Prediction Skill of Global Numerical Weather Prediction Model Deterministic Forecast over India during North East Monsoon

V.R. Durai[1], *Rashmi Bhardwaj[2] and Y.V. Rama Rao
[1]India Meteorological Department, New Delhi
[2]Guru Gobind Singh Indraprastha University
Dwarka, New Delhi
E-mail: durai.imd@gmail.com

ABSTRACT

This study provides a concise documentation of the current level of skill of the deterministic NWP model during North East monsoon 2010; making detailed inter-comparison with daily rainfall analysis from the use of rain gauge observations and satellite (TRMM) derived quantitative precipitation estimates (QPE) obtained from NASA web site. Model performance is evaluated for day-1 to day-5 forecasts of 24-h accumulated precipitation in terms of several accuracy and skill measures. Forecast quality and potential value are found to depend strongly on the verification dataset, geographic region, and precipitation threshold. Precipitation forecasts of the model, when accumulated over the whole season (October to December), reproduce the observed pattern. However, the model predicted rainfall is comparatively higher than the observed rainfall over most parts of the NE Monsoon regions during the season. The model showed considerable skill in predicting the daily and seasonal mean rainfall over southern peninsular India and also over five broad regions (Tamilnadu, Kerala, South Interior Karnataka, Royalseema and Coastal Andra Pradesh) of Indian North East monsoon areas. Various skill score and categorical statistics for the deterministic global model of IMD GFS rainfall forecast for NE Monsoon 2010 are prepared. The model bias for rainfall prediction changes from overestimation to underestimation at the threshold of 20 mm/day except for day-5 forecast. Model skill falls dramatically for occurrence rainfall thresholds greater than 15 mm/day. This implies that the model is much better at predicting the occurrence of rainfall than they are at predicting the magnitude and location of the peak values.

Keywords: *Global forecast system, quantitative precipitation estimation, Indian summer monsoon, rainfall rediction skill.*

1. Introduction

The summer monsoon season (June through September) is meteorologically most important for India because more than 80% of the land area gets about 90% of its annual precipitation during this period. In India, where rainfall is seasonal and the agriculture is mostly rainfall dependent for planning its day-to-day operations. The main crops i.e. rice, maize are cultivated during this seasons. However, during the autumn, the zone of maximum rainfall migrates to southernmost India, Sri Lanka and the neighboring seas. Hence, the main season of rainfall over the southern parts of India is the northeast (NE) monsoon during the October through to December period. This season is also termed the retreating (southwest) monsoon season or the post-monsoon season. It is a major period of rainfall in south peninsular India, in particular the south-eastern parts. The rainfall in this season determines the agricultural production in this area. The rain shadow due to the Western Ghats spreads over a large area of the Indian peninsular. The states in the northwest and southeast get the least amount of rainfall. This entire region is sheltered by the Western Ghats (orographic barrier along the west coast of India) from the full blast of the rain-bearing winds of the southwest monsoon. Therefore, there is not much rain over this region during the southwest monsoon. With the retreat of the southwest monsoon and the reversal of the pressure and wind distribution, which occurs at the beginning of October, a trough of low pressure becomes established in the south Bay of Bengal. This type of situation forces equatorial maritime air in sufficient depth towards south India, causing widespread rainfall.

Most of the NWP performance studies over India were carried during south west monsoon. To date; there have been no systematic studies on deterministic model performance skill over India during NE monsoon season in the short to medium range time scale. One of the most critical issues in modeling the global atmosphere by Atmospheric general circulation models is the simulation and initialization of precipitation processes. The main aim of this study is to evaluate the rainfall prediction skill of deterministic NWP model over India in short to medium range time scale during NE monsoon 2010. Performance statistics for the precipitation forecast of many NWP models have been documented by various authors Mc Bride and Ebert, (2000), Doswell et al (1990), ete.

2. NWP Model

The Global forecasting system (GFS) run at IMD is a primitive equation spectral global deterministic model with state of art dynamics and physics (Kanamitsu, et al, 1991; Moorthi et al, 2001). This GFS model is conforming to a dynamical framework known as the Earth System Modeling Framework (ESMF) and its code was restructured to have many options for updated dynamics and physics. Details about the NCEP GFS are available at http://www.emc.ncep.noaa.gov/GFS/doc.php. The details about model physics and dynamics are discussed in the recent study by Durai and Roy Bhowmik, 2014. The model physics changes from its previous version to current version at T574 are mainly in radiation, gravity wave drag, planetary boundary layer processes, shallow and deep convection schemes and an introduction of tracer transport scheme in the vertical (Saha et al, 2010).

The assimilation system (for GFS) is a global 3-dimensional variational technique, based on NCEP Grid Point Statistical Interpolation (GSI *3.0.0*; Kleist *et al 2009*) scheme, which is the next generation of Spectral Statistical Interpolation (SSI; David et al 1992). The T574 Global Data Assimilation System (GDAS) use variational quality control, flow dependent re-weighting of background error statistics, use of new version of Community Radiative Transfer Model (CRTM 2.0.2), improved tropical cyclone relocation algorithm. In the operational mode at IMD, the GDAS cycle runs 4 times a day (00 UTC, 06 UTC, 12 UTC and 18 UTC) and GFS model runs 2 times a day (00 and 12 UTC). The analysis and forecast for seven days are performed using the High Power Computing System (HPCS) installed in IMD Delhi. Details of data presently being processed for GFS at IMD are available at http://www.imd.gov.in/section/nhac/dynamic/data_coverage.pdf

3. Methodology

In this study rainfall verifications were carried out for IMD GFS model runs at 00 UTC against daily rainfall analysis over Indian NE monsoon regions i.e. south peninsular India for the NE Monsoon 2010 at the resolution of 50 km based on the merged rainfall data combining gridded rain gauge observations prepared by IMD Pune for the land areas and TRMM 3B42RT data for the Sea areas (Mitra et al 2014). The quality of satellite estimated rainfall over Indian region is found to be good and useful for validation of NWP model forecast (Durai et al 2010).

