SCIENCE CHINA
Earth Sciences

•RESEARCH PAPER•

https://doi.org/10.1007/s11430-022-9961-7

Predictability and skill of convection-permitting ensemble forecast systems in predicting the record-breaking "21·7" extreme rainfall event in Henan Province, China

Kefeng ZHU[1], Chenyue ZHANG[2], Ming XUE[2,3*] & Nan YANG[2]

[1] *Key Laboratory of Transportation Meteorology of China Meteorological Administration, Nanjing Joint Institute for Atmospheric Sciences, Nanjing 210041, China;*

[2] *Key Laboratory of Mesoscale Severe Weather/Ministry of Education and School of Atmospheric Sciences, Nanjing University, Nanjing 210093, China;*

[3] *Center for Analysis and Prediction of Storms and School of Meteorology University of Oklahoma, Norman Oklahoma 73072, USA*

Received February 24, 2022; revised May 11, 2022; accepted June 8, 2022; published online August 9, 2022

Abstract During 19–21 July 2021, an extreme rainfall event occurred in Henan Province, China, during which a record-breaking maximum hourly rainfall of 201.9 mm was recorded in Zhengzhou at 09 UTC July 20. In this study, the predictability of this extreme rainfall event is investigated using two convection-permitting ensemble forecast systems (CEFSs): one initialized from NCEP GEFS (named CEFS_GEFS) and the other initialized from time-lagged ERA5 data (named CEFS_ERA). Both are able to reproduce the daily heavy rainfall along the Taihang Mountains, but most members have significant position biases for the extreme rainfall in Zhengzhou. For the hourly rainfall, a few members are able to capture the evolution and propagation of extreme rainfall. However, all ensemble members underestimate the extreme hourly rainfall and have position errors of a few tens to a few hundreds of kilometers. Such results suggest that the predictability of the extreme hourly rainfall at the accuracy of city scale in Zhengzhou is low, especially by deterministic forecasting models, and the occurrence of the extreme requires that many favorable conditions to happen simultaneously. In terms of the Brier score, CEFS_GEFS performs better than CEFS_ERA. The latter lacks spread, especially in regions with scarce rain, resulting in less dispersion in precipitation distributions and larger probability forecast error. When a neighborhood is applied, the probability of precipitation (POP) is significantly increased over Zhengzhou. While the traditional POP shows almost no skill for hourly rainfall ≥ 25 mm h^{-1}, the neighborhood POP significantly improves the forecast skill score, for both daily and hourly rainfall, suggesting higher predictability when spatial error among the ensemble members is allowed.

Keywords Convective-permitting ensemble forecasts, Neighborhood precipitation probability, Extreme rainfall

Citation: Zhu K, Zhang C, Xue M, Yang N. 2022. Predictability and skill of convection-permitting ensemble forecast systems in predicting the record-breaking "21·7" extreme rainfall event in Henan Province, China. Science China Earth Sciences, 65, https://doi.org/10.1007/s11430-022-9961-7

1. Introduction

During 19–21 July 2021, a once-in-a-millennium rainfall event occurred in Henan Province, China. The daily rainfall reached as much as 680 mm from 12 UTC 19 to 12 UTC 20 July 2021. As a comparison, the average annual rainfall amount in Zhengzhou, the capital city of Henan Province, is 640.8 mm. The downpour was the heaviest on an hourly basis over mainland China since records began in 1951. Zhengzhou received 201.9 mm h^{-1} at 09 UTC 20 July 2021 (Yin et al., 2021). The total number of people recorded as dead or missing was 398, including 14 passengers trapped in a flooded subway (Su et al., 2021), and the economic losses

* Corresponding author (email: mxue@ou.edu)

reached nearly twenty billion dollars. Ahead of this event, the Zhengzhou Meteorological Bureau issued a red rainstorm warning on the night of 19 July 2021, which is the highest level. However, the actual amount of extreme hourly rainfall was not expected by forecasters.

Operational global NWP models predicted the general rain pattern well, but its rainfall center was located to the northwest of the observed position and its intensity was greatly underestimated (Shi et al., 2021). The maximum daily rainfall predicted by ECMWF and NCEP GFS was 317 and 161 mm, respectively. Convection-permitting resolution (CPR) models greatly improve intensity forecasts. The operational Precision Weather Analysis and Forecasting System running at Jiangsu Meteorological Bureau predicted daily maximum rainfall of 518 mm. The maximum hourly rainfall reached as much as 140 mm h^{-1} (Shi et al., 2021). In Yin et al. (2021), after much tuning, the maximum hourly rainfall simulated by the 1 km resolution WRF model reached as much as 233 mm h^{-1}, which was even greater than the observed maximum. Moreover, the predicted rainfall center was close to Zhengzhou, with a position error of only a few tens of kilometers. Also, the time of the simulated hourly rainfall peak was only one hour earlier than observed —a much better performance than the operational global and CPR forecasts. The frustration is that their model settings cannot be fulfilled in real-time; NCEP-FNL data were used for the initial and boundary conditions (ICBCs). Also, in order to reduce the large-scale forecast error, Four-Dimension Data Assimilation was used to assimilate the analysis field throughout the integration. In general, CPR modeling strengthens confidence in successfully predicting extreme rainfall of this type, albeit with large amounts and position of uncertainty.

Convection-permitting/convection-resolving forecast systems (CEFS), which are based on CPR modeling, provide an effective way to estimate the uncertainty for a given forecast and deliver probabilistic guidance for the desired forecast element. Forecasters can identify a 'most likely' forecast from an ensemble forecast and quantify the uncertainty associated with that forecast (Evans et al., 2014). With the rapid growth in computing power, the CEFS approach has gained popularity and has been used for forecast guidance in real-time and operational setting (Clark et al., 2012). Center for Analysis and Prediction of Storms (CAPS) developed the first real-time, large-domain, multi-physics CEFS as part of the NOAA Hazardous Weather Testbed Spring Forecast Experiment in 2007 (Xue et al., 2007; Kong et al., 2008). Some examples for operational CEFSs are the 2.8 km COSMO-DE-EPS at Deutscher Wetterdienst (Peralta et al., 2012), the 2.2 km MOGR-EPS at the UK Met Office (Golding et al., 2016), the 2.5 km AROME-EPS at Météo-France (Nuissier et al., 2016), and the 4 km time-lagged NCASE at NCEP (Du et al., 2014). The progress in operational CEFSs and their model perturbation methods were well summarized in Wang and Shen (2019).

Many studies have already shown that a CEFS can provide superior forecast guidance for predicting severe weather events, as compared with deterministic forecasts, over the continental U.S. (Clark et al., 2009; Schumacher et al., 2013; Loken et al., 2017; Schwartz et al., 2017). Over China, the horizontal resolution of real-time regional ensemble forecasts is ~10 km (Wang et al., 2021), and studies using CEFSs are relatively fewer in number. Of those studies that have been conducted, Zhu and Xue (2016) used CEFSs to study the predictability of an extreme rainfall event that took place on 21 July 2012 in Beijing. With a high probability of precipitation (POP) over the Beijing area, CEFSs have greatly enhanced forecasters' confidence in predicting the occurrence of very heavy rainfall in Beijing. Some good members can reproduce the warm-sector heavy rainfall, which has tended to be missed by most operational deterministic forecasts (Zhang et al., 2013). Wu et al. (2020) used a CEFS to investigate the predictability of a warm-sector heavy rainfall event over southern China, and demonstrated an intrinsic predictability limit for this type of rainfall. Accurate prediction of small-scale convection that leads to heavy rainfall remains a great challenge (Fritsch and Carbone 2004). In this study, the predictability of the so-called "21·7" Henan extreme rainfall event is investigated using CEFSs. We pay particular attention to whether the CEFS members can capture the location and timing of the precipitation system associated with the historical hourly rainfall of this event.

Insufficient ensemble dispersion might be one of the reasons for the limited predictability of some extreme rainfall events by CEFSs (Novak et al., 2008). Many methods have been developed to address this under-dispersion in CEFSs, including perturbing the ICBCs (Stensrud et al., 1999; Gebhardt et al., 2011) and using different combinations of multi-physics schemes (Gallus and Bresch, 2006; Clark et al., 2010) or even multiple parameters (Yussouf and Stensrud, 2011). An important aspect is to understand how the ensemble spread grows within CEFSs (Romine et al., 2014; Schwartz et al., 2014; Loken et al., 2019). Previous studies have investigated the ensemble spread in CEFSs using a large domain (covering the main precipitation area and also a large area without precipitation), which is helpful for understanding the uncertainty in the prediction of synoptic precipitation systems. However, it may not help when it comes to an understanding the variation in small-scale precipitation systems that lead to extreme rainfall. How the spread grows within the region of high-impact weather, and the roles of different ICBC perturbation methods, remain unclear. In this study, we compare two commonly used initial perturbation methods—namely, downscaled and time-lagged ICBCs. Also, to understand the role of physics schemes in the growth of ensemble spread, two small regions are se-

lected: one that covers the main precipitation region of this extreme rainfall event, and another that was less affected by the precipitation. The spread in those two regions is investigated and compared.

Additionally, to improve the predictability of CEFSs, we introduce a neighborhood POP method. This is because, for this extreme rainfall event, most of the CEFS members show large biases in the location and timing of maximum hourly rainfall. The traditional POP product, which is based on a point-to-point algorithm, shows a very low probability of hourly heavy rainfall in Zhengzhou, resulting in almost no forecast skill. For current CPR models, spatial and temporal errors are inevitable. Therefore, a neighborhood method, which has been used in the prediction of tornado occurrence (Sobash et al., 2011), is applied to the precipitation field. The practical predictability and skill of such a neighborhood approach is documented and evaluated.

2. Case overview

Figure 1 shows the synoptic weather charts from 00 UTC 20 July to 12 UTC 21 July with 12 hour intervals. The location of Henan Province is indicated by the blue rectangle labeled "L" in Figure 1a, and its capital, Zhengzhou, is marked by the black circle. The Taihang Mountains run along the boundary of Henan and Shanxi provinces. Zhengzhou is located about a hundred kilometers southeast of the Taihang Mountains.

During the period of extreme rainfall, the center of the western Pacific subtropical high was located around 35°N, which is north of its climatological position (usually between 25°N and 30°N in July). Meanwhile, typhoon In-Fa was located over the East China Sea, and typhoon Cempaka was making landfall in Guangdong Province in southern China. It can be seen that two channels of water vapor transportation into Zhengzhou were established (Figure 1b): one associated with the southeasterly flows and strengthened by typhoon In-fa as it moved landward, which brought water vapor from the East China Sea; and the other associated with the southerly flows that enhanced as typhoon Cempaka made landfall in Guangdong Province, which transported water vapor through southern China from the South China Sea. A similar water transportation pattern was found in association with the Beijing "7·21" extreme rainfall event of the year 2012 (Zhu and Xue, 2016). The synoptic weather pattern and Taihang Mountains created favorable conditions for the formation of extreme rainfall in Zhengzhou. One can see that water vapor was flowing to Zhengzhou during 20 July, resulting in this record-breaking extreme rainfall event. The next day, as typhoon In-Fa moved slowly northwest, the water vapor concentrated at the foot of the Taihang Mountains (Figure 1c, 1d), and so did the rainfall center.

3. Experimental design and verification methods

3.1 CEFS configurations

The CEFSs were designed by employing WRF version 3.9.1 (Skamarock et al., 2005). The forecast domain had 1150×800×50 grid points with 4 km grid spacing, covering typhoon In-Fa in the East China Sea and Cempaka in the South China Sea. The center of the simulation domain was set to Zhengzhou.

Two CEFS experiments were conducted. The first was downscaled from the coarser-resolution NCEP Global Ensemble Forecast System (GEFS, Wei et al., 2008; Zhu et al., 2018), which has 30 members and a gridded horizontal resolution of 0.5° and 31 vertical levels. For more detailed information and the settings of GEFS, readers are referred to https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gefs.php. We refer to this experiment as CEFS_GEFS. The use of ICBC uncertainty alone may not be able to produce sufficient spread for CEFSs (Romine et al., 2014; Schwartz et al., 2014). Given this, representations of uncertainty in model physical processes were considered using multi-physics schemes. Table 1 lists the physics configurations of CEFS_GEFS, which followed those of CAPS CEFS (Kong et al., 2007; Xue et al., 2007) and the configurations used for the Beijing "7·21" extreme rainfall event (Zhu and Xue, 2016), but with some differences. Here, we tried to make the physics configurations as diverse as possible. For this experiment, all the ensemble members started from 1200 UTC 19 July 2021 and were integrated for 48 hours.

The other set of CEFS experiments was performed using ERA5 data as the ICBCs. The primary purpose was to see whether the forecast would be improved by using more accurate ICBCs. ERA5 is the successor to ERA-Interim (Hersbach et al., 2020). It has been used in many studies as the observational reference to evaluate the atmospheric circulation simulated by CPR models (Li et al., 2020; Cai et al., 2021). The horizontal grid resolution of upper-air variables is 0.25°, while it is 0.125° for the surface variables. There are 37 vertical levels. ERA5 is produced at hourly intervals and is available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. The expectation is that the CEFS_ERA experiment receives better ICBC from ERA5, whereas the CEFS_GEFS experiment uses only the GEFS forecast but not the analysis for its ICBC. However, unified ICBCs may cause under-dispersion in CEFSs. Therefore, we used a time-lagged method together with multi-physics configurations to increase the ensemble spread. An extended goal of this experiment was to see whether the time-lagged method could help increase the ensemble dispersion. We refer to this experiment as CEFS_ERA, and the details of the settings are presented in Table 2. CEFS_ERA has 20 en-

