

Multi-model ensemble schemes for predicting northeast monsoon rainfall over peninsular India

NACHIKETA ACHARYA¹, S C KAR^{2,*}, MAKARAND A KULKARNI¹, U C MOHANTY¹ and L N SAHOO³

¹Centre for Atmospheric Sciences, Indian Institute of Technology, New Delhi, India.

²National Centre for Medium Range Weather Forecasting, Noida, India.

³Department of Statistics, Utkal University, Bhubaneswar, India.

*Corresponding author. e-mail: sckar@ncmrwf.gov.in

The northeast (NE) monsoon season (October, November and December) is the major period of rainfall activity over south peninsular India. This study is mainly focused on the prediction of northeast monsoon rainfall using lead-1 products (forecasts for the season issued in beginning of September) of seven general circulation models (GCMs). An examination of the performances of these GCMs during hindcast runs (1982–2008) indicates that these models are not able to simulate the observed interannual variability of rainfall. Inaccurate response of the models to sea surface temperatures may be one of the probable reasons for the poor performance of these models to predict seasonal mean rainfall anomalies over the study domain. An attempt has been made to improve the accuracy of predicted rainfall using three different multi-model ensemble (MME) schemes, *viz.*, simple arithmetic mean of models (EM), principal component regression (PCR) and singular value decomposition based multiple linear regressions (SVD). It is found out that among these three schemes, SVD based MME has more skill than other MME schemes as well as member models.

1. Introduction

The northeast (NE) monsoon season (October, November and December) is the period of major rainfall activity over south peninsula of India, particularly in Andhra Pradesh, Rayalaseema, Tamil Nadu and Pondicherry. This season is also known as the winter monsoon (Nageswara Rao 1999) and post-monsoon season (Singh and Sontakke 1999). The NE monsoon season contributes to about 50% of annual rainfall in the east coast of Indian peninsula (Kumar *et al* 2007). During southwest monsoon, there is not much rain over this region as this is the rain shadow region of the Western Ghats mountains along the west coast of India. When

the southwest monsoon retreats, pressure and wind distribution reverses at the beginning of October, a trough of low pressure becomes established in the south Bay of Bengal. The passage of easterly low-pressure waves occasionally intensifies this trough. Depressions and cyclonic storms also occasionally form in the trough of low pressure over the south Bay of Bengal. Due to this type of situations, equatorial maritime air moves towards south India, and causes widespread rainfall (Kripalani and Kumar 2004). The rainfall during this period is of immense societal significance to 150 million people as it supports the main cultivation season known as Maha in Sri Lanka and Rabi in southern India (Zubair 2002).

Keywords. Northeast monsoon; multi-model ensemble; rainfall; prediction; principal component regression; single value decomposition.

Eventhough interannual variation and prediction of the southwest monsoon has been widely studied, as documented by numerous publications, the NE monsoon over south peninsular India has received much less attention, with a limited number of studies (e.g., Doraiswamy 1946; Rao and Jagannathan 1953; Rao 1963; Ramaswamy 1972; Srinivasan and Ramamurthy 1973; Dhar and Rakecha 1983; Krishnan 1984; Raj and Jamadar 1990; Sridharan and Muthusamy 1990; Singh and Sontakke 1999; Kripalani *et al* 2004; Zubair and Ropelewski 2006; Kumar *et al* 2007; Nayagam *et al* 2009). Some of these studies have discussed about global teleconnection patterns related to NE monsoon. Kripalani and Kumar (2004) found a direct relationship between Indian Ocean Dipole (IOD) and NE monsoon, suggesting that the positive (negative) phase enhances (suppresses) the north-east monsoon activity. Studies have shown that the rainfall over the southern part of India, during the NE monsoon season increases with El Niño events (Dhar and Rakecha 1983; Singh and Sontakke 1999). In their study, Zubair and Ropelewski (2006) have explained the relationship between El-Niño Southern Oscillation (ENSO) and NE monsoon over peninsular India and Sri Lanka. Balachandran *et al* (2006) defined a zonal temperature anomaly gradient index (ZTAGI) between eastern equatorial Pacific and western equatorial Pacific and found a stable significant inverse relationship with NE monsoon rainfall. In a recent study, the spatial and temporal variability of rainfall over peninsular India during the NE monsoon season has been discussed (Nayagam *et al* 2009). In this study, the dominant modes of the NE monsoon rainfall were identified using Empirical Orthogonal Function (EOF) analysis and the power over the identified scales was extracted using wavelet analysis. They found that the dominant frequencies in the 2–8 year band are the 4-year period and the 8-year period. Zubair and Ropelewski (2006) concluded that if peninsular India is considered as a single entity, there has been a remarkable rise in the correlation of the NE monsoon rainfall and ENSO. This regional rise may be attributed to a slight enhancement of the low-level circulation in recent decades during the NE monsoon that leads to relative increase in the orographic component of rainfall. Further Zubair and Ropelewski (2006) found that the intensification of the ENSO–NE monsoon rainfall relationship is modest and within the historical record but stands in contrast to the weakening relationship in summer. The intensification of the circulation is consistent with the warming of surface temperatures over the tropical Indian Ocean in recent decades.

For predicting southwest monsoon rainfall over India, many statistical/empirical forecasting

models (Thapliyal 1981; Gowariker *et al* 1991; Sahai *et al* 2003; Rajeevan *et al* 2006a) have been developed and used. But there are no such studies about the prediction of NE monsoon. For long range forecasting purpose, IMD has used a statistical model based on east Pacific wind at 850 hPa, north India wind at 200 hPa, south Indian Ocean wind at 200 hPa and Nino 3.4 SST indices as predictors (source: IMD). There are several studies on prediction of south–west monsoon using general circulation models (GCMs) and their multimodel ensemble (MME) forecast (Palmer *et al* 2004; Kumar *et al* 2005; Wang *et al* 2005; Krishnamurti *et al* 2006; Chakraborty and Krishnamurti 2009). However, no systematic study has been made on the skill of global dynamic models as far as seasonal NE monsoon rainfall predictions are concerned. No attempt has been made so far to examine the usability of MME techniques for the prediction of NE monsoon rainfall over India.

