Forecasting Hourly Water Temperature Across Four Tennessee Valley Authority Water Sites

Ryan Blanchette and Patrick Schulkers

May 3, 2021

Abstract

Abstract – Having access to early and reliable forecasts for water temperature is critical to plant performance for the Tennessee Valley Authority (TVA). 55% of TVA's power portfolio is dependent on water from reservoirs and rivers. Using historical water temperatures and nearby air temperatures for four sites – Cumberland Fossil Plant, Gallatin Fossil Plant, Kingston Fossil Plant, and Tims Ford Dam – time series models were trained to produce two-week hourly water temperature forecasts. Linear regression was the top-performing model for each site, with average mean absolute percentage errors (MAPE) between 0.13% and 0.29%. The most important predictors in the models included lagged 6 and 12 hour mean and max water temperatures and time of day.

1. Introduction

Tennessee Valley Authority (TVA) power portfolio includes 82% of energy production; plants, 41% nuclear, 27% from natural gas, and 14% from coal[1]. 89% of all-electric power worldwide is produced by steam-electric power plants. The process heats water into high-pressure steam that spins a steam turbine which drives an electrical generator. After spinning the turbines, the steam exits the turbine into a condenser where the steam is cooled. In both operations, intake water is the critical resource for operations to create the high-pressure steam used to spin the turbines and cool the steam in the condenser after it passes the turbine. Once cooled, the water is returned to energy production or discharged back into the initial water resource based on the plant's production process. The plants' optimal energy-efficient operation and meeting state and federal downstream discharge temperature regulations rely on accurate intake water temperature prediction. TVA monitors the water temperature with automatic sensors above and below their plants to meet the state and federal regulations.

During the spring to fall months, the reduced water flow and higher water temperatures create two operations challenges. The intake water used in energy production to drive turbines uses high-pressure steam generated in a high-temperature atmospheric pressure furnace fired by gas, coal, or a nuclear reactor. The first challenge is that a higher intake temperature generally reduces the vacuum level for the high-pressure steam and decreases the turbine's energy production efficiency. Requiring more steam to generate the same amount of energy is a major inefficiency.

After the steam leaves the turbine entering the condenser, a higher intake water temperature means a greater volume of water is required to cool the steam before returning to energy production[2]. Additionally, the second is the discharge water temperature limited by federal law and enforced by state agencies. TVA has a limit for the discharge temperature of 32.8 degrees Celsius, where the primary constraint strain takes place during the summer months. Higher intake water temperature increases the amount of water needed to cool the steam to meet the maximum Tennessee limit of 32.8 degrees Celsius before being discharged back into the water resource. Knowledge of future water temperatures allows the plants to adjust the cooling tower use to maintain optimization.

Currently, TVA provides a 7-day hourly water temperature forecast using hydrodynamic models. They use dam release patterns and upstream water temperature as model inputs. Their modeling approach is time-consuming and does not achieve the desired accuracy as flat-lining the water temperature forecast can be wildly inaccurate. Flat-lining cannot capture diurnal fluctuations, and it is unable to account for temperature trends from changing weather patterns.

Using machine learning and times series models will allow for better forecasting of intake water temperatures. Using these models can create realistic forecasts with diurnal fluctuations that react to daily, monthly, and seasonal weather changes. The aim is to develop models for four TVA locations; two reservoirs and two free-flowing rivers. TVA needs improved water temperature prediction for TVA to reduce energy consumption and water resource use.

Each site is unique due to the nearby dam's location or absence of a dam since upstream vs. downstream reservoirs, and rivers will behave differently. The first of the locations is the TVA Gallatin Fossil Plant, natural gas and coal-fired power plant located downstream of the Cordell Hull Dam northeast of Nashville, TN. The plant can produce 976 MW of electricity using four GE combustion turbine units with a combined capacity of 600 MW net. The second location is the Cumberland Fossil Plant, the largest generating asset in the TVA coal fleet located downstream of the Cheatham Dam. It can produce 2,470 MW net with all the waste heat returned to the Cumberland River water. The third location is the Kingston Fossil Plant, located on the Emory River where it meets the Clinch River upstream from the Watts Bar Dam. The plant has nine generating turbines producing a total capacity of 1,700 MW. The fourth location is the Tims Ford Dam on the Elk River that generates power generation for south-central Tennessee. Downstream of the Tims Ford Dam is a variety of warm-water fish thriving with at least two endangered species, the boulder darter and several mussels species, due to the rising water temperatures. The Buffalo River near Flatwoods is used to mimic the target temperature of Elk River to provide operational information for Tims Ford Dam water release models.

