ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 31, MARCH 2014, 293–304

Improvement of 6–15 Day Precipitation Forecasts Using a Time-Lagged Ensemble Method

JIE Weihua[1,2], WU Tongwen*[2], WANG Jun[3], LI Weijing[2], and LIU Xiangwen[2]

[1]*University of Chinese Academy of Sciences, Beijing* 100049

[2]*Beijing Climate Center, China Meteorological Administration, Beijing* 100081

[3]*Department of Earth and Atmospheric Sciences, University of Nebraska–Lincoln, NE* 68588 – 0340*, USA*

(Received 19 February 2013; revised 4 May 2013; accepted 15 May 2013)

ABSTRACT

A time-lagged ensemble method is used to improve 6–15 day precipitation forecasts from the Beijing Climate Center Atmospheric General Circulation Model, version 2.0.1. The approach averages the deterministic predictions of precipitation from the most recent model run and from earlier runs, all at the same forecast valid time. This lagged average forecast (LAF) method assigns equal weight to each ensemble member and produces a forecast by taking the ensemble mean. Our analyses of the Equitable Threat Score, the Hanssen and Kuipers Score, and the frequency bias indicate that the LAF using five members at time-lagged intervals of 6 h improves 6–15 day forecasts of precipitation frequency above 1 mm d^{-1} and 5 mm d^{-1} in many regions of China, and is more effective than the LAF method with selection of the time-lagged interval of 12 or 24 h between ensemble members. In particular, significant improvements are seen over regions where the frequencies of rainfall days are higher than about 40%–50% in the summer season; these regions include northeastern and central to southern China, and the southeastern Tibetan Plateau.

Key words: time-lagged ensemble system, lagged average forecast, 6–15 day forecasts, precipitation

Citation: Jie, W. H., T. W. Wu, J. Wang, W. J. Li, and X. W. Liu, 2014: Improvement of 6–15 day precipitation forecasts using a time-lagged ensemble method. *Adv. Atmos. Sci.*, **31**(2), 293–304, doi: 10.1007/s00376-013-3037-8.

1. Introduction

Numerical weather prediction (NWP) models have demonstrated high levels of skill in terms of short-range (1–6 day) forecasts of geopotential heights, temperature and precipitation, but less accuracy for medium- to long-range (6 days or longer) forecasts (Qin and van den Dool, 1996; Schmeits and Kok, 2010). In order to address this problem, ensemble forecasting methods have been proposed in many studies and are routinely used in operational centers to improve weather forecasts, subseasonal and seasonal prediction, and even the long-range prediction of climate change (Sivillo et al., 1997; Krishnamurti et al., 2000).

Can ensemble techniques be used to improve medium-range forecasts of precipitation? It is a research topic that has not been explored very well because the focus of most past studies on precipitation prediction has been on short-range (rather than medium-range) ensemble forecasting. Success for improving short-range precipitation forecasts has been demonstrated through the multi-model ensemble method (Ebert, 2001), initial perturbations method (Walser et al., 2004) and the operational centers ensemble prediction system (Schmeits and Kok, 2010). In contrast, Liu (2003) and Vitart and Molteni (2009) showed the value of ensemble methods for forecasting the mean of 30- and 5-day precipitation in the medium- to long-term. Overall, attempts to improve 1–2 week forecasts of daily precipitation using ensemble techniques, so far, have been limited.

The purpose of the present paper is to demonstrate how an ensemble-mean method based on a time-lagged ensemble system (as described in section 2) can produce significantly better forecasts of daily precipitation than a deterministic method for lead times of 6–15 days. The evaluation is conducted through: (1) numerical experiments using version 2.0.1 of the Beijing Climate Center's Atmospheric General Circulation Model (BCC_AGCM2.0.1) for the summers of 1996–2005; and (2) comparisons of model results with ground-based observations (section 2). While the time-lagged ensemble forecast is not a new concept, it is commonly used to produce an ensemble mean or probability for short-range and medium-range forecasts of geopotential heights, temperature, relative humidity, and other fields (e.g., van den Dool and Rukhovets, 1994; Lu et al., 2007). To our knowledge, this concept has not been applied to medium-range forecasts of daily precipitation, although its value for

* Corresponding author: WU Tongwen
Email: twwu@cma.gov.cn

short-range precipitation forecasts has been demonstrated, such as in Yuan et al. (2009).

2. Model, data and ensemble method

The global spectral model, BCC_AGCM2.0.1, is based on the National Center for Atmospheric Research (NCAR) Community Atmosphere Model, with subsequent development by the National Climate Center (NCC) at the China Meteorological Administration (CMA). New features of the BCC_AGCM2.0.1 include different numerical schemes for atmospheric dynamics and several new physical parameterizations (e.g., a new convection scheme and dry adiabatic adjustment scheme). The model has a horizontal resolution of T42 (approximately $2.8125^\circ \times 2.8125^\circ$ transformed grid) and 26 levels in a hybrid sigma/pressure vertical coordinate system (Wu et al., 2008, 2010). Previous studies have shown that BCC_AGCM2.0.1 reproduces the present-day climate fairly well (Wu et al., 2010; Chen et al., 2011).

Owing to the lack of data assimilation systems for BCC_AGCM, an initial coordinated integration method (ICIM) was applied to overcome the problems with the BCC_AGCM's initial condition for the hindcast simulations (Jie and Wu, 2010). In ICIM, the model requires a spin-up period of 10 days (before forecast lead time) by using NCEP-II reanalyses data with a horizontal resolution of $2.5^\circ \times 2.5^\circ$ and 17 levels; the atmosphere temperature and wind fields at each BCC_AGCM grid box and level are interpolated respectively from their counterparts in the reanalyses. During the spin-up, the model boundary conditions of sea surface temperature and sea ice are specified according to version 2 of the NCEP optimal interpolation SST analysis (NCEP-OI2) data (horizontal resolution of T42; 1981–2006; http://www-pcmdi.llnl.gov/projects/amip/AMIP2EXPDSN/BCS/amipbc_dwnld_files/T42/nc/).

The time-lagged ensemble forecast system used in the present work is shown in Fig. 1. The ensemble forecast valid for a particular time was constructed from individual forecasts initialized at different times, but valid for the same time. The lagged average forecast (LAF) assigns equal weights to all ensemble members, and then computes the mean as the final forecast. The total number of ensemble members depends on the time lagging interval, ΔT, between each ensemble member. Thus, if all the model runs initialized during the preceding day are used to construct the time-lagged ensemble, a total of five members can be generated if ΔT is 6 h (e.g., four runs per day), but only two members if ΔT is 24 h (e.g., one run per day). To evaluate the impact of the length of ΔT on forecast performance, we consider experiments using $\Delta T = 24$, 12 and 6 h, respectively, for precipitation forecasts.

Hindcasts were initially performed using BCC_AGCM 2.0.1 with outputs every six h for June—July–August (JJA) 1998 (runs from nine days before 1 June were needed in this work) when El Niño and the anomalous Western Pacific Subtropical High resulted in the strongest persistent precipitation for a century in China (Ding and Hu, 2003). To further

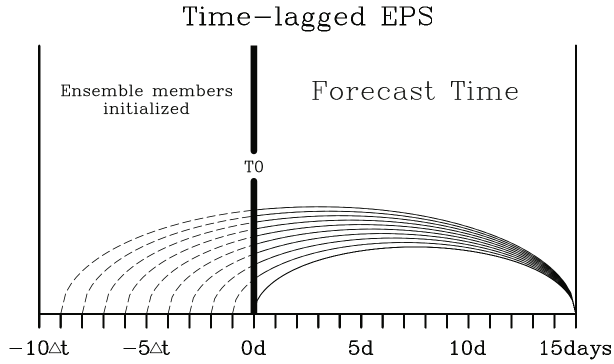

Fig. 1. Diagram showing how a time-lagged ensemble forecast system is constructed in this study.

evaluate the value of the ensemble method, similar simulations were also conducted for cases that were respectively forecasted from 1 June, 1 July and 1 August in other summers during 1996–97 and 1999–2005 (runs from three days earlier in each case with outputs every six hours were needed, as detailed in section 4.2). The forecast lead time for every hindcast was 15 days.

