Bayesian nonparametric adjustment of confounding

Chanmin Kim | Mauricio Tec | Corwin Zigler

1Department of Statistics, SungKyunKwan University, Seoul, South Korea
2Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
3Department of Statistics and Data Science, The University of Texas, Austin, Texas, USA

Correspondence
Chanmin Kim, Department of Statistics, SungKyunKwan University, Seoul, Korea. Email: chanmin.kim@skku.edu

Funding information
Environmental Protection Agency, Grant/Award Number: EPA 83587201; National Research Foundation of Korea, Grant/Award Numbers: NRF-2020R1F1A1A01048168, NRF-2022R1F1A1062904; National Institutes of Health, Grant/Award Numbers: NIHRO1ES026217, NIHRO1ES029950, NIHRO1ES030616, NIHRIAG074372

Abstract
Analysis of observational studies increasingly confronts the challenge of determining which of a possibly high-dimensional set of available covariates are required to satisfy the assumption of ignorable treatment assignment for estimation of causal effects. We propose a Bayesian nonparametric approach that simultaneously (1) prioritizes inclusion of adjustment variables in accordance with existing principles of confounder selection; (2) estimates causal effects in a manner that permits complex relationships among confounders, exposures, and outcomes; and (3) provides causal estimates that account for uncertainty in the nature of confounding. The proposal relies on specification of multiple Bayesian additive regression trees models, linked together with a common prior distribution that accrues posterior selection probability to covariates on the basis of association with both the exposure and the outcome of interest. A set of extensive simulation studies demonstrates that the proposed method performs well relative to similarly-motivated methodologies in a variety of scenarios. We deploy the method to investigate the causal effect of emissions from coal-fired power plants on ambient air pollution concentrations, where the prospect of confounding due to local and regional meteorological factors introduces uncertainty around the confounding role of a high-dimensional set of measured variables. Ultimately, we show that the proposed method produces more efficient and more consistent results across adjacent years than alternative methods, lending strength to the evidence of the causal relationship between $SO_2$ emissions and ambient particulate pollution.

KEYWORDS
air pollution, BART, Bayesian nonparametrics, causal inference, confounder selection

1 | INTRODUCTION

Analysis of observational studies always relies on the so-called “no unmeasured confounding” or conditional ignorability assumption that all confounders are measured and included in the model or adjustment procedure. Ideally, the optimal set of variables would be known based on the knowledge of the underlying causal structure of the problem under investigation. However, in many practical settings, the underlying causal structure is not known with certainty, especially when the number of potential confounders is large or the possibility of nonlinearities or interactions in relationships with exposures and outcomes is incompletely understood. Such is the case in a motivating investigation of the causal effect of sulfur dioxide ($SO_2$) emissions from power plants on the ambient concentration of fine particulate matter ($PM_{2.5}$) during 2013 and 2014 in the United States, where the possibility of confounding...
due to local and regional meteorologic conditions (among other factors) generates a large number of features that may interact in complex ways to confound the relationship of interest. Of particular note, the regional nature of meteorologic confounding in this context means that not only local conditions but also meteorological variables in surrounding areas should be considered as potential confounding variables (Tec et al., 2023), which greatly expands the number of available potential confounders. As a result, our analysis entails a total of 104 potential confounding variables measured on each of 12,943 observational units.

Increasing appreciation that many variable selection methods based on regularized regression (such as the LASSO (Tibshirani, 1996) and its variations) are suboptimal for causal effect estimation has motivated a variety of methods for so-called “confounder selection,” as a distinct endeavor to the more common setting of variable selection. A common thread of this research is to orient the prioritization of variables to consider both associations with the exposure of interest and the outcome of interest. Wang et al. (2012) proposed a method called Bayesian adjustment for confounding (BAC) to conduct variable selection and model averaging on both the exposure and outcome models, linking the two models using unknown nuisance indicators for the inclusion of each potential confounding covariate. Extensions to BAC have included those to accommodate binary covariates (Lefebvre et al., 2014) and generalized linear models (Wang et al., 2015). BAC can account for uncertainty around confounder selection in the estimation of causal effects; however, it heavily depends on relatively simplistic parametric model assumptions. Wilson and Reich (2014) proposed a similarly-motivated decision-theoretic method that works well for a variety of sample sizes, but poses difficulties in choosing the final solution path for causal estimation. Shortreed and Ertefaie (2017) proposed the outcome-adaptive LASSO for selecting appropriate covariates for inclusion in propensity score, similarly relying on a parametric linear model and restricted to the case of a binary treatment. Häggström (2018) developed a method to learn the causal structure using a probabilistic graphical model and to estimate the causal effect based on the estimated graph. However, this method estimates a single set of covariates and cannot account for uncertainty in confounder selection. Shi et al. (2019) offered a neural-network approach to jointly learn variables that predict both the treatment and outcome toward inclusion in a targeted regularization scheme, but the finite-sample performance of methods like this in moderately sized datasets is unknown.

As a related thread of research, decision tree ensembles (Breiman, 2001) are a powerful tool to predict nonlinear dose–response relationships that have received increased attention in the causal inference literature. Among many others, the Bayesian additive regression trees (BART; Chipman et al. 2010) method and its modifications have been widely used for obtaining causal estimates. Aside from the implicit variable selection in the construction of the component regression trees, there has been no BART method with an explicit feature for confounder selection when there is a very large set of confounders.

In this paper, we endeavor to unite the objectives of confounder selection procedures such as BAC with the promise of flexible tree-based methods such as BART. The proposed method orient the flexibility, predictive power, and implicit regularization involved in BART toward a procedure explicitly defined to weight confounders (and other variables) in a manner consistent with principles suggested in, for example, VanderWeele (2019), for identifying adjustment variables for estimation of causal effects. Specifically, the exposure and outcome models are jointly fitted via the BART method with a common prior on a vector of selection probabilities. The joint estimation of the two models generates a selection probability vector that is updated in such a way that more posterior selection probability accrues to variables prioritized in the models for the exposure and/or outcome, based in part on a version of the sparsity-inducing prior proposed in Linero (2018). Key benefits of the proposed approach relative to existing methodologies include its avoidance of parametric modeling assumptions, its ability to capture complex relationships and interactions among covariates, exposures, and outcomes, the inherent account of uncertainty when regularizing the large set of covariates in accordance with confounding principles, and the ability to de-emphasize the inclusion of so-called instruments that are associated with the exposure but otherwise unrelated to the outcome of interest. Importantly, the method is not designed in the vein of “causal discovery” research in that it cannot learn the true causal structure underlying the observed data. While individual covariates will be discussed as “confounders,” “instruments,” “predictors,” or “noise,” the iterative nature of the model computation implies that each covariate’s apparent role is learned conditional on other quantities in the model, which will be shown to produce satisfactory estimates of causal effects without any absolute declaration of individual covariates’ (marginal) status as a confounder, instrument, predictor, or noise variable. The usefulness of the method is evidenced—for both a binary and continuous version of exposure—in simulation studies and in the aforementioned analysis of the causal effects of power plant emissions on ambient air pollution, where the proposed method displays more consistent estimates of causal effects across neighboring years than alternatives.
2 | MODEL

