Spatial statistics and causal inference:

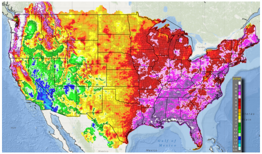
Spatial confounding and interference in air pollution research

Georgia Papadogeorgou

UNC Environmental Epidemiology seminar - Feb 8, 2019

Spatial data and causal inference in air pollution research

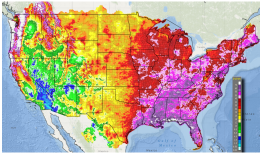
- Variables are expected to have spatial structure
  - Exposure, outcome, covariates



https://kcstormfront.wordpress.com/2015/01/11/2014-in-review/

Spatial data and causal inference in air pollution research

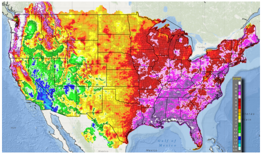
- Variables are expected to have spatial structure
  - Exposure, outcome, covariates
- Questions of interest are often causal
  - What is the effect of a specific intervention on polluting sources?



https://kcstormfront.wordpress.com/2015/01/11/2014-in-review/

Spatial data and causal inference in air pollution research

- Variables are expected to have spatial structure
  - Exposure, outcome, covariates
- Questions of interest are often causal
  - What is the effect of a specific intervention on polluting sources?



- Integration of spatial data and causal inference
  - Spatial correlation of confounding variables
  - Interference, spillover effects

https://kcstormfront.wordpress.com/2015/01/11/2014-in-review/

NO_x emission control technologies

- Regulations such as the Clear Air Act enforce stricter rules on emissions
- Power plants follow different compliance strategies
- We focus on the installation of NO_x emission reduction control technologies

NO_x: Nitric oxide and nitrogen dioxides, precursors of ozone, reacting with other compounds in the presence of sunlight to create ozone

NO_x emission control technologies

- Regulations such as the Clear Air Act enforce stricter rules on emissions
- Power plants follow different compliance strategies
- We focus on the installation of NO_x emission reduction control technologies
- Selective Catalytic Reduction (SCR) and Selective Non-Catalytic Reduction (SNCR) are the most effective in reducing NO_x

---

NO_x: Nitric oxide and nitrogen dioxides, precursors of ozone, reacting with other compounds in the presence of sunlight to create ozone

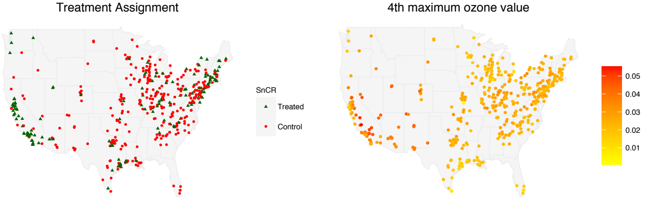
NO_x emission control technologies

- Regulations such as the Clear Air Act enforce stricter rules on emissions
- Power plants follow different compliance strategies
- We focus on the installation of NO_x emission reduction control technologies
- Selective Catalytic Reduction (SCR) and Selective Non-Catalytic Reduction (SNCR) are the most effective in reducing NO_x
- Are SCR/SNCR more effective than alternative strategies in reducing ambient ozone concentrations?

NO_x: Nitric oxide and nitrogen dioxides, precursors of ozone, reacting with other compounds in the presence of sunlight to create ozone

Data

- Coal and natural gas power plants during June-August 2004
- $A = 1$ if at least half of facility heat input is used by units with installed SCR/SNCR technologies, $A = 0$ otherwise
- 152 treated facilities, 321 controls
- Y: NO_x emissions / 4^{th} maximum ambient ozone concentration
- Covariates: Power plant characteristics, demographics, weather



Publicly available data sources: Air Markets Program Data, 2000 Census, EPA monitoring sites

Statistical challenges

- Unmeasured spatial confounding
    - Volatile organic compounds and sunlight is necessary for the creation of ozone
    - They may confound the relationship of NO_x control strategies and ambient ozone
    - Weather and atmospheric covariate information varies spatially

