Replies to Franziska Clerc-Schwarzenbach

The line numbers and sections are based on the version of the manuscript that the referee commented on. As the referee wished, we have only commented on changes that were not proofreading comments or other similar minor changes. We went trough all of the minor comments and modified the text appropriately.

Comment

- My main concern is that data from ERA5-Land were used for potential evaporation. These are unrealistically high (e.g., Clerc-Schwarzenbach et al., 2024; Klingler et al., 2021) and do not contain meaningful data regarding potential evaporation. To illustrate the problem, I am plotting below the potential evaporation time series (from the file pet.csv, i.e., the PET product aggregated by the authors from different sources) for one of the catchments in CAMELS-FI, for a randomly selected four-year period (Figure 1). Especially towards the end of the calendar year (around October), the PET values are way too high, going up to more than 10 mm/day. Furthermore, the plot also illustrates that the aggregated time series has very abrupt changes which is also not optimal. I suggest that the authors re-think what data they want to use for potential evaporation. A potentially helpful source could be the paper by Singer et al. (2021) and references therein. Note that Caravan now also uses the data compiled by Singer et al. (2021) instead of the unrealistic ERA5-Land data.

Actions:

- Replaced ERA5-Land potential evaporation with PET estimate of Singer et al. 2021 (called pet_singer), as suggested.
- As a result, the pet timeseries as well as climatic attributes pet_mean and aridity were also updated. Pet timeseries values reduced in the late autumn, and as a consequence both pet_mean and aridity reduced considerably across almost all catchments.
- Changed text in various locations to reflect the change

Reply:

- We thank the referee greatly for suggesting a better data source, in addition to pointing out the issues with ERA5-Land. ERA5-Land was originally chosen, because it covered the time period from 1961 to 1980, which many other datasets don’t. Additionally, the temporal pattern is reasonable (obviously not when combined), even if the quantities were too high. In hindsight, these alone were not good enough reason for using the data, and after inspecting the suggested alternative source decided that replacing ERA5-Land potential evaporation with Singer et al. 2021 was the most reasonable course of action.

Comment

- My other concern is that for several variables (see annotations in the manuscript), there are two different data sources essentially representing the same. I think that this can be confusing to the user and leads to the situation in which every user makes a more or less random choice on what data to use. If you are planning to keep the different data sources for one reason or the other, I would suggest that you include clear guidelines for the users on what data to use (and, if possible, why). In my opinion, it is helpful if we as a community try to use the same data so studies on a dataset stay as comparable as possible.

Action

- The text was modified in multiple locations to make the preferred data source, and reasons for the choice clearer for the end users.

Reply:

- We understand the concern for the ”selection at random” issue. We included multiple datasets only if there was no one clearly best dataset, and think that further research should be conducted to determine the optimal dataset.
- We would also like to note multiple datasets can also be a benefit, as it has been shown that data-driven models can perform better with meteorological forcings from multiple different data sources (Kratzert et al., 2021).
- Finnish meteorological institute provides only snow depth as a gridded product, but many hydrological applications would actually want SWE. CCI3 is able to better capture snow timing better and performs well as long as snow cover isn't too thick. ERA5-Land does not capture the timing of start and end of snow period as well, but performs otherwise better especially in Northern Finland, where snow is even more important for hydrology than in the south. This is why we suggest using snow depth if possible, and otherwise preferring ERA5-Land if one source of SWE has to be given.
- After the adoption of (Singer et al., 2021) as a data source, we think the combined PET is now a reasonable suggestion for the primary PET product. This is because researchers have to worry less about observation gaps when using pet_fmi, or effects of snow when using pet_fmi or pet_singer alone. We also suppose fmi_pet to be higher quality than pet_singer, since it has been specifically developed for Finland, altough confirming that would require a more thorough comparison, that is out of scope of this article.
- The first author is currently conducting research onto importances of different variables, which will also clarify the quality of different products presenting the same variable. The initial results support the decisions described here.

Comment
- I suggest to add an additional readme file in the folder “artificial-cross-catchment-bifurcations” that describes clearly what the data contained in this folder are and mean. I think this would be valuable, as this folder is for sure a special item of the CAMELS-FI dataset.

Action
- Added a readme

Reply
- After consideration a readme was added, since the item is indeed not common in CAMELS.

