Coding Challenge: Hydropower Generation on the Columbia River

In the Pacific Northwest of the United States, the Bonneville Power Authority (BPA) operates a number of large hydropower facilities along the Columbia River and tributaries to the river. Total generation from these projects is published along with discharge data. **Discharge** is the volume of water per unit time flowing through the hydropower project. Most of the reservoirs are run-of-river, while a few have reservoirs upstream.

The data from BPA can be delayed, but the Unites States Geologic Survey (USGS) published discharge data at several locations along the river in real time. Using this data, the challenge is to develop a tool to predict generation from the observed USGS observations.

Figure 1 Columbia River Projects

Goal

Below you are provided links to published data from the BPA, USACE, and the USGS. The data is Total System Generation data, or historic data of discharge and generation at the hydropower projects, along with discharge from a several USGS stations. Using this data, you are to write a program to forecast daily generation from the Columbia River. It is up to you to decide how to use the datasets to train and validate your model. Assume that the datasets you collect would be available if the model is put into production.

There are many hydropower projects along the Columbia River, and data is available for individual projects if you wish to provide more granularity in your analysis. However, you need not run the analysis for all. At a minimum, develop a tool to forecast **Columbia River Hydropower Generation** based on observed **Discharge** at The Dalles, Oregon.

Data

There are three key data providers. The Bonneville Power Authority (BPA) provides Total System Generation data. However, note this data includes generation from other parts of the region the are **not** included in the Columbia River. The United States Army Corps of Engineers (USACE) runs most of the large and more significant hydropower projects. They publish data for each of the projects directly. The United States Geologic Survey (USGS) provides hydrologic data.

BPA System Generation

Total BPA system generation is published at the following page:

https://transmission.bpa.gov/Business/Operations/Wind/default.aspx

Item #4 shows the generation over the last 7 days.

Item #5 provides links to historic generation in separate files for each year.

NOTE: *the published Total System Generation includes several projects not on the Columbia River.*

Although on this page the data is updated every 5 minutes, for the purposes of this exercise, assume it is delayed.

USACE Hydropower Project Data

Monthly power summaries are published at this link:
https://www.nwd-wc.usace.army.mil/ftppub/power/

TIP: You can use the Monthly Power summaries to scale Total System Generation.

More granular data for some locations is also available via the **USACE DataQuery 2.0 Portal** if you require.

Relevant dams/projects on the Columbia River include those below, others are not included in Columbia River generation. The following should account for a significant majority of the generation.

BPA Project Name	USACE ID / DataQuery 2.0 ID
Dworshak	DWR
Albeni Falls	ALF
Bonneville	BON
John Day	JDA
McNary	MCN
The Dalles	TDA
Chief Josepth	CHJ
Ice Harbor	IHR
Little Goose	LGS
Lower Monumental	LWM
Libby	LIB
Lower Granite	LGL
Grand Coulee	GCL
Hungry Horse	HGH

USGS Discharge

The USGS has a National Water Dashboard. The link for The Dalles is:

https://dashboard.waterdata.usgs.gov/app/nwd/en/?aoi=usgs-14105700

The Dalles station ID is: 14105700.

Once the page is loaded, clicking on the green circular icon for the station will pop-up a new page with links to data (circled in red in Figure 2).

Click this link to go to the data page.

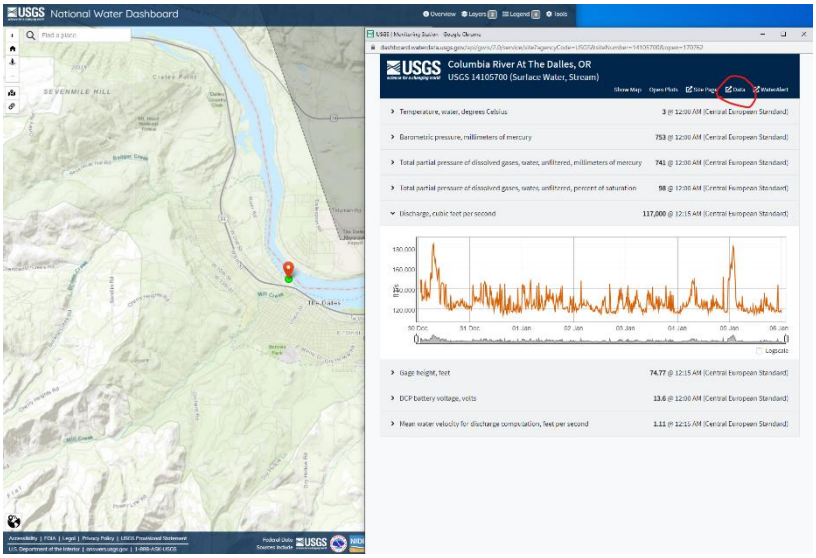

Figure 2 National Water Dashboard

Once on the data page, select the ‘Current / Historical Observations’ link (circled in red in Figure 3) which will load another data selection page for historical and real-time data. You can use the ‘Change time span’ feature to pull longer histories.

