Multi-Model Synthetic Superensemble Algorithm for Seasonal Climate Prediction using DEMETER Forecasts

W. T. Yun[1,2], L. Stefanova[1], A. K. Mitra[1,3], T. S. V. Vijaya Kumar[1], W. Dewar[4] and T. N. Krishnamurti[1]

[1]Department of Meteorology, The Florida State University
[2]Korea Meteorological Administration
[3]NCMRWF, New Delhi, India
[4]Department of Oceanography, The Florida State University

March, 2004

(Submitted to Tellus, DEMETER Special Issue)

Corresponding author address: Dr. W. T. Yun
Department of Meteorology
The Florida State University
Tallahassee, FL 32306 USA
E-mail: wtyun@io.met.fsu.edu

Abstract

In this paper, multi-model superensemble and a new weighted multi-model ensemble approach named synthetic superensemble for long-range climate prediction using coupled DEMETER (Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction) output is presented. Despite the continuous improvement of coupled models, they have serious systematic errors in terms of the mean, the annual cycle and the interannual variability, and consequently the predictive skill of extended forecasts remains quite low. The purpose of the synthetic superensemble approach is to improve long-range prediction skill.

The superensemble algorithm entails the division of a time line into two parts, a training phase and a forecast phase. In this technique, the different model forecasts are statistically combined during the training phase using multiple linear regression, with the skill of each ensemble member implicitly factored into the superensemble forecast. The idea of the synthetic algorithm is generating a new data set from actual data set for multiple linear regression. The synthetic data are created form the original data set by finding a consistent spatial pattern between observed analysis and the forecast data set. This procedure is a multiple linear regression problem in EOF space.

Our experiments show that the unbiased multi-model ensemble of forecasts, such as the DEMETER data, improves the prediction of spatial patterns, i.e., the anomaly correlation, but it shows poor skill of categorical forecast. Due to the removal of biases of the different models, the forecast errors of the bias corrected multi-model ensemble and superensemble are already quite small. Based on various skill measures, the forecast produced by the proposed method outperforms the other conventional forecasts.

1. Introduction

A major stumbling block to the improvement of the skill of forecast is model error, as seen in long-term (monthly or longer) simulations. All coupled models have serious systematic errors in terms of the mean, the annual cycle or the statistics of interannual variability and, in some cases, all three of these characteristics (Kirtman et al. 2003). For overcoming this problem, there are some statistical or empirical approaches. In this paper we introduce an empirical multi-model synthetic superensemble method for seasonal climate forecast using the DEMETER (Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction) coupled model output.

The ensemble approach, single or multi-model, is a relatively recent contribution to the general area of weather and climate forecasting. Most deterministic and probabilistic ensemble forecasts are produced with a single dynamical model, although sometimes a set of multi-models is used. The skill of single and multi-model ensembles has been reported by many studies (Doblas-Reyes et al. 2000, Graham et al. 2000, Palmer et al. 2000, Palmer et al., 2003). Such ensemble techniques are nowadays routinely used at operational weather forecasting centers (Molteni et al. 1996, Buizza et al. 1998, Toth and Kalnay 1997, Houtekamer et al. 1996, Stephenson and Doblas-Reyes 2000) and are also applied in seasonal timescale climate studies (Brankovic and Palmer 1997, Zwiers 1996, Pavan and Doblas-Reyes 2000, Kharin and Zwiers 2002, Peng et al. 2002). Each different model has its own variability generated by internal dynamics (Straus and Shukla 2000). As a result, the performance of a multi-model ensemble is generally more reliable than that of a single model.

The main objective of this paper is to design a multi-model ensemble for seasonal climate prediction using ocean-atmosphere coupled models. An approach to produce seasonal climate forecast using multi-models is the weighted multi-model superensemble named by Krishnamurti et al. (1999, 2000a,b, 2001, 2003). In the sense of its construction, the superensemble is a postprocessing product of multi-model forecasts. This superensemble can be used as a tool for making both deterministic and probabilistic predictions.

The superensemble algorithm entails the division of a time line into two parts, a training phase and a forecast phase. In this technique, the different model forecasts are statistically combined during the training phase using multiple linear regression, with the skill of each ensemble member implicitly factored into the superensemble forecast.

The forecast resulting from the projection of these solutions into a forecast phase has smaller errors and higher skill than most conventional models and conventional ensemble techniques. The ensemble mean assigns a weight of *1/N* to each of the *N* member models everywhere (and for all variables), regardless of their relative performance. As a result, assigning the same weight of *1/N* to some poorer models has been noted to degrade the skill of the ensemble mean. It is possible to remove the bias of models individually and to compute an ensemble mean of the bias-removed models. This too has somewhat lower skill compared to the superensemble, which carries selective weights distribution in space, multi-models, and variables.

Many enhancements of the superensemble technique have been made in past studies (Krishnamurti et al. 1999, 2000a,b, 2001, 2003, 2003; Stefanova and Krishnamurti, 2002, Yun et al., 2003) and it has been shown that this technique provides higher skill forecasts compared to all participating member models and the ensemble mean. Various studies have discussed extensively the multi-model seasonal predictions. Pavan and Doblas-Reyes (2000) combined seasonal forecasts from four different AGCMs and found minimal skill improvement. Kharin and Zwiers (2002) assessed different ways of constructing multimodel forecasts and found a disagreement with the results of Krishnamurti et al, in that their regression-improved multimodel forecast (i.e. the superensemble) performed worse than the multimodel ensemble. This discrepancy is due to the fact that in their calculations the seasonal mean is removed only after the regression coefficients are calculated, while in case of Krishnamurti's superensemble the seasonal mean is removed prior to the calculation of regression coefficients (Yun et al. 2003). Yun et al. reported skill improvement of superensemble forecast applying the Singular Value Decomposition (SVD) technique. They constructed a multiple regression model based on SVD technique for the generation of multi-model superensemble forecasts. The regression model was constructed using covariance matrices where the bias and the annual cycle were removed. For obtaining the optimal regression coefficients, the squared uncertainties of the estimated parameter are minimized by setting the small singular values to zero, based on the premise that since smaller squared uncertainties of estimated parameter explain the relative variance better, that would enhance the multi-model superensemble forecast.

Krishnamurti et al. (2003) noted that the superensemble skill during the forecast phase could be degraded if the training was executed with either poorer analysis or poorer forecasts. This indicates that the forecast will be improved when higher quality training data sets are deployed for the evaluation of the multi-model statistics. This present paper focuses on improving the seasonal-timescale climate prediction skill through the generation of synthetic data set from actual multi-model data. The synthetic superensemble algorithm is an alternative method for obtaining high quality data set for superensemble forecast and to improving the prediction skill of extended forecasts. In this study, the multi-model synthetic ensemble/superensemble prediction algorithm for deterministic seasonal climate prediction is presented.

