

Assignment 3

final testing forecasts and insights into sources of predictability/skill

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. 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 in the Summer and Fall and to reduce diatom growth is to keep the average temperature of 0-5m (“integrated depth”) greater than 14°C in the Spring and early Summer, which is a bit of a “Goldilocks” problem.

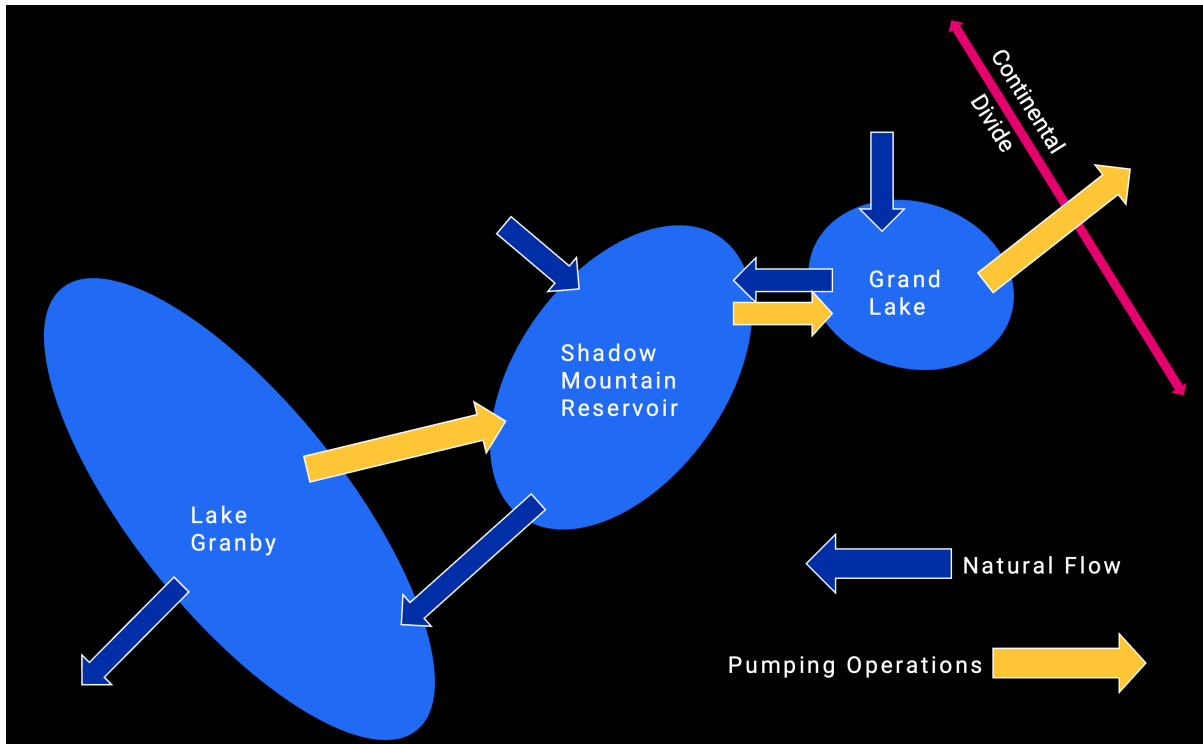


Figure 1: Cartoon schematic of water flow in the Three Lakes System

The overarching goal of this forecast system 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 at two depths (near-surface 0-1 meter, integrated 0-5 meter). The operational neural network . When operationl, we would add an operational “knob” that would alter pumping operations (while maintaining water balance) 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 out of scope for this class, so instead, I will focus on reliable 7-day forecasts using an auto-regressive neural network.

