Data-Driven Forecasting: Assignment 2

prototype of the forecast system; initial evaluation of the data with baselines for comparison

B Steele

GitHub link

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. Despite the importance of water quality to determine water-system health, consistent and frequent monitoring of waterbodies (by physically visiting a site) or sensor network deployment to monitor water temperature are both costly endeavors (CITE). 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. Clarity in Grand is regulated by Senate Document 80 which dates back to 1937 and the inception of the Colorado Big-Thompson project, however in 2016 stakeholders and operators adopted a system of "goal qualifiers" for Grand. The goal qualifiers are defined through Secchi disc depth measurements (a measure of water clarity), requiring a 3.8-meter Secchi depth average and 2.5-meter Secchi depth daily minimum to be met throughout the July 1 to September 11 regulatory season.

Water in the Three Lakes System naturally flows from Grand into Shadow Mountain into Granby, but pumping operations reverse that by pumping hypolimnetic water (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.

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.

Northern Water has collected extensive data at the Three Lakes System for many years. In 2015, they deployed an instrumented buoy in Shadow Mountain near Chipmunk Lane, the connection between Grand and Shadow Mountain. The data-driven forecast relies on aggregated daily data from the buoy, as well as volume data from inflows (North Fork into Grand Lake, North Inlet into Shadow Mountain) as well as the interflow between Shadow and Grand (Chipmunk Lane). In addition to the hydrologic data, I use summaries of meteorology from a met station at the southern end of Shadow Mountain, which is gap-filled with NCEI data from a local weather station on the north side of Grand Lake, which I assert are similar enough to be interchangable.

An operational 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, CITE) 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. This leads me to believe that the operations could be used as a "knob" to control water temperature to some extent in the Three Lakes System. Further, setting the pumping operations to zero (but not accounting for water balance within the system) shows sensitivity to pumping operations (Figure 6, Figure 7) through increased temperature for both outputs, which is expected.

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 the 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 an operational "knob" that would alter pumping operations as a mechanism to mitigate water temperature within the forecast application and attempt to reach the "Goldilocks" range during the regulatory period. Adding that knob is likely out of scope for this class, so instead, I will focus on reliable 7-day forecasts using an auto-regressive neural network.

Example Forecast

- an example forecast with the forecast system

  1) operational forecast: use actual pumping/inflow for future days/forecast outlook to show the skill added with those data

2) persistence forecast: use static pump/inflow data but otherwise rollout the forecast like operational - this will start to get at whether or not the 'pump' lever has any impact on our forecasts

3) null forecast: yesterday is today and tomorrow and the day after, but with noise

Metrics and baselines

- discussion of the metrics you will use to evaluate the forecasts
  - CRPS and forecast skill - e.g. Woelmer et al 2024 - $CRPS_o$ = *CRPS of operational forecast;* $CRPS_n$ = CRPS of null forecast (could also replace $CRPS_n$ with $CRPS_p$ - the CRPS of persistence model)

$$ForecastSkill = 1 - CRPS_o/CRPS_n$$

- discussion of the relevant baselines
  - persistence is probably the 'best' baseline. In order to actually assess skill at the pump operations, though, we need to address the value of including pumping operations

Forecast Evaluation

First step: assess the data-driven model's ability to predict tomorrow's temperature based on observed data, aka, can we even trust this?

Model	MSE 1m	MAE 1m	RMSE 1m	MAPE 1m	MSE 0-5m	MAE 0-5m	RMSE 0-5m	MAPE 0-5m
Operational Model (with today's met)	0.27	0.41	0.49	2%	0.08	0.24	0.30	2%
Persistence Baseline	0.24	0.40			0.11	0.28		

Operational forecast model performs slightly better than the preliminary model for the test year 2022 at 1mdepth, but not better than persistence in model development. This is to be expected, as 2022 had some of the highest water temperatures on record. Additionally, sensical that this improves the 1m predictions slightly as those are driven by met variables a bit more. [[I wanted to add some SHAP plots, but my SHAP code is broken… sooooo no SHAP here.]]

Next step: assess the rollout/forecast skill... but I need to roll it out first... and then will address:

- evaluation of the forecasts on a validation/testing set using chosen metrics
- comparison of forecast skill compared to the baselines
- discussion of whether the output is or is not "reasonable" in your judgment (e.g. physically consistent, overfitting, values within known physical ranges, ...)