The precipitation data used in this study are based on reports from 2466 rain gauges. Before they are interpolated to model grids [using the Cressman interpolation method (Cressman, 1959)] for model evaluation, these precipitation data undergo quality control processes, such as processing of climate and gauge outliers and homogenization, conducted by the Chinese National Meteorological Information Center.

3. Verification methods

For all numerical experiments, the skill of the daily precipitation forecasts is evaluated using the Equitable Threat Score (ETS), frequency bias (BIA), and the Hanssen and Kuipers Score (HK) (Hanssen and Kuipers, 1965; Schaefer, 1990; Wilks, 1995, respectively). These verification scores are based on a categorical dichotomy between the rain forecasts and observations (e.g., a yes/no statement). For each precipitation threshold, four categories of hits, false alarms, misses, and correct no-rain forecasts are defined in a 2×2 contingency table (Schaefer, 1990). The BIA score indicates underestimation (overestimation) of rainfall frequency with a value lower (higher) than 1.0. The ETS score is used to assess the skill of the predicted rainfall location without random forecast. The range of ETS is from $-1/3$ to 1. An ETS equal to 1 indicates a perfect forecast, while an ETS close to 0.0 or negative indicates poor rain forecasting skill. The HK score is a measure of the accuracy both for events and non-events. A perfect forecast has an HK score equal to 1.0, while a score of -1.0 means that the precipitation forecast capability is worse than a random forecast that receives a score of 0.0.

Further, the reliability of ensemble forecasts is assessed using Rank Histograms (RHs) (Hamill and Colucci, 1998;

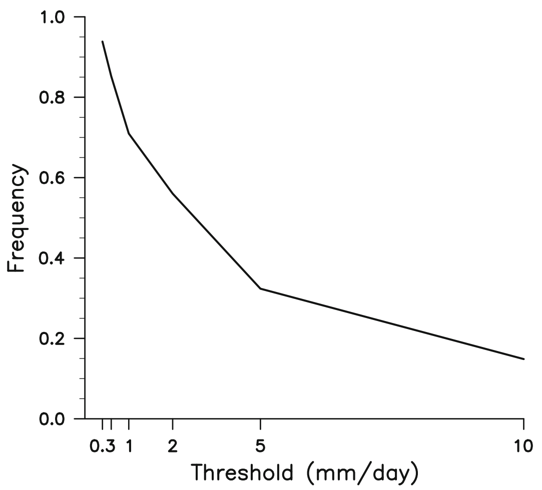

Fig. 2. The mean observed frequency of daily precipitation amount averaged for all grid boxes over China with different precipitation intensities (0.3+ mm, 0.5+ mm, 1+ mm, 2+ mm, 5+ mm and 10+ mm) during summer 1998.

Hamill, 2001), which are generated by computing the rank of observed precipitation relative to values from an ensemble sorted from lowest to highest for all grid boxes. Thus, the number of possible ranks is the number of ensemble members plus 1. A U-shaped rank indicates a lack of variability in the ensemble, and a uniform rank shows the ensemble is dispersed and reliable.

In this study, we mainly focus on forecasts for rainfall of 1 mm and above per day (hereafter 1+ mm) and 5 mm and above per day (hereafter 5+ mm). As shown in Fig. 2, the occurrence frequencies of 1+ mm and 5+ mm rainfall in China accounted for more than 73% and 33% of all rainfall days in summer 1998, respectively. Thus, the 1+ mm rainfall is taken as a relative high frequency event and the 5+ mm as a relative low frequency event.

4. Results

4.1. *Verification for summer 1998*

4.1.1. *Overall skill*

To gain a first estimate of what is the longest time that an ensemble member can be lagged (from the initial time) to maintain its usefulness for improving the precipitation forecast, we evaluate the LAF-based prediction of precipitation by the selection of the time-lagged interval of 24 h. Figure 3a shows the ETS scores of the LAF method (including the deterministic forecast) using members (with $\Delta T = 24$ h) for forecasting rainfall of 1+ mm over China during JJA as a function of forecast time length for up to 15 days in 1998. It supports a conclusion that the skill for forecasting 6–15 day 1+ mm precipitation can be improved by LAF. The ETS score for the deterministic forecast (black line) decreases from about 0.3 to 0.15 during the first six days, and thereafter gradually declines for lead time beyond six days (Fig. 3a). In comparison, the LAF method (dashed colored lines) using $\Delta T = 24$ h generally shows a significant enhancement of ETS scores for forecasts beyond 5 days. The enhancement is especially noticeable for the LAF forecasts using four members (time lagging of three days), and further inclusion of other ensembles that are lagged by more than $4 \times \Delta T$ neither degrades nor dramatically increases the ETS score. These results suggest that an ensemble within four members is sufficient for an improved forecast of precipitation location.

The conjecture that four ensemble members is sufficient for improving precipitation forecasts is also supported by the analysis of HK score, which accounts for the accuracy of both events and non-events. Figure 3b shows that the HK score for deterministic forecasts (black line) beyond five days is less than 0.2 and continues to decrease with forecast lead time. In contrast, the HK scores for LAF with four members or more (dashed colored lines) are generally close to 0.2 during 6th–15th days, again highlighting the value of lagged ensemble members for medium-range forecasts of precipitation. Furthermore, while the deterministic forecast has a persistent underestimation with BIAs of 0.8–0.9, the LAF using four members generally shows a smaller frequency bias (but overestimation), except beyond 10 days. However, in comparison, an increasing overestimation (BIAs > 1) is presented as more than two members are incorporated in LAF (Fig. 3c). A synthesis of Figs. 3a–c suggests that LAF, using four members, can significantly improve forecasts of daily precipitation for lead times of 6–15 days.

In this work, we take more than 5+ mm rainfall as medium-range heavier precipitation. As shown in Figs. 4a–c, the 6–15 day forecast skill of 5+ mm rainfall is similar to the prediction for 1+ mm rainfall, and the LAF method using four members or more also improves the 6–15 day forecasts of daily precipitation. The corresponding ETS scores are generally higher than the deterministic forecast by about 0.025 (Fig. 4a) and the HK scores increase by about 0.05 beyond six days (Fig. 4b). However, BIA values are larger than 1.0 and increase (indicating more overestimation) when five or more ensemble members at $\Delta T = 24$ h are used (Fig. 4c).

The above results of LAF with $\Delta T = 24$ h suggest that time-lagged forecasts made within the last three days could generally contain useful information for medium-range forecasting. Hence, three days was considered as the maximum length for time lagging in the experiments described next.

