<u>**Authors' response to referees**</u>

Dear Anonymous Referee#2,

We would like to thank you for your thoughtful reviews of the manuscript. You have pointed out important issues, and your comments are very helpful to enhance the quality of our manuscript. In the following, a response to the general and specific comments by the refree#2 is provided. In this report, first comments from referees are given then the corresponding author's response and changes in the manuscript (if there are changes) are followed. The page numbers both for the comments and in our response refer to the original manuscript that was reviewed.

General Comments:

Referee comment:

In this paper, an extensive validation of satellite precipitation products with rain gauge measurements in Ethiopia (Upper Blue Nile region) is performed. The study is done thoroughly, and much statistical comparison information has been collected. This is a useful validation paper.

The presentation needs, however, improvements:
- Please carefully check the English language throughout the paper: for example, check plural - singular, check articles (a – the are sometimes interchanged).
- Please add in the figure caption the location and period of the measurements
- Please add information in the figure legends instead of in the x/y axis labels

In Section 5, please shorten the description of the results. In the text, mostly the results of the Tables are repeated. Please only give the highlights, and focus on discussing the physical causes and mechanisms.

Authors' response:

We are very glad to know that referee# 2 has stated the importance of our validation paper. All the general comments were considered and we have tried our best to revise accordingly. The author's have made proper English editing and re-checked grammatical errors available throughout the manuscript and made a correction when appropriate. We have focused on important results only and tried to shorten the description section (provided in Section 5) as much as possible (this issue is addressed in the revised manuscript). The location and period of measurement has also been added for all figure captions as suggested by the referee and all figure captions contain the following phrase to indicate the region and study periods: "…the Upper Blue Nile basin for the period of 2000-2015". In addition, all the figures information that were presented in the x/y axis labels (in the original manuscript) have been removed and provided as a legend in the modified figure (we have discussed this matter below in the specific comments section).

Specific comments:

Referee comment:
Abstract: Biases: what is the unit? What is the bias in percent? More quantitative results on the CHIRPS accuracy should be added to the abstract, since that was the aim of the paper.

Authors' response:
The bias in our study was calculated using the ratio of satellite rainfall estimates and ground observed rainfall values. Therefore, the bias values, in this case, has no unit and the value of 1 indicate the perfect score in which satellite rainfall manages to capture an equivalent amount of rainfall recorded by rain gauge stations. A bias value above and below 1 indicate that an aggregate satellite overestimation and underestimation, respectively, of the ground precipitation amounts. However, it is still possible to express the bias values in percent as well. For example, bias values of 1.04 and 0.76 can be expressed in percent, one is overestimating ground observed rainfall by 4% and the other one is underestimating by 24%, respectively. The author's have also expressed bias values in percent in the manuscript when it is needed.

In addition, quantitative results have been added to the abstract as suggested.

Authors' change in the manuscript: The changes are highlighted with yellow color.

