Tellus (2005), 57A, 280–289
Printed in UK. All rights reserved

TELLUS

A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts

By 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, Florida State University;* [2]*Korea Meteorological Administration, Seoul, Korea;* [3]*National Centre for Medium-Range Weather Forecasting (NCMRWF), Noida, India;* [4]*Department of Oceanography, Florida State University*

(Manuscript received 31 March 2004; in final form 17 January 2005)

ABSTRACT

In this paper, a multi-model ensemble approach with statistical correction for seasonal precipitation forecasts using a coupled DEMETER model data set 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; consequently, the predictive skill of extended forecasts remains quite low. One of the approaches to the improvement of seasonal prediction is the empirical weighted multi-model ensemble, or superensemble, combination. In the superensemble approach, 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 skill of a superensemble relies strongly on the past performance of the individual member models used in its construction. The algorithm proposed here involves empirical orthogonal function (EOF) filtering of the actual data set prior to the construction of a multi-model ensemble or superensemble as an alternative solution for seasonal prediction. This algorithm generates a new data set from the input multi-model data set by finding a consistent spatial pattern between the observed analysis and the individual model forecast. This procedure is a multiple linear regression problem in the EOF space. The newly generated EOF-filtered data set is then used as an input data set for the construction of a multi-model ensemble and superensemble. The skill of forecast anomalies is assessed using statistics of categorical forecast, spatial anomaly correlation and root mean square (RMS) errors. The various verifications show that the unbiased multi-model ensemble of DEMETER forecasts improves the prediction of spatial patterns (i.e. the anomaly correlation), but it shows poor skill in categorical forecast. Due to the removal of seasonal mean biases of the different models, the forecast errors of the bias-corrected multi-model ensemble and superensemble are already quite small. Based on the anomaly correlation and RMS measures, the forecasts produced by the proposed method slightly outperform 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 orthogonal function (EOF) based empirical multi-model ensemble/superensemble method for seasonal climate forecast using the DEMETER (Development of a European multi-model ensemble system for seasonal to interannual 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 in many studies (Doblas-Reyes et al., 2000; Graham et al., 2000; Palmer et al., 2000; Palmer et al., 2004). Such ensemble techniques are nowadays routinely used at operational weather forecasting centers (Houtekamer et al., 1996; Molteni et al., 1996; Toth and Kalnay, 1997; Buizza et al., 1998; Stephenson and Doblas-Reyes, 2000) and are also applied in seasonal time-scale climate

*Corresponding author:
e-mail: wtyun@kma.go.kr
†Present address: Korea Meteorological Administration, 460-18 Shindaebang dong dongjak gu, Seoul, Korea.

studies (Zwiers, 1996; Brankovic and Palmer, 1997; Pavan and Doblas-Reyes 2000; Kharin and Zwiers 2002; Peng et al., 2002).

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 post-processing 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; Stefanova and Krishnamurti, 2002; Krishnamurti and Sanjay, 2003; 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 atmospheric GCMs and found minimal skill improvement. Kharin and Zwiers (2002) assessed different ways of constructing multi-model forecasts and found a disagreement with the results of Krishnamurti et al., in that their regression-improved multi-model forecast (i.e. the superensemble) performed worse than the multi-model 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 the superensemble of Krishnamurti et al. 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 the 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 because 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 present paper focuses on improving the seasonal time-scale climate prediction skill through the generation of an EOF-based data set from actual multi-model data prior to the construction of the multi-model ensemble/superensemble prediction.

2. Multi-model data set

The multi-model data set used in this study is the output of seven global coupled ocean–atmospheric models from the DEMETER project (Palmer et al., 2004). The DEMETER hindcasts were started from 1 February, 1 May, 1 August and 1 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 1987–2001 is evaluated in this paper. We use the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis as verification data. All calculations are made using cross-validation, with each year being successively withheld from the training data set, and the remaining 14 yr used for calculation of the model and observed statistics. A complete description of the DEMETER data set can be found on the website http://www.ecmwf.int/research/demeter.

3. Algorithm for proposed multi-model 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 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. In this section, we introduce the computational algorithm for the creation of a new EOF-based data set for multi-model ensemble prediction.