To evaluate the sensitivity of the LAF-based prediction to the use of different time-lagged intervals (ΔT) for generating ensemble members, we further analyze the LAF results using $\Delta T = 6$ and 12 h within the maximum lagging length (3 days). Figures 3d and e and 3g and h show that the improvements of 1+ mm rainfall from LAF with either $\Delta T = 6$ or 12 h are similar to the improvements by LAF using $\Delta T = 24$ h (Figs. 3a and b); the corresponding ETS and HK scores are generally higher than the deterministic forecast beyond five days. This result also indicates that the time-lagged ensemble members with lags less than or equal to three days are sufficient for an improved medium-range forecast of precipitation. However, for lead times of 6–15 days, an increasing overestimation (BIAs > 1) is also shown as more members

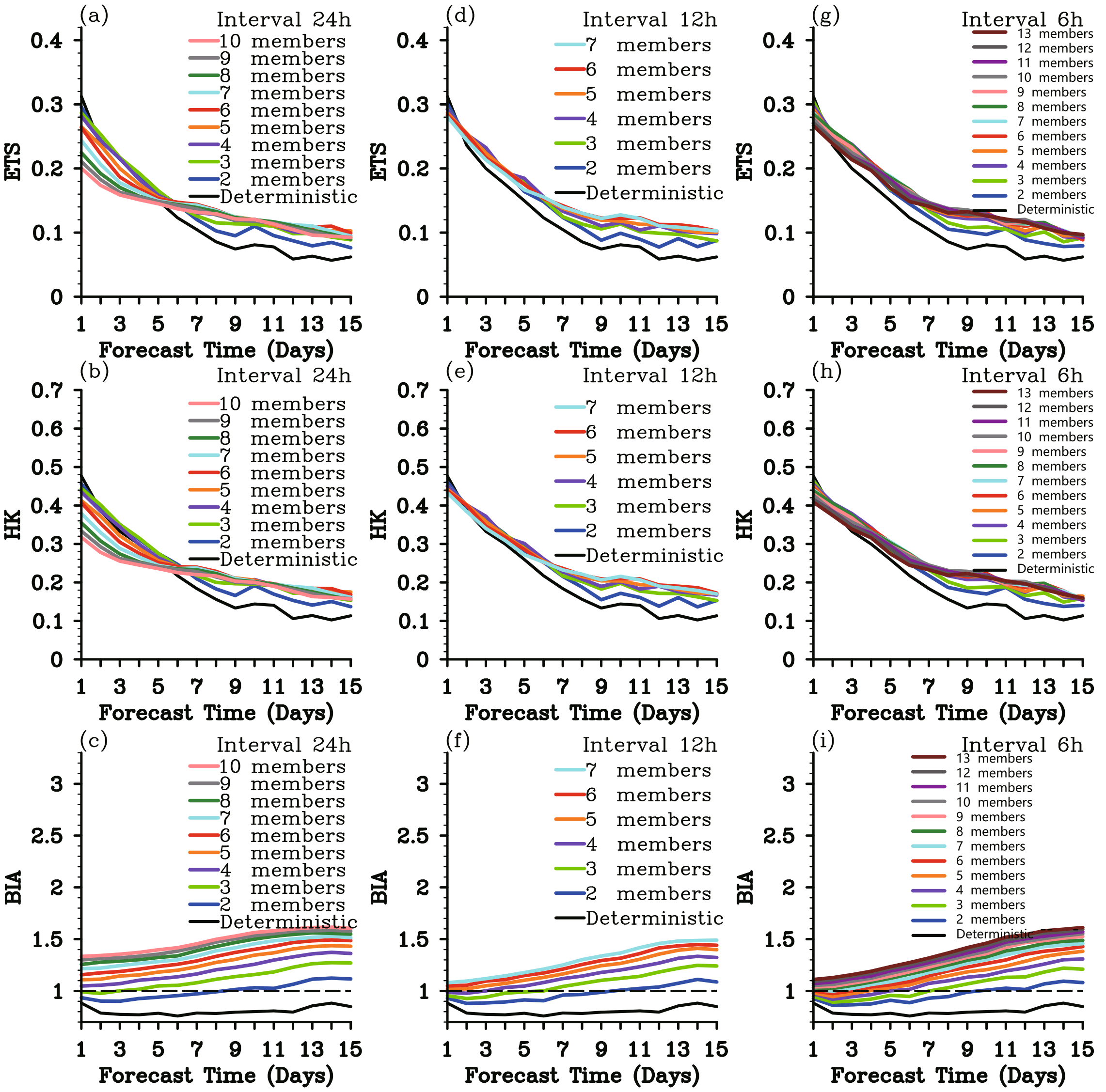

Fig. 3. Evaluation of LAF-based prediction of 1+ mm rainfall with time-lagged intervals of 24 h (a–c), 12 h (d–f) and 6 h (g–i). The averaged ETS scores for prediction of 1+ mm and 5+ mm rainfall as a function of forecast length for up to 15 days during Jun, Jul and Aug 1998 in China are respectively shown as panels (a), (d) and (g) in the first row. The second and third rows are respectively the same as the first row, except that they show the HK (b, e, h) and BIA (c, f, i) scores.

separated by $\Delta T = 6$ or 12 h are included (Figs. 3f and i). On the whole, LAF with five members at $\Delta T = 6$ or 12 h can significantly improve 6–15 day precipitation forecasts. For predicting medium-range 5+ mm rainfall, the improvement from ensembles members with $\Delta T = 6$ or 12 h (Figs. 4d–i) is virtually the same as the counterparts with $\Delta T = 24$ h (Figs. 4a–c).

Figure 5 shows a careful comparison of the three best ensembles from LAF using different time-lagged intervals. The analyses of the ETS, HK and BIA scores consistently indicate that the LAF method using five ensemble members with $\Delta T = 6$ h (orange dashed line) generally shows a slight improvement for 1+ mm precipitation when compared with the five-member $\Delta T = 12$ h LAF (blue dashed line) and the four-member $\Delta T = 24$ h LAF (green dashed line) (Figs. 5a–c). For heavier precipitation, the ETS and HK results indicate that the improvement of 6–15 day precipitation forecasts is relatively insensitive to LAF with different ΔT values (Figs. 5d and e), but the frequency biases from five-member $\Delta T = 6$ h LAF are smallest (Fig. 5f). In addition, despite LAF showing less credence for improving 1–2 day precipitation forecasts for both 1+ mm and 5+ mm, it nevertheless also shows marginal improvement for 3–5 day forecasts of precipitation using the LAF method (Fig. 5).

4.1.2. *Spatial distribution of forecast accuracy rates*

The above analyses of the ETS, HK and BIA scores consistently indicate that the LAF method using five ensemble

