Improved Seasonal Climate Forecasts of the South Asian Summer Monsoon Using a Suite of 13 Coupled Ocean–Atmosphere Models

ARINDAM CHAKRABORTY AND T. N. KRISHNAMURTI

Department of Meteorology, The Florida State University, Tallahassee, Florida

(Manuscript received 29 December 2004, in final form 6 October 2005)

ABSTRACT

Several modeling studies have shown that the south Asian monsoon region has the lowest skill for seasonal forecasts compared with many other domains of the world. This paper demonstrates that a multimodel synthetic superensemble approach, when constructed with any set of coupled atmosphere–ocean models, can provide improved skill in seasonal climate prediction compared with single-member models or their ensemble mean for the south Asian summer monsoon region. However, performance of the superensemble tends to improve when a better set of input member models are used. As many as 13 state-of-the-art coupled atmosphere–ocean models were used in the synthetic superensemble algorithm. The merit of this technique lies in assigning differential weights to the member models. The rms errors, anomaly correlations, case studies of extreme events, and probabilistic skill scores are used here to assess these forecast skills. It was found that over the south Asian region the seasonal forecasts from the superensemble are, in general, superior to the forecasts of the individual member models, and their bias-removed ensemble mean at a significance level of 95% or more (based on a Student's *t* test) during the 13 yr of forecasts. Moreover, the skill of the superensemble was found to be better than those of the ensemble mean over smaller domains as well as during extreme events that were monitored, especially during the switch on and off of the Indian Ocean dipole, which seems to modulate the Indian monsoon rainfall. The results of this paper suggest that the superensemble provides somewhat consistent forecasts on the seasonal time scale. This methodology needs to be tested for real-time seasonal climate forecasting over the south Asian region.

1. Introduction

The agriculture-based economy in many south Asian countries is dependent on the seasonal southwest summer monsoon precipitation. Therefore, an accurate prediction of the monsoon precipitation over this region at least one season in advance is extremely important. Unfortunately, the present-day atmospheric general circulation models (AGCMs) fail to simulate the mean and interannual variation of the south Asian summer monsoon (Gadgil and Sajani 1998; Kang et al. 2002). The seasonal forecasting skills of the AGCMs were found to be poorer in simulating the Asian–Australian monsoon as compared with the El Niño regions (Wang et al. 2004a). This is because prescribing sea surface temperature (SST) anomalies in AGCMs fails to provide realistic measures of atmosphere–ocean coupling away from the El Niño regions.

On the other hand, the state-of-the-art coupled general circulation models (CGCMs) fail in reproducing the exact annual cycle, mean, and interannual variability of the atmosphere when seasonal forecast time scales are considered (Yang and Anderson 2000). Some of these shortcomings originate from model bias, which is inevitable in general circulation models (GCMs). Furthermore, small perturbations of the initial condition also result in very different final states (Lorenz 1969). These are some of the main sources of errors in a GCM. One method of reducing the forecast errors is the ensemble simulation (Brankovic et al. 1990; Brankovic and Palmer 1997). Several approaches were attempted to combine multimodel ensemble forecasts to a single reliable forecast that carries higher skills when compared to the individual member models. These include the simple ensemble mean (Peng et al. 2002; Pavan and Doblas-Reyes 2000; Doblas-Reyes et al. 2000; Stephenson and Doblas-Reyes 2000; Palmer et al.

Corresponding author address: Arindam Chakraborty, Department of Meteorology, The Florida State University, Tallahassee, FL 32306.
E-mail: arch@io.met.fsu.edu

2004), regression-improved ensemble mean (Peng et al. 2002; Kharin and Zwiers 2002), bias-removed ensemble mean (Kharin and Zwiers 2002), and the multimodel superensemble (Krishnamurti et al. 1999). The multimodel superensemble technique showed higher skill for short range and seasonal forecasting compared with member model forecasts (Krishnamurti et al. 2000a,b, 2001, 2002, 2003; Yun et al. 2003, 2005). There is one major difference between a bias-removed ensemble mean of multimodels and the superensemble. The former assigns a weight $1/N$ for all models after bias removal (where N is the number of models), and the latter assigns fractional weights (positive or negative) at different locations for different models. Some models are very poor compared with the best model. Assigning the same weights $1/N$ (after bias removal) does not make the very poor model at the same weight as the best model (Stefanova and Krishnamurti 2002). This merit of the superensemble seems to provide somewhat higher skill even over the bias-removed ensemble mean forecast (Stefanova and Krishnamurti 2002).

How this collective bias-removal procedure of the superensemble affects the forecast both for the total and anomaly fields over the south Asian monsoon region is examined in this paper. An accurate forecast of the total field is as important as the anomaly forecast because a model with improved climatology (i.e., the total fields) provides better seasonal means and interannual variations (Sperber and Palmer 1996). It was shown by Krishnamurti et al. (2003) that the forecast skill of the superensemble is higher if high quality model datasets and high quality observations (or analysis) were used in the training phase. But how the superensemble forecasts depend on the number of input models has not been explored so far. Moreover, a comprehensive study on the performance of the superensemble during extreme climatic events has not been addressed yet. In this paper, 13 state-of-the-art coupled ocean–atmosphere models are being used to compare the skill of the superensemble forecasts with that of the conventional ensemble mean and of the individual models over the south Asian monsoon region in seasonal time scale. The main objective of the present work is to compare the forecasting skill of the superensemble for this suite of coupled models, and to statistically quantify the spread of forecast skill over the south Asian monsoon region. In addition to that, this study intercompares the forecasting skill of the 13 CGCMs over the south Asian monsoon domain. This study also addresses the issue of the dependence of forecast skill on the number of input models used for constructing the superensemble and its performance during extreme events on the seasonal time scale. Section 2 describes the methodology in creating the superensemble, the models used in the analysis, and the datasets. Section 3 provides an example of the bias of the simulations of the CGCM. Section 4 describes the results assessed in terms of skill scores and case studies. How well the superensemble captures the extreme events is illustrated in section 5. The dependence of the superensemble forecast on the input datasets is illustrated in section 6. Section 7 summarizes the major results of this study.

2. Methodology, models, and datasets

a. Conventional superensemble methodology

In the superensemble technique (Krishnamurti et al. 1999, 2000a) a single consensus forecast is obtained from a set of multimodel forecasts. At first, the time line of the available dataset is divided into two parts: the training phase and the forecast phase. The performance of the individual models is obtained in the training phase using multiple linear regression against observed (analysis) fields. The outcome of this regression is the weights assigned to the individual models in the ensemble, which are then passed on to the forecast phase to construct the superensemble forecasts.

The temporal anomalies of a variable, rather than the full fields, are used in the multiple regression technique. Hence, in formulating the superensemble forecast, the weights are multiplied to the corresponding model anomalies. The constructed forecast is

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

where $\overline{O}$ is the observed climatology; a_i is the weight for the ith member in the ensemble; and F_i and $\overline{F}_i$ are the forecasts and forecast climatological values, respectively, for the ith model's forecast. The summation is taken over the N member models of the ensemble. Note here that the climatology $\overline{O}$ and the anomaly for a particular season are from different sources. We have calculated S for every season (52 in total) of the forecast period and then calculated the climatology $\overline{S}$ that was used in this study. Now, since the anomalies calculated by the superensemble for different years are independent to each other, their sum over the available years of forecasts need not equate to zero. Therefore, the climatology of a particular season created from S [of Eq. (1)] need not be identical to $\overline{O}$. Poor performance of the superensemble during only a single year of the forecast period can make $\overline{S}$ different than $\overline{O}$.

Therefore, the superensemble climatology is not equal to the observed climatology.

The weights a_i are obtained by minimizing the error term G, written as

$$G = \sum_{t=1}^{N_{\text{train}}} (S'_t - O'_t)^2, \tag{2}$$

where N_{train} is the number of time samples in the training phase, and S'_t and O'_t are the superensemble and observed field anomalies respectively at training time t. This exercise is performed for every grid point and vertical level in the dataset during every forecast phase. In other words, one weight is given to every model at every grid point in the three-dimensional space for each forecast. All together there were about 1.7×10^6 weights obtained for the forecasting of a single-level parameter. It was found that the weight for a model over a grid and for a particular variable does not vary significantly from one year to the other year of forecast. In other words, the superensemble weights are stable and this methodology can be considered robust for prediction in seasonal time scale.

The superensemble forecast differs from the conventional bias-removed ensemble mean forecast in that the weights for the superensemble are not $1/N$. Putting $a_i = 1/N$, for $i = 1, 2, \ldots, N$ in Eq. (1), one would obtain the bias-removed ensemble mean. The superensemble forecasts are made in selecting the assigning weights and seen to perform better than the bias-removed ensemble mean forecast (Stefanova and Krishnamurti 2002). Although superensemble provides a deterministic forecast, it has been shown by Stefanova and Krishnamurti (2002) that the superensemble algorithm carries an equivalent probabilistic forecast as well.

b. Synthetic superensemble methodology

This variant of the superensemble methodology was shown to provide better forecast skill than that of the conventional superensemble (Yun et al. 2005). The synthetic superensemble forecast is obtained when synthetic datasets are used in the superensemble algorithm in place of the member model forecast datasets. One set of synthetic data is created corresponding to each model by a linear regression with the observed (analysis) field in the EOF space. This technique is described below.

The observational data can be written as a linear combination of the principal component (PC) and the spatial EOFs:

$$O(x, t) = \sum_{k=1}^{M} P_k(t)\Phi_k(x), \tag{3}$$

where $P_k(t)$ is the PC time series, $\Phi_k(x)$ denotes the EOF component of the kth mode, and M is the number of modes selected. In the present study, M was chosen such that the EOFs explain 99% of the variance in the actual dataset. The results are not dependent on this number as long as the variance explained by the EOFs is high (>95%).

Similar to the observations, the model datasets are also expanded in terms of PC time series and EOFs:

$$F_i(x, t) = \sum_{k=1}^{M} F_{i,k}(t)\phi_{i,k}(x), \tag{4}$$

where i denotes the ith-member model in the ensemble, $F_{i,k}(t)$ and $\phi_{i,k}(x)$ are the PC time series and spatial EOF, respectively, for the kth mode, and M is the number of modes selected.

Now, a consistent pattern of the forecast and observation data is created by multiple linear regression. The kth mode of the PC time series of the observations is written as a linear combination of the kth mode of the PC time series of different models with an error term:

$$P_k(t) = \sum_{i=1}^{N} \alpha_{i,k}F_{i,k}(t) + \varepsilon_{i,k}, \tag{5}$$

where i denotes the ith member in the ensemble, $\alpha_{i,k}$ is the weight of the ith model for mode k, and $\varepsilon_{i,k}$ is the error term. The αs are estimated using a multiple linear regression, which minimizes the error variance $E(\varepsilon^2)$. After the weights are computed, the regression-improved PC time series of the kth mode for the ith model is defined as

$$F^{\text{reg}}_{i,k}(t) = \alpha_{i,k}F_{i,k}(t),$$

and the synthetic data for the ith-member model is defined as

$$F^{\text{reg}}_i(x, t) = \sum_{k=1}^{M} F^{\text{reg}}_{i,k}(t)\Phi_k(x), \tag{6}$$

where $\Phi_k(x)$ represents the EOF components of the observations. Once the synthetic data $[F^{\text{reg}}_i(x, t)]$ are generated in this manner for every model of the ensemble, these are then passed on to the input of the conventional superensemble algorithm to construct what is called a synthetic superensemble forecast. The synthetic superensemble algorithm is schematically described in Fig. 1. It was found that the synthetic superensemble performs better than the conventional superensemble (Yun et al. 2005). In this paper the phrases *superensemble* and *synthetic superensemble* will be used interchangeably to refer to the synthetic superensemble forecasts. One of the main purposes of this paper is to

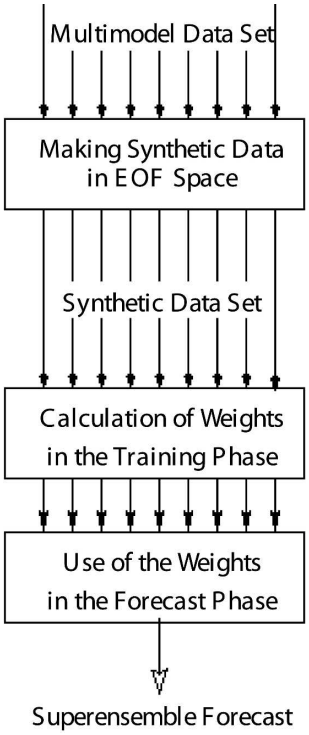

FIG. 1. The algorithm for making the SSE forecast. The multimodel dataset is passed on to a multiple regression in the EOF space to create the synthetic data for every member model. These newly created synthetic data are used in the training phase to calculate the weights of different models. These weights are used in Eq. (1) to construct the superensemble forecast.

show that this procedure seems to provide a rather major improvement over those provided by single-member models and the ensemble mean.

c. Models

As many as 13 state-of-the-art atmosphere–ocean CGCMs were used for the construction of the superensemble. The characteristics of these models, along with the acronyms in referencing them, are listed in Table 1.

