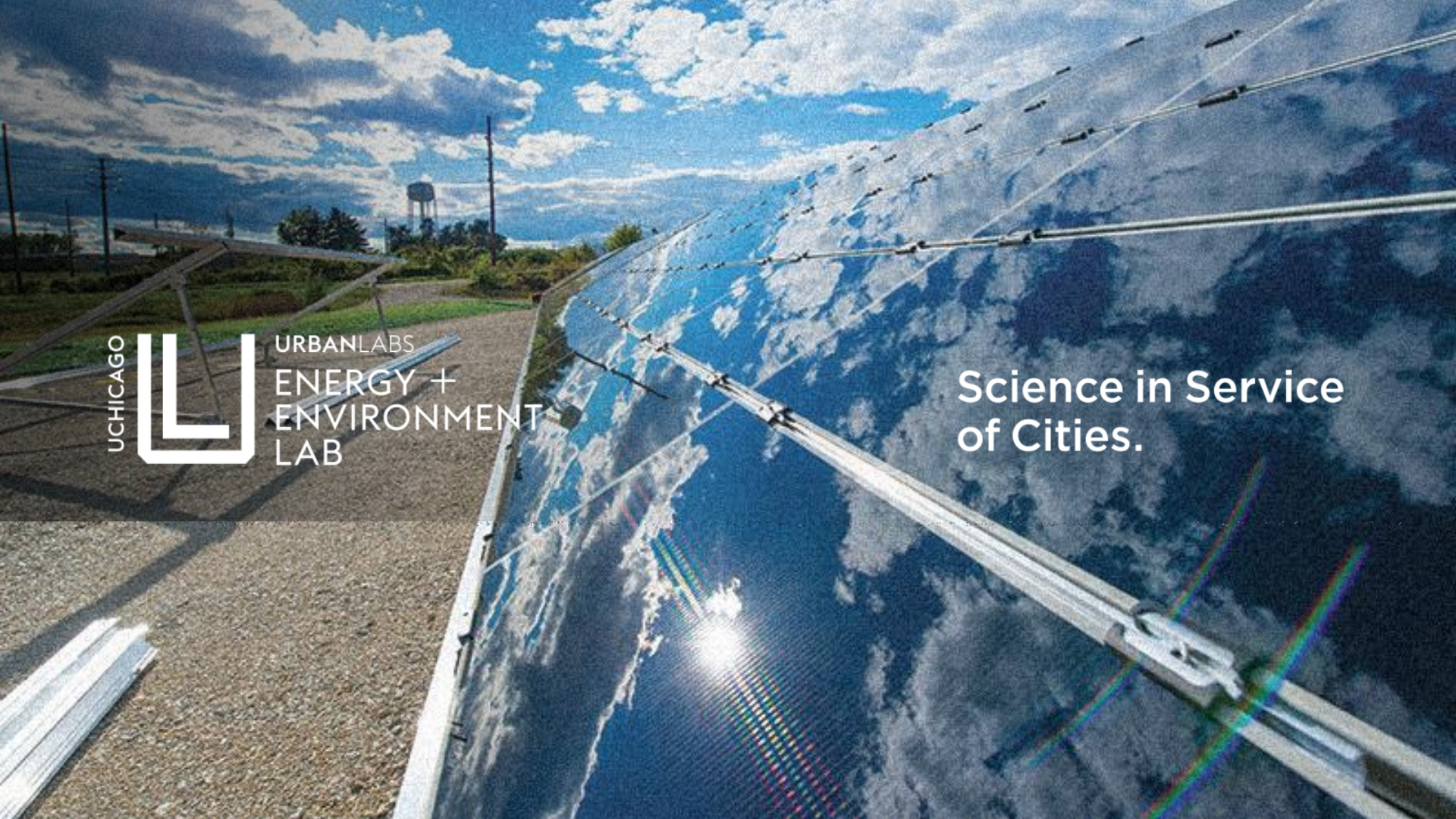

UCHICAGO
URBANLABS
ENERGY +
ENVIRONMENT
LAB
Science in Service
of Cities.

