Probabilistic Seasonal Prediction of Summer Rainfall over East China Based on Multi-Model Ensemble Schemes*

LI Fang† (李 芳)

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

(Received March 30, 2010; in final form March 22, 2011)

ABSTRACT

The skill of probability density function (PDF) prediction of summer rainfall over East China using optimal ensemble schemes is evaluated based on the precipitation data from five coupled atmosphere-ocean general circulation models that participate in the ENSEMBLES project. The optimal ensemble scheme in each region is the scheme with the highest skill among the four commonly-used ones: the equally-weighted ensemble (EE), EE for calibrated model-simulations (Cali-EE), the ensemble scheme based on multiple linear regression analysis (MLR), and the Bayesian ensemble scheme (Bayes). The results show that the optimal ensemble scheme is the Bayes in the southern part of East China; the Cali-EE in the Yangtze River valley, the Yangtze-Huaihe River basin, and the central part of northern China; and the MLR in the eastern part of northern China. Their PDF predictions are well calibrated, and are sharper than or have approximately equal interval-width to the climatology prediction. In all regions, these optimal ensemble schemes outperform the climatology prediction, indicating that current commonly-used multi-model ensemble schemes are able to produce skillful PDF prediction of summer rainfall over East China, even though more information for other model variables is not derived.

Key words: multi-model ensemble, uncertainty, probability density function, seasonal prediction, rainfall

Citation: Li Fang, 2011: Probabilistic seasonal prediction of summer rainfall over East China based on multi-model ensemble schemes. *Acta Meteor. Sinica*, **25**(3), 283–292, doi: 10.1007/s13351-011-0304-4.

1. Introduction

The dynamic climate model is an important tool for seasonal prediction. The "two-tier" and "one-tier" approaches (Wang et al., 2009) are commonly used in the dynamical seasonal prediction. For the former, SST is first predicted by a coupled model, and then the atmospheric anomalies are predicted using the atmospheric model forced by the predicted SST. For the latter, the coupled atmosphere-ocean general circulation model (CGCM) is employed, and the prediction is essentially an initial value problem (Palmer et al., 2004). Though there are large errors in current CGCMs, many studies have demonstrated that the CGCMs are the most promising tools for seasonal prediction of monsoon precipitation due to their capability in reproducing the monsoon-ocean interaction (Zeng et al., 1990; Wu and Kirtman, 2005; Wang et al., 2005, 2009).

Uncertainty is inevitable in the dynamical seasonal prediction, so the prediction is essentially probabilistic and the predicted information should be expressed in manner of a probability density function (PDF) (Palmer et al., 2005; Gneiting, 2008; Doblas-Reyes et al., 2009; Lavers et al., 2009). The uncertainty includes initialization uncertainty and model uncertainty (Huang, 1993; Palmer et al., 2005; Weigel et al., 2009). The former arises from the lack of observations, measurement errors, or inappropriate data-assimilation procedures. The latter arises from errors and simplification in model itself, such as the defects in physical process parameterizations, improper model parameter settings, and imperfect boundary conditions. Different from short-range weather forecast, both the initialization uncertainty and model uncertainty are important and must be taken into account

*Supported by the National Natural Science Foundation of China (40830103).

†Corresponding author: lifang@mail.iap.ac.cn.

(Chinese version to be published)

in seasonal prediction.

The multi-model ensemble, whose product is probabilistic, is an effective approach to reduce and quantitatively estimate the prediction uncertainty (Doblas-Reyes et al., 2009; Alessandri et al., 2011) and to improve the skill of seasonal prediction (World Climate Reaserch Program (WCRP) Strategic Framework 2005–2015; WCRP Position Paper on Seasonal Prediction in 2008; Report of the 12th Session of the JSC/CLIVAR Working Group on Seasonal to Interannual Prediction (WGSIP) in 2009). First, the multi-model ensemble is a statistical post-process of dynamical model outputs, and hence a statistical-and-dynamical combined prediction method. A good multi-model ensemble scheme can reduce the prediction uncertainty by a simpler way than improving the model itself, and provide more accurate reference information for the operational seasonal prediction than raw model data (Palmer et al., 2005). Second, a probabilistic prediction produced by multi-model ensemble can provide more reference information for the operational seasonal prediction than a single-value (deterministic) prediction product (Doblas-Reyes et al., 2009; Lavers et al., 2009). In addition, the multi-model ensemble scheme is based on the multiple initial-condition ensemble datasets from several models, so its reduced and quantitatively estimated uncertainty includes not only the initialization uncertainty but also the model uncertainty.

So far, there are four commonly-used multi-model ensemble schemes. The simplest one is equal-weight ensemble (EE), which assigns equal weights to ensemble members regardless of their relative performance (Kharin and Zwiers, 2002; Kang and Yoo, 2006). However, all dynamical models inevitably have systematic errors in terms of the mean, the annual cycle or the interannual variability, and in some cases, all three of these characteristics (Kirtman et al., 2003). To overcome this problem, the model outputs are calibrated before EE, which is referred to as Cali-EE hereafter (Peng et al., 2002; Ke et al., 2009). In addition, prediction skills of different ensemble members are generally different, so Krishnamurti et al. (1999) proposed the multiple linear regression ensemble scheme (MLR) to assign unequal weights to ensemble members by the multiple linear regression. The above three schemes are mainly used to produce single-value (deterministic) prediction, and the versions for probabilistic prediction have not appeared until recent several years (Gneiting et al., 2005; Weigel et al., 2009). Besides the above three schemes based on the classical statistics, the ensemble scheme based on Bayesian statistics (Bayes) is also commonly used (Coelho et al., 2004; Luo et al., 2007; Li et al., 2009).

