AUTHORS' RESPONSE TO REVIEWER AND 2

1st REVIEWER

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:

1) 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?

Reply: In this study, the analysis of the quality of each dataset is limited to the precipitation in the Indus basin. We do not make presumptions about the quality of the datasets in other areas, nor other variables for the reanalyses. We have removed the term "rank", particularly when performing the cross-correlation analysis (Section 3.3.2) as the reader may understand it in an absolute sense. We instead specify which datasets perform best for which measures.
Nevertheless, we can infer that, for example, reanalyses that represent the precipitation in the Indus basin better also represent circulation patterns and the processes involved in the generation of precipitation better.
Lastly, we specifically reorganised the conclusion around the key messages we want to convey. These are as follows:

- **The method we have used gives detailed information on the strengths and limitations of each of the 20 datasets investigated.**
- **There are large uncertainties, especially if considering all datasets equally. However, by evaluating the strengths and limitations of each dataset, we have found that some stand out as being of much higher quality which eventually helps to reduce the uncertainty.**
- **Particularly, progress in reanalysis products is real, with one (ERA5) scoring as high as the observations for all measures.**

These reanalyses offer a different point of view than the observations, which is useful for estimating the uncertainty, and can even be used to some extent to validate observations.

- **We also emphasise on the need to systematically adjust rain gauge measurements to account for precipitation undercatchment.**

TEXT MODIFIED

- **For changes in Section 3.3.2 on cross-validation of Daily variability, which was part of Section 3.2 Daily variability in the reviewed document, see answer to general comment 8 of the first reviewer**

- **Changes in the Conclusion are shown below in bold:**

"In this study, we have compared a large number of precipitation datasets of different types across two distinct zones of the Indus watershed: six datasets are based only on rain gauges, four are derived from satellite observations, and ten from reanalysis. We have shown that the number and diversity of the datasets help to identify and quantify the limitations and abilities of each of them, which in turn enables a better estimation of the **uncertainty**.

We have compared the datasets on the basis of annual mean precipitation, the seasonal cycle, as well as the variability over time scales from one day to 10 years. We have relied on the literature to evaluate the different sources of uncertainty **and have interpreted the mean differences between datasets in terms of their quality. We have suggested that the similarities in variability can directly be interpreted in terms of quality, especially when comparing datasets with no common methods or data source. Most reanalyses do not assimilate precipitation observations, which makes it possible to cross-validate between observational and reanalysis data based on variability. Regardless of the observational datasets used as a reference, we have found that some reanalyses have significantly higher correlation with that reference than other reanalyses, which we have interpreted as a sign of good quality. Conversely, when using a reanalysis as a reference, some observational datasets have significantly higher correlation than others. The use of reanalyses to validate observational datasets is justified by the quality of reanalysis products demonstrated in this study. Specifically, at the scale of the Indus basin, and for the daily variability, the same level of similarity between the reanalyses and observations is also seen between the observational datasets themselves.**

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 lead 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 rates can explain some of the biases.

One of our findings concerns the important uncertainty in fine scale spatial patterns of precipitation, particularly in the upper Indus, where precipitation is the most heterogeneous. Important discrepancies remain between datasets, which explain part of the differences in mean precipitation. This issue needs to be tackled in observational datasets by including more measurements and by updating the climatology used in the interpolation methods. In reanalysis products, higher resolution and better modelling of the small scale processes are likely needed to improve confidence in the spatial pattern of precipitation. In this study, we have deliberately selected two large study areas, which has increased the confidence in the datasets. Area-wide correlation particularly improves the significance of the variability analysis, compared to a point-wise correlation.

We have also found that the quality of the datasets depends on the season. Rain gauge measurements suffer from important underestimations in winter for the upper Indus. Most satellite-derived datasets even further amplify this bias. By contrast, reanalyses perform best during winter. Particularly, the most recent reanalyses produce a very similar amount of winter precipitation and its variability is similar to the observations at all timescales. We have suggested that this amount of precipitation is closer to reality than the observations, although some overestimations are possible, due to, for example, misrepresentation of the lowest precipitation rates. Summer precipitation, in both study areas, is much more uncertain in the reanalyses in total amount, seasonality, and variability. In contrast, satellite observations perform better in summer than in winter and seem to bring additional information to rain gauge measurements.

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

More specifically, APHRODITE is the best **observational** dataset for daily and monthly variability, thanks to a large number of observations in the whole basin. However, it also exhibits drier conditions than most of the other datasets, which is partially caused by the interpolation method it uses and possibly by a lower quality of the data. Surprisingly, APHRODITE-2 is not as good, especially for the longer term variability, as it removes some observations in areas with an already lower density of measurements. **CPC is the least reliable observational dataset, particularly for the upper Indus, with a large dry bias compared to GPCC-monthly, the lowest correlation scores at all time scale, and an error on the dates before 1998. However, its quality significantly improves after 2005, which, we**

suspect, is due to a change in the quality of the data source. GPCC-monthly is one of the most reliable datasets **both in terms of amount and variability. GPCC-daily relies on GPCC-monthly for its monthly mean. The very low number of daily measurements included in the early part of the covered period limits its quality, but this quickly improves as more observations are included.**

Satellite-based datasets are very dependent on the quality of the rain-gauge product they integrate. The added-value of satellite observations remains limited at the basin scale. **The signal is degraded during winter for the upper Indus, while better results in the lower Indus suggest slightly wetter conditions than the rain gauge-based datasets. Importantly, the quality of satellite-based datasets resides in their near real time availability as well as their higher temporal and spatial resolution than rain gauge based datasets.**

The quality of reanalysis datasets has clearly improved since the first datasets were released. ERA5 is the latest reanalysis and clearly stands out as the one representing best the observations, in terms of amount, seasonality, and variability at all time scales investigated. Remarkably, it is the only reanalysis representing the decadal variability of the summer precipitation that is seen in the observations in both study areas. Furthermore, for the daily to inter-annual variability, the best performing observational dataset has often a better level of similarity with ERA5 than with other observational datasets. Some of these qualities can be derived from its high resolution, which allows the representation of interesting fine scale features, as well as the assimilation of precipitation measurements.

After ERA5, ERA-Interim, MERRA1 and MERRA2 have relatively similar performance. **Reichle et al. (2017) showed that the soil moisture content was not improved over South Asia from MERRA1 to MERRA2, neither in terms of variability nor biases, despite the use of CPC to correct the precipitation input to the land surface model of MERRA2. Given the difficulties of CPC to represent precipitationin the Indus basin, correcting the modelled precipitation with this dataset probably does not improve the signal. In this study, we were able to show that the correction with CPC feeds back locally on the modelled precipitation, particularly at the monthly scale for the upper Indus. We have also suggested that the dry bias of MERRA2 in the lower Indus, and the decreased score on the daily variability compared to MERRA1, is also due to that correction.**

The confidence in JRA's precipitation in the upper Indus is generally high, but drops for the daily and monthly variability in the 1990s. By contrast, it represents overly wet conditions for the lower Indus. CFSR has problems reproducing the daily variability and the seasonality of the monsoon, especially in the upper Indus. This is probably improved by the latest version that started in April 2011. However, it would likely be better to treat the two versions separately as it seems the new version produces somewhat different statistics of precipitation. The twentieth century reanalyses, **which includes** only surface **observations,** are not as good as the others, especially in winter. However, while 20CR barely reproduces any of the variability depicted by the observation, ERA-20C has much better capabilities, close to NCEP1 and CFSR, especially during summer. Neither 20CR nor ERA-20C represent the decadal variability as shown by the observation before 1980.

Finally, large uncertainties remain about precipitation in the upper Indus, but one should not treat all datasets equally. We have demonstrated that specific datasets represent the precipitation better, which helps to narrow down the uncertainty. Particularly, we have argued that reanalyses and observational datasets can both be useful for cross-validation. They can also be used for quality monitoring. Daily correlation of precipitation for key areas can be performed between a series of datasets with near real time updates. Changes in correlation between one or several datasets would therefore highlight a change in quality that would need to be investigated."

2) 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.

Answer: analysing a multi-dataset mean does not provide more information about the quality of the datasets considered, and is therefore slightly outside the scope of the study. Such a mean is often used when uncertainties are very large and a best dataset cannot be selected (e.g. a multi-model mean for future simulation of the climate). In our study, however, we found that both APHRODITE and GPCC-monthly for the observation and ERA5 for the reanalyses were performing significantly better than other datasets. Furthermore, all datasets covered different periods, which complicates the use of a mean in practical situations.
Nevertheless, we checked if a mean can bring better results in terms of daily variability for the period 1998-2007. We considered the mean of the 6 observational datasets available at daily resolution, as well as the mean of ERA5, JRA, MERRA2 and CFSR (which are the most recent reanalyses). The correlation of these means against a reference are higher than that of most of the datasets composing them. However, the best datasets still have better scores. There is one exception in the lower Indus, as the mean of the observations performs significantly better than any of the individual observational datasets. In that domain, all observational datasets have very similar scores, and the mean is able to further improve these scores. These results seem to be too specific to be included in the study.

3) 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?

Answer: This study does not evaluate the underestimation of rain gauge measurements, but rather suggests that this underestimation is the main cause of the differences between reanalyses and observations. Precipitation estimates from reanalyses are likely closer to the real amounts, but they may also be overestimated, and we cannot quantify this with our method. We rather urge a systematic correction of rain gauge measurements, using similar techniques as was used by Dahri et al. (2018), as this seems to be the best way to evaluate underestimations.

TEXT MODIFIED

See 6th paragraph of the updated conclusion in answer to general comment 1 of 1st reviewer

4) 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.

Answer: We added a new figure that gives the elevation, as suggested in comment 3 of the 2nd reviewer. It is used to introduce the domain of study.

TEXT MODIFIED:

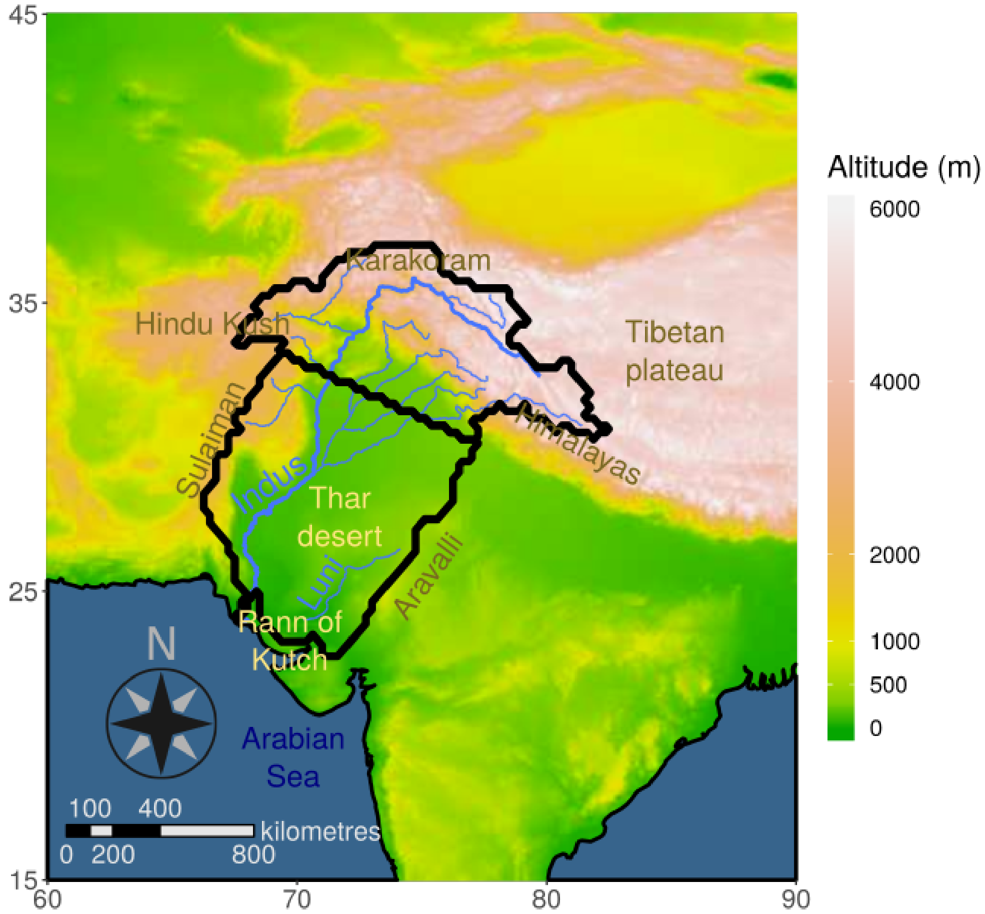
- **caption to figure 1:**

"Relief and topographical features in and around the area investigated. The thick outer black contour represent the watershed on the Indus and Luni rivers. This area is split to form the two study areas: the upper Indus to the north, and the lower Indus to the south."

- **At the start of section 2.1 "Study areas", we have added (in bold):**

"The Indus River basin extends across the north-westernmost part of the South Asian sub-continent, and is an area of various topographic features, as indicated in figure 1."

FIGURE ADDED



5) 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?

Answer: All rain-gauge datasets are based on a bi-linear interpolation of the station measurements, which justify the use of a bi-linear interpolation here. The only difference is that it is not the direct measurements that are interpolated in these observational datasets, but its anomaly against a climatology. However, using a specific climatology here would bias the validation.
Another, more conservative approach is to select the grid points whose centre is within the area of study. However, this would lead to small changes on the area being considered and precipitation biases. These biases are partly eliminated by the bi-linear interpolation.

TEXT MODIFIED

- **The whole of Section 2.3 Methods has been updated in light of other comments. The changes are in bold, the one considering this specific comment are in the first paragraph:**

"For each dataset, the time series of precipitation **are** averaged over the two **study areas (upper and lower Indus) and** calculated at a monthly resolution, and daily if possible. **The datasets have different spatial resolution which causes a problem when calculating the precipitation averages over the study areas. Simply selecting the cells whose centre is within these areas leads to small biases in the extent of the region considered. These biases are reduced by bi-linearly interpolating all data to a 0.25° grid, common to APHRODITE, APHRODITE-2, and GPCC-monthly. This choice is further discussed in section 3.1.1."**

The analysis is performed over the 10-year period from 1998-2007, which is common to all datasets, **except when analysing the trends and inter-annual to decadal variability, for which we use all data available. We focus on the two wet seasons of the upper Indus.** Summer is defined from June to September, which matches the monsoon precipitation peak. Winter is defined from December to March. This fits the snowfall peak rather than the precipitation peak, but makes it possible to focus on issues of snowfall estimation (Palazzi et al. 2013). **In the lower Indus, we use the same definition of summer, but winter is not analysed, as it is a dry season.**

We first compare the mean and seasonal cycle of each dataset in sections 3.1 and 3.2. There, for illustrative purposes, we make quantitative statements using GPCC-monthly as a reference. However, in section 3.1.3, we use the precipitation dataset from Dahri et al. (2018) as reference instead. This dataset cannot be used in other parts of the study, as it is limited to one part of the upper Indus, and only provides annual means.

Then, in section 3.3 we compare the daily variability of the precipitation using the Pearson correlation. The correlation significance is discussed at the 95% probability level. To reduce the impact of abnormally large rainfall events when investigating the trend (cf. Section 3.3.4), we use the Spearman correlation. Lastly, in section 3.4, other time scales of variability of the precipitation are investigated: monthly, seasonal, inter-annual, and decadal, still using the Pearson correlation at the 95% confidence interval."

- **We also added a paragraph at the start of the Result section, on the differences of annual mean precipitation among rain gauges datasets (3.1.1), just after the introduction of the figures (cf. answer on comment 8 of the first reviewer)**

6) 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 “variability”? Is that year-to-year variability? Daily? Something else? This must be better specified in the Methods before going to the Results section.

Answer: the different time scales investigated are presented one paragraph above in that section, which led to some confusion for the readers. This has been corrected in the updated methodology (cf. Answer to the general comment 5 of the first reviewer).

7) 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)

Answer: We added information about the use of references in the method section. We have also clarified in which part of the result sections which reference is used. Lastly, the discussion of figure 2 (now figure 3) on seasonality is performed differently and focuses only on seasonality and not on annual mean differences.

TEXT MODIFIED

- **3rd paragraph of the reviewed Method section (cf. Answer to the general comment 5 of the first reviewer)**
- **The figure 3 on seasonality is discussed in a specific section (3.2, see answer to general comment 8 of the first reviewer)**

8) 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).

Answer: As suggested by the second reviewer, we split section 3.1 and 3.2 into several subsections (See comment 11 of the second reviewer).

The headings are now:

3.1 Annual mean
3.1.1 Differences between rain gauges-based datasets
3.1.2 Considering satellite and reanalysis datasets
3.1.3 Impact of rain gauge biases in mountainous terrains

3.2 Seasonal cycle

3.3 Daily variability
3.3.1 Lag analysis
3.3.2 Cross-validation
3.3.3 Influence of the seasonality
3.3.4 Trends

We particularly disentangled the analysis of the differences in annual mean on one hand, and the seasonal and monthly mean on the other. This improves the discussion of figure 3 on seasonality (cf general comment 7 and specific comment on Page 14, Line 27 of the first reviewer). We also put the comparison with Darhi et al. 2018 results in a specific subsection. Lastly, we deleted some redundant text.

TEXT MODIFIED

- **We mention the use of Dahri et al. 2018 as a reference in the method section (cf. answer to general comment 5 of the first reviewer)**
- **New section 3.1 and 3.2 are now as follow, with the changes in bold:**

"

3. Results

3.1 Annual mean

3.1.1 Differences among rain gauges-based datasets

Annual mean precipitation in both domains and for each dataset are given in Table 4 (last two columns). We first focus on the rain gauge-

based datasets (upper part of the table). Spatial pattern differences are shown in Figure 2-A to E.
First, we should mention that the bi-linear method we use to interpolate each dataset to the same grid (cf. Subsection 2.3) leads to some differences between datasets. The two GPCC products can be used to evaluate the impact of our interpolation method, as they have a different spatial resolution but are based on the same climatology. Hence, the small underestimation of GPCC-daily compared to GPCC-monthly (about 1% in the upper Indus and 5% in the lower Indus) should be related to the interpolation method. However, these differences are small enough to justify the use of our method.

More generally, annual mean differences can be explained by methods and data that each dataset uses. Particularly, the interpolation of station measurements to a grid differs from one dataset to the other. APHRODITE's interpolation method, for instance, considers the orientation of the slope to quantify the influence of nearby stations. This greatly reduces the amount of precipitation falling in the inner mountains compared to GPCC-monthly. An example of this pattern is evident to the North of the Himalayas where only very few observations exist (Figure 2-D; Yatagai et al. (2012). In CRU, the interpolation method (triangulated linear interpolation of anomalies; Harris et al. (2014) seems to smooth areas of strong gradients such as near the foothills of the Himalayas (Figure 2-B). This smoothing might explain a slightly drier upper Indus, and slightly wetter lower Indus, compared to GPCC-monthly (Table 4).

Differences can also be explained by the dramatic change in location and number of stations used to compute the statistics (Figure 2-A, C, D, and E, Table 2). **For example, CPC is by far the driest dataset in the upper Indus and the second driest in the lower Indus. This is likely related to the low number of observations it includes,** leaving vast areas with no or very few observations, including the wettest regions (Figure 2-E). However, there is no linear **relationship** between precipitation **amount** and number of observations. GPCC-daily includes the lowest number of observations**, but this does not impact its climatology, because the climatology is derived from GPCC-monthly**. On the contrary, APHRODITE **comprises** a much higher number of observations than other datasets, but remains much drier than GPCC-monthly **(about 20% drier in both study areas)**.

Yatagai et al. (2012) pointed out that differences in quality checks compared to the other datasets might explain APHRODITE’s **dry bias. They noted that APHRODITE partly** relies on GTS data that are sent in near real time to the global network**, with the risk of misreporting. This risk particularly concerns misreported zero values, which are hard to detect and lead to a dry bias. The large dry bias seen in CPC data migh be associated to the same issue, since CPC is entirely based on GTS data**. In GPCC-monthly (and daily), only stations with at least 70% of data per month are retained (Schneider et al 2014), while in CRU this number is 75% (Harris et al, 2014). Thus, limiting the analysis to the most reliable weather stations can help build confidence in recorded total precipitation amount.

Interestingly, APHRODITE-2 is **more than 10% wetter** than APHRODITE **in both study areas**. Several changes have been performed in the methodology: quality control of extreme high values, station-value conservation after interpolation, merging daily observation with different definitions of **End of Day time (cf. Section 3.1.1)**, and an updated climatology. However, the difference in mean precipitation is **most likely** related to the change in observations from rain gauges. **Although APHRODITE-2 comprises more observations basin-wide,** this increase mainly occurs over the Indian territory, whereas Pakistan is presented with fewer precipitation measurements, especially in the dry southern central part (Figure 2-D). This decrease in observations in a drier area can reasonably explain the increase in mean precipitation in the lower Indus. In the upper Indus, the increase is mainly due to the inclusion of measurements from one isolated weather station in the northernmost part of **the area.**

3.1.2 Considering the other datasets

We now consider satellite-based datasets (middle part of the table 4). In the upper Indus, CMAP stands out as being the wettest observational datasets, 13% wetter than GPCC-monthly. By contrast, the other three (TMPA, GPCC-1DD, GPCP-SG), are drier than GPCC-monthly (between 10 and 5%), despite being calibrated by this GPCC-monthly. In the lower Indus, all satellite-derived datasets are wetter than the rain gauge products (between 10 and 30% more than GPCC-monthly). The complexity of the algorithm used to produce the satellite-based datasets makes determining the reasons for their differences with each other or with rain gauge products difficult. According to previous studies, their ability to represent precipitation over rough terrain is limited (e.g. Hussain et al. (2017). **In fact, figure 2-F shows that the strongest differences between TMPA and GPCC-monthly occurs near mountain ranges, such as the upper Indus**. In contrast, precipitation estimates over flat terrain with sparse observations and mostly convective precipitation benefit from satellite observations (Ebert et al. 2007). This suggests that **the** higher precipitation mean of the satellite-derived datasets for the lower Indus is possibly due to better consideration of locally higher precipitation rates during convective events.