Four versions of the Florida State University Global Spectral Model (FSUGSM) were run with four different combinations of two cumulus convection schemes (viz., the Arakawa–Schubert and the Kuo) and two radiation schemes. The version of the Arakawa–Schubert scheme used here is a simplified form of the original scheme (Arakawa and Schubert 1974), introduced by Grell (1993). The Kuo convection scheme (Krishnamurti et al. 1980; Krishnamurti and Bedi 1988) is a modified version of that described in Anthes (1977). The *new radiation scheme* used in the FSUGSM considers band models. In addition to that, an *old radiation scheme*, that is, an emissivity–absorptivity-based one (Krishnamurti et al. 2002), is also being used here. The atmospheric component of the FSUGSM was run at the horizontal resolution of T63 (triangular truncation at 63 waves) and 14 vertical levels. The oceanic component was adapted from Hamburg Ocean Primitive Equation (HOPE) global model (Wolff et al. 1997), which has a resolution of 5° in longitude and 0.5°–5.0° in latitude with higher resolution over the equatorial latitudes. The oceanic initial conditions were obtained from a comprehensive coupled assimilation where the initialized SSTs were relaxed to observed SST (Krishnamurti et al. 2002).

A coupled version of the Community Climate Model, version 3 (CCM3), from the National Center for Atmospheric Research (NCAR) was also included in this suite of CGCMs. The atmospheric component of this model was run at the T63 resolution and with 18 vertical levels. Construction of the oceanic dataset was carried out in the same manner as for the four FSU models.

A Predictive Ocean Atmosphere Model for Australia-1 (POAMA1) is also included in this multimodel suite. The atmospheric component of the POAMA1 is run at an R47 resolution (rhomboidal truncation at 47 waves) and it has 17 vertical levels. A detailed description of this model can be found in Wang et al. (2004b).

Finally, the seven member models of the Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction (DEMETER) project were included in this suite. The DEMETER consists of a variety of models from different parts of Europe. A description of this project is provided in Palmer et al. (2004). There were nine members in an ensemble simulation for each model of the DEMETER family. An average over these nine members was taken before using these data in the superensemble.

d. Datasets, construction of the superensemble, and the regions of study

The monthly averages of the different parameters from the 13 CGCMs were available for this study. The common time span for these forecast datasets covers the years 1989–2001. Seasonal mean values were used here. The forecast of a particular model was averaged over four seasons of a year (March–May, June–August, September–November, and December–February). Thus, 52 seasonal forecasts were obtained from each model over the 13 yr of simulations. The DEMETER forecasts started on the 1st of February, May, August, and November of each year and was carried out for 6 months. The forecasts for the first month were not used in our analysis. The seasonal forecasts used in the present study were obtained by averaging the next

TABLE 1. Salient features of the 13 coupled atmosphere–ocean general circulation models used in the analysis.

Name (source)	Atmospheric component			Oceanic component		
	Model	Resolution	Initial condition	Model	Resolution	Initial condition
ANR (FSU)	FSUGSM with Arakawa–Schubert convection and new radiation (band model)	T63L14	ECMWF with physical initialization	HOPE global	5° longitude, 0.5°–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
AOR (FSU)	FSUGSM with Arakawa–Schubert convection and old radiation (emissivity–absorptivity based)	T63L14	ECMWF with physical initialization	HOPE global	5° longitude, 0.5°–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
KNR (FSU)	FSUGSM with Kuo convection and new radiation (band model)	T63L14	ECMWF with physical initialization	HOPE global	5° longitude, 0.5°–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
KOR (FSU)	FSUGSM with Kuo convection and old radiation (emissivity–absorptivity based)	T63L14	ECMWF with physical initialization	HOPE global	5° longitude, 0.5°–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
CCM3 (NCAR)	CCM3 atmospheric model	T63L18	ECMWF	Slab Ocean Model SOM	2.4° × 1.2°–2.4°	Coupled assimilation
POAMA1 (Australia)	Bureau of Meteorology Research Centre (BMRC) Atmospheric Model (BAM3)	R47L17	From latest atmosphere and ocean conditions from Global Atmospheric Sampling Program	Australian Community Ocean Model 2 (ACOM2)	−2° × 0.5°–1.5°, 25 levels	From ocean assimilation that was based on optimum interpolation (OI) technique.
CERFACS (France)	ARPEGE	T63L31	ECMWF 40-yr Re-analysis (ERA-40)	OPA 8.2	2° × 2°, 31 levels	Forced by ERA-40
ECMWF (Europe)	IFS	T95L40	ERA-40	HOPE-E	1.4° × 0.3°–1.4°, 29 levels	Forced by ERA-40
INGV (Italy)	ECHAM-4	T42L19	Coupled AMIP type	OPA 8.1	2° × 0.5°–1.5°, 31 levels	Forced by ERA-40
LODYC (France)	IFS	T95L40	ERA-40	OPA 8.2	2° × 2°, 31 levels	Forced by ERA-40
MPI (Germany)	ECHAM-5	T42L19	Coupled run relaxed to observed SST	MPI Open Model Interface (MPI-OMI)	2.5° × 0.5°–2.5°, 23 levels	Coupled run relaxed to observed SST
MetFr (France)	ARPEGE	T63L31	ERA-40	OPA 8.0	182 × 152 GP, 31 levels	Forced by ERA-40
UKMO (United Kingdom)	HadAM3	2.5° × 3.75°, 19 levels	ERA-40	GloSea OGCM Third Hadley Centre Coupled Ocean–Atmosphere GCM (HadCM3) based	1.25° × 0.3°–1.25°, 40 levels	Forced by ERA-40

three months of the forecast. Hence, these seasonal forecasts can be termed as one-month lead time forecasts. The observational datasets correspond to Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997) for precipitation and the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for other variables. All the datasets were interpolated to a common 2.5° × 2.5° horizontal grid prior to the construction of the superensemble.

A cross-validation technique (Déqué 1997) was used for constructing the superensemble forecasts for a particular year. The year to be forecast (4 seasons) was excluded from the dataset and the weights of the superensemble were computed based on the performance of the models during the rest of the years (48 seasons), called the training period. These weights were then applied to the individual member models for the construction of the superensemble forecasts [Eq. (1)]. While comparing anomalies of different models, annual cycle and bias of individual models have been removed. In the case of superensemble, at first forecasts were constructed for all the 52 seasons. The annual cycle and bias were removed from this time series (likewise for the member models) at the time of comparison of anomalies.

Figure 2 shows the south Asian monsoon region (covering the area 10°S–35°N, 50°–110°E) over which the forecast skills of the coupled models and the superensemble are compared. All the major features of the south Asian monsoon are encountered over this domain. The shading in this figure shows the observed (Xie and Arkin 1997) climatology in precipitation during June–August. Guided by this spatial distribution of climatological precipitation we have selected four subdomains (indicated by a, b, c, and d in the figure) of the monsoon domain to assess the performance of the superensemble on regional scales. These four regions have different characteristics for the monsoon precipitation. Region a mainly encompasses the central Indian landmass, region b encompasses an area of heavy rainfall over the Bay of Bengal, region c is dominated by the orographic rainfall and the onset phase of Indian monsoon, and region d is at the equatorial Indian Ocean where the intertropical convergence zone forms and propagates north toward the Indian subcontinent. Another domain (partially outside of this monsoon region), which covers 20°S–0°, 80°–110°E, was chosen to examine the performance of the superensemble during extreme Indian Ocean dipole (IOD)-related events.

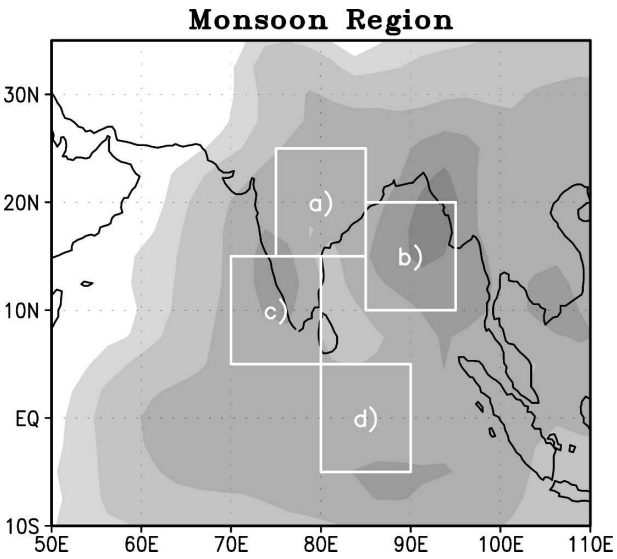

FIG. 2. The monsoon region (10°S–35°N, 50°–110°E) and the four subregions selected for precipitation forecast analysis. The shadings represent climatological June–August precipitation from CMAP (Xie and Arkin 1997). Shadings are at 1, 2, 5, 10, and 15 mm day^{-1} (darker shades for higher precipitation rates).

3. Mean monsoon and model bias

It is important to assess the simulated climatology of a GCM before investigating its forecast for a particular season. Figure 3 shows the spatial pattern of climatological precipitation during June–August from observations (Xie and Arkin 1997), 13 CGCMs, their ensemble mean (EM), and the synthetic superensemble (SSE). The spatial correlation coefficient with observations for different models is indicated at the top-right corner of each panel. These were calculated using the following formula:

$$\mathrm{SC} = \frac{\sum_{i=1,j=1}^{i=N_x,j=N_y} (F_{i,j} - \langle F\rangle)(O_{i,j} - \langle O\rangle)}{\sqrt{\sum_{i=1,j=1}^{i=N_x,j=N_y} (F_{i,j} - \langle F\rangle)^2}\sqrt{\sum_{i=1,j=1}^{i=N_x,j=N_y} (O_{i,j} - \langle O\rangle)^2}}, \tag{7}$$

where $F_{i,j}$ and $O_{i,j}$ is the value of the parameter at grid location (i, j) from model output and observations, respectively; $\langle F\rangle$ and $\langle O\rangle$ are the area average of model output and observations, respectively, over the domain; and N_x and N_y are the number of grids in the x and y directions, respectively. The spatial correlation is a measure of the forecast efficiency of the models in simulating the spatial pattern of the parameter over the domain. Note from Fig. 3 that most of the models fail to capture the key spatial features of precipitation over the monsoon region. For example, the precipitation maxima over the northern Bay of Bengal and over the south–west coast of the Indian peninsula is missing in some of the models. Some models overestimate these precipitation peaks. The models that could capture the climatological features of rainfall distribution over this region realistically [viz., POAMA1, Istituto Nazionale di Geofisica e Vulcanologia (INGV), the U.K. Met Office model (UKMO), and the EM] carry higher spatial correlation (more than 0.80). The overall spread from

these coupled models is more than what was noted for the Atmospheric Model Intercomparison Project (AMIP) runs (Krishnamurti et al. 2000b). In the case of the SSE, the spatial correlation is very close to one. Sperber and Palmer (1996) argued that the mean and the interannual variation of a GCM will be better if its climatology is close to the observations. This suggests that the seasonal forecast obtained from the superensemble might be more reliable than any other model of the suite and their ensemble mean since it represents the climatology extremely well (see the first and the last panels of Fig. 3).

4. Seasonal forecasting

a. Precipitation

In Fig. 4 seasonal precipitation during June–August of 1989–2001 from the 13 CGCMs, their EM, and the superensemble are correlated with the observations for the south Asian monsoon domain. The correlation coefficients between the observed and predicted precipitation are indicated inside the panels. For most of the models and the ensemble mean the correlation coefficient ranges from 0.4 to 0.8, whereas for the superensemble it is 0.94. In particular, large spread in the data was noticed for ANR, AOR, KNR, KOR (see Table 1 for acronym definitions), and CCM3 (coefficient $\leq$0.45). All these models tend to predict higher rainfall when the observed rainfall is low and lower rainfall when the observed rainfall is high. The DEMETER models showed somewhat better forecast skill compared with the other models in the suite (coefficient $\geq$0.60). Note that the rainfall distribution pattern for the Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS) and Météo-France (MetFr) models are similar. Both of these models predict rainfall too high when the actual value (observed) is low ($<$10 mm day^{-1}) and reach a saturation for higher values of observed rainfall ($>$10 mm day^{-1}). It appears that this similarity is due to the use of similar atmosphere [Action de Recherche Petite Echelle Grande Echelle (ARPEGE)] and ocean [Océan Parallélisé (OPA)] components in these two CGCMs (see Table 1). The UKMO, which used the Third Hadley Centre Atmospheric Model (HadAM3) and a global seasonal (GloSea) ocean GCM (OGCM), showed a different pattern (and somewhat higher skill) in the simulated precipitation. On the other hand, rainfall distributions from ECMWF and Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC) show a similar pattern. For both of these models, a higher than observed rainfall is noticed in the lower ranges ($<$10 mm day^{-1}) and lower to equal amount of rainfall forecast was noticed for higher ranges ($>$15 mm day^{-1}) of observed precipitation. Note that both these models used the same atmospheric component [Integrated Forecast System (IFS); Table 1], but a different oceanic component. For the other two DEMETER members [INGV and Max Planck Institute (MPI)], the frequency of rainfall at different intensities is somewhat similar. Both of these models used two different versions of the same atmospheric model (ECHAM 4 and 5; see Table 1). Hence, it appears that the model biases of the CGCMs in predicting precipitation are dependent on the atmospheric component of the model. Figure 4 also shows that the EM of these models tends to predict higher than observed rainfall in the low rainfall regimes ($<$5 mm day^{-1}) and lower than observed rainfall in the high rainfall regimes ($>$15 mm day^{-1}). The prediction of the superensemble (correlation coefficient = 0.94) was better than the member models and the EM in all intensity ranges of seasonal precipitation over the south Asian monsoon domain.