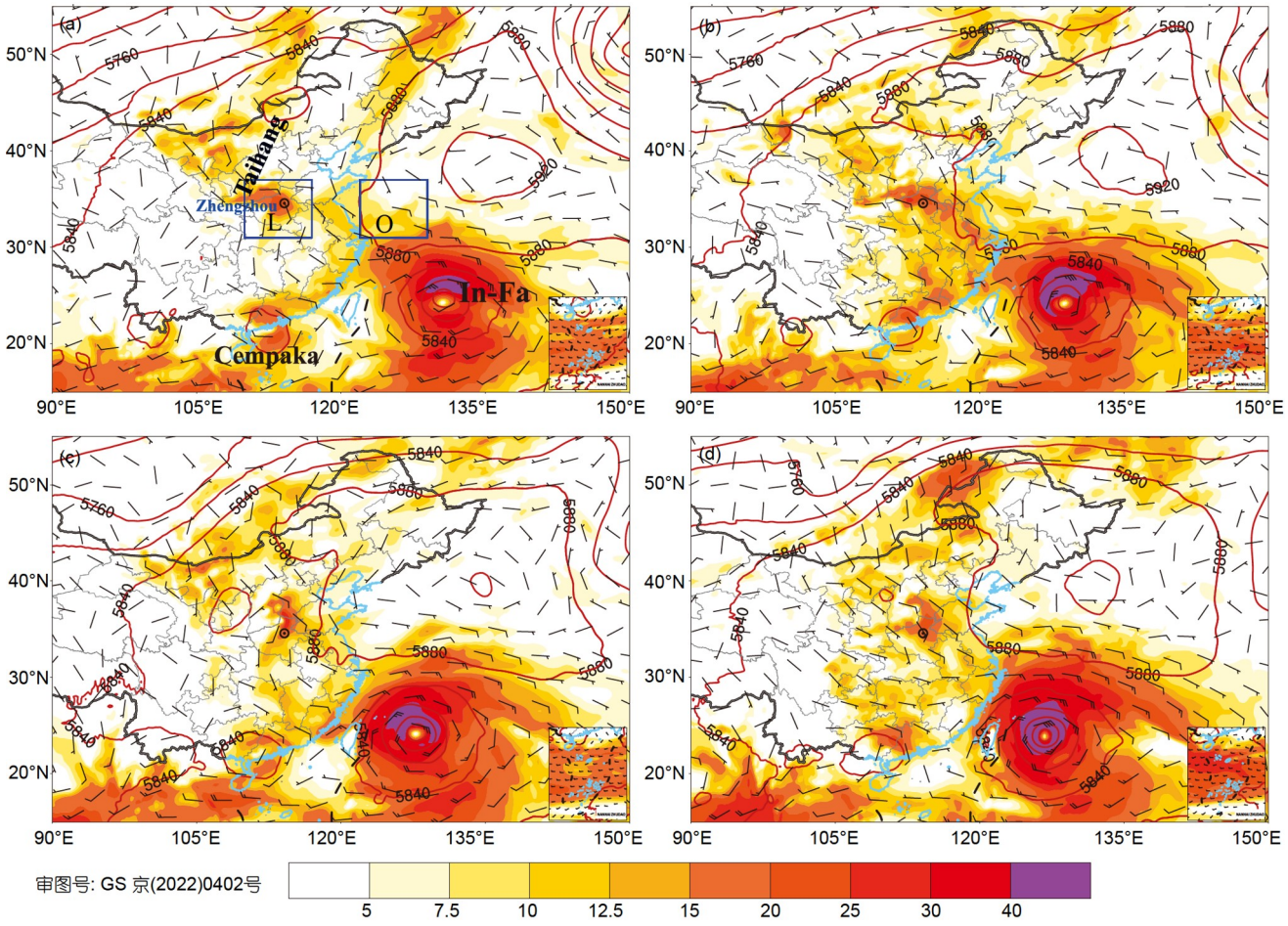

Figure 1 Synoptic weather chart of ERA5 reanalysis for the 500 hPa geopotential height (brown contours), 850 hPa winds (barbs) and horizontal water vapor flux (colored shading) at (a) 0000 UTC 20 July, (b) 1200 UTC 20 July, (c) 0000 UTC 21 July, and (d) 1200 UTC 21 July 2012. The black circle indicates the location of Zhengzhou; typhoon In-Fa is located over the East China Sea; and typhoon Cempaka is making landfall in South China. Here, "L" in (a) is the verification domain over land which covers the main precipitation in Henan. "O" is the verification domain over ocean with the same area as "L", but has scarce rainfall.

semble members. We have also tested 30 ensemble members. In terms of the Brier score, the current setting with 20 members is slightly better than that with 30 members. The experiment of CEFS_ERA could be divided into three groups. For the first group (members 1 to 10), the forecasts started at 1200 UTC 19 July as CEFS_GEFS. To increase the ensemble spread, we used as many different physics schemes as possible. Then, the latter two groups used time-lagged initials. The forecasts of the second group (members 11 to 15) started 3 hours earlier (that is, at 0900 UTC 19 July), while the forecasts of the third group (members 16 to 20) started 3 hours later (that is, at 1500 UTC 19 July). All the forecasts ended at 1200 UTC 21 July 2021. The physics configurations of the two groups were set to the same as for members 1 to 5 of the first group. Similar to the study of Zhu and Xue (2016), we do not claim that the current configurations of the above experiments are optimal—an important purpose is to evaluate the performance of the CEFS as configured.

3.2 Temporal- and spatial-neighborhood POP

For the observed extreme rainfall amount, as presented later, the overall predictability is low, especially in Zhengzhou. In this study, a neighborhood POP method, which has been used previously for tornado probability forecasting, is suggested and compared with the traditional POP approach. The new algorithm is described as follows. For a given threshold P_0, the traditional POP of a grid point (i, j) is computed as

$$p_{i,j} = \sum_{l=1}^{l=N} \text{where}\left(x_{l,i,j} \geq P_0, 1, 0\right) / N. \quad (1)$$

Here, the function "where" represents if the condition is true, the value is set to 1, and otherwise it is set to 0; l represents the lth member; and N is the total number of ensemble members.

For the spatial-neighborhood method, the POP is redefined within a neighborhood square box with radius R_g in terms of the number of grid points. In practical, R_g could be inter-

Table 1 Parameter settings of CEFS_GEFS[a)]

Case name	ICs and BCs	Microphysics	PBL	Surface layer	Radiation scheme (long)	Radiation scheme (short)
mem01	GEFS01	8, Thompson	1, YSU	1, Revised MM5	1, rrtm	1, Dudhia
mem02	GEFS02	8, Thompson	2, MYJ	2, Janjic Eta	4, rrtmg	4, rrtmg
mem03	GEFS03	10, Morrison	7, ACM2	1, MM5	3, CAM	3, CAM
mem04	GEFS04	8, Thompson	2, MYJ	2, Janjic Eta	5, Goddard	5, Goddard
mem05	GEFS05	10, Morrison	1, YSU	1, Revised MM5	1, rrtm	1, Dudhia
mem06	GEFS06	10, Morrison	1, YSU	1, Revised MM5	5, Goddard	5, Goddard
mem07	GEFS07	10, Morrison	7, ACM2	1, Revised MM5	1, rrtm	1, Dudhia
mem08	GEFS08	10, Morrison	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem09	GEFS09	10, Morrison	5, MYNN2	1, Revised MM5	1, rrtm	1, Dudhia
mem10	GEFS10	10, Morrison	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem11	GEFS11	10, Morrison	4, QNSE-MF	4, QNSE	1, rrtm	1, Dudhia
mem12	GEFS12	10, Morrison	1, YSU	1, Revised MM5	3, CAM	3, CAM
mem13	GEFS13	9, Milbrandt-Yau	1, YSU	1, Revised MM5	1, rrtm	1, Dudhia
mem14	GEFS14	10, Morrison	1, YSU	1, Revised MM5	5, Goddard	5, Goddard
mem15	GEFS15	10, Morrison	5, MYNN2	1, Revised MM5	1, rrtm	1, Dudhia
mem16	GEFS16	8, Thompson	1, YSU	1, Revised MM5	1, rrtm	2, Goddard(old)
mem17	GEFS17	8, Thompson	1, YSU	1, Revised MM5	1, rrtm	1, Dudhia
mem18	GEFS18	10, Morrison	7, ACM2	1, MM5	3, CAM	3, CAM
mem19	GEFS19	8, Thompson	2, MYJ	2, Janjic Eta	5, Goddard	5, Goddard
mem20	GEFS20	10, Morrison	1, YSU	1, Revised MM5	1, rrtm	1, Dudhia
mem21	GEFS21	10, Morrison	1, YSU	1, Revised MM5	5, Goddard	5, Goddard
mem22	GEFS22	10, Morrison	7, ACM2	1, Revised MM5	1, rrtm	1, Dudhia
mem23	GEFS23	10, Morrison	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem24	GEFS24	10, Morrison	5, MYNN2	1, Revised MM5	1, rrtm	1, Dudhia
mem25	GEFS25	10, Morrison	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem26	GEFS26	10, Morrison	4, QNSE-MF	4, QNSE	1, rrtm	1, Dudhia
mem27	GEFS27	10, Morrison	1, YSU	1, Revised MM5	3, CAM	3, CAM
mem28	GEFS28	8, Thompson	1, YSU	1, Revised MM5	1, rrtm	2, Goddard(old)
mem29	GEFS29	8, Thompson	1, YSU	1, Revised MM5	1, rrtm	1, Dudhia
mem30	GEFS30	10, Morrison	5, MYNN2	1, Revised MM5	1, rrtm	1, Dudhia

a) Each member uses different ICs and lateral BCs from GEFS and different WRF physics combinations.

preted as the maximum tolerance for the position error. An extra fraction threshold is needed for determining a "Yes" or "No" event ($n_{l,i,j}$) for a given point in the defined neighborhood square box. Considering the small size for heavy rain, we used a fixed number of grid points F_g as the fraction threshold:

$$n_{l,i,j} = \sum_{k=-R_g}^{k=R_g} \sum_{k=-R_g}^{k=R_g} \text{where}\left(x_{l,i-k,j-k} \geq P_0, 1, 0\right), \quad (2)$$

$$p_{i,j} = \sum_{l=1}^{l=N} \text{where}\left(n_{l,i,j} \geq F_g, 1, 0\right) / N. \quad (3)$$

In this study, R_g was set to 12 grid points, corresponding to a 48 km position bias. F_g is related to the area of precipitation. Generally speaking, the area of accumulated rainfall should be larger than at of hourly rainfall. Here, F_g was set to 20 grid points for the hourly rainfall and 50 grid points for the 24 hour accumulated rainfall. For example, a forecasted "Yes" event within a radius of 12 grid points in the square box requires at least 20 grid points with a hourly precipitation amount larger than P_0. As soon as the "Yes" event is calculated for the lth member, the POP is calculated as per eq. (3). There is room for further refinement on the choice of the parameters, but it is left for future work.

For the temporal- and spatial-neighborhood method, forecast biases are allowed not only in space but also in time–space. The POP is calculated as follows:

$$n_{t,l,i,j} = \sum_{k=-R_g}^{k=R_g} \sum_{k=-R_g}^{k=R_g} \text{where}\left(x_{t,l,i-k,j-k} \geq P_0, 1, 0\right), \quad (4)$$

Table 2 Parameter settings of CEFS_ERA[a)]

Case name	ICs and BCs	Microphysics	PBL	Surface layer	Radiation scheme (long)	Radiation scheme (short)
mem01	ERA5	8, Thompson	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem02	ERA5	8, Thompson	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem03	ERA5	10, Morrison	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem04	ERA5	19, NSSL 1-moment	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem05	ERA5	7, Goddard	1, YSU	1, Revised MM5	5, Goddard	5, Goddard
mem06	ERA5	9, Milbrandt-Yau	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem07	ERA5	16, WDM 6-class	5, MYNN	5, MYNN	5, Goddard	5, Goddard
mem08	ERA5	8, Thompson	2, MYJ	2, Janjic Eta	4, rrtmg	4, rrtmg
mem09	ERA5	10, Morrison	7, ACM2	1, Revised MM5	3, CAM	3, CAM
mem10	ERA5	10, Morrison	1, YSU	1, Revised MM5	5, Goddard	5, Goddard
mem11	ERA5-3h	8, Thompson	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem12	ERA5-3h	8, Thompson	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem13	ERA5-3h	10, Morrison	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem14	ERA5-3h	19, NSSL1-moment	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem15	ERA5-3h	7, Goddard	1, YSU	1, Revised MM5	5, Goddard	5, Goddard
mem16	ERA5+3h	8, Thompson	1, YSU	1, Revised MM5	4, rrtmg	4, rrtmg
mem17	ERA5+3h	8, Thompson	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem18	ERA5+3h	10, Morrison	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem19	ERA5+3h	19, NSSL1-moment	5, MYNN	5, MYNN	4, rrtmg	4, rrtmg
mem20	ERA5+3h	7, Goddard	1, YSU	1, Revised MM5	5, Goddard	5, Goddard

a) Each member uses different WRF physics combinations. The time-lagged method is applied to members 11–20.

$$\mathrm{FY}_{l,i,j} = \sum_{t=-t_w}^{t=t_w} \mathrm{where}\left(n_{t,l,i,j} \geq F_g, 1, 0\right), \tag{5}$$

$$p_{i,j} = \sum_{l=1}^{l=N} \mathrm{where}\left(FY_{l,i,j} \geq 1, 1, 0\right) / N. \tag{6}$$

A forecasted "Yes" event of the lth member ($\mathrm{FY}_{l,i,j}$) is calculated within an extra temporal neighborhood width t_w. Similarly, t_w could be interpreted as the maximum tolerance for the time phase error. Here, t_w is set to 1, that is extra ±1 hour forecast of current time are also used for the calculation of POP. As long as one moment within time windows met the condition of a "Yes" event, the forecast of that grid point was computed as "Yes". In general, the smaller F_g, the larger R_g and t_w, the higher of POP.

3.3 Verification metrics

In this study, the ensemble forecasts were verified against hourly gridded precipitation (Shen et al., 2014; Pan et al., 2015), which is a 0.01°×0.01° merged precipitation analysis product (over 70°–140°E and 15°–60°N). This dataset has been widely used for the evaluation of CPR model performance in China (Wu et al., 2018; Zhao et al., 2020; Cai et al., 2021; Li et al., 2021). For the verification, we employed the Model Evaluation Tools software (Brown et al., 2009), which contains a comprehensive suite of verification metrics for ensemble-based probabilistic forecasts. Verification scores, such as the rank histogram (Hamill 2001), Brier score (Murphy, 1973; Wilks, 2010), relative position (RELP) and receiver operating characteristic (ROC, Marzban 2004), were used to evaluate the ensemble probabilistic forecasting skill. Two domains were chosen for our ensemble forecast verification: one covering Henan Province (31°–37°N, 110°–117°E), labeled as box "L" in Figure 1a; and the other a small domain just covering Zhengzhou (34°–35°N, 112.5°–114.5°E). The former is used to evaluate the overall performance of the ensemble forecast for the main precipitation system over Henan Province and its surrounding area, while the latter is used more for the predictability of the CEFS with respect to the extreme rainfall over Zhengzhou. Additionally, to examine the sensitivity of the spread to the ICBCs and physics schemes, a verification domain labeled "O" in Figure 1a over the ocean (31°–37°N, 122°–129°E), which was less affected by precipitation, was chosen.

