



## Comment on **essd-2021-71**

Anonymous Referee #1

---

Referee comment on "CCAM: China Catchment Attributes and Meteorology dataset" by Zhen Hao et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-71-RC1>, 2021

---

Large Sample hydrological data sets are a basis for data-driven hydrological analysis. The authors compiled a large-scale catchment attributes and meteorological time series dataset of contiguous China. The creation of such datasets is time- and resource-consuming, therefore the dataset can be of great benefit for hydrological studies in that region.

The authors announce that they will also publish the source code for compiling the dataset. Since the compiled dataset should also be updated when the underlying base datasets are updated, the publication of the program code would be of great benefit to the scientific community. However, at this time, the source code is not yet available. If possible, however, the source code should already be published along with the data, which would add significant value to the paper.

Many of the normalized discharge time series seem to have gaps, which reduces their usability for further analysis. It would be useful for the reader if the mean length of complete discharge time series could be quantified, or the gaps could be addressed in general; in the moment, from the text the reader assumes that the time series have a regular 7 day resolution without gaps. Also, the reason for the normalization should be addressed; as from the normalized time series, e.g. no mean annual discharge can be derived; and this parameter is also not given in the metadata of the catchments.

Specific Comments:

line 28: Either name the main processes of the hydrological cycle, or focus on specific processes in terrestrial catchments. Careful with the right terms: rainfall instead of raindrop

line 33: "[...] it is possible for the hydrological model to learn [...]": this is only true for a spacial type of models (machine-learning models)

line 37: citation style: instead of "(Kratzert, Klotz et al. 2019) shows" it should be "Kratzert, Klotz et al. (2019) shows". also applies for similar citations later in the text

Figures 3,4,5: as it is already written in a community comment, the text in the figures is

hardly readable.

Appendix B: Assumption for the use of Pearsons Correlation Coefficient is normal distribution; and linear relationship is assumed. is this really always the case here?