README Snowcast Showdown by UltimateHydrology

Requirements needed for running submitted code

In the development we used Conda package manager, the other managers were not tested:

- Miniconda https://docs.conda.io/en/latest/miniconda.html
- conda version : 4.11.0
- Python 3.9.5.final.0

The attached file **environments.yml** contains all necessary libraries and dependencies, to install them please use:

```
conda update conda=4.11.0
```

```
conda env create --name ultimatehydrology python=3.9 --file=environments.yml
```

```
conda activate ultimatehydrology
```

Data sources used in training and inference

Most of the features (attributes) can be classified into 4 main groups according to the different data sources. All of them are used in the training and inference code.

1. Meteorological parameters and their aggregates for 3 time periods from atmospheric model High-Resolution Rapid Refresh (HRRR NOAA):

- Average for the day of the forecast
- Aggregates by weekly period (week before the date of the forecast)
- Aggregates by winter period (from December of the previous year)

The High-Resolution Rapid Refresh (HRRR) is a NOAA real-time 3km resolution, hourly updated, cloud-resolving, convection-allowing atmospheric model, initialized by 3km grids with 3km radar assimilation.

Data is updated weekly for historical and future predictions. We connected to the Approved data access location: https://noaahrrr.blob.core.windows.net/hrrr .

List of created features:
Daily period:

Name in the dataset	Description	Unit
si10	daily mean wind speed	m/s
dswrf	daily mean radiation	W/m²
t2m	daily mean temperature	°C
tp	daily sum precipitation	mm
tp_pls	daily sum of liquid precipitation	mm
tp_mns	daily sum of solid precipitation	mm
t2m_pls	daily sum of warm temperature	°C
t2m_mns	daily sum of cold temperature	°C
rain_nrg	daily sum of t2m_pls*tp_pls	mm · °C

Weekly period:

Name in the dataset	Description	Unit
temp_mean	mean of temperature over the last 7 days	°C
temp_sum	sum of temperature over the last 7 days	°C
temp_sum_cold	sum of cold temperature over the last 7 days	°C
temp_sum_warm	sum of warm temperature over the last 7 days	°C
temp_sum_cold_hours	count of cold hours over the last 7 days	°C
temp_sum_warm_hours	count of warm hours over the last 7 days	°C
tp_mean	mean of precipitation over the last 7 days	mm
tp_sum	sum of precipitation over the last 7 days	mm
tp_sum_liquid	sum of liquid precipitation over the last 7 days	mm
tp_sum_solid	sum of solid precipitation over the last 7 days	mm
rain_enrg	last 7 days sum of t2m_pls*tp_pls	mm · °C
thaw_count	count of thaws over the last 7 days	dimensionless
dswrf_mean	mean of radiation over the last 7 days	W/m²
dswrf_sum	sum of radiation over the last 7 days	W/m²
si10_mean	mean of wind speed over the last 7 days	m/s
si10_sum	sum of wind speed over the last 7 days	m/s
sdwe_mean	mean of swe over the last 7 days	mm
sdwe_sum	sum of swe over the last 7 days	mm
sdwe_range	delta swe per week	mm
sdwe_last	swe 7 days ago	mm
sdwe_first	swe for real-time evaluation date	mm

Winter period:

Name in the dataset	Description	Unit
si10_cumsum	cumsum of wind speed from the beginning of winter	m/s
si10_mean_sws	cummean of wind speed from the beginning of winter	m/s
dswrf_cumsum	cumsum of radiation from the beginning of winter	W/m²
dswrf_mean_sws	cummean of radiation from the beginning of winter	W/m²
t2m_cumsum	cumsum of temperature from the beginning of winter	°C
t2m_mean_sws	cummean of temperature from the beginning of winter	°C
tp_cumsum	cumsum of precipitation from the beginning of winter	mm
tp_mean_sws	cummean of precipitation from the beginning of winter	mm
tp_pls_cumsum	cumsum of liquid precipitation from the beginning of winter	mm
tp_pls_mean_sws	cummean of liquid precipitation from the beginning of winter	mm
tp_mns_cumsum	cumsum of solid precipitation from the beginning of winter	mm
tp_mns_mean_sws	cummean of solid precipitation from the beginning of winter	mm
t2m_pls_cumsum	cumsum of warm temperature from the beginning of winter	°C
t2m_pls_mean_sws	cummean of warm temperature from the beginning of winter	°C

