Verification of short range forecasts of extreme rainfall during monsoon

RAGHAVENDRA ASHRIT, KULDEEP SHARMA, ANUMEHA DUBE, GOPAL IYENGAR,

ASHIS MITRA and E. N. RAJAGOPAL

NCMRWF, MoES, A-50, Institutional Area-II, Sector-62, Noida, (U. P.), India

e mail : raghu@ncmrwf.gov.in, raghu.ashrit@nic.in

सार – भारत वर्ष में मॉनसून ऋतु (जून-सितम्बर) के दौरान दैनिक वर्षा बड़े पैमाने पर सिनॉप्टिक और मेसोस्केल विक्षोभों के परस्पर प्रभाव पर निर्भर करती है जिसमें छुटपुट वर्षा के दौर और कभी कभी भारी वर्षा के दौर आते हैं। इस प्रकार के दौर से अक्सर कुछ ही दिनों में अत्यधिक मात्रा में वर्षा हो जाती है जिसका ऋतुनिष्ठ औसत वर्षा के आकलन पर काफी बड़ा प्रभाव पड़ सकता है। मुंबई में 26 जुलाई, 2005 को प्रतिदिन 100 सें. मी. से अधिक वर्षा का रिकार्ड होना एक बाहरी चरम घटना है जो किसी शहर के लिए जून-सितम्बर में प्रतिदिन 18 मि. मी. की औसत वर्षा से 20 मानक विचलन अधिक हैं एवं 28 मि. मी./प्रतिदिन के दैनिक मानक विचलन से भी अधिक है।

हालांकि इस प्रकार की बाहरी घटनाएं भारत में मॉनसून ऋतु के दौरान असामान्य नहीं हैं, ये घटनाएं उच्च विभेदन के पूर्वानुमान मॉडल्स के लिए भी गंभीर चुनौती प्रस्तुत करती हैं। हाल की सात मॉनसून ऋतुओं (2007-2013) में प्रेक्षित और दैनिक वर्षा के मॉडल पूर्वानुमान दोनों के लिए इन बाहरी घटनाओं के आँकड़ों की जाँच की गई है। एक दिन में हुई वर्षा की कुछ बाहरी घटनाएं (पूर्वी भारत के मैदान क्षेत्र में) ऋतु की कुल वर्षा की 30 प्रतिशत वर्षा करती हैं । इस शोध पत्र में परंपरागत सत्यापन अंकों जैसे:- प्रोबेबिलीटी ऑफ डिटेक्शन (POD), इक्यूटेबल थ्रीट स्कोर (ETS), क्रिटीकल सक्सेस इंडेक्स (CSI) इत्यादि की विभिन्न श्रेणियों के लिए उपयोग करके वर्षा का सत्यापन प्रस्तुत किया गया है। इन नई विधियों की संक्षिप्त समीक्षा में अत्यधिक उग्र घटनाओं का सत्यापन कहा गया है, जैसा कि ओड्स रेशियो (OR), एक्सट्रीम डेपेन्डेन्सी स्कोर (EDS), सिमेट्रीक एक्सट्रीम डेपेन्डेन्सी स्कोर (SEDS), एक्सट्रीमल डिपेन्डेन्स इंडेक्स (EDI) और सिमेट्रीक इ डी आई (SEDI) को भारतीय संदर्भ में एक उदाहरण के रूप में प्रस्तुत किया गया है।

ABSTRACT. The daily rainfall over India during the monsoon season (June-September) is governed by the interplay of the large-scale, synoptic and mesoscale disturbances, many of which are sporadic rainfall spells and extremely intense. These spells often bring extreme amounts of rain over only a few days, which can have sizable impacts on the estimated seasonal mean rainfall. The record rainfall of over 100 cm/day in Mumbai on 26th July, 2005 is an outlier/extreme at over 20 standard deviations for activity of typical June-September average rainfall of 18 mm/day with daily standard deviation of 28 mm/day.

While such outliers are not uncommon in India during the monsoon season, they pose serious challenge to even the high resolution forecast models. The statistics of these outlier events are examined both for observed and model-forecast daily rainfall for recent seven monsoon seasons (2007-2013). Some of the extreme one day rainfall events (over the plains of eastern India) contribute up to 30% of the seasonal total rain. This study presents rainfall verification over India using traditional verification scores such as Probability of Detection (POD), Equitable Threat Score (ETH), Critical Success Index (CSI) etc. for various categories. Further, the statistical challenges associated with the verification of the extreme events are discussed. A brief review of the new methods suggested in literature for verification of the extreme events, such as Odds Ratio (OR), Extreme Dependency Score (EDS), Symmetric Extreme Dependency Score (SEDS), Extremal Dependence Index (EDI) and Symmetric EDI (SEDI) is provided with example application to Indian context.

Key words – Rainfall, Observed rainfall, Unified model (UKMO), Observed and forecast mean, Forecast verification.

1. Introduction

The rainfall during the months of June-September (JJAS) forms the life line not just for India but for the entire subcontinent. The monthly and seasonal mean rainfall amounts are generally used to describe the general behaviour of the monsoon. However, such a description can give the misleading impression that the monsoon is a robust slowly evolving system. The mean monsoon rainfall, however, is the rainfall averaged over many sporadic rainfall spells having spatial scales from 100 to 1000 km. As per the India Meteorological

Department (IMD) the average (period: 1961-1990) JJAS accumulated rainfall amount over Mumbai is 216.5 cm. On 27th July 2005 the city of Mumbai (Santa Cruz) recorded 94.4 cm of rainfall in a single day. The same day Colaba observatory recorded 7 cm which is located within 30 km. With a significant portion (43%) of seasonal rainfall in Mumbai (Santa Cruz) received in a single day, *i.e.*, on 27th July, 2005, this event represents an extreme event or an outlier. Such outliers are not uncommon in Indian monsoon rainfall. The most recent events are the heavy rainfall observed in Maharashtra and Uttarakhand. During 1st to 16th June, Mumbai and adjoining areas received rainfall approximately 300% more than the average. Similary over Uttarakhand, the rainfall received during 13th to 19th June, 2013 is approximately 800% more than the average in Kedarnath and adjoining areas. They are caused by embedded convective systems associated with monsoon depressions (Sikka, 1977) and mid tropospheric cyclones (Keshavamurthy, 1973). Thus, issuing a reliable short to medium range (3-7 days) forecast is of utmost importance for heavy rainfall events leading to catastrophic floods, disruption of transport over the affected regions. Early warnings could help the authorities to take necessary measures to reduce the damage to life and property.

