Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-65-AC2, 2019

Geoscientific Model Development Discussions

Open Access

EGU

Interactive comment on "Evaluation of Unified Model Rainfall Forecasts over the Western Ghats and North East states of India" *by* Kuldeep Sharma et al.

Kuldeep Sharma et al.

kuldeep@ncmrwf.gov.in

Received and published: 5 August 2019

General Comments: 1. This paper presents an evaluation of general model advances which have occured in the Met Office UM since 2007, in the context of summer rainfall over India, and so the subject area is within the scope of GMD and EGU. The paper does not present any novel modelling or verification tools, nor does it report on any new model experiments. The novelty is rather in bringing together recently-developed, as well as more commonly-used, verification metrics to assess the UM in the specific context of rainfall over mountainous regions of India, using recently-available observations.

2. The overall conclusion of the paper, that UM rainfall forecasts have improved steadily over the last 12 years, for two key regions of India, is a substantial result. However, the authors provide no discussion as to how this result could have come about for reasons other than improvements in the UM. For example, it may be that different years have been easier or more difficult to forecast: to conduct a thorough evaluation, it would be necessary to run different versions of the UM for the same year and compare the forecasts. Another approach, using only the existing forecasts, might be to see whether improvements across years between which an upgrade has been made are larger than improvements across years between which an upgrade has not been made.

Reply to General comment 1 & 2: The authors thank the reviewer's observation regarding the scope and novelty in the approach. The authors agree with the suggestion given by the reviewer and detailed set of experiments using a frozen version of the model needs to be done to carry out the sensitivity studies. This can be considered a separate study. The discussion presented in this paper forms a background to carry out such sensitivity studies. Following the suggestion (in the context of another approach), we have found that during 2011-2013, when there was no major modeling upgrades, the SEDI over NE-states (90th percentile threshold) fluctuated between 0.45 to 0.49 which crossed 0.5 and reached to 0.53 in 2015.

3. The methods are quite clearly described, although the authors should make clear exactly which data were used to define the percentiles (e.g. was the same dataset used for both observations and forecasts, or were the observations and forecasts assigned to percentiles separately, based on their respective dataset, so that the percentiles for any given observation/forecast pair correspond to different absolute thresholds?). In terms of reproducibility of the results, the authors provide codes and data (although these could be more clearly documented, perhaps even to the extent of saying which code and data are used to produce which figure in the paper). I haven't tried to apply these, but I think that the methods and data are sufficiently clearly described in the paper that one could in principle reproduce the results from this.

Printer-friendly version

Discussion paper

Reply to General Comment 3. The authors wish to clarify that the observations and forecasts use their own dataset to calculate the percentiles (different absolute threshold).

4. The initial review brought up the issue that the 0.5 degree grid was not of sufficiently fine resolution to allow a full assessment of the forecasts in terms of orographic processes, and the authors do not seem to have addressed this. It should at least be mentioned that there are some processes that are not captured by the grid being used. On the other hand, it probably is an appropriate grid for assessing the UM (being a few times coarser than the UM grid spacing, so corresponding to the UM's "effective" resolution) and could still be appropriate for assessing larger-scale orographic processes, and the upscale effects of smaller-scale processes. Is it possible to look at the verification scores as a function of horizontal location? Then it might be possible to determine how the improvement in the forecast (as well as the quality itself of the forecasts) varies with the steepness of the orography. Going in the opposite direction, it may be worthwhile to apply neighborhood-based verification metrics (e.g. Fractional Skill Score) to determine how the improvements vary with scale going to coarser scales.

Reply to General Comment 4. The authors are very much thankful about the feedback of choosing the horizontal resolution of the model. The text now included the comment on how model successfully captures larger-scale orographic processes, and the upscale effects of smaller-scale processes in the section 4.1, even when assessed at 0.5x0.5 grid resolution.

To assess the skill as a function of grid resolution, we will require some of the fuzzy verification methods (Fraction Skill Score, Scale decomposition etc). But these methods are often effective while working with very high resolution (1x1 km grid) observation and forecast hence this has not been attempted.

The exercise of computing the verification scores as a function of horizontal scale is beyond the scope of this paper.

