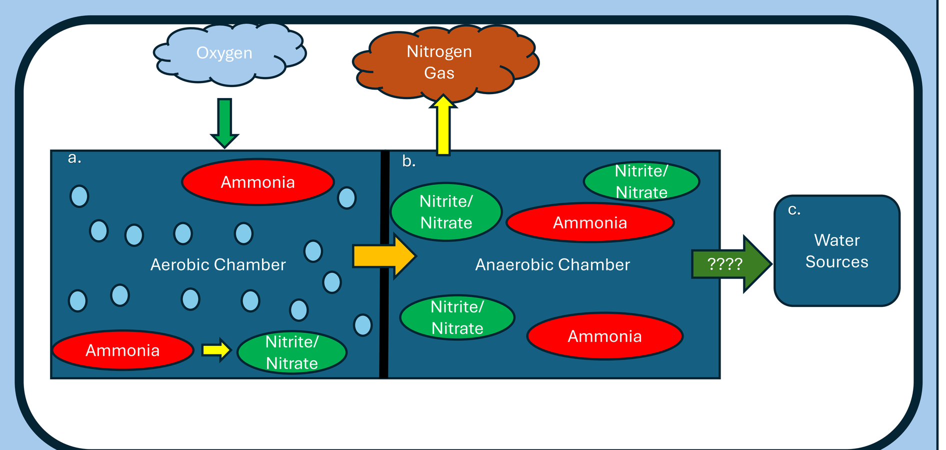
Machine Learning application in predicting nitrogen discharge from a local wastewater treatment plant

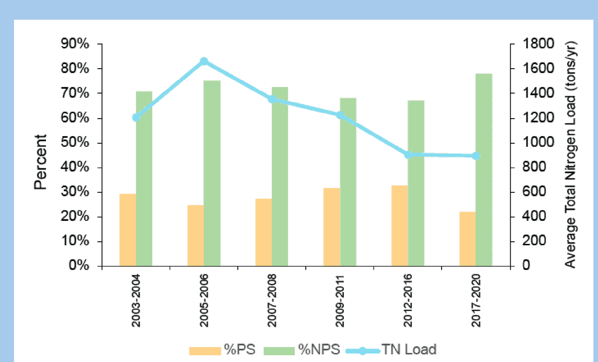
Liam Wistran and Dr. Matthew Ferby
Department of Civil & Environmental Engineering, University of New Hampshire, Durham, NH 03824

University of New Hampshire

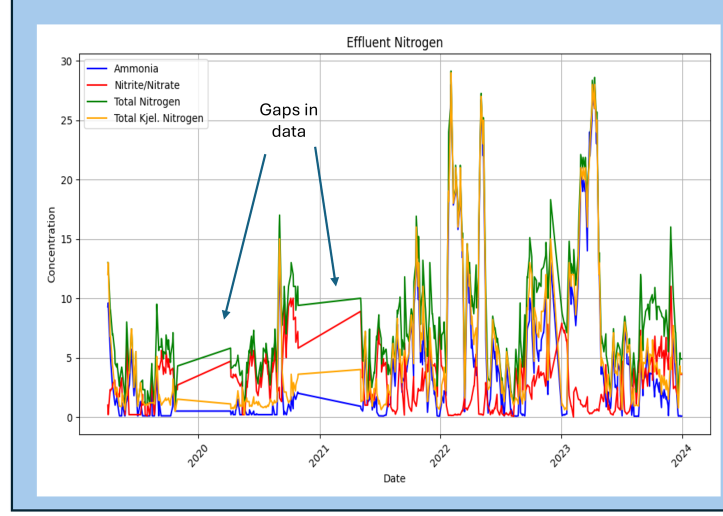
Introduction

- Wastewater Treatment Plant (WWTP) processes remove nitrogen by converting nitrogen species (e.g., ammonia, nitrate, nitrite) into nitrogen gas
- In order determine its effectiveness, WWTPs must collect samples of the effluent on how much nitrogen is going out and alter their processes as needed





- Currently, the Great Bay receives about 800 tons of total nitrogen every year (point and non-point sources)
- Nitrogen can cause eutrophication, which can create dead zones void of oxygen and affect the health of the aquatic ecosystem



- Infrequent sampling can result in inaccurate estimation of WWTP nitrogen discharge
- While nitrogen data gaps exist, other important features (e.g., flow) could provide insight to nitrogen levels
- Using machine learning (ML) and WWTP data can be used to predict the gaps in nitrogen data
- Machine Learning (ML) models can accurately predict the gaps in nitrogen discharge data, so they can be utilized by wastewater treatment plants to help meet EPA discharge limits