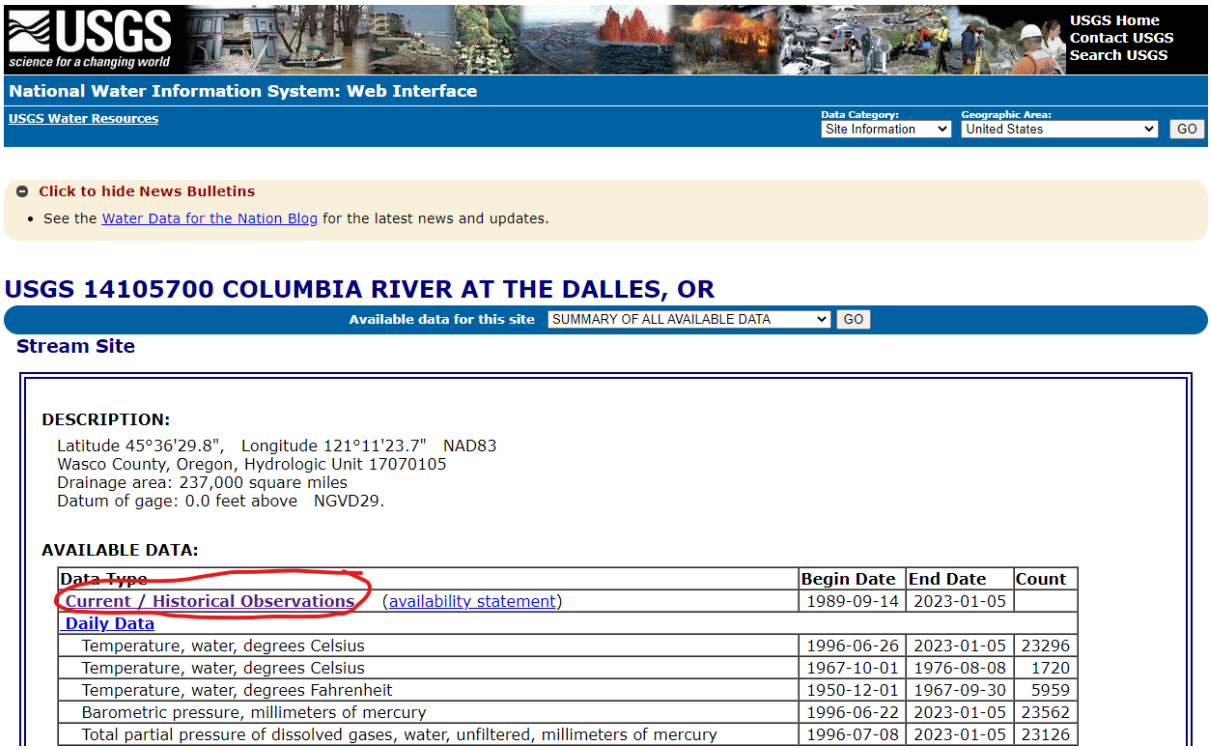

Figure 3 USGS Water Data Page

TIP:

Knowing the Station ID, one can use the following URL structure to download the data directly where the parameter **&site_no=14105700** is the Station ID:

https://nwis.waterdata.usgs.gov/usa/nwis/uv/?cb_00060=on&cb_00065=on&format=rdb**&site_no=14105700**&legacy=1&period=&begin_date=2013-12-29&end_date=2023-01-01

Station IDs may be retrieved by zooming out on the National Water Dashboard to find other relevant Streamflow / Discharge sites.

Deliverable and Evaluation

You will provide your code in a format ready to be run (please do not post the challenge to an open repository). A private repository or an archive attached to email are simple approaches. The solution should provide instructions for running the code.

Datasets used for training and evaluation of the model should be included in the repository or archive. Do not provide code that retrieves data as part of the evaluation, though if you have written 'Data Scraping' code, it could be provided to be assessed as part of the evaluation.

Additionally, please provide a write-up describing your approach, issues you encountered, and anything that may be relevant regarding the analysis. Consider this as a 'prototype' for something that may be put into operation. What issues would you raise in advance?

The analysis is to be evaluated on four metrics:

- Approach – is the choice of model correct? Are the results skillfull?
- Comprehensive – does the solution address the main challenge?
- Communication – does the write-up communicate the findings succinctly and adequately?
- Coding – is the provided solution well written, maintainable, and able to be put into operation?

Write-up and Instructions

You should provide instructions for running you code and provide references to any non-standard dependencies that may be required.

The primary development languages at Statkraft are C++, C#, and Python -- with the latter being the main software used on our desks for analytics. You may use any language of your choosing, but it will be helpful in the evaluation if you use Python to demonstrate your own familiarity.

Please remember, the evaluation is not only about the solution or performance of the model you develop, but also provides an opportunity for you to demonstrate the way you structure and write code, use of tests, and general coding standards.

In your findings, feel free to identify what you see as gaps or shortcomings in your analysis, and what you would do to address this in a more operational environment.