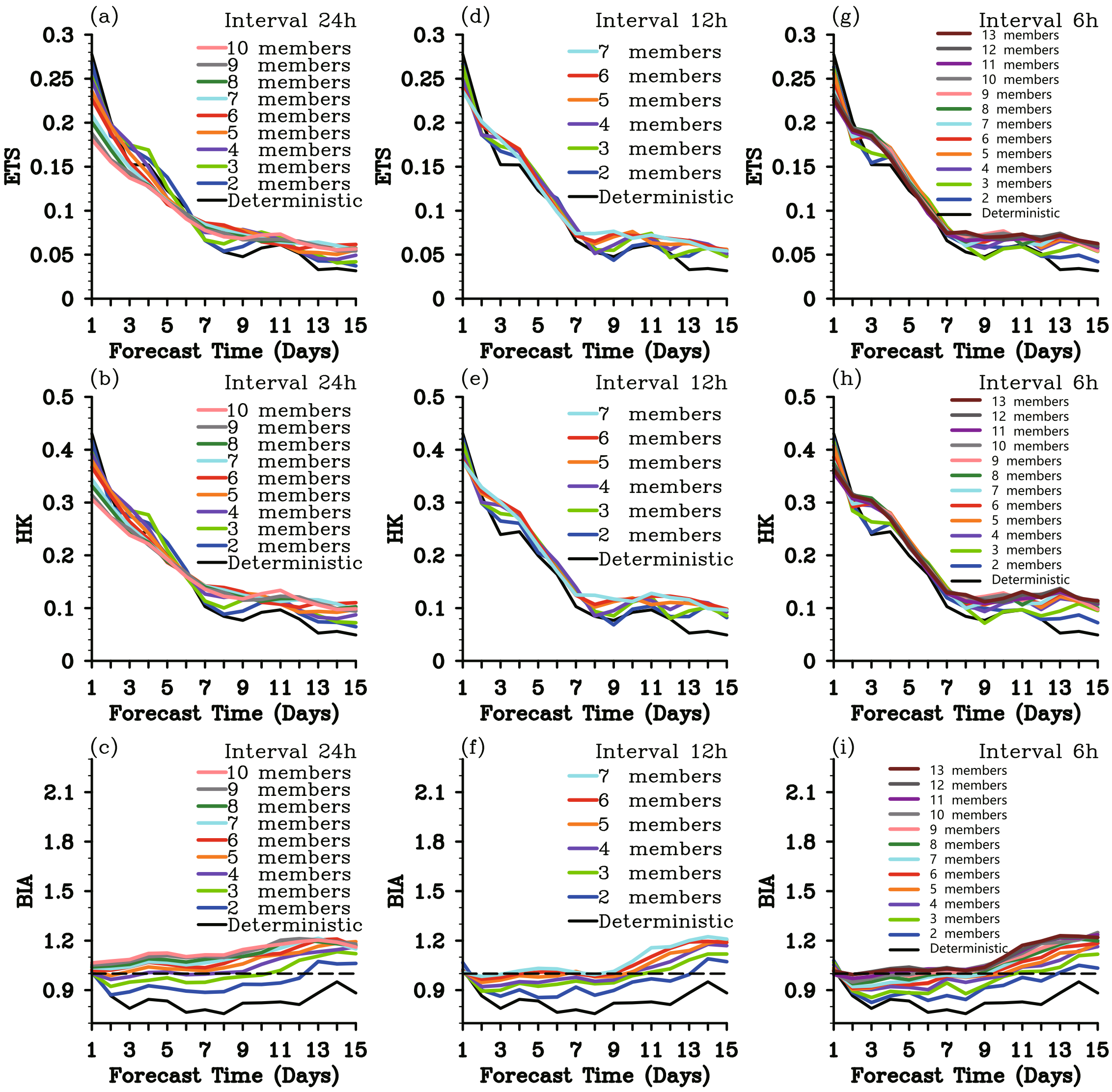

Fig. 4. The same as Fig. 3, but for 5+ mm rainfall.

members with $\Delta T = 6$ h shows a significant improvement over the deterministic forecast for 6–15 day precipitation forecasts. To assess the geographical performance of five-member $\Delta T = 6$ h LAF, we carry out an evaluation of time series of forecasts in each model grid box. Figures 6a–c show the differences of forecast accuracy rates for 1+ mm precipitation between the LAF and deterministic forecast methods at lead times of 8, 11 and 14 days, in which the accuracy rates are computed as (rain day hits + correct no-rain days)/(total days) $\times$ 100% for the ensemble forecasts in summer 1998, and the rain day hits and correct no-rain days in each model grid box for a given lead time are counted based on modeled daily precipitation (valid at the corresponding lead time) in that grid box during 1 June to 31 August. It features as an extended area with positive values of difference located over most regions in northeastern and central to southern China for all the lead times, where the corresponding accuracy rates are generally enhanced by up to 5%–15% higher than their counterparts from the deterministic forecast. The results show that the LAF method provides significant improvements in the prediction of rainfall in these areas, especially the southeastern Tibetan Plateau and part of northeastern China. It is interesting that these regions of significant improvement are located exactly in the areas of 1+ mm precipitation frequency with higher than about 40%–50% occurrence from rain-gauge observations across China (Figs. 6d–f).

In Fig. 6, we can also see that there is a large region of negative values in the northwest of China and middle reaches of the Huanghe River that belong to arid and semi-arid drought regions where there is a lower frequency of 1+ mm precipitation (< 40%) occurrence in summer 1998. This implies no dramatic improvements and indeed a decrease in accuracy rates in these drier areas, as the number of correct no-rain forecasts in less rainy areas are reduced by using the LAF method, although the hits are a little higher (not shown). As for the ensemble forecasts of medium-range

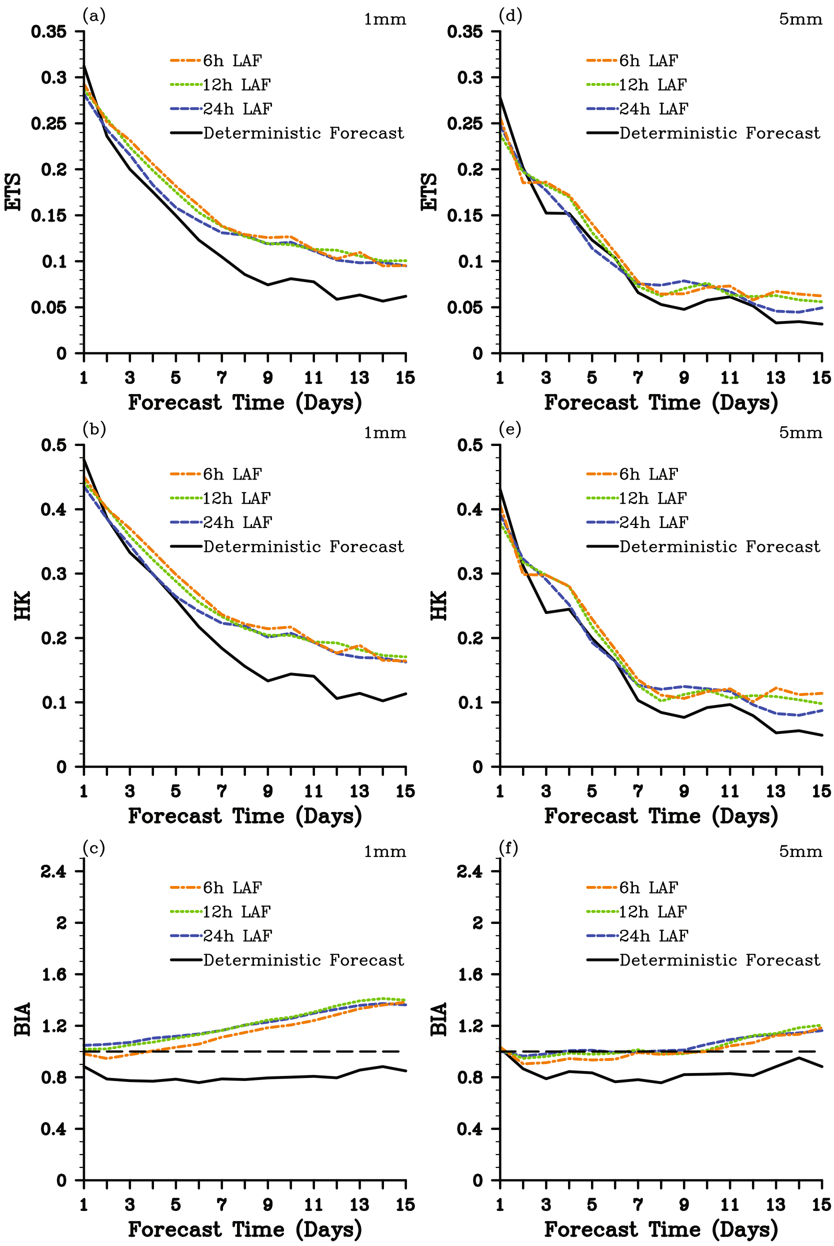

Fig. 5. The same as Fig. 3, but only for the LAF method using five members at 6-h and 12-h time-lagged intervals, and four members at 24-h time-lagged intervals.

heavier precipitation, the LAF accuracy rates, in comparison with their counterparts using the deterministic method, do not increase evidently, although some improvements for hits can be found in the northeast of China, as well as the reaches of the Huanghe and Yangtze rivers (not shown).

