ORIGINAL PAPER

Forecasting quantitative rainfall over India using multi-model ensemble technique

V. R. Durai · Rashmi Bhardwaj

Abstract A new approach to ensemble forecasting of rainfall over India based on daily outputs of four operational numerical weather prediction (NWP) models in the medium-range timescale (up to 5 days) is proposed in this study. Four global models, namely ECMWF, JMA, GFS and UKMO available on real-time basis at India Meteorological Department, New Delhi, are used simultaneously with adequate weights to obtain a multi-model ensemble (MME) technique. In this technique, weights for each NWP model at each grid point are assigned on the basis of unbiased mean absolute error between the bias-corrected forecast and observed rainfall time series of 366 daily data of 3 consecutive southwest monsoon periods (JJAS) of 2008, 2009 and 2010. Apart from MME, a simple ensemble mean (ENSM) forecast is also generated and experimented. The prediction skill of MME is examined against observed and corresponding outputs of each constituent model during monsoon 2011. The inter-comparison reveals that MME is able to provide more realistic forecast of rainfall over Indian monsoon region by taking the strength of each constituent model. It has been further found that the weighted MME technique has higher skill in predicting daily rainfall compared to ENSM and individual member models. RMSE is found to be lowest in MME forecasts both in magnitude and area coverage. This indicates that fluctuations of day-to-day errors are relatively less in the MME forecast. The inter-comparison of domain-averaged skill scores for different rainfall thresholds further clearly demonstrates that the MME algorithm improves slightly above the ENSM and member models.

V. R. Durai
India Meteorological Department, New Delhi 110003, India

R. Bhardwaj (✉)
Guru Gobind Singh Indraprastha University,
Dwarka, New Delhi 110078, India
e-mail: rashmib@ipu.ac.in

1 Introduction

Forecasting of rainfall over the Indian region is a challenging task, since monsoon constitutes the major weather system that affects the economy of a large population. During last two decades, numerical weather prediction (NWP) models have acquired greater skill and played an important role in the weather forecasting. The rainfall prediction skill of NWP models is still not adequate to address satisfactorily detailed aspects of Indian summer monsoon because of large spatial and temporal variability of rainfall and some inherent limitations of NWP models. The foundation of these models is deterministic model based on some initial conditions, which neglect small-scale effects and also approximate complicated physical processes and interactions. During the 1960s, Lorenz (1963, 1965) investigated the fundamental aspects of atmospheric predictability in the study of butterfly effect and showed that no matter, how good the observations or how good the forecasting techniques, there is almost certainly an insurmountable limit as to how far into the future one can forecast. He demonstrated that weather, even when viewed as a deterministic system, may have a finite prediction time (1963). Further, predictability varies with different weather situations (1965). NWP models lose skill due to model errors arise as a result of errors both in the initial conditions and in the forecast model itself.

In order to overcome the shortcomings in observing systems and model physics, a new approach known as ensemble prediction systems (EPSs) have been used in operations at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Molteni et al. 1996) and the US National Centers for Environmental Prediction (NCEP) (Tracton and Kalnay 1993) and the Canadian Meteorological Centre (CMC) (Houtekamer et al. 1996, 2009). Buizza and Palmer (1998) have shown in their study that the greater the number of ensemble members the greater the skill of the final ensemble prediction. As a result, some operational centers run large numbers of model integrations to produce their ensemble forecasts (Buizza et al. 1999; Kalnay et al. 1998). The implicit assumption with most single-model EPSs is that errors result primarily from uncertainties in the initial conditions. A drawback with this single-model EPS approach is that any biases present in the model itself will also be present in the ensemble.

Another ensemble forecasting approach that has been taken to address the single-model EPS issues are to combine forecasts from more than one NWP model. In this approach, the ensemble is composed of output from different NWP models and/or initial times, rather than a single model with perturbed initial conditions to take into account the uncertainty in the model formulation and initial conditions. Hamill and Colucci (1997, 1998) combined ensembles from the NCEP Eta model and regional spectral model to generate improved short-range forecasts of probability of precipitation. Krishnamurti et al. (1999, 2000) introduced a multi-model super ensemble technique that shows a major improvement in the prediction skill. The consideration of multiple NWP models is based on considering various possible weather scenarios which give the uncertainty in initial conditions, and model formulations lead to more accurate forecasts (Fritsch et al. 2000). The precision of the consensus forecasts should be better as more of the uncertainties are accounted for, and the value of individual consensus members is determined not only by their accuracy but also the relationships between members.

A multi-model multi-analysis ensemble system was reported to evaluate the deterministic forecasts from UKMO and ECMWF ensemble data (Evans et al. 2000) and depicts the superiority of the multi-model system over the single-model data. Richardson (2001) showed that simple ensemble mean and simple bias correction provide better forecasts as compared to the individual member models. Multi-model multi-analysis data were also studied using ECMWF and UKMO ensemble outputs for quasi-operational medium-range forecasting (Mylne et al. 2002) in both probabilistic and deterministic manner. It was observed that the MME is more beneficial than a single-model ensemble prediction system (EPS). Johnson and Swinbank (2009) used the ECMWF, UKMO and GFS

global model data to prepare multi-model ensemble products in medium-range prediction and discussed the forecasts using bias correction, model-dependent weights and variance adjustments. It was observed that multi-model ensemble gives an improvement in comparison with simple ensemble mean. Krishnamurti et al. (2009) claimed that the super ensemble is able to produce the lowest root mean square error (RMSE), providing roughly 20 % improvement over the best model.

Roy Bhowmik and Durai (2010, 2012) used the correlation coefficient (CC) method to find weights for member model and then make a multi-model ensemble (MME) product. However, the benefit of giving weights to member models over a simple ensemble mean (giving equal weights) was not documented there. In recent studies, the MME precipitation forecast for Indian monsoon was reported by Mitra et al. (2011) and Kumar et al. (2012). By downscaling the model data to a uniform resolution and by use of ensemble technique, they have reported skill enhancement of precipitation forecasts in medium range for the Indian monsoon region. Recently, Engel and Ebert (2012) have studied the operational consensus forecasts of 2-m temperature over Australia region using regional and global model outputs and discussed that the bias-corrected weighted average consensus forecast at 1.25° outperforms all models at that scale.

In the present study, a new MME technique is introduced to forecast rainfall over India using four operational global NWP models data available on near real time at India Meteorological Department (IMD), New Delhi. The main purpose of this study was to evaluate the MME rainfall forecast skill improvement over India in short- to medium-range timescale during summer monsoon 2011. Apart from MME, a simple ensemble mean (ENSM) forecast is also generated and experimented. The prediction skill of the member models, ENSM and MME technique is examined and discussed in terms of different statistical skill scores in spatial and temporal scales. This paper comprises of five sections. Section 2 gives a brief description of observed and model data. The MME methodology used in this work is described in Sect. 3. Results of prediction skills are presented in Sect. 4. Finally, the summary and concluding remarks are given in Sect. 5.

2 Data source

2.1 Numerical models

In this study, the day 1 to day 5 rainfall forecast data from four operational global NWP models, namely European Centre for Medium-Range Weather Forecasts (ECMWF), the US National Centers for Environmental Prediction (NCEP)'s global forecasting system (GFS), UK Met Office

Table 1 Characteristics of NWP model used for MME forecasts at IMD

Model (global)	Type	Lag (h)	Spatial resolution	Temporal resolution
ECMWF	Spectral model	00 UTC	0.25	0–168 h, 6 hourly
UKMO	Finite-difference model	00 UTC	1.0	0–144 h, 12 hourly
JMA	Spectral model	00 12 UTC	0.25	0–84 h, 6 hourly
			0.25	0–264 h, 6 hourly
NCEP–GFS	Spectral model	00 UTC	1.0	0–171 h, 3 hourly

(UKMO) and Japan Meteorological Agency (JMA), are used. Details of the models used for this study are provided in Table 1. ECMWF global model (Persson and Grazzini 2007) runs at 20-km horizontal grid resolution and 91 vertical layers. NCEP/GFS (Kanamitsu 1989) runs at 27-km horizontal grid and 64 vertical layers. UKMO model runs at 40-km horizontal grid and 50 vertical layers (Davies et al. 2005). JMA (Saito et al. 2006) model runs at 20-km horizontal grid and 60 vertical layers. Here, the 00 UTC run of all models (except JMA) is used to generate 24-h accumulated rainfall forecast valid for 00 UTC of the next 5 days. For JMA, 00 UTC run is used to generate forecast valid for 00 UTC of day 1, day 2 and day 3, and JMA 12 UTC run the previous day is used to generate 24 h rainfall accumulation of 84–108 h forecast valid for 00 UTC of day 4, and 108–132 h forecast valid for 00 UTC of day 5. The model data used are at 1° × 1° uniform latitude/longitude resolution, to represent the large-scale aspect of the monsoon rainfall. These models were being run at their respective centers (countries) at a higher horizontal and vertical resolution. Thus, the individual NWP model simulation is done at higher spatial scales than the MME results which have been analyzed at 1° × 1° uniform latitude/longitude resolution in this study.

2.2 Observational data

The daily observed (rain gauge) rainfall data from the India Meteorological Department (IMD) are quality controlled and objectively analyzed at 1° × 1° latitude–longitude grid (Rajeevan et al. 2005). The objective technique used for this rainfall analysis is based on the Cressman interpolation method (Cressman 1959). The Cressman weight function used in the objective rainfall analysis is defined by

$$W(r_{i,m}) = \left[\frac{R^2 - r_{i,m}^2}{R^2 + r_{i,m}^2}\right]^2, \quad \text{for} \quad r_{i,m} < R$$
$$= 0, \qquad\qquad \text{for} \quad r_{i,m} \geq R$$

where R is the radius of rainfall influence ($R = 200$ km), and $r_{i,m}$ is the distance of the station from the grid point in km.

The analyzed observed rainfall (rain gauge) used for the study is accumulated rainfall in the 24 h ending 0830 h IST

(0300 UTC). The final daily rainfall analysis data at the resolutions of 1° × 1° are prepared by merging rain gauge observations data for the land areas and Tropical Rainfall Measuring Mission (TRMM) 3B42V6 data for the sea areas (Durai et al. 2010). The temporal and spatial distributions of observed and model predicted rainfall have been studied. The accumulated values of seasonal rainfall, seasonal mean errors and root mean square errors have been compared. High-quality rainfall observations are key elements of MME because the gridded rainfall analysis data are used to update bias correction and compute weights for the member models. In India, the synoptic observation of rainfall is performed over a 24-h period ending at 0300 UTC. Hence, we have verified the model forecast valid at 00 UTC against the observed rainfall analysis (rain gauge) ending at 0300 UTC. The error caused by the timing mismatch is usually negligible compared to model forecast errors.

