Atmósfera 21(3), 225-239 (2008)

Multi-model ensemble forecasting of rainfall over Indian monsoon region

S. K. ROY BHOWMIK and V. R. DURAI
India Meteorological Department
New Delhi-110003, India
Corresponding author: S. K. Roy Bhowmik; e-mail: skrb@imdmail.gov.in

Received May 16, 2007; accepted October 24, 2007

RESUMEN

En este trabajo se propone un método para el ensamble del pronóstico de lluvia sobre la región del monzón de la India con base en las salidas diarias de cuatro modelos operacionales de predicción numérica del tiempo (NWP, por sus siglas en inglés) a escala temporal corta (hasta 48 horas). El método se utiliza para preparar un ensamble de predicción de lluvia de 24 y 48 horas en el modo de prueba diario durante el monzón del verano de 2006, utilizando la predicción de precipitación de los modelos constituyentes con los pesos de los puntos preasignados de la matriz. La habilidad de predicción del ensamble se prueba contra las observaciones y las salidas correspondientes a cada uno de los modelos constituyentes. La intercomparación revela que el método es capaz de mejorar el pronóstico al considerar la fuerza de cada uno de los modelos. El método tiene potencial para ser aplicado operacionalmente.

ABSTRACT

In the present study, a method is proposed for the ensemble forecasting of rainfall over the Indian monsoon region based on daily outputs of four operational numerical weather prediction (NWP) models in the short-range time scale (up to 48 hours). The method is applied to prepare 24 and 48 hours ensemble forecastings of rainfall in the test mode daily during the summer monsoon 2006, using the rainfall prediction of constituent models with the pre-assigned grid point weights. The prediction skill of the ensemble forecasts is examined against observations and corresponding outputs of each constituent model. The intercomparison reveals that the method is capable to improve the forecast by taking the strength of each constituent model. The method has the potential for operational application.

Keywords: Numerical weather prediction, ensemble forecasting, rainfall prediction skill.

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 the last two decades, numerical weather prediction (NWP) methods have acquired greater skill and are playing an increasingly important role in the weather forecasting. Nevertheless rainfall prediction skill of NWP models is still not adequate to address satisfactorily detailed aspects of Indian summer monsoon. This is because of large spatial and temporal variability of rainfall and some inherent limitations of NWP models. Since these models are built on the foundation of deterministic modeling which start with some initial conditions, the inherent limitation of these NWP models is that they neglect small scale effects and they approximate complicated physical processes and interactions. The models lose skill because of the growth of the inevitable uncertainty in the initial conditions. In order to overcome these shortcomings, a new approach known as ensemble forecasting was introduced in the 1990s (Molteni *et al*., 1996; Toth and Kalnay, 1997; Zhang and Krishnamurti, 1997; etc.). In this method, forecasts are made either with different models or different initial conditions or both and are combined into a single forecast to take into account the uncertainty in the model formulation and initial conditions.

The notion of ensemble forecasting was first introduced in the studies of Lorenz (1963, 1965), where he examined the initial state uncertainties and the well known butterfly effect. He noted that the atmosphere is essentially chaotic, because the processes involved in its evolution are non-linear. The study of Lorenz (1963, 1965) showed that no matter how good the observations are, or how good the forecasting techniques, there is almost certainly an insurmountable limit as to how far into the future one can forecast.

In ensemble forecasting, the main issue relates to the removal of the collective errors of multi-models. The major drawback of the straight average approach of assigning an equal weight of 1.0 to each model is that it may include several poor models. The average of these poor models degrades the overall results. To address this problem of ensemble forecasting, Krishnamurti *et al.* (1999, 2000) introduced a multimodel super ensemble technique that shows a major improvement in the prediction skill. In the super ensemble approach, weight is assigned to each model based on spatial and temporal performance of respective models. The strategy for the multimodel super ensemble involves two phases. In the first phase, known as training period, one utilizes the multimodel and observed fields to derive statistics. Weights are generated from the least square minimization of difference between the analysis and the model utilizing a training dataset of 120 days. Daily rainfall analysis prepared on the basis of Tropical Rainfall Measurement Mission (TRMM) and Special Sensor Microwave Imager (SSMI) dataset are used as observed field. In the second phase, also called forecast phase, one utilizes the multimodel forecast and aforementioned statistics to obtain the final super ensemble forecast. A post processing algorithm based on multiple regression of multimodel solutions towards observed fields during the training period shows promising results. The procedure can be used for the basic variables such as wind, temperature, pressure, precipitation and humidity. The resulting super ensemble reduces forecast errors below those of multimodels. They claimed that the super ensemble is able to produce the lowest root mean square error (RMSE), providing roughly 20% improvement over the best model.

In the present study, a method is proposed for the ensemble forecasting of rainfall over Indian monsoon region in short-range time scale (up to 48 hours) based on daily datasets from four operational NWP models available on real time basis at India Meteorological Department (IMD) New Delhi. The method is applied in test mode during the summer monsoon 2006. The prediction skill of the method is examined and discussed in this paper.

2. Methodology and data sources

The NWP models considered for this study are: Limited Area Model (LAM) and Mesoscale Model Version 5 (MM5) operational at IMD New Delhi, and the MM5 and T-80 (grid space ~176 km over tropics) models operational at the National Centre for Medium Range Weather Forecasting (NCMRWF), Noida. The horizontal resolution of NCMRWF MM5 model is 30 km and is run with the initial and boundary conditions from T-80 model. IMD MM5 (hereafter MM5) model is run at the horizontal resolution of 45 km with the initial and boundary conditions from the outputs of Global Forecast System (GFS) of the National Centre for Environmental Prediction (NCEP), available at the horizontal resolution of $1° \times 1°$ lat./long. The horizontal resolution of LAM is 75 km, which is run with initial and boundary conditions from the NCMRWF T-80 model forecasts. As the NCMRWF forecast outputs are available in one day time lag, 48 hours forecast of NCMRWF (T-80 as well as MM5) and MM5 is considered as day-1 forecast and the 72 hours forecast as the day-2 forecast. For the LAM, 24 hours forecast is taken as day-1 forecast and 48 hours forecast as day-2 forecast. A recent study (Roy Bhowmik *et al*., 2007) showed that though LAM, in general, is able to capture three regions of climatologically heavy rainfall domains, viz western Ghats, north-east India and along the monsoon trough, the location and magnitude of predicted rainfall differs considerably from the actual ones. The rainfall performance statistics of NCMRWF T-80 model over Indian monsoon region is documented in recent studies (Basu, 2003, 2005).

