

Verification and comparison of probabilistic precipitation forecasts using the TIGGE data in the upriver of Huaihe Basin

L.-N. Zhao^{1,2}, F.-Y. Tian², H. Wu^{1,2}, D. Qi², J.-Y. Di^{1,2}, and Z. Wang^{1,2}

¹Public Weather Service Centre, China Meteorological Administration, Beijing 100081, China

²National Meteorological Centre, China Meteorological Administration, Beijing 100081, China

Received: 15 July 2010 – Revised: 12 December 2010 – Accepted: 13 December 2010 – Published: 4 March 2011

Abstract. The precipitation forecasts of three ensemble prediction systems (EPS) and two multi-model ensemble prediction systems (MM EPS) were assessed by comparing with observations from 19 rain gauge stations located in the *Dapoling-Wangjiaba* sub-catchment of Huaihe Basin for the period from 1 July to 6 August 2008. The sample Probabilistic Distribution Functions (PDF) of gamma distribution, the Relative Operating Characteristic (ROC) diagrams, the percentile precipitation and a heavy rainfall event are analyzed to evaluate the performances of the single and multi-model ensemble prediction system (EPS).

The three EPS were from the China Meteorological Administration (CMA); the United States National Centre for Environment Predictions (NCEP); and the European Centre for Medium-Range Weather Forecasts (ECMWF), all were obtained from the TIGGE-CMA archiving centre (THORPEX Interactive Grand Global Ensemble, TIGGE). The MM EPS were created using the equal weighting method for every ensemble member over the test area, the first (MM-1) consisted of all three EPS, the second (MM-2) consisted of the ECMWF and NCEP EPS.

The results demonstrate the level of correspondence between deterioration in predictive skill and extended lead time. Compared with observations and with a lead time of one day, ECMWF performs a little better than other centre's. With over five days in advance, all the three EPS and the two MM EPS don't give reliable probabilistic precipitation forecasts. Both MM EPS can outperform CMA and NCEP for most of the forecasted days, but still perform a little worse than ECMWF. Though variation of daily percentile precipita-

tion and ROC areas show MM-2 outperforms MM-1, gamma distribution indicates much similar performances for all 10-day forecast, and neither is superior to ECMWF.

1 Introduction

Ensemble forecasting was developed as a result of attempts to understand the limits of deterministic prediction of the atmospheric state by the setting of initial state conditions. After decades of development, short-range and, especially, medium-range precipitation forecasts, have been greatly improved by employing ensemble prediction systems (Charba and Klein, 1980; Charba et al., 2003). However, single EPS usually have restrictions in capturing specific atmospheric conditions; consequently multi-model prediction system (MMS) and probabilistic prediction were developed by considering the characteristics of many EPS. Thus TIGGE, a key component of THORPEX (The Observing System Research and Predictability Experiment) was established, providing a very good basis for probabilistic precipitation, which also facilitates the establishment of the hydrologic ensemble prediction experiment (HEPEX) (Schaake et al., 2007). Detailed descriptions of the characteristics of the 10 ensemble systems collected by TIGGE are made by Park et al. (2008) and Matsueda and Tanaka (2008). Buizza (2008) summarized two of the main advantages of ensemble-based probabilistic forecasts as: the ability of an EPS to predict the most likely scenario; and the ability of an EPS to predict the probability of occurrence of any event, and provide more consistent successive forecasts. Many other meteorologists (Wandishin and Mullen, 2001; Verbunt 2007; Pappenberger 2008; Fowler et al., 2007) have documented the advantages



Correspondence to: F.-Y. Tian
(tianfy@cma.gov.cn)

of ensemble-based precipitation forecasts but they also note that many problems need to be solved.

Rainfall is one of the most important weather phenomena which can result in severe floods and huge economic loss. Providing timely and accurate quantitative precipitation forecast (QPF) is a primary goal of operational prediction, and it is a major factor that affects the issuance of flood warning (Gourley and Vieux, 2005). Unfortunately, quantitative precipitation forecasts lose skill more rapidly with range than forecasts of any other surface elements (Sanders, 1986; Roebber and Bosart, 1998). Precipitation obtained from the TIGGE-CMA portal was broadly verified and compared with observations over the *Dapoling-Wangjiaba* sub-catchment.

This paper is organized as follows. The test area and the data are described in Sect. 2. The methodology used to access the ensemble skill is discussed in Sect. 3. The performance of the different ensemble configurations are validated and compared in Sect. 4, and Sect. 5 is the summary and discussion.

