DISSERTATION

BAYESIAN TREED DISTRIBUTED LAG MODELS

Submitted by

Daniel S. Mork

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2021

Doctoral Committee:

Advisor: Ander Wilson

Julia Sharp
Josh Keller
Andreas Neophytou

ABSTRACT

BAYESIAN TREED DISTRIBUTED LAG MODELS

In many applications there is interest in regressing an outcome on exposures observed over a previous time window. This frequently arises in environmental epidemiology where either a health outcome on one day is regressed on environmental exposures (e.g. temperature or air pollution) observed on that day and several proceeding days or when a birth or children's health outcome is regressed on exposures observed daily or weekly throughout pregnancy. The distributed lag model (DLM) is a statistical method commonly implemented to estimate an exposure-time-response function by regressing the outcome on repeated measures of a single exposure over a preceding time period, for example, mean exposure during each week of pregnancy. Inferential goals include estimating the exposure-time-response function and identifying critical windows during which exposures can alter a health endpoint.

In this dissertation, we develop novel formulations of Bayesian additive regression trees that allow for estimating a DLM. First, we propose treed distributed lag nonlinear models to estimate the association between weekly maternal exposure to air pollution and a birth outcome when the exposure-response relation is nonlinear. We introduce a regression tree-based model that accommodates a multivariate predictor along with parametric control for fixed effects. Second, we propose a tree-based method for estimating the association between repeated measures of a mixture of multiple pollutants and a health outcome. The proposed approach introduces regression tree pairs, which allow for estimation of marginal effects of exposures along with structured interactions that account for the temporal ordering of the exposure data. Finally, we present a framework to estimate a heterogeneous DLM in the presence of a potentially high dimensional set of modifying variables. We present simulation studies to validate the models. We apply these methods to esti-

mate the association between ambient pollution exposures and birth weight for a Colorado, USA birth cohort.

ACKNOWLEDGEMENTS

There are milestones in life when we are given the opportunity to reflect on our journey and the people who have guided us along the way. For me, the completion of my doctoral degree is one such milestone. While there are truly thousands of individuals who shape our paths, a few stand out with respect to this accomplishment and I would name them here.

Undoubtedly the reason I am writing this today, is due to the loving encouragement of my parents, Barb and Steve Mork, and my most wonderful partner in life, Samantha Sickbert. Together with my brothers, Thomas and Patrick Mork, these people have continuously believed in me and I am infinitely thankful for their support.

I am grateful to my advisor, Dr. Ander Wilson, for being an exceptional mentor and advocate. His guidance, patience, and insights have shaped me as a statistician and researcher. I am most thankful for the countless hours he spent helping me hone my writing and ideas.

The backbone of this academic accomplishment is also a result of a lifetime of supportive teachers or mentors—people who pushed me to realize my potential. Those who stand out include several grade school teachers: Marilyn Peterson, Sue Howard, and Ron Moberg; college professor Dr. Jeremy Loebach; and the many faculty at CSU who have supported me. Specifically, I wish to thank my committee, Dr. Julia Sharp, Dr. Josh Keller, and Dr. Andreas Neophytou, for their feedback and thoughtful discussion on this dissertation. I am also very grateful to have had the opportunity to work with Dr. Julia Sharp in the CSU Graybill Statistics and Data Science Laboratory; she has been an invaluable mentor and role model of what it means to be a professional statistician.

Finally, I am thankful for friends and family, both in and out of the Statistics program. They offered motivation and reprieve from the intense focus required to pursue a doctoral degree. In particular, Nicholas and Leah Ramberg, who always offer inspiration and a reminder of the best parts of life, as well as Jan and Al Sickbert, who are a constant source of encouragement and adventure.

DEDICATION

I dedicate this dissertation to the scientists, medical professionals, and essential workers who pulled us through the COVID-19 pandemic.

TABLE OF CONTENTS

Chapter 1

Introduction

Humans are continuously exposed to thousands of chemicals and pollutants. Many chemicals are known to be detrimental to human health. Harmful chemicals are present in our indoor home and workplace environments (e.g. cook stove exhaust, lead paint, cleaning products, industrial compounds) as well as in the outdoor air we breath (e.g. automotive and industrial exhausts, wildfire smoke, dust, agricultural products, smog). Some pollutants, such as fine particulate matter ($PM_{2.5}$) or semi-volatile organic compounds, are small enough that they can enter our bloodstream through inhalation or skin absorption (Weschler and Nazaroff, 2012; Morrison et al., 2016). For an expectant mother, chemical exposures may pose risks to a developing fetus and alter birth and health outcomes. In particular, air pollution exposure during pregnancy has been linked to detrimental birth outcomes including decreased birth weight (Bell et al., 2007), increased risk of preterm birth (Stieb et al., 2012), increased risk of asthma (Lee et al., 2018; Bose et al., 2017), and altered neurological outcomes (Chiu et al., 2016), among others (Šrám et al., 2005).

1.1 Perinatal Critical Windows

Fetal development occurs in distinct stages throughout gestation. As a specific example in lung development, bronchial tree formation during the pseudoglandular stage in weeks $5 - 17$ builds from trachea and lung analge developed during late embryonic stage in weeks $4 - 7$ (Schittny, 2017). Each developmental stage incorporates specific biological mechanisms, which if interrupted may alter future processes and later outcomes. In the context of lung development, an expectant mother's exposure to pollutants may induce non-regular mechanisms including hypoxia, oxidative stress, and epigenetic influences (Veras et al., 2017). These mechanisms can alter the trajectory of lung development and predispose an unborn child to certain diseases such as asthma and wheezing.

To better understand the nuance in how maternal exposure to air pollution during pregnancy is related to changes in developmental processes, research focuses on identifying critical windows, or

time periods when exposure can alter phenotypes observed in birth or health outcomes of the child. Critical windows can be as short as a week or span many months and are hypothesized to align with sensitive developmental stages. Identifying these windows can shed light on the mechanism by which an exposure effects an outcome. An equally important goal is to estimate the exposure-response relationship to give insight into the magnitude of the effects. The effect magnitude at different time points throughout the exposure period will influence the overall exposure-response function.

Traditionally, most research on perinatal exposures estimates the relationship between air pollution exposure and birth outcomes using pregnancy average or trimester average exposure to a single pollutant or mixtures of chemicals. While this approach gives a broad understanding of the impact of pollution on fetal development, the true relationship may be more nuanced and there is potential for bias in the results when critical windows do not align with clinically defined trimesters. To explore developmentally driven windows, Hazlehurst et al. (2021) modeled $PM_{2.5}$ exposure aligned to four specific stages of lung development and found exposure during the saccular phase ($24 - 36$ weeks gestation) to be associated with increased risk of asthma. However, the true biological critical windows are unlikely to perfectly align with predefined windows such as clinically defined trimesters or developmentally identified windows. In contrast to the lung development example provided here, there are not clearly defined developmental stages for many endpoints. Moreover, methods that rely on exposure averaged over predefined time periods such as trimesters or the entire pregnancy are likely to mischaracterize or completely miss critical windows (Wilson et al., 2017b).

Advances in exposure assessment have allowed for more flexible models that leverage high resolution exposure data to better identify windows and estimate the exposure-time-response function (Wright, 2017; Buckley et al., 2019). Here, high resolution refers to weekly or daily exposure measurement taken in close proximity to the mother's residence. These data allow for more precise identification of periods during the gestation process when exposure is associated with altered health outcomes for the child. Distributed lag models (DLMs), which are more flexible and al-

low the data to determine the location of the critical windows can more precisely identify critical windows and more completely characterize the exposure-time-response relation.

In this dissertation, we consider additional complexities around estimating perinatal critical windows. First we address is the assumption that pollution effects are linear in the exposure concentration. Second, we consider the effect of simultaneous exposure to multiple pollutants, which may create a more realistic picture of the exposure-response relationship (Hoskovec et al., 2021; Davalos et al., 2017). For multiple pollutants, critical windows may exist as interactions across time, where an early exposure to one pollutant increases susceptibility to another exposure at a later time point. Finally, individual differences may alter exposure susceptibility during gestation. We consider critical window and exposure effect heterogeneity due to a range of the characteristics of the mother and child.

1.2 Distributed Lag Models

A key challenge in estimating critical windows is dealing with correlation in high temporal resolution exposure data. A commonly applied method to identify perinatal critical windows and estimate the exposure-response relationship between maternal exposures and an outcome is the DLM. Wilson et al. (2017b) showed that a constrained DLM outperforms more naive methods such as using average exposure over each of the trimesters because DLMs adjust for exposures at other time points throughout pregnancy and provide a data driven approach to identify critical windows even when they do not align with clinically defined trimesters.

Figure 1.1 presents an example of the correlation across weekly $PM_{2.5}$ exposures measurements taken during gestation for a Colorado birth cohort. Correlations in this example range from -0.04 to 0.42. We note higher correlations at nearby weeks, but also increased correlation at around 24 weeks due to seasonal trends in pollution.

The single exposure DLM involves estimating the effects of a time-resolved vector of exposures to an individual pollutant on a scalar health endpoint. For a sample $i = 1, \ldots, n$, let y_i denote a continuous response, $\mathbf{x}_i = [x_{i1}, \ldots, x_{iT}]'$ represent a vector of exposures measurements taken at

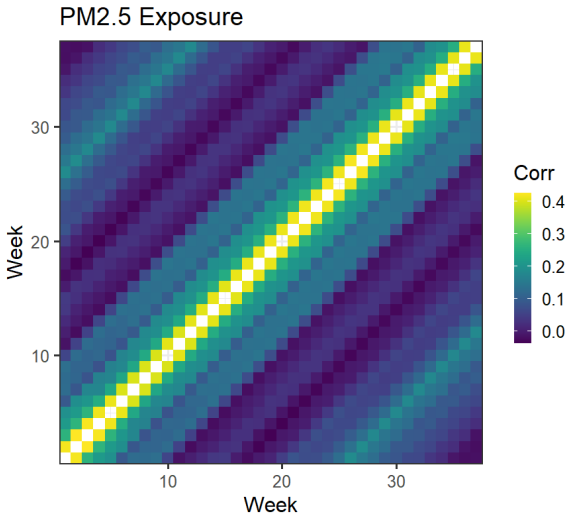


Figure 1.1: Correlation of weekly $PM_{2.5}$ exposures from a Colorado birth cohort. The color represents the correlation between exposures at the weeks given by the x- and y-coordinates with lighter color indicating stronger positive correlation.

equally spaced times $t \in \{1, \ldots, T\}$, and $\mathbf{z}_i$ represent a vector of covariates including model intercept. The single exposure DLM is

$$y_i = \sum_{t=1}^{T} x_{it}\theta_t + \mathbf{z}_i'\boldsymbol{\gamma} + \epsilon_i. \tag{1.1}$$

Here, θ_t is the linear effect of exposure at time t; $\boldsymbol{\gamma}$ is a vector of regression coefficients; and ϵ_i represents independent errors distributed $\mathcal{N}(0, \sigma^2)$. In the context of exposure during pregnancy, as used throughout this dissertation, we consider weekly exposure measurements for the first $T = 37$ weeks of gestation. We limit analyses to full term births, which we define as at least 37 weeks of gestation at birth, and make note of the limitation in our analyses that stratifying by time of birth may introduce potential collider bias (Hernán et al., 2014). The distributed lag effects, θ_t, correspond to each measurement period. An example is visualized in Figure 1.2. Time points when the confidence interval of θ_t does not include zero designates a critical window and identifying these are of key scientific interest. To account for the temporal correlation in the exposure data, θ_t is constrained to vary smoothly in time. Methods for estimating DLMs include splines (Zanobetti et al.,

2000; Gasparrini et al., 2010), Gaussian processes (Warren et al., 2012), and principal components (Wilson et al., 2017a).

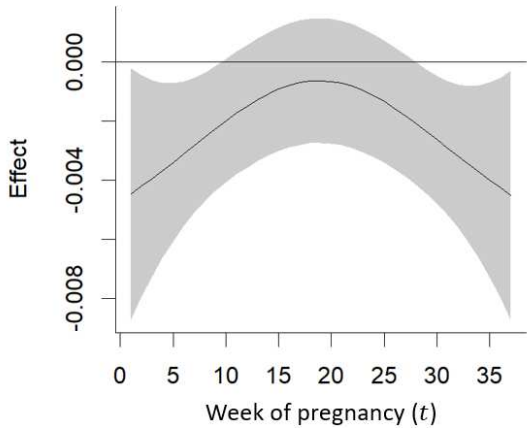


Figure 1.2: Visualization of a distributed lag model. The solid line describe the mean effect at each week and the grey area captures the 95% credible interval.

The distributed lag nonlinear model (DLNM) is an extension of the DLM that allows for non-linear effects in exposure-concentration at each time point (Gasparrini et al., 2010, 2017). The single exposure DLNM is written

$$y_i = \sum_{t=1}^{T} w(x_{it}, t) + \mathbf{z}_i'\boldsymbol{\gamma} + \epsilon_i. \tag{1.2}$$

The exposure-response function $w(x_{it}, t)$ relates the exposure concentration x_{it} for individual i at week t of gestation to the outcome. Methods for estimating a DLNM typically represent w as a bivariate basis expansion in the exposure concentration and time dimensions. Estimation with splines or penalized spline regression allows for a range of assumptions to be made regarding the structure of the exposure-time-response function.

A common issue with current smoothing methods for estimating DLMs and DLNMs, such as splines, is that estimates may generalize critical windows beyond the true period of effect (Figure 1.3). Critical windows are hypothesized to be defined by biological mechanisms during a developmental process that may be altered by environmental exposures. Methods that can identify the

discrete time spans of these events will be more precise in estimating critical windows. To improve critical window identification for DLMs, Warren et al. (2020a) proposed a hierarchical Bayesian framework for DLMs that performs variables selection to select weeks in or out of the critical windows. No similar approach has been proposed to avoid oversmoothing in DLNMs. In Chapter 2 of this dissertation we propose a method that relaxes the smoothness constraint for a DLNM and show that it outperforms spline-based methods and reduces overgeneralizing the critical windows.

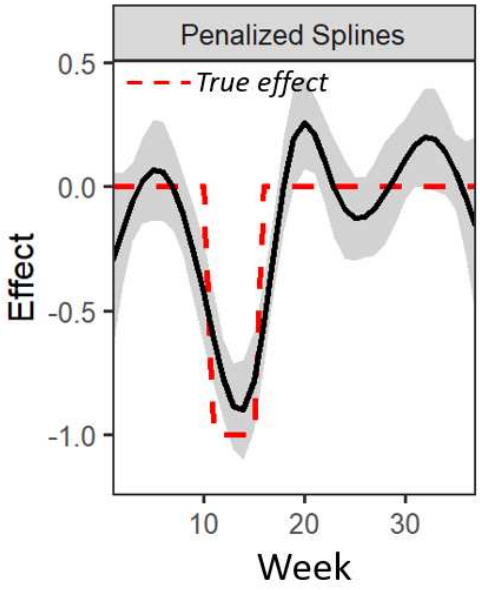


Figure 1.3: Penalized spline estimation (solid line) and confidence interval (grey area) of distributed lag effect, compared to true distributed lag effect (dashed line). Critical windows are identified when confidence interval does not contain zero.

The DLM allows for estimation of the exposure-time-response for a single environmental exposure. While DLMs can be used additively to account for multiple pollutants, this approach leaves out possible interactions between these exposures. Interactions between exposures may occur at the same time period or at different time periods where early exposure to one pollutant increases susceptibility to the same or another pollutant at a later time period. Adding flexibility to accommodate time-sensitive interactions poses a challenge of increasing dimensionality in the required parameter space. As an example, for vectors of two time-resolved exposure measurements $\mathbf{x}_{i1}$ and

$\mathbf{x}_{i2}$, the two-exposure DLM with interactions is

$$y_i = \sum_{t=1}^{T} x_{it1}\theta_{t1} + \sum_{t=1}^{T} x_{it2}\theta_{t2} + \sum_{t_1=1}^{T}\sum_{t_2=1}^{T} x_{it_1 1}x_{it_2 2}\theta_{t_1 t_2} + \mathbf{z}_i'\boldsymbol{\gamma} + \epsilon_i. \tag{1.3}$$

Here, θ_{t1} and θ_{t2} parameterize the main effects of exposures 1 and 2, respectively, while $\theta_{t_1 t_2}$ parameterizes the interactions over time. Chen et al. (2019) proposed a method for estimating the joint effects and interactions of two exposures observed at high temporal resolution, but their method does not extend to mixtures of three or more pollutants. Chapter 3 of this dissertation defines a generalized distributed lag mixture model (DLMM), which considers a larger number of exposures and their interactions across time. We propose a method for estimating the DLMM that performs exposure component selection and effect shrinkage. We show in simulation that this method can precisely identify critical windows and outperforms single exposure methods in a single exposure setting while allowing for additional potential exposures in the data.

There is evidence that exposure effects, including critical window timing and effect magnitude, vary across a population. Several methods have sought to estimate DLMs that allow for effect heterogeneity. Wilson et al. (2017a) proposed a Bayesian distributed lag interaction model (BDLIM) to estimate differences in the exposure-response function for a set of predetermined subgroups. Warren et al. (2020b) developed a spatially-varying DLM to account for changes in pollution composition or demographics over a study region. Both papers highlight the bias incurred due to an assumption of homogeneous effects across a population when the effect are truly heterogeneous. BDLIM has been applied in several epidemiological analyses, including two that found an increased risk of asthma due to prenatal exposures to both fine particulate matter and nitrate for a subgroup of boys concurrently exposed to high prenatal stress (Lee et al., 2018; Bose et al., 2017). Despite these advances in estimating DLM heterogeneity, the methods of Wilson et al. (2017a) and Warren et al. (2020b) remain limited in their scope and interpretability. A spatially-varying DLM is unable to identify the specific features associated with changes in the underlying distributed lag function. BDLIM is limited to predefined categorical subgroups. In an age of big data, we have

access to a potentially high dimensional set of categorical and continuous modifiers where the true modifiers responsible for differences in distributed lag effects are unknown. Harnessing this information in precision models can lead to personalized environmental health decision making. In Chapter 4, we define the heterogeneous DLM (HDLM) and propose three methods allowing for estimation of individualized distributed effects and critical windows. We show in simulation that our proposed HDLM approach precisely identifies critical windows for a individual, estimates the distributed lag effects and selects the correct modifiers responsible for changes in the DLM structure and effects.

1.3 Bayesian Additive Regression Trees

In this dissertation, we propose methods for DLMs using Bayesian additive regression trees (BART). BART is popular supervised machine learning method developed by Chipman et al. (2010) for estimating a non-parametric function due to a vector of covariates, $\mathbf{x}_i$. The BART model for an ensemble of A regression trees is written

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{1.4}$$

$$f(\mathbf{x}_i) = \sum_{a=1}^{A} g(\mathbf{x}_i, \mathcal{T}_a). \tag{1.5}$$

The function g is a weak-learner based on tree $\mathcal{T}_a$ and the regression function f is constructed as an additive ensemble of weak-learners, which are each based off unique tree structures $\mathcal{T}_1, \ldots, \mathcal{T}_A$. Priors on the regression trees and tree-specific effects regularize the model by encouraging shallow trees to prevent over fitting.

A regression tree divides the data into subgroups defined by binary splitting rules on internal nodes resulting in a partition of the sample located at tree terminal nodes. Figure 1.4 diagrams a regression tree. Each terminal node receives a partial effect estimate. Specifically, for tree $\mathcal{T}_a$ with

B_a terminal nodes denoted η_{ab},

$$g(\mathbf{x}_i, \mathcal{T}_a) = \mu_{ab} \text{ if } \mathbf{x}_i \in \eta_{ab}. \tag{1.6}$$

For each observation, the effect estimate is determined by the sum of partial effects due to all trees in the ensemble, that is

$$f(\mathbf{x}_i) = \sum_{a=1}^{A} \sum_{b=1}^{B_a} \mu_{ab} \mathbb{I}(\mathbf{x}_i \in \eta_{ab}) \tag{1.7}$$

where $\mathbb{I}$ is the indicator function.

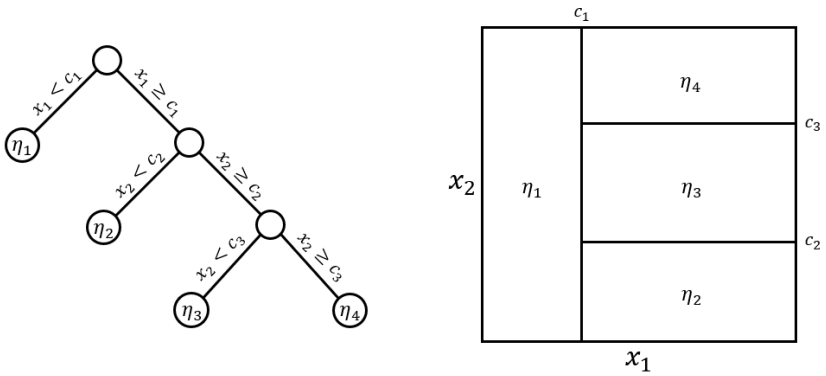


Figure 1.4: A regression tree (left) divides a sample based on two covariates, x_1 and x_2. The terminal nodes, denoted η_b for $b \in \{1, 2, 3, 4\}$, correspond to the resulting partition of the covariate space (right). An observation at a given point in the covariate space will be associated with a single terminal node and terminal node effects.

Estimation of BART proceeds through Bayesian backfitting (Hastie and Tibshirani, 2000) and standard MCMC procedures. Trees are updated via a stochastic process that considers four possible steps: grow, prune, change, and swap. Steps are accepted using the independent Metropolis-Hastings algorithm after integrating over terminal node parameters to deal with changes in dimensionality.

The general BART framework is well documented (Tan and Roy, 2019), shown to have optimal posterior concentration properties (Ročková et al., 2020) and extended for application to a range of different problems. Several notable BART extensions include a method for high dimensional

sets of predictors (Linero and Yang, 2018), causal inference (Hahn et al., 2020), varying coefficient models (Deshpande et al., 2020) and estimation of a smooth function in a single covariate (Starling et al., 2020). The BART framework has also been adapted to accommodate various data types including logistic and count outcomes (Murray, 2020), survival analysis (Sparapani et al., 2016), longitudinal data (Spanbauer and Sparapani, 2021), and heteroscedastic errors (Pratola et al., 2020).

1.4 Outline

This remainder of this dissertation will proceed as follows: In Chapter 2 we develop a novel regression tree-based method for identifying critical windows and incorporating exposure measurement uncertainty in a DLNM. We show our treed DLNM outperforms spline-based methods for estimating distributed lag effects and identifying critical windows. We apply our method to a Colorado-based birth cohort to estimate the association between fine particulate matter exposure and birth weight, and find critical windows throughout pregnancy.

In Chapter 3 we define the distributed lag mixture model and propose a BART method based on an ensemble of tree pairs for estimating the main and interaction effects of environmental mixtures observed at high temporal resolution. Our proposed approach performs variable selection to identify exposures correlated with the outcome and effect shrinkage to reduce the variance of estimated main and interaction effects not associated with the outcome. We apply this method to estimate the relationship between weekly exposures to a mixture of temperature and four ambient pollutants and birth weight. Our analysis finds critical windows due to four exposures and an interaction between fine particulate matter and temperature.

We next propose a method for estimating heterogeneous critical windows and distributed lag effects in Chapter 4. Our approach utilizes regression trees for modification and considers three methods for estimating the DLM. We show our proposed approach can develop individualized distributed lag effect estimates based on a set of modifying covariates and performs well in two simulation scenarios. We apply our heterogeneous DLM to identify susceptible populations in the association between fine particulate matter exposure and birth weight. We consider ten modifying

covariates, including continuous, categorical, and binary modifiers. The analysis suggests that Hispanic designation, mother's age, body mass index, and educational attainment are responsible for changes in the distributed lag effects and critical windows.

Finally, we summarize our work in Chapter 5. We propose several future directions for research in the area of Bayesian treed distributed lag models and describe the potential impact of this work on environmental health epidemiology.

Chapter 2

Treed Distributed Lag Nonlinear Models

2.1 Introduction

In many applications there is interest in regressing an outcome on exposures observed over a previous time window. This frequently arises in environmental epidemiology applications where either a health outcome on one day is regressed on exposures (e.g. temperature or air pollution) observed on that day and several proceeding days or when a birth or children's health outcome is regressed on exposures observed daily or weekly throughout pregnancy (Stieb et al., 2012).

In the context of maternal exposure to air pollution, which we consider in this paper, there are generally two inferential goals. The first is to estimate the critical windows of susceptibility– periods in time during which an exposure can alter a future phenotype. The second goal is to estimate the exposure-time-response function. Recent studies have identified critical windows and associations between maternal exposure to air pollution and several outcomes including preterm birth (Chang et al., 2012, 2015) adiposity (Chiu et al., 2017), asthma and wheeze (Bose et al., 2018; Lee et al., 2018), neurodevelopment (Chiu et al., 2016), among other outcomes (Stieb et al., 2012; Šrám et al., 2005). This includes studies that have found that the linear (Chiu et al., 2017; Chang et al., 2015) and nonlinear (Wu et al., 2018) association vary across weeks of gestation.

A popular approach to estimate the association between maternal exposure to air pollution during pregnancy and a birth outcome is a distributed lag model (DLM) (Schwartz, 2000; Zanobetti et al., 2000). In a DLM, the outcome is regressed on the exposures at each of the time points, e.g. mean exposure during each week of pregnancy. Most commonly, the model is constrained so that the exposure effect varies smoothly over time. Constraining the model adds stability to the estimator in the presence of typically high temporal correlation in the exposure. Methods of regularization include penalized spline regression (Zanobetti et al., 2000), Gaussian processes (Warren et al., 2012), principal components or splines (Wilson et al., 2017a). Wilson et al. (2017b)

showed that a constrained DLM outperforms more naive methods such as using average exposure over each of the trimesters because DLMs adjust for exposures at other time points throughout pregnancy and provide a data driven approach to identify critical windows even when they do not align with clinically defined trimesters.

To extend the DLM to estimate non-linear associations in the exposure-response function at any given time, a class of distributed lag non-linear models (DLNMs) has been proposed (Gasparrini et al., 2010, 2017). DLNM methods typically operate by cross-basis smoothing with splines or penalized spline regression. This results in a unique non-linear exposure-response function at each time point that varies smoothly over the lagged exposures.

A consequence of imposing smoothness over time in a DLM or DLNM is that estimates may generalize the critical window(s) to a wider set of times than is appropriate. Critical windows are hypothesized to be defined by biological events in the fetal developmental process that may be altered by environmental exposures. Methods that can adapt to the discrete time spans of these events are needed to better estimate critical windows. Motivated by this, Warren et al. (2020a) proposed a hierarchical Bayesian framework to improve critical window characterization for DLMs using a variable selection approach that selected weeks in or out of the critical window. However, there are no DLNM methods that relax the smoothness constraint for a non-linear exposure-response function.

In this paper, we propose a method for DLNM that relaxes the smoothness assumption and can more precisely identify critical windows. The proposed approach, which we call treed distributed lag non-linear models (TDLNM), is based on the Bayesian additive regression trees (BART) framework developed by Chipman et al. (2010). Applied to estimating a distributed lag function, TDLNM treats the time series of exposures as a single multivariate predictor and uses a tree structure to partition the exposure concentration and time dimensions to construct a flexible exposure-time-response surface.

We propose two forms of TDLNM. The first form uses a dichotomous tree structure to form a piecewise constant exposure-time-response function. By using an ensemble of trees, the model

can approximate both smooth and non-smooth functions. The second form imposes smoothness only in the exposure-concentration dimension but not over time. This forces smoothness in the exposure-response, while maintaining precision in critical window identification. We also discuss how the smooth version can be used to incorporate exposure uncertainty into the model.

Following development of TDLNM, we perform a simulation study that compares our proposed method to spline-based methods across a variety of settings. These simulations demonstrate that our method excels in the estimation of the exposure-time-response function for non-smooth settings, but also adapts well to estimating scenarios with a smooth exposure-time-response. Importantly, we find that TDLNM more precisely identifies critical windows and has an extremely low rate of critical window misspecification. In addition, simulations show that TDLNM has narrower confidence intervals, especially near the boundaries, while maintaining nominal coverage. Finally, we apply TDLNM to estimating the association between the fine particulate matter ($PM_{2.5}$) experienced by a mother during pregnancy and the resulting birth weight of the child. Software to implement this method is available in the `R` package `dlmtree`.

2.2 Data

We analyze birth records from Colorado, USA, vital statistics data. The data includes live, singleton, full term ($\geq$ 37 weeks gestation) births from Colorado with estimated conception dates between 2007 and 2015, inclusive, with no known birth defects. We limited the analysis data to the northern front range counties (those immediately east of the Rocky Mountains roughy extended from Colorado Springs to the Wyoming border). This area contains the majority of the Colorado population. We further limited the analysis to census tracts with elevation less than 6000 feet above sea level. This both reduces the potential confounding by altitude and the impact of mountainous terrain on exposure data.

The primary outcome of interest in this paper is birth weight for gestational age z-score (BWGAZ). We obtained BWGAZ using the Fenton birth charts (Fenton and Kim, 2013). BWGAZ measures birth weight as the number of standard deviations above or below the expected birth weight of

a child with the same fetal age and sex. The data contain individual level covariate information including mother's age, weight, height, income, education, marital status, prenatal care habits and whether they smoked before or during pregnancy, as well as race and Hispanic designations. We limit the analysis to observations with complete covariate information, resulting in 300,463 births.

We use $PM_{2.5}$ exposure data from the US Environmental Protection Agency fused air quality surface using downscaling data. This data is publicly available at www.epa.gov/hesc/rsig-related-downloadable-data-files. The statistical methodology for construction of the data files has been described in Berrocal et al. (2010). We linked the exposure data to the birth records based on the census tract of maternal residence at birth. We then constructed weekly average exposures for each week of gestation. A map detailing the number of births in each county is shown in Figure A.1.

This study was approved by the Institutional Review Board of Colorado State University.

2.3 Methods

2.3.1 DLNM Framework

Before introducing our proposed method we briefly recap the DLNM framework and standard methodology. Let y_i be the continuous outcome for person i from a sample $i = 1, \ldots, n$. Let $\mathbf{x}_i = [x_{i1}, \ldots, x_{iT}]^T$ denote a vector of exposures observed at equally spaced times $t = 1, \ldots, T$. In our case, y_i indicates BWGAZ while x_{it} represents the i^{th} mother's exposure to $PM_{2.5}$ in week t of pregnancy. We control for a vector of covariates, denoted $\mathbf{z}_i$. The Gaussian DLNM model is

$$y_i \sim \mathcal{N}\left[f(\mathbf{x}_i) + \mathbf{z}_i^T\boldsymbol{\gamma}, \sigma^2\right], \tag{2.1}$$

where $f(\mathbf{x}_i)$ is the distributed lag function and $\boldsymbol{\gamma}$ is a vector of regression coefficients.

The distributed lag function $f(\mathbf{x}_i)$ can take several linear as well as non-linear forms. The DLNM allows a unique non-linear association between exposures at each time point and the out-

come. In general, the distributed lag function is defined as

$$f(\mathbf{x}_i) = f(x_{i1}, \ldots, x_{iT}) = \sum_{t=1}^{T} w(x_{it}, t) \tag{2.2}$$

where $w(x, t)$ is the exposure-response function relating exposure at week t of gestation to the outcome. Existing frameworks for the DLNM (Gasparrini et al., 2010, 2017) utilize a cross-basis where w is represented as a bivariate basis expansion in the exposure concentration and time dimensions. Penalized spline implementations allow for a range of assumptions to be made regarding the structure of the exposure-time-response. For example, varying ridge penalties target shrinkage at specific times, while varying difference penalties control the smoothness along the curve. Basis expansion methods, such as splines, regularize the model to improve stability of the estimated effect in the presence of multicollinearity in the predictor. However, these methods also impose the assumption of smoothness in the DLNM.

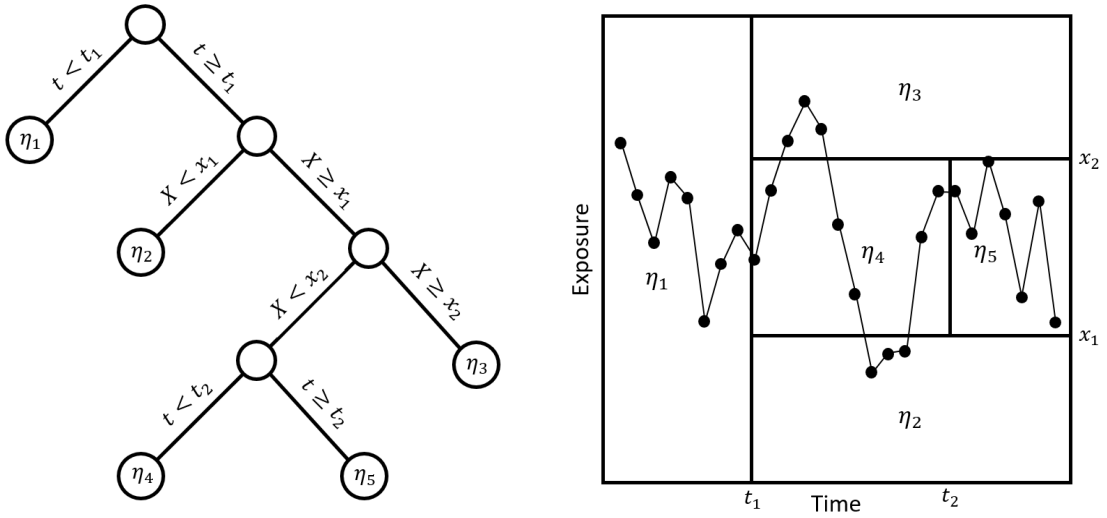
2.3.2 Treed DLNM Approach

We introduce a sum-of-trees model based on the BART framework (Chipman et al., 2010) to estimate the exposure-time-response function, $f(\mathbf{x}_i)$. The general approach is to build dichotomous trees that partition the time-varying exposure $\mathbf{x}_i$ in both the exposure concentration and time dimensions. Figure 2.1 illustrates the approach for a single tree. Figure 2.1(a) is a diagram of a tree showing binary rules defined on the exposure and time values. These rules divide the exposure-time space into five terminal nodes, denoted $\eta_1, \ldots, \eta_5$. Figure 2.1(b) shows the exposure-time space partitioned into five regions with each region corresponding to a single terminal node. A tree and corresponding parameters define a piecewise constant exposure-response function,

$$w(x_{it}, t) = \mu_b \quad \text{if } x_{it} \in \eta_b. \tag{2.3}$$

The distributed lag function for tree $\mathcal{T}$ takes a form similar to that in (2.2) and is defined as

$$g(\mathbf{x}_i, \mathcal{T}) = \sum_{t=1}^{T} w(x_{it}, t). \tag{2.4}$$



(a) (b)

Figure 2.1: Example of tree, $\mathcal{T}$, with terminal nodes η_b, $b \in \{1, 2, 3, 4, 5\}$. Panel (a) diagrams the tree with dichotomous splits on time or exposure concentration while panel (b) represents the resulting partition of the exposure-time space for a single observation.

In our TDLNM framework, we consider an ensemble of A regression trees. For tree $\mathcal{T}_a$, $a \in \{1, \ldots, A\}$, denote the B_a terminal nodes as $\eta_{a1}, \ldots, \eta_{aB_a}$. Each terminal node, η_{ab} has a corresponding set of limits in time and exposure concentration given by the rules defined as splits of the tree and a corresponding parameter μ_{ab}. Collectively, the terminal nodes of each tree define a partition of the exposure-time-space and allow for flexible estimation of the exposure-time surface. As in (2.3), we define the effect of each exposure-time combination in tree $\mathcal{T}_a$ to be $w_a(x_{it}, t) = \mu_{ab}$ if $x_{it} \in \eta_{ab}$. Each regression tree in the ensemble provides a partial estimate of the distributed lag

non-linear function f. Formally, the exposure-time-response function for TDLNM is

$$f(\mathbf{x}_i) = \sum_{a=1}^{A} g(\mathbf{x}_i, \mathcal{T}_a), \tag{2.5}$$

where $g(\mathbf{x}_i, \mathcal{T}_a)$ represents the partial estimate contributed by tree a given in (2.4).

TDLNM foregoes the basis-imposed smoothness assumption. However, when different time and exposure breaks are staggered across trees the ensemble of trees can approximate smooth functions. Model regularization is a result of the tree prior, which prefers trees having only a few splits. Smaller trees ensure that the model is stable in the presence of temporal correlation because each terminal node averages across multiple time points.

2.3.3 Smoothing in exposure concentration

Most epidemiological studies assume that the exposure-response relationship is smooth in exposure concentration. The TDLNM methods presented above assume a piecewise linear structure that can approximate a smooth function, but it is never truly smooth. In this subsection we propose a TDLNM model that is truly smooth in exposure (TDLNMse). Importantly, TDLNMse does not force smoothness in time to allow for accurate critical window estimation.

To allow for smoothing in the exposure-response, we introduce a weight function on the terminal-node specific effects. A similar idea was introduced by Linero and Yang (2018), which assigned a node-specific probability to each observation using a gating function at each dichotomous split on a covariate. TDLNMse differs in that we desire smoothing only in the exposure-concentration dimension. To accomplish this, we define smoothing parameter σ_x and modify (2.3) to be

$$w_a\left(x_{it}, t\right) = \sum_{b=1}^{B_a} \mu_{ab} \cdot \psi(x_{it}; \eta_{ab}, \sigma_x). \tag{2.6}$$

The weight function $\psi(x_{it}; \eta_{ab}, \sigma_x)$ allows each observation x_{it} to be distributed across all terminal nodes that contain time point t. For the weight function we use a normal kernel with bandwidth

σ_x. Hence, the weight for x_{it} assigned to node η_{ab} is

$$\psi(x_{it}; \eta_{ab}, \sigma_x) = \left\{ \Phi\left(\frac{\lceil x_{ab} \rceil - x_{it}}{\sigma_x} \right) - \Phi\left(\frac{\lfloor x_{ab} \rfloor - x_{it}}{\sigma_x} \right) \right\} \cdot \mathbb{I}(t \in \eta_{ab}), \tag{2.7}$$

where $\lceil x_{ab} \rceil$ and $\lfloor x_{ab} \rfloor$ refers to the maximum and minimum exposure concentration limits of node η_{ab}, respectively, and Φ is the standard normal cumulative distribution function. The inclusion of the indicator function allows TDLNMse to retain a piecewise constant effect in time at each exposure concentration value. The kernel smoother requires fewer terminal nodes to estimate a smooth effect in exposure-concentration as observations near the boundary of two terminal nodes will have an estimated effect that is in between estimates of observations located centrally in the nodes.

For TDLNMse, we propose fixing the bandwidth σ_x a priori. Alternatively, we could assign a prior to σ_x and estimate the bandwidth.

2.3.4 Incorporating exposure uncertainty with TDLNMse

A situation that has not been addressed in the DLNM literature is uncertainty in the exposure. Many exposure models including climate models and spatially kriged exposure models provide measurement of uncertainty. Most commonly these occur in one of two forms—standard errors for the exposure data or multiple realizations from a model such as draws from a posterior predictive distribution or an ensemble method. This uncertainty is not accommodated for in the health effect estimates from standard DLNMs.

Exposure uncertainty can be incorporated into TDLNM by using a weight function to spread the exposure across multiple terminal nodes according to the probability that the exposures is in each of those nodes. The result is similar to TDLNMse, using a weight function corresponding to the uncertainty in each observation. In the case of reported standard errors for the exposure data, we use (2.7) with observation specific smoothing parameter σ_{xi} that is equal to the standard error for each observation. If instead we have multiple draws of exposures from an ensemble or Bayesian model we replace $\mathbf{\Phi}$ in (2.7) with the empirical cumulative distribution function.

2.3.5 Interpretation of TDLNM and relation to spline-based DLNMs

To gain some insight into the exposure-response function characterized by TDLNM we consider the DLNM relation at a single time point. The distributed lag function in TDLNM is given by combining (2.4) and (2.5), i.e. the sum over trees and the sum of each tree over time. Reversing the order of the summation we get

$$f(\mathbf{x}_i) = \sum_{t=1}^{T} \sum_{a=1}^{A} w_a(x_{it}, t). \tag{2.8}$$

At time t, the exposure-response function, $\sum_{a=1}^{A} w_a(x_{it}, t)$, is equivalent to the BART model with univariate predictor x_{it}. In the case of TDLNM this implies a piecewise-constant exposure-response function across the exposure concentration levels at time t. For TDLNMse, the weight function ψ acts as a linear smoother over exposure concentration for the exposure-response function at each time.

2.4 Prior Specification and Computation

Our prior specification is based on that of Chipman et al. (2010); however, some modifications and a different MCMC algorithm are needed to accommodate or improve performance with the multivariate predictor and parametric control for covariates. In this section we specify key differences in the priors and computation approach from that of Chipman et al. (2010), including a horseshoe-like shrinkage prior on tree-specific effects and an altered prior for tree splits on a multivariate predictor. Full details on the priors and computation are in Appendix A.2.

2.4.1 Prior Specification

We apply a tree-specific, horseshoe-like prior to the effects at the terminal nodes μ_{ab} (Carvalho et al., 2010). The prior for terminal node b on tree a is

$$\mu_{ab} | \sigma^2, \omega^2, \tau_a^2 \sim \mathcal{N}\left(0, \sigma^2 \omega^2 \tau_a^2\right). \tag{2.9}$$

Here $\tau_a \sim \mathcal{C}^+(0,1)$ and $\omega \sim \mathcal{C}^+(0,1)$ define the horseshoe prior on trees. We specify prior $\sigma \sim \mathcal{C}^+(0,1)$ and $\boldsymbol{\gamma} \sim \mathcal{MVN}(\mathbf{0}, \sigma^2 c\mathbf{I})$, where c is fixed at a large value.

For the half-Cauchy priors on all variance parameters we adopt the hierarchical framework of Makalic and Schmidt (2015), where $r^2|s \sim \mathcal{IG}(1/2, 1/s)$ and $s \sim \mathcal{IG}(1/2, 1)$ gives that marginally $r \sim \mathcal{C}^+(0,1)$. This allows for Gibbs sampling of all variance components.

The tree-specific shrinkage prior on μ_{ab} results in better mixing throughout the MCMC sampler. This occurs by allowing shrunken trees with a small variance component and smaller effects $\mu_{a1}, \ldots, \mu_{aB_a}$ to more easily explore splitting locations in the exposure-time space. After reconfiguration, these trees have the ability to contribute larger partial estimates.

Our stochastic tree generating process largely follows Chipman et al. (1998). The probability a tree splits at node η with depth d_η equals $p_{\text{split}}(\eta) = \alpha(1+d_\eta)^{-\beta}$, where hyperpriors $\alpha \in (0,1)$ and $\beta \geq 0$. In our data setting the number of potential splits in the time direction is $T-1$ while the number of potential split points in the exposure direction is equal to the number of unique exposure values minus one which is substantially larger than $T-1$. To address this imbalance we limit the potential exposure split points a priori and propose an alternative prior on potential split points. By limiting the exposure split points we also avoid situations where a split in one dimension limits future splits in another dimension due to empty nodes. For example, if TDLNM has a tree that splits on an extreme value in the exposure-concentration dimension, it may be unable to further split on the time dimension due to lack of data in one partition of the exposure-time space. We restrict the potential split locations in the exposure dimension to a predefined set of quantiles or values. Specification of potential splitting values also improves computational efficiency by allowing for precalculation of counts or weights for the limited number of potential splits.

We assign prior probabilities uniformly across potential time splits and uniformly across potential exposure splits such that there is a 0.5 probability of selecting either a time or exposure split as the first splitting rule in a tree. Hence, for s_x and s_t total potential exposure and time splitting values, respectively, the probability of the first split in a tree being on exposure equals $1/(2s_x)$ or $1/(2s_t)$ if the split is on time. For a splitting decision further down the tree, the splitting

rule probability is proportional to the probabilities of the potential remaining splits in a selected node. Following a split in time, there are fewer potential remaining splits in time, increasing the probability that the next split will take place in the exposure dimension.

2.4.2 MCMC Sampler

We estimate TDLNM using MCMC. The MCMC approaches used for BART do not apply to the current model for two reasons. First, the algorithm of Chipman et al. (2010) relied on the fact that any specific vector of predictors $\mathbf{x}_i$ is contained in a single terminal node on each tree, whereas TDLNM divides the exposures related to each observation across the terminal nodes. Second, we modify the algorithm to allow for parametric control of the confounding variables $\mathbf{z}$. Due to these differences we propose an alternative MCMC approach for the TDLNM model. In particular, we integrate out $\boldsymbol{\gamma}$ using standard analytical techniques. Then, we apply Bayesian backfitting (Hastie and Tibshirani, 2000) to simultaneously estimate the effects of the partial exposure-time-response based on the partition defined by each tree, $\mathcal{T}_a$. Full details of the MCMC sampler can be found in Appendix A.2.

2.4.3 Hyperprior selection and tuning

Tree splitting hyperpriors were set to the defaults used in Chipman et al. (2010) with $\alpha = 0.95$ and $\beta = 2$; different settings did not improve results. Trees in TDLNM explore only two dimensions, which requires fewer trees to adequately explore the predictor space. In preliminary work, we found that 10 to 20 trees was sufficient. Results did not change using more than 20 trees. We used $A = 20$ trees for our simulation and data analysis. We assigned the stochastic tree process to grow or prune with probability 0.3 each and change with probability 0.4. The fixed smoothing parameter in TDLNMse, σ_x, is data dependent: too large and the estimated effect will appear linear, too small and the model reverts to TDLNM (no smooth effect). For our simulation and data analysis, we set σ_x to half the standard deviation of the exposure data. We found this setting to balance a smooth effect while also clarifying non-linearity in the exposure-concentration effects.

2.4.4 Estimating the exposure-time-response function

The distributed lag non-linear function f includes the model intercept. To ease interpretation we remove the intercept by centering f at a reference exposure value, x_0, at each time. As f is estimated as a sum of tree, we center each tree at the reference value and we use the centered trees for posterior inference.

2.5 Simulation

We conduct a simulation study to compare the empirical performance of TDLNM and TDLNMse to established DLNM methods that use penalized and unpenalized splines. Key to the simulation is that we compare performance on simulation scenarios representing both smooth and non-smooth exposure-time-response functions.

We simulate data according to (2.1) and (2.2), using a sample size equivalent to our exposure data ($n = 300,463$). To accurately represent the autocorrelation found in air pollution exposure data, we use $PM_{2.5}$ exposures from our data analysis. We simulate the DLNM surface using 37 consecutive weeks from each observation to simulate a full-term pregnancy. We consider four simulated exposure-time-response functions. Each corresponds to a different true model (TDLNM, TDLNMse, smooth DLNM with splines, and linear DLM). The four DLNM scenarios are: A) piecewise constant effect in exposure across weeks $11-15$; B) linear effect in exposure across weeks $11-15$; C) smooth, non-linear effect (logistic shape) in exposure across weeks $11-15$; D) smooth, non-linear effect (logistic shape) in exposure with a smooth effect in time peaking at week 13 and extending approximately five weeks in either direction. We generate the outcomes using log-transformed exposure data. All scenarios are centered at log-exposure value 1. Several cross-sections of the exposure-time surfaces are shown in Figures 2.2 and 2.3. Algebraic details of the DLNM surface for each scenario and a graphic representation can be found in Appendix A.3.1.

