Topics in spatial causal inference

Brian Reich, NC State

SRCOS, 2022, Jekyll Island

Collaborators: Can Cui, Andrew Giffin, Yawen Guan, Garritt Page, Ana Rappold, Massimo Ventrucci, Lili Wu, Shu Yang

Adjusting for unmeasured spatial confounders

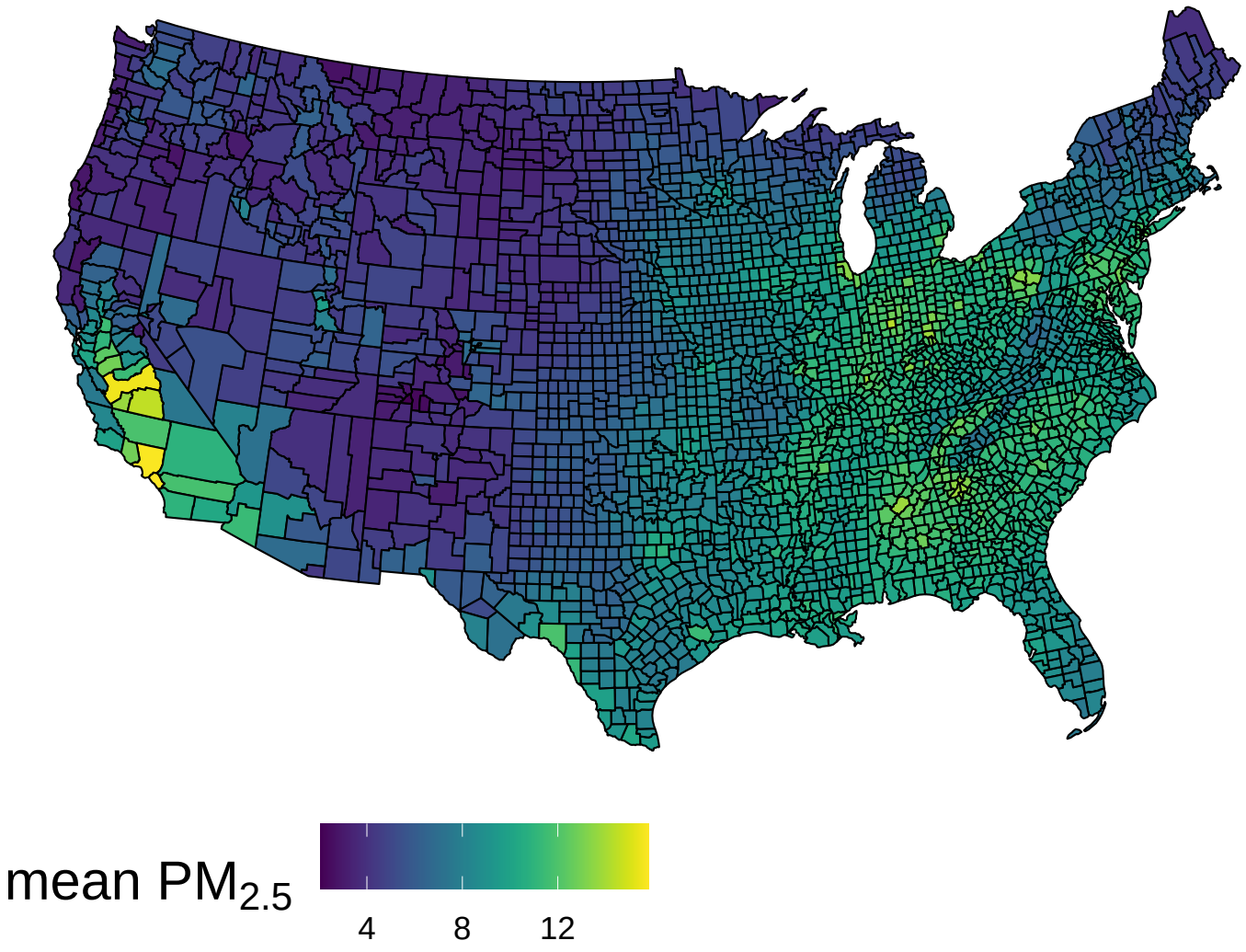
- Environmental and epidemiological data are often observational and spatially correlated
- A fundamental task is to estimate the effect of a treatment variable on a response variable
- The key assumption that there are no missing confounders is generally impossible to verify
- However, it may be possible to remove the effects of unmeasured confounders that are smooth spatial functions
- We examine several methods to adjust for missing spatial confounding variables

Motivating Example

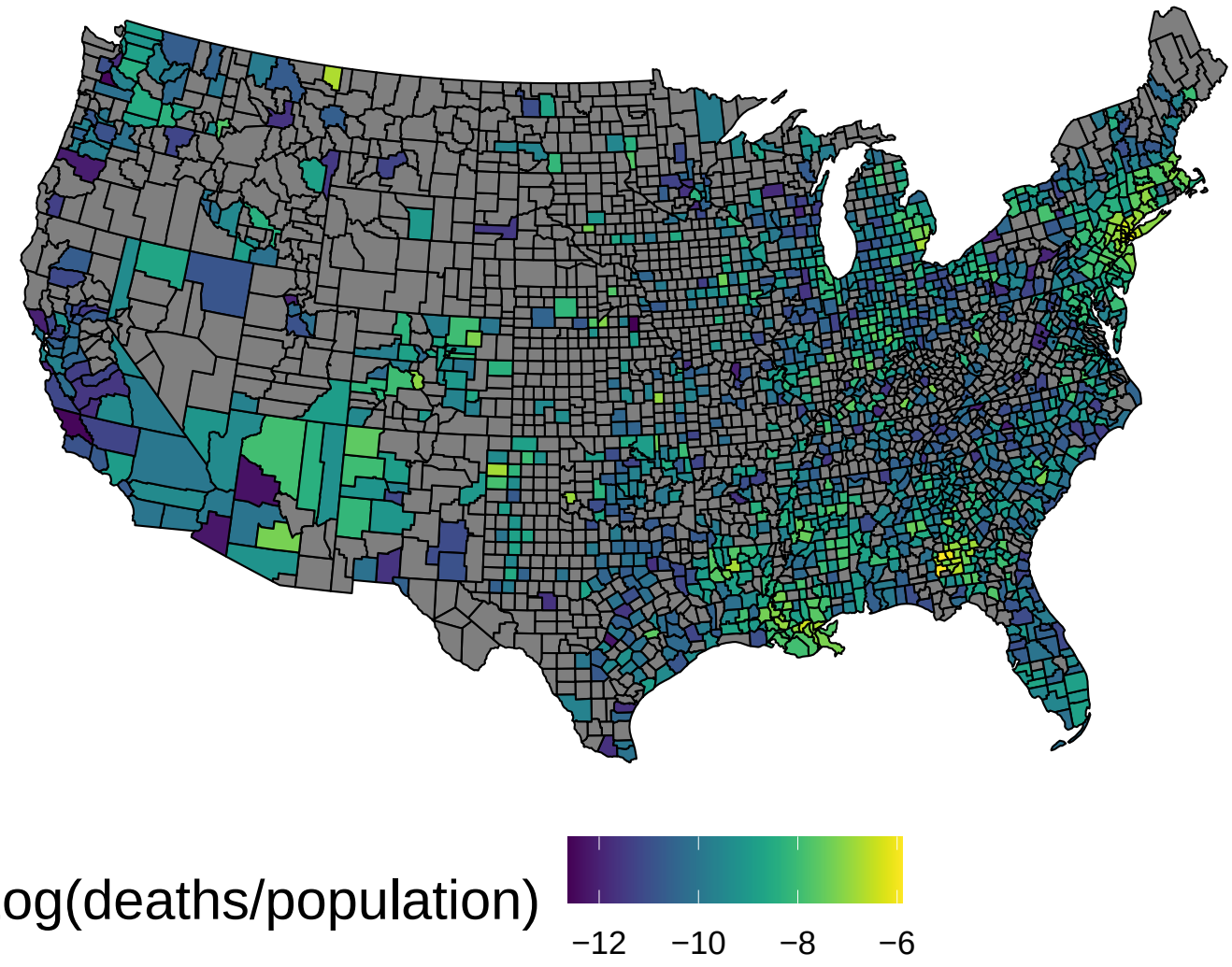
- Wu et al (2020) [1] found that an increase of 1 $\mu g/m^3$ in ambient fine particulate matter ($PM_{2.5}$) is associated with a 15% increase in the COVID-19 mortality rate.
- The response variable Y is the cumulative COVID-19 mortality counts through May 12, 2020 for US counties.
- The exposure variable X is county-level average exposure to $PM_{2.5}$ for 2000-2016

[1] Wu, Nethery, Sabath, Braun, Dominici (2020). Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis, *Science Advances.*

Exposure variable, $PM_{2.5}$ concentration



Outcome variable, log mortality rate



Other applications

- Wildland fires $\rightarrow$ increased air pollution?
- Using an app $\rightarrow$ reduced air pollution exposure?
- Fishing regulations $\rightarrow$ ecological diversity?

Standard spatial linear regression model

- The response at location $\mathbf{s} \in \mathcal{R}^2$ is $Y(\mathbf{s})$
- The treatment variable of interest is $A(\mathbf{s})$
- The model is $Y(\mathbf{s}) = \beta_0 + A(\mathbf{s})\beta_A + \theta(\mathbf{s}) + \varepsilon(\mathbf{s})$
- Our goal is to estimate β_A
- The residuals have two components: $\theta(\mathbf{s})$ is spatially correlated and $\varepsilon(\mathbf{s}) \overset{iid}{\sim} \text{Normal}(0, \tau^2)$

Standard spatial linear regression model

- The spatial term $\theta(\mathbf{s})$ is a Gaussian process with $\mathrm{E}\{\theta(\mathbf{s})\} = 0$ and $\mathrm{Var}\{\theta(\mathbf{s})\} = \sigma^2$

- The correlation decays with distance

$$\mathrm{Cor}\{\theta(\mathbf{s}_i), \theta(\mathbf{s}_j)\} = \rho(d_{ij})$$

  where d_{ij} is the distance between $\mathbf{s}_1$ and $\mathbf{s}_j$

- For example, the exponential correlation is

$$\rho(d) = \exp(-d/\phi)$$

Potential confounding variables

- Variable $X(\mathbf{s})$ is a **confounder** if it is correlated with $A(\mathbf{s})$ and $Y(\mathbf{s})$
- It is a **spatial confounder** if has spatial correlation
- It is an **missing spatial confounder** it is unknown or unobserved

Missing confounders and residual correlation

- Say the truth is $Y(\mathbf{s}) = \beta_0 + A(\mathbf{s})\beta_A + X(\mathbf{s})\beta_X + \varepsilon(\mathbf{s})$
- If X is observed, we can use non-spatial regression
- This would be fine even if X and A are correlated
- If X is unobserved, it would contribute to spatial component of the error term $X(\mathbf{s}) + \varepsilon(\mathbf{s})$
- We might fit the spatial model

$$Y(\mathbf{s}) = \beta_0 + A(\mathbf{s})\beta_A + \theta(\mathbf{s}) + \varepsilon(\mathbf{s})$$

and hope that θ accounts for X

Simulation with a missing spatial confounder

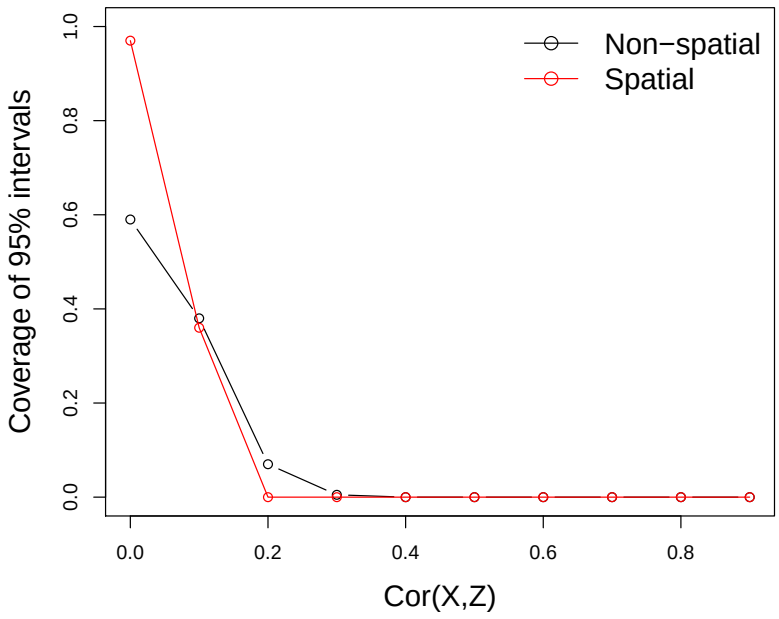
- We generate $X(\mathbf{s})$ and $A(\mathbf{s})$ as spatial processes
- The correlation between X and A ranges from 0 to 1
- The data are generated as

