

Experimental real-time multi-model ensemble (MME) prediction of rainfall during monsoon 2008: Large-scale medium-range aspects

A K MITRA^{1,*}, G R IYENGAR¹, V R DURAI², J SANJAY³,
T N KRISHNAMURTI⁴, A MISHRA⁴, D R SIKKA⁵

¹*National Centre for Medium Range Weather Forecasting (NCMRWF),
Ministry of Earth Sciences, Noida 201 307, India.*

²*India Meteorological Department, New Delhi 110 003, India.*

³*Indian Institute of Tropical Meteorology, Pune 411 008, India.*

⁴*Department of Meteorology, Florida State University, Tallahassee, USA.*

⁵*40 Mausam Vihar, New Delhi 110 051, India.*

*e-mail: ashis.mitra@nic.in

Realistic simulation/prediction of the Asian summer monsoon rainfall on various space–time scales is a challenging scientific task. Compared to mid-latitudes, a proportional skill improvement in the prediction of monsoon rainfall in the medium range has not happened in recent years. Global models and data assimilation techniques are being improved for monsoon/tropics. However, multi-model ensemble (MME) forecasting is gaining popularity, as it has the potential to provide more information for practical forecasting in terms of making a consensus forecast and handling model uncertainties. As major centers are exchanging model output in near real-time, MME is a viable inexpensive way of enhancing the forecasting skill and information content. During monsoon 2008, on an experimental basis, an MME forecasting of large-scale monsoon precipitation in the medium range was carried out in real-time at National Centre for Medium Range Weather Forecasting (NCMRWF), India. Simple ensemble mean (EMN) giving equal weight to member models, bias-corrected ensemble mean (BCEMn) and MME forecast, where different weights are given to member models, are the products of the algorithm tested here. In general, the aforementioned products from the multi-model ensemble forecast system have a higher skill than individual model forecasts. The skill score for the Indian domain and other sub-regions indicates that the BCEMn produces the best result, compared to EMN and MME. Giving weights to different models to obtain an MME product helps to improve individual member models only marginally. It is noted that for higher rainfall values, the skill of the global model rainfall forecast decreases rapidly beyond day-3, and hence for day-4 and day-5, the MME products could not bring much improvement over member models. However, up to day-3, the MME products were always better than individual member models.

1. Introduction

The Asian monsoon is one of the major components of the earth climate system. Realistic modelling, simulation and prediction of monsoon

are challenging scientific tasks for the Earth System Science Community. For India, the monsoon rain is of enormous importance as it gives shape to its agriculture, economy and rhythms of life. The science pertaining to monsoon has progressed

Keywords. India monsoon; rainfall forecast; global-models; multi-model ensemble (MME).

significantly in the last two decades because of the increased wealth of new data from satellite observations, improved understanding of the processes and enhanced computing power. Numerical models and data assimilation algorithms have been further improved at all major international centers across the globe. The accuracy of weather forecasts has improved steadily in last three decades, and the systematic error with forecast length in the medium range has reduced. However, compared to mid-latitudes, the skill for tropics is still lower, and is particularly of concern for rainfall forecast for the Indian monsoon region (Webster *et al* 1998; Gadgil 2003; Krishnamurti *et al* 2006a, b; Woods 2006). The errors even in large-scale rainfall pattern forecast still remains. These errors are due to uncertainties in the assimilation–forecast system. These uncertainties could be due to the errors in the prescribed initial states, which may arise from observational instruments, satellite estimates and the data assimilation method. The forecast skill is also dependent on the synoptic situation, flow regime region and known (or unknown) teleconnections. The accuracy of a model also varies depending on the formulation, horizontal–vertical resolutions and the parameterization schemes representing the small-scale processes in the model. The process of improving forecast skill of an individual modelling system through research and development (R&D) in modelling and data assimilation is rather a slow process. For example, at European Centre for Medium-Range Weather Forecasting (ECMWF), the R&D of approximately ten years resulted in skill improvement of one day. Therefore, from a practical forecasting aspect, we have to learn to live with these uncertainties and the current available skill of the models. The point is how to make the best use of the forecasts from any (or many) centers, when each center is having data from many member models.

In the context of different model data from different centers and each center producing many ensemble members, the use of ensemble method in short, medium range and even for short-term climate prediction has become popular in recent years. Because of the aforementioned uncertainties in the modelling system, the forecast error increase as the forecast length increases, until it becomes no longer useful. If we have an ensemble of forecasts (many forecasts), we can have a greater reliability of the forecasts. Forecast can be represented in a probabilistic sense also. Clustering (or tubing) of several similar forecasts is also useful. From many ensemble forecasts, we can obtain clues for possible extreme/severe events. One single forecast (from one single model) may fail to catch the likelihood of an extreme event, but the ensemble might give

some extra clue on the extreme (or anomalous) episode. The current practices at major centers are:

- (i) to perturb the initial conditions (scientifically based) and make many member runs with these different initial conditions,
- (ii) make many runs by altering the model formulations, physical parameterizations, and
- (iii) MME members using (i) and (ii).

As many centers are exchanging or providing model data in real-time, in principle, the MME can be attempted with very little computing resources. This MME is also popularly known as the poor man’s ensemble, as it uses meager amount of resources. One assumption for MME is that the member models do provide some signal, and the noise (and error) is less compared to signal. While dealing with the monsoon, we are aware of the limitations of the models to capture the change in weather patterns, from dry to wet, or from wet to dry. In monsoon, the formation and movement of lows and depressions have some limitations. Correct simulation of the monsoon trough’s intensity is also an outstanding issue in modelling. If the gross errors are higher in intensity, track, position and timings of the aforementioned monsoon systems, then the MME may not be able to improve much above the crude member model’s skill. Even if the skill score improves in a statistical sense, the forecast may not be of much use, as the system in MME output may be away from the real system with a different intensity. World Meteorological Organization’s (WMO) ‘The Observing System Research and Predictability Experiment’ (THORPEX) is planning to provide eight global model’s data at a coarser resolution under ‘THORPEX Interactive Grand Global Ensemble’ (TIGGE) and each model is having around 15 ensemble members, making it in total around 120 members. The forecasting community has started experimenting with the multi-model data products, to examine the benefits in terms of skill enhancement and preparing probabilistic forecasts, particularly for extreme events. In India, the usefulness of the MME forecast was identified and it was planned during 2007 that on an experimental basis, the monsoon rainfall MME forecast will be attempted from monsoon 2008 season, using whatever model data available. A small team comprising of scientists from National Centre for Medium Range Weather Forecasting (NCMRWF), India Meteorological Department (IMD) and Indian Institute of Tropical Meteorology (IITM) was constituted by the Ministry of Earth Sciences (MoES) to start work on the experimental MME forecast.

At major weather/climate forecast centers, calibration of model outputs with respect to

observations have been done to obtain an improved skill. Different statistical post-processing techniques have been applied to model output parameters for the scale, region and phenomenon of interest. These post-processing methods have enabled the forecasters to obtain enhanced skill and value from models. MME is another post-processing (calibration) technique having the potential to enhance the skill of rainfall prediction. In this report, the performance of the experimental MME forecast of rainfall during monsoon 2008 is studied. As the first attempt, only the large-scale aspects of monsoon rain from member models and the multi-model products are documented. We are aware of the difficulties in producing rainfall forecasts for smaller regions by the state-of-the-art global models. Significant errors are obviously expected if one decides to come down to smaller meso-scales below the ‘large scale organized convective rainfall’ associated with monsoon. Therefore, as a first attempt, we study here the one by one degree latitude/longitude grid rainfall data from four member models and the associated multi-model products in the medium range. It is well known that the simple average made from many models (simple ensemble mean, giving equal weight to each member model) generally produces higher skill score. Our interest here is to document the improvement of the MME output (by giving different weights to different models) over the simple ensemble mean.

2. Data and MME methodology

In this study, we use the daily medium range (day-1 to day-5) rainfall prediction data from four state-of-the-art operational global models, namely Global Forecast System of National Centers for Environmental Prediction, USA (NCEP/GFS), United Kingdom Meteorological Office, UK (UKMO), Japan Meteorological Agency, Japan (JMA) and NCMRWF (India) for monsoon 2008 season (June–September). These model data were available up to day-7, but we restricted to test the MME only up to day-5. NCMRWF global assimilation–forecast system is an adapted version of the NCEP/GFS system implemented in the year 2007. Therefore, we expect, the basic characteristics of rainfall data from NCEP/GFS and NCMRWF to be broadly similar. The model data used was at $1^\circ \times 1^\circ$ 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. NCEP/GFS model data was from runs made at 35 km

horizontal grid and 64 vertical layers. UKMO model data was from runs made at 40 km horizontal grid and 50 vertical layers. JMA model data was from runs made at 20 km horizontal grid and 60 vertical layers. NCMRWF/GFS model data was from runs made at 50 km horizontal grid and 64 vertical layers. Therefore, what it simulated was of higher (finer) spatial scales than the results we have analyzed at $1^\circ \times 1^\circ$ resolution. As the purpose of this study is to note the skill enhancement coming from the multi-model algorithm, we thought it can be experimented at even this $1^\circ \times 1^\circ$ latitude/longitude resolution at which the data was provided from the respective centers, and depending upon the benefits, the algorithm can again be tested on much higher resolution data. Some recent studies on monsoon have shown the benefits of interpolating the coarser resolution data to a finer grid (sort of downscaling) and then do the MME on the finer grids. We also plan to test that step in our future work. Therefore, in this study, we use the original $1^\circ \times 1^\circ$ data we received from the operational centers *via* ftp. The corresponding daily observed gridded rainfall data at same $1^\circ \times 1^\circ$ resolution was prepared by merging rain-gauge values with the satellite estimates (Mitra *et al* 2003). The gridded rainfall analysis data used in model calibration (training) has to be of good quality. Otherwise, it might degrade the MME results.

Early works by Krishnamurti *et al* (1999) showed that it is possible to obtain skill improvements both in weather and climate scales by the use of the multi-model technique named as ‘Super Ensemble’ forecasting. Later it was extended for tropical precipitation by incorporating multi-analysis concept along with the use of multi-models (Krishnamurti *et al* 2000). Later Mishra and Krishnamurti (2007) applied the super-ensemble algorithm for Indian monsoon and showed the skill enhancement using seven global models data. 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 they showed the superiority of the multi-model system over the individual model data. Richardson (2001) used multi-model and multi-analysis data to produce both deterministic and probabilistic ensemble forecasts using four global models from UKMO, ‘German Met Service’ (DWD), Meteo France and NCEP and showed that simple ensemble mean and simple bias correction produce useful products. The probability of precipitation and rainfall distribution by using multi-model data from seven global models for Australia region was studied by Ebert (2001). Multi-model multi-analysis data was also tried by using ECMWF and UKMO ensemble outputs

for quasi-operational medium range forecasting (Mylne *et al* 2002) in both probabilistic and deterministic sense. They noted that the MME is more beneficial than a single model Ensemble Prediction System (EPS) and the skill of winter was higher compared to summer season. Operational consensus forecast by including several models is seen to outperform ‘direct model output’ (DMO) and ‘model output statistics’ MOS forecasts (Woodcock and Engel 2005). Multi-model product prepared by the use of NCEP/GFS and ECMWF data also shows improvements in week-2 forecasts (Whitaker *et al* 2006), improving over the MOS forecast of individual models. Most of these studies (algorithms) used simple ensemble mean or a mean of calibrated data from member models. However, by analyzing the past performance of model for a region, it might be interesting to examine if model dependent weights help further to compute the final multi-model forecast. In a very recent study (Johnson and Swinbank 2009), the ECMWF, UKMO and GFS global model data were used to prepare multi-model ensemble products in medium-range prediction. Here forecasts from bias-correction, model-dependent weights, and variance adjustments were studied. It was found that the multi-model ensemble gives an improvement in comparison to calibrated single model ensemble. They also noted that only small improvements from the use of model dependent weights and variance adjustments were achieved.

Many studies mentioned above have reported results of MME from wind, MSLP, Z500 and t2 m, etc. Reports on MME of rainfall particularly for tropical regions are relatively less. Hamill *et al* (2008) have shown that precipitation is a different type of weather parameter and has to be dealt differently in calibration and statistical post-processing. It is highly variable in space and time. It is non-continuous, has non-negative values and non-Gaussian in nature. They also discussed the training period (sample size) is important and particularly so for very heavy rainfall amounts. It was also shown that the MME combination gives up to a two-day increase in predictability at day-7 for mid-latitudes. MME for meso-scale weather systems and associated summer heavy precipitation were attempted for south-east US using high resolution six meso-scale model data (Cartwright and Krishnamurti 2007), which showed improvements with the super-ensemble algorithm. Roy Bhowmik and Durai (2008, 2010) used the linear regression method to find weights for member model and then make a multi-model ensemble product. However, the benefits of giving weights to member models over a simple ensemble mean were not documented there. In a recent study, the high resolution precipitation MME forecast for Indian monsoon was

reported by Krishnamurti *et al* (2009). By down-scaling the data to a higher resolution and use of super-ensemble technique, they have reported skill enhancement of precipitation forecasts in medium range for the Indian monsoon systems.