2.1 | Causal model

The analysis goal is to quantify the causal effect of an exposure, denoted $A$, on an outcome of interest, denoted by $Y$, in a manner that adjusts for observed confounders. Let $X = \{X_1, X_2, \ldots, X_P\}$ be a set of $P$ observed pre-treatment variables that are regarded as potential confounders which are assumed to contain a subset of the minimal set of confounders required to satisfy ignorability. In many applications including our motivating power plant investigation, the number of potential confounders is large and we do not have a priori knowledge about which should be adjusted for in the causal estimation, in part due to the possibility of complex inter-relationships among true confounders and their relationships with exposures and outcomes. For illustration, we focus development of main ideas with a binary exposure, but will also consider the continuous exposure case. We assume continuous outcomes throughout, but generalizations to other outcome types are available.

We formalize the estimation within a potential outcomes framework (Rubin, 1974). Among a sample of $i = 1, 2, \ldots, N$ observational units, the potential outcome for unit $i$ is defined as $Y_i(a)$ denoting the potential value of the outcome $Y$ that could be observed under the exposure status $A = a$. Then, the target causal estimand is $\Delta(a, a') = E[Y(a) - Y(a')]$, which is estimated by the following equation with observed data under the strongly ignorable treatment assignment assumption, that is, $Y_i(1), Y_i(0) \perp A_i | X_i$ (Rosenbaum & Rubin, 1983)

$$\Delta(a, a'; x) = E[Y|A = a, X = x] - E[Y|A = a', X = x].$$

(1)

This quantity can be marginalized over confounders $X$ to obtain the population effect, $\Delta(a, a')$. It is worth noting that our method adjusts for all available covariates but each will be weighted differently in the posterior inference for causal effects in accordance with certain principles of confounder selection, which will be discussed later. Weighting covariates $X$ properly while estimating their role in the model $E[Y|A, X]$ are the most important tasks. In the next section, we will provide an overview of BART before getting into the task.

2.2 | Overview of Bayesian additive regression tree method

The general BART model (Chipman et al., 2010) is expressed as follows: $y_i = f(X_i) + \varepsilon_i \approx \sum_{h=1}^{H} g(X_i; T_h, M_h) + \varepsilon_i \sim N(0, \sigma^2)$, for $i = 1, \ldots, N$, where $y_i$ and $X_i$ are the $i$th response and vector of predictors, respectively, and $g(X_i; T_h, M_h)$ is dictated by an estimated tree structure. Specifically, each of $H$ distinct tree structures is denoted by $T_h (h = 1, \ldots, H)$ and the parameters denoting the mean outcome at the terminal nodes of the $h$th tree are denoted by $M_h = \{\mu_{h,1}, \ldots, \mu_{h,n_h}\}$ where $n_h$ is the number of terminal nodes of $T_h$. Each tree structure $T_h$ contains internal nodes with corresponding “splitting rule” (e.g., $X_j > c$) consisting of the “splitting variable” $X_j$ and the “splitting value” $c$.

The BART method uses “Bayesian backfitting” (Hastie et al., 2000) for a Metropolis-within-Gibbs sampler, where each tree $j$ is fit iteratively through the residual responses. In each iteration, a new tree structure is proposed via three distinct tree alterations: GROW, PRUNE, and CHANGE (Kapelner & Bleich, 2016) (see Appendix B). In GROW and CHANGE alterations, a new predictor is randomly sampled from a pool of $P$ predictors according to assumed priors. Chipman et al. (2010) originally proposed a uniform prior, $1/P$, on selection probabilities $s = (s_1, \ldots, s_P)$ of $P$ predictors. Recently, Linero (2018) replaced it with a Dirichlet prior to obtain adaptivity to sparsity, $(s_1, \ldots, s_P) \sim D(\alpha/P, \ldots, \alpha/P)$.

2.3 | Bayesian additive regression tree confounder selection

In the spirit of the BAC approach of Wang et al. (2012), we assume two models for estimating the effect of exposure $A$ on outcome $Y$ along with a pool of high-dimensional potential confounders $X = \{X_1, \ldots, X_P\}$: $E(A_i|X_i) = f_a(X_i)$ and $E(Y_i|A_i, X_i) = f_y(A_i, X_i)$ where $f_a(\cdot)$ and $f_y(\cdot)$ are some link functions. In Wang et al. (2012), Zigler and Dominici (2014), and Wang et al. (2015), parametric generalized linear models are assumed with vectors of indicators $\alpha^A$ and $\alpha^Y$ indicating whether each covariate is selected into the treatment and outcome models, respectively. High posterior probability of including confounders in the adjustment is accomplished through a joint prior on $\alpha^A$ and $\alpha^Y$ that renders variables selected into the treatment model more likely to be selected into the outcome model (thus prioritizing covariates associated with both $A$ and $Y$). The average causal effect is estimated by Bayesian model averaging (BMA).

A drawback of the aforementioned methods is they are not able to explore all possible model forms such as nonlinear terms in $f_a(\cdot)$ or $f_y(\cdot)$. We propose a new flexible method to account for uncertainty in confounder selection using BART models that do not depend on parametric model structures:

$$P(A_i = 1) = \Phi\left(\sum_{h=1}^{H} g_a(X_i; T_h, M_h)\right)$$

(2)
\[ y_i = \sum_{h=1}^{H} g_{\delta}(a_i; \mathbf{X}_i; \mathbf{T}'_h, \mathcal{M}'_h) + \epsilon_i, \]

where \( \epsilon_i \sim N(0, \sigma^2) \)  

