K., Kumar, S., Dube, A., Karunasagar, S., Arulalan, T. et al. (2020) Prediction of the August 2018 heavy rainfall events over Kerala with high-resolution NWP models. *Meteorological Applications*, 27, e1906. Available from: https://doi .org/10.1002/met.1906

Baisya, H. & Pattnaik, S. (2019) Orographic effect and multiscale interactions during an extreme rainfall event. *Environmental Research Communications*, 1, 051002. Available from: https://doi .org/10.1088/2515-7620/ab2417

Baran, S. & Nemoda, D. (2016) Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting. *Environmetrics*, 27, 280–292. Available from: https://doi.org/10.1002/env.2391

Basher, R. (2006) Global early warning systems for natural hazards: systematic and people-centred. *Philosophical Transactions of the Royal Society A*, 364, 2167–2182. Available from: https://doi.org /10.1098/RSTA.2006.1819

Befort, D.J., Leckebusch, G.C. & Cubasch, U. (2016) Intra-seasonal variability of the Indian summer monsoon: wet and dry events in COSMO-CLM. *Climate Dynamics*, 47, 2635–2651. Available from: https://doi.org/10.1007/s00382-016-2989-7

Brier, G.W. (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1–3. Available from: https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Chakraborty, P., Sarkar, A., Bhatla, R. & Singh, R. (2021) Assessing the skill of NCMRWF global ensemble prediction system in predicting Indian summer monsoon during 2018. *Atmospheric Research*, 248, 105255. Available from: https://doi.org/10.1016/j.atmosres.2020.105255

Chandra, R.R. & Mujumdar, P.P. (2019) Hydrologic impacts of climate change: quantification of uncertainties. *Proceedings of the Indian National Science Academy*, 85, 77–94. Available from: https://doi.org/10.16943/ptinsa/2018/49506

Dirmeyer, P.A., Cash, B.A., Kinter, J.L., Jung, T., Marx, L., Satoh, M. et al. (2012) Simulating the diurnal cycle of rainfall in global climate models: resolution versus parameterization. *Climate Dynamics*, 39, 399–418. Available from: https://doi.org/10.1007/s00382-011-1127-9

Eden, J.M., Widmann, M., Grawe, D. & Rast, S. (2012) Skill, correction, and downscaling of GCM-simulated precipitation. *Journal of Climate*, 25, 3970–3984. Available from: https://doi.org/10.1175/JCLI-D-11-00254.1

Falga, R. & Wang, C. (2022) The rise of Indian summer monsoon precipitation extremes and its correlation with long-term changes of climate and anthropogenic factors. *Scientific Reports*, 12, 11985. Available from: https://doi.org/10.1038/s41598-022-16240-0

Francis, P. & Gadgil, S. (2006) Intense rainfall events over the west coast of India. *Meteorology and Atmospheric Physics*, 94, 27–42. Available from: https://doi.org/10.1007/s00703-005-0167-2

Ghosh, S., Das, D., Kao, S.-C. & Ganguly, A.R. (2012) Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nature Climate Change*, 2, 86–91. Available from: https://doi.org/10.1038/nclimate132.7

Ghosh, S., Karmakar, S., Saha, A., Mohanty, M.P., Ali, S., Raju, S.K. et al. (2019) Development of India's first integrated expert urban flood forecasting system for Chennai. *Current Science*, 117, 741–745.

Gneiting, T., Raftery, A.E., Westveld, A.H., III & Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133, 1098–1118. Available from: https://doi.org/10.1175/MWR2904.1

Goswami, B.N., Venugopal, V., Sengupta, D., Madhusoodanan, M.S. & Xavier, P.K. (2006) Increasing trend of extreme rain events over India in a warming environment. *Science*, 314, 1442–1445. Available from: https://doi.org/10.1126/science.1132027

Gupta, A. & Takahashi, H.G. (2022) Long-term changes in spatially coherent extreme precipitation systems over Central India. *Atmospheric Science Letters*, 23, e1118. Available from: https://doi.org/10.1002/asl.1118

Gupta, A.K. & Nair, S.S. (2011) Urban floods in Bangalore and Chennai: risk management challenges and lessons for sustainable urban ecology. *Current Science*, 100, 1638–1645.