East China (east of 105°E) lies in the East Asian monsoon region. The strong interannual variability of East Asian summer monsoon makes floods and droughts occur frequently in the region, causing huge economic loss (Huang et al., 2003; Lau et al., 2004). Up to now, the operational prediction skill of summer rainfall in East China has still been low, though operational seasonal prediction skill has been overall improved for the past decades. Many dynamical climate models have been used to produce single-value (deterministic) reference prediction of summer rainfall over East China (Zeng et al., 1990; Lin et al., 1998; Chen, 2003; Ding et al., 2004; Li et al., 2004; Lang et al., 2004; Li et al., 2005; Wei et al., 2005; Liu and Guo, 2005; Wang et al., 2008). Recently, some Chinese meteorologists have begun to apply or develop the single-value (deterministic) multi-model ensemble schemes (Feng and Fu, 2007; Ke, 2007; Qin, 2007; Ke et al., 2009). However, the single-value (deterministic) reference prediction would lose a mass of information because of large uncertainty in the dynamical seasonal prediction of China summer rainfall (Wang et al., 1997; Wang, 2001). In the present study, PDF seasonal prediction of summer rainfall over East China is performed by the above four commonly-used multi-model ensemble schemes using 46-yr (1960–2005) hindcast rainfall data from the ENSEMBLES project (http://www. ensemble-eu.org). Then, based on the analysis of their PDF prediction product, the optimal scheme with the highest skill is identified and used to assess the current status of probabilistic seasonal prediction of summer rainfall over East China.

2. Data

2.1 Rainfall data

Observed rainfall data at 160 rain gauge stations of China are provided by the National Climate Center, China Meteorological Administration. The model-output rainfall data are from the ENSEMBLES stream

2 dataset (Weisheimer et al., 2009). The ENSEMBLES project is the follow-on project of Development of a European Multi-model Ensemble System for Seasonal to Interannual Prediction (DEMETER). Its stream 2 dataset includes outputs from five CGCMs, and each of them has nine ensemble members with different initial conditions. The five CGCMs are the IFS/HOPE from the European Centre for Medium-Range Weather Forecasts (ECMWF), the HadGEM2 from the UK Met Office (UKMO), the ARPEGE/OPA8.2 from the Météo-France (MF), the ECHAM5/MPI-OM1 from the Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR), and the ECHAM5/OPA8.2 from the Euro-Mediterranean Centre for Climate Change (CMCC-INGV) in Bologna. Compared with models in DEMETER, these models have improved physical parameterizations by including additional components (e.g., sea-ice or land-surface modules) and interannual variability of the greenhouse gas forcing. Improvments are also made in model resolution and in the initialization. For each model, nine ensemble members are formed with different ocean initial conditions that are derived from perturbations of wind stress and SST. The atmospheric and land surface initial conditions are taken from the ECMWF 40-yr reanalysis (ERA-40) dataset. The hindcast data are from 1960 to 2005, and have a horizontal resolution of 2.5° × 2.5°. The four start dates are 1 February, 1 May, 1 August, and 1 November at 0000 GMT per year. The summer (JJA) rainfall data from the hindcast initiated on 1 May are used in the present study.

2.2 Preprocessing of data

To simplify the analysis, East China is divided into five regions by use of the regionalization scheme of Wang and Tu (2002) and Chen et al. (2009). First, the Rotated Empirical Orthogonal Function (REOF) (Wei, 1999; Von Storch and Zwier, 1999) is applied to the standardized rainfall observations at 120 stations to obtain five rotated patterns (which explain 43% of the total variance) with well-separated centers. The number of EOFs in maximum variance rotation is determined by the Rule-of-Thumb (North et al., 1982). Based on the rotated loading values, then, East China is divided into five regions: South China (R1), the Yangtze River valley (R2), the Yangtze-Huaihe River basin (R3), the eastern part of northern China (R4), and the central part of northern China (R5) (Fig. 1). After the regionalization, the averaged station rainfall in each region from 1960 to 2005 is defined as the time series of observed regional summer rainfall. The model rainfall data are firstly bilinearly interpolated to 120 stations over East China in Fig. 1, and then the rainfall data averaged over the stations of each region are regarded as time series of simulated regional summer rainfall from CGCMs.