This study is mainly focused on prediction of NE monsoon rainfall in seasonal timescale using three different MME schemes using seven different GCMs for 27 years (1982–2008) with lead-1 products, i.e., predictions for the season made in the beginning of September. Section 2 describes the models used and the three different MME schemes. Section 3 has the results and discussions. The study has been concluded in section 4.

2. Data and methodology

2.1 Observed data

For observed rainfall, the $1.0^{\circ} \times 1.0^{\circ}$ gridded datasets (Rajeevan *et al* 2006b) have been used. Area averaged time series are made on seven meteorological subdivisions, *viz.*, north interior Karnataka, coastal Karnataka, south interior Karnataka, Kerala, Tamil Nadu, Rayalaseema and coastal Andhra Pradesh. The intercorrelations between the rainfall of these seven meteorological subdivisions are positive and highly significant which indicate that the rainfall over the different meteorological subdivisions varies in the same sense. The NE monsoon rainfall averaged over the region is also significantly correlated with those of subdivisional rainfall. This shows that the region selected represents the core of NE monsoon rainfall (Nayagam *et al* 2009). Thus, the whole of this region can be considered to be fairly coherent. The observed sea surface temperatures (SST) are from Reynolds and Smith (1994) and wind data used in the study are based on NOAA NCEP-DOE Reanalysis-2 (Kanamitsu *et al* 2002).

2.2 GCM/AOGCM products

Hindcast runs (lead-1 for October–November–December of 1982 to 2008) from seven global models have been used in this study. These models are NCEP climate forecast system (CFS), the latest version of the NCAR Community Climate Model (CCM3.6), International Research Institute of Climate and Society (IRI) ECHAM4.5 (E4p5), IRI mixed layer coupled model ECHAM4.5-GML (GML), IRI coupled models ECHAM4P5-MOM3-AC1 (CPL-AC1) and ECHAM4P5-MOM3-DC2 (CPL-DC2) and IRI 2 tier ECHAM4.5-CFS SST (E4p5-CFS). The details of these models are discussed in Kar *et al* (2011). Table 1 represents a brief summary of each model including references of relevant literature. The CFS runs were made at NCEP, USA and other models were run at International Research Institute (IRI). The initial conditions for CFS model runs (ensemble members) are from August and September of each year. The two-tier models run at IRI use monthly SST forecasts based on August SST. The atmospheric initial conditions are from September. The AGCMs use initialization of atmospheric fields from long historical runs with observed SST on September 1 of other years/ensemble members. There is no atmospheric or land surface initialization to analyzed observations, and the atmospheric initial conditions are essentially random for the calendar date. The one-tier models use ocean data assimilation, but again there is no atmospheric or land surface data assimilation. The model data have been downloaded from IRI (<http://portal.iri.columbia.edu/portal/server.pt>). For this study the average of multiple ensemble members has been used for each model. Before applying MME methods, the ensemble mean prediction of rainfall by each model was bilinear interpolated to a common $1.0^0 \times 1.0^0$ grid which

is comparable to the resolution of the observed data.

2.3 Multi-model ensemble schemes

In this study, three different MME schemes are utilized. The following are the brief description of the MME methods utilized here.

2.3.1 MME using simple arithmetic mean

Simplest of the MME schemes is based on simple averaging of all the individual models. Hagedorn *et al* (2005) have described the rationale behind the success of such MME techniques for seasonal prediction. In the present study, the method of carrying out MME by simple arithmetic mean of individual models is being referred to as EM. In this method, all the individual member models have been assigned same weight while carrying out ensemble average.

2.3.2 MME using principal component regression

This method involves principal component regression (PCR) for finding a weighted multi-model ensemble mean. The main advantage of principal component regression is that it orthogonalises the predictors in which multi-collinearity between predictors automatically get removed. Fekedulegn *et al* (2002) have explained this method in detail. PCR procedure is a multiple linear regression problem in principal component space. In this method, at first, principal component (PC) analysis is done on models' rainfall and only such models which explain a high amount of variance are retained for

Table 1. Description of member models used in MME.

Model	Resolution	AGCM	OGCM	Ensemble member	Reference
CFS	(T62) $1.8^0 \times 1.8^0$	GFS	MOM3	15	Saha <i>et al</i> (2006)
CCM3.6	(T42) $2.7^0 \times 2.8^0$	CCM3.6	Constructed analog SST	24	Hurrell <i>et al</i> (1998)
E4p5 (ca sst)	(T42) $2.7^0 \times 2.8^0$	ECHAM4p5	Constructed analog SST	24	Roeckner <i>et al</i> (1996)
GML	(T42) $2.7^0 \times 2.8^0$	ECHAM4p5	CFS-predicted SSTs prescribed over the tropical Pacific basin (semi-coupled)	12	Roeckner <i>et al</i> (1996); Lee and De Witt (2009)
CPL-AC1	(T42) $2.7^0 \times 2.8^0$	ECHAM4p5	MOM3 (anomaly-coupled)	24	Roeckner <i>et al</i> (1996); Pacanowski and Griffes (1998)
CPL-DC2	(T42) $2.7^0 \times 2.8^0$	ECHAM4p5	MOM3 (direct-coupled)	12	Roeckner <i>et al</i> (1996); Pacanowski and Griffes (1998)
E4p5-CFS	(T42) $2.7^0 \times 2.8^0$	ECHAM4p5	CFS-predicted SST	24	Roeckner <i>et al</i> (1996)

further simple multiple linear regression. In our case, first four PCs (explaining about 80% variance) have been selected for the regression. The above-mentioned procedure is carried out in leave one out cross-validation where forecasted year has been successively withheld from the training dataset, and the remaining 26 years (as total length of data is 27 years) have been used for calculation.