As per the IMD nomenclature (Table 1) 6.45-12.44 cm/day rainfall at any location is termed as '*heavy rain*'; 12.45-24.44 cm/day is '*very heavy*'; and rainfall exceeding 24.45 cm/day is termed as '*extremely heavy*'. For cases of rainfall amounts excess of 12 cm/day and close to the highest recorded rainfall for that location, the event is termed '*exceptionally heavy*'. The extreme events tend to be rare and severe and have huge social-economic impact as they bring damage to life and property. Historical records indicate increase in the frequency and intensity of the extreme rainfall events (Roy and Balling, 2004; Goswami *et al.*, 2006).

The forecast accuracy of the Numerical Weather Prediction (NWP) models has steadily improved in the last couple of decades as indicated in numerous studies (Kalnay *et al.*, 1998; Vitart 2013; Magnusson and Kallen 2013; Walters *et al.*, 2014). The quality of a 6-day forecast in 2010 is about the same as the quality of a 3-day forecast was in 1980 for the northern hemisphere (Magnusson and Kallen, 2013). Verification of rainfall forecasts over India is also reported in several of the recent studies (Mandal, *el al.*, 2007; Das *et al.*, 2008; Ashrit and Saji, 2010; Durai and Bharadwaj, 2013). All the earlier studies use the standard verification metrics that include correlation coefficient (CC), Root Mean Squared Error (RMSE), Probability of Detection (POD), False Alarm Ratio (FAR), Threat Score (TS) and Equitable Threat Score (ETS) etc., which are well established and widely used in the

TABLE 1

Nomenclature of rainfall intensity for any station over India

Descriptive Term	Rainfall Amounts (cm/day)
No Rain	0
Very Light Rain	0.01 - 0.24
Light Rain	0.25 - 0.75
Moderate Rain	0.76 - 3.55
Rather Heavy	3.56 - 6.44
Heavy Rain	6.45 - 12.44
Very Heavy Rain	12.45 - 24.44
Extremely Heavy Rain	>24.45
Exceptionally Heavy Rain	>12cm/day and approaching the record highest rainfall amount for the month

literature. However, most of the standard verification metrics are not adequate / suitable for rare and extreme events since the magnitude of these scores 'degenerate to vanishingly low values' (Stephenson *et al.*, 2008).

The models now include many complex physical processes and advanced data assimilation schemes. With the availability to high performance computing (HPC) the NWP models now feature very high horizontal and vertical resolution. The improved accuracy of forecasting in recent years is mainly attributed to improved (*i*) forecast model (resolution and physics), (*ii*) data assimilation and (*iii*) observations. Although these recent advances have reflected in the improved accuracy of forecast in the NWP models, accurate prediction of extreme events is still a challenge. Extreme events like the '*very heavy*' or '*extremely heavy*' or the '*exceptionally heavy rain*' often occur over small and isolated regions which can be captured using high resolution models. Verification of such extremes is a greater challenge since they are rare, occur over isolated locations and at times go unreported. On the other hand the high resolution NWP models have biases and tend to produce rather too many of such heavy rain events.

Assessing the quality of predictions of extreme weather events, however, is complicated by the fact that measures of forecast quality typically degenerate to trivial values as the rarity of the predicted event increases (Ferro and Stephenson, 2011). Extreme Value Theory (EVT) provides a robust, well established framework for extreme value analysis. It studies the properties of extreme values, and enables them to be fit with theoretical distributions (or *probability models*). Such theoretical distributions enable one to describe the behaviour of extremes
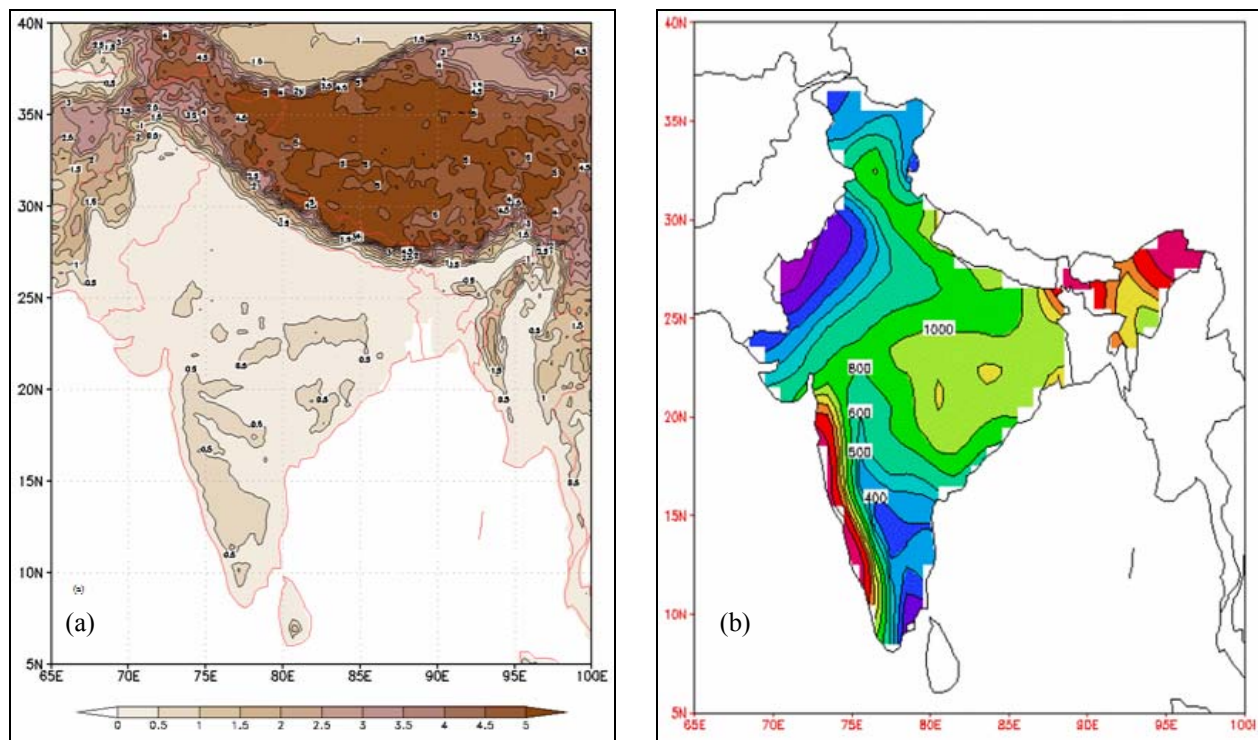
Figs. 1(a&b).   (a) Geographical domain of analysis with terrain height (km) over India (b) Climatological normal (1951-2003) monsoon (June-September) accumulated rainfall (mm/season)

