Statistical Methods for Estimating the Effects of Multi-Pollutant Exposures in Children's Health Research

Statistical Methods for Estimating the Effects of Multi-pollutant Exposures in Children's Health Research

A dissertation presented

by

Shelley Han Liu

to

The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts

August 2016

Dissertation Advisor: Brent Coull Shelley Han Liu

Statistical Methods of Estimating the Effects of Multi-pollutant Exposures in Children's Health Research

Abstract

We develop statistical strategies to explore how time-varying exposures to heavy metal mixtures affects cognition and cognitive trajectories in children. In chapter 1, we develop a Bayesian model, called Lagged Kernel Machine Regression (LKMR), to identify time windows of susceptibility to exposures of metal mixtures. In chapter 2, we develop a Mean Field Variational Bayesian (MFVB) inference procedure for LKMR. We demonstrate large computational gains under MFVB as opposed to Markov chain Monte Carlo (MCMC) inference for LKMR, which allows for the analysis of large datasets while maintaining accuracy. In chapter 3, we present a Bayesian hierarchical model, called Bayesian Varying Coefficient Kernel Machine Regression, to investigate the impact of exposure to heavy metal mixtures on cognitive growth trajectories in children. Simulation studies demonstrate the effectiveness of these methods, and the methods are used to analyze data from two prospective birth cohort studies in Mexico City.

Contents

Acknowledgments

I would like to first thank my advisor, Brent Coull, for being an excellent mentor who has shaped me into the biostatistician I am today. Thank you for always having the time and patience to support me in my endeavors, and provide advice and guidance. I would also like to thank Matt Wand, who hosted me last summer in Australia and taught me a great deal about coding and statistics. Thank you to Victor DeGruttola, with whom I worked on my first summer project with on infectious disease statistics. Lastly, thank you to my committee members Xihong Lin and Robert Wright, for their invaluable feedback - their sharp insights and suggestions have greatly improved my work.

As I look back, there are so many mentors and programs that have shaped my path. Thank you to Roger Kroes at Northwestern who gave me the opportunity to work in a molecular biology lab freshman year. Thank you to the University of Wisconsin-Madison's Department of Biostatistics and Medical Informatics, which gave me my first foray into biostatistics through the SIBS program, and gave me the opportunity to return the following summer through the IBS-SRP. Those experiences helped solidify my desire to study biostatistics. Thank you to the Rose Traveling Fellowship, which allowed me to step outside of biostatistics and travel to Chile, and see how my quantitative skills could help inform policy. Thank you to the National Science Foundation EAPSI-Australia and EAPSI-China grants which allowed me to extend my dissertation work and develop global collaborations.

Thank you to the Biostatistics department and Shattuck International House, which have made Boston feel like home these five years, and introduced me to myriad of friends at HSPH.

The deepest gratitude to my parents, Yanming and Yuming and my little brother Michael, for their perpetual love and support. It's because of them that I had access to so many opportunities. Lastly, thank you to my fiancé Thomas Yu, who has been a steadfast part of this journey. Words cannot describe how much his support has meant to me since freshman year of college - thank you for putting up with countless Boston-Chicago flights during graduate school, and for believing in me.

Lagged Kernel Machine Regression for Identifying Time Windows of Susceptibility to Exposures of Complex Metal Mixtures

Shelley Han Liu

Department of Biostatistics

Harvard Graduate School of Arts and Sciences

Jennifer F. Bobb

Biostatistics Unit

Group Health Research Institute

Kyu Ha Lee

Biostatistics and Epidemiology Core

The Forsyth Institute

Chris Gennings

Department of Environmental Medicine and Public Health

Department of Population Health Science and Policy

Icahn School of Medicine at Mount Sinai

Birgit Claus Henn
Department of Environmental Health
Boston University School of Public Health

David Bellinger
Department of Environmental Health
Harvard T.H. Chan School of Public Health

Christine Austin
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Lourdes Schnaas
Division for Research in Community Interventions
National Institute of Perinatology, Mexico

Martha Tellez-Rojo
Center for Research in Nutrition and Health
National Institute of Public Health, Mexico

Robert Wright
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Manish Arora
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Brent Coull

Department of Biostatistics

Harvard T.H. Chan School of Public Health

1.1 Introduction

Neurodevelopment and cognitive function are important outcomes in public health. A critical public health concern is the impact of neurotoxic chemicals on children's health. There is a large body of literature on the impact of exposure to individual chemicals, such as lead, on neurodevelopment (Bellinger, 2008; Tellez-Rojo et al., 2006). However, exposure to chemical mixtures, rather than to individual chemicals, are more reflective of real-world scenarios. Accordingly, the National Institute of Environmental Health Sciences (NIEHS) has placed a priority on quantification of the health impacts of exposure to environmental mixtures (Carlin et al., 2013).

For the estimation of the health effect of metal mixtures on neurodevelopment, the exposure-response relationship can be complex, exhibiting both nonlinearity and non-additivity. The effect of some metals, such as trace elements like manganese, can be nonlinear as they are essential nutrients at low doses but neurotoxic at high exposure levels. These dual roles can result in an inverted-u relationship with neurodevelopment (Claus Henn et al., 2010). Moreover, existing work on metal mixtures provides evidence of interactions between individual metals. For instance, Claus Henn et al. (2014) found increased lead toxicity in the presence of higher levels of manganese, arsenic, and cadmium.

Another layer of complexity in the identification of environmental effects on children's health is that health effects can be highly-dependent on exposure timing. There exist many sequential developmental processes in early life, as development is unidirectional and well-timed (Stiles and Jernigan, 2010; Tau and Peterson, 2010). For instance, pregnancy is a state of sequential physiologic changes, such that an infant may be particularly susceptible to exposure during a certain developmental stage, which we call a critical exposure window. Metal mixture exposures may be especially harmful during prenatal and early life periods. Several metals cross the placental barrier, potentially causing injury to the fetal brain. A previous study reported that the interaction of lead and cadmium may depend on the stage of pregnancy (Kim et al., 2013). In such cases,

measuring exposure either in the wrong critical window or averaging exposure over the entire pregnancy when only a specific window is most relevant is a form of exposure misclassification.

There is a lack of statistical methods to simultaneously accommodate the complex exposure-response relationship between metal mixtures and neurodevelopment while analyzing data on critical exposure windows of these exposures. Traditionally, these two research questions - (1) assessing health effects of complex mixtures and (2) identifying time windows of susceptibility - have been studied separately. Methods to address the complex exposure-response relationship include classification and regression trees, random forest, cluster analysis, nonparametric Bayesian shrinkage, Bayesian mixture modeling and weighted quantile sum regression (Billionnet et al., 2012; Herring, 2010; de Vocht et al., 2012; Diez et al., 2012; Roberts and Martin, 2006; Gennings et al., 2013). Bobb et al. (2015) developed Bayesian kernel machine regression (BKMR) for estimating health effects of complex mixtures and conducting variable selection for exposures at a single time point. Meanwhile, methods for identifying time windows of susceptibility are focused on using single pollutant distributed lag models to study the effect of a single toxicant assuming no interaction between time windows (Hsu et al., 2015; Warren et al., 2012, 2013; Darrow et al., 2011). One exception is the work of Heaton and Peng (2013), who developed a higher degree distributed lag model to account for cross-time interaction. However, this model still only related to exposure of a single pollutant. To our knowledge, there are no existing methods to identify critical exposure windows of multi-pollutant mixtures.

To address this gap in the statistical literature, we develop methodology to investigate how exposures to heavy metal mixtures during early childhood affect long-term cognitive function, and identify specific critical windows of exposure. We introduce a new method, Lagged Kernel Machine Regression (LKMR), to estimate the health effects of time-varying exposures to environmental mixtures, and identify critical exposure windows of a mixture. We adopt a Bayesian paradigm for inference of LKMR. We use the

kernel machine regression (KMR) framework, which is popular in the statistical genetics literature where it is used primarily to test the significance of gene sets and predict risk for health outcomes (Cai et al., 2011; Maity and Lin, 2011), Bayesian KMR has also been shown to effectively estimate complex exposure-response functions associated with metal mixtures (Bobb et al., 2015).

We develop LKMR to handle time-varying mixture exposures. By incorporating methods from the single time point BKMR and the single exposure distributed lag model, LKMR estimates nonlinear and non-additive effects of mixtures of exposures while assuming these effects vary smoothly over time. To accomplish these goals, we develop a novel Bayesian penalization scheme that combines the group and fused lasso. The group lasso regularizes the exposure-response function at each time point, whereas the fused lasso shrinks the exposure-response functions from timepoints close in time towards one another. Notably, we show this can be achieved by seamlessly using the kernel matrix relying on a given similarity matrix into the penalty term for the grouped lasso component. We implement the method using Bayesian lasso methods (Yuan and Lin, 2006; Huang et al., 2012; Kyung et al., 2010). Although Bayesian grouped lasso and Bayesian fused lasso have been used individually, our new proposal combines these penalization schemes together with kernel machine methods, resting in a novel model formulation.

We apply this model to data from the ongoing Early Life Exposures in Mexico and Neurotoxicology (ELEMENT) study. In ELEMENT, a prospective birth cohort study, teeth dentine captures exposure to barium (Ba), chromium (Cr), lithium (Li), manganese (Mn), and zinc (Zn) over time. Through imaging of teeth, fine temporal resolution of metal exposure from the second trimester of pregnancy to early childhood is obtained. In our data application, we were interested in Mn-Zn interactions at three exposure windows of early development (second and third trimesters of pregnancy and mos. 0-3 after birth), and their associations with neurodevelopment.

The paper is organized as follows: Section 1.2 provides a review of KMR; Section 1.3 introduces the statistical model LKMR; Section 1.4 describes the simulation studies for evaluating performance of LKMR; Section 1.5 addresses an application of the method to the ELEMENT dataset; and Section 1.6 provides discussion and concluding remarks.

1.2 Review of kernel machine regression

To set notation we first review kernel machine regression as a framework for estimating the effect of a complex mixture when exposure is measured only at a single time point. Suppose we observe data from n subjects. We describe the model for a continuous, normally distributed outcome. For each subject $i = 1, \ldots, n$, KMR relates the health outcome (y_i) to M components of the exposure mixture $\mathbf{z}_i = (z_{1i}, ..., z_{Mi})$ through a nonparametric function, $h(\cdot)$, while controlling for C relevant confounders $\mathbf{x}_i = (x_{1i}, ..., x_{Ci})$. The model is

$$y_i = h(z_{1i}, ..., z_{Mi}) + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{1.1}$$

where $\boldsymbol{\beta}$ represents the effects of the potential confounders, and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. $h(\cdot)$ can be estimated parametrically or non-parametrically. We employ a kernel representation for $h(\cdot)$ in order to accommodate the possibly complex exposure-response relationship.

The unknown function $h(\cdot)$ can be specified through basis functions or through a positive definite kernel function $K(\cdot, \cdot)$. Under regularity conditions, Mercer's theorem (Cristianini and Shawe-Taylor, 2000) showed that the kernel function $K(\cdot, \cdot)$ implicitly specifies a unique function space, H_k, that is spanned by a set of orthogonal basis functions. Thus, any function $h(\cdot) \in H_k$ can be represented through either a set of basis functions under the primal representation, or through a kernel function under the dual representation. The kernel function uses a similarity metric $K(\cdot, \cdot)$ to quantify the distance between the exposure profiles $\mathbf{z}_i$ between any two subjects in the study. For example, the Gaussian kernel quantifies similarity through the Euclidean distance; the polynomial kernel, through the inner product. Through specifying different kernels, one is able to control the complexity of the exposure-response function.

Liu et al. (2007) developed least-squares kernel machine semi-parametric regression for studying genetic pathway effects. They make the connection between kernel machine methods and linear mixed models, demonstrating that (1) can be expressed as the mixed model

$$y_i \sim N(h_i + \mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2) \tag{1.2}$$

$$\mathbf{h} = (h_1, ..., h_n)^T \sim N\left[\mathbf{0}, \tau \mathbf{K}(\cdot, \cdot)\right], \tag{1.3}$$

where **K** is a kernel matrix with *i,j* element $K(\mathbf{z}_i, \mathbf{z}_j)$.

1.3 Lagged kernel machine regression

1.3.1 Model formulation

Now assume exposures to a complex mixture are measured at multiple timepoints with the goal of identifying critical windows of exposure, such that we have data on the multi pollutant exposures $\mathbf{z}_{it} = (z_{1i,t}, ..., z_{Mi,t})$. We define the lagged kernel machine regression (LKMR) as

$$y_i = \beta_0 + \sum_t h_t(z_{1i,t}, ..., z_{Mi,t}) + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{1.4}$$

where $\boldsymbol{h}_t(\cdot) = (h_{1,t}, ..., h_{n,t})^T$ represents the (potentially complex) exposure-response function for the exposures $\mathbf{z}_{1,t}, ..., \mathbf{z}_{n,t}$ measured at time t, controlling for exposures at all other timepoints. We use the mixed model representation proposed by Liu et al. (2007) for each $\boldsymbol{h}_t$, $t = 1, \ldots, T$, yielding

$$y_i = \beta_0 + \sum_t h_{i,t} + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i. \tag{1.5}$$

Model (1.4) represents a multiple kernel learning model. However, straightforward model fitting $\boldsymbol{h}_t(\cdot)$ for all $t = 1, \ldots, T$, will yield unstable estimates of each time-specific exposure-response function because of the unavoidable correlation among the time-varying measures of each exposure. We therefore employ a strategy that regularizes the individual exposure-response estimates by shrinking the time-specific $\boldsymbol{h}_t(\cdot)$ that are adjacent in time towards one another. Accordingly, in LKMR, we impose penalization through a novel Bayesian, grouped, fused lasso. The group lasso component regularizes

each $\boldsymbol{h}_t$ individually by penalizing the kernel surface, as exposures at individual time windows can share similarities, and also provides a framework for incorporating kernel machine regression. Meanwhile, the fused lasso component pools shrinks differences in $\boldsymbol{h}_t$ across neighboring time windows, and smoothes adjacent kernel surfaces. To incorporate the possibility of complex exposure-response functions at each time t, we incorporate kernel distance functions within the group lasso implementation.

Let $\boldsymbol{h} = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_T)^T$. The conditional prior of $\boldsymbol{h}|\sigma^2$ is

$$\pi(\boldsymbol{h}|\sigma^2, \lambda_1, \lambda_2) \propto exp\left[\frac{-\lambda_1}{\sigma^2}\sum_{t=1}^{T}\|\boldsymbol{h}_t\|_{\mathbf{G}_t} - \frac{\lambda_2}{\sigma^2}\sum_{t=1}^{T-1}|\boldsymbol{h}_{t+1} - \boldsymbol{h}_t|_1\right], \tag{1.6}$$

where $\|\boldsymbol{h}_t\|_{\mathbf{G_t}} = (\boldsymbol{h_t}^T\mathbf{G_t}\mathbf{h_t})^{\mathbf{1/2}}$. We define $\mathbf{G}_t = \mathbf{K}_t^{-1}$, where $\mathbf{K}_t$ denotes the kernel matrix for time t with *i,j* element $K_t(\mathbf{z}_i, \mathbf{z}_j)$. Depending on a particular application, one can choose any of many different kernel functions. Previous literature has shown an inverted-*u* relationship between metals and neurodevelopment (Claus Henn et al., 2010); thus, we choose $\mathbf{K}_t$ to be a quadratic kernel, such that $K(\mathbf{z, z'}) = (\mathbf{zz'} + 1)^2$.

An advantage of the model is that we can formally specify it in a hierarchical fashion, which allows for a Gibbs sampler implementation. We introduce latent parameters $\boldsymbol{\tau} = (\tau_1^2, ..., \tau_T^2)$ and $\boldsymbol{\omega} = (\omega_1^2, ..., \omega_{T-1}^2)$, as we prefer this conditional prior for implementation using a Gibbs sampler: $\boldsymbol{h}|\boldsymbol{\tau}, \boldsymbol{\omega}, \sigma^2 \sim N_P(\mathbf{0}, \sigma^2\Sigma_h)$, where $P = n * T$.