Hamill, T.M. (1997) Reliability diagrams for multicategory probabilistic forecasts. *Weather and Forecasting*, 12, 736–741. Available from: https://doi.org/10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2

Hamill, T.M. (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129, 550–560. Available from: https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2

Hay, L.E. & Clark, M.P. (2003) Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *Journal of Hydrology*, 282, 56–75. Available from: https://doi.org/10.1016/S0022-1694(03)00252-X

Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15, 559–570. Available from: https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Javanshiri, Z., Fathi, M. & Mohammadi, S.A. (2021) Comparison of the BMA and EMOS statistical methods for probabilistic quantitative precipitation forecasting. *Meteorological Applications*, 28, e1974. Available from: https://doi.org/10.1002/met.1974

Jha, S.K., Shrestha, D.L., Stadnyk, T.A. & Coulibaly, P. (2018) Evaluation of ensemble precipitation forecasts generated through post-processing in a Canadian catchment. *Hydrology and Earth System Sciences*, 22, 1957–1969. Available from: https://doi.org/10.5194/hess-22-1957-2018

Konduru, R.T. & Takahashi, H.G. (2020) Effects of convection representation and model resolution on diurnal precipitation cycle over the Indian monsoon region: toward a convection-permitting regional climate simulation. *Journal of Geophysical Research*, 125, e2019JD032150. Available from: https://doi.org/10.1029/2019JD032150

Mahanta, R. & Das, D. (2017) Flood induced vulnerability to poverty: evidence from Brahmaputra Valley, Assam, India. *International Journal of Disaster Risk Reduction*, 24, 451–461. Available from: https://doi.org/10.1016/j.ijdrr.2017.04.014

Mahapatra, B., Walia, M. & Saggurti, N. (2018) Extreme weather events induced deaths in India 2001–2014: trends and differentials by region, sex and age group. *Weather and Climate Extremes*, 21, 110–116. Available from: https://doi.org/10.1016/j.wace.2018.08.001

Mamgain, A., Sarkar, A. & Rajagopal, E.N. (2020a) Medium-range global ensemble prediction system at 12 km horizontal resolution and its preliminary validation. *Meteorological Applications*, 27, e1867. Available from: https://doi.org/10.1002/met.1867

Mamgain, A., Sarkar, A. & Rajagopal, E.N. (2020b) Verification of high resolution (12 km) global ensemble prediction system. *Atmospheric Research*, 236, 104832. Available from: https://doi.org/10.1016/j.atmosres.2019.104832

Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M. et al. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7, 764–773. Available from: https://doi.org/10.1038/nclimate3418

Maraun, D. & Widmann, M. (2018) *Statistical downscaling and bias correction for climate research*. Cambridge: Cambridge University Press. Available from: https://doi.org/10.1017/9781107588783

Mason, S.J. & Graham, N.E. (1999) Conditional probabilities, relative operating characteristics, and relative operating levels. *Weather and Forecasting*, 14, 713–725. Available from:

https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2

Mitra, A.K., Iyengar, G.R., Durai, V.R., Sanjay, J., Krishnamurti, T.N., Mishra, A. et al. (2011) Experimental real-time multi-model ensemble (MME) prediction of rainfall during monsoon 2008: large-scale medium-range aspects. *Journal of Earth System Science*, 120, 27–52. Available from: https://doi.org/10.1007/s12040-011-0013-5

Mohandas, S., Francis, T., Singh, V., Jayakumar, A., George, J.P., Sandeep, A. et al. (2020) NWP perspective of the extreme precipitation and flood event in Kerala (India) during august 2018. *Dynamics of Atmospheres and Oceans*, 91, 101158. Available from: https://doi.org/10.1016/j.dynatmoce.2020.101158

Mohanty, M.P., Mudgil, S. & Karmakar, S. (2020) Flood management in India: a focussed review on the current status and future challenges. *International Journal of Disaster Risk Reduction*, 49, 101660. Available from: https://doi.org/10.1016/j.ijdrr.2020.101660

Mukherjee, S., Aadhar, S., Stone, D. & Mishra, V. (2018) Increase in extreme precipitation events under anthropogenic warming in India. *Weather and Climate Extremes*, 20, 45–53. Available from: https://doi.org/10.1016/j.wace.2018.03.005

Mukhopadhyay, P., Bechtold, P., Zhu, Y., Murali Krishna, R.P., Kumar, S., Ganai, M. et al. (2021) Unraveling the mechanism of extreme (more than 30 sigma) precipitation during august 2018 and 2019 over Kerala, India. *Weather and Forecasting*, 36, 1253–1273. Available from: https://doi.org/10.1175/WAF-D-20-0162.1

Nanditha, J.S. & Mishra, V. (2021) On the need of ensemble flood forecast in India. *Water Security*, 12, 100086. Available from: https://doi.org/10.1016/j.wasec.2021.100086

Pai, D.S., Rajeevan, M., Sreejith, O.P., Mukhopadhyay, B. & Satbha, N.S. (2014) Development of a new high spatial resolution (0.25° × 0.25°) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam*, 65, 1–18. Available from: https://doi.org/10.54302/mausam.v65i1.851

