Reviewer 3 – Hörður Bragi Helgason

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

Comments on the manuscript

1. Quality control for daily and hourly streamflow:

The manuscript states that both daily and sub-daily streamflow data are quality controlled by the source institutions prior to release, and that the hourly data originate from the quality-controlled UK-Flow15 dataset. At the same time, detailed quality control flags are only provided for the hourly data, while no equivalent flags or diagnostics are supplied for the daily streamflow. It would be helpful to clarify the rationale for this asymmetry.

I suggest clarifying how users should interpret the statement that no data were removed or modified by the quality control process for the hourly data, given that the data have already undergone quality control within UK-Flow15. Is the term "quality control" used differently for the daily and hourly products? For example screening or filtering at the daily scale versus the provision of diagnostic flags without filtering at the hourly scale.

The sources of the daily and hourly data are different and therefore you are right that the term 'quality control' is used differently for the two datasets.

Daily flows are quality controlled by the original measurement agencies and the UK National River Flow Archive in a process that removes data values that are perceived to be erroneous. Therefore the daily data is the best estimate available and considered suitable for analysis and modelling without reference to flags. The quality control process for the hourly data, applied in production of the UK-Flow15 dataset (Fileni et al, 2025; https://catalogue.ceh.ac.uk/documents/211710ac-f01b-4b52-807f-373babb1c368) acknowledges that some flow data may be incorrect, potentially substantially so in some cases, and addresses this by applying flags to alert users of data issues, rather than removing data. It is therefore important that users have access to these flags in order to make their own judgements about which data are suitable for use for a specific application.

We have now clarified the quality control process in Section 3.1.2 for the daily data:

"The streamflow data undergo quality control by the measuring authorities and the UK NRFA before being uploaded to the NRFA site. This quality control process removes flow values that are perceived to be erroneous, while retaining as complete a series as possible through the use of a range of infilling techniques (Dixon et al, 2012)."

And modified Section 3.2.2 to more clearly highlight the differences in the quality control process:

"The hourly flow and level timeseries are also provided with quality control flags. The hourly flow and level data have been quality controlled using both visual/manual inspection and automated quality control, including novel quality assessment techniques to assess the plausibility of extreme flow events (Fileni et al, 2026). This quality control process is different to the quality control process applied to the daily flow data. Erroneous daily flow data are removed as part of the quality control process by measuring authorities and the UK NRFA, thus the daily flow data are the best estimate available and considered suitable for analysis and modelling without reference to flags. However, the quality control process for the hourly flow and level data acknowledges that some of these data may be incorrect and addresses this by applying flags to alert users of data issues, rather than removing data. It is therefore important that users use the flag-based system on the hourly flow and level data to identify, remove, or interpolate potentially problematic data as per their study requirements."

In the manuscript, users are strongly encouraged to use the diagnostic flags for the hourly data to decide how to treat potentially problematic observations. While this flexibility is valuable, it may present a barrier for some users. I therefore suggest considering the provision of an additional, ready-to-use pre-filtered hourly streamflow product based on a simple and clearly documented flag selection. This would improve accessibility for users who prefer a conservative default dataset, while still retaining the full flagged dataset for more advanced or customized analyses.

We understand there is some complexity around this that some users might find difficult to navigate. While it is challenging to define a "one size fits all" approach when it comes to suitable data quality for an application, we can make it simpler for users to start using a set of reliable stations. There is no clear default list of stations suitable for all purposes, and removing anything considered unsuitable could dramatically reduce the amount of data available. We therefore have included additional information that will allow users to select a subset of stations to use for different applications, rather than a subset of data values within stations. This has the benefit of retaining complete series for these stations.

We have revised the dataset to include new categorical information on (1) the quality of high flow data for each station, and (2) whether a station has significant issues in their historic record alongside information about these issues. This categorical information is sourced from the National River Flow Archive and the following station attributes : Indicative suitability for QMED" and "Indicative suitability for Pooling" – these have definitions defined here: https://nrfa.ceh.ac.uk/data/about-data/peak-flow-data/indicative-suitabilities, while the methodologies used for identifying issues in the historical record are described in Fileni et al (2026; https://doi.org/10.5194/essd-2026-152).

We have revised Section 3.2, 3.3 and 5.7 of the paper to include guidance for users on how to use this new categorical information as follows:

"Users seeking a simpler way to select a subset of stations that are suitable for a range of applications are recommended to use the "station quality" categorical information contained within the hydrometry attribute file (see Section 5.7)."