4.1.3. *Case study*

The five-member $\Delta T = 6$ h LAF method was used to predict the spatial distribution of daily precipitation during a continuous heavy rain event beginning 23 June 1998 (Jie and Wu, 2010). Figure 7a shows the observed northeast-to-southwest rain belt on 30 June 1998 (the 8th day of forecasting). The deterministic forecast lacks the skill to predict 1+ mm rainfall in the southeast of China, over the middle and lower reaches of the Yangtze River, and slightly underestimates the rainfall in the northeast of China and south of the Tibetan Plateau (Fig. 7b). In contrast, the results from the LAF method show: (1) the prediction of 1+ mm rainfall over the southeast and middle and lower reaches of the Yangtze River is significantly better; (2) the observed 5+ mm rainfall in the southeast of the Tibetan Plateau is slightly better predicted; but (3) a false prediction of 5+ mm rainfall appears in the northeast of China (Fig. 7c).

On 3 July 1998 (the 11th day of forecasting), the observed 1+ mm rainfall is distributed in the south, west and northeast of China, and the 5+ mm main rain belt is located in the south and northeast of China (Fig. 7d). Unfortunately, the deterministic forecast dramatically underestimates rainfall in

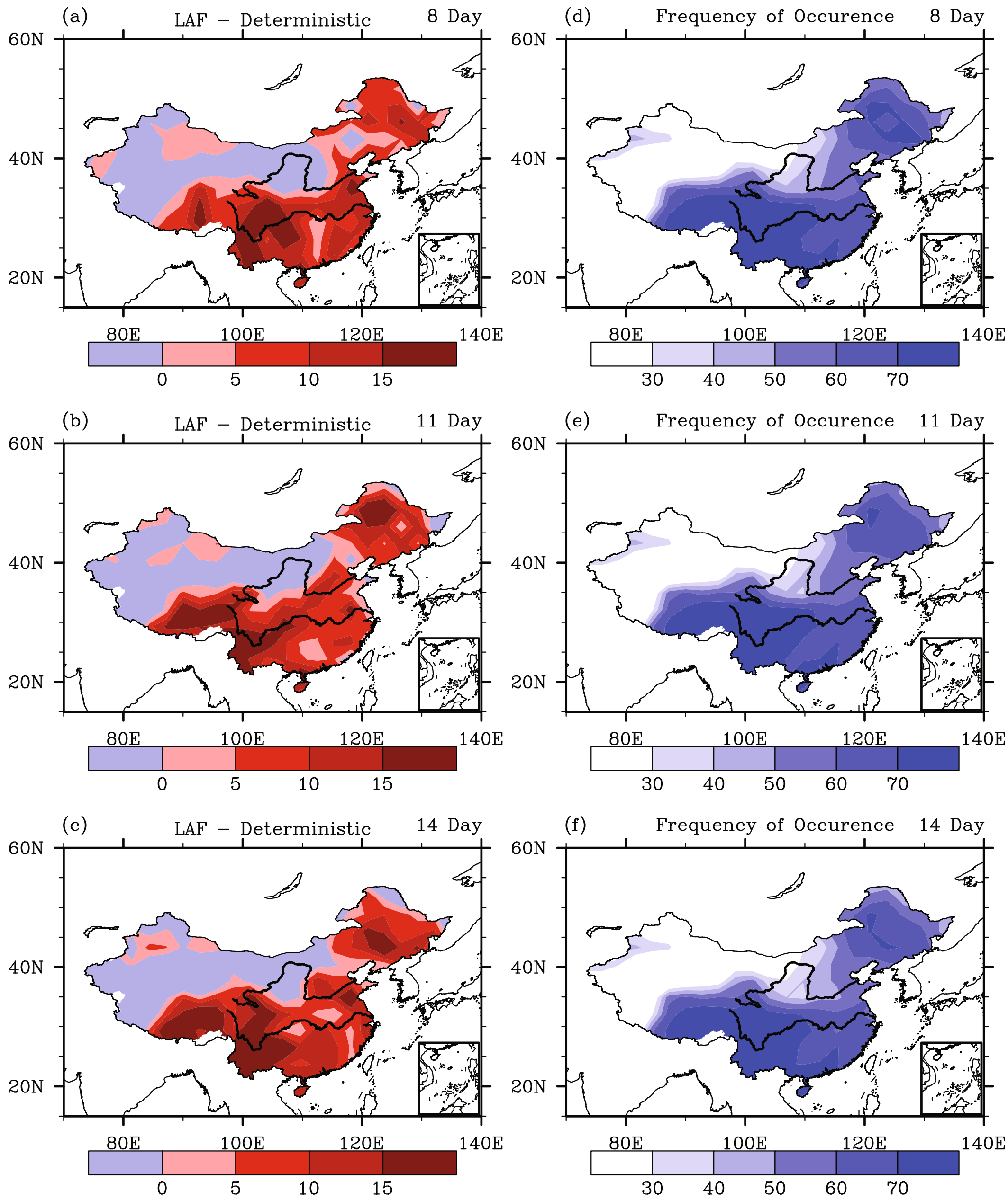

Fig. 6. Geographic distribution of the differences of forecast accuracy rates (a–c) for 1+ mm precipitation between the LAF method using five members at 6-h time-lagged intervals and the deterministic forecast at lead times of 8 (a, d), 11 (b, e) and 14 (c, f) days during Jun to Aug 1998, and the frequency of occurrence for 1+ mm precipitation from observations on the corresponding date. See text for details.

these areas (Fig. 7e). However, the LAF method significantly improves both 1+ mm and 5+ mm predicted rainfall, except for a missed event in the northwest (Fig. 7f).

On the 14th day (6 July 1998), the observed rainfall is concentrated in the south and the northeast of China, over the upper and middle reaches of the Yangtze River, and over the Tibetan Plateau (Fig. 7g). However, the missed event from the deterministic forecast appears in the northeast and over the Tibetan Plateau, and a false prediction occurs over the lower reaches of the Yangtze River (Fig. 7h). By using LAF, the results show the following: (1) the forecast skill of 1+ mm rainfall in the northeast and over the Tibetan Plateau (despite the predicted rainfall coverage being large) is enhanced; (2) the 5+ mm rainfall begins to occur in the northeast even though the predicted coverage is too small compared to the observation; and (3) there is no dramatic improvement in rainfall prediction over the lower reaches of the Yangtze River (Fig. 7i).

4.2. *Verification for other years*

Finally, to verify how LAF performs in "average" years with more normal precipitation conditions, we applied the

