Dear Reviewer, we thank you very kindly for the constructive comments and time spent in reviewing the manuscript. We have carefully revised the manuscript according to the comments and suggestions. Below we provide a point-by-point response.

- *R1. The dataset is a compilation of several data sources that are quite well-known and acclaimed in the international hydrological community, however bringing this data together undoubtedly was an effort. At the same time, these sources of data ensure consistency and universality for the resulting dataset under review a priori. It was already mentioned in the above comment, that the regridding procedures are not documented in the manuscript, hence its' consistency still needs validation.*

  A1. We agree that this information should be contained in the manuscript itself rather than just the preprocessing code that we provide. We substantially extended the section on grid harmonization to address these concerns while also discussing limitations more explicitly. Please find below the new content section

  Grid Harmonization:

  As a common grid format, we decided to use the grid of the ERA5-Land reanalysis dataset (Muñoz Sabater, 2019; CopernicusClimateChangeService, 2022), which covers the earth's surface with a $0.1° \times 0.1°$ resolution, corresponding to roughly $9km \times 9km$ for the case of our study area. ERA5-Land contains a vast number of meteorological variables and has an hourly resolution, spanning from 1950 onward. It has been widely used and is actively maintained and updated. This means the dataset we provide with this paper remains easily extendable, should a user like to e.g. include an additional meteorological variable in their experiments, extend the study area or increase the temporal resolution. Our study area comprises a total of 7169 grid cells on this grid.

  The spatial data originally comes in different formats (vector or grid), projections and resolutions. All data sources were thus harmonized to the grid of ERA5-Land. The first step consisted in converting the maps to a common coordinate system. For the sake of compatibility with ERA5, we decided to use the geographic coordinate system WGS 84. Then for each map separately, polygons covering 0.1▪ in latitude and longitude with the grid cell at the center were extracted. For categorical maps like hydrogeology and land cover, consisting of classes such as "Artificial surfaces: Urban fabric", the fractions covered by each class within each polygon were calculated. This results in a distributional description of class occurrence maximally conserves the original information, as no averaging or other kind of aggregation is necessary. Quantitative information, like e.g. clay content in topsoil, was however aggregated in a final step within the polygon. As a downside to this approach, note that both cases require calculating coverage areas in a geographic coordinate system. This treats the surfaces as flat instead of accounting for the earth's curvature, making the calculations imprecise. We consider this effect negligible here since the surfaces contained in the grid cells are so small that they can be safely considered approximately flat. A more severe limitation is the fact that at high latitudes, polygons defined by a given latitudinal and longitudinal extent become substantially more narrow on the side facing the pole, which also impairs the calculation of area. We could not change the polygons to counteract this effect since they are implicitly dictated by the ERA5 grid. At the moderate latitudes of our study area and especially the small polygons used in the grid, this distortion can still be considered acceptable for the sake of harmonization with ERA5, depending on the application. However, applying this approach in polar regions for example would necessitate an intermediate step of choosing a suitable projected coordinate system for calculating the areas in order to make the distortion explicit and thus better understand its effects. Boundary effects are not an issue with this vector approach as no interpolation is required, however. All maps covered regions well exceeding the study area.

  Specifically for the four map types, the hydrogeological map already came in the WGS 84 reference system. The land cover map came in vector format and LAEA coordinate system, so the polygon's coordinates could simply be calculated point-wise using the geopandas Python package (den Bossche et al., 2025) with pyproj (Snow et al., 2025) as a backend. This was followed by calculating the fraction of each class within the polygon described above. The soil map came in LAEA reference system but as a raster. We decided against interpolating as it would not contribute any additional information. Instead, we simply considered each cell in the raster as a separate polygon and calculated the fractions using the vector method described above. The situation was the same for the elevation map, with the added initial step of down-sampling the map from 25 m resolution to 500 m resolution using weighted average resampling implemented in rasterio (Mapbox, 2024).

- *R2. The authors state that the dataset is more suitable for distributed hydrological models' testing rather than CARAVAN dataset since it's not lumped. While this is obviously true, some* *uncertainty is still possible originating in interpolation from the variables' sources and eventually in interpolation on the particular model's grid.*

A2. We included a discussion of potential shortcomings of the employed algorithms in the aforementioned new subsection, see above for details. We chose the ERA5 dataset's grid as a target to unify onto since it is very widely used, continuously updated and offers the possibility to add further variables for other research questions.

- *R3. The name of the manuscript refers to rainfall streamflow modelling, however the domain and especially its' southernmost part of is located in snowmelt runoff area. The authors are advised to address this issue, since no snow-related data is given in the dataset.*

  A3. Thank you for pointing this out, we made according changes to the manuscript that are listed in the difference file. When compiling the dataset, we tried to ensure maximum comparability to related datasets (please also refer to our answer A2 to the other reviewer. Hence, we did not include snow specifically, although ERA5 offers several snow-related products which could be integrated into our dataset with the code with provide. We also consider snow-related variables to be somewhat derived rather than raw input variables. In "Hydrological concept formation inside long short-term memory (LSTM) networks" (Lees et al., 2022)for example, the authors use snow depth as a "diagnostic probe" to show that neural network models create physically plausible representations of the world based on "raw" input variables.

References

CopernicusClimateChangeService: ERA5-Land Hourly Data from 1950 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/Cds.E2161bac (Accessed on 23-Oct-2021), 2022.

den Bossche, J. V., Jordahl, K., Fleischmann, M., Richards, M., McBride, J., Wasserman, J., Badaracco, A. G., Snow, A. D., Roggemans, P., Ward, B., Tratner, J., Gerard, J., Perry, M., Taves, M., Hjelle, G. A., carsonfarmer, Tan, N. Y., Bell, R., ter Hoeven, E., Caria, G., Cochran, M. D., rraymondgh, Culbertson, L., Bartos, M., Chai, C. P., Eubank, N., sangarshanan, Flavin, J., and Rey, S.: Geopandas/Geopandas: V1.1.2, Zenodo, https://doi.org/10.5281/zenodo.18024156, 2025.

Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., and Dadson, S. J.: Hydrological Concept Formation inside Long Short-Term Memory (LSTM) Networks, Hydrology and Earth System Sciences, 26, 3079–3101, https://doi.org/10.5194/hess-26-3079-2022, 2022.

Mapbox: Rasterio v1.4.3, Mapbox, 2024.

Muñoz Sabater, J.: ERA5-Land Hourly Data from 1950 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/Cds.E2161bac (Accessed on 17-Sep-2024), 2019.

Snow, A. D., Cochran, M., Miara, I., Hoese, D., den Bossche, J. V., Mayo, C., Lucas, G., Cochrane, P., de Kloe, J., Karney, C., Shaw, J. J., Anh, T. Q., Filipe, Ouzounoudis, G., Couwenberg, B., Lostis, G., Dearing, J., Jurd, B., Gohlke, C., Schneck, C., McDonald, D., Taves, M., Itkin, M., May, R., Stewart, A. J., de Bittencourt, H., Little, B., Hugonnet, R., and Rahul, P. S.: Pyproj4/Pyproj: 3.7.2rc1, Zenodo, https://doi.org/10.5281/zenodo.16817340, 2025.