

We thank the anonymous referee for such a detailed review. The insights provided will definitely improve the quality of the manuscript.

The referee's primary concern is regarding the hydrologic evaluation of IMERG over Indian basins. We agree that the novelty of this study lies in the hydrologic evaluation. However, the availability of streamflow data for Indian basins for the time period of IMERG data availability (starting from March 2014) is limited. WRIS, the website (<http://www.india-wris.nrsc.gov.in>) which provides streamflow dataset for India, is not updated and contains data for only a few gaging sites from March 2014 onwards. On going through the WRIS portal again (in January 2017) expecting better streamflow data availability, we found streamflow for the 'Barman' Gaging station in Upper Narmada basin, 'Ashti' gaging site for Wainganga river sub-basin of Lower Godavari from March 2014 apart from the gaging sites in Mahanadi basin that we have already used. We did the hydrological evaluation over Wainganga river sub-basin and will include the results in the revised manuscript. In case of Upper Narmada basin we found the flow was regulated through a reservoir and in the absence of reservoir discharge data it is extremely difficult to calibrate the model, hence we will not include it.

The referee reported the need to draw comparisons of our conclusions with those of the previous studies done for a pan-India scale. We will include them in our revised manuscript.

Another issue was regarding the length of the manuscript along with a large number of figures. The referee gave detailed suggestions, to which we have responded subsequently. We will move some figures into the supplementary and condense the text.

We have clubbed the responses to minor grammatical or language correction. To other questions pointed by the referee, we have answered them in point-by-point answer format.

Title: Slight misplacement of punctuation, I believe this should read: "Does the GPM mission improve the systematic error component in satellite rainfall estimates over TRMM? An evaluation at the pan-India scale."

Lines 47-69: Interesting, and I see why this has been included, but this much detail is maybe not required as not all of these example are directly relevant to this study; this paragraph could easily be shortened.

Lines 75-77: This is almost a repeat of lines 44-46.

Lines 120 & 142: I would suggest replacing the word "scanty" with "scarce", which is much more widely used and less colloquial.

Section 2.1: While background information (and especially the maps) on the study area is always appreciated, I would recommend condensing section 2.1 - not all of the information is relevant or referred to later in the paper.

Lines 201-202: This is a repeat of lines 78-79.

Section 3: Throughout the results section, there are a lot of statements along the lines of "IMERG outperforms TRMM in x out of y basins, but they are similar in z basins" – the authors may be able to reduce the text and number of figures by constructing a table of the number of basins in which IMERG outperforms TRMM, the number in which they are similar, and vice versa, for each skill measure evaluated in the paper.

This would also be interesting for the reader to give a quick overview of these numbers without needing to read the entire text and pick them out. Of course, it is still worth discussing these and the regional differences etc. as the authors have done, but this text could be reduced.

Lines 279-280: *The authors state that the two datasets show similar skills, and immediately then state that IMERG is better in 70% of the basins - this is somewhat of a contradiction.*

Section 3.3: *Throughout the section on basin-wise bias, the results are difficult to follow. Typically in the literature, a positive bias indicates over-estimation, and a negative bias indicates under-estimation. I would recommend that the authors amend the presentation of the results here to also use this convention, making it more intuitive for the*

reader and more consistent with the literature. This is simply a case of reversing the sign in the results, i.e. using bias = simulated - observed, instead of bias = observed - simulated.

Some of the language chosen in this section is also confusing, see specific comments following:

Line 352: *The authors use the term “increased” bias - it is not clear if this refers to a larger negative or positive bias.*

Line 408: *Does section e refer to section 3.5?*

Line 543: *The term “slightly” is ambiguous - how much worse are they? How much better is the NSE? How much larger the bias?*

We will make the suggested modifications in the revised manuscript.

