ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 33, NOVEMBER 2016, 1240–1258

Evaluation of WRF-based Convection-Permitting Multi-Physics Ensemble Forecasts over China for an Extreme Rainfall Event on 21 July 2012 in Beijing

Kefeng ZHU[1] and Ming XUE*[1,2]

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

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

(Received 29 July 2016; revised 22 August 2016; accepted 2 September 2016)

ABSTRACT

On 21 July 2012, an extreme rainfall event that recorded a maximum rainfall amount over 24 hours of 460 mm, occurred in Beijing, China. Most operational models failed to predict such an extreme amount. In this study, a convective-permitting ensemble forecast system (CEFS), at 4-km grid spacing, covering the entire mainland of China, is applied to this extreme rainfall case. CEFS consists of 22 members and uses multiple physics parameterizations. For the event, the predicted maximum is 415 mm d^{-1} in the probability-matched ensemble mean. The predicted high-probability heavy rain region is located in southwest Beijing, as was observed. Ensemble-based verification scores are then investigated. For a small verification domain covering Beijing and its surrounding areas, the precipitation rank histogram of CEFS is much flatter than that of a reference global ensemble. CEFS has a lower (higher) Brier score and a higher resolution than the global ensemble for precipitation, indicating more reliable probabilistic forecasting by CEFS. Additionally, forecasts of different ensemble members are compared and discussed. Most of the extreme rainfall comes from convection in the warm sector east of an approaching cold front. A few members of CEFS successfully reproduce such precipitation, and orographic lift of highly moist low-level flows with a significantly southeasterly component is suggested to have played important roles in producing the initial convection. Comparisons between good and bad forecast members indicate a strong sensitivity of the extreme rainfall to the mesoscale environmental conditions, and, to less of an extent, the model physics.

Key words: convective-permitting, ensemble forecasts, moist low-level flows, orographic lifting, extreme rainfall

Citation: Zhu, K. F., and M. Xue, 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**(11), 1240–1258, doi: 10.1007/s00376-016-6202-z.

1. Introduction

On 21 July 2012, an extreme rainfall event occurred in Beijing, China. The average 24-h accumulated rainfall across rain gauge stations in the city of Beijing was 190 mm, which is the highest in the recorded history of Beijing since 1951 (Chen et al., 2012). The maximum rainfall among all meteorological and hydrologic sites was 460 mm, at a hydrological station in Hebei Township of Fangshan District in the southwest suburb of Beijing. The maximum recorded hourly rainfall was 100.3 mm (Chen et al., 2012). Such excessive rainfall caused major urban flooding in Beijing; 79 people died and millions of people were affected. Direct financial loses were estimated to be about 2 billion U.S dollars. As a reference, the mean annual rainfall amount from 1949 to 2010 is only 600 mm in the Beijing region. For this event, the Beijing Meteorological Bureau issued an orange-color rainstorm warning, which is the second from highest level. The actual, extreme amount of rainfall was, however, not expected by forecasters. The city population was not well prepared.

Operational NWP models predicted a general rain pattern and high probability of heavy rain in the Beijing area for that day. However, the rain intensity predicted varied greatly by forecast model, forecast lead time, as well as model resolution (Tao and Zheng, 2013; Jiang et al., 2014). The rain intensity was significantly under-predicted by global models from various NWP centers (maximum <200 mm d^{-1}) (Zhang et al., 2013; Yu and Meng, 2016). The convection-permitting Rapid Update Cycle of Beijing Meteorological Bureau (BJ-RUC) (Fan et al., 2009), which has a horizontal resolution of 3 km, did predict 24-h rain intensity of more than 300 mm (Jiang et al., 2014). However, the predicted maximum was outside Beijing. Moreover, BJ-RUC is a deterministic

* Corresponding author: Ming XUE
Email: mxue@ou.edu

forecasting system. The prediction by a single model of an extreme rainfall amount that was about half of the average annual rainfall amount did not give decision-makers much confidence without additional probabilistic guidance.

Ensemble forecasting is well-established as an effective way of providing uncertainty estimates of weather forecasting, and for forecasting the probability of certain events occurring. Most operational NWP centers, including the ECMWF (Palmer et al., 1993; Molteni et al., 1996; Buizza et al., 2003), NCEP (Toth and Kalnay, 1993; Du et al., 1997; Toth and Kalnay, 1997), UKMO (Bowler et al., 2008) and MSC (Ritchie and Beaudoin, 1994; Houtekamer et al., 1996; Pellerin et al., 2003), have developed, and are running, global and regional operational ensemble prediction systems (EPSs), although the resolutions are generally too coarse to resolve convection. The China Meteorological Administration (CMA) has also developed global (Ren et al., 2011; Liu et al., 2013) and regional (Zhang et al., 2014) EPSs. Both have 14 members. For this case, the ensemble forecasts of the NCEP global EPS are more accurate than those of the ECMWF and CMA global EPSs for the peak rainfall stage (Yu and Meng, 2016).

Studies have shown that ensemble-derived quantitative precipitation forecasts are often more skillful than a single forecast (e.g., Buizza and Hollingsworth, 2002; Theis et al., 2005; Charles and Colle, 2009). Li et al. (2015) conducted a series of mesoscale (horizontal grid spacing of 9 km) ensemble forecasts for this extreme event. Among all the ensemble forecasts, the experiment that used initial perturbation, multiple physics and multiple initial and boundary conditions was the best and was much better than the deterministic forecast in the prediction of rain intensity as well as the location. However, the rain intensities were still underestimated. For precipitation forecasting, it is desirable to use model resolutions that allow direct prediction of convection instead of relying on convective parameterization, which is a great source of uncertainty. Since 2007, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma has been producing experimental storm-scale ensemble forecasts (SSEFs) over the continental United States at 3- to 4-km grid spacing (considered convection-permitting resolutions) (Xue et al., 2007), for evaluation at NOAA's Hazardous Weather Testbed (Clark et al., 2012). The SSEF system has been increasing in sophistication over the years (Xue et al., 2011; Kong et al., 2014), and has included multiple dynamic cores, and multiple physics options in combination with perturbed initial and boundary conditions. Such convection-permitting ensembles have been shown to have better abilities in providing severe weather forecasting guidance over the continental United States (e.g., Clark et al., 2009; Schwartz et al., 2009; Xue et al., 2013; Iyer et al., 2016).

To the best of the authors' knowledge, ensemble forecasting at convection-permitting resolutions has so far not been applied to the forecasting of heavy or extreme precipitation in China—at least not over continental China as a whole. For extreme but low-probability rainfall events in Beijing, it is even more important to assess the ability of state-of-the-art non-hydrostatic models running at a convection-permitting resolution in predicting such events, and to assess the uncertainty of such predictions.

Even though BJ-RUC (which was based on the WRF model) predicted extreme rainfall amounts associated with this event, it did not place the maximum precipitation in Beijing. Zhang et al. (2013) pointed out that most operational model forecasts failed to capture the precipitation associated with the convection in the warm sector ahead of an approaching cold front, which was a key contributor to the total extreme precipitation (Tao and Zheng, 2013). The prediction of warm-sector convection remains a major challenge (Zhong et al., 2015). An extended goal of our study is to understand the physical processes responsible for producing the historical rainfall amount at a particular location and time, by analyzing output from an ensemble of forecasts that have different initial and boundary conditions, and different model physics. As the first step towards this goal, we document and evaluate in this paper the performance of a 4-km ensemble of forecasts, in both the probabilistic and deterministic sense. For brevity, the goal of obtaining a physical understanding is left for a separate paper.

Specifically, we apply a similar strategy used by CAPS's SSEF to the Beijing case. The WRF model (Skamarock et al., 2005) is used with its various physics parameterizations. As part of an ongoing effort to establish and evaluate a convection-permitting/convection-resolving ensemble forecasting system (CEFS) suitable for warm-season precipitation forecasting over China, we use a model domain that covers the whole of continental China (see Fig. 1). The initial and boundary conditions are interpolated from those of an experimental version of the NCEP Global Ensemble Forecasting System (GEnFS), whose initial conditions were produced by the ensemble Kalman filter data assimilation method. To help understand the sensitivity of the prediction of extreme rainfall to model physics (relative to the initial and boundary conditions), we perform a second set of ensemble forecasts in which only the physics parameterization schemes differ. To the best of our knowledge, this high-impact event has not been studied from the perspective of a convection-permitting ensemble.

The rest of this paper is organized as follows: Section 2 provides an overview of the 21 July 2012 extreme rainfall event in Beijing. In section 3, the model configurations and ensemble experiments are described, and the probability matching (PM) ensemble mean algorithm and verification metrics are described. The ensemble forecasts are examined in section 4, together with an investigation of the warm-sector rain. The sensitivity of precipitation to the model physics is examined in section 5. Finally, we conclude in section 6 with a summary and suggestions for future work.

2. Overview of the extreme rainfall event

The extreme rainfall event in Beijing on 21 July 2012 started at 0200 UTC [1000 LST (local standard time)] and

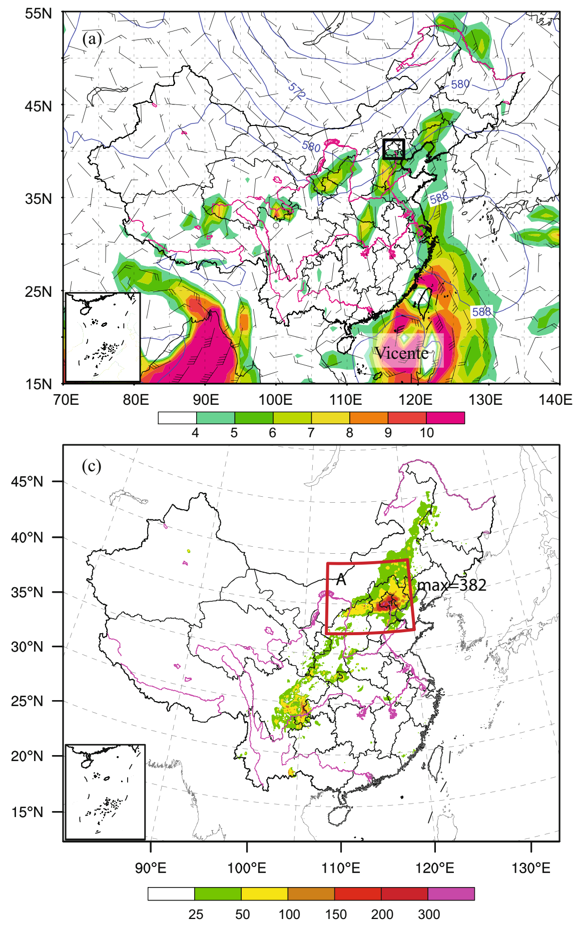

Fig. 1. The 500-hPa geopotential height (blue lines; units: gpm), 925-hPa winds (barbs; units: m s^{-1}) and 925-hPa horizontal water vapor flux [color–shaded; untis: g (hPa cm $s)^{-1}$] at (a) 0600 UTC 21 July and (b) 1200 UTC 21 July 2012 (black box indicates the location of Beijing, typhoon Vicente is located in the south China sea), and the 24-h accumulated rainfall (units: mm) from 0000 UTC 21 July to 0000 UTC 22 July 2012 from (c) rain gauge observations and (d) NCEP GFS forecast. The magenta box in (c) indicates one of the two verification areas, with the other being mainland China. For the observation plot, we use distance-weighted interpolation to obtain the gridded data, and the maximum after interpolation is 382 mm. The corresponding maximum value of GFS is 195 mm.

ended at 1800 UTC (0200 LST). According to Li et al. (2013), there were two main stages of the heavy rainfall. The first stage was from 0200 UTC to 1200 UTC, which is defined as the period of warm-sector rainfall. During this period, a cold front was slowly moving into the Beijing region from the west. Southeasterly air with very high moisture content converged onto the windward slope of the Taihang Mountain Range west of Beijing. A large number of storm cells formed in the southwest part of Beijing near the mountain range and moved northeastwards, and organized into a quasi-linear mesoscale convective system (MCS) that was more or less parallel to the southwest–northeast-oriented mountain range (Yu, 2012). New cells continuously formed at the southern end of the linear MCS, and moved northeast along the MCS, exhibiting back-building and echo-training processes (Doswell et al., 1996; Yu, 2012). As intense storm cells continually moved over similar areas, extreme precipitation was produced. The maximum hourly rainfall was observed around 1300 UTC (2100 LST), when it exceeded 100 mm h^{-1} (Chen et al., 2012). After 1200 UTC, the cold front moved into Beijing, producing a smaller amount of rainfall, and the precipitation system moved out of the Beijing region by 1800 UTC.

