

Response to Referee Comment #1

Thank you for your insightful comments, which have helped strengthen the manuscript. We have revised the manuscript accordingly, and please find below our itemized responses.

In the conclusions, the authors states that “The use of IMERG Early to calibrate GFS would improve GFS in terms of correlation and variability, but not in terms of bias”. How do you come to this conclusion? Just via comparing the performance of GFS forecasts and IMERG Early products? I think it is inadequate. I strongly suggest that the authors should add some more analysis on the comparison of performance between raw GFS forecasts and calibrated GFS forecasts by IMERG Early products. In addition, the authors also evaluate the performance of some other Satellite Precipitation Products, such as CHIRP, IMERG Early and IMERG Early Cal, against the IMERG Final rainfall products. However, I do not understand why you evaluate these products? You evaluated these products but did not do any analysis on using these products to improve the GFS forecasts.

Response:

- We acknowledge the importance of comparing the performance of raw forecasts with calibrated GFS forecasts. However, this would require developing new appropriate methodologies, which is outside the scope of this study. We have added a new paragraph (paragraph 4 of the Conclusions Section) indicating the need for developing such a methodology, and suggested potential methodologies that could be explored. We plan to pursue this in future research.
- As far as the purpose of evaluating the performance of different datasets (CHIRP, IMERG Early, and IMER Early Cal) is concerned:
 - **IMERG Early Cal:** The climatological bias correction approach used to generate the ‘IMERG Early Cal’ dataset did not improve performance compared to IMERG Early. Thus, we have decided to remove the ‘IMERG Early Cal’ dataset from our evaluation (as suggested by the Reviewer in a separate comment below).
 - **CHIRPS:** In this study, IMERG Final has been used as a reference to evaluate the performance of GFS forecasts. We conducted additional assessment to evaluate the performance of IMERG Final with respect to another independent and high-quality (i.e. satellite-gauge merged) rainfall product (i.e. CHIRPS). Agreement between the reference (IMERG Final) and CHIRPS would indicate that the IMERG Final estimates are robust. We have added a new paragraph in the Conclusions section (paragraph 2) as well as additional texts in Section 2.3 to clarify this.
 - **IMERG EARLY:** Post-processing calibration of GFS forecasts (in order to improve the accuracy of GFS) requires the use of “relatively better performing” and “available in near-real-time” independent rainfall observations to correct real-time dynamical GFS model forecasts. In this study, we conducted additional assessment to evaluate the performance of the near-real-time product, IMERG Early. Comparison of the performance of IMERG Early with the performance of GFS would indicate to what extent the IMER Early products could be used for

calibration of GFS forecasts. We have added a new paragraph in the Conclusions section (paragraph 3) as well as additional texts in Section 2.3 to clarify this.

- The quoted statement, “The use of IMERG Early to calibrate GFS would improve GFS in terms of correlation and variability, but not in terms of bias”, was removed as it was not conveying adequate information.
- The entire Conclusions Section has been edited to improve clarity.

In section 3.4, what’s the forecast uncertainty? How to evaluate or quantify the uncertainty? I think the “uncertainty” in section 3.4 is only the different performance, but not uncertainty. The authors stated that the GFS forecasts show large underestimation bias for heavy rain rates. I suggest to add some explanations for the poor performance, by evaluating other variables related to the physical mechanism that affect the precipitation over the study region, or citing some relevant references.

Response:

- We accept the correction. We have revised it to read “Dependence of Forecast Performance on Precipitation Rate”.
- There are very few studies on evaluation of GFS forecasts (see our Introduction section). We found one paper that examined the bias of GFS at high rates. The paper reported the difficulty of capturing high rain rates in GFS models. We have added this reference (see Section 3.6). Evaluation of the different error sources of GFS forecasts is outside the scope of this study, as our approach focuses on evaluation of total GFS performance (lumping together all error sources) due to limitation in our ground reference data.

In section 3.5, the authors states that the climatological bias correction approach is not effective in removing the bias in IMERG Early estimates. Why do you present the results? It is not meaningful for this manuscript. I think you could do some analysis for the effective method of bias correction to improve the IMERG Early estimates and thus to improve the GFS performance by calibration.

Response: We agree that it is not meaningful to present results that do not improve performance. Hence, we have removed the climatological bias correction from our evaluation. In the Conclusions section (see last paragraph), we have recommended the need for identifying appropriate bias calibration techniques, and suggested some potential methodologies that could be explored.

In addition, the introduction should be improved seriously. For example, the current studies on the evaluation of GFS forecasts and its performance on global scale or other regions should be added.

Response: We have added literature review of GFS performance evaluation in other regions of the world (see Paragraph 4 of the Introduction Section).

Minor comments:

The abstract should be carefully revised. For example, it should not include the detail introduction of study basin.

Response: We have revised the abstract, and also removed the details in the introduction of the study basin.

The resolution of GFS forecasts and Satellite Precipitation Products are not consistent, how do you deal with them? The authors do not describe any information about this.

Response: We have added statements describing the methodology used to bring both products to the same resolution (see paragraph 1 of Section 2.5).

Line 76: remove “)”

Response: Done. Thank you.

Figure 1: I suggest to add legend for the drainage basin, or use the appropriate color for the boundary of the sub-basin

Response: Done. Thank you.

What is R in Figure 2? Please classify.

Response: It was correlation, however, we have removed it now as it is not that meaningful.

Line 235: “How well is the annual precipitation total forecasted in each dam watershed?” I do not understand, please classify.

Response: Here, we aggregate the 1-day lead GFS forecasts to annual time scale and compared the results against corresponding annual precipitation estimates from IMERG Final. To improve clarity of the text, we have revised the section heading as “How well do GFS forecasts capture annual rainfall?” and also added a clarifying text (see first sentence of that section).