The annual mean precipitation in reanalysis datasets is listed in the lower part of table 4. In the lower Indus, the range of values is very high: the wettest dataset, JRA, is five times wetter than the driest dataset, 20CR. This range shows the significant difficulties for reanalyses to represent precipitation in an area were convection dominates. Among the most recent reanalyses, ERA5 has the closest precipitation estimates to the observational datasets, yet above the estimates from rain gauges. Figure 2-H suggests that these wetter conditions mainly comes from the north-western edge of the Suleiman range, an area with sparse precipitation observation (cf. Figure 2-A), therefore increasing confidence in ERA5 estimation. The two twentieth century reanalysis (20CR and ERA-20C) are amongst the driest reanalysis datasets, suggesting that their models have difficulties to

propagate the monsoon precipitation into the lower Indus region, when only surface observations are assimilated. Lastly, MERRA2 exhibits a severe drop of precipitation compared to the previous version, MERRA1. Summer monsoon precipitation **is known to be** strongly affected by surface moisture content, especially in flat areas like **the lower Indus** (Douville et al. 2001)**. MERRA2 uses CPC data to constrain the precipitation flux at the surface. Due to the dry bias of CPC, soil moisture is reduced for most of India (Figure 3 in Reichle et al. 2017), explaining the drop in precipitation.**

For the upper Indus, the most striking features is that all reanalysis datasets except MERRA1 and ERA-20C predict higher precipitation amounts than GPCC-monthly, about 20% higher on average. In the following we investigate whether this difference can be explained by an underestimation of rain gauge measurements.

3.1.3 Impact of rain gauge biases in mountainous terrains

Rain gauge measurements are known to potentially underestimate precipitation and particularly snowfall (Sevruk et al. 1984, Goodison et al. 1989). This is an important issue for mountainous regions such as the upper Indus. However, among the six rain gauge-based datasets, only GPCC's products consider a correction of the data. Based on a study by Legates et al (1990), a correction factor, which depends on the month, is applied at each grid cell. Most of these factors vary between 5 and 10% (Figure 4 in Schneider et al. 2014), and explain why GPCC's products are wetter than most of the other rain gauge-based datasets. Recently, Dahri et al. (2018, hereafter Dahri2018) compiled the measurements from over 270 rain gauges in the upper Indus and adjusted their **values** to undercatchment, following WMO guidelines. They found a basin-wide adjustment of 21%, but this varies from 65% for high altitude stations, to around 1% for the stations in the plains.

The Dahri2018 dataset has both the advantage of considering a very large number of observations and correcting rain gauge measurements. However, its result is based on a study area somewhat smaller than the upper Indus **region** presented here, and only covers the period from 1999 to 2011. **For comparison purpose**, we recomputed the annual mean of several of the most recent and highest resolution datasets to fit these definitions (Table 5).

Table 5 shows that none of the observational datasets is able to reproduce the Dahri2018 precipitation estimates. They all have a dry bias, from 30% for TMPA, to 10% for GPCC-monthly. Furthermore, APHRODITE-2 and TMPA even underestimate the unadjusted value of Dahri2018, which suggests that the underestimation is not only related to rain gauge measurements, but also to the representation of the spatial pattern. By contrast, GPCC-monthly is 7% higher than the Dahri2018 unadjusted values, which corresponds to the correction factor used in GPCC. This suggests that the unadjusted values in both datasets are very close, **and highlights quality of GPCC. Nevertheless, we** also found discrepancies in the spatial patterns between GPCC-monthly and Dahri2018. Particularly, in the northernmost part of the upper Indus **region**, in the Karakoram range, GPCC-monthly exhibits lower precipitation means than Dahri2018, which cannot be explained by the difference in **correction factors** between the two datasets alone. The nearest stations used in GPCC-monthly are all located in the dry and more accessible Indus River valley to the south of the mountain range (Figure 2-A). Those drier conditions extend to the north due to the interpolation method used by GPCC, while Dahri2018 includes station measurements with wetter conditions than in the valley. This difference illustrates the impact of biased weather station locations mentioned in the introduction and in several other studies (e.g. Archer and Fowler, 2004; Ménégoz et al., 2013; Immerzeel et al., 2015).

Still in the Karakoram range, figures 2-G and E show that MERRA2 and ERA5 are wetter than GPCC-monthly, and therefore closer to Dahri2018. However, spatial discrepancies remain. Particularly, the maximum of precipitation in MERRA2 is shifted to the North, which leads to important biases when averaging on the Dahri2018's study area. Our domain of study, which does not overlap with the highest precipitation rates, is less affected by shifts and is better fitted to compare the large scale precipitation patterns. Nevertheless, the four selected reanalysis datasets in Table 5 overestimate the Dahri2018 adjusted values, by 20% on average. This suggests that part but not all of the differences between reanalysis and observational data can be explained by biases from the latter. This overestimation of modelled precipitation in reanalyses for the upper Indusis corroborated by previous studies (e.g. Palazzi et al., 2015).

To conclude, all rain gauge-based datasets suffer from **an underestimation of** annual mean precipitation for the upper Indus when compared to Darhi2018. **This results from biases in rain gauge locations and measurements**. Quality control and interpolation **methods** also impact **precipitation amount in** both parts of the basin. Satellite observations probably improve precipitation estimates in flat areas with sparse observations. However, **they** cannot correct observational biases since they use them for calibration, and biases remain unchanged **or even amplify** for the upper Indus. **Reanalyses do not include rain gauge measurement**, except for ERA5 and MERRA2, and are therefore not affected by observational **biases. However, model biases can also be significant as suggested by the spread of the annual mean precipitation values. Reanalyses tend to be wetter than observational datasets in the upper Indus, which is partly explained by the underestimation of the observations. Lastly, all datasets suffer from spatial discrepancies, which are** detrimental to small-scale comparisons, especially near mountains, but justify our choice to use a larger study **area**.

3.2 Seasonal cycle

The seasonal cycle of precipitation for each dataset is presented in Figure 3. Analysing the seasonality is particularly interesting in the

upper Indus, as it is characterised by two wet seasons. The mean precipitation of each season is presented in table 4 (second and third column). The rain gauge-based datasets exhibit a very similar seasonality for both study areas. In the upper Indus, the maxima of precipitation occur in February and July, the minima in May and November. The differences between the datasets vary little from one month to another, which suggests that the causes of the differences identified in the previous section (e.g. misreporting, station location and number, interpolation method) are independent of the seasonality. The satellite-based datasets represent the summer precipitation almost exactly as GPCC-monthly. The annual mean differences are explained by biases during the winter season, which suggests that winter precipitation is more difficult to estimate for those datasets.

The reanalyses represent the dry and wet seasons of the upper Indus, but with a larger spread than in the observations and some differences in seasonal cycle (Figure 3-B). On average, winter precipitation is 30% higher than in GPCC-monthly, with the notable exception of ERA-20C (Table 4). Those wetter conditions also extend to the surrounding drier months: April/May and October/November. **However**, the mean summer precipitation in reanalyses is not significantly **different from** GPCC-monthly (Table 4). Only ERA-Interim stands out with a wet summer precipitation bias, mainly in the north-west corner of the upper Indus domain, a bias partly corrected in ERA5 (Figure 2-H). **The winter wet bias is not surprising after the comparison with the Dahri2018 dataset in the section 3.1.3. Indeed, Dahri2018 found that the most important rain gauge underestimations happen in winter when precipitation mostly falls as snow. More interestingly, we found that the latest reanalyses (ERA5, JRA, MERRA2, and CFSR) represent winter precipitation in similar ways. We haven't been able to investigate the seasonality of the Dahri2018 dataset, but we suggest that the latest reanalyses better represent winter precipitation than the observational datasets.**

We noted another discrepancy **in seasonality** between a majority of the reanalyses and the observations for the upper Indus: a delay of the summer precipitation starting from the pre-monsoon season (Figure 3-B). The observations show that May is the driest month of that season followed by a sharp increase in precipitation in June. Only ERA5, ERA25 Interim, and MERRA1 reproduce this behaviour. In contrast, NCEP2 and CFSR are much drier in June than in May. For other reanalyses, precipitation during May and June are comparable. This delay continues into the summer monsoon period: while the observations clearly show a wetter July than August, this is only the case for ERA5, ERA-Interim, and both MERRA reanalyses. A similar delay can be found over the Ganges plain and along the Himalayas, which suggests wider uncertainties on the monsoon propagation in the reanalyses. **By contrast, no such delay is found in the lower Indus, despite the large uncertainty on the amount of precipitation (Figure 3-D)**

9) 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.

Answer: we added or changed the sentences that introduce each figures before discussing them, if this was not done properly

TEXT MODIFIED:

- **Figure 1: "**The Indus River basin extends across the north-westernmost part of the South Asian sub-continent, **an area of various topographic features, as represented in figure 1."**
- **Figure 2: "Precipitation amount varies across the basin as shown in Figure 2-A."**
- **Table 1 and 2: "**We have selected five commonly used and one newly available gridded dataset based only on rain gauge data. **These are the** first six datasets **presented** in Table 1**. The mean number of stations used in the two study areas are available for five of the datasets and presented in Table 2."**
- **Table 3: "Table 3 shows the ensemble of the ten reanalysis datasets that have been used in this study."**
- **Table 4 and rest of Figure 2: "Annual mean of precipitation in both domains and for each dataset are given in Table 4 (last two columns). We first focus on the rain gauge-based datasets (upper part of the table). We found an important range of values in both study areas, which are related to discrepancies in the precipitation spatial pattern as presented in Figure 2-A to E."**
- **Figure 3: "The seasonal cycle of precipitation for each datasets is presented in figure 3."**
- **Table 6: "Table 6 presents the daily correlation of precipitation between the different datasets, for the upper Indus. The upper part of the table focuses on the cross-correlation between the observational datasets."**
- **Table 7: "The same correlation analysis is performed for the lower Indus domain (Table 7)."**
- **Figure 5: "Figure 5 presents the seasonality, for the upper Indus, of the correlations between the reanalyses and APHRODITE-2."**

10) 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.

Answer: We have found 3 occurrences of the words “consistent” and “consistency”. The first disappeared during the transformation of the first result section. In the other two cases, it had a meaning of stability.

TEXT MODIFIED

- “Correlations between JRA and APHRODITE remains mostly between 0.8 and 0.85. ERA-20C is also **fairly stable** over time, generally above NCEP1. 20CR, by contrast, exhibits a much higher variability with correlation dropping as low as 0.4 at times, and sometimes reaching NCEP1”
- “We found **relatively stable correlations** with APHRODITE and CRU during the twentieth century”

11) 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.

Answer: This is one of the key messages of the conclusion and is further discussed there (cf. answer to comment 1 of the 1st reviewer). In essence, both reanalysis and gridded observational data are estimates of the actual precipitation. Reanalyses rely the understanding of processes and a large ensemble of observations, while observational datasets only rely on precipitation measurements in specific locations. Their difference of approach makes them differently dependent on different sources of uncertainty, which allow a cross-validation. Importantly, the cross-validation is limited in our study to the precipitation variability at daily and monthly time scale and at the scale of the study areas. It is justified by the fact the most recent reanalyses represent variability within the range of uncertainty given by the different observational datasets. That is, these reanalyses represent the daily and monthly variability of the precipitation at least as well as the observational datasets.
Validation of the amount of precipitation is more complicated and needs a thorough understanding of the source of uncertainty in each dataset. Furthermore, we do not assume that the validation can be performed for other variables.

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 [...]”
Answer: Agreed, changed as suggested.

Page 1, Line 5: Add “data” after “reanalysis”
Answer: added

Page 1, Line 10: Please replace “most able” with “most performing”, or similar
Answer: we further clarified the sentence to “ERA5 is the reanalysis that offers estimates of precipitation closest to observations, “

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
Answer: “Correction” corresponds here to a factor by which the raw value are multiplied. “small” refers to that factor. We added the term “factor” twice in the sentence:
“GPCC products are the only datasets that include a correction **factor** of the rain gauge measurements but **this factor** remains likely too small”

Introduction

Page 2, Line 13: Add “of” between “use” and “rain gauges”
Answer: added

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
Answer: “2” is now a superscript.

Page 3, Line 14: Add “the” between “and” and “heterogeneity”
Answer: added

Page 3, Line 17: Remove “a” between “over” and “flat”
Answer: removed

Page 3, Line 18: Replace “those” at the end of the sentence with “station data”
Answer: replaced

Page 3, Line 21: Replace “case” with “cases”
Answer: replaced

Page 3, Line 23: Replace “consider” with “highlight” ; “reanalysis” with “reanalyses”and “observation” with “observations”
Answer: all replaced

Page 3, Line 25: Please change “has it made possible” with “has made it possible”
Answer: changed

Page 3, Line 26: Add “product” after “reanalysis”
Answer: added

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.
Answer: Agreed, deleted.

Page 3, Line 30: Please replace “qualities” with e.g. “strengths and limitations”
Answer: changed as suggested

Page 3, Line 31: Please replace “have” with “has”
Answer: replaced

Page 3, Line 34: Please replace “method” with “methods”
Answer: replaced

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.
Answer: we have added a reference to the precipitation, as well as a sentence at the end of the paragraph in the conclusion section. We have updated the text to conform with the four (instead of three) subsections that now compose the Results section.

TEXT MODIFIED
“ The subsequent result section is split into **four** parts, which review, **for the precipitation: i) the annual mean, ii) the seasonality, iii) the** daily variability, and **iv) the** monthly and longer term variability. **The final section concludes with the main results, the potential of the method, and future research priorities.”**

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 10-11above)*

Answer: This is a misunderstanding, we do not provide the contour of the watershed of the Luni River. However, the Luni River is represented in Figure 1 (new figure, see answer to general comment 4 of the 1st reviewer), which illustrates our point.

The reference to the figure was misplaced and confusing. We have added a new sentence at the end of that paragraph to clarify the reference to the whole watershed.

TEXT MODIFIED

"**It may also** forms seasonal rivers, such as the Luni River, which has been included in the **study areas**. This particular river flows into the Rann of Kutch, which is a flat salt marsh with complex connections with the Arabian Sea and the mouth of the Indus River (Syvitski et al., 2013), and is bounded on the west by the Aravalli Range. Although not strictly a part of the Indus watershed, it provides a clear and steady boundary for the domain of study. **The total watershed considered for the study is represented by the outer black line shown in figure 1**"

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

Answer: replaced

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"

Answer: replaced

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

Answer: replaced as suggested

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 Harden-berg, 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)

Answer: indeed, replaced as suggested. The reference has been added

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

Answer: good point, removed

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.

Answer: wrong phrasing indeed. We have modified the sentence.

TEXT MODIFIED

"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 km²) is characterised **by a single wet season during summer, as wintertime precipitation is negligible**"

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"

Answer: Agreed, changed accordingly

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

Answer: thanks, removed

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

Answer: replaced

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

Answer: the sentence has been changed to: "A new dataset has been issued in 2019 from the same institute **extending the period covered up to 2015**"

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

Answer: replaced

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

Answer: replaced

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

Answer: we referred to the amount and type of observations included in the datasets.

TEXT MODIFIED

"It has the highest temporal and spatial resolution of the selection (sub-daily, and 0.25° like APHRODITE and GPCC-monthly) and **includes a large diversity of satellite observations**"

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?

Answer: It does not make sense indeed, all datasets are compared, not particularly those ones. That part of the sentence is removed.

TEXT MODIFIED

"The precipitation dataset from the Climate Research Unit has a similar resolution and time coverage as GPCC-monthly."

Page 6, Line 18: At the beginning of the sentence, please rephrase "All three datasets described above use GPCC …."
Answer: We specified the three datasets in the sentence.

TEXT MODIFIED
"All three of these datasets **(TMPA, GPCP-1DD, and GPGP-SG)** use GPCC for calibration"

Reanalysis

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

Page 9, Line 3: better to say "can vary" rather than "varies"
Answer: agreed, replaced

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".
Answer: Replaced, this is part of the general comment 9 of the 1st reviewer

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".
Answer: We updated the sentence.

TEXT MODIFIED
"ERA5 currently starts in 1979 but future releases are expected to extend this back to 1950."

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

Methods

Page 12, Line 6: Please add "of precipitation" after "seasonality"
Answer: Does not apply in the improved version of that section

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

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"
Answer: we did not add "extremely" here. Being dry is enough not to have analysed that season

TEXT MODIFIED
“In the lower Indus, we use the same definition for summer, but winter is not analysed, as it is a dry season”

Page 12, Line 10: Please add “precipitation” before “time series”
Answer: Does not apply in the improved version of that section

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.
Answer: agreed, the passage is removed

Page 12, Line 22: “as we will discuss....” –> “as discussed in the Results section”
Answer: Does not apply in the improved version of that section

Results

Subsection 3.1

Page 13, Line 3: Add “precipitation” before “seasonal cycle”
Answer: Does not apply in the improved version of that section

Page 13, Line 3: 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.
Answer: Agreed, we have now justified the use of qualitative statement using GPCC as a reference for illustrative purposes. This is now explained in the methods section (cf. answer to general comment 5 of the 1st reviewer).

TEXT MODIFIED
In the method section: ”We first compare the mean and seasonal cycle of each datasets in sections 3.1 and 3.2. For quantitative statements we use GPCC-monthly as a reference”

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”
Answer: Does not apply in the improved version of that section

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.

Answer: To clarify the description, this is now presented in a dedicated section (3.2, cf. answer to general comment 8 of the 1st reviewer). We also focus on the months with a minimum or a maximum of precipitation, as well as on bias specific to a season, or more stable throughout the year.

TEXT MODIFIED
See answer to general comment 8 of the 1st reviewer

Page 13, Line 11: Please replace "inferior to" with "less than"
Answer: Does not apply in the improved version of that section

Page 13, Line 13: I would replace "of mean precipitation" with "of the precipitation annual cycle"
Answer: Does not apply in the improved version of that section

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.
Answer: This referred to the fact that we could have used the grid points whose centre was in the domain, instead of a bi-linear interpolation, which leads to further biases (cf. answer to general comment 5 of the 1st reviewer). But the word "carefully" does not reflect this idea, and is therefore removed. This point is further discussed in a specific paragraph (cf. general comment 5 of the 1st reviewer)

TEXT MODIFIED
See second paragraph of section 3.1.1 in answer to general comment 8 of the 1st reviewer

Page 13, Line 16: I would say that GPCC-daily "uses" GPCC-monthly and not that "is based"
Answer: This sentence is slightly modified in the improved version of the manuscript, and takes account of the comment.

TEXT MODIFIED
"The two GPCC products [...] uses the same climatology."

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.
Answer: Indeed, we have rephrased this point. We expect the datasets to be similar, but some differences exist, which should be related to the interpolation method we used, and this can be investigated.

TEXT MODIFIED
See second paragraph of section 3.1.1 in answer to general comment 8 of the 1st reviewer

Page 13, line 25: CPC is "drier" or driest? Maybe driest is the correct term and this refers to the upper Indus only.
Answer: CPC is the driest dataset for the upper Indus, and the second driest for the lower Indus. This is now explicitly stated in the text.

TEXT MODIFIED
"For example, CPC is by far the driest dataset in the upper Indus and the second driest in the lower Indus. This is likely related to the low number of observations it includes, leaving vast areas with no or very few observations, including the wettest regions (Figure 2-E)"

Page 13, Line 26: "linear relation"? I would rather say "clear correlation"
Answer: Correlation refers to a specific statistical tool, which we did not use here. We change the word "relation" to "relationship"

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

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.
Answer: There is a misuse of the word "treating", we have replaced it with "misreporting". The whole sentence was not very clear either. So, zero values can be reported by mistake, and the quality checks do not identify them as missing values. This is clarified in the text.

TEXT MODIFIED
"They noted that APHRODITE partly relies on GTS data that are sent in near real time to the global network**, with risks of misreporting. The risk particularly concerns misreported zero values, harder to detect and which could lead to a dry bias. The large dry bias seen in CPC data could be associated with the same issue, since CPC is entirely based on GTS data"**

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

Answer: replaced

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.
Answer: This is a minor point and it is removed from the text.

Page 14, Line 8: Please change "Different" with "Several" or "Various"
Answer: We used the word "several" as suggested.

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

TEXT MODIFIED
"However, the difference in mean precipitation is most likely related to the change in observations from rain gauges. **Although the APHRODITE-2 comprises more observations basin-wide**, this increase mainly happens over Indian territory."

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

Page 14, Line 24: Remove "somewhat"
Answer: Does not apply in the improved version of that section

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.
Answer: We modified the sentence so we actually compare JRA with 20CR, making the point about the large spread of values.

TEXT MODIFIED
"the wettest dataset, JRA, is five times wetter than the driest dataset, 20CR"

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.
Answer: This comment lead us to split the subsection "Seasonal cycles and annual means" in two, one addressing the annual means and biases, and the second the seasonal cycles and seasonal biases. See answer to comment 7 of the 2nd reviewer.

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. ClimDyn (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.
Answer: Agreed, and this is also suggested by the fact reanalysis overestimate the annual amount found in Dahri2018.