Figure 1 shows the synoptic weather charts at 0600 UTC and 1200 UTC, together with the observed (Fig. 1c) and predicted (NCEP GFS; Fig. 1d) 24-h accumulated precipitation between 0000 UTC 21 July and 0000 UTC 22 July 2012. At 0600 UTC, Beijing (enclosed in the black box in Fig. 1a) was located ahead of a major 500 hPa trough, which moved eastward to be closer to Beijing over the next 6 h (Fig. 1b). From the wind and horizontal water vapor flux fields, one can see two channels of water vapor transportation into the Beijing region: a stronger channel associated with the south-southeasterly flows, which brought in water vapor from the East China Sea; and a weaker one associated with the south-southwesterly flows, which transported water vapor through Southwest China from the Bay of Bengal. At this time, the northwestern Pacific subtropical high was located off the East

China Sea, while the intensifying tropical storm Vicente was present over the South China Sea, moving towards the coastal province of Guangdong. The northward moisture transportation towards northern China is believed to have been strengthened by Vicente (Yu, 2012; Wen et al., 2015), which created a strong channel of flow between its cyclonic circulation and the anticyclonic circulation of the subtropical high (Fig. 1b).

The observed 24-h accumulated rainfall (Fig. 1c) shows a band of precipitation stretching from Southwest China through Northeast China, with amounts mostly below 50 mm. This band was associated with a cold front. Over most of the Beijing area, total rainfall exceeded 100 m, and the heaviest rain was east of the frontal rainband. Based on this objectively analyzed precipitation map (with smoothing effect), the maximum 24-h rainfall was 382 mm. For this event, the NCEP GFS successfully predicted heavy rain over the Beijing area (Fig. 1d), although the maximum amount of 190 mm was not enough to indicate the occurrence of a historical extreme rainfall event.

3. Experimental design and verification methods

3.1. *CEFS*

The design of our 4-km CEFS follows the CAPS SSEF (Kong et al., 2007; Xue et al., 2007), but with some differences. Only a single WRF model (WRF3.3.1) (Skamarock et al., 2005) dynamic core is used. The model domain, with 4-km horizontal grid spacing, has 1500×1100 grid points in the horizontal direction and 50 levels in the vertical direction (see Fig. 1)— large enough to cover all of mainland China. The ensemble forecasts start from 0000 UTC 21 July and end at 0000 UTC 22 July 2012, covering both precipitation stages.

The initial and lateral boundary conditions are from the experimental global ensemble forecast system (referred to as GEnFS to distinguish it from the operational NCEP GEFS) produced by NOAA's ESRL, which is initialized by a global ensemble Kalman filter (EnKF) data assimilation system, as described in Whitaker et al. (2008). This EnKF system has been shown to produce better initial conditions and subsequent tropical cyclone forecasts (Whitaker et al., 2008) than the operational GSI three-dimensional variational system used at that time.

GEnFS has 80 members and the data assimilation cycle is 6 h. The gridded horizontal resolution of the data is 0.5°, and there are 27 vertical levels. The first 20 members are used here to initialize our ensemble forecasts and to provide the boundary conditions (BCs). Additionally, we add two special members: one uses the operational NCEP GFS analysis and forecast as the initial conditions (ICs) and BCs, and the other uses the GEnFS ensemble mean analysis and corresponding forecast for the same purposes. We name these two members g0 and e0, respectively (Table 1). Therefore, there are a total of 22 members in the 4-km CEFS.

As with the CAPS SSEF systems, multiple physics suites are used within CEFS to form a multi-physics ensemble. Table 1 lists the physics configurations of the CEFS members. We follow two simple principles to select the physics configurations: one is to make the physics configurations as diverse as possible, and the other is to have two or more differences in the physics configurations between any two members. In the early stages of our study, we also performed additional experiments based on other physics configurations. The Pleim–Xiu land surface model (LSM) became the preferred choice as it produced better precipitation forecasts for this case in general. We do not claim that the current configurations of the ensemble members are optimal—an important purpose of this study is to evaluate the performance of the ensemble system as configured.

As will be seen later, the skills of the individual members of the CEFS ensemble in predicting the extreme rainfall in the Beijing region are quite different. To gain some understanding on the relative sensitivity of the extreme rainfall forecast to the ICs and BCs versus the model physics, we perform another set of forecasts that share the same ICs and BCs as the most skillful member (mem13) of CEFS, but differ in the physics packages used. Table 2 lists these experiments, which can be divided into three groups using: different microphysics (MPY), different planetary boundary layer (PBL), and different radiation (RAD) parameterizations. To distinguish from the CEFS members, we refer to these experiments with the prefix "mpy". So, experiments mpy01 through mpy09 differ in the microphysics schemes used; mpy10 through mpy13 differ in the surface layer and PBL schemes used; and mpy14 through mpy18 differ in the radiation schemes used. Here, mpy01, with Goddard Lin microphysics (Tao et al., 1989), a Pleim–Xiu surface layer (Pleim, 2006), an Asymmetrical Convective Model version 2 (ACM2) PBL (Pleim, 2007), the Pleim–Xiu LSM (Pleim and Xiu, 1995), and the GFDL short- and longwave radiation schemes (Fels and Schwarzkopf, 1975), is considered the control experiment for the purpose of physics configuration.

3.2. *PM ensemble mean*

Unlike forecast variables, such as geopotential height, which often show near Gaussian ensemble distributions, precipitation forecasts tend to have a positively skewed distribution (Hamill et al., 2008). With non-Gaussian distributions and often-present location errors, a simple ensemble mean of precipitation fields tends to smooth out precipitation peak values and results in underestimation of maximum rainfall, especially for extreme rainfall events. Ebert (2001) proposed several alternatives to generating ensemble mean precipitation, including weighted averaging, median forecasts, bias reduction and the probability matching (PM) approach. Among the proposed methods, the PM mean gives the best prediction of size, shape, intensity and location of heavy rain (Ebert, 2001). It is calculated as follows:

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_m, x_{m+1}, x_{m+2}, \ldots, x_{2m}, \ldots, x_{(n-1)m}, x_{(n-1)m+1}, \ldots, x_{nm}) , \quad (1)$$

$$\bar{\boldsymbol{x}}_{\rm PM} = (\bar{x}_{1,{\rm PM}}, \bar{x}_{2,{\rm PM}}, \ldots, \bar{x}_{n,{\rm PM}}) , \quad (2)$$

Table 1. Parameter settings of the 4-km CEFS. Each member uses different ICs and lateral BCs from GEnFS and different WRF physics combinations.

Case name	ICs and BCs	Microphysics	LSM	Surface layer and PBL	Radiation scheme
g0	GFS	Milbrandt–Yau	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	CAM longwave CAM shortwave
e0	GEnFS mean	Milbrandt–Yau	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mem01	GEnFS mem01	Milbrandt–Yau	Pleim–Xiu LSM	Monin–Obukhov MYJ	GFDL longwave GFDL shortwave
mem02	GEnFS mem02	Morrison	Noah LSM	MM5 Monin–Obukhov YSU	RRTM longwave Dudhia shortwave
mem03	GEnFS mem03	Morrison	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	RRTM longwave Dudhia shortwave
mem04	GEnFS mem04	Thompson	Noah LSM	MM5 Monin–Obukhov YSU	Goddard longwave Goddard shortwave
mem05	GEnFS mem05	Morrison	Noah LSM	QNSE QNSE	CAM longwave CAM shortwave
mem06	GEnFS mem06	Ferrier	Pleim–Xiu LSM	TEMF TEMF	CAM longwave CAM shortwave
mem07	GEnFS mem07	Milbrandt–Yau	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	CAM longwave CAM shortwave
mem08	GEnFS mem08	Thompson	Pleim–Xiu LSM	Monin–Obukhov MYJ	RRTM longwave Goddard shortwave
mem09	GEnFS mem09	Thompson	RUC LSM	Monin–Obukhov MYJ	RRTM longwave Dudhia shortwave
mem10	GEnFS mem10	Morrison	RUC LSM	MM5 Monin–Obukhov YSU	RRTM longwave Dudhia shortwave
mem11	GEnFS mem11	WDM 5-class	RUC LSM	MYNN MYNN 2.5 level TKE	RRTM longwave Dudhia shortwave
mem12	GEnFS mem12	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	RRTMG longwave RRTMG shortwave
mem13	GEnFS mem13	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	CAM longwave CAM shortwave
mem14	GEnFS mem14	Goddard Lin	RUC LSM	MYNN MYNN 2.5 level TKE	RRTMG longwave RRTMG shortwave
mem15	GEnFS mem15	WSM 6-class	Noah LSM	MYNN MYNN 2.5 level TKE	Goddard longwave Goddard shortwave
mem16	GEnFS mem16	Morrison	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	GFDL longwave GFDL shortwave
mem17	GEnFS mem17	Morrison	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	CAM longwave CAM shortwave
mem18	GEnFS mem18	WDM 6-class	Noah LSM	MYNN MYNN 2.5 level TKE	Goddard longwave Goddard shortwave
mem19	GEnFS mem19	WDM 6-class	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	Goddard longwave Goddard shortwave
mem20	GEnFS mem20	WDM 6-class	RUC LSM	MYNN MYNN 2.5 level TKE	GFDL longwave GFDL shortwave

where

$$\bar{x}_{j,\mathrm{PM}} = \mathrm{Median}(x_{(j-1)m+1}, x_{(j-1)m+2}, \ldots, x_{(j-1)m+m}),\quad j = 1,\ldots,n\,, \tag{3}$$

$$\bar{\boldsymbol{x}}_{\mathrm{SM}} = (\bar{x}_{1,\mathrm{SM}}, \bar{x}_{2,\mathrm{SM}}, \ldots, \bar{x}_{n,\mathrm{SM}})\,. \tag{4}$$

Here, m is the ensemble size and n is the total number of grid points. The vector $\boldsymbol{x}$ contains the forecast rainfall amounts at all grid points and for all ensemble members, and the amounts have been sorted in descending order from largest to smallest values. The operators $\overline{()}_{\mathrm{SM}}$ and $\overline{()}_{\mathrm{PM}}$ denote the simple ensemble mean and an intermediate PM mean, respectively. To calculate the PM mean, one value, $\bar{x}_{j,\mathrm{PM}}$, is picked out of every m values from $\boldsymbol{x}$ by taking the median value according to Eq. (3). The $\bar{x}_{j,\mathrm{PM}}$ values make $\bar{\boldsymbol{x}}_{\mathrm{PM}}$ up the intermediate PM mean vector. Then, the simple ensemble rainfall amounts at every grid point are similarly ranked from largest to smallest into the vector $\bar{\boldsymbol{x}}_{\mathrm{SM}}$, but with the location information of

Table 2. Parameter settings of sensitivity tests using different physics schemes. All the experiments use the GEnFS mem13 as the ICs and BCs.

