NGFS rainfall forecast verification over India using the contiguous rain area (CRA) method

KULDEEP SHARMA, RAGHAVENDRA ASHRIT, ELIZABETH EBERT*,
GOPAL IYENGAR and ASHIS MITRA


*Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne

e mail : kuldeep@ncmrwf.gov.in

ABSTRACT. The real time medium range rainfall forecasts of NCMRWF's Global Forecast System (NGFS) are evaluated over India (land only) against 0.5 degree gridded rainfall (IMD-NCMRWF) observations during JJAS of 2010-2013 using the features-based Contiguous Rain Area (CRA) method. The model resolution is about ~35 km in 2010 and ~25 km during 2011-2013. The emphasis of this study is the spatial verification of model rainfall forecasts associated with the Bay of Bengal low pressure systems that cause widespread rainfall over eastern and central parts of India. Based on IMD reports, 45 episodes of rainfall events related to low pressure systems were identified during JJAS 2010-2013 over India. This paper focuses on the verification of NGFS's global forecast system (NGFS) using the contiguous rain areas (CRA) method (Ebert and McBride, 2000) during monsoon season (June, July, August, September, JJAS) of 2010-2013 over India. This technique has been applied to the heavy rainfall events (low pressure, depression and deep depression) as reported by IMD during the season. It is a features-based method which compares the properties of the matched forecast and observed features, where a “feature” is any weather event that can be drawn as a closed contour on a grid.

Key words – NGFS, CRA, Forecast verification, QPF.

1. Introduction

Traditional skill scores are commonly used for the verification of NWP model rainfall forecasts over the Indian monsoon region during the last few years (Ranade et al., 2014; Mitra et al., 2011 and Mandal et al., 2007). However, as the spatial and temporal resolution of forecasts from numerical weather prediction (NWP) models grows increasingly finer, there is a need for spatial verification approaches that adequately reflect the quality of these forecasts without over penalizing the errors at the grid scale. There are many spatial verification techniques which are used in verification including neighborhood or fuzzy verification, scale decomposition, features-based verification, and field deformation approaches [for reviews of these methods - Casati et al. (2008) and Gilleland et al. (2009)]. These techniques focus on different aspects of forecast quality such as scale-dependent accuracy, location errors, intensity errors and the realism of the spatial pattern. The majority of these spatial methods require forecasts and observations matched on a common grid.

This paper focuses on the verification of NCMRWF’s global forecast system (NGFS) using the contiguous rain areas (CRA) method (Ebert and McBride, 2000) during monsoon season (June, July, August, September, JJAS) of 2010-2013 over India. This technique has been applied to the heavy rainfall events (low pressure, depression and deep depression) as reported by IMD during the season. It is a features-based method which compares the properties of the matched forecast and observed features, where a “feature” is any weather event that can be drawn as a closed contour on a grid.
Rainfall over India during the monsoon season is mainly brought about by the moisture buildup due to southwesterly flow in the lower troposphere. However, the rainfall activity over certain regions is enhanced due to the formation of low pressure systems over the Bay of Bengal, which sometimes intensify (into depression and deep depression) and move inland over eastern parts of India. These systems often develop deep convection and cause heavy rain along the track as they move northwestwards along the monsoon trough. These low pressure systems play a crucial role in producing rainfall over central India. This study focuses on the verification of rainfall activity over this region (Eastern India) with special emphasis on low pressure systems. In the next section, data and methodology will be discussed. Further results and discussion produced by using the CRA method on heavy rainfall events during JJAS 2010-2013 will be discussed.

2. Data and methodology

In the present study, we have used the observed and NGFS model forecast rainfall data over India during the monsoon (June-September) of four years from 2010-2013. The NGFS model is NCMRWF’s global forecast system (NGFS; T574L64). During 2010 the NGFS system features a horizontal grid resolution of about ~35 km (T382L64). However, since 2011 the horizontal resolution is about ~25 km (T574L64). The evaluation of the rainfall forecasts is carried out at 0.5 degree grid resolution for day-1 through day-5. However, for brevity, only the verification results corresponding to day-1 forecasts are presented.

