Understanding Environmental Impact of Transportation Systems With Causal AI

Anais Barja, Alexander Flores | Faculty Advisor: Xinghui Zhao | School of Engineering and Computer Science | WSU Vancouver

Pairing a causal ai model with real time environmental analysis could help identify developing trends early as well as show how things could've gone differently through counterfactuals

Introduction

- We are analyzing the causal effects of transportation related variables on the environment.
- Chemicals from vehicles on the road are known to pollute the water via run off.
- The chemicals in our water sources can affect us and wildlife.

DoWhy

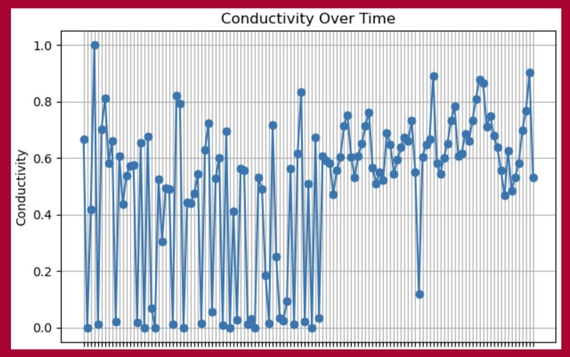
- Causal AI is a type of machine learning model that uses a causal approach to understand why something happened.
- We are using DoWhy, a python library from Microsoft, to perform causal analysis.
- We are first focusing on water quality variables but causality analysis can be used in many areas.
- Causal analysis is often used in the medical field to understand certain effects of treatments.

Previous Work

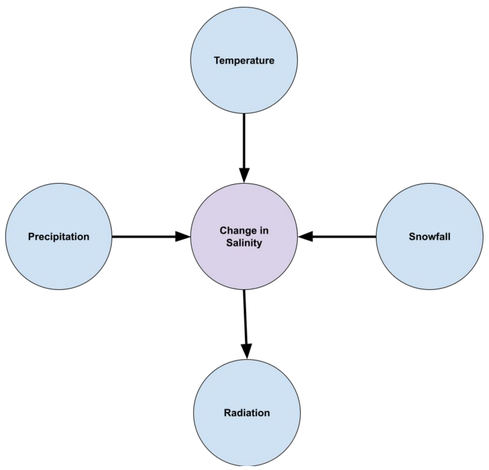
- Previously we created a remote sensing system to collect local water quality data in real time.
- We used a variety of sensors, solar panel, battery, and an Arduino UNO
- Due to insufficient data, we are suing synthetic and open source water quality datasets.

Method

- With Causal AI, "what if" data known as counterfactuals can be generated to understand what could have happened if something was different.
- We are using them to help understand trends in salinity levels and how they could have been different based on other variables



- The illustration below shows the causal relationships in the model we are using to analyzes changes in salinity or other water properties..

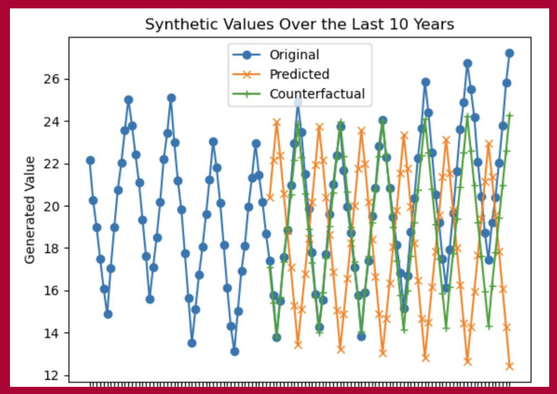


Datasets

- A neural network model was trained to predict both the true values and false values based on modified variables that affect the final result.

Synthetic Salinity

- For the years of 2000 - 2023, monthly salinity values were generated. The data mimics similar behavior to real salinity where levels tend to fluctuate during certain parts of the year.
- This data was generated to help compare results when analyzing real salinity values.

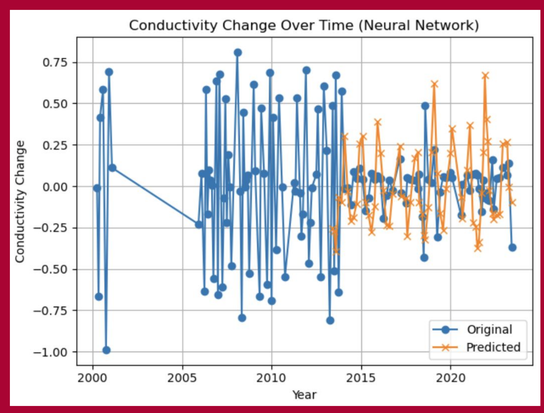


United States Geological Survey

- We are also using monthly water quality data from USGS for the years of 2000 - 2023. This data was combined with weather data from NOAA and solar radiation data from AgriMet.
- All this data has been grouped and averaged by month.

Results

- We found that the counterfactual and the neural network both succeeded in closely predicting correct values based on the patterns.
- The graph on the left shows a purposeful shift in pattern for the values being predicted by the neural network.
- The counterfactual was able to correctly determine what the correct value should be based on the modified synthetic data generated from the neural network.
- Our initial tests with the water quality data from USGS shows a reasonable estimate in predicting changes in conductivity using a neural network model.



Future Work

- We plan to further test and compare the neural network with the causal one to determine the benefits and limitations of using each.