Figures:

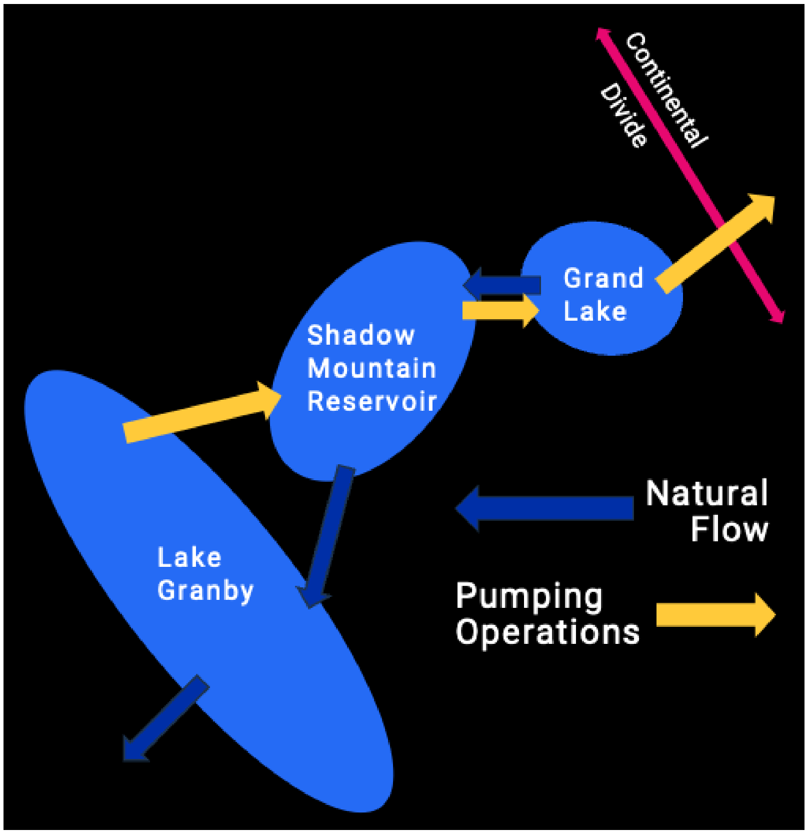

Figure 1: Cartoon schematic of water flow in the Three Lakes System

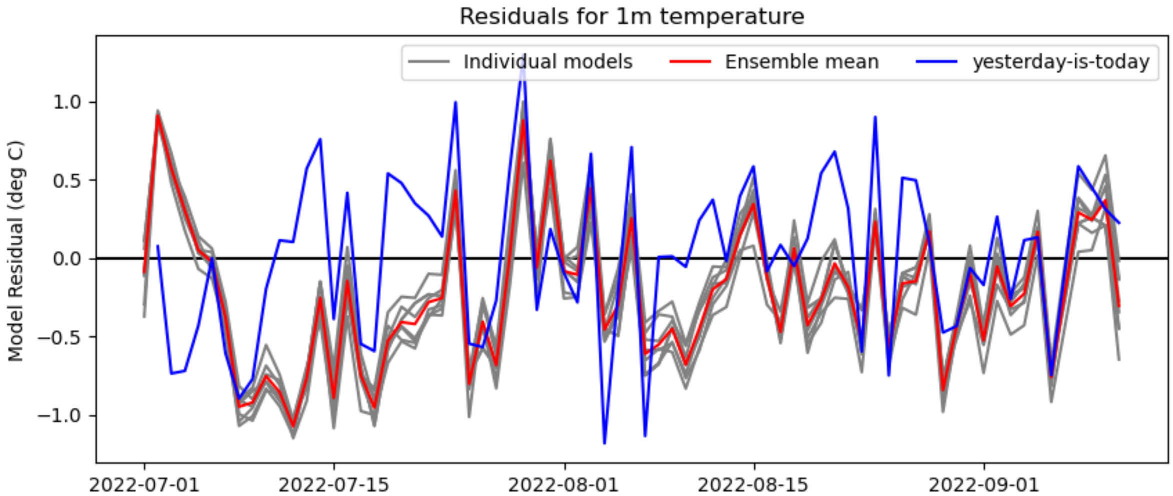

Figure 2: Residuals (predicted minus observed) at upper 1m of auto-regressive neural network ensemble (grey/red) and persistence model (yesterday is today, blue). Initial model performance for the test set was slightly worse than persistence with a MSE of 0.27°C (persistence 0.24°C), MAE of 0.41°C (persistence 0.40°C), MAPE: 2.49% (not calculated for persistence).

