Interactive comment on "Cross-validating precipitation datasets in the Indus River basin" *by* Jean-Philippe Baudouin et al.

Anonymous Referee #2

Received and published: 12 August 2019

The manuscript entitled "Cross-validating precipitation datasets in the Indus River basin" compares a collection of twenty rain gauge, satellite and reanalysis precipitation data sets in the upper and lower Indus river basin using a cross-validation methodology. This paper is a valuable study for academics and practitioners who use precipitation data sets in the area. My recommendation is that the paper is published after revision to the comments and questions below.

1) Abstract Line 14. "These findings highlight the need for a systematic characterisation of the underestimation of rain gauge measurements" Whilst you raise this issue in the abstract it is not discussed at all in the conclusions, either comment on this in the conclusion or remove from the abstract.

2) P.g.5. You provide a brief description of the catchment, but I think this could be

improved by stating actual elevation values of the catchment alongside the size of the catchment and the two subcatchments.

3) P.g.5. You use Figure 1 (A) as the reference in the description of the catchment, but I think more value would be obtained by making a separate larger figure to discuss the catchment. I think that the map should include elevation as well.

4) Section 2.2. You provide a very good description and rationale for why you selected certain rain gauge and reanalysis data sets. However, for the satellite data sets the section is very short. Was alternative satellite products considered, and if so why were they not picked? What was the advantage of selecting the data sets you do choose to include?

5) Page 6. Line 8 "which is useful for comparison" what do you mean by this comment? Are you saying that due to the CRU having a similar resolution and time coverage it was useful to compare to just the GPCC-monthly or for the entire analysis?

6) Page 6. Line 16 "and the largest variety of input" what do you mean by this comment?

7) Page 6. Line 17 "is useful for comparison" why is this data set in particular useful for comparison?

8) Page 6. Line 18 "All three datasets use GPCC for calibration" Which three data sets? Why is this important? Does this have any further implications in the analysis since the GPCC is used as the comparative data set?

9) Page 12. You use bi-linear interpolation to estimate the grids, why? Where other methods considered?

10) Page 12. How were abnormally large rainfall events (outliers) considered when you calculated the mean? As this may have skewed the mean?

11) Section 3. Whilst the results section is very extensive and detailed, it also is very

difficult to read due to it not having many (only 3) subsections. I think to improve you should split each of the subsections into subsubsections with their own theme.

12) Section 3. You use the GPCC-monthly data as the base to compare against however this was never justified in the text. I think this should be at least mentioned in Section 2.3 (methods) section.

13) Section 3. Partway through you change to compare against a different data set Dahri2018, why? Again this should be added into the methods section.

14) Page 40. Line 26 "Particularly, correlations are greatly impacted by extreme values". Why was this not discussed earlier in the text?

15) Page 40. Line 27 "Moreover, we deliberately selected a large domain of study to improve the confidence in the datasets" Why was this not discussed earlier in the text?