# A New Way to Improve Seasonal Prediction by Diagnosing and Correcting the Intermodel Systematic Errors

ZONGJIAN KE

*Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, and Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*

PEIQUN ZHANG

*Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing, China*

WENJIE DONG

*State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China*

LAURENT LI

*Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing, China, and Laboratoire de Météorologie Dynamique, CNRS, UPMC, Paris, France*

(Manuscript received 16 June 2008, in final form 12 December 2008)

ABSTRACT

Seasonal climate prediction, in general, can achieve excellent results with a multimodel system. A relevant calibration of individual models and an optimal combination of individual models are the key elements leading to this success. However, this commonly used approach appears to be insufficient to remove the intermodel systematic errors (IMSE), which represent similar error properties in individual models after their calibration. A new postprocessing method is proposed to correct the IMSE and to increase the prediction skill. The first step consists of carrying out a diagnosis on the calibrated errors before constructing the multimodel ensemble. In contrast to previous studies, the calibrated errors here are treated directly as the investigation target, and temporal correlation coefficients between the calibrated errors and other meteorological variables are calculated. In the second stage, mathematical and statistical tools are applied in an effort to forecast the IMSE in individual models. Then, the IMSE are removed from the calibrated results and the new corrected data are used to construct the multimodel ensemble. The hindcast of the European Union–funded Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER) multimodel system is used to test the method. The simulated Southern Oscillation index is used to diagnose and to correct the calibrated errors of the simulated precipitation. The prediction qualities of the corrected data are assessed and compared with those of the uncorrected dataset. The results show that it is feasible to improve seasonal precipitation prediction by forecasting and correcting the IMSE. This improvement is visible not only for the individual models, but also for the multimodel ensemble.

## 1. Introduction

For seasonal climate prediction, a multimodel ensemble average built over several general circulation models from different operational centers is generally more capable than the individual forecasts (Palmer et al. 2000; Pavan and Doblas-Reyes 2000). Hagedorn et al. (2005) demonstrated that the basic idea of the multimodel concept is to take into account a number of independent and skillful models with the hope of achieving better coverage of the whole possible climate phase space. With idealized and real-world cases, they demonstrated that better coverage improves the prediction and the improvement can be mainly attributed to greater consistency and a higher reliability of the multimodel over the long term. Many studies have

*Corresponding author address:* Dr. Zongjian Ke, National Climate Center, CMA, Hai Dian District, Beijing, 100081, China.
E-mail: kezj@cma.gov.cn

DOI: 10.1175/2008MWR2676.1

concentrated on the combination of different single models in order to achieve the maximum consistency and to obtain a more reliable forecast system (Krishnamurti et al. 1999, 2000; Yun et al. 2003, 2005; Doblas-Reyes et al. 2005; Barnston et al. 2003). Doblas-Reyes et al. (2005) compared the forecast skills of both multiple linear regression (MLR) and simple multimodel mean (SMM) ensembles in terms of probability scores and determined that much longer time series and colinearity reduction between the single models was required for the MLR ensemble to achieve a significant accuracy improvement. To reduce the colinearity among the single models, principal component analysis and empirical orthogonal function methods are often used to produce new datasets for the construction of the multimodel ensemble, but the SMM still remains one of the best methods (Doblas-Reyes et al. 2005; Yun et al. 2005). Yoo and Kang (2005) also indicated that the skill of a multimodel composite appears to depend on numerous factors: the ensemble size of the individual models, the number of participating models, the particular spatial patterns, and the capability of the combined models. In addition, together with the multimodel ensemble, Bayesian methods are also reported to have the potential to improve climate prediction skill (Coelho et al. 2004, 2006; Stephenson et al. 2005).

Although the multimodel ensemble shows some advantages over single models, it does not help very much to reduce errors with similar patterns among the single models (Hagedorn et al. 2005). In fact, the commonly used calibration procedure with bias and variance correction cannot remove intermodel systematic errors (IMSE), which represent similar error properties among individual models. If simulated errors were stochastic after the calibration, the multimodel (either SMM or MLR) would be an appropriate approach to removing them by increasing the number of models; however, this is far from the reality as errors from individual models are often not random. Rather, they are systematic behaviors of individual models. It is obvious that the multimodel ensemble cannot remove such IMSE because they are often related to the nonlinearity of the climate system. This issue was largely ignored in previous studies, and no substantial work on the IMSE can be found in the literature, although many studies have addressed systematic errors in individual models (Feddersen et al. 1999; Yang and Anderson 2000; Kang et al. 2004; Guldberg et al. 2005) or on the multimodel ensemble with statistical downscaling approaches (Feddersen et al. 2005; Pavan et al. 2005; Gutiérrez et al. 2005; Dolores Frías et al. 2005; Díez et al. 2005). Therefore, we concentrate our efforts on the IMSE and error diagnosis is performed in the individual models and in the ensemble forecast.

Because the IMSE cannot be removed with a multimodel ensemble, a correction technology based on error diagnosis is used to reduce the IMSE. In this paper, the IMSE are regarded as a direct objective, and therefore the predictands are converted to the IMSE, which differs from the methods used in previous studies. The key element is to diagnose and forecast the IMSE in an efficient manner. When the IMSE in individual models is predicted accurately, an improvement of the prediction skill can be obtained. It is important to choose an appropriate diagnostic factor for different simulated meteorological parameters, which is helpful to predict and correct the IMSE.