Case name	Microphysics	LSM	Surface layer and PBL	Radiation scheme
mpy01	Goddard Lin (GLin)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (PlX)	GFDL longwave GFDL shortwave
mpy02	Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy03	Thompson (Thom)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy04	SBU_YLIN (SY)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy05	Morrison (Mor)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy06	WSM 6-class (WSM6)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy07	WDM 5-class (WDM5)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy08	WDM 6-class (WDM6)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy09	Milbrandt–Yau (MY)	Pleim–Xiu LSM	Pleim–Xiu ACM2 (Pleim)	GFDL longwave GFDL shortwave
mpy10	Goddard Lin	Pleim–Xiu LSM	MYNN MYNN 2.5 level TKE	GFDL longwave GFDL shortwave
mpy11	Goddard Lin	Pleim–Xiu LSM	MM5 Monin–Obukhov YSU	GFDL longwave GFDL shortwave
mpy12	Goddard Lin	Pleim–Xiu LSM	Monin–Obukhov MYJ	GFDL longwave GFDL shortwave
mpy13	Goddard Lin	Pleim–Xiu LSM	QNSE QNSE	GFDL longwave GFDL shortwave
mpy14	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	RRTM longwave Dudhia shortwave (RRTM-Dud)
mpy15	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	RRTMG longwave RRTMG shortwave
mpy16	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	RRTM longwave Goddard shortwave (RRTM-God)
mpy17	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	Goddard longwave Goddard shortwave (God)
mpy18	Goddard Lin	Pleim–Xiu LSM	Pleim–Xiu ACM2 Pleim	CAM longwave CAM shortwave

each grid-point stored along with its rank. Finally, $\bar{x}_{j,\mathrm{SM}}$ is reassigned the value of $\bar{x}_{j,\mathrm{PM}}$ and put back to its original grid point location, and this is done for all values of j. This procedure yields the PM mean rainfall field, $\bar{\boldsymbol{x}}_{\mathrm{PM}}$. In the PM field, at the grid-point where the largest (smallest) simple ensemble mean is found, the largest (smallest) value from the intermediate PM mean vector $\bar{\boldsymbol{x}}_{\mathrm{PM}}$ is assigned. The PM mean has been found to be more skillful than the simple ensemble mean for the CAPS SSEF forecasts (e.g., Kong et al., 2008; Clark et al., 2009). In this study, the PM mean of 24-h accumulated rainfall will be presented and compared with the simple ensemble mean precipitation.

3.3. *Verification methods*

In this study, the ensemble forecasts of CEFS are verified against surface and upper-air observations. The GEnFS forecasts are used as a reference. For the verification, we employ the Model Evaluation Tools developed by the U.S. Developmental Testbed Center (Brown et al., 2009), which contains a comprehensive suite of verification metrics for both deterministic and ensemble-based probabilistic forecasts. Two domains are chosen for our verification: one is the entire forecast domain covering the whole of mainland China (referred to as "FULL"), which serves to verify the forecasts of environmental conditions for the Beijing extreme rainfall event; and the other is enclosed by the magenta box labeled A in Fig. 1c. This domain covers Beijing and its surrounding areas, and is mainly used to verify precipitation forecasts in the Beijing region.

Over 20 000 surface rain gauge observations over China are collected, quality controlled and used to verify the precipitation forecasts. The verifications are carried out in both the "FULL" and "Box-A" domain. Ensemble-based verification scores, such as the rank histogram (Hamill, 2001) and Brier

Score (Murphy, 1973), are used to evaluate the probabilistic forecasting skills of CEFS. Upper-air sounding observations are used to verify the forecast RH, temperature (T), and model zonal and meridional wind components (U and V, respectively), while surface observations are used to verify the 2-m RH and T, and the 10-m U and V; and these verifications are performed in the "FULL" domain.

4. Evaluation of CEFS ensemble forecasts

4.1. *Subjective evaluation of precipitation forecasts*

Figure 2 shows the postage stamp charts of 24-h accumulated rainfall of individual ensemble members together with observed rainfall, from 0000 UTC 21 July through 0000 UTC 22 July 2012, within the "Box-A" domain. The observed

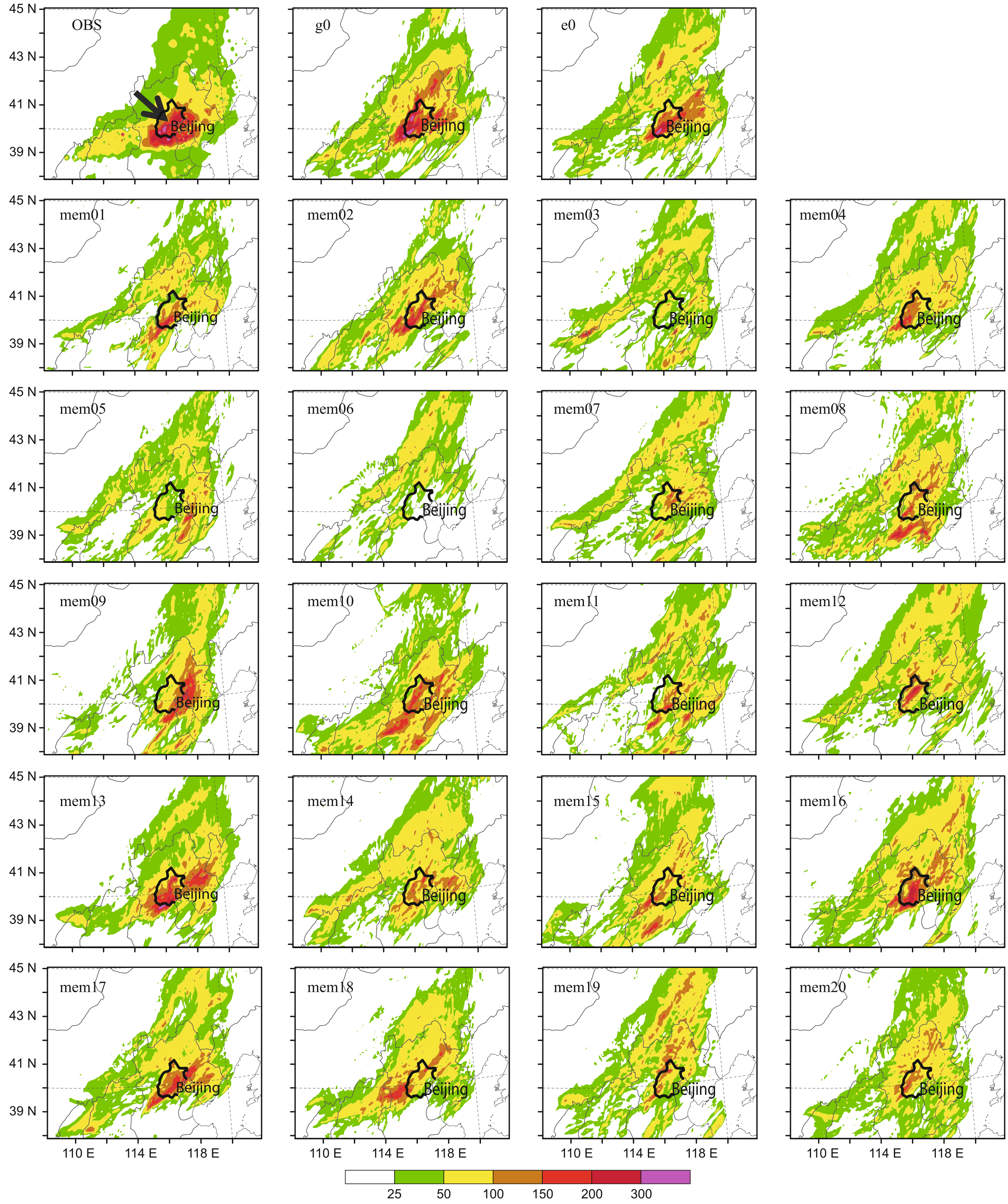

Fig. 2. Postage stamp plots of 24-h accumulated rainfall (units: mm) from 0000 UTC 21 July to 0000 UTC 22 July 2012, observed (OBS), and from the 4-km CEFS members (see Table 1 for the definition of members). In this, and some of the later figures, Beijing City is shown by the bold-black border.

heavy rain center, with a maximum of 382 mm, is located in southwest Beijing (as pointed to by the black arrow in Fig. 2a). The forecasts show a high level of diversity among the members. Most members predict rainfall of more than 100 mm d^{-1}, but the location and spatial coverage vary. Heavy rain exceeding 250 mm d^{-1} is captured by members g0, e0, mem13 and mem16, and the maximum centers are generally located in south or southwest Beijing, close to the observed maximum location. A maximum of 452 mm is predicted by member g0, which is the member initialized from the operational GFS analysis. We will refer to these four members as "good" members.

In contrast, members such as mem06 and mem03 only predict small patches of rainfall exceeding 100 mm, and very few patches are located within the city of Beijing. In fact, the maximum amount predicted by mem06 over Beijing is less than 50 mm, while the maximum barely exceeds 100 m over Beijing according to mem03. Other members predict maximum rainfall amounts of between 100 and 250 mm, but the maximum centers have various displacement errors. In mem05, the precipitation system seems to have moved too rapidly towards the southeast, while that in mem12 seems to have moved a little too slow, placing the maximum accumulated precipitation too far southeast and northwest of Beijing, respectively. Overall, about half of the members predict heavy precipitation of more than 200 mm over a large enough area close to Beijing, while the other half predict precipitation that is too weak.

The spaghetti plots of 100 and 150 mm 24-h rainfall contours from the ensemble members are shown in Fig. 3. For

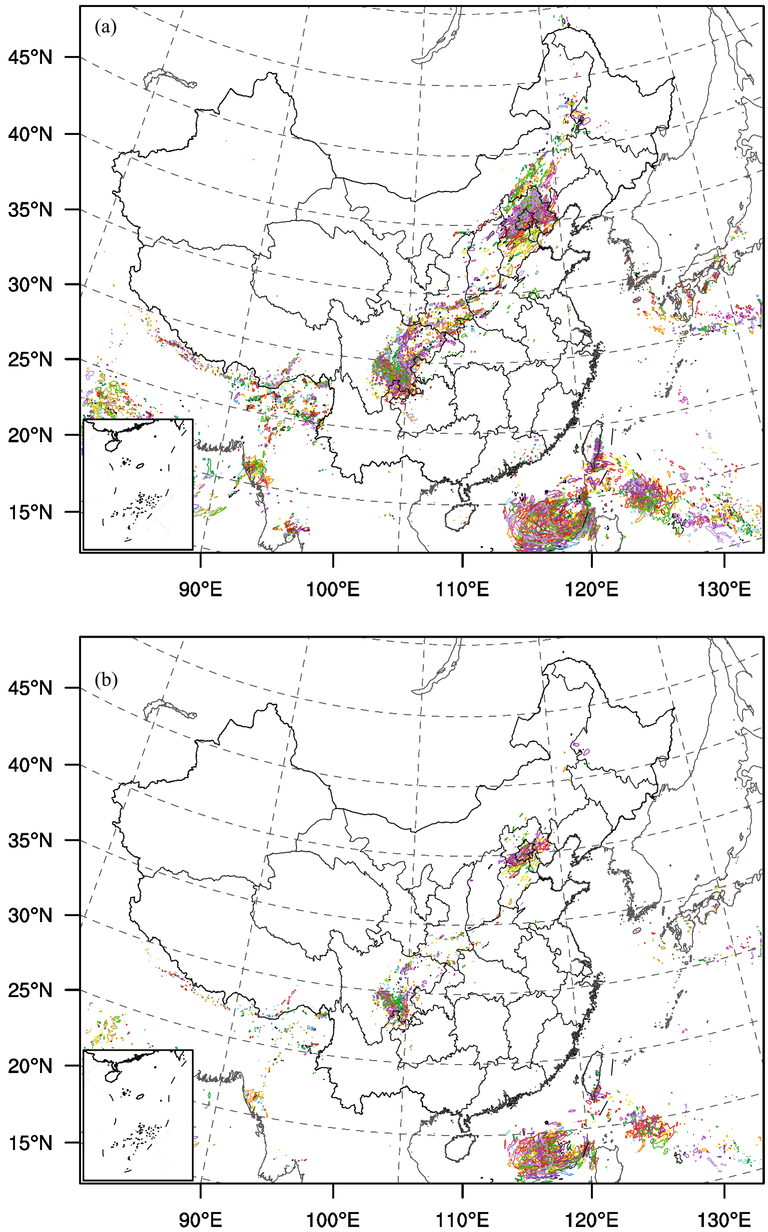

Fig. 3. Spaghetti plots of 24-h accumulated rainfall contours of (a) 100 mm and (b) 150 mm. Different colors represents different members.