5. As far as I am aware the authors give proper credit to previous work. An issue was raised in the initial review that two previous papers were very similar to this work. The main innovation of the current paper is that it looks at percentile-based thresholds, and that it looks at more mountainous areas of India, and this is clear from the Introduction. I do wonder, however, if it would be worth comparing the percentile and absolute threshold scores directly, since the latter did not show such a pronounced improvement in the previous paper. Similarly, it may be worth applying the methods of the current paper to more regions over India.

Reply to General Comment 5. The authors are thankful for the comments. Comparison of absolute thresholds (5 cm/day which was taken in my previous paper, Sharma et al 2017) with percentile based thresholds would be tricky since even within small domain chosen for study, absolute amounts corresponding to 90th percentile vary. Still a comparison of CSI from 2007-2018 in Day-1 forecast over NE-states is provided (Fig 1 , this no is to only for this response). CSI computed at 5cm/day threshold also shows the increasing trend from 2007 to 2018.

6. The paper is well presented, with an appropriate title, abstract, amount of figures, supplementary material and use of mathematics. It would, however, benefit from some editing for the English: it is generally always possible to ascertain the meaning of the text, but some improvements would make it easier to read (particularly for non-native speakers of English) and therefore increase its impact. It would also be good to see Figures 2 and 3 for the other years (probably in supplementary material), and some discussion as to whether they corroborate the authors' conclusions with respect to these Figures. This is touched on in lines 227-229, where the authors say that no significant change is seen in 2007-2012, but nothing is mentioned about 2014, 2016 and 2017.

Reply to General Comment 6: Thank you for your feedback about the presentation of the paper. As the reviewer suggested, we have included the supplementary figures (S3-S8). Figure S6 and S7 shows that model hardly produced the peak amounts of rainfall during the season. Also Figure S8 shows, model shows an improvement in capturing the maximum rainfall (season's highest rainfall) that after 2013.

Specific Comments: 1.Lines 107-130: Are the satellite products not used at all for 2007-2011? This is not clear from the text. Additionally, using different observational datasets for different periods could contribute to the change in verification scores seen over different years. Is it possible to apply all three observation datasets at least to 2016-2018, to see if the scores change significantly with different datasets?

Reply to specific Comment 1: Thank you for your feedback. The authors wish to clarify that we have utilized only one data source (IMD-NCMRWF merged rainfall data). The text has been modified to reflect this. "The observed rainfall data is the IMD-NCMRWF merged rainfall product (0.5 x 0.5grid). This rainfall analysis is based on merging of gauges measurements with satellite based rainfall estimates (TRMM: Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis (TMPA)-3B42 and GPM: Global Precipitation Measurement). The satellite rainfall estimates are based on TRMM during 2007-2015 and on GPM from 2016. This merged data set represents the Indian monsoon rainfall realistically and is superior to other available rainfall data sets over the Indian monsoon region (Mitra et al. 2013, Reddy et al 2019). " it must be noted that this study uses the best possible rainfall product available for the entire period over Indian region. Evaluation using one or two most recent monsoon seasons will have very small sample size. GPM derived IMERG can't be used directly since it has biases in rainfall estimation (Reddy et al 2019). Also over mountains this rainfall product shows large biases. (Krishna et al 2017). Since TRMM observations are not available from 2016 . So, this exercise is also not possible at this stage.

2.Lines 190-191: As mentioned in the General Comments, are these percentiles with respect to observations or forecasts? Lines 182-184 suggests that each of observations and forecasts use their own dataset to calculate the percentiles (so that the absolute thresholds are different). But this would then mean that the number of observed and forecast events is the same, so a+b=a+c and b=c. This means that BIAS=1, FAR=1- POD and CSI=POD/(2-POD). This seems to be the case from Figures 6 and 7: the quantities FAR, POD and CSI are all fairly simply related to each other and BIAS is always very close to 1. It may therefore be sufficient to only report one of these four quantities.

Reply to specific Comment 2: Yes, observations and forecasts use their own dataset to calculate the percentiles (different absolute threshold). Since the number and observed forecast events is same , BIAS close to 1, plot is avoided. However, FAR, POD and CSI are retained.