Direct comparison can be made of mean values of *seasonal rain rate* and daily *rain rate and seasonal mean error.* In addition to simple point by point comparisons like *root-mean-square error, mean absolute error,* and *coefficient of linear correlation* between forecast and analysis can be computed.

In addition to these simple measures, a number of categorical statistics are applied. The term *categorical* refers to the yes/no nature of the forecast verification at each grid point. Some threshold (i.e., 0.1, 1, 2, 5, 10, 15, 35, 65 mm/day) is considered to define the transition between rain versus no-rain event. Then at each grid point (at the resolution of 50 km), each verification time is scored as falling under one of the four categories of *correct no-rain forecasts (Z), false alarms (F), misses (M),* or hits *(H)* **.** A number of categorical statistical skill measures are used, computed from the elements of rain/no-rain contingency **Table.1**

TABLE 1

Rain contingency table applied at each grid point

Observed	Predicted	
	Rain	No Rain
No Rain	H	F
	M	Z

They include bias score (bias):

$$BS = \frac{F+H}{M+H} \quad (1)$$

The bias score is equal to the number of rain forecasts divided by the total number of observations of rain. Thus the bias score is a measure of the relative frequency of rain forecasts compared with observations.

Threat score (critical success index):

$$TS = \frac{H}{H+M+F} \quad (2)$$

The threat score (TS) measures the fraction of observed and/or forecast events that were correctly predicted.

In additional to above three score, probability of detection (POD and false alarm ratio (*FAR*) could be generated easily by defining;

Probability of detection (POD)

$$POD = \frac{H}{H+M} \quad (3)$$

The probability of detection (POD) is equal to the number of hits divided by the total number of rain observations; thus it gives a simple measure of the proportion of rain events successfully forecast by the model.

4. Results and Discussion

4.1 Characteristics of observed and model predicted precipitation

In India, where rainfall is seasonal and the agriculture is mostly rainfall dependent for planning its day-to-day operations. The main crops i.e. rice is cultivated over most parts of south peninsular India, particularly in tamilnadu during October to December i.e. NE monsoon season. We begin with a description of observed cumulative rainfall field for the Indian North East monsoon season (1 October – 31 December 2010). The observed rainfall (Fig.1; OBS) distribution during NE monsoon 2010 shows a north south oriented belt of heavy rainfall along the East coast with a peak of more than 100 cm. Another heavy rainfall belt is observed along the kerala coast. For a numerical model of the atmosphere to be successful in predicting monsoon precipitation over India, the first step is to reproduce the observed characteristic patterns in the seasonal accumulated values.

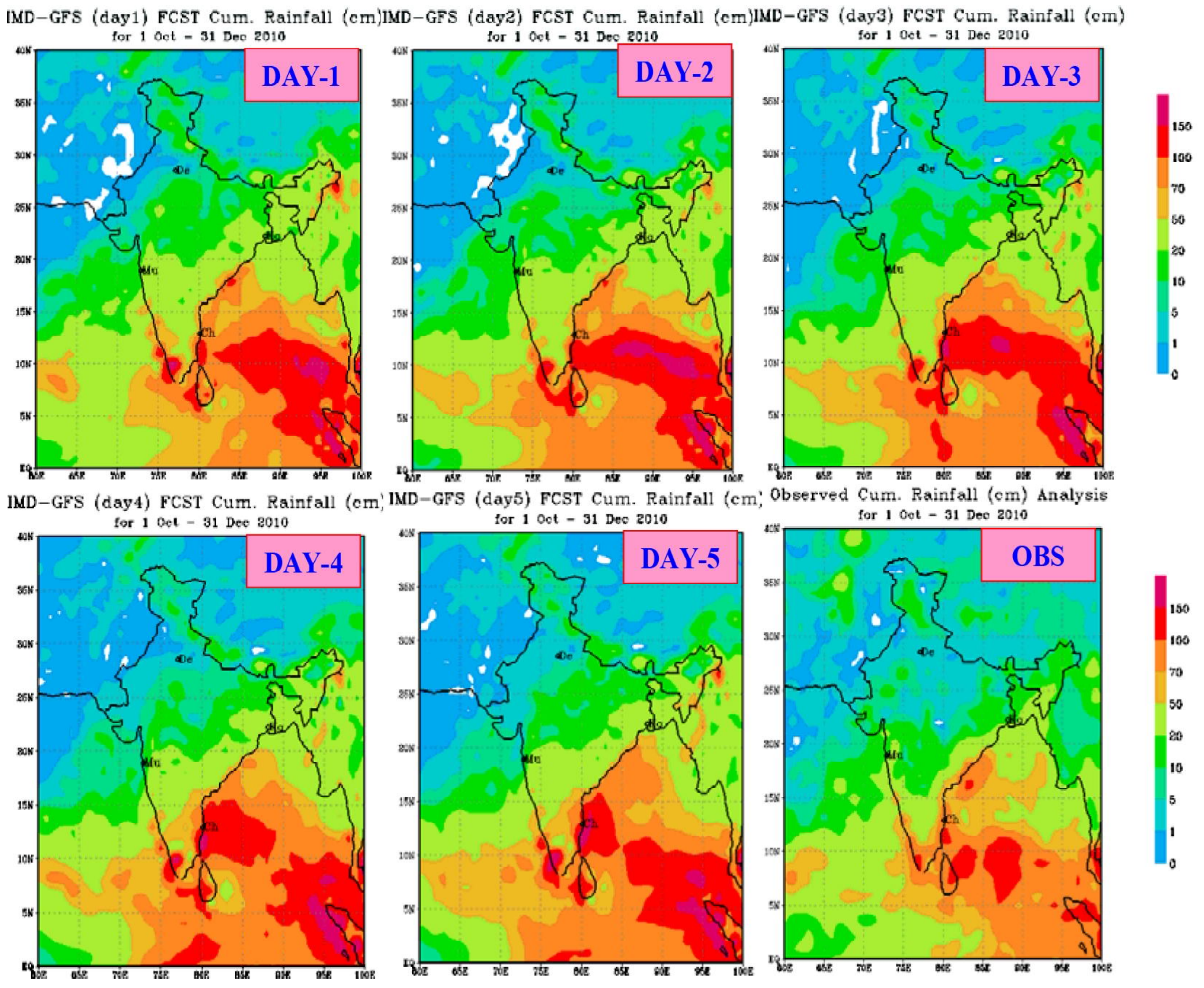


Fig.1 Spatial distribution of model predicted cumulative rainfall based on GFS day-1 -to day-5 forecasts and observed (OBS) cumulative rainfall for the period from 1 October to 31 December 2010.

