Biostatistics (2022) **23**, 3, *pp.* 754–771
doi: 10.1093/biostatistics/kxaa051
Advance Access publication on February 2, 2021

Treed distributed lag nonlinear models

DANIEL MORK, ANDER WILSON*
Statistics Department, Colorado State University, 1877 Campus Delivery, Fort Collins, CO, USA 80523
Ander.Wilson@colostate.edu

Summary

In studies of maternal exposure to air pollution, a children's health outcome is regressed on exposures observed during pregnancy. The distributed lag nonlinear model (DLNM) is a statistical method commonly implemented to estimate an exposure–time–response function when it is postulated the exposure effect is nonlinear. Previous implementations of the DLNM estimate an exposure–time–response surface parameterized with a bivariate basis expansion. However, basis functions such as splines assume smoothness across the entire exposure–time–response surface, which may be unrealistic in settings where the exposure is associated with the outcome only in a specific time window. We propose a framework for estimating the DLNM based on Bayesian additive regression trees. Our method operates using a set of regression trees that each assume piecewise constant relationships across the exposure–time space. In a simulation, we show that our model outperforms spline-based models when the exposure–time surface is not smooth, while both methods perform similarly in settings where the true surface is smooth. Importantly, the proposed approach is lower variance and more precisely identifies critical windows during which exposure is associated with a future health outcome. We apply our method to estimate the association between maternal exposures to $PM_{2.5}$ and birth weight in a Colorado, USA birth cohort.

Keywords: Air pollution; Children's health; Critical windows; Distributed lag; Regression trees.

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 *and others*, 2012).

In the context of maternal exposure to air pollution, which we consider in this article, 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 *and others*, 2012, 2015) adiposity (Chiu *and others*, 2017), asthma and wheeze (Bose *and others*, 2018; Lee *and others*, 2018), neurodevelopment (Chiu *and others*, 2016), among other outcomes (Stieb *and others*, 2012; Šrám *and others*, 2005). This

*To whom correspondence should be addressed.

© The Author 2021. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

includes studies that have found that the linear (Chiu *and others*, 2017; Chang *and others*, 2015) and nonlinear (Wu *and others*, 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 *and others*, 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 *and others*, 2000), Gaussian processes (Warren *and others*, 2012), and principal components or splines (Wilson *and others*, 2017*a*). Wilson *and others* (2017*b*) 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 nonlinear associations in the exposure–response function at any given time, a class of distributed lag nonlinear models (DLNMs) has been proposed (Gasparrini *and others*, 2010, 2017). DLNM methods typically operate by cross-basis smoothing with splines or penalized spline regression. This results in a unique nonlinear 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 *and others* (2020) 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 nonlinear exposure–response function.

In this article, 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 nonlinear models (TDLNM), is based on the Bayesian additive regression trees (BART) framework developed by Chipman *and others* (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 and 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. DATA

We analyze birth records from CO, 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 6,000 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 article 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. These data are 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 *and others* (2009). 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 1 of supplementary material available at *Biostatistics* online.

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

3. METHODS

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 ith 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

$$y_i \sim \mathcal{N}\left[f(\mathbf{x}_i) + \mathbf{z}_i^T \boldsymbol{\gamma}, \sigma^2\right], \tag{3.1}$$

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 nonlinear forms. The DLNM allows a unique nonlinear association between exposures at each time point and the outcome. In general, the distributed lag function is defined as

$$f(\mathbf{x}_i) = f(x_{i1}, \ldots, x_{iT}) = \sum_{t=1}^{T} w(x_{it}, t) \tag{3.2}$$

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 *and others*, 2010, 2017) utilize a cross-basis where w is represented as a bivariate basis expansion in the exposure concentration and time dimensions. Penalized