Accurate measurement of rainfall is vital to analyze the spatial and temporal patterns of precipitation at various scales. However, the conventional rain gauge observations in many parts of the world such as Ethiopia are sparse and unevenly distributed. An alternative to traditional rain gauge observations could be satellite-based rainfall estimates. Satellite rainfall estimates could be used as a sole product (e.g. in areas with no (poor) ground observations) or through integrating with rain gauge measurements. In this study, the newly available Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) data has been evaluated in comparison to rain gauge data for the period of 2000 to 2015 across the Upper Blue Nile basin in Ethiopia. Besides, the Tropical Applications of Meteorology using SATellite and ground-based observations (TAMSAT) version 2 and 3 (TAMSAT 2 and TAMSAT 3) and the African Rainfall Climatology (ARC 2) products have been used as a benchmark and compared with CHIRPS. The TAMSAT version 2 rainfall estimates were used in this study mainly to assess the improvements made with the recent version of a TAMSAT product (TAMSAT 3). From the overall analysis at dekadal and monthly temporal scale, CHIRPS exhibited the best performance in comparison to TAMSAT and ARC2 products. An evaluation based on categorical/volumetric and continuous statistics indicated that CHIRPS has the greatest skills in detecting rainfall events (POD=0.99, 1.00) and measure of volumetric rainfall (VHI=1.00, 1.00), the highest correlation coefficients (r=0.81, 0.88), the best bias values (0.96, 0.96), and the lowest RMSE (28.45 mm/dekad, 59.03 mm/month) than TAMSAT and ARC2 products at dekadal and monthly analysis, respectively. CHIRPS overestimates the frequency of rainfall occurrence (up to 31% at dekadal scale), particularly during the dry months, although the volume of rainfall recorded during those events was very small. Indeed, TAMSAT 3 has shown very comparable performance with that of CHIRPS product, mainly with regards to bias. CHIRPS has underestimated ground observed rainfall only by 4% for both dekadal and monthly temporal scale while TAMSAT 3 has overestimated just by 4% and 3%, respectively. The ARC 2 product was found to have the weakest performance underestimating rainfall amounts by about 24%. In addition, the skill of CHIRPS is less affected by variation in elevation in comparison to TAMSAT and ARC 2 products. This validation study also shows that the TAMSAT 3 has overcome the main weaknesses of TAMSAT 2, which is an underestimation of high rainfall amounts by up to 31% in this study. Overall, the finding of this validation study indicates the potentials of CHIRPS product to be used for various operational applications such as rainfall pattern and variability study in the Upper Blue Nile basin in Ethiopia.

Referee comment:

p. 5, l. 9: which satellite data was used?

Authors' response:

Thank you for the comment. The "Satellite data" indicated in p. 5, l. 9 refers to the two Thermal Infrared (TIR) satellite observations archives used for the generation of CHIRPS. The first one is Globally Gridded Satellite (GriSat) archive produced by NOAA's National Climate Data Center and the other is NOAA Climate Prediction Center dataset (CPC TIR).

Authors' change in the manuscript: Based on the comments, this has been clearly indicated in the revised manuscript as follows:

…First, Infrared Precipitation (IRP) pentad (5-day) rainfall estimates are created from two TIR satellite observations archives (i.e., Globally Gridded Satellite (GriSat) and NOAA Climate Prediction Center dataset (CPC TIR)) using cold cloud duration (CCD) and calibrate using the Tropical Rainfall Measuring Mission Multi-Satellite Precipitation Analysis (TMPA 3B42) precipitation pentads…

Referee comment:

p. 6, l. 12: clarify this sentence

Authors' response:

The comment is accepted and the sentence is re-written.

Authors' change in the manuscript: The sentence is re-written as follows:

… The comparison between gridded satellite rainfall estimates and ground rainfall observations can be made using either grid to grid or point to grid comparison methods. However, an attempt made to convert point ground observations to gridded interpolated dataset lead to poor result due to uneven geospatial distributions of gauge stations. Thus, this study has used point-to-grid comparison approaches….

Referee comment:

Eqs. 1 - 5: Please write the sums clearer: summation from i=1 to i=n. What does the index i mean?

Authors' response:

The comments are accepted and corrected accordingly.

Authors' change in the manuscript:
Equations are corrected according to the comment as follows. Only one equation is presented here to show the changes we made otherwise the comments are applied to all equations in the revised manuscript.

$$\text{VHI} = \frac{\sum_{i=1}^{n}(S_i|(S_i>t\&G_i>t))}{\sum_{i=1}^{n}(S_i|(S_i>t\&G_i>t))+\sum_{i=1}^{n}(G_i|(S_i\leq t\&G_i>t))},$$

Where S is satellite rainfall estimates, G is gauge observations, i= 1 to n and n is the sample size, and t is the threshold values (t=1mm in this study).

Referee comment:
p. 7, l. 16: Which unit has t?

Authors' response:
The threshold value t has a unit in mm.

Referee comment:
p. 7, l. 22: What is the relation between r and R^2 used frequently in the paper? If it is the same quantity, then use only one.

Authors' response:
Thank you for the comment. It is the same and we now used only r throughout the revised manuscript.