2.3.3 MME using singular value decomposition

For carrying out weighted multi-model ensemble mean, multiple regression method has been employed. Singular value decomposition (SVD) has been employed for the computation of the regression coefficients (referred to as SVD scheme in the following text). The advantage of SVD method is that it removes the singular matrix problem while calculating covariance among models which cannot be entirely solved with the Gauss–Jordan elimination method. The details of the method are explained in Yun *et al* (2003). In this method, the model forecasts are regressed in the training period with the observed counterpart to obtain weights:

$$E = \sum_{t=1}^N (S'_t - O'_t)^2,$$

where E is the error term that is minimized to obtain the weights, N is the length of the training dataset. S'_t and O'_t are the predicted and observed rainfall anomalies, respectively, at training time t . The outcome of this regression is statistical weights W_i ($i = 1, 2, \dots, M$; M being the number of models) assigned to every model. These weights are then passed on to the forecast phase to construct the final SVD predicted anomalies (S):

$$S = \sum_{i=1}^M W_i F_i,$$

where F_i are the forecasted anomalies from individual models.

Leave one out cross-validation technique has been used in which, each year has been successively withheld from the training dataset, and the remaining 26 years have been used for calculation of the model and observed statistics. This means and regression coefficients are used for calculating the forecast for the verification year (the year that was withheld). The model weights are the most important outcome of this SVD based multiple regression method. Similar method has been employed by Kar *et al* (2011) for the computation of weights for the individual models for July rainfall.

3. Results and discussion

3.1 Skill of individual GCMs

The observed spatial distribution of seasonal rainfall climatology over the above-stated (in section 2) seven meteorological subdivisions is shown in figure 1. During October, November December months, most parts of peninsular India get more than 1.5 mm/day rainfall. Tamil Nadu receives the largest amount of rainfall (about 3 to 5 mm/day) among all the subdivisions. The climatology of area averaged rainfall over peninsular India from observation and individual models are shown in figure 2(a). The observed climatological (based on 27 years) rainfall is 3.5 mm/day. Among the models, the rainfall climatology of CFS is more closer to observation. Except CCM3, remaining models underestimate the rainfall amounts. The interannual variability (standard deviation) of rainfall from observation and individual models have also been calculated for the region and are shown in figure 2(b). The observed interannual variability of rainfall is about 0.85 mm/day. Interannual variability simulated by all the models is much less than that of observation. It is expected that the IAV in the ensemble mean rainfall would be smaller than that observed, because the average of multiple ensemble members has been taken which reduces the total variance. In order to examine the mean standard deviation (IAV) of the ensemble members, for each model and ensemble members, interannual variability has been computed (figure not shown). It is seen that the IAV values of ensemble members of the models are closer to that of observation. For, the CFS model are marginally larger than that of the observation while other models have IAV values marginally smaller than that of observed IAV.

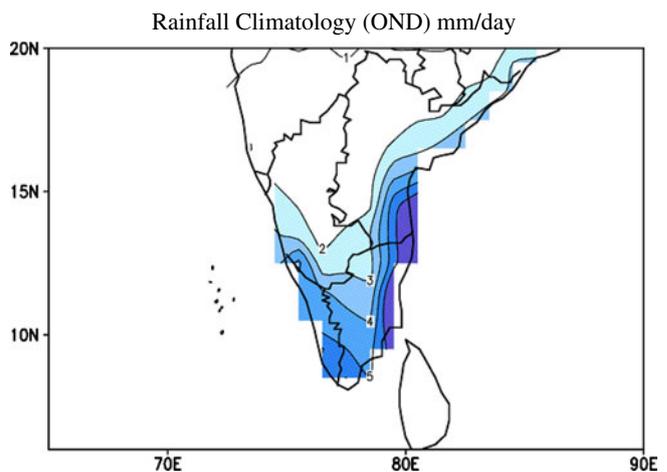


Figure 1. Climatology of observed rainfall over the south peninsular India during NE monsoon season.

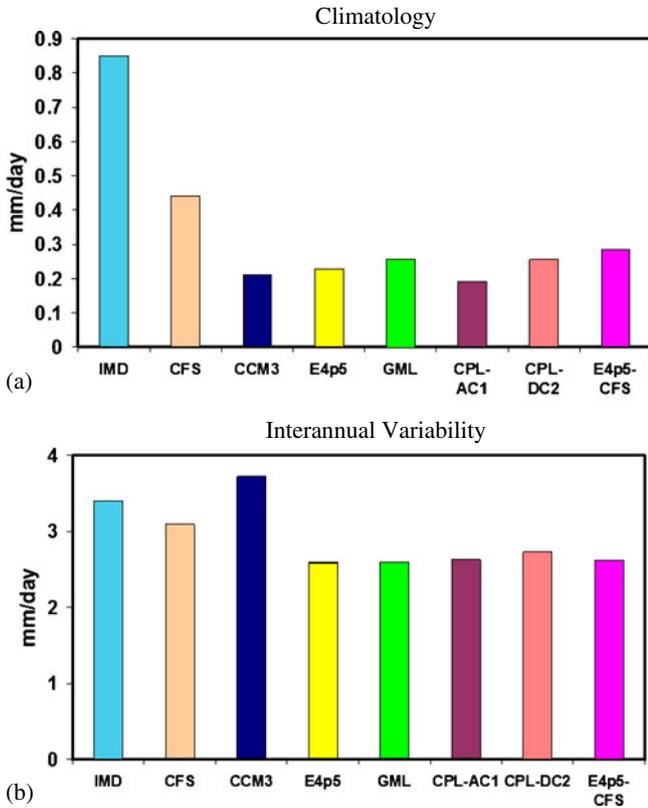


Figure 2. (a) Climatology (mm/day) and (b) interannual variability (IAV) (mm/day) of rainfall over peninsular India in October–November–December from IMD observed data and the model hindcast runs for 1982 to 2008.

The temporal correlations between the observed seasonal rainfall (area averaged) and the model results have been computed for the 27 (1982–2008) years of hindcast runs and are shown in figure 3(a). It may be noted that for statistically significant correlation (95% confidence interval) for 27 years of study, it is expected that the correlation magnitude should be at least 0.32. Except CCM3 and E4p5-CFS model, all other models show negative correlation. Root mean square errors (RMSE) from individual model is shown in figure 3(b) which indicates that all the models have large errors. Average RMSE of all the models is 1.5 mm/day. Equitable threat score (ETS) measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance and has been computed for each model for the domain averaged rainfall. The ETS is defined as:

$$ETS = \frac{(H - Hr)}{(H + M + F - Hr)}$$

where

$$Hr = (H + M) * (H + F) / T$$

H , M , and F are hits, misses and false alarm for each category. Hr is the hits due to random chance and T is the total number of events. ETS ranges from -0.33 to 1 with $ETS = 0$ indicating no skill in predictions. Several threshold values have been used in the study and the ETS values are shown in figure 3(c). It is seen from the figure that individual member models have very poor skill in predicting the seasonal mean rainfall values.

