

Treed Distributed Lag Mixture Models

Daniel Mork and Ander Wilson



Colorado State University

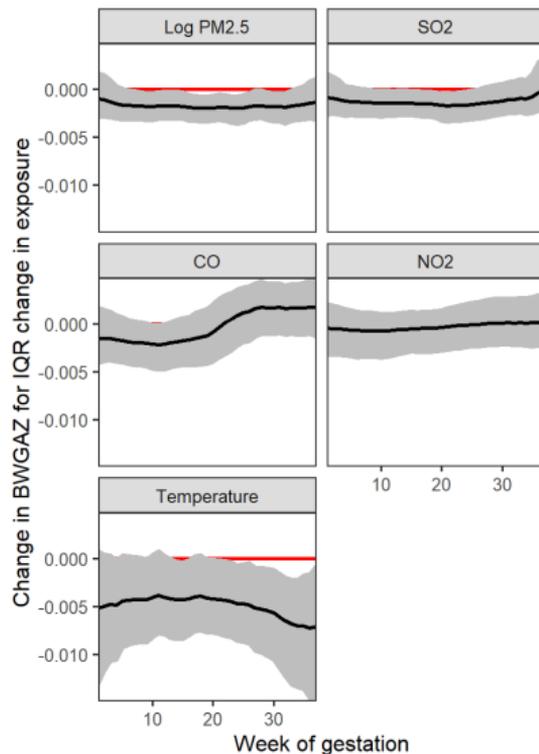
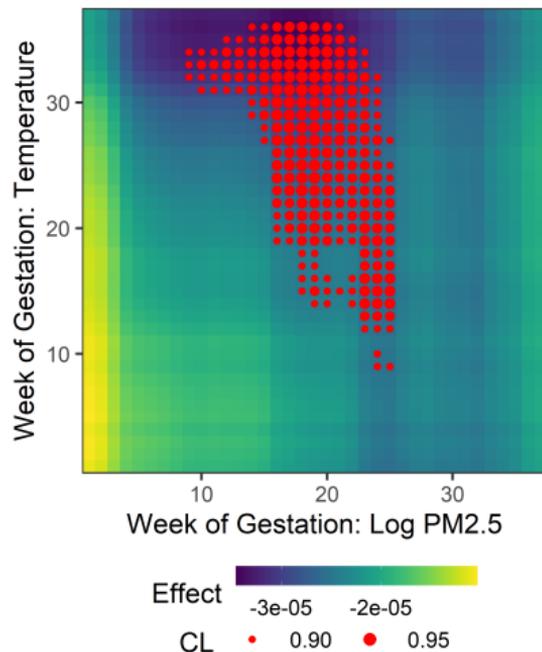
April 28, 2021

Presented at ISGlobal Exposome Data Challenge Event

Treed Distributed Lag Mixture Models

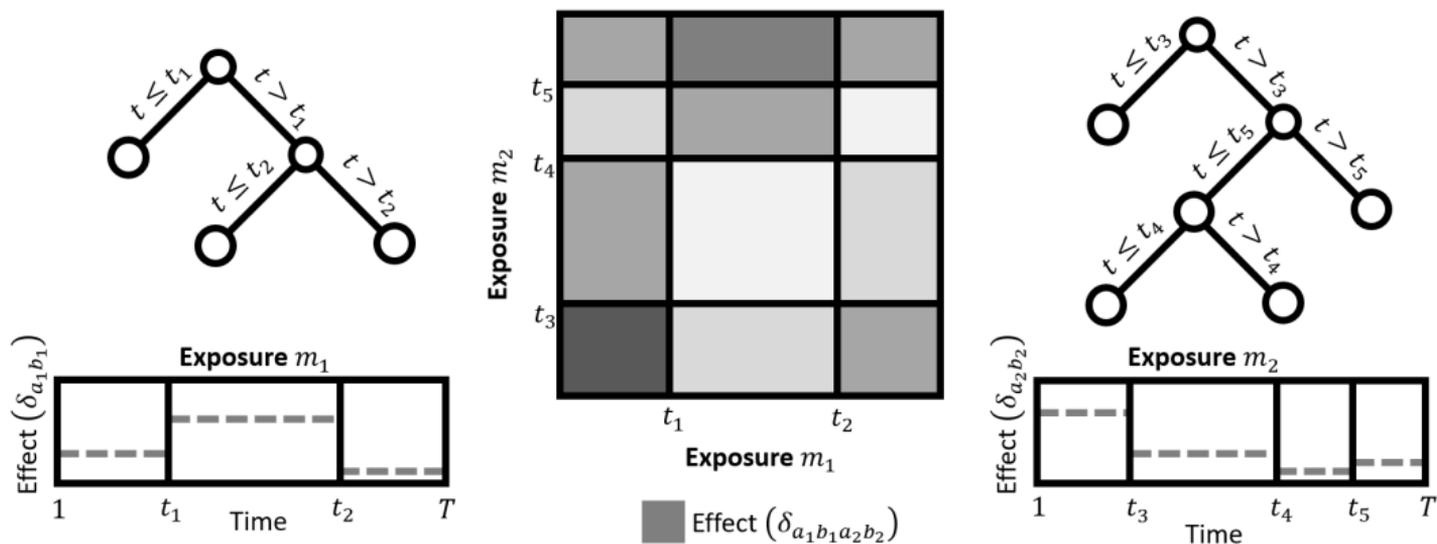
- ▶ Estimate the association between **repeated measures of exposures** and an outcome
- ▶ Use BART style model that:
 - fits an interpretable parametric model
 - uses trees to add structure in the effects of repeated measures of exposures
 - performs exposure selection of main effects and interactions
- ▶ Mork & Wilson (2021). Estimating Perinatal Critical Windows to Environmental Mixtures via Structured Bayesian Regression Tree Pairs. arXiv:2102.09071.

