Reviewer 2 – Alexander Dolich

Thank you for your work in compiling CAMELS-GB v2 and this manuscript. The data is very nicely formatted and the addition of hourly hydro-meteorological timeseries data is extremely valuable to expand the field of large-sample hydrology to sub-daily resolutions. The addition of groundwater data and the extensive quality control of hourly streamflow data is also very valuable and outstanding among other CAMELS datasets. The extension of the daily timeseries length and the expansion of the catchment attributes, especially the changing land cover attributes, are also of great value and show the author's motivation and dedication to maintain and update CAMELS-GB.

The manuscript gives a very detailed and thorough description of the dataset while highlighting its importance and also describing its limitations. I recommend accepting the manuscript after minor revisions.

I think the accessibility of the dataset could be enhanced by offering the option to easily download a compressed file containing the entire dataset. A more detailed explanation of this as well as minor comments and technical corrections can be found in the attached PDF file.

Many thanks Alexander for the positive assessment of the paper and the helpful feedback. Please see our detailed responses below.

The advantage of having multiple meteorological data sources could be highlighted better. There are more advantages for sure (e.g. the possibility to fill gaps in the hourly radar data with the interpolated station data), but especially deep learning methods can make use of multiple forcing data sources at once (see Kratzert et al., 2021), which could be especially useful when mixing radar and station data as these data sources have different strengths.

Thanks, we have added a new section 3.3 detailing guidance on the use of the hydro-meteorological datasets and added this nice advantage here.

"CAMELS-GB v2 contains hydro-meteorological timeseries from different products, over different timescales and at different resolutions. Multiple products enable users to reflect some of the uncertainty in these data and make use of multiple forcing data at once (e.g. Kratzert et al, 2021). To help users to select the most appropriate dataset for their study, we recommend: ..."

L29-L30: The DOI URL is not working (Error: DOI not found), maybe this needs to be updated? → This is a formatting error, when I click on the link, the URL ends before the line break (the DOI URL in the Data availability section works)

Thanks, we will ensure the link works in the final manuscript.

L52: You could add that the meteorological timeseries are spatially aggregated to the catchment boundaries, I think this would be more clear then “catchment timeseries”.

Thanks, this has been revised to:

“hydro-meteorological time series (spatially aggregated to catchment boundaries)”

L84: Is this the correct citation of the exactextract package? Also in the References section (L514). I think extactextract was not published yet, so you would have to refer to it like a website? Maybe refer to one of the Github releauses or to a version on pypi?

Many thanks for spotting this. We have added the website link on pypi (https://pypi.org/project/exactextract/0.2.0/) to the version of the exactextract package we used to the references.

The updated citation is:

“Baston, D.: exactextract, version 0.2.0, [code], https://pypi.org/project/exactextract/0.2.0/ (last access: 22 January 2026), 2025.”

We also added the link to the Code availability section.

“The exactextract Python package (https://pypi.org/project/exactextract/0.2.0/, last access: 22 January 2026, Baston, 2025) is used to extract catchment average data from gridded datasets based on the catchment boundary polygons described in Section 2.”

L85: suggestion: “used to extract and spatially aggregate"

Thanks, we have revised as suggested.

L86: I would change “by” to “while” here

Thanks, we have revised as suggested.

L154: Label c) is missing in Figure 2

Thanks, we have added the label.

L168: I think it would be good to add information about the timezone of your data. Maybe this is not strictly necessary for Great Britain as it is in GMT / UTC+0, but once there are more CAMELS datasets in sub-daily resolution, the time zone information becomes very important for interoperability. So I would also add that information for CAMELS-GB, just to be clear about it. I guess that you also do not change between summer and winter time, that information should also be added.

Thanks, we added a sentence in the revised paper to clarify the time zone (UTC) for the timeseries data. In addition, all references to specific times (e.g., 09:00) throughout the text, as well as in Tables 3 and 5, have been updated to explicitly indicate UTC (e.g., 09:00 UTC).

"Hourly hydro-meteorological time series of rainfall and river flow are provided for the 671 catchments from 1st October 1990 09:00 UTC to 1st October 2022 08:00 UTC (Table 3). All timeseries data in CAMELS-GB are reported in UTC with no daylight saving time adjustment."

L258: "Figure 4" is bold

Thanks, we have revised this.

Figure 4: The text size of the legend is very small and should be increased.

Thanks, we have done this.

L276-L279: The length of the hydro-meteorological data was also extended in v2, so the hydrological and meteorological attributes were also calculated for that longer period, right? This could be added here.

A sentence about why these attributes were calculated for the daily data only and not the hourly data could also be added (and what this implies / do you expect major differences when calculating these attributes based on hourly or daily data?). Could also be added to section 5.2 / 5.3.

Thanks, we have clarified the time period that the climatic indices are calculated over and addressed the lack of attributes calculated from the hourly data.

"The climatic indices in CAMELS-GB v2 are derived over the time period 1970-2022 using the HadUK-Grid catchment daily rainfall…"

"Climatic indices calculated from the hourly data are not provided as the long term and seasonal climate indices are very similar regardless of the temporal resolution of the data. While the short-term (i.e. frequency, duration and timing of high and low precipitation events) climatic indices would be different when calculated using hourly data for some catchments, analysis of the impacts of temporal resolution on catchment attributes (including careful consideration of choice of indices) is beyond the scope of this paper."