2. Multi-model Data Set

The multi-model data set in this study used is the output of seven global coupled ocean-atmospheric models from DEMETER (Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction) project (Palmer et al., 2003). The DEMETER hindcasts were started from 1st February, 1st May, 1st August, and 1st November initial conditions. Each model was itself run in ensemble mode, based on nine different initial conditions from each start date. Each hindcast has been integrated for six months and comprises an ensemble of nine members. The multi-model data set for the period from 1987 to 2001 is evaluated in this paper. We use ECMWF reanalysis as verification data. All calculations are done using cross-validation, with each year being successively withheld from the training data set, and the remaining 14 years used for calculation of the model and observed statistics, including the computation of all 14 years excepting the one for which a forecast is performed. A complete description of the DEMETER data set can be found on the website: http://www.ecmwf.int/research/demeter.

3. Algorithm for Multi-model Synthetic Ensemble/Superensemble

Despite the continuous improvement of both dynamical and empirical models, the predictive skill of extended forecasts remains quite low. Multi-model ensemble predictions have relied primarily on empirical techniques. Such prediction methods rely on statistical relationships established from an analysis of past observations (Chang et al, 2000). This means that the multi-model ensemble prediction depends strongly on the past performance of individual member models. The prediction may be improved if higher quality data set is used for the evaluation of the multi-model bias statistics in the training phase. In this chapter, we introduce the computational algorithm for the creation of a synthetic data set for climate prediction.

In the context of seasonal climate forecasts, many studies (Krishnamurti et al., 1999, 2000a,b, 2001, 2002, 2003; Peng et al., 2002; Pavan and Dobals-Reyes, 2000; Doblas-Reyes et al., 2000; Stephenson 2000; Kharin and Zwiers, 2002; Stefanova and Krishnamurti, 2002; Yun et al., 2003; Palmer, 2003) have discussed various multi-model approaches for forecasting of anomalies, such as the bias corrected ensemble mean, eq. (1), the biased ensemble mean, eq. (2), and the superensemble forecast, eq. (3). The superensemble contains no bias since the model climatology have been considered.

$$E_c = \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{F_i}) \tag{1}$$

$$E_b = \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{O}) \tag{2}$$

$$S = \sum_{i=1}^{N} a_i (F_i - \overline{F_i}) \tag{3}$$

Where F_i is the i^{th} model forecast out of N models, $\overline{F_i}$ is the mean of the i^{th} forecast over the training period, $\overline{O}$ is the observed mean over the training period, and a_i is the regression coefficient of the i^{th} model. The difference between these approaches comes from mean status and the weights. A major aspect of the superensemble forecast is the training of the forecast data set. The superensemble prediction skill during the forecast phase could be improved when higher quality data set is available for training.

The synthetic data generating algorithm is used to obtain high quality forecast data set from actual multi-model data set for superensemble prediction. Fig.1 is a schematic chart illustrating the synthetic algorithm. The synthetic data set is generated from the original data set by finding a consistent spatial pattern between the observed analysis and the forecast data set. This procedure is a linear regression problem in EOF space. A set of created synthetic data is then used as a multi-model data set for an ensemble/superensemble forecast. The computational procedure for generating the synthetic data set is described below.

The observation data (O) and the multi-model forecast data set (F_i) can be written as a linear combination of EOFs (Empirical Orthogonal Functions), which describe the spatial and temporal variability.

$$O(x,t) = \sum_n \tilde{O}_n(t)\phi_n(x) \tag{4}$$

$$F_i(x,T) = \sum_n \tilde{F}_{i,n}(T) \cdot \varphi_{i,n}(x) \tag{5}$$

where $\tilde{O}_n(t)$, $\tilde{F}_{i,n}(t)$ and $\phi_n(x)$, $\varphi_{i,n}(x)$ are the PC (Principal Component) time series and the corresponding EOFs of the n^{th} mode for the observation and model forecast, respectively. Index i indicates a particular member model. The PCs in equations (4, 5) represent the evolution of spatial patterns during the training period (t) and the whole forecast time period (T). We can now estimate a consistent pattern between the observation and the forecast data, which evolves according to the PC time series of observation data. The procedure of finding this consistent pattern is a regression problem in EOF space. The regression relationship between the observation PC time series and a number of PC time series of individual model forecast data can be written as

$$\tilde{O}(t) = \sum_n \alpha_{i,n} \tilde{F}_{i,n}(t) + \varepsilon_{i,n}(t)\,. \tag{6}$$

With equation (6) we can express the observation time series as a linear combination of the predictor time series. To obtain the regression coefficients $\alpha_{i,n}$ the regression is performed in the EOF domain. The regression coefficients $\alpha_{i,n}$ are found such that the residual error variance $E(\varepsilon^2)$ is minimized. The errors $\varepsilon_{n,m}$ are assumed to have a mean of zero and each pair ε_n, ε_m , $n \neq m$, is assumed uncorrelated. The covariance matrix is constructed with the PC time series of each model. The principal component time series are uncorrelated with one another. For obtaining the weights, the covariance matrix is built with the seasonal cycle-removed anomaly. The covariance matrix of the PCs is given by

$$C_{n,m} = \tilde{F}'_n(t)\tilde{F}'_m(t)\,, \tag{7}$$

where the prime represents seasonal mean subtracted anomaly at a given time, and n and m indicate n^{th} and m^{th} EOF modes. Once the regression coefficients $\alpha_{I,n}$ are found, the PC time series of synthetic data set is written as

$$\tilde{F}_i^{reg}(T) = \sum_n \alpha_{i,n} \tilde{F}_{i,n}(T)\,. \tag{8}$$

The synthetic data set is now generated by reconstruction with corresponding EOFs and PCs.

$$F_i^{syn}(x,T) = \sum_n \tilde{F}_{i,n}^{reg}(T) \cdot \phi_n(x) \tag{9}$$

What is unique about the synthetic data set, which is generated from ocean-atmosphere coupled multi-model simulations, is that it minimizes the residual error variance by finding a consistent pattern from the seven member models (Fig. 1). The residual error variance (Appendix) is minimized using least square error approach.

$$\frac{\partial e^2}{\partial a} = 0 . \tag{10}$$

This generated synthetic data set is used as an input data set for a superensemble system that produces a deterministic forecast. To illustrate the performance of the synthetic superensemble, a comparison of precipitation forecasts for the month of July 2001 over the tropics (30ºS-30ºN) is shown in Fig. 2. The observed analysis is shown in the top panel. OBS, EM, SEM, SF and SSF in plots indicate observation, bias corrected ensemble mean, synthetic ensemble mean, superensemble based on SVD and synthetic superensemble forecast, respectively. Due to the removal of biases of the models, the forecast errors of the bias corrected multi-model ensemble are already quite small. Though all forecasts in Fig. 2 show almost the same spatial patterns and almost same amplitude, the ensemble mean forecast generally does not capture the strength of amplitude very well.