There have been many successful applications of statistical and machine learning models for predicting water temperature. The two primary forecasting methods are time series/stochastic modeling and machine learning. Autoregressive Integrated Moving Average (ARIMA) models, which use autoregressive and moving average components of historical data [3], were the most accurate 3-5 day forecasting method in a study of Delaware River water temperatures [4]. Faruk also showed that using ARIMA modeling is highly accurate in forecasting water temperatures by applying them to the Buyuk Menderes River in Turkey [5].

Similarly, Caissie, et al. used Autoregressive modeling to accurately forecast three days of water temperatures for the Little Southwest Miramichi River in New Brunswick, Canada [6]. Using the autoregressive nature of time series data, Rabi, et al. [7], Toffolon and Piccolroaz [8], and Caissie, et al. [9] successfully forecasted various water bodies using stochastic methods. Rather than using raw past data, other studies have used machine learning models to forecast water temperatures. These models include k-nearest neighbors [10], linear regression [11], decision trees [12], random forests [13], and deep learning methods such as multilayer perceptrons [14] and Long Short Term Memory (LSTM) models [15].

Along with different algorithms, there are also differences in model inputs. While some of the referenced studies use only past water temperatures as input and [9], [14], [15], most find success in using exogenous variables such as air temperature [7], [11], [12] and various atmospheric variables [13], [5]. Interestingly, Zhu also noted that when including temporal features such as day of the year to the model, forecasts were dramatically improved [12].

2. Methods

2.1 Data

Three different pieces of data for each location were collected. TVA provided the water temperature and flow rate data from four different water sources in Tennessee. The hourly water temperature data included seven years from the Buffalo River near Flatwoods, nine years from the Emory River at Oakdale, twelve years from the Cheatham Dam, and twelve years from the Cordell Hull Dam.

In addition to water temperature data, historical air temperature data were collected and downloaded from Iowa State University's Environmental Mesonet (IEM). IEM collects weather data worldwide using Automated Airport Weather Station (AWOS) and Automated Surface Observing System (ASOS) sensors. Two of the sites, Cheatham Dam and Emory River, included hourly air temperature data. The other two sites included data at a 20-minute interval. Figure 1 shows the locations of the data collection sites, and Appendix A provides descriptive statistics.

Figure 1: Locations of the four water sites

2.2 Data Preprocessing

The raw data contained several extreme values, random conversions to Celsius as a measure for water temperature, and several long runs of missing values. First, the extreme values were smoothed out by replacing each one with the previously observed data point. Next, the temperatures that jumped to Celsius were converted back to Fahrenheit. Finally, the missing values were imputed using the imputeTS R library. Specifically, the na_seadec function was used, which uses the average value for each season of a series to fill in missing data. [16] For example, if a series had an average water temperature of 50 degrees on January 1 each year, if January 2017 was missing, the na_seadec function would impute a value of 50 degrees. The Buffalo River water temperature contained the additional step of removing the second half of the data due to an unexplained pattern shift in the data. Similar steps were taken to clean air temperature data. Once all of the data were cleaned, air temperature and water temperature were joined by hour and date.