Poor performances by these models to simulate NE monsoon rainfall could be due to several reasons. Most of the models used in the present study have coarse horizontal resolution (either T42 or T62). Whereas, it is expected that large-scale anomalies can get predicted by such coarse resolution models, the details of rainfall variability over the peninsular region of India may not get properly resolved by them. In a 2-tier modelling system, the predicted sea surface temperatures (SSTs) are

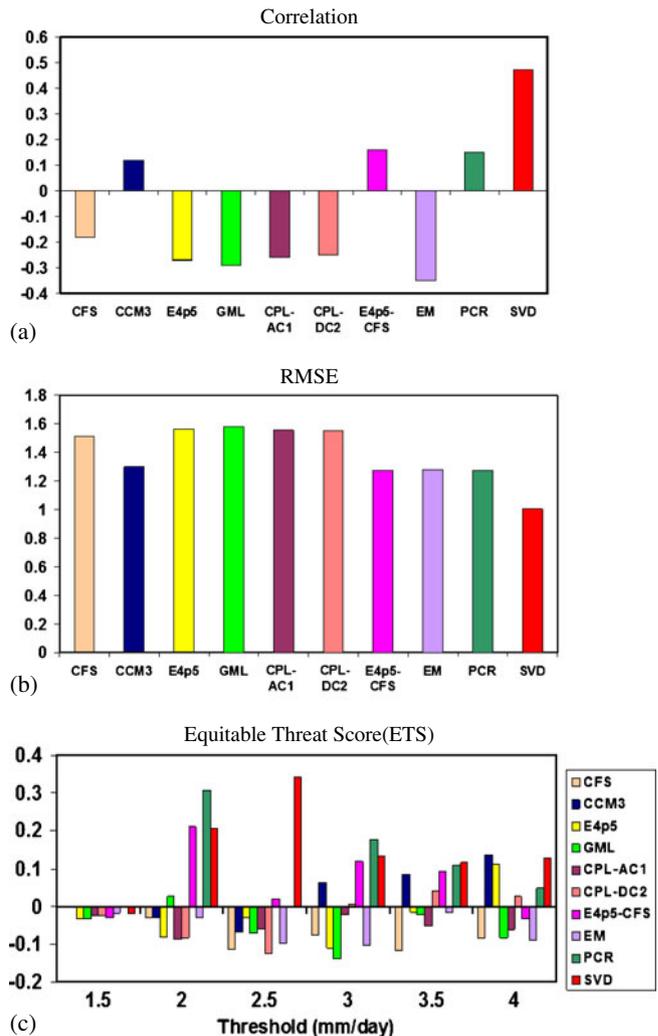


Figure 3. (a) Correlation, (b) root mean square error, and (c) the equitable threat score between rainfall from observation and models as well as MME schemes.

prescribed to atmospheric GCM and in 1-tier system, the SSTs evolve as a part of coupled ocean–atmosphere system. One of the factors that affect IAV of model simulated rainfall is the way a model responds to interannually varying SSTs. In addition, the internal dynamics of the models and some other boundary forcing may also affect the IAV in rainfall (Kar *et al* 2001).

Correlations between observed area averaged northeast monsoon rainfall over peninsular India and global SST have been calculated. The same field for each model with corresponding SSTs have also been computed and are shown in figure 4. It is noticed that there is a significant positive relationship between eastern equatorial Pacific SSTs and observed rainfall. Zubair and Ropelewski (2006) have noted about the increase in strength of this positive correlation since 1980. Positive correlation is also noticed in the Bay of Bengal and in the South China Sea. It is noticed that even if the SST predictions by individual models are reasonably good (figure not shown), the SST–rainfall (over peninsular India) remote response is not well

simulated by these models as shown in figure 4. Most of the models fail to capture the positive relationship between rainfall over southern India and SST over eastern equatorial Pacific except CPL-DC2. In order to examine the reason for such behaviour of this model, rainfall anomaly of each ensemble member of this model has been examined (figure not shown). It is seen that the ensemble members show wide range of variability, indicating the internally generated variability in the model is quite large and dominate the SST-forced variability. As a result, the model does not produce large variability in its ensemble mean. Moreover, the model’s rainfall has large correlations with central north Pacific SST and southwest Indian Ocean which are not observed. Therefore, while the ensemble mean (the signal part as a result of externally forced variability) shows a good correlation with the SST forcing, interannual variability of the model’s rainfall does not agree with that of observation.

In this study, years when rainfall anomaly (normalized) is between ± 1 standard deviation are considered as normal years, and all other years are