In the context of seasonal climate forecasts, many studies (Krishnamurti et al., 1999, 2000a,b, 2001, 2003; Doblas-Reyes et al., 2000; Pavan and Doblas-Reyes 2000; Stephenson and Doblas-Reyes 2000; Kharin and Zwiers 2002; Peng et al., 2002; Stefanova and Krishnamurti, 2002; Yun et al., 2003; Palmer et al., 2004) have discussed various multi-model approaches for forecasting of anomalies, such as the ensemble mean, the unbiased ensemble mean and the superensemble forecast. These are

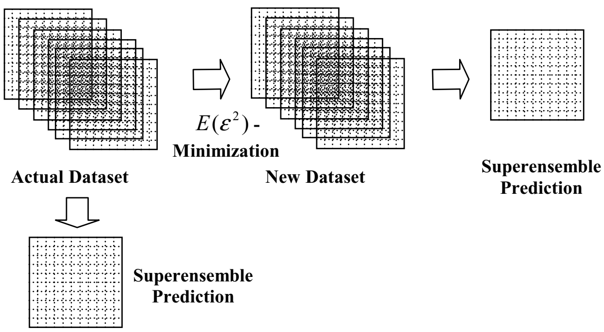

Fig. 1. Schematic chart for the proposed superensemble prediction system. The new data set is generated from the original data set by minimizing the residual error variance $E(\varepsilon^2)$ for each model.

defined as follows:

$$E_{\mathrm{b}} = \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{O}) \tag{1}$$

$$E_{\mathrm{c}} = \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{F_i}) \tag{2}$$

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

Here, E_{b} is the ensemble mean, E_{c} is the unbiased ensemble mean, S is the superensemble, F_i is the ith model forecast out of N models, $\overline{F_i}$ is the monthly or seasonal mean of the ith forecast over the training period, $\overline{O}$ is the observed monthly or seasonal mean over the training period, and a_i is the regression coefficient of the ith model. The difference between these approaches comes from the mean bias and the weights. Both the unbiased ensemble mean and the superensemble contain no mean bias because the seasonal climatologies of the models have been considered. The difference between the unbiased ensemble and the superensemble comes from the differential weighting of the models in the latter case. 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 the input multi-model predictions are statistically corrected to reduce the model errors.

Figure 1 is a schematic chart illustrating the proposed algorithm. The new data set is generated from the original data set by finding a consistent spatial pattern between the observed

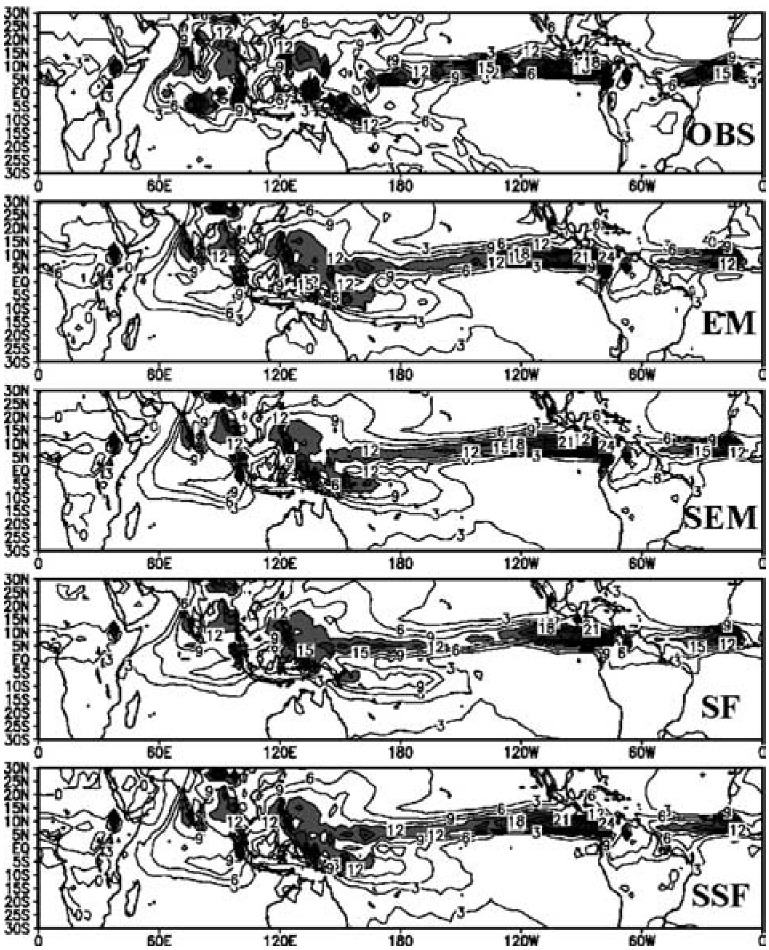

Fig. 2. A comparison of July precipitation (mm d^{-1}) forecasts in the tropics (30°S–30°N) for 2001. OBS, EM, SEM, SF and SSF in the plots indicate observation, unbiased ensemble mean of the original data set, unbiased ensemble mean of the EOF-filtered data set, superensemble of the original data set and superensemble of the EOF-filtered data set, 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)

analysis and each model. This procedure is a linear regression problem in EOF space. The newly generated set of EOF-filtered data is then used as an input multi-model data set for ensemble/superensemble forecast. The computational procedure for generating the new data set is described below.

