To the Graduate Council:

I am submitting herewith a dissertation written by Rachel Wood-Ponce entitled "Machine Learning for Urban Water Runoff Prediction and Water Temperature Forecasting." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Systems Engineering.

_____

Anahita Khojandi, Major Professor

We have read this dissertation
and recommend its acceptance:

_____

Jon Hathaway

_____

Hugh Medal

_____

Bing Yao

Accepted for the Council:

_____

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

To the Graduate Council:
I am submitting herewith a dissertation written by Rachel Wood-Ponce entitled "Machine Learning for Urban Water Runoff Prediction and Water Temperature Forecasting." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Systems Engineering.

Anahita Khojandi, Major Professor

We have read this dissertation
and recommend its acceptance:

Jon Hathaway

Hugh Medal

Bing Yao

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Machine Learning for Urban Water Runoff Prediction and Water Temperature Forecasting

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Rachel Wood-Ponce

May 2025

This dissertation is dedicated to my daughter, Alejandra, my husband, Moises, my parents, Sean and Elizabeth, and my sisters, Bethany, Shawna, and Sarah for their love and belief in me for the last 4 and a half years.

Acknowledgements

I extend my gratitude to my advisor, Dr. Khojandi, and the members of my committee for their support throughout this journey. I am also very grateful to my graduate school professors, mentors, and friends, whose encouragement has been invaluable. To my family, thank you for your unwavering support and encouragement. To my husband Moises, your love and steadfast support have been my anchor and to my daughter Alejandra, your radiant joy has been a source of light and inspiration during challenging times.

The research in Chapter 1 is based upon work supported by the National Science Foundation under Grant No. CMMI-1634975. This chapter has been published in near-identical form in the Urban Water Journal, Volume 21, Issue 5. https://doi.org/10.1080/1573062X.2024.2312514

Abstract

In this dissertation, we investigate the application of machine learning (ML) methods in two areas of hydrology.

Chapter 1 presents the development of data-driven models to predict urban stormwater runoff volume. Machine learning (ML) models are used to approximate output from the Stormwater Management Model (SWMM), a hydrological model used for runoff simulation. SWMM is powerful but computationally intensive, so ML models are developed to expedite runoff predictions. A case study for the First Creek watershed in Knoxville, Tennessee, is performed using rainfall data and subcatchment characteristics. Feature engineering and clustering are applied to compare SWMM and ML model outputs. Results show that random forests accurately predict runoff volume almost instantly, significantly reducing computational requirements.

Chapter 2 explores trends, fluctuations, and correlations of water temperatures and other variables at Tennessee Valley Authority (TVA) fossil plants. The aim of this research is to utilize real-world historical water temperature data to analyze the environmental changes at fossil plants.

Chapter 3 focuses on forecasting hourly water temperature for the efficiency of three TVA fossil plants and the impact on aquatic life. The findings of this study can be beneficial for fossil plants that need to consider the sensitivity of water temperatures for their operations.

Table of Contents

List of Tables

List of Figures

Chapter 1

Developing Data-Driven Learning Models to Predict Urban Stormwater Runoff Volume

1.1 Introduction

Climate change has become one of the most severe environmental anomalies around the globe, resulting in extreme hydrological events, and threatening infrastructure. In addition, the continuing urbanization and densification of land also poses a threat to surface water quality. More than 50 percent of the world's population lives in urban areas, this number expected to increase in the coming years (2; 3). Densification is a common strategy for accommodating rapid growth in urban areas, resulting in more impervious areas (4; 5). Therefore, the combination of climate change and urbanization/densification leads to more vulnerability to flooding and stormwater runoff pollution. It is becoming progressively important to find sustainable solutions to the additional capacity required to convey flows as conditions change.

In an effort to help support stormwater management and planners, the Storm Water Management Model (SWMM) was created by the United States Environmental

Protection Agency (6). SWMM is used for the planning, analysis, and design of stormwater runoff systems. Given the land use, geological features, and historic precipitation, SWMM models the hydrologic processes of a particular area. A given watershed may include varying percentages of development, pastures, forested areas, or other land use / land covers. With this level of land cover heterogeneity, the watershed is divided into appropriate number of subcatchments that have similar characteristics. Despite the popularity of SWMM, this approach can be computationally expensive for the typical user. This can be exemplified by three studies. First, a recent study by (7) used a times series spanning only one year expressed in hourly units in a medium-sized watershed with 140 subcatchments on a 2.4 GHz CPU. A single simulation took approximately 25 minutes. While 25 minutes for a single simulation may seem reasonable, this study required rerunning a SWMM simulation hundreds of times to achieve the desired outcome, meaning that getting the results for all scenarios required an incredibly lengthy run-time. Next, in a study by (8) sought to conduct municipal scale analyses that included hydraulic networks for the continuous simulation of rainfall timeseries. However, to use SWMM for these analyses would take hours to days to complete the simulation. This led them to develop SWMM for Low Impact Technology Evaluation. Finally, a study by (9) developed an efficient simulation-optimization methodology using the Markov Chain with the multi-objective shuffled leaping frog algorithm to bypass the use of the comparatively computationally expensive SWMM.

One potential solution to time intensive models based on physical processes is the use of data-driven models such as machine learning (ML). ML constructs artificial intelligence algorithms that learn knowledge from existing datasets, and predict output from new datasets (10). Specifically, ML algorithms detect and describe patterns in datasets, attempt to generalize them, and then make predictions on the system state (11). ML predictions have been used in many areas of hydrology including stormwater quality (12), water treatment and monitoring (13), rainfall-runoff modeling (14), and water drainage systems (15). ML in hydrology has

developed and expanded in recent years, as the input-output systems in hydrology are well-suited for these approaches (11).

This study investigates the possibility of streamlining hydrologic modeling in order to pave the way for more vigorous optimization efforts in the future. This is done without the need to simplify or reduce assumptions. To accomplish this, non-linear ML models are created to act as function approximators to estimate the sophisticated underlying hydrologic sub-routines used in SWMM. Specifically, we seek to use ML models to accurately predict the output from a SWMM simulation. Using SWMM allows us to generate enough data to make ML possible. The use of ML in this context is novel as most urban watersheds do not have enough field collected raw data to apply these methods. This study also looks to create ML models that can be generalized to other areas, within this watershed, that have different geophysical features. The area of interest for this study is the First Creek Watershed in Knoxville, TN, USA. This watershed is comprised of 140 subcatchments. The map of the area and the subcatchments are shown in Figure 1.1.

The contribution of this study is four-fold. First, we develop a novel method that utilizes ML algorithms as a complement to SWMM to predict the stormwater runoff volume without compromising the prediction accuracy. Second, the proposed method exploits the knowledge learned by the ML model to provide faster predictions than SWMM. Third, we use real world data to evaluate the potential performance of our proposed approach. Fourth, we demonstrate the generalizability of our proposed approach by investigating it in areas with differing characteristics.

1.2 Background

SWMM models have wide application in literature to simulate complex hydrology of urban catchments pre-and post-development (16), to model water quality (17), and to study the shift in the hydrologic regime as a result of the informed and random spatial placement of green infrastructure (18; 19) under precipitation uncertainty (7).

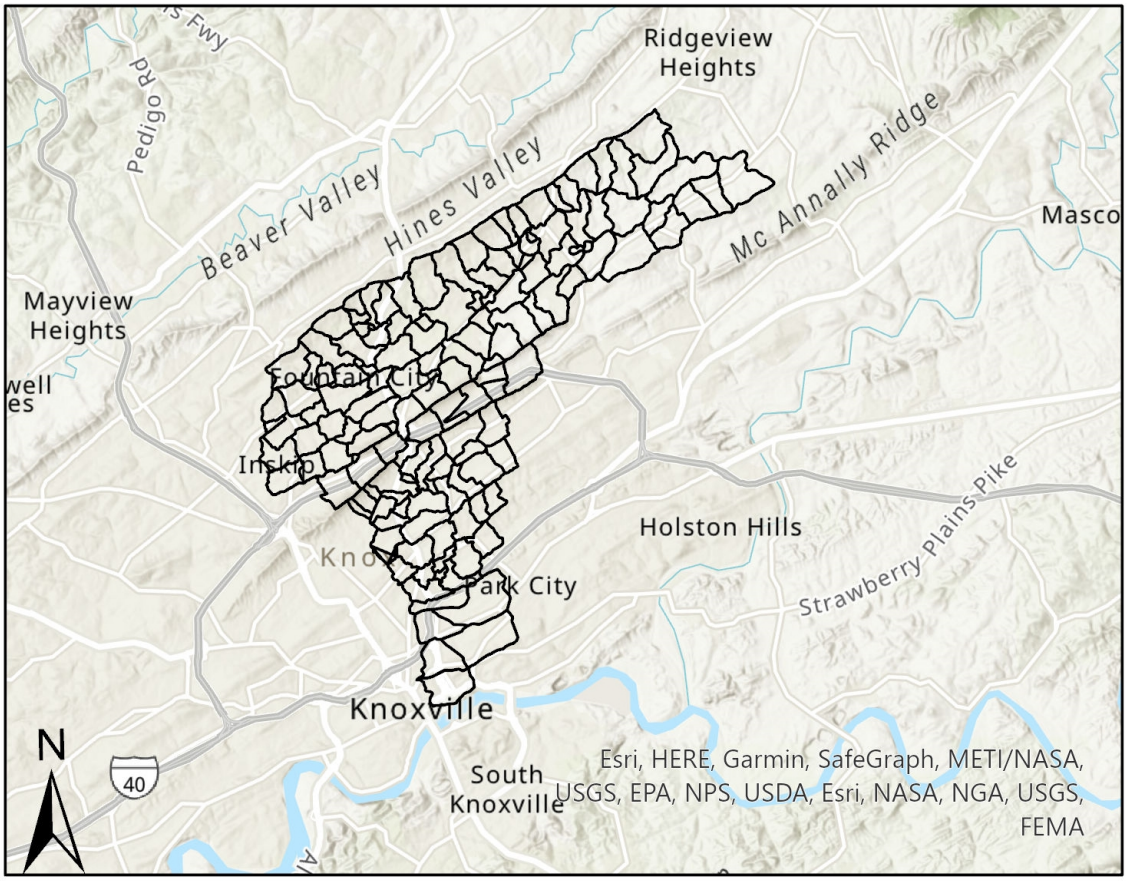

Figure 1.1: Map of hydrological subcatchments of the First Creek, Knoxville, Tennessee

ML techniques are increasingly being implemented in hydrology and environmental engineering (20; 21). This approach is appealing for these areas as ML models can make predictions without deep knowledge of underlying physical parameters (22). In recent years, many different ML methods have been used for prediction in hydrologic research domains. In a study by (23), a regression technique called Support Vector Machine (SVM) is applied to 10 sites for soil moisture estimation in the Lower Colorado River Basin. The SVM results are compared to an Artificial Neural Network (ANN) model and a Multivariate Linear Regression (MLR) model, showing that SVM outperformed the other models. MLR and ANN are also used in a study by (24). These models, in addition to a random forest (RF) model, were used to predict the performance of stormwater biofilters removing heavy metals from stormwater runoff. The results of this study demonstrated that the RF model was more robust than the other models in predicting metal removal.

Other examples include a study by (25) compared eight ML methods to estimate the Total Suspended Solids (TSS) in the national stormwater quality database. Some of the models used include Linear Regression, Support Vector Regression (SVR), ANN, Regression Tree, Adaptive Boosting (AdBoost), and RF. Of all the models tested, AdBoost and RF gave the best fit. There have also been many studies that forecast stormwater runoff in order to predict flooding and water levels (26; 27; 28; 29; 30; 31). A study by (32) uses a multi-layer perceptron (MLP) network, which is a popular type of ANN, to predict stormwater runoff quantity.

The comparison of ML models and process-based and physically based models is seen in many different areas of study, such as crop yield (33), healthcare (34), and heating, ventilation and air conditioning (HVAC) systems (35) in addition to hydrology (36; 37; 38; 39). In each area, the studies found that the ML models proved to be more flexible and efficient, bypassing computationally expensive models and allowing for real-time predictions. A study by (40) used ML models and compared their output to the output of two hydrological models, LRHM and HYMOD2. They found that the ML approach helped to unveil anomalies in the catchment data that

did not fit in the structure of the hydrological models. The study argues that while the usefulness of hydrological models is indisputable, the ML models proved to be a helpful tool to evaluate these models.

SWMM model performance has been compared to SVR modeling rainfall-runoff in two basins in Italy (41). Notably, the performance of both models was comparable with SVR having a lower root mean square error and higher coefficient of determination. SWMM has overestimated the peak flow by 20% while SVR underestimated the peak flow by 10%. In another example, the output of the SWMM model has been used for data augmentation to train a Deep Neural Network model to predict urban flooding in a basin in Seoul (42).

Although there have been many uses of SWMM and ML in the literature, to the best of our knowledge, there has been no previous study that trains ML models as substitutes for SWMM. A similar study compared eight different ML models, including MLR, MLP, and RF, with a process-based model called advanced hydrologic prediction system (AHPS) at Lake Erie (43). MLR exhibited a more accurate prediction of the actual water level in Lake Erie than AHPS. The study demonstrates that in addition to their time-saving advantage, ML models can be used to accurately predict water levels.

1.3 Data and Methods

In this section, we explore the data, methods, and location used in this study. A discussion of the data used is given in Section 1.3.1. Section 1.3.2 describes the ML models. An in depth description of the watershed used for this case study and the associated SWMM model is given in Section 1.3.3. Section 1.3.4 introduces the three experiments conducted. Sections 1.3.5, 1.3.7, 1.3.8 explains the methods used for clustering and interpreting the ML models, the model tuning steps taken and modifications made before conducting each of the experiments, and explains each of

the evaluation metrics used in the experiments. Figure 1.2 gives a visual overview of the methods used for this study.

1.3.1 Data and Data Preprocessing

Rainfall data for this study was collected at the McGee Tyson Airport in Knoxville, TN. It contains hourly rainfall depth for January 1st, 2010, to December 31st, 2020. These data are used in the SWMM simulation, which then generates outputs from each of the subcatchments of First Creek Watershed. The total runoff in 10^6 gallons is the target value for this study, meaning that the ML models are trying to predict this value and then comparisons are made to the actual output of this variable from the SWMM model. Additionally, there is input data for each of the subcatchments. This includes the characteristics of the 140 subcatchments, as shown in Table 1.1.

Before beginning the experiments, pre-processing of the data was required. As previously stated, there are three sets of data: hourly rainfall data, data containing the geographic characteristics of the subcatchments, and output data containing daily output runoff volumes from the SWMM simulation. We want to predict the daily amount of runoff due to precipitation in a given subcatchment with ML. The pre-processing required to accomplish this included combining each day of the hourly precipitation data with each subcatchment (and thus its associated characteristics). This creates the independent variables. Additional pre-processing removed days where no precipitation occurred. This is so the model can more accurately predict output on days with precipitation.

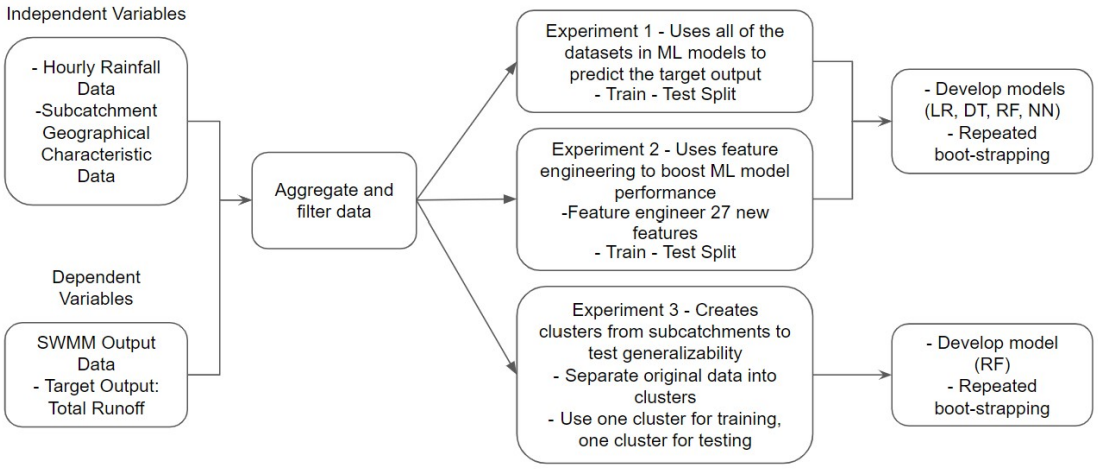

Figure 1.2: Overview of the methods for this study. We use three sets of data. All of these datasets are combined and preprocessed. Next, three experiments are conducted. The first experiment takes the filtered combined dataset and splits the data 75-25 for training and testing. The second experiment uses feature engineering for the dataset. The final experiment separates the subcatchment into several clusters. The clusters are then trained and tested against all the other clusters, where one cluster is used for training and the other is used for testing. This experiment only uses the RF model. After each of the experiments are completed, the results are evaluated, where the predicted value is compared to the actual target output.

Table 1.1: Description of Subcatchment data

Feature	Description	Average (standard deviation)
Area	The total area of the subcatchment	99.63 (55.96) acres
%Imperv	Percent of the land area which is impervious	13.12 (7.57)%
Width	The width of the overflow path. Given by the subcatchment area divided by the average maximum overland flow length	3730.24 (2168.98) feet
%Slope	Percent of slope of the land area	4.40 (2.77)%
N-Perv	Manning's n for overland flow over the pervious portion of the subcatchment	0.25 (0.03)
S-Perv	Depth of depression storage	0.12 (0.01) in
MaxRate	Maximum rate of infiltration	5.99 (0.96) in/hour
MinRate	Minimum rate of infiltration	0.15 (0.03) in/hour
MaxInfil	Maximum volume of infiltration	2.55 (0.41) inch

1.3.2 ML models

This study compares three different types of popular ML algorithms (44). These are linear regression algorithm, decision tree, and back propagation algorithm. There are then at least one ML model for each type of algorithm (two for decision tree). These ML models are linear regression, decision tree, random forest, and neural networks. This section provides an explanation of these models.

Linear Regression

Linear regression (LR) is a supervised learning algorithms in ML that assumes that there exists a linear relationship between the input and output. It then attempts to fit a linear equation to the observed data. That is, if y represents a single output variable, and x represents the input variables, then y can be calculated as a linear combination of x:

$$y = c_0 + c_1 x \tag{1.1}$$

where c_0 is the intercept and c_1 is the slope of the line (45).

LR is rather straightforward to implement and interpret. However, as stated, this model requires the assumption of a linear relationship between the independent and dependent variables and thus can be sensitive to outliers. This model was selected for this study as a 'measuring stick' for the other models as we expect the results to be meaningful, but we anticipate the model to potentially underperform.

Decision Tree

Decision Tree (DT) is a supervised learning model, which can be used for both classification and regression. This study uses DT for regression. DT uses a tree-like model of decisions to predict the value of the target variable. It is comprised of three basic elements: root node, branches, and leaf node. The root node is the ultimate objective. Branches correspond to different options (i.e., the different

possible attribute values) that flow from the root and leaf nodes to other leaf nodes. And, finally, the leaf nodes represent the possible outcome for each branch, or action, that is taken. In regression, the model looks to find a point to split the data into two parts, where the error is minimized at this point. It does this repetitively and a tree-like structure is created (46; 47). DT models are intuitive and easy to interpret. However, the calculations in these models can be complex, requiring more computational time for training.

Random Forest

Random forest (RF) is a supervised learning algorithm that consists of many decision trees. RF investigates all of the features of dataset and determines which features are contributing the most after training. It builds these decision trees by using bagging and feature randomness to create an uncorrelating forest. In RF for regression, each tree in the forest is created from different and random sample of rows. Then a different sample of features is selected for splitting at each node. The individual trees make their own prediction, and the predictions are averaged to produce one result (48).

RF models are not as influenced by outliers and can handle both linear and non-linear relationships. They generally preform very well with tabular data and provide accurate results. Because RF is comprised of many DTs, these models are more computationally intensive than DTs, and the results are not as easily interpretable.

Neural Networks

A Neural network (NN) contains nodes that communicate with each other. There are three basic layers, each of which consists of nodes. These layers are input, hidden, and output. The nodes in each layer are connected to the nodes in the next layer. Each layer can contain a different number of nodes. The number of nodes in the input layer will be determined by the number of features in the dataset, and the number of nodes in the output layer will be determined by the number of target values. The

output of a single node is calculated by summing up the multiplication of an input with the corresponding weight. The bias is then added, i.e., a value which provides each node a trainable constant. The resulting value is passed through an activation function which gives the output. This can be expressed by the equation below:

$$Y = \varphi\left(\sum_{i=1}^{n} x_i \times w_i + b\right) \tag{1.2}$$

where Y represents the output, $\varphi(\cdot)$ represents the activation function, x_i represents an input neuron, w_i represents the corresponding weight, and b represents the bias (49). NN models are very flexible, as any number of layers can be used for training, and can easily be used for large amounts of input data. NN models are notoriously difficult to interpret and are prone to overfitting.

1.3.3 SWMM Model

Training examples for this model are generated based on output metrics from SWMM (which have been calibrated using a relatively smaller amount of observed data required than that needed to develop ML models). This allows training samples for the ML model that may be outside of events captured through hydrologic monitoring. This is often not performed for a long enough time frame to capture all possible scenarios, and is not performed at all for many urban watersheds.