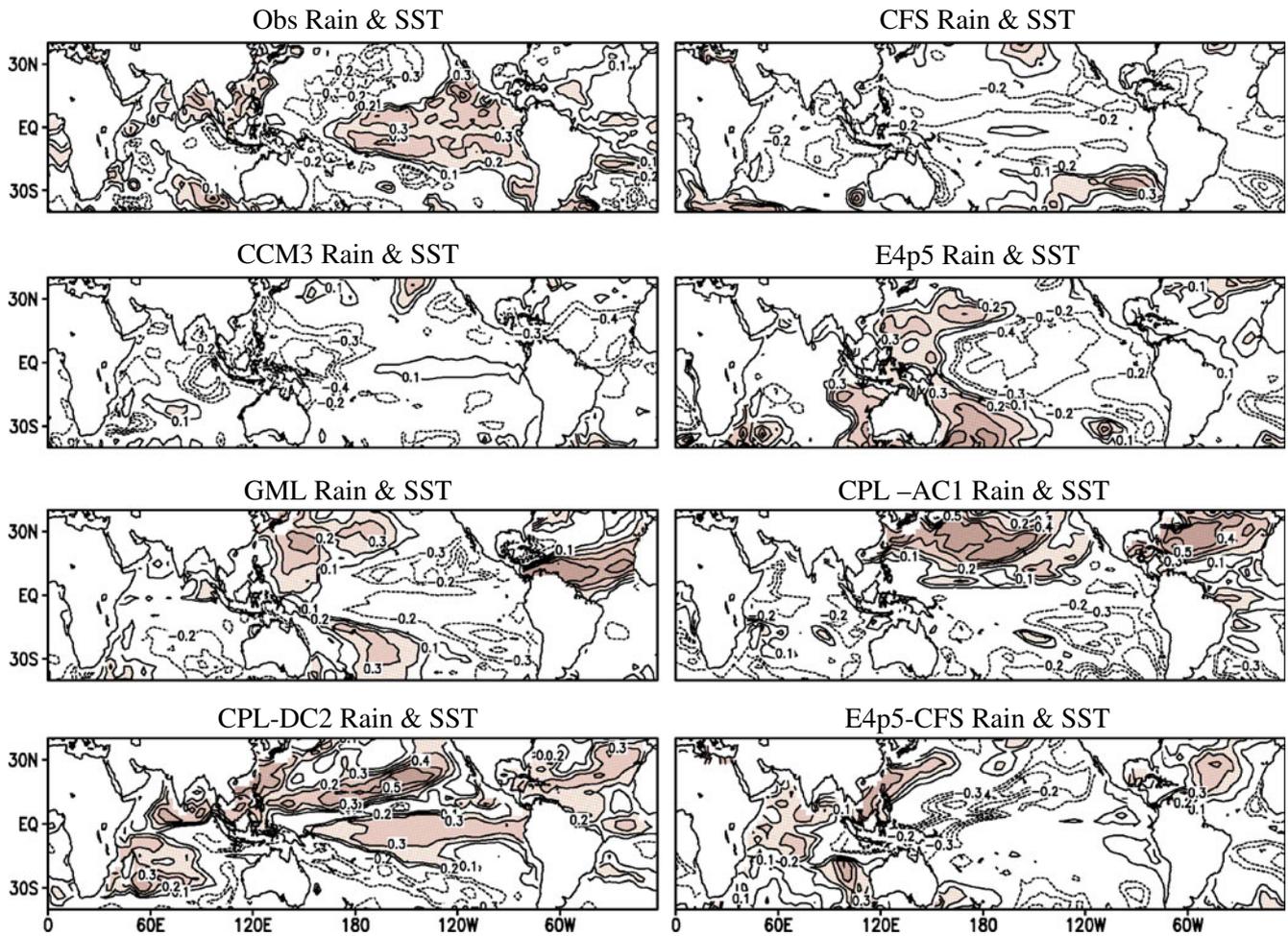


Figure 4. Correlation of the October–November–December rainfall over peninsular India with observed SST and model predicted rainfall with model SSTs. Areas with correlation values with >0.2 have been shaded.

considered as typical years when either flood or drought occurred. Some of these are extreme north-east monsoon years with significantly less rainfall (1982, 1988, 1989 and 2005), and in some years, more than normal rainfall occurred (1987, 1993, 1996, 1997 and 1998). Some of these typical years may also be characterized as El Niño year (1987), La Niña year (1988) and Indian Ocean dipole year (1982, 1996, 1997, 1998). In order to

examine the large-scale circulation features predicted by the models for these years, wind anomalies at 200 hPa from observation and models were studied in detail. Composite zonal wind anomalies (U) at 200 hPa for the above-normal years, below-normal years and their differences are shown in figure 5. Westerly winds in the mid-latitude regions of both the hemispheres and easterlies in the tropical belt is the major climatological observed pattern at