(3)

for subjects \( i = 1, \ldots, N \) where \( \Phi \) is the standard normal cdf. Note that it is straightforward to accommodate a continuous exposure by replacing Equation (2) with \( A_i = \sum_{h=1}^{H} g_{d}(\mathbf{X}_i; \mathbf{T}_h, \mathcal{M}_h) + \omega_i, \omega_i \sim N(0, \tau^2) \). In a manner analogous to BAC, the exposure and outcome models are fitted jointly to assign posterior weight to models that emphasize the “right” confounders. Instead of introducing explicit covariate inclusion parameters (e.g., \( \alpha^A, \alpha^Y \) in BAC), the exposure and outcome models are linked via common prior on selection probabilities \( s = (s_1, \ldots, s_P) \) of \( P \) potential confounders in GROW and CHANGE alterations. Specifically, we use a common sparsity-inducing Dirichlet prior \( (s_1, \ldots, s_P) \sim D(\alpha/P, \ldots, \alpha/P) \) which gives a conjugate Gibbs-sampling update, \( (s_1, \ldots, s_P) \sim D(\alpha/P + m_1 + n_1, \ldots, \alpha/P + m_P + n_P) \), where \( m_j \) and \( n_j \) denote the number of splits on confounder \( X_j \) in Models (2) and (3), respectively. This conjugate update is based on the following likelihood of \( s \),

\[ \prod_{h=1}^{H} \prod_{b \in T_h} s_b \prod_{c \in T'_h} s_c, \]

where \( b \) and \( c \) denote nodes of trees \( T_h \) and \( T'_h \), respectively, and \( j_b \) and \( j_c \) denote the predictors used to split nodes \( b \) and \( c \), respectively. This likelihood is correct under a certain assumption (Assumption 2.2 in Linero 2018) which is acceptable whenever the number of unique values of each confounder is sufficiently large or the trees are typically very shallow, which is often reasonable.

If a given covariate, \( X_j \), is used often as a splitting variable in either the model for \( A \) or the model for \( Y \), the model will accumulate posterior mass on the selection probability \( s_j \) through larger values of \( m_j \) or \( n_j \), with the most selection probability accumulating for variables that are important in the prediction of \( A \) and/or \( Y \) (which will accumulate larger values of \( m_j \) and/or \( n_j \)). Thus, selection probabilities will tend to favor the \( X_j \) that are associated with \( A \), associated with \( Y \), or associated with both \( A \) and \( Y \). The variables ultimately used for effect estimation in the model for \( Y \) will be those that are proposed for splitting via this prior and accepted in the updating step of the model for \( Y \), which will further prioritize variables associated with \( Y \). We discuss this confounder prioritization in detail in Section 4.

### 2.3.1 Separate models for each exposure arm

In the case of a binary treatment, a modified version of Model (3) is proposed for each \( A = a \in \{0, 1\} \),

\[ y_i = \sum_{h=1}^{H} g_{\delta}^{a}(\mathbf{X}_i; \mathbf{T}'_h, \mathcal{M}'_h) + \epsilon_i^a, \quad \epsilon_i^a \sim N(0, \sigma^2) \text{ for } i \in I_a, \]

(4)

where \( I_a \) denotes a set of observations under \( A = a \in \{0, 1\} \). Practically, three models, Models (2) and (4) for \( A = 0 \) and \( A = 1 \), are fitted with a common prior on the selection probabilities \( s = (s_1, \ldots, s_P) \) with a conjugate sampling update \( (s_1, \ldots, s_P) \sim D(\alpha/P + m_1 + n_1^0, \ldots, \alpha/P + m_P + n_P^0) \), where \( n_1^0 \) and \( n_P^0 \) are the number of splits on confounder \( X_j \) in Model (4) for \( A = a \in \{0, 1\} \).

### 2.3.2 Single model

As first highlighted in Hill (2011) and similarly considered in Hahn et al. (2020), if there exist parts of the covariate space with little or no observed representation from both exposure groups (which is referred to as lack of common support), estimating two separate outcome models for exposures \( A = 0 \) and \( A = 1 \) may produce highly biased estimates. Furthermore, specification of two separate outcome models is moot when \( A \) is not binary. This motivates consideration of a single outcome Model (3), but doing so introduces a key challenge for the confounder selection case. Note that the outcome model has an additional selection probability parameter \( s_q \) that is tied to the exposure variable in the vector of selection probabilities \( s = (s_0, s_1, \ldots, s_P) \), while the exposure model only uses the subset \( s' = (s'_1, \ldots, s'_P) \) in the selection of \( P \) predictors. Thus, it is not straightforward to specify a common prior on the vectors of selection probabilities \( s \) and \( s' \).

We link two vectors using a common prior by reparameterizing \( s' = (s'_1, \ldots, s'_P) = (s_1/(1-s_0), s_2/(1-s_0), \ldots, s_P/(1-s_0)) \) based upon neutrality of the Dirichlet distribution. With the Dirichlet distribution \( D(\alpha/P, \alpha/P, \ldots, \alpha/P) \) for the common prior for \( s = (s_0, s_1, \ldots, s_P) \), the update of \( s \) is based on the following likelihood \( \times \text{prior} (Q) \): \( Q = \prod_{h=1}^{H} \prod_{b \in T_h} s_b \prod_{c \in T'_h} s_c \prod_{j=0}^{P} s_j^{a/P-1} \), which is valid under the same assumption used in Section 2.3. This is equivalent to \( Q = \left( \frac{1}{1-s_0} \right)^{P} \prod_{j=1}^{P} m_j (s_j + \alpha/P - 1) m_j + n_j + \alpha/P - 1 \), where \( m_j \) and \( n_j \) denote the number of splits on confounder \( X_j \) in the exposure and outcome models, respectively. Since we no longer entertain an efficient conjugate update, we use the Metropolis-Hastings algorithm to update \( s \) with a Dirichlet proposal distribution based on \( Q \), that is, \( D(n_0 + \alpha/P, m_1 + n_1 + \alpha/P, \ldots, m_P + n_P + \alpha/P) \), through the following acceptance ratio \( P_{AR}(s \rightarrow s^\text{new}) = \min\{1, (\frac{\sum_{j=1}^{P} s_j^{a/P}}{\sum_{j=1}^{P} s_j^{a/P}}} \) \}, where \( s^\text{new} \) is a proposed vector.
One potential drawback of this approach is that the exposure variable is now a part of the variables to be selected based on $s$. To avoid a situation that the exposure variable is rarely used in the tree structure, we can consider a modified Dirichlet proposal distribution $D(n_0 + c + \alpha/P, m_1 + n_1 + \alpha/P, m_2 + n_2 + \alpha/P, ..., m_J + n_J + \alpha/P)$ where $c$ is some positive number. Since other probabilities $(s_1, ..., s_J)$ are subject to additional powers based on the numbers of splits (i.e., $m_j$'s) used in the exposure models, the selection probability $s_0$ may be underweighted in the original $Q$. In our simulation and application studies, we use $c = n_0$, which results in $s_0^{2n_0 + \alpha/P - 1}$ component in the Dirichlet proposal distribution.