4. Ensemble forecast results

4.1 Predictability of 24 hour accumulated rainfall forecasts

Figure 2 shows postage stamp plots for the ensemble forecast of experiment CEFS_GEFS. In general, most ensemble

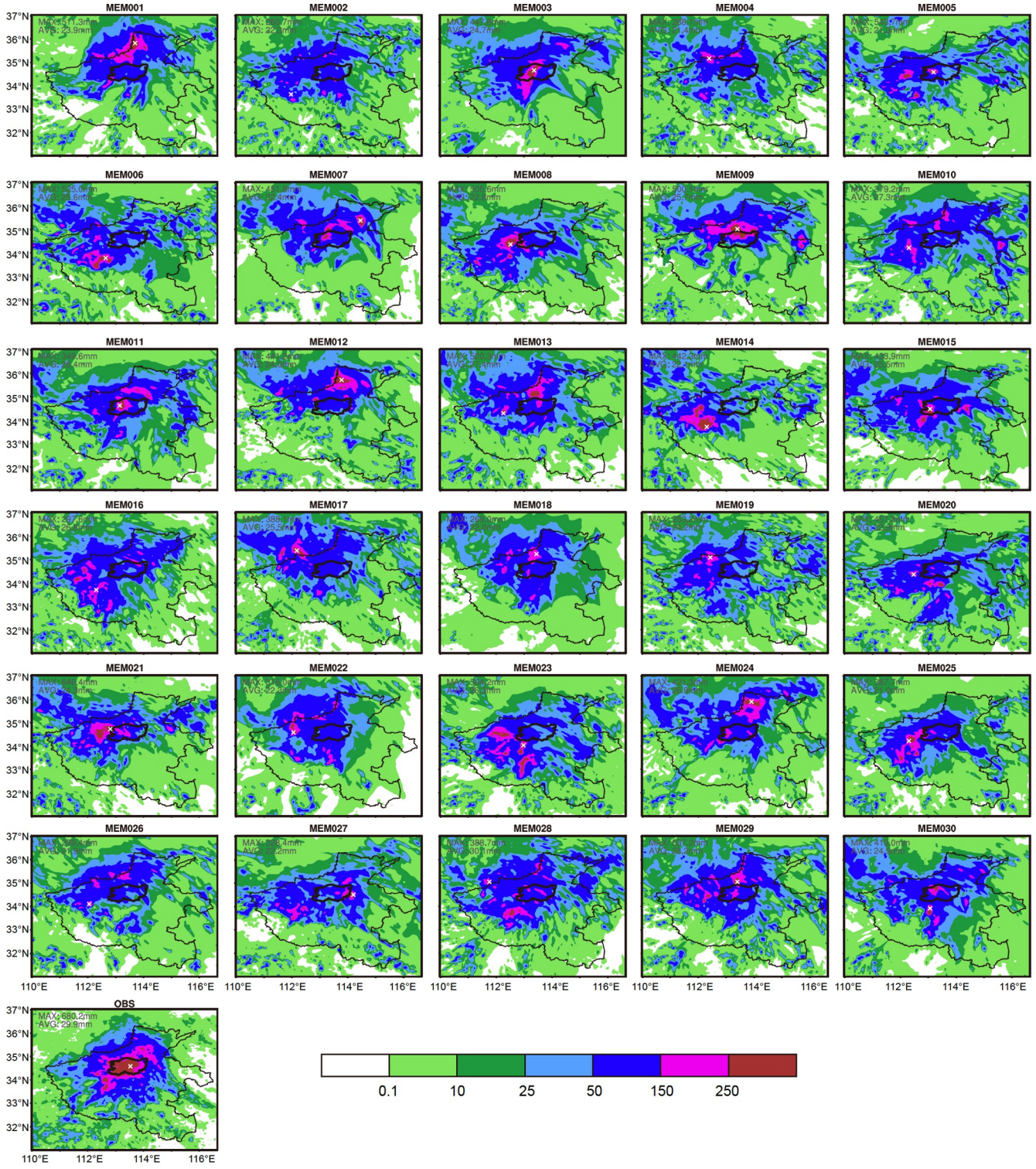

Figure 2 Postage stamp plots of 24 hour accumulated rainfall (unit: mm) from 1200 UTC 20 July to 1200 UTC 21 July 2021 of the 4 km CEFS members initialized from GEFS (see Table 1 for the definition of members) and observation. In this figure, and some subsequent figures, Zhengzhou is indicated by the bold-black border. The white "X" symbol represents the location of maximum rainfall.

members have no difficulty in predicting the heavy rainfall in Henan Province. The predicted rainfall centers (see the white "X" symbols in each plot) are distributed in all directions of Zhengzhou, but mostly along the Taihang Mountains—that is, northwest and west of the observed center. Only a few members (such as members 3, 5, 11 and 15) are able to reproduce the extreme rainfall center in Zhengzhou.

The overall performance of CEFS_ERA is similar to ex-

periment CEFS_GEFS (see Figure 3). It has no difficulty in predicting the heavy rainfall along the Taihang Mountains. The predicted rainfall centers are mostly distributed to the west of Zhengzhou, resulting in higher POP values. This is because time-lagged CEFS_ERA uses almost the same ICBCs, which leads to almost consistent westward location bias of precipitation forecast among most members.

The position bias of each member is much better reflected in the ensemble mean forecast. With extreme points distributed in all directions, the ensemble mean of CEFS_GEFS is able to predict the center of heavy rainfall in Zhengzhou (see Figure 4b). In comparison, the predicted center of the CEFS_ERA ensemble mean is located to the southwest of Zhengzhou (see Figure 4c). For the precipitation amount, CEFS_ERA, with less variability for the precipitation distribution, obtains a higher maximum ensemble mean value than CEFS_GEFS. Unsurprisingly, both CEFSs greatly underestimate the precipitation amount simply in terms of the ensemble mean. The observed maximum is 680 mm, compared to a maximum of 323 mm in the two experiments. The probability-matched (PM; (Kong et al., 2008; Clark et al., 2009) rain was therefore calculated. The PM rain is a deterministic product that can improve the underestimation of rainfall amounts that are typically associated with using a simple ensemble mean. Details of the calculation formula are given in Zhu and Xue (2016). As expected, both CEFSs demonstrate significant improvement in the forecast precipitation amount. The maximum PM rain amounts are 513 and 746 mm for CEFS_GEFS and CEFS_ERA, respectively (see Figure 4d, 4e). For the extreme rainfall event, the use of PM rain is practical.

Figure 4f and 4g shows POP values exceeding 150 mm. For this extreme event, the maximum POP is 50% for CEFS_GEFS, and 100% for CEFS_ERA. The higher POP of CEFS_ERA, as explained previously, is due to its lower variability in terms of the precipitation distribution (see also Figure 3). This will lead to a poorer verification skill score. For both experiments, the highest POP values are along the Taihang Mountains and to the west of Zhengzhou, suggesting low predictability of extreme rainfall in Zhengzhou.

To gain some understanding as to why the CEFSs have difficulty in predicting the extreme rainfall in Zhengzhou, the ensemble forecast products for the next day are shown in Figure 5. There are two observed heavy rainfall centers: one near the Taihang Mountains ("A" in Figure 5a), and another to the southeast of Zhengzhou ("B" in Figure 5a). Both ex-

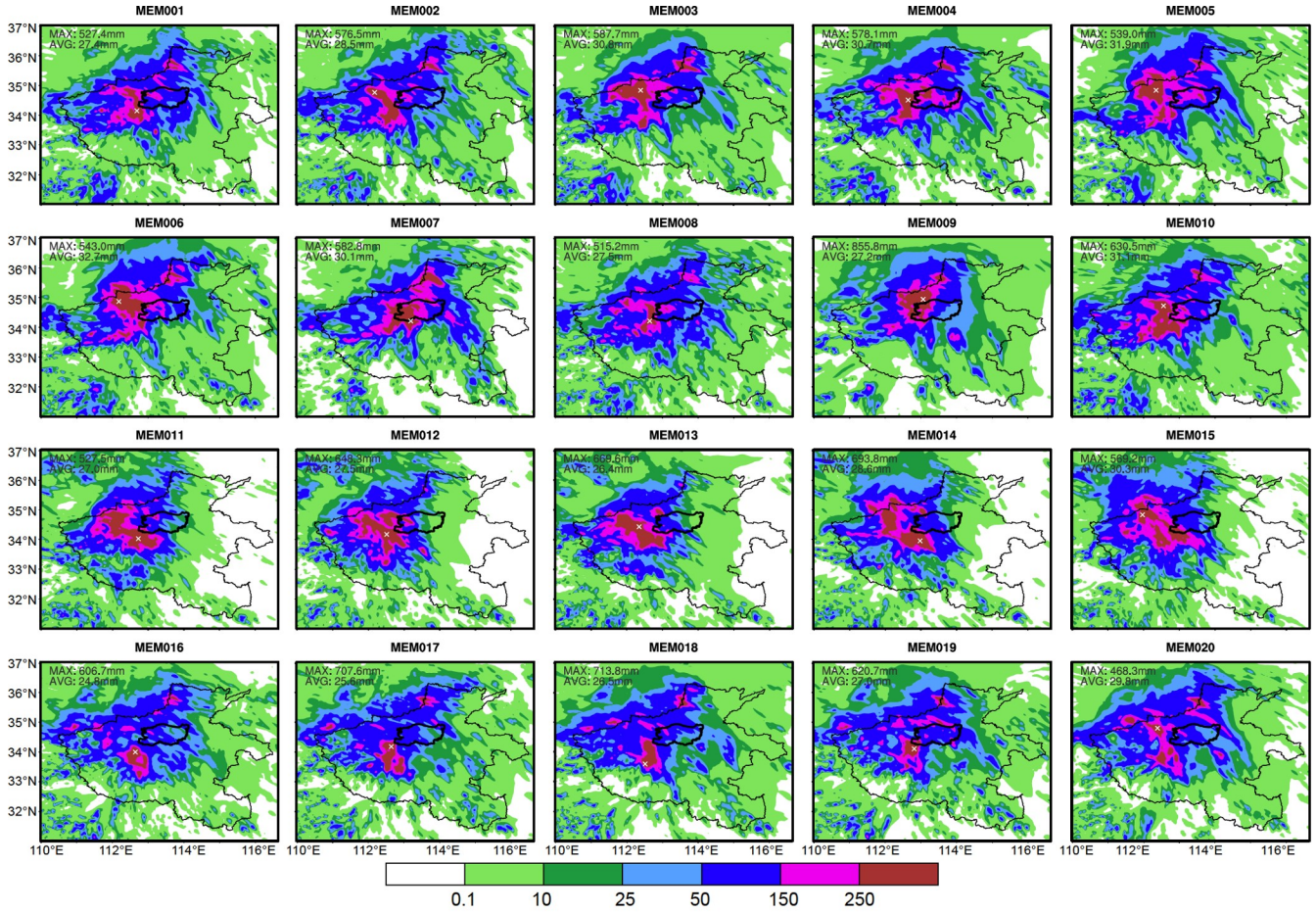

Figure 3 As in Figure 2 but for experiment CEFS_ERA (see Table 2 for the definition of members).

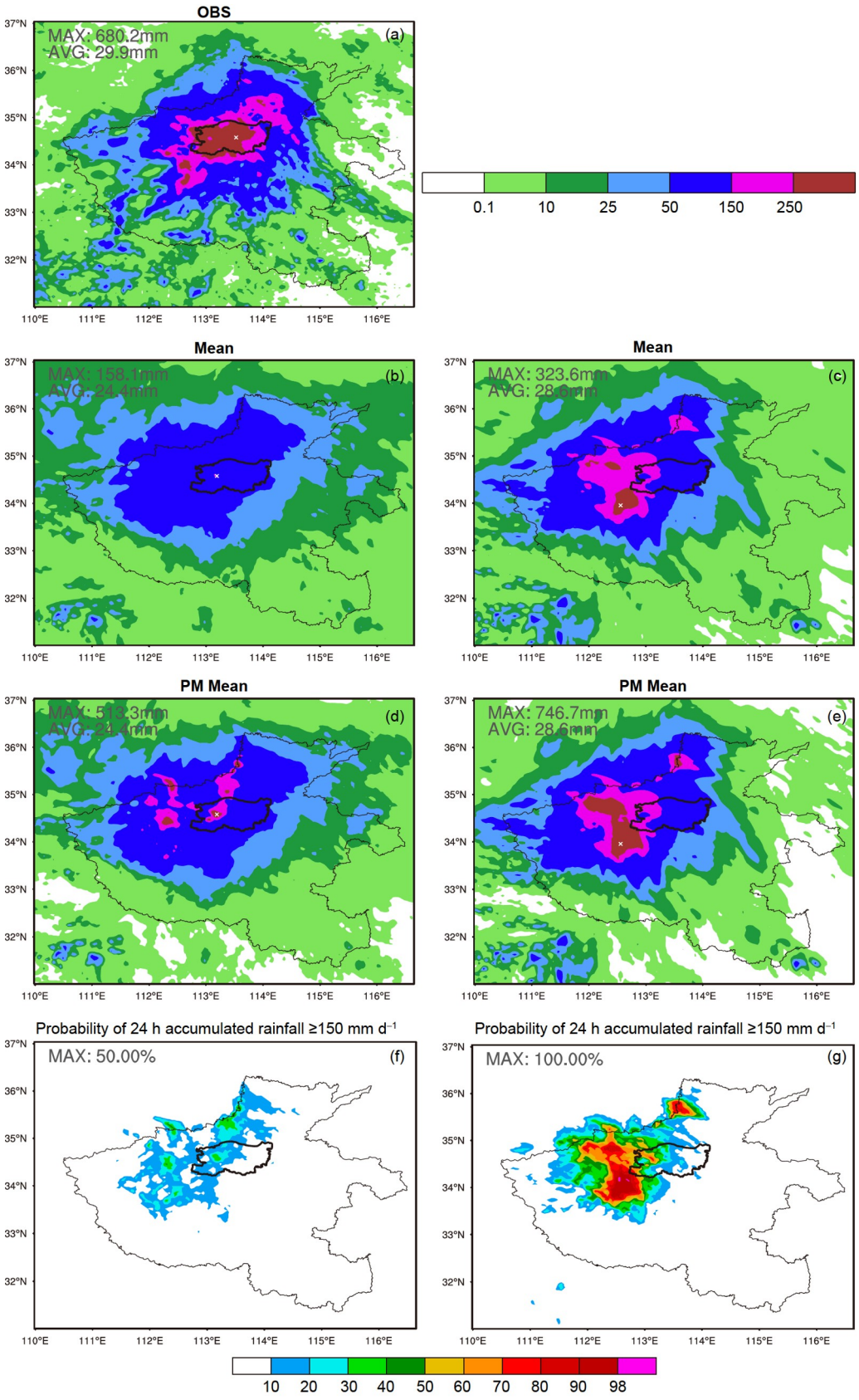

Figure 4 The 24 hour accumulated rainfall (unit: mm) from 1200 UTC 19 July to 1200 UTC 20 July 2021 for the (a) observation, (b) ensemble mean, (d) PM mean, and (f) POP for the threshold of larger than 150 mm. Panels ((b), (d), (f)) are calculated from CEFS_GEFS, while ((c), (e), (g)) are calculated from CEFS_ERA.