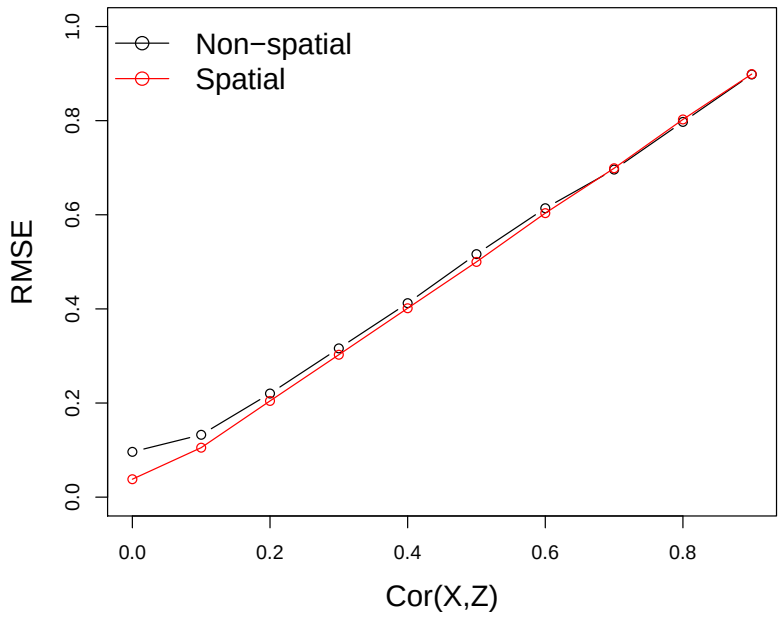
$$Y(\mathbf{s}) = A(\mathbf{s})\beta_A + X(\mathbf{s})\beta_X + \varepsilon(\mathbf{s})$$

where $\beta_A = \beta_X = 1$ and $\varepsilon(\mathbf{s}) \overset{iid}{\sim} \text{Normal}(0, 0.1^2)$

- We fit non-spatial and spatial regression models that exclude X

Coverage and root mean squared error for β_A





- Spatial regression works well when $\text{Cor}\{X(\mathbf{s}), A(\mathbf{s})\} = 0$
- However, even a small correlation causes huge problems

Possible remedies for a missing spatial confounder[1]

1. Scientific knowledge
2. Propensity scores
3. Matching
4. Mean adjustments

[1] Reich, Yang, Guan, Giffin, Miller, Rappold. (2020). A review of spatial causal inference methods for environmental and epidemiological applications. https://arxiv.org/abs/2007.02714

Adding known confounders

- Wu et al include several potential confounder variables:
  - Climate variables such as temperature
  - SES variables including income and race
  - Access to ICU beds
- Call the known confounders $W_1(\mathbf{s}), ..., W_p(\mathbf{s})$
- The fitted model is

$$Y(\mathbf{s}) = \beta_0 + \sum_{j=1}^{p} W_j(\mathbf{s})\alpha_j + A(\mathbf{s})\beta_A + \theta(\mathbf{s}) + \varepsilon(\mathbf{s})$$

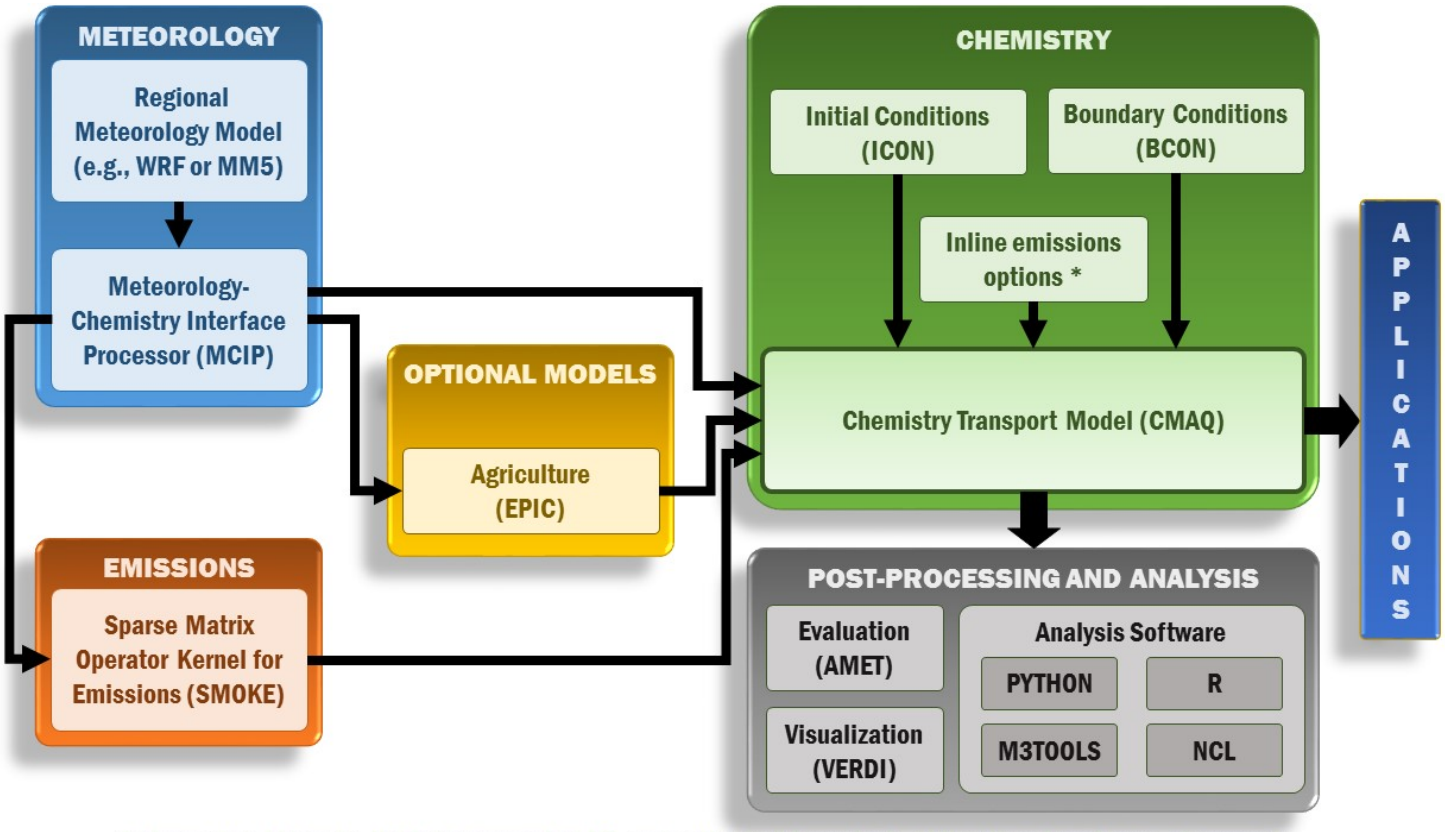
- A logical extension to environmental application is to replace the $W_j(\mathbf{s})$ with numerical model output

EXAMPLE: Adding mathematical/process models

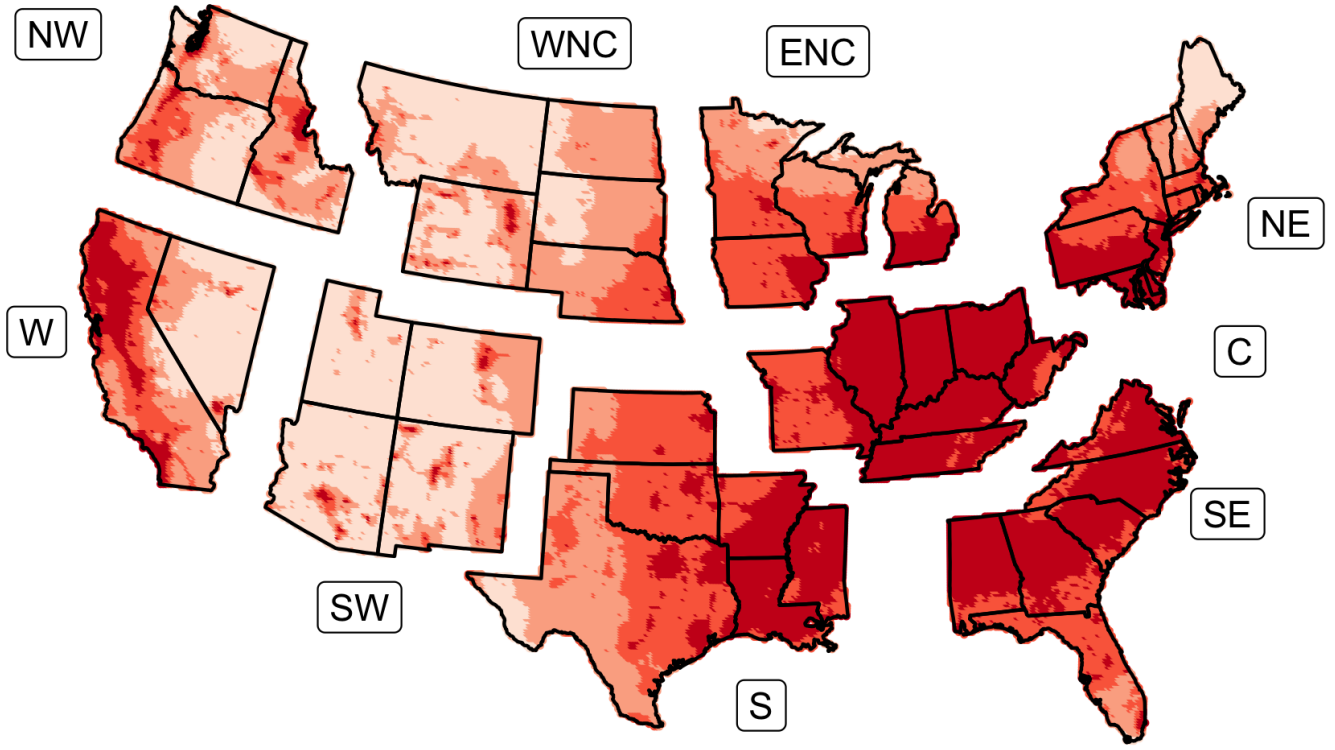
- Fine particular matter ($PM_{2.5}$) is a criteria pollutant monitored by the EPA to protect human health
- In the US, emissions of $PM_{2.5}$ from most sources are steadily declining except for forest fire smoke
- How much $PM_{2.5}$ and health burden is causally-attributed to wildfires?
- This is difficult to answer directly because only total $PM_{2.5}$ (background + fire) can be measured
- We [2] combine numerical models (CMAQ), observational data and causal inference

[2] Larsen et al. A spatial causal analysis of wildland fire-contributed PM2.5 using numerical model output. *Annals of Applied Statistics*.

