A Spectral Confounder Adjustment for Spatial Regression with Multiple Exposures and Outcomes

Shih-Ni Prim

August 5, 2026

Collaborators

- Brian Reich, Shu Yang [1]
- Yawen Guan [2]
- Ana G Rappold, Lloyd Ana G Rappold, K. Lloyd Hill, Wei-Lun Tsai, Corinna Keeler[3]

[1]NC State
[2]Colorado State University
[3]US EPA, Center for Public Health and Environmental Assessment

Table of contents

Section 1

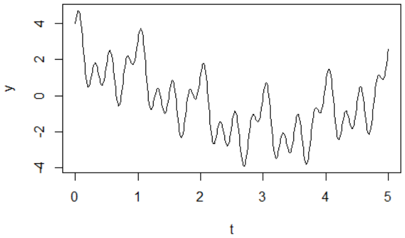
Background & Motivation

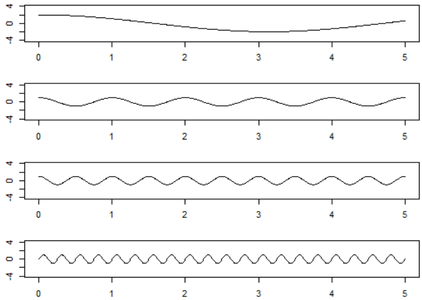
Spatial confounding

- Spatial confounding can arise from the presence of spatially-structured missing confounders.
- $Y(s) = \beta_0 + \beta_1 X(s) + \theta(s) + \epsilon(s).$
- Example: With **air pollution** as exposure and **cancer** as outcome, a possible missing confounder is **access to healthcare**, which has a spatial structure.
- Accounting for spatial confounding is crucial for providing unbiased effect estimates.

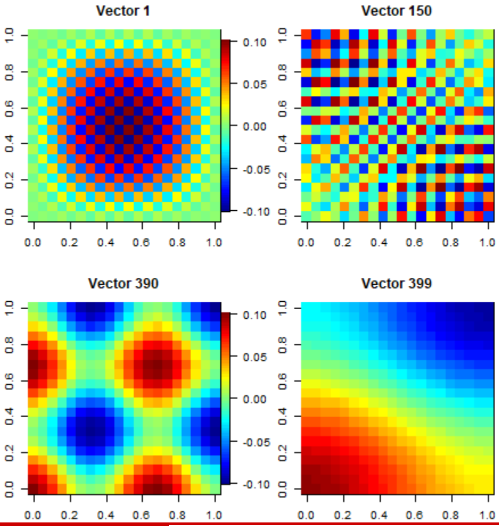
Decomposition of signal

A spatial process can be decomposed into multiple spatial scales.





Eigenvectors (from spectral decomposition) represent different spatial scales



Spatial confounding adjustment

- **Intuition**: Confounders have smooth patterns; differences in exposures and outcomes in neighboring cities are unlikely caused by confounders.
- **Fact**: Spatial information can be decomposed into different spatial scales.
- **Inherent assumption**: Confounding exists in smooth spatial trend/global spatial scales.
- Methods for spatial confounding adjustment
  - Geoadditive structural equation modeling (gSEM) removes an estimated spatial trend and uses residuals (Thaden & Kneib 2018, Wiecha et al. 2025).
  - Spatial+ includes columns of $\mathbf{\Gamma}$ as covariates (Dupont et al. 2022, Urdangarin et al. 2024).

Spectral method and multi-scale coefficent estimates

- Spectral method (Guan et al. 2023) projects spatial data onto spectral domain.
    - Project the spatial data onto the spectral domain (using the eigenvector matrix of the conditional autoregressive (CAR) model).
    - Spatial scales are independent from one another.
    - Each spatial scale is given its own effect estimate.
    - The no-unmeasured-confounder assumption is relaxed to allow for **global confounding**, which is possible because of the multiscale coefficients.
    - This study only allows for one exposure and one outcome.

Multiple exposures and multiple outcomes

- We are exposed to many harmful chemicals, e.g. different PFAS compounds.
- These exposures affect multiple outcomes, e.g. hypertension and hyperlipidemia.
- Multivariate models allow for efficiency gain and borrowing strength across outcome.