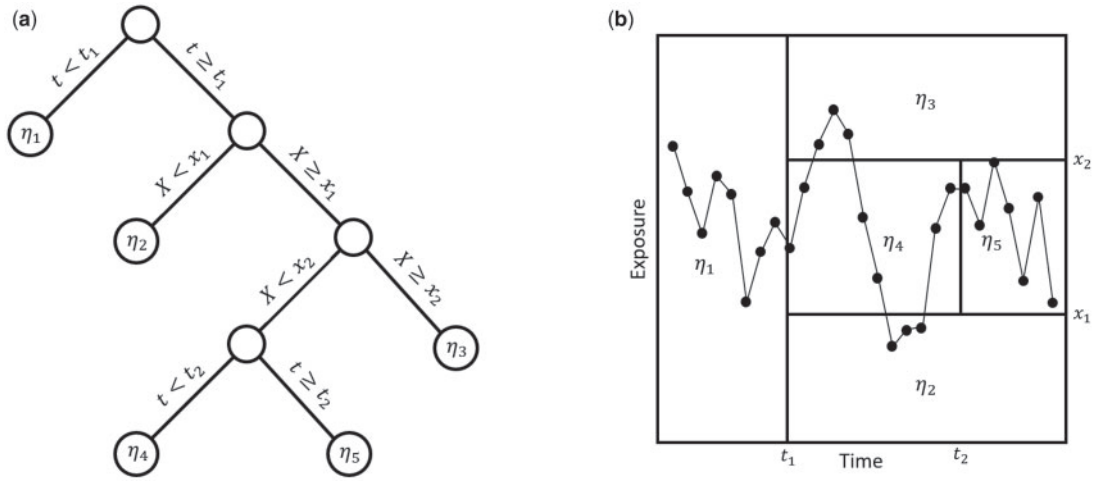

Fig. 1. Example of tree, $\mathcal{T}$, with terminal nodes η_b, $b \in \{1, 2, 3, 4, 5\}$. Panel (a) diagrams the tree with dichotomous splits on time or exposure concentration, whereas panel (b) represents the resulting partition of the exposure-time space for a single observation.

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.

3.2. *Treed DLNM approach*

We introduce a sum-of-trees model based on the BART framework (Chipman *and others*, 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 1 illustrates the approach for a single tree. Figure 1a 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 1b 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,

$$w(x_{it}, t) = \mu_b \quad \text{if } x_{it} \in \eta_b. \tag{3.3}$$

The distributed lag function for tree $\mathcal{T}$ takes a form similar to that in (3.2) and is defined as

$$g(\mathbf{x}_i, \mathcal{T}) = \sum_{t=1}^{T} w(x_{it}, t). \tag{3.4}$$

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 (3.3), 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 nonlinear function f. Formally, the exposure–time–response function for TDLNM is

$$f(\mathbf{x}_i) = \sum_{a=1}^{A} g(\mathbf{x}_i, \mathcal{T}_a), \tag{3.5}$$

where $g(\mathbf{x}_i, \mathcal{T}_a)$ represents the partial estimate contributed by tree a given in (3.4).

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.

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 (3.3) to be

$$w_a(x_{it}, t) = \sum_{b=1}^{B_a} \mu_{ab} \cdot \psi(x_{it}; \eta_{ab}, \sigma_x). \tag{3.6}$$

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

$$\psi(x_{it}; \eta_{ab}, \sigma_x) = \left\{ \Phi\left(\frac{\lceil x_{ab} \rceil - x_{it}}{\sigma_x}\right) - \Phi\left(\frac{\lfloor x_{ab} \rfloor - x_{it}}{\sigma_x}\right) \right\} \cdot \mathbb{I}(t \in \eta_{ab}), \tag{3.7}$$

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.

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 (3.7) 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 Φ in (3.7) with the empirical cumulative distribution function.

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 (3.4) and (3.5), i.e. the sum over trees and the sum of each tree over time. Reversing the order of the summation we get

$$f(\mathbf{x}_i) = \sum_{t=1}^{T} \sum_{a=1}^{A} w_a(x_{it}, t). \tag{3.8}$$

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.

