

Article ID: 1006-8775(2024)04-0361-12

0–12 Hour QPFs of HRRR-TLE Using Optimized Probability-Matching Method: Taking Hunan Province as an Example

LIU Jin-qing (刘金卿)¹, MAO Zi-yi (毛紫怡)², DAI Guang-feng (戴光丰)³, YANG Zhao-li (杨兆礼)³,
PENG Xuan (彭 轩)^{1,4}

(1. College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073 China; 2. Hunan Meteorological Observatory, Hunan Meteorological Bureau, Changsha 410118 China; 3. Guangzhou Institute of Tropical and Marine Meteorology/Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, China Meteorological Administration, Guangzhou 510641 China; 4. Key Laboratory of High Impact Weather (Special), China Meteorological Administration, Changsha 410073 China)

Abstract: In real-time operations, the minutely/hourly updated high-resolution rapid refresh (HRRR) system is one of the most expensive numerical weather prediction (NWP) models. Based on a twenty-member HRRR-time-lagged-ensemble (HRRR-TLE) system developed from two real-time convection-permitting HRRR models, CMA-GD(R3) and CMA-SH3, from the China Meteorological Administration (CMA), this study proposes an optimized probability-matching (OPM) technique to improve 0–12 h quantitative precipitation forecasts (QPFs) based on the correlation and error relationships between ensemble forecasts and observations during the training window. Then, a series of sensitivity experiments using different cost functions and optimized ratios was conducted to further improve OPM predictions. The results indicate that: (1) In the HRRR-TLE system, there is no always optimal member in both weak rain and severe rain forecasts, as measured by the equitable threat score (ETS) and bias extent (BE) at four thresholds (1+, 5+, 10+, and 20+ mm h⁻¹; e.g., “1+” means ≥ 1). (2) Compared with the HRRR-TLE system, the QPFs generated by the traditional PM technique showed a notable increase in ETS and a decrease in BE at all of the above thresholds. Compared with the traditional probability-matching method (PM), OPM can generate more skillful forecasts on both spatial representations and rain rates by using the sliding-weight method and optimized ensembles, respectively. (3) In particular, in the 20+ mm h⁻¹ forecasts, which are often difficult to predict, the ETS of the optimal OPM test, with a 20% optimization ratio and symmetric mean absolute percentage error cost function, increased by 64.6%, and the BE decreased by 5.7%, relative to PM. Moreover, OPM shows good stability in both daytime and nighttime periods.

Key words: high-resolution rapid refresh; convection-permitting; time-lagged; probability-matching; CMA-GD(R3); CMA-SH3

CLC number: P456.7

Document code: A

Citation: LIU Jin-qing, MAO Zi-yi, DAI Guang-feng, et al. 0–12 Hour QPFs of HRRR-TLE Using Optimized Probability-Matching Method: Taking Hunan Province as an Example [J]. *Journal of Tropical Meteorology*, 2024, 30(4): 361–372, <https://doi.org/10.3724/j.1006-8775.2024.032>

1 INTRODUCTION

The development of numerical weather prediction (NWP) is guided by the goal of generating forecasts with greater accuracy (Palmer ^[1]; Bauer et al. ^[2]; Swinbank et al. ^[3]; Powers et al. ^[4]; Wong et al. ^[5]; Cafaro et al. ^[6]; Wolff et al. ^[7]). The minutely/hourly updated convection-permitting high-resolution rapid refresh (HRRR) system aims to achieve this goal by swiftly assimilating a more significant number of the latest weather observations (Benjamin et al. ^[8]; Bytheway et al. ^[9]; Slivinski et al. ^[10]). In the United

States, the National Centers for Environmental Prediction (NCEP), an arm of the NOAA's National Weather Service, released the first convection-permitting HRRR version (3-km grid spacing and 1-h cycling) in 2014 (<https://rapidrefresh.noaa.gov/hrrr/>), and the latest version HRRRv4 was released in late 2020 (Dowell et al. ^[11]; James et al. ^[12]). In China, a series of HRRR forecasts are currently operational at the China Meteorological Administration (CMA), such as CMA-BJ, CMA-GD(R3), and CMA-SH3 (Li et al. ^[13]; Chen et al. ^[14]; Liu et al. ^[15]). These HRRR systems have become increasingly vital for short-range weather guidance and early warning, especially for rapidly developing severe convective weather events with short lead time (Turner et al. ^[16]; Pinto et al. ^[17]; Dougherty et al. ^[18]).

However, such high-frequency HRRR systems face limitations in realistic applications. One challenge is that the latest model outputs are not always the optimal predictions among massive outputs from different run times because of the models' initial spin-up problem (Short and Petch ^[19]; Warner et al. ^[20]; Khattiwala ^[21]).

Received 2023-03-31; **Revised** 2024-08-15; **Accepted** 2024-11-15

Funding: Natural Science Foundation of Hunan Province (2021JC0009); National Natural Science Foundation of China (42375002, 42105146); Key Projects of Hunan Meteorological Bureau (XQKJ22A004)

Biography: LIU Jin-qing, Ph.D. candidate, Senior Engineer, primarily undertaking research on FengYun satellites and numerical weather prediction.

Corresponding author: LIU Jin-qing, e-mail: liujq_0912@126.com

Another issue is that due to the rapid error growth at the convection-permitting scale in HRRR, their forecasts are usually limited to a very short duration, making long-term predictions unpredictable. Therefore, effectively maximizing the use of HRRR outputs remains a challenge.

In recent years, the time-lagged ensemble (TLE) has emerged as a competitive approach in attempts to address the above issues at a low computational cost (Jie et al. [22]; Raynaud and Bouttier [23]; DelSole et al. [24]). Each HRRR-TLE member originates from a series of model's earlier runs with different initial conditions (ICs) in the same or different HRRR model. The simplest ensemble mean (EM) could correct forecast errors resulting from the model's initial spin-up and convective-scale rapid error growth thanks to its smoothing effect, which tends to reduce intense model features and reduce the model's random errors (Lu et al. [25]; Porson et al. [26]). However, whether the TLE approach improves the skill of HRRR forecasts is not always clear. Given the high computational cost and the long-term unpredictability, their forecasts are usually within 24 hours. Therefore, the membership size of TLE is not too large. Thus, if no sufficient assimilation data is added in each HRRR run, the TLE members will be similar, and any weighting method will fail to improve QPFs relative to a single member. For instance, Raynaud et al. [27] demonstrated that the use of the flow-dependent Bayesian weighting of TLE forecasts does not lead to a noticeable gain in the forecast quality.

Based on two domestic real-time HRRR models, CMA-GD(R3) and CMA-SH3, this study aims to optimize the use of massive HRRR outputs to improve predictions on both spatial representations and rain rates by employing an optimized probability-matching (OPM) method. Section 2 provides an overview of the data and introduces the OPM technique. Section 3 discusses the results of the experiments. The summary and conclusions are presented in the final section.

2 DATA AND METHODOLOGY

2.1 Introduction to CMA-GD(R3) and CMA-SH3

The precipitation products from two real-time convection-permitting HRRR models, CMA-GD(R3) and CMA-SH3, provided by the CMA were used in this paper (Fig. 1), with a horizontal grid spacing of 3 km from March 1 to August 31, 2021. Note that only the data from May 1 to August 31 were used for algorithm verification, and the data from March 1 to April 28 were only used as a training dataset.

The CMA-GD model was developed by the Guangdong Meteorological Service based on the Global/Regional Assimilation and Prediction System (GRAPES), with horizontal resolutions of 1 km (official name: CMA-GD(R1)) and 3 km (official name: CMA-GD(R3)), respectively (Zhong et al. [28]; Chen et al. [29]). The 3-km version is used in this study because the 1-km version is only for Guangdong (18° – 27° N, 107° – 119° E) and does not include our study area.

