

# Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China

Feng Ma,<sup>a,b</sup> Aizhong Ye,<sup>a,b\*</sup> Xiaoxue Deng,<sup>a,b</sup> Zheng Zhou,<sup>a,b</sup> Xiaojie Liu,<sup>c</sup> Qingyun Duan,<sup>a,b</sup>  
Jing Xu,<sup>a,b</sup> Chiyuan Miao,<sup>a,b</sup> Zhenhua Di<sup>a,b</sup> and Wei Gong<sup>a,b</sup>

<sup>a</sup> State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, China

<sup>b</sup> Joint Center for Global Change Studies, Beijing, China

<sup>c</sup> Institute of Geographic Sciences and Natural Resource Research (IGSNRR), Chinese Academy of Sciences (CAS), Beijing, China

**ABSTRACT:** There is an increasing focus on the usefulness of climate model-based seasonal precipitation forecasts as inputs for hydrological applications. This study reveals that most models from the North American Multi-Model Ensemble (NMME) have potential to forecast seasonal precipitation over 17 hydroclimatic regions in continental China. In this paper, we evaluated the NMME precipitation forecast against observations. The evaluation indices included the correlation coefficient ( $R$ ), relative root-mean-square error (RRMSE), rank histogram (RH), and continuous ranked probability skill score (CRPSS). We presented the RRMSE- $R$  diagram to distinguish differences between the performances of individual models. We find that the predictive skill is seasonally and regionally dependent, exhibiting higher values in autumn and spring and lower values in summer. Higher predictive skill is observed over most regions except the southeastern monsoon regions, which may be attributable to local climatology and variability. Among the 11 NMME models, CFS, especially CFSv2, exhibits the best predictive skill. The GFDL and NASA models, which are followed by CMC, perform worse than CFS. The performances of IRI and CCSM3 are relatively worse than that of the other models. The forecast skills are significantly improved in multi-model mean forecasts based on simple model averaging (SMA). The improvement is more obvious for Bayesian model averaging (BMA), which is employed to further improve the forecast skill and address model uncertainty using multiple model outputs, than individual model and SMA.

**KEY WORDS** seasonal precipitation predictions; NMME; BMA; RRMSE- $R$  diagram; China

Received 16 October 2014; Revised 5 March 2015; Accepted 10 March 2015

## 1. Introduction

Precipitation is an active and critical variable in climate dynamics. Furthermore, assessing retrospective climate model forecasts, i.e. the predictive skill of climatic variables, can give valuable baseline information for many applications, including streamflow, flood and drought monitoring, and agriculture and water resource management (He *et al.*, 2010; Pappenberger *et al.*, 2011). Although individual weather events are generally not predictable more than 14 days ahead due to the chaotic climate system, forecasting at seasonal lead times is potentially possible because of interactions between the atmosphere and the slowly varying components (Wang *et al.*, 2007), such as the ocean and land surfaces. The predictability of precipitation depends not only on the investigated region but also on the time scale and the relevant internal atmospheric dynamical processes (Unganai and Mason, 2002; Korecha and Barnston, 2007; Verbist

*et al.*, 2010; Lorenzo *et al.*, 2011; Nicholson, 2014). Therefore, precipitation prediction, model biases, and observational uncertainties have recently attracted substantial attention (Phillips and Gleckler, 2006; Yuan *et al.*, 2011; Liu *et al.*, 2013; Sun *et al.*, 2014).

Recently, multi-model ensemble prediction has become a powerful tool for monthly-to-seasonal time scale predictions to address uncertainties (Peng *et al.*, 2002; Palmer *et al.*, 2004; Doblus-Reyes *et al.*, 2005; Hagedorn *et al.*, 2005; Kirtman and Min, 2009; Lavers *et al.*, 2009). Compared with single-value deterministic weather predictions, ensemble predictions provide not only predictions of the most likely events but also uncertainty information (Park *et al.*, 2008). Moreover, ensemble predictions have the added advantage of extended prediction lead times by considering the uncertainty in initial conditions and in model physics. Furthermore, recent studies have shown that a multi-model ensemble, even using a simple equal weight combination, has higher prediction skill scores than that of any individual model in the prediction of climate variables. The benefits stemming from the use of a multi-model approach have been extensively documented for seasonal to interannual predictions and for long-term climate projections (Palmer *et al.*, 2004; Lewis,

\* Correspondence to: Aizhong Ye, State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China. E-mail: azye@bnu.edu.cn

Table 1. The NMME model properties.

Model name	Abbreviation	Institute	Period (monthly)	Ensemble size	Lead time (months)	Reference
CMC1-CanCM3	CMC1-3	Canadian Meteorological Centre (CMC) – Canada	1981–2010	10	0–11	Merryfield <i>et al.</i> (2013)
CMC2-CanCM4	CMC2-4	Canadian Meteorological Centre (CMC) – Canada	1981–2010	10	0–11	Merryfield <i>et al.</i> (2013)
COLA-RSMAS- CCSM3	CCSM3	National Center for Atmospheric Research (NCAR) – United States	1982–2009	6	0–11	Collins <i>et al.</i> (2006)
GFDL-CM2p1	GFDL	Geophysical Fluid Dynamics Laboratory (NOAA/GFDL) – United States	1982–2010	10	0–11	Zhang <i>et al.</i> (2007)
GFDL-CM2p1- aer04	GFDL-a	Geophysical Fluid Dynamics Laboratory (NOAA/GFDL) – United States	1982–2009	10	0–11	Zhang <i>et al.</i> (2007)
IRI-ECHAM4p5- Anomaly Coupled	IRI-A	International Research Institute for Climate and Society (IRI) – United States	1982–2009	12	0–7	DeWitt (2005)
IRI-ECHAM4p5- Direct Coupled	IRI-D	International Research Institute for Climate and Society (IRI) – United States	1982–2009	12	0–7	DeWitt (2005)
NASA-GMAO	NASA	National Aeronautics and Space Administration (NASA) – United States	1981–2012	10 (6)	0–8	Holdaway <i>et al.</i> (2014)
NASA-GMAO- 062012	NASA-0	National Aeronautics and Space Administration (NASA) – United States	1981–2013	12 (11)	0–8	Holdaway <i>et al.</i> (2014)
NCEP-CFSv1	CFSv1	National Centers for Environmental Prediction (NOAA/NCEP) – United States	1981–2009	15	0–8	Saha <i>et al.</i> (2006)
NCEP-CFSv2	CFSv2	National Centers for Environmental Prediction (NOAA/NCEP) – United States	1982–2009	24	0–9	Saha <i>et al.</i> (2014)

2005; Tebaldi and Knutti, 2007; Weisheimer *et al.*, 2009). Furthermore, Xu *et al.* (2007) found that ensemble predictions perform better than single deterministic control simulations in forecasting precipitation amounts from severe storms. Therefore, the simple model averaging (SMA) (Hagedorn *et al.*, 2005; Duan and Phillips, 2010) and Bayesian model averaging (BMA) (Duan and Phillips, 2010) methods have been utilized to merge climate forecasts from multiple models to produce more skillful predictions. For example, seasonal hydrologic predictions are determined using seasonal climate forecasts from a combination of several climate models with observed climatology in a Bayesian framework over the Ohio River basin in the eastern United States (Luo and Wood, 2008). Miao *et al.* (2013) also found that the performance of BMA in simulating annual temperature dynamics is significantly better than single climate models and SMA.

The North American Multi-Model Ensemble (NMME) is a newly formed multi-model ensemble system for intra-seasonal to interannual (ISI) prediction. The NMME Phase I archive consists of hindcast data from 11 models representing six modelling centres (Kirtman *et al.*, 2014). Table 1 lists the models included in NMME phase I. The motivation for the development of NMME was increasing studies indicating that a multi-model strategy may be a

practical approach for adequately resolving forecast uncertainties (Palmer *et al.*, 2004; Hagedorn *et al.*, 2005; Palmer *et al.*, 2008). The NMME, which was previously reviewed by Kirtman *et al.* (2014), provides a new opportunity to improve the seasonal predictability of precipitation using multi-model approaches (Infanti and Kirtman, 2014). A first look at the NMME system for monthly and seasonal predictions was presented by Zhang *et al.* (2011). Misra and Li (2014) compared seasonal predictability of the Atlantic Warm Pool and seasonal rainfall anomalies over the continental United States using subsets of the NMME. The NMME can also capture some aspects of ENSO variability, including sea surface temperature (SST) anomalies and rainfall at short lead times (Kirtman *et al.*, 2014). Currently, Yuan and Wood (2013) have investigated the capability of seasonal forecasting of global drought onset using NMME predictions. The NMME system is also used for seasonal drought prediction in the United States by the Climate Prediction Center (CPC) (Kirtman *et al.*, 2014). And given its global coverage, it would need to be examined if the NMME predictions can be effectively applied to other countries, including China. This application is important because we are focused on enhancing seasonal drought prediction capability for China using climate predictions from multiple dynamic models, e.g. the NMME.