4. Prior Specification and Computation

Our prior specification is based on that of Chipman *and others* (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 *and others* (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 Section B of supplementary material available at *Biostatistics* online.

4.1. *Prior specification*

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

$$\mu_{ab}|\sigma^2, \omega^2, \tau_a^2 \sim \mathcal{N}\left(0, \sigma^2\omega^2\tau_a^2\right). \tag{4.9}$$

Here, $\tau_a \sim \mathcal{C}^+(0,1)$ and $\omega \sim \mathcal{C}^+(0,1)$ defines 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 *and others* (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 one minus the number of unique exposure values 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.

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 *and others* (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 Section B of supplementary material available at *Biostatistics* online.

4.3. *Hyperprior selection and tuning*

Tree splitting hyperpriors were set to the defaults used in Chipman *and others* (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 10–20 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 nonlinearity in the exposure-concentration effects.

4.4. *Estimating the exposure–time–response function*

The distributed lag nonlinear 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 trees, we center each tree at the reference value and we use the centered trees for posterior inference.

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 (3.1) and (3.2), 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 (i) piecewise constant effect in exposure across weeks $11-15$; (ii) linear effect in exposure across weeks $11-15$; (iii) smooth, nonlinear effect (logistic shape) in exposure across weeks $11-15$; (iv) smooth, nonlinear effect (logistic shape) in exposure with a smooth effect in time peaking at week 13 and extending approximately 5 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 and 3. Algebraic details of the DLNM surface for each scenario and a graphic representation can be found in Section C.1 of supplementary material available at *Biostatistics* online.

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 5th 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`.

5.1. *Simulation estimators and comparisons*

TDLNM and TDLNMse used the prior settings described in section 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 10th 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 *and others* (2017).