Comment:
- I think it is handy that you provide the timeseries per catchment as well as per attribute. You clearly state in the support file that the two folders contain the same data, just organized differently. Potentially, you could add another readme file in the “timeseries-by-attribute” folder stating this again. There, you could also let the user know which of the timeseries is recommended if there are several timeseries for the same variable (see general comments).

Action
- Added recommendations to support file, did not add additional READMEs

Reply
- After consideration, we decided not to add additional readme files to the suggested data folders. This was for two reasons. First, those folders contain a lot of files, meaning that Readme can be easily missed. Secondly, we would like to avoid splitting the documentation as much as possible, since that makes future updates easier and less error-prone.

Comment
- Line 37: Are there more examples than EStreams? Maybe the Italian dataset FOCA (double-check!)?

Action
- Found another example

Reply

- We found another example, Chen et al., (2023). FOCA is indeed also purely indice based. However, this section of the text focuses on datasets larger than one country, so we did not cite it.

Comment

- Line 38: This needs some more explanation. What do you mean here? In addition, note that the now cited EStreams dataset provides a catalogue that makes it possible to access the Q data.

Action

- Clarified the text

Reply

- We did our best to clarify the text. While Estreams does provide links for accessing the different streamflows, this can still be considerable amount of work and thus a barrier to adoption compared to providing streamflow values directly.
- For example, here are the instructions provided for Finland. Lets just say that there was a reason we used the API instead of the web interface (https://rajapinnat.ymparisto.fi/api/Hydrologiarajapinta/1.0/) . Significant obstacles bolded by us. We would like to note that the data maximum is very low, only a few catchments at a time.
  - 1. First **make an account** using the link: https://www.syke.fi/fi-FI/Avoin_tieto/Ymparistotietojarjestelmat/Rekisteroityminen;
  - 2. Once logged in: click on the "Ympäristötiedon hallintajärjestelmä Hertta" to get to the database;
  - 3. Translate to English by first clicking on "Asetukset" and clickon “Kieli” and choose “ Englanti” for English. “Suomi” means finish. Click ok “Hyvasky” and close the pop-up ‘Sulje”. ;
  - 4. Then you click “Hydrology > Hydrological observations > Data search” ;
  - 5. **The site has a maximum amount of data you can download in one go.** So you need to choose an area and then search up all the stations in that area, download and then move on to the next area. One tip is to choose "county" in the tab "Area";
  - 6. Now you should seelct the discharge. Click on "Observation station", select "Parameter", then "Virtaama", then click on "Accept";
  - 7. After selecting the "Area" and the "Observation station", click on "Places on list" on the right side of the screen. You will be redirected to another page;
  - 8. Now **select the desired stations (there is a maximum**) and click on "select action" > "resultlist", and then select the desired time-period;
  - 9. Now click on "to file", select the desired file format (EXCEL), then click ion "to file" again. Now you will be able to download the data;
  - 10. **Repeat until you are done with all the desired stations**.

Comment

- Line 44: There are now also other temporal resolutions (e.g., CAMELS-LUX, CAMELS-GB v2).

Action

- mentioned the hourly resolutions

Reply

- Updated.

Comment

- Line 47 The list does not contain all CAMELS datasets (e.g., CAMELS-LUX, CAMELS-COL, LamaH-ICE etc. missing). As the family is growing so fast, I would just give a few examples here.
- In addition, if you mention Brazil, consider also citing the CABra dataset by Almagro et a

Action

- Rewrote the section, searched for existing datasets and added those to appendix C. Also created a github repository, that we intend to keep updating, and hope that the community will find it.

Reply

- We wanted to include a list of earlier works, but decided that appendix would probably be the best place for that. Considering the rapid change, we thought that an updating github repo could be a good place to keep up as an community. At least we would have found that helpful when starting this project. The first author plans to keep updating the list for the forseeable future, but in case something were to happen to him somebody can fork the repository and keep updating it.

Comment

- Line 58: The study by Clerc-Schwarzenbach et al. only looked at the quality of the meteorological time series, not of the catchment attributes.

Action

- Reformat

Reply

- The sentence has now been reformatted to only make claims about meteorological timeseries

Comment

- Line 81: How does this ensure high data quality?

