ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 25, NO. 1, 2008, 72–82

Multimodel Ensemble Forecasts for Precipitations in China in 1998

KE Zongjian*[1,2,3] (柯宗建), DONG Wenjie[2] (董文杰), and ZHANG Peiqun[2] (张培群)

[1] *Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

[2] *Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing* 100081

[3] *Graduate University of Chinese Academy of Sciences, Beijing* 100049

(Received 17 October 2006; revised 24 January 2007)

ABSTRACT

Different multimodel ensemble methods are used to forecast precipitations in China, 1998, and their forecast skills are compared with those of individual models. Datasets were obtained from monthly simulations of eight models during the period of January 1979 to December 1998 from the "Climate of the 20th Century Experiment" (20C3M) for the Fourth IPCC Assessment Report. Climate Research Unit (CRU) data were chosen for the observation analysis field. Root mean square (RMS) error and correlation coefficients (R) are used to measure the forecast skills. In addition, superensemble forecasts based on different input data and weights are analyzed. Results show that for original data, superensemble forecasting based on multiple linear regression (MLR) performs best. However, for bias-corrected data, the superensemble based on singular value decomposition (SVD) produces a lower RMS error and a higher R than in the MLR superensemble. It is an interesting result that the SVD superensemble based on bias-corrected data performs better than the MLR superensemble, but that the SVD superensemble based on original data is inferior to the corresponding MLR superensemble. In addition, weights calculated by different data formats are shown to affect the forecast skills of the superensembles. In comparison with the MLR superensemble, a slightly significant effect is present in the SVD superensemble. However, both the SVD and MLR superensembles based on different weight formats outperform the ensemble mean of bias-corrected data.

Key words: precipitation, multimodel ensemble, China

DOI: 10.1007/s00376-008-0072-y

1. Introduction

Initial errors and model errors are the two key types of error which can lead to uncertainties in numerical weather and climate prediction. In general, initial errors are considered to be the main problem, while model errors are more relevant to short-term climate prediction (Mu et al., 2002). In order to reduce the effects of initial errors, different methods are used to generate initial perturbations (Leith, 1974; Hoffman and Kalnay, 1983; Molteni et al., 1996; Toth and Kalnay, 1993, 1997), with breeding and singular vectors being two of the more popular methods at the present time. For model errors, some scientists believe that multimodel ensembles could decrease the uncertainties of single models, and thus this is an approach which has become popular in recent years (Kalnay, 2003).

It has long been known that an ensemble average of operational global forecasts from different operational centers is generally more skillful than the best individual forecast (Fritsch et al., 2000). Based on an ensemble average, a multimodel superensemble was proposed (Krishnamurti et al., 1999), and many subsequent studies have since been made (Krishnamurti et al., 2000, 2001; Pavan and Doblas-Reyes, 2000; Yun et al., 2003, 2005). Some of this research has indicated that multimodel superensembles can improve the forecasting skills (Krishnamurti et al., 1999, 2000, 2001), however other results have shown few advantages or improvements when using this approach (Pavan and Doblas-Reyes, 2000; Peng et al., 2002; Kharin and

*Corresponding author: KE Zongjian, zongjke@mail.iap.ac.cn

Zwiers, 2002).

Pavan and Doblas-Reyes (2000) combined four different AGCMs for seasonal forecasting and found minimal improvement. Similarly, Peng et al. (2002) indicated that the use of more sophisticated techniques for constructing multimodel ensembles may not be any more advantageous than the use of simpler approaches. Furthermore, Kharin and Zwiers (2002) found that superensemble forecasting obtained by optimally weighting the individual ensemble members did not perform well, either in the simple ensemble mean or in the regression-improved ensemble mean. Yun et al. (2003) suggested that the disagreement between Kharin and Zwiers (2002) and Krishnamurti et al. (2000) is due to the fact that, in the former of the two studies, the seasonal mean was removed only after the regression coefficients were calculated, whereas in Krishnamurti's work, the seasonal mean was removed prior to the calculation of the regression coefficients. How, therefore, does the order in which weights are derived affect the ability of the ensemble? Crucially, previous studies have tended to focus on the effect of data formats on the ensemble mean, but have paid little attention to data format differences affecting superensemble skills. Furthermore, atmospheric predictability is low in the extratropics. How would a multimodel ensemble perform in China? These are the main issues to be addressed by the present study, and to achieve this eight models from the "Climate of the 20th Century Experiment" (20C3M) were chosen to assess the abilities of different methods of ensemble forecast and to analyze the effects of different input data on these forecasting skills.

Three ensemble methods are used, including the ensemble mean, multiple linear regression (MLR) and singular value decomposition (SVD). In terms of Krishnamurti et al. (1999) and Yun et al. (2003), the MLR and SVD methods are equivalent to the MLR and SVD superensembles, respectively. Original and bias-corrected data will be treated as the input data. Finally, superensemble results regarding weights derived from different data formats will be analyzed.