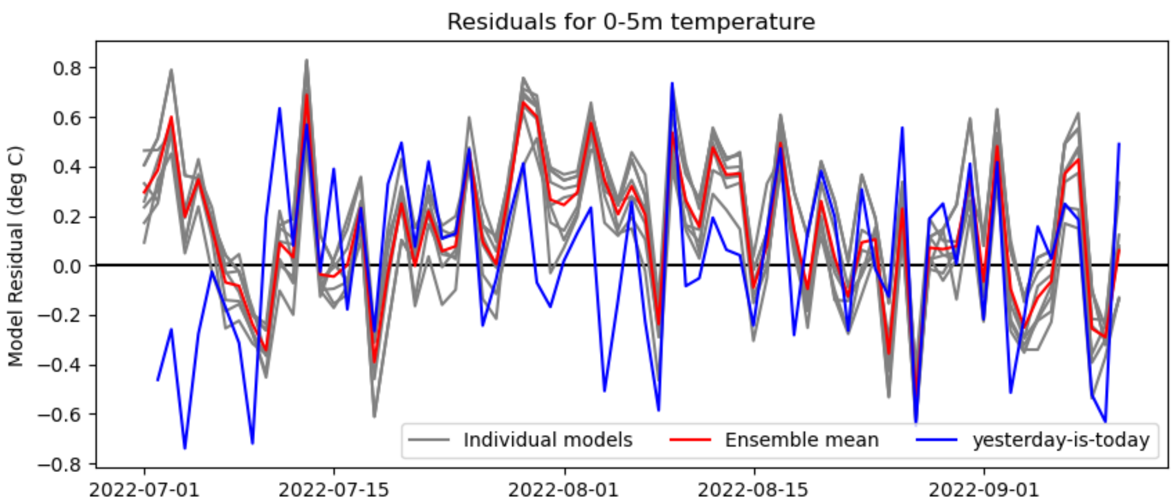

Figure 3: Residuals (predicted minus observed) at the integrated depth of auto-regressive neural network ensemble (grey/red) and persistence model (blue). Initial model performance for the test set was better than persistence with a MSE of 0.09°C (persistence 0.11°C), MAE of 0.24°C (persistence 0.24°C), MAPE: 1.73% (not calculated for persistence).