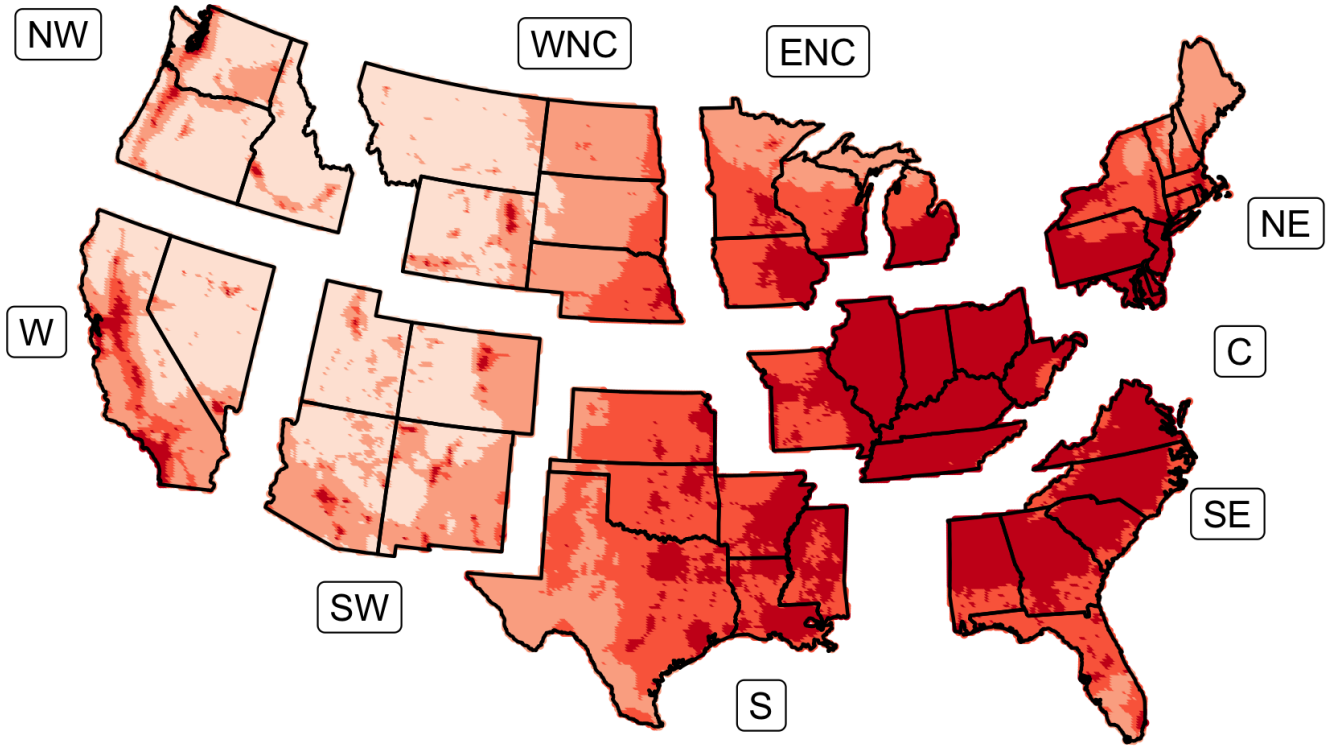
CMAQ: The Community Multiscale Air Quality Modeling System



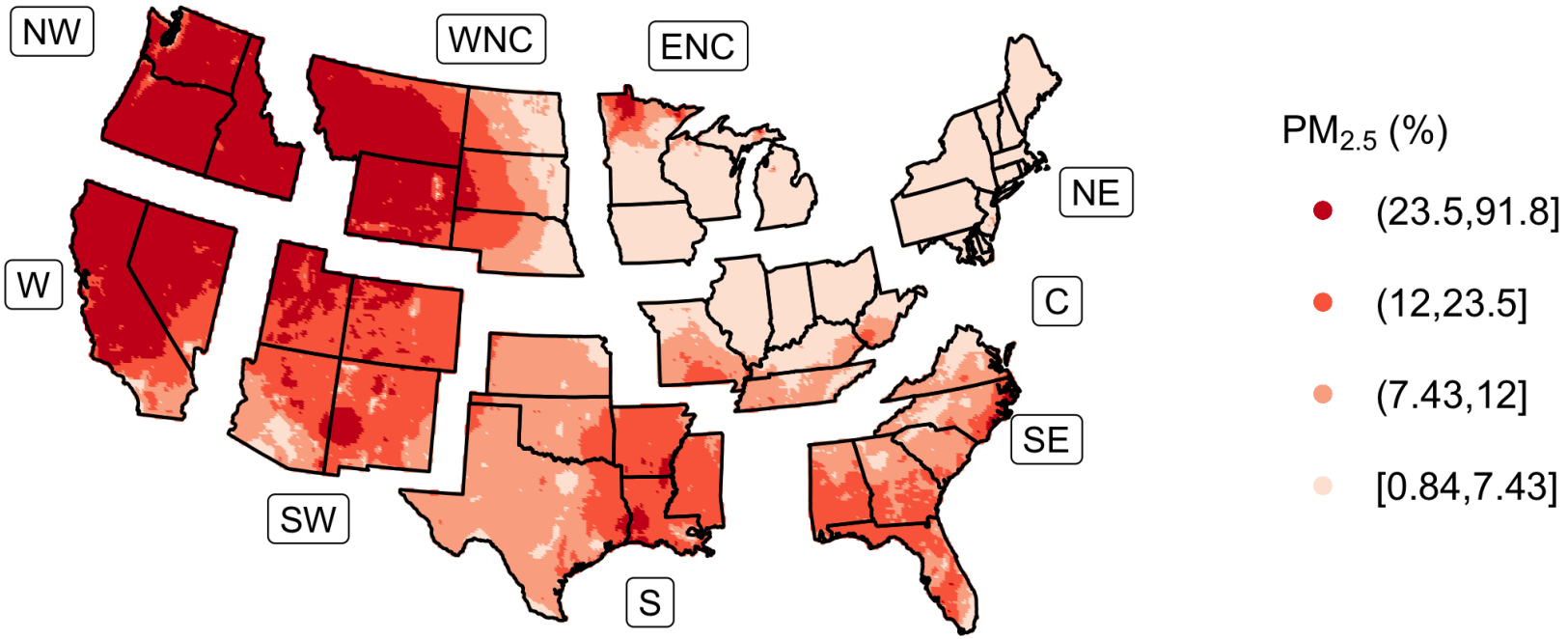
Annual average CMAQ (12km × 12km)



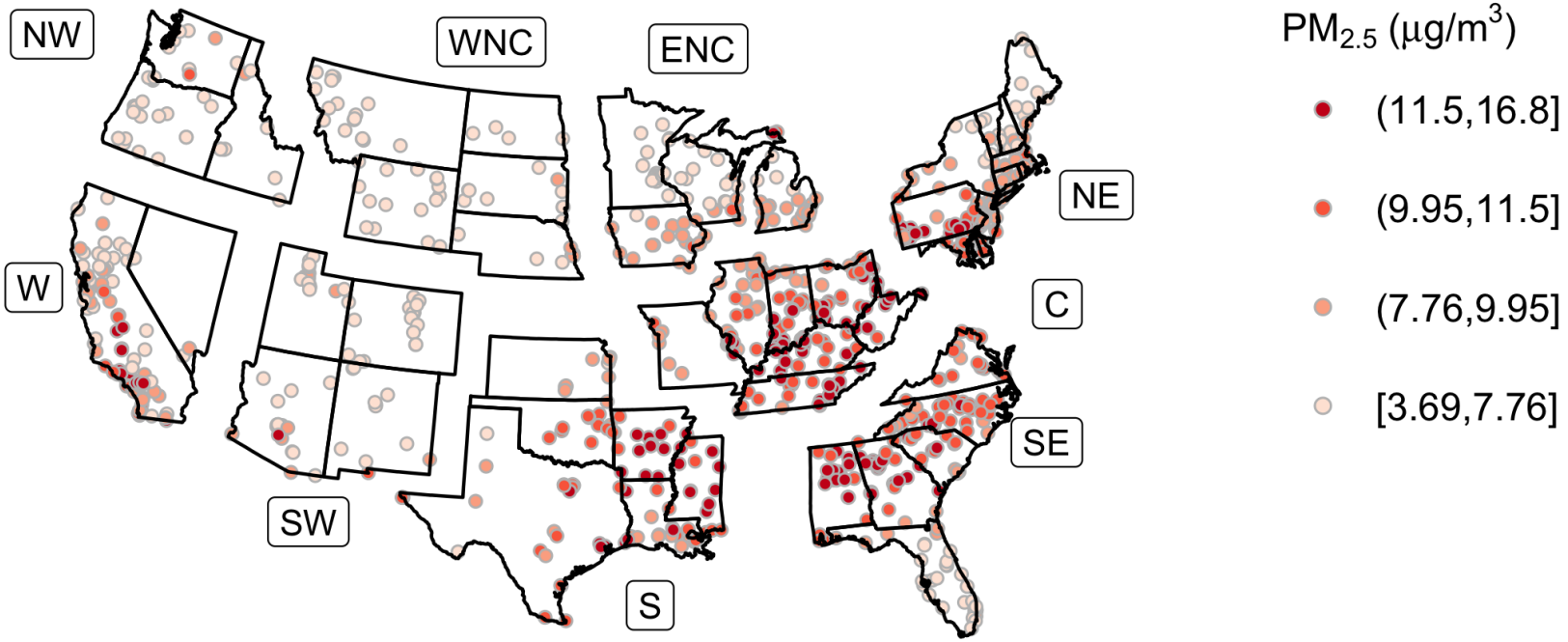
CMAQ run without fires



Difference between the runs (as % of total)

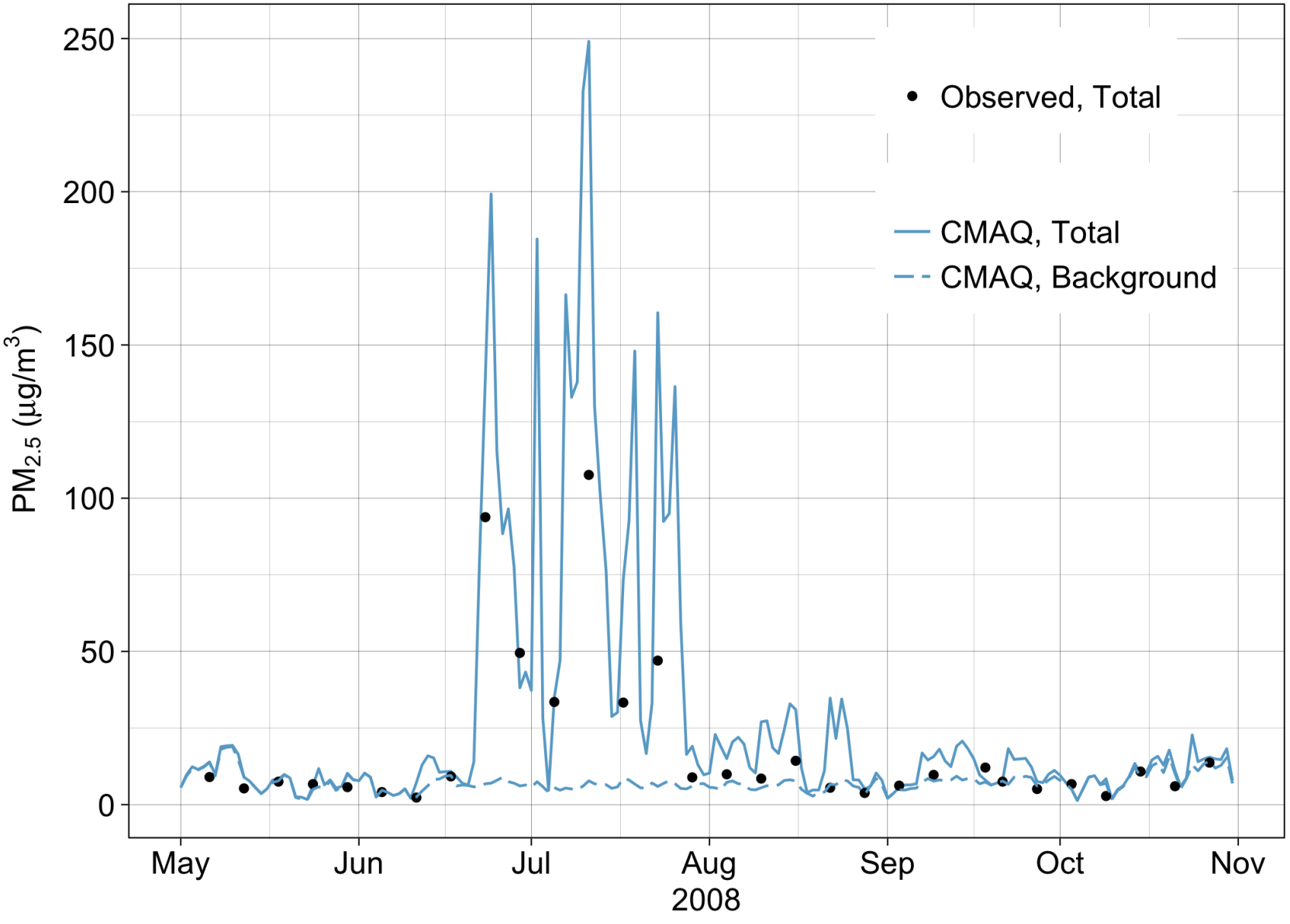


EPA Monitoring Stations (background + fire)