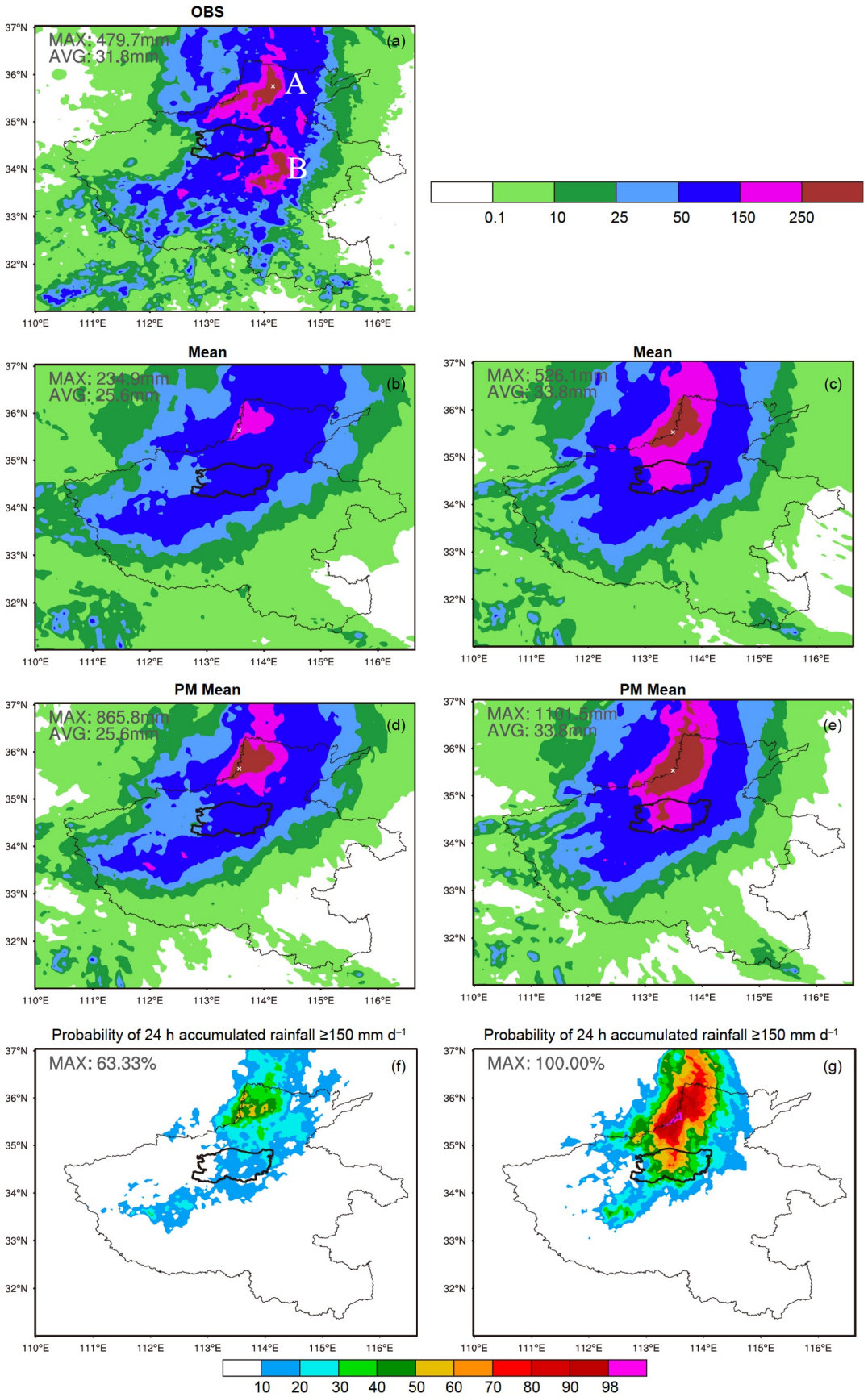

Figure 5 As in Figure 4 but for the period from 1200 UTC 20 July to 1200 UTC 21 July 2021.

periments have no difficulty in predicting heavy rain center A, but show low predictability for B. This is because B is propagated from the previous day precipitation at Zhengzhou. In general, the CPR model is able to simulate synoptic weather systems and external forcings due to terrain blocking and lifting. Therefore, for the precipitation system A along the Taihang Mountains, the CEFS has a greater chance to make a successful forecast. However, the precipitation at Zhengzhou as well as precipitation center B in the next day are due to the low-level convergence of a mesoscale barrier jet and a meso-β-scale vortex (to be discussed in Section 4.2). Predicting the exact timing and location of such mesoscale convergence is challenging for CPR models, hence the low predictability of precipitation at Zhengzhou and B. Forecast biases in time and space, as suggested in Section 5, should be expected in POP.

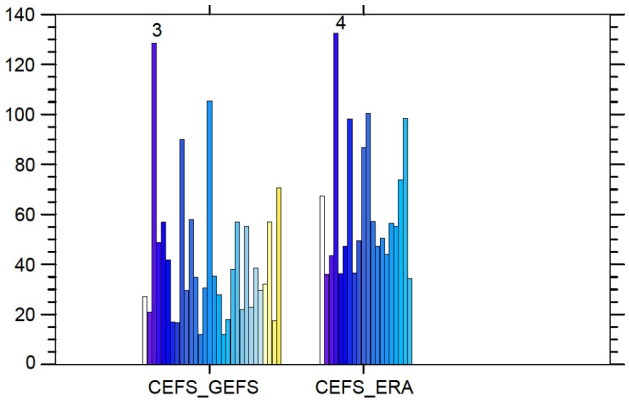

Figure 6 RELP of each ensemble member for the 3 hour accumulated rainfall (from 0700 UTC to 0900 UTC 20 July 2021) for the domain surrounding Zhengzhou. The numbers are the best member in terms of RELP for each CEFS.

4.2 Predictability of the extreme hourly rainfall forecasts

As mentioned above, the observed maximum hourly rainfall amount reached 201.9 mm h^{-1} at 0900 UTC 20 July 2021, which created a new record for hourly rainfall over mainland China. For the extreme hourly rainfall of this event, most CEFS members failed to reproduce its location and amount (not shown). However, to see whether some members could capture the extreme rainfall event, RELP was used to select the best member (see Figure 6). The RELP value represents the member obtaining *N* grid points with the lowest RMSE of all members. The larger the RELP value, the closer the member forecast is to the observation when compared to the rest of the members. The best members of CEFS_GEFS and CEFS_ERA are mem003 and mem004, respectively, using the RELP calculated from the 3 h accumulated rainfall at 0900 UTC 20 July 2021 within Zhengzhou.

Figure 7 shows the hourly rainfall of each best member of CEFS_GEFS and CEFS_ERA. Note that, due to the interpolation (originally from rain gauge stations (201.9 mm) to 1 km hourly gridded observations (169.3 mm), and also from the 1 km grid to a 4 km model grid (133.4 mm)), the maximum hourly gridded observation is lower than the rain gauge station mentioned above. Nonetheless, these intensity biases of a few extreme points do not affect the general conclusions, including the verification results in the following section. The observed extreme rainfall stayed in Zhengzhou from 0800 UTC to 1000 UTC. The best member of CEFS_GEFS is able to predict the extreme rainfall in Zhengzhou, but it is located to the west of the observed extreme point and its intensity is clearly underestimated. The intensity of the best member from CEFS_ERA is improved, but the location of the extreme point is relatively far away from the observed point. Both observations and forecasts show the extreme rainfall to have been formed at the southwest edge of the main precipitation, where the southeast and northeast flow converged. The northeast flow was originated from a mesoscale barrier jet on the eastern slope of the Taihang Mountain due to terrain blocking of large-scale easterly flows. The southeast flow is in association with a low-level meso-β-scale vortex located to the west of Zhengzhou and the large-scale easterly inflows caused by the typhoon In-Fa over the East China Sea (Wei et al., 2022). We also compared the forecasts with observations throughout the two forecasted days. In general, the best members were able to reproduce the general spatial patterns and the evolution of the event (not shown). However, the exact amount and the location of extreme precipitation are still hard to predict, due to the difficulty in the forecasting the location of meso-scale low-level convergence, as well uncertainties in the simulated microphysical processes.

Figure 8 shows the time series of hourly rainfall for domains from the large-scale area to the point of extreme rainfall. For the large domain covering Henan Province (see domain "L" in Figure 1a), most ensemble members are able to reproduce the diurnal variation of precipitation (Figure 8a, 8b), indicating the synodical pattern of precipitation and its evolution are captured well by the ensemble forecasts. The biggest issue is the amount. Most members from CEFS_GEFS underestimate the mean precipitation over domain "L" (Figure 8a), while it is overestimated by most members from CEFS_ERA (Figure 8b). Compared to ERA5, the initials from GEFS underestimate the water vapor content transported to that region (not shown). For the small domain covering Zhengzhou, due to the position biases, the mean rainfall amounts are greatly underestimated during the period of extreme rainfall (18–24 hour forecasts in Figure 8c, d) for both CEFSs. For the point of extreme rainfall, all ensemble forecasts fail to reach the observed maximum (201.9 mm h^{-1}) (Figure 8e, 8f). The predicted hourly maximum of the two CEFSs' members is around 180 mm h^{-1}.

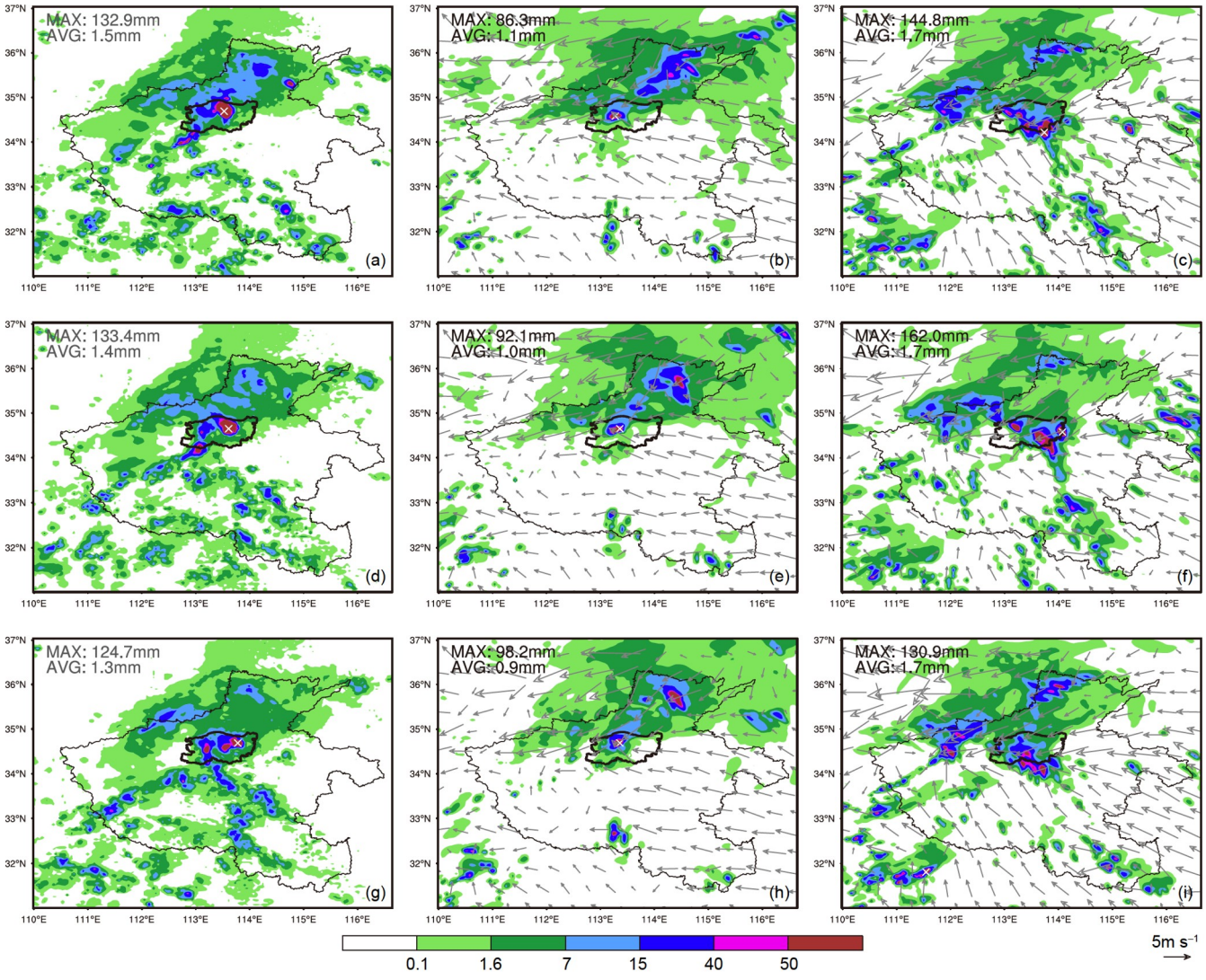

Figure 7 Hourly rainfall (unit: mm) for (a) 0800 UTC, (d) 0900 UTC, and (g) 1000 UTC 21 July 2021. The observed maximum hourly rainfall is at 0900 UTC. The second and third columns are the corresponding forecasts from the best member of CEFS_GEFS (member 003) and CEFS_ERA (member 004), respectively. The observations are interpolated to the 4 km WRF model grid. Therefore, the maximum is not the same as the site observations in Figure 8.

The timing of the maximum is either advanced or delayed from the observed peak time. Figure 8g and 8h show the corresponding accumulated rainfall for the extreme point. The general evolutionary trend is well captured. Most members are able to reach the observed maximum accumulated rainfall amount during the two days of the forecast, if neglecting the timing and position biases. Overall, the CEFSs are able to reproduce the variation in precipitation in the large domain, but have limited forecast skill for the variation and amount in Zhengzhou and at the point of extreme rainfall.

To see whether the ensemble forecast products can provide some useful information, the ensemble mean and POP values at the time of the observed maximum are shown in Figure 9. Unsurprisingly, the simple ensemble mean greatly underestimates the precipitation amount (Figure 9b, 9c). The PM rain improves the forecast precipitation amount, but the heavy rain clusters ($\geq$40 mm h^{-1}) are scattered mostly outside Zhengzhou (Figure 9b, 9c), indicating great uncertainty in the location of heavy rain. The POP values are almost zero in Zhengzhou for a threshold larger than 25 mm h^{-1}. Without consideration of position biases, the traditional ensemble forecast products fail to provide useful information for the extreme rainfall in Zhengzhou at the time of the observed maximum. Therefore, timing and position biases need to be considered for this extreme rainfall event. In this respect, the neighborhood products are presented and discussed in section 5.