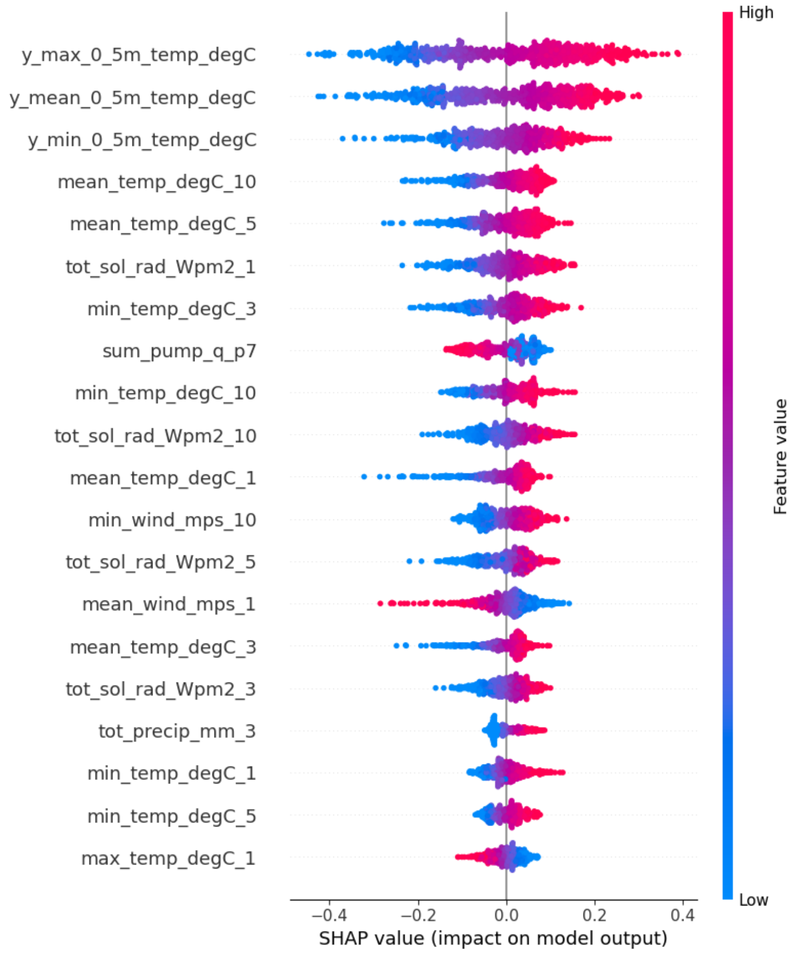

Figure 4: SHAP analysis for predicting the top 1m water temperature at Shadow Mountain from a fully-connected, auto-regressive neural network. Note “sum_pump_q_p7” (sum of pumping volume over the previous seven days), which indicates some sensitivity of near-surface temperature to pumping operations.

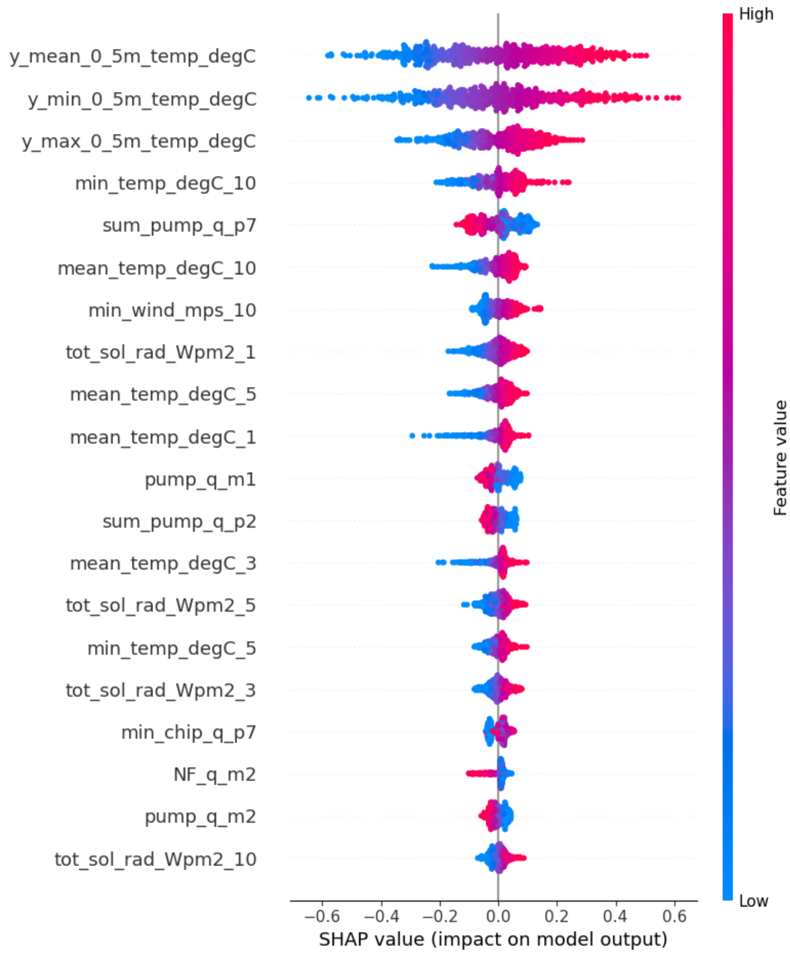

Figure 5: SHAP analysis for predicting the average water temperature (0-5m) at Shadow Mountain from a fully-connected, auto-regressive neural network. Note "sum_pump_q_p7" (sum of pumping volume over the previous seven days), which indicates a relatively strong response in predicted integrated depth water temperature to pumping operations.

