Bayesian Kernel Machine Regression Distributed Lag Models

Ander Wilson, Hsiao-Hsien Leon Hsu, Yueh-Hsiu Mathilda Chiu, Robert O. Wright, Rosalind J. Wright, Brent A. Coull

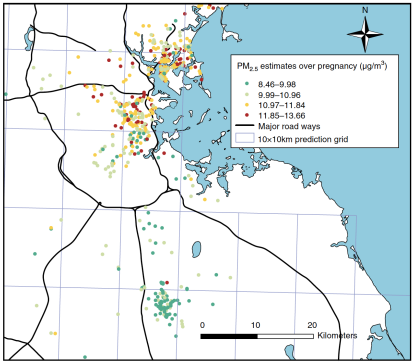
August 24, 2021
International Society of Environmental Epidemiology

Overall Motivation

- Substantial research on mixture methods but mostly focused on exposure observed at a single time point
- Many methods to identify critical windows, mostly based on distributed lag models (DLMs), but generally for only a single exposure
- Our motivation is to identify critical windows and estimate the exposure-response relationship for mixtures that are assessed longitudinally

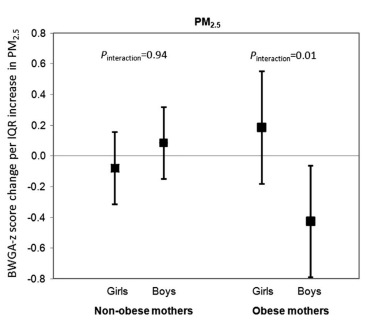
ACCESS Prospective Birth Cohort

- 997 Boston-area births between 8/2002 and 1/2007
- $PM_{2.5}$ and components at maternal residence for each week of pregnancy
- Primary outcome is birth weight for gestational age *z*-score (BWGAZ)
- Baseline covariates: maternal, age, education, race/ethnicity, pre-pregnancy BMI, atopy, self reported smoking during pregnancy, stress index, neighborhood disadvantage index, season of birth

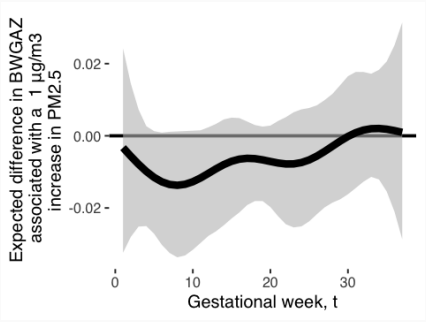


Source: Hsu et al. *Am J Respir Crit Care Med*.

Existing Evidence



Source: Lakshmanan et al. (2015) *Env. Res.*

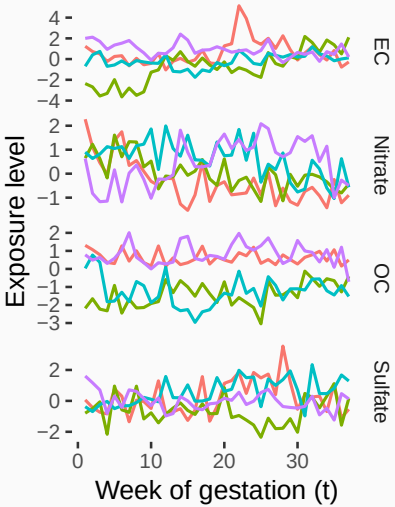


Source: Wilson et al. (2017) *Biostatistics*

- Evidence of an association between $PM_{2.5}$ exposure and BWGAZ for male babies with obese mothers ($n = 109$)
- What about nitrate, organic carbon (OC), elemental carbon (EC) and sulfate?

Data Setting and Objectives

- Resulting data structure is 4 exposures at 37 time points
  - High dimensional exposure data
  - Can be highly multicollinear
- Want a model that allows for:
  - Nonlinear associations
  - Interaction effects
  - Identification of critical windows



Bayesian Kernel Machine Regression (BKMR)

- BKMR estimates a high dimensional exposure-response function

$$Y_i = h(E_{i1}, \ldots, E_{iM}) + Z_i'\gamma + \epsilon_i$$

- For repeated measures of exposures we can:
  1. Reduce the dimension of the exposure by using pregnancy average exposure
     - Ignores variation in exposure and the magnitude of the effect over pregnancy
     - Cannot identify critical windows
     - May obscure effects that are limited to a small number of weeks
  2. Include all measures of exposures
     - May result in model instability
     - Fails to account for biological understand that exposure effect at proximal time points will likely be similar in sign and magnitude
     - Hard to interpret results

Distributed Lag Model (DLM)

- Our approach is to use the constrained DLM framework within BKMR
  - Reduces the dimension of the exposure data
  - Adds structure so that the exposure-effect varies smoothly across gestational weeks
  - Improves model stability
- The DLM model for a single exposure is

$$Y_i = \alpha + \sum_{t=1}^{T} X_{it}\delta_t + \mathbf{Z}_i'\boldsymbol{\gamma} + \epsilon_i$$

- Regression coefficient δ_t is constrained to vary smoothly over time
- Alternatively can be represented in a functional form using continuous time

$$\sum_{t=1}^{T} X_{it}\delta_t \quad \Longleftrightarrow \quad \int_{\mathcal{T}} x_i(t)\delta(t)dt$$