4.3 Skill scores of the CEFSs

In this subsection, the skill scores of the two CEFSs are compared. Figure 10 shows the Brier score of the two CEFSs. The Brier score is the mean-squared probability er-

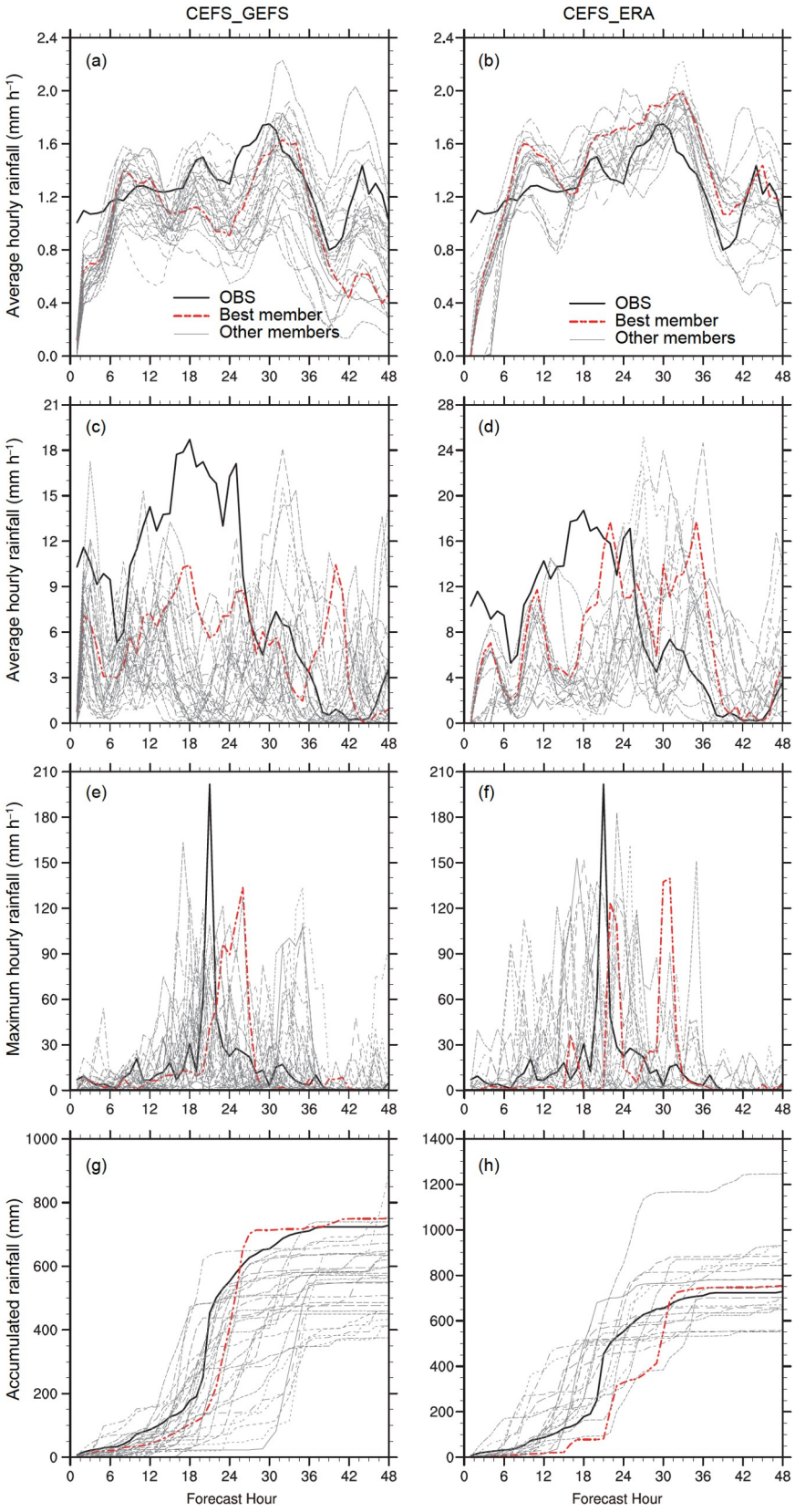

Figure 8 ((a), (b)) Time series of average hourly rainfall in a domain surrounding Henan Province (see domain "L" in Figure 1a for the observation and ensemble members. ((c), (d)) Similar to ((a), (b)) but for a small domain in Zhengzhou. ((e), (f)) Similar to ((a), (b)) but for the point of extreme rainfall within domain "L". ((g), (h)) Time series of observed and predicted accumulated rainfall at the grid point of the largest 24 hour accumulated rainfall (from forecast hours 12–36). The left-hand column is from CEFS_GEFS, while the right-hand column is from CEFS_ERA.

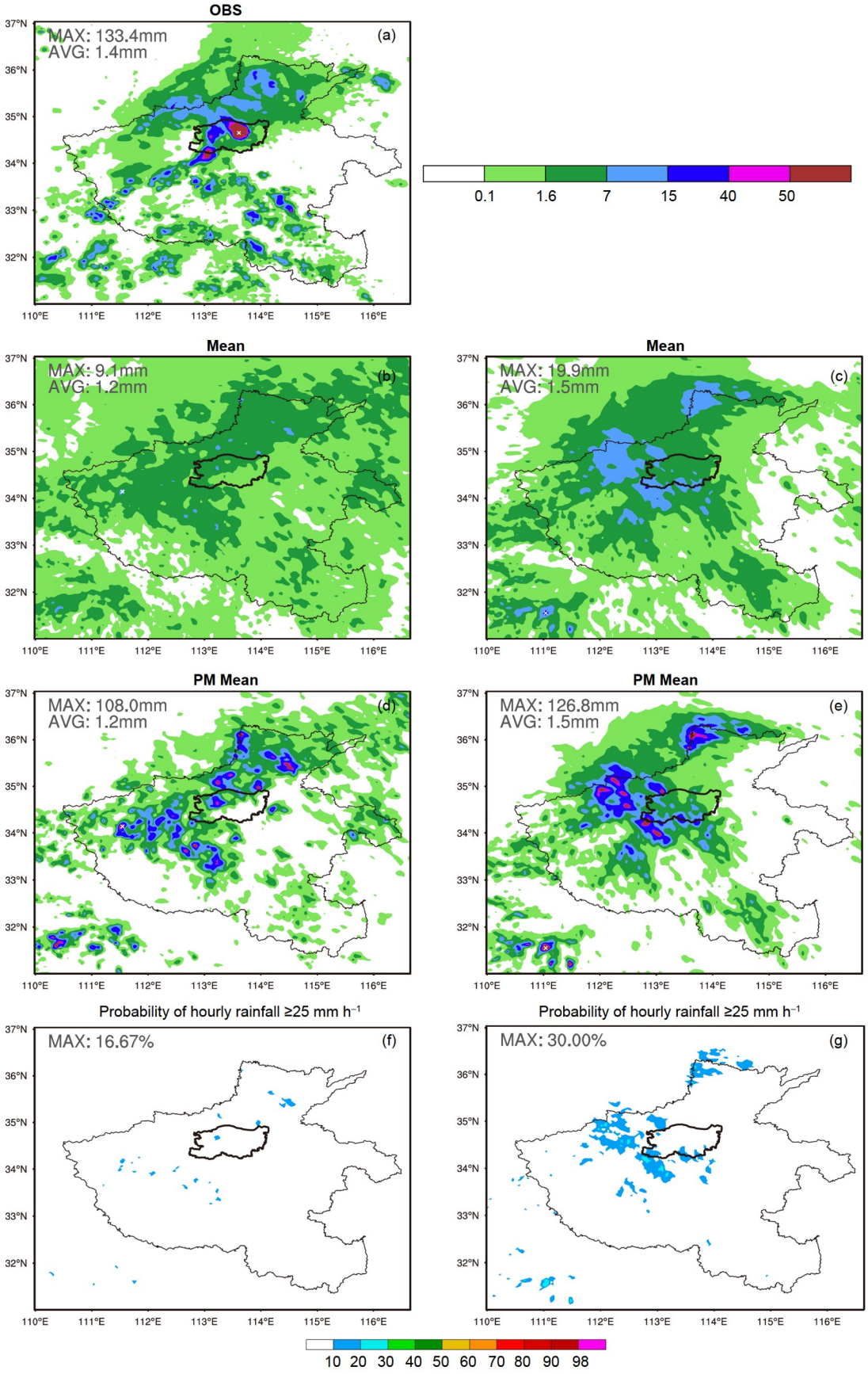

Figure 9 Similar to Figure 4 but for the hourly rainfall at the time of observed maximum hourly rainfall. Here, the threshold for POP is set to 25 mm h^{-1}.

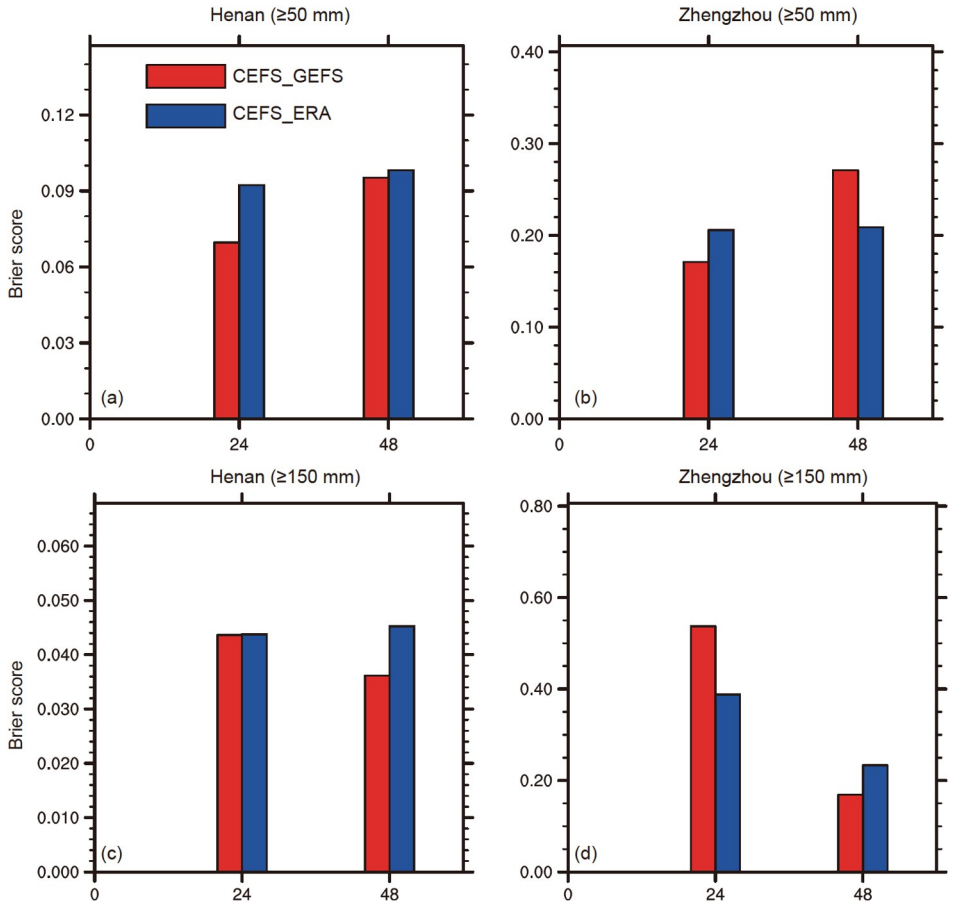

Figure 10 Brier score for the 24 hour accumulated rainfall from different CEFS experiments in a domain covering Henan (left column) and Zhengzhou (right column). The top and bottom rows represent result with 50 and 150 mm rainfall thresholds, respectively.

ror. The smaller the Brier score, the better the ensemble forecast. For the verification domain of Henan Province (left-hand column of Figure 10), CEFS_GEFS performs better than CEFS_ERA for both thresholds of ≥50 and ≥150 mm. The precipitation pattern and its location are largely determined by the initials (Zhu and Xue, 2016). As shown later, CEFS_ERA, which uses the time-lagged approach, lacks sufficient spread in ICBCs and hence produced less variation in the precipitation location and thus relatively poorer skill scores. For the smaller verification domain, the Brier score increases twofold, indicating increased uncertainty of the probability forecast within Zhengzhou. Predicting the exact location of the extreme rainfall is still a great challenge for CEFSs.

The rank histogram, which measures the ensemble forecast distribution as compared to that of observations, is investigated in Figure 11. For this verification score, the uniform distribution is the best. Both CEFSs show U-shaped distributions in the verification domain of Henan, indicating overestimation of low precipitation values and underestimation of high precipitation values (Zhu and Xue, 2016). The current CEFSs are seriously under-dispersive for rainfall. The underestimation of high values is more obvious when the verification domain is limited to Zhengzhou. It can be seen that both CEFSs show a J-shaped distribution (see Figure 11b) for the first 24 hours of accumulated rainfall. This is consistent with the previous objective evaluation in which the CEFSs greatly underestimated the precipitation values in Zhengzhou (see Figure 4). As the heavy rainfall center moves out of Zhengzhou during the second 24 hours of the forecast, the rank distributions of the two CEFSs improve.

Overall, for the large verification domain of Henan, which covers the main rain band, CEFS_GEFS initialized from the GEFS forecast is generally better than CEFS_ERA initialized from the time-lagged ERA5 forecast in terms of Brier score and rank histogram. However, for the small verification domain of Zhengzhou, due to the low predictability for the extreme rainfall, it is hard to tell subjectively from the verification scores which method is better.