The Shanghai Meteorological Service has been running the CMA-SH model for more than 10 years, based on the Weather Research and Forecasting (WRF) model. It has horizontal resolutions of 9 km (official name: CMA-SH9) and 3 km (official name: CMA-SH3), respectively (Zhang et al. [30]). The 3-km version is used in this study because the 9-km version only provides twice-daily products and does not meet our short-range forecasting needs.

Table 1 shows the configurations of these two HRRR models. They have different microphysics, planetary boundary layers, initial conditions, and lateral boundary conditions, as well as different numbers of vertical levels and forecast hours.

Stational observations were used as the “truth” in this study. The HRRR gridded data were interpolated to 420 selected stations in Hunan Province (24.5° – 30.1° N, 108.7° – 114.5° E) with the nearest neighborhood method, which is

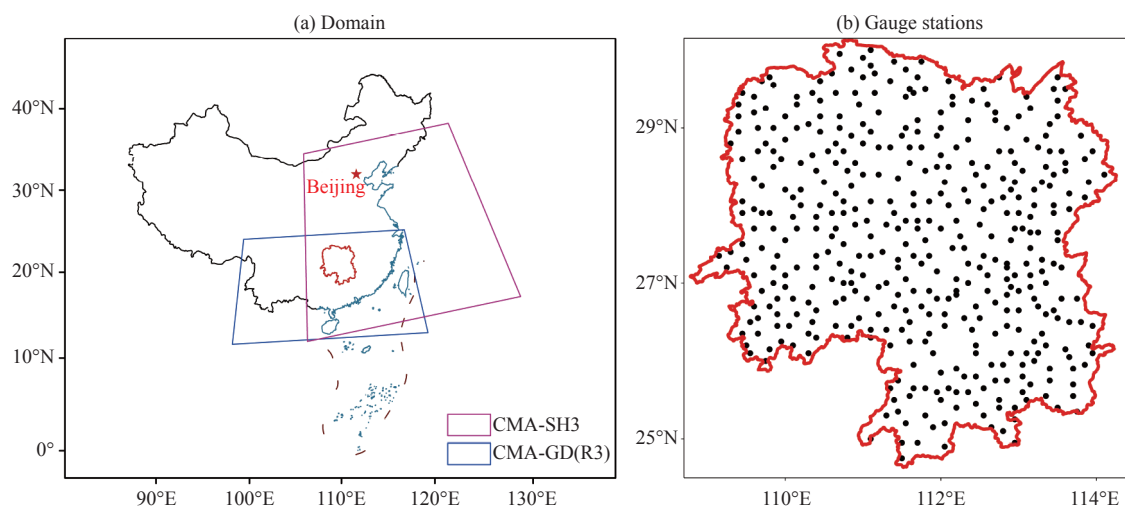


Figure 1. (a) Domain of CMA-GD(R3) (blue box: 16.6° – 30.76° N, 96.6° – 122.76° E) and CMA-SH3 (pink box: 17.38° – 43.41° N, 106.61° – 136.03° E) in the Lambert map projection. The red line represents the location of Hunan Province (24.5° – 30.1° N, 108.7° – 114.5° E). (b) Locations of 420 automatic rain gauge stations (black dots) in Hunan.

Table 1. Model and physics configuration for convection-permitting HRRR models CMA-GD(R3) and CMA-SH3.

Configuration	CMA-GD(R3)	CMA-SH3
Microphysics	WSM6	Thompson
Planetary boundary layer	New-MRF	Yonsei University (YSU)
Initial conditions	ECMWF	Global Forecasting System (GFS)
Lateral boundary conditions	ECMWF	Global Forecasting System (GFS)
Vertical levels	65	51
Grid spacing	~3 km	~3 km
Forecast duration	36 h	24 h
Operation frequency	Per hour	Per hour
Organization	Guangdong Meteorological Service	Shanghai Meteorological Service

applicable to HRRR (James et al. [12]) to be consistent with the observations. This interpolation was performed before the development of the HRRR-TLE system.

Moreover, given that the HRRR model needs some spatial spin-up to simulate realistic precipitation, at least 150 km grid spacing on the lateral boundary needs to be removed in the application (Brisson et al. [31]), as the driving model did not consider the morphogenesis of hydrometeors. Therefore, data from CMA-GD(R3) on the northern edge of Hunan was slightly affected by the spatial spin-up, but the overall dataset for most areas of Hunan is reliable.

2.2 Development of the time-lagged ensemble and the HRRR-TLE system

The TLE is developed from a series of high-frequency runs with different initial conditions, which uses the most recent model outputs to forecast the next 0–12 h. For example, Fig. 2 shows a typical TLE structure with ten members from a lag time of 1 h to 10 h. Given that the latency of the model is about 40 minutes, if the current time is $T+0$ h, then the latest updated run and the farthest run were thus generated by the previous hour ($T-1$ h model run, namely T1) and the previous 10th hour ($T-10$ h model run, namely T10), respectively.

The HRRR-TLE system used in this study is a twenty-member size system, including ten members from CMA-GD(R3) and ten members from CMA-SH3. The details of each member are shown in Table 2; for each lag time, members come from these two models. After the system was constructed, four initial times, at 0000, 0600, 1200, and 1800 UTC, were used for the calculation and verification of the algorithm.

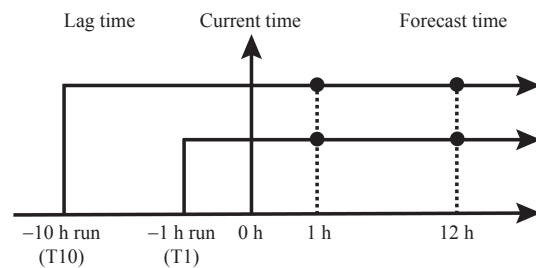


Figure 2. Illustration of a typical time-lagged ensemble (TLE) with a size of ten members (T1–T10). Taking a single model CMA-GD(R3) as an example, if the current time is $T+0$ h, then the latest outputs are generated by the $T-1$ h model run (namely T1) since the latency of this model is about 40 minutes, and the farthest run is the $T-10$ h model run (namely T10).

2.3 Optimized probability-matching method and sensitivity experiments

The traditional probability-matching method (PM) is an equal-weight bias-correction technique based on EM. Technically, it uses the ensemble average as the precipitation spatial distribution and the median of ranked ensemble predictions as precipitation intensity (Fig. 3a).

Compared with PM, OPM aims to improve precipitation predictions on both distribution and intensity based on the performance of each ensemble member during the training window (Fig. 3b). In order to have enough samples for heavy rain, this study adopted the previous 60 days as the training window.

In rainfall distributions, a dynamic sliding-weight average method that is demonstrated as effective for multi-model bias correction (Liu et al. [15]) was employed to

Table 2. Details of two latest and two farthest HRRR-TLE members from CMA-GD(R3) and CMA-SH3, respectively, taking the hourly precipitation from 1200 to 1300 UTC as an example of prediction.

