

Earth Syst. Sci. Data Discuss., author comment AC6
<https://doi.org/10.5194/essd-2021-71-AC6>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.



Point-by-point reply to Referee 1

Zhen Hao et al.

Author 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-AC6>, 2021

To make the response more structured, we have organized the response.

Point-by-point reply to Referee 1

Date: 2021-08-16

Manuscript Number: ESSD-2021-71

Title of Article: Catchment attributes and meteorology for large sample study in
contiguous China

Name of the Author: Zhen Hao

Email Address of the Author: zhen.hao18@alumni.imperial.ac.uk

RC: 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.

AC: The code is now available at: <https://github.com/haozhen315/CCAM-China-Catchment-Attributes-and-Meteorology-dataset>

To further improve the usability and influence of the code, we have reformulated the code such that the user can generate a basin's characteristics with just a "one-click" when the required source data are prepared, which will significantly improve the accessibility of the catchment attributes.

In the past few days, we have made extensive efforts to reorganize our code. Combined with the data set that has been released, we are aiming to achieve two goals:

(1) Researchers can quickly obtain catchment attributes and meteorological time series of

the local catchment from our data set.

(2) If the local catchment has a custom boundary, using our code can calculate the catchment attributes and meteorological time series quickly based on the given boundary. Our code currently supports one-click generation of all static attributes, as long as the required source data has been prepared according to the instruction, and the generation of the meteorological time series can also be quite effortless.

RC: 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

AC: The original streamflow observations are not continuous. The average record length is 11.3 years. In the revised version of the dataset, we separately provide continuous streamflow observations with an average record length of 8.3 years.

RC: 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.

AC: Due to the strict redistribution policy of streamflow data, we are afraid not to be able to release the original streamflow data and the mean discharge. We must ensure that the source data is not released, but we want the released data to be useful, so we present the current solution. The current data can be used in such a situation: when it is desirable to verify the generalization ability of a machine learning model on a global scale, HydroMLYR (new name) can support the verification of the performance in the Yellow River Basin.

RC: 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

AC: This sentence has been reformulated as "Rainfall, interception, evaporation and evapotranspiration, groundwater flow, subsurface flow and surface runoff are the main components of the terrestrial hydrological cycle."

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

AC: This sentence has been reformulated as "However, by examining a large sample of catchments, it is possible for a data-driven model to learn the similarities and differences of hydrological behaviours across catchments (Kratzert, Klotz et al. 2019)."

RC: 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

AC: "(Kratzert, Klotz et al. 2019) shows" should be "Kratzert, Klotz et al. (2019) shows". Other similar issues have also been fixed.

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

AC: We have fixed it by redrawing these figures.

RC: 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?

AC: Although the Pearson's correlation can only provide a complete description of the association when both the two variables are standard, we think the most doubtful part of using it is that it assumes a linear relationship (a change in one variable will cause a proportional change to the other). Because there are too many variables, we can't plot the scatterplot one by one to check whether the relationships are linear. We think Spearman or Kendall's Tau may be more suitable in this case due to its wider application range. Even if the relationship between the two variables is linear, Spearman or Kendall can also return a very close result to Pearson. However, if the relationship between two variables is only monotonic, Pearson will have information loss.

Best,

Zhen Hao