Treed Distributed Lag Mixture Models



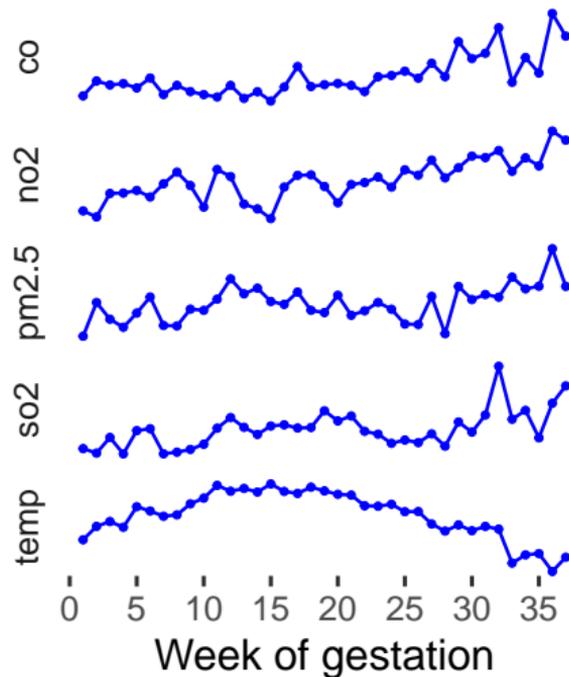
$$E(Y_i) = \underbrace{\sum_{m=1}^M \sum_{t=1}^T x_{imt} \theta_{mt}}_{\text{main effects}} + \underbrace{\sum_{m_1=1}^M \sum_{m_2=m_1}^M \sum_{t_1=1}^T \sum_{t_2=1}^T x_{im_1 t_1} x_{im_2 t_2} \theta_{m_1 m_2 t_1 t_2}}_{\text{pairwise interactions}} + \underbrace{z_i' \gamma}_{\text{Covariates}}$$

Treed Distributed Lag Mixture Models



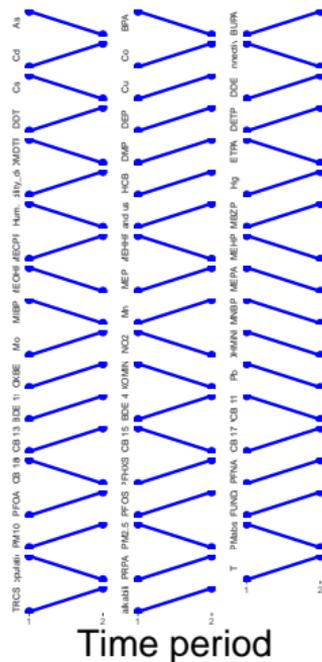
- ▶ Tree splits over time
- ▶ Each time segment has a constant exposure effect
- ▶ Interactions between time segments defined by two separate trees
- ▶ Tree structure and exposures applied to are learned from the data
- ▶ Use additive ensemble of tree pairs

