ESTIMATING THE HEALTH EFFECTS OF ENVIRONMENTAL MIXTURES USING BAYESIAN SEMIPARAMETRIC REGRESSION AND SPARSITY INDUCING PRIORS

By Joseph Antonelli[1], Maitreyi Mazumdar[2,*], David Bellinger[2,†], David Christiani[3,‡], Robert Wright[4] and Brent Coull[3,§]

[1]*Department of Statistics, University of Florida, jantonelli@ufl.edu*

[2]*Harvard Medical School,* [*]*maitreyi.mazumdar@childrens.harvard.edu;* [†]*david.bellinger@childrens.harvard.edu*

[3]*Harvard T.H. Chan School of Public Health,* [‡]*dchris@hsph.harvard.edu;* [§]*bcoull@hsph.harvard.edu*

[4]*Icahn School of Medicine at Mount Sinai, robert.wright@mssm.edu*

Humans are routinely exposed to mixtures of chemical and other environmental factors, making the quantification of health effects associated with environmental mixtures a critical goal for establishing environmental policy sufficiently protective of human health. The quantification of the effects of exposure to an environmental mixture poses several statistical challenges. It is often the case that exposure to multiple pollutants interact with each other to affect an outcome. Further, the exposure-response relationship between an outcome and some exposures, such as some metals, can exhibit complex, nonlinear forms, since some exposures can be beneficial and detrimental at different ranges of exposure. To estimate the health effects of complex mixtures, we propose a flexible Bayesian approach that allows exposures to interact with each other and have nonlinear relationships with the outcome. We induce sparsity using multivariate spike and slab priors to determine which exposures are associated with the outcome and which exposures interact with each other. The proposed approach is interpretable, as we can use the posterior probabilities of inclusion into the model to identify pollutants that interact with each other. We utilize our approach to study the impact of exposure to metals on child neurodevelopment in Bangladesh and find a nonlinear, interactive relationship between arsenic and manganese.

1. Introduction. The study of the health impacts of environmental mixtures, including both chemical and nonchemical stressors, on human health is a research priority in the environmental health sciences (Braun et al. (2016), Carlin et al. (2013), Taylor et al. (2016)). In the past, the majority of health effect studies upon which environmental regulation is based estimate the independent association between an outcome and a given exposure, either marginally or while controlling for co-exposures. However, exposure to mixtures of chemicals is a real world scenario, and it has been argued that chemical mixture effects are possible even when each chemical exposure is below its no observable adverse effect level (Kortenkamp et al. (2007)).

Accordingly, there has recently been a renewed focus on the development of statistical methods for the analysis of the effects of chemical mixtures, or, more generally, environmental mixtures defined to include both chemical and nonchemical (such as diet or psychosocial factors) stressors on health. Recent work on this topic has recognized that investigations on these effects may focus on a variety of research questions of interest, including the identification of individual exposures driving this effect, identification of interactions among chemicals on health and the estimation of the cumulative health effects of exposure to multiple agents. Recent reviews of relevant statistical methods useful for analyzing the resulting data have

highlighted that the choice of statistical method will depend on the primary objective of the study (Braun et al. (2016)), and recent work in this area has developed both Bayesian and frequentist methods that implement forms of regularization, variable selection, dimension reduction, statistical learning or smoothing in various combinations to address these questions in the presence of correlated exposures (see Lazarevic et al. (2019) for an excellent review). Notably, there is a gap in this literature in terms of methods that directly quantify the strength of evidence supporting interactions among individual chemicals on a health outcome in the data. For instance, Bayesian kernel machine regression (BKMR; Bobb et al. (2015), Bobb et al. (2018)) can estimate cumulative health effects of exposure to multiple agents, provide variable importance scores for each individual chemical and estimate selected summary statistics that reflect interactions among some chemicals at given levels of others. However, it does not yield overall variable importance scores for interactions among chemicals within the larger mixture of exposures.

Other approaches allow for interactions between exposures, perhaps with some variable selection technique applied, but these approaches are often based on models assuming simple, linear associations between exposures or other strong parametric assumptions. It has been recently shown that in some settings, such as in studies on the impacts of exposure to metal mixtures on neurodevelopment in children, the exposure-response relationship between the multivariate exposure of interest and a health outcome can be a complex, nonlinear and non-additive function (Bobb et al. (2015), Henn, Coull and Wright (2014), Henn et al. (2010)). Therefore, there is a need for statistical approaches that can handle complex interactions among a large number of exposures. Even in settings with a relatively small number of exposures, the number of potential interactions can become large, making some form of variable selection or dimension reduction necessary. In this paper we provide a Bayesian approach to this problem that allows for nonlinear effects on the outcome, can identify important exposures among a large mixture and can identify important interactions among the elements of a mixture.

1.1. *Review of statistical literature.* There exists a vast statistical literature on high-dimensional regression models, including approaches to reduce the dimension of the covariate space. The LASSO (Tibshirani (1996)) utilizes an L1 penalty on the absolute value of regression coefficients in a regression model which effectively selects which variables should be included in the model by setting some coefficients equal to zero. Many approaches have built upon this same methodology (Zou (2006), Zou and Hastie (2005)); however, these all typically rely on relatively simple, parametric models. More recently, a number of papers (Bien, Taylor and Tibshirani (2013), Hao, Feng and Zhang (2014), Lim and Hastie (2015), Zhao, Rocha and Yu (2009)) have adopted these techniques to identify interactions, although these approaches are only applicable to linear models with pairwise interactions. While similar ideas for nonlinear models were used by Radchenko and James (2010), these analyses were restricted to pairwise interactions. Moreover, quantification of estimation uncertainty can be quite challenging in high dimensions.

A number of approaches have been introduced that embed variable selection within nonlinear regression models. Shively, Kohn and Wood (1999) utilized variable selection in Bayesian nonparametric regression based on an integrated Wiener process for each covariate in the model and then extended these ideas in Wood et al. (2002) to nonadditive models that allowed for interactions between covariates. They utilized the BIC approximation to the marginal likelihood in order to calculate posterior model probabilities which requires enumerating all potential models. Even when the number of covariates in the model is not prohibitively large, inclusion of just two-way interactions can lead to a massive number of possible models. Reich, Storlie and Bondell (2009) utilized Gaussian process priors for all main effects and

two-way interactions within a regression model and performed variable selection in order to assess which covariates are important predictors of the outcome. They do not allow for explicit identification of higher-order interactions, as all remaining interactions are absorbed into a function of the covariates that accounts for higher-order interactions. Yau, Kohn and Wood (2003) and Scheipl, Fahrmeir and Kneib (2012) utilized variable selection in an additive regression model framework with interaction terms within a Bayesian framework, though one must specify the terms that enter the model a priori and there can be a large number of potential terms to consider. Qamar and Tokdar (2014) used Gaussian processes to separately model subsets of the covariate space and used stochastic search variable selection to identify which covariates should enter into a given Gaussian process function, implying that they interact with each other. While this approach can be used to capture and identify complex interactions, it relies on Gaussian processes and therefore does not scale to larger data sets.

Complex models that allow for nonlinearities and interactions are becoming popular in the machine learning literature, with Gaussian process (or kernel machine) regression (O'Hagan (1978)), support vector machines (Cristianini and Shawe-Taylor (2000)), neural networks (Bengio et al. (2003)), Bayesian additive regression trees (BART, Chipman, George and Mc-Culloch (2010)) and random forests (Breiman (2001)) being just a few examples. Bobb et al. (2015) used Bayesian kernel machine regression to quantify the health effects of an environmental mixture. This approach imposed sparsity within the Gaussian process to flexibly model multiple exposures. While these approaches and their many extensions can provide accurate predictions in the presence of complex relationships among exposures, they typically only provide a variable importance measure for each feature included in the model and do not allow for formal inference on higher-order effects. An important question remains how to estimate the joint effect of a set of exposures that directly provides inference on interactions while scaling to larger data sets without requiring a search over such a large space of potential models involving higher-order interactions. Further, full Markov chain Monte Carlo (MCMC) inference in kernel machine regression does not scale with the sample size, limiting its applicability in large sample settings. While one can potentially use approximate Bayesian methods, such as variational inference to simplify computation, these methods are only approximate and can yield biased estimators for some parameters (Guo et al. (2016), Miller, Foti and Adams (2016)).