The hierarchical model is represented as:

$$\mathbf{y}|\boldsymbol{h}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N_n(\mathbf{X}\boldsymbol{\beta} + \sum_t \boldsymbol{h}_t, \sigma^2\boldsymbol{I}_n) \tag{1.7}$$

$$\boldsymbol{h}|\tau_1^2, ..., \tau_T^2, \omega_1^2, ..., \omega_{T-1}^2, \sigma^2 \sim N_P(\mathbf{0}, \sigma^2\Sigma_h) \tag{1.8}$$

$$\tau_1^2, ..., \tau_T^2 \overset{iid}{\sim} \text{gamma}(\frac{n+1}{2}, \frac{{\lambda_1}^2}{2}) \tag{1.9}$$

$$\omega_1^2, ..., \omega_{T-1}^2 \overset{iid}{\sim} \frac{{\lambda_2}^2}{2}e^{\frac{-\lambda_2{}^2\omega_t^2}{2}}, \tag{1.10}$$

where $\tau_1^2, ..., \tau_T^2, \omega_1^2, ..., \omega_{T-1}^2, \sigma^2$ are mutually independent. The form of Σ_h^{-1} follows from representing the Laplace (double exponential) conditional prior of $\boldsymbol{h}|\sigma^2$ as a scale mixture of a normal distribution with an exponential mixing density (Andrews and Mallows, 1974). Figure 1.1 presents the full form of the variance covariance matrix Σ_h^{-1} . The diagonal blocks of size n x n arise due to the kernel structure placed on each $\boldsymbol{h}_t(\cdot)$, whereas the off-diagonals involving the ω_t^2 parameters serve to shrink random effects adjacent in time towards one another.

1.3.2 Prior specification

To complete the model specification, we define prior distributions for the regression parameters $\boldsymbol{\beta}$ and σ^2. We take the prior for σ^2 to gamma, and $\boldsymbol{\beta}$ to have an independent flat prior, $\pi(\boldsymbol{\beta}) \propto 1$. In this article, we use a gamma prior for tuning parameters λ_1 and λ_2, as suggested by Park and Casella (2008). The gamma priors are placed on λ_1^2 and λ_2^2 for convenience based on their appearance in the posterior distribution. Priors are of the form:

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)}(\lambda^2)^{r-1}e^{-\delta\lambda^2}, \lambda^2 > 0, r > 0, \delta > 0 \tag{1.11}$$

As motivated by simulation studies described in Section 4, we evaluated a grid of r and δ values to identify hyperparameters which optimize estimation of the exposure-response surface. Our simulation studies identified hyperparameter values for λ_1^2, r = 150 and δ = 10, and hyper parameters for λ_2^2, r = 60 and δ = 10, as values that performed well under a range of simulation scenarios and maximize the posterior inference of the exposure-response surface.

1.3.3 MCMC sampler

In this section, we describe the Gibbs sampler implementation for the hierarchy of (7) - (10). For convenience, we denote $\sum_t \boldsymbol{h}_t$ as $\boldsymbol{W}\boldsymbol{h}$. The joint density is:

$$f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{h}, \sigma^2, \boldsymbol{\tau}, \boldsymbol{\omega}) \propto \frac{1}{(2\pi\sigma^2)^{n/2}}\text{exp}\left[\frac{-1}{2\sigma^2}(\mathbf{y} - \mathbf{W}\boldsymbol{h} - \boldsymbol{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{W}\boldsymbol{h} - \boldsymbol{X}\boldsymbol{\beta})\right] \times$$

Figure 1.1: Form of covariance matrix Σ_h^{-1}

$$\Sigma_h^{-1} = \begin{bmatrix}
\frac{G_{1,11}}{\tau_1^2} + \frac{1}{\omega_1^2} & \frac{G_{1,12}}{\tau_1^2} & \cdots & \frac{G_{1,1n}}{\tau_1^2} & \frac{-1}{\omega_1^2} & 0 & \cdots & 0 & \cdots \\
\frac{G_{1,12}}{\tau_1^2} & \frac{G_{1,22}}{\tau_1^2} + \frac{1}{\omega_1^2} & \cdots & \frac{G_{1,2n}}{\tau_1^2} & 0 & \frac{-1}{\omega_1^2} & 0 & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \cdots \\
\frac{G_{1,1n}}{\tau_1^2} & \frac{G_{1,2n}}{\tau_1^2} & \cdots & \frac{G_{1,nn}}{\tau_1^2} + \frac{1}{\omega_1^2} & 0 & 0 & 0 & \frac{-1}{\omega_1^2} & \cdots \\
\frac{-1}{\omega_1^2} & 0 & \cdots & 0 & \frac{G_{2,11}}{\tau_2^2} + \frac{1}{\omega_1^2} + \frac{1}{\omega_2^2} & \frac{G_{2,12}}{\tau_2^2} & \cdots & \frac{G_{2,1n}}{\tau_2^2} & \ddots \\
0 & \frac{-1}{\omega_1^2} & \ddots & \vdots & \frac{G_{2,12}}{\tau_2^2} & \frac{G_{2,22}}{\tau_2^2} + \frac{1}{\omega_1^2} + \frac{1}{\omega_2^2} & \cdots & \frac{G_{2,2n}}{\tau_2^2} & \cdots \\
\vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \vdots & \cdots \\
0 & 0 & \cdots & \frac{-1}{\omega_1^2} & \frac{G_{2,1n}}{\tau_2^2} & \frac{G_{2,2n}}{\tau_2^2} & \cdots & \frac{G_{2,nn}}{\tau_2^2} + \frac{1}{\omega_1^2} + \frac{1}{\omega_2^2} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}$$

$$\frac{1}{\sigma^2}\prod_{t=1}^{T}\left(\frac{1}{2\pi\sigma^2\tau_t^2}\right)^{n/2}\exp\left(-\frac{\|\boldsymbol{h_t}\|^2_{\mathbf{G}_t}}{2\sigma^2\tau_t^2}\right)\frac{\left(\frac{\lambda_1^2}{2}\right)^{\frac{n+1}{2}}\left(\tau_t^2\right)^{\frac{n+1}{2}-1}}{\Gamma\left(\frac{n+1}{2}\right)}\exp\left(-\lambda_1^2\tau_t^2/2\right)\times$$
$$\prod_{t=1}^{T-1}\left(\frac{1}{2\pi\sigma^2\omega_t^2}\right)^{1/2}\exp\left[-\frac{\sum_{i=1}^{n}\left(h_{i,t+1}-h_{i,t}\right)^2}{2\sigma^2\omega_t^2}\right]\frac{\lambda_2^2}{2}\exp\left(-\lambda_2^2\omega_t^2/2\right) \tag{1.12}$$

For brevity, we detail the full conditional distributions of a few key parameters:
The full conditional distribution of $\boldsymbol{h}$ is:

$$\boldsymbol{h}|\sigma^2,\boldsymbol{\tau},\boldsymbol{\omega},X,\boldsymbol{\beta},\mathbf{y}\sim N_P((\boldsymbol{W}^T\boldsymbol{W}+\Sigma_h^{-1})^{-1}\boldsymbol{W}^T\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right),\sigma^2(\boldsymbol{W}^T\boldsymbol{W}+\Sigma_h^{-1})^{-1})$$

The full conditional of σ^2 is:

$$\sigma^2|\boldsymbol{h},\boldsymbol{\tau},\boldsymbol{\omega},\boldsymbol{X},\boldsymbol{\beta},\mathbf{y}\sim \tag{1.13}$$
$$IG(n/2+T/2+(T-1)/2+1,\frac{(\mathbf{y}-\mathbf{W}\boldsymbol{h}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{W}\boldsymbol{h}-\mathbf{X}\boldsymbol{\beta})}{2}+\frac{\boldsymbol{h}^T\Sigma_h^{-1}\boldsymbol{h}}{2})$$

Also, the full conditional of β is:

$$\boldsymbol{\beta}|rest\sim N((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y}-\mathbf{W}\boldsymbol{h}),\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}) \tag{1.14}$$

The Gibbs sampler is implemented to cyclically sample from the distributions of $\boldsymbol{h},\boldsymbol{\beta},\sigma^2$, $\boldsymbol{\tau^2}$, $\boldsymbol{\omega^2}$, λ_1^2, and λ_2^2 conditional on the current values of the other parameters. We note that for several parameters, such as $\boldsymbol{h}$, $\boldsymbol{\beta}$, $\boldsymbol{\tau}^2$, and $\boldsymbol{\omega}^2$, the Gibbs sampler is a block update.

1.3.4 Predicting health effects at new time-varying exposure profiles

It is often important to estimate and visualize the exposure-response surface, in order to ascertain health effects of time-varying toxicant mixtures. Suppose we are interested in predicting the exposure-response profile for a new profile of metal mixture exposures at time t. There are currently n subjects in our study for $t=1,...,T$ time points, and we are interested in predicting the response for the n_{new} subjects. Thus, we are interested in estimating $h_{n+1,t},...,h_{n+n_{new},t}$.

In order to reduce computation time, we approximate the posterior mean and variance of $\boldsymbol{h}_{new,t}$ = $(h_{n+1,t},...,h_{n+n_{new},t})$, by using the estimated posterior mean of the other parameters in the formulas below. First, we define $\tilde{\boldsymbol{h}}$, which the reordered $\boldsymbol{h}$ vector,

such that the new time point of interest, t, is at the end of the vector. This step aids in estimating the $\boldsymbol{h}_{new}$. Thus,

$$\tilde{\boldsymbol{h}} = (h_{1,1}, ..., h_{n,t-1}, h_{1,t+1}, ..., h_{n,t+1}, ..., h_{1,T}, ..., h_{n,T}, h_{1,t}, ..., h_{n,t}, h_{n+1,t}, ..., h_{n+n_{new},t})^T$$

Because we have reordered the $\tilde{\boldsymbol{h}}$ vector, we need to similarly reorder the corresponding covariance matrix, and denote the reordered matrix as $\tilde{\Sigma}_h^{-1}$. The joint distribution of observed and new exposure profiles is:

$$\begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{h}_{new} \end{pmatrix} \sim N\left\{ \mathbf{0}, \tilde{\Sigma}_h = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{12}^T & \tilde{\Sigma}_{22} \end{pmatrix} \right\} \tag{1.15}$$

where $\tilde{\Sigma}_{11}$ denotes the $2n$ x $2n$ matrix with $(i,j)^{th}$ element $K(\mathbf{z}_i, \mathbf{z}_j)$, $\tilde{\Sigma}_{12}$ denotes the n x n_{new} matrix with $(i, j_{new})^{th}$ element $K(\mathbf{z}_i, \mathbf{z}_{j_{new}})$, and $\tilde{\Sigma}_{22}$ denotes the n_{new} x n_{new} matrix with $(i_{new}, j_{new})^{th}$ element $K(\mathbf{z}_{i_{new}}, \mathbf{z}_{j_{new}})$. It follows that the conditional posterior distribution of $\boldsymbol{h}_{new}$ is:

$$\boldsymbol{h}^{\boldsymbol{new}} | \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\tau}^{\mathbf{2}}, \sigma^2 \sim N_{n_{new}} \Bigg\{ \tilde{\Sigma}_{12}^T \tilde{\Sigma}_{11}^{-1} \Big\{ \frac{1}{\sigma^2} \mathbf{W}^T \mathbf{W} + \tilde{\Sigma}_{11}^{-1} \Big\}^{-1} \frac{1}{\sigma^2} \mathbf{W}^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\boldsymbol{b}), \tag{1.16}$$

$$\tilde{\Sigma}_{12}^T \tilde{\Sigma}_{11}^{-1} \Big\{ \frac{1}{\sigma^2} \mathbf{W}^T \mathbf{W} + \tilde{\Sigma}_{11}^{-1} \Big\}^{-1} \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12} + \tilde{\Sigma}_{22} - \tilde{\Sigma}_{12}^T \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12} \Bigg\}$$

In order to reduce computation time, we approximate the posterior mean and variance of $\boldsymbol{h}_{new}$ based on the estimated posterior mean of the other parameters.

1.4 Simulation studies

We conducted simulation studies to evaluate the performance of the proposed LKMR model for estimating critical exposure windows of environmental mixtures. Our simulation study considered a three-toxicant scenario: two toxicants (out of three) exert a gradual non-additive, non-linear effect over four time windows that are representative of early life. We used the following model: $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_t \boldsymbol{h}_t(\mathbf{z}_{it}) + e_i$, where $e_i \sim N(0,1)$, $\mathbf{x}_i = (x_{1i}, x_{2i})$ and $x_{1i} \sim N(10,1)$ and $x_{2i} \sim Bernoulli(0.5)$. We simulated auto-correlation within toxicant exposures z_m across time, and correlation between toxicants, using the

Kronecker product for the exposure correlation matrix. Three choices for auto-correlation within toxicants were considered: high (0.8), medium (0.5) and low (0.2). The exposure-response function $h(z_i)$ was simulated as quadratic with two-way interactions. We assumed there is no effect of exposure to the environmental mixture at Time 1, and a gradual increasing effect was simulated from Time 2 to 4. We assume $h_t(\mathbf{z}) = \alpha_t h(\mathbf{z})$, where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 0.5, 0.8, 1.0)$ and $h(\mathbf{z}) = z_1^2 - z_2^2 + 0.5 z_1 z_2 + z_1 + z_2$. In conducting the analysis, exposure covariates are logged, centered and scaled. Confounder variables are also centered and scaled.

Table 1.1 presents the results of this simulation. We compared the performance of LKMR to that of Bayesian kernel machine regression (BKMR) applied using exposures from each time window separately. For each simulated data set, to assess the performance of the model for the purposes of estimating the time-specific exposure-response function, we regressed the predicted $\widehat{\boldsymbol{h}}$ on $\boldsymbol{h}$ for each time point. We present the intercept, slope and R^2 of the regressions over 100 simulations. Good estimation performance occurs when the intercept is close to zero, and the slope and R^2 are both close to one. We also present the root mean squared error (RMSE) and the coverage (the proportion of times the true $h_{i,t}$ is contained in the posterior credible interval).

Table 1.1 Simulation results, regression of $\widehat{\boldsymbol{h}}$ on h

h function	Time window	Intercept	Slope	R^2	RMSE	Coverage
LKMR	1	-0.02	N/A	N/A	0.45	1.00
	2	-0.04	0.97	0.86	0.50	1.00
AR-1 = 0.8	3	-0.08	0.96	0.94	0.54	0.99
	4	-0.09	0.99	0.97	0.51	1.00
BKMR	1	0.25	N/A	N/A	2.49	0.65
	2	0.15	2.87	0.90	2.60	0.60
AR-1 = 0.8	3	0.03	2.22	0.97	2.57	0.50
	4	0.03	1.73	0.96	2.05	0.59
LKMR	1	-0.02	N/A	N/A	0.39	1.00
	2	-0.06	0.96	0.91	0.42	1.00
AR-1 = 0.5	3	-0.07	0.97	0.96	0.45	1.00
	4	-0.09	0.98	0.98	0.43	1.00
BKMR	1	0.20	N/A	N/A	0.99	0.92
	2	0.11	1.46	0.75	1.29	0.87
AR-1 = 0.5	3	0.10	1.55	0.90	1.53	0.73
	4	0.10	1.27	0.93	1.18	0.83
LKMR	1	-0.02	N/A	N/A	0.38	1.00
	2	-0.05	0.97	0.93	0.38	1.00
AR-1 = 0.2	3	-0.05	0.98	0.97	0.41	1.00
	4	-0.07	0.97	0.98	0.41	1.00
BKMR	1	0.14	N/A	N/A	0.59	0.98
	2	0.12	0.80	0.63	0.91	0.94
AR-1 = 0.2	3	0.10	1.04	0.85	0.96	0.92
	4	0.06	0.99	0.91	0.84	0.94

Performance of estimated $h_t(\mathbf{z}_i)$ across 100 simulated datasets. RMSE denotes the root mean squared error of the $\hat{h}$ as compared to h. Coverage denotes the proportion of times that the true h falls within in the posterior credible interval of each time point. At time window 1, there is no effect; thus, slope and R^2 are not applicable to the regression of $\widehat{\boldsymbol{h}}$ on $\boldsymbol{h}$.

The results in Table 1.1 suggest that the LKMR significantly outperforms BKMR applied using a single exposure time point when there is high autocorrelation for individual mixture components across time. Specifically, as compared to BKMR, LKMR provides reduc-

tions in RMSE on the order of 75-82% when the autocorrelation in a given exposure is 0.8, and by 36 - 51% when the autocorrelation is 0.2. Credible interval coverage for LKMR is consistently around 100%, as compared to 50-65% for high exposure autocorrelation and 92-98% for low correlation. Furthermore, as demonstrated by a slope of $\widehat{\boldsymbol{h}}$ on $\boldsymbol{h}$ greater than one, BKMR estimates of $\widehat{\boldsymbol{h}}$ at a given time point tends to be biased when the exposure autocorrelation is high, whereas estimates from the LKMR model are approximately unbiased under all autocorrelation scenarios. Collectively, these results demonstrate that naive application of BKMR in this setting suffers from the fact that it estimates the association between exposure at a given time but does not control for exposure at other time points. When autocorrelation in exposure among multiple exposure times is high, this lack of adjustment leads to biased estimates of an exposure effect at the time of interest, whereas when the exposures are roughly uncorrelated, there is less potential for confounding by exposure at different times. In contrast, because LKMR uses penalization to borrow information from neighboring time windows, it performs well under both high and low AR-1 scenarios, and is capable of handling time-varying mixture exposures. These simulations demonstrate that LKMR is less biased than BKMR at estimating the exposure-response function across multiple time points.