3 | ESTIMATION

3.1 | Posterior computation for the Bayesian additive regression tree models

To draw posterior samples from $P(T'_1, ..., T'_H, \mathcal{M}'_1, ..., \mathcal{M}'_H | \sigma^2 | D)$ for the outcome model (3) (or 4), we use “Bayesian backfitting” (Hastie et al., 2000) for a Metropolis-within-Gibbs sampler, where each tree $T'_h$ is fit iteratively via the residual responses: $R_{i,-j} = y_i - \sum_{j \neq j} g_j(X_i; T'_h, \mathcal{M}'_h)$. For each tree $T'_h$, we propose a new tree structure $T'_h$ from the full conditional $[T'_h | R_{1,-j}, ..., R_{n,-j}, \sigma^2]$ (i.e., tree alterations) and the parameter within the tree updated through the full conditional $[\mathcal{M}'_h | T'_h, R_{1,-j}, ..., R_{n,-j}, \sigma^2]$ with acceptance ratios for three alteration steps. The exact forms of the acceptance ratios for three alteration steps and the details of the posterior computation steps are provided in Appendix C. Note that for the separate model, we draw posterior samples from the outcome model for each $A = a \in \{0, 1\}$. To draw posterior sample for the binary exposure model (2), we introduce latent variable $Z$: set $I(Z_i > 0) = A_i$ for $i = 1, ..., n$ and apply general BART for continuous data to $Z$. If $A$ is continuous, the updating is analogous to Bayesian backfitting described for model (3). After updating all tree structures and the corresponding parameters, we update all remaining parameters based on the Gibbs sampler or the Metropolis-Hastings algorithm. Finally, for the single-model scheme, we use the Metropolis-Hastings algorithm to update the vector of selection probabilities $s$ through the acceptance ratio $P_{AR}(s \rightarrow s_{\text{new}}) = \min \{1, \left( \frac{1 - \sum_{j=1}^J s_j}{1 - \sum_{j=1}^J m_j} \right)^{\sum_{j=1}^J m_j} \}$. For the separate model approach, $s$ is updated with a conjugate sampling update as in Section 2.3.1.

3.2 | Estimation of the causal effect

To estimate the target causal estimand $\Delta(\alpha, \alpha')$ in Equation (1) for binary exposure (i.e., $a = 1, \alpha' = 0$), we use the posterior samples drawn from Section 3.1. For the separate model, we estimate $\Delta(1,0)$ through the following equation:

$$\hat{\Delta}(1,0) = 1 \over N \sum_{i=1}^N \left[ \frac{1}{R} \sum_{r=1}^R \left\{ \sum_{h=1}^H g_y^1(r)(X_i; T'_h, \mathcal{M}'_h) - \sum_{h=1}^H g_y^0(r)(X_i; T_0, \mathcal{M}_0) \right\} \right],$$

(5)

where $g_y^1(r)$ and $g_y^0(r)$ are the $r$th posterior samples for Model (4) with $A = 1$ and $A = 0$. For the single model, we evaluate the following equation instead

$$\hat{\Delta}(1,0) = 1 \over N \sum_{i=1}^N \left[ \frac{1}{R} \sum_{r=1}^R \left\{ \sum_{h=1}^H g_y^1(r)(X_i; T'_h, \mathcal{M}'_h) - \sum_{h=1}^H g_y^0(r)(X_i; T'_h, \mathcal{M}'_h) \right\} \right],$$

(6)

where $g_y^a(r)$ is the $r$th posterior samples for the model (3). Like BAC, our approach does not require a separate BMA step to estimate the causal effects averaged across all possible outcome models. Since each posterior samples of $g_y^a(r)$ or $g_y^0(r)$ only include selected confounders in their nodes, averaging over posterior samples of $g_y^a(r)$ and $g_y^0(r)$ suffices for estimation of causal effects averaged over the different model specifications, provided that the outcome model includes at least the minimal set of confounders required to satisfy the assumption of ignorability. We elaborate on the methods ability to ensure this in the following section.

4 | CONFOUNDER PRIORITIZATION WITH BAYESIAN ADDITIVE REGRESSION TREE METHOD

The proposed method entails two features for prioritizing adjustment variables for estimation of causal effects: the splitting prior and the posterior updating of the outcome model in Equation (3) (or Model (4)). To illustrate, we focus on the separate model in Section 2.3.1 involving Model (4). Denote by $X_A$ and $X_Y$ the sets of covariates in the true exposure model and the true outcome model, respectively, which would contain as a subset the minimal set of
confounders required to satisfy ignorability, as well as some additional variables. Among the selection probabilities \( s = (s_1, ..., s_p) \), \( s_j \)'s will have values approaching 0 if the corresponding predictors \( X_j \)'s are neither in \( \mathcal{X}_A \) nor in \( \mathcal{X}_Y \) when the number of predictors is large relative to \( \alpha \) in the Dirichlet prior (see Appendix D for a detailed explanation). Those covariates with 0 selection probabilities are not available as splitting variables in tree alteration steps and will not enter into the Models \((2)\) and \((4)\) with probability 1. If \( X_j \) is a predictor in either \( \mathcal{X}_A \) or \( \mathcal{X}_Y \) or both, then \( s_j \) will have non-zero posterior probability. Thus, the \( H \) and \( I \) is a predictor in either \( \mathcal{X}_A \) or \( \mathcal{X}_Y \) or both, then \( s_j \) will have non-zero posterior probability. Thus, the proposed prior ensures that the covariates available for tree construction meet the “disjunctive cause criterion” (VanderWeele, 2019), where the set of predictors controlled for is related to the exposure, the outcome, or both. In general, this criterion is a good alternative to two others: (a) the pre-exposure criterion to control for any predictor that is prior to the exposure (which would pertain to all pre-exposure predictors including every \( X_j \in \mathcal{X}_A \cup \mathcal{X}_Y \)); and (b) the common cause criterion to adjust for all pre-exposure covariates that are common causes of exposure and outcome (which would pertain to \( X_j \in (\mathcal{X}_A \cap \mathcal{X}_Y) \) only).

A better alternative to all of the aforementioned criteria is to exclude from the “disjunctive cause criterion” subset any variables exhibiting association with \( A \) without exhibiting an association with \( Y \), known as the “disjunctive cause criterion without instruments” (VanderWeele, 2019). More information about instruments can be found in Appendix F. As shown previously, the proposed joint splitting prior prioritizes the “disjunctive cause criterion,” where \( s_j \) can accumulate nonzero mass for instrumental variables owing to many accepted splits in the model for \( A \). However, in the tree alteration steps of the outcome models, even if an instrument \( X_j \) with \( s_j > 0 \) is randomly sampled from the pool of covariates and suggested for a next splitting variable, its acceptance with the Metropolis-Hastings step for the next tree structure will relate to whether the proposed split results in improved prediction of \( Y \); an instrumental variable not directly related to the outcome will rarely be accepted for a tree split in the model used for estimating causal effects. Thus, the combination of the joint splitting prior and the posterior updating steps for the outcome model renders our method closely linked to the “disjunctive cause criterion without instruments” criterion.