We generate a set of covariates (five standard normal, five binomial with probability 0.5) and corresponding coefficients from standard normal. We include a seasonal trend by using ozone data. Specifically, we add a random ozone effect for every 5^{th} week (5, 10, ..., 35), where ozone

measurement at each time is centered mean zero, scaled to have standard deviation one and multiplied against a draw from $\mathcal{N}(0, \sigma^2 = 0.04)$. This allows for a different seasonal trend for each simulated dataset that is correlated with both the exposure, $PM_{2.5}$, and the outcome. We set the error variance σ^2 such that $\text{Var}[f(\mathbf{x}_i)]/\sigma^2 = 1/1000$ to represent a realistic signal to noise ratio and run 500 simulation replicates in each scenario. The simulation design can be reproduced with the `R` package `dlmtree`.

2.5.1 Simulation estimators and comparisons

TDLNM and TDLNMse used the prior settings described in section 2.4.3. Thirty evenly spaced values ranging between the 0.01 percentile and the 99.9 percentile of all log-exposure value were designated as potential splits in the exposure-dimension. After a burn-in period of 5,000 iterations, we ran each model for 15,000 iterations, thinning to every tenth draw.

We compare TDLNM and TDLNMse to several spline-based penalized and unpenalized DLNM models. The models are described as follows with more detail given in Gasparrini et al. (2017).

- GAM: base model defined by penalized cubic B-spline smoothers of rank 10 in both exposure and time dimensions, with second-order penalties, estimated with REML;
- DLM: using GAM with a linear assumption in exposure concentration;
- GLM-AIC: optimal number of unpenalized, quadratic B-splines in both exposure and time dimensions (df 1 to 10) selected by minimizing AIC;
- GAMcr: defined by replacing the cubic B-spline basis in GAM with cubic regression splines and penalties on the second derivatives;
- GAM-exp: GAM, replacing the second-order penalties with a varying ridge penalty.

To assess model performance, we center the DLNM for each model at log-exposure value 1 and evaluate the estimated DLNM over a grid of points.

In each model we include all 10 simulated covariates as well as indicators for year and month to control for the additional seasonal trend. We log-transform the exposure concentration values

to reduce skew in the exposure data and allow for equally spaced knots in the spline basis models. The decision to log-transform the response has no impact on TDLNM as the model will produce identical results with or without a log-transform; it does impact the smoothing with TDLNMse.

2.5.2 Simulation Results

Summary measures of model performance are shown in Table 2.1.

Here, we compare each model by the root mean square error (RMSE) of the entire exposure-time surface and broken down to the RMSE within and outside the simulated critical windows. We also show the empirical coverage of 95% confidence intervals along with average confidence interval width. In addition, the models are compared on the probability of identifying a non-zero effect across grid points inside the simulated critical window (TP), the probability of incorrectly placing a non-zero effect across grid points outside the simulated critical window (FP), and the precision of correct identification of a non-zero effect: TP/(TP+FP). We designate a non-zero effect in the true exposure-time surface as any effect outside of the interval from -0.005 to 0.005 to account for scenario B and C, which have a non-zero effect everywhere between weeks 11-15 and scenario D, which has a non-zero effect everywhere. Figures 2.2 and 2.3 show cross-sections of the exposure-time-response surface using estimates from models TDLNM, TDLNMse, and GAMcr. A non-zero estimate in the plots indicates a change in the response for any observation with that particular time and exposure-concentration value.

TDLNM and TDLNMse have as good or better overall RMSE in scenarios A, C and D. In all scenarios, the tree-based methods have the lowest RMSE in areas of zero effect in the exposure-time-response surface. Figure 2.2 highlights the ability of TDLNM and TDLNMse to find a sharp distinction between times with or without effects. The shrinkage prior on the tree-specific parameters reduces variance leading to lower RMSE in areas of no effect. In areas of non-zero effect, our models have lower RMSE than spline based models in scenario A and are comparable in scenarios C and D. In scenario B the RMSE in areas of non-zero effect is higher for TDLNM and TDLNMse, as the spline based models do a better job interpolating into the extreme exposure values

Table 2.1: Simulation results, showing RMSE for estimation of the exposure-time-surface with no-effect and effect separated. Coverage and CI width is based on 95% confidence intervals. Effect identification considers the likelihood of identifying a non-zero effect (TP) or incorrectly designating a non-zero effect (FP) over the DLNM surface. Precision is calculated within each simulation as TP/(TP+FP). Standard errors are available in Table A.2.

	DLNM RMSE			DLNM Coverage		Effect Identification		
Model	Overall	No Effect	Effect	Overall	CI Width	TP	FP	Precision
Scenario A: Piecewise in Exposure and Time								
TDLNM	0.086	0.066	0.213	1.00	0.43	0.87	0.00	1.00
TDLNMse	0.100	0.077	0.252	0.99	0.46	0.82	0.00	0.98
GAM	0.294	0.258	0.584	0.95	1.08	0.47	0.03	0.90
DLM	0.370	0.342	0.626	0.68	0.53	1.00	0.30	0.77
GLM-AIC	1.531	1.536	1.462	0.84	3.35	0.49	0.15	0.55
GAMcr	0.263	0.241	0.454	0.98	1.10	0.62	0.01	0.96
GAM-exp	0.241	0.165	0.669	0.94	0.67	0.32	0.01	0.87
Scenario B: Linear in Exposure								
TDLNM	0.292	0.081	0.768	0.87	0.37	0.56	0.01	0.99
TDLNMse	0.270	0.073	0.712	0.87	0.34	0.64	0.01	0.99
GAM	0.312	0.257	0.547	0.73	0.48	0.90	0.18	0.84
DLM	0.299	0.257	0.489	0.64	0.36	1.00	0.26	0.79
GLM-AIC	0.267	0.253	0.346	0.79	0.46	0.99	0.18	0.85
GAMcr	0.248	0.206	0.426	0.84	0.54	0.87	0.09	0.90
GAM-exp	0.283	0.226	0.518	0.76	0.37	0.94	0.15	0.86
Scenario C: Smooth in Exposure								
TDLNM	0.077	0.033	0.223	0.94	0.18	0.58	0.01	0.99
TDLNMse	0.070	0.032	0.201	0.97	0.17	0.67	0.01	0.99
GAM	0.142	0.126	0.241	0.91	0.36	0.60	0.06	0.91
DLM	0.138	0.120	0.245	0.64	0.18	1.00	0.31	0.77
GLM-AIC	0.186	0.167	0.309	0.82	0.40	0.53	0.14	0.80
GAMcr	0.113	0.104	0.176	0.95	0.37	0.64	0.03	0.96
GAM-exp	0.126	0.103	0.255	0.92	0.28	0.62	0.05	0.93
Scenario D: Smooth in Exposure and Time								
TDLNM	0.105	0.041	0.203	0.80	0.26	0.40	0.00	0.99
TDLNMse	0.098	0.038	0.190	0.95	0.24	0.45	0.01	0.99
GAM	0.120	0.100	0.171	0.97	0.44	0.54	0.01	0.98
DLM	0.122	0.090	0.193	0.69	0.23	0.94	0.23	0.80
GLM-AIC	0.284	0.277	0.306	0.81	0.52	0.45	0.14	0.77
GAMcr	0.110	0.092	0.156	0.97	0.41	0.57	0.01	0.98
GAM-exp	0.099	0.068	0.164	0.97	0.35	0.47	0.00	0.99

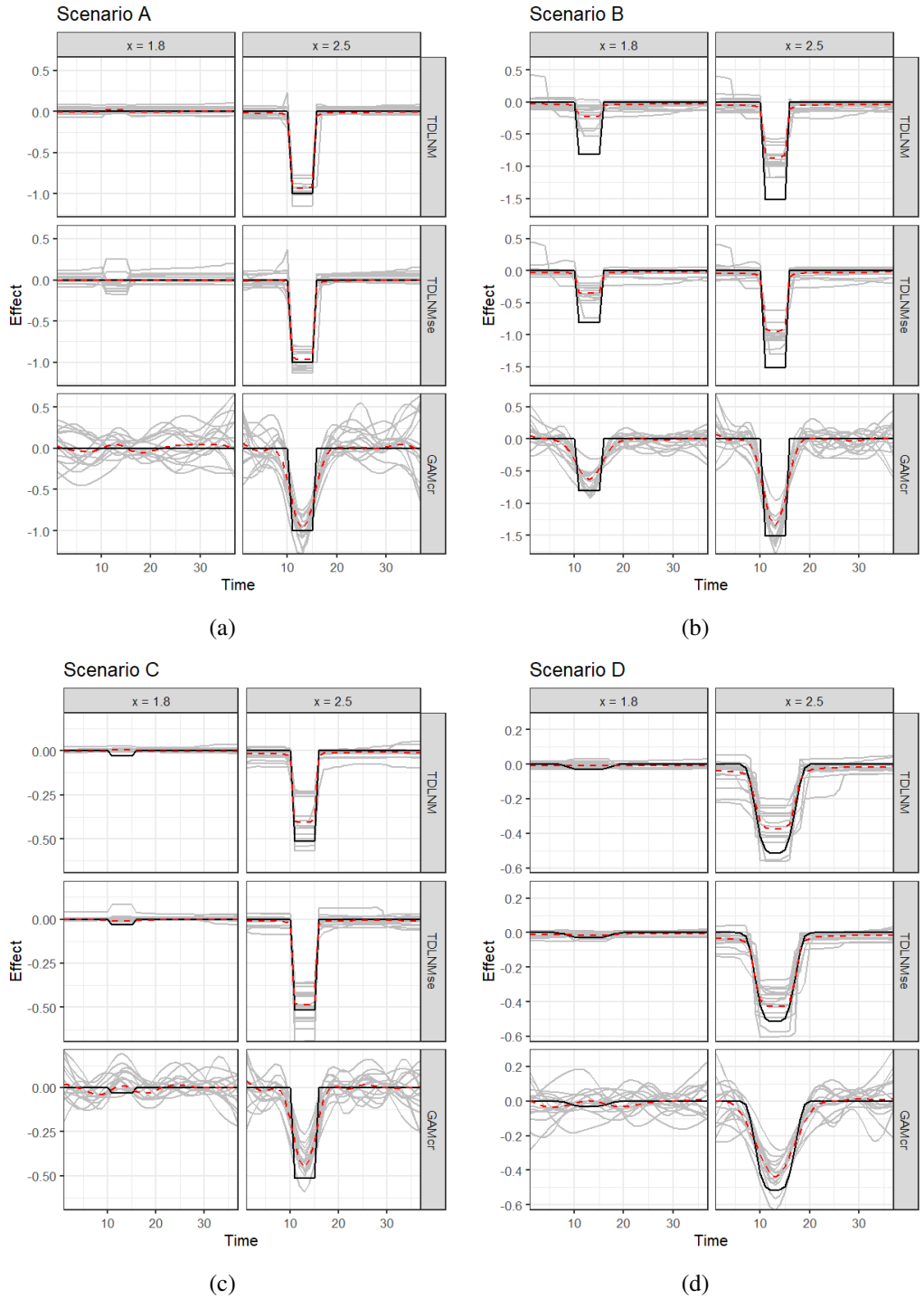


Figure 2.2: Simulation results of TDLNM, TDLNMse, and GAMcr comparing the exposure-time-response (y-axis) cross-section across all times (x-axis) fixed at two different exposure concentrations. Grey lines show 15 random simulation replicates, red dashed line indicates average across all simulations and solid black lines indicates the true simulated response.

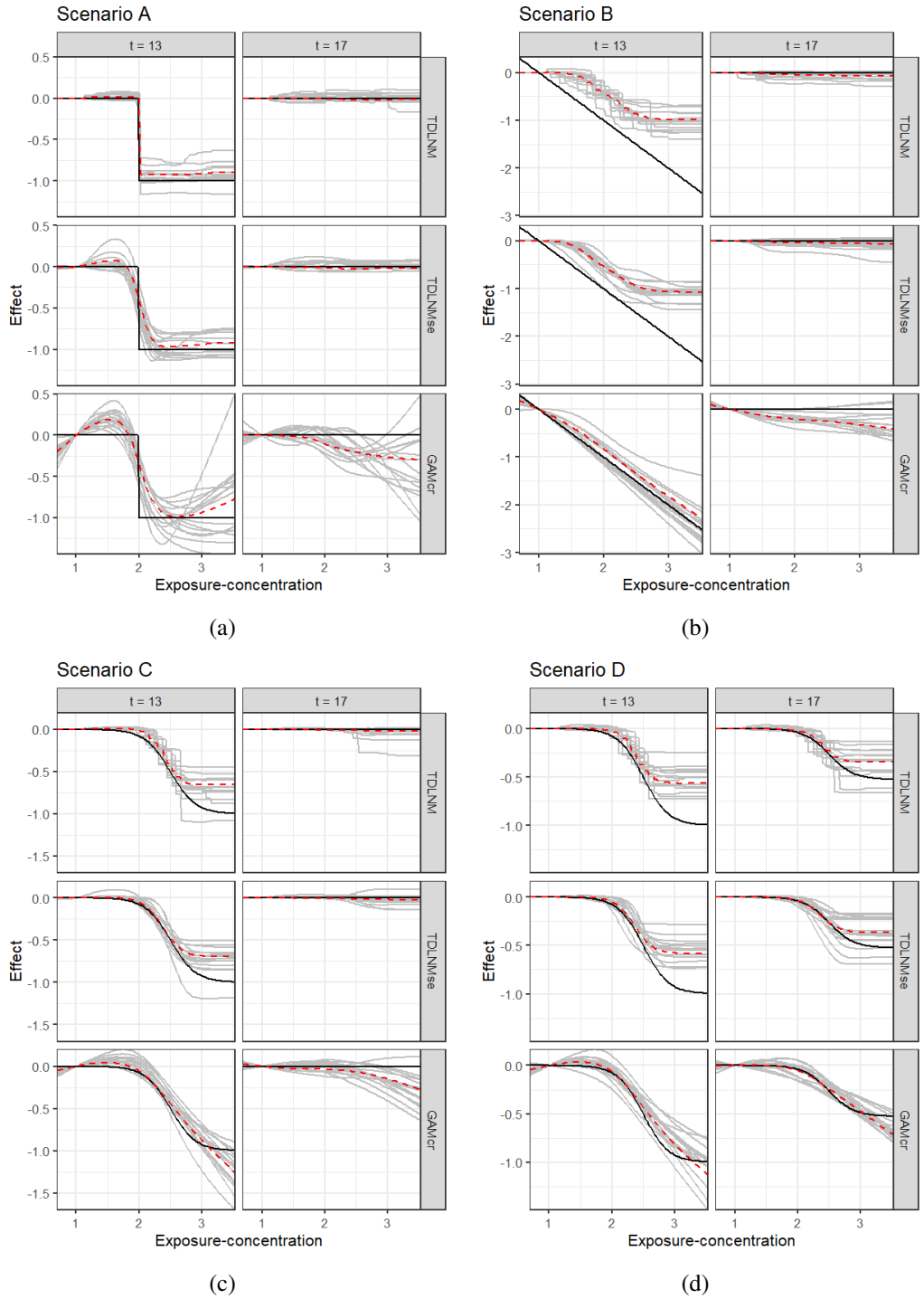


Figure 2.3: Simulation results of TDLNM, TDLNMse, and GAMcr comparing the exposure-time-response (y-axis) cross-section across all exposure concentrations (x-axis) fixed at two different times. Grey lines show 15 random simulation replicates, red dashed line indicates average across all simulations and solid black lines indicates the true simulated response.

where few data points reside. Figure 2.3 contrasts how tree-based models attenuate the effect at the boundaries of exposure values, while GAMcr continues the trend linearly.

The tree-based models have near nominal coverage, except in scenario B. All models show below nominal coverage in scenario B, however, TDLNM and TDLNMse perform best, each having 87% surface coverage. In addition, our models have the smallest average confidence interval width, which is particularly notable at the boundaries in time or extreme exposure concentration where the 'wiggliness' of spline-based models becomes more pronounced (Figures 2.2 and 2.3). The lack of 'wiggliness' in the tree-based model estimates contributes to narrow confidence intervals as well as decreased RMSE, especially in areas of zero-effect. Furthermore, the variation between simulation replicates is much smaller for TDLNM and TDLNMse.

Scenario B, while seemingly natural for a DLM, poses several difficult situations. First, a proper estimate by TDLNM would require trees with many breaks spanning the exposure concentration during the correct critical window. Second, TDLNM attenuates the effect when data is sparse (e.g. high and low concentrations in this scenario). Third, at high concentration, there is a jump from zero to a large effect that smooth methods cannot accommodate; in particular, DLM extends the critical window well beyond the true period of effect as a result of the smoothness assumption.

Precision with TDLNM and TDLNMse is the highest across all simulation scenarios (Table 2.1). The high precision is a result of near zero FP, but with a tradeoff of lower TP in scenarios B and D. The cross-sectional plots in Figure 2.2 shows the ability of TDLNM and TDLNMse to adapt to non-smooth exposure-time response surfaces. Figure A.3 indicates the probability detecting a non-zero effect in at least one exposure value in each week. These results shows that the spline-based methods have a much higher probability of misclassifying weeks just outside of the true critical windows. On the other hand, the tree-based models adapt to changing smoothness in the exposure-time-response surface and rarely detect non-zero effects outside of the true critical window. The key takeaway is that the critical windows detected by TDLNM and TDLNMse have a high probably of being correct.

2.6 Data Analysis

We use TDLNM and TDLNMse to estimate the relationship between a mother's exposure to $PM_{2.5}$ during the first 37 weeks of pregnancy and child BWGAZ. By using weekly exposures, we limit the temporal resolution at which critical windows can be identified with any method to correspond to weeks. For comparison, we also apply DLNM using penalized cubic regression splines (GAMcr) and DLM. We control for maternal baseline characteristics and season and long-term trends. The maternal characteristics are: pre-pregnancy age (quadratic fit), weight, smoking (if done before or during pregnancy), income, education, prenatal care (when first received), race and Hispanic designations, elevation, and county of residence. We do not control for fetal sex or gestational age as the outcome, BWGAZ, is already adjusted for these factors. In addition, we adjust for seasonal effects using indicators for year and month of conception.

For TDLNM and TDLNMse, we use the same hyperparameters as in our simulation, running the models for a burn-in period of $5,000$ iterations followed by $15,000$ iterations retaining every tenth draw from our MCMC sampler. We specify 30 equally spaced potential splits in the exposure dimension ranging from the 0.1 percentile to the 99.9 percentile of log-exposure values. Different numbers of potential splits were considered, but showed no differences in the result. In TDLNMse we set the smoothing parameter σ_x equal to half the standard deviation of the log-exposures. Models GAMcr and DLM used the same settings as in simulation. The DLNM estimates for all models are centered at the median exposure value (approximately 7 μg/m^3). Critical windows are defined as any week containing a region in the exposure-time-response where the 95% confidence interval does not contain zero.

2.6.1 DLNM Results

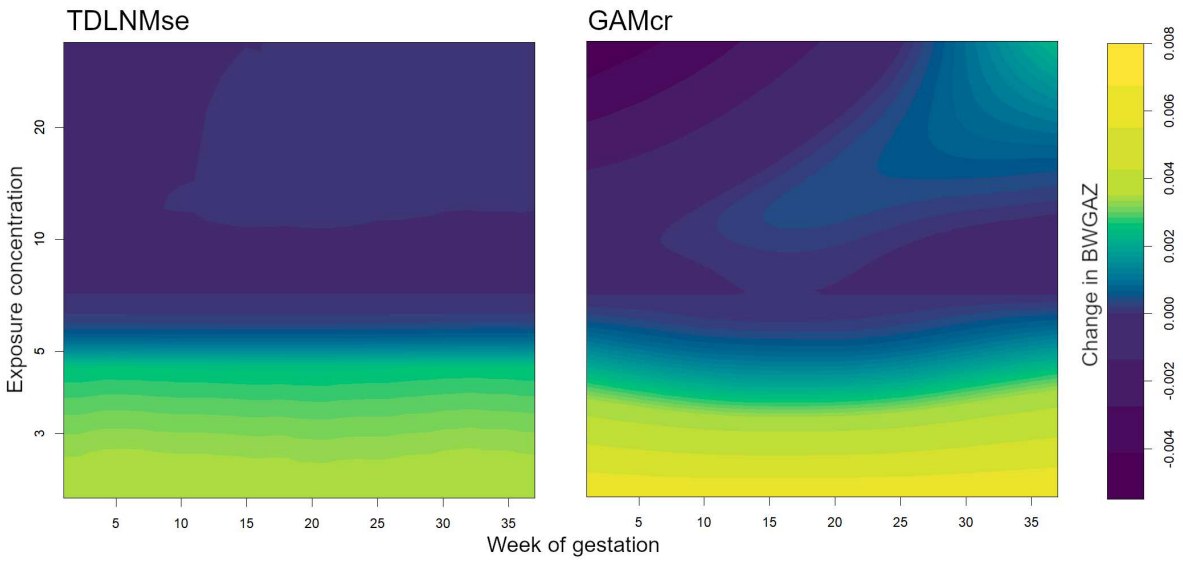
The posterior mean exposure-time-response estimates for TDLNMse is shown in Figure 2.4(a). $PM_{2.5}$ exposure below the median is associated with an increase in BWGAZ. Exposure concentration above the median value indicate a slight decrease in BWGAZ, but the 95% credible intervals do not give reason to believe this is different from zero. This pattern is present across all gesta-

tional weeks. Cross-sections of the exposure-time-response surface at weeks 5, 15, 25, and 35 are shown in Figure 2.4(b) and indicate a critical window spanning the entire pregnancy.

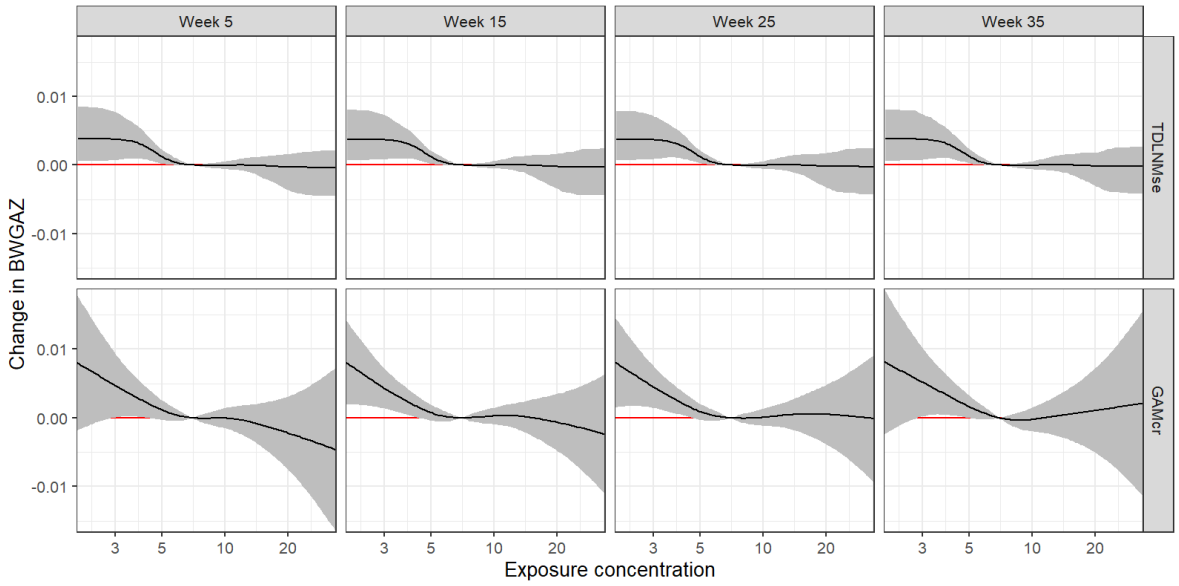
Based on TDLNMse, a change from median (7.0 $\mu g/m^3$) to the 25th percentile of $PM_{2.5}$ exposure (5.89 $\mu g/m^3$) across the pregnancy would result in a cumulative mean increase in BWGAZ of 0.0132 (95% CI: $0.0003, 0.0354$) or approximately 5.74g (95% CI: $0.11, 15.41$) when translated to actual birth weight (this is approximate because BWGAZ accounts for gestational age and fetal sex). The non-linear association shows that a further decrease in $PM_{2.5}$ exposure to the 10th percentile (5.02 $\mu g/m^3$) would result in a 0.055 (95% CI: $0.016, 0.090$) mean increase in BWGAZ, or an approximate increase of 24.1g (95% CI: $7.155, 39.10$). These results suggest that decreasing $PM_{2.5}$ below the current national ambient air quality standards would result in higher average birth weights in this population.

The mean exposure-time-response estimate for GAMcr, shown in Figure 2.4(a), closely resembles the estimates of TDLNMse. As in our simulations, we see a difference in the tail behavior. GAMcr continues the trend in the effect with large intervals. Despite the large point estimate with GAMcr at low exposure levels the larger confidence intervals include zero. In contrast, TDLNMse tapers off and estimates a smaller effect with substantially smaller intervals that do not contain zero. The smaller intervals found in TDLNMse near the boundaries are a result of these boundary regions being grouped in terminal nodes that also contain internal regions and therefore receive the same estimates.

Our findings of an association between increased $PM_{2.5}$ and decreased BWGAZ are consistent with previous literature. A meta-analysis by Sun et al. (2016) found a 10 μg/m^3 increase in $PM_{2.5}$ across pregnancy to be associated with 15.9g decrease in birth weight (95% CI: $-26.8, -5$); increased exposures in the second and third trimesters were also determined to have a nonzero negative association with birth weight. Zhu et al. (2015) reported similar results in a separate meta-analysis. Strickland et al. (2019) found that the magnitude of associations between $PM_{2.5}$ and birth weight increased for higher percentiles of the birth weight distribution across all trimesters. Finally, a study investigating individual chemical components of $PM_{2.5}$ found non-zero increased



(a) Exposure-time-response surface



(b) Exposure-response function at weeks 5, 15, 25, and 35

Figure 2.4: Panel (a) shows the estimated exposure-time-response surface for models TDLNMse and GAMcr. Panel (b) shows cross sections of the estimated exposure-time-response with columns showing the estimated effect at four times: $t = 5, 15, 25, 35$, while rows compare models TDLNMse and GAMcr. All plots indicate the exposure-time-response relative to the median exposure concentration value ($7.0 \mu g/m^3$).

risk of low birth weight for maternal exposures during each trimester of pregnancy (Ebisu and Bell, 2012).

2.6.2 Comparing less flexible model alternatives

For comparison, we fit TDLNM, a DLM and several linear models to compare results. Each of these models was consistent with the TDLNMse results. More details on these methods can be found in Appendix A.4.

2.7 Discussion

In this work we have proposed a tree-based method for a DLNM to estimate the association between a time-resolved series of pollution exposures and a continuous birth outcome. TDLNM eliminates the smoothness assumption in the exposure-time response surface. TDLNMse imposes smoothness only in the exposure-concentration dimension but not over time. TDLNM also has the potential to account for measurement error within the exposure-response function. By relaxing the smoothness assumption in the time dimension, our new methods more precisely identify critical windows of susceptibility.

TDLNM provides several extensions to tree-based regression models. First, we allow for a multivariate predictor with temporal correlation. Second, we provide a computationally efficient method for estimating a tree-based function while controlling for a fixed effect. Finally, we eliminate the need for cross-validation to select variance hyperpriors through the application of a horseshoe-like prior on tree-specific effects.

In simulation scenarios, we show that TDLNM and TDLNMse have a low false positive rate of critical window identification, while spline-based DLNMs have a tendency to over-generalization the time periods containing critical windows. Furthermore, our tree-based methods can approximate both smooth and non-smooth exposure-time-response functions. As the smoothness assumption in time changes, TDLNM and TDLNMse allow for information sharing at the same exposure levels across time, so that the piecewise constant steps are distributed across adjacent times allow-

ing for near-smooth estimates. The shrinkage priors reduce the variance of estimates, reducing RMSE in areas of no effect and decreasing the rate of false positives. In the presence of a linear trend, DLNM models are overly flexible. While penalized spline DLNM can revert to an approximately linear model, TDLNM requires a large number of splits in the exposure-concentration dimension to accomplish the same results. As seen in simulation Scenario B, TDLNMse attenuated the linear trend in areas of few exposure observations. The simulations indicate that TDLNM and TDLNMse have high precision in identifying critical windows.

We applied TDLNM and TDLNMse to a Colorado birth cohort. We found a non-linear effect of $PM_{2.5}$ on BWGAZ. Specifically, we found that below median levels of $PM_{2.5}$ throughout pregnancy were associated with higher BWGAZ. We found no change in BWGAZ due to above median $PM_{2.5}$ exposure.

Chapter 3

Estimating Perinatal Critical Windows to Environmental Mixtures via Structured Bayesian Regression Tree Pairs

3.1 Introduction

Maternal exposure to environmental chemicals during pregnancy is an important public health concern due to potential impact on children's health. Increased exposure to environmental chemicals has been linked to decreased birth weight, increased risk of asthma, and altered neurological development, among other outcomes (Bosetti et al., 2010; Stieb et al., 2012; Jacobs et al., 2017). There is also evidence that changes in temperature are associated with birth weight (Jakpor et al., 2020; Kloog et al., 2015). Recent research has focused on leveraging exposure data that is observed at high temporal resolution throughout pregnancy to identify critical windows of susceptibility during the gestational process (Wright, 2017). Critical windows are periods in time when an exposure can alter a future health outcome and could be as short as a week or span many months. Estimating the exposure-response relationship over the course of pregnancy is an equally important goal. Most studies that leverage repeated measurements of exposure consider the effects of only a single environmental chemical. While understanding the effects of a single exposure is essential, considering a mixture is necessary to develop a more realistic picture of the exposure-response relationship because it accounts for interactions and controls for confounding by co-exposures (Davalos et al., 2017). In this paper, *mixture* refers to simultaneous exposure to multiple environmental factors. As an example of the importance of considering interactions and mixtures, Anenberg et al. (2020) documented evidence of a synergistic effect of heat and air pollution on cardiovascular and respiratory disease among adults. However, estimating critical windows for mixtures has been elusive due to an absence of statistical methods.

A distributed lag model (DLM) is commonly used to estimate the association between a single time-resolved exposure and a health outcome (Zanobetti et al., 2000; Warren et al., 2012; Wilson et al., 2017a; Gasparrini et al., 2017). A DLM regresses the outcome on the exposure measurements at multiple time points, e.g. regressing birth weight on weekly mean exposure during gestation. Because of high temporal correlation between repeated measures of exposure there is a need to regularize the DLM. This is most commonly done with a constrained DLM that smooths the effect over the exposure time and allows for sharing of information between parameters at different times during gestation. Common constraints include splines (Zanobetti et al., 2000; Gasparrini et al., 2017), Gaussian processes (Warren et al., 2012) and principal components (Wilson et al., 2017a). Compared to using average exposure over pregnancy or each of the trimesters, the DLM has been shown to reduce bias in estimates as well as improve critical window estimation (Wilson et al., 2017b). To account for multiple exposures, the DLM can be used additively and extended to include interactions between two time-resolved predictors (Chen et al., 2019; Muggeo, 2007). However, methods to identify critical windows and estimate exposure-time-response functions with mixtures of more than two time-resolved predictors are lacking.

Estimating DLMs with interactions requires flexibility to identify interactions between exposures across time. Interactions at different time points correspond to the 'priming' hypothesis. Priming posits that prenatal exposure leads to phenotype changes (Bolton et al., 2014). These changes may result in altered biological mechanisms that increase susceptibility to later exposures. Extending DLMs to include time-sensitive interactions poses a challenge of increasing dimensionality in the required parameter space. As the number of exposures and temporal resolution of measurements increase, the number of possible interactions increases at a quadratic rate. In addition, data for mixtures observed at multiple time points typically exhibit high collinearity both over repeated measures and across mixture components at a single time point. A key challenge is, therefore, regularizing the model while allowing for flexibility to identify critical windows.

We propose a regression tree approach to estimate a constrained DLM for a single exposure or mixture of exposures observed at multiple time points. Regression trees have been applied in

numerous fields including the study of chemical mixtures observed at a single time point in environmental epidemiology (Park et al., 2017). Bayesian additive regression trees (BART), introduced by Chipman et al. (2010), has been adapted for a wide variety of data generating situations such as high dimensional prediction (Linero and Yang, 2018) or for causal inference (Hahn et al., 2020). While BART is generally focused on out-of-sample prediction, our goal is to adapt this framework to the estimation of distributed lag effects. Applied to exposure observations taken repeatedly over time, current regression tree techniques are lacking in several respects. First, the regression trees would treat the measurements for a single exposure at adjacent times as independent predictors. This is equivalent to fitting an unconstrained DLM which is unstable due to high collinearity. Second, regression trees would not account for the structure in mixture data where one measurement from each exposure is taken at the same time point. To account for temporal ordering in a single time-resolved predictor, Mork and Wilson (2021b) proposed a tree based model to estimate a distributed lag nonlinear model that subdivides the time and exposure-concentration dimensions of the exposure-time response surface, but fails to generalize to mixtures.

In this paper, we define the distributed lag mixture model (DLMM), which extends the DLM to estimate the main effects of multiple exposures along with all two-way interactions. We propose regression-tree based methods for estimating both DLM and DLMM. Our DLM method uses regression trees that split on time to construct a constrained DLM that is piecewise constant with breakpoints determined by binary splits in the tree. Combined in an additive ensemble of trees, the resulting DLM can be approximately smooth or have piecewise constant distributed lag effects. Our proposed DLMM further builds on the tree literature. We introduce the concept of tree pairs—two trees that collectively define DLM structures of main effects and interactions between two time-resolved exposure measurements. We develop a computational framework to estimate an additive ensemble of tree pairs that allows for both the tree structures and the exposures to which the tree structures are applied to be learned from the data. Furthermore, our method conducts exposure selection and effect shrinkage to remove time-resolved predictors or interactions that do not influence the outcome.

We apply our models to a Colorado-based administrative birth cohort. This analysis investigates changes to birth weight associated with five environmental exposures measured weekly during gestation. Software is made available in the `R` package `dlmtree`.

3.2 Colorado Birth Cohort Data

We analyze birth weight for gestational age z-score, BWGAZ, using birth vital statistics records from Colorado, USA. The data include all births from Colorado with estimated conception dates between 2007 and 2015, inclusive. We limit the analysis to the Denver metropolitan area. Besides birth outcomes, the data include individual covariates including mother's age, weight, height, income, education, marital status, prenatal care habits, smoking habits, as well as race and Hispanic indicators.

We are interested in the association between birth weight and a mother's weekly exposure to particulate matter smaller than 2.5 microns in diameter ($PM_{2.5}$), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and temperature. We limited our analysis to singleton, full-term births ($\geq$ 37 weeks) and observations with complete covariate and exposure data, resulting in 195,701 births. This study was approved by the Institutional Review Board of Colorado State University. Demographic breakdowns and additional data details are described in Appendix B.1.

3.3 Model

3.3.1 Distributed lag mixture models

For a sample $i = 1, \ldots, n$, let y_i denote a continuous response, $\mathbf{x}_i = [x_{i1}, \ldots, x_{iT}]'$ represent a vector of exposure measurements taken at equally spaced times $t \in \{1, \ldots, T\}$, and $\mathbf{z}_i$ represent a vector of covariates including model intercept. The single exposure DLM is

$$y_i = \sum_{t=1}^{T} x_{it}\theta_t + \mathbf{z}_i'\boldsymbol{\gamma} + \epsilon_i. \tag{3.1}$$

In (3.1), θ_t is the linear effect of exposure at time t; $\boldsymbol{\gamma}$ is a vector of regression coefficients; and ϵ_i represents independent errors distributed $\mathcal{N}(0, \sigma^2)$.

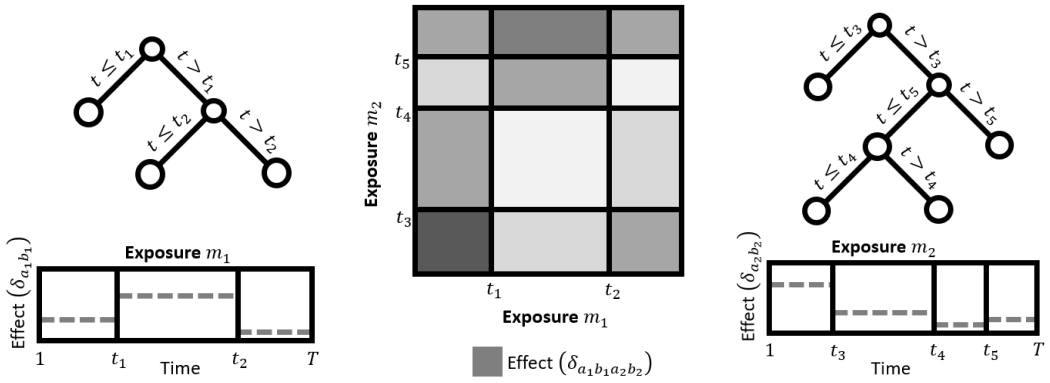
We consider a model involving M exposures. Let $\mathbf{x}_{im} = [x_{im1}, \ldots, x_{imT}]'$ represent the vector of measurements for exposure m corresponding to individual i. A DLMM with pairwise interactions can be written

$$y_i = \sum_{m=1}^{M}\sum_{t=1}^{T} x_{imt}\theta_{mt} + \sum_{m_1=1}^{M}\sum_{m_2=m_1}^{M}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} x_{im_1t_1}x_{im_2t_2}\theta_{m_1m_2t_1t_2} + \mathbf{z}_i'\boldsymbol{\gamma} + \epsilon_i. \tag{3.2}$$

Here, θ_{mt} is the main effect of exposure to pollutant m at time t. Interactions are considered at every time combination t_1 for exposure m_1 with t_2 for exposure m_2 and parameterized by $\theta_{m_1m_2t_1t_2}$. This includes interactions within the same exposure. Within-exposure interactions at $t_1 = t_2$ represents a quadratic main effect. In total, the DLMM requires $MT + \binom{M+1}{2}T^2$ parameters, which quickly becomes a 'large-p' problem as the number of exposures grows. For example, the DLMM in our data analysis involving 5 exposures and 37 time points requires estimating 20,720 parameters.

3.3.2 Treed DLM

We first introduce our proposed method of estimating a DLM for a single exposure with no interactions. In our treed distributed lag model (TDLM), binary trees partition the entire exposure time span, T, into non-overlapping segments. Figure 3.1(a) illustrates the approach. Each binary tree is characterized by a set of internal nodes that split on available time points and a set of terminal nodes, which are the endpoints of the tree structure. The terminal nodes define the time partition and are denoted η_b for $b \in \{1, \ldots, B\}$. We define the linear relationship between each of the exposure measurements in terminal node η_b and the outcome by a single coefficient, denoted δ_b. The distributed lag effects in (3.1) are, therefore, defined by the terminal nodes and node specific effects as $\theta_t = \delta_b$ if $t \in \eta_b$. This represents a piecewise constant DLM such that all times within the same terminal node have the same effect on the outcome. In our proposed model, both the tree structure and the effects are learned from the data.



(a) (b) (c)

Figure 3.1: Diagram of structured regression trees. Binary tree structures in panels (a) and (c) partition the time period for an individual exposure and each terminal node $\eta_{a_i b_j}$ corresponds to an estimated effect $\delta_{a_i b_j}$ (dashed lines). The interaction surface in panel (b) depicts how the tree structures intersect; the intersection of terminal nodes $\eta_{a_1 b_1}$ in panel (a) and $\eta_{a_2 b_2}$ in panel (c) has corresponding effect $\delta_{a_1 b_1 a_2 b_2}$ (with independent magnitude indicated by shading).

By keeping the number of terminal nodes small, TDLM introduces a necessary constraint on the distributed lag function to account for temporal correlation by assuming that the effect of exposures within the same terminal node are equal. This framework also allows for sharp changes in effect estimates across times at the boundary between different terminal nodes where the effect estimates can quickly change in magnitude or sign.

We propose an additive model with an ensemble of A regression trees. This allows for smoothness in the DLM as each tree may partition the time span of exposure differently. For tree $\mathcal{T}_a$, $a \in \{1, \ldots, A\}$, denote terminal nodes $\{\eta_{ab}\}_{b=1}^{B_a}$ with corresponding effects $\{\delta_{ab}\}_{b=1}^{B_a}$. The distributed lag effect $\theta_t = \sum_{a=1}^{A} \sum_{b=1}^{B_a} \delta_{ab} \mathbb{I}(t \in \eta_{ab})$, where $\mathbb{I}(\cdot)$ is the indicator function.

3.3.3 Treed DLMM

The treed distributed lag mixture model (TDLMM) extends TDLM to multiple time-resolved predictors by replacing each individual tree with a pair of trees. A pair of trees is composed of two binary trees, as described in Section 3.3.2, which define the main effects of two exposures and the interaction between these same exposures. The interactions are structured based on the time partitions of each tree. Exposures in each time-segment of the first tree are interacted with exposures

in each time-segment of the second tree. A diagram representing the structured regression trees and interaction surface is shown in Figure 3.1. The exposures partitioned by a tree pair may be two different exposures or the same exposure twice. In the case both trees partition the same exposure, we are able to estimate within-exposure interactions as well as potential nonlinear effects via quadratic terms created by same-time interactions. This maintains a hierarchical structure where an interaction is only included when the corresponding main effects are part of the model. Both the structures of the trees and exposures are learned in our proposed model.

In TDLMM we use an ensemble of A tree pairs. The ensemble allows for different pairs of exposures to be included in the model, which correspond to the main effects and interactions that may be present. Consider tree-pair $\{\mathcal{T}_{a_1}, \mathcal{T}_{a_2}\}$ with sets of terminal nodes $\{\eta_{a_1 b}\}_{b=1}^{B_{a_1}}$ and $\{\eta_{a_2 b}\}_{b=1}^{B_{a_2}}$, respectively. Let $S_{a_1} = m$ if exposure m is partitioned by tree $\mathcal{T}_{a_1}$. Similarly $S_{a_2} = m$ if exposure m is partitioned by tree $\mathcal{T}_{a_2}$. The main effect of exposure m at time t is

$$\theta_{mt} = \sum_{a=1}^{A} \left[\sum_{b=1}^{B_{a_1}} \delta_{a_1 b_1} \mathbb{I}(S_{a_1} = m, t \in \eta_{a_1 b_1}) + \sum_{b=1}^{B_{a_2}} \delta_{a_2 b_2} \mathbb{I}(S_{a_2} = m, t \in \eta_{a_2 b_2}) \right] \tag{3.3}$$

and the interaction between exposure m_1 and m_2 at times t_1 and t_2 is

$$\theta_{m_1 m_2 t_1 t_2} = \sum_{a=1}^{A} \sum_{b_1=1}^{B_{a_1}} \sum_{b_2=1}^{B_{a_2}} \delta_{a_1 b_1 a_2 b_2} \mathbb{I}(S_{a_1} = m_1, S_{a_2} = m_2, t_1 \in \eta_{a_1 b_1}, t_2 \in \eta_{a_2 b_2}). \tag{3.4}$$

The main effects for terminal nodes $\eta_{a_1 b_1}$ and $\eta_{a_2 b_2}$ are parameterized by $\delta_{a_1 b_1}$ and $\delta_{a_2 b_2}$, respectively. The interaction effect between exposures in terminal nodes $\eta_{a_1 b_1}$ and $\eta_{a_2 b_2}$ is given by $\delta_{a_1 b_1 a_2 b_2}$.

TDLMM can be reduced into two simpler models. The first drops within-exposure interaction by fixing these interactions to zero. The second simplification drops all interactions which is equivalent to an additive DLM. In the remainder of this paper we refer to these simplifications as TDLMMns (no-self interactions) and TDLMMadd (additive).

3.4 Prior Specification and Computation

3.4.1 TDLM priors and posterior computation

The prior for TDLM consists of two parts: a prior on trees and a prior on the regression parameters conditional on the trees. We apply a normal prior to the regression parameters:

$$\delta_{ab}|\tau_a^2, \nu^2, \sigma^2 \sim \mathcal{N}(0, \tau_a^2\nu^2\sigma^2). \tag{3.5}$$

Here, $\tau_a \sim \mathcal{C}^+(0,1)$ and $\nu \sim \mathcal{C}^+(0,1)$ define a horseshoe-like prior on tree-specific effects (Carvalho et al., 2010). This is a global-local shrinkage prior that performs shrinkage overall and at the tree level. The additional tree-specific variance component τ_a allows for the effects on poor fitting trees to be shrunk. When shrunk, the trees can more easily reconfigure before regaining larger terminal node effects. In practice, we find including this tree-specific variance results in more precise estimates of the distributed lag function. We specify global prior $\sigma \sim \mathcal{C}^+(0,1)$ and $\boldsymbol{\gamma} \sim MVN(\mathbf{0}, \sigma^2 c\mathbf{I})$, where c is fixed at a large value.

The prior on trees is a stochastic tree generating process based off the approach of Chipman et al. (1998). Complete details are given in Appendix B.2.

3.4.2 TDLMM priors and posterior computation

For TDLMM with multiple predictors and tree pairs the prior involves three components: 1) the prior on trees; 2) the prior on node effects; and 3) the prior on which exposures appear in each tree pair. One goal of TDLMM is to shrink or remove exposure and interaction effects that are uncorrelated with the response. We specify a prior that uses two approaches to achieve this goal. First, we add a prior on node specific effects that allows for effects of unique exposures and interactions to be shrunk. Second, we specify a prior on which exposures are included into each of the A tree pairs. This avoids having to pre-specify pairs of exposures into each tree pair and allows for variable selection because an exposure that is not included into any tree pair is selected out of the model.

In TDLMM each of the trees in tree-pair a is defined by a tree structure $\mathcal{T}_{a_i}$, $i \in \{1, 2\}$, and an exposure S_{a_i} that the tree structure is applied to. The prior on $\mathcal{T}_{a_i}$ is the same as described in Section 3.4.1. The prior distribution on exposure S_{a_i} is

$$S_{a_i}|\mathcal{E} \sim \text{Categorical}(\mathcal{E}), \quad \mathcal{E} \sim \text{Dirichlet}(\kappa, \ldots, \kappa). \tag{3.6}$$

Here, $\mathcal{E} = [E_1, \ldots, E_M]$, where E_m is the probability that a tree splits on exposure m and κ is a hyperprior that controls the sparsity of exposures. This prior is motivated by Linero and Yang (2018) but differs in that we select an exposure for the entire tree, while the former selects a variable to split on at a particular node of a tree. Details on setting κ and a Bayes factor approach for exposure selection are give in Appendix B.3.

Both $\mathcal{T}_{a_i}$ and S_{a_i} are updated via Markov chain Monte Carlo (MCMC). New proposals for the structure of each tree in a tree-pair are the same as in TDLM. We also introduce a new proposal mechanism that switches the exposure, S_{a_i}, considered by a tree. For each tree, we select with equal probability one of four different proposals: grow, prune, change, and switch-exposure. When switch-exposure is the update step we propose a new exposure S'_{a_i} from (3.6). The decision to accept any of the four possible moves is made with the Metropolis-Hastings algorithm. Alternatively, the exposure, S_{a_i}, can be updated with Gibbs sampler but at a high computational cost.

The node-specific priors described in (3.5) shrinks tree-specific effects, but would apply the same variance parameters for all exposures. We introduce an alternative variance component pertaining specifically to each exposure or pair of two exposures. The prior on the node-specific effect, $\delta_{a_i b}$, at terminal node $\eta_{a_i b}$, which splits exposure S_{a_i}, is defined as

$$\delta_{ab}|\mu^2_{S_{a_i}}, \nu^2, \sigma^2 \sim \mathcal{N}(0, \mu^2_{S_{a_i}} \nu^2 \sigma^2). \tag{3.7}$$

Here, $\mu_{S_{a_i}}$ is a local variance prior corresponding to the exposure, S_{a_i}, being used by tree $\mathcal{T}_{a_i}$. The prior for an interaction effect between terminal nodes $\eta_{a_1 b_1}$ and $\eta_{a_2 b_2}$ that split exposures S_{a_1} and

S_{a_2}, respectively, is

$$\delta_{a_1 b_1 a_2 b_2} | \mu^2_{S_{a_1} S_{a_2}}, \nu^2, \sigma^2 \sim \mathcal{N}(0, \mu^2_{S_{a_1} S_{a_2}} \nu^2 \sigma^2). \tag{3.8}$$

Here, $\mu_m \sim \mathcal{C}^+(0,1)$ and $\mu_{m_1 m_2} \sim \mathcal{C}^+(0,1)$ represent the variance components for the main effects of exposure m or the interaction of exposures m_1 and m_2, respectively. Because the same variance components are used in every tree or tree-pair that estimates the same exposure main or interaction effects, it allows the model to separate shrinkage on the main and interaction effects. Priors on $\boldsymbol{\gamma}, \nu$, and σ are the same as in TDLM. Full details on the fitting of TDLMM are given in Appendix B.2.