PDF prediction usually needs to assume that the target variable satisfies a classic probabilistic distribution, e.g., normal distribution. According to the Jarque-Bera test (Bera and Jarque, 1980), all the time series of regional summer rainfall in R2, R3, R4, and R5 satisfy the assumption of normal distribution at the 0.05 significance level. For time series of summer

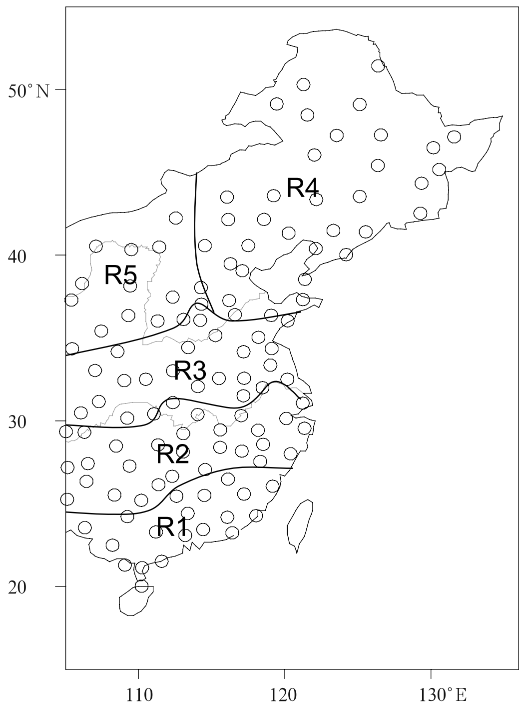

Fig. 1. Geographical distribution of 120 rain gauge stations (circles) and regionalization of East China (thick lines). The five regions are South China (R1), the Yangtze River valley (R2), the Yangtze-Huaihe River basin (R3), the eastern part of northern China (R4), and the central part of northern China (R5).

rainfall in R1, the rainfall data are not normally distributed, and the power transformation is used in the present study to normalize the rainfall data. Results show that the simulated and observed time series of rainfall in R1 normalized by the cubic root transformation can pass the Jarque-Bera test at the 0.05 significance level.

3. Methodology

Due to the short simulation record, in the present study, the buildup and skill evaluation of ensemble prediction models are within a leave-one-out cross-validation framework (Wilks, 1995). In this framework, for a certain target year, the data in other years are used as a training set to estimate the parameters of ensemble prediction models. Prediction skill obtained in this framework is a cross-validated skill.

3.1 Climatology prediction

The climatology prediction is a special PDF prediction, in which the prediction is equal to the PDF fitted by history observations of the target variable. It is the benchmark of the seasonal PDF prediction of rainfall. A skillful prediction must be superior to the climatology prediction. For time series of regional rainfall that are normally distributed or normalized, the predictive PDF by the climatology prediction for each year is normally distributed. Its mean and variance are estimated by climatological mean and variance of observations in the training set.

3.2 EE

According to Tippet et al. (2007) and Weigel et al. (2009), for ensemble members that are normally distributed, the predicted PDF of the EE can be assumed to be normal. At each target year t, the mean and variance of ensemble prediction are estimated by the mean and variance of ensemble members as

$$Y_t \sim N(\frac{1}{m}\sum_{i=1}^{m} X_{it}, \frac{1}{m}\sum_{i=1}^{m}(X_{it} - \frac{1}{m}\sum_{i=1}^{m} X_{it})^2), \quad (1)$$

where Y_t is the predicted rainfall at target year $t = 1, 2, \cdots, n$; X_{it} is the model rainfall for ensemble member i in year t; and m and n are the total numbers of ensemble members and target years, respectively.

3.3 Cali-EE

The difference between the Cali-EE and EE lies in ensemble members. The ensemble members in the former are calibrated model data. In the present study, the bias and variance correction are used to calibrate the model data, and the ensemble members of the Cali-EE is expressed as follows:

$$X1_{it} = \overline{O} + (X_{it} - \overline{X}_i)(S_O/(S_X)_i), \quad (2)$$

where $\overline{O}$ and $\overline{X}_i$ represent the climatological mean of the observations and the i-th raw ensemble member in the training set; S_O and $(S_X)_i$ are the standard deviations of the observations and the i-th raw ensemble member, respectively. Through the prior calibration, the statistical average properties (mean and variance) of each ensemble member are equal to those of the observations.

3.4 MLR

For MLR, the predicted rainfall is expressed as

$$Y_t \sim N(a + b_1 X_{1t} + \cdots + b_m X_{mt}, c + d(S_X^2)_t). \quad (3)$$

Here, a and $b_1, \cdots b_m$ are regression coefficients of regression equation

$$\boldsymbol{u} = a + b_1\boldsymbol{X}_1 + \ldots + b_m\boldsymbol{X}_m. \quad (4)$$

In Eq. (3), $c+d(S_X^2)_t$ is used to estimate the variation of residual error, where $(S_X^2)_t$ is the ensemble spread in year t. In the study of Gneiting et al. (2005), all the coefficients in Eq. (3) are derived by a nonlinear optimization method, which is sensitive to the initial values of these coefficients. Since the number of ensemble members is 45 in the present study and much larger than 5 in Gneiting et al. (2005), identification of the initial values of so many coefficients is difficult and could lead to a fallacious solution. Therefore, in the present study, we first estimate the regression coefficients in Eq. (4) by the principle component regression with O as dependent variable and $X_i (i = 1, 2, \cdots, m)$ as regressors. After the identification of regression coefficients, the coefficients c and d are estimated with nonlinear optimization to minimize the temporal-averaged continuous ranked probability score as in Gneiting et al. (2005).