$PM_{2.5}$ is measured every 3-6 days; this is the 2008-2012 average

Time series plot for one site in Northern CA



Data sources and potential outcomes notation

- Monitor data at location $\mathbf{s}$ and day t: $Y_t(\mathbf{s})$
- CMAQ run without fires: $\hat{\theta}_t(\mathbf{s})$
- CMAQ run with fires: $\hat{\theta}_t(\mathbf{s}) + \hat{\delta}_t(\mathbf{s})$
- The "treatment" is the regime
  - $A = 0$: world without wildland fires
  - $A = 1$: current world with wildland fires
- $Y_t(\mathbf{s}, 0)$ and $Y_t(\mathbf{s}, 1)$ are the potential $PM_{2.5}$ outcomes
- Model: $Y_t(\mathbf{s}, 0) = \theta_t(\mathbf{s})$ and $Y_t(\mathbf{s}, 1) = \theta_t(\mathbf{s}) + \delta_t(\mathbf{s})$
- θ_t and δ_t are stochastic processes

Potential outcomes framework

- The causal effect is

$$\Delta(\mathbf{s}) = \mathrm{E}[Y_t(\mathbf{s}, 1) - Y_t(\mathbf{s}, 0)]$$

  where the average is over the distribution of fires over the entire spatial domain

- **Challenge**: we never observe data under $A = 0$

- To address this we use
  - CMAQ output
  - Causal assumptions

Assumptions

We assume there exist

- $C_t(\mathbf{s}) \in \{0, 1\}$ where $\mathbf{s}$ is affected by smoke iff $C_t(\mathbf{s}) = 1$
- Bias-correction functions B_0 and B_1

so that the following assumptions hold:

(A1) Consistency: $Y_t(\mathbf{s}) = Y_t[\mathbf{s}, C_t(\mathbf{s})]$

(A2) No unmeasured confounders given model output:

$$\theta_t(\mathbf{s}) = B_0[\hat{\theta}_t(\mathbf{s})] + e_{1t}(\mathbf{s}) \text{ and } \delta_t(\mathbf{s}) = B_1[\hat{\delta}_t(\mathbf{s})] + e_{2t}(\mathbf{s}),$$

where $\mathbf{e}_t(\mathbf{s}) = [e_{1t}, e_{2t}]$ is independent of A and $\mathbf{C}$

Assumptions

Are these assumptions reasonable?

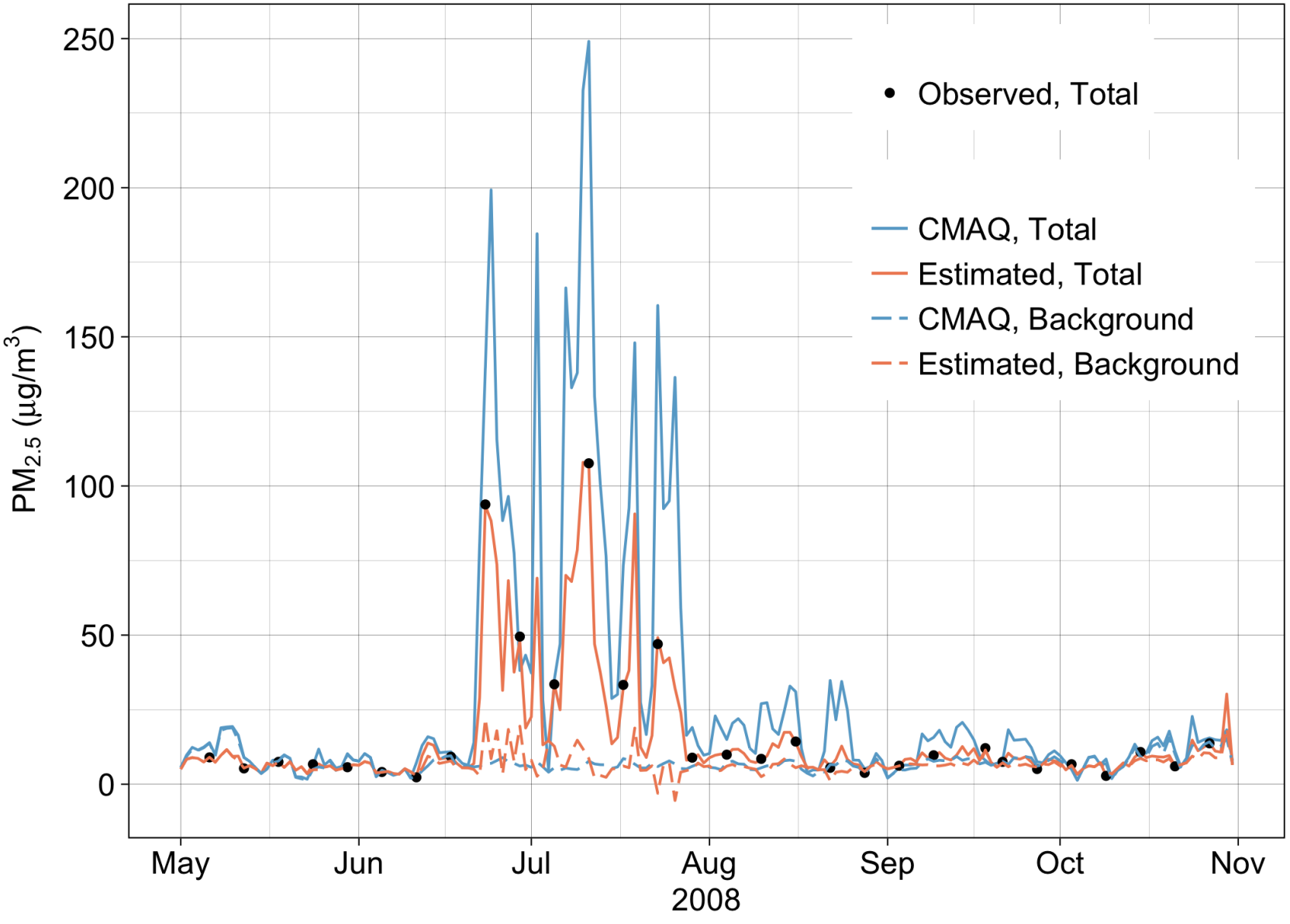
- (A1) assumes that we have some observations we are sure are not affected by fire smoke ... *probably OK?*

- (A2) assumes that the CMAQ modelers have included the important drivers of fine particulate matter .... *maybe OK? Have we accounted for all feedback loops?*

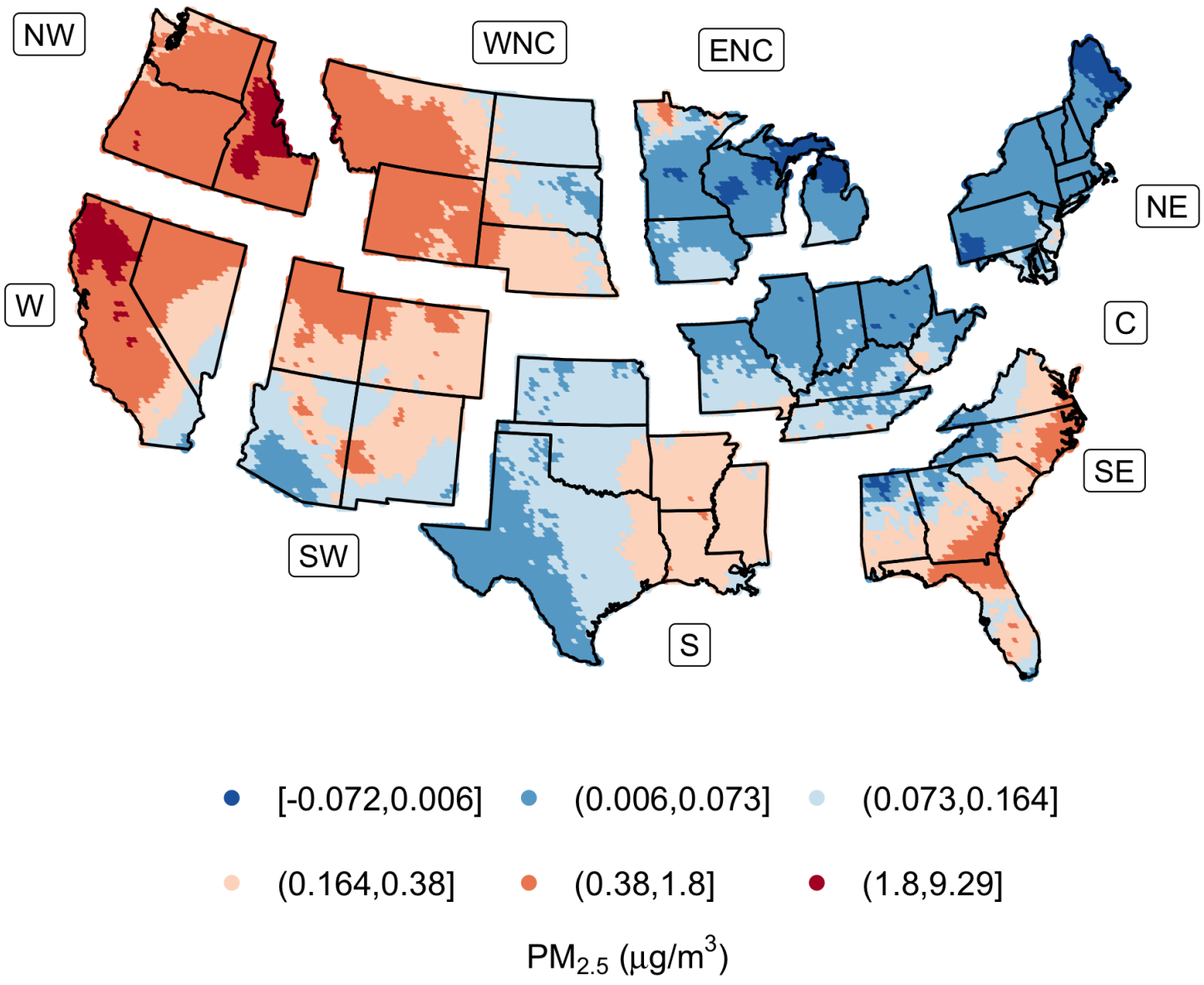
Causal interpretation

- Theory: We prove that under these assumptions the estimate effect has a causal interpretation
- We specify parametric models for the bias correction functions B_0 and B_1 and the spatial process $\mathbf{e}_t(\mathbf{s})$
- Theory: We show that all parameters, including the correlation between counterfactuals, are identifiable
- This serves as a basis for using a Bayesian approach to estimating $\Delta(\mathbf{s})$ and quantify its uncertainty