HRRR-TLE members	Description	Model initial time	Lag time	Forecast hour
T1(GD*)	Latest outputs of CMA-GD(R3)	1100 UTC	1 h	1–2 h
T1(SH**)	Latest outputs of CMA-SH3	1100 UTC	1 h	1–2 h
T10(GD)	Farthest outputs of CMA-GD(R3)	0200 UTC	10 h	10–11 h
T10(SH)	Farthest outputs of CMA-SH3	0200 UTC	10 h	10–11 h

*CMA-GD(R3) based on GRAPES model

**CMA-SH3 based on WRF model

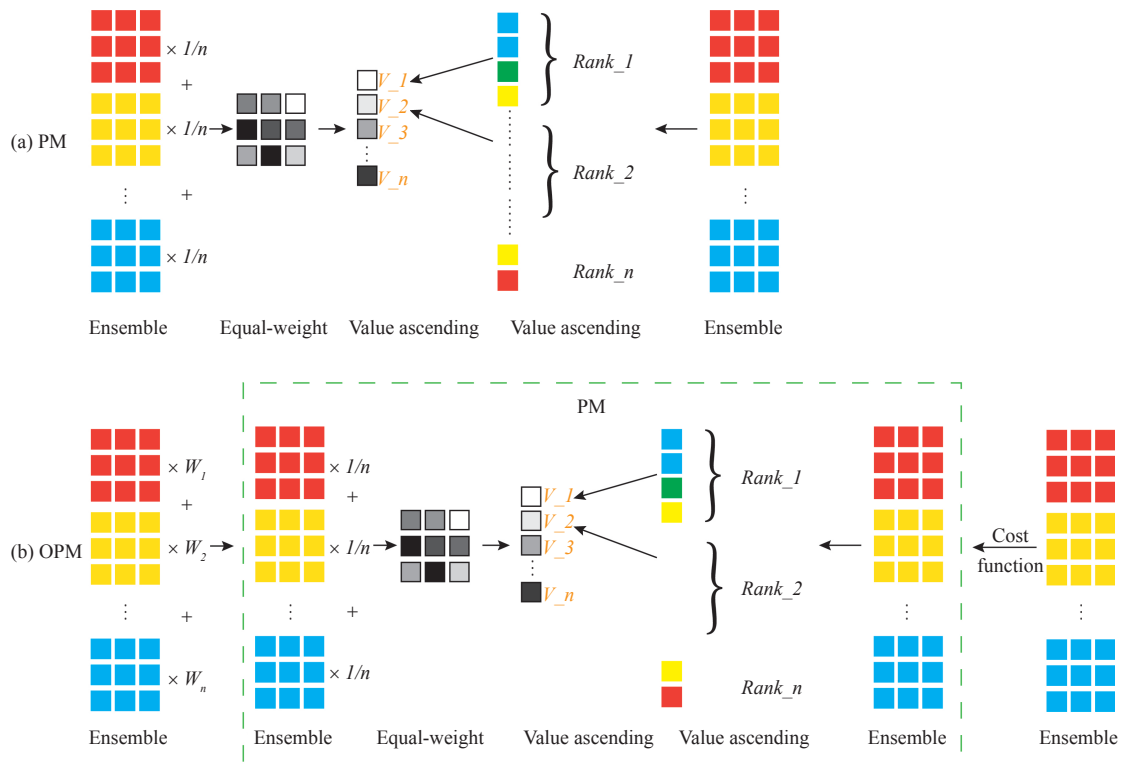


Figure 3. An illustration of (a) PM and (b) OPM. OPM has two upgrades relative to PM: one is the weight of the ensemble used for precipitation distribution, and the other is the selection of ensembles used for precipitation intensity.

replace the equal-weight EM of PM. The weight of each HRRR-TLE member is calculated as follows.

Step 1: For a given training window, we calculate the Spearman correlation coefficient (SCC) of each ensemble member at the same forecast time every day.

Step 2: For each forecast time, only the maximum SCC (SCC-max) and corresponding ensemble members are retained. Then, the sum of SCC-max (SSCC-max) retained by each member during the entire training period is calculated.

Step 3: The weight of each member is determined by dividing its SSCC-max by the SSCC-max of all members.

After the above steps, the improved distribution is obtained via a weighted average method.

With respect to rainfall intensity, a subset of successful members with higher accuracy is selected as the source of rain rates. Generally, there are many widely used performance indicators for quantitative precipitation forecasts (QPFs), such as the mean absolute error (MAE), root mean square error (RMSE), equitable threat score (ETS), and frequency bias. However, the scale-dependent measures (MAE and MSE) are sensitive to forecasting outliers and may produce extremely biased results when the dataset includes outliers (Chen et al. [32]), while the dichotomous metrics (ETS and bias) cannot simultaneously assess the performance of different thresholds. Therefore, here, two indicators that theoretically meet our optimization requirements were used as the cost function for selecting a subset of best-performing members. One is the commonly used correlation coefficient (CC), and the

other is the symmetric mean absolute percentage error (SMAPE, Armstrong [33]). CC can effectively measure the linear relationship between the observation and predictions; while SMAPE is informative in representing rainfall accuracy, is resistant to outliers, and is favored for rainstorm forecasting. The equations for SMAPE and CC are shown as follows.

$$\text{SMAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{F_i - O_i}{F_i + O_i} \right|,$$

$$\text{CC} = \frac{\sum_{i=1}^N (F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (F_i - \bar{F})^2 \sum_{i=1}^N (O_i - \bar{O})^2}},$$

where O and F represent the observation and model outputs, respectively, while i represents the i th sample, and N represents the number of valid samples.

Theoretically, SMAPE has the following advantages: (1) it is scale-independent and avoids being dominated by a single outlier; (2) it has a bias favoring heavy rainfall forecasting, which is effective in selecting predictions with a higher ETS and lower bias.

Furthermore, a series of sensitivity experiments using different optimized ratios were designed to test the sensitivity of the prediction relative to the optimization size. The steps of OPM tests are as follows.

Step 1: For a given training window, the cost function of each ensemble member is calculated during the entire training period.

Step 2: Ranking the values of the cost function, we obtain the sequence numbers of each ensemble member. Then, a subset of optimized members is constructed by the best N members (Table 3).

2.4 Verification methods

Common scores (Rossa et al. [34]) of the equitable threat score (ETS), bias extent (BE), and precipitation intensity (PI) were used to assess the skill in predicting hourly precipitation at four given thresholds (1+, 5+, 10+, and 20+ mm h⁻¹; e.g., “1+” means ≥ 1). The evaluation expressions are presented in Table 4, that ETS verifies the prediction of precipitation distribution, BE verifies the prediction of precipitation frequency, and PI verifies the average intensity of those rainfall samples (0.1+ mm h⁻¹). The perfect score for ETS and BE is 1 and 0, respectively, while an ETS close to zero or negative indicates poor location prediction skill, and a BE close to zero indicates poor frequency prediction at a given threshold.

3 RESULTS

3.1 Results of OPM

Figure 4 compares the performance of HRRR-TLE members, PM, and OPM with different cost function. On average, the ETS skills of the HRRR-TLE system are 0.118, 0.046, 0.019, and 0.005 at thresholds 1+, 5+, 10+, and 20+ mm h⁻¹, respectively, and the BE skills are 0.501, 0.544, 0.542, and 0.550. Specifically, among all twenty ensemble members, T1(GD), T6(GD), T7(GD), and T4(GD) show the best performance at the above thresholds, with ETS scores of 0.142, 0.058, 0.026, and 0.009; the skills of T1(GD) rank 1st, 5th, 7th, and 4th at the above four thresholds, respectively (Figs. 4a and 4b). With respect to frequency bias, T10(GD) have the lowest BE scores at all thresholds, with scores of 0.253, 0.293, 0.240, and 0.056, respectively (Figs. 4c and 4d). Therefore, we

conclude that the two latest members T1(GD) and T1(SH) are not always the best-performing members in the HRRR-TLE system, and there is not a consistently best member for different indicators. However, the traditional PM can exceed the overall skill of the system, with an increase of 23.6%, 22.9%, 14.4%, and 4.8% in ETS, and a decrease of 0.1%, 0.7%, 1.3%, and 8.9% in BE (Table 5).

As shown in the OPM100% test, after changing the distribution, the skills of all indicators are improved or maintained compared with PM, indicating that the dynamic sliding-weight method of OPM is effective for the HRRR-TLE system. Furthermore, if both spatial distribution and rain rates are changed at the same time (eg., OPM10%–90%), the improvement will be greater. In sensitivity experiments with optimization rates between 10% and 90%, both OPM(SMAPE) and OPM(CC) showed a notable advantage over PM in ETS and BE skills. Additionally, the cost function SMAPE works better than CC in increasing the ETS skill for the 20+ mm h⁻¹ precipitation; therefore, OPM(SMAPE) tests are used in the following discussion.