Name in the dataset	Description	Unit
t2m_mns_cumsum	cumsum of cold temperature from the beginning of winter	°C
t2m_mns_mean_sws	cummean of cold temperature from the beginning of winter	°C
rain_nrg_cumsum	cumsum of rain_nrg from the beginning of winter	mm · °C
rain_nrg_mean_sws	cummean of rain_nrg from the beginning of winter	mm · °C
si10_m7	daily mean wind speed 7 days ago	m/s
dswrf_m7	daily mean radiation 7 days ago	W/m²
t2m_m7	daily mean temperature 7 days ago	°C
tp_m7	daily sum precipitation 7 days ago	mm
tp_pls_m7	daily sum of liquid precipitation 7 days ago	mm
tp_mns_m7	daily sum of solid precipitation 7 days ago	mm
t2m_pls_m7	daily sum of warm temperature 7 days ago	°C
t2m_mns_m7	daily sum of cold temperature 7 days ago	°C
rain_nrg_m7	daily sum of t2m_pls*tp_pls 7 days ago	mm · °C
si10_cumsum_m7	cumsum of wind speed from the beginning of winter 7 days ago	m/s
si10_mean_sws_m7	cummean of wind speed from the beginning of winter 7 days ago	m/s
dswrf_cumsum_m7	cumsum of radiation from the beginning of winter 7 days ago	W/m²
dswrf_mean_sws_m7	cummean of radiation from the beginning of winter 7 days ago	W/m²
t2m_cumsum_m7	cumsum of temperature from the beginning of winter 7 days ago	°C
t2m_mean_sws_m7	cummean of temperature from the beginning of winter 7 days ago	°C
tp_cumsum_m7	cumsum of precipitation from the beginning of winter 7 days ago	mm
tp_mean_sws_m7	cummean of precipitation from the beginning of winter 7 days ago	mm
tp_pls_cumsum_m7	cumsum of liquid precipitation from the beginning of winter 7 days ago	mm
tp_pls_mean_sws_m7	cummean of liquid precipitation from the beginning of winter 7 days ago	mm
tp_mns_cumsum_m7	cumsum of solid precipitation from the beginning of winter 7 days ago	mm
tp_mns_mean_sws_m7	cummean of solid precipitation from the beginning of winter 7 days ago	mm
t2m_pls_cumsum_m7	cumsum of warm temperature from the beginning of winter 7 days ago	°C
t2m_pls_mean_sws_m7	cummean of warm temperature from the beginning of winter 7 days ago	°C
t2m_mns_cumsum_m7	cumsum of cold temperature from the beginning of winter 7 days ago	°C
t2m_mns_mean_sws_m7	cummean of cold temperature from the beginning of winter 7 days ago	°C
rain_nrg_cumsum_m7	cumsum of rain_nrg from the beginning of winter 7 days ago	mm · °C

Name in the dataset	Description	Unit
rain_nrg_mean_sws_m7	cummean of rain_nrg from the beginning of winter 7 days ago	mm · °C

2. MODIS Terra MOD10A1 satellite imagery product (Snow Cover Daily L3 Global 500m SIN Grid).

The snow cover algorithm calculates NDSI for all land and inland water pixels in daylight using MODIS band 4 (visible green) and band 6 (shortwave near-infrared). We used three main features:

Name in the product	Name in the dataset
NDSI_Snow_Cover	sc
NDSI	ndsi1
Snow_Albedo_Daily_Tile	sa1

Detailed information and descriptions of the values are given here https://nsidc.org/data/MOD10A1.