The land use across the watershed of interest ranges from rural areas (mostly agricultural lands, 23.8%) upstream to urban, highly developed land (residential and commercial, 52.6%) in the middle and downstream reaches. The SWMM model used in this study is obtained from the Stormwater Engineering Division of the City of Knoxville. The watershed is modeled in SWMM as a set of hydrologic units with different geometric physical and soil characteristics that discharge to a node. Nodes are connected via links that represent pipes, conduits, and open channels. The First Creek watershed SWMM model consists of 140 subcatchment units delineated from topography for a total area of 5645 ha. A set of properties were defined for

each subcatchment including total area, slope, width, impervious area percentage, Manning roughness for the pervious area (N-Perv), and depth of depression storage (S-Perv). The area of subcatchments ranges from 3.3 ha to 122 ha with slope from 1% to 15% and imperviousness from 1.3% to 40.4% . The flow from pervious and impervious areas in each subcatchment was routed to the subcatchment outlet. Dynamic wave model was used to route flow using a time step of one second while the runoff time step was 5 minutes. Infiltration rate was accounted for in the model using the Horton Infiltration model. The input parameters defined for each subcatchment include maximum infiltration rate, minimum infiltration rate, and decay constant.

SWMM Model Validation

Model calibration and validation was performed by the original model creators (50) consulting with the city of Knoxville. The procedure followed for model validation is to compare the model output to observed data. The output used for this purpose is total annual flow volume, the flow frequency exceedance curve, flow hydrographs, and hourly and daily flow output. The goodness of fit was evaluated using visual inspection of graphs, error percentage, and the Nash-Sutcliffe efficiency (NSE, defined by (51)), which is a metric that measures how well a simulation can predict the outcome variable. (NSE further described in section 1.3.8). The daily precipitation of three rain gages located around the First Creek Watershed was used to account for spatial and temporal variability. Data from 2007 to 2009 are used to collaborate and validate the performance of the First Creek SWMM Model. The error of the annual flow volume predicted by the model compared to measured flow is less than 1% throughout the three-year period. Hydrograph plots indicated that the model is able to capture the temporal trend in observed data. The flow frequency-exceedance curves shows favorable agreement between the observed and measured flow. The NSE for the entire simulation period is 0.57 and 0.62 for the hourly and daily flow, respectively. Accordingly, the model performance is deemed satisfactory and appropriate for predicting observed data (52). After the model was validated using

the observed data, the model was run for a longer period of time to get input data for the ML models.

1.3.4 Description of Experiments

This section gives a description of each of the experiments performed in this study. A summary of each experiment is shown in Table 1.2.

Experiment 1: Base Model

The first experiment examines how accurately the ML models can predict total runoff. First, all of the input, subcatchment, output data from SWMM are gathered as input for the ML models. This is true for all following experiments. Then the total runoff at the watershed outlet data are split into training and testing data, where 75% of the data are used to train the models, and 25% of the data are used to test the model. Each of the models go through repeated boot-strapping, where they are run ten times with randomly partitioned data in order to accurately gauge the robustness of each model. As stated above, all of the ML models described are used in this experiment.

Experiment 2: Feature Engineering

In an attempt to boost the accuracy of predictions, a second experiment was performed as an extension of the first involving feature engineering. The purpose of feature engineering is to create meaningful and useful features from historical data that will boost performance of ML models (53). In this study, this is done by transforming the historical data into new features by combining and averaging the existing features. The features that are created include: combining 3 hours the hourly precipitation data (i.e. combining hours 1-3, hours 4-6, etc.), combining 6 hours of hourly precipitation (combining hours 1-6, hours 7-12, etc.), combining the morning and evening precipitation hours, total daily rainfall, average

Table 1.2: Summary of Experiments 1-3

Experiment number	Experiment name	Description of Experiment
Experiment 1	Base Model	LR, DT, RF, and NN are used to predict the target variable from the SWMM simulation: total runoff
Experiment 2	Feature Engineering	New engineered features are added to the data. LR, DT, RF, and NN are used to predict total runoff.
Experiment 3	Clustering Subcatchments	Subcatchments are grouped into clusters. These clusters are used, one for training and one for testing, in RF to predict total runoff.

hourly rainfall, standard deviation of hourly rainfall, minimum and maximum hourly rainfall, and the average of 3, 7, and 14 days of rain. Hydrologically, this feature engineering allows the ML models to incorporate the "history" of the watershed, which can act as a surrogate for critical conditions such as antecedent moisture in the system at the time of a given rain event. A table describing each of the newly created features is shown in Table 1.3.

Experiment 3: Clustering Subcatchments

The purpose of Experiment 3 is to begin to answer the question: can a ML model trained on one area predict the output of another area with differing geophysical features? That is, can the model be generalized for other catchments with different characteristics? To do this, we separate the subcatchments into multiple clusters and then compare the clusters against each other. A clustering algorithm is used to create clusters that contain subcatchments of similar characteristics. Thus, each cluster is geophysically different from one another. After the number of clusters has been finalized (see details below), one cluster is used as the training set for the RF model and another cluster is used for the testing and the total runoff for this cluster is predicted. When testing one cluster against itself, 75% of the data in this cluster is used for training, and 25% of the data are used for testing.

1.3.5 Clustering Methods

In this section, an explanation of the methods used for clustering subcatchments in Experiment 3 is given.

k-means Clustering

Specific to Experiment 3, the algorithm k-means clustering is used to create clusters. While there are different methods for clustering data, k-means clustering is one of the most commonly used algorithms (54). The algorithm begins by specifying the

Table 1.3: Description of the new features created in Experiment 2

Created Feature	Description
Hours 1 - 3	Sum of the amount of rainfall in hours 1 through 3
Hours 4 - 6	Sum of the amount of rainfall in hours 4 through 6
Hours 7 - 9	Sum of the amount of rainfall in hours 7 through 9
Hours 10 - 12	Sum of the amount of rainfall in hours 10 through 12
Hours 13 - 15	Sum of the amount of rainfall in hours 13 through 15
Hours 16 - 18	Sum of the amount of rainfall in hours 16 through 18
Hours 19 - 21	Sum of the amount of rainfall in hours 19 through 21
Hours 22 - 24	Sum of the amount of rainfall in hours 22 through 24
Hours 1 - 6	Sum of the amount of rainfall in hours 1 through 6
Hours 7 - 12	Sum of the amount of rainfall in hours 7 through 12
Hours 13 - 18	Sum of the amount of rainfall in hours 13 through 18
Hours 19 - 24	Sum of the amount of rainfall in hours 19 through 24
AM Rain	Sum of the amount of rainfall in hours 1 through 12
PM Rain	Sum of the amount of rainfall in hours 13 through 24
Total Rain	Sum of the amount of rainfall in all hours
Average Hourly Rainfall	Average amount of rainfall in all hours
Standard Deviation of Rainfall	Standard deviation of rainfall in all hours
Maximum Hourly Rainfall	The maximum amount of rainfall in all hours
Minimum Hourly Rainfall	The minimum amount of rainfall in all hours
Month	The number representation of the month
Day	The number representation of the day
Week	The number representation of the week
Quarter	The number representation of the quarter
Average of Previous 3 Days	The average amount of rainfall in the previous 3 days
Average of Previous 7 Days	The average amount of rainfall in the previous 7 days
Average of Previous 14 Days	The average amount of rainfall in the previous 14 days
Average of Previous 21 Days	The average amount of rainfall in the previous 21 days

number of clusters that are desired, that is, k. Next, k number of centroids are randomly assigned. The datapoints closest to the centroids are then assigned to those clusters. Then an updated centroid is found by finding the mean of each of the clusters. Datapoints can be reassigned to clusters as the centroids are updated. The algorithm finishes when the position of the centroid does not change. The quality of the cluster is determined by minimizing the distance between every pair of datapoints assigned to the same cluster (55). However, k-means is sensitive to the number of clusters chosen, meaning that choosing a random number of clusters could result in high errors and poor cluster results. Another common problem with this clustering method is the creation of clusters that are empty or have very few datapoints. Due to these problems, there is a necessity for additional methods: the elbow method and constrained k-means.

Elbow Method

In an attempt to remedy high errors and poor cluster results, the Elbow method is often implemented. The Elbow method tests a range of k values and seeks to find the optimal number k based on minimized spread between datapoints in the clusters (56).

Constrained k-means

This algorithm is a special case of k-means and has been developed based on a paper by (57). Constrained k-means clustering allows for a minimum number of datapoints to be specified for each of the clusters. This can remedy the problem of clusters that are empty or have very few datapoints.

1.3.6 SHAP Values

SHapley Additive exPlanation (SHAP) values is a method used to increase interpretability of a ML model by measuring the importance of each feature in a dataset.

While many models opt for accuracy above interpretation, SHAP gives both accurate and interpreatable results. This method is based on cooperative game theory, where the SHAP value determines the impact of each feature (or player). It does this by calculating the contribution of each feature for the prediction. The numerical values given to each of the features when summed is equal to the difference of the expected, or baseline, model output and the current model output (58). SHAP values are used to increase interpretability of each feature in the dataset in Experiments 1 and 2.

1.3.7 Model Training and Tuning for Each Experiment

All of the experiments are developed in Python using the Sci-kit Learn (59) and Keras libraries (60). Preliminary testing is performed to determine key parameters, such as number of estimators in RF, and both the learning rate and number of nodes in the hidden layer of NN. The number of estimators tested for the RF model ranged from 50-1000. The results from this test show that there is a negligible difference between the number of estimators after 200. Thus, 200 estimators are used in each of the experiments. A NN model with a single layer is used, with a Rectified Linear Unit activation function in the hidden layer and an Adam optimizer. In Experiments 1 and 2, 75% of the data are used for training and 25% are used for testing and each model is run ten times. In Experiment 3, the RF model is used 49 times, where one cluster is used for training and another is used for testing. When the same cluster is used for training and testing, 75% of the data in that cluster are used for training and 25% are used for testing. The RF model used in this experiment is also run ten times.

Experiment 1

To determine the number of nodes in the hidden layer of the NN and the optimal learning rate, a grid search was conducted. This test used a validation split of 0.2. The number of nodes tested ranged from a minimum of 32 nodes and a maximum of

500 nodes. The learning rates tested were 0.01, 0.001, and 0.0001. After 30 trials, the optimal number of nodes in the hidden state is determined to be 128 with a learning rate of 0.0001.

Experiment 2

For this experiment, all of the models are trained on the data with the engineered features. The same number of estimators is used for the RF model. However, because there are engineered features, a new feature selection was performed. After the features are ranked and selected, these features were used in the NN model. Just as in Experiment 1, the parameters are tuned with the same values, that is the number of nodes in the hidden layer tested ranged between 32 and 500, and the learning rates used are 0.01, 0.001, and 0.0001. After the trials, it was determined that the optimal number of nodes in the hidden state are 256 with a learning rate of 0.0001.

Experiment 3

As stated previously, the objective of this experiment is to begin exploring if the model can be generalized to other watersheds that are geophysically different from the watershed the model is trained on. In order to do this, the watershed is split into clusters. The method used for clustering is k-means. In order to see how many clusters the subcatchments should be separated into, k-means Elbow method was performed. The results from this method show that 7 clusters is the optimal number, as can be seen in Figure 1.3. The 140 subcatchments are then split into 7 clusters. However, when using k-means as the method of clustering, it is common for one or more clusters to be empty or summarize very few datapoints. This results in a very small cluster with few datapoints that is difficult to train on and generalize for other clusters. To make sure that the number of subcatchments in each cluster are at least 5% of the data (i.e., 7 subcatchments), Constrained k-means is used (as described above). The minimum

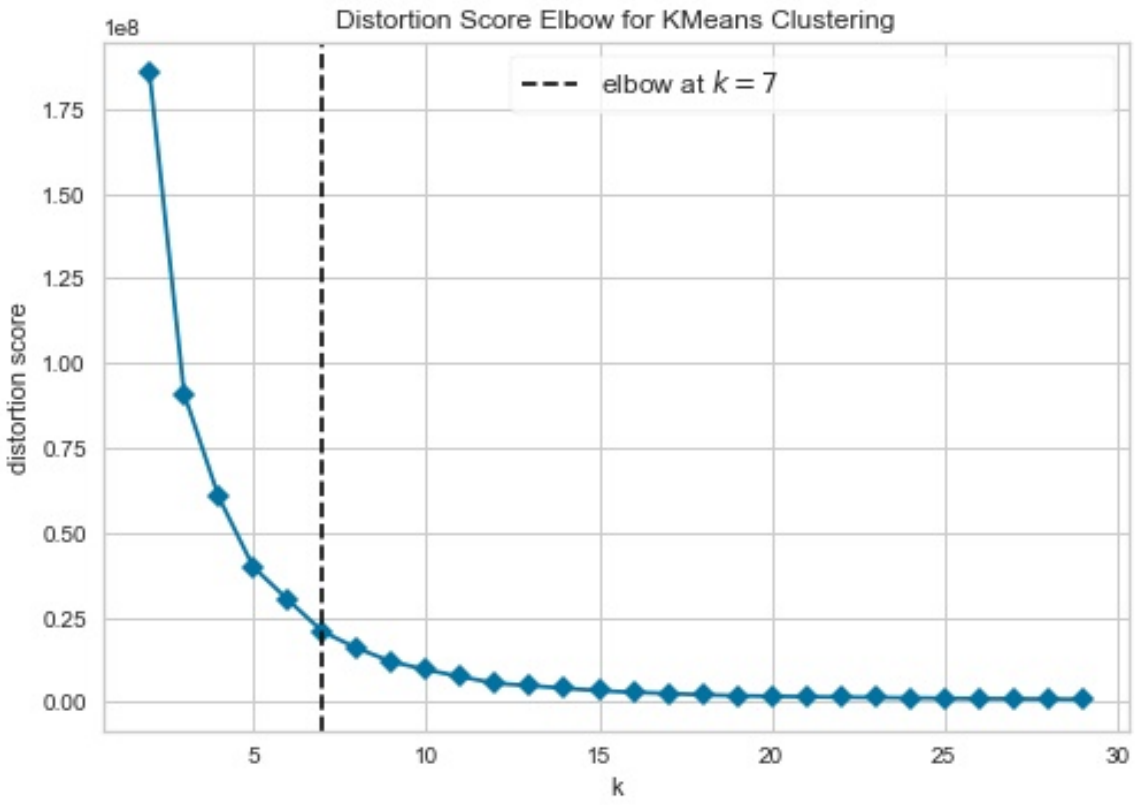

Figure 1.3: Elbow Method for k-means.

size of the cluster was set to 7 subcatchments, and a maximum size was not specified. This would allow for the algorithm to mostly cluster as it normally would, only ensuring that the smallest cluster would not be less than 5% of the data. The means of the resulting clusters are shown in Table 1.4. As can be seen in the table, the area and width of each cluster varies greatly. While %Imperv does not vary quite as much, we do see that Cluster 5 has significantly less impervious area compared to other clusters.

Because this experiment examines the accuracy of clusters trained and tested against each other, the hypothesis would be that the clusters that are trained and tested against themselves would result in the highest accuracy. Additionally, a model trained and tested on clusters with similar geophysical data will perform better than a model trained and tested on dissimilar areas. Figure 1.4 presents the map of the subcatchments separated into the seven different clusters.

1.3.8 Evaluation Metrics

In order to evaluate the results of the models in the experiments, three metrics are used. These metrics are Mean Absolute Error (MAE), Mean Square Error (MSE), and NSE. MAE and MSE were selected as they are widely used to explain the performance of ML models and are easily interpretable (61). NSE was selected as it is a reliable statistic for evaluating the goodness of fit of hydrologic models (62).

MAE

MAE calculates the magnitude of the difference of the predicted value and the true value of a given observation. It then averages the total sum of the absolute errors. The closer this number is to 0, the more accurate the predictions of the model are. The equation for MAE is shown below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y| \tag{1.3}$$

Table 1.4: Average (standard deviation) values for each category in Cluster data

Category	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Area	85.87 (44.37)	125.00 (46.45)	89.91 (32.86)	113.22 (37.10)	164.50 (51.37)	54.96 (33.75)	116.45 (82.34)
%Imperv	14.14 (7.91)	10.72 (6.61)	14.12 (7.98)	12.03 (6.50)	6.10 (4.50)	15.23 (9.19)	13.51 (6.54)
Width	1954.60 (232.00)	6857.67 (376.70)	4374.05 (332.01)	5474.39 (350.95)	8621.71 (1448.24)	1058.67 (356.16)	3071.28 (365.27)
%Slope	4.94 (2.55)	3.60 (2.24)	4.25 (2.32)	3.43 (1.50)	2.40 (1.16)	6.07 (4.17)	4.29 (2.92)
N-Perv	0.25 (0.03)	0.28 (0.03)	0.25 (0.03)	0.25 (0.03)	0.28 (0.02)	0.25 (0.03)	0.25 (0.03)
S-Perv	0.12 (0.01)	0.12 (0.01)	0.12 (0.01)	0.12 (0.01)	0.13 (0.02)	0.12 (0.01)	0.12 (0.01)
MaxRate	6.03 (0.96)	5.92 (0.88)	5.75 (0.90)	6.06 (0.83)	6.52 (0.62)	5.89 (1.24)	6.07 (1.01)
MinRate	0.15 (0.03)	0.15 (0.03)	0.15 (0.03)	0.16 (0.03)	0.17 (0.02)	0.15 (0.04)	0.16 (0.31)
MaxInfil	2.57 (0.40)	2.51 (0.40)	2.44 (0.38)	2.56 (0.38)	2.77 (0.26)	2.50 (0.52)	2.62 (0.45)
# of Subcatch	35	15	22	18	7	18	25

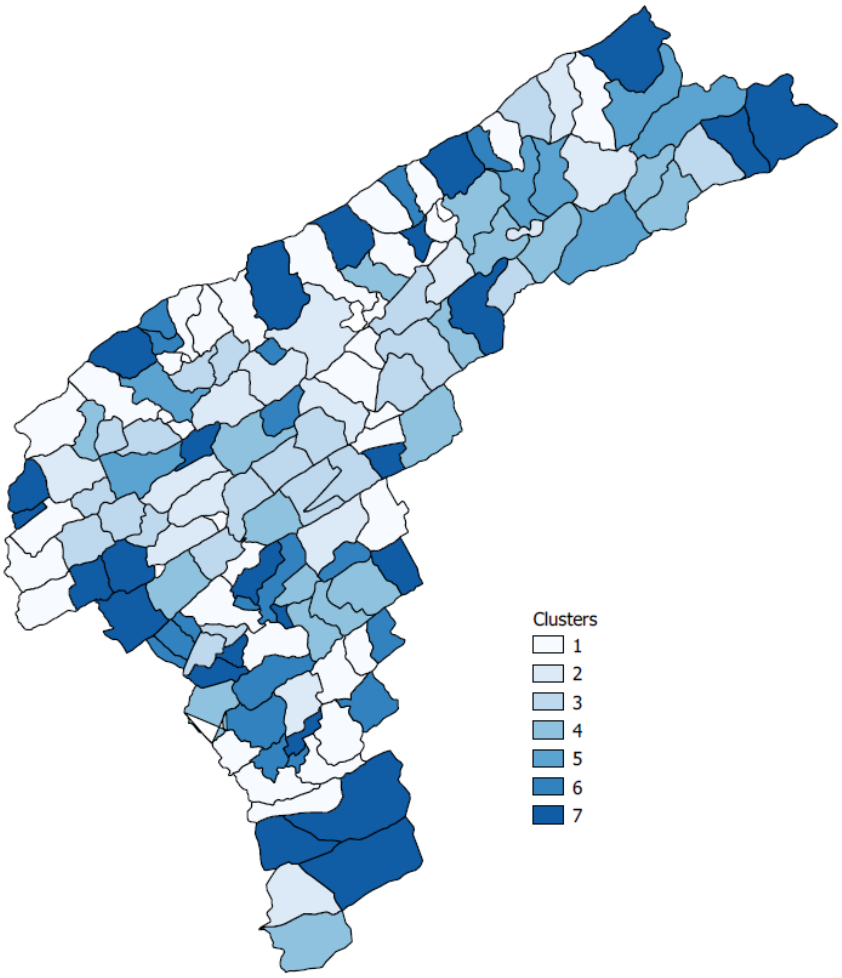

Figure 1.4: First Creek Watershed after subcatchments have been clustered together

where y_i is the predicted value, y is the true value, and n is the total number of data points. This is the main evaluation metric for this study. It will determine and evaluate how well a model predicts. MAE is a more direct representation of sum of error terms. It also treats all errors the same.

MSE

MSE squares the difference of the predicted value and the true value of a given observation. It then averages the total sum of the square errors.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y)^2 \tag{1.4}$$

The squaring in this equation has the effect of inflating or magnifying large errors. This has the effect of punishing the model for larger errors.

NSE

The Nash Sutcliffe Effeciency (NSE) is an assessment of the goodness of fit for a hydrological model and is widely used in this area of study (51), i.e.,

$$NSE = 1 - \frac{\sum(Q_i - \hat{Q}_i)^2}{\sum(Q_i - \bar{Q})^2}, \tag{1.5}$$

where Q_i is the observed flow, $\hat{Q}_i$ is the simulated flow, and $\bar{Q}$ is the mean of the flow. NSE ranges from $-\infty$ to 1 with an NSE of 1 indicating the best match between observed and modeled data.