4.4 Spread of the CEFSs

The current CEFSs are still under-dispersive in terms of rank histogram. Note that the larger spread does not mean better ensemble forecast if the forecast error is large. The conclu-

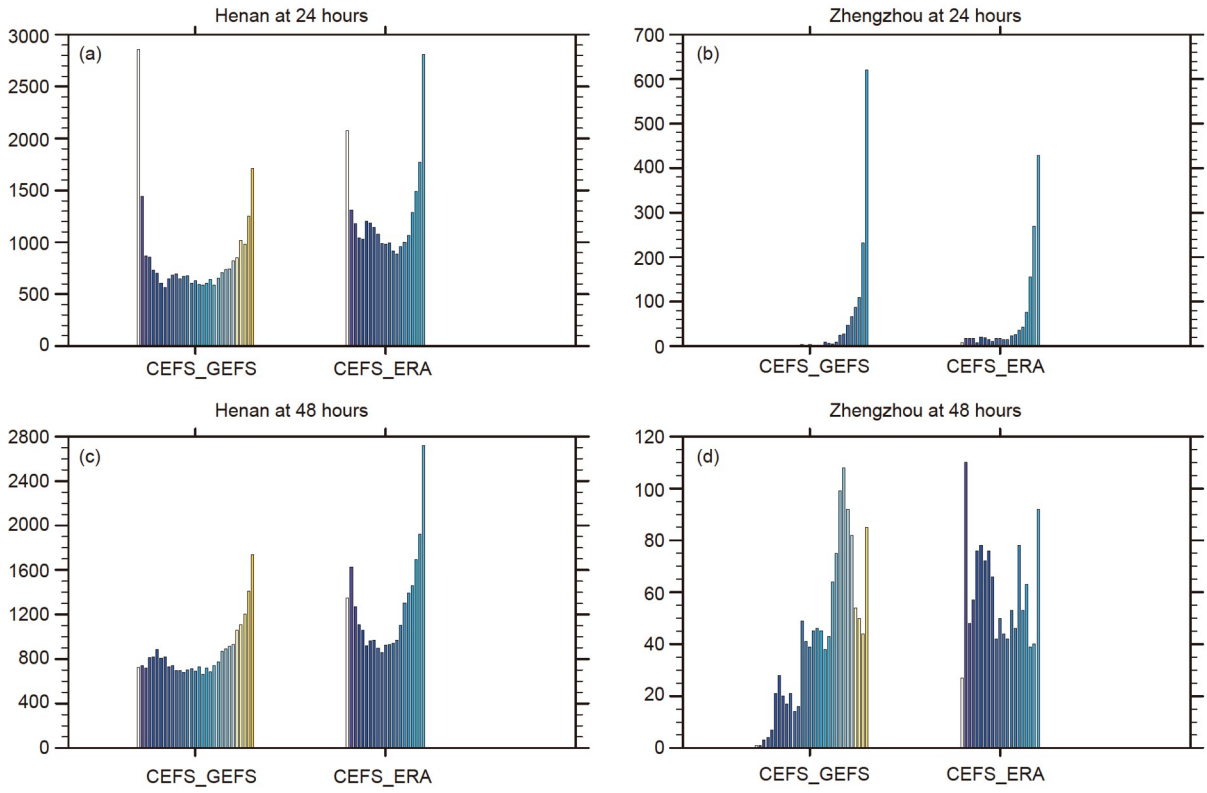

Figure 11 Rank histogram for the 24 hour accumulated rainfall for different ensemble forecasts in a domain covering (a) Henan Province and (b) Zhengzhou, for the first 24 hours of the forecast. ((c), (d)) as in ((a), (b)) but for the accumulated rainfall during hours 25–48.

sion of the relatively insufficient spread of atmospheric variables of CEFS_ERA in this subsection is based on the evaluation of the precipitation forecasts, where CEFS_ERA clearly exhibits less dispersion of rainfall distribution than CEFS_GEFS. To gain some understandings of how the spread grows as well as the role of multiple ICBCs, multi-physics schemes and the time-lagged method, the spreads of the two CEFSs are calculated. Here, the sensitivity of different numbers of ensemble members is used to examine the role of multiple ICBCs (to make the same ensemble numbers for the two CEFSs) and the time-lagged method (to see the difference with and without the time-lagged method). Also, two verification domains (labeled "L" and "O" in Figure 1a) were selected to compare the role of multi-physics schemes in the regions of heavy rainfall and scarce rainfall. The spreads using different ensemble members are referred to as "GN30", "GN20", "GN10", "GN05", "EN20", "EN10" and "EN05", in which the letter "G" represents the CEFS initialized from GEFS and the letter "E" the CEFS initialized from ERA5. The number that follows represents the total number of ensemble members used for the calculation of the spread from the corresponding CEFS. For example, "GN20" is the spread calculated from the first 20 members of CEFS_GEFS. Thus, "GN30" is an experiment of CEFS_GEFS while EN20 is CEFS_ERA.

Figure 12 shows the spread of the U (zonal) and V (meridional) wind components, temperature, and water vapor at different upper levels and at the surface for the verification domain "L", which is the location of extreme rainfall in Henan. In general, the spread of the "G" series of experiments is higher than that of the "E" series of experiments for both upper levels and the surface, indicating the great importance of the perturbation ICBCs for the design of the CEFS. For the "G" series of experiments, the increase in the number of ensemble members increases the spread if the number of ensemble members is small. It can be seen that GN20 obtains a mostly larger spread than GN10, and GN10 obtains mostly larger spread than GN05. However, the spread does not increase when the number of ensemble members reaches 20, as we see from the spread curves of GN30 (red lines in Figure 12) and GN20 (black lines in Figure 12), which mostly overlap one another. That indicates that the spread reaches a saturation state at around 20 members. For the "E" series of experiments, since the first 10 members share the same ICBCs, the growth of spread is totally dependent on the different physics schemes. The spread of the ensemble system needs some time before it reaches a saturation state. Taking the variable of temperature as an example, it needs around 18 to 24 hours to reach a saturation state (first row in Figure 12), and other variables behave similarly. The time-lagged method increases the spread considerably (see the difference between the blue and

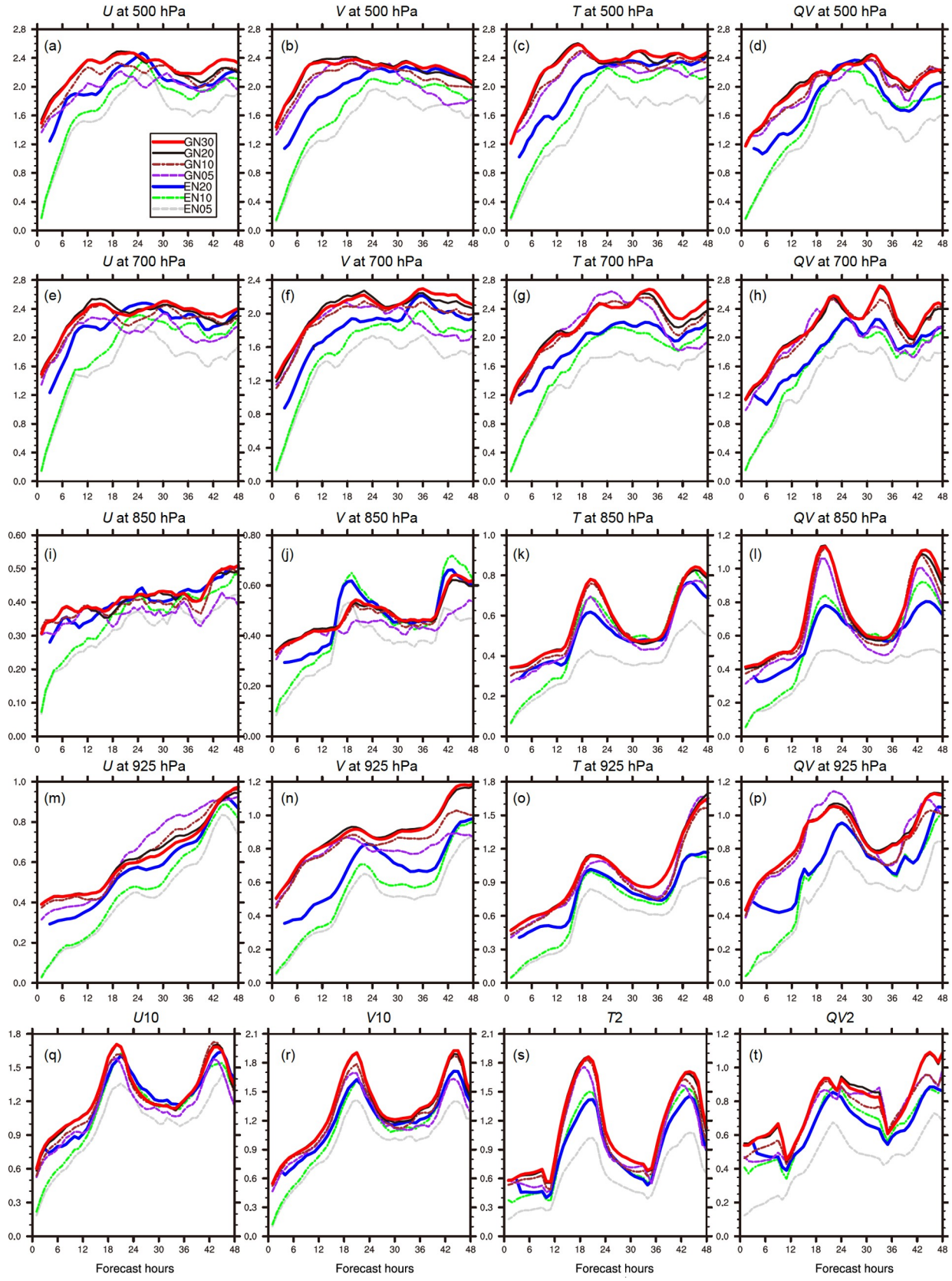

Figure 12 Spread for different variables at different levels over domain "L" in Figure 1a, where Henan heavy rainfall located. The first to fifth rows depict the results at 500, 700, 850, 925 hPa and the surface, respectively. From left to right, the columns present the variables of *U*- and *V*-wind, temperature and water vapor mixing ratio, respectively. In the legend in (a), "G" represents initial ICBCs from GEFS, and "E" from ERA5. The proceeding number represents the number of members used for the calculation of spread.

green lines in Figure 12), but is still not comparable to the corresponding "G" series with the same number of ensembles, especially for the first 24 hours of the forecast. The spread at 3 hours of the forecast for EN20 is lower than that of GN20 at the beginning, which may suggest the current ±3 hour time window is not long enough for the growth of ensemble spread. Larger time windows are needed for the design of time-lagged CEFSs.

To further understand how the physics affects the spread, we selected a region that was not directly affected by the heavy rainfall (box "O" in Figure 1)—a region that is also over the ocean, so it will have relatively stable underlying surface forcing. The "E" series of experiments, which depend mainly on different physics schemes, fail to obtain sufficient spread. This is well illustrated in Figure 13. Compared to the heavy rainfall region of Henan, the spread differences between the "G" and "E" series of experiments are enlarged, especially in the upper-air levels. It can be seen that the spread of the "E" series of experiments is clearly not comparable to that of the "G" series (see Figure 13), while it is sometimes comparable to the "G" series in the region of Henan for some variables after 24 hours of the forecast (see *U*-wind at 700 hPa and *V*-wind and temperature at 500 hPa in Figure 12). The release of latent heat within the heavy-rain area will have significant impacts on the in-cloud environment. Levels around melting levels, such as 500 hPa, are greatly affected by the heavy rainfall, and that explains why the spread of the "E" series of experiments could grow to an equivalent level as the "G" series for some variables at some levels. The above results may suggest that the usage of different physics schemes will increase the spread of CEFSs but that the region affected will be limited. The heavy rainfall region will be affected more than the region without. The insufficient spread outside the heavy-rainfall region is possibly one of the reasons why CEFS_ERA shows less variation in the distribution of precipitation (see Figure 3), since the location of precipitation and its intensity are largely determined by the environment.

5. Neighborhood POP results

Large position and intensity error ultimately lead to low predictability of CEFSs. The traditional POP method mentioned above, since the position bias is not considered, fails to provide instructive forecast guidance. In this section, the neighborhood POP method introduced in Section 3.2 is investigated. Experiment CEFS_GEFS, which has better performance, is used for the test.

Figure 14 shows the spatial-neighborhood POP for the 24 hour accumulated rainfall. Here, the neighborhood radius is set to 12 grid spacing, and we therefore refer to this experiment as N_POP_W12_T0. For the 50 mm threshold, N_POP_W12_T0 greatly increases the precipitation probability. It can be seen that the POP for most of Zhengzhou is larger than 98%. By comparison, the POP of CEFS_GEFS without the neighborhood method is mostly larger than 70% (not shown). This difference is well reflected in the likelihood diagram (see Figure 15), which can be used to visualize the ability of the forecast system to distinguish situations leading to the occurrence of an event of interest from those leading to nonoccurrence of the event. The larger the separation of the means of the conditional distributions between observed "Yes" and "No" events, the better the discrimination of the probability forecast. Also, a better system is expected to have larger variance within the conditional distributions. For CEFS_GEFS, the number of correct occurrences is evenly distributed. By comparison, N_POP_W12_T0 shows a larger variance for the correct occurrences. The higher the forecast probability, the larger the number of correct occurrences. Clearly, N_POP_W12_T0 provides a more meaningful probability forecast than CEFS_GEFS. For the 150 mm threshold, CEFS_GEFS shows a POP of no more than 50%, and is mostly located to the west of Zhengzhou (see Figure 4f). N_POP_W12_T0 greatly increases the forecast precipitation probability. The area of POP larger than 10% (see Figure 14b) matches well with observed rainfall amounts larger than 150 mm (see Figure 4a), indicating the extreme rainfall area is well captured by the neighborhood precipitation probability. The distribution of observed correct occurrence is improved, as we see the peak value moves to the right (see a difference in Figure 15c, 15d). The ROC diagram is also presented in Figure 14. Clearly, N_POP_W12_T0 is better than CEFS_GEFS for both the 50 and 150 mm threshold.

For the hourly rainfall, both the temporal- and spatial-neighborhood methods are employed, and we refer to the experiments as N_POP_W12_T1 and N_POP_W12_T0, respectively. Figure 16 shows the POP for the threshold larger than 25 mm h^{-1} at the time of the observed hourly maximum (0900 UTC 20 July 2021). At that time, the POP of CEFS_GEFS without the neighborhood method is almost zero (Figure 9f). Both neighborhood methods greatly increase the POP value, covering the heavy-rainfall area well, although the centers of high probability are still not in Zhengzhou (Figure 16a, 16b). For the area under the ROC curve, the traditional POP method shows a value close to 0.5 for most forecast hours, indicating almost no forecast skill score for the threshold larger than 25 mm h^{-1}.The neighborhood POP method greatly increases the verification skill scores. Compared to N_POP_W12_T0, N_POP_W12_T1 with the addition of the temporal neighborhood method is slightly better.