2. Datasets and superensemble construction

Datasets were obtained from the monthly simulations of eight models during the period January 1979 to December 1998 from the "Climate of the 20th Century Experiment" (20C3M) for the Fourth IPCC Assessment Report. The eight models chosen for the construction of the superensemble presented by Krishnamurti et al. (1999) are shown in Table 1. Climate Research Unit (CRU) data were used for the observation analysis field. In terms of the format of the CRU data, all multimodel forecast fields were interpolated to a common resolution of $0.5^\circ \times 0.5^\circ$ for the monthly mean time intervals. Various multimodel approaches, including the ensemble mean, MLR and SVD superensembles (SE) for precipitation prediction have been discussed in the literature (Krishnamurti et al., 2001, 2003; Stephenson and Doblas-Reyes, 2000; Yun et al., 2003). These are defined as follows:

$$\mathrm{Eb} = \frac{1}{N}\sum_{i=1}^{N}(F_i) \ , \tag{1}$$

$$\mathrm{SE} = \overline{O} + \sum_{i=1}^{N} a_i(F_i - \overline{F}_i) \ , \tag{2}$$

where Eb is the bias ensemble mean. By substituting original data for the seasonal-cycle removed data, then Eq. (1) can be written as:

$$\mathrm{Ec} = \overline{O} + \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{F}_i) \ . \tag{3}$$

Here, Ec represents the bias-corrected ensemble mean; F_i is the ith model forecast out of N models; $\overline{F}_i$ is the monthly mean of the ith forecast over the training period; $\overline{O}$ is the monthly mean of the observed state over the training period, and corresponds to $\overline{F}_i$; and a_i is the regression coefficient of the ith model, which is computed at each grid point; SE represents the superensemble.

In the conventional mulitimodel superensemble forecast (Krishnamurti et al., 2000), a_i is obtained using Gauss-Jondan elimination with pivoting, and in the SVD superensemble a_i is calculated by using singular value decomposition of the covariance matrix and selecting only the largest singular value (Yun et al., 2003). If original data are used to perform the superensemble forecast, Eq. (2) could be written as:

$$\mathrm{SE} = \sum_{i=1}^{N} a_i \times F_i \ . \tag{4}$$

3. Verification metrics

In order to measure forecasting ability, either the bias between the forecast and observation can be calculated, or the degree of correlation between them can be obtained. The correlation coefficient (R) is often used to assess the relationship between two meteorological variables, and root mean square (RMS) denotes the mean state of difference between a variable and the mean value. The correlation coefficient is a good

Table 1. Selected 20C3M Project models.

Acronym	Horizontal resolution (Atmosphere)	20C3M group
CCSM3	1.40625° × 1.40625°	National Center for Atmospheric Research, Boulder, Colorado, USA
CGCM3.1	2.8125° × 2.8125°	Canadian Climate Centre, Downsview, Ontario, Canada
CNRM-CM3	2.8125° × 2.8125°	Centre National de Recherches Météorologiques, Toulouse, France
ECHAM5	1.875° × 1.875°	Max Planck Institute for Meteorology, Hamburg, Germany
FGOALS1.0	2.8125° × 2.7906°	Institute of Atmospheric Physics, Beijing, China
GFDL-CM2.1	2.5° × 2.0225°	Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA
MRI-CGCM2.3.2	2.875° × 2.875°	Meteorological Research Institute, Tsukuba, Japan
UKMO-HadGEM1	1.875° × 1.875°	United Kingdom Meteorological Office, Bracknell, England, UK

measure of phase error (Déqué, 1997), but forecasts with large errors of magnitude still have the possibility of good correlation coefficients. Therefore, it is necessary to use RMS to evaluate the magnitude of errors. In this paper, both RMS and R are used to compare different ensemble skills and are defined as follows:

$$\mathrm{RMS} = \sqrt{\frac{1}{G}\sum[(F-\overline{F})-(O-\overline{O})]^2}\,, \qquad (5)$$

$$R = \frac{\sum(F-\overline{F})(O-\overline{O})}{\sqrt{\sum(F-\overline{F})^2}\sqrt{\sum(O-\overline{O})^2}}\,. \qquad (6)$$

Here, the overbar denotes time average, and G denotes the number of grid points. It is necessary to note that $\overline{F}$ represents the forecast average over the forecast period, and the R only means the correlation coefficient between the forecast and observation during the forecast period. In order to quantify the difference between different forecasts, the following definition can be adopted. The improvement of the R of forecast A over forecast B can be defined as R_A/R_B-1, where R_A and R_B are the R of forecasts A and B, respectively. Similarly, with regard to RMS, the improvement of forecast A over forecast B can be defined as $1-\mathrm{RMS}_A/\mathrm{RMS}_B$ (Yun et al., 2005). These definitions are used to compare the skills of any two forecasts.

4. Results

4.1 *Performance of multimodels and ensemble forecasts*

Given the definition of the proposed superensemble (Krishnamurti et al., 1999), a multimodel superensemble based on MLR has been used to study the ensemble forecast. Some scientists believe that the superensemble technique can improve forecasting ability because of the collective information of all the models used in the statistical algorithm (Krishnamurti et al., 2000), although other results have shown that superensembles do not perform better than simple ensemble mean (Peng et al., 2002; Kharin and Zwiers, 2002). In addition, it has been found that the use of SVD for constructing superensembles provides an incremental improvement in forecasting ability over conventional (MLR) superensembles (Yun et al., 2003). In the present study, different performances of individual models and their ensembles are analyzed.