Objectives

(1) Organize Weather and Water Treatment Data
(2) Code in Python to create a Correlation Matrix and Machine Learning model Comparison
(3) Collect data from the code to create a more accurate prediction model
(4) Predict gaps in Nitrogen Discharge Data

Methodology

Data Acquisition

Raw Data
- Collected data 2019-2023 at the Durham Wastewater Treatment Plant such as Wastewater flow (MGD), Biological Oxygen Demand (mg/L), and various Nitrogen Species (mg/L)
- Collected data from Durhams 2019-2023 Temperature (°F) and Precipitation (Inches) Data

Preprocessing Data
- Combined wastewater and meteorological data set into one data set (excel)
- Interpolated missing data for Biological Oxygen Demand (BOD)

Output Data and Results

Correlation Matrix
- Coded a Correlation Matrix through Python
- Found which features had the highest correlations
- Used features with highest correlation to predict the models

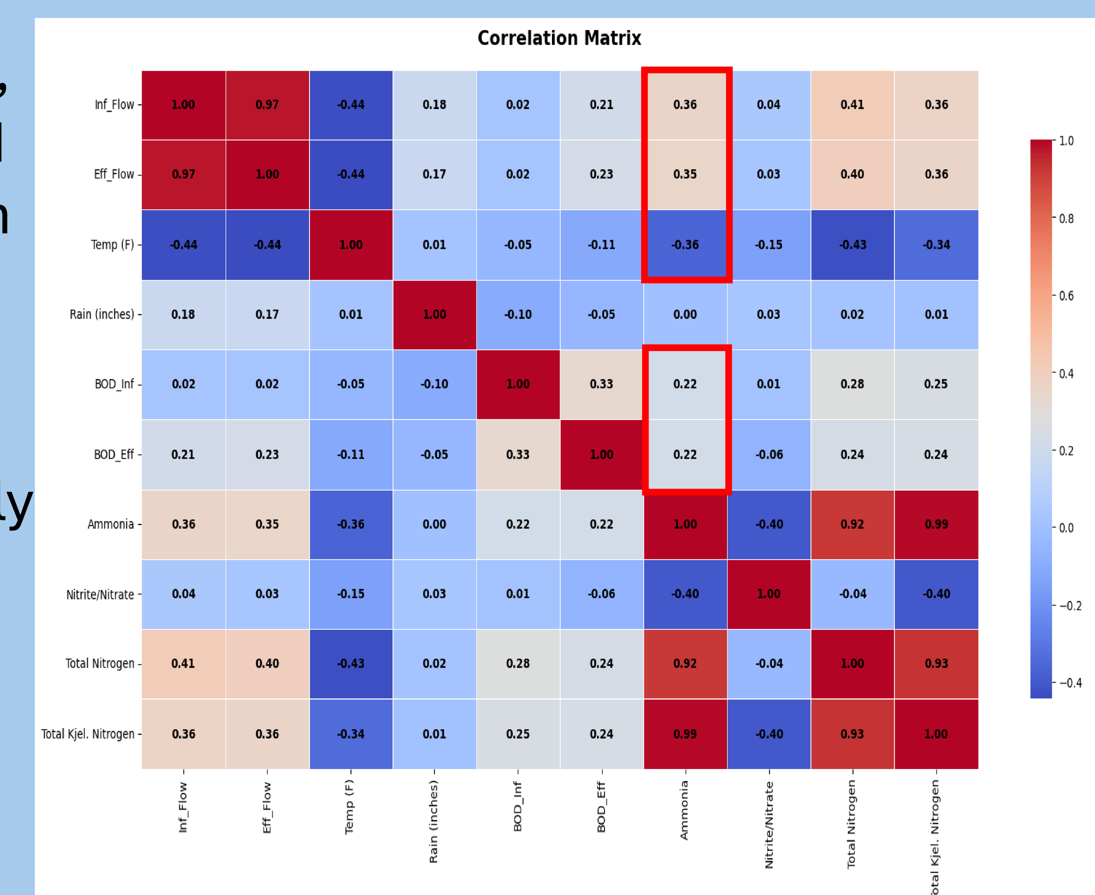
R^2 Graphs
- Created a code to test multiple machine learning models (OLS, AdaBoost, SVM, Decision Tree, Random Forest, KNN, XG Boost, and ANN) for R^2 value
- Used features found in correlation matrix
- Top three models used to predict the values