The forecast fields (day-1 to day-5) of accumulated rainfall for the NE monsoon season 2010 based on the model is shown in Fig.1. The forecasts by this model, in general, could reproduce the heavy rainfall along the east coast of Tamilnadu and along the west coast of Kerala. The other large seasonal total precipitation due to dynamical forcing produced by the generation of cyclonic circulations over the South eastern Bay of Bengal region is also seen in the model prediction. The regions of less precipitation to the north of Southern Peninsular India are also noticed in model forecasts. However, some spatial variations in magnitude are noticed. The spatial distribution pattern of model predicted rainfall is closer to the corresponding observed field

The observed and model predicted NE monsoon seasonal (1October -31 December) mean rain rate (mm/day) over the entire Southern Peninsular India and also over sub-divisions (Fig.2) of Coastal Andra Pradesh (CAP), Tamilnadu (TN), Kerala (KL), South Interior Karnataka (SIK), Royalseema (RSM) based on the model day-1 to day-5 forecasts for the NE monsoon 2010 (1 Oct - 31 Dec) is plotted in Fig 3. It shows that the model day-1 to day-3 forecasts of seasonal mean rainfall (mm/day) over Indian NE monsoon regions are close to the observed rainfall, while the day-5 rainfall over estimate the seasonal mean rainfall during the season.

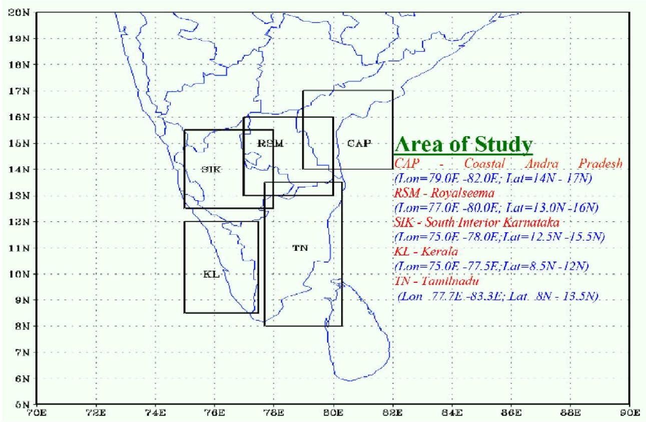


Fig.2 Five meteorological sub-divisions over southern peninsular India i.e. Tamilnadu (TN), Kerala (KL), South Interior Karnataka (SIK), Royalseema (RSM) and Coastal Andra Pradesh (CAP).