For the experimental rainfall MME during monsoon 2008, the scheme used here is very similar to that of Krishnamurti *et al* (2000) and Mishra and Krishnamurti (2007). It has a training phase and a forecast phase. During training phase, the member model forecasts were regressed (compared) against observation/analysis to obtain different weights. A multiple linear regression is performed in this phase to estimate the relative performance of member models. These weights are then passed on to the forecast phase to create the MME forecasts. The weights for the models are now based on their past performance. In this process, it combines a set of multi-model forecasts to construct a single consensus forecast. The weights of the MME vary geographically and with forecast lead time. The MME forecast is constructed as:

$$\text{MME} = \bar{O} + \sum_{i=1}^N a_i (F_i - \bar{F}_i),$$

where F_i is the i th model forecast, \bar{F}_i is the mean of the i th model forecast over the training period, \bar{O} is the observed mean of the training period, a_i is the regression coefficient obtained by a minimization procedure during the training period, and N is the number of forecast models involved. The coefficient a_i is derived from estimating the minimum of function to satisfy the mean square error criteria:

$$G = \sum_{i=1}^{N_{\text{tr}}} \left(\text{MME}'_t - O_t \right)^2.$$

Ensemble mean bias removed is defined as:

$$\text{BCE} = \bar{O} + \sum_{i=1}^N \frac{1}{N} (F_i - \bar{F}_i).$$

In addition to removing the bias, MME scales the individual model forecasts contributions according to their relative performance in the training period in a way that is equivalent to weighing them mathematically. In the initial days of June 2008 during monsoon 2008 season MME forecasting daily in real time, we used the daily day-1 to day-5 rainfall forecasts data from same models for monsoon 2007. For training, the observed monsoon 2007 merged satellite gauge data was used. A running 92 days training period was always maintained with march of days during the monsoon 2008 season. As the days progressed in 2008 season, we started including recent past model and observation data to the 92 days training period. For example, on 20

Table 1. Dates considered as training period during monsoon 2008 (real time) MME forecast.

1 June	JJA 2007 – 92 days
4 June	Start including data from June 2008, 1 June 2008 Remove days from August 2007, 31 August 2007
1 July	29–30 June 2007, JA07, 1–28 June 2008
4 July	JA07, 2–30 June 2008, 1 July 2008
31 July	JA07, 29–30 June 2008, 1–28 July 2008
4 August	3–31 July 2007, A07, July 2008, 1 August 2008
31 August	30–31 July 2007, A07, July 2008, 1–28 August 2008
1 September	13–30 June 2007, September 2007, June 2008, 15–28 August 2008
7 September	19–30 June 2007, September 2007, June 2008, 15 August–3 September 2008
30 September	September 2007, 12–30 June 2008, 15 August–26 September 2008

June 2008, the latest data from 1–17 June 2008, were included in training. Similarly on 15 August 2008, in addition to July, August data of 2007, we included July and August (up to 12th) data from 2008 itself in training. However, the training period was maintained to 92 days. Table 1 shows the training period dates for reference.

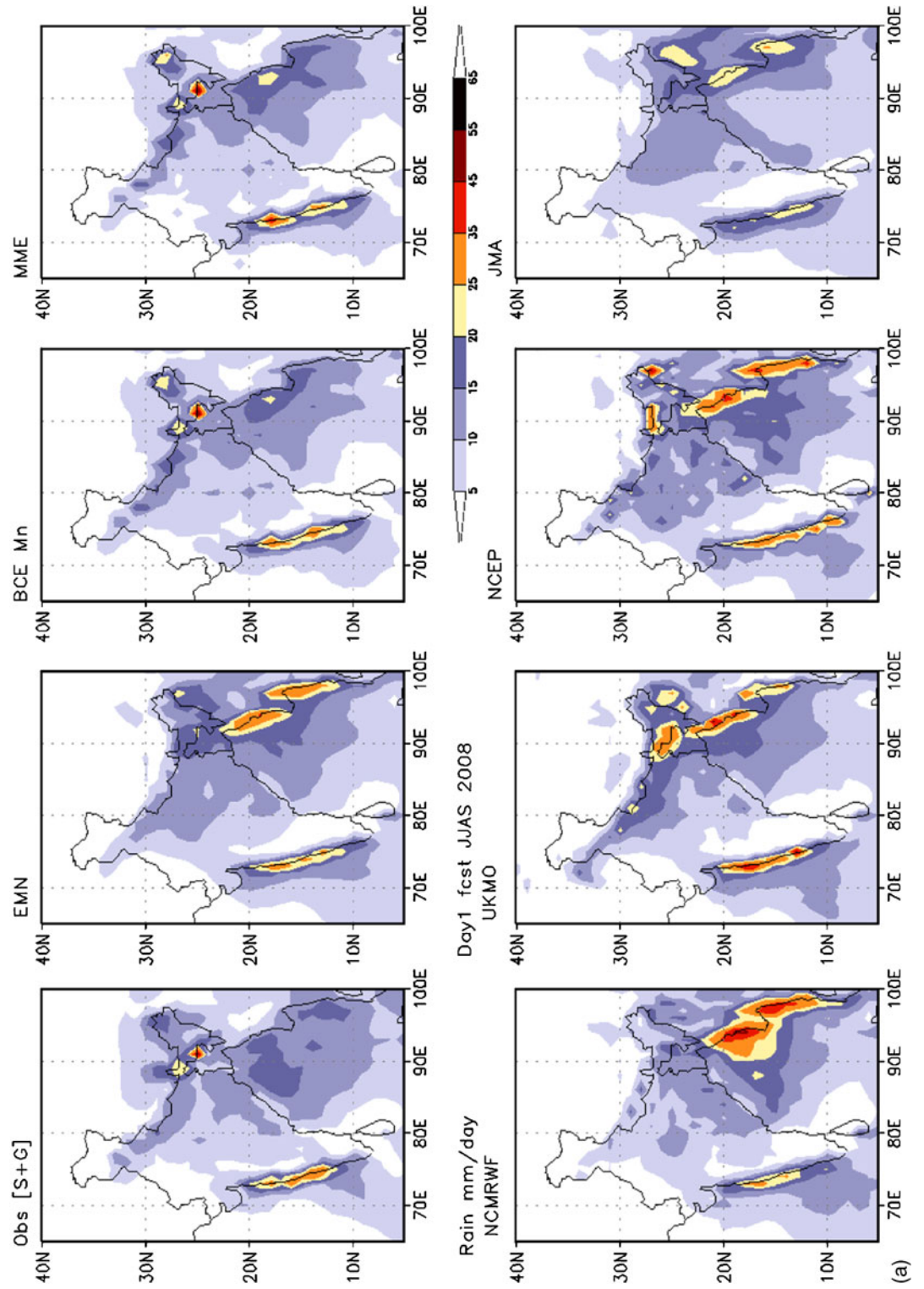
3. Results from monsoon 2008

As described in the previous section, during the monsoon 2008 (1 June– 30 September) the daily real time global model data from four centers were collected and used in this multi-model ensemble forecast preparation on an experimental basis. Forecasts of rainfall were made from day-1 through day-5. All member global model forecast and the multi-model forecast data for the forecast length of day-1 through day-5 were used to document the skill of rainfall forecast during the monsoon 2008 season for the Indian region. The performance of rainfall forecasts from member models and the multi-model algorithm are evaluated in terms of error statistics and threshold statistics. In this section, we describe the error and threshold statistics. In the discussions, the skill for day-1, day-3 and day-5 will be shown and discussed for brevity. Figure 1(a, b and c) shows the member-model and multi-model produced total rainfall during the 2008 season made up from day-1, day-3 and day-5 forecasts, respectively. In the figure, the observed rainfall is produced by merging rain-gauge and satellite estimates from METEOSAT IR data. At first glance, all the four member models seem to reproduce the observed monsoon rain closely. However, when the details are compared, we see each member model differs from the observations in different ways, and at different regions. For example, in the day-1 forecasts (figure 1a) the NCMRWF and NCEP model produces too much rain in the Arakan coast region to the east of Bay of Bengal. In NCEP model on the west coast of India

the north–south rainfall band extends too much to the south and in northern plains (monsoon trough region) the model produces more rain than observations. In the UKMO model, it is seen to rain more on the west coast, Himalayan foothills and northern Bangladesh. In contrast to all the models, the JMA model produces the least rainfall. The multi-model products in the upper row when compared with observations, look closer and more realistic. The BCEMn and MME look superior to simple ensemble mean (EMN). The EMN is superior to member models, but still maintains the signature of the strong biases of the member model. For example, the heavy rain along Arakan coast is seen in the EMN. The BCEMn and MME are able to reduce the biases and look closer to observations. The MME products look superior to member models and much closer to observations.

Similar to day-1 (figure 1a), the day-3 and day-5 seasonal total rainfall from observations, member models and multi-model ensembles are shown in figure 1(b and c). The systematic biases of member models continue to remain on day-3 forecasts with changing magnitudes. However, the pattern of biases continues to be of similar nature. NCMRWF and NCEP models almost maintain the similar biases with slight enhancement. In UKMO model, the west coast rain gradually becomes closer to observations. However, over the foothills and northern Bangladesh, the rainfall positive biases are seen to increase with time (from day-1 to day-3). JMA model shows gradual decreasing rainfall over all the regions around India. In contrast, in both day-3 and day-5, the multi-model products look closer to observations. Both the MME and BCEMn look better than simple EMN. Even in day-5 forecasts, the relatively lower biases in multi-model products are good signs of encouragement to continue R & D in MME forecasting algorithms.

To examine more clearly the rainfall biases of the four member models from the observations, the difference of observed amounts from the predicted amounts are computed for day-1, day-3 and day-5



(a)

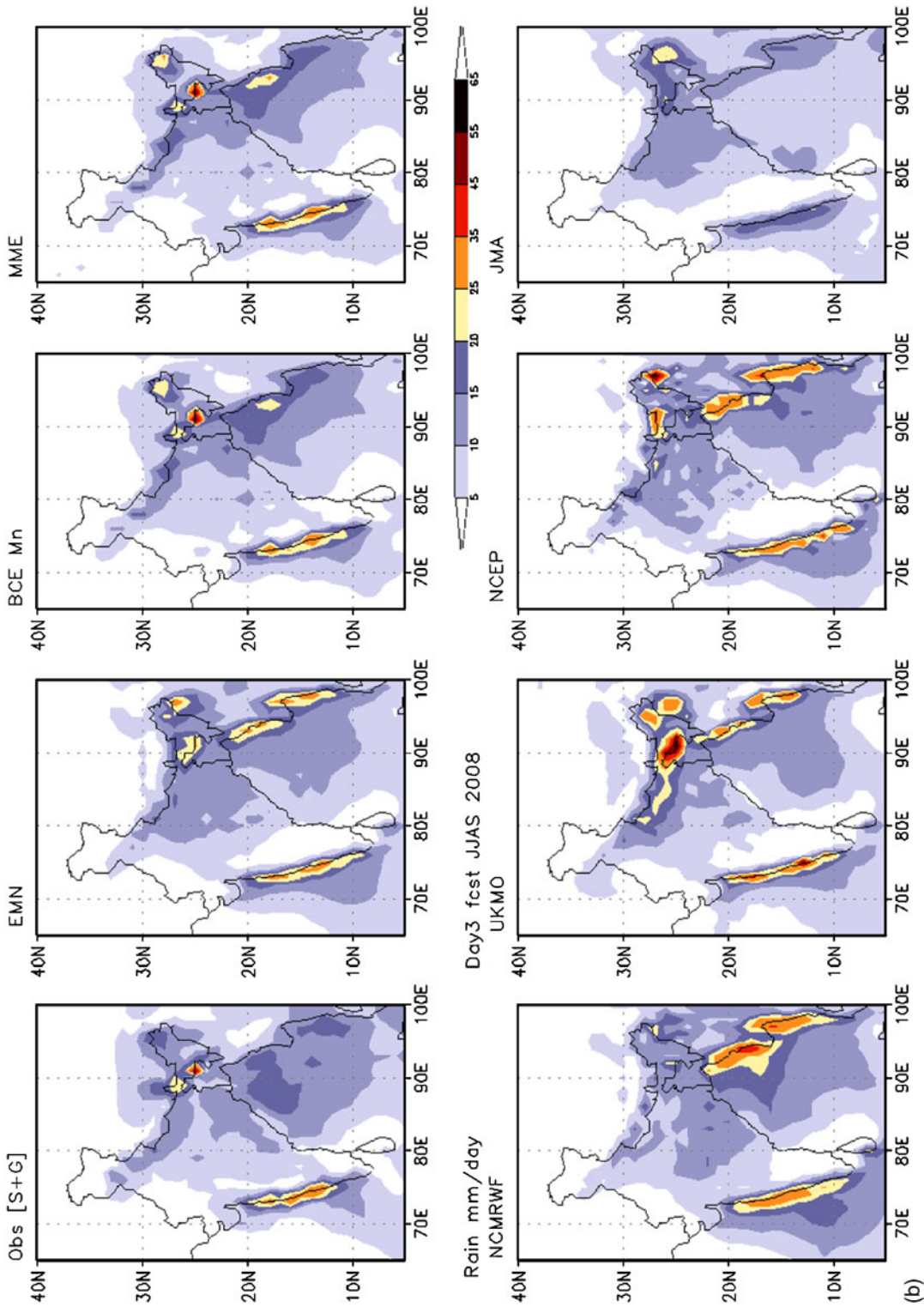


Figure 1. Total rainfall during monsoon 2008 (JJAS) from observations, member models and multi-model products: (a) day-1, (b) day-3, and (c) day-5.