The observation data (O) and the multi-model forecast data set (F_i) can be written as linear combinations of EOFs, which describe the spatial and temporal variability:

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

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

Here, $\tilde{O}_n(t)$, $\tilde{F}_{i,n}(t)$ and $\phi_n\ (x)$, $\varphi_{i,n}(x)$ are the principal component (PC) time series and the corresponding EOFs of the nth mode for the observation and model forecast, respectively. Index i indicates a particular member model. The PCs in eqs. (4) and (5) represent the time 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 the training observations. The regression relationship between the observation PC time series and the 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 eq. (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 is minimized. The covariance matrix is constructed with the PC time series of each model. For obtaining the regression coefficients $\alpha_{i,n}$, the covariance matrix is built with the seasonal cycle-removed anomaly. Once the regression coefficients $\alpha_{i,n}$ are found, the PC time series of new data set is written as

$$\tilde{F}_i^{\mathrm{reg}}(T) = \sum_n \alpha_{i,n}\tilde{F}_{i,n}(T). \tag{7}$$

The new data set is now generated by reconstruction with corresponding EOFs and PCs:

$$F_i^{\mathrm{syn}}(x, T) = \sum_n \tilde{F}_{i,n}^{\mathrm{reg}}(T)\phi_n(x). \tag{8}$$

This EOF-filtered data set generated from the DEMETER coupled multi-model is used as an input data set for both multi-model ensemble and superensemble prediction systems that produce deterministic forecasts. What is unique about the new data set is that it minimizes the variance of the residual error between the observations and each of the member models (Fig. 1). The residual error variance is minimized using a least-squares error approach.

To illustrate the performance of the proposed multi-model ensemble, a comparison of precipitation forecasts for the month of July 2001 over the tropics (30°S–30°N) is shown in Fig. 2. The observed analysis is shown in the top panel. OBS, EM, SEM, SF and SSF in the plots indicate observations, bias-corrected ensemble mean of the original data, ensemble mean of EOF-filtered data set, superensemble of the original data and superensemble of the EOF-filtered data set, respectively. Figure 2 shows that the proposed superensemble can capture the spatial patterns and the strength of amplitude as well as the original ensemble mean (EM) and superensemble (SF). All four forecasts are visually very similar. In order to assess their relative accuracy, we use the objective verification metrics described in the following section.