Action

- Reformatted

Reply

- The section has been reformatted to better reflect the intended meaning.

Comment

- Line 91: I think that this is great. However, consider giving a reason why you chose to do so. Furthermore, as far as I understand, the scripts are from CAMELS (Addor et al.), maybe you can explain this a bit more.

Action

- Added reasons behind the decisions, cited (Addor et al., 2017) as the source for scripts

Reply

- Updated. The scripts are indeed from (Addor et al., 2017). This had a large influence to the provided signatures.

Comment

- I think it is very important that you decide for one term: outflow / streamflow / runoff / discharge, etc. Except of course if you want to use the different terms for different meanings. In that case, please clearly define what you mean by what. This makes it easier to understand if you are always referring to the same variable.

Action

- Unified naming, addded explanation to lines to the beginning of section 4.1 Hydrologic time series.

Reply

- We decided to use the term discharge, and divide it to volumetric or specific discharge, if the separation was needed. This was consistent with the naming of the timeseries variables.

Comment
- Line 124: A map or some description on where these places are would be helpful here. (Also for the place names further below.)

Action
- Added map of used toponyms to appendix D

Reply
- After rereading the article, relatively many toponyms were indeed used. In addition to used names, we also decided to include named gauges in the map.

Comment
- Line 129: This may need some more explanation, e.g. that you computed the streamflow at this gauge yourself (if I understand correctly) and that you are therefore calling it like this.

Action
- Expanded the explanation, reorganized the text for clarity.

Reply
- We hope that the text is now clearer.

Comment
- Line 133: In the support document, you use xxxx-xxxx and xxx or xxxx, instead of the zeroes, which I actually think is the better way of indicating it, so I would consider replacing here. In addition, there's also at least one gauge id of the format xxxx-xxxx-xxxx, maybe it's worth mentioning this here, too.

Action
- Reformatted

Reply
- Reformatted

Comment
- Line 168: Only considering the catchments included in CAMELS-FI?

Action
- Clarified

Reply
- Yes. Clarified in the text

Comment
- Table 1 (gauge_name Description): Are lakes included as "catchments" in the dataset? This comes a bit as a surprise here.

Action
- no action

Reply
- Many of the gauges are indeed named after a lake. The observations are conducted by measuring the water level of the lake, and calculating the discharge of the lake to the downstream from that. Even though the measurement might be from some other part of the lake, the location of gauge in both SYKE’s original data and CAMELS-FI are marked at the outlet of the lake. Thus, for many catchments, a portion of the catchment closest to the marked location of the gauge is covered mostly by a lake, which gives its name to the gauge. We hope that this clarifies things.

Comment

- Table 1: Just like with the different terms for streamflow, I think it would be helpful for the reader if only one term (basin / catchment / watershed etc.) was used, unless different things should be described with the different terms.

Action

- Unified naming convention

Reply

- We decided to use catchment to describe the upstream area from a particular gauge, and basin for streams that share an outlet to sea. We renamed watershed to catchment boundaries for clarity and consistency.

Comment

- Table 2: Potential evaporation from ERA5-Land is not reliable. See for example Clerc-Schwarzenbach et al. (2024, already cited) or Klingler et al. (2021, LamaH-CE data description paper, 10.5194/essd-13-4529-2021).
  I do not think that it is a good solution to use ERA5-Land potential evaporation when there are no other data available. For example please look in Singer et al. (2021, 10.1038/s41597-021-01003-9) and references therein to find suitable alternatives.
- Table 2: Wouldn't this (pet_fmi) already be a much more suitable way to include potential evaporation?

Action

- See first reply

Reply

- See first reply

Comment

- Table 2: It is not clear to me how the lines in this table are organized, what is the data source for temperature?

Action

- Made lines more understandable

Reply

- We hope that the table is more readable now. FMI is the source for temperature.

Comment

- Table 2: I am a bit critical when it comes to including different data sources for the same variable without clearly indicating what should be used. Please consider adding a clear guideline for the users of the data.

Action

- Added guidelines to Section 4.2 Meteorologic timeseries for PET and at the end of section 7 Uncertainty and consistency of hydrometeorologicla observations for SWE.

Reply

- We added recommendations for default dataset. Please also see reply for the second comment.