TEXT MODIFIED

- **In section 3.1.3 (See answer to general comment 8 of the 1st reviewer)**

"Nevertheless, the four selected reanalysis datasets in Table 5 overestimate the Dahri2018 adjusted value, by 20% on average. This suggests that part but not all of the differences between reanalysis and observational data can be explained by biases from the latter. Modelled precipitation in reanalysis are likely overestimated in the upper Indus, which corroborates results from previous studies (e.g. Palazzi et al., 2015). "

- **Further in that section:**

"Reanalyses tend to be wetter than observational datasets in the upper Indus, which is partly explained by the underestimation of the observations."

Page 15, Line 33: "maxima" –> "maximum precipitation values"
Answer: Does not apply in the improved version of that section

Page 16, Lines 4-5: I don't understand the meaning of the sentence "those errors are[...] yearlong", in particular of the term "consistent"
Answer: We meant that the errors are relatively independent of the season. It is removed in the improved version of that section

Page 16, Line 5: Please consider to change this part "low density observations" with"a low density of observations"
Answer: Does not apply in the improved version of that section

Page 16, Line 18: "over estimations" should be "overestimations"; "overruled", maybe better "avoided"?
Answer: Does not apply in the improved version of that section

Page 16, Line 19: I would avoid qualitative expressions like "high to very high", just leave the percent values reported subsequently
Answer: Agreed, this is removed in the improved version of that section

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?
Answer: Does not apply in the improved version of that section

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.
Answer: Agreed, the last sentence of section 3.1 is now an updated version of this one.

TEST MODIFIED
Last sentence of section 3.1:
“Lastly, all datasets suffer from spatial discrepancies, which are detrimental to small-scale comparisons, especially near mountains, but justify our choice to use a larger study **area”**

Subsection 3.2

Note that this corresponds to subsection 3.3 in the new update manuscript

Page 20, Line 2: Please add “precipitation” between “daily” and “variability”.
Answer: added

Same line: it is not clear to me what the concept of “dependency between each dataset” means
Answer: Our sentence is not related to the lag analysis that follows, it is moved to the new section 3.3.2 “Cross-validation in the upper Indus”. We replaced the word “dependency” with “co-variability”, which is investigated in that section using correlations.

TEXT MODIFIED

- **The section 3.3.1 starts by:**

“Investigating the daily precipitation variability helps to better quantify the quality of each dataset.”

- **The section 3.3.2 starts by:**

We now start the comparison of the daily variability **between each dataset. Particularly, we aim to understand whether the co-variability exhibited between datasets is** coming from the use of common methods or **data source**, or from a good representation of the **precipitation variability.**

Page 20, Line 4: Please replace “most of the reanalyses” with “of most of the reanalyses”
Answer: Does not apply in the improved version of that section

Page 20, Line 4-5: the sentence in parentheses is unclear.
Answer: TMPA has a sub-daily resolution and we can compute a 24-hour accumulation ending at different times of the day (e.g. 0h, 3h, 6h ... UTC). APHRODITE has daily-accumulated precipitation values, and we would expect the accumulation period to end at 00h UTC. To check this, we test whether APHRODITE values correlate better with TMPA values when these ones are accumulated up to 21h, 0h, or 3h, and so

on. The problematic sentence is removed from the paragraph, and we have revised the paragraph discussing the sub-daily lag.

TEXT MODIFIED
In the paragraph discussing the sub-daily lag:
"Possible differences in the End of Day times of the observational datasets are investigated using the sub-daily resolution of TMPA. We compute TMPA daily accumulation with different End of Day time and determine which one maximises the correlation with the other observational datasets. APHRODITE and CPC (after 1998) maximise the correlation with TMPA when for the latter an End of Day at 03h UTC is used."

Page 20, Lines 9-10: Please replace "from APHRODITE" to "APHRODITE-2" with "in both APHRODITE products"
Answer: There is a misunderstanding here, the two APHRODITE datasets have actually an opposite characteristic (on the consideration of different End of Day time). We have rephrased the sentence.

TEXT MODIFIED
"Neither GPCC-daily nor APHRODITE documentation mention this issue, while a specific effort has been made to homogenise all observations in APHRODITE-2."

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?
Answer: The correlation between the two types of datasets is without doubt related to the fact that they represent the precipitation variability at this scale in the same way, or rather, the correlation is a measure of that similarity. The question here is about the cause of that similarity: it is either because they are based on a similar method or input data, or solely because they both try to estimate precipitation. This is further explained at the start of section 3.3.2

TEXT MODIFIED

- **1st paragraph of section 3.3.2:**

"We now start the comparison of the daily variability **between each dataset. Particularly, we aim to understand whether the co-variability exhibited between datasets is** coming from the use of a common method or **data source**, or from a good representation of the **precipitation variability. All datasets are estimates of precipitation, but they use different methods and input data to achieve this (cf. section 2.2). If two datasets share a similar method or data source, this could at least partly explain the co-variability between the datasets. If, on the contrary, the two datasets are independent, then the co-variability**

they share is most likely due to the precipitation signal they estimate. Therefore, the higher the correlation between two independent datasets, the better the estimate of precipitation in both datasets."

- The problematic sentence is modified:

"Therefore, the correlations between the two types of datasets **is not affected by common data or method, and is rather a measure of their quality,** which helps identifying the best datasets in each group**."**

Page 23, Line9: "analysis of the correlation" –> "correlation analysis".
Answer: This sentence is removed

Same line:I would say "ERA-Interim ranks second and is the best performing reanalysis among those which do not assimilate precipitation observations"
Answer: we further modified this sentence to avoid the use of the word "rank"

TEXT MODIFIED
"ERA-Interim has the **second highest correlations**, and is the best performing reanalysis among those that do not assimilate precipitation observations"

Page 23, Line 11: Please add "version" between "first" and "outperforms".
Answer: added

Same line:"century reanalysis" –> "20th century reanalysis" or the correct term for this product.
Answer: we added the word "twentieth"

Subsection 3.3
Note that this corresponds to section 3.4 in the new version

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

Page 32, Line 8: "good", should be justified.
Answer: The good quality is demonstrated in that section

TEXT MODIFIED
"Those two datasets present a more stable quality and good **correlations as we demonstrate below.**

Page 32, Line 12: Please add "the correlation" before "continues" (subject missing here). Same at Line 14 ("it rises" or "the correlation rises")
Answer: added

Page 32, Line 19: Remove “feedback”. This sentence should be rephrased since it is not easily readable.
Answer: The sentence has been removed

Conclusions

Page 39, Line 3: “six” –> “six datasets are”; “four” –> “four are”
Answer: changed

Page 39, Line 4: “of datasets” –> “of the datasets”; “each”–> “each of them”
Answer: changed

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.
Answer: We replaced “true value” with “uncertainty”

Page 39, Line 14: is there any reference to be cited in support to the statement about teleconnections?
Answer: Removed from the updated version of the conclusion. We want to highlight the fact that ERA5 does represent decadal variability.

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”
Answer: we rephrased this sentence.

TEXT MODIFIED
“We also found that the quality of the datasets depends on the season.”

Page 39, Line 32: “CPC is also a dry dataset”, I would rather say the “CPC exhibits a dry bias compared to ....”
Answer: The new sentence is further improved, but we have considered the comments.

TEXT MODIFIED
“CPC [...] with a large dry bias compared to GPCC-monthly”

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”.
Answer: Does not apply in the improved version of that section

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.*

Answer: We changed the last paragraph to emphasise several points: large uncertainty remains, but some datasets perform better, cross-validation between reanalysis and observational datasets are possible, and we suggest future possibilities such as quality monitoring.

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 km²**) includes the maxima of precipitation along the Himalayas and most of the winter precipitation, while the southern part (hereafter the lower Indus, **785000 km²**) 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 setDahri2018, 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 1st 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.

Cross-validating precipitation datasets in the Indus River basin

Jean-Philippe Baudouin[1], Michael Herzog[1], and Cameron A. Petrie[2]

[1]Department of Geography, University of Cambridge

[2]Department of Archaeology, University of Cambridge

Correspondence: Jean-Philippe Baudouin (jpb88@cam.ac.uk, baudouin.jeanphilippe@gmail.com)

Abstract.

Large uncertainty remains about the amount of precipitation falling in the Indus River basin, particularly in the more mountainous northern part. While rain gauge measurements are often considered as a reference~~they are only punctual and~~ , they provide information for specific, often sparse, locations (point observations) and are subject to underestimation, particularly in mountain areas. Satellite observations and reanalysis ~~output~~ data can improve our knowledge but validating their results is often difficult. In this study, we offer a cross-validation of 20 gridded datasets based on rain gauge, satellite and reanalysis data, including the most recent and ~~little~~ less studied APHRODITE-2, MERRA2, and ERA5. This original approach to cross-validation alternatively uses each dataset as a reference and interprets the result according to their dependency with the reference. Most interestingly, we found that reanalyses represent the daily variability of precipitation as well as any observational datasets, particularly in winter. Therefore, we suggest that reanalyses offer better estimates than non-corrected rain gauge-based datasets where underestimation is problematic. Specifically, ERA5 ~~has proven to be the most able reanalysis for representing the amounts of precipitation~~ is the reanalysis that offers estimates of precipitation closest to observations, in terms of amounts, seasonality as well as ~~its variability~~variability, from daily to multi-annual scale. By contrast, satellite observations bring limited improvement at the basin scale. For the rain gauge-based datasets, APHRODITE has the finest temporal representation of the precipitation variability, yet ~~importantly it~~ it importantly underestimates the actual amount. GPCC products are the only datasets that include a correction ~~of the measurements but remain~~ factor of the rain gauge measurements but this factor remains likely too small. These findings highlight the need for a systematic characterisation of the underestimation of rain gauge measurements.

1 Introduction

Throughout the Holocene, the Indus River and its tributaries have provided much of the water needed by the people living in its basin for various purposes (e.g. food, energy, industry). The diversity of use and the risks associated with scarcity or excess of water under variable and changing climatic and socio-economic conditions highlight the importance of water management in both Pakistan and north-west India (Archer et al., 2010; Laghari et al., 2012). Moreover, the Indus headwaters are an important locus of water storage with numerous glaciers whose current and future change remains uncertain (Hewitt, 2005; Gardelle et al., 2012). Therefore, a comprehensive evaluation of the basin wide water cycle is needed. Studies that have addressed this issue have stressed the uncertainties inherent in the observed precipitation (Singh et al., 2011; Gardelle et al., 2012; Immerzeel et al., 2015; Wang et al., 2017; Dahri et al., 2018).

Gridded products allow ~~a homogeneous spatialisation~~ for a homogeneous spatial representation of precipitation at a river basin-scale for statistical purposes (Palazzi et al., 2013). They can be derived from rain gauges, satellite imagery or atmospheric models (e.g. reanalysis), but need validation to assess their quality. Most studies that validate precipitation products in Pakistan, India, or in the adjacent mountainous areas (Hindu-Kush / Karakoram / Himalayas) make use of rain gauge data as a reference, either directly from ~~the~~ weather stations (Ali et al., 2012; Khan et al., 2014; Ghulami et al., 2017; Hussain et al., 2017; Iqbal and Athar, 2018), or after gridding (Palazzi et al., 2013; Rajbhandari et al., 2015; Rana et al., 2015, 2017). However, some authors have pointed out that these reference datasets also suffer from limitations that could dramatically reduce correlation and increase biases, incorrectly lowering the confidence in the dataset validated (Tozer et al., 2012; Ménégoz et al., 2013; Rana et al., 2015, 2017).

The first issue of validating gridded precipitation products with rain gauge measurements is simply the uncertainty of the measurements. Beside the risk of corruption or missing values in the reporting process, it has been demonstrated that rain gauges can underestimate precipitation (Sevruk, 1984; Goodison et al., 1989). The main source of underestimation is wind-driven under-catchment that can reach up to 50% in case of snowfall (Goodison et al., 1989; Adam and Lettenmaier, 2003; Wolff et al., 2015; Dahri et al., 2018), but also includes wetting of the instrument, evaporation before measuring, and splashing out ~~(WMO 2008)~~(WMO, 2008). Dahri et al. (2018) used the guidelines from the World Meteorological Organization (WMO) to re-evaluate the precipitation measured from hundreds of rain gauges in the upper Indus and found the underestimation to be between 1 and 65% for each station, and 21% basin wide. The second issue is the one of spatial representativeness. A rain gauge records a measurement at a specific location whereas in a gridded dataset, each value represents the mean over all the grid box. ~~The~~ Thus, the two types of data ~~thus~~ have a different spatial representativeness. This discrepancy in representativeness increases when considering shorter timesteps and areas with strong heterogeneity such as mountainous terrains, which is especially impactful when studying extreme events. Some methods exist to quantify and tackle this issue (e.g. Tustison et al., 2001; Habib et al., 2004; Wang and Wolff, 2010).

~~The gridding method is~~ Gridding methods are used to spatially homogenise point measurements and ~~also has~~ they also have limitations. Firstly, the specificity of the interpolation method can impact the result (Ensor and Robeson, 2008; Newlands et al., 2011). Secondly, the sparsity of the weather stations increases the uncertainties, which can range from 15 to 100% in areas with a low number of rain gauges (Rudolf and Rubel, 2005). This last point is especially problematic in the Indus River basin. For climatological purposes, the WMO has published guidelines for the density of rain gauges: from one station per ~~900km2~~ 900 km^2 in flat coastal areas, to one every 250 ~~km2~~ km^2 in mountains (WMO, 2008). However, the Meteorological Department of Pakistan have recently published a 50-year climatology of precipitation for the country based on 56 stations, ~~which uses~~ that is around one station per 15,000 ~~km2~~ km^2 (Faisal and Gaffar, 2012). Gridded rain gauge-based datasets rely on a similar density of observations in the Indus River basin (cf. Figure ~~1, Table 4~~2, Table 2). The situation in India is better as the Indian Meteorological Department produces a country-wide dataset of precipitation that is used for monsoon monitoring and includes up to 6300 stations. This distribution makes around one station per ~~500km2~~500 km^2, which is well within the WMO guideline. However, areas of lower density remain, especially in the western Himalayas and the Thar Desert, which are both in the Indus River basin (Kishore et al., 2016). Rain gauges are not only scarce in mountainous areas, but their location is also biased. In order to be accessible all year long, they are generally situated at the bottom of valleys, and these locations appear to be significantly drier than locations at altitude (Archer and Fowler, 2004; Ménégoz et al., 2013; Immerzeel et al., 2015; Dahri et al., 2018), which means that the interpolation method underestimates precipitation in the surrounding mountains.

There are a number of ways of overcoming the limitations of gridded rain gauge data, including the use of data derived from satellites and ~~reanalysis~~reanalyses. Satellite imagery can help to reduce both the lack and the heterogeneity of surface measurements. Satellite-based products generally make use of global infrared observations of cloud cover and microwave measurements along a swath (the narrow band where the observations are made as the satellite passes). However, their abilities over a heterogeneous terrain are more limited than over a flat and homogeneous ~~areas~~ one (Khan et al., 2014; Hussain et al., 2017; Iqbal and Athar, 2018). Moreover, these products still need rain gauges for calibration and are therefore dependent on the quality of ~~those~~station data.

Reanalyses of the atmosphere offer another way to estimate precipitation. ~~There are many~~ Many valuable variables in a reanalysis ~~, which~~ are the result of the assimilation of observations with model outputs, but ~~their~~ estimates of precipitation are, in most ~~case~~cases, a pure model product. That is, the precipitation is a forecast generated by the model used for the reanalysis, and is not constrained by direct observations in the way that other assimilated quantities are. Models are known to predict precipitation with difficulty and most studies ~~consider~~ highlight that precipitation from ~~reanalysis~~ reanalyses is less reliable than that based on ~~observation~~ observations (Rana et al., 2015; Kishore et al., 2016). The reasons often invoked include discrepancies in spatial patterns and important model biases. However, recent progress in assimilation techniques has ~~it made~~ made it possible to integrate precipitation observations in the most recent reanalysis product (ERA5, Hersbach et al., 2018), and significant improvements are possible (e.g. Beck et al., 2019).

This study aims to better understand the quality and limitations of 20 precipitation datasets that are available for a ~~domain of study~~ study area encompassing the Indus River basin. ~~Specifically, it focuses on the underestimation of rain gauge-based products, the possible improvement that can be obtained from remote sensing, and the capability of reanalysis products to represent the basin wide precipitation variability.~~ Previous studies have investigated the ~~qualities~~ strengths and limitations of precipitation datasets in this area (e.g. Ali et al., 2012; Palazzi et al., 2013; Khan et al., 2014; Hussain et al., 2017), but none ~~have~~ has looked at such a large number of datasets nor at the most recent ones. Moreover, our method slightly differs, as we offer a cross-validation, thereby avoiding the problems that come from the selection of a unique reference. We cross-compare each of the datasets, identify their similarities and discrepancies, and using the diversity of data source and ~~method~~methods, assess their strengths and weaknesses. After presenting the datasets selected for the study, we give a general description of the methods. The subsequent result section is split into ~~three~~ four parts, which review, for the precipitation: i) the ~~seasonal cycle and annual means~~annual mean, ii) the seasonality, iii) the daily variability, and ~~iii)~~ iv) the monthly and longer term variability. The final section concludes with the main results, the potential of the method, and future research priorities.

2 Data and Methods

2.1 ~~Domain of study~~Study areas

The Indus River basin extends across the north-westernmost part of the South Asian sub-continent, and is an area of various topographic features, as indicated in figure 1. It is bounded from the north-east to the west by high mountain ranges, including the Himalayas, Karakoram, Hindu Kush and ~~Suleiman~~ Sulaiman Ranges. To the south, the Indus ~~river~~ River flows into the Arabian Sea. The eastern border is the most ambiguous as it extends into the flat dune-fields of the Thar desert. Much of the precipitation that falls in this extensive area evaporates before reaching the Indus River or the sea. ~~Once on the ground, it~~ It may also forms seasonal rivers, such as the Luni River, which has been included in the ~~domain of study (outer dark blue contour in figure 2-A)~~study area. This particular river flows into the Rann of Kutch, which is a flat salt marsh with complex connections with the Arabian Sea and the mouth of the Indus River (Syvitski et al., 2013), and is ~~bound on~~ bounded to the west by the Aravalli Range. Although not strictly a part of the Indus watershed, it provides a clear and steady boundary for the ~~domain of study .~~ study area. The total watershed considered for the study is represented by the outer black line shown in figure 1.

~~Differences in relief and precipitation seasonality and pattern suggest that the basin can be separated into two distinct domains~~Precipitation amount varies across the basin as shown in Figure 2-A, as well as its seasonality. In the flat southern part, most of the precipitation occurs in July and August, under the influence of the South-Asian summer monsoon propagating from the Indian Ocean and India, while during the rest of the year ~~it~~ the basin remains dry (e.g. Ali et al., 2012; Khan et al., 2014; Rana et al., 2015). By contrast, the northern ~~domain~~ region is much more mountainous and encompasses a steep maximum of precipitation along the Himalayan front~~(Figure 2-A)~~. This precipitation falls throughout the year, ~~but exhibits~~ exhibiting a seasonal bi-modality explained by ~~a differences in process (e.g. Archer and Fowler, 2004; Singh et al., 2011; Palazzi et al., 2013). As it does~~ differences in circulation patterns (e.g. Archer and Fowler, 2004; Singh et al., 2011; Palazzi et al., 2013; Filippi et al., 2014). As in the southern part of the basin, a sharp peak in precipitation occurs in July-August related to the summer monsoon, but a second, broader peak also occurs in winter, from January to April, triggered by mid-latitude, extra-tropical western disturbances (Cannon et al., 2015; Dimri and Chevuturi, 2016; Hunt et al., 2018). ~~We have divided the basin~~

Those differences in relief and precipitation seasonality and pattern suggest that the basin can be split into two distinct areas, along a line between 68.75°E-33.5°N and 77.5°E-30°N (inner dark blue contour in figure ~~2-A~~1), which broadly corresponds to the 100mm isohyet of winter precipitation (defined from December to March). ~~The~~ Thus, the northern part of the basin (hereafter the upper Indus~~) thus~~ , 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~~) focuses only on the summerprecipitation (~~, 785000 km^2) is characterised by a single wet season during summer, as wintertime precipitation is negligible (cf. Figure 3).

2.2 ~~Data~~

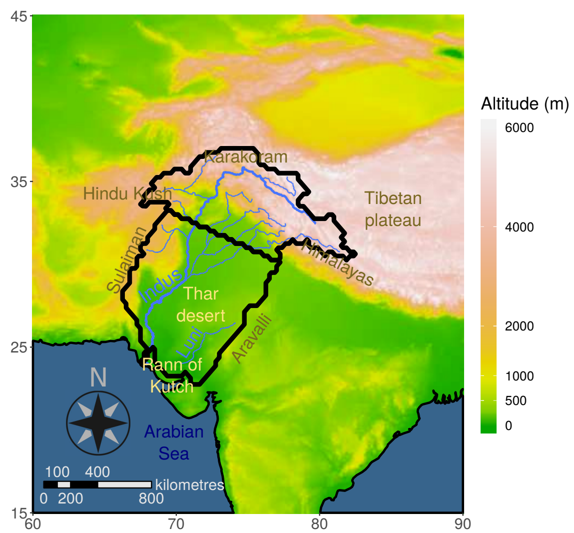


Figure 1. Relief and topographical features in and around the area investigated. The thick outer black contour represent the watershed on the Indus and Luni rivers. This area is split to form the two study areas: the upper Indus to the north, and the lower Indus to the south.