1.4 Results and Discussion

In this section, the results from the three experiments are presented. Each of the ML models, after some time required for training, were able to make predictions almost instantaneously. These experiments were conducted on a 1.30 GHz CPU.

1.4.1 Experiment 1

The purpose of this experiment is to compare ML and determine how well ML models can estimate total runoff volumes generated by SWMM for the whole watershed (i.e., all subcatchments combined). To accomplish this, the combined data that is created and filtered, as described in Section 1.3.4, is split into training and testing data. The results are presented with the standard deviation for each model to gauge their robustness.

As mentioned before, there are two NN models used: a network with a single hidden and a network with two hidden layers. The optimal number of nodes in each hidden layer is 128 with a learning rate of 0.0001, as determined in Section 1.3.4. When the results of the two NN models are compared, the model with a single hidden layer performed better. These are the results represented in Table 1.5.

Shown in the results, LR performs the worst of all the models in two of the three evaluation metrics. A lower performance is to be expected as LR is trying to fit a linear model to data that cannot be completely described in this manner. Although NN models have shown good performance in other hydrologic studies, this model performed poorly in this study. This could be because only a very small NN was considered. A study by (63) found that their results improved when the number of nodes in the hidden layer increased. However, herein, only one setup for the NN was used as determined by the grid search conducted in Section 1.3.7.

The best performing model is RF, showing the best results in each of the evaluation metrics. RF performing better than DT is expected as RF utilizes many decision trees to provide more accurate results. These results show that, on average, the RF model predicts 0.021 MAE, which differs by 21,000 gallons across the entire watershed. After the model is trained, these predictions are made almost instantly, about eight seconds. However, because this is the average, there could be events skewing the results. For example, if large precipitation events rarely happen, then they may be interpreted

Table 1.5: Comparison of ML models for Experiment 1

Model	MAE [10^6 gallons]	MSE [$(10^6$ gallons$)^2$]	NSE
LR	0.114 (0.003)	0.070 (0.001)	0.541 (0.010)
DT	0.027 (0.001)	0.021 (0.001)	0.861 (0.003)
RF	**0.021 (0.001)**	**0.015 (0.001)**	**0.903 (0.002)**
NN (single layer)	0.095 (0.001)	0.111 (0.004)	-0.002 (0.001)

as outliers and can potentially distort the results of ML models (64). Thus, the predictions during these storms would result in a higher overall MAE.

In order to examine how the amount of rainfall impacts the prediction of the total runoff, the dataset is separated into quartiles (Q1, Q2, Q3, Q4) based on the daily amount of rainfall, where Q1 contains the lowest 25% of daily rainfall, Q2 contains the next lowest 25%, and so on for Q3 and Q4. The statistics of the total daily rain and daily runoff for each of these quartiles are shown in Table 1.6. The resulting MAE for each quartile is also reported.

From these results, we gain more insights about how the amount of rain impacts the predictions of the RF model. Although days with no rain are removed from the data, if there is a small amount of rain in a given hour of day, this datapoint is kept. Thus, this is the type of data found in Q1. This results in an average runoff amount of 100 gallons with a maximum amount of 10,000 gallons. The MAE for Q1 is 0.0008, or an error of 800 gallons. On the other hand, Q4 contains the more extreme precipitation events. It has an average runoff of 393,100 gallons, with an extreme event resulting in 17,300,000 gallons. Therefore, the MAE for Q4 is much higher, 0.0610 (610,000 gallons).

To get a deeper understanding of the contribution of the features in the dataset, SHAP values are plotted in Figure 1.5. These values show that of the top 20 features of the dataset, only four features from the subcatchment data are contributors. These features are Area, %Imperv, MaxInfil, and N-Perv, with Area and %Imperv the only ones in the top 10. The other 16 features are comprised of hourly precipitation data. Runoff volume can only occur due to rainfall, meaning that the amount of rainfall that occurs is highly influential. However, the ranking of the hourly data (i.e., Hour 10 being the most impactful for the results) was interesting and not immediately explainable, prompting further analysis. Extensive descriptive statistics were calculated to see if these could give insight to the manner of this ranking. After thorough examination of the statistics, a concrete justification was not observed. This

Table 1.6: Daily rain amount (in), daily runoff (10^6 gallons), and MAE for Q1-Q4 in Experiment 1

	Q1		Q2		Q3		Q4	
Statistic	Rain	Runoff	Rain	Runoff	Rain	Runoff	Rain	Runoff
Mean	0.0093	0.0001	0.0917	0.0144	0.3265	0.0880	1.0640	0.3931
S.D	0.0098	0.0011	0.0423	0.0219	0.0995	0.0896	0.6223	0.7100
Min	0.0001	0.0000	0.0304	0.0000	0.1802	0.0300	0.5201	0.0000
Max	0.0303	0.0100	0.1801	0.2600	0.5109	0.9000	5.0702	17.3000
MAE	0.0008 (0.0000)		0.0032 (0.0000)		0.0130 (0.0001)		0.0610 (0.0003)	

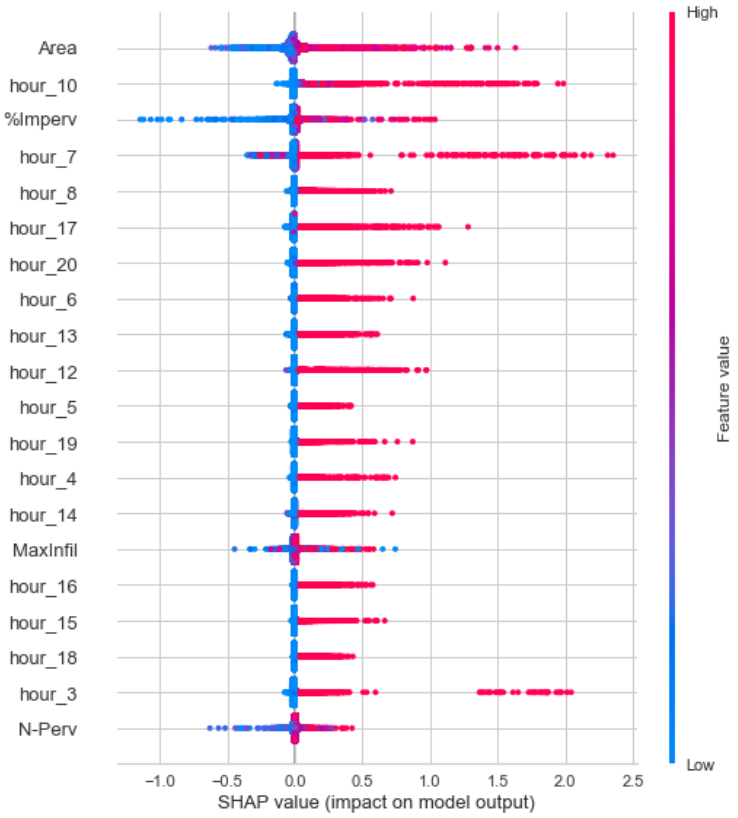

Figure 1.5: Feature ranking with random forest for Experiment 1

remains a point of possible future exploration. These descriptive statistics are shown in the Appendix 3.6.

1.4.2 Experiment 2

There are 27 features engineered for this experiment. A complete description of each of the created features is described in Table 1.3 in Section 1.3.4.

Overall, the ranking of model performance are similar to that observed for Experiment 1, as can be seen in Table 1.7. However, each of the models' predictions are more accurate with the addition of the engineered features. The RF model, shows the most accurate results of the ML models in all of the evaluation metrics. By creating these new features, the accuracy of predicted values from RF are significantly better. This is within 6,000 gallons of the actual output from the SWMM simulation, and compares favorably to the observed difference of 21,000 gallons in Experiment 1.

With the new features, the quartiles are compared as they were in the previous experiment. The statistics of the quartiles in Experiment 2 are displayed in Table 1.8. In Q1, the results seem to indicate that the RF model predicted exceptionally well, with a MAE of 0. And Q4, when compared to the MAE of 0.0610 in Experiment 1, predicts significantly better. The results show a MAE of 0.0209, or 20,900 gallons of water.

As in Experiment 1, a plot of the SHAP values is created and shown in Figure 1.6. The results show that the top 20 features are comprised of nine features from the engineered features, eight features from the subcatchment data, and three features from the hourly precipitation data. The engineered feature total rain gives the most positive impact on the prediction, which is logical.

The first two experiments in this study implement four ML models and attempt to accurately predict the target variable, total runoff, originally generated from the SWMM simulation. While Experiment 1 uses only the three datasets as they are given, Experiment 2 uses feature engineering to create 27 new features for the dataset.

Table 1.7: Comparison of ML models for Experiment 2

Model	MAE [10^6 gallons]	MSE [$(10^6$ gallons$)^2$]	NSE
LR	0.093 (0.001)	0.061 (0.002)	0.551 (0.014)
DT	0.007 (0.001)	0.009 (0.003)	0.933 (0.023)
RF	**0.006 (0.001)**	**0.005 (0.002)**	**0.962 (0.014)**
NN (single layer)	0.120 (0.004)	0.103 (0.010)	-0.001 (0.001)

Table 1.8: Daily rain amount, daily runoff, and MAE for Q1-Q4 in Experiment 2

	Q1		Q2		Q3		Q4	
Statistic	Rain	Runoff	Rain	Runoff	Rain	Runoff	Rain	Runoff
Mean	0.0093	0.0001	0.0917	0.0144	0.3265	0.0880	1.0640	0.3931
S.D	0.0098	0.0011	0.0423	0.0219	0.0995	0.0896	0.6223	0.7100
Min	0.0001	0.0000	0.0304	0.0000	0.1802	0.0300	0.5201	0.0000
Max	0.0303	0.0100	0.1801	0.2600	0.5109	0.9000	5.0702	17.3000
MAE	0.0000 (0.0000)		0.0001 (0.0001)		0.0015 (0.0001)		0.0209 (0.0002)	

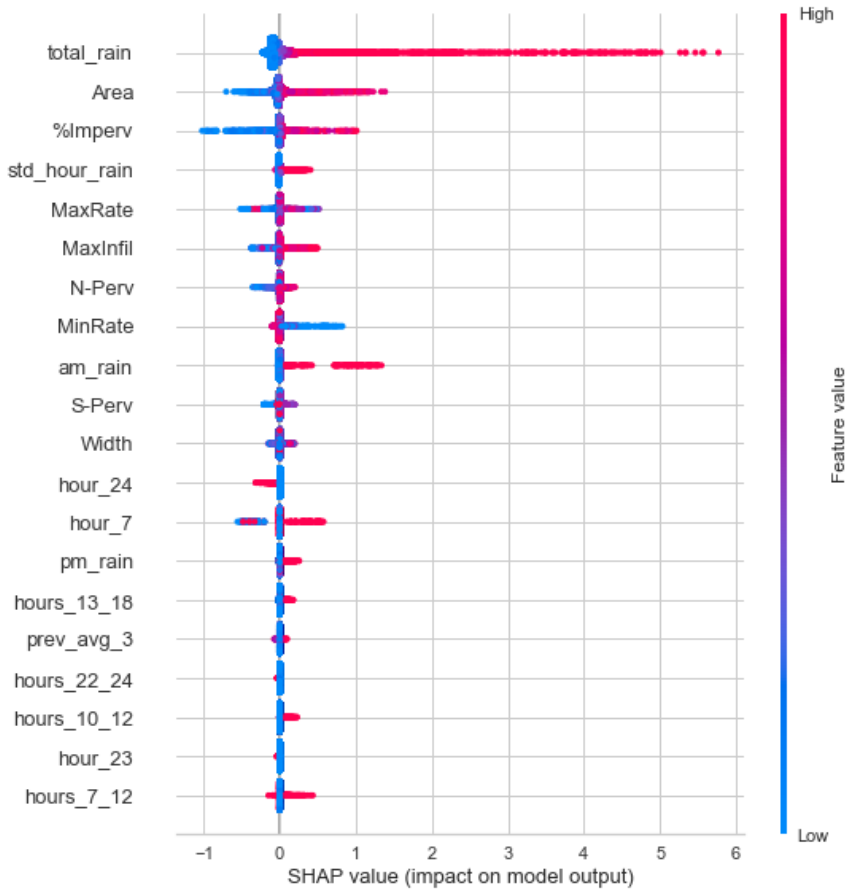

Figure 1.6: Feature ranking with random forest for experiment 2

The four ML models perform in a similar manner in each of the experiments, with RF performing the best followed by DT, NN, and LR respectively. Because LR indicates a low performance, it is assumed the data cannot be completely described with a linear shape. DT and RF are similar models in that they both utilize decision trees to illustrate possible outcomes. However, RF uses a specified number of these trees in order to take a mean of the values from the trees. This usually results in higher accuracy than a single tree, as can be demonstrated in Experiment 1 and 2. NN performs only slightly better than LR in MAE and MSE, but significantly worse in NSE. This indicates that the NN model has a low error when predicting values, but there is little correlation between the variables.

When the SHAP values for Experiments 1 and 2 are evaluated, the results of Experiment 1 depict the hourly precipitation data as the biggest contributors of the observed outcomes. However, this is not the case in Experiment 2 where the subcatchment features and the engineered features have the largest positive impact on the results. It is interesting that while Experiment 1 only considers a few of the subcatchment characteristics as having more positive impacts on the results, Experiment 2 considers significantly more subcatchment features, even with the addition of the new features. Of the three hours of precipitation identified in the SHAP values of Experiment 2, two of them are not recognized as significant features in SHAP values of Experiment 1.

1.4.3 Experiment 3

Because RF performed the best in both of the prior experiments, this model was carried forward to Experiment 3. As mentioned, this model takes one cluster as the training data and a second as the testing data. For example, Cluster 1 is used to train an RF model. Cluster 2 is then used to test the model. This continues for Clusters 3-7. When a cluster is trained and tested against itself, then the 75-25 train-test split

is implemented. Table 1.9 shows the results of Experiment 3 using RF as our ML model and MAE as our evaluation metric.

When the clusters are tested against themselves, in every case, they perform the best. However, when the clusters are tested against other clusters, the accuracy of their predictions vary. For instance, when Cluster 4 is tested on Cluster 2, a MAE of 0.054 (0.009) is given. Of the instances where Cluster 4 is tested against the other clusters, this is the best MAE. The reason for this could be because Cluster 4 is geophysically similar to Cluster 2. When looking at Table 1.4, Cluster 2 and Cluster 4 have closer average values in their Area, %Imperv, Width, and %Slope categories than Cluster 4 has with the other clusters. On the other hand, when Cluster 6 is tested on Cluster 2, a MAE of 0.128 (0.002) is given. This could be because these clusters are more dissimilar in their characteristics. For example, Cluster 6 has the smallest area of the clusters with the highest %Imperv. While the variables of the subcatchment data could explain the differences of performance in some of the instances, it could be that the characteristics still do not fully capture the complexities of the whole area.

1.5 Conclusion

This study investigates applying machine learning models to bypass extensive SWMM simulations which can be computationally expensive. This is particularly critical for planning efforts, where a large number of scenarios are tested and compared, during which the computational time required for SWMM makes efforts implausible. Three datasets were utilized in this effort: hourly rainfall data in Knoxville, TN, from January 1st, 2010, through December 31st, 2020, characteristics of the subcatchments in the First Creek Watershed, and output data (total daily runoff volume) from SWMM in this watershed. Three different experiments were performed to see how accurate the ML models are compared to the SWMM output and to see if these models can be generalized to other areas with different geophysical characteristics.

Table 1.9: Results of Experiment 3 (MAE [10^6 gallons]). Row cluster used for training, column cluster used for testing.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Cluster 1	**0.018 (0.001)**	0.074 (0.002)	0.066 (0.001)	0.067 (0.002)	0.043 (0.002)	0.057 (0.001)	0.097 (0.001)
Cluster 2	0.069 (0.001)	**0.023 (0.001)**	0.056 (0.009)	0.054 (0.009)	0.040 (0.006)	0.128 (0.002)	0.118 (0.005)
Cluster 3	0.056 (0.006)	0.070 (0.001)	**0.016 (0.001)**	0.064 (0.001)	0.045 (0.003)	0.068 (0.001)	0.091 (0.001)
Cluster 4	0.074 (0.003)	0.067 (0.001)	0.069 (0.001)	**0.026 (0.017)**	0.041 (0.014)	0.096 (0.001)	0.103 (0.004)
Cluster 5	0.063 (0.001)	0.048 (0.001)	0.048 (0.001)	0.064 (0.001)	**0.015 (0.001)**	0.098 (0.001)	0.097 (0.001)
Cluster 6	0.067 (0.001)	0.074 (0.001)	0.070 (0.001)	0.081 (0.001)	0.057 (0.001)	**0.019 (0.001)**	0.100 (0.001)
Cluster 7	0.089 (0.001	0.062 (0.001)	0.051 (0.001)	0.076 (0.001)	0.035 (0.001)	0.088 (0.001)	**0.024 (0.001)**

The real world data is utilized to evaluate the novel method developed and assess the ML models predictions.

The first two experiments examine the overall watershed in four different ML models, with NN also being tested with different layer numbers. The second experiment utilizes feature engineering, a technique that creates new features from historical data, to boost the performance of a ML model. The results from this study indicate that with the addition of engineered features, RF can predict the total runoff with a MAE of 0.006 (0.001). These first two experiments show that ML models can be used as a compliment for SWMM with a great deal of accuracy. The third experiment tests the generalizability of the ML models by clustering subcatchments into seven clusters. These clusters are used in RF, where one cluster is used for training and another is used for testing. While the clusters performed the best when tested against themselves, the experiment may show promise of generalizing ML models to geophysically different areas.

The conclusions from this study are three-fold: (*i*) Our proposed ML models can accurately predict total runoff in a fraction of the time compared with SWMM; (*ii*) The reduction in computational time realized in this study can pave the way for more robust scenario analysis and optimization in strom water management; and (*iii*) This study demonstrates that ML models may be used as a complement for monitored data in watersheds that are not instrumented for field data collection.

Chapter 2

Examining System Changes at TVA Fossil Plants

2.1 Introduction

The power sector is highly sensitive to changes in water temperature and availability, especially at fossil power plants, which depend on water for multiple phases of energy generation (65). In these plants, water from nearby rivers or reservoirs is drawn in, heated to create steam that drives turbines, cooled, and then either recycled or discharged back into the environment. This reliance on water makes fossil plants vulnerable to shifts in water temperature, with significant implications for their efficiency and environmental impact. As water temperatures rise, particularly in the summer months, plant efficiency declines, requiring more coal and leading to higher emissions of pollutants such as CO_2 and SO_2 (66). While efforts have been made to reduce emissions and transition away from fossil fuels, there are still over 3,400 operational fossil plants in the U.S. (67). These plants remain critical to current energy production, forming the backbone of many power systems.

Fossil plants are subject to regulatory limits on the temperature of water they discharge, with state environmental agencies mandating that downstream water

temperatures must not exceed 90°F to protect aquatic life. These limits place additional constraints on the plants, particularly during the summer when intake water temperatures can approach or exceed regulatory thresholds. When this happens, plants are forced to reduce output or temporarily shut down, impacting both their efficiency and the broader power grid. Given the critical role water temperature plays in plant operations, fossil plants must have as much information about environmental and temperature trends as possible.

Several studies have analyzed hydrologic trends (68; 69; 70; 71; 72; 73). The Mann-Kendall test is a widely recognized, robust, non-parametric method for detecting monotonic trends in environmental time series (74; 75; 76). Other papers have examined relationships with other variables such as air temperature and flow rate to see how these impact changes in water temperature (77; 78; 79). Webb, Clack, Walling (78) found that water temperatures closely follow air temperatures, though higher flow rates could reduce the water temperature's sensitivity to air temperature.

The main objectives of this study are to employ robust statistical methods, including the Mann-Kendall test and correlation analysis, to detect long-term trends and fluctuations in water temperature. Additionally, it examines the relationships between water temperature, air temperature, and flow rate to identify the drivers behind observed thermal changes. Through this approach, the study provides valuable insights into the changing thermal environments of fossil plants, with implications for their operational resilience and environmental sustainability.

2.2 Data and Methods

The Tennessee Valley Authority (TVA), which operates five fossil power plants across Tennessee and surrounding areas, relies on nearby rivers and reservoirs for water intake (80). These plants generate enough power to supply 3.3 million homes annually. Water from these sources is drawn in, heated to about 1,000°F to create high-pressured steam, which drives turbines. The steam is then cooled and either reused or released

back into the environment. Efficiently managing the temperature of both intake and discharged water is crucial for maintaining operational efficiency and compliance with environmental regulations.

This study analyzes three fossil plant locations operated by TVA in Tennessee. The first is the Gallatin Fossil Plant located upstream from the Old Hickory Reservoir northeast of Nashville, TN. The reservoir contains 22,500 acres of water. The second location is the Cumberland Coal Plant, located on the Barkley Reservoir upstream of the Cheatham Dam. All the waste heat at this location is returned to the Cumberland River water. The third location is the Kingston Fossil Plant and is located upstream from the Watts Bar Reservoir, but the distance so great from the Watts Bar Dam that the Emory River operates as the feed source for water intake. The map of these locations are shown in Figure 2.1.