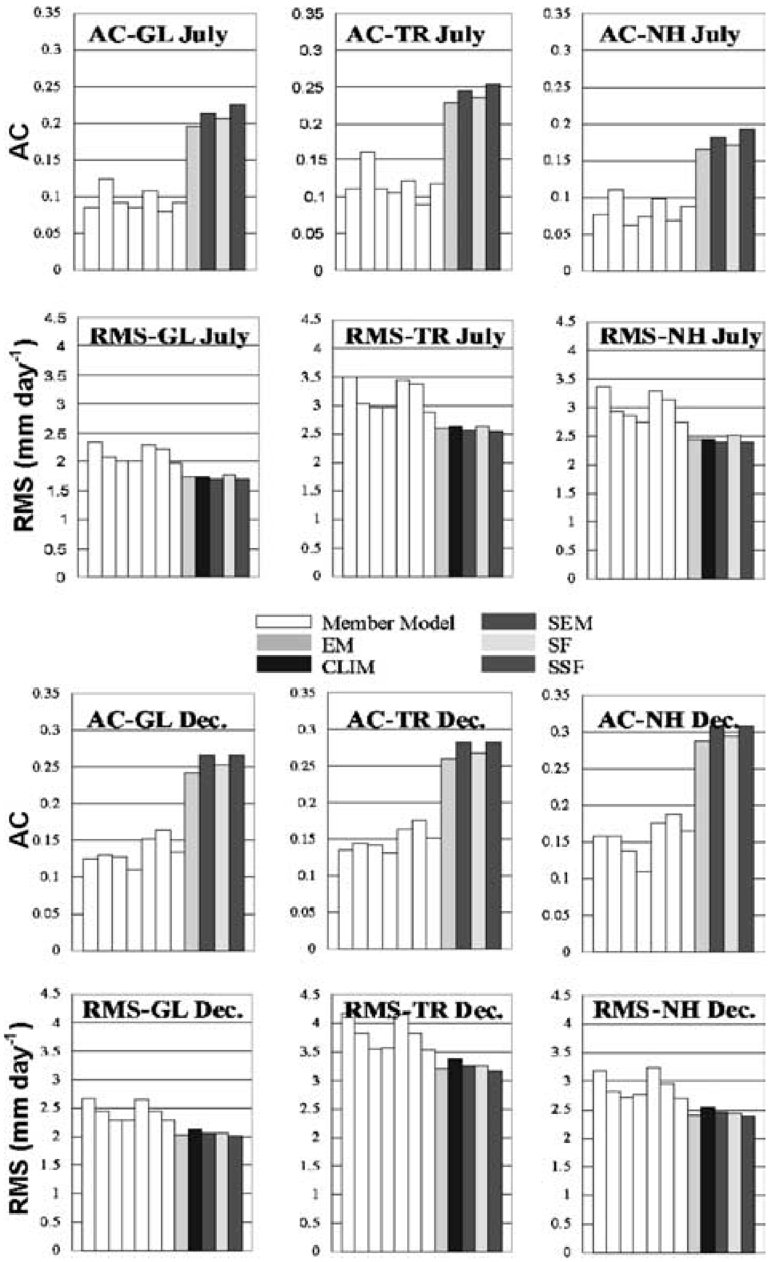

Fig. 3. The 15-yr (1987–2001) averaged precipitation AC and RMS for July and December for global, tropical (30°S–30°N) and North Hemispheric (0°–60°N) domains. The bars in the diagram indicate the seven member models, unbiased ensemble mean (EM) of the original data set, climatology (CLIM; just for RMS), unbiased ensemble mean of the EOF-filtered data set (SEM), superensemble of the original data set (SF) and superensemble of the EOF-filtered data set (SSF).

4. Verification metrics

The spatial anomaly correlation (AC) and root mean square (RMS) 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. The AC is a good measure of phase error that does not take bias into account (Déqué, 1997). It is possible for a forecast with large errors in magnitude to still have good correlation coefficients. It is therefore necessary to evaluate the phase and magnitude errors separately:

$$\mathrm{AC} = \frac{\sum (F - \bar{F})(O - \bar{O})}{\sqrt{\sum (F - \bar{F})^2}\sqrt{\sum (O - \bar{O})^2}} \tag{9}$$

$$\mathrm{RMS} = \sqrt{\frac{1}{G}\sum [(F - \bar{F}) - (O - \bar{O})]^2}. \tag{10}$$

Here, the overbar denotes time average, G denotes the number of grid points, the summation is performed over space, and F is the forecast whose errors are being assessed.

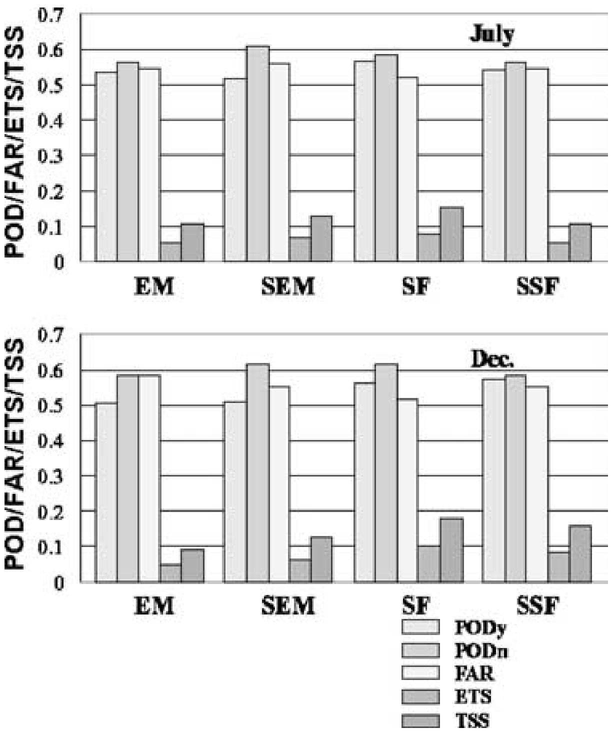

Fig. 4. Verification of July and December categorical forecasts for positive precipitation anomaly. The skills are averaged 15 yr (1987–2001). The bars in the diagram show PODy, PODn, FAR, ETS and TSS from left to right. EM, SEM, SF and SSF indicate unbiased ensemble mean of the original data set, unbiased ensemble mean of the EOF-filtered data set, superensemble of the original data set, and superensemble of the EOF-filtered data set, respectively.