First, the period of 1979–1997 was chosen for the training time, and 1998 for the forecast time. Monthly mean precipitation was adopted for the calculation. In this subsection, ensemble skills based on three different methods are compared. Figure 1 shows the observed and forecasted total precipitation distributions by individual models and ensemble methods in China for 1998. From the observation field, it can be seen that the most significant rainfall center occurs in East China, and in the southern part of the Yangtze River. In addition, South China shows heavy precipitation. Most of the individual models show the heaviest rainfall in the east of the Tibetan Plateau, but only Figs. 1a, 1d, 1e, and 1g partly reflect the heaviest rainfall center shown in the observation field. However, the rainfall magnitude is remarkably underestimated, and the position of the rainfall center differs between each model. Ensemble mean forecasting based on bias data does not remove the overestimated rainfall center located on the Tibetan Plateau, but the MLRb and SVDb superensemble forecasts realize it. Moreover, compared with individual models, superensemble forecasts show the position of the rainfall center more accurately, although the rainfall magnitude is underestimated. Compared with bias data, it is found that ensemble forecasts based on bias-corrected data reflect more information similar to the observation field, although they also underestimate strong precipitation centers in the observation field. One of the probable reasons for underestimating rainfall is that the grids of individual models are significantly coarser than those

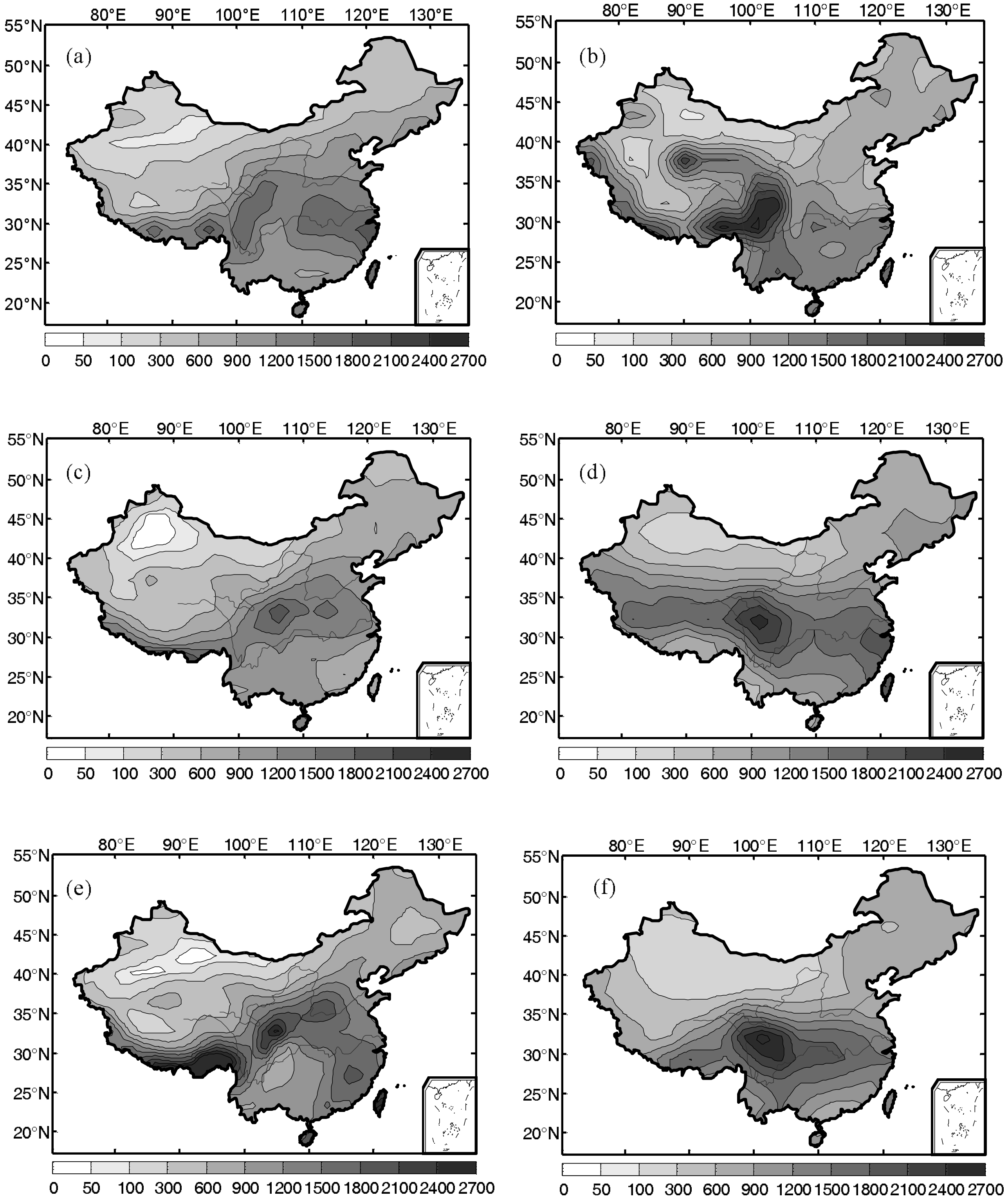

Fig. 1. Total precipitations for individual models, ensemble forecasts and the observation field in China, 1998 (unit: mm): (a)–(h) are the individual models (CCCMA, CNRM, GDFL, IAP, MPI, MRI, NCAR, and UKMO); (i)–(j), (k)–(l) and (m)–(n) are, respectively, the ensemble mean, MLR, and SVD superensembles for original and bias-corrected data; and (o) represents observation.