Statistical challenges

- Unmeasured spatial confounding
  - Volatile organic compounds and sunlight is necessary for the creation of ozone
  - They may confound the relationship of NO_x control strategies and ambient ozone
  - Weather and atmospheric covariate information varies spatially
- Interference
  - Pollution travels with the wind
  - "Upwind" pollution sources can affect ambient concentrations in the area surrounding power plants at long distances
  - Discussed in vaccine trials, herd immunity, spillover effects

Adjusting for Unmeasured Spatial Confounders

Notation

- For unit i
  - Treatment $A_i \in \{0, 1\}$
  - Potential outcomes $Y_i(1), Y_i(0)$ (SUTVA)
  - Covariates $L_i = (L_{i1}, L_{i2}, \ldots, L_{ip})$

Notation

- For unit i
  - Treatment $A_i \in \{0, 1\}$
  - Potential outcomes $Y_i(1), Y_i(0)$ (SUTVA)
  - Covariates $L_i = (L_{i1}, L_{i2}, \ldots, L_{ip})$
- Average Treatment Effect on the Treated:

$$ATT = E[Y(1) - Y(0)|A = 1]$$

Notation

- For unit i
  - Treatment $A_i \in \{0,1\}$
  - Potential outcomes $Y_i(1), Y_i(0)$ (SUTVA)
  - Covariates $L_i = (L_{i1}, L_{i2}, \ldots, L_{ip})$
- Average Treatment Effect on the Treated:

$$ATT = E[Y(1) - Y(0)|A = 1]$$

- Positivity: $P(A = 1|L) \in (0,1)$
- Ignorability: $Y(1), Y(0) \amalg A|L$

Notation

- For unit i
  - Treatment $A_i \in \{0, 1\}$
  - Potential outcomes $Y_i(1), Y_i(0)$ (SUTVA)
  - Covariates $L_i = (L_{i1}, L_{i2}, \ldots, L_{ip})$
- Average Treatment Effect on the Treated:

$$ATT = E[Y(1) - Y(0)|A = 1]$$

- Positivity: $P(A = 1|L) \in (0, 1)$
- Ignorability: $Y(1), Y(0) \perp\!\!\!\perp A|L$
- Propensity score matching
  - PS model $P(A = 1|L)$
  - Match treated units to controls with similar PS estimates

Unmeasured spatial confounding

- Confounders $L = (X, U)$
  - X are observed, U are unobserved
- If U varies spatially, can we adjust for it?
  - If a matched pair is sufficiently close, the treated and control units will have similar values of U

Rosenbaum and Rubin [1983]

Unmeasured spatial confounding

- Confounders $L = (X, U)$
  - X are observed, U are unobserved
- If U varies spatially, can we adjust for it?
  - If a matched pair is sufficiently close, the treated and control units will have similar values of U
- Observed variables X:
  - Use the propensity score to adjust for the observed confounders
  - $P(A_i = 1 | X_i) = f(X_i) = \text{expit}\left(X_i^T \beta\right)$

Rosenbaum and Rubin [1983]

Unmeasured spatial confounding

- Confounders $L = (X, U)$
  - X are observed, U are unobserved
- If U varies spatially, can we adjust for it?
  - If a matched pair is sufficiently close, the treated and control units will have similar values of U
- Observed variables X:
  - Use the propensity score to adjust for the observed confounders
  - $P(A_i = 1|X_i) = f(X_i) = \text{expit}\left(X_i^T \beta\right)$
- Navigate the tradeoff between:
  1. Making matches as similar as possible with respect to X
  2. Small distance of matched pairs to capture similarity in U

Rosenbaum and Rubin [1983]

Distance Adjusted Propensity Score Matching

- For a treated unit i and a control unit j define

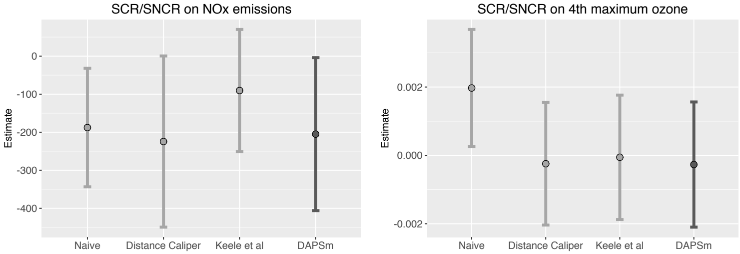
$$DAPS_{ij} = w|PS_i - PS_j| + (1 - w) * Dist_{ij},\ w \in [0, 1]$$

  where PS propensity score estimates, and $Dist$ spatial proximity.
- w: relative importance of the observed and unobserved confounders
  - High values of w - most matching weight on observed covariates
  - Lowe values of w - most matching weight on spatial proximity