the 100-mm contours, most members cluster over the Beijing area and several are scattered around Beijing, indicating a high degree of certainty in predicting medium to heavy precipitation by the ensemble, although uncertainty does exist (Fig. 3a). There are two other clusters of 100-mm contours in the forecast domain—one in the eastern part of Sichuan Province, which is at the southern end of the frontal system stretching through the Beijing region; and the other over the South China Sea, which is associated with typhoon Vincente. For the 150-mm rainfall, the contours cluster quite tightly over the Beijing area, and there are more contours over the southern part of Beijing (Fig. 3b), suggesting a high probability of heavy rain in Beijing. The ensemble probabilities of precipitation (POPs) exceeding 100 mm and 150 mm are shown in Fig. 4 for the zoomed "Box-A" domain. While the spaghetti plots give a subjective view of the precipitation patterns and distributions, the POP provides quantitative probabilistic forecasts. For this event, the maximum POP is 82% for 100 mm d^{-1} and 50% for 150 mm d^{-1}. The highest

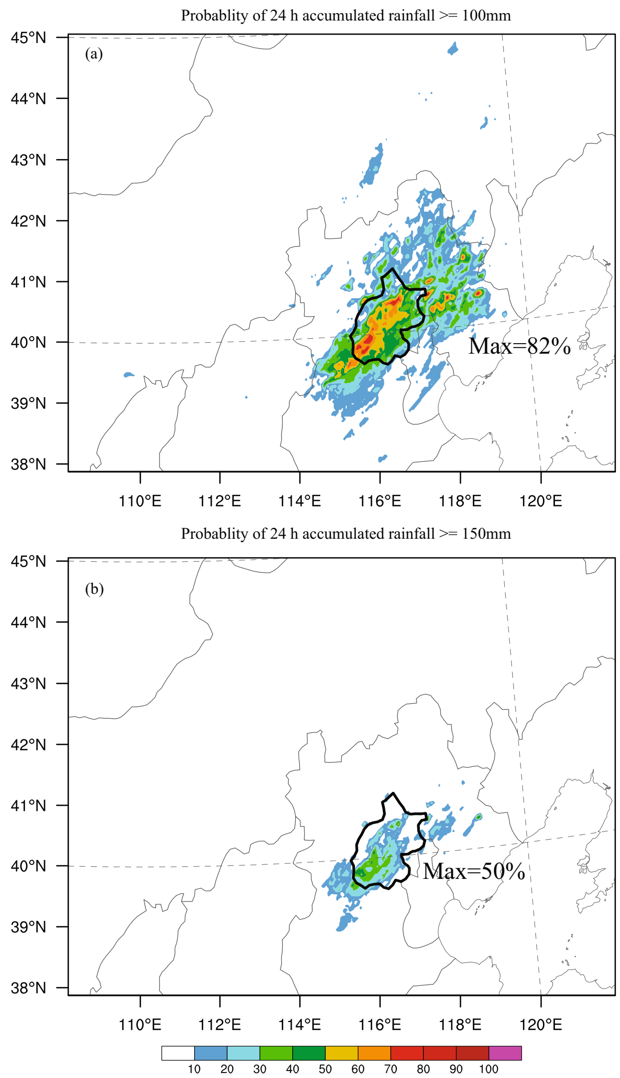

Fig. 4. Probability of 24-h accumulated rainfall of (a) ⩾100 mm and (b) ⩾150 mm.

probability is located in the southwest part of Beijing for both thresholds, which is consistent with the observed heaviest rainfall. Such ensemble forecast products, if available in real time, would have greatly enhanced forecasters' confidence in the occurrence of very heavy rainfall in Beijing.

Figure 5 shows the simple ensemble mean and PM mean 24-h precipitation forecasts together with observed precipitation. Given the diversity of the members in predicting the intensity and location of the maximum rainfall, the simple ensemble mean significantly underestimates the rain intensity, giving an ensemble mean maximum of only 151 mm d^{-1} (Fig. 5b). The PM ensemble mean, however, is able to capture the maximum intensity as well as the spatial distribution of the heavy precipitation much better; the PM mean maximum is as high as 415 mm d^{-1} and is located correctly near the southwest corner of Beijing, matching the observations very well. Based on the PM mean algorithm, the PM mean maximum is usually close to the maximum value predicted by all ensemble members at all grid-points. For this reason, the choice of domain used for calculating the PM mean does matter, and the calculation domain should be chosen to cover only the relevant precipitation systems, as we do here with the "Box-A" domain.

4.2. *Quantitative evaluation of the ensemble forecasts*

In this subsection, CEFS forecasts are evaluated using the GEnFS as a reference (the much coarser-resolution GEnFS is not expected to out-perform the convection-permitting CEFS ensemble, but it helps to have a reference when evaluating the performance. NCEP global forecast products are used routinely by forecasters in China as one of the references when producing operational forecasts). Figure 6 shows the maximum and the 5th-percentile rainfall amounts within the "Box-A" region, and the rank histograms of 24-h precipitation forecasts of the CEFS members and the corresponding GEnFS members used to provide the CEFS ICs and BCs. The use of the 5th-percentile rainfall amount here helps alleviate the effects of precipitation biases among the members, as is discussed in Zhu et al. (2015).

For the maximum rainfall, GEnFS, not surprisingly, significantly underestimates the amount. The gridded maximum value of observations is 382 mm, while very few members of GEnFS predict more than 150 mm of rainfall (Fig. 6b). The intensity is greatly improved in CEFS. All except one member predict more than, or very close to, 200 mm of rainfall, though only 6 members predict close to or more than 300 mm. Member e0 predicts a maximum that is closest to the maximum in the objectively analyzed rainfall field, while g0 and mem13 predict maximum values of 440–450 mm, which are close to the station-observed maximum of 460 mm (see Chen et al., 2012).

For the 5th-percentile rainfall, which represents the lower end of the rainfall intensity, all GEnFS members over-predict the light-rain amount (Fig. 6d), while the predictions of CEFS are distributed around the observed value of about 0.7 mm (Fig. 6c), indicating that CEFS members also predict light rain better. For the rank histograms of rainfall, GEnFS shows

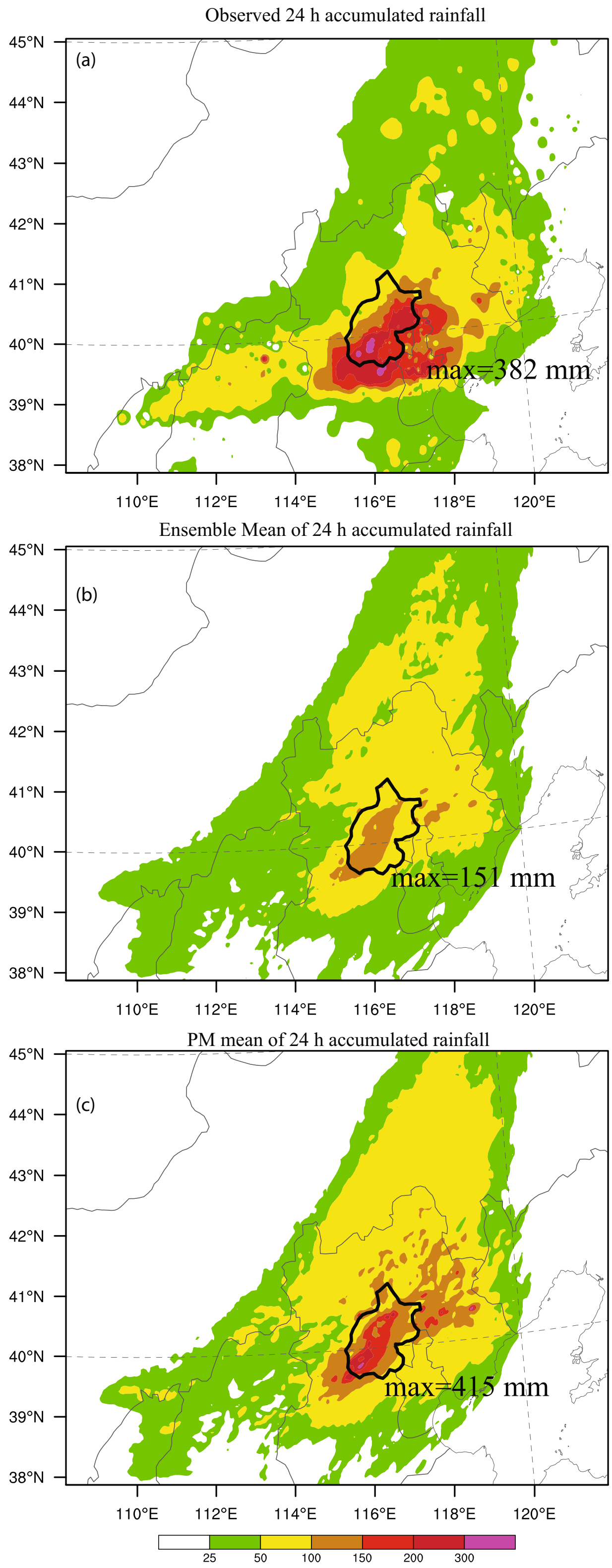

Fig. 5. 24-h accumulated rainfall (units: mm) from (a) rain gauge observations, (b) the simple ensemble mean, and (c) the PM ensemble mean.

pronounced and more or less symmetric "U" shapes (Fig. 6f) for both the "FULL" and "Box-A" domains, indicating underprediction of high precipitation values and over-prediction of low precipitation values. The ensemble system is seriously under-dispersive for rainfall. For CEFS (Fig. 6e), the occurrence frequencies near both ends of the histograms are greatly reduced for both domains, and more so for the "Box-A" region. Most interestingly, the histogram of CEFS shows an almost uniform distribution for the "Box-A" region, apart from a slight underestimation of high values. Larger intensity forecast errors in other parts of the domain contribute to more underestimation of high values for the "FULL" domain.

The rank histogram informs us of the ensemble forecast distribution as compared to that of observations; it does not, however, reveal the sharpness of the ensemble forecasts, which is also an important property of an ensemble. Hamill (2001) suggests that it should be used in conjunction with other probabilistic ensemble skill scores. Here, the Brier score and its components for reliability, resolution and uncertainty (Stephenson et al., 2008) are calculated for the 24-h accumulated rainfall for the "Box-A" region and given in Fig. 7. The Brier score is similar to the RMSE but is calculated as the mean-square difference of forecast probability and that of corresponding observations. It can be decomposed into three components representing reliability, resolution and uncertainty of the forecast. For the Brier score and its reliability component, a smaller value is better; while for resolution, a larger value is better. The value of uncertainty is not related to the forecast but is a function of the frequency of the events occurring. For the 50 mm d^{-1} threshold, the Brier scores and reliabilities of CEFS are all smaller (better) than those of GEnFS, while its resolution is higher. This indicates that the precipitation probability forecast of CEFS is better than that of GEnFS. For the 100 mm d^{-1} threshold, CEFS again has a smaller Brier score overall (∼0.1 versus ∼0.157) and higher resolution (∼0.07 versus ∼0.01) when compared to GEnFS. However, the reliability value is somewhat higher than that of GEnFS. This appears to be due to the overestimation of the rainfall area above 100 mm d^{-1} (see Fig. 3a), but the reliability difference is much smaller than that of the resolution. Overall, the Brier score and its components show that CEFS is better than GEnFS for the probabilistic forecasting of heavy precipitation for the Beijing region in this case.