Answer to detailed comments:

Lines 99-104: *I would like to see more justification of the choice to focus on the basin level, to make it clear what the benefit of this study is over the previous studies the authors have mentioned. The authors state “most” of the previous studies - what about the remaining? How does this study improve on this? Why is the basin scale more useful for water resources and policy makers? It is not clear at the moment why this would be much more useful than the grid-scale analyses.*

We specifically focused on basin scale because it is more relevant hydrologically. The results of a basin scale study can be directly used by the watershed managers. Most of the previous studies (as cited in the manuscript) focus on gridscale but we see a gradually changing trend to analysis on basin scale (Bisht et al., 2017; Kneis et al., 2014). It becomes easier to compare the statistical and hydrologic results when the analyses are carried out at a basin scale. Thus, we used basin scale as the reference in this manuscript.

Line 262: *Could the authors clarify this statement?*

The hydrologic model was calibrated twice, once with IMD as the rainfall forcing and once with TRMM. The model was not calibrated with IMERG as the data period was too short

(March 2014 – December 2014). Instead, the two variants of the calibrated model were validated separately using IMERG and TRMM as the rainfall forcings for the year 2014.

Regarding the warmup period, the calibration period was from 2000-2011. The year 2000 was taken as a spinup period and the results for 2000 was excluded while computing calibration statistics.

Lines 270-275: Some of this explanation should be included in the datasets section. It is not clear why this is done like this - why were the TRMM statistics obtained for 2 periods? Also implied here is that IMERG data is only available for March – December 2014, but later in the conclusions the authors state that a longer timeseries is available. This is confusing and should be clarified. If a longer timeseries of IMERG is available, why did the authors choose to use only 2014?

There seems to be a misunderstanding in IMERG timeseries availability. We meant to say that the IMERG is still a very young mission having started in March 2014, and as more data becomes available with time, they will lead to a clearer picture as to how IMERG compares with TRMM.

Lines 309-310: Could the authors expand on what the implications of this result are; why is it worth noting?

The comparison is drawn between the retrospective (1998-2013) and current (2014) time period of TRMM. Over a long period, there is a lot of temporal smoothing which may not be true for a shorter time scale. We just pointed it out in the manuscript, it doesn't really have any other significance.

Line 354: Surely, in the 20 basins that now exhibit a positive bias which did not before, this is indeed a decay in skill for these basins? Please clarify this statement.

As mentioned in the text, although the number of basins with positive bias increased, it wasn't a fall in skill as the basins with relatively unbiased results ($-10\% \leq \text{Pbias} \leq 10\%$) increased. What really happened was some of the more negatively biased basins went to the unbaised category, thus improving the overall skill.

Line 356: What do the authors mean by an increase in the variability of the bias? This is not clear.

This line is really ambiguous and has no clear meaning. We would remove it in the revised manuscript.

Lines 354-365: The terms “lower” and “higher” when referring to bias are ambiguous; it would be better to refer to “smaller” and “larger” biases. Again, it is not clear in this paragraph whether the authors refer to positive or negative biases. Please also check the rest of the section / paper for further use of these terms.

We agree that the use of "smaller" and "larger" biases instead of "lower" and "higher" biases make more sense. We would take it into account in the revised manuscript.

Lines 474-475: What is the reason behind this part of the evaluation? What do the authors aim to gain from this analysis? This may have been mentioned earlier in the paper but is not completely clear and it would be good to see clarification at the start of section 3.5.

We performed a correlation analysis of skill with climatology and topography to understand the systematic biases in satellite products. We will reemphasize it in section 3.5.

Line 488: Again, the use of "high/low" when referring to bias is confusing.

High/low bias will be changed with large/small bias.

Lines 533-537: This reads as though it should be part of the methodology of the paper, rather than results.

This was included to quickly recap the calibration and validation time durations. We feel this is a good practise as the reader doesn't have to go back in text and he/she can get the relevant information in brief.