3.5 Bayes

The Bayesian probabilistic ensemble scheme proposed by Coelho et al. (2004) is used here. The scheme includes three steps: (1) selecting the prior distribution, (2) modeling the likelihood function, and (3) determining the posterior distribution. First, the probabilistic distribution of the climatology prediction is used as the prior distribution of regional summer rainfall in the present study. Since the model and observed summer rainfall are normally distributed or normalized, according to Coelho et al. (2004), the likelihood can be assumed to be normal. The likelihood, then, is modeled by performing a simple linear regression between the ensemble mean prediction $\overline{X}_i$ and observations O_t as $N(aO_t + b, \gamma V_t)$. Here, a and b are regression coefficients estimated from the training set; $V_t = (S^2_X)_t/m$; and the dependency factor γ is the weighted mean of square regression residuals.

Finally, according to the Bayes' theorem, the posterior distribution is also normal for a normal prior distribution and normal likelihood (Lee, 1997). The resulting normal posterior distribution for the target year t is given by

$$Y_t|\overline{X}_t \sim N(u_t, S_t^2), \tag{5}$$

where the mean u_t and variance S_t^2 are

$$\frac{1}{S_t^2} = \frac{1}{S_O^2} + \frac{a^2}{\gamma V_t}, \tag{6}$$

$$\frac{u_t}{S_t^2} = \frac{\overline{O}}{S_O^2} + \frac{a^2}{\gamma V_t}(\frac{\overline{X}_t - b}{a}). \tag{7}$$

Equations (6) and (7) state that the precision of posterior distribution $\frac{1}{S_t^2}$ is exactly equal to the sum of precisions for the prior distribution $\frac{1}{S_O^2}$ and the ensemble system $\frac{a^2}{\gamma V_t}$, while posterior ensemble mean u_t is the weighted average of prior mean and ensemble mean.

4. Results

4.1 Calibration and sharpness

Calibration and sharpness are two desirable properties of probabilistic prediction; and the goal of probabilistic prediction is to maximize the sharpness of the predicted PDF subject to calibration (Raftery et al., 2005; Gneiting et al., 2005, 2007). Calibration (or reliability) refers to the statistical consistency between the predicted PDFs and the observations, and is a joint property of predictions and observations. Sharpness (or resolution) refers to the concentration of the predictive distributions and is a property of the predictions only. In the present study, according to Raftery et al. (2005) and Gneiting et al. (2005, 2007), probability integral transform (PIT) histogram and average width of central 95% prediction interval relative to that of climatology prediction (RW) are used to assess the calibration and the sharpness, respectively.

4.1.1 Calibration

The PIT value is the predictive cumulative distribution at an observation value. PIT histogram is a continuous analog of the verification rank histogram. Its calculation process and relative theorems are introduced by Gneiting et al. (2007) in detail. When the sample size is infinite, if the predictive PDF is calibrated, the PIT values should be uniformly distributed (i.e., the relative frequency in PIT histograms is uniform). In practice, the sample size is generally not large enough, and the uniform distribution of the PIT values is tested by the nonparametric χ^2 goodness-of-fit test (Pearson, 1900).

As shown in Fig. 2, the relative frequencies of the Cali-EE, the MLR, and the Bayes are more uniform than that of the EE in all regions. That is, their PDF predictions are much better calibrated than those of the EE. According to the χ^2 goodness-of-fit test (Table 1), in all regions, the PIT values of the Cali-EE, the MLR, and the Bayes are significantly uniformly distributed ($\alpha = 0.05$), so their PDF predictions are well calibrated. On the contrary, in all regions except R3, the PIT values of the EE have large χ^2 statistic and do not pass the χ^2 goodness-of-fit test, so their PDF predictions are poorly calibrated. These results indicate that the Cali-EE, the MLR, and the Bayes are able to calibrate the prediction products. The good calibration of the Cali-EE prediction suggests that the calibration of statistical average properties of each