Overall, the neighborhood POP method greatly improves the performance of the CEFSs. Since precise predictions of the location and amount of extreme rainfall are still beyond

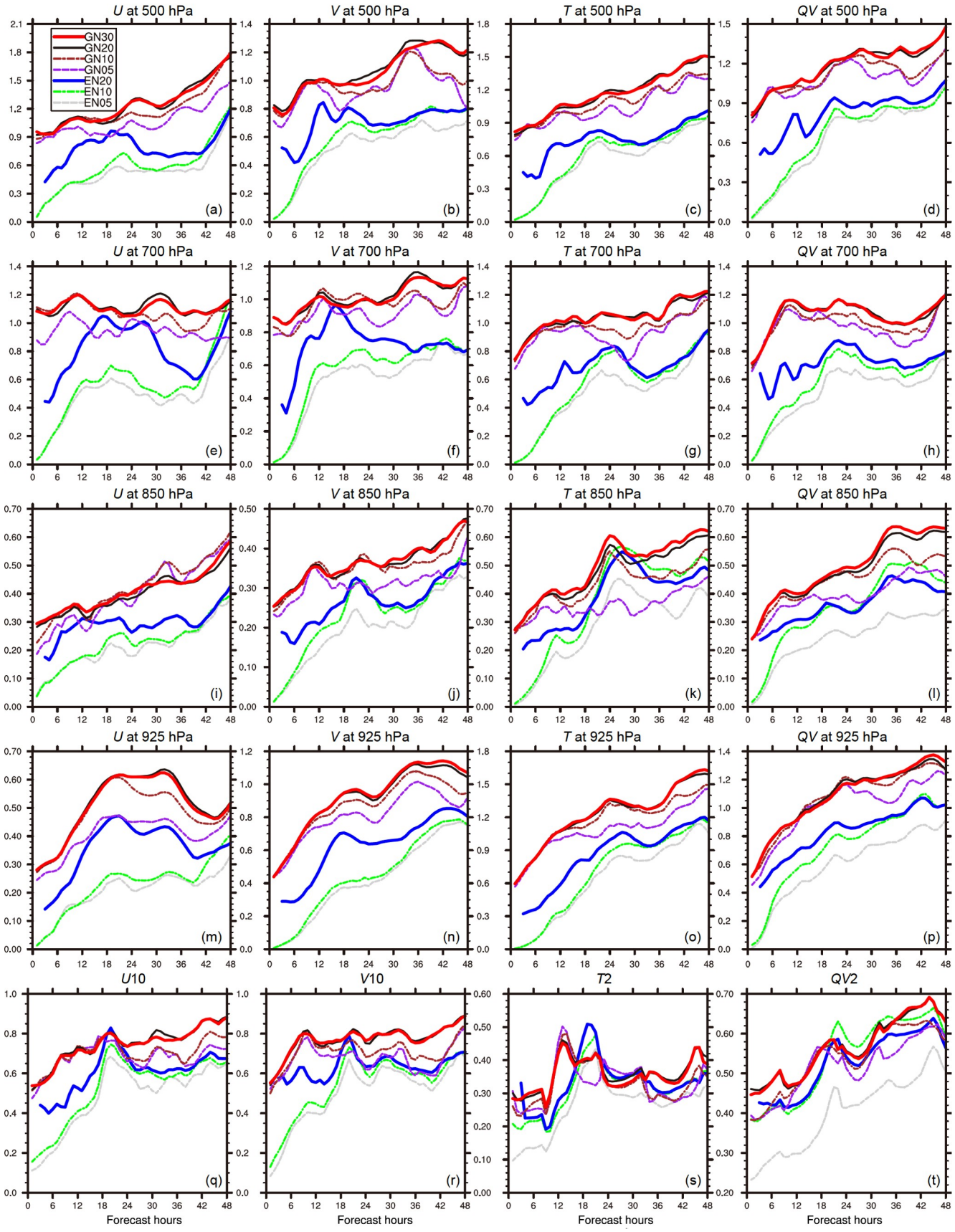

Figure 13 As in Figure 12 but for the domain "O" in Figure 1a over the ocean, which has scarce rainfall during the two days of the forecast.

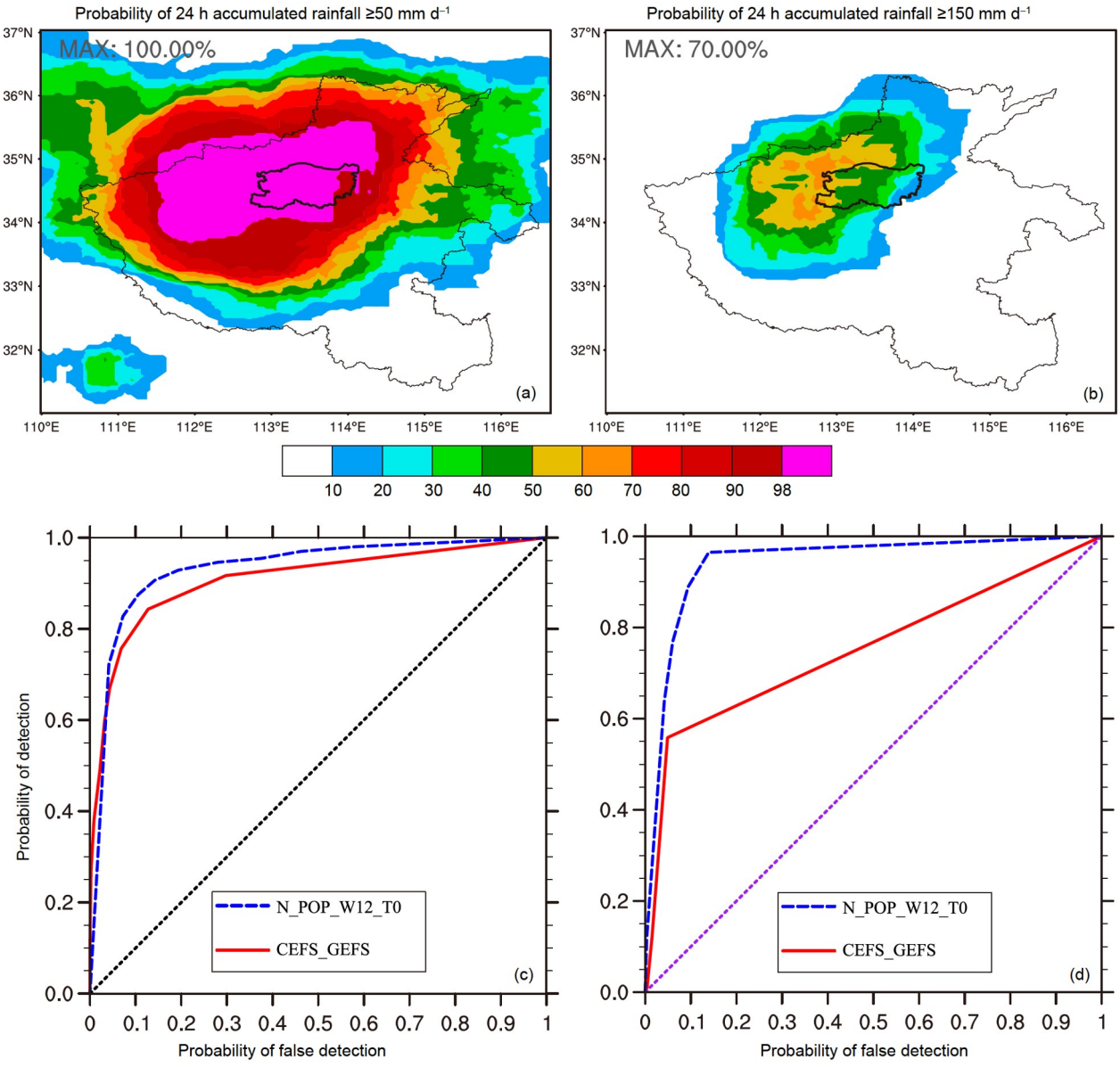

Figure 14 Neighborhood POP of 24 hour accumulated rainfall (from 1200 UTC 19 to 1200 UTC 20 July 2021) for rainfall (a) ≥50 mm and (b) ≥150 mm. The neighborhood width is set to 12 grid points and the number of points for determining a forecast "Yes" event within the neighborhood area is set to 50 grid points. ((c), (d)) Corresponding ROC diagrams of (a, b) for the verification domain "L".

the capability of CPR models, such neighborhood products, if available in real time, will greatly increase the confidence of forecasters in the prediction of highly extreme rainfall.

6. Summary and future work

This paper reports on a study of the extreme rainfall event that occurred in Henan Province on 19–21 July 2021 using CEFSs. The recorded maximum hourly rainfall reached 201.9 mm h^{-1} in Zhengzhou, which is a record over mainland China. Two 4 km CEFSs based on the WRF model were employed: one initialized from GEFS (CEFS_GEFS) and the other from time-lagged ERA5 data (CEFS_ERA). Both CEFSs used multi-physics schemes and employed WRF models.

First, the predictability of this extreme rainfall was investigated. For the two days of extreme rainfall, the observed rainfall centers moved from Zhengzhou to the foot of the Taihang Mountains. Results showed the CEFSs to produce high POP values of daily heavy rainfall along the Taihang Mountains throughout the 48 hours of the forecast, indicting the heavy rainfall induced by orographic lift and blocking is predictable. However, for the extreme daily rainfall in Zhengzhou, which is located a few hundred kilometers away from the mountains, the CEFSs show large uncertainty, with

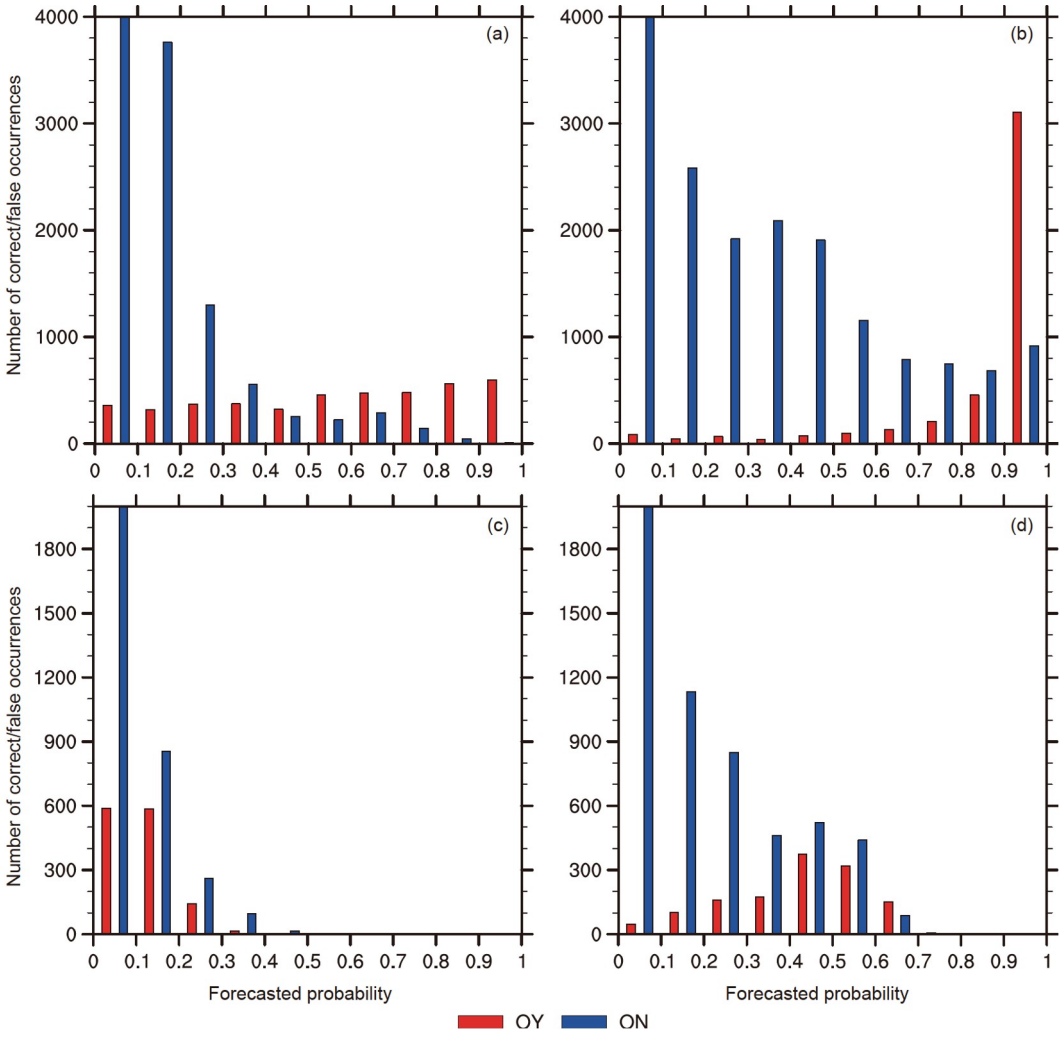

Figure 15 Likelihood diagrams for the 24 hour accumulated rainfall (≥50 mm) for the POP (a) without and (b) with the neighborhood method. ((c), (d)) As in ((a), (b)) but for the threshold ≥150 mm. The time period is from 1200 UTC 19 to 1200 UTC 20 July 2021.

only a few members predicting correct rainfall centers. The predictability of extreme daily rainfall in Zhengzhou is limited if the position biases are not considered. For the hourly rainfall, only a few members capture the extreme rainfall in Zhengzhou. Most of members have large position errors. The predicted maximum hourly rainfall of all members is around 180 mm h^{-1}, lower than the observed maximum. The predictability of extreme hourly rainfall at the accuracy of city scale in Zhengzhou is low.

The performances of the two CEFSs were evaluated. Both CEFSs show under-dispersion for the precipitation forecast in terms of RANK histograms. In terms of Brier score, CEFS_GEFS performs better than CEFS_ERA for the large verification domain covering Henan Province. For the small verification domain of Zhengzhou, the Brier score of both CEFSs increase greatly, highlighting the limited intrinsic predictability for the location of city-scale extreme rainfall.

The roles of multi-physics schemes, multiple ICBCs, and the time-lagged method in the growth of ensemble spread were further examined through comparison of the two CEFSs using different numbers of ensemble members and different verification domains. Two regions were selected: one was the same as the previous verification domain, which was greatly influenced by the extreme rainfall in Henan; and the other was over the ocean which has scarce rainfall during the two days of the forecast. In general, the spreads of all investigated forecast variables (*U*-wind, *V*-wind, temperature, and water vapor) for CEFS_GEFS are larger than those of CEFS_ERA, indicating the key role played by multiple ICBCs. The current ±3 hour time window is insufficient for the forecast error growth, and the unified boundary is another factor for the relatively smaller spread of CEFS_ERA. The difference in spread between the two CEFSs is more obvious in the no-precipitation region, suggesting the multi-physics schemes mostly affect the region of heavy rainfall. With less spread of forecast variables in the no-precipitation region, CEFS_ERA shows less variation in the location of precipitation, resulting in poor forecast skill scores.

