HESSD

Interactive
comment

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

Anonymous Referee #1

Received and published: 31 July 2019

Review of the paper: "Cross-validating precipitation datasets in the Indus River basin", by Jean-Philippe Baudouin, Michael Herzog, and Cameron A. Petrie

This paper presents a comprehensive assessment of several (twenty) state-of-the-art datasets for precipitation in the Upper Indus Basin region. These datasets cover different sources including station-based observational data, satellite products and re-analyses. The paper provides important and useful information on the strengths and uncertainties of the different datasets which could serve as a reference for precipitation in this area and be used for validation purposes.

However, in order to be really of use, the paper requires some revisions. Some sentences are hard to be read and should be rephrased or better explained. Some grammatical errors need to be fixed. I think that the paper deserves publication in this journal

only after the comments/requests listed below are carefully addressed.

*General comments/questions

- Is it possible from the analysis presented in the paper to really identify the best and/or the worst performing dataset, though a possible dataset "rank" probably depends on the variable/process one is looking at? The authors present advantages, disadvantages and strengths of the various datasets but the main message which remains to the reader, I think, is that many uncertainties remain. The conclusion, as it is, is not really "positive". What is the main message that the authors want to convey?

- Given the analysis presented in the paper, do the authors think that using the average of all datasets (a multi-dataset mean) could be a further output to be provided, along with the individual datasets themselves? I suggest to add the mean of all observation-based datasets, and of all reanalysis dataset indeed.

- One of the findings of the paper, which corroborates previous studies, is the fact that precipitation estimates from rain gauges are underestimated. The last sentence of the abstract highlights the need to account for this bias. Is it possible/reasonable, based on the study and results presented in the paper, to suggest "correction factors" to be applied to rain gauge estimates for the study area?

- In the description of the domain of study (section 2.1), it would be nice to have information on the average elevation of the two study domains and the range of elevations at least in the upper Indus.

- I have some concerns about the methodology used to interpolate the different datasets at the same spatial resolution. Is bilinear interpolation correct when dealing with precipitation fields? Wouldn't be preferable to use a more conservative approach? Did the authors test other approaches?

- In the section "Methods", the authors say that the comparison among the different datasets is performed in terms of mean and variability. What do they mean with "vari-

ability"? Is that year-to-year variability? Daily? Something else? This must be better specified in the Methods before going to the Results section.

- The discussion of Figure 2 (for the observation-based datasets, both in-situ and satellite), though containing many elements and considerations is, in my opinion, slightly confused. One reason is that the authors do not state in a very clear way that, e.g., they are taking one dataset as the reference (GPCC-monthly) against which to compare the other datasets. I agree that one reference is used, but this should be clearly stated (for example already in the "Methods" section)

- Each subsection of the "Results" section is very long (especially 3.1 and 3.2) and there is a risk that the reader gets "lost", in combination to the fact that there is a lot of information delivered. I suggest to try reducing these subsections a little bit and make it clear what the final message to the reader is. For example, one confusing thing, at least for me, is that on the one hand the "reference" dataset against which the other products are compared is GPCC-monthly (if I understood well), while, on the other hand, another cited paper (Dahri2018) is taken as a reference (long discussion in Section 3.1).

- Tables and Figures are not (always) correctly introduced in the text. In my opinion, when a figure/table is cited, it should be briefly described to say what it shows/displays (leaving the technical details to the caption and legend). Also check all Figures and Tables captions.

- Words like "consistency" or "consistent" are often used but I think that they are too generic. Please try to find other ways to convey the message.

- I'm not completely comfortable with the message that the reanalyses are useful to validate observations (as stated for example in the Conclusions). To be better discussed.

*Specific Comments

Abstract

- Page 1, Lines 3-4: In my opinion, the sentence "While rain gauge ... underestimation" would need to be rephrased. I would add a comma (,) after the word "reference" and I would change the subsequent sentence as follows, ", they provide information for specific, often sparse, locations (point observations) and are subject to underestimation in mountain areas", rather than "they are only punctual [...]"

- Page 1, Line 5: Add "data" after "reanalysis"

- Page 1, Line 10: Please replace "most able" with "most performing", or similar

- Page 1, Line 14: I don't understand whether "small" refers to "correction"; also, the term "correction" should be better explained. Please try to rephrase this sentence, if possible.

