Probabilistic Precipitation Forecasting over East Asia Using Bayesian Model Averaging

LUYING JI

Key Laboratory of Meteorological Disasters, Ministry of Education (KLME)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing, China

XIEFEI ZHI

Key Laboratory of Meteorological Disasters, Ministry of Education (KLME)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, and Nanjing Joint Center for Atmospheric Research, Nanjing, China

SHOUPENG ZHU

Key Laboratory of Meteorological Disasters, Ministry of Education (KLME)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing, China, and Max Planck Institute for Meteorology, and Climate Service Center Germany (GERICS), Helmholtz Centre for Materials and Coastal Research, Hamburg, Germany

KLAUS FRAEDRICH

Max Planck Institute for Meteorology, Hamburg, Germany

ABSTRACT

Bayesian model averaging (BMA) was applied to improve the prediction skill of 1–15-day, 24-h accumulated precipitation over East Asia based on the ensemble prediction system (EPS) outputs of ECMWF, NCEP, and UKMO from the TIGGE datasets. Standard BMA deterministic forecasts were accurate for light-precipitation events but with limited ability for moderate- and heavy-precipitation events. The categorized BMA model based on precipitation categories was proposed to improve the BMA capacity for moderate and heavy precipitation in this study. Results showed that the categorized BMA deterministic forecasts were superior to the standard one, especially for moderate and heavy precipitation. The categorized BMA also provided a better calibrated probability of precipitation and a sharper prediction probability density function than the standard one and the raw ensembles. Moreover, BMA forecasts based on multimodel EPSs outperformed those based on a single-model EPS for all lead times. Comparisons between the two BMA models, logistic regression, and raw ensemble forecasts for probabilistic precipitation forecasts illustrated that the categorized BMA method performed best. For 10–15-day extended-range probabilistic forecasts, the initial BMA performances were inferior to the climatology forecasts, while they became much better after preprocessing the initial data with the running mean method. With increasing running steps, the BMA model generally had better performance for light to moderate precipitation but had limited ability for heavy precipitation. In general, the categorized BMA methodology combined with the running mean method improved the prediction skill of 1–15-day, 24-h accumulated precipitation over East Asia.

1. Introduction

Precipitation forecasts have significant impacts on our daily life (L. Zhang et al. 2015; Surcel et al. 2017). However, it is difficult for deterministic forecasts to reproduce the future atmospheric reality due to the chaotic characteristics of the atmospheric dynamics and the

---

ⓐ Denotes content that is immediately available upon publication as open access.

Corresponding author: Xiefei Zhi, zhi@nuist.edu.cn

uncertainties associated with models and initial conditions (Lorenz 1963, 1969; Thompson 1957; Smagorinsky 1969). Thus, probabilistic precipitation forecasts are becoming inevitable for weather forecast systems, especially for extended-range (10–30 days) precipitation forecasting. Since 1995 the Meteorological Development Laboratory of the U.S. National Weather Service has produced an extended-range forecasting guidance for probability of precipitation (PoP) using model output statistics (MOS; Carrol and Maloney 2004).

Developing extended-range weather forecasts is necessary to fill the gap between short- to medium-range weather forecasts (1–9 days) and short-term climate predictions (above 30 days), which is also important for disaster prevention and mitigation (Xie et al. 2007; Zeng and Wang 2009; Zhuang et al. 2010). Numerical weather prediction (e.g., ensemble forecasts; Barnett and Preisendorfer 1987; Miyakoda et al. 1983; Yang et al. 2001) and prediction methods using low-frequency atmospheric oscillation signals (Waliser et al. 1999; Plaut and Vautard 1994; Jones et al. 2000; Goswami and Xavier 2003) are considered as efficient and essential methods for extended-range forecasts. In recent decades, multimodel ensemble forecasting methodologies have been developed by many researchers (Chen et al. 2010; Hagedorn et al. 2012; Zhi et al. 2012, 2013; Zhang and Zhi 2015; H. B. Zhang, et al. 2015; Slater et al. 2017) and may play an important role for improving extended-range weather forecasts.

Recently, ensemble forecasting has been one of the key technical concepts for the transition from deterministic to probabilistic forecasts (Räisänen and Ruokolainen 2006; Majumdar and Torn 2014; Scheuerer et al. 2017). The WMO established a 10-yr international research program—The Observing System Research and Predictability Experiment (THORPEX)—in order to further improve the accuracy of 1–15-day weather forecasts (Shapiro and Thorpe 2004). The THORPEX Interactive Grand Global Ensemble (TIGGE) is a key component of the program, which collects ensemble forecasting products from several weather centers such as the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Centers for Environment Prediction (NCEP), the Met Office (UKMO), etc. Different centers use different models, initial perturbation schemes, and ensemble sizes to generate ensemble forecasts (Herrera et al. 2016). In fact, the TIGGE datasets have provided an important basis for the production of probabilistic precipitation forecasts (Park et al. 2008; Bougeault et al. 2010).

Several methods are currently used to derive probabilistic precipitation forecasts such as linear regression (Bermowitz 1975), binning techniques (Yussouf and Stensrud 2006), analog methods (Hamill et al. 2004, 2015), extended logistic regression (Walker and Duncan 1967; Roulin and Vannitsem 2012), and neural networks (Koizumi 1999; Hall et al. 1999). However, some of them are not able to take advantage of the complete information available in the ensemble and only give probabilities for specific events instead of the full predictive probability density function (PDF; Sloughter et al. 2007), and some of them are unable to provide quantitative forecast uncertainties. The main task of ensemble forecasts is to quantitatively describe forecast uncertainty and provide a reliable PDF rather than a better deterministic forecast (Fraedrich et al. 2003; Langmack et al. 2012).

Bayesian model averaging (BMA; Raftery et al. 2005; Sloughter et al. 2007, 2010; Fraley et al. 2010; Liu and Xie 2014) and ensemble MOS (EMOS; Gneiting et al. 2005; Scheuerer 2014; Scheuerer and Hamill 2015) are two state-of-the-art approaches developed for ensemble-based probabilistic precipitation forecasts. EMOS uses a single parametric PDF and effectively calibrates precipitation forecasts (Hemri et al. 2014; Baran and Nemoda 2016; Vogel et al. 2018). BMA can provide weighted average prediction PDFs based on the relative prediction skills of the individual ensemble members (Sloughter et al. 2007; Fraley et al. 2010) and then produce corresponding probabilistic forecasts. Raftery et al. (2005) originally applied BMA to the prediction of temperature and sea level pressure, whose PDFs were approximately normal, yielding well-calibrated and sharp PDFs. However, precipitation is a discontinuous variable, whose PDF does not follow the normal distribution and has a nonnegligible probability of being equal to zero. The BMA method was further developed by Sloughter et al. (2007) and applied to skewed weather variables [e.g., precipitation and wind; Sloughter et al. (2010, 2013)]. Fraley et al. (2010) successfully extended the BMA approach to postprocess multimodel ensembles of any composition. Subsequently, BMA was employed in more studies on daily weather forecasts and climate predictions (Min and Hense 2007; Casanova and Ahrens 2009; Schmeits and Kok 2010; Wang et al. 2012; Erickson et al. 2012; Kim and Suh 2013; Ji et al. 2017). Many analyses demonstrated that the BMA method performed superior to unprocessed forecasts, but with limited capacity for heavy precipitation (Sloughter et al. 2007; Liu and Xie 2014).

So far, relatively few studies applied BMA to extended-range weather forecasts, particularly to forecasts of moderate and heavy precipitation. In our study, we analyze the performance of the BMA method for probabilistic precipitation forecasts, especially for the extended-range forecasts and for heavy precipitation. We apply the

TABLE 1. Ensemble forecast systems used in this study.

Prediction center	Model spectral resolution	Initial perturbation scheme	Ensemble members
ECMWF	T399L62/T255L62	Singular vectors and ensemble of data assimilations (EDA)	51
NCEP	T126L28	Ensemble transform and rescaling	21
UKMO		Ensemble transform Kalman filter (ETKF)	24

BMA method to three single-model ensemble prediction systems (EPSs) (ECMWF, NCEP, and UKMO) and the multimodel ensemble system comprising all three EPSs, and venture to improve the BMA capacity for moderate- and heavy-precipitation forecasts and for forecasts at longer lead times.

The paper is organized as follows. The datasets and methods are described in section 2. In section 3, we provide and compare the deterministic and probabilistic precipitation forecasting results for lead times of 1–15 days using the standard and modified BMA methods. Finally, a summary and discussion is given in section 4.

2. Data and methods

a. Data

We use the 24-h accumulated precipitation forecasts from the ECMWF (51 members), NCEP (21 members), and UKMO (24 members) EPSs with a resolution of $1° × 1°$ (Table 1) initialized at 1200 UTC for lead times of 1–15 days. The multimodel ensemble [exchangeable grand EPS (EGE)] includes all three EPSs and contains 96 members. The datasets are obtained from the TIGGE–ECMWF portal.

The hourly precipitation products merged from the precipitation analyses of the U.S. National Oceanic and Atmospheric Administration Climate Prediction Center morphing technique (CMORPH) and observed values from automatic weather stations (AWS) are used for the evaluation. This merged gauge–satellite precipitation product with a resolution of $0.1° × 0.1°$ was produced by combining probability density functions of both products and optimal interpolation, which integrate the advantages of gauge observed data and retrieved satellite precipitation data (Janowiak and Xie 1999; Xie and Xiong 2011).