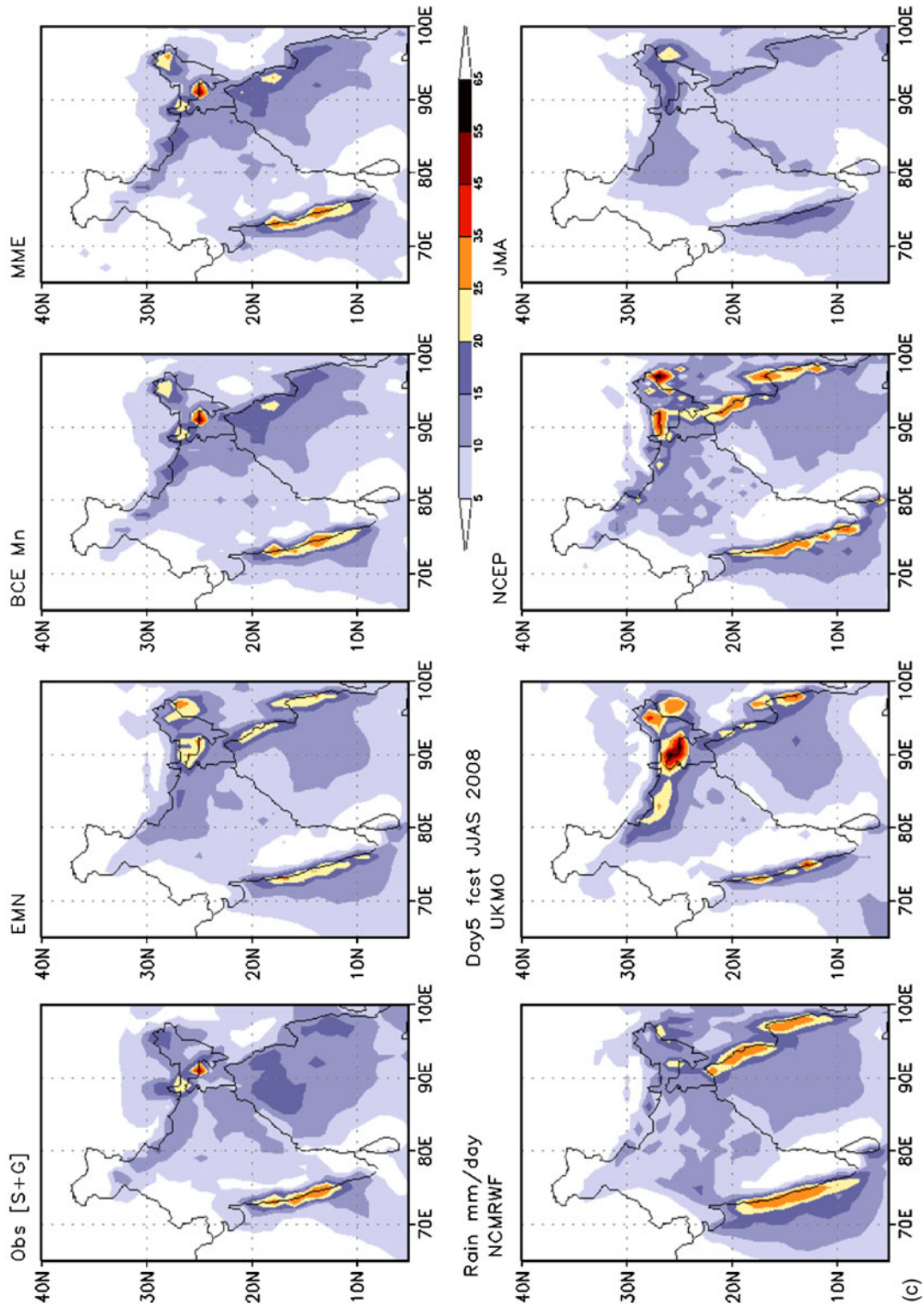


Figure 1. (Continued).

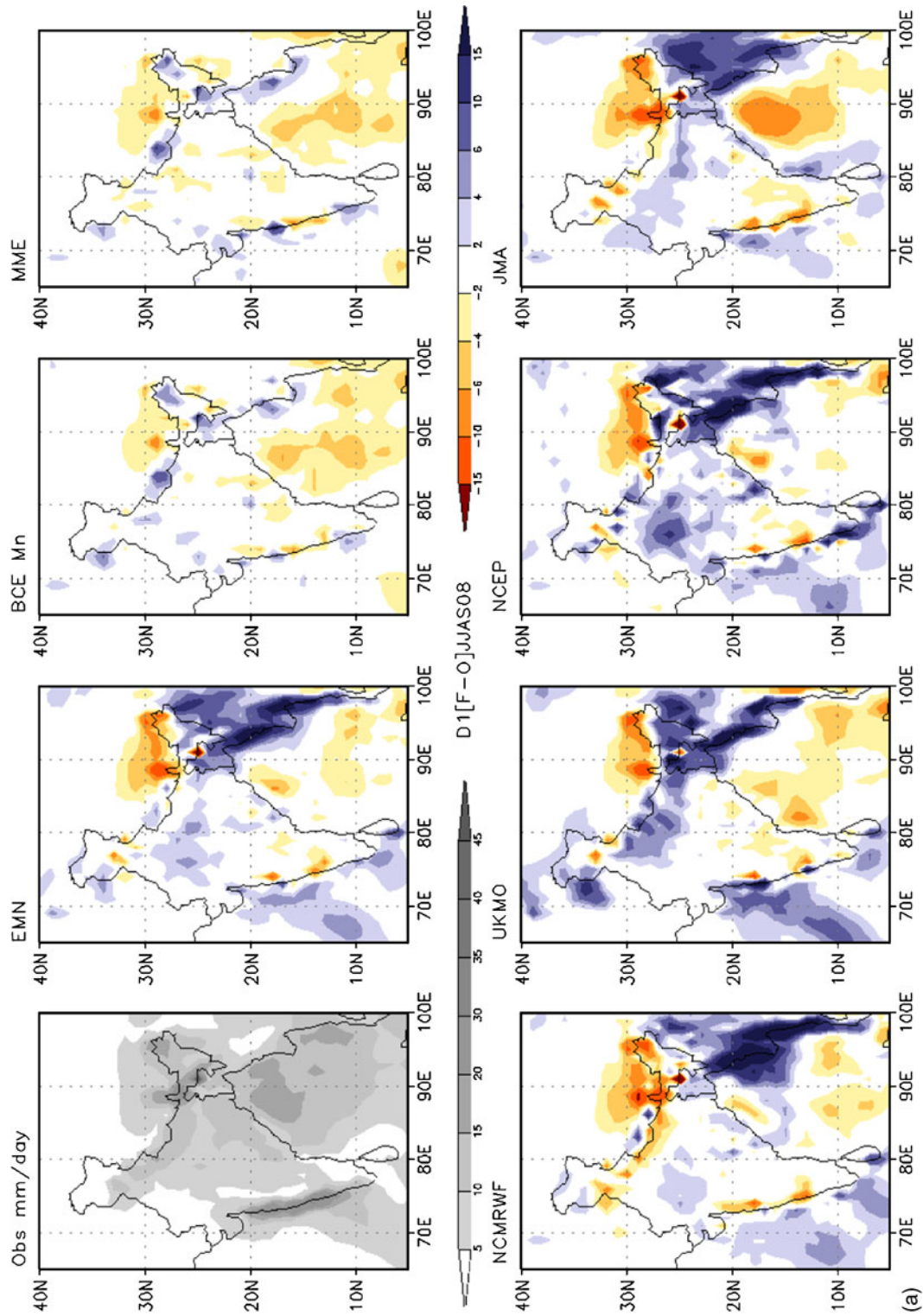
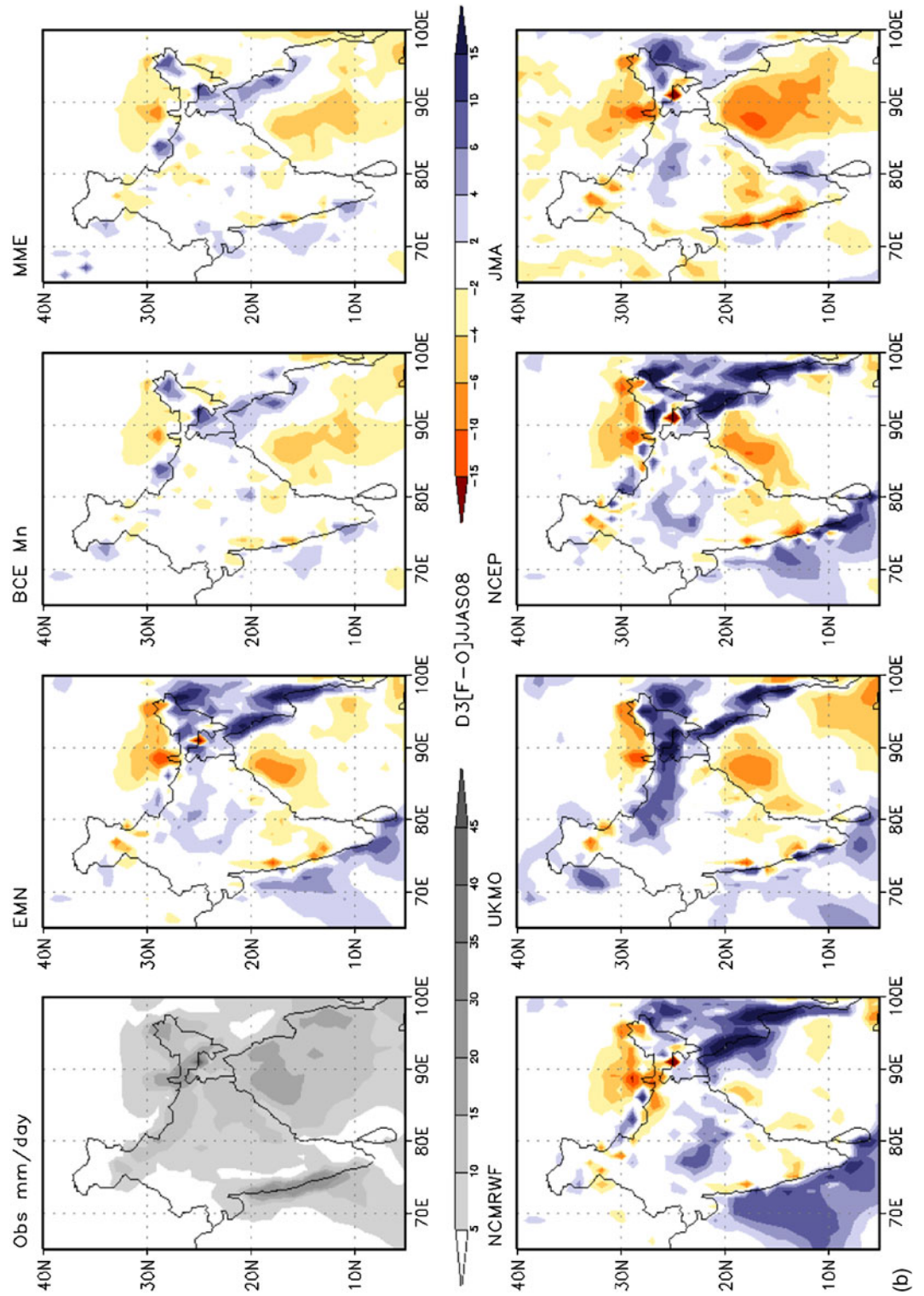


Figure 2. Difference of observations from forecasts of total rainfall (mean error) during monsoon 2008 (JJAS) from member models and multi-model products: (a) day-1, (b) day-3, and (c) day-5.



