Reviewer 1

Precipitation Pattern in the Western Himalayas revealed by Four Datasets

Hong Li, Jan Erik Haugen, Chong-Yu Xu

HESS manuscript: **hess-2017-296**

Recommendation: **Reconsider after major corrections**

This paper provides a limited examination of four precipitation datasets over the Western Himlayan region of India, and is primarily performed by comparing the spatial pattern and trends of the datasets to identify which datasets are suitable for hydrological modelling. The results are short, but there is a conclusion that WRF and APHRODITE are useful tools. While the paper address an important issue faced by hydro-meteorologists, there are still major problems in the strength and sufficiency of the methodology and of the analyses. While there are improvements in this revised version, there are major issues in the justification of the methodology and the analysis of the results, as well as a lack of detail regarding the WRF modelling hinders the replicability of the results. In my opinion, the current manuscript does not offer sufficiently thorough support to amount to a substantial advancement of understanding precipitation patterns in this region. Therefore, I recommend that the manuscr ipt should be reconsidered for publication after major revisions.

Reply: We first thank the editor and the reviewers for your positive evaluation of our work in general, and we appreciate very much the professional and constructive comments raised by reviewers. In the revised version, we add more information about the WRF model setup. Additionally, we do a Kolmogorov–Smirnov test about difference between four datasets. More importantly, we add data from a rain gauge as a benchmark and compare it with four gridded datasets at various time scales. Last but not least, we carefully revise and improve the manuscript from both scientific presentation as well as English.

Major comments:

1. The major limitation of this manuscript is that there is still a lack of benchmarking to determine which is the “best” dataset for use in hydrological modelling. While the discharge and runoff measurements are a step in this direction, I have to question

why the authors do not simply use *in situ* observations as the benchmark, and compare the four data sets to this data. Daily and monthly observed precipitation is freely available from the National Climatic Data Center (https://gis. ncdc. noaa. go v/maps/nce i/cdo/da ily) for this region. It would be prudent for this data to be included in this manuscript, given that there is currently no direct benchmark for precipitation used. Although the authors identify the lack of precipitation measurements and their quality is limited in the Himalayan region, including this data and the MODIS data as the benchmark will dramatically improve the veracity of the analysis and conclusions. This will also help determine what added value the "best" dataset has over the others.

Reply: We thank the reviewer for recognizing our efforts by including MODIS data and for suggestion to compare with the other available dataset. We visit the website of the National Climatic Data Center and indeed find many stations there. We download all precipitation data in the study area for the study period from 1981.01.01 to 2007.12.31. In total, there are 38 stations which have measurements during this period. However, many measurements are missing as shown in Figure A2. The maximum missing rate is as high as 99.99%, and the minimum is 26%, which makes the use of these data somewhat problematic. However, we include *in situ* measurements from a rain gauge, Bhuntar. We get the data from our research partner in India in a previous research project and the data have been used in hydrological modelling for the Beas Basin. More information of this benchmark data and analysis is given in "3.5 Gauge data" and "5.1 Comparison gridded precipitation datasets with gauge data". Of course, we also update other text to make more readable, and all changes are shown in the color of blue.

2. As well as the spatial pattern and trends in precipitation, whether or not extreme precipitation events are captured by the individual data sets would also be of value to hydro-meteorologists. The "best" data set cannot be identified just from the mean spatio-temporal characteristics, when extreme events are important for accurate hydrological modelling.

Reply: Thanks for suggestion. We agree that extreme events are important for accurate hydrological modelling. In the revised version, we include extreme precipitation analysis based on annual maximum daily precipitation. Results show that the WRF dataset gives the best estimations of different quantiles. This part is presented in "5.1 Comparison gridded precipitation datasets with gauge data".