We also check the ensemble forecasts of other atmospheric variables. They are verified against surface and sounding observations within the "FULL" domain. In general, for RH, temperature, and wind components, CEFS yields a larger ensemble spread but smaller RMSEs than GEnFS, especially at the surface and lower levels (not shown). The rank histogram distributions of surface variables show that CEFS, while still under-dispersive, greatly reduces the forecast sampling biases at both low and high ends, especially for the longer range forecasts (again not shown).

4.3. *Precipitation in the warm-sector region*

As mentioned in the introduction, most operational models missed the precipitation in the pre-frontal warm sector over Beijing before 1200 UTC 21 July 2012. Figure 8 shows the observed and CEFS forecasted hourly rainfall amounts valid at 0600 UTC 21 July 2012, together with the forecasted

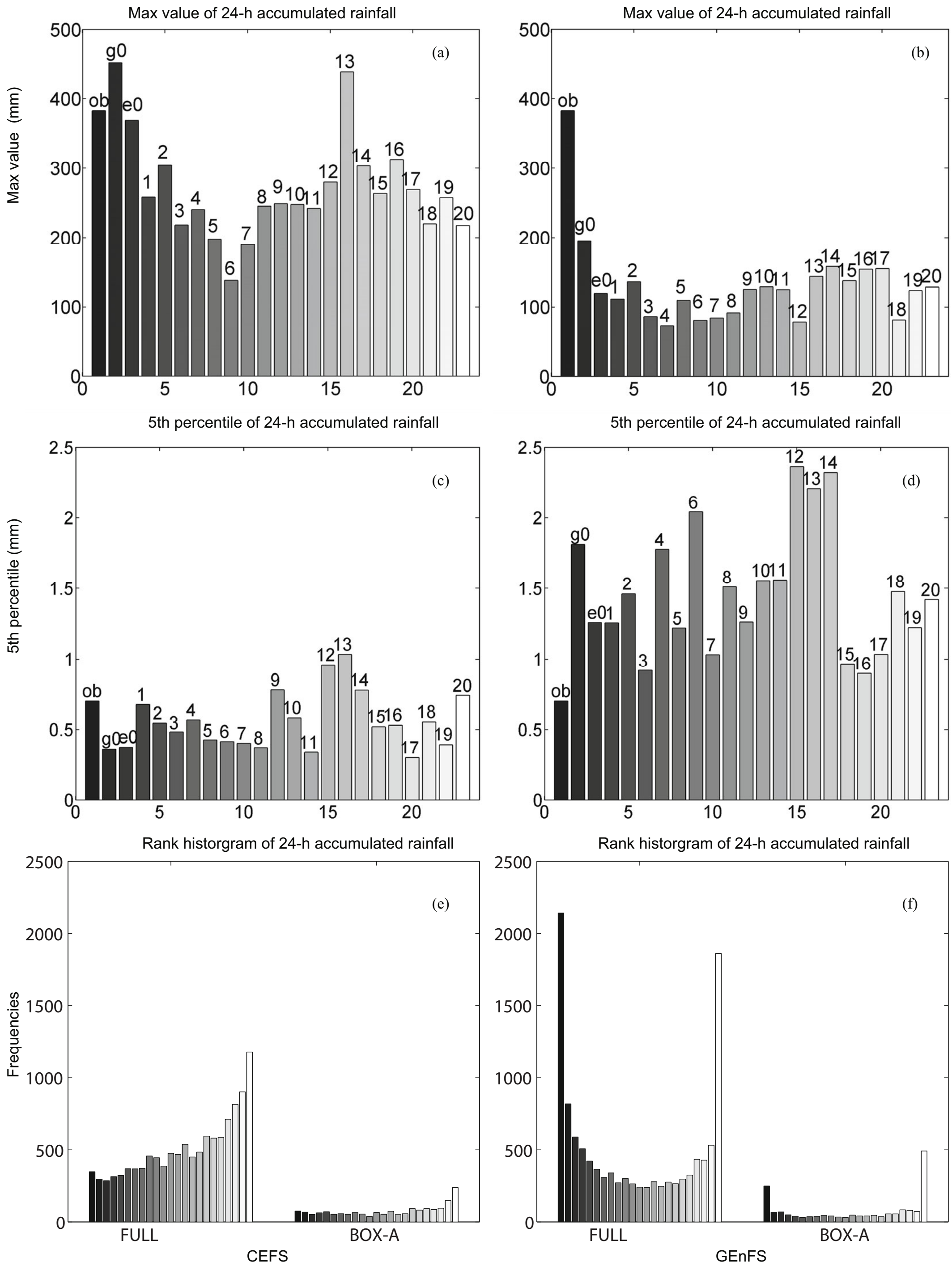

Fig. 6. 24-h accumulated rainfall amounts for (a, c, e) CEFS and (b, d, f) GEnFS members for the (a, b) maximum values and (c, d) 5th-percentile values within the "Box-A" region, and (e, f) the rank histograms for CEFS and GEnFS. The rank histograms are plotted for both the "Box-A" and "FULL" regions.

surface winds. The forecasts for the two "control" members g0 and e0, and from those of the "good" member mem13 and "bad" member mem06, are shown. As can be seen from Fig. 6d, mem13 and mem06 have the best and worst maximum and light rainfall forecasts overall among the CEFS members, and are therefore examined in more detail here. At this time, a cold front is located about 200 km west of Beijing, and a band of light to moderate rain can be found along the front (Fig. 8a). The band of frontal precipitation is predicted by all of the members shown in Fig. 8, including mem06. In the observations, however, the most significant precipitation at this time is found over and to the southwest of Beijing (Fig. 8a), and this area of precipitation has been referred to as the (pre-frontal) warm-sector precipitation (Chen et al., 2012,

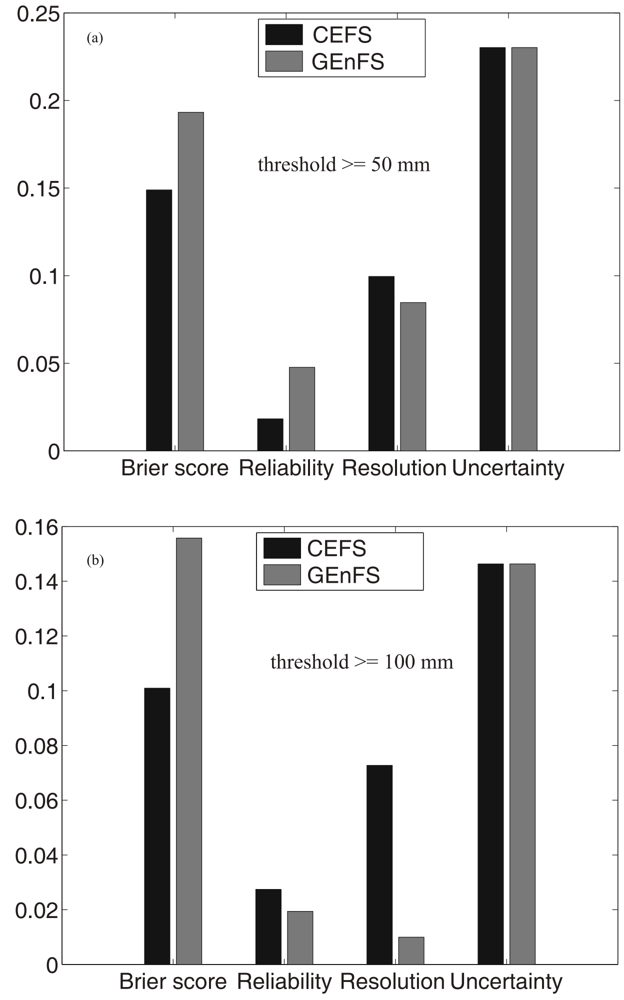

Fig. 7. Brier scores and their components for 24-h accumulated rainfall for the "Box-A" domain for (a) threshold ⩾50 mm and (b) threshold ⩾100 mm.

Zhang et al., 2013), which marked the onset of heavy precipitation over Beijing. It, together with the later, quasi-linearly organized MCS, contributed significantly to the total amount of precipitation.

The warm-sector precipitation at this stage is best predicted by mem13 (Fig. 8d), followed by g0 (Fig. 8b), and is almost completed missed by mem06 (Fig. 8e). The prediction of e0 is also relatively poor for the warm-sector rain at this time. Yu and Meng (2016) study showed that the low-level (850 hPa) low contributed most to the heavy rain in Beijing. The heavy rainfall was likely caused by strong low-level lifting. To gain some further understanding, we show in Fig. 9 the 850-hPa wind and horizontal water vapor flux fields of mem13 and mem06 at 0000 and 0600 UTC. The star symbol in Fig. 9 indicates the location of the observed maximum 24-h accumulated rainfall, which is in the southwest part of Beijing City, at the eastern edge of the mountain range that has a mean height of about 1.5 km (the mountain range is part of the large Taihang Mountain Range).

At 0000 UTC, both mem13 and mem06 show a concentrated band of water vapor flux coming from the south into the Beijing region. In mem13, the winds associated with the band are mostly southerly, with a slight easterly component (Fig. 9a), but those in mem06 have a noticeable westerly component, making the flow more parallel to the southwest–northeast-oriented Taihang Mountain Range (Fig. 9b), reducing the range-normal wind component. As a result, there is more convergence of flows on the eastern slope of the Taihang Mountain Range, south of the maximum precipitation spot, due to mountain blocking, in mem13, giving rise to strong moisture flux convergence there (Fig. 9a). Meanwhile, in mem06, the more-or-less mountain-range parallel flow moves further north until it encounters the eastward-extending Yanshan Mountains north of Beijing City, creating large moisture flux convergence south of the Yanshan Mountains. For such flow configurations, one would expect much more precipitation on the upwind slope of the Taihang Mountains in mem13 at the location of the largest moisture flux convergence, as is the case in the forecast.

By 0600 UTC, there is a significant enhancement to the southerly and southeasterly flows at 850 hPa, with a corresponding increase in the moisture flux, in mem13 (Fig. 9c). To the southeast of the maximum precipitation point, the flows are almost all southeasterly, creating an almost 90° incident angle to the Taihang Mountain Range, and strong lifting of low-level moist air. As a result, a more-or-less southwest–northeast-oriented band of heavy precipitation forms near the foot of the mountain range (Fig. 8d), in a way very similar to observations (Fig. 8a). In comparison, the much more southerly flows found in mem06 (Fig. 9d) create little rainfall along the Taihang Mountain Range; rainfall stronger than observed is instead created along the cold front to the northwest, apparently due to less depletion of moisture by precipitation before the air stream reaches the cold front. These examinations suggest that the environmental flows south and southeast of the Beijing region, and especially the flow directions before 0600 UTC 21 July, together with the associated orographic lift, played important roles in producing the pre-frontal, warm-sector heavy rainfall at this stage, which marked the onset of the extreme rainfall event in Beijing. Properly sampling the uncertainties in the environmental conditions is important for convective-scale ensemble forecasting. In the next section, we further explore the sensitivity of the precipitation forecast to the model physics.