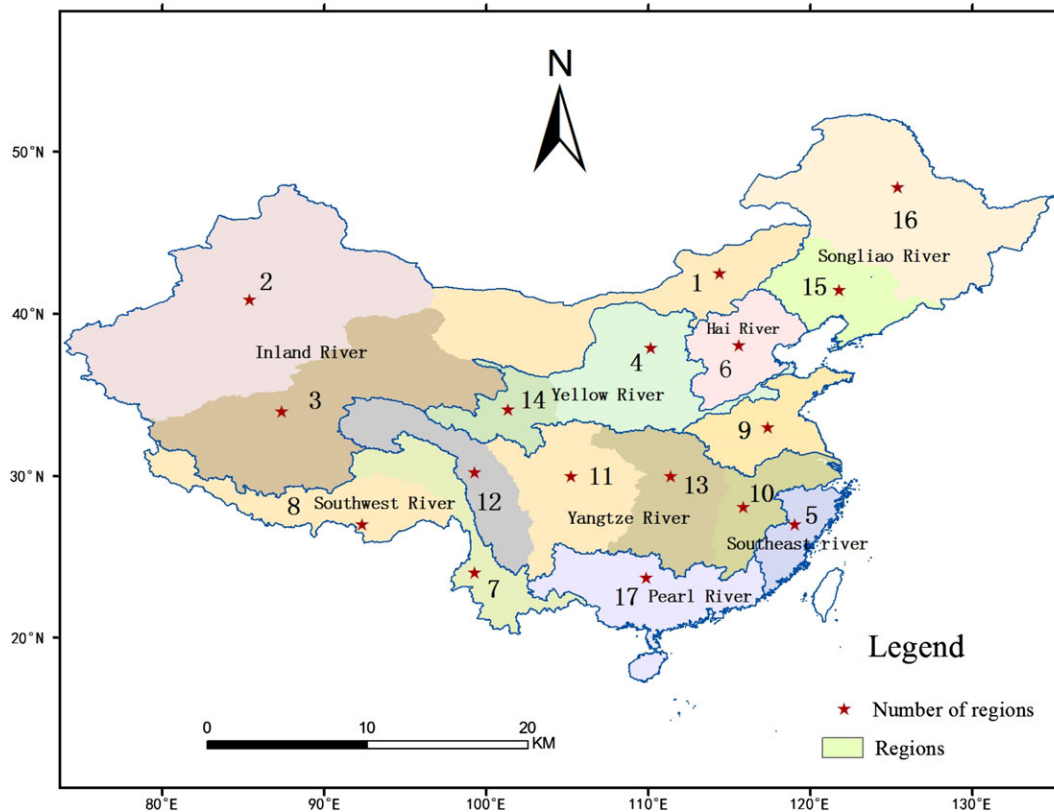


Figure 1. Seventeen hydroclimatic regions in continental China.

Therefore, the primary purpose of this paper is to demonstrate the potential usefulness of the NMME forecasts in a Chinese setting by analyzing precipitation forecasts for 17 hydroclimatic regions in continental China. This paper is organized as follows: Section 2 introduces the 17 hydroclimatic regions in continental China, NMME datasets and observational data used in this study. Section 3 presents the simple model averaging (SMA), Bayesian model averaging (BMA), and verification methods. Section 4 evaluates and discusses the predictive skill of the NMME precipitation forecasts; a summary is provided in Section 5.

## 2. Study area and data

### 2.1. Seventeen hydroclimatic regions in continental China

In this study, based on the standard watershed divisions and climate categories, such as Köppen-Geiger climate types (Peel *et al.*, 2007), continental China is divided into 17 hydroclimatic regions (Figure 1). The list of information about these regions is provided in Table 2, including the area and annual precipitation.

#### 2.1.1. Climate model forecasts and observational data

Two types of data are used in this study: the retrospective NMME precipitation forecasts and gauge-based daily precipitation data. Forecast datasets are available from <http://iridl.ldeo.columbia.edu/SOURCES/Models/>

Table 2. Information for the 17 large hydroclimatic regions (Lang *et al.*, 2014).

ID	Region name	Annual precipitation (mm)	Area (km <sup>2</sup> )
1	Inland rivers in Inner Mongolia	220.4	1 537 520
2	Inland rivers in Xinjiang	168.3	1 104 104
3	Inland rivers in northern Tibet	199.9	694 413
4	Yellow River	391.0	448 864
5	Southeast rivers	1705.9	226 496
6	Hai River	515.9	578 092
7	Southwest rivers in Yunnan	882.2	316 057
8	Southwest rivers in southern Tibet	876.7	908 881
9	Huai River	819.5	415 287
10	Lower Yangtze River	1606.5	324 061
11	Yangtze River	1001.1	323 970
12	Upper Yangtze River	795.2	399 541
13	Middle Yangtze River	1276.5	567 237
14	Upper Yellow River	169.3	504 092
15	Liao River	566.1	310 117
16	Songhua River	535.5	199 721
17	Pearl River	1700.7	567 520

.NMME/. Some pertinent information of the NMME, including the individual model names (as well as abbreviation), predicted period, number of ensembles and model lead times, are shown in Table 1. More detailed descriptions of the NMME model datasets can be found in Kirtman *et al.* (2014). Gauge-based daily precipitation datasets (Xie *et al.*, 2007) over East Asia for the

period 1978–2007 at a horizontal resolution of  $0.5^\circ \times 0.5^\circ$  latitude-longitude grid are used to evaluate the NMME forecasts. The evaluation focuses on comparing spatially averaged NMME predictions and observations of each hydroclimatic region.

### 3. Methodology

#### 3.1. Simple model averaging method

The simplest multi-model mean is determined using simple model averaging (SMA) method, i.e. each model in the multi-model forecasts gets the same weight ( $w_k = 1/K$ , where  $K$  is the number of models), without considering the performance of the model (Miao *et al.*, 2013). In this study, we use the weighted average of single-model ensemble (Katz and Ehrendorfer, 2006) to improve the skill relative to using any individual ensemble member (Atger, 1999).

#### 3.2. Bayesian model averaging method

The Bayesian model Averaging (BMA) method (Duan and Phillips, 2010) is a common solution to the problem of reducing model uncertainty, considering a predicted climatological variable  $y$ , the corresponding observational data  $y_T$ , and  $K$  model forecasts  $\{f_1, f_2, \dots, f_K\}$  of variable  $y$ . In the method of BMA, we assumed that each model forecast,  $f_k$ , is associated with a conditional PDF,  $g(y|f_k)$ , that can be represented by a Gaussian distribution. Under the law of total probability, the BMA predictive model, given observational data  $y_T$ , can be expressed as

$$p(y|f_1, f_2, \dots, f_k) = \sum_{k=1}^K p(y|f_k) * p(f_k|y_T) \quad (1)$$

where,  $p(y|f_k)$  follows that

$$p(y|f_k) = g(y|f_k) \quad (2)$$

and  $p(f_k|y_T)$  is the posterior probability of forecast  $k$  as the best one. We can identify  $p(f_k|y_T)$  as a statistical weight  $w_k$ , whose magnitude reflects an individual model's relative contribution to predictive skill over the validity period. They are non-negative and follow that  $\sum w_k = 1$ . To estimate unknown  $w_k$ , the likelihood function (Equation (3)) is maximized using Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977; McLachlan and Krishnan, 1997).

$$l(w_1, \dots, w_k; c_0; c_1) = \sum_t \log(p(y_t|f_{1t}, \dots, f_{kt})) \quad (3)$$

where,  $\sum_t$  denotes a summation over all period time  $t$ . Subsequently, The BMA can be computed by combining different models with corresponding weights ( $w_k$ ) into a multi-model prediction as Equation (4).

$$\begin{aligned} E[y|f_1, \dots, f_k] &= E[p(y|f_1, \dots, f_k)] \\ &= E\left[\sum_{k=1}^K w_k g(y|f_k)\right] = \sum_{k=1}^K w_k f_k \end{aligned} \quad (4)$$

Recently, the BMA method has been improved (Fraley *et al.*, 2010; Tian *et al.*, 2011.) and applied in many studies (Duan *et al.*, 2007; Sloughter *et al.*, 2007; Duan and Phillips, 2007; Zhao *et al.*, 2011; Sloughter *et al.*, 2013).