We analyzed the data for the time period from 1 May to 31 August 2013 over an area located in East Asia covering the region 15.05°–58.95°N, 70.15°–139.95°E. For both the observed and predicted, 24-h accumulated precipitation below 0.01 mm day^{-1} is defined as "no precipitation."

b. Bayesian model averaging

We followed Sloughter et al. (2007), who extended the BMA method (which we explain in detail in the next

paragraphs) to 24-h accumulated precipitation, by using the cube root of 24-h accumulated precipitation y as the predictor variable. In the case of multimodel ensembles usually two or more of the member forecasts from the same EPS, which lack individually distinguishable physical features, are exchangeable (Fraley et al. 2010).

Assuming that there are S EPSs and each EPS has m_s exchangeable members. The total number of members is given by $\sum_{s=1}^{S} m_s = m$. The variable $f_{s,t}$ represents the tth member of the sth EPS. Then the BMA prediction PDF for exchangeable ensemble forecasts can be written as

$$p\left(y|\{f_{s,t}\}_{s=1,...,S,t=1,...,m_s}\right) = \sum_{s=1}^{S}\sum_{t=1}^{m_s} \omega_{s,t}[P(y=0|f_{s,t})$$
$$I(y=0) + P(y>0|f_{s,t})g_{s,t}$$
$$(y|f_{s,t})I(y>0)], \quad (1)$$

where $\omega_{s,t}$ is the posterior probability of forecast $f_{s,t}$. The indicator function $I[]$ is equal to 1 if the condition in brackets is satisfied; otherwise, it is zero.

The first part $P(y=0|f_{s,t})$ is the probability of no precipitation as a function of $f_{s,t}$ by a logistic regression model:

$$\log it P(y=0|f_{s,t}) = \log\frac{P(y=0|f_{s,t})}{P(y>0|f_{s,t})}$$
$$= a_{0s,t} + a_{1s,t}f_{s,t}^{1/3} + a_{2s,t}\delta_{s,t}, \quad (2)$$

where $\delta_{s,t}$ is an indicator and is equal to 1 if $f_{s,t}=0$; otherwise, it is equal to 0. Here, $P(y>0|f_{s,t})$ is the probability of nonzero precipitation given the forecast $f_{s,t}$. The parameters a_{0k}, a_{1k}, and a_{2k} are estimated by using the Newton–Rapson iterative method against the training period.

The second part is the gamma distribution PDF of the precipitation amount $g_{s,t}(y|f_{s,t})$:

$$g_{s,t}(y|f_{s,t}) = \frac{1}{\beta_{s,t}^{\alpha_{s,t}}\Gamma(\alpha_{s,t})}y^{\alpha_{s,t}-1}\exp(-y/\beta_{s,t}), \quad (3)$$

where the shape parameter $\alpha_{s,t}$ and scale parameter $\beta_{s,t}$ of the gamma distribution depends on $f_{s,t}$ through the following relationships:

$$\mu_{s,t} = b_{0s,t} + b_{1s,t}f_{s,t}^{1/3} = \alpha_{s,t}\beta_{s,t}, \quad \text{and} \quad (4)$$

$$\sigma_{s,t}^2 = c_{0s,t} + c_{1s,t}f_{s,t} = \alpha_{s,t}\beta_{s,t}^2, \quad (5)$$

where $\mu_{s,t}$ and $\sigma_{s,t}^2$ are the mean and variance of the gamma distribution, respectively. The parameters $b_{0s,t}$ and $b_{1s,t}$ are estimated by linear regression from the training period. The parameters $\omega_{s,t}$, $c_{0s,t}$, and $c_{1s,t}$ are estimated by the expectation maximization (EM) algorithm based on the training data. The details of parameter estimations associated with the EM algorithm can be found in Fraley et al. (2010) and are not repeated here. Last but not least, members that are treated as exchangeable will have equal BMA weight and equal BMA parameter values.

As mentioned in the previous studies (Sloughter et al. 2007; Liu and Xie 2014), the BMA method, which is called the standard BMA method in the following, has a limited ability to successfully forecast heavy precipitation, which could be associated with uncertainties in the parameter estimation. The fewer samples of the heavy-precipitation events than the light to moderate ones may affect the bias correction for heavy precipitation. To overcome this problem, considering that the intensity of precipitation may affect the parameter estimation, we construct different BMA models based on different precipitation categories. In our study, the 24-h accumulated precipitation is separated into three categories based on the ensemble mean [i.e., light precipitation (<10 mm), moderate precipitation (~10–24.9 mm), and heavy precipitation (≥25 mm)]. We call this extension in the following the categorized BMA model. First, we select samples of various categories of precipitation according to the daily accumulative precipitation amount during a training period. Then the BMA models for different precipitation amounts are established, respectively. Consequently, the most appropriate BMA model can be determined for the forecast period based on the ensemble mean. A spatial sliding window of $1° \times 1°$ is used to increase the number of moderate- and heavy-precipitation samples in order to decrease the sampling variability of the BMA results. That is, within the $1° \times 1°$ spatial window centered at the target point, grid points with the same precipitation magnitude are also additionally taken as samples for the respective precipitation category.

Parameter estimation for the BMA is obtained from a training period. Thus, it is important to choose the optimal length of the BMA training period. Here we adopt a sliding temporal window (following Raftery et al. 2005), using a training sample period of the N previous days, where $N = 10, 15, 20, \ldots, 50$. Then we construct different BMA prediction models for a single-model EPS (i.e., BMA model based on ECMWF, NCEP, and UKMO EPS, hereafter abbreviated as E-BMA, N-BMA and U-BMA, respectively) and multimodel EGE (i.e., the standard BMA and the categorized BMA model, hereafter abbreviated as s-BMA and c-BMA, respectively). Evaluations for the performances of different BMA models using different training lengths are also carried out. The certain time when the verification metrics tend to be stable is taken as the optimal sliding training period. As a result, a 30-day sliding training period is selected for all BMA models (not shown).

c. Running mean method

Previous studies (Lorenz 1982; Chou 1989) pointed out the difficulty in producing daily forecasts for extended-range periods due to the limited atmospheric predictability. However, for extended-range weather forecasting precipitation data averaged over several days are very valuable, as they contain relevant information for future precipitation. Therefore, we apply also the multiday running mean of precipitation for the 10–15-day extended-range weather forecasts with

$$y_j = \frac{1}{k}\sum_{i=1}^{k}x_{i+j-1}(j = 1, 2, \ldots, n - k + 1), \quad (6)$$

where x is the observed daily precipitation or the daily precipitation of the model forecasting, and k is the running step. The index of x indicates the serial number of time series. We use $k = 3, 5$, and 7 to preprocess the data for every lead time of the extended-range period and then reconduct the BMA model.

d. Verification methods

The anomaly correlation coefficient (ACC), the mean absolute error (MAE), and the equitable threat score (ETS) are used to evaluate the BMA deterministic forecasts (i.e., the median of the BMA prediction PDF). For the probabilistic forecasts, the Brier score (BS) is adopted to assess forecasts exceeding specific thresholds and the continuous ranked probability score (CRPS) is applied to measure the BMA forecast distributions. Additionally, skill scores of some of these metrics are also employed to verify the improvements in comparison with the reference forecasts.

1) ANOMALY CORRELATION COEFFICIENT AND MEAN ABSOLUTE ERROR

The anomaly correlation coefficient and the mean absolute error are defined by

TABLE 2. Precipitation test classification.

Observation	Forecast	
	Yes	No
Yes	N_A	N_C
No	N_B	N_D

TABLE 3. Results from logistic regression PoP, s-BMA, c-BMA, and the corresponding observations for two example grids ($mm^{1/3}$).

Ensemble member	ECMWF	NCEP	UKMO
	s-BMA model		
BMA weight	0.55	0.22	0.23
Member PoP	0.86	0.81	0.76
BMA PoP	0.83		
BMA forecast	2.35		
	c-BMA model		
BMA weight	0.53	0.22	0.25
Member PoP	0.88	0.88	0.88
BMA PoP	0.88		
BMA forecast	2.82		
Observation	2.98		

$$\text{ACC} = \frac{\frac{1}{N}\sum_{i=1}^{N}(f_i - \overline{f})(o_i - \overline{o})}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - \overline{f})^2 \frac{1}{N}\sum_{i=1}^{N}(o_i - \overline{o})^2}}, \quad \text{and} \quad (7)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|f_i - o_i|, \quad (8)$$

where N is the total number of grid samples in the spatial field; f_i means the median forecast value of the ith sample; o_i represents the corresponding observation; and $\overline{f}$ and $\overline{o}$ denote the averages of the f_i ($i = 1, 2, \ldots, N$) and o_i, respectively. ACC and MAE can be used to assess the correlations and differences between model forecasts and observed values, respectively. Higher ACC and lower MAE always demonstrate better model forecasts.

2) EQUITABLE THREAT SCORE

The ETS represents the forecasting skill of precipitation in a certain area relative to stochastic forecasts and is specified as follows:

$$\text{ETS} = \frac{N_A - R(a)}{N_A + N_B + N_C - R(a)}, \quad \text{and} \quad (9)$$

$$R(a) = \frac{(N_A + N_B)(N_A + N_C)}{N_A + N_B + N_C + N_D}. \quad (10)$$

Variables in the formula are defined in Table 2. ETS, by definition, has penalties for both inaccuracy and misses and ranges from $-1/3$ to 1. Larger ETS values represent higher forecasting skills, and the forecast has no skill relative to random forecasts if the ETS is less than 0.