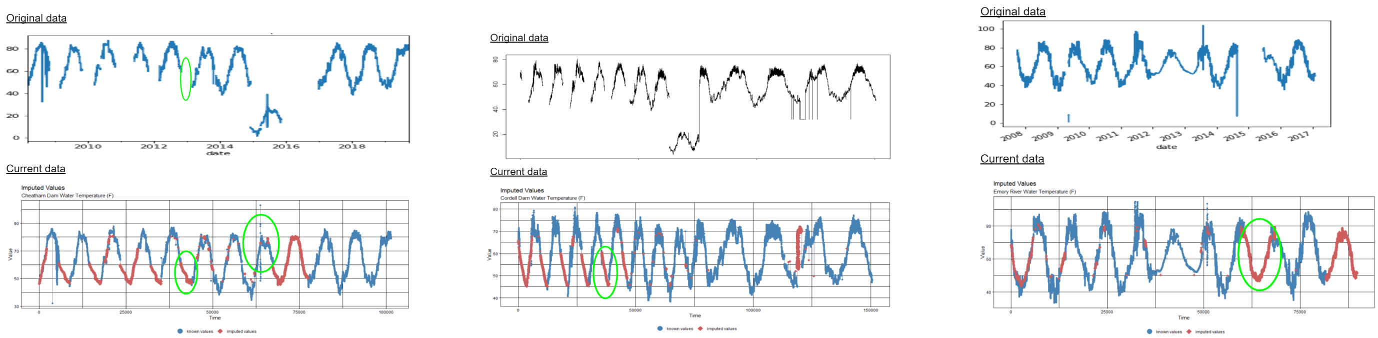
2.3 Models

Two types of models will be used: statistical time series models and a linear regression model. For time series modeling, seasonal naive and STLM models were used. Seasonal naive forecasting uses the last observed value from the previous season to make predictions. STLM models use multiple seasonal decomposition to create the forecasts. Two STLM models were used, one with air temperatures, and one without. In addition, an average model was used, which uses the series average as a forecast. Each model was fit and used to forecast using the forecast R library. [17]

3. Results

3.1 Missing Value Imputation

The Cheatham Dam, Cordell Hull Dam and Emory River sites each had many odd data shifts and long streaks of missing water temperature data. Figure 2 shows the before and after results of the Fahrenheit corrections and missing data imputations.



(a) Cheatham Dam (b) Cordell Hull Dam (c) Emory River

Figure 2: Water temperature before and after Fahrenheit correction and missing value imputations

3.2 Model Formulation

Each water site was fit with the five different model: mean model, Seasonal Naive, STLM, STLM with Air Temperature, and Linear Regression. The mean model, Seasonal Naive, and STLM models only used hourly water temperature as input and output data. The STLM with Air Temperature model used both hourly water and air temperatures as input data and hourly water temperatures as output. For the linear regression model, several transformations such as date features, lagged running averages, and lagged running maximums of hourly water and air temperatures were used as input data.

3.3 Model Performance

The models were tested using June through the end of September of 2016, and the previous data were used for training. A rolling-window cross-validation method was used where the models were used to forecast two weeks of data, evaluated, and then moved forward two weeks until all nine weeks during the testing period were covered. For example, the first iteration of training forecasted June 1 - June 14. Each model was evaluated using the Mean absolute percentage error (MAPE), which is shown is equation 1.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{A_t - F_t}{A_t}\right| \quad (1)$$

For each site, the linear regression performed the best, having both the lowest average MAPE and tightest MAPE standard deviation Table 1 shows the linear regression fit statistics for each site. Appendix B contains the fit statistics for the rest of the models. In addition to model performance, it is important to study which linear regression features were the most important. Figure 3 shows a plot of feature importances for the Buffalo River forecast. Note that lagged rolling average and means and hour of the day (denoted by hour_part_X) were the most critical variables, while air temperature did not play much of a role at all.

Table 1: Linear Regression Performance

Model	**Avg. MAPE**	**Std. Dev. MAPE**
Cheatham Dam	0.13%	0.03%
Cordell Hull Dam	0.29%	0.11%
Emory River	0.16%	0.02%
Buffalo River	0.17%	0.03%