4. On SVD based Superensemble Prediction System

For computing the synthetic superensemble forecasts, the singular value decomposition (SVD) method is applied to the computation of the regression coefficients for a set of synthetic data. The regression model is constructed using covariance matrices where the bias and the annual cycle are removed. The SVD of the covariance matrix C is its decomposition into a product of three different matrices. The covariance matrix C can be rewritten as a sum of outer products of columns of a matrix U and rows of a transposed matrix V^T, represented as $C = UWV^T$. U and V are matrices that obey the orthogonality relations. W is an diagonal matrix, which contains real positive singular values arranged in decreasing magnitude.

The pointwise regression model using the SVD method removes the singular matrix problem that can’t be entirely solved with the Gauss-Jordan elimination method. Moreover, zeroing of the small singular values gives better regression coefficients than the SVD solution where the small singular values are left nonzero (Yun et al., 2003). If the small singular values are retained as nonzero, it usually makes the residual error larger (Press et al., 1992).

The computed weights vary in space based on the models’ past performance skills and have positive and negative fractional values. Those models that exhibit a better performance during the training phase over a given region carry positive weights and those that performed poorly acquire negative weights (Krishnamurti et al. 2003).

5. Verification Metrics

The spatial AC (anomaly correlation) and RMS (root mean square) of one month lead seasonal mean anomalies are used as objective skill measures. These skill metrics describe the average magnitude of the errors and the phase errors of the forecast anomalies corresponding to the observed anomalies. Though AC is a good measure of phase error and doesn’t take bias into account (Déqué, 1997), it is possible for a forecast with large errors in magnitude to still have a good correlation coefficients. It is therefore also necessary to evaluate the phase and magnitude errors separately.

$$AC = \frac{\sum (F-\overline{F})(O-\overline{O})}{\sqrt{\sum (F-\overline{F})^2}\sqrt{\sum (O-\overline{O})^2}} \tag{11}$$

$$RMS = \sqrt{\frac{1}{G}\sum[(F-\overline{F})-(O-\overline{O})]^2} \qquad (12)$$

where the overbar denotes time average, G denotes the number of grid points, and the summation is done over space.

We also use categorical forecasts to quantify forecast skill (Brown et al. 2001). Categorical forecasts indicate whether a particular category of observations will occur, such as positive/negative precipitation anomaly. These categorical forecasts are verifiable with examination of the frequencies of occurrence of various pairs of forecasts and observations (Table 1). Some of verification measures are listed in Table 1. Probability of "Yes" or "No" detection (PODy, PODn) is estimates of the proportions of "Yes" or "No" observations that were correctly forecasted. These measure the ability of the forecasts to discriminate between Yes and No observations. The true skill score (TSS) summarizes this discrimination ability. False alarm ratio (FAR) estimates the frequency of "Yes" forecasts that did not verify. The Gilbert skill score (GSS), also known as equitable threat score (ETS) is the proportion of correct "Yes" forecasts, relative to the number of times the event was forecasted to occur, minus the fraction of correct "Yes" forecasts that would be expected to occur by chance.

6. Results of Multimodel Synthetic Ensemble/Superensemble Forecasts

This section describes and compares the skill and performance of the aforementioned ensemble, superensemble based on SVD, and synthetic ensemble/superensemble forecasts. All calculations are done with cross-validation. The forecasts of ensemble and synthetic ensemble/superensemble are demonstrated by applying them to the DEMETER models. Each model of DEMETER was itself run in ensemble mode, based on nine different initial conditions. In a large sample, the ensemble mean provides on average better skill than an individual forecast (Leith, 1974), but it represents just a part of the information contained in the ensemble (Doblas-Reyes, 2000). Kharin and Zwiers (2002) examined the ensemble mean forecast skills. They find that the skill of ensemble mean forecast is dependent on ensemble size. We assessed the skill scores of mutli-model ensemble based on our experiment.

At first, we took one run from each member model and performed monthly synthetic superensemble forecast with 168 months training. Fig. 3 illustrates 15 years (1987-2001) averaged AC and RMS skill of unbiaseed ensemble mean (EM), climatology (CLIM), unbiased synthetic ensemble mean (SEM), superensemble based on SVD (SF), and synthetic superensemble (SSF) of global, tropical (30°S-30°N) and north hemispheric (0°-60°N) precipitation forecasts for July and December. The range of AC (EM, SEM, SF, SSF) is between 0.17 and 0.32. The synthetic ensemble/superensemble forecasts show best skill in terms of AC and RMS measures. In the sence of categorical forecast of positive precipitation anomaly, the superensemble based on SVD shows less FAR than the other forecasts and has best TSS and ETS skill (Fig. 4). Even though the unbiased ensemble mean performs better than any individual model, the forecast skill is very low.

Secondly, we used all the 9 ensemble members for each model for extended seasonal superensemble forecast. The superensemble is constructed by training with 56 seasons of forecast data set. Fig. 5 shows the 1-month lead Summer(June, July, August) and Winter (December, January, February) AC and RMS of global precipitation. All multi-model forecasts (EM, SEM, SF, SSF) show better AC than individual models in most years. Averaged over 15 years, the synthetic ensemble mean shows the best AC skill scores except for Spring (March, April, May) (Fig. 6). The range of AC (EM, SEM, SF, SSF) is between 0.29 and 0.47. Fig. 7 illustrates the precipitation anomaly forecasts for Summer 2001. The observed analysis is shown in the top panel. EM, SEM, SF and SSF indicate the unbiased ensemble mean, synthetic ensemble mean, superensemble based on SVD and synthetic superensemble forecast, respectively. The anomalies of the unbiased ensemble mean and the synthetic ensemble mean are very small, due to the averaging of a large number of realizations. The synthetic superensemble forecast gives more realistic precipitation anomaly compared to both the ensemble mean and the superensemble based on SVD. The verification of categorical forecast of positive precipitation anomaly is shown in Fig. 8. Both superensemble forecasts are better than both ensemble mean forecast in terms of ETS and TSS. The FARs of both superensembles are smaller than those of both ensemble means.

In the case of monthly forecasts, the bias corrected multi-model ensemble shows some improvement in skill compared to the individual models. The forecasts produced by the superensemble and synthetic algorithms show better scores than those of the bias corrected ensemble mean and individual model forecasts in terms of AC, RMS, and in terms of categorical forecast measures. The synthetic superensemble forecast shows the lowest RMS error (highest skill) on average. In most years the AC and RMS of the synthetic ensemble/superensemble are better than those of individual model, bias corrected ensemble mean, climatology and superensemble based on SVD method. For categorical forecasts, the superensemble based on SVD shows the best skill. In the case of seasonal forecast, the synthetic ensemble and superensemble based on SVD have the best anomaly correlation. The synthetic superensemble seasonal forecasts show less AC improvement compared to the monthly forecast. This may be caused by shortage of training. In the case of categorical forecast, both superensembles show the best skills.