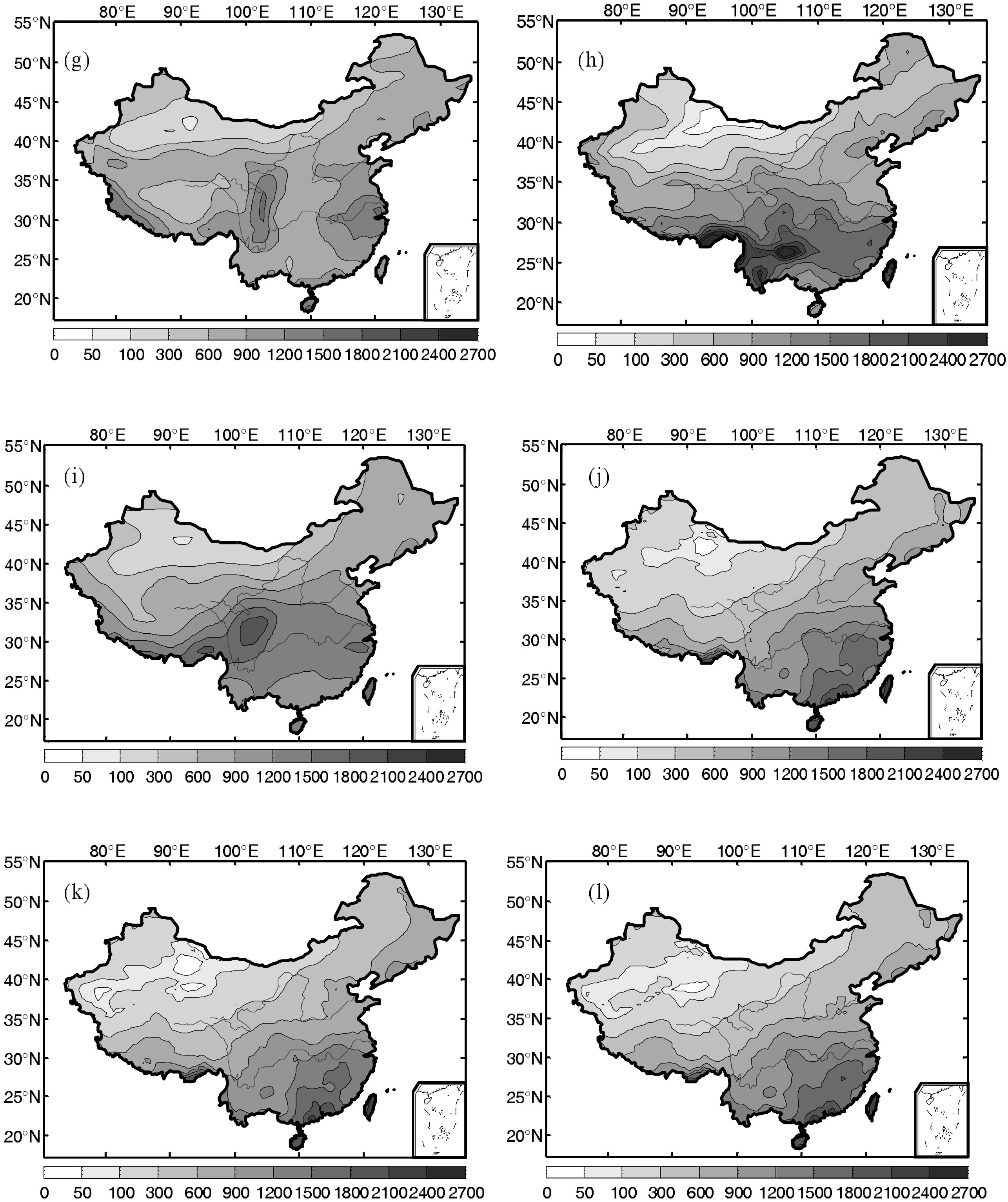

Fig. 1. (Continued).

