

Response to Reviewer #1's comments:

This paper aims to enhance both the quantity and quality of the streamflow data and metadata on gauging stations in India. The author integrated multiple datasets to update the information of gauging stations and compile historical streamflow data through a rigorous quality control process. The information and datasets provided in this study will prove valuable to hydrologic communities, while the methodologies presented can be adopted by both the administrative organizations and research communities to homogenize the hydrological datasets in India. I found this paper is generally well written. I recommend this paper to be accepted after few minor comments below are addressed.

The author would like to thank the reviewer for the positive feedback and the suggestions made to improve the readability of the manuscript. The author agrees with all of the comments, and revisions were made accordingly. Please see below the specific responses to the individual comments.

1. The GHI dataset is compiled from multiple datasets. It will be good to add a flowchart showing the steps of how different datasets are processed and the relationships between GHI and these datasets. Figure 4 and Table 5 serve such purpose but the usage and relations of different datasets are not clear.

In order to address this comment, Figure 4 was revised to include a chart illustrating the input datasets, the data preparation methods, the specific quality control measures associated with each method, and the final outputs from this study. The beginning of Section 3 (“Methodology”) was revised accordingly.

2. The base maps in figures are indistinct (e.g., figures 3 and 5). It seems that the author used screenshots of Google Maps and enlarged the images without properly dealing with the resolution. The letters of street names and landmark names are blurry. Please make sure to use high-resolution images for base map when making these figures.

In the original manuscript screenshots from Google Maps and OpenStreetMap were used within several figures to provide spatial reference. The displayed screenshots turned out to be blurry in print. In the revised manuscript, such screenshots were replaced with high resolution original graphics (original tiles from the above map vendors). Figures 3, 5, 6, and figures A3 and A5 in the appendix were revised and the author made sure that the printed version of these figures are not blurry.

3. In Section 3.5, the method of how hydrometeorological variables are compiled is not explicitly mentioned. Are these variables compiled directly from available datasets as listed in Table 5? Are there any newly generated time series estimates from existing data by, such as, model outputs or statistical calculations?

The time series estimates within GHI are based on existing observations (streamflow) and grid-based products (e.g., IMD and ERA5-Land). Using the newly created catchment boundaries, grid-based products were aggregated to the catchment scale using an area-weighted procedure. The time series of such aggregated values is one of the outputs from GHI. Section 3.5 was revised to make the reader better understand the output created by GHI. Moreover, a new graphic has been added in the Appendix (Figure A6) to describe the creation of area-weighted hydrometeorological data.

4. Some of the citation formats are incorrect. I see many places having the citation ends with double right parenthesis. For example, in Line 43, “River Discharge Data Set (RivDIS, Vorosmarty et al. (1998))”. It can be written as “River Discharge Data Set (RivDIS) (Vorosmarty et al., 1998)”. There are many other places with this issue (i.e., Lines 44, 54, 59, 115, 190, 242, 248, etc.). Please check throughout the manuscript.

As suggested by the reviewer, the double right parentheses issue was corrected in the revised manuscript, and relevant citations were revised accordingly. The revised manuscript was checked to make sure there are no occurrences of right or left double parentheses.