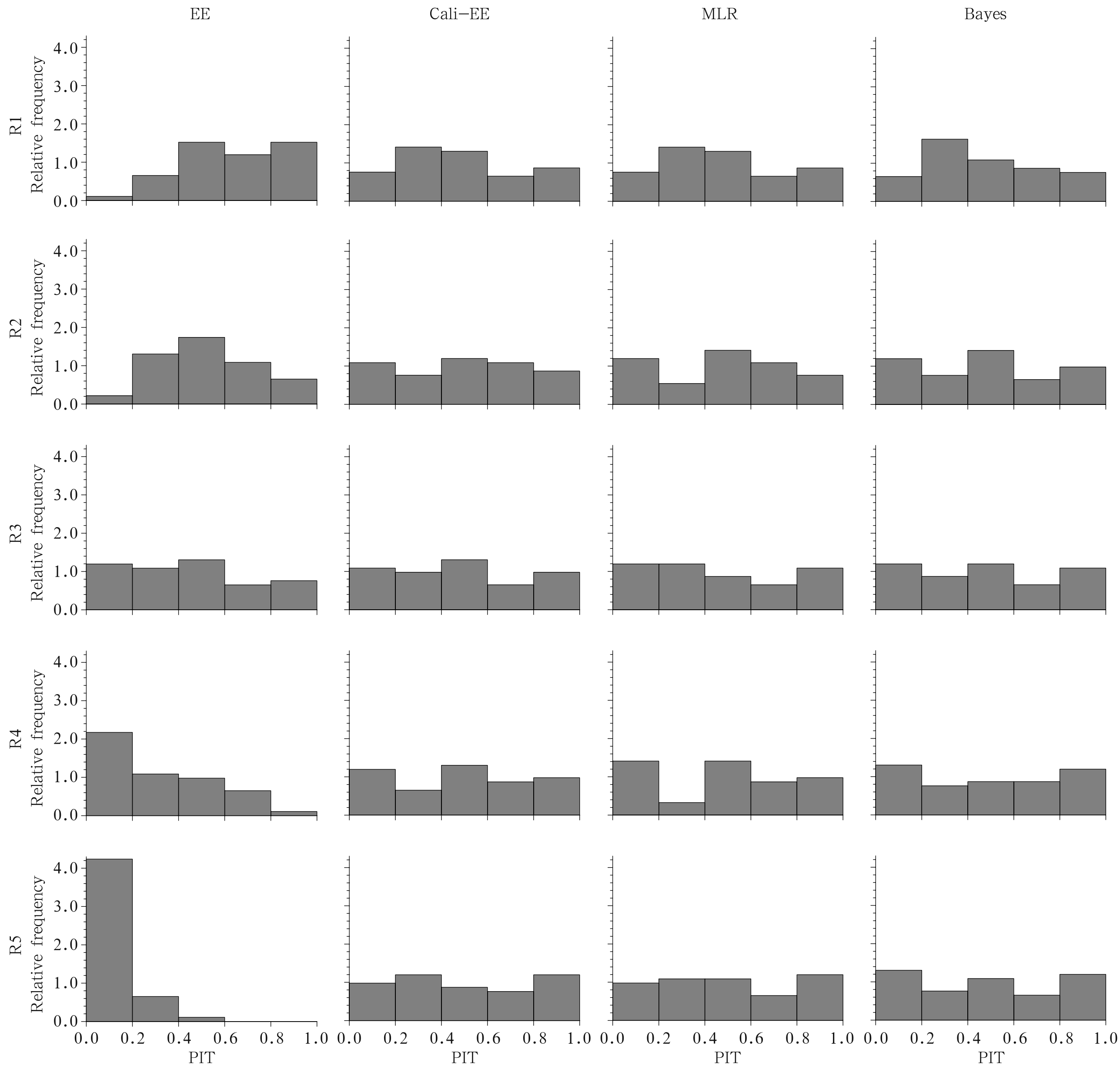

Fig. 2. Probability integral transform (PIT) histograms for different regions (rows) and different ensemble schemes (columns). The four ensemble schemes are the equally-weighted ensemble (EE), EE for calibrated model-simulations (Cali-EE), the ensemble scheme based on multiple linear regression analysis (MLR), and the Bayesian ensemble scheme (Bayes).

ensemble member can lead to well calibrated ensemble PDF prediction. The good calibration of the MLR and the Bayes are consistent with the results in Gneiting et al. (2005) and Coelho et al. (2004).

The reason for the poor calibration of EE is investigated based on the shape of PIT histograms in Fig. 2. In R1, the PIT histogram of the EE is right biased, which indicates the climatological mean of the mathematical expectations of the predicted PDFs in this region is lower than the climatological mean of observations, and hence leading to less rainfall. In R2, the PIT histogram of the EE is distributed with a peak, which accentuates the over-dispersive raw ensemble. In addition, the PIT histogram of EE is left biased in

R4 and R5, especially in the latter, which indicates the PDF predictions in these two regions have more rainfall than observations.

4.1.2 Sharpness

For rainfall in all regions that is normally distributed or normalized, the width of central 95% ($\alpha = 0.05$) predicted interval is equal to $2 \times z_{1-\alpha/2}\sigma$, where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution, and σ is the standard deviation of predicted PDF. Consequently, RW is the average width of central 95% prediction interval relative to that of the climatology prediction. That is, $\mathrm{RW}=(\frac{1}{n}\sum_{t=1}^{n}\sigma_{Y,t})\Big/(\frac{1}{n}\sum_{t=1}^{n}\sigma_{\mathrm{clim},t})$, where $\sigma_{Y,t}$ and $\sigma_{\mathrm{clim},t}$ are the standard deviations of ensemble prediction and climatology prediction, respectively, in the target year t. The smaller the RW, the sharper the predictive PDF, and vice versa. The predicted PDF is sharper than that of the climatology prediction when RW is smaller than 1.

Table 1. Statistics for the χ^2 goodness-of-fit test

	EE	Cali-EE	MLR	Bayes
R1	13.8	**4.2**	**4.2**	**5.5**
R2	12.7	**1.2**	**4.4**	**3.6**
R3	**2.9**	**2.0**	**2.0**	**2.0**
R4	21.2	**2.5**	**7.5**	**2.0**
R5	123.3	**1.4**	**1.6**	**2.9**

Note: The values in boldface denote that the probabilistic distributions of probability integral transform (PIT) values are significantly uniform ($\alpha = 0.05$) and hence their corresponding probability density function (PDF) predictions are well calibrated.