3.Line 200: Since the SEDI is a relatively new technique, it may be worth including the Ferro & Stephenson reference here as well.

Reply to specific Comment 3: The reference has been added in the text as well as in the reference list.

4.Lines 237-239: It would be useful to see a value for how the bias has changed overall for each of the two regions. This can be ascertained to some extent from the figures, but not quantitatively.

Reply to specific Comment 4: One table (Table 4) has been introduced and also text has been modified as " The mean error in 2013, 2015 and 2018 over WGs and NE-states are displayed in Table 4."

5.Lines 258-264: Do you have a reason why you think the improvement in mean rainfall and highest rainfall is linked to the specific upgrades you state? It is of course likely to be the case: indeed, you could probably just change "linked" to "likely to be linked" and this would be fine.

Reply to specific Comment 5: Thank you very much for this comment. The text has been modified as suggested.

6. Lines 286-296: See comment above for lines 190-191. Given an improvement in POD you are guaranteed to get an improvement in FAR, if you use the percentile

method where you are excluding bias effects.

Reply to specific Comment 6: Thank you for your observation. When we consider percentile based threshold, BIAS becomes 1. Hence, BIAS is implicitly removed.

7.Lines 317-321: Is it worth trying even higher thresholds? The more traditional methods seem to work quite well for 80% and 90%; is not the point of the SEDI to assess such extreme thresholds that insufficient counts exceeding the threshold are available to usefully apply the traditional methods?

Reply to specific Comment 7: Thank you for your feedback on this. As the reviewer has pointed out the use of SEDI for higher thresholds, we have computed SEDI for 95th and 99th percentile threshold also. We found the same increasing trend. The figures show the SEDI for Day-1, Day-2 and Day-3 forecasts for 95th and 99th percentiles over WGs (Fig 2, this no is to only for this response) and NE-states (Fig 3, this no is to only for this response).

8.Line 324: Does this sentence refer to the improvement from 2007-2008? This improvement in the data assimilation is not listed in Table 1.

Reply to specific Comment 8: Thanks for pointing this out. We have now included it in the Table 1. The revised version of Table 1 has now new Table 1.

9.Lines 346-352: As mentioned earlier, please provide plots for other years in the Supplementary Material to support these conclusions.

Reply to specific Comment 9: The plots for earlier monsoon seasons from 2007-2012 has been included in the supplementary material which also confirms the consistent wet bias over Indo-Gangetic plains.

10.Table 1: It would also be useful to list the different UM configurations used (e.g. GA6.1). Figure 4: For the NE states, the numbers of counts in more recent years are clearly higher than those in earlier years. Could this be to do with the different datasets used in different years?

Reply to specific Comment 10: The different UM configurations has been added in Table 1. As clarified earlier, there are no different data sets. It is the merged (satellite +gauge) rainfall product. These higher counts in some years may be due to the interannual variability.

Technical Corrections:

1.Line 112: The two numbers should be multiplied by 122 (based on text later in the manuscript). Reply to Technical Correction 1:

As indicated by the reviewer, The numbers now have been multiplied by 122. Also, The text has been modified accordingly.

2.Lines 162-163: Some earlier years of Met Office operational forecasts used an earlier version (e.g. GA3.1). I think the current version used operationally is GA6.1 but please check this. Reply to Technical Correction 2:

Thank you for pointing out. Yes, It is GA6.1 and it has been modified in the text also.

3.Lines 256-257: "The number of forecast counts is too high over ...". Currently the text implies that the number of counts increases each year.

Reply to Technical Correction 3: The text has been modified as suggested.

4.Line 317: "Figures 8 and 9".

Reply to Technical Correction 4: Thank you for pointing out. The text has been modified

5.Figure 2: Please make it clear that, for the right panels (c,f,i), you are subtracting the observations from the forecast.

Reply to Technical Correction 5: Yes, the mean error is computed by subtracting observation from the forecasts. The formula has been added in the Table 3.

6.Figure 4 caption: "... forecast of rainfall above a threshold of 10cm/day ..."?