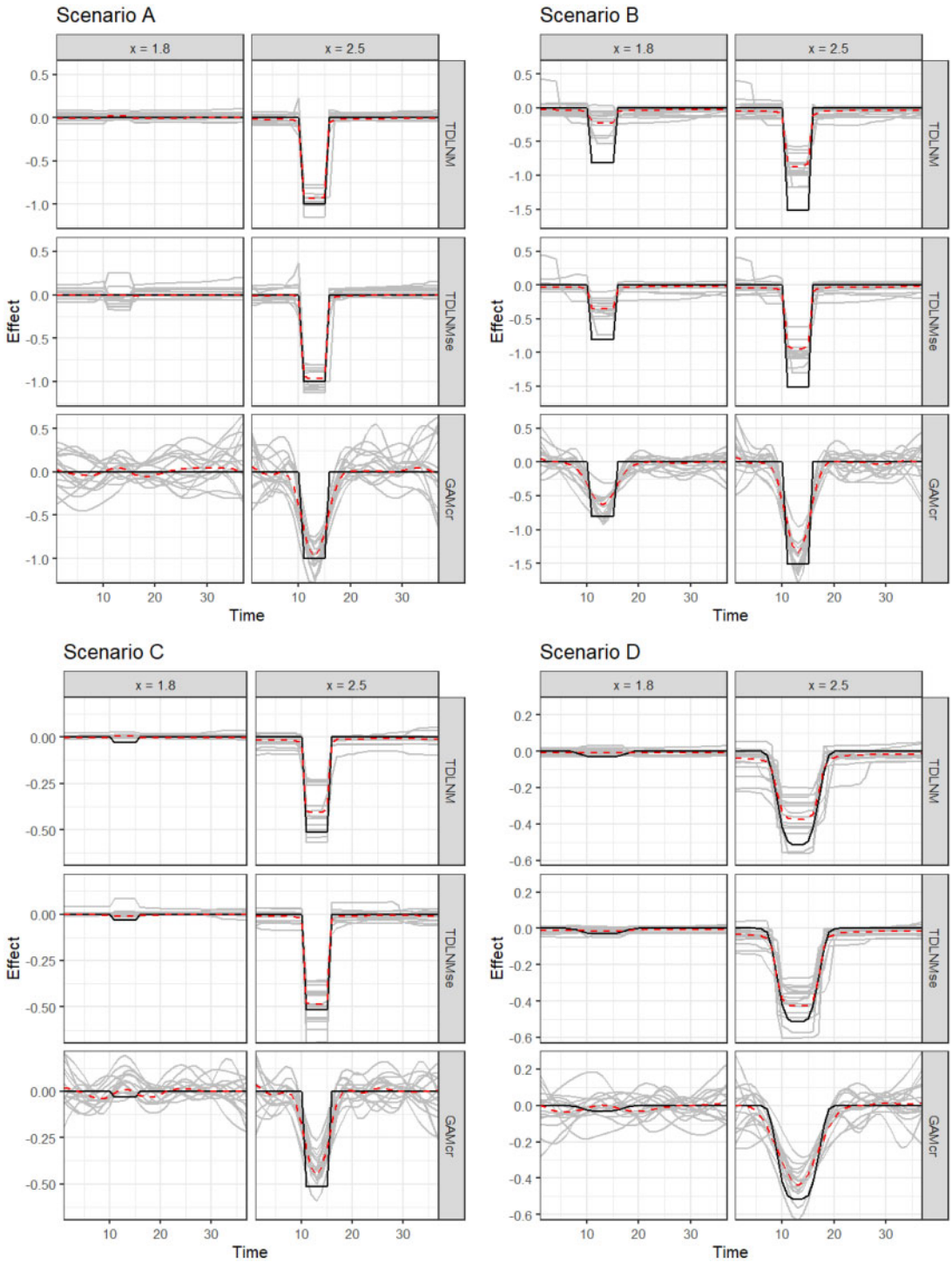

Fig. 2. Simulation results of TDLNM, TDLNMse, and GAMcr comparing the exposure-time-response (y-axis) cross-section across all times (x-axis) fixed at two different exposure concentrations. Grey lines show 15 random simulation replicates, red dashed line indicates average across all simulations, and solid black lines indicates the true simulated response.