Comment

- Line 189: This paragraph is difficult for me to understand. Does this have anything to do with the "virtual gauges" described before?
  Maybe this becomes clearer when using the data, but potentially it could be worth re-writing this paragraph.

Action

- Clarified the text

Reply

- We did our best to clarify the text. This is not related to the virtual gauges. Based on other comments, it seems that the concept became clearer after looking at the data, and we added the asked readme to this section of data for further clarification.

Comment
- Figure 2: Make sure that each value is only represented once (currently, a time series lenght of 30 years could be both, light brown as well as light turquoise).
- The font size on the left part of this figure is quite huge, you could consider saving some space there and then adding some space between the two legends to make the figure more easier to separate.

Action
- Made the changes

Reply
- Thanks for noticing the issue.

Comment
- Line 202: Is there a reason for the missing data?

Action
- Investigated

Reply
- After investigation, we did not find a clear reason for this in the metadata or article describing the dataset. We also contacted the corresponding author of the dataset, but never received a reply.

Comment
- Line 212 : Does this end up in a consistent time series (one of the aims stated in the beginning)? (Combined pet)

Action
- See first reply

Reply
- The changes applied to PET have already been described. We feel that after these changes the combined PET is the best dataset, as we suspect that FMIs higher resolution product made specifically for Finland provides better results than Singer et al. (2021). We agree that this is not ideal for consistency, but the new datasets seem to merge relatively well, especially in the autumn. Springtime combined PET sees rapid increase (sometimes even a jump) after snowmelt, when the combined PET changes from snow evaporation to FMIs PET. However, since snow and ground albedo are quite different, this might be actually more realistic than gradual increase seen in Singer et al. Usually the presence of snow also prevents temperatures from rising too much above zero, limiting evapotranspiration.

Comment
- Line 215: This makes sense, as thanks to the other data included, the overestimation of ERA5-Land is slightly buffered. Still, the fact that we see arid catchments in Finland based on the PET data used here (Fig. 3) supports the fact that the PET data need some more work.

Action
- See first reply, updated text

Reply
- These catchments are not arid anymore after the latest update. See also the first reply.

Comment
- Line 225: If you trust in all the (SWE) products, would an ensemble be a good solution to provide the user with one time series that they should prefer?

Action

- no action

Reply

- Ensemble would be problematic, because CCI3 is known to underestimate SWE if snowpack gets too deep. This would create too low values for Lapland and eastern Finland. With shallower snowpacks, the differences between the products are smaller, and we didn't find conlusive evidence that would clearly show that one or the other would be better. ERA5-Land also gives too long snow periods compared to FMI's product. Because ERA5 provides reasonable estimates for the whole country, we decided to recommend it. See also the second reply.

Comment

- Table 3 high_prec_timing description: (if two seasons register the same number of events, a value of NaN is given)
- Could there be a solution for these cases that avoids the loss of all information?

Action

- No action

Reply

- There could be ways to provide this, but it would cause CAMELS-FI to deviate from one of the more standard parts of previous CAMELS (climate indices calculated with scripts from Addor et al., 2017), which is why we are hesitant to change this behavior (and cause a compatibility deviation) without especially good reason. We would also like to point that a value of NaN does not equal loss of **all** information, as it tells the end user that at least two seasons are equal. Thus a value of NaN is a sign to the user that this specific division of months might not be the optimal one for that gauge or that the seasonality of rainfall is not that strong. In CAMELS-FI, the most suitable months for high_prec_timing with NaN (n=3) values would likely be either July-September or Augustus-October, as all other catchments have the most high precipitation events on either summer (most catchments) or autumn (some south coast catchments) months. There are 7 catchments with nan in the season of low precipitation events. All other catchments have their low precipitation timings in the spring, and so the NaNs are also likely to be close to that (no recognizable spatial pattern)

Comment

- Table 3, elev_gauge Description: Can this be explained in the text somewhere?

Action

- Slight modifications to the text

Reply

- Yes, at section 5.3 Elevation

Comment

- Table 3, elev_range description: Would it make more sense here to use maximum and minimum? Or how are the parts of the catchment below the gauge location accounted for?
- 249: (due to mines) Contributing to the catchment?