Figure 5 shows the root-mean-square (rms) errors of the precipitation anomaly forecast over the south Asian monsoon region (10°S–35°N, 50°–110°E) during June–August of 1989–2001 from the 13 different models, for the bias-removed EM (BREM) and for the SSE. Also shown are the mean rms errors from 13 yr of simulations. The significance level at which the rms error of the superensemble is superior to that of the BREM is expressed in percent and is indicated at the top of the panels for the respective year of the forecast. A Student's t test was used to calculate the significance level (see the appendix). The anomalies of all forecasts were calculated by removing the respective forecast climatology for that season from the actual forecast. In calculating the forecast climatology, the year to be forecast was not included since that might bias the climatology. The BREM was calculated by averaging all the model anomaly forecasts [applying a weight of $1/N$ to all models in the Eq. (1)]. It is noted that during the Northern Hemisphere (NH) summer season individual member models show poor performance compared with the BREM or SSE forecasts during most of the years of simulations. The BREM is generally better than the member models of the suite. However, it is worth noting that the BREM always has skill lower than those of the multimodel superensemble. This is because poorer models have same weight as the better models (after the bias removal) when the BREM is calculated. During 7 out of these 13 yr, the difference of rms error between the BREM and the SSE was significant at more than 95% confidence level. It was found that during the years when the difference between the BREM

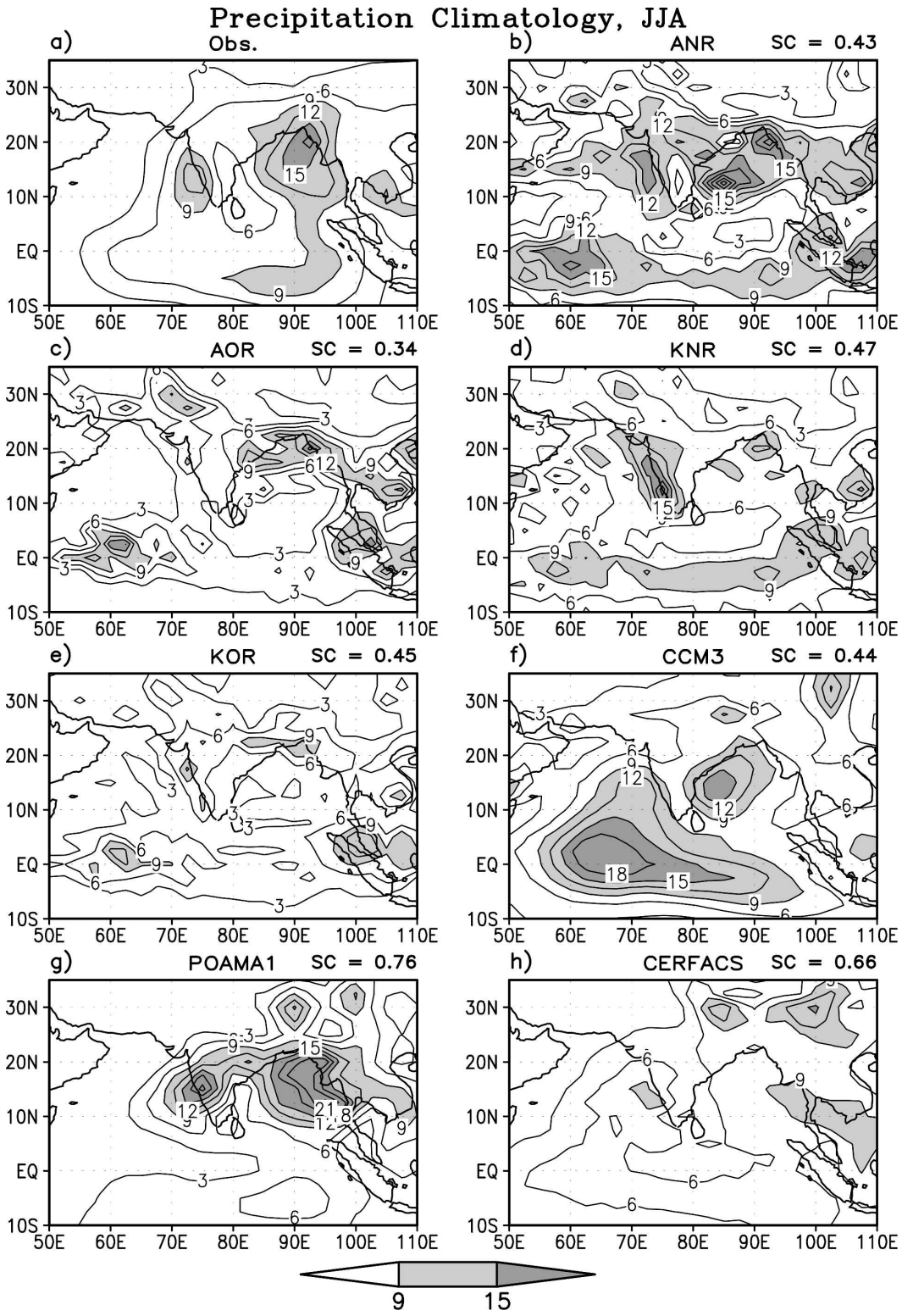

FIG. 3. Precipitation climatology (mm day^{-1}) during June–August 1994 from observations and forecasts of 13 coupled models, their EM, and the SSE. The spatial correlation with observations for different models is shown as numbers at the top-right corner of each panel.

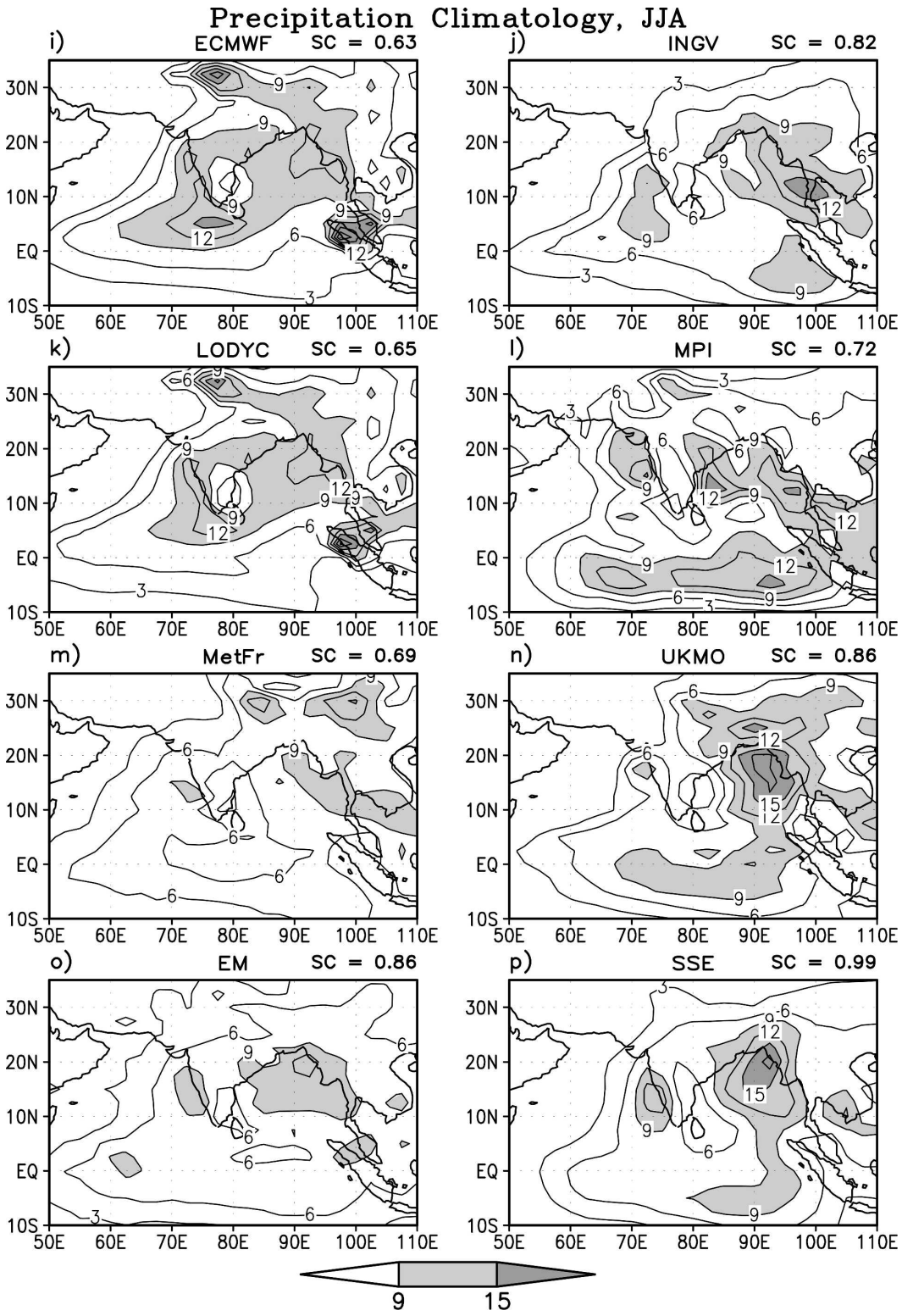

FIG. 3. (*Continued*)

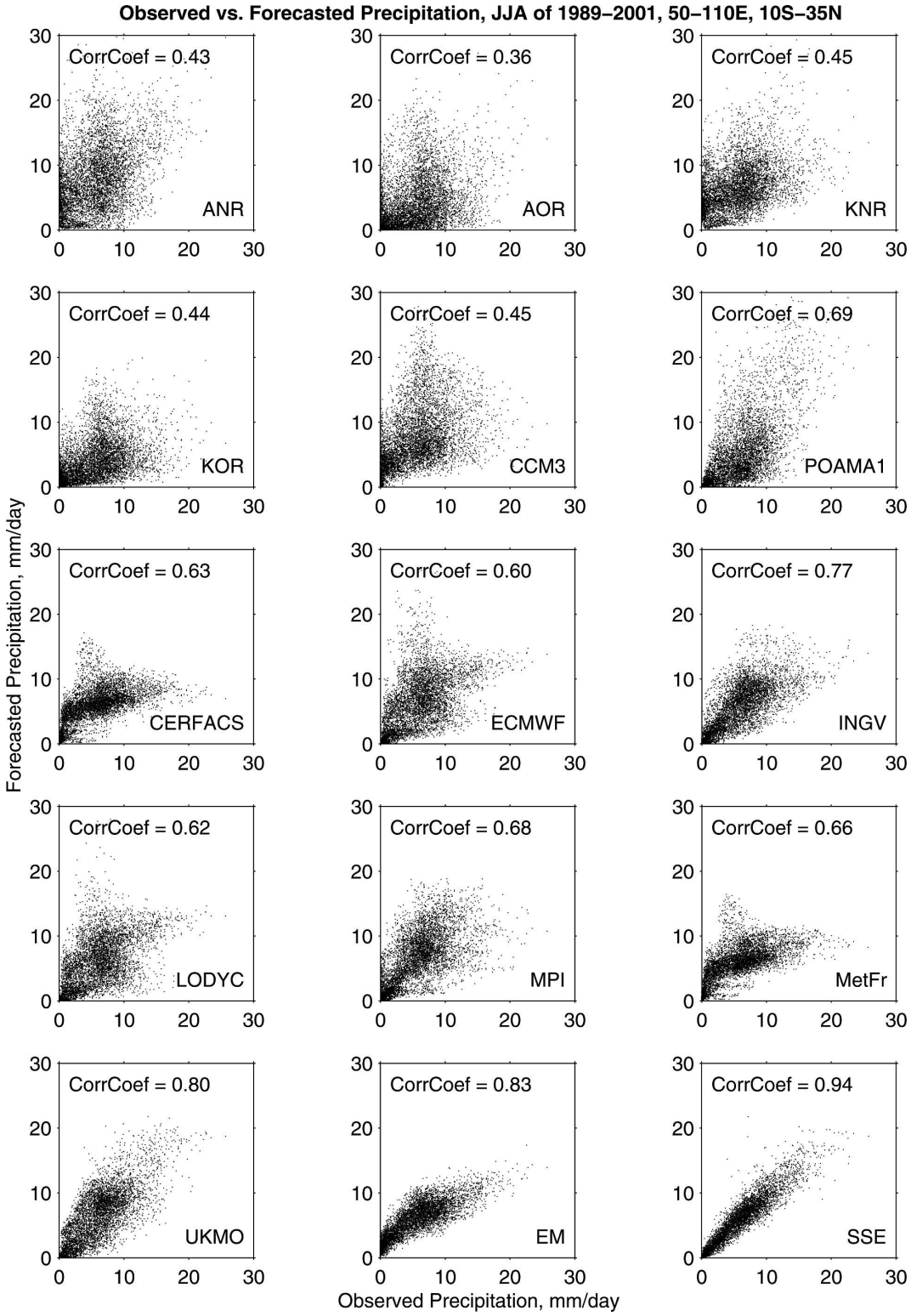

FIG. 4. Relation between simulated and observed precipitation (grid by grid) over the south Asian monsoon domain during June–August of 1989–2001 from 13 CGCMs, their EM, and the superensemble. The correlation coefficient between the simulated and observed precipitation is indicated at the top of the panels.

