Optimal Policy Learning Under Spatial Dependence With Applications to Groundwater in Wisconsin

Xindi Lin†, Christopher Zahasky‡, Hyunseung Kang†

†Department of Statistics, University of Wisconsin--Madison, ‡ Department of Geoscience, University of Wisconsin--Madison

Introduction

- ***Motivation: Data-Driven, Cost-Efficient Groundwater Policy***
  - **A necessary trade-off**: increasing water well depth improves groundwater quality but increases installation costs.
  - **Goal:** Determine the *minimum well depth* required to meet the *public health standards* for contaminants in groundwater.
- ***Method: Spatial Minimum Resource Threshold Policy (spMRTP)***
  - A Gaussian process model for **spatial dependence** of contaminants in groundwater.
  - Policy learning via risk minimization with a novel, doubly robust loss function.
  - **Computational efficiency** via the *Vecchia approximation*.
- ***Application: Nitrates in Wisconsin Groundwater***

Framework

- ***Notations***
  - $\mathscr{S} = \{s_1, s_2, \ldots, s_n\} \subset \mathscr{D} \subset \mathbb{R}^2$: set of spatial locations where n observations are measured.
    - Y_s: observed concentration of contaminate at location s
    - $A_s \in \mathscr{A} \subset \mathbb{R}$: observed well depth at location s
    - $Y_s(a)$: potential concentration of contaminants at location s and depth $a \in \mathscr{A} \subset \mathbb{R}$
    - X_s: measured spatial covariates at location s
    - U_s: unmeasured spatial covariates at location s
- ***Assumptions***
  - Causal consistency: $Y_s = Y_s(A)$ almost surely.
  - Strong ignorability: $A_s \perp Y_s(a) \mid X_s$ and $p(a \mid X_s = x) > 0$ for all a, x .
  - Spatial unconfoundedness: $U_s \perp A_s \mid X_s$.
  - Additive, semiparametric, spatial structural model: $Y_s = \mu(X_s, A_s) + U_s + \epsilon_s$.
    - $\mu(X, A)$: nonparametric, monotonically decreasing function w.r.t. A for all X.
    - U_s: mean-zero, Gaussian process, i.e., $E[U_s \mid A_s, X_s] = 0$
    - ϵ_s: mean-zero, i.i.d. measurement error, i.e., $E[\epsilon_s \mid A_s, X_s, U_s] = 0$
- ***Definition of spMRTP $\theta^*(x_{s_0})$***
  - $\mathscr{T}$: target threshold for the outcome (e.g., nitrate concentration is less than $\mathscr{T}$=10 mg/L)
  - Given measured covariates at new location s_0 (i.e., x_{s_0}), the *spMRTP is defined as*

$$\theta^*(x_{s_0}) = \arg\min_{a \in \mathscr{A}} \; a, \text{ such that } E[Y_{s_0}(a) \mid X_{s_0}] + U_{s_0} \leq \mathscr{T} .$$

Identification and Estimation

Identification

- Under assumptions above, we have the following identification results
  - $E[Y_{s_0}(a) \mid X_{s_0}] = E[Y_{s_0} \mid A_{s_0} = a, X_{s_0}] = \mu(X_{s_0}, a)$
  - U_{s_0} can be approximated by conditional mean $E[U_{s_0} \mid \{U_s\}_{s \in \mathscr{S}}]$

Doubly Robust Nonparametric Estimation

- Assume $\theta \in \Theta$ where Θ is a function class defined on $\mathscr{D} \mapsto \mathscr{A}$.
- STEP 1: **Estimation of nuisance parameters**
  - Estimate $\hat{\mu}$ and $\hat{p}(a \mid x)$ —> Estimate covariance function with $Y - \hat{\mu}$—> Estimate $\hat{E}[U(s) \mid \mathscr{S}]$ by kriging.
- STEP 2: **Doubly robust risk minimization**