Three different pieces of data for each location were collected. The water temperature and flow rate data from three different water sources in Tennessee are provided by TVA. The hourly water temperature data included nine years (2008 - 2016) from the Emory River at Oakdale, which is where the water intake happens for the Kingston plant, 17 years (2008 - 2024) from the Cheatham Dam, which the Cumberland plant is upstream from, and 17 years (2008 - 2024) from the Old Hickory Reservoir at Cordell Hull Dam, which the Gallatin plan is upstream from. To get air temperature data, Meteostat Python Library was used to retrieve data from weather stations close to the fossil plants (81). Meteostat uses data provided by national weather services such as the National Oceanic and Atmospheric Administration (NOAA). The data for each site is given in hourly increments. A summary of this data is shown in Table 2.1.

After evaluating the datasets from each site, the following years of data were used in our experimentation: the Gallatin plant used data from 2008-2024, the Cumberland plant used data from 2008-2024, and the Kingston plant used data from 2008-2016. The three types of data are combined to create one dataset per site.

Figure 2.1: Map of the three TVA fossil plant locations. (1)

Table 2.1: Types of data collected for each location.

Site Location	Measure	Measure Location	Time Frame
Gallatin Fossil Plan	Water Temperature	Cordell Hull Dam	2008 - 2018
	Air Temperature	Nashville MET Data	2008 - 2020
	Water Flow Rate	Cordell Hull Dam	2008 - 2018
Cumberland Fossil Plant	Water Temperature	Cheatham Dam	2008 - 2019
	Air Temperature	Nashville MET Data	2008 - 2020
	Water Flow Rate	Cheatham Dam	2008 - 2019
Kingston Fossil Plant	Water Temperature	Emory River at Oakdale	2008 - 2016
	Air Temperature	Oak Ridge MET Data	2008 - 2019
	Water Flow Rate	Emory River at Oakdale	2008 - 2016

2.3 Methods

2.3.1 Data-Preprocessing

Before the data can be used in analyses, there are some outliers that must be taken care of. These outliers can occur for several reasons. First, the measurements rely on gauges that may malfunction, be vandalized (as has happened at the Kingston fossil plant), or cease functioning altogether. Second, when it is noticed that a gauge is not working properly, there is opportunity for these values to be inputted manually. When this happens, there are known instances where the temperature is inputted as Celsius instead of Fahrenheit, which is what the temperature is typically recorded in. An example of gaps in the data and an instance of inputting temperature in Celsius is shown in Figure 2.2. It is imperative to correct these outliers before continuing to analyze the data so that the results can be more meaningful.

Outliers are identified through statistical analysis and manual inspections. Once outliers are confirmed, they are either corrected (if a known error was present, such as Celsius recording) or removed.

For the purposes of this study, we focused only on the summer months, from June 1st to September 30th. These months are of particular concern, as high water temperatures during this period have the greatest impact on the efficiency of fossil plants and the surrounding ecosystems.

2.3.2 Study Design

To investigate how climate change is impacting the system, we employ a series of statistical methods to analyze water temperature trends, examine relationships with environmental variables, and identify patterns indicative of a changing environment.

Descriptive Statistical Analysis As an initial step, basic descriptive statistics were calculated to each summer of our dataset (June 1st to September 30th) including

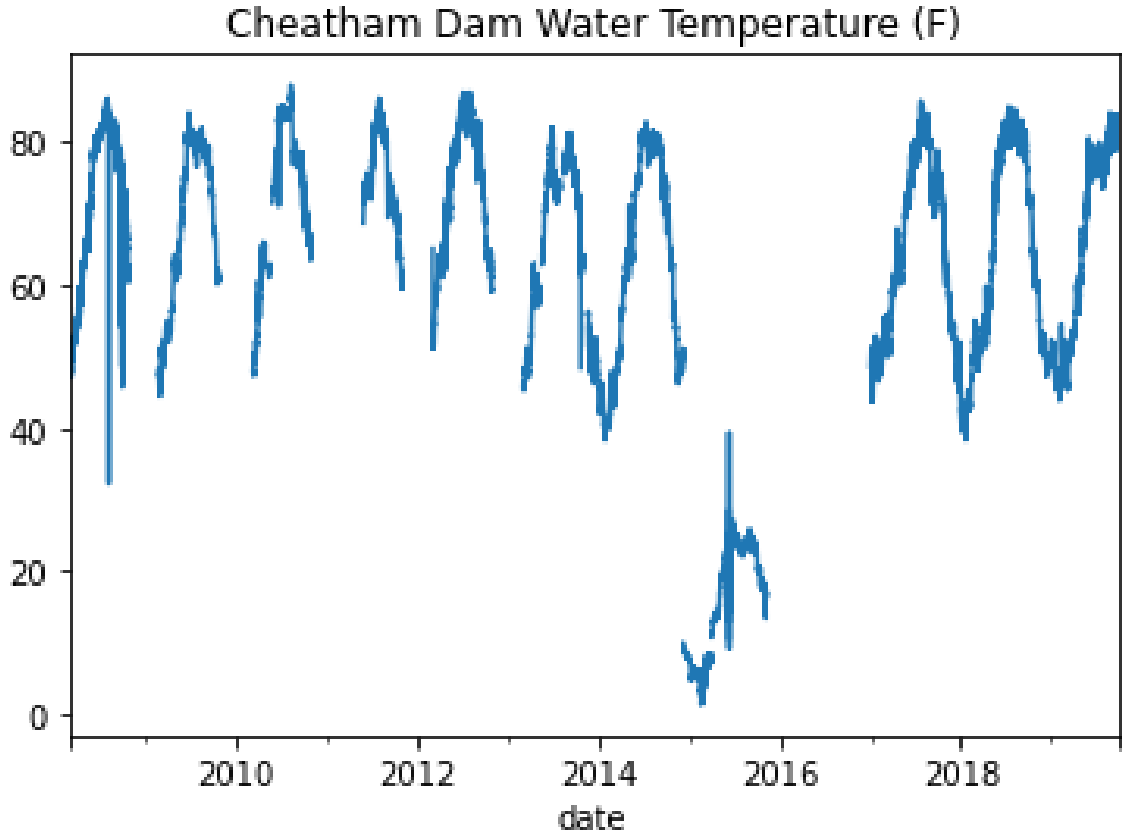

Figure 2.2: Example of data with missing and incorrect values from a TVA fossil power plant location

the mean, quartiles, minimum, and maximum water temperatures. By comparing these statistics across the 9-17 years of data available at each location, we aim to observe any significant changes over time that could indicate trends in temperatures. This analysis serves as a preliminary step in understanding how the water temperature distribution has shifted over the years, highlighting any potential increases in the extremes (minimum or maximum values) or overall mean temperatures. To better visual this analysis, box plots were created for each dataset, offering a comparative view of the data's range, central tendencies, and potential outliers across different locations.

Time Series Decomposition To examine the underlying structure of the data, ARIMA decomposition is performed on each dataset. This technique separates the time series into its three primary components: trend, seasonality, and residuals. The decomposition provided insights into the long-term trends and seasonal variations in water temperatures, air temperatures, and flow rates, enabling a clearer interpretation of temporal patterns (82).

Mann-Kendall Test The Mann-Kendall test calculates a test statistic S that measures the direction and significance of a trend. The significance of S is evaluated using a p-value, which tells us if the trend is statistically significant or not (83; 84). This method is used to evaluate whether a monotonic trend exists over time. The results of this test provided quantitative evidence of increasing or decreasing trends in water temperature.

Correlation Analysis Pearson's correlation coefficient was calculated to investigate the relationships between key variables: water temperature, air temperature, and flow rate. This analysis measures linear correlation between two variables. It quantifies the strength and direction of these relationships, offering insights into how external factors influence water temperature dynamics (85).

2.4 Results

2.4.1 Descriptive Statistical Analysis

In this section, we present the results of the descriptive analytics for water temperature at each of the fossil plant locations. Tables of these statistics for Gallatin, Cumberland, and Kingston are shown in A.5, A.6, and A.7, respectively.

The analysis of the annual box plots, shown in Figure 2.3, reveals notable variations in water temperatures across different years and sites, highlighting key years and patterns of significance. The year 2011 stands out across all the fossil plant locations with the largest interquartile range observed. This indicates a year of substantial temperature variability, suggesting significant environmental or operational dynamics that affected all the sits. Both Gallatin and Cumberland exhibit show a high median temperature accompanied by frequent extreme temperature readings that fall outside the typical range in 2012. Because we do not see this at Kingston, it could be attributed to the fact that Gallatin and Cumberland are closer in distance. At Gallatin, water temperatures show a pattern of rising and falling over several years, followed by periods of more subtle fluctuations. Cumberland displays an initial trend of decreasing temperatures, which then stabilizes in the later years. The peak temperatures in 2008 could be attributed to residual impacts from the 2007 drought, suggesting a delayed recovery in water temperature stability. Due to the shorter dataset available for Kingston, discerning a clear trend is challenging. The available data does not show significant fluctuations, which may limit the ability to draw comprehensive conclusions for this site.

2.4.2 Time Series Decomposition

By employing ARIMA-based seasonal decomposition, we have elucidated the underlying trends in water temperature, air temperature, and flow rates at the fossil plants

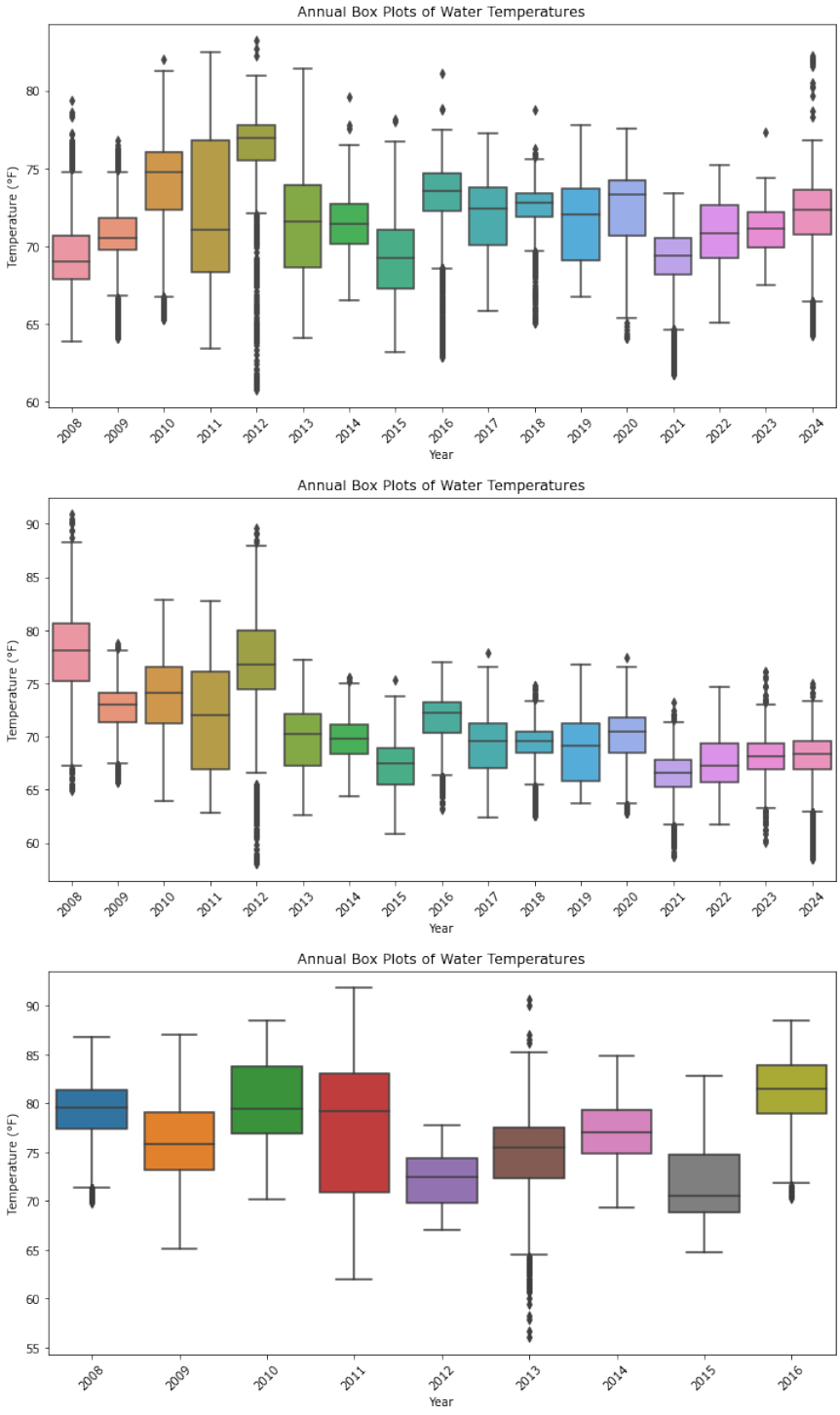

Figure 2.3: Box plots to visualize the descriptive statistics at Gallatin (top), Cumberland (middle), and Kingston (bottom).

during summer months. The period length for the ARIMA model was set to 2928, representing the total hours (24 hours $\times$ 122 days from June 1 to September 30).

In Figure 2.4, the trend component for water temperatures echoes the patterns observed in the box plots, exhibiting a general decrease followed by stabilization in more recent years. Conversely, the flow data show a distinct upward trend over the years, indicating an increase in flow rates over time. At Cumberland, as depicted in Figure 2.5, there is a pronounced trend of decreasing water temperatures over the years, coupled with an increase in flow rates. The trends observed at Kingston, presented in Figure 2.6, are less clear-cut, making them challenging to interpret definitively. However, these trends—or the apparent lack thereof—can be quantitatively assessed using the Mann-Kendall test to determine the statistical significance of observed changes. These decomposed components reveal that while air and water temperatures have shown specific directional trends at the sites, the flow rates present contrasting behaviors.

2.4.3 Mann-Kendall Test

In analyzing the water temperature data from Gallatin, Cumberland, and Kingston, the Mann-Kendall test was employed to detect the presence of any significant trends. The test results, summarized in Table 2.2, reveal a statistically significant decreasing trend in water temperatures at all locations. This is evidenced by the p-values of 0.0 across the board. This indicates that the observed declines in temperature are highly unlikely to be due to random variations in the data.

The S statistics, with values of -0.046 for Gallatin, -0.377 for Cumberland, and -0.080 for Kingston, confirm the visual downward trends identified through seasonal decomposition analysis. Cumberland exhibits the most substantial decrease in water temperatures over the period studied, with the highest magnitude of the S statistic.

This trend analysis highlights the need for ongoing monitoring and investigation into the factors contributing to these temperature decreases.

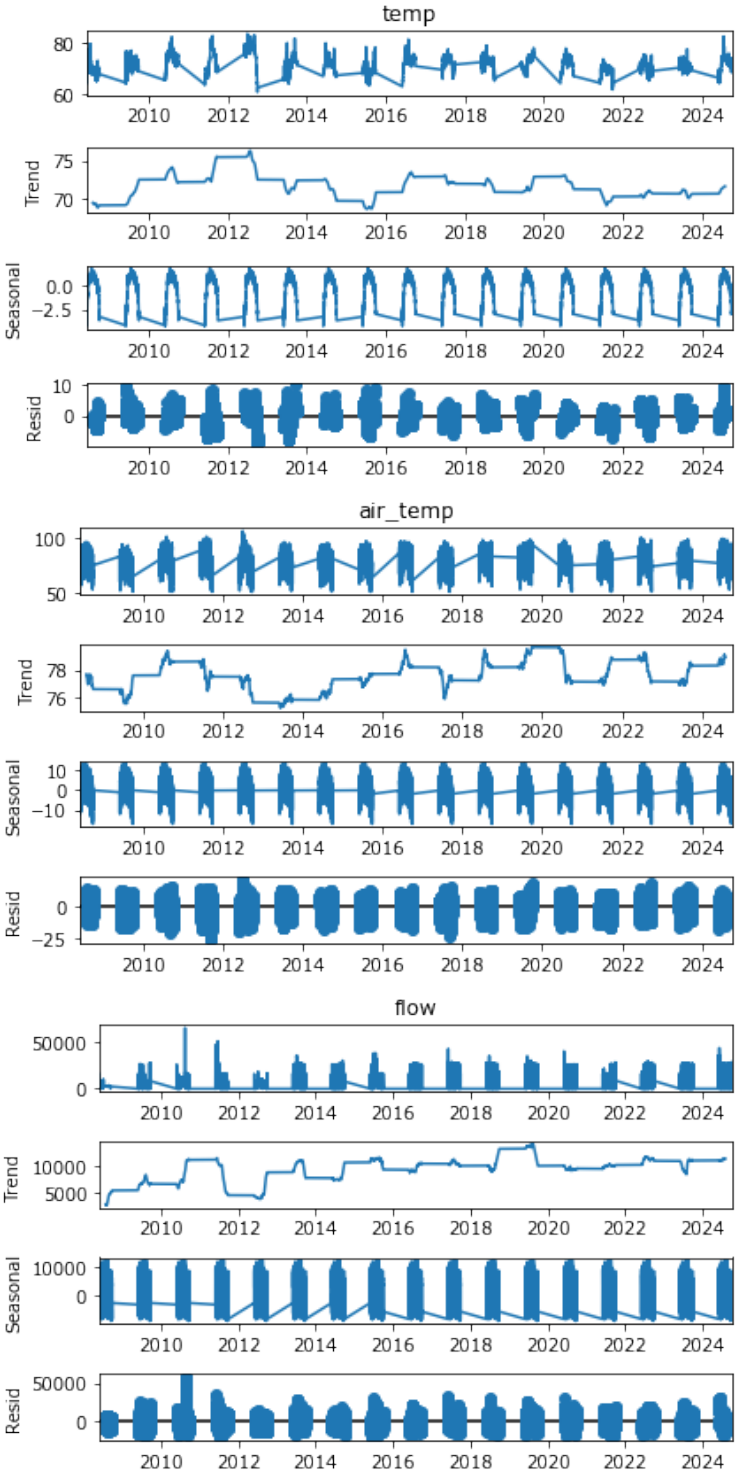

Figure 2.4: Gallatin time series decomposition shows the trends, seasonality, and residuals of water temperature data (top four plots), air temperature data (middle four plots), and flow data (bottom four plots).

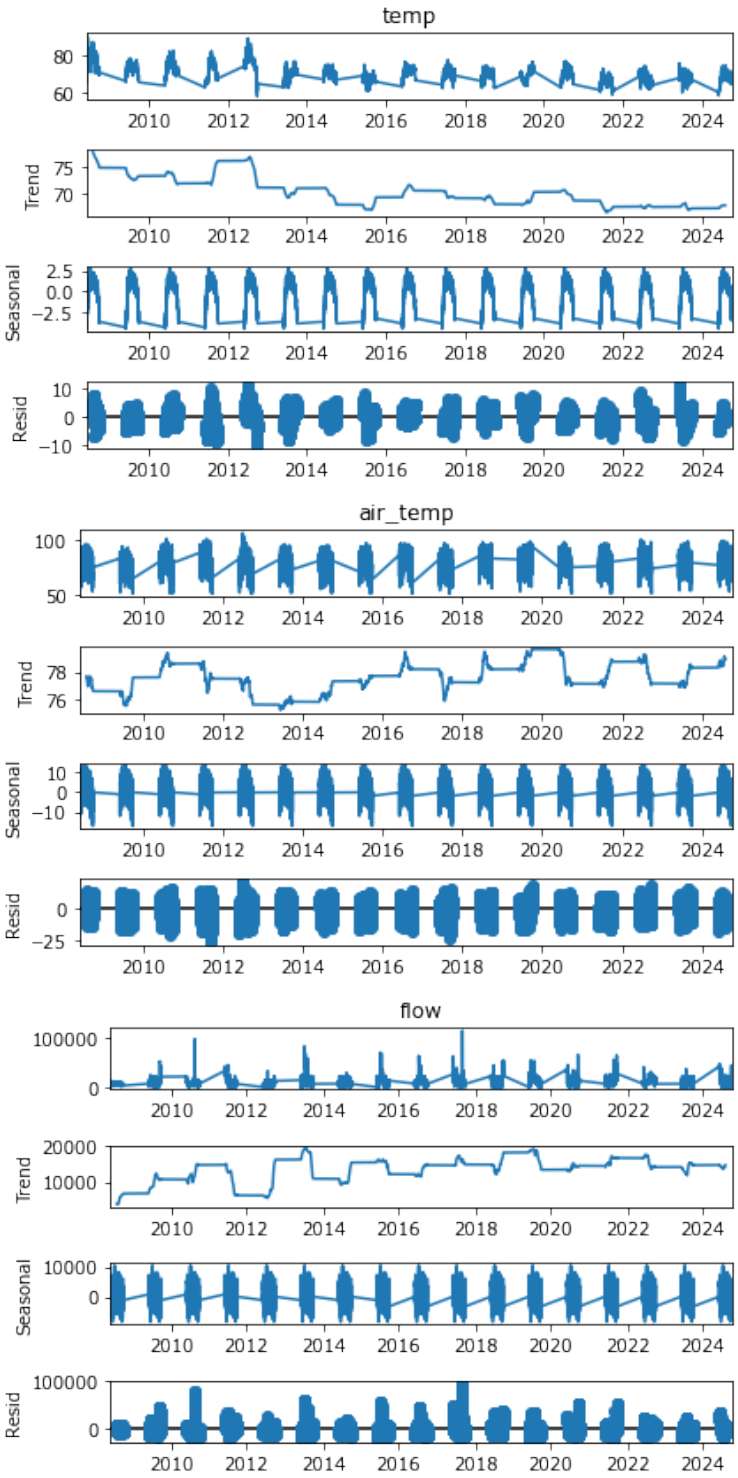

Figure 2.5: Cumberland time series decomposition of water temperature data, air temperature data, and flow data