1.2. *Our contribution.* In this paper we introduce a Bayesian semiparametric regression approach that addresses several of the aforementioned issues as it allows for nonlinear interactions between variables, identifies interactions between variables and scales to larger data sets than those handled by traditional Gaussian process regression. We will introduce sparsity via spike and slab priors (George and McCulloch (1993), Mitchell and Beauchamp (1988)) within a semiparametric regression framework that reduces the dimension of the parameter space while yielding inference on interactions among individual exposures in the mixture. Section 2 introduces the motivating data set looking at exposure to metals in Bangladesh. Section 3 introduces our model formulation and discusses various estimation issues. Section 4 discusses some results about our prior distribution. Section 5 illustrates the proposed approach through simulated data and compares it with existing approaches. Section 6 estimates the health effects of environmental exposures on neurodevelopment in Bangladeshi children. We will conclude with some final remarks in Section 7.

2. Chemical mixtures and neurodevelopment in Bangladesh. To study the health impact of chemical mixtures on neurodevelopment, we examine a study of 375 children in Bangladesh whose mothers were potentially exposed to a variety of chemical pollutants during pregnancy. The outcome of interest is the z-score of the motor composite score (MCS),

a summary measure of a child's psychomotor development derived from the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) (Bayley (2006)). The exposure measures are log-transformed values of arsenic (As), manganese (Mn), and lead (Pb) measured in umbilical cord blood which is taken to represent prenatal exposure to these metals. As potential confounders of the neurodevelopment-exposure associations, we will also control for the following covariates: gender, age at the time of neurodevelopment assessment, mother's education, mother's IQ, an indicator for which clinic the child visited and the HOME score, which is a proxy for socioeconomic status. Previous studies have examined these data (Bobb et al. (2015), Valeri et al. (2017)) and found harmful associations between the exposures and the MCS score. Further, these studies have suggested a possible interaction between arsenic and manganese, though it was unclear whether this was simply random variation or a true signal. The assessment of interactions within environmental mixture data is an important problem (Henn, Coull and Wright (2014)), as the identification of exposures that interact with one another can help identify groups of exposures, or exposure profiles, that have particularly strong effects on health. The ability to identify them and assign measures of uncertainty around the existence of interactions is important and will represent an important aspect of our proposed approach.

2.1. *Preliminary analyses.* Before examining any associations with the outcome, we first look at the Pearson correlation among the log exposure values. Manganese and lead are uncorrelated, manganese and arsenic have a correlation of 0.58 and arsenic and lead have a correlation of -0.37. Next, we evaluate marginal associations between the exposures and outcome, unadjusted for any additional covariates or co-exposures. Table 1 shows these results. We see that all three exposures have significant marginal correlations with the MCS score. We also evaluate whether a linear or nonlinear relationship between the MCS score, and the exposures was favored by the data by looking at AIC values. The nonlinear AIC value is obtained by regressing the MCS score against a three degree of freedom spline basis representation of the exposures. We see that for arsenic the nonlinear model is favored, while the linear model is favored for the other two exposures. Arsenic concentrations are positively associated with MCS score, which could possibly be due to unmeasured confounding by, say, nutrients, due to the fact that a significant exposure pathway of arsenic exposure in this population is diet.

Next, to highlight the importance of the new methodology, we propose to address the presence of interactions between environmental exposures, and we analyze the data using current state of the art approaches for the analysis of chemical mixtures. Specifically, we use the Bayesian Kernel Machine Regression approach (BKMR, Bobb et al. (2015)) which is now widely used within the environmental epidemiological literature (Braun (2017), Harley et al. (2017)). BKMR allows for nonlinear relationships, interactions among exposures and adjusts for additional covariates measured in the population. We use the bkmr package in R, generating 25,000 Markov chain Monte Carlo (MCMC) scans and the default prior specification in

TABLE 1

Marginal relationship between the exposures and the MCS score in the Bangladesh data. Linear AIC represents the AIC of the linear model between the MCS score and the exposures, while nonlinear AIC represents the AIC of a linear model between the MCS score and a three degree of freedom spline basis representation of the exposures

Exposure	Correlation with outcome	Linear AIC	Nonlinear AIC
Arsenic	0.46 (0.37, 0.53)	972.67	966.93
Lead	-0.17 (-0.27, -0.07)	1050.05	1053.51
Manganese	-0.39 (-0.47, -0.3)	999.67	1001.81

TABLE 2
Posterior inclusion probabilities for the three exposures analyzed in the Bangladesh data

Exposure	Posterior inclusion probability
Pb	0.69
Mn	0.74
As	0.81

the package. Table 2 shows the posterior inclusion probabilities for the three metal exposures generated by this model.

These measures quantify the amount of uncertainty regarding the importance of each of the three exposures in predicting MCS score but do not reflect whether or not the three exposures interact in their effects on the outcome. One could get posterior inclusion probabilities such as these if all three exposures have distinct, important main effects or if there is a three-way interaction between them. To begin to address this question, one can visualize the estimated exposure response curves between these exposures. We visually inspect the relationship that manganese and arsenic have on the MCS score. In particular, Figure 1 shows the relationship between manganese and the MCS score while fixing arsenic at different quantile levels (0.1, 0.5 and 0.9). If there is an interaction between these two variables then these curves would differ across the different quantiles of arsenic, whereas if there were no interaction, then the shape of the three curves would not differ. We see that there is some variability in the shape of the curves, particularly for the 0.5 quantile, which has an inverted U shape. This suggests the possibility of an interaction between manganese and arsenic; however, this can not be concluded definitively given the combination of the posterior inclusion probabilities and the exposure-response curves. This shows the importance of new statistical methodology to identify these interactions

3. Model formulation. Throughout, we assume that we observe $D_i = (Y_i, X_i, C_i)$ for $i, \ldots, n$, where n is the sample size of the observed data, Y_i is the outcome, X_i is a p-dimensional vector of exposure variables and C_i is an m-dimensional vector of covariates for
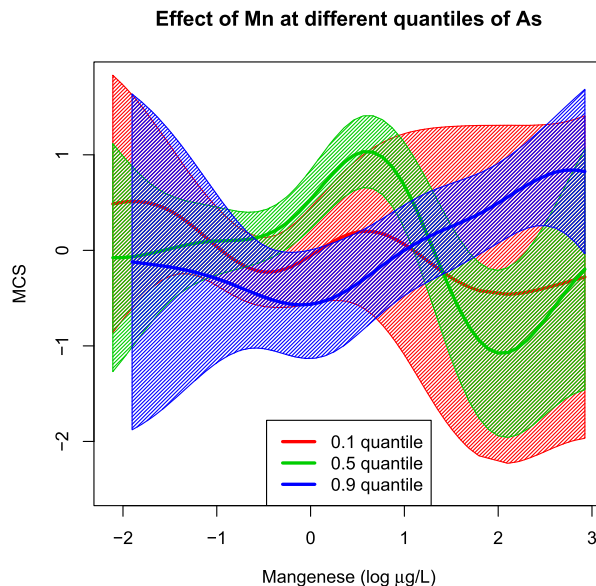
Effect of Mn at different quantiles of As

FIG. 1. *Exposure response curve between manganese and MCS score at different quantiles of arsenic.*

subject i. Building on the construction of Wood (2006), we assume that we have low rank bases that represent smooth functions for each of the exposures. For instance, the main effect of X_1 and an interaction effect between X_1 and X_2 could be modeled as follows:

$$f(X_1) = \widetilde{X}_1 \boldsymbol{\beta}_1 \qquad f(X_1, X_2) = \widetilde{X}_1 \boldsymbol{\beta}_1 + \widetilde{X}_2 \boldsymbol{\beta}_2 + \widetilde{X}_{12} \boldsymbol{\beta}_{12},$$

respectively, where $\widetilde{X}_1 = [g_1(X_1), \ldots, g_d(X_1)]$ and $\widetilde{X}_2 = [q_1(X_2), \ldots, q_d(X_2)]$ represent d-dimensional basis function expansions of X_1 and X_2, respectively. $\widetilde{X}_{12} = [g_1(X_1)q_1(X_2), g_1(X_1)q_2(X_2), \ldots, g_d(X_1)q_d(X_2)]$ represents a d^2-dimensional basis expansion of the interaction between X_1 and X_2. In this work, we specify $g()$ and $q()$ to be natural spline basis functions of X_1 and X_2, respectively. For simplicity, we let the degrees of freedom for each exposure all be set to d, though this can easily be extended to allow the degrees of freedom to vary across exposures. Importantly, the construction above enforces that main effect terms are a subset of the terms included in the interaction terms. As we include higher-order interactions, we impose the constraint that all lower level interactions between the exposures in that interaction term are included in the basis expansion. To ease the presentation of our model, we use subscripts to denote the order of the interaction to which a coefficient applies. For instance, $\boldsymbol{\beta}_{12}$ refers to parameters associated with two-way interaction terms between X_1 and X_2 (excluding main effect terms), while $\boldsymbol{\beta}_1$ refers to the main effect parameters for X_1. This process could be trivially extended to any order interaction up to a p-way interaction which is the maximum possible interaction in our model. We will use the following model formulation:

$$Y_i \sim \text{Normal}\big(f(X_i) + C_i \boldsymbol{\beta}_c, \sigma^2\big),$$

(3.1)
$$f(X_i) = \sum_{h=1}^{k} f^{(h)}(X_i),$$

$$f^{(h)}(X_i) = \sum_{j_1=1}^{p} \widetilde{X}_{ij_1} \boldsymbol{\beta}_{j_1}^{(h)} + \sum_{j_1=2}^{p} \sum_{j_2 < j_1} \widetilde{X}_{ij_1 j_2} \boldsymbol{\beta}_{j_1 j_2}^{(h)} + \cdots,$$

where the dots in the final line indicate that the summation extends all the way through p-way interactions. Our main goal is the estimation of $f(X_i)$ and the identification of important exposures within X_i. We are representing $f(X_i)$ as the sum of k different functions $f^{(h)}(X_i)$, where k is a number chosen to be sufficiently large so that all exposure effects can be captured by the model. This model is both over parameterized and not identifiable without any additional constraints. This over parameterization occurs because the model allows for all possible interactions of any order which includes far too many parameters to reasonably estimate. It is not identifiable because each of the k functions has the same functional form and the regression coefficients for a particular function $f^{(h)}(X_i)$ are only identifiable up to a constant. To reduce the dimension of the parameter space and to identify parameters of interest, we will adopt the following prior specification for the regression coefficients:

$$P(\boldsymbol{\beta}_S^{(h)} | \boldsymbol{\zeta}) = \left(1 - \prod_{j \in S} \zeta_{jh}\right) \delta_0 + \left(\prod_{j \in S} \zeta_{jh}\right) \psi_1(\boldsymbol{\beta}_S^{(h)}),$$

where S is some subset of $\{1, 2, \ldots, p\}$,

$$P(\zeta_{jh}) = \tau_h^{\zeta_{jh}} (1 - \tau_h)^{1 - \zeta_{jh}} \mathbf{1}(A_h \not\subset A_m \forall m \text{ or } A_h = \{\}),$$

where $A_h = \{j : \zeta_{jh} = 1\}$.

Here, $\boldsymbol{\zeta} = \{\zeta_{jh}\}$ is a matrix of binary indicators of whether the exposure j is included in the hth function in the model. One of our main goals is to identify important exposure effects and

interactions between exposures, and these can be seen immediately from $\{\zeta_{jh}\}$ which tells us both which exposures and which interactions are included in a particular function. We use spike and slab priors to eliminate exposures from the model, and in this paper we consider the spike to be a point mass at $\mathbf{0}$, while the slab, $\psi_1()$, is a multivariate normal distribution centered at $\mathbf{0}$ with covariance $\Sigma_{\boldsymbol{\beta}}$. We let $\Sigma_{\boldsymbol{\beta}}$ be a diagonal matrix with $\sigma^2 \sigma_{\boldsymbol{\beta}}^2$ on the diagonals.

The prior distribution for ζ_{jh} is a Bernoulli distribution with parameter τ_h. This prior is accompanied by an indicator function, $\mathbf{1}(A_h \not\subset A_m \ \forall m \text{ or } A_h = \{\})$, which reflects the fact that each function either contains information from a unique set of exposures or is removed from the model completely. If a function only contained exposures that were all included in another function, then it would be providing redundant information and it is not needed, so we preclude this from happening by introducing this indicator function. More details on this indicator function and its implications for sampling can be found in the following section. Full details of posterior sampling from this model can be found in Section 3 of the Supplementary Material (Antonelli et al. (2020)).

3.1. *Identifiability of variable inclusion parameters.* Two scientific goals in the study of environmental mixtures are to identify which exposures are associated with the outcome and to identify interactions among exposures. To this end, our prior encodes the condition that, for any two sets of variables A_h and A_m included in functions h and m, either $A_h \not\subset A_m$ and $A_m \not\subset A_h$ or one of the two is an empty set. An illustrative example of why this condition is necessary is a scenario in which all elements of $\boldsymbol{\zeta}$ are zero with the exception of ζ_{11} and ζ_{12}. In this scenario the first exposure has a nonzero effect in both the first and second functions of the model, that is, $\boldsymbol{\beta}_1^{(1)}$ and $\boldsymbol{\beta}_1^{(2)}$ are both nonzero. Regardless of the values of $\boldsymbol{\beta}_1^{(1)}$ and $\boldsymbol{\beta}_1^{(2)}$, this model could be equivalently expressed with one function, rather than two. For instance, while $\boldsymbol{\beta}_1^{(1)} = -\boldsymbol{\beta}_1^{(2)}$ represents the same model as $\boldsymbol{\beta}_1^{(1)} = \boldsymbol{\beta}_1^{(2)} = \mathbf{0}$, inference regarding $\boldsymbol{\zeta}$ would differ across these two models. In the first scenario it would appear that the first exposure is associated with the outcome, even, if in truth, it is not. Our conditions avoid this scenario by ensuring that there are no functions that simply contain subsets of the exposures found in another function, and, therefore, $\boldsymbol{\zeta}$ will provide a meaningful and interpretable measure of the importance of each exposure and potential interaction.

Another issue affecting our model selection indicators is that low-order interactions are subsets of the models defined by higher-order interactions. For example, if the true model is such that there exists a main effect of both X_1 and X_2 but no interaction between them, we could mistakenly estimate that there is an interaction between exposure 1 and 2 simply due to the main effects. (Wood (2006), Reich, Storlie and Bondell (2009)) addressed this issue in similar contexts by utilizing constraints that force the two-way interaction between X_1 and X_2 to only include information that is orthogonal to the main effects of the individual exposures. In our context, this is not so much an issue of identifiability but rather an issue with MCMC chains getting stuck in local modes. When exploring the model space via MCMC, we could accept a move that includes the two-way interaction, which is a local mode, and never be able to leave this model, even if the simpler model is the true model.

To avoid this scenario, we simply need to search the model space in a particular manner. We will illustrate this in the case of a two-way interaction as follows:

$$\boldsymbol{\zeta}_{p_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\zeta}_{p_2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\zeta}_{p_3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Imagine that, at a given iteration of an MCMC, $\boldsymbol{\zeta} = \boldsymbol{\zeta}_{p_1}$, and we look to update ζ_{21} which indicates whether exposure 2 enters into function 1. If we do this naively, we simply could compare models $\boldsymbol{\zeta}_{p_1}$ and $\boldsymbol{\zeta}_{p_2}$, which would result in a choice of $\boldsymbol{\zeta}_{p_2}$, even if there were

only main effects of X_1 and X_2 and no interaction between them. If, however, we update ζ_{22} at the same time, we also evaluate model $\boldsymbol{\zeta}_{p_3}$. If the true relationship is a main effect of the two exposures, then model $\boldsymbol{\zeta}_{p_3}$ will be favored by the data, and we will estimate $\boldsymbol{\zeta}$ accurately. If the truth is a two-way interaction between X_1 and X_2, then $\boldsymbol{\zeta}_{p_2}$ will be favored by the data. This setting generalizes to higher-order interactions and illustrates how we can avoid misleading conclusions about interactions. Whenever we propose a move to a higher-order interaction, we must also evaluate all lower-order models that could lead to incorrectly choosing the higher-order model. While proposing moves in this manner keeps the MCMC from getting stuck at incorrect models, it needs to be carefully constructed to maintain a valid sampler. In Section 6 of the Supplementary Material (Antonelli et al. (2020)), we prove that this strategy can be tailored in such a way that the sampler is reversible and, therefore, maintains the correct target distribution.

3.2. *Selection of spline degrees of freedom.* An important tuning parameter in the proposed model is the number of degrees of freedom, d, used to flexibly model the effects of the p exposures. Throughout, we will make the simplifying assumption that the degrees of freedom is the same for all exposures, though this can be relaxed to allow for differing amounts of flexibility across the marginal functions. We propose the use of the WAIC model selection criterion (Watanabe (2010), Gelman et al. (2004)) to select d. If we let $s = 1, \ldots, S$ index samples from the posterior distribution of the model parameters and let $\boldsymbol{\theta}$ represent all parameters in the model, the WAIC is defined as

$$WAIC = -2\left(\sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(Y_i|\boldsymbol{\theta}^s)\right) - \sum_{i=1}^{n} \mathrm{var}(\log p(Y_i|\boldsymbol{\theta}))\right).$$

The WAIC approximates leave one out cross validation and, therefore, provides an assessment of model fit based on out of sample prediction performance. We first draw samples from the posterior distributions of the model parameters under a range of values for d and then select the model that minimizes WAIC. While this is not a fully Bayesian approach to choosing d, we will see in Section 5 that this approach still leads to credible intervals with good frequentist properties.

