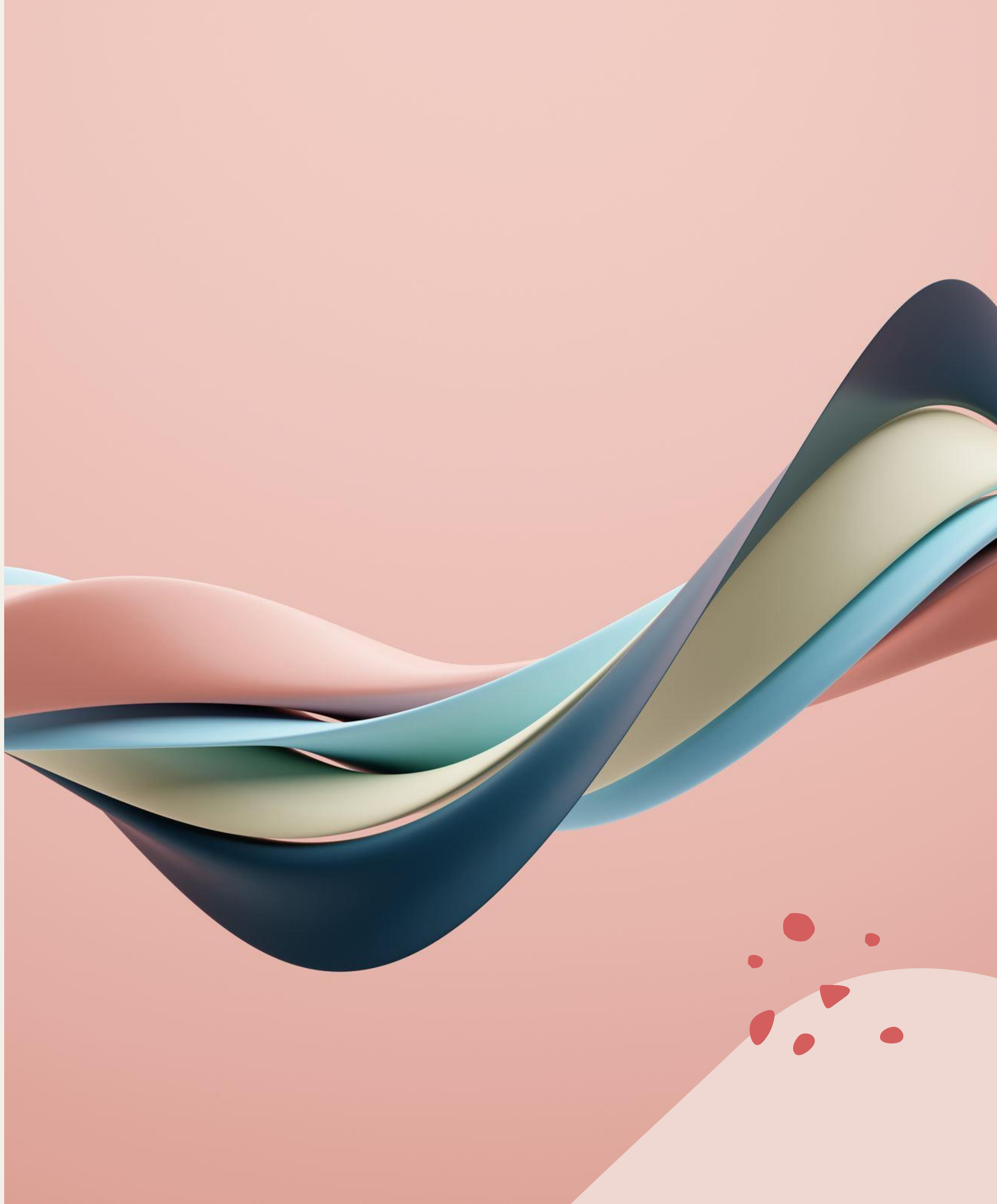
Drawing Causal Relationships Between Climate Variables Utilizing Causal Forests

Kelsey Hawkins

CPSC 541 – Spring 2025



Background

- Using water usage and climate variables to find which variables have causal effects and their magnitude of causality
- 100,000 residential houses
- Data from 2018-2023
- 6 Climate Variables:
    - Precipitation
    - Evapotranspiration
    - Average Solar Radiation
    - Average Air Temperature
    - Average Relative Humidity
    - Average Wind Speed

Method – Causal Forest

- Double Machine Learning (DML) :
  - Estimation of causal effects while controlling for high-dimensional confounders.
  - Separates estimation of treatment and outcome variables for the final causal estimate.
  - Uses Causal Forests and XGBoost Regressor combined
- Causal Forest:
  - Ensemble non-parametric model for estimating heterogeneous treatment effects
  - Model non-linear relationships and complex interactions. Output is Average Treatment Effect and Conditional Average Treatment Effect

Results

- Evapotranspiration is the most causally significant variable
  - ATE: 1.33 CCF, 95% Confidence interval (0.87, 1.78)

- Precipitation is second most causally significant variable
  - ATE: -0.1561, 95% confidence interval (-0.19, -0.12)

- Average Wind speed is third most causally significant variable
  - ATE: -0.044, 95% confidence interval (-0.054, -0.034)

Conclusions

- Climate variables contribute heavily to water usage
- Not all climate variables contribute equally
- This can be used for predictive modeling, and marketing tactics for water conservation

- Causal forests and DML are a robust way to estimate causal effects with climate variable and water usage data.