3) BRIER SCORE AND CONTINUOUS RANKED PROBABILITY SCORE

The BS (Brier 1950) is kind of a mean square probabilistic error, which has been widely used in the assessment of probabilistic precipitation forecasts of exceeding specific thresholds (Ferro 2007):

$$\text{BS} = \frac{1}{n}\sum_{i=1}^{n}(f_i - O_i)^2, \quad (11)$$

where f_i is the forecasting probability, and O_i is the actual outcome of the event at instance i. If it is true,

$O_i = 1$; otherwise, $O_i = 0$, and n is the number of forecasting instances. The Brier score is negatively oriented and has the range $0 \le \text{BS} \le 1$.

For a predictive cumulative distribution function $f(y)$ and an observation o, the CRPS is defined as

$$\text{CRPS} = \frac{1}{n}\sum_{i=1}^{n}\int\left[f(y_i) - H(y_i - o_i)\right]^2 dx, \quad (12)$$

where $H(y_i - o_i)$ is the Heaviside function that jumps from 0 to 1 at the observed value. If $y_i < o_i$, $H(y_i - o_i) = 0$; otherwise, $H(y_i - o_i) = 1$. CRPS can be considered as an integral over BS for all thresholds and measures the overall probabilistic performance.

4) SKILL SCORE

To evaluate the prediction skill of the BMA model compared to the reference forecasts, the skill score is proposed:

$$\text{SS} = \frac{\text{Score(forecast)} - \text{Score(reference)}}{\text{Score(perfect)} - \text{Score(reference)}}, \quad (13)$$

where Score(perfect) and Score(reference) are the skills achieved by the perfect forecasts and reference forecasts, respectively. In our study, the raw ensemble and climatology forecast will be used as references and Score(forecast) represents the scores obtained from the BMA forecasts. This metric is positively oriented. The skill scores of ETS and BS will be calculated to detect the BMA improvements in comparison with the raw ensembles and climatology forecasts in the following sections.

3. BMA probabilistic precipitation forecasting

a. BMA PDF

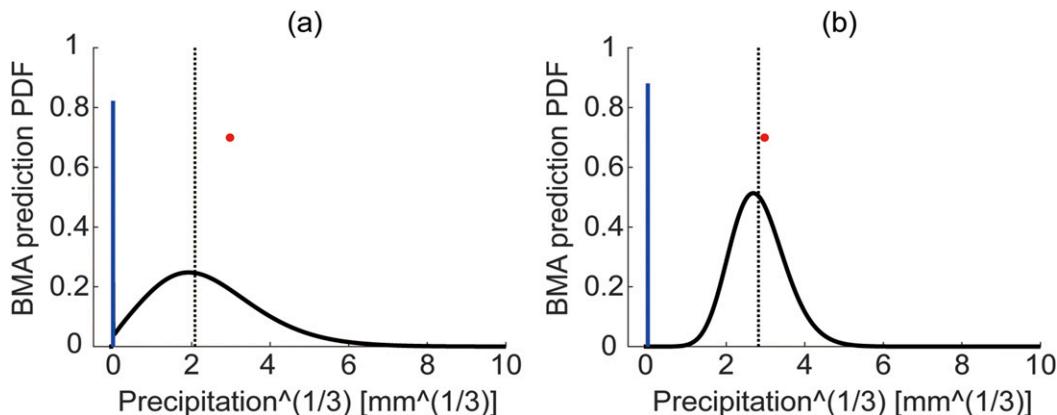To illustrate how the BMA method works for probabilistic precipitation forecasts, the 1-day lead time

FIG. 1. The prediction PDF of 24-h accumulated precipitation for 19°N, 90°E on 29 Jun 2013 obtained from the (a) standard BMA model and (b) categorized BMA model with a lead time of 1 day. The blue vertical line at zero represents the BMA estimation of the probability of precipitation, and the curve is the BMA PDF of the precipitation amount given that is nonzero. The dashed vertical line represents the BMA deterministic forecast (median forecast), and the red dot represents the verifying observation.

forecast of the heavy-precipitation event on 29 June 2013 at grid 19°N, 90°E is taken as an instance. Table 3 shows the results from the raw ensembles, the logistic regression, s-BMA, c-BMA, and the corresponding observations for this sample. The prediction PDFs obtained from the two BMA methods are shown in Fig. 1. The probability of precipitation exceeding a certain threshold is the proportion of the area under the BMA PDF to the right of the given threshold, multiplied by the probability of the nonzero precipitation. The deterministic forecast
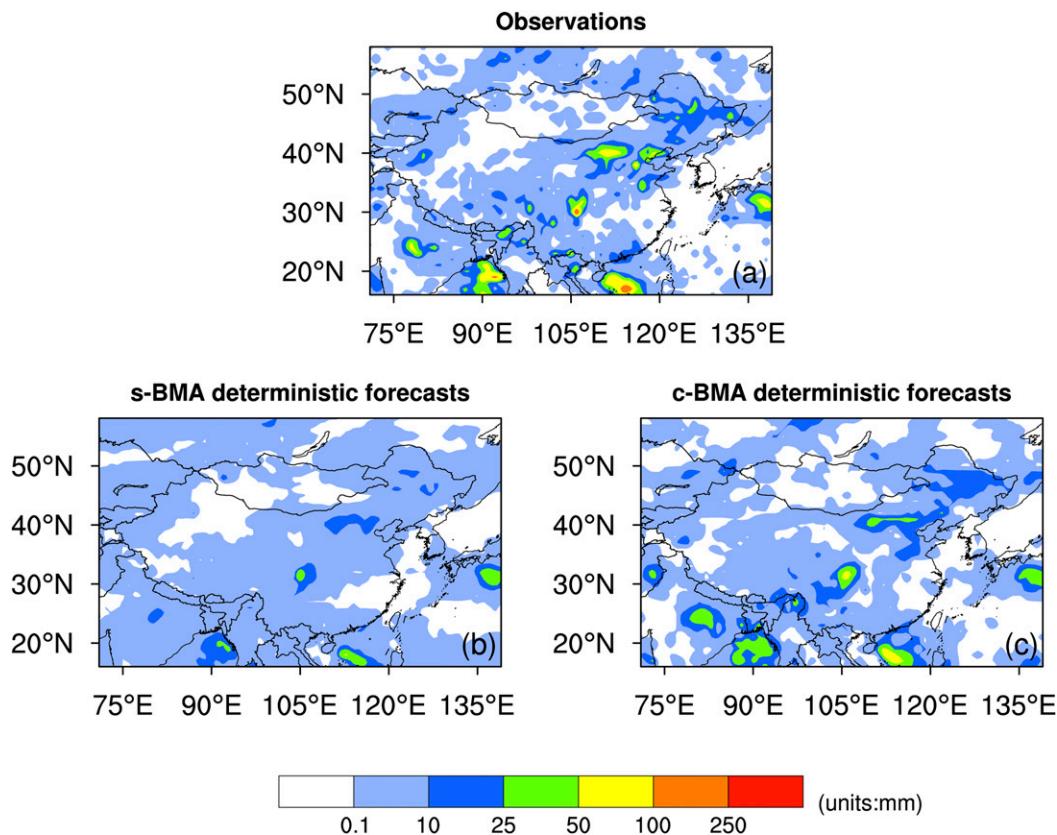
FIG. 2. Spatial distributions of 24-h accumulated precipitation (mm) from (a) observations, (b) s-BMA deterministic forecasts, and (c) c-BMA deterministic forecasts for 29 Jun 2013.
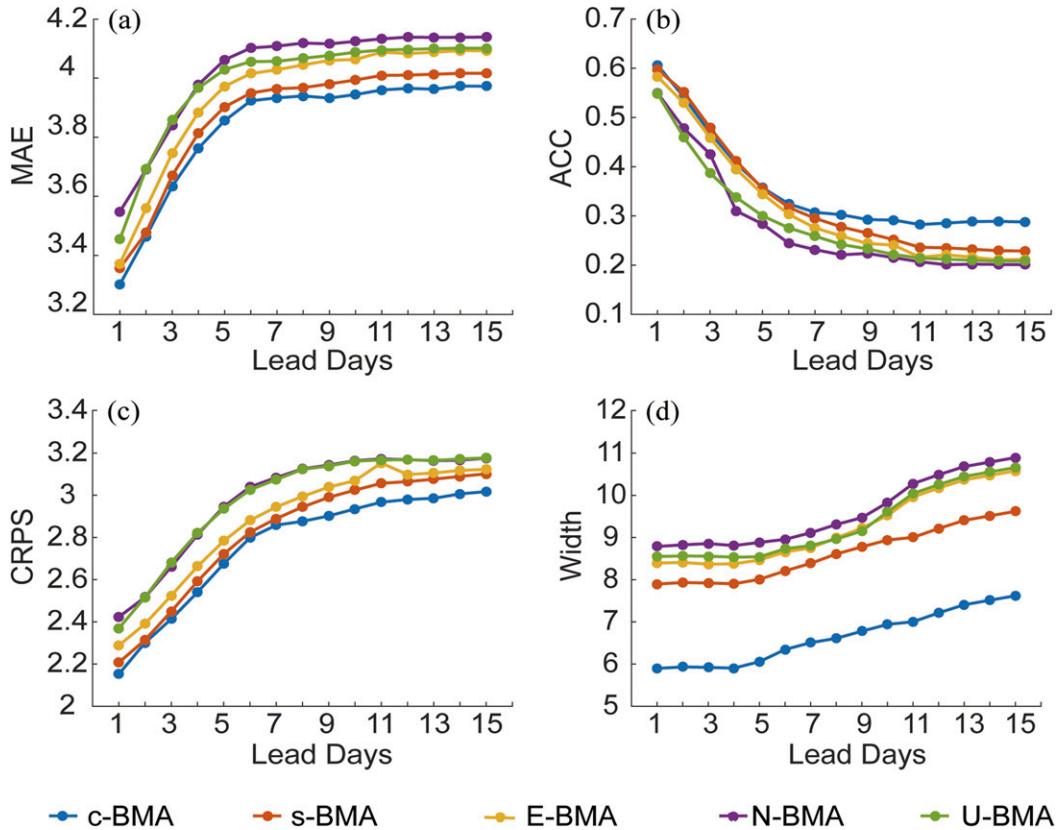
FIG. 3. Mean verification metrics for different BMA models of 24-h accumulated precipitation with lead times of 1–15 days on different EPSs. (a) MAEs of BMA deterministic forecasts, (b) ACCs of BMA deterministic forecasts, (c) CRPSs, and (d) average widths of lower 90% prediction intervals.