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \left(\mathscr{T} - \hat{\mu}(X_i, \theta(s_i)) - \hat{E}[U(s_i) \mid \mathscr{S}] - \frac{\{Y_i - \hat{\mu}(X_i, \theta(s_i)) - \hat{E}[U(s_i) \mid \mathscr{S}]\} K_\delta(A_i - \theta(s_i))}{\hat{p}(A_i \mid X_i)} \right)^2 .$$

Simulation Studies

- **Simulation Setting**
  - Let $\mathscr{D} = [0,1]^2$ and $\mathscr{S}$ be a uniform 50×50 grid on $\mathscr{D}$.
  - $(X_1, \ldots, X_5)$ are i.i.d. normally distributed.
  - $A \mid X_1, \ldots, X_5$: beta distributed with logistic mean model.
  - $U(s)$: gaussian process with range = 0.2.
  - $Y \mid A, X_1, \ldots, X_5$: normal distribution with non-linear mean.
- **Completing Methods**
  - Indirect method: $\hat{\theta}(s) = inf_{a \in \mathscr{A}} \left\{ a \mid \hat{\mu}\{X(s), a\} + \hat{E}\left[U_s \mid \{U_s\}_{s \in \mathscr{S}}\right] \right\}$.
  - Non-spatial MRTP: $\hat{\theta}(s) = inf_{a \in \mathscr{A}} \left\{ a \mid \hat{\mu}\{X(s), a\} \right\}$.
- **Implementation**
  - Tuning parameters are selected via cross-fitting.
  - Use generalized linear regression to estimate $p(a \mid x)$.
  - Use linear regression to estimate $\mu(x, a)$.

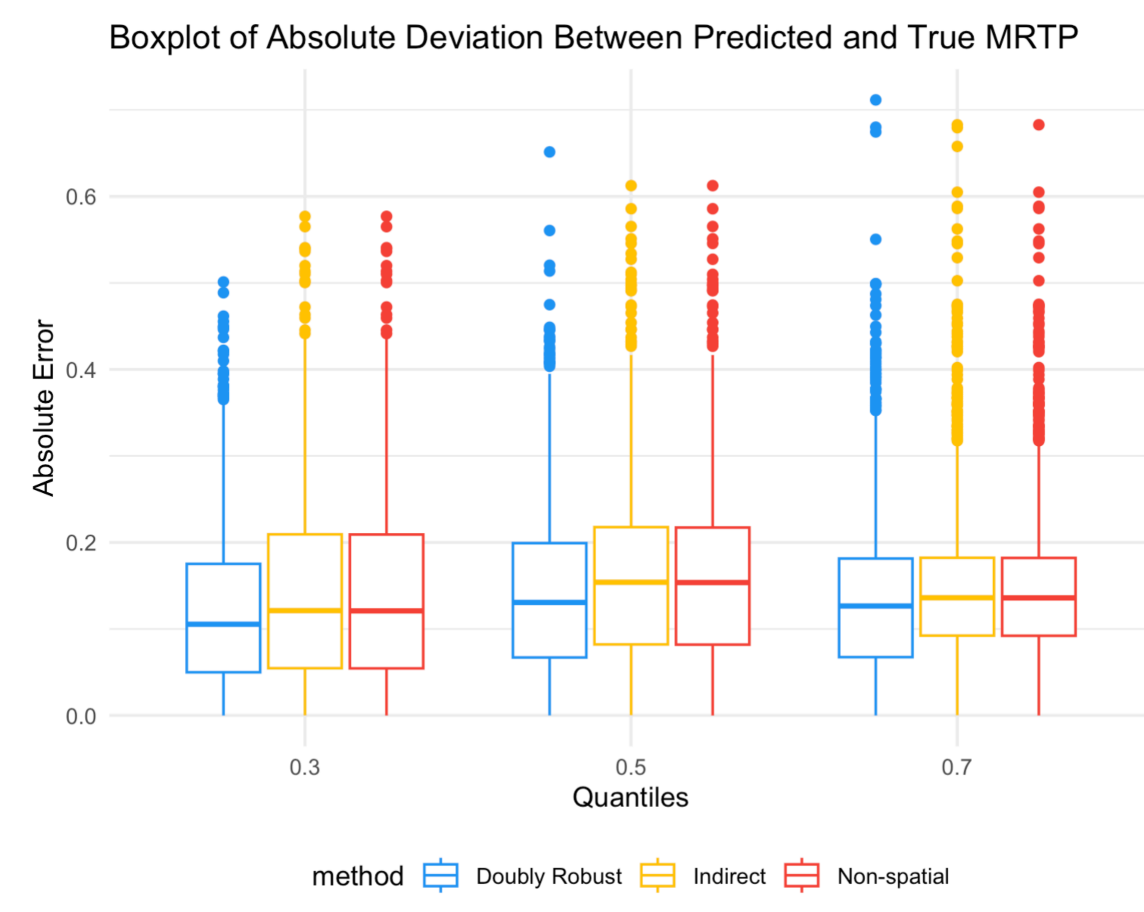
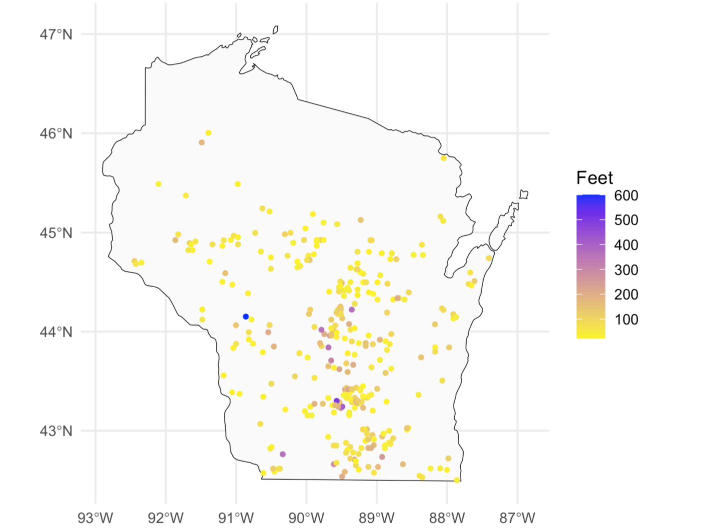