2 The test catchment and the datasets

2.1 The datasets preparation

The European Centre for Medium-Range Weather Forecasts (ECMWF), the United States National Centre for Environment Predictions (NCEP), and the China Meteorological Administration (CMA) multi-member 1–10 day total precipitation forecasting with initial time 0000 GMT obtained from the TIGGE-CMA portal were used in this study. Detailed information of the ensemble systems are listed (Table 1). The ECMWF, NCEP and CMA have 51, 21 and 15 ensemble members respectively. Only 1–10 days predicted precipitation data are analyzed for comparing the EPS's performance in this paper.

There are 19 rain gauges in the test catchment providing hourly accumulations of precipitation. The daily accumulations of precipitation from the rain gauges were calculated with quality control implemented by examining the coherence in time and space, and consistency with the synoptic situations. The total predicted precipitation data of ECMWF, NCEP and CMA are used, and two multi-model ensemble systems designed: one was composed with the EPS of ECMWF, NCEP and CMA (MM-1); and the other consisted of ECMWF EPS and NCEP EPS only (MM-2). All five EPSs were evaluated against observations of the 19 rain gauge stations in the test catchment. The test period lasts from 1 July to 6 August, 2008. One heavy rainfall event occurred on 22 July during the test period. It was hoped that the multi-model ensemble procedure would remove existing biases in the single-model EPS, therefore, no quality control and bias correction were implemented for EPS. The equal weighted factor was arbitrarily imposed on the form of MM-1 and MM-2 without considering the forecasting skills of individ-

ual EPS. The results will undoubtedly be affected by the ensemble size of a single EPS used for the multi-model. However, as that is a complex problem related to a specific EPS, it is beyond the scope of this study. Since the model resolutions of different EPS vary from, in order to facilitate comparison of the performances of different EPS, the predicted precipitation of each EPS was interpolated to the 19 individual rain gauge stations using the bilinear interpolation method. The predicted total precipitation of the EPS over the catchment is then computed from the interpolated predicted precipitation.

2.2 The Dapoling-Wangjiaba catchment

The *Dapoling-Wangjiaba* catchment is located at the southwest of the Huaihe Basin, East-China, with a catchment area of about 30 630 km², and altitude ranging from 200 to 500 m. The test catchment is also the origin of the Huaihe River. Figure 1 illustrates the catchment and the locations of the 19 rain gauge stations. The main reason to choose this region as the test catchment is that severe precipitation events usually happen over this area in summer and severe floods occur frequently. Additionally, there is a dense rain gauge network distributed in the test area.

3 Forecast validation techniques

Predicted precipitation at the rain gauge locations of EPS was validated with methods such as the predicted probability distribution (Gamma function distribution), the Relative Operating Characteristic (ROC) and the variation of daily percentile precipitation. The probability distribution of a heavy rainfall event was also analyzed. To obtain these statistical characteristics, all the 19 rain gauge observations and ensemble members were treated equally.

3.1 Gamma distribution

The statistical distributions of precipitation are distinctly asymmetric and skewed (Ison et al. 1971; Wilks, 1990, 2006), and the fact that the gamma distribution provides a flexible representation involving only two parameters for describing precipitation data makes it widely used. The gamma distribution is defined by the Probability Distribution Function (PDF) as following:

$$f(x) = \frac{(x/\beta)^{\alpha-1} \exp(-x/\beta)}{\beta \Gamma(\alpha)} \quad x \geq 0, \alpha > 0, \beta > 0 \quad (1)$$

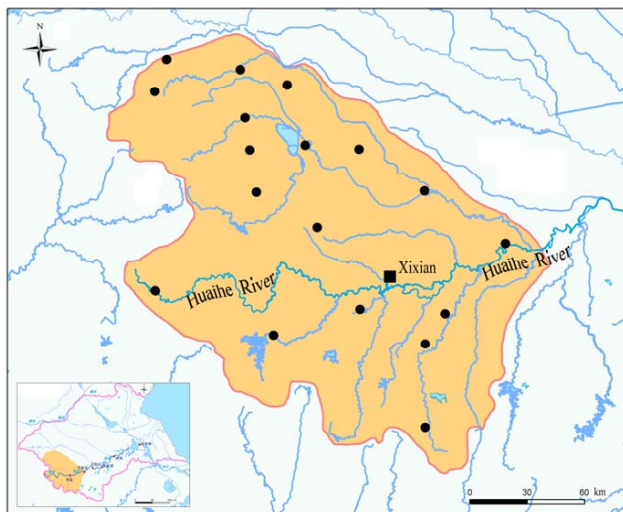
In which α and β are shape and scale parameter, respectively. $\Gamma(\alpha)$ is the gamma function defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (2)$$

To estimate the two parameters fitting for the gamma distribution, a maximum likelihood approach is usually

Table 1. Comparison of the ensemble systems used in this study.