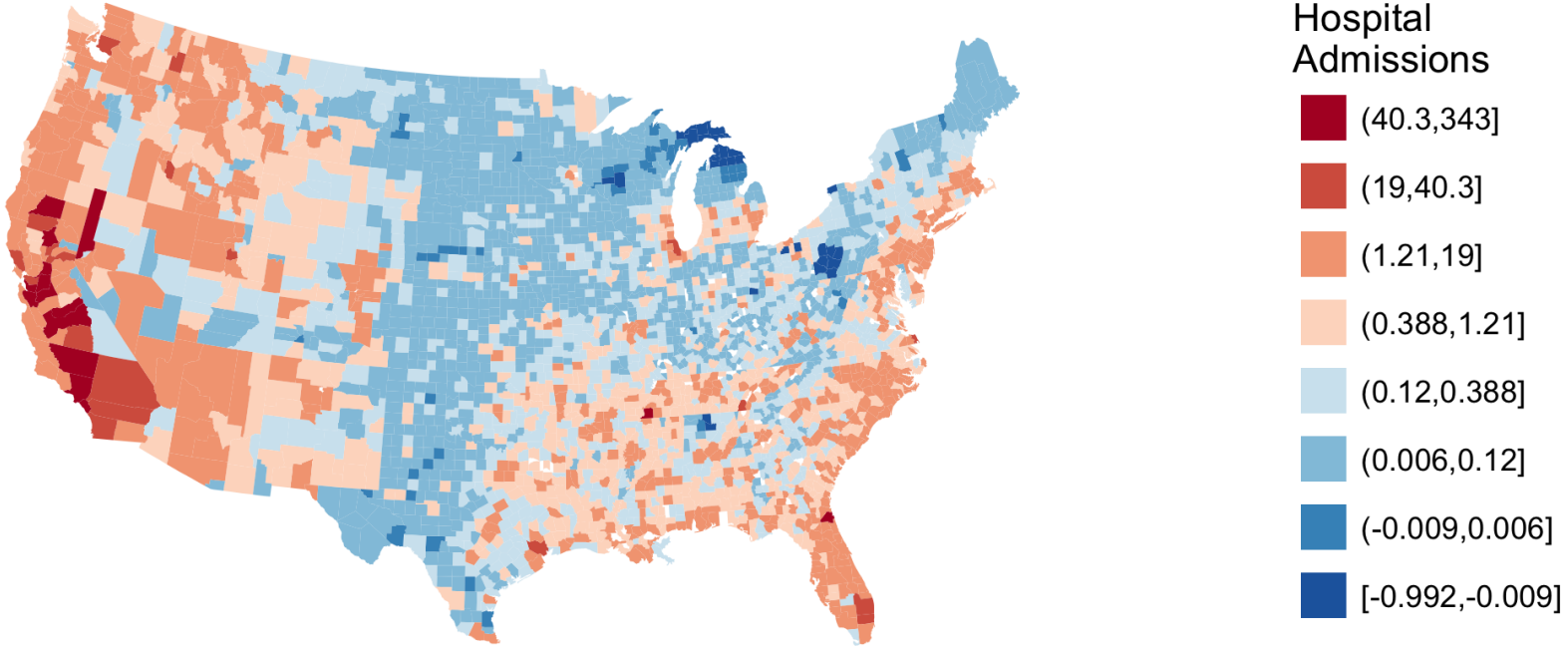
Data and estimates for one site in CA



Causal estimate, $\Delta(\mathbf{s})$, posterior mean



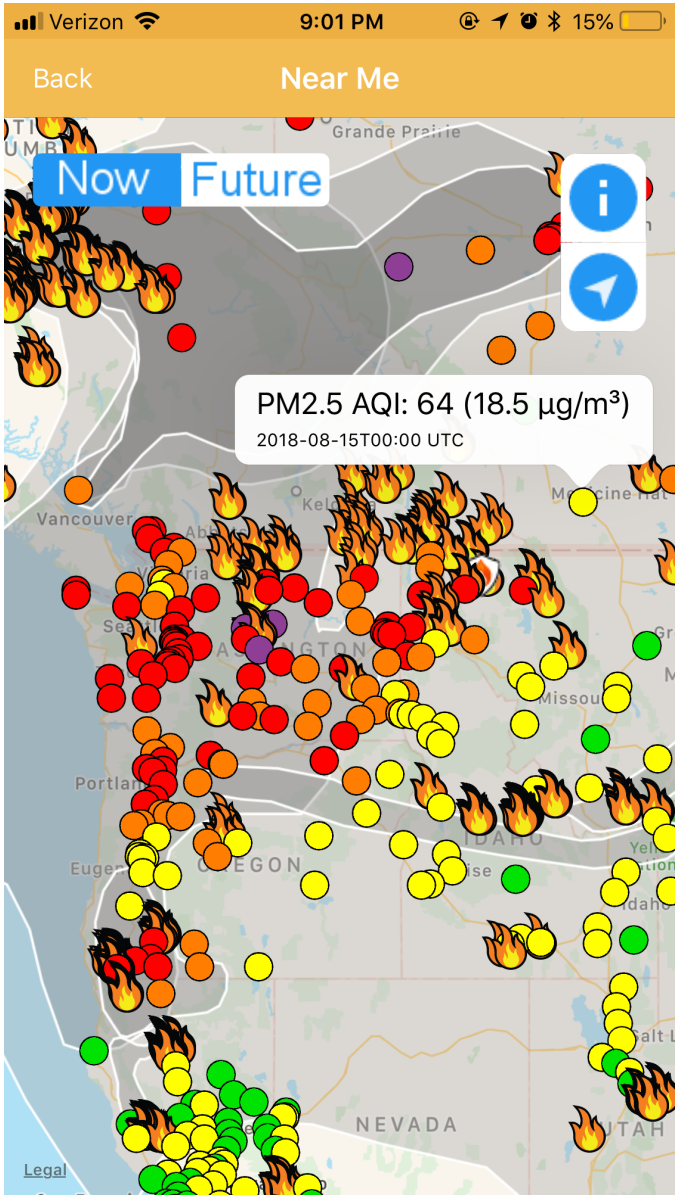
Cumulative health burden by county



EXAMPLE: Propensity scores

- Smoke Sense is a cell phone app design by the US EPA to educate and engage citizens
- The app provides users current and forecasted air quality and alerts of fires
- Users record their smoke observations, health symptoms and preventative actions
- Users also play educations games, earn badges, etc.

The smoke Sense app



Smoke Sense data and notation

- **Response**: Y_{it} is the number of self-reported symptoms (asthma attack, chest pain, etc.) by user i in week t
- **Treatment**: A_{it} is a binary indicator that user i took preventive measures (stay indoors, use mask, etc.)
- **Covariates**: baseline variables (age, gender, etc.) and time-varying variables (smoke exposure, visibility, etc.)
- We use $n = 1076$ users from 2019 and the number of observations ranges from 1-15
- Our objective is determine if using the app leads to a reduction in symptoms

Challenges

App-based platforms provide unprecedented opportunities to reach users, but pose challenges:

- **Self-selection**: Protective behaviors were left to the user and may depend on their characteristics
- **Informative missingness**: Participants are more likely to self-report when they experience smoke or symptoms
- **Spatial variation**: The causal effect of treatment may vary over the study's large and diverse domain

Potential outcomes (PO)

- For simplicity, we[3] drop the user subscript (i) and assume regular observation times $t \in \{1, 2, ...\}$
- History is denoted by overbar, so $\bar{A}_m$ and $\bar{X}_m$ are the set of values before time m
- POs are denoted by $\bar{a}_m$ is the set of treatments before time m and a_m is the set of treatment taken after time $m < t$
- The PO at time t is $Y_t(\bar{a}_m, a_m)$
- Our goal is to estimate the causal effect which is the mean difference between $Y_t(\bar{a}_m, a_m)$ and $Y_t(\bar{a}_m, 0)$
- We do this within the structured nested mean model of Robins (1992)

[3] Wu et al, https://arxiv.org/abs/2005.12017

Structured nested mean model

- The causal effect γ as a function of parameters $\boldsymbol{\theta}$ is

  $$\mathrm{E}\left\{Y_t(\bar{a}_m, a_m) - Y_t(\bar{a}_m, 0) | \bar{a}_m, \bar{x}_m\right\} = \gamma_{tm}(\bar{a}_m, \bar{x}_m; \boldsymbol{\theta}) = \gamma_{tm}(\boldsymbol{\theta})$$

- An example with covariate and lagged treatment effects is

  $$\gamma_{tm}(\boldsymbol{\theta}) = x_m \boldsymbol{\beta} + \psi \exp\left\{-\frac{(t-m-\mu)^2}{2\sigma^2}\right\} a_m$$

  and $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \psi, \mu, \sigma\}$

- The treatment at time $m = t - \mu$ has the most impact

- The magnitude of treatment at lag μ is ψ

Estimation

- Sequential randomization assumption: treatment at time m is a function only of the propensity score

$$e_m = \text{Prob}(A_m = 1 | \bar{A}_{m-1}, \bar{X}_m, \bar{Y}_{m-1})$$

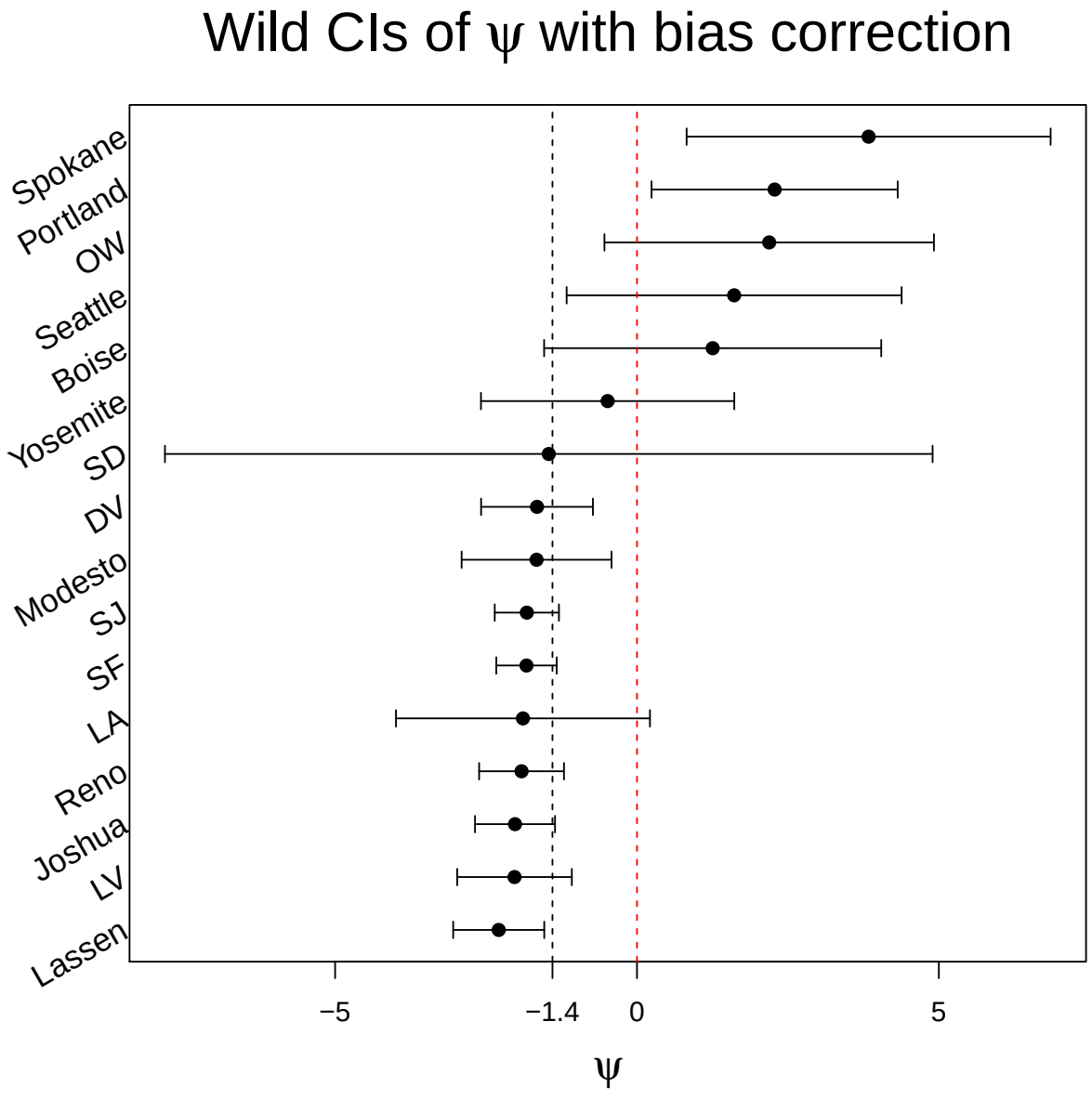
- To estimate the lag m effect, we must remove effects for time $l \in \{m, ..., t-1\}$
- Let $R_{tm}(\theta) = Y_t - \sum_{l=m}^{t-1} \gamma_{tl}(\theta)$ and $\mu_{tm}(\theta) = \text{E}(R_{tm})$
- The weighted (by w_{tm}) estimating equation is

$$G(\theta) = \sum_{t=1}^{n_t} \sum_{m=1}^{t-1} w_{tm} \{R_{tm}(\theta) - \mu_{tm}(\theta)\} (A_t - e_t)$$