3. A missing component of the work as a whole is that there is no proven identification of the real source of the differences. The differences are just explained, without any digging into the simulated processes and forcing data to really find the causes. The key to a valuable comparison study is to at least isolate the source of the differences so that others may understand and build on the work. While the authors identify that differences between the data sets may be attributed to differences in grid size resolution, why not regrid the four data sets onto the same grid and then compare and analyse? On line 26, line 5, the authors state that horizontal resolution is the reason for APHRODITE's better performance. Is this true, or is it an artefact of the individ ual observations? The authors later attribute ERA-Interim's poor performance to individual observations, rather than horizontal resolution. So which is it, horizontal resolution or an artefact of the number/distribution/quality of

observations? This is contradictory and confusing and does not isolate the true reason for differences between data sets.

Reply: We did not re-grid, as we thought re-gridding coarse resolution to finer resolution only adding more colors rather than really adding information. We have improved the writing for this part. The reasons for differences between datasets are data source and used methodology to produce the data. The spatial resolution is important in showing spatial variability, which should be at a hydrological scale. We state that APHRODITE is much better than IMD, because its resolution is good to more clearly show rain belt at the mountain's foothill. However, we also state that the APHRODITE dataset shows too little precipitation (less than 300 mm/year) at the northeast of the study area, as shown in Figure 2 and "4.1 Spatial variations". All the four datasets are widely used in climate and hydrological studies, but most researchers only select one or two datasets. Here we select four types and they are representative in their own type. We aim to present the differences.

On the other side, we agree that comparing the same spatial resolution is a good way to present the result. Therefore, we interpolate other three datasets to the IMD grids as shown in Figure A1. The general patterns are still the same, but some information is missing at the coarse spatial resolution.

4. A big question is whether the differences between the data sets are statistically significant. The absence of this is a real deficiency and should be addressed. The reader just doesn't know whether the small differences (e.g. in the spatial differences) have any statistical meaning and which of the different data sets are truly better. It may be that it cannot be concluded that any of these are statistically significantly different. But, even to know that would be of value. Without significance testing results, the reader can't conclude either way.

Reply: To address the question of statistical significance, we add Table 3 "P-value of the tailed Kolmogorov–Smirnov test on differences of on annual precipitation (mm/year) among the datasets. The p-value indicates strong evidence against the null hypothesis. It is typically to reject the null hypothesis, which is two datasets are the same here, when p-value is not greater than 0.05." The test is a nonparametric test that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test.

In addition, we include the Bhuntar rain gauge, we not only compare monthly anomaly, but also maximum and minimum, as shown in Figure 12. This is also a type of significance analysis to show how reliable the diffidence is.

5. Some very crucial information is required in the WRF section to clarify how the simulations were conducted. It is the only dataset to have been produced by the authors, and the scant detail currently given means that there is no way of replicating their results. Details that need to be clarified in the text are as follows:
   a. What justification is there for using such a precise resolution of 16306 km,

rather than 16 km? Maussion *et al*. and Li *et al*. are cited for their model setups, despite both studies using a nested model approach (with the innermost domain less than 3 km). The authors have assumed that model performance (of the other two studies) will scale appropriately to 16306 km. What sensitivity simulat ions were conducted to test for this? Since this region is surrounded by complex topographic features, and the aim of this paper is to emphasise the benefit of WRF over the other coarser gridded data sets, it would seem appropriate to use a nested model approach in this study going to a resolution of > 5km.

Reply: The spatial resolution of the WRF dataset is 16 km, and we have corrected the mistake in the manuscript. We have also explained that there are two reasons that we did not use a nested approach. The first one is that the forcing data are at 0.75 degree, and the step from forcing data to the WRF model is only 4. This step is fine in climate simulation. The second reason is consideration for the following hydrological model work. There are several catchments and we are trying to collect more runoff data. It would be too many small nested domains. Therefore, we do not use a nest approach.

b. The authors state that it is not easy to determine the optimised selection of parameterization schemes – this is correct, and setting the model up for a region often involves intensive testing of the physics options and sensitivit y simulations. So please justify why some options from Maussion *et al.* and some options from Li *et al*. have been selected without proper testing? This is contradictory, and there is no adequate justification for doing this (just use one setup or the other).