3 Methodology

In this MME approach, the first step is to estimate the biases of the individual model. The second step is to apply bias correction to each individual model and generate bias-corrected ensemble mean (ENSM) forecast. In the ENSM forecast, all the individual member models have been assigned same weight while carrying out ensemble mean. Initially, equal weight averaging appears to be a poor algorithm because it assigns the same weight to all forecasts irrespective of their relative historical or theoretical merit. It is likely that some models exhibit more skill than other models in certain situations, and the best individual model forecasts are contaminated by the worst model. Therefore, it is sensible to estimate model-dependent weights. In view of this, in the third step, we have computed the model-dependent weights and generated weighted MME forecast.

3.1 Systematic bias estimation

For obtaining the bias-corrected ENSM forecast, the first step is to remove the systematic error (bias) in the

models at each grid point. The bias parameter $(b_{i, j, k})$ at each grid point (i, j) for model k is provided by the best easy systematic estimator (BES). Many meteorological studies (Fritsch et al. 2000; Engel and Ebert 2012) used the BES as the measure of the center of the historical sample in operational consensus forecasting. For the computation of systematic bias at each grid point, the global model (k) forecasts $F_{i,j,k}$ and observed rainfall analysis $O_{i,j}$ of the same forecast hour from three consecutive monsoon periods, i.e., from 1 June to 30 September 2008, 2009 and 2010 (366 days) have been used. The best easy systematic bias (BES) at each grid for each model (k) is given as

$$b_{i,j,k} = (Q_1 + 2Q_2 + Q_3)/4$$

where Q_1, Q_2 and Q_3 are the forecast minus observed $(F_{i,j,k} - O_{i,j})$ error of first, second and third quartiles of the distribution of daily bias errors, respectively.

3.2 Ensemble mean (ENSM)

For obtaining the bias-corrected ENSM forecast, the first step is to remove the systematic error (bias) in the models at each grid. At each grid point, the bias parameter (b_k) for model k is estimated by the best easy systematic (BES) estimator. The bias correction scheme attempts to estimate the current systematic error at each grid using a modified mean of past errors. The bias-corrected ensemble mean (ENSM) forecast is computed as

$$\text{ENSM} = \frac{1}{N} \sum_{K=1}^{4} [F_{i,j,k} - b_{i,j,k}]$$

where N is the total number of models used in the ensemble forecast.

3.3 Multi-model ensemble (MME)

In this method, the random errors of the member models (after bias correction) are then used to generate the weighting parameters to make a MME product. Roy Bhowmik and Durai (2010) used the linear correlation coefficient method to find weights for member model and then make a multi-model ensemble product. In order to generate a MME method, firstly, the model outputs of constituent member models are interpolated at the uniform grid resolution of $1° \times 1°$ lat/long for the domain from lat $0°$ to $40°$ N and long $60°$E to $100°$E. Secondly, the weight for each model and for each grid is determined objectively by computing the unbiased mean absolute error (UMAE) between the bias-corrected forecast and observed rainfall time series of 366 daily data of 3 consecutive southwest monsoon periods (1 June–30 September of 2008, 2009 and

2010). Weights are lower for component forecasts with greater UMAE during the training periods ($t = 1, 2…., 366$ days), and the sum of weights is constrained to equal one. The unbiased mean absolute error (UMAE) is given as

$$\text{UMAE}_{i,j,k} = \frac{1}{N} \sum_{t=1}^{n} \left| F_{i,j,k} - b_{i,j,k} - O_{i,j} \right|$$

where $i = 1, 2, …, 41, j = 1, 2, …, 41$.

The model-dependent weighting parameters $W_{i,j,k}$ for each grid point (i, j) of each model (k) are then obtained using an inverse UMAE from the following equation:

$$W_{i,j,k} = \frac{a_{i,j,k}}{\sum_{k=1}^{4} a_{i,j,k}}$$
$$\text{and } a_{i,j,k} = \frac{1}{(\text{UMAE})_{i,j,k}}$$

where $i = 1, 2, …, 41, j = 1, 2, …, 41$.

For the computational consistency, the UMAE is taken as 0.01 in case UMAE is equal to 0. Deficiency in either bias removal or weighting may result in a suboptimal consensus forecast. Another practical point of consideration is the normalization of the weighting. With not all models available at all hours, the influence of individual models can vary between forecast hours. Normalized weighting of model forecasts also varies between forecast hours due to the relative skill of the models at each hour through the forecast period.

The final MME forecast is generated from the model-dependent weight and bias-corrected component forecasts using the equation.

$$\text{MME} = \frac{1}{N} \sum_{k=1}^{k=4} W_{i,j,k} F_{i,j,k}$$

where $i = 1, 2, …, 41, j = 1, 2, …, 41$.

This method is applied to prepare day 1 to day 5 forecasts of daily rainfall during the summer monsoon (1 June–30 September 2011) using the rainfall prediction of constituent models with the pre-assigned grid point weights.

4 Evaluation of MME prediction skill

The standard procedure for the model rainfall forecast verification (WMO No 1992) is to compute mean error, root mean square error (RMSE) and correlation coefficient (CC) between forecast and analyzed fields valid for the same verification time. In this study, the error statistics (mean error, RMSE, CC and skill score) based on the daily rainfall analysis and corresponding day 1 to day 5 ensemble forecasts have been computed. A quantitative
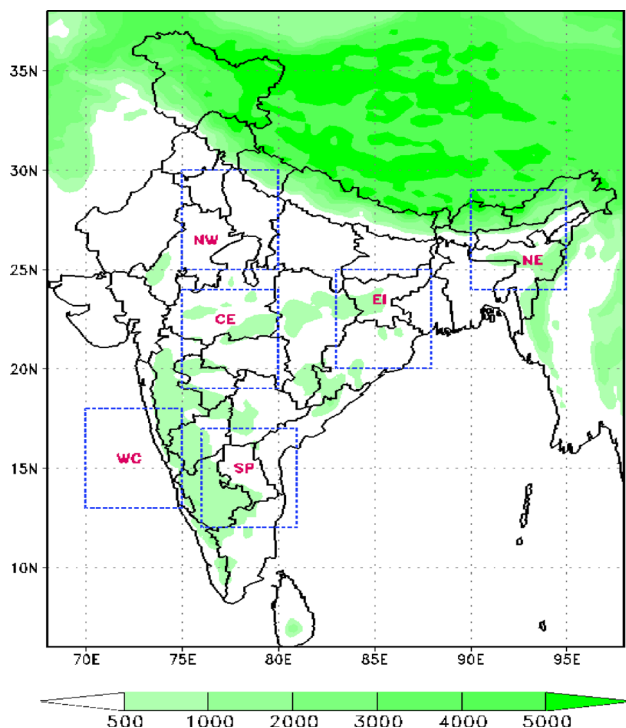
Fig. 1 Topography of India; *shading* indicates areas of terrain elevation greater than 500 m above mean sea level; and the regions of study is shown in outline: NW, northwest India; CE, Central India; EI, eastern India; NE, northeast India; WC, west coast of India and SP, south peninsula India

inter-comparison of error statistics among the constituent models and the bias-corrected ensemble mean and multi-model ensemble forecast are discussed. In order to examine the performance of the model in different homogeneous regions of the country, we selected (Fig. 1) six representative regions for (1) northwest India (NW: lon: 75E–80E, lat: 25N–30N), covering Rajasthan and Haryana, (2) Central India (CE: lon: 75E–80E, lat: 19–24N), covering Vidarbha and neighborhoods, (3) East India (EI: lon: 83E–88E, lat: 20–25N), covering Orissa and neighborhoods, (4) northeast India (NE: lon: 90E–95E, lat: 24N–29N), covering Assam and neighborhoods (5) west coast of India (WC: lon: 70E–75E, lat: 13N–18N), covering Konkan-Goa and (6) south peninsular India (SP: lat 12°N–17°N; lon 76°E–81°E), covering Royalseema and neighborhood. The reason for selecting these domains is related to the known important synoptic circulation features of the large-scale Indian monsoon system and the associated rainfall over the regions.

An approach based on hypothesis testing (Wilks 1995; Hamill 1999) for comparing the performances of two forecast systems is adopted. The hypothesis for difference in RMSE between MME and member models, namely ECMF, UKMO, JMA, GFS and ENSM, is written as

$$H_0 : \mathrm{RMSE}_{\mathrm{MME}} - \mathrm{RMSE}_{\mathrm{MEM}} = 0$$
$$H_a : \mathrm{RMSE}_{\mathrm{MME}} - \mathrm{RMSE}_{\mathrm{MEM}} \neq 0$$

where H_0 is the null hypothesis, and H_a is the alternate hypothesis. $\mathrm{RMSE}_{\mathrm{MME}}$ and $\mathrm{RMSE}_{\mathrm{MEM}}$ are the population RMSEs of the MME and member model.

The paired t test requires a vector of the daily differences in RMSE between MME and individual models. From the sample estimates of the mean and variances, the t test parameter can be written as

$$t_s = \left(|\mathrm{RMSE}_{MME} - \mathrm{RMSE}_{MEM}| / \overline{s_D} \right),$$

where $\overline{s_D} = S_D / \sqrt{n}$, and S_D is the standard deviation of differences in RMSE between MME and individual models. The average RMSE and its sample standard deviation are calculated from the daily differences of 122 (n) days of forecast during monsoon periods (1 June–30 September 2011). The null hypothesis is that there is no underlying difference in quality between the two forecasts, and the alternative hypothesis is that one of the forecasts is better than the other for the user application considered. The significance level (p value) of the difference between the two estimated values is evaluated. If the p value is very small, the data support the alternative hypothesis. If the p value is large, the data support the null hypothesis. The hypothesis H_0 is rejected with $(1 - \alpha)\,100\,\%$ confidence if $|ts| > t\ (1 - \alpha/2),\ (n-1)$. That is $\mathrm{RMSE}_{\mathrm{MME}}$ and $\mathrm{RMSE}_{\mathrm{MEM}}$ differ with a significance level more than $(1 - \alpha)\,100\,\%$ when the above condition is satisfied. The significance level is calculated using the paired t *test*.