Matches

- Average distance of matched pairs
  - Naïve: 1066 miles
  - DAPSm: 141 miles

Results



- Reduction by 205 tons of NO_x emissions (95% CI: $4 - 406$)
- -0.27 (95% CI: -2.1 to 1.56) parts per billion in ambient ozone

– The national ambient air quality standard for ozone is 70 parts per billion.
– Keele et al. [2015]

Conclusions

- We propose a method to reduce bias from spatial unmeasured confounding
- SCR/SNCR control technologies lead to
  - Reductions in NO_x emissions
  - Their effect on ozone is not significant

- Additional information in the paper:
  - How to pick the tuning parameter w
  - Robustness to the choice of w as an indication of no unmeasured spatial confounding
  - Comparison with other methods for incorporating spatial information

Relaxing the no interference assumption

Interference

- Previously, we assumed that each unit had $Y(0), Y(1)$
  - *Your* outcome has nothing to do with *my* treatment
- Treatment effects with "interference"
  - *Your* outcome may depend on your and *my* treatment
  - Potential outcomes $Y(0, 0, \ldots, 0), Y(0, 0, \ldots, 0, 1)$, etc

Interference

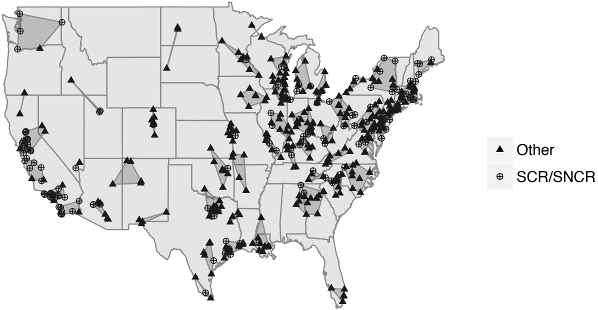
- Previously, we assumed that each unit had $Y(0), Y(1)$
  - *Your* outcome has nothing to do with *my* treatment
- Treatment effects with "interference"
  - *Your* outcome may depend on your and *my* treatment
  - Potential outcomes $Y(0, 0, \ldots, 0), Y(0, 0, \ldots, 0, 1)$, etc
- Vaccine trials, infectious diseases

Interference

- Previously, we assumed that each unit had $Y(0), Y(1)$
  - *Your* outcome has nothing to do with *my* treatment
- Treatment effects with "interference"
  - *Your* outcome may depend on your and *my* treatment
  - Potential outcomes $Y(0, 0, \ldots, 0), Y(0, 0, \ldots, 0, 1)$, etc
- Vaccine trials, infectious diseases
- Ambient pollution concentrations are affected by multiple sources
  - Pollution emitted "locally"
  - Pollution that is *transported* from nearby sources

Partial interference

- *Partial interference*: Partition of units in interference clusters
  - A unit's outcome can depend on the treatment level of units in their cluster
  - Does not depend on treatment of units in other clusters



Defining estimands in the presence of interference

- Causal inference with interference was introduced in the context of two-stage randomized trials [Hudgens and Halloran, 2008]
- Extensions to observation studies consider estimands for two-stage randomized design (Tchetgen Tchetgen and VanderWeele [2012], Perez-Heydrich et al. [2015])

Defining estimands in the presence of interference

- Causal inference with interference was introduced in the context of two-stage randomized trials [Hudgens and Halloran, 2008]
- Extensions to observation studies consider estimands for two-stage randomized design (Tchetgen Tchetgen and VanderWeele [2012], Perez-Heydrich et al. [2015])
- Such estimands represent
  - What would we observe if treatment was assigned randomly to units with probability α?