7. Summary and Conclusion

In this study, a new weighted multi-model ensemble approach for long-range (monthly and longer) forecast named synthetic superensemble for ocean-atmosphere coupled model is presented. The purpose of the synthetic superensemble approach is to improve long-range forecast skill. The sensitivity to initial conditions leads naturally to one of the fundamental problems in dynamical seasonal climate forecasting. The prediction skills of conventional single or multi-model ensemble have relied primarily on quality of actual input model data set. This means simply that the skill of ensemble prediction could be improved when the higher quality data set is deployed. This idea is tested with synthetic data generating algorithm. The synthetic superensemble presented in this paper is an empirical technique. The weighted multi-model synthetic ensemble technique relies on statistical relationships between individual member model data set and past observations. We generated a synthetic dataset from the actual DEMETER data and used it for synthetic superensemble forecast.

In the application here, the prediction skills are examined and compared with bias corrected ensemble, climatology, superensemble based on SVD method, synthetic ensemble, and synthetic superensemble forecasts. Due to the removal of biases of the different models, the forecast errors of the bias corrected multi-model ensemble and superensemble are already quite small. The forecast produced by the proposed method outperforms the other conventional forecasts. Our experiments show that the unbiased multi-model ensemble of ensemble forecasts, such as the DEMETER data, improves the prediction of spatial patterns i.e. the anomaly correlation, but it shows poor skill of categorical forecast. Based on our experiments we can summarize that the weighted multi-model ensemble/superensemble based on synthetic data set is one of the best solutions for long-range prediction, even though the skill of long-range prediction is still low.

Acknowledgments.

The research reported here was supported by NSF grant number ATM-0108741, and NOAA grant number NA06GPO512, FSURF grant number 1338-831-45, and NASA grant NAGS-13563. We gratefully acknowledge ECMWF for providing DEMETER data set.

Appendix

The linear equation for solving the error minimization problem is given by

$$C \cdot \mathrm{a} + \mathrm{e} = \tilde{O}''$$

where *C*, • and • indicate covariance matrix, vector of regression coefficients and vector of error, respectively, and $\tilde{O}'' = \sum_{t=0}^{Train} \tilde{O}'\tilde{F}'$. *Train* denotes training period and prime represents seasonal mean subtracted anomaly. Error variance is given by

$$\mathrm{e}^2 = \mathrm{e}^T \mathrm{e}$$

$$= (\tilde{O}'' - C\mathrm{a})^T (\tilde{O}'' - C\mathrm{a})$$

$$= \tilde{O}''^T \tilde{O}'' + \mathrm{a}^T C^T C\mathrm{a} - \tilde{O}''^T C\mathrm{a} - \mathrm{a}^T C^T \tilde{O}''^T$$

least square error of regression coefficients is obtained by solving

$$\frac{\partial \mathrm{e}^2}{\partial \mathrm{a}} = 2C^T C\mathrm{a} - 2C^T \tilde{O}'' = 0$$

And best linear unbiased estimator is

$$\mathrm{a} = (C^T C)^{-1} C^T \tilde{O}''$$

T denotes transpose of a vector.

References

Brown , B. G., Mahoney, J. L., Fowler, T. L and Henderson, J. 2001. Approaches for verification of ceiling and visibility diagnosis and forecasts. *Submitted to FAA Aviation Weather Research Program.*

Brankovic, C. and Palmer, T. N. 1997. Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.*, **125**, 859-874.

Buizza, R., Petroliagis, T., Palmer, T., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A. and Wedi, N. 1998. Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Met. Soc.*, **124**, 1935-1960.

Chang, Y., Schubert, S. D. and Suarez, M. J. 2000. boreal winter predictions with the GEOS-2 GCM: The role of boundary forcing and initial conditions. *Quart. J. Roy. Met. Soc.*, **126**, 2293-2321.

Doblas-Reyes, F. J., Déqué, M. and Piedelievre, J.-P. 2000. Multi-model spread and probabilistic forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.,***126**, 2069-2087.

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

Graham, R. J., Evans, A. D. L., Mylne, K. R., Harrison, M. S. J. and Robertson K. B. 2000. An assessment of seasonal predictability using atmospheric general circulation models. *Quart. J. Roy. Meteor. Soc.,***126**, 2211-2240.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

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

Kirtman, B. P., Min, D., Schopf, P. S. and Schneider, E. K. 2003. A new approach for coupled GCM sensitivity studies. *COLA Technical Report*, **CTR 154**, 48pp

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

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T. E., Bachiochi, D. R., Williford, C. E., Gadgil, S. and Surendran, S. 2000a. Improving tropical precipitation forecasts from a multi analysis superensemble*. J. Climate*, **13**, 4217-4227.

Krishnamurti, T. N., Kishtawal, C. M., Shin, D. W. and Williford, C. E. 2000b. Multi-model superensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196-4216.

Krishnamurti, T. N., Surendran, S., Shin, D. W., Correa-Torres, R. J., Kumar, T. S. V., Williford, C. E., Kummerow, C., Adler, R. F., Simpson, J., Kakar, R., Olson, W. S. and Turk, F. J. 2001. Real time multianalysis/multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.*, **129**, 2861-1883.

Krishnamurti, T. N., Rajendran, K., Vijaya Kumar, T. S. V., Lord, S., Toth, Z., Zou, X., Cocke, S., Ahlquist, J. E. and Navon, I. M. 2003. Improved skill for the anomaly correlation of geopotential heights at 500 hPa. *Mon. Wea. Rev.*, **131**, 1082-1102.

Krishnamurti, T. N., K. and Sanjay, J. 2003. A new approach to the cumulus parameterization issue. *Tellus,* **55A**, 275-300.

Leith, C. E. 1974. Theoretical skill of Mote Carlo forecasts. *Mon. Weather Rev.*, **102**, 409-418.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Met. Soc.*, **122**, 73-119.

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

Palmer, T. N. and Coauthors. 2003. Development of a European Multi-Model Ensemble System for seasonal to inter-annual prediction (DEMETER), *submitted to BAMS.*

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

Peng, P., Kumar, A., Van den Dool, A. H. and Barnston, A. G. 2002. An analysis of multi-model ensemble predictions for seasonal climate anomalies, *J. Geophys. Res*, 107, 4710, doi:10.1029/2002JD002712.

Press, W. H., Teukolsky, S. A., Vettering, W. T., Flannery, B. P. 1992. Numerical Recipes in Fortran, 2nd ed., *Cambridge Univ. Press,* 963pp

Stefanova, L. and Krishnamurti, T. N. 2002. Interpretation of seasonal climate forecast using Brier skill score, FSU superensemble, and the AMIP-1 dataset. *J. Climate*, **15**, 537-544.