2.1. Observed rainfall data over India

Rainfall analysis based on quality controlled observations is very useful and critical for verification of the NWP forecasts. In this study, the India Meteorological Department (IMD) and NCMRWF’s merged gridded daily rainfall data set for the period 2010-2013 is used. IMD’s rain gauges provide data over the Indian mainland while TRMM 3B42 data over the ocean has been used to prepare the merged product (Mitra et al., 2009). These correspond to 24-hour rainfall accumulations valid at 0300 UTC.
2.2. NGFS model forecast rainfall data over India

The NGFS (T382L64 system) was operational in May, 2010. Later a parallel upgraded system was implemented at a resolution of T574L64 in November, 2010 with all of the latest developments in the data decoding, assimilation, model and pre/post processing available at that time. Further details can be obtained at http://www.ncmrwf.gov.in/gfs_report_final.pdf. The forecasts were generated based on initial condition of 0000 UTC. The rainfall is accumulated for a 24 hour period ending at 0300 UTC on the next day. For the uniformity with the observations, the model quantitative precipitation forecasts (QPFs) are regridded to 0.5° resolutions by using bilinear interpolation of the total rainfall volume, including ocean grid points.

The domain chosen for the present study is 7°-38.5° N, 67-100.5° E for the period of JJAS 2010-2013. The spatial distribution of mean monsoon rainfall from IMD-NCMRWF observed and Day-1 forecast is shown in Figs. 1(a-h). The eastern part of India (78° E-89° E, 15-28° N) shown by the red box in Figs. 1(a-h) is mostly affected by monsoon troughs, lows and depressions during the season. The forecast rainfall (distribution and amount) over eastern India closely follows the forecast position and movement of low pressure systems. The errors in tracks and location of low pressure systems are directly related to the errors in rainfall distribution. Object-based spatial verification methods are used to quantify the errors in rainfall location and distribution. In the next section, the CRA method of spatial verification used in the present study is described briefly.

2.3. CRA verification method

The CRA method is an object-based verification procedure suitable for gridded forecasts that was developed to estimate the systematic errors in forecasts for rainfall systems (Ebert and McBride, 2000; Ebert and Gallus, 2009). It was one of the first methods to measure errors in predicted location and to separate the total error into components due to errors in location, volume and pattern. The steps involved in CRA technique are described in Ebert and Gallus (2009). A brief summary of the procedure is given here.

Firstly a CRA is defined for an observation/forecast pair based on a user-specified isohyet (rain rate contour) in the forecast and/or the observations. It is the union of the forecast and observed rain entities as illustrated in Fig. 2. The forecast and observed entities need not overlap, but they must be associated with each other, that is, they must be nearby and associated with a common synoptic situation. During the monsoon season large parts of India regularly receive rainfall in the range up to 10 mm/day, with embedded areas of much higher rainfall so CRAs defined by thresholds of 10, 20, and 40 mm/day were tested to identify and isolate the heavy rain events.

In the next step a pattern matching technique is used for estimating the location error. Here the forecast field is horizontally translated over the observed field in a series of iterations until the best match is obtained. The location error is then simply the vector displacement of the forecast.

The best match between the two entities can be determined either (a) by maximizing the correlation coefficient, (b) by minimizing the total squared error, (c) by maximizing the overlap of the two entities, or (d) by overlaying the centres of gravity of the two entities. For a good forecast all of the methods will give very similar location errors. In the present study the best match is determined by maximizing the correlation. The mean squared error (MSE) of the forecast and its decomposition (location error, volume error and pattern error) are computed as shown below (Grams et al., 2006, for details of the derivation).

$$MSE_{\text{Total}} = MSE_{\text{Displacement}} + MSE_{\text{Volume}} + MSE_{\text{Pattern}} \quad (1)$$

where, the component errors are estimated as

$$MSE_{\text{Displacement}} = 2S_r S_o (r_{\text{OPT}} - r),$$

$$MSE_{\text{Volume}} = (F' - O'),$$

$$MSE_{\text{Pattern}} = 2S_r S_o (1 - r_{\text{OPT}}) + (S_F - S_O)^2 \quad (2)$$

and are often presented as fractions of the total MSE. In the above expressions $F'$ and $O'$ are the mean
418 MAUSAM, 66, 3 (July 2015)

Fig. 3. Verification of NGFS Day-1 rainfall forecasts valid for 1 August, 2013

forecast and observed precipitation values after shifting the forecast to obtain the best match, $S_F$ and $S_O$ are the standard deviations of the forecast and observed precipitation, respectively, before shifting. The spatial correlation between the original forecast and observed features ($r$) increases to an optimum value ($r_{OPT}$) in the process of correcting the location via pattern matching.