Selection and shrinkage are complementary in TDLMM. Shrinkage can apply to either the interaction effect, main effect, or both. Selection removes exposures from the model entirely.

3.4.3 Marginal DLM effects with TDLMM

Due to the interactions between exposures in TDLMM, the effects of each exposure are dependent on the levels of the other exposures. To estimate the main effect of each exposure m on the outcome while accounting for co-exposures, we marginalize the DLMM at specified levels of all exposures. Fixing the levels of all exposures to be $\widetilde{x}_1, \ldots, \widetilde{x}_M$, the marginal effect of exposure m at time t is defined as

$$\widetilde{\theta}_{mt}(\widetilde{x}_1, \ldots, \widetilde{x}_M) = \theta_{mt} + \sum_{m'=1}^{m} \sum_{t'=1}^{T} \widetilde{x}_{m'} \theta_{m'mt't} + \sum_{m'=m}^{M} \sum_{t'=1}^{T} \widetilde{x}_{m'} \theta_{mm'tt'}. \tag{3.9}$$

Marginalization by integrating out other exposure effects amounts to evaluating (3.9) at the empirical mean of each exposure. Because the within-exposure interaction represents a possible nonlinear effect, this can cause interpretation problems when calculating $\widetilde{\theta}_{mt}$ (a linear effect). In this case, estimating a contrast, such as effect due to an inter-quartile range (IQR) change in exposure, is a more reasonable approach.

3.4.4 Logistic regression

We propose a logistic regression method for TDLM and TDLMM. Details are available in Appendix B.2.5.

3.4.5 Prior selection

For tree structure parameters we follow Chipman et al. (2010) setting $\alpha = 0.95$ and $\beta = 2$. Altering these parameters did not improve performance. For TDLM we set $A = 20$ trees. For TDLMM we set A and κ to achieve a desired prior inclusion probability for exposure m. Settings $A = 20$ and $\kappa = 1.089$ gives prior inclusion probability of 0.9. Additional details are provided in Appendix B.3.

3.5 Simulation

We considered two simulation scenarios. The first validated TDLM and compared to established DLM methods. The second evaluated TDLMM for a mixture of five exposures. In Appendix B.4 we provide additional simulations to justify shrinkage components of our models. All simulations can be reproduced with `R` package `dlmtree`.

3.5.1 Single exposure and binary outcome

We first considered a binary outcome with a single exposure ($PM_{2.5}$) generated by

$$y_i|p_i \sim \text{Bernoulli}(p_i), \quad \text{logit}(p_i) = c_1 + 0.1 \cdot [f_1(\mathbf{x}_i) + \mathbf{z}_i^T\boldsymbol{\gamma}], \tag{3.10}$$

with simulated DLM effect $f_1(\mathbf{x}_i) = \sum_{t=s}^{s+7} x_{it}$. This defines a DLM such that $\theta_t = 0.1$ from times s to $s+7$ and zero otherwise. Starting time s was drawn uniformly from $\{1, \ldots, T-7\}$. We set c_1 for each simulation replicate such that the mean of p_i, denoted $\overline{p}$, was approximately 0.5 or 0.05. In Appendix B.4.3 we replicated this simulation using a smooth DLM effect.

We simulated 100 data sets of sample size $n =$ 5,000. For each data set we sampled exposure observations from the cohort described in Section 3.2 with lengths $T = 37$, and centered and

scaled all exposure data. We also generated ten covariates (five standard normal, five binomial with probability 0.5) and corresponding regression coefficients from standard normal.

In this scenario, we compared single exposure models TDLM to penalized spline DLM (Gasparrini et al., 2017) with cubic regression splines and the critical window variable selection (CWVS) method (Warren et al., 2020a). In addition, we compared to TDLMM with all 5 exposures from our data analysis. For TDLMM, we included time-resolved measurements of four additional exposures that were not included in any other model and have no effect on the outcome in this design. Using the multi-exposure model in a single exposure setting explored the loss of efficiency that results when only one of five correlated predictors affects the outcome. The results for CWVS are based on the DLM estimation when a critical window is identified (DLM|cw) and when the probability of a critical window is greater than 0.5 (Pr(cw)$>$.5), as described by Warren et al. (2020a). MCMC chains were run for 10,000 iterations after a burn-in period of 5,000 and thinned to every 5th iteration.

The objective of our method is to estimate the distributed lag effect and identify critical windows. We therefore focus our simulation on estimation of these quantities rather than predictive performance, which is the focus of many BART models.

Results from scenario one are given in Table 3.1. We calculated the marginal DLM root mean square error (RMSE) $= \sqrt{\sum_{t=1}^{37}(\widetilde{\theta}_t - \widehat{\widetilde{\theta}}_t)^2/37}$ and coverage based on the estimated marginal distributed lag effect, $\widehat{\widetilde{\theta}}_t$. For TDLMM we calculated RMSE and coverage of $PM_{2.5}$ only. We also evaluated the probability that a model detects a true critical window (TP) or places a critical window where the true effect is zero (FP). For TDLMM, FP considered the marginal effects of all exposures. Finally, we evaluated precision of critical window identification, calculated TP$/$(TP$+$FP).

The tree-based methods were the most accurate in estimating the distributed lag function; they had lowest RMSE and maintained near 95% coverage. The penalized DLM had low RMSE, but poor coverage due to the wiggliness of splines, which also resulted in misclassification of critical windows. The CWVS model maintained high coverage of the distributed lag function, but was not as accurate in estimating the effects as evidenced by RMSE.

Table 3.1: Results for simulation scenario one: binary outcome with single exposure DLM effect. We compare Gaussian process (CWVS) and penalized cubic regression spline (DLMcr) DLMs with our treed DLM and DLMM methods under conditions of a frequent ($\overline{p} = 0.5$) or infrequent ($\overline{p} = 0.05$) binary outcome. The first two columns describe DLM estimation, which refers to estimation of the marginal exposure effects, $\widetilde{\theta}_t$, for the active exposure. The final three columns describe critical window identification, which is summarized by the probability DLM 95% credible intervals do not contain zero at correct (TP) or incorrect (FP) time periods, as well as Precision = TP/(TP + FP).

	DLM Estimation		Window Identification		
Model	RMSE $\times 100$	Coverage	TP	FP	Precision
$\overline{p} = 0.5$					
CWVS: dlm|cw	2.10	0.97	0.99	0.03	0.97
CWVS: p>0.5	2.10	0.97	0.99	0.07	0.93
DLMcr	1.83	0.77	1.00	0.17	0.85
TDLM	1.22	0.98	0.98	0.01	0.99
TDLMMadd	1.31	0.97	0.99	0.02	0.98
TDLMMns	1.47	0.96	1.00	0.03	0.97
TDLMM	1.39	0.96	0.99	0.03	0.97
$\overline{p} = 0.05$					
CWVS: dlm|cw	3.36	0.99	0.72	0.02	0.97
CWVS: p>0.5	3.36	0.99	0.94	0.11	0.90
DLMcr	2.64	0.72	1.00	0.21	0.83
TDLM	2.19	0.96	0.88	0.02	0.98
TDLMMadd	2.39	0.93	0.91	0.03	0.97
TDLMMns	2.49	0.93	0.88	0.04	0.96
TDLMM	2.59	0.91	0.83	0.04	0.95

In both $\overline{p}$ settings, the mixture models outperformed the single exposure spline-based DLM and CWVS models in terms of RMSE while maintaining near nominal coverage. These are important findings because the cost of including additional exposures into the treed models is minimal in terms of distributed lag function estimation. In TDLMMns when $\overline{p} = 0.5$ the posterior inclusion probability of $PM_{2.5}$ was 1. This decreased slightly (0.89) when $\overline{p} = 0.05$. For other exposures, the posterior inclusion probability was < 0.01 in both $\overline{p}$ settings.

Comparing models in terms of critical window detection, TDLM, TDLMM, and CWVS: dlm|cw were the most precise models at identifying critical windows. In the more difficult $\overline{p} = 0.05$ setting, the increased complexity of TDLMM was associated with a slight decrease in TP, but the method retained high precision due to the very low FP rate. CWVS: $p > 0.5$ had higher TP at the cost of higher FP. Spline methods were less precise with higher FP.

Results from the supplemental simulation with a smooth distributed lag function were similar and are presented in Appendix B.3. In both the current and smooth simulations, TDLM and TDLMM had lower RMSE, closer to nominal coverage, and higher precision in identifying critical windows compared to the alternative estimators.

3.5.2 Multiple exposures and a continuous outcome

We considered a continuous response with an exposure main effect from $PM_{2.5}$ and an interaction effect between $PM_{2.5}$ and NO_2. The outcome was generated as

$$y_i = c_2 f_2(\mathbf{x}_{i1}, \mathbf{x}_{i2}) + \mathbf{z}_i^T \boldsymbol{\gamma} + \epsilon_i \tag{3.11}$$

where c_2 is a scalar such that the variance of $c_2 f_2$ equals one. The simulated effect was

$$f_2(\mathbf{x}_{i1}, \mathbf{x}_{i2}) = \sum_{t=s_1}^{s_1+7} x_{i1t} + 0.025 \sum_{t_1=s_1}^{s_1+7} \sum_{t_2=s_2}^{s_2+7} x_{i1t_1} x_{i2t_2} \tag{3.12}$$

for starting times s_1 and s_2 each drawn uniformly from $\{1, \ldots, T-7\}$. This scenario consists of a main effect with a critical window of length eight for the $PM_{2.5}$ exposure and an interaction

Table 3.2: Results for simulation scenario two: main effect of $PM_{2.5}$ with $PM_{2.5}-NO_2$ interaction. The first four columns describe RMSE and coverage of the estimated marginal distributed lag effects, $\widetilde{\theta}_t$, for $PM_{2.5}$ and NO_2 compared to the true simulated marginal effects. The final four columns compare critical window detection for marginal effects of $PM_{2.5}$ and NO_2, which are summarized by the probability that the 95% CI does not include zero at a correct (TP) or incorrect (FP) time period.

	RMSE $\times 100$		Coverage		TP		FP		
Model	PM	NO_2	PM	NO_2	PM	NO_2	PM	NO_2	Other
$\sigma^2 = 25$									
TDLMMadd	3.59	4.19	0.95	0.84	0.94	0.82	0.03	0.07	0.00
TDLMMns	3.57	4.27	0.97	0.96	0.91	0.63	0.02	0.03	0.00
TDLMM	3.55	4.33	0.98	0.97	0.87	0.51	0.01	0.02	0.00
$\sigma^2 = 50$									
TDLMMadd	4.62	4.64	0.92	0.83	0.86	0.52	0.04	0.05	0.00
TDLMMns	4.61	4.75	0.96	0.97	0.76	0.31	0.02	0.02	0.00
TDLMM	4.50	4.79	0.97	0.97	0.62	0.22	0.01	0.01	0.00
$\sigma^2 = 100$									
TDLMMadd	5.75	5.07	0.89	0.84	0.47	0.22	0.03	0.03	0.00
TDLMMns	5.85	5.21	0.93	0.95	0.29	0.10	0.02	0.01	0.00
TDLMM	5.73	5.21	0.94	0.96	0.24	0.05	0.01	0.01	0.00

effect between exposure to $PM_{2.5}$ at times s_1 to $s_1 + 7$ and NO_2 at times s_2 to $s_2 + 7$. We drew ϵ_i independently for each observation from $\mathcal{N}(0, \sigma^2)$ such that $\sigma^2 \in \{25, 50, 100\}$. In this scenario, we compared TDLMM, TDLMMns (no-self interactions) and TDLMMadd (additive DLMs) using all exposure measurements from our data analysis. No other methods are currently available that would offer a direct comparison.

Table 3.2 compares the marginal DLM RMSE and coverage for active exposures $PM_{2.5}$ and NO_2. As in scenario one, we describe TP and FP for critical window detection. We found that all variants of TDLMM had low RMSE for estimating the marginalized effects of $PM_{2.5}$ and NO_2. The additive model had below nominal coverage for the marginal effect of NO_2 due to lack of appropriate interaction terms. We found that TDLMMns had higher power than full TDLMM. Additive TDLMM had the highest TP rate, but increased FP for NO_2. All TDLMM variants show zero FP for nonactive exposures. Overall, the tree-based DLMMs had high precision for identifying critical windows.

Table 3.3: Exposure and interaction posterior inclusion probabilities. The values in this table describe the probability at least one tree or tree-pair in the ensemble estimates the effects for a specific exposure or interaction, respectively. Nonactive exposures or interactions are summarized together as 'Other'.

	Main Effect			Interaction	
Model	PM	NO_2	Other	PM$-NO_2$	Other
$\sigma^2 = 25$					
TDLMMadd	1.00	1.00	0.78	-	-
TDLMMns	1.00	1.00	0.17	0.94	0.06
TDLMM	1.00	0.99	0.39	0.85	0.22
$\sigma^2 = 50$					
TDLMMadd	1.00	0.99	0.82	-	-
TDLMMns	0.99	0.95	0.26	0.83	0.10
TDLMM	0.99	0.96	0.43	0.80	0.23
$\sigma^2 = 100$					
TDLMMadd	0.98	0.96	0.85	-	-
TDLMMns	0.88	0.81	0.38	0.61	0.15
TDLMM	0.89	0.85	0.52	0.58	0.25

Table 3.3 summarizes exposure and interaction posterior inclusion probabilities averaged across simulation replicates. For the correct exposures, all variants of TDLMM had posterior inclusion probability above the prior inclusion probability (0.9) in low and medium noise settings. For non-active exposures, the posterior inclusion probability was below that of the correct exposures. The differences were less pronounced in TDLMMadd. Posterior inclusion probabilities for the correct interaction were consistently higher than for other interactions.

3.6 Analysis of Colorado Birth Cohort Data

We applied TDLMM to BWGAZ for our Denver, CO metro area dataset. The analysis included five exposures and covariates described in Section 3.2 with $T = 37$ weeks of exposure measurements corresponding to the first 37 weeks of gestation. In addition, our analysis controlled for year and month of conception, census tract elevation, and a county-specific intercept. TDLMM used the prior specification described in Section 3.4.5. The model ran for 100,000 iterations after 10,000 burn-in and was thinned to every 5th iteration. For this analysis, we used TDLMMns for

ease of interpretation of the linear effects. We compared to TDLMM and found few differences (see Appendix B.5.2).

3.6.1 Exposure and interaction selection

Four exposures had posterior inclusion probabilities above the prior inclusion probability (0.9): $PM_{2.5}$ (> 0.99), SO_2 (> 0.99), CO (0.989), and temperature (> 0.99). NO_2 was included into the model at a much lower rate (0.428). The posterior inclusion probability for the $PM_{2.5}-$temperature interaction was 0.988. The next highest posterior inclusion probabilities for interactions were CO$-$temperature (0.69) and SO_2-CO (0.6). The seven remaining interactions had posterior inclusion probabilities below 0.5.

3.6.2 Marginal exposure associations with BWGAZ

The top row of Figure 3.2 shows the estimated marginal distributed lag function of each exposure with other exposures fixed to their empirical means. The results identify critical windows to $PM_{2.5}$ during weeks $6-34$, SO_2 during weeks $10-29$, CO during weeks $9-11$ and temperature across the entire pregnancy. The critical windows show that increased exposure is associated with decreased BWGAZ.

TDLMM allows us to estimate interactions between time-resolved predictors. Figure 3.3(a) shows the presence of an interaction between log $PM_{2.5}$ and temperature. This interaction indicates that elevated exposure to $PM_{2.5}$ during weeks $15-25$ results in an increase in the magnitude of the temperature effect during weeks $19-35$. These interactions occur at the same times (e.g. weeks $19-25$) as well as different times (e.g. log $PM_{2.5}$ in week 20 and temperature in week 30) and may be explained by the priming hypothesis. Figure 3.3(b) shows the marginal distributed lag functions for $PM_{2.5}$ and temperature at varying percentiles of co-exposures. We see a larger effect of $PM_{2.5}$ and changes in critical windows when temperature and other exposures are fixed at a higher percentile. The effect of temperature also shows changes in the magnitude and timing of critical windows at increased levels of other exposures. Additional interaction plots are included in Appendix B.5 and show no evidence of meaningful interactions.

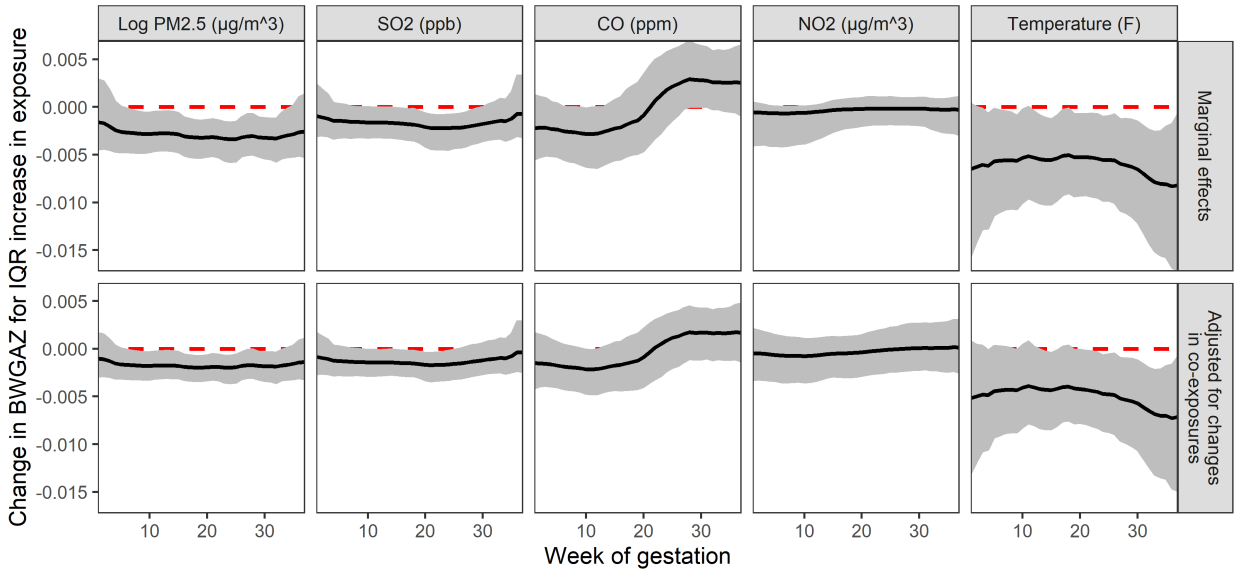


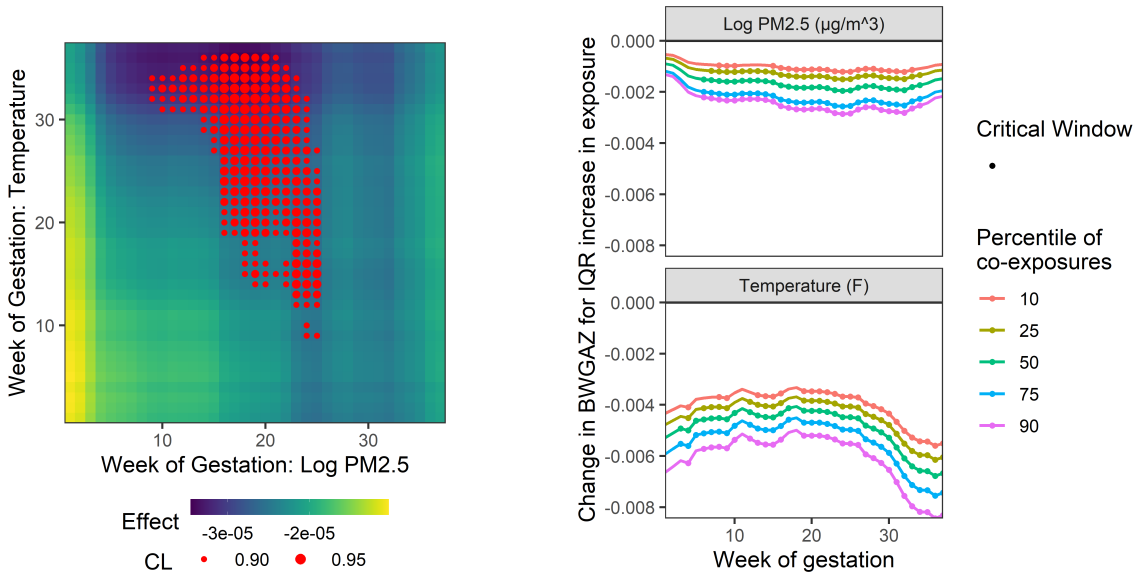
Figure 3.2: Posterior mean distributed lag distributed lag function (black line) for each exposure (columns) with 95% credible interval (grey area) of the effect. The top row shows the estimated marginal effect for an IQR increase in exposure, holding other exposures at their empirical mean. The bottom row shows the estimated change in BWGAZ for a first to third quartile change in one exposure along with the expected changes in all other exposures due to correlation with the exposure of interest.

3.6.3 Accounting for changes in co-exposures

Due to high correlation among exposures in our analysis (ranging from -0.55 to 0.69 at the same week), a change in any one exposure will likely be accompanied by simultaneous changes in co-exposures. This suggests that the marginal results in Section 3.6.2, which assumes that co-exposures are held at their empirical means, should only be interpreted in a narrow range. As an alternative, we estimate distributed lag functions that account for simultaneous changes in co-exposures. This provides another way to look at the results from the same model.

Let $x_{m(q)}$ represent the qth quantile of the empirical distribution of exposure m. The expected change in the outcome for an IQR change in exposure m at time t, conditional on the expected co-occurring changes in all other exposures can be written

$$\mathrm{E}\left[Y\middle|\widetilde{\mathbf{X}}_t = \mathrm{E}\left\{\mathbf{X}_t\middle|x_{mt} = x_{m(0.75)}\right\}, \widetilde{\mathbf{X}}_{[t]} = \overline{\mathbf{x}}, \mathbf{z} = \mathbf{z}_0\right] - \mathrm{E}\left[Y\middle|\widetilde{\mathbf{X}}_t = \mathrm{E}\left\{\mathbf{X}_t\middle|x_{mt} = x_{m(0.25)}\right\}, \widetilde{\mathbf{X}}_{[t]} = \overline{\mathbf{x}}, \mathbf{z} = \mathbf{z}_0\right]. \quad (3.13)$$



(a) $PM_{2.5}-$temperature interaction surface.

(b) Marginal DLMs showing interaction.

Figure 3.3: Panel (a) shows the estimated $PM_{2.5}-$temperature interaction effects. The color of each cell indicates the sign and direction of the interaction effect between $PM_{2.5}$ at one time and temperature at another time. The points indicate estimated interactions where the credible interval does not contain zero and the point size represents the credible interval probability. Panel (b) shows the estimated marginal distributed lag function for an IQR increase in the indicated exposure, when all other exposures are fixed at a given percentile (color of lines) of their empirical distributions. Points on each line indicate where the 95% credible interval of the marginal effect does not include zero.

Here, $\widetilde{\mathbf{X}}_t = \{\widetilde{x}_{1t}, \ldots, \widetilde{x}_{Mt}\}$ defines the values of all exposures at time t while $\widetilde{\mathbf{X}}_{[t]}$ is the collection of exposure measurements at all time points except t. We set each element of $\widetilde{\mathbf{X}}_{[t]}$ equal to the empirical mean for that exposure to isolate the effect only at the time of interest. The value of $\mathbf{z}$ does not influence the expected change because it does not interact with exposure measurements. To estimate the values of co-exposures at time t, given by $\mathrm{E}[\mathbf{X}_t | x_{mt} = x_{m(q)}]$, we fit penalized spline models (cubic splines with 5 degrees of freedom) between pairs of exposures, estimated using restricted maximum likelihood as described by Wood (2017). Specifically, we considered measurements for exposure m as the only predictor and fit separate models for each co-exposure (e.g. log $PM_{2.5}$ was used as the predictor for NO_2, SO_2, CO, and temperature in four separate models). Using the model fits, we predicted all co-exposures at the 25^{th} and 75^{th} percentiles of exposure m. This process was repeated for every $m \in \{1, \ldots, 5\}$.

The results of this analysis adjusting for changes in co-exposures are shown in the bottom row of Figure 3.2. This can be interpreted as the expected change in BWGAZ associated with an IQR change in one pollutant and the expected change in the four other co-exposures at the same time point. There are several important takeaways. First, the shape of estimated distributed lag functions are similar. Second, the confidence intervals take into account the uncertainty in the estimated main effects and interactions of co-exposures. That uncertainty is not included in the estimates presented in the top row of Figure 3.2 because the level of co-exposures remain fixed. Third, the associations between BWGAZ and exposure to $PM_{2.5}$, SO_2 and temperature are persistent after accounting for expected changes in co-exposures.

A change in $PM_{2.5}$ from the 25^{th} to 75^{th} percentile of exposure is an increase from 6.12 to 8.67 $\mu g/m^3$. Considering the association between $PM_{2.5}$ and BWGAZ after adjusting for co-exposures finds a critical window during weeks $6 - 33$, which is similar to the critical window found in the marginal distributed lag function. The cumulative effect of $PM_{2.5}$, or the effect of a simultaneous IQR increase in every week of pregnancy, adjusting for co-exposures is associated with an average change in BWGAZ of -0.064 (95% CI: $-0.094, -0.035$). The cumulative effect of an IQR increase in SO_2 (0.94 to 1.90 ppb) is associated with a change in BWGAZ of -0.049 (95% CI: $-0.076, -0.019$), with a critical window during weeks $9 - 25$.

Our findings of an inverse relationship between $PM_{2.5}$ and birth weight are consistent with a meta-analysis done by Stieb et al. (2012). Similar findings exist for a relationship between SO_2 and birth weight (Dugandzic et al., 2006), but are less consistent (Stieb et al., 2012). We note that the majority of comparison studies do not account for effects of co-occurring exposures or interactions over time and no previous study has considered time-resolved measures of five exposures including interactions.

3.7 Discussion

In this work we propose TDLMM to estimate the association between mixtures of environmental exposures observed at high-temporal resolution on birth and children's health outcomes.

TDLMM, and the reduced version TDLM, have several innovative statistical features. The models introduce structured regression trees that estimate a constrained DLM. The mixtures approach, TDLMM, uses an ensemble of tree pairs that parameterize both main effects and pairwise interactions between time-resolved predictors.

In a simulation study we demonstrated that our single exposure tree-based DLM outperformed established methods for estimating the DLM in terms of estimation of the exposure-response function and precision in identifying critical windows. Moreover, our tree-based mixture approach, TDLMM, that included additional exposures not associated with the outcome also outperformed the established single exposure methods in the single exposure setting. This is in large part due to the methods' ability to select out or shrink the effects of exposures that are not associated with the outcome. Using real exposure data, TDLMM identifies critical windows, selects the proper exposures, and estimates the exposure-response function in a simulation scenario with five time-resolved predictors.

We applied TDLMM to analyze associations between Denver, CO area birth weight and five environmental exposures experienced weekly by mothers during gestation. This data set included 195,701 full term births with estimated conception dates from 2007 through 2015. This data analysis to estimate the main and interaction effects due to a mixture of five environmental exposures observed weekly during the first 37 weeks of gestation produced several key takeaways. We identified $PM_{2.5}$, SO_2 and temperature effects based on the effect of one exposure conditional on fixed co-exposures and based on a change in one exposure accounting for simultaneous changes in co-exposures. In addition, we found substantial evidence of an interaction between $PM_{2.5}$ and temperature. Independently, birth weight changes due to exposure appear small, but when combined with other risk factors (including those often correlated with increased pollution) may have a substantial impact on birth weight, which may increase the prevalence of associated comorbidities. Furthermore, there is evidence that the effects of air pollution are larger at the lower quantiles of birth weight—arguably the more susceptible population (Lamichhane et al., 2020).

Our analysis of five pollutants observed weekly throughout pregnancy and birth weight is the first analysis to identify critical windows to a mixture observed at high-temporal resolution within a distributed lag mixture model framework. Previous studies have either estimated critical windows for a mixture observed at high temporal resolution in an additive DLM setting (Figueroa-Romero et al., 2020; Horton et al., 2018) or to a mixture observed at one or a small number of time points (Levin-Schwartz et al., 2019; Chiu et al., 2018). Data driven methods that allow for precision in identifying critical windows due to environmental mixtures have the potential to open doors to discovery and understanding in biological science (Wright, 2017). As the size and resolution of the exposure data available continues to grow, our method fills a critical research gap in statistical methods for epidemiology and environmental health in being able to estimate the effects of time-resolved measures of a mixture on a continuous or binary health outcome.

Chapter 4

Heterogeneous Distributed Lag Models to Estimate Personalized Effects of Maternal Exposures to Air Pollution

4.1 Introduction

A growing body of research has found maternal exposure to environmental chemicals during pregnancy to be associated with changes in children's birth and health outcomes. Detrimental outcomes linked to increased exposure include decreased birth weight (Bell et al., 2007), increased risk of preterm birth (Stieb et al., 2012), increased risk of asthma (Lee et al., 2018; Bose et al., 2017), and altered neurological outcomes (Chiu et al., 2016), among others (Šrám et al., 2005). Recently, research has focused on determining the time periods during fetal development, or critical windows of susceptibility, when increased exposure can alter future health outcomes. Identification of critical windows gives insight into how biological processes involved in development may be impacted by exposure to environmental chemicals (Wright, 2017). However, exposure effects, including critical window timing and effect magnitude, are likely to vary across a population. Effect heterogeneity may be governed by biological (e.g., sex of child), socioeconomic (e.g., maternal income), or other non-chemical environmental factors (e.g., neighborhood characteristics). Estimating individualized exposure effects will better inform precision environmental health interventions. Furthermore, this brings attention to vulnerable populations, which the Environmental Protection Agency (EPA) is required to protect through the 2016 update to the Toxic Substances Control Act (Krimsky, 2017).

A commonly applied method to identify perinatal critical windows and estimate the exposure-response relation between maternal exposures and an outcome is the distributed lag model (DLM). In a DLM, an outcome is regressed on repeated measures of an exposure assessed over a time

period prior to the outcome. DLMs are typically constrained so the exposure-response function varies smoothly over the time period of exposure. These constraints yield effect estimates that are more biologically plausible as well as add stability to the estimator in the presence of high autocorrelation in the exposure data, which is typical with repeated measures of environmental exposures, such as the air pollution exposure considered in this paper. Methods for constraining DLMs include splines (Zanobetti et al., 2000; Gasparrini et al., 2010), Gaussian processes (Warren et al., 2012), principal components (Wilson et al., 2017a), and regression trees (Mork and Wilson, 2021a). The majority of studies that apply these methods assume a homogeneous exposure-response relationship across the population.

Several methods have sought to estimate DLMs that vary across a population. Wilson et al. (2017a) proposed a Bayesian distributed lag interaction model (BDLIM) to estimate differences in the exposure-response function for a parsimonious set of predetermined subgroups. Warren et al. (2012) and Warren et al. (2020b) developed spatially-varying DLMs to account for changes in pollution composition or demographics over a study region. These papers highlight the bias incurred by a homogeneous effects assumption when effects are truly heterogeneous. BDLIM has been applied in several epidemiological analyses, including two that found an increased risk of asthma due to prenatal exposures to fine particulate matter and nitrate for a subgroup of boys concurrently exposed to high prenatal stress (Lee et al., 2018; Bose et al., 2017). Despite these advances in estimating DLM heterogeneity, the methods of Wilson et al. (2017a), Warren et al. (2012), and Warren et al. (2020b) remain limited in their scope and interpretability. A spatially-varying DLM is unable to identify the subject characteristics associated with changes in the underlying distributed lag function. BDLIM is limited to predefined categorical subgroups and is most reasonably applied to data having only a small number of subgroups. In the age of big data, we have access to a potentially high dimensional set of categorical and continuous modifiers, and the true modifiers responsible for differences in distributed lag effects are unknown. Harnessing this information can lead to personalized environmental health decision making.

We propose a Bayesian additive regression tree (BART) method for estimating distributed lag effect heterogeneity due to a set of modifying covariates. The BART framework of Chipman et al. (2010) is a popular method for estimating non-parametric functions. Extensions of BART allow for estimating distributed lag nonlinear (Mork and Wilson, 2021b) and mixture models (Mork and Wilson, 2021a). The treed DLM approach of Mork and Wilson (2021a) outperforms competing spline and Gaussian process methods when the goal is distributed lag effect estimation and critical window identification. However, both of these approaches assume a common distributed lag effect for all individuals.

Chipman et al. (2002) proposed treed regression models that modify a vector of regression coefficients using a single Bayesian tree. More recently, several approaches have used an ensemble of multiple Bayesian trees to modify a single predictor. For example, Starling et al. (2020) introduced BART with targeted smoothing (tsBART), which allows a smooth risk function of a single univariate predictor to vary across a population. Deshpande et al. (2020) proposed a varying coefficient BART model that uses a separate ensemble of regression trees to modify each regression coefficient in the model. In a non-Bayesian approach to estimating effect heterogeneity, Odden et al. (2020) applied a random forest algorithm to identify heterogeneous exposure associations. However, no method has been proposed to modify a structured vector of regression coefficients, such as a constrained distributed lag function, using an ensemble of trees.

In this paper, we define the heterogeneous DLM (HDLM). HDLM extends the DLM to estimate a personalized distributed lag function that varies across the population according to a possibly high dimensional set of potential modifying factors. Our additive regression tree method introduces individualized functional predictors through a vector of structured regression coefficients on the terminal nodes of a modifier tree. Specifically, an ensemble of regression trees divides the sample based on a set of modifying covariates and estimates a DLM unique to each subgroup. We introduce three methods for estimating the distributed lag effects on the modifier tree terminal nodes. The first considers a Gaussian process DLM for each subgroup and assumes the same smoothness in the distributed lag effects for all observations. The second method incorporates the

treed DLM method of Mork and Wilson (2021a) and defines a nested tree structure where the treed DLM is used to estimate unique critical windows and effect sizes for the subgroup at each terminal node of the modifier tree. This approach relaxes the smoothness assumption imposed by Gaussian processes and improves performance when the magnitude of the effect varies over the population or there are subgroups with no exposure effect. The third method also uses a treed DLM whose structure is shared across all subgroups and allows for variation in only the magnitude of the effect, but not the location of the critical window. This approach shares information on the structure of the DLM across subgroups. We develop a computational framework for our methods that selects modifying covariates responsible for changes in the exposure-response relationship. In addition, our method extends the tree literature by introducing the concept of a nested tree model, where a tree-based functional estimator is given to subgroups defined by a traditional regression tree.

We provide a comprehensive simulation of our method and show it is able to estimate individualized critical windows and effects as well as identify the covariates responsible for modifying the distributed lag function. The simulation also shows that the nested and shared tree HDLMs outperform the Gaussian Process HDLM in terms of HDLM estimation and critical window detection. We apply our method to a Colorado-based administrative birth cohort and explore differences in the relationship between fine particulate matter ($PM_{2.5}$) and birth weight across a range of continuous, categorical, and binary modifiers specific to individuals. We identify age, body mass index (BMI), Hispanic designation, and education as potential modifiers of the distributed lag effects. The analysis indicates that non-Hispanics with increased BMI are more susceptible to $PM_{2.5}$ exposures and early- and late-gestation are potential critical windows. In addition, we find individual variability within subgroups due to other modifying characteristics. Software to replicate our simulation and use our method in other applications is available in the `R` package `dlmtree`.

4.2 Colorado Birth Cohort Data

We analyze birth weight for gestational age z-score (BWGAZ) using birth vital statistics records from Colorado, USA. BWGAZ is the birth weight adjusted for gestational age and fe-

tal sex using a standard reference table (Fenton and Kim, 2013). Besides birth outcomes, the data include individual covariate information including mother's age, weight, height, income, education, marital status, prenatal care habits, smoking before and during pregnancy, as well as race and Hispanic designations. The data include all births from Colorado with estimated conception dates between 2007 and 2015, inclusive. We limit the data to the northern front range counties (those immediately east of the Rocky Mountains roughy extending from Colorado Springs to the Wyoming border). This area contains the majority of the Colorado population. We further restrict our analysis to census tracts with elevation lower than 6000 feet above sea level. This restriction both reduces the potential confounding by altitude and the impact of mountainous terrain on exposure data. We analyze singleton, full-term births ($\geq$ 37 weeks) with complete covariate and exposure data, resulting in 310,236 births.

We are interested in the association between a mother's weekly exposure to particulate matter smaller than or equal to 2.5μm in diameter ($PM_{2.5}$) and resulting birth weight. In addition, we wish to identify vulnerable populations with increased susceptibility to $PM_{2.5}$ exposures and estimate each individual's critical windows and distributed lag effects. To answer these questions, we consider heterogeneous distributed lag effects that differ across a set of modifying covariates specific to each observation. Potential modifiers in our analysis include continuous variables: maternal age and body mass index (BMI); ordinal variables: income classification, highest educational attainment, and smoking (never, former, less than 10 cigarettes/day, at least 10 cigarettes/day); nominal variables: marital status, prenatal care, and race; and binary variables: sex and Hispanic indicators. The distribution of the modifiers are described in Appendix C.1.

At each census tract in our study, daily $PM_{2.5}$ measurements were obtained from EPA community multiscale air quality modeling system (CMAQ) using downscaled data (Berrocal et al., 2010). We then created 37 consecutive weekly average exposures for each pregnancy beginning on the date of conception for the census tract of residence. The weekly average $PM_{2.5}$ data were log-transformed to reduce skew. The weekly average log-transformed $PM_{2.5}$ measurements range from 0.68 to 3.53 with mean 1.97 and standard deviation 0.30. Pregnancy average log $PM_{2.5}$ ex-

posure ranged from 1.34 to 2.33. To account for potential spatial and temporal confounding, we included a fixed intercept for county, year, and month of conception, and controlled for census tract elevation and trimester average temperature. This study was approved by the Institutional Review Board of Colorado State University.

4.3 Methods

4.3.1 Heterogeneous DLM Framework

We consider a sample $i = 1, \ldots, n$ with continuous outcome y_i, a vector of exposure measurements $\mathbf{x}_i = [x_{i1}, \ldots, x_{iT}]'$ taken at equally spaced times $t \in \{1, \ldots, T\}$, and a vector of covariates, $\mathbf{z}_i$, which includes the model intercept. In this paper, y_i refers to BWGAZ while x_{it} is the $PM_{2.5}$ exposure measurement for individual i during week t of pregnancy (considering $T = 37$ weeks of gestation). When there is no heterogeneity in the exposure effect across the sample population, the discrete time DLM takes the form

$$y_i = \sum_{t=1}^{T} x_{it}\theta_t + \mathbf{z}_i'\boldsymbol{\gamma} + \varepsilon_i. \tag{4.1}$$

Here, θ_t represents the linear effect due to exposure at time t, $\boldsymbol{\gamma}$ is a vector of regression coefficients and ε_i represents independent errors distributed $\mathcal{N}(0, \sigma^2)$.

We consider a DLM where the exposure effects $\theta_1, \ldots, \theta_T$ vary across the population. Let the modifiers, $\mathbf{m}_i$, be a subset of the covariates, $\mathbf{z}_i$. The heterogeneous DLM is

$$y_i = \sum_{t=1}^{T} x_{it}\theta_t(\mathbf{m}_i) + \mathbf{z}_i'\boldsymbol{\gamma} + \varepsilon_i \tag{4.2}$$

where $\theta_t(\mathbf{m}_i)$ is the exposure effect at time t specific to an observation with modifiers $\mathbf{m}_i$.

Our method for modeling heterogeneity due to a set of modifying covariates uses an ensemble of regression trees combined with functional estimators of the distributed lag effect. We denote a modifier regression tree by $\mathcal{M}_a$ for $a \in \{1, \ldots, A\}$ that partitions the population based on

a set of modifiers $\mathbf{m}_i$. Terminal nodes are denoted η_{ab} where $b = 1, \ldots, B_a$ distinguishes the subgroups partitioned by tree a. Each terminal node is associated with a T-dimensional vector of parameters $\boldsymbol{\theta}_{ab}$. Considering all trees in the ensemble, the distributed lag function for observation i, $\boldsymbol{\theta}(\mathbf{m}_i) = [\theta_1(\mathbf{m}_i), \ldots, \theta_T(\mathbf{m}_i)]'$, is

$$\boldsymbol{\theta}(\mathbf{m}_i) = \sum_{a=1}^{A} \sum_{b=1}^{B_a} \boldsymbol{\theta}_{ab} \mathbb{I}(\mathbf{m}_i \in \eta_{ab}) \tag{4.3}$$

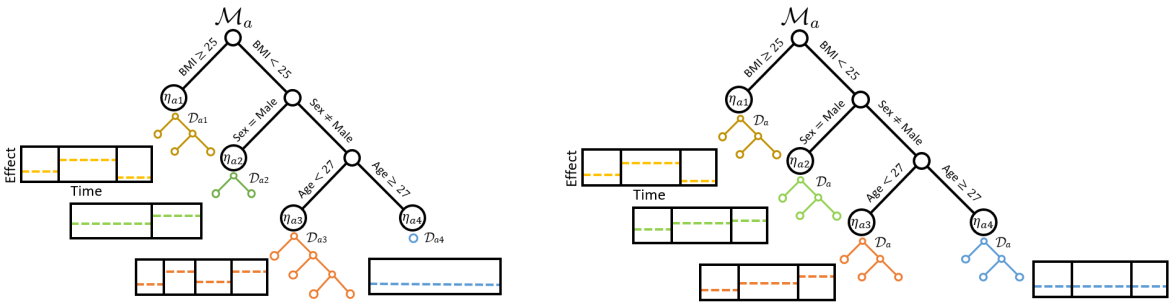
where $\boldsymbol{\theta}_{ab} = [\theta_{ab1}, \ldots, \theta_{abT}]'$ parameterizes the partial distributed lag function for the subgroup represented by terminal node η_{ab} in tree $\mathcal{M}_a$ and $\mathbb{I}(\cdot)$ is the indicator function.

4.3.2 Gaussian Process HDLM

We first propose a method for estimating the HDLM using Gaussian processes, which is an established approach for DLMs (e.g. Warren et al. (2012)). Consider a single modifier tree terminal node, η_{ab}, containing a subset of observations. Let the corresponding $T-$dimensional set of DLM parameters $\boldsymbol{\theta}_{ab}$ follow a Gaussian process with covariance function $\Sigma_{\boldsymbol{\phi}}(t, t')$ where $\boldsymbol{\phi}$ are parameters defining the covariance. We consider the exponential covariance, $\Sigma_{\phi}(t, t') = \exp\{-\phi|t - t'|\}$. An inherent assumption of this approach is that all observations share the same value of ϕ and therefore have the same smoothness over time in their distributed lag effects. Having a common ϕ for all nodes on a tree is required due to model computation, which requires integration over the distributed lag effects for each subgroup. The equal smoothness assumption may be beneficial if the distributed lag effects for all observations have a similar magnitude and window length. In the case where there is a nonzero exposure effect in only a small proportion of the population, the smoothness of the distributed lag effect imposed by the remaining sample will make this critical window difficult to estimate. This is because estimation of the smoothing parameter will largely reflect the null group and cause oversmoothing of the active group.

We define the Gaussian process prior for the distributed lag effects as

$$\boldsymbol{\theta}_{ab} | \tau_a, \nu, \sigma, \phi \sim \mathcal{GP}[\mathbf{0}, \tau_a^2 \nu^2 \sigma^2 \boldsymbol{\Sigma}(\phi)]. \tag{4.4}$$



(a) Nested Tree HDLM

(b) Shared Tree HDLM

Figure 4.1: Panel (a) diagrams the nested tree HDLM. Modifier tree $\mathcal{M}_a$ is structured with binary splitting rules on modifiers BMI, sex, and age. Each terminal node η_{ab} has a unique treed DLM structure $\mathcal{D}_{ab}$ with corresponding piecewise effects given by δ_{abc}, shown as dashed lines. Panel (b) diagrams the shared tree HDLM. Here, each terminal node of the modifier tree η_{ab} uses the same treed DLM structure $\mathcal{D}_a$. The time points where the distributed lag effects change are shared across all trees while the effect magnitude is unique to each subgroup.

The variance parameters follow a half-Cauchy prior, $\tau_a, \nu \sim \mathcal{C}^+(0,1)$, to define a local-global horseshoe-like estimator on tree specific effects (Carvalho et al., 2010). This differs from previous BART implementations (Chipman et al., 2010; Starling et al., 2020), which apply a uniform variance across all trees. The horseshoe variance prior will shrink the effects of misspecified trees reducing variance and false window detection. A similar prior specification was shown to improve DLM estimation in the treed DLM method of Mork and Wilson (2021a). We restrict the range of ϕ to $\exp\{-\phi\} \in (0.05, 0.95)$ and assign prior $\phi \sim \text{Gamma}(1/2, 1/2)$, which gives higher probability to smoother distributed lag effects.

4.3.3 Nested Tree HDLM

Mork and Wilson (2021a) showed that treed DLMs outperform competing spline and Gaussian process-based methods in terms of distributed lag effect estimation and precision of critical window identification. We propose using a treed DLM as a parametric model at each terminal node of the modifier tree. This results in a nested tree structure visualized in Figure 4.1(a). Here, each subgroup of the modifier tree is paired with a unique distributed lag tree structure and corresponding effects.

Before describing the nested tree HDLM we give an overview of the treed DLM. The notation presented here distinguishes the DLM regression tree from the modifier regression tree presented in Section 4.3.1. Consider binary tree $\mathcal{D}$, which partitions the exposure time span, T, into C non-overlapping time segments. Internal nodes of $\mathcal{D}$ are assigned binary rules at time points within the period of exposure (e.g. $t < t_1$ and $t \geq t_1$ for $t_1 \in \{2, \ldots, T\}$). The terminal nodes of $\mathcal{D}$, denoted λ_c for $c \in \{1, \ldots, C\}$, bin together time points to define a piecewise constant distributed lag effect (see Figure 4.1(a)). That is, $\theta_t = \delta_c$ if $t \in \lambda_c$, where δ_c is a distributed lag effect for all exposure observations within the time period of λ_c. The binning of distributed lag effects adds structure to the distributed lag function and stabilizes the model in the presence of autocorrelation in the exposure data.

In the nested tree DLM, each modifier tree subgroup has a unique treed DLM structure. As a result, there is no sharing of information across terminal nodes concerning the timing of the critical window, the effect size, or the smoothness of the DLM. This creates a highly flexible model that allows structures of the treed DLM to adapt to the subgroup whose exposure-response is being estimated. Keeping the number of treed DLM terminal nodes small introduces constraints in the distributed lag function to account for temporal correlation in the exposure measurements.