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

Strauss, D. M. and Shukla, J. 2000. Distinguishing between the SST forced variability and internal variabilityin mid-latitudes: Analysis of observations and GCM simulations. *Quart. J. Roy. Meteor. Soc.*, **126**, 2323-2350.

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

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

Zwiers, F. W. 1996. Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. *Climate Dynamics*, **12**, 825-847.

List of tables

List of figures

Fig. 1. Schematic chart for the synthetic superensemble prediction system. The synthetic data are generated from the original data set by minimizing the residual error variance $E(e^2)$.

Fig. 2. A comparison of July precipitation (mm/day) forecasts in the tropics (30ºS-30ºN) for 2001. OBS, EM, SEM, SF and SSF in plots indicate observation, unbiased ensemble mean, synthetic ensemble mean, superensemble based on SVD and synthetic superensemble forecast, respectively.

Fig. 3. 15 years (1987-2001) averaged precipitation AC and RMS for July and December for global, tropical (30°S-30°N), and north hemispheric (0°-60°N) domains. The bars in the diagram indicate the 7 member models, unbiased ensemble mean (EM), climatology (CLIM; just for RMS), synthetic ensemble mean (SEM), superensemble based on SVD (SF), synthetic superensemble (SSF).

Fig. 4. Verification of July and December categorical forecasts for positive precipitation anomaly. The skills are averaged 15 years (1987-2001). The bars in the diagram show PODy, PODn, FAR, ETS, and TSS from left to right. EM, SEM, SF, and SSF indicate ensemble mean, synthetic ensemble mean, superensemble based on SVD, and synthetic superensemble, respectively.

Fig. 5. Cross-validated AC skill scores for the 1-month lead Summer (June, July, August) and Winter (December, January, February) global precipitation forecasts in 1987-2001. The bars in diagram indicate skill scores of the 7 individual member models, bias corrected ensemble mean (EM), synthetic ensemble mean (SEM),

superensemble based on SVD (SF), and synthetic superensemble (SSF) from left to right.

Fig. 6. 15 years (1987-2001) averaged AC precipitation skill scores of all seasons (MAM, JJA, SON, DJF) for global, tropical (30°S-30°N), and north hemispheric (0°-60°N) domains. The bars in the diagram indicate the 7 member models, unbiased ensemble mean (EM), synthetic ensemble mean (SEM), superensemble based on SVD (SF), synthetic superensemble (SSF).

Fig. 7. Anomaly forecast of precipitation for summer (June, July, August) 2001. OBS, EM, SEM, SF, and SSF indicate anomaly of observation, ensemble mean, synthetic ensemble mean, superensemble based on SVD, and synthetic superensemble, from top to bottom.

Fig. 8. Verification of categorical forecasts of positive precipitation anomaly for 15 years (1987-2001) averaged for all seasons (MAM, JJA, SON, DJF) for the tropics (30°S-30°N). The bars in the diagram show PODy, PODn, FAR, ETS, and TSS from left to right. EM, SEM, SF, and SSF indicate ensemble mean, synthetic ensemble mean, superensemble based on SVD, and synthetic superensemble, respectively.

Table 1. Contingency table for evaluation of categorical forecasts and verification measures for categorical forecasts.

Forecasts	**Observation**	
	Yes	**No**
Yes	YY	YN
No	NY	NN
Verification measures		
PODy	YY/(YY+NY)	Probability of “Yes” observations
PODn	NN/(YN+NN)	Probability of “No” observations
TSS	PODy+PODn-1	True Skill Statistics
FAR	YN/(YY+YN)	False Alarm Ratio
ETS (GSS)	(YY-C)/(YY+NY+YN-C), where C=(YY+YN)(YN+NY)/N N=YY+YN+NY+NN	Equitable Threat Score (Gilbert Skill Score)

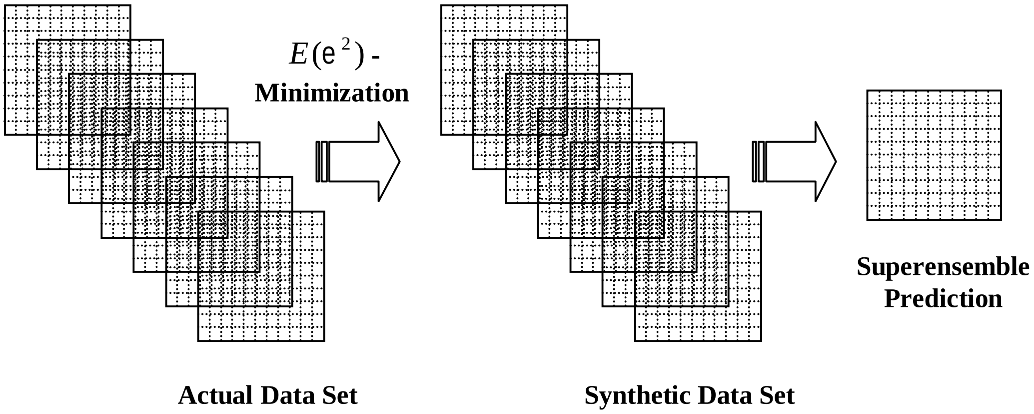

Fig. 1. Schematic chart for the synthetic superensemble prediction system. The synthetic data are generated from the original data set by minimizing the residual error variance $E(e^2)$.

