Team: Pow

Problem Space: We would like to predict Snow Water Equivalent (SWE) as a proxy for drought and flood risk. As a metric, SWE is used by hydrologists, policy-makers, and others to estimate the availability of water resources, an important input in flood planning, reservoir management, and much more. We plan to take two different approaches to predicting SWE, linear regression and time series forecasting, and compare the results.

Dataset: Natural Resources Conservation Service (NRCS), US Dept. of Agriculture Air and Water Database (AWDB), including Snow Telemetry (SNOWTEL).

Number of entries: ~730 stations, located across 11 western states

Number of features per entry: ~105 available features/station (feature data). We plan to use ~5 or 6 features in our training data.

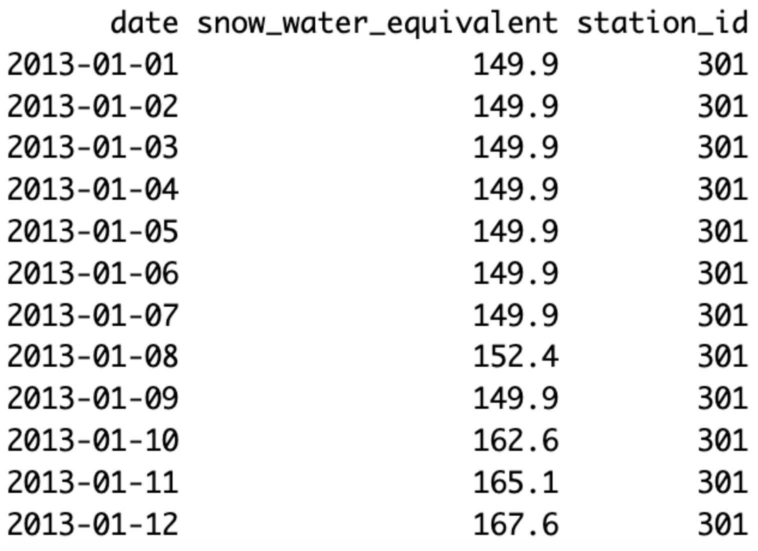
Description of some of the most interesting or important features:

- ❖ **Multiple linear regression model:**
  - ➢ Air Temperature Observed (TOBS) (-avg/-min/-max)
  - ➢ Ground Surface Interface Temperature (TGSI) (-tgsv/-tgsx/-tgsn)
  - ➢ Net Solar Radiation (NTRDC) (-ntrdv/-ntrdx/-ntrdn)
  - ➢ Precipitation Accumulation (PREC) (-prcp)
  - ➢ Snow Fall (SNOW)
  - ➢ Soil Temperature (STO) (-stv/-stx/-stn)
  - ➢ Solar Radiation Total (SRADT) (-adv/-adx/-adn)
- ❖ **Time series forecasting models**:
  - ➢ Snow Water Equivalent

Variable(s) you've chosen as your outcomes:

- ❖ **Multiple linear regression model:**
  - ➢ Snow Depth Average (SNWD) (-snwdv/-snwdx/-snwdn)
  - ➢ Snow Water Equivalent (WTEQ) (-wteqv/-wteqx/-wteqn)
- ❖ **Time series forecasting models:**
  - ➢ Snow Water Equivalent

First few rows: The feature data we are interested in for the multiple linear regression model are behind a SOAP API. We will need to write a script to scrape this data, and we will put it, perhaps, into dataframes with `pandas` (or similar). For the time series models we will only need the snow water equivalent data from each of the stations. The first few rows of that data set are pictured below. As you can see there’s a row for each day at each station.



```
         date snow_water_equivalent station_id
2013-01-01                   149.9        301
2013-01-02                   149.9        301
2013-01-03                   149.9        301
2013-01-04                   149.9        301
2013-01-05                   149.9        301
2013-01-06                   149.9        301
2013-01-07                   149.9        301
2013-01-08                   152.4        301
2013-01-09                   149.9        301
2013-01-10                   162.6        301
2013-01-11                   165.1        301
2013-01-12                   167.6        301
```

Methodology:

Proposed Model:

- ARIMA and Exponential smoothing for time series data
    - ARIMA uses a weighted linear sum of past observations and exponential smoothing uses an exponentially decreasing weighted sum of past observations.
- Multiple Linear Regression for feature data.

How our idea differs from / builds on ideas from the research field: We would like to potentially open the selection of features up and allow the model to choose a set of features and weights that minimize its mean square error on new (unseen) data. Our methods may not forge new frontiers in our understanding of hydrology, but it does promise a well-defined exercise for practicing our hand at machine-learning techniques, and if we can have the model optimize itself without overfitting that would be neat, too. Plus, this is an interesting problem and one of key importance to states like Colorado.

Stretch Goals: In addition to the two time series models already proposed, we could additionally try to use recurrent neural networks as those are also said to be good for time series forecasting.

Other Projects we Drew inspiration from:

https://github.com/drivendataorg/snowcast-showdown/blob/main/2nd%20Place/reports/Model%20report%20UltimateHydrology.pdf

https://github.com/hanis-z/Snow-water-equivalent/blob/main/README.md