In order to develop the method, in the first step, model outputs of constituent models (NCMRWF as well as IMD models) are interpolated at the uniform grid resolution of $1° \times 1°$ lat./long. for the domain from 0 to 40° N and to 100° E. In the second step, the weight for each model and for each grid is determined objectively by computing the correlation coefficient between the predicted and observed rainfall. Daily rainfall analysis (Roy Bhowmik and Das, 2007) at the same resolution ($1° \times 1°$) based on rain gauge observations and satellite estimates (KALPANA-1) is considered as the observed rainfall.

The weights $W_{i,j,k}$ for each grid (i, j) of each model (k) are obtained from the following equation:

$$W_{i,j,k} = \frac{C_{i,j,k}}{\sum_{k=1}^{4} C_{i,j,k}}, \qquad i = 1, 2, \ldots, 41; \quad j = 1, 2, \ldots, 41$$

$C_{i,j,k}$ = Correlation coefficient between rainfall analysis and forecast rainfall for the grid (i, j) of model (k). For the computational consistency, $C_{i,j,k}$ is taken as 0.0001 in case $C_{i,j,k}$ is less than or equal to 0.

The pre-assigned grid point weights are determined for each model using time series of 120 days daily data of south west monsoon (1 June to 30 September) of 2005.

The method is applied to prepare day-1 (24 hours) and day-2 (48 hours) ensemble forecasting of rainfall daily during the summer monsoon (1 June to 30 September) 2006 using the rainfall prediction of constituent models with the pre-assigned grid point weights. In Figure 1, the method is illustrated as a schematic diagram.

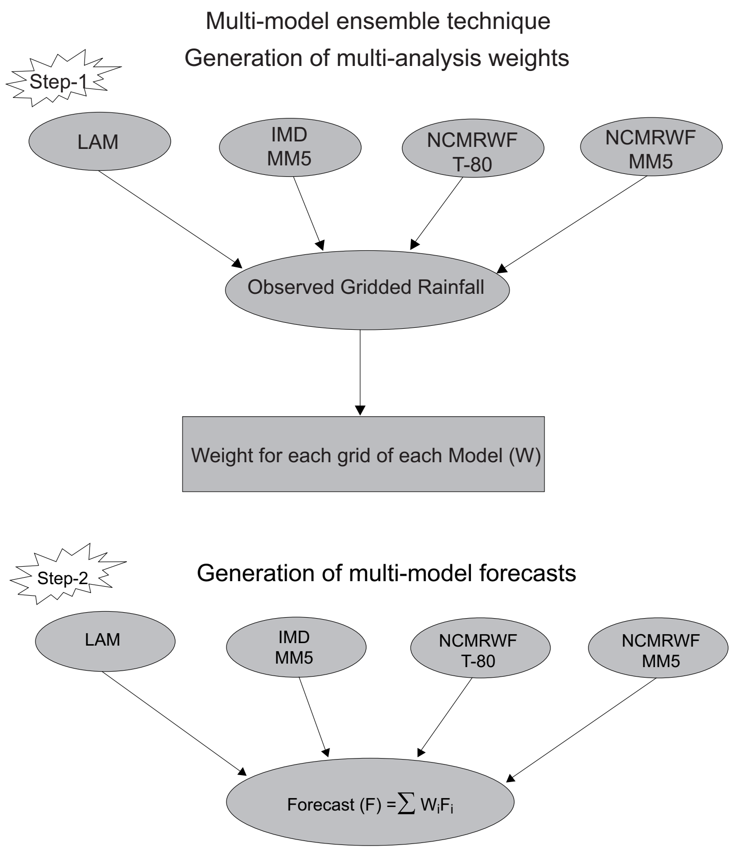

Fig. 1. A schematic diagram showing the ensemble approach.

3. Broad features of summer monsoon 2006

The 2006 summer monsoon was a normal monsoon year with the seasonal rainfall of 100% of the long term average normal. The rainfall was unevenly distributed in space and time. Monsoon advanced over Kerala (extreme south-east state of India) on 26 May, almost a week prior to the normal date. The monsoon covered the entire country on 24 July, about 9 days later than normal date of 15 July. Withdrawal of monsoon from northwest India took place on 21 September, 20 days later than normal date of 1 September. During the season, as many as sixteen low pressure systems formed and out of them eight systems were depressions. One depression formed in July, four depressions in August and three depressions in September. Most of them had long tracks across the central mostly in west/northwest direction. For the purpose of description of the performance of the models over the region, the Indian subcontinent is shown in Figure 2.

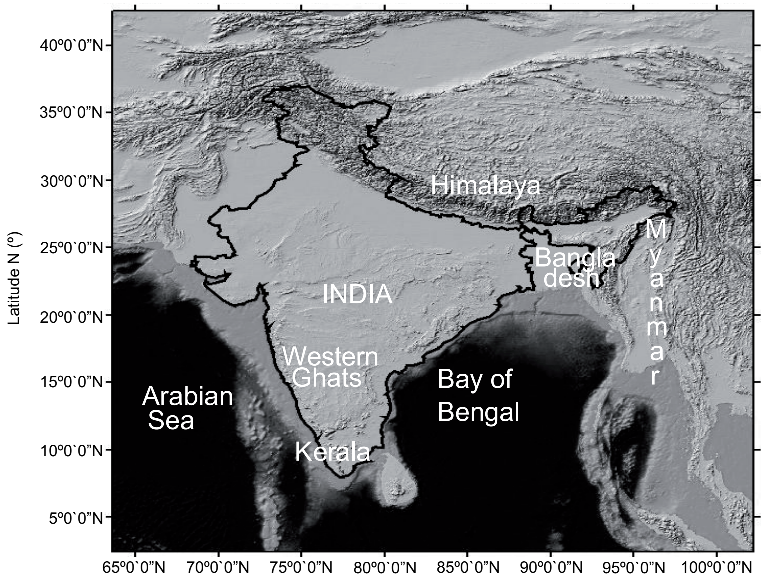

Fig. 2. The Indian sub-Continent.

4. Evaluation of prediction skill

The standard procedure for the model rainfall forecast verification (WMO, 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, we have computed the error statistics (mean error, RMSE, CC and skill score) based on the daily rainfall analysis and corresponding day-1 and day-2 ensemble forecasts. A quantitative inter-comparison of error statistics among LAM, MM5, NCMRWF MM5, T-80 and the ensemble forecast is discussed below.

4.1 Mean rainfall

We begin with the description of seasonal mean rainfall distribution (1 June to 30 September 2006) based on daily analysis (observations) and corresponding day-1 day-2 forecasts of LAM, MM5 (IMD), MM5 (NCMRWF), T-80 and the ensemble. Figures 3(a-f) and 4(a-e) illustrate the corresponding verification maps for the entire monsoon season 2006. The observed rainfall distribution shows a north-south oriented belt of heavy rainfall along the west coast between latitude 8 and 20° N with two peaks (20-25 mm) centered near 10 and 15° N. The sharp gradient of rainfall between the west coast heavy rain region and the rain shadow region to the east is brought out realistically in the objective analysis. Two heavy rainfall pockets (10-15 mm) are observed over the extreme northeastern parts of the country. The rainfall over the domain of monsoon trough has been 10-15 mm.