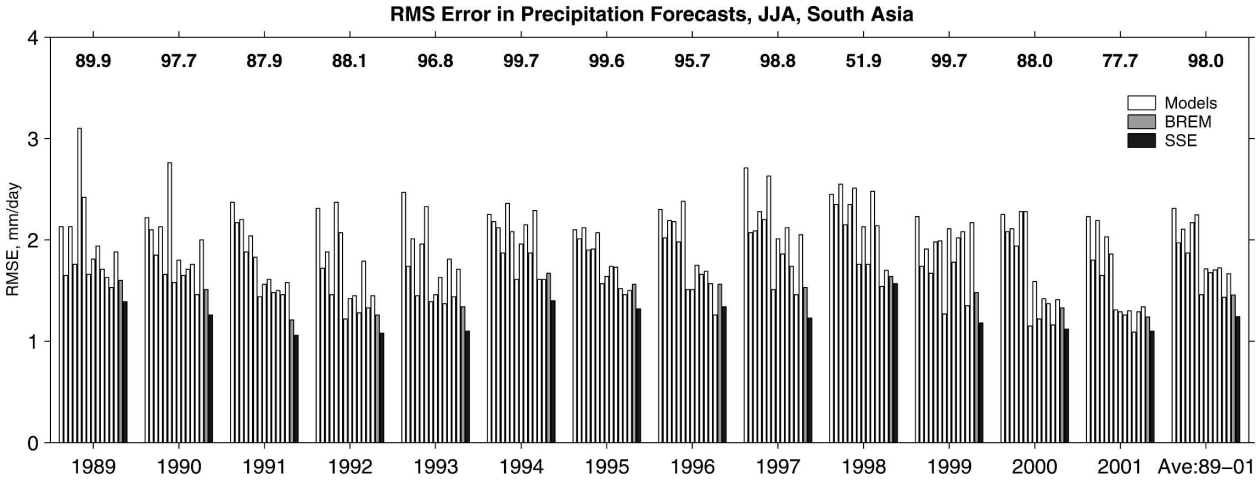

FIG. 5. The rms error in precipitation anomaly forecasts (in mm day^{-1}) during June–August over the region 10°S–35°N, 50°–110°E from 13 coupled models, their BREM, and the SSE. Also shown is the significance level (in percent) by which the SSE rms error differs from that of the BREM using a Student's *t* test (see appendix).

and superensemble was not highly significant, either the standard deviation within the member models was high (e.g., in 1989 and 1992) or the difference between the BREM and the superensemble was low (e.g., in 1998 and 2001). On an average (Ave:1989–2001) the superensemble anomaly forecast was better than the BREM forecast with a confidence level of 98%. Overall, this result suggests that the seasonal anomaly forecasting skill from the superensemble approach is higher compared with the member models within the ensemble and it is robust with the interannual variation of seasonal precipitation.

While rms error of the anomaly is a skill to measure the amplitude of the predicted anomaly, anomaly correlation (AC) is a measure of the skill to predict the sign of the anomaly. In Fig. 6a the AC over the south Asian region during June–August of 1989–2001 from 13 CGCMs, their EM, the conventional superensemble (CSE), and the SSE are shown. No single model was able to show a consistent high AC during the 13 yr. This result clearly shows that high confidence in seasonal forecasts cannot be achieved using a single model. Considering the average over 13 yr, the highest AC among the models was 0.24 (not shown). Even the EM of the

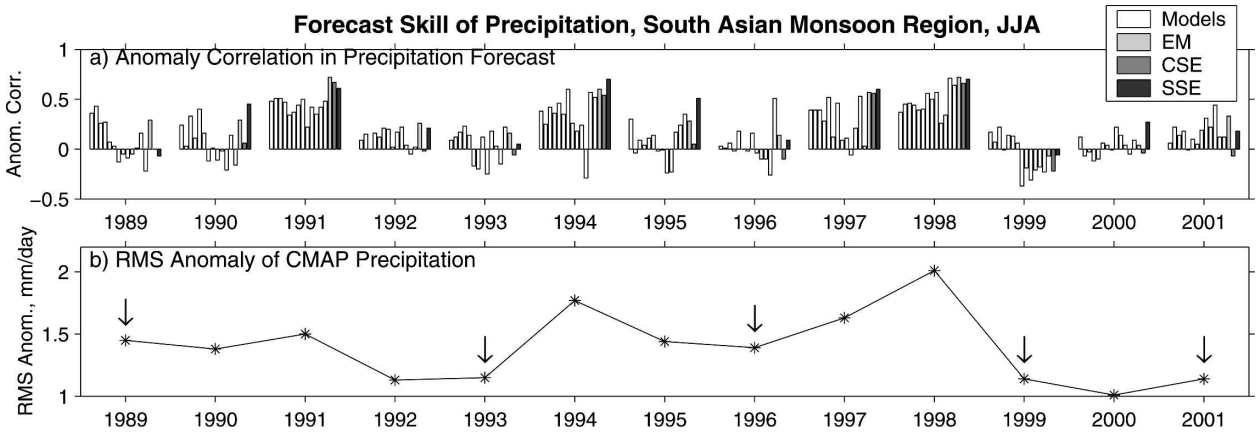

FIG. 6. (a) The AC in precipitation forecasts (mm day^{-1}) during June–August over the region 10°S–35°N, 50°–110°E from 13 coupled models, their EM, the CSE, and the SSE; (b) rms anomaly of observed (CMAP; Xie and Arkin 1997) precipitation over the same domain. A larger value of this parameter indicates a larger anomaly (positive or negative). The years marked by the arrows are characterized by particular low anomalies. During those years, the SSE AC is not high. Whereas during years such as 1994, 1997, and 1998 when the anomaly in the observations was high over the region, the SSE AC was also significantly higher compared with the other years.

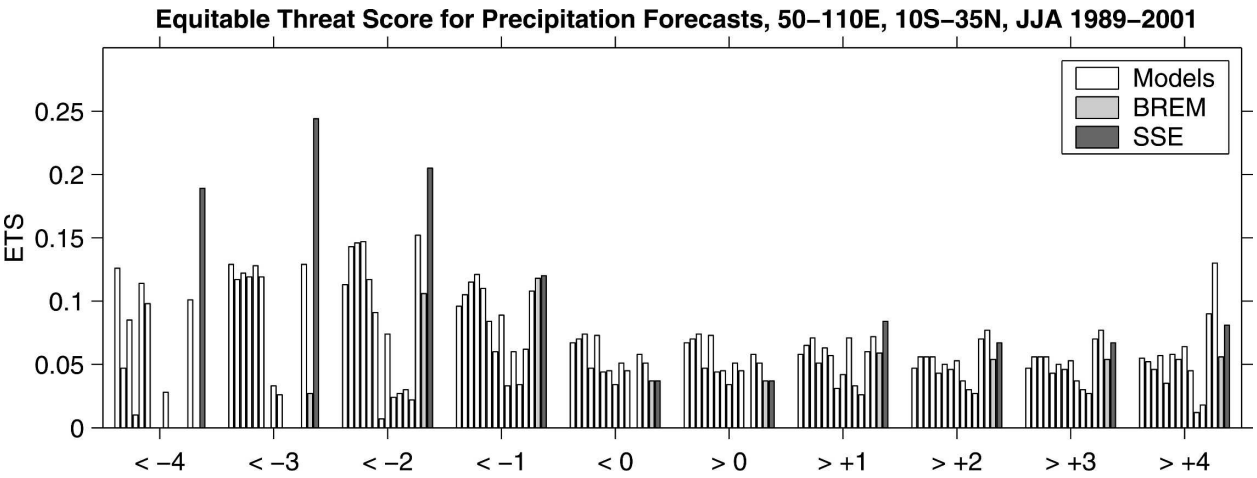

FIG. 7. The ETS of precipitation forecasts during June–August of 1989–2001 over the region 10°S–35°N, 50°–110°E from 13 coupled models, their BREM, and the SSE. The ETS was calculated from the anomaly of every dataset. The anomaly for a season was calculated by subtracting the climatology of that season from the actual value, while the climatology for that particular season was calculated by averaging over all available years but the forecast year. ETS from SSE shows particular high skill when the precipitation anomaly is highly negative (<-2 mm day^{-1}) and positive (>1 mm day^{-1}). This is consistent with the results of Fig. 6 that the SSE forecast skill is higher than other forecasts over the south Asian region particularly when the anomaly of precipitation is high.

models does not show a consistent and high AC during all the years (average AC = 0.33). There were 3 yr during this period when EM showed the highest AC among all the forecasts. On the other hand, during 5 yr out of this 13 yr of simulations, the AC from the SSE was higher than the member models, the CSE, and the EM. The 13-yr mean AC for the SSE was 0.33 (not shown). Note that the SSE performed better than the CSE forecast in 12 out of the 13 yr of simulations (except 1991). This shows the superiority of the SSE forecasts over that of the CSE, which is consistent with that reported in Yun et al. (2005).

It was found that AC from the SSE was particularly high when the anomaly of precipitation over this region was high. This is illustrated in Fig. 6b. This figure shows the rms anomaly of observed (CMAP) precipitation during June–August of the 13 yr. The rms anomaly was defined as

$$\text{rms anomaly} = \sqrt{\frac{\sum_{\text{Domain}} (O - \overline{O})^2}{N}},$$

where $\overline{O}$ is the climatology of observations O, and N is the total number of grid points. The summation is taken over all the grid points in the domain. A high value of this number indicates a high precipitation anomaly (positive or negative) year. It was found that during the years (such as 1989, 1993, 1996, 1999, and 2001; indicated by arrows in the figure) when the anomaly in the observed precipitation was low, the performance of the SSE was not satisfactory. The SSE performed particularly well when this anomaly was high and the AC from the SSE was high (e.g., during 1994, 1997, and 1998). Therefore, it is possible to state that the skill of the SSE was particularly high when the precipitation over this region experienced some extrema (positive or negative). It is very intriguing to say that one can depend on the SSE forecast more than any other single model forecasts or their EM forecast if the season is expected to show a high precipitation anomaly. In other words, the SSE performs better during an extreme year as compared with a normal year for precipitation forecasts over the south Asian monsoon region. This is further confirmed by analyzing the probabilistic skill scores of precipitation forecasts (Fig. 7).

In Figs. 8 and 9, the spatial patterns of the anomalies in precipitation are examined. Figure 8 shows the precipitation anomaly during June–August 1994 from observations, all the member models and their EM, and the superensemble. The AC of the forecast from different simulations with the observations over this domain is written at the top-right corner of each panel. The year 1994 was chosen because it was characterized by a positive IOD event (Saji et al. 1999), which is known to have impacts on precipitation over the Indian landmass (Ashok et al. 2001; Gadgil et al. 2003). The main features of the precipitation pattern during this year were the anomalous lower precipitation over the east equatorial Indian Ocean, anomalous higher precipitation over the west equatorial Indian Ocean, and anomalous higher precipitation over the Indian land-

mass and northern Bay of Bengal (Fig. 8a). It is clear from this figure that most of the models were not able to capture these anomalous patterns of precipitation over the south Asian region for this year. For some models the precipitation minimum over the eastern Indian Ocean was too close to the equator (ECMWF, LODYC), whereas for some other models a weak minima (CERFACS, MetFr) was noted. Moreover, many models were unable to capture the positive precipitation anomaly over the western Indian Ocean (ECMWF, LODYC, MPI, UKMO, KOR). By comparing Figs. 8a and 8p it can be said that the superensemble forecasts were closer to observations over both the east and west equatorial Indian Ocean compared with any other model and their EM. Figure 8 also shows that the AC during 1994 from most of the models was poorer (<0.50). Only 3 of the 13 models and the EM have an AC more than 0.50. Note again that the superensemble outperformed the skill of all the models and their EM with an AC of 0.62.

In Fig. 9 another example is shown for the seasonal forecasts of monsoon precipitation anomalies for the year 2000. The forecast skill measured by the rms error for most of the models is around 4 to 5 mm day^{-1}. The results for the superensemble are around 1.05 mm day^{-1}. The AC for most models varies from around ± 0.05 to ± 0.10, whereas that of the superensemble is around 0.27. Although the superensemble skill is higher than those of the member models, an AC of 0.27 is somewhat low. The corresponding correlation for the total rainfall for the superensemble is of the order of 0.9. Thus, the superensemble is able to predict the climatology of total rains quite closely. The magnitude of the anomaly for the observed rains over the land area during the 2000 summer monsoon season was around 1 mm day^{-1}. Over the region of the equatorial eastern Indian Ocean, the anomaly rain is of the order of 1 mm day^{-1}. The superensemble simulations are realistically accurate over the equatorial eastern Indian Ocean. However, this positive anomaly belt stretches westward to 60°E where observed negative anomalies were prevalent. None of the models handled this western region accurately. In summary, we cannot say that those seasonal forecasts for the superensemble may be very useful to the agricultural community but that these are the best compared with what is available from the member models.