September 15th, 2020

Using Predictive Analytics to Improve Inspection Targeting

Michael Greenstone
Pritzker Faculty Director, Energy & Environment Lab
University of Chicago

UCHICAGO URBANLABS ENERGY + ENVIRONMENT LAB | Science in Service of Cities.

Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program

Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

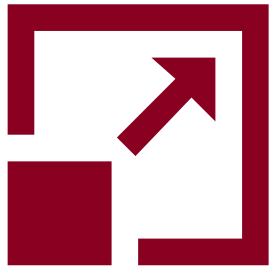
3) Improving Inspection Targeting and Efficiency in Your Program

OUR APPROACH

WE PARTNER WITH POLICYMAKERS TO:

IDENTIFY
Promising solutions to environmental challenges

TEST
The most promising policies and programs



SCALE UP
The most effective and cost-efficient policies and programs

Our Toolkit

- Leverage **economic theory** to help design efficient policies, regulations, and enforcement strategies
- Conduct **impact evaluations** to determine causal impact of programs and policies using:
  - Randomized control trials (RCTs)
  - Quasi-experimental methods
- Apply **machine learning** methodology to help regulators best allocate finite resources.

U.S. EPA and UChicago E&E Lab Partnership

- OECA and the University of Chicago Energy & Environment Lab began our **partnership in early 2015**
- **Goal:** Collaborate to develop and evaluate the effectiveness of novel techniques for supporting compliance assurance efforts
- UChicago **Research Fellow** joined EPA HQ in June 2017 (embedding is part of our model)

Selected Partners

We partner with policymakers at local, state and federal levels, including

Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program

What is Predictive Analytics?

- **Predictive analytics** is the practice of using data to determine patterns and predict future probabilities
  - Advantage of this approach is to discern and use subtle relationships in data to predict future events
- **Also known as:** Machine Learning, Artificial Intelligence, Data Science, Big Data, etc.

What is Predictive Analytics?

Private companies have been using predictive analytics to **inform decisions and improve performance for years**

- Detects potential credit card fraud
- Product recommendations at Amazon and Netflix

This revolution is coming to the public policy sphere, where government agencies are learning how to use data to **provide better services and increase their cost-effectiveness**

- New York Fire Department prioritizes building inspections by predicting which buildings are more likely to catch fire
- Chicago Department of Public Health predicts children at high risk for lead poisoning for screening

Transforming Environmental Regulation

- This wave of technological change is making the world much more efficient
- Yet these methods are not yet widely applied in environmental regulation – leaving potentially substantial efficiency gains on the table
- **Predictive analytics has the potential to transform environmental regulation by achieving greater efficiency in an increasingly resource-constrained environment**

The Challenge: How to Improve Inspection Targeting

- Across sectors, all branches of government **conduct inspections** to enforce laws that protect the environment, health, food, worker safety, etc.
- In most cases, only **a small fraction** of facilities are inspected each year due to resource constraints
- How can regulators **identify more violators** with **fewer resources?**

Proposed Solution: Identify High-Risk Facilities

Solution: Use machine learning to identify facilities **most likely to violate** RCRA regulations

Built model based on **15 years of historical data**, including thousands of variables such as:

- Facility characteristics (location, industry, etc.)
- Historical enforcement and compliance data: RCRA + other regulations

Proposed Solution: Identify High-Risk Facilities

Solution: Use machine learning to identify facilities **most likely to violate** RCRA regulations

We **evaluate model performance** by:

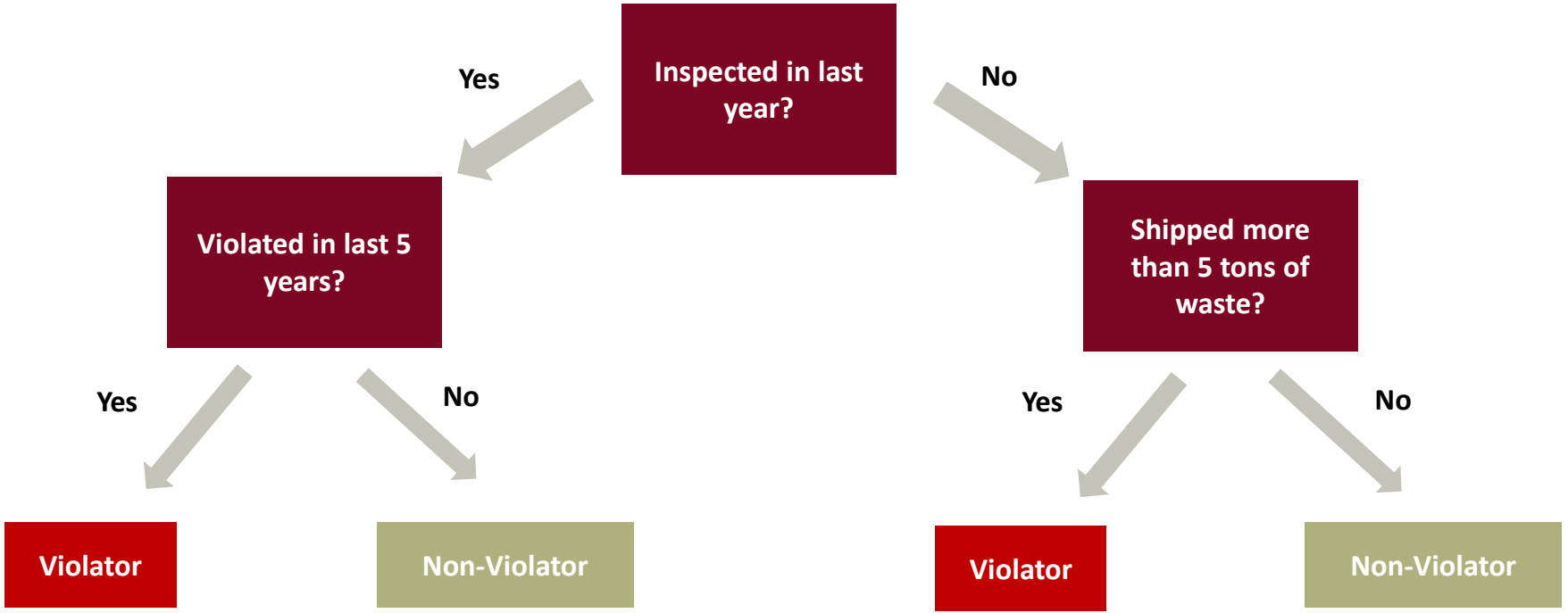
- Training the model on historical data (2000-2018)
- Predicting facility risk level in recent past (2019)
- Comparing to actual violations (in 2019)

Proposed Solution: Identify High-Risk Facilities

Solution: Use machine learning to identify facilities **most likely to violate** RCRA regulations

Model generates **risk scores** for each facility, forecasting likelihood that inspecting each facility detect a severe RCRA violation.

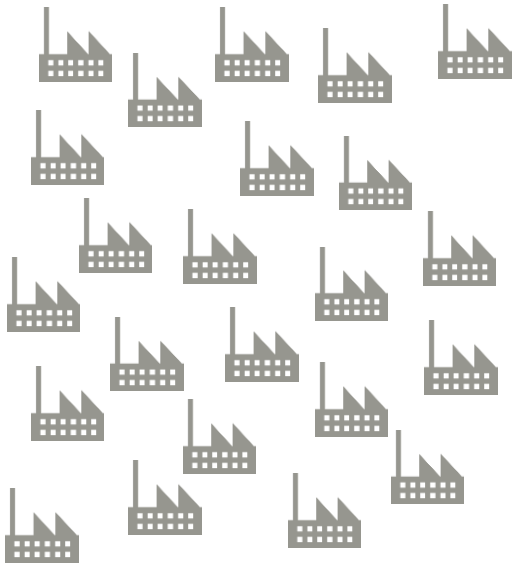
Decision Trees and Random Forest Algorithms