The method’s adherence to the above confounder selection principles implies that posterior estimates of the causal effects with the expressions in Equation \((5)\) or \((6)\) will be based on the relevant variables for unbiased effect estimation. To illustrate, let \( \mathcal{X}_G = (\mathcal{X}_A \cap \mathcal{X}_Y) \) be the set of covariates that satisfy the common cause criterion, and \( \mathcal{X}_* = (\mathcal{X}_A \cap \mathcal{X}_Y) \cup \mathcal{X}_Y \) be the set of covariates that satisfy the disjunctive cause criterion without instruments. The set of variables in \( \mathcal{X}_G \) represents the minimal set of confounders required to satisfy the ignorability assumption, and will suffice to control for confounding for the effect of the exposure on the outcome in the absence of unmeasured confounders. While \( \mathcal{X}_A \) represents a minimal set, the set \( \mathcal{X}_* \) is preferred. Denote \( I, I_{\gamma}, \) and \( I_{\star} \) the set of all possible confounders configurations, a subset of covariate configurations that include at least all the covariates in \( \mathcal{X}_G \) (and possibly others) as a splitting variable at least once, and a subset of covariate configurations that use all the covariates in \( \mathcal{X}_* \) as a splitting variable at least once, respectively. Note that \( I_{\star} \subset I_{\gamma} \subset I \). Denote \( R, R_{\gamma}, \) and \( R_{\star} \) the corresponding posterior samples from models in \( I, I_{\gamma}, \) and \( I_{\star} \), respectively. Then, the posterior mean estimate of the target causal estimand in Equation \((5)\) can be decomposed into three parts as follows:

\[
\Delta(1, 0) = \frac{1}{|R_{\star}|} \sum_{r \in R_{\star}} \left[ \frac{1}{N} \sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} g_{y, 0}^{1, (r)}(X_i; T_h^1, \mathcal{M}_h^1) \right\} \right] + \frac{1}{|R_{\gamma}| - |R_{\star}|} \sum_{r \in R_{\gamma} \setminus R_{\star}} \left[ \frac{1}{N} \sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} g_{y, 0}^{1, (r)}(X_i; T_h^1, \mathcal{M}_h^1) \right\} \right] + \frac{1}{|R| - |R_{\gamma}|} \sum_{r \in R \setminus R_{\gamma}} \left[ \frac{1}{N} \sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} g_{y, 0}^{1, (r)}(X_i; T_h^1, \mathcal{M}_h^1) \right\} \right],
\]  

where the first two terms of summations represent models that will provide unbiased estimation of the causal effect (in the absence of unmeasured confounding); the first is the sum over models that include the set of confounders in \( \mathcal{X}_G \), the second term is the sum over models that include the confounders in \( \mathcal{X}_* \). The third term represents models that omit at least one confounder and will not provide unbiased estimation of the causal effect.

The BART confounder selection procedure proposed above is designed precisely to allocate posterior support to models in the first two terms of expression \((7)\) (i.e., large \(|R_{\gamma}|\) and \(|R_{\star}|\) relative to \(|R|\), particularly the first. The joint splitting prior will ensure that splits are proposed in accordance with models in the first two terms, the posterior updates of the model in Equation \((3)\) (or \((4)\)) will be accepted to focus posterior weight on the first term.
5 | SIMULATION

We test our model performance based on a toy example with \( N = 300 \) (\( N = 500 \) for the last scenario) observations. In six different scenarios, 100 potential confounders (\( X_1 - X_{100} \)) are independently generated from \( N(0, 1) \) where only five of them (\( X_1 - X_5 \)) are true confounders: (Scenario 1) A model contains only true confounders (\( X_1 - X_5 \)). Y model contains true confounders (\( X_1 - X_5 \)) and predictors (\( X_6 - X_7 \)); (Scenario 2) A model contains only true confounders (\( X_1 - X_5 \)). Y model contains true confounders (\( X_1 - X_5 \)) and ‘instrumental variables’ (\( X_6 - X_7 \)). Y model contains true confounders (\( X_1 - X_5 \)); (Scenario 3) A model contains true confounders (\( X_1 - X_5 \)) and ‘instrumental variables’ (\( X_6 - X_7 \)). Y model contains true confounders (\( X_1 - X_5 \)); (Scenario 4) A model contains only true confounders (\( X_1 - X_5 \)). Y model contains true confounders (\( X_1 - X_5 \)) and predictors (\( X_6 - X_{17} \)); (Scenario 5) A model contains only true confounders (\( X_1 - X_5 \)). Y model contains true confounders (\( X_1 - X_5 \)) and predictors (\( X_6 - X_{17} \)). The outcome and exposure models are the same in Scenario 4, but the sample size is \( N = 500 \); (Scenario 6) the same settings in Scenario 4 except the coefficients of the true confounders in the Y model are relatively smaller. See Appendix E.1 for the detailed specification.

We generate \( m = 200 \) replicates under each scenario. For our proposed approach, we consider both methods discussed in Section 2, the separate and single models. The MCMC chain runs for 20,000 iterations and the first half is discarded as burn-in. To reduce autocorrelation among samples and to save computer memory, the thinning interval is set to 10. Convergence is examined for each replicate data using the Gelman–Rubin (G–R) diagnostics.

5.1 | Alternative methods for comparison

To compare our model performance to others, we consider six similarly-motivated comparison methods: (a) Bayesian causal forest (BCF, Hahn et al. 2020); (b) BAC (Wang et al. 2015); (c) Markov/Bayesian network confounder selection (CovSelHigh, Häggström 2018); (d) Bayesian penalized credible region (BayesPen, Wilson and Reich 2014); (e) BART with a sparsity-inducing Dirichlet prior on the selection probabilities (DBART, Linero 2018); and (f) Neural networks (Dragonnet, Shi et al. 2019). The detailed explanation of these methods is available in Appendix E.2.

5.2 | Simulation results, \( P < N \)

Table 1 summarizes the results from the six simulation settings. In terms of bias and MSE, the proposed single model outperforms other methods across all scenarios. Note in particular the single model produces biases (and MSEs) three times less than those from the BCF model, a somewhat surprising result since BCF has been shown to perform very well relative to other similar methods for causal effect estimation (Hahn et al., 2020). However, it is worth noting that the present study does not consider heterogeneous treatment effects (a key strength of BCF), and the default prior specification in BCF is not intended to focus posterior support on a set of necessary covariates that is a small subset of those available for analysis, as is the case here. Appendix E.3 examines different simulation scenarios designed for the settings, where BCF is known to perform well, and illustrates situations where the proposed model performs well compared to BCF or vice versa. DBART, which adds a variable selection prior to the BART outcome model, outperforms the separate model in terms of bias and MSE, but has particularly low coverage as the number of predictors increases (Scenarios 4, 5, and 6).

