Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-65-RC2, 2019

Geoscientific Model Development Discussions
Open Access

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

Anonymous Referee #2

Received and published: 10 July 2019

General Comments:

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 mountaineous regions of India, using recently-available observations.

The overall conclusion of the paper, that UM rainfall forecasts have improved steadily

Printer-friendly version

Discussion paper

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

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.

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.

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 mountaneous 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.

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.

Specific Comments:

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?

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.

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

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.

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.

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.

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?

Line 324: Does this sentence refer to the improvement from 2007-2008? This improve-

Printer-friendly version

Discussion paper

ment in the data assimilation is not listed in Table 1.

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

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?

Technical Corrections:

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

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.

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

Line 317: “Figures 8 and 9”.

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

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

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).

There are numerous minor grammatical errors, but these should be picked up during the copy-edit.

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).

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