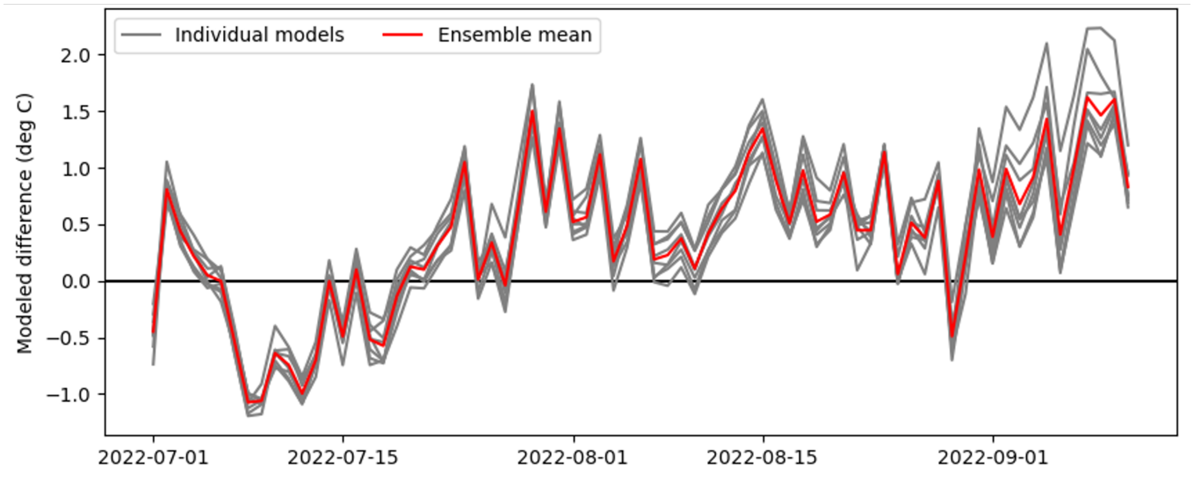

Figure 6: Residuals (predicted minus observed) of test set when pumping operations set to zero for water temperature at 0-1m. This indicates that the model is sensitive to pumping and that the result is sensical (if there is no introduction of cold water, the tempearture is predicted to be higher).

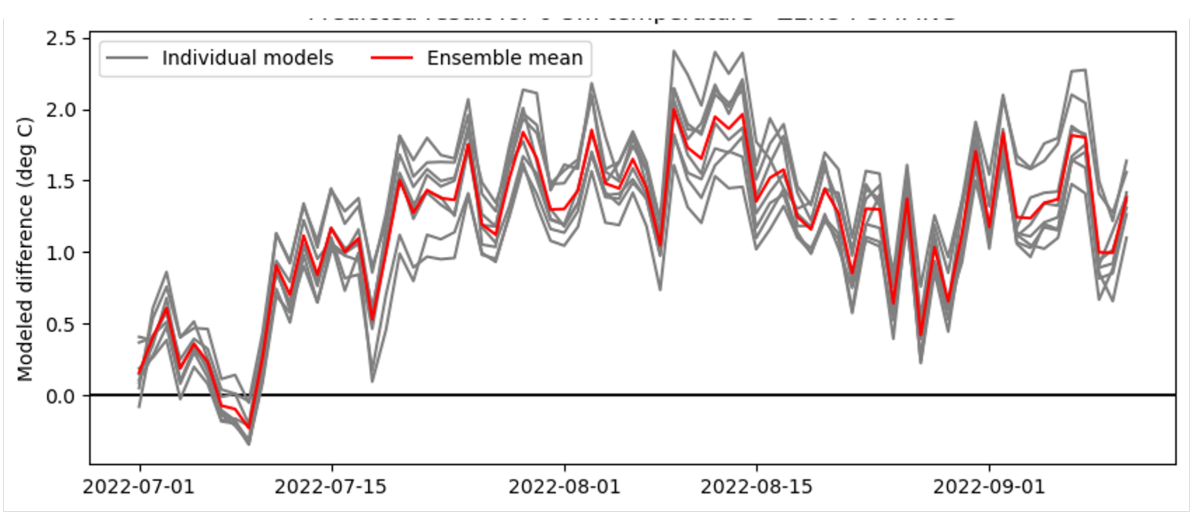

Figure 7: Residuals (predicted minus observed) of test set when pumping operations set to zero for water temperature at 0-5m. This indicates that the model is sensitive to pumping and that the result is sensical (if there is no introduction of cold water, the tempearture is predicted to be higher), additionally the integrated temperature is more dramatically impacted by no pumping operations than the upper 1m temperature in the previous figure.