5. Sensitivity of precipitation to physics schemes

The second set of sensitivity experiments listed in Table 2 aims to examine the relative sensitivity of extreme precipitation to model physics. Figure 10 shows the average hourly precipitation within a domain from 38.5°N to 41.5°N and from 114.5°E to 117.5°E, where heavy rainfall occurred. The area is a little larger than the Beijing region in order to account for the forecast position error. For the observed rainfall, there are two rapid rainfall intensification periods: from

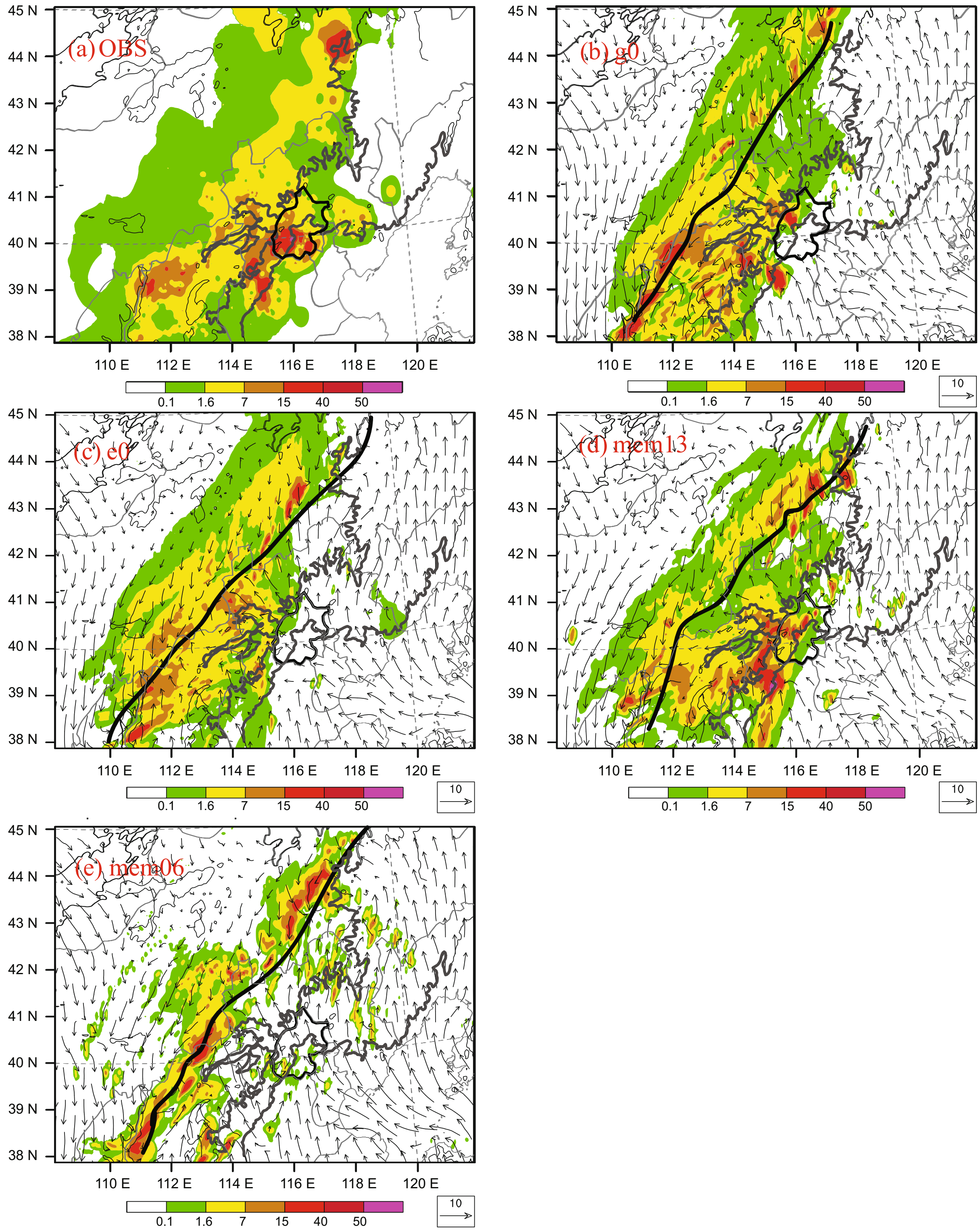

Fig. 8. (a) Observed hourly rainfall at 0600 UTC 21 July 2012 and (b–e) forecasted hourly rainfall and surface winds for members g0, e0, mem13 and mem06, respectively. The single thick black line in each panel indicates the convergence line in the forecasted surface winds, which roughly corresponds to a surface cold front. The black contours of medium thickness represent the terrain heights of 100 m and 1000 m.

0300 to 0700 UTC and from 0900 to 1300 UTC. The former is related to the first stage of warm-sector rainfall from less-organized convective cells, while the latter occurred when the well-organized quasi-linear MCS with back-building and echo-training characteristics was established (Zhang et al., 2013). Member mem13 of CEFS successfully captures the rapid intensification in both periods with a very good timing of onset. It ends the rapid intensification at 1000 UTC,

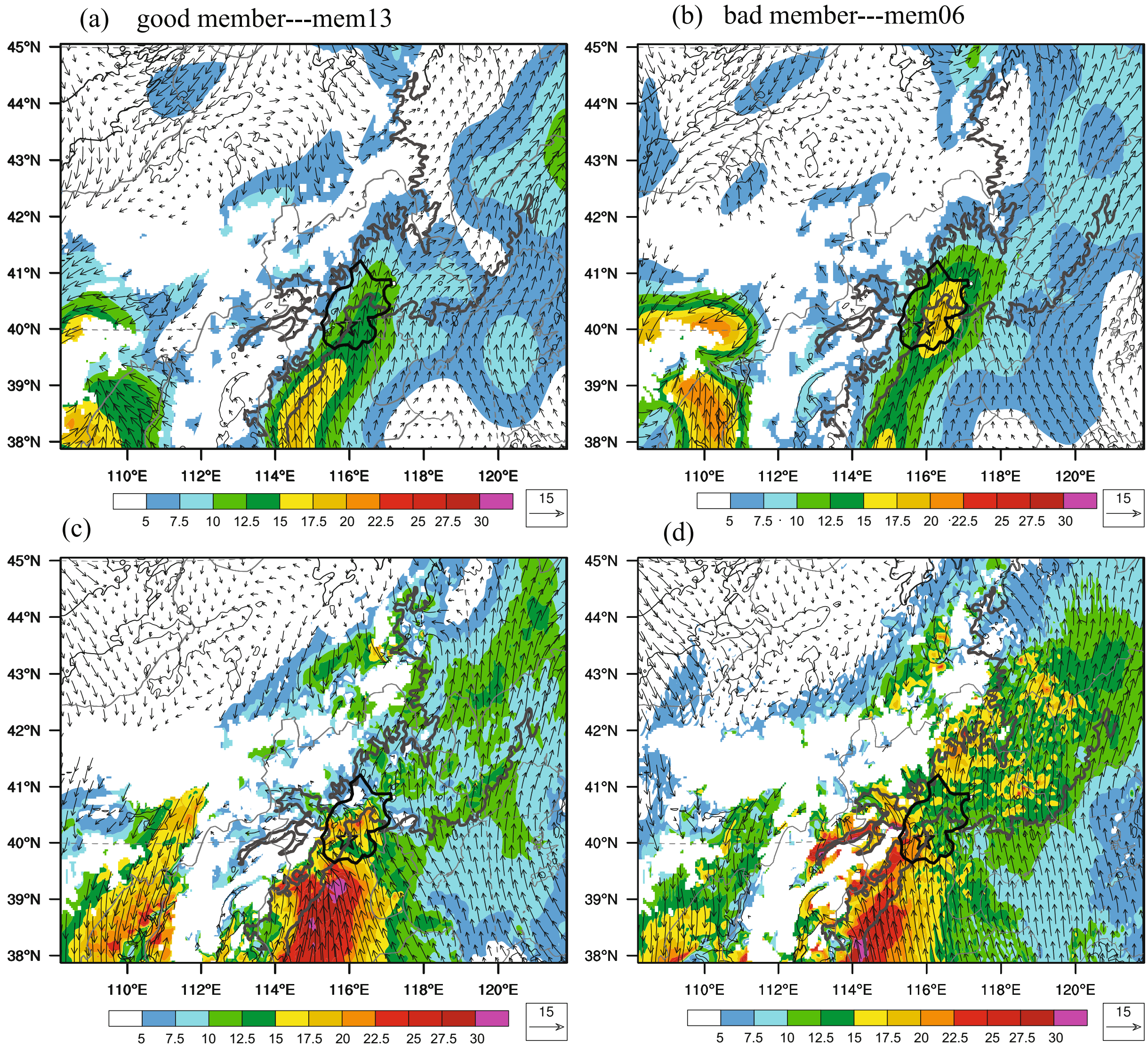

Fig. 9. The 850-hPa wind vectors (units: m s^{-1}) and horizontal water vapor flux (color-shaded) fields [units: g (hPa cm s)$^{-1}$] at the (a, b) IC time of 0000 UTC 21 July 2012 and (c, d) 6-h forecast time of 0600 UTC 21 July 2012, of the (a, c) "good" member mem13 and (b, d) "bad" member mem06. The thick dark-gray contours indicate the terrain heights of 100 m and 1000 m, respectively.

one hour earlier than observed. Member g0 produces a larger domain-average maximum hourly precipitation than mem13, but the rapid intensification phase is delayed by two to three hours. It produces a total amount of precipitation within the average domain that is closer to observation. For most other members, the total amount of precipitation is significantly underestimated, and the onset of heavy precipitation is also significantly delayed.

Figures 10b–d show the results of the sensitivity experiments with different physics schemes but with the same ICs and BCs as mem13. It can be seen that, among all the physics sensitivity experiments, the average precipitation differences in the first nine hours of the forecast are small. As a comparison, the average hourly precipitation is very similar up to 1000 UTC when the forecast intensity peak is reached, except for the WDM5 (mpy07) and WDM6 (mpy08) members, which reach lower peaks one hour earlier (Fig. 10b). There is little sensitivity of precipitation to radiation physics throughout the forecast period (Fig. 10d). Among the PBL schemes, the Pleim–Xiu scheme produces more sustained precipitation (Fig. 10c). There is a much larger sensitivity of precipitation after 0900 UTC among the microphysics members, which is not very surprising (Fig. 10b) since microphysics has a much bigger opportunity to affect the precipitation process when the precipitation system is fully developed. Despite the noticeable sensitivity to the microphysics during the later period, the overall precipitation timing and amount, however, have much smaller variability across the physics-difference members than the members of CEFS that have both IC and BC differences and physics differences. These results clearly indicate that, for the Beijing extreme precipitation case, the synoptic and mesoscale environmental conditions, as deter-