Introduction

- Page 2, Line 13: Add "of" between "use" and "rain gauges"

- Page 3, Line 1: 900km2 –> 900 km2

- Page 3, Line 2: 250 km2 –> 250 km2

- Page 3, Line 3: 15,000 km2 –> 15,000 km2

- Page 3, Line 7: 500 km2 –> 500 km2

- Page 3, Line 14: Add "the" between "and" and "heterogeneity"

- Page 3, Line 17: Remove "a" between "over" and "flat"

- Page 3, Line 18: Replace "those" at the end of the sentence with "station data"

- Page 3, Line 21: Replace "case" with "cases"

- Page 3, Line 23: Replace "consider" with "highlight" ; "reanalysis" with "reanalyses" and "observation" with "observations"

- Page 3, Line 25: Please change "has it made possible" with "has made it possible"

- Page 3, Line 26: Add "product" after "reanalysis"

- Page 3, Lines 28-30: I would remove the sentence starting with "Specifically" and ending with "variability". These are like results and conclusions of the study which is going to be presented, not useful here.

- Page 3, Line 30: Please replace "qualities" with e.g. "strengths and limitations"

- Page 3, Line 31: Please replace "have" with "has"

- Page 3, Line 34: Please replace "method" with "methods"

- Page 4, I would specify somewhere that the analyses described in items i), ii), and iii) concern precipitation. For example, "[...] which review the precipitation i) seasonal cycle [...], ii) daily variability [...], and iii) monthly and longer term [...]". Moreover, I would expect another sentence at the end of the paragraph for the Conclusions section. As it is, the sentence seems like suspended.

Data and methods

Domain of study

- Page 5. Line 8: I'm not sure to correctly identify the contour indicating the Luni River, as mentioned in the text. Is it the dark blue thick contour which also indicates the upper border of the study area? If so, as I also understand from the sentence at lines 10-11, I would better specify this at this point (maybe moving the sentence at lines 10-11 above)

- Page 5, Line 9: Please replace "bound" with "bounded"

- Page 5, Line 14: Please change "while the rest of the year it remains dry" with "while during the rest of the year the basin remains dry"

- Page 5, Line 16: I would replace "but exhibit" with "exhibiting"

- Page 5, Line 17: Rather than "process", I would say "circulation patterns"; also

please add another reference in parentheses which is significant for explaining the wintertime precipitation, i.e. Filippi et al., 2014 (Filippi, L., E. Palazzi, J. von Hardenberg, and A. Provenzale: Multidecadal Variations in the Relationship between the NAO and Winter Precipitation in the Hindu Kush-Karakoram. J. Climate, 27, 7890-7902, https://doi.org/10.1175/JCLI-D-14-00286.1, 2014)

- Page 5, Line 18: Please remove "it does"; the sentence is okay also in this way "As in the southern part of the basin"

- Page 5, Line 24: Not true, from Fig. 2, that there is no precipitation at all in winter in the lower Indus. I would rather say that "the southern part [...] is mostly characterized by summer precipitation (wintertime precipitation is negligible)", or something similar.

Data

- I would change the title of Section 2.2 in "Datasets", Section 2.2.1 in "Rain gauge data", Section 2.2.2 in "Satellite data", Section 2.2.3 in "Reanalysis data"

- Page 5, Line 29: Typo, there is a double parenthesis after the citation Yatagai et al., 2012.

- Page 5, Line 32: Please replace "the fact that" with "because"

- Page 6, Line 1: Delete "with a period covered". The sentence should be ... by the same institute extending up to 2015"

- Page 6, Line 7: Please replace "to" after "coverage" with "as"

- Page 6, Line 15: Please replace "of the available" with "among the available"

- Page 6, Line 15: What does "variety of input" mean? Please be more specific.

- Page 6, Lines 16-17. I don't understand what the sentence about GPCP_1DD means. Why do the authors specify that this specific dataset is valid only for comparison? What does this mean?

- Page 6, Line 18: At the beginning of the sentence, please rephrase "All three datasets described above use GPCC ...."