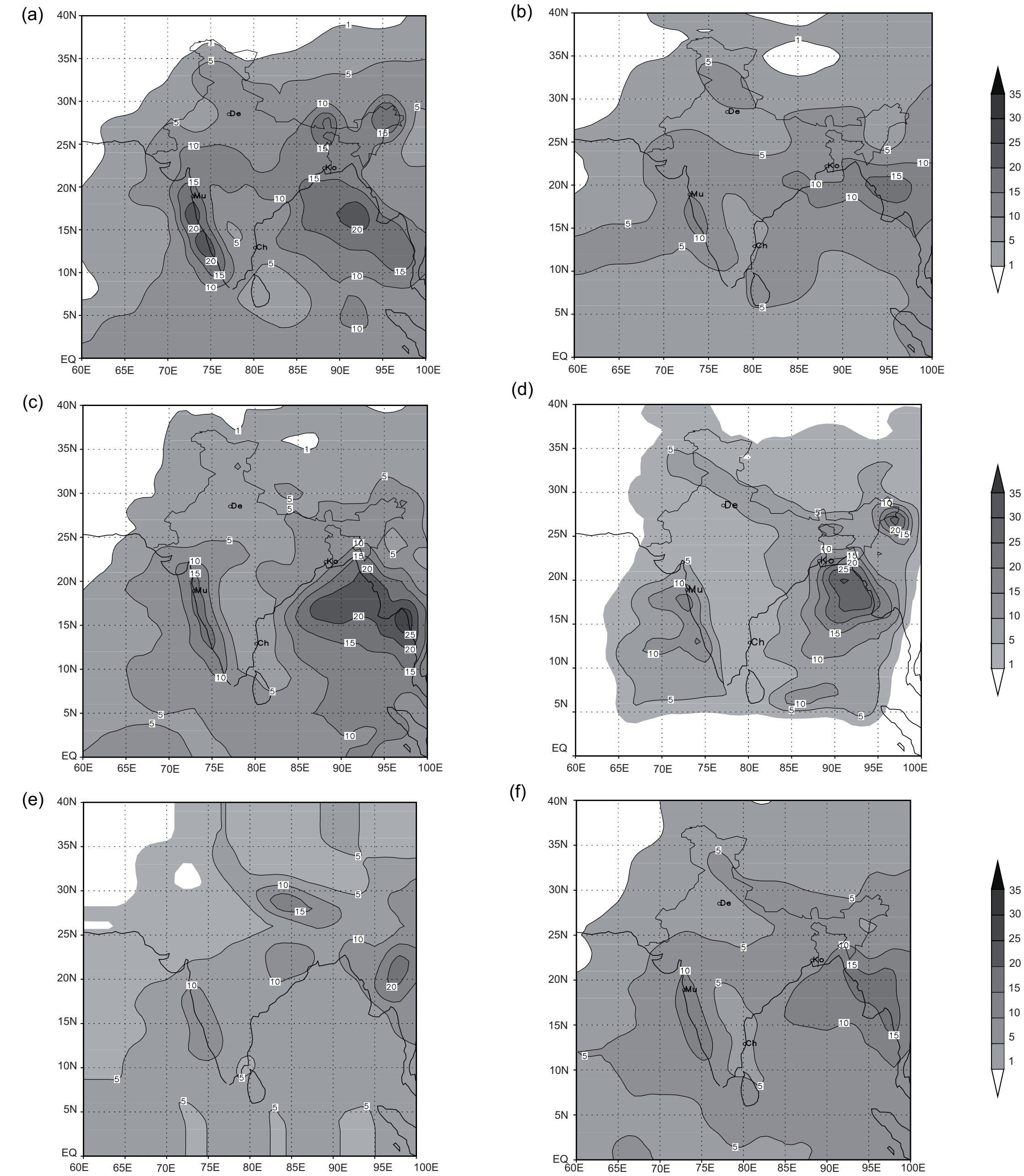

Fig. 3. Spatial distribution of season's (1 June to 30 September 2006) mean rainfall (mm day-1) based on daily (a) rainfall analysis and corresponding day-1 forecasts of (b) LAM, (c) MM5, (d) MM5 (NCMRWF), (e) T-80 and (f) ensemble.