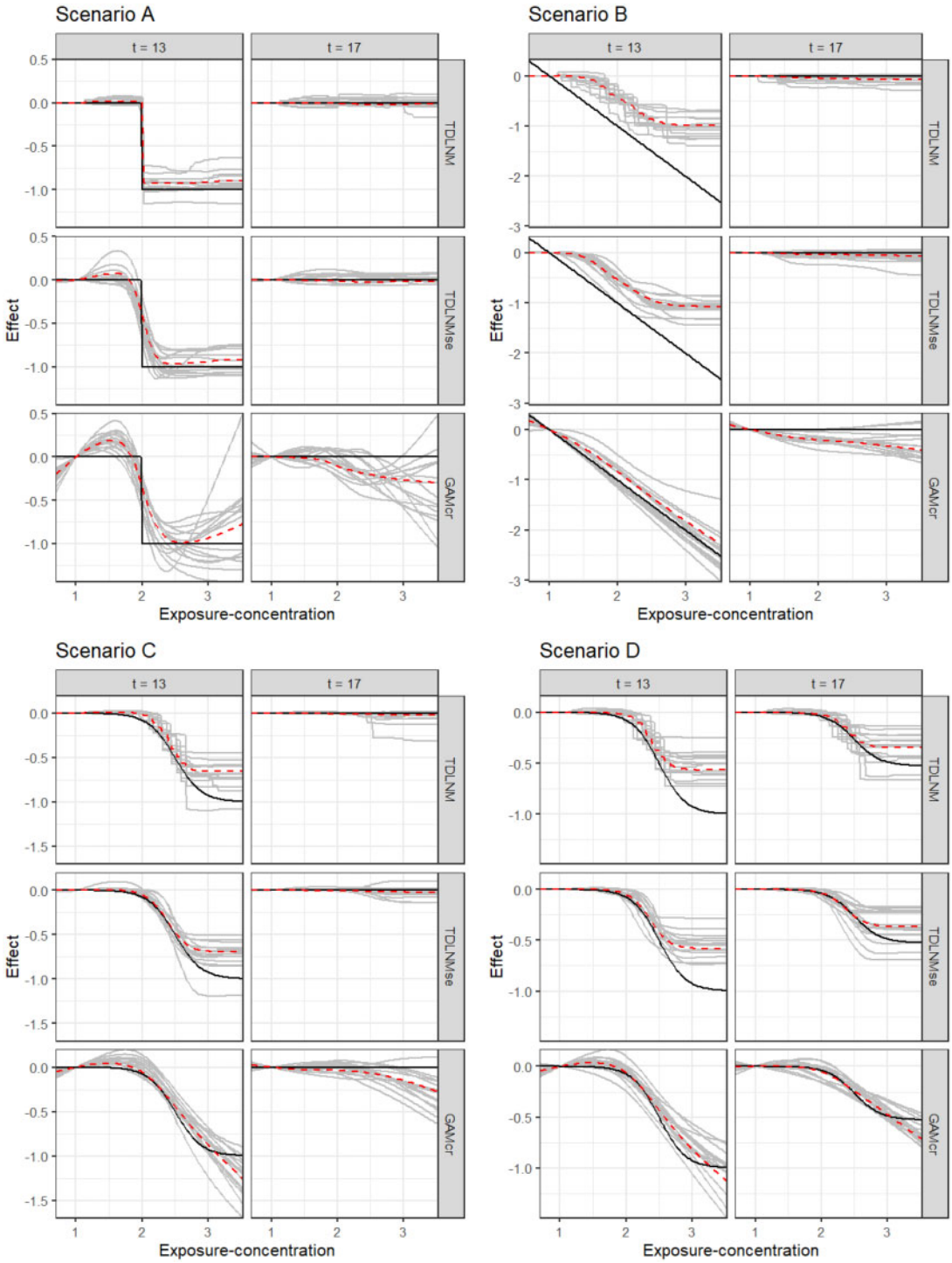

Fig. 3. Simulation results of TDLNM, TDLNMse, and GAMcr comparing the exposure-time-response (y-axis) cross-section across all exposure concentrations (x-axis) fixed at two different times. Grey lines show 15 random simulation replicates, red dashed line indicates average across all simulations, and solid black lines indicates the true simulated response.

- 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 1–10) 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.

5.2. *Simulation results*

Summary measures of model performance are shown in Table 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 nonzero effect across grid points inside the simulated critical window (TP), the probability of incorrectly placing a nonzero 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 nonzero 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 nonzero effect everywhere between weeks 11–15 and scenario D, which has a nonzero effect everywhere. Figures 2 and 3 show cross-sections of the exposure–time–response surface using estimates from models TDLNM, TDLNMse, and GAMcr. A nonzero 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 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 nonzero 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 nonzero effect is higher for TDLNM and TDLNMse, as the spline based models do a better job interpolating into the extreme exposure values where few data points reside. Figure 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

Table 1. *Simulation results, showing RMSE for estimation of the exposure-time-surface with no-effect and effect separated*

	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

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 Supplemental Materials Table 2.

notable at the boundaries in time or extreme exposure concentration where the 'wiggliness' of spline-based models becomes more pronounced (Figures 2 and 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 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 show the ability of TDLNM and TDLNMse to adapt to non-smooth exposure–time response surfaces. Figure 3 of supplementary material available at *Biostatistics* online indicates the probability detecting a nonzero 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 nonzero 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.

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 $\mu g/m^3$). Critical windows are defined as any week containing a region in the exposure–time–response where the 95% confidence interval does not contain zero.

