Precipitation forecast over China for different thresholds using the multimodel bias-removed ensemble mean

Yang Lv¹,²,³, Xiefei Zhi¹,² and Shoupeng Zhu¹,²

¹Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science & Technology, Nanjing, China;
²Weather Online Institute of Meteorological Applications, Wuxi, China
³Email: 1325819177@qq.com

Abstract. Based on the daily accumulated precipitation data obtained from the ensemble forecasts by three meteorological agencies and the CMORPH observational data, the experiments of bias-removed ensemble mean (BREM) towards classified samples of different precipitation thresholds are carried out with results as follows: (1) The Classified BREM (CBREM) is characterized by higher skill in precipitation forecast in contrast to BREM. Most visible improvements can be observed for light precipitation, but there is a negative impact for a moderate threshold. (2) The CBREM after choosing optimal grading thresholds for each grid point further improves forecast skill of precipitation, which shows greatest advancement for moderate precipitation with the threat score improving percentage of over 20% on average.

1. Introduction
Precipitation is considered as one of the most important factors impacting agriculture and human society [1-3]. With the climate background of conspicuous changes, increasing demand of precipitation forecast has been addressed [4-6]. In recent decades, numerical models have been developed to improve simulations and forecasts of various atmospheric systems [7-8]. To make the best use of outputs from numerical weather prediction from multiple meteorological agencies, statistical post-processing methods such as multimodel ensemble have been proposed, which improve forecast skills of temperature, precipitation, wind and many other factors [9-13]. Moreover, considering the discontinuity and locality of precipitation, the forecast skill of precipitation can also be improved by dividing the precipitation samples to different thresholds and afterwards carrying out the calibration separately.

In this study, the two typical multimodel ensembles of simple ensemble mean (EMN) and bias-removed ensemble mean (BREM) [14] are employed for precipitation forecast. Additionally, the BREM is also performed towards samples of different precipitation thresholds. On the other hand, the influences of different reference forecasts on splitting samples and grading thresholds are further studied.

The structure of this paper is organized as follows. The used data and methods are briefly introduced in Section 2. In Section 3, we assess forecasts of the raw models, EMN and BREM. The BREM for different precipitation thresholds and different threshold grading criteria are also described in this section. Finally, the conclusion and discussion are presented in Section 4.
2. Data and methods

2.1. Data
The used forecast dataset of daily accumulated precipitation are obtained from the ensemble forecasts initialized at 0000 UTC produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), and the United Kingdom Meteorological Office (UKMO) under the THORPEX Interactive Grand Global Ensemble (TIGGE) framework with a horizontal resolution of $1^\circ \times 1^\circ$. The lead times range from 1 to 7 days. In addition, the daily accumulated precipitation data provided by CPC MORPHing technique (CMORPH) are used as observation. It is produced with a horizontal resolution of $0.1^\circ \times 0.1^\circ$ and is widely used in precipitation-related researches. In this study, the used resolution is $1^\circ \times 1^\circ$ and the common period is 1 January-31 August 2016. To be noted, the two month of July and August are used for forecast verifications after the post-processing.

2.2. The multimodel ensemble methods
Multimodel ensemble forecast are always characterized by higher skills compared to individual models, in which, the EMN is the most common method:

$$\text{EMN} = \sum_{i=1}^{N} F_i$$

where $\text{EMN}$ is the output of EMN, $N$ is the number of ensemble members, and $F_i$ is the forecast of model member $i$.

Besides, assuming that the model systematic errors can be removed through a training phase, the BREM has been proposed as the following formula:

$$\text{BREM} = \bar{O} + \frac{1}{N} \sum_{i=1}^{N} (F_i - \bar{F}_i)$$

where $\text{BREM}$ is the output of BREM, $\bar{O}$ is the mean of observation in the training phase, $F_i$ is the forecast of the model member $i$, $\bar{F}_i$ is the forecast mean of the model $i$ in the training phase. Importantly, the running training phase is applied in our study to adjust the latest forecast performance of the models.

Considering that the precipitation with different thresholds varies in spatial and temporal characteristics, we classify the experiment samples of the 24 h accumulated precipitation into four thresholds according to the criteria by the China Meteorological Administration (CMA), i.e. light precipitation ($< 10$ mm), moderate precipitation (10-25 mm), heavy precipitation (25-50 mm), and rainstorm ($\geq 50$ mm). Afterwards, they are calibrated with the BREM method separately to obtain an overall better forecast.

2.3. The verification methods
The multimodel ensemble results are verified via methods of the root-mean-square error (RMSE) and the threat score (TS), which are widely used in precipitation related studies [14].

RMSE describes the magnitudes that the forecast data deviates from the observation. The smaller RMSE indicates the higher forecast skill.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F_i - O_i)^2}$$

where $n$ is the total number of grid points, $F_i$ is the forecast of the grid $i$, and the $O_i$ is the observation of the grid $i$.