L324: The decrease in urban land cover from 2021 to 2022 in Figure S9 is really suspicious. Do you think this is an error in the data? Maybe you could reach out to the dataset's authors for clarification and if you get a timely response you could add the explanation to the manuscript.

Thanks, we did reach out to the dataset author prior to the paper being submitted and they listed a number of reasons for the decrease in urban land cover which can be found in the paper in Section 5.4.

L358-L372: I think the comparison of the maximum gauged flow and the maximum flows in the timeseries is very important and can give a good estimate of the uncertainties in high flows. But maybe you could add a sentence about why this is important for uncertainty estimation somewhere at the start of the section. It is about the rating curve and at which point extrapolating beyond the measured values of the rating curve begins, right?

Thanks, we have added a sentence about why this is important for uncertainty estimation.

"These new attributes contain important information about the uncertainty at high flows as users can identify where (and how often) the daily/hourly flow timeseries exceed the highest manual measurement of flow taken at a gauging station and therefore are based on extrapolation of the rating curve which can result in significant uncertainties (Juston et al, 2014)."

Table 1: "No longer providing wind speed, humidity, short-wave radiation and long-wave radiation as these were rarely used." → it' s okay if you made this decision, but in my opinion this information could be very valuable in certain tasks, so I do not really understand why you do not provide this information anymore, if it is still available.

When producing CAMELS-GB v2, we decided to prioritise providing climate timeseries (rainfall and potential evapotranspiration) from multiple datasets (i.e. CHESS and HadUK-Grid). While we agree that climate timeseries of wind speed, humidity and short/long-wave radiation could be useful, these climate variables are only available at monthly timescales from HadUK-Grid and therefore wouldn't be consistent between datasets. These variables from the CHESS datasets are already provided as part of CAMELS-GB v1.

Furthermore, the provision of multiple estimates of potential evapotranspiration in CAMELS-GB v2 means wind speed and humidity are less important as they would primarily be used to estimate PET for hydrological applications.

S4: Labels a), b), c) are missing

Thanks, we have updated for Figures S4 and S5.

Technical corrections

Thanks – we have revised the paper to address all the technical corrections as suggested.

L24: “1970- 2022” → missing white space, should be “1970 - 2022”

L45: “finable” → “findable”

L61: the second “are” is incorrect here

L120 and L127: time spans are not consistent, e.g. “1890 – 2019” and “1961-2019”,also the case in other Lines

L263: You suddenly use “CAMELS-GB-v2” in this chapter, to be consistent this should be changed to “CAMELS-GB v2” (also in the caption of Figure 4)

L264: “aggregated to monthly” → “aggregated to monthly values”

L300: “low date” should be “flow date”

L289, L327, L340, L355, L357: citation format (remove “,” before year) “Coxon et al., (2020)” → “ Coxon et al. (2020)”, also in Table 5: “Harrigan et al., (2018)” → “Harrigan et al. (2018)” and “Salwey et al., (2023)” → “Salwey et al. (2023)”

L392: “is based the same” → “is based on the same”

L447: “dataset available” → “dataset is available”

Dataset

I opened all of the CSV files in CAMELS-GB v2 in Python with pandas and the catchment boundary shapefile with geopandas and did some basic checks (e.g. all IDs are present everywhere). I also visually inspected all attribute CSV files and some of the timeseries CSV files. During these checks I did not encounter any issues, the data is very nicely formatted and, together with the data description and the paper, very easy to work with, great work!

Thanks – we spent a lot of time formatting and checking the dataset so appreciate this comment!

I think that the accessibility of the dataset could be enhanced. When I click on “Download the data”, I end up at the eidchub datastore where I can download individual files manually by clicking on them. This is only helpful if users are only interested in individual stations or attributes, which would not be very common. Usually, users of CAMELS-GB want to

download the entire dataset. To achieve this, the user has to use wget (bulk download), which can already be challenging for users who are not very familiar with programming and/or Linux and HTTPS servers. I had to do some research on Google to get all the correct wget options to download the entire datasets and exclude the index.html files and I would consider myself quite experienced with things like this. If interested users cannot manage to easily download the dataset, they will just turn away. So I think having the option to just download the entire CAMELS-GB v2 dataset e.g. as a zip file would be very helpful and a lot more user friendly. The compressed zip file of CAMELS-GB v2 is also only 1.8 GB. Maybe I missed something and the option to easily download the entire dataset in one go already exists, in this case it should be directly visible on the data centre page.

We appreciate that different users will have different expectations for data access. We have received feedback on this dataset and the previous CAMELS-GB release that many users just want data for one or a few stations (and therefore would not want to download the entire dataset). We need to balance the needs of these users with users that do want to download the entire dataset. Currently, it is not possible to enable both these options in the data centre, although they are considering future developments to enable this.

Our expectation is that most users wanting to analyse the entire dataset will have programming skills and an example wget statement is provided on the download page, as well as further guidance to help: https://eidc.ac.uk/help/getdata/downloadData#programmaticAccess. Furthermore, we are developing online notebook services to allow direct access to and analysis of CAMELS-GB and other datasets, which would not be possible with a large zip file.