

Reviewer 1

Dear Alexandre,

Thank you very much for your positive and helpful review of UK-Flow15.

We have addressed all the **Minor Comments** point by point. Our responses are provided in the Response column of Table 1, alongside the corresponding changes made to the manuscript in the Changes column.

We also greatly appreciated the **Technical Corrections**, which were very helpful in identifying errors. These have been corrected, as detailed in Table 2.

Your comments on the **Dataset** were also very valuable. Where applicable, we contacted EIDC and made minor changes to the dataset, including corrections to typographical errors and column headers. Some of the broader suggestions have been considered for the development of future products, while the difficulties encountered when downloading the data have been reported to EIDC. Further details are provided in Table 3.

We thank you for the time and care you put into this review. Your comments have helped us improve both the dataset and the data descriptor manuscript.

Table 1: Minor Comments Addressed

Nbr	Comment	Response	Changes
1	title: maybe use e.g. “river flow dataset” or “streamflow dataset” instead of just “flow dataset”. At the moment, it is not clear from the title that this is a hydrological dataset.	We agree that this would improve clarity. Thank you for the suggestion.	“River flow” has been added to the title.
2	L14: “assessing hydrological change” → what do you mean here? For which type of changes do we need 15-minute data? Maybe include here that working with 15-minute data is especially necessary in	We have clarified that 15-min data is particularly useful for assessing changes in high flow characteristics.	“ <i>changes in high flow characteristics</i> ” replaced “ <i>hydrological change</i> ”

	small catchments, where dynamics are rapid.		
3	L18: state the exact number of gauging stations (instead of “over 1,300”)	We agree.	The exact number of stations, 1,369, has been added.
4	L49: I think you could cite a few more CAMELS datasets here (of your choice, I don’t usually see the one from Brazil cited as an example, so you could add one or two more)	A global south country was deliberately chosen to showcase the extent that these datasets are available and used. We have complemented with one additional CAMELS from the global north (Germany)	<i>Loritz et al.,2024</i> has been added as reference
5	L52: CAMELS-GB v2 (containing hourly time series for UK) is already available (preprint paper, the dataset is already accessible)	We clarify that UK-Flow15 data is the basis of the CAMELS-GBv2 hourly flow data. Still, we agree with the contradiction in the sentence.	We have replaced the sentence (L53-54) with: <i>However, most of these products are limited to coarser, daily temporal resolution, highlighting a clear need for the development of quality-controlled sub-daily flow datasets to support large-sample hydrological analyses.</i>
6	L66-L88: I think these two paragraphs could be merged into one / restructured, as some aspects are repeated (floods, high-intensity events, climate change), the opening of the two paragraphs are very similar (“the lack of ...).	We have made small changes to the second paragraph to ensure that the two paragraphs address distinct topics. P1: the importance of high-resolution data for hydrological research, and P2: how the limited availability of such data has affected flood-risk assessment and policy-making.	Changed the start of P81 to <i>Beyond hydrological research, the availability of sub-daily flow data would benefit flood risk policy and planning, where the absence of high-resolution data is often circumvented rather than addressed directly.</i>

7	L97-100: maybe add that also the comparability of results from different studies can be improved by offering a consistent data base (instead of each researcher collecting slightly different data again)	The benefits of a consistent high-resolution flow dataset are discussed in the paragraph beginning at line 101. To clarify that a consistent database also adds to comparability of results the word comparability has been added in L102.	L102: Reproducibility, <i>comparability</i> , and transparency remain persistent challenges in the scientific community
8	L133-135: the numbers of the hydrometric areas don't tell me anything here, I think they don't need to be mentioned here	We have retained the hydrometric-area numbers because they illustrate the structure of the NRFA station identifiers described in the preceding paragraph. The purpose of having them highlighted is linking to P125, where we present the structure of the stations <i>"The NRFA ID is a six-digit code: the first three digits corresponding to the hydrometric area"</i> with how it looks practically <i>"e.g. The Thames Catchment corresponds to hydrometric area 039"</i>.	A correction in the areas number: 39 ->039 1->001
9	L144: "the closer the fill is to real size the denser the region is" → I do not understand	We have rewritten the caption and clarified that the cartogram	(b) Spatial distribution: <i>cartogram representing station density (stations/km²) and the number of stations in each UK</i>