#### 3.3. Verification of forecast skill

A relatively skillful model should be able to accurately simulate both the pattern and magnitude of variability (Taylor, 2001). To compare the forecast skill of the NMME forecasts, we utilize two types of statistical measures. One type is the accuracy of the ensemble prediction means, which measures the performance of the ensemble prediction means against the observations. The relative root-mean-squared error and correlation coefficient (RRMSE-R) diagram is used for this purpose. The second type is the skill of the ensemble forecasts, which include rank histograms (RHs) and continuous ranked probability skill scores (CRPSSs).

##### 3.3.1. The RRMSE-R diagram

The statistic most often used to quantify pattern similarity is the correlation coefficient. Here, the correlation coefficient  $R$  between forecasts and observation is defined considering Taylor diagram (Taylor, 2001), but with some modifications as Equation (5).

$$R = \frac{R' - R_{\min}}{R_{\max} - R_{\min}}, r = R_{\min} + r'(R_{\max} - R_{\min}) \quad (5)$$

We can calculate  $\theta$  ( $\cos \theta = R$ ) to plot the RRMSE-R diagram using both  $R$  and  $r$ . Here,  $\theta$  is the angle corresponding to the correlation coefficient,  $R$  and  $R'$  are the correlation coefficients of RRMSE-R and Taylor diagram for calculating  $\theta$ , respectively,  $r$  and  $r'$  denote the correlation coefficients labelled in the RRMSE-R and Taylor diagrams, respectively, and  $R_{\min}$  and  $R_{\max}$  are the minimum and maximum correlation coefficients in the RRMSE-R diagram, respectively.

A statistic often used to quantify differences in two fields is the RRMSE, which is defined as the ratio of RMSE to the average of the observation. With the two measures, it is possible to construct a diagram that statistically quantifies the performance of forecasts. A point is plotted on a polar style graph determined by correlation coefficient and RRMSE. The radial distances from the origin to the points are the RRMSE, and the azimuthal positions ( $\cos \theta$ ) give the correlation coefficient. A point with small  $\text{RRMSE} * \sin \theta$  corresponds to a good forecast.

##### 3.3.2. Efficiency criteria of ensemble forecast (RH and CRPSS)

The rank histogram (RH) (Hamill, 2001) is a tool to evaluate the spread of an ensemble, which is useful for assessing the reliability of ensemble forecasts and ensemble dispersion. The underlying assumption is that the ensemble forecasts are distributed with easily delineated ranges or 'bins' of the predicted variable such that the probability of occurrence of the observation within each bin is constant.



Rank histograms are generated by repeatedly tallying the rank of the verification (usually an observation) relative to values from an ensemble that has been sorted from the lowest to the highest (Hamill, 2001). The shape of the rank histogram typically implies ensemble qualities in the forecasts. For example, a flat rank histogram is typically indicative of reliability, a U-shaped one denotes a lack of spread and the observations fall outside the ensemble, a Dome-shaped one indicates too much spread and the observations fall within the ensemble centre. Bias (wet/dry) excessively populates the (left/right) of a sloped histogram. A skill score for the rank histogram is the mean error ( $Err$ ), which is calculated as the mean of the differences between the model and perfect histograms. The  $Err$  is defined as

$$Err = \sqrt{\frac{1}{N+1} \sum_{i=1}^{N+1} \left[ R_i - \frac{1}{N+1} \right]^2}, \quad (6)$$

where  $R_i$  is value of  $i$ th rank in rank figure,  $N$  is the member's number. To make a fair comparison between models with different ensemble size, the  $Err$  is normalized by the value for a perfect 'flat' rank histogram as Equation (7):

$$FracErr = \frac{Err}{1/(N+1)} = Err^* (N+1), \quad (7)$$

where  $FracErr$  is the mean fractional error, which is not biased by the number of bins.

The continuous ranked probability score (CRPS) applies to probabilistic forecasts that exhibit predictive cumulative distribution functions. The CRPS (Hersbach, 2000) is the integrated squared error between the cumulative distribution function (CDF) of the forecasts and the CDF of the observations. The CRPS is defined as

$$CRPS = \int_{-\infty}^{+\infty} [P_F(x) - P_O(x)]^2 dx, \quad (8)$$

where  $P_F$  and  $P_O$  are the CDFs of the forecasts and observations, respectively.  $x$  denotes the events to be analyzed. Note also that a small CRPS indicates a good forecast. The continuous rank probability skill score (CRPSS) is a skill score based on the CRPS with a normal skill score format:

$$CRPSS = 1 - \frac{CRPS}{CRPS^*} \quad (9)$$

where  $CRPS^*$  is the CRPS for a reference system, which is a climatological forecast of the predictand. Because the CRPS measures the average square error in probability, values approaching zero are preferred. It follows that a CRPSS closer to 1 is preferred because this value indicates a low CRPS of the forecasting system relative to the CRPS of the reference system.

#### 4. Results and discussion

The purpose of this study is to assess the predictive skill of the NMME precipitation forecasts over 17 hydroclimatic regions in continental China. Both the mean and spread of the ensemble forecasts are verified compared with the observations.

##### 4.1. Evaluation of ensemble mean precipitation

In this section, the NMME ensemble mean precipitation is evaluated over 17 hydroclimatic regions in continental China. The performance measures include the Pearson correlation coefficient ( $R$ ), and relative root-mean-square error (RRMSE). We calculated the two performance indices to compare the monthly observation with the ensemble mean of each model in NMME.

The correlation coefficients for the 17 hydroclimatic regions based on the CMC1-CanCM3 forecasts are provided in Figure 2.

The correlation coefficients between the forecasts and the observations are computed for each month of an individual year (from January to December) (vertical axis in Figure 2) and for different lead months (horizontal axis in Figure 2) during the period 1981–2009. The sample size (29 years) is small to be more precise in its correlation coefficient calculations. Therefore, the target months are defined for spanning 3 months, where the months before and after the target month within the same lead are added. For example, target month 1 at lead month 1 is December at lead month 1, January at lead month 1 and February at lead month 1; target month 2 at lead month 1 is January at lead month 1, February at lead month 1 and March at lead month 1; and so on. The approach expands the sample size to remove random sampling noise from extreme events and to better estimate the correlation coefficients (Schaake *et al.*, 2007).

In all regions, the correlation coefficients are high in autumn (September to November) and spring (April to June), while the correlation coefficients are low in summer (July to August) and winter (December to March). Therefore, CMC1-CanCM3 has a high predictive skill in autumn and spring over the 17 hydroclimatic regions. The predictive skill is low in summer and winter, indicating that some challenges in predicting precipitation in summer and winter are present over these regions. This phenomenon may be attributed to the presence of the strong weather noise variability in summer, including strong blocking events in the extratropics and the intra-seasonal oscillation events (Madden–Julian Oscillation, MJO) in the tropics (Zheng *et al.*, 2000). Madden and Shea (1978) suggested that there will be better potential predictability when small climatic noise variance exists. As shown in Zhao *et al.* (2008), climatic noise varies with pronounced seasonal variations. The largest variations typically occur in summer, and the seasonal variations are larger in inland areas compared with coastal regions. These appear consistent with Zheng *et al.* (2004), who reported that the potentially predictable variability is low in summer. IPCC AR5 (Kirtman *et al.*, 2013) also showed that the effect of the external forcing on predictability increases with time. Therefore, the potential predictability of precipitation in China exhibits large seasonal differences, with better predictability in spring and autumn and reduced predictability in summer over most regions. These findings may be attributable to the effects of the climatic noise variations and atmospheric circulation.