We also use categorical forecasts to quantify forecast skill. 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 the verification measures are also listed in Table 1. The probabilities of 'yes' or 'no' detection (PODy, PODn) are estimates of the proportions of 'yes' or 'no' observations that were correctly forecast. These measure the ability of the forecasts to discriminate between 'yes' and 'no' observations. The true skill score (TSS) summarizes the ability of the forecasts. The TSS is a more objective measure of the skill of categorical forecast, because it combines the PODy and PODn measures so that models that have high PODn by virtue of underforecasting events (and thus have a low PODy) and models that have high PODy by virtue of overforecasting (and thus a low PODn) are no longer unduly rewarded. The 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.

5. Results of multi-model synthetic ensemble/superensemble forecasts

In this section we describe and compare the skill and performance of the unbiased ensemble of the original data set, superensemble of the original data set, and the ensemble/superensemble of EOF-filtered forecasts. The anomaly forecasts and scores have been computed using cross-validation. The original and EOF-filtered 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 et al., 2000). Kharin and Zwiers (2002) reported that the skill of the ensemble mean forecast is dependent on ensemble size. We assessed the skill of multi-model ensemble and superensemble based on our experiment.

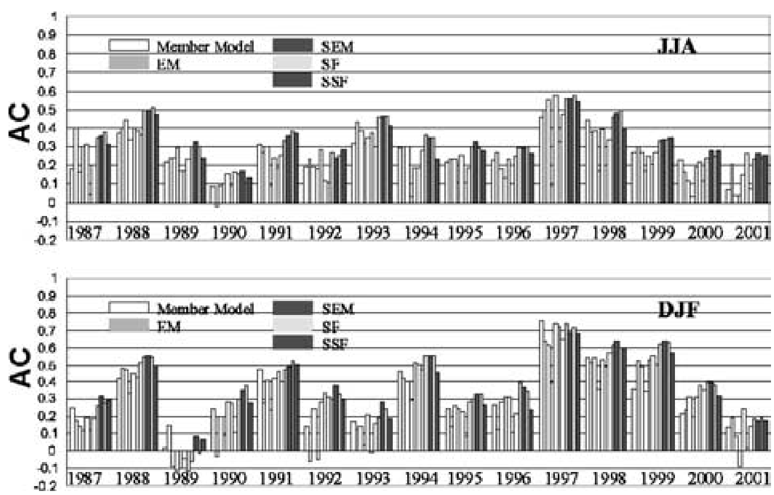

Fig. 5. Cross-validated AC for the one-month lead summer (June, July, August) and winter (December, January, February) global precipitation forecasts in 1987–2001. The bars in diagram indicate skill scores of the seven individual member models, unbiased ensemble mean of the original data set (EM), unbiased ensemble mean of the EOF-filtered data set (SEM), superensemble of the original data set (SF) and superensemble of the EOF-filtered data set (SSF) from left to right.