Action

- No action

Reply

- It does not make hydrologic sense to use minimum instead of gauge elevation, because minimums that are significantly lower than gauge elevation are caused by mining operations. The mines are often open pit mines, and are kept dry with pumps. The Hydrologically the depth of the mine is not as relevant as the gauge due to this pumping. If there are no mines, minimum and gauge elevation are typically very close to each other. However, we wanted to include both variables and some explanation to avoid confusion that could arise if only one was provided.

Comment

- Table 3, stream_elas: How is this calculated?

Action

- Cited the source for the method, (Sankarasubramanian et al., 2001).

Reply

- Here is the algorithm used by the code from https://github.com/naddor/camels:
  - Calculate the mean specific discharge and precipitation for the whole signature period (mp_tot, mq_tot)
  - Calculate mean precipitation and specific discharge for each year (mp, mq; lists of values for each year)
  - Calculate yearly deviations from the long term mean for precipitation and specific discharge (dp, dq)
  - stream_elas = median((dq / mq) / (dp/mp))

Comment

- Why is it helpful to include both baseflow indices?

Action

- removed baseflow_index_lfstat

Reply

- We didn't offer this much tought when initially constructing the dataset. We used the code from https://github.com/naddor/camels, which produces the two by default. CAMELS-GB (Coxon et al., 2020) had also included two baseflow indices, so including two ways to calculate the indice didn't feel that odd. With further thought, if somebody is interested at the differences of baseflow indices or they require a specific method, they will likely calculate it anyway by themselves. With the added concerns expressed by the reviewer towards the possibility of end users selecting between two options randomly, we decided that removing the baseflow_index_lfsat was the right decision. The approach of modified Lyne-Hollick filter used by (Ladson et al., 2013) seemed to be more common according to a brief and nonsystematic literature review.

Comment

- Table 3, high_q_freq description: why the median here but the mean for low_q?

Action

- No action

Reply

- The code used https://github.com/naddor/camels/blob/master/hydro/hydro_signatures.R for the relevant function hass the following comment: "In dry conditions, the median of q can be 0...". This would obviously be problematic, as 0 * 9 is still 0. There are some catchments in CAMELS-FI with discharge under the detection limit in the summer (zero discharge), but only for short periods. Even though this is not an issue in Finland, we don't see a reason to break compatibility with other CAMELS using the same code.

Comment

- Table 3, zero_q_freq: Were only the days with streamflow data taken into account here?

Action

- No action

Reply

- Yes, (https://github.com/naddor/camels time/timetools.R → find_avail_data_array & hydro/hydro_signatures.R → compute_no_flow)

Comment

- Table 3 Land cover data source: Please indicate the naming convention for the different years

Action

- Added the naming convention to section 4.4

Reply

- Modified the text, but not the table due to wanting to keep the table compact. Full list and format is also available in the support document.

Comment

- Line 346: What are the measurements from the dams based on then, if not on rating curves?

Action

- Clarified

Reply

- Modified the sentence to clarify it's intended meaning, that the non dammed results are based on **more uncertain** rating curves.

Comment

- Line 362: How do you distinguish between PE and PET?

Action

- Removed mention of PE

Reply

- Realized that there is no need to distinguish potential evaporation (PE) and PET from each other in that sentence, and thus removed the mention of PE.

Comment

- Line 418 Not quite sure about this, but in my understanding "conceived" is a bit mis-leading here. I would suggest to use "designed".

Action

- No action

Reply

- The cambridge dictionary gives three definitions to the word conceive
  - To imagine something
  - To invent a plan or an idea
  - to become pregnant, or to cause a baby to begin to form
- The first author was thinking especially of the first two meanings while writing, and wants to distinguish those tasks from the design, which was done more collaboratively. We think that it is clear enough from the context that the word is not used in its third meaning.

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Chen, X., Jiang, L., Luo, Y., and Liu, J.: A global streamflow indices time series dataset for large-sample hydrological analyses on streamflow regime (until 2022), Earth Syst. Sci. Data, 15, 4463–4479, https://doi.org/10.5194/essd-15-4463-2023, 2023.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling, Hydrol. Earth Syst. Sci., 25, 2685–2703, https://doi.org/10.5194/hess-25-2685-2021, 2021.

