Modeling PFAS in New Hampshire Background Soils

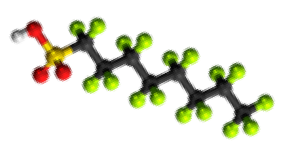
Camila Cuellar Quiroga, William Fortin, Kylie Krivis, William Stark
Department of Environmental Engineering, University of New Hampshire, Durham, NH 03824

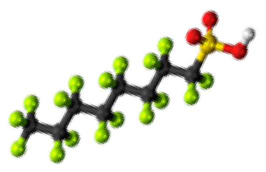
Introduction

This project aims to quantify PFAS levels in background soils across New Hampshire to help distinguish ambient levels, from contamination due to localized sources. To achieve this, an analysis of the data from 100 soil samples collected by USGS was used to investigate how soil properties may influence PFAS distribution. Based on these relationships, a **random forest decision tree model** was made to quantify the correlation the soil features pH, total organic carbon, moisture, and soil protein, vs measured PFOS concentrations.

What are PFAS?

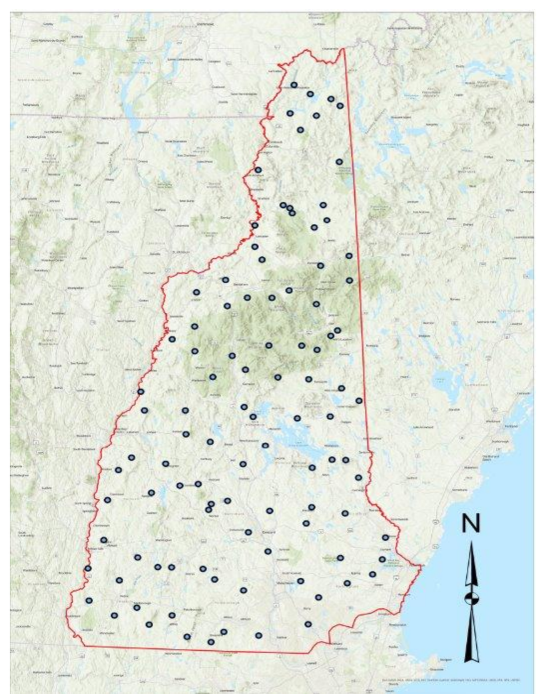
- Per- and polyfluoroalkyl substances are a large group of complex synthetic chemicals known for their resistance to heat, water and oil.
- They do not degrade in the environment, leading to long term contamination.
- Human exposure is linked to potential health effects such as immune disruption, developmental issues, and cancer.





USGS Study

•Sampling conducted in 2021 by USGS at 100 sites across the state.
•Analyzed for 34 PFAS analytes and key soil characteristics
•Equal-area grid approach used to ensure spatial coverage
•500-meter buffer maintained around known PFAS sources to avoid direct contamination
•Composite samples collected at two depths: 0–6 inches and 6–12 inches
•Strict PFAS-free protocols followed to minimize contamination during sampling