To formally define the nested tree DLM in the context of an ensemble of modifier trees, denote $\mathcal{D}_{ab}$ as the treed DLM associated with terminal node η_{ab} in modifier tree $\mathcal{M}_a$. Treed DLM $\mathcal{D}_{ab}$ contains C_{ab} terminal nodes, denoted λ_{abc}, with corresponding distributed lag effects δ_{abc}, $c = 1, \ldots, C_{ab}$. To calculate the HDLM, as in (4.3), let $\theta_{abt} = \delta_{abc}$ if $\mathbf{m}_i \in \eta_{ab}$ and $t \in \lambda_{abc}$. Then the distributed lag effect for individual i at time t is

$$\theta_t(\mathbf{m}_i) = \sum_{a=1}^{A} \sum_{b=1}^{B_a} \sum_{c=1}^{C_{ab}} \delta_{abc} \mathbb{I}(\mathbf{m}_i \in \eta_{ab}, t \in \lambda_{abc}). \tag{4.5}$$

4.3.4 Shared Tree HDLM

The shared tree HDLM is a simplification of the nested tree HDLM that pairs a single treed DLM with each modifier tree. All subgroups of a modifier tree receive the same treed DLM

structure but distributed lag effects with different magnitudes. The shared tree HDLM is visualized in Figure 4.1(b). The modifier tree and treed DLM structures are learned from the data, but a change in the treed DLM structure is applied to all subgroups of a modifier tree. This assumes that the DLM has the same change points for critical windows. For our data problem involving prenatal development, this could relate to an assumption that developmental stages always occurring during the same weeks, although the magnitude of effects may differ. In contrast, the nested tree HDLM allows for the possibility that the onset and duration of the effects varies for each subgroup.

Notation for the shared tree HDLM is similar to the nested tree HDLM. For each modifier tree $\mathcal{M}_a$ with terminal nodes η_{ab}, we consider a single treed DLM $\mathcal{D}_a$ with terminal nodes λ_{ac}, $c \in \{1, \ldots, C_a\}$. The treed DLM $\mathcal{D}_a$ is utilized as the distributed lag function at all terminal nodes η_{ab} of $\mathcal{M}_a$. The distributed lag effects, δ_{abc}, are specific to each modifier tree terminal node and treed DLM terminal node. As in the nested tree DLM, we calculate the heterogeneous DLM by setting $\theta_{abt} = \delta_{abc}$ if $t \in \lambda_{ac}$.

4.3.5 Prior Specification

Here we detail the prior specification of nested tree HDLM. The approach for other methods are similar and details are available in Appendix C.2. The nested tree HDLM prior consists of five components: modifier tree structure, treed DLM structure, distributed lag effects, fixed effects of covariates, and the error variance. We have several goals in mind when shaping priors. First, as with BART, trees with fewer terminal nodes will help to stabilize the model. This is particularly true for the treed DLM where few terminal nodes provide a necessary constraint on the distributed lag effects. Second, the model should prioritize rules on modifiers that result in different DLM structures or effects to remove modifiers that do not differentiate subgroups. Third, we want to lower false window detection rates by shrinking the effects pertaining to subgroups and treed DLMs that poorly fit the data.

Modifier Tree Priors

The prior on modifier tree structures is defined in two parts: a prior probability that a node will have a split and a prior probability of the binary rule defined at that split. We adopt the BART prior for a node split. That is, for node η with depth d_η (the first node in a tree has depth zero) the probability the node is an internal node equals $p_{\text{split}}(\eta) = \alpha(1 + d_\eta)^{-\beta}$ where $\alpha \in (0, 1)$ and $\beta > 0$. Following Chipman et al. (2010) we set $\alpha = 0.95$ and $\beta = 2$, which encourages smaller trees. Changes to these priors did not improve performance in simulations. Priors on tree splitting rules in the modifier tree follow Linero and Yang (2018). Complete details are given in Appendix C.2.

Treed DLM Priors

The treed DLM also uses the BART prior for node splits, with $\alpha = 0.95$ and $\beta = 2$. For the splitting rule prior, we assign a uniform prior across all available time points, resulting in $T - 1$ possible splits for a tree with a single node.

The distributed lag effects are assigned the conjugate normal prior,

$$\delta_{abc} | \tau_a, \nu, \sigma \sim \mathcal{N}(0, \tau_a^2 \nu^2 \sigma^2) \tag{4.6}$$

where $\tau_a, \nu \sim \mathcal{C}^+(0, 1)$ define a horseshoe-like estimator on tree specific effects. We include the error variance σ^2 in this prior, allowing it to be integrated out during tree updates. This prior specification differs from previous BART implementations (Chipman et al., 2010; Starling et al., 2020), which apply a uniform variance prior across all trees. The modifier-tree-specific variance prior improves performance in treed DLM because it shrinks the effects of misspecified trees, which serves to reduce variance and false window detection (Mork and Wilson, 2021a).

Other Priors

To complete a fully Bayesian specification of the nested tree DLM we assign a non-informative prior to the fixed effects, $\boldsymbol{\gamma} \sim \mathcal{MVN}(\mathbf{0}, d\sigma^2 I_p)$, where I_p is a $p \times p$ identity matrix and d is fixed at a large value. Finally, we specify prior $\sigma \sim \mathcal{C}^+(0, 1)$.

4.3.6 Computation

The nested tree DLM is estimated by sampling from the posterior distribution using MCMC. As in BART, we apply Bayesian backfitting (Hastie and Tibshirani, 2000) to estimate the effects for each modifier tree and apply the independent Metropolis-Hastings (MH) algorithm to update modifier trees and treed DLMs. Our algorithm differs from BART in several ways. First, we control for a set of fixed effects when estimating the heterogeneous DLMs. Second, each terminal node of the modifier tree has a unique treed DLM structure that is learned from the data and must be updated. We briefly outline our algorithm and provide full details in Appendix C.3.

To improve estimates of the distributed lag function we integrate the fixed effect parameters, $\boldsymbol{\gamma}$, out of the data likelihood. Updates to the modifier tree occur through the four proposal steps described in Chipman et al. (1998): grow, prune, change, and swap. The grow step adds an additional split at a terminal node, and prune removes a split from an internal node connected to two terminal nodes. The change step modifies a binary splitting rule at an internal node. Swap reverses the order of rules in two adjacent internal nodes.

Each terminal node of the modifier tree has a unique treed DLM. In a grow proposal, a terminal node becomes an internal node and the corresponding treed DLM is eliminated and replaced by two new treed DLMs at the new modifier tree terminal nodes. For each new terminal node, a new treed DLM is drawn from the tree prior. Likewise for a prune proposal, an internal node becomes a terminal node. In this case the two existing treed DLMs are discarded and a new treed DLM is drawn from the tree prior. During a change or swap step in the modifier tree the terminal nodes retain the same treed DLM. To account for the potential change in dimensionality during modifier

tree updates we integrate over all terminal node effects, δ_{abc}, and model variance, σ^2. The update can then proceed with a MH step similar to that used in previous BART implementations.

After an update to the modifier tree, we update the treed DLM associated with each subgroup. Updates to treed DLMs use only grow, prune, and change proposals. A swap results in empty treed DLM terminal nodes and is not used. Conditionally conjugate normal priors allow for Gibbs updates of terminal node effects. Following updates to tree structures and terminal node effects, the remaining parameters are updated with standard MCMC procedures.

4.4 Simulation

We developed three simulation scenarios to evaluate estimation of the heterogeneous distributed lag function and the ability to identify correct modifiers. The first simulation scenario relates to the nested tree HDLM and considered subgroups with three different distributed lag effects: an early window, a late window, and no effect. The second scenario mimics the shared tree HDLM and had two groups: a distributed lag effect that does not change in time but was scaled by a continuous modifier and a group with no effect. Scenario 3 compared the HDLM to traditional DLM methods when there was no effect modification.

In general, we found that the nested and shared tree HDLMs outperform the Gaussian process HDLM in all settings and performed similarly to a standard DLM in scenario 3 with no effect modification. For subgroups with no distributed lag effects, the nested tree HDLM was better than the shared tree HDLM. In a low noise setting we saw a greater distinction between the nested and shared tree approaches, while they were comparable in higher noise scenarios.

Each scenario involved 13 covariates. Two of these covariates were responsible for the DLM heterogeneity in scenarios 1 and 2. All covariates were included as potential modifiers in the HDLMs. Covariates $\mathbf{z}_i = [z_{i1}, \ldots, z_{i13}]'$ were independently generated: $z_{i1} \sim \mathcal{N}(0,1)$, $z_{i2} \sim$ Bernoulli(0.5), $z_{i3} \sim$ Uniform$(0,1)$, $z_{ip} \sim \mathcal{N}(0,1)$ for $p \in \{4, \ldots, 8\}$, $z_{ip} \sim$ Bernoulli(0.5) for $p \in \{9, \ldots, 13\}$. We also include $z_{i0} = 1$ as a model intercept. A set of 37 consecutive weekly

$PM_{2.5}$ exposures, denoted $\mathbf{x}_i = [x_{i1}, \ldots, x_{i37}]'$ was drawn for each observation using real exposure measurements from our data analysis.

The simulation generated continuous outcomes under the model

$$y_i = r \cdot \mathbf{x}_i' \boldsymbol{\theta}(\mathbf{m}_i) + \mathbf{z}_i' \boldsymbol{\gamma} + \varepsilon_i \tag{4.7}$$

where $\boldsymbol{\gamma}$ are parameters drawn from standard normal, and $\boldsymbol{\theta}(\mathbf{m}_i)$ represent the HDLM from each simulation scenario. In the HDLM we let $\mathbf{m}_i$ be all variables in $\mathbf{z}_i$ except for the intercept, although only two modifiers are actually responsible for the effect heterogeneity. Scaling factor r was defined such that $\text{Var}[r \cdot \mathbf{x}_i' \boldsymbol{\theta}(\mathbf{m}_i)] = 1$, and ε_i was drawn independently from $\mathcal{N}(0, \sigma^2)$ under three different noise settings, $\sigma^2 \in \{10, 25, 50\}$. Each simulation scenario and σ^2 combination was run 100 times with $n =$ 5,000 observations. An additional 5000 observations were generated as a testing dataset to evaluate out-of-sample model performance but not used for model fitting.

Each simulation replicate was estimated by three methods for HDLM: nested tree HDLM, shared tree HDLM, and Gaussian process HDLM. We also used two methods for a standard DLM: treed DLM and Gaussian process DLM. Simulation scenario 1 was also estimated using the nested tree and Gaussian process HDLM where the modifier tree was fixed to use the true subgroups. All models used 20 modifier trees in the ensemble and were run for 10,000 MCMC iterations thinned to every 5^{th} iteration, following 5,000 burn-in iterations. All simulations can be reproduced with `R` package `dlmtree`.

4.4.1 Scenario 1: Early/Late Window

We simulated the heterogeneous distributed lag function with two true modifiers, z_{i1} and z_{i2},

$$\theta_t(\mathbf{m}_i) = \begin{cases} \mathbb{I}(t \in [11, 18]) & \text{if } z_{i1} > 0 \text{ and } z_{i2} = 1 \\ \mathbb{I}(t \in [17, 26]) & \text{if } z_{i1} > 0 \text{ and } z_{i2} = 0 \cdot \\ 0 & \text{if } z_{i1} \leq 0 \end{cases} \tag{4.8}$$

The first group ($z_{i1} > 0, z_{i2} = 1$) had a nonzero distributed lag effect during weeks $11 - 18$. The second group ($z_{i1} > 0, z_{i2} = 0$) had a nonzero distributed lag effect during weeks $17 - 26$, overlapping with the first group by two weeks. The third group ($z_{i1} \leq 0$) had no distributed lag effect.

Pointwise DLM results averaged across simulation replicates are given in Table 4.1. We separately analyzed subgroups with an effect ($z_{i1} > 0$) from subgroups with no effect ($z_{i1} \leq 0$). We report DLM root mean square error (RMSE) $= \sqrt{\sum_{t=1}^{37}[\theta_t(\mathbf{m}_i) - \hat{\theta}_t(\mathbf{m}_i)]^2/37}$ and coverage of the distributed lag effects by 95% pointwise credible intervals, averaged across observations in effect and no effect subgroups. We also calculated the probability that the model detects a true critical window (true positive, TP) when $\theta_t(\mathbf{m}_i)$ is nonzero as well as the probability a critical window is identified where the true effect is zero (false positive, FP), both using the 95% pointwise credible interval for the distributed lag effects. Finally we report the mean square prediction error (MSPE), $n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. We report MSPE as the ratio of MSPE for each model relative to the treed DLM without modification using 5,000 additional out-of-sample observations for each simulation replicate. The MSPE differences are small across all models because the signal of the HDLM is small relative to the fixed effects and residual variance.

The nested tree HDLM yields distributed lag estimates with smaller RMSE than those from shared tree and Gaussian process HDLMs across all error settings. For critical window identification, the nested tree HDLM yields a similar or higher TP than the other models in all error settings. The FP of the nested tree HDLM ranges from 0.02 in the lowest error setting to 0.08 in the highest error setting; the shared tree HDLM has FP ranging from 0.04 to 0.09, respectively. The added flexibility of the nested tree HDLM allowed for different change points for critical windows and varying smoothness in the effect and no effect groups, features lacking in the shared tree and Gaussian process HDLMs. Coverage of the distributed lag function was near nominal levels except in the highest error setting, where coverage decreased in all HDLMs. The decreased coverage was only evident in the subgroups with a true exposure effect and is due to two factors: shrinkage of the exposure effect and in some cases combining the two groups with different critical windows

Table 4.1: Simulation results for estimating the DLM in scenario 1 (early/late effect). Results are considered pointwise across the DLM for each individual and broken down for individuals with a zero versus nonzero effect. MSPE is calculated for the response using 5,000 out of sample observations and divided by the MSPE for treed DLM without modification.

	Effect ($z_{i1} > 0$)				No Effect ($z_{i1} \leq 0$)			
Model	RMSE*	Coverage	TP	FP	RMSE*	Coverage	FP	MSPE
$\sigma^2 = 10$								
Nested Tree HDLM	4.36	0.95	0.94	0.02	2.32	1.00	0.00	0.938
Shared Tree HDLM	5.30	0.93	0.94	0.04	2.70	0.99	0.01	0.940
Gaussian Process HDLM	5.63	0.95	0.97	0.02	3.51	0.99	0.01	0.941
Nested Tree: Truth	3.08	0.97	0.98	0.02	1.21	1.00	0.00	0.933
Gaussian Process: Truth	5.24	0.95	1.00	0.03	3.05	1.00	0.00	0.941
Treed DLM	11.26	0.63	0.92	0.20	5.58	0.64	0.36	1.000
Gaussian Process DLM	11.28	0.65	0.84	0.17	5.61	0.69	0.31	0.998
$\sigma^2 = 25$								
Nested Tree HDLM	6.72	0.91	0.88	0.04	2.67	1.00	0.00	0.977
Shared Tree HDLM	7.39	0.90	0.86	0.05	3.18	0.99	0.01	0.978
Gaussian Process HDLM	7.14	0.95	0.90	0.02	3.94	1.00	0.00	0.978
Nested Tree: Truth	5.29	0.96	0.94	0.02	1.96	1.00	0.00	0.974
Gaussian Process: Truth	6.43	0.96	0.97	0.02	3.63	1.00	0.00	0.987
Treed DLM	11.45	0.64	0.84	0.18	5.52	0.68	0.32	1.000
Gaussian Process DLM	11.49	0.70	0.61	0.11	5.55	0.78	0.22	0.999
$\sigma^2 = 50$								
Nested Tree HDLM	9.70	0.80	0.71	0.08	3.37	1.00	0.00	0.993
Shared Tree HDLM	9.86	0.79	0.71	0.09	3.64	0.99	0.01	0.993
Gaussian Process HDLM	9.29	0.90	0.50	0.02	4.30	1.00	0.00	0.992
Nested Tree: Truth	7.26	0.92	0.91	0.04	2.56	1.00	0.00	0.989
Gaussian Process: Truth	7.53	0.96	0.83	0.01	4.04	1.00	0.00	0.990
Treed DLM	11.70	0.66	0.77	0.16	5.39	0.71	0.29	1.000
Gaussian Process DLM	11.73	0.74	0.40	0.06	5.46	0.87	0.13	1.000

RMSE* = RMSE$\times$100

into a single group. Coverage in the no effect group was above the nominal level in all settings. The nested tree DLM also had similar or lower MSPE than competing methods.

The treed DLM model with subgroups fixed at the truth outperformed the HDLM models. However, the fixed subgroup Gaussian process approach had higher RMSE and MSPE than the nested tree HDLM in the lowest error scenario. This is first due to the fact that the true distributed lag function is not smooth. Second, the smoothness assumption of the Gaussian process was homogeneous across all subgroups, leading to over-smoothing in the effect subgroups and under-smoothing in the no effect subgroup. These results motivate the use of the treed DLM approaches when considering heterogeneity. The DLM methods with no effect modification were consistently

the worst performing models with low coverage of the distributed lag function, higher FP rates, and the highest RMSE and MSPE.

In Appendix C.4 we present modifier posterior inclusion probabilities (PIP) for an individual modifier or interactions of modifiers. The PIP for an individual modifier is the probability that the modifier is used in at least one splitting rule across the ensemble of trees. We define a modifier interaction to be when two modifiers are used as consecutive splitting rules in the same tree. The average modifier PIPs will be a function of the number of modifiers and number of trees in the model with larger PIPs for fewer modifier or more trees. Modifiers with a larger PIP relative to other modifiers represent possible modification of the distributed lag effects and critical windows, and give a starting point to comparing the exposure effects for individuals or subgroups.

In scenario 1, the true modifiers (z_1 and z_2) have PIPs that range from 0.97 to 1 in the low and middle noise scenarios and 0.79 to 0.99 in the high noise scenario. The other modifiers have PIPs ranging from 0.59 to 0.63, on average. The modifier that determines the critical window placement, z_2 has slightly lower PIP than the modifier that determines effect versus no effect groups, z_1. For scenario 1, we found an interaction between modifiers z_1 and z_2 to have PIP of 1 in the low error setting, 0.88 in the middle error setting, and 0.46 in the highest error setting. The average PIP for other modifier interactions was 0.11. We note that a model with more trees or fewer modifiers would likely find higher PIPs for irrelevant modifiers because there are more chances for a splitting rule to use each modifier.

4.4.2 Scenario 2: Scaled Effect

In scenario 2 we simulated the heterogeneous distributed lag function based on two continuous covariates z_{i1} and z_{i3},

$$\theta_t(\mathbf{m}_i) = \begin{cases} z_{i3}\mathbb{I}(t \in [11, 18]) & \text{if } z_{i1} > 0 \\ 0 & \text{if } z_{i1} \leq 0 \end{cases}. \tag{4.9}$$

The distributed lag function is nonzero during weeks $11 - 18$ for the first group ($z_{i1} > 0$) and scaled by the modifier z_{i3}. The second group ($z_{i1} \leq 0$) had no distributed lag effects. Here, we did

not compare to a fixed subgroups model because the continuous modification does not allow for true subgroups.

Table 4.2 summarizes model performance in term of estimation of the distributed lag function and out-of-sample MSPE. The shared tree HDLM yielded lower RMSE on the distributed lag function in low and middle error settings, higher TP rate for identifying windows, and similar MSPE compared to other HDLM approaches. The added flexibility of the nested tree model was not needed, which leads to slightly lower performance of this approach in this scenario. In order for the Gaussian process model to perform better, the smoothness of the DLM would need to vary due to the modifier z_3, which is not possible in the current method.

Table 4.2: Simulation results for estimating the DLM in scenario 2 (scaled effect). Results are considered pointwise across the DLM for each individual and broken down for individuals with a zero versus nonzero effect. MSPE is calculated for the response using 5,000 out of sample observations.

	Effect ($z_{i1} > 0$)				No Effect ($z_{i1} \leq 0$)			
Model	RMSE*	Coverage	TP	FP	RMSE*	Coverage	FP	MSPE
$\sigma^2 = 10$								
Nested Tree HDLM	5.05	0.92	0.77	0.01	2.30	0.99	0.01	0.920
Shared Tree HDLM	4.60	0.93	0.86	0.01	2.37	0.99	0.01	0.919
Gaussian Process HDLM	6.23	0.94	0.69	0.02	3.58	1.00	0.00	0.923
Treed DLM	10.37	0.81	0.97	0.01	6.04	0.78	0.22	1.000
Gaussian Process DLM	10.71	0.82	0.99	0.02	5.94	0.77	0.23	1.000
$\sigma^2 = 25$								
Nested Tree HDLM	6.62	0.92	0.66	0.01	2.59	1.00	0.00	0.969
Shared Tree HDLM	6.53	0.93	0.76	0.01	2.85	0.99	0.01	0.969
Gaussian Process HDLM	7.53	0.94	0.63	0.01	3.98	1.00	0.00	0.971
Treed DLM	10.92	0.82	0.92	0.02	5.72	0.78	0.22	1.000
Gaussian Process DLM	11.12	0.84	0.86	0.01	5.71	0.80	0.20	1.000
$\sigma^2 = 50$								
Nested Tree HDLM	8.48	0.91	0.62	0.01	3.10	1.00	0.00	0.986
Shared Tree HDLM	8.59	0.91	0.64	0.02	3.42	0.99	0.01	0.987
Gaussian Process HDLM	9.02	0.93	0.53	0.00	4.26	1.00	0.00	0.988
Treed DLM	11.54	0.81	0.86	0.03	5.30	0.79	0.21	1.000
Gaussian Process DLM	11.54	0.85	0.58	0.01	5.44	0.87	0.13	1.000

RMSE* = RMSE$\times$100

Appendix C.4 presents PIPs for individual modifiers and modifier interactions. The modifier tree HDLMs correctly distinguish the true modifiers in all error settings with PIPs ranging from 0.87 to 1. The other modifiers have PIPs ranging from 0.55 to 0.63, on average. In the highest

error setting the modifier responsible for effect size change, z_3, has slightly lower PIP (0.88) than the modifier responsible for effect or no effect subgroups, z_1 (1). We found the interaction between modifiers z_1 and z_3 to have PIP of 1 in the low error setting, 0.91 in the middle error setting, and 0.65 in the highest error setting. The average PIP for other modifier interactions was 0.11.

4.4.3 Scenario 3: No Effect Heterogeneity

In the final simulation scenario we considered a DLM without effect modification. The distributed lag effects were defined $\theta_t(\mathbf{m}_i) = \max[0, (t-s)(s-t+9)]$, where s is a starting time drawn uniformly from $\{1, \ldots, T-9\}$. That is, the distributed lag function is smooth with a critical window over an 8-week time period. The distributed lag effect is identical for all observations in a given dataset.

Table 4.3 presents results for estimation of the distributed lag function. The HDLM performs on par with standard DLM methods without heterogeneity. The HDLM methods have higher RMSE than the standard TDLM and similar to the Gaussian process DLM. Coverage of the true distributed lag effects by 95% pointwise credible intervals meet the nominal level. The TP rates from the HDLM methods are slightly below those from the DLM methods without effect heterogeneity while the FP rate is near zero for all methods. The tree-based HDLMs outperformed the Gaussian process DLM in the higher noise setting; these results encourage the use of tree-based approaches for estimating a DLM or HDLM in general. In terms of modifier selection, the average PIP of any given modifier ranges from 0.63 to 0.69. This is similar to the non-active modifiers in simulation scenarios 1 and 2. Full details of modifier inclusion are given in Appendix C.4.

4.5 Data Analysis

We applied the nested tree, shared tree, and Gaussian process HDLMs to estimate the relationship between BWGAZ and a mother's exposure to $PM_{2.5}$ during the first 37 weeks of pregnancy. We allowed for effect heterogeneity due to ten modifiers and controlled for all demographic, spatial, and temporal covariates described in Section 3.2. The covariates and modifiers are outlined

Table 4.3: Simulation results for estimating the DLM in scenario 3 (no effect heterogeneity). Results are considered pointwise across the DLM for each individual. MSPE is calculated for the response using 5,000 out of sample observations and divided by the MSPE for TDLM without modification.

Model	RMSE×100	Coverage	TP	FP	MSPE
$\sigma^2 = 10$					
Nested Tree HDLM	2.90	0.93	0.83	0.00	1.001
Shared Tree HDLM	2.83	0.94	0.83	0.00	1.001
Gaussian Process HDLM	2.86	1.00	0.94	0.00	1.001
Treed DLM	2.37	0.97	0.89	0.01	1.000
Gaussian Process DLM	2.81	0.99	0.96	0.01	1.007
$\sigma^2 = 25$					
Nested Tree HDLM	3.42	0.97	0.76	0.00	1.001
Shared Tree HDLM	3.35	0.97	0.76	0.00	1.001
Gaussian Process HDLM	3.84	1.00	0.76	0.00	1.001
Treed DLM	3.01	0.98	0.82	0.01	1.000
Gaussian Process DLM	3.64	0.99	0.89	0.01	1.003
$\sigma^2 = 50$					
Nested Tree HDLM	4.31	0.98	0.65	0.00	1.001
Shared Tree HDLM	4.21	0.98	0.67	0.00	1.001
Gaussian Process HDLM	4.73	1.00	0.54	0.00	1.001
Treed DLM	3.76	0.97	0.79	0.01	1.000
Gaussian Process DLM	4.41	0.99	0.80	0.01	1.001

in Table 4.4. We did not include a fixed effect for fetal sex as the outcome, BWGAZ, was already adjusted for this factor. Each model ran for 15,000 iterations after 5,000 burn-in and was thinned to every 5^{th} iteration. Following a 10-fold cross-validation, the shared tree HDLM had slightly lower MSPE. Because the results from the nested tree HDLM were similar, we defaulted to the shared tree HDLM for the data analysis presented in the main text because it is a simpler model. The Gaussian process HDLM had higher MSPE and higher uncertainty in the DLM estimates. Results of our cross-validation and results from the nested tree and Gaussian process HDLMs are provided in Appendix C.5. For comparison, we also modeled the exposure-time-response using a treed DLM without effect modification.

4.5.1 DLM without effect modification

Figure 4.2 shows the estimated exposure effect using a treed DLM with no effect modification. We found that increased $PM_{2.5}$ exposure was associated with decreased BWGAZ during each week of gestation. The DLM identifies critical windows during weeks 5-6 and 31-34. The cumulative effect of an inter-quartile range (IQR) increase in $PM_{2.5}$ at every week of pregnancy corresponds

Table 4.4: Covariates included as fixed effects or modifiers (indicated by checks) of the HDLM. For covariates included as modifiers we report the posterior inclusion probability (PIP), which is defined as the probability the modifier is used in at least one splitting rule in the ensemble.

Covariate	Type	Mean (IQR)	Categories	Fixed effect	Modifier	PIP
Age at conception	Continuous	28.7 (24 – 33)		✓	✓	0.93
Height	Continuous	64.4 (62 – 66)		✓		–
Prior weight	Continuous	151.3 (126 – 169)		✓		–
Body mass index	Continuous	25.7 (21.6 – 28.4)		✓	✓	0.95
Avg. temp/trimester	Continuous	51.6 (38.2 – 65.1)		✓		–
Income range	Ordinal		6	✓	✓	0.74
Highest education	Ordinal		5	✓	✓	0.90
Smoking habits	Ordinal		4	✓	✓	0.78
Marital status	Categorical		6	✓	✓	0.50
Prenatal care	Categorical		3	✓	✓	0.48
Race	Categorical		4	✓	✓	0.61
County of residence	Categorical		12	✓		–
Month of conception	Categorical		12	✓		–
Year of conception	Categorical		9	✓		–
Hispanic	Binary		2	✓	✓	0.95
Sex of child	Binary		2		✓	0.64

to a decrease in BWGAZ of -0.026 (95% CI: $-0.044, -0.006$). In a more interpretable context, an increase from 5.9 to 8.5 $\mu g/m^3$ $PM_{2.5}$ (the 25^{th} and 75^{th} weekly exposure percentiles) relates to an approximate decrease in birth weight of 11.3g; this is approximate because BWGAZ adjusts for sex and gestational age.

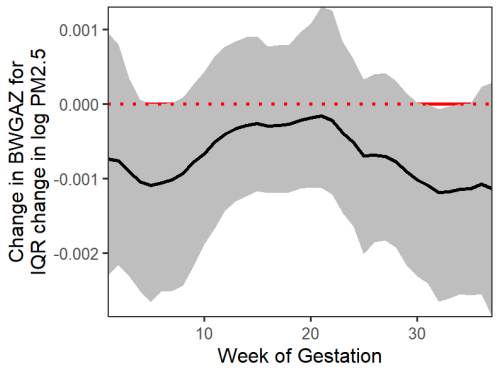
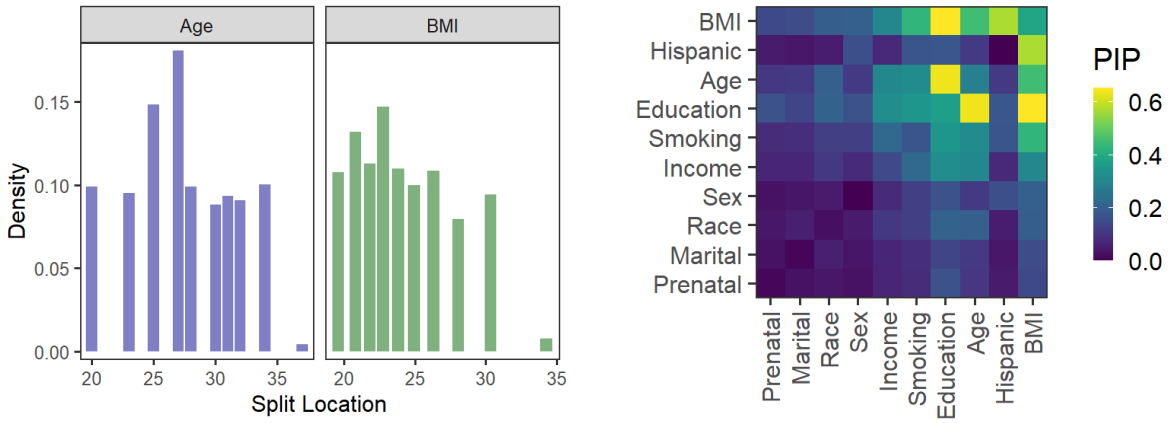


Figure 4.2: Estimated distributed lag effects due to an IQR increase in $PM_{2.5}$ using the treed DLM with no effect modification. The solid line indicates the posterior mean while the gray area represents a 95% credible interval. Weeks where the credible interval does not contain zero represent critical windows.



(a) (b)

Figure 4.3: Panel (a) shows the density of splitting locations for two continuous modifiers: maternal age and BMI. Panel (b) shows posterior inclusion probability (PIP) of modifier interactions.

4.5.2 Modifier selection to determine susceptible populations

The modifier PIPs from the shared tree HDLM are presented in Table 4.4. The modifiers with the highest PIP include maternal BMI (0.95), Hispanic (0.95), age (0.93) and education (0.90). The next highest PIP modifiers were smoking (0.78) and income (0.74). Considering empirical evidence from our simulation scenarios, with a similar number of modifiers and trees, PIPs below 0.7 do not provide significant evidence of effect modification. For continuous modifiers, Figure 4.3(a) describes the density of splitting values for age and BMI, with modes at 27 and 22.8, respectively.

Figure 4.3(b) illustrates the PIP of interactions between modifiers. In simulation, the average PIP for non-active modifier interactions was 0.11. We found the highest interaction PIPs to be education–BMI (0.65), education–age (0.64), and BMI–Hispanic (0.57).

When we consider effect modification, we can interpret the effect heterogeneity as an interaction between the modifiers and the exposure effect. For instance, a subgroup that is the result of a rule on a single modifier (59% of tree-specified subgroups) is a two-way interaction between that modifier and $PM_{2.5}$. The distributed lag effects for a subgroup that is the result of rules on two modifiers (33% of tree-specified subgroups) is a three-way interaction between the two modifiers as well as exposure to $PM_{2.5}$. While the ensemble of trees should be able to incorporate interactions via the additive nature of multiple trees, some complex interactions, such as the effect/no

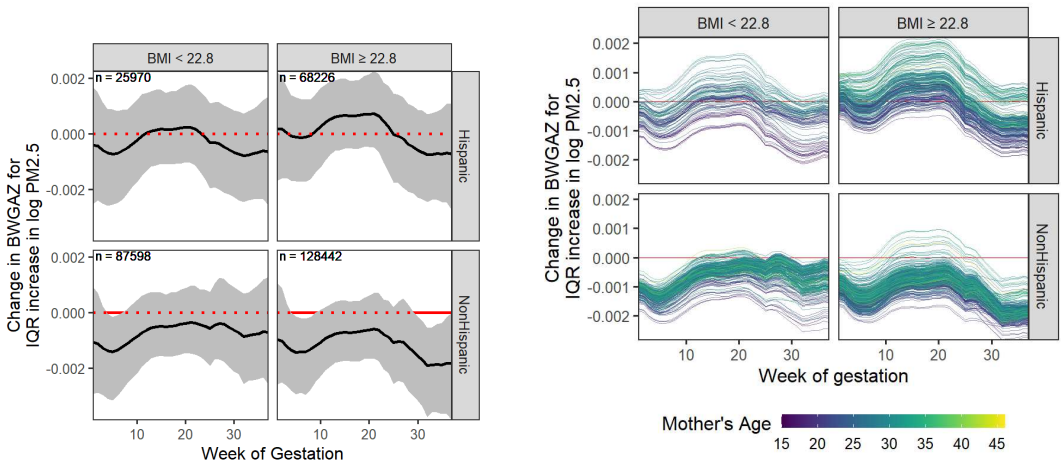
effect partition in our simulations, may only be captured with a single tree that splits on multiple modifiers.

The modifier PIPs provide a starting point to explore potential susceptible populations, which are characterized by their differential effects to exposure. In this analysis, we focus on four modifiers: maternal BMI, Hispanic, age, and education, along with interactions between these modifiers. To simplify visualization of the results, we divided continuous modifiers, BMI and age, at the splitting value modes. We present subgroup specific average effects along with personalized distributed lag effect estimates for a sample of individuals in each subgroup to give a sense of the remaining within-group heterogeneity.

4.5.3 Subgroup-specific distributed lag effects

Figures 4.4(a) and 4.5(a) compare subgroup-specific DLM estimates across pairs of modifiers using the posterior analysis technique described in Appendix C.3.5. These subgroups are based on some of the largest modifier PIPs. Other two-way combinations showed little or no evidence of different subgroup distributed lag effects. More complex multi-way interactions are presented in Appendix C.5. A consistent theme across subgroup analyses is a differential effect for Hispanic and non-Hispanic subgroups. We see a consistent negative exposure effect at all time points for the non-Hispanic groups and an early and late critical window that is present in many non-Hispanic subgroups. For Hispanic subgroups the exposure effect hovers around zero, which is generally consistent with no exposure effect.

Figure 4.4(a) offers evidence of early-gestation susceptibility among all non-Hispanics. For non-Hispanics with BMI above 22.8, we detect a second critical window later in gestation during which increased $PM_{2.5}$ exposure is associated with lower birth weight. We also find differences in critical windows among non-Hispanics based on their level of education, seen in Figure 4.5(a). Non-Hispanics with less than a college education show a late-gestation critical window. Having at most a high school education suggests the presence of an early-gestation critical window for non-Hispanics. For Hispanics, there is no evidence of an association at any level of education.



(a) (b)

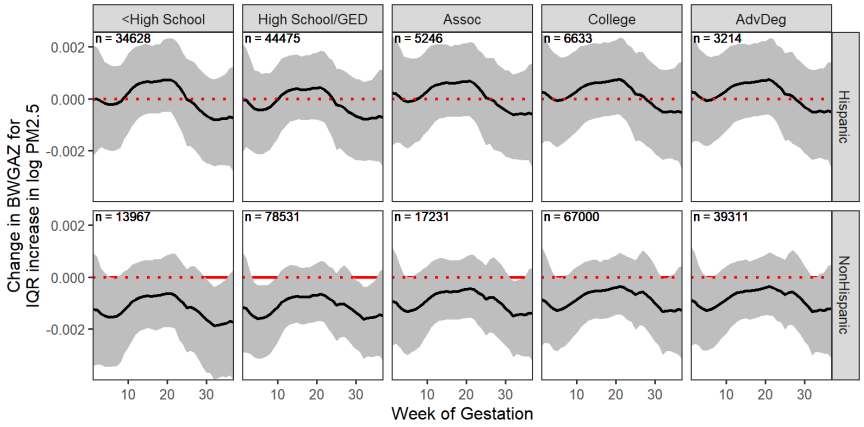
Figure 4.4: Panel (a) shows subgroup-specific DLM estimates with 95% credible intervals. Panel (b) shows DLM estimates for 1,000 individuals from our data analysis. Both panels are grouped by Hispanic designation (rows) and BMI above/below 22.8 (columns). Panel (b) DLMs are colored according to a Mother's age at conception with lighter color representing older individuals.

4.5.4 Personalized distributed lag effect estimates

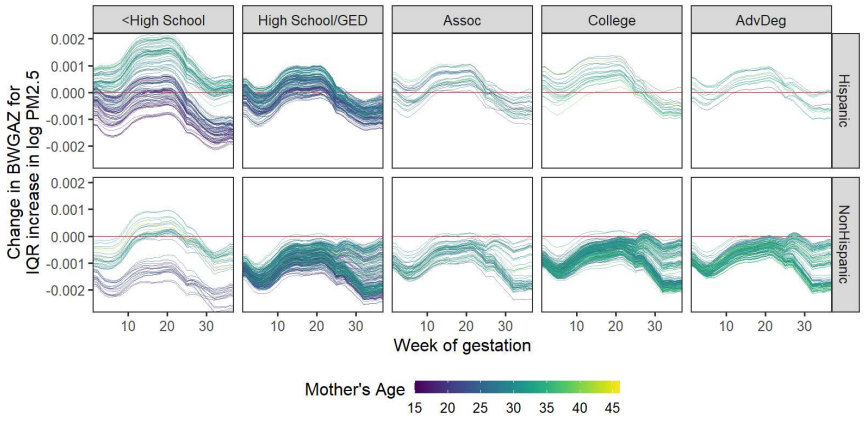
The subgroup analysis in Section 4.5.3 highlights broad trends across modifiers and identifies several susceptible subgroups: non-Hispanic, BMI above 22.8, and less than college education. However, so far we have considered only two variables at a time. In reality, the individuals in these subgroups vary across the entire observed range of the other modifiers considered in this analysis. In this section, we explore remaining variability in the distributed lag effects among individuals within these subgroups.

Figures 4.4(b) and 4.5(b) show the estimated DLMs for 1,000 individuals randomly selected from our data set. Estimates for these individuals have been grouped according to the subgroup analysis in Section 4.5.3 and colored to highlight differences in the distributed lag effects according to a third modifier. This third modifier was selected based on noticeable variation in distributed lag effects among individuals; however, other modifiers are also responsible for these differences.

Figure 4.4(b) colors the average distributed lag effects based on the age of the individual with darker color representing younger mothers. We see a trend towards younger individuals having a larger negative effect. This difference by age is more noticeable for Hispanics with younger



(a)



(b)

Figure 4.5: Panel (a) shows subgroup-specific DLM estimates with 95% credible intervals, divided by Hispanic and education modifiers. Panel (b) shows DLM estimates for 1,000 individuals from our data analysis divided by Hispanic and education modifiers and colored by age, with lighter color representing older women.

individuals showing a more consistently negative effect. Figure 4.5(b) visualizes differences by education and Hispanic designation with color again representing continuous modification by maternal age. The distributed lag effects show pronounced differences between younger and older individuals with less than a high school degree. Compared to their older counterparts, young Hispanic and non-Hispanic women with less than a high school education have consistently larger negative effects. Younger women with less education appear to be a highly susceptible group to $PM_{2.5}$ exposure.

4.5.5 Cumulative effect estimates and four-way interactions

We next explored the differences in total exposure susceptibility across three modifiers simultaneously: maternal BMI, age, and Hispanic designation. We randomly selected 5,000 individuals from our data set and calculated the cumulative effect, or the sum of week-specific effects associated with an IQR increase in $PM_{2.5}$ exposure throughout pregnancy. The estimated cumulative effect was averaged across bins of the continuous modifiers, BMI and age. Results are visualized in Figure 4.6. For non-Hispanic individuals, higher BMI and lower age correspond to the larger negative effects. Individuals who are older and have lower BMI had the smallest cumulative effect of $PM_{2.5}$. Consistently across the non-Hispanic subgroup, increased exposure was related to lower BWGAZ. For Hispanic individuals, the cumulative effect of exposure was centered around zero. We see evidence of a larger negative effect of exposure for younger Hispanic women; however this effect is less than the cumulative effect for non-Hispanics of the same age and BMI. There is little evidence of modification by BMI for Hispanics.

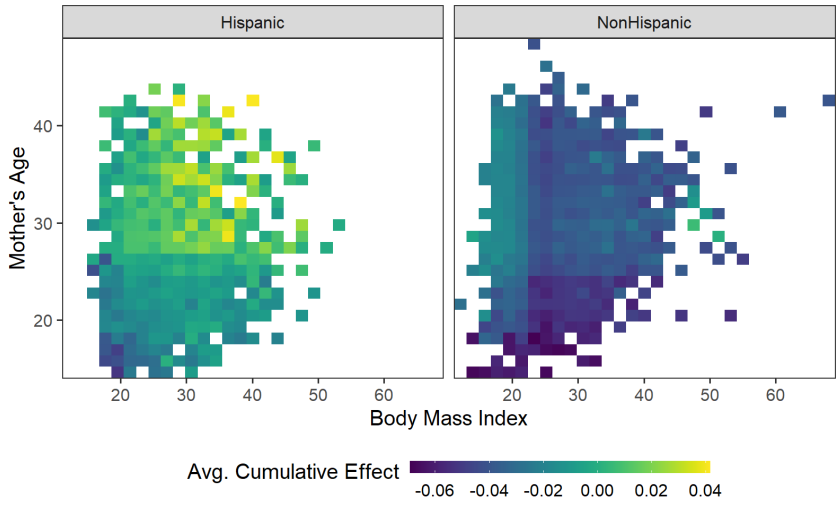


Figure 4.6: Heat map of average cumulative exposure effect for 5,000 individuals from data analysis. Each colored block is the average cumulative effect for individuals with a particular body mass index (x-axis), age (y-axis) and Hispanic designation (panels). Darker color indicates a larger negative effect on BWGAZ associated with an IQR increase in $PM_{2.5}$ exposure throughout pregnancy.

The cumulative effect of $PM_{2.5}$ exposure on BWGAZ for non-Hispanic individuals in this sample ranged from -0.079 to 0.007, while for Hispanic individuals the cumulative effect ranged from -0.052 to 0.051. In the context of birth weight, the cumulative effect of an IQR increase in $PM_{2.5}$ exposure ranges from -34.3g to 3.2g for non-Hispanic individuals in the sample and from -22.8g to 22.1g for Hispanic individuals. These ranges are approximate because BWGAZ is adjusted for sex and gestational age.

4.5.6 Model utility

In contrast to the population average effect described in Section 4.5.1, the HDLM analysis finds evidence of heterogeneity across the population. In particular, we identified several subgroups with increased susceptibility to $PM_{2.5}$ exposure. These susceptible subgroups include non-Hispanic mothers who are younger or have higher BMI as well as non-Hispanic mothers with lower educational attainment. Using HDLM, we isolated critical windows for these subgroups finding early- and late-gestation to be time periods of higher susceptibility. A personalized analysis found additional variability within the subgroups due to the remaining modifiers. Critical windows for some individuals were much wider than suggested by the population average effect, and for other individuals there was little evidence of any exposure effects. The cumulative effect estimates showed a wide range of variability in the total change in birth weight due to increased $PM_{2.5}$ exposure. Compared to the population average, some individuals showed a 3 times larger decrease in birth weight due to an IQR increase in $PM_{2.5}$ exposure across pregnancy. For susceptible individuals or populations, the HDLM gives the ability to create targeted interventions for precision environmental health. Personalized understanding of exposure effects on health also gives rise to streamlined scientific understanding of biological mechanisms due to exposure.

4.6 Discussion

We proposed a framework for estimating effect heterogeneity in a distributed lag function due to a possibly high dimensional set of individual modifiers. HDLM can estimate personalized criti-

cal windows and effect sizes as well as perform modifier variable selection. In addition, we extend the BART framework by allowing for modification in a multivariate predictor and define the nested tree model composed of one regression tree for modification and a second regression tree unique to a subgroup for estimating the distributed lag effects.

The nested tree and shared tree HDLMs outperform the Gaussian process HDLM in a simulation study. When effect heterogeneity exists, all HDLM methods outperform methods that assume homogeneity in the distributed lag effects, especially in subgroups with no effect. The difference between standard DLM and HDLM highlights the bias incurred in estimating the distributed lag function both for individuals with and without exposure effects. We also show that our methods consistently estimate high PIPs for modifiers responsible for changes in the distributed lag effects relative to the other potential modifiers. This differentiation in PIPs provides a pathway for selecting modifiers to explore when applying this method. Additional testing of our model suggested modifier selection is insensitive to misspecification of the fixed effects.

We applied the shared tree HDLM to estimate personalized distributed lag effects due to $PM_{2.5}$ exposure in a Colorado, USA birth cohort. This analysis considered ten modifiers of different data types. The model identified four potential modifiers and posterior analysis showed variations in the distributed lag effects for different levels of these modifiers. In particular, we note changes in a late-term critical window due to BMI, overall changes in magnitude due to Hispanic designation, and larger negative effects for young individuals with less than a high school education. We also explored strategies to summarize subgroup-specific distributed lag effects to compare broad differences across the population.

Our proposed methods for estimating an HDLM provide new approaches to exploring individual differences in an exposure-time-response. These tools can lead to personalized environmental health decision making and pinpoint at-risk individuals for intervention. Furthermore, personalized exposure-response functions provide stakeholders with detailed information regarding the burden of air pollution on health.

Chapter 5

Conclusion

Maternal exposure to environmental chemicals has been linked to numerous detrimental birth and children's health outcomes. To better understand how exposure impacts fetal development, research has focused on identifying critical windows, or time periods of development during which increased exposure is associated with changes in the resulting health outcomes of the child. Understanding critical windows can give insight into the biological mechanisms impacted by exposure. The primary tool for identifying critical windows is the distributed lag model (DLM), which can account for temporal correlation in high resolution exposure data and identify critical windows that may not align with clinically-defined periods of development. Research and statistical methods able to show how and when changes in exposure induce changes in fetal development can inform new policy regarding air pollution standards, which have the potential to improve health across a population.

In this dissertation we proposed novel methods extending DLMs using the Bayesian additive regression trees (BART) framework. In Chapter 2, we proposed the treed distributed lag nonlinear model (TDLNM) for precisely identifying critical windows when the effects are assumed to be nonlinear at each time point of exposure. We showed that TDLNM outperformed competing spline-based approaches. Furthermore, TDLNM can incorporate exposure measurement uncertainty into the model. Previous implementations of DLNMs have not been able to accommodate exposure uncertainty into the model.

In Chapter 3, we developed a novel tree-based method for DLM estimation that outperformed spline and Gaussian process methods. We also defined the distributed lag mixture model (DLMM) for estimating interactions between a mixture of exposures, and extended our treed DLM method to estimating a DLMM. The DLMM incorporates interactions at different time points for two exposures, which allows for study of the priming or two-hit hypotheses in environmental health epidemiology. Our proposed method for estimating the DLMM performs exposure selection to

determine which exposures are associated with the outcome and shrinks the effects of irrelevant exposures or interactions. We showed that our method precisely identifies critical windows for the marginal effect of exposures and selects the correct exposures. Furthermore, our model allows for nonlinear exposure effects using within-exposure interactions at the same time points.

We defined the heterogeneous DLM (HDLM) in Chapter 4, which allows the distributed lag effects and critical windows to vary across a population due to individual demographic or neighborhood characteristics. We proposed three methods for estimating the HDLM using traditional regression tree framework combined with Gaussian process or our treed DLM approach from Chapter 3 in what we define as the nested tree or shared tree HDLM. We show that the nested tree HDLM outperforms other methods when the critical windows vary across the population while the shared tree method often does better when the distributed lag effects vary and the critical windows stay the same. Our methods have high precision in identifying critical windows. In addition, our methods for estimating the HDLM perform variable selection to determine characteristics responsible for changes to the distributed lag effects from a possibly high dimensional set of effect modifiers.