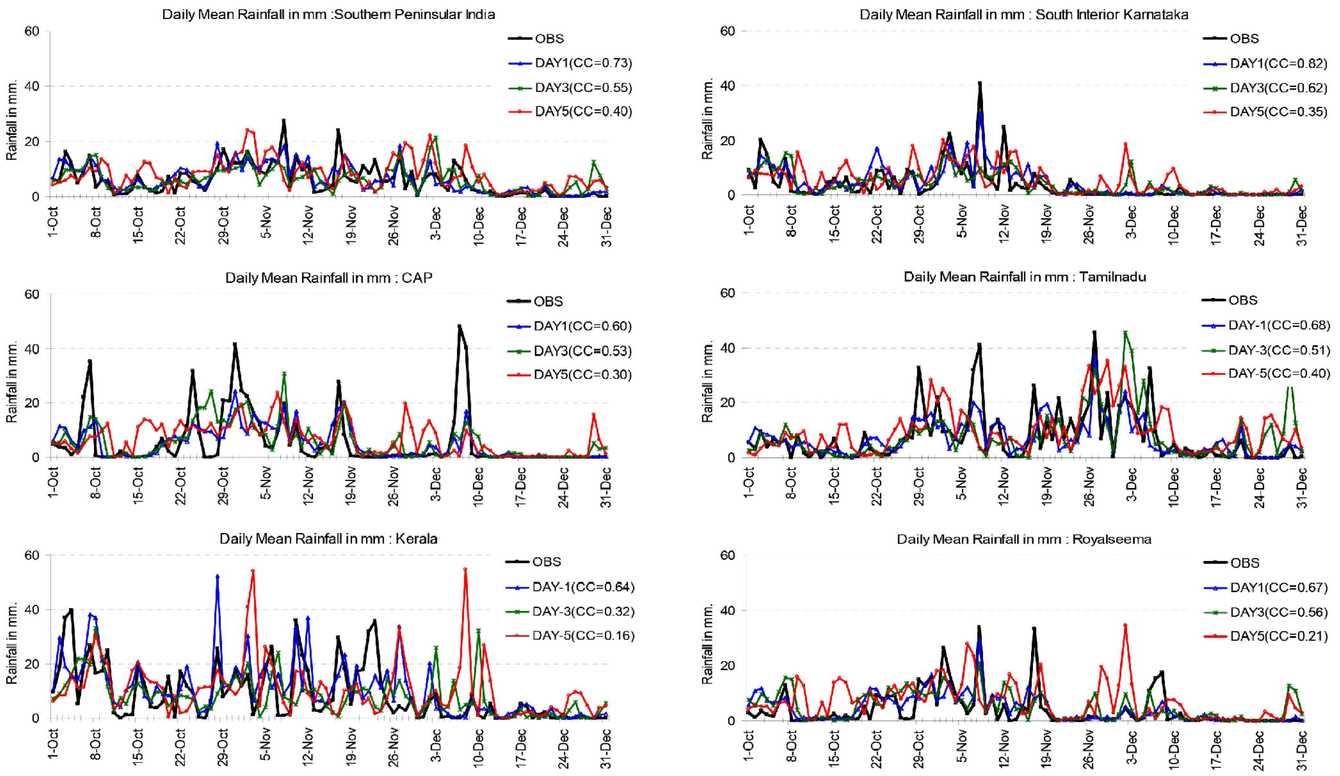


Fig.3 Time series of daily domain mean observed versus corresponding day-1 to day-5 rainfall (mm) forecasts by GFS model over Southern Peninsular India and also over five broad homogeneous regions Coastal Andra Pradesh (CAP), Tamilnadu (TN), Kerala (KL), South Interior Karnataka (SIK), Royalseema (RSM) based on the model day-1 to day-5 forecasts for the NE monsoon 2010 (1 October -31 December).

The temporal correlation coefficient (CC) between domain mean observed rainfall and forecast rainfall is more than 0.5 up to day-3 forecast and then it is high in the day-1 forecast in all the regions over south peninsular India. It also shows that the day-1, day-3 and day-5 forecast are in phase with each other, indicating a consistency in the model forecast.

The correlation coefficient between trends in the forecast and observation is a measure of the phase relationship between them. The *CC* between daily domain mean observed and forecasted precipitation of day-1 to day-5 for southern Peninsular India and for five broad regions is shown in Fig 4. From Fig 4, it is seen that the all India domain mean *CC* for all day-1 to day-5 forecasts has values greater than 0.60, with a higher value 0.90 for day-1, followed by 0.85 for day-2 and 0.75 for day-3. The inter-comparison of domain mean CC of day1-day5 rainfall forecasts by the model over five regions of India is more consistent with the daily rainfall time series as shown in Fig 4. The skill of the model is dependent on both the time scale 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 it raining over a large area than a small one, and when the rainfall is widespread rather than localized.

4.2 Verification of precipitation forecast

The precipitation forecast skills are highly depended on the resolutions of verified grids/boxes (spatial) and time period (temporal). There is higher skill if the verified grids/boxes are very large or the time period is very long. The average of the forecast errors over a long period of time is a measure of the systematic part of the forecast error, while root-mean-square error (rmse) is a measure of the random component of the forecast error. The correlation coefficient between trends in the forecast and observation is a measure of the phase relationship between them.

Spatial distribution of root mean errors (rmse) of rainfall for day-1 to day-5 forecasts by the model for NE monsoon 2010 (Fig 5) shows that the magnitude of rmse is between 5 to 15 mm/day for all the day-1 to day-5 forecast over most parts of the NE monsoon regions except over South East Bay of Bengal, where it is in the order of 15-20 mm/day.

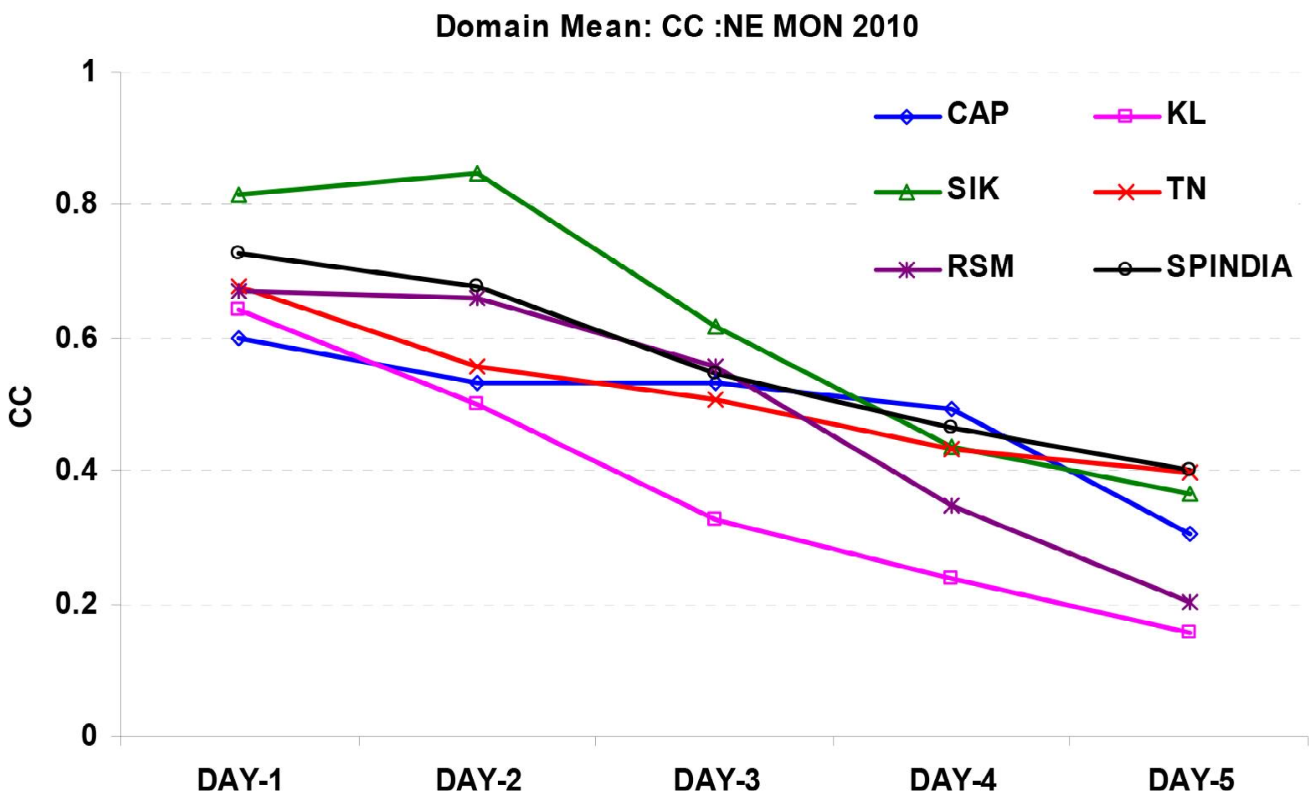


Fig.4 The correlation coefficient (cc) between daily domain mean observed and forecasted precipitation of day-1 to day-5 for southern Peninsular India and for five broad homogeneous regions.