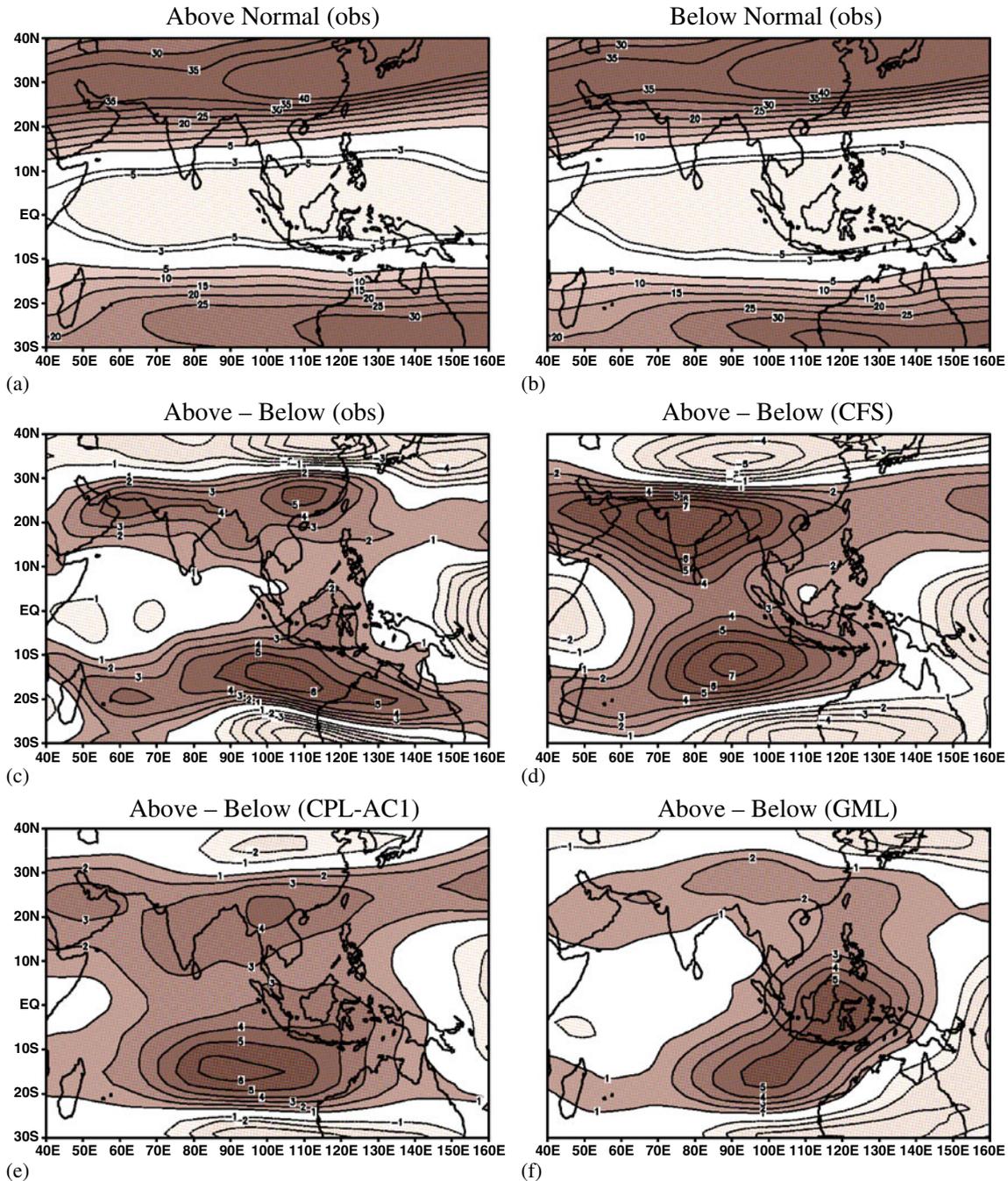


Figure 5. Composite of observed zonal winds at 200 hPa for above and below normal rainfall years (a and b), their observed difference (c) and difference seen in CFS, CPL-AC1 and GML models respectively (d, e, f).

200 hPa. In the above-normal years, stronger westerlies prevail over the latitude band of 20° – 30° N, as compared to below-normal years. At 850 hPa, stronger easterlies over the Indian Ocean and Bay of Bengal extending over to southern Indian peninsula are noticed in the composite difference plot (figure not shown). Corresponding to stronger convergence and convection in the region, a series of cyclonic and anti-cyclonic circulations are noticed at 200 hPa in the northern hemisphere. The composite difference of U at 200 hPa from CFS, CPL-AC1 and GML are also shown in figure 5. It is seen that location and intensity of westerly maxima are not simulated by these models correctly. Westerly winds from CFS model are very strong, especially over India in above-normal years but these are weaker than observation in other two models. This indicates that the lower level convergence and intense convection in these models are not brought out well for these contrasting years.

From the above discussions, it is concluded that the models have large errors in predicting rainfall over the peninsular India correctly in hindcast mode. Therefore, it is considered appropriate to use statistical post-processing techniques to properly correct the rainfall forecasts from these models and combine them using MME techniques to make more reliable forecasts.

3.2 Skill of MME schemes

Leave one out cross-validation technique has been used for all these above-mentioned methods (section 2.3) in which, each year has been successively withheld from the training dataset, and the

remaining 26 years (as data length is 27 years) have been used for calculation of the model and observed statistics, i.e., the seasonal means and regression coefficients. These means and regression coefficients are used for calculating the forecast for the verification year (the year that was withheld).

Skill of individual member models in terms of temporal correlation, RMSE and ETS discussed in the earlier sections showed that the member models have poor skill in simulating the northeast monsoon. The temporal correlation coefficients and RMSE obtained from the three different MME schemes, *viz.*, EM, PCR and SVD are shown in figure 3(a and b), respectively. It is seen from figure 3(a) that the correlation of EM is still very low. EM scheme is based on simple averaging of model, i.e., giving equal weight to each model. As a result, a good model gets the same weight as a model with low skill in this method. It may be noted that as five out of seven models show negative correlation in predicting NE monsoon rainfall, the EM scheme is not able to enhance the skill. PCR based MME provides some improvement over the EM scheme and shows positive correlation skill but it is not significant. The SVD scheme has the best skill among all the three MME schemes. The correlation coefficient obtained by using the SVD scheme is not only higher than member models and other MME schemes, it is also statistically significant. From figure 3(b), it is seen that the RMSE of rainfall obtained from SVD based MME scheme is also lesser than all the models as well as EM and PCR schemes. Examination of ETS values for the MME schemes in figure 3(c) shows that the MME schemes have improved upon the individual member models. The PCR and SVD schemes

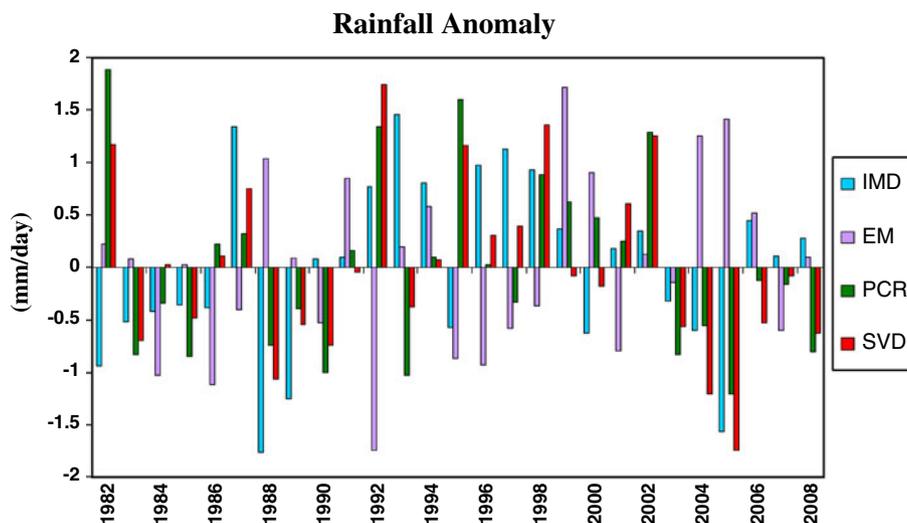


Figure 6. Rainfall anomalies (mm/day) from observation and three MME schemes for NE monsoon season over peninsular India during 1982–2008.

perform better than the simple MME carried out using equal weights to the member models. Overall, the SVD scheme has the best skill during the period of study. In PCR and SVD methods, weights are given to member models on the basis of their skill in training period (using leave one out cross-validation). The main difference is, PCR orthogonalises predictors (here models) and reduces the number of predictors by selecting first few modes which explain most of the variance, while in SVD based regression all the predictors (here the models) are used to find regression coefficients using decomposition of covariance matrix.