A description of the dataset used is given in section 2. The method and corresponding measurement tools are presented in section 3. The results of error diagnosis and correction of the simulated precipitation are presented in section 4. The conclusions and discussion are summarized in section 5.

## 2. Data

The Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER) hindcast datasets are chosen to study the precipitation errors and to construct the multimodel ensemble. The DEMETER system (Palmer et al. 2004) comprises seven state-of-the-art global coupled ocean–atmosphere models, run under a common protocol. Each hindcast has four initialization dates, 1 February, 1 May, 1 August, and 1 November, with a duration of 6 months and an ensemble of nine members. Because our focus is on the seasonal predictability, monthly-mean precipitation data are first converted into a seasonal-mean format. Given that the length of the seasonal forecast is not uniform for the seven models, only the models of the European Centre for Medium-Range Weather Forecasts (ECMWF), Météo-France (MF), and the Met Office (UKMO), which have the same longer period (up to 43 yr), are chosen for the period from 1959 to 2001. An ensemble mean of nine members is first performed for the individual models, respectively. The climate variable of interest is the precipitation rate, which is important for societal applications. Precipitation is also an integrated indicator of model performance. The Climate Prediction Center Merged Analysis of Precipitation (CMAP) dataset (Xie and Arkin 1997; Chen et al. 2002, 2003, 2004), downloaded from the official public Web site (ftp://ftp.cpc.ncep.noaa.gov) is used for verification. These data, obtained from both simulations and observations, are first interpolated to a common $2.5° \times 2.5°$ horizontal grid through use of a bilinear interpolation.

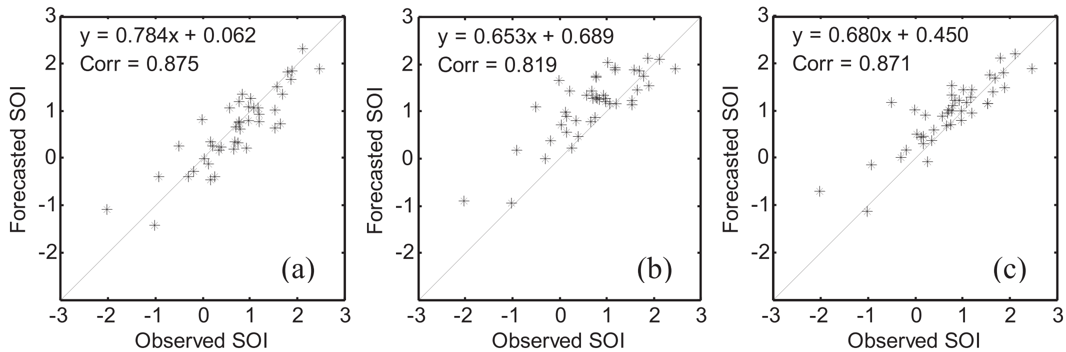


FIG. 1. Scatterplots of the forecast and observed SOI in DJF over the 43 yr: (a) ECMWF, (b) MF, and (c) UKMO.

Given the exploratory character of this study in reducing the IMSE of precipitation in multimodel ensemble, we propose, with some subjectivity, to use the sea surface temperature (SST) to improve forecast skill. The idea behind this choice is that the SST is an important parameter that depicts the interaction between ocean and atmosphere. However, the simulated SST data cannot be obtained from the ECMWF official Web site. Because there is a strong relationship between the sea level pressure (SLP) and SST, the SLP is considered as a diagnostic factor in our study.

The El Niño–Southern Oscillation is one of the most prominent sources of interannual variations in weather and climate around the world. The Southern Oscillation index (SOI), obtained as the Tahiti–Darwin SLP difference, is used to represent the reverse variation of pressure between the eastern Pacific Ocean and the tropical Indian Ocean, which is a global-scale teleconnection pattern in the atmosphere. It has been indicated that the SOI has a dominant effect on seasonal precipitation throughout the tropical Pacific (Trenberth and Caron 2000). Moreover, the correlation patterns of seasonal precipitation with the SOI are strongest in December–February (DJF). Therefore, the simulated SOI from DEMETER is chosen to investigate the variation of the simulated precipitation errors in DJF.

## 3. Method

### a. Initial calibration

Because the main shortcoming of single-model predictions is the lack of reliability that results from systematic biases in individual models, a calibration procedure has to be applied to individual models to remove or reduce such biases. The calibration procedure can be realized through the utilization of observations and can be considered as a way of obtaining predictions with average statistical properties similar to those of the reference dataset. A simple inflation of the model variance is employed to adjust the simulated spread to the observed one.

An initial calibration on the model outputs consists of performing a variance inflation and shifting the mean values to make the average statistical properties of the models close to those of the observations. This calibration is carried out before the diagnosis of the IMSE of the precipitation. The expression for this is as follows:

$$F_{\mathrm{it}}^{\mathrm{cl}} = \overline{O} + (F_{\mathrm{it}} - \overline{F_i})(S_o/S_F), \tag{1}$$

where $F_{\mathrm{it}}$ is the $i$th model forecast in the $t$th time, $\overline{F_i}$ represents the climatological forecast of the $i$th model, $\overline{O}$ is the climatological observation, and $S_o$ and $S_F$ represent the standard deviation of the observation and simulation, respectively. The forecast errors after the calibration may therefore be written as follows:

$$\mathrm{Err}_{\mathrm{it}}^{\mathrm{cl}} = F_{\mathrm{it}}^{\mathrm{cl}} - O_t, \tag{2}$$

where $O_t$ is the $t$th time observation. As mentioned above, $\mathrm{Err}_{\mathrm{it}}^{\mathrm{cl}}$ comprises mainly the IMSE that are unlikely to be removed by the calibration process or by the multimodel ensemble.