3.2 Results of sensitivity experiments

Figure 5 compares the performances of different optimization ratios on the OPM(SMAPE) (Table 2). The results reveal that all tests have a higher ETS than PM, but only those with a ratio below 50% can decrease the BE skills. Specifically, OPM20% with four members and OPM10% with two members achieve the best ETS and BE skills in 20+ mm h⁻¹ predictions. Given the operational demand of severe convection forecasting, the OPM20% test is selected as the optimal test in the following further analysis. Its ETS and BE skills in 20+ mm h⁻¹ are 0.009 and 0.472, respectively, with a 64.6% and 5.7% improvement relative to PM (Table 6). According to the principle of the OPM method, it has been improved in

Table 3. Design of sensitivity experiments. N represents the optimization ratios for OPM, including 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%, respectively. Note that when the optimization ratio is 100%, the OPMs with different cost functions are the same, and the only difference between OPM and PM is the spatial distribution.

Experiment	Distribution	Intensity	Training window	Cost function
PM	Equal-weight ensemble mean	Median of ranked ensemble predictions	None	None
OPM(SMAPE)	Sliding-weight average	Same as PM, but using optimized ensembles with the best N members	Past 60 days	SMAPE
OPM(CC)	Same as OPM(SMAPE)	Same as OPM(SMAPE)	Past 60 days	CC

Table 4. Metrics used in this study. a , b , c , and d represent the number of hits, false alarms, misses, and correct negatives, respectively; a_{random} represents the number of hits due to random chance; i represents the i th precipitation sample (0.1+ mm h⁻¹); F and N represents the intensity and number of valid precipitation samples, respectively.

Metrics	Expression	Perfect score	Description
ETS	$\frac{a - a_{\text{random}}}{a + b + c - a_{\text{random}}}$	1	Equitable threat score
BE	$\left \frac{a+b}{a+c} - 1 \right $	0	Bias extent
PI	$\frac{1}{N} \sum_{i=1}^N F_i$		Precipitation intensity

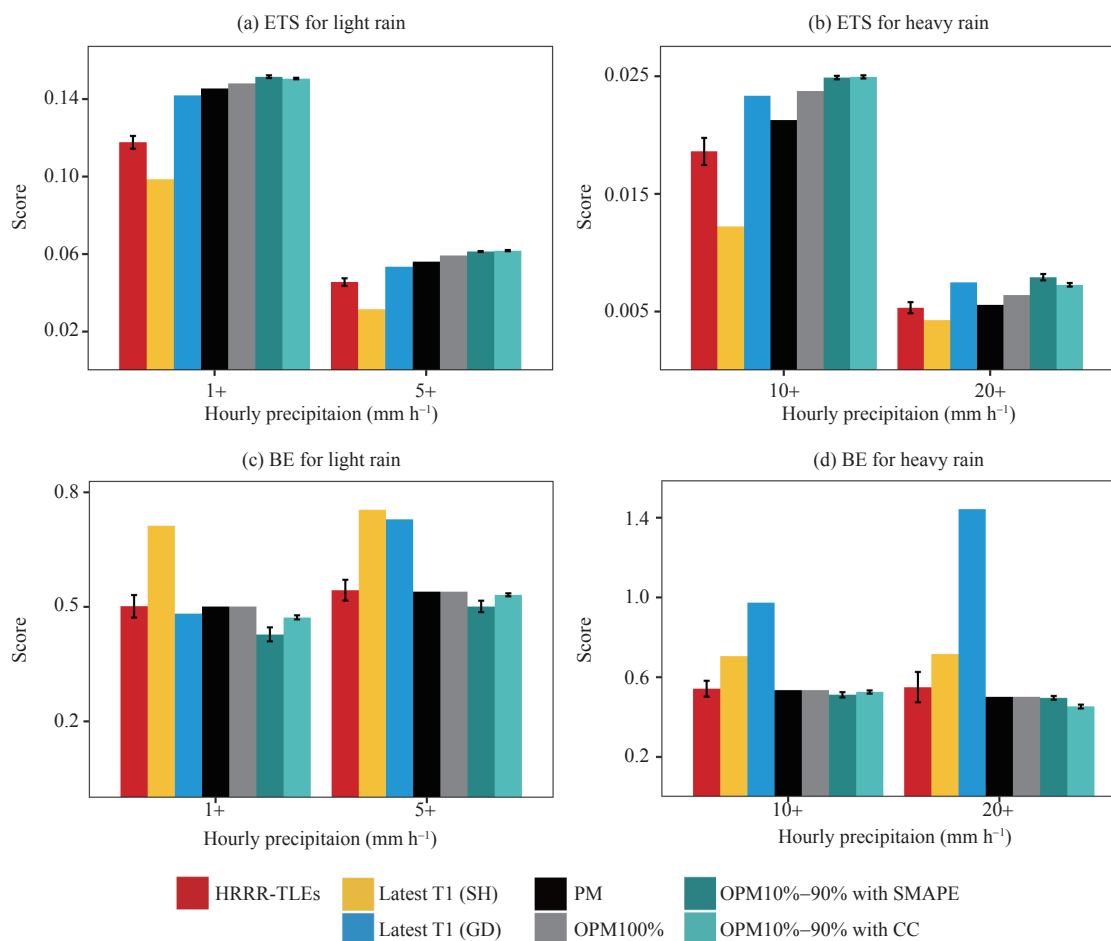


Figure 4. Evaluation histograms of each HRRR-TLE member (red), the latest two members (yellow and blue), PM (black), OPM100% (gray), and OPM20%–90% with different cost function (dark green and light green) from 1 April to 31 August 2021 for 0–12 h forecasts. (a, b) ETS; (c, d) BE at given thresholds (1+, 5+, 10+, and 20+ mm h⁻¹). The PM and OPM methods are shown in Fig. 3. Note that the OPM100% tests with different cost functions are the same, and the only difference between OPM100% and PM is the spatial distribution. The error bar represents the mean $\pm$ standard error. The perfect score is 1 and 0 for ETS and BE, respectively.

Table 5. Improvement of PM relative to the average of HRRR-TLE system.

PM	ETS				BE			
	1+	5+	10+	20+	1+	5+	10+	20+
Score	0.145	0.056	0.021	0.006	0.501	0.540	0.535	0.501
Improvement	23.6%	22.9%	14.4%	4.8%	0.1%	0.7%	1.3%	8.9%

aspects distribution and intensity (Fig. 3).

3.3 Diurnal variation

The above results indicate that OPM works better than PM in long-term statistics. Here, we further explore the time of day during which OPM performs well and the times during which it performs poorly. Fig. 6 compares the results for PM and OPM20% in 24-h forecasting.

Overall, OPM works better in most time periods of daily forecasting than PM, with higher ETS skills in both weak rain (Fig. 6a) and heavy rain (Fig. 6b), and it exhibits a lower SMAPE (Fig. 6d). In particular, the results also show that the rainfall intensity of OPM has a notable increase relative to PM and observations (Fig. 6c), which is consistent with the unique advantages of the SMAPE

cost function in favor of heavy rainfall forecasting.

Furthermore, during the daytime (0000–1100 UTC), OPM improves PM during nine moments (75%), nine moments (75%), and eleven moments (92%) in the ETS of weak rain, ETS of heavy rain and SMAPE, respectively. In contrast, during nighttime (1200–2300 UTC), it works for nine moments (75%), ten moments (83%), and eleven moments (92%) with respect to the above metrics. Briefly, OPM can steadily improve PM in both daytime and nighttime periods.