Referee comment:
p. 8, l. 20: Why do these discrepancies occur?

Authors' response:
Basically, TIR based satellite rainfall estimates are assuming that a significant portion of rainfall in the monitoring area is convective and there is a linear relationship between the length of time that cold cloud duration (CCD) and amount of rain that falls. It considers warm orographic clouds as non-precipitating. However, this might not be true for complex topographic areas such as the Upper Blue Nile basin (in our study area) which could produce orographic rainfall and possibly dominated by warm rain processes. In this case, such an approach will lead to underestimation of rainfall in high elevation areas. Therefore, the large discrepancy for TAMSAT2 and ARC 2 rainfall pattern in our study area could be attributed to the orographic effect on rainfall, warm rain process, and the calibration practice using gauge stations. TAMSAT 3 and CHIRPS product have shown a better performance in this aspect that could be because of improved calibration process using gauge stations and inclusion of elevation in the development of CHIRPS dataset, respectively.

Referee comment:
p. 10, l. 28-29: please clarify this sentence

Authors' response:
The comment is accepted and the sentence is re-written.

Authors' change in the manuscript: The sentence is re-written as follows:

…However, the competencies of TAMSAT and ARC2 products in detecting rainfall events seem to reduce with elevations.

Referee comment:
Figures: For all figures: please indicate the region and the time period to which the data belong.

Authors' response:
The comment is accepted and already addressed in the general comments section above.

Referee comment:
Figure 2: Why are the data sets TAMSAT2 and ARC2 so poor in horizontal structure of precipitation patterns? Is this due to the binning scale? Is it still useful to consider these low-resolution datasets?

Authors' response:
As indicated in the figure (Figure 2) the value of TAMSAT 2 and ARC2 reduces horizontally (from west to east regions or from low to high elevation areas (as indicated in Figure 1)). This is actually due to the effect of warm orographic clouds in high elevation areas as well stated above. Such an effect will lead to underestimation of rainfall in high elevation areas. This issue has been indicated in the manuscript. These datasets are obsolete now and they are outperformed by the CHIRPS and the recent TAMSAT version (TAMSAT 3), at least for our study area.

Referee comment:
Figure 2: The figure suggests that the Kiremt season has more precipitation than the total year for the TAMSAT3 dataset. That cannot be correct. Figure 2: For these important maps, can you please a give lat/long grid like in Fig. 1?

Authors' response:
Thank you for the comment. We agree with the comment. Initially, there were cartographic symbolization errors in presenting the legend and we have corrected it accordingly. In addition, we have added the lat/long grids to the figure as commented by the referee.

Authors' change in the manuscript:
We have modified Figure 2 as follows:

