Manuscript: Evaluation of satellite rainfall estimates over Ethiopian river basins

Major remarks

The authors are evaluating different satellite-based rainfall estimates over Ethiopia. Such satellite estimates are very helpful over data sparse regions so that studies evaluating their accuracy over specific regions are important. In order to understand biases in the satellite estimates, these studies should also include analyses and speculations on why biases occur in the precipitation estimates. In my opinion the current manuscript is good in its descriptive part, but it has several shortages (see also several minor remarks) in the background and speculation part (too much number crunching without providing explanatory details leading to better understanding). A good result is that the authors found a relation of bias and elevation. But no explanation or speculation is provided what is causing this relation. Here, also the knowledge obtained in previous studies/publications may be taken into account.

Partially it is unclear, which reference data are used in the data comparisons. Partially the SREs itself are used as reference data, even though this paper aims to evaluate them (e.g. Table 3). You cannot use data as reference in a study where you want to evaluate these data. In the conclusions section, for example, it is written: "This study has found that all three products capture the varied spatial precipitation pattern over the mountainous northwest, and the homogeneous mean annual precipitation and rainy days in the lower elevation of the southeast."
But no rainfall pattern based on rain gauge data is shown, which would be necessary to support this statement.

The question on the limitations/uncertainty in the results of the present study is omitted up to now. For example, are there any shortcomings or limitations of the study that are induced by the fact that the rain gauge data are only monthly and that they also comprise missing data in the rather short time period of 5 years?

In summary, I suggest that the paper may be accepted after major revisions are conducted.

Minor remarks

In the following suggestions for editorial corrections are marked in *Italic*.

Abstract – par. 1 - p. 7670 – line 16-20
This part is very descriptive ("performs better than"), but no indication on why one estimate performs better than another is given.

Introduction – par. 3 - p. 7671 – line 15
... and *western mountains* during ...

Introduction – par. 3 - p. 7671 – line 17
... in the *western mountains* and...

Sect. 2.1 - p. 7672 – 7674
This section is partially redundant with Table 2. Please focus on important characteristics of the basins, and leave out information on ranking or the mere repetition (without any specific comment) of numbers that can be easily obtained from the table.

Sect. 2.1 – par. 2 - p. 7673 – line 8
... and *provide* six ...

Sect. 2.2
Sect. 2.2 provides numbers of annual precipitation, thereby pretending that these are the truth observed values. But looking at Table 3, it becomes obvious that these number are derived by averaging the three satellite estimates, despite the fact that the paper aims at evaluating these estimates in the following sections and despite of any biases of specific estimates that are not known in this part of the paper. This does seem adequate. It would be better to base these estimates on rain gauge measurement, and later on compare the satellite estimates.

Again, ranking is rather unimportant as long as no specific comments are made for this, e.g. with regard to climatic situation.

Sect. 2.3 – par. 2 - p. 7675 – line 26
... for the *six* basins ...

Sect. 3.1 – par. 1 - p. 7677 – line 21
It is written:
... to the other SREs (correlation of 0.912) …

Correlation of what with what? You refer to two other SREs, but only give one value.

Same remark applies to Sect. 3.1 – par. 2 – p. 7678 – line 28.

Sect. 3.1 – par. 1 - p. 7677 – line 22
It is written:
…, whereas PERSIANN underestimates …

This implies that PERSIANN has a low bias, so you pretend that the other two estimates are the truth. This is not justified in the text. Thus, you can only state that PERISANN estimate is lower than the other two SREs.

Sect. 3.2 – par. 1 - p. 7678 – line 14
The *Rift* Valley …

Sect. 3.3 – par. 1 - p. 7679 – line 4-6
In this section, the three SRE estimates are compared to rain gauge data. But correlation numbers seem to be given for the SREs to each other, and not to the rain gauge data. The reason for this is not clear.

Sect. 3.4 – par. 1 - p. 7679 – line 24-26
It is written:
A bias ratio greater than one indicates overestimation by SRE, a bias ratio less than one indicates underestimation by SRE, and a bias ratio of one indicates no bias in the SRE.

Why there is no range around 1 that considers that the bias is within the uncertainty of the rain gauge measurements? In the current statement, it is assumed that a) rain gauge data are perfect, and b) no acceptable range of bias is defined around 1 where the estimate can still be considered as good.
This is certainly not adequate!

Sect. 3.4 – par. 2 - p. 7680 – line 11
… basins *dominated by* less …

Fig. 1
Use discrete colour steps! For black/white figures, not more than 5 steps are recommended.

Fig. 2
Legend is too small. Please use 5 instead of 6 or 7 colour steps.
In addition I don't understand the colour scales ranging from 4 to 1 to 1.6 to 0.4 (upper panels) and 4 to 1 to 16 to 4 (lower panels).

Fig. 3, 4, 5, 6
Legends are too small.

Fig. Caption 5, 6
It is written:
Based on river basins with at least 10 or more rain gauges.

This statement is confusing as in the whole paper, only 6 river basins are studied.