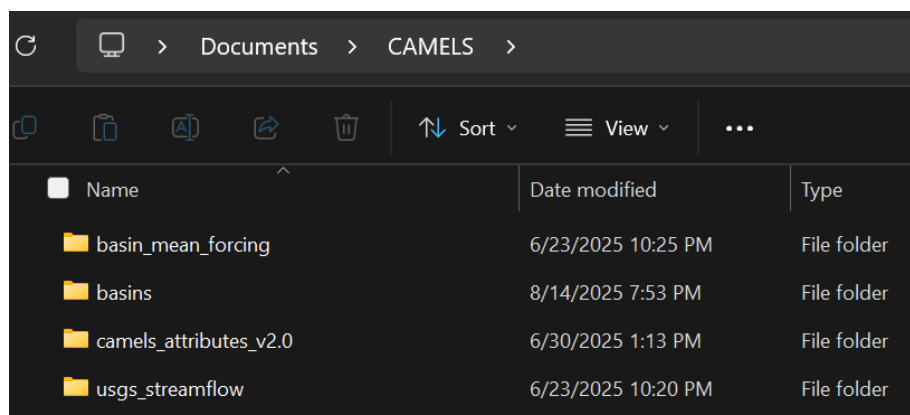


# Running Kratzert's Model on CAMELS

I started off by downloading the CAMELS dataset from the NSF website at <https://gdex.ucar.edu/dataset/camels/file.html>. The most important files for this project are

- basin\_timeseries\_v1p2\_metForcing\_obsFlow.zip
- camels\_clim.txt
- camels\_geol.txt
- camels\_hydro.txt
- camels\_name.txt
- camels\_soil.txt
- camels\_topo.txt
- camels\_vege.txt

and other files like `camels_attributes_v2.0.pdf` and `readme.txt` are helpful for understanding attributes. I created a folder called `CAMELS` to store the dataset in. After extracting `basin_timeseries_v1p2_metForcing_obsFlow.zip`, I moved the `basin_mean_forcing` and `usgs_streamflow` folders (from `basin_dataset_public_v1p2`) into the `CAMELS` folder. I created a folder within `CAMELS` called `basins` (for later), then another folder called `camels_attributes_v2.0` and moved all seven `camels_****.txt` (static attributes) files into `camels_attributes_v2.0`. Eventually, the main directory looked like this:



Next, I downloaded a NeuralHydrology model that worked for the CAMELS dataset. I decided to use the regional models mentioned in “HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin” (Kratzert, Gauch, Klotz, and Nearing). For this, I needed `basin_list.txt`, a list of the 531 basins their model is trained on, which I have uploaded on [Google Drive](#). I stored it in `CAMELS/basins`. After downloading `never-paper.tar.gz` from <https://doi.org/10.5281/zenodo.10139248>, I extracted it and went to the subdirectory `single-basin-vs-regional-model/run_dirs/regional_model`, where there were several folders. Each folder contains the data for a pretrained model. I chose the first option, `531_basins_multi_forcings_temporal_split_ensemble_member_7_0510_195908`, copied it to a more accessible place (for example, within the `CAMELS` folder), and renamed the folder to `nh_model`. To ensure that the model could run on my system, I had to modify the `config.yml` file within the `nh_model`:

- `data_dir` was changed from `camels-data` to the path to my `CAMELS` folder (so in this case, `/Users/*****/Documents/CAMELS`)
- `device` was changed from `cuda:2` to `cpu` since my laptop didn't have a CUDA GPU
- `run_dir` (much further down in the file) was changed from the original (very long) path to the new run directory, i.e. `/Users/*****/Documents/CAMELS/nh-model`
- `test_basin_file`, `train_basin_file`, and `validation_basin_file` went from `/home/gsnearing/hypertuning_experiments/basin_lists/531_basin_list.txt` to the location of `basin_list.txt` on my machine, which was `/Users/*****/Documents/CAMELS/basins/basin_list.txt`
- `train_dir` was changed from its original long path to the `train_data` folder within the new run directory, `/Users/*****/Documents/CAMELS/nh-model/train_data`

To evaluate the model on CAMELS, I installed the NeuralHydrology library for Python. On a computer with Python installed, I ran `pip install neuralhydrology` (sometimes, `python -m pip install neuralhydrology` or `py -m pip install neuralhydrology` works instead). The full documentation is available at <https://neuralhydrology.readthedocs.io>, but since I only needed to evaluate a pretrained model I ran the command (NOTE: won't work yet) `nh-run evaluate --run-dir /Users/*****/Documents/CAMELS/nh-model` (the last argument would be wherever the pretrained model from `never-paper.tar.gz` was stored). The evaluation ran smoothly at first but crashed roughly a quarter of the way through.

This error was due to a few misformatted files within the CAMELS database. I debugged by modifying the `neuralhydrology/datasetzoo/camelsus.py` file within my locally installed NeuralHydrology library to print out which file it was reading as it ran. The culprit files were all located within the `CAMELS/basin_mean_forcing/maurer` folder, specifically at

- `03/02108000_lump_maurer_forcing_leap.txt`
- `09/05120500_lump_maurer_forcing_leap.txt`
- `15/09492400_lump_maurer_forcing_leap.txt`

Each file contains an incorrect header (in the fourth row) that leaves out the labels for the `Year`, `Mnth`, `Day`, and `Hr` columns, as well as missing some capitalization in the other columns. I replaced the faulty headers with the correct header that I found from other files: `Year Mnth Day Hr Day1(s) PRCP(mm/day) SRAD(W/m2) SWE(mm) Tmax(C) Tmin(C) Vp(Pa)`. I couldn't fit it on one line in this report, but make sure it is only one line in the `.txt` file. After fixing this, I was able to use `nh-run evaluate --run-dir /Users/*****/Documents/CAMELS/nh-model` with no issues.