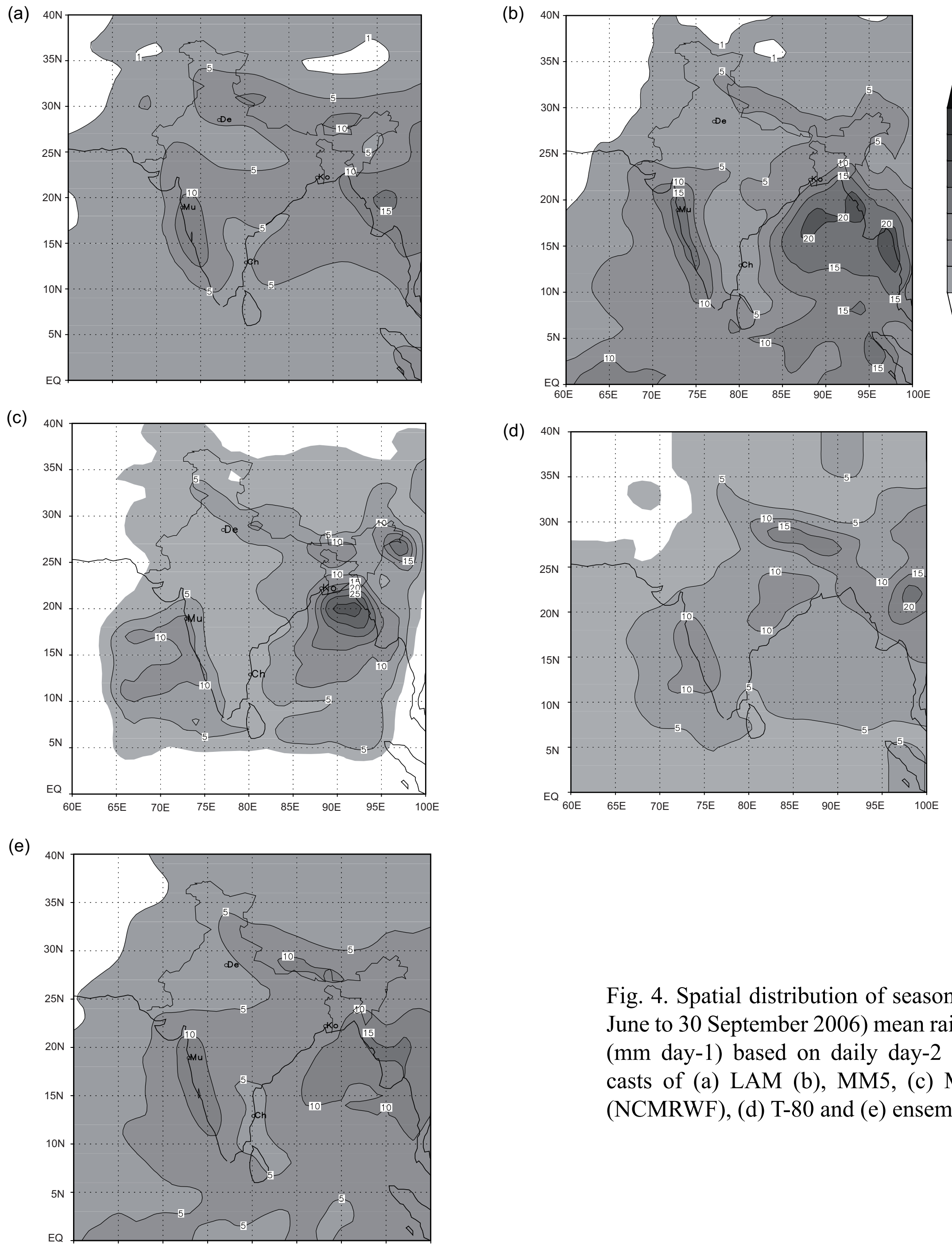

Fig. 4. Spatial distribution of season's (1 June to 30 September 2006) mean rainfall (mm day-1) based on daily day-2 forecasts of (a) LAM (b), MM5, (c) MM5 (NCMRWF), (d) T-80 and (e) ensemble.

The corresponding day-1 as well as day-2 forecast mean fields of LAM show north-south oriented heavy rainfall along the west coast with a peak of order 10-15 mm centred at 15° N. Another heavy rainfall belt (15-20 mm) lies over the northeast Bay of Bengal and adjoining Myanmar areas. Over the monsoon trough region rainfall has been 5-10 mm. The day-2 rainfall pattern broadly remains the same where additionally another rainfall belt (5-10 mm) prevailed along the foothills of Himalayas.

The MM5 day-1 and day-2 forecast fields describe a north-south oriented belt of heavy rainfall along the west coast between 8 and 20° N with a peak centered near 15° N (15-20 mm). The pattern is found to be closer to the observation. Another heavy rainfall belt (20-25 mm) is located over the north Bay of Bengal and adjoining Myanmar areas.

The MM5 (NCMRWF) day-1 and day-2 forecast fields describe a north-south oriented belt of heavy rainfall along the west coast between latitude 8 and 20° N with a peak centered near 15° N (10-15 mm). Other heavy rainfall belt (20-25 mm) is located over the northwest Bay of Bengal, and another one is noticed along the foothills of the Himalayas. A small pocket of heavy rainfall (20 mm) is located over the extreme northeast India.

The corresponding day-1 as well as day-2 forecast mean fields of T-80 show north-south oriented heavy rainfall along the west coast with a peak of order 10 mm centered at lat 15° N. Another heavy rainfall belt (15-20 mm) lies along the foot hills of Himalayas. A small pocket over east-central shows rainfall of order 10 mm. The day-2 rainfall pattern broadly remains the same.

The day-1 as well as day-2 forecast mean fields of the ensemble forecast could reproduce north-south oriented heavy rainfall along the west coast with a peak of order 10-15 mm extending between 10 and 20° N. The heavy rainfall belt (10-15 mm) over the north Bay of Bengal is found to extend from Myanmar coast to west-central Bay of Bengal. Over the monsoon trough region the rainfall has been 5-10 mm.

The comparison reveals that LAM performs better over the domain of monsoon trough, but significantly underestimates orographic rainfall along the Western Ghats and over the north Bay of Bengal. On the contrary, the MM5 models, particularly IMD MM5 model, perform nearly realistic over the north Bay of Bengal, along the Western Ghats extending northwards in the area of offshore trough and over the domain of mid tropopheric cyclone (MTC), but fails to capture heavy rainfall belt over the domain of monsoon trough. The rainfall by T-80 model is found to be underestimated, particularly over the domain of monsoon heavy rainfall belts. Both the NCMRWF models (T-80 as well as MM5) show a rainfall belt along the foot hills of Himalayas extending from extreme northeast India to northwest India. This feature is also reflected in the ensemble forecast. It is interesting to note that the ensemble forecast could take the advantage of LAM for the domain of monsoon trough and advantage of MM5 for the domain of monsoon low pressure area, and for the orographic rainfall along the Western Ghats.

4.2 Mean error

Figures 5 (a-e) and 6(a-e) present spatial pattern of mean error (forecast-analysis) for the entire monsoon season for LAM, MM5, MM5 (NCMRWF), T-80 and the ensemble forecast based on day-1 and day-2 forecasts. The mean errors of LAM, both for day-1 and day-2 show that the rainfall along the Western Ghats and adjoining sea areas, over the extreme northeastern parts of the country, over the parts of southeast Arabian Sea and over the north Bay of Bengal, is significantly under-estimated (~10 mm). Rainfall over Myanmar and adjoining areas are over-estimated. The MM5 (IMD) forecast for both day-1 and day-2 shows underestimation of rainfall over Western Ghats (~5 mm), and central and eastern parts of the country. The over-estimation of rainfall is noticed over the Myanmar areas. The MM5 (NCMRWF) forecast for both day-1 and day-2 show underestimation of rainfall (~5 mm) over Western Ghats and over the domain of monsoon trough. The over-estimation of rainfall is noticed over the northwest Bay of Bengal. T-80 shows overall underestimation of rainfall over most parts of the country. In the ensemble forecast mean error is found to be reduced considerably compared to the individual models. Along the Western Ghats and adjoining sea areas an underestimation of rainfall of order 5 mm took place. Over the extreme northeastern parts of the country and over the north Bay of Bengal the amount of underestimation has been 5-10 mm. Rainfall over Myanmar and adjoining areas are over-estimated. Otherwise, for the remaining parts of the region the forecast is close to observation. Similar feature persisted in the day-2 forecast also.