Spatially-varying effects

- We extend this structured nested mean model to allow for spatially-varying θ
- This could lead to tailoring the app to individual environments
- We use geographically weighted (w_{tm})local polynomial estimation
- The local effect at $\mathbf{s}^*$ is taken to be polynomial in $\mathbf{s}$ near $\mathbf{s}^*$
- Observations are kernel weighted by distance $||\mathbf{s} - \mathbf{s}^*||$
- Double robustness: We prove consistency if either the potential-outcome or mean models are correct

Smoke Sense – Local estimate of $\psi(\mathbf{s})$



EXAMPLE: Spatial matching

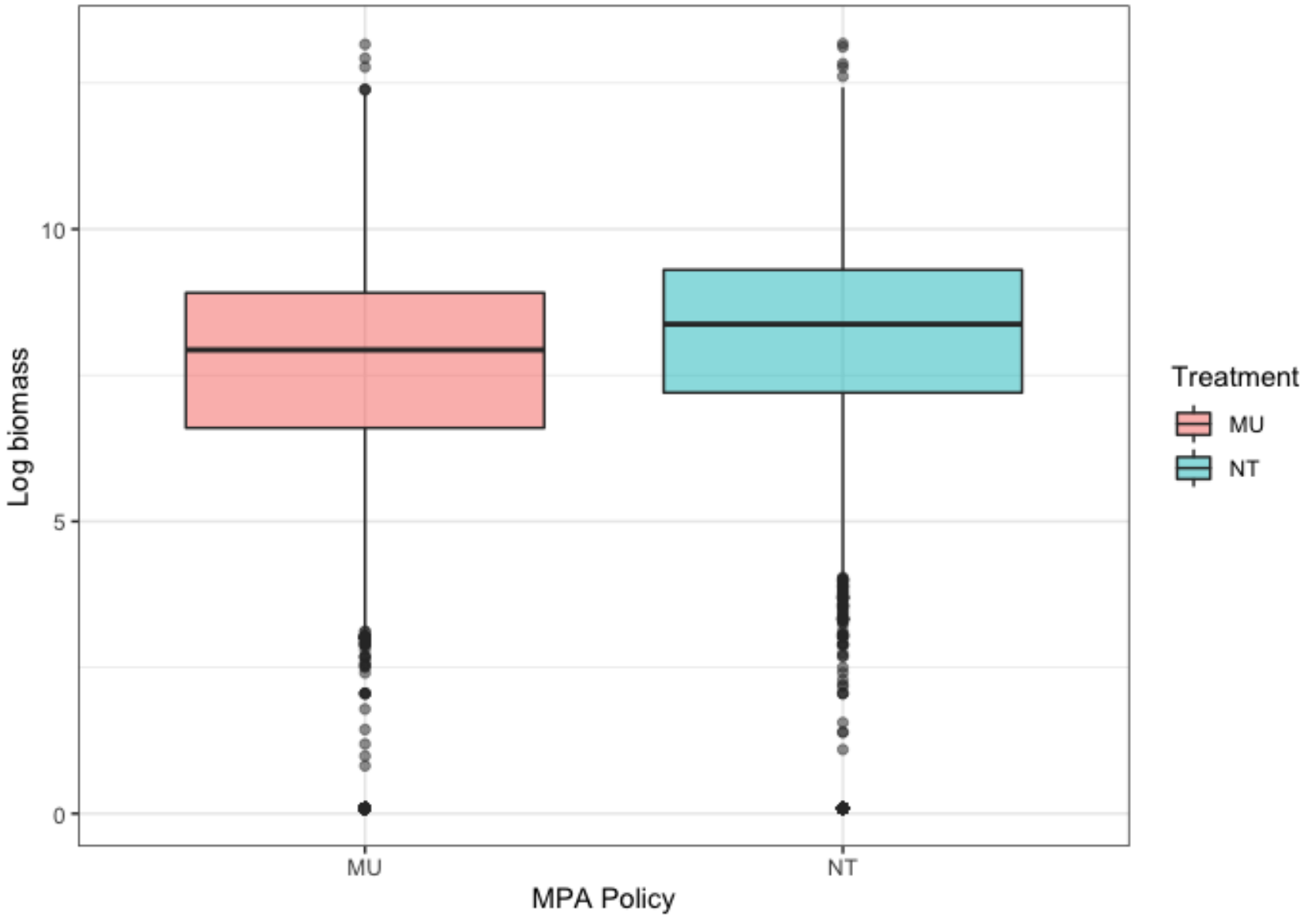
- Commercial fishing is highly regulated for conservation
- There is variation in the type and degree of regulation
- Using data from global Gill et al [4] we compare two broad classes of regulations: no take (NT) and multi-use (MU)
- Our objective is to compare the relative effectiveness of these two policies

[4] Gill et al. (2017). Capacity shortfalls hinder the performance of marine protected areas globally. *Nature*, **543**, 665–669

Data and notation

- Response: Y_i is the log biomass at site $i = 1, ..., n = 9987$
- Treatment: $A_i = 1$ for NT and $A_i = 0$ for MU
- Covariates: $\mathbf{X}_i$ are known covariates
- Dependence structure: the n observations are nested in 215 Marine Protection Areas (MPAs)

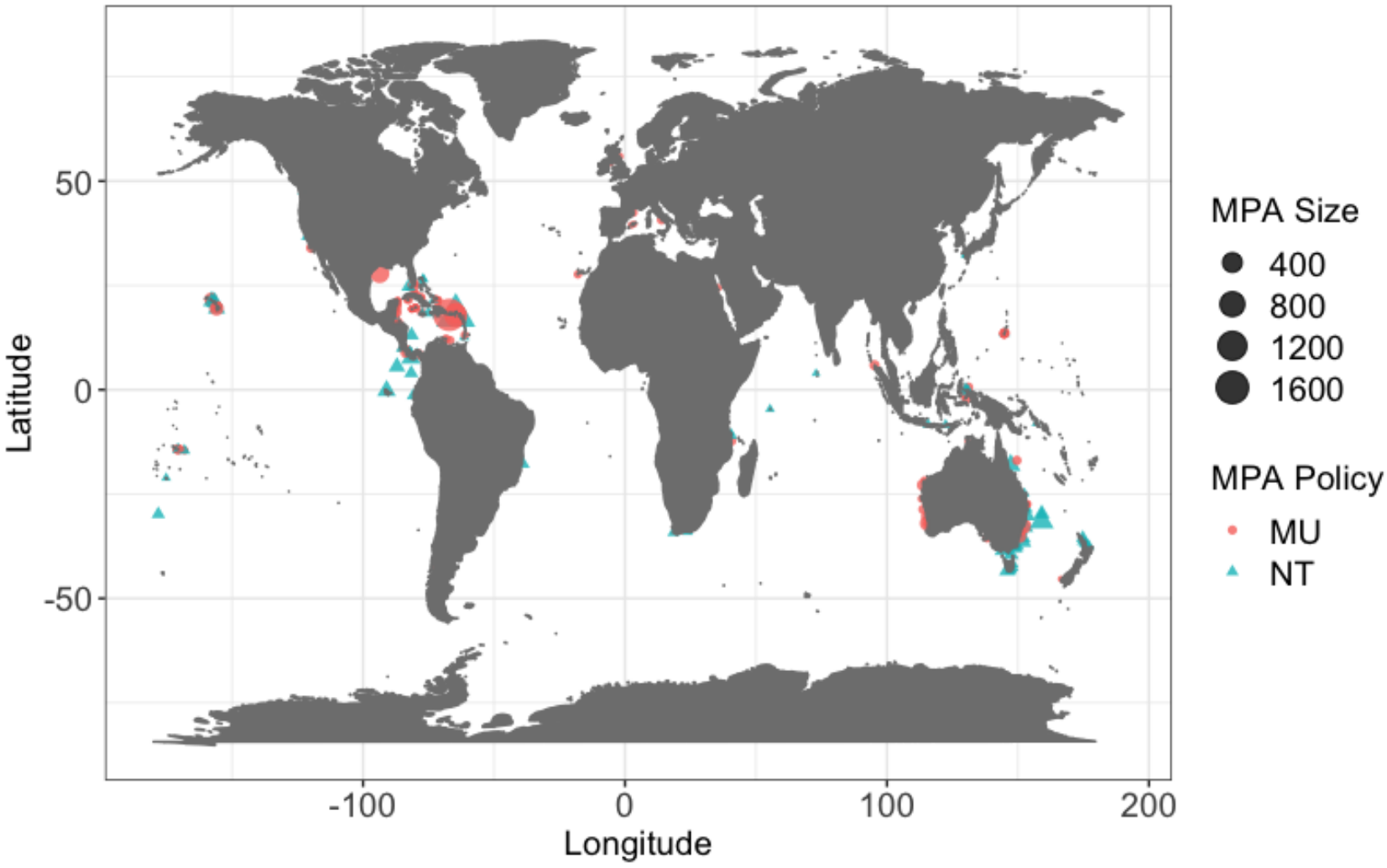
Response (Y) by treatment (A)



Covariates (X)

	Site-level	MPA-level
Continuous	Latitude (degree) Longitude (degree) Depth (m) Wave exposure (kW/m) Distance to shoreline (km) Distance to market (km) Coastal population (million/100km^2) Sample date (year) Minimum sea surface temperature ($^\circ$C) Chlorophyll-A (mg/m^3) Reef area with 15km (km^2) MPA age (years)	MPA size (km^2)
Categorical	Habitat type (16) Marine ecoregion (56) Sampling protocol (6)	Country (43)

Dependence structure



Challenges

- The treatment variables are not randomly assigned
- The allocation of treatments is likely influenced by biomass
- Sites are nested within MPA, giving (spatial) dependence

Potential outcomes (POs) and assumptions

- $Y_i(a)$ is the PO under treatment $a \in \{0, 1\}$
- The PO model for a site in MPA r is

$$Y_i(a) = \mu_a(\mathbf{X}_i) + \alpha_r + \varepsilon_i$$

  where α_r is an MPA random effect

- Assumption: the POs for each unit are not influenced by the treatment assigned to other units
- Assumption: conditional independence between treatment and the POs when accounting for the covariates

Matching estimator

- We propose a matching estimator to account for spatial dependence and missing confounders
- Each site is matched by space and covariates with M observations with the opposite treatment
- The missing PO (i.e., $Y_i(0)$ if $A_i = 1$) are imputed using the mean of these M observations
- The naive matching estimator for the average treatment effect (ATE) is

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} Y_i(1) - Y_i(0)$$

- The average treatment effect on the treated (ATT) is similar

Extensions

- The naive estimator is biased if $\mu_a(\mathbf{X})$ varies by a
- For fit several regression models including random forests to remove this bias
- Deriving the standard error is challenging because of MPA effects and sites are members of multiple pairs
- We develop a wild bootstrap for this purpose
- **Theory**: We prove that our estimator is consistent and asymptotically normal

Results

Table: Summary of the average treatment effect (ATE) and the average treatment effect on the treated (ATT) with estimated standard errors in parentheses when comparing the multi-use (MU) policy and no-taken (NT) policy in MPAs where MU is considered as treatment group; Response is log (Fish Biomass).