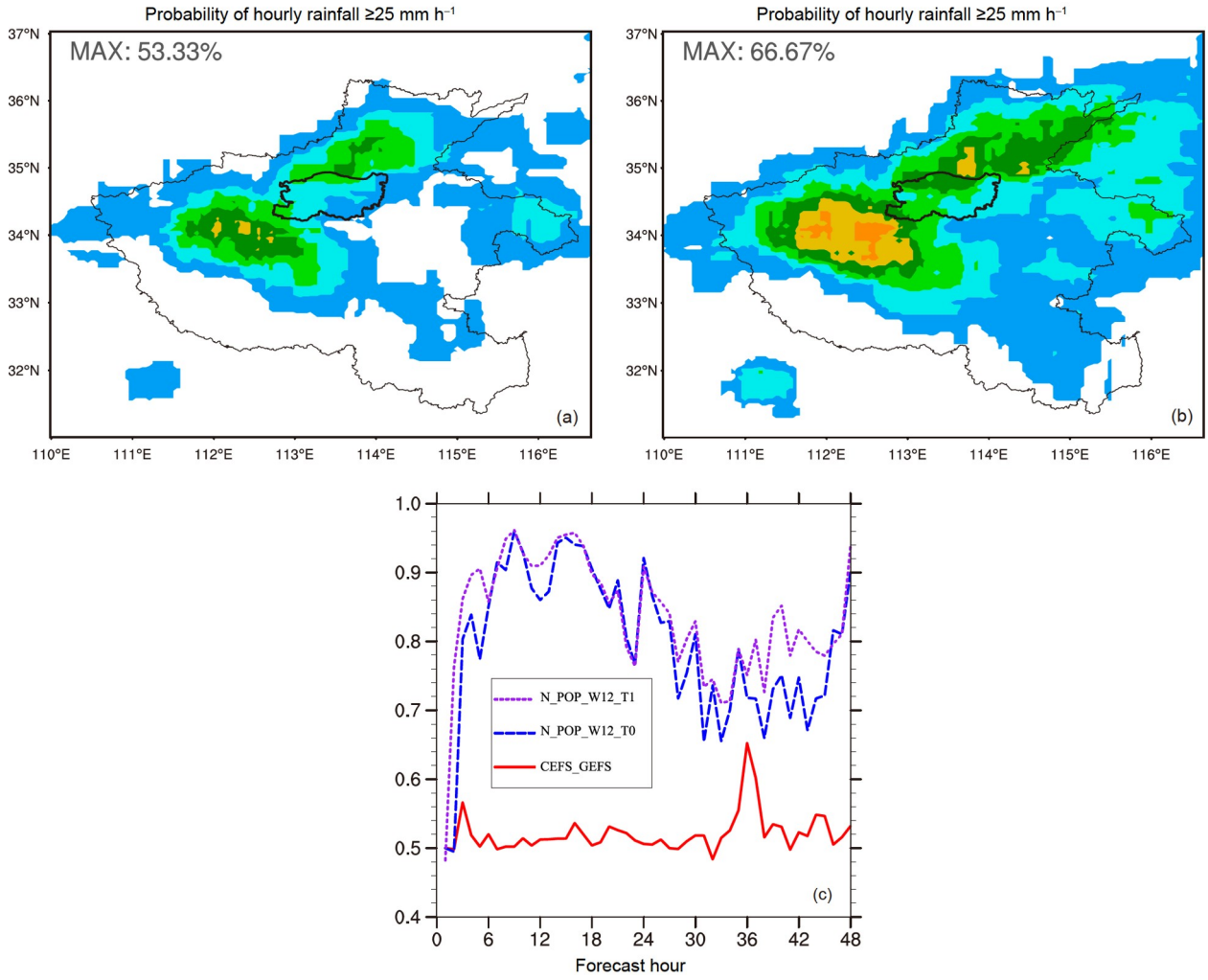

Figure 16 POP of hourly rainfall (≥25 mm h^{-1}) for the (a) spatial- and (b) temporal- and spatial-POP method. The neighborhood width is set to 12 grid points and the number of points for determining a forecast "Yes" event within the neighborhood area is set to 20 grid points. The time window is set to ±1 hour for the temporal neighborhood method. (c) The corresponding area under the ROC curve for the POP with and without the neighborhood method.

Additionally, a neighborhood POP method is introduced, with which CEFS_GEFS shows moderate POP values for heavy rainfall, at both daily and hourly scales, over Zhengzhou. The predictability of this extreme rainfall is improved by using this method. Objective verification shows that the neighborhood method improves the forecast probability distribution and significantly improves the forecast skill score.

In this study, we focus on the predictability of extreme rainfall using CEFSs. We do not attempt to fully analyze and understand the role of ICBCs and physical processes responsible for the extreme rainfall, or the key factors for a successful forecast. Some "good" members are able to reproduce the general evolution and propagation of the extreme rainfall, albeit with some intensity and position errors. In the future, we plan to further analyze this dataset to gain some insights along these lines. Also, some methods for representing errors of ICBCs and uncertainty in the model physics schemes are worthy of examination. The impact of different perturbation methods on the ensemble spread at convection-permitting resolution will be compared and investigated in the future.

Acknowledgements *This work was primarily supported by the National Natural Science Foundation of China (Grant Nos. 41975124, 41730965) and the National Key Research and Development Program of China (Grant No. 2018YFC1507604).*

References

Brown B G, Gotway J H, Bullock R, Gilleland E, Ahijevych D. 2009. The Model Evaluation Tools (MET): Community tools for forecast evaluation. In: 25th Conference on International Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology. Paper 9A.6

Cai S X, Huang A N, Zhu K F, Yang B, Yang X Y, Wu Y, Mu X Y. 2021. Diurnal cycle of summer precipitation over the Eastern Tibetan Plateau and surrounding regions simulated in a convection-permitting model.

Clim Dyn, 57: 611–632
Clark A J, Gallus Jr. W A, Xue M, Kong F. 2009. A comparison of precipitation forecast skill between small convection-permitting and large convection-parameterizing ensembles. Weather Forecast, 24: 1121–1140
Clark A J, Gallus William A. J, Xue M, Kong F. 2010. Growth of Spread in Convection-Allowing and Convection-Parameterizing Ensembles. Weather Forecast, 25: 594–612
Clark A J, Weiss S J, Kain J S, Jirak I L, Coniglio M, Melick C J, Siewert C, Sobash R A, Marsh P T, Dean A R, Xue M, Kong F, Thomas K W, Wang Y, Brewster K, Gao J, Wang X, Du J, Novak D R, Barthold F E, Bodner M J, Levit J J, Entwistle C B, Jensen T L, Correia James J. 2012. An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. Bull Amer Meteorol Soc, 93: 55–74
Du J, DiMego G, Zhou B B, Jovic D, Ferrier B, Yang B, Benjamin S. 2014. NCEP Regional Ensembles: Evolving toward hourly-updated convectionallowing scale and storm-scale predictions within a unified regional modeling system. In: 26th Conference on Weather Analysis and Forecasting/22nd Conference on Numerical Weather Prediction. Paper J1.4
Evans C, Van Dyke D F, Lericos T. 2014. How do forecasters utilize output from a convection-permitting ensemble forecast system? Case study of a high-impact precipitation event. Weather Forecast, 29: 466–486
Fritsch J M, Carbone R E. 2004. Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. Bull Amer Meteorol Soc, 85: 955–966
Gallus Jr. W A, Bresch J F. 2006. Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. Mon Weather Rev, 134: 2632–2641
Gebhardt C, Theis S E, Paulat M, Ben Bouallègue Z. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmos Res, 100: 168–177
Golding B, Roberts N, Leoncini G, Mylne K, Swinbank R. 2016. MOGREPS-UK convection-permitting ensemble products for surface water flood forecasting: Rationale and first results. J Hydrometeorol, 17: 1383–1406
Hamill T M. 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon Weather Rev, 129: 550–560
Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De C G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan R J, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de R P, Rozum I, Vamborg F, Villaume S, Thépaut J N. 2020. The ERA5 global reanalysis. Q J R Meteorol Soc, 146: 1999–2049
Kong F Y, Droegemeier K K, Hickmon N L. 2007. Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. Mon Weather Rev, 135: 759–782
Kong F Y, Xue M, Xue M, Droegemeier K K, Thomas K W, Wang Y, Kain J S, Weiss S J, Bright D, Du J. 2008. Real-time storm-scale ensemble forecast experiment-Analysis of 2008 spring experiment data. In: 24th Conference on severe local storms. Paper 12.13
Li P X, Furtado K, Zhou T J, Chen H M, Li J, Guo Z, Xiao C. 2020. The diurnal cycle of East Asian summer monsoon precipitation simulated by the Met Office Unified Model at convection-permitting scales. Clim Dyn, 55: 131–151
Li P X, Furtado K, Zhou T J, Chen H M, Li J. 2021. Convection-permitting modelling improves simulated precipitation over the central and eastern Tibetan Plateau. Q J R Meteorol Soc, 147: 341–362
Loken E D, Clark A J, Xue M, Kong F Y. 2017. Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. Weather Forecast, 32: 1403–1421
Loken E D, Clark A J, Xue M, Kong F. 2019. Spread and skill in mixed- and single-physics convection-allowing ensembles. Weather Forecast, 34: 305–330
Marzban C. 2004. The ROC curve and the area under it as performance measures. Weather Forecast, 19: 1106–1114
Murphy A H. 1973. A new vector partition of the probability score. J Appl Meteorol, 12: 595–600
Novak D R, Bright D R, Brennan M J. 2008. Operational forecaster uncertainty needs and future roles. Weather Forecast, 23: 1069–1084
Nuissier O, Marsigli C, Vincendon B, Hally A, Bouttier F, Montani A, Paccagnella T. 2016. Evaluation of two convection-permitting ensemble systems in the HyMeX Special Observation Period (SOP1) framework. Q J R Meteorol Soc, 142: 404–418
Pan Y, Shen Y, Yu J J, Xiong A Y. 2015. An experiment of high-resolution gauge-radar-satellite combined precipitation retrieval based on the Bayesian merging method (in Chinese). Acta Meteorol Sin, 73: 177–186
Peralta C, Ben Bouallègue Z, Theis S E, Gebhardt C, Buchhold M. 2012. Accounting for initial condition uncertainties in COSMO-DE-EPS. J Geophys Res, 117: D07108
Romine G S, Schwartz C S, Berner J, Fossell K R, Snyder C, Anderson J L, Weisman M L. 2014. Representing forecast error in a convection-permitting ensemble system. Mon Weather Rev, 142: 4519–4541
Schumacher R S, Clark A J, Xue M, Kong F Y. 2013. Factors influencing the development and maintenance of nocturnal heavy-rain-producing convective systems in a storm-scale ensemble. Mon Weather Rev, 141: 2778–2801
Schwartz C S, Romine G S, Fossell K R, Sobash R A, Weisman M L. 2017. Toward 1-km ensemble forecasts over large domains. Mon Weather Rev, 145: 2943–2969
Schwartz C S, Romine G S, Smith K R, Weisman M L. 2014. Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble kalman filter. Weather Forecast, 29: 1295–1318
Shen Y, Zhao P, Pan Y, Yu J J. 2014. A high spatiotemporal gauge-satellite merged precipitation analysis over China. J Geophys Res-Atmos, 119: 3063–3075
Shi W R, Li X, Zheng M J, Zhang B, Wang H B, Zhu K F, Zhuge X Y. 2021. Multi-model comparison and high-resolution regional model forecast analysis for the “7.20” Zhengzhou severe heavy rain (in Chinese). Trans Atmos Sci, 44: 688–702
Sobash R A, Kain J S, Bright D R, Dean A R, Coniglio M C, Weiss S J. 2011. Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. Weather Forecast, 26: 714–728
Stensrud D J, Brooks H E, Du J, Tracton M S, Rogers E. 1999. Using ensembles for short-range forecasting. Mon Weather Rev, 127: 433–446
Su A F, Lv X N, Cui L M, Li Z, Xi L, Li H. 2021. The basic observational analysis of “7.20” extreme rainstorm in Zhengzhou (in Chinese). Torrent Rain Disast, 40: 445–454
Wang J Z, Chen F J, Chen J, Liu X Q, Li H Q, Deng G, Li X L, Wang Y Z. 2021. Verification of GRAPES-REPS model precipitation forecasts over China during 2019 flood season (in Chinese). Chin J Atmos Sci, 45: 664–682
Wang L, Shen X. 2019. Review on the representation of model uncertainty in convection-allowing ensemble prediction system (in Chinese). Meteorol Mon, 45: 1158–1168
Wei M Z, Toth Z, Wobus R, Zhu Y J. 2008. Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. Tellus A-Dynamic Meteor Oceanography, 60: 62–79
Wei P, Xu X, Xue M, Zhang C Y, Wang Y, Zhao K, Zhou A, Zhang S S, Zhu K F. 2022. On key dynamical processes supporting the 21.7 Zhengzhou record-breaking hourly rainfall in China. Adv Atmos Sci, https://doi.org/10.1007/s00376-022-2061-y
Wilks D S. 2010. Sampling distributions of the Brier score and Brier skill score under serial dependence. Q J R Meteorol Soc, 136: 2109–2118
Wu N G, Zhuang X R, Min J Z, Meng Z Y. 2020. Practical and intrinsic predictability of a warm-sector torrential rainfall event in the south China monsoon region. J Geophys Res-Atmos, 125: e31313

Wu Y, Huang A N, Huang D Q, Chen F, Yang B, Zhou Y, Fang D X, Zhang L J, Wen L J. 2018. Diurnal variations of summer precipitation over the regions east to Tibetan Plateau. Clim Dyn, 51: 4287–4307

Xue M, Kong F Y, Weber D, Thomas K W, Wang Y, Brewster K, Droegemeier K K, Weiss J S K S J, Bright D R, Wandishin M S, Coniglio M C, Du J. 2007. CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. In: 22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction, CDROM 3B.1

Yin J F, Gu H D, Liang X D, Yu M, Sun J S, Xie Y X, Li F, Wu C. 2021. A possible dynamic mechanism for rapid production of the extreme hourly rainfall in Zhengzhou city on 20 July 2021. J Meteorol Res, 36: 6–25

Yussouf N, Stensrud D J. 2011. Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the ensemble kalman filter. Mon Wea Rev, 140: 562–586

Zhang D L, Lin Y H, Zhao P, Yu X D, Wang S Q, Kang H W, Ding Y H. 2013. The Beijing extreme rainfall of 21 July 2012: "Right results" but for wrong reasons. Geophys Res Lett, 40: 1426–1431

Zhao Y, Huang A N, Kan M Y, Dong X N, Yu X J, Wu Y, Zhang X D, Cai S X. 2020. Characteristics of hourly extreme precipitation along the Yangtze River Basin, China during warm Season. Sci Rep, 10: 5613

Zhu K F, Xue M. 2016. Evaluation of WRF-based convection-permitting multi-physics ensemble forecasts over China for an extreme rainfall event on 21 July 2012 in Beijing. Adv Atmos Sci, 33: 1240–1258

Zhu Y J, Zhou X Q, Li W, Hou D C, Melhauser C, Sinsky E, Pena M, Fu B, Guan H, Kolczynski W, Wobus R, Tallapragada V. 2018. Toward the improvement of subseasonal prediction in the National Centers for Environmental Prediction global ensemble forecast system. J Geophys Res-Atmos, 123: 6732–6745

(Responsible editor: Zhiyong MENG)