(b)

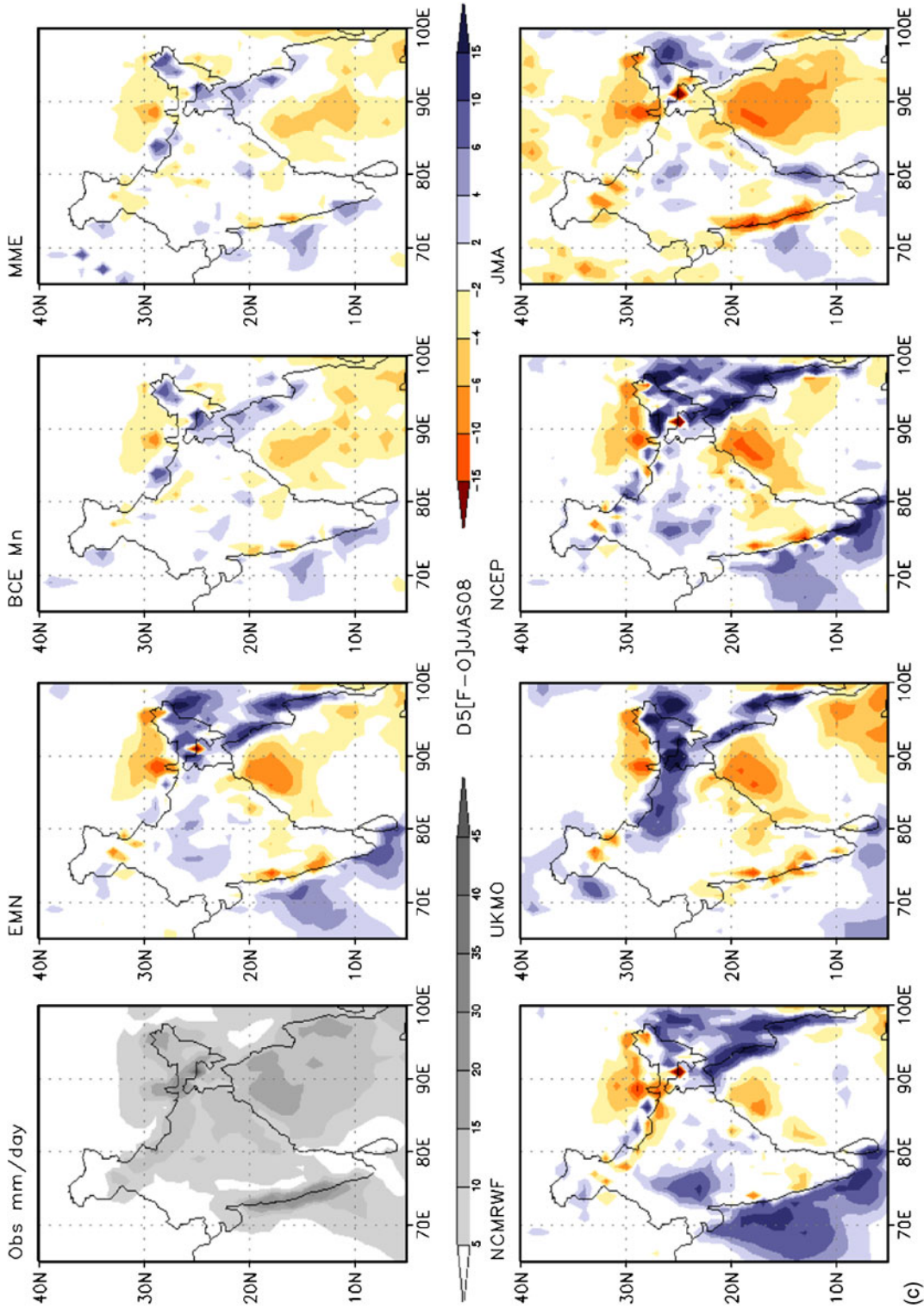


Figure 2. (Continued).

forecasts and are shown in figure 2(a), 2(b) and 2(c), respectively. And for reference, the seasonal rainfall in grey-scale is also given at the top left corner of the plot. The difference plot is interpreted as mean error during the monsoon 2008 season. In these difference (mean error) plots, the positive biases (model is wetter) are shown in shades of blue and the negative biases (model is drier) are shown in shades of orange. Regions of dark blue shades show that the model produces more rain compared to observations. These difference (forecast minus observation) plots clearly bring out the regions of mean biases for day-1, day-3 and day-5 forecasts. For all the days, the biases from the multi-model products like EMN, BCEMn and MME (upper row) show the least biases compared to member models. Among the multi-model products, the least biases are seen in both the MME and BCEMn as compared with EMN for day-1, day-3 and day-5 forecasts.

The above described mean error in a sense shows the systematic error of the model. However, while summing for the season, errors of opposite sign might be getting cancelled in some regions and the bias representation may not be fully informative. Hence, the mean error (bias) plots have to be examined in conjunction with the root mean square error (RMSE) values for the aforementioned regions also. The RMSE of the member model and the multi-model forecast products in comparison to the observations are given in figure 3(a), 3(b) and 3(c) for day-1, day-3 and day-5, respectively. Again for reference, the observed plot is shown at top left corner of the diagram. For day-1 (figure 3a) among member models NCMRWF and NCEP have higher RMSEs. UKMO RMSE is lesser than NCEP and NCMRWF. Another feature of the RMSE is that the errors are more in models where the rainfall amounts are also more. For example, the west coast of India, monsoon trough region and the Arakan coast region have the higher RMSE values. The forecasts from the multi-model products have less RMSE compared to member models. Again, among them the BCEMn and MME have lesser RMSE compared to EMN. When we examine the day-3 (figure 3b) and day-5 (figure 3c) RMSEs we see for all member models the values are more compared to day-1. The errors grow gradually from day-1 to day-5. However, again in both day-3 and day-5, the multi-model forecast products have less RMSEs compared to their respective member models. On day-5, the BCEMn has the least RMSE.

During monsoon, in medium range forecasts, it is important and challenging to predict the day-to-day rainfall associated with the passing transient weather systems or the fluctuating strength

of the monsoon. Therefore, the rainfall anomaly in forecasts and observations has to be examined in terms of their similarity. The anomalies of the observation and forecasts are computed from their respective seasonal means during 2008 monsoon for day-1, day-3 and day-5 forecasts and are shown in figure 4(a), 4(b) and 4(c), respectively. The anomaly correlation coefficients (ACC) are plotted in a scale of 0 to 1 in figure 4. Shades of yellow, orange or red shows gradual increase of ACC from 0.5 to 0.9, and ACC below 0.5 are shown in shades of blue. For day-1 forecasts we see in general the ACCs are higher for all member models. But the multi-model products like the EMN and BCEMn have higher scores of ACC compared to member models. Going from day-1 (figure 4a) to day-3 (figure 4b), immediately we see a drop in skill in ACC for the member models. Only UKMO model shows some skill. However, for day-3 forecasts the EMN and BCEMn have higher skills than member models and even the MME. For day-5 forecasts the skill of member models all come below 0.2 in terms of ACC. UKMO again has higher skill than other member models. The multi-model products from day-5 of EMN and BCEMn have higher skill compared to all member models and the MME. From this it can be concluded that for monsoon rainfall forecasts in medium range, the current state-of-the-art models have some skill till day-3, which is slightly enhanced by the multi-model technique. But for day-5, all the model and multi-model skills are very low, and may not be having any forecast value. The modelling community have to focus research on how to improve the skill of rainfall associated with transient weather systems beyond day-3. In the seamless concept this may lead to improved monthly and seasonal forecast of monsoon rainfall.

All the above results discussed give some general idea of the quality of rainfall forecasts in terms of error statistics for monsoon for the member global models and the products from the multi-model algorithm. But we are aware of the limitations of the numerical models in simulating the final products in model, i.e., the rainfall quantity at the right regions. Therefore, it is relevant to examine and document the skill of rainfall forecasts in terms of rainfall amounts in different categories (different threshold amounts of rainfall) in terms of threshold statistics. Standard statistical parameters like equitable threat score (ETS), hit rate (HR) and bias score are computed for the comparisons in different categories of rainfall amounts. A brief description of these categorical statistics is given in Ebert *et al* (2007). ETS is commonly used as an overall efficiency measure for inter-comparison of precipitation products. ETS gives the fraction