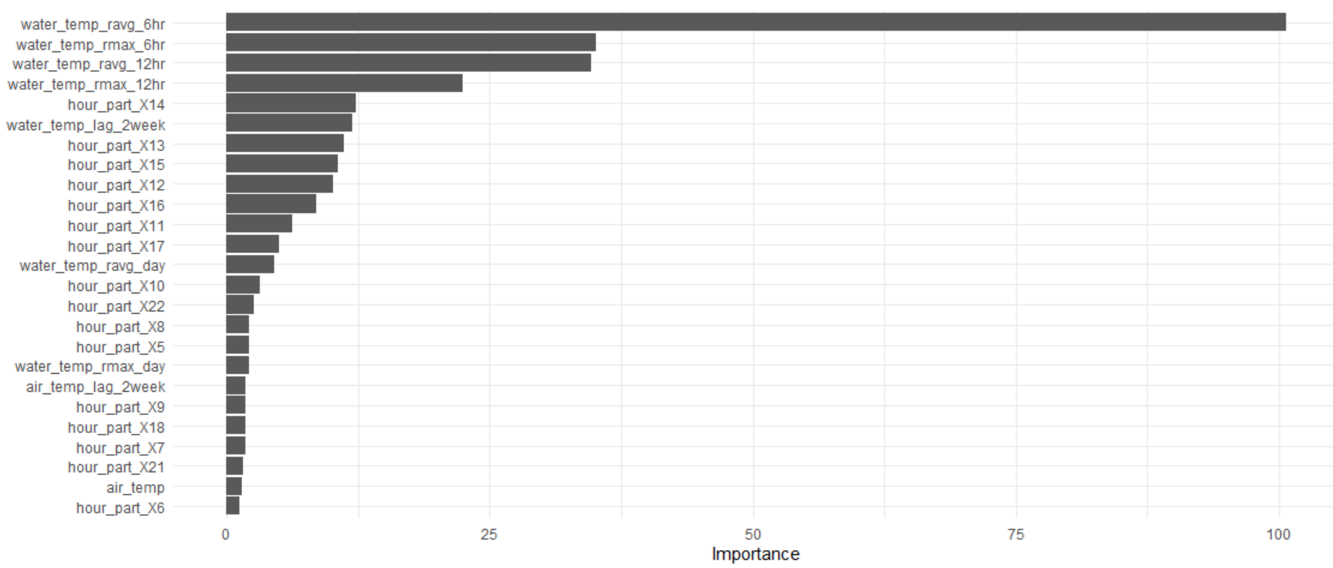


Figure 3: Buffalo River Feature Importance

4. Discussion

4.1 Impact

Moving forward, the forecasts should be further tested using longer forecast periods and newer data. By extending the forecasting period to three, four, or even more than four weeks, the models would gain greater value by allowing more time to react to the forecasts. The eventual goal of these forecasts would be to predict water temperatures in real-time, i.e., if today were January 1, 2021, one would want forecasts for January 14, 2021. An additional value-added layer would be forecast intervals, which add an upper and lower bound to the individual forecasts. Having these bounds adds a degree of confidence, and it is preferred to have the intervals be as tight as possible. Finally, further research could be applied to additional water sites, both in-state and out of state, to test the strength of using past water temperature and air temperature values in a linear regression model.

4.2 Future Work

The research demonstrates the predictable nature of water temperature using machine learning. Given the varying waterways evaluated, the model produced an average MAPE of 0.18%, demonstrating its viability as a predictor for intake water temperature for energy production plants. The research also showed its ability to impact areas where the water temperature can be altered to sustain endangered marine life. The three dam locations all affect existing fossil plants servicing the state of Tennessee's energy needs. Since TVA is to "deliver power at the lowest feasible cost," cost savings in production cost with knowledge of the intake water temperature will benefit all the residents of the state[1]. The efficient operations will bring cost savings on the upstream and downstream operations side of the spinning turbine. TVA was the focus of the research, but the forecasting methodology could be applied to energy production plants throughout the world.

References

[1] Our power system. https://www.tva.com/energy/our-power-system. Accessed: 2021-03-07.

[2] JoAnna L Lessard and Daniel B Hayes. Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. *River research and applications*, 19(7):721–732, 2003.

[3] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[4] Jeffrey C Cole, Kelly O Maloney, Matthias Schmid, and James E McKenna Jr. Developing and testing temperature models for regulated systems: A case study on the upper delaware river. *Journal of hydrology*, 519:588–598, 2014.

[5] Durdu Ömer Faruk. A hybrid neural network and arima model for water quality time series prediction. *Engineering applications of artificial intelligence*, 23(4):586–594, 2010.

[6] Daniel Caissie, Nassir El-Jabi, and Mysore G Satish. Modelling of maximum daily water temperatures in a small stream using air temperatures. *Journal of hydrology*, 251(1-2):14–28, 2001.

[7] Anamarija Rabi, Marijana Hadzima-Nyarko, and Marija Šperac. Modelling river temperature from air temperature: case of the river drava (croatia). *Hydrological sciences journal*, 60(9):1490–1507, 2015.

[8] Marco Toffolon and Sebastiano Piccolroaz. A hybrid model for river water temperature as a function of air temperature and discharge. *Environmental Research Letters*, 10(11):114011, 2015.

[9] Daniel Caissie, Maria E Thistle, and Loubna Benyahya. River temperature forecasting: case study for little southwest miramichi river (new brunswick, canada). *Hydrological Sciences Journal*, 62(5):683–697, 2017.

[10] André St-Hilaire, Taha BMJ Ouarda, Zoubeida Bargaoui, Anik Daigle, and Laurent Bilodeau. Daily river water temperature forecast model with ak-nearest neighbour approach. *Hydrological Processes*, 26(9):1302–1310, 2012.

[11] GB Sahoo, SG Schladow, and JE Reuter. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. *Journal of Hydrology*, 378(3-4):325–342, 2009.