(i.e., the median of the BMA prediction PDF) obtained from the s-BMA model is 2.35 mm$^{1/3}$, while it is 2.82 mm$^{1/3}$ for the c-BMA model, which is much closer to the observation of 2.98 mm$^{1/3}$ (Table 3). The PDF from the c-BMA model (Fig. 1b) also indicates that the deterministic forecast (vertical dash line) of c-BMA model is relatively close to the observation (red dot).

The spatial distributions of 24-h accumulated precipitation from the observations and the two BMA models on 29 June 2013 are shown in Fig. 2. The forecasts of the s-BMA model have more false alarms and misses than the c-BMA model forecasts. Thus, the c-BMA model can generate a more accurate deterministic forecast not only for moderate and heavy precipitation but also for light precipitation.

b. Comparisons between different BMA models and raw ensembles

The assessments of different BMA models are described in the following section based on an individual EPS and a multimodel EGE for 24-h accumulated precipitation with lead times of 1–15 days. The BMA model, logistic regression, and raw ensemble forecasts are evaluated. In addition, the calibration of the BMA and raw ensemble is presented.

The verification metrics (i.e., MAE, ACC, CRPS, and average width of the lower 90% prediction intervals) are given in Fig. 3 for different BMA models. The average width of the lower 90% prediction intervals assesses the sharpness of the BMA prediction PDF. The c-BMA model performs better than the s-BMA especially for longer lead times. The EGE BMA (i.e., s-BMA and c-BMA) models outperform the single EPS BMA models (i.e., E-BMA, N-BMA, and U-BMA), with smaller MAEs, CRPSs, average width of the lower 90% prediction intervals, and higher ACCs. In addition, the EPS prediction skill decreases with increasing lead times.

The comparison between the c-BMA model and the raw ensemble is presented in Fig. 4. The BMA model performs much better than the raw ensemble for all lead times, in terms of not only deterministic but also probabilistic forecasts. Noticeably, there is a "mutation" at
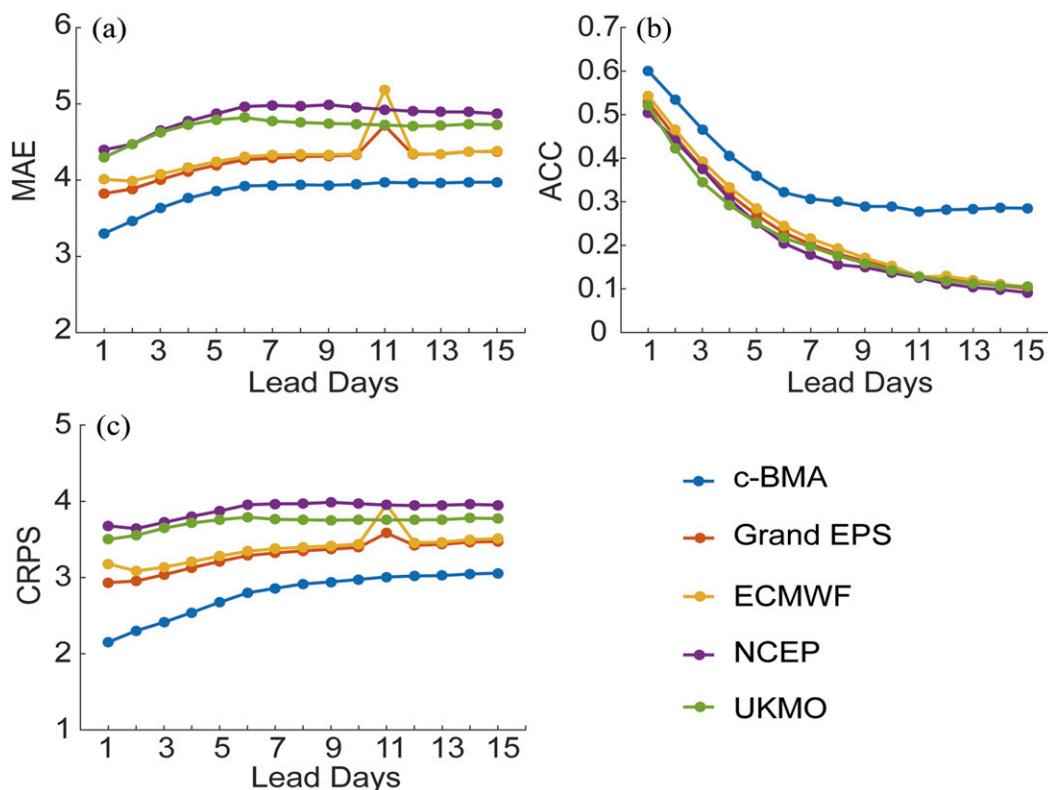
FIG. 4. Mean verification metrics for the categorized BMA model and raw ensembles with lead times of 1–15 days. (a) MAEs of BMA deterministic forecasts, (b) ACCs of BMA deterministic forecasts, and (c) CRPSs.

day 11 of E-BMA (Fig. 3), raw ECMWF EPS, and raw grand EPS (Fig. 4), which is due to the coarser resolution of the ECMWF EPS and the atmosphere–ocean model coupling from day 10 onward (Andersson 2015). To avoid a too significant mutation, there exists a 24-h overlap period between days 9 and 10, especially for large-scale precipitation. The raw grand EPS contains the ensemble members from the ECMWF EPS without bias correction. As a result, the BMA model of ECMWF EPS, raw ECMWF EPS, and raw grand EPS all have relatively poor performances at the lead time of 11 days. But for the EGE BMA (i.e., s-BMA and c-BMA) models, the abrupt phenomenon is not as obvious as in the BMA model of ECMWF EPS, which can be attributed to the bias correction and the different weights of different ensemble members from different EPSs.

Taking the raw ensemble as the reference, the ETS skill scores for four-category precipitation forecasts are presented in Fig. 5. The performances of the s-BMA deterministic forecasts for moderate and heavy rainfall are even inferior to the raw ensembles for all lead times. The skill scores for moderate and heavy precipitation are improved after the categorization, but the improvements for moderate and heavy precipitation decrease with increasing lead time. The c-BMA model also improves the capacity of deterministic forecast for light precipitation at longer lead times.