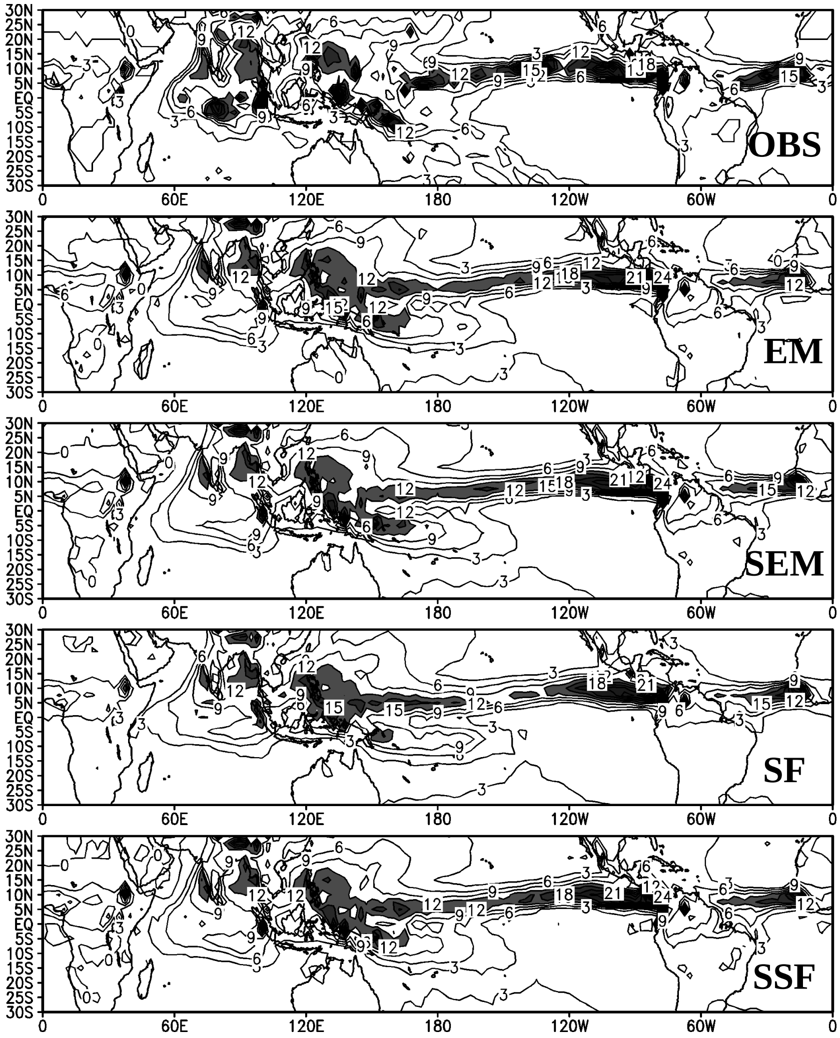

Fig. 2. A comparison of July precipitation (mm/day) forecasts in the tropics (30ºS-30ºN) for 2001. OBS, EM, SEM, SF and SSF in plots indicate observation, unbiased ensemble mean, synthetic ensemble mean, superensemble based on SVD and synthetic superensemble forecast, respectively.

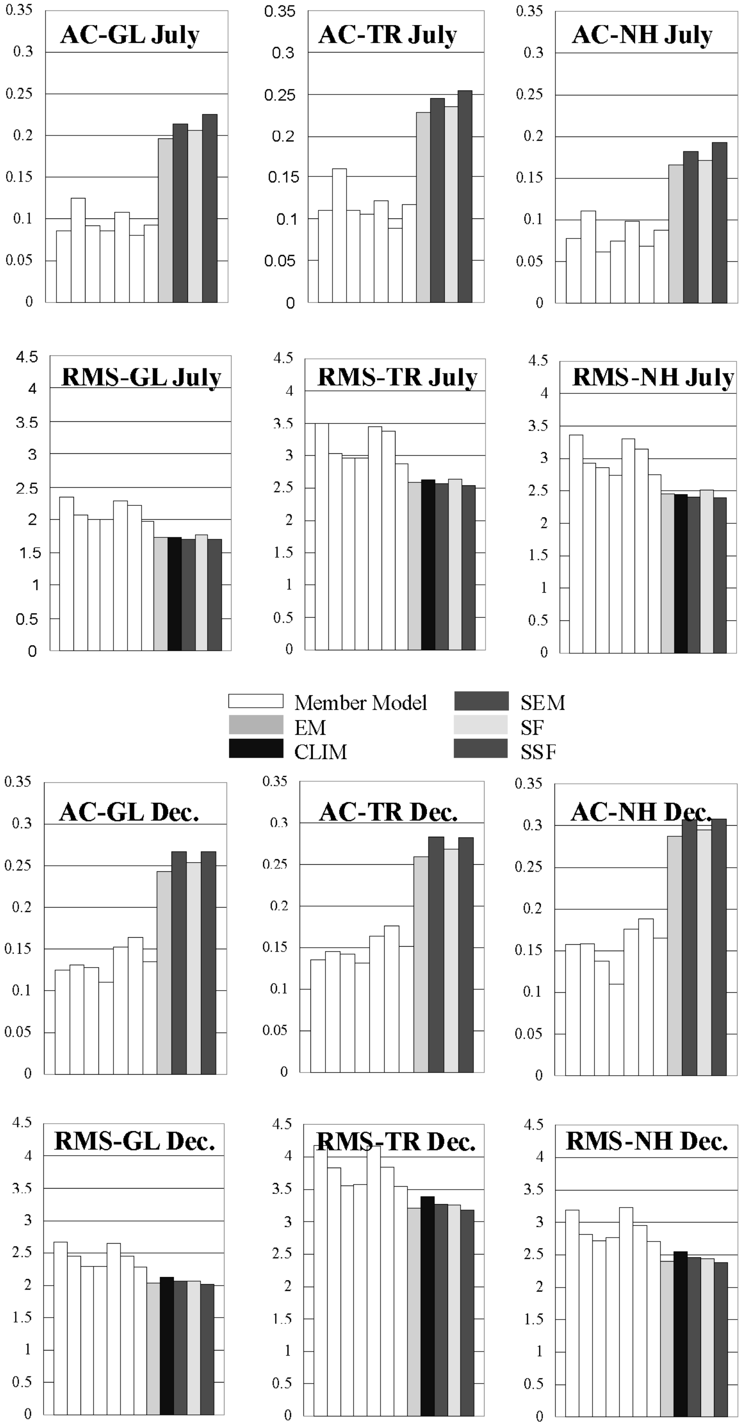

Fig. 3. 15 years (1987-2001) averaged precipitation AC and RMS for July and December for global, tropical (30°S-30°N), and north hemispheric (0°-60°N) domains. The bars in the diagram indicate the 7 member models, unbiased ensemble mean (EM), climatology (CLIM; just for RMS), synthetic ensemble mean (SEM), superensemble based on SVD (SF), synthetic superensemble (SSF).

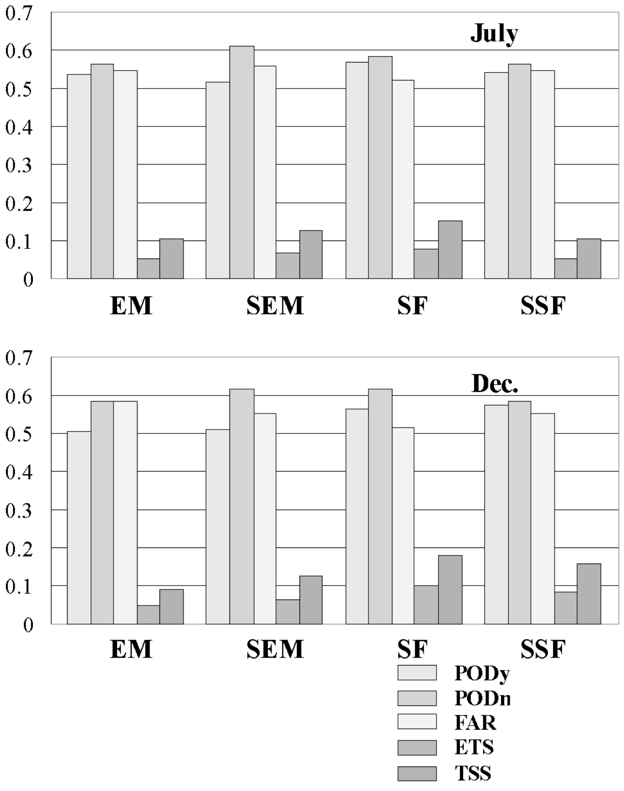

Fig. 4. Verification of July and December categorical forecasts for positive precipitation anomaly. The skills are averaged 15 years (1987-2001). The bars in the diagram show PODy, PODn, FAR, ETS, and TSS from left to right. EM, SEM, SVD, and SSF indicate ensemble mean, synthetic ensemble mean, superensemble based on SVD, and synthetic superensemble, respectively.