Section 3.6: This section is presented in the introduction as a major part of the novelty of this study, but in comparison to the proportion of the paper spent discussing the rainfall results, very little discussion is offered in terms of the hydrology. The implications of the findings are not discussed, and with only one basin used in this experiment, it is not possible to say whether the results would be similar for other basins in India or elsewhere. The aim of this experiment is left unclear and while I think it could be a very interesting part of the study, it seems somewhat unfinished. I would like to see, as the authors state would indeed be interesting, a comparison of these results for other basins in different regions in the study area.

A key limitation to our study is that the hydrologic evaluation hasn't been carried out over more basins. As mentioned in the beginning of this response, there is not much that we can do about it due to the limitation in available recent streamflow data for majority of Indian basins. We will include hydrologic evaluation over Wainganga basin.

Conclusion 1: To which parameter do the authors refer to with the quoted values?

We referred to skill in terms of correlation. We will mention it in the revised manuscript.

Conclusion 5: Use of "higher" bias, as before.

We will modify the high/low bias terminology to larger/smaller in the revised manuscript.

Conclusion 7: If a longer timeseries of IMERG is available - why was this not used? This should be clarified / justified.

As mentioned before, there seems to be a misunderstanding in IMERG timeseries availability. We meant to say that the IMERG is still a very young mission and as more data becomes available with time, they will lead to a clearer picture as to how IMERG compares with TRMM. We will clarify that in our revised manuscript

Lines 601-604: These statements are somewhat contradictory. The authors state throughout that IMERG outperforms TRMM in various aspects, and here state that there is a reasonable improvement, and also that the improvement is only incremental and not ground-breaking, but also that IMERG is a worthy successor of TRMM. These statements leave the reader somewhat confused as to what the overall conclusion of the study is.

We will modify the revised manuscript to improve the clarity of this message.

Line 611: "post forecast data assimilation scheme" - do the authors refer to postprocessing?

We indeed meant postprocessing of streamflows.

Figures: I'm afraid there are too many figures included in this manuscript, particularly considering the majority of figures are multi-panel. I would suggest moving a number of the figures into supplementary material. Please see below my detailed comments and recommendations for each figure:

Figure 1: Thank you for including this map, this is incredibly helpful for those readers who are not as familiar with the geography of the region. I would recommend splitting Figure 1 into two figures, one containing the two geographical maps (a) and (d), and the second comprising of (b) and (c). Also, the colour scales used for (b) and (c) are confusing - please modify these; the best way to present these would be a colour bar with just one colour for each map, ranging from light to dark with increasing values.

We will split the figures as suggested by the referee.

The reason for selecting multiple colorbar for figure 1 (b) and (c) is to highlight the spatial heterogeneity in the study area. When we used a simple one colorbar, a lot of information was lost in the contrast (for instance the contrast between low rainfall in Rajasthan and medium rainfall in the Western part of Indo-Gangetic plain in figure 1(b)).

Figures 2.1 and 2.2. Firstly, it is strange to label two separate figures as 2.1 and 2.2 - surely these should be figures 2 and 3. Secondly, what exactly is the precipitation shown here? Is it daily precipitation? Is it averaged over a time period? This should be clarified and included on the axes / in the caption, for both 2.1 and 2.2.

We will split the figures 2.1 and 2.2 into two figures. The scatterplots show daily precipitation which we will mention in the figure description.

Figures 3 and 5: These figures are very difficult and confusing to interpret - this data is not continuous (it represents the independent basins, rather than e.g. a continuous time period), and this is not the best way to present it. I would in fact recommend removing figures 3 and 5, and just discussing their results in the text as you have done.

We will move the figures in supplementary material and retain the discussion in the text.

Figures 4 and 6, the corresponding spatial maps, are a much clearer way of presenting the data.