The correlation coefficients are high over most regions, but are relatively low over region 2 (Inland rivers in

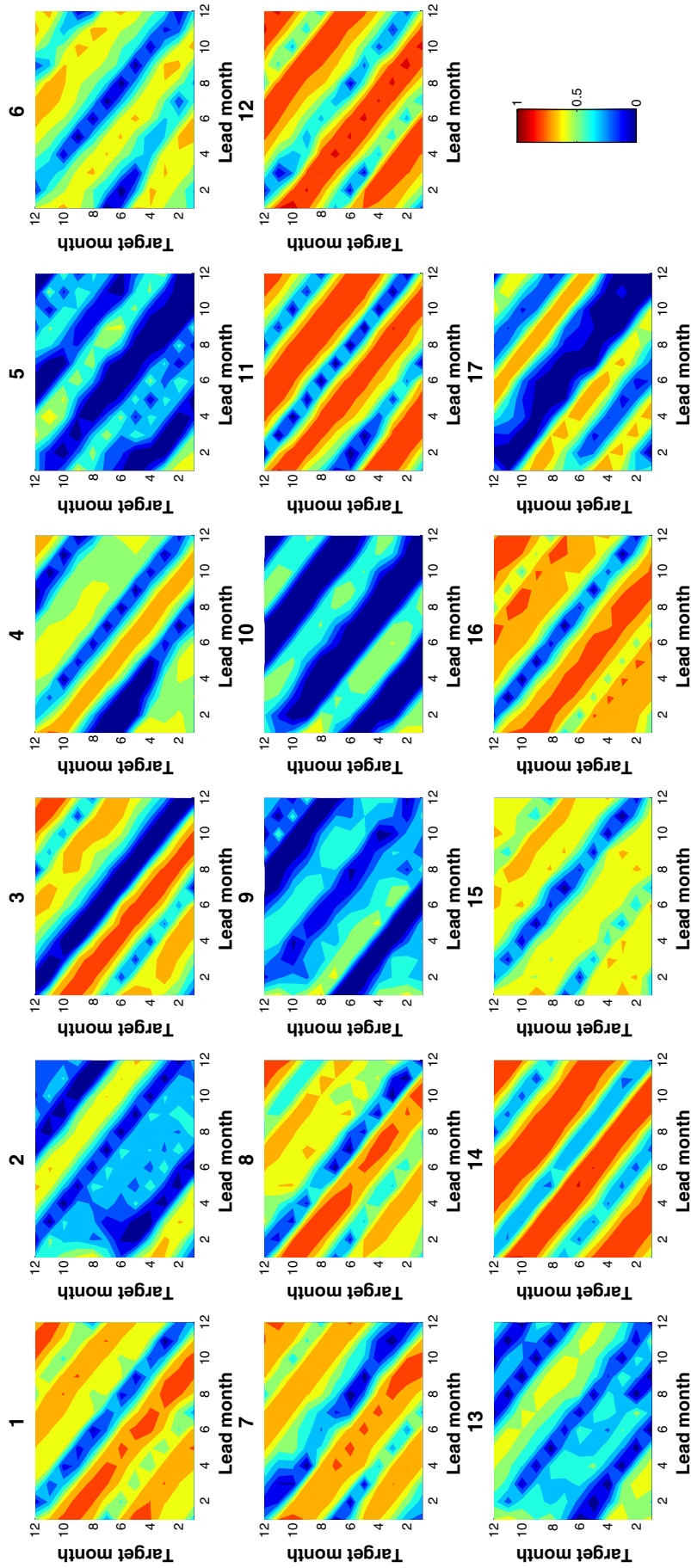


Figure 2. Correlation coefficients between precipitation forecasts and observations for CMC1-CanCM3 over the 17 hydroclimatic regions in continental China.

Xinjiang), 5 (Southeast rivers), 9 (Huai River), 10 (Lower Yangtze River), and 13 (Middle Yangtze River). It can be indicated that CMC1-CanCM3 has relatively better predictability over some regions, such as upstream Yangtze River, Yellow Rivers, and Southwest Rivers. The upstream Yangtze River and Yellow Rivers are located in the Tibetan Plateau climate zones, which are located in the middle troposphere (Shao and Zhang, 1998). However, limited precipitation predictability appears in the Huai River basin, the downstream Yangtze Rivers, and southeast rivers. In particular, the variability of precipitation in these regions has been recognized as one of the prominent features associated with the behaviour of the East Asia summer monsoon (EASM) (Ying *et al.*, 2013). Regional climate anomalies can arise from many sources of external forcings, such as ocean and land anomalies, and internal processes inherent in the atmosphere (Kang *et al.*, 2003). There are two dominant factors influencing the predictable precipitation signal in the Yangtze–Huai River basin. The first is the variations in the Kuroshio-related SST in the northwest Pacific from MAM to JAS, the second is the ENSO-related SST in the eastern tropical Pacific from ASO to DJF (Ying *et al.*, 2013). However, internal processes (e.g. Madden–Julian oscillation), which are known to have a predictability of only a few days, also affect the activity of EASM. Thus, these regions are less predictable due to the large contribution of the internal atmospheric processes to the seasonal mean (Kang *et al.*, 2003). Bei and Zhang (2007) also suggests that mesoscale predictability of heavy precipitation event along Mei–Yu front of China is inherently limited. However, the predictability of South China winter rainfall has not been documented in the literature (Yang *et al.*, 2014).

The close relationship with varying climatic noise may make predictability of China precipitation seasonal and regional dependent. As stated in Liu *et al.* (2000), the climatic noise of monthly precipitation over China has obvious seasonal variation and it is greater in summer than in winter; in most areas, the climatic noise is prominently decreasing from south to north and from coast to inland.

The other models have similar distributions of correlation coefficient; these results are shown in Supporting Information (Figures S1–S10). In addition, the correlation coefficients for SMA and BMA are shown in Supporting Information (Figures S11–S12). The mean values of the correlations averaged over all target and lead months and all 17 regions in Figures S11 and S12 are 0.56 (SMA) and 0.58 (BMA), respectively. The averaging results suggest that SMA performs better than most models, and BMA exhibits the best performance.

In this section, we apply cross validation to calculate the skill scores in RRMSE-R diagram, preventing artificially high in-sample skill scores. The datasets are separated into two periods, called the calibration (Jan, Mar, May, Jul, Sep, Nov 1982–2009) and validation (Feb, Apr, Jun, Aug, Oct, Dec 1982–2009) periods (Wilks, 1995). The cross validation was applied to the BMA, which was verified for the calibration and validation periods. For this approach, the weights in Equation (4) were computed from (Jan,

Mar, May, Jul, Sep, Nov) for (Feb, Apr, Jun, Aug, Oct, Dec) forecasts; and from (Feb, Apr, Jun, Aug, Oct, Dec) for (Jan, Mar, May, Jul, Sep, Nov) forecasts. Then the (Feb, Apr, Jun, Aug, Oct, Dec) forecasts were used for calibration (BMA-cal in Figure 3) and (Jan, Mar, May, Jul, Sep, Nov) forecasts for validation (BMA-val in Figure 3). The predictive skills of 11 models and multi-model forecasts (lead month 1) based on SMA and BMA over the 17 hydroclimatic regions for the entire period are also presented in Figure 3. As shown in Figure 3, the performance of multi-model forecasts based on the BMA for calibration and validation periods are similar.

According to Figure 3, the overall differences between the 11 models are small over the 17 regions. For the forecasts at lead month 1, the CFSv2 model exhibits significantly higher skill than the other 10 models over most regions. Moreover, the CFSv1 model skill exceeds the skill of nine models over 4 of the 17 regions. The GFDL models exhibits middle skills over 14 of the 17 regions. The NASA and NASA-062012 models have similar predictive skills that are analogous to the predictive skill of the CFS over 9 of the 17 regions, but perform poorly over regions 8, 11, 12, and 17. The CMC models show moderate skills over 7 of the 17 regions. IRI models exhibit bad skills over 9 of 17 regions. The CCSM3 exhibits worse skills 11 of the 17 regions, in part because CCSM3 model has only ocean initialization (Becker *et al.*, 2014). Therefore, the CFS model, especially CFSv2, outperforms the other models, GFDL and NASA have relatively better performance. The CMC models also perform well. The IRI and CCSM3 are worse than the other models over most regions. In addition, the SMA method performs better than most models. This finding indicates that a multi-model average improves the overall predictive ability over all regions. The BMA method performs the best over 17 regions, indicating that the BMA approach can reduce model uncertainty and significantly improve predictive skill of precipitation. Most correlation coefficients are concentrated between 0.6 and 0.95 over regions except regions 2, 5, 9, and 10. Most models perform better in central China, nevertheless, humid regions and monsoon regions, such as regions 5 (Southeast rivers), 10 (Lower Yangtze River), and 9 (Huai River) tend to have relatively low predictive skills for precipitation. It can be indicated that these models, whose correlation coefficients exceed 0.6, produce reliable seasonal precipitation forecasts for most regions in continental China. These results are at least partially in accordance with the aforementioned analysis.