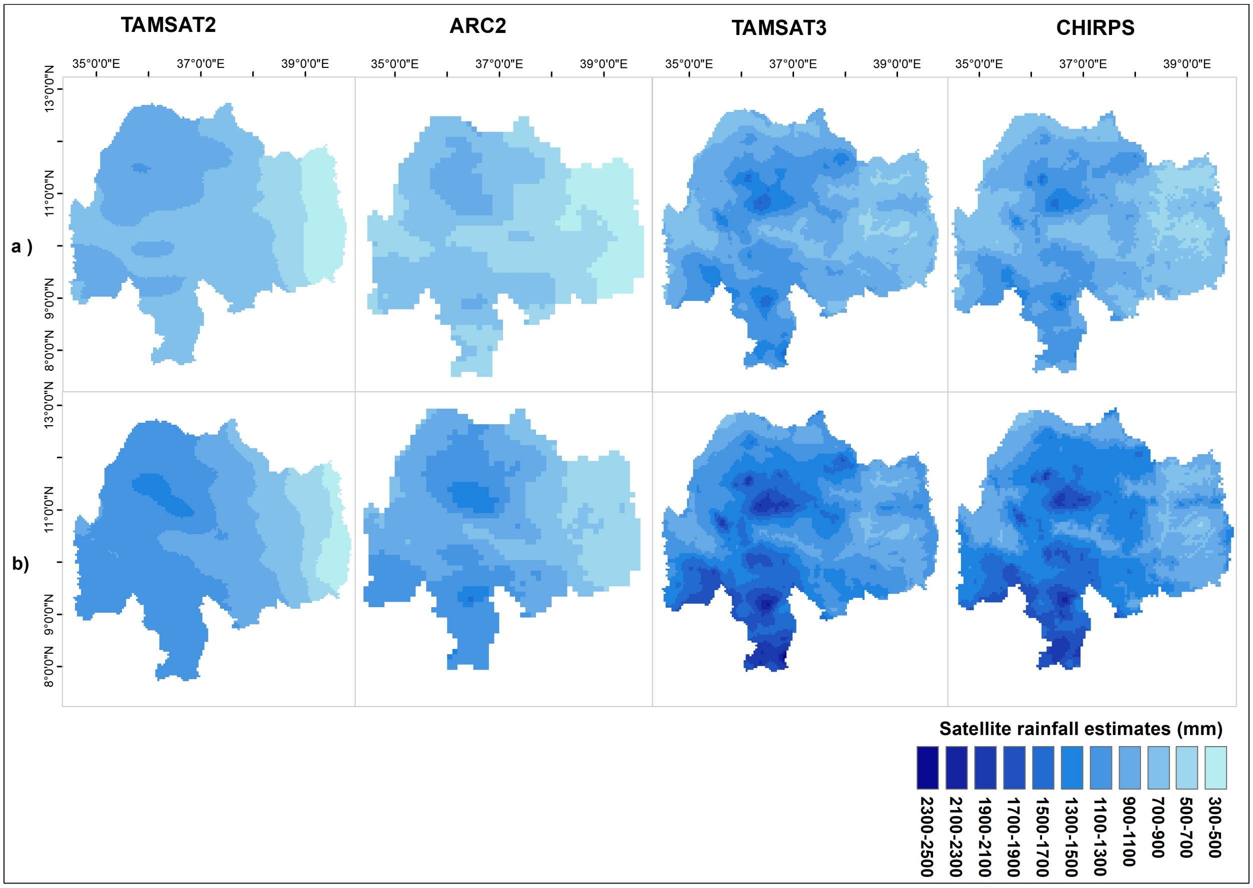

Figure 2: Comparison of mean satellite rainfall estimates for: (a) Kiremt season (June-September), and (b) annual rainfall over the Upper Blue Nile basin for the period of 2000-2015. Years with missed values were not considered in the mean analysis.

Referee comment:
Tables 1 and 3: What do the bold numbers mean?

Authors' response:
It was just by mistake and it is now avoided.

Referee comment:
Fig. 3: please zoom in on the 0 - 200 mm part, which is most interesting.

Authors' response:
We have accepted the comment and added additional figure which magnifies the values from 0 t 200 mm part.

Authors' change in the manuscript:
We have modified Figure 3 as follows:

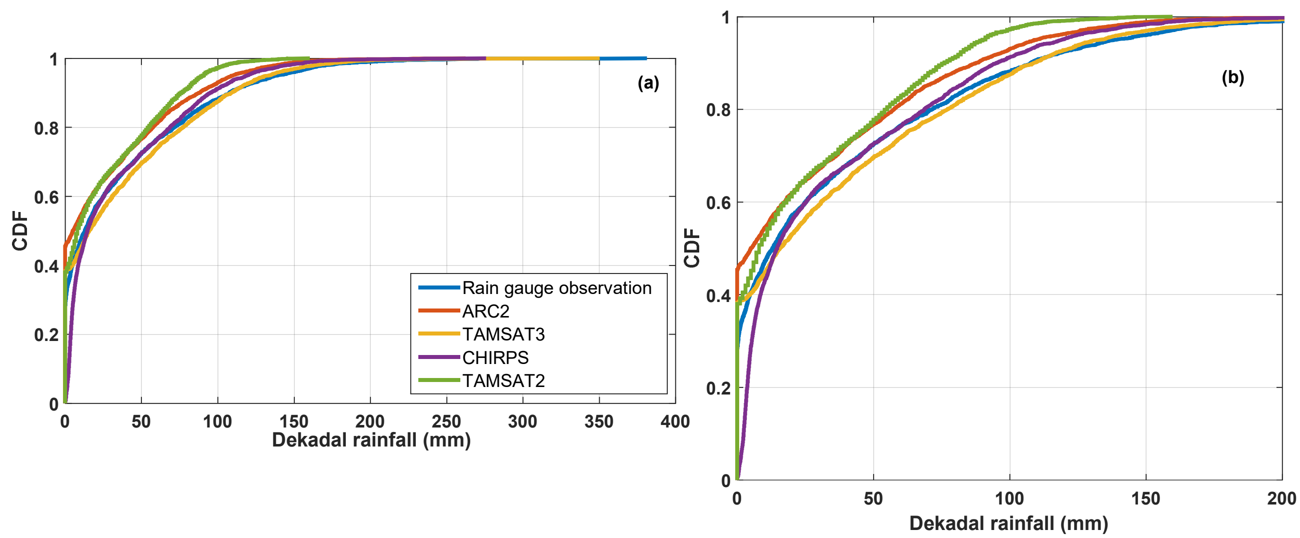

Figure 3: Cumulative distribution function (CDF) of dekadal rainfall for ground rainfall observation, ARC 2, TAMSAT, and CHIRPS rainfall estimates (a) and magnified view of their CDF for 0 to 200 mm part (b) over the Upper Blue Nile basin for the period of 2000-2015.

Referee comment:
Fig. 6: what is the unit of the Bias?

Authors' response:
It is already addressed under the general comments section above.

Referee comment:
Fig. 7: Please give the station name and altitude in the figure legend, not in the y-axis label.

Authors' response:
We have accepted the comment and the station name and altitude is given in the legend.

Authors' change in the manuscript:
We have modified Figure 7 as follows:

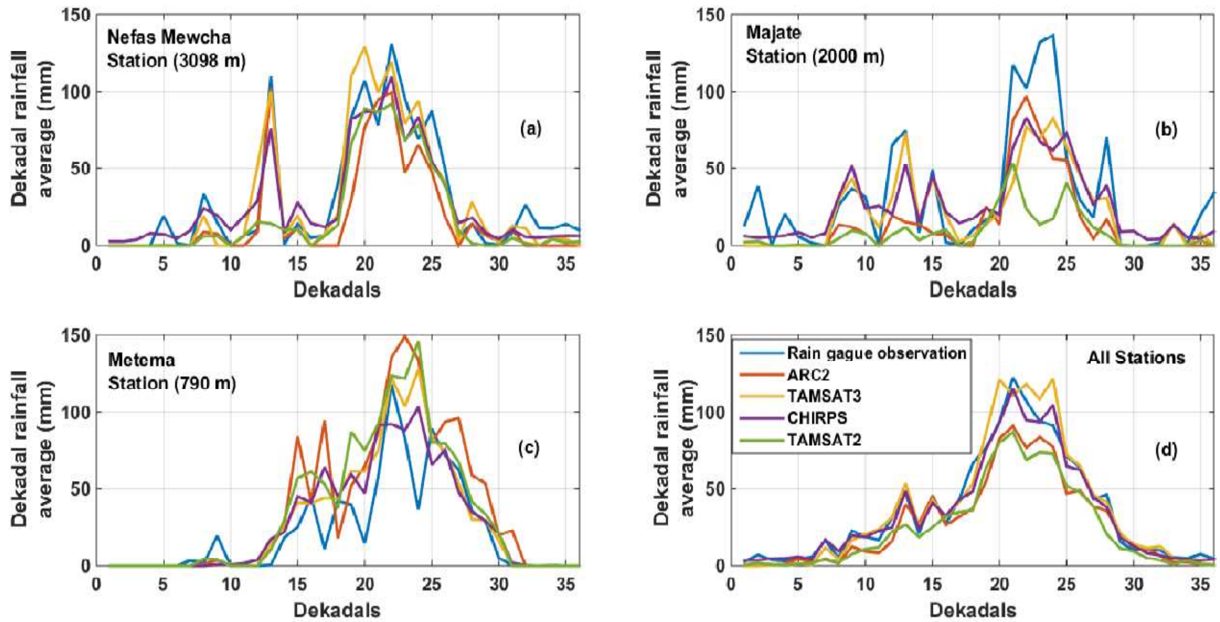

Figure 7: Comparison of the satellite rainfall products at rain gauge stations with wider difference in elevation values (e.g., > 2000m) based on dekadal average for over the Upper Blue Nile basin for the period of 2000-2015, (a) at "Nefas Mewucha" station with an elevation of 3098 m a.s.l, (b) at "Majate" stations with an elevations of 2000 m a.s.l, (c) at "Metema" stations with an elevation of 790 m a.s.l, and d) at dekadal rainfall average from all rain gauge stations. The x-axis represents the 36 dekadals of a year.

Referee comment:
Fig. 8: for clarity, please put the data set name in the legend or above the figure, but not in the x-axis label.

Authors' response:
We have accepted the comment and the data set names are provided in the legend.

Authors' change in the manuscript: We have modified Figure 8 as follow. The same modification has been made on Figure 4 as well.

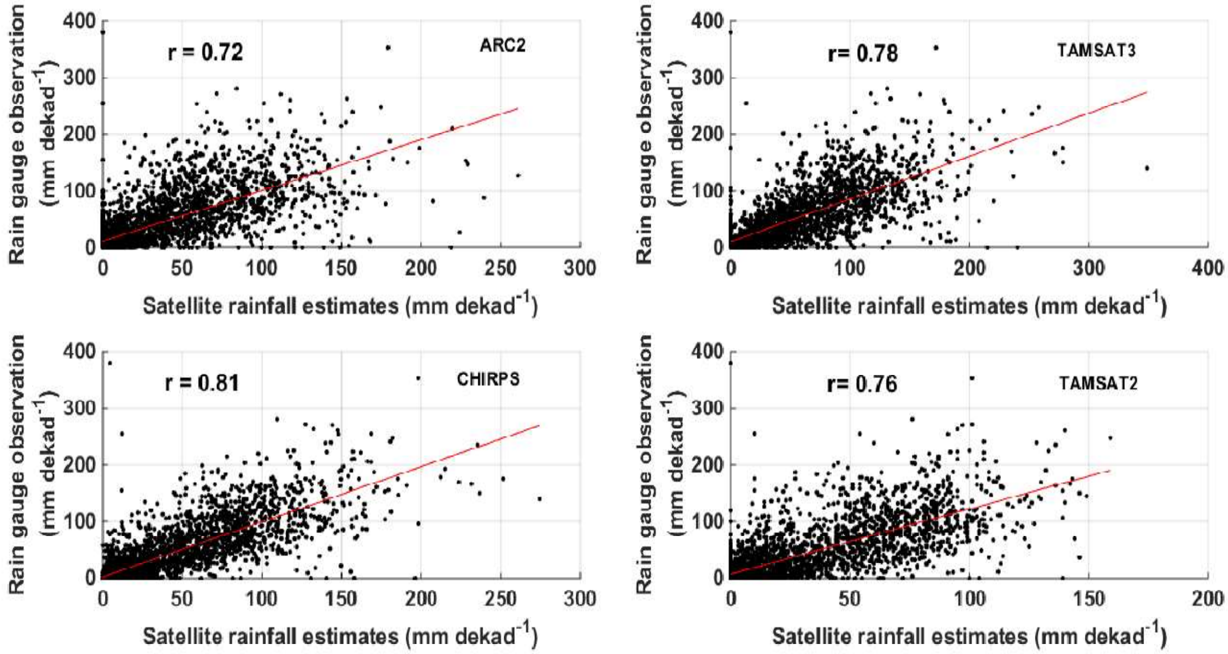

Figure 4: Scatter plot between rain gauge observations and satellite rainfall estimates at dekadal temporal scale over the Upper Blue Nile basin for the period of 2000-2015.