Figure: boxplot of absolute error between the predicted spMRTP and true spMRTP. Doubly robust : our doubly robust estimation. Indirect: indirect method using outcome regression. Non-spatial: indirect method that ignores the spatial dependency term $U(s)$.

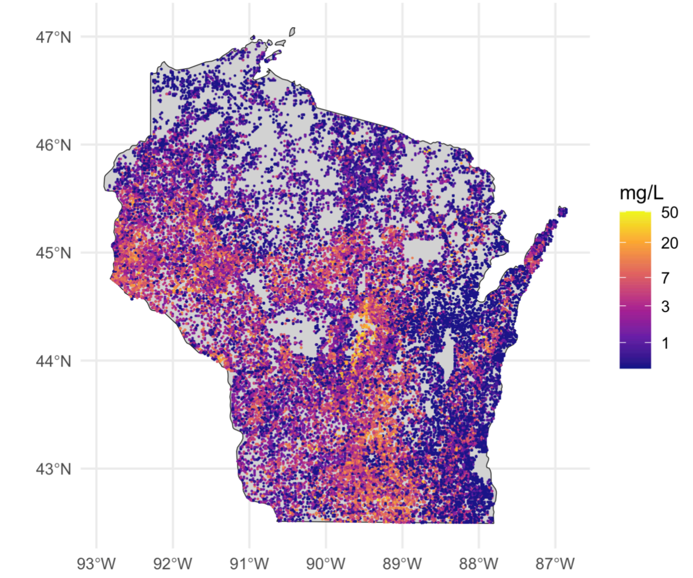
Application: Nitrate in Wisconsin Groundwater

- Nitrate remains the most widespread groundwater contaminant in Wisconsin.
- Millions of dollars[6] are spent to meet the 10mg/L public health standard.
- We aim to estimate ***minimum well depth to meet the 10mg/L health standard.***
- We use publicly available groundwater nitrate measurements and environmental variables that are hypothesized to predict nitrate contamination.