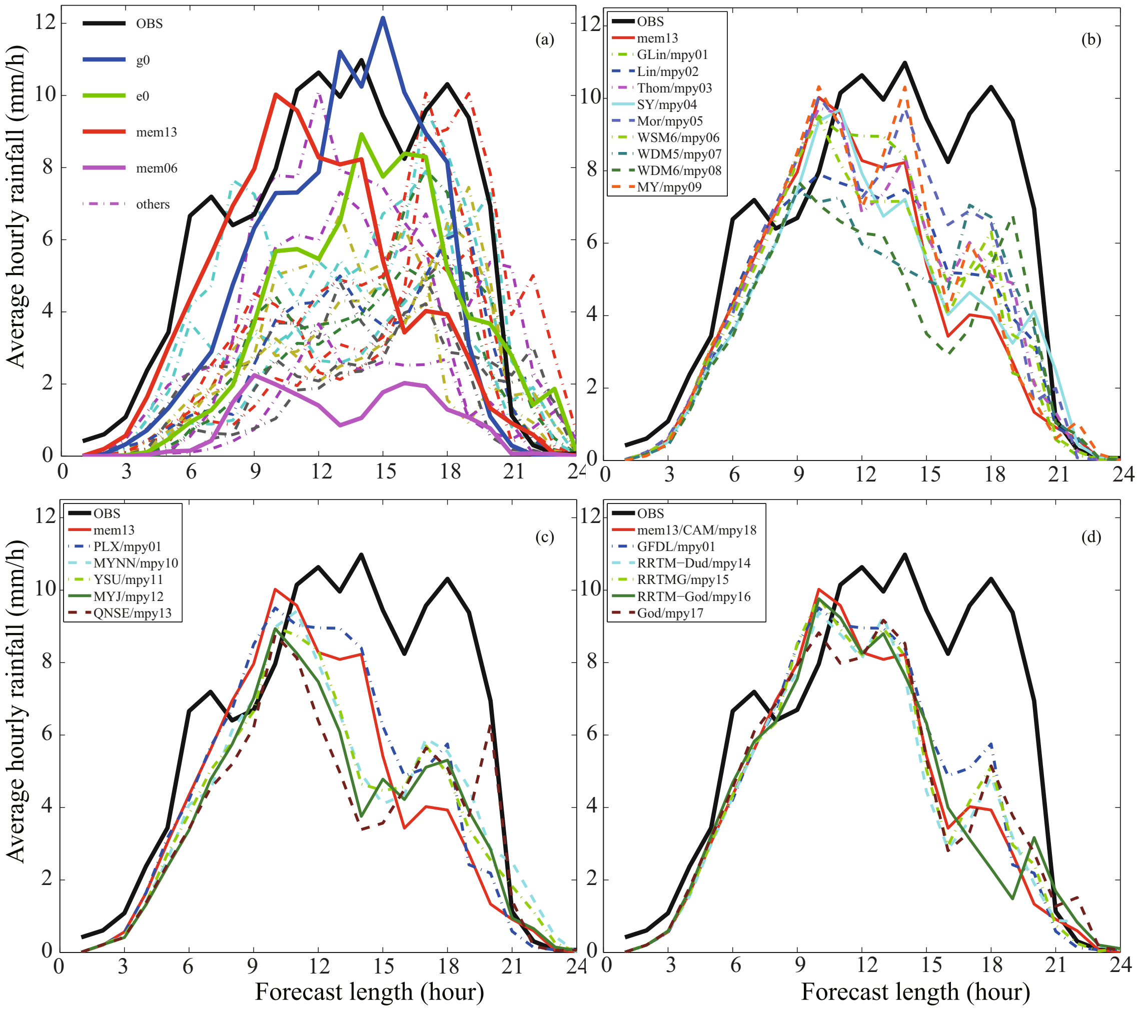

Fig. 10. Time series of average hourly rainfall in a domain from 38.5°N to 41.5°N and 114.5°E to 117.5°E for (a) observation and ensemble members of CEFS, and for mem13 and members of the set of sensitivity forecasts (see Table 2) with (b) different microphysics schemes, (c) different PBL schemes, and (d) different radiation schemes. The sensitivity forecasts all use the same ICs and BCs as mem13.

mined by the ICs and BCs, have a much larger influence on the forecast precipitation, which is consistent with our earlier discussions when comparing the results of CEFS members. Given the rather large (full-China domain) computational domain used and the relatively short forecast time examined, the ICs should have a much larger impact than the BCs in this case.

The time series of the observed maximum accumulated rainfall at a meteorological station (the 460-mm maximum was reported by a hydrological station whose time series data are not available), and the predicted accumulated rainfall at the grid-point with the largest 24-h rainfall (of each member) within a 100 km radius of the observed maximum, are plotted in Fig. 11 for CEFS members and the physics sensitivity forecasts. We tested search radii up to 800 km; most members have the same maximum locations, except for a few "bad" members such as mem06.

Consistent with the hourly rainfall, the observed station maximum has two main rain accumulation stages (Fig. 11a). Among all CEFS members, mem13 performs best. It successfully captures both stages, though there are delays of two to four hours. The second best member is g0. It predicts the largest rainfall amount, although most is contributed by the second stage. The precipitation of the first stage is significantly underestimated. Most other CEFS members perform similarly to member g0 for the first stage, but significantly under-predict the rainfall amount in the second stage. Therefore, most of them predict no more than 300 mm of rainfall. Figures 11b–d are the results of different physics schemes. The maximum difference among all the members in each group is also given in each panel. For the precipitation at the maximum accumulation locations, different microphysics and PBL schemes produce a high level of diversity, even in the first nine hours of the forecast. The maximum differences in these two groups are 164 mm and 130 mm, respectively (Figs. 11b and c). The differences among the forecasts with different radiation physics are much smaller (Fig. 11d).

Figure 12 shows the locations of the maximum accumu-

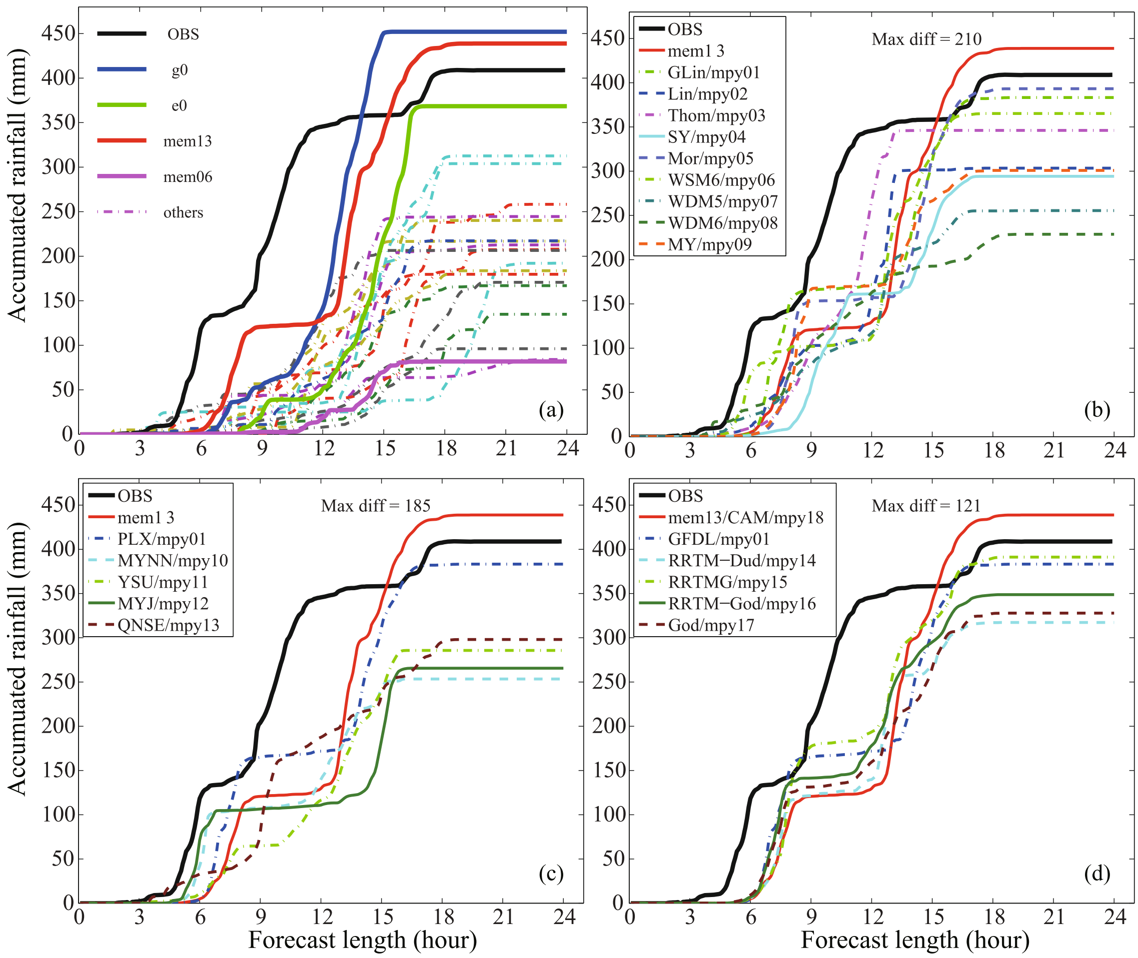

Fig. 11. Time series of observed and predicted accumulated rainfall at the grid-point with largest 24-h accumulated rainfall (of each member) within a 100-km radius of the observed maximum, from (a) members of CEFS, and for mem13 and members of the set of sensitivity forecasts (see Table 2) with (b) different microphysics schemes, (c) different PBL schemes, and (d) different radiation schemes.

lated rainfall for all the forecast members. The observed maximum is located at the foot of the Taihang Mountains (marked by the star in Fig. 12). For most CEFS members, the maximum locations are generally close to the Taihang Mountains, although there is significant scattering in the locations (Fig. 12a). The "good" member g0 has a better maximum location than the "good" member 13. There is also scattering in the maximum locations with the microphysics and PBL members, but overall the scattering is much smaller than that of full perturbation members in CEFS. Compared to domain-average rainfall, the intensity and location of the maximum rainfall have clearly larger forecast uncertainties.

6. Summary and future plan

This paper studies the extreme rainfall event that occurred in Beijing on 21 July 2012 (maximum rainfall of 460 mm) using a convection-permitting ensemble forecasting system, CEFS. The system consists of 22 members based on the WRF model and uses multiple physics schemes. Its domain covers the whole of mainland China and has a 4-km horizontal grid spacing and 51 vertical levels. The forecasts of this extreme rainfall event are evaluated in terms of both deterministic and probabilistic forecasting.

For this event, CEFS predicts a high probability of torrential rain in the Beijing region, especially in its southwestern part, consistent with observations. The predicted highest probability of 100 mm d^{-1} and 150 mm d^{-1} precipitation is 82% and 50%, respectively. The highest predicted value of the PM ensemble mean is 415 mm d^{-1}, while the simple ensemble mean gives only 151 mm d^{-1}. Note that we did not perform any calibration for the ensemble. Such a high forecast probability, together with other ensemble forecast products, if available in real-time, would have been very useful for decision-making and public warning for this historically extreme heavy rain event that caused the loss of 79 lives.

The precipitation forecasts of CEFS are evaluated using ensemble-based verification scores. The precipitation forecasts of a global ensemble forecasting system initialized from EnKF data assimilation (GEnFS) are used as a reference. Two verification domains are used. One is the "FULL" domain covering the whole of mainland China, and the other is

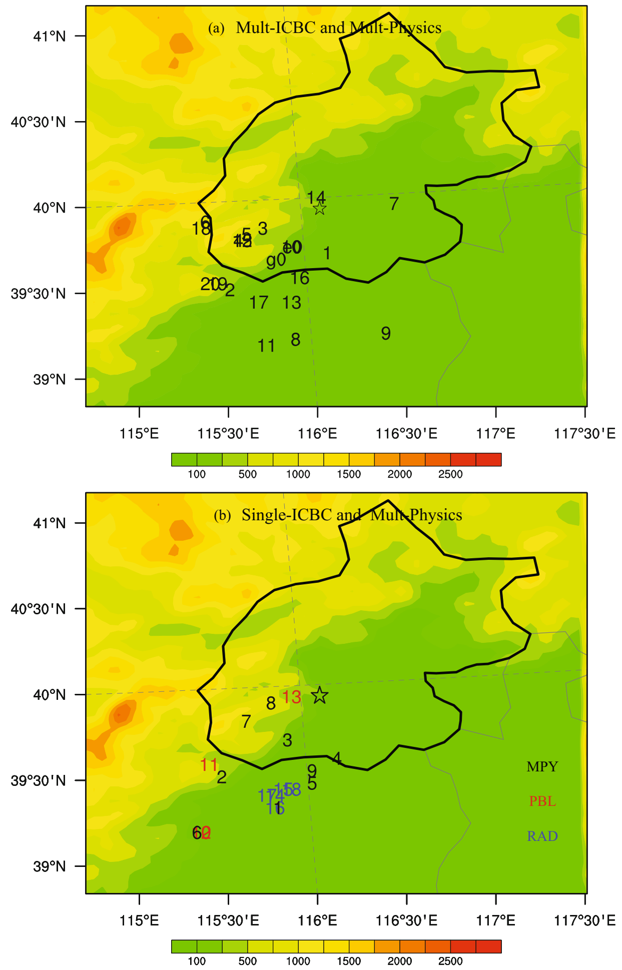

Fig. 12. Locations of maximum 24-h accumulated rainfall of (a) observed and ensemble forecasts of CEFS, and (b) ensemble forecasts with different physics schemes. In (b), the color of the numbers represents different groups.

a small domain covering Beijing and its surrounding areas. For both verification domains, GEnFS displays pronounced "U"-shaped rank histograms, while CEFS exhibits shapes that are almost flat, especially for the small verification domain, though there is still a small level of under-prediction of high values. In terms of probability verification scores, CEFS achieves a lower (higher) Brier score and a higher resolution than GEnFS, indicating that the probability forecasts of CEFS are more reliable than those of GEnFS.