	this sentence, what density is shown here? number of stations / km ² ?	represents station density in stations per square kilometre.	<i>hydrometric area; fills that remain closer to the true size of the region indicate greater station density.</i>
10	L185-L188: What did you do in cases of lower resolution, e.g. hourly / 2 hours?	We have clarified how records available at coarser temporal resolutions were handled. Thank you!	<i>Added: For records available only at resolutions coarser than 15 minutes, the original values were flagged and replaced with NAs in the harmonised series.</i>
11	L254 (Fig. 2): When looking at the figure alone, I would get the following impressions, and I am not sure if they are correct	We agree with the reviewer.	The figure has been corrected
11.1	The timestamp 13:30:00 is duplicated, so you kept those? From Chapter 3.2 I think duplicated timestamps were removed	Duplicate timestamps were removed	
11.2	There is a 2 hour gap from 13:45 to 15:45, so you kept these “time jumps” in the datetime column? From the text I got the impression that all data is in 15 minute resolution, interpolating higher resolutions? What did you do in cases of lower resolution?	The missing timestamps were inserted and the corresponding values recorded as NAs.	
11.3	14:45:00 has a flag of 111 → “Mismatch >5% between 15-min values and NRFA daily values” → in these cases, all 15- minute values of that day would need to have a code of 1YZ, right?	Yes, that is correct.	
11.4	The value at 13:30:00 should be flagged as a negative value (Y=1)	Yes	
11.5	The value at 13:45:00 should not be flagged as a negative value	Yes	

11.6	So in general, I think the figure could nicely show the quality control codes, but at the moment it raises more questions than it answers. This can maybe be solved by a more extensive figure caption, explaining the details, or by changing the figure itself	We apologise for these errors and thank the reviewer for identifying them. The figure corresponded to an earlier prototype of the framework and has now been corrected.	
12	L271 (Table 2): It would be good to add details on what a relative spike, absolute spike, drop, fluctuation, truncated low and high flows actually are, i.e. how they are (mathematically) identified or what threshold values were used	We agree that the definitions and thresholds used for these checks are important. These parameters were extensively tested and are described in detail in the companion paper, including the statistical tests, equations, thresholds, and limitations: https://doi.org/10.5194/egusphere-2026-277	None
13	L280 (Table 3): similar to Table 2, it would be helpful to know how e.g. unrealistically high and exceptionally high events are identified / what are the thresholds? Also, Flag 4 can realistically very well occur, e.g. for small scale thunder storms / convective cells, maybe this can be mentioned in the text	To keep the present manuscript focused on the description and use of the dataset, we refer readers to the companion paper for the complete methodological details. This organisation allows the dataset and the quality-control procedures to be documented in sufficient detail for their respective audiences. We also agree that events corresponding to Flag 4 may occur	

		naturally, particularly in small catchments affected by convective storms. The flag therefore identifies observations requiring further assessment rather than observations that should automatically be removed.	
14	L290-L293: It also provides basic metadata like the river name, and the coordinates of the gauge	We agree that these are useful metadata fields. River names and gauge coordinates are provided in the station identification metadata, as described in Table 4.	None
15	L286 (Table 4):	We thank the reviewer for carefully reviewing the table.	
15.1	“UK national grid coordinates” → add the EPSG code to make it easier working with it	We agree	We have updated the metadata with both the British National Grid and WGS84 EPSG Coordinates. And added the corresponding EPSG code for each of them.
15.2	number of missing values → should be percentage of missing values	We agree	Table 4 has been corrected to specify the <i>percentage</i> of missing values.
15.3	what is the “NRFA quality status”, what are the “Daily”, “Pooling”, “Qmed” values in the csv?	Clarified in text	Added in section 4.2.1: The indicative suitability of each station for high-flow hydrological applications, classified into four ordered categories: pooling, qmed, daily, and not_nrfa. Suitability for pooling indicates the highest level of confidence in the gauged high-flow record, followed by suitability for QMED estimation and daily-flow analysis. The not_nrfa category identifies stations outside

			the NRFA network for which no formal suitability assessment is available.
15.4	“visual_inspection_anomalies_meta” this file does not exist in the dataset, I think it is called “01_common_sense_anomalies_meta.csv” there → I would prefer “visual_inspection_anomalies_meta.csv”	We agree	The filename has been standardised in both the manuscript and the data archive as 01_visual_inspection_anomalies_meta.csv
16	L295: see comment above, the file is named differently in the dataset	We agree	See 15.4
17	L353-368: This comment is about entire Section 5; the usage example of the quality flags is very helpful, but I think the entire chapter would be a more exciting read and be less technical if you would formulate an exemplary research question / use case and actually make use of the quality-control high-flow data, just to tell a story. At the end of the chapter, you could compare the results of the quality controlled and the original data to make a point why the quality check is important, based on a practical example	We agree that a practical demonstration of the quality-control flags is important. However, we deliberately structured this section to demonstrate how each of the available files and quality-control outputs can be used. Although this makes the section necessarily technical, we consider this comprehensive overview important for a data descriptor manuscript. Practical examples showing how the quality-control procedures improve the data and their use in hydrological modelling are presented in Section 5.1 of the	None