The poor performance of BAC in this simulation study is not particularly surprising given the fact that the method depends on (generalized-)linear models such that it is hard to capture nonlinear terms of the confounders simulated in these scenarios. Similarly for the poor performance of BayesPen in this simulation study, as this method also depends on parametric model structures. The CovSelHigh model based on all causes of treatment and outcome (CovSelHigh\(X,T,Y\)) performs relatively well compared to BAC, BayesPen and its two siblings (CovSelHigh\(Q\), CovSelHigh\(Z\)) as illustrated in Häggström (2018). However, the resulting biases are larger than 0.1 for Scenarios 1, 3, 4, 5, and 6, whereas the proposed single model produces biases much smaller than 0.1 across the scenarios. Except for Scenario 1, the Dragonnet produces results with larger biases and higher MSE’s than our model. Furthermore, this method yields no uncertainty quantification, so coverage cannot be evaluated.

To illustrate how the various methods prioritize different variables, Figures 1–3 show the posterior probabilities of inclusion for all covariates in the proposed, BAC, CovSelHigh, and DBART methods across simulation scenarios 1, 3, and 6. The posterior probability of inclusion is defined as \( \frac{1}{200} \sum_{m=1}^{200} P_m(j) \) for each \( j \) potential confounder, where \( P_m(j) \) denotes the posterior probability of inclusion of the \( j \)th confounder in the model for the \( m \)th simulation replicate. Across all three of these scenarios, the proposed methods (the top two panels in each figure) select true confounders (red points) with posterior probability 1, with the single exception in Scenario 6, where the confounder with the weakest association with the outcome (\( X_4 \)) has average posterior inclusion probability around 0.8. In terms of confounder selection, the DBART model...
produces the most similar results to the proposed models; however, as the number of predictors increases (Scenario 6), other variables begin to have non-trivial posterior inclusion probabilities. The BAC method similarly exhibits very high posterior probability of inclusion for the main effects of the true confounders, but struggles to select interaction terms involving these confounders, likely a consequence of its reliance on the need to specify a functional form of the interaction a priori that, in this case, does not reflect the data generation. The ability of the CovSelHigh algorithms to include the true confounders is more variable across these scenarios, indicating substantial posterior probability of missing confounders in Scenario 6, when the associations are smaller than in the other scenarios. Figure 1 shows how the proposed methods are more likely than the comparison methods to include predictors of $Y$ unassociated with $A$. Figure 2 illustrates how the proposed methods satisfy the disjunctive cause without instruments criterion. The proposed methods give lower than a 50% posterior probability of inclusion for each of the two instruments, but assign posterior probability 1 to all of true confounders. Specifically, our proposed models assigns posterior probability of 1 to the first two terms in Equation (7) (i.e., $I_\cap$), and the BAC also assigns posterior probability of 1 to the models in $I_\cap$. If we further examine the nature of the models constituting $I_\cap$, our
FIGURE 1 [Scenario 1] The posterior probability of inclusion in the model. Five plots are for the single, separate, BAC, CovSelHigh_{XTY}, and DBART, respectively. Red triangles (△) indicate five true confounders and blue crosses (+) indicate two additional predictors in the outcome model. In the single model and the BAC model, a black star (⋆) indicates the exposure variable. In the BAC model, the last 100 points are for interactions between the exposure variable and each confounder. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version. BAC, Bayesian adjustment for confounding.

proposed separate model (and single model) assigns 0.70 (and 0.60) to the models in \( I_s \) (i.e., the first term in Equation (7)), while the BAC assigns 0 posterior probability to the models in that class, meaning that in the presence of an unmeasured confounder and the threat of Z-bias, our methods can weight more the models that produce unbiased effect estimates. Figure 3 shows how the methods perform when associations are comparatively lower than the other scenarios.

Despite the clear performance advantage in these data generations, the proposed method exhibits undercoverage of 95% uncertainty intervals. This may be due to the methods’ accrual of nonzero weight to models that miss at least one confounder (Equation (7)), but more likely in this case due to the models’ extrapolation to areas of where covariate distributions do not “overlap” in the exposure groups. Methods based on BART are known produce overly precise inferences in areas, where extrapolation is required due to non-overlap. Appendix E.6 provides a detailed analysis of uncertainty quantification as it relates to non-overlap. Also, Li et al. (2022) provided more detailed discussions.

We also perform an additional simulation study to evaluate the ability of the proposed model when the number of potential confounders (\( P = 100 \)) is strictly larger than the sample size (\( n = 90 \)). Both of our proposed models outperform in terms of biases and MSEs. The results are provided in Appendix E.4. The average computation time of the proposed model for each combination of \( P = (50, 200, 600) \) and \( n = (100, 500, 1000) \) is provided in Appendix E.5. We set the number of trees (\( h \)) to 100 as the default. If the number of true confounders is expected to be large in comparison to \( h \), then larger \( h \) may be needed.

6 | ESTIMATING CAUSAL EFFECTS OF POWER PLANT EMISSIONS ON AMBIENT PM\(_{2.5}\)

A variety of federal regulations are designed to protect public health through limiting exposure to harmful pollution. Key regulatory efforts target SO\(_2\) emissions from power plants specifically, as these emissions are known to interact in the atmosphere to elevate ambient concentrations of fine particulate matter (PM\(_{2.5}\)). Owing to the outsized public health impact of PM\(_{2.5}\) exposure and the major contributions from power plants, there is
growing interest in quantifying the air quality impacts of emissions (or interventions to reduce emissions) from power plants (Kim et al., 2020) using large-scale data on emissions, meteorology, air quality, and population demographics.

We consider an analysis of emissions from 397 coal-fired power plants (comprised of 935 electricity generating units) active during 2013 for their relationship with ambient PM$_{2.5}$ concentrations during the same year, focusing on states in the Eastern United States where pollution from coal plants is of particular consequence (see Figure 4). To measure each ZIP code’s exposure to power plant emissions, we use a reduced-complexity atmospheric model, called HYSPLIT Average Dispersion (HyADS) (Henneman et al., 2019). Viewing the ZIP code as the observational unit, potential confounding factors for each ZIP code come from several sources. Appendix A summarizes the construction of the data, which ultimately comprises data on 12943 ZIP codes, each having a measure of HyADS emissions exposure ($A$), ambient PM$_{2.5}$ concentration ($Y$), and 104 potential confounders ($X$). We produce analogous data for 2014 to investigate consistency of results across consecutive years.