As shown in Fig. 3, in all regions, the predictive PDFs based on the Cali-EE, the MLR, and the Bayes are sharper than that of the EE. In addition, compared with the PDF of climatology prediction, the predictive PDFs based on the EE in all regions and the MLR in R3 are evidently flatter while the predictive PDFs based on the MLR in R4 and R5 and the Bayes in R5 are evidently sharper; the sharpness of the others is close to that of the climatology prediction. Note that Eq. (6) decides that the PDFs predicted by the Bayes are always sharper than that of the climatology prediction.

4.2 Skill evaluation

The prediction skill is decided by calibration and sharpness together. It is directly proportional to the

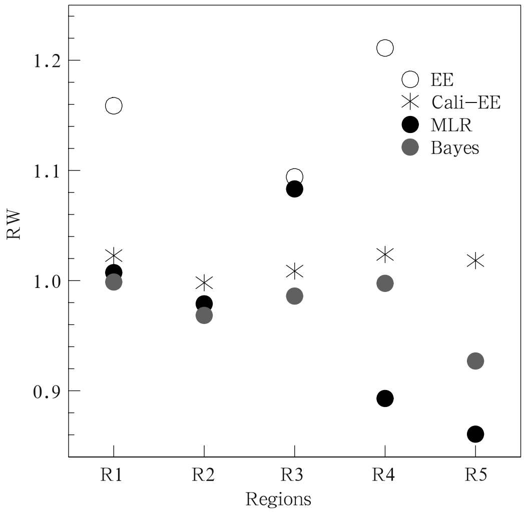

Fig. 3. Average width of cantral 95% prediction interval relative to that of the climatology prediction (RW) for different regions and different ensemble schemes.

calibration when the sharpness is fixed, but generally not to sharpness when the calibration is fixed unless the PDF prediction is perfectly calibrated. Gneiting et al. (2007) recommended temporal-averaged continuous ranked probability score (CRPS) as the score rule to address the calibration and sharpness simultaneously. The CRPS is the integral of the Brier scores at all possible threshold values of predictand, and represents the difference between cumulative distribution functions of predictions and observations (Hersbach, 2000). Because the climatology prediction is the benchmark of the PDF prediction of seasonal rainfall, in the present study, we use the predicted CRPS relative to that of the climatology prediction (RCRPS) instead of CRPS as skill score. The PDF prediction is more skillful than the climatology prediction when RCRPS is smaller than 1. The smaller the RCRPS, the more skillful the prediction.

As shown in Fig. 4, in R1, the Bayes performs the best, followed by the climatology prediction, the MLR, the Cali-EE, and the EE. In R2 and R5, the Cali-EE performs the best, followed by the MLR, the Bayes, the climatology prediction, and the EE. In R3, the Cali-EE is the most skillful, followed by the Bayes, the EE, the climatology prediction, and the MLR. In R4, the MLR is the most skillful, followed by the Cali-EE, the Bayes, the climatology prediction, and the

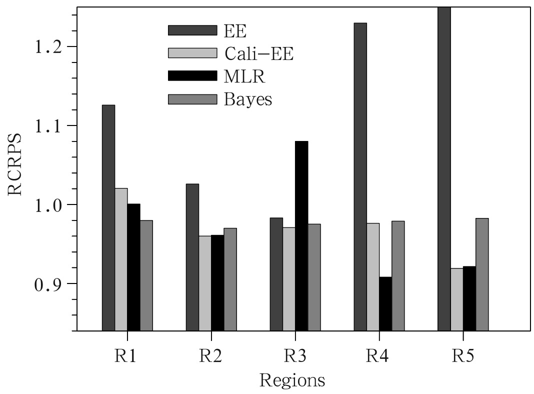

Fig. 4. Predicted temporal-averaged continuous ranked probability score (CRPS) relative to that of climatology prediction (RCRPS) for different regions and different ensemble schemes.

EE. That is, the optimal ensemble schemes with the smallest RCRPS are the Bayes in R1, the Cali-EE in R2, R3, and R5, and the MLR in R4. Obviously, in all regions, they are more skillful than the EE, and hence more suitable for the PDF prediction of summer rainfall over East China.

Compared with the climatology prediction, in all regions, these optimal ensemble schemes are more skillful, which indicates that they do have some prediction skill. Moreover, according to sub-sections 4.1 and 4.2, these optimal ensemble schemes are well calibrated, and are sharper than or approximately equal to the interval-width of the climatology prediction.

