Assignment 1 Check In
design the forecast system approach; get and process the data

B Steele

Progress:

- well-defined problem

- functional preliminary NN architecture to work from

Challenges:

- Forecasting requires acquisition and integration of NOAA GEFS forecasts, since many of the important features in the preliminary models are weather-related.

- How to represent spread in model? Use MAE and assume Gaussian distribution as a seed?

Scientific Motivation and Problem Statement

Water temperature is often an indicator of water quality, as it governs much of the biological activity in freshwater systems. While temperature is an important parameter to monitor in freshwater lakes, manual monitoring of waterbodies (by physically visiting a site) and sensor network deployment to monitor water temperature, are costly endeavors. Northern Water, the municipal subdistrict that delivers drinking water to approximately 1 million people in northern Colorado and irrigation water for ~600,000 acres of land, has had recurring issues with water clarity in Grand Lake, the deepest natural lake in Colorado. They believe that the clarity issues in Grand Lake are primarily due to algal and diatom growth in Shadow Mountain reservoir which are pushed into Grand when they initiate pumping operations. The regulatory period for clarity in Grand Lake is July 1 - September 11.

Water in the Three Lakes System naturally flows from Grand into Shadow Mountain into Granby, but pumping operations reverse that by pumping cold water from Granby reservoir into Shadow Mountain and then into Grand and into the tunnel to serve the Front Range (Figure 1). Northern suspects there is a biological "sweet spot" for water temperature in Shadow Mountain Reservoir that may reduce algal and diatom growth and therefore mitigate clarity impacts during pumping operations.

Turns out the optimal temperature for reducing algal growth is to keep the upper 1m of water less than 15°C and to reduce diatom growth is to keep the average temperature of 0-5m ("integrated depth") greater than 14°C, which is a bit of a "Goldilocks" problem. A preliminary auto-regressive neural network model predicted tomorrow's water temperature reliably and better than or similar to a persistence model for the regulatory season during 2022 (Figure 2, Figure 3). Using SHAP analysis (a method of explainable AI for neural networks) we found that operational pumping has an impact on tomorrow's temperature (Figure 4, Figure 5), though the impact of operations on the integrated depth is stronger than the upper 1m.

The overarching goal is to create a decision support system that forecasts water temperature in Shadow Mountain Reservoir on a daily timestep to a horizon of seven days, since the operations of the pump over the previous seven days are the most influential operations variable from this preliminary model. Initially, this application will assume a constant operational pumping regime (where the previous day's pumping is continued throughout the forecast horizon), but the intention is to eventually add a 'knob' that would alter pumping operations as a mechanism to mitigate water temperature and attempt to reach the "Goldilocks" zone during the regulatory period. Adding that knob is likely out of scope for this class.

Description of the Data

- buoy data
- met data
- inflow/outflow data
- pumping operations

Description of pre-processing

- description of any data pre-processing performed and why you did it

Data split

Must account for timeseries, so split by regulatory year. When developing preliminary model, I intentionally held out 2023, assuming I'd use it for forecasting. There are discrete gaps in data over the winter, so I can assert some independence between each of the years of data.

- training/validation will be leave one year out for years 2014-2022, creating an ensemble prediction
- testing: 2023