Ladson, A. R., Brown, R., Neal, B., and Nathan, R.: A Standard Approach to Baseflow Separation Using The Lyne and Hollick Filter, Australas. J. Water Resour., 17, 25–34, https://doi.org/10.7158/13241583.2013.11465417, 2013.

Sankarasubramanian, A., Vogel, R. M., and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, Water Resour. Res., 37, 1771–1781, https://doi.org/10.1029/2000WR900330, 2001.

Singer, M. B., Asfaw, D. T., Rosolem, R., Cuthbert, M. O., Miralles, D. G., MacLeod, D., Quichimbo, E. A., and Michaelides, K.: Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present, Sci. Data, 8, 224, https://doi.org/10.1038/s41597-021-01003-9, 2021.

Replies to the second referee

Major comments

1.

- Comment: L 106. “A Minimum of five years (1826 days) of observations required. We considered this sufficient to calculate streamflow signatures.” Many CAMELS data sets, including the first one for the USA, require about 20 or 30 years of Q observations for inclusion. For long-term studies, WMO recommends at least 30 years of hydrometeorological observations. The reason for these stricter requirements is that hydrometeorological variables often have long-term persistence (e.g. O’Connell et al., 2022) and temporal clustering (e.g., Chagas et al., 2024; Lun et al., 2020), thus, 5 years of data may not be representative of long-term hydrological behavior. In order to include the hydrological indices of catchments with fewer than about 20 years of data, it would be interesting for the manuscript to assess whether these values are representative of long-term conditions, for example by comparing with neighboring catchments that have longer data.
- Actions: Modified the corresponding line, cautioning that short timeseries may not represent long-term conditions, and that the end user needs to evaluate suitable timeseries length for their application. Added comparison between the indices of short and long timeseries to section 6.2 of the revised manuscript (section 7 of the preprint).
- Reply: We agree that five years is not long enough for representing long term state of a catchment robustly. The previous formatting was not particularly informative of our intents and could be read in a misleading way. The text now explains that we wanted to include catchments with shorter time periods as they can still be very useful for studying individual events or calibrating/training hydrological models, for example. We consider it reasonable to provide indices also for these catchments, so that the end user can get a quick overview of the hydrologic behavior of the catchment during the period of available data. The length of time series used for calculating the indices is also clearly indicated in the hydrologic indices data file, allowing for easy filtering based on user requirements. As requested, we compared indices of short timeseries to longer neighbouring catchments. The results are now included in the section 6.2 of the revised manuscript (section 7 in the preprint).

2.

- Comment: L 211. “In addition to providing the aforementioned data, we combined the data into one convenience PET attribute from 1981, which was also used for calculating climatic signatures (see section 5.2). This was done by using snow evaporation when the snow depth of the catchment was over zero, and filling the snow-free days with FMI PET if it was available; the remainder was completed by ERA5-Land potential evaporation.” Combining the time series from two different sources can be problematic because each source is usually

built from completely different data and models, particularly evapotranspiration variables which often have high uncertainties. In order to keep the combined PET time series, it would be interesting to include a more in-depth investigation of whether that is appropriate, including comparisons with other data sources, where possible, and assessing if it does not introduce step changes in PET and other spurious artifacts.

- Action: Investigated step changes, modified text at sections 3.2 and 6.3 of the revised manuscript (4.2 and 7 in the preprint)
- Reply: Combining different data sources for evaporation is indeed not ideal, and investigating the step changes gives users important information to decide whether this combination suits their needs. We want to emphasize, that all presented evaporation datasets are available separately in CAMELS-FI, so that users are free to select a dataset that suits their requirements, and are not forced to use the combined pet. Comparison to a reference dataset would have been interesting, but unfortunately the regular evaporation measurements conducted in Finland have been conducted as pan evaporation, and only for the summer months. This prevents direct comparisons to the pet methods utilizing Penman-Monteith, and also seasonal dynamics outside of summer. According to our analysis, all the other changes are typically stepless, except changes between fmi and snow evaporation. These are almost fully limited to springtime. There, the median change is 1.8 mm d^{-1} (3 x typical variation), and can be even larger. However, it should be noted that even though the change is too quick (one day), this sort of increase spread over slightly longer time, such as a week, could be somewhat realistic, since melting and evaporating snow 1) has high albedo 2) Prevents temperature from rising too much above zero and 3) increases the relative humidity of air above. After the snow has melted, temperature is free to rise and relative humidity is often quite low, leading to rapid increase in potential evaporation (Betts et al. 2001).