3.3 Skill of MME prediction in some typical years

During the study period (i.e., 1982–2008), less than normal (below normal) rainfall occurred over the NE monsoon domain in 1982, 1988, 1989 and 2005 and rainfall was above normal in 1987, 1993, 1996, 1997 and 1998. In the previous section, it was noted that SVD based MME has significantly more skill than other MME schemes. The main purpose of the present section is to compare the rainfall anomalies from the three MME schemes for these below-normal and above-normal northeast monsoon years. Observed rainfall anomalies (mm/day) and anomalies from the MME schemes for the NE monsoon seasons over peninsular India for each year are shown in figure 6. The signs of rainfall anomalies from the EM scheme for most of the below-normal as well as above-normal years do not match with that of observed anomalies while PCR and SVD schemes have higher skill in predicting rainfall anomalies for these years. But all the MME schemes fail to predict rainfall anomalies of 1982. Magnitude-wise rainfall anomalies from SVD is closer to the observation than that from the PCR.

4. Summary and conclusion

The post-monsoon or northeast monsoon season is the major period of rainfall activity over south peninsula, particularly in north interior Karnataka, coastal Karnataka, south interior Karnataka, Kerala, Tamil Nadu, Rayalaseema and coastal Andhra Pradesh. Agricultural activity of these subdivisions (mainly Tamil Nadu) is extremely dependent on the amount of rainfall received in this season. Although there are some studies related to variability and teleconnections, no such studies exist for the prediction of NE monsoon rainfall over peninsular India. This study is mainly focused on prediction of seasonal mean rainfall anomalies during NE monsoon seasons using products from seven global

dynamical models with 1-lead (forecasts prepared in beginning of September).

Examination of performances of these GCMs during hindcast runs (1982–2008) indicate that they are not able to simulate the observed interannual variability. This could be because of the fact that most of the models used in the present study have coarse resolution (T42 or T62). Whereas it is expected that large scale anomalies can get predicted by such coarse resolution models, the details of rainfall variability over the peninsular Indian region can not get resolved by them. Another probable reason for the poor performance is the fact that the atmospheric model do not respond in a correct manner to the prescribed predicted SSTs used in 2-tier systems, or the evolution of SSTs in coupled models do not agree with the observed SSTs. It is also seen that the large-scale anomaly patterns are not brought out by these models. Therefore, statistical post-processing was carried out for bias correction and optimum prediction by combination of all these available models. Three different MME schemes, *viz.*, EM, PCR and SVD have been evaluated. It is noticed that as EM is simple arithmetic mean of model, its performance is not good compared to other schemes based on weighted multi-model ensemble. SVD schemes are more skillful than PCR in the hindcast period. There are few typical (extremely above or below normal rainfall) NE monsoon years during the study period. SVD scheme has higher skill to capture the pattern of these typical years than remaining methods.

The objective of post-processing the model products using SVD scheme is to remove the statistical bias if any in model predictions and to reduce mean square error so that the weighted multi-model ensemble predictions are closer to the observed values in the training period. The regression coefficients are then used to build the predictions as if entire training period decides the mean bias of the model. If the statistical properties of the model is such that the weight computed by the model during the training period is valid for the forecasted period, the weighted multi-model ensemble mean would provide the best predictions. It is seen in our study that this is indeed the case and the SVD scheme provides the best skill.

It may be noted that total period of data used in this study is only from 1982 to 2008 which may not be very large for the estimation of regression coefficients. The large set of hindcasts helps to get optimum weight for the models in SVD scheme. Limited size of the hindcasts is also a challenge for training in multiple regression schemes. However, regressions will suffer from the decadal changes associated with global warming and associated teleconnections if a longer hindcast dataset

is used. Therefore, it is not just the hindcast length, but other options such as improving the models, data assimilation, etc. should also be considered which enhanced the skill of SVD based MME schemes.

Acknowledgements

This research has been conducted as part of the project entitled 'Development and Application of Extended Range Forecast System for Climate Risk Management in Agriculture' at IIT, Delhi sponsored by the Department of Agriculture and Cooperation, Government of India. India Meteorological Department (IMD) is thanked for providing observed gridded rainfall datasets. The model datasets used in this study have been provided by the International Research Institute of Climate and Society (IRI), USA. The computing for the GCM simulations made by IRI was partially provided by a grant from the NCAR Climate System Laboratory (CSL) program to the IRI.