The TS reflects the forecast skill of precipitation over different thresholds. The higher TS indicates the higher forecast skill.

$$\text{TS} = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{alarms}}$$

where $\text{hits}$ is the number of grid points where both forecast and observation reach a certain precipitation threshold, $\text{misses}$ is the number of grid points where observation reaches a threshold but forecast does not. Similarly, $\text{alarms}$ is the number of grid points where forecast reaches a threshold but observation does not.
3. Results

3.1. Improvement of precipitation forecast by BREM

The RMSEs of daily accumulated precipitation from different models and multimodel ensembles are displayed with lead times from 1 to 7 days in Figure 1, which are averaged over the period of 1 July-31 August 2016. In general, the RMSE increases over the lead time. The ECMWF forecast shows lowest RMSEs among the three raw models, showing its superiority to the other two centers. Besides, the EMN and BREM are characterized by lower RMSEs, indicating the significant advantages of multimodel ensembles, with the latter performing better.

![Figure 1](image1.png)

**Figure 1.** The RMSEs of daily accumulated precipitation from different models and multimodel ensembles for lead times of 1-7 days.

![Figure 2](image2.png)

**Figure 2.** The frequency of daily accumulated precipitation averaged over the period from 1 July to 31 August 2016 over China for different thresholds of (a) < 10 mm; (b) 10-25 mm; (c) 25-50 mm and (d) > 50 mm.

3.2. Frequency distribution of precipitation over different thresholds

There are huge differences in precipitation across China, and the probability of precipitation for different thresholds are also different. Figure 2 shows the spatial distributions of precipitation frequency for different thresholds over China. It is obvious that the precipitation in China is dominated...
by light rain. The frequencies are mainly greater than 0.6. The moderate precipitation is also common in some areas, with the frequency ranging from 0.3 to 0.5. In contrast, the frequencies of high precipitation and rainstorm are relatively lower, especially for the rainstorm. Therefore, the samples of high precipitation and rainstorm could be too few for the forecast post-processing. Correspondingly, a spatial window of $3° \times 3°$ is determined to increase the samples of high precipitation and rainstorms.

3.3. BREM based on different precipitation thresholds
Precipitation with different thresholds are characterized by different spatial and temporal features. Therefore, the Classified BREM (CBREM) is used to take precipitation samples of different thresholds into consideration. With respect to the CBREM method, it would have great influences on the obtained results using different criteria of classification based on different forecasts, i.e. the raw forecasts or the ensembles. According to the results obtained from Section 3.1, the ECWMF, EMN and BREM are superior to the others in the precipitation forecasts, which are used as classification references in the following CBREM experiments. The precipitation forecasts towards different thresholds determined by ECWMF, EMN and BREM are hereafter abbreviated as c-ECMWF, c-EMN and c-BREM, respectively. Importantly, during the forecast of rainstorm, considering the small number of samples and their possible serious effects on the human society, overestimations are always considered more acceptable than underestimations. Therefore, if a rainstorm occurs in the classification reference forecast, due to the possible smoothing effect of the BREM, it is directly taken as the final forecast without any post-processing.

Figure 3 shows the TS by CBREM under different classification reference for different precipitation thresholds of (a) < 10 mm; (b) 10-25 mm; (c) 25-50 mm and (d) > 50 mm. The c-ECMWF, c-EMN and c-BREM represent forecasts with threshold classification criteria based on ECMWF, EMN and BREM, respectively.

Figure 3 shows the TS by CBREM under different classification reference for different precipitation thresholds. In general, the TSs of the forecast decrease with the increasing lead times. For light precipitation, the TSs by CBREM are always higher than those by BREM. The c-EMN shows the most skillful performance among the results. The improvement magnitudes decrease with the
increasing lead times from ~0.4 to ~0.2. For moderate precipitation, the CBREM forecasts are inferior to the initial BREM, showing a negative impact of CBREM in forecasting precipitation with a moderate threshold. For heavy precipitation, the CBREM forecasts generally perform better than the BREM, especially for lead times of 4-7 days, although the improvement magnitudes are relatively limited. Additionally, the CBREM forecasts still indicate higher skills in forecasting rainstorms, with the c-ECMWF showing best performance.

3.4. The optimal grading thresholds

On the basis of the considerable results of CBREM, different criteria of precipitation thresholds could also affect the forecast results, which are further investigated in the following section.