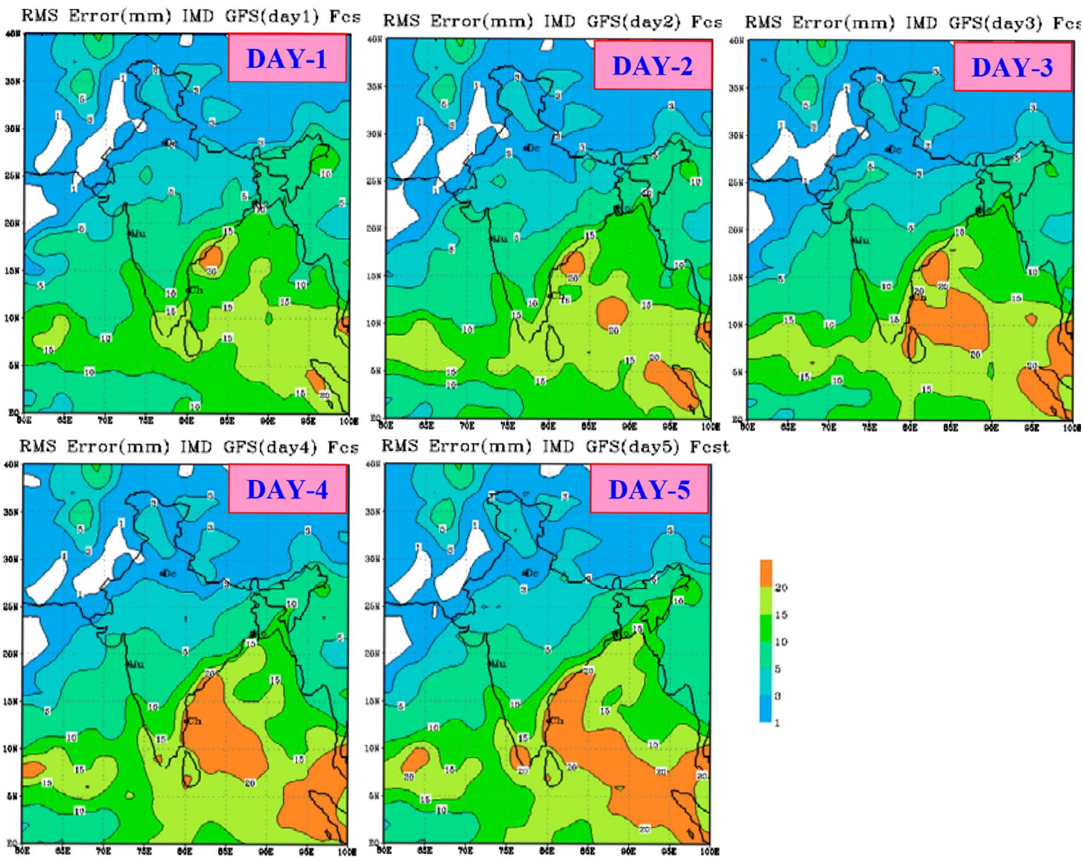
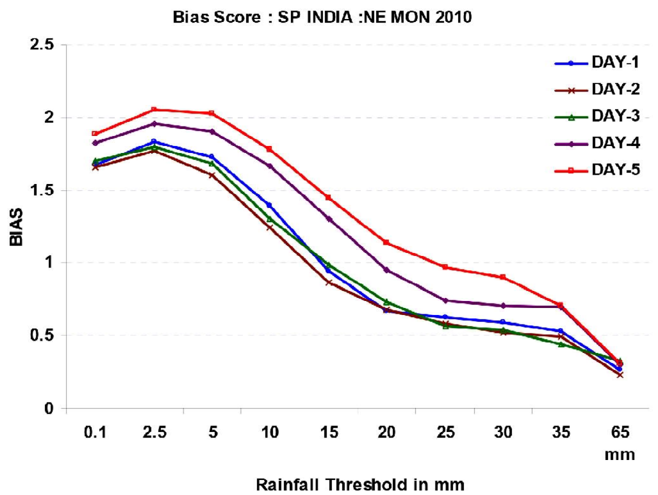


Fig.5 Spatial distribution of root mean error (forecast-observed) rainfall (mm/day) based on IMD GFS day-1 to day-5 forecasts for the period from 1 October to 31 December 2010.



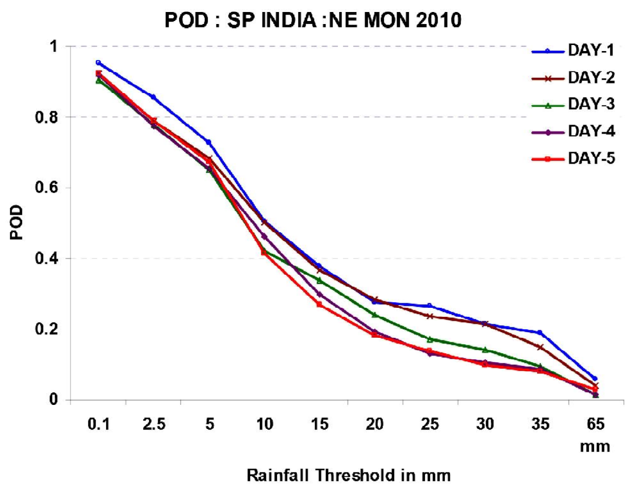


Fig.6 The Bias Score between daily domain mean observed and forecasted precipitation of day-1 to day-5 for southern Peninsular India during 1October -31 December 2010.