through the estimation of few key parameters. Stephenson *et al.* (2008) proposed a new verification measure, the Extreme Dependency Score or EDS, for summarizing the performance of deterministic forecasts of rare binary events. Instead of degenerating, the EDS values converge to a meaningful limit for rare events. Further, the EDS is found to have several drawbacks, (being susceptible to hedging by over forecasting and being base-rate dependent) as discussed in Ferro and Stephenson (2011) where in an improved measure Symmetric EDS (SEDS), Extreme Dependence Index (EDI) and Symmetric Extremal Dependence Index (SEDI) are proposed as improved family of score for verification of extreme and rare events.

This study presents rainfall verification over India using traditional verification scores such as Probability of Detection (POD), Critical Success Index (CSI) and Equitable Threat Score (ETS) for various rainfall categories. Further, the statistical challenges associated with the verification of the extreme events are discussed. The new sets of scores are briefly reviewed (EDS, SEDS, EDI and SEDI) and are used for verification to assess the impact on the verification of extreme rain amounts.

2.   Data and methodology

2.1.   *Observed and model forecast rainfall data*

(a)   *Observed rainfall data over India*

Figs. 1(a&b) shows the geographical domain chosen for the present study, 7°-38.5° N, 67°-100.5° E. Rainfall analysis based on quality controlled observations is very useful and critical for verification of the NWP forecasts. In this study we use two observation data sets, (a) the Tropical Rainfall Measuring Mission (TRMM) 3B42 (V7) daily multi-sensor 0.25° × 0.25° gridded rainfall and (b) the India Meteorological Department (IMD) and NCMRWF merged satellite + gauge gridded rainfall data at 1°, denoted NSGM.

TRMM rainfall data has biases over the Indian land regions which require correction (Mitra *et al*., 2009; Chen *et al*., 2013). The NSGM objectively analyses IMD daily raingauge observations onto a 1° grid using a successive corrections technique with the TRMM 3B42 satellite precipitation providing the first guess field, thus, providing spatially continuous rainfall over land and

TABLE 2

Contingency table representing the frequencies of forecast-observation pairs for which the event and non-event were forecasted and observed

		Observed		Total
		Yes	No	
Forecast	Yes	Hits	False alarms	Forecast yes
	No	missed	Correct negatives	Forecast no
	Total	Observed yes	Observed no	total

TABLE 3

List of scores used for evaluation of categorical rainfall forecasts

$$POD = \frac{hits}{hits + misses} \quad \text{(also known as hit rate } H\text{)}$$

$$TS = \frac{hits}{hits + misses + false\ alarms} \quad \text{(also called Critical Success Index-CSI)}$$

$$ETS = \frac{hits - hits_{random}}{hits + misses + false\ alarms - hits_{random}}$$

$$hits_{random} = \frac{(hits + misses)(hits + false\ alarms)}{total}$$

$$OR = \frac{hits + correct\ negatives}{misses + false\ alarms} = \frac{\left(\dfrac{POD}{1 - POD}\right)}{\left(\dfrac{POFD}{1 - POFD}\right)}$$

$$EDS = \frac{2 \log\left(\dfrac{hits + misses}{total}\right)}{\log\left(\dfrac{hits}{total}\right)} - 1 = \frac{\ln p - \ln H}{\ln p + \ln H} \qquad H\text{-hitrate}$$

$$SEDS = \frac{\ln\left(\dfrac{hits_{random}}{hits}\right)}{\ln\left(\dfrac{hits}{total}\right)} = \frac{\ln q - \ln H}{\ln p + \ln H} \qquad H\text{-hitrate}$$

$$EDI = \frac{\ln F - \ln H}{\ln F + \ln H} \qquad H\text{-hitrate, } F\text{-False alarm rate}$$

$$SEDI = \frac{\ln F - \ln H + \ln (1 - H) - \ln (1 - F)}{\ln F + \ln H + \ln (1 - H) + \ln (1 - F)}$$

$$POFD = \frac{False\ alarms}{correct\ negatives + false\ alarms} \quad \text{(also known as false alarm rate } F\text{)}$$

where, p = (hits + misses)/total is the base rate (climatology), q = (hits + false alarms)/total is the frequency with which the event is forecast, H is the hit rate and F is the false alarm rate, also known as the probability of false detection
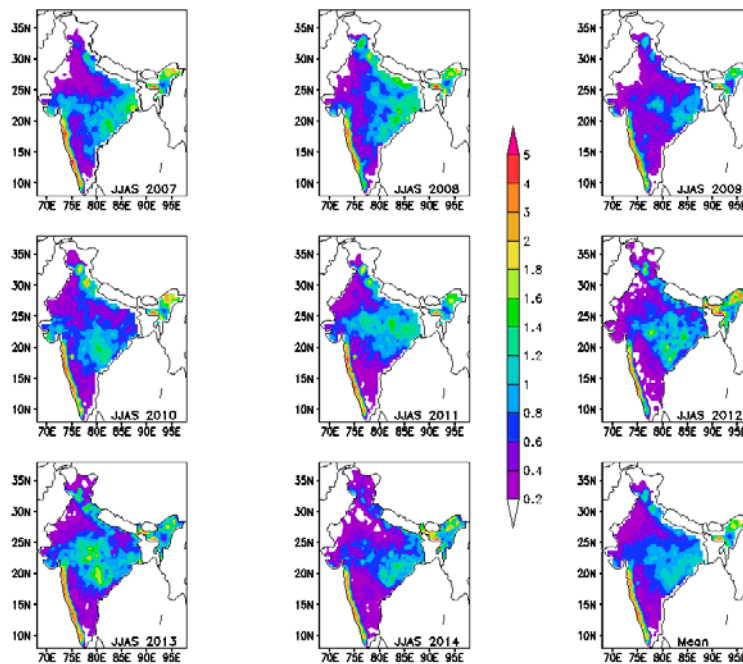
Fig. 2. Observed seasonal (JJAS) mean rainfall (cm/day) over India from 2007-2014
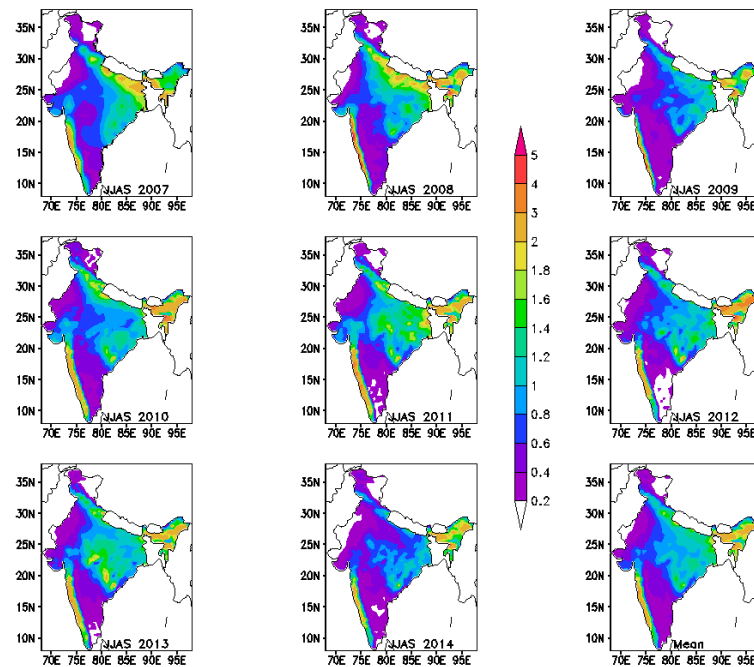

