General Comment

Yin et al generated a comprehensive hydrometeorological dataset at global scales. Compared to existing dataset, GSHA includes more variables and corresponding uncertainty estimates. It represents a significant contribution for large sample hydrology datasets and is very useful for data-driven hydrological applications, and calibration/validation of large-scale hydrological models and Earth system models. It is suitable to be published at Earth System Science Data. However, I recommend a major revision is needed before it is ready to be published. Please find my detailed comments in the following.

Major Comments #1

Line 163 – Line 170: The authors should clarify how they aggregated the daily streamflow to annual streamflow. There could be missing data for the streamflow in any year, and the missing days and the number of the missing days are not the same in different years. How the authors addressed the data gaps in the daily streamflow? For example, is there a criterion for the number of available days in a year that used to filter "good" years?

In addition, I think monthly streamflow indices are more useful for modelers to calibrate and validate models. For example, previous studies have used the monthly time series from GSIM to calibrate large scale hydrological models and Earth system models.

Major Comments #2

There is a lack of validation of watershed delineation. The watershed delineation could be one of the most important characteristics of GSHA, as many other variables were extracted based on watershed boundary. I think the flow directions may be carefully validated in previous study, but it is important to validate the delineated watershed boundary. For example, most gauges reported watershed boundary or drainage area, which can be used as benchmark.

Major Comments #3

One of the novelties of GSHA comparing to existing large sample hydrology datasets is GSHA provide the uncertainty analysis for the selected variables. But I think current description of uncertainty estimate is not clear, and the method is not comprehensive. Specifically,

Line 354-Line358: I don't think Eq (1) represents the uncertainty of the meteorological variables. As X_{max} and X_{min} represent the maximum and minimum values of the extracted variables from each individual dataset, $X_{max} - X_{min}$ is more linked to the natural variability instead of uncertainty of that dataset. For the example of temperature, if we have a dataset give us $X_{max} = 35℃$, $X_{min} = -5℃$, and $\bar{X} = 10℃$. Is the uncertainty of this dataset being $\frac{35-(-5)}{10} \times 100\% = 400\%$? And please further explain why the range of Eq (1) is between 0 and 200%.

However, based on the results in Figure 6, I think X_{max} and X_{min} are derived from all the datasets? I think the authors should further clarify the definition of uncertainty. In addition, $X_{max} - X_{min}$ cannot capture the uncertainty in the temporal variability. It is possible for two

datasets capture exact the same X_{max} and X_{min}, but have different distribution. Thus, $X_{max} - X_{min}$ is not a good metric for analyzing the uncertainties from different datasets. I suggest the authors to include more metrics in the uncertainty analysis.

Specific Comments

Line 160: upstream drainage ~~basin~~ area

The section numbers from 3.2 to 3.7 were wrong.

Line 212: Need to define "shorter record length" explicitly.

Line 216: CHP is defined in Table 3. But I think it is better to give the full name in the main text as well.

Line 318-Line 325: How you match the dams in GeoDAR to GSHA? Is it possible for a watershed to have several dams? How about the watershed that doesn't have a dam from GeoDAR?

Table 4: I think MSWEP is at spatial resolution of $0.1° \times 0.1°$. Please double check.

Figure 5a, b, and c: Are the X and Y axis normalized? I suggest the authors to plot the original data (e.g., in m^3/s) to demonstrate that no system errors were introduced during the processing of GSHA. In addition, the authors should explain why the comparison of some watersheds are very off from the 1:1 line. In my understanding, both GSIM and GSHA were derived from gauge observations for the streamflow indices. Therefore, same gauge measurements should be used at the same watershed in both GSIM and GSHA. Is the significant difference caused by (1) different gauges were used, or (2) different method was applied to address the data gaps in the gauge measurements (see my Major Comments #1), etc. Overall, I think it is useful for the authors to further explain the significant discrepancies in those gauges.

Figure 7: I don't think the decline of uncertainty as the watershed area increases is obvious for longwave radiation.

Line 517 – Line 529: The analysis of runoff coefficient and its changing trend in the past few decades is very interesting and is very critical for us to understand response of hydrological cycle to global warming. Such analysis with observed streamflow is more convincing than model simulations, which can be highly biased. I believe there exist some other studies focusing on this topic, such as runoff trend in this historical period. I wonder if the authors can give more discussion for this analysis and include more references.