Country/Domain	ECMWF Europe	NCEP USA	CMA China
Ensemble members	51	21	15
Forecast length (in days)	15	16	10
Perturbation method	Singular Vectors	Ensemble Transform	Bred Vectors
Spatial Resolution	1° × 1°	1° × 1°	0.5625° × 0.5625°
Number of levels	62	28	31

**Fig. 1.** Illustration of the test catchment and the locations of 19 stations in the test area.

recommended (Wilks, 1990). A maximum likelihood approximation for the gamma distribution presented by Thom (1958) was employed in this study. By which the Thom estimator for the shape parameter could be given as

$$\hat{\alpha} = \frac{1}{4B} \left(1 + \sqrt{1 + \frac{4B}{3}} \right) \quad (3)$$

Where B is the difference between the natural log of the sample mean and the mean of the logs of the data, derived as

$$B = \ln(\bar{x}) - \frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (4)$$

And then the scale parameter is estimated by

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \quad (5)$$

For cases where the scale parameter $\hat{\beta}$ does not equal to 1.0, the transformation of $x' = x/\hat{\beta}$ should be performed to obtain standardized variables. In this study, the sample size (N)

varied greatly as ensemble members varied from EPS to EPS. N can be given as:

$$N = n \cdot M \cdot D \quad (6)$$

Where n is the station number in the test area, M is the ensemble member used in each EPS, and D is the total tested days. Both the observation and prediction whose daily precipitation amount is less than 0.01 mm day⁻¹ are treated as “no precipitation”.

3.2 ROC curve

The ROC curve has become increasingly popular as a measurement of forecast discrimination to distinguish between an event and non-event (Buizza and Palmer, 1998; Kharin and Zwiers, 2003). The area under the ROC curve is a scalar measure that is frequently used to summarize the resolution. The perfect value is 1.0 and the no-skill value is 0.5. Here, hits and false alarm rate pairs were computed according to the thresholds of 0.1 mm, 5.0 mm, 10.0 mm, 15.0 mm and 25.0 mm, to reveal the performance of the five EPS over the test area.

3.3 Areal percentile precipitation

An established percentile method presented by Hyndman (1996) was adopted for the areal percentile precipitation. The equation can be given as

$$Q_i(p) = (1 - \gamma)A_{(j)} + \gamma A_{(j+1)} \quad (7)$$

Where $j = \text{int}(p \cdot n + (1 + p)/3)$, $\gamma = p \cdot n + (1 + p)/3 - j$, p is the percentile, $Q_i(p)$ is the i -th percentile areal precipitation, A is the array of the forecasted areal precipitation in ascending order, and n the ensemble members. The areal precipitation was obtained by averaging the records of 19 observations or simulated precipitation values.

4 Results

The probability of predicted daily precipitation of different EPS and MM-1 and MM-2 is compared in Fig. 2 using fitted gamma distribution functions. Furthermore, to evaluate how