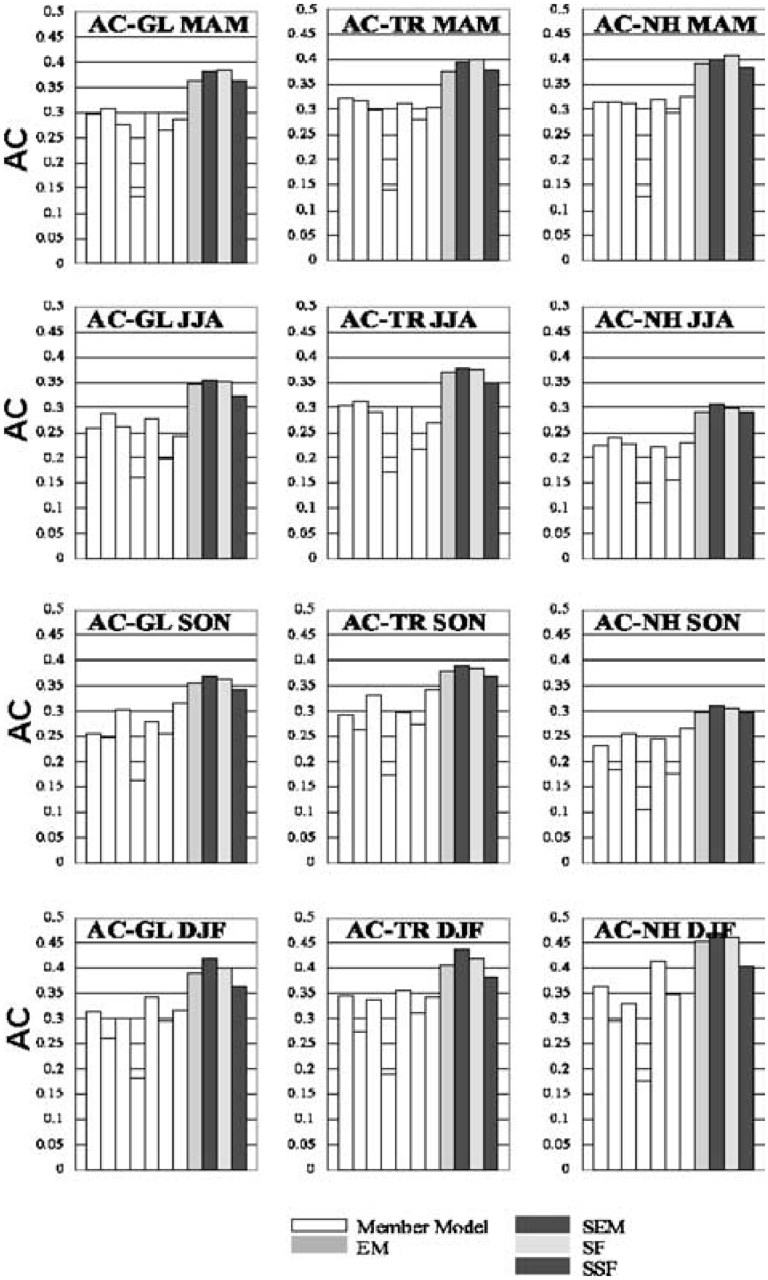

Fig. 6. The 15-yr (1987–2001) averaged precipitation AC for all seasons (MAM, JJA, SON, DJF) for global, tropical (30°S–30°N) and North Hemispheric (0°–60°N) domains. The bars in the diagram indicate the seven member models, unbiased ensemble mean of the original data set (EM), unbiased ensemble mean of the EOF-filtered data set (SEM), superensemble of the original data set (SF), and superensemble of the EOF-filtered data set (SSF).

At first, we took one run from each member model and performed monthly superensemble forecasts with 168 months training. Figure 3 illustrates the 15 yr (1987–2001) averaged AC and RMS skill of the unbiased ensemble mean of the original data set (EM), climatology (CLIM), unbiased ensemble mean of the EOF-filtered data set (SEM), superensemble of the original data set (SF) and superensemble of the EOF-filtered data set (SSF) for 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 EOF-based ensemble/superensemble forecasts show best skill in terms of AC and RMS measures. The improvement of the AC of forecast A over forecast B can be defined as $AC_A/AC_B - 1$, where AC_A is the AC of forecast A and AC_B is the AC of forecast B. Similarly, in terms of the RMS errors, the improvement of forecast A over forecast B can be defined as $1 - RMS_A/RMS_B$, where RMS_A and RMS_B are the RMS errors of forecasts A and B, respectively. Using these definitions, the improvement of the EOF-based ensemble/superensemble over the original ensemble and superensemble is of the order of 10% for the anomaly correlations, and 1–2% for the RMS. In the sense of categorical forecast of positive precipitation anomaly, the superensemble of