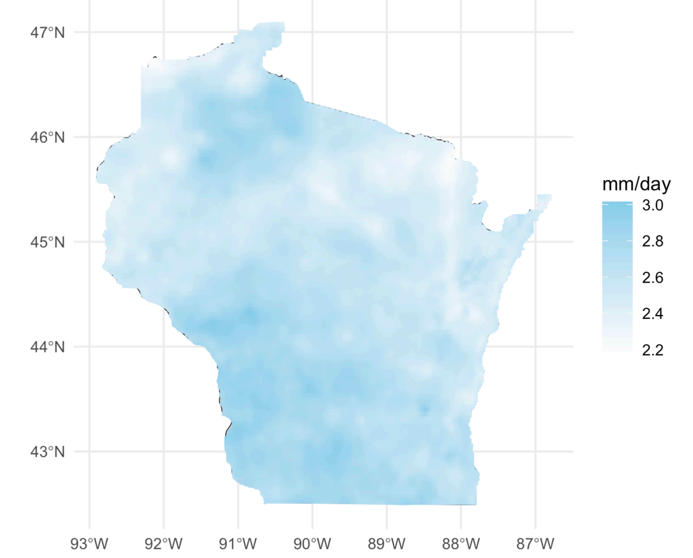
- **Required well depth** in Wisconsin.



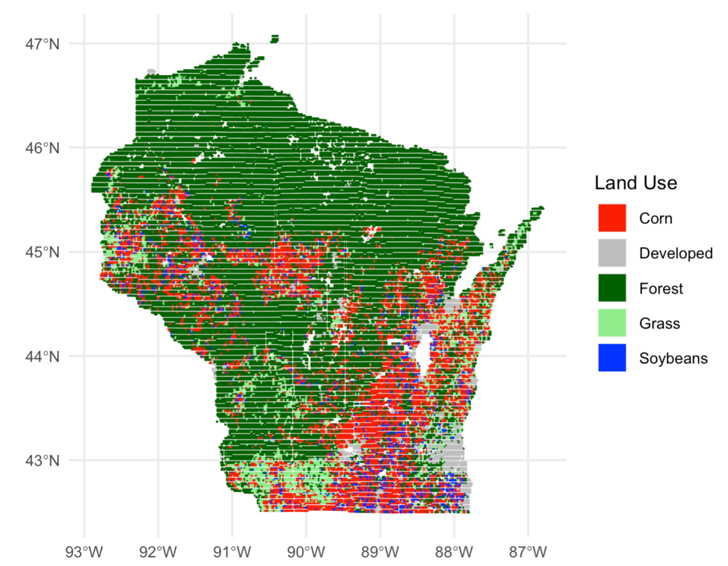
- **Nitrate measurements**[1] taken between 2014 and 2024 in Wisconsin.



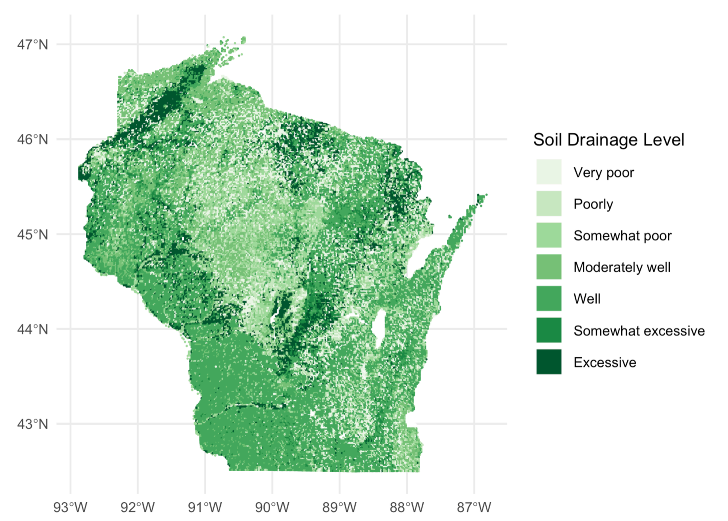
- **Daily average precipitation**[3] at each site from 2014 to 2024 (Total: 18,880 sites)