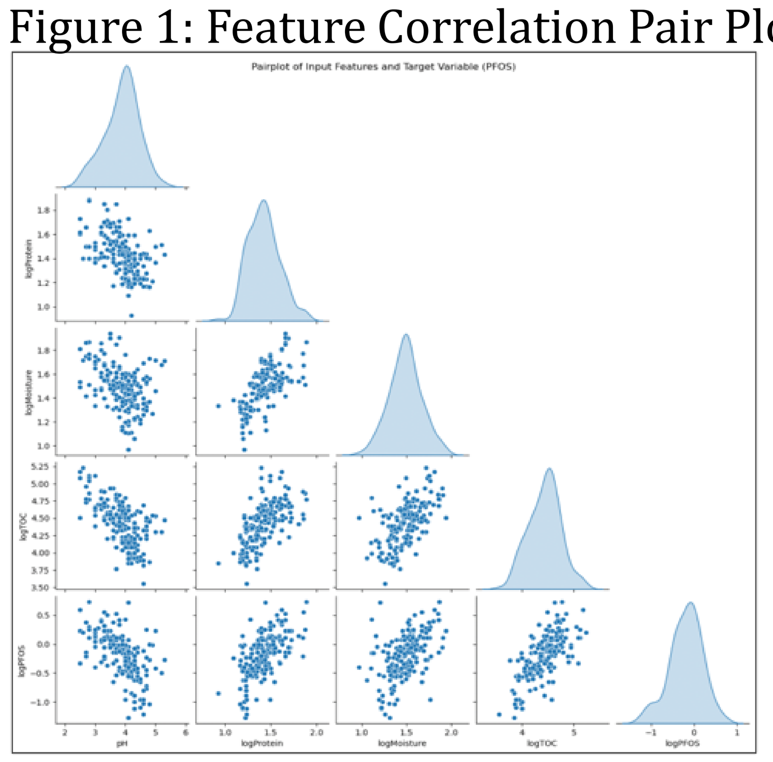


PFAS Deposition and Background Sources

•Background PFAS refers to diffuse, non-localized contamination—not tied to a direct spill or site release.
•Primary transport in this study was assumed to be atmospheric deposition.
•PFAS emissions from sources like manufacturing plants and landfills can enter the air.
•These compounds settle onto soil through dust and precipitation.
•Once deposited, PFAS can adsorb onto undisturbed soils such as forests and wetlands.

Model Framework

Figure 1: Feature Correlation Pair Plots



Inputs:
Soil Parameters
- pH
- TOC
- Soil Moisture
- Soil Protein

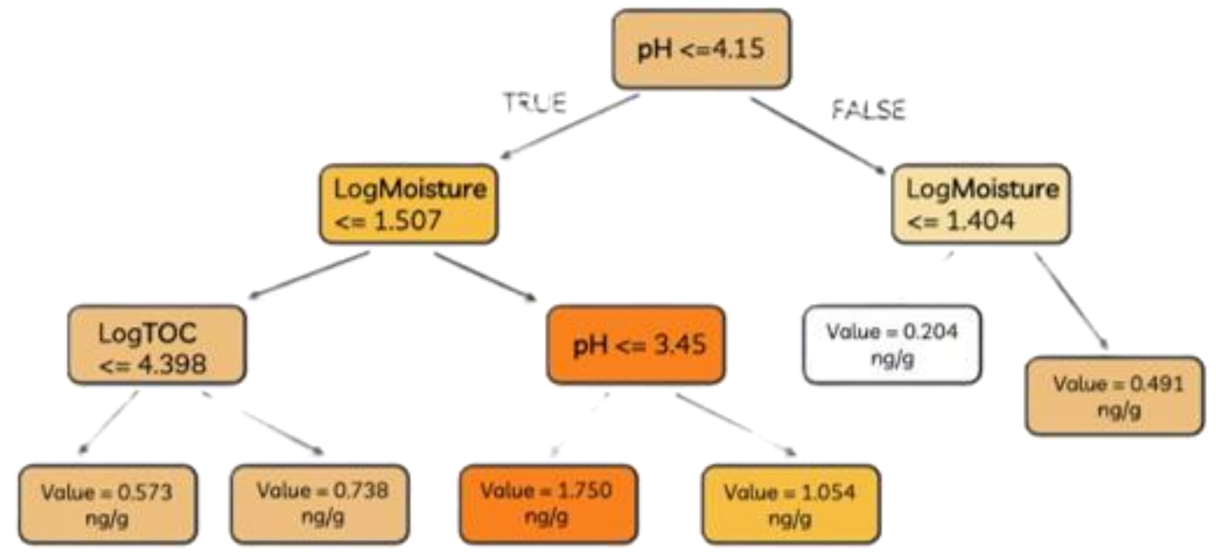
Measured PFOS

Output:
Predicted PFOS concentration based on relationships with given soil parameters

Modeling Design Steps

1. Feature Correlation Analysis with Pair Plots
2. Model Training Using 75% of Data Points
3. Parameter Tuning and Model Fitting
4. Random Forest Construction
5. Model Validation Using 25% of Data Points

Figure 2: Decision Tree Visualization



Modeling Results

The model is trained on 75% of the data to find patterns between the soil parameters and PFOS. But the model must next be evaluated with unseen data to quantify its true performance, which is why a 25% test set is used.

After completing a 4-Fold cross validation on our test set,
The overall performance of the model was found to be an R^2 value of 0.51. **This means the model understands 51% of the variation in the data set.**