From many to two time points



← Intended use
 - 1 to many exposures
 - many time points

Current use →
 - many exposures
 - few time points



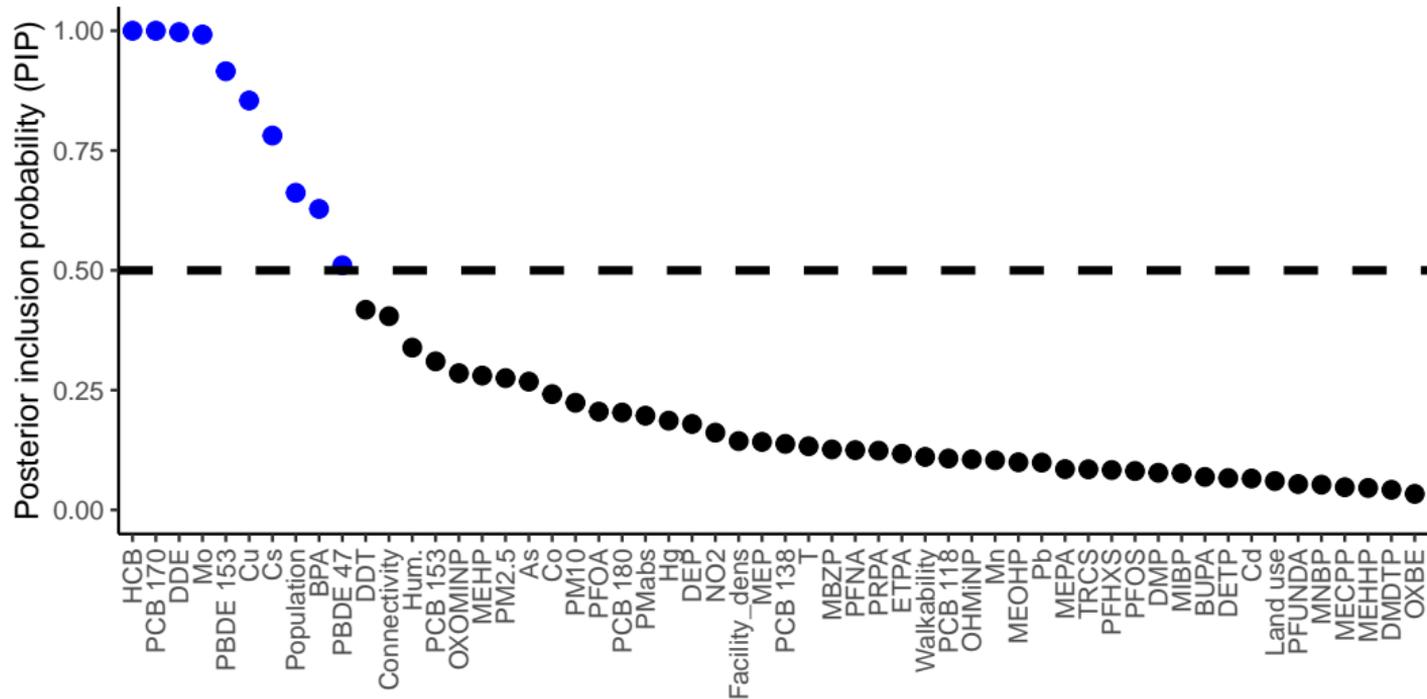
TDLMM with two time points

- ▶ Exposure enters the model as either:
 - two separate predictors
 - cumulative exposure as a single predictor
- ▶ Interactions use the same measure(s) of exposures
- ▶ Exposures can be selected in and out of the model
- ▶ Interactions only included if both main effects are included
- ▶ Shrinkage applied to regression coefficients