6.1. *DLNM results*

The posterior mean exposure–time–response estimates for TDLNMse is shown in Figure 4a. $PM_{2.5}$ 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 gestational weeks. Cross-sections of the exposure–time–response surface at weeks 5, 15, 25, and 35 are shown in Figure 4b 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 5.74 g (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 nonlinear

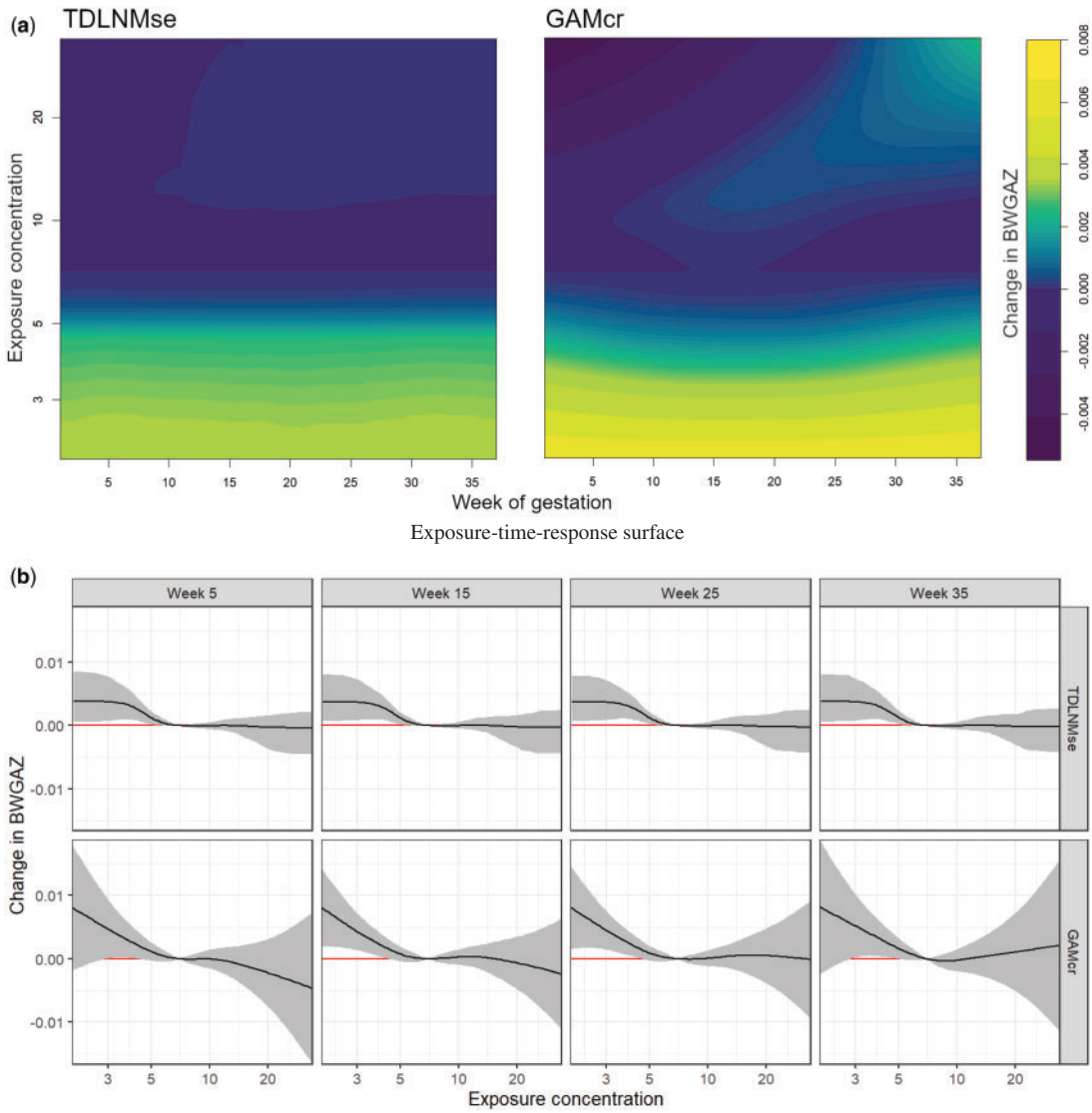

Fig. 4. Panel (a) shows the estimated exposure–time–response surface for models TDLNMse and GAMcr. Panel (b) shows cross sections of the estimated exposure-time-response with columns showing the estimated effect at four times: $t = 5, 15, 25, 35$, while rows compare models TDLNMse and GAMcr. All plots indicate the exposure-time-response relative to the median exposure concentration value ($7.0 \mu g/m^3$).