### b. Correction of the IMSE

The IMSE are the main target of our study, and we need to find a relevant correction to reduce errors that are persistent after the initial calibration. We cannot elucidate all physical causes for such errors; we can, however, apply statistical methods to quantify the influencing factors.

The performance of all three models before carrying out any diagnosis in simulating the SOI in DJF is shown in Fig. 1. A generally good performance for the SOI can be found in individual models with correlation coefficients higher than 0.8. The distribution of scattering points indicates that the simulation of the SOI in individual models is steady. The SOI is therefore confirmed as a good choice to make a first diagnosis, and we calculate the temporal correlation coefficients between the calibrated precipitation errors and the simulated SOI. We also need to define a threshold for the correlation coefficient to be statistically significant.

The threshold is fixed at 0.301 in our study, which corresponds to a confidence level of 95% for a $t$ test. Regions with temporal correlation coefficients of larger than 0.301 in the individual models are eligible for additional error correction. As an exploratory study, we want to use a correction method that is as simple as possible. A linear regression is thus performed between the precipitation errors and the simulated SOI during the training period. Therefore, the precipitation errors within the forecast phase can be obtained with a simulated SOI derived from models in the forecasting time and the regression coefficients calculated in the training time. A cross-validation method is adopted in the process of error correction (Michaelsen 1987). Each year is successively withheld, and the remaining 42 yr are used to calculate the regression coefficients. Last, the obtained precipitation errors are removed from the calibrated precipitation, and the corrected precipitation can be written as follows:

$$F_{\mathrm{it}}^{\mathrm{cr}} = F_{\mathrm{it}}^{\mathrm{cl}} - \mathrm{Err}_{\mathrm{it}}^{\mathrm{cr}}, \tag{3}$$

where $\mathrm{Err}_{\mathrm{it}}^{\mathrm{cr}}$ and $F_{\mathrm{it}}^{\mathrm{cr}}$ are the calculated forecast error and the corrected forecast of the $i$th model in the $t$th time, respectively.

In an operational seasonal prediction system, the statistics, such as the linear regression, for evaluating the IMSE can be constructed through previously calibrated model outputs. The diagnostic factors, such as the simulated SOI, in the forecasting phase can directly serve as predictors to forecast the IMSE.

It is important to note that the threshold of 0.301 is applied in our study to the temporal correlation coefficient calculated over the whole period to keep the studied regions unchanged in the cross-validation process. This arrangement induces a slight variation in the estimation of the confidence level of the correlation coefficients during the training time, but it is believed to have no significant impact on our study. Furthermore, this approximation can be avoided in an operational system, as mentioned earlier. The operation of error diagnosis and correction is carried out for each grid point separately.

### c. Validation tools

Forecast quality and capability assessment can be achieved through different measurement tools that emphasize different aspects of the forecast (Jolliffe and Stephenson 2003). The anomaly correlation coefficient (ACC), which assesses the relationship between two meteorological variables, is a good measure of the phase errors. It is used not only to diagnose the calibrated precipitation errors, but also to measure the skill of different forecasts. The ACC can be formally written as

$$\mathrm{ACC} = \frac{\sum (F - \overline{F})(O - \overline{O})}{\sqrt{\sum (F - \overline{F})^2}\sqrt{\sum (O - \overline{O})^2}}, \tag{4}$$

where $F$ and $O$ are the forecast and observation, respectively, and $\overline{F}$ and $\overline{O}$ are their temporal averages.

The Brier score (BS) is a quadratic measure of errors for probabilistic forecasts (Brier 1950). It plays a similar role to the root-mean-square in deterministic forecasts. The average BS for a set of $n$ forecasts is defined as

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2, \tag{5}$$

where $f_i$ is the forecast probability for the $i$th forecast and $o_i$ is the $i$th outcome of the observation probability.

The Brier skill score (BSS) is defined as (Wilks 1995; Toth et al. 2003)

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{ref}}}. \tag{6}$$

This quantity represents the level of improvement of the Brier score relative to that of a reference forecast strategy $\mathrm{BS}_{\mathrm{ref}}$. It is a positively oriented accuracy measurement, meaning that larger BSS values indicate better forecasts, and vice versa. The most widely used reference strategy for calculating the BSS is that of "climatology," in which the climatological probability of the forecast variable is issued perpetually (Mason 2004). Climatology is an appealing reference strategy because it is intended to provide an indication of whether the forecasts are better than having no forecast information at all, apart from knowledge of the historical likelihood of the event. Given that the BSS can be used in multievent situations, the precipitation probability is divided into three categories—above normal, normal, and below normal—with similar numbers of years in each (14, 15, and 14, respectively) over the 43-yr period of interest.

