**Hydrology and
Earth System
Sciences**

Discussions

Open Access

EGU

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

Jean-Philippe Baudouin et al.

jpb88@cam.ac.uk

Received and published: 1 October 2019

2nd REVIEWER

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.

Answer: We have added a paragraph concerning this issue in the conclusion as it is one of the key messages we want to convey (cf. answer to general comment 1 of the 1st reviewer)

TEXT MODIFIED

   • **In the conclusion:**

"As mentioned above, rain gauge-based datasets underestimate precipitation. Only GPCC products use a correction factor to account for measurement underestimation, but this one is still too small. We emphasise the need to correct directly the measured values before interpolation to a grid, using, for example, methods similar to those developed by Dahri et al. (2018)"

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 sub-catchments.

Answer: we added the size of the domain considered in that section. We also added a figure with the elevation (cf. answer to comment 4 of the 1st reviewer).

TEXT MODIFIED

- **On the size of the domains (section 2.1):**

"Thus, the northern part of the basin (hereafter the upper Indus, **595000 km^2**) includes the maxima of precipitation along the Himalayas and most of the winter precipitation, while the southern part (hereafter the lower Indus, **785000 km^2**) is characterised **by a unique wet season, in summer, as wintertime precipitation is negligible"**

- **Regarding the elevation map, see answer to general comment 4 of the first reviewer.**

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.

Answer: The figure has been added and is used to introduce the study areas (cf answer to comment 4 of the 1st reviewer and 2 of 2nd reviewer)

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?

Answer: The main reason we selected those datasets is that they enable the study over a common period of 10 years with the other dataset as they start in 1998 or earlier. We added a sentence at the start of that subsection to discuss this point.

TEXT MODIFIED

- **At the start of section 2.2.2:**

"Various satellite-based gridded precipitation products are available, but we have only selected datasets providing data from 1998, to ensure a long enough common period with the rain gauge-based datasets (the common period reaches years due to APHRODITE ending in 2007)."

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?

Answer: this has been removed as the aim of the study is to compare all datasets, not specifically those (same specific comment by 1st reviewer)

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

Answer: we have referred to the amount and type of observations included in the datasets (same specific comment by 1st reviewer).

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

Answer: Similarly as for comment 5 of the 2nd reviewer, this sentence has been modified.

TEXT MODIFIED

"We also selected the daily product from the Global Precipitation Climatology Project (GPCP-1DD; Huffman and Bolvin, 2013) as well as the monthly product issued by the same group (GPGP-SG Adler et al., 2016)"

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

Answer: We added the name of three datasets in parenthesis. It does impact the analysis when using GPCC as reference. This characteristic is used to explain the results where appropriate. We added a sentence here to clarify this point. We also changed the following sentence to highlight the reason for the selection of CMAP (unlike the other it is calibrated by another dataset, this further addresses the comment 4 of the 2nd reviewer).

TEXT MODIFIED

- **At the end of section 2.2.2:**

"All three of these datasets (TMPA, GPCP-1DD, and GPGP-SG) use GPCC for calibration**, which could introduce some similarities. By contrast, the last dataset** included, CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997), uses CPC for calibration**. It** has the same time coverage and resolution as GPCP-SG**. This version** does not include reanalysis data, to simplify the analysis**."**

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

Answer: Same comment as general comment 5 of the 1st reviewer, see answer there.

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

Answer: Abnormally large rainfall events are not treated separately. If limited to a small area, their effect is mitigated by taking the average over the study areas. We justify this as we are not interested in such fine scale phenomena in this study. Some abnormal events remain at the basin scale. They cause problems when considering correlation on a moving window (e.g. Figure 5 and 6). In that case we have used the Spearman coefficient. We also checked that the main result remained the same when the Pearson coefficient is used. A sentence is added to the method section about the use of Spearman correlation. We also further discussed the limitation of the Pearson correlation regarding extreme values in the conclusion

TEXT MODIFIED

- **in the Method section:**

"To reduce the impact of abnormally large rainfall events when investigating the trend (cf. Section 3.3.4), we use the Spearman correlation."

- **in the conclusion**

"We have used the Pearson correlation to compare the datasets, although it has some limitations. For example, it is affected by extreme values, that is, in our context, abnormally large precipitation events. These led to some difficulties in interpreting trends and we preferred the Spearman formula in this context (cf. Figures 6 and 7). By contrast, the Pearson correlation is less affected by the difficulties in representing the lowest precipitation rates, although these one could explain some of the biases."

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.

Answer: We agree with the comment and split the subsection 1 and 2 of the result section into several sub-subsection. See answer to the general comment 8 of the first reviewer

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.

Answer: Similar to comment 7 and 8 of the first reviewer. We added information about the use of references in the method 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.

Answer: The Dahri2018 dataset offers a very interesting assessment of the rain gauge undercatchment and demonstrates that part of the difference in mean precipitation between reanalyses and observational datasets can be explained by this undercatchment. However, we cannot use this dataset as a reference in the whole study, as it covers only a small fraction of the Indus watershed, which is itself included in our upper Indus domain. Furthermore, the paper does not provide the monthly values that we could have used to assess the seasonality. The part of the analysis using this dataset is now enclosed in a subsection, and we refer to it in the method section. See also answer general comment 8 of the first reviewer, and the changes in the method section in the answer to general comment 5 of the first reviewer

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

Answer: This sentence has been removed. We actually found similar results using both Spearman and Pearson correlation coefficients. What is more problematic is the heavy tail towards 0, and we further discuss this in the conclusion (see answer to comment 10 of the 2nd reviewer)

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?

Answer: we discuss this point now in the conclusion in relation to the important uncertainty in the fine scale spatial patterns of precipitation. See 4th paragraph of the updated conclusion in the answer to the general comment 1 of the first reviewer.

C9