1.5 Application

We applied the proposed LKMR model to analyze the association between neurodevelopment and metal mixture exposures in the ELEMENT study conducted in Mexico City. In a pilot study nested within this larger cohort study (n=81), we estimated as the primary outcome the visual spatial subtest score measured at eight years of age. Exposures to metals Ba, Cd, Li, Mn, Zn were measured in teeth dentine, which provides time-specific measures of exposure over both the pre- and post-natal period for each child. These time-varying exposures were averaged to reflect three biologically relevant time windows: second and third trimesters of pregnancy, and 0-3 months after birth). We controlled for child gender, gestational age of infant at birth, maternal IQ, and child hemoglobin at year two.

As a preliminary analysis, we first fit a linear regression model using metal mixture exposures from all three time windows to identify exposures and time point(s) of significance. Zn had a positive association with neurodevelopment at the second trimester (p=0.003), and a negative one at the 3rd trimester (p = 0.0009). Mn was positively associated with neurodevelopment at the third trimester (p = 0.013), and negatively associated during months 0-3 (p = 0.003). Cr had a positive association at the second trimester (p=0.017), and a negative one at the 3rd trimester (p = 0.039) and months 0-3 of early life (p=0.017). Lastly, there was suggestion of an interaction between Zn and Mn at the second trimester (p = 0.051). These results suggest that under assumptions of linearity and additivity, there is some evidence of an exposure-response relationship across multiple timepoints, which warrants further exploration using LKMR.

We then applied LKMR to study time-varying metal mixture exposure effects during early life on visual spatial ability. We first estimated the relative importance of each metal, as shown in Figure 1.2. Relative importance is quantified by the difference in the estimated main effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures. The results are similar to that of the simple linear regression. We detect a negative association of Zn with neurodevelopment in the 3rd trimester. The results also suggest evidence of a positive association of Mn with neurodevelopment at the 3rd trimester, which shifts to a negative association after birth. This qualitatively different (positive and negative) association between Mn and the outcome for Mn exposure pre- and post-natally is particularly intriguing, as Mn is both an essential nutrient and a toxicant. It could be that the developing fetus needs Mn prenatally and receives it via the mother, whereas post-natal exposure reflects environmental exposures that are more harmful.

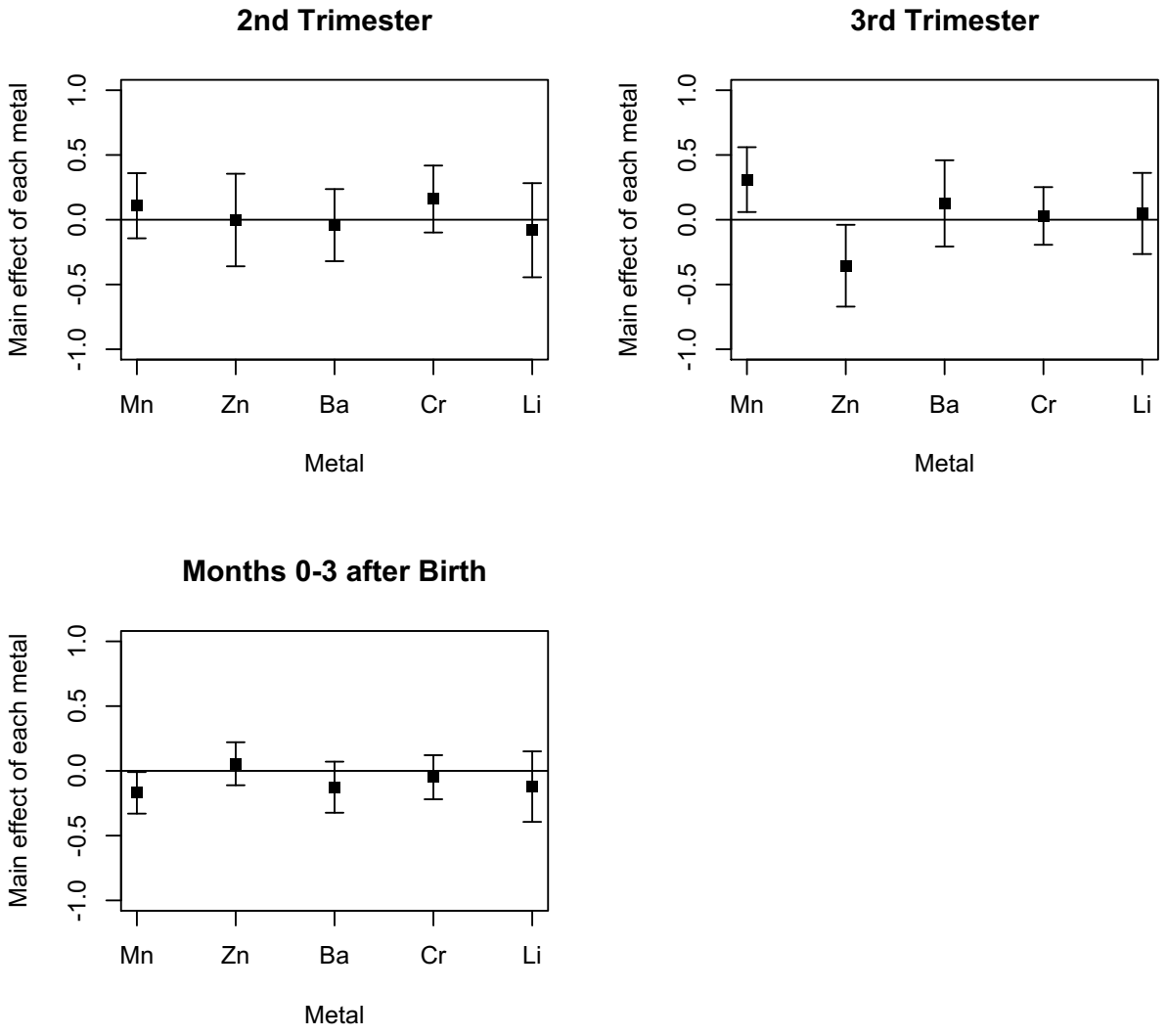

Figure 1.2: LKMR estimated main effect of each metal at three critical windows for ELEMENT data. Plot of the estimated relative importance of each metal, as quantified by the difference in the estimated effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures.

As both the LKMR estimated relative importance and the linear model indicate effects of Mn and Zn, we focus on those two metals when exploring the exposure-response relationship. Because the exposure response surface is five-dimensional, we use heat maps and cross-sectional plots to reduce dimensionality and graphically depict the exposure-response relationship. Figure 1.3 presents the plot of the posterior mean of the exposure-response surface of Mn and Zn at the median of Ba, Cd, Li estimated using LKMR. The

shape of the surface at the second trimester suggests an interaction between Mn and Zn, which will be further explored below. Also, the results suggest that the direction of the association changes at birth. At the third trimester, high Mn and moderate Zn exposures are associated with higher scores, while after birth, low Mn and a range of Zn exposures are associated with higher scores.

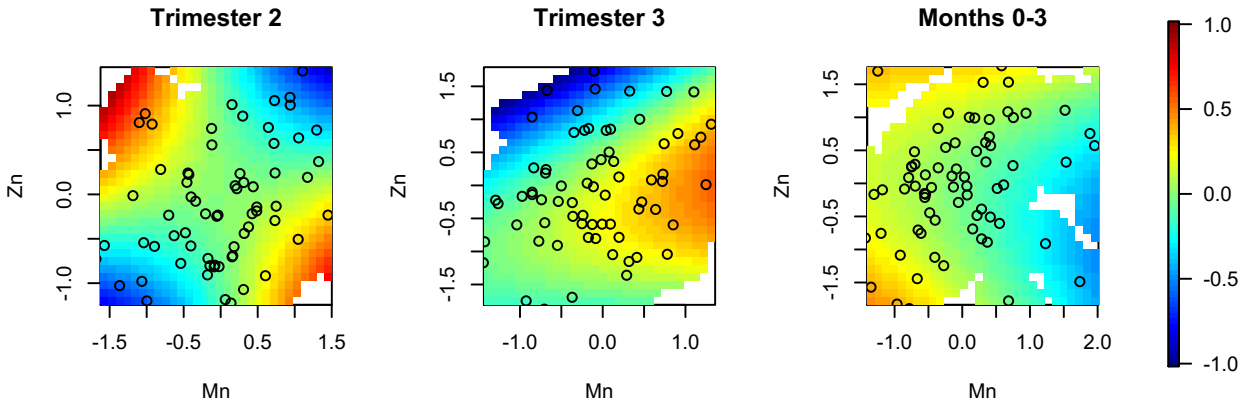

Figure 1.3: LKMR estimated time-specific exposure response functions applied to ELEMENT data. Plot of the estimated posterior mean of the exposure-response surface for Mn and Zn, at the median of Ba, Cd, Li.

To further reduce dimensionality, Figure 1.4 depicts the plot of the predicted cross-section of the exposure-response surface for Mn, at low and high Zn and median Ba, Cr, Li exposures. These results suggest that the association between Mn exposure and visual spatial score depends on exposure timing. Comparing the top panel (low Zn) to the bottom panel (high Zn), we detect a possible suggestion of a Mn-Zn interaction, specifically effect modification in the presence of higher Zn levels at the second trimester of pregnancy. At the second trimester, there is a positive association between Mn exposure and visual spatial score in the presence of low Zn levels. However, the association becomes negative in the presence of high Zn levels. Notably, this interaction is not suggested in the plot of the relative importance, where Mn and Zn are both non-significant at the 2nd trimester. In the cross-sectional plot, we also note evidence of a positive association between Mn and Zn before birth, and a negative one after birth. Lastly, the cross-sectional graphs suggest the effects are mainly linear, indicating that a quadratic kernel is sufficient to capture the

exposure-response relationship.

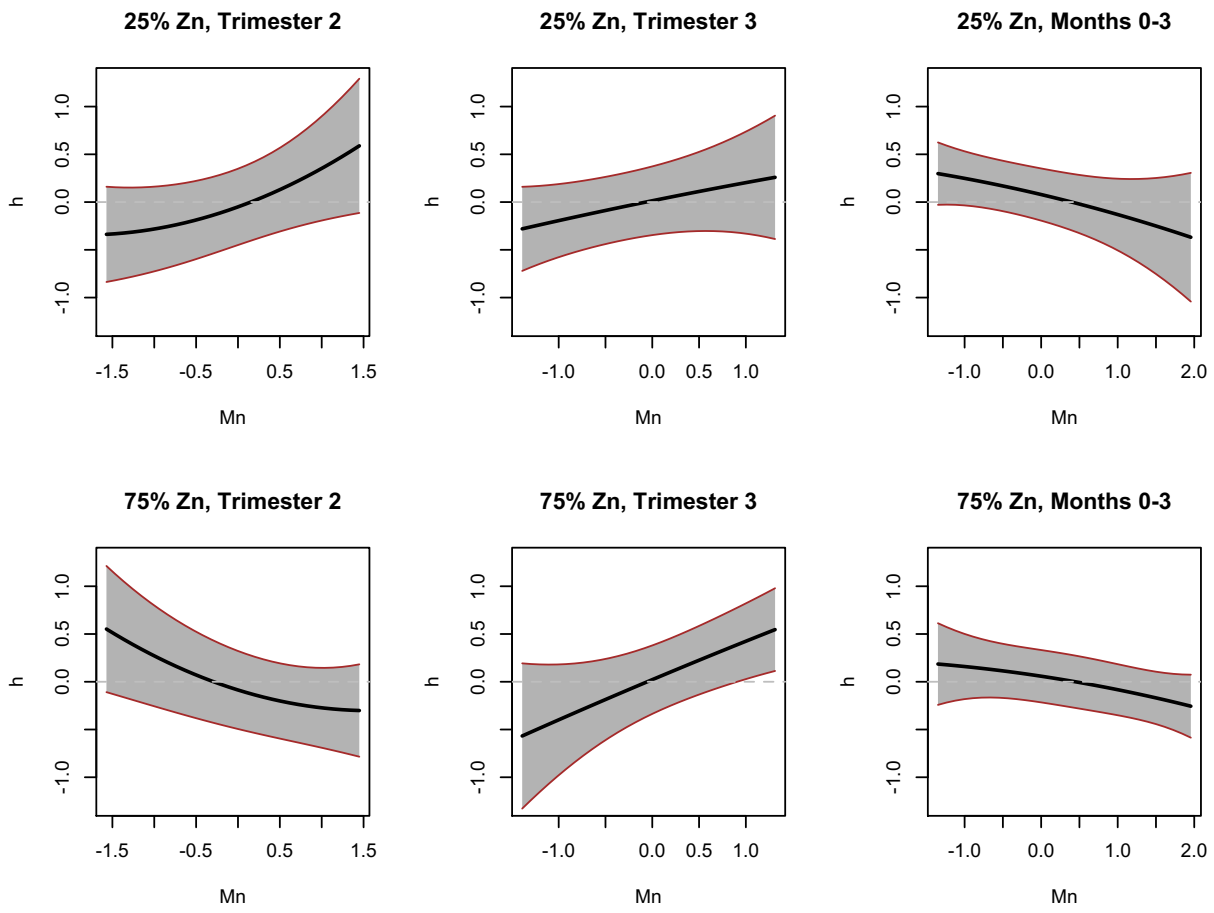

Figure 1.4: LKMR estimated time-specific exposure-response functions for Mn at low and high Zn levels applied to ELEMENT data. Plot of the cross-section of the estimated exposure-response surface for Mn, at low Zn exposure of 25th percentile (top panel) and high Zn exposure of 75th percentile (bottom panel), holding Ba, Cr, and Li constant at median exposures.

In Figure 1.5, we focus on the estimated interaction effect between Mn and Zn at the three critical time windows. This was quantified by estimating difference in effects for high (75th percentile) and low (25th percentile) Mn-Zn exposures. The results indicate that there is a significant Mn-Zn interaction for the second trimester, which was also evidenced by the linear model.

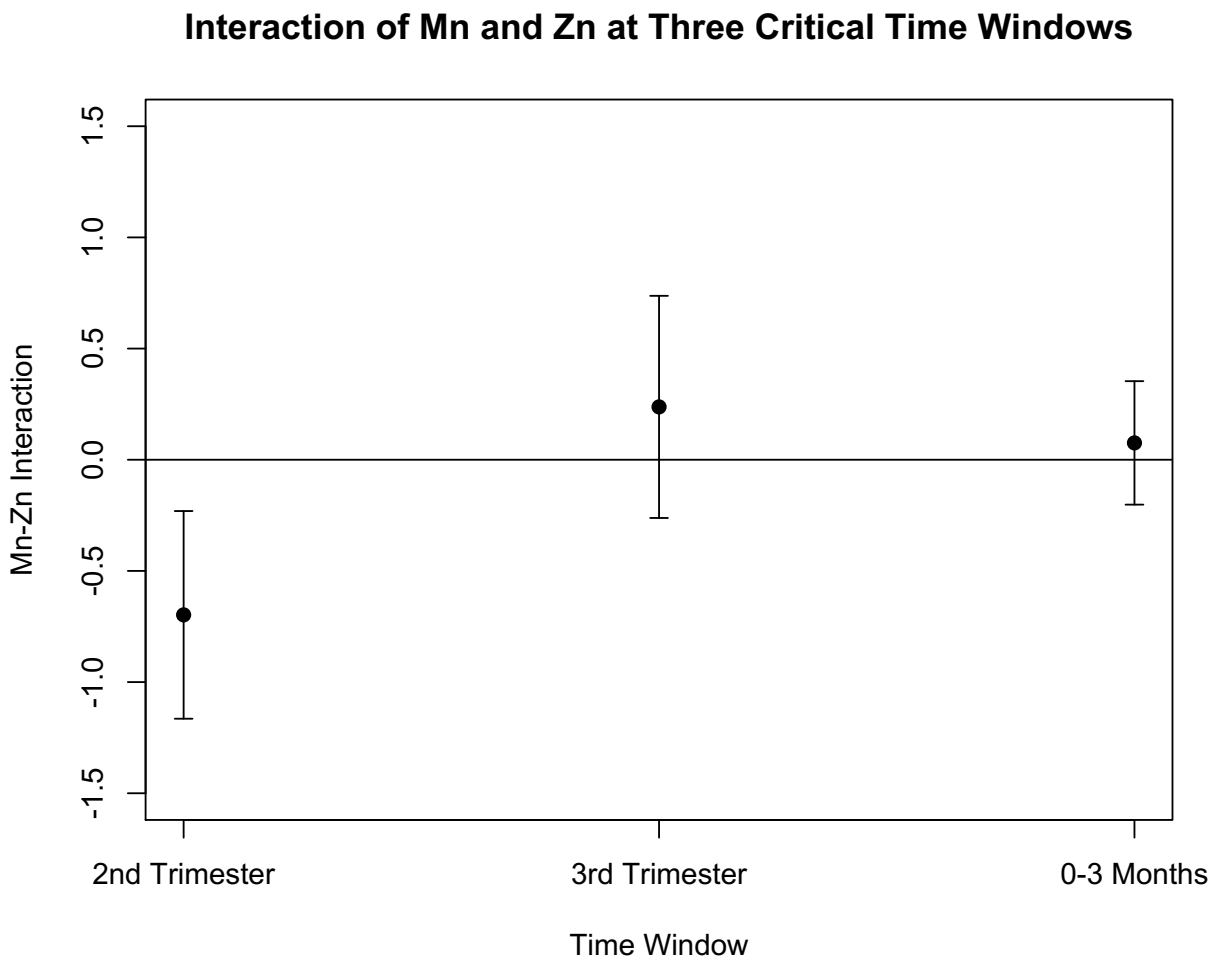

Figure 1.5: LKMR estimated Mn-Zn interaction at three critical windows for ELEMENT data. Plot of the estimated interaction effect between Mn and Zn, holding Ba, Cr, and Li constant at median exposures. This was quantified by estimating effects for high (75th percentile) and low (25th percentile) Mn-Zn exposures.

