*From Space to Place: Advancing Environmental Health with Remote Earth Observations*

# Uncertainty Quantification for Satellite
# and Environmental Health Applications


Ellen M. Considine

Assistant Professor of Geography, CU Boulder

Fellow of the Cooperative Institute for Research in Environmental Sciences (CIRES)

Public Health Program Affiliate


April 28, 2026

# High-level overview of methods used to quantify uncertaint quality measurements using satellite data

I will focus on statistical and machine learning methods:

1. Is uncertainty modeled/estimated internally, or after the fact?
2. Are uncertainty estimates evaluated and/or calibrated to different conditions?
3. What metrics are available for downstream health analyses?

# Is uncertainty modeled/estimated internally, or after the fa

<u>Internally:</u>

- Bayesian statistical models
  - Work directly with probability distributions
  - Explicitly model spatial structure
- Machine learning-based
  - Predict quantiles instead of a single point estimate
  - Repeatedly perturb the model parameters to generate a distribution (“Monte Carlo dropout”)

<u>After the fact:</u>

- Model / interpolate the residuals
  - [Residuals = observations minus model predictions]
- Fit the model many times on resampled data (bootstrap)

# Is uncertainty modeled/estimated internally, or after the fa

**Internally:**

- Bayesian statistical models
  - Work directly with probability distributions
  - Explicitly model spatial structure
- Machine learning-based
  - Predict quantiles instead of a single point estimate
  - Repeatedly perturb the model parameters to generate a distribution ("Monte Carlo dropout")

Typically considered best practice

Sometimes combined with modeling the residuals

**After the fact:**

- Model / interpolate the residuals
  - [Residuals = observations minus model predictions]
- Fit the model many times on resampled data (bootstrap)

Traditionally less computationally efficient; some modern adaptations

# How are uncertainty estimates evaluated and/or calibrated?

- For evaluation: are spatial and temporal autocorrelation being accounted for?
  - e.g., spatial and/or temporal cross-validation folds
- Are prediction intervals calibrated? (ideally on held-out data)
  - e.g., does the true exposure fall within a 95% interval 95% of the time?
  - If not, can scale the interval width ("conformal prediction")
- Across conditions of interest?
  - Region / land use type
  - Time period / season / meteorology
  - Pollution concentration
  - Factors affecting satellite retrieval
  - …

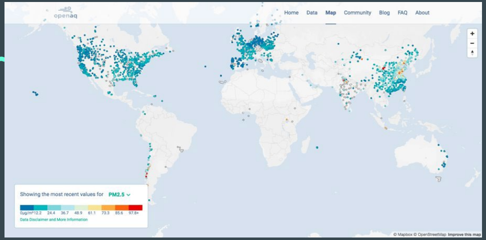
What metrics are available for downstream health analyses?

- Quantile estimates
- Interval width
- …

# Different flavor from my recent work: policy impact evaluat

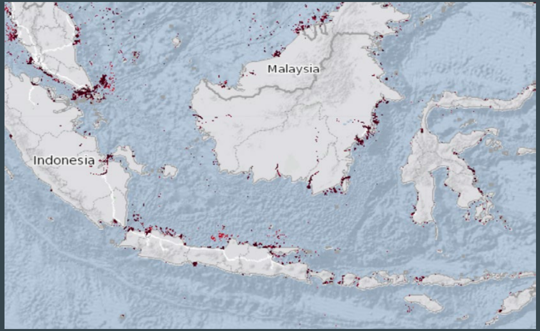
Can credibly estimate a policy effect and its uncertainty over a large region and time period

Even in regions with little to no groundlevel air quality monitoring



By pairing:

- Modern causal inference methods
- Existing satellite-derived air quality datasets
- Complementary remotely-sensed data products which add critical context

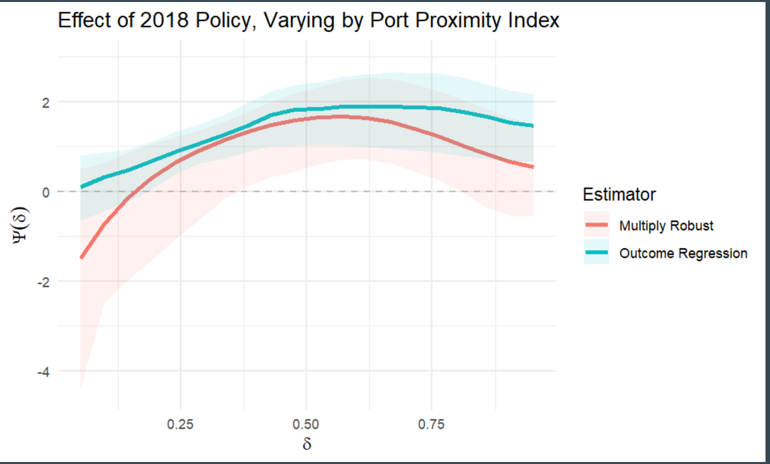



Cargo ship loitering density

Case study: estimating air quality impacts of plastic waste burning in Indonesia, before & after China stopped taking plastic waste exports