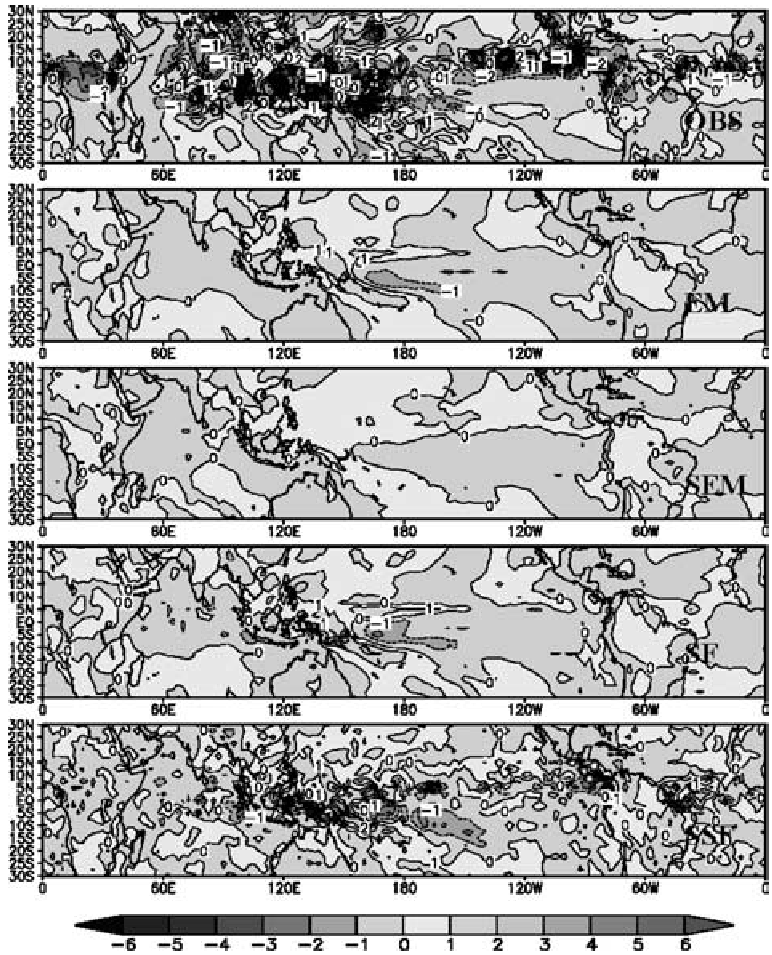

Fig. 7. Anomaly forecast of precipitation for summer (June, July, August) 2001. OBS, EM, SEM, SF and SSF indicate anomaly of observation, unbiased ensemble mean of the original data set, unbiased ensemble mean of the EOF-filtered data set, superensemble of the original data set, and superensemble of the EOF-filtered data set, from top to bottom.

the original data set 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 nine ensemble members for each model for the extended seasonal superensemble forecast. The superensemble was constructed by training with 56 seasons of forecast data set. Figure 5 shows the one-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 the individual models in most years. Averaged over 15 yr, the ensemble mean of the EOF-filtered data set shows the best AC, better than the original data ensemble mean by up to about 10%, except for spring (March, April, May) (Fig. 6). The range of AC (EM, SEM, SF, SSF) is between 0.29 and 0.47. Figure 7 illustrates the precipitation anomaly forecasts for summer 2001. The observed analysis is shown in the top panel. All the ensemble anomalies are quite small due to the averaging of a large number of realizations. However, the superensemble forecast using the EOF-filtered data set gives a more realistic magnitude and pattern of precipitation anomaly compared to the other multi-model methods.

The verification of categorical forecast of positive precipitation anomaly is shown in Fig. 8. Both superensemble forecasts are better than both ensemble mean forecasts 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 original superensemble and by the EOF-filtered ensemble and superensemble show better scores than those of the original bias-corrected ensemble mean and individual model forecasts in terms of AC, RMS, and in terms of categorical forecast measures. The EOF-filtered superensemble forecast shows the lowest RMS error (highest skill) on average. In most years the AC and RMS of the EOF-filtered ensemble/superensemble are better than those of individual model, bias-corrected ensemble mean of the original data set, climatology and superensemble of the original data set. For categorical forecasts, the