The percentage of skill improvement by the MME forecast in terms of RMSE with respect to reference forecasts is given by

$$\mathrm{SKILL}(\%) = \frac{(\mathrm{RMSE}_{\mathrm{REF}} - \mathrm{RMSE}_{\mathrm{MME}})}{\mathrm{RMSE}_{\mathrm{REF}}} \times 100$$

where $\mathrm{RMSE}_{\mathrm{MME}}$ is the RMSE of the MME forecast, and $\mathrm{RMSE}_{\mathrm{REF}}$ represents the RMSE of the reference, i.e., ECMF or ENSM forecasts.

4.1 Spatial distribution of error

In this section, the error statistics and spatial distribution pattern of mean error (ME), mean absolute error (MAE) and RMSE of rainfall in mm/day for day 3 forecast have been discussed. Excellent reviews of forecast verification scores and methods have been carried out by Fuller (Fuller 2004) and Wilks (1995). Bias is the difference between the model forecast and its observation, MAE uses the mean absolute value of the bias, and RMSE is the mean of square root of the sum of the squared bias. The MAE of rainfall for MME and member models, namely ECMF, UKMO, JMA, GFS and ENSM, for day 3 forecast during monsoon
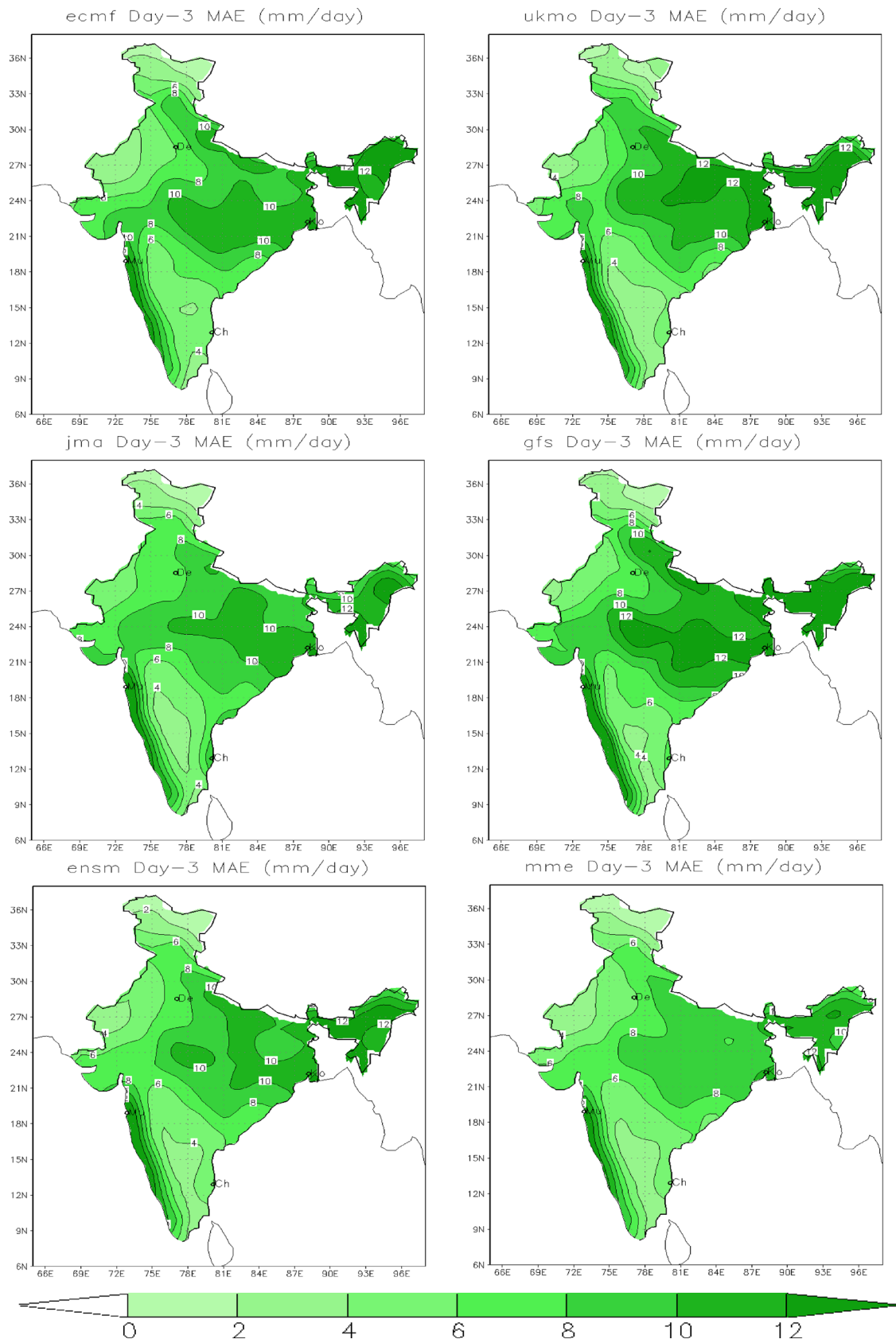
Fig. 2 Spatial distribution of season's (1 June–30 September 2011) mean absolute error (MAE) in mm day^{-1} of ECMF, UKMO, JMA, GFS, ENSM and MME for day 3 forecast

2011 is shown in Fig. 2. At a first glance, the ENSM and MME seem to have less MAE as compared to all the four member models in the day 3 forecast. The GFS model produces larger MAE in all the regions over India as compared to member models and MME forecast. ECMF has smaller MAE among the four member models. All the models produce almost the similar error pattern of larger error over west coast of India in more realistic way. The west coast of India, which is on the windward side of the Western Ghats, interacts with the lower troposphere moist winds of the Arabian Sea to produce heavy rainfall amounts during the summer monsoon season (Rao 1976). The MAE is relatively high over the west coast of India region as compared to other regions of India in all the member models and MME forecast. The NWP model skills are very low for forecasting heavy rainfall amount. Hence, the individual NWP model skill for heavy rainfall amounts has to be improved further to reap any benefit from the MME technique. MME is able to improve upon the individual models for light and moderate rainfall. The MME and the four member models have smaller MAE over NW and SP India. The magnitude of MAE for MME forecast is less as compared to ENSM and the four member models in all the regions of India.

The values of MAE are increased in day 3 forecast as compared to day 1 and day 2 forecasts (figure not shown) in all the models. Day 3 forecast also produce almost the similar error pattern of day 1 forecast with larger error over west coast of India, NE India and East India and smaller error over NW and SP India, but the magnitude of MAE is little larger as compared to day 1 forecast. The MAE biases of member models continue to remain higher than MME in all the days of forecasts with changing magnitudes. In general, MME provides less MAE than the entire member models in all the regions of India.

The spatial pattern of seasonal mean error (bias: forecast − analysis) for the entire 2011 monsoon season for day 3 forecast is shown in Fig. 3. The mean errors of ECMWF, UKMO and JMA model show that the rainfall along the west coast of India is significantly underestimated (by 6–10 mm) in the forecasts. Except ECMWF, all the models overestimate rainfall (positive bias) in day 3 forecast over the northeastern regions except a small portion where model gives negative bias. The magnitude of mean error in the MME forecast is less as compared to member models over all the regions of study. Most of the member models (UKMO, JMA and GFS) show under estimation (negative bias) of rainfall over East India, parts of Central India, west coast of India and southern peninsular India in the forecast. The spatial distribution of ME plot clearly depicts the regions of mean biases in the MME and member model forecasts over India. The MME

products show the least biases compared to member models and ENSM in all the forecast days.

RMSE of the member model and MME forecast in comparison with the observations are given in Fig. 4 for day 3 forecast. For ECMWF, UKMO and JMA, the RMSE ranges between 20 and 25 mm along the west coast in the forecast. The magnitude of RMSE over parts of northeast India and Central India is in the order of 15–20 mm for all the models. GFS has higher RMSE values among member models in all the regions over India. The higher RMSE values are seen over the west coast of India, monsoon trough region. RMSE for MME is smaller than member models and ENSM. MME again has less RMSE as compared to respective member models in day 3 forecast. The magnitude of RMSE values is of the order 10–15 mm for MME over most parts of the country. Broadly, similar pattern of RMSE is observed for ENSM in all the days of forecast. In general, MME have smaller RMSE compared to ENSM and member models in the short-range timescale during the monsoon periods.