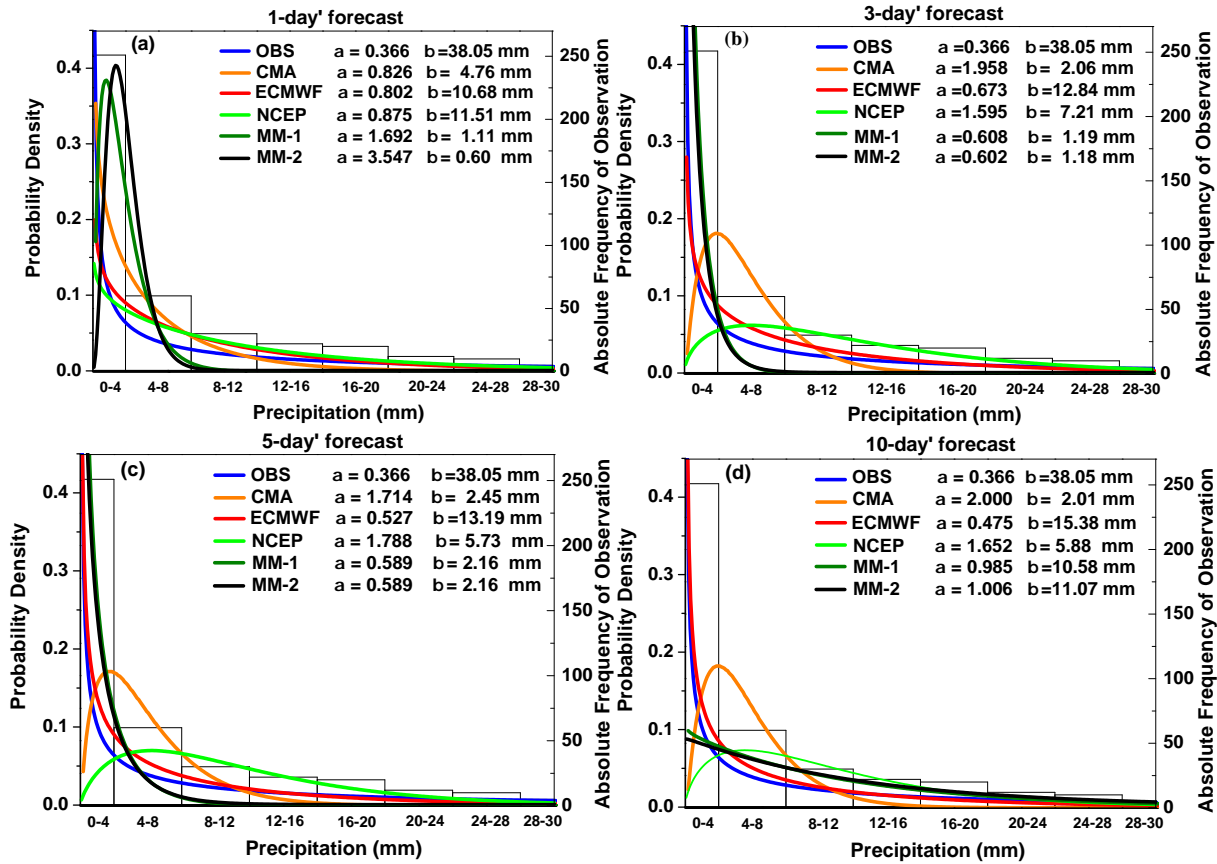


Fig. 2. Gamma distribution density function of daily precipitation of 5 EPSs, CMA (orange line), ECMWF (red line), NCEP (green line), MM-1 (olive line) and MM-2 (black line) with a lead time of 1-day (a), 3-day (b), 5-day (c), and 10-day (d). The Gamma distribution density function of observation is blue line. The histogram is observation sample frequencies of the daily precipitation.

closely the probability density function of gamma distribution fits with the observation, the sample frequency curve of daily precipitation records is shown here too.

For 1-day forecasts, all the five EPS give high probability of predicted precipitation which are less than 8.0 mm day⁻¹. Though the three single EPS show similar results, the curve of probability density function for ECMWF is close to the frequency of the observations' daily precipitation. Both the MM-1 and MM-2 which are slanted to the left strongly show no apparent superiority compared to the three single EPS. Compared to observation, both ECMWF and NCEP slightly underestimate showers and light rain and overestimate moderate rain. CMA tends to overestimate the daily precipitation less than 8.0 mm day⁻¹ and underestimate heavy rainfall especially heavier than 20 mm day⁻¹.

For 3-day forecasts, ECMWF shows a similar performance to the 1-day forecasts. Both PDFs of NCEP and CMA illustrate strong skew to the left, indicating overestimates of moderate rain. The skew to the left for MM-1 and MM-2 were both greatly improved compared to the 1-day forecasts, however, both show overestimates for show-

ers and light rain and underestimates for precipitation heavier than 4.0 mm day⁻¹. The results of the 5-day forecasts are very similar to the 3-day forecasts, except that ECMWF gives an overestimate for all precipitation less than approximately 16.0 mm day⁻¹. However, for 10-day forecasts, only ECMWF has a similar PDF to the PDF of observed precipitation albeit with overestimates for light and moderate rain and underestimates for heavier rain. Both NCEP and CMA have similar gamma distributions compared to 3-day and 5-day forecasts. The two multi-model MM-1 and MM-2 EPSs have $\hat{\alpha}$ approach 1.0 and $\hat{\beta}$ approximately 10.0 mm, and the function intersects the vertical axis at about 0.1 for $x = 0.0$ mm. However, the two MM EPS PDFs apparently overestimate daily rainfall intensities higher than 3.0 mm day⁻¹, in contrast to any of the three single EPS.

Comparing the all 1-day, 3-day, 5-day and 10-day lead times (from Fig. 2a to d), it is clear that $\hat{\alpha}$ for ECMWF uniformly decreases from 0.802 (1-day) to 0.475 (10-day), while $\hat{\beta}$ uniformly increases from 10.68 mm (1-day) to 15.38 mm (10-day), which is closest to the observation values as the lead time extends from 1-day to 10-day.