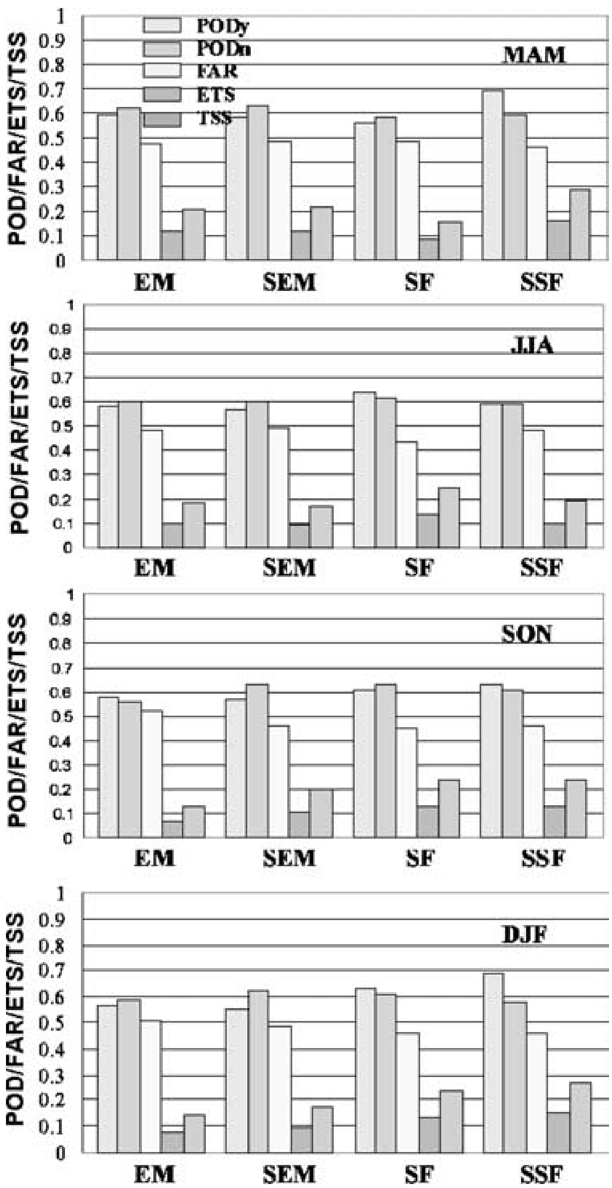

Fig. 8. Verification of categorical forecasts of positive precipitation anomaly for 15 yr (1987–2001) averaged for all seasons (MAM, JJA, SON, DJF) for the tropics (30°S–30°N). The bars in the diagram show PODy, PODn, FAR, ETS and TSS from left to right. EM, SEM, SF and SSF indicate the unbiased ensemble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

original superensemble shows the best skill. In the case of seasonal forecast, the AC of the EOF-filtered ensemble and the original superensemble are better than the EOF-filtered superensemble seasonal forecasts. This may be caused by shortage of training. In the case of categorical forecasts, both superensembles show the best skills.

6. Summary and conclusions

In this paper, we present a new addition to the multi-model ensemble approach for long-range (monthly and longer) forecast for the ocean–atmosphere coupled model. The new algorithm consists of EOF filtering of the individual models by finding a consistent pattern between the model forecast and the observations. The generated EOF-filtered data set is then used as input for the multi-model ensemble and superensemble systems. The purpose of the proposed algorithm is to reduce the model forecast errors and to improve long-range forecast skill. The prediction skills of single or multi-model ensembles rely primarily on quality of actual input model data set. This means that the skill of ensemble prediction could be improved when a data set with reduced errors is deployed for the calculation of multi-model statistics. This idea is tested with the proposed algorithm. The EOF-filtered ensemble/superensemble presented in this paper is an empirical post-processing technique that relies on the statistical relationships between individual member model forecasts and past observations established during a training period. We generated the new EOF-filtered data set from the actual DEMETER data and used it for multi-model ensemble and superensemble forecasts.

The prediction skills of the proposed algorithm are examined and compared with those of the unbiased ensemble and superensemble of the original data set and the climatology. Due to the removal of biases of the different models, the forecast errors of the bias-corrected multi-model ensemble and superensemble are quite small. Our experiment shows that the multi-model ensemble/superensemble forecast produced with EOF-filtered data slightly outperforms the other conventional forecasts. The unbiased multi-model ensemble of ensemble forecasts, such as the DEMETER data, improves the prediction of spatial patterns (i.e. the AC), but it shows poor skill of categorical forecast. Based on our experiments we can summarize that the multi-model ensemble/superensemble based on EOF-filtered data may contribute towards the improvement of the long-range prediction skill, even though the skill of long-range prediction is still low.

7. Acknowledgments

The research reported here was supported by the National Science Foundation (NSF) grant number ATM-0108741, the National Oceanic and Atmospheric Administration (NOAA) grant number NA06GPO512, the Florida State University Research Foundation (FSURF) grant number 1338-831-45, and the National Aeronautics and Space Administration (NASA) grant NAG5-13563. We gratefully acknowledge ECMWF for providing the DEMETER data set.

References

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. and co-authors. 1998. Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Q. J. R. Meteorol. 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. *Q. J. R. Meteorol. Soc.* **126**, 2293–2321.

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

Doblas-Reyes, F. J., Déqué, M. and Piedelievre, J.-P. 2000. Multi-model spread and probabilistic forecasts in PROVOST. *Q. J. R. Meteorol. Soc.* **126**, 2069–2087.

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. *Q. J. R. Meteorol. 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 multi-model 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. and Sanjay, J. 2003. A new approach to the cumulus parametrization issue. *Tellus* **55A**, 275–300.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z. and co-authors. 1999. Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* **285**, 1548–1550.

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T. E., Bachiochi, D. R. and co-authors. 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. and co-authors. 2001. Real-time multianalysis/multi-model 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. and co-authors. 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 Mote Carlo forecasts. *Mon. Wea. Rev.* **102**, 409–418.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119.

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

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. and co-authors. 2004. Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85**, 853–872.

Pavan, V. and Doblas-Reyes, J. 2000. Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Climatic Dyn.* **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. and Flannery, B. P. 1992. *Numerical Recipes in Fortran*. 2nd edn. Cambridge Univ. Press, Cambridge, 963 pp.

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

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

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

Yun, W.-T., Stefanova, L. 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 Dyn.* **12**, 825–847.