Reply: There are many options in terms of setting the WRF model. We used a setting that we have used before and it also gave reliable results. We have clarified this part in the revision.

c. What was the height of the lowest vertical level (usually ~25m)?

Reply: We apologies we did not give the details. The height of the lowest model level depends on the surface pressure. The height of the lowest model level varies between 15 and 27 meters depending on the terrain height. We have added this information in “3.4 WRF dataset” and Table 1.

d. Did the authors consider using other reanalysis products (e.g. MEERA, GFS)?

Reply: We do not consider other analysis products. We use ERA-interim to force the WRF model and would like to see if the WRF model at high resolution can add more values. As the reviewer mentioned, there are indeed quite many reanalysis products available at present. However, we really cannot include all of them her. Otherwise, it is too much on reanalysis products. We have clarified this part.

e. Is there any erroneous data at the boundary edges that run through the Himalayas?

Reply: We interpret the question as the error at the WRF boundary. The errors are usually shown at three to five grids at the boundary edges. Here we use a very large domain and the study area is a small central area (Figure 1). In addition, the large domain covers a considerable area of ocean, and therefore, the inflow boundary is good.

f. What topography data set was used in WRF? Did the authors compare the modelled topography to the real observed topography? Even good elevation agreement, such as being within 10m, can affect the results at the precision reported.

Reply: We have clarified that we use the 10m topography and land use provided by WRF pre-processing. The source is listed at http://www2.mmm.ucar.edu/wrf/users/download/get_sources_wps_geog.html ("3.4 WRF dataset"). We agree that elevation is very important and topography plays an important role in precipitation generation mechanism as well as in climate model simulations. We believe that the NCAR and UCAR have compared and concluded that they are of high quality.

g. At no point in this section have the authors explained how the data has been extracted from WRF. Was the data extracted from the nearest grid point or an interpolation?

Reply: We use the native model grids within our study area, and we do not do interpolation except Figure A1. To compare with the Bhuntar rain gauge, we extract data of the nearest point. This description lies in "3.5 Comparison with gauge data" as well as titles of two figures and a table (Figure 11, Figure 12 and Table 5).

h. Were these runs continuous, restarted, or reinitialised, and what was done to check for model spin up and drift? It is really important to make this clear.

Reply: Continuous run of climate models requires too much time and the jobs cannot start in the queue system. Therefore, the model results are restarted about every five years. The simulation is from 1979 and results since 1981 are used here. We have added this description in "3.5 WRF dataset".

We do not mention the drift problem in the manuscript since it usually appears in global climate models. But from the results, the non-parametric Mann-Kendell test for annual precipitation (Table A1), for monthly precipitation (Figure 7), and seasonal precipitation (Figures 8-11 and 13), the WRF simulation does not show additional long term trends compared with other datasets.

6. There is still a lack of verification indicators, and including the data from available rain gauges could easily be done. As highlighted above, WMO rain gauge data is freely available. Daily precipitation may also be available from the Bhakra Beas Management Board (see e.g. Norris *et al*. 2017)

Norris, J., Carvalho, L. M., Jones, C., Cannon, F., Bookhagen, B., Palazzi, E., & Tahir, A. A. (2017). The spatiotemporal variability of precipitation over the Himalaya: evaluation of one-year WRF model simulation. *Climate Dynamics*, *49*(5-6), 2179-2204.

Reply: We checked the website of Bhakra Beas Management Board and did not find any possibility to download data, and there is even no contact information. It is likely that Norris *et al*. 2017 got the data because they have cooperation with their Indian partners. For WMO rain gauge, as mentioned before, we have downloaded and analyzed them, but there are too many missing data and it is difficult to use here. However, we include observation data from a rain gauge, Bhuntar. We get the data from our research partner in India in a previous research project along with the discharge data, and the data have been used in hydrological modelling for the Beas Basin. More information of this benchmark data and analysis is given in "3.5 Gauge data" and "5.1 Comparison gridded precipitation datasets with gauge data". Of course, we also update other text to make more readable, and all changes are marked by blue color.