Reply to Technical Correction 6: This figure now has changed to figure 5 and the caption has been modified as suggested.

7.Figures 6-9: Please use a range of 0 to 1 for the y-axes (except for BIAS, where the y-axis could be zoomed in further, unless you want to emphasize that this is always nearly equal to 1; see also comment for lines 190-191).

Reply to Technical Correction 7: As suggested by the reviewer, we have changed the range of yaxis (0 to 1). The plot of BIAS has been removed in context to reply of specific comment 2.

8.One of the previous papers is referenced in at least one place as "Kuldeep et al. (2017)", but appears in the reference list as "Sharma K., ...". The other previous paper is referenced as Kuldeep et al. (2019), but does not appear at all in the reference list. I found two further references that do not appear in the references list: Grant (2001) and Donaldson et al. (1975).

Reply to Technical Correction 8: Thank you for pointing out. All the references are correctly included in the text and reference list.

Please also note the supplement to this comment:
https://www.geosci-model-dev-discuss.net/gmd-2019-65/gmd-2019-65-AC2-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-65, 2019.

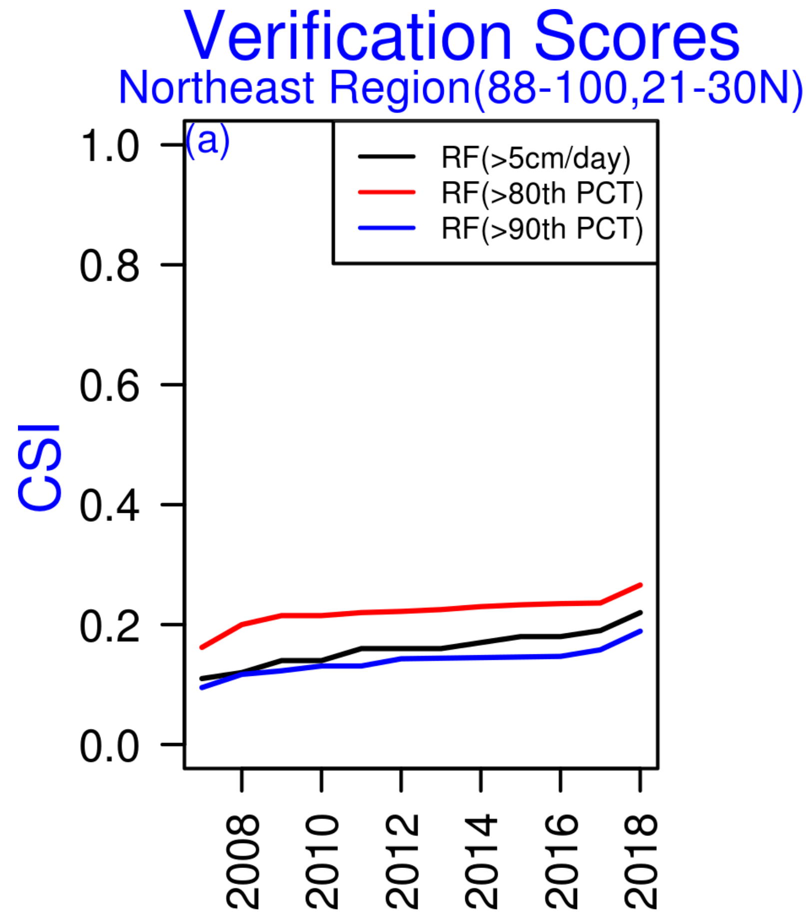

Fig. 1.