Predicted Models
- Used the features from the correlation matrix and the models from the R^2 graphs to create a graph predicting the values
- Placed it against the real graph to compare
- Root Mean Squared Error calculated to find which model is closed to the R^2 value

Results

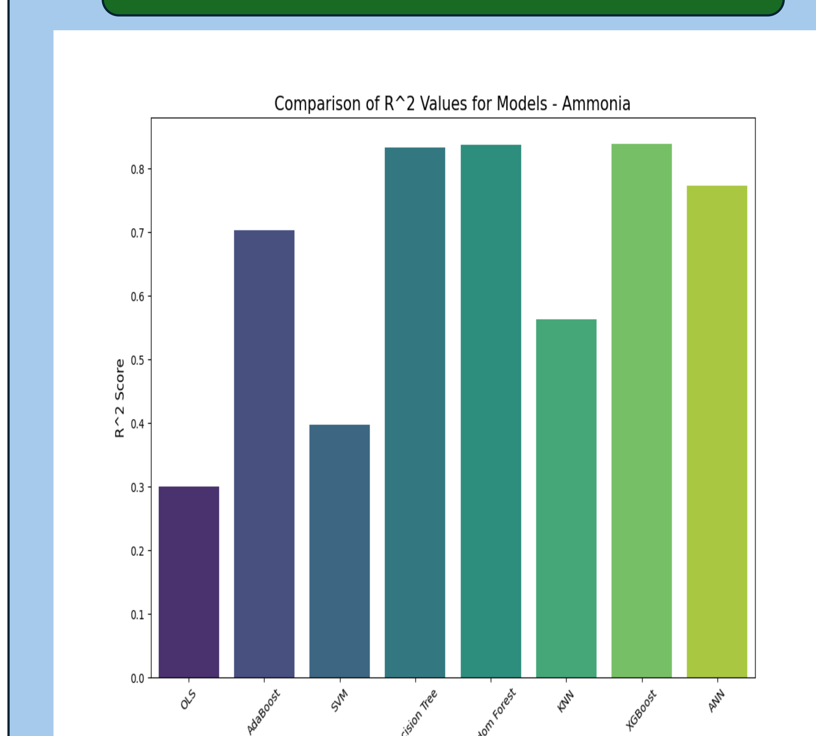
Correlation Matrix

- Wastewater Flow, Temperature, and Biological Oxygen Demand **had the highest correlations**
- Nitrate/Nitrite only had low inverse correlations
- Precipitation had very little correlation