# Uncertainty quantification aspects of this work

- "Multiply robust" estimator:
  - Separately models the outcome ($PM_{2.5}$) and exposure using flexible machine learning
    - Cross-fitting
  - Targets confounding with respect to both the time period and the level of port proximity (continuous)
- Accounts for residual spatial correlation through a novel spatial weighted bootstrap



# Thank you!

Contact: ellen.considine@colorado.edu

Website: ellenconsidine.github.io

Bluesky: @emconsidine.bsky.social

Dr. Rachel Nethery
Assoc. Prof. of Biostatistics
Harvard School of Public Health

# Resources for additional exploration

- [Our 2026 preprint] “A Spatiotemporal, Quasi-experimental Causal Inference Approach to Characterize the Effects of Global Plastic Waste Export and Burning on Air Quality Using Remotely Sensed Data”
- [2026 paper] “Enhancing Estimation of Fine Particulate Matter Chemical Composition Across North America by Including Geophysical A Priori Information in Deep Learning with Uncertainty Quantification”

General:

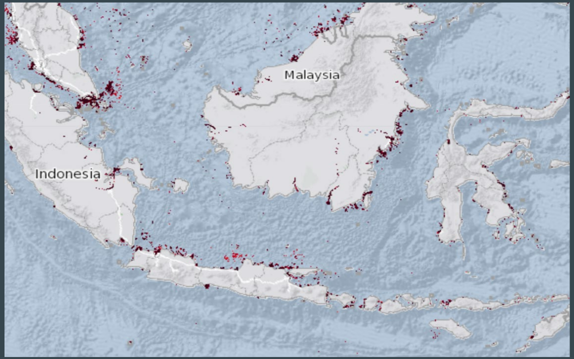
- [2022 short course slides] “Uncertainty quantification in Machine Learning” – *methods overview*
- [2026 paper] “Confidently Uncertain: Validating Satellite ECV Measurement Uncertainty Estimates”
- [2024 paper] “Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction”
- [2026 paper] “Uncertainty Quantification of Satellite-Based Essential Climate Variables Derived from Deep Learning”

# *Extra slides*

# Triangulating Air Quality Impacts o

# Plastic Waste Burning

- Open burning of plastic wastes is an urgent global health issue
- Case study: in 2018, China stopped importing plastic waste, and other countries such as Indonesia received an influx

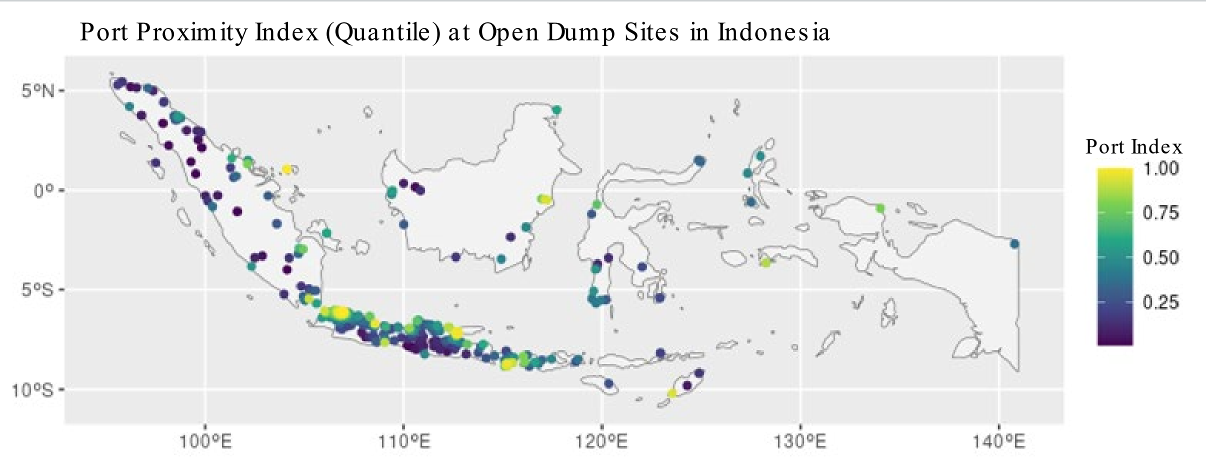



Cargo ship loitering density

Increase up to 1.68 μg/m$^3$ (95% CI = [0.72, 2.48]) at dump sites with medium-high port proximity