For this extreme rainfall event, most of the rainfall came from convection in MCSs that occurred in the warm sector east of an approaching cold front. A few members of CEFS successfully reproduced the MCS precipitation. A "good" member from CEFS is specifically compared with a "bad" member that produced too little rain in the warm sector. It is believed that the first-stage precipitation in the warm sector was mostly induced by orographic lift. For the "good" member, there is more of an easterly component in the low-level southeasterly flow that brought in rich moisture to converge onto the southeast-facing slope of the southwest–northeast-oriented Taihang Mountain Range on the west side of Beijing, producing strong orographic lift. In contrast, the low-level flow in the "bad" member is mostly southerly, and is weaker. The low-level moisture is therefore transported further north of Beijing, producing little rain in the Beijing area. These results indicate a strong sensitivity of the extreme rainfall in this event to the mesoscale environmental conditions.

The relative sensitivity of the forecasted precipitation to model physics is investigated by running another set of 18 forecasts using the ICs and BCs as the best member of CEFS, but with different microphysics, PBL and radiation schemes. The precipitation forecast is found to be not very sensitive to the radiation scheme used, but sensitive to the microphysics scheme after the peak rainfall intensity is reached. The Plaim–Xiu PBL scheme produces more precipitation than other PBL schemes. There is more sensitivity to the physics parameterizations in terms of the maximum precipitation amount and location than domain-average rainfall. Overall, environmental conditions as given by the ICs produce more precipitation diversity than physics parameterizations.

In this study, although some of the members of our 4-km CEFS perform quite well in capturing the extreme rainfall that occurred in Beijing on 21 July 2012, demonstrating higher skill than the coarse-resolution GEnFS in terms of ensemble forecasting, large uncertainties still exist across the ensemble members. Some members completely miss the warm-sector rainfall and even the "good" members have intensity and position errors. Ensemble-based methods assimilating high-resolution local observations will likely both increase the forecast accuracy and reduce the forecast uncertainty. Also, ensemble probabilistic calibration, if properly performed, can further improve the probabilistic forecasting skill. However, ensemble probabilistic calibration of precipitation suitable for predicting extreme rainfall will require large data samples. In this paper, we did not attempt to fully analyze and understand the physical processes responsible for the extreme rainfall, or the successes and failures of individual members in predicting them. We plan to further analyze this dataset in future work to gain greater insights along these lines.

Acknowledgements. This work was primarily supported by the National Fundamental Research (973) Program of China (Grant No. 2013CB430103) and the Special Foundation of the China Meteorological Administration (Grant No. GYHY201506006). The work was also supported by the National Science Foundation of China (Grant No. 41405100).

REFERENCES

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722.

Brown, B. G., J. H. Gotway, R. Bullock, E. Gilleland, and D. Ahijevych, 2009: The model evaluation tools (MET): Community tools for forecast evaluation. *Proc. 25th Conf. Int. Inter-*

active Information and Processing Systems (*IIPS*) *for Meteorology*, *Oceanography*, *and Hydrology*, Phoenix, AZ, US., American Meteorological Society, Paper 9A.6.

Buizza, R., and A. Hollingsworth, 2002: Storm prediction over Europe using the ECMWF ensemble prediction system. *Meteorological Applications*, **9**, 289–305.

Buizza, R., D. S. Richardson, and T. N. Palmer, 2003: Benefits of increased resolution in the ECMWF ensemble system and comparison with Poor-man's ensembles. *Quart. J. Roy. Meteor. Soc*., **129**, 1269–1288.

Charles, M. E., and B. A. Colle, 2009: Verification of extratropical cyclones within the NCEP operational models. Part II: The short-range ensemble forecast system. *Wea. Forecasting*, **24**, 1191–1214.

Chen, Y., J. Sun, J. Xu, S. N. Yagn, Z. P. Zong, T. Chen, C. Fang and J. Sheng, 2012: Analysis and thinking on the extremes of the 21 July 2012 torrential rain in Beijing Part I: Observation and thinking. *Meteorological Monthly*, **38**, 1255–1266. (in Chinese)

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Y. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 hazardous weather testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc*., **93**, 55–74.

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev*., **125**, 2427–2459.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev*., **129**, 2461–2480.

Fan, S. Y., M. Chen, J. Q. Zhong, and Z. F. Zheng, 2009: Performance tests and evaluations of Beijing local high-resolution rapid update cycle system. *Torrential Rain and Disasters*, **28**, 119–125. (in Chinese)

Fels, S. B., and M. D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci*., **32**, 1476–1488.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev*., **129**, 550.

Hamill, T. M., R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev*., **136**, 2620–2632.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev*., **124**, 1225–1242.

Iyer, E. R., A. J. Clark, M. Xue, and F. Y. Kong, 2016: A comparison of 36-60-h precipitation forecasts from convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **31**, 647–661.

Jiang, X. M., H. L. Yuan, M. Xue, X. Chen, and X. G. Tan, 2014: Analysis of a torrential rainfall event over Beijing on 21–22 July 2012 based on high resolution model analyses and forecasts. *Acta Meteorologica Sinica*, **72**, 207–219.

Kong, F. Y., K. K. Droegemeier, and N. L. Hickmon, 2007: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. *Mon. Wea. Rev*., **135**, 759–782.

Kong, F. Y., and Coauthors, 2008: Real-time storm-scale ensemble forecast 2008 spring experiment. *Proc. 24th Conf. Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., Paper 12.13.

Kong, F. Y., and Coauthors, 2014: An overview of CAPS storm-scale ensemble forecast for the 2014 NOAA HWT spring forecasting experiment. *Proc. 27th Conf. Severe Local Storms*, Madison WI, Amer. Meteor. Soc., Paper 43.

Li, J., J. Du, and Y. Liu, 2015: A comparison of initial condition-, multi-physics- and stochastic physics-based ensembles in predicting Beijing "7.21" excessive storm rain event. *Acta Meteorologica Sinica*, **73**, 50–71. (in Chinese)

Liu, Y. Z., X. S. Shen, and X. L. Li, 2013: Research on the singular vector perturbation of the GRAPES global model based on the total energy norm. *Acta Meteorologica Sinica*, **71**, 517–526. (in Chinese)

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc*., **122**, 73–119.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor*., **12**, 595–600.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble Prediction. *ECMWF Seminar Proceedings on Validation of Models over Europe*, Vol. I., Reading, U.K., ECMWF, 21–66.

Pellerin, G., L. Lefaivre, P. Houtekamer, and C. Girard, 2003: Increasing the horizontal resolution of ensemble forecasts at CMC. *Nonlinear Processes in Geophysics*, **10**, 463–468.

Pleim, J. E., 2006: A simple, efficient solution of flux–profile relationships in the atmospheric surface layer. *J. Appl. Meteor. Climatol.*, **45**, 341–347.

Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395.

Pleim, J. E., and A. J. Xiu, 1995: Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models. *J. Appl. Meteor*., **34**, 16–32.

Ren, Z., J. Chen, and H. Tian, 2011: Research on T213 ensemble prediction system stochastic physics perturbation. *Meteorological Monthly*, **37**, 1049–1059. (in Chinese)

Ritchie, H., and C. Beaudoin, 1994: Approximations and sensitivity experiments with a baroclinic semi-Lagrangian spectral model. *Mon. Wea. Rev*., **122**, 2391–2399.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev*., **137**, 3351–3372.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2. NCAR Technical Note NCAR/TN–468+STR, 88 pp.

Stephenson, D. B., C. A. S. Coelho, and I. T. Jolliffe, 2008: Two extra components in the brier score decomposition. *Wea. Forecasting*, **23**, 752–757.

Tao, W.-K., J. Simpson, and M. McCumber, 1989: An ice-water saturation adjustment. *Mon. Wea. Rev*., **117**, 231–235.

Tao, Z. Y., and Y. G. Zheng, 2013: Forecasting issues of the extreme heavy rain in Beijing on 21 July 2012. *Torrential Rain and Disasters*, **32**, 193–201. (in Chinese)

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorological Applications*, **12**, 257–268.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC:

The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Wen, Y. R., L. Xue, Y. Li, N. Wei, and A. M. Lü, 2015: Interaction between typhoon Vicente (1208) and the Western Pacific subtropical high during the Beijing extreme rainfall of 21 July 2012. *J. Meteor. Res.*, **29**, 293–304.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. C. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482.

Xue, M., F. Y. Kong, K. W. Thomas, J. D. Gao, Y. H. Wang, K. A. Brewster and K. K. Droegemeier, 2013: Prediction of convective storms at convection-resolving 1 km resolution over continental United States with radar data assimilation: An example case of 26 May 2008 and precipitation forecasts from spring 2009. *Advances in Meteorology*, **2013**, Article ID 259052, doi: 10.1155/2013/259052.

Xue, M., and Coauthors, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. *22nd Conf. Weather Analysis and Forecasting and 18th Conf. Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., CDROM 3B. 1.

Xue, M., and Coauthors, 2011: CAPS realtime storm scale ensemble and high resolution forecasts for the NOAA hazardous weather testbed 2010 spring experiment. *24th Conf. Weather Forecasting and 20th Conf. Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., Paper 9A.2.

Yu, H. Z., and Z. Y. Meng, 2016: Key synoptic-scale features influencing the high-impact heavy rainfall in Beijing, China, on 21 July 2012. *Tellus A*, **68**, 31045.

Yu, X. D., 2012: Investigation of Beijing extreme flooding event on 21 July 2012. *Meteorological Monthly*, **38**, 1313–1329. (in Chinese)

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

Zhang, H. B., J. Chen, X. F. Zhi, Y. L. Li, and Y. Sun, 2014: Study on the application of GRAPES regional ensemble prediction system. *Meteorological Monthly*, **40**, 1076–1087. (in Chinese)

Zhong, L. Z., R. Mu, D. L. Zhang, P. Zhao, Z. Q. Zhang, and N. Wang, 2015: An observational analysis of warm-sector rainfall characteristics associated with the 21 July 2012 Beijing extreme rainfall event. *J. Geophys. Res.*, **120**, 3274–3291.

Zhu, K. F., Y. Yang, and M. Xue, 2015: Percentile-based neighborhood precipitation verification and its application to a landfalling tropical storm case with radar data assimilation. *Adv. Atmos. Sci.*, **32**, 1449–1459, doi: 10.1007/s00376-015-5023-9.