Fig. 3. UKMO Day-1 forecast seasonal (JJAS) mean rainfall (cm/day) over India from 2007-2014

ocean. As noted by Mitra *et al*. (2009), the 1° grid resolution is appropriate for capturing the large scale rain features associated with the monsoon. The merging of the IMD gauge data into TRMM 3B42 not only corrects the mean biases in the satellite estimates but also improves the large-scale spatial patterns in the satellite field, which is affected by temporal sampling errors (Mitra *et al*., 2009).

(b) *Model forecast rainfall over India*

    The Unified Model is the numerical modelling system developed and used at the U. K. Met Office and will be denoted UKMO in this paper. In this 'seamless' prediction system different configurations of the same model are used across all time and space scales, with each
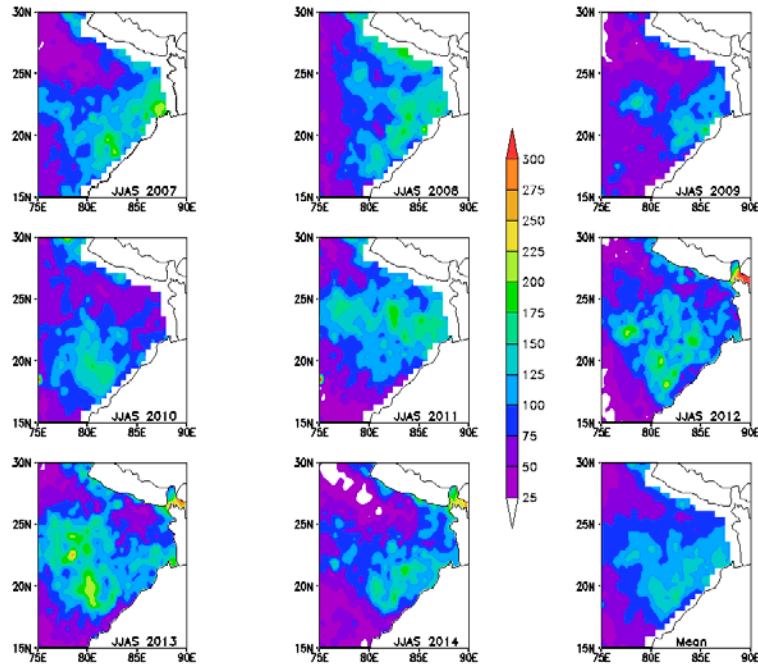
Fig. 4. Observed seasonal (JJAS) accumulated rainfall (cm/season) over eastern India from 2007-2014
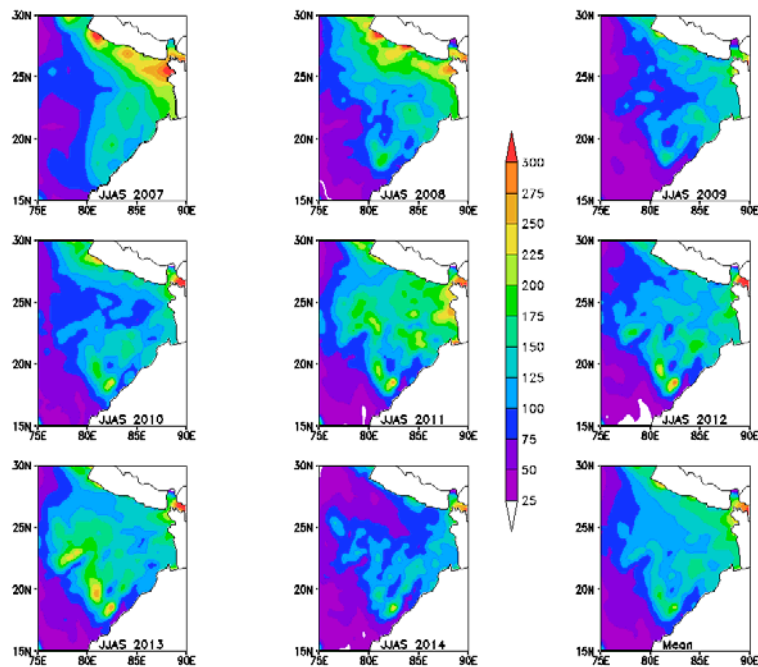