2.2 Datasets

2.2.1 Rain gauge data

We have selected five commonly used and one newly available gridded dataset based only on rain gauge data~~(~~. These are the first six datasets presented in Table 1~~)~~. The mean number of stations used in the two study areas are available for five of the datasets and presented in Table 2. The Asian Precipitation Highly Resolved Observed Data Integration Towards Evaluation of water resources (APHRODITE; Yatagai et al., 2012) ~~)~~ was developed by the Research Institute for Humanity and Nature (RIHN) and the Meteorological Research Institute of Japan Meteorological Agency (MRI/JMA). Specific to Asia, it is one of the best datasets available for the area (Rana et al., 2015), both in term of resolution (0.25° and daily, it includes a large number of rain gauges; Table 2) and ~~the fact that~~ because it covers an extended period (over 50 years). However, it does not provide data after 2007. A new dataset has been issued in 2019 from the same institute ~~with a period covered extending~~ extending the period covered up to 2015 and using a new algorithm (APHRODITE-2), though its quality has not yet been investigated. Covering the whole twentieth century at a monthly resolution, the Global Precipitation Climatology Center monthly dataset (GPCC-monthly; Schneider et al., 2018) is widely used in climatology and for calibration purposes (e.g. satellite-based datasets, Table 1). GPCC-daily (Ziese et al., 2018) offers a better temporal resolution (daily), but at a lower spatial resolution and has a much-reduced time coverage compared to GPCC-monthly. It uses a smaller number of rain gauges (Table 2), but is constrained by GPCC-monthly. The precipitation dataset from the Climate Research Unit (CRU; Harris and Jones, 2017) ~~,~~ has a similar resolution and time coverage ~~to~~ as GPCC-monthly~~, which is useful for comparison~~. We also selected another daily dataset from NOAA's Climate Prediction Center (CPC; Xie et al., 2010). Although CPC uses a lower number of rain gauges compared to APHRODITE (Table 2), its availability extends to the present with near real time updates, which means that it can be used for calibrating other near real time products (e.g. CMAP in Table 1 and MERRA2 in Table 3).

2.2.2 Satellite data

~~We selected four of the available~~ 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 10 years due to APHRODITE ending in 2007). Four were eventually selected (last four datasets in Table 1). The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007) is the most widely used satellite-based datasets. It has the highest temporal and spatial resolution of the ~~available satellite datasets~~ selection (sub-daily, and 0.25° like APHRODITE and GPCC-monthly) and ~~the largest variety of input. The~~ includes a large diversity of satellite observations. We also selected the daily product from the Global Precipitation Climatology Project (GPCP-1DD; Huffman and Bolvin, 2013) ~~is useful for comparison. The same project also issued a monthly dataset~~ as well as the monthly product issued by the same group (GPGP-SG Adler et al., 2016). All three of these datasets (TMPA, GPCP-1DD, and GPGP-SG) use GPCC for calibration~~. We have also included~~, which could introduce some similarities. By contrast, the last dataset included, CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997), ~~which~~ uses CPC for calibration~~, and~~

Table 1. Observational datasets of precipitation selected for this study, derived from rain gauges or satellites

Name	Version	Time coverage	Time resolution	Spatial resolution	Based on	Reference
APHRODITE	V1101	1951-2007	Daily	0.25°	Rain gauge only	Yatagai et al. (2012)
APHRODITE-2	V1901	1998-2015	Daily	0.25°	Rain gauge only	
CPC	V1.0	1979 (monthly) / 1998 (daily) -2018	Daily	0.5°	Rain gauge only	Xie et al. (2010)
GPCC-daily	V2	1982-2016	Daily	1°	Rain gauge and GPCC-monthly	Ziese et al. (2018)
GPCC-monthly	V8	1891-2016	Monthly	0.25°	Rain gauge only	Schneider et al. (2018)
CRU	TS4.02	1901-2017	Monthly	0.5°	Rain gauge only	Harris and Jones (2017)
TMPA	3B42 V7	1998-2016	3-hourly	0.25°	GPCC, satellites	Huffman et al. (2007)
GPCP-1DD	V1.2	1996-2015	Daily	1°	GPCC, satellites	Huffman and Bolvin (2013)
GPCP-SG	V2.3	1979-2018	Monthly	2.5°	GPCC, satellites	Adler et al. (2016)
CMAP	V1810	1979-2018	Monthly	2.5°	CPC, satellites	Xie and Arkin (1997)

. It has the same time coverage and resolution as GPCP-SG. ~~We have selected the version that~~ This version does not include reanalysis data, to simplify the analysis.

Table 2. Number of stations used on average for the rain-gauge-based datasets (except CRU for which this information was not directly available), per time step, for the two study areas, and over the period 1998-2007.

Datasets	Upper Indus	Lower Indus
APHRODITE	55	48
APHRODITE-2	88	65
CPC	15	21
GPCC-daily	11	16
GPCC-monthly	35	33

2.2.3 Reanalysis data

Unlike the observation datasets, reanalysis ~~datasets~~ data can be quite different from one another. They generally use their own atmospheric model and assimilation scheme, and the type and number of observations assimilated ~~varies. Therefore, we selected a large set of~~ can vary. Table 3 shows the ensemble of the ten reanalysis datasets ~~(Table 3)~~that have been used in this study. The four reanalyses of the latest generation are, from most recent to oldest: ERA5 (Hersbach et al., 2018) from the European Centre for Medium-Range Weather Forecasts (ECMWF), the Modern Era Retrospective-analysis for Research and Applications version 2 (MERRA2; Gelaro et al., 2017) from the NASA, the Japanese 55-year Reanalysis (JRA; Kobayashi et al., 2015) from the JMA, and the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010, 2014) from the National Center for Environmental Prediction (NCEP). These are still regularly updated, ~~and~~ they all include the latest observations from satellites and cover the full satellite era from at least 1980. JRA goes back to 1958, when the global radiosonde observing system was established~~, while~~ . ERA5 ~~will eventually cover the whole second half of the twentieth century.~~ currently starts in 1979 but future releases are expected to extend this back to 1950.

In terms of technical differences, ERA5 uses a more complex assimilation scheme than the others reanalysis (4DVAR), which allows for better integration of the observations. It is also the only one that assimilates precipitation measurements. MERRA2 also uses observations, but takes them from a gridded dataset (CPC) and only uses them to correct the precipitation field before analysing the atmospheric impact on the land surface; this changes land surface feedbacks on the atmosphere. CFSR is an Ocean-Atmosphere coupled reanalysis, that is, the sea surface is modelled and provides feedback to the atmospheric model, instead of being prescribed by an analysis from observations. ERA5 and MERRA2 are the most recent of the reanalysis datasets to be published, and not many studies have looked at the improvement from their predecessor, respectively ERA-Interim (Dee et al., 2011) and MERRA1 (Rienecker et al., 2011). Both have stopped being updated or will be very shortly, but they are included in this study for comparison purposes.

Reanalyses for the whole twentieth century have also been produced, but to retain the homogeneity of the type of observations assimilated they only include surface observations. The twentieth century reanalysis from NCEP (20CR; Compo et al., 2011), only assimilates surface pressure, but more recently, the ECMWF produced ERA-20C (Poli et al., 2016), which has surface wind assimilated along with surface pressure.

We have also made use of older generation reanalysis datasets that are still being updated, including: the NCEP/NCAR reanalysis (NCEP1; Kalnay et al., 1996) and the NCEP/NDOE reanalysis (NCEP2; Kanamitsu et al., 2002). Both are useful to quantify the progress in reanalysis systems as well as to compare them with more observation-limited century long reanalyses.

Table 3. Datasets of precipitation selected for this study, derived from reanalysis

Name	Time coverage	Spatial resolution	Remarks	Reference
ERA5	1979-2018	0.25°	4DVAR, precipitation assimilated	Hersbach et al. (2018)
ERA-Interim	1979-2018	0.75°	4DVAR assimilation scheme	Dee et al. (2011)
JRA	1958-2018	0.5°		Kobayashi et al. (2015)
MERRA2	1980-2018	0.5° / 0.625°	Correction of the precipitation with CPC for land interaction. Assimilate aerosol observations	Gelaro et al. (2017)
MERRA1	1979-2010	0.5° / 0.66°		Rienecker et al. (2011)
CFSR	1979-2018	0.5°	Coupled reanalysis (atmosphere, ocean, land, cryosphere). Same analyses as MERRA1. Version 2 starting in 04/2011	Saha et al. (2010, 2014)
NCEP2	1979-2018	1.875°	Fixed errors and updated model since NCEP1 No satellite radiance assimilated	Kanamitsu et al. (2002)
NCEP1	1948-2018	1.875°	No satellite radiance assimilated	Kalnay et al. (1996)
20CR	1871-2012	1.875°	Assimilate surface pressure only	Compo et al. (2011)
ERA-20C	1900-2010	1°	Assimilate surface pressure and marine wind only	Poli et al. (2016)

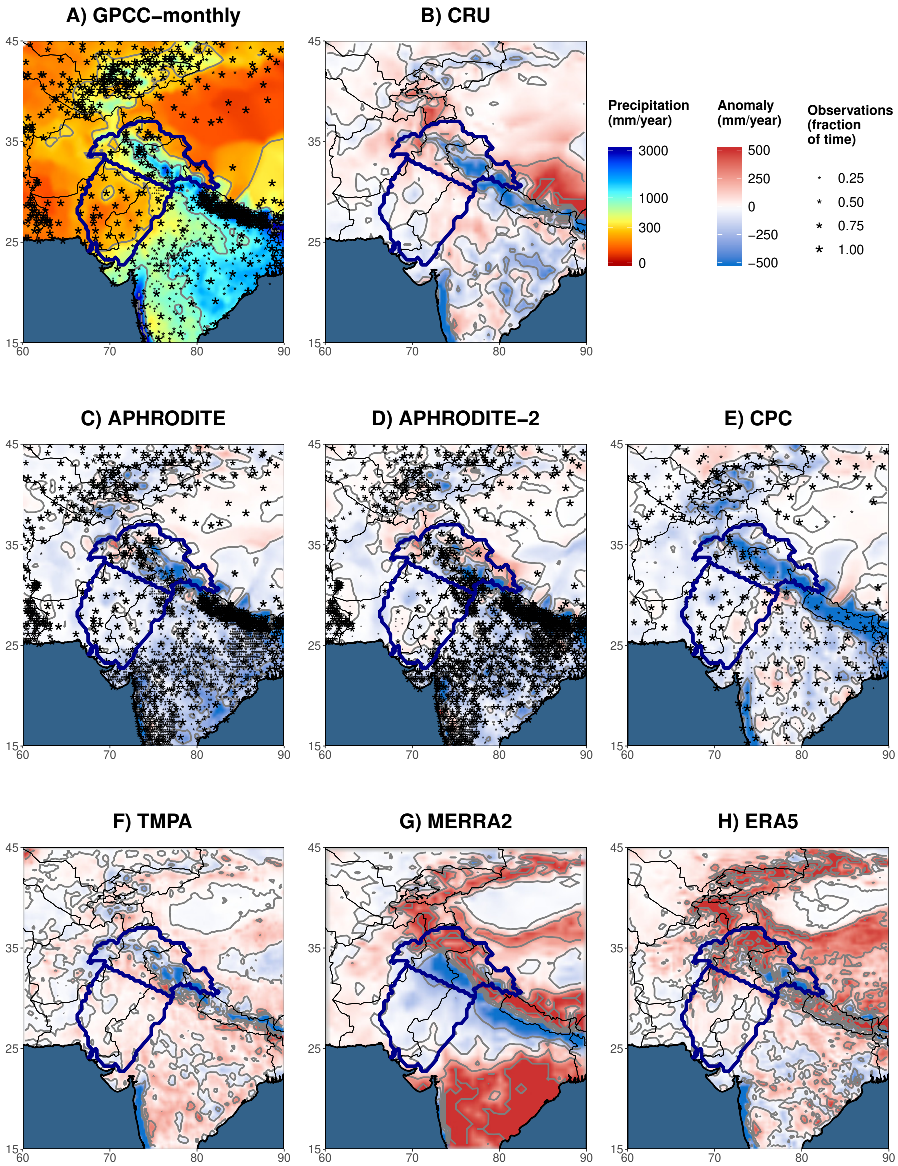


Figure 2. Map of annual mean precipitation for different datasets. The annual mean is computed over the period 1998-2007. GPCC monthly (A) is used as a reference to compute the anomaly for the other datasets (B to H). The grey lines are the isohyets whose level corresponds to the labels in the legend. The boundaries of the two study areas are displayed in dark blue on each map. The stars mark the grid cells that include at least one gauge observation. The size of the stars represents the number of time steps with at least one observation over that cell, relative to the total number of time steps needed to compute the annual mean (120 for A, 3652 for C,D and E). This information was not available for CRU (B) nor ERA5 (H), and does not apply to the satellite-based TMPA (F), and ~~the two reanalyses~~ MERRA2 (G~~and H~~).

2.3 Methods

For each dataset, the time series of precipitation are averaged over the two ~~domains of study~~ study areas (upper and lower Indus) ~~were~~ and calculated at a monthly resolution, and daily if possible. ~~In order to keep the boundaries of the domains of study fixed, all data were~~ The datasets have different spatial resolution which causes a problem when calculating the precipitation averages over the study areas. Simply selecting the cells whose centre is within these areas leads to small biases in the extent of the region considered. These biases are reduced by bi-linearly ~~interpolated~~ interpolating all data to a 0.25° grid, common to APHRODITE, APHRODITE-2, and GPCC-monthly. This choice is further discussed in section 3.1.1.

~~Most of the~~ The analysis is performed over the 10-year period from 1998-2007, which is common to all ~~of the datasets. Different timescales are investigated: daily, monthly, inter-annual and decadal~~datasets, ~~as well as the seasonality~~except when analysing the trends and inter-annual to decadal variability, for which we use all data available. We focus on the two wet seasons of the upper Indus. Summer is defined from June to September~~for both domains~~, which matches the monsoon precipitation peak. Winter is defined from December to March~~for the upper Indus, which~~. This fits the snowfall peak rather than the precipitation peak, but ~~this~~ makes it possible to focus on ~~the issues on~~ issues of snowfall estimation (Palazzi et al., 2013). ~~Winter is not defined in~~ In the lower Indus, we use the same definition for summer, but winter is not analysed, as it is a dry season.

~~The study mainly focuses on comparing~~ We first compare the mean and ~~variability of the time series. For the variability, the Pearson coefficient is mainly used, along with the Spearman coefficient in case of issues with extreme values. The differences between the time series are explained by the differences in the way the datasets are produced: either from the method or from the raw data used in input. This helps with identifying and quantifying the key sources of uncertainty. Conversely, if two datasets share similarities, it is often due to similarities in the methods or in the input data. Taking those datasets into account to evaluate the uncertainty will lead to its underestimation. However, similarities are also shared by unrelated datasets. When two time series with no common input data and different methods exhibit a common behaviour, it is likely due to the fact that behaviour is present in the real signal. Therefore, we assume that the two time series showing common behaviour better represent the reality than a more dissimilar third one, which eventually helps reducing the uncertainty.~~

~~The use of observational datasets and reanalyses in combination is key in this study. Reanalyses use very different methods in their production than the gridded observational datasets, and most do not include precipitation observations in their input. Reanalyses and observational datasets are thereby independent and can be used to validate one another. This is made possible by the increase in quality of the most recent reanalyses, as we will now discuss in the result section.~~ seasonal cycle of each dataset in sections 3.1 and 3.2. For quantitative statements, we use GPCC-monthly as a reference. However, in section 3.1.3, we use the precipitation dataset from Dahri et al. (2018) as reference instead. This dataset cannot be used in other parts of the study, as it is limited to one part of the upper Indus, and only provides annual means.

Then, in section 3.3 we compare the daily variability of the precipitation using the Pearson correlation. The correlation significance is discussed at the 95% probability level. To reduce the impact of abnormally large rainfall events when investigating trends in daily variability (cf. section 3.3.4), we use the Spearman correlation. Lastly, in section 3.4, other time scales of variability of the precipitation are investigated: monthly, seasonal, inter-annual, and decadal, still using the Pearson correlation at the 95% confidence interval.

3 Results

3.1 ~~Seasonal cycles and annual means~~Annual mean

~~The study of the seasonal cycle has been performed through the analysis of the monthly (Figure 3), seasonal, and annual mean (Table 4) of precipitation over the two domains of study. The results are split between observations (based on both rain gauges and satellites) and reanalyses. To simplify the analysis, we compare the datasets to GPCC-monthly data, as it likely provides good estimates (see below).~~

3.1.1 Differences between rain gauges-based datasets

~~All the datasets record the wet and dry seasons, although with different magnitudes. For the upper Indus, the three~~ Annual mean precipitation in both study areas and for each dataset are given in Table 4 (last two columns). We first focus on the rain gauge-based datasets ~~: CPC, APHRODITE and GPCC-monthly, are ranked in the same order for each month (Figure 3~~(upper part of the table). Spatial pattern differences are shown in Figure 2-A ~~). CPC is the driest (-51% in annual mean compared to GPCC-monthly), followed by APHRODITE (-23%). Similarly, CPC and APHRODITE are drier than GPCC-monthly in the lower Indus, although in a different order (-15% and -22% respectively, Figure 3-C). By contrast, GPCC-daily and CRU are much closer to GPCC-monthly (difference inferior to 5 and 8% respectively).~~ to E.

~~Several reasons explain the differences of mean precipitation between these datasets. First, GPCC products include a correction of the rain gauge measurements of about +5 to +10% (cf. Figure 4 in Schneider et al., 2014). Second, the datasets have different resolutions that affect the spatial representativeness, despite having carefully interpolated all datasets to the same grid. For instance,~~ First, we should mention that the bi-linear method we use to interpolate each dataset to the same grid (cf. subsection 2.3) leads to some differences between datasets. The two GPCC products can be used to evaluate the impact of our interpolation method, as they have a different spatial resolution but uses the same monthly climatology. Hence, the small underestimation of GPCC-daily ~~is based on~~ compared to GPCC-monthly ~~, therefore their differences in climatology should mainly be due to the difference in spatial resolution. Third,~~ (about 1% in the upper Indus and 5% in the lower Indus) is related to the interpolation method~~affects the spatial patterns and can lead to basin-scale differences. APHRODITEincludes a methodthat~~ . However, these differences are small enough to justify the use of our method.

More generally, annual mean differences can be explained by methods and data that each dataset uses. Particularly, the interpolation of station measurements to a grid differs from one dataset to the other. APHRODITE's interpolation method, for instance, considers the orientation of the slope to quantify the influence of nearby stations. This greatly reduces the amount of precipitation falling in the inner mountains compared to GPCC-monthly. An example of this pattern is evident ~~for the northern side~~ to the north of the Himalayas where only very few observations exist (Figure 2-D; Yatagai et al., 2012). In CRU, the interpolation method (triangulated linear interpolation of anomalies; Harris et al., 2014) seems to smooth areas of strong gradients such as near the foothills of the Himalayas (Figure 2-B). This smoothing ~~would~~ might explain a slightly drier upper Indus, and slightly wetter lower Indus, compared to GPCC-monthly (Table 4). ~~Fourth, the~~

Differences can also be explained by the dramatic change in location and number of stations used to compute the statistics ~~varies dramatically from one dataset to another, particularly in the upper Indus where the precipitation patterns are the most heterogeneous~~ (Figure 2-A, C, D, and E, Table 2). ~~CPC is likely drier because of a lower~~ For example, CPC is by far the driest dataset in the upper Indus and the second driest in the lower Indus. This is likely related to the low number of observations it includes, leaving vast areas with no or very few observations, including the wettest regions (Figure ~~3~~2-E). However, there is no linear ~~relation between precipitation quantity~~ relationship between precipitation amount and number of observations. GPCC-daily includes the lowest number of observations~~and still has an annual mean precipitation over the domain of study that is similar to GPCC-monthly, which is due to its being constrained by~~, but this does not impact its climatology, because the climatology is derived from GPCC-monthly. On the contrary, APHRODITE ~~has~~ comprises a much higher number of observations than others datasets, but remains much drier than GPCC-monthly ~~. APHRODITE's creator~~ (about 20% drier in both study areas).

Yatagai et al. (2012) pointed out that ~~the~~ differences in quality checks ~~could explain this behaviour (Yatagai et al., 2012). This possibility is especially so as the product partially~~ compared to the other datasets might explain APHRODITE's dry bias . They noted that APHRODITE partly relies on GTS data that are sent in near real time to the global network, ~~which means that there is a significant risk that missing valuesare reported as no precipitation, thus leading to underestimation of the total precipitation. In that case, the issue should also affect CPC, which is~~ with the risk of misreporting. This risk particularly concerns misreported zero values, which are hard to detect and lead to a dry bias. The large dry bias seen in CPC data might be associated to the same issue, since CPC is entirely based on GTS data~~only~~. In GPCC-monthly (and daily), only stations with at least 70% of data per month are retained ~~(Schneider et al 2014)~~(Schneider et al., 2014), while in CRU this number is ~~increased to~~ 75% (Harris et al., 2014). Thus, limiting the analysis to the most reliable weather stations ~~could~~ can help build confidence in ~~the~~ recorded total precipitation amount. ~~However, we were not able to prove the existence of spurious null or close to zero values at grid points with observations that are compared to TMPA, due to too low correlation. Rather, we found strong differences on the monthly value of GPCC, APHRODITE and CPC, at grid points where the three have integrated observations. This could suggest that the three datasets use differently consolidated precipitation measurements from the same weather station, or handle the interpolation at a grid point with observations differently.~~

Interestingly, APHRODITE-2 ~~has a higher mean than APHRODITE (+15% and +13% in the upper and lower Indus respectively), which is closer to GPCC-monthly. Different~~ is more than 10% wetter than APHRODITE in both study areas. Several changes have been performed ~~on~~ in the methodology: quality control of extreme high values, station-value conservation after interpolation, merging daily observation with different definitions ~~for the start and end of a day, and~~ of End of Day time (cf. section 3.3.1), and an updated climatology. However, the difference in mean precipitation is ~~mainly~~ most likely related to the change in observations from rain gauges~~: they are basin-wide more numerous, but~~. Although APHRODITE-2 comprises more observations basin-wide, this increase mainly ~~happens over Indian territory. Pakistan actually sees a decrease in~~ occurs over the Indian territory, whereas Pakistan is presented with fewer precipitation measurements, especially in the dry southern central part (Figure 2-D). This decrease in observations in ~~the drier area explains~~ a drier area can reasonably explain the increase in mean precipitation in the lower Indus. In the upper Indus, the increase is mainly due to the inclusion of measurements from one isolated weather station in the northernmost part of ~~India~~the area.