HyADS represents a continuously-scaled exposure and will be analyzed as such, but for illustrative purposes and for comparison with other methods, it will also be analyzed as a binary exposure, in which case it is dichotomized to define each ZIP code as either high or low exposed to SO$_2$ emissions based on the mean of the 2013 HyADS coal emissions exposure levels (980,249,908) as in Figure 4.

Table 2 includes posterior mean estimates (with 95% posterior intervals) for the causal effect of (binary) HyADS exposure on ambient PM$_{2.5}$ from six different methods: our single and separate models, BCF, BAC, and DBART. Table 2 also includes estimates from Dragonnet, reported with standard deviations of the effect estimate based on 20 random initializations of the neural network. We exclude CovSelHigh from Häggström (2018) and BayesPen from Wilson and Reich (2014) from this analysis because the former evaluates all possible network connections among variables and easily exceeds the memory limit with a larger set of confounders, and the latter requires computation of $(X^TX)^{-1}$ for the model with all available covariates, which is difficult to invert with multicollinearity (e.g., multiple temperature measures for each locations). For
**FIGURE 3** [Scenario 6] The posterior probability of inclusion in the model. Five plots are for the single, separate, BAC, CovSelHigh\_XTY, and DBART, respectively. Red triangles (△) indicate five true confounders and blue crosses (+) indicate 12 additional predictors in the outcome model. In the single model and the BAC model, a black star (∗) indicates the exposure variable. In the BAC model, the last 100 points are for interactions between the exposure variable and each confounder. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version. BAC, Bayesian adjustment for confounding.

**FIGURE 4** The map of the zip code locations: the treated (low HyADS coal-emissions exposure; blue locations) vs. the control (high HyADS coal-emissions exposure; red locations) locations. The analysis includes the zip codes located in the Eastern US. The histogram shows frequencies of high or low HyADS coal-emissions exposures dichotomized at the mean (980249908). This figure appears in color in the electronic version of this paper, and any mention of color refers to that version. HyADS, HYSPLIT Average Dispersion.
our proposed models and DBART model, the number of distinct trees is set to 200 for each model and for the (hyper-)parameters, we use the recommended settings in Kapelner and Bleich (2016). We run two MCMC chains each with 100,000 iterations and discard the first half as burn-in. To check MCMC convergence, we examine the Gelman–Rubin diagnostics for multiple chains and indicates that no critical convergence issues are detected (1.2 for the single model and 1.1 for the separate model).

Our approaches estimate that low coal-emissions’ exposure causes reduction in ambient PM$_{2.5}$ concentrations by $-0.06 (-0.09, -0.01)$ mg/m$^3$ and $-0.61 (-0.64, -0.58)$ mg/m$^3$ for the single model and the separate model, respectively, for year 2013. For 2014 (using the same cut-off value to dichotomize the exposure), the single and separate model estimate that the low HyADS emissions exposure causes reduction in ambient PM$_{2.5}$ concentrations by $-0.08 (-0.15, -0.01)$ mg/m$^3$ and $-0.26 (-0.37, -0.15)$ mg/m$^3$, respectively, representing close agreement with the analysis from 2013 and suggesting adequate control for confounding. These results indicate that ambient PM$_{2.5}$ concentrations are affected by the HyADS coal-emissions exposure on average, especially when allowing the model for PM$_{2.5}$ to differ in the low and high HyADS exposure areas.

In this application, since the treatment variable is significantly affected by region/climate, the degree of non-overlap region appears to be severe when the overlap through the estimated propensity score is analyzed (Appendix E.6). It is well known that general BART in non-overlap regions is inadequately overconfident in uncertainty quantification. However, as a result of our analysis, we find that the single model can alleviate these issues to some extent, and based on this, we can consider the results of the single model to be more reliable.

TABLE 2 Estimates of the causal effect of low HyADS coal-emissions exposure on annual ambient PM$_{2.5}$ concentrations from five different methods for years 2013 and 2014.

| Year | Ours (single) | Ours (separate) | BCF | BAC | DBART | Dragonnet |
|------|--------------|----------------|-----|-----|-------|-----------|
| 2013 | Posterior Mean | $-0.06$ | $-0.61$ | $-0.28$ | $0.59$ | $-0.07$ | $-0.08$ |
|      | Posterior 95% C.I. | $(-0.09, -0.01)$ | $(-0.64, -0.58)$ | $(-0.39, -0.19)$ | $(0.47, 0.71)$ | $(-0.15, 0.01)$ | $0.09^\dagger$ |
| 2014 | Posterior Mean | $0.08$ | $0.26$ | $0.04$ | $-0.28$ | $-0.06$ | $-0.26$ |
|      | Posterior 95% C.I. | $(-0.15, -0.01)$ | $(-0.37, -0.15)$ | $(-0.08, 0.16)$ | $(-0.39, -0.19)$ | $(-0.11, -0.02)$ | $0.09^\dagger$ |

$^\dagger$In the case of the Dragonnet model, a standard deviation derived from 20 random initializations of neural network is reported. BAC, Bayesian adjustment for confounding; BCF, Bayesian causal forest; HyADS, HYSPLIT average dispersion.

is likely due in this case to the algorithm frequently involving MCMC iterations with an exposure model producing complete separation of exposure levels.

6.1 Inference for variable importance

Figures S7 and S8 in Appendix H depict the posterior probability of inclusion for each covariate in the proposed models (the single and separate models). Considering variables with at least a 100% posterior probability of appearing in the trees as variables that satisfy the disjunctive cause without instruments criterion, the method identifies 58 variables from the single model and 91 covariates from the separate model. The single model prioritizes a smaller set of confounders and a more restrictive response surface, which results in a narrower posterior interval. Even though the two methods have a similar pattern in terms of which variables are included, some of variables play different roles in the two models. Table S5 in Appendix H shows the variables included in each model, grouped by the type of variable.

Examining the posterior inclusion probabilities provides a sense of the importance of both local and regional meteorological factors, with both of our proposed models including many weather variables at various regional directions. The separate model includes almost all neighboring weather variables (70 regional variables with posterior inclusion probability = 100%) as well as 12 (6 winter and 6 summer) local weather variables. The single model prioritizes a smaller set of regional weather variables, with 37 regional weather variables having posterior inclusion probability = 100%. From the results, we can deduce that the importance of the regional weather variable in this causal estimation is quite large relative to the local meteorological conditions.