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 24.1 g (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 4a, 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 *and others* (2016) found a 10 μg/m^3 increase in $PM_{2.5}$ across pregnancy to be associated with 15.9 g 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 *and others* (2015) reported similar results in a separate meta-analysis. Strickland *and others* (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 risk of low birth weight for maternal exposures during each trimester of pregnancy (Ebisu and Bell, 2012).

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 Section D of supplementary material available at *Biostatistics* online.

7. DISCUSSION

In this article, 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 allowing 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 nonlinear 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.

8. Software

The R package `dlmtree` used in the simulation and data analysis can be found at https://github.com/danielmork/dlmtree.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgment

Conflict of Interest: None declared.

Funding

National Institutes of Health (ES028811). These data were supplied by the Center for Health and Environmental Data Vital Statistics Program of the Colorado Department of Public Health and Environment, which specifically disclaims responsibility for any analyses, interpretations, or conclusions it has not provided. This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The RMACC Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

References

Berrocal, V. J., Gelfand, A. E., Holland, D. M. and Statisti-Cian, S. (2009). A spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 176–197.

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 others*. (2018). Prenatal nitrate air pollution exposure and reduced child lung function: timing and fetal sex effects. *Environmental Research* **167**, 591–597.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 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**, 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**, 509–521.

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

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

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 others*. (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.

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**, 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**, 1–13.

GASPARRINI, A., ARMSTRONG, B. AND KENWARD, M. G. (2010). Distributed lag nonlinear models. *Statistics in Medicine* **29**, 2224–2234.

GASPARRINI, A., SCHEIPL, F., ARMSTRONG, B. AND KENWARD, M. G. (2017). A penalized framework for distributed lag nonlinear models. *Biometrics* **73**, 938–948.

HASTIE, T. AND TIBSHIRANI, R. (2000). Bayesian backfitting. *Statistical Science* **15**, 196–223.

LEE, A., LEON H., HSIAO-HSIEN, M. C., YUEH-HSIU, B., SONALI, R., MARIA J., KLOOG, I., WILSON, A., SCHWARTZ, J., COHEN, S., COULL, B. A., WRIGHT, R. O. *and others*. (2018). Prenatal fine particulate exposure and early childhood asthma: effect of maternal stress and fetal sex. *Journal of Allergy and Clinical Immunology* **141**, 1880–1886.

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**, 1087–1110.

MAKALIC, E. AND SCHMIDT, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters* **23**, 179–182.

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

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**, 375–382.

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**, 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.

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**, 1157–1167.

WARREN, J. L., KONG, W., LUBEN, T. J. AND CHANG, H. H. (2020). Critical window variable selection: estimating the impact of air pollution on very preterm birth. *Biostatistics* **21**, 790–806.

WILSON, A., CHIU, Y.-H. M., HSU, H.-H. L., WRIGHT, R. O., WRIGHT, R. J. AND COULL, B. A. (2017*a*). Bayesian distributed lag interaction models to identify perinatal windows of vulnerability in children's health. *Biostatistics* **18**, 537–552.

WILSON, A., CHIU, Y.-H. M., HSU, H.-H. L., WRIGHT, R. O., WRIGHT, R. J. AND COULL, B. A. (2017*b*). Potential for bias when estimating critical windows for air pollution in childrens health. *American Journal of Epidemiology* **186**, 1281–1289.

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 nonlinear model. *Environmental Research Letters* **13**, 24023.

ZANOBETTI, A., WAND, M. P., SCHWARTZ, J. AND RYAN, L. M. (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* **1**, 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.

[*Received June 11, 2020; revised October 7, 2020; accepted for publication November 1, 2020*]