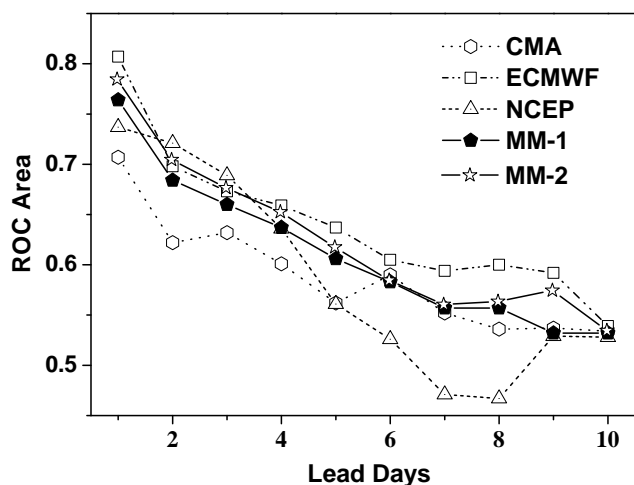


Fig. 3. Variation of ROC areas as forecast lead days ranging from 1 to 10 days.

Comparison of the ROC curves (not shown) with a lead time of 1-day, 3-day, 5-day, and 10-day, and calculating the areas under the ROC curves using the simple trapezoid method, yields conclusions largely consistent with those drawn from the previously analysis. One notable exception is that NCEP and MM-2 outperform ECMWF for 2-day and 3-day forecast, but NCEP also gives the worst result for lead times longer than 5-day. Note that except NCEP, all EPS show areas below the ROC curves greater than 0.5 for all 10 day forecasts, thus showing the ability to discriminate precipitation events. The areas under the fitted ROC curves displayed very similar behavior (Fig. 3). Even though they were constructed with two or three of the EPS, the MM-1 and MM-2 showed no improvement compared to ECMWF with all 10 day's forecast. MM-2 is a little superior to MM-1 as the ROC area is always higher.

Figure 4 shows the variation of daily areal observed and predicted rainfall with lead times of 1-day, 5-day, and 10-day for all EPSs. All EPSs illustrate the temporal variation of the precipitation well compared to the observations. For 1-day forecasts, most of the observations are well included within the 5-th and 99-th percentile precipitation of ECMWF and MM-2. CMA shows a severe underestimation of the probability of precipitation compared to observation. The probability of precipitation from NCEP is a little worse than that from ECMWF, but much better than that from CMA. MM-1 which is the combination of three EPSs usually shows large extension of the box-and-whisker plots. An inspection of Fig. 4 shows that, as lead times increase from 1-day to 10-day, the extension of the box-and-whisker plots increases indicating an increase of spread. With MM-1 and MM-2, a combination of three and two EPSs respectively, more probabilities were presented: e.g. on 13, 16, 18, July, for 1-day forecasts, the observations were contained in the 5-th to 99-th percentile precipitation, but only part was contained by

a single EPS. With the lead time of 10 days, the 50-th percentile precipitation decreases to 0.0 mm while most of the maximum precipitations have apparent overestimates compared to observations, this means that at least 50 percent of ensemble members give underestimates; much worse than the 1 and 5 day forecasts. MM-2 provides an advantage for 1-day and 5-day forecasts as the spread shows, but not for 10-day forecasts.

The heavy rainfall event on 22 July was further analyzed. For all EPSs 1-day, 5-day and 10-day forecasts the possibility of a heavy rainfall event is well indicated by the variation of daily forecasts, even though the observation was not included between the 5-th and 99-th percentiles of predicted precipitation. Fig. 5 shows, for 22 July, the spatial distribution of precipitation probability exceeding 50.0 mm day^{-1} with a 1-day lead time. The heavy rainfall event on 22 July registered extreme rainfall of approximately $200.0 \text{ mm day}^{-1}$ from one station. It can be seen from Fig. 5a that the whole Huaihe Basin was covered by heavy rainfall, with the maxima lying east-west oriented, and a little to the north of the test catchment. All of the three EPS captured this precipitation event, but the high probability area varied greatly. CMA gave two high probability centres, both lying far from the test area. ECMWF performs better than NCEP for the heavy rainfall event prediction with one day lead time. MM-1 and MM-2 give similar results with feasible high probability and better spatial distribution than ECMWF albeit with a lower probability.