A forecast is considered ‘good’ when the observation-forecast pair match very well (are very similar). For all such ‘good’ cases, the CRA error decomposition is easily done (‘good’ CRAs). A forecast is considered ‘bad’ when the observation-forecast pair do not match well (are very different). For all such ‘bad’ cases CRA error decomposition is likely to be erroneous (‘bad’ CRAs). In such CRAs involving ‘bad’ forecasts, the displacement is considered incorrect since the correlation ($r_{OPT}$) is not significant or the CRA is shifted out of domain (Ebert and McBride, 2000 for discussion of CRA quality). Verification statistics of error decomposition are based on the ‘good’ CRAs.

3. Results and discussion

3.1. Mean monsoon rainfall: Observation and forecast

Figs. 1(a-h) show observed rainfall and NGFS model predicted rainfall. The four upper (lower) panels show the observed (Day-1 forecast) rainfall. The upper panels show rainfall amounts exceeding 1.5 cm/day over the west coast of India. Similarly, over central and eastern India, rainfall amounts exceeding 0.6 cm/day (core monsoon) cover a large area with isolated pockets of higher (>1.5 cm/day) mean rainfall amounts. High rainfall over north-east India is also observed.

The model captures the rainfall over west coast of India. It is able to capture large scale rainfall features like higher rainfall over the monsoon trough region and north-east India, and lower rainfall over southeast peninsular India and northwest India. The year to year variability in monsoon rainfall over India is also represented by the model. The Day-1 forecast is in good agreement with the observations over the west coast while the model overestimates over eastern coast and central India region in all the seasons. Also, the model slightly overestimates rain over the peninsula and north-east parts of the country as compared with observed rainfall. During JJAS 2010, the model over-predicts rainfall along the monsoon trough region whereas it under-predicts it during JJAS 2011-2013. The NGFS model shows a dry bias over the north-west parts of the country in almost all seasons.

3.2. Verification of a depression (30 July-1 August, 2013)

Within the season there are active and weak rainfall spells and rainfall associated with lows and depressions.
Evaluation of rainfall associated with such systems is of interest, including their displacement, pattern, and intensity errors. We start with an example of a depression that formed in Bay of Bengal and produced heavy rain in central India on 1 August, 2013.

The national scale rainfall verification starts with several commonly used QPF statistics as shown in Fig. 3 for Day-1 forecasts from the NGFS model (For definitions of the statistics shown in Fig. 3, Wilks, 2011). The Day-1 forecast valid for 1st August 2013 indicates that the forecast (as compared to observation) has smaller number of raining grids 582 (636), slightly lower average rain rate of 34 mm/day (37 mm/day) and lower maximum rain rate 248 mm/day (280 mm/day). The spatial correlation of 0.44 and the categorical skill scores for rain exceeding 1 mm (equitable threat score of 0.28, Hanssen and Kuipers score of 0.43) indicate moderate skill at predicting the location of the rain. Also, the value of mean absolute error is about 11 mm/day and the root mean square error (RMSE) is about 22 mm/day. This is mainly due to the observed rainfall over central India (20° - 25° N, 70° - 75° E) being severely underestimated by the model. Also, the rainfall over some pockets of the west coast is also underestimated. The value of false alarm ratio is 0.35, probability of detection is 0.59 and bias score is 0.91. This is because the model predicts false rainfall over Gujarat region and some parts of North-west of Uttar Pradesh (27.5° - 30° N, 70° - 75° E). We can also see that the rainfall over eastern and central India is largely underestimated by the model.