In this study, 66 schemes of threshold selection are employed. That is, for the light precipitation threshold of < 10 mm, we determine 6 new upper levels of 5-10 mm by 1 mm. For the moderate precipitation, the thresholds are adjusted to 15-25 mm by 1 mm composing 11 levels. In total, the combination of light and moderate precipitation thresholds composes 66 schemes. To be noted, these 66 schemes are carried out towards all grid points to determine an optimal one for each point. The changes in percentage of the TSs after applying the optimal threshold scheme for each grid point in China are displayed in Figure 4. To be noted, although the CBREM is performed by newly generated threshold schemes, the evaluation is carried out toward the traditional precipitation thresholds (by CMA) of 10 mm, 25 mm and 50 mm. Due to that the rainstorm threshold of 50 mm maintains, the distribution of TS changes is omitted in Figure 4. It shows conspicuous improvements in TS for the whole area and for all thresholds. For the light precipitation, the improvement magnitudes range between 5-15%, with generally lower values over the west. As for the moderate precipitation, the most visible improvements with percentages greater than 20% are located on the east of China, which shows more significant advancements than those of light precipitation. For the heavy precipitation, the improvements are the least among the three thresholds, with magnitudes of 20% located over only few parts of Southeast China.

![Figure 4](image.png)

**Figure 4.** The percentage improvement of TS under the optimal thresholds over the TS under the traditional thresholds for each grid point in China.

4. Summary

Using the daily accumulated precipitation data provided by the ensemble forecasts of ECMWF, JMA, UKMO, and the CMORPH observational data, the experiments of ensemble methods such as EMN,
BREM, CBREM are performed. The obtained results are as follows: (1) Multimodel forecast improves the forecast skill of precipitation conspicuously. BREM has better performance than EMN. In addition, ECMWF provides the best precipitation forecast among the three single-center models used in this study. (2) The CBREM by traditional grading thresholds shows more skillful forecast of precipitation in contrast to BREM. The c-ECMWF has the best performance on the whole. The improvement of forecast for light precipitation is most significant, and there is a negative impact of CBREM in forecasting precipitation with a moderate threshold. In addition, the advancement of CBREM for heavy precipitation is mainly reflected during lead times of 4-7 days. (3) The forecast skill can be further improved by optimal grading thresholds. Improvements are most significant for moderate precipitation forecast. The advancement of optimal grading thresholds is relatively limited for Northwest China. More efforts on determining optimal grading thresholds for each grid point are necessary in the future.

Acknowledgments
This work has been jointly supported by the National Key R & D Program of China (Grant No. 2017YFC1502002), Fujian Key Laboratory of Severe Weather (Grant No. 2020KFKT02), and LCP/CMA (Grant No. 2017Z016).

References
[1] Zhang L, Zhu X, Fraedrich K, et al 2014 Interdecadal variability of winter precipitation in Southeast China Clim Dyn 43 2239-2248
[2] Zhang L, Sielmann F, Fraedrich K, et al 2017 Atmospheric response to Indian Ocean Dipole forcing: changes of Southeast China winter precipitation under global warming Clim Dyn 48 1467-1482
[3] Fan Y, Fan K, Xu Z, et al 2018 ENSO-South China Sea summer monsoon interaction modulated by the Atlantic Multidecadal Oscillation J Clim 31 3061-3076
[4] Zhu S, Ge F, Fan Y, et al 2020 Conspicuous temperature extremes over Southeast Asia: seasonal variations under 1.5 °C or 2 °C global warming Clim Chang 160 343-360
[5] Zhang L, Fraedrich K, Zhu X, et al 2015 Interannual variability of winter precipitation in Southeast China Theor Appl Climatol 119 229-238
[6] Ge F, Zhu S, Peng T, et al 2019 Risks of precipitation extremes over Southeast Asia: does 1.5 °C or 2 °C global warming make a difference? Environ Res Lett 14 044015
[7] Zhu S, Remedio A R C, Sein D, et al 2020 Added value of the regionally coupled model ROM in the East Asian summer monsoon modeling Theor Appl Climatol 140 375-387
[8] Zhu S, Ge F, Sielmann F, et al 2020 Seasonal temperature response over the Indochina Peninsula to a worst-case high emission forcing: a study with the regionally coupled model ROM Theor Appl Climatol 1-10 https://doi.org/10.1007/s00704-020-03345-7
[9] Ji L, Zhi X, Zhu S 2019 Probabilistic precipitation forecasting over East Asia using Bayesian model averaging Wea Forecasting 34 377-392
[10] Zhang L, Zhi X 2015 Multimodel consensus forecasting of low temperature and icy weather over central and southern China in early 2008 J Trop Meteorol 21 67-75
[11] Ji L, Zhi X, Simmer C, et al 2020 Multimodel ensemble forecast of precipitation based on an object-based diagnostic evaluation Mon Wea Rev 148 2591-2606
[12] Peng T, Zhi X, et al 2020 Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning post-processing methods Atmosphere 11 823
[13] Zhi X, Tian X, Liu P, et al 2019 Interdecadal variations in winter extratropical anticyclones in East Asia and their impacts on the decadal mode of East Asian surface air temperature Meteor Atmos Phys 131 1763-1775
[14] Zhi X, Qi H, Bai Y, et al 2012 A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data J Meteor Res 26 41-51