3.

- Comment: Section 4.1 Hydrologic time series. The manuscript could benefit from a more in-depth description of how the streamflow measurements are conducted by the data providers (SYKE and ELY), including how the measurements have changed over time (if such information is available), how the data were quality checked, and whether data quality flags are available and included. Some description is already present in Section 7, but I believe that it could be expanded. If the data providers have documentation describing how the measurements are or were conducted, it may be worth citing it in the manuscript.
- Action: Added data quality flags to the data. Revised text appropriately at section 3.1 and 6.1 of the revised article (4.1 and 7 in the preprint) by adding more detailed information on how the measurements were conducted and expanded quality descriptions.
- Reply: Thanks for suggesting the addition of quality flags! We added them to the dataset, as well as related remarks associated with some of the quality flags. We added information on the equipment used for the measurements to the text, along with the changes to those. We also added information on the quality check process. We were unable to find detailed citable documentation that was peer-reviewed or in English, but cited a relevant publication by SYKE, in Finnish. Large portions of the metadata provided are based on internal knowledge from SYKE, provided by Dr Jari Uusikivi (one of the authors), whose team is responsible for the observations.

4.

- Comment: I could not find geological indices. The manuscript mentions that hydrolithological data are not available. However, the manuscript could include other variables such as geological type, to align the data set more closely with the CAMELS data sets.
- Action: Added geological rock type to the dataset
- Reply: The original decision to exclude lithology was made based on discussions with two Finnish geology professors (Esa Heilamo and Antti Ojala), who estimated that the fracture

density would be more relevant than the rock type, as most of Finland is covered in either metamorphic or plutonic rocks, which conduct water poorly outside of cracks. Unfortunately, no high quality crack density data covering the entire Finland exists. However, including the rock type data was relatively simple, so we decided to add that as suggested, in case somebody finds it useful. We would like to note that available geological data already exhibits large variations between different CAMELS datasets, so it is basically impossible to align CAMELS-FI to all of them. After deliberation, we decided that CAMELS-CH includes a rock type classification, that is similar to available rock type data for Finland, so we used that as a basis of inspiration.

Minor comments

5.

- Comment: Fig. 1 and Section 5.6 (Human influence). Does the regulation referred to here indicate regulation through artificial dams? What are the criteria for defining the regulation classes (not regulated, minor and major regulation)? It is not clear in either the figure caption or the section.
- Action: clarified section 5.6 (4.7 in the revised manuscript)
- Reply: The active regulation is mostly conducted trough dams, but there are also exceptions, such as adjustable flood levees, extractions, or permanent lowering of lake levels. It should be noted that catchments from class 1 (no active regulation) can still have dams, but they often only have the role of preventing the level of the body of water behind them from going below a certain level. Some small hydropower stations without a reservoir are also included in class 1 (after consultation with topic experts from SYKE), because their discharge is not dependent on regulation limits or electricity prices.

6.

- Comment: Table 1. The difference between gauge_id and basin_id is not clear. The gauge_id is described as “catchment identifier”.
- Action: Added definition of a basin to line 154 of the revised manuscript (line 142 of the preprint)
- Reply: Basins are defined as all the locations that share an outlet to sea. Each gauge in CAMELS-FI has a corresponding catchment, and since their relationship is one-to-one, the gauge_id can be used to refer to both the gauge and its catchment.

7.

- comment: Table 3. How was the attribute “regulation_level” calculated?
- Action: The description was already revised based on the other referee’s comment
- Reply: It was classified based on the criteria laid out at lines 315 to 320 in the revised manuscript (287 to 291 in the preprint).

8.

- comment L128: “Kammonen, Luiro (1358)”. Is 1358 the gauge_id? It may be worth clarifying.
- Action: clarified
- Reply: yes, clarified.

References

Betts, A. K., Ball, J. H., and McCaughey, J. H.: Near-surface climate in the boreal forest, Journal of Geophysical Research: Atmospheres, 106, 33529–33541, https://doi.org/10.1029/2001JD900047, 2001.