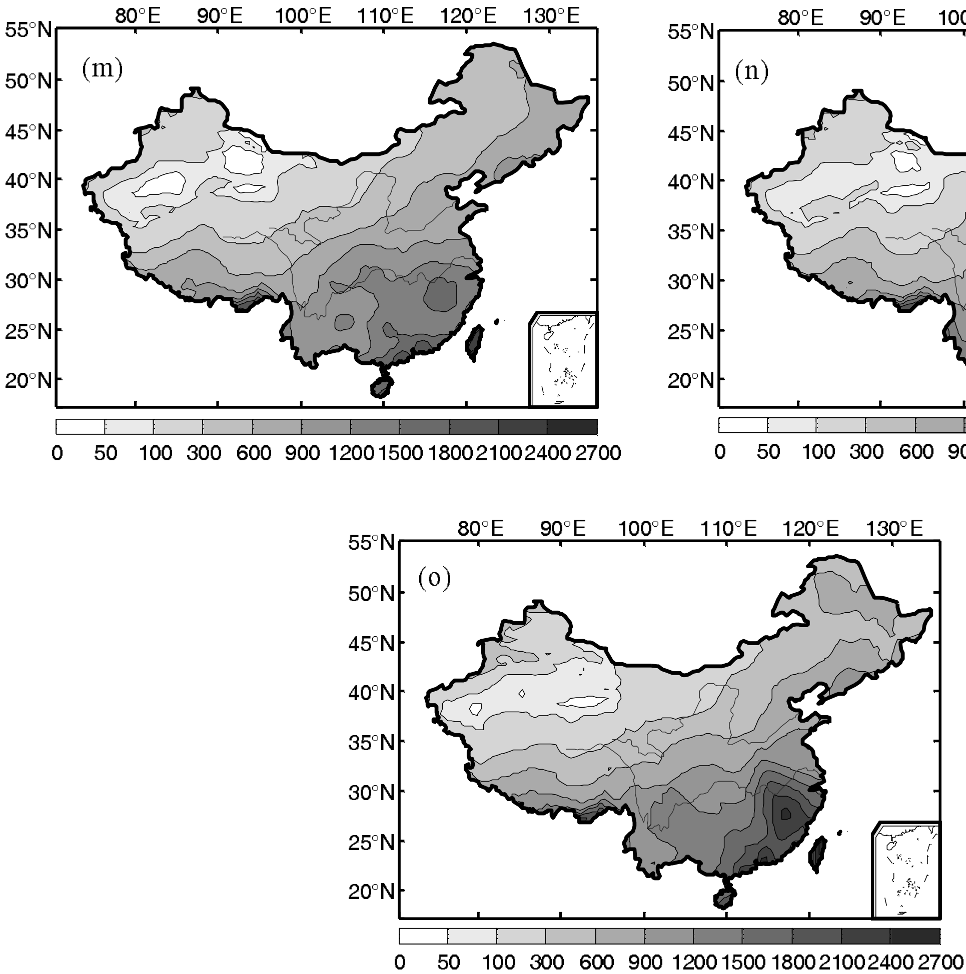

Fig. 1. (Continued).

Table 2. Mean RMS and R of precipitation events for individual models and ensemble forecasts based on different data inputs. A R value of 0.576 represents a coefficient significant at the 95% level ($n = 12$).

	RMS	R
CCCMA	1.755	0.650
CNRM	1.910	0.627
GFDL	1.843	0.405
IAP	2.029	0.363
MPI	2.031	0.457
MRI	1.819	0.425
NCAR	1.635	0.508
UKMO	1.933	0.637
Eb	1.344	0.667
MLRb	1.233	0.775
SVDb	1.326	0.654
Ec	1.181	0.789
MLRc	1.117	0.815
SVDc	1.089	0.842

of the observation field, and strong rainfall centers which occur in the observation field are negligible in the individual models. Finally, for heavy precipitations in South China, most ensemble forecasts perform well, except for Eb in Fig. 1.

Figure 2 shows the averaged RMS and R of individual models and ensemble forecasts for the forecast period. It can be seen that individual models show different abilities to forecast precipitations, and that ensemble forecasts perform better than individual models. For bias data, it can be seen that the ensemble mean (Eb) is superior to the best individual model, and the superensemble based on MLR shows the best forecasting ability. As for the forecasting skill of the SVD superensemble, both the RMS and R values are slightly lower than for Eb. However, for bias-corrected data, the SVD superensemble performs best and significantly better than other ensemble forecasts.

Table 2 lists the values of RMS and R for all fore-

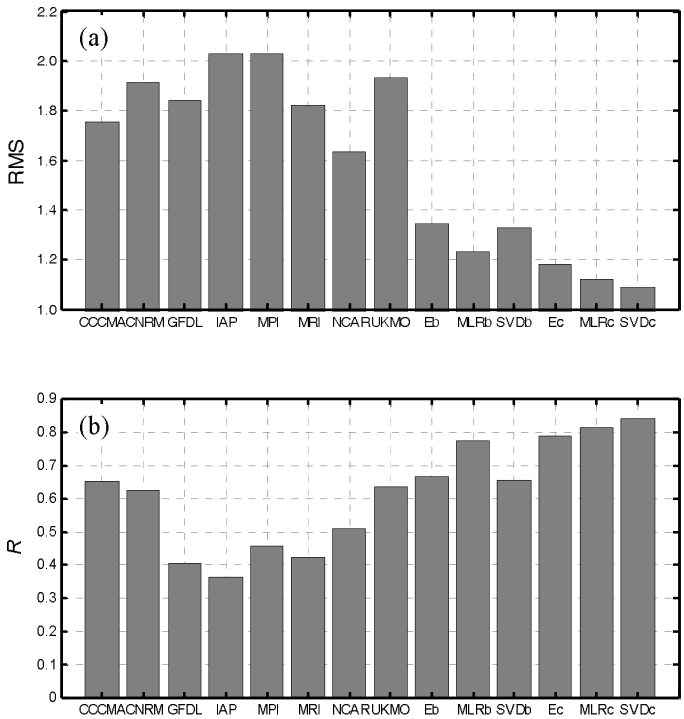

Fig. 2. Mean RMS and R of precipitations for the individual models and ensemble forecasts during the forecast period from January to December 1998 in China. Units for RMS is mm d^{-1}. (a) RMS; (b) R.

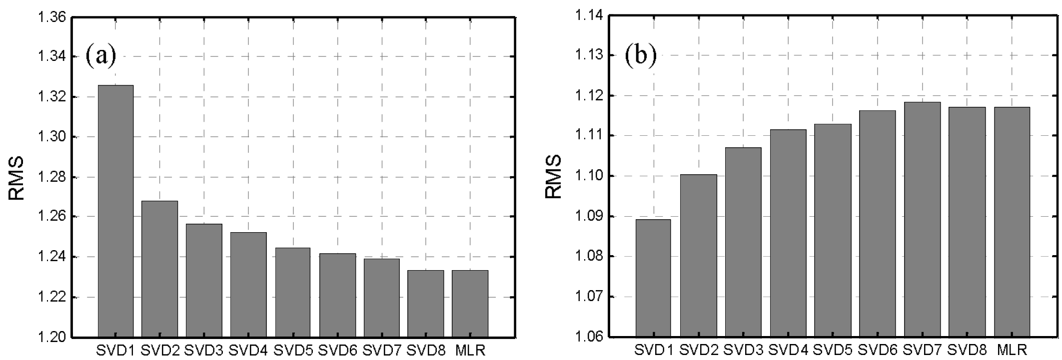

Fig. 3. Mean RMS of precipitations for the SVD superensemble forecasts derived from different inputs of data during the forecast period from January to December 1998. SVDn represent the SVD superensemble with n larger singular values held in calculating weights (units: mm d^{-1}). (a) SVDb; (b) SVDc.

casts. The RMS range of individual models is between 1.635 and 2.031. Similarly, the R range of individual models is between 0.363 and 0.650. In terms of the above definition of improvement, the RMS improvements of the Ec, MLRc, and SVDc forecasts based on bias-corrected over those of bias data are 12.1%, 9.4% and 17.9%, respectively. The corresponding R improvements are 18.3%, 5.2%, and 28.8%. It is obvious that the SVD superensemble is the most sensitive to the format of input data. By comparing SVDc with Ec and MLRc, 7.8% and 2.5% improvements in RMS, and 6.7% and 3.3% improvements in the R are found.