"Finally, users should carefully consider the quality of the hourly flow and level timeseries before use in studies. We recommend using the flag-based system on the hourly flow and level timeseries to identify, remove, or interpolate potentially problematic data as per their study requirements to ensure that only good-quality data are utilised (see Fileni et al, 2026 for a more detailed assessment of the quality control process). Users can also select stations to use based on the station quality information (see Section 5.7). For low flow studies, we would suggest the daily flow data are best to use as they provide a good representation of slowly changing low flows and has been given additional quality control by the UK National River Flow Archive. For mean flow analyses, we recommend removing any stations where major issues have been identified. For high flow analyses, we recommend only using stations that have good quality high-flow data and for flood event quantification and flood frequency analysis we recommend only using stations with the highest quality flow data (see Section 5.7)."

"New hydrometry attributes were added describing the quality of the flow data collected at each station. This includes categorical information from the UK NRFA to help understand the quality of high flow measurements and qualitative information on any issues found during analysis of the 15minute flow data (Fileni et al, 2026). Stations that are suitable for QMED (station_quality_qmed) can be used for high-flow estimation; however, measurements are more likely to deviate from the true values, particularly for the largest flows, while stations that are suitable for pooling (station_quality_pooling) have the highest quality high-flow measurements and are recommended for flood event quantification and flood frequency analysis."

2. Daily mean temperature definition (Line 134):

The description of how daily mean temperature is calculated from HadUK-Grid data is potentially confusing. A brief explanation of the rationale for this convention, related to the timing of minimum temperature observations and the diurnal cycle, would help readers understand why this approach is used instead of averaging same-day maxima and minima.

Sorry, we did not explain this clearly. Based on the Hydro-PE dataset paper (Robinson et al., 2023), the HadUK-Grid daily climate variables are defined such that the maximum air temperature (tasmax, Tmax, °C) is measured between 09:00 UTC on day D and 09:00 UTC on day D+1, while the minimum air temperature (tasmin, Tmin, °C) is measured between 09:00 UTC on day D−1 and 09:00 UTC on day D. Therefore, using tasmax on day D together with tasmin on day D+1 represents temperatures over the same 24-hour period, from 09:00 UTC on day D to 09:00 UTC on day D+1. This is also

how daily mean temperature is calculated in their paper and dataset, and we adopt the same approach.

We revised Section 3.1.1:

"In the HadUK-Grid daily climate variables, the maximum air temperature is measured between 09:00 UTC on day D and 09:00 UTC on day D+1, while the minimum air temperature is measured between 09:00 UTC on day D−1 and 09:00 UTC on day D (Robinson et al., 2023). Therefore, daily mean temperatures have been calculated by averaging maximum air temperature on day D and minimum air temperature on day D+1 to ensure both values represent the same 24-hour period for each day."

3. Interpretation of daily PET from Hydro-PE (Line 151):

The manuscript notes that several meteorological variables in the Hydro-PE HadUK-Grid dataset are temporally downscaled from monthly to daily resolution using smooth interpolation. While this approach is understandable for achieving long temporal coverage, it implies that daily PET and PETI values do not represent true day-to-day variability in all controlling variables. I suggest explicitly highlighting this limitation and clarifying that the Hydro-PE daily PET is most appropriate for seasonal to long-term analyses rather than event-scale or short-term applications.

Thanks, we have added a new Section 3.3 of guidance for users on dataset selection in the paper and we included this as one of the suggestions.

"Many of the variables used to derive Hydro-PE daily PET and PETI were temporally downscaled from monthly to daily resolution. Therefore, these products may be more suitable for seasonal to long-term analyses, and less appropriate for event-scale or short-term applications."

4. Groundwater well coverage:

The inclusion of groundwater level time series for 55 boreholes is a valuable new component of CAMELS-GB v2. At present, wells are restricted to those located within CAMELS-GB catchments. I suggest considering the inclusion of additional groundwater wells, if available, even if they do not fall strictly within CAMELS-GB catchment boundaries. Groundwater level time series can be highly valuable for hydrological studies even when not directly associated with a specific CAMELS-GB catchment.