Reanalysis

- Page 9, Line 2: To avoid repetitions, "reanalysis datasets" should be replaced with "reanalysis data"

- Page 9, Line 3: better to say "can vary" rather than "varies"

- Page 9, Lines 3-4: I would rephrase this sentence in this way "Table 3 shows the ensemble of the ten reanalysis datasets which we used in this study".

- Page 9, Lines 10-11. I would change a little bit the sentence, for example: "ERA5 currently starts in 1979 (see Table 3) but future releases are expected to extend back to 1950".

- Page 9, Line 12: please change "than the others" with "than the other products"

Methods

- Page 12, Line 6: Please add "of precipitation" after "seasonality"

- Page 12, Line 9: "issues on" –> "issues of"

- Page 12, Line 9: For the last sentence, I would rather say "Winter is not analysed in the lower Indus as it is an extremely dry season"

- Page 12, Line 10: Please add "precipitation" before "time series"

- Page 12, Lines 11-18. For me all this doesn't fit here. This is already a result or, better, a possible explanation of the reasons why the different datasets show different behaviours. This should be discussed in the Results section, or in a dedicated Discussions section (to be eventually added) or in the Conclusions.

- Page 12, Line 22: "as we will discuss...." –> "as discussed in the Results section"

Results

Subsection 3.1

- Page 13, Line 3: Add "precipitation" before "seasonal cycle"

- Page 13, Line 13: The sentence of GPCC needs to be rewritten, I suggest: "... we compare the datasets to GPCC-monthly data, taken as a reference for this analysis". I think that it is more correct to state that GPCC-monthly is considered here as the reference rather than saying that it provides "good precipitation estimates", unless the authors add some references in support of this statement.

- Page 13, Line 7: I would start the sentence in a different way: "Figure 2 overall shows that all different datasets are able to capture the seasonality of precipitation in the two areas, though with different magnitudes"

- Page 13, Lines 7-12. I don't like the description of Figs. 2 A) and C) made in this paragraph. I would avoid sentences like "are ranked in the same order"; I would try to describe the climatology of precip. as seen by the different datasets taking one of them as the reference (as the authors do, if I understand well). Basically, the figure needs to be better described, also highlighting the performances of the various datasets in summer and winter.

- Page 13, Line 11: Please replace "inferior to" with "less than"

- Page 13, Line 13: I would replace "of mean precipitation" with "of the precipitation annual cycle"

- Page 13, Line 15: Regridding can be source of uncertainty, depending also on the kind of interpolation which is applied. The bilinear interpolation could not be the most appropriate method for precipitation. So the term "carefully" is questionable in this sentence, in my opinion.

- Page 13, Line 16: I would say that GPCC-daily "uses" GPCC-monthly and not that "is

based"

- Page 13, Lines 16-17: GPCC-daily and GPCC-monthly are not so different, as expected. Though GPCC-daily uses less stations, it incorporates GPCC-monthly analysis which uses more stations. I expect that these 2 products are very similar.

- Page 13, line 25: CPC is "drier" or driest? Maybe driest is the correct term and this refers to the upper Indus only.

- Page 13, Line 26: "linear relation"? I would rather say "clear correlation"

- Page 13, Line 30: I would replace "creator" with "developers"

- Page 13, Lines 31-33: please rephrase the entire sentence. Here is (only) a suggestion: "In particular, APHRODITE underestimation of total precipitation (compared to GPCC products) might be related to the fact that it partly relies on GTS data, in which missing values could be treated as no precipitation values. The large dry bias seen in CPC data could be associated with the same issue, since CPC is entirely based on GTS." I still don't understand, however, why a missing value in CPC would be treated as no precipitation.

- Page 14, Line 2: "build" –> "building"

- Page 14, Lines 2-5: This sentence needs to be rewritten, it is not really understandable especially when referring to TMPA, and to correlations (what datasets ?); this is really not clear to me.

- Page 14, Line 8: Please change "Different" with "Several" or "Various"

- Page 14, line 11: The sentence "they are basin-wide more numerous [...] territory" should be rephrased and improved.

- Page 14, Line 13: Rather than "explains" I would say "could reasonably explain"

- Page 14, Line 24: Remove "somewhat"

- Page 14, Line 25: "wetter by a factor of two". With respect to what? GPCC-monthly? To the observations in general? The reference has to be always indicated in a comparative sentence like this one.