1) PROBABILISTIC SKILL SCORE

Categorical forecast metrics were calculated to find the probabilistic skill of the superensemble as compared with the member models and their unbiased EM. Equitable threat score (ETS; also known as the Gilbart skill score) is one such metric. ETS is defined in Table 2.

The ETS was calculated from the anomaly of every forecast and the observations of precipitation over the south Asian monsoon region. Anomaly for a season was created by subtracting the climatology of that season from the actual values. Whereas the climatology for the season was calculated by averaging over the values of all possible years excluding the current (forecast) year (i.e., for the year the anomaly was calculated). This method ensures that the year to be forecasted does not influence the climatology. Then precipitation anomalies of June–August seasons during the 13 yr were used to calculate the ETS. Ten different categories in the anomaly were considered. Those are anomalies less than -4, -3, -2, -1, 0 mm day^{-1}, and greater than 0, $+1$, $+2$, $+3$, $+4$ mm day^{-1}. The results from the 13 member models, their BREM, and the SSE are shown in Fig. 7. The SSE outperformed the member models and the BREM by a huge margin in extreme dry events of precipitation (<-4, <-3, <-2 mm day^{-1}). It performed as well as the BREM when precipitation anomalies were considered to be less than -1 mm day^{-1}. For more than normal rainfall, the SSE outperformed all the models and the BREM when the threshold was considered to be $+1$ mm day^{-1}. For $+2$ and $+3$ mm day^{-1} thresholds, the SSE performed as good as two other member models of the suite, although for both these cases ETS from the SSE was much higher than the other models and their BREM. If only positive or negative precipitation anomalies are to be considered (>0 and <0 cases), the performance of all member models, the BREM, and the SSE were comparable.

These results are consistent with that of Fig. 6. Both these figures show that in anomaly forecasts of precipitation, the SSE outperforms the member models and their EM particularly when the anomaly is high. For near-normal precipitation, it performs as good as the member models and the EM. Therefore, the SSE can be used for seasonal precipitation forecasts with high confidence when the forecasted season is expected to have a large departure from the normal values.

2) FORECAST SKILLS OVER SUBDOMAINS

Figure 10 shows the time series of error in seasonal precipitation forecasts over four smaller domains (10° latitude $\times$ 10° longitude) within the monsoon region. These subregions are labeled by a, b, c, and d in Fig. 2. The time series were constructed using 52 seasonal forecasts during the 13 yr of simulations from different models, their EM, and the superensemble. The observed precipitation (Xie and Arkin 1997) was subtracted from these time series to assess the errors of the

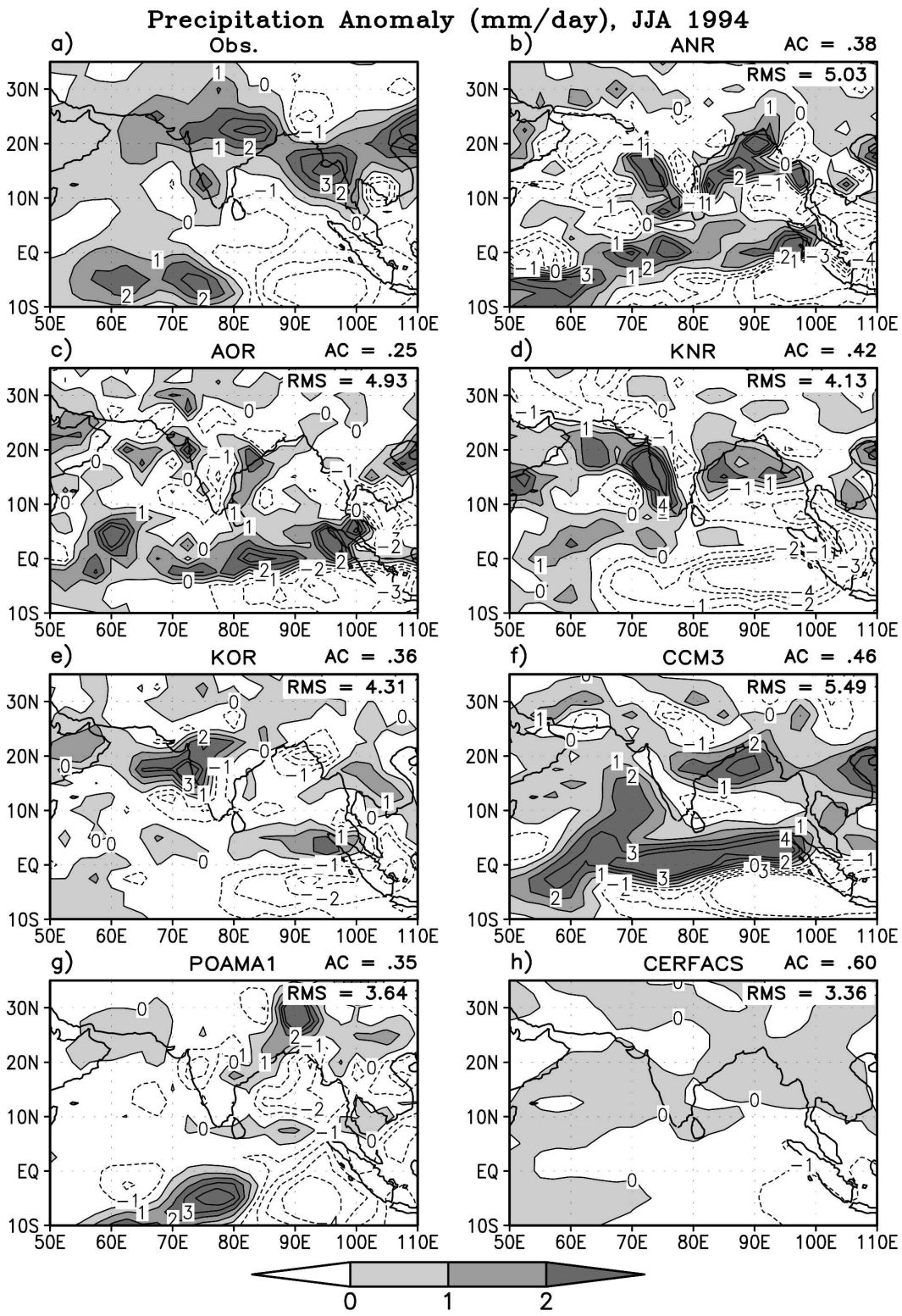

FIG. 8. Precipitation anomalies (mm day^{-1}) during June–August 1994 from observations and the forecast of 13 coupled models, their EM, and the SSE. The contour interval is 1 mm day^{-1} and positive anomalies are shaded. The anomaly correlations and rms errors for different models are given in the top-right corner of each panel.

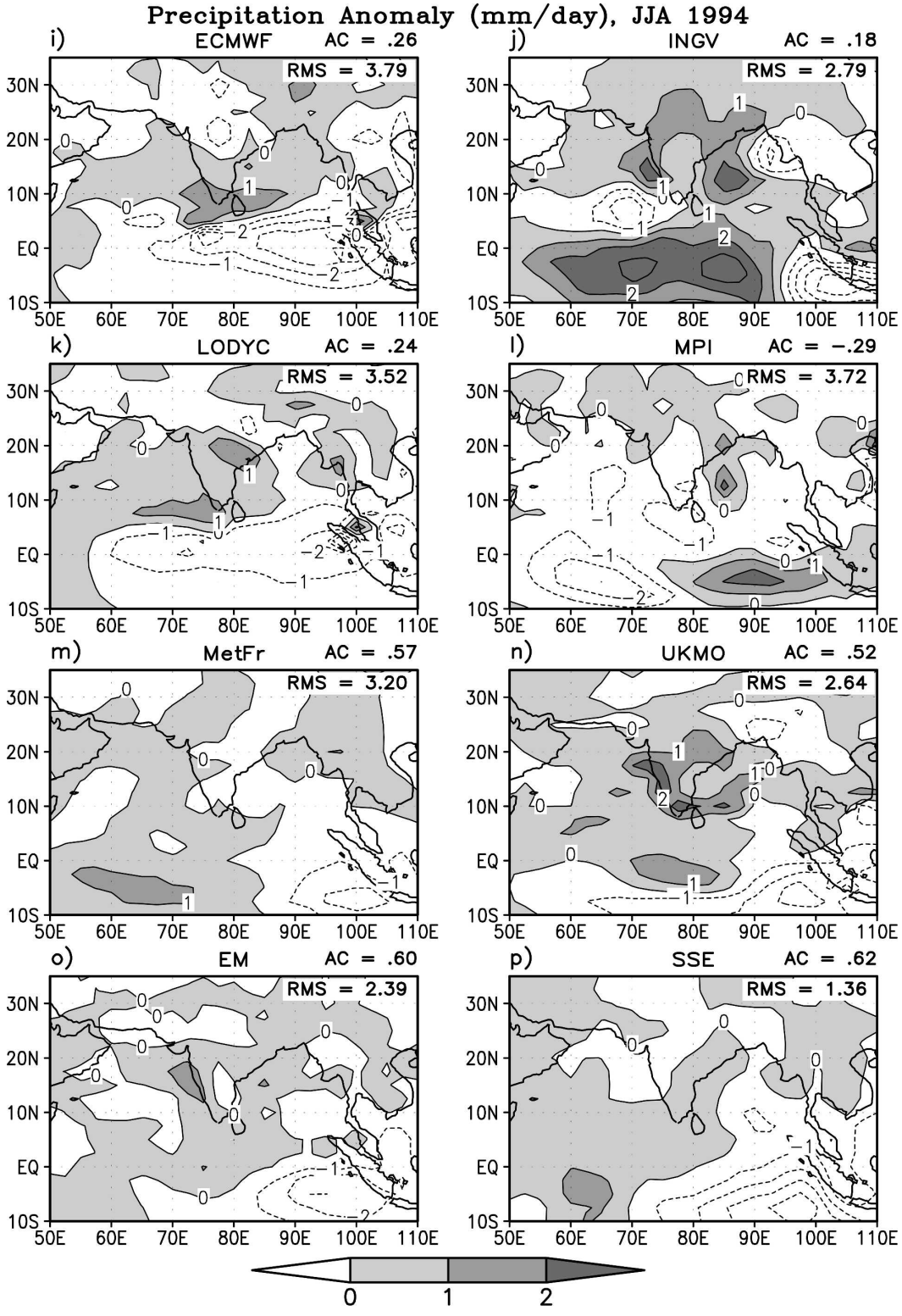

FIG. 8. (*Continued*)

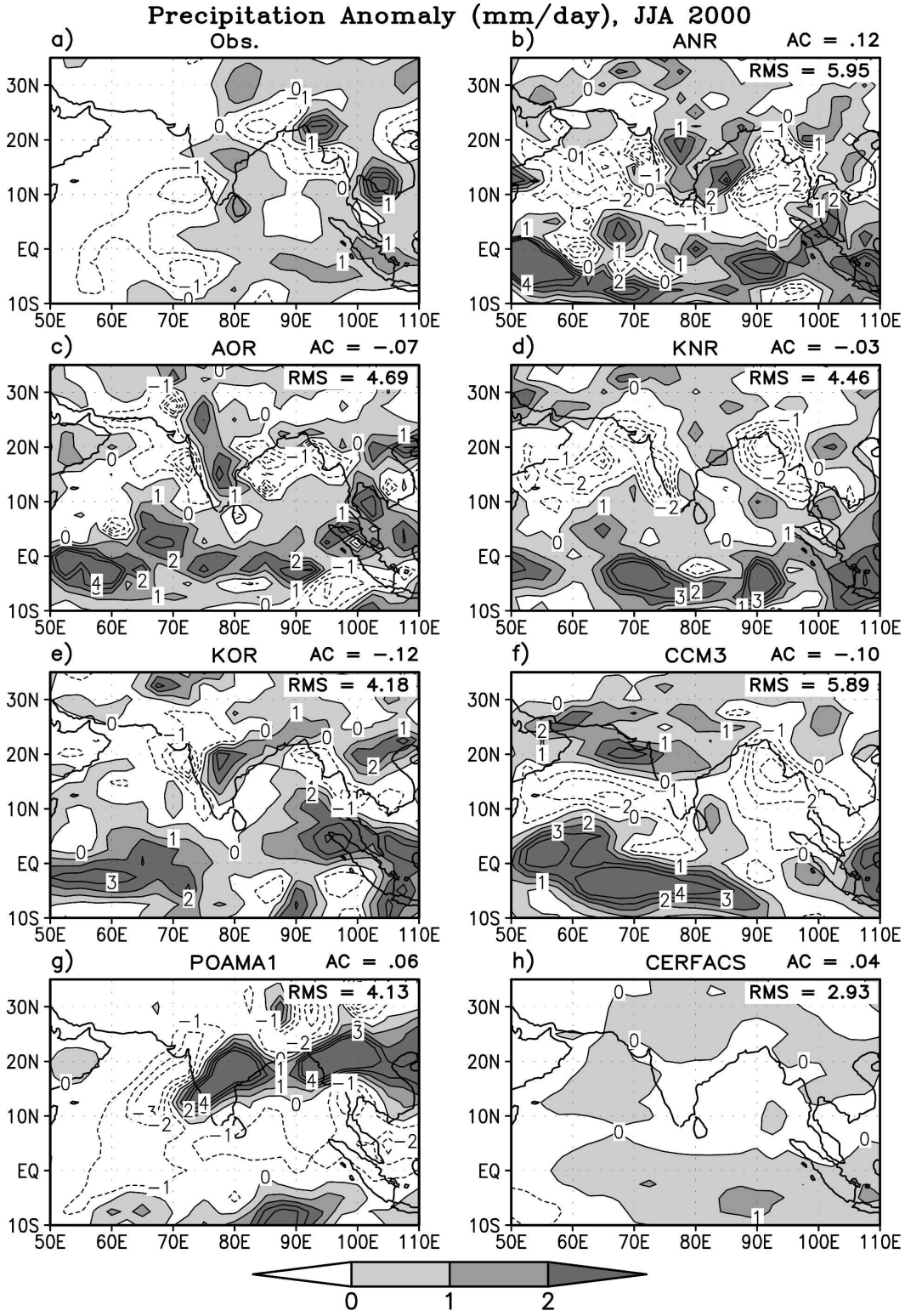

FIG. 9. As in Fig. 8, but for June–August 2000.