Fig. 5. UKMO forecast (Day-1) seasonal (JJAS) accumulated rainfall (cm/season) over eastern India from 2007-2014 and Mean

configuration designed to best represent the processes which have most influence on the timescale of interest. The atmospheric model uses non-hydrostatic dynamics with semi-Lagrangian advection and semi-implicit time stepping. It is a grid point model with the ability to run with a rotated pole and variable horizontal grid. A number

of sub-grid scale processes are represented, including convection, boundary layer turbulence, radiation, cloud, microphysics and orographic drag. During 2007-2014 the horizontal and vertical resolution of the global configuration improved from about 40 km and 50 levels in 2007 to about 25 km and 70 levels in 2010.
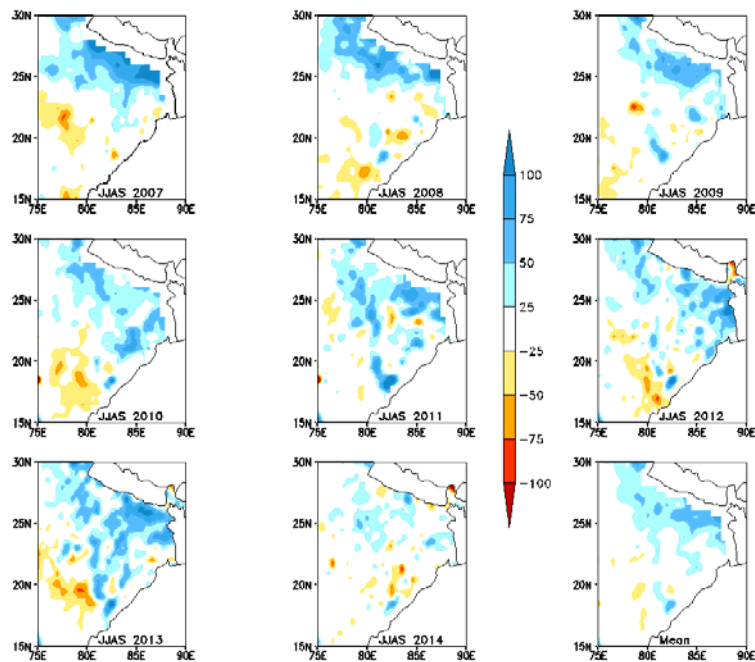
Fig. 6. Accumulated rainfall difference (Forecast-Observed) for Day-1 forecast over eastern India
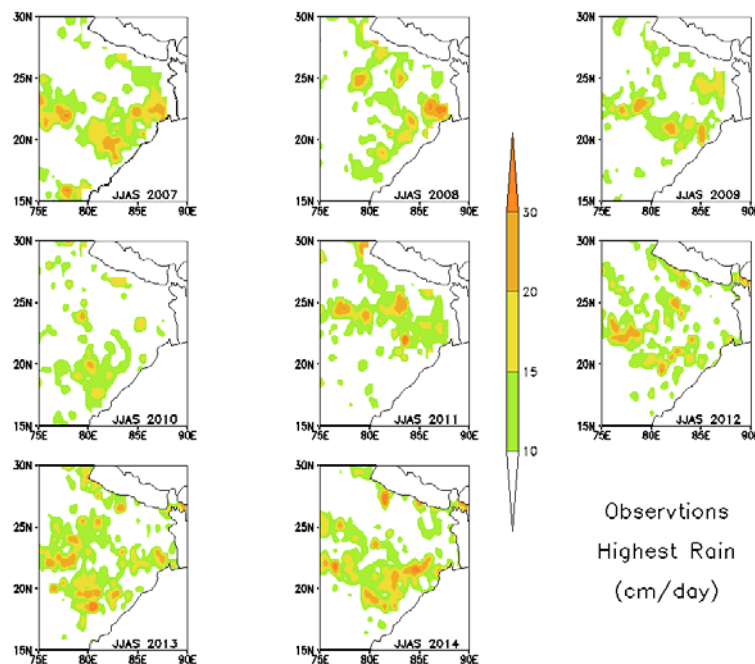

Fig. 7. Observed highest one day rainfall of the season over eastern India

This study uses the rainfall forecasts over India from the operational global forecast model configuration. The verification is presented for Day-1 forecasts. The model forecasts are interpolated to a common grid resolution of 0.5° × 0.5° for verification. The verification of 24 hour accumulated rainfall during the monsoon seasons is based on eight seasons (976 days).

2.2. *Verification methodology*

The forecast daily rainfall fields are first verified using standard categorical verification scores. The contingency table is a useful way to see what types of errors are being made (Table 2). A perfect forecast system would produce only *hits* and *correct negatives* and
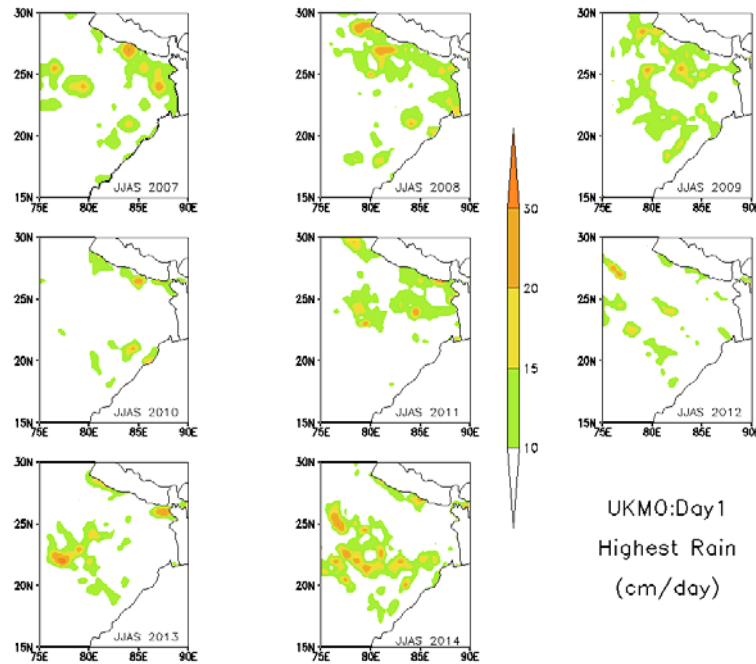
Fig. 8. UKMO forecast (Day-1) highest one day rainfall of the season over eastern India

no *misses* or *false alarms*. A large variety of *categorical statistics* are computed from the elements in the contingency table to describe particular aspects of forecast performance. Jolliffe and Stephenson (2012) and Wilks (2011) provide detailed descriptions of these scores. Table 3 provides list of some of the scores used in this study.

The verification is based on the observations and forecasts from 976 days for 8 seasons (2007-2014). The robustness and significance of the verification scores established using the bootstrap estimation of 95% confidence interval for each of the scores.