References

- Balachandran S, Asokan R and Sridharan S 2006 Global surface temperature in relation to northeast monsoon rainfall over Tamil Nadu; *J. Earth Syst. Sci.* **115**(3) 349–362, doi: 10.1007/BF02702047.
- Chakraborty A and Krishnamurti T N 2009 Improving global model precipitation forecasts over India using downscaling and the FSU super-ensemble. Part II: Seasonal Climate; *Mon. Weather Rev.* **137** 2736–2757.
- Dhar O N and Rakecha P R 1983 Foreshadowing northeast monsoon rainfall over Tamil Nadu, India; *Mon. Weather Rev.* **111** 109–112.
- Doraiswamy I V 1946 Forecasting the northeast monsoon rainfall of south Madras; Scientific Note No. **98**, India Meteorological Department, pp. 147–154.
- Fekedulegn B D, Colbert J J, Hicks R R Jr and Schuckers M 2002 Coping with multicollinearity: An example on application of principal components regression in dendroecology; U.S. Department of Agriculture, Forest Service, 43p.
- Gowariker V, Thapliyal V, Kulshrestha S M, Mandal G S, Sen Roy N and Sikka D R 1991 A power regression model for long-range forecast of southwest monsoon rainfall over India; *Mausam* **42** 125–130.
- Hagedorn R, Doblas-Reyes F J and Palmer T N 2005 The rationale behind the climate prediction using DEMETER forecast; *Tellus* **57A** 280–289.
- Hurrell J W, Hack J J, Boville B A, Williamson D L and Kiehl J T 1998 The dynamical simulation of the NCAR Community Climate Model version 3 (CCM3); *J. Climate* **11** 1207–1236.
- Kanamitsu M, Ebisuzaki W, Woollen J, Yang S K, Hnilo J J, Fiorino M and Potter G L 2002 NCEP-DOE AMIP-II Reanalysis (R-2); *Bull. Am. Meteorol. Soc.* **83** 1631–1643.
- Kar S C, Sugi Masato and Sato Nobuo 2001 Interannual variability of the Indian summer monsoon and internal variability in the JMA global model simulations; *J. Meteor. Soc. Japan* **79** 607–623.
- Kar S C, Acharya N, Mohanty U C and Kulkarni M A 2011 Skill of monthly rainfall forecasts over India using multi-model ensemble Schemes; *Int. J. Climatol.* **31** 1–16, doi: 10.1002/joc.2334.
- Kripalani R H and Kumar P 2004 Northeast monsoon rainfall variability over south peninsular India vis-à-vis the Indian Ocean dipole mode; *Int. J. Climatol.* **24** 1267–1282.
- Krishnamurti T N, Mitra A K, Yun W-T and Kumar T S V 2006 Seasonal climate forecasts of the Asian monsoon using multiple coupled models; *Tellus* **58A** 487–507.
- Krishnan A 1984 An analysis of trends in the rainfall and droughts occurring in the southwest and northeast monsoon systems in the southern peninsular India; *Mausam* **35** 379–386.
- Kumar K K, Hoerling M and Rajagopalan B 2005 Advancing dynamical prediction of Indian monsoon rainfall; *Geophys. Res. Lett.* **32** L08704, doi: 10.1029/2004GL021979.
- Kumar P, Rupa Kumar K, Rajeevan M and Sahai A K 2007 On the recent strengthening of the relationship between ENSO and northeast monsoon rainfall over south Asia; *Climate Dynamics* **28** 649–660.
- Lee D E and De Witt D G A 2009 New hybrid coupled forecast system utilizing the CFSSST, 2009 Forecasts http://portal.iri.columbia.edu/portal/server.pt/gateway/PTARGS_0_49_72_5734_0_0_18/NOAA-abstract2009.
- Nageswara Rao G 1999 Variations of the SO relationship with summer and winter monsoon rainfall over India: 1872–1993; *J. Climate* **12** 3486–3495.
- Nayagam L R, Janardanan R and Ram Mohan H S 2009 Variability and teleconnectivity of northeast monsoon rainfall over India; *Global Planet. Change* **69** 225–231.
- Pacanowski R C and Griffes S M 1998 MOM 3.0 manual; NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, 608p.
- Palmer T N and Coauthors 2004 Development of a European multimodel ensemble system for seasonal to interannual prediction (DEMETER); *Bull. Am. Meteorol. Soc.* **85** 853–872.
- Raj Y E A and Jamadar S M 1990 Normal dates of onset and withdrawal of southwest and northwest monsoons over southern peninsular India; *Vayu Mandal* **20** 76–84.
- Rajeevan M, Pai D S, Anil Kumar R and Lal B 2006a New statistical models for long-range forecasting of southwest monsoon rainfall over India; *Climate Dynamics* **28** 813–828.
- Rajeevan M, Bhate J, Kale J and Lal B 2006b High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells; *Curr. Sci.* **91** 296–306.
- Ramaswamy C 1972 The severe drought over Tamil Nadu during the retreating monsoon period of 1968 and its association with anomalies in the upper level flow pattern over the Northern Hemisphere; *Ind. J. Meteorol. Geophys.* **23** 303–316.
- Rao Krishna P R and Jagannathan P 1953 A study of the northeast monsoon rainfall of Tamil Nadu; *Ind. J. Meteorol. Geophys.* **4** 22–43.
- Rao K V 1963 A study of the Indian northeast monsoon season; *Ind. J. Meteorol. Geophys.* **14** 143–155.
- Reynolds R W and Smith T M 1994 Improved global sea surface temperature analyses; *J. Climate* **7** 929–948.
- Roeckner E and Coauthors 1996 The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. Max-Planck-Institut für Meteorologie Rep. 218, Hamburg, Germany, 90p.
- Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van Den Dool H M, Pan H-L, Moorthi S, Behringer D,

- Stokes D, Pena M, Lord S, White G, Ebisuzaki W, Peng P and P Xie 2006 The NCEP climate forecast system; *J. Climate* **19** 3483–3517.
- Sahai A K, Grimm A M, Satyan V and Pant G B 2003 Long-lead prediction of Indian summer monsoon rainfall from global SST evolution; *Climate Dynamics* **20** 855–863.
- Singh N and Sontakke N A 1999 On the variability and prediction of the rainfall in the post-monsoon season over India; *Int. J. Climatol.* **19** 300–309.
- Sridharan S and Muthusamy A 1990 Northeast monsoon rainfall in relation to El Niño, QBO and Atlantic hurricane frequency; *Vayu Mandal* **20** 108–111.
- Srinivasan V and Ramamurthy K 1973 Northeast monsoon. Comprehensive articles on selected topics. Forecasting Manual Part IV, FMU Report No IV-18.4, India Meteorological Department, Poona, India.
- Thapliyal V 1981 ARIMA model for long-range prediction of monsoon rainfall in Peninsular India; India Meteorological Department Monograph Climatology No. 12/81.
- Wang B, Ding Q, Fu X, Kang I-S, Jin K, Shukla J and Doblas-Reyes F 2005 Fundamental challenge in simulation and prediction of summer monsoon rainfall; *Geophys. Res. Lett.* **32** L15711, doi: 10.1029/2005GL022734.
- Yun W-T, Stefanova L and Krishnamurti T N 2003 Improvement of the superensemble technique for seasonal forecasts; *J. Climate* **16** 3834–3840.
- Zubair L 2002 El Niño–southern oscillation influences on rice production in Sri Lanka; *Int. J. Climatol.* **22** 242–250.
- Zubair L and Ropelewski C F 2006 The strengthening relationship between ENSO and northeast monsoon rainfall over Sri Lanka and southern India; *J. Climate* **19** 1567–1575.

MS received 2 October 2010; revised 25 March 2011; accepted 31 March 2011