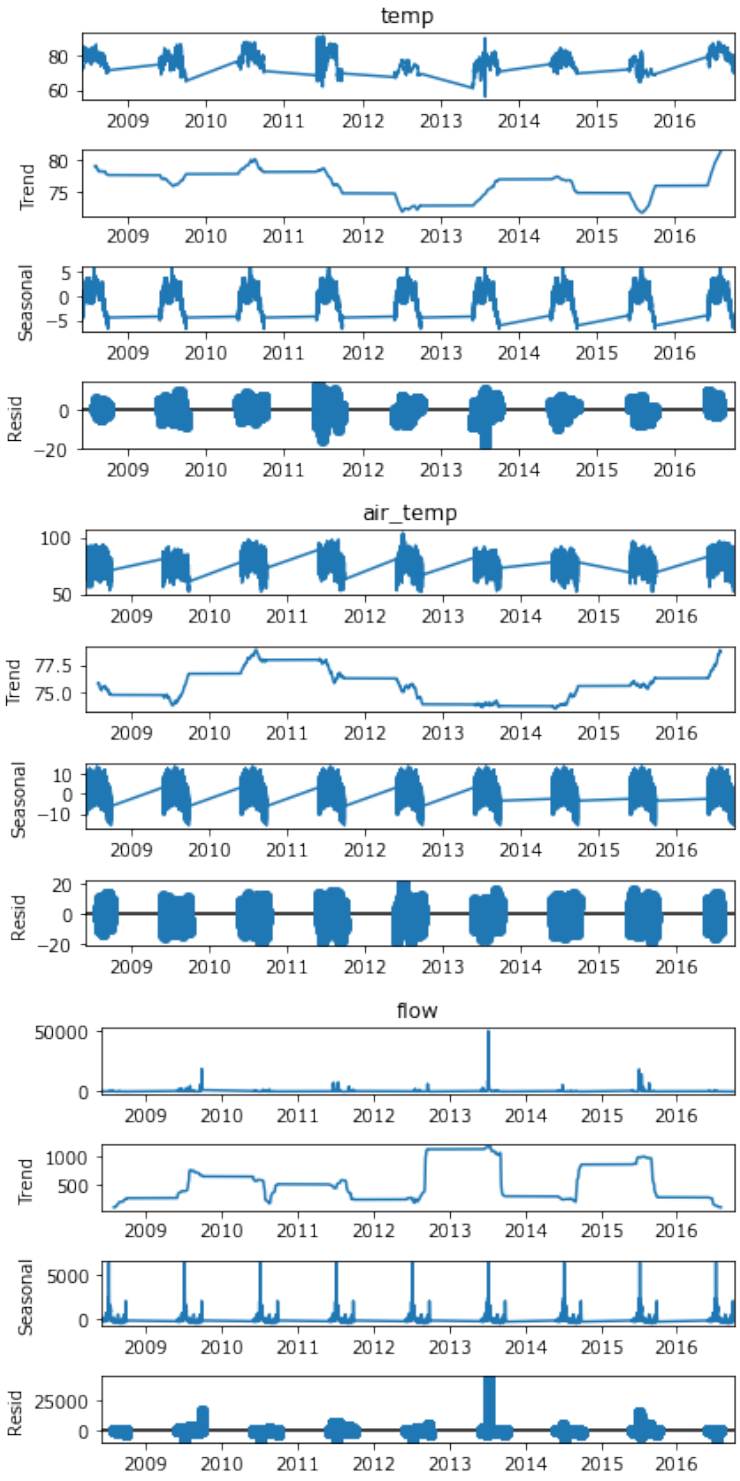

Figure 2.6: Kingston time series decomposition of water temperature data, air temperature data, and flow data

Table 2.2: Mann-Kendall test on water temperature data at all fossil plant locations

Gallatin		Cumberland		Kingston	
Trend:	decreasing	**Trend:**	decreasing	**Trend:**	decreasing
p-value:	0.0	**p-value:**	0.0	**p-value:**	0.0
S**:**	-0.046	S**:**	-0.377	S**:**	-0.080

2.4.4 Correlation Analysis

Correlation analysis was conducted to examine the relationships between water temperature, air temperature, and flow rate at three fossil plant locations. The results of this analysis can be seen in Figure 2.7 Our findings reveal variable degrees of association among these parameters across the sites.

At Gallatin, the results show weak to moderate correlations, with air temperature positively related to both water temperature and flow rate. Interestingly, air temperature had a stronger correlation with flow rate than water temperature with a Pearson's coefficient of 0.23 and 0.16 respectively. Water temperature is inversely related to flow rate with a coefficient of -0.19. Cumberland displayed the strongest negative correlation between water temperature and flow rate, with a coefficient of -0.42 indicating a pronounced cooling effect of higher flow rates on water temperatures. The coefficient for the correlation between air and water temperature at this location is also higher than that of Gallatin. Kingston exhibits a more substantial positive correlation between water and air temperatures compared to other sites, with a coefficient of 0.4, suggesting a slightly stronger linkage between atmospheric conditions and aquatic temperatures at this location. Conversely, water temperature and flow rate are negatively correlated, with a coefficient of -0.26, consistent with expected physical dynamics where increased flow can lead to reduced water temperatures.

2.5 Conclusion

This study utilized robust statistical methods to analyze long-term trends and fluctuations in water temperature, alongside examining the relationships between water temperature, air temperature, and flow rate across three TVA fossil plant locations: Gallatin, Cumberland, and Kingston. We utilized two methods to analyze the trends of the water temperatures: Time series decomposition and Mann-Kendall

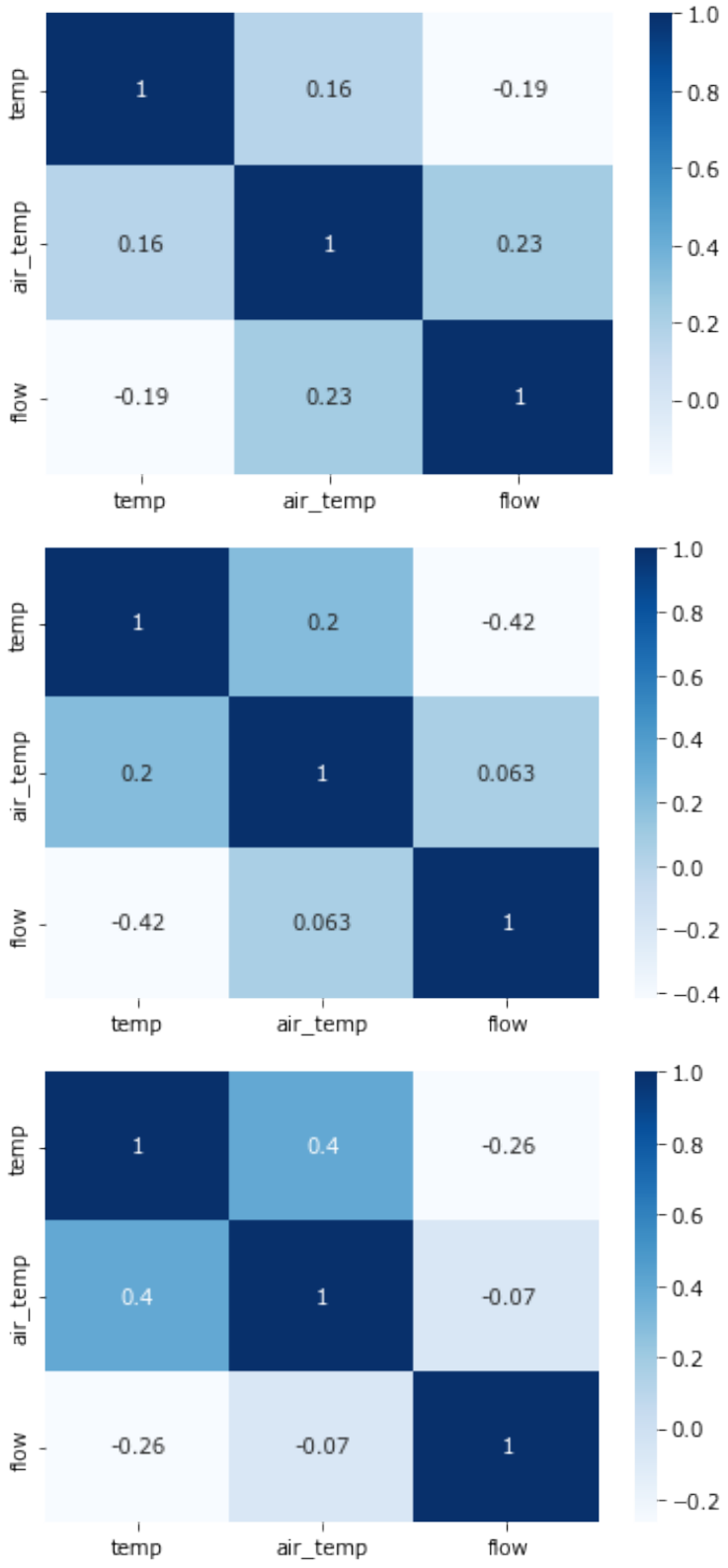

Figure 2.7: Pearson's Correlation Coefficient on data from Gallatin (top), Cumberland (middle), and Kingston (bottom)

test. The result of the Mann-Kendall consistently revealed a statistically significant decreasing trend in water temperatures at all sites, as evidenced by p-values of 0.0. The trend components extracted from ARIMA models supported these findings, showing general decreases in water temperatures, with some stabilization in recent years at Gallatin and Cumberland.

Correlation analysis with Pearson's Correlation Coefficient was conducted to understand the relationships between the available data from the fossil plants. Moderate to weak positive correlations were observed at all sites, suggesting that air temperature somewhat influences water temperature but is not the sole driver. Negative correlations across the sites indicate that increased flow rates contribute to cooling water temperatures.

Understanding the drivers behind water temperature changes is crucial for operations at these plants and predicting their ecological impact. This study highlights the interplay of factors impacting temperatures at fossil plants. The results aid in providing essential information to TVA fossil plants in their consideration of plant operations and their impacts.

Chapter 3

Forecasting Hourly Water Temperatures for Fossil Plant Efficiency and Aquatic Life

3.1 Introduction

Fossil power plants play a vital role in the energy landscape, yet their operations are highly sensitive to changes in water temperature (65). These plants rely on water drawn from nearby rivers and reservoirs for cooling and steam generation, the largest amount of required for cooling (86). Because of this dependency on nearby bodies of water, it makes temperature fluctuations a critical operational variable. Regulatory requirements further complicate this issue, as state environmental agencies mandate that discharged water temperatures must not exceed 90°F to protect aquatic ecosystems. During the summer, when intake water temperatures approach or exceed these thresholds, plants often face reduced output or temporary shutdowns, impacting efficiency and grid reliability.

Given the pivotal role water temperature plays, reliable forecasting of future conditions is essential for planning and decision-making in plant operations. The

Tennessee Valley Authority (TVA) currently uses a simple model to forecast water temperatures, predicting only a single daily temperature and failing to capture hourly fluctuations or longer-term warming trends. These limitations hinder their utility for real-time operational decisions during critical periods.

Because of the growing influence of climate change on water temperatures, more advanced methods are needed to predict future trends with greater accuracy. Machine Learning (ML) presents a promising solution for this challenge. ML algorithms learn from historical datasets to make accurate predictions about future conditions. In the context of water temperature forecasting, ML has gained increasing attention and success in recent years (87; 88). Various studies have demonstrated the efficacy of ML models in predicting water temperatures with higher precision than traditional hydrodynamic models (89; 90; 91; 92; 93). Many studies employ neural network models such as Long Short-Term Memory (LSTM) models as these models are flexible and able to predict complex patterns (94; 95; 96). Zhang et al. (97) compared LSTM results to a support vector regression and a multi-layer perceptron regression. They found that their LSTM model outperformed the other models. Another method often used with time series data for prediction is Autoregressive Integrated Moving Average (ARIMA) model, which uses autoregressive and moving average components of historical data (98). ARIMA was the most accurate 3-5 day forecasting method in a study of Delaware River water temperatures (99). Faruk also showed that using ARIMA modeling is highly accurate in forecasting water temperatures by applying them to the Buyuk Menderes River in Turkey (100). 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 (101). Using the autoregressive nature of time series data, Rabi, et al. (102), Toffolon and Piccolroaz (103), and Caissie, et al. (104) 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 (105), linear regression (106), decision trees (107), and random forests and (108).

Our case study will focus on three fossil power plants in Tennessee that are in operation under the Tennessee Valley Authority (TVA). The first fossil plant location is the Gallatin Fossil Plant, located on the Cumberland River near Nashville, TN. This site operates four turbo-generator units with a combined summer net generating capacity of 976 MW (109). A summer net capacity refers to the amount of power the power plant can produce in an average day in the summer, minus the amount of energy the plant itself uses. The second fossil plant is the Cumberland Fossil Plant and is downstream from the Cheatham Dam. This location is the largest generating asset in the fleet of TVA plants, utilizing two turbo-generator units with a maximum gross output of 2,470 MW (110). The final fossil plant location is the Kingston Fossil Plant, which is located on the Clinch River arm of Watts Bar near Kingston, TN. At this location, there are nine turbo-generators in operation and they produce a summer net capability of 1,398 MW (111).

The purpose of this study is to create a model, utilizing ML, that can be used at each of the sites to provide accurate hourly water temperature forecasts. Having this information available to the site will allow them to [1)]Make adjustments to their operations so they can ensure they are optimizing their efficiency and [2)]Guarantee they remain within their discharge water temperature limit to protect the aquatic life surrounding them.

3.2 Data and Methods

This section introduces the methods used in this study. An overview of these methods can be seen in Figure 3.1.

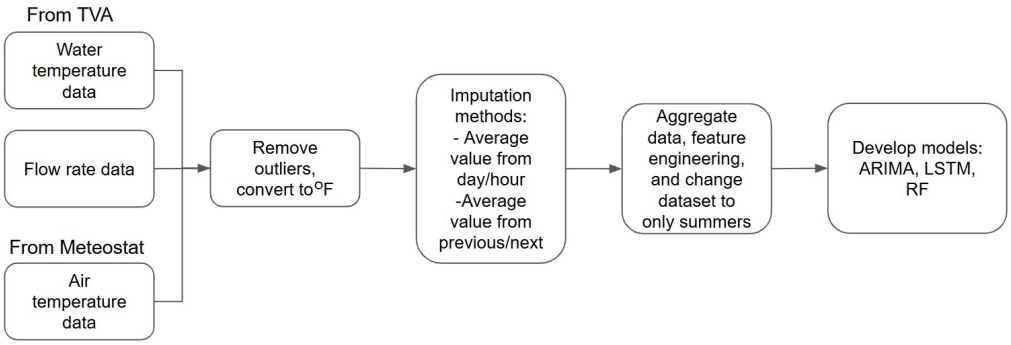

Figure 3.1: Overview for the methods of this study. We use three sets of data. We acquired water temperature and flow data from TVA and air temperature data from Meteostat, which uses data from NOAA. Next, the data must be cleaned by removing outliers, converting instances of metric data, and imputing missing values. After this, we can aggregate data, create new features, and reduce the datasets to contain only summer data. Finally, we can use our forecasting models to predict the next week of temperatures.

3.3 Data and Data Preprocessing

This study utilizes water temperature, air temperature, and flow data from each of the fossil plant locations. At Gallatin and Cumberland, we have 17 years worth of data, and at Kingston we have 9 years of data. The data for this study is the same that are summarized in Tables A.5, A.6, A.7.

ML models require extensive amounts of data to produce meaningful results. As mentioned, we have many years of data at each site that we can utilize. However, the water temperature data at our locations have proven to be undependable when compared to the other types of data. As this is to be our target variable in the models, it is crucial that this data is reliable. Though there is water temperature data from specific sites, there are many missing, extreme, and unreliable values in the data which will require preprocessing before it can be used. There are many reasons why the data may currently have these attributes, such as malfunctioning or failing sensors, manual input of temperatures in Celsius as opposed to Faharenheit, or vandalism. It is important to correct these before continuing to create a useful tool with ML models. This is a challenge that we must overcome so that there are a sufficient number of 'good' training examples for the ML models.

There are two methods used for imputation, depending upon what makes sense for the data and pattern of missing data points.

Imputing Average Value from Date and Hour The main method that was used to impute missing datapoints. When a missing datapoint is found, the average of the available data for that date and hour in other years will be used. By using this method, we are able to retain the daily fluctuations of the temperatures. This method was used in water temperature and air temperature for every location.

Imputing Average Value from Previous and Next Datapoint This method looks at the two surrounding datapoints, the hour previous and the next hour, and imputes the average of these values. This imputation was used only on the flow data

at the Cumberland fossil plant. This is because all of the missing values happened regularly at the same hour on certain days of every year. This is a unique situation in which the other imputation method would not work.

After we impute missing values, only the summer months of the combined datasets will be examined. That is, only data from June 1st to September 30th will be considered. This is because (*i*) this is the timeframe when making sure the water stays below the given temperature limits is most important and (*ii*) the spring and fall months are influenced by reservior fill and drawdown, respectively, as well as cold fronts leading to these seasons not being ideal for analysis.

Feature Engineering Once the data has been cleaned, we can add more variables through feature engineering. The purpose of feature engineering is transform data into meaningful features that will help the accuracy of ML models (112). After this step, we are able to add 13 features which include year, month, day, hour, season, monthly average temperature, and one-hot encoding of the days of the week. Now that the datasets are complete, we are able to use them in our models. We use these models to predict one week of data, as this is the goal of TVA's current forecasting model.

3.4 Forecasting Models

Seasonal Autoregressive Integrated Moving Average with Exogenous Factors ARIMA is the combination of two processes: Autoregressive (AR) and moving average (MA). It was introduced by Box and Jenkins in 1976 (82). AR models measure the dependencies between observations at previous time steps to forecast the value at the next time step. MA models use a linear combination of past time steps' white noise terms, or residual error terms. The ARIMA combines these and add the Integrated (I) element. This step involves differencing the time-series to convert a non-stationary time-series to a stationary time-series (98; 113). ARIMA is a standard

tool for time series data. However, because it assumes linear relationships between variables, it often does not perform well structurally complex data (114). SARIMAX is an extension of ARIMA where seasonality and exogenous variables are accounted for in the prediction.

Long Short-Term Memory LSTM is a special type of Recurrent Neural Network (RNN). They were created to remedy the problem of exploding or vanishing gradients that can happen in RNNs, which causes them to not remember Long term dependencies. While RNNs are able to make more accurate predictions from recent information, they do not retain information presented earlier meaning their performance will decrease the larger the gap becomes. LSTMs are explicitly designed to remember long term dependencies, making them ideal for processing and forecasting for time-series data (115).

Random Forest Random Forest (RF) is a machine learning algorithm, introduced by Breiman in 2001 (48), which can be used for classification or regression. It operates by training each tree on a random subset of the data and using a random selection of features at each split, introducing diversity among trees and reducing overfitting. RF regression has become very popular as it can be applied to a wide variety of problems by capturing complex interactions in the data. Because of this, it makes the model well suited for time series forecasting (116).

3.5 Results and Discussion

In this section, we go over the results of the data preprocessing step and the forecasts of each of the models.

3.5.1 Data Preprocessing

To evaluate the imputation method, an experiment was conducted. The goal of the experiment was to see how well the imputed values compared to true values in the dataset and see if the daily fluctuations of the dataset were captured. By comparing to available data, we can get a sense of how good the data imputed will be.

To conduct the experiment, first, all missing values in the datasets were removed. Then a random 100-hour chunk of data was selected. Values for this chunk were calculated as the average for the corresponding date and hour of the remaining available data. The imputed values were then compared to the true values in plots. While the imputation was not perfect, it effectively captured daily fluctuations. Images of these imputation are included in Figure 3.2 and MAE values for each of these figures are shown in Table 3.1

The number of datapoints that required imputation in each type of data at each fossil plant is represented in Table 3.2. By knowing how many missing points will require imputation, we can more accurately assess how this imputation may impact our datasets used in our models.

In the table, Gallatin exhibits the fewest missing water temperature data points. While there are some missing air temperature records over the years at this location, the gaps are relatively minor, generally not exceeding the data equivalent of a day. However, the challenge with Gallatin lies in its flow data, where we observe a consistent pattern of missing data points, corresponding to the same day and hour each year.

Cumberland, similarly, shows minimal missing data points. The most significant data gap occurred in 2016, with 108 missing temperature recordings. Since Cumberland utilizes the same air temperature data as Gallatin, the number of missing data points matches that of the latter location.