4.2 Anomaly correlation coefficient (ACC)

Anomaly correlation coefficient (ACC) of the observation and forecasts are computed from their respective seasonal mean during 2011 monsoon for day 3 (Fig. 5) forecast. The ACC lies between 0.3 and 0.5 over a large part of Central and East India in all the models. The magnitude of ACC decreases with the forecast lead time, and by day 5, ACC values over most parts of India are between 0.2 and 0.4, except in pockets near the east coast and south peninsular India where the ACC values are below 0.2. The magnitude of ACC is higher for MME forecast as compared to ENSM and ECMF, UKMO, JMA and GFS over most of the country. For a sample size of 122 (monsoon days), the ACC is statistically significant at the 99 % confidence level for ACC values exceeding 0.239. Inter-comparison reveals that MME has relatively higher ACC than ensemble mean and member models in all day 1 to day 5 forecasts. MME, ENSM and member models show higher values of ACC along the monsoon trough region and smaller values over NW and SP India in all the 5 days forecasts. The magnitude of ACC decreases with the forecast lead time and by day 3 ACC values over most of India lies between 0.3 and 0.5, except MME and ENSM forecast. MME has higher skill than ENSM and member models in day 3 also. It has been observed that the value of ACC for ECMF, UKMO, JMA and GFS is very low in day 5 (Figure not shown), and may not be having any forecast value. It concludes that for monsoon rainfall forecasts, the current state-of-the-art global models have some skill till day 3, which is enhanced by the MME technique.
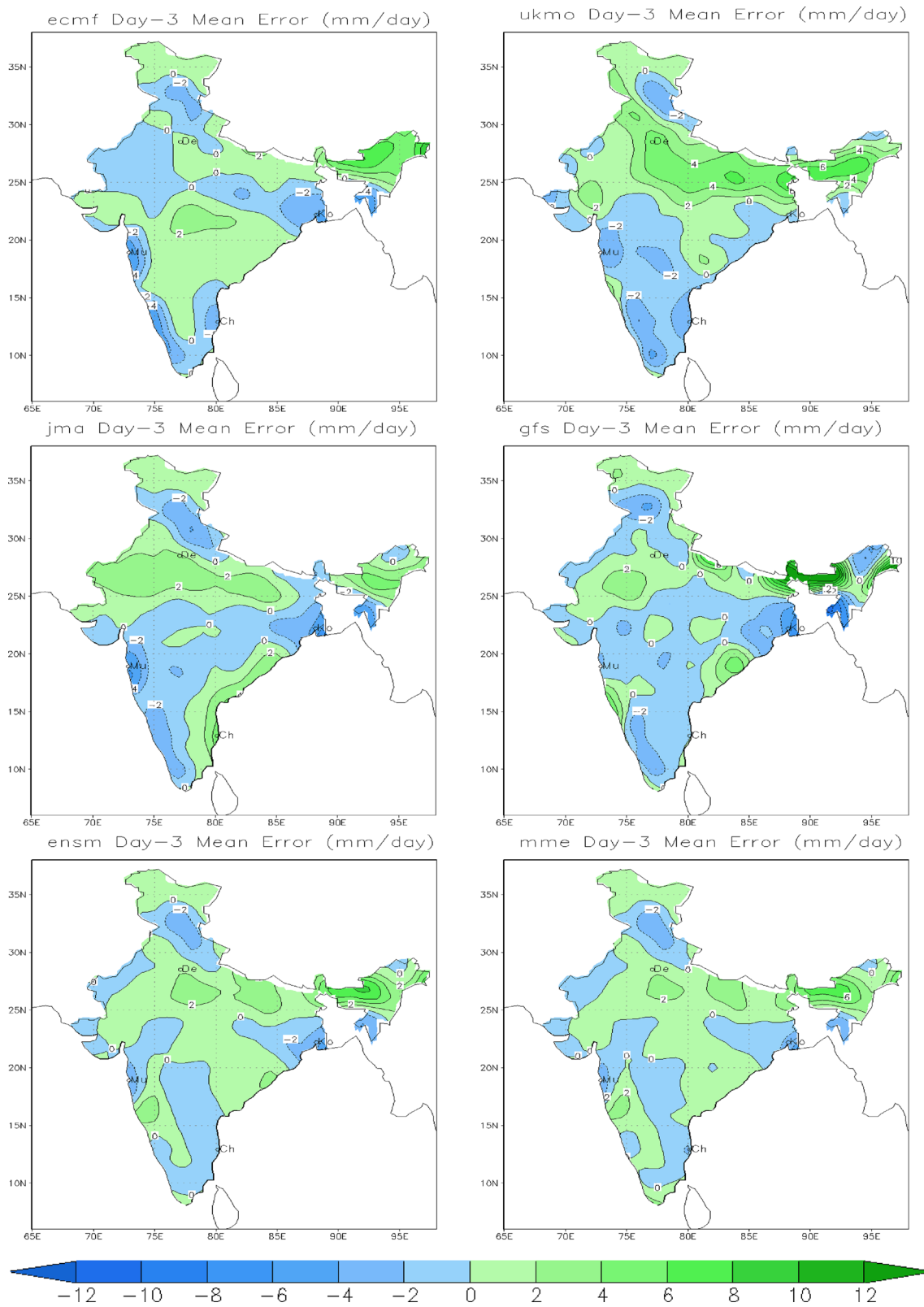
Fig. 3 Spatial distribution of season's (1 June–30 September 2011) mean error (ME) in mm day^{-1} of ECMF, UKMO, JMA, GFS, ENSM and MME for day 3 forecast
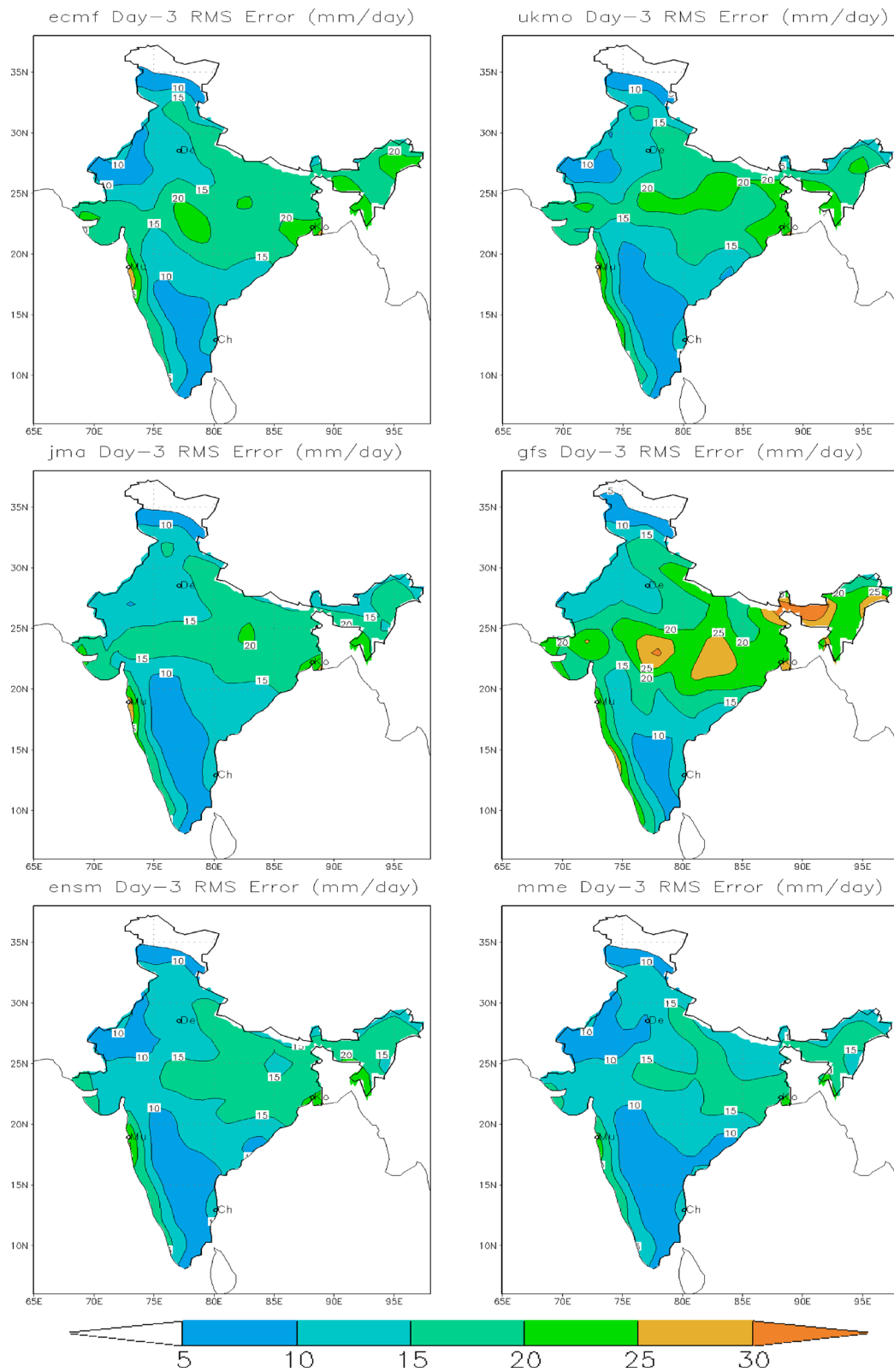
Fig. 4 Spatial distribution of season's (1 June–30 September 2011) root mean square error (RMSE) in mm day^{-1} of ECMF, UKMO, JMA, GFS, ENSM and MME for day 3 forecast
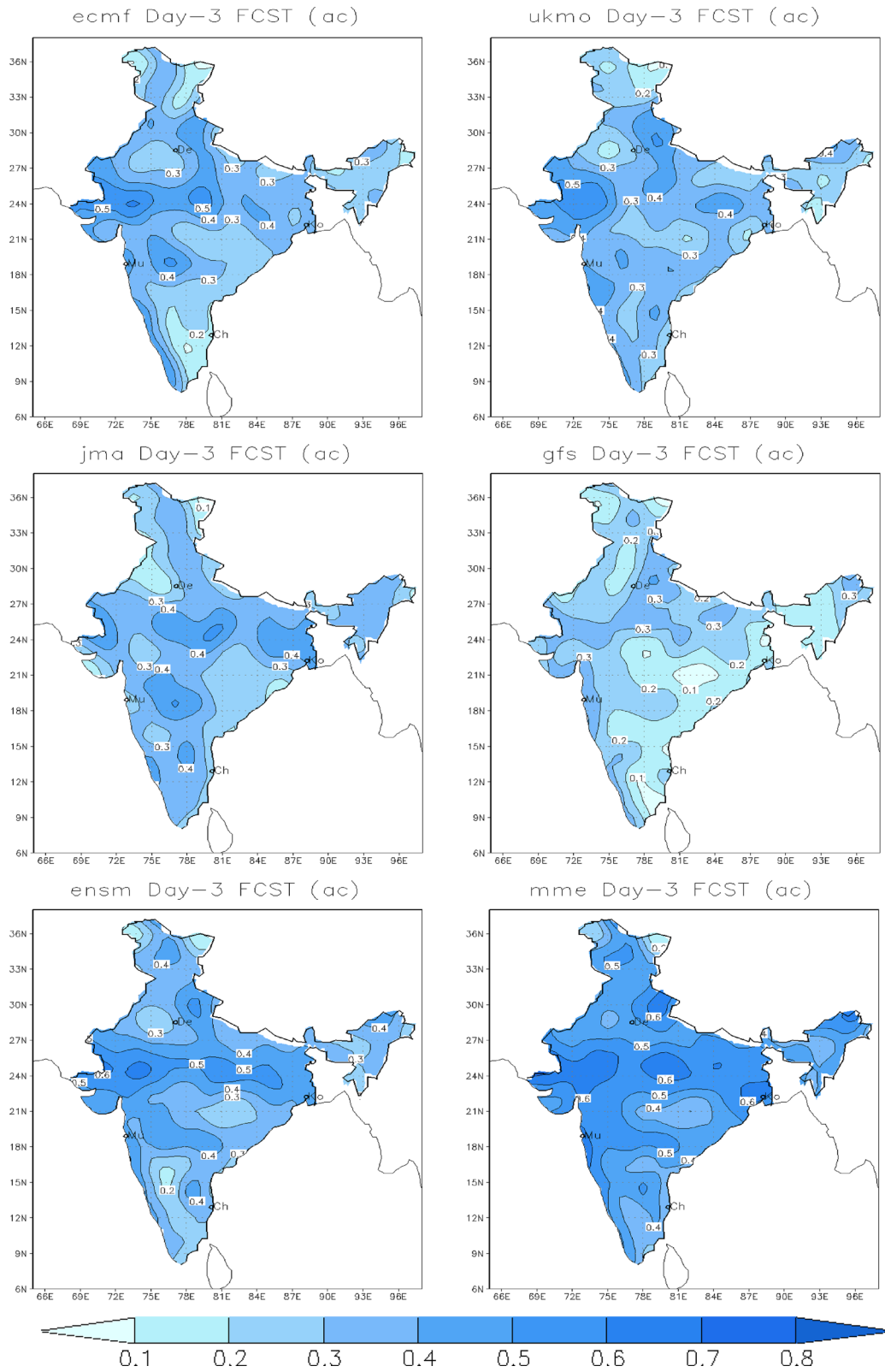
Fig. 5 Spatial distribution of season's (1 June–30 September 2011) anomaly correlation coefficient (ACC) of ECMF, UKMO, JMA, GFS, ENSM and MME for day 3 forecast