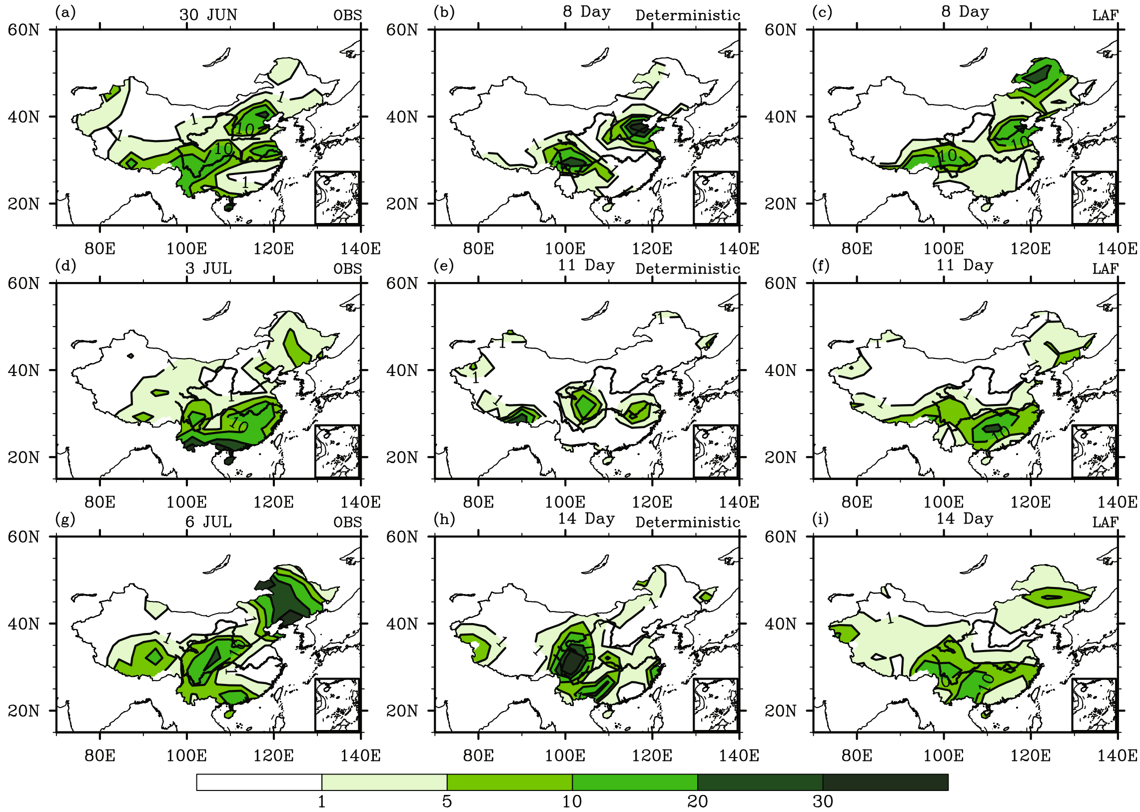

Fig. 7. Spatial distribution of daily precipitation over China on 30 Jun (top row), 3 Jul (middle row), and 6 Jul (bottom row) 1998 from observations (a, d, and g in the left panel), and precipitation forecasts begun on 23 June 1998 from the deterministic forecast method (b, e, and h in the middle panel) and the LAF method using five members at time-lagged intervals of 6 h (c, f, and i in the right panel). Units: mm d^{-1}.

five-member $\Delta T = 6$ h LAF to the summers of 1996–2005. Figure 8 shows the averaged ETS, HK and BIA scores for 30 cases that are from every prediction for 15-day lead times beginning 1 June, 1 July and 1 August in 10 years (thick lines). When compared with the deterministic forecast for lead times of 6–15 days, the ETS and HK scores are increased by the LAF (Figs. 8a and b), although the frequency biases are not reduced as the corresponding BIAs are a little higher (Fig. 8c). This supports the conclusion that the LAF method yields improvement for 6–15 day 1+ mm precipitation forecasting that is consistent with the results described in section 4.1 for the wet summer of 1998. For the prediction of 5+ mm rainfall, the improvements are concentrated only for lead times of 4–9 days (Figs. 8d–f). Generally, they are not as significant as that for the summer of 1998, suggesting that the LAF improves the 5+ mm precipitation forecast more substantially in wet years. We also note that, in Fig. 8, the ETS and HK scores of the LAF method are still higher than those of the deterministic forecast in the typical drier summer of 2004 in most rainy areas of China (refer to http://cmdp.ncc.cma.gov.cn/Monitoring/Bulletin/200408/monitoringc/schinarrc.gif).

Similar to Fig. 6, Figs. 9a–c show the spatial distribution of the differences of forecast accuracy rates for 1+ mm precipitation for the 30 cases in 10 years. It also shows that significant improvements in rainfall prediction for days 8, 11 and 14 are still located in northeastern and central to southern China and over the southeastern Tibetan Plateau, where the corresponding accuracy rates are generally about 5%–15% higher than their counterparts of the deterministic forecast. Consistent with results for the typical summer of 1998, there are no improvements in arid and semi-arid regions (masked by white color), such as the northwest of China and the middle reaches of the Huanghe River, where the frequencies of occurrence for 1+ mm precipitation are generally lower than 40% (Figs. 9d–f). It is notable that there is a north-to-south belt of decrease in the forecast accuracy rate in central China at the lead time of the 14th forecast day (Fig. 9c). It is possibly related to the lower occurrence of precipitation in this region at the lead time of the 14th day in these 30 cases (Fig. 9f). The partial improvements in the medium-range heavier precipitation are primarily reflected in the hits for northeastern China, and over the reaches of the Huanghe River and Yangtze River, rather than in the correct no-rain forecasts (not shown).

Finally, to examine the reliability of the time-lagged ensemble system, we plotted rank histograms (RHs) for the statistics of 30 cases (1996–2005) of ensemble forecasts with different initial forecast time separately for 1 June, 1 July and 1 August for 10 years. Figure 10a shows the RHs of the five $\Delta T = 6$ h optimal ensemble members, for the lead time of

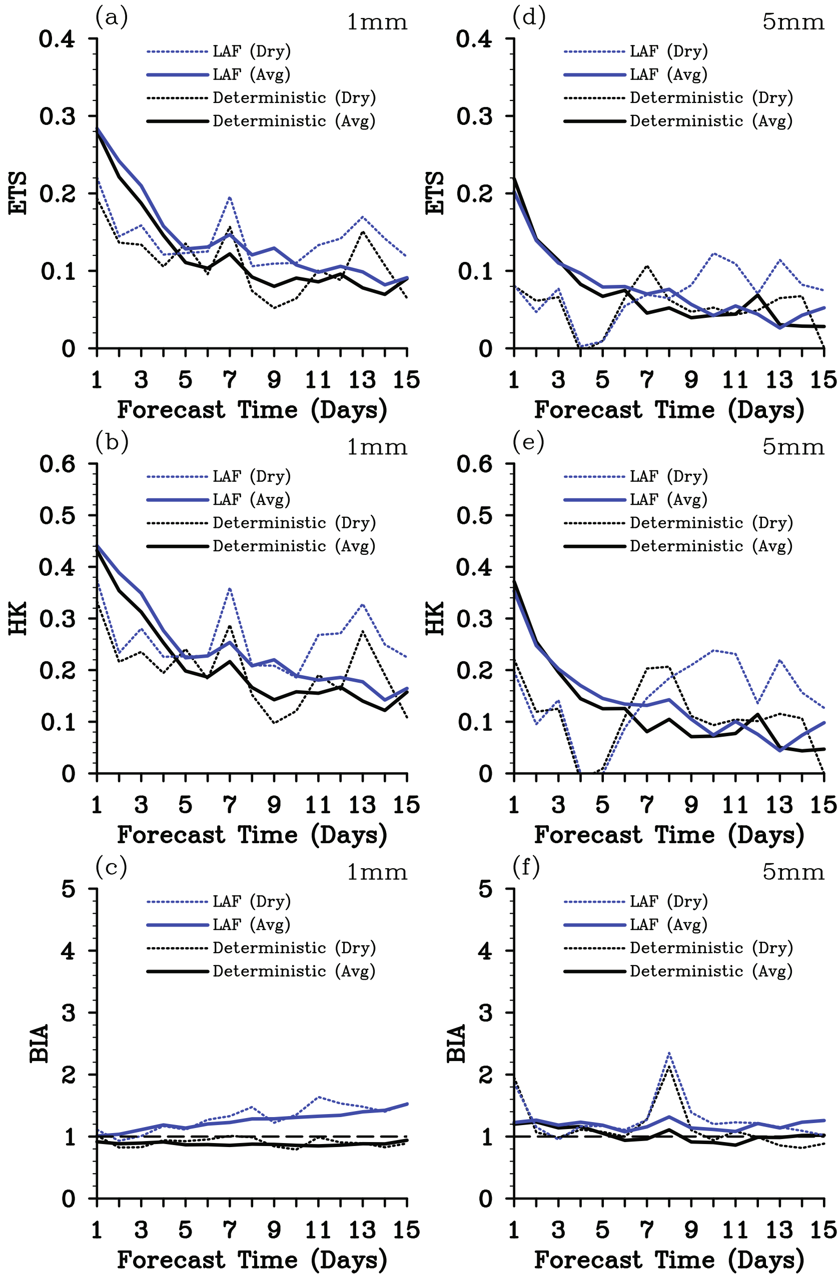

Fig. 8. The same as Fig. 3, but for the LAF method using five members at 6-h time-lagged intervals for 30 cases during 1996–2005. The thick lines are "average" years, and the thin lines are the dry summer year of 2004. See text for details.