Our methods were applied to a Colorado-based administrative birth cohort. We found critical windows of susceptibility for birth weight due to changes in air pollution and temperature exposure. Our DLNM analysis found significant decreases in birth weight associated with increased fine particulate matter ($PM_{2.5}$) levels. Importantly, we observed an association between $PM_{2.5}$ exposure and birth weight at low levels of exposure including those below the current national ambient air quality standards. An analysis of weekly exposure to four pollutants and temperature found critical windows for $PM_{2.5}$, sulfur dioxide (SO_2), carbon monoxide (CO) and temperature and an interaction between $PM_{2.5}$ and temperature suggesting a priming effect. We also developed an approach to adjust for the temporal correlation between exposures to assess the impact of an individual exposure on the outcome. We found persistent decreases in birth weight due to increases in $PM_{2.5}$, SO_2 and temperature. Finally, our heterogeneous DLM analysis accounted for ten potential individual-specific modifiers of the distributed lag effect of $PM_{2.5}$ exposure on birth weight.

We found non-Hispanic and higher body mass index to be populations with higher susceptibility to $PM_{2.5}$ exposure, and individual in these groups had larger negative changes in birth weight due to increased $PM_{2.5}$ during early- and late-gestation compared to other individuals in the study. The results of our data analyses showed overwhelming evidence that increased air pollution is associated with decreased birth weight. These findings reinforce the public health impact of maternal air pollution exposure during pregnancy.

5.1 Future Work

We propose several directions for future work within the context of Bayesian treed distributed lag models. First, our treed DLNM allows for exposure measurement uncertainty. However, incorporating exposure measurement uncertainty has not been formally tested in simulation and validation of this technique is essential before application in an epidemiological (or other) study. A further extension of the DLNM would assume monotonic effects in the exposure-response at each time-point, which is consistent with our understanding of the effect of many air pollutants on birth outcomes. Our DLMM currently allows only quadratic effects. Incorporating increased nonlinearity and higher order interactions into the DLMM is a logical next step. One challenge in nonlinear mixture effects is the increased complexity of the problem when considering nonlinear interactions across a large number of exposures with many repeated measurements. Another consideration to extend the DLMM in incorporating exposures with differing measurement schedules or thinking about measurements in continuous time versus discrete time. With regard to the heterogeneous DLM, we are also limited to a single exposure with linear effects. Allowing for heterogeneity in a nonlinear exposure-time-response presents challenges in identifiability of the effect modifiers when the fixed effects are misspecified. Further work is required to untangle the fixed effects and modifiers responsible for heterogeneity in a DLNM. Developing a method for heterogeneity in a DLMM is also possible within our current Bayesian treed DLM framework.

5.2 Impact

Precise critical window identification is paramount to identifying stages of development and biological mechanisms impacted by air pollution exposure. Our Bayesian treed methods consistently outperform competing DLM techniques in identifying critical windows with high probability as well as estimating exposure effects. Furthermore, we have extended the standard DLM to take advantage of complex data scenarios involving multiple exposures, measurement uncertainty, and individual effect modifiers. Our proposed methods open doors to environmental epidemiology research not previously possible. The methods developed in this thesis have the potential to unlock deeper understanding of the impact of pollution on human development and drive public health policy that leads to beneficial birth outcomes and improved health at the population level.

Bibliography

Anenberg, S. C., Haines, S., Wang, E., Nassikas, N., and Kinney, P. L. (2020). Synergistic health effects of air pollution, temperature, and pollen exposure: a systematic review of epidemiological evidence. *Environmental Health: A Global Access Science Source*, 19(1).

Bell, M. L., Ebisu, K., and Belanger, K. (2007). Ambient Air Pollution and Low Birth Weight in Connecticut and Massachusetts. *Environmental Health Perspectives*, 115(7):1118–1124.

Berrocal, V. J., Gelfand, A. E., Holland, D. M., and Statisti-Cian, S. (2010). A Spatio-Temporal Downscaler for Output From Numerical Models. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(2):176–197.

Bolton, J. L., Auten, R. L., and Bilbo, S. D. (2014). Prenatal air pollution exposure induces sexually dimorphic fetal programming of metabolic and neuroinflammatory outcomes in adult offspring. *Brain, Behavior, and Immunity*, 37:30–44.

Bose, S., Chiu, Y. H. M., Hsu, H. H. L., Di, Q., Rosa, M. J., Lee, A., Kloog, I., Wilson, A., Schwartz, J., Wright, R. O., Cohen, S., Coull, B. A., and Wright, R. J. (2017). Prenatal nitrate exposure and childhood asthma. Influence of maternal prenatal stress and fetal sex. *American Journal of Respiratory and Critical Care Medicine*, 196(11):1396–1403.

Bose, S., Rosa, M. J., Mathilda Chiu, Y.-H., Leon Hsu, H.-H., Di, Q., Lee, A., Kloog, I., Wilson, A., Schwartz, J., Wright, R. O., Morgan, W. J., Coull, B. A., and Wright, R. J. (2018). Prenatal nitrate air pollution exposure and reduced child lung function: Timing and fetal sex effects. *Environmental Research*, 167:591–597.

Bosetti, C., Nieuwenhuijsen, M. J., Gallus, S., Cipriani, S., La Vecchia, C., and Parazzini, F. (2010). Ambient particulate matter and preterm birth or birth weight: A review of the literature. *Archives of Toxicology*, 84(6):447–460.

Buckley, J. P., Hamra, G. B., and Braun, J. M. (2019). Statistical approaches for investigating periods of susceptibility in children's environmental health research. *Current Environmental Health Reports*, 6(1):1–7.

Carlin, B. P. and Chib, S. (1995). Bayesian model choice via markov chain monte carlo Methods. *Journal of the Royal Statistical Society. Series B*, 57(3):473–484.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Chang, H. H., Reich, B. J., and Miranda, M. L. (2012). Time-to-event analysis of fine particle air pollution and preterm birth: Results from North Carolina, 2001-2005. *American Journal of Epidemiology*, 175(2):91–98.

Chang, H. H., Warren, J. L., Darrow, L. A., Reich, B. J., Waller, L. A., and Chang, H. H. (2015). Assessment of critical exposure and outcome windows in time-to-event analysis with application to air pollution and preterm birth study. *Biostatistics*, 16(3):509–521.

Chen, Y. H., Mukherjee, B., and Berrocal, V. J. (2019). Distributed lag interaction models with two pollutants. *Journal of the Royal Statistical Society. Series C*, 68(1):79–97.

Chipman, H. A., George, E. I., and McCulloch, R. E. (1998). Bayesian CART model search. *Journal of the American Statistical Association*, 93(443):935–948.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2002). Bayesian treed models. *Machine Learning*, 48(1-3):299–320.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 6(1):266–298.

Chiu, Y. H., Bellavia, A., James-Todd, T., Correia, K. F., Valeri, L., Messerlian, C., Ford, J. B., Mínguez-Alarcón, L., Calafat, A. M., Hauser, R., and Williams, P. L. (2018). Evaluating effects

of prenatal exposure to phthalate mixtures on birth weight: A comparison of three statistical approaches. *Environment International*, 113:231–239.

Chiu, Y.-H. M., Hsu, H.-H. L., Coull, B. A., Bellinger, D. C., Kloog, I., Schwartz, J., Wright, R. O., and Wright, R. J. (2016). Prenatal particulate air pollution and neurodevelopment in urban children: Examining sensitive windows and sex-specific associations. *Environment International*, 87:56–65.

Chiu, Y.-H. M., Hsu, H.-H. L., Wilson, A., Coull, B. A., Pendo, M. P., Baccarelli, A., Kloog, I., Schwartz, J., Wright, R. O., Taveras, E. M., and Wright, R. J. (2017). Prenatal particulate air pollution exposure and body composition in urban preschool children: Examining sensitive windows and sex-specific associations. *Environmental Research*, 158:798–805.

Davalos, A. D., Luben, T. J., Herring, A. H., and Sacks, J. D. (2017). Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures. *Annals of Epidemiology*, 27(2):145–153.

Deshpande, S. K., Bai, R., Balocchi, C., and Starling, J. E. (2020). VCBART: Bayesian trees for varying coefficients. *arXiv preprint arXiv:2003.06416*, pages 1–44.

Dugandzic, R., Dodds, L., Stieb, D., and Smith-Doiron, M. (2006). The association between low level exposures to ambient air pollution and term low birth weight: A retrospective cohort study. *Environmental Health*, 5:1–8.

Ebisu, K. and Bell, M. L. (2012). Airborne PM2.5 chemical components and low birth weight in the Northeastern and mid-Atlantic regions of the United States. *Environmental Health Perspectives*, 120(12):1746–1752.

Fenton, T. R. and Kim, J. H. (2013). A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatrics*, 13(1).

Figueroa-Romero, C., Mikhail, K. A., Gennings, C., Curtin, P., Bello, G. A., Botero, T. M., Goutman, S. A., Feldman, E. L., Arora, M., and Austin, C. (2020). Early life metal dysregulation in amyotrophic lateral sclerosis. *Annals of Clinical and Translational Neurology*, 7(6):872–882.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234.

Gasparrini, A., Scheipl, F., Armstrong, B., and Kenward, M. G. (2017). A penalized framework for distributed lag non-linear models. *Biometrics*, 73(3):938–948.

Hahn, P. R., Murray, J. S., and Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis*, 15(3):965–1056.

Hastie, T. and Tibshirani, R. (2000). Bayesian backfitting. *Statistical Science*, 15(3):196–223.

Hazlehurst, M. F., Carroll, K. N., Loftus, C. T., Szpiro, A. A., Moore, P. E., Kaufman, J. D., Kirwa, K., LeWinn, K. Z., Bush, N. R., Sathyanarayana, S., Tylavsky, F. A., Barrett, E. S., Nguyen, R. H. N., and Karr, C. J. (2021). Maternal exposure to PM2.5 during pregnancy and asthma risk in early childhood. *Environmental Epidemiology*, 5(2):e130.

Hernán, M. A., Schisterman, E. F., and Hernández-Díaz, S. (2014). Invited commentary: Composite outcomes as an attempt to escape from selection bias and related paradoxes. *American Journal of Epidemiology*, 179(3):368–370.

Horton, M. K., Hsu, L., Henn, B. C., Margolis, A., Austin, C., Svensson, K., Schnaas, L., Gennings, C., Hu, H., Wright, R., Rojo, M. M. T., and Arora, M. (2018). Dentine biomarkers of prenatal and early childhood exposure to manganese, zinc and lead and childhood behavior. *Environment International*, 121:148–158.

Hoskovec, L., Benka-Coker, W., Severson, R., Magzamen, S., and Wilson, A. (2021). Model choice for estimating the association between exposure to chemical mixtures and health outcomes: A simulation study. *Plos one*, 16(3):e0249236.

Jacobs, M., Zhang, G., Chen, S., Mullins, B., Bell, M., Jin, L., Guo, Y., Huxley, R., and Pereira, G. (2017). The association between ambient air pollution and selected adverse pregnancy outcomes in China: A systematic review. *Science of the Total Environment*, 579:1179–1192.

Jakpor, O., Chevrier, C., Kloog, I., Benmerad, M., Giorgis-Allemand, L., Cordier, S., Seyve, E., Vicedo-Cabrera, A. M., Slama, R., Heude, B., Schwartz, J., and Lepeule, J. (2020). Term birthweight and critical windows of prenatal exposure to average meteorological conditions and meteorological variability. *Environment International*, 142(June):105847.

Kapelner, A. and Bleich, J. (2016). bartMachine: Machine learning with bayesian additive regression trees. *Journal of Statistical Software*, 70(1):1–40.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):319–323.

Kloog, I., Melly, S. J., Coull, B. A., Nordio, F., and Schwartz, J. D. (2015). Using satellite-based spatiotemporal resolved air temperature exposure to study the association between ambient air temperature and birth outcomes in Massachusetts. *Environmental Health Perspectives*, 123(10):1053–1058.

Krimsky, S. (2017). The unsteady state and inertia of chemical regulation under the US Toxic Substances Control Act. *PLoS Biology*, 15(12):1–10.

Lamichhane, D. K., Lee, S. Y., Ahn, K., Kim, K. W., Shin, Y. H., Suh, D. I., Hong, S. J., and Kim, H. C. (2020). Quantile regression analysis of the socioeconomic inequalities in air pollution and birth weight. *Environment International*, 142(January):105875.

Lee, A., Leon Hsu, H.-H., Mathilda Chiu, Y.-H., Bose, S., Rosa, M. J., Kloog, I., Wilson, A., Schwartz, J., Cohen, S., Coull, B. A., Wright, R. O., and Wright, R. J. (2018). Prenatal fine particulate exposure and early childhood asthma: Effect of maternal stress and fetal sex. *Journal of Allergy and Clinical Immunology*, 141(5):1880–1886.

Levin-Schwartz, Y., Gennings, C., Schnaas, L., Del Carmen Hernández Chávez, M., Bellinger, D. C., Téllez-Rojo, M. M., Baccarelli, A. A., and Wright, R. O. (2019). Time-varying associations between prenatal metal mixtures and rapid visual processing in children. *Environmental Health*, 18(92):18.

Linero, A. R. and Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 80(5):1087–1110.

Makalic, E. and Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.

Mork, D. and Wilson, A. (2021a). Estimating perinatal critical windows to environmental mixtures via structured Bayesian regression tree pairs. *arXiv preprint arXiv:2102.09071*.

Mork, D. and Wilson, A. (2021b). Treed distributed lag non-linear models. *Biostatistics, in press*.

Morrison, G. C., Weschler, C. J., Bekö, G., Koch, H. M., Salthammer, T., Schripp, T., Toftum, J., and Clausen, G. (2016). Role of clothing in both accelerating and impeding dermal absorption of airborne SVOCs. *Journal of Exposure Science and Environmental Epidemiology*, 26(1):113–118.

Muggeo, V. M. R. (2007). Bivariate distributed lag models for the analysis of temperature-by-pollutant interaction effect on mortality. *Environmetrics*, 18(3):231–243.

Murray, J. S. (2020). Log-linear Bayesian additive regression trees for multinomial logistic and count Regression Models. *Journal of the American Statistical Association*, pages 1–35.

Odden, M. C., Rawlings, A. M., Khodadadi, A., Fern, X., Shlipak, M. G., Bibbins-Domingo, K., Covinsky, K., Kanaya, A. M., Lee, A., Haan, M. N., Newman, A. B., Psaty, B. M., and Peralta, C. A. (2020). Heterogeneous exposure associations in observational cohort studies: The example of blood pressure in older adults. *American Journal of Epidemiology*, 189(1):55–67.

Park, S. K., Zhao, Z., and Mukherjee, B. (2017). Construction of environmental risk score beyond standard linear models using machine learning methods: Application to metal mixtures, oxidative stress and cardiovascular disease in NHANES. *Environmental Health*, 16(1):102–118.

Pratola, M. T., Chipman, H. A., George, E. I., and McCulloch, R. E. (2020). Heteroscedastic BART via multiplicative regression trees. *Journal of Computational and Graphical Statistics*, 29(2):405–417.

Ročková, V., van der Pas, S., et al. (2020). Posterior concentration for bayesian regression trees and forests. *Annals of Statistics*, 48(4):2108–2131.

Schittny, J. C. (2017). Development of the lung. *Cell and tissue research*, 367(3):427–444.

Schwartz, J. D. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology*, 11(3):320–326.

Spanbauer, C. and Sparapani, R. (2021). Nonparametric machine learning for precision medicine with longitudinal clinical trials and Bayesian additive regression trees with mixed models. *Statistics in Medicine*, (January 2020):1–27.

Sparapani, R. A., Logan, B. R., McCulloch, R. E., and Laud, P. W. (2016). Nonparametric survival analysis using Bayesian Additive Regression Trees (BART). *Statistics in Medicine*, 35(16):2741–2753.

Šrám, R. J., Binková, B., Dejmek, J., and Bobak, M. (2005). Ambient air pollution and pregnancy outcomes: A review of the literature. *Environmental Health Perspectives*, 113(4):375–382.

Starling, J. E., Murray, J. S., Carvalho, C. M., Bukowski, R. K., and Scott, J. G. (2020). BART with targeted smoothing: An analysis of patient-specific stillbirth risk. *Annals of Applied Statistics*, 14(1):28–50.

Stieb, D. M., Chen, L., Eshoul, M., and Judek, S. (2012). Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. *Environmental Research*, 117:100–111.

Strickland, M. J., Lin, Y., Darrow, L. A., Warren, J. L., Mulholland, J. A., and Chang, H. H. (2019). Associations between ambient air pollutant concentrations and birth weight: A quantile regression analysis. *Epidemiology*, 30(5):624–632.

Sun, X., Luo, X., Zhao, C., Zhang, B., Tao, J., Yang, Z., Ma, W., and Liu, T. (2016). The associations between birth weight and exposure to fine particulate matter (PM2.5) and its chemical constituents during pregnancy: A meta-analysis. *Environmental Pollution*, 211:38–47.

Tan, Y. V. and Roy, J. (2019). Bayesian additive regression trees and the general bart model. *Statistics in medicine*, 38(25):5048–5069.

Veras, M. M., de Oliveira Alves, N., Fajersztajn, L., and Saldiva, P. (2017). Before the first breath: prenatal exposures to air pollution and lung development. *Cell and tissue research*, 367(3):445–455.

Warren, J., Fuentes, M., Herring, A. H., and Langlois, P. H. (2012). Spatial-temporal modeling of the association between air pollution exposure and preterm birth: Identifying critical windows of exposure. *Biometrics*, 68(4):1157–1167.

Warren, J. L., Kong, W., Luben, T. J., and Chang, H. H. (2020a). Critical window variable selection: Estimating the impact of air pollution on very preterm birth. *Biostatistics*, 21(4):790–806.

Warren, J. L., Luben, T. J., and Chang, H. H. (2020b). A spatially varying distributed lag model with application to an air pollution and term low birth weight study. *Journal of the Royal Statistical Society Series C*, 69(3):681–696.

Weschler, C. J. and Nazaroff, W. W. (2012). SVOC exposure indoors: Fresh look at dermal pathways. *Indoor Air*, 22(5):356–377.

Wilson, A., Chiu, Y.-H. M., Hsu, H.-H. L., Wright, R. O., Wright, R. J., and Coull, B. A. (2017a). Bayesian distributed lag interaction models to identify perinatal windows of vulnerability in children's health. *Biostatistics*, 18(3):537–552.

Wilson, A., Chiu, Y.-H. M., Hsu, H.-H. L., Wright, R. O., Wright, R. J., and Coull, B. A. (2017b). Potential for bias when estimating critical windows for air pollution in children's health. *American Journal of Epidemiology*, 186(11):1281–1289.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

Wright, R. O. (2017). Environment, susceptibility windows, development, and child health. *Current Opinion in Pediatrics*, 29(2):211–217.

Wu, H., Jiang, B., Zhu, P., Geng, X., Liu, Z., Cui, L., and Yang, L. (2018). Associations between maternal weekly air pollutant exposures and low birth weight: A distributed lag non-linear model. *Environmental Research Letters*, 13(2):24023.

Zanobetti, A., Wand, M. P., Schwartz, J., and Ryan, L. M. (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*, 1(3):279–292.

Zhu, X., Liu, Y., Chen, Y., Yao, C., Che, Z., and Cao, J. (2015). Maternal exposure to fine particulate matter (PM 2.5) and pregnancy outcomes: a meta-analysis. *Environmental Science and Pollution Research*, 22:3383–3396.

Appendix A

Treed Distributed Lag Nonlinear Models

A.1 Additional figures on birth data

Table A.1 provides demographic breakdowns of the data and covariates used in our data analysis. Figure A.1 details the number of births by census tract.

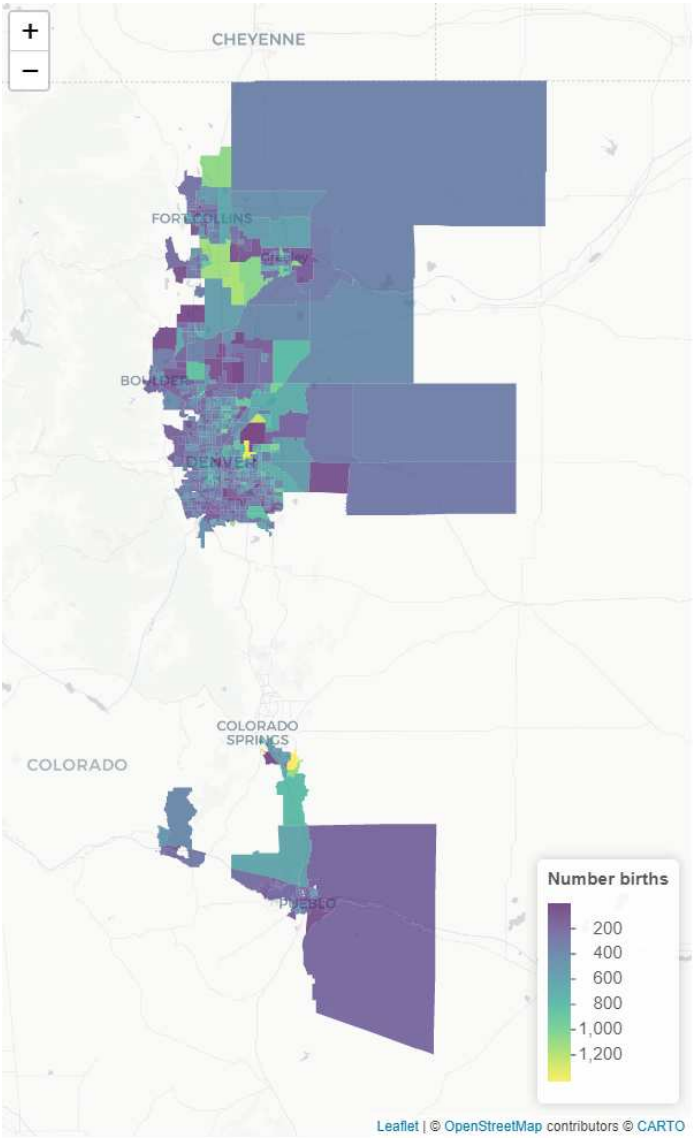


Figure A.1: Map of CO census tracts highlighted with color scale indicating number of births from each census tract used in our data analysis.

Table A.1: Demographic breakdown of the covariates used in the analysis of Colorado birth weights and estimated effect from analysis using TDLNMse.

Demographic Category	N	%	Estimate (95% CI)
Complete sample	300,463	100	
Age at beginning of pregnancy			26.7 (23.444, 29.826)
Quadratic term			-6.058 (-8.102, -4.07)
< 20	18,800	6.3	
$20 - 29$	144,597	48.1	
$30 - 39$	128,394	42.7	
≥ 40	8,672	2.9	
Weight (lb) at beginning of pregnancy			0.004 (0.004, 0.004)
< 100	4,971	1.7	
$100 - 149$	164,966	54.9	
$150 - 199$	100,148	33.3	
≥ 200	30,378	10.1	
Height			0.033 (0.032, 0.034)
Annual income ($\times$\$1000 USD)			
< 15	74,432	24.8	-0.038 (-0.05, -0.025)
$15 - 24$	36,308	12.1	
$25 - 34$	27,909	9.3	0.024 (0.003, 0.045)
$35 - 49$	27,788	9.2	0.023 (0.006, 0.041)
$50 - 74$	43,205	14.4	0.035 (0.02, 0.051)
≥ 75	90,821	30.2	0.035 (0.022, 0.049)
Race			
American Indian	2,080	0.7	
Asian or Pacific Islander	30,623	10.2	-0.133 (-0.242, -0.027)
Black	14,699	4.9	-0.283 (-0.313, -0.253)
White	253,061	84.2	-0.064 (-0.073, -0.056)
Hispanic			
No	210,019	69.9	-0.057 (-0.071, -0.043)
Yes	90,354	30.1	
Highest Education			
Less than high school	46,506	15.5	0.071 (0.044, 0.101)
High school diploma	118,986	39.6	0.046 (0.03, 0.063)
Associate's degree	21,808	7.3	0.036 (0.009, 0.063)
Bachelor's degree	71,789	23.9	0.034 (0.023, 0.045)
Advanced degree	41,374	13.8	
Marital Status			
Married	229,492	76.4	
Never married	63,379	21.1	-0.068 (-0.085, -0.051)
Other	7,592	2.5	-0.051 (-0.072, -0.029)
Prenatal Care			
Yes	293,237	97.6	0.104 (0.085, 0.124)
No	3,625	1.2	
Unknown	3,601	1.2	0.061 (-0.115, 0.242)
Smoking Before or During Pregnancy			
No	281,093	93.6	
Yes	19,370	6.4	-0.269 (-0.288, -0.25)
Child's Sex			
Female	147,693	49.2	
Male	152,770	50.8	

A.2 Additional details on computation

A.2.1 Preprocessing

Before running the TDLNM algorithm, we perform the following operations to promote computational precision and mitigate numerical overflow issues:

- The response, $\mathbf{y}$, is centered to have mean zero and scaled to have a range equal to 1.
- Continuous covariates are centered to have mean zero and all covariates are scaled by their ℓ_2 norm such that $\mathbf{Z}^T\mathbf{Z}$ has a diagonal of ones.

Initial values for the MCMC algorithm are drawn from full conditional distributions where possible and prior distributions otherwise.

A.2.2 Tree update

Prior to updating trees, we integrate out the fixed effect, $\boldsymbol{\gamma}$, using standard analytical techniques,

$$p(\mathbf{y}|\mathbf{f}, \sigma^2) = \int_{\boldsymbol{\gamma}} p(\mathbf{y}|\mathbf{f}, \boldsymbol{\gamma}, \sigma^2) p(\boldsymbol{\gamma}|\sigma^2) d\boldsymbol{\gamma}. \tag{A.1}$$

This results in the distribution

$$\mathbf{y}|\mathbf{f}, \sigma^2 \sim \mathcal{MVN}_n\left(\mathbf{f}, \sigma^2 \mathbf{V}_{\mathbf{Z}}\right) \tag{A.2}$$

where

$$\mathbf{V}_{\mathbf{Z}} = \left(\mathbf{I} - \mathbf{Z}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z}^T\right)^{-1} \tag{A.3}$$

and

$$\mathbf{V}_{\boldsymbol{\gamma}} = (\mathbf{Z}^T\mathbf{Z} + \mathbf{I}/c)^{-1}. \tag{A.4}$$

Here, $\mathbf{y} = [y_1, \ldots, y_n]^T$ is a vector of our continuous response; $\mathbf{f} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)]$ where $f(\mathbf{x}_i)$ incorporates the estimates $\boldsymbol{\mu}_a$ from all trees $\mathcal{T}_a$ for $a = 1, \ldots, A$; and $\mathbf{Z}$ is a matrix of covariates such that row i equals $\mathbf{z}_i^T$.

The update of each tree, $a = 1, \ldots, A$, proceeds as follows. First, we calculate $\mathbf{R}_a$, the partial residuals after removing the effects of all other trees. We define $\mathbf{R}_a$ as

$$\mathbf{R}_a = \mathbf{y} - \sum_{\substack{a'=1 \\ a' \neq a}}^{A} g(\mathbf{X}, \mathcal{T}_{a'}). \tag{A.5}$$

Here, $g(\mathbf{X}, \mathcal{T}_a)$ is the vector of partial distributed lag estimates, $g(\mathbf{X}, \mathcal{T}_a) = [g(\mathbf{x}_1, \mathcal{T}_a), \ldots, g(\mathbf{x}_n, \mathcal{T}_a)]^T$. We parameterize $g(\mathbf{x}_i, \mathcal{T}_a) = \mathbf{c}_{ia}^T \boldsymbol{\mu}_a$ where $\mathbf{c}_{ia} = [\#(\mathbf{x}_i \in \eta_{a1}), \ldots, \#(\mathbf{x}_i \in \eta_{aB_a})]^T$ is the counts of exposures in each terminal node η_{ab} and $\boldsymbol{\mu}_a = [\mu_{a1}, \ldots, \mu_{aB_a}]^T$ is the node-specific effects. In the case of TDLNMse, $\#(\mathbf{x}_i \in \eta_{ab})$ is replaced by $\sum_{t=1}^{T} \psi(x_{it}; \eta_{ab}, \sigma_x)$, the sum of weights across all observed times. Then for matrix $\mathbf{C}_a$ with rows $\mathbf{c}_{ia}$, we describe the distribution of the partial residuals as

$$\mathbf{R}_a | \boldsymbol{\mu}_a, \sigma^2 \sim \mathcal{MVN}_{B_a} \left(\mathbf{C}_a \boldsymbol{\mu}_a, \sigma^2 \mathbf{V}_\mathbf{Z} \right). \tag{A.6}$$

Second, we update $\mathcal{T}_a$ using a Metropolis-Hastings algorithm. Following Kapelner and Bleich (2016) we consider a proposal distribution with three options: grow, prune, or change as defined in the Bayesian CART algorithm (Chipman et al., 1998). The transition steps from current tree $\mathcal{T}_a$ to new tree $\mathcal{T}_a^*$ are described as follows:

- Grow: Randomly select a terminal node, η, to grow. Randomly select a splitting rule according to $p_{\text{rule}}(\rho|\eta)$ and create two new terminal nodes using the new splitting rule along with all previous rules.
- Prune: Randomly select an internal node with exactly two terminal nodes descending from it and remove the splitting rule.
- Change: Randomly select any internal node, η, and define a new splitting rule according to $p_{\text{rule}}(\rho|\eta)$. Update the limits of all terminal nodes that branch from η.

It is important to note that the grow and prune steps are counterparts to one another, while change is its own counterpart that can reverse the Markov chain. The transition kernel $p(\mathcal{T}^*|\mathcal{T})$ is given by the probability of selecting a step, multiplied by the probabilities associated with that step. For our simulations and data analysis we set the step probabilities as $p(\text{grow}) = p(\text{prune}) = 0.3$ and $p(\text{change}) = 0.4$.

After a new tree proposal, $\mathcal{T}_a^*$, is made, we accept it by a standard Metropolis-Hastings ratio. To eliminate the need for complicated procedures due to the change in parameter dimension and to make the trees invariant to the covariates and variance, we integrate over $\boldsymbol{\mu}_a$ as well as σ^2. In BART, integrating out the vector $\boldsymbol{\mu}_a$ can be done one parameter at a time, as each observation is restricted to a single terminal node. However, in TDLNM, the exposure-time observations may reside in several terminal nodes requiring us to simultaneously integrate over the entire vector, $\boldsymbol{\mu}_a$. The marginal likelihood of $\mathbf{R}_a$ is given as

$$\begin{aligned} p(\mathbf{R}_a|\mathcal{T}_a, \omega^2, \tau_a^2) &= \int_{\sigma^2}\int_{\boldsymbol{\mu}_a} p(\mathbf{R}_a|\boldsymbol{\mu}_a, \mathcal{T}_a, -)p(\boldsymbol{\mu}_a|\mathcal{T}_a, -)p(\sigma^2)\; d\boldsymbol{\mu}_a\; d\sigma^2 \\ &= \left(\omega^2\tau_a^2\right)^{-B_a/2} \left|\mathbf{V}_{\boldsymbol{\mu}_a}\right|^{1/2} \\ &\times \left[\frac{\mathbf{R}_a^T\left(\mathbf{I}_n - \mathbf{Z}^T\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z} - \mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{C}_a\mathbf{V}_{\boldsymbol{\mu}_a}\mathbf{C}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\right)\mathbf{R}_a}{2} + \frac{1}{\xi_{\sigma^2}}\right]^{-(n+1)/2} \end{aligned} \tag{A.7}$$

where ξ_{σ^2} is from the hierarchy $\sigma^2|\xi_{\sigma^2} \sim \mathcal{IG}(1/2, 1/\xi_{\sigma^2})$ and $\xi_{\sigma^2} \sim \mathcal{IG}(1/2, 1)$ such that $\sigma \sim \mathcal{C}^+(0, 1)$. We then calculate $p(\mathbf{R}_a, \mathcal{T}_a) = p(\mathbf{R}_a|\mathcal{T}_a)p(\mathcal{T}_a)$ and accept $\mathcal{T}_a^*$ according to the Metropolis-Hastings ratio given by

$$r = \min\left\{1, \frac{p(\mathcal{T}_a^*)p(\mathbf{R}_a|\mathcal{T}_a^*, \omega^2, \tau_a^2)p(\mathcal{T}_a|\mathcal{T}_a^*)}{p(\mathcal{T}_a)p(\mathbf{R}_a|\mathcal{T}_a, \omega^2, \tau_a^2)p(\mathcal{T}_a^*|\mathcal{T}_a)}\right\}. \tag{A.8}$$

A.2.3 Full conditionals

After the tree update, we draw $\boldsymbol{\mu}_a$ from the full conditional distribution,

$$\boldsymbol{\mu}_a|\mathcal{T}_a, \mathbf{R}_a, \sigma^2, \omega^2, \tau_a^2 \sim \mathcal{MVN}_{B_a}\left(\mathbf{V}_{\boldsymbol{\mu}_a}\mathbf{C}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{R}_a, \sigma^2\mathbf{V}_{\boldsymbol{\mu}_a}\right), \tag{A.9}$$

where

$$\mathbf{V}_{\boldsymbol{\mu}_a} = \left(\mathbf{C}_a^T \mathbf{V}_{\mathbf{Z}}^{-1} \mathbf{C}_a + \omega^{-2}\tau_a^{-2}\mathbf{I}_{B_a}\right)^{-1}. \tag{A.10}$$

We note here that the computation time of $\mathbf{V}_{\boldsymbol{\mu}_a}$ is no more than $\mathcal{O}(nB_a^3)$. Since B_a is limited by the size of the trees, computation time of our algorithm grows at the reasonable rate of $\mathcal{O}(n)$.

The update of each tree and corresponding partial DLNM is followed by a draw from the full conditional for hyperparameter ξ_{τ_a},

$$\xi_{\tau_a}|- \sim \mathcal{IG}\left(1, 1 + \frac{1}{\tau_a^2}\right) \tag{A.11}$$

and a draw from the full conditional for τ_a^2,

$$\tau_a|- \sim \mathcal{IG}\left(\frac{B_a + 1}{2}, \frac{\boldsymbol{\mu}_a^T \boldsymbol{\mu}_a}{2\sigma^2\omega^2} + \frac{1}{\xi_{\tau_a}}\right). \tag{A.12}$$

We update hyperparameter ξ_σ from full conditional

$$\xi_\sigma|- \sim \mathcal{IG}\left(1, 1 + \frac{1}{\sigma^2}\right), \tag{A.13}$$

as well as σ^2 from full conditional

$$\sigma^2|- \sim \mathcal{IG}\left[\frac{n + u + 1}{2}, \frac{\|\mathbf{V}_{\mathbf{Z}}^{-1/2}(\mathbf{y} - \mathbf{f})\|_2^2}{2} + \frac{U}{2\omega^2} + \frac{1}{\xi_\sigma}\right], \tag{A.14}$$

where $u = \sum_{a=1}^{A} B_a$ and $U = \sum_{a=1}^{A} \tau_a^{-2}\|\boldsymbol{\mu}_a\|_2^2$. We also update hyperparameter ξ_ω from full conditional

$$\xi_\omega|- \sim \mathcal{IG}\left(1, 1 + \frac{1}{\omega^2}\right), \tag{A.15}$$

and ω^2 from full conditional

$$\omega^2|- \sim \mathcal{IG}\left(\frac{u + 1}{2}, \frac{U}{2\sigma^2} + \frac{1}{\xi_\omega}\right). \tag{A.16}$$

Finally we draw $\boldsymbol{\gamma}$ from its full conditional,

$$\boldsymbol{\gamma}|\mathbf{y}, \mathbf{f}, \sigma^2 \sim \mathcal{MVN}\left[\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z}(\mathbf{y}-\mathbf{f}), \sigma^2\mathbf{V}_{\boldsymbol{\gamma}}\right]. \tag{A.17}$$

A.3 Additional simulation details

A.3.1 Simulation scenarios

The exposure-time-response functions used in our simulations are constructed as follows:

- Scenario A:

$$w(x_{it}, t) = \begin{cases} -1 & \text{if } t \in [11, 15] \text{ and } x_{it} > 2 \\ 0 & \text{otherwise} \end{cases}$$

- Scenario B:

$$w(x_{it}, t) = \begin{cases} 1 - x_{it} & \text{if } t \in [11, 15] \\ 0 & \text{otherwise} \end{cases}$$

- Scenario C:

$$w(x_{it}, t) = \begin{cases} \left\{\frac{1}{1+\exp(5x_{it}-12.5)} - 1\right\} & \text{if } t \in [11, 15] \\ 0 & \text{otherwise} \end{cases}$$

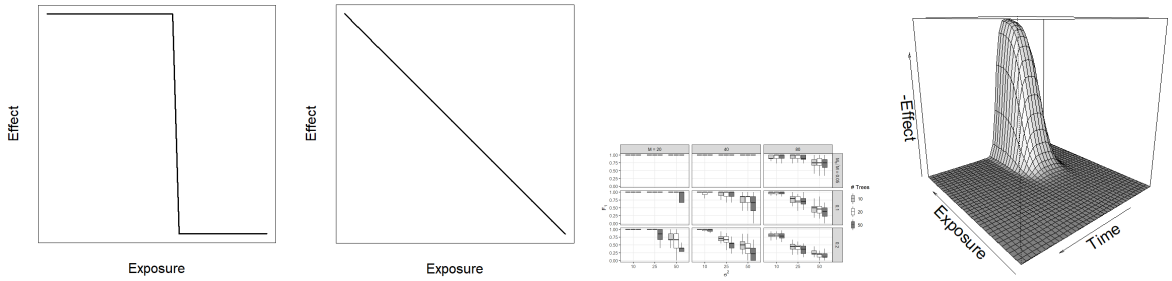
- Scenario D:

$$w(x_{it}, t) = \left\{\frac{1}{1+\exp(5x_{it}-12.5)} - 1\right\} \cdot \exp\left\{-\frac{(t-13)^4}{400}\right\}$$

A graphical representation of the design is shown in Figure A.2.

A.3.2 Additional simulation results

Table A.2 shows simulation results from the main text, including standard errors for the simulation replicates.



(a) Scenario A (b) Scenario B (c) Scenario C (d) Scenario D

Figure A.2: Simulation scenarios representing the exposure-time-response functions. Panels (a), (b) and (c) show the exposure-response function at times $11 - 15$, while the DLNM surface is zero everywhere else. Panel (d) shows a smooth exposure-response.

Table A.2: Simulation results with standard errors for simulation replicates, showing RMSE for estimation of the exposure-time-surface with no-effect and effect separated. Coverage and CI width is based on 95% confidence intervals. Effect identification considers the likelihood of identifying a non-zero effect (TP) or incorrectly designating a non-zero effect (FP) over the DLNM surface. Precision is calculated within each simulation as TP/(TP+FP).

	DLNM RMSE			DLNM Coverage		Effect Identification		
Model	Overall	No Effect	Effect	Overall	CI Width	TP	FP	Precision
Scenario A: Piecewise in Exposure and Time								
TDLNM	0.086 (0.013)	0.066 (0.013)	0.213 (0.049)	1.00 (0.01)	0.43 (0.01)	0.87 (0.26)	0.00 (0.01)	1.00 (0.01)
TDLNMse	0.100 (0.016)	0.077 (0.011)	0.252 (0.110)	0.99 (0.02)	0.46 (0.02)	0.82 (0.29)	0.00 (0.01)	0.98 (0.10)
GAM	0.294 (0.044)	0.258 (0.037)	0.584 (0.281)	0.95 (0.04)	1.08 (0.04)	0.47 (0.30)	0.03 (0.04)	0.90 (0.24)
DLM	0.370 (0.041)	0.342 (0.028)	0.626 (0.237)	0.68 (0.07)	0.53 (0.07)	1.00 (0.00)	0.30 (0.07)	0.77 (0.04)
GLM-AIC	1.531 (4.470)	1.536 (4.790)	1.462 (1.670)	0.84 (0.11)	3.35 (0.11)	0.49 (0.46)	0.15 (0.10)	0.55 (0.39)
GAMcr	0.263 (0.031)	0.241 (0.029)	0.454 (0.177)	0.98 (0.02)	1.10 (0.02)	0.62 (0.26)	0.01 (0.02)	0.96 (0.14)
GAM-exp	0.241 (0.026)	0.165 (0.020)	0.669 (0.222)	0.94 (0.03)	0.67 (0.03)	0.32 (0.25)	0.01 (0.02)	0.87 (0.30)
Scenario B: Linear in Exposure								
TDLNM	0.292 (0.027)	0.081 (0.016)	0.768 (0.179)	0.87 (0.04)	0.37 (0.04)	0.56 (0.07)	0.01 (0.04)	0.99 (0.05)
TDLNMse	0.270 (0.026)	0.073 (0.014)	0.712 (0.174)	0.87 (0.05)	0.34 (0.05)	0.64 (0.07)	0.01 (0.05)	0.99 (0.05)
GAM	0.312 (0.017)	0.257 (0.016)	0.547 (0.159)	0.73 (0.07)	0.48 (0.07)	0.90 (0.15)	0.18 (0.07)	0.84 (0.04)
DLM	0.299 (0.010)	0.257 (0.014)	0.489 (0.098)	0.64 (0.07)	0.36 (0.07)	1.00 (0.00)	0.26 (0.08)	0.79 (0.05)
GLM-AIC	0.267 (0.154)	0.253 (0.170)	0.346 (0.073)	0.79 (0.07)	0.46 (0.07)	0.99 (0.06)	0.18 (0.08)	0.85 (0.06)
GAMcr	0.248 (0.019)	0.206 (0.014)	0.426 (0.112)	0.84 (0.05)	0.54 (0.05)	0.87 (0.15)	0.09 (0.04)	0.90 (0.04)
GAM-exp	0.283 (0.019)	0.226 (0.015)	0.518 (0.213)	0.76 (0.06)	0.37 (0.06)	0.94 (0.15)	0.15 (0.06)	0.86 (0.04)
Scenario C: Smooth in Exposure								
TDLNM	0.077 (0.004)	0.033 (0.002)	0.223 (0.033)	0.94 (0.03)	0.18 (0.03)	0.58 (0.11)	0.01 (0.03)	0.99 (0.03)
TDLNMse	0.070 (0.003)	0.032 (0.002)	0.201 (0.030)	0.97 (0.03)	0.17 (0.03)	0.67 (0.09)	0.01 (0.02)	0.99 (0.03)
GAM	0.142 (0.004)	0.126 (0.006)	0.241 (0.033)	0.91 (0.03)	0.36 (0.03)	0.60 (0.08)	0.06 (0.03)	0.91 (0.03)
DLM	0.138 (0.002)	0.120 (0.003)	0.245 (0.018)	0.64 (0.07)	0.18 (0.07)	1.00 (0.00)	0.31 (0.08)	0.77 (0.05)
GLM-AIC	0.186 (0.054)	0.167 (0.056)	0.309 (0.068)	0.82 (0.06)	0.40 (0.06)	0.53 (0.11)	0.14 (0.06)	0.80 (0.08)
GAMcr	0.113 (0.004)	0.104 (0.004)	0.176 (0.015)	0.95 (0.02)	0.37 (0.02)	0.64 (0.06)	0.03 (0.02)	0.96 (0.02)
GAM-exp	0.126 (0.003)	0.103 (0.004)	0.255 (0.042)	0.92 (0.03)	0.28 (0.03)	0.62 (0.12)	0.05 (0.03)	0.93 (0.03)
Scenario D: Smooth in Exposure and Time								
TDLNM	0.105 (0.005)	0.041 (0.002)	0.203 (0.020)	0.80 (0.03)	0.26 (0.03)	0.40 (0.09)	0.00 (0.01)	0.99 (0.02)
TDLNMse	0.098 (0.004)	0.038 (0.002)	0.190 (0.018)	0.95 (0.03)	0.24 (0.03)	0.45 (0.08)	0.01 (0.02)	0.99 (0.03)
GAM	0.120 (0.005)	0.100 (0.005)	0.171 (0.012)	0.97 (0.03)	0.44 (0.03)	0.54 (0.09)	0.01 (0.02)	0.98 (0.03)
DLM	0.122 (0.004)	0.090 (0.004)	0.193 (0.009)	0.69 (0.08)	0.23 (0.08)	0.94 (0.06)	0.23 (0.10)	0.80 (0.06)
GLM-AIC	0.284 (0.275)	0.277 (0.333)	0.306 (0.120)	0.81 (0.08)	0.52 (0.08)	0.45 (0.10)	0.14 (0.09)	0.77 (0.12)
GAMcr	0.110 (0.004)	0.092 (0.004)	0.156 (0.010)	0.97 (0.03)	0.41 (0.03)	0.57 (0.09)	0.01 (0.02)	0.98 (0.03)
GAM-exp	0.099 (0.004)	0.068 (0.003)	0.164 (0.016)	0.97 (0.03)	0.35 (0.03)	0.47 (0.08)	0.00 (0.01)	0.99 (0.02)

Figure $A.3$ indicates the probability of detecting a non-zero effect in at least one exposure value in each week, for models TDLNM, TDLNMse, GAMcr, and DLM.

Figure A.4 shows slices of the DLNM surface and estimates by GAM and DLM. Figure A.5 shows slices of the DLNM surface and estimates by GLM-AIC and GAM-exp.

A.3.3 Tree-specific variance prior

To better understand the operating characteristics of the ensemble of trees, we analyzed the tree structures from our simulation study. Specifically, we investigate the size of trees in the ensemble and how influential each tree is. To investigate tree size we ordered the trees by number of terminal nodes at each MCMC iteration. Then we found the median, 2.5, and 97.5 percentiles of terminal nodes of the ordered trees across all MCMC iterations. Averages of each metric was calculated from simulation replicates for each scenario. The results are shown in Figure A.6(a).

The largest tree had a median of 5 to 6 nodes in each of the scenarios. Other trees had medians of between 2 and 4 nodes. Hence, all of the trees in the ensemble tend to have more than one terminal node and therefore contribute the the outcome in some way.

We repeated the analysis looking at the ordered rank of τ_a^2, which is the tree-specific variance component. Results are shown in Figure A.6(b). This shows that there tends to be one tree that is dominant and has the largest variance for node-specific effects and likely contains the major shape of the exposure-time-response function. The other trees tend to have smaller effects and likely add detail to the estimated function.

A.3.4 Impact on tree structure by smoothing

We investigate the impact of smoothing on the tree structure. The exposure interval within a terminal node means that, for example, an exposure at the boundary of a terminal node will have an effect that is primarily determined by the two adjacent terminal nodes. However, if the smoothing parameter is large, then other terminal nodes (whose exposure-concentration interval is further away) will also make a large contribution to the estimated effect. In practice, fewer terminal nodes are needed to estimate smooth exposure-response functions with TDLNMse than with TDLNM.