3.4 Severe weather prediction examples

Figures 7 and 8 present two severe convective weather (SCW, ≥ 20 mm h⁻¹) examples over Hunan Province from 2021 to illustrate the effectiveness of

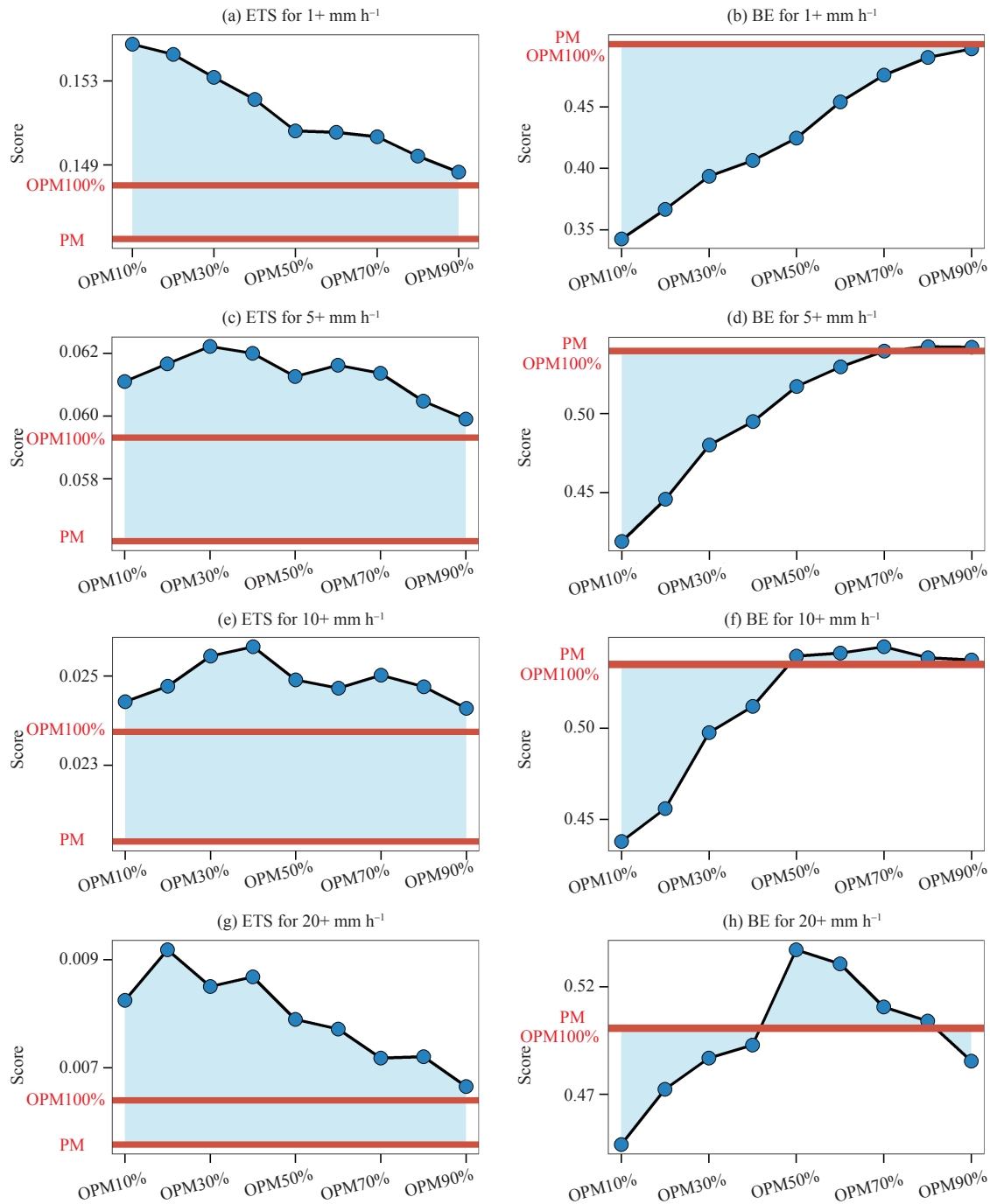


Figure 5. Comparison of 0–12 h ETS (left panels) and BE (right panels) in the OPM(SMAPE) sensitivity experiments at thresholds of (a, b) 1+ mm h⁻¹, (c, d) 5+ mm h⁻¹, (e, f) 10+ mm h⁻¹, and (g, h) 20+ mm h⁻¹ from April 1 to August 31, 2021. Note that OPM100% only has a spatial distribution different from PM. The perfect score is 1 and 0 for ETS and BE, respectively. The details of the experiments are shown in Table 3.

Table 6. Improvement of OPM20% relative to PM.

OPM20%	ETS				BE			
	1+	5+	10+	20+	1+	5+	10+	20+
Score	0.154	0.062	0.025	0.009	0.367	0.446	0.456	0.472
Improvement over PM	6.1%	10.1%	16.4%	64.6%	26.8%	17.4%	14.8%	5.7%

OPM, which is initially at 1800 UTC May 6 (Case I) and 1200 UTC May 2 (Case II), respectively. In these two

cases, the 0–12 h prediction trends of CMA-SH3 and CMA-GD(R3) differ; that is, as time approaches,

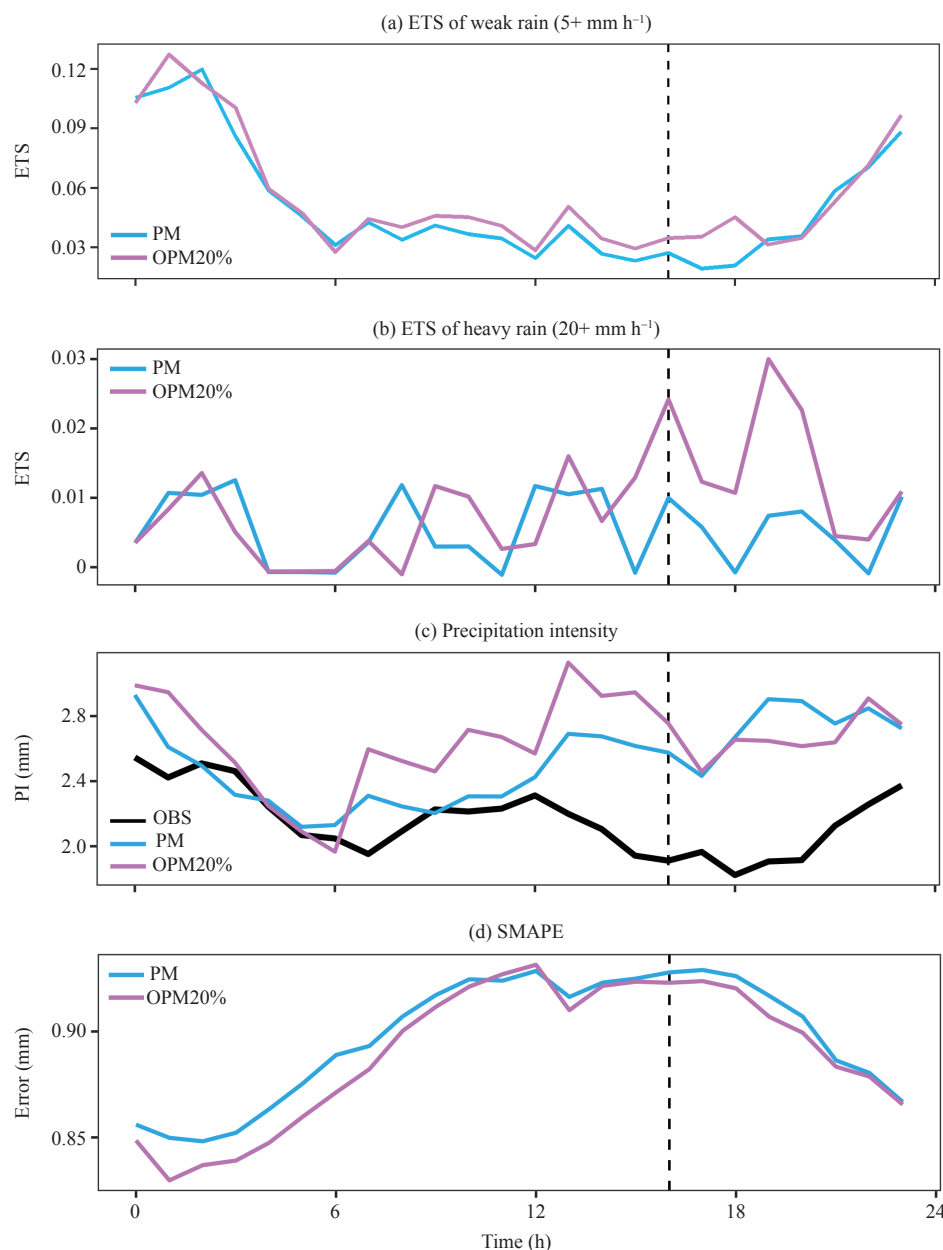


Figure 6. Diurnal variation for PM and OPM20% from April 1 to August 31, 2021. (a) ETS of weak rain $5+ \text{ mm h}^{-1}$, (b) ETS of heavy rain $10+ \text{ mm h}^{-1}$, (c) PI, and (d) SMAPE. The black, blue, and pink lines represent the observation, PM, and OPM20%, respectively. The dotted lines indicate the results of 1600 UTC.

precipitation weakens in one and strengthens in the other (Figs. 8e–8x).