## 4. Results

The temporal correlation coefficient is used to illustrate the IMSE in different models. Figure 2 displays maps of the temporal correlation coefficients between the calibrated precipitation errors and the simulated SOI during DJF for the three individual models and the simple multimodel (SMM) ensemble, respectively. It is

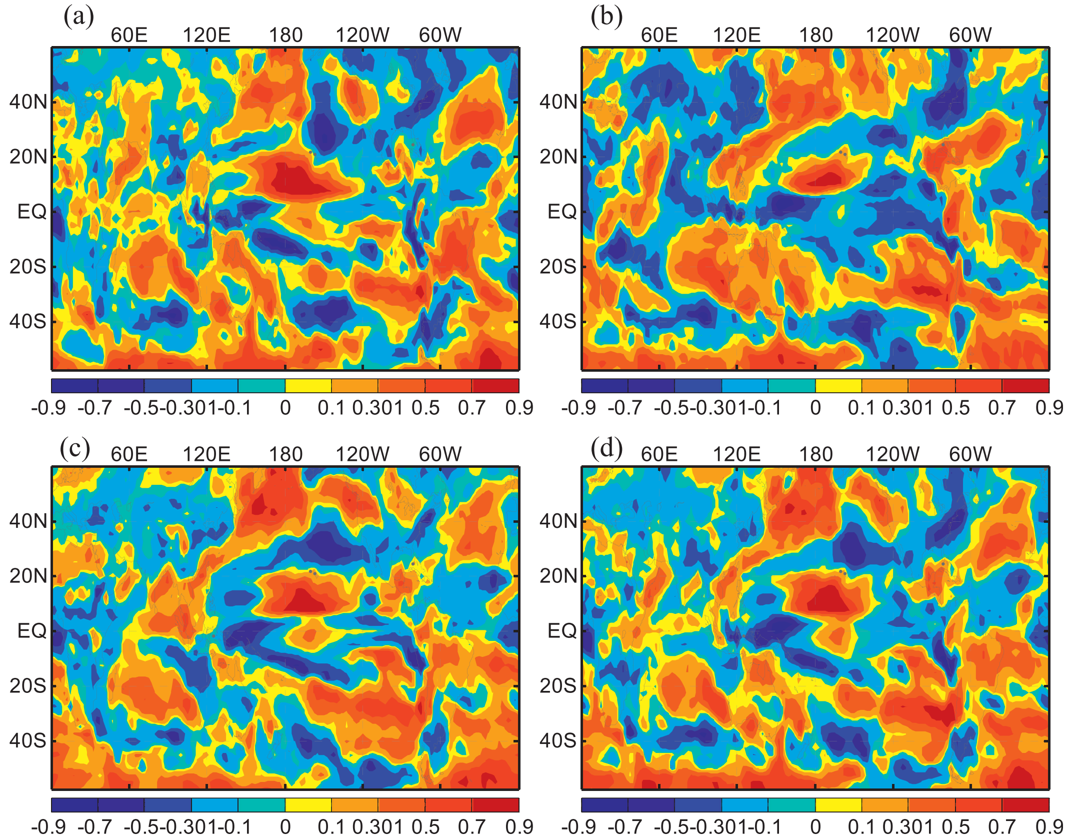


FIG. 2. Temporal correlation coefficients between the calibrated precipitation errors and the simulated SOI during DJF over the 43 yr: (a) ECMWF, (b) MF, (c) UKMO, and (d) SMM.

clear that a strong similarity exists among the three models. Negative correlation is found over the tropical Pacific, especially in the western and eastern parts. The northeast Pacific (30°N, 150°W) gives consistently negative correlations, as does the Southern Ocean at 40°S from the Indian Ocean to the middle of the Pacific. The subtropical zone of the North Atlantic Ocean also exhibits a negative correlation. Positive correlation is consistently present over a few oceanic domains: the subtropics of the Southern Ocean at about 30°S, the high latitudes of the Southern Ocean at about 50°S, the subtropics of the North Pacific at 10°N, the subtropics of the Atlantic at about 30°N, and the high latitudes of the North Pacific. Although there is a slight difference in the location and intensity of the high correlation centers among different models, the geographical shape of the correlation coefficients is generally similar. Furthermore, we can see in the following (left panels of Fig. 3) that most regions with a high correlation between the precipitation errors and the SOI are also regions in which simulated precipitation is less similar to the observed precipitation. These IMSE are indeed related to the nonlinearity (such as an asymmetric response for an equal-amplitude SST anomaly with opposite signs), which has a very similar manifestation among different models.

SMM with equal weighting are used for the ensemble forecast, and the corresponding temporal correlation distribution is presented in Fig. 2d. The high-correlation regions are still persistent, which means that the multimodel ensemble is hardly capable of removing the similar properties of the calibrated precipitation errors among various models, which is in accord with the results of a previous study (Hagedorn et al. 2005). A similar conclusion can be drawn from the MLR ensemble proposed by Krishnamurti et al. (1999).