We thank the reviewer for this suggestion. We have considered the possibility of a wider inclusion of wells that are part of the National Groundwater Level Archive (NGLA: National Groundwater Level Archive | British Geological Survey (BGS)**), but have decided that for the CAMELS-GB v2 data release we are best restricting the groundwater timeseries to boreholes within the CAMELS-GB catchments. The key spatial unit of the CAMELS-GB v2 dataset is the surface water catchment, as defined by the surface topography. Deciding which wells from the NGLA to include/exclude**

from outside the CAMELS-GB v2 catchments would require defining groundwater basins and conceptualising how these relate to the surface water catchments. This would be a complex undertaking and is beyond the scope of CAMELS-GB v2. The NGLA, of which the CAMELS-GB v2 wells form a subset, is maintained by the British Geological Survey and data are available on request for the researcher/practitioner wishing to consider the role of groundwater in surface water catchment hydrology in more detail. CAMELS-GB v2 still provides a step forward building on previous work (e.g. CAMELS-GB) with the inclusion of groundwater level measurements.

5. Interpretation and flagging of land cover change data:
In Section 5.4, the manuscript notes that most catchments show a decrease in urban land cover from 2021 to 2022 that is unlikely to be reflected in the real world. If this artefact is known a priori, perhaps removing these data for the year 2022 would make sense. If not, it would at least be helpful to clarify how users should interpret these data. Options could be either explicitly flagging the affected years or transitions in the dataset or metadata, or providing clearer guidance in the manuscript and documentation on how these land cover time series should and should not be used for change detection analyses.

Thanks for this comment. As noted in the paper, some of the land cover change between years is real, but there is also noise in the variability between years, which reflects uncertainty in how the land cover classifications are derived. Consequently, it is difficult to identify (or remove) affected years as we do not know what proportion of the change between years is real or simply noise. We feel it is important to provide all the years to reflect this variability/uncertainty, rather than removing data. We have added some further guidance into Section 5.4 for users:

"Consequently, we recommend users carefully consider how these data are incorporated in change detection analyses and consider pooling data for recent years together to reflect the uncertainty in the change signal."

6. File naming conventions:
The manuscript does not explicitly describe the file naming convention for the time series files. A short note explaining the format would improve the ease of use, for example:
• camels_gb_v2_hydromet_daily_timeseries_{gauge_id}_{start_date}-{end_date}.csv
• camels_gb_v2_groundwater_daily_timeseries_{well_id}_{start_date}-{end_date}.csv

Thanks, the file naming conventions are provided in the supporting documents on the Environmental Information Data Centre. We have now made this clear in the paper.

Minor technical corrections in the manuscript

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

1. Table 5, hydrometry attributes:
The description of structurefull_flow reads “flow at which the river begins to the wingwalls of a structure”, which is missing a verb. This likely should read “flow at which the river begins to overtop / reach the wingwalls of a structure”.

2. Line 37: Remove the extra “and” in “Global Streamflow and Indices and Metadata Archive”.

3. Line 45: “finable” should be “findable”.

4. Line 258: “Figure 4” is in bold, which is inconsistent with formatting elsewhere in the manuscript.

5. Benchmark catchments (Line 387):
The sentence
“All CAMELS-GB catchments are identified as either being part of this network or not to provide users with an indication of ‘near-natural’ catchments and suitable for studies where human impacts need to be minimal.”
is confusing. I suggest rephrasing to something like:
“All CAMELS-GB catchments are flagged according to whether they belong to the UK Benchmark Network, providing users with an indication of which catchments are relatively near-natural and therefore more suitable for studies requiring minimal human impact.”

6. Undefined abbreviation:
The abbreviation “CEH” is used throughout the manuscript but is never defined. Please spell it out at first use as UK Centre for Ecology and Hydrology.

7. Equation numbering: Please consider numbering equations (in Section 5.8.3.

8. Baseflow index description in Table 5:
The description of baseflow_index_ceh refers to “the Gustard et al. (1992) method described in Appendix A”. This appears to have been carried over from the CAMELS-GB v1 paper, as there is no Appendix A in the current manuscript.

Comments regarding the dataset

1. Data access and bulk download
For ease of use, consider providing an option for bulk download that does not require users to create an account on the Environmental Information Platform or use wget, if this is feasible within platform constraints.

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.