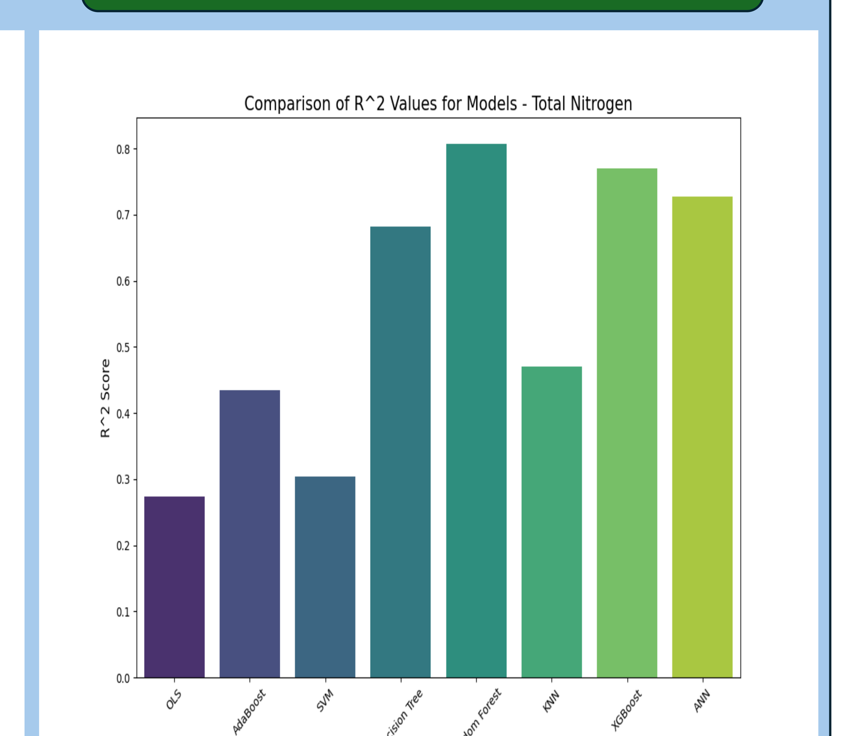


Machine Learning Model Comparison of various Nitrogen Species

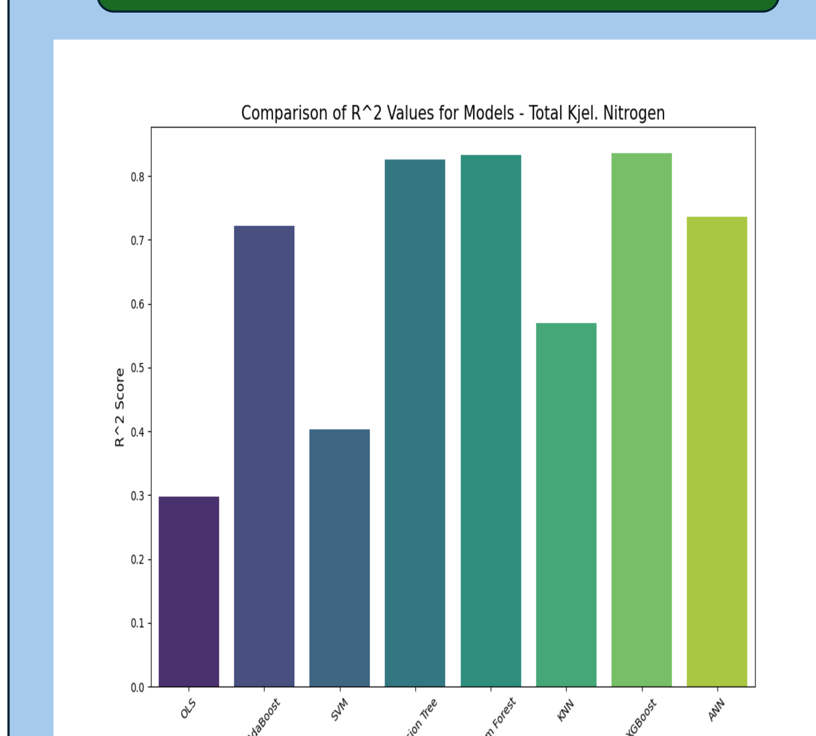
Ammonia



Total Nitrogen



Total Kjeldahl Nitrogen



- Random Forest, Extreme Gradient Boosting (XG Boost), and Artificial Neural Network (ANN) yielded the highest R^2 Value
- More complex models produced the highest correlation
- Nitrite and Nitrate is not well predicted by the data
- Though is still possible to predict through subtracting predicted Total Nitrogen from TKN values

Results

Prediction Graphs: (1) RMSE between 3-4, (2) XG Boost and ANN had the highest lowest error, and (3) predicted values close to actual Values

Conclusions

This work proved that machine learning can be utilized to handle infrequent sampling of nitrogen discharge from local water treatment plants. The models were able to correctly predict and match the flow of nitrogen species. Self correcting models like XG Boost and ANN preformed the best, while simpler models fell behind performance wise. With additional data points Wastewater Treatment Plants can more accurately predict missing data. This research demonstrates how machine learning can be used to predict nitrogen discharge and other effluent factors in water treatment plants more accurately. The results from this work will help guide a larger effort in understanding economical and energy requirements for decreasing point source nitrogen contributions to the Great Bay.

Acknowledgments

Robert (Bob) Henry Student Assistance Fund for funding this research project

References: (1) Xu, Y., et al. '*Artificial Intelligence-Assisted Prediction of Effluent Phosphorus in a Full-Scale Wastewater Treatment Plant with Missing Phosphorus Input and Removal Data*' ACS EST Water 2024,4, 880-889 (2) *Figure 3. Brief explanation of the eutrophication phenomenon*. (n.d.). iGem Thessaly (3) *Section 2 - basics of wastewater treatment*. MASSTC. (2017, June 16). https://www.masstc.org/homeowners/ia-septic-system-maintenance/ia-compendium/2-basics-of-wastewater-treatment (4) State of our estuaries | indicators - nutrient loading. (n.d.). https://www.stateofourestuaries.org/indicators/nitrogen-loading (5) Wastewater Treatment Facility Monthly Operations Report. (n.d.). Durham. (6) National Centers for Environmental Information (NCEI). (n.d.). *Daily summaries station details*. Daily Summaries Station Details: DURHAM, NH US, GHCND:USC00272174 | Climate Data Online (CDO) | National Climatic Data Center (NCDC). https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USC00272174/detail