4.3 Skill score and statistical significance

The skill of the model is dependent on both the timescale over which the forecasts are being examined and the spatial coverage of the rain itself, i.e., it is easier to predict with reasonable accuracy the probability of rainfall over a large area than a small one, and when the rainfall is widespread rather than localized. The MAE is one of the common measures of forecast accuracy. A box-and-whiskers plot of MAE (mm/day) for day 3 forecast of ECMF, UKMO, JMA, GFS, ENSM and MME over different homogeneous regions of India is shown in Fig. 6. The box-and-whisker plot is a very widely used graphical tool introduced by Tukey (1977). It is a simple plot of five quintiles: the minimum, the lower quartile, the median, the upper

quartile and the maximum. Using these five numbers, the box plot essentially presents a quick sketch of the distribution of the underlying data. Box-and-whiskers plots show the range of data falling between the 25th and 75th percentiles, mean (dot circle), median (horizontal line inside the box) and the whiskers showing the complete range of the data, i.e., lowest and highest values in the data. The mean, median and the range of MAE values are smaller for MME as compared to ENSM and other member models. MME forecast also reduces the error variance significantly in all the regions of study, i.e., CE, EI, NE, NW, WC and SP. The mean, median and error range of MAE for GFS day 3 forecast are larger as compared to other models over almost all the regions of India. The high mean MAE is noticed over NE and WC India, and it is in
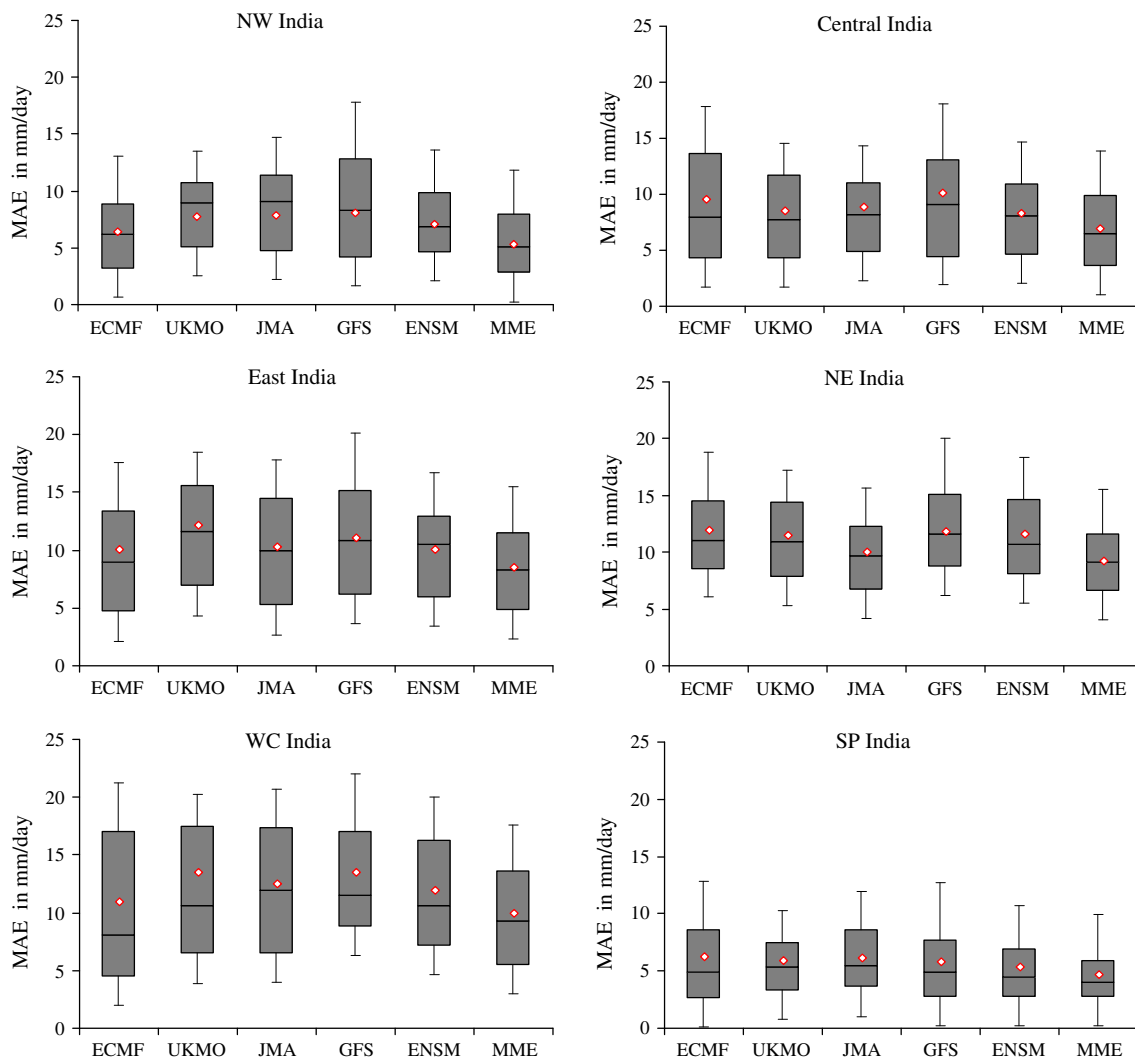

Fig. 6 Box-and-whiskers plots of MAE (mm/day) for day 3 forecast of ECMF, UKMO, JMA, GFS, ENSM and MME over different homogeneous region of India during 1 June–30 September 2011. The *boxes* show the median, 25th and 75th percentiles, and the *red dot* inside the box represents the mean, while the *whiskers* showing the lowest and highest values

the order of 10–15 mm/day, while it is less over SP India in all the member models and MME forecast. In general, MME provides less mean, median and error range in all the regions of India.

The daily mean RMSE of day 1 to day 5 forecasts of member models and MME products for different homogeneous regions of India, i.e., NW India, Central India, East India, NE India, west cost of India and southern peninsula India during monsoon 2011 are shown in Fig. 7. For Central India, the RMSE is in the order of 10–15 mm/day among member models. ENSM and MME have lower RMSE. RMSE for UKMO is lesser than GFS. Also RMSE is more in models where the rainfall amounts is more. The higher value of RMSE (more than 20 mm/day) is observed over NE India and the west coast of India. ENSM and MME have lesser RMSE compared to member models over NE and west coast of India. The magnitude of RMSE errors vary from 2 mm/day over East India to 10–15 mm/day over northeast and west coast of India. In all the regions, the day 1 to day 5 forecasts of MME have less RMSE compared to member models and ENSM.

The performance of both competing forecast model on a given day is related to the synoptic conditions on that day.

So, the hypothesis test should treat the forecast data as paired. On days where the weather is dominated by high pressure, both forecast models are likely to correctly forecast large areas of no precipitation, but on stormy days both forecasts are likely to exhibit generally higher than normal error. Pairing of the data reduces the variance and results in a more powerful hypothesis test. A *paired t test* is carried out for differences in RMSE between MME and member models, namely ECMF, UKMO, JMA, GFS and ENSM (Table 2), over different homogeneous regions of India during monsoon periods (1 June–30 September 2011). The average differences in RMSE (DIFF) and its sample standard deviations (SD) are calculated from the daily differences of 122 days for the day 3 forecast. The differences in RMSE between MME and member models are significantly greater than zero, as indicated by the p value in Table 2. It was found from the significant test on difference in RMSE that the improvement of MME forecast over member models and the ensemble mean forecast is statistically significant at 99 % confidence level mostly over all the homogeneous regions of India in the forecasts.

As shown in Fig. 8, the mean RMSE of MME was lower than both the best individual model (ECMF) and the
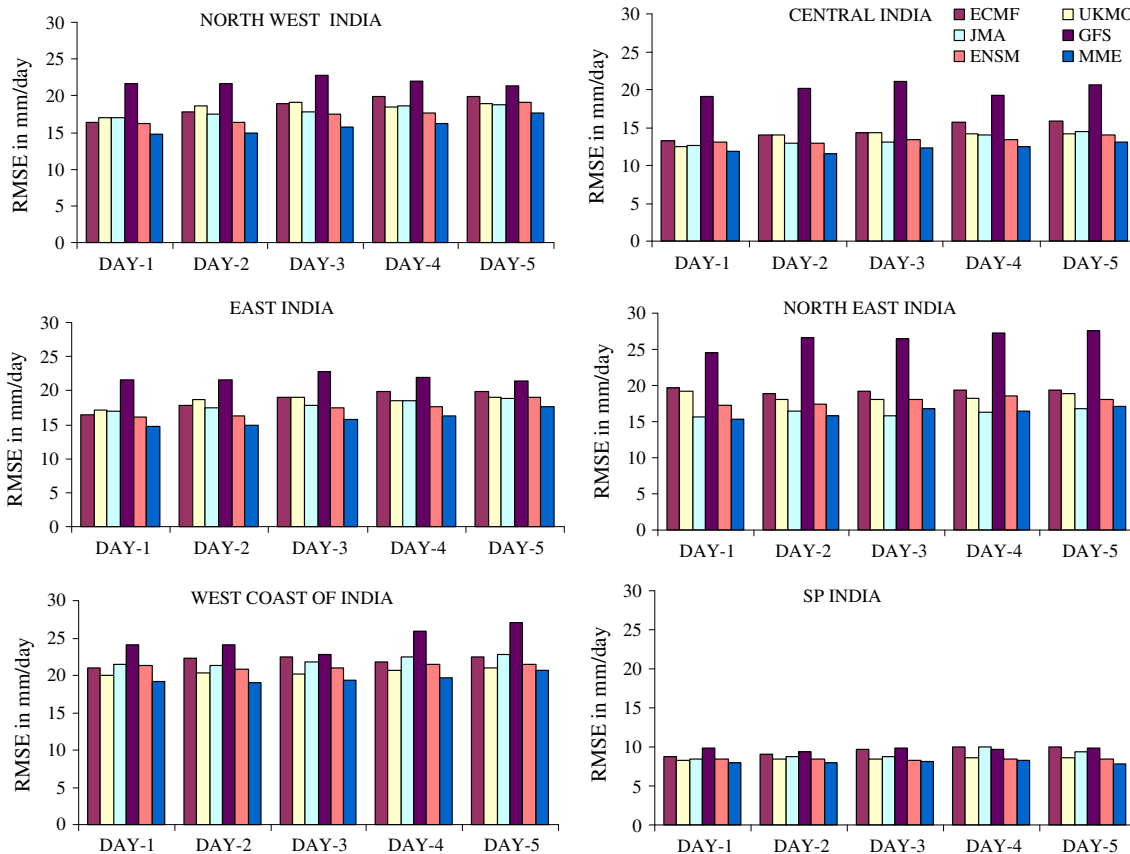