3.   Results and discussion

3.1. *Observed and forecast mean and accumulated rainfall*

A comparison of the observed and forecast mean JJAS rainfall is shown in Figs. 2 and 3 during each of the eight years (2007-2014) and an average of eight years. The observed mean rainfall in each of the season is compared against the Day-1 forecasts. This comparison is presented to highlight the systematic biases in the Day-1 rainfall forecast. The forecasts show dry (*wet*) bias over the peninsula (Himalayas). The mean seasonal rainfall over India exhibits large spatial variability. Rainfall along the west coast of the peninsula and over northeast India exceeds 2 cm/day under the influence of orography. On

the other hand, rainfall amounts over northwestern India and eastern peninsula is lower than 0.4 cm/day. The forecast means (Fig. 3) also successfully capture the observed spatial variations (Fig. 2). The focus of this study is the large region over the plains in eastern India (Figs. 4 & 5). Rainfall over this region mainly comes from the monsoon trough and from the Bay of Bengal low pressure systems independent of the orographic influences. The rainfall over this region forms the core monsoon rainfall. The observed accumulated rainfall over this region exceeds 100 cm/season. Forecasts successfully capture this feature. However, the model forecasts have higher rainfall amounts excess of 250 cm/season which are not evident in the observations. The difference (forecast-observed) of accumulated rainfall is shown in Fig. 6. The forecasts show rainfall amounts higher than observations by over 50 to 100 cm/season over Gangetic Plains. This prominent in each of the years and is also reflected in the mean difference. It is pertinent to note that this region is affected by the Bay of Bengal low pressure systems during the monsoon. There are year to year wide variations in the rainfall peak amounts as well as spatial coverage. This can be mainly attributed to varying number of low pressure systems from year to year.

The observed and forecast highest rainfall of the season at each grid is shown in Figs. 7 and 8. Cases exceeding 10 cm/day are shown. Fig. 7 shows the rainfall hot-spots during each season along with the areal extent over the region. The model forecasts underestimate the
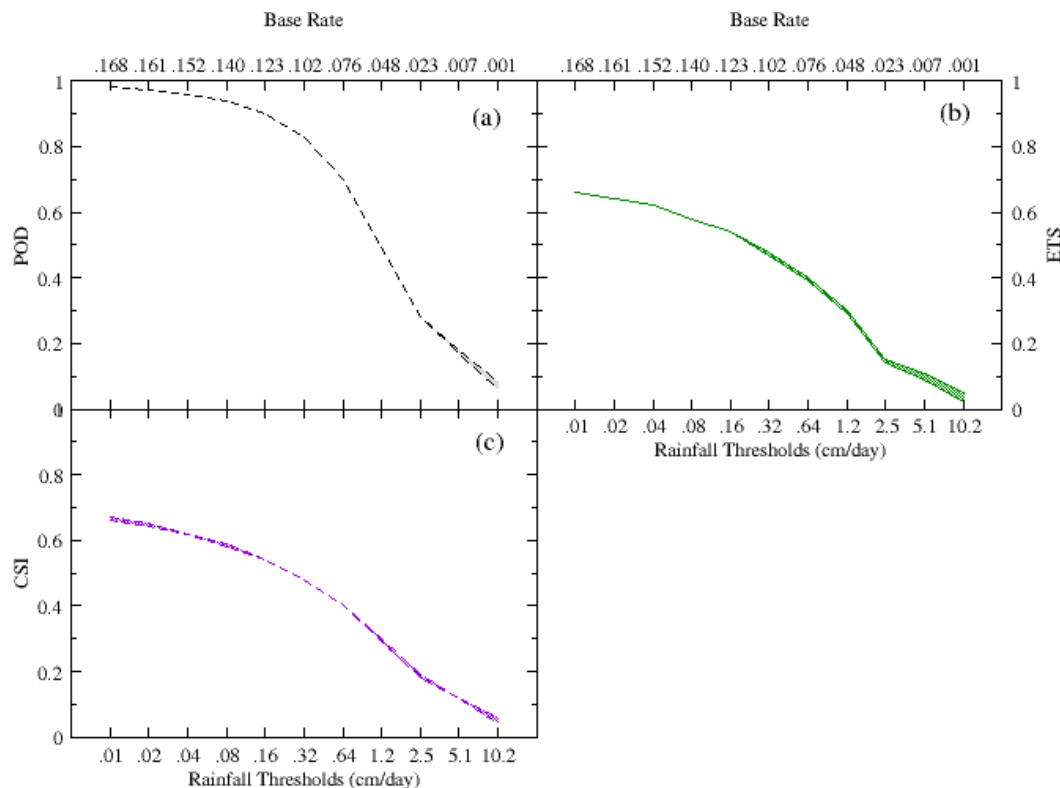
Figs. 9(a-c). (a) Probability of detection (POD), (b) Equitable threat Score (ETS) and (c) Critical Success Index (CSI) computed for Day-1 rainfall forecasts for eight monsoon seasons (2007-2014). The shaded area is bounded by scores significant at 95% CI. (Base rate is indicated on the secondary *x*-axis on top panels of a and b

observed rainfall peaks and this is also reflected in the spatial extent of the coverage. Thus, on the one hand the accumulated rainfall in Figs. 4 and 5 suggest that the forecasts *overestimate* the rainfall amounts; while the isolated rainfall peaks (Figs. 7 and 8) suggest that the forecasts *underestimate* the peak rainfall amounts. This has implication on the forecast rainfall frequency and skill. In the following section forecastsare evaluated with special emphasis on the high rainfall values (>10 cm/day). The verification scores are based on the observed and forecast rainfall at corresponding grids in the domain over eastern India shown in Figs. 4-8. The verification statistics are computed based on data from all eight seasons (976 days) for a 31 × 31 grid.The sample includes close to 50000 (14000 and 2366) cases of rainfall ≥2.5 cm (5.1 cm and 10.2 cm). The sample size used for this analysis is indeed impressive. The verification scores are presented with bootstrap estimation of 95% confidence intervals for the scores.