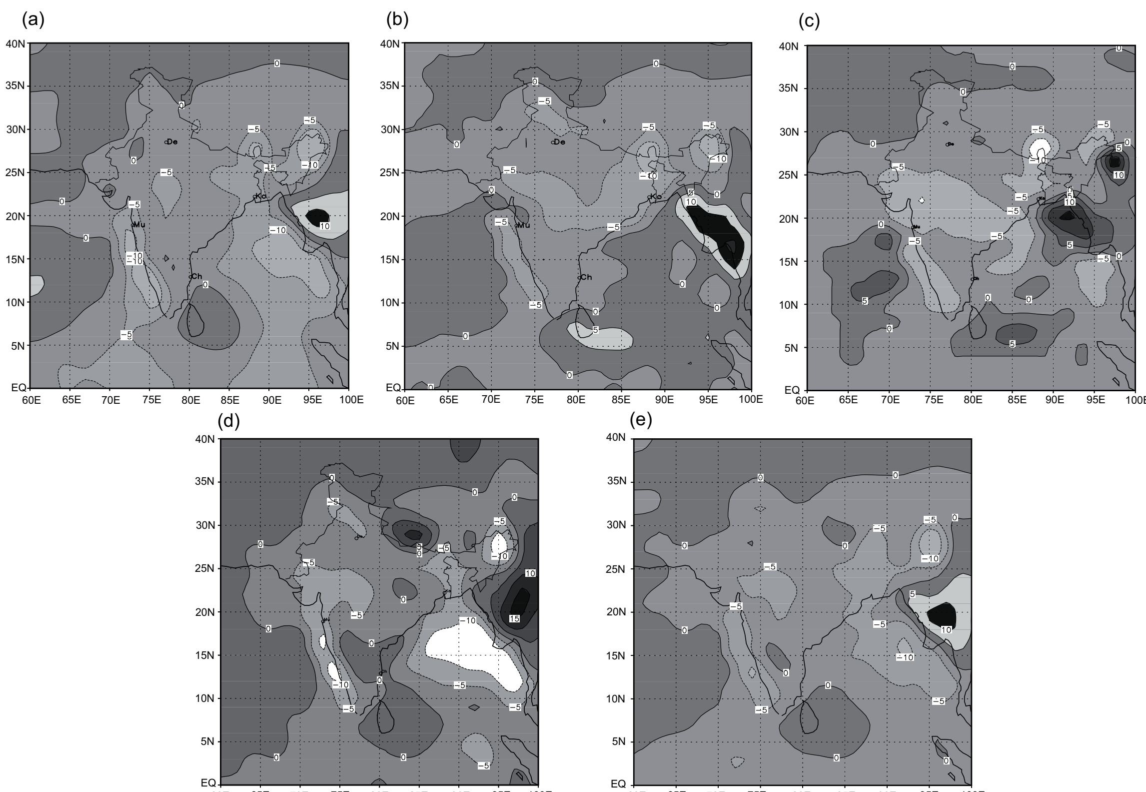

Fig. 5. Spatial distribution of season's (1 June to 30 September 2006) mean error (Forecast-analysis) in mm day-1 based on daily day-1 forecasts of (a) LAM, (b) MM5 (c) MM5, (NCMRWF), (d) T-80 and (e) ensemble.

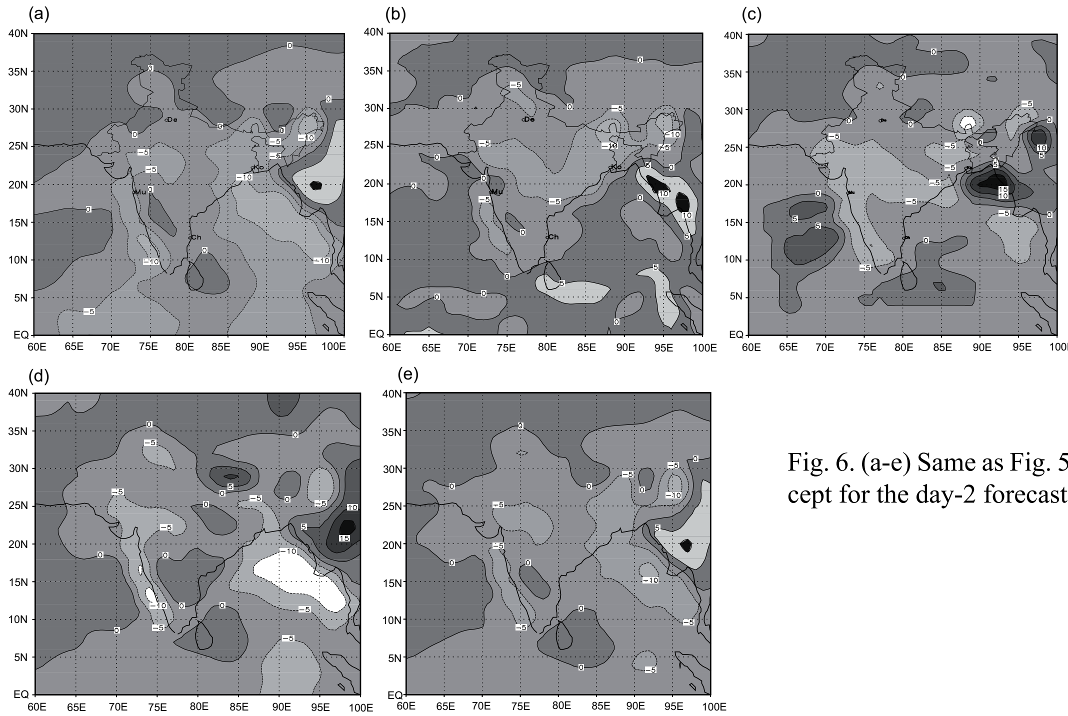

Fig. 6. (a-e) Same as Fig. 5 except for the day-2 forecasts.

4.3 Root mean square error (RMSE)

Figures 7(a-e) and 8(a-e) illustrate the RMSE for day-1 and day-2 forecasts of LAM, MM5, MM5 (NCMRWF), T-80 and the ensemble forecast. For LAM, both day-1 and day-2 RMSE ranges between 20-30 mm along the west coast, over parts of the north Bay of Bengal, Gangetic West Bengal and in some pockets along the foothills of Himalayas. Broadly similar pattern is noticed in case of MM5 and T-80 models. The comparison very clearly shows that the RMSE has been lowest in the ensemble forecasts, particularly in the spatial distribution. This feature is noticed in both day-1 and day-2 forecasts.