#### 4.2. Verification of ensemble forecasts

The above results are focused on deterministic and ensemble mean forecasts. In the following discussion, the ensemble spread and probabilistic predictive skill of the NMME are evaluated using rank histograms (RHs) and continuous ranked probability skill scores (CRPSSs) at lead month 1.

Almost all of model forecasts show sloped (left) rank histograms except region 17 (Pearl River), indicating wet bias in the ensembles (not shown). It can be concluded

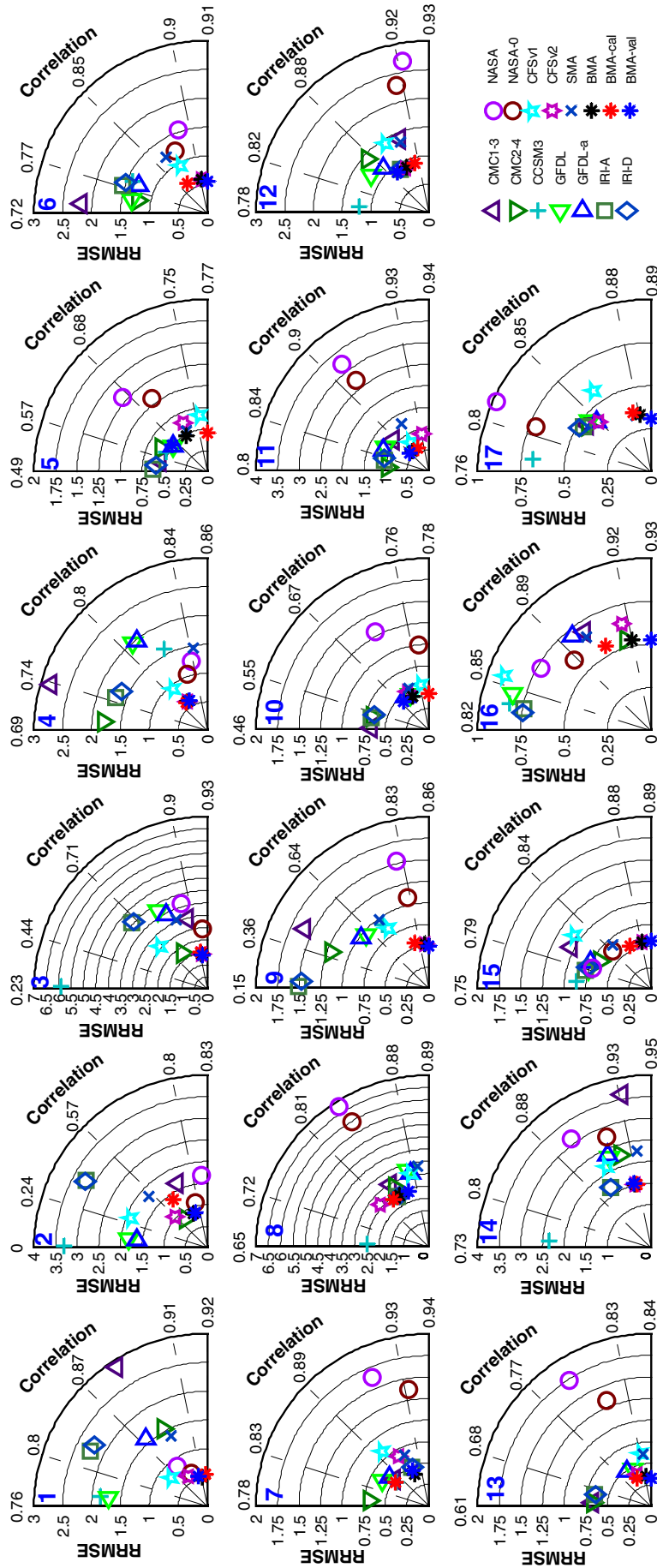


Figure 3. Comparison based on the RRMSE-R diagrams among the 11 models and the multi-model forecasts based on SMA and BMA from the NMME. The 11 models and the SMA and BMA are for the entire period. The BMA-cal and BMA-val are for the calibration (Jan, Mar, May, Jul, Sep, Nov 1982–2009) and validation (Feb, Apr, Jun, Aug, Oct, Dec 1982–2009) period, respectively. The radial distances from the origin to the points are the RRMSE, and the azimuthal positions give the correlation coefficient. The point with smaller RRMSE\* $\sin\theta$  corresponds to better forecast.



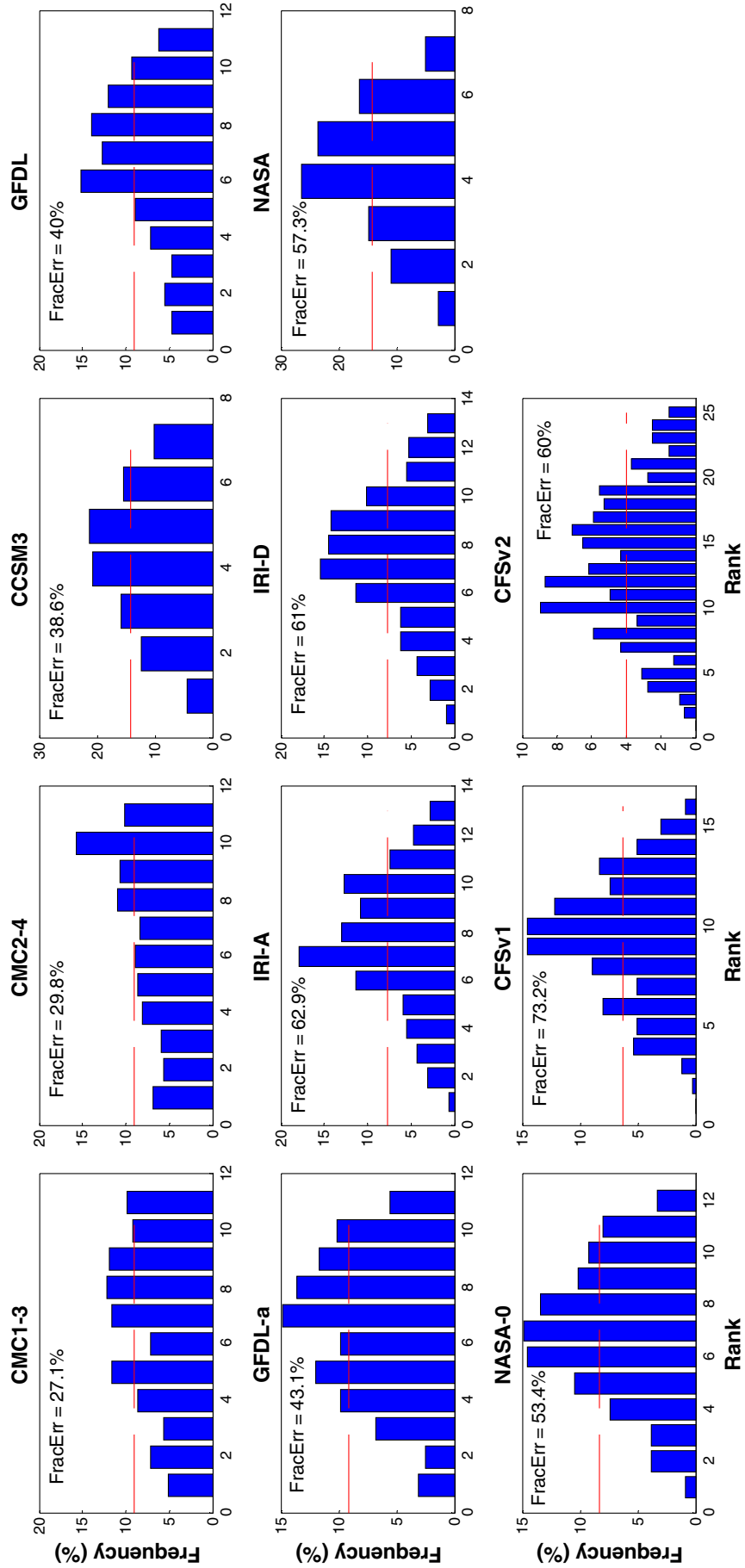


Figure 4. Rank histograms of the ensemble forecast anomalies from the 11 NMME models at lead month 1 over region 1 (Hai River). The horizontal line represents a perfect 'flat' rank histogram.