Other comments:

7. I recommend that the authors carefully proofread the manuscript again. There are numerous grammatical and typographical, none of which I have corrected. Occasionally, the clarity of the arguments is lost by poor sentence structure. This makes the manuscript hard to follow at times.

Reply: Thanks for suggestion. We will carefully proofread and correct typo and grammatical errors.

8. Revise lines 5 – 9 on page 2, the paragraph is hard to follow

Reply: They are revised.

9. Although the weaknesses are discussed, I would like to see the advantages of each of the datasets also included in Table 1, as well as in the corresponding paragraph, for a two-sided comparison.

Reply: They are added as the reviewer suggested.

10. Line 5, page 3 – "This interaction brings plenty of precipitation" – This is a qualitative statement which is not useful here. Use the WMO data to state exactly how much; be precise.

Reply: Due to the reason we mentioned before, we refer to India rainfall report 2016 (http://hydro.imd.gov.in/hydrometweb/(S(0rlpja453hfxgw45x041xyqr))/PRODUCTS/Publications/Rainfall%20Statistics%20of%20India%20-%202016/Rainfall%20Statistics%20of%20India%20-%202016.pdf) and revised to be more accurate.

11. Lines 16 – 20, page 3 – What is the relevance of this paragraph in this section? What is happening to the glacier? Consider removing, or expand to make relevant.

Reply: we moved this paragraph to "5.3 Implications for glaciers" and expanded.

12. Have you considered also including TRMM (3B42V7) as well as/instead of APHRODITE in the analysis? TRMM is the most reliable decadal dataset of gridded precipitation estimates in the Himalaya (Norris *et al.*), and APHRODITE is mostly based on low elevation sites.

Reply: We indeed have considered TRMM. However, TRMM is launched in 1997 and there are fewer overlap years with other datasets. Therefore, we do not include her.

13. Explicitly state where the discharge measurements were sourced. This information appears to be missing.

Reply: The discharge stations are operated by Central Water Commission regional office in India. We have added this information in the revised version.

14. Lines 10 – 12, page 6 – This detail is more appropriate in the method section

Reply: They are moved to "3.4 WRF dataset".

15. Lines 16 – 21, page 6 – this paragraph does not make sense at all. Be concise and analytical.

Reply: We removed this in the revised version.

16. Lines 3 – 7, page 7 – This section should also be in the method section. If you are not using *in situ* observations as the benchmark for "best" then use this as your justification in the methodology section.

Reply: They are moved to "3.6 Discharge".

17. Figure 1 – please show the full WRF domain and river catchments (in panelled plots) so that Figures 2, 6, 8, 10 are more easily understood. You cannot discern ice thickness differences in the figure, so consider removing.

Reply: The figure is changed as the reviewer suggestions.

18. Figure 2 – Without a clearer geographical setting, it is difficult to understand what is being shown here. What has been masked? A discrete color bar is needed, as well as a panel of topography.

Reply: The geographical setting shown as the x and y axis. We add elevation and modify figure to be seen more clearly.

19. Figure A1 – This is not useful. Include the catchment basins in Figure 1 and remove this figure.

Reply: Thanks for suggestion. We removed Figure A1 and revised Figure 1. The catchment basins are very small compared to the whole study area. In the new version, we add also the boundary of India and disputed area as well as small window map to show their locations more clearly.

Reviewer 2

Figure 1: This figure is not following proper international boundaries. Wrong depiction of map of India (shaded region)

Reply: We are sorry for disputed boarders and we change the label to "India and disputed area".

Figure A1: Very small map, and geographical location of Beas is not clear from this map.

Reply: Thanks for suggestion. We removed Figure A1 and revised Figure 1. The catchment basins are very small compared to the whole study area. In the new version, we add also the boundary of India and disputed area as well as small window map to show their locations more clearly.