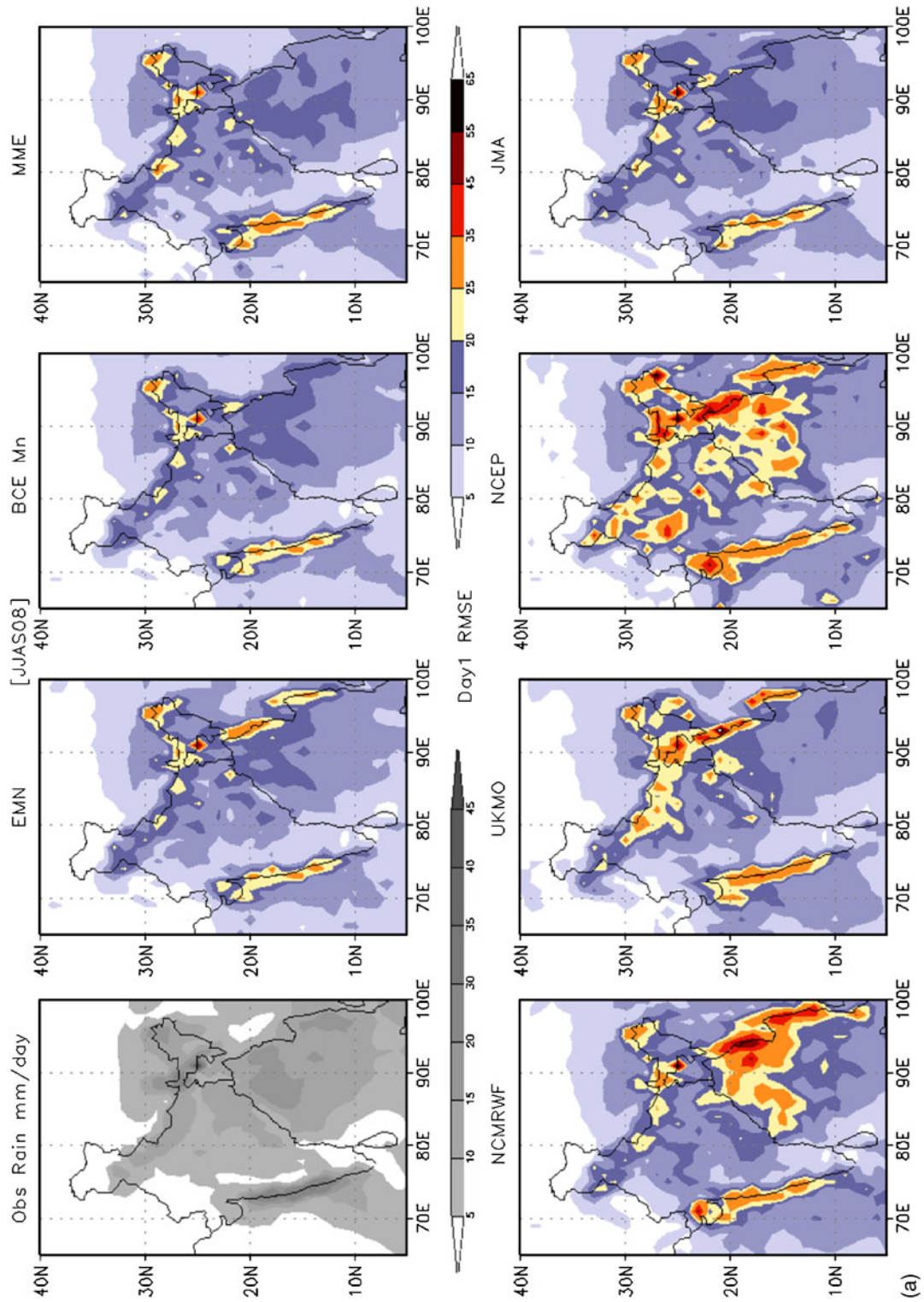
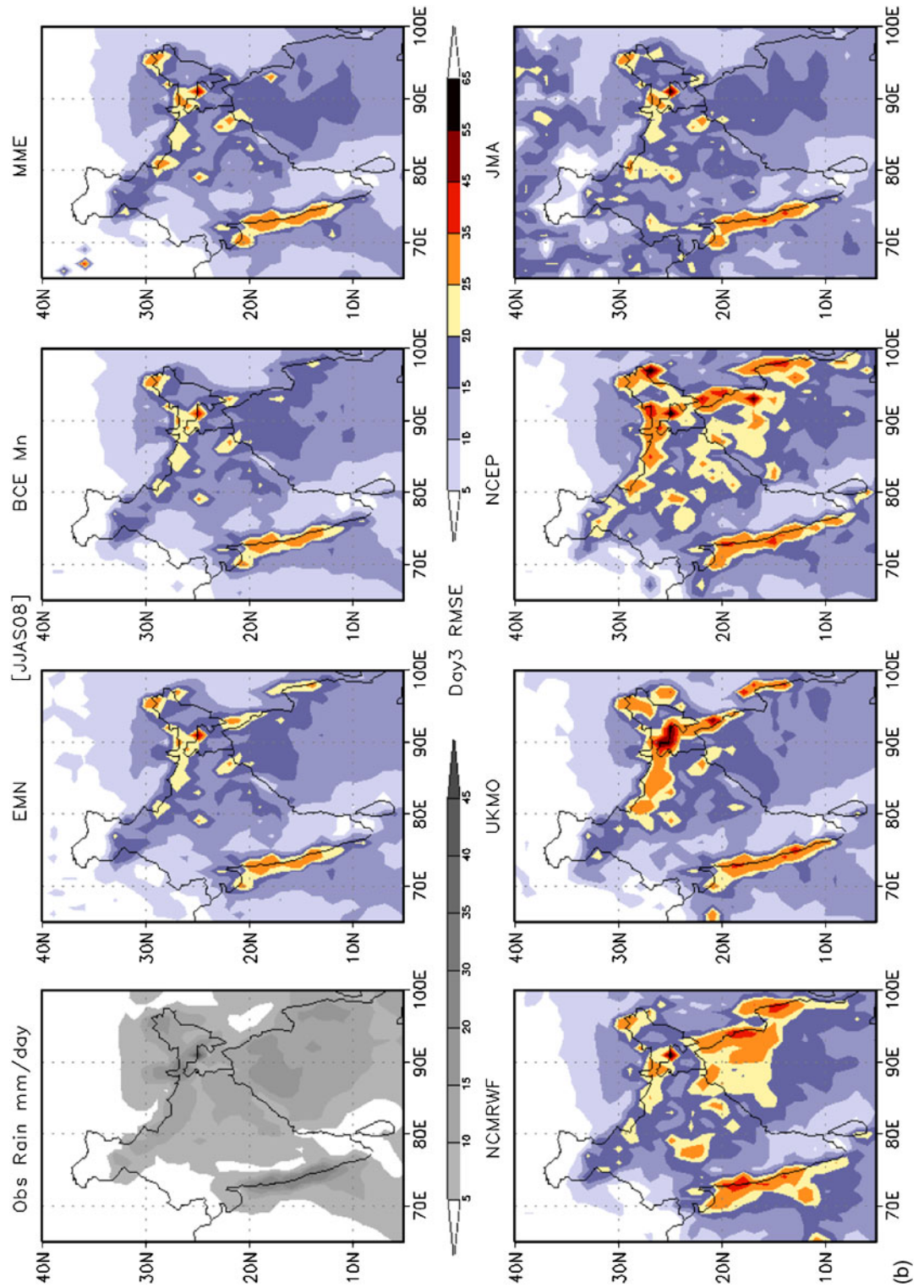


Figure 3. Root mean square error (RMSE) rainfall during monsoon 2008 (JJAS) for member models and multi-model products: (a) day-1, (b) day-3, and (c) day-5.



(b)

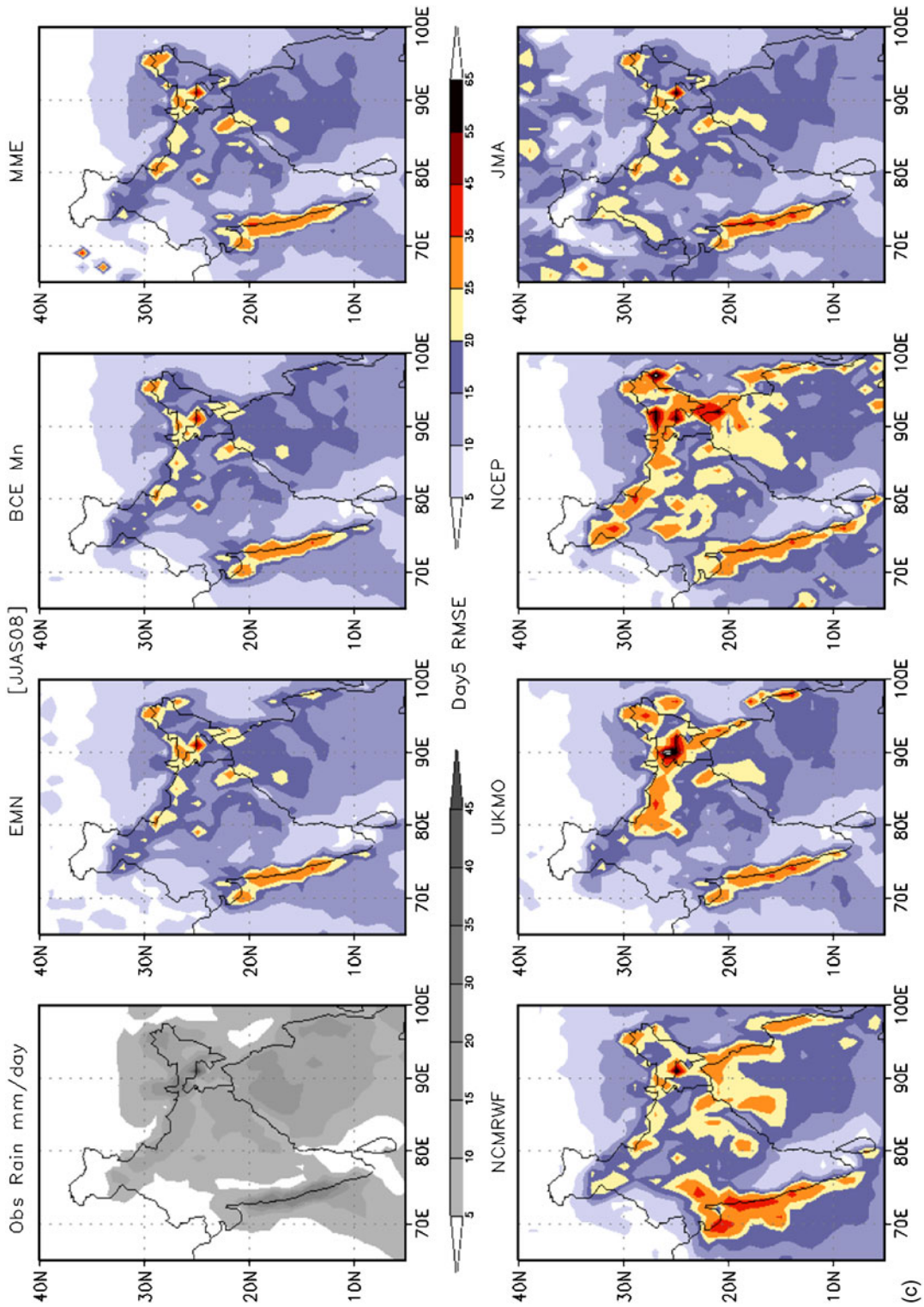
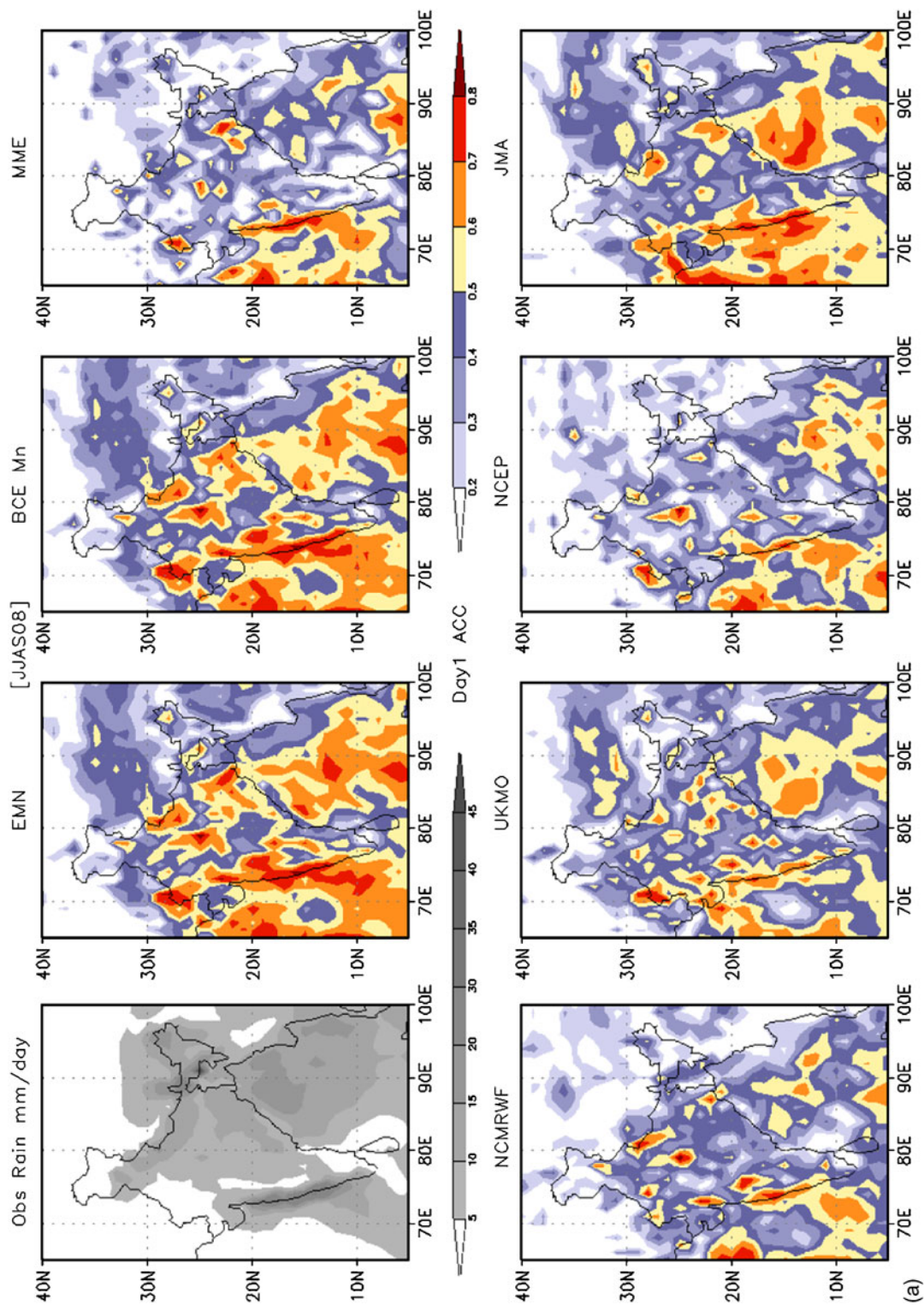


Figure 3. (Continued).