3.3. *Prior choice.* There are a number of hyperparameters that need to be chosen to complete the model specification. First, one must choose an appropriate value for $\boldsymbol{\Sigma}_\beta$, which controls the amount of variability and shrinkage of $\boldsymbol{\beta}_S^{(h)}$ for those coefficients originating from the slab component of the prior. We will make the simplifying assumption that $\boldsymbol{\Sigma}_\beta$ is a diagonal matrix with $\sigma^2\sigma_\beta^2$ on the diagonals. It is well known that variable selection is sensitive to the choice of this parameter (Reich, Storlie and Bondell (2009)), so it is crucially important to select an appropriate value. We propose the use of an empirical Bayes strategy where we estimate σ_β^2 and then obtain posterior samples conditional on this value. A description of the algorithm for finding the empirical Bayes estimate is available in Section 4 of the Supplementary Material (Antonelli et al. (2020)). One could alternatively put a hyper prior on σ_β^2 for a fully Bayesian analysis, though we have found that the empirical Bayes strategy works better in practice, so we restrict attention to that case.

We also must select a value of τ_h, the prior probability of an exposure being included in function h. We will assume that $\tau_h \sim \mathrm{Beta}(M, \gamma)$ and set M and γ to fixed values. One can choose values to induce a desired level of sparsity in the model; however, in practice, the appropriate degree of sparsity is typically not known and, therefore, in order to accommodate higher-dimensional data, we choose values that scale with the number of exposures. We will set $\gamma = p$ which allows the prior amount of sparsity to increase as the number

of exposures increases. This strategy closely resembles other existing approaches in high-dimensional Bayesian models, such as those proposed by Zhou et al. (2015) and Ročková and George (2018). We will discuss the implications of scaling the prior with p in Section 4.

The only remaining prior distributions to specify are for $\boldsymbol{\beta}_C$ and σ^2. For both parameters we specify diffuse, conjugate priors. In the case of $\boldsymbol{\beta}_C$, this means independent normal prior distributions with very large variances. For the residual variance σ^2 we specify an inverse gamma distribution with both parameters set to 0.001.

3.4. *Lower bound on slab variance.*

One issue with the aforementioned empirical Bayes estimation of $\sigma^2_{\boldsymbol{\beta}}$ is that it will be estimated to be very small if there is little to no association between the exposures and the outcome. This result can lead to problems in the estimation of $\boldsymbol{\zeta}$. If the prior variance for the slab is very small, then it is almost indistinguishable from the spike, since in this case it approximates a point mass at zero. In this situation, updating ζ_{jh} becomes essentially a coin flip for all j and h which will lead to high posterior inclusion probabilities, even when there is no true association. To avoid this, we estimate a lower bound, above which we believe that any high posterior inclusion probabilities reflect true signals in the data.

To estimate this lower bound, we first permute the rows of Y, thereby breaking any associations between the exposures and the outcome. We run our MCMC algorithm for a small number of iterations on the permuted data and keep track of the probability that $\zeta_{jh} = 1$ when $\tau_h = 0.5$ for all h. We do this for a decreasing grid of potential values for $\sigma^2_{\boldsymbol{\beta}}$, and we take as our lower bound for $\sigma^2_{\boldsymbol{\beta}}$ the smallest value of $\sigma^2_{\boldsymbol{\beta}}$ such that a prechosen threshold is met. Examples of such criteria include the smallest value of $\sigma^2_{\boldsymbol{\beta}}$ such that the probability of including a main effect for any of the exposures is less than 0.25 or the smallest value such that the probability of any two-way interaction is less than 0.05. Alternatively, because one can calculate the average false discovery rate for each value of $\sigma^2_{\boldsymbol{\beta}}$, one can use this strategy to implement Bayesian counterparts to traditional multiple testing adjustments that control frequentist error rates, such as the false discovery rate. It is important to note that this false discovery rate control is conditional on a fixed value of τ_h, which we choose to be 0.5. In practice, we let τ_h be random and estimated from the data. Nonetheless, this procedure should provide a reasonable lower bound for $\sigma^2_{\boldsymbol{\beta}}$. If strict control of false discovery rate is required, then one could specify a truncated prior for τ_h such that it is never greater than the value of τ_h used for the calculation of the lower bound (i.e., 0.5). It is important to note that this procedure is very fast computationally, generally taking only a small fraction of the computation time as the full MCMC for one data set.

To illustrate how this approach works for a given analysis, we simulated data with no associations between the exposure and outcome. In this case, the empirical Bayes estimate of $\sigma^2_{\boldsymbol{\beta}}$ will be very small as it is roughly estimated to be the average squared value of the nonzero coefficients in $\boldsymbol{\beta}$. Figure 2 shows the posterior inclusion probabilities averaged over all exposures in this example as a function of $\sigma^2_{\boldsymbol{\beta}}$. The gray region represents the area where false positives start to arise because $\sigma^2_{\boldsymbol{\beta}}$ is too small, and we see that the empirical Bayes estimate is contained in this region. If we proceed with the empirical Bayes estimate, then we would conclude that all of the exposures had posterior inclusion probabilities near 0.5, incorrectly indicating they are important for predicting the outcome. Our permutation strategy to identifying a lower bound is able to find a value of $\sigma^2_{\boldsymbol{\beta}}$ that avoids high posterior inclusion probabilities for exposures with null associations with the outcome. In practice, we estimate both the lower bound and empirical Bayes estimate and proceed with the maximum of these two values which, in this simulated case, is the lower bound.
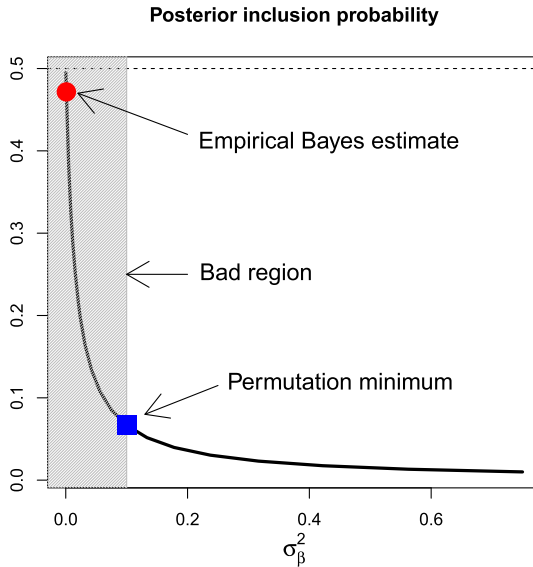
Posterior inclusion probability

FIG. 2.   *Illustration of empirical Bayes estimate and lower bound estimate of σ_β^2.*

4. Properties of prior distributions.   A primary goal in many analyses of the health effects of an environmental mixture is to identify interactions among exposures on risk. Therefore, it is important to understand the extent to which the proposed prior specification induces shrinkage on the interaction terms. Both τ_h and $\boldsymbol{\Sigma_\beta}$ dictate the amount of shrinkage. We utilize empirical Bayes to select $\boldsymbol{\Sigma_\beta}$. Therefore, here we focus on the role of τ_h which directly controls the probability that a jth order interaction is nonzero in the model. The prior for this parameter can be controlled through selection of M and γ. We examine the probability that a jth order interaction is zero by deriving the probability that the first j covariates are included in an interaction term, as this probability will be the same for any other interaction of the same order.

LEMMA 4.1.   *Assuming Model* 3.1, *the prior probability that there does not exist a function*, k' *in the model such that* $\zeta_{1k'} = \zeta_{2k'} = \cdots = \zeta_{jk'} = 1$ *and* $\zeta_{j+1k'} = \cdots = \zeta_{pk'} = 0$ *is*

$$(4.1) \qquad \left(1 - \frac{\Gamma(j+M)\Gamma(p+\gamma-j)}{\Gamma(p+\gamma+M)}\right)^k.$$

Section 5 of the Supplementary Material (Antonelli et al. (2020)) contains a proof of this result which is useful because it expresses the probability of an interaction of any order being zero as a function of M and γ. This result can guide the user in specifying values of M and γ that best represents prior beliefs. It has been noted in the literature that a desirable feature of the prior is to place more shrinkage on higher-order interactions (Gelman et al. (2008), Zhou et al. (2015)). Therefore, we use (4.1) to derive conditions that accomplish this type of shrinkage. Specifically, Theorem 4.2 provides a default choice of γ that specifies the amount of shrinkage on an interaction term to be an increasing function of the order of that term.

THEOREM 4.2.   *Assuming Model* 3.1, *such that* $\tau_h \sim \text{Beta}(M, \gamma)$, *then the probability in Lemma* 4.1 *increases as a function of j if* $\gamma \geq p + M - 1$.