Exposome Analysis: Overview

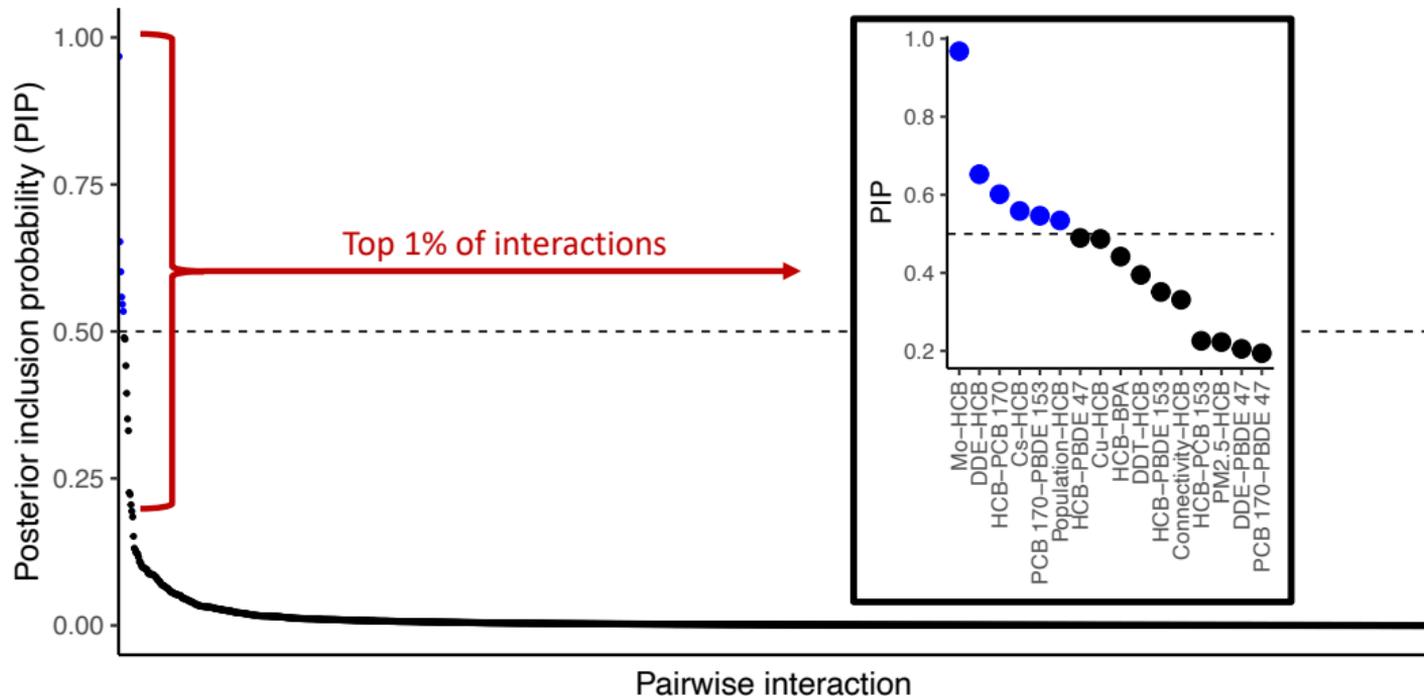
- ▶ Outcome: BMI z-score
- ▶ Covariates: cohort, gestational age at birth, maternal BMI, maternal weight gain, maternal education, parity, native to country of birth
- ▶ Exposures: 56 exposures measured at 2 time points
 - 1540 pairwise interactions
 - 6272 parameters for exposures ($n = 1301$)