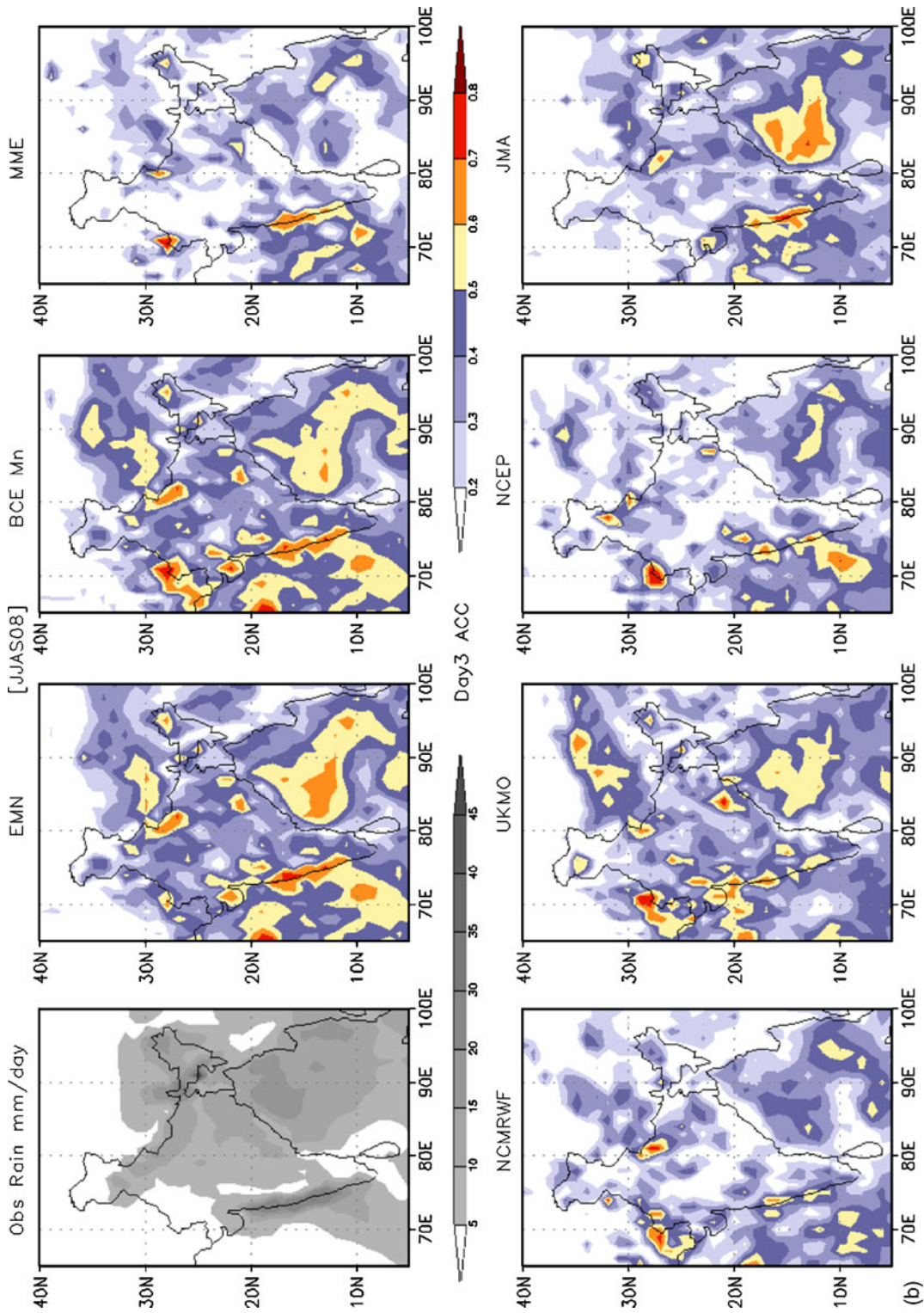


Figure 4. Anomaly correlation coefficient (ACC) rainfall during monsoon 2008 (JJAS) for member models and multi-model products: (a) day-1, (b) day-3, and (c) day-5.

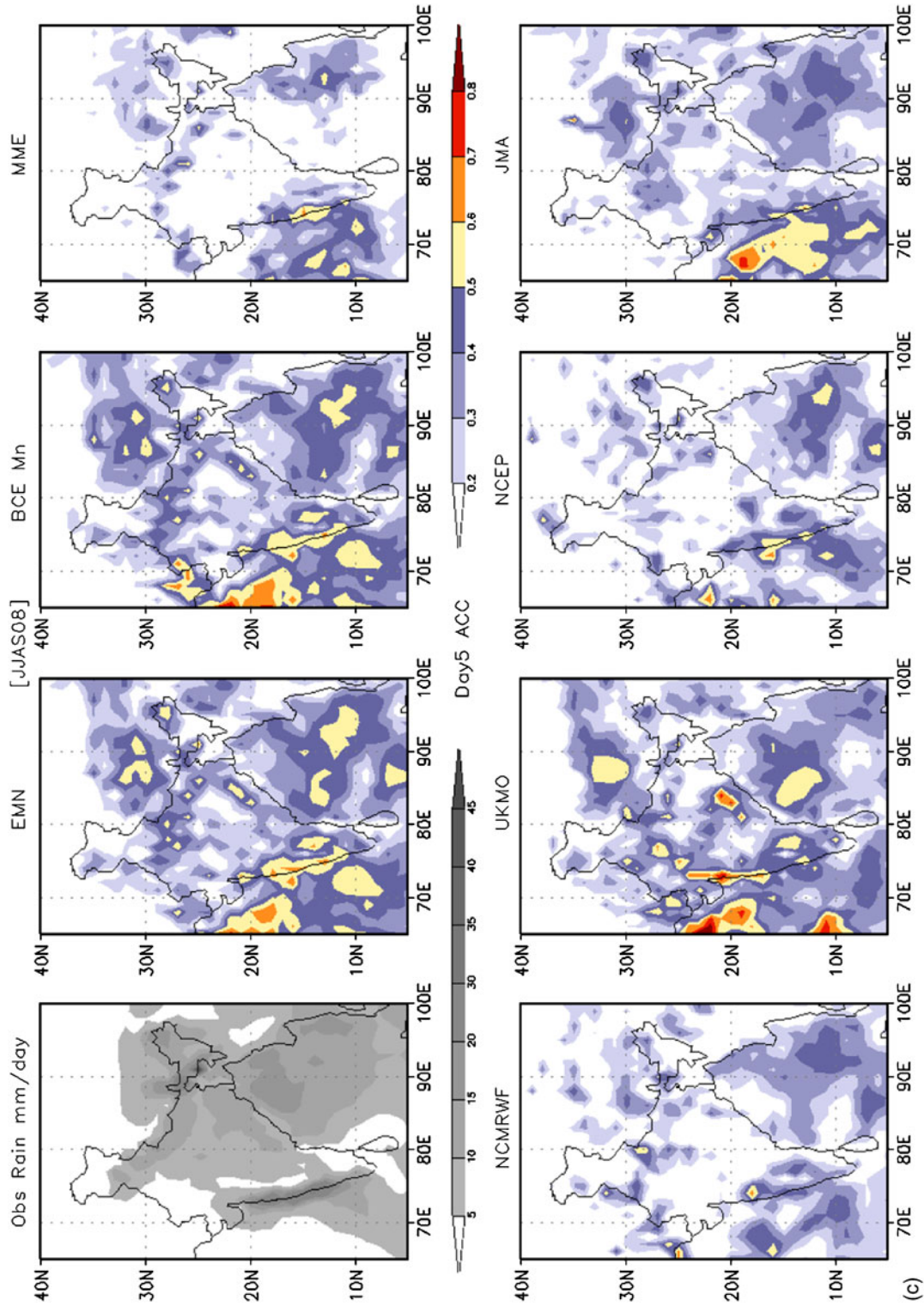


Figure 4. (Continued).

of observed and/or detected rain that was correctly detected and is adjusted for the number of hits that could be expected purely due to random chance. HR (success rate) 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 bias score is the ratio of the estimated to predicted rain areas. For four different regions, namely, all of India, central India, peninsular India, and the west coast were selected to compute these categorical skill scores. These scores are obtained from the data covering the daily values for the entire 2008 monsoon season of 122 days (JJAS). 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 land regions. All of India covers all the land grid points between 67°E to 100°E longitude, and 7°N to 37°N latitude. The central India domain covers the monsoon trough region, extending from 73°E to 90°E longitude and 22°N to 28°N latitude. The west coast domain of India extends from 70°E to 78°E longitude and 10°N to 20°N latitude. The peninsular India domain extends from 74°E to 85°E longitude and 7°N to 21°N latitude. The performance measures used are in terms of different threshold statistics. Here, six thresholds from 1 to 6 cm per day (interval of 1 cm per day, on x-axis) are considered. For all four domains, for the 2008 monsoon season the skills are shown in figures 5–8, for different length of forecasts like day-1, day-3 and day-5.

The left panel in figure 5(a, b, c) shows the ETS for thresholds 1 to 6 cm per day, for day-1, day-3 and day-5 forecasts for the aforementioned all India region. As well known, for any given day (forecast length) the ETS generally gradually decreases with increasing threshold. Again, with increasing length of forecast period (day-1 to day-5) for each threshold rainfall category, the ETS skill score decreases gradually. The ETS for all member models look similar for all forecast length and thresholds. However, in general, the ETS of UKMO is slightly higher than other member models for higher thresholds even with increasing forecast lengths and seen to have even higher (or comparable) skills than the multi-model products at higher thresholds. The multi-model products like the EMN, BCEMn and MME show higher skill in ETS compared to member models for different length of forecasts and thresholds. In general, among the multi-model products the BCEMn shows higher skill than EMN and MME. Therefore, in terms of multi-model products for rainfall

in terms of ETS skill, we can conclude that the use of multi-model has some benefits compared to using single independent models for issuing rainfall forecasts. The middle panel in figure 5(a, b, c) shows the bias scores for different threshold and forecast lengths for the aforementioned all India region box under consideration. If the bias value is greater than 1, then the model has a positive bias (tendency to rain at more points compared to observed points/regions for a particular threshold under consideration). It is interesting to see in day-1 forecasts that the member models are doing very good for lower thresholds (up to 3 cm per day), and the multi-model products are not able to improve upon them. And for higher thresholds, the multi-model is bringing the biases to the negative side. The same is true for the bias score roughly for day-3 and day-5 forecasts. The right panel in figure 5(a, b, c) shows the hit rate of rainfall forecast for different threshold and forecast lengths. For the hit rate, the multi-model products are mostly seen to perform much better than the four member models. For all days, the hit rate from MME and BCEMn are superior to simple EMN and the member models for all thresholds. This also shows the power of the multi-model algorithm to be able to produce rainfall at right grids for different thresholds. It is able to correct the rainfall amounts for different thresholds. MME is producing the best skill for all days and thresholds for the all India region box.

After examining the all India region box, we now study three more different and typical regions of the monsoon. The central India box considered is the most crucial for the monsoon forecasting point, as it is closely related to the monsoon trough (heart of the Indian monsoon system) position and intensity forecast. For state-of-the-art global models, even in medium range, correctly simulating the monsoon trough position and intensity (along with embedded monsoon lows and depressions) is a very challenging task. The manifested rainfall forecast is even more tough for the monsoon trough region (central India box here). Figure 6(a, b, c) shows the ETS, bias score and the hit rate for different days of forecast length and different threshold rainfall values. We notice an immediate drop in ETS skill in central India region as compared to the earlier discussed all India region skills. Beyond 3 cm/day threshold the ETS skill is as low as 0.05 or even lower for all days. Only for threshold of 1 and 2 cm, we find some skill for different days. Here, again the multi-model products are able to enhance the skill over the member models. Almost for all threshold and days, the