According to the observation, in Case I, there is convective cloud growing in the middle of Hunan (Fig. 7c), and the convective rain happens under the influence of cold air and shear lines, with a divergence reaching $-1 \times 10^{-5} \text{ s}^{-1}$ on the 850 hPa weather chart (Fig. 7a) and a maximum intensity of hourly precipitation of 36.6 mm h^{-1} (Fig. 8a1). In Case II, a convective system is moving into Hunan (Fig. 7c), and the convective rainfall occurs under the influence of a developing southwest vortex in the Sichuan Basin and strong low-level jet ($\geq 12 \text{ m s}^{-1}$) on the 850 hPa weather chart (Fig. 7b), with a maximum hourly precipitation of 40.5 mm h^{-1} (Fig. 8a).

Furthermore, the difference in predictions between PM and OPM indicates that the new approach effectively eliminates the false alarms of redundant precipitation in the rain belt and decreases the number of missing rainfall (Figs. 8b–8d). For example, in Case I, the redundant precipitation of PM in western Hunan is weakened, and the missing severe precipitation in eastern Hunan is enhanced by the OPM (Figs. 8b1 and 8c1). The results of Case II are consistent with Case I in that the new method can also strengthen the severe precipitation in the south and eliminate the redundant precipitation in the northwest. As a result, the ETS skill of the convective precipitation of OPM increases by 31.6% and 112.2% relative to PM in two cases, respectively, and the BE decreases by 99.4%

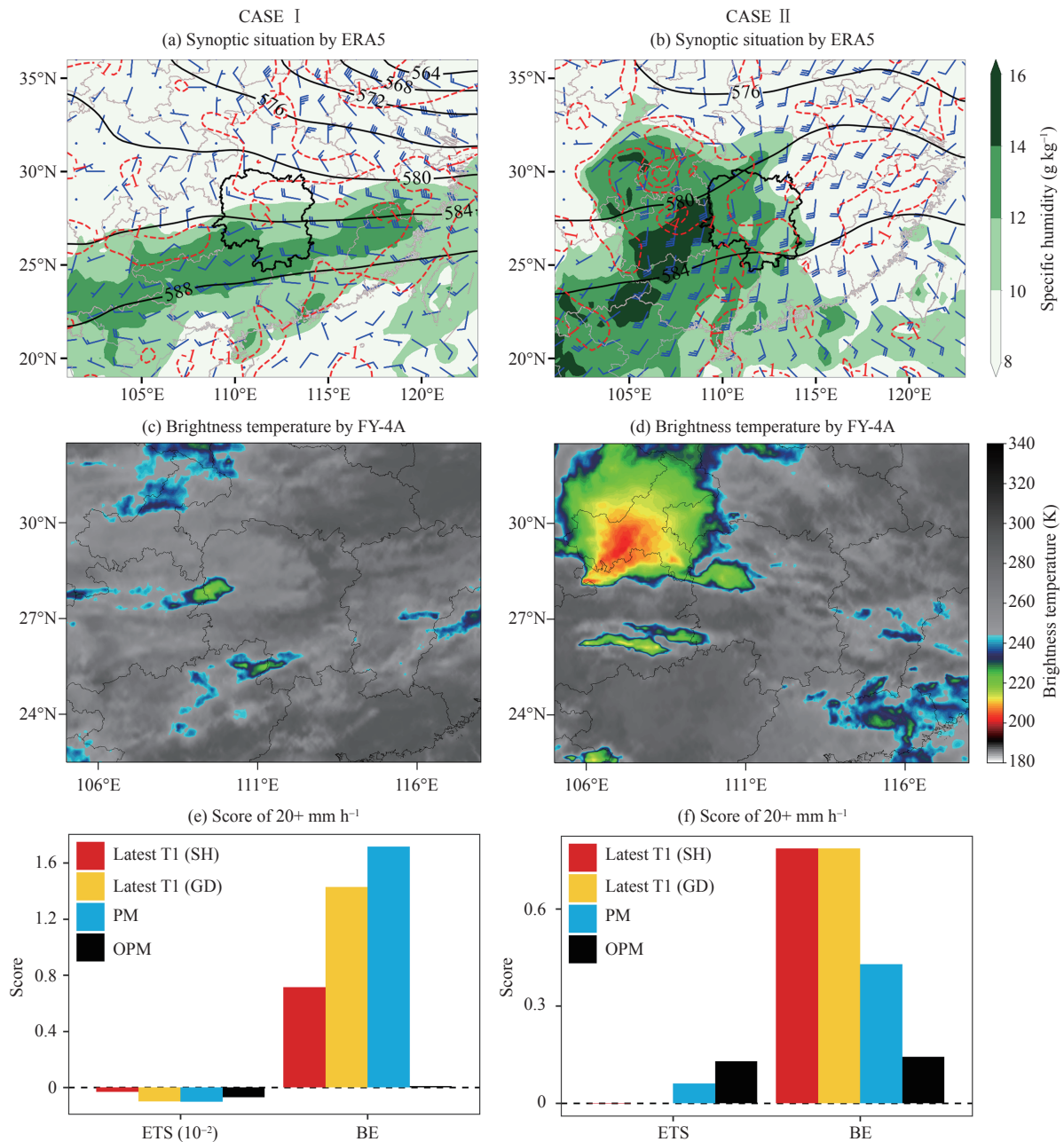


Figure 7. Observation and forecasting performance for two severe convective weather cases ($\geq 20 \text{ mm h}^{-1}$) over Hunan Province. (a, b) Synoptic situation by ERA5, including geopotential height (black contour, contour interval is 4, units: dagpm) at 500 hPa, wind (blue barb, units: m s^{-1}), divergence (red contour, units: 10^{-5} s^{-1}), and specific humidity (shading, units: g kg^{-1}) at 850 hPa. (c, d) Brightness temperature (units: K) at 10.8 μm by FengYun-4A. (e, f) 0–12h ETS and BE skills of 20+ mm h^{-1} by two latest HRRR-TLE members, PM and OPM20%. The initial forecasting times are 1800 UTC May 6 (Case I) and 1200 UTC May 2, 2021 (Case II), respectively, and the observed time is 6 h after the initial time.

and 66.7%, respectively (Figs. 7e and 7f).

4 CONCLUSIONS

Based on a twenty-member HRRR-TLE system developed from two domestic real-time HRRR models, i. e., CMA-GD(R3) and CMA-SH3, this study proposed an OPM method to improve QPFs of both spatial representations and rain rates, based on the relationship between ensemble forecasts and the observations during the training window. Furthermore, a series of sensitivity experiments using different indicators and different

optimized ratios were conducted to test the sensitivity of the OPM prediction relative to the cost function and optimization size. The following is a summary of the findings.