Field Testing the Model

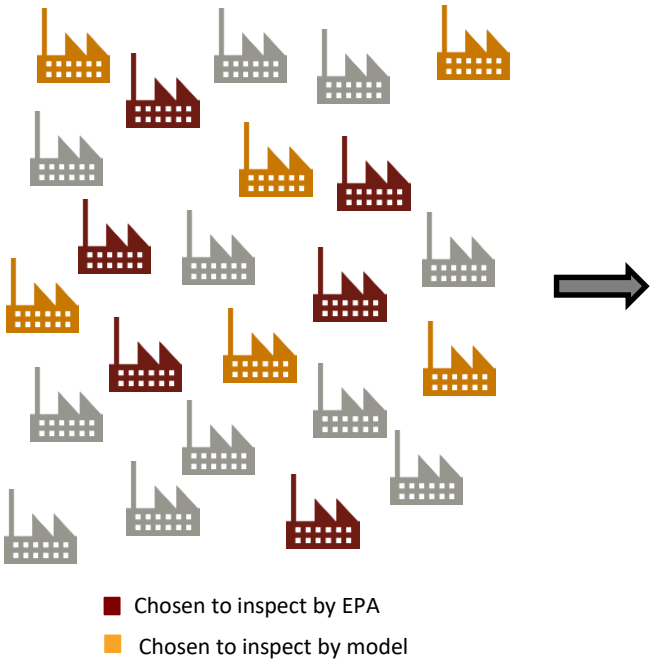
- We are working with EPA Regional offices to conduct a **rigorous field-test** of the model in the field.
- This on-the-ground testing serves to evaluate the **potential of predictive analytics to target inspections in the real world**.
- Results will benchmark the efficacy of using predictive analytics to target high-risk facilities relative to status quo practices.

Validation of the Model in Practice



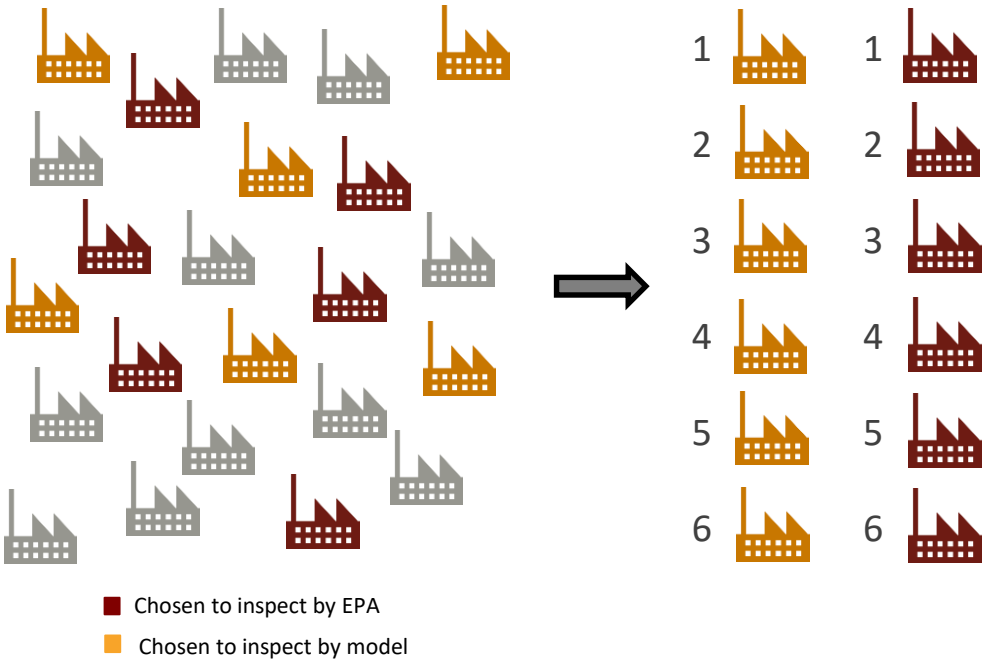
1. **Determine eligible universe of facilities with Regions and generate risk scores.**