3.2. *Verification of rainfall forecasts*

Probability of Detection (POD), Critical Success Index (CSI) and Equitable Threat Score (ETS) form some

of the standard verification measures widely used in rainfall forecast verification. The definitions of these scores are listed in Table 3. Panels in Figs. 9(a-c) show these scores for Day-1 forecast rainfall. POD gives the correctly predicted fraction of observed 'yes' events while CSI gives the degree correspondence between the observed and forecast 'yes' events. Similar to CSI, ETS gives the degree of correspondence between the observed and forecast 'yes' events, after accounting for random hits. All these scores range from 0 to 1 (-1/3 to 1 for ETS); with value of 0 meaning no skill and a value of 1 meaning perfect forecast. While POD is a good measure for rare events, it is sensitive to hits and ignores false alarms. It is also very sensitive to climatological frequency of events. CSI on the other hand is not considered all that good for rare events since it is concerned with forecasts that count. While it penalizes misses and false alarms, it ignores the correct negatives. It is also sensitive to climatological frequency of the events. It can be seen from Figs. 9(a-c) that the values of ETS, POD and CSI for rainfall amounts of under 2 cm/day suggest relatively better accuracy. However, for rainfall amounts greater than 2cm the value of scores decrease diminishes drastically. The base rate is indicated in the secondary *x*-axis shown at the top of the
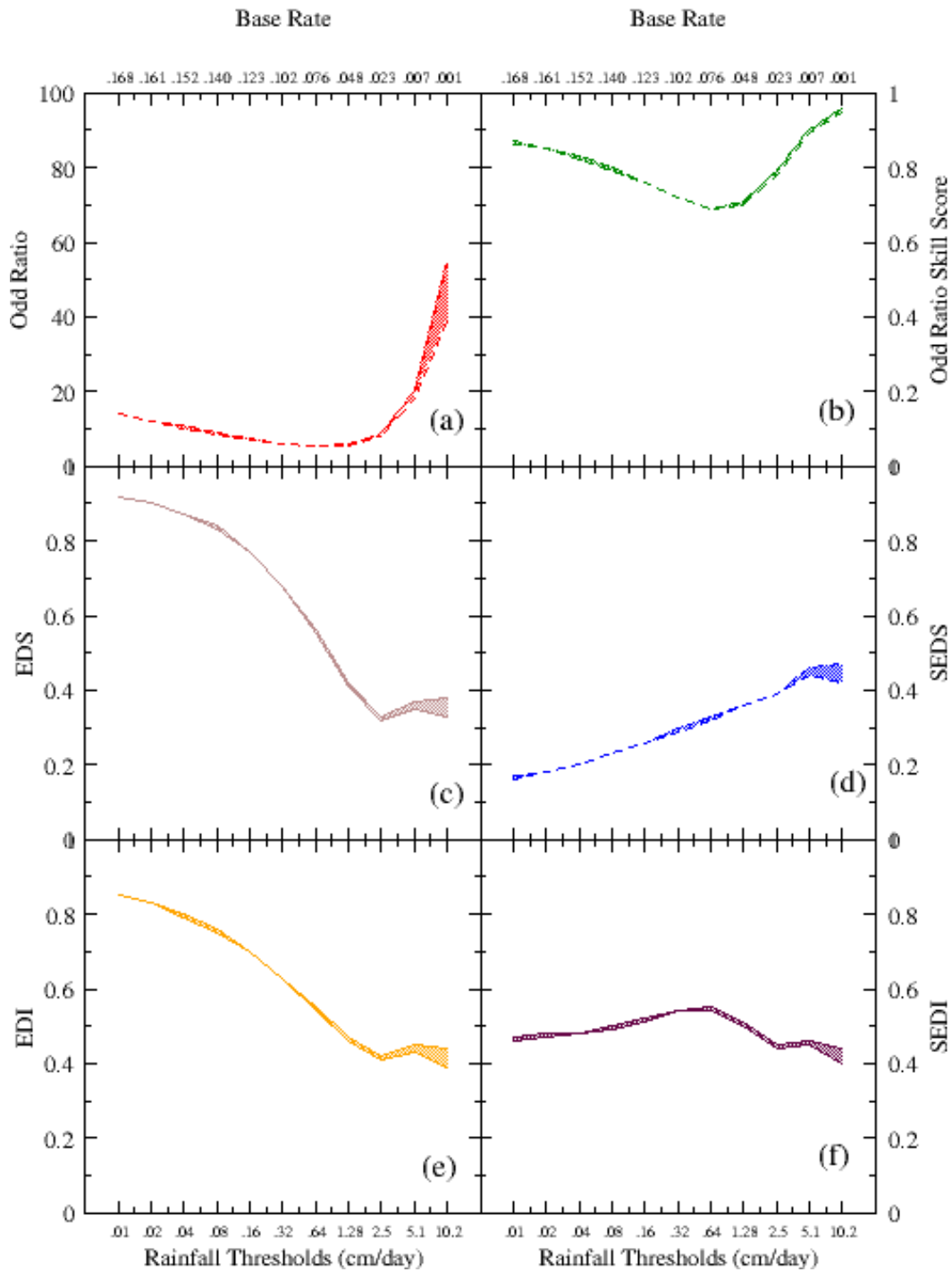
Figs. 10(a-f). Verification scores for Day-1 Rainfall forecasts over eastern India based on eight monsoon seasons (2007-2014).
(a) Odds Ratio (OR) (b) Odds Ratio Skill Score (ORSS) (c) Extreme Dependency Score (EDS) (d) Symmetric Extreme Dependency Score (e) Extremal Dependence Index (EDI) (f) Symmetric Extremal Dependence Index (SEDI). The shaded area is bounded by scores significant at 95% CI. (Base rate is indicated on the secondary *x*-axis on top panels of a and b

panels of 9 (a and b) in Figs. 9(a-c). The asymptotic nature of these scores makes it difficult for a forecaster to judge

on the quality of the forecasts or relative performance of the different models for higher rainfall thresholds.

The Odds Ratio (OR) approaches infinity for high rainfall threshold (as the event becomes rare). Odds Ratio gives the ratio of the odds of a "yes" forecast being correct, to the odds of a "yes" forecast being wrong. It ranges from 0 to ∞ with a value of 1 indicating no skill and ∞ indicating perfect forecast. This score provides better measure for rarer events and is less sensitive to hedging. It can also be expressed in terms of Odds Ratio Skill Score (ORSS) which is also known as Yule's Q. This score indicates improvement of the forecast over random chance. It ranges from -1 to 1; 0 indicating no skill and 1 denoting perfect forecast. Figs. 10 (a&b) show OR and ORSS for different rainfall thresholds. Both OR and ORSS give better scores for rare events, less sensitive to hedging and are independent of base rate.