5. Conclusions and discussion

Based on the model-output rainfall data from the ENSEMBLES project, we have analyzed the calibration, sharpness, and skill of PDF predictions by four commonly-used multi-model ensemble schemes, and have identified the optimal ensemble scheme with the highest skill in seasonal prediction of summer rainfall in each sub-region of East China. Furthermore, by comparing the optimal ensemble schemes with the climatology prediction, the status of the multi-model PDF prediction of summer rainfall over East China has been assessed. The results show that compared with the EE, the PDF predictions of the Cali-EE, the MLR, and the Bayes are better calibrated and sharper in all sub-regions and more skillful except for the MLR in R3. The optimal ensemble scheme is the Bayes in the southern part of East China; the Cali-EE in the Yangtze River valley, the Yangtze-Huaihe River basin, and the central part of northern China; and the MLR in the eastern part of northern China. Their predictive PDFs are well calibrated, and are sharper than or close to that of the climatology prediction. In all regions, the optimal ensemble schemes outperform the climatology prediction, indicating that the commonly-used ensemble schemes are able to produce skillful PDF predictions.

In fact, the actual skill of the multi-model ensemble PDF prediction for regional summer rainfall over East China shall not be lower than the skill of optimal ensemble schemes shown in the present study for three reasons. First, the model data in the present study are from only five CGCMs in the ENSEMBLES dataset. There may be other combinations of CGCMs which provide more prediction information for the summer rainfall over East China. Second, there are other ensemble prediction schemes besides the four schemes used in the present study. The other ensemble schemes may be able to identify more prediction information, for example, the ensemble scheme that assigns unequal weights to ensemble members according to their skill scores; more sophisticated calibrating methods in Cali-EE; and more sophisticated likelihood model in Bayesian ensemble schemes. In addition, identifying information from more model variables to optimize ensemble members may also improve the skill of multi-model ensemble PDF prediction further (Li et al., 2009).

Acknowledgments. The author thanks two anonymous reviewers and the editors for helpful comments. Drs. Z. D. Lin and C. F. Li from the Institute of Atmospheric Physics, Chinese Academy of Sciences are appreciated for valuable suggestions.

REFERENCES

Alessandri, A., A. Borrelli, A. Navarra, et al., 2011: Evaluation of probabilistic quality and value of the ENESMBLES multi-model seasonal forecast comparison with DEMETER. *Mon. Wea. Rev.*, in press.

Bera, A. K., and C. M. Jarque, 1980: Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letter*, **6**, 255–259.

Chen Hong, 2003: IAP dynamical extraseasonal and interannual climate prediction system and its real-time prediction. Ph. D. dissertation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 62 pp. (in Chinese)

Chen, L. J., D. L. Chen, H. J. Wang, et al., 2009: Regionalization of precipitation regimes in China. *Atmos. Oceanic Sci. Lett.*, **2**, 301–307.

Coelho, C. A. S., S. Pezzulli, M. Balmaseda, et al., 2004: Forecast calibration and combination: A simple bayesian approach for ENSO. *J. Climate*, **17**, 1504–1516.

Ding Yihui, Li Qingquan, Li Weijing, et al., 2004: Advance in seasonal dynamical prediction operation in China. *Acta Meteor. Sinica*, **62**, 598–612. (in Chinese)

Doblas-Reyes, F. J., A. Weisheimer, M. Déqué, et al., 2009: Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. *Quart. J. Roy. Met. Soc.* **135**, 1538–1559.

Feng Jinming and Fu Congbin, 2007: Inter-comparison of long-term simulations of temperature and precipitation over China by different regional climate models. *Chinese J. Atmos. Sci.*, **31**, 805–814. (in Chinese)

Gneiting, T., 2008: Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society (Series A)*, **171**, 319–321.

——, A. E. Raftery, A. H. Westveld, et al., 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118.

——, F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society (Series B)*, **69**, 243–268.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570.

Huang Jiayou, 1993: *Statistical and Dynamical Analysis and Forecast.* China Meteorological Press, Beijing, 1–32. (in Chinese)

Huang Ronghui, Li Chongyin, and Wang Shaowu, 2003: *Research on the Formation Mechanism and Prediction Theory of Severe Climate Disasters in China.* China Meteorological Press, Beijing, 483 pp. (in Chinese)

Kang, I. S., and J. H. Yoo, 2006: Examination of multi-model ensemble seasonal prediction methods using a simple climate system. *Clim. Dyn.*, **26**, 285–294.

Ke Zongjian, 2007: Multimodel ensemble analysis in seasonal climate prediction. Ph. D. dissertation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 173 pp. (in Chinese)

Ke, Z., P. Zhang, W. Dong, et al., 2009: A new way to improve seasonal prediction by diagnosing and correcting the intermodel systematic errors. *Mon. Wea. Rev.*, **137**, 1898–1907.

Kharin, V. V., and F. W. Zwiers, 2002: Climate predictions with multimodel ensembles. *J. Climate*, **15**, 793–799.

Kirtman, B. P., D. Min, P. S. Schopf, et al., 2003: A new approach for coupled GCM sensitivity studies. COLA Technical Report CTR 154, America, 48 pp.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRaw, et al., 1999: Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, **285**, 1548–1550.

Lang Xianmei, Wang Huijun, and Jiang Dabang, 2004: Extraseasonal short-term predictions of summer climate with IAP9L-AGCM. *Chinese J. Geophys.*, **47**, 19–24. (in Chinese)

Lau, K. M., K. M. Kim, and J. Y. Lee, 2004: Interannual variability, global teleconnection and potential predictability associated with the Asian summer monsoon. *East Asian Monsoon*, C. P. Chang, Ed., World Scientific Publisher Co., Singapore, 153–176.