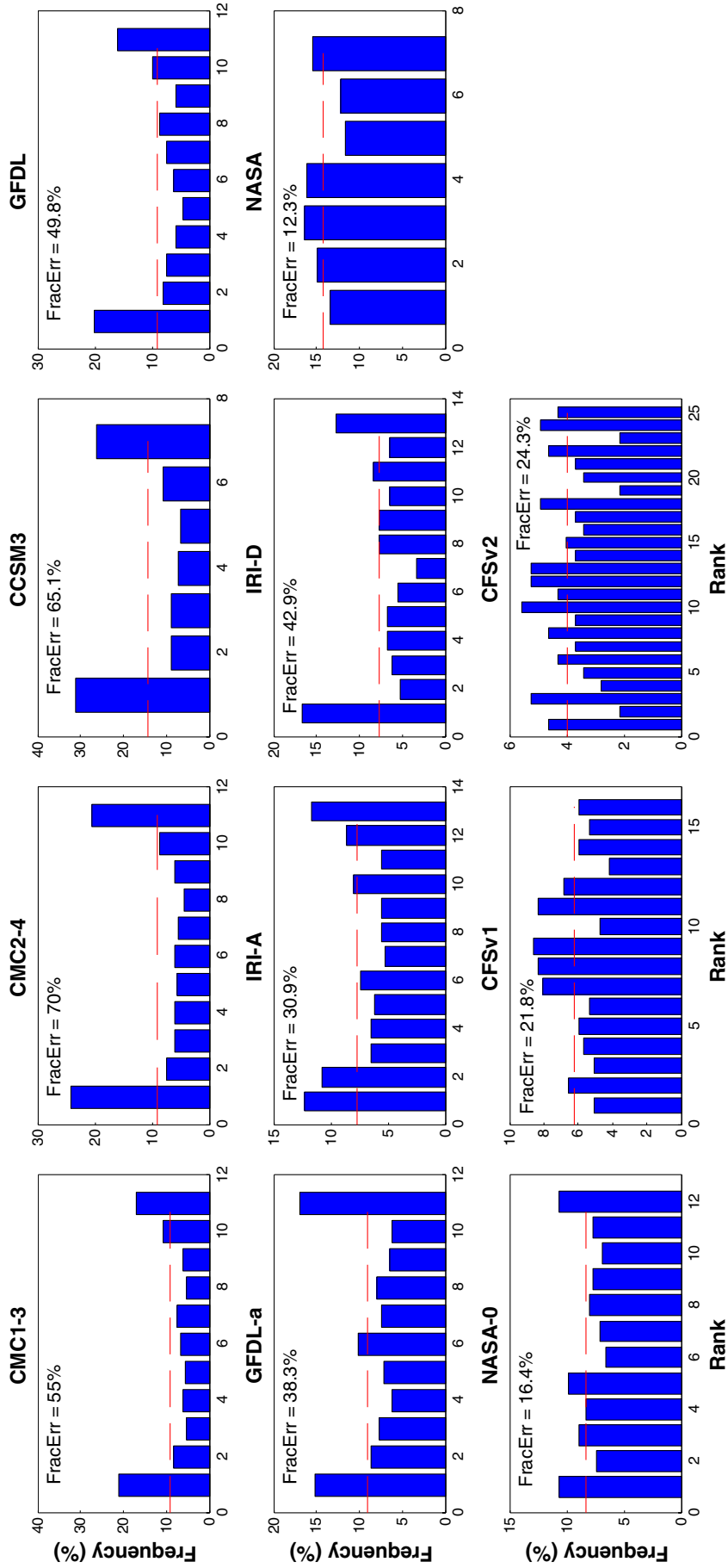


Figure 5. Rank histograms of the ensemble forecast anomalies from the 11 NMME models at lead month 1 over region 1 (Pearl River). The horizontal line represents a perfect 'flat' rank histogram.

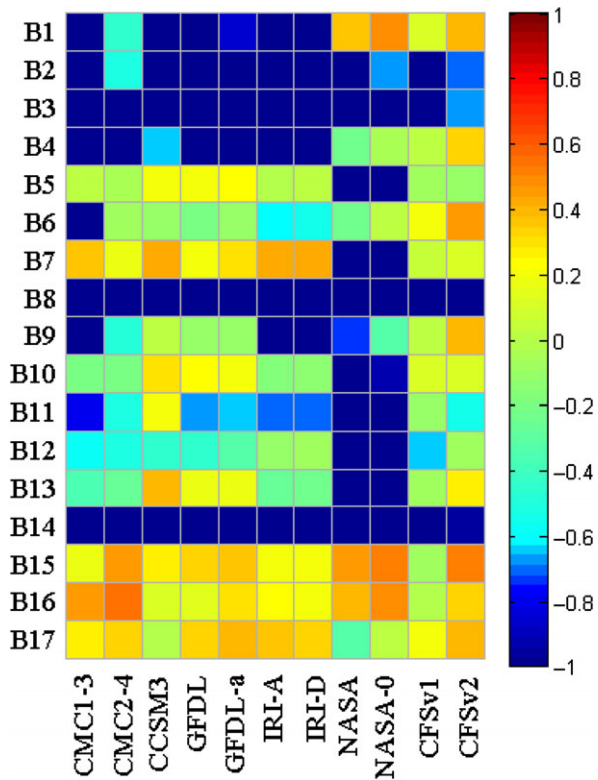


Figure 6. Continuous ranked probability skill scores (CRPSS) for the ensemble forecasts from the 11 NMME models at lead month 1 over the 17 regions in continental China.

that models forecasts typically overestimate the precipitation amount over most regions. In order to tell more information about the forecast ensemble distribution, the rank histograms are constructed using the forecast and observed anomalies. Figure 4 shows the rank histograms of the NMME ensemble forecast anomalies over region 6 (Hai River basin). As shown in Figure 4, most models (except for the CMC models) show Dome-shaped rank histogram, indicating too much spread in these models ensembles over Hai River. Over region 17 (presented in Figure 5), some models, such as CMC, CCSM3, GFDL, and IRI, show U-shaped rank histogram, indicating lack of spread in these ensembles. The NASA and CFS models show flatter rank histogram, indicating better forecasts than other models. These results may be helpful to improve models ensembles and thus forecast skill.

Figure 6 shows the continuous ranked probability skill scores (CRPSSs) of the ensemble precipitation forecasts from the 11 NMME models over the 17 regions. The skill scores are less than 0.6 for the entire models and regions. The skill scores of CFSv2 are positive over most regions, indicating that CFSv2 produces reliable forecasts over most regions. In the case of 17 regions, CFSv2 produces higher predictive skill. IRI and CCSM3 perform worse than most models. Also, CRPSSs over regions 15, 16, 17 are positive, showing reasonable forecasts over these regions. The CRPSSs of most models over regions 2, 3, 4, 8, 9, 14 are less than  $-0.6$ , demonstrating unreliable ensemble forecasts over these regions.

## 5. Conclusions

In this paper, we compare the NMME seasonal precipitation forecasts. Specifically, the predictive skills of 11 models and the SMA and BMA methods are assessed over 17 hydroclimatic regions in continental China. Various methods were used in this study to analyze the data, including correlation coefficient, RRMSE (indicated by the RRMSE-R diagram), rank histograms, and CRPSSs. The results can be summarized as follows.