Fig. 7 The POD Score between daily domain mean observed and forecasted precipitation of day-1 to day-5 for southern Peninsular India during 1October -31 December 2010.

The aspect of model behaviour is further explored in Fig. 7, the bias, Fig 8 probability of detection (POD) and Fig 9 false-alarm rate (FAR) for classes with class marks of 0.1, 1, 2, 5, 10, 15, 35, ...65 mm/day are presented.

The bias of a model forecast is the ratio of the predicted number of occurrences of an event to the number of occurrences of the same event actually realized in nature. It measures the ratio of the frequency of forecast events to the frequency of observed events. Indicates whether the forecast system has a tendency to under forecast (*BIAS*<1) or over forecast (*BIAS*>1) events. It does not measure how well the forecast corresponds to the observations, only measures relative frequencies.The day-1 bias (Fig 6) of the model continuously over predicts (bias >1) rainfall event in all the threshold ranges up to 65 mm, while, the day-2 to day-5 bias over predict (bias <1) rainfall event only up to 20 mm and above 20 mm the bias score down to below 1.0. And also, the values of day-1 bias are high as compared to day-2 to day-5 bias in all the threshold ranges. In general, the model rainfall forecasts (except day-1) over predicts events of a lower magnitude, but the crossover to under prediction occurs at a higher value close to 25 mm.

The probability of detection (POD) is equal to the number of hits divided by the total number of rain observations; thus it gives a simple measure of the proportion of rain events successfully forecast by the mode. From Fig 7, it is seen that the probability of detection(POD) is more than 50% 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. From Fig 7 it is seen that skill is a strong function of threshold as well as forecast lead time (day-1 to day-5) , with the probability of detection (POD) (Fig 7) decreasing from about 80-90% for rain/no rain (> 0.1 mm/day) to about 20% or 30% for rain amounts above 30 mm/day .Consistent with this the false alarm ratio increases with threshold, from about 20 or 30% at low threshold to 70%–80% at high thresholds.

Threat score (TS), also known as the critical success index (CSI, e.g., Schaefer, 1990); or equitable threat score (ETS) which is a modification of the threat score to account for the correct forecasts due to chance (Gilbert, 1884), is for verification of the skill in precipitation forecasting. The threat score (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. It can be thought of as the accuracy when correct negatives have been removed from consideration, that is, *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 threat score (Fig 8) starts close to 0.65 for rainfall threshold of 0.1 mm/day and then decreases to 0.3 near the 10 mm mark. Interestingly, the day-2 to day-5 threat score remains relatively constant as a function of threshold for low and moderate threshold values and the day-1 score is slightly higher in all the threshold ranges.

The observed and model predicted seasonal (1October-31December) mean rain rate (mm/day) over Southern Peninsular (SP) India and five regions i.e. CAP, KL, RSM,SIK, TN for the NE monsoon 2010 is shown in Fig.9. It is noticed from figure. 9 that model overestimate the seasonal mean rainfall over all the regions of NE monsoon areas in all day-1 to day-5 forecast. An inter-comparison of time series of daily country mean weekly (seven days) cumulative observed and corresponding (seven days) forecasted rainfall of GFS over different regions of NE monsoon regions of India are presented in Fig 10. The active and weak spell of rainfall activity is well reflected in the weekly rainfall of different of NE monsoon areas of India and also the weekly rainfall over the entire south peninsular India. The seven days cumulative forecast of both GFS are in phase (active / weak spell of rainfall) with the corresponding observed rainfall, indicating the predictability of rainfall in weekly time scale. The active rainfall activity (positive anomaly) and weak or break condition of rainfall activity (negative anomaly) is well captured by GFS in weekly scale.

5. Conclusions

From the result presented above, the following, is concluded. (1) Model day-1 to day-5 precipitation forecasts, when accumulated over the whole season (1 October to 31 December), reproduce the observed pattern (Fig.1) with two large areas (along the Tamilnadu Coast, Kerala coast) with a total precipitation in excess of 100 cm. The observed variability of daily mean precipitation over NE monsoon region is reproduced remarkably well by the day-1 to day-5 forecasts of the model. This