To complete our case study, we compare the results under LKMR to those obtained by BKMR applied using data from each critical window separately. First, in Figure 1.6, we estimate the relative importance of each metal under BKMR, which is analogous to Figure 1.2 under LKMR. The results suggest that when focused on Mn and Zn, only Zn exposure at the third trimester is significantly negatively associated with visual spatial score. This is markedly different from the results under LKMR and the linear model, which indicate a positive association of Mn at the third trimester, and negative associations for Mn and

Zn at Months 0-3. Figure 1.7 depicts the posterior mean of the exposure-response surface of Mn and Zn at the median of Ba, Cd, Li, which is analogous to Figure 1.3 under LKMR. In Trimester 2, there is little association between Mn and Zn exposure with neurodevelopment. However, under LKMR and the linear model, there was suggestion of an Mn-Zn interaction. In months 0-3 after birth, however, findings between LKMR and BKMR generally correspond, with higher scores associated with low Mn exposure across a range of Zn exposures. Lastly, Figure 1.8 depicts the predicted cross-sectional plot for BKMR, analogous to Figure 1.4 under LKMR, and suggests that no time windows have a significant interaction effect. Taken together, these findings further suggest that BKMR lacks the ability to detect a signal, which may be due to confounding by exposure at the other time points that the method does not account for.

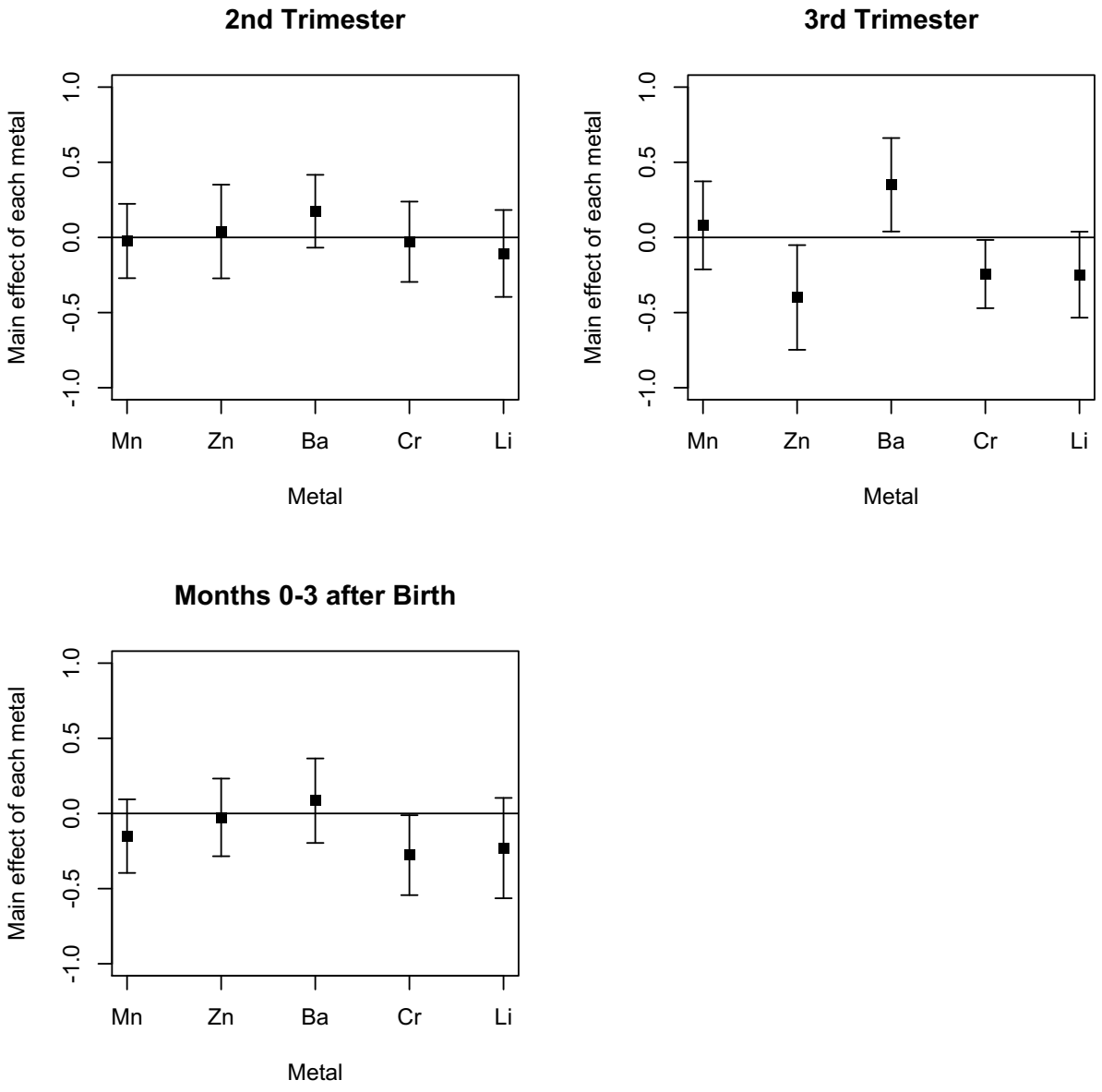

Figure 1.6: BKMR estimated main effect of each metal at three critical windows for ELEMENT data. Plot of the estimated relative importance of each metal, as quantified by the difference in the estimated effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures.

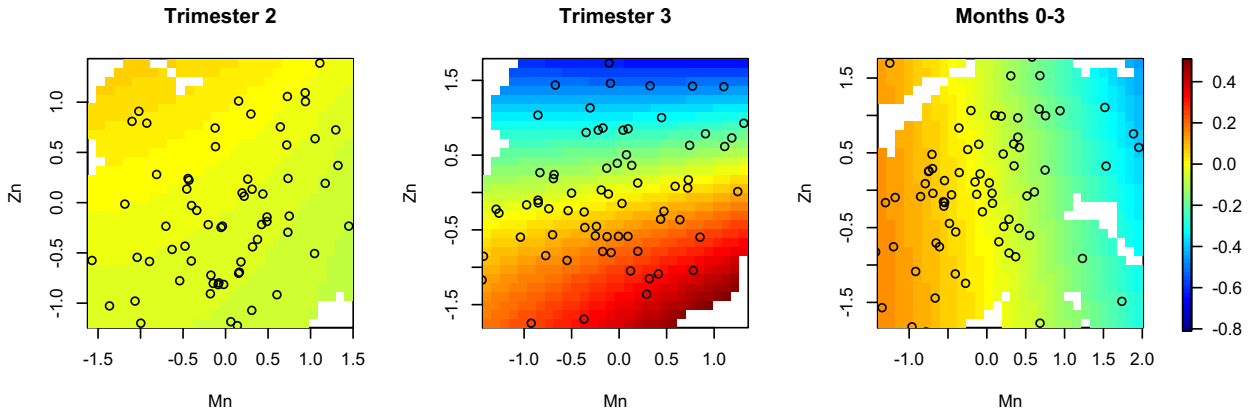

Figure 1.7: BKMR estimated time-specific exposure response functions applied to ELEMENT data. Plot of the estimated posterior mean of the exposure-response surface for Mn and Zn, at the median of Ba, Cr, Li.

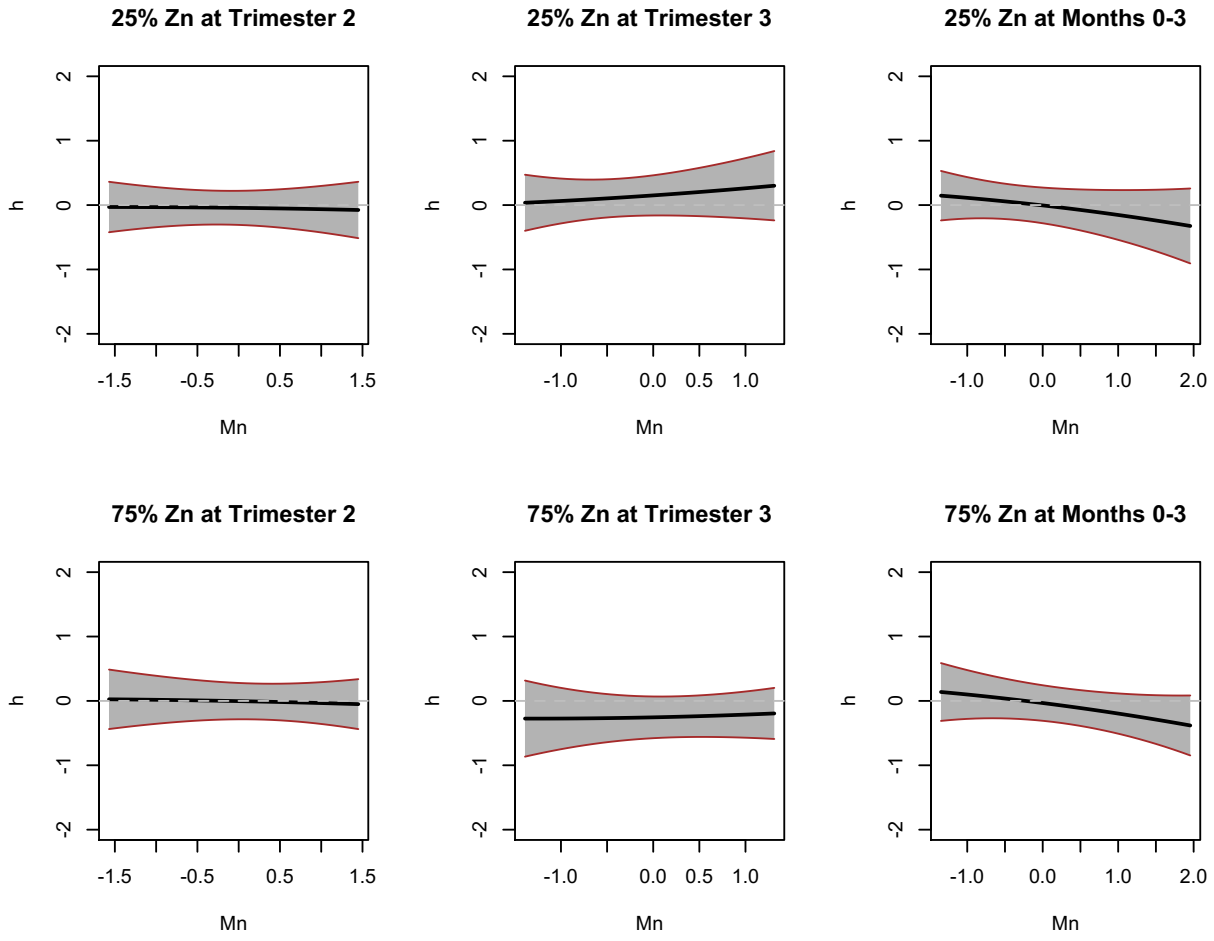

Figure 1.8: BKMR estimated Mn-Zn interaction at three critical windows for ELEMENT data. Plot of the estimated interaction effect between Mn and Zn, holding Ba, Cr, Li constant at median exposures. This was quantified by estimating effects for high (75th percentile) and low (25th percentile) Mn-Zn exposures.

1.6 Discussion and Conclusion

In this article, we have developed a lagged kernel machine regression model that uses Bayesian regularization to analyze data on time-varying exposures of environmental mixtures to identify critical windows of exposure in children's health. The kernel framework allows for a flexible specification of the unknown exposure-response relationship. We use a Bayesian formulation of the group lasso, which regularizes each kernel surface, and the fused lasso, which smoothes individual multivariate exposure-response surfaces

over time. Our method can account for auto-correlation of mixture components over time while exploring for the possibility of non-linear and non-additive effects of individual exposures. A key contribution of this article is the incorporation of the kernel machine framework into distributed lag modeling.

We demonstrated that the LKMR method achieves large gains over approaches that consider each critical window separately, particularly when serial correlation among the time-varying exposures is high. We applied LKMR to analyze associations between neurodevelopment and metal mixtures in the ELEMENT cohort. In the presence of complex exposure-response relationships that can vary with the timing of exposures, LKMR is a promising method to quantify health effects and identify time windows of susceptibility. LKMR, which uses information from neighboring time windows through penalization, is able to detect effect modification that was missed by BKMR. In the application of LKMR to the ELEMENT study, we detect an interesting interaction between manganese and zinc. At low levels of zinc, manganese exposure at the second trimester of pregnancy is positively associated with neurodevelopment. However, this positive association shifts after birth, at which point it is negatively associated with cognition. This suggests manganese functions as a trace element and an essential nutrient before birth, and is a toxicant after birth. Furthermore, this effect is not present under high exposure levels of zinc at the second trimester. The finely detailed interaction effect is captured by LKMR but not by BKMR, suggesting the potential existence of nuanced effects among other metals as well, which warrants further investigation.

As LKMR focuses on health outcomes at a single time point, a logical extension of the model would be to model the longitudinal health impact of exposures to time-varying metal mixtures. One may also be interested in adding variable selection to the model to identify the most important subsets of toxicants in their effects on health. With increasing sample size and complexities of the model, computationally efficient methods for fitting the model, such as variational Bayes (Ormerod and Wand, 2010), may be appropriate for improving computational efficiency.

To our knowledge, this is the first article on statistical methods for identifying critical exposure windows of multi-pollutant mixtures. The development of statistical methods that can handle the complexity of multi-pollutant mixtures whose effects may vary over time contributes to the limited knowledge on health effects of chemical mixtures, shedding light on interaction, effect modification and toxicity.

Mean Field Variational Bayesian Inference for Lagged Kernel Machine Regression in Children's Environmental Health

Shelley Han Liu

Department of Biostatistics

Harvard Graduate School of Arts and Sciences

Jennifer F. Bobb

Biostatistics Unit

Group Health Research Institute

Lourdes Schnaas

Division for Research in Community Interventions

National Institute of Perinatology, Mexico

Martha Tellez-Rojo

Center for Research in Nutrition and Health

National Institute of Public Health, Mexico

Manish Arora

Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Robert Wright
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Brent Coull
Department of Biostatistics
Harvard T.H. Chan School of Public Health

Matt P. Wand
School of Mathematical and Physical Sciences
University of Technology Sydney

2.1 Introduction

There is growing interest from environmental health institutes and regulatory agencies to quantify and assess the health impacts of exposure to toxicant mixtures. The National Institute for Environmental Health Sciences has set the study of environmental mixtures as a priority area (Carlin et al., 2013; Billionnet et al., 2012). It is hypothesized that exposure to mixtures of toxicants, such as heavy metals, may play a significant role in neurodevelopment in early life. There may be certain time windows of susceptibility, also called critical exposure windows, during which vulnerability to metal mixture exposures is increased. As there are many sequential developmental processes in fetal life and early childhood (Stiles and Jernigan, 2010), the health effects of heavy metal mixture exposures can be highly-dependent on exposure timing.

Liu et al. (2016) proposed Lagged Kernel Machine Regression (LKMR) to estimate the health effects of time-varying exposures to heavy metal mixtures, and identify critical exposure windows. Under LKMR, the non-linear and non-additive effects of time-varying mixture exposures are estimated while allowing for the effects to vary smoothly over time, similar to a distributed lag model. This was accomplished using a novel Bayesian penalization scheme that combines the group and fused lasso (Kyung et al., 2010; Yuan and Lin, 2006; Park and Casella, 2008; Huang et al., 2012) within a Bayesian kernel machine regression framework (Bobb et al., 2015). The flexible symmetric kernel within each group term, or time point, allows for the identification of sensitive time windows. Meanwhile, the penalization across time points for each subject functions to fuse each individual's time-varying exposures. The authors describe a Markov chain Monte Carlo (MCMC) algorithm using Gibbs sampling for LKMR. Due to the complexity of the LKMR model, computational time for updating parameters in the MCMC algorithm dramatically increases with the number of subjects or time points studied. As many iterations in the MCMC algorithm are needed to ensure that the chain converges to a stable posterior distribution for the parameters of interest, this can be computationally burdensome.