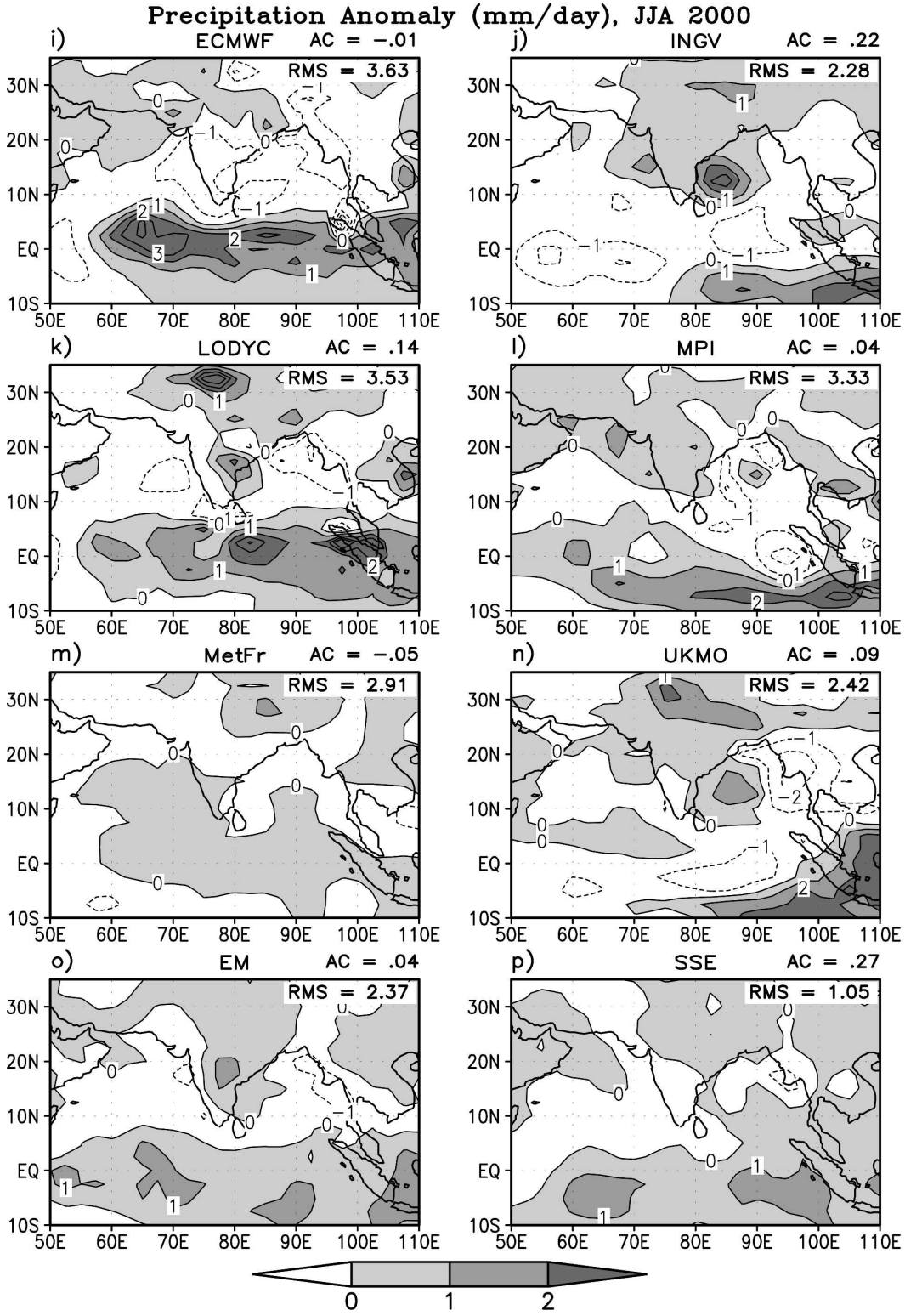

FIG. 9. (*Continued*)

TABLE 2. The definition of ETS (a probabilistic skill score for categorical forecasts).

		Observation		
		Yes	No	
Forecast	Yes	YY	YN	$\text{ETS} = \dfrac{\text{YY} - C}{\text{YY} + \text{NY} + \text{YN} - C}$, where $C = \dfrac{(\text{YY} + \text{YN})(\text{YN} + \text{NY})}{\text{N}}$, and
	No	NY	NN	$\text{N} = \text{YY} + \text{YN} + \text{NY} + \text{NN}$

seasonal forecasts. The main feature of this figure is the large spread in the forecast errors for the member models and relatively less errors for the EM and for the superensemble over all the four subdomains. Moreover, the magnitude of the errors varied during different seasons of the year. For the member models, the highest errors in forecasts are noticed during June–August over the Indian landmass (Fig. 10a), over the Bay of Bengal (Fig. 10b), and over the southwestern India (Fig. 10c). Consequently, the EM shows larger errors during June–August as compared with the other seasons. However, the forecast errors of the superensemble did not show any trends of being the largest during the NH summer season compared with other seasons over these regions. Moreover, superensemble forecasts were closer to the observational estimates as compared with all the member models and the EM during all of the seasons for the 13 yr of simulations over these three regions. Over the equatorial Indian Ocean (Fig. 10d) the member models show the largest

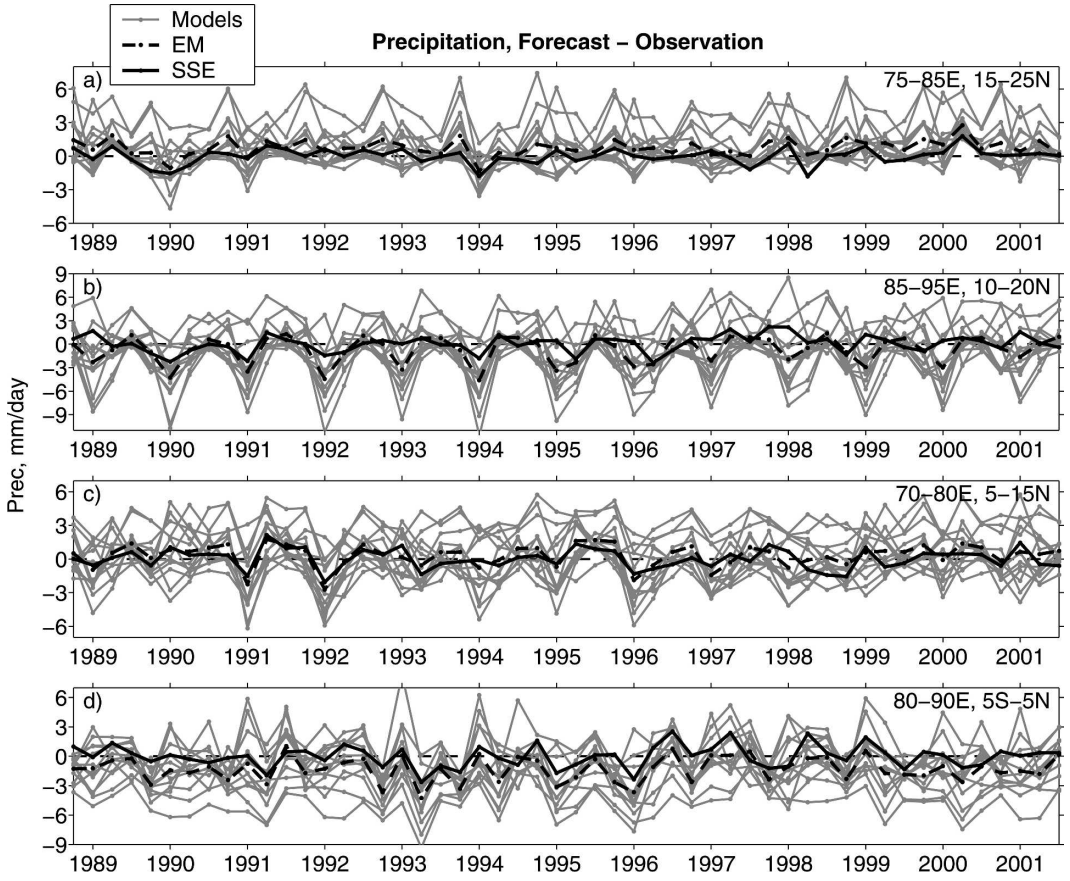

FIG. 10. Error in precipitation forecasts (mm day^{-1}) during 13 yr of seasonal simulations (four seasons per year) over four different (10° latitude × 10° longitude) regions (shown in Fig. 2). All the member models are shown as gray solid lines. The EM is the thick dashed line, and the SSE is the thick solid line. The year labels along the horizontal axes indicate the June–August season of that year.

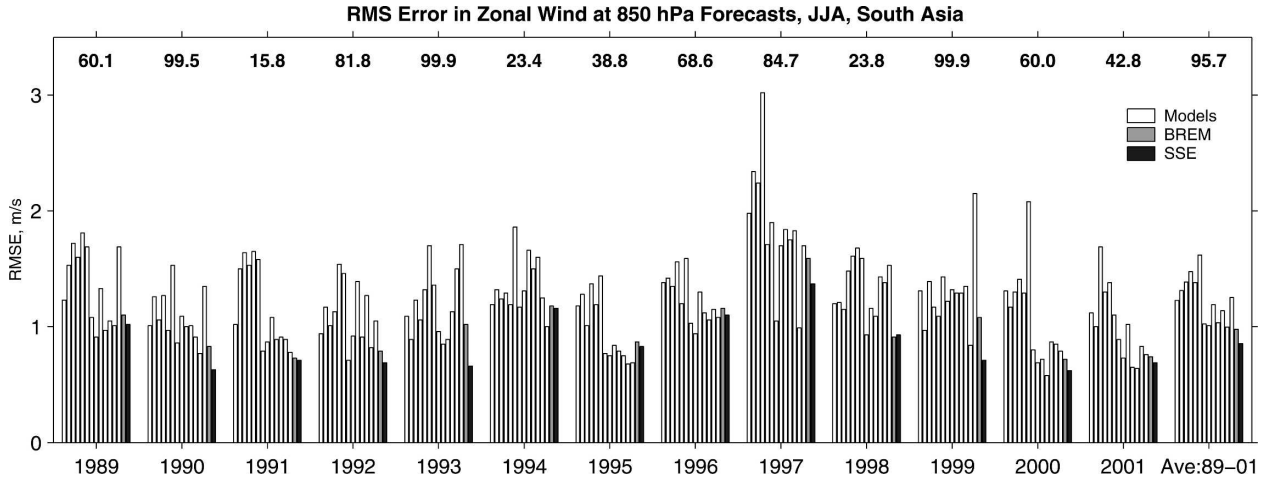

FIG. 11. The rms error in the zonal wind anomaly forecasts at 850 hPa (in m s^{-1}) during June–August over the region 10°S–35°N, 50°–110°E from 13 coupled models, their BREM, and the SSE. Also shown is the significance level (in percent) by which the SSE rms error differs from that of the BREM using a Student's *t* test (see the appendix).

spread for errors in seasonal forecasts during all four seasons of a year. The EM was closer to observations as compared to all the member models. However, the superensemble forecast outperforms all the member models and the EM over this oceanic region as well.