After correction of the IMSE, the results (not shown) clearly demonstrate that the precipitation errors no longer have high correlations with the SOI. Figure 3 displays the temporal ACC between the simulated and observed precipitation during DJF in each individual model and in the ensemble forecast. Figures 3a–d show the results of the calibrated data, and Figs. 3e–h depict the results of the corrected data. In the regions in which the IMSE are corrected, such as the northwestern and central Pacific, the North Atlantic, and the southeastern Pacific in the Southern Hemisphere, the temporal

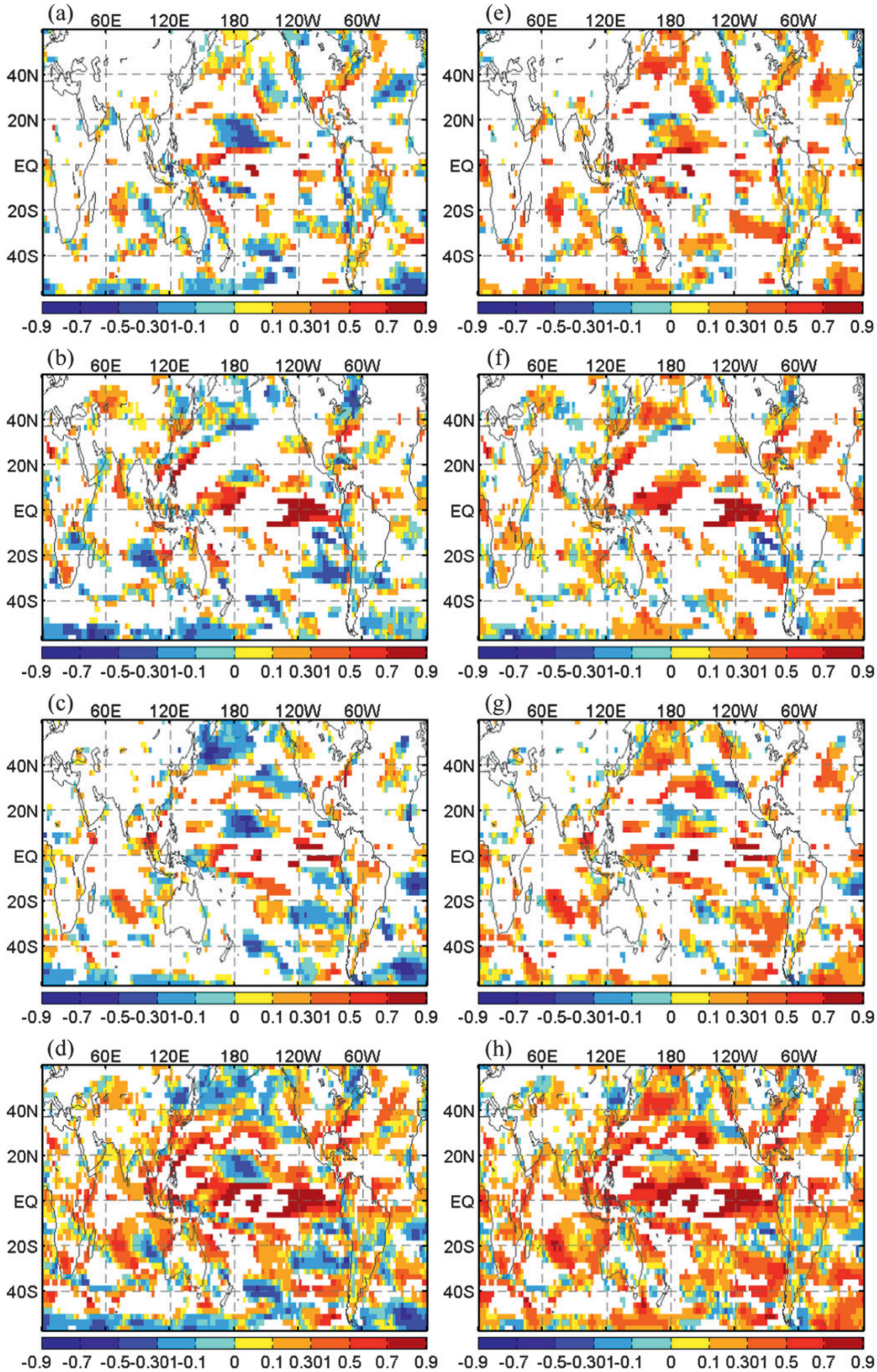


FIG. 3. Temporal ACC between the simulated and the observed precipitation in the corrected regions during DJF for (a) ECMWF, (b) MF, (c) UKMO, and (d) SMM, before correction. (e)–(h) As in (a)–(d), but with the corrected data.