Defining estimands in the presence of interference

- Causal inference with interference was introduced in the context of two-stage randomized trials [Hudgens and Halloran, 2008]
- Extensions to observation studies consider estimands for two-stage randomized design (Tchetgen Tchetgen and VanderWeele [2012], Perez-Heydrich et al. [2015])
- Such estimands represent
  - What would we observe if treatment was assigned randomly to units with probability α?
- Are these estimands interpretable?
  - Covariates can be predictors of treatment allocation [Barkley et al., 2017]
  - Dependence between units

Counterfactual treatment allocation under realistic interventions

- Let $P_{\alpha,L}$ be the counterfactual treatment allocation
  - How treatment is assigned in a hypothesized world
  - α represents the cluster-average propensity of treatment

Counterfactual treatment allocation under realistic interventions

- Let $P_{\alpha,L}$ be the counterfactual treatment allocation
  - How treatment is assigned in a hypothesized world
  - α represents the cluster-average propensity of treatment
- How would treatment arise in cluster i if
  - The cluster-average propensity of treatment was set to α?
  - Individual treatment adoption depended on a covariate L with log-odds δ_L?

Notation

- Clusters $i \in \{1, 2, \ldots, N\}$ with n_i units
- For unit j in cluster i
  - Treatment $A_{ij} \in \{0, 1\}$
  - Potential outcomes $Y_{ij}(\cdot) = \{Y_{ij}(\mathbf{a}_i), \mathbf{a}_i \in \{0, 1\}^{n_i}\}$,
  - Unit covariates $L_{ij} = (L_{ij1}, L_{ij2}, \ldots, L_{ijp})$

Notation

- Clusters $i \in \{1, 2, \ldots, N\}$ with n_i units
- For unit j in cluster i
  - Treatment $A_{ij} \in \{0, 1\}$
  - Potential outcomes $Y_{ij}(\cdot) = \{Y_{ij}(\mathbf{a}_i), \mathbf{a}_i \in \{0, 1\}^{n_i}\}$,
  - Unit covariates $L_{ij} = (L_{ij1}, L_{ij2}, \ldots, L_{ijp})$
- For cluster i
  - Cluster treatment $\mathbf{A}_i = (A_{i1}, A_{i2}, \ldots, A_{in_i})$
  - Cluster treatment excluding unit j: $\mathbf{A}_{i,-j}$
  - Cluster potential outcomes $\mathbf{Y}_i(\cdot)$
  - Cluster covariates $\mathbf{L}_i$

Covariate dependent counterfactual treatment allocation

How would treatment arise in cluster i if

- The cluster-average propensity of treatment was set to α?
- Individual treatment adoption depended on a covariate L with log-odds δ_L?

Covariate dependent counterfactual treatment allocation

How would treatment arise in cluster i if

- The cluster-average propensity of treatment was set to α?
- Individual treatment adoption depended on a covariate L with log-odds δ_L?

Define

$$\mathrm{logit} P_{\alpha,L}(A_{ij} = 1 | L_{ij}; \delta_L) = \xi_i^{\alpha} + L_{ij}\delta_L$$

where

$$\frac{1}{n_i}\sum_{j=1}^{n_i} P_{\alpha,L}(A_{ij} = 1 | L_{ij};\ \xi_i^{\alpha}, \delta_L) = \alpha$$

Average potential outcome

- Individual average potential outcome

$$\overline{Y}_{ij}(a;\alpha) = \sum_{\mathbf{s}} Y(A_{ij} = a, \mathbf{A}_{i,-j} = \mathbf{s}) P_{\alpha,L}(\mathbf{A}_{i,-j} = \mathbf{s} | A_{ij} = a)$$

  Average all possible treatment allocations where
  - Observation ij gets treatment a
  - Cluster-level treatment probability is α

Average potential outcome

- Individual average potential outcome

  $$\overline{Y}_{ij}(a;\alpha) = \sum_{\mathbf{s}} Y(A_{ij} = a, \mathbf{A}_{i,-j} = \mathbf{s}) P_{\alpha,L}(\mathbf{A}_{i,-j} = \mathbf{s} | A_{ij} = a)$$

