Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-49-RC1, 2020

Open Access Earth System Science Data Discussions

Interactive comment on "CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain" *by* Gemma Coxon et al.

Anonymous Referee #1

Received and published: 13 May 2020

The creation of the paper and dataset was motivated by the lack of having one consistent and comprehensive large sample hydro-meteorological dataset for Great Britain. As is outlined in the objectives of the paper, such a large sample dataset would be of great value for many different research purposes. The paper then describes how an impressive amount of data and meta data were combined into one single data set: CAMELS-GB. Furthermore, limitations of the different data and meta data sources were mentioned.

The authors have put in some great effort to produce a very comprehensive hydrometeorological data and meta data set that will be a valuable resource for many hydrological

Printer-friendly version

Discussion paper

studies and more. The meta-data descriptions are elaborate but could sometimes be more accurate. The mentioning of the limitations of several meta-data sets at the end of different sections is useful. Below my suggestions and comments, which are presented in order of appearance in the paper as I do not have any major criticisms.

Line 31: You could add here that these discharge uncertainty estimates are made for different flow quantiles, which is a nice thing to have.

Line 71: are not allowed to instead of "cannot"

Line 85; I agree that “subject to change” negatively affects the repeatability of analyses. However, “subject to change” can also mean subject to improvement, for example in the quality of the streamflow records. The latter might speak in favor of sometimes directly using the most up to date data from the NRFA. A comment on that, and on whether it is planned to occasionally create a new version of the CAMELS-GB dataset, might be useful. Related to the latter comment; The dataset is fixed for a certain time period (1970-2015) for good reasons; however, some researchers might want to include some of the more recent events. In that case, they might either extent the CAMELS-GB data set with NRFA data or directly download complete time series of the NRFA. Provide a remark somewhere what the preferred option would be.

Line 154: Add a note why it was not possible to derive suitable surface area for these catchments.

Line 177 (out of personal interest): Why were shapefiles transformed to ASCII grids? You could also overlay shapefiles and gridded data to derive (weighted) catchments averages. Or is this less accurate / consistent?

Line 205: You could add a note here on human-induced non-stationarities, as you specifically included human influenced time series in the CAMELS-GB dataset.

Lines 232-252: Great that both PET and PETI are included. Both products are derived for grassland, which is of course perfectly fine. However, a note on how representative

Printer-friendly version

Discussion paper

grassland PET is for some of the catchments where e.g. forestry or agriculture are the dominant land use class (as shown in Figure 2) would be nice.

Lines 232-252: How does the PET estimation method for CAMELS-GB compare to PET estimation methods of the other CAMELS datasets? Is it the aim here to use the best possible method that the data allows or to use a method that is consistent across all CAMELS datasets (which favors a comparison across datasets)?

Line 264: “97% (654) of the gauges have at least 20 years” this cannot be seen in Figure 1a.

The comments below refer to either the section or the dataset that is described in this section:

Section 6.1: Provide a a reason for some rare but substantial differences between gauge elevation and minimum elevation.

Section 6.1: Why do two catchments have NA values in their mean elevation, but do have values for e.g., min and max elevation. Please check.

Section 6.2: High and low prec timing; Instead of providing NAs for tied values, you could provide both seasons.

Section 6.2: Definition of seasonality unclear. Provide a reference to the exact method.

Section 6.2: Why an absolute definition for low precipitation frequency and a relative definition for high precipitation frequency (and why these thresholds)? Figure S4f makes sense according to the relative definition but is a bit counter intuitive.

Section 6.2: The provided meta data could be extended with annual averages of the other meteorological variables, at least with average annual temperature.

Section 6.3: Provide a reference to the method used to calculate streamflow elasticity.

Section 6.3: Good that two base flow indices are provided. I personally liked using the

BFIHOST, but the latter is not directly derived from the streamflow record and therefore might not fit in this sub dataset. Might it fit somewhere else? Or was there another reason that it was excluded?

Section 6.3: I think zero_q_freq is the fraction and not the percentage of time with zero flow. Please check. This might also explain why you have some NAs in the slope of the FDC. Please check as well.

Section 6.3: Any reasoning / reference on why you chose 9 times median flow or 0.2 times mean flow as thresholds for high flow / low flow events?

Section 6.5, line 332: Mention the depth range of the top soil layers.

Section 6.5: Nice that you provide ranges of e.g. the tawc! Clarify in table 2 that tawc is calculated over the soil depth available for roots (if that is the case).

Section 6.7: Nice that discharge uncertainty estimates for different quantiles are provided!

Section 6.8: Weren't UKBN catchments also labeled as suitable for low- medium and high flow assessments? Why isn't this information included in the current data set?

Section 6.8: I completely understand the uncertainties with regard to the human interventions, which are nicely outlined in the limitations. However, what should be commented on is that some of the benchmark catchments seem to be relatively heavily impacted by a human intervention of some sort, e.g., the occurrence of a significant amounts of abstractions or the presence of several reservoirs. As a user of the dataset, does this mean that I should interpret some of the benchmark catchments with care? Or that I should be extra careful when interpreting the abstraction and reservoir information?

Section 6.8: It might be useful to additional add the Factors affecting runoff codes, which are presented in the UK hydrometric register, as indicative information on the type of human influence that might have been present at some point in time in the

catchment? Factors affecting runoff codes might also be highly uncertain, but together with the already presented data on abstractions and reservoirs, they might provide some additional clues on possible human influences.

Line 514-518: State that this is annual average precipitation. For me, it would be enough to just mention mm / year (and delete mm / day).

Line 524: Add space between number and unit (mm)

Figure 1a: As you already have a second y-axis on the right, you might also consider adding a second x-axis on the top that indicates the accumulated amount of years with missing data.

Figure 2e: The distribution of total reservoir volume might be more informative (although prob. not very different from the number of reservoirs).

Supplement (S2): Nice that all these maps are added. It would be helpful if they had titles, so you do not always need to read the caption. . .

Dataset: Clear and easy to process (in my case with R, but I am sure that this will be the same for other software). It would be nice to have one .zip file in the parent directory, which allows you to directly download all the time series data at once.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-49, 2020.

Printer-friendly version

Discussion paper