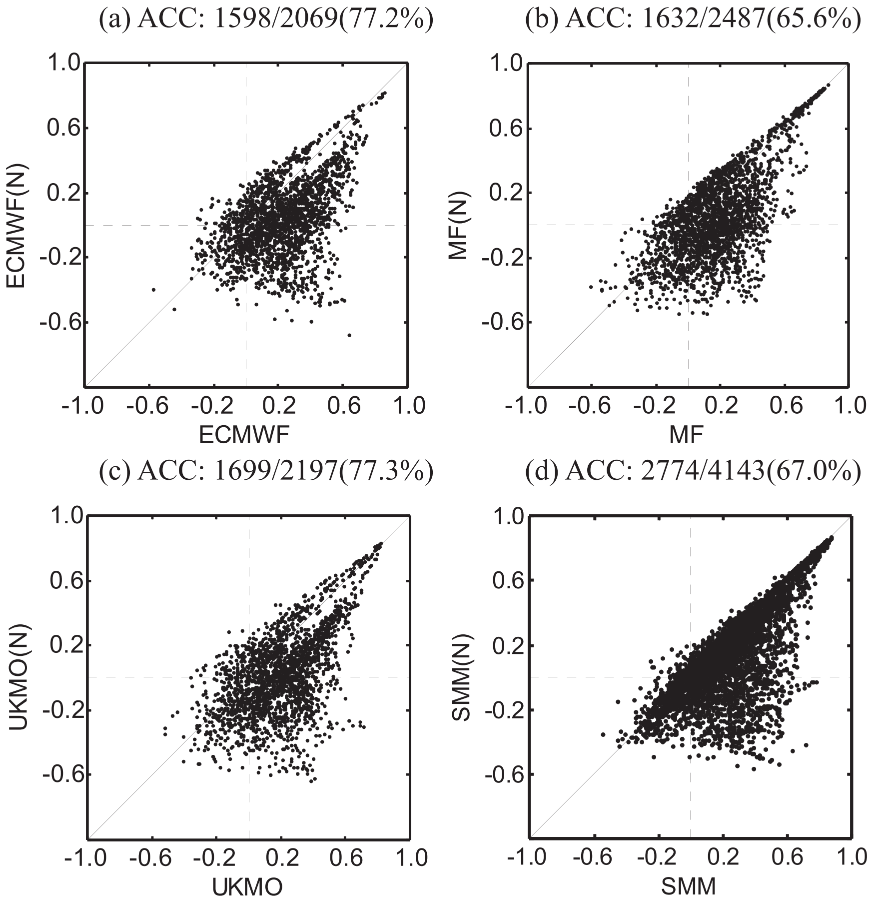


FIG. 4. Scatterplots of the temporal ACC between the simulated and observed precipitations in the corrected regions during DJF: (a) ECMWF, (b) MF, (c) UKMO, and (d) SMM. The "N" in the parentheses of the ordinates represents the results before correcting errors. The numbers on top of (a)–(d) correspond to the number of times the score of the system in the abscissa has a larger value than the one in the ordinate (to the left of the slash), the total number of cases (to the right of the slash), and, in parentheses, the percentage.

anomaly correlation coefficients of the three models increase significantly. Moreover, in the high-latitude regions of the Southern Hemisphere, similar improvements can be found over oceans after the IMSE correction. In comparison with the multimodel ensemble result based on the calibrated data, the corrected ensemble forecast also shows significant improvements in the individual models, especially over oceans.

Figure 4 shows scatterplots of the temporal ACC in the corrected regions. A general improvement can be seen in the individual models and in the ensemble forecast. The amelioration is particularly important in the ECMWF and UKMO models, whereas the improvement is less important in the MF model. This is because the corrected regions are mainly over oceans in the ECMWF and UKMO but are also over the Eurasian continent in the MF. Our IMSE correction, through a simple linear regression between the precipitation errors and the SOI, may not be entirely appropriate over continents. Figure 5 shows the mean temporal ACC averaged over different regions. It is clear that a general improvement is present for all models and all regions.

Figure 6 depicts the BSS for the individual models and the ensemble forecast based on the calibrated and corrected datasets in the corrected regions. Obvious differences exist in the BSS after the IMSE correction. The corrected BSS values are generally higher than the uncorrected values, which indicates that the probabilistic skill scores of the forecast can also be improved with IMSE correction. The SMM ensemble forecast demonstrates a performance that is similar to that of the individual models.

The IMSE of the precipitation in other seasons are also studied in the same manner, as presented here for DJF. The results are not shown for the sake of conciseness, but the same process of diagnosing, forecasting, and