Description of the forecast approach

- detailed description of the data-driven forecast approach (e.g. the method)
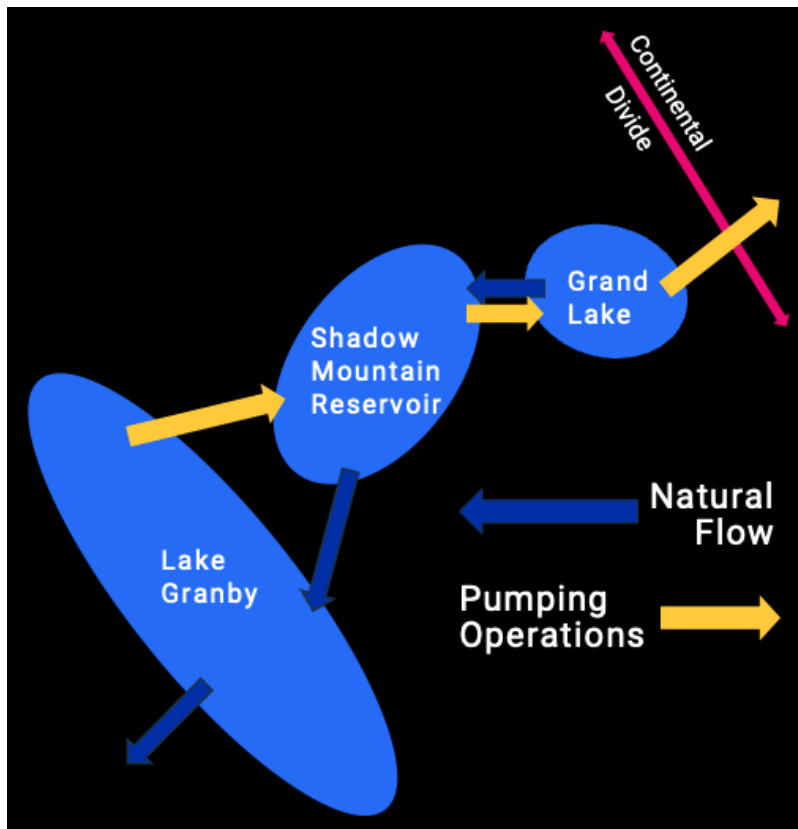

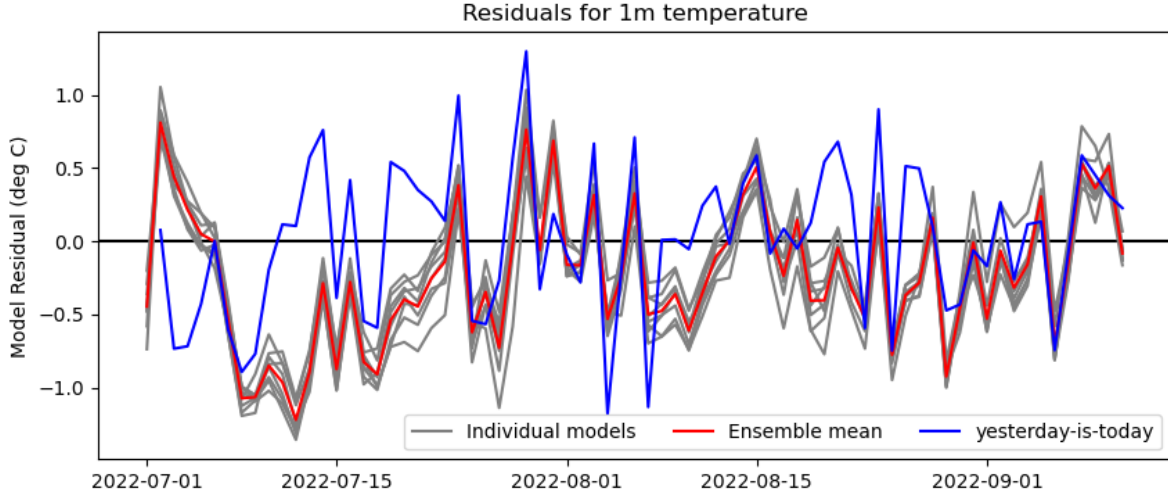Figure 1: Cartoon schematic of water flow in the Three Lakes System