5 Summary and discussion

In this paper, the performance of the ECMWF, NCEP, CMA, and two MM EPS combining the EPS during the verified period over the test area are investigated using the total precipitation data obtained from the TIGGE-CMA portal. Validation and comparison was performed between each other and the observations of 19 rain gauges located in the test catchment, and the main results were summarized.

The CMA usually shows poor performance for forecasting moderate and heavier rainfall while the probability of weak precipitation was usually enhanced. The approximate distribution function, variation of ROC areas indicate forecast accuracy was highly dependant on the lead time. Forecast accuracy deteriorates quickly as lead time extends. The two MM EPS don't show much improvement with comparison to ECMWF, but outperform CMA and NCEP for almost all 10 day forecasts. Effects of CMA to the MM EPS should be overlooked as MM-1 and MM-2 usually give similar results. The percentile areal rainfall shows that the observed precipitation values were usually included by the 25-th and 75-th percentile precipitation with short time forecasts, and for 10 days in advance, moderate and heavier precipitation were captured. However, the box-and-whisker plots spread wider as the forecast time extended, indicating an increase in

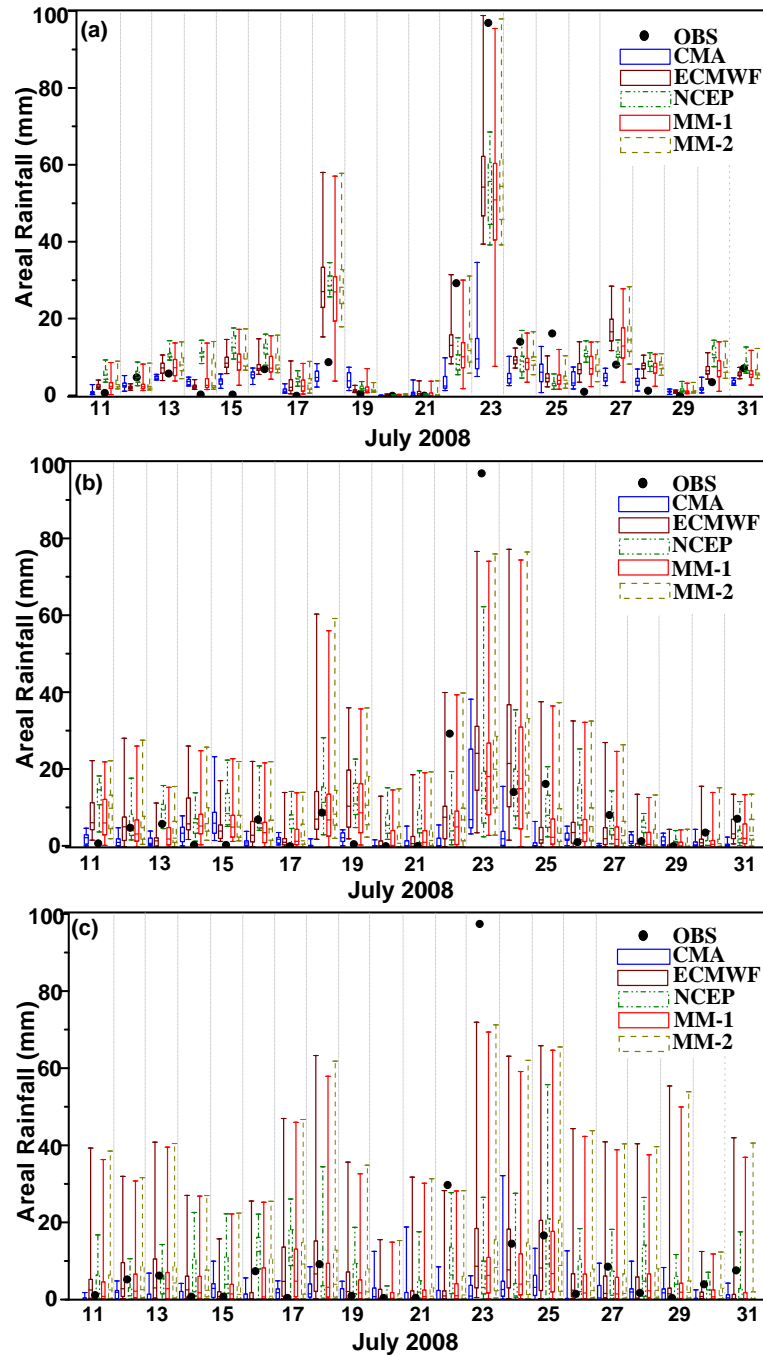


Fig. 4. Percentile area-averaged-precipitation of CMA, NCEP, ECMWF, MM-1 and MM-2 over *Dapoling-Wangjiaba* catchment for lead time of (a) 1-day, (b) 5-day, and (c) 10-day during the study period. The five short lines in every box-and-whisker plot indicate the 5-th, 25-th, 50-th, 75-th, and 99-th percentile precipitation, respectively.

uncertainty. Further analysis of the case of heavy rainfall exceeding 50 mm day^{-1} shows that all EPS could forecast the heavy precipitation event with a short lead time, however, tending to fall outside or skewing the affected area compared to observation.