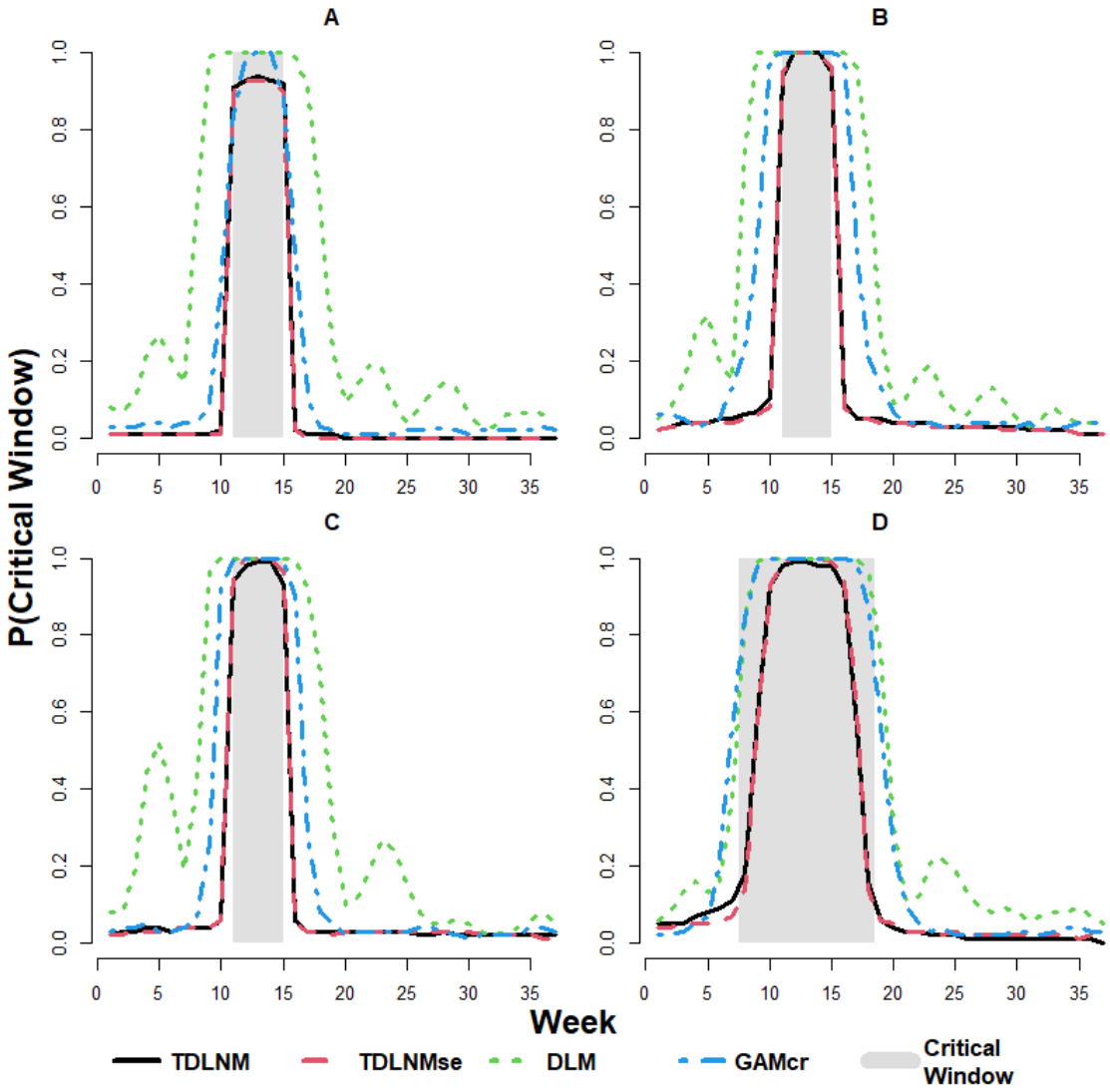


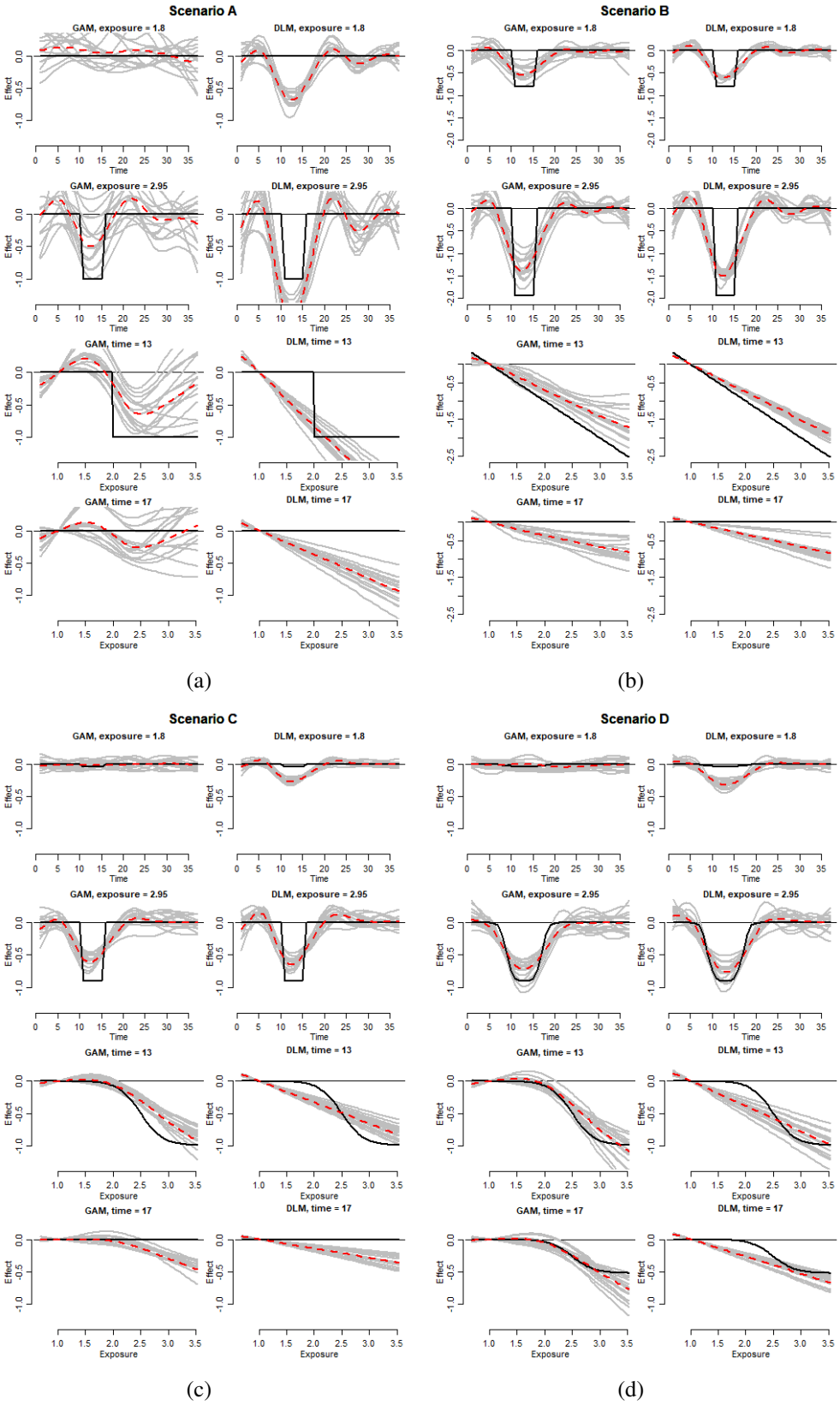
Figure A.3: Probability of detecting a critical window at each week of the exposure-time-response surface. Panels A through D indicate simulation scenario with gray area indicating true critical window and lines showing the probability of each model detecting a non-zero difference from the centering value in at least one exposure concentration.

As a result, smaller trees are used on average with TDLNMse. This is shown in Figure A.7, which compares the average terminal nodes for TDLNM and TDLNMse for 100 simulation replicates. The difference in terminal nodes is evident for scenarios B, C, and D, which have a smooth effect exposure-response effect.

A.4 Additional data analysis results

For comparison, we fit TDLNM, a DLM and several linear models to compare results. Each of these models was consistent with the TDLNMse results. A linear regression model found mean log-exposure of $PM_{2.5}$ across the entire pregnancy to be associated with decreased BWGAZ (effect $= -0.081$, 95% CI: $-0.129, -0.034$). Analyzing the mean exposure throughout pregnancy as a smooth effect (Figure A.8(c)) resembles the association found by TDLNMse, where below median exposure concentration is associate with increase BWGAZ, while the effect is near zero for above median exposure concentration.

Findings from TDLNM (Figure A.8(a)) show results similar to that of TDLNMse in the data analysis, but with a more defined threshold of effect at around $5\mu g/m^3$ $PM_{2.5}$ concentration. The DLM (Figure A.8(b)) shows evidence of a negative association between increased log-exposure and BWGAZ across the pregnancy. However, the 95% confidence intervals in the DLM indicate the effect is non-zero in only the first and third trimester. A linear model with log-exposure analyzed by trimester finds similar results (trimester 1 effect $= -0.043$, 95% CI: $-0.074, -0.012$; trimester 2 effect $= -0.002$, 95% CI: $-0.035, 0.031$; trimester 3 effect $= -0.039$, 95% CI: $-0.068, -0.010$). While the DLM only indicates non-zero effects in the first and third trimester, the ability of the DLNM methods to borrow strength across the more extreme exposure concentration values may lead to improved power in finding critical windows across the entire pregnancy. In addition, the DLM and trimester model do not account for the nonlinearity present in the DLNM results.



(a) (b)

(c) (d)

Figure A.4: Results of our simulation study comparing GAM and DLM.

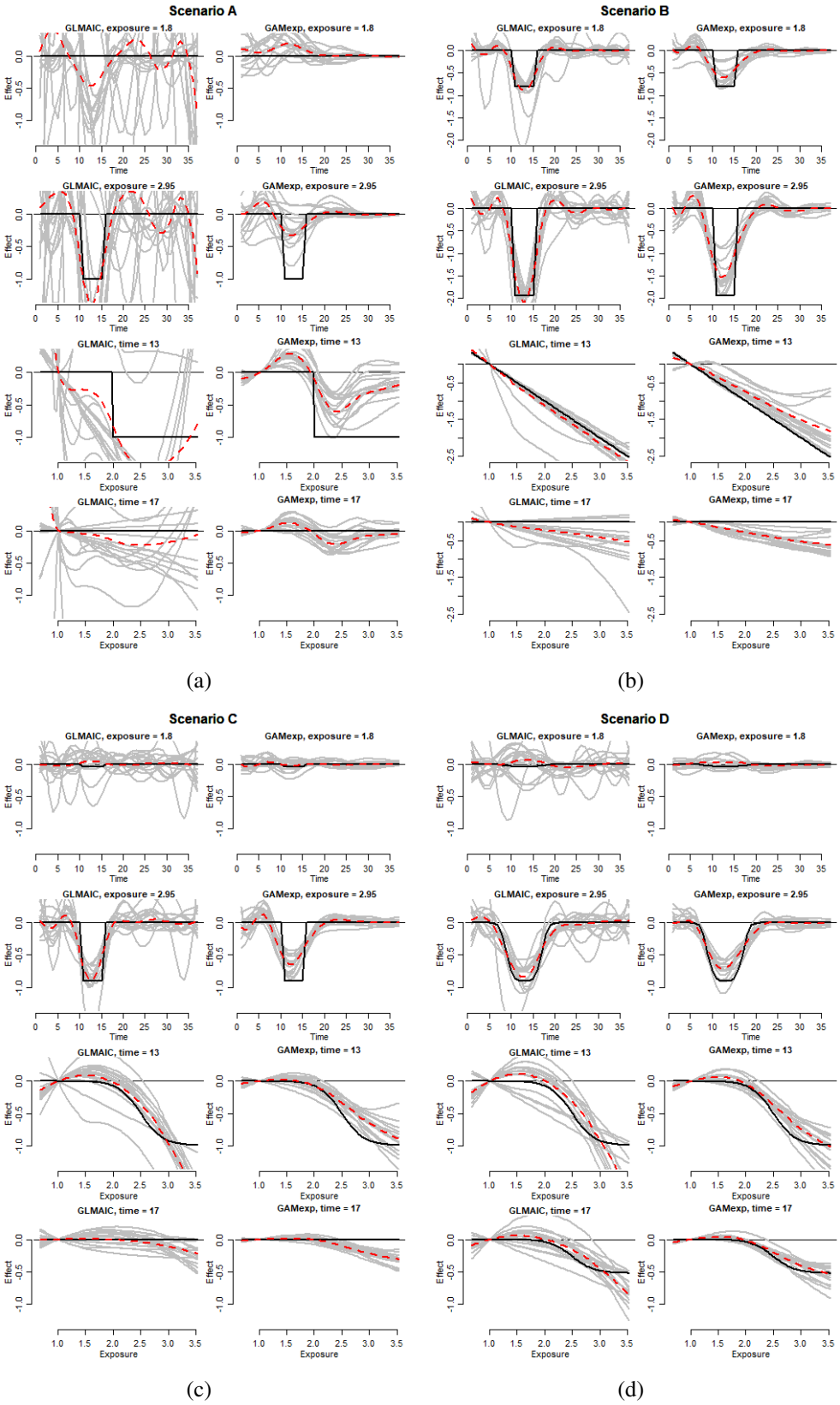
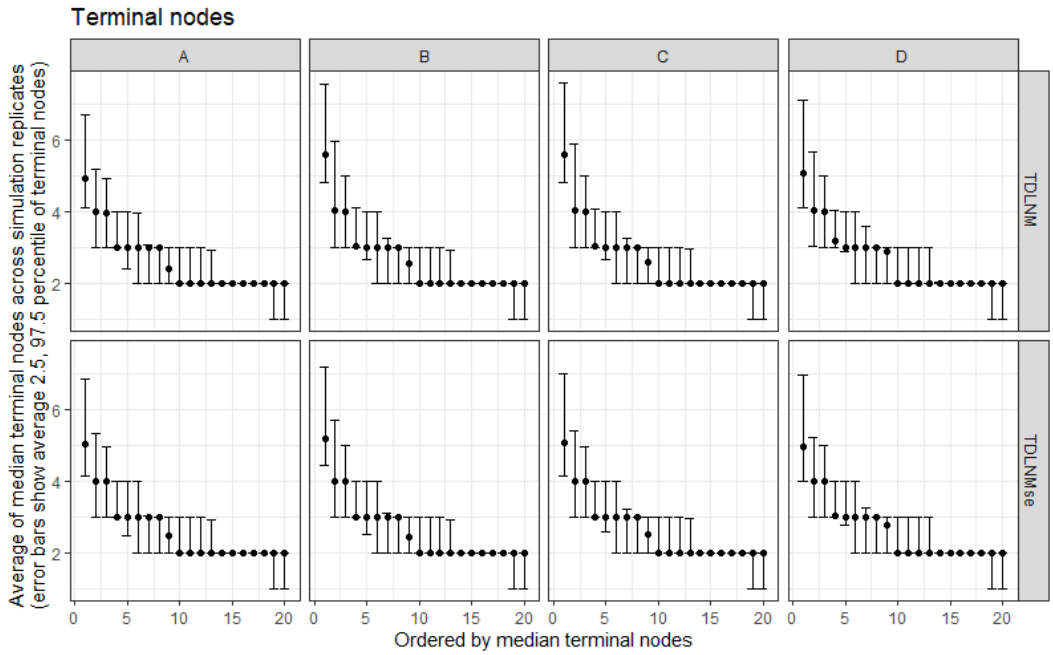
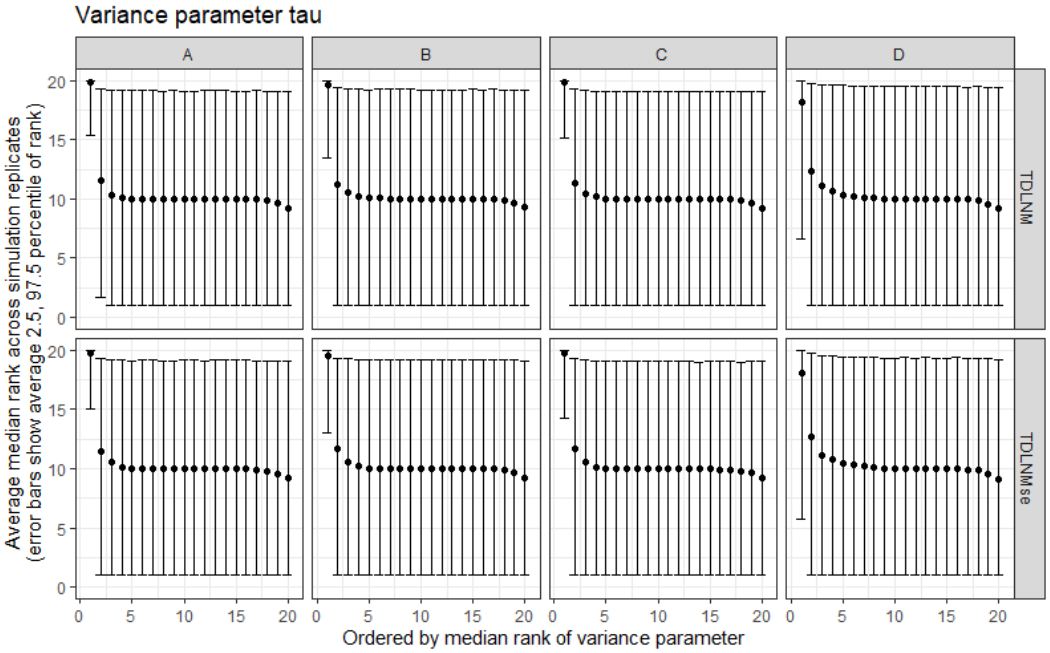


Figure A.5: Results of our simulation study comparing GLM-AIC and GAM-exp.



(a)



(b)

Figure A.6: Tree structure and variance metrics from simulation analysis. At each MCMC iteration, trees were ordered by the number of terminal nodes (panel a) and size of variance parameter τ_a^2 (panel b). The median, 2.5, and 97.5 percentiles of these metrics were calculated across all MCMC iterations. The panels show the average of these metrics across simulation replicates for each simulation scenario.

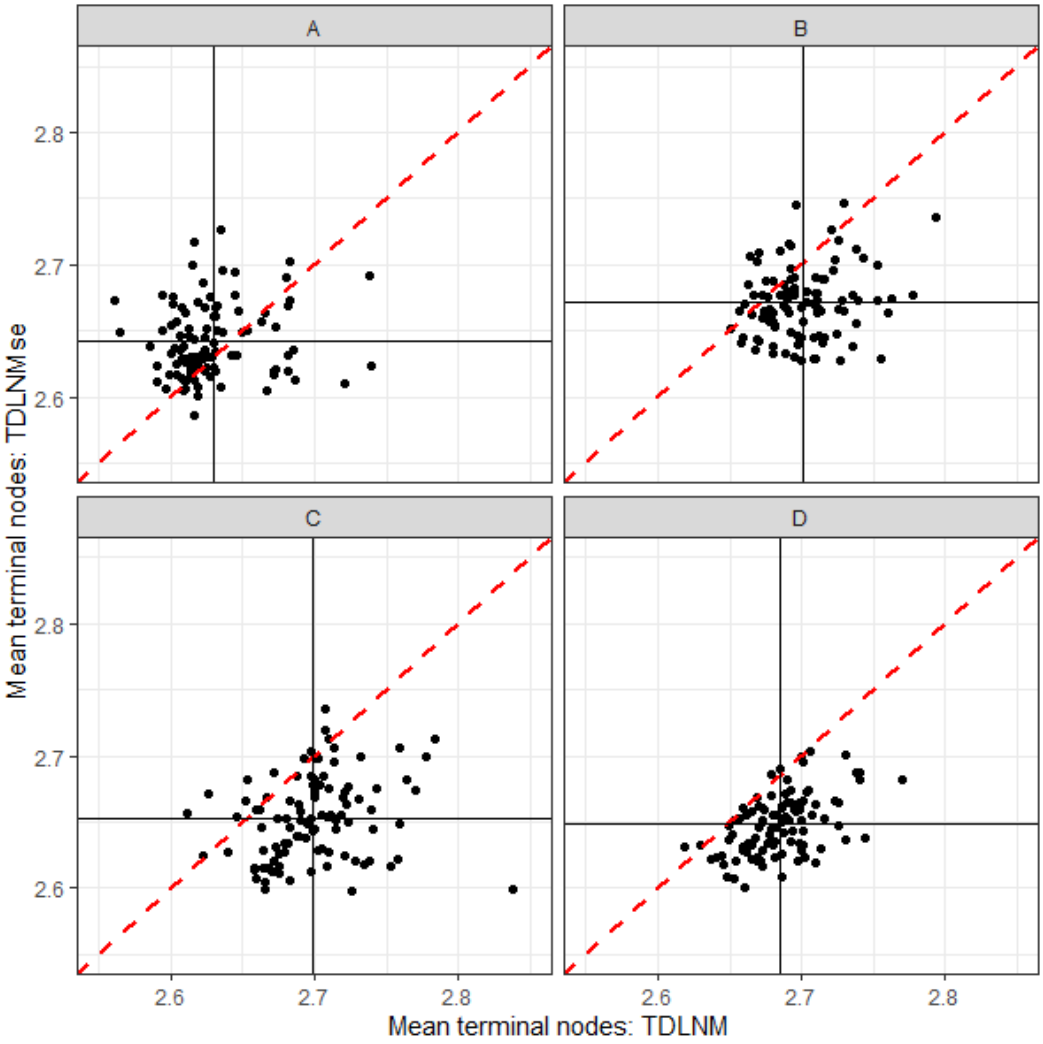
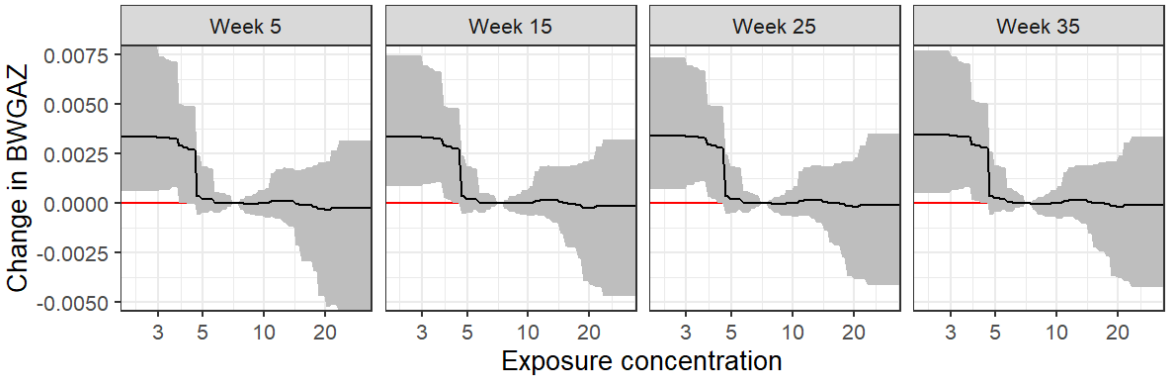
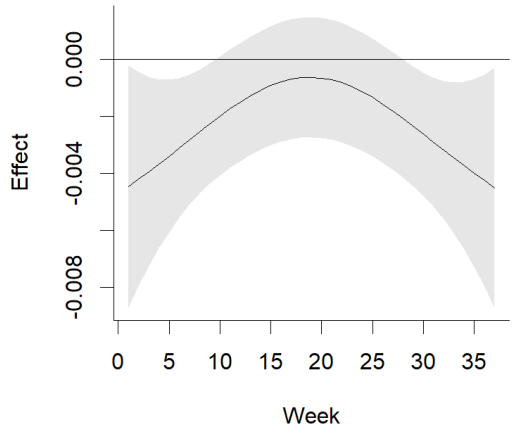


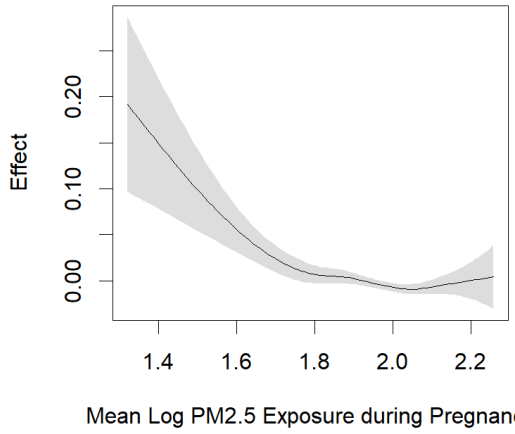
Figure A.7: Average terminal nodes from 100 TDLNM and TDLNMse simulation replicates for each simulation scenario. The black lines indicate average across replicates while the red dashed line is $y = x$.



(a) TDLNM



(b) DLM



(c) Penalized Cubic Spline

Figure A.8: Estimated exposure-time-response between log-$PM_{2.5}$ exposure and BWGAZ, using models TDLNM (a), DLM (b), and penalized cubic spline on mean log-exposure throughout pregnancy (c) relative to the median exposure-concentration value.

Appendix B

Estimating Perinatal Critical Windows to Environmental Mixtures via Structured Bayesian Regression Tree Pairs

B.1 Colorado Birth Cohort Data

We analyze birth weight for gestational age z-score, BWGAZ, using birth vital statistics records from Colorado, USA. BWGAZ is the birth weight adjusted for gestational age and fetal sex using a standard reference table (Fenton and Kim, 2013).

Our data set is limited to births in the Denver metropolitan area, including the counties Adams, Arapahoe, Broomfield, Denver, Douglas, and Jefferson. Tables B.1 and B.2 provide demographic breakdowns of the covariates used in our data analysis and estimates from TDLMMns of their linear association with BWGAZ.

$PM_{2.5}$ data was obtained from US Environmental Protection Agency (EPA) CMAQ models. Other exposure measurements came from all available US EPA monitors in the study area. Daily exposure measurements were assigned to each census track based on inverse distance weighting. We then created weekly average exposures for each pregnancy based on the date of conception and census tract of residence. The weekly average $PM_{2.5}$ data was log-transformed to reduce skew.

Table B.1: Demographic breakdown of the covariates used in the analysis of Colorado birth weights along with estimated linear association with change in BWGAZ.

Demographic Category	N	%	Est (95% CI)
Complete sample	195,702	100	
Age at beginning of pregnancy			0.0259 (0.0202, 0.0316)
(quadratic term)			-0.0003 (-0.0004, -0.0002)
< 20	11,301	5.8	
$20 - 29$	90,828	46.4	
$30 - 39$	87,636	44.8	
≥ 40	5,936	3.0	
Height (in)			0.0346 (0.0331, 0.0361)
< 60	6,215	3.2	
$60 - 64$	99,866	51.0	
$65 - 69$	82,330	42.1	
≥ 70	7,290	3.2	
Weight (lb) at beginning of pregnancy			0.0038 (0.0037, 0.0039)
< 100	3,338	1.7	
$100 - 149$	107,662	55.0	
$150 - 199$	65,206	33.3	
≥ 200	19,495	10.0	
Annual income ($\times$\$1000 USD)			
< 15	50,073	25.6	-0.0312 -0.0447 -0.0178
$15 - 24$	22,383	11.4	
$25 - 34$	17,076	8.7	0.0142 (-0.0026, 0.0311)
$35 - 49$	16,464	8.4	0.0189 (0.0017, 0.0360)
$50 - 74$	26,430	13.5	0.0284 (0.0122, 0.0445)
≥ 75	63,275	32.3	0.0230 (0.0071, 0.0387)

Table B.2: Demographic breakdown (continued) of the covariates used in the analysis of Colorado birth weights along with estimated linear association with change in BWGAZ.

Demographic Category	N	%	Est (95% CI)
Race			
American Indian	1,349	0.7	
Asian or Pacific Islander	20,707	10.6	-0.1518 (-0.1986, -0.1053)
Black	12,896	6.6	-0.2801 (-0.3281, -0.2325)
White	160,749	82.1	-0.0775 (-0.1238, -0.0326)
Hispanic			
No	135,339	69.2	-0.0697 (-0.0796, -0.0597)
Yes	60,362	30.8	
Highest Education			
Less than high school	31,028	15.9	0.0711 (0.0535, 0.0886)
High school diploma	75,510	38.6	0.0467 (0.0326, 0.0607)
Associate's degree	13,180	6.7	0.0319 (0.0135, 0.0500)
Bachelor's degree	48,055	24.6	0.0331 (0.0208, 0.0458)
Advanced degree	27,928	14.3	
Marital Status			
Married	151,200	77.3	
Never married	40,160	20.5	-0.0716 (-0.0824, -0.0607)
Other	4,341	2.2	-0.0385 (-0.0644, -0.0128)
Prenatal Care			
Yes	191,067	97.6	0.1378 (0.0954, 0.1807)
No	1,516	0.8	
Unknown	3,118	1.6	0.0959 (0.0447, 0.1474)
Smoking			
Never	181,957	93	0.3310 (0.2856, 0.3765)
Former	3,073	1.6	0.2484 (0.1942, 0.3032)
$<$ 10 Cig/day	9,356	4.8	0.0684 (0.0194, 0.1166)
$\geq$ 10 Cig/day	1,315	0.7	
Child's Sex			
Female	96,036	49.1	
Male	99,665	50.9	

B.2 Additional details on computation

B.2.1 Preprocessing

Before running the TDLMM algorithm, we perform the following operations to promote computational precision and mitigate numerical overflow issues:

- The response, $\mathbf{y}$, is centered to have mean zero and scaled to have a range equal to 1.
- Continuous covariates are centered to have mean zero and all covariates are scaled by their ℓ_2 norm such that $\mathbf{Z}^T\mathbf{Z}$ has a diagonal of ones.
- All exposure data is scaled to have standard deviation 1.

B.2.2 Tree update

Consider the distribution of the data $\mathbf{y}_i \sim \mathcal{N}[f(\mathbf{X}_i) + \mathbf{z}^T\boldsymbol{\gamma}, \sigma^2]$, where f represents the distributed lag function as described in equations (1) or (2) of the main text which is evaluated for the set of predictors associated with observation i. The full posterior of $\boldsymbol{\gamma}$ is

$$p(\boldsymbol{\gamma}|\mathbf{y}, \mathbf{f}, \sigma^2) \propto \sigma^{-p/2}|\mathbf{V}_{\boldsymbol{\gamma}}|^{-1/2} \exp\left\{-\sigma^{-2}(\mathbf{y}-\mathbf{f})^T\mathbf{Z}^T\mathbf{V}_{\boldsymbol{\gamma}}^{-1}\mathbf{Z}(\mathbf{y}-\mathbf{f})\right\}. \tag{B.1}$$

Here, $\mathbf{y} = [y_1, \ldots, y_n]^T$ is a vector of our continuous response; $\mathbf{f} = [f(\mathbf{X}_1), \ldots, f(\mathbf{X}_n)]$ where $f(\mathbf{X}_i)$; and $\mathbf{Z}$ is a matrix of covariates such that row i equals $\mathbf{z}_i^T$. In addition,

$$\mathbf{V}_{\boldsymbol{\gamma}} = (\mathbf{Z}^T\mathbf{Z} + \mathbf{I}/c)^{-1}, \tag{B.2}$$

where c is a fixed at a large value indicating a non-informative prior on $\boldsymbol{\gamma}$. In TDLMM, the distributed lag function $f(\mathbf{X}_i)$ is calculated using the parameter estimates based on equations (3) and (4) in the main text. In order to account for the effect of covariates $\mathbf{z}_i$ when estimating the trees, we integrate over the parameters $\boldsymbol{\gamma}$. This results in the marginal distribution for our data,

$$\mathbf{y}|\mathbf{f}, \sigma^2 \sim \mathcal{MVN}(\mathbf{f}, \sigma^2\mathbf{V}_{\mathbf{Z}}), \tag{B.3}$$

where

$$\mathbf{V_Z} = (\mathbf{I} - \mathbf{Z}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z}^T)^{-1}. \tag{B.4}$$

The update of each tree, $a = 1, \ldots, A$, proceeds as follows. First, we calculate $\mathbf{R}_a$, the partial residuals after removing the effects of all other trees. We define $\mathbf{R}_a$ as

$$\mathbf{R}_a = \mathbf{y} - \sum_{\substack{a'=1 \\ a' \neq a}}^{A} g(\mathbf{X}, \mathcal{T}_{a'_1}, \mathcal{T}_{a'_2}, S_{a'_1}, S_{a'_2}). \tag{B.5}$$

Here, $g(\mathbf{X}, \mathcal{T}_{a'_1}, \mathcal{T}_{a'_2}, S_{a'_1}, S_{a'_2})$ is the vector of partial distributed lag estimates which is parameterized as

$$\begin{aligned} g(\mathbf{X}; \mathcal{T}_{a_1}, \mathcal{T}_{a_2}, S_{a_1}, S_{a_2}) &= \sum_{b_1=1}^{B_{a_1}} x^*_{S_{a_1} a_1 b_1} \delta_{a_1 b_1} + \sum_{b_2=1}^{B_{a_2}} x^*_{S_{a_2} a_2 b_2} \delta_{a_2 b_2} && \text{(B.6)} \\ &\quad + \sum_{b_1=1}^{B_{a_1}} \sum_{b_2=1}^{B_{a_2}} x^*_{S_{a_1} a_1 b_1} x^*_{S_{a_2} a_2 b_2} \delta_{a_1 b_1 a_2 b_2}. && \text{(B.7)} \end{aligned}$$

with $x^*_{mab} = \sum_{t \in \eta_{ab}} x_{mt}$. Then for matrix $\mathbf{X}_a$ with columns representing the parameterization from (B.6) and rows representing individual observations, the conditional distribution of the partial residuals is

$$\mathbf{R}_a | - \sim \mathcal{MVN}_{B_a} \left(\mathbf{X}_a \boldsymbol{\delta}_a, \sigma^2 \mathbf{V_Z} \right) \tag{B.8}$$

where $\boldsymbol{\delta}_a$ represents a vector of parameters corresponding to g.

Second, we update each $\mathcal{T}_{a_j}$, $j \in \{1, 2\}$, using a Metropolis-Hastings algorithm. We consider a proposal distribution with transition steps as follows:

- Grow: Randomly select a terminal node, η, to grow. Randomly select a splitting rule according to $p_{\text{rule}}(\rho|\eta)$ and create two new terminal nodes using the new splitting rule along with all previous rules.
- Prune: Randomly select an internal node with exactly two terminal nodes descending from it and remove the splitting rule.

- Change: Randomly select any internal node, η, and define a new splitting rule according to $p_{\text{rule}}(\rho|\eta)$. Update the limits of all terminal nodes that branch from η.

- Switch-exposure: Randomly select a new exposure from $S|\mathcal{E}$ and use exposure measurements within existing tree structure.

The grow and prune steps are counterparts to one another, while change is its own counterpart that can reverse the Markov chain. The transition kernel $p(\mathcal{T}^*|\mathcal{T})$ is given by the probability of selecting a step, multiplied by the probabilities associated with that step. We also consider a switch-exposure proposal, which exchanges S with new exposure data S'. The switch-exposure proposal reverses itself. For our simulations and data analysis we draw a new proposal from the four options (grow, prime, change, switch-exposure) with equal probability. For TDLM, we only consider grow, prune, and change proposals.

After a new tree proposal or switch-exposure proposal is made, we accept it by a standard Metropolis-Hastings ratio. To eliminate the need for complicated procedures due to the change in parameter dimension and to make the trees invariant to the covariates and variance, we integrate over $\boldsymbol{\delta}_a$ as well as σ^2. In BART, integrating out the vector $\boldsymbol{\delta}_a$ can be done one parameter at a time, as each observation is restricted to a single terminal node. However, in TDLMM, the exposure observations reside across all terminal nodes requiring us to simultaneously integrate over the entire vector, $\boldsymbol{\delta}_a$. The marginal likelihood of $\mathbf{R}_a$ is calculated to be

$$\begin{aligned} p(\mathbf{R}_a|\mathcal{T}_{a_j}-) &= \int_{\sigma^2}\int_{\boldsymbol{\delta}_a} p(\mathbf{R}_a|\boldsymbol{\delta}_a,\mathcal{T}_{a_j},-)p(\boldsymbol{\delta}_a|\mathcal{T}_a,-)p(\sigma^2)\,d\boldsymbol{\delta}_a\,d\sigma^2 \\ &= \left(\nu^2\right)^{-(B_{a_1}+B_{a_2}+B_{a_1}B_{a_2})/2}\left(\mu^2_{S_{a_1}}\right)^{-B_{a_1}/2}\left(\mu^2_{S_{a_2}}\right)^{-B_{a_2}/2}\left(\mu^2_{S_{a_1}S_{a_2}}\right)^{-B_{a_1}B_{a_2}/2} \\ &\quad\times|\mathbf{V}_{\boldsymbol{\delta}_a}|^{1/2}\left[\frac{\mathbf{R}_a^T\left(\mathbf{I}_n-\mathbf{Z}^T\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z}-\mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{X}_a\mathbf{V}_{\boldsymbol{\delta}_a}\mathbf{X}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\right)\mathbf{R}_a}{2}+\frac{1}{\xi_{\sigma^2}}\right]^{-(n+1)/2} \end{aligned} \tag{B.9}$$

where ξ_{σ^2} is from the hierarchy $\sigma^2|\xi_{\sigma^2} \sim \mathcal{IG}(1/2, 1/\xi_{\sigma^2})$ and $\xi_{\sigma^2} \sim \mathcal{IG}(1/2, 1)$ such that $\sigma \sim \mathcal{C}^+(0,1)$. We then calculate $p(\mathbf{R}_a, \mathcal{T}_{a_j}|-) = p(\mathbf{R}_a|\mathcal{T}_{a_j}, -)p(\mathcal{T}_{a_j})$ and accept $\mathcal{T}^*_{a_j}$ according to the

Metropolis-Hastings ratio given by

$$r = \min\left\{1, \frac{p(\mathcal{T}_{a_j}^*)p(\mathbf{R}_a|\mathcal{T}_{a_j}^*, -)p(\mathcal{T}_{a_j}|\mathcal{T}_{a_j}^*)}{p(\mathcal{T}_{a_j})p(\mathbf{R}_a|\mathcal{T}_{a_j}, -)p(\mathcal{T}_{a_j}^*|\mathcal{T}_{a_j})}\right\}. \tag{B.10}$$

The same technique is used to update exposure S to S', however the MH ratio simplifies to the ratio of marginal likelihoods: $p(\mathbf{R}_a|S', -)$ and $p(\mathbf{R}_a|S, -)$.

B.2.3 Full conditionals

After the tree update, we draw $\boldsymbol{\delta}_a$ (all parameters specific to tree pair a) from the full conditional distribution,

$$\boldsymbol{\delta}_a|- \sim \mathcal{MVN}_{B_a}\left(\mathbf{V}_{\boldsymbol{\delta}_a}\mathbf{X}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{R}_a, \sigma^2\mathbf{V}_{\boldsymbol{\delta}_a}\right), \tag{B.11}$$

where

$$\mathbf{V}_{\boldsymbol{\delta}_a} = \left(\mathbf{X}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{X}_a + \nu^{-2}\tau_a^{-2}\mathbf{U}_a\right)^{-1}. \tag{B.12}$$

Here, $\mathbf{U}_a$ is a diagonal matrix of exposure-specific variances. The first B_{a_1} diagonal elements equal $\mu_{S_{a_1}}^{-2}$, the next B_{a_2} elements equal $\mu_{S_{a_2}}^{-2}$ and the final $B_{a_1}B_{a_2}$ diagonal elements equal $\mu_{S_{a_1}S_{a_2}}^{-2}$. We note here that the computation time of $\mathbf{V}_{\boldsymbol{\delta}_a}$ is no more than $\mathcal{O}(n^2(B_{a_1} + B_{a_2} + B_{a_1}B_{a_2})^3)$. Since B_{a_j} is limited by the size of the trees, computation time of our algorithm grows at the rate of $\mathcal{O}(n^2)$. Pre-computing many of the values reduces model run time.

The update of each tree and corresponding partial DLM is followed by a draw from the full conditional of remaining parameters and hyperparameters:

$$\xi_\sigma|- \sim \mathcal{IG}\left(1, 1 + \frac{1}{\sigma^2}\right); \tag{B.13}$$

$$\sigma^2|- \sim \mathcal{IG}\left[\frac{n+b+1}{2}, \frac{\|\mathbf{V}_{\mathbf{Z}}^{-1/2}(\mathbf{y}-\mathbf{f})\|_2^2}{2} + \frac{D}{2\nu^2} + \frac{1}{\xi_\sigma}\right]; \tag{B.14}$$

where $b = \sum_{a=1}^{A} B_{a_1} + B_{a_2} + B_{a_1} B_{a_2}$ and $D = \sum_{a=1}^{A} \boldsymbol{\delta}_a^T \mathbf{U}_a \boldsymbol{\delta}_a$. Also,

$$\xi_\nu | - \sim \mathcal{IG}\left(1, 1 + \frac{1}{\nu^2}\right); \tag{B.15}$$

$$\nu^2 | - \sim \mathcal{IG}\left(\frac{b+1}{2}, \frac{D}{2\sigma^2} + \frac{1}{\xi_\nu}\right); \tag{B.16}$$

$$\xi_{\mu_m} | - \sim \mathcal{IG}\left(1, 1 + \frac{1}{\mu_m^2}\right); \tag{B.17}$$

$$\mu_m | - \sim \mathcal{IG}\left(\frac{b_m + 1}{2}, \frac{D_m}{2\sigma^2\nu^2} + \frac{1}{\xi_{\mu_m}}\right); \tag{B.18}$$

where $b_m = \sum_{a=1}^{A} B_{a_1} \mathbb{I}(S_{a_1} = m) + B_{a_2} \mathbb{I}(S_{a_2} = m)$, $D_m = \sum_{a=1}^{A} \boldsymbol{\delta}_{a_1}^T \boldsymbol{\delta}_{a_1} \mathbb{I}(S_{a_2} = m) + \boldsymbol{\delta}_{a_2}^T \boldsymbol{\delta}_{a_2} \mathbb{I}(S_{a_2} = m)$ and $\boldsymbol{\delta}_{a_1}$ is the portion of $\boldsymbol{\delta}_a$ corresponding to tree $\mathcal{T}_{a_1}$. Updates for $\mu_{m_1 m_2}$ follow similarly.

Updates of $\mathcal{E}$ come from full conditional

$$\mathcal{E} | - \sim \text{Dirichlet}\left(\kappa + N_{\{S=1\}}, \ldots, \kappa + N_{\{S=M\}}\right) \tag{B.19}$$

where $N_{\{S=m\}}$ is the number of trees that use exposure m.

Finally we draw $\boldsymbol{\gamma}$ from its full conditional,

$$\boldsymbol{\gamma} | \mathbf{y}, \mathbf{f}, \sigma^2 \sim \mathcal{MVN}\left[\mathbf{V}_{\boldsymbol{\gamma}} \mathbf{Z}(\mathbf{y} - \mathbf{f}), \sigma^2 \mathbf{V}_{\boldsymbol{\gamma}}\right]. \tag{B.20}$$

B.2.4 Posterior analysis of exposure and interaction variance

The exposure variance parameter μ_m^2 is subject to inferential issues due to differences in the scale of exposure measurements and because it is included in the effect estimates across possibly many trees. To make inference on the effect size of exposures or interactions, we scale each exposure in TDLMM to have standard deviation one and devise an effect size metric based on the prior distributions of the node specific effects. We define the exposure effect size metric for

exposure m as

$$\mu_m^* = \mu_m^2 \sum_{a=1}^{A} \left[\mathbb{I}(S_{a_1} = m) + \mathbb{I}(S_{a_2} = m) \right], \tag{B.21}$$

which is the sum of variances for all trees using exposure m. After calculating μ_m^* for each exposure m we determine the relative rank of exposure effect sizes, which is the ordered rank of the effect size metrics scaled to fall in the range $[0, 1]$. The relative rank allows for interpretation of the effect size of an exposure relative to that of other exposures. A relative rank near one represents the largest effect size, a relative rank near 0.5 indicates average effect size and a relative rank near zero indicates the smallest effect size.

For the interaction of exposures m_1 and m_2, the interaction effect size metric is defined similarly as

$$\mu_{m_1 m_2}^* = \mu_{m_1 m_2}^2 \sum_{a=1}^{A} \mathbb{I}(S_{a_1} = m_1, S_{a_2} = m_2) \tag{B.22}$$

where the indicator function equals 1 if exposures m_1 and m_2 are both included in the tree pair. We determine the relative ranks of interaction effect sizes separately from the relative ranks of exposure effects.

B.2.5 Logistic model

Consider a binomial response, y_i, ($i \in \{1, \ldots, N\}$) with log-odds of success ψ_i,

$$y_i | \psi_i \sim \text{Bin} \left[n_i, \frac{1}{1 + \exp(-\psi_i)} \right] \tag{B.23}$$

where

$$\psi_i = f(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iM}) + \mathbf{z}_i^T \boldsymbol{\gamma} \tag{B.24}$$

indicates the DLMM plus a fixed effect. By Theorem 1 of Polson, Scott, and Windel (2013),

$$p(y_i | \psi_i) = \binom{n_i}{y_i} \frac{\exp(\psi_i)^{y_i}}{(1 + \exp(\psi_i))^{n_i}} = \binom{n_i}{y_i} \frac{e^{k_i \psi_i}}{2^{n_i}} \mathbb{E}_{\omega_i | n_i, 0}[\exp(-\omega_i \psi^2 / 2)] \tag{B.25}$$

where $k_i = y_i - n_i/2$ and $\omega_i | n_i, 0 \sim \text{PG}(n_i, 0)$ refers to the Pólya Gamma distribution.

We define a latent variable, ω_i, specified by the distribution

$$\omega_i|\psi_i \sim \text{PG}(n_i, \psi_i). \tag{B.26}$$

The density of ω_i can be written

$$p(\omega_i|\psi_i) = \frac{e^{-\omega_i\psi^2/2}p(\omega_i|n_i, 0)}{\mathbb{E}_{\omega_i|n_i,0}[\exp(-\omega_i\psi^2/2)]}. \tag{B.27}$$

Note that y_i and ω_i are conditionally independent. Consider their joint density,

$$p(y_i, \omega_i|\psi_i) = p(y_i|\psi_i)p(\omega_i|\psi_i) \propto \exp\left(-\frac{\omega_i}{2}(\lambda_i - \psi_i)^2\right) \tag{B.28}$$

where $\lambda_i = (y_i - n_i/2)/\omega_i$. Also,

$$p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\psi}) = \prod_{i=1}^{N} p(y_i, \omega_i|\psi_i) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{\psi})^T\Omega(\boldsymbol{\lambda} - \boldsymbol{\psi})\right\} \tag{B.29}$$

where $\Omega = \text{diag}\{\omega_1, \ldots, \omega_N\}$ and $\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}$ are column vectors of y_i, ω_i, and λ_i, respectively. We may now develop a Bayesian model for $\boldsymbol{\gamma}$ and $\mathbf{f}$ conditional on observed $\mathbf{y}$ and latent variable $\boldsymbol{\omega}$, where an MCMC algorithm consists of iterating updates of $\boldsymbol{\omega}$ followed by $\boldsymbol{\psi}$.

B.3 Prior inclusion probability

The prior inclusion probability is the probability an exposure is used in at least one tree in the model. Let $C_m = \sum_{a=1}^{A}[\mathbb{I}(S_{a_1} = m) + \mathbb{I}(S_{a_2} = m)]$ be the count of trees using exposure m. Then C_m follows the Beta-Binomial distribution with $2A$ draws and prior selection probability based on the Dirichlet distribution with fixed parameter κ. The prior probability of inclusion, $I_m = \mathbb{I}(C_m > 0)$ is

$$\mathbb{P}(I_m = 1) = 1 - \frac{\Gamma(2A + M\kappa) \cdot \Gamma\left[(M + 1)\kappa\right]}{\Gamma\left[2A + (M + 1)\kappa\right] \cdot \Gamma(M\kappa)}. \tag{B.30}$$

Higher prior inclusion probability corresponds to a less informative prior belief regarding the number of exposures relevant to the outcome, while lower prior inclusion probability indicates more sparsity. Examples of prior inclusion probabilities are shown in Figure B.1.