It is interesting that the SVD superensemble skill for bias data is inferior to the corresponding MLR superensemble, but that the SVD superensemble skill for bias-corrected data is superior. It indicates that for the same ensemble method, the SVD superensemble based on bias-corrected data significantly improves forecasting. The SVD method selects only the largest singular value and makes the residual less when weights are solved (Yun et al., 2003), although this is only for a SVD superensemble of bias-corrected data. However, it can clearly be seen that the forecasting skills of SVDb is worse than MLRb. This indicates that different formats of data not only influence the forecasting skills based on the same method, but also have differing degrees of effect on various ensemble methods.

Figure 3 shows the averaged RMS of forecasted precipitations by the SVD method under different conditions where different numbers of singular values are held in calculating weights. For bias data, RMS shows a tendency to decrease with an increasing number of singular values held (Fig. 3a). It is worthwhile noting that the SVD superensemble is the same as the MLR superensemble when all singular values are held. In other words, if the character string "SVDn" is used to represent the SVD superensemble with n larger singular values held in calculating weights, the performance of SVD8 is equal to that of the MLR superensemble. However, the SVD superensemble based on bias-corrected data reflects an opposite phenomenon (Fig. 3b), and the RMS of SVD1 is sharply less than that of SVD8. Its RMS distribution shows an approximate tendency to increase with an increasing number of singular values held. This can be used to explain that the SVD superensemble performs worse than the MLR superensemble for bias data, but better than the MLR superensemble for bias-corrected data. This indicates that for bias-corrected data, only choosing the largest singular value to solve weights could perform best in the SVD method, which is in accordance with the results of Yun et al. (2003).

In order to understand the performance of ensemble forecasts in detail, monthly RMS errors during the forecast period are shown in Fig. 4. RMS error in summer is significantly higher than that of the other seasons, which is in accordance with the heaviest precipitation in China taking place in summer. The ensemble for bias-corrected data performs generally better than the ensemble based on bias data, except in September and October. For bias-corrected data, the SVD superensemble in general shows a lower RMS than other ensembles.

Figure 5 shows bias percentages of summer precipitations forecasted by several ensemble methods in China for 1998. In comparison with Eb, all other ensemble forecasts show obvious improvement. In the south of Xinjiang Province, MLRc and SVDc show less bias than Ec. However, it is also found that Ec performs best for Northeast China. This indicates that different levels of performance of the ensemble methods could be present in different regions, although in general, the superensemble shows more advantages than the ensemble mean in China. It could be used to partly explain disagreement existing in previous results. For a certain grid, the superensemble being superior to the simple ensemble mean mainly depends on whether or not weights obtained during the training time fit various models during the forecast time.

4.2 *Different order of obtaining weights*

Kharin and Zwiers (2002) found a disagreement with the results of Krishnamurti et al. (2000), in that the superensemble performed worse than the ensemble mean. Yun et al. (2003) considered that this discrepancy was due to the fact that the seasonal mean was removed only after the regression coefficients were calculated, while the seasonal mean was removed prior to the calculation of regression coefficients in the work by Krishnamurti et al. (2000). In other words, they used a different data format to calculate regression coefficients of the superensemble. How different, then, is it when different data formats are used to calculate weights? Is this difference enough to obtain an opposite forecast result? How great an effect do weights obtained with different data formats have on forecasting ability? To answer this, regression coefficients with different data formats are calculated and superensemble forecasts are made. In order to distinguish between them, the term "wr" is used to represent the process of first calculating weights, and then removing seasonal cycle information; and "rw" to represent the process in reverse order. Indeed, the result of "rw" is the same as that of bias-corrected data. Figure 6 illustrates the mean RMS and R of the MLR and SVD superensembles based on different regression coefficients, and it can be seen that different regression coefficients have a slight effect on the MLR and SVD superensembles. The RMS and R improvements of MLRrw over MLRwr are 4.0% and 0.3%, respectively. In comparison with this, slightly obvious differences derived from the discrepancy of calculating weights are shown for the SVD superensemble. The RMS and R improvements for SVDrw over SVDwr are 6.0% and 2.6%, respectively. Therefore, the SVD superensemble is slightly more sensitive to the order in which weights are derived than the MLR superensemble, and calculating weights after removing the seasonal mean could improve forecasting skills. For the MLR superensemble, weights derived from bias-corrected data also improve forecasting skills, however, the degree of improvement is inferior to that of the SVD superensemble.