With the use of EPS and methodology of probabilistic forecasts, especially our understanding of the scientific bases on forecasting extreme precipitation events, providing 3–10 day' probability flood forecast is well developed. Due to advances in numerical weather prediction, the feasibility

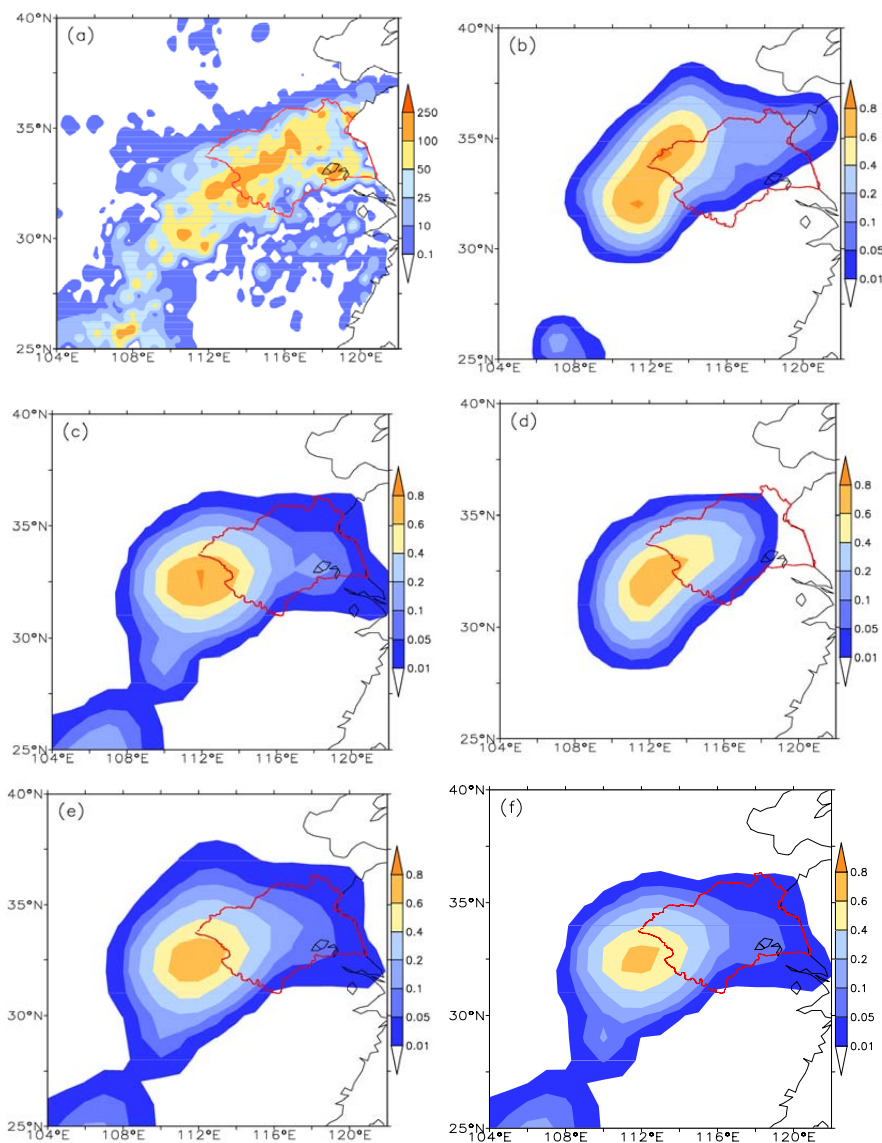


Fig. 5. Precipitation spatial distribution in millimetres of observation (a) and predicted probability of exceeding 50.0 mm day^{-1} with the EPS of (b) CMA, (c) ECMWF, (d) NCEP, (e) MM-1, and (f) MM-2 with a 1-day lead time.

and opportunities for developing probability forecasts have been well documented (Krzysztofowicz, 2001; Thielen et al., 2009). It should be noted that, in this study, the bilinear interpolation method was implemented, and the effect of topography was not considered. In addition, Buizza and Palmer (1998) and Verbunt et al. (2007) pointed out that precipitation is usually affected by the total number of ensemble members, consequently, the effect of member size of the EPS and MM EPS used in this study should be examined. Finally, more work should be done on how best to set the weight of each member of the Multi-Model ensemble forecast.