Figure 2: Residuals at upper 1m of auto-regressive neural network ensemble (grey/red) and persistence model (yesterday is today). Initial model performance for the test set was slightly worse than persistence with a MSE of 0.27°C (persistence 0.24°C), MAE of 0.43°C (persistence 0.40°C), MAPE: 2.49% (not calculated for persistence).
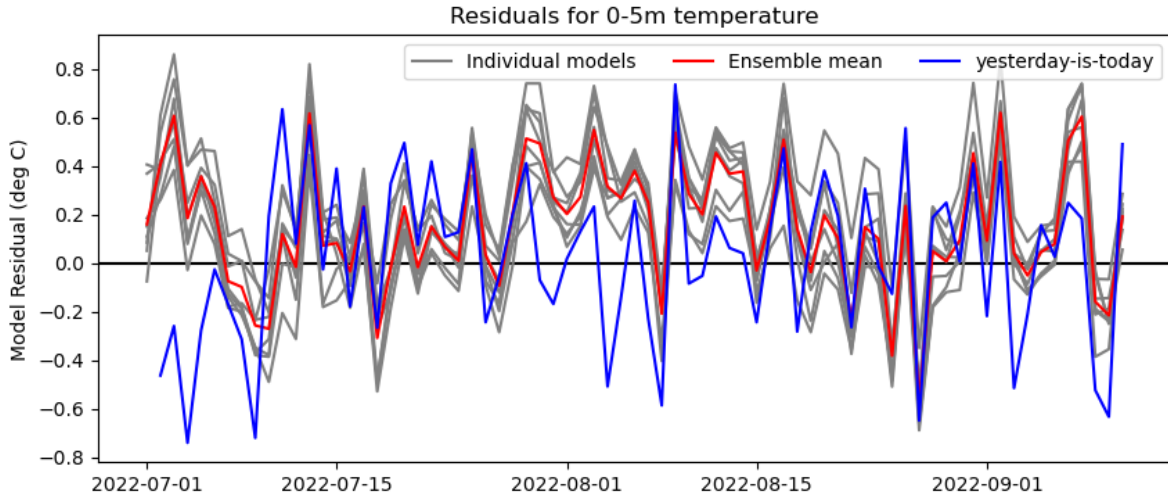