To reduce computation time, this article implements an approximation method, called mean field variational Bayes (MFVB) for LKMR analysis. Variational approximations are useful when standard sampling-based approaches to posterior approximation are impractical or infeasible. Variational Bayes, shorthand for variational approximate Bayesian inference, is a computationally efficient alternative to MCMC (Faes et al., 2011; Wand, 2014; Pham et al., 2013; Wand et al., 2011; Wand and Ormerod, 2012; Menictas and Wand, 2013; Goldsmith et al., 2011; Hall et al., 2011). Unlike MCMC, variational Bayes is a deterministic technique. While MCMC tends to converge slowly, variational Bayes provides a fast approximation to the true posterior.

The central idea behind variational Bayes is that the posterior densities of interest are approximated by other densities for which inference is more tractable. Suppose in a Bayesian model, we observe data $\mathbf{y}$, and are interested in the parameter vector $\boldsymbol{\theta}$. The density transform variational approach involves approximating the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ by another density, $q(\boldsymbol{\theta})$, and minimizing the Kullback-Liebler divergence (Ormerod and Wand, 2010). A common type of restriction for the q density is a non-parametric mean field approximation, which assumes $q(\boldsymbol{\theta})$ factorizes into $\prod_{i=1}^{M} q_i(\theta_i)$, for some partition $\theta_1, ..., \theta_M$ of $\boldsymbol{\theta}$. Under this restriction, we can derive explicit solutions for updating each product component, and develop a iterative process for obtaining simultaneous solutions.

In this paper, we focus on the implementation of mean field variational Bayes in the LKMR model. The paper is developed as follows: Section 2.2 provides a review of MFVB and kernel machine regression; Section 2.3 details the LKMR model; Section 2.4 describes the simulation studies; Section 2.5 applies the method to a children's environmental health study, and Section 2.6 provides the discussion and conclusion.

2.2 Review of mean field variational Bayes

Suppose we use a Bayesian paradigm to model the continuous parameter vector $\boldsymbol{\theta} \in \Theta$ corresponding to an observed data vector $\mathbf{y}$. The posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}) \equiv p(\mathbf{y}, \boldsymbol{\theta})/p(\mathbf{y})$ is used for Bayesian inference, where $p(\mathbf{y})$ is known as the marginal likelihood. It can be shown that the logarithm of the marginal likelihood is bound by:

$$log\, p(\mathbf{y}) = \int q(\boldsymbol{\theta}) log\left\{\frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right\} d\boldsymbol{\theta} + \int q(\boldsymbol{\theta}) log\left\{\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{y})}\right\} d\boldsymbol{\theta} \geq \int q(\boldsymbol{\theta}) log\left\{\frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right\} d\boldsymbol{\theta} \tag{2.1}$$

The integral,

$$\int q(\boldsymbol{\theta}) log\left\{\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{y})}\right\} d\boldsymbol{\theta} \geq 0 \tag{2.2}$$

is known as the Kullback-Leibler divergence between density q and $p(\cdot|\mathbf{y})$. This quantity is greater or equal to zero for all densities q, and equal to zero if and only if $q(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})$ almost everywhere. Therefore, the q-dependent lower bound on the marginal likelihood is:

$$\underline{p}(\mathbf{y}; q) = exp \int q(\boldsymbol{\theta}) log\left\{\frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right\} d\boldsymbol{\theta} \tag{2.3}$$

In variational approximation, we approximate the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ using a $q(\boldsymbol{\theta})$ for which $\underline{p}(\mathbf{y}; q)$ is more tractable than $p(\mathbf{y})$. By minimizing the Kullback-Liebler divergence between q and $p(\cdot|\mathbf{y})$, we are maximizing $\underline{p}(\mathbf{y}; q)$. We use approximate Bayesian inference under product density restrictions, called mean field variational Bayes. Under this non-parametric restriction, we assume that $q(\boldsymbol{\theta})$ can be factored into $\prod_{i=1}^{M} q_i(\theta_i)$ for some partition $\{\theta_1, ..., \theta_M\}$ of $\boldsymbol{\theta}$.

By maximizing the $log\, \underline{p}(\mathbf{y}; q)$ over each of the $q_1, ...q_M$, we obtain the optimal densities:

$$q_i^*(\theta_i) \propto exp[E_{-\theta_i} log\, p(\mathbf{y}, \boldsymbol{\theta})], i = 1, ..., M \tag{2.4}$$

$E_{-\theta_i}$ indicates expectation with respect to the density $\prod_{j \neq i} q_j \theta_j$. Using iteration, one can update each $q_i^*(\cdot)$ for $i = 1, ..., M$.

2.2.1 Review of kernel machine regression

We first review the kernel machine regression framework for estimating the effect of a complex environmental mixture at a single exposure time point. Suppose we observe data from *n* subjects. For each subject $i = 1, \ldots, n$, kernel machine regression (KMR) relates the continuous, normally distributed health outcome (Y_i) to M components of the exposure mixture $\mathbf{z}_i = (z_{1i}, ..., z_{Mi})$ through a nonparametric function, $h(\cdot)$, while controlling for p relevant confounders $\mathbf{x}_i = (x_{1i}, ..., x_{pi})$. The model is

$$Y_i = h(z_{1i}, ..., z_{Mi}) + \mathbf{x}_i^T \beta + \epsilon_i, \tag{2.5}$$

where $\boldsymbol{\beta}$ represents the effects of the potential confounders, and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. $h(\cdot)$ can be estimated parametrically or non-parametrically. We employ a kernel representation for $h(\cdot)$ in order to accommodate the possibly complex exposure-response relationship.

The unknown function, $h(\cdot)$, can be specified either through basis functions or through a positive definite kernel function $K(\cdot, \cdot)$. Under regularity conditions, Mercer's theorem (Cristianini and Shawe-Taylor, 2000) shows that the kernel function, $K(\cdot, \cdot)$, implicitly specifies a unique function space, H_k, that is spanned by a set of orthogonal basis functions. Thus, any function $h(\cdot) \in H_k$ can be represented through either a set of basis functions under the primal representation, or through a kernel function under the dual representation. The kernel function uses a similarity metric $K(\cdot, \cdot)$ to quantify the distance between the exposure profiles $\mathbf{z}_i$ between any two subjects in the study. For example, the Gaussian kernel quantifies similarity through the Euclidean distance; the polynomial kernel, through the inner product. Through specifying different kernels, one is able to control the complexity of the exposure-response function.

Liu et al. (2007) developed least-squares kernel machine semi-parametric regression for studying genetic pathway effects. The paper connects kernel machine methods and linear mixed models, demonstrating that (1) can be expressed as the mixed model

$$y_i \sim N(h_i + \mathbf{x}_i^T \beta, \sigma^2) \tag{2.6}$$

$$\mathbf{h} = (h_1, ..., h_n)^T \sim N\left[\mathbf{0}, \tau \mathbf{K}(\cdot, \cdot)\right], \tag{2.7}$$

where **K** is a kernel matrix with *i,j* element $K(\mathbf{z}_i, \mathbf{z}_j)$.

2.3 Lagged Kernel Machine Regression

Now we assume that exposures to a complex mixture are measured at multiple time-points, with the goal of identifying critical windows of exposure. Suppose we observe data from n subjects, each with an unique multi-pollutant exposure profile $\mathbf{z}_{it} = (z_{1i,t}, ..., z_{Mi,t})$. For each subject $i = 1, \ldots, n$ exposed to multi-pollutant mixtures at time intervals $t = 1, ..., T$, we use the following model to relate the health outcome to the clinical covariates and exposure covariates:

$$Y_i = \beta_0 + \sum_t h_t(z_{1i,t}, ..., z_{Mi,t}) + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i \tag{2.8}$$

$$Y_i = \beta_0 + \sum_t h_{i,t} + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i \tag{2.9}$$

The unknown function, $h(\cdot)$, represents the relationship between multi-pollutant exposures and the health outcome; each individual has an unique h_{it} at each time point.

Liu et al. (2016) details the LKMR model and the estimation of $h(\cdot)$; for brevity, we present a brief description of the hierarchical model here, with a primary focus on the MFVB approximation. LKMR uses Bayesian regularization to account for collinearity of mixture components while exploring for the possibility of non-linear and non-additive effects of individual exposures. Through a kernel machine framework which is incorporated into distributed lag modeling, the model allows for a flexible specification of the unknown exposure-response relationship. LKMR is the solution to this grouped, fused Lasso optimization:

$$\hat{\boldsymbol{h}}_{group,fused} = \arg\min_h (\mathbf{Y} - \mathbf{W}\boldsymbol{h} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{W}\boldsymbol{h} - \mathbf{X}\boldsymbol{\beta}) + \lambda_1 \sum_{t=1}^{T} \|\boldsymbol{h}_t\|_{G_t} + \lambda_2 \sum_{t=1}^{T-1} |\boldsymbol{h}_{t+1} - \boldsymbol{h}_t|_1$$

where $\|\boldsymbol{h}_t\|_{\mathbf{G_t}} = (\boldsymbol{h}_t^T \mathbf{G_t} \mathbf{h_t})^{\mathbf{1/2}}$. We define $\boldsymbol{h}_t = (h_{t,1}, ..., h_{t,n})$ and $\mathbf{G}_t = \mathbf{K}_t^{-1}$, where $\mathbf{K}_t$ denotes the kernel matrix for time t with *i,j* element $K_t(\mathbf{z}_i, \mathbf{z}_j)$. We choose $\mathbf{K}_t$ to be a

quadratic kernel, such that $K(\mathbf{z}, \mathbf{z'}) = (\mathbf{zz'} + 1)^2$.

The hierarchical model is represented as:

$$\mathbf{Y}|\boldsymbol{h}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N_n(\mathbf{X}\boldsymbol{\beta} + \sum_t \boldsymbol{h}_t, \sigma^2 \boldsymbol{I}_n) \tag{2.10}$$

$$\boldsymbol{h}|\tau_1^2, ..., \tau_T^2, \omega_1^2, ..., \omega_{T-1}^2 \sim N(\mathbf{0}, \Sigma_h) \tag{2.11}$$

$$\tau_1^2, ..., \tau_T^2 \sim gamma(\frac{n+1}{2}, \frac{{\lambda_1}^2}{2}) \tag{2.12}$$

$$\omega_1^2, ..., \omega_{T-1}^2 \sim \prod_{t=1}^{T-1} \frac{{\lambda_2}^2}{2} e^{\frac{-{\lambda_2}^2 \omega_t^2}{2}} \tag{2.13}$$

where $\tau_1^2, ..., \tau_T^2, \omega_1^2, ..., \omega_{T-1}^2, \sigma^2$ are mutually independent.

Figure 2.1 depicts the directed acyclic graph (DAG) of the Bayesian statistical model.

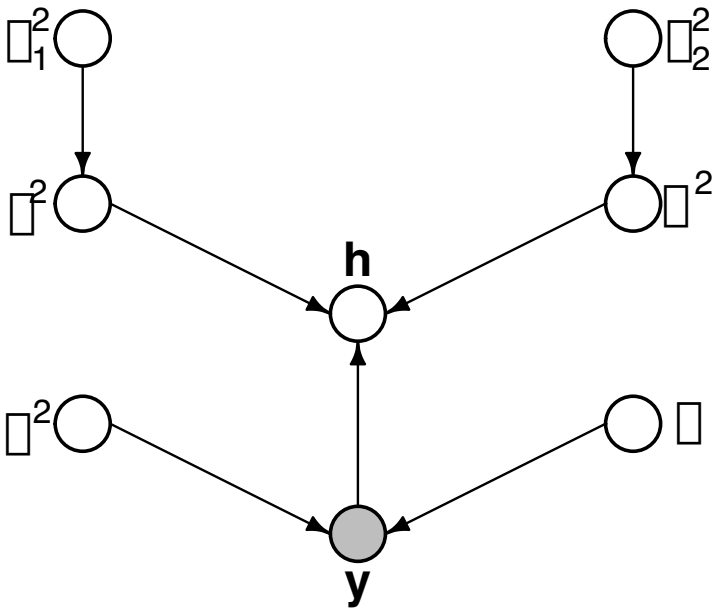

Figure 2.1: Distributed acyclic graph representation of Bayesian hierarchical model

It can be shown via standard algebraic manipulations that the full conditional distributions for this model are given by the following, from which Gibbs sampling can be readily implemented:

$$\boldsymbol{h}|rest \sim N\left\{(\frac{1}{\sigma^2}\boldsymbol{W}^T\boldsymbol{W} + \Sigma_h^{-1})^{-1}\frac{1}{\sigma^2}\boldsymbol{W}^T\,(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), (\frac{1}{\sigma^2}\boldsymbol{W}^T\boldsymbol{W} + \Sigma_h^{-1})^{-1}\right\} \tag{2.14}$$

$$\sigma^2|rest \sim \textit{Inverse Gamma}\left\{\frac{n+1}{2} + T, \frac{(\mathbf{Y} - \mathbf{W}\boldsymbol{h} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{W}\boldsymbol{h} - \mathbf{X}\boldsymbol{\beta})}{2} + \frac{\boldsymbol{h}^T\Sigma_h^{-1}\boldsymbol{h}}{2}\right\} \tag{2.15}$$

$$\boldsymbol{\beta}|rest \sim N\left\{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} - \mathbf{W}\boldsymbol{h}), \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right\} \tag{2.16}$$

$$\frac{1}{\tau_t^2}|rest \sim \textit{Inverse Gaussian}\left\{\sqrt{\frac{\lambda_1^2\sigma^2}{\|\boldsymbol{h}\|_{G_t}^2}}, \lambda_1^2\right\} \tag{2.17}$$

$$\frac{1}{\omega_t^2}|rest \sim \textit{Inverse Gaussian}\left\{\sqrt{\frac{\lambda_2^2\sigma^2}{\sum_{n=1}^{N}(h_{t+1,n} - h_{t,n})^2}}, \lambda_2\right\} \tag{2.18}$$

$$\lambda_1^2|rest \sim Gamma\Big\{T + r, \sum_{t=1}^{T} \tau_t^2/2 + \delta\Big\} \tag{2.19}$$

$$\lambda_2^2|rest \sim Gamma\Big\{T - 1 + r, \sum_{t=1}^{T-1} \omega_t^2/2 + \delta\Big\} \tag{2.20}$$

We now consider a MFVB approximation based on the following factorization for approximation of the joint posterior density function:

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{h}, \boldsymbol{\omega}^2, \boldsymbol{\tau}^2, \lambda_1^2, \lambda_2^2|\mathbf{Y}) \approx q(\boldsymbol{\beta})q(\sigma^2)q(\boldsymbol{h})q(\boldsymbol{\omega}^2)q(\boldsymbol{\tau}^2)q(\lambda_1)q(\lambda_2) \tag{2.21}$$

This leads to the following forms of the optimal q-densities:

$$q^*(\boldsymbol{\beta}) \sim N\Big\{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} - \mathbf{W}\mu_{q(h)}), (\mu_{q(1/\sigma^2)}\mathbf{X}^T\mathbf{X})^{-1}\Big\} \tag{2.22}$$

$$q^*(\sigma^2) \sim \textit{Inverse Gamma}\Big\{\frac{n(T+1)}{2}, \tag{2.23}$$

$$\frac{(\mathbf{Y} - \mathbf{X}\mu_{q(\beta)} - \mathbf{W}\mu_{q(h)})^T(\mathbf{Y} - \mathbf{X}\mu_{q(\beta)} - \mathbf{W}\mu_{q(h)}) + \mu_{q(h)}^T\mu_{q(\Sigma_{\tau^2,\omega^2})^{-1}}\mu_{q(h)}}{2}\Big\}$$

$$q^*(\boldsymbol{h}) \sim N\Big\{\big\{\mu_{q(1/\sigma^2)}\mathbf{W}^T\mathbf{W} + \mu_{q(\Sigma_{\tau^2,\omega^2})^{-1}}\big\}^{-1}\mu_{q(1/\sigma^2)}\mathbf{W}^T(\mathbf{Y} - \mathbf{X}\mu_{q(\beta)}), \tag{2.24}$$

$$\big\{\mu_{q(1/\sigma^2)}\mathbf{W}^T\mathbf{W} + \mu_{q(\Sigma_{\tau^2,\omega^2})^{-1}}\big\}^{-1}\Big\}$$

$$q^*(\frac{1}{\tau_t^2}) \sim \textit{Inverse Gaussian}\Big\{\Big\{\frac{\mu_{q(\lambda_1^2)}}{\mu_{q(\|h_t\|^2_{G_t})}}\Big\}^{1/2}, \mu_{q(\lambda_1^2)}\Big\} \tag{2.25}$$

$$q^*(\frac{1}{\omega_t^2}) \sim \textit{Inverse Gaussian}\Big\{\Big\{\frac{\mu_{q(\lambda_2^2)}}{\mu_{q(\sum_{n=1}^{N}(h_{t+1,n} - h_{t,n})^2)}}\Big\}^{1/2}, \mu_{q(\lambda_2^2)}\Big\} \tag{2.26}$$

$$q^*(\mu_{q(\lambda_1^2)}) \sim \textit{Gamma}\Big\{\frac{T(n+1)}{2} + r_1, \sum_{t=1}^{T}\frac{\mu_{q(\tau_t^2)}}{2} + \delta_1\Big\} \tag{2.27}$$

$$q^*(\mu_{q(\lambda_2^2)}) \sim \textit{Gamma}\Big\{T - 1 + r_2, \sum_{t=1}^{T-1}\mu_{q(\omega_t^2)} + \delta_2\Big\} \tag{2.28}$$

where the parameters are updated according to the algorithm in Figure 2.2.