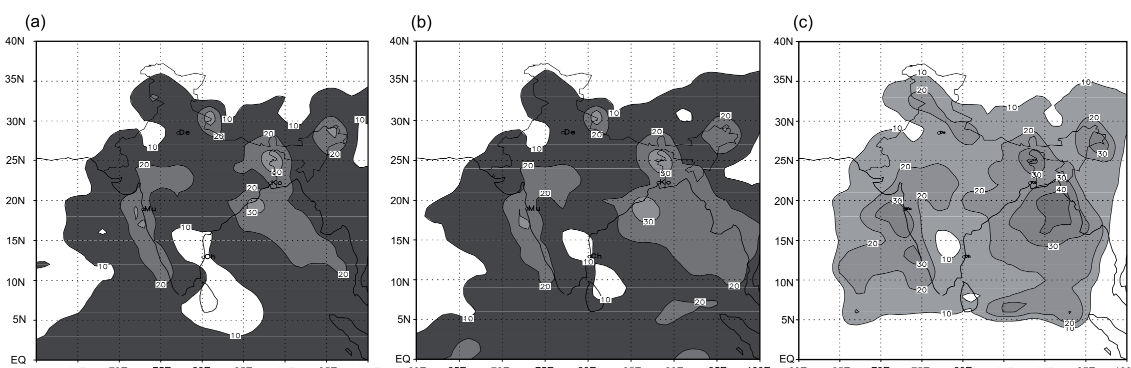

Fig. 7. Spatial distribution of season's (1 June to 30 September 2006) RMSE in mm day-1 based on daily day-1 forecasts of (a) LAM (b) MM5 , (c) MM5 (NCMRWF), (d) T-80 and (e) ensemble.

Continues in the next page.

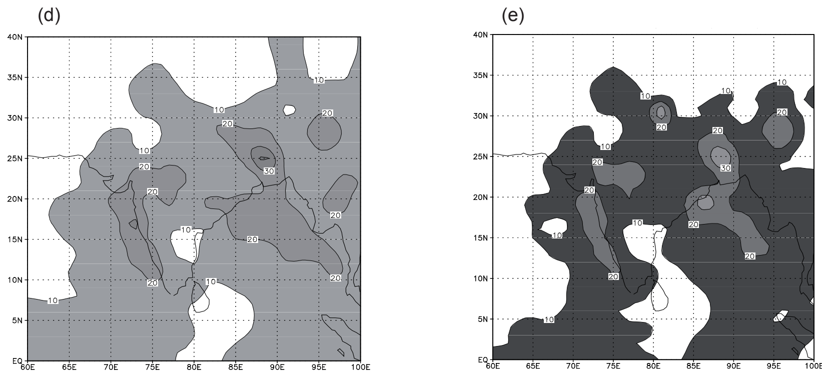

Fig. 7. Spatial distribution of season's (1 June to 30 September 2006) RMSE in mm day-1 based on daily day-1 forecasts of (a) LAM (b) MM5 , (c) MM5 (NCMRWF), (d) T-80 and (e) ensemble. Continued.

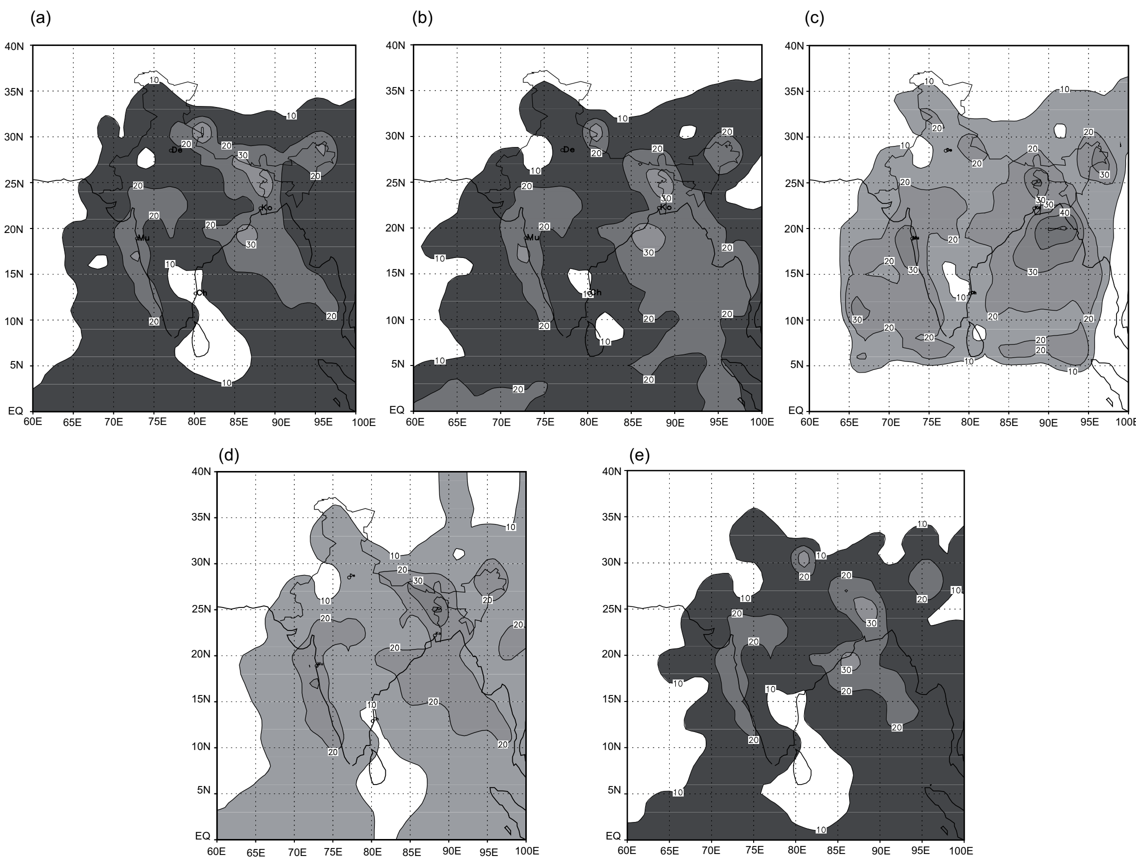

Fig. 8 (a-e) Same as Fig. 7 except for the day-2 forecasts.

4.4 Correlation coefficient (CC)

Figures 9(a-e) and 10(a-e) show the spatial distribution of CC between forecast and observations for day-1 and day-2 forecasts of LAM, MM5, MM5 (NCMRWF), T-80 and ensemble. The distribution of CC for LAM day-1 forecast ranges between 0.4 to 0.5 over the domain of monsoon trough, MTC and over the northwest Bay of Bengal. In the day-2 forecast, CC reduces, but the spatial pattern remain the same over these areas. The CC for the MM5 day-1 forecast has been 0.4 to 0.5 over the parts of Arabian Sea, Western Ghats, and north Bay of Bengal. The CC has been around 0.3 over some pockets over the land. The CC slightly reduces in the day-2 forecast. The CC for the MM5 (NCMRWF) has been good (0, 4) over the domain of MTC and for T-80 it has been good over the foot hills of Himalayas. The CC slightly reduces in the day-2 forecast.

It is very encouraging to note that in the ensemble forecast higher CC values (0.4 to 0.5) occupied larger areas covering Arabian Sea, entire west coast, central India, north Bay of Bengal and some pockets over north India. This feature is found to be persistently prevailed in the day-2 forecast.