Exposome Analysis: Main effect selection

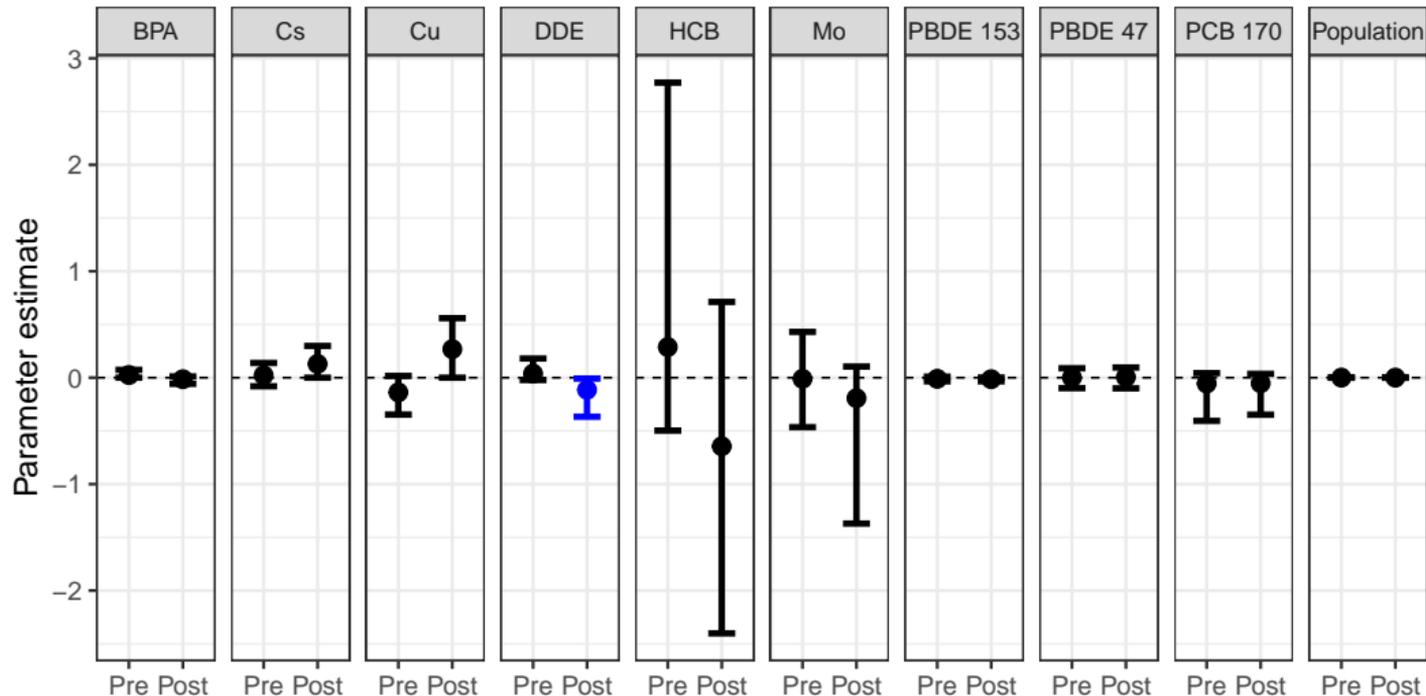


Can also obtain Bayes factor for main effects

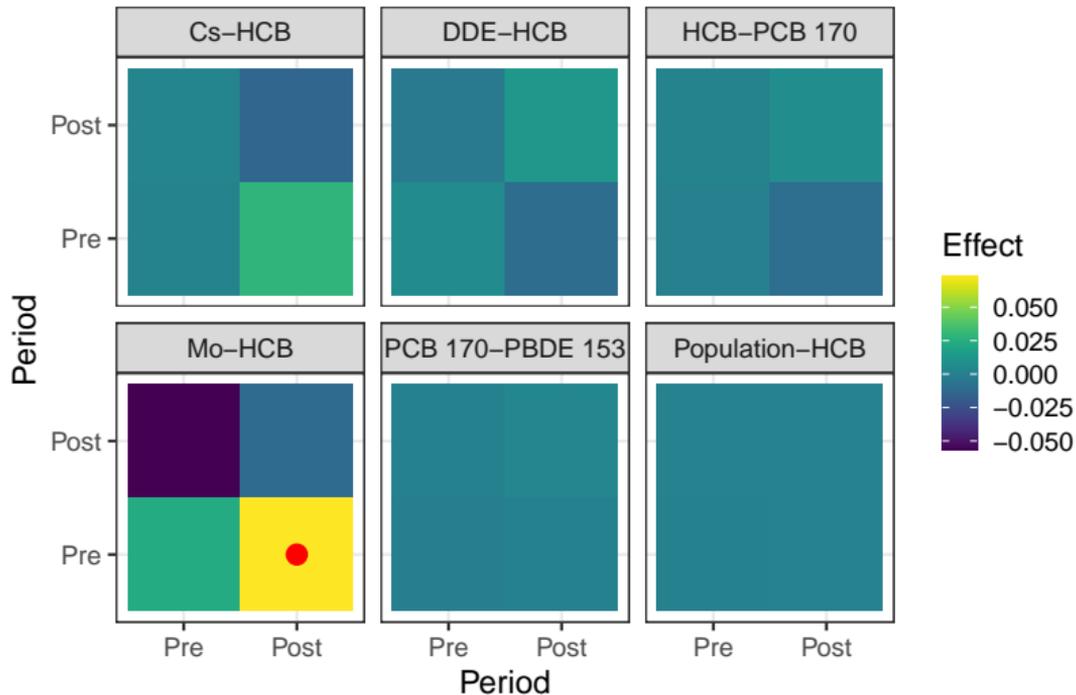
Exposome Analysis: Interaction effect selection



Exposome Analysis: Main effect estimates



Exposome Analysis: Interaction effect estimates



Method overview

- ▶ Most useful when exposures are observed with many repeated measures
- ▶ Allows for quadratic main effects and pairwise interactions
- ▶ Interactions can be time-sensitive or time-lagged
- ▶ Hierarchical variable selection
- ▶ Works with $p \gg n$
- ▶ No higher-order interactions or highly complex exposure response
- ▶ Computationally efficient
- ▶ Easy to interpret

Thank you!

- ▶ Mork & Wilson (2021). Estimating Perinatal Critical Windows to Environmental Mixtures via Structured Bayesian Regression Tree Pairs. arXiv:2102.09071.
- ▶ R package: github.com/danielmork/dlmtree
- ▶ Contact:
 - ander.wilson@colostate.edu
 - anderwilson.github.io
- ▶ Thanks to Glen McGee and Brent Coull for help on the exposome application
- ▶ This work was supported by NIH grants ES029943 and ES028811. This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder and Colorado State University.