Section 2

Model Details

Proposed method: Multivariate Spectral Model (MSM)

- We extend the spectral model in Guan et al. (2023) to multiple exposures and outcomes.
- The multiscale estimates allow for filtering out confounded information.
- The coefficients are now structured in a three-way tensor.

Multivariate Spectral Model (MSM)

Spatial model ("naive"):

$$Y_{sr} = \eta_{0r} + \sum_{p=1}^{P} Z_{sp}\eta_{pr} + \sum_{e=1}^{E} X_{se}\beta_{er} + \theta_{sr} + \varepsilon_{sr}$$

- Y_{sr} is the r^{th} outcome at location s.
- Z_{sp} is the p^{th} adjusting covariates at location s.
- X_{se} is the e^{th} exposure at location s.
- θ_{sr} is the random effect associated with r^{th} outcome at location s. (θ_{sr} is modeled using a multivariate CAR.)
- ϵ_{sr} is the residual associated with r^{th} outcome at location s.

Multivariate Spectral Model (MSM)

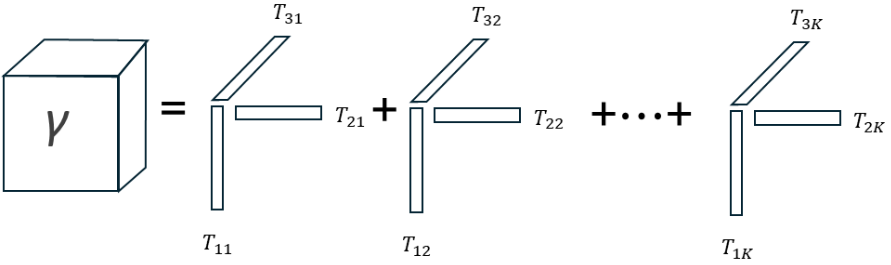
$$\underbrace{\mathbf{\Gamma}^\top Y_r}_{Y_r^*} = \mathbf{\Gamma}^\top \mathbf{1}\eta_{0r} + \sum_{p=1}^{P} \underbrace{\mathbf{\Gamma}^\top Z_p}_{Z_p^*} \eta_r + \sum_{e=1}^{E} \underbrace{\mathbf{\Gamma}^\top X_e}_{X_e^*} \boldsymbol{\beta}_{er} + \underbrace{\mathbf{\Gamma}^\top \boldsymbol{\theta}_r}_{\theta_r^*} + \underbrace{\mathbf{\Gamma}^\top \boldsymbol{\varepsilon}_r}_{\epsilon_r^*},$$

where $\mathbf{Y}_r = (Y_{r1}, \cdots, Y_{rn})^\top$ and similarly for $\mathbf{Z}_p, \mathbf{X}_e, \boldsymbol{\theta}_r$, and $\boldsymbol{\varepsilon}_r$.

- MSM allows the coefficient to vary by spatial scale, which means $\boldsymbol{\beta}_{er}$ is now an $n \times 1$ vector.
- $\boldsymbol{\beta}$ is a three-way tensor with dimension $N \times E \times R$, with β_{ier} being the effect estimate for spatial scale i, exposure e, and outcome r.

Canonical polyadic tensor decomposition

- We assume **the coefficients vary smoothly across spatial scales** and **a low-rank structure** of the tensor coefficient.
- Two steps of dimension reduction:
  - **Basis function**: $\beta_{ier} = \sum_{l=1}^{L} B_{il}\gamma_{ler}$, $L << N$, where $B_{\cdot 1}, \cdots, B_{\cdot L}$ are L basis functions.
  - Canonical polyadic (CP) **tensor decomposition** (Kolda & Bader 2009): $\gamma_{ler} = \sum_{k=1}^{K} T_{1lk} T_{2ek} T_{3rk}$.