Section 5 of the Supplementary Material (Antonelli et al. (2020)) contains a proof of this result, which shows that higher-order interactions are more aggressively forced out

of the model under either the standard Beta$(1, p)$ prior, which is often used in high-dimensional Bayesian models (Ročková and George (2018), Zhou et al. (2015)) or, if $\gamma \geq p + M - 1$, which shrinks higher-order interactions even more aggressively than this beta prior. In practice, a common prior choice for τ_h is Beta(M, p), with M being some constant that does not change with the number of exposures. Nonetheless, this result shows that one can achieve stronger penalization of higher-order interaction terms as long as we scale the prior with p.

5. Simulation study.
Here, we present results from a simulation study designed to assess the operating characteristics of the proposed approach in a variety of settings. In all cases we take the maximum of the empirical Bayes and lower bound estimates of σ_β^2. For each simulated data set, we fit the model for values of $d \in \{1, 2, 3, 4, 5, 6\}$ and run a Gibbs sampler for 10,000 iterations with two chains, keeping every eighth sample and discarding the first 2000 as burn-in. We present results for each value of d, the value that is selected based on the WAIC, the spike and slab GAM (ssGAM) approach of Scheipl, Fahrmeir and Kneib (2012) and for the Bayesian kernel machine regression (BKMR) approach of Bobb et al. (2015). We assess the performance of the various models by calculating the mean squared error (MSE) and 95% credible interval coverages of the true relationship between the exposures and the outcome, $f(X_1, \ldots, X_p)$, at all observed values of the exposure. We also assess the ability of the proposed model to identify nonzero exposure effects and interaction terms.

5.1. *Polynomial interactions.* First, we generate data with a sample size of $n = 200$ and generate $m = 10$ covariates, C, from independent standard normal distributions. We generate $p = 10$ exposures, X, from a multivariate normal distribution with mean μ_X. We will let the mean of X be a linear function of the covariates, that is, $\mu_X = C\alpha$. We let all pairwise correlations between the exposures be 0.6 which is the average pairwise correlation among the exposures in the NHANES data analyzed in Section 1 of the Supplementary Material (Antonelli et al. (2020)). The marginal variances are set to 1 for all exposures. We set the residual variance to $\sigma^2 = 1$. For the outcome, we use a polynomial model that contains both a linear and a nonlinear interaction term as follows:

$$Y = 0.7X_2X_3 + 0.6X_4^2X_5 + C\beta_C + \epsilon.$$

For exact values of α and β_C, see Section 2 of the Supplementary Material (Antonelli et al. (2020)). Figure 3 presents the results of the simulation study. The top left panel of Figure 3 shows the matrix of posterior probabilities that a pair of exposures interact with each other. Specifically, cell (i, j) of this heatmap shows the posterior probability that exposures i and j interact with each other in the model averaged over all simulations. Results show that the model correctly identifies the pairs (X_2, X_3) and (X_4, X_5) as interacting with each other. The top right panel shows the average 95% credible interval coverage of $f(X_i)$ for each approach considered. Our model, based on WAIC, achieves nearly the nominal coverage indicating that we are accounting for all sources of uncertainty. The bottom left panel highlights the ability of our procedure to estimate the marginal effect of X_4, which has a quadratic effect on the outcome, while holding constant the values of other exposures. The grey lines, which represent individual simulation estimates, reflect the true relationship, denoted by the red line. Finally, the bottom right panel of Figure 3 presents the mean squared error of the global effect estimates of $f(X_i)$. Results show that well chosen values of d, including the automated choice using WAIC, lead to the best performance among the methods considered.
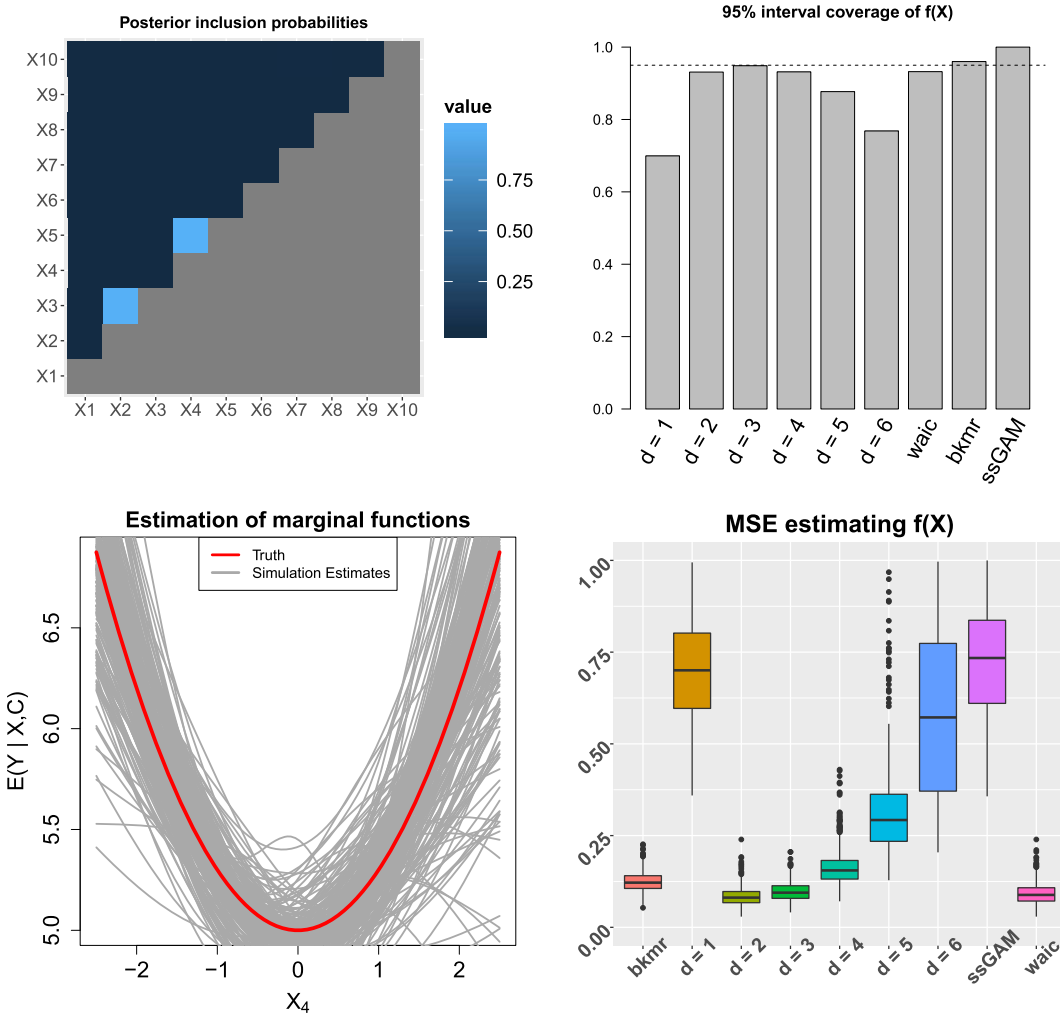
FIG. 3. *Results of the quadratic simulation. The top left panel shows the posterior probabilities of exposures interacting with each other for the model selected by the WAIC. The top right panel shows the 95% credible interval coverages of $f(X_i)$. The bottom left panel shows the estimated marginal functions of X_4 on Y for the model selected by the WAIC, where the red line represents the truth and the grey lines represent individual estimates from separate simulations. The bottom right panel shows the MSE in estimating $f(X)$ for all models considered.*

5.2. *High-dimensional data with three-way interaction.* Now, we present a scenario with $n = 10,000$ and $p = 100$ to highlight that our approach can handle large sample sizes and large exposure dimensions. We will generate the exposures from independent uniform random variables. We generate one additional covariate, C, from a standard normal distribution, and we set the residual variance to $\sigma^2 = 1$. The outcome model is closely related to one that was explored in Qamar and Tokdar (2014) and corresponds to a model that includes a three-way, nonlinear interaction as follows:

$$Y = 2.5\sin(\pi X_1 X_2) + 1.5\cos(\pi(X_3 X_4 + X_5)) + 2(X_6 - 0.5) + 2.5X_7 + \epsilon.$$