For precipitation forecasts, the skill of the superensemble was highest compared with all the 13 state-of-the-art coupled models and their EM. These forecast skills were assessed from the rms errors, anomaly correlations, case studies of anomaly predictions, and from the errors of total precipitation over the full monsoon domain and over four subdomains. This result suggests that the superensemble forecasts can be considered as one of the best current methods over the south Asian monsoon region for the seasonal time scale and can perhaps be used for real-time seasonal prediction with some degree of confidence.

b. Winds

The diabatic heat source in the Tropics is related to circulation features (Matsuno 1966; Webster 1972; Gill 1980). The heat source largely arises from the latent heat release during precipitation. The induced circulation, in turn, can affect the precipitation in a positive feedback mechanism (Srinivasan and Nanjundiah 2002; Li and Zeng 2002) by supplying enhanced moisture to the system (Chakraborty et al. 2002). Hence, forecasts of the horizontal winds of the lower troposphere are also equally important to ensure a better skill for the seasonal monsoon rainfall prediction. Figures 11 and 12

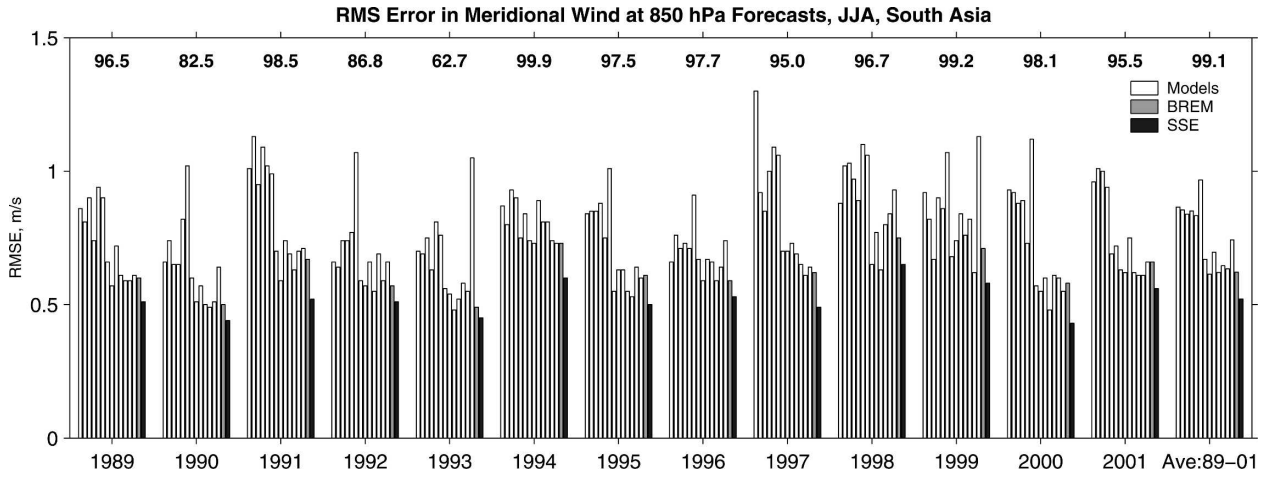

FIG. 12. As in Fig. 11, but for the meridional wind anomaly forecasts.

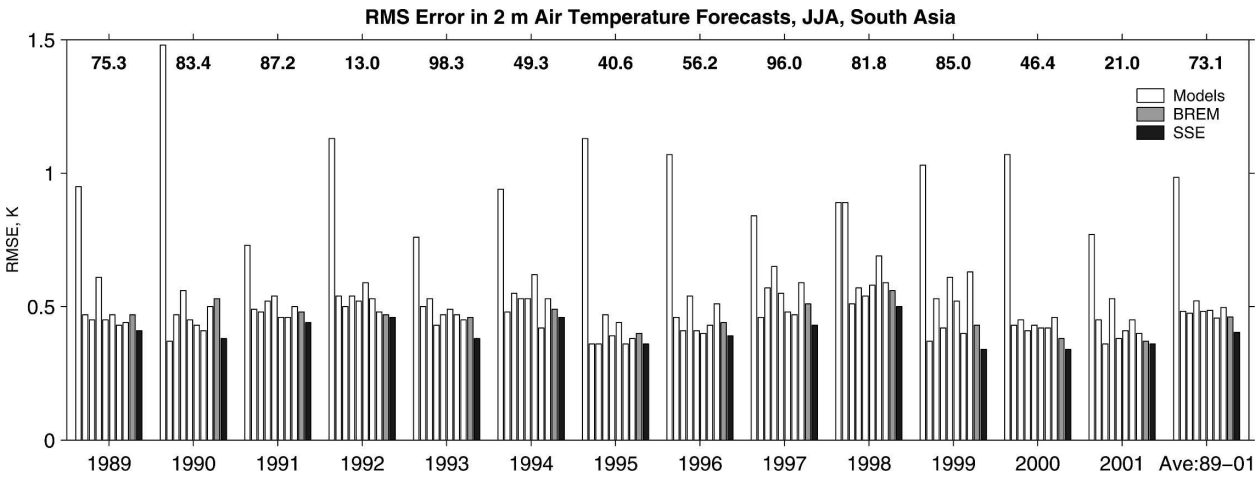

FIG. 13. The rms error in 2-m air temperature anomaly forecasts (K) during June–August over the region 10°S–35°N, 50°–110°E from eight coupled models, their BREM, and the SSE. Also shown is the significance level (in percent) by which the SSE rms error differs from that of the BREM using a Student's *t* test (see the appendix).

show the rms error in the seasonal mean (June–August) zonal and meridional wind anomaly forecast at 850 hPa over 10°S–35°N, 50°–110°E from 13 different CGCMs, their BREM, and the SSE. The numbers at the top of the figure for each year indicate the significance level, by which the superensemble forecast differs from that of the BREM using a Student's *t* test (appendix). A large variation in rms error of the zonal wind forecasts by the member models is shown in Fig. 11. The BREM of the models does not show any higher skill compared with several of the member models during many years of forecasts. However, the models that show higher skill compared with the BREM vary from one year to the other. Hence, in overall skill measure, performance of the BREM is better than the individual member models. But note that the BREM (almost) always (except for 1998) has skill lower than those of the SSE. This is because, unlike the SSE, poorer models have the same weight as the best models (after bias removal) when the BREM is calculated. The significance level, by which the superensemble forecast differs from that of the BREM, was more than 95% in 3 of the 13 yr of simulations, and in the average over these years (labeled as Ave: 89–01 in the figure).

The rms errors of the meridional wind anomaly forecasts (Fig. 12) have much lower values compared with those of the zonal wind forecasts (Fig. 11). This may be related to the fact that over this domain during June–August the magnitude of the zonal wind anomaly is much higher (more than double) than the magnitude of the meridional wind anomaly, and that the rms error tends to be lower if the magnitude of the variable is less as compared with the case when the magnitude is high. Member models show rms errors generally in the range of 0.75 to 1.25 m s^{-1}. The BREM forecast rms errors are generally lower than the individual member models' rms errors. However, from Fig. 12 it is evident that the superensemble performs better than any other models and their BREM for the meridional wind forecasts over this domain for all of the years. The difference in the rms errors for the superensemble as compared with the BREM is significant at more than 95% level (based on a Student's *t* test) during 10 of the 13 yr of simulations. On average (Ave: 1989–2001), rms error of the SSE was better than the rms error of the BREM with a significant level of more than 99%. These results show that applying different weights to different models in the suite according to their performance can make a better forecast product in the seasonal mean scale over the south Asian summer monsoon region.

c. Surface air temperature

One more important parameter in the context of seasonal prediction is the 2-m air temperature. The 2-m air temperature from only 8 of the 13 coupled models (the Australian model and seven DEMETER members) was accessible for this study. The rms error of the seasonal anomaly forecasts of this parameter, averaged over the domain 10°S–35°N, 50°–110°E, is shown in Fig. 13. The rms error of most of the models remains within 0.50–0.75 K. Surprisingly, the rms error of the BREM does not show a very low value compared with those of the member models. However, the rms error of the superensemble forecast was, in general, lower than the BREM forecast during these 13 yr of simulations. The low difference in confidence level during 11 of the 13 yr of forecast was a result of the high standard deviation of the member models because of the very poor perfor-

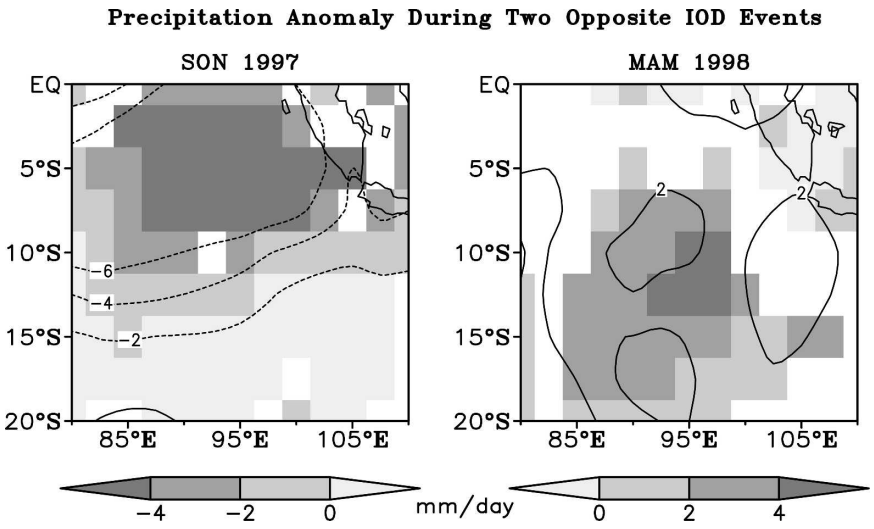

FIG. 14. Anomaly in precipitation forecasts (mm day^{-1}) during September–November (SON) 1997 and March–May (MAM) 1998 (two opposite Indian Ocean dipole events) from observations (contour) and superensemble forecast (shaded). Only those grids are shaded where the superensemble anomaly forecast is superior to the EM anomaly forecast with more than 95% confidence using a Student's *t* test.

mance of one or two models. If this superensemble were constructed more frequently during a day it would be possible to predict surface temperature, especially for the maximum–minimum temperature, more accurately.

5. Precipitation anomaly during two opposite phases of IOD events

In this section the performance of the superensemble in predicting some extreme events is presented. The IOD mode (Saji et al. 1999) is considered to be one of the strong elements of the air–sea interactions over the Indian Ocean. During a positive IOD event, the SST anomaly over the eastern equatorial Indian Ocean becomes negative and that over the western equatorial Indian Ocean becomes positive. An opposite SST anomaly pattern is seen during a negative IOD event. This seesaw of the SST anomaly has a strong coupling with the zonal wind along the equator and precipitation surrounding this region. Further studies (e.g., Ashok et al. 2001; Gadgil et al. 2003) showed that the IOD events can have major impacts on the Indian summer monsoon. In this section, two extreme IOD events were chosen for comparison of the observed and superensemble-based precipitation forecasts.

During September–November of 1997 the SST over the eastern Indian Ocean was below normal and during March–May 1998 it was above normal. The precipitation over this region was below normal during the months of September–November of 1997 and above normal during the months of March–May of 1998. Figure 14 shows the anomaly in precipitation forecasts from the superensemble (shaded) and the observed estimates (contour) for these two seasons over the eastern equatorial Indian Ocean. Only those grid boxes were shaded where the superensemble forecasts were better than the EM forecasts of the 13 CGCMs at 95% significance level using a Student's *t* test. Note that the superensemble predicts the seasonal precipitation anomaly over this region both in location and magnitude reasonably well. Also note that in most of the places where the precipitation anomaly is extreme, the superensemble forecasts are better than those of the EM forecasts at the 95% significance level. Therefore, it is possible to conclude that the superensemble is able to predict the total rains over the domain 10°S–35°N, 50°–110°E as well as the anomaly of extreme events over a smaller spatial domain better in comparison with the EM.

6. Member models and the superensemble

It was pointed out by Krishnamurti et al. (2003) that the forecasts from the superensemble improve if better-quality model datasets or observed analysis fields are used in the training phase. This section tries to verify this notion for several climate models and explores a possible reason for the improvement in forecast skills when better models are added to the suite.

Three different combinations of models were considered from the ensemble of 13 coupled models. First, superensemble forecasts were obtained using only two models, namely, CCM3 and POAMA1. Next, the seven member models of the DEMETER project were added to these two models, and superensemble forecasts were obtained for this group of nine models. Finally these results are compared with the results obtained from the superensemble forecasts constructed using all 13 models.

The skill of the superensemble in seasonal forecasts of precipitation for these three combinations of input models were examined over four regions, namely, global (60°S–60°N, 0°–360°E), Tropics (30°S–30°N, 0°–360°E), North America (0°–40°N, 130°–60°W), and south Asia (10°S–35°N, 50°–110°E). Figure 15 shows the results of the precipitation forecasts. The mean of the rms errors in precipitation over the 13 yr of forecasts during June–August were taken for brevity. It can be noted from this figure that the forecast errors were largest over the four regions when two models were used in constructing the superensemble. (We have used two different combinations of two-member SSE, namely, CCM3-POAMA1 and AOR-CCM3. The rms errors were similar in both the cases. Here we present results from the CCM3-POAMA1 combination.) Adding the seven DEMETER member models to the suite (9 models in total) and further adding the four FSU models (13 models in total) continually improves the rms errors in seasonal forecasts. The highest improvement was noticed over the North American region. Note that for all the combinations, the lowest errors were found for the global precipitation and highest errors were for the south Asian region. Irrespective of this fundamental difference in the order of rms errors, adding more models to the suite always improves the superensemble forecasts over these regions. This is because of the fact that when a set of better data is available for multiple regression, the weights assigned to the poor datasets (models) decrease (even regionally) and so does their contribution to the final forecasts. Note that in Fig. 15 adding more models did not improve the rms error over a region. This is because of the fact that adding more models increases the degrees of freedom and there is a greater chance of getting errors from the added degrees of freedom (weights of the models in this case). Another important message that can be received from Fig. 15 is that adding more models can improve the forecast skill (whatever may be its margin), but the order of rms errors over different regions remains the same. This probably relates to the different predictive skills over various domains. Better models in the training phase can improve the forecast over every region but this improvement is limited by the predictability over the region.