Figs. 10 (c-f) show Extreme Dependency Score (EDS) family of scores. EDS Symmetric (SEDS), Extremal Dependence Index (EDI) and Symmetric Extremal Dependence (SEDI), can be collectively called EDS family of scores. These scores measure association between the observed and forecast rare events. They range from -1 to 1 with 0 meaning no skill and 1 indicating perfect score. Though EDS does not approach zero, it has several undesirable properties like it is base-rate dependent, sensitive to hedging, varies from -1 to 1 etc. EDI and SEDI overcome most of the drawbacks since they have non-degenerate limit, are base-rate independent, insensitive to hedging etc Ferro and Stephenson (2011). As can be seen in Figs. 10(a-f), for higher rainfall amounts, these scores do not converge to trivial values. Further, these scores allow one to examine the relative difference in the forecast accuracy. Both EDS and SEDS seem to form improvement over the scores discussed in Figs. 9(a-c). EDS and SEDS tend to have large difference for lower rainfall thresholds (<2 cm/day) while for high rainfall thresholds (>5 cm/day) they tend to both tend to show some variability. However, as discussed in Ferro and Stephenson (2011) these scores are sensitive to base rate and hedging. Thus EDI and SEDI make very useful candidates for forecast verification and model inter comparison for extreme rainfall forecasts. Similar to EDS and SEDS, both these scores also show large difference for rainfall ≤ 2 cm and attain comparable magnitudes for rainfall > 2 cm/day.

Standard traditional verification scores could be used to monitor the forecast performance and model improvements. However for impact of model improvements like increased grid or vertical resolution on the extreme rainfall amounts is often obscured due to sample size and standard scores which converge to trivial values. This can be effectively overcome by appropriate use of the OR, EDI and SEDI.

4. Summary

This study summarizes the results of the rainfall forecast verification over eastern Indian region using the Day-1 rainfall forecasts from eight monsoon seasons. The verification shows:

(*i*) Rainfall wet bias over Gangetic plains in each of the years and also averaged over eight years.

(*ii*) However, the extreme rain in monsoon over the plains forms a significant portion of the seasonal total rainfall. Although the forecasts indicated that the models *overestimate* the accumulated seasonal rainfall, they *underestimate* the highest one day rainfall amounts and spatial extent.

(*iii*) Verification of the rainfall forecasts using the standard skill scores such as ETS and CSI show good skill in the forecasts for rainfall thresholds less than 2 cm/day. These values degenerate to vanishingly low values for rainfall thresholds >8 or 10 cm/day. For a data sample that includes close to 50000 (14000 and 2366) cases of rainfall ≥2.5 cm (5.1 cm and 10.2 cm) a better measure of the performance at the tails is essential.

(*iv*) Verification using Odds Ratio (OR) and Odds Ratio Skill Score (ORSS) are also presented. These scores demonstrate meaningful values for the heavy rainfall thresholds. Additionally the EDS family of scores are also tested and presented. EDI and SEDI tend to show very similar pattern for rainfall amounts >5 cm or so. EDI on the one hand degenerates very similar to many of the standard scores like ETS for low rainfall thresholds. However both EDI and SEDI indicate converging to meaningful values (0.45 in the present case.)

Acknowledgements

References

Ashrit, R. and Mohandas, Saji, 2010, "Mesoscale model forecast verification during monsoon 2008", *J. Earth Syst. Sci*., **119**, 4, 417-446.

Chen, Y., Ebert, E. E., Walsh, K. J. E. and Davidson, N. E., 2013, "Evaluation of TRMM 3B42 daily precipitation estimates of tropical cyclone rainfall over Australia", *J. Geophys. Res*., **118**, 1-13.

Das, S., Ashrit, R., Iyengar, G. R., Mohandas, Saji, Das Gupta, M., George, John P., Rajagopal, E. N. and Dutta, Surya Kanti, 2008, "Skill of different mesoscale models over Indian region during monsoon season: Forecast errors", *J. Earth Syst., Sci.*, **117**, 5, 603-620.

Durai, V. R. and Bharadwaj, Rashmi, 2013, "Improving precipitation forecast skill over India usinf a multi-model ensemble technique", *Geofizika*, **30,** 2, 119-141.

Ferro, C. A. T. and Stephenson, D. B., 2011, "Extremal Dependence Indices: improved verification measures for deterministic forecasts of rare binary events", *Wea. forecasting*, **26**, 699-713.

Goswami B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S. and Xavier, Prince K., 2006, "Increasing trend of extreme rain events over India in a warming environment", *Science*, **314**, 1442-1445.

Jolliffe, I. T. and Stephenson, D. B., 2011, "Introduction, in Forecast Verification: A Practitioner's Guide in Atmospheric Science", Second Edition (eds. I. T. Jolliffe and D. B. Stephenson), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781119960003.

Kalnay, E., Lord, S. J. and McPherson, R. D., 1998, "Maturity of Operational Numerical Weather Prediction: Medium Range", *Bull. Amer.Meteor. Soc.*, **79**, 12, 2753-2769.

Keshavamurty, R. N., 1973, "Power spectra of large-scale disturbancesof the Indian southwest monsoon", *Ind. J. Meteor. Geophys.*, **24,**117-124.

Magnusson, L. and Kallen, Erland, 2013, "Factors influencing skill improvements in the ECMWF forecasting system", *Mon. Wea. Rev.*, **141**, 3142-3153.

Mandal, V., De, U. K. and Basu, B. K., 2007, "Precipitation Forecast Verification of the Indian Summer Monsoon with intercomparison of three diverse regions", *Weather and Forecasting*, **22**, 428-443.

Mitra, A. K., Bohra, A. K., Rajeevan, M. N. and Krishnamurti, T. N., 2009, "Daily Indian precipitation analyses formed from a merged of rain-gauge with TRMM TMPA satellite derived rainfall estimates", *J. of Met. Soc. of Japan*, **87**A, 265-279.

Roy, S. S. and Balling, R., 2004, "Trends in extreme daily precipitation indices in India", *Int. J. Climatol.*, **24**, 457-466.

Sikka, D. R., 1977, "Some aspects of the life history, structure and movement of monsoon depressions", *Pure Appl. Geophys.*, **115**, 1501-1529.

Stephenson, D. B., Casati, B., Ferro, C. A. T. and Wilson, C. A., 2008, "The extreme dependency score: A non-vanishing measure for forecasts of rare events", *Meteorol. Appl.*, **15**, 41-50.

Vitart, Frederic, 2013, "Evaluation of ECMWF sub-seasonal forecast skill scores over the past 10 years", *Technical Memorandum*, p694, ECMWF.

Walters, D., Wood, Nigel, Vosper, Simon and Milton, Sean, 2014, "End Game: A new dynamical core for seamless atmospheric prediction", Met Office.

Wilks, D. S., 2011, "Statistical Methods in Atmospheric Sciences", *International Geophysical Series*, **100**, Academic Press.