The errors in the rainfall distribution can be quantified using CRA technique. Fig. 4 shows the CRA verification using a 40 mm/day rainfall threshold to isolate the heavy rainfall over the eastern part of India. The CRA is bounded by the domain 18° - 24° N and 75° - 83° E. The spatial map shows the observed and forecast rainfall with the 40 mm/day contour shown in bold. The scatter plot on the right indicates reasonably good agreement.
### TABLE 1
Table represents the dates and duration of rainfall systems that formed during JJAS 2010-2013 over Bay of Bengal

(LP : Low pressure, D : Depression, DD : Deep Depression, WLP : Well marked Low)

<table>
<thead>
<tr>
<th>Year/Month</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>9-13(LP), 24-26 (LP)</td>
<td>2-5 (LP), 6-8 (LP), 24-26 (WLP), 28-1 Aug(L)</td>
<td>4-8 (WLP), 12-13 (LP), 23-27 (LP), 30-31 (LP)</td>
<td>3-6 (LP), 8-13 (LP), 17-20 (WLP)</td>
</tr>
<tr>
<td>2011</td>
<td>11-12(D), 16-23 (DD), 29-30 (LP)</td>
<td>6-7 (LP), 13-16 (LP), 22-23 (D)</td>
<td>8-11 (LP), 11-17 (WLP)</td>
<td>29 Aug-10 (WLP), 22-23 (D), 6-15 (LP), 13-19 (LP)</td>
</tr>
<tr>
<td>2012</td>
<td>-</td>
<td>7-11(LP), 20-22(LP)</td>
<td>3-9 (LP), 12-14 (LP), 17-22 (LP), 25-27 (LP), 30-31(LP)</td>
<td>3-10 (WLP), 10-11 (LP)</td>
</tr>
<tr>
<td>2013</td>
<td>4-5 (LP), 5-7 (LP), 12-17 (LP)</td>
<td>10-13 (LP), 15-17 (LP), 19-25 (LP), 25-29 (LP), 30-1 Aug (D)</td>
<td>20-23 (D), 9-11 (LP)</td>
<td>23-29 (LP)</td>
</tr>
</tbody>
</table>

### TABLE 2
The total number of CRAs for rainfall threshold of 40 mm/day in NGFS forecasts and the number of CRAs with correct displacements over India and Eastern coast of India

<table>
<thead>
<tr>
<th>For rainfall threshold 40 mm/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of CRA’s (including good and bad CRA’s) all over India</td>
</tr>
<tr>
<td>Total number of good CRA’s all over India</td>
</tr>
<tr>
<td>Total number of good CRA’s during LP days</td>
</tr>
<tr>
<td>Total Number of good CRA’s over East India during LP days (78-89E,15-28N)</td>
</tr>
<tr>
<td>Total number of good CRA’s during Non-LP days</td>
</tr>
<tr>
<td>Total Number of good CRA’s over East Coast during Non-LP days (78-89E,15-28N)</td>
</tr>
</tbody>
</table>

between observed and forecast rainfall after the shifting the forecast rainfall to correct the location error. The numbers below the scatter plot show (i) number of grids with rainfall excess of 40 mm/day (ii) the average rain rate (mm/day) (iii) the maximum rain (mm/day) and (iv) the rain volume (km$^3$) in the observations and forecasts. In the forecasts the maximum rain (highest rain amount) is lower (235 mm/day) than the observed value (280 mm/day). The number of grids with rainfall exceeding 40 mm/day in the forecasts is 44 as against 87 in observations; the forecast average rain rate and volume in the CRA are 61 mm/day and 22 km$^3$ as against observed values of 32 mm/day and 11 km$^3$, respectively. The forecast has a RMSE of 55 mm/day which is mainly contributed by errors in pattern (46%) and displacement (41%).

#### 3.3. CRA results for low pressure systems during JJAS 2010-2013

This section describes the performance of the NGFS model during the heavy rainfall episodes and the rest of the days (other than heavy rainfall episodes) which occurred during JJAS 2010-2013 over eastern parts of India (78-89° E, 15-28° N). Based on India Meteorological Department (IMD) reports, we have identified 45 cases of Lows, Depressions and Deep Depressions (Table 1). As seen in the table, there are 33 low pressure systems, 6 well marked lows, 5 depressions, and 1 deep depression which caused wide spread rainfall during JJAS 2010 - 2013. We will abbreviate these 45 heavy rainfall days as Low Pressure days (LP) and rest as Non- Low Pressure days (Non-LP days).

CRA statistics are obtained for each day of the four seasons for rainfall thresholds of 40 mm/day. This threshold was chosen to focus on the heaviest rain regions which may be expected to have a large impact on the underlying population. Subsequently, the CRA statistics are summarized using box-whisker plots separately for the LP and Non-LP days.