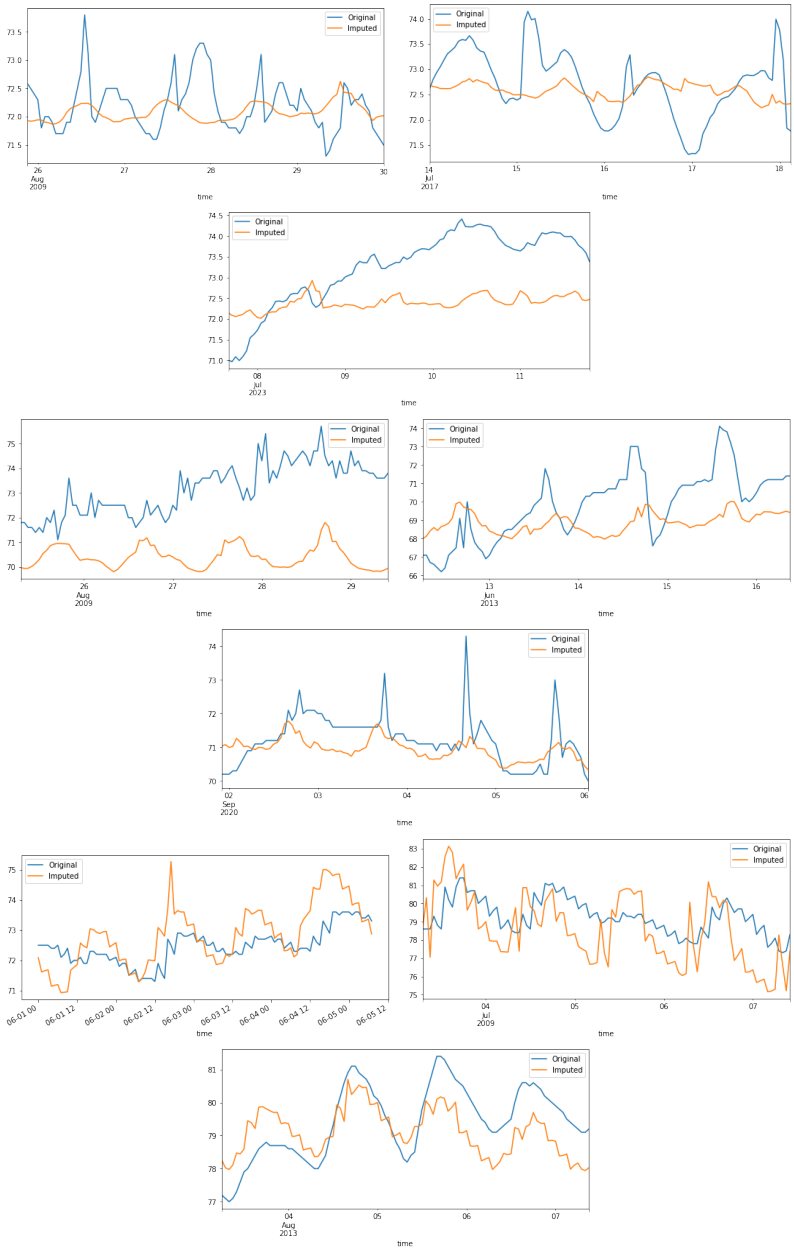

Figure 3.2: Comparison of true and imputed values in water temperature data at Gallatin (top three plots), Cumberland (middle three plots), and Kingston (bottom three plots). MAE values for each of these figures are shown in Table 3.1

Table 3.1: MAE values for the imputation experiment at each of the fossil plant locations.

Gallatin		Cumberland		Kingston	
August 2009	0.379	August 2009	2.750	June 2008	0.680
July 2017	0.532	June 2013	1.737	July 2009	1.536
July 2023	1.039	September 2020	0.492	August 2013	0.834

Table 3.2: Number of missing datapoints for water temperature, air temperature, and flow at each of the fossil plant locations

	Gallatin			Cumberland			Kingston		
Year	**Water**	**Air**	**Flow**	**Water**	**Air**	**Flow**	**Water**	**Air**	**Flow**
2008	11	0	122	9	0	0	16	0	0
2009	0	6	122	0	6	0	0	4	0
2010	0	5	122	0	5	0	63	0	0
2011	0	10	122	5	10	0	122	0	0
2012	0	24	122	0	24	0	0	20	0
2013	0	0	122	9	0	0	17	0	0
2014	0	2	122	0	2	0	1031	2	0
2015	0	2	122	0	2	0	558	10	0
2016	0	5	122	108	5	0	0	2	0
2017	0	2	122	0	2	0			
2018	0	0	122	0	0	0			
2019	0	0	122	0	0	0			
2020	0	6	122	0	6	0			
2021	0	3	122	0	3	0			
2022	0	16	122	0	16	0			
2023	0	0	122	0	0	0			
2024	0	0	122	0	0	0			

Kingston represents the most problematic site in terms of data reliability. Vandalism at this location has severely compromised its temperature recording capabilities, resulting in the most substantial data inconsistencies.

3.5.2 Forecasting

This section evaluates the performance of three different predictive models — SARIMAX, LSTM, and RF—to forecast one week of water temperatures at three TVA fossil plant locations: Gallatin, Cumberland, and Kingston. The effectiveness of each model is assessed based on their ability to predict daily water temperatures during the summer months, and their performance was quantitatively measured using the Mean Absolute Error (MAE).

SARIMAX As can be seen in Figure 3.3, the SARIMAX model consistently begins with predictions that closely match the actual temperature values, indicating its ability to initialize accurately and follow the initial trend effectively. A key weakness observed across all three locations is the monotone nature of the SARIMAX forecasts. While the model captures some daily fluctuations, it falls short of representing the full dynamic variations observed in the actual data.

LSTM Using data from Gallatin, the overall trend of the data is captured in the LSTM's predictions, however the daily variations are not modeled. This can be seen in Figure 3.4. While the true temperatures exhibit a fluctuating pattern with multiple peaks and valleys, the LSTM model predicts a relatively flat, consistently increasing trend, which contrasts sharply with the observed values at Cumberland. At Kingston, the LSTM starts fairly close to the true temperatures, but diverges significantly as the time goes on.

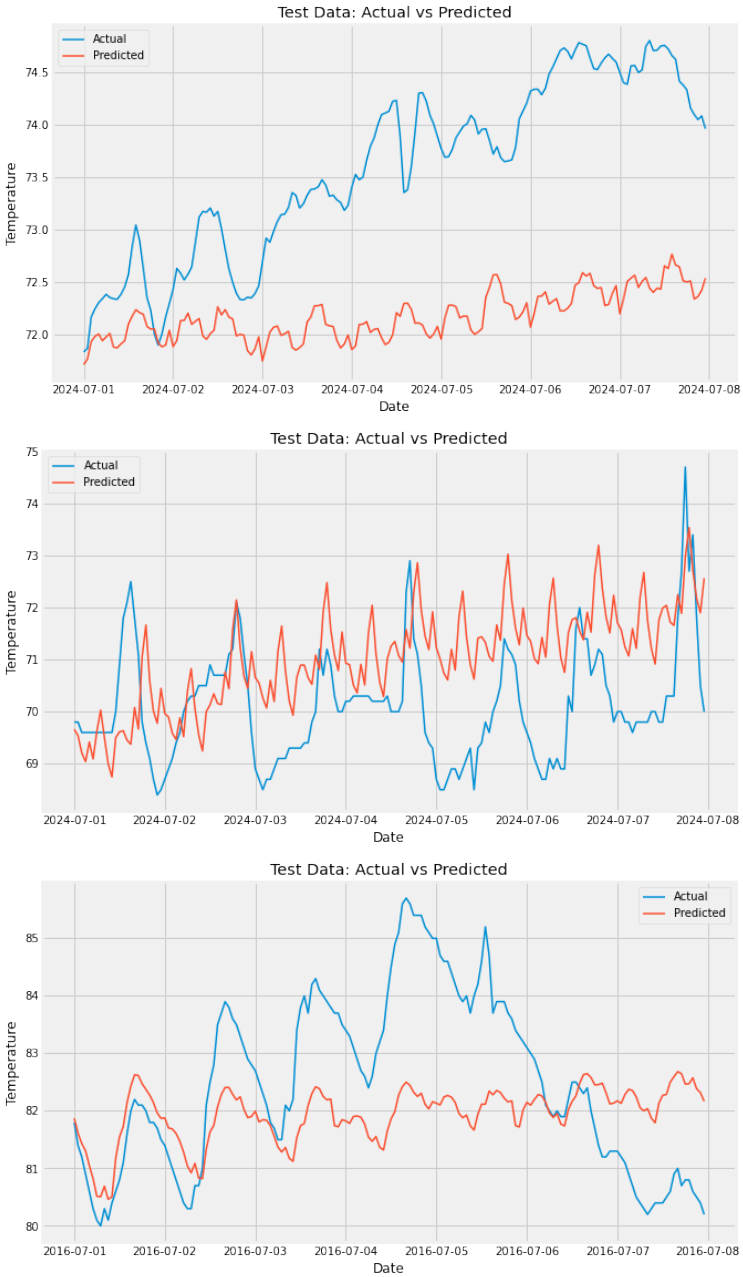

Figure 3.3: Comparison of SARIMAX forecasts at Gallatin (top), Cumberland (middle), and Kingston (bottom). MAE values for the models are given in Table 3.3

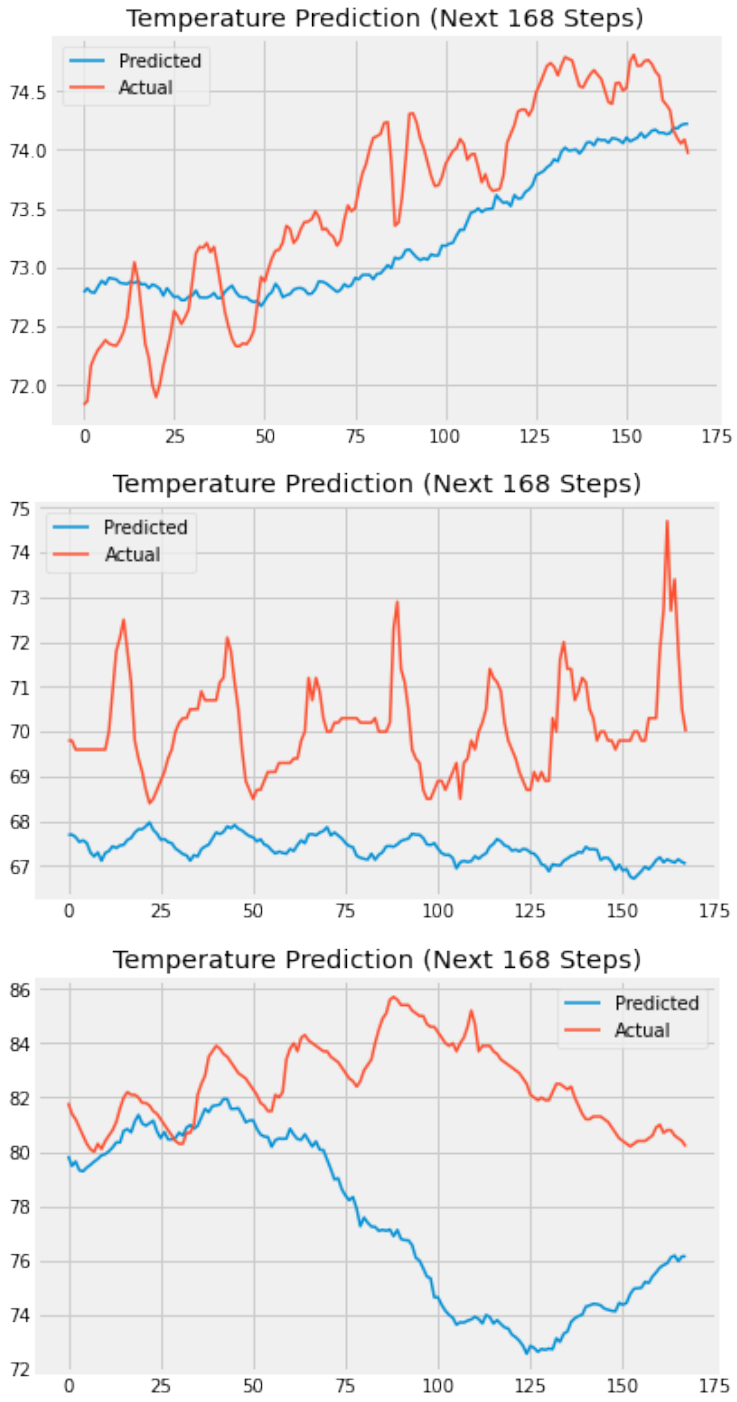

Figure 3.4: Comparison of LSTM forecasts at Gallatin (top), Cumberland (middle), and Kingston (bottom). MAE values for the models are given in Table 3.3

RF The RF model at Gallatin does a very poor job of predicting the temperatures and instead predicts nearly the same value for the entire forecasting period. However, despite this, this model has the lowest error at this location. This gives us an indication that additional metrics of evaluation may need to be investigated to get a better idea of how a model forecasts at a location. The predictions for RF at Cumberland are much lower than the true value and does not capture the overall trend. Finally, at Kingston, this seems to be the best forecasting out of the models, where the trend and fluctuations are captured. The predictions from the RF model are shown in Figure 3.5.

Overall, we see that SARIMAX has a smallest amount of error between all the forecasting models at Cumberland. This is interesting as the model did not appear to accurately capture fluctuations in the data. The RF has the smallest error at Gallatin and Kingston. Because of how drastically different the results were for RF at these locations, it is surprising that the error is low for both of them. It may have been expected to see higher error for the Kingston fossil plant as this location had the smallest amount of data requiring quite a bit of preprocessing. However, this is not evident in this evaluation metric. The comparison of each of the MAE values for the models are shown in Table 3.3.

Looking at the results of these models, it is clear to see that are not doing the best job at modeling the behavior and fluctuations of the true water temperatures. These results are more granular and could be useful to TVA, however they are not accurate enough to be entirely reliable. It is surprising that while the parameters of these models have been tuned, their outputs are less than ideal and may not be as meaningful as we may hope for the TVA fossil plants. Because of these results, it brings up an interesting topic that needs to be discussed. We must look deeper into the forecasting models and discuss how and when models fail, how to examine the models to pinpoint the issue, and how to improve results if possible. By engaging in this conversation, we can get a better idea of why the results may have come out the way that they did and give us insight in how to improve them in the future.

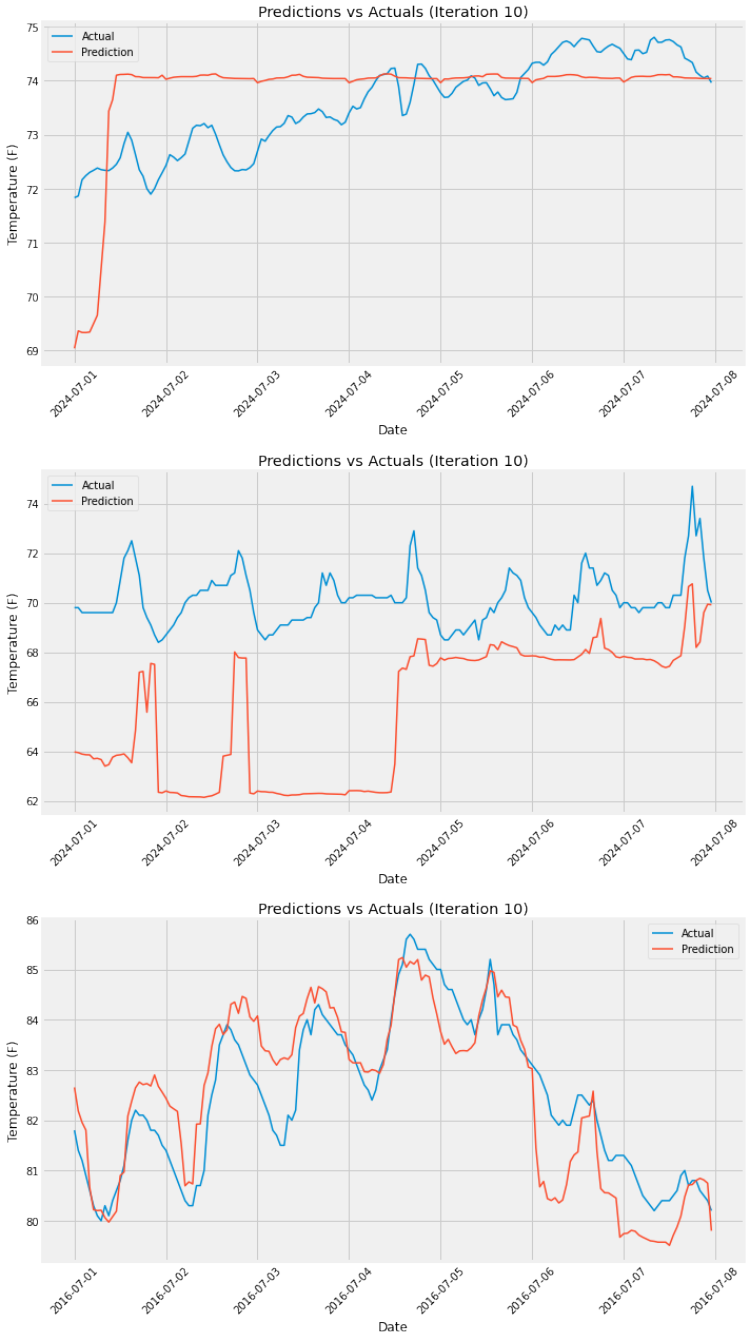

Figure 3.5: Comparison of RF forecasts at Gallatin (top), Cumberland (middle), and Kingston (bottom). MAE values for the models are given in Table 3.3

Table 3.3: MAE values for SARIMAX, LSTM, and RF model at Gallatin, Cumberland, and Kingston

	Gallatin	Cumberland	Kingston
SARIMAX	1.407	**1.298**	1.267
LSTM	1.544	2.095	3.354
RF	**0.794**	4.489	**0.706**

When it comes to utilizing machine learning models, there is no one-size-fits-all model that can accomplish every task. Selecting the appropriate algorithm is crucial and begins with a thorough understanding of the problem at hand and how various algorithms solve it. This understanding aids in predicting which algorithms will be most effective (117). An essential benefit of a well-chosen algorithm is its ability to discern complex and subtle patterns in the data. However, the efficacy of a good model is contingent on the quality of the data available; reliable and comprehensive data is necessary for a model to generate accurate predictions (118).

Once algorithms are selected, it's imperative to assess the structure and/or hyperparameters of the model. Sometimes, a simple model with default settings may perform better than a more complex one that has been extensively tuned (119). In other instances, it might be beneficial to conduct a grid search to determine the most effective structure or parameters. The best approach depends on the specific data and the task required by the model. This iterative process of evaluation and adjustment is crucial for optimizing model performance.

Machine learning models face numerous challenges that can compromise their effectiveness and accuracy. While these models generally improve with more data, excessively large datasets can overwhelm an algorithm, significantly degrading performance or even rendering it inoperative (120). Models are susceptible to overfitting, where they perform well on training data but poorly on unseen data, limiting their practical applicability. Other issues models may encounter include underfitting, where a model fails to learn effectively from the training data and cannot generalize well; and scalability concerns, where increasing data size and complexity disproportionately escalate computational demands and processing time (121).

Models based on deep learning, despite their advanced capabilities, often suffer from being "black boxes," where their decision-making processes are not transparent, making them difficult to trust and verify, particularly in fields requiring rigorous validation. This lack of interpretability can be a critical barrier, especially when models need to be explained or justified in real-world applications (122).

In the realm of forecasting, the complexity of real-world systems often presents substantial challenges for both ML and traditional statistical models. Real-world data, characterized by its chaotic behaviors, introduces unpredictable fluctuations that complicate the modeling process significantly. Such complexities necessitate sophisticated approaches to forecasting, where the context is a critical determinant of a model's success.

Machine learning models, with their advanced algorithms, are adept at navigating complex, nonlinear relationships found in various domains such as inventory management (123; 124) and financial markets (125; 126). These models leverage large datasets to uncover underlying patterns that are not immediately apparent, providing insights that can significantly enhance forecasting accuracy in these fields.

However, the sophistication of ML models can sometimes be a double-edged sword. In contexts where the data do not support complex modeling, or where the interpretive demands of the model outputs are stringent, ML models can underperform compared to simpler, traditional statistical methods. This phenomenon was highlighted in a study by (127), which compared the efficacy of various ML models against traditional statistical methods in time series forecasting. The study found that all the ML models assessed performed worse than the statistical approaches, emphasizing that ML models were also more computationally expensive.

A subsequent study expanded on these findings by including deep learning (DL) models in the comparison (128). It was observed that while DL models did perform better than both the traditional ML models and statistical methods, the marginal improvement in accuracy came at the cost of significantly increased computational resources and time. This indicates that while DL can offer enhanced forecasting capabilities, the substantial investment in computational power and time may not always be justified, especially when the gains are minimal. Another aspect of forecasting that can complicate the model is the incorporation of covariates in time series forecasting. In (129), it was found that in most cases, including more

variables in an LSTM model that these covariates become ineffective or a hindrance to performance as forecast horizons extend.