- 1 From the correlation coefficient between the forecasts and observation, the performance is seasonally dependent. The forecast reliability is higher in autumn and spring compared with summer, which may be primarily related to seasonal differences in ocean-land-atmosphere interactions and the predictability of atmospheric circulation.
- 2 The 11 models perform differently over different regions. The correlation coefficient of regions 1 (Inland rivers in Inner Mongolia), 7, 8 (Southwest rivers), 11 (Yangtze River), 12 (Upper Yangtze River), 14 (Upper Yellow River), 16 (Songhua River) are larger than those of the other regions, especially regions 2 (Inland rivers in Xinjiang), 5 (Southeast rivers), 9 (Huai River), and 10 (Lower Yangtze River). The Huai River, the downstream Yangtze and southeast rivers (i.e. south-eastern monsoon regions) have worse predictive skill. In contrast, the northeast, southwest and upstream Yangtze and Yellow Rivers have better predictive skill. Most models show wet bias in forecasting monthly precipitation over all regions except region 17 (Pearl River). Despite these biases, models show much spread or lack of spread over different regions. Some models (e.g. NASA and CFS) also show reliable ensemble forecasts over some regions (such Pearl River). Predictability of precipitation in one region corresponds to the contribution of local internal atmospheric processes and external boundary condition.
- 3 Most measurements demonstrate that most models tend to be potentially skillful for use as inputs to hydrological models. In general, the CFS model, especially CFSv2, outperforms the other models according to multiple verification measurements, such as the correction coefficient, RRMSE, rank histogram, and CRPSS. The GFDL and NASA models also exhibit relatively high quality results over most regions. The CMC models have lower reliability than the aforementioned models. The IRI and CCSM3 models trail behind the other models. SMA improves the predictive skill. The improvement is more obvious for BMA, which combines the advantages of the individual models.

In brief, this preliminary comparison assists in further recognizing the forecast skill of the 11 models from the NMME over 17 hydroclimatic regions in continental China. To use the NMME dataset more extensively, the above results will be a very important reference. For example, the results can be used as input to hydrological models to provide insights into drought monitoring

in China. With developments over the past decade, great progress has been made. However, the NMME system continues to have considerable imperfections in China. To improve the usefulness of the NMME dataset in China, much work is needed, such as the re-calibration of model outputs to the given observed values.

### Acknowledgements

This study was supported by the Natural Science Foundation of China (No. 41475093), National Science and Technology Support Plan Program (NO. 2013BAB05B04), and the Fundamental Research Funds for the Central Universities (NO. 2013YB32).

### Supporting Information

The following supporting information is available as part of the online article:

**Figure S1.** Correlation coefficients between precipitation forecasts and observations for CMC2-CanCM4 over the 17 hydroclimatic regions in continental China.

**Figure S2.** Correlation coefficients between precipitation forecasts and observations for CCSM3 over the 17 hydroclimatic regions in continental China.

**Figure S3.** Correlation coefficients between precipitation forecasts and observations for GFDL over the 17 hydroclimatic regions in continental China.

**Figure S4.** Correlation coefficients between precipitation forecasts and observations for GFDL-aer04 over the 17 hydroclimatic regions in continental China.

**Figure S5.** Correlation coefficients between precipitation forecasts and observations for IRI-A over the 17 hydroclimatic regions in continental China.

**Figure S6.** Correlation coefficients between precipitation forecasts and observations for IRI-D over the 17 hydroclimatic regions in continental China.

**Figure S7.** Correlation coefficients between precipitation forecasts and observations for NASA over the 17 hydroclimatic regions in continental China.

**Figure S8.** Correlation coefficients between precipitation forecasts and observations for NASA-062012 over the 17 hydroclimatic regions in continental China.

**Figure S9.** Correlation coefficients between precipitation forecasts and observations for CFSv1 over the 17 hydroclimatic regions in continental China.

**Figure S10.** Correlation coefficients between precipitation forecasts and observations for CFSv2 over the 17 hydroclimatic regions in continental China.

**Figure S11.** Correlation coefficients between precipitation forecasts and observations for SMA over the 17 hydroclimatic regions in continental China.

**Figure S12.** Correlation coefficients between precipitation forecasts and observations for BMA over the 17 hydroclimatic regions in continental China.

### References

Atger F. 1999. The skill of ensemble prediction systems. *Mon. Weather Rev.* **127**: 1941–1953.

- Becker E, van den Dool H, Zhang Q. 2014. Predictability and forecast skill in NMME. *J. Clim.* **27**: 5891–5906.
- Bei N, Zhang F. 2007. Impacts of initial condition errors on mesoscale predictability of heavy precipitation along the Mei-Yu front of China. *Q. J. R. Meteorol. Soc.* **133**: 83–99.
- Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, McKenna DS, Santer BD, Smith RD. 2006. The community climate system model version 3 (CCSM3). *J. Clim.* **19**: 2122–2143.
- Dempster A, Laird N, Rubin D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.* **39**: 1–38.
- DeWitt DG. 2005. Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly Coupled Atmosphere–Ocean General Circulation Model. *Mon. Weather Rev.* **133**: 2972, doi: 10.1175/MWR3016.1.
- Doblas-Reyes FJ, Hagedorn R, Palmer TN. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus A* **57**: 234–252.
- Duan Q, Phillips TJ. 2010. Bayesian estimation of local signal and noise in multimodel simulations of climate change. *J. Geophys. Res.* **115**: D18, doi: 10.1029/2009JD013654.
- Duan Q, Ajami NK, Gao X, Sorooshian S. 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* **30**: 1371–1386.
- Fraley C, Rafery AE, Gneiting T. 2010. Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Weather Rev.* **138**: 190–202.
- Hagedorn R, Doblas-Reyes FJ, Palmer TN. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus A* **57**: 219–233.
- Hamill TM. 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.* **129**: 550–560.
- He Y, Wetterhall F, Bao H, Cloke H, Li Z, Pappenberger F, Hu Y, Manful D, Huang Y. 2010. Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: a case study. *Atmos. Sci. Lett.* **11**: 132–138.
- Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast.* **15**: 559–570.
- Holdaway D, Errico R, Gelaro R, Kim JG. 2014. Inclusion of linearized moist physics in NASA's Goddard Earth Observing System Data Assimilation Tools. *Mon. Weather Rev.* **142**: 414–433.
- Infanti JM, Kirtman BP. 2014. Southeastern U.S. rainfall prediction in the North American Multi-Model Ensemble. *J. Hydrometeorol.* **15**: 529–550.
- Kang IS, Lee JY, Park CK. 2003. Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *J. Clim.* **17**: 834–844.
- Katz R, Ehrendorfer M. 2006. Bayesian approach to decision making using ensemble weather forecasts. *Weather Forecast.* **21**: 220–231.
- Kirtman BP, Min D. 2009. Multimodel ensemble ENSO prediction with CCSM and CFS. *Mon. Weather Rev.* **137**: 2908–2930.
- Kirtman B, Power SB, Adedoyin JA, Boer GJ, Bojariu R, Camilloni I, Doblas-Reyes FJ, Fiore AM, Kimoto M, Meehl GA, Prather M, Sarr A, Schär C, Sutton R, van Oldenborgh GJ, Vecchi G, Wang HJ. 2013. Climate change 2013: the physical science basis. In *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, UK and New York, NY, 953–1028.
- Kirtman BP, Min D, Infanti JM, Kinter JL, Paolino DA, Zhang Q, van den Dool H, Saha S, Mendez MP, Becker E, Peng P, Tripp P, Huang J, DeWitt DG, Tippet MK, Barnston AG, Li S, Rosati A, Schubert SD, Rienecker M, Suarez M, Li ZE, Marshak J, Lim YK, Tribbia J, Pegion K, Merryfield WJ, Denis B, Wood EF. 2014. The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intra-seasonal prediction. *Bull. Am. Meteorol. Soc.* **95**: 585–601.
- Korecha D, Barnston AG. 2007. Predictability of June–September rainfall in Ethiopia. *Mon. Weather Rev.* **135**: 628–650.
- Lang Y, Ye A, Gong W, Miao C, Di Z, Xu J, Liu Y, Luo L, Duan Q. 2014. Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China. *J. Hydrometeorol.* **15**: 1546–1559.
- Lavers D, Luo L, Wood EF. 2009. A multiple model assessment of seasonal climate forecast skill for applications. *Geophys. Res. Lett.* **36**: L23711, doi: 10.1029/2009GL041365.