Fig. 7 Mean root mean square error (RMSE) of day 1 to day 5 forecasts of ECMF, UKMO, JMA, GFS, ENSM and MME over different homogeneous region of India during 1 June–30 September 2011

Table 2 Two-sided p values of the significance of differences in RMSE between MME and individual models, namely ECMF, UKMO, JMA, GFS and ENSM, for day 3 forecast over different homogeneous regions of India during summer monsoon (1 June–30 September 2011) season

Model	T test	Central India	East India	NE India	NW India	SP India	WC India
ECMF	DIFF	2	3.24	2.39	2.13	1.57	3.21
	SD	7.35	5.42	5.82	4.31	3.52	6.39
	t (121)	2.99	6.58	4.52	5.42	4.9	5.52
	p value	0.0034	<0.00001	0.000014	<0.00001	<0.00001	<0.00001
UKMO	DIFF	1.97	3.31	1.2	3.26	0.39	0.81
	SD	3.5	5.7	4.64	4.21	1.83	4.6
	t (121)	6.19	6.39	2.83	8.54	2.32	0.93
	p value	<0.0000	<0.00001	0.0054	<0.00001	0.0220	0.0552
JMA	DIFF	0.84	2.09	1.02	2.29	0.71	2.47
	SD	3.03	4.19	3.85	4.25	2.8	5.17
	t (121)	3.07	5.5	2.91	5.92	2.79	5.25
	p value	0.00264	<0.00001	0.0043	<0.00001	0.0061	<0.00001
GFS	DIFF	8.77	7.1	9.56	4.39	1.76	3.5
	SD	10	8.25	8.72	7.93	2.73	7.88
	t (121)	9.64	9.48	12.06	6.09	7.09	4.88
	p value	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
ENSM	DIFF	1.09	1.7	1.2	1.02	0.27	1.64
	SD	2.63	3.63	3.33	2.62	1.37	3.81
	t (121)	4.54	5.15	3.96	4.3	2.15	4.72
	p value	0.000013	<0.00001	0.00013	0.000035	0.0335	<0.00001

ENSM in all day 1 to day 5 forecasts. The skill score ranges from 0 to 100 with value of zero indicating no skill improvement and a value of 100 is for perfect skill. The MME forecast shows improvement in RMSE in the range of 15–20 % as compared to the best individual model (ECMF; Fig. 8a), while the improvement in RMSE is in the range of 5–10 % as compared to ENSM (Fig. 8b) over most of the regions of India in all day 1 to day 5 forecasts. The equal weight ensemble average (ENSM) always improves the RMSE, but the improvement is more significant for weighted ensemble average, i.e., MME forecast. The RMSE skill improvement of MME over west coast of India is low as 10 % and over CE and EI regions are high (15–20 %) compared to the best model. The inter-comparison of RMSE skill improvement reveals that MME have significant skill improvement in RMSE over different homogeneous regions of India by taking the strength of each constituent model and has the potential for operational applications.

The CC between trends in the forecast and observation is a measure of the phase relationship between them. The mean values of daily spatial CC of observed rainfall and day 1 to day 5 forecasts of MME, ENSM and member models over different homogeneous regions of India during summer monsoon 2011 are shown in Fig. 9. It is seen from the Fig. 9 that the mean spatial CC over CE, NW, WC and SP regions is higher compared to EI and NE regions for all models, but MME have higher scores of spatial CC
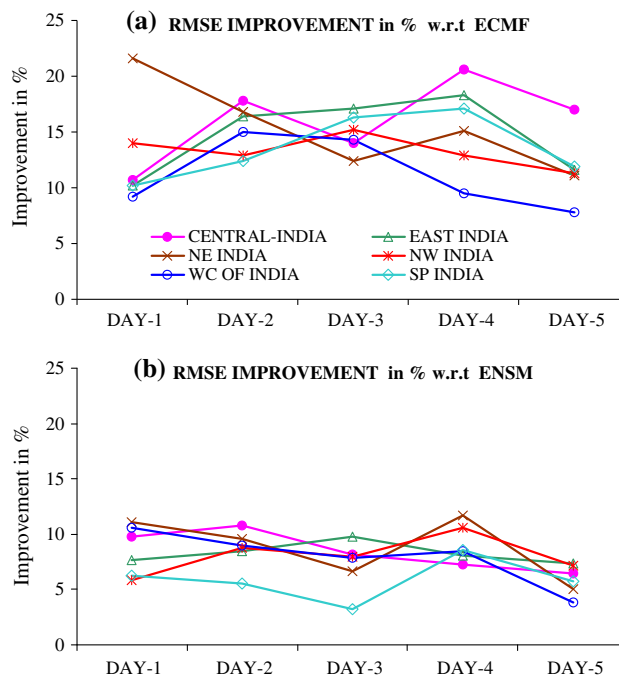

Fig. 8 a Improvement of MME forecast in terms of RMSE (in %) with respect to ECMF; **b** improvement of MME forecast w.r.t ENSM over different homogeneous region of India during monsoon 2011

compared to ENSM and member models. GFS has very low spatial CC over all the regions of study. MME and all the member models have lower spatial CC in day 4 and day

5 as compared to day 1 to day 3 forecasts over CE, EI and NW India. ENSM shows spatial CC similar to the best individual model occasionally over some regions of India (Fig. 9). The lower spatial CC for ENSM as compared to the best individual model is due to the straight average approach of assigning an equal weight of 1.0 to each individual model in the ENSM, which may include several poor models. The average of these poor models degrades the overall results. But, the spatial CC for all day 1 to day 5 forecasts of MME (weighted ensemble) is always higher than the ENSM and the best individual model over almost all the homogeneous regions of India during summer monsoon 2011.

4.4 Threshold statistics

All the above results give some general idea of the quality of rainfall forecasts in terms of error statistics. The statistical parameters based on the frequency of occurrences in various classes are more suitable for determining the skill of a model in predicting precipitation. Therefore, it is relevant to examine the skill of rainfall forecasts in terms of rainfall amounts in different thresholds. Standard statistical parameters such as threat score (TS) also called the critical success index, (CSI, e.g., Schaefer, 1990) and probability of detection (POD) or hit rate (HR) are computed for the comparisons in different categories of rainfall amounts. TS is the ratio of the number of correct model prediction of an event to the number of all such events in both observed and predicted data. HR is the ratio of the number of correctly forecast points above a threshold to that of the number of forecast points above the corresponding threshold.

The HR and TS skill scores are obtained from the data covering the daily values for the entire 2011 monsoon season of 122 days. HR for rainfall threshold of 0.1, 2.5, 5, 10, 15, 20, 25, 30, 35 and 65 mm/day for day 1 to day 5 forecasts of MME and member models over all India during monsoon 2011 is shown in Fig. 10. It is observed that the HR is more than 60 % for class marks below 10 mm/day for day 1, day 2 and day 3 forecast, while it is further below for day 3 and day 5 forecast for all the models. Also, it is shown that skill is a strong function of threshold as well as forecast lead time (day 1 to day 5), with HR decreasing from about 80–90 % for rain/no rain
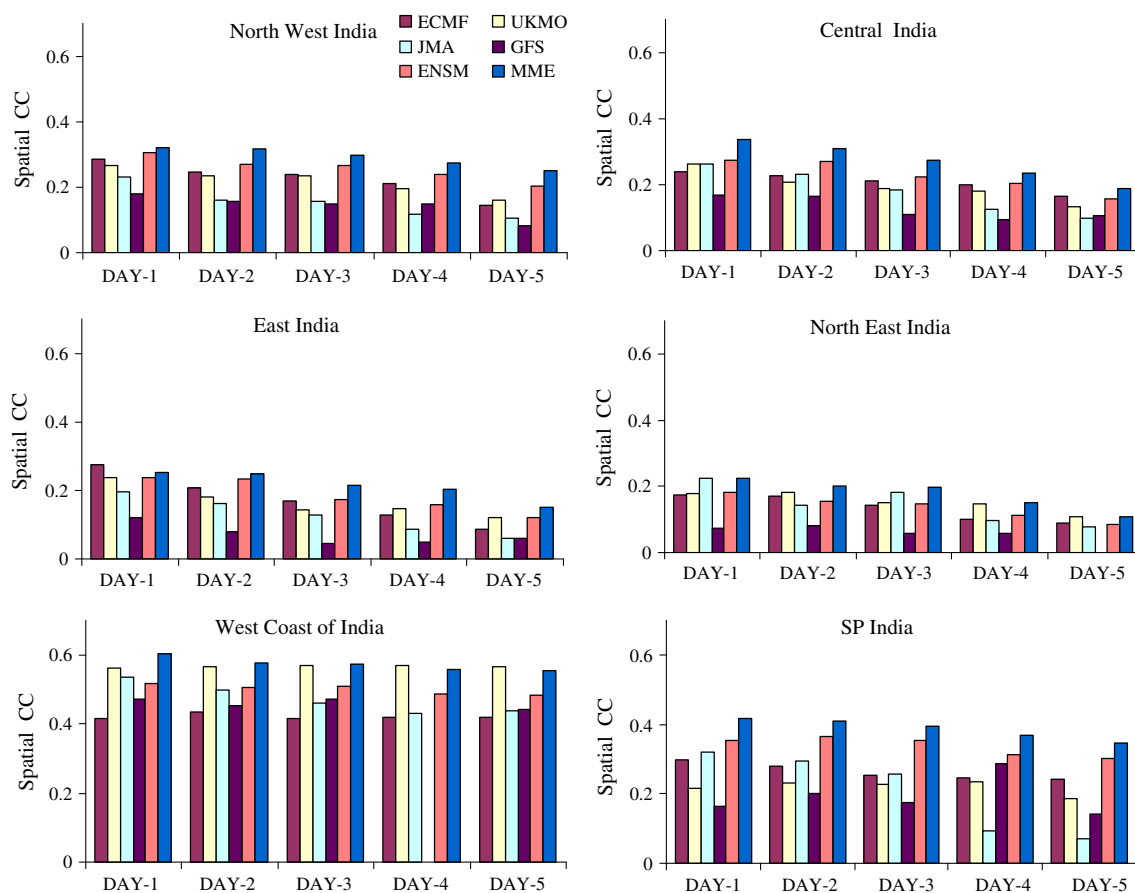