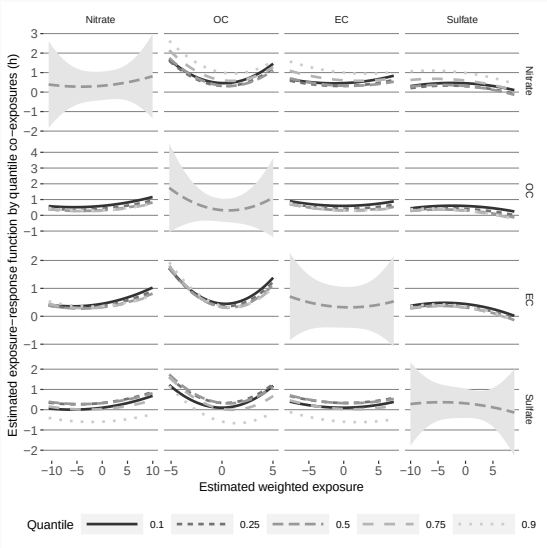
BKMR-DLM

- Represent each exposure as a weighted exposure
$$E_{im} = \int_{\mathcal{T}} w_m(t) x_{im}(t) dt$$
- Weight function $w_m(t)$ identifies critical windows similar to a constrained DLM
- The BKMR model estimates the exposure-response surface as a function of the weighted exposures
$$Y_i = h(E_{i1}, \ldots, E_{iM}) + \mathrm{Z}_i'\boldsymbol{\gamma} + \epsilon_i$$
- Bayesian model fitting estimates the weight functions $w_m(t)$ and the mixture effect $h()$ simultaneously.
- Weight functions are constrained to have norm one and positive integral for identifiability and parameterized with natural splines

BKMR-DLM Properties

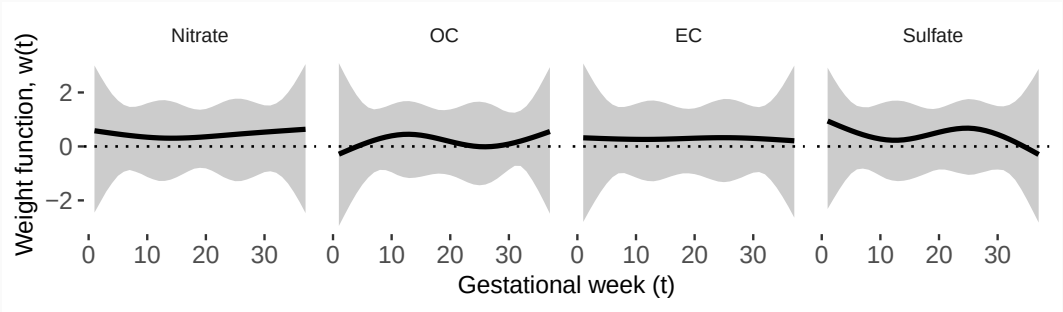
- Allows for nonlinear associations and higher order interactions
- Identifies critical windows through the weight function
- Exposures may have positive effect of some time periods and negative over other time periods
- Simulation results show
  - Ability to estimate nonlinear exposure-response functions and interactions
  - Low power to detect critical windows
  - Improved estimation of the exposure-response function even when a critical window cannot be identified
  - Very low frequency of identifying incorrect critical windows

Illustration on ACCESS Data



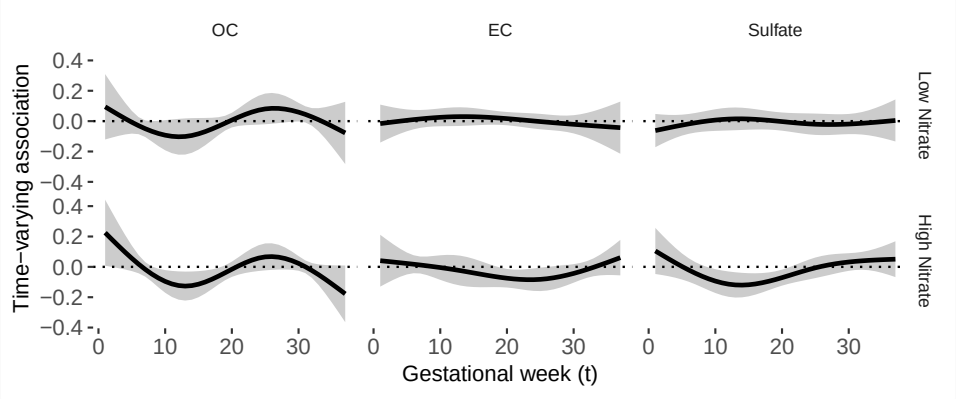
- Similar interpretation of exposure-response function to BKMR
- Key difference: x-axis is weighted exposure (estimated)

Illustration on ACCESS Data



- Areas of the weight function that deviate from zero identify critical windows
- The sign of the weight function does not determine the direction of the association

Illustration on ACCESS Data



- DLM for each exposure stratified by mean nitrate level over pregnancy
- BKMR-DLM is a potentially powerful exploratory tool to identify possible interactions to investigate with more parsimonious models

Summary

- BKMR-DLM allows for repeated measures of exposure to a mixture
- Improved estimation of exposure-response function compared to BKMR with pregnancy average exposures or an additive DLM or DLNM
- Allows for nonlinear associations and interactions
- Low power to identify critical windows but can in high-signal settings
- Software available: anderwilson.github.io/regimes

Thanks You!

This work was supported in part by NIH grants R01ES028811, R01ES013744, P30ES000002, P30ES023515, and UH3OD023337 and US EPA grant RD-83587201. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, US EPA does not endorse the purchase of any commercial products or services mentioned in the publication. The ACCESS cohort has been supported by NIH grants R01ES010932, U01HL072494, and R01HL080674.

Thanks You!

anderwilson.github.io

ander.wilson@colostate.edu

Wilson, A, Hsu, H-HL, Chiu, Y-HM, Wright, RO, Wright, RJ, Coull, BA (In press). Kernel Machine and Distributed Lag Models for Assessing Windows of Susceptibility to Environmental Mixtures in Children's Health Studies. *Annals of Applied Statistics*. ArXiv:1904.12417.

software: anderwilson.github.io/regimes