Further assessments are carried out for the nine-category precipitation forecasts from the c-BMA predicted PDF, s-BMA predicted PDF, logistic regression, ensemble consensus voting, and climatology forecasts, as shown in Table 4. The EGE BMA models, including the s-BMA and c-BMA models, generally outperform the logistic regression, raw ensemble, and climatology forecasts. However the performance of the s-BMA model at the threshold of 0 mm is not as good as the logistic regression. Furthermore, the c-BMA model performs better than the s-BMA, especially at higher thresholds, indicating that the categorization significantly improves the BMA capacity for moderate and heavy precipitation. Besides, climatology forecasts are inferior to the logistic regression forecasts, but superior to the raw ensemble forecasts. Similar results are obtained for other lead times (not shown), but the advantages of the BMA models to the climatology forecasts do not always persist for longer lead times.
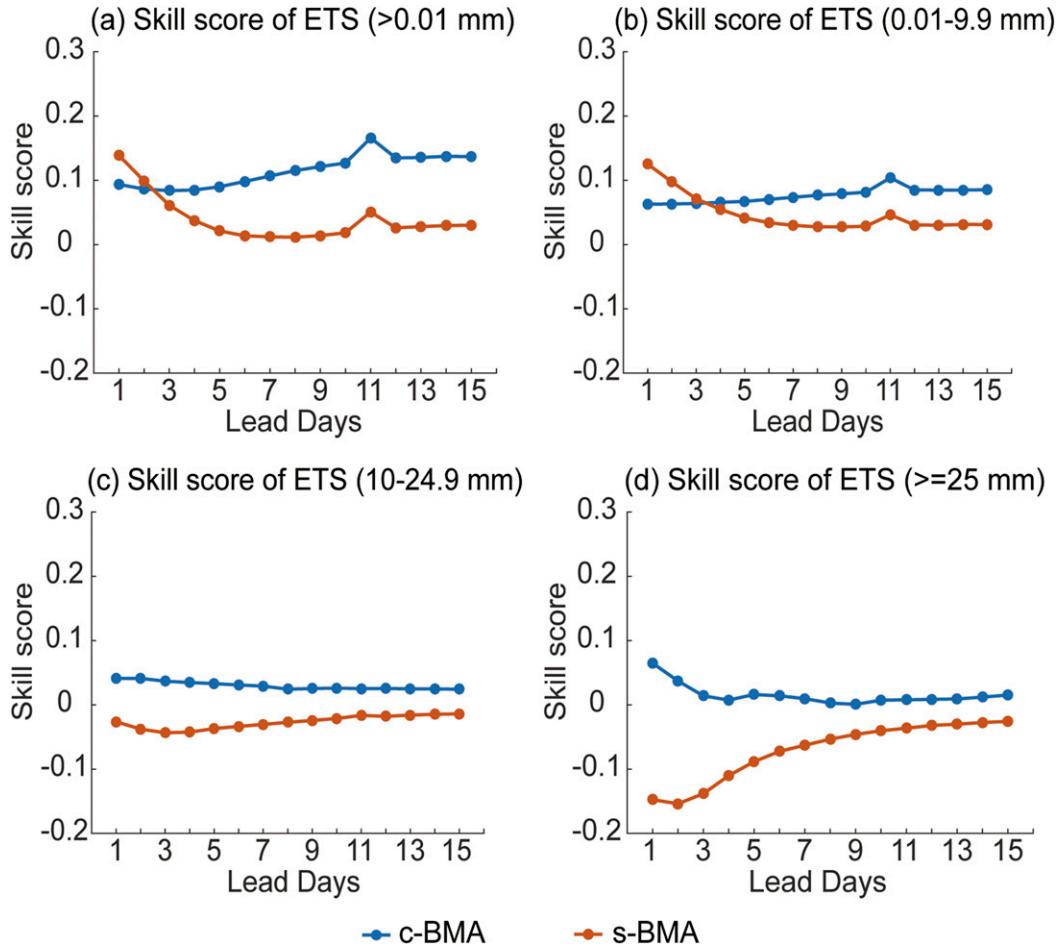
FIG. 5. Skill scores of ETS for four-category precipitation forecasts obtained from the categorized BMA model and standard BMA model with respect to the raw ensemble with lead times of 1–15 days: (a) >0.01, (b) ~0.01–9.9, (c) ~10–24.9, and (d) ≥25 mm.

Probabilistic forecasting is always applied with the advantage of maximizing the sharpness of the prediction subject to calibration (Raftery et al. 2005; Gneiting et al. 2007). We use the verification rank histogram (Talagrand et al. 1997) and the probability integral transform (PIT) histogram (Raftery et al. 2005; Gneiting

et al. 2007) to evaluate the calibration of an original EPS and the BMA forecast distributions, respectively. In both cases, a more uniform histogram characterizes the better calibration. The verification rank histogram for the raw ECMWF ensemble forecasts is displayed in Fig. 6 for the 1-day lead time, as well as the PIT

TABLE 4. Mean BS values for probabilistic precipitation forecasts exceeding some specific thresholds using different ensemble methods with lead time of 1 day.

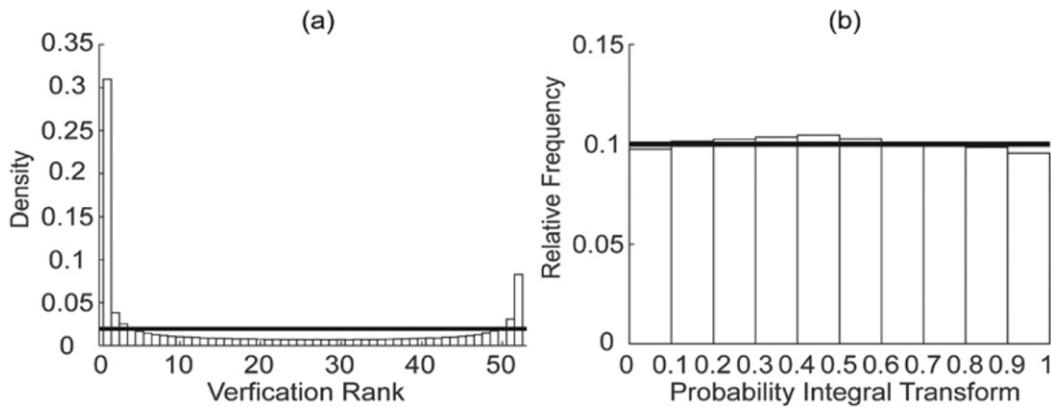| Threshold (mm)	c-BMA	s-BMA	Logistic	Ensemble	Climatology
0	0.2523	0.2797	0.2695	0.3196	0.2809
5	0.1037	0.1130	0.1174	0.2118	0.1227
10	0.0590	0.0789	0.0883	0.1429	0.0857
15	0.0360	0.0572	0.0654	0.0986	0.0624
25	0.0158	0.0330	0.0435	0.0517	0.0357
30	0.0112	0.0260	0.0373	0.0392	0.0278
50	0.0035	0.0108	0.0237	0.0149	0.0116
100	0.0006	0.0019	0.0158	0.0022	0.0021

FIG. 6. (a) Verification rank histogram for ECMWF EPS raw ensemble forecasts and (b) PIT histogram for the categorized BMA model forecast distributions of precipitation accumulation with lead time of 1 day.

histogram of the c-BMA forecasts. The raw ensemble exhibits an L shape, indicating that the original EPS has a positive deviation. By contrast, the c-BMA forecast distributions are well calibrated. The calibration of PoP forecasts in the form of the reliability diagram is shown in Fig. 7. The reliability diagram also indicates that c-BMA performs better in predicting the probability of nonzero precipitation than the s-BMA model and the raw ensemble. Similar results hold for the other EPSs and other lead times (not shown).

c. Extended-range probabilistic precipitation forecasting

In this section, we focus on precipitation forecasts for extended-range lead times of 10–15 days. As mentioned in section 2c, more attention will be paid to the averaged accumulated precipitation over several days during the extend-range period instead of the daily forecast. To this goal the running mean method is applied to preprocess the data for every lead time of the extended-range period. The BMA model is reconducted for extended-range precipitation forecasts with preprocessed data resulting from different running steps (e.g., 3 days, 5 days, and 7 days). Then the BMA performances for different running steps are evaluated via the MAE, ACC, CRPS, and Brier skill score (BSS) to find the optimal running step.

Figure 8 displays the performance of the BMA models for extended-range probabilistic precipitation forecasts based on different preprocessed data. The BMA prediction using running mean data shows lower MAE and CRPS as well as higher ACC than the initial BMA prediction, indicating the enhanced forecast capability induced by the running mean preprocessing. Additionally, it is indicated by the BS for the six-category precipitation forecasts obtained from the four models

(Fig. 9), that the BMA model based on running mean data is consistently more skillful than the BMA model on original data for extended-range probabilistic precipitation forecasts. Moreover, the BMA model performs better as the running step increases while the forecast skills decrease with increasing lead times. The improvements of forecast skills compared to the reference forecast (i.e., climatology forecast hereby) are shown in Fig. 10. Considering the differences between the climatological event frequencies for different regions within East Asia, the calculations of BSS values are referred by Hamill and Juras (2006). The sample climatology probability is determined by each month and each grid point, and then the reference Brier scores of the climatology are calculated separately for subsets with different climatologies.
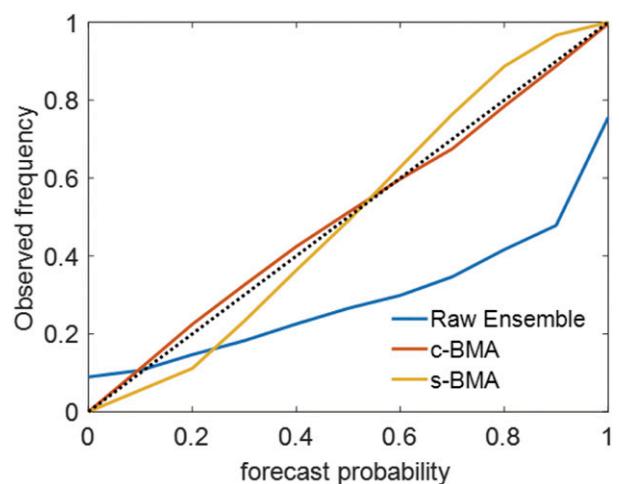