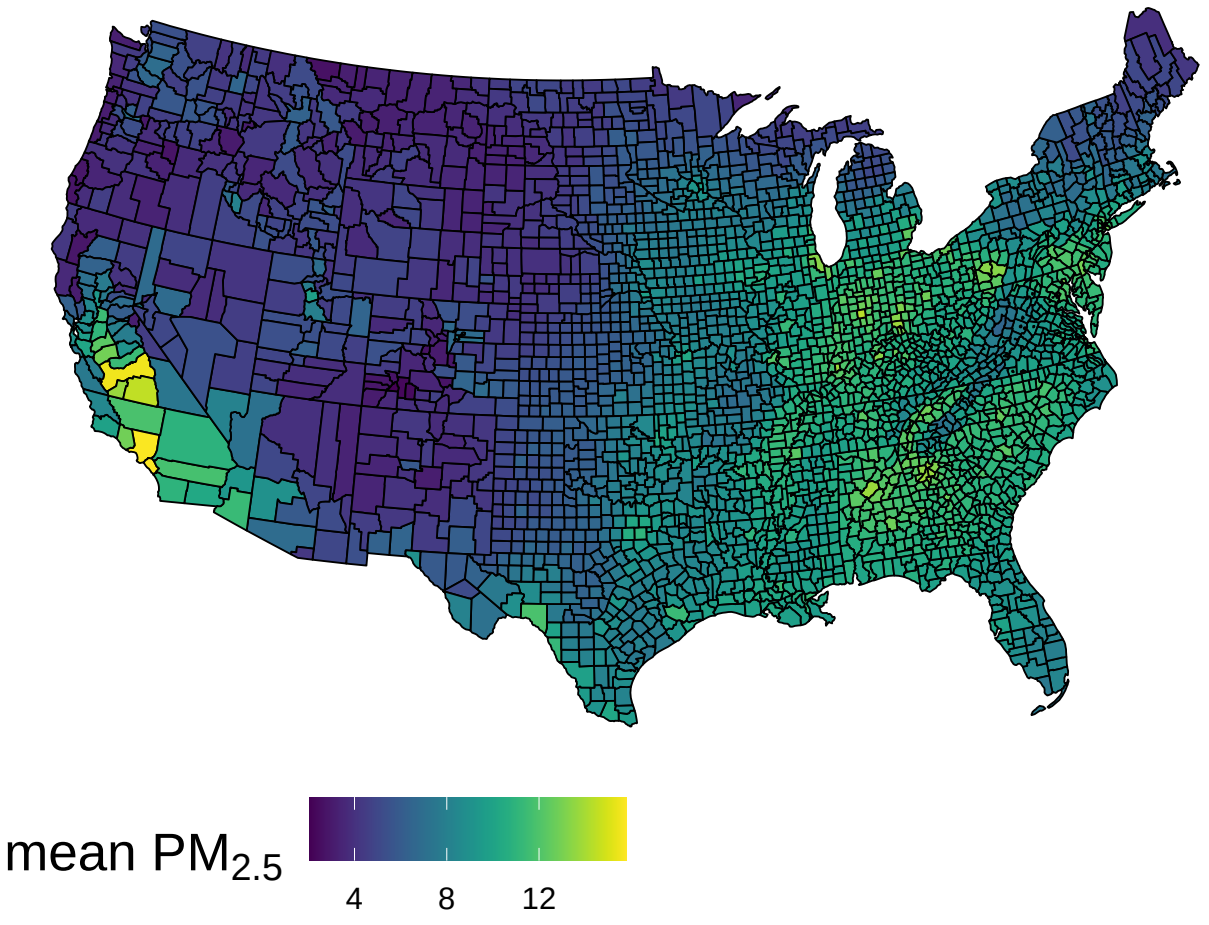
	ATE		ATT	
	Point Estimate	95% CI	Point Estimate	95% CI
Matching on MPA-level covariates				
Sieve Method	-0.49 (0.31)	(-1.10, 0.13)	-0.67 (0.50)	(-1.64, 0.30)
Smooth Spline	-0.27 (0.25)	(-0.77, 0.23)	-0.19 (0.38)	(-0.93, 0.56)
Regression Forest	-0.57 (0.35)	(-1.26, 0.12)	-0.82 (0.49)	(-1.77, 0.13)
Matching on all covariates				
Sieve Method	-0.41 (0.17)	(-0.76, -0.07)	-0.58 (0.26)	(-1.10, -0.06)
Smooth Spline	-0.34 (0.30)	(-0.93, 0.26)	-0.41 (0.41)	(-1.22, 0.40)
Regression Forest	-0.70 (0.32)	(-1.32, -0.07)	-1.03 (0.53)	(-2.06, 0.00)

EXAMPLE: Adjusting for unmeasured spatial confounders

- The key assumption that there are no missing confounders is generally impossible to verify
- However, it may be possible to remove the effects of unmeasured confounders that are smooth spatial functions
- We propose a method to adjust for missing spatial confounding variables using spectral methods.
- **Motivating example** Wu et al (2020) [5] found that an increase of 1 $\mu g/m^3$ in $PM_{2.5}$ is associated with a 15% increase in the COVID-19 mortality rate.

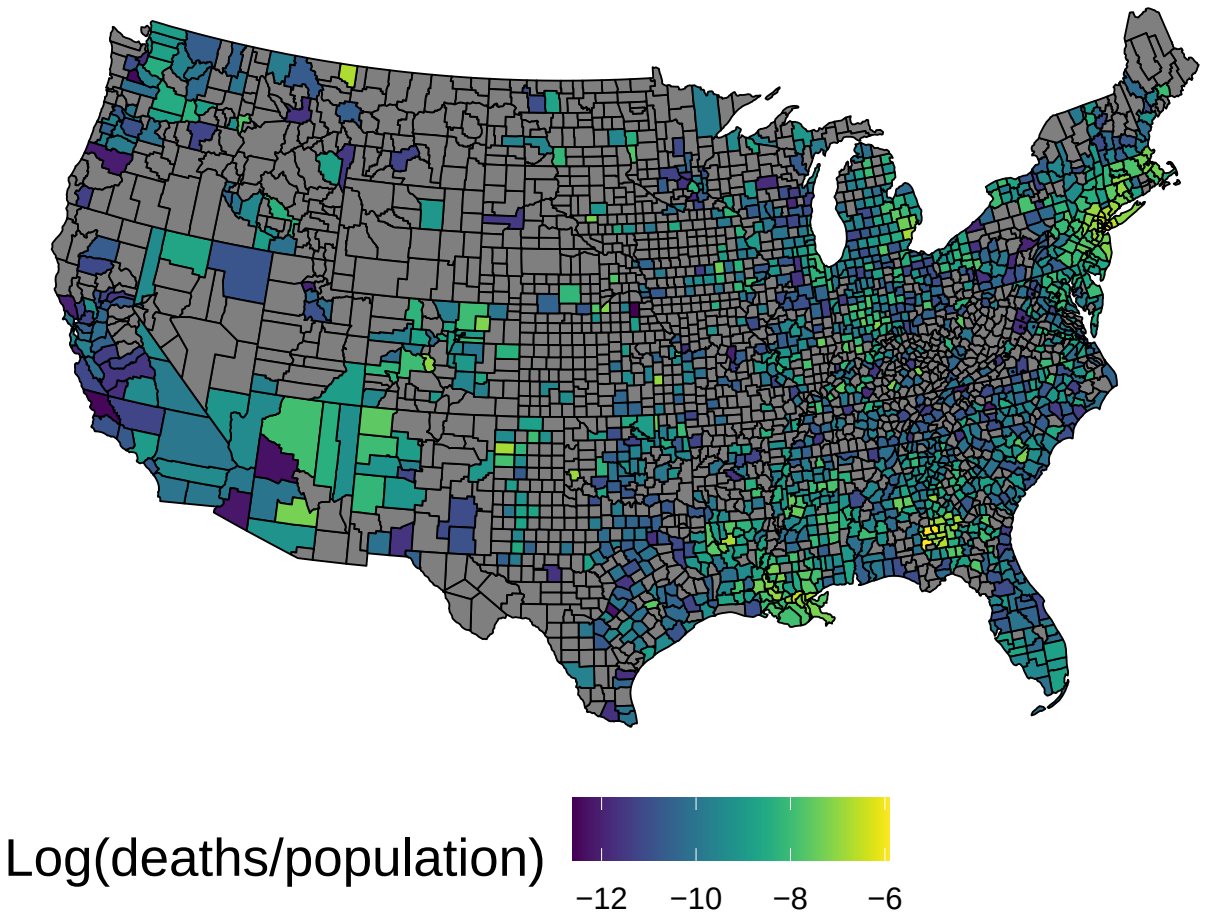
[5] Wu, Nethery, Sabath, Braun, Dominici (2020). Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis, *Science Advances.*

Exposure variable, $PM_{2.5}$ concentration



The exposure variable X is county-level average exposure to $PM_{2.5}$ for 2000-2016

Outcome variable, log mortality rate



The response variable Y is the cumulative COVID-19 mortality counts through May 12, 2020 for US counties.

Our approach

- We assume that local information is more informative than global information
- If you could design the experiment you would assign different $A(\mathbf{s})$ to neighbors, right?
- We propose a joint model for A and the unobserved spatial confounder, X[6]
- We derive the optimal confounder adjustment under the joint model
- We then discuss assumptions needed to identify the parameters in the counfounder adjustment

[6] See Schnell and Papadogeorgou (2019) for areal data, Dupont, Wood and Augustin (2020) for splines and Stokes and Purdon (2017) for time series.

Unmeasured confounders

- $X(\mathbf{s})$ is unknown or not measured
- Obviously, additional assumptions are required to estimate its correlation with $A(\mathbf{s})$
- We assume $[A(\mathbf{s}), X(\mathbf{s})]$ follows a bivariate spatial GP
- Then X can be written $X(\mathbf{s}) = \widehat{X}(\mathbf{s}) + \delta(\mathbf{s})$ where
  - $\widehat{X}(\mathbf{s}) = \mathrm{E}\{X(\mathbf{s})|\mathbf{A}\}$
  - $\delta(\mathbf{s})$ is a Gaussian process that is independent of $\mathbf{A}$

Unmeasured confounders

The observation equation then becomes

$$Y(\mathbf{s}) = \alpha + A(\mathbf{s})\beta_A + \widehat{X}(\mathbf{s})\beta_X + \delta(\mathbf{s}) + \epsilon(\mathbf{s})$$

- $\widehat{X}(\mathbf{s})$ is the required confounder adjustment
- The spatial term $\delta(\mathbf{s})$ is independent of $\mathbf{A}$
- Therefore, if we know $\widehat{X}$ we eliminate spatial confounding
- But how to estimate $\widehat{X}$?

We model A and X using spectral methods

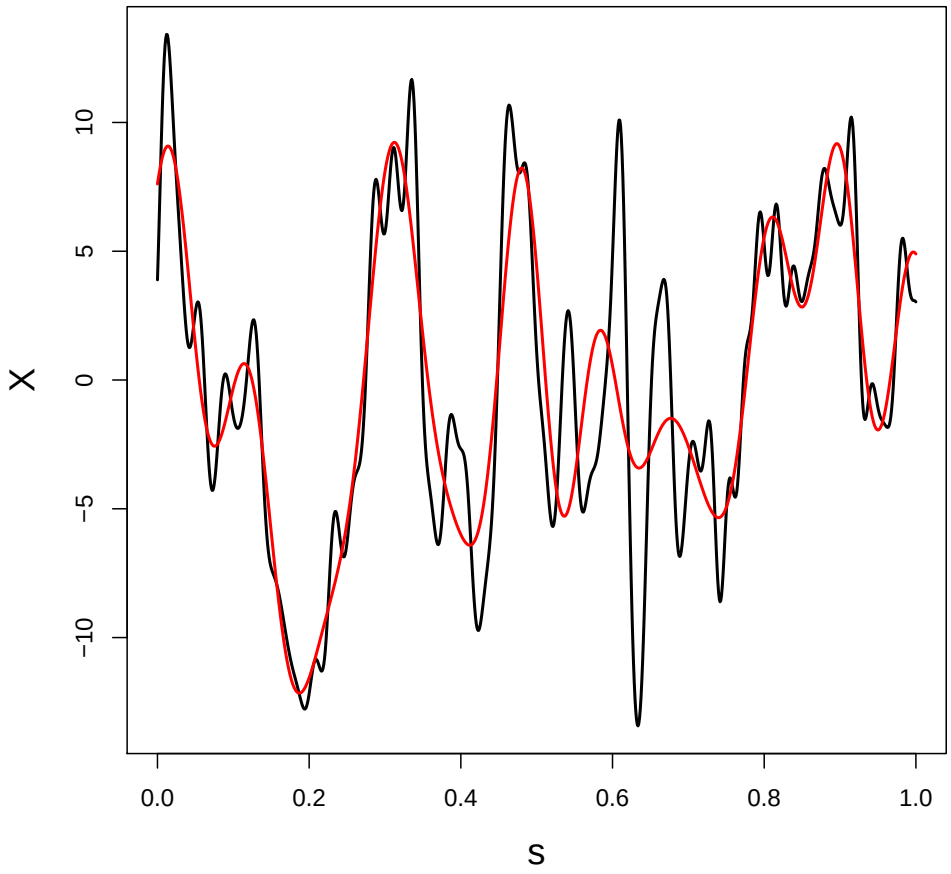
- If A and X are stationary they have spectral representations
- The spectral representations are