~~The mean summer precipitation of the~~

3.1.2 Considering satellite and reanalysis datasets

We now consider satellite-based datasets ~~overall differs little from GPCC-monthly estimate for the upper Indus (less than 3%, Table~~ (middle part of the table 4). In ~~winter, TMPA, GPCP-1DD, and GPCP-SG have slightly lower estimates~~ the upper Indus, CMAP stands out as being the wettest observational datasets, 13% wetter than GPCC-monthly~~, and are more comparable to APHRODITE. CMAP stands out, reporting the largest winter precipitation amount (+36% compared to~~ . By contrast, the other three (TMPA, GPCC-1DD, GPCP-SG), are drier than GPCC-monthly (between 10 and 5%), despite being calibrated by this GPCC-monthly~~)~~. In the lower Indus, all ~~satellite products~~ satellite-derived datasets are wetter than the rain ~~gauge-based products (from~~ gauge products (between 10 ~~to~~ and 30% more than ~~in~~ GPCC-monthly~~in the annual mean), especially for both GPCP products~~). The complexity of the algorithm used to produce the satellite-based datasets makes determining the reasons for their differences with each other or with rain gauge products ~~unclear~~difficult. According to previous studies, their ability to represent precipitation over rough terrain is limited (e.g. Hussain et al., 2017). In fact, figure 2-F shows that the strongest differences between TMPA and GPCC-monthly occurs near mountain ranges, such as the upper Indus~~is limited (e.g. Hussain et al., 2017)~~. In contrast, precipitation estimates over flat terrain with sparse observations and mostly convective precipitation ~~like the lower Indus~~ benefit from satellite observations (Ebert et al., 2007).

~~The reanalyses represent the dry and wet seasons, but with a somewhat larger spread than the observations (Figure 3-B and D). This is especially true~~ This suggests that the higher precipitation mean of the satellite-derived datasets for the lower Indus ~~, with two outliers: JRA, which is wetter by a factor of two; and 20CR, which has almost no wet season. The monthly variations are, on the other hand, generally well captured, with a maximum in July. For the upper Indus, some discrepancies are evident in the seasonality. The most striking one is a wetter winter. On average winter precipitation~~ is ~~30% higher than in GPCC-monthly, with the notable exception of ERA-20C (Table 4). Those wetter conditions also extend to the surrounding drier months: April-May and October-November. Interestingly,~~ possibly due to better consideration of locally higher precipitation rates during convective events.

The annual mean precipitation in reanalysis datasets is listed in the lower part of table 4. In the lower Indus, the ~~mean summer precipitation for the reanalyses is not significantly wetter than GPCC-monthly (Table 4). Only Era-Interim stands out with a wet bias, mainly in the north-west corner of the upper Indus domain, a bias partly corrected in ERA5 (Figure 2-H)~~range of values is very high: the wettest dataset, JRA, is five times wetter than the driest dataset, 20CR. This range shows the significant difficulties for reanalyses to represent precipitation in an area were convection dominates. Among the most recent reanalyses, ERA5 has the closest estimates of precipitation to the observational datasets, yet above the estimates from rain gauges. Figure 2-H suggests that these wetter conditions mainly comes from the north-western edge of the Suleiman range, an area with sparse precipitation observation (cf. Figure 2-A), therefore increasing confidence in ERA5 estimation. The two twentieth century reanalysis (20CR and ERA-20C) are amongst the driest reanalysis datasets, suggesting that their models have difficulties to propagate the monsoon precipitation into the lower Indus region, when only surface observations are assimilated. Lastly, MERRA2 exhibits a severe drop of precipitation compared to the previous version, MERRA1. Summer monsoon precipitation is known to be strongly affected by surface moisture content, especially in flat areas like the lower Indus (Douville et al., 2001). MERRA2 uses CPC data to constrain the precipitation flux at the surface. Due to the dry bias of CPC, soil moisture is reduced for most of India (Figure 3 in Reichle et al., 2017), explaining the drop in precipitation. ~~Overall, the annual precipitation for the upper Indus is approximately~~

For the upper Indus, the most striking features is that all reanalysis datasets except MERRA1 and ERA-20C predict higher precipitation amounts than GPCC-monthly, about 20~~% higher in the reanalysis products than in GPCC-monthly.~~ % higher on average. In the following we investigate whether that this difference can be explained by an underestimation of rain gauge measurements.

~~A second discrepancy is visible between a majority of the reanalyses and the observations for the upper Indus : a delay in the seasonality starting during the pre-monsoon season. The observations show that May is the driest month of that season and is followed by a sharp increase in precipitationin June, but only ERA5, ERA-Interim, and MERRA1 represent this behaviour. In contrast, NCEP2 and CFSR are much drier in June than in May. For the other reanalyses, the precipitation for May and June are comparable. This delay continues into the summer monsoon period: while the observations clearly show a wetter July than August, this is only the case for ERA5, ERA-Interim, and both MERRA reanalyses. A similar delay can be found over the Ganges plain and all along the Himalayas, which suggests wider uncertainties on the monsoon propagation in the reanalysis.~~

3.1.3 Impact of rain gauge biases in mountainous terrains

~~Difference in winter precipitation between reanalysis and observational datasets can at least in part be explained by a dry bias in the observations~~Rain gauge measurements are known to potentially underestimate precipitation and particularly snowfall (Sevruk, 1984; Goodison et al., 1989). This is an important issue for mountainous region such as the upper Indus. However, among the six rain gauge-based datasets, only GPCC's products consider a correction of the data. Based on a study by Legates and Willmott (1990), a correction factor, which depends on the month, is applied at each grid cell. Most of these factors vary between 5 and 10% (Figure 4 in Schneider et al., 2014), and explain why GPCC's products are wetter than most of the other rain gauge-based datasets. Recently, Dahri et al. (2018, hereafter Dahri2018) compiled the measurements from over 270 rain gauges in the upper Indus and adjusted their ~~results~~ values to undercatchment, following WMO guidelines. They found a basin-wide adjustment of 21%, but this varies from 65% for high altitude stations, to around 1% for the ~~ones in the plain. GPCC-monthly, the only observational dataset to take undercatchment into consideration, adjusts the values by around 5 to 10% (Schneider et al., 2014). Therefore, the Dahri2018 study indicates that GPCC correction factors are largely underestimated in the mountains. Dahri2018 selected a domain of study~~ stations in the plains.

The Dahri2018 dataset has both the advantage of considering a very large number of observations and correcting rain gauge measurements. However, its result is based on a study area somewhat smaller than the upper Indus ~~domain~~ region presented here, and only covers the period ~~1999-2011. To make comparison with its results~~ from 1999 to 2011. For comparison purpose, we recomputed the annual mean of several of the most recent and highest resolution datasets to fit these ~~requirements~~ definitions (Table 5). ~~We found that the unadjusted precipitation mean in~~

Table 5 shows that none of the observational datasets is able to reproduce the Dahri2018 precipitation estimates. They all have a dry bias, from 30% for TMPA, to 10% for GPCC-monthly. Furthermore, APHRODITE-2 and TMPA even underestimate the unadjusted value of Dahri2018, which suggests that the underestimation is not only related to rain gauge measurements, but also to the representation of the spatial pattern. By contrast, GPCC-monthly is 7% ~~lower than the GPCC value, a percentage around~~ higher than the Dahri2018 unadjusted value, which corresponds to the correction factor used in GPCC. This suggests that the unadjusted values in ~~the two~~ both datasets are very close, ~~while GPCC-monthly uses much fewer stations measurements. While this could highlight GPCC quality~~ and highlights the quality of GPCC. Nevertheless, we also found ~~some~~ discrepancies in the spatial patterns ~~that offset each other at the basin scale. In the Karakoram Range first, at~~ between GPCC-monthly and Dahri2018. Particularly, in the northernmost part of the upper Indus ~~domain~~region, in the Karakoram range, GPCC-monthly exhibits lower precipitation means than ~~in~~ Dahri2018, which cannot be explained by the difference ~~of correction factor~~ in correction factors between the two datasets alone. The nearest stations used in GPCC-monthly are all located in the dry and more accessible Indus River valley to the south of the mountain range (Figure 2-A). Those drier conditions ~~are extended~~ extend to the north ~~by~~ due to the interpolation method used by GPCC, while Dahri2018 ~~integrate station measurements suggesting~~ includes station measurements with wetter conditions than in the valley. This ~~approach~~ difference illustrates the impact of biased weather station locations mentioned in the introduction and in several ~~previous~~ other studies (e.g. Archer and Fowler, 2004; Ménégoz et al., 2013; Immerzeel et al., 2015). ~~In contrast, in the western part of the domain,~~

Still in the Karakoram range, figures 2-G and H show that MERRA2 and ERA5 are wetter than GPCC-monthly~~extends the precipitation maximum to the north~~, and therefore closer to Dahri2018. However, spatial discrepancies remain. Particularly, the maximum of precipitation in MERRA2 is shifted to the North, which leads to ~~an overestimation compared to the precipitation measurements used in Dahri2018. Eventually, those two biases seem to offset each other, making the area averaged estimate of GPCC-monthly close to Dahri2018. APHRODITE-2 and TMPA similarly fail at representing the differences between valleys and mountains and underestimate significantly the adjusted value from~~ important biases when averaging on the Dahri2018~~(-21% and -31% respectively).~~

~~By contrast~~'s study area. Our study area, which does not overlap with the highest precipitation rates, is less affected by shifts and is better fitted to compare the large scale precipitation patterns. Nevertheless, the four selected reanalysis datasets in Table 5 overestimate the Dahri2018 adjusted ~~value. We noted that the south border of the domain of study in Dahri2018 overlaps with the highest precipitation rates, which occur in summer. Depending on the different representation of the relief and its modelled impact on weather in the reanalyses , a small shift to the north or the south can occur, with a potentially important impact on the mean over Dahri2018's domain. In contrast, the upper Indus domain we have delineated includes all the highest rates of precipitation falling in the Indus River basin, which reduces the impact of a possible small shift to the north or the south of the maxima. A shift to the north is especially evident in MERRA2 (Figure 3-G) and explains the stronger overestimate found in Table 5 than in Table 4, compared to GPCC-monthly~~values, by 20% on average. This suggests that part but not all of the differences between reanalyses and observational data can be explained by biases from the latter. This overestimation of modelled precipitation in reanalyses for the upper Indus is corroborated by previous studies (e.g. Palazzi et al., 2015).

~~All~~ To conclude, all rain gauge-based datasets ~~thus suffer from underestimating~~ suffer from an underestimation of annual mean precipitation for the upper Indus when compared to Darhi2018. ~~The main reason is linked to the adjustment of undercatchment, which is either underestimated or is not sufficiently considered. This is especially critical in winter. Differences in quality~~ This results from biases in rain gauge locations and measurements. Quality control and interpolation ~~method also likely explain differences between the rain gauge-based datasets and impact~~ methods also impact precipitation amount in both parts of the basin. ~~Those errors are more consistent yearlong. Interestingly, interpolation uncertainty seems not limited to areas with low density observations. Further, satellite observations might bring improvement in~~ Satellite observations probably improve precipitation estimates in flat areas with sparse observations. ~~For example, a higher precipitation mean for the lower Indus is possibly due to better consideration of local higher rates during convective events. However, satellite products cannot correct observation~~ However, they cannot correct observational biases since they use them for calibration, and biases remain unchanged or even amplify for the upper Indus~~domain. Overall, GPCC-monthly has the closest estimate to Dahri2018, thanks to a large ensemble of good quality rain gauge measurements. A study with a high density of bias-corrected rain gauge measurements would be needed in the lower Indus to identify which dataset better represents the precipitation mean. The IMD (Indian Meteorological Department) dataset can provide such reference for the Indian part of the basin. Kishore et al. (2016) showed that the IMD product is closer to GPCC-monthly than to APHRODITE in the north-west part of India, while in Prakash et al. (2015) TMPA biases are limited between -10 and 10% in that domain.~~

~~Except~~. Reanalyses do not include rain gauge measurement, except for ERA5 and MERRA2, ~~reanalysis data are independent from rain gauge measurements~~ and are therefore not affected by ~~these underestimates. They suggest a much wetter winter season for the upper Indus, while the latest reanalyses (ERA5, JRA, MERRA2, and CFSR) converge towards a similar amount, which is possibly closer to the reality than observation-based estimates, although some over estimations cannot be overruled. Summer is a more complicated season,~~ observational biases. However, model biases can also be significant as suggested by the spread of ~~precipitation intensity in reanalysis products is high to very high, varying between -29% and +48% of GPCC-monthly mean for~~ the annual mean precipitation values. Reanalyses tend to be wetter than observational datasets in the upper Indus, ~~and between -72% and +164% for the lower Indus. Even when only considering the latest reanalyses, the summer mean does not converge. Moreover, their seasonality does not always fit the observations, nor their representation of spatial patterns. These latest difficulties~~ which is partly explained by the underestimation of the observations. Lastly, all datasets suffer from spatial discrepancies, which are detrimental to small-scale comparisons, especially near mountains, but justify our choice to use a larger study ~~domain. MERRA2 and ERA5/ERA-Interim have estimates of the pre-monsoon and monsoon seasonality closer to observations in the upper and the lower Indus, respectively~~area.

Table 4. Mean annual and seasonal precipitation (in mm) falling over the two study areas, for the period 1998-2007. Winter is defined from December to March and summer from June to September. The first ten datasets are observations, the second ten are reanalyses.

Datasets	Upper Indus			Lower Indus
	Winter	Summer	Annual	Annual
APHRODITE	154	237	484	198
APHRODITE-2	179	272	555	223
CPC	98	200	355	216
GPCC-daily	201	297	607	243
GPCC-monthly	201	301	613	255
CRU	166	281	565	267
TMPA	156	298	555	286
GPCP-1DD	161	305	569	317
GPCP-SG	167	309	583	325
CMAP	273	307	696	279
ERA5	280	380	828	300
ERA-Interim	289	445	931	305
JRA	299	325	810	586
MERRA2	265	310	724	177
MERRA1	205	267	598	355
CFSR	282	214	656	162
NCEP2	274	259	703	276
NCEP1	372	343	915	239
20CR	244	319	746	116
ERA-20C	175	276	551	175

Table 5. Mean annual precipitation (in mm) for various datasets over the study area defined in Dahri et al. (2018) for the period 1999-2011. Both adjusted and unadjusted values (the latter in parenthesis) from Dahri et al. (2018) are reported in the second line

Datasets	Revised Upper Indus
Dahri2018	697 (574)
APHRODITE-2	548
GPCC-monthly	612
~~TRMM~~ TMPA	480
ERA5	835
JRA	827
MERRA2	929
CFSR	783

3.2 Seasonal cycle

The seasonal cycle of precipitation for each dataset is presented in Figure 3. Analysing the seasonality is particularly interesting in the upper Indus, as it is characterised by two wet seasons. The mean precipitation of each season is presented in table 4 (second and third column). The rain gauge-based datasets exhibit a very similar seasonality for both study areas. In the upper Indus, the maxima of precipitation occur in February and July, the minima in May and November. The differences between the datasets vary little from one month to another, which suggests that the causes of the differences identified in the previous section (e.g. misreporting, station location and number, interpolation method) are independent of the seasonality. The satellite-based datasets represent the summer precipitation almost exactly as GPCC-monthly. The annual mean differences are explained by biases during the winter season, which suggests that winter precipitation is more difficult to estimate for those datasets.

The reanalyses represent the dry and wet seasons of the upper Indus, but with a larger spread than in the observations and some differences in seasonal cycle (Figure 3-B). On average, winter precipitation is 30% higher than in GPCC-monthly, with the notable exception of ERA-20C (Table 4). Those wetter conditions also extend to the surrounding drier months: April/May and October/November. However, the mean summer precipitation in reanalyses is not significantly different from GPCC-monthly (Table 4). Only ERA-Interim stands out with a wet summer precipitation bias, mainly in the north-west corner of the upper Indus, a bias partly corrected in ERA5 (Figure 2-H). The winter wet bias is not surprising after the comparison with the Dahri2018 dataset in the section 3.1.3. Indeed, Dahri2018 found that the most important rain gauge underestimations happen in winter when precipitation mostly falls as snow. More interestingly, we found that the latest reanalyses (ERA5, JRA, MERRA2, and CFSR) represent winter precipitation in similar ways. We haven't been able to investigate the seasonality of the Dahri2018 dataset, but we suggest that the latest reanalyses better represent winter precipitation than the observational datasets.

We noted another discrepancy in seasonality between a majority of the reanalyses and the observations for the upper Indus: a delay of the summer precipitation starting from the pre-monsoon season (Figure 3-B). The observations show that May is the driest month of that season, followed by a sharp increase in precipitation in June. Only ERA5, ERA-Interim, and MERRA1 reproduce this behaviour. In contrast, NCEP2 and CFSR are much drier in June than in May. For other reanalyses, precipitation during May and June are comparable. This delay continues into the summer monsoon period: while the observations clearly show a wetter July than August, this is only the case for ERA5, ERA-Interim, and both MERRA reanalyses. A similar delay can be found over the Ganges plain and along the Himalayas, which suggests wider uncertainties on the monsoon propagation in the reanalyses. By contrast, no such delay is found in the lower Indus, despite the large uncertainty on the amount of precipitation (Figure 3-D).

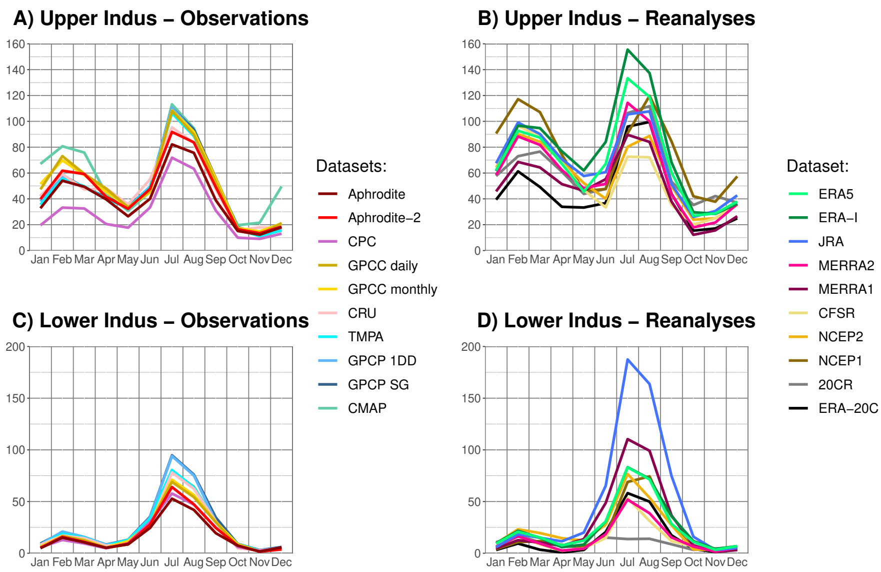


Figure 3. Monthly mean of precipitation, over the period 1998-2007, representing the seasonal cycle. Results are split between upper Indus (A and B) and lower Indus (C and D) as well as observation datasets (A and C) and reanalysis (B and D).

3.3 Daily variability

~~Comparing the daily variability helps to quantify the dependency between each dataset, either coming from the use of common methods or input data, or from the representation of the true variability. Before computing the daily correlation, we checked for possible lags between the datasets, using especially the sub-daily resolution of TMPA and most of the reanalyses (e. g. is the daily value of APHRODITE more correlated with the daily value of TMPA accumulated from 21h, 00h, 03h, etc.).~~

3.3.1 Lag analysis

Investigating the daily precipitation variability helps to better quantify the quality of each dataset. Before computing the daily correlation, we checked for possible lags between the datasets. Lags can have different origins. The first is the accumulation period considered for the rain gauge measurements. CPC documentation (Xie et al. 2010) points out that the official period is different from one country to another (in our case, Afghanistan, Pakistan, and India all ~~have~~ use different periods, ~~starting at~~ or "End of Day time": 00hUTC, 06hUTC, 03hUTC, respectively~~, cf. End of Day time for CPC~~), which could impact precipitation estimates. Neither GPCC-daily ~~documentation does not~~ nor APHRODITE documentation mention this issue, while ~~from APHRODITE to APHRODITE-2,~~ a specific effort has been made to homogenise all observations in APHRODITE-2. Secondly, the TMPA algorithm uses the 00h imagery for the following day accumulation, and therefore, ~~is~~ could be more representative of an accumulation starting at 22:30h UTC (Huffman et al., 2007). Thirdly, biases in the daily cycle are possible in the reanalyses.