- Page 14, Line 27: The sentence "some discrepancies are evident in the seasonality" could be misleading here, since only at line 33 the authors really report on changes in the seasonality, i.e., monthly shifts in some precipitation characteristics.

- Page 15, Line 5 (the whole paragraph). Besides the dry bias of rain gauges (rain-gauges are known to underestimate solid precipitation), one further reason for the wet bias of the reanalysis products (again, compared to GPCC-monthly) could be related to the "model component" of the reanalyses themselves. Models, in fact, are also known to have a wet (and cold) bias in mountains and in the cold season particularly (e.g., Palazzi et al., 2015; Palazzi, E., von Hardenberg, J., Terzago, S. et al. Clim Dyn (2015) 45: 21. https://doi.org/10.1007/s00382-014-2341-z). This should be added somewhere in the text. Reanalyses are a combination of observations+model which means that they can inherit drawbacks and advantages of both of them.

- Page 15, Line 33: "maxima" –> "maximum precipitation values"

- Page 16, Lines 4-5: I don't understand the meaning of the sentence "those errors are [...] yearlong", in particular of the term "consistent"

- Page 16, Line 5: Please consider to change this part "low density observations" with "a low density of observations"

- Page 16, Line 18: "over estimations" should be "overestimations"; "overruled", maybe better "avoided"?

- Page 16, Line 19: I would avoid qualitative expressions like "high to very high", just leave the percent values reported subsequently

- Page 16, Lines 20-21: "The summer mean does not converge", please rephrase this sentence. Do the authors mean that the spread among the various product is large?

- Page 16, Lines 23-24: In my opinion, the sentence starting with "These latest" and ending with "study domain" would be suitable as a final statement of this section.

Subsection 3.2

- Page 20, Line 2: Please add "precipitation" between "daily" and "variability". Same line: it is not clear to me what the concept of "dependency between each dataset" means

- Page 20, Line 4: Please replace "most of the reanalyses" with "of most of the reanalyses"

- Page 20, Line 4-5: the sentence in parentheses is unclear.

- Page 20, Lines 9-10: Please replace "from APHRODITE" to "APHRODITE-2" with "in both APHRODITE products"

- Page 22, Lines 28-29: I would rephrase this sentence: "common dependency of the true variability", in particular I'm not really comfortable with the term "true". The correlation between the two types of datasets can be related to the fact that they represent the precipitation variability at this scale in the same way?

- Page 23, Line9: "analysis of the correlation" –> "correlation analysis". Same line: I would say "ERA-Interim ranks second and is the best performing reanalysis among those which do not assimilate precipitation observations"

- Page 23, Line 11: Please add "version" between "first" and "outperforms". Same line: "century reanalysis" –> "20th century reanalysis" or the correct term for this product.

Subsection 3.3

- Page 32, Lines 3-4: Delete the part of the sentence after "time scale", not useful.

- Page 32, Line 8: "good", should be justified.

- Page 32, Line 12: Please add "the correlation" before "continues" (subject missing

here). Same at Line 14 ("it rises" or "the correlation rises")

- Page 32, Line 19: Remove "feedback". This sentence should be rephrased since it is not easily readable.

Conclusions

- Page 39, Line 3: "six" –> "six datasets are"; "four" –> "four are"

- Page 39, Line 4: "of datasets" –> "of the datasets"; "each"–> "each of them"

- Page 39, Line 5: "true values", an expression that should be avoided. It is quite clear, also from the analysis presented in this paper, that it is not possible to define a ground truth for precipitation, at least in this area.

- Page 39, Line 14: is there any reference to be cited in support to the statement about teleconnections?

- Page 39, Line 16: I would express the concept the other way around. For example "The quality of the datasets also depends on the season which is analysed"

- Page 39, Line 32: "CPC is also a dry dataset", I would rather say the "CPC exhibits a dry bias compared to ...."

- Page 40, Line 2. Is the word "There" at the beginning of the sentence used to say "In this case" (i.e., in the lower Indus)? I prefer "In this case" than "There".

- Page 40, last sentence: I suggest to rephrase this sentence, especially avoiding expressions like "while reanalyses are even worse". There are other ways to say that uncertainties remain. I would point more toward the lesson learned in this paper, with a more, let's say, positive view. That sentence is really sharp.