In the HRRR-TLE system, the latest two members T1 (GD) and T1(SH) were not always the best-performing members in long-term statistical verification, as measured by ETS and BE. The overall skills at thresholds of 1+, 5+, 10+, and 20+ mm h^{-1} were 0.118, 0.046, 0.019, and 0.005 in ETS, respectively, and 0.501, 0.544, 0.542, and 0.550 in BE. Specifically, T1(GD), T6(GD), T7(GD), and T4(GD)

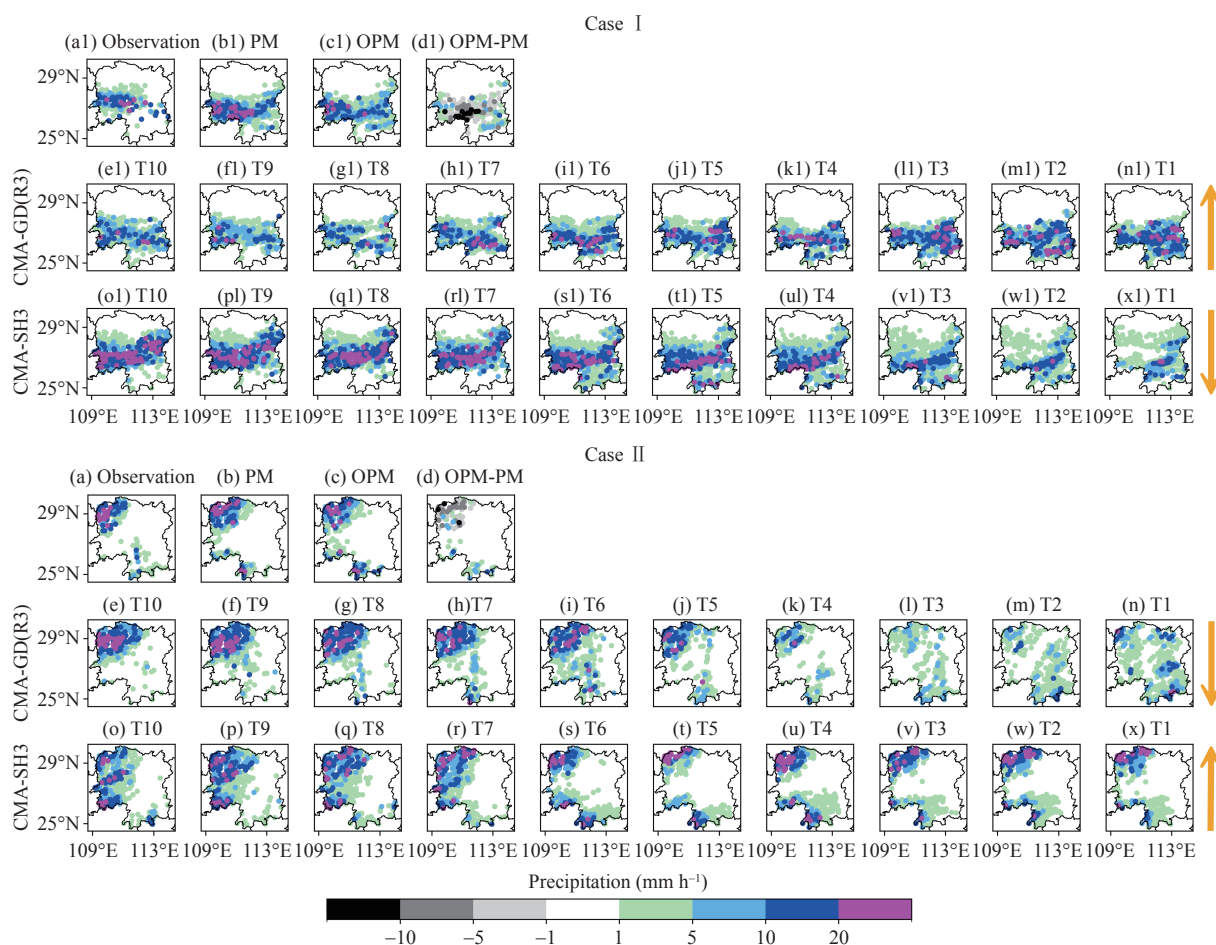


Figure 8. The 0–12 h maximum hourly precipitation valid for two severe convective weather cases ($\geq 20 \text{ mm h}^{-1}$) over Hunan Province, showing (a) observation, (b) PM, (c) OPM20%, (d) difference between OPM20% and PM, and (e–x) each HRRR-TLE member. Case I initiates at 1800 UTC May 6, 2021. Case II initiates at 1200 UTC May 2, 2021. The yellow upward and downward arrows represent strong convective weather that increases and decreases as time approaches.

were the most skilled members at the above thresholds, respectively. With respect to frequency bias, T10(GD) worked best at all thresholds. Therefore, there is not always the best ensemble member for different indicators, but the traditional PM can exceed the overall skill of the system.

In the sensitivity experiments, OPM was effective in improving short-range forecasts on both the distribution and rain rates for the HRRR-TLE system. The results of the OPM100% test illustrated that the dynamic sliding-weight method for ensemble members was effective in improving rainfall spatial representations. The OPM10%–90% tests showed that the optimization method by both cost functions SMAPE and CC worked in selecting skilled members with higher accuracy and can improve predictions. This is because the new approach effectively eliminated the false alarms of redundant precipitation and decreased the number of missing rainstorms, which means that the optimized HRRR-TLE members were closer to the real prediction. Moreover, OPM showed good stability in both daytime and nighttime periods.

In particular, in the $20+ \text{ mm h}^{-1}$ convective precipitation forecasts, which are often difficult to

predict, SMAPE worked better than CC, and only those with a ratio below 50% could improve ETS and BE skills at the same time. Among them, OPM20% with four optimized members performed the best, with a 64.6% and 5.7% improvement in ETS and BE, respectively, relative to PM.

Data Availability Statement: Data and code for this work is available in Zenodo (Liu ^[35]).

REFERENCES

- [1] PALMER T N. Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction [J]. Quarterly Journal of the Royal Meteorological Society, 2012, 138(665): 841–861, <https://doi.org/10.1002/qj.1923>
- [2] BAUER P, THORPE A, BRUNET G. The quiet revolution of numerical weather prediction [J]. Nature, 2015, 525(7567): 47–55, <https://doi.org/10.1038/nature14956>
- [3] SWINBANK R, KYOUDA M, BUCHANAN P, et al. The TIGGE project and its achievements [J]. Bulletin of the American Meteorological Society, 2016, 97(1): 49–67, <https://doi.org/10.1175/BAMS-D-13-00191.1>
- [4] POWERS J G, KLEMP J B, SKAMAROCK W C, et al. The