Initialize: $\mu_{q(1/\sigma^2)} > 0, \mu_{q(\beta)} = 1, \mu_{q(1/\tau^2)} = 1, \mu_{q(1/\omega^2)} = 1.$

Cycle:

$$\Sigma_{q(h)} \leftarrow \left\{\mu_{q(1/\sigma^2)} W^T W + \mu_{q(\Sigma_{\tau^2,\omega^2})^{-1}}\right\}^{-1}$$

$$\mu_{q(h)} \leftarrow \mu_{q(1/\sigma^2)} \Sigma_{q(h)} W^T (Y - X\mu_{q(\beta)})$$

$$\mu_{q(\beta)} \leftarrow (X^T X)^{-1} X^T (Y - W\mu_{q(h)})$$

$$\mu_{q(1/\sigma^2)} \leftarrow \frac{n(T+1)}{(Y - X\mu_{q(\beta)} - W\mu_{q(h)})^T (Y - X\mu_{q(\beta)} - W\mu_{q(h)}) + \mu_{q(h)}^T \mu_{q(\Sigma_{\tau^2,\omega^2})^{-1}} \mu_{q(h)}}$$

$$\mu_{q(1/\tau_t^2)} \leftarrow \left\{\frac{\mu_{q(\lambda_1^2)}}{\mu_{q(\|h_t\|^2_{G_t})}}\right\}^{1/2}$$

$$\mu_{q(1/\omega_t^2)} \leftarrow \left\{\frac{\mu_{q(\lambda_2^2)}}{\mu_{q(\sum_{n=1}^N (h_{t+1,n} - h_{t,n})^2)}}\right\}^{1/2}$$

$$\mu_{q(\lambda_1^2)} \leftarrow \frac{T(N+1)+2r_1}{\sum_{t=1}^T \mu_{q(\tau_t^2)} + 2\delta_1}$$

$$\mu_{q(\lambda_2^2)} \leftarrow \frac{T-1+r_2}{\frac{1}{2}\sum_{t=1}^{T-1} \mu_{q(\omega_t^2)} + \delta_2}$$

until the increase is negligible.

Figure 2.2: MFVB algorithm for lagged kernel machine regression

2.3.1 Prediction at new exposure profiles

An important aim of environmental health studies is the characterization of the exposure-response surface. It is often of interest to predict health effects at unobserved exposure profiles. Suppose we are interested in predicting the exposure-response relationship for new profiles of metal mixture exposures, $\mathbf{z}_{new} = (\mathbf{z}_{new,1}, ..., \mathbf{z}_{new,M})$, for n_{new} subjects, where $\boldsymbol{h}_{new} = (h_{t,n+1}, ..., h_{t,n+n_{new}})^T$, $t = 1, ..T$ represent the desired predictions. In order to estimate $\boldsymbol{h}_{new}$, we first re-arrange the $\boldsymbol{h}$ vector so that
$\tilde{\boldsymbol{h}} = (h_{1,1}, ..., h_{1,n}, h_{2,1}, ..., h_{2,n}, h_{2,n+1}, ..., h_{2,n+n_{new}}, h_{1,n+1}, ..., h_{1,n+n_{new}})^T$. Because we have reordered the $\tilde{\boldsymbol{h}}$ vector, we need to similarly reorder the covariance matrix to correspond, and denote the reordered matrix by $\tilde{\Sigma}_h^{-1}$.
The joint distribution of observed and new exposure profiles is:

$$\begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{h}_{new} \end{pmatrix} \sim N\left\{ \mathbf{0}, \tilde{\Sigma}_h = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{12}^T & \tilde{\Sigma}_{22} \end{pmatrix} \right\} \tag{2.29}$$

where $\tilde{\Sigma}_{11}$ denotes the $2n$ x $2n$ matrix with $(i,j)^{th}$ element $K(\mathbf{z}_i, \mathbf{z}_j)$, $\tilde{\Sigma}_{12}$ denotes the n x n_{new} matrix with $(i, j_{new})^{th}$ element $K(\mathbf{z}_i, \mathbf{z}_{j_{new}})$, and $\tilde{\Sigma}_{22}$ denotes the n_{new} x n_{new} matrix with $(i_{new}, j_{new})^{th}$ element $K(\mathbf{z}_{i_{new}}, \mathbf{z}_{j_{new}})$. It follows that the conditional posterior distribution of $\boldsymbol{h}_{new}$ is:

$$\boldsymbol{h}^{\boldsymbol{new}} | \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\tau}^{\mathbf{2}}, \sigma^2 \sim N_{n_{new}} \Bigg\{ \tilde{\Sigma}_{12}^T \tilde{\Sigma}_{11}^{-1} \Big\{ \frac{1}{\sigma^2} \mathbf{W}^T \mathbf{W} + \tilde{\Sigma}_{11}^{-1} \Big\}^{-1} \frac{1}{\sigma^2} \mathbf{W}^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{2.30}$$

$$\tilde{\Sigma}_{12}^T \tilde{\Sigma}_{11}^{-1} \Big\{ \frac{1}{\sigma^2} \mathbf{W}^T \mathbf{W} + \tilde{\Sigma}_{11}^{-1} \Big\}^{-1} \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12} + \tilde{\Sigma}_{22} - \tilde{\Sigma}_{12}^T \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12} \Bigg\}$$

In order to reduce computation time, we approximate the posterior mean and variance of $\boldsymbol{h}_{new}$ based on the estimated posterior mean of the other parameters.

2.4 Simulation study

We conducted simulation studies to evaluate the performance of the proposed MFVB inference procedure for estimating critical exposure windows of environmental mixtures. Our simulation study considered a three-toxicant scenario, where two toxicants exerted

a gradual non-additive and non-linear effect over four time windows. We used the following model: $y_i = x_i^T\beta + \sum_t h_t(\mathbf{z}_{it}) + e_i$, where $e_i \sim N(0,1)$ and $x_{1i} \sim N(10,1)$ and $x_{2i} \sim Bernoulli(1, 0.5)$. We simulated auto-correlation within toxicant exposures Z_m across time, and correlation between toxicants, using the Kronecker product for the exposure correlation matrix. Three choices for auto-correlation within toxicants were considered: high (0.8), medium (0.5) and low (0.2). The exposure-response function $h(z_i)$ was simulated as quadratic with two-way interactions. We simulated no effect of exposure to the environmental mixture at Time 1, and a gradual increasing effect was simulated from Time 2 to 4. We assume $h_t(\mathbf{z}) = \alpha_t h(\mathbf{z})$, where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 0.5, 0.8, 1.0)$ and $h(\mathbf{z}) = z_1^2 - z_2^2 + 0.5z_1z_2 + z_1 + z_2$. In conducting the analysis, exposure covariates and confounder variables were centered and scaled.

Table 2.1 presents the results of this simulation, for sample size of 100. We compared the performance of MFVB approximation to that of Bayesian MCMC. For each simulated data set, to assess the performance of the model for the purposes of estimating the time-specific exposure-response function, we regressed the predicted $\widehat{\boldsymbol{h}}$ on $\boldsymbol{h}$ for each time point. We present the intercept, slope and R^2 of the regressions over 100 simulations. Good estimation performance occurs when the intercept is close to zero, and the slope and R^2 are both close to one. We also present the root mean squared error (RMSE) and the coverage (the proportion of times the true $h_{i,t}$ is contained in the posterior credible interval). Furthermore, we present the width of the 95% posterior credible interval. Notably, the RMSE is generally smaller under MFVB as compared with MCMC. The reduction is most apparent situations of high autocorrelation among mixture components, where the reduction in RMSE ranges from 9-24%. We also see that the intercept, slope and R^2 tend to be very similar under MFVB and MCMC inference.

Table 2.1 Simulation results, regression of $\widehat{\boldsymbol{h}}$ on h for MFVB vs. MCMC

h function	Time window	Intercept	Slope	R^2	RMSE
MFVB	1	-0.02	N/A	N/A	0.38
	2	-0.01	0.93	0.90	0.41
AR-1 = 0.8	3	0.00	0.99	0.95	0.48
	4	0.00	0.98	0.97	0.43
MCMC	1	0.00	N/A	N/A	0.47
	2	0.00	0.98	0.85	0.51
AR-1 = 0.8	3	0.00	1.00	0.93	0.56
	4	0.00	1.01	0.99	0.46
MFVB	1	0.00	N/A	N/A	0.32
	2	-0.01	0.93	0.91	0.38
AR-1 = 0.5	3	0.00	0.97	0.96	0.38
	4	0.00	0.99	0.98	0.37
MCMC	1	0.00	N/A	N/A	0.37
	2	0.00	0.96	0.90	0.40
AR-1 = 0.5	3	0.00	0.98	0.96	0.39
	4	0.00	1.01	0.98	0.38
MFVB	1	-0.01	N/A	N/A	0.32
	2	-0.01	0.93	0.92	0.36
AR-1 = 0.2	3	0.01	0.95	0.96	0.36
	4	0.00	0.97	0.98	0.37
MCMC	1	0.01	N/A	N/A	0.35
	2	0.01	0.96	0.92	0.36
AR-1 = 0.2	3	0.01	0.97	0.97	0.35
	4	-0.01	0.99	0.98	0.36

Performance of estimated $h_t(\mathbf{z}_i)$ across 100 simulated datasets. RMSE denotes the root mean squared error of the $\hat{h}$ as compared to h. Coverage denotes the proportion of times that the true h falls within in the 95% posterior credible interval of each time point.

We next conduct a simulation to study the effect of varying sample sizes on estimated posterior credible interval width and coverage. In Figure 2.3, the same three-toxicant scenario was considered as in Table 2.1, but for sample sizes of N = 100, 200, 300, 500, 800. Because of the computational infeasibility of applying the MCMC procedure to larger datasets, it was not used for sample sizes of N = 500 and 800. The $\boldsymbol{h}$ contains the ag-

gregated information for h_1, h_2, h_3, h_4. We note that for $\boldsymbol{h}$, the estimated 95% posterior credible interval width is about half as small under MFVB as under MCMC. The interval width is also shorter for $\boldsymbol{\beta}$ and σ^2. As sample sizes increase, the interval widths estimated under both MFVB and MCMC shrink. Coverage, the proportion of times the true parameter falls into the posterior credible interval, is high for $\boldsymbol{h}$ across the range of sample sizes. It ranges from 98% for N = 100, to 100% for N = 800. We note that the coverage of σ^2 increases substantially under MFVB for increasing sample sizes, changing from 32% for N = 100 to 86% for N = 800. Coverage of $\boldsymbol{\beta}$ increases from 86% for N = 100 to 96% for N = 800.

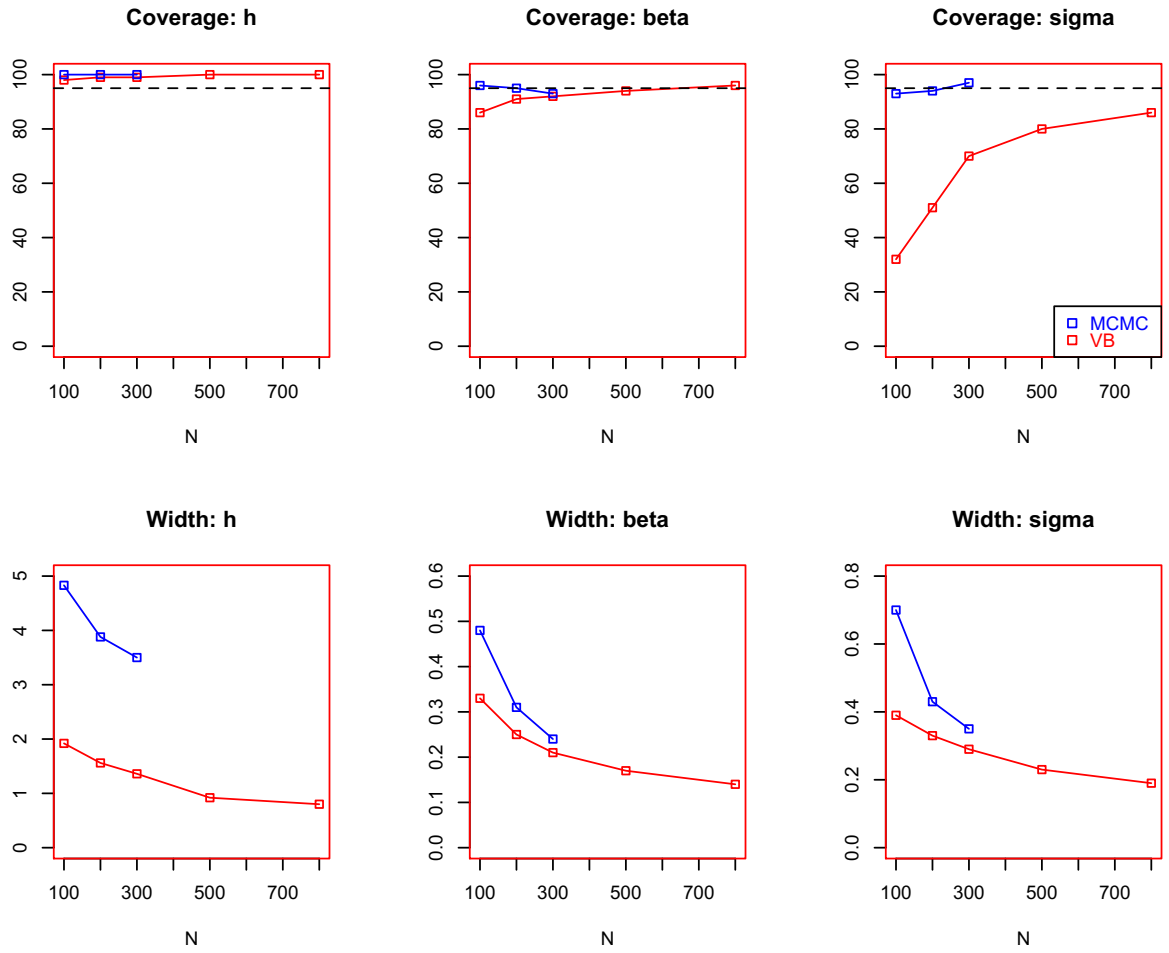

Figure 2.3: Coverage and posterior credible interval width of key parameters, using MFVB and MCMC. Performance of estimated across 100 simulated datasets for key parameters $\boldsymbol{h}, \boldsymbol{\beta}$ and σ^2 across a range of sample sizes (N = 100, 200, 300, 500, 800). Width denotes the length of the 95% posterior credible interval. Coverage denotes the proportion of times that the true parameter falls within in the 95% posterior credible interval. The dotted horizontal line marks 95%.

Lastly, Table 2.2 records the average computation time for MFVB and MCMC methods under the simulations in Table 2. In general, the MFVB procedure is about three hundred times faster than MCMC estimation. For example, for a sample size of N = 300, only 27 minutes is required under MFVB, whereas 3.5 days is required under MCMC.

Table 2.2 Average time in minutes for MFVB and MCMC methods applied to simulation case

Method	N = 100	N = 200	N = 300	N = 500	N = 800
MFVB	0.49	4.71	26.6	81.4	365
MCMC	175	1409	4990	N/A	N/A

Performance across 100 simulated datasets.