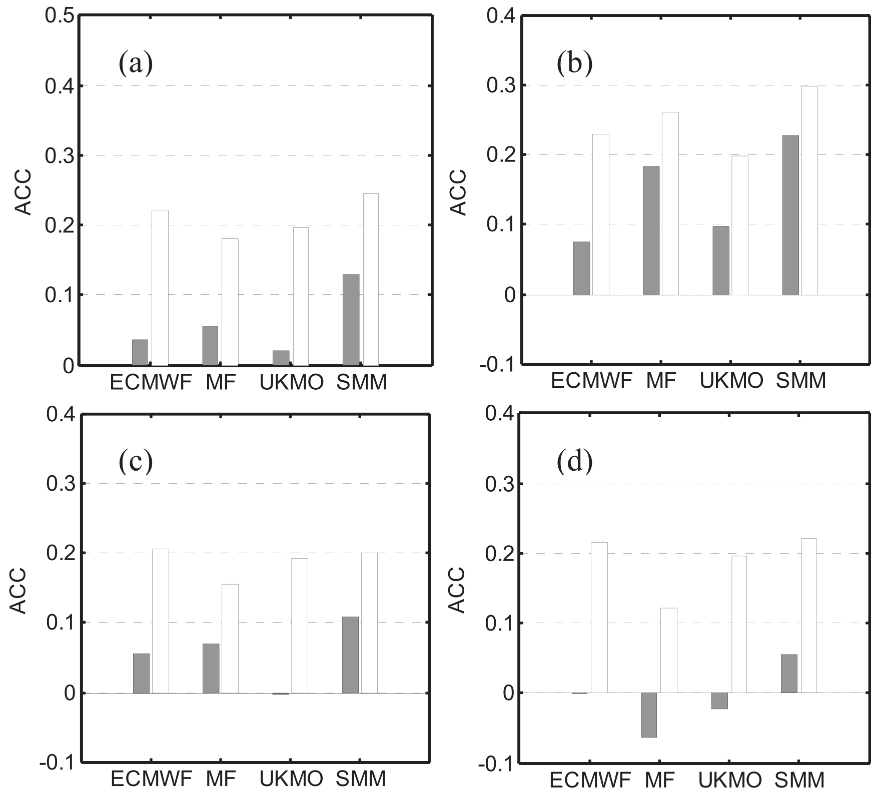


FIG. 5. The mean temporal ACC calculated by the calibrated and corrected datasets over different corrected regions during DJF: (a) global, (b) tropics (20°S–20°N), (c) extratropics in the Northern Hemisphere (20°–60°N), and (d) extratropics in the Southern Hemisphere (60°–20°S). The gray and white bars represent the calibrated and corrected forecasts, respectively.

removing the IMSE, followed by a correction of calibrated precipitation errors, can improve the forecast skills for other seasons.

Last, we point out that the simulated SOI is not the only variable that is useful for reducing the IMSE. In some regions across the globe, the SOI is probably an unsuitable diagnostic factor for the IMSE of precipitation. It is then feasible to use other meteorological variations to diagnose the calibrated errors and to correct the IMSE. In addition, this method of reducing the IMSE is not in conflict with the previous studies of error correction through different statistical methods, including statistical downscaling (Feddersen et al. 1999; Yang and Anderson 2000; Kang et al. 2004). It can serve as a preprocessing operation before carrying out other statistical correction methods.

## 5. Conclusions and discussion

The multimodel ensemble has been regarded for a long time as a pragmatic approach to reduce the simulation errors of single models. However, multimodel ensembles hardly show improvement for IMSE. Some statistical correction methods have been proposed to reduce these systematic errors in previous studies, but little attention has been paid to the systematic errors themselves and little work about IMSE diagnosis relevant to multimodel ensembles can be found. Therefore, we propose a method of error diagnosis and forecast to reduce the IMSE.

Our study differs from previous ones, because the IMSE is the direct forecast object, and the main goal to improve seasonal prediction is assessed by a proper forecasting of the IMSE. An error diagnosis and a forecast procedure were performed after calibrating the simulation data. Error correction was then applied to the single models through a simple mathematical manipulation. A multimodel ensemble could thus be reconstructed using the corrected data. Let us note that the applied method to correct the IMSE is not in conflict with previously reported statistical correction methods. Moreover, this diagnosis and correction can be used as a preprocessing step for statistical correction or downscaling methods.

As an example, the simulated SOI was used to diagnose and correct the calibrated precipitation errors. The results showed that a strong similarity among models existed in the geographical patterns of the correlation

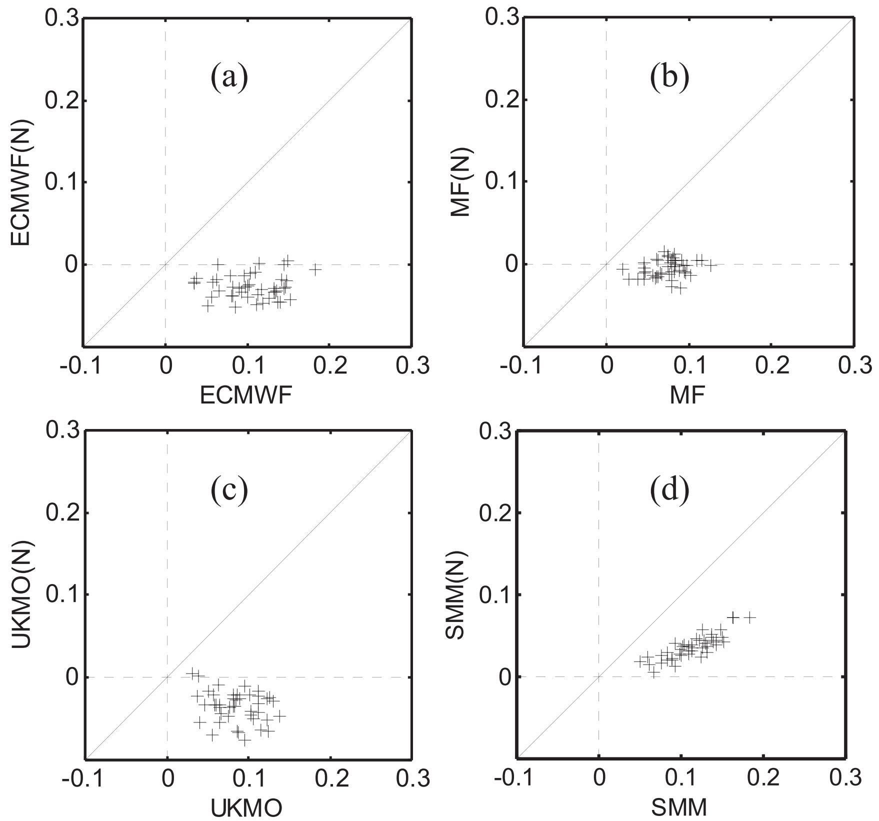


FIG. 6. BSS for the individual models and the ensemble forecast based on the calibrated (represented by the "N") and corrected data in the corrected regions during DJF: (a) ECMWF, (b) MF, (c) UKMO, and (d) SMM.