FIG. 7. Reliability diagram of binned PoP forecast vs observed relative frequency of precipitation, for consensus voting of the raw ensemble, the standard BMA model, and the categorized BMA model with a lead time of 1 day. The dotted line refers to the perfect forecast.
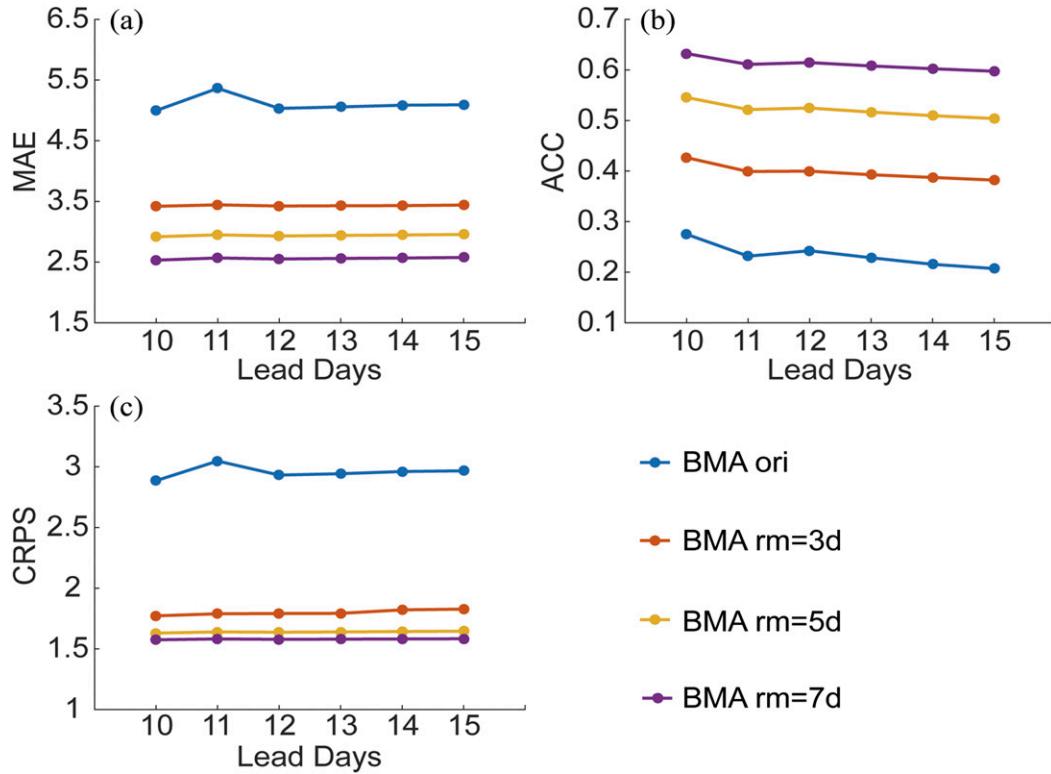
FIG. 8. Mean verification metrics of BMA models based on initial data and 3-, 5-, and 7-day running mean data for lead times of 10–15 days. (a) MAEs of BMA deterministic forecasts, (b) ACCs of BMA deterministic forecasts, and (c) CRPSs.

The results show that the BMA model based on running mean data performs better than that based on the original data for light to moderate precipitation. However, for heavy precipitation, the running mean method does not work as expected and is even inferior for extended-range forecast in terms of the BSS values, which is mainly due to the fact that the sample size of heavy precipitation decreases rapidly after using the running mean method, especially for longer running steps. In addition, because of the often limited duration of a weather process, the running step should not be too long.

4. Summary and discussion

In this study, the BMA method was applied to improve the ensemble forecasts of 24-h accumulated precipitation with lead times of 1–15 days based on the TIGGE multimodel ensembles. The BMA prediction models were established for different EPSs, including single-model EPSs (ECMWF, NCEP, and UKMO) and the multimodel EGE.

The standard BMA (s-BMA) deterministic forecasts were relatively accurate for light-precipitation events,

but became inaccurate for moderate- and heavy-precipitation events. We proposed the categorized BMA (c-BMA) model, which separated the 24-h accumulated precipitation into three categories according to the ensemble mean forecast (i.e., light precipitation below 10 mm, moderate precipitation from 10 to 24.9 mm, and heavy precipitation above 25 mm). Samples of these precipitation categories were selected to establish different BMA models and estimate their respective BMA parameters from the training period. Thus, the most appropriate BMA model can be chosen for the forecast period based on the forecast ensemble mean.

The c-BMA forecasts showed larger ACCs between forecasts and observations and smaller MAEs and CRPSs than those of the s-BMA models. Besides, the ETS values of the c-BMA models for all precipitation categories were improved. There were differences between model forecast skills for different categories of precipitation. Thus, the c-BMA method can take all ensemble information into full use to achieve better forecasts. Generally, the c-BMA forecasts were clearly superior to the s-BMA not only for deterministic forecasts but also for probabilistic forecasts. Additionally,
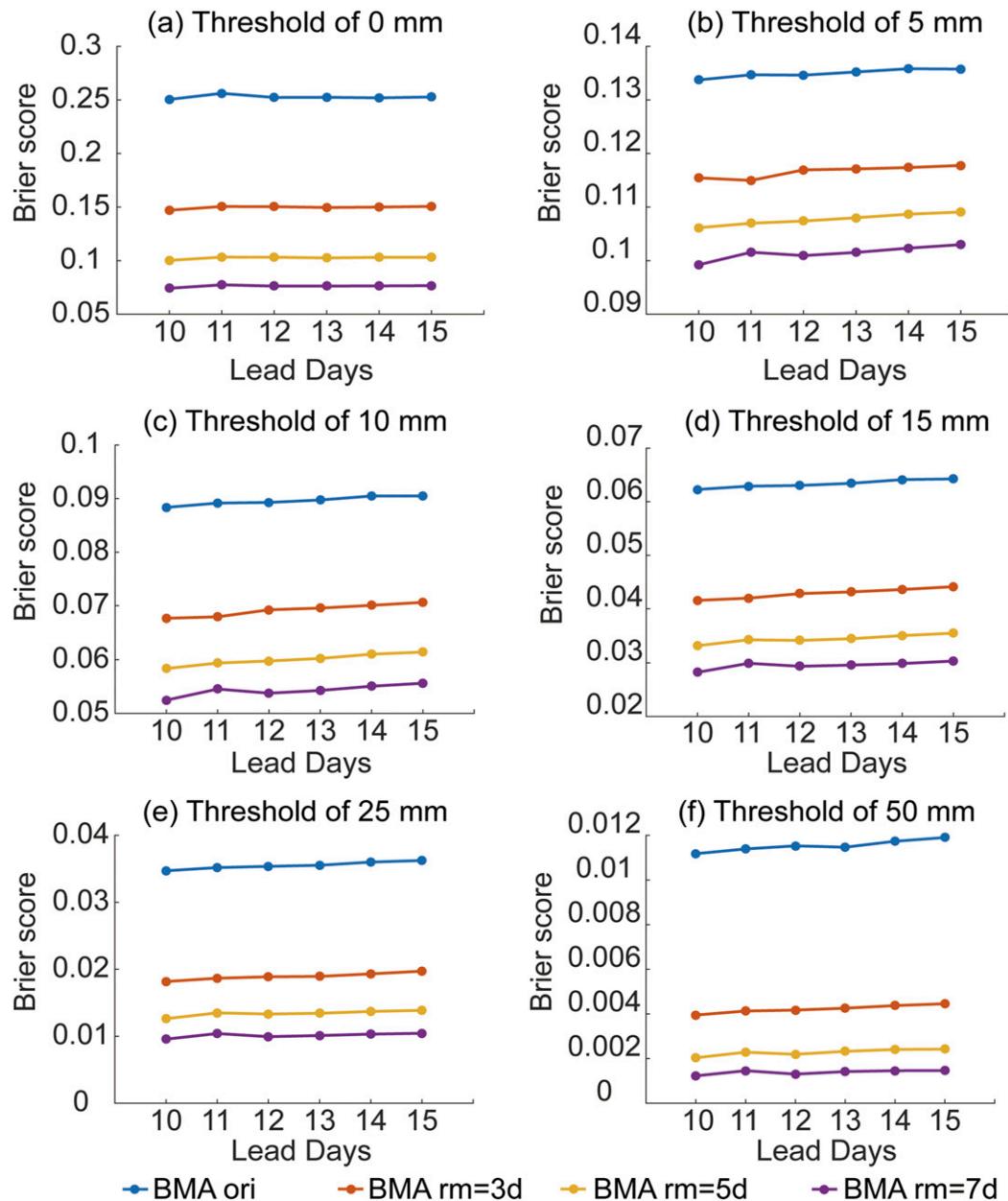
FIG. 9. Mean Brier scores of the six-category precipitation forecasts obtained from BMA models on initial data and 3-, 5-, and 7-day running mean data with lead times of 10–15 days: (a) 0, (b) 5, (c)10, (d) 15, (e) 25, and (f) 50 mm.

the BMA models for the EGE (i.e., s-BMA and c-BMA) outperformed those of any other single-model EPS (i.e., E-BMA, N-BMA, and U-BMA) for all lead times. The probabilistic forecasts of the BMA models performed better than the raw ensemble forecasts and the logistic regression results. As a statistical postprocessing method, BMA yields a full prediction PDF, which comprises two parts: the probability of zero precipitation and the PDF for the accumulated precipitation

above zero. Furthermore, the BMA PDF is better calibrated than the raw ensemble.