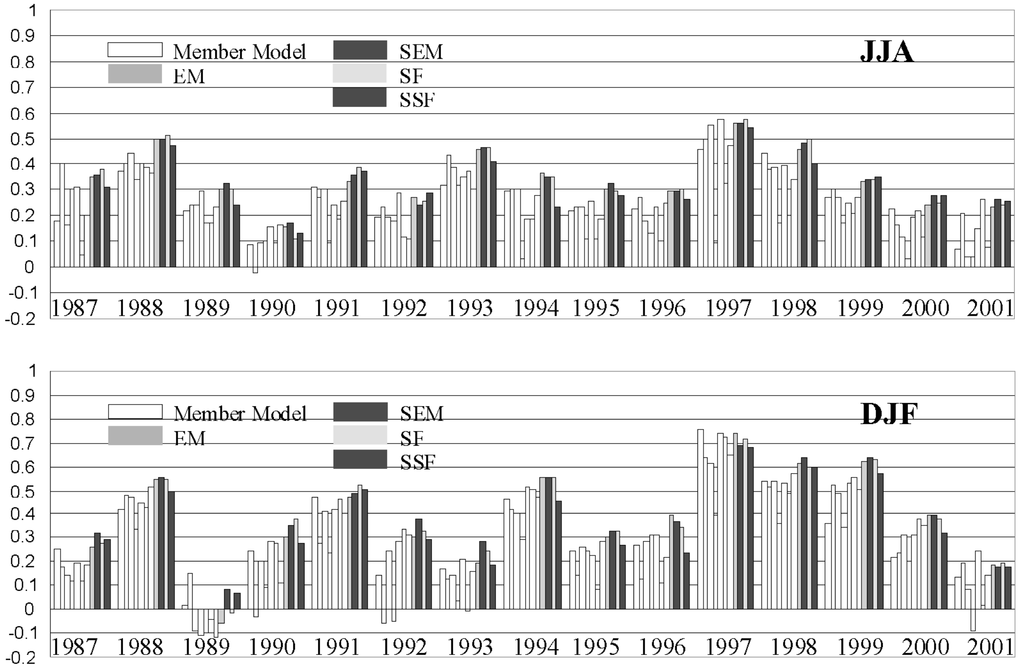

Fig. 5. Cross-validated AC skill scores for the 1-month lead Summer (June, July, August) and Winter (December, January, February) global precipitation forecasts in 1987-2001. The bars in diagram indicate skill scores of the 7 individual member models, bias corrected ensemble mean (EM), synthetic ensemble mean (SEM), superensemble based on SVD (SF), and synthetic superensemble (SSF) from left to right.

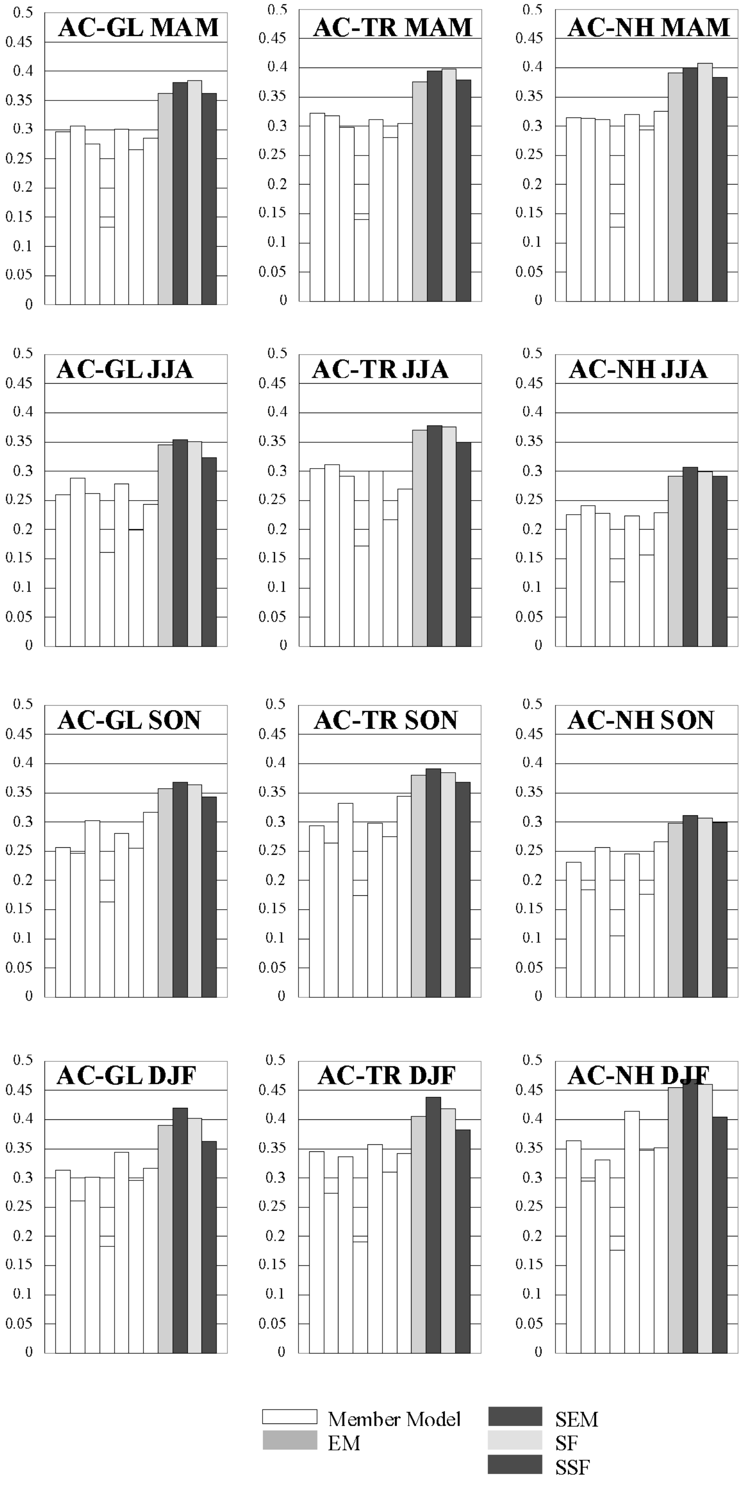

Fig. 6. 15 years (1987-2001) averaged AC precipitation skill scores of all seasons (MAM, JJA, SON, DJF) for global, tropical (30°S-30°N), and north hemispheric (0°-60°N) domains. The bars in the diagram indicate the 7 member models, unbiased ensemble mean (EM), synthetic ensemble mean (SEM), superensemble based on SVD (SF), synthetic superensemble (SSF).