Is it possible that this difference could lead to different conclusions about ensemble methods? To address this, the forecasting skills of MLRwr and SVDwr are compared with Ec, whose RMS and R are shown

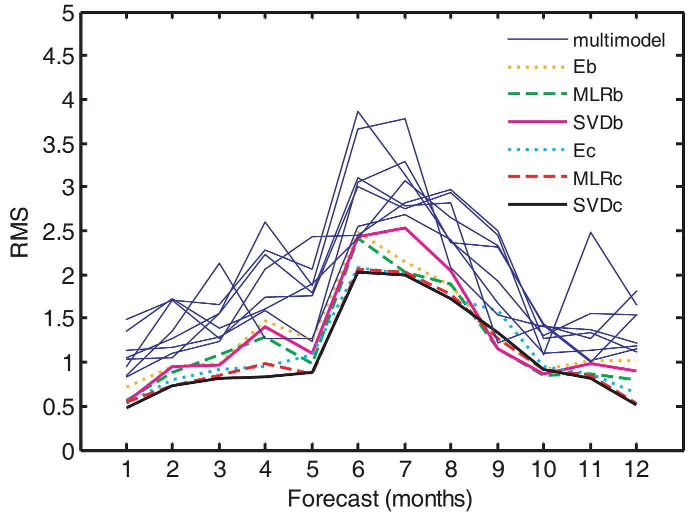

Fig. 4. RMS error of precipitations comparing the SVD superensemble (thick solid lines) to the individual models (thin solid lines), the ensemble mean (doted lines), and the MLR superensemble (dashed lines) for two data input formats during the forecast period (units: mm d^{-1}).

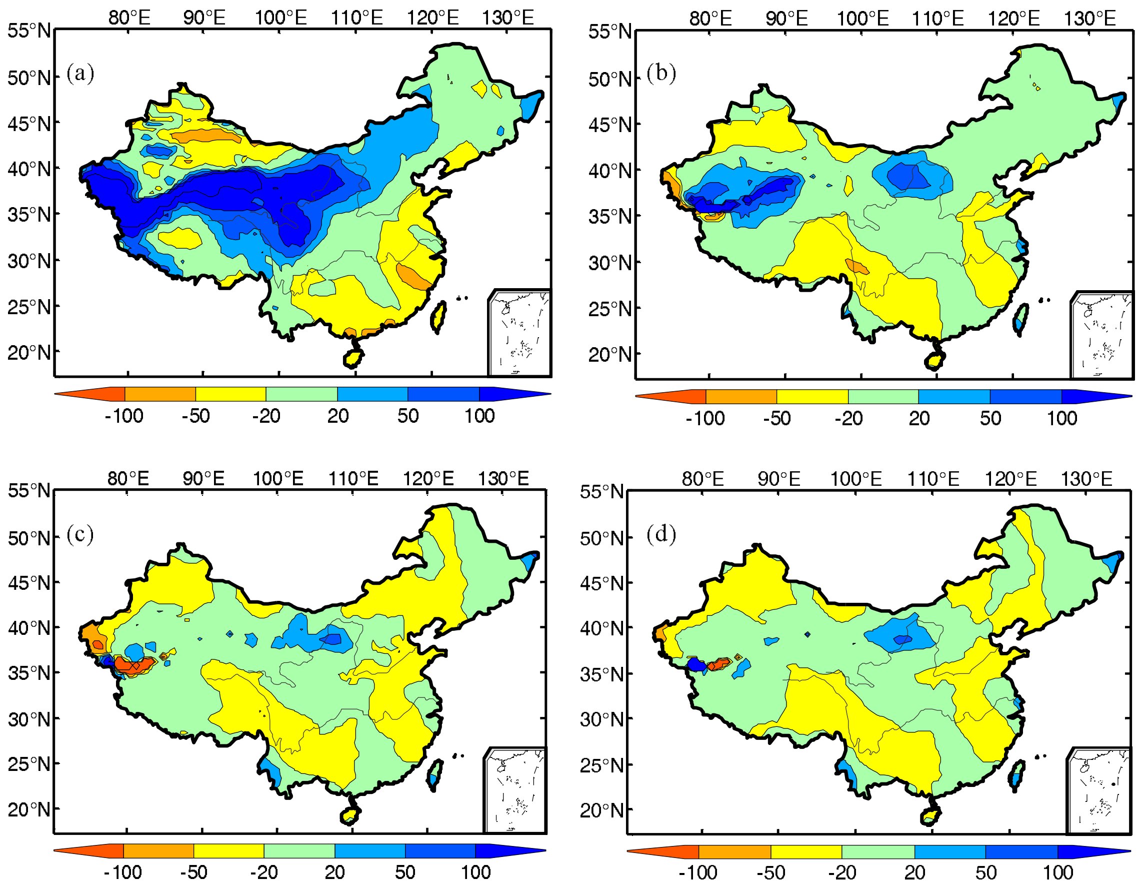

Fig. 5. Bias percentages of summer precipitations forecasted by different ensemble methods in China, 1998 (unit: %). (a) Eb; (b) Ec; (c) MLRc; (d) SVDc.

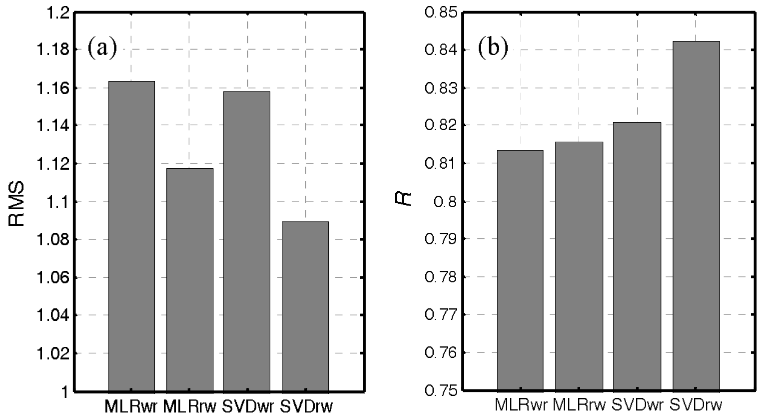

Fig. 6. Mean RMS and R of precipitations of the superensemble forecasts during the forecast period in China, where "rw" represents removing the seasonal mean and then calculating the weights; and "wr" represents calculating the weights and then removing the seasonal mean (units for RMS: mm d^{-1}). (a) RMS; (b) R.