Interactive comment

Printer-friendly version

Discussion paper

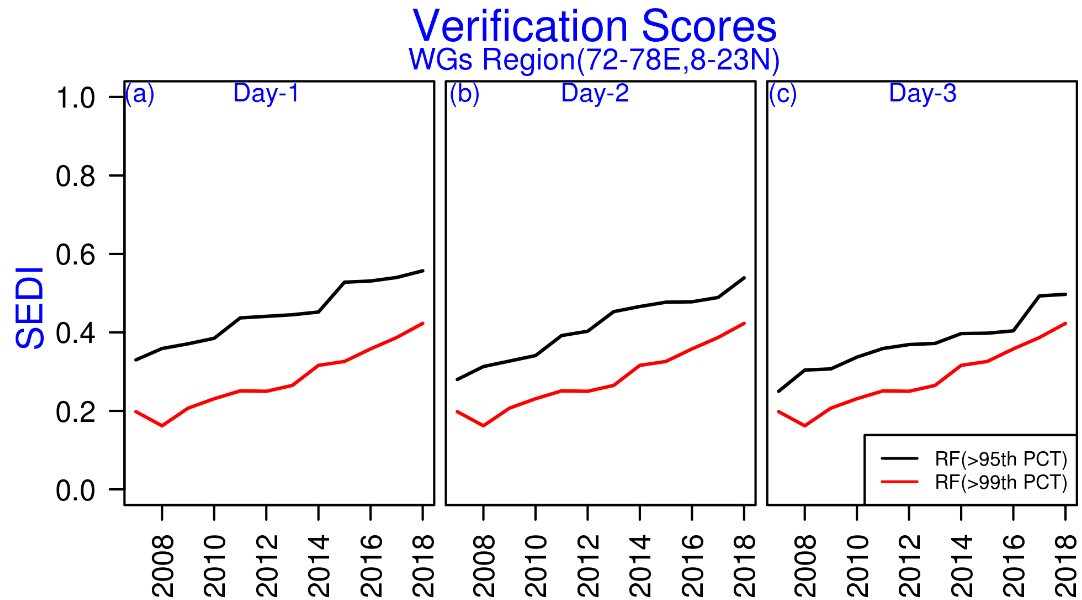
Verification Scores
WGs Region(72-78E,8-23N)
SEDI
(a) Day-1
(b) Day-2
(c) Day-3
1.0
0.8
0.6
0.4
0.2
0.0
2008
2010
2012
2014
2016
2018
RF(>95th PCT)
RF(>99th PCT)

Fig. 2.

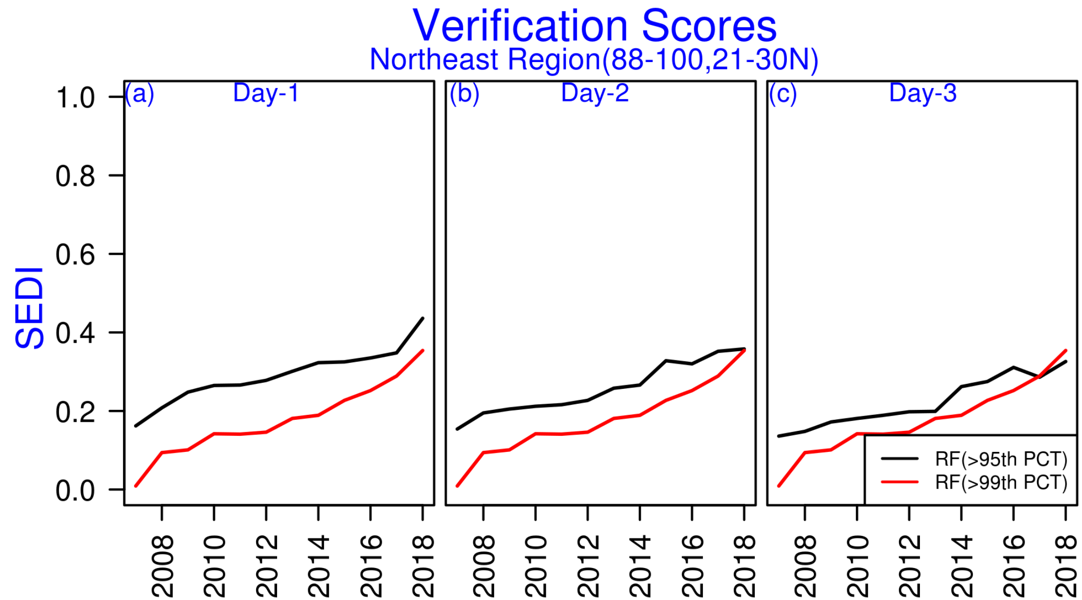

Fig. 3.

Printer-friendly version

Discussion paper