As expected, the probabilistic precipitation forecasts became less skillful with increasing lead time. To achieve better performance, the BMA model for extended-range probabilistic precipitation forecasts was reoperated with the running mean data for a precipitation process during the extended-range period. As a result, the obtained BMA probabilistic precipitation
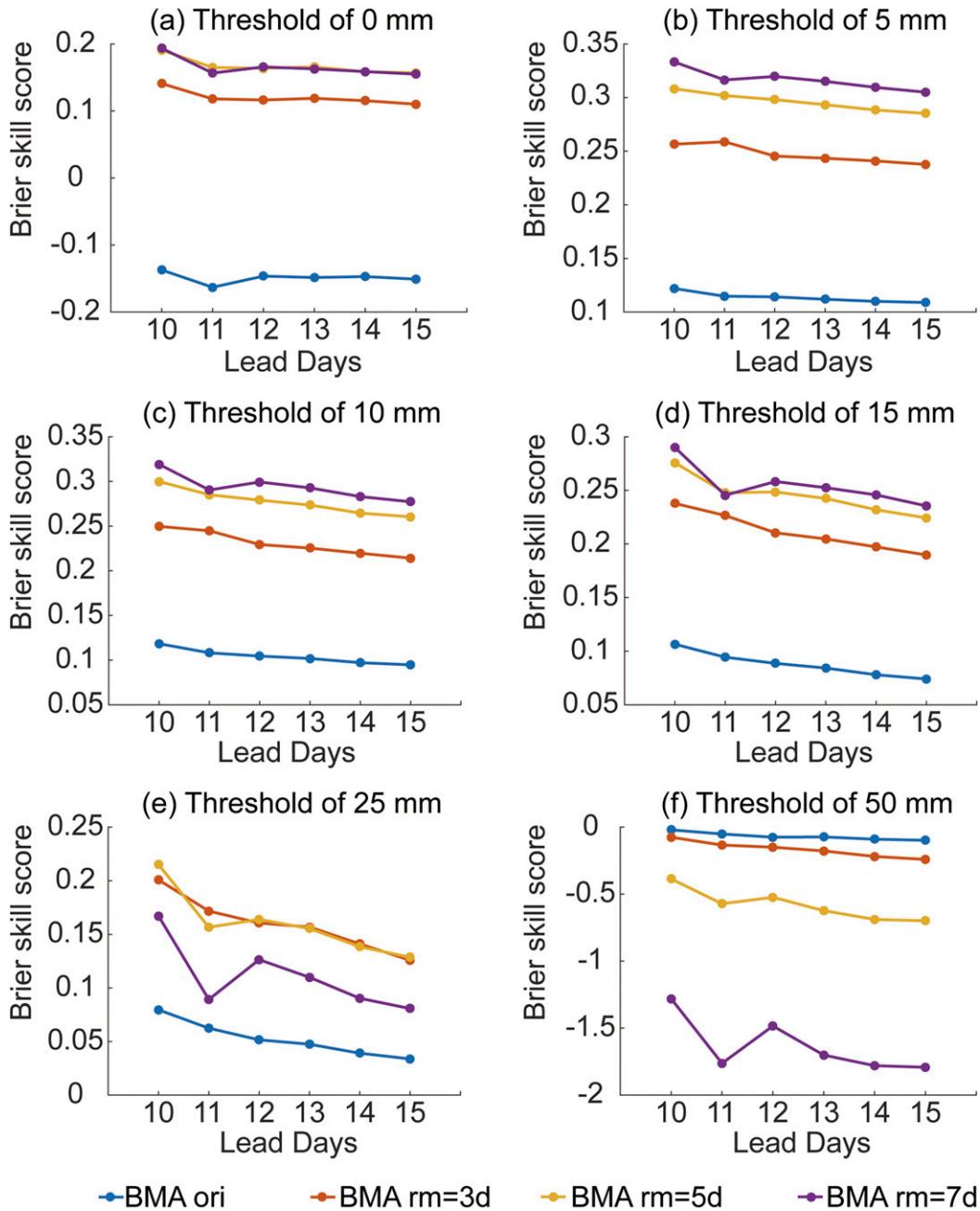
FIG. 10. As in Fig. 9, but for mean Brier skill scores.

forecasts became more skillful then and performed better than the climatology forecasts for light to moderate precipitation, but had limited or lower skill for heavy precipitation.

In conclusion, the categorized BMA probabilistic precipitation forecasts based on multi-EPSs with lead times of 1–15 days greatly improved the quality of probabilistic precipitation forecasts. On the other hand, with the development of probabilistic forecasts and

multimodel ensemble forecasts, both short- to medium-range and extended-range precipitation forecasts will be further advanced.

In our experiment, we used three EPSs, which provided extended-range forecasts and already had relatively high prediction skills. As more centers provide extended-range numerical weather predictions in the future, more EPSs can be used to conduct the experiment, which may further improve the BMA performance. But it should be noted the

EPSs devoted to the BMA model need to be selected further, not following "the more the better" paradigm as, apparently, not all EPSs are sufficiently efficient. In addition, the EMOS method (Scheuerer 2014) and the optimal weight method (Wanders and Wood 2016) can also be employed for extended-range precipitation forecasts and further compared to the BMA method, which will be investigated in the near future.

REFERENCES

Andersson, E., 2015: Availability and interpolation of the NWP output: Interpolation techniques. User guide to ECMWF forecast products, 32–33, https://confluence.ecmwf.int/display/FUG/Forecast+User+Guide.

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

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850, https://doi.org/10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2.

Bermowitz, R. J., 1975: An application of model output statistics to forecasting quantitative precipitation. *Mon. Wea. Rev.*, **103**, 149–153, https://doi.org/10.1175/1520-0493(1975)103<0149:AAOMOS>2.0.CO;2.

Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, https://doi.org/10.1175/2010BAMS2853.1.

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

Carrol, K. L., and J. C. Maloney, 2004: Improvements in extended-range temperature and probability of precipitation guidance. *Symp. 50th Anniversary of Operational Numerical Weather Prediction*, College Park, MD, NWS/Amer. Meteor. Soc., P4.6.

Casanova, S., and B. Ahrens, 2009: On the weighting of multi-model ensembles in seasonal and short-range weather forecasting. *Mon. Wea. Rev.*, **137**, 3811–3822, https://doi.org/10.1175/2009MWR2893.1.

Chen, C. H., C. Y. Li, Y. K. Tan, and T. Wang, 2010: Research of the multi-model super-ensemble prediction based on cross-validation. *J. Meteor. Res.*, **68**, 464–476.

Chou, J. F., 1989: Predictability of the atmosphere. *Adv. Atmos. Sci.*, **6**, 335–346, https://doi.org/10.1007/BF02661539.

Erickson, M. J., B. A. Colle, and J. J. Charney, 2012: Impact of bias-correction type and conditional training on Bayesian model averaging over the northeast United States. *Wea. Forecasting*, **27**, 1449–1469, https://doi.org/10.1175/WAF-D-11-00149.1.

Ferro, C. A. T., 2007: Comparing probabilistic forecasting systems with the Brier score. *Wea. Forecasting*, **22**, 1076–1088, https://doi.org/10.1175/WAF1034.1.

Fraedrich, K., C. C. Raible, and F. Sielmann, 2003: Analog ensemble forecasts of tropical cyclone tracks in the Australian region. *Wea. Forecasting*, **18**, 3–11, https://doi.org/10.1175/1520-0434(2003)018<0003:AEFOTC>2.0.CO;2.

Fraley, C., A. E. Raftery, and T. Gneiting, 2010: Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev.*, **138**, 190–202, https://doi.org/10.1175/2009MWR3046.1.

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1.

——, F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. B*, **69**, 243–268, https://doi.org/10.1111/j.1467-9868.2007.00587.x.

Goswami, B. N., and P. K. Xavier, 2003: Potential predictability and extended range prediction of Indian summer monsoon breaks. *Geophys. Res. Lett.*, **30**, 1966, https://doi.org/10.1029/2003GL017810.

Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2012: Comparing TIGGE multimodel foreasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 1814–1827, https://doi.org/10.1002/qj.1895.

Hall, T., H. E. Brooks, and C. A. Doswell III, 1999: Precipitation forecasting using a neural network. *Wea. Forecasting*, **14**, 338–345, https://doi.org/10.1175/1520-0434(1999)014<0338:PFUANN>2.0.CO;2.

Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, https://doi.org/10.1256/qj.06.25.

——, J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2.

——, M. Scheuerer, and G. T. Bates, 2015: Analog probabilistic precipitation forecasts using GEFS reforecasts and climatology-calibrated precipitation analyses. *Mon. Wea. Rev.*, **143**, 3300–3309, https://doi.org/10.1175/MWR-D-15-0004.1.

Hemri, S., M. Scheuerer, F. Pappenberger, K. Bogner, and T. Haiden, 2014: Trends in the predictive performance of raw ensemble weather forecasts. *Geophys. Res. Lett.*, **41**, 9197–9205, https://doi.org/10.1002/2014GL062472.

Herrera, M. A., I. Szunyogh, and J. Tribbia, 2016: Forecast uncertainty dynamics in the THORPEX Interactive Grand Global Ensemble (TIGGE). *Mon. Wea. Rev.*, **144**, 2739–2766, https://doi.org/10.1175/MWR-D-15-0293.1.

Janowiak, J. E., and P. P. Xie, 1999: CAMS-OPI: A global satellite-rain gauge merged product for real-time precipitation monitoring applications. *J. Climate*, **12**, 3335–3342, https://doi.org/10.1175/1520-0442(1999)012<3335:COAGSR>2.0.CO;2.