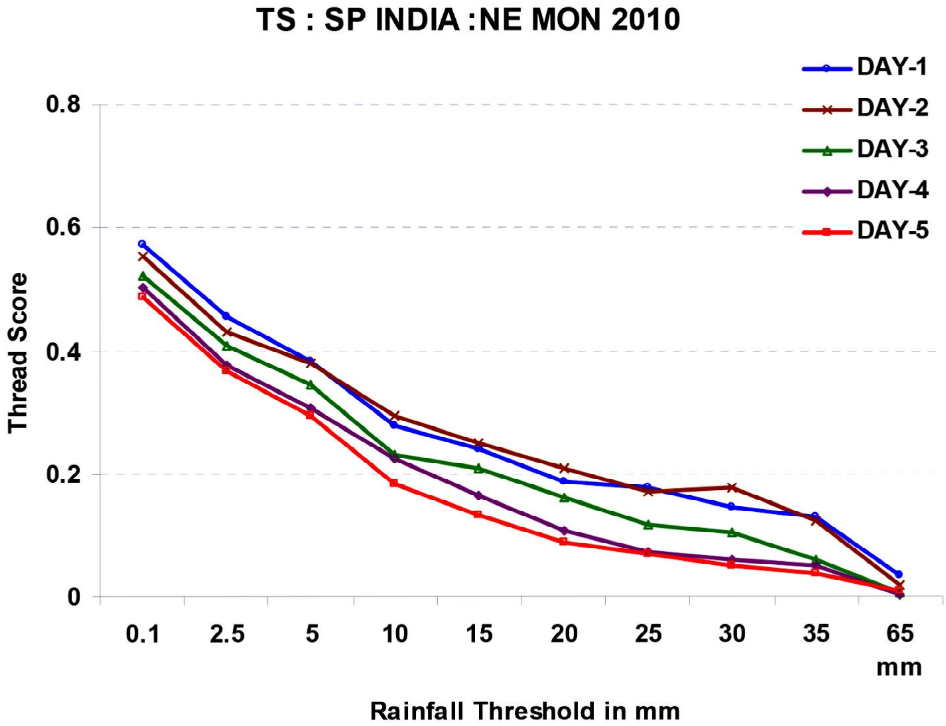


Fig.8 The Thread Score (TS) between daily domain mean observed and forecasted precipitation of day-1 to day-5 for southern Peninsular India during 1October -31 December 2010.

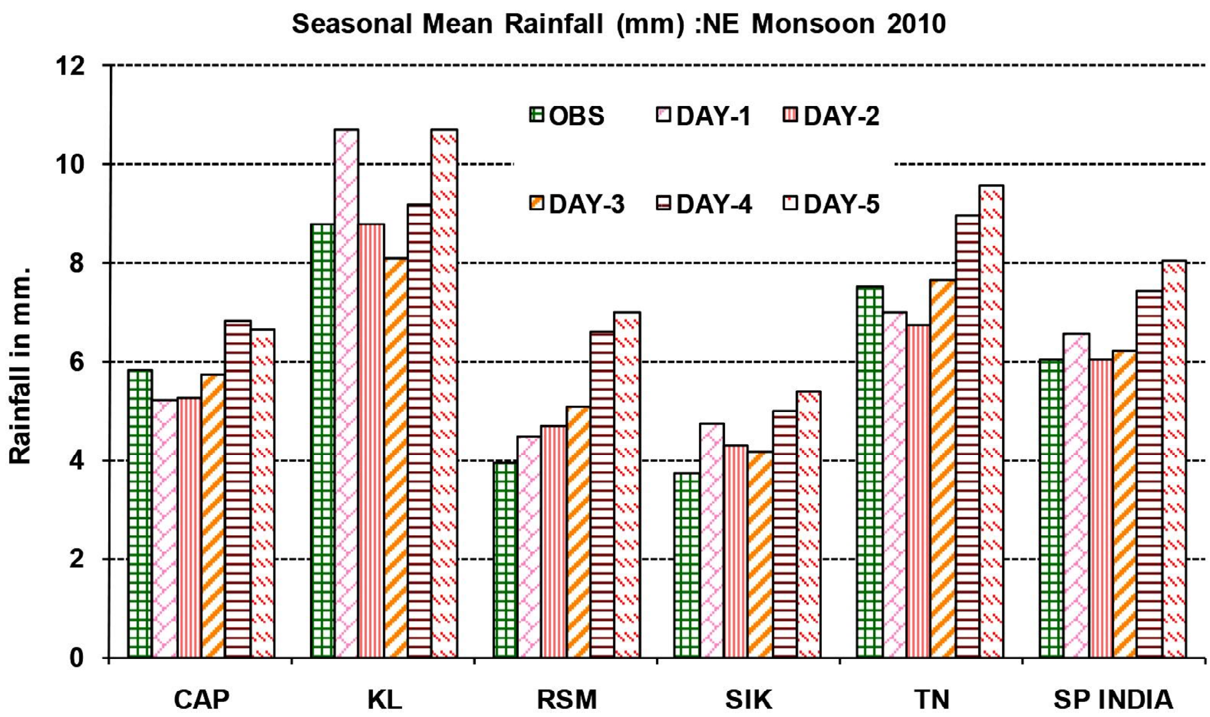


Fig.9 The observed and model predicted seasonal (1Oct -31 Dec) mean rain rate (mm/day) over Coastal Andra Pradesh (CAP), Kerala (KL), Royalseema (RSM), South Interior Karnataka (SIK), Tamilnadu (TN) and Southern Peninsular (SP) India based on the model day-1 to day-5 forecasts for the NE monsoon 2010.