Our main finding relates to CPC. Figure 4 shows the daily correlation year per year of CPC against APHRODITE and MERRA2, for two lags: 0h and -24h (previous day for CPC). We found that the two lags switch their behaviour somewhere around 1997/1998, which we interpret as an error in the data processing for CPC. That is, in CPC before 1998, precipitation values correspond to those for the following day. This should not have an important impact on monthly and longer accumulations, but we limited the daily analysis of CPC to the period from 1998 to 2018. Moreover, similar errors might have happened earlier during the 1980s as the curves in Figure 4 come closer or invert again. This error also propagates to the corrected precipitation of MERRA2. That is, before 1998, the land surface in the model receives the precipitation of the following day. Theoretically, this could enhance precipitation by increasing surface moisture supply before the precipitation actually falls. However, we have not been able to find a significant change before and after 1998. The error has been reported to NOAA's CPC.

~~Daily correlation, per year, between CPC and Aphrodite (A and C), and MERRA2 (B and D) for both upper Indus (A and B) and lower Indus (C and D). The green line is the correlation between the same days in each datasets. For the red line, the previous day of CPC is used instead. The black vertical line is the start of the year 1998, around where the main error should be.~~

~~The comparison between TMPAand the rain gauge-based datasets~~Possible differences in the End of Day times of the observational datasets are investigated using the sub-daily resolution of TMPA. We compute TMPA daily accumulation with different End of Day time and determine which one maximises the correlation with the other observational datasets.

APHRODITE and CPC (after 1998) ~~shows a maximum of correlation delayed by around +3h, while no delays are evident for GPCC-daily and APHRODITE-2~~maximise the correlation with TMPA when for the latter an End of Day at 03h UTC is used. This behaviour suggests that both CPC and APHRODITE are more representative of an accumulation period ~~starting~~ ending at 03hUTC, influenced by the Indian rain gauge network. APHRODITE-2 successfully corrected this delay, ~~for a start~~ maximising correlation with TMPA for a End of Day at 00hUTC, like GPCC-daily. ~~We also~~

A similar analysis can be performed for the reanalyses, to investigate the possibility of a delay in the daily cycle of the precipitation. We found that most reanalyses have a negligible (~~<~~$\lesssim$3h) delay with ~~GPCC-daily~~TMPA. However, the reanalyses of the twentieth century have a different behaviour: both have a +12h delay~~with TMPA~~. For those two, only surface observations are assimilated. It is possible that 12h is the time needed by the troposphere to adjust to those surface constraints.

Finally, we decided to take the accumulation period starting at 00h for all sub-daily datasets. Indeed, it is not straightforward to correct the delay in APHRODITE or CPC for instance, since only a daily resolution is available. Moreover, the correlation coefficients are not ~~significantly~~ too importantly affected by those sub-daily lags.

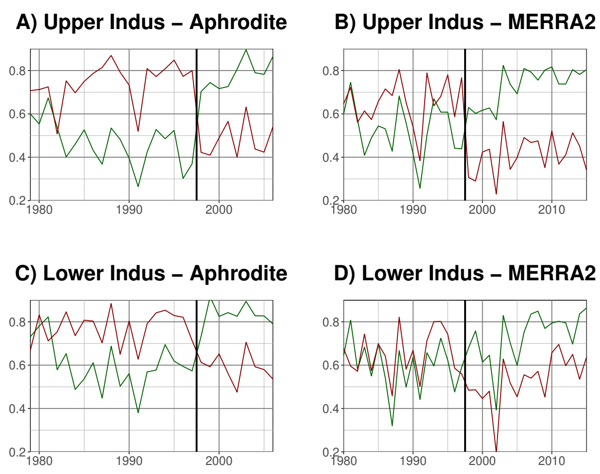


Figure 4. Daily correlation, per year, between CPC and Aphrodite (A and C), and MERRA2 (B and D) for both upper Indus (A and B) and lower Indus (C and D). The green line is the correlation between the same days in each dataset. For the red line, the previous day of CPC is used instead. The black vertical line is the start of the year 1998, around where the main error should be.

3.3.2 Cross-validation

We now start the comparison of the daily variability between each ~~observational datasets for the upper Indus, using the correlation coefficients (upper part of Table 6). Note that the significance of the correlation (or of their difference) mentioned hereafter is defined for a 95% level. Two groups can be identified with higher correlation between the members of each. In the first group, TMPA and GPCP-1DD have a correlation of~~ dataset. Particularly, we aim to understand whether the co-variability exhibited between datasets is coming from the use of a common method or data source, or from a good representation of the precipitation variability. All datasets are estimates of precipitation, but they use different methods and input data to achieve this (cf. section 2.2). If two datasets share a similar method or data source, this can at least partly explain the co-variability between the datasets. If, on the contrary, the two datasets are independent, then the co-variability they share is most likely due to the precipitation signal they estimate. As a consequence, the higher the correlation between two independent datasets, the better is the estimate of precipitation of both datasets.

Table 6 presents the daily correlation of precipitation between the different datasets, for the upper Indus. The upper part of the table focuses on the cross-correlation between the observational datasets. The highest correlation coefficient, almost 0.9, is between TMPA and GPCP-1DD, showing how close those two datasets are, likely due to the satellite observations they have in common and the similarity of retrieval procedures (Rahman et al., 2009; Palazzi et al., 2013; Rana et al., 2017). ~~In the second group, the~~ The rain gauge-based datasets APHRODITE, CPC, and GPCC-daily have ~~a correlation of around 0.8~~ also a high correlation between one another ~~.~~ about 0,8. The two versions of APHRODITE are even closer, due to their similarities of conception. When comparing GPCP-1DD and TMPA's correlation coefficients ~~against~~ using the rain gauge-based datasets as reference, it turns out that the TMPA coefficients are systematically significantly higher than those for GPCP-1DD (at the level 95%). That is, TMPA variability is closer to the rain gauge-based datasets than GPCP-1DD is. It could be either because TMPA includes more information from the rain gauge measurements than GPCP-1DD or because it has better quality (better algorithm, better ~~raw data~~ data source). Similarly, ~~if we compare the rain gauge-based dataset coefficients against satellites,~~ we note that APHRODITE ~~has significantly higher values~~ and APHRODITE-2 have significantly higher correlation with the satellite-based datasets than CPC and GPCC-daily ~~, as does APHRODITE-2~~ do.

~~We can argue that APHRODITE and APHRODITE-2 both represent the daily variability in the upper Indus better than CPC and GPCC-daily using~~ In the lower part of table 6, the correlation between the reanalyses and the ~~observations (lower part of Table 6). Precipitation~~ observational datasets are about as high as between the observational datasets, suggesting that reanalyses are as good as observational datasets in representing the daily variability. Moreover, precipitation from reanalysis and observational data are independent from each other, in the sense that they do not share the same ~~input data~~ data source (except ERA5, which assimilates precipitation observations, and MERRA2, which integrate CPC data; the two need to be treated separately). ~~Therefore~~Hence, the correlations between the two types of datasets ~~can only be explained by a common dependency on the true variability~~is not affected by common data or method, and is rather a measure of their quality, which helps identifying the best datasets in each group. ~~Moreover, those correlations have the same order of magnitude as the correlations between each observational dataset. That is, reanalyses are as good as observational datasets in representing the daily variability.~~

~~For each reanalysis dataset, it is possible to rank the observational datasets depending on the correlation with the reanalysis (~~We continue the comparison of the observational datasets using reanalyses as a reference (comparison along the rows of ~~Table~~ table 6). ~~We found that~~ APHRODITE-2 ~~systematically ranks first, not far from APHRODITE, but with a correlation each time significantly higher~~ has systematically higher correlation with the reference, regardless of the reanalysis used, than the other observational datasets. It is followed by APHRODITE. Both have significantly higher correlation than GPCC-daily, in third position. By contrast, CPC has systematically a lower ~~rank~~ correlation than GPCC-daily. ~~We~~ Interpreting these results in terms of quality, we attribute the lower performance of CPC and GPCC-daily to the much lower number of observational inputs than in APHRODITE and APHRODITE-2 (Table 2). Despite a slightly higher number of measurements, CPC performs worse than GPCC-daily, likely due to issues on the quality of those measurements, discussed in ~~Section 3.2~~section 3.1.1. Regarding satellite-based datasets, TMPA systematically outperforms GPCP-1DD, but the two, along with CPC, ~~rank the lowest~~ have the lowest correlations with the reanalyses. That is, satellite measurements seem to ~~perform poorly over the upper Indus~~degrade the signal from rain gauge measurements.

We can also ~~rank the correlation of the reanalysis products for each observational dataset~~ compare the reanalyses quality using observational datasets as a reference (along the columns of table 6). ERA5 ~~systematically ranks first~~has systematically higher correlations with the observations. However, this reanalysis assimilates ~~observations~~rain gauge measurements, such that it is not completely independent from the observational datasets. It is certainly a sign of good quality that the reanalysis output resembles the observations, but the reanalysis data could also include some of the observation errors~~, which is an issue that cannot be quantified by this analysis of the correlation~~. ERA-Interim ~~ranks second~~ has the second highest correlations, and is the best ~~of the reanalysis not integrating~~ performing reanalysis among those that do not assimilate precipitation observations. It is closely followed by MERRA2, while CFSR has ~~the lowest correlation with the observations of~~ poorer results among the latest generation of reanalyses. Interestingly for NCEP's reanalyses, the first version outperforms the second version. The two twentieth century reanalyses also show interesting behaviour: while 20CR has the lowest correlations with the observations, ERA20C performance is between CFSR and NCEP1, despite only assimilating surface observations. This behaviour clearly shows the progress made in reanalysis processing (e.g. in atmospheric modelling and data assimilation) over the last decades.

~~For~~ The same correlation analysis is performed for the lower Indus ~~domain, the results are quite similar~~ (Table 7). The results are quite similar, but we also note some interesting differences. The correlations between the observations are all higher for this ~~domain~~study area. In this flat ~~domain~~area, precipitation is less heterogeneous, and observations are more representative of their surrounding (i.e. larger spatial representativeness). In contrast, the reanalyses have lower correlations with observations than for the upper Indus. The lower Indus only receives precipitation during the summer monsoon, which is less well represented in models than the winter precipitation in the upper Indus (see ~~below~~ following section on seasonality). More in details, APHRODITE-2 and APHRODITE still ~~rank first~~perform best among the observational datasets, but the four other datasets rank in ~~different orders~~a different order: satellite products are possibly better in that flatter ~~domain~~area. For the reanalyses, we noticed that MERRA2 does not outperform MERRA1. ~~Summer monsoon precipitation , especially in flat areas like this one, is strongly affected by~~ It echoes the large change in precipitation amount between the two discussed above (Table 4), and, similarly, could be related to the integration of CPC in MERRA2. Indeed, table 7 suggests that CPC does not perform as well as the other observational datasets in terms of variability, and, indeed, surface moisture content ~~(Douville et al. 2001). However, this parameter~~ variability was not improved from MERRA1 to MERRA2 in the area ~~as its authors would have expect by correcting the precipitation seen by the land surface model with CPC (Reichle et al., 2017, Figure 1 in). On the contrary, we have shown that CPC likely underestimates precipitation (Table 4) and is not very good at representing the daily variability either (Table 7). This use of CPC possibly explains the slightly worse capability of MERRA2 compared to MERRA1, where improvement would have been expected.~~ (Figure 1 in Reichle et al., 2017). As for ERA5 and ERA-Interim, they remains the two reanalysis datasets with the highest correlation with the observations.

Table 6. Daily correlation between different datasets, in the upper Indus for the period 1998-2007.

Datasets	APHRODITE	APHRODITE-2	CPC	GPCC-daily	TMPA	GPCP-1DD
APHRODITE-2	0.92					
CPC	0.797	0.775				
GPCC-daily	0.819	0.836	0.816			
TMPA	0.76	0.762	0.687	0.712		
GPCP-1DD	0.735	0.725	0.665	0.676	0.898	
ERA5	0.888	0.903	0.743	0.81	0.741	0.727
ERA-Interim	0.854	0.87	0.722	0.777	0.733	0.727
JRA	0.843	0.86	0.677	0.759	0.702	0.697
MERRA2	0.846	0.862	0.714	0.778	0.708	0.699
MERRA1	0.834	0.849	0.683	0.76	0.698	0.688
CFSR	0.795	0.82	0.64	0.74	0.641	0.625
NCEP2	0.706	0.731	0.552	0.661	0.577	0.545
NCEP1	0.76	0.769	0.606	0.687	0.619	0.598
20CR	0.596	0.635	0.512	0.567	0.481	0.478
ERA20C	0.754	0.746	0.646	0.691	0.644	0.643

Table 7. Same as Table 6 for the lower Indus

Datasets	APHRODITE	APHRODITE-2	CPC	GPCC-daily	TMPA	GPCP-1DD
APHRODITE-2	0.887					
CPC	0.838	0.825				
GPCC-daily	0.864	0.841	0.87			
TMPA	0.829	0.869	0.79	0.809		
GPCP-1DD	0.771	0.801	0.72	0.74	0.906	
ERA5	0.858	0.871	0.805	0.826	0.835	0.772
ERA-Interim	0.828	0.837	0.763	0.794	0.79	0.744
JRA	0.719	0.76	0.709	0.708	0.76	0.73
MERRA2	0.777	0.794	0.723	0.763	0.725	0.677
MERRA1	0.782	0.796	0.749	0.76	0.775	0.741
CFSR	0.7	0.69	0.626	0.657	0.672	0.618
NCEP2	0.601	0.632	0.572	0.618	0.576	0.523
NCEP1	0.635	0.643	0.605	0.623	0.596	0.545
20CR	0.442	0.4	0.35	0.393	0.345	0.308
ERA20C	0.655	0.712	0.643	0.663	0.678	0.673

~~We investigated the possible seasonalityin the correlation coefficients detailed so far~~

3.3.3 Influence of the seasonality

Figure 5 presents the seasonality, for the upper Indus~~. We did not find notable variation between the observational datasets. Seasonal variation does occur when comparing the reanalysesto the observations. In Figure 5, we compare the daily variability of each reanalysis with APHRODITE-2 for each month. The data~~ , of the correlations between the reanalyses and APHRODITE-2. This reference is chosen because of its higher correlation with the reanalyses, but the other rain gauge-based datasets give a similar seasonality. The figure shows that the reanalyses are altogether more similar to APHRODITE-2 during the winter season than during summer~~, while the differences between the reanalyses vary~~. From December to April, all reanalysis products have a similarly high correlation with the observational dataset (>0.9), except for the two century reanalyses, and to a lesser extent the older NCEP reanalyses. From May ~~onwards~~onward, all correlations drop to various degrees. Both NCEP reanalyses drop the most, followed by CFSR. ERA5 shows the highest correlations, just above ERA-Interim, JRA, ~~MERRA-1, and MERRA-2.~~ MERRA1, and MERRA2. For the century reanalyses, 20CR drops to very low values (<0.5 and even <0.2 in September and October), while ERA-20C remains at acceptable levels, around CFSR. ~~We therefore~~ Accordingly, we have very high confidence in the capability of most reanalyses to represent the daily variability in winter. In summer, the confidence is more dependent on the reanalysis, and overall lower than in winter. However, it is unclear if the seasonality of the correlation between ~~the reference~~ APHRODITE-2 and the best reanalyses (ERA5, ERA-Interim, JRA, MERRA1, MERRA2) is due to a changing ability of the reanalyses or of ~~the reference dataset,~~ APHRODITE-2. The seasonality for those reanalyses disappears when using TMPA as a reference, but mainly due to a drop in winter correlation, which rather suggests that satellite observations are not suited for that season (not shown). The analysis of the seasonality is less interesting in the lower Indus~~domain~~, since it is mainly dominated by the monsoon. The results resemble what was just discussed for summer in the upper Indus.

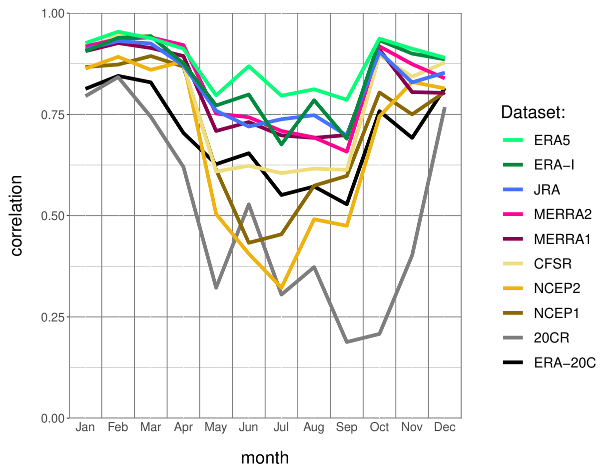


Figure 5. Daily correlation, per month, between APHRODITE-2 and each reanalysis, in the upper Indus. The period considered is 1998-2007.

3.3.4 Trends

We also looked at possible trends in the representation of the daily variability, due to a change in the type, quantity, or quality of input data in each dataset. We computed the time series of correlations between observations and reanalyses using a two year moving window~~for the upper Indus (Figure 6)~~. However, the Pearson correlation we used so far is also known to be sensitive to extreme values. This leads to jumps in the correlation when an extreme value (abnormally large precipitation event) passes in the moving window and is well ~~predicted~~represented. In order to have a clearer signal, without jumps, we used instead the Spearman correlation. This coefficient is based on the rank rather than on the absolute value of each observation and is therefore not sensitive to extreme values. We checked that most of the results presented above are valid with the Spearman correlation as well.

In Figure 6-A, we compare the observational dataset using ERA-Interim as a reference, for the upper Indus. We first notice that APHRODITE and APHRODITE-2 always have significantly higher correlation scores than the others, except around 2004-2006, and relatively stable values between 0.85 and 0.9. The quality of those two datasets found over the period 1998-2007 can therefore be extended to the whole period 1979-present. GPCC-daily exhibits stronger variability during the first 20-years, but then its score increases and stabilises around 0.85. This behaviour is likely due to an increase of the number of observations that are between 5 and 10 before 2000, but above 15 after 2005. CPC is in general very close to GPCC-daily, except around the year 2000, which explains the differences between the two datasets over the period 1998-2007 previously investigated. The two satellite products TMPA and GPCP-1DD are very similar to each other, relatively stable, but at a lower level than the rain gauge-based datasets.

We now investigate in Figure 6-B the quality of the most recent reanalysis ~~(Figure 6-B)~~ using as reference APHRODITE (plain line) and APHRODITE-2 (dotted line). ~~This reference is~~ These references are justified by the stability of ~~the good result~~ their good results discussed above. ~~The two references~~ They give similar results over their common period, which helps when analysing the whole time period. ERA5 and ERA-Interim are the two most stable reanalyses and have the highest correlations. JRA is also one of the best reanalysis datasets in the ~~first decade~~1980's, but its correlation drops by about 0.05 compared to ERA5 after 1990 and never recovers. MERRA1 and 2 exhibit similar variability to each other, but the first version often has better results than the latter. CFSR is the most problematic reanalysis with the strongest variability and much lower correlation. However, it shows much better results at the end of the time period, with the release of its second version.

~~Finally~~Lastly, over the second half of the twentieth century, the large change in number and type of observations assimilated could impact the quality of the reanalysis ~~(~~and is therefore investigated in Figure 6-C~~)~~. However, no trend can be found. Correlations between JRA and APHRODITE remains mostly between 0.8 and 0.85. ERA-20C is also ~~quite consistent~~ fairly stable over time, generally above NCEP1. 20CR, by contrast, exhibits a much higher variability with correlation dropping as low as 0.4 at times, and sometimes reaching NCEP1.

There are some differences in the results for the lower Indus ~~(Figure 7)~~as shown in Figure 7. First, for the observation, CPC and GPCC-daily reach the quality of APHRODITE-2 around 2005, despite including half the number of observations (Figure 7-A). Certainly, after 2005, the more homogeneous coverage of observations in CPC and GPCC-daily than in APHRODITE-2 counterbalances the reduction in number (Figure 2-D and E). Before 2005, the cause of the improvement of GPCC-daily can again be tracked to the increase in observations included, while the rise in quality of CPC remains of uncertain origin, since the number and location of observations are constant. TMPA shows correlation very close to CPC, with a similar unexplained rise between 2000 and 2005, almost reaching the quality of the rain gauge-based datasets. GPCP-1DD has lower scores than TMPA, but also sees a rising trend during the two decades it covers. Comparing the differences between the reanalyses (Figure 7-B), we found much smaller differences than when using the Pearson correlation (Table 7), which suggest that the difference in quality resides in the representation of the extreme events. No clear change can be observed during the period 1979-2015, however.