Fig. 9 Mean spatial correlation coefficient (CC) of daily observed rainfall and day 1 to day 5 (24 h accumulated) forecasts of ECMF, UKMO, JMA, GFS, ENSM and MME over different homogeneous region of India during 1 June–30 September 2011
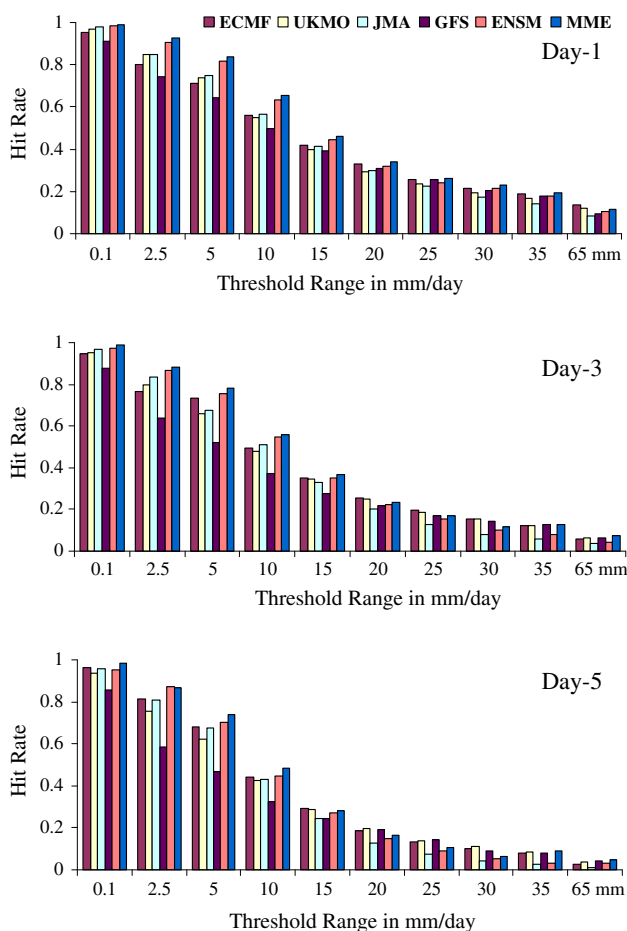
Fig. 10 Hit rate (HR) for rainfall threshold of 0.1, 2.5, 5, 10, 15, 20, 25, 30, 35 and 65 for day 1, day 3 and day 5 forecasts of ECMF, UKMO, JMA, NCEP, ENSM and MME forecasts during monsoon 2011

Fig. 11 Threat score (TS) for rainfall threshold of 0.1, 2.5, 5, 10, 15, 20, 25, 30, 35 and 65 mm/day for day 1, day 3 and day 5 forecasts of ECMF, UKMO, JMA, GFS, ENSM and MME forecasts over All India during monsoon 2011

(>0.1 mm/day) to about 20 or 30 % for rain amounts above 30 mm/day. For HR, the ENSM and MME products are mostly seen to perform better than the member models. For all days, HR from MME and ENSM are superior to the member models in all thresholds range. MME shows slightly higher values of HR skill than ENSM and other member models for all days and for all thresholds over all India regions. The number of data point (n) over Indian land area is 4,485 (x axis 69 and y axis 65), and the standard deviation of HR for MME and member models is in the range from 0.174 to 0.226 for different threshold ranges. The statistical significance test for difference in HR (Fig. 10) is carried out using student's t test over all India and found that this difference in HR between MME and member models, at all threshold range is statistically significant at 95 % confidence level.

Since the formula for HR contains reference to misses and not to false alarms, the hit rate is sensitive only to missed events ra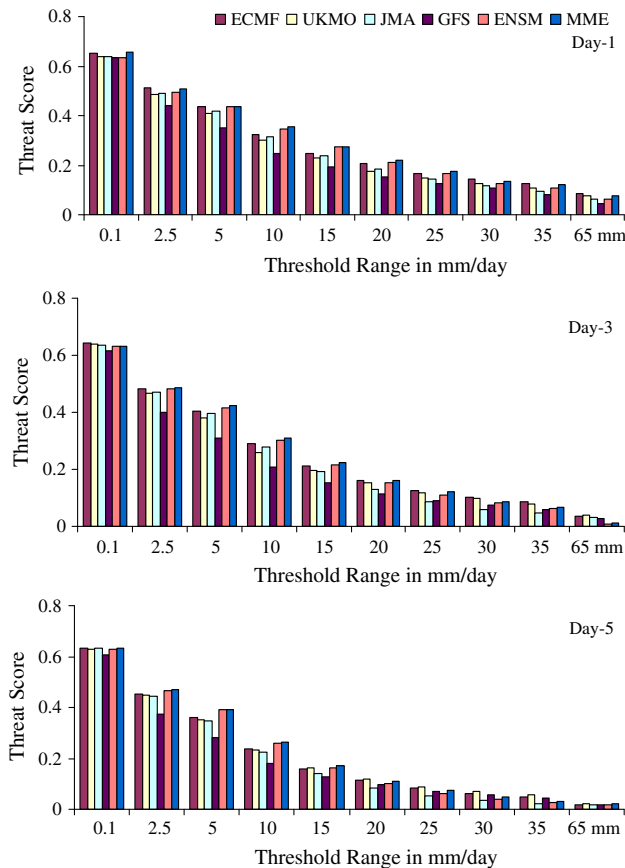ther than false alarms. This means that one can always improve the hit rate by forecasting the event more often, which usually results in higher false alarms and, especially for rare events, is likely to result in an over forecasting bias. However, the TS is typically lower for rare events than for more common events for a particular hit rate. For accuracy, correct negatives have been removed from consideration, i.e., TS is only concerned with forecasts that count. It does not distinguish the source of forecast error and just depends on climatological frequency of events (poorer scores for rarer events) since some hits can occur purely due to random chance. The higher value of a threat score indicates better prediction, with a theoretical limit of 1.0 for a perfect model. TS skill score (Fig. 11) starts close to 0.65 for rainfall threshold of 0.1 mm/day and then decreases to 0.35 near 10 mm mark. TS skill gradually decreases with increase in threshold. Also, with increase in length of forecast period (day 1 to day 5) for each threshold rainfall category, TS skill score falls gradually. TS for all member models looks similar for all forecast length and thresholds. However, in general, TS of MME is slightly higher than ENSM and other member
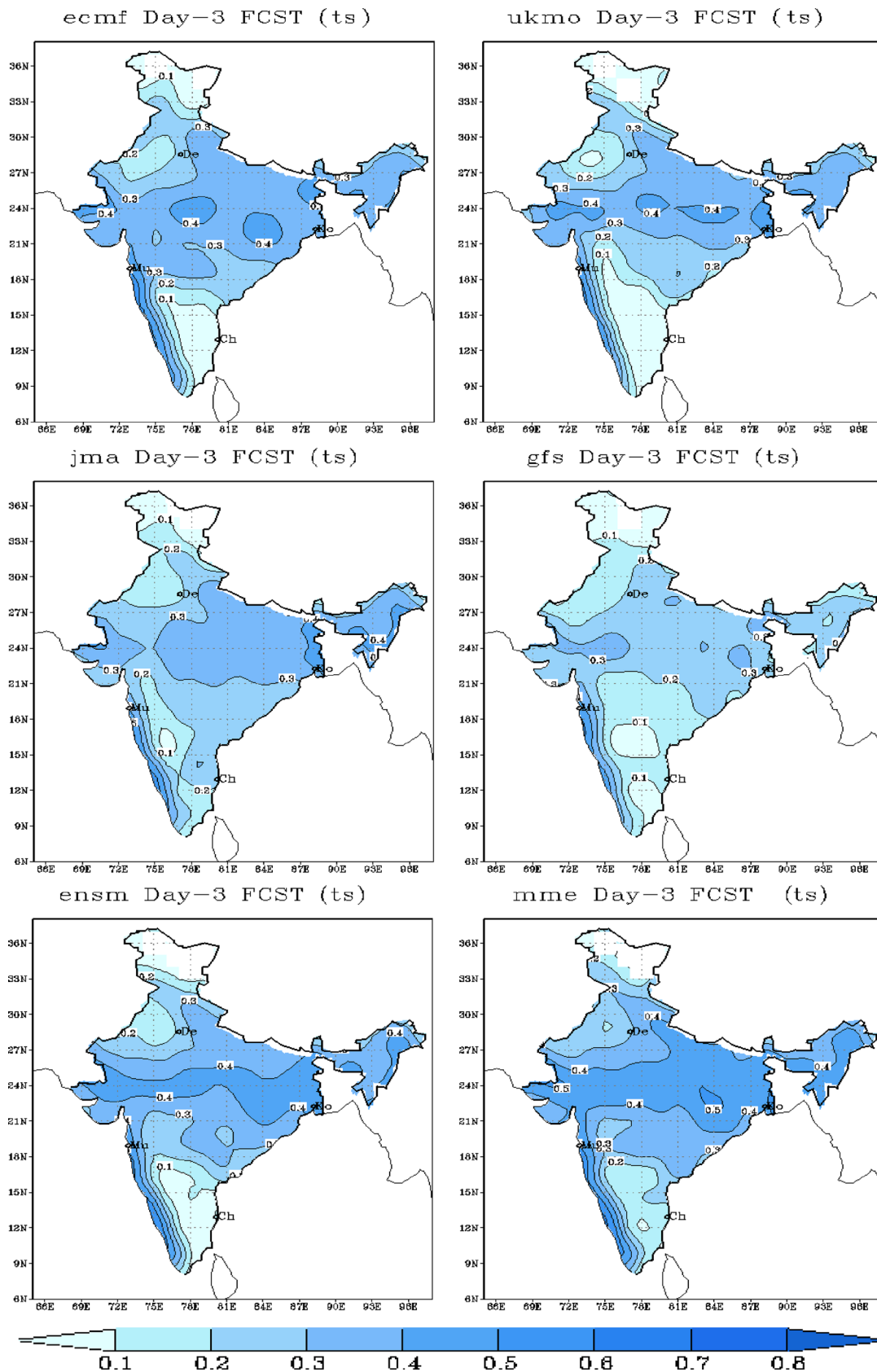
Fig. 12 Spatial distribution of threat score (TS) for rainfall threshold of 15 mm/day for day 3 forecast of ECMF, UKMO, JMA, GFS, ENSM and MME forecasts during monsoon 2011