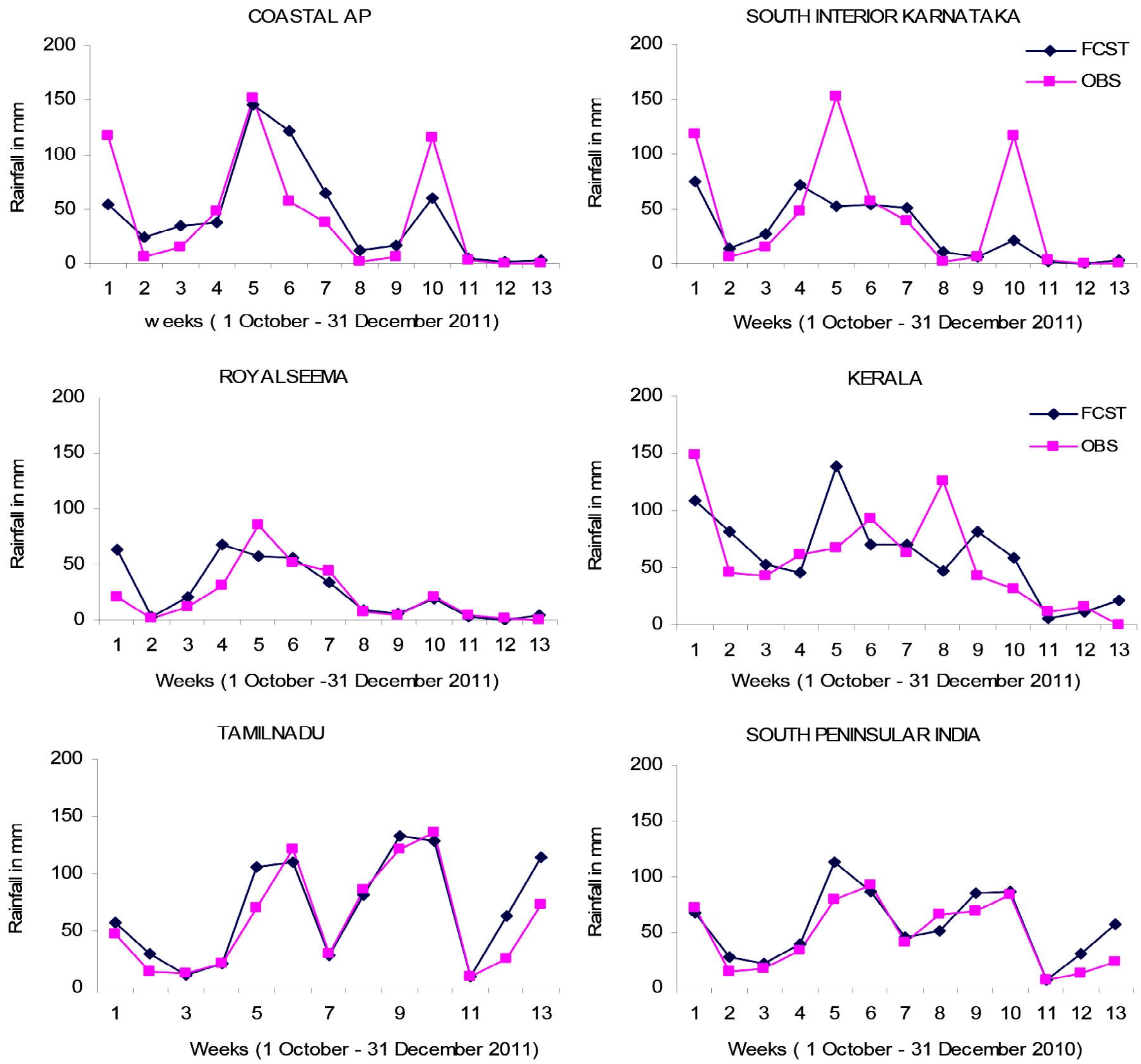


Fig.10 Weekly Domain Mean Observed and model predicted Rainfall (mm/week) over Southern Peninsular India and five broad homogeneous regions i.e. Tamilnadu (TN), Kerala (KL), South Interior Karnataka (SIK), Royalseema (RSM) and Coastal Andra Pradesh (CAP) based on the model day-1 to day-5 forecasts for the NE monsoon 2010.

implies that though the short- to medium-range forecasts have errors in the spatial distribution, the spatial average of daily precipitation over the NE Monsoon region and over five broad regions of Southern Peninsular India is in reasonable agreement with that observed. (2) The model has a tendency to over predict the frequency of occurrence of precipitation events in the light and moderate categories and to under predict events in higher categories. The model bias changes from overestimation to underestimation at the threshold of 25 mm/day except for day-1 forecast (Fig 6). (3) Model skill falls dramatically for occurrence thresholds greater than 20 mm/ day (Fig 8). This implies that the model is much better at predicting the occurrence of rain than they are at predicting the magnitude and location of the peak during the North East monsoon season over India.

Acknowledgements

Authors are grateful to the Director General of Meteorology and Deputy Director General of Meteorology (NWP Division), India Meteorological Department for their encouragements and supports to complete this research work. Also, Authors are thankful to Guru Gobind Singh Indraprastha University for providing research facilities. Acknowledgements are due to NCEP, USA for providing the source codes and NCMRWF for technical support for the implementation of the upgraded version GFS T574 at IMD.

References

Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting,* **5,** 576–585.

David F. Parrish and John C. Derber, (1992):The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. Monthly Weather Review, **120**, 1747-1763.

Durai V R, Roy Bhowmik S K and Mukhopadhaya B., 2010b : "Evaluation of Indian summer monsoon rainfall features using *TRMM and KALPANA-1* satellite derived precipitation and rain gauge observation" *Mausam , **61(3)**, 317-336*

Durai, V. R. and Bhowmik, S. R. (2014): Prediction of Indian summer monsoon in short to medium range time scale with high resolution global forecast system (GFS) T574 and T382. *Climate Dynamics*, *42*(5-6), 1527-1551.

Gilbert, G. F.: Finley's tornado predictions, *Amer. Meteorol. J., 1, 166–172, 1884.*

Kanamitsu, M., J.C. Alpert, K.A. Campana, P.M. Caplan, D.G. Deaven, M. Iredell, B. Katz, H.-L. Pan, J. Sela, and G.H. White, 1991: Recent changes implemented into the global forecast system at NMC. *Weather and Forecasting*, **6**, 425-435.

Kleist, Daryl T., David F. Parrish, John C. Derber, Russ Treadon, Wan-Shu Wu, Stephen Lord, (2009): Introduction of the GSI into the NCEP Global Data Assimilation System. Weather Forecasting, **24**, 1691–1705.

Mitra, A. K., Momin, I. M., Rajagopal, E. N., Basu, S., Rajeevan, M. N., & Krishnamurti, T. N. (2013). Gridded daily Indian monsoon rainfall for 14 seasons: Merged TRMM and IMD gauge analyzed values. *Journal of Earth System Science*, *122*(5), 1173-1182.

Moorthi, S., H. L. Pan and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. *NWS Technical Procedures Bulletin*, **484**, pp14.

McBride J L and Ebert E E 2000 Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia; *Wea. Forecasting* **15** 103–121

Saha, Suranjana, and Coauthors, (2010): The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., **91**, 1015–1057

Schaefer, J. T.: 1990 The critical success index as an indicator of warning skill, *Wea. Forecasting, **5**, 570–575 pp*