calculated between the calibrated systematic errors of precipitation and the simulated SOI in different DEMETER models. The forecast skills of individual models and their ensemble were all improved through the forecast and correction of the IMSE. It was also found that the improvement of precipitation forecasts over oceans was more significant than over continents, which indicates that the simulated SOI is a more appropriate variable to diagnose the calibrated precipitation errors over oceans. It is speculated that different meteorological variations in different regions may involve different variables that are more relevant than others and can be used to diagnose and correct the IMSE. Such studies will be pursued in the future to further improve the forecast skills of individual models and their ensembles in seasonal climate prediction. We are also planning to use a similar approach with other atmospheric oscillation indexes, such as the North Atlantic Oscillation, to perform further correction of the IMSE. We will also extend our study to temperatures.

*Acknowledgments.* We gratefully acknowledge the ECMWF and CMAP for providing the DEMETER-coupled model datasets and the observation data. This study was supported by the National Key Technology R&D Program (Grant2006BAC02B04) and the Major State Basic Research Development Program of China (Grant 2006CB400503). Comments from the anonymous reviewers helped to clarify the manuscript.

REFERENCES

Barnston, A. G., S. J. Mason, L. Goddard, D. G. Dewitt, and S. E. Zebiak, 2003: Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.,* **84,** 1783–1796.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78,** 1–3.

Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global land precipitation: A 50-year monthly analysis based on gauge observations. *J. Hydrometeor.,* **3,** 249–266.

——, ——, ——, ——, and T. M. Smith, 2003: Reconstruction of the oceanic precipitation from 1948 to the present. Preprints, *14th Symp. on Global Change and Climate Variations*, Long Beach, CA, Amer. Meteor. Soc., J6.1. [Available online at http://ams.confex.com/ams/pdfpapers/70083.pdf.]

——,——, ——, ——, and ——, 2004: Verifying the reanalysis and climate models outputs using a 56-year data set of reconstructed global precipitation. Preprints, *14th Conf. on*

*Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., J6.1. [Available online at http://ams.confex.com/ams/pdfpapers/70083.pdf.]

Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Climate,* **17,** 1504–1516.

——, D. B. Stephenson, F. J. Doblas-Reyes, M. Balmaseda, A. Guetter, and G. J. van Oldenborgh, 2006: A Bayesian approach for multi-model downscaling: Seasonal forecasting of regional rainfall and river flows in South America. *Meteor. Appl.,* **13,** 73–82.

Díez, E., C. Primo, J. A. García-Moya, J. M. Gutiérrez, and B. Orfila, 2005: Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus,* **57A,** 409–423.

Doblas-Reyes, F. J., R. Hagedorn, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting-II. Calibration and combination. *Tellus,* **57A,** 234–252.

Dolores Frías, M., J. Fernández, J. Sáenz, and C. Rodríguez-Puebla, 2005: Operational predictability of monthly average maximum temperature over the Iberian Peninsula using DEMETER simulations and downscaling. *Tellus,* **57A,** 448–463.

Feddersen, H., and U. Andersen, 2005: A method for statistical downscaling of seasonal ensemble predictions. *Tellus,* **57A,** 398–408.

——, A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate,* **12,** 1974–1989.

Guldberg, A., E. Kaas, M. Deque, S. Yang, and S. Vester Thorsen, 2005: Reduction of systematic errors by empirical model correction: Impact on seasonal prediction skill. *Tellus,* **57A,** 575–588.

Gutiérrez, J. M., R. Cano, A. S. Cofiño, and C. Sordo, 2005: Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps. *Tellus,* **57A,** 435–447.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting-I. Basic concept. *Tellus,* **57A,** 219–233.

Jolliffe, I. T., and D. B. Stephenson, 2003: Introduction. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 1–12.

Kang, I. S., J. Y. Lee, and C. K. Park, 2004: Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *J. Climate,* **17,** 834–844.

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.

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

Mason, S. J., 2004: On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.,* **132,** 1891–1895.

Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.,* **26,** 1589–1600.

Palmer, T. N., C. Brankovic, and D. S. Richardson, 2000: A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Quart. J. Roy. Meteor. Soc.,* **126,** 2013–2034.

——, and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal- to-Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.,* **85,** 853–872.

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

——, S. Marchesi, A. Morgillo, C. Cacciamani, and F. J. Doblas-Reyes, 2005: Downscaling of DEMETER winter seasonal hindcasts over Northern Italy. *Tellus,* **57A,** 424–434.

Stephenson, D. B., C. A. S. Coelho, F. J. Doblas-Reyes, and M. Balmaseda, 2005: Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus,* **57A,** 253–264.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 137–163.

Trenberth, K. E., and J. M. Caron, 2000: The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation. *J. Climate,* **13,** 4358–4365.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. 1st ed. Academic Press, 467 pp.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.

Yang, X. Q., and J. L. Anderson, 2000: Correction of systematic errors in coupled GCM forecasts. *J. Climate,* **13,** 2072–2085.

Yoo, J. H., and I. S. Kang, 2005: Theoretical examination of a multi-model composite for seasonal prediction. *Geophys. Res. Lett.,* **32,** L18707, doi:10.1029/2005GL023513.

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

——, ——, 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.