the 8th day. It presents a reverse L-shaped RH, suggesting that the ensemble is slightly lacking of variability (Hamill, 2001) and has dry biases [dry biases often cause observations to rank highest with a reversed L shape, while wet biases in ensemble forecasts can lead to an L-shaped RH (Yuan et al., 2009)]. As the lead time increases, the ensemble for the lead times of the 11th and 14th days (Figs. 10b and c) tend to be a uniform rank, but with slightly sloped distributions of the RHs toward the right side, indicating that the ensemble is more reliable in the longer lead time and tends to under-predict precipitation amounts.

A few factors may contribute to this slightly insufficient variability of the time-lagged ensemble system. First, the time-lagged ensemble forecast system is only a single-model ensemble system: it does not account well for model error. Second, all of the initialization times are too close to each other (Lu et al., 2007). It is interesting to note that the under-dispersion and bias is common in many other ensemble forecast systems as well, especially for precipitation forecasting, such as the European Centre for Medium Range Weather Forecasts (ECMWF) ensemble prediction system (Mullen and Buizza, 2002), the NCEP regional spectral model ensem-

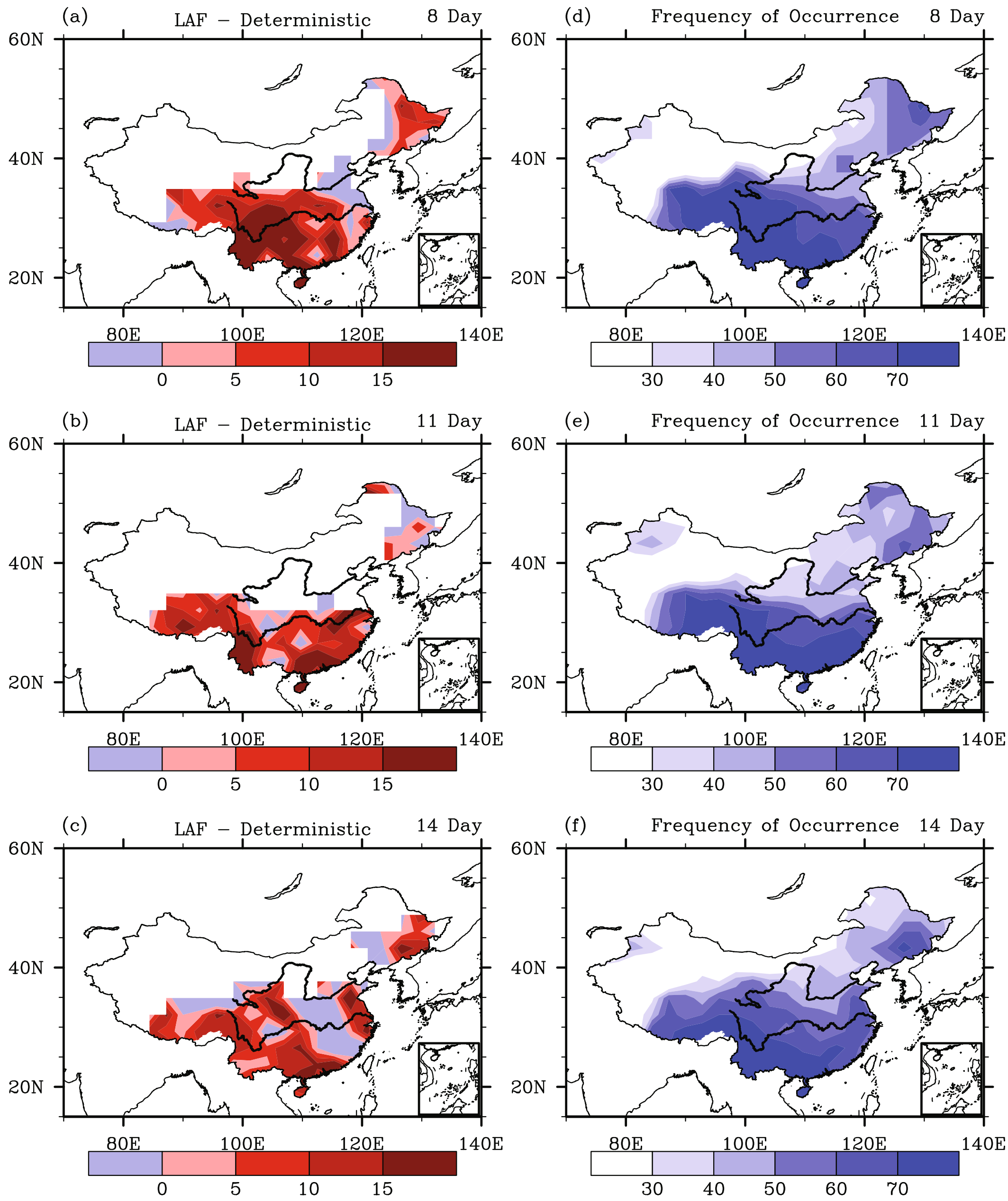

Fig. 9. The same as Fig. 6, but only for the LAF method for 30 cases during 1996–2005. See text for details.

bles (Yuan et al., 2007), the NCEP multi-model short-range ensemble forecast system (Stensrud and Yussouf, 2007), and the Canadian ensemble forecast system (Peel and Wilson, 2008). For example, the RH diagram of the NCEP multi-model short-range ensemble forecast system is similar to that of our ensemble system. It is noted that such a kind of bias is an indicator of the deficiency of the overall modeling system, and hence does not necessarily occur for every single forecast in each model grid box. An empirical correction of such bias in the ensemble forecast, through an increase of the model horizontal resolution (Mullen and Buizza, 2002), the development of a multi-model ensemble method (Yuan et al., 2009), or the improvement of model physical parameterizations, is an active area for research, and will be our focus in future studies.

5. Summary and discussion

While the time-lagged ensemble technique has been instrumental in improving medium-range forecasts of geopotential heights, this paper is among the first to demonstrate its value for the improvement of 6–15 day precipitation forecasts. The LAF with $\Delta T = 24$ h suggests that time-lagged forecasts made within the last three days could generally contain useful information valuable for improving medium-range forecasts of precipitation. According to the sensitivity