Validation of the Model in Practice



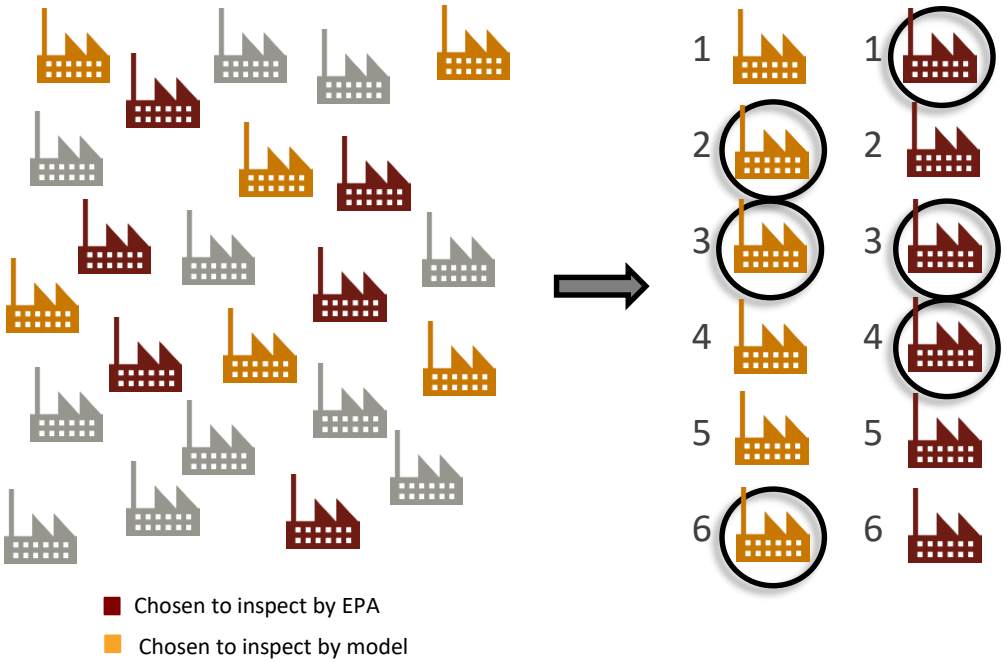
1. Determine eligible universe of facilities with Regions and generate risk scores.
2. **Model and EPA each choose top facilities to inspect.**

Validation of the Model in Practice



1. Determine eligible universe of facilities with Regions and generate risk scores.
2. **Model and EPA each choose top facilities to inspect.**

Validation of the Model in Practice



1. Determine eligible universe of facilities with Regions and generate risk scores.
2. Model and EPA each choose top facilities to inspect.
3. **Randomly sample from both lists of facilities.**

Increased Efficiency

- Results to-date suggest that the model increases the likelihood of determining a severe violation from 24.8 to 36.4 percentage points – an **increase of 47%**
- In FY17, the EPA conducted ~1,190 RCRA inspections
- If the RCRA predictive analytics model was used nationwide, the EPA:
  - Could have **found 138 additional severe violators** with the same inspection resources
- **Indicates that predictive analytics can substantially improve efficiency of environmental inspections programs**

Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program

Taking Predictive Analytics to Scale

- These promising preliminary findings have motivated the EPA to **roll out this targeting framework nationwide**.
- An **ECHO Gov targeting tool driven by predictive analytics** to be launched Fall 2020.
- ECOS/ASTSWMO member states, EPA Regions, and EPA HQ, have provided invaluable guidance on predicted outcomes that **best match regulator priorities** and interpretability of model output.
- The E&E Lab has delivered a **next-generation model** to support targeting of FY2021 inspections that will predict the likelihoods of multiple enforcement outcomes, including violation detection and formal enforcement.

Taking Predictive Analytics to Scale

- This approach is not limited to supporting inspection targeting for non-compliance related to hazardous waste
- **Scalable across programs and environmental media**, conditional on availability and quality of data
- Other possible applications include:
  - Targeting inspections of facilities regulated under the Clean Water Act, Clean Air Act, etc.
  - Identifying high-risk facilities to direct compliance assistance
- **The E&E Lab is currently developing a model to predict which facilities are most likely to be found in violation of the Clean Water Act at inspection**

What else do we do?

The UChicago Energy & Environment Lab also:

- Provides **technical assistance** to regulators and state agencies
- Conducts **impact evaluations** to determine causal impact of programs and policies using
  - Randomized control trials (RCTs) and
  - Quasi-experimental methods

Please reach out to us to discuss how we can help you design and test effective and efficient policies, regulations and enforcement strategies.

Interested in working with us?

Contact us!

Michael Greenstone
Faculty Director, UChicago Energy & Environment Lab
mgreenst@uchicago.edu

Olga Rostapshova
Executive Director, UChicago Energy & Environment Lab
oro@uchicago.edu

Thank you.