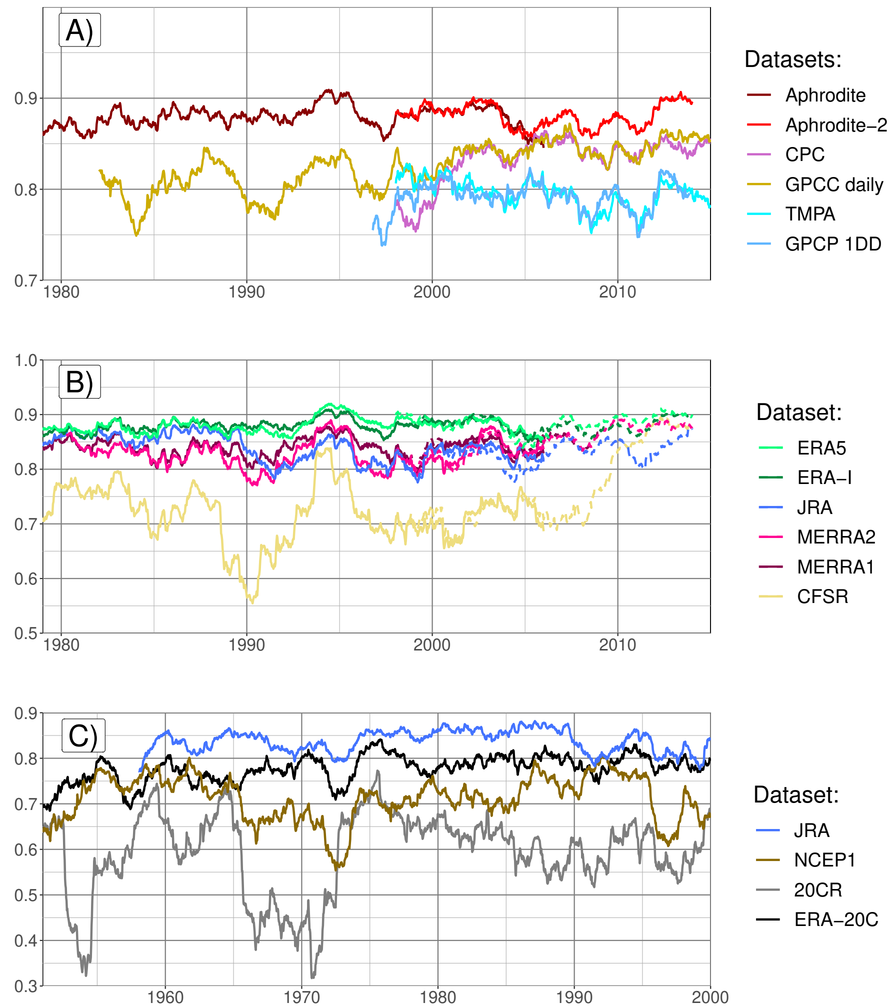


Figure 6. Daily correlation using the Spearman formula, on a running two-year window, between a reference and different datasets, for the upper Indus. The years on the x-axis is the start of the two-year window. In A) observational datasets are tested against ERA-Interim. Figure B) shows the correlation between a selection of reanalysis and APHRODITE over the period ~~1979-2007~~ 1979-2005 (plain line) and APHRODITE-2 over the period 1998-2013 (dotted line). Finally, C) presents the reanalyses covering the second half of the 20th century, with APHRODITE as reference

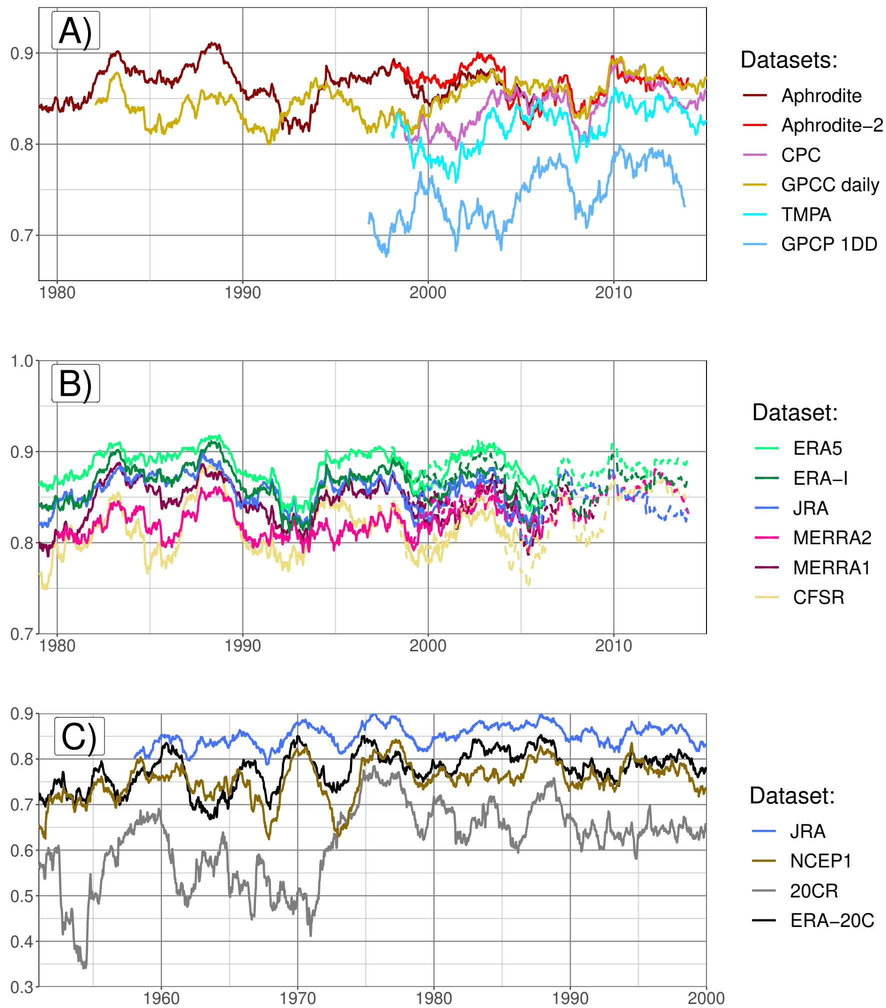


Figure 7. Same as Figure 6 but for the lower Indus

3.4 Monthly, seasonal, and inter-annual variability

A good representation of daily precipitation variability does not ensure a good representation of monthly or longer period variability. Moreover, all the observational datasets selected for this study can be analysed at a monthly time scale~~, such as GPCC-monthly or CRU, which only have a monthly resolution~~. In Figure 8, we present the trend in monthly correlation between a reference and each type of dataset for the upper Indus. The correlation is calculated with the Pearson formula and over a ten-year moving window. It uses the monthly anomaly of precipitation, relative to a monthly mean computed over the same ten-year moving window. The reference to validate the observational datasets is ERA-Interim (A), and to validate the reanalyses GPCC-monthly (B). Those two datasets present ~~lower variability in their~~ a more stable quality and good ~~results~~correlations as we demonstrate below. They also cover the whole period 1979 to the present. However, we checked the main results with other references to validate them.

The best observational dataset for representing monthly variability for the upper Indus is APHRODITE (Figure 8-A). By contrast to the daily variability analysis, APHRODITE-2 has a ~~much~~ significantly lower correlation with ERA-Interim on the common period with APHRODITE (1998-2007) and the correlation continues to drop after it. The difference in correlation between the two datasets is quite dependent on the reference, but all show the subsequent decrease. By contrast, CPC starts with the lowest correlation, but the correlation rises in the last decade at the level of the other datasets. CMAP, based on CPC also presents lower correlation, but is more variable, and it depicts a similar rise around the year 2000. All the other datasets are very close to each other. ~~CRU is slightly below GPCC-monthly, while GPCP products and TMPA, all including GPCC data, are slightly above it.~~

Still for monthly variability, the closest reanalysis to the observations is ERA5 (Figure 8-B), except when using CPC and CMAP as reference: then, MERRA2 has higher correlation at times, likely due the use of CPC data in both CMAP and MERRA2. ~~The feedback impact of fixing the precipitation input in the surface model with CPC seems much more important at this time scale than with the daily variability.~~ Several datasets show a decrease in correlation during the 1990s: JRA, has a drop more pronounced than what is observed for the daily variability, and a drop appears for NCEP1, NCEP2 and ERA-20C. 20CR has the lowest correlation, while MERRA2, MERRA1, and ERA-Interim are quite similar, with correlation just below ERA5. CFSR also has relatively high values, but exhibits a decreasing trend, especially in the last 10 years, which is even more pronounced when testing with the other observational datasets. It is possible that version 2 of CFSR gives better results, but it has not been running long enough to evaluate the monthly variability over a 10-year period. Instead, the correlations in Figure 8-B include both versions toward the end of the time period, which could add discrepancies when computing the monthly mean anomaly.

We also tested the datasets with the longest time coverage against GPCC-monthly (Figure 8-C). We found ~~a relative consistency in the correlation~~ relatively stable correlations with APHRODITE and CRU during the twentieth century: the ~~timeseries~~ time series do not diverge, despite the lowering number of observations. However, since the datasets are not independent, we cannot say that the quality of those datasets remains constant. The reanalyses present fluctuating ~~correlation~~ correlations with the reference. ERA-20C has lower ~~correlation~~ correlations in the first half of the century, which could be due to a lowering confidence in either the reference or the reanalysis. However, ERA-20C ~~correlation gets~~ correlations get closer to 20-CR during that period, which suggests the variation in the reanalysis quality is the most important factor.

The lower Indus shows somewhat different results in terms of monthly variability (Figure 9). For the observations, APHRODITE does not have the highest correlations, as it is bypassed by GPCP-SG during the 1980s. After 2000, all datasets perform very similarly with two exceptions: CRU, which always has lower correlations, and APHRODITE-2 whose correlations drop during the last two years. ~~CPC exhibits the same rising trend as for the upper Indus, but it is closer to the other datasets.~~ For the reanalysis, ERA5 still has the highest correlation but is joined by ERA-Interim just before the year 2000. MERRA2 ~~also exhibits a rising trend. Surprisingly, MERRA2~~ does not show specifically higher correlation with CPC, as it does for the upper Indus~~domain~~, except for the two first years, where CPC has the lowest values. It is possible that the smaller difference in quality between CPC and the other observational datasets is not important enough to influence MERRA2's quality significantly. ~~A drop in correlation is still observe for JRA, but it occurs latter, between 1995 and 2005. CFSR does not show any trend. NCEP1 has surprisingly high correlations, especially during the 1980s when it is above CFSR. Finally~~Lastly, for the century-long datasets, correlations between CRU and GPCC-monthly show a decreasing trend, that could be related to an increasing difference in the observations included in each dataset. By contrast, ERA-20C correlation ~~drops at~~ are as low as 20CR ~~level~~ before 1950.

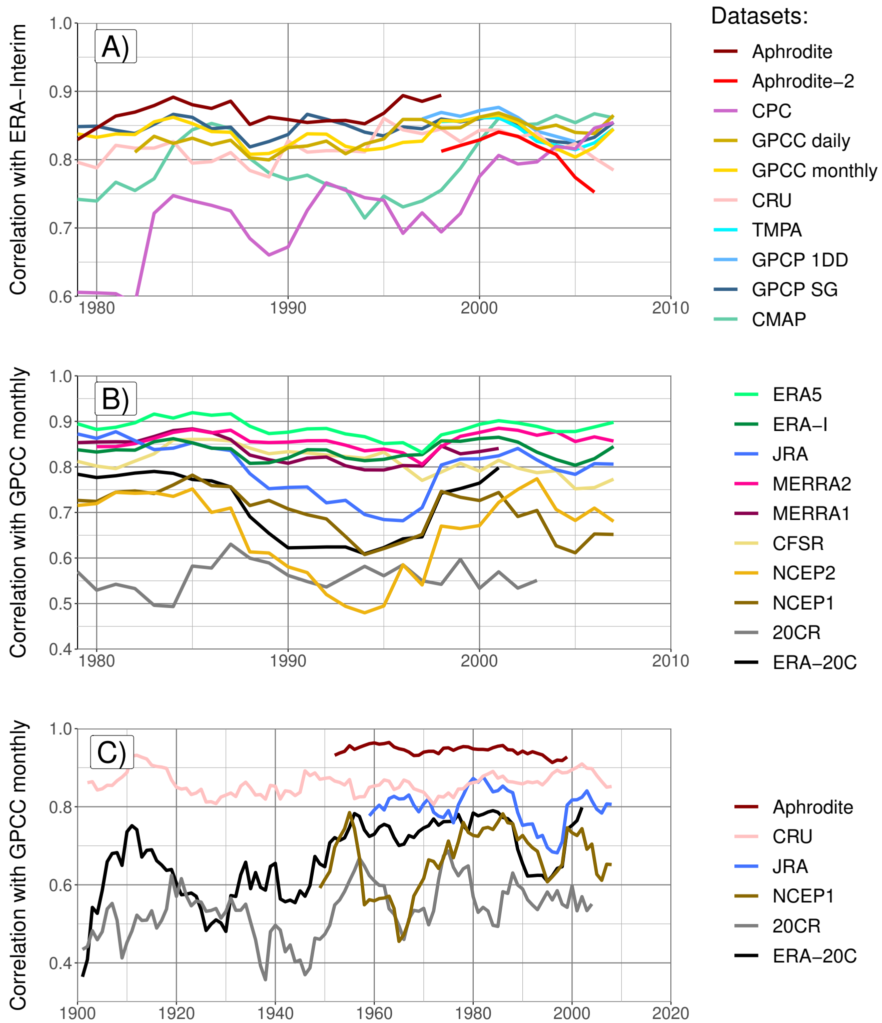


Figure 8. Correlation of monthly anomaly on a running ten-year window for the upper Indus. The monthly mean needed for the anomaly is computed relatively to the ten-year window. The years on the x-axis is the start of the ten-year window. Similarly as in Figure 5, a set of datasets is tested against a reference. In A) observational datasets are tested against ERA-Interim. B) shows the correlation between the reanalysis and GPCC-monthly. ~~Finally~~Lastly, C) presents the longest datasets, except GPCC-monthly which is used as reference.

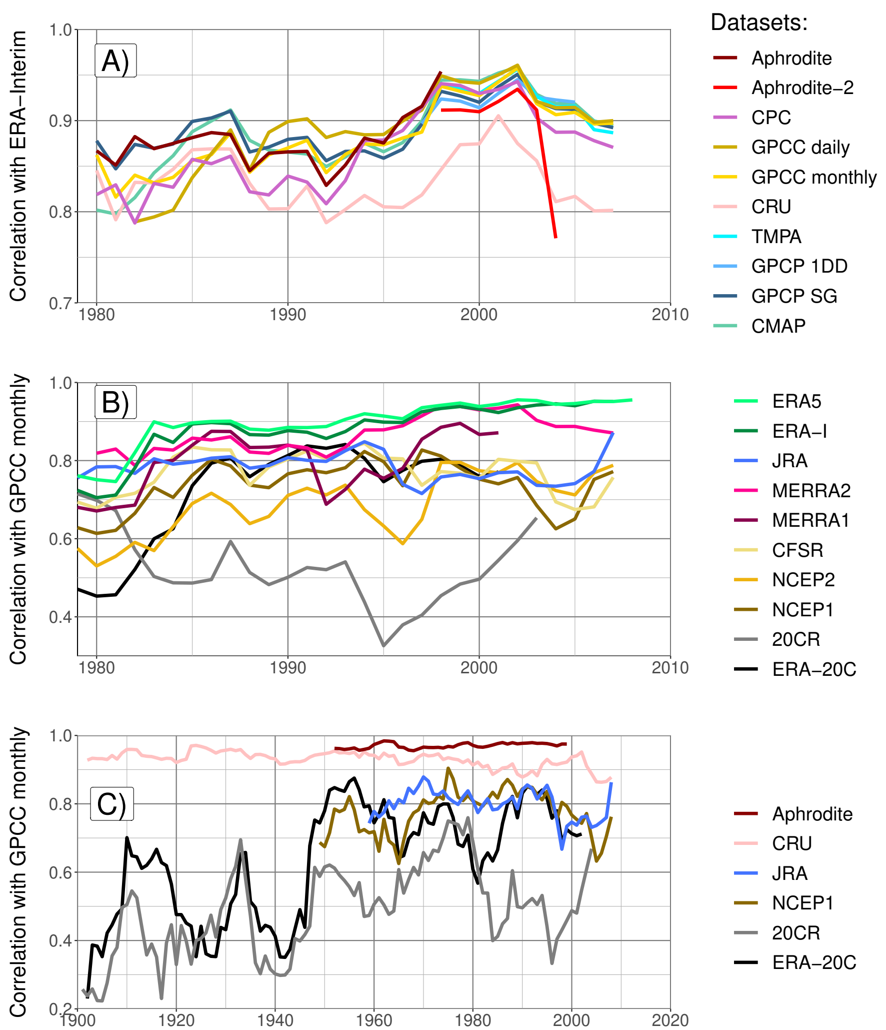


Figure 9. Same as Figure 8 but for the lower Indus

In Figure 10, we compare the inter-annual variability of GPCC-monthly to the reanalyses over the period 1981-2010 and to the other observational datasets covering that period. In Figure 11, we look at the ~~10-years~~ 10-**year** moving mean for each of these datasets. Note that the years we mention in the text correspond to the start of that 10-year window. The results are split by season and ~~domain~~study area. GPCP inter-annual variability is almost identical to that of GPCC-monthly, due to the inclusion of GPCC-monthly data (Figure 10). By contrast, CPC has a much lower correlation with GPCC-monthly, especially in the upper Indus. This agrees with the lower capabilities found for the daily and monthly variability of CPC. Moreover, CPC is the most dissimilar observational dataset for the decadal variability, ~~and is similar to~~ particularly for the upper Indus, along with CMAP and APHRODITE-2 (Figure 11). In contrast, the other datasets show a very similar behaviour.

The reanalyses in winter have a ~~behaviour~~ decadal variability similar to the observation for the period 1980-2010 (Figure 11-D). Moreover, the most recent reanalyses tend to converge towards the same amount of precipitation after 2000. By contrast, the reanalyses that run before 1980 do not represent the decadal variability depicted by the observations. For summer in both ~~domains, only~~ study areas, none but ERA-5 represents the decadal variability observed. ~~It also has an inter-annual correlation with GPCC-monthly that is higher than the correlation between GPCC and CRU. The other reanalyses have significantly lower correlation and miss all or some of the decadal variability.~~ For example, in the upper Indus during summer, the precipitation amount increases after 2000 in the observations (Figure 11-B). While MERRA2 and CFSR show an increase of precipitation 2 or 3 times more important, ERA-interim and NCEP1 and 2 show instead a decrease ~~. Moreover, only ERA5 and ERA-interim represent the peak around 1990 clearly. Similarly, in the lower Indus the 1990 peak is only reproduced by ERA5, CFSR, NCEP2, and ERA-20C, and the early twenty first century rise by ERA5 and ERA-Interim~~(Figure 11)-E. Interestingly, while the observations show similar decadal variability for summer between the upper and the lower Indus, this is not the case for the reanalyses, except maybe for the twentieth century reanalyses, and ~~of course~~ ERA-5~~, which represent well the observations~~. Notably, ERA5 has an inter-annual correlation with GPCC-monthly that is higher than the correlation between GPCC and CRU for all three panel in (Figure 10), suggesting it is at least as able as observational datasets.

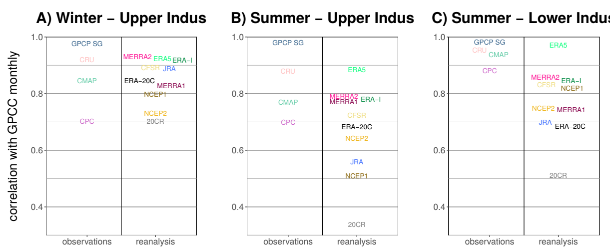


Figure 10. Inter-annual correlation on the period 1981-2010 between GPCC-monthly and the other datasets covering that period. The correlations are computed for specific seasons and domains. We split the result by type of dataset (Observation and renalysis)

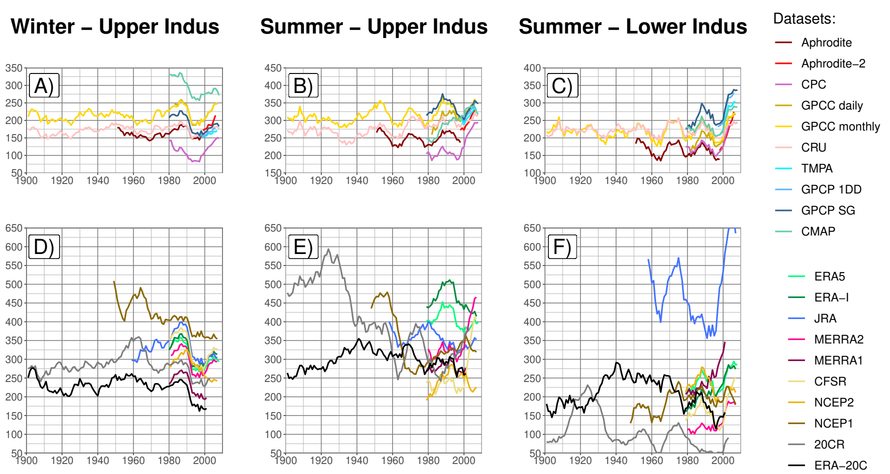


Figure 11. Decadal variability of precipitation using a 10-year running mean for different seasons and domains (Winter in the upper Indus: A and D; summer in the upper Indus: B and E; summer in the lower Indus: C and F) and the different datasets (Observational datasets: A, B and C; Renalysis datasets: D, E and F)

4 Conclusions

In this study, we have compared a large number of precipitation datasets of different types across two distinct zones of the Indus watershed: six datasets are based only on rain gauges, four are derived from satellite observations, and ten from reanalysis. We have ~~showed~~ shown that the number and diversity of the datasets help to identify and quantify the limitations and abilities of each of them, which in turn enables a better estimation of the ~~true values~~uncertainties.

