Earth Syst. Sci. Data Discuss., author comment AC7
https://doi.org/10.5194/essd-2021-71-AC7, 2021

Point-by-point reply to Referee 2

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-AC7, 2021

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

Point-by-point reply to Referee 2

Date: 2021-08-16

Manuscript Number: ESSD-2021-71

Title of Article: CCAM: China Catchment Attributes and Meteorology dataset

Name of the Author: Zhen Hao

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

RC: In the abstract, the authors state their "dataset provides numerous opportunities for comparative hydrological research, such as examining the difference in hydrological behaviours across different catchments and building general rainfall-runoff modelling frameworks for many catchments instead of limited to a few". My concern is that the scope of this dataset (in its current form) might be more limited, because of the following restrictions imposed to the streamflow data: […]

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 streamflow. 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. The abstract has also been rewritten to avoid those claims.

RC: I understand that releasing the true streamflow time series is challenging, but some

decisions made by the authors are puzzling. For instance, “for confidentiality, the names of these basins have not been announced”, but shapefiles are provided and give (presumably) the exact location of the catchments. Likewise, the mean streamflow can be inferred quite readily from catchment descriptors, which makes me feel that the normalisation of the timeseries is unnecessary.

AC: In fact, the location of hydrological observation stations can be observed through remote sensing satellite images and then combined with HydroSHEDS's River network to determine their location. Then the boundaries of the basin can be determined based on the publicly available DEM, but we cannot release the names of these hydrological observation stations; these are sensitive information. In addition, indicators that can be used to calculate the mean discharge have not been released.

RC: Hence, I recommend that the authors do not use the name CAMELS, as all the CAMELS datasets provide daily streamflow timeseries for their hundreds of catchments, which many in the community see as their most important characteristic. There are many alternative naming options, including exotic animal names (as illustrated by the recent LamaH dataset - https://essd.copernicus.org/preprints/essd-2021-72/).

AC: The paper title sounds like the title of a CAMELS dataset, and the use of "Normal-Camels-YR" might be misleading. We suggest using CCAM to stand for "China Catchment Attributes and Meteorology dataset" and HydroMLYR to stand for "Hydrology dataset for Machine Learning of the Yellow River Basin." The new names may avoid readers' wrong expectation of the data set and more clearly indicates the purpose.

RC: Data availability/reproducibility: The abstract mentions that “complement code for generating the dataset will be open-sourced such that the user can generate meteorological series and catchment attributes for any watershed within contiguous China”, yet the conclusion makes it clear that the forcing dataset SURF_CLI_CHN_MUL_DAY is only freely available for Chinese researchers (L444). This is a non-negligible constraint. Furthermore, I don’t see a paper documenting the SURF_CLI_CHN_MUL_DAY dataset, and the link provided (http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_DAY.html) leads to page in Chinese.

AC: We are sorry that we made a false statement about the facts at the beginning. We found on the registration page that foreign researchers can also register, but the interface is still in Chinese, which is out of our control. The SURF_CLI_CHN_MUL_DAY data was issued by the National Meteorological Information Center of the China Meteorological Administration (NMIC/CMA). The data is quality controlled, and it is widely used in research in China. However, it does not have a related paper.

Best,

Zhen Hao