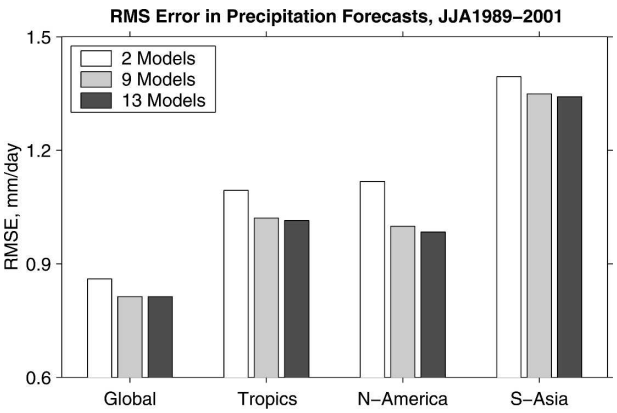

FIG. 15. Rms errors of precipitation forecasts (in mm day^{-1}) from the superensemble of 2 models (CCM3 and POAMA1), 9 models (CCM3, POAMA1, and seven DEMETER members), and 13 models (all those mentioned in Table 1) during June–August and averaged over 1989–2001 over four different regions. It can be noted that the superensemble forecasts continue to improve as better models are used in the analysis.

One obvious question arises at this point: Does the removal of a poor model from the suite improve the superensemble forecasts? From the spatial distribution of weights that are assigned to the member models, it was found that no single model has superior forecast skill compared to the other models over every region of the domain. A poor model over a region is a good model over another region, and the notion of "poor" and "good" in terms of the forecasts skills is only appropriate to the overall performance of the model. Hence, removing a poor model from the suite will degrade the skill of the superensemble forecast by reducing its performance over some region where the overall poor model performs better than the other models in the suite.

The key to the success of the superensemble over the conventional multimodel forecasts lies in assigning the differential weights among the models. In addition, assigning different weights to the same model in different locations, different variables, and for different forecasting seasons takes into consideration the inhomogeneity of the GCM forecast skills along these dimensions. Assigning more weights to a good model and less weights to a poor model, which is unlike the conventional ensemble forecasts, improves the forecast skill of the superensemble over that of the EM forecasts.

7. Conclusions

The multimodel superensemble is a powerful approach to improve upon the results of member models.

This approach removes the collective bias of all of the member models. This is based on assigning weights to each model based on their past performance. These weights vary along the three coordinates (horizontal and vertical), by the variables being improved and by the model. There are no less than 10^7 weights being used to construct the superensemble. A BREM assigns a weight of $1/N$ (where N is the number of models). Here all models carry the same weights. This assumes that a poor model upon bias removal is equivalent to the best model upon bias removal. It was noted that this is not entirely correct. This is, however, not necessarily the case if poorer models are added to the suite of models. Adding a better model to the suite decreases the weights of a poor model and its contribution to the superensemble forecast. The results from the use of just 2 models versus 9 models versus 13 models were compared and it was found that the skill does continue to improve as better models are added to the suite. Since the weights vary over space, this methodology incorporates the variations of the skills of the models over different regions of a domain by assigning a different set of weights to the models according to their local performance.

Figure 16 summarizes the forecast skill of the member models, the BREM, and the superensemble over the south Asian monsoon region for four different variables during June–August of the 13 yr of simulations. The rms errors for the best model in the suite in every season were chosen as representative of the multimodel ensemble. Also shown are the rms errors from the BREM and the superensemble. It was noted that for any variable even the best model differs from one year of the simulation to another year. Moreover, for a particular season the designated best model differs from one variable to another. Therefore, it is not possible to always rely on a single model for seasonal forecasts over the south Asian region since the model performance varies from year to year for different variables. The BREM usually performs better than the best model during most of the years. The rms errors from the superensemble are, in general, lower than those of the best model and the BREM, especially for precipitation and meridional wind anomaly forecasts.

In all of these forecasts it is safe to state that the best current forecasts are possible from the SSE that carries a somewhat better forecast skill compared with (a) member models; (b) the EM of member models; and (c) the BREM. A Student's t test was carried out in each case to show that the improvement of the superensemble over the EM is statistically significant at a level generally >95% for all of the years of simulations. This is true for all the parameters (namely, precipitation, low-level horizontal winds, and surface air temperature) studied in this paper. The negative anomalous precipitation over the east Indian Ocean and the positive anomalous precipitation over the west Indian Ocean during 1994, which characterizes the Indian Ocean dipole, was best simulated by the SSE methodology. A case study over the eastern Indian Ocean during two opposite phases of the dipole events shows that the superensemble forecasts could realistically capture the anomalous precipitation during these extreme events.

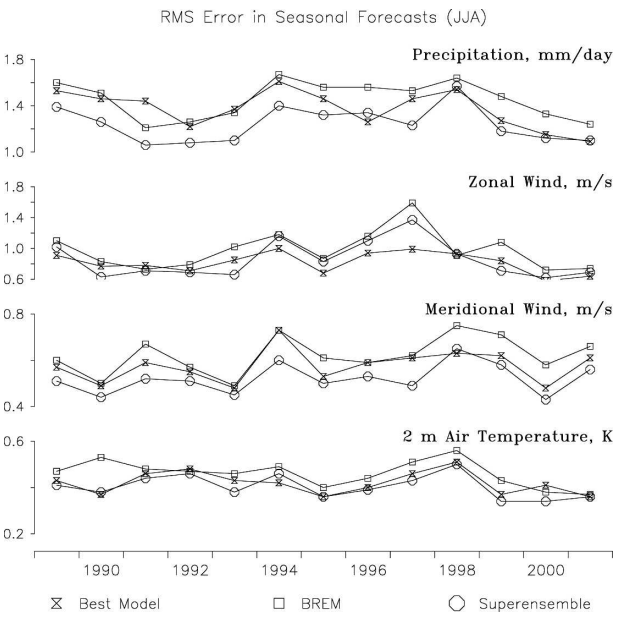

FIG. 16. The forecast skills (rms error of anomalies) during 13 yr of June–August simulations over the domain covering 10°S–35°N, 50°–110°E for the best model in the suite consisting of 13 coupled models, for the BREM and superensemble for four different variables. It was noted that for any variable the best-member model of one year was different than that of the other year. In general, the superensemble performed better than the best-member model and the BREM.

Therefore, the superensemble shows the consistent and best forecast skills compared with the member models and the EM for seasonal time scale. This suggests that the multimodel superensemble methodology can be used for real-time applications of monsoon forecasts over the south Asian region with a greater confidence as compared with the conventional multimodel forecasts.

Acknowledgments. We gratefully acknowledge the ECMWF for providing observed analysis and seven DEMETER coupled model datasets and the BMRC, Australia, for providing the POAMA-1 dataset. The research reported here was supported by NSF Grant

ATM-0108741, NOAA Grant NA06GPO512, FSURF Grant 1338-831-45, and NASA Grant NAG5-13563.

APPENDIX

Significance Test on Difference in rms Error

Here we assume that the rms error of the member models is normally distributed. The compound hypothesis is (see Green and Margerison 1978, p. 130)

$$H_0 : \mathrm{rms}_{\mathrm{em}} - \mathrm{rms}_{\mathrm{sse}} = 0$$
$$H_1 : \mathrm{rms}_{\mathrm{em}} - \mathrm{rms}_{\mathrm{sse}} \neq 0 \text{ ,}$$

where H_0 is the null hypothesis and H_1 is the alternate hypothesis. Here $\mathrm{rms}_{\mathrm{em}}$ and $\mathrm{rms}_{\mathrm{sse}}$ are the population rms errors of the EM and the SSE. Note here that we take *rms error of the EM* instead of the mean rms error of the members in the ensemble to compute the significance level. This is because, in general, the rms error of the EM is lower than the mean rms error of the individual members in the ensemble.

From the sample estimates of the mean and variances, the *t*-test parameter can be written as

$$t_s = \frac{|\mathrm{rms}_{\mathrm{em}} - \mathrm{rms}_{\mathrm{sse}}|}{s_{\overline{D}}},$$

where $s_{\overline{D}} = s_D/\sqrt{n}$ with s_D being the standard deviation of rms error within the ensemble of n members. There are $n - 1$ degrees of freedom in this case. The hypothesis H_0 is rejected with $(1 - \alpha)$ 100% confidence if

$$|t_s| > t_{(1-\alpha/2),(n-1)}.$$

In other words, we state that the results of $\mathrm{rms}_{\mathrm{em}}$ and $\mathrm{rms}_{\mathrm{sse}}$ differ with a significance level more than $(1 - \alpha)$ 100% when the above condition is satisfied. The significance level was calculated by solving the above equation for α with known values of $\mathrm{rms}_{\mathrm{em}}$, $\mathrm{rms}_{\mathrm{sse}}$, s_D, and n.

REFERENCES

Anthes, R. A., 1977: A cumulus parameterization scheme utilizing a one-dimensional cloud model. *Mon. Wea. Rev.,* **105,** 270–286.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.,* **31,** 674–701.

Ashok, K., Z. Guan, and T. Yamagata, 2001: Influence of the Indian Ocean Dipole on the relationship between the Indian monsoon rainfall and ENSO. *Geophys. Res. Lett.,* **28,** 4499–4502.

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

——, ——, F. Molenti, S. Tibaldi, and U. Cubasch, 1990: Extended range predictions with ECMWF models: Time lagged ensemble forecasting. *Quart. J. Roy. Meteor. Soc.,* **116,** 867–912.

Chakraborty, A., R. S. Nanjundiah, and J. Srinivasan, 2002: Role of Asian and African orography in Indian summer monsoon. *Geophys. Res. Lett.,* **29,** 1989, doi:10.1029/2002GL015522.

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

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

Gadgil, S., and S. Sajani, 1998: Monsoon precipitation in the AMIP runs. *Climate Dyn.,* **14,** 659–689.

——, P. N. Vinayachandran, and P. A. Francis, 2003: Droughts of the Indian summer monsoon: Role of clouds over the Indian Ocean. *Curr. Sci.,* **85,** 1713–1719.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.,* **106,** 447–462.

Green, J. R., and D. Margerison, 1978: *Statistical Treatment of Experimental Data.* 3d ed. Elsevier Scientific, 382 pp.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterization. *Mon. Wea. Rev.,* **121,** 764–787.

Kang, I.-S., and Coauthors, 2002: Intercomparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs. *Climate Dyn.,* **19,** 383–395.

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

Krishnamurti, T. N., and H. S. Bedi, 1988: Cumulus parameterization and rainfall rates: Part III. *Mon. Wea. Rev.,* **116,** 583–599.

——, Y. Ramanathan, H. L. Pan, R. J. Pasch, and J. Molinary, 1980: Cumulus parameterization and rainfall rates I. *Mon. Wea. Rev.,* **108,** 465–472.

——, 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.

——, ——, D. W. Shin, and C. E. Williford, 2000a: Multimodel superensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

——, ——, Z. Zhang, T. E. LaRow, D. R. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran, 2000b: Improving tropical precipitation forecasts from a multianalysis superensemble. *J. Climate,* **13,** 4217–4227.

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

——, L. Stefanova, A. Chakraborty, T. S. V. V. Kumar, S. Cocke, D. Bachiochi, and B. Mackey, 2002: Seasonal forecasts of precipitation anomalies for North American and Asian monsoons. *J. Meteor. Soc. Japan,* **80,** 1415–1426.

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

Li, J., and Q. Zeng, 2002: A unified monsoon index. *Geophys. Res. Lett.,* **29,** 1274, doi:10.1029/2001GL013874.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* **26,** 636–646.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan,* **44,** 25–43.

Palmer, T. N., and Coauthors, 2004: Development of a European Multi-Model Ensemble System for Seasonal-to-Interannual

Prediction (DEMETER). *Bull. Amer. Meteor. Soc.,* **85,** 853–872.

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

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

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature,* **401,** 360–363.

Sperber, K. R., and T. N. Palmer, 1996: Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project. *J. Climate,* **9,** 2727–2750.

Srinivasan, J., and R. S. Nanjundiah, 2002: The evolution of Indian summer monsoon in 1997 and 1983. *Meteor. Atmos. Phys.,* **79,** 243–257.

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

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

Wang, B., I.-S. Kang, and J.-Y. Lee, 2004a: Ensemble simulations of Asian–Australian monsoon variability by 11 AGCMs. *J. Climate,* **17,** 803–818.

Wang, G., O. Alves, A. Zhong, N. Smith, A. Schiller, G. Meyers, F. Tseitkin, and S. Godfrey, 2004b: POAMA: An Australian ocean–atmosphere model for climate prediction. *Extended Abstracts, 15th Symp. on Global Change and Climate Variations,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, J13.18.

Webster, P. J., 1972: Response of the tropical atmosphere to local, steady forcing. *Mon. Wea. Rev.,* **100,** 517–540.

Wolff, J.-O., E. Maier-Reimer, and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model: HOPE. DKRZ Tech. Rep. 13, 98 pp.

Xie, P., and P. A. Arkin, 1997: 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.

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

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