- Data:
  - $PM_{2.5}$ estimates (from van Donkelaar et al. 2021)
  - Meteorology from ERA5; active fire points from VIIRS
  - Sites where open dumping has been detected by Global Plastic Watch in satellite imagery
  - Cargo ship signal density (while loitering)
- Statistical method: robust extension of Diff-in-Diff with a continuous exposure
  - Uncertainty quantification accounting for residual spatial correlation
- Results: observed strong evidence that monthly $PM_{2.5}$ increased at Indonesian dump sites in the wake of China's policy change

# Motivation for methodological adaptation




We expect the long-term trend in air pollution at non-dump sites to be *different* than that at dump sites because domestic waste is increasing and nearly half of this is openly burned

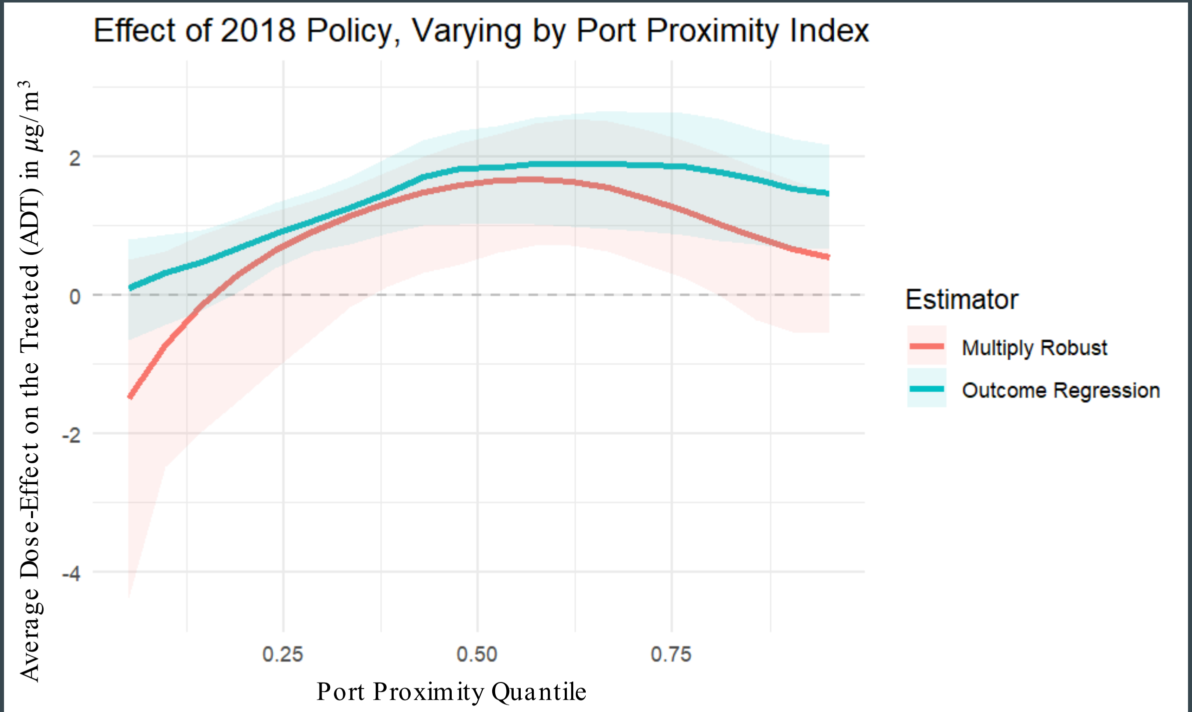
# Adapted cutting-edge quasi-experimental statistical methods

Combined two relevant strains of causal inference:

1. Estimating a heterogeneous pre-post policy effect using distance as an induced continuous exposure – Hettinger et al. (2025)
2. Using pre-intervention years as controls for post-intervention years when the policy is ubiquitous – demonstrated in several recent papers

Also extended Hettinger's weighted bootstrap approach (to obtain 95% CI) to account for residual spatial correlation

# Main results




Interpretation: We observe strong evidence that monthly $PM_{2.5}$ concentrations increased after China's ban took effect (2018-2019) compared to concentrations expected under business-as-usual (2012-2017), with increases as large as 1.68 μg/$m^3$ (95\% CI = [0.72, 2.48]) at dump sites with medium-high port proximity. Interestingly, the effect appears more modest at dump sites with very high port proximity, possibly reflecting smaller increases in dumping/burning where there is more government oversight.