[12] Senlin Zhu, Emmanuel Karlo Nyarko, Marijana Hadzima-Nyarko, Salim Heddam, and Shiqiang Wu. Assessing the performance of a suite of machine learning models for daily river water temperature prediction. *PeerJ*, 7:e7065, 2019.

[13] Mario Álvarez-Cabria, José Barquín, and Francisco J Peñas. Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors. *Science of the Total Environment*, 545:152–162, 2016.

[14] Renata Graf, Senlin Zhu, and Bellie Sivakumar. Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *Journal of Hydrology*, 578:124115, 2019.

[15] Senlin Zhu, Bahrudin Hrnjica, Mariusz Ptak, Adam Choiński, and Bellie Sivakumar. Forecasting of water level in multiple temperate lakes using machine learning models. *Journal of Hydrology*, 585:124819, 2020.

[16] Steffen Moritz and Thomas Bartz-Beielstein. imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1):207–218, 2017.

[17] Rob J Hyndman and Yeasmin Khandakar. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22, 2008.

Appendices

A. Descriptive Statistics of Data

Site Location Requiring Water Temperature Predictions	Measure	Measure Location	Time Frame	Frequncy
Gallatin Fossil Plant	Water Temperature	Cordell Hull Dam	2007-2020	Hourly
	Water flow rate	Cordell Hull Dam	2007-2020	Hourly
	Air temperature	Nashville MET data	2014-2019	Hourly
Cumberland Fossil Plant	Water temperature	Cheatham Dam	2008-2019	Hourly
	Water flow rate	Cheatham Dam	2007-2019	Hourly
	Air temperature	Nashville MET data	2007-2020	Hourly
Kingston Fossil Plant	Water temperature	Emory River at Oakdale	2007-2017	Hourly
	Water flow rate	Emory River at Oakdale	2007-2017	Hourly
	Air temperature	Oak Ridge MET data	2007-2019	Hourly
Tims Ford Dam	Water temperature	Buffalo River near Flat Woods, TN	2013-2020	Hourly
	Water flow rate	Buffalo River near Flat Woods, TN	2013-2020	Hourly
	Air temperature	Fayetteville MET data	2014-2019	Hourly

Site Location Requiring Water Temperature Predictions	Measure	Mean	Standard Deviation	Min	Max
Gallatin Fossil Plant	Water temperature	55.69	16.59	3.50	80.70
	Water flow rate	14,002.67	12,554.43	0.00	130,100.00
	Air temperature	57.90	17.82	3.20	96.80
Cumberland Fossil Plant	Water temperature	61.37	20.18	1.40	87.90
	Water flow rate	23,756.15	22,983.27	0.00	240,000.00
	Air temperature	59.94	17.82	1.04	107.06
Kingston Fossil Plant	Water temperature	61.49	13.34	1.00	103.70
	Water flow rate	1,551.57	3,429.68	2.00	74,900.00
	Air temperature	58.67	16.85	0.00	105.08
Tims Ford Dam	Water temperature	59.76	11.96	32.00	88.90
	Water flow rate	845.86	1,368.60	126.00	32,586.00
	Air temperature	60.48	17.21	1.40	95.00

B. Model Performance

Buffalo River

Model	Avg. MAPE	Std. Dev. MAPE
Linear Regression	0.17%	0.03%
Seasonal Naive	3.85%	1.65%
STLM with Air Temp.	4.11%	1.10%
STLM	4.11%	1.10%
Mean Model	20.22%	4.16%

Cheatham Dam

Model	Avg. MAPE	Std. Dev. MAPE
Linear Regression	0.13%	0.03%
Seasonal Naive	2.82%	2.14%
STLM with Air Temp.	2.82%	2.14%
STLM	3.19%	1.74%
Mean Model	17.55%	3.99%

Cordell Hull Dam

Model	Avg. MAPE	Std. Dev. MAPE
Linear Regression	0.29%	0.11%
Seasonal Naive	2.48%	1.13%
STLM with Air Temp.	2.48%	1.13%
STLM	3.59%	1.24%
Mean Model	14.59%	2.97%

Emory River

Model	Avg. MAPE	Std. Dev. MAPE
Linear Regression	0.16%	0.02%
Seasonal Naive	2.85%	1.29%
STLM with Air Temp.	2.85%	1.29%
STLM	12.28%	4.07%
Mean Model	24.38%	3.34%