General Comment:

Comments: I appreciate the authors' efforts for addressing my previous comments. However, I cannot recommend the publication in its current form because I found there are still some further validations and clarifications needed. Please find my comments in the following.

Thank you for your comment. We aggregated the daily streamflow by calculating their annual indices, such as annual mean, maximum, percentiles, as well as temporal characteristics such as maximum flood occurrence date, duration of high and low flow events. Therefore, we describe our data as streamflow characteristics instead of annual streamflow. **The dataset includes a "valid observation days" field**, which describes the number of days with available daily streamflow in the corresponding year, **as well as a "Q=0 days" field** representing the number of days with runoff measurement equal to 0. The data were not filtered or selected based on any criterion set by the authors, because **we would like to let the users decide how many available or non-zero measurements define a "good" year to them based on their research purposes and scales.** To make this clearer, we modified the sentence in lines 170-171 to "We also include numbers of zero observations and valid samples to allow flexible data screening by the users."

Comments: Thank you for addressing my comments. However, I think the annual indices with number of available daily/sub-daily data should be reported as 'NaN' or removed. If we have both Year A and Year B has the same number of daily data, e.g., ~180 days, to derive the annual indices. But Year A is available from Jan to Jun, while Year B is available from Jul to Dec. Such inconsistency will result in bias for calibrating/validating model simulation because they represent streamflow characteristic from different seasons. Therefore, although the users can decide if a year is "good" or not, they will not know if all the good years are consistent in the time period.

Thanks for your comment. Since we found out we do not have access to officially reported areas of all watersheds from agency websites, we validated our watershed areas for Australian Bureau of Meteorology 2022 (BOM), Canada National Water Data Archive 2022 (HYDAT), and The Global Runoff Data Centre 2022 (GRDC). The validation results are plotted in **Figure R1**.

Comments: I appreciate the authors' efforts to validate the watershed delineation based on my comment. I believe the contributing area should be reported for each gauge, at least from USGS. For example, the author can find the drainage area at this USGS gauge: https://waterdata.usgs.gov/monitoring-location/07374000/#parameterCode=00065&period=P7D&showMedian=true. I understand that GRDC gauge coordinates may be highly uncertain, so USGS gauges could be a good benchmark, which is at higher quality.

Thanks for the comment. The uncertainty we calculate represents the discrepancy between long-term means of the datasets, instead of the differences of each value in the time series. The

$X_{!"\#}$ and $X_{!\$\%}$ values are the maximum and minimum values in the **dataset ensembles (in our dataset two to three members included), rather than the max and min values in the temporal series**. We use this estimate to represent uncertainty of the mean value. Therefore, the distributions and variances inside each dataset are not considered. We understand that uncertainty should be represented by a range around the true value of the variable, but we do not know the true values of each variable at each particular date, and daily estimates from the datasets can be very biased. Therefore, we believe uncertainty range represented by discrepancy of the long term mean can be more meaningful compared to a time series of daily differences. 200% uncertainty occurs when one dataset $X_{\min} = 0$ and $X_{\max} > 0$. As we use K as temperature unit, there will be no negative value in the data.

Comments: Do you mean for the case $X_{\min} = 0$ and $X_{\max} > 0$, the $\bar{X} = \frac{X_{max}}{2}$? Do you assume the variable X varies linearly from $X_{\min}$ to $X_{\max}$?

We matched the gauges by their latitudes and longitudes, each point should represent the pair of the same gauge. However, the location matching might confuse a small proportion of very close gauges. Therefore, it is possible that the different gauges used cause deviations of validation results, and **we think locational error is the most significant factor causing the problem**. However, currently we do not have a proper method to find out which gauge pairs are wrong based on ids and locations, thus we plotted all pairs in the validation figures. For data selection, GSIM suggested that "Given that data quality requirements can vary substantially, it will remain the work of individual users to establish selection criteria for each study, thereby finding a trade-off between data quantity (number of gauges) and data quality (record length, missing periods)" (Gudmundsson et al., 2018), which is consistent with our decision not to filter the observations as mentioned in the reply of Major Comments #1. However, according to the time step in the GSIM file, the first time step and last time step are usually 31st Dec., apart from some missing values, while we did not process our data that way. This might cause some discrepancies, but with monthly indices provided, we believe more accurate analysis can be carried out. We added these two reasons in the 4.1 Technical Validation section to inform the readers of these causes of differences.

Comments: Except the location and id, the contributing area can be used as the third criterion for paring the gauge in both GSIM and GSHA. Specifically, if the contributing area are not the same, there is a high probability that not the same gauge is used in GSIM and GSHA for comparison.

In addition, I don't understand how the time step impact the annual streamflow indices, e.g., p90 that is reported in Figure R2. Should the estimate of the annual streamflow indices have based on 365 or 366 daily streamflow (if the data is available for the whole year)?