Pastén-Zapata, E., Jones, J.M., Moggridge, H. & Widmann, M. (2020) Evaluation of the performance of euro-CORDEX regional climate models for assessing hydrological climate change impacts in Great Britain: a comparison of different spatial resolutions and quantile mapping bias correction methods. *Journal of Hydrology*, 584, 124653. Available from: https://doi.org/10.1016/j.jhydrol.2020.124653

Pattanaik, D.R. & Rajeevan, M. (2010) Variability of extreme rainfall events over India during southwest monsoon season. *Meteorological Applications*, 17, 88–104. Available from: https://doi.org/10.1002/met.164

Piani, C., Haerter, J.O. & Coppola, E. (2010) Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99, 187–192. Available from: https://doi.org/10.1007/s00704-009-0134-9

Ray, K., Giri, R.K., Ray, S.S., Dimri, A.P. & Rajeevan, M. (2021) An assessment of long-term changes in mortalities due to extreme weather events in India: a study of 50 years' data, 1970-2019. *Weather and Climate Extremes*, 32, 100315. Available from: https://doi.org/10.1016/j.wace.2021.100315

Revadekar, J.V. & Preethi, B. (2012) Statistical analysis of the relationship between summer monsoon precipitation extremes and food grain yield over India. *International Journal of Climatology*, 32, 419–429. Available from: https://doi.org/10.1002/joc.2282

Rotunno, R. & Houze, R.A. (2007) Lessons on orographic precipitation from the mesoscale alpine Programme. *Quarterly Journal of the Royal Meteorological Society*, 133, 811–830. Available from: https://doi.org/10.1002/qj.67

Roxy, M.K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R. et al. (2017) A threefold rise in widespread extreme rain events over central India. *Nature Communications*, 8, 708. Available from: https://doi.org/10.1038/s41467-017-00744-9

Rupa, R.C. & Mujumdar, P.P. (2019) Flood modelling: recent Indian contributions. *Proceedings of the Indian National Science Academy*, 85, 705–722. Available from: https://doi.org/10.16943/ptinsa/2019/49648

Sarkar, A., Kumar, S., Dube, A., Prasad, S.K., Mamgain, A., Chakraborty, P. et al. (2021) Forecasting of tropical cyclone using global and regional ensemble prediction systems of NCMRWF: a review. *Mausam*, 72, 77–86. Available from: https://doi.org/10.54302/mausam.v72i1.131

Schefzik, R. (2017) Ensemble calibration with preserved correlations: unifying and comparing ensemble copula coupling and member-by-member postprocessing. *Quarterly Journal of the Royal Meteorological Society*, 143, 999–1008. Available from: https://doi.org/10.1002/qj.2984

Scheuerer, M. (2014) Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society*, 140, 1086–1096. Available from: https://doi.org/10.1002/qj.2183

Scheuerer, M. & Hamill, T.M. (2015) Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Monthly Weather Review*, 143, 4578–4596. Available from: https://doi.org/10.1175/MWR-D-15-0061.1

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L.V., Hegerl, G. et al. (2017) Understanding, modeling and predicting weather and climate extremes: challenges and opportunities. *Weather and Climate Extremes*, 18, 65–74. Available from: https://doi.org/10.1016/j.wace.2017.10.003

Singh, D., Ghosh, S., Roxy, M.K. & McDermid, S. (2019) Indian summer monsoon: extreme events, historical changes, and role of anthropogenic forcings. *WIREs Climate Change*, 10, e571. Available from: https://doi.org/10.1002/wcc.571

Sooraj, K.P., Terray, P. & Xavier, P. (2016) Sub-seasonal behaviour of Asian summer monsoon under a changing climate: assessments using CMIP5 models. *Climate Dynamics*, 46, 4003–4025. Available from: https://doi.org/10.1007/s00382-015-2817-5

Sreenath, A.V., Abhilash, S., Vijaykumar, P. & Mapes, B.E. (2022) West coast India's rainfall is becoming more convective. *NPJ Climate and Atmospheric Science*, 5, 36–37. Available from: https://doi.org/10.1038/s41612-022-00258-2

Srinivas, C.V., Yesubabu, V., Hari Prasad, D., Hari Prasad, K.B.R.R., Greeshma, M.M., Baskaran, R. et al. (2018) Simulation of an extreme heavy rainfall event over Chennai, India using WRF: sensitivity to grid resolution and boundary layer physics. *Atmospheric Research*, 210, 66–82. Available from: https://doi.org/10.1016/j.atmosres.2018.04.014