Lavers, D., L. Luo, and E. F. Wood, 2009: A multiple model assessment of seasonal climate forecast skill for applications. *Geophys. Res. Lett.*, **36**, L23711, doi: 10.1029/2009GL041365.

Lee, P. M., 1997: *Bayesian Statistics: An Introduction.* 2nd ed., Arnold, 344 pp.

Li, F., Q. C. Zeng, and C. F. Li, 2009: A Bayesian scheme for probabilistic multi-model ensemble prediction of summer rainfall over the Yangtze River valley. *Atmos. Oceanic Sci. Lett.*, **2**, 314–319.

Li Qingquan, Ding Yihui, and Zhang Peiqun, 2004: A rough test and assessment of summer prediction over season of the global coupled ocean-atmosphere model. *Acta Meteor. Sinica*, **62**, 740–751. (in Chinese)

Li Weijing, Zhang Peiqun, Li Qingquan, et al., 2005: Research and operational application of dynamical climate model prediction system. *Chinese J. Appl. Meteor.*, **16**, 1–11. (in Chinese)

Lin Zhaohui, Li Xu, Zhao Yan, et al., 1998: An improved short-term climate prediction system and its application to the extraseasonal prediction of rainfall anomaly in China for 1998. *Clim. Environ. Res.*, **3**, 339–348. (in Chinese)

Liu Yanxiang and Guo Yufu, 2005: Experiment of interannual climate prediction for 2003 with coupled model. *Clim. Environ. Res.*, **10**, 257–264. (in Chinese)

Luo, L., E. F. Wood, and M. Pan, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions, *J. Geophys. Res.*, **112**, D10102, doi: 10.1029/2006JD007655.

North, G. R., T. L. Bell, R. F. Cahalan, et al., 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706.

Palmer, T. N., A. Alessandri, U. Andersen, et al., 2004: DEMETER: Development of a European multimodel ensemble system for seasonal to interannual prediction. *Bull. Amer. Meteor. Soc.*, **85**, 853–872.

—–, F. Doblas-Reyes, R. Hagedorn, et al., 2005: Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Phil. Trans. R. Soc. B.*, **360**, 1991–1998.

Pearson, K., 1900: On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine*, **50**, 157–175.

Peng, P., A. Kumar, H. Van den Dool, et al., 2002: An analysis of multimodel ensemble predictions for seasonal climate anomalies. *J. Geophys. Res.*, **107**, 4710, doi: 10.1029/2002JD002712.

Qin Zhengkun, 2007: Bias correction and superensemble method for seasonal-interannual dynamical climate prediction. Ph. D. dissertation, Nanjing University of Information Science & Technology, Nanjing, 140 pp. (in Chinese)

Raftery, A. E., T. Gneiting, F. Balabdaoui, et al., 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.

Tippett, M. K., A. G. Barnston, and A. W. Robertson, 2007: Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *J. Climate*, **20**, 2210–2228.

Von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, UK, 455 pp.

Wang, B., Q. H. Ding, X. H. Fu, et al., 2005: Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophys. Res. Lett.*, **32**, L15711. doi:10.1029/2005GL022734.

—–, J. Lee, I. Kang, et al., 2009: Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004). *Clim. Dyn.*, **33**, 93–117.

Wang Huijun, Xue Feng, and Bi Xunqiang, 1997: A preliminary study on the uncertainty of short-term climate prediction. *Clim. Environ. Res.*, **2**, 217–222. (in Chinese)

—–, Sun Jianqi, Lang Xianmei, et al., 2008: Some new results in the research of the interannual climate variability and short-term climate prediction. *Chinese J. Atmos. Sci.*, **32**, 806–814. (in Chinese)

Wang Shaowu, 2001: *Research Progress in Modern Climatology.* China Meteorological Press, Beijing, 453 pp. (in Chinese)

Wang Xiaoling and Tu Qipu, 2002: Regional distribution of annual variations of dekad precipitation in China. *J. Nanjing Inst. Meteor.*, **25**, 518–524. (in Chinese)

Wei Fengying, 1999: *Modern Statistical Technology in Climatological Diagnoses and Prediction.* China Meteorological Press, Beijing, 128–134. (in Chinese)

Wei Jie, Zhang Qingyun, and Tao Shiyan, 2005: The ensemble seasonal climate prediction for 2004 summer and its verification. *Clim. Environ. Res.*, **10**, 19–31. (in Chinese)

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2009: Seasonal ensemble forecasts: Are recalibrated single models better than multimodels? *Mon. Wea. Rev.*, **137**, 1460–1479.

Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, et al., 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, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, American, 467 pp.

Wu, R., and B. Kirtman, 2005: Roles of Indian and Pacific Ocean air-sea coupling in tropical atmospheric variability. *Clim. Dyn.*, **25**, 155–170.

Zeng Qingcun, Yuan Congguang, Wang Wanqiu, et al., 1990: Experiments in numerical extra-seasonal prediction of climate anomalies. *Chinese J. Atmos. Sci.*, **14**, 10–25. (in Chinese)