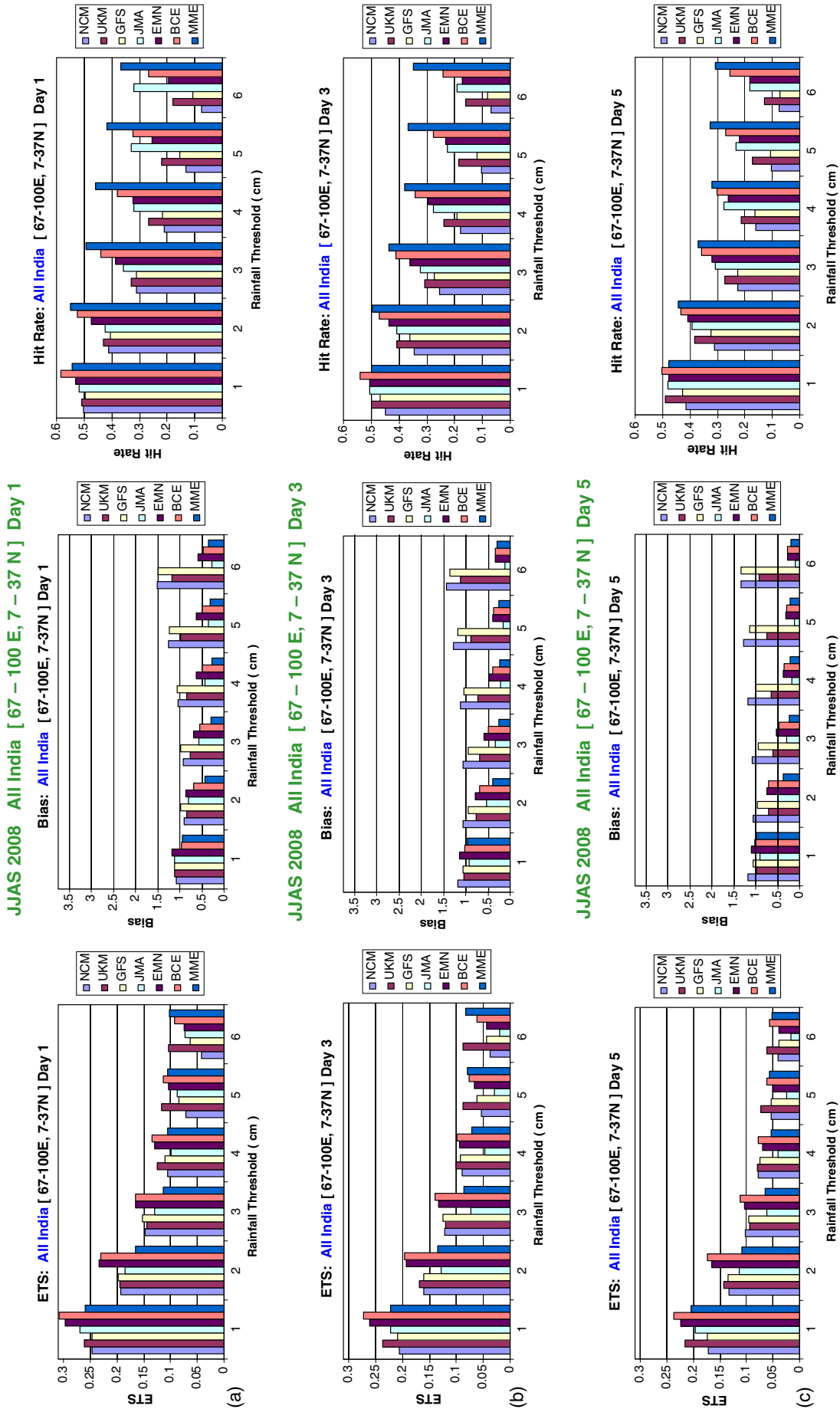


Figure 5. Categorical rainfall forecast skill score (ETS, bias, hit rate) during monsoon 2008 (JJAS) for member models and multi-model products, for the All India region: (a) day-1, (b) day-3, and (c) day-5.

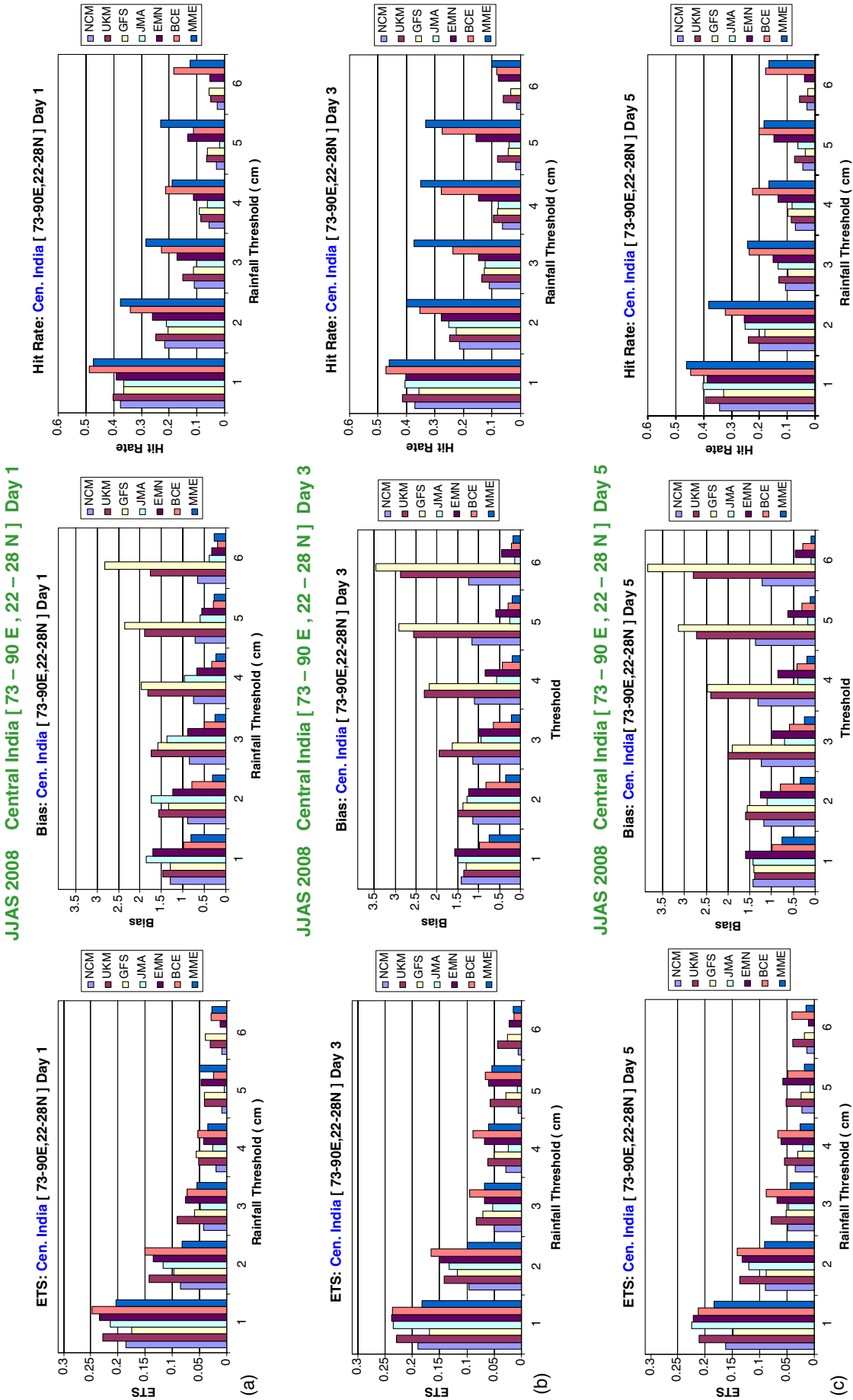


Figure 6. Same as figure 5, except for central India region.

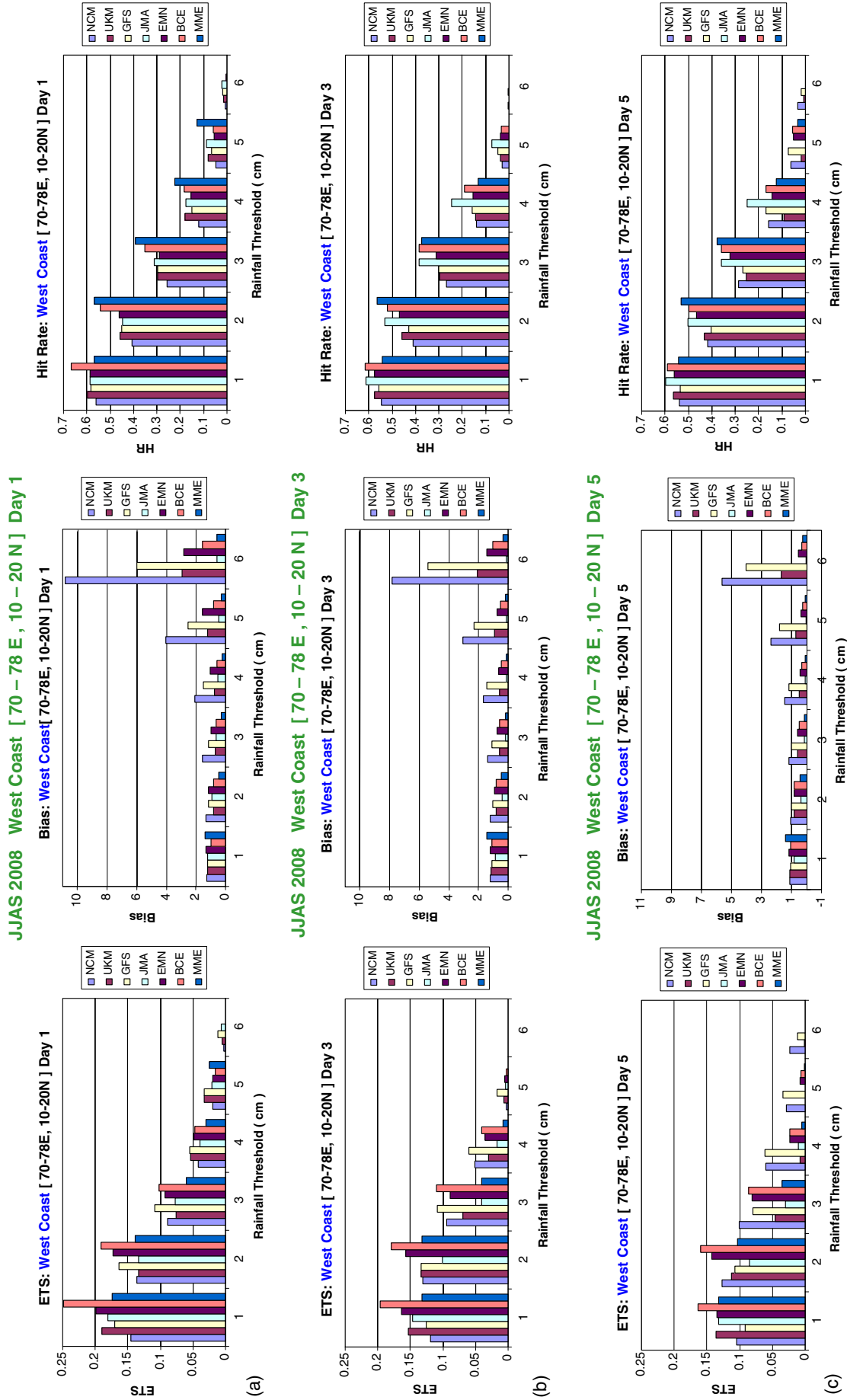


Figure 7. Same as figure 5, except for west coast region.

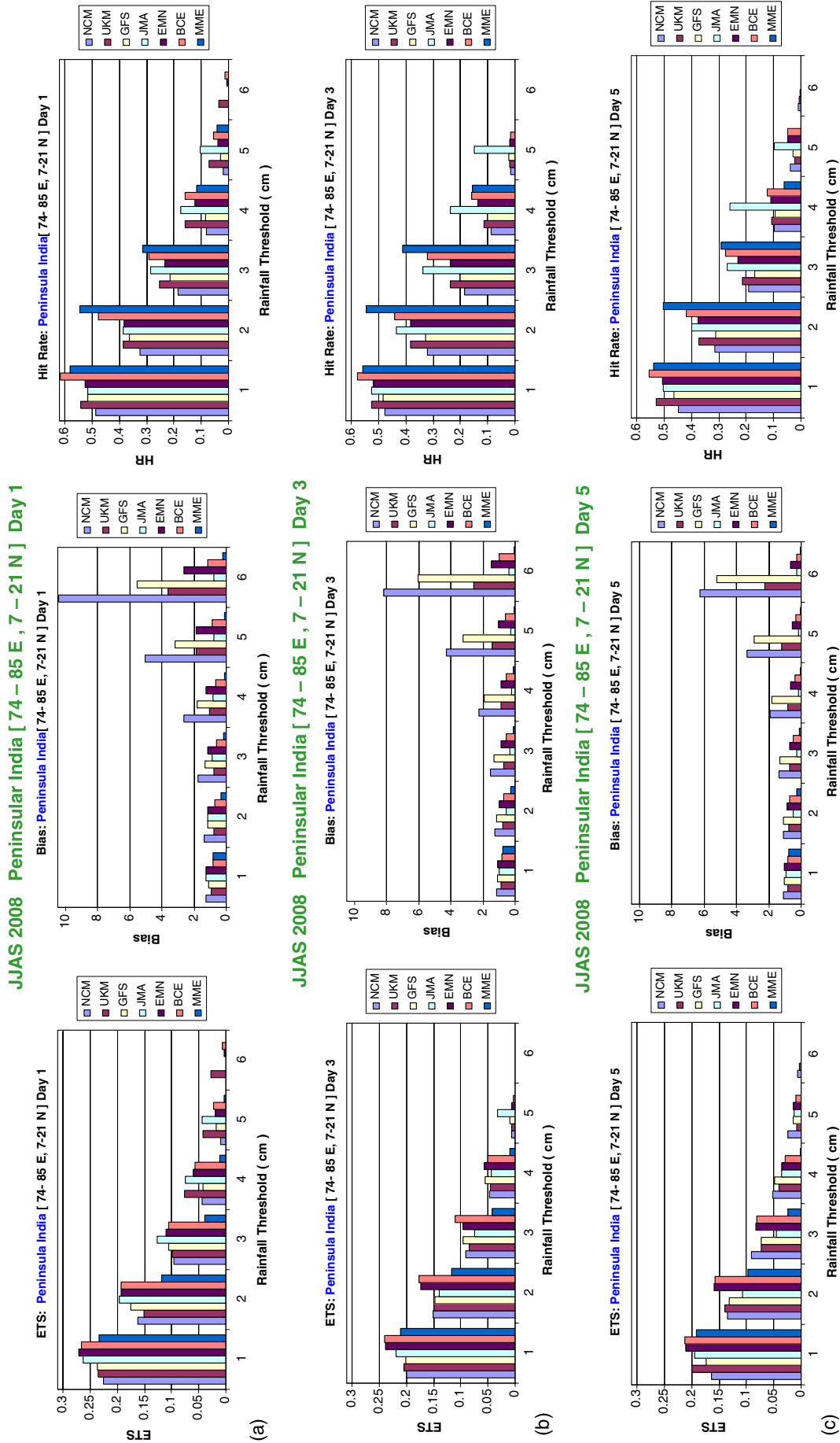


Figure 8. Same as figure 5, except for peninsular region.