Data-Driven Forecast System

Model development

Northern Water has collected extensive data at the Three Lakes System for many years. In 2014, they deployed an instrumented buoy in Shadow Mountain near Chipmunk Lane, the connection between Grand and Shadow Mountain. Each season the buoy is deployed in late may and removed from the lake in early October, creating relatively independent years of data. The data-driven forecast relies on aggregated daily data from the buoy, volume data from inflows (North Fork into Grand Lake, North Inlet into Shadow Mountain), and volume of the interflow between Shadow and Grand (Chipmunk Lane). In addition to these hydrologic data, I use lagged meteorology data from a met station at the southern end of Shadow Mountain.

To determine the optimal use of the meteorological data, I experimented with two feature sets to predict the 2022 season (where 2022 was a hold-out test set). One feature set used each individual 3h data point (mimicking the GEFS 0.25 output) and another used aggregated daily values from those 3h data points. The observed met data were summarized to mimic the 3 hour GEFS data for instantaneous air temperature, wind speed, and relative humidity, minimum and maximum of the previous 3 hours for temperature, and mean of the previous 3 hours for solar radiation. Using each of the 3h data as features created a very large feature set (>100 features). I compared the performance of two feature sets, one where the 3h data were each used as features and one where the 3h data were used to create daily summaries (Table 1). From this, I determined that I should use the daily aggregations instead of the individual 3h data as it performed most similar to our baseline of a persistence forecast (‘yesterday-is-today’).

Table 1: Summary of model performance for aggregated and unaggregated met variables for 2022 test data.

Met Features	MSE 0-1m	MAE 0-1m	percent bias 0-1m	MSE 0-5m	MAE 0-5m	percent bias 0-5m
persistence (baseline)	0.23	0.37	2.44	0.12	0.28	2.15
3h data	0.34	0.44	3.00	0.20	0.37	2.87
3h aggregated to daily	0.29	0.43	2.90	0.16	0.33	2.53

The architecture of the initial model to determine input features and the operational forecast system is a fully-connected neural network comprised of a 10% drop out layer followed by 2 layers of 10 nodes each with leaky relu activations. The model was optimized through an Adam optimization function and the loss function was mean square error. Batch size was 64 and learning rate 0.001. To assure that the model was not overfit, I aimed to use hyperparameters that balanced losses across all cross validation data sets, assuring that the validation error was similar to the training error. For the 2022 testing for feature sets and the 2023 operational model, I used an ensemble method for prediction resulting from the leave-one-out validation (resulting in 8 members for 2022 and 9 for 2023).

Operational Model Testing Results

On a one-day time horizon, the operational model performs similarly to the persistence model, indicating that it should have reasonable skill upon roll out (Table 2). The near-surface (0-1 meter) prediction was more accurate during the entire buoy deployment period (mid May until early October) when compared to the persistence model than the integrated depth (0-5 meter depth), which only was comparable to the persistence when you only consider the regulatory period (Jul 1-Sept 11).

Table 2: Testing metrics for operational forecast model at 1 day time horizon across the full season of data and the regulatory period (Jul 1- Sept 11) for the 2023 test set.

model	MSE 0-1m	MAE 0-1m	percent bias 0-1m	MSE 0-5m	MAE 0-5m	percent bias 0-5m
persistence full season	0.26	0.42	2.45	0.12	0.28	2.00
operational full season	0.25	0.40	2.89	0.22	0.37	3.18

model	MSE 0-1m	MAE 0-1m	percent bias 0-1m	MSE 0-5m	MAE 0-5m	percent bias 0-5m
persistence regulatory period only	0.24	0.38	2.25	0.11	0.26	1.79
operational regulatory period only	0.20	0.37	2.20	0.14	0.30	2.20

Implementing the Forecast System

Results and Skill of Forecast System

- final testing results, forecast skill (i.e. metrics) and comparisons to baselines

Discussion of predictability

Particularly poor in early season. Unaccounted for variables controlling warm up period.

- detailed discussion of sources of predictability (this could be for a single forecast, or an overview of the sources of skill for all forecasts) and the approaches you used to gain this insight
- description of whether you think these sources of predictability make sense