Figure 4: I like these maps, it is clear what they show and intuitive to interpret. However, the colours used are very confusing - please amend the colour scale to use just one colour from 0 to 1 (light to dark), and avoid rainbow colours. In the case of (j), (k) and (l) it is not immediately obvious that there is a negative correlation in one or more of the basins and it is hard to spot. So on these maps, two colours should be used – the same as (a-i) for 0 to +1, with white at 0, and a different colour for the negative values. For example, the colour scale the authors have chosen for figure 8 would be perfect for figure 4, with white at 0.

The rationale behind using a rainbow color to represent correlation was to focus on the spatial heterogeneity in correlation values. When we used a single color in the colorbar (ranging from light to dark), most of the spatial features of correlation were lost. For instance, it became really difficult to decipher correlation value of 0.3 from 0.6, which is rather substantial. That's why we used rainbow color bars instead of a single color bar.

Figure 6: Again, I like this figure, but the colour scale should be improved. I would recommend again a scale such as that used in figure 8, where 0 is white and the darker the colour, the larger the value. Please note that the colour scale has a big impact on the way the reader interprets the data, and incorrectly used colour scales can be misleading.

We will modify the color scale as recommended by the referee.

Figure 7: Again, the colour scale here is not the best option. For this data, the best would be to use one single colour, from light at 0 to dark at 1. For example, in figures (j-l), at first glance it seems that the blue basins have an opposite result to the red basins, but this is not the case.

Figure 8: While this scale would be perfect for the results shown in figure 4(j-l) and figure 6, it is not the good choice for the data presented in figure 8. As with figure 7, the best option would be one colour from light at 0 to dark at 1.

We opted for the multi color bar to highlight the spatial variability. If there is one colorbar which varies from light to dark, a lot of information is lost as the contrast decreases significantly. For instance, in the case of figure 8, FAR of anything more than 0.5 can be taken as high error, which will be lost if the colorbar had only one color scale. We will use the same color bars in figure 7 and 8 but incorporate the referee's suggestion in other figures.

Figure 9: This graph could be removed and just discussed in the text.

The figure will be moved to the supplementary section in the revised manuscript.

figures 10 - 17: While I can see why the authors have presented the data in this way, again, there is the issue that this data is not continuous so this type of graph is not really correct, and also this is confusing for the reader. There are also a large number of similar plots here, I would suggest to pick one or two which show the most interesting results to present in the main body of the paper, and move the rest to supplementary material. Most readers would not analyse all the information in all of these figures and would appreciate the highlights, but the interested reader could easily find all the graphs in the supplementary material. This would solve the problem of the overwhelming number of figures included in this paper. Also, I would recommend that the authors display all of these graphs (whether in the main body or the supplementary material) instead as scatter plots of the rainfall/elevation vs. bias/correlation. This would be a much more accurate and easy-to-understand way of displaying the data.

We will move the majority of the figures to the supplementary section. Also, we will show the respective scatterplots along with the line plots in the supplementary section.

Figure 18: What is “data points”? Is it time on the x axis? Please change this label, and if it is indeed time as I suspect, please display the dates.

Data points indeed here means dates. We will modify the label.

Recommendation: I believe that the work is of interest / useful, and warrants publication, but the manuscript indeed requires some work in terms of the descriptions and presentation of the study and its results, and clarification of some confusing aspects of the paper. Ideally, I would like to see the rainfall-runoff exercise extended.

We would like to again thank the referee for giving such a detailed feedback. We will try to incorporate the suggestions in our revised manuscript.

References:

Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S. and Sridhar, V.: Spatio-temporal trends of rainfall across Indian river basins, Theor. Appl. Climatol., 1–18, doi:10.1007/s00704-017-2095-8, 2017.

Kneis, D., Chatterjee, C. and Singh, R.: Evaluation of TRMM rainfall estimates over a large Indian river basin (Mahanadi), *Hydrol. Earth Syst. Sci.*, 18(7), 2493–2502 [online] Available from: <http://www.hydrol-earth-syst-sci-discuss.net/11/1169/2014/hessd-11-1169-2014.pdf> (Accessed 20 October 2014), 2014.