multi-model ETS skill show higher value than the member models. Among multi-model products, in terms of ETS, the EMN and BCEMn show higher skill than MME. However, when the skill of the models are only up to the threshold of 2 cm per day, we can not expect to use the model forecast for higher rainfall amount events. The multi-model algorithm improved slightly above the member models. The basic information content has to come from member models only. For central India region, the bias score indicates a higher bias (raining at more points) for almost all thresholds and days by the UKMO and GFS models. With increasing thresholds (higher rainfall amount events) the bias of UKMO and GFS are seen to gradually increase. The bias of the NCMRWF model look modest compared to other member models. For lower thresholds (up to 3 cm) for all days the multi-model products do a good job, by reducing the bias of the models. BCEMn is superior to EMN and MME. But for higher thresholds (above 3 cm) the multi-model products bring the bias to the negative side, and again BCEMn is best among multi-model products in terms of bias score for central India region. For the hit rate again (for central India) the scores fall rapidly with increasing thresholds. The multi-model products have higher skill of hit rate compared to member models. MME and BCEMn have high skill compared to simple ensemble mean EMN. For all thresholds and days, the hit score from multi-model products are higher compared to four member global models.

West coast of India having the Western Ghats mountains is another important region for the monsoon studies. The higher orography of the Western Ghats interacts with the lower troposphere moist winds of the Arabian Sea (low-level westerly jet of the monsoon) to produce heavy rainfall amounts. The strength of the large-scale monsoon flow is manifested in the low-level flow and the associated rainfall over the west coast region of India. The ETS, bias score and the hit rate score for the west-coast region box are shown in figure 7(a, b, c) for different days and rainfall thresholds. For the west-coast region, the ETS score show rapid fall in skill with increasing threshold for all forecast lengths. The ETS skills are very low beyond the threshold of 3 cm. For higher rainfall amounts the model ETS skill are very low for all days. For threshold up to 3 cm, the multi-model products are able to improve upon the member models. The EMN and the BCEMn are better than MME. As the skill of member models beyond 3 cm are very low, what the multi-model algorithm enhances are unimportant from forecasting point of view. Basically, the model simulation for the higher rainfall amounts has to be improved further to reap any benefit from the multi-model

algorithm. The bias score for the west coast region look better than the central India and the all India scores. For higher rainfall thresholds (beyond 5 cm) particularly the NCMRWF and GFS produce higher positive biases, and the multi-model products seem to improve the bias score there for all days. Here, the simple EMN look better than BCEMn and the MME. In the hit rate scores, again for all thresholds and days, generally the multi-model products are better than the member models. For hit rate MME and BCEMn are better than the simple EMN.

The last region where the skill of rainfall forecast from member models and the multi-model products are examined is the peninsular India region (figure 8a, b, c). Generally, this region rains less when the monsoon trough is active. Peninsular India gets rainfall if some monsoon lows or depressions encroach into this area, or sometimes during the passage of rainfall bands from south to north. Therefore, this is also a sensitive region to monitor the model rainfall skill during monsoon. Similar to other regions discussed earlier, here also, the ETS skill score decreases with increasing threshold and increasing length of forecast days. Here, the member model skills are very low beyond 3 cm threshold for day-3 and day-5. At lower thresholds (3 cm and below) the multi-model products namely the BCEMn and the EMN show higher skill compared to member models. Except day-5, for higher thresholds (4 cm and above) also we see marginal improvements in ETS from the multi-model products, however, all the member model and the multi-model skills are so low that it can not be used for forecasting. For peninsular India, the bias score for higher rainfall thresholds (above 4 cm) the member models show high positive biases. The BCEMn is able to reduce the biases for all days. The hit rate score for the peninsular region from the multi-model products (MME and BCEMn) also show improvement up to 3 cm threshold for all days. Beyond 3 cm threshold the scores are very low for the hit rate for all days, and the multi-model is unable to enhance the scores much above the member models.

4. Summary and future work

During monsoon 2008 on an experimental basis MME forecasting of large-scale monsoon precipitation in medium range was carried out in real-time at NCMRWF/MoES, India. Apart from simple ensemble mean and bias corrected ensemble mean, linear regression based weights were given to four global models to obtain the multi-model ensemble forecasts. In general, the multi-model ensemble

forecast products have higher skill than individual model product. The skill score for Indian domain and other sub-regions indicate that the bias corrected ensemble mean produces the best skill. Giving weights to different model to get an MME product helps to improve over individual member models marginally. The basic skills of the global model rainfall decrease rapidly beyond day-3. To get real benefits of the MME algorithm the basic single model skill of rainfall has to be improved beyond day-3. It was the first step in the learning experiences in MME. It has worked like a feasibility study of the MME experiment for rainfall forecast. While looking for enhanced skill in MME, as a by-product, the member models also get verified and we are able to keep track of the state-of-the-art model's performance. This feedback is useful for continuous model development and other modelling related research. From poor man's MME, we are progressing towards great grand MME under WMO/TIGGE. In coming years we plan to include other newer model data into this MME system. In this study we use rather simple techniques and standard scores to assess the usefulness and benefits of the MME forecast against member models. However, with increasing number of member models and ensemble members from each model, to be able to understand and document the full potential and usefulness of the MME products both in deterministic and probabilistic sense various skill scores have to be used (Cusack and Arribas 2008). Probabilistic ensemble forecasting has to be taken up for tropics. Bowler *et al* (2008) have shown the usefulness of a short-range ensemble prediction system which will be made operational at UKMO. They show that the regional ensemble is more skillful than the global ensemble, and compares favourably to the ECMWF ensemble for many variables. In India also a regional ensemble system for probabilistic forecasts has to be experimented.

Recent approach used by Krishnamurti *et al* (2009) on downscaling the global model data to a regional scale and then applying MME algorithm will be experimented in a future study.

Acknowledgements

This report is the outcome of the MoES MME project. Steering committee members of the project had provided useful suggestions during the course of this study. Thanks are due to Head, NCMRWF for his constant support and encouragement. Data used in this study were obtained from NCEP (via ftp), UKMO (through NCMRWF) and JMA, Japan (through IMD). Daily METEOSAT-IR data available through ftp was used to prepare the observed rainfall estimates.

References

- Bowler N E, Arribas A, Mylne K R, Robertson K B and Beare S E 2008 The MOGREPS short range ensemble prediction system; *Quart. J. Roy. Meteor. Soc.* **134** 703–722.
- Cartwright T J and Krishnamurti T N 2007 Warm season mesoscale superensemble precipitation forecasts in the southeastern United States; *Weather Forecast.* **22** 873–886.
- Cusack S and Arribas A 2008 Assessing the usefulness of probabilistic forecasts; *Mon. Weather Rev.* **136** 1492–1504.
- Ebert E E 2001 Ability of a poor man's ensemble to predict the probability and distribution of precipitation; *Mon. Weather Rev.* **129** 2461–2480.
- Ebert E E, Janowiak J and Kidd C 2007 Comparison of near real time precipitation estimates from satellite observations and numerical models; *Bull. Am. Meteorol. Soc.* **88** 47–64.
- Evans R E, Harrison M S J, Graham R J and Mylne K R 2000 Joint medium-range ensembles from the Met. Office and ECMWF system; *Mon. Weather Rev.* **128** 3104–3127.
- Gadgil S 2003 The Indian monsoon and its variability; *Annu. Rev. Earth Planet. Sci.* **31** 429–467.
- Hamill T M, Hagedorn R and Whitaker J S 2008 Probabilistic forecast calibration using ECMWF and GFS ensemble forecasts, part II: precipitation; *Mon. Weather Rev.* **136** 2620–2632.
- Johnson C and Swinbank R 2009 Medium-range multimodel ensemble combination and calibration; *Quart. J. Roy. Meteor. Soc.* **135** 777–794.
- Krishnamurti T N, Kishtawal C M, LaRow T E, Bachiochi D R, Zhang Z, Williford C E, Gadgil S and Surendran S 1999 Improved weather and seasonal climate forecasts from multi-model super-ensemble; *Science* **285** 1548–1550.
- Krishnamurti T N, Kishtawal C M, Shin D W and Williford C E 2000 Improving tropical precipitation forecasts from a multi-analysis super-ensemble; *J. Climate* **13** 4217–4227.
- Krishnamurti T N, Vijaya Kumar T S, Yun W T, Chakraborty A and Stefanova L 2006a Weather and seasonal climate forecasts using the superensemble approach; In: *Predictability of Weather and Climate* (eds) Tim Palmer and Renate Hagedorn (ECMWF), Cambridge University Press, pp. 532–560.
- Krishnamurti T N, Vijay Kumar T S and Mitra A K 2006b Seasonal climate prediction of Indian summer monsoon; In: *The Asian Monsoons* (ed.) Bin Wang, pp. 553–582.
- Krishnamurti T N, Mishra A K, Chakraborty A and Rajeevan M 2009 Improving Global Model Precipitation Forecasts over India using downscaling and the FSU superensemble. Part I: 1–5-Day forecasts; *Mon. Weather Rev.* **137**(9) 2713–2735.
- Mishra A K and Krishnamurti T N 2007 Current status of multi-model super-ensemble operational NWP forecast of the Indian summer monsoon; *J. Earth Syst. Sci.* **116**(5) 1–16.
- Mitra A K, Dasgupta M, Singh S V and Krishnamurti T N 2003 Daily rainfall for Indian monsoon region from merged satellite and raingauge values: Large-scale analysis from real-time data; *J. Hydrometeorol. (AMS)* **4**(5) 769–781.
- Mylne K R, Evans R E and Clark R T 2002 Multi-model multi-analysis ensembles in quasi-operational

- medium-range forecasting; *Quart. J. Roy. Meteor. Soc.* **128** 261–384.
- Richardson D S 2001 Ensembles using multiple models and analyses; *Quart. J. Roy. Meteor. Soc.* **127** 1847–1864.
- Roy Bhowmik S K and Durai V R 2008 Multi-model ensemble forecasting of rainfall over Indian monsoon region; *Atmosfera* **21**(3) 225–239.
- Roy Bhowmik S K and Durai V R 2010 Application of multi-model ensemble techniques for real time district level rainfall forecasts in short range time scale over Indian region; *MAP* **106**(1–2) 19–35.
- Webster P J, Magana V O, Palmer T N, Shukla J, Tomas R A, Yanai M and Yasunari T 1998 Monsoons: Processes, predictability and the prospects for prediction; *J. Geophys. Res.* **103C** 14,451–14,510.
- Whitaker J S, Wei X and Vitart F 2006 Improving week-2 forecasts with multimodel reforecast ensembles; *Mon. Weather Rev.* **134** 2279–2284.
- Woodcock F and Engel C 2005 Operational consensus forecasts; *Weather Forecast.* **20** 101–111.
- Woods A 2006 Medium range weather prediction (USA: Springer Publications) 261 pp.

MS received 10 March 2010; revised 15 September 2010; accepted 21 September 2010