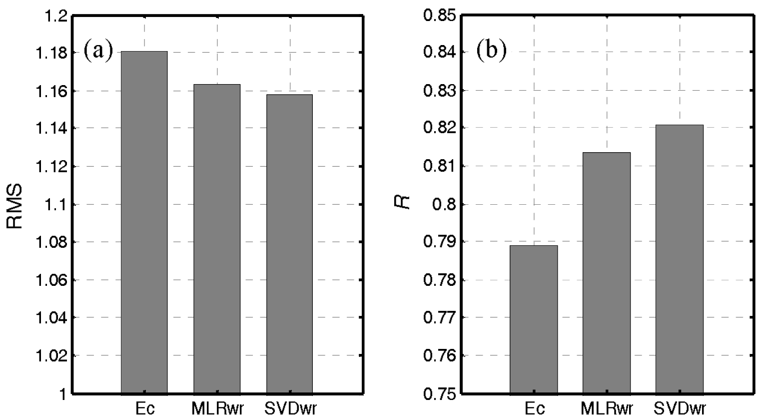

Fig. 7. Mean RMS and R of precipitations of Ec, MLRwr and SVDwr during the forecast period in China (units for RMS: mm d^{-1}). (a) RMS; (b) R.

in Fig. 7. Both MLRwr and SVDwr perform better than Ec. According to this study, therefore, it is found that the order in which weights are obtained could influence ensemble ability to a certain extent, but this might not be the main reason for disagreement presented in previous studies (Kharin and Zwiers, 2002; Krishnamurti et al., 2000).

5. Summary

Eight model outputs from 20C3M were used to study the forecasting skills of multimodel ensembles for precipitations in China, 1998. Different ensemble methods were compared with individual models. Moreover, the effects of different formats of input data on superensemble forecasting were analyzed. Finally, the order in which weights were derived was also considered.

Forecasting skills of superensembles were superior to those of all individual models and the simple ensemble mean. Different formats of input data had obvious effects on forecasting skills. Superensemble forecasts based on bias-corrected data performed better than those of original data. In contrast, the SVD superensemble was slightly more sensitive to the format of input data. Therefore, an interesting phenomenon occurred in that the SVD superensemble based on bias-corrected data performed better than the MLR superensemble, although a performance to the contrary was shown for the original data. The cause of this needs to be studied further in the future. In addition, the order in which weights were derived influenced slightly the forecasting skills, with it being optimal to remove the seasonal mean first, before calculating the weights. It is worthwhile noting that different performance levels of ensemble methods may be present in different regions. For a superensemble, it is important that more valuable information of individual models is obtained during the training time. How to calculate weights more suitably in individual models for the forecast period is the key to improving forecasting skills.

Acknowledgements. We gratefully acknowledge the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the Climate Research Unit (CRU) for providing simulation datasets and observation data, respectively. This study was supported by the Major State Basic Research Development Program of China (Grant No. 2006CB400503) and National Key Technology Research and Development Program (Grant No. 2006BAC02B04).

REFERENCES

Déqué, M., 1997: Ensemble size for numerical seasonal forecasts. *Tellus*, **49A**, 74–86.

Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislochy, 2000: Model consensus. *Wea. Forecasting*, **15**, 571–582.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, UK, 248pp.

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

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.

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

Krishnamurti, T. N., and Coauthors, 2001: Real-time multianalysis-multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.*, **129**, 2861–2883.

Krishnamurti, T. N., and Coauthors, 2003: Improved skill for the anomaly correlation of geopotential heights at 500 hPa. *Mon. Wea. Rev.*, **131**, 1082–1102.

Leith, C, E., 1974: Theoretical skill of Monte Carto forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The new ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

Mu, M., W. Duan, and J. Wang, 2002: The predictability problems in numerical weather and climate prediction. *Adv. Atmos. Sci.*, **19**, 191–204.

Pavan, V., and J. Doblas-Reyes, 2000: Multimodel seasonal hindcasts over the Euro-Atlantic: Skill scores and dynamic features. *Climate Dyn.*, **16**, 611–625.

Peng, P., A. Kumar, H. van den Dool, and A. G. Barnston, 2002: An analysis of multimodel ensemble predictions for seasonal climate anomalies. *J. Geophys. Res.*, **107**(D23), 4710, doi: 10.1029/2002JD002712.

Stephenson, D. B., and F. J. Doblas-Reyes, 2000: Statistical methods for interpreting Monte Carlo ensemble forecasts. *Tellus*, **52A**, 300–322.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at the NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP: The breeding method. *Mon. Wea. Rev.*, **125**, 3297–3318.

Yun, W. T., L. Stefanova, and T. N. Krishnamurti., 2003: Improvement of the superensemble technique for seasonal forecasts. *J. Climate.*, **16**, 3834–3840.

Yun, W. T., L. Stefanova, A. K. Mitra, T. S. V. Vijaya Kumar, W. Dewar, and T. N. Krishnamurti, 2005: A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus*, **57A**, 280–289.