models for all thresholds even with increase in forecast lengths. A multi-model product for rainfall in terms of TS skill concludes that the use of multi-model has some benefits compared to using single independent models for rainfall forecasts. The number of data point (n) over Indian land area is 4,485 (x axis 69 and y axis 65), and the standard deviation of differences in TS between MME and member models is in the range from 0.134 to 0.186 for different threshold ranges. The statistical significance test for TS (Fig. 11) is carried out using paired student's t test over all India and found that this difference in TS between MME and member models at all threshold range is statistically significant at 95 % confidence level.

The spatial distribution of TS for rainfall threshold of 15 mm/day for day 3 forecast of ECMF, UKMO, JMA, GFS, ENSM and MME forecast is given in Fig. 12. The spatial pattern of TS skill score for all member models looks similar in all days of forecasts. The TS skill score is more than 0.4 over a large part of Central and East India in all the member models except GFS, but the same for MME is slightly higher in day 1 forecast (Fig not shown). The magnitude of TS value for MME forecast is slightly higher in all the regions over India as compared to ENSM and member models. The TS skill score distribution pattern of day 3 forecast looks almost similar to day 1 forecast with larger TS skill score value over west coast of India, NE India and East India and smaller over NW and SP India. The magnitude of TS skill score is smaller in day 3 forecast as compared to day 1 forecast in all the models. In general, the TS skill score of MME remains relatively higher than ENSM and other member models in all the regions of study.

5 Summary and concluding remarks

This study assesses the performance of MME and four NWP models (ECMWF, GFS, UKMO and JMA) to provide rainfall forecasts over Indian region in spatial and temporal scales during summer monsoon season of 2011. The verification of rainfall is done in the spatial scale of 100 km, in a regional spatial scale and also country as a whole in terms of skill scores, such as mean error, root mean square error, correlation efficient and categorical statistics such as HR and TS. The inter-comparison of rainfall prediction skill of MME forecast against the constituent models reveals that MME forecast is able to provide more realistic spatial distribution of rainfall over the Indian monsoon region by taking the strength of each constituent model. The spatial pattern of RMSE and ACC clearly indicates that MME forecast is superior to the forecast of constituent models. RMSE is found to be lowest in MME forecasts both in magnitude and in the area

coverage. This indicates that fluctuations of day-to-day errors are relatively less in case of MME forecast.

The domain-averaged categorical skill scores for different rainfall thresholds clearly demonstrate that the MME algorithm improves slightly above the ENSM and member models in all the day 1 to day 5 forecasts. The method is found to be very useful in forecasting spatial distribution of rainfall over Indian monsoon region. MME forecasting provides better information for practical forecasting in terms of making a consensus forecast and the model uncertainties. For all forecast periods, the MME is more skillful than the individual models according to most statistical measures. The results are statistically significant at 95 % confidence level. The method with three season training data has shown sufficiently promising results for operational applications. The rainfall forecast skill of MME improves by taking the strength of each constituent model and has the potential for operational applications. It remains to be observed what further improvement in MME prediction skill is possible from the construction of the downscaled MME and use of monthly weights for constituent models on the basis of the longer training periods.

References

Buizza R, Palmer TN (1998) Impact of ensemble size on ensemble prediction. Mon Weather Rev 126:2503–2518

Buizza R, Hollingsworth A, Lalaurette F, Ghelli A (1999) Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. Weather Forecast 14:168–189

Cressman GP (1959) An operational objective analysis system. Mon Weather Rev 87:367–374

Davies T, Cullen MJP, Malcolm AJ, Hawson MH, Stanisforth A, White AA, Wood N (2005) A new dynamical core for the Met Office's global and regional modeling of the atmosphere. Quart J Roy Meteor Soc 131:1759–1794

Durai VR, Roy Bhowmik SK, Mukhopadhaya B (2010) Evaluation of Indian summer monsoon rainfall features using TRMM and KALPANA-1 satellite derived precipitation and rain gauge observation. Mausam 61(3):317–336

Engel C, Ebert E (2012) Gridded operational consensus forecasts of 2-m temperature over Australia. Weather Forecast 27(2):301–332

Evans RE, Harrison MSJ, Graham RJ (2000) Joint medium range ensembles from the UKMO and ECMWF systems. Mon Weather Rev 128(9):3104–3127

Fritsch JM, Hilliker J, Ross J, Vislocky RL (2000) Model consensus. Weather Forecast 15:571–582

Fuller, S. R. (2004). Forecast verification: A practitioner's guide in atmospheric science. In: Joliffe IT, Stephenson DB (eds) Wiley, Chichester, 2003. Xiv +240 pp, ISBN 0 471 49759 2, Weather, 59: 132. doi:10.1256/wea.123.03

Hamill TM (1999) Hypothesis tests for evaluating numerical precipitation forecasts. Weather Forecast 4:155–167

Hamill TM, Colucci SJ (1997) Verification of eta-RSM short-range ensemble forecasts. Mon Weather Rev 125:1312–1327

Hamill TM, Colucci SJ (1998) Evaluation of eta-RSM ensemble probabilistic precipitation forecasts. Mon Weather Rev 126:711–724

Houtekamer PL, Lefaivre L, Derome J, Richie H, Mitchell HL (1996) A system simulation approach to ensemble prediction. Mon Weather Rev 124:1225–1242

Houtekamer PL, Mitchell HL, Deng X (2009) Model error representation in an operational ensemble Kalman filter. Mon Weather Rev 137:2126–2143

Johnson C, Swinbank R (2009) Medium-range multimodel ensemble combination and calibration. Quart J Roy Meteor Soc 135:777–794

Kalnay E, Lord SJ, McPherson RD (1998) Maturity of operational numerical weather prediction: medium range. Bull Am Meteor Soc 79:2753–2769

Kanamitsu M (1989) Description of the NMC global data assimilation and forecast system. Weather Forecast 4:335–342

Krishnamurti TN, Kishtawal CM, Larow T, Bachiochi D, Zhang Z, Willford EC, Gadgil S, Surendran S (1999) Improved weather and seasonal climate forecasts from multimodel super ensemble. Science 285:1548–1550

Krishnamurti TN, Kishtawal CM, Shin DW, Willford EC (2000) Improving tropical precipitation forecasts from a multi analysis super ensemble. J Clim 13:4217–4227

Krishnamurti TN, Mishra AK, Chakraborty A, Rajeevan M (2009) Improving global model precipitation forecasts over India using downscaling and the FSU superensemble. Part I: 1–5-Day forecasts. Mon Wea Rev 137:2713–2735

Kumar A, Mitra AK, Bohra AK, Iyengar GR, Durai VR (2012) Multi-model ensemble (MME) prediction of rainfall using neural networks during monsoon season in India. Meteorolog Appl 19(2):161–169

Lorenz EN (1963) Deterministic non-periodic flow. J Atmos Sci 20:130–141

Lorenz EN (1965) A study of the predictability of a 28 variable atmospheric model. Tellus 17:321–333

Mitra AK, Iyengar GR, Durai VR, Sanjay J, Krishnamurti TN, Mishra A, Sikka DR (2011) Experimental real-time multi-model ensemble (MME) prediction of rainfall during monsoon 2008: large-scale medium-range aspects. J Earth Syst Sci 120(1):27–52

Molteni F, Buizza R, Pamer TN, Petroliagis T (1996) The ECMWF ensemble prediction system. Quart J Roy Meteorolg Soc 122:73–119

Mylne K, Evans R, Clark R (2002) Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. Q J R Meteorolg Soc 128:361–384

Persson A, Grazzini F (2007) User guide to ECMWF forecast products, Meteor. Bull, M3.2, ECMWF, Reading, UK, 71 pp

Rajeevan M, Bhate J, Kale JD, Lal B (2005) Development of a high resolution daily gridded rainfall data for Indian region; Met. Monograph Climatology No. 22/2005, India Meteorological Department, Pune, p 26

Rao YP (1976) Southwest monsoon. India Meteorological Department, New Delhi, p 376

Richardson DS (2001) Ensembles using multiple models and analyses. Quart J Roy Meteor Soc 127:1847–1864

Roy Bhowmik SK, Durai VR (2010) Application of multi-model ensemble techniques for real time district level rainfall forecasts in short range time scale over Indian region. Meteorolg Atmos Phys 106(1–2):19–35

Roy Bhowmik SK, Durai VR (2012) Development of multimodel ensemble based district level medium range rainfall forecast system for Indian region. J Earth Syst Sci 121(2):273–285

Saito K et al (2006) The operational JMA non-hydrostatic mesoscale model. Mon Wea Rev 134:1266–1298

Schaefer JT (1990) The critical success index as an indicator of warning skill. Weather Forecast 5:570–575

Tracton S, Kalnay E (1993) Ensemble forecasting at NMC: operational implementation. Weather Forecast 8:379–398

Tukey JW (1977) Exploratory data analysis: reading. Addison-Wesley, Massachusetts 688

Wilks DS (1995) Statistical methods in the atmospheric sciences: an introduction. Academic Press, New York, p 467

WMO No 485 (1992) Manual on the global data processing system, attachment II