It is clear that the relationship between the exposures is nonlinear and nonadditive. We do not show results for BKMR due to the sample size limitation from using Gaussian processes. Further, there are a large number of exposures considered here, as there are 4950 possible two-way interactions and 161,700 possible three-way interactions. For this reason, we do not include ssGAM in this simulation as it requires specifying the form of the linear predictor,
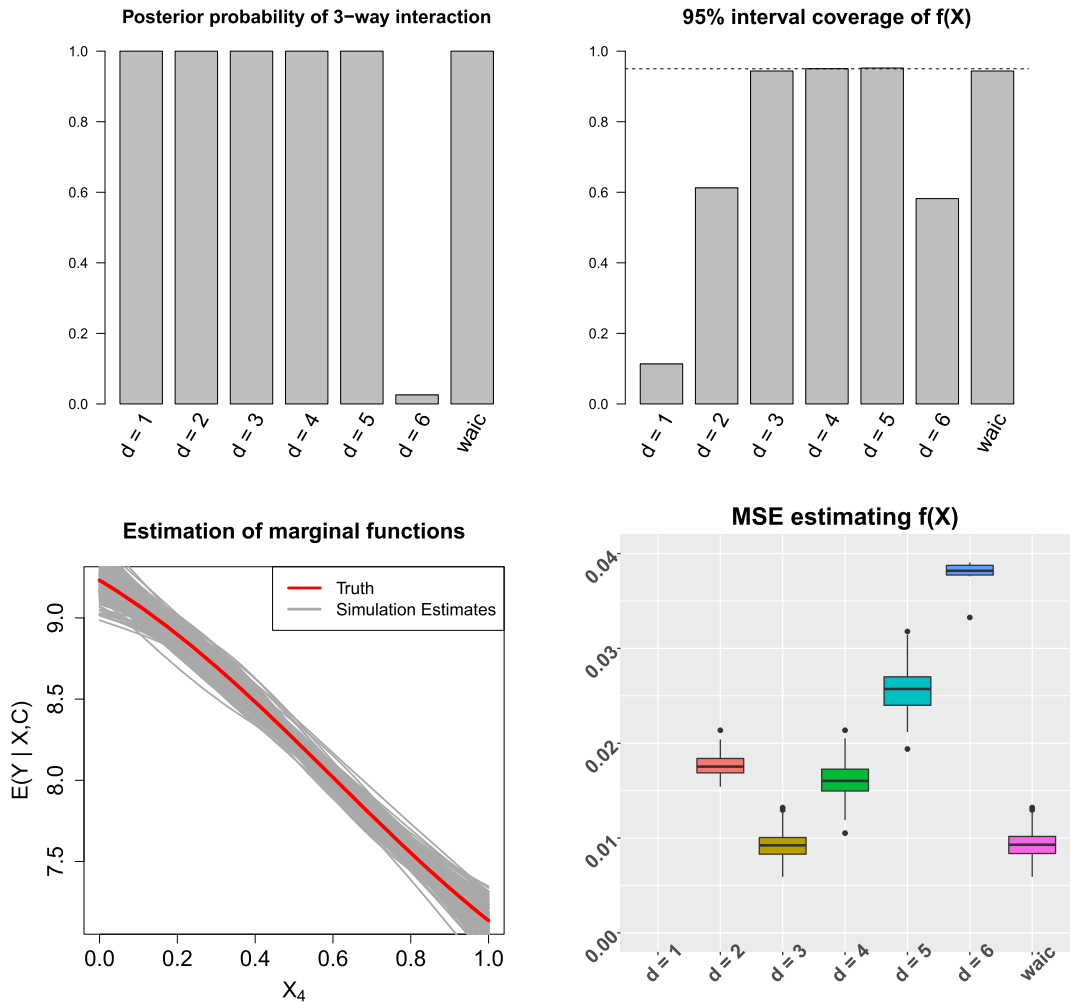
FIG. 4. *Results of the high-dimensional simulation. The top left panel shows the posterior probabilities of the three-way interaction between X_3, X_4 and X_5 for all models. The top right panel shows the 95% credible interval coverages of $f(X_i)$. The bottom left panel shows the estimated marginal functions of X_4 on Y for the model selected by the WAIC, where the red line represents the truth and the grey lines represent individual estimates from separate simulations. The bottom right panel shows the MSE in estimating $f(X)$ for all models considered. The MSE for $d = 1$ is too large to be seen on the plot.*

and there are too many terms to consider. Our approach considers interactions of any order, though it does not require prespecification of which terms to consider and aims to find the correct model among all possible interactions. Figure 4 shows results from this scenario. The top left panel shows the posterior probability of a three-way interaction between X_3, X_4 and X_5. This probability is 1 for all models except for $d = 6$, which does not capture this interaction. The top right panel shows the 95% credible interval coverages of $f(X_i)$. We see that the coverage is low for $d = 1$ and $d = 2$, which are not sufficiently flexible to capture the true function, while the coverage is low for $d = 6$ since it does not capture the three-way interaction. Importantly, the model chosen by WAIC achieves the nominal coverage level. The bottom left panel of Figure 4 shows the estimated marginal effect of X_4 on the outcome, fixing the other exposures. Results suggest that in almost all cases the method accurately estimates the true effect which is close to linear in the range of the data we are looking at. The bottom right panel of Figure 4 shows that the MSE of the model chosen by WAIC is

d	WAIC	Pb	Mn	As	Interaction between Mn and As
1	922.79	0.83	0.79	0.86	0.21
2	922.16	0.76	0.82	0.97	0.43
3	910.77	0.78	0.99	1.00	0.82
4	902.96	0.94	0.98	1.00	0.96
5	901.39	0.92	1.00	1.00	0.94

competitive, or better, than any individual choice of d, suggesting that WAIC is an effective method for choosing the degrees of freedom for the spline terms.

6. Analysis of Bangladesh data. Here, we analyze the data on metal mixtures and neurodevelopment introduced in Section 2. We applied our proposed approach to identify the complex relationship between the three metal exposures and neurodevelopment. We ran the Markov chain for 40,000 iterations with $M = 3$, $\gamma = p$, $a_0 = 0.001$, $b_0 = 0.001$ and a noninformative, normal prior on β_c. We ran the model for values of $d \in \{1, 2, 3, 4, 5\}$. Convergence of the MCMC was confirmed using trace plots and PSR values, and these results can be seen in Section 7 of the Supplementary Material (Antonelli et al. (2020)).

6.1. *The importance of the degrees of freedom.* Here, we evaluate the impact that the degrees of freedom d had on the analysis. Table 3 shows the WAIC values for each of the values of d considered, the corresponding main effect posterior inclusion probabilities for each exposure as well as the posterior inclusion probability for the two-way interaction between manganese and arsenic. The WAIC values strongly indicate that linear models are not appropriate for this data, as higher degrees of freedom lead to lower WAIC values and, therefore, better model fits. One of the most telling takeaways from Table 3 is the correspondence between WAIC values and the posterior inclusion probabilities for the two-way interaction between manganese and arsenic. Models with higher degrees of freedom lead to lower WAIC values and higher posterior inclusion probabilities for the two-way interaction. It appears that the lower values of d are not adequate for capturing the interaction between manganese and arsenic, and this leads to poorer fits to the data. The marginal inclusion probabilities for each exposure are more stable across d; however, these go up slightly from $d = 1$ to $d = 5$ as well. Overall, the WAIC indicates that $d = 5$ provides the best fit to the data, and we will therefore present results from this model from this point onward.

6.2. *Posterior inclusion probabilities.* One feature of the proposed approach is the ability to examine the posterior distribution of ζ in order to identify whether individual exposures are important for outcome prediction, whether there is evidence of an interaction between two metals or any higher-order interactions. Figure 5 shows the posterior probability of two-way interactions between all metals in the model and the marginal posterior inclusion probability for each metal exposure. Figure 5 provides strong evidence that each exposure is associated with a child's MCS score. This confirms the preliminary results of Section 2 that showed associations between each of the three exposures and the MCS score. A key distinction of the results here and those seen in Section 2 is that it also provides strong evidence of an interaction between Mn and As, as well as moderate evidence of an interaction between Pb and As. Not shown in the figure is the posterior probability of a three-way interaction among Pb, Mn and As, which was estimated to be 0.0005.
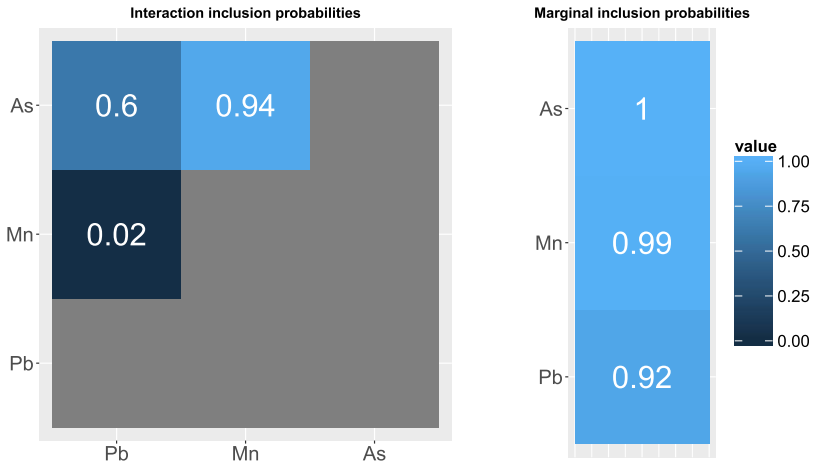
FIG. 5. *The left panel contains posterior inclusion probabilities for two-way interactions, and the right panel shows the marginal inclusion probabilities for each exposure.*

6.3. *Visualization of identified interactions.* Now that we have identified interactions in the model, we plot posterior predictive distributions to assess the form of these associations. Figure 6 plots the relationship between Mn and MCS for different values of As while fixing Pb at its median. At the median level of As, we estimate an inverted U-shaped effect of Mn on MCS scores. This matches results seen previously in a kernel machine regression analysis of these data Bobb et al. (2015) which provided graphical evidence of an interaction between Mn and As on the MCS score but didn't provide a measure of the strength of this interaction.

We also examine surface plots that show the exposure-response relationship between As, Mn and MCS across the range of As and Mn values, not just specific quantiles. Figure 7 shows both the posterior mean and pointwise standard deviations of this relationship. We see that the median levels of Mn and As are the most beneficial toward children in terms of their MCS score which corresponds to the hump in the middle panel of Figure 6. This is a somewhat unexpected result given that As is not beneficial, even at low levels, and could be due to residual confounding (Valeri et al. (2017)). The posterior standard deviations of this estimate indicate that the uncertainty around this surface is the smallest in the center, where most of the data is observed, and greater on the fringes of the observed data, where we are effectively extrapolating effects.
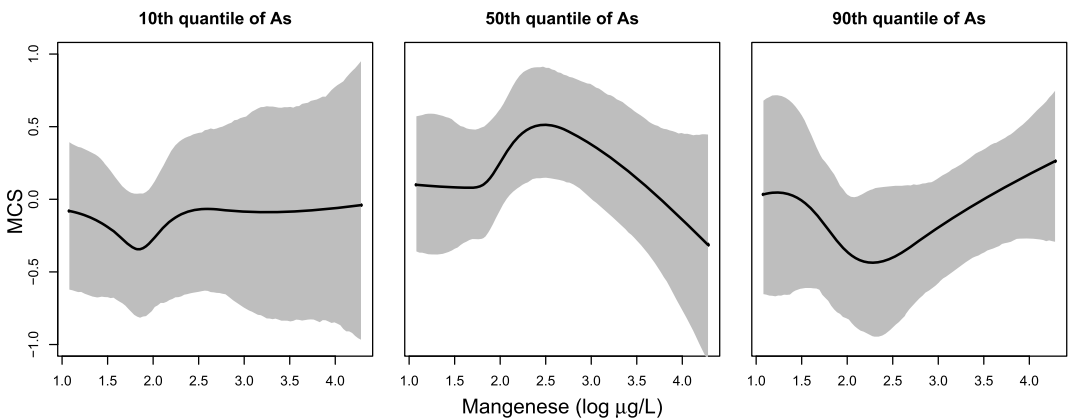

FIG. 6. *Posterior means and pointwise 95% credible intervals of the effect of manganese on MCS at different quantiles of arsenic while fixing lead at its median.*
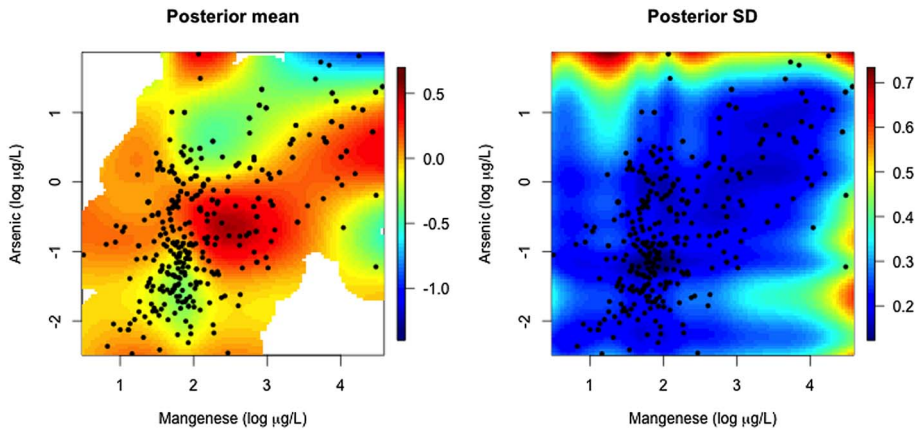
FIG. 7.   *Posterior means and standard deviation of the relationship between manganese and arsenic on MCS. The black dots represent the location of the observed data points.*

7. Discussion.   We have proposed a flexible method for analyzing complex, nonlinear associations between multiple exposures and a continuous outcome. Our method improves upon existing methodology for environmental mixtures in several ways. By not relying on Gaussian processes to introduce flexibility in the exposure-response relationship, our approach can handle larger data sets on the order of tens of thousands of subjects. The spike and slab formulation of our model also allows for formal inference on the presence of interactions among exposures, an important question in environmental epidemiology. Previous studies have used graphical tools and inference on summary statistics characterizing the effects of one pollutant given the levels of others to identify interactions between pollutants. While useful, these approaches do not provide uncertainty assessment of the interaction as a whole. Moreover, it can be difficult to visualize higher-order interactions. Once we have samples of the ζ matrix, we can easily perform inference on more complex features of the model such as higher-order interactions and whether or not exposures are more likely to enter the model together or separately. We have illustrated the strengths of this approach through simulated data and a study of the health effects of metal mixtures on child neurodevelopment in Bangladesh.

A natural question that arises in the application of our method is how does it perform on higher-dimensional mixtures in real-world settings. We evaluated a situation with both a large sample size and number of exposures in Section 5.2 of our simulation study and found that it performs well at identifying higher-order interactions. To further illustrate this point, we have analyzed a secondary data set from the National Health and Nutrition Examination Survey (NHANES), which has a larger sample size and more exposures than the Bangladesh data. The results of this study can be found in Section 1 of the Supplementary Material (Antonelli et al. (2020)).

Throughout the manuscript we have assumed that the relationship between the exposures and outcome is constant across levels of the covariate space C. In some cases, there is interest in understanding if the exposure effect varies across levels of particular covariates such as age or gender, which is frequently referred to as treatment effect heterogeneity. This could easily be explored using our model if the covariates are continuous. The covariates of interest can be placed with the exposures inside of the main $f()$ function, and our approach will be able to identify if any of these covariates modify the exposure effect. Identifying treatment effect heterogeneity in the presence of categorical covariates is not immediately applicable within our modeling framework, but it provides a topic of future research.

In the modeling framework we focused on a high-dimensional X and assumed that the number of covariates in C was relatively small, meaning we do not need to impose any

shrinkage on $\boldsymbol{\beta}_C$. If the number of covariates is large as well, then shrinkage or spike and slab priors would need to be adopted. If interest is in estimation of $f(\boldsymbol{X})$, this could lead to what is called regularization induced confounding (RIC, Hahn, Murray and Carvalho (2017), Hahn et al. (2018)). RIC occurs because shrinkage leads to biased estimates of the coefficients in $\boldsymbol{\beta}_C$, and this bias leads to confounded estimates of the exposure effects of interest. Future work would look to extend the ideas here to the more difficult setting of a high-dimensional $\boldsymbol{C}$ when RIC could affect inference of the effects of interest. This would require reducing the shrinkage of the elements in $\boldsymbol{C}$ that are most associated with the elements of $\boldsymbol{X}$ in a manner that removes RIC bias.

While our formulation enables the analysis of much larger data sets, there are some disadvantages relative to fully nonparametric techniques such as Gaussian processes. Our model requires us to specify the functional form of the relationships between the predictors and the outcome a priori. In the current implementation, we employed natural splines to flexibly model these relationships. While other formulations are possible, any formulation would require a subjective choice. Further, one must select the number of degrees of freedom of the splines, and we have restricted this number to be the same for all exposures across all features. This can be a limitation if some of the functions to be estimated are quite smooth while others are more complicated. Our approach is also not fully Bayesian, as we use the WAIC to select the degrees of freedom. However, our simulations show the proposed estimation procedure yields uncertainty estimates that are accurate in all of the scenarios we considered. Future extensions could consider a data-adaptive approach to the selection of the degrees of freedom for each function or the utilization of other flexible structures that are both computationally fast and do not require prespecification of the functional forms of the exposure-response functions.

Software. An R package implementing the proposed methodology is available at github.com/jantonelli111/NLinteraction.

SUPPLEMENTARY MATERIAL

Additional data analyses, technical details and convergence diagnostics (DOI: 10.1214/19-AOAS1307SUPP; .pdf). These additional materials contain the additional data analysis looking at environmental exposures in the NHANES study, additional simulation results, derivations of empirical Bayes estimates, full details of the MCMC sampling, a proof of reversibility for our MCMC algorithm, a proof of Theorem 4.2, and convergence diagnostics for the data analyses.

REFERENCES

ANTONELLI, J., MAZUMDAR, M., BELLINGER, D., CHRISTIANI, D., WRIGHT, R. and COULL, B. (2020). Supplement to "Estimating the health effects of environmental mixtures using Bayesian semiparametric regression and sparsity inducing priors." https://doi.org/10.1214/19-AOAS1307SUPP.

BREIMAN, L. (2001). Random forests, decision trees, and categorical predictors: The "absent levels" problem. *Mach. Learn.* **45** 5–32.

BAYLEY, N. (2006). *Bayley Scales of Infant and Toddler Development-Third Edition*: *Administration Manual*.

BENGIO, Y., DUCHARME, R., VINCENT, P. and JAUVIN, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.* **3** 1137–1155.

BIEN, J., TAYLOR, J. and TIBSHIRANI, R. (2013). A LASSO for hierarchical interactions. *Ann. Statist.* **41** 1111–1141. MR3113805 https://doi.org/10.1214/13-AOS1096

BOBB, J. F., VALERI, L., CLAUS HENN, B., CHRISTIANI, D. C., WRIGHT, R. O., MAZUMDAR, M., GODLESKI, J. J. and COULL, B. A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* **16** 493–508. MR3365442 https://doi.org/10.1093/biostatistics/kxu058

BOBB, J. F., HENN, B. C., VALERI, L. and COULL, B. A. (2018). Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environ. Health* **17** 67.

BRAUN, J. M. (2017). Early-life exposure to EDCs: Role in childhood obesity and neurodevelopment. *Nat. Rev. Endocrinol.* **13** 161–173. https://doi.org/10.1038/nrendo.2016.186

BRAUN, J. M., GENNINGS, C., HAUSER, R. and WEBSTER, T. F. (2016). What can epidemiological studies tell us about the impact of chemical mixtures on human health? *Environ. Health Perspect.* **124** A6–A9.

CARLIN, D. J., RIDER, C. V., WOYCHIK, R. and BIRNBAUM, L. S. (2013). Unraveling the health effects of environmental mixtures: An NIEHS priority. *Environ. Health Perspect.* **121** A6–A8. https://doi.org/10.1289/ehp.1206182

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172 https://doi.org/10.1214/09-AOAS285

CRISTIANINI, N. and SHAWE-TAYLOR, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge Univ. Press, Cambridge.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. *Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. MR2027492

GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2** 1360–1383. MR2655663 https://doi.org/10.1214/08-AOAS191

GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.

GUO, F., WANG, X., FAN, K., BRODERICK, T. and DUNSON, D. B. (2016). Boosting variational inference. ArXiv Preprint. Available at arXiv:1611.05559.

HAHN, P. R., MURRAY, J. and CARVALHO, C. M. (2017). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. Confounding, and Heterogeneous Effects (October 5, 2017).

HAHN, P. R., CARVALHO, C. M., PUELZ, D. and HE, J. (2018). Regularization and confounding in linear regression for treatment effect estimation. *Bayesian Anal.* **13** 163–182. MR3737947 https://doi.org/10.1214/16-BA1044

HAO, N., FENG, Y. and ZHANG, H. H. (2014). Model selection for high dimensional quadratic regression via regularization **85721** 26.

HARLEY, K. G., BERGER, K., RAUCH, S., KOGUT, K., HENN, B. C., CALAFAT, A. M., HUEN, K., ESKENAZI, B. and HOLLAND, N. (2017). Association of prenatal urinary phthalate metabolite concentrations and childhood BMI and obesity. *Pediatr. Res.* **82** 405.

HENN, B. C., COULL, B. A. and WRIGHT, R. O. (2014). Chemical mixtures and children's health. *Curr. Opin. Pediatr.* **26** 223.

HENN, B. C., ETTINGER, A. S., SCHWARTZ, J., TÉLLEZ-ROJO, M. M., LAMADRID-FIGUEROA, H., HERNÁNDEZ-AVILA, M., SCHNAAS, L., AMARASIRIWARDENA, C., BELLINGER, D. C. et al. (2010). Early postnatal blood manganese levels and children's neurodevelopment. *Epidemiology* **21** 433.

KORTENKAMP, A., FAUST, M., SCHOLZE, M. and BACKHAUS, T. (2007). Low-level exposure to multiple chemicals: Reason for human health concerns? *Environ. Health Perspect.* **115** 106–114.

LAZAREVIC, N., BARNETT, A. G., SLY, P. D. and KNIBBS, L. D. (2019). Statistical methodology in studies of prenatal exposure to mixtures of endocrine-disrupting chemicals: A review of existing approaches and new alternatives. *Environ. Health Perspect.* **127** 026001.

LIM, M. and HASTIE, T. (2015). Learning interactions via hierarchical group-lasso regularization. *J. Comput. Graph. Statist.* **24** 627–654. MR3397226 https://doi.org/10.1080/10618600.2014.938812

MILLER, A. C., FOTI, N. and ADAMS, R. P. (2016). Variational boosting: Iteratively refining posterior approximations. ArXiv Preprint. Available at arXiv:1611.06585.

MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1036. MR0997578

O'HAGAN, A. (1978). Curve fitting and optimal design for prediction. *J. Roy. Statist. Soc. Ser. B* **40** 1–42. MR0512140

QAMAR, S. and TOKDAR, S. (2014). Additive Gaussian process regression. ArXiv Preprint. Available at arXiv:1411.7009.

RADCHENKO, P. and JAMES, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. *J. Amer. Statist. Assoc.* **105** 1541–1553. MR2796570 https://doi.org/10.1198/jasa.2010.tm10130

REICH, B. J., STORLIE, C. B. and BONDELL, H. D. (2009). Variable selection in Bayesian smoothing spline ANOVA models: Application to deterministic computer codes. *Technometrics* **51** 110–120. MR2668168 https://doi.org/10.1198/TECH.2009.0013

ROČKOVÁ, V. and GEORGE, E. I. (2018). The spike-and-slab LASSO. *J. Amer. Statist. Assoc.* **113** 431–444. MR3803476 https://doi.org/10.1080/01621459.2016.1260469

SCHEIPL, F., FAHRMEIR, L. and KNEIB, T. (2012). Spike-and-slab priors for function selection in structured additive regression models. *J. Amer. Statist. Assoc.* **107** 1518–1532. MR3036413 https://doi.org/10.1080/01621459.2012.737742

SHIVELY, T. S., KOHN, R. and WOOD, S. (1999). Variable selection and function estimation in additive nonparametric regression using a data-based prior. *J. Amer. Statist. Assoc.* **94** 777–806. MR1723272 https://doi.org/10.2307/2669990

TAYLOR, K. W., JOUBERT, B. R., BRAUN, J. M., DILWORTH, C., GENNINGS, C., HAUSER, R., HEINDEL, J. J., RIDER, C. V., WEBSTER, T. F. et al. (2016). Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: Lessons from an innovative workshop. *Environ. Health Perspect.* **124** A227–A229. https://doi.org/10.1289/EHP547

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

VALERI, L., MAZUMDAR, M. M., BOBB, J. F., HENN, B. C., RODRIGUES, E., SHARIF, O. I., KILE, M. L., QUAMRUZZAMAN, Q., AFROZ, S. et al. (2017). The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20–40 months of age: Evidence from rural Bangladesh. *Environ. Health Perspect.* **125**.

WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WOOD, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* **62** 1025–1036. MR2297673 https://doi.org/10.1111/j.1541-0420.2006.00574.x

WOOD, S., KOHN, R., SHIVELY, T. and JIANG, W. (2002). Model selection in spline nonparametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 119–139. MR1883129 https://doi.org/10.1111/1467-9868.00328

YAU, P., KOHN, R. and WOOD, S. (2003). Bayesian variable selection and model averaging in high-dimensional multinomial nonparametric regression. *J. Comput. Graph. Statist.* **12** 23–54. MR1965210 https://doi.org/10.1198/1061860031301

ZHAO, P., ROCHA, G. and YU, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist.* **37** 3468–3497. MR2549566 https://doi.org/10.1214/07-AOS584

ZHOU, J., BHATTACHARYA, A., HERRING, A. H. and DUNSON, D. B. (2015). Bayesian factorizations of big sparse tensors. *J. Amer. Statist. Assoc.* **110** 1562–1576. MR3449055 https://doi.org/10.1080/01621459.2014.983233

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x