Ji, L. Y., X. F. Zhi, and S. P. Zhu, 2017: Extended-range probabilistic forecasts of surface air temperature over East Asia during boreal winter (in Chinese). *Trans. Atmos. Sci.*, **40**, 346–355.

Jones, C., D. E. Waliser, J. K. E. Schemm, and W. K. M. Lau, 2000: Prediction skill of the Madden and Julian oscillation in dynamical extended range forecasts. *Climate Dyn.*, **16**, 273–289, https://doi.org/10.1007/s003820050327.

Kim, C., and M. S. Suh, 2013: Prospects of using Bayesian model averaging for the calibration of one-month forecasts of surface air temperature over South Korea. *Asia-Pac. J. Atmos. Sci.*, **49**, 301–311.

Koizumi, K., 1999: An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network. *Wea. Forecasting*, **14**, 109–118, https://doi.org/10.1175/1520-0434(1999)014<0109:AOMTMN>2.0.CO;2.

Langmack, H., K. Fraedrich, and F. Sielmann, 2012: Tropical cyclone track analog ensemble forecasting in the extended Australian basin: NWP combinations. *Quart. J. Roy. Meteor. Soc.*, **138**, 1828–1838, https://doi.org/10.1002/qj.1915.

Liu, J. G., and Z. H. Xie, 2014: BMA probabilistic quantitative precipitation forecasting over the Huaihe basin using TIGGE multimodel ensemble forecasts. *Mon. Wea. Rev.*, **142**, 1542–1555, https://doi.org/10.1175/MWR-D-13-00031.1.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

——, 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513, https://doi.org/10.3402/tellusa.v34i6.10836.

Majumdar, S. J., and R. D. Torn, 2014: Probabilistic verification of global and mesoscale ensemble forecasts of tropical cyclogenesis. *Wea. Forecasting*, **29**, 1181–1198, https://doi.org/10.1175/WAF-D-14-00028.1.

Min, S. K., and A. Hense, 2007: Hierarchical evaluation of IPCC AR4 coupled climate models with systematic consideration of model uncertainties. *Climate Dyn.*, **29**, 853–868, https://doi.org/10.1007/s00382-007-0269-2.

Miyakoda, K., T. Gordon, R. Caverly, W. Stern, J. Sirutis, and W. Bourke, 1983: Simulation of a blocking event in January 1977. *Mon. Wea. Rev.*, **111**, 846–869, https://doi.org/10.1175/1520-0493(1983)111<0846:SOABEI>2.0.CO;2.

Park, Y. Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.*, **134**, 2029–2050, https://doi.org/10.1002/qj.334.

Plaut, G., and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere. *J. Atmos. Sci.*, **51**, 210–236, https://doi.org/10.1175/1520-0469(1994)051<0210:SOLFOA>2.0.CO;2.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.

Räisänen, J., and L. Ruokolainen, 2006: Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus*, **58A**, 461–472, https://doi.org/10.1111/j.1600-0870.2006.00189.x.

Roulin, E., and S. Vannitsem, 2012: Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, **140**, 874–888, https://doi.org/10.1175/MWR-D-11-00062.1.

Scheuerer, M., 2014: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quart.*

J. Roy. Meteor. Soc., **140**, 1086–1096, https://doi.org/10.1002/qj.2183.

——, and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.*, **143**, 4578–4596, https://doi.org/10.1175/MWR-D-15-0061.1.

——, S. Gregory, T. M. Hamill, and P. E. Shafer, 2017: Probabilistic precipitation-type forecasting based on GEFS ensemble forecasts of vertical temperature profiles. *Mon. Wea. Rev.*, **145**, 1401–1412, https://doi.org/10.1175/MWR-D-16-0321.1.

Schmeits, M. J., and K. J. Kok, 2010: A comparison between raw ensemble output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF ensemble precipitation reforecasts. *Mon. Wea. Rev.*, **138**, 4199–4211, https://doi.org/10.1175/2010MWR3285.1.

Shapiro, M. A., and A. J. Thorpe, 2004: THORPEX international science plan. WMO/TD-1246, WWRP/THORPEX 2, 57 pp., https://www.wmo.int/pages/prog/arep/wwrp/new/documents/CD_ROM_international_science_plan_v3.pdf.

Slater, L. J., G. Villarini, and A. A. Bradley, 2017: Weighting of NMME temperature and precipitation forecasts across Europe. *J. Hydrol.*, **552**, 646–659, https://doi.org/10.1016/j.jhydrol.2017.07.029.

Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220, https://doi.org/10.1175/MWR3441.1.

——, T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, **105**, 25–35, https://doi.org/10.1198/jasa.2009.ap08615.

——, ——, and ——, 2013: Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Wea. Rev.*, **141**, 2107–2119, https://doi.org/10.1175/MWR-D-12-00002.1.

Smagorinsky, J., 1969: Problems and promises of deterministic extended range forecasting. *Bull. Amer. Meteor. Soc.*, **50**, 286–311, https://doi.org/10.1175/1520-0477-50.5.286.

Surcel, M., I. Zawadzki, M. K. Yau, M. Xue, and F. Kong, 2017: More on the scale dependence of the predictability of precipitation patterns: Extension to the 2009–13 CAPS Spring Experiment ensemble forecasts. *Mon. Wea. Rev.*, **145**, 3625–3646, https://doi.org/10.1175/MWR-D-16-0362.1.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. Workshop on Predictability*, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 1–25.

Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295, https://doi.org/10.1111/j.2153-3490.1957.tb01885.x.

Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting, 2018: Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa. *Wea. Forecasting*, **33**, 369–388, https://doi.org/10.1175/WAF-D-17-0127.1.

Waliser, D. E., C. Jones, J. K. E. Schemm, and N. E. Graham, 1999: A statistical extended-range tropical forecast model based on the slow evolution of the Madden–Julian oscillation. *J. Climate*, **12**, 1918–1939, https://doi.org/10.1175/1520-0442(1999)012<1918:ASERTF>2.0.CO;2.

Walker, S. H., and D. B. Duncan, 1967: Estimation of the probability of an event as a function of several independent variables. *Biometrika*, **54**, 167–179, https://doi.org/10.1093/biomet/54.1-2.167.

Wanders, N., and E. F. Wood, 2016: Improved sub-seasonal mete-orological forecast skill using weighted multi-model ensemble simulations. *Environ. Res. Lett.*, **11**, 94007, https://doi.org/10.1088/1748-9326/11/9/094007.

Wang, Q. J., A. Schepen, and D. E. Robertson, 2012: Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging. *J. Climate*, **25**, 5524–5537, https://doi.org/10.1175/JCLI-D-11-00386.1.

Xie, P. P., and A. Y. Xiong, 2011: A conceptual model for constructing high-resolution gauge-satellite merged precipi-tation analyses. *J. Geophys. Res.*, **116**, D21106, https://doi.org/10.1029/2011JD016118.

Xie, S. P., C. H. Chang, Q. Xie, and D. X. Wang, 2007: Intra-seasonal variability in the summer south China sea: Wind jet, cold filament, and recirculations. *J. Geophys. Res.*, **112**, C10008, https://doi.org/10.1029/2007JC004238.

Yang, H., D. Zhang, and L. Ji, 2001: An approach to extract ef-fective information of monthly dynamical prediction—The use of ensemble method. *Adv. Atmos. Sci.*, **18**, 283–293, https://doi.org/10.1007/s00376-001-0020-6.

Yussouf, N., and D. J. Stensrud, 2006: Prediction of near-surface variables at independent locations from a bias-corrected ensemble forecasting system. *Mon. Wea. Rev.*, **134**, 3415–3424, https://doi.org/10.1175/MWR3258.1.

Zeng, L., and D. X. Wang, 2009: Intraseasonal variability of latent-heat flux in the South China Sea. *Theor. Appl. Climatol.*, **97**, 53–64, https://doi.org/10.1007/s00704-009-0131-z.

Zhang, H. B., X. F. Zhi, J. Chen, Y. N. Wang, and Y. Wang, 2015: Study of the modification of multi-model ensemble schemes for tropical cyclone forecasts. *J. Trop. Meteor.*, **21**, 389–399.

Zhang, L., F. Sielmann, K. Fraedrich, X. H. Zhu, and X. F. Zhi, 2015: Variability of winter extreme precipitation in Southeast China: Contributions of SST anomalies. *Climate Dyn.*, **45**, 2557–2570, https://doi.org/10.1007/s00382-015-2492-6.

——, and X. F. Zhi, 2015: Multi-model consensus forecasting of low temperature and icy weather over central and southern China in early 2008. *J. Trop. Meteor.*, **21**, 67–75.

Zhi, X. F., and Coauthors, 2013: Multi-model ensemble forecasts of surface air temperature and precipitation using TIGEE da-tasets (in Chinese). *Trans. Atmos. Sci.*, **36**, 257–266.

——, H. X. Qi, Y. Q. Bai, and C. Z. Lin, 2012: A comparison of three kinds of multi-model ensemble forecast techniques based on the TIGGE data. *J. Meteor. Res.*, **26**, 41–51.

Zhuang, W., S. P. Xie, D. X. Wang, B. Taguchi, H. Aiki, and H. Sasaki, 2010: Intraseasonal variability in sea surface height over the South China Sea. *J. Geophys. Res.*, **115**, C04010, https://doi.org/10.1029/2009JD013165.