2. Hourly resampling and timestamp conventions:
The manuscript states that 15-minute streamflow data were aggregated to hourly values using a next-hour resampling convention. It is unclear which timestamp convention is used for hourly precipitation. I suggest explicitly documenting the timestamp convention used for all hourly variables and ensuring consistency between streamflow and precipitation. More generally, adopting start-of-hour timestamps for all hourly variables would be consistent with the definition of the daily data, where values for a given date represent the mean or sum over that calendar day rather than a shifted window. While timestamp conventions differ between regions and institutions, clear documentation and internal consistency are essential for correct interpretation and use of the data.

Thanks. We have added the timestamp convention to the start of Section 3.2 to make this clear.

"Rainfall on a given hour refers to the amount accumulated over the previous hour, whereas river level and flow are reported as hourly averages, representing the mean value over the previous hour."

3. Units in Table 5:
In Table 5, max_daily_flow and max_hourly_flow are listed with units of percent. These appear to be absolute flow values and should likely have units of mm day^{-1}, mm $hour^{-1}$, or m^3 s^{-1}. Please clarify and correct if needed.

Thanks for spotting this – we have revised this.

4. Consistency of flow units:
Flow-related attributes in Table 5 use a mix of mm day^{-1} and m^3 s^{-1}. While this reflects common practice, it may be worth considering whether greater standardisation would improve usability.

Thanks for this comment – as noted, this reflects common practice and so we would like to keep the mix of units for flow.

5. Runoff ratios greater than one:
Seven catchments have runoff ratios exceeding 1.0, notably:
• gauge_id 26006 with a runoff ratio of 3.01
• gauge_id 27038 with a runoff ratio of 2.74
I suggest explicitly identifying these catchments in the manuscript and briefly discussing potential explanations, to alert users.

Thanks, we have added the following text to Section 5.3:

"Seven catchments have a runoff ratio greater than one with gauge ID 26006 and 27038 with the highest runoff ratios of 3.01 and 2.74 respectively. Catchments with a runoff ratio greater than one are either groundwater dominated catchments where the groundwater catchment greatly exceeds the topographical catchment or are impacted by human activities with water being imported into the catchment for water supply or hydro-electricity."

6. Missing value conventions:
For the attribute high_prec_timing, Table 5 states that NaN is used when two seasons register the same number of events. However, the file camels_gb_v2_climatic_attributes.csv uses "NA" instead of "NaN". I suggest changing this to NaN for consistency.

Thanks for spotting this. We have revised this in the updated dataset.

7. Attributes with high fractions of missing values:
Some hydrometry attributes have more than 50 percent missing values, notably bankfull_flow (53 percent missing) and structurefull_flow (65 percent missing). A brief explanation of the reasons for this level of missingness would help users assess the reliability and appropriate use of these attributes.

This metadata is collated by the UK National River Flow Archive and made available alongside the daily flow data. It is produced by the original measuring agencies and has been accumulated over many years. Its availability is dependent on a number of factors: these metrics may not be relevant to many gauging stations, e.g. an open channel station has no structure and therefore no structure-full flow; metrics such as bank-full flow are often modelled, and this modelling may not have not been undertaken at many stations; there may not have been focused efforts within each agency to produce these values, which are often estimates. The information is considered the best available for this large selection of gauging stations, and despite the level of completeness, is expected to be useful for many types of studies, in

particular helping identify where flows significantly exceed those able to be accurately measured and where flood plain flows are likely to be occurring.

We have revised Section 5.7 to include:

"The channel characteristics were also re-extracted from the NRFA with changes for a small number of stations. These data are useful for flood analyses, but it is worth noting there is a large proportion of missing data for these attributes as they (1) may not be relevant for some gauging stations (e.g. an open channel station has no structure and therefore no structure-full flow), and (2) are often modelled and this modelling may not have not been undertaken at many stations."

8. Shapefiles for monitoring locations:
The dataset includes catchment boundary shapefiles but does not appear to include shapefiles of streamflow gauge locations or groundwater well locations. Providing point shapefiles for both river gauging stations and groundwater wells would be beneficial. While the relevant coordinates are available in tabular form, distributing these locations as shapefiles would be consistent with the provision of catchment boundaries and common practice in large-sample hydrology datasets. Including key attributes directly in the shapefiles (also the catchment boundaries shapefile) would further enhance usability.

Thanks for this suggestion. We believe that these shapefiles can be easily generated by users from the data provided and so have decided not to include these.