  Average all possible treatment allocations where
  - Observation ij gets treatment a
  - Cluster-level treatment probability is α
- Group average potential outcome $\overline{Y}_i(a;\alpha) = \frac{1}{n_i} \sum_{j=1}^{n_i} \overline{Y}_{ij}(a;\alpha)$
- Population average potential outcome $\overline{Y}(a;\alpha) = E_{G_0}\left[\overline{Y}_i(a;\alpha)\right]$, for super-population of clusters G_0

Average potential outcome

- Individual average potential outcome

  $$\overline{Y}_{ij}(a;\alpha) = \sum_{\mathbf{s}} Y(A_{ij} = a, \mathbf{A}_{i,-j} = \mathbf{s}) P_{\alpha,L}(\mathbf{A}_{i,-j} = \mathbf{s} | A_{ij} = a)$$

  Average all possible treatment allocations where
  - Observation ij gets treatment a
  - Cluster-level treatment probability is α
- Group average potential outcome $\overline{Y}_i(a;\alpha) = \frac{1}{n_i} \sum_{j=1}^{n_i} \overline{Y}_{ij}(a;\alpha)$
- Population average potential outcome $\overline{Y}(a;\alpha) = E_{G_0}\left[\overline{Y}_i(a;\alpha)\right]$, for super-population of clusters G_0
- Direct effect for fixed cluster-average treatment propensity

  $$DE(\alpha) = \overline{Y}(1,\alpha) - \overline{Y}(0,\alpha)$$

- Indirect effect between two fixed cluster-average treatment propensity

  $$IE(\alpha_1,\alpha_2) = \overline{Y}(0,\alpha_1) - \overline{Y}(0,\alpha_2)$$

Group and population potential outcome estimators

Assumptions

- **Positivity**: $P(\mathbf{A}_i = \mathbf{a}_i | \mathbf{L}_i) > 0$, for all $\mathbf{a}_i \in \{0, 1\}^{n_i}$
- **Ignorability**: $\mathbf{Y}_i(\cdot) \amalg \mathbf{A}_i | \mathbf{L}_i$

Group and population potential outcome estimators

Assumptions

- **Positivity**: $P(\mathbf{A}_i = \mathbf{a}_i | \mathbf{L}_i) > 0$, for all $\mathbf{a}_i \in \{0,1\}^{n_i}$
- **Ignorability**: $\mathbf{Y}_i(\cdot) \amalg \mathbf{A}_i | \mathbf{L}_i$

Define

$$\widehat{Y}_i(a,\alpha) = \sum_{j=1}^{n_i} \frac{P_{\alpha,L}(\mathbf{A}_{i,-j} | A_{ij} = a, \mathbf{L}_i; \boldsymbol{\delta})}{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{A}_i | \mathbf{L}_i; \boldsymbol{\gamma}) n_i} I(A_{ij} = a) Y_{ij}$$

and

$$\widehat{Y}(a;\alpha) = \frac{1}{N} \sum_{i=1}^{N} \widehat{Y}_i(a,\alpha)$$

where $f_{\mathbf{A}|\mathbf{L},i}(\mathbf{A}_i | \mathbf{L}_i; \boldsymbol{\gamma})$ is the propensity score of the observed treatment vector

Theoretical and practical results

- All results assume that positivity and ignorability hold
- $\widehat{Y}_i(a, \alpha)$ is unbiased for $\overline{Y}_i(a, \alpha)$ (for known propensity score)
- $\widehat{Y}(a; \alpha)$ is consistent (for correctly-specified estimated propensity score)
- Asymptotic results are derived for increasing number of clusters
- Asymptotic or bootstrap CIs are acquired
- Performance was checked in an extensive simulation study
- Coverage of bootstrap CIs was better than asymptotic CIs for a small number of clusters

Propensity score and counterfactual treatment allocation

- Propensity score of *observed* treatment

$$\mathrm{logit} P(A_{ij} = 1 | L_{ij}) = \delta_0 + b_i + L_{ij}^T \boldsymbol{\delta},\ \ b_i \sim N\left(0, \sigma_b^2\right)$$

Cluster propensity score:

$$f_{\mathbf{A}|\mathbf{L},i}(\mathbf{A}_i | \mathbf{L}_i; \boldsymbol{\gamma}) = \int \prod_{j=1}^{n_i} P(A_{ij} | L_{ij}, \boldsymbol{\delta}, b_i) f(b_i | \sigma_b^2) \mathrm{d}b_i$$