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

Uccellini, L.W. & Ten Hoeve, J.E. (2019) Evolving the national weather service to build a weather-ready nation: connecting observations, forecasts, and warnings to decision-makers through impact-based decision support services. *Bulletin of the American Meteorological Society*, 100, 1923–1942. Available from: https://doi.org/10.1175/BAMS-D-18-0159.1

UNDRR. (2020) Human cost of disasters: an overview of the last 20 years 2000-2019. In: *United Nations Office for disaster risk reduction (UNDRR)*. Geneva.

van Oldenborgh, G.J., Otto, F.E.L., Haustein, K. & Achuta Rao, K. (2016) The heavy precipitation event of December 2015 in Chennai, India. *Bulletin of the American Meteorological Society*, 97, S87–S91. Available from: https://doi.org/10.1175/BAMS-D-16-0129.1

Vannitsem, S., Bremnes, J.B., Demaeyer, J., Evans, G.R., Flowerdew, J., Hemri, S. et al. (2021) Statistical postprocessing for weather forecasts: review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*, 102, E681–E699. Available from: https://doi.org/10.1175/BAMS-D-19-0308.1

Varikoden, H. & Reji, M.J.K. (2022) Features of regional Indian monsoon rainfall extremes. In: *Climate impacts on extreme weather*. Amsterdam: Elsevier, pp. 83–100.

Viswanadhapalli, Y., Srinivas, C.V., Basha, G., Dasari, H.P., Langodan, S., Ratnam, M.V. et al. (2019) A diagnostic study of extreme precipitation over Kerala during August 2018. *Atmospheric Science Letters*, 20, e941. Available from: https://doi.org/10.1002/asl.941

Volosciuk, C., Maraun, D., Vrac, M. & Widmann, M. (2017) A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrology and Earth System Sciences*, 21, 1693–1719. Available from: https://doi.org/10.5194/hess-21-1693-2017

Wallemacq, P., Herden, C. & House, R. (2015) The human cost of natural disasters 2015: a global perspective. Technical Report, Centre for Research on the Epidemiology of Disasters.

Webster, S., Uddstrom, M., Oliver, H. & Vosper, S. (2008) A high-resolution modelling case study of a severe weather event over New Zealand. *Atmospheric Science Letters*, 9, 119–128. Available from: https://doi.org/10.1002/asl.172

Widmann, M., Bedia, J., Gutiérrez, J.M., Bosshard, T., Hertig, E., Maraun, D. et al. (2019) Validation of spatial variability in downscaling results from the VALUE perfect predictor experiment. *International Journal of Climatology*, 39, 3819–3845. Available from: https://doi.org/10.1002/joc.6024

Widmann, M., Blake, R., Sooraj, K.P., Orr, A., Sanjay, J., Karumuri, A. et al. (2019) Current opportunities and challenges in developing hydro-climatic Services in the Himalayas. Report of India-UK Water Center Pump Priming Project 2017, Center for Ecology & Hydrology, Wallingford and Indian Institute of Tropical Meteorology, Pune.

Wilks, D.S. (2006) Comparison of ensemble MOS methods in the Lorenz 96 setting. *Meteorological Applications*, 13, 243–256. Available from: https://doi.org/10.1017/S1350482706002192

Wilks, D.S. (2011) *Statistical methods in the atmospheric sciences*, 3rd edition. New York: Academic Press.

Wilks, D.S. & Hamill, T.M. (2007) Comparison of ensemble-MOS methods using GFS reforecasts. *Monthly Weather Review*, 135, 2379–2390. Available from: https://doi.org/10.1175/MWR3402.1

Willetts, P.D., Marsham, J.H., Birch, C.E., Parker, D.J., Webster, S. & Petch, J. (2017) Moist convection and its upscale effects in simulations of the Indian monsoon with explicit and parametrized convection. *Quarterly Journal of the Royal Meteorological Society*, 143, 1073–1085. Available from: https://doi.org/10.1002/qj.2991

Yuen, R.A., Gneiting, T., Thorarinsdottir, T.L. & Fraley, C. (2013) Ensemble model output statistics to create probabilistic forecasts from ensemble forecasts and weather observations. R package version 0.7′. http://CRAN.R-project.org/package=ensembleMOS (accessed: October 2020).

How to cite this article: Angus, M., Widmann, M., Orr, A., Ashrit, R., Leckebusch, G.C. & Mitra, A. (2024) A comparison of two statistical postprocessing methods for heavy-precipitation forecasts over India during the summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, 150(761), 1865–1883. Available from: https://doi.org/10.1002/qj.4677