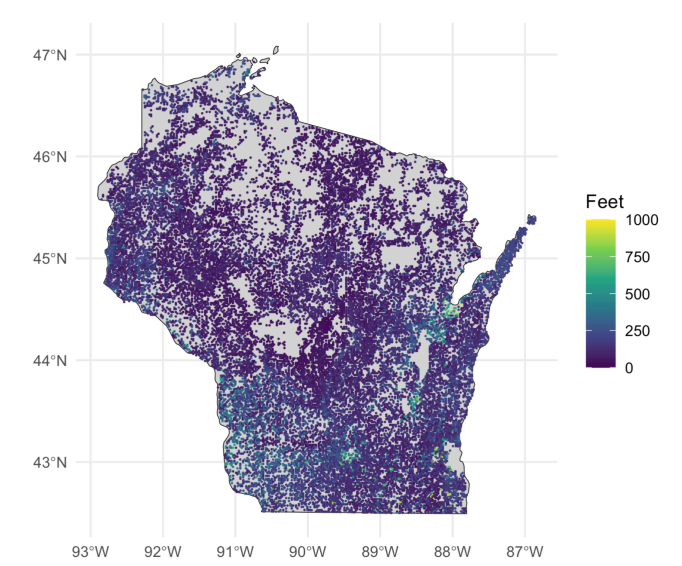
- **Land cover**[2] in Wisconsin.



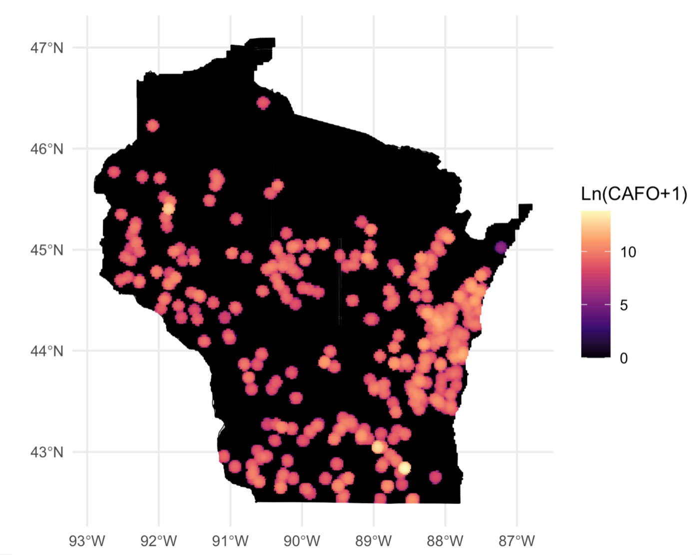
- **Seven levels** of soil drainage[4].



- **Well depth**[1] in Wisconsin.



- Concentrated animal feeding operations[5]



References: [1] Groundwater Retrieval Network (GRN). [2] Cropland Data Layer (CDL). [3] United States Geological Survey (USGS). [4] Soil Survey Geographic Database (SSURGO). [5] Wisconsin Pollutant Discharge Elimination System (WPDES). [6] Estimates of Recoverable and Non-Recoverable Manure Nutrients Based on the Census of Agriculture, USDA. [6] Wisconsin Groundwater Coordinating Council Report to the Legislature, 2024

Funding: Advanced statistical models for large datasets of pesticide and emerging contaminants in groundwater, Wisconsin Groundwater Coordinating Council, Department of Agriculture, Trade, and Consumer Protection.