Figure 3: Residuals at the integrated depth of auto-regressive neural network ensemble (grey/red) and persistence model (blue). Initial model performance for the test set was better than persistence with a MSE of 0.09°C (persistence 0.11°C), MAE of 0.24°C (persistence 0.24°C), MAPE: 1.73% (not calculated for persistence).
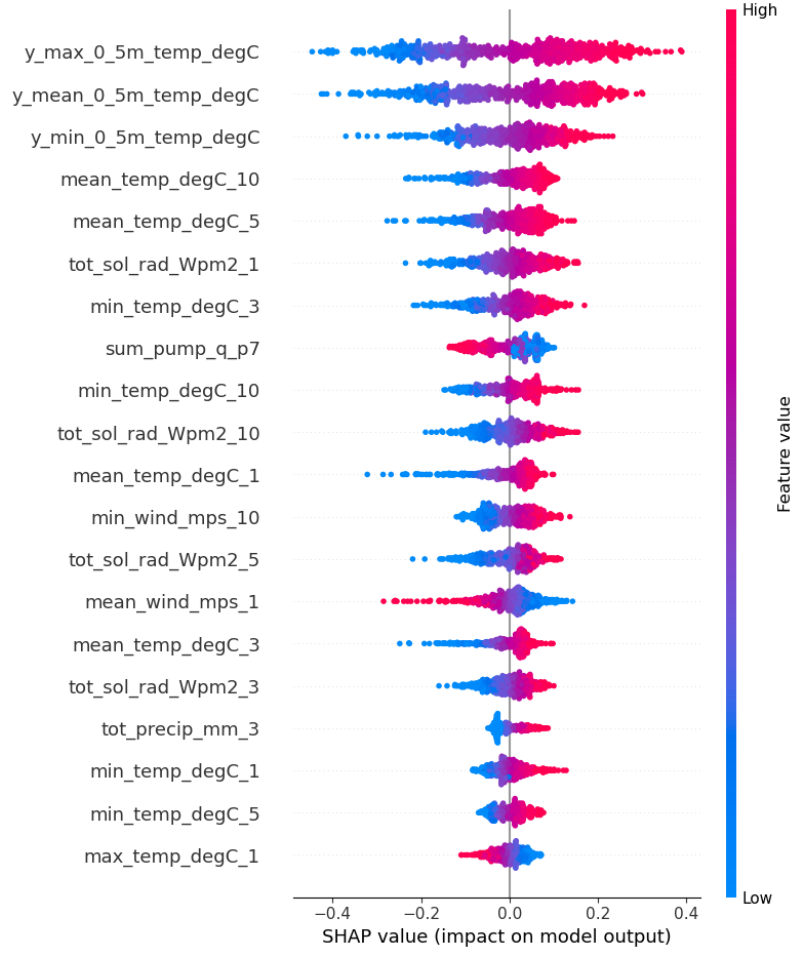
Figure 4: SHAP analysis for predicting the top 1m water temperature at Shadow Mountain from a fully-connected, auto-regressive neural network. Note "sum_pump_q_p7" (sum of pumping volume over the previous seven days), which indicates some sensitivity of near-surface temperature to pumping operations.
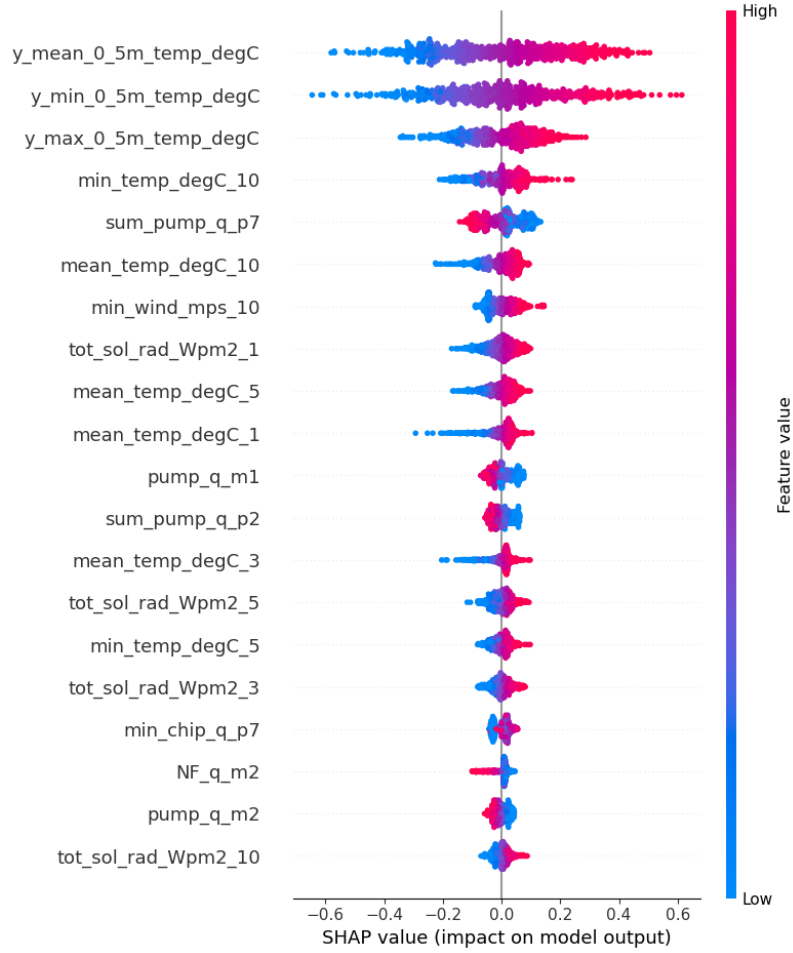
Figure 5: SHAP analysis for predicting the average water temperature (0-5m) at Shadow Mountain from a fully-connected, auto-regressive neural network. Note "sum_pump_q_p7" (sum of pumping volume over the previous seven days), which indicates a relatively strong response in predicted integrated depth water temperature to pumping operations.