- Lewis JM. 2005. Roots of ensemble forecasting. *Mon. Weather Rev.* **133**: 1865–1885.
- Liu Y, Ma K, Lin Z. 2000. Potential predictability of monthly precipitation over China. *J. Meteor. Res.* **14**: 316–329.
- Liu Y, Duan Q, Zhao L, Ye A, Tao Y, Miao C, Mu X, Schaake JC. 2013. Evaluating the predictive skill of post-processed NCEP GFS ensemble precipitation forecasts in China's Huai river basin. *Hydrol. Processes* **27**: 57–74.
- Lorenzo MN, Taboada JJ, Iglesias I, Gómez-Gesteira M. 2011. Predictability of the spring rainfall in Northwestern Iberian Peninsula from sea surfaces temperature of ENSO areas. *Clim. Change* **107**: 329–341.
- Luo L, Wood EF. 2008. Use of Bayesian merging techniques in a multi model seasonal hydrologic ensemble prediction system for the eastern United States. *J. Hydrometeorol.* **9**: 866–884.
- Madden RA, Shea DJ. 1978. Estimates of the natural variability of time-averaged temperatures over the United States. *Mon. Weather Rev.* **106**: 1695–1703.
- McLachlan G, Krishnan T. 1997. *The EM Algorithm and Extensions*. New York, NY: John Wiley.
- Merryfield WJ, Lee WS, Boer GJ, Kharin VV, Scinocca JF, Flato GM, Ajayamohan RS, Fyfe JC, Tang Y, Polavarapu S. 2013. The Canadian seasonal to interannual prediction system. Part I: models and initialization. *Mon. Weather Rev.* **141**: 2910–2945.
- Miao C, Duan Q, Sun Q, Li J. 2013. Evaluation and application of Bayesian multi-model estimation in temperature simulations. *Prog. Phys. Geogr.* **37**: 727–744.
- Misra V, Li H. 2014. The seasonal climate predictability of the Atlantic Warm Pool and its teleconnections. *Geophys. Res. Lett.* **41**: 661–666.
- Nicholson SK. 2014. The predictability of rainfall over the Greater Horn of Africa. Part I. Prediction of seasonal rainfall. *J. Hydrometeorol.* **15**: 1011–1027.
- Palmer TN, Doblus-Reyes FJ, Hagedorn R, Alessandri A, Gualdi S, Andersen U, Feddersen H, Cantelaube P, Terres JM, Davey M, Graham R, Décléuse P, Lazar A, Déqué M, Guérémy JF, Díez E, Orfila B, Hoshen M, Morse AP, Keenlyside N, Latif M, Maisonnave E, Rogel P, Marletto V, Thomson MC. 2004. Development of a European multi-model ensemble system for seasonal-to-interannual prediction (Demeter). *Bull. Am. Meteorol. Soc.* **85**: 853–872.
- Palmer TN, Doblus-Reyes FJ, Weisheimer A, Rodwell MJ. 2008. Toward seamless prediction: calibration of climate change projections using seasonal forecast. *Bull. Am. Meteorol. Soc.* **89**: 459–470.
- Pappenberger F, Thielen J, Medico MD. 2011. The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European flood alert system. *Hydrol. Processes* **25**: 1091–1113.
- Park YY, Buizza R, Leutbecher M. 2008. Tigge: preliminary results on comparing and combining ensembles. *Q. J. R. Meteorol. Soc.* **134**: 2029–2050.
- Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci. Discuss.* **11**: 1633–1644.
- Peng P, Kumar A, van den Dool H, Barnston AG. 2002. An analysis of multimodel ensemble predictions for seasonal climate anomalies. *J. Geophys. Res.* **107**: 18-1–18-2, doi: 10.1029/2002JD002712.
- Phillips TJ, Gleckler PJ. 2006. Evaluation of continental precipitation in 20th century climate simulations: the utility of multimodel statistics. *Water Resour. Res.* **42**: W03202, doi: 10.1029/2005WR004313.
- Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van den Dool HM, Pan HL, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP climate forecast system. *J. Clim.* **19**: 3483–3517.
- Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou Y-T, H-Y C, Iredell M, Ek M, Meng J, Yang R, Mendez MP, van den Dool H, Zhang Q, Wang W, Chen M, Becker E. 2014. The NCEP climate forecast system version 2. *J. Clim.* **27**: 2185–2208.
- Schaake J, Demargne J, Hartman R, Mullusky M, Welles E, Wu L, Herr H, Fan X, Seo DJ. 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrol. Earth Syst. Sci. Discuss.* **4**: 655–717.
- Shao Y, Zhang H. 1998. Climate characteristics in the source regions of the Yangtze River and Yellow River. *J. Qinghai Environ.* **2**: 68–72 (in Chinese).
- Sloughter JML, Raftery AE, Gneiting T, Fraley C. 2007. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Weather Rev.* **135**: 3209–3220.
- Sloughter JM, Gneiting T, Raftery AE. 2013. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Weather Rev.* **141**: 2107–2119.
- Sun Q, Miao C, Duan Q, Kong D, Ye A, Di Z, Gong W. 2014. Would the 'real' observed dataset stand up? A critical examination of eight observed gridded climate datasets for China. *Environ. Res. Lett.* **9**: 015001, doi: 10.1088/1748-9326/9/1/015001.
- Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**: 7183–7192, doi: 10.1029/2000JD900719.
- Tibaldi C, Knutti R. 2007. The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc. A* **365**: 2053–2075.
- Tian X, Xie Z, Wang A, Yang X. 2011. A new method for solving the Bayesian model average. *Sci China Press* **41**: 1679–1687 (in Chinese).
- Unganai LS, Mason SJ. 2002. Long-range predictability of Zimbabwe summer rainfall. *Int. J. Climatol.* **22**: 1091–1103.
- Verbist K, Robertson AW, Cornelis WM, Gabriels D. 2010. Seasonal predictability of daily rainfall characteristics in central Northern Chile for dry-land management. *J. Appl. Meteorol. Climatol.* **49**: 1938–1955.
- Wang B, Lee JY, Kang IS, Shukla SJ, Kug JS, Kumar A, Schemm J, Luo JJ, Yamagata T, Park CK. 2007. How accurately do coupled climate models predict the leading modes of Asian-Australian monsoon interannual variability? *Clim. Dyn.* **30**: 605–619.
- Weisheimer A, Doblus-Reyes FJ, Palmer TN, Alessandri A, Arribas A, Déqué M, Keenlyside N, Mac Vean M, Navarra A, Rogel P. 2009. ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions – skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett.* **36**: L21711, doi: 10.1029/2009GL040896.
- Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press: San Diego, CA, 467 pp.
- Xie P, Chen M, Yang S, Yatagai A, Hayasaka T, Fukushima Y, Liu C. 2007. A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeorol.* **8**: 607–626.
- Xu G, Zhao S, Wang Y, Yang Y. 2007. Experiment of ensemble forecast of heavy rainfall in the Huaihe River during rainy season of 2003 (in Chinese). *Climatic Environ. Res.* **12**: 481–488.
- Yang S, Li C, Lu R. 2014. Predictability of winter rainfall in South China as demonstrated by the coupled models of ensembles. *Adv. Atmos. Sci.* **31**: 779–786.
- Ying K, Zheng X, Quan X, Frederiksen CS. 2013. Predictable signals of seasonal precipitation in the Yangtze–Huaihe River Valley. *Int. J. Climatol.* **33**: 3002–3015.
- Yuan X, Wood EF. 2013. Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.* **40**: 4900–4905.
- Yuan X, Wood EF, Luo L, Pan M. 2011. A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys. Res. Lett.* **38**: L13402, doi: 10.1029/2011GL047792.
- Zhang S, Harrison MJ, Rosati A, Wittenberg A. 2007. System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Weather Rev.* **135**: 3541–3564.
- Zhang Q, Dool H, Saha S, Mendez M, Becker E, Peng P, Huang J. 2011. Preliminary evaluation of multi-model ensemble system for monthly and seasonal prediction. In *the 36th NOAA Annual Climate Diagnostics and Prediction Workshop*. NOAA: Silver Spring, MD, 124–131.
- Zhao X, Wu H, He L. 2008. Estimate of climatic noise and potential predictability of seasonal precipitation in China. *J. Nanjing Inst. Meteor.* **31**: 819–827.
- Zhao L, Liang L, Cheng C, Zhang Y, Liu Y. 2011. Bias correction for ensemble predictions of precipitation based on Bayesian model averaging. In *28th Meteorological Conference on S3 Weather Prediction*, Meteorological Society China, Xiamen, 14 pp.
- Zheng X, Nakamura H, Renwick JA. 2000. Potential predictability of seasonal means based on monthly time series of meteorological variables. *J. Clim.* **13**: 2591–2604.
- Zheng X, Sugi M, Frederiksen CS. 2004. Interannual variability and predictability in an ensemble of climate simulations with the MRI-JMA AGCM. *J. Meteor. Soc. Jpn. Ser. II* **82**: 1–18.