- weather research and forecasting model: Overview, system efforts, and future directions [J]. *Bulletin of the American Meteorological Society*, 2017, 98(8): 1717–1737, <https://doi.org/10.1175/BAMS-D-15-00308.1>
- [5] WONG M, ROMINE G, SNYDER C. Model improvement via systematic investigation of physics tendencies [J]. *Monthly Weather Review*, 2020, 148(2): 671–688, <https://doi.org/10.1175/MWR-D-19-0255.1>
- [6] CAFARO C, WOODHAMS B J, STEIN T H M, et al. Do convection-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical East Africa? [J]. *Weather and Forecasting*, 2021, 36(2): 697–716, <https://doi.org/10.1175/WAF-D-20-0172.1>
- [7] WOLFF J K, FOSSELL K R, HARROLD M, et al. Enhancing the student experience with numerical weather prediction: Advantages of integrating container and cloud technologies into course curricula [J]. *Bulletin of the American Meteorological Society*, 2022, 103(2): E531–E535, <https://doi.org/10.1175/BAMS-D-21-0248.1>
- [8] BENJAMIN S G, JAMES E P, HU M, et al. Stratiform cloud-hydrometeor assimilation for HRRR and RAP Model short-range weather prediction [J]. *Monthly Weather Review*, 2021, 149(8): 2673–2694, <https://doi.org/10.1175/MWR-D-20-0319.1>
- [9] BYTHEWAY J L, HUGHES M, CIFELLI R, et al. Demonstrating a probabilistic quantitative precipitation estimate for evaluating precipitation forecasts in complex terrain [J]. *Weather and Forecasting*, 2022, 37(1): 45–64, <https://doi.org/10.1175/WAF-D-21-0074.1>
- [10] SLIVINSKI L C, LIPPI D E, WHITAKER J S, et al. Overlapping Windows in a Global Hourly Data Assimilation System [J]. *Monthly Weather Review*, 2022, 150(6): 1317–1334, <https://doi.org/10.1175/MWR-D-21-0214.1>
- [11] DOWELL D C, ALEXANDER C R, JAMES E P, et al. The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model, Part I: Motivation and system description [J]. *Weather and Forecasting*, 2022, 37(8): 1371–1395, <https://doi.org/10.1175/WAF-D-21-0151.1>
- [12] JAMES E P, ALEXANDER C R, DOWELL D C, et al. The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model, Part II: Forecast performance [J]. *Weather and Forecasting*, 2022, 37(8): 1397–1417, <https://doi.org/10.1175/WAF-D-21-0130.1>
- [13] LI X R, FAN K, YU E T. A heavy rainfall event in autumn over Beijing-atmospheric circulation background and hindcast simulation using WRF [J]. *Journal of Meteorological Research*, 2018, 32(3): 503–515.
- [14] CHEN Z T, DAI G F, WU K X, et al. Development of 1km-scale operational model in South China [J]. *Journal of Tropical Meteorology*, 2021, 27(4): 319–329, <https://doi.org/10.46267/j.1006-8775.2021.028>
- [15] LIU J Q, LI Z L, WANG Q Q. Quantitative precipitation forecasting using an improved probability-matching method and its application to a typhoon event [J]. *Atmosphere*, 2021, 12(10): 1346, <https://doi.org/10.3390/atmos12101346>
- [16] TURNER D D, CUTLER H, SHIELDS M, et al. Evaluating the Economic Impacts of Improvements to the High-Resolution Rapid Refresh (HRRR) Numerical Weather Prediction Model [J]. *Bulletin of the American Meteorological Society*, 2022, 103(2): E198–E211, <https://doi.org/10.1175/BAMS-D-20-0099.1>
- [17] PINTO J O, GRIM J A, STEINER M. Assessment of the High-Resolution Rapid Refresh model's ability to predict mesoscale convective systems using object-based evaluation [J]. *Weather and Forecasting*, 2015, 30(4): 892–913, <https://doi.org/10.1175/WAF-D-14-00118.1>
- [18] DOUGHERTY K J, HOREL J D, NACHAMKIN J E. Forecast skill for California heavy precipitation periods from the High-Resolution Rapid Refresh model and the Coupled Ocean–Atmosphere Mesoscale Prediction System [J]. *Weather and Forecasting*, 2021, 36(6): 2275–2288, <https://doi.org/10.1175/WAF-D-20-0182.1>
- [19] SHORT C J, PETCH J. Reducing the spin-up of a regional NWP system without data assimilation [J]. *Quarterly Journal of the Royal Meteorological Society*, 2022, 148(745): 1623–1643, <https://doi.org/10.1002/qj.4268>
- [20] WARNER J L, PETCH J, SHORT C J, et al. Assessing the impact of a NWP warm-start system on model spin-up over tropical Africa [J]. *Quarterly Journal of the Royal Meteorological Society*, 2023, 149(751): 621–636, <https://doi.org/10.1002/qj.4429>
- [21] KHATIWALA S. Fast Spin-up of geochemical tracers in ocean circulation and climate models [J]. *Journal of Advances in Modeling Earth Systems*, 2023, 15(2): e2022MS003447, <https://doi.org/10.1029/2022MS003447>
- [22] JIE W H, WU T W, WANG J, et al. Improvement of 6–15 day precipitation forecasts using a time-lagged ensemble method [J]. *Advances in Atmospheric Sciences*, 2014, 31(2): 293–304, <https://doi.org/10.1007/s00376-013-3037-8>
- [23] RAYNAUD L, BOUTTIER F. The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts [J]. *Quarterly Journal of the Royal Meteorological Society*, 2017, 143(709): 3037–3047, <https://doi.org/10.1002/qj.3159>
- [24] DELSOLE T, TRENARY L, TIPPETT M K. The weighted-average lagged ensemble [J]. *Journal of Advances in Modeling Earth Systems*, 2017, 9(7): 2739–2752, <https://doi.org/10.1002/2017MS001128>
- [25] LU C G, YUAN H L, SCHWARTZ B E, et al. Short-range numerical weather prediction using time-lagged ensembles [J]. *Weather and forecasting*, 2007, 22(3): 580–595, <https://doi.org/10.1175/WAF999.1>
- [26] PORSON A N, CARR J M, HAGELIN S, et al. Recent upgrades to the Met Office convective-scale ensemble: An hourly time-lagged 5-day ensemble [J]. *Quarterly Journal of the Royal Meteorological Society*, 2020, 146(732): 3245–3265, <https://doi.org/10.1002/qj.3844>
- [27] RAYNAUD L, PANNEKOUCKE O, ARBOGAST P, et al. Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective scale [J]. *Quarterly Journal of the Royal Meteorological Society*, 2015, 141(687): 459–468, <https://doi.org/10.1002/qj.2366>
- [28] ZHONG S X, CHEN Z T, XU D S, et al. A review on GRAPES-TMM operational model system at Guangzhou Regional Meteorological Center [J]. *Journal of Tropical Meteorology*, 2020, 26(4): 495–504, <https://doi.org/10.46267/j.1006-8775.2020.043>
- [29] CHEN Z T, DAI G F, WU K X, et al. Development of 1km-Scale Operational Model in South China [J]. *Journal of Tropical Meteorology*, 2021, 27(4): 319–329, <https://doi.org/10.46267/j.1006-8775.2021.028>

- org/10.46267/j.1006-8775.2021.028
- [30] ZHANG X, YANG Y H, CHEN B D, et al. Operational precipitation forecast over China Using the Weather Research and Forecasting (WRF) Model at a gray-zone resolution: Impact of convection parameterization [J]. *Weather and Forecasting*, 2021, 36(3): 915–928, <https://doi.org/10.1175/WAF-D-20-0210.1>
- [31] BRISSON E, DEMUZERE M, VAN LIPZIG N P. Modelling strategies for performing convection-permitting climate simulations [J]. *METEOROLOGISCHE ZEITSCHRIFT*, 2016, 25: 149–163, <https://doi.org/10.1127/metz/2015/0598>
- [32] CHEN C, TWYCROSS J, GARIBALDI J M. A new accuracy measure based on bounded relative error for time series forecasting [J]. *PLoS ONE*, 2017, 12(3): e0174202, <https://doi.org/10.1371/journal.pone.0174202>
- [33] ARMSTRONG J S. Long-Range Forecasting: From Crystal Ball to Computer [M]. New Jersey: A Wiley-Interscience Publication, 1985: 346–354.
- [34] ROSSA A, NURMI P, EBERT E. Precipitation: Advances in Measurement, Estimation and Prediction [M]. Berlin/Heidelberg: Springer Berlin Heidelberg, 2008: 419–452.
- [35] LIU J Q. 0–12 h QPFs of HRRR-TLE using optimized probability-matching method: Taking Hunan Province as an example [EB/OL]. [2024-10-28], <https://zenodo.org/records/10852686>

Citation: LIU Jin-qing, MAO Zi-yi, DAI Guang-feng, et al. 0–12 Hour QPFs of HRRR-TLE Using Optimized Probability-Matching Method: Taking Hunan Province as an Example [J]. *Journal of Tropical Meteorology*, 2024, 30(4): 361–372, <https://doi.org/10.3724/j.1006-8775.2024.032>