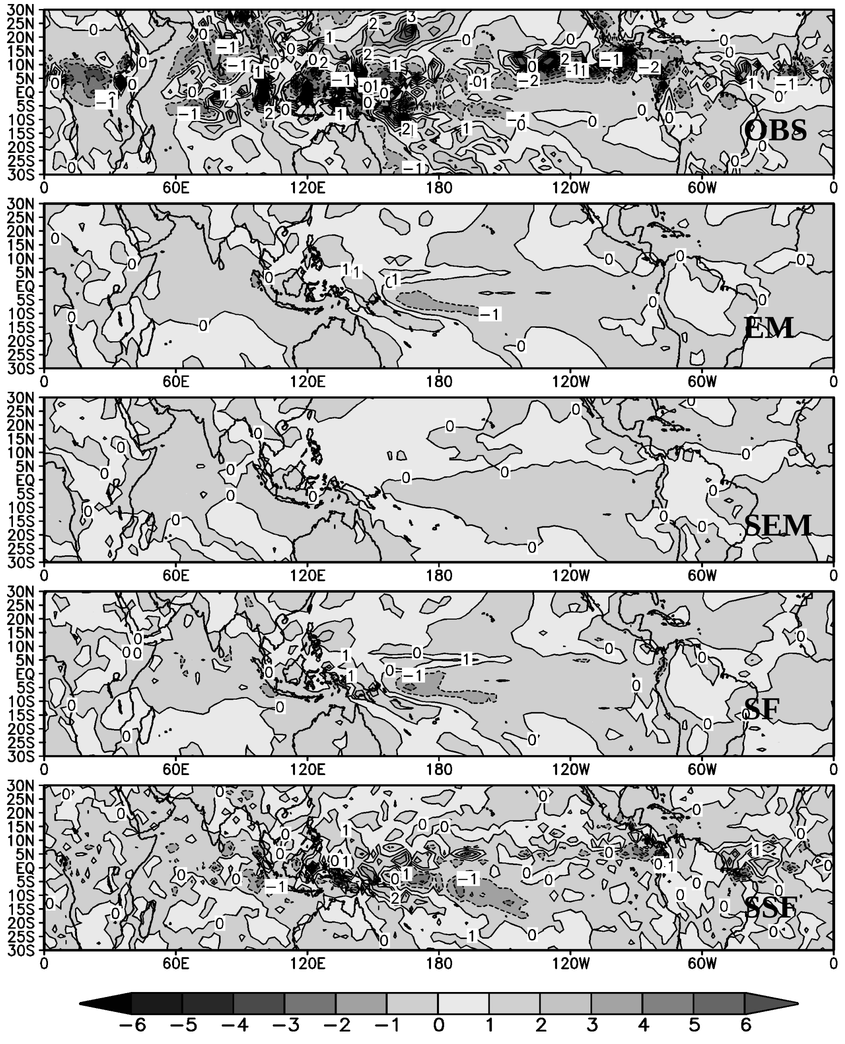

Fig. 7. Anomaly forecast of precipitation for summer (June, July, August) 2001. OBS, EM, SEM, SF, and SSF indicate anomaly of observation, ensemble mean, synthetic ensemble mean, superensemble based on SVD, and synthetic superensemble, from top to bottom.

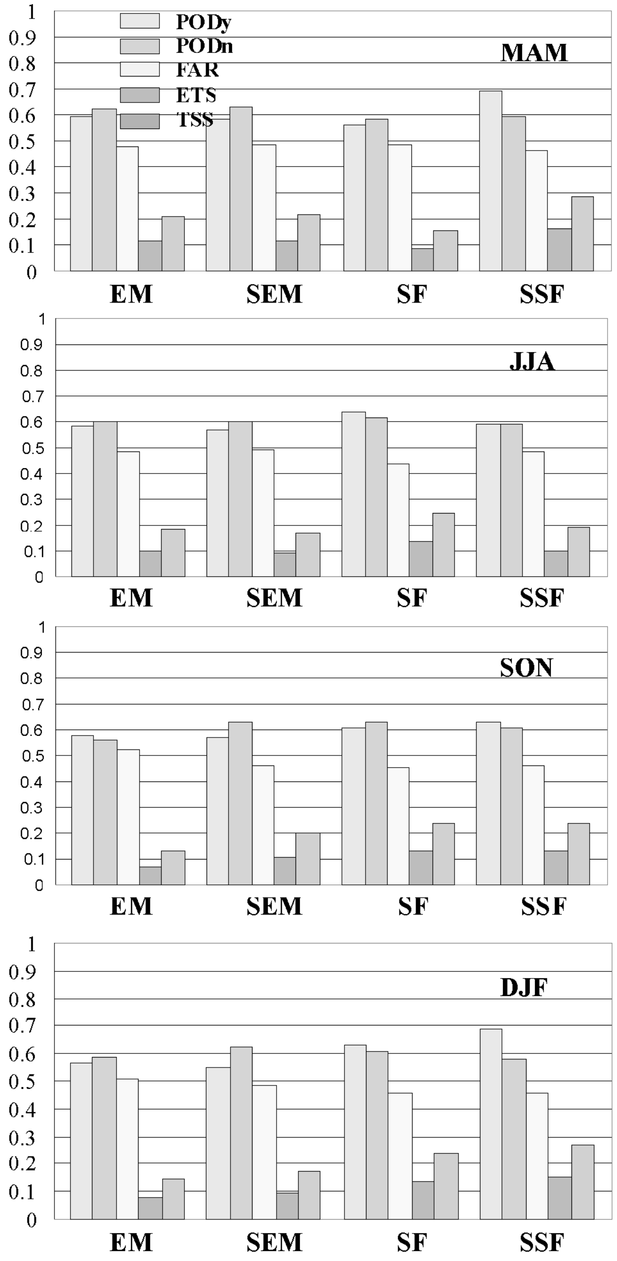

Fig. 8. Verification of categorical forecasts of positive precipitation anomaly for 15 years (1987-2001) averaged for all seasons (MAM, JJA, SON, DJF) for the tropics (30°S-30°N). The bars in the diagram show PODy, PODn, FAR, ETS, and TSS from left to right. EM, SEM, SF, and SSF indicate ensemble mean, synthetic ensemble mean, superensemble based on SVD, and synthetic superensemble, respectively.