Data is updated daily, but we use only the necessary dates. We connected to Approved data access location: https://modissa.blob.core.windows.net/modis-006 .

3. Relief parameters, produced on Copernicus DEM (90 meter resolution).

Relief parameters were calculated by usage of Copernicus DEM for each grid cell, as well as their aggregates in 200 and 500 m geodetic buffers (min, max, average, median) with the opensource libraries (GDAL, RichDEM). These attributes were calculated in advance for the whole grid and do not change over time.
We used Approved data access location:
https://planetarycomputer.microsoft.com/api/stac/v1/collections/cop-dem-glo-90

List of created features:

Name in the dataset	Description
alt * alt_min_200 alt_max_200 alt_mean_200 alt_median_200 alt_min_500 alt_max_500 alt_mean_500 alt_median_500	**Height, meters** alt_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters.
slope * slope_mean_200 slope_median_200 slope_mean_500 slope_median_500	**Slope, degrees** slope_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters. Without suffix value at point. https://richdem.readthedocs.io/en/latest/python_api.html
aspect * aspect_mean_200 aspect_median_200 aspect_mean_500 aspect_median_500	**Aspect** aspect_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters. https://richdem.readthedocs.io/en/latest/python_api.html
curv_prof * curv_prof_mean_200, curv_prof_median_200 curv_prof_mean_500 curv_prof_median_500	**Profile curvature** curv_prof_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters. https://richdem.readthedocs.io/en/latest/python_api.html

Name in the dataset	Description
curv * curv_mean_200 curv_median_200 curv_mean_500 curv_median_500	**Curvature** curv_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters. https://richdem.readthedocs.io/en/latest/python_api.html
curv_plan * curv_plan_mean_200 curv_plan_median_200 curv_plan_mean_500 curv_plan_median_500	**Planform curvature** curv_plan_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters. https://richdem.readthedocs.io/en/latest/python_api.html
tri * tri_mean_200 tri_median_200 tri_mean_500 tri_median_500	**TRI (Terrain Ruggedness Index)** tri_{agg}_{buffer}, where {agg} is aggregate function (min, max, mean, median), {buffer} is buffer radius in meters. https://gdal.org/programs/gdaldem.html

* without suffix -- value at the cell centroid

4. Additional sources:

- Coordinates of cell centroids
- Height in the cell centroid
- Ordinal date from the beginning of the year (Day of year)
- Spatial-temporal interpolation based on Ground measure data (SNOTEL, CDEC) by Random Forest model. 4 parameters are used:
    - lat
    - lon
    - alt
    - time (day_of_year)

Instructions for running inference, using the inference source code and model weights above

We provide a main point of entry to our code as the Jupyter Notebook that runs all steps of the pipeline to run inference source code and model weights.

Brief description of the model. The solution uses gradient boosting models and their stacking.

1. First level models:

- XGBRegressor – CV Score 5 Folds RMSE = 3.75
- CatBoostRegressor – CV 5 Folds Score RMSE = 3.65
- LGBMRegressor – CV 5 Folds Score RMSE = 3.78

2. Second level model (meta-model):

- XGBRegressor – CV 5 Folds Score RMSE = 3.49

Running the model

To run the weekly forecast you just need to run **main_runner.ipynb** in the created environment.

```
ipython kernel install --user --name= ultimatehydrology
```

Running the steps of model:

1. Downloading and processing the meteorological data for the previous week
2. Downloading and processing of MODIS space images for the previous week
3. Data collection from new (meteo, space images) and stable(relief, etc.) resources
4. Loading model weights from the pretrained model
5. Making prediction and preparing a weekly submission for appropriate week.

Instructions for training the model from scratch using the training source code above

To run the training model pipeline you just need to run all cells in **train_model.ipynb** in the created environment.