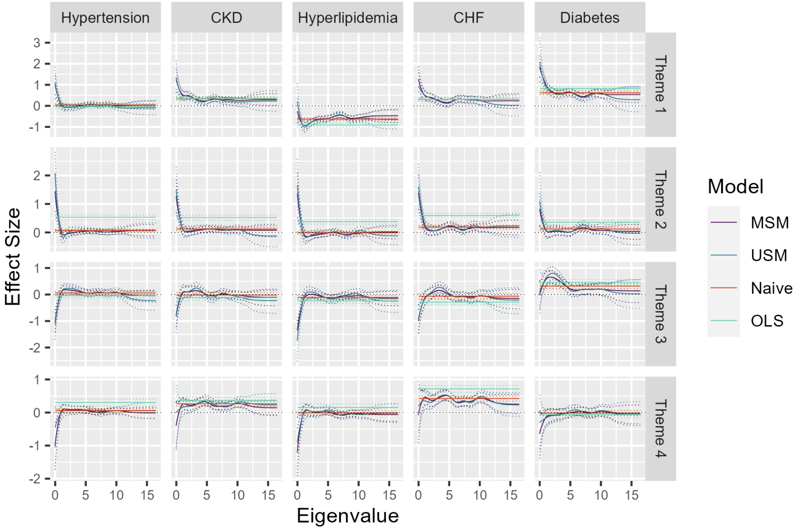
Section 3

Data analysis

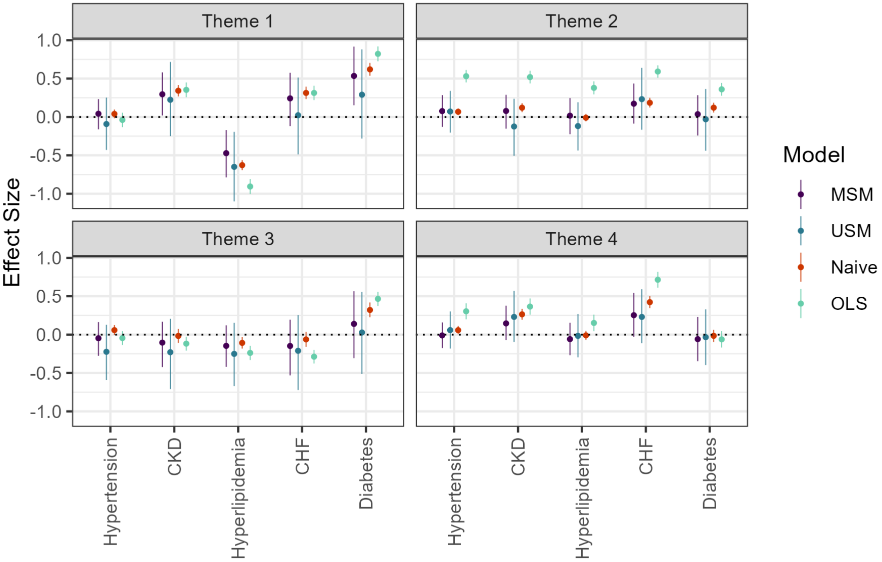
Data analysis

- Use Social Vulnerabillity Index (SVI) to predict health outcomes in 10,149 ZIP Code Tabulation Areas (ZCTAs) in the south region.
- **Multiple exposures**: SVI (from 2018): between 0 and 1; higher score means worse preparation for disasters
  - Theme 1: lack of economic resilience
  - Theme 2: household level adoptive capacity
  - Theme 3: fraction of minority or non-English speaking population
  - Theme 4: housing type and transportation
- **Multiple outcomes**: 5 Chronic health conditions (incidence rates in ZCTAs) from 2017-2019: Hypertension, Chronic Kidney Disease (CKD), Hyperlipidemia, Congestive Heart Failure (CHF), Diabetes

Multiscale coefficient estimates



Data analysis result



Section 4

Conclusion

Conclusion

- We extend spectral confounding adjustment tool for settings with multiple exposures and multiple outcomes.
- MSM provides multiscale effect estimates through a tensor, which allows for borrowing strengths across exposures and outcomes.
- Assuming no local confounding, MSM provides unbiased estimates.

Thank you!

Website: https://snprim.github.io/
Email: prims@wfu.edu
Email: sprim@ncsu.edu

Section 5

References

Dupont, E., Wood, S. N. & Augustin, N. H. (2022), 'Spatial+: A novel approach to spatial confounding', *Biometrics* **78**(4), 1279–1290.