Acknowledgements. We appreciate the thoughtful and careful reviews by two anonymous referees, and the very valuable proofing of English language offered by B. Parkes. We are also grateful to Editors of Y. He and F. Pappenberger for their great patience and encouragement. This work has partially supported by the Special Funds for Public Welfare of China under Grant No. GYHY201006037, Grant No. GYHY200906007 and by the Huaihe River Meteorological Opening Research Fund (HRM200701).

Edited by: F. Pappenberger

Reviewed by: two anonymous referees

References

- Buizza, R.: The value of probabilistic prediction, *Atmos. Sci. Lett.*, 9, 36–42, 2008.
- Buizza, R. and Palmer, T. N.: Impact of ensemble size on ensemble prediction, *Mon. Weather Rev.*, 126, 2503–2518, 1998.
- Charba, J. P. and Klein, W. H.: Skill in precipitation forecasting in the National Weather Service, *B. Am. Meteorol. Soc.*, 61, 1546–1555, 1980.
- Charba J. P., Reynolds, D. W., Mcdonald, B. E., and Carter, G. M.: Comparative verification of recent quantitative precipitation forecasts in the National Weather Service: a simple approach for scoring forecast accuracy, *Weather Forecast.*, 18, 161–183, 2003.
- Fowler, H. J., Ekstrom, M., Blenkinsop, S., and Smith, A. P.: Estimating change in extreme European precipitation using a multimodel ensemble, *J. Geophys. Res.*, 112, D18104, doi:10.1029/2007JD008619, 2007.
- Gourley, J. and Vieux, B. E.: A method for evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective, *J. Hydrometeor.*, 6, 115–132, 2005.
- Hyndman, R. J. and Fan, Y.: Sample quantiles in statistical packages, *Am. Stat.*, 50(4), 361–365, 1996.
- Ison, N. T., Feyerherm, A. M., and Bark, L. D.: Wet period precipitation and the gamma distribution, *J. Appl. Meteorol.*, 10, 658–665, 1971.
- Kharin, V. V. and Zwiers, F. W.: On the ROC score and probability forecasts, *J. Climate*, 16, 4145–4150, 2003.
- Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, *J. Hydrol.*, 249, 2–9, 2001.
- Matsueda, M. and Tanaka, H. L.: Can MCGE outperform the ECMWF ensemble? *SOLA*, 4, 77–80, 2008.
- Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L., Buizza, R., and de Roo, A.: New dimensions in early flood warning across the globe using grand-ensemble weather predictions, *Geophys. Res. Lett.*, 35, L10404, doi:10.1029/2008GL033837, 2008.
- Park, Y.-Y., Buizza, R., and Leutbecher, M.: TIGGE: preliminary results on comparing and combining ensembles, *Q. J. Roy. Meteor. Soc.*, 134, 2029–2050, 2008.
- Roebber, P. J. and Bosart, L. F.: The sensitivity of precipitation to circulation details, Part I: an analysis of regional analogs, *Mon. Weather Rev.*, 126, 437–455, 1998.
- Sanders, F.: Trends in skill of Boston forecasts made at MIT 1966–84, *B. Am. Meteorol. Soc.*, 67, 170–176, 1986.
- Schaake, J., Hamill, T. M., Buizza, R., and Clark, M.: HEPEX: the hydrological ensemble prediction experiment, *B. Am. Meteorol. Soc.*, 88, 1541–1547, 2007.
- Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System - Part 1: Concept and development, *Hydrol. Earth Syst. Sci.*, 13, 125–140, doi:10.5194/hess-13-125-2009, 2009.
- Thom, H. C. S.: A note on the gamma distribution, *Mon. Weather Rev.*, 86, 117–122, 1958.
- Verbunt, M., Walser, A., Gurtz, J., Montani, A., and Schär, C.: Probabilistic flood forecasting with a limited-area ensemble prediction system: selected case studies, *J. Hydrometeorol.*, 8, 897–909, 2007.
- Wandishin, M. S. and Mullen, S. L.: Evaluation of a short-range multimodel ensemble system, *Mon. Weather Rev.*, 129, 729–747, 2001.
- Wilks, D. S.: Maximum likelihood estimation for gamma distribution using data containing zeros, *J. Climate*, 3, 1495–1501, 1990.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, 2nd edn., Academic Press, United States, 627 pp., 2006.