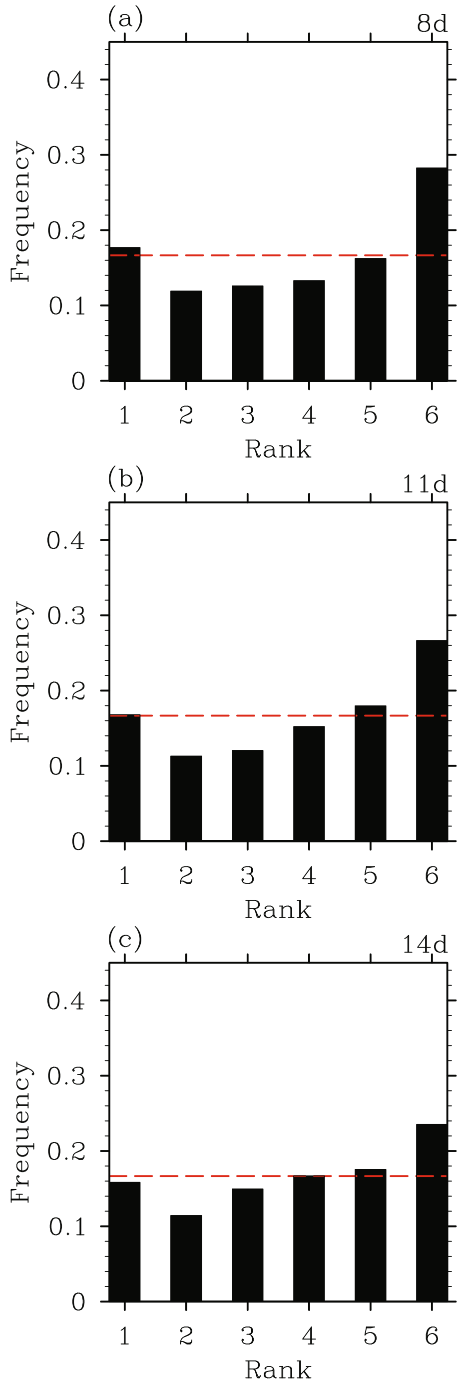

Fig. 10. Rank histograms of time-lagged ensemble forecasts with 6-h time-lagged intervals for the (a) 8th; (b) 11th; and (c) 14th day precipitation forecasts. The histograms are based upon the results of 30 cases in 10 years. The abscissa indicates the rank of the observation among all ensemble members, while the ordinate indicates the frequency of the total sample for each rank. The red lines denote an averaged rank. See text for details.

analyses for the selection of the time-lagged intervals of 6, 12 and 24 h within a lagging of 3 days, the LAF method using five members at $\Delta T = 6$ h was found to be able to most significantly improve the 6–15 day forecasts of daily precipitation. Meanwhile, the geographic distribution of the accuracy rates for all daily forecasts during summer 1998 further supports that the LAF method can enhance the forecast skill in rainy areas where the frequencies of rainfall days are higher than about 40%–50%, such as northeastern and central to southern China, and over the southeastern Tibetan Plateau. The application of LAF for "average" years further verifies that the LAF is effective for 6–15 day precipitation, although the improvement in 5+ mm precipitation forecasting is less significant in "average" years than wet years (e.g., 1998).

The reason for the improvements by LAF relative to the deterministic forecast method is very complicated. It is possibly attributed to the time-lagged ensemble method, which is analogous to a method creating initial perturbation at the forecast time. The ensemble for initial perturbations can, to some extent, decrease the influence due to uncertainty of the model initial state. However, this work also shows that the LAF method still has its limitations in improving precipitation forecasts in arid and semi-arid drought regions over China in summer. This might be linked to the deficiency of describing the precipitation process in the model, which can be overcome by improving the model initial condition Once there are some members of forecast precipitation occurrence in the period of no rainfall from observations, the averaged precipitation for all ensemble members is easily overestimated in arid and semi-arid drought regions.

In this work, ensemble techniques for medium-range forecasts were only tested for summer rainfall in eastern Asia. Admittedly, care must be exercised, and further study is needed to evaluate the LAF method for the prediction of winter precipitation. Nevertheless, the encouraging results shown in this study suggest the feasibility of developing ensemble techniques for the medium-range prediction of precipitation. The LAF method developed in this study has the potential to be used for other regions. In addition, to increase the variability of the time-lagged system, further investigations into the design of the time-lagged multi-model ensemble system, or improving its physical parameterizations, are possibly needed. And finally, if under the constraint of practical limitations models can only run once or twice per day (instead of four times per day), the LAF method using the selection of the time-lagged interval (24 or 12 h) is also recommended.

Acknowledgements. This work was supported by the National Basic Research Program of China (973 Program: Grant No. 2010CB951902), the Special Program for China Meteorology Trade (Grant No. GYHY201306020), and the Technology Support Program of China (Grant No. 2009BAC51B03). The authors also wish to thank Dr. Timothy HUME for technical comments and for proofreading the manuscript.

REFERENCES

Chen, H., R. Yu, J. Li, X. Xin, Z. Wang, and T. Wu, 2011: The coherent interdecadal changes of East Asia climate in mid-summer simulated by BCC_AGCM2.0.1. *Climate Dyn.*, doi: 10.1007/s00382-011-1154-6.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.

Ding, Y., and G. Hu, 2003: A study on water vapor budget over China during the 1998 severe flood periods. *Acta Meteorologica Sinica*, **61** (2), 129–145. (in Chinese)

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.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560.

Hamill, T. M., and S. J. Colucci, 1998: Evaluation of Eta–RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.*, **126**, 711–724.

Hanssen, A. W., and W. J. A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Mededelingen en verhandelingen*, **81**, 2–15.

Jie, W., and T. Wu, 2010: Hindcast for the 1998 Summer Heavy Precipitation in the Yangtze and Huaihe River valley using BCC_AGCM2.0.1 model. *Chinese J. Atmos. Sci.*, **34** (5), 962–978. (in Chinese)

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.

Liu, Y., 2003: Prediction of monthly-seasonal precipitation using coupled SVD patterns between soil moisture and subsequent precipitation. *Geophys. Res. Lett.*, **30**, 1827, doi: 10.1029/2003GL017709.

Lu, C., H. Yuan, B. E. Schwartz, and S. G. Benjamin, 2007: Short-range numerical weather prediction using time-lagged ensembles. *Wea. Forecasting*, **22**, 580–595.

Mullen, S. L., and R. Buizza, 2002: The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system. *Wea. Forecasting*, **17**, 173–191.

Peel, S., and L. J. Wilson, 2008: A diagnostic verification of the precipitation forecasts produced by the Canadian ensemble prediction system. *Wea. Forecasting*, **23**, 596–616.

Qin, J., and H. M. van den Dool, 1996: Simple extensions of an NWP model. *Mon. Wea. Rev.*, **124**, 277–287.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.

Schmeits, J. M., 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.

Sivillo, J. K., J. E. Ahlquist, and Z. Toth, 1997: An ensemble forecasting primer. *Wea. Forecasting*, **12**, 809–818.

Stensrud, D. J., and N. Yussouf, 2007: Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system. *Wea. Forecasting*, **22**, 3–17.

van den Dool, H., and L. Rukhovets, 1994: On the weights for an ensemble-averaged 6–10 day forecast. *Wea. Forecasting*, **9**, 457–465.

Vitart, F., and F. Molteni, 2009: Dynamical extended-range prediction of early monsoon rainfall over India. *Mon. Wea. Rev.*, **137**, 1480–1492.

Walser, A., D. Luthi, and C. Schar, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.*, **132**, 560–577.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Wu, T., and Coauthors, 2008: A modified dynamic framework for atmospheric spectral model and its application. *J. Atmos. Sci.*, **65**, 2235–2253.

Wu, T., and Coauthors, 2010: The Beijing climate center for atmospheric general circulation model (BCC-AGCM2.0.1): Description and its performance for the present-day climate. *Climate. Dyn.*, **34**, 123–147.

Yuan, H., S. L. Mullen, X. Gao, S. Sorooshian, J. Du, and H. H. Juang, 2007: Short-range probabilistic quantitative precipitation forecasts over the southwest United States by the RSM ensemble system. *Mon. Wea. Rev.*, **135**, 1685–1698.

Yuan, H., C. Lu, J. A. McGinley, P. J. Schultz, B. D. Jamison, L. Wharton, and C. J. Anderson, 2009: Evaluation of short-range quantitative precipitation forecasts from a time-lagged multimodel ensemble. *Wea. Forecasting*, **24**, 18–38.