		companion paper: https://doi.org/10.5194/egusphere-2026-277	
18	L393-L401: Maybe add that one has to be cautious when removing these spikes from the data when interested in high flows (we do not want to remove flood peaks), and why your approach and QC flagging is not dangerous / cautious enough	We are confident that interpolating flagged spikes does not remove genuine high-flow events because of how the flagging criteria are defined. We have clarified this point in the manuscript.	Added: Visual inspection provided confidence that interpolating all flagged spikes is appropriate. The automated criteria identify isolated single timestep excursions, defined by a rapid increase immediately followed by a rapid decrease, or vice versa, rather than the coherent sequence of values across a rising limb, peak, and recession that characterises a genuine high flow event. Visual inspection showed that these flagged values can be categorised into three distinct groups:
19	L430 (Fig. 6): Panel labels a-d are missing, explain what True and False in the title mean, or e.g. station kept / removed in the title	Thank you. We have corrected the panel labels and clarified the meaning of “True” and “False.”	Clarification in legend: True indicates that the measurement was considered physically plausible given the characteristics of the station, whereas False indicates that the measurement was considered physically unrealistic.
20	L453: Give the exact number of stations, not “over 1300”	We agree.	1369 has been clarified
21	L459-L466: you could also mention here that the dataset can serve as a common data source to make studies and results comparable	We have clarified the point	Added: This level of transparency supports FAIR and reproducible data use, facilitates comparison between studies, and supports informed decision-making across a range of hydrological applications.

22	L474: “stage is an important metric” → should be “ stage is an important variable”?	We agree	Replaced “metric” with “variable”.
----	---	----------	------------------------------------

Table 2: Technical corrections addressed

Nbr	Comment	correction
23	L17: I would prefer to not use the abbreviation 15-min here, maybe change to “... flow dataset in 15 minute resolution for the UK, ...”	corrected
24	L24: “QC”: even though it is a common abbreviation, I think it would be better to write it out at the first use “quality control (QC) outcomes”	corrected
25	L70: whitespace after “>” and after “hours”	corrected
26	L129: what is (c.38)? Is it possible to write it out?	corrected
27	L140: Remove one of the full stops (“.”)	corrected
28	L144: formulation “with data in a row” → maybe better something like “years with consecutive data”	corrected
29	L295: Remove one of the full stops (“.”)	corrected
30	L308: whitespace missing before “Additional”	corrected
31	L344: “has a max a gap of 240 minutes.” → “has a maximum gap length of 240 minutes.”	corrected
32	L370-L371: “The following paragraphs present examples of our approach.” Is this sentence correct here? It feels like it should be in the beginning of the Section	corrected
33	L380: “differentfrom” → “different from”	corrected
34	L405 (Fig. 5): The Q99 line is barely visible	corrected
35	L457: “an readily” → “a readily”	corrected
36	L480: “dataset” → “datasets”	corrected

Table 3: Dataset issues addressed

Nbr	Comment	Response
37	An option to bulk download the data without using a terminal (e.g. zip file) would improve the accessibility of the dataset	We have fed back these comments to EIDC, host of the data. These are all outside our control.
38	Also, the wget option displayed when clicking Download on the UKCEH website results in an error, what worked for me was: wget -r -np -nH --cut-dirs=3 -R "index.html*" https://catalogue.ceh.ac.uk/datastore/eidchub/211710ac-f01b-4b52-807f-373babb1c368/	
39	Even though that's probably not on your side, the wget downloaded instructions should be improved	
40	Would it be possible to also include the catchment geometries / shapes in the dataset? I think this could potentially be very useful to e.g. derive meteorological catchment timeseries from raster data or catchment attributes (e.g. about soils)	We agree that catchment geometries would be a valuable addition to the dataset. The purpose of this initial release was to provide the much-needed 15-minute river-flow data in a standardised format for the UK. We consider this primary objective to have been fulfilled. We appreciate this suggestion. The development of related products is ongoing, and catchment geometries will be considered for inclusion in a future product.
41	02_uk_products_meta.csv: typo in column names "Fraction_of_values_that_are_diffrent_from_daily",	Thank you. We have corrected the typos

	“Fraction_of_values_that_are_diffrent_from_annual_maximas”, and “Fraction_of_values_that_are_diffrent_from_pot_peaks” → diffrent → different	
42	file name visual_inspection_anomalies_meta.csv (see comment above)	Thank you. The file has been renamed 01_visual_inspection_anomalies_meta.csv.
43	Generally, the dataset is well formatted, I was able to open all .csv files without any issues. I checked all time series .csv files for duplicate datetimes and timestamp differences that are not 15 minutes, I did not find any issues.	Thank you for confirming this.