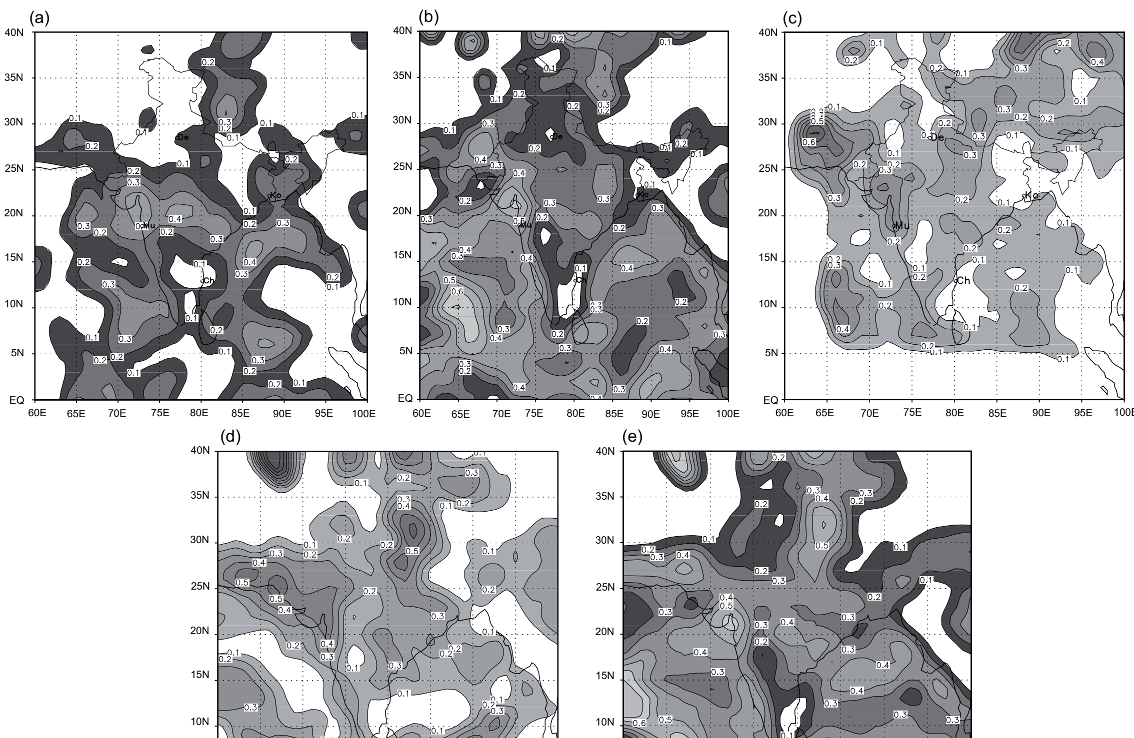

Fig. 9. Spatial distribution of season's (1 June to 30 September 2006) CC based on daily day-1 forecasts of (a) LAM (b) MM5, (c) MM5 (NCMRWF), (d) T-80 and (e) Ensemble.

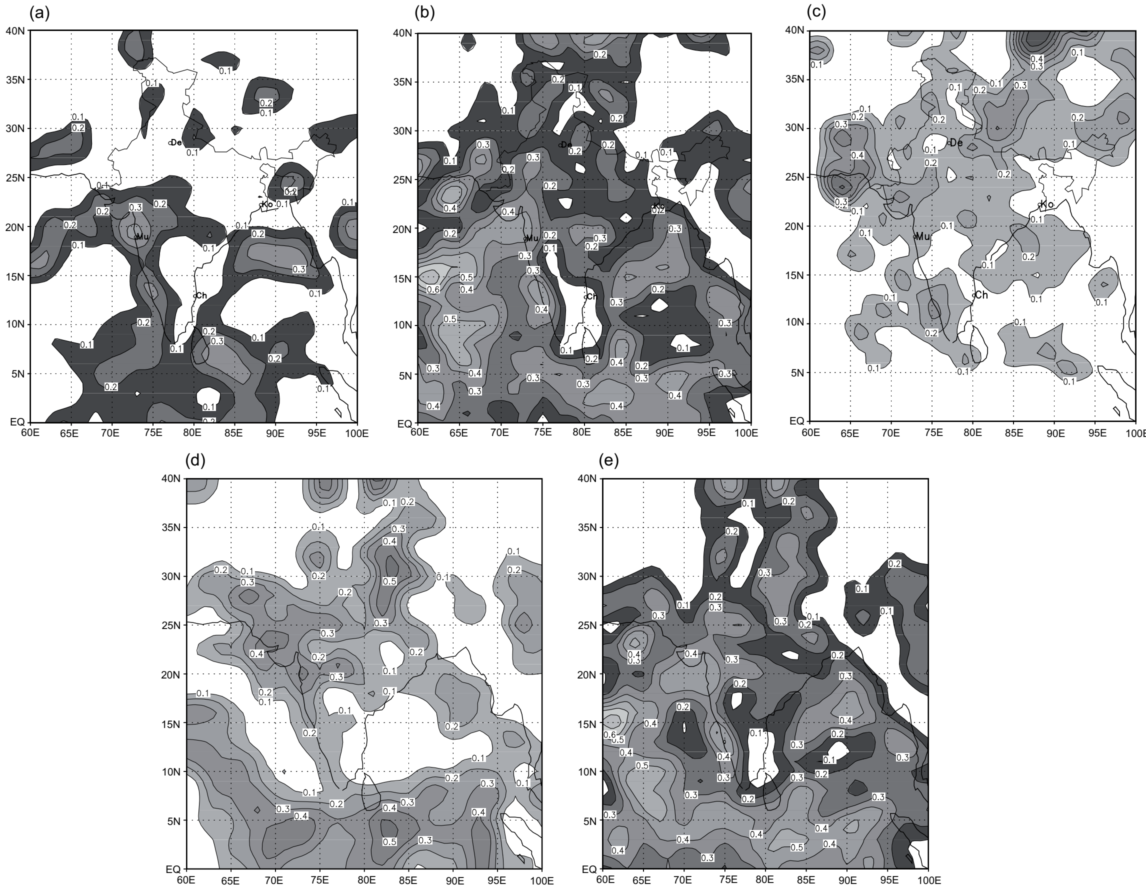

Fig. 10. (a-e) Same as Fig. 9 except for the day-2 forecasts.