$$X(\mathbf{s}) = \int \exp(i\omega^T \mathbf{s}) x(\omega) \mathrm{d}\omega$$

$$A(\mathbf{s}) = \int \exp(i\omega^T \mathbf{s}) a(\omega) \mathrm{d}\omega$$

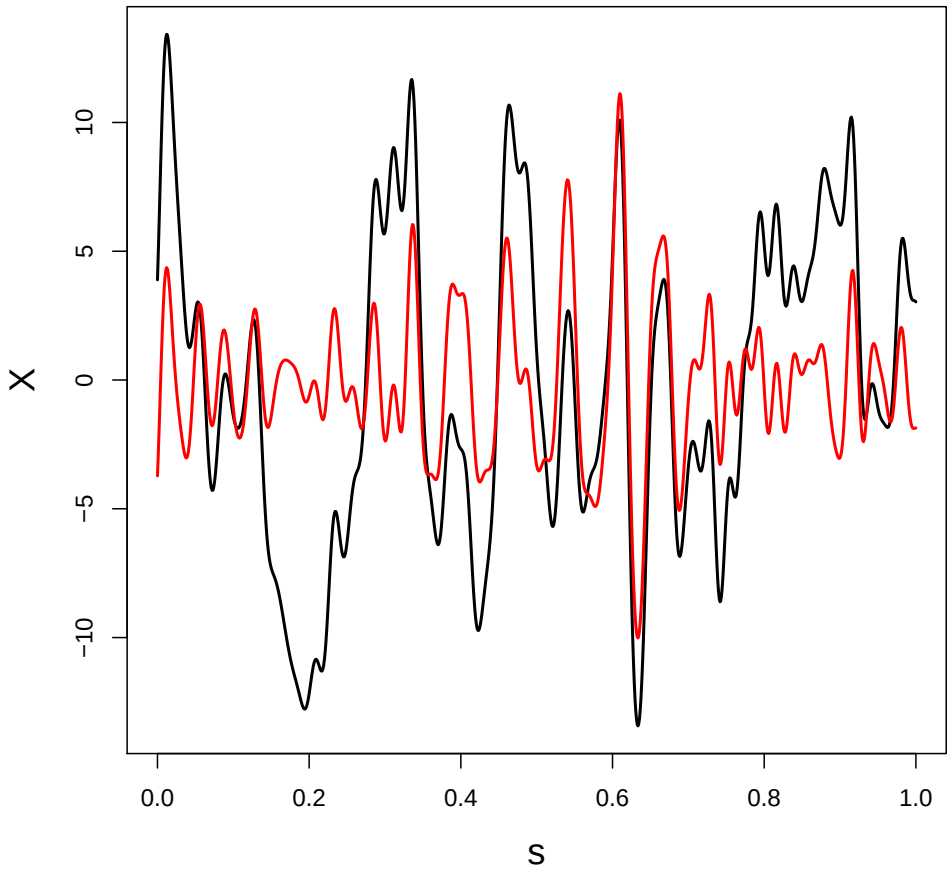
- $\omega \in \mathcal{R}^2$ is a frequency
- $x(\omega)$ and $a(\omega)$ are mean zero and independent across ω

Spectral decomposition in 1D



- Full signal (black): $X(\mathbf{s}) = \int \exp(i\omega^T \mathbf{s}) x(\omega) \mathrm{d}\omega$
- Low-frequencies (red): $\tilde{X}(\mathbf{s}) = \int_{||\omega|| < T} \exp(i\omega^T \mathbf{s}) x(\omega) \mathrm{d}\omega$

Spectral decomposition in 1D



- Full signal (black): $X(\mathbf{s}) = \int \exp(i\omega^T \mathbf{s}) x(\omega) \mathrm{d}\omega$
- High-frequencies (red): $\tilde{X}(\mathbf{s}) = \int_{||\omega|| > T} \exp(i\omega^T \mathbf{s}) x(\omega) \mathrm{d}\omega$

We model X and A using spectral methods

- Dependence is measured by the coherence[1]:

$$\alpha(\omega) = \text{Cor}\{x(\omega), a(\omega)\}$$

  which we allow to vary by frequency

- However, we assume that

$$\lim_{||\omega|| \to \infty} \alpha(\omega) = 0$$

  and prove that we can identify the treatment effect

- This allows for confounding of large-scale features (small ω) but assumes independence at small scales (large ω)

- This is a similar motivation for matching spatial neighbors

[1] Kleiber (2017) has a good review

Model in the spatial domain

- The model in the spatial domain is

$$Y(\mathbf{s}) = \beta_0 + A(\mathbf{s})\beta_x + \widehat{X}(\mathbf{s})\beta_X + \delta(\mathbf{s}) + \epsilon(\mathbf{s})$$

- The confounder adjustment is

$$\widehat{X}(\mathbf{s}) = \int K(\mathbf{u} - \mathbf{s})A(\mathbf{u})\mathrm{d}\mathbf{u},$$

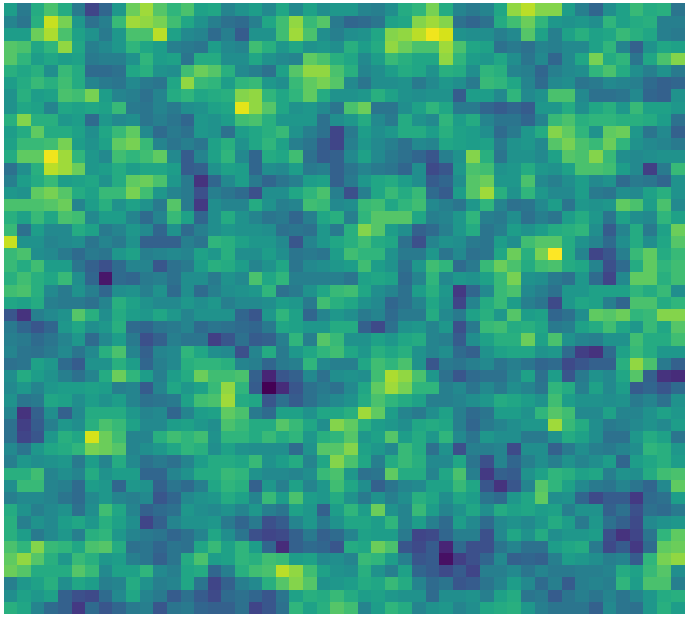
  where $K(\mathbf{u} - \mathbf{s})$ is the inverse Fourier transform of $\alpha(\boldsymbol{\omega})$

- $\widehat{X}$ is a smoothed version of A that is included to adjust for spatial confounding

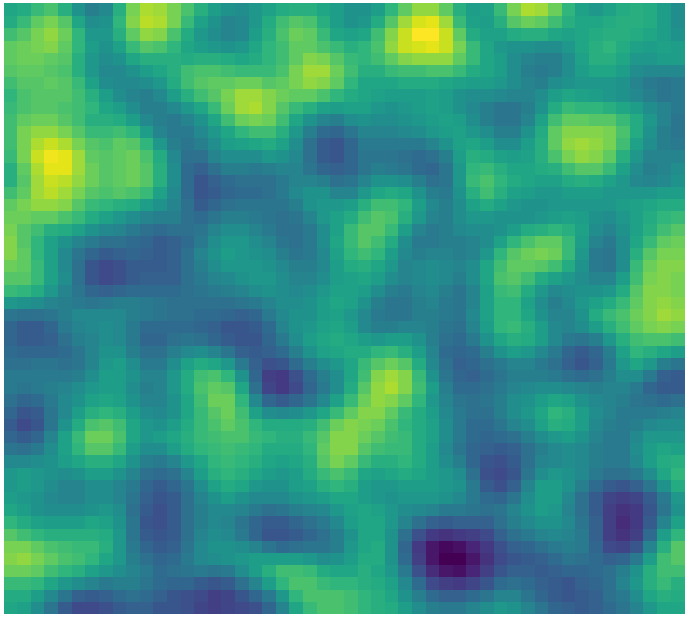
- The methods can be fit using standard MCMC/ML methods

Plot of the covariate A (left) versus $\hat{\mathbf{X}}$ (right)

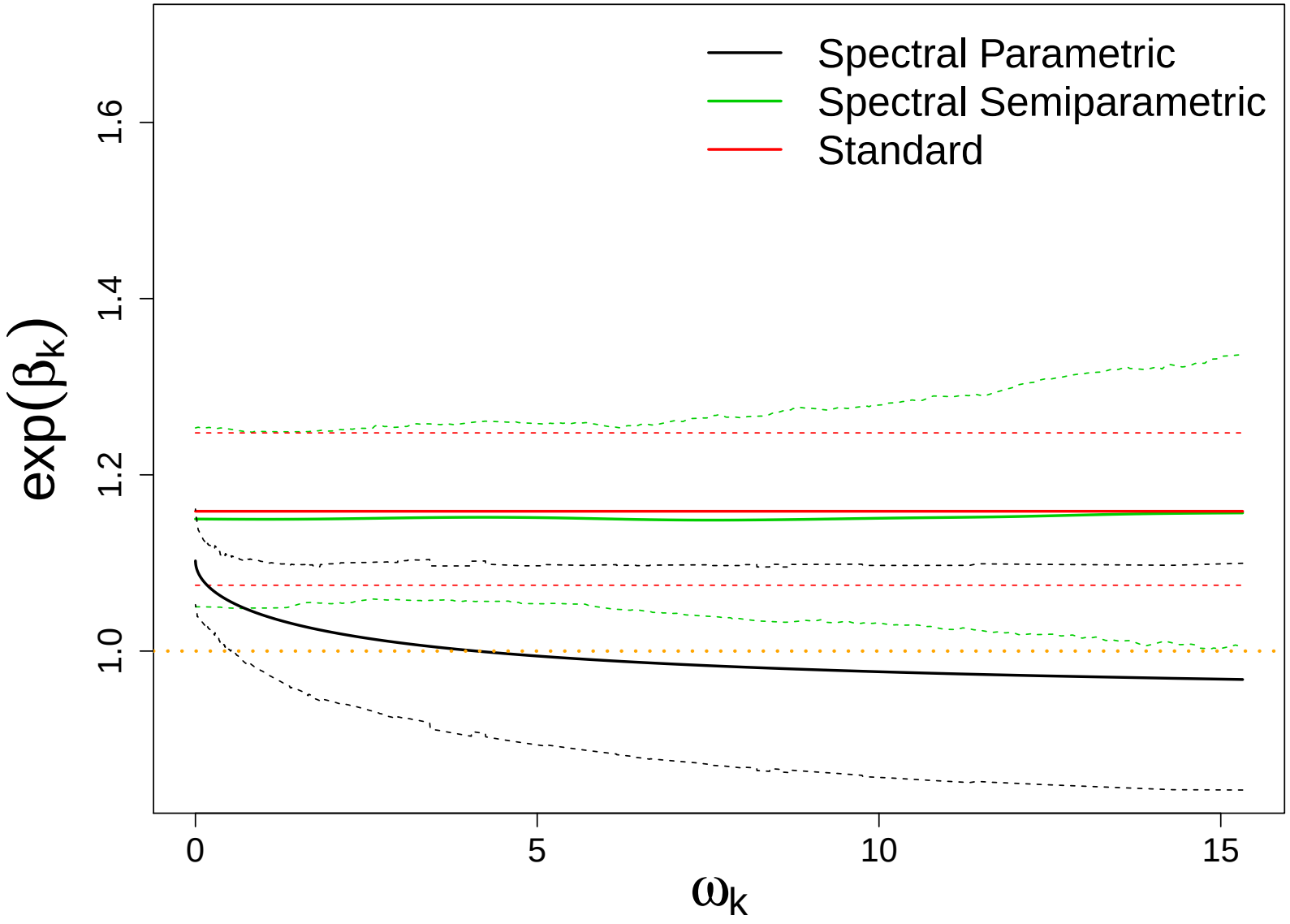
$\nu_{XZ} = 1\nu_X$



$\nu_{XZ} = 3\nu_X$



COVID mortality/PM



- Spatial causal inference is an important and rapidly-evolving area
- We have covered a wide variety of methods, and this is just the surface
- For a more complete review of spatial causal inference, see Reich, et al. (2020). A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review*
- Thanks to the NIH (R01ES031651-01,R01ES027892-01) and NSF (DMS-2152887)