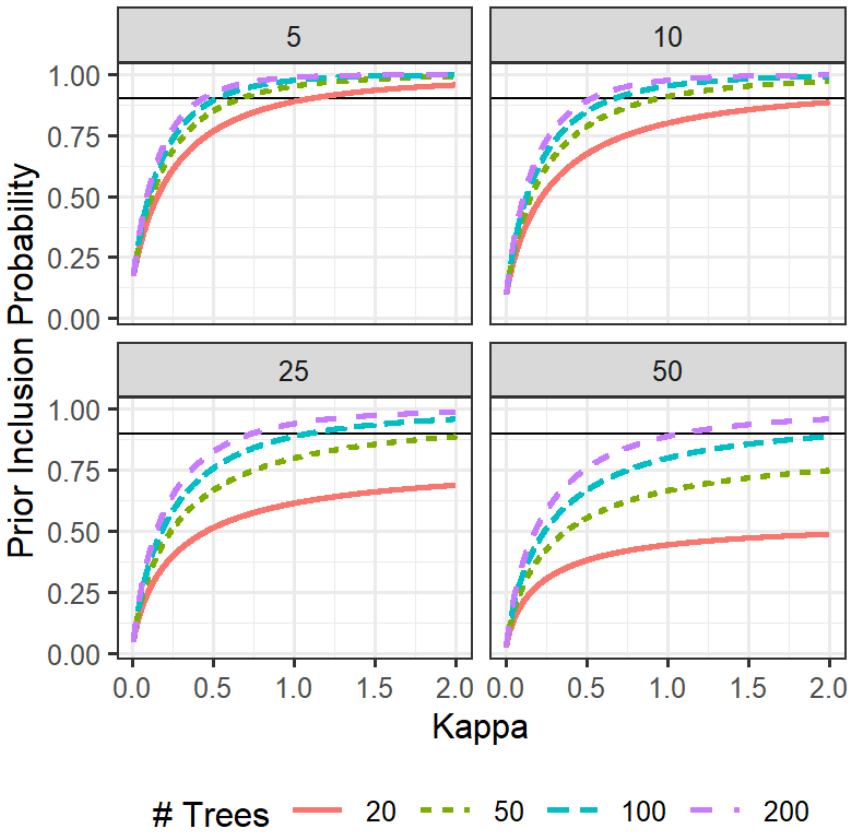


Figure B.1: Exposure prior inclusion probabilities (y-axis) for different settings of κ (x-axis), number of trees (color and style of line), and number of exposures (panel). The horizontal black line at $y = 0.9$ corresponds to the prior inclusion probability used in our simulation and data analysis.

B.3.1 Setting A and κ

For a given number of exposures, M, we can set the number of trees, A, and Dirichlet parameter κ to give the desired prior inclusion probability. A smaller number of trees will result in lower variance estimates and it may be desirable to use a smaller number of trees when possible as we have validated that coverage is acceptable. We note that the Bayes factor selection approach described

in Appendix B.3.2 is not sensitive to these settings, however inference using posterior inclusion probability should be performed with more care and relate to the prior inclusion probabilities.

For our simulation and data analysis with 5 exposures and prior inclusion probability of 0.9, we used 20 trees with kappa = 1.089 (in Figure B.1 note intersection of black and red solid lines in top left panel). On the other hand, a model with 25 exposures would likely require 100 trees to achieve the same prior inclusion probability.

B.3.2 Bayes factor method for exposure selection

Let $I_m^{(r)}$, $r \in \{1, \ldots, R\}$ refer to the inclusion of exposure m in MCMC sample r. Using the method of Carlin and Chib (1995), the marginal posterior probability that exposure m is included in the model can be estimated by

$$\widehat{\mathbb{P}}(I_m = 1|\mathbf{y}) = R^{-1} \sum_{r=1}^{R} \mathbb{I}(I_m^{(r)} = 1). \tag{B.31}$$

We apply the marginal posterior probability to approximate Bayes factors for exposure variable selection. Specifically, the Bayes factor that exposure m is included in the model is estimated by

$$\widehat{\text{BF}}_m = \frac{\widehat{\mathbb{P}}(I_m = 1|\mathbf{y})/\widehat{\mathbb{P}}(I_m = 0|\mathbf{y})}{\mathbb{P}(I_m = 1)/\mathbb{P}(I_m = 0)} \tag{B.32}$$

where the denominator refers to the prior odds given in (B.30). We consider $\log_{10} \widehat{\text{BF}}_m > 0.5$ as a criteria for substantial evidence that exposure m is associated with the response (Kass and Raftery, 1995).

B.4 Additional simulation results

B.4.1 Scenario 1: Comparing TDLM without tree-specific shrinkage

To justify the tree-specific shrinkage parameters (τ_a) in TDLM, we repeated simulation scenario 1 fixing $\tau_a = 1$. The results are given in Table B.3. Including this additional term results in lower RMSE without changing coverage and precision.

Table B.3: Simulation scenario 1 comparing TDLM with and without tree-specific shrinkage parameters τ_a.

	DLM Est.		Critical Windows		
Model	RMSE $\times 100$	Coverage	TP	FP	Precision
$\overline{p} = 0.5$					
Original TDLM	1.22	0.98	0.98	0.02	0.98
Fix $\tau_a = 1$	1.55	0.96	1.00	0.03	0.97
$\overline{p} = 0.05$					
Original TDLM	2.17	0.96	0.88	0.03	0.97
Fix $\tau_a = 1$	2.32	0.95	0.94	0.03	0.97

$^* \times 100$

B.4.2 Scenario 2: Comparing TDLMM with different shrinkage priors

We repeat simulation scenario 2 to compare TDLMM without exposure- and interaction-specific shrinkage priors as well as including tree-pair-specific shrinkage priors. Results are given in Tables B.4 and B.5. We note that excluding the exposure- and interaction-specific shrinkage priors increases RMSE and creates below nominal coverage. In addition the FP rate increases and the posterior inclusion probability on nonactive exposures and interactions increases. Including a tree-pair-specific shrinkage parameter does not improve performance across all measures.

B.4.3 Single exposure with smooth DLM and binary outcome

We replicated simulation scenario one with a smooth distributed lag effect, defined

$$f_1(\mathbf{x}) = \frac{2}{21} \sum_{t=s}^{s+7} x_t \cdot \max\{0, -(t-s)(t-s-8)\}. \tag{B.33}$$

Here, s is a random starting time and was drawn uniformly from $\{1, \ldots, T-7\}$. The distributed lag effect, f_1 is a quadratic function and the cumulative effect is equal to the cumulative effect in simulation scenario one. Results are given in Table B.6. We see the tree-based models again have lower RMSE than established single exposure DLMs. We also see that TDLM and TDLMM maintain high coverage and TP with precision near one.

B.4.4 Multiple exposures with continuous outcome

Appendix Table B.7 describes the average relative rank of the effect size metric for main and interaction effects as described in Appendix B.2.4. The effect size for active exposures was higher than for other exposures; the same was true comparing interactions.

Figures B.2 and B.3 show the variation of posterior inclusion and effect size results from simulation scenario two. These figures show the active exposures or interactions compared to the non-active exposures or interactions with maximum or minimum posterior inclusion probability.

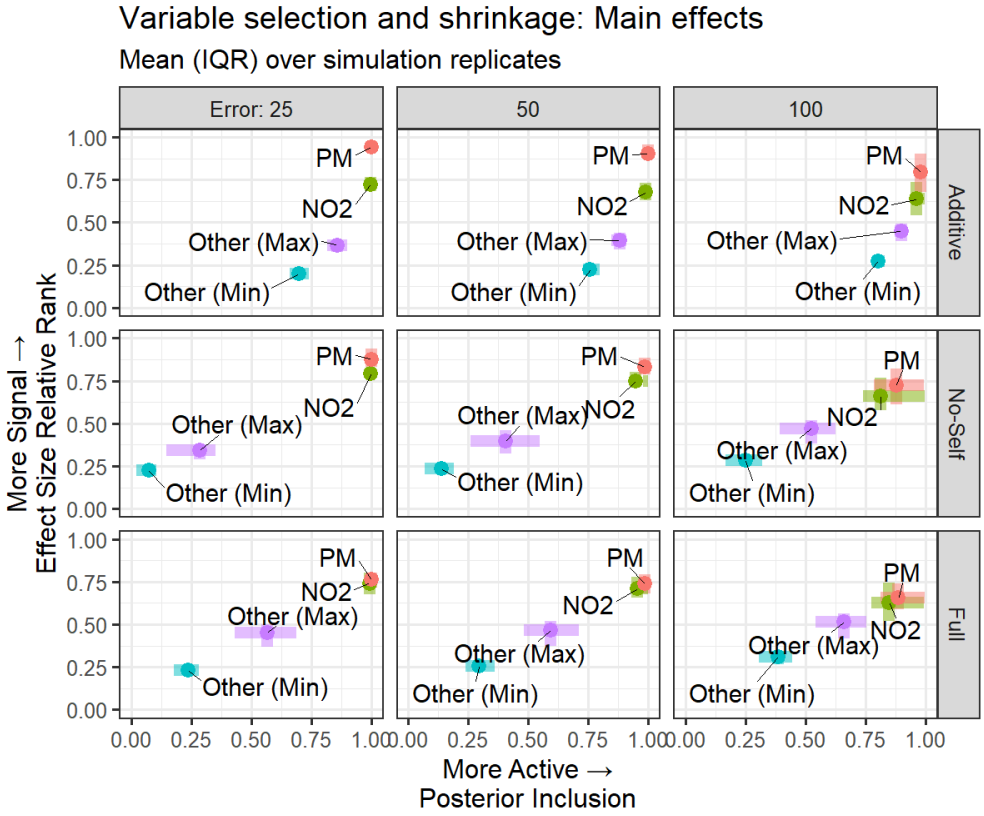


Figure B.2: Variation in exposure selection and shrinkage results from simulation scenario two. Points show the average posterior inclusion (x-axis) and effect size (y-axis) while the error bars indicate the IQR of the metric across simulation replicates. Other max and min refer to the non-active exposure in each replicate with the largest or smallest posterior inclusion, respectively.

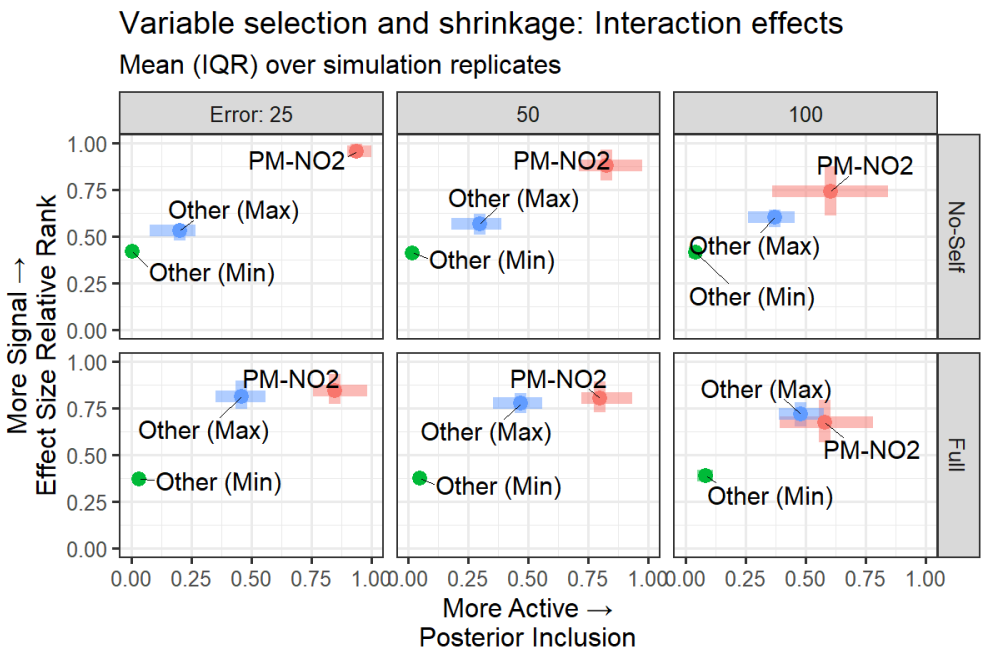


Figure B.3: Variation in interaction selection and shrinkage results from simulation scenario two. Points show the average posterior inclusion (x-axis) and effect size (y-axis) while the error bars indicate the IQR of the metric across simulation replicates. Other max and min refer to the non-active interactions in each replicate with the largest or smallest posterior inclusion, respectively.

Table B.4: Scenario 2, comparing TDLMM and variants under different shrinkage parameters. Originally specified is as described in the main text, tree-specific variance adds an additional local shrinkage parameter to each tree pair, fix $\mu_m, \mu_{m_1 m_2} = 1$ removes local shrinkage parameters from the exposure.

	RMSE $\times 100$		Coverage		TP		FP		
Model	PM	NO_2	PM	NO_2	PM	NO_2	PM	NO_2	Other
$\sigma^2 = 25$									
Originally specified models									
TDLMMadd	3.59	4.19	0.95	0.84	0.94	0.82	0.03	0.07	0.00
TDLMMns	3.57	4.27	0.97	0.96	0.91	0.63	0.02	0.03	0.00
TDLMM	3.55	4.33	0.98	0.97	0.87	0.51	0.01	0.02	0.00
Add tree-specific variance parameter, τ_a									
TDLMMadd	3.38	4.17	0.96	0.83	0.91	0.84	0.02	0.08	0.00
TDLMMns	3.56	4.29	0.98	0.94	0.90	0.66	0.02	0.04	0.00
TDLMM	3.52	4.32	0.98	0.96	0.85	0.56	0.01	0.02	0.00
Fix $\mu_m, \mu_{m_1 m_2} = 1$									
TDLMMadd	3.88	4.25	0.91	0.85	0.98	0.78	0.05	0.06	0.00
TDLMMns	4.02	4.46	0.93	0.91	0.95	0.51	0.05	0.03	0.00
TDLMM	4.45	4.27	0.86	0.92	0.93	0.63	0.02	0.03	0.00

	RMSE $\times 100$		Coverage		TP		FP		
Model	PM	NO_2	PM	NO_2	PM	NO_2	PM	NO_2	Other
$\sigma^2 = 50$									
Originally specified models									
TDLMMadd	4.62	4.64	0.92	0.83	0.86	0.52	0.04	0.05	0.00
TDLMMns	4.61	4.75	0.96	0.97	0.76	0.31	0.02	0.02	0.00
TDLMM	4.50	4.79	0.97	0.97	0.62	0.22	0.01	0.01	0.00
Add tree-specific variance parameter, τ_a									
TDLMMadd	4.51	4.65	0.93	0.81	0.82	0.56	0.03	0.07	0.00
TDLMMns	4.51	4.73	0.97	0.94	0.72	0.34	0.02	0.03	0.00
TDLMM	4.52	4.82	0.97	0.97	0.62	0.21	0.01	0.01	0.00
Fix $\mu_m, \mu_{m_1 m_2} = 1$									
TDLMMadd	4.90	4.73	0.87	0.82	0.93	0.48	0.06	0.06	0.00
TDLMMns	5.17	4.90	0.89	0.91	0.83	0.32	0.06	0.04	0.00
TDLMM	5.34	4.73	0.83	0.89	0.80	0.36	0.03	0.03	0.00

	RMSE $\times 100$		Coverage		TP		FP		
Model	PM	NO_2	PM	NO_2	PM	NO_2	PM	NO_2	Other
$\sigma^2 = 100$									
Originally specified models									
TDLMMadd	5.75	5.07	0.89	0.84	0.47	0.22	0.03	0.03	0.00
TDLMMns	5.85	5.21	0.93	0.95	0.29	0.10	0.02	0.01	0.00
TDLMM	5.73	5.21	0.94	0.96	0.24	0.05	0.01	0.01	0.00
Add tree-specific variance parameter, τ_a									
TDLMMadd	5.75	5.09	0.90	0.84	0.43	0.23	0.03	0.03	0.00
TDLMMns	5.82	5.30	0.93	0.94	0.32	0.10	0.02	0.01	0.00
TDLMM	5.75	5.33	0.94	0.97	0.20	0.04	0.01	0.01	0.00
Fix $\mu_m, \mu_{m_1 m_2} = 1$									
TDLMMadd	5.97	5.12	0.84	0.81	0.50	0.24	0.04	0.04	0.00
TDLMMns	6.25	5.19	0.86	0.91	0.33	0.18	0.03	0.03	0.00
TDLMM	6.26	5.10	0.82	0.88	0.35	0.17	0.02	0.02	0.00

Table B.5: Scenario 2, exposure and interaction posterior inclusion probability for TDLMM and variants using different shrinkage priors. Originally specified is as described in the main text, tree-specific variance adds an additional local shrinkage parameter to each tree pair, fix $\mu_m, \mu_{m_1 m_2} = 1$ removes local shrinkage parameters from the exposure.

	Main Effect			Interaction	
Model	PM	NO_2	Other	PM$-NO_2$	Other
$\sigma^2 = 25$					
Originally specified models					
TDLMMadd	1.00	1.00	0.78	-	-
TDLMMns	1.00	1.00	0.17	0.94	0.06
TDLMM	1.00	0.99	0.39	0.85	0.22
Add tree-specific variance parameter, τ_a					
TDLMMadd	1.00	1.00	0.80	-	-
TDLMMns	1.00	1.00	0.28	0.93	0.11
TDLMM	1.00	1.00	0.43	0.90	0.23
Fix $\mu_m, \mu_{m_1 m_2} = 1$					
TDLMMadd	1.00	1.00	0.66	-	-
TDLMMns	1.00	1.00	0.74	0.68	0.24
TDLMM	1.00	1.00	0.70	0.96	0.40

	Main Effect			Interaction	
Model	PM	NO_2	Other	PM$-NO_2$	Other
$\sigma^2 = 50$					
Originally specified models					
TDLMMadd	1.00	0.99	0.82	-	-
TDLMMns	0.99	0.95	0.26	0.83	0.10
TDLMM	0.99	0.96	0.43	0.80	0.23
Add tree-specific variance parameter, τ_a					
TDLMMadd	1.00	0.99	0.83	-	-
TDLMMns	0.99	0.96	0.36	0.85	0.14
TDLMM	0.99	0.96	0.49	0.79	0.25
Fix $\mu_m, \mu_{m_1 m_2} = 1$					
TDLMMadd	1.00	0.99	0.77	-	-
TDLMMns	0.99	0.98	0.77	0.79	0.36
TDLMM	1.00	0.99	0.79	0.91	0.45

	Main Effect			Interaction	
Model	PM	NO_2	Other	PM$-NO_2$	Other
$\sigma^2 = 100$					
Originally specified models					
TDLMMadd	0.98	0.96	0.85	-	-
TDLMMns	0.88	0.81	0.38	0.61	0.15
TDLMM	0.89	0.85	0.52	0.58	0.25
Add tree-specific variance parameter, τ_a					
TDLMMadd	0.98	0.97	0.86	-	-
TDLMMns	0.91	0.85	0.47	0.65	0.19
TDLMM	0.90	0.88	0.56	0.60	0.27
Fix $\mu_m, \mu_{m_1 m_2} = 1$					
TDLMMadd	0.99	0.97	0.85	-	-
TDLMMns	0.98	0.97	0.84	0.81	0.47
TDLMM	0.98	0.97	0.86	0.83	0.49

Table B.6: Results for simulation scenario one with smooth DLM effect. The DLM estimation refers to estimation of the active exposure effect only. Effect identification indicates when the DLM 95% CI does not contain zero at correct (TP) or incorrect (FP) time periods and Precision $= \text{TP}/(\text{TP} + \text{FP})$.

	DLM Estimation		Effect Identification		
Model	RMSE$\times$100	Coverage	TP	FP	Precision
$\overline{p} = 0.5$					
CWVS: dlm|cw	2.16	0.98	0.94	0.02	0.98
CWVS: p>0.5	2.16	0.98	0.97	0.05	0.95
DLMcr	1.67	0.83	1.00	0.13	0.88
TDLM	1.38	0.98	0.87	0.01	0.99
TDLMMadd	1.37	0.98	0.88	0.01	0.99
TDLMMns	1.46	0.98	0.93	0.01	0.99
TDLMM	1.45	0.96	0.91	0.03	0.97
$\overline{p} = 0.05$					
CWVS: dlm|cw	3.58	0.99	0.72	0.01	0.99
CWVS: p>0.5	3.58	0.99	0.90	0.09	0.91
DLMcr	2.86	0.78	1.00	0.17	0.85
TDLM	2.21	0.98	0.78	0.02	0.98
TDLMMadd	2.43	0.96	0.82	0.02	0.98
TDLMMns	2.53	0.95	0.86	0.03	0.97
TDLMM	2.62	0.95	0.85	0.03	0.97

Table B.7: Exposure effect size results for simulation scenario two: main effect of $PM_{2.5}$ with $PM_{2.5}-NO_2$ interaction. Relative rank of the effect size describes the average rank of the effect size metric on a scale from 0 to 1, separately for main and interaction effects.

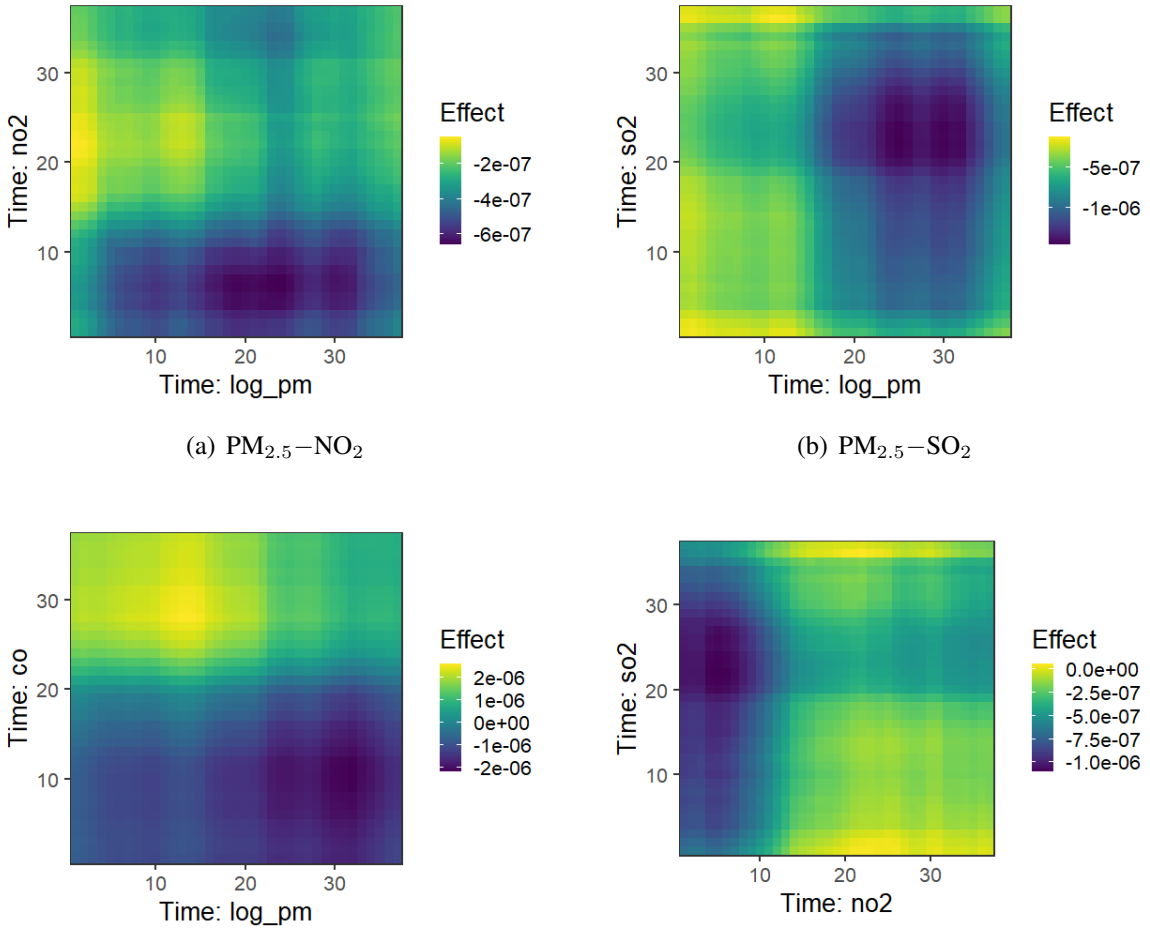
	Effect Size Relative Rank				
	Main Effect			Interaction	
Model	PM	NO_2	Other	PM$-NO_2$	Other
$\sigma^2 = 25$					
TDLMMadd	0.93	0.72	0.28	-	-
TDLMMns	0.80	0.70	0.33	0.88	0.46
TDLMM	0.72	0.69	0.37	0.81	0.48
$\sigma^2 = 50$					
TDLMMadd	0.89	0.67	0.31	-	-
TDLMMns	0.74	0.65	0.37	0.79	0.47
TDLMM	0.67	0.64	0.40	0.73	0.48
$\sigma^2 = 100$					
TDLMMadd	0.79	0.63	0.36	-	-
TDLMMns	0.66	0.60	0.41	0.67	0.48
TDLMM	0.58	0.60	0.44	0.62	0.49

B.5 Additional data analysis results

B.5.1 Additional figures from data analysis

Figures B.4 and B.5 show additional interaction plots from the data analysis.

Figure B.6 describes variable selection and effect size results for the data analysis. We see that NO_2 had the smallest effect size, on average, while the other four exposures had similar effect sizes. $PM_{2.5}-$temperature had the largest effect size on average.



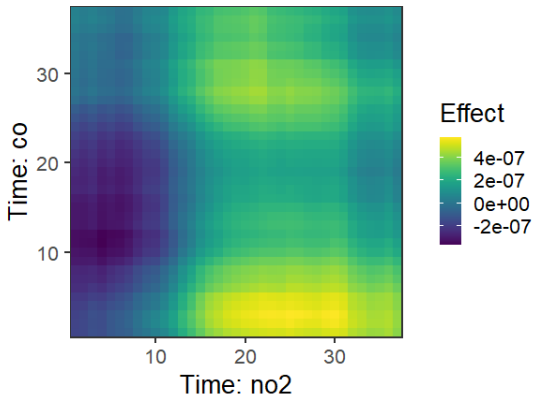
(a) $PM_{2.5}-NO_2$

(b) $PM_{2.5}-SO_2$

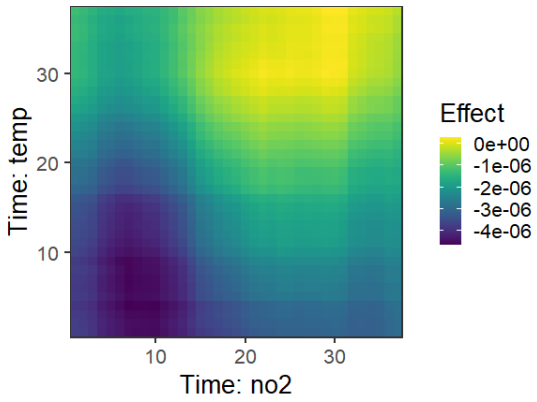
(c) $PM_{2.5}-CO$

(d) NO_2-SO_2

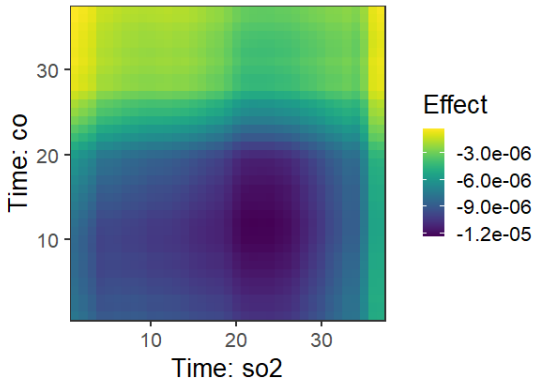
Figure B.4: Additional interaction plots from data analysis.



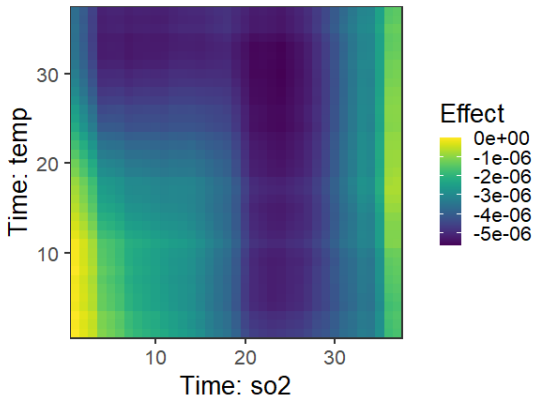
(a) NO_2-CO



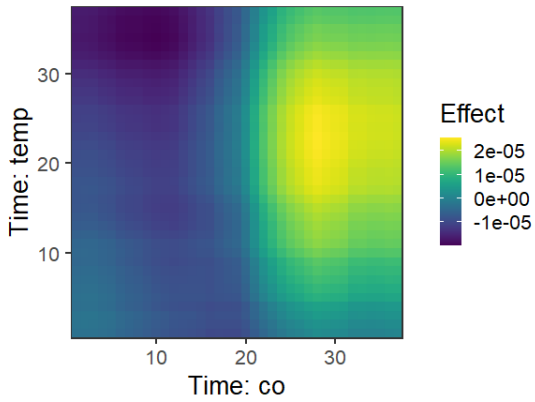
(b) NO_2-temperature



(c) SO_2-CO



(d) SO_2-temperature



(e) CO−temperature

Figure B.5: Additional interaction plots from data analysis.

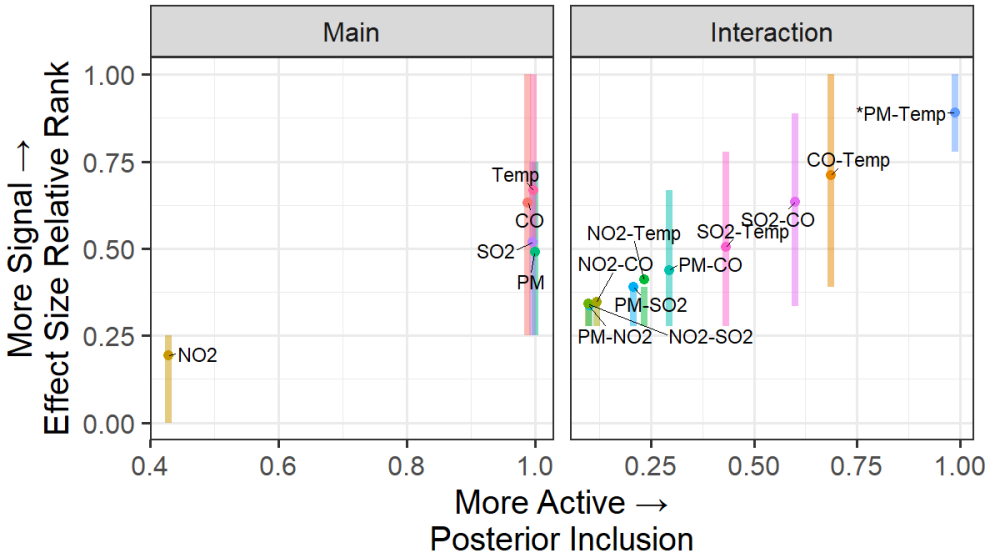


Figure B.6: Posterior inclusion probability (x-axis) and effect size relative rank (y-axis) results from the data analysis. The left panel shows posterior inclusion and effect size results for the five main effects and the right panel contains the same for the ten possible interactions. Exposures or interactions marked with an '*' indicate selection based the Bayes factor analysis.

B.5.2 TDLMM including within-exposure interactions

We reanalyze the data in the main text using TDLMM including all within-exposure interactions. This model was run for 50,000 iterations thinned to every fifth iteration following 5,000 burn-in iterations. Figure B.7 shows the estimated marginal DLM (top row) and DLM adjusting for changes in co-exposures (bottom row). We note the shape and effect sizes from TDLMM are similar to the original data analysis. The DLM and effect of $PM_{2.5}$ is nearly unchanged. Several differences include:

- Shorter critical window for SO_2: weeks 17-28
- No critical windows due to CO

In terms of exposure and interaction selection, $PM_{2.5}$ and temperature have a posterior inclusion of 1.00, while SO_2 (0.9992) and CO (0.9851) also show very high posterior inclusion. NO_2 has a posterior inclusion of 0.5511, similar to the data analysis with TDLMMns. For interactions, $PM_{2.5}-$temperature has the highest posterior inclusion (0.9955) followed by CO$-$temperature (0.8453) and temperature-temperature (0.7664). SO_2-temperature (0.6697) and SO_2-CO (0.5630) also have posterior inclusion above 0.5.

Figure B.8 describes the posterior inclusion and effect size of the main and interaction effects from the data analysis using TDLMM.

TDLMM also estimates within-exposure interaction effects, shown in Figure B.9. These plots are folded to show interaction effects between an earlier and later time period. Effects along the diagonal, at the time time period, represent possible non-linearities in the main effect of an exposure.

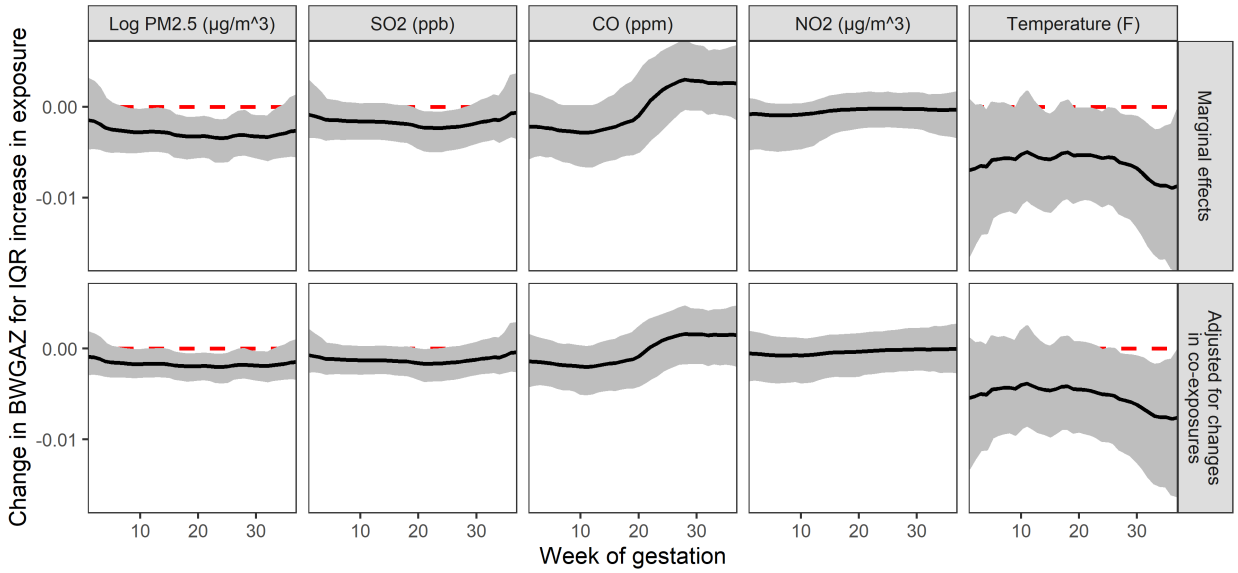


Figure B.7: Posterior mean distributed lag function (black-line) for each exposure (columns) with 95% credible interval (grey area) of the effect from TDLMM (including within-exposure interactions). The top row shows the marginal effect of an IQR increase in exposure, holding other exposures at their empirical mean. The bottom row shows the estimated change in BWGAZ for a first to third quartile change in one exposure along with expected changes in all other exposures due to correlation with the exposure of interest.

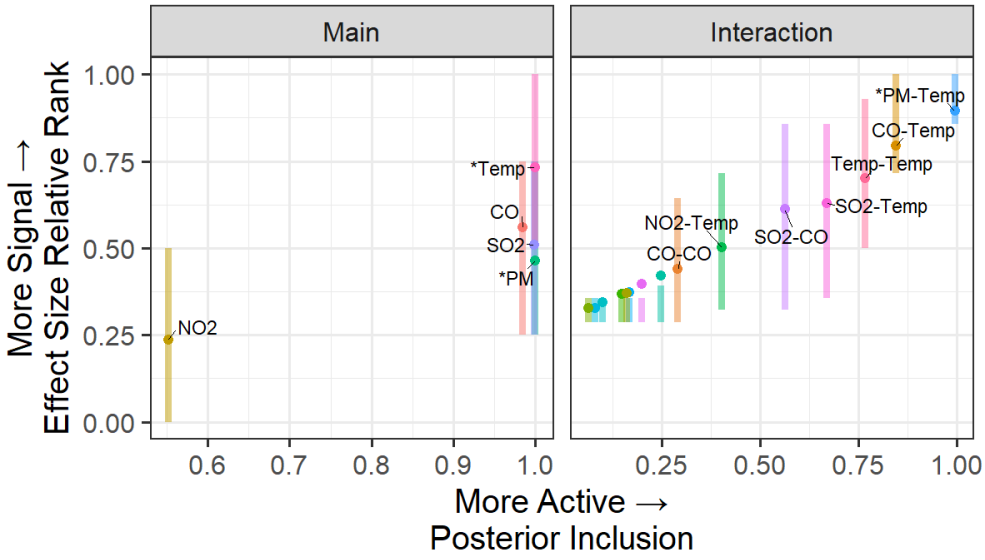
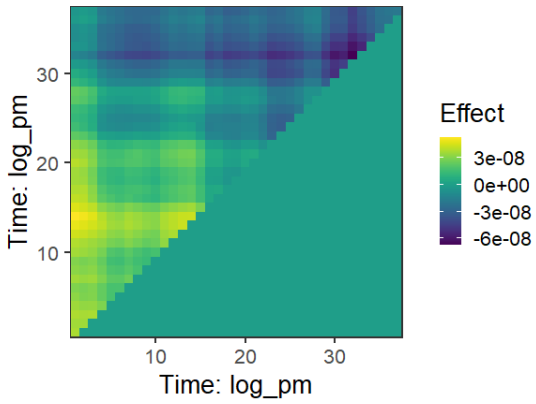
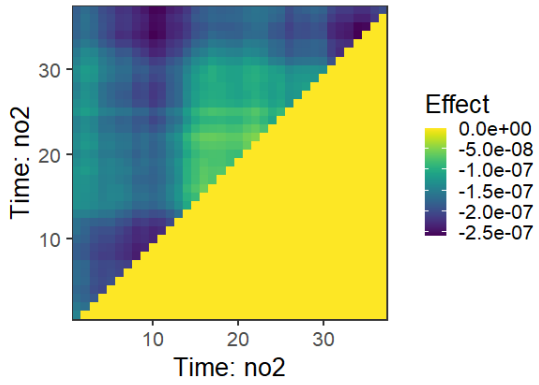


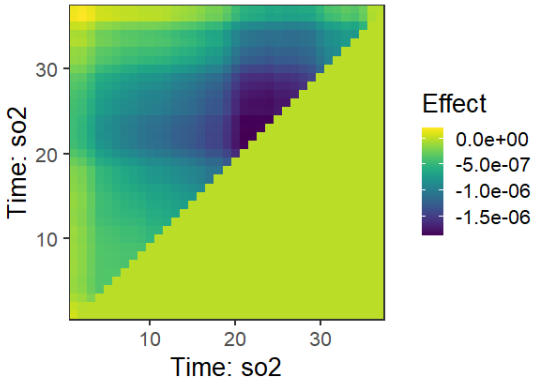
Figure B.8: Exposures selection (x-axis) and effect size (y-axis) results from TDLMM (including within-exposure interactions). The left panel shows posterior inclusion and effect size results for main effects while the right panel shows interaction effects. Exposures marked with an '*' indicate selection based on the Bayes factor analysis.



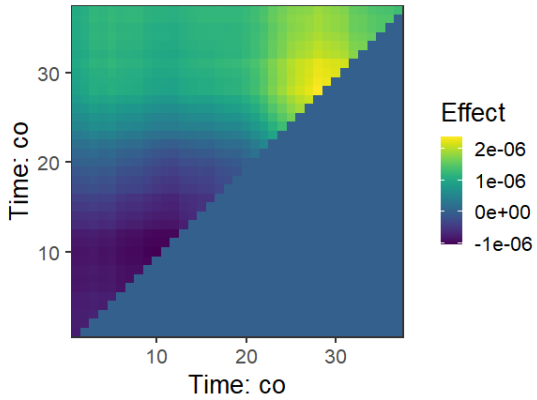
(a) $PM2.5-PM_{2.5}$



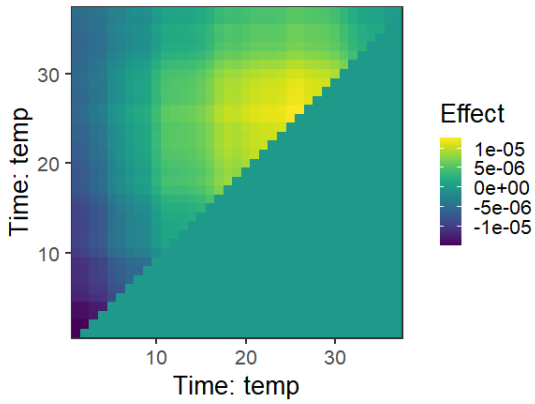
(b) NO_2-NO_2



(c) SO_2-SO_2



(d) CO−CO



(e) Temperature−temperature

Figure B.9: Within-exposure interaction plots from data analysis using TDLMM. Plots are folded to show the full interaction effect between an earlier and later time point. Effects at the same time point represent a non-linear effect.

Appendix C

Heterogeneous Distributed Lag Models to Estimate Personalized Effects of Maternal Exposures to Air Pollution

C.1 Colorado birth cohort data

Figures C.1 and C.2 present the distribution of modifying covariates used in the data analysis. Age and BMI are continuous modifiers; income, education, and smoking are ordinal modifiers; marital status, prenatal care, and race are categorical modifiers; Hispanic and sex are binary modifiers.

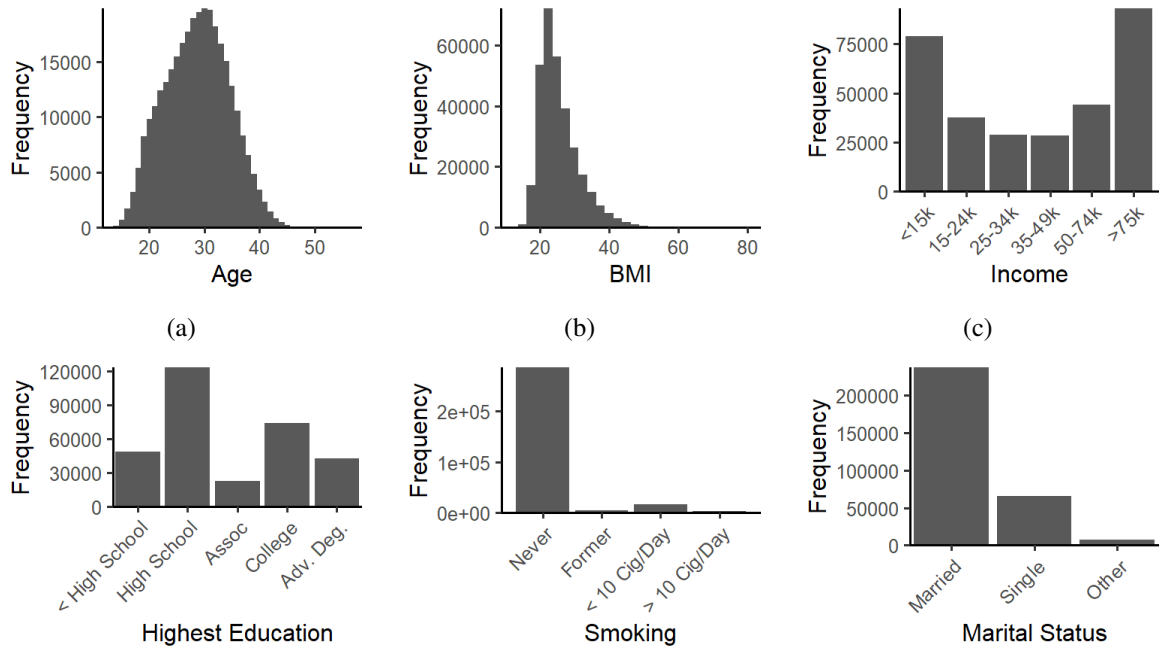


Figure C.1: Distribution of modifying covariates used in HDLM.

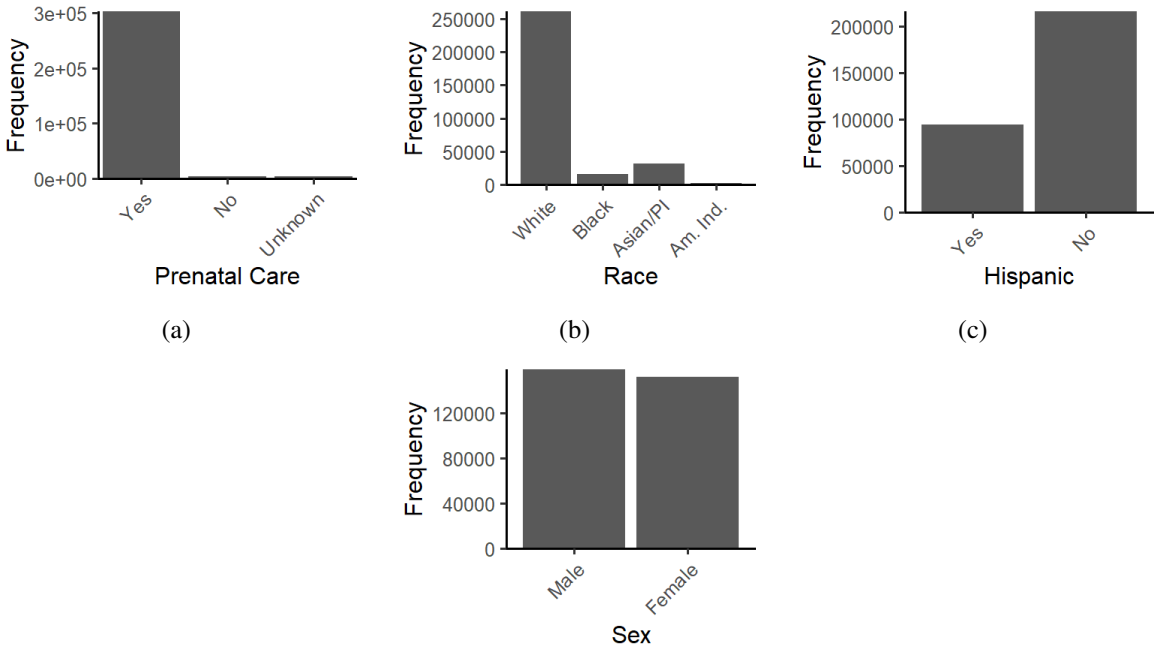


Figure C.2: Distribution of modifying covariates used in HDLM (continued).

C.2 Model Specification

In the shared tree and Gaussian process HDLMs, the modifier tree specification is identical to the nested tree HDLM. That is, the probability of a split at terminal node η with depth d_η is $p_{\text{split}}(\eta) = \alpha(1 + d_\eta)^{-\beta}$ where $\alpha \in (0, 1)$ and $\beta > 0$. We set $\alpha = 0.95$ and $\beta = 2$.

