Dear Editor,

Thank you for your comments on the manuscript, which have helped us improve its quality. We are sorry that we could not resubmit the modified manuscript before Christmas break. We planned to keep the dataset available on GEUS repository. The problem of GEUS repository has now been resolved, and we have tested that the dataset can be successfully downloaded from the website via the link. Our responses to the other comments are highlighted in blue.

Best regards,

Jun Liu (on behalf of all co-authors)

I cannot access the data at the GEUS repository. For single files, I receive the following error:

{"status":"ERROR","code":404,"message":"Datafile 84632: Failed to locate and/or open physical file."}

For downloading the full dataset, I receive an empty zip file. Please carefully check that the repository is stable, complete, and up to date. I would also guess that linking your paper will be a good addition. If the title can be updated, it might be worth considering.

Reply: There were some issues with the repository, which have been fixed now. We have tested the dataset can be downloaded from the link:

I find your data availability statement confusing, plus GEUS is misspelled there. Here is a suggestion:

The CAMELS-DK dataset is available at the GEUS Dataverse under a Creative Commons Zero license at https://doi.org/10.22008/FK2/AZXSYP (Koch et al., 2024). Since all data is continuously updated by the respective providers, we list the respective sources of their free availability: ID15 catchment shapefile is updated. The newest version of the shapefile can be requested by writing to id15@ecos.au.dk. High-resolution DEM data is available here https://dataforsyningen.dk/data/928. The gridded climate data is created by the Danish Meteorological Institute (DMI) and can be downloaded through the DMI Data API (https://opendatadocs.dmi.govcloud.dk/DMIOpenData). The observed streamflow data can be downloaded from the Danish Environmental Protection Agency (https://mst.dk/erhverv).

Simulation results of surface water and groundwater dynamics from the National Hydrological model of Denmark are available at (https://dennationalehydrologiskemodel.dk/).

Reply: we have modified the section, see lines: 520 – 528.

There are missing spaces (e.g. L23, L297, L436, and many more) and at least one twisted abbreviation (L514). This gives the manuscript a sloppy appearance. Please carefully correct this. This will also ease the further copyediting with Copernicus.

Reply: we have carefully checked the space errors and typos of the manuscript.

L128ff: ID15v2.5 vs. ID15 I find this very confusing. Obviously, there are several versions of the basins. But would it suffice to clarify this once and stick to the term "ID15-catchments"? L280 would be a good spot to refer to the specific version again as part of the citation.

Reply: agreed, we have uniformed the name of ID15 catchments as 'ID15 catchments' in the manuscript, dataset, and dataset description.

L228ff: Please carefully check the argumentation structure of these paragraphs. I find it hard to follow. Moreover, I am under the impression that the numbers are merely to show off. Could the numbered list L253ff simply become such a list? As block text, I get lost. And I have difficulties keeping track of the reasoning for each of the 8 simulated variables as part of the CAMELS-DK dataset. Exactly this could guide the argumentation in this subsection?

Reply: We removed some of the details surrounding the calibration of the DK-model, following your suggestion that they might partly have been unnecessary information here. Our argumentation in the latter part is that, as streamflow is tightly coupled with groundwater in Denmark, it is relevant to include groundwater-related variables (and others, such as soil moisture) from the DK-model. We cleaned up that section a bit, removing some details, hoping to make this more clear to the reader, the corrections can be found in lines: 258-277

I am even more confused by the setting that the model is used at 500 m lateral resolution but in L356 you refer to the 100 m grid plus vertical definitions comprised in 5 variables. Why did you separate subsection 3.3 from 4.6? I know that my comments are very late (and I am sorry for this), but I hoped that the comments of the referees would solve this.

Reply: We understand your confusion. The hydrological model exists in version with 100m and 500m, where we use results from the 500m version, as mentioned in line 233. In line 356, we refer to the resolution of the hydrologeological model – which is 100 m. The same hydrogeological model is used as basis for both the 500m and 100m version of the

hydrological model (as applies for much of the other input data; both the 100m and 500m versions of the hydrological model share the same input; the data then just is resampled internally by the modelling software to the actual hydrological model resolution. To clarify this, we added a sentence where the 500m and 100m versions of the hydrological model are mentioned, see lines 242-244.

L341ff: If you include these data, cite it here!

Reply: the references of the data has been added to the manuscript.

L470ff: Since I could not access your data, I could also not check your data description readme. First of all, you could and should refer to this readme here.

Reply: data description file has been mentioned in the first sentence, see line 476.

L473: Tables 2-5 (plural)

Reply: the typos has been corrected.

L476: I guess shapefiles are plural too.

Reply: the typos has been corrected.

L486f: I am still a bit unsure about the strong emphasis on groundwater – which is not a standard feature in other CAMELS publications. Moreover, you have not used a ML model to simulate the ungauged basins but a "physics-based grid model" which is also something "special". I suggest clarifying these specifics very early in the manuscript as part of the speciality of Denmark (as all CAMELS datasets have regional specialities) and to make it tangible as an interesting addition.

Reply: yes, we have added some sentences to explain why we want to have simulated groundwater dynamics in the dataset (see Lines 106-113).

We have developed ML models to simulated discharge in all catchments in another study (10.5194/hess-28-2871-2024), which has been cited in the current manuscript to explain the benefit of involving simulated variables, such as simulated streamflow, soil moisture, and shallow groundwater levels for streamflow modelling with Long Short-term Memory (LSTM) model (see lines 98 -101).

In the summary you give the quite generic statement, that your CAMELS dataset can "assist machine learning studies" – sure, this is exactly why CAMELS and more, so CARAVAN came into existence. But with your addition about

groundwater and the MIKE-SHE* references (since I could not access the data: are they existing for the catchments with observations as well?) this could be pointed out more specifically?

Reply: We have revised the conclusion/summary section of the manuscript. The reasons for including DK-model simulations are also discussed in this section; please refer to lines 486–501.

I am sorry to come with these points at a late state of the process. I hope it will be very easy for you to still do these small corrections before the Christmas break to have the manuscript swiftly accepted thereafter. If you disagree, maybe there is another minor detail which should be clarified so that readers like myself do not derail during reading?

Reply: We appreciate your comments on the manuscript and the dataset, they really helped us to improve the quality of the work.