Selecting the optimal model for forecasting is crucial for accuracy. As, already examined, the context and specific characteristics of the problem often dictate which models might perform better. Once a model is chosen, it's essential to rigorously evaluate its effectiveness for the intended task. Various methods can be employed to assess a model's performance, ultimately allowing us to choose the model that yields the most precise forecasts. Some ways to validate a model can be done quantitatively with an evaluation metric, using a benchmark model for comparison, and validating by utilizing a validation dataset. An article by Hewamalage, Ackermann, and Bergmeir (130) stresses the importance of evaluation metrics and cross-validation in order to get a better understanding of the performance of a model. Depending upon the characteristics of the time series and the goal of the forecasting task, an evaluation metric and cross-validation method should be carefully selected. Other sources recommend considering multiple accuracy metrics, illuminating the fact that multiple steps must be taken in order to ensure that a model is the best fit (131; 132; 133)

After identifying a forecasting problem, selecting appropriate hyperparameters, and effectively evaluating a model, it may still occur that the model does not perform as expected. In such cases, additional factors need to be examined to resolve the issues. One critical area to revisit is the quality of the dataset. The integrity and relevance of data play a crucial role in the performance of machine learning models. If the results are unsatisfactory, the initial steps should include reevaluating the dataset. This process involves examining the data for anomalies, removing noisy data, imputing missing values, and conducting feature selection and extraction to eliminate irrelevant or redundant data, as discussed by Hall (1999) (134). Further methods for ensuring data is adequately prepared can be explored in the comprehensive reviews by Garcia et al. and Khalid et al. (135; 136).

Another effective strategy to enhance ML model performance is through the application of ensemble methods. These methods aim to combine multiple models

to form a more robust and accurate prediction system. The strength of ensemble methods lies in their ability to harness the predictive power of various models, thereby improving the overall forecast accuracy. The successful implementation of these approaches in forecasting has been documented in several studies, including (137; 138; 139).

Depending on the specific machine learning model being used, several additional steps might be necessary to optimize performance. Normalization, for instance, can be crucial for ensuring that data within different scales are treated equally, preventing biases towards variables with larger scales. Regularization techniques, such as L1 and L2, are also instrumental in preventing overfitting by adding a penalty to the loss function associated with large coefficients, thereby promoting simpler models that perform better on new, unseen data. Moreover, exploring more advanced model architectures can provide significant benefits, especially in complex scenarios where traditional models falter. (140; 141; 142).

Sometimes, even after extensive data preprocessing and model adjustments, the accuracy of the forecasts may not meet expectations. This often suggests that deeper, systemic factors might be missing from the data, which no amount of minor adjustments can compensate for. In such cases, a thorough investigation is required to identify and address these gaps.

Integrating domain knowledge into the forecasting process can be particularly beneficial. By incorporating expert insights into the model, it becomes possible to identify key drivers and interactions within the data that are not immediately apparent. This domain knowledge not only aids in enhancing the model's predictive accuracy but also helps in hypothesizing about potentially missing data that could improve model performance.

For instance, in forecasting scenarios like water temperature forecasting, understanding the surrounding geophysical and environmental characteristics can potentially improve forecasts. Experts from the locations can provide insights into

how these factors interact, offering a richer, more comprehensive dataset for the models to learn from.

Ultimately, if a model continues to perform poorly, it may be necessary to revisit the fundamental assumptions of the model, the quality of the data being used, and the appropriateness of the model type for the task. Incorporating a multidisciplinary approach that leverages both data science and subject-matter expertise can significantly enhance the robustness and accuracy of the predictions.

The predictive accuracy of our water temperature forecasting model appears to be compromised by several interrelated factors, even after rigorous data cleaning efforts. To begin with, the quality of the underlying dataset might not be optimal. Despite meticulous processes to remove statistical outliers and impute missing entries, these modifications could inadvertently affect the fidelity of the final results. The lack of comprehensive variables within the dataset is another significant limitation. Critical factors such as water depth, sunlight or shade percentage, along with other geophysical and environmental characteristics (143; 144), are clearly absent from our analysis. These variables likely play crucial roles in determining water temperatures but are not represented in our available data. Including such factors could dramatically improve the model's predictive capabilities. Furthermore, without incorporating this extended domain knowledge, it remains challenging to gauge whether the forecasts produced by our model or TVA's can be reliably utilized to inform operational decisions at fossil plants. The absence of these critical data points hinders our ability to develop a model that fully captures the complexities of the ecosystem, thereby limiting the potential utility of our forecasts in real-world applications.

3.6 Conclusions and Future Work

This study embarked on developing a forecasting model aimed at reliably predicting hourly water temperatures across three TVA fossil power plants. Accurate predictions

are vital for optimizing plant efficiency, ensuring compliance with regulatory standards, and protecting local aquatic ecosystems. Such precision in forecasting enables plant operators to make well-informed operational decisions, thereby preventing disruptions and minimizing environmental damage.

The research was concentrated on data from June through September, a critical period for maintaining water temperature limits essential for operational compliance and environmental safety.

Despite the advancements over existing forecasting tools at TVA—which traditionally provide only daily temperature predictions—the results of our machine learning models, although more detailed, did not achieve the level of accuracy required for complete reliability in operational planning. While our models offer a more granular view than TVA's current methods, there remains a significant gap in their ability to consistently predict temperatures with the precision needed for practical applications.

Future work could aim to enhance the accuracy and utility of these forecasts. One approach is to experiment with ensemble methods that combine multiple modeling techniques to improve prediction accuracy. Additionally, incorporating more comprehensive domain knowledge into the models could address some of the current shortcomings. Comparing the results of our models with simpler benchmark models, such as ARIMA and SARIMA, will also help to quantify the improvements our methods offer. An ablation study could be conducted next to determine whether the removal of certain exogenous variables might lead to more accurate predictions. Moreover, adopting a tailored evaluation metric that penalizes underestimations more severely could provide more meaningful insights to TVA, especially if there is a greater concern about forecasts underestimating water temperatures. These steps may help refine our models to meet the operational needs of TVA and ensure environmental compliance.

Bibliography

[1] GISGeography, "Tennessee lakes and rivers map," May 2022. x, 42

[2] A. K. Kacyira, "Addressing the sustainable urbanization challenge," 2012. 1

[3] R. Liang, M. A. Thyer, H. R. Maier, G. C. Dandy, and M. Di Matteo, "Optimising the design and real-time operation of systems of distributed stormwater storages to reduce urban flooding at the catchment scale," *Journal of Hydrology*, vol. 602, p. 126787, 2021. 1

[4] L. Rosenberger, J. Leandro, S. Pauleit, and S. Erlwein, "Sustainable stormwater management under the impact of climate change and urban densification," *Journal of Hydrology*, vol. 596, p. 126137, 2021. 1

[5] S.-S. Baek, D.-H. Choi, J.-W. Jung, H.-J. Lee, H. Lee, K.-S. Yoon, and K. H. Cho, "Optimizing low impact development (lid) for stormwater runoff treatment in urban area, korea: Experimental and modeling approach," *Water Research*, vol. 86, pp. 122–131, 2015. 1

[6] L. A. Rossman, "Storm water management model user's manual version 5.1 - manual," Nov 2015. 2

[7] M. Barah, A. Khojandi, X. Li, J. Hathaway, and O. Omitaomu, "Optimizing green infrastructure placement under precipitation uncertainty," *Omega*, vol. 100, p. 102196, 2021. 2, 3

[8] T. Dell, M. Razzaghmanesh, S. Sharvelle, and M. Arabi, "Development and application of a swmm-based simulation model for municipal scale hydrologic assessments," *Water*, vol. 13, no. 12, 2021. 2

[9] G. Liu, L. Chen, Z. Shen, Y. Xiao, and G. Wei, "A fast and robust simulation-optimization methodology for stormwater quality management," *Journal of Hydrology*, vol. 576, pp. 520–527, 2019. 2

[10] T. Xu and F. Liang, "Machine learning for hydrologic sciences: An introductory overview," *WIREs Water*, vol. 8, no. 5, p. e1533, 2021. 2

[11] H. Lange and S. Sippel, "Machine learning applications in hydrology," in *Forest-water interactions*, pp. 233–257, Springer, 2020. 2, 3

[12] C. B. Guzman, R. Wang, O. Muellerklein, M. Smith, and C. G. Eger, "Comparing stormwater quality and watershed typologies across the united states: A machine learning approach," *WATER RESEARCH*, vol. 216, JUN 1 2022. 2

[13] M. Lowe, R. Qin, and X. Mao, "A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring," *WATER*, vol. 14, MAY 2022. 2

[14] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (lstm) networks," *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 6005–6022, 2018. 2

[15] S. H. Kwon and J. H. Kim, "Machine learning and urban drainage systems: State-of-the-art review," *WATER*, vol. 13, DEC 2021. 2

[16] S. Jang, M. Cho, J. Yoon, Y. Yoon, S. Kim, G. Kim, L. Kim, and H. Aksoy, "Using swmm as a tool for hydrologic impact assessment," *Desalination*, vol. 212, no. 1-3, pp. 344–356, 2007. 3

[17] C. Tuomela, N. Sillanpää, and H. Koivusalo, "Assessment of stormwater pollutant loads and source area contributions with storm water management model (swmm)," *Journal of Environmental Management*, vol. 233, pp. 719–727, 2019. 3

[18] P. M. Avellaneda, A. J. Jefferson, J. M. Grieser, and S. Bush, "Simulation of the cumulative hydrological response to green infrastructure," *Water Resources Research*, vol. 53, no. 4, pp. 3087–3101, 2017. 3

[19] T. Epps and J. Hathaway, "Using spatially-identified effective impervious area to target green infrastructure retrofits: A modeling study in knoxville, tn," *Journal of hydrology*, vol. 575, pp. 442–453, 2019. 3

[20] T. Xu and F. Liang, "Machine learning for hydrologic sciences: An introductory overview," *WIREs Water*, vol. 8, no. 5, p. e1533, 2021. 5

[21] C. Shen, X. Chen, and E. Laloy, "Editorial: Broadening the use of machine learning in hydrology," *FRONTIERS IN WATER*, vol. 3, MAY 11 2021. 5

[22] S. Sahoo, T. A. Russo, J. Elliott, and I. Foster, "Machine learning algorithms for modeling groundwater level changes in agricultural regions of the u.s.," *Water Resources Research*, vol. 53, no. 5, pp. 3878–3895, 2017. 5

[23] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Advances in Water Resources*, vol. 33, no. 1, pp. 69–80, 2010. 5

[24] H. Fang, B. Jamali, A. Deletic, and K. Zhang, "Machine learning approaches for predicting the performance of stormwater biofilters in heavy metal removal and risk mitigation," *Water Research*, vol. 200, p. 117273, 2021. 5

[25] M. Moeini, A. Shojaeizadeh, and M. Geza, "Supervised machine learning for estimation of total suspended solids in urban watersheds," *Water*, vol. 13, no. 2, 2021. 5

[26] A. Mosavi, P. Ozturk, and K.-w. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, 2018. 5

[27] S. Zhu, H. Lu, M. Ptak, J. Dai, and Q. Ji, "Lake water level fluctuation forecasting using machine learning models: a systematic review," *Environmental Science and Pollution Research*, vol. 27, 12 2020. 5

[28] Z. Jiang, S. Yang, Z. Liu, Y. Xu, Y. Xiong, S. Qi, Q. Pang, J. Xu, F. Liu, and T. Xu, "Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in yangtze river basin," *Environmental Modelling Software*, vol. 155, p. 105436, 2022. 5

[29] S. Xie, W. Wu, S. Mooser, Q. Wang, R. Nathan, and Y. Huang, "Artificial neural network based hybrid modeling approach for flood inundation modeling," *Journal of Hydrology*, vol. 592, p. 125605, 2021. 5

[30] H. Tamiru and M. O. Dinka, "Application of ann and hec-ras model for flood inundation mapping in lower baro akobo river basin, ethiopia," *Journal of Hydrology: Regional Studies*, vol. 36, p. 100855, 2021. 5

[31] L.-C. Chang, H.-Y. Shen, Y.-F. Wang, J.-Y. Huang, and Y.-T. Lin, "Clustering-based hybrid inundation model for forecasting flood inundation depths," *Journal of Hydrology*, vol. 385, no. 1, pp. 257–268, 2010. 5

[32] J. He, C. Valeo, A. Chu, and N. F. Neumann, "Prediction of event-based stormwater runoff quantity and quality by anns developed using pmi-based input selection," *Journal of Hydrology*, vol. 400, no. 1, pp. 10–23, 2011. 5

[33] G. Leng and J. W. Hall, "Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models," *Environmental Research Letters*, vol. 15, p. 044027, apr 2020. 5

[34] V. Mišić, K. Rajaram, and E. Gabel, "A simulation-based evaluation of machine learning models for clinical decision support: application and analysis using hospital readmission," *npj Digital Medicine*, vol. 4, p. 98, 06 2021. 5

[35] A. Warey, S. Kaushik, B. Khalighi, M. Cruse, and G. Venkatesan, "Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results," *International Journal of Heat and Mass Transfer*, vol. 148, p. 119083, 2020. 5

[36] J. Chen, A. C. Pillai, L. Johanning, and I. Ashton, "Using machine learning to derive spatial wave data: A case study for a marine energy site," *Environmental Modelling Software*, vol. 142, p. 105066, 2021. 5

[37] L. Sankarrao, D. K. Ghose, and M. Rathinsamy, "Predicting land-use change: Intercomparison of different hybrid machine learning models," *Environmental Modelling Software*, vol. 145, p. 105207, 2021. 5

[38] C. Pylianidis, V. Snow, H. Overweg, S. Osinga, J. Kean, and I. N. Athanasiadis, "Simulation-assisted machine learning for operational digital twins," *Environmental Modelling Software*, vol. 148, p. 105274, 2022. 5

[39] A. Essenfelder and C. Giupponi, "A coupled hydrologic-machine learning modelling framework to support hydrologic modelling in river basins under interbasin water transfer regimes," *Environmental Modelling Software*, vol. 131, p. 104779, 2020. 5

[40] E. Rozos, P. Dimitriadis, and V. Bellos, "Machine learning in assessing the performance of hydrological models," *Hydrology*, vol. 9, no. 1, 2022. 5

[41] F. Granata, R. Gargano, and G. De Marinis, "Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the epa's storm water management model," *Water*, vol. 8, no. 3, p. 69, 2016. 6

[42] H. I. Kim and K. Y. Han, "Urban flood prediction using deep neural network with data augmentation," *Water*, vol. 12, no. 3, p. 899, 2020. 6

[43] Q. Wang and S. Wang, "Machine learning-based water level prediction in lake erie," *Water*, vol. 12, no. 10, p. 2654, 2020. 6

[44] S. Ray, "A quick review of machine learning algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 35–39, 2019. 10

[45] A. J. Dobson, *An introduction to generalized linear models / Annette J. Dobson.* Chapman Hall/CRC Boca Raton, 2nd ed. ed., 2002. 10

[46] I. Jenhani, N. B. Amor, and Z. Elouedi, "Decision trees as possibilistic classifiers," *International Journal of Approximate Reasoning*, vol. 48, no. 3, pp. 784–807, 2008. Special Section on Choquet Integration in honor of Gustave Choquet (1915–2006) and Special Section on Nonmonotonic and Uncertain Reasoning. 11

[47] L. Breiman, *Classification and regression trees.* Abingdon: Routledge, 2017 - 1984. 11

[48] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001. 11, 63

[49] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016. http://www.deeplearningbook.org. 12

[50] C. Knoxville, "Water quality model for first and whites creeks," 2012. 13

[51] J. Nash and J. Sutcliffe, "River flow forecasting through conceptual models part i — a discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970. 13, 25

[52] D. Moriasi, M. Gitau, N. Pai, and P. Daggupati, "Hydrologic and water quality models: Performance measures and evaluation criteria," *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)*, vol. 58, pp. 1763–1785, 12 2015. 13

[53] T. Verdonck, B. Baesens, M. Óskarsdóttir, and S. vanden Broucke, "Special issue on feature engineering editorial - machine learning," Aug 2021. 14

[54] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005. 16

[55] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979. 18

[56] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," *IOP Conference Series: Materials Science and Engineering*, vol. 336, p. 012017, apr 2018. 18

[57] P. S. Bradley, K. P. Bennett, and A. Demiriz, "Constrained k-means clustering," *Microsoft Research, Redmond*, vol. 20, no. 0, p. 0, 2000. 18

[58] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017. 19

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. 19

[60] F. Chollet *et al.*, "Keras." https://keras.io, 2015. 19

[61] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," *arXiv preprint arXiv:1809.03006*, 2018. 22

[62] R. H. McCuen, Z. Knight, and A. G. Cutter, "Evaluation of the nash-sutcliffe efficiency index," *Journal of Hydrologic Engineering*, vol. 11, no. 6, pp. 597–602, 2006. 22

[63] N. S. Raghuwanshi, R. Singh, and L. Reddy, "Runoff and sediment yield modeling using artificial neural networks: Upper siwane river, india," *Journal of Hydrologic Engineering*, vol. 11, no. 1, pp. 71–79, 2006. 26

[64] S. Chakravarty, H. Demirhan, and F. Baser, "Fuzzy regression functions with a noise cluster and the impact of outliers on mainstream machine learning methods in the regression setting," *Applied Soft Computing*, vol. 96, p. 106535, 2020. 28

[65] F. Petrakopoulou, A. Robinson, and M. Olmeda-Delgado, "Impact of climate change on fossil fuel power-plant efficiency and water use," *Journal of Cleaner Production*, vol. 273, p. 122816, 2020. 39, 57

[66] N. H. Teguh, L. Yuliati, and D. B. Darmadi, "Effect of seawater temperature rising to the performance of northern gorontalo small scale power plant," *Case Studies in Thermal Engineering*, vol. 32, p. 101858, 2022. 39

[67] EPA Website, "Learn about green infrastructure," 2021. 39

[68] J. R. Jones, J. S. Schwartz, K. N. Ellis, J. M. Hathaway, and C. M. Jawdy, "Temporal variability of precipitation in the upper tennessee valley," *Journal of Hydrology: Regional Studies*, vol. 3, pp. 125–138, 2015. 40

[69] G. Lischeid, R. Dannowski, K. Kaiser, G. Nützmann, J. Steidl, and P. Stüve, "Inconsistent hydrological trends do not necessarily imply spatially heterogeneous drivers," *Journal of Hydrology*, vol. 596, p. 126096, 2021. 40

[70] V. M. Brown, B. D. Keim, and A. W. Black, "Trend analysis of multiple extreme hourly precipitation time series in the southeastern united states," *Journal of Applied Meteorology and Climatology*, vol. 59, no. 3, pp. 427 – 442, 2020. 40

[71] M. Fooladi, M. H. Golmohammadi, H. R. Safavi, R. Mirghafari, and H. Akbari, "Trend analysis of hydrological and water quality variables to detect anthropogenic effects and climate variability on a river basin scale: A case study of iran," *Journal of Hydro-environment Research*, vol. 34, pp. 11–23, 2021. 40

[72] R. Noori, M. Sabahi, A. Karbassi, A. Baghvand, and H. Taati Zadeh, "Multivariate statistical analysis of surface water quality based on correlations and variations in the data set," *Desalination*, vol. 260, no. 1, pp. 129–136, 2010. 40

[73] H. G. Orr, G. L. Simpson, S. des Clers, G. Watts, M. Hughes, J. Hannaford, M. J. Dunbar, C. L. R. Laizé, R. L. Wilby, R. W. Battarbee, and R. Evans, "Detecting changing river temperatures in england and wales," *Hydrological Processes*, vol. 29, no. 5, pp. 752–766, 2015. 40

[74] Z. W. Kundzewicz and A. J. Robson, "Change detection in hydrological records—a review of the methodology / revue méthodologique de la détection de changements dans les chroniques hydrologiques," *Hydrological Sciences Journal*, vol. 49, no. 1, pp. 7–19, 2004. 40

[75] O. Kisi and M. Ay, "Comparison of mann–kendall and innovative trend method for water quality parameters of the kizilirmak river, turkey," *Journal of Hydrology*, vol. 513, pp. 362–375, 2014. 40

[76] A. Gadedjisso-Tossou, K. I. Adjegan, and A. K. M. Kablan, "Rainfall and temperature trend analysis by mann–kendall test and significance for rainfed cereal yields in northern togo," *Sci*, vol. 3, no. 1, 2021. 40

[77] B. W. WEBB and F. NOBILIS, "Long-term perspective on the nature of the air–water temperature relationship: A case study," *Hydrological Processes*, vol. 11, no. 2, pp. 137–147, 1997. 40

[78] B. W. Webb, P. D. Clack, and D. E. Walling, "Water–air temperature relationships in a devon river system and the role of flow," *Hydrological Processes*, vol. 17, no. 15, pp. 3069–3084, 2003. 40

[79] J. C. White, K. Khamis, S. Dugdale, F. L. Jackson, I. A. Malcolm, S. Krause, and D. M. Hannah, "Drought impacts on river water temperature: A process-based understanding from temperate climates," *Hydrological Processes*, vol. 37, no. 10, p. e14958, 2023. 40

[80] "Our power system." 40

[81] C. S. Lamprecht, "Meteostat python." 41

[82] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015. 46, 62

[83] H. B. Mann, "Nonparametric tests against trend," *Econometrica: Journal of the econometric society*, pp. 245–259, 1945. 46

[84] M. G. Kendall, "Rank correlation methods.," 1948. 46

[85] P. Sedgwick, "Pearson's correlation coefficient," *Bmj*, vol. 345, 2012. 46

[86] P. Behrens, M. T. van Vliet, T. Nanninga, B. Walsh, and J. F. Rodrigues, "Climate change and the vulnerability of electricity generation to water stress in the european union," *Nature Energy*, vol. 2, Jul 2017. 57

[87] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," *Journal of Economic Surveys*, 2021. 58

[88] A. R. S. Parmezan, V. M. Souza, and G. E. Batista, "Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model," *Information sciences*, vol. 484, pp. 302–337, 2019. 58