As described in section 2.3, the CRA verification uses pattern matching technique to decompose the total
Fig. 5. RMSE in NGFS rainfall along with different components of errors for LP and Non-LP days during JJAS 2010-2013 for 40 mm/day threshold. RMSE is in mm/day and error components are in percentage but plotted using the same vertical scale. The divisor represents median and box represents percentiles (25 and 75) and outliers are represented by dots.

error into components. In the analysis carried out for the whole season, the fraction of rejected CRAs is direct indicator of the model performance, that is, better performing models should have relatively lower rejection rates of CRAs because forecast heavy rain systems more frequently correspond well with observed heavy rain systems. The total number of CRAs over India along with number of good CRAs for rainfall thresholds of 40 mm/day for Day-1 forecasts is shown in Table 2. Also total number of good CRAs over eastern parts of India is also shown in Table 2.

The percentage of CRAs rejected for 40 mm/day thresholds is 62% all over India. The percentage of CRAs rejected during LP days (Non-LP days) over eastern parts of India is 67 (65)%. Accurate prediction of amount and location of heavy rains associated with embedded convection is still a challenge. Comparing to observations, the forecast location and intensity show a mismatch for heavy rains. This is reflected in high number of rejections in CRA analysis. This issue can be addressed in two possible ways and is planned for future work. One way is to verify 48 hour and 72 hour accumulated rainfall for higher thresholds (40 mm and above). Generally heavy rains are spread over a large area. It is also possible to revise the size criterion (based on which CRA objects are constructed) for higher threshold.

The RMSE and different components of error (volume, pattern and displacement expressed as percentage of total MSE) for good CRAs using a rainfall threshold of 40 mm/day for LP days and Non-LP days in eastern India are summarized in box-whisker plots in Fig. 5. The number of good CRAs for LP days (Non-LP days) for rainfall threshold of 40 mm/day over eastern parts of India is 84 (118). The average RMSE in rainfall forecasts is found to be approximately 50 mm/day during both LP and Non-LP days for CRAs with rainfall thresholds of 40 mm/day. The similarity of the RMSE magnitude to the heavy rain threshold was also found by Ashrit et al. (2015), who verified monsoon rainfall forecasts from the Met Office global model over a six-year period.

For these CRAs the contributions from pattern error are dominant (47% and 44% for LP and Non-LP days, respectively) and the contribution from volume error is generally least (15% during LP and Non-LP days, respectively). The contributions from displacement error are 37 (40) % during LP days (Non-LP days). The similarity of the relative error contributions between LP and Non-LP days suggests that the causes of model error for LP situations may not be very different than in Non-LP cases.

4. Summary

We have evaluated the performance of NGFS model rainfall forecast during four monsoon seasons (2010-2013) using CRA technique with special focus on low pressure systems that formed over Bay of Bengal which caused wide spread rainfall over eastern and central parts of India (78°-89° E, 15°-28° N). We have identified 45 cases of rainfall based on IMD reports.

(i) Forecast mean monsoon rainfall has dry (wet) bias over the north-west and west-coast (north-east) part of the country by 0.5 cm/day. The mean rain amounts are overestimated in Eastern India by more than 0.5 cm/day during 2010-2012 while rainfall is largely underestimated by more than 0.3 cm/day except for some pockets of eastern coast of India.

(ii) The best matching observed-forecast pairs yield “good” CRAs for which error decomposition results are meaningful and significant. For poorly matching observation-forecast pairs (“bad” CRAs), the error decomposition results are not significant. Better performing models should have relatively large number of “good” CRAs. In the present study, the percentage of “bad” CRA’s for 40 mm/day threshold during LP days (Non-LP days) is 67 (65)% over eastern part of India. This suggests that model still has a significant challenge in accurate prediction (location and intensity) of heavy rain events.

(iii) In the eastern part of India which is mainly affected by lows and depressions, the average RMSE in rainfall forecasts on LP days is found to be approximately...
50 mm/day with similar values found during Non-LP days when a CRAs rainfall threshold of 40 mm/day is used.

(iv) The CRA statistics shows the main contribution to the errors is because of pattern, followed by displacement, while the contribution due to volume is least in both LP days and Non-LP days.

Displacement and pattern errors are associated with errors in dynamics (predicted flow) while volume error is associated with errors in physics (moisture) treatment. These components provide guidance for model developers when the statistics of error components are studied for large samples of cases.

References