In Appendix G, we also conduct analyses for years 2013 and 2014 with the exposure, $A$, regarded as continuous where, as illustrated in Section 2.3. In addition, we provide the exposure–response function $E[Y(a)]$ of varying $A = a$ values for the single models fit to data from both 2013 and 2014 (the partial dependence plots).
7 | DISCUSSION

In this paper, we present a BARTs model to estimate causal effects from observational data when a number of potential confounders is large relative to the sample size, indicating potential benefits of concerted efforts to prioritize which of a large set of potential confounders should be included in the analysis. The proposed method allows estimation of causal effects with such prioritization of the relevant variables, while accounting for uncertainty in confounder selection. Since the proposed method does not depend on parametric model assumptions, it can handle any data model having complex nonlinear dose–response functions. The simulation results show that the proposed method outperforms many existing methods, and consistency of results in analyses of power plant pollution in consecutive years points toward successful adjustment for confounding relative to other similar methods.

Note that ignorability cannot be empirically confirmed, and the method's ability to infer a specific causal structure may be limited. For example, in the presence of M-bias (i.e., there are additional unmeasured confounders), our method may open a backdoor path by selecting a certain covariate that should not be chosen. However, the rationale here is that this method takes a set of covariates (that are assumed to include all confounders as a subset) and adjusts for them in an outcome model that weights each covariate differently in accordance with principles of confounder selection. This weighting serves as useful regularization when the set of covariates is large, and avoids the need to manually choose in an ad-hoc fashion which of the set of covariates to include in the model—all covariates are included, but weighted differently.

There are several avenues for future work to refine the methodology proposed here. First, the methods proposed here use the default settings for the component BART model (hyper-) parameters. Even though we find no evidence that causal estimates in the cases considered here heavily depend on the parameter settings, we could alternatively use cross-validation to tune the (hyper-) parameters for more precise estimates. The principles outlined here might also extend to settings, where there are more causal variables of interest, for example, in the mediation analysis context. Because the data in the application are spatially correlated, it is also possible to add a spatial random effect to the proposed model (e.g., Kim 2022) and test its performance.

There are also potential extensions to incorporate the confounder selection capabilities explored here into the BCF framework. In principle, a common splitting prior could be incorporated for confounder selection in both the propensity score and BCF outcome model, but a key challenge is simultaneously estimating the propensity score function and the prognostic function since the latter is a function of the former.

ACKNOWLEDGMENTS

The authors gratefully acknowledge that this work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2020R1F1A1048168, NRF-2022R1F1A1062904), US National Institutes of Health (NIHR01ES026217, NIH01ES029950, NIH01ES030616, NIHRF1AG074372), and US Environmental Protection Agency (EPA 83587201). Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the supporting agencies. The views expressed in this paper are those of the authors and do not necessarily represent the view or policies of the US Environmental Protection Agency. Further, the US government does not endorse the purchase of any commercial products or services mentioned in the publication.

DATA AVAILABILITY STATEMENT

The data that support the findings of this paper are openly available in the GitHub repository at https://github.com/lit777/csBART.

ORCID

Chanmin Kim https://orcid.org/0000-0002-2588-6704
Mauricio Tec https://orcid.org/0000-0002-1853-5842
Corwin Zigler https://orcid.org/0000-0002-9611-5894

REFERENCES

Breiman, L. (2001) Random forests. Machine Learning, 45(1), 5–32.
Chipman, H.A., George, E.I., & McCulloch, R.E. (2010) BART: Bayesian additive regression trees. The Annals of Applied Statistics, 4(1), 266–298.
Häggström, J. (2018) Data-driven confounder selection via Markov and Bayesian networks. Biometrics, 74(2), 389–398.
Hahn, P.R., Murray, J.S. & Carvalho, C.M. (2020) Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects (with discussion). Bayesian Analysis, 15(3), 965–1056.
Hastie, T., & Tibshirani, R. (2000) Bayesian backfitting (with comments and a rejoinder by the authors). Statistical Science, 15(3), 196–223.
Henneman, L.R., Choirat, C., Ivey, C., Cummiskey, K. & Zigler, C.M. (2019) Characterizing population exposure to coal emissions sources in the United States using the HyADS model. Atmospheric Environment, 203, 271–280.
Hill, J.L. (2011) Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics, 20(1), 217–240.
Kapelner, A. & Bleich, J. (2016) bartmachine: Machine learning with Bayesian additive regression trees. Journal of Statistical Software, 70(4), 1–40.
Kim, C. (2022) Bayesian additive regression trees in spatial data analysis with sparse observations. *Journal of Statistical Computation and Simulation*, 92(15), 3275–3300.

Kim, C., Henneman, L.R., Choirat, C. & Zigler, C.M. (2020) Health effects of power plant emissions through ambient air quality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(4), 1677–1703.

Lefebvre, G., Delaney, J.A. & McClelland, R.L. (2014) Extending the Bayesian adjustment for confounding algorithm to binary treatment covariates to estimate the effect of smoking on carotid intima-media thickness: the multi-ethnic study of atherosclerosis. *Statistics in Medicine*, 33(16), 2797–2813.

Linero, A.R. (2018) Bayesian regression trees for high-dimensional prediction and variable selection. *Journal of the American Statistical Association*, 113(522), 626–636.

Rosenbaum, P.R. & Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

Rubin, D.B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688.

Shi, C., Blei, D. & Veitch, V. (2019) Adapting neural networks for the estimation of treatment effects. *Advances in Neural Information Processing Systems*, Vol. 32, Vancouver, Canada.

Shortreed, S.M. & Ertefaie, A. (2017) Outcome-adaptive lasso: variable selection for causal inference. *Biometrics*, 73(4), 1111–1122.

Tec, M., Scott, J. & Zigler, C. (2023) Weather2vec: Representation learning for causal inference with non-local confounding in air pollution and climate studies. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37, Washington, DC, USA.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

VanderWeele, T.J. (2019) Principles of confounder selection. *European Journal of Epidemiology*, 34(3), 211–219.

Wang, C., Dominici, F., Parmigiani, G. & Zigler, C.M. (2015) Accounting for uncertainty in confounder and effect modifier selection when estimating average causal effects in generalized linear models. *Biometrics*, 71(3), 654–665.

Wang, C., Parmigiani, G. & Dominici, F. (2012) Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics*, 68(3), 661–671.

Wilson, A. & Reich, B.J. (2014) Confounder selection via penalized credible regions. *Biometrics*, 70(4), 852–861.

Zigler, C.M. & Dominici, F. (2014) Uncertainty in propensity score estimation: Bayesian methods for variable selection and model-averaged causal effects. *Journal of the American Statistical Association*, 109(505), 95–107.

**SUPPORTING INFORMATION**

Web Appendix A–H referenced in Sections 2–6 are available with this paper at the Biometrics website on Wiley Online Library. R code files for the simulation studies and data analysis are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Kim, C., Tec, M. & Zigler, C. (2023) Bayesian nonparametric adjustment of confounding. *Biometrics*, 79, 3252–3265. [https://doi.org/10.1111/biom.13833](https://doi.org/10.1111/biom.13833)