Propensity score and counterfactual treatment allocation

- Propensity score of *observed* treatment

$$\text{logit}P(A_{ij} = 1|L_{ij}) = \delta_0 + b_i + L_{ij}^T \boldsymbol{\delta},\ b_i \sim N\left(0, \sigma_b^2\right)$$

Cluster propensity score:

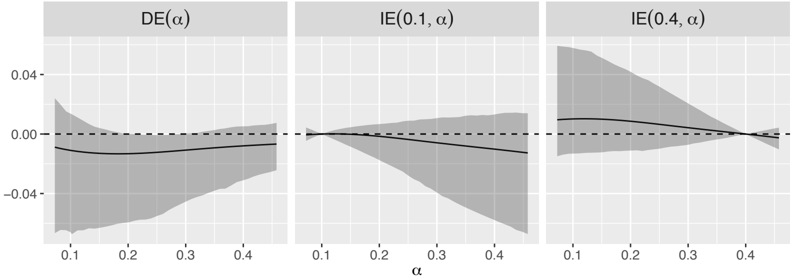
$$f_{\mathbf{A}|\mathbf{L},i}(\mathbf{A}_i|\mathbf{L}_i;\boldsymbol{\gamma}) = \int \prod_{j=1}^{n_i} P(A_{ij}|L_{ij},\boldsymbol{\delta},b_i) f(b_i|\sigma_b^2)\mathrm{d}b_i$$

- Use the observed treatment allocation to inform $P_{\alpha,L}$

$$\text{logit}P_{\alpha,L}(A_{ij} = 1|L_{ij};\boldsymbol{\delta}) = \xi_i^{\alpha} + L_{ij}^T \boldsymbol{\delta},\ \text{where}$$

$$\frac{1}{n_i}\sum_{j=1}^{n_i} P_{\alpha,L}(A_{ij} = 1|L_{ij};\ \boldsymbol{\delta}) = \alpha$$

Direct and indirect effect of SCR on ambient ozone



$$DE(\alpha) = \overline{Y}(1, \alpha) - \overline{Y}(0, \alpha)$$

$$IE(\alpha_1, \alpha_2) = \overline{Y}(0, \alpha_1) - \overline{Y}(0, \alpha_2)$$

Ozone is measured in parts per million
The national ozone air quality standard of 0.07 parts per million.

Concluding remarks

- Estimands for realistic public health interventions
  - Cluster-average propensity of treatment
  - Distribution of cluster-average propensity of treatment
- Proposed consistent estimators and derived asymptotic variances
- SCR/SNCR technologies are more effective in decreasing ozone against alternatives
  - In the surrounding area
  - In the surrounding area of other power plants

Papadogeorgou, Mealli, and Zigler [2018b]

References

Brian G Barkley, Michael G Hudgens, John D Clemens, Mohammad Ali, and Michael E Emch. Causal Inference from Observational Studies with Clustered Interference. 2017. URL `https://arxiv.org/pdf/1711.04834.pdf`.

Michael G Hudgens and M. Elizabeth Halloran. Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103(482):832–842, jun 2008.

Luke Keele, Rocío Titiunik, and José Zubizarreta. Enhancing a Geographic Regression Discontinuity Design Through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout. *Journal of Royal Statistical Society A*, 178:223–239, 2015.

Georgia Papadogeorgou, Christine Choirat, and Corwin Zigler. Adjusting for Unmeasured Spatial Confounding with Distance Adjusted Propensity Score Matching. *Biostatistics*, 2018a. arXiv:1610.07583.

Georgia Papadogeorgou, Fabrizia Mealli, and Corwin Zigler. Causal inference for interfering units with cluster and population level treatment allocation programs. arXiv:1711.01280, 2018b.

Carolina Perez-Heydrich, Michael G Hudgens, M. Elizabeth Halloran, John D. Clemens, Mohammad Ali, and Michael E. Emch. Assessing Effects of Cholera Vaccination in the Presence of Interference. *Biometrics*, 33(4):395–401, 2015.

P R Rosenbaum and D B Rubin. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55, 1983. doi: 10.1093/biomet/70.1.41.

E. J. Tchetgen Tchetgen and T. J. VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012.