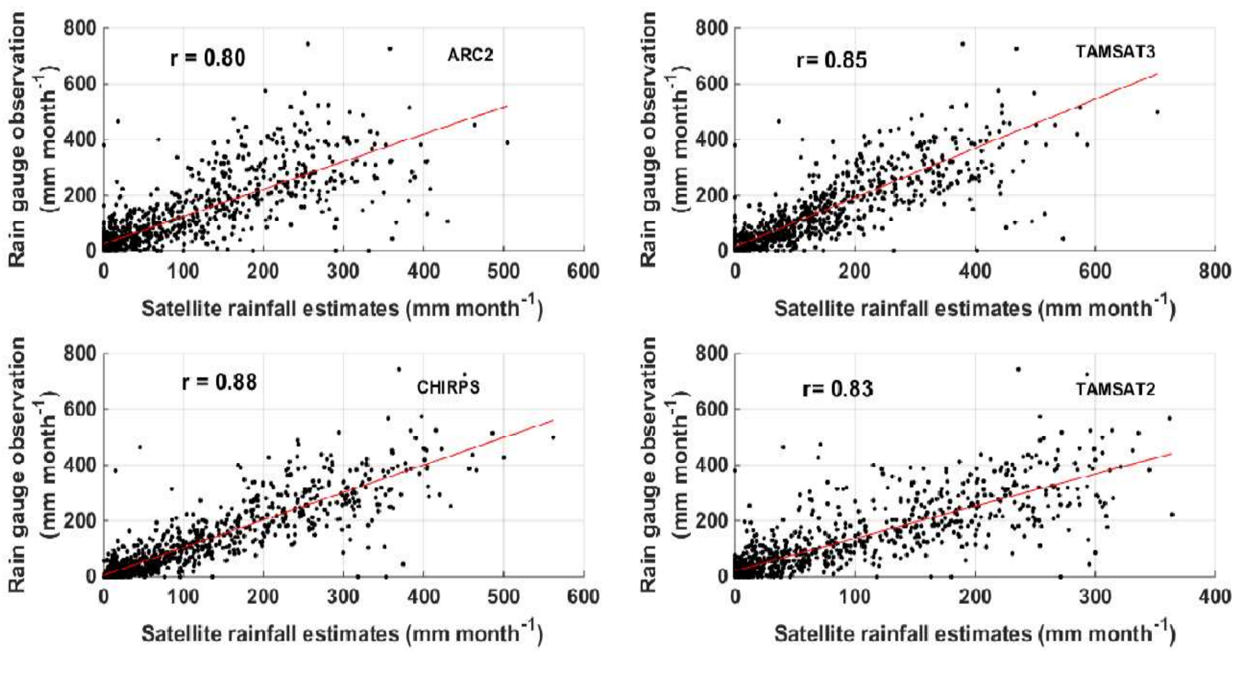

Figure 8: Scatter plot between rain gauge observations and satellite rainfall estimates at monthly temporal scale over the Upper Blue Nile basin for the period of 2000-2015

Minor syntxs and typos corrections:

Referee comment:

1. p. 2, l. 24-2: spaces at the end of sentences are missing
2. p. 5, l. 10: calibrated
3. p. 6, l. 19: m > M
4. Please use italics for all symbols: H, M, F, i, t, n.
5. Eqs. 1 – 3: Please use spaces in these conditional relations.
6. P. 7, l. 16: Where > Here
7. P. 8, l. 6: Where > Here
8. p. 8, l. 24: at a different elevation > per elevation
9. p. 10, l. 27: more prominent
10. p. 11, l. 20: TAMSAT
11. Caption Fig. 1: …are with high elevation …
12. Fig. 7: caption: of year > of a year
13. Fig. 9, 10: caption: at each twelve months > for each month
14. Fig. 10, caption : 2025 > 2015

Authors' response:

All the technical corrections (listed from 1 to 14) commented by the referee has been accepted and addressed accordingly in the revised version of the manuscript