4.5 Skill scores

The statistical parameter based on the frequency of occurrence of any event is more suitable for determining the rainfall prediction skill of a model. In Figure 11, an inter-comparison of the domain averaged (equator to 40° N and 60° to 100° E) values of threat score for different rain thresholds (0, 2.5, 5, 7.5, 10 and 15 mm) are shown for day-1 and day-2 forecasts of LAM, MM5, MM5 NCMRWF, T-80 and the ensemble. The threat score is the ratio of the number of successful predictions of an event to the total number of predictions made for occurrence or non-occurrence of the event. Higher value of threat score indicates better prediction with a theoretical limit of 1.0 for the perfect model. The comparison clearly shows that the ensemble has the highest skill at all the thresholds both for day-1 and day-2 forecasts. In the day-1 forecast, the score starts from 0.85 and then becomes 0.57 at the 2.5 rain threshold, 0.42 at 5 mm rain threshold, 0.23 at 10 mm rain threshold and becomes 0.7 at 15 mm rain threshold. In the day-2 forecast, the score begins from 0.84 and then decreases to 0.51, 0.40, 0.21 and 0.1 at the threshold values of 00, 2.5, 5, 10 and 15 mm, respectively.

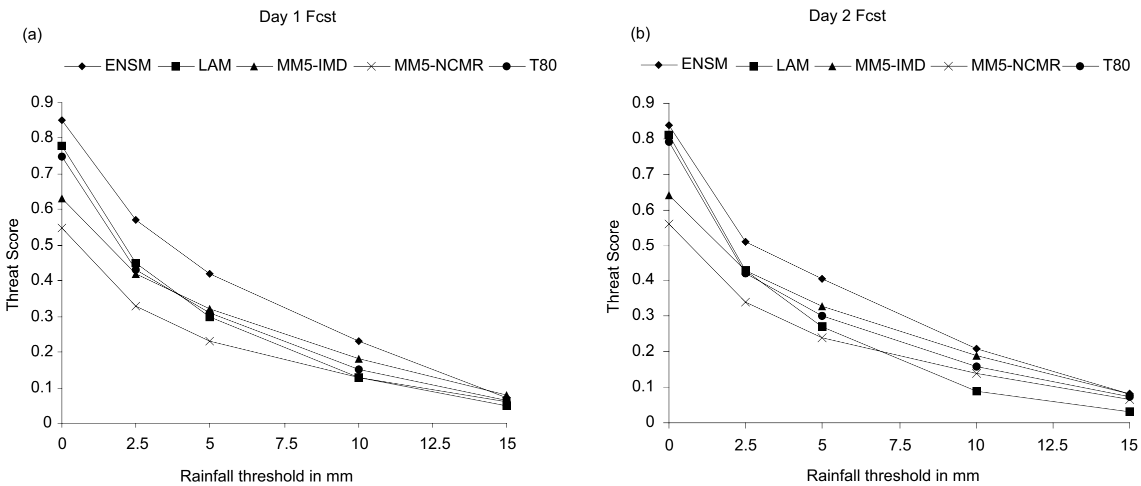

Fig. 11. A comparison of domain averaged threat score for day-1 and day-2 forecasts based on LAM, MM5, MM5 (NCMRWF), T-80 and ensemble.

5. Conclusions

In this paper a method is proposed for the ensemble forecasting of rainfall over Indian monsoon region in the short-range time scale on the basis of daily forecast fields from four operational NWP models available on real time basis at IMD, New Delhi. The comparison of rainfall prediction skill of the ensemble forecast against the constituent models, namely LAM and MM5 of IMD and MM5 and T-80 of NCMRWF, and observations reveals that the ensemble 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 strength of the MM5 model is that it shows better skill along Western Ghats and over the north Bay of Bengal. But fails to capture rainfall over the domain of monsoon trough. On the contrary, LAM shows better skill over the domain of monsoon trough and performs poor along Western Ghats and over the north Bay of Bengal. The ensemble forecast could take the advantage of each of the constituent models. The pattern of the systematic errors remains broadly the same for the day-1 and day-2 forecasts. The skill of day-1 forecast is found slightly better compared to day-2 forecast. The spatial pattern of RMSE and CC clearly indicated that the ensemble forecast is superior to the forecast of constituent models. The RMSE is found to be lowest in the ensemble forecasts both in magnitude and in the area coverage. This indicates that fluctuations of day to day errors are relatively less in case of ensemble forecast. The comparison of domain averaged threat scores (Fig. 11) for different rain thresholds clearly demonstrated that the ensemble forecasts have the highest skill. The method is found to be very useful in forecasting of spatial distribution of rainfall over Indian monsoon region. The method with one season data has shown sufficiently promising results for operational applications.

As we do more and more days of forecasts, we can pass the data sets of the forecast periods to the training period, thus increasing the length of the training period. It remains to be seen what further improvement in the ensemble prediction skill is possible from the use of monthly weights for constituent models on the basis of the longer training period. The future extension of this work would also require to increase the forecast period up to 5 days replacing regional models by global models as the members of the ensemble.

Acknowledgements

Authors are grateful to the Director General of Meteorology, India Meteorological Department for the encouragements and all supports to complete this research work. Authors acknowledge the use of MM5 and T-80 model rainfall outputs of NCMRWF, Noida, in this work.

References

Basu B. K., 2003. Verification of model predicted precipitation over India during monsoon 1997. *Mausam* **54**, 359-376.

Basu B. K., 2005. Some characteristics of model produced precipitation during summer monsoon over India. *J. Appl. Met*. **44**, 324-339.

Krishnamurti T. N., C. M. Kishtawal, T. Larow, D. Bachiochi, Z. Zhang, E. C. Willford, S. Gadgil and S. Surendran, 1999. Improved weather and seasonal climate forecasts from multimodel super ensemble. *Science* **285**, 1548-1550.

Krishnamurti T. N., C. M. Kishtawal, D. W. Shin and E. C. Willford, 2000. Improving tropical precipitation forecasts from a multi analysis super ensemble. *J. Clim*. **13**, 4217-4227.

Lorenz E. N., 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* **20**, 130-141.

Lorenz E. N., 1965. A study of the predictability of a 28 variable atmospheric model. *Tellus* **17**, 321-333.

Molteni F., R. Buizza, T. N. Pamer and T. Petroliagis, 1996. The ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.* **122**, 73-119.

Roy Bhowmik S. K. and K. Das Ananda, 2007. Rainfall analysis for Indian monsoon region using the merged rain gauge observations and satellite estimates: Evaluation of monsoon rainfall features. *J. Earth Sys. Sci*, **116**, 187-198.

Roy Bhowmik S. K., D. Joardar and H. R. Hatwar, 2007. An evaluation of rainfall prediction skill of IMD operational NWP system. *Meteorol. Atmos. Phys*. **95**, 205-221.

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

WMO, 1992. Manual on the global data processing system. Attachment No. 485, II-7.

Zhang Z. and T. N. Krishnamurti, 1997. Ensemble forecasting of hurricane tracks. *Bull. Amer. Meteor. Soc*. **78**, 2785-9275.