For selecting a splitting rule in the modifier tree, let $\rho = \{m_j, K\}$ define a splitting rule on modifying covariate m_j with splitting set K. A splitting set refers to a binary rule that places observations into one of two groups. For continuous m_j, K is an inequality and for categorical m_j, K is a proper subset of the categories. Define ψ_j to be the probability of selecting modifying covariate m_j, $j \in \{1, \ldots, J\}$, to be used for a binary rule. Then, the splitting rule prior is written

$$p_{\text{rule}}(\rho|\eta) = p(m_j|\eta)p(K|m_j, \eta). \tag{C.1}$$

Here, $p(m_j|\eta) = \psi_j / \sum_j \psi_j \mathbb{I}(m_j \in \mathcal{E}_\eta)$, where $\mathcal{E}_\eta$ is the set of eligible modifiers at node η. A modifier is eligible if the subgroup at node η can be divided into two nonempty subgroups using a binary

rule based on that modifier. Also, $p(K|m_j, \eta)$ is the probability of splitting set K for modifier m_j at node η. For a continuous modifier, $p(K|m_j, \eta) = 1/(n_{m_j,\eta} - 1)$ where $n_{m_j,\eta}$ is the number of splitting locations available for m_j at node η. If m_j is categorical, $p(K|m_j, \eta) = 1/(2^{n_{m_j,\eta}-1} - 1)$, which is based on the the number of partitions of $n_{m_j,\eta}$ unique values into two non-empty groups. Following Linero and Yang (2018), the prior on probabilities $\boldsymbol{\psi} = \{\psi_1, \ldots, \psi_J\}$ is

$$\boldsymbol{\psi} \sim \text{Dirichlet}(\kappa/J, \ldots, \kappa/J) \tag{C.2}$$

$$\frac{\kappa}{\kappa + J} \sim \text{Beta}(\zeta, 1) \tag{C.3}$$

where $\zeta \in (0.5, 1)$. Smaller values of ζ correspond to increased sparsity of modifiers. We use $\zeta = 0.5$ in our simulations and data analysis.

The fixed effects receive a non-informative prior, $\boldsymbol{\gamma} \sim \mathcal{MVN}(\mathbf{0}, c\sigma^2 I_p)$ where c is fixed at a large value and I_p is a $p \times p$ identity matrix. The variance parameter σ is given a half-Cauchy prior distribution and included in the fixed effects for integration during computation.

C.2.1 Shared Tree HDLM

As in the nested tree HDLM, the DLM effects are assigned the conjugate normal prior,

$$\delta_{abc}|\tau_a, \nu, \sigma \sim \mathcal{N}(0, \tau_a^2\nu^2\sigma^2). \tag{C.4}$$

Variance parameters τ_a, ν, σ follow a half-Cauchy distribution with scale $= 1$ and define a global-local horseshoe-like estimator on tree specific effects. Although the DLM structure and variance is the same for all modifier tree terminal nodes in the shared tree HDLM, each subgroup receives unique distributed lag effect estimates.

C.2.2 Gaussian Process HDLM

The Gaussian process prior for DLM effects is

$$\boldsymbol{\theta}_{ab}|\tau_a, \nu, \sigma, \phi \sim \mathcal{GP}[\mathbf{0}, \tau_a^2\nu^2\sigma^2\Sigma(\phi)]. \tag{C.5}$$

Variance parameters τ_a, ν, σ follow a half-Cauchy distribution with scale $= 1$ and define a global-local horseshoe-like estimator on tree specific effects. We use the exponential covariance matrix $\Sigma(\phi)$ with range parameter ϕ and restrict the range such that $\exp\{-\phi\} \in (0.05, 0.95)$. That is, the lag-1 covariance is between 0.05 and 0.95. We assign prior $\phi \sim$ Gamma$(1/2, 1/2)$, which gives higher probability to more smoothness in the HDLM.

C.3 Computational Approach

We describe the computation approach of the treed HDLMs in general and note where the algorithm changes depending on the method. The algorithm is based on the approach described by (Chipman et al., 2010) with differences to accommodate fixed effects, multivariate predictors, and the treed DLM for estimation of a vector of structured regression coefficients.

C.3.1 Preprocessing

Before running the treed HDLM algorithm, we perform the following operations to promote computational precision and mitigate numerical overflow issues:

- The response, $\mathbf{y}$, is centered to have mean zero and scaled to have a range equal to 1.
- Continuous covariates are centered to have mean zero and all covariates are scaled by their ℓ_2 norm such that $\mathbf{Z}^T\mathbf{Z}$ has a diagonal of ones.
- Exposure data is scaled to have standard deviation 1.

C.3.2 Modifier tree update

Marginalizing out fixed effect parameters

Consider the distribution of the data $\mathbf{y}_i \sim \mathcal{N}[f(\mathbf{x_i}, \mathbf{m}_i) + \mathbf{z}^T\boldsymbol{\gamma}, \sigma^2]$, where $f(\mathbf{x_i}, \mathbf{m}_i) = \boldsymbol{\theta}(\mathbf{m}_i)$ is the heterogeneous distributed lag effects for individual i. The posterior distribution of $\boldsymbol{\gamma}$ is

$$p(\boldsymbol{\gamma}|\mathbf{y}, \mathbf{f}, \sigma^2) \propto \sigma^{-p/2}|\mathbf{V}_{\boldsymbol{\gamma}}|^{-1/2} \exp\left\{-\sigma^{-2}(\mathbf{y}-\mathbf{f})^T\mathbf{Z}^T\mathbf{V}_{\boldsymbol{\gamma}}^{-1}\mathbf{Z}(\mathbf{y}-\mathbf{f})\right\}. \tag{C.6}$$

Here, $\mathbf{y} = [y_1, \ldots, y_n]^T$ is a vector of our continuous response; $\mathbf{f} = [f(\mathbf{x}_1, \mathbf{m}_1), \ldots, f(\mathbf{x}_n, \mathbf{m}_n)]'$ where $\mathbf{Z}$ is a matrix of covariates such that row i equals $\mathbf{z}_i^T$. In addition,

$$\mathbf{V}_{\boldsymbol{\gamma}} = (\mathbf{Z}^T\mathbf{Z} + \mathbf{I}/c)^{-1}, \tag{C.7}$$

where c is a fixed at a large value indicating a non-informative prior on $\boldsymbol{\gamma}$.

In the treed HDLMs, the heterogeneous distributed lag function $f(\mathbf{x}_i, \mathbf{m}_i)$ is estimated by the sum of partial distributed lag functions. In order to account for the effect of covariates $\mathbf{z}_i$ when estimating the trees, we integrate over the parameters $\boldsymbol{\gamma}$. This results in the marginal distribution for our data,

$$\mathbf{y}|\mathbf{f}, \sigma^2 \sim \mathcal{MVN}(\mathbf{f}, \sigma^2\mathbf{V}_{\mathbf{Z}}), \tag{C.8}$$

where

$$\mathbf{V}_{\mathbf{Z}} = (\mathbf{I} - \mathbf{Z}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z}^T)^{-1}. \tag{C.9}$$

Bayesian backfitting

The update of each modifier tree, $a = 1, \ldots, A$, proceeds using Bayesian backfitting (Hastie and Tibshirani, 2000). First, we calculate $\mathbf{R}_a$, the partial residuals after removing the effects of all other trees. We define $\mathbf{R}_a$ as

$$\mathbf{R}_a = \mathbf{y} - \sum_{\substack{a'=1 \\ a' \neq a}}^{A} g(\mathbf{X}, \mathcal{T}_{a'}). \tag{C.10}$$

Here, $g(\mathbf{X}, \mathcal{T}_a)$ is the vector of partial distributed lag estimates due to modifier tree $\mathcal{T}_a$ and corresponding distributed lag effects for each modifier tree terminal node. In the case of nested tree HDLM, $\mathcal{T}_a$ encompasses the nested treed DLMs: $\mathcal{D}_{a1}, \ldots, \mathcal{D}_{aB_a}$. In the case of shared tree HDLM, $\mathcal{T}_a$ includes the shared treed DLM $\mathcal{D}_a$.

Let $\mathbb{X}_a = [\mathbf{X}_{a1}, \ldots, \mathbf{X}_{aB_a}]$ denote a block-style exposure data matrix for modifier tree a. Each $\mathbf{X}_{ab}$ is a $n \times T$ matrix that corresponds to the exposure data for modifier terminal node η_{ab}. The non-zero rows of $\mathbf{X}_{ab}$ are exposure observations for the subgroup defined by the rules leading to η_{ab}; other rows are equal to zero.

The distribution of the partial distributed lag effects is

$$\mathbf{R}_a|- \sim \mathcal{MVN}\left(\mathbb{X}_a\Theta_a, \sigma^2\mathbf{V}_\mathbf{Z}\right) \tag{C.11}$$

where $\Theta_a = [\boldsymbol{\theta}'_{a1}, \ldots, \boldsymbol{\theta}'_{aB_a}]'$ represents a vector of distributed lag parameters parameters corresponding to the subgroups of modifier tree $\mathcal{T}_a$. In the case of nested and shared tree HDLMs, sets of the $\boldsymbol{\theta}_{ab}$ parameters are equal based on the piecewise constant structure of the treed DLM (Mork and Wilson, 2021a).

Tree prosposal

We update each $\mathcal{T}_a$, using a Metropolis-Hastings algorithm. We consider a proposal distribution with transition steps as follows:

- Grow: Randomly select a terminal node, η, to grow. Randomly select a splitting rule according to $p_{\text{rule}}(\rho|\eta)$ and create two new terminal nodes using the new splitting rule along with all previous rules.

- Prune: Randomly select an internal node with exactly two terminal nodes descending from it and remove the splitting rule.

- Change: Randomly select any internal node, η, and define a new splitting rule according to $p_{\text{rule}}(\rho|\eta)$. Update the limits of all terminal nodes that branch from η.

- Swap: Randomly select two connected internal nodes and reverse the rule ordering. That is, the parent node splitting rule and the child node splitting rule trade places.

The grow and prune steps are counterparts to one another, while change and swap are their own counterparts that can reverse the Markov chain. The transition kernel $p(\mathcal{T}^*|\mathcal{T})$ is given by the probability of selecting a step, multiplied by the probabilities associated with that step. For our simulations and data analysis we draw a new proposal from the four options (grow, prime, change, swap). The probability of using a grow or prune proposal is 0.25, change is 0.4, and swap 0.1.

Nested tree proposal

In the nested tree HDLM, a grow and prune proposal also requires new nested treed DLMs. In particular, a modifier tree grow proposal creates two new terminal nodes, requiring a new treed DLM at each node. A prune proposal removes two previous terminal nodes and requires a new treed DLM at the resulting pruned node.

A nested treed DLM is drawn from the tree prior using a stochastic growing process. At each terminal node, λ_c of the treed DLM, a split occurs with probability $p_{\text{split}}(\lambda_c) = \alpha(1 + d_\lambda)^{-\beta}$ (see Section C.2). If a split occurs, a time-point split in the DLM is selected with uniform probability over the remaining time points. This process repeats until no splits occur or there are no remaining time points. The hyperparameters α and β encourage small trees, which is necessary to retain constraints on the distributed lag effects.

Accepting a modifier tree proposal

After a new modifier tree proposal is made, we accept it by a standard Metropolis-Hastings ratio. To eliminate the need for complicated procedures due to the change in parameter dimension and to make the trees invariant to the covariates and variance, we integrate over Θ_a as well as σ^2. In BART, integrating out the vector Θ_a can be done one parameter at a time, as each observation is restricted to a single terminal node. However, in the treed HDLMs, the resulting covariance after integrating out fixed effect parameters $\boldsymbol{\gamma}$, requires us to simultaneously integrate over all distributed

lag parameters for modifier tree $\mathcal{T}_a$. The marginal likelihood of $\mathbf{R}_a$ is calculated to be

$$\begin{aligned} p(\mathbf{R}_a|\mathcal{T}_{a_j}-) &= \int_{\sigma^2}\int_{\Theta_a} p(\mathbf{R}_a|\Theta_a,\mathcal{T}_a,-)p(\Theta_a|\mathcal{T}_a,-)p(\sigma^2)\; d\Theta_a\; d\sigma^2 \\ &= \left(\nu^2\tau_a^2\right)^{-P_a/2} \\ &\quad \times |\mathbf{V}_{\Theta_a}|^{1/2}\left[\frac{\mathbf{R}_a^T\left(\mathbf{I}_n - \mathbf{Z}^T\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z} - \mathbf{V}_{\mathbf{Z}}^{-1}\mathbb{X}_a\mathbf{V}_{\Theta_a}\mathbb{X}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\right)\mathbf{R}_a}{2} + \frac{1}{\xi_{\sigma^2}}\right]^{-(n+1)/2} \end{aligned} \tag{C.12}$$

where ξ_{σ^2} is from the hierarchy $\sigma^2|\xi_{\sigma^2} \sim \mathcal{IG}(1/2, 1/\xi_{\sigma^2})$ and $\xi_{\sigma^2} \sim \mathcal{IG}(1/2, 1)$, which results in $\sigma \sim \mathcal{C}^+(0,1)$; and P_a is the number of unique parameters in Θ_a (for the Gaussian process HDLM, $P_a = TB_a$, for the shared tree HDLM, $P_a = C_aB_a$, and for the nested tree HDLM, $P_a = \sum_{b=1}^{B_a} C_{ab}$). Finally,

$$\mathbf{V}_{\Theta_a} = \left(\mathbb{X}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\mathbb{X}_a + \nu^{-2}\tau_a^{-2}\mathbf{U}_a\right)^{-1}, \tag{C.13}$$

where $\mathbf{U}_a$ is a diagonal block matrix of $\Sigma(\phi)$ for the Gaussian process HDLM, or an identity matrix for the shared and nested HDLMs. We note that use of the Woodbury matrix identity gives a more efficient method of calculating $\mathbf{V}_{\Theta_a}$.

After integrating over tree-specific parameters, we calculate $p(\mathbf{R}_a, \mathcal{T}_a|-) = p(\mathbf{R}_a|\mathcal{T}_a,-)p(\mathcal{T}_a)$ and accept $\mathcal{T}_a^*$ according to the Metropolis-Hastings ratio given by

$$r = \min\left\{1, \frac{p(\mathcal{T}_a^*)p(\mathbf{R}_a|\mathcal{T}_a^*,-)p(\mathcal{T}_a|\mathcal{T}_a^*)}{p(\mathcal{T}_a)p(\mathbf{R}_a|\mathcal{T}_a,-)p(\mathcal{T}_a^*|\mathcal{T}_a)}\right\}. \tag{C.14}$$

C.3.3 Treed DLM update

For the shared tree HDLM, a single treed DLM, $\mathcal{D}_a$, is updated for every modifier tree, $\mathcal{T}_a$. In the nested tree HDLM, a nested treed DLM, $\mathcal{D}_{ab}$ is updated for each modifier tree terminal node η_{ab}. Proposals for the treed DLMs occur through grow, prune, and change steps, and use a uniform prior over the remaining time splitting points. See Mork and Wilson (2021a) for more details.

After a new treed DLM has been proposed, it is accepted with the independent Metropolis-Hastings algorithm. This process is identical to that described for modifier trees in Section C.3.2.

C.3.4 Full conditionals

After each modifier tree update (and treed DLM updates for shared and nested HDLMs), we draw Θ_a (all parameters specific to modifier tree a) from the full conditional distribution,

$$\Theta_a|- \sim \mathcal{MVN}_{P_a}\left(\mathbf{V}_{\Theta_a}\mathbb{X}_a^T\mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{R}_a, \sigma^2\mathbf{V}_{\Theta_a}\right). \tag{C.15}$$

The update of trees and corresponding partial HDLM is followed by a draw from the full conditional of remaining parameters and hyperparameters:

$$\xi_\sigma|- \sim \mathcal{IG}\left(1, 1+\frac{1}{\sigma^2}\right); \tag{C.16}$$

$$\sigma^2|- \sim \mathcal{IG}\left[\frac{n+P+1}{2}, \frac{\|\mathbf{V}_{\mathbf{Z}}^{-1/2}(\mathbf{y}-\mathbf{f})\|_2^2}{2}+\frac{D}{2\nu^2}+\frac{1}{\xi_\sigma}\right]; \tag{C.17}$$

where $P=\sum_{a=1}^{A}P_a$ and $D=\sum_{a=1}^{A}\Theta_a^T\mathbf{U}_a\Theta_a/\tau_a^2$. Also,

$$\xi_\nu|- \sim \mathcal{IG}\left(1, 1+\frac{1}{\nu^2}\right); \tag{C.18}$$

$$\nu^2|- \sim \mathcal{IG}\left(\frac{P+1}{2}, \frac{D}{2\sigma^2}+\frac{1}{\xi_\nu}\right); \tag{C.19}$$

$$\xi_{\tau_a}|- \sim \mathcal{IG}\left(1, 1+\frac{1}{\tau_a^2}\right); \tag{C.20}$$

$$\tau_a^2|- \sim \mathcal{IG}\left(\frac{P_a+1}{2}, \frac{\Theta_a^T U_a\Theta_a}{2\sigma^2\nu^2}+\frac{1}{\xi_{\tau_a}}\right). \tag{C.21}$$

Updates of modifier selection probabilities $\boldsymbol{\psi}$ come from full conditional

$$\boldsymbol{\psi}|- \sim \text{Dirichlet}\left(\kappa/J+N_{\{m_j=1\}}, \ldots, \kappa/J+N_{\{m_j=M\}}\right) \tag{C.22}$$

where $N_{\{m_j=m\}}$ is the number of splitting rules using modifier m. The hyperprior κ is updated by the Metropolis-Hastings algorithm.

Finally we draw $\boldsymbol{\gamma}$ from its full conditional,

$$\boldsymbol{\gamma}|\mathbf{y}, \mathbf{f}, \sigma^2 \sim \mathcal{MVN}\left[\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{Z}(\mathbf{y}-\mathbf{f}), \sigma^2\mathbf{V}_{\boldsymbol{\gamma}}\right]. \tag{C.23}$$

C.3.5 Subgroup posterior analysis

The HDLM framework allows for inference on personalized distributed lag functions for a specific level of modifiers or subgroup specific distributed lag estimates averaged over the levels of modifiers within a particular subgroup. Let S be a set of observations based on a subgroup of interest. For example, S may the the set of all babies born to Hispanic mothers or all boy babies with obese mothers. Denote $S_{\eta_{ab}} \subset S$ as the observations from S contained in modifier tree terminal node η_{ab}. Then, we define weights based on the proportion of S in each terminal node, $w(S, \eta_{ab}) = |S_{\eta_{ab}}|/|S|$. The DLM for subgroup S is calculated

$$\boldsymbol{\theta}_S = \sum_{a=1}^{A}\sum_{b=1}^{B_a}\boldsymbol{\theta}_{ab}w(S, \eta_{ab}) \tag{C.24}$$

where $\boldsymbol{\theta}_{ab}$ is the vector of DLM effects for modifier tree terminal node η_{ab}.

C.4 Additional Simulation Results

C.4.1 Scenario 1: Early/Late Window

Table C.1 presents modifier posterior inclusion probabilities for Scenario 1.

Table C.2 presents modifier interaction posterior inclusion probabilities (PIP). An modifier interaction occurs when two modifiers are used in consecutive splitting rules in the same tree. For simulation scenario 1, we would expect modifiers z_1 and z_2 to be used in consecutive splitting rules due to how the distributed lag effects are defined. We see in the lowest error setting the modifier interaction PIP for $z_{1/2}$ (interaction between modifiers z_1 and z_2) is 1 for all HDLMs. In

Table C.1: Simulation results for modifier inclusion in scenario 1 (early/late effect). Results describe the average posterior inclusion probability across simulation replicates for each potential modifier, (z_i) or group of modifiers ($z_{[i-j]}$) in any tree of the model. A value of one indicates the modifier was always present in at least one tree of the model.

Model	z_1^*	z_2^*	z_3	$z_{[4-8]}$	$z_{[9-13]}$
$\sigma^2 = 10$					
Nested Tree HDLM	1.00	1.00	0.60	0.61	0.59
Shared Tree HDLM	1.00	1.00	0.60	0.61	0.58
Gaussian Process HDLM	1.00	1.00	0.62	0.62	0.60
$\sigma^2 = 25$					
Nested Tree HDLM	1.00	0.97	0.62	0.62	0.60
Shared Tree HDLM	1.00	0.97	0.62	0.62	0.60
Gaussian Process HDLM	1.00	0.99	0.62	0.62	0.60
$\sigma^2 = 50$					
Nested Tree HDLM	0.99	0.80	0.63	0.63	0.61
Shared Tree HDLM	0.99	0.79	0.63	0.63	0.61
Gaussian Process HDLM	0.99	0.85	0.63	0.63	0.61

* active modifiers in scenario 1

the medium error scenario this decreases slightly to a PIP of 0.88 for the nested tree HDLM; for the highest error scenario this is 0.46.

Table C.2: Simulation results for modifier interaction posterior inclusion probabilities (PIP) in scenario 1 (early/late effect). Interaction PIP indicates the probability a pair of modifiers was used for two consecutive splitting rules in the same tree at least once in the ensemble. We report the average across simulation replicates of the correct interaction PIP, $z_{1/2}$, the average PIP of other interactions, $z_{i/j}$ and the maximum PIP of other interactions, $\max z_{i/j}$.

Model	$z_{1/2}$	$z_{i/j}$	$\max z_{i/j}$
$\sigma^2 = 10$			
Nested Tree HDLM	1.00	0.11	0.41
Shared Tree HDLM	1.00	0.11	0.44
Gaussian Process HDLM	1.00	0.12	0.41
$\sigma^2 = 25$			
Nested Tree HDLM	0.88	0.12	0.40
Shared Tree HDLM	0.89	0.12	0.42
Gaussian Process HDLM	0.93	0.12	0.41
$\sigma^2 = 50$			
Nested Tree HDLM	0.46	0.12	0.41
Shared Tree HDLM	0.43	0.12	0.42
Gaussian Process HDLM	0.54	0.12	0.40

We also tracked the average PIP of other modifier interactions as well as the maximum PIP of other modifier interactions. The average was around 0.12 while the maximum PIP of modifier interactions ranged from 0.40 to 0.44. These results give a relative measure for modifier interaction importance in our data analysis. That is, modifier interactions larger than 0.44 may represent meaningful differences in the distributed lag effects.

C.4.2 Scenario 2: Scaled Effect

Table C.3 presents posterior inclusion probabilities for an individual modifier.

Table C.3: Simulation results for modifier inclusion in scenario 2 (scaled effect). Results describe the average posterior inclusion probability across simulation replicates for each potential modifier (or group of modifiers) in any tree of the model. A value of one indicates the modifier was always present in at least one tree of the model.

Model	z_1^*	z_2	z_3^*	$z_{[4-8]}$	$z_{[9-13]}$
$\sigma^2 = 10$					
Nested Tree HDLM	1.00	0.56	1.00	0.59	0.57
Shared Tree HDLM	1.00	0.57	1.00	0.60	0.58
Gaussian Process HDLM	1.00	0.55	1.00	0.59	0.56
$\sigma^2 = 25$					
Nested Tree HDLM	1.00	0.59	0.98	0.61	0.59
Shared Tree HDLM	1.00	0.59	0.98	0.62	0.59
Gaussian Process HDLM	1.00	0.58	0.98	0.61	0.59
$\sigma^2 = 50$					
Nested Tree HDLM	1.00	0.60	0.88	0.62	0.61
Shared Tree HDLM	0.99	0.60	0.88	0.62	0.60
Gaussian Process HDLM	0.99	0.60	0.87	0.63	0.61

* active modifiers in scenario 2

Table C.4 presents modifier interaction PIPs for simulation scenario 2. Here, we show the PIP of the active interactions, $z_{1/3}$ and $z_{3/3}$. Because scenario 2 had a continuous effect, it is reasonable that the tree may split on the modifier z_3 multiple times. However, other trees in the ensemble that split on z_1 may split on z_3 at a different location to build the scaled distributed lag effect.

We find that the modifier interaction PIP for $z_{1/3}$ is 1 in the lowest error scenario, 0.91 in the middle error case, and 0.65 in the largest error setting. The $z_{3/3}$ modifier is larger than other

Table C.4: Simulation results for modifier interaction posterior inclusion probabilities (PIP) in scenario 2 (scaled effect). Interaction PIP indicates the probability a pair of modifiers was used for two consecutive splitting rules in the same tree at least once in the ensemble. We report the average across simulation replicates of the correct interaction PIPs, $z_{1/3}$ and $z_{3/3}$, the average PIP of other interactions, $z_{i/j}$ and the maximum PIP of other interactions, $\max z_{i/j}$.

Model	$z_{1/3}$	$z_{3/3}$	$z_{i/j}$	$\max z_{i/j}$
$\sigma^2 = 10$				
Nested Tree HDLM	1.00	0.46	0.11	0.41
Shared Tree HDLM	1.00	0.58	0.11	0.38
Gaussian Process HDLM	1.00	0.49	0.11	0.45
$\sigma^2 = 25$				
Nested Tree HDLM	0.91	0.34	0.11	0.35
Shared Tree HDLM	0.94	0.36	0.11	0.38
Gaussian Process HDLM	0.91	0.37	0.11	0.41
$\sigma^2 = 50$				
Nested Tree HDLM	0.65	0.26	0.12	0.38
Shared Tree HDLM	0.66	0.27	0.12	0.39
Gaussian Process HDLM	0.60	0.26	0.12	0.40

modifiers on average, but does not always exceed the maximum value of other modifiers. Here, we see average and maximum PIPs for other modifiers similar to results in scenario 1.

C.4.3 Scenario 3: No Effect Heterogeneity

Table C.5 presents modifier interaction PIPs for simulation scenario 3 (no effect heterogeneity). In this scenario, no modifier interaction plays a role in modeling the distributed lag effect. We find that the average PIP for modifier interactions is 0.12 and the maximum PIP ranges from 0.24 to 0.27 across all error settings. The fact that no one modifier interaction stands out bodes well for the model estimating distributed lag effects without heterogeneity.

C.5 Additional Data Analysis Results

C.5.1 Effect modification

Table C.6 compares modifier posterior inclusion probabilities among shared, nested, and Gaussian process HDLMs. We note that age, BMI, education, and Hispanic designation modifiers have the highest posterior inclusion probabilities across all three models.

Table C.5: Simulation results for modifier interaction posterior inclusion probabilities (PIP) in scenario 3 (no effect heterogeneity). Interaction PIP indicates the probability a pair of modifiers was used for two consecutive splitting rules in the same tree at least once in the ensemble. We report the average across simulation replicates of the average PIP of all interactions, $z_{i/j}$, and the maximum PIP of all interactions, $\max z_{i/j}$.

Model	$z_{i/j}$	$\max z_{i/j}$
$\sigma^2 = 10$		
Nested Tree HDLM	0.12	0.24
Shared Tree HDLM	0.12	0.26
Gaussian Process HDLM	0.12	0.24
$\sigma^2 = 25$		
Nested Tree HDLM	0.12	0.26
Shared Tree HDLM	0.12	0.25
Gaussian Process HDLM	0.12	0.27
$\sigma^2 = 50$		
Nested Tree HDLM	0.12	0.27
Shared Tree HDLM	0.12	0.27
Gaussian Process HDLM	0.12	0.27

Table C.6: Modifying covariate posterior inclusion probabilities for three HDLM methods.

Modifier	Shared Tree	Nested Tree	Gaussian Process
Age at conception	0.93	0.90	0.83
Body mass index	0.95	0.96	0.96
Income range	0.74	0.71	0.72
Highest education	0.90	0.88	0.86
Smoking habits	0.78	0.76	0.81
Marital status	0.50	0.46	0.45
Prenatal care	0.48	0.60	0.54
Race	0.61	0.48	0.54
Hispanic	0.95	0.98	0.90
Sex of child	0.64	0.43	0.53

C.5.2 Additional Figures

Figure C.3 shows the MSPE from 10-fold cross-validation of the three HDLMs relative to a treed DLM without effect modification. On average, the MSPE of the shared tree HDLM was smallest (0.99975) followed by nested tree HDLM (0.99979) and Gaussian process HDLM (0.99982). The signal of the exposure effect is very small relative to the signal from the fixed effects and residual error, leading to very small differences between the MSPE for these methods.

Figure C.4 shows subgroup-specific DLMs for Hispanic vs non-Hispanic.

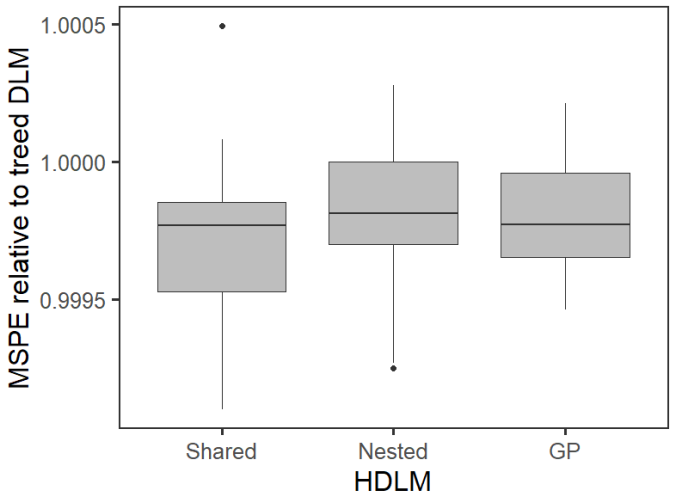


Figure C.3: MSPE relative to treed DLM without modification.

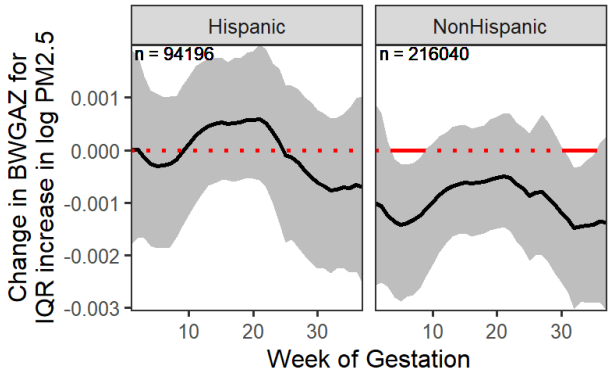


Figure C.4: Hispanic and non-Hispanic subgroup-specific DLMs.

Figures C.5 and C.6 show the Hispanic subgroup broken down by education and age, as well as education and BMI, respectively.

Figures C.7 and C.8 show the non-Hispanic subgroup broken down by education and age, as well as education and BMI, respectively.

The next largest modifier PIP was Smoking and the next largest modifier interaction PIP was between smoking and BMI. Figures C.9 and C.10 show subgroup-specific DLMs for Hispanic and non-Hispanic subgroups broken down by smoking (never vs. former or current) and BMI.

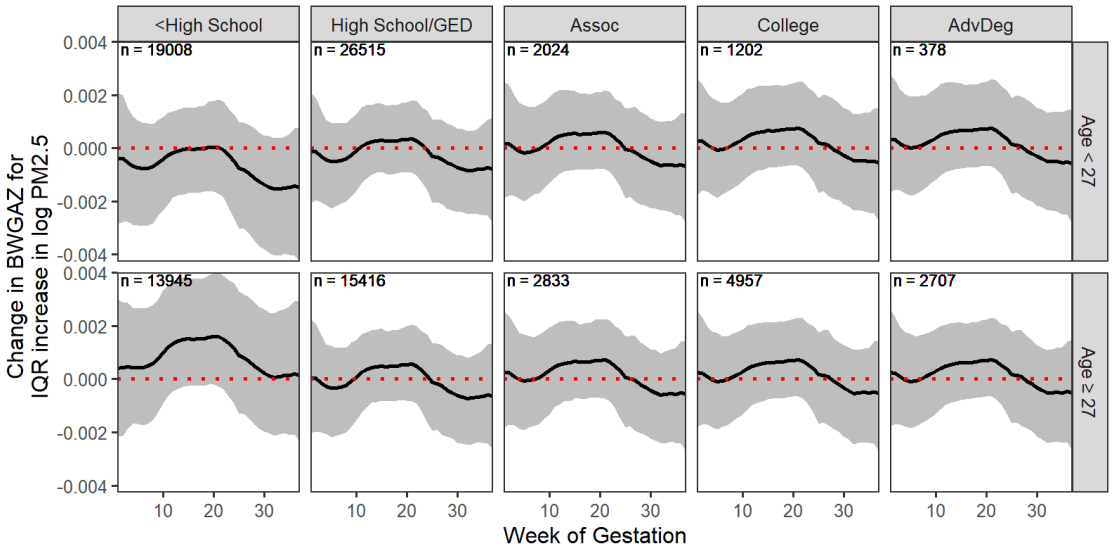


Figure C.5: Hispanic subgroup broken out by education and age.

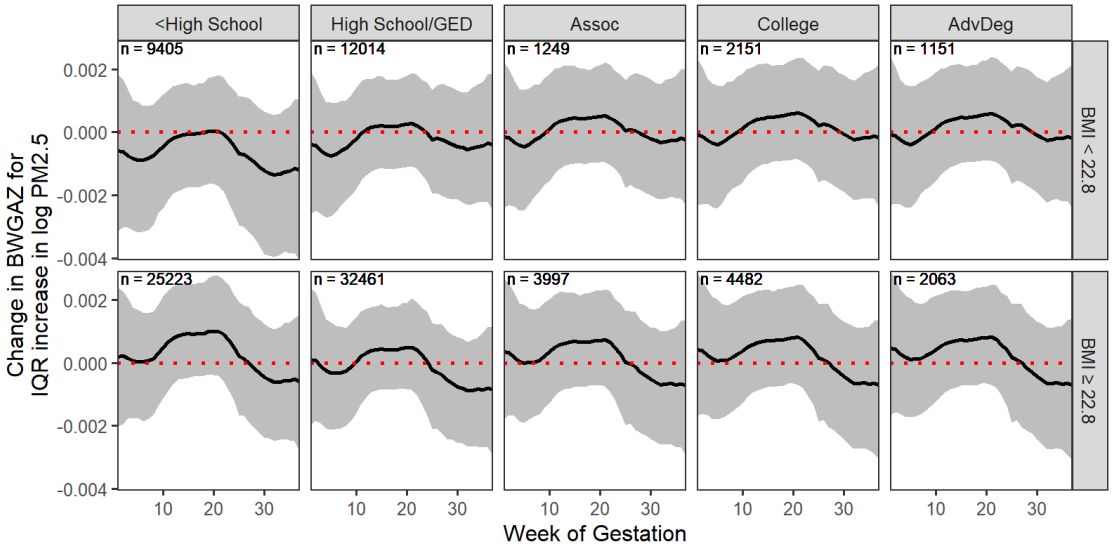


Figure C.6: Hispanic subgroup broken out by education and BMI.

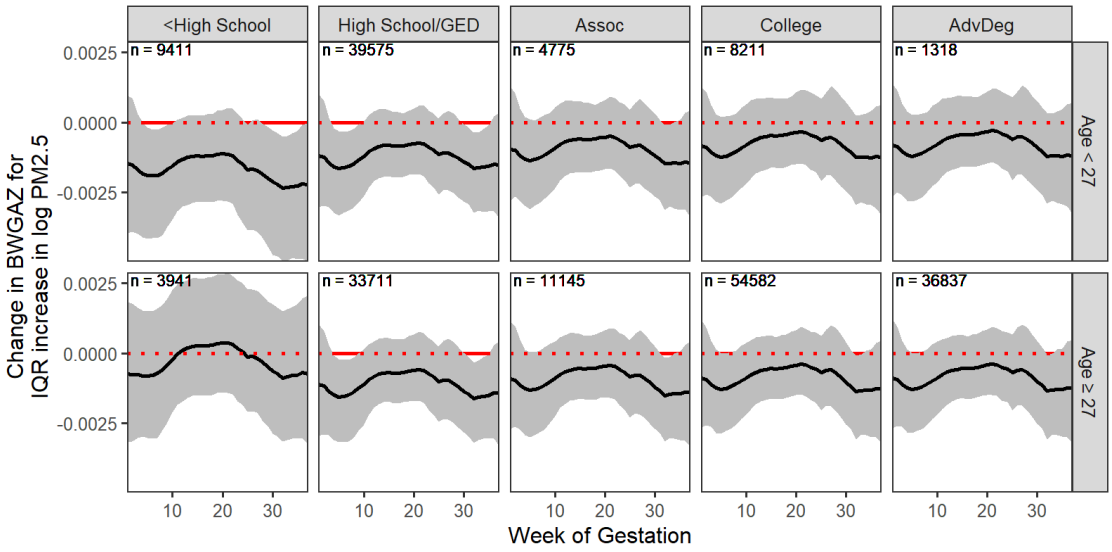


Figure C.7: Non-Hispanic subgroup broken out by education and age.

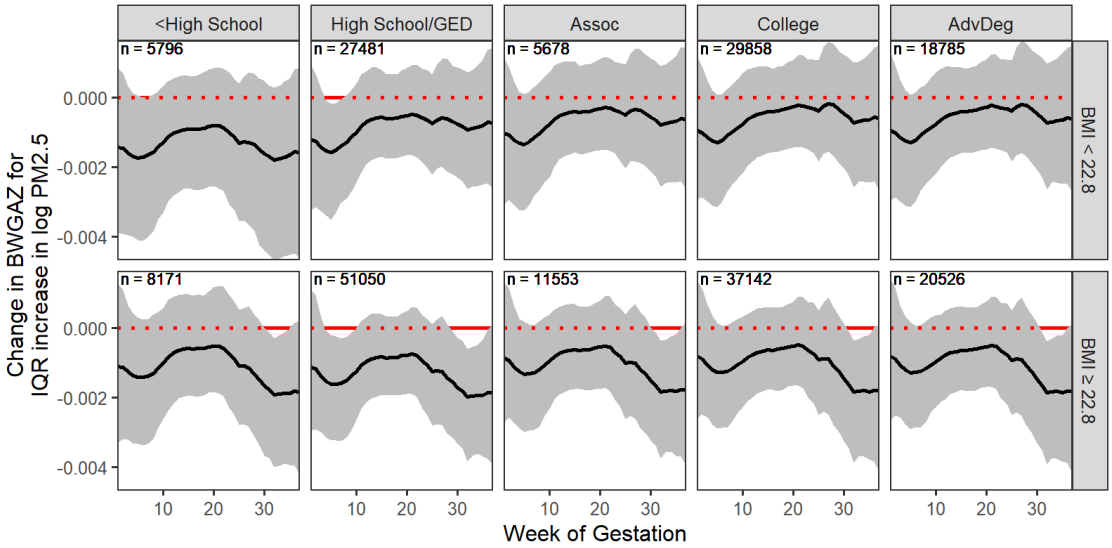


Figure C.8: Non-Hispanic subgroup broken out by education and BMI.

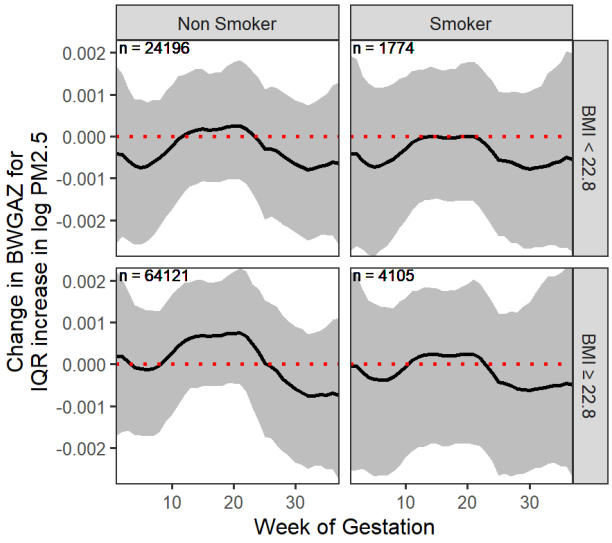


Figure C.9: Hispanic subgroup broken out by smoking (never vs former or current) and BMI.

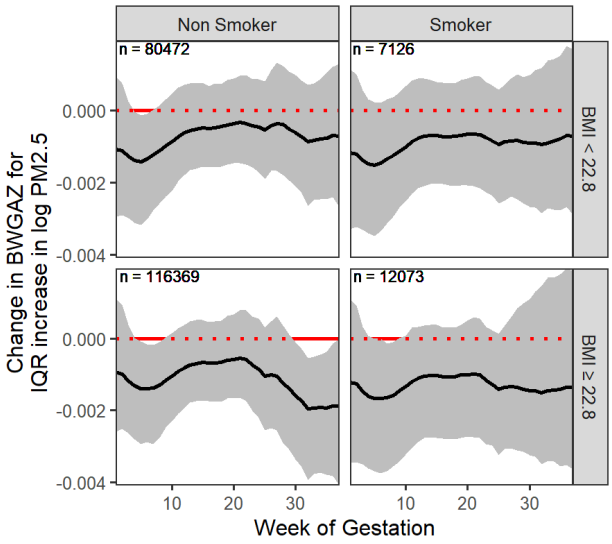


Figure C.10: Non-Hispanic subgroup broken out by smoking (never vs former or current) and BMI.

C.5.3 Nested Tree HDLM

We replicate the subgroup-specific analyses of the paper using the nested tree HDLM. Figure C.11 shows subgroup specific DLMs broken down by Hispanic and BMI modifiers. Figure C.12 shows subgroup specific DLMs broken down by Hispanic and education modifiers.

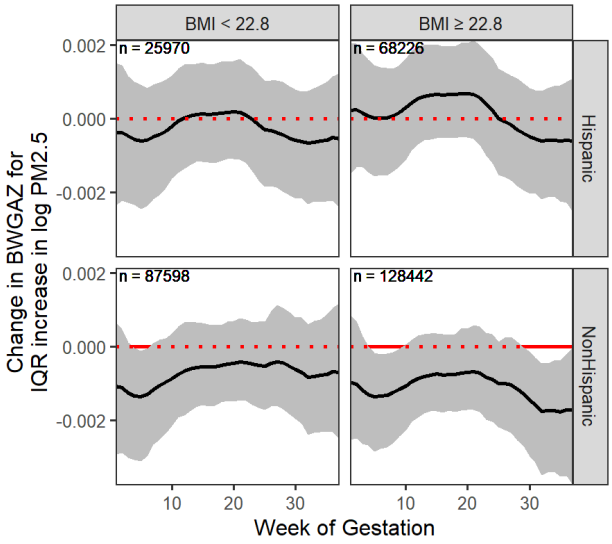


Figure C.11: Subgroup-specific DLMs broken down by Hispanic and BMI, using the nested tree HDLM.

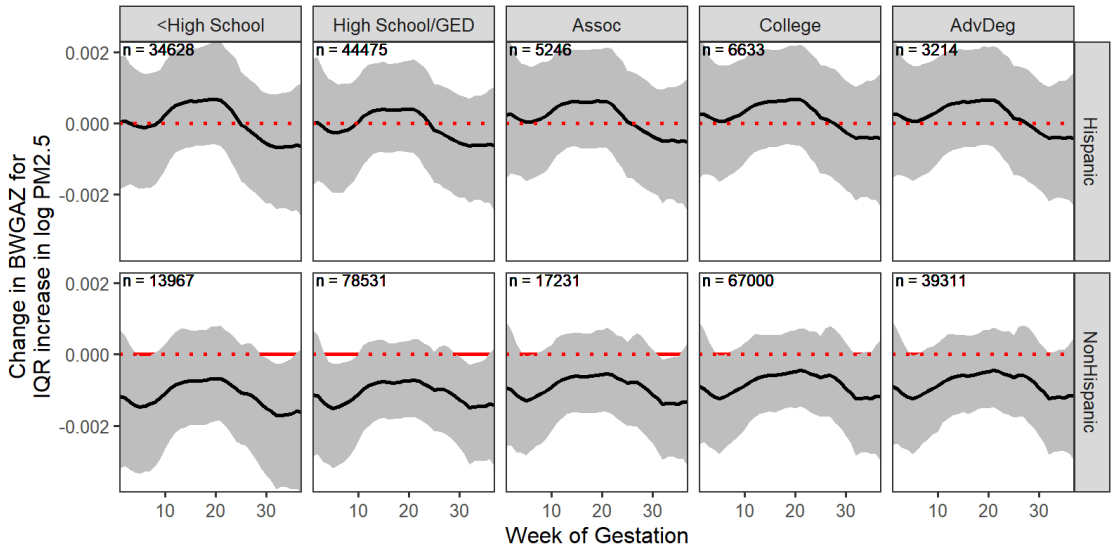


Figure C.12: Subgroup-specific DLMs broken down by Hispanic and education, using the nested tree HDLM.

C.5.4 Gaussian process HDLM

We replicate the subgroup-specific analyses of the paper using the Gaussian process HDLM. Figure C.13 shows subgroup specific DLMs broken down by Hispanic and BMI modifiers. Figure C.14 shows subgroup specific DLMs broken down by Hispanic and education modifiers. We note the increased variance of the distributed lag effect estimates as well as decreased smoothness in the DLM, which does not allow for critical window identification.

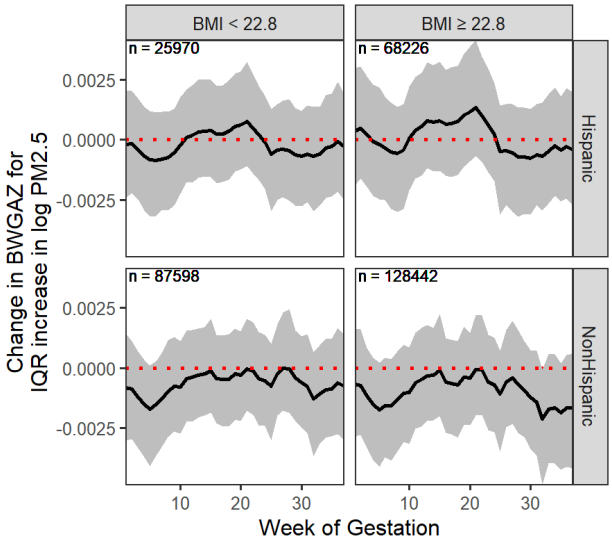


Figure C.13: Subgroup-specific DLMs broken down by Hispanic and BMI, using the Gaussian process HDLM.

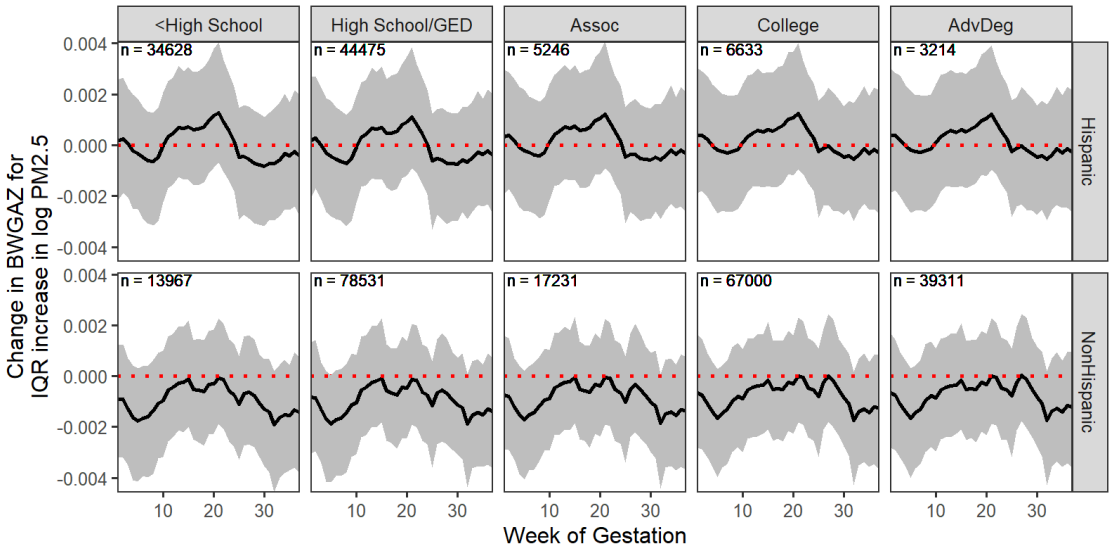


Figure C.14: Subgroup-specific DLMs broken down by Hispanic and education, using the Gaussian process HDLM.