~~It is quite usual to validate reanalysis data using observations~~We have compared the datasets on the basis of the annual mean precipitation, the seasonal cycle, as well as the variability over time scales from one day to 10 years. We have relied on the literature to evaluate the different sources of uncertainty and have interpreted the mean differences between datasets in terms of their quality. We have suggested that the similarities in variability can directly be interpreted in terms of quality, especially when comparing datasets with no common methods or data source. Most reanalyses do not assimilate precipitation observations, which makes it possible to cross-validate between observational and reanalysis data based on variability. Regardless of the observational datasets used as a reference~~. When using daily correlation over a domain much larger than the spatial resolution, we found that most reanalyses are as similar to the observations as observational datasets are to each other. That is, reanalyses are as good as observational datasets for representing~~ , we have found that some reanalyses have significantly higher correlation with that reference than other reanalyses, which we have interpreted as a sign of good quality. Conversely, when using a reanalysis as a reference, some observational datasets have significantly higher correlation than others. The use of reanalyses to validate observational datasets is justified by the quality of reanalysis products demonstrated in this study. Specifically, at the scale of the Indus basin, and for the daily variability, ~~thereby justifying the possibility of using reanalyses as references for validating observations . The datasets validated can be ranked depending on the correlation with the reference, and, if those datasets are independent from the reference, as are most reanalyses compared to observations, the ranking can be interpreted in terms of quality and proximity to~~ the ~~true variability. We found that the ranking is generally not~~ same level of similarity between the reanalyses and observations is also seen between the observational datasets themselves.

We have used the Pearson correlation to compare the datasets, although this has some limitations. For example, it is affected by extreme values, that is, in our context, abnormally large precipitation events. These lead to 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 ~~reference, which increases the confidence in the results. Interestingly, the similarity between observations and reanalyses tends to decrease at lower frequencies, which we interpret as a difficulty of the reanalysis in representing the impact of larger scale drivers of precipitation(i.e. teleconnections). However, some reanalyses remain good enough to be used as references at the monthly and inter-annual timescale~~difficulties in representing the lowest precipitation rates, although these rates can explain some of the biases.

One of our findings concerns the important uncertainty in fine scale spatial patterns of precipitation, particularly in the upper Indus, where precipitation is the most heterogeneous. Important discrepancies remain between datasets, which explain part of the differences in mean precipitation. This issue needs to be tackled in observational datasets by including more measurements and by updating the climatology used in the interpolation methods. In reanalysis products, higher resolution and better modelling of small scale processes are likely needed to improve confidence in the spatial pattern of precipitation. In this study, we have deliberately selected two large study areas, which has increased the confidence in the datasets. Area-wide correlation particularly improves the significance of the variability analysis, compared to a point-wise correlation.

~~The season also impacts~~ We have also found that the quality of ~~datasets. All datasets represent the variability at all timescales very similarly during winter in~~ the datasets depends on the season. Rain gauge measurements suffer from important underestimations in winter for the upper Indus. ~~Only the satellite-based datasets represent the winter daily variability with more difficulty. In summer, however, in either of the domains, the correlations drop, primarily due to less reliable reanalyses , but also possibly less reliable rain gauge-based datasets. The dissimilarity is especially evident for the decadal variability, where only ERA5 represent the same behaviour as the observations.~~

~~The method applied for the variability cannot be used to quantify the actual precipitation amount. Despite similar representation of the variability , the datasets exhibit important differences in the precipitation amount~~Most satellite-derived datasets even further amplify this bias. By contrast, reanalyses perform best during winter. Particularly, the most recent reanalyses produce a very similar amount of winter precipitation and its variability is similar to the observations at all timescales. We have ~~relied on the literature to evaluate the different sources of uncertainty. Winter is an interesting season in the upper Indus, as the rain gauge-based datasets are known to underestimate snowfall significantly, especially in areas that are difficult toaccess. The reanalyses are not affected by those biases, and, in light of their very good quality on the variability, we consider the precipitation amount in the reanalyses more realistic than in the observations. By contrast, precipitationamount during summer~~ suggested that their amount of precipitation is closer to reality than the observations, although some overestimations are possible, due to, for example, misrepresentation of the lowest precipitation rates. Summer precipitation, in both study areas, is much more uncertain ~~, echoing the results on the variability. Notably, in some reanalyses the seasonality is delayed~~in the reanalyses in total amount, seasonality, and variability. In contrast, satellite observations perform better in summer than in winter and seem to bring additional information to rain gauge measurements.

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

More specifically, APHRODITE is the best observational dataset for daily and monthly variability, thanks to a large number of observations in the whole basin. However, it also exhibits drier conditions than most of the other datasets, which is partially caused by the interpolation method it uses and possibly by a lower quality of the data. Surprisingly, APHRODITE-2 is not as good, especially for the longer term variability, as it removes some observations in areas with an already lower density of measurements. CPC is ~~also a dry dataset, although this bias is reduced towards the end of the period covered. Similarly, the quality is much lower than other observational datasets during the 1980s and 1990s, and includes~~ the least reliable observational dataset, particularly for the upper Indus, with a large dry bias compared to GPCC-monthly, the lowest correlation scores at all time scales, and an error on the dates ~~, but improves significantly after 2000. Since the number and location of measurements does not change~~before 1998. However, its quality significantly improves after 2005, which, we suspect, is due to a change in the quality of ~~those measurements. GPCC-daily also sees an increase of daily correlation after 2000, and reaches APHRODITE-2 levels in the lower Indus. There, the improvement can be related to an increase in the number of measurements. Although it uses a very low number of measurements, its monthly mean is constrained by GPCC-monthly, which proves to be a good approach. Indeed,~~ the data source. GPCC-monthly is one of the most reliable datasets ~~in term of variability, but also~~ both in terms of amount ~~as it is the only rain gauge-based dataset to include a correction for the measurement under-catchment, although the correcting factor is probably underestimated~~and variability. GPCC-daily relies on GPCC-monthly for its monthly mean. The very low number of daily measurements included in the early part of the covered period limits its quality, but this quickly improves as more observations are included.

Satellite-based datasets are very dependent on the quality of the rain-gauge product they integrate. The added-value of satellite observations remains limited at the basin scale~~, and is probably more important in the flatter lower Indus , the summer season, and for longer term variability. Significantly, they do not correct the winter bias and show lower quality than rain-gauge based datasetson the daily correlation~~. The signal is degraded during winter for the upper Indus, while better results in the lower Indus suggest slightly wetter conditions than the rain gauge-based datasets. Importantly, the quality of satellite-based datasets resides in their near real time availability as well as their higher temporal and spatial resolution than rain gauge based datasets.

The ~~reanalysis that represents best the observations is ERA5. It represents the variability at all timescales and the seasonality as well. It is the newest reanalysis and the only one to assimilate precipitation measurements. For that latter reason, however, it is somewhat problematic to use it to validate the observational datasets, although it does not give different results than other references.~~ quality of reanalysis datasets has clearly improved since the first datasets were released. ERA5 is the latest reanalysis and clearly stands out as the one representing best the observations, in terms of amount, seasonality, and variability at all time scales investigated. Remarkably, it is the only reanalysis representing the decadal variability of the summer precipitation for both study areas as it is seen in the observations. Furthermore, for the daily to inter-annual variability, the best performing observational dataset has often a better level of similarity with ERA5 than with other observational datasets. Some of these qualities can be derived from its high resolution, which allows the representation of interesting fine scale features, as well as the assimilation of precipitation measurements.

After ERA5, ERA-Interim, MERRA1, and MERRA2 have relatively similar performance. ~~We found some dependency between~~ Reichle et al. (2017) showed that the soil moisture content was not improved over South Asia from MERRA1 to MERRA2~~and CPC, due to~~, neither in terms of variability nor biases, despite the use of CPC to correct the precipitation input ~~on~~ to the land surface ~~. This correction unlikely improved MERRA2's~~ model of MERRA2. Given the difficulties of CPC to represent precipitation in the ~~domain of study. JRA is relatively good over the period it covers, but exhibits a decrease in quality around 1990 in~~ Indus basin, correcting the modelled precipitation with this dataset probably does not improve the signal. In this study, we were able to show that the correction with CPC feeds back locally on the modelled precipitation, particularly at the monthly scale for the upper Indus~~and 1995~~. We have also suggested that the dry bias of MERRA2 in the lower Indus~~that lasts for around 10 years in each case. The summer seasonality is not very well reproduced either, especially in the lower Indus where JRA~~, and the decrease score on the daily variability compared to MERRA1, is also due to that correction.

The confidence in JRA's precipitation in the upper Indus is generally high, but drops for the daily and monthly variability in the 1990's. By contrast, it represents overly wet conditions for the lower Indus. CFSR has problems reproducing the daily variability and the seasonality of the monsoon, especially in the upper Indus. This is probably improved by the latest version that started in April 2011. However, it would likely be better to treat the two versions separately as it seems the new version produces somewhat different statistics of precipitation. The twentieth century reanalyses, ~~including only surface observation~~which includes only surface observations, are not as good as the others, especially in winter. However, while 20CR barely reproduces any of the variability depicted by the observation, ERA-20C has much better capabilities, close to NCEP1 and CFSR, especially during summer. Neither 20CR nor ERA-20C represent the decadal variability ~~as shown by~~ seen in the observation before 1980.

~~This study has focused on the analysis of precipitation using basic tools such as mean and correlation. More complex tools also exist for a more thorough analysis of the precipitation statistics~~Finally, large uncertainties remain about precipitation in the upper Indus, but one should not treat all datasets equally. We have demonstrated that specific datasets represent the precipitation better, which helps to narrow down the uncertainty. Particularly, ~~correlations are greatly impacted by extreme values, while important biases could also occur from inaccurate representations of the lowest precipitation rates. Moreover, we deliberately selected a large domain of study to improve the confidence in the datasets . Observation-based datasets still miss important patterns due to a lack of measurements in key areas, while reanalyses are even worse. However, either datasets could be used to constrain finer scale analysis.~~

we have argued that precipitation from reanalyses and observational datasets can both be useful for cross-validation. They can also be used for quality monitoring. Daily correlation of precipitation for key areas can be performed between a series of datasets with near real time updates. Changes in correlation between one or several datasets would therefore highlight a change in quality that would need to be investigated.

Code availability. The code in R used to produce the figures and the tables is available upon request to the first author

Author contributions. Original idea, analysis and text by JPB, guidance and review by MH and CP

Competing interests. The authors declare that they have no conflict of interest

Disclaimer. TEXT

Acknowledgements. We acknowledge all the teams involved in the producing of the precipitation datasets used in this study and gratefully thanks them for making the datasets available online. We also thank our editor and the two anonymous reviewers for comments that improved the paper. This research was carried out as part of the *TwoRains* project, which is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 648609)~~. We also thank our editor and reviewers for comments that improved the paper~~.

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, Journal of Geophysical Research: Atmospheres, 108, 2003.

Adler, R., Sapiano, M., Huffman, G., Bolvin, D., Wang, J., Gu, G., Nelkin, E., Xie, P., Chiu, L., Ferraro, R., et al.: New Global Precipitation Climatology Project monthly analysis product corrects satellite data shifts, GEWEX News, 26, 7–9, 2016.

Ali, G., Rasul, G., Mahmood, T., Zaman, Q., and Cheema, S.: Validation of APHRODITE precipitation data for humid and sub humid regions of Pakistan, Pakistan Journal of Meteorology, 9, 2012.

Archer, D. R. and Fowler, H. J.: Spatial and temporal variations in precipitation in the Upper Indus Basin, global teleconnections and hydrological implications, Hydrology and Earth System Sciences Discussions, 8, 47–61, 2004.

Archer, D. R., Forsythe, N., Fowler, H. J., and Shah, S. M.: Sustainability of water resources management in the Indus Basin under changing climatic and socio economic conditions, Hydrology and Earth System Sciences, 14, 1669–1680, 2010.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207–224, 2019.

Cannon, F., Carvalho, L. M., Jones, C., and Bookhagen, B.: Multi-annual variations in winter westerly disturbance activity affecting the Himalaya, Climate dynamics, 44, 441–455, 2015.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al.: The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, 2011.

Dahri, Z. H., Moors, E., Ludwig, F., Ahmad, S., Khan, A., Ali, I., and Kabat, P.: Adjustment of measurement errors to reconcile precipitation distribution in the high-altitude Indus basin, International Journal of Climatology, 38, 3842–3860, 2018.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Dimri, A. and Chevuturi, A.: Western disturbances-an Indian meteorological perspective, Springer, 2016.

Douville, H., Chauvin, F., and Broqua, H.: Influence of soil moisture on the Asian and African monsoons. Part I: Mean monsoon and daily precipitation, Journal of Climate, 14, 2381–2403, 2001.

Ebert, E. E., Janowiak, J. E., and Kidd, C.: Comparison of near-real-time precipitation estimates from satellite observations and numerical models, Bulletin of the American Meteorological Society, 88, 47–64, 2007.

Ensor, L. A. and Robeson, S. M.: Statistical characteristics of daily precipitation: comparisons of gridded and point datasets, Journal of Applied Meteorology and Climatology, 47, 2468–2476, 2008.

Faisal, N. and Gaffar, A.: Development of Pakistan's new area weighted rainfall using Thiessen polygon method, Pakistan Journal of Meteorology, 9, 2012.

Filippi, L., Palazzi, E., von Hardenberg, J., and Provenzale, A.: Multidecadal variations in the relationship between the NAO and winter precipitation in the Hindu Kush–Karakoram, Journal of Climate, 27, 7890–7902, 2014.

Gardelle, J., Berthier, E., and Arnaud, Y.: Slight mass gain of Karakoram glaciers in the early twenty-first century, Nature geoscience, 5, 322, 2012.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, 2017.

Ghulami, M., Babel, M. S., and Singh Shrestha, M.: Evaluation of gridded precipitation datasets for the Kabul Basin, Afghanistan, International journal of remote sensing, 38, 3317–3332, 2017.

Goodison, B., Sevruk, B., and Klemm, S.: WMO solid precipitation measurement intercomparison: Objectives, methodology, analysis, Atmos. Depos, 179, 57–64, 1989.

Habib, E., Ciach, G. J., and Krajewski, W. F.: A method for filtering out raingauge representativeness errors from the verification distributions of radar and raingauge rainfall, Advances in Water Resources, 27, 967–980, 2004.

Harris, I. and Jones, P.: CRU TS4. 00: Climatic Research Unit (CRU) Time-Series (TS) version 4.00 of high resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2015), Centre for Environmental Data Analysis, 25, 2017.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, International journal of climatology, 34, 623–642, 2014.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Munoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E., and Zuo, H.: Operational global reanalysis: progress, future directions and synergies with NWP, ERA-report, Serie 27, https://doi.org/10.21957/tkic6g3wm, https://www.ecmwf.int/node/18765, 2018.

Hewitt, K.: The Karakoram anomaly? Glacier expansion and the 'elevation effect,'Karakoram Himalaya, Mountain Research and Development, 25, 332–341, 2005.

Huffman, G. J. and Bolvin, D. T.: Version 1.2 GPCP one-degree daily precipitation data set documentation, NASA, Goddard Space Flight Center, Greenbelt, MD, USA https://rda. ucar. edu/datasets/ds728, 3, 2013.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of hydrometeorology, 8, 38–55, 2007.

Hunt, K. M., Turner, A. G., and Shaffrey, L. C.: The evolution, seasonality and impacts of western disturbances, Quarterly Journal of the Royal Meteorological Society, 144, 278–290, 2018.

Hussain, S., Song, X., Ren, G., Hussain, I., Han, D., and Zaman, M.: Evaluation of gridded precipitation data in the Hindu Kush–Karakoram–Himalaya mountainous area, Hydrological sciences journal, 62, 2393–2405, 2017.

Immerzeel, W., Wanders, N., Lutz, A., Shea, J., and Bierkens, M.: Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff, Hydrology and Earth System Sciences, 19, 4673–4687, 2015.

Iqbal, M. F. and Athar, H.: Validation of satellite based precipitation over diverse topography of Pakistan, Atmospheric Research, 201, 247–260, 2018.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–472, 1996.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., and Potter, G.: Ncep–doe amip-ii reanalysis (r-2), Bulletin of the American Meteorological Society, 83, 1631–1644, 2002.

Khan, S. I., Hong, Y., Gourley, J. J., Khattak, M. U. K., Yong, B., and Vergara, H. J.: Evaluation of three high-resolution satellite precipitation estimates: Potential for monsoon monitoring over Pakistan, Advances in Space Research, 54, 670–684, 2014.

Kishore, P., Jyothi, S., Basha, G., Rao, S., Rajeevan, M., Velicogna, I., and Sutterley, T. C.: Precipitation climatology over India: validation with observations and reanalysis datasets and spatial trends, Climate dynamics, 46, 541–556, 2016.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al.: The JRA-55 reanalysis: General specifications and basic characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93, 5–48, 2015.

Laghari, A., Vanham, D., and Rauch, W.: The Indus basin in the framework of current and future water resources management, Hydrology and Earth System Sciences, 16, 1063, 2012.

Legates, D. R. and Willmott, C. J.: Mean seasonal and spatial variability in gauge-corrected, global precipitation, International Journal of Climatology, 10, 111–127, 1990.

Ménégoz, M., Gallée, H., and Jacobi, H.: Precipitation and snow cover in the Himalaya: from reanalysis to regional climate simulations, Hydrology and Earth System Sciences, 17, 3921–3936, 2013.

Newlands, N. K., Davidson, A., Howard, A., and Hill, H.: Validation and inter-comparison of three methodologies for interpolating daily precipitation and temperature across Canada, Environmetrics, 22, 205–223, 2011.

Palazzi, E., Von Hardenberg, J., and Provenzale, A.: Precipitation in the Hindu-Kush Karakoram Himalaya: observations and future scenarios, Journal of Geophysical Research: Atmospheres, 118, 85–100, 2013.

Palazzi, E., von Hardenberg, J., Terzago, S., and Provenzale, A.: Precipitation in the Karakoram-Himalaya: a CMIP5 view, Climate Dynamics, 45, 21–45, https://doi.org/10.1007/s00382-014-2341-z, https://doi.org/10.1007/s00382-014-2341-z, 2015.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J.-N., et al.: ERA-20C: An atmospheric reanalysis of the twentieth century, Journal of Climate, 29, 4083–4097, 2016.

Prakash, S., Mitra, A. K., AghaKouchak, A., and Pai, D.: Error characterization of TRMM Multisatellite Precipitation Analysis (TMPA-3B42) products over India for different seasons, Journal of Hydrology, 529, 1302–1312, 2015.

Rahman, S., Sengupta, D., and Ravichandran, M.: Variability of Indian summer monsoon rainfall in daily data from gauge and satellite, Journal of Geophysical Research: Atmospheres, 114, 2009.

Rajbhandari, R., Shrestha, A., Kulkarni, A., Patwardhan, S., and Bajracharya, S.: Projected changes in climate over the Indus river basin using a high resolution regional climate model (PRECIS), Climate Dynamics, 44, 339–357, 2015.

Rana, S., McGregor, J., and Renwick, J.: Precipitation seasonality over the Indian subcontinent: An evaluation of gauge, reanalyses, and satellite retrievals, Journal of Hydrometeorology, 16, 631–651, 2015.

Rana, S., McGregor, J., and Renwick, J.: Wintertime precipitation climatology and ENSO sensitivity over central southwest Asia, International Journal of Climatology, 37, 1494–1509, 2017.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P., Koster, R. D., and De Lannoy, G. J.: Assessment of MERRA-2 land surface hydrology estimates, Journal of Climate, 30, 2937–2960, 2017.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., et al.: MERRA: NASA's modern-era retrospective analysis for research and applications, Journal of climate, 24, 3624–3648, 2011.

Rudolf, B. and Rubel, F.: Global precipitation, in: Observed global climate, chap. 11, pp. 1–24, Springer, 2005.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al.: The NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1058, 2010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., et al.: The NCEP climate forecast system version 2, Journal of Climate, 27, 2185–2208, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theoretical and Applied Climatology, 115, 15–40, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., and Ziese, M.: GPCC Full Data Monthly Product Version 2018 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges Built on GTS-Based and Historical Data, Deutscher Wetterdienst: Offenbach am Main, Germany, 2018.

Sevruk, B.: International comparison of national precipitation gauges with a reference pit gauge., WMO Instrument and Observing Methods Report, No. 17, 111, 1984.

Singh, S. P., Bassignana-Khadka, I., Singh Karky, B., and Sharma, E.: Climate change in the Hindu Kush-Himalayas: the state of current knowledge, Tech. rep., International Centre for Integrated Mountain Development (ICIMOD), 2011.

Syvitski, J. P., Kettner, A. J., Overeem, I., Giosan, L., Brakenridge, G. R., Hannon, M., and Bilham, R.: Anthropocene metamorphosis of the Indus Delta and lower floodplain, Anthropocene, 3, 24–35, 2013.

Tozer, C., Kiem, A., and Verdon-Kidd, D.: On the uncertainties associated with using gridded rainfall data as a proxy for observed, Hydrology and Earth System Sciences, 16, 1481–1499, 2012.

Tustison, B., Harris, D., and Foufoula-Georgiou, E.: Scale issues in verification of precipitation forecasts, Journal of Geophysical Research: Atmospheres, 106, 11 775–11 784, 2001.

Wang, J. and Wolff, D. B.: Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurements, Journal of Applied Meteorology and Climatology, 49, 310–324, 2010.

Wang, Q., Yi, S., and Sun, W.: Precipitation-driven glacier changes in the Pamir and Hindu Kush mountains, Geophysical Research Letters, 44, 2817–2824, 2017.

WMO: Guide to Hydrological Practices. Volume I. Hydrology–From Measurement to Hydrological Information, 2008.

Wolff, M., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T., and Brækkan, R.: Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study, Hydrology and Earth System Sciences, 19, 951–967, 2015.

Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, Bulletin of the American Meteorological Society, 78, 2539–2558, 1997.

Xie, P., Chen, M., and Shi, W.: CPC unified gauge-based analysis of global daily precipitation, in: Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc, vol. 2, 2010.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Bulletin of the American Meteorological Society, 93, 1401–1415, 2012.

Ziese, M., Rauthe-Schöch, A., Becker, A., Finger, P., Meyer-Christoffer, A., and Schneider, U.: GPCC full data daily version 2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and historic data, 2018.