[89] R. Qiu, Y. Wang, B. Rhoads, D. Wang, W. Qiu, Y. Tao, and J. Wu, "River water temperature forecasting using a deep learning method," *Journal of Hydrology*, vol. 595, p. 126016, 2021. 58

[90] S. Zhu, E. K. Nyarko, M. Hadzima-Nyarko, S. Heddam, and S. Wu, "Assessing the performance of a suite of machine learning models for daily river water temperature prediction," *PeerJ*, vol. 7, p. e7065, 2019. 58

[91] X. Yu, S. Shi, L. Xu, Y. Liu, Q. Miao, and M. Sun, "A novel method for sea surface temperature prediction based on deep learning," *Mathematical Problems in Engineering*, vol. 2020, 2020. 58

[92] M. Feigl, K. Lebiedzinski, M. Herrnegger, and K. Schulz, "Machine-learning methods for stream water temperature prediction," *Hydrology and Earth System Sciences*, vol. 25, no. 5, pp. 2951–2977, 2021. 58

[93] D. Jiang, Y. Xu, Y. Lu, J. Gao, and K. Wang, "Forecasting water temperature in cascade reservoir operation-influenced river with machine learning models," *Water*, vol. 14, no. 14, p. 2146, 2022. 58

[94] F. Farhangi, A. Sadeghi-Niaraki, J. Safari Bazargani, S. V. Razavi-Termeh, D. Hussain, and S.-M. Choi, "Time-series hourly sea surface temperature prediction using deep neural network models," *Journal of Marine Science and Engineering*, vol. 11, no. 6, 2023. 58

[95] S. Zhu, B. Hrnjica, M. Ptak, A. Choiński, and B. Sivakumar, "Forecasting of water level in multiple temperate lakes using machine learning models," *Journal of Hydrology*, vol. 585, 03 2020. 58

[96] R. Graf, S. Zhu, and B. Sivakumar, "Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach," *Journal of Hydrology*, vol. 578, 11 2019. 58

[97] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE geoscience and remote sensing letters*, vol. 14, no. 10, pp. 1745–1749, 2017. 58

[98] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018. 58, 62

[99] J. Cole, K. Maloney, M. Schmid, and J. Jr, "Developing and testing temperature models for regulated systems: A case study on the upper delaware river," *Journal of Hydrology*, vol. 519, Part A, pp. 588–598, 11 2014. 58

[100] D. Faruk, "A hybrid neural network and arima model for water quality time series prediction," *Eng. Appl. of AI*, vol. 23, pp. 586–594, 06 2010. 58

[101] D. Caissie, N. El-Jabi, and M. Satish, "Modeling of maximum daily water temperatures in a small stream using air temperatures," *Journal of Hydrology*, vol. 251, pp. 14–28, 09 2001. 58

[102] A. Rabi, M. Hadzima-Nyarko, and M. Šperac, "Modelling river temperature from air temperature: case of the river drava (croatia)," *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, vol. 60, 09 2015. 58

[103] M. Toffolon and S. Piccolroaz, "A hybrid model for river water temperature as a function of air temperature and discharge," *Environmental Research Letters*, vol. 10, p. 114011, nov 2015. 58

[104] D. Caissie, M. E. Thistle, and L. Benyahya, "River temperature forecasting: case study for little southwest miramichi river (new brunswick, canada)," *Hydrological Sciences Journal*, vol. 62, no. 5, pp. 683–697, 2017. 58

[105] A. St-Hilaire, T. Ouarda, Z. Bargaoui, A. Daigle, and L. Bilodeau, "Daily water temperature forecast model with a k-nearest neighbour approach," *Hydrological Processes*, vol. 26, pp. 1302–1310, 04 2012. 59

[106] G. Sahoo, S. Schladow, and J. Reuter, "Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models," *Journal of Hydrology*, vol. 378, no. 3, pp. 325–342, 2009. 59

[107] G. Sahoo, S. Schladow, and J. Reuter, "Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models," *Journal of hydrology*, vol. 378, no. 3-4, pp. 325–342, 2009. 59

[108] M. Álvarez Cabria, J. Barquín, and F. 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*, vol. 545-546, p. 152–162, 2016. 59

[109] "Gallatin fossil plant." 59

[110] "Cumberland fossil plant." 59

[111] "Kingston fossil plant." 59

[112] A. Zheng and A. Casari, *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc.", 2018. 62

[113] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of arima and lstm in forecasting time series," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 1394–1401, IEEE, 2018. 62

[114] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks," *BMC bioinformatics*, vol. 15, pp. 1–9, 2014. 63

[115] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. 63

[116] H. Tyralis and G. Papacharalampous, "Variable selection in time series forecasting using random forests," *Algorithms*, vol. 10, no. 4, p. 114, 2017. 63

[117] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, no. 1, pp. 381–386, 2020. 74

[118] I. El Naqa and M. J. Murphy, *What is machine learning?* Springer, 2015. 74

[119] H. J. Weerts, A. C. Mueller, and J. Vanschoren, "Importance of tuning hyperparameters of machine learning algorithms," *arXiv preprint arXiv:2007.07588*, 2020. 74

[120] A. L'heureux, K. Grolinger, H. F. Elyamany, and M. A. Capretz, "Machine learning with big data: Challenges and approaches," *Ieee Access*, vol. 5, pp. 7776–7797, 2017. 74

[121] S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, "Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms," *Electronics*, vol. 12, no. 8, p. 1789, 2023. 74

[122] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics*, vol. 10, no. 5, p. 593, 2021. 74

[123] M. Dekker, K. Van Donselaar, and P. Ouwehand, "How to use aggregation and combined forecasting to improve seasonal demand forecasts," *International Journal of Production Economics*, vol. 90, no. 2, pp. 151–167, 2004. 75

[124] G. Zotteri and M. Kalchschmidt, "A model for selecting the appropriate level of aggregation in forecasting processes," *International Journal of Production Economics*, vol. 108, no. 1-2, pp. 74–83, 2007. 75

[125] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European journal of operational research*, vol. 270, no. 2, pp. 654–669, 2018. 75

[126] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, "Stock market index prediction using artificial neural network," *Journal of Economics, Finance and Administrative Science*, vol. 21, no. 41, pp. 89–93, 2016. 75

[127] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PloS one*, vol. 13, no. 3, p. e0194889, 2018. 75

[128] S. Makridakis, E. Spiliotis, V. Assimakopoulos, A.-A. Semenoglou, G. Mulder, and K. Nikolopoulos, "Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward," *Journal of the Operational Research Society*, vol. 74, no. 3, pp. 840–859, 2023. 75

[129] G. Davies, "Evaluating the effectiveness of predicting covariates in lstm networks for time series forecasting," *arXiv preprint arXiv:2404.18553*, 2024. 75

[130] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," *Data Mining and Knowledge Discovery*, vol. 37, no. 2, pp. 788–832, 2023. 76

[131] J. Rožanec, E. Trajkova, K. Kenda, B. Fortuna, and D. Mladenić, "Explaining bad forecasts in global time series models," *Applied Sciences*, vol. 11, no. 19, p. 9243, 2021. 76

[132] N. Mehdiyev, D. Enke, P. Fettke, and P. Loos, "Evaluating forecasting methods by considering different accuracy measures," *Procedia Computer Science*, vol. 95, pp. 264–271, 2016. 76

[133] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: An empirical study on performance estimation methods," *Machine Learning*, vol. 109, no. 11, pp. 1997–2028, 2020. 76

[134] M. A. Hall, *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999. 76

[135] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big data analytics*, vol. 1, pp. 1–22, 2016. 76

[136] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 science and information conference*, pp. 372–378, IEEE, 2014. 76

[137] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez, "Multi-step forecasting for big data time series based on ensemble learning," *Knowledge-Based Systems*, vol. 163, pp. 830–841, 2019. 77

[138] H. L. Cloke and F. Pappenberger, "Ensemble flood forecasting: A review," *Journal of hydrology*, vol. 375, no. 3-4, pp. 613–626, 2009. 77

[139] H. Wu and D. Levinson, "The ensemble approach to forecasting: A review and synthesis," *Transportation Research Part C: Emerging Technologies*, vol. 132, p. 103357, 2021. 77

[140] S. Bhanja and A. Das, "Impact of data normalization on deep neural network for time series forecasting," *arXiv preprint arXiv:1812.05519*, 2018. 77

[141] S. Rahman and M. S. U. Zaman, "Time series sales forecasting: A hybrid deep learning regularization approach," in *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, pp. 1–6, IEEE, 2024. 77

[142] G. Kariniotakis, G. Stavrakakis, and E. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE transactions on Energy conversion*, vol. 11, no. 4, pp. 762–767, 1996. 77

[143] D. Caissie, "The thermal regime of rivers: a review," *Freshwater biology*, vol. 51, no. 8, pp. 1389–1406, 2006. 78

[144] B. W. Webb, D. M. Hannah, R. D. Moore, L. E. Brown, and F. Nobilis, "Recent advances in stream and river temperature research," *Hydrological Processes: An International Journal*, vol. 22, no. 7, pp. 902–918, 2008. 78

Appendix

A.1 Extensive descriptive statistics for each hour of the hourly precipitation data - Chapter 1

Table A.4

	Hour 1	**Hour 2**	**Hour 3**	**Hour 4**	**Hour 5**	**Hour 6**	**Hour 7**	**Hour 8**
Mean	0.015	0.013	0.013	0.017	0.015	0.017	0.017	0.017
Std	0.065	0.053	0.057	0.063	0.055	0.062	0.057	0.064
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	1.450	0.950	0.960	1.180	0.930	0.820	0.600	1.410
Kurtosis	192.276	139.157	92.930	109.755	87.825	56.917	36.565	157.600
Skewness	11.293	9.934	8.263	8.462	7.793	6.693	5.485	9.618
Entropy	1.303	1.323	1.371	1.445	1.440	1.476	1.502	1.515
	Hour 9	**Hour 10**	**Hour 11**	**Hour 12**	**Hour 13**	**Hour 14**	**Hour 15**	**Hour 16**
Mean	0.015	0.015	0.014	0.015	0.013	0.017	0.018	0.017
Std	0.054	0.059	0.058	0.061	0.048	0.068	0.069	0.073
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	0.850	0.910	0.860	0.940	0.580	1.040	1.090	1.230
Kurtosis	60.495	81.434	86.728	71.395	41.521	72.145	83.152	113.854
Skewness	6.549	7.721	8.105	7.352	5.842	7.354	7.824	9.199
Entropy	1.455	1.391	1.347	1.324	1.339	1.426	1.484	1.450
	Hour 17	**Hour 18**	**Hour 19**	**Hour 20**	**Hour 21**	**Hour 22**	**Hour 23**	**Hour 24**
Mean	0.018	0.018	0.017	0.015	0.014	0.0178	0.014	0.014
Std	0.072	0.071	0.077	0.069	0.054	0.077	0.061	0.056
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	1.270	1.370	1.250	1.140	0.700	1.470	1.100	1.040
Kurtosis	93.019	106.225	111.156	123.742	48.506	145.363	119.618	103.819
Skewness	8.089	8.317	9.333	9.777	6.088	10.146	9.279	8.314
Entropy	1.462	1.456	1.397	1.361	1.364	1.384	1.368	1.328

A.2 Descriptive Statistics for Gallatin, Cumberland, and Kingston, respectively - Chapter 2

Table A.5

Gallatin Fossil Plant

Water Temperatures																	
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Mean	69.454	70.508	74.138	72.228	76.041	71.185	71.174	68.969	73.063	71.884	72.518	71.586	72.323	69.111	70.764	71.078	71.614
Std	2.460	2.143	2.723	4.566	3.282	3.226	2.009	2.388	2.544	2.582	1.480	2.699	2.738	2.295	2.350	1.393	2.886
Min	63.900	64.100	65.310	63.405	60.785	64.100	66.558	63.218	62.905	65.835	65.068	66.720	64.133	61.723	65.115	67.488	64.255
25%	67.900	69.800	72.324	68.363	75.523	68.627	70.131	67.246	72.244	70.076	71.930	69.058	70.661	68.154	69.272	69.898	70.768
50%	69.032	70.500	74.763	71.095	76.988	71.625	71.433	69.280	73.540	72.453	72.835	72.010	73.325	69.396	70.843	71.124	72.330
75%	70.700	71.800	76.052	76.791	77.815	73.923	72.706	71.045	74.688	73.775	73.440	73.721	74.250	70.505	72.663	72.186	73.660
Max	79.400	76.800	82.023	82.468	83.210	81.435	79.633	78.160	81.088	77.253	78.753	77.788	77.580	73.388	75.205	77.340	82.225
Air Temperatures																	
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Mean	77.647	75.549	79.031	77.482	77.334	75.736	76.299	77.050	79.175	76.301	79.150	79.071	77.519	77.339	79.028	77.202	78.914
Std	7.886	7.621	8.788	9.613	9.752	7.822	8.306	8.420	8.196	8.602	7.569	7.898	8.376	8.082	8.655	8.250	8.388
Min	53.060	51.080	51.080	51.080	51.080	51.980	51.080	51.080	51.080	51.080	55.940	53.060	51.080	51.080	51.080	53.960	53.060
25%	71.960	71.060	73.040	71.060	71.060	69.980	71.060	71.960	73.940	71.060	73.940	73.040	71.960	73.040	73.940	71.060	73.040
50%	77.000	75.020	78.080	77.000	77.000	75.020	75.920	75.920	78.980	75.920	78.080	78.080	77.000	75.920	78.980	77.000	78.800
75%	84.020	80.960	86.000	84.920	84.020	80.960	82.040	82.940	86.000	82.940	84.920	84.920	84.020	82.940	84.920	82.940	84.920
Max	96.080	93.920	100.940	100.940	105.980	96.080	95.000	96.980	96.980	96.980	98.060	98.060	98.060	96.980	100.040	98.060	98.960
Flow Rates																	
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Mean	2615.898	7504.457	6913.986	10174.513	3676.486	10993.289	7413.444	11487.849	8784.443	11238.167	9245.602	13517.303	9901.375	10287.882	11551.573	8388.173	11411.557
Std	1670.053	6567.345	6744.185	8981.033	3423.306	10871.450	8519.466	8174.354	9350.548	9240.562	9606.160	10486.246	9540.018	7737.591	9811.125	9278.125	10797.461
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	1500.000	0.000	4450.000	3900.000	1000.000	0.000	0.000	7450.000	0.000	2600.000	0.000	2575.000	0.000	2100.000	0.000	0.000	0.000
50%	2500.000	7500.000	7400.000	7500.000	2200.000	7600.000	7350.000	10065.000	7390.000	7755.000	7430.000	15570.000	7605.000	8010.000	8480.000	7830.000	8240.000
75%	3000.000	11700.000	7500.000	15800.000	7500.000	17300.000	15600.000	16200.000	15800.000	16260.000	16260.000	25540.000	16470.000	15510.000	17050.000	16330.000	17200.000
Max	16000.000	27300.000	64500.000	50400.000	18200.000	35400.000	29950.000	37870.000	26850.000	42320.000	34935.000	34390.000	39812.000	28150.000	29410.000	27765.000	43275.000

Table A.6

Cumberland Fossil Plant

Water Temperatures

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Mean	77.760	72.650	73.931	71.563	76.594	69.786	69.588	67.001	71.570	69.193	69.202	68.674	69.897	66.333	67.429	68.142	67.707
Std	4.212	2.153	3.448	4.794	5.005	3.173	1.956	2.473	2.570	2.859	2.030	2.902	2.750	2.212	2.517	1.766	2.984
Min	64.900	65.700	63.900	62.800	58.100	62.600	64.400	60.800	63.140	62.400	62.600	63.700	62.800	58.800	61.700	60.100	58.500
25%	75.200	71.400	71.200	66.900	74.500	67.300	68.400	65.480	70.415	67.100	68.500	65.800	68.500	65.300	65.700	66.900	66.900
50%	78.100	73.000	74.100	72.000	76.800	70.300	69.800	67.460	72.222	69.600	69.600	69.100	70.500	66.600	67.300	68.200	68.400
75%	80.600	74.100	76.600	76.100	79.950	72.170	71.100	68.900	73.220	71.200	70.500	71.200	71.800	67.800	69.400	69.400	69.600
Max	90.900	78.800	82.900	82.800	89.600	77.200	75.600	75.380	77.000	77.900	74.800	76.800	77.400	73.200	74.700	76.100	75.000

Air Temperatures

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Mean	77.647	75.549	79.031	77.482	77.334	75.736	76.299	77.050	79.175	76.301	79.150	79.071	77.519	77.339	79.028	77.202	78.914
Std	7.886	7.621	8.788	9.613	9.752	7.822	8.306	8.420	8.196	8.602	7.569	7.898	8.376	8.082	8.655	8.250	8.388
Min	53.060	51.080	51.080	51.080	51.080	51.980	51.080	51.080	51.080	51.080	55.940	53.060	51.080	51.080	51.080	53.960	53.060
25%	71.960	71.060	73.040	71.060	71.060	69.980	71.060	71.960	73.940	71.060	73.940	73.040	71.960	73.040	73.940	71.060	73.040
50%	77.000	75.020	78.080	77.000	77.000	75.020	75.920	75.920	78.980	75.920	78.080	78.080	77.000	75.920	78.980	77.000	78.800
75%	84.020	80.960	86.000	84.920	84.020	80.960	82.040	82.940	86.000	82.940	84.920	84.920	84.020	82.940	84.920	82.940	84.920
Max	96.080	93.920	100.940	100.940	105.980	96.080	95.000	96.980	96.980	96.980	98.060	98.060	98.060	96.980	100.040	98.060	98.960

Flow Rates

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Mean	3963.422	11432.275	9859.255	13070.687	6156.301	18778.876	9589.003	16096.549	11464.991	17315.839	13454.862	17342.367	15217.122	17021.141	14821.124	11803.586	14670.594
Std	2188.529	9405.387	10299.126	10243.026	5456.632	17594.634	5714.268	11218.028	8950.947	12701.602	9135.690	10442.128	9811.366	9623.151	6371.057	7495.927	10263.641
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	3000.000	6500.000	4000.000	6600.000	0.000	6700.000	6600.000	6900.000	6700.000	12800.000	660.000	12150.000	6700.000	13400.000	13000.000	6600.000	7100.000
50%	3000.000	6700.000	9700.000	6700.000	6600.000	14000.000	6900.000	13400.000	6900.000	14200.000	13200	14000.000	13600.000	14000.000	13700.000	13200.000	14200.000
75%	3000.000	14050.000	13200.000	20900.000	6700.000	21600.000	13600.000	21350.000	13800.000	21100.000	20400.000	22200.000	21000.000	21725.000	20400.000	20400.000	21900.000
Max	14200.000	52300.000	98200.000	45100.000	33200.000	83500.000	29600.000	70700.000	63400.000	115200.000	54800.000	54900.000	66200.000	65000.000	40800.000	36400.000	47400.000

Table A.7

Kingston Fossil Plant

Water Temperatures									
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Mean	79.024	75.909	80.063	77.297	72.317	74.883	76.917	71.679	81.293
Std	3.444	4.147	4.079	6.513	2.659	4.008	3.051	4.040	3.287
Min	69.800	65.100	70.200	62.000	67.100	61.300	69.343	64.700	70.300
25%	77.400	73.200	76.900	70.932	69.800	72.400	74.813	68.900	79.000
50%	79.550	75.800	79.500	79.200	72.400	75.500	77.000	70.500	81.500
75%	81.400	79.100	83.800	83.000	74.400	77.500	79.300	74.697	83.900
Max	86.800	87.000	88.500	91.800	77.800	90.600	84.900	82.757	88.500
Air Temperatures									
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Mean	75.805	74.078	78.378	76.842	75.232	73.941	74.068	75.587	78.626
Std	7.922	6.755	8.242	9.143	8.698	6.876	7.116	7.984	8.039
Min	53.060	48.020	51.980	53.060	46.040	53.060	50.000	46.940	50.000
25%	69.980	69.080	73.040	69.980	69.980	69.080	69.080	69.980	73.040
50%	75.020	73.040	78.080	75.920	75.020	73.040	73.040	75.020	78.080
75%	82.040	78.980	84.920	84.020	80.960	78.980	78.980	80.960	84.920
Max	93.920	91.940	96.980	96.980	96.980	91.940	91.940	96.980	96.980
Flow Rates									
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Mean	94.754	751.862	219.813	579.417	195.665	1097.710	234.184	998.401	94.895
Std	124.596	1539.667	270.143	1090.701	555.364	3985.839	376.282	2149.838	134.087
Min	7.000	56.000	18.000	6.000	6.000	25.000	19.000	51.000	6.000
25%	23.000	124.000	58.000	72.000	25.000	124.000	58.000	117.000	25.000
50%	44.000	300.000	101.000	191.000	49.000	324.000	135.000	249.000	52.000
75%	111.000	682.000	269.000	571.000	145.000	614.000	294.000	806.000	96.000
Max	903.000	18836.000	1816.000	7780.000	6380.000	50300.000	5440.000	18400.000	1000.000

Vita

Rachel Wood-Ponce graduated from Lee University in 2020 with a Bachelor of Science in Mathematics with a Pre-Engineering Emphasis. She then started the PhD program in Industrial and Systems Engineering at the University of Tennessee at Knoxville. In August of 2024, she received her concurrent Master of Science degree in Industrial and Systems Engineering before finishing her PhD in December of 2024. Her doctoral research focuses on the application of machine learning for prediction and forecasting in hydrology.