Guan, Y., Page, G. L., Reich, B. J., Ventrucci, M. & Yang, S. (2023), 'Spectral adjustment for spatial confounding', *Biometrika* **110**(3), 699–719.

Kolda, T. G. & Bader, B. W. (2009), 'Tensor Decompositions and Applications', *SIAM Review* **51**(3), 455–500.

Leroux, B. G., Lei, X. & Breslow, N. (2000), Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence, *in* M. E. Halloran & D. Berry, eds, 'Statistical Models in Epidemiology, the Environment, and Clinical Trials', Springer New York, New York, NY, pp. 179–191.

Thaden, H. & Kneib, T. (2018), 'Structural equation models for dealing with spatial confounding', *The American Statistician* **72**(3), 239–252.

Urdangarin, A., Goicoa, T., Kneib, T. & Ugarte, M. D. (2024), 'A simplified spatial+ approach to mitigate spatial confounding in multivariate spatial areal models', *Spatial Statistics* **59**, 100804.

Wiecha, N., Hoppin, J. A. & Reich, B. J. (2025), 'Two-stage estimators for spatial confounding with point-referenced data', *Biometrics* .

Section 6

Appendix

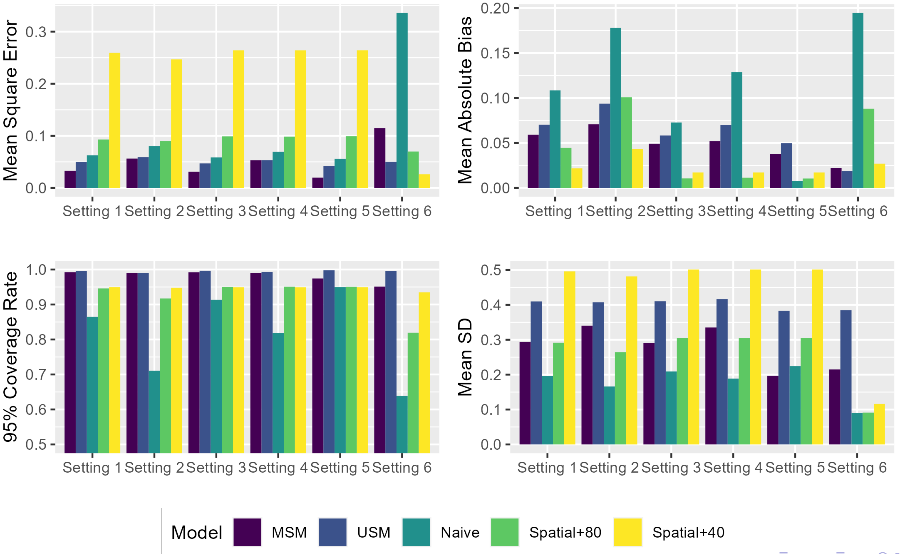
Appendix: Conditional autoregressive (CAR) model

- The CAR model is used as a spatial model for the outcome.
- For an arbitrary random vector $\mathbf{V} = (V_1, ..., V_S)^\top$, the CAR model with Leroux parameterization (Leroux et al. 2000) is multivariate normal with mean zero and covariance matrix

$$\sigma_V^2 \left\{(1-\lambda_V)\mathbf{I}_S + \lambda_V \mathbf{Q}\right\}^{-1}.$$

- We denote this model by $\mathbf{V} \sim \mathrm{CAR}(\sigma_V^2, \lambda_V)$.
- Adjacency matrix $\mathbf{A}$ contains $a_{ij} = 1$ if areas i and j are adjacent and 0 otherwise. $a_{ii} = 0$.
- The precision matrix $\mathbf{Q}$ has values $\sum_t a_{st}$ on the diagonal and $-a_{st}$ on the off diagonals.
- Let $\boldsymbol{\Gamma}\mathbf{W}\boldsymbol{\Gamma}^T$ be the spectral decomposition of $\mathbf{Q}$.

Appendix: Additional simulation results



Appendix: Tensor shows latent structure