2.5 Application

We applied the MFVB procedure to analyze the association between birthweight and time-varying metal mixture exposures in the PROGRESS study conducted in Mexico City. The primary outcome was the Fenton z-scored birthweight in kilograms. Exposures to 11 metals were measured in the mother's blood at the second and third trimesters of pregnancy as well as birth. These metals include arsenic (As), cadium (Cd), cobalt (Co), chromium (Cr), cesium (Cs), copper (Cu), manganese (Mn), lead (Pb), antimony (Sb), selenium (Se) and zinc (Zn). Through exploratory analysis, we found high levels of correlation between Cu-Se, Cu-Zn, Zn-Se at the third trimester, which were 0.83, 0.87 and 0.85, respectively. Furthermore, at the second trimester, Cu-Zn had correlation of 0.91. This led us to remove Se and Zn from the analysis, leading to a final panel of 9 metals (As, Cd, Co, Cr, Cs, Cu, Mn, Pb, Sb).

We controlled for socioeconomic status (3 categories: low, middle, high), mother's hemoglobin during the second trimester of pregnancy, mother's educational level ($<$ high school, high school, $>$ high school), child gender, mother's WASI IQ, mother's age and mother's pre-pregnancy BMI. In our analysis, metal exposure levels were logged, then centered and scaled. Confounder variables were also centered and scaled. We considered all subjects with complete data in confounder variables, metal exposures and outcome, which resulted in N = 391.

As a primary analysis, we considered a linear regression model that simultaneously regressed birthweight on confounders and metal exposures at all time points. Several metals were identified as significant at the $\alpha = 0.05$ level. At the second trimester of pregnancy, Co was positively associated with birthweight (p = 0.031). At the third trimester, Mn had a positive association with birthweight (p = 0.018) while Pb and Cs had a negative ones (p = 0.021 and p = 0.047, respectively). Lastly, at birth, Cd was negatively associated with birthweight (p = 0.044). Through this linear model, there was also evidence of several two-way interactions: some examples include a positive Co-Cs interaction (p = 0.010) and a negative Cu-As interaction (p = 0.034) at the second trimester. There was also suggestion of a positive Cu-As interaction (p = 0.063) at birth.

Using the MFVB inference procedure, we first estimated the relative importance of each metal, as shown in Figure 2.4. Relative importance is quantified by the difference in the estimated main effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures. The results are similar to that of the simple linear regression, although the effects of several metals appear to be suggested but non-significant, as the confidence intervals cover zero. For example, there are suggestions of the positive effect of Co at the 2nd trimester, positive effect of Mn at the third trimester, negative effect of Pb at the third trimester, as well as negative effect of Cd at birth.

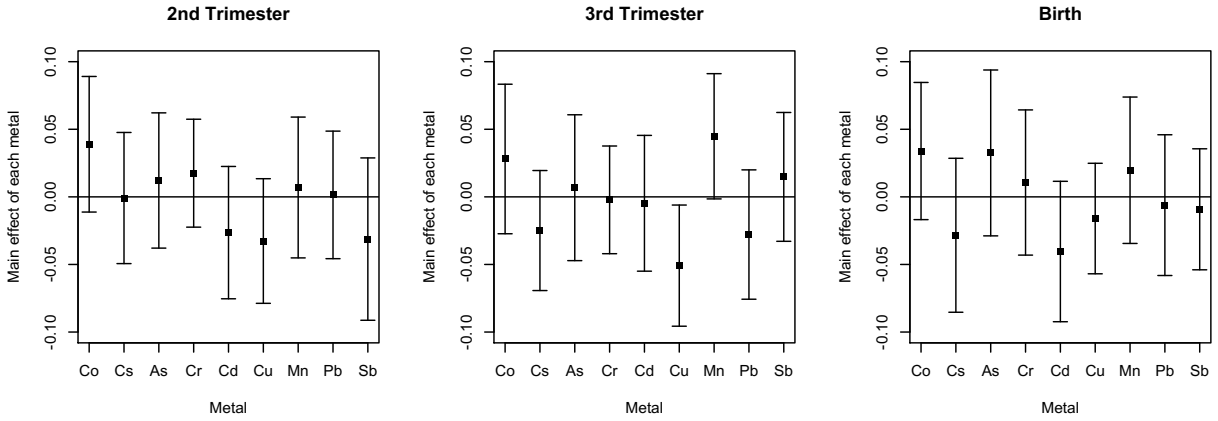

Figure 2.4: Estimated time-specific main effects applied to PROGRESS data. Plot of the relative importance of Co, Cs, As, Cd, Cr, Cu, Mn, Pb, and Sb. Relative importance is quantified by the difference in the estimated main effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures.

As both the MFVB procedure at the linear model suggested significant effects of Co and Cs, we next focus on those two metals when exploring the exposure-response relationship. Because the exposure response surface is nine-dimensional, we use heat maps and cross-sectional plots to reduce dimensionality and graphically depict the exposure-response relationship. Figure 2.5 presents the plot of the predicted exposure-response surface for Co and Cs, estimated using MFVB at the median of all other metal exposures. The shape of the surface at the second and third trimesters suggests an interaction effect.

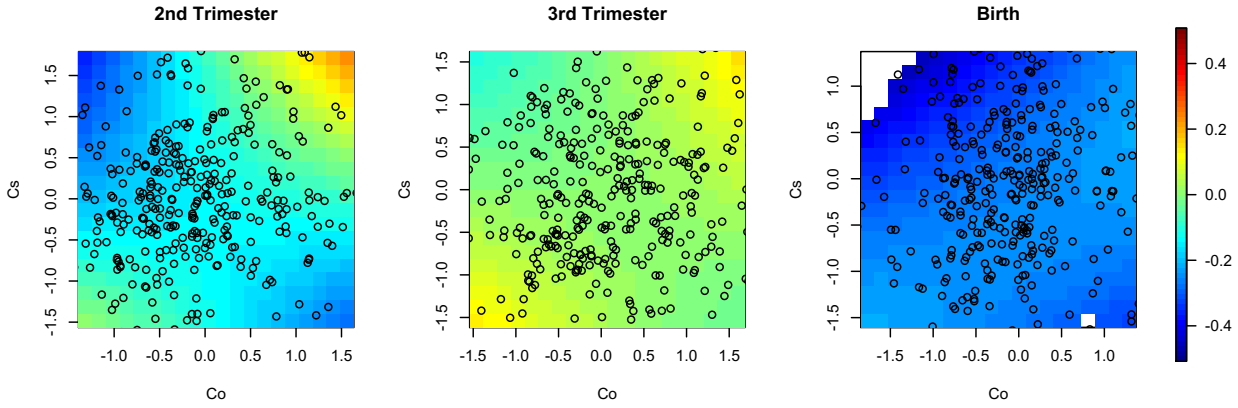

Figure 2.5: Estimated time-specific Co-Cs exposure response functions applied to PROGRESS data. Plot of the estimated posterior mean of the exposure-response surface for Co and Cs, at the median of As, Cd, Cr, Cu, Mn, Pb, and Sb.

To explore this, in Figure 2.6 we plot the predicted cross-section of the exposure-response surface for Co, at low and high Cs, holding all other metal exposures at their median. Comparing the top panel (low Cs) to the bottom panel (high Cs), we detect a possible suggestion of a Co-Cs interaction, specifically effect modification in the presence of higher Cs levels at the second and third trimesters of pregnancy. In the presence of higher Cs levels, there appears to be a positive linear relationship between Co exposures and birthweight.

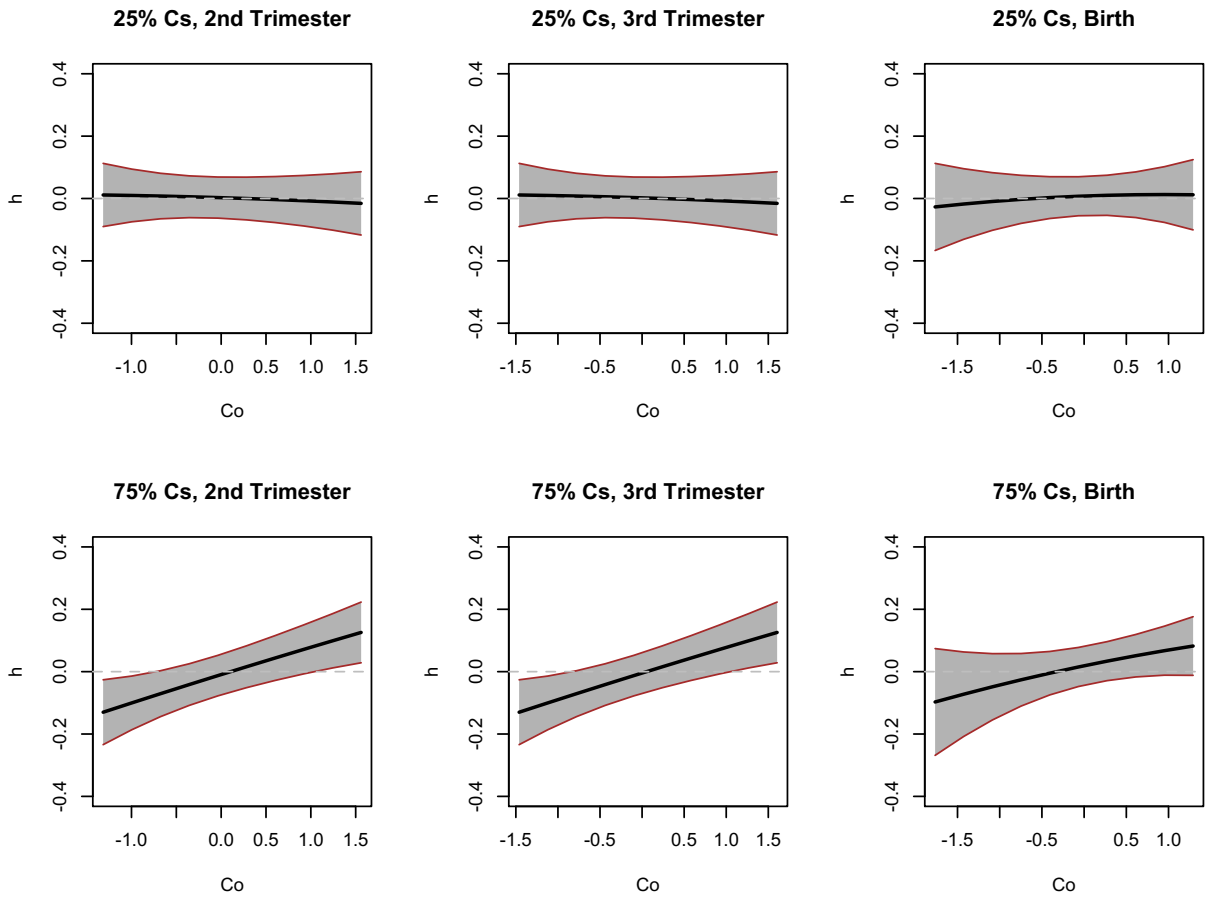

Figure 2.6: Estimated time-specific exposure-response functions for Co at low and high Cs levels applied to PROGRESS data. Plot of the cross-section of the estimated exposure-response surface for Co, at low Cs exposure of 25th percentile (top panel) and high Cs exposure of 75th percentile (bottom panel), holding As, Cd, Cr, Cu, Mn, Pb, and Sb constant at median exposures.

The significant Co-Cs interaction effect at the second and third trimesters is further evidenced by Figure 2.7, which quantifies this interaction by estimating effects for high (75th percentile) and low (25th percentile) Co-Cs exposures. Lastly, as the cross-sectional graphs suggest the effects are mainly linear, a quadratic kernel appears to sufficiently capture the exposure-response relationship.

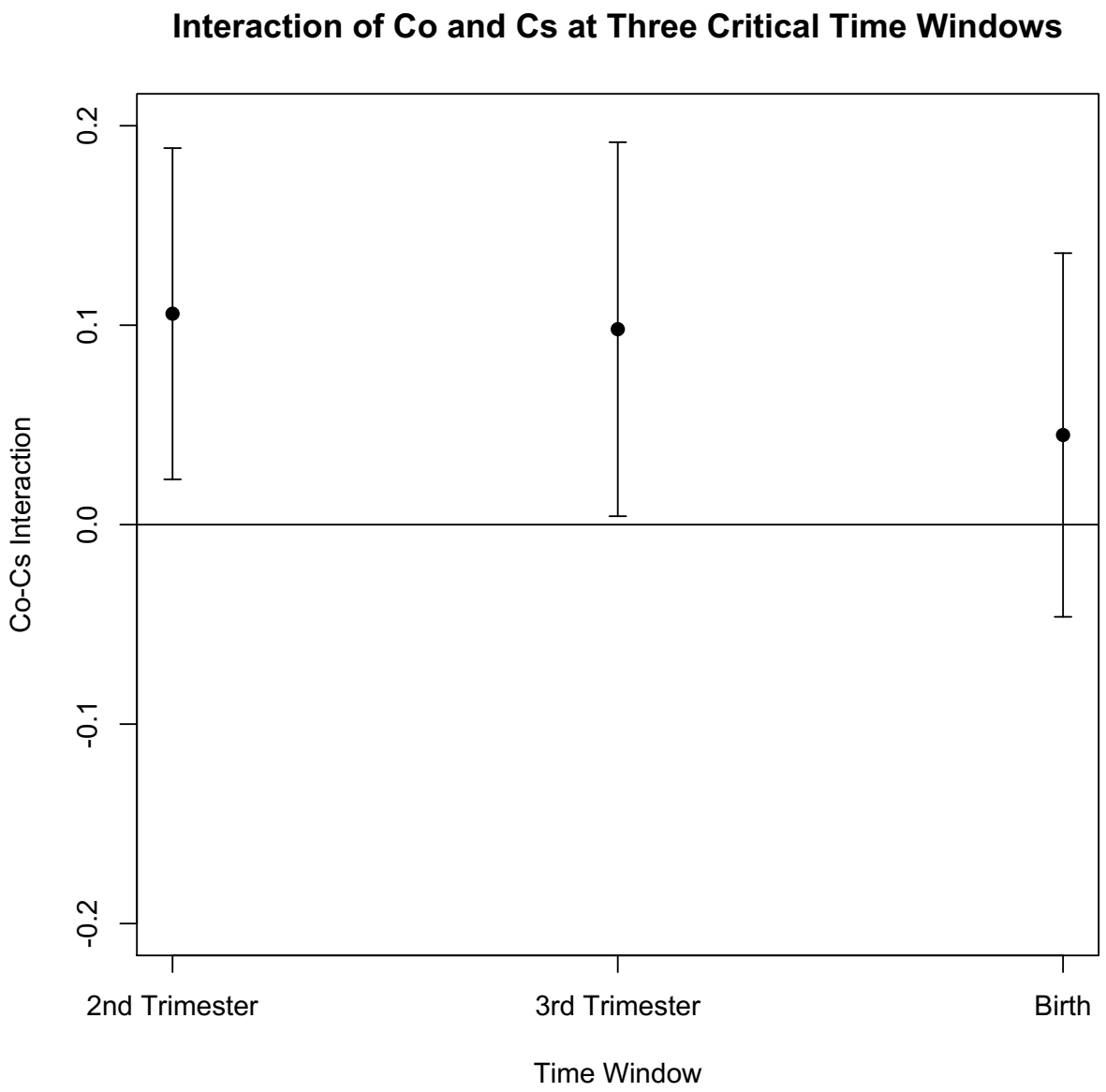

Figure 2.7: Estimated Co-Cs interaction at three critical windows for PROGRESS data. Plot of the estimated interaction effect between Co and Cs, holding As, Cd, Cr, Cu, Mn, Pb, and Sb constant at median exposures. This was quantified by estimating effects for high (75th percentile) and low (25th percentile) Co-Cs exposures.

The analysis also suggested a possible weaker interaction between Cu and As. The estimated Cu-As exposure-response surface, at the median of all other metal exposures, is plotted in Figure 2.8. Figure 2.9 illustrates the cross-sectional plot, while Figure 2.10 plots the estimated interaction effects. Taken together, they hint at an interaction effect

at the second trimester of pregnancy. Cu exposure in the presence of higher As levels seems to be negatively associated with birthweight at the second trimester of pregnancy. This negative interaction effect was found to be significant in the linear model during the preliminary analysis.