Figure 4: Soil Parameter Importance

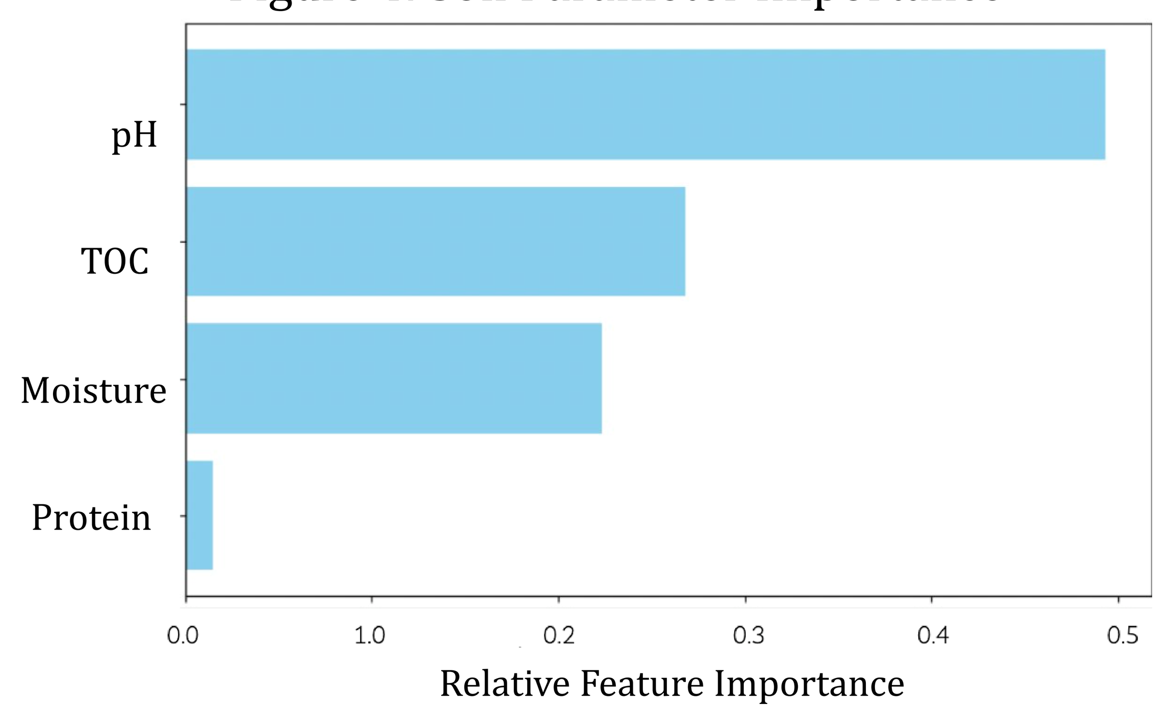
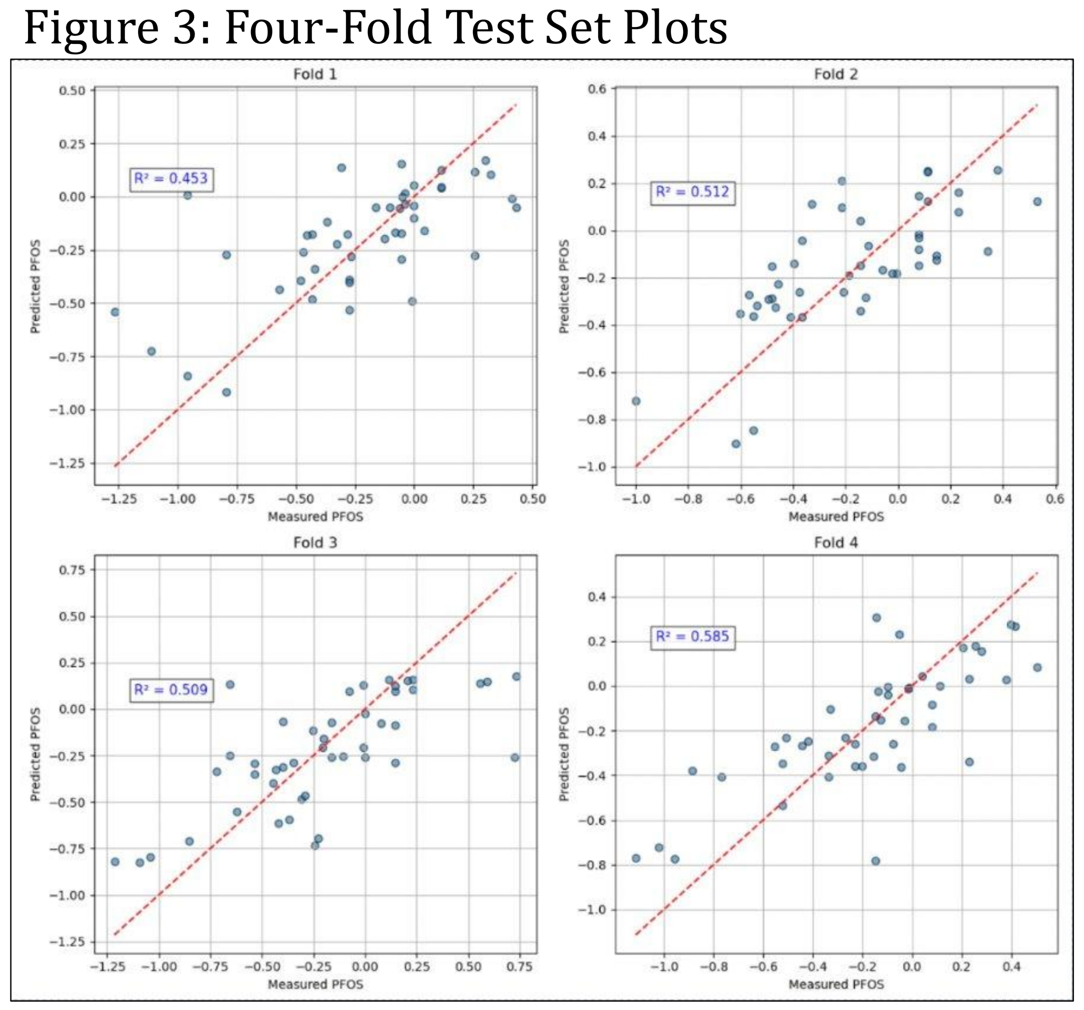


Figure 3: Four-Fold Test Set Plots



Feature Importance is a measure of how much a variable reduces prediction error across all the decision trees in the forest.

The feature importance chart from our Random Forest model shows which soil characteristics had the greatest impact on predicting PFOS concentrations. It is clear that pH was the most influential, followed by total organic carbon (TOC), and moisture. **These variables likely affect how PFAS compounds bind to or move through the soil in the environment.**

Industry Application

1. With new NHDES regulations, and increased significance industry, greater understanding of the **fate and transport** of PFAS is extremely important.
2. Quantifying which soil parameters have a higher probability of containing elevated PFAS can make more **selective sampling plans**, and decrease cost.
3. Understanding how background PFAS fluctuates with different soil parameters makes **more efficient site characterization**, and easier detection of local exposure.

Conclusion

This study shows the potential for using machine learning to predict background PFAS concentrations in soil based on key soil properties. Understanding which factors influence PFAS distribution can help inform future sampling strategies and improve risk assessment in areas without known contamination sources. While our research focused specifically on PFOS, the same framework could be applied to any PFAS.

The overall performance of our model showed a **51%** understanding of the variation between the inputted soil parameters and measured PFOS. The limitations of this model were not due to a lack of fit, but rather other soil characteristics that are likely influencing background concentration throughout the state. The predictive power could be increased by including more soil parameters, such as soil type, as well as larger data set for the model to train on.

Overall, this project is an early step in understanding how PFAS sorb to soil, and variation in background due to changing soil conditions. While the model's predictive power is only moderate, there were multiple useful insights found in this research.

References

Jon Dewitz. National Land Cover Database (NLCD) 2019 Products (Ver. 3.0, February 2024). U.S. Geological Survey, 29 Feb. 2024. DOI.org (Datacite), https://doi.org/10.5066/P9KZCM54.

Santangelo, Leah M., et al. Statewide Survey of Shallow Soil Concentrations of Per- and Polyfluoroalkyl Substances (PFAS) and Related Chemical and Physical Data across New Hampshire, 2021. U.S. Geological Survey, 2022. DOI.org (Datacite), https://doi.org/10.5066/P9KG38B5

Technical Summary Report: Proposed Soil Remediation Standards (SRS) for Perfluorooctanoic Acid (PFOA), Perfluorooctane Sulfonic Acid (PFOS), Perfluorohexane Sulfonic Acid (PFHxS), and Perfluorononanoic Acid (PFNA). NHDES, Oct. 2023, https://www.des.nh.gov/sites/g/files/ehbemt341/files/documents/r-wmd-23-03.pdf.

US EPA, O. (2024, June 6). Random forest modeling [Data and Tools]. https://www.epa.gov/streamflow-duration-assessment/random-forest-modeling

Acknowledgements

Project Sponsors: Harrison Roakes and Gregory Bush

SANBORN | HEAD
University of New Hampshire College of Engineering and Physical Sciences

Faculty Advisors: Dr. James Malley and Dr. Fei Han

This poster template provided courtesy of
UNH ESRC Poster Printing Services

Trust us to make your poster look **GREAT!**

Website: http://posters.unh.edu
Poster Guide: http://goo.gl/1E7TJY

DELETE THIS SLIDE BEFORE PRINTING