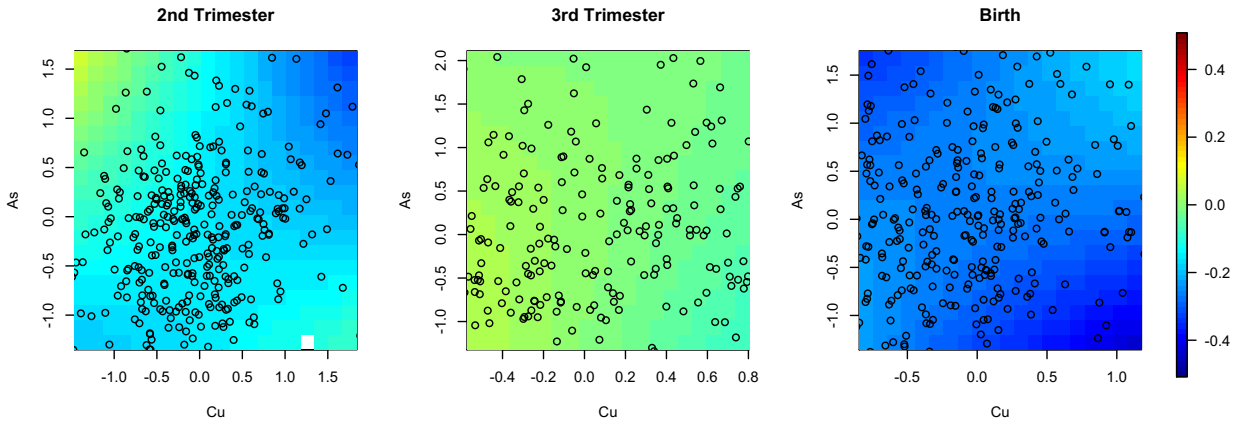

Figure 2.8: Estimated time-specific Cu-As exposure response functions applied to PROGRESS data. Plot of the estimated posterior mean of the exposure-response surface for Cu and As, at the median of Cd, Cr, Co, Cs, Mn, Pb, and Sb.

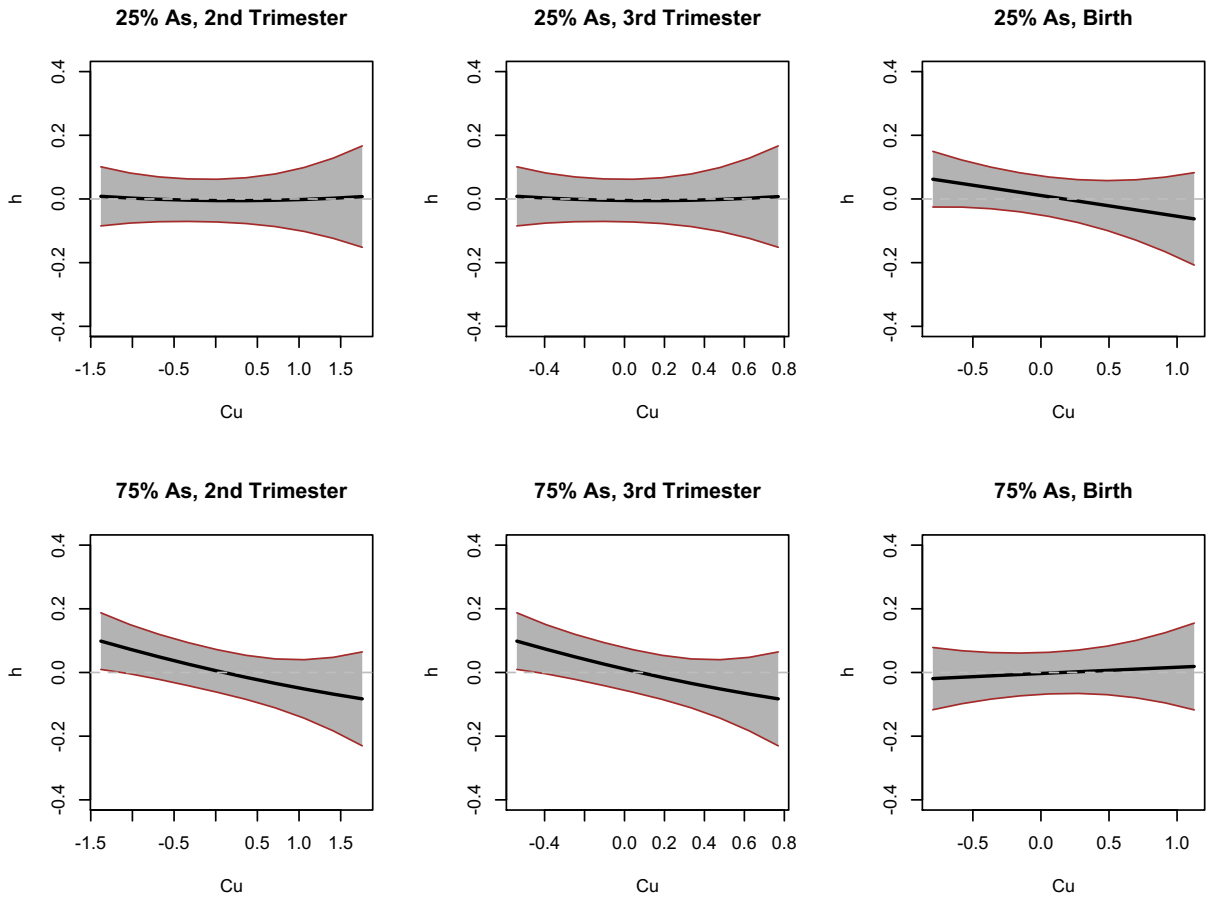

Figure 2.9: Estimated time-specific exposure-response functions for Cu at low and high As levels applied to PROGRESS data. Plot of the cross-section of the estimated exposure-response surface for Co, at low Cs exposure of 25th percentile (top panel) and high Cs exposure of 75th percentile (bottom panel), holding Cd, Cr, Co, Cs, Mn, Pb, and Sb constant at median exposures.

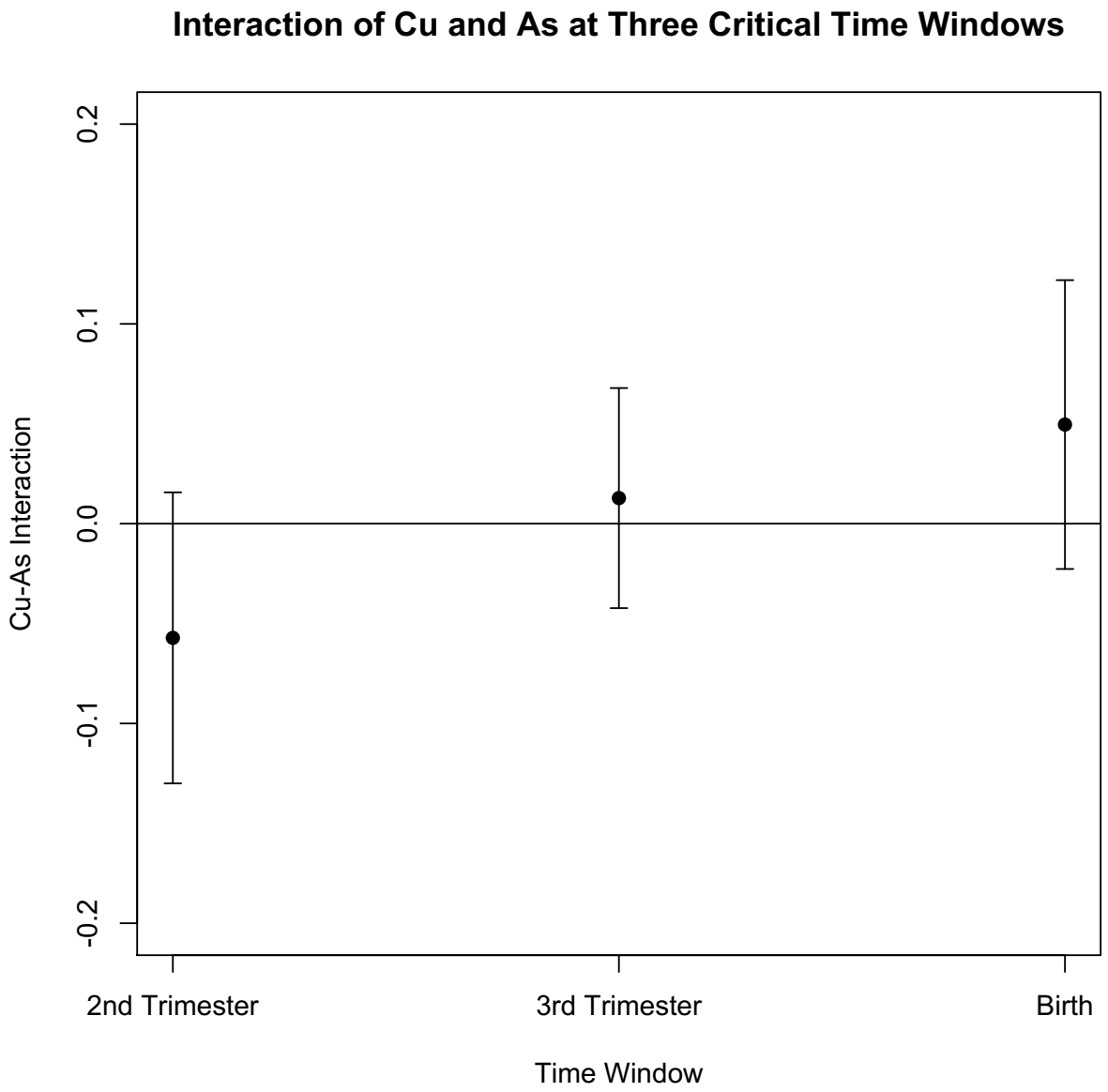

Figure 2.10: Estimated Cu-As interaction at three critical windows for PROGRESS data. Plot of the estimated interaction effect between Cu and As, holding Cd, Cr, Co, Cs, Mn, Pb, and Sb constant at median exposures. This was quantified by estimating effects for high (75th percentile) and low (25th percentile) Cu-As exposures.

2.6 Discussion and Conclusion

In this article, we have developed a MFVB inference procedure for the LKMR model, which allows for computationally efficient analysis of large environmental health

datasets. A key contribution of the article is the dramatic decrease in computational time required for MFVB as compared with MCMC, while also maintaining high accuracy. We applied the MFVB algorithm to analyze a prospective cohort study of children's environmental health in Mexico City, where we found evidence of interaction and effect modification between multiple pairs of metal exposures. Specifically, we identified a possible positive interaction between cobalt and cesium at the second and third trimesters of pregnancy, as well as a potential positive association between cobalt and birthweight at the second trimester. As cobalt is part of the vitamin B12 complex, evidence of cobalt exposure may be suggestive of healthy dietary habits which could lead to a higher birthweight. We also found a possible hint of Cu-As interaction at the second trimester of pregnancy, suggesting that Cu exposure in the presence of higher As levels is negatively associated with birthweight. As the literature on metal mixture exposures is sparse, these detected potential effects could serve as a starting point for future environmental health studies.

We note that the computational efficiency of the algorithm allowed us to analyze a prospective cohort study of moderate size with ease, which in turn provides us power to investigate a large panel of 9 metal exposures. In a previous study, we were limited by the computational burden to analyze a study of 81 subjects, which limited us to exploring only exploring 5 metal exposures due to a lack of power to detect effects of a larger panel (Liu et al., 2016).

Simulations demonstrated that MFVB is on average over 300 times faster than MCMC. This translates into dramatic computational gains for large datasets. For example, for a dataset of N = 800, which is of reasonable size for prospective cohort studies, inference using MFVB takes 6.1 hours. Estimation using MCMC would take 76 days - an impossible time hindrance.

Notably, we showed that MFVB inference maintains high accuracy for the key parameter of interest, $\boldsymbol{h}$, which quantifies the unknown exposure-response relationship.

Under varying simulation scenarios, MFVB estimates $\boldsymbol{h}$ well, often with smaller RMSE than using MCMC estimation. Posterior credible interval coverage is consistently 98-100% for $\boldsymbol{h}$ under MFVB for a range of sample sizes.

We also demonstrated that the coverage of the variance parameter, σ^2, varies significantly across the range of sample sizes. For small datasets, coverage can be poor, but increases considerably for larger sample sizes, reaching 86% under N = 800. When sample sizes are small, one could easily use MCMC estimation; thus, the coverage would not be a concern. However, MFVB inference was specifically created for large sample sizes, when it would be nearly impossible to conduct analysis under MCMC without a significant time burden. In the situation of large sample sizes, the coverage of σ^2 is much higher, albeit still smaller than than of MCMC. This is a small drawback to the MFVB inference. However, we believe the minor reduction in coverage of σ^2 in MFVB as compared with MCMC approaches is over-ridden by the dramatic increase in computational efficiency.

We also note that the width of the posterior credible intervals for $\boldsymbol{\beta}, \sigma^2, \boldsymbol{h}$ are smaller under MFVB inference than under MCMC estimation. For $\boldsymbol{h}$, the intervals are approximately half as small. It is known that due to the form of the Kullback-Leibler divergence used in the variational Bayes framework (Bishop, 2006; Rue et al., 2009; Wang and Titterington, 2005), the procedure can underestimate the posterior variance. While this may be a contributing factor to the reduced coverage of the σ^2 parameter, in which interval widths may have been too short to properly capture the truth, this factor does not seem to hinder our analyses as $\boldsymbol{h}$ is estimated well. If one was interested in adjusting for the sometimes underestimated posterior variance estimates, a grid-based method proposed by Ormerod (2011) could be used. However, for large p and small n, as is in the case described in this article, the approach of calculation over the grid values can easily render the GBVA method computationally infeasible.

A potential extension of this article is to develop variable selection within the MFVB for

LKMR inference procedure. As the method can efficiently analyze studies with large sample sizes, there is greater power to investigate a larger panel of toxicants and detect specific ones that exert an undue effect on health. We suggest the addition of a variable selection term, denoted by $\boldsymbol{r} = (r_1, ..., r_M)^T$, which corresponds to M components of the mixture. This vector could incorporated into the overall kernel matrix $\mathbf{K}(\mathbf{z}, \mathbf{z}'; \boldsymbol{r})$, such that at each (i, j)th element we have $K(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{r}) = (1 + (\boldsymbol{r}\boldsymbol{z}_i)(\boldsymbol{r}\boldsymbol{z}_j))^2$. However, estimation of $\boldsymbol{r}$ is not straightforward. Because $\boldsymbol{r}$ is embedded into the kernel matrix and the individual r_m cannot be separated, we are unable to find a closed form density for $q^*(\boldsymbol{r})$ in the MFVB procedure. Therefore, quadrature may be required, which is a classical numerical technique for evaluation definite integrals that do not have analytical solutions. Wand et al. (2011) proposed univariate quadrature for evaluating non-analytic integrals which correspond to non-Gibbsian updates, which could be extended to multivariate quadrature in this case. Another possibility could be to use adaptive rejection sampling (ARS). To our knowledge, the only example of ARS in the variational Bayes framework is mentioned briefly in Winn and Bishop (2005).

In conclusion, we note that the complexity of the LKMR Bayesian model, coupled with the computational burden of estimation using MCMC, warrants and necessitates the variational Bayes approach. Furthermore, the high accuracy and similar coverage of key parameters between MCMC and MFVB illustrates the usefulness of the variational Bayes approach.

Bayesian Varying Coefficient Kernel Machine Regression for Longitudinal Data to Assess Health Effects of Exposures to Complex Metal Mixtures

Shelley Han Liu

Department of Biostatistics

Harvard Graduate School of Arts and Sciences

Jennifer F. Bobb

Biostatistics Unit

Group Health Research Institute

Lourdes Schnaas

Division for Research in Community Interventions

National Institute of Perinatology, Mexico

Martha Tellez-Rojo

Center for Research in Nutrition and Health

National Institute of Public Health, Mexico

David Bellinger

Department of Environmental Health
Harvard T.H. Chan School of Public Health

Manish Arora
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Robert Wright
Department of Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai

Brent Coull
Department of Biostatistics
Harvard T.H. Chan School of Public Health

3.1 Introduction

Child neurodevelopment and cognitive function are critically important to public health. The National Institute of Environmental Health Sciences (NIEHS) has placed a priority on the quantification of the health impacts of exposure to environmental mixtures (Carlin et al., 2013). Numerous articles have reported on the neurodevelopmental effects of perinatal exposure to toxicants, such as pesticides, heavy metals, insecticides, lead and methyl mercury (Gonzalez-Alzaga et al., 2015; Kalkbrenner et al., 2014; Rauh et al., 2006; Canfield et al., 2003; Marques et al., 2014). In this article, we investigate how exposures to heavy metal mixtures affects cognitive growth trajectories. During fetal life and early childhood, neurodevelopmental processes are rapid, sequential and well-timed, such that exposure to toxicant exposures can cause lifelong effects (Stiles and Jernigan, 2010). Many metals cross the placental barrier, potentially causing injury to the fetal brain. Environmental toxicants may impose signals on the central nervous system's development; even weak signals could alter a normal cognitive growth trajectory to a maladaptive one.

Statistical methods for analyzing the longitudinal health impact of exposure to toxicant mixtures must address several key issues. First, it is important to account for potential interactive and synergistic relationships among mixture components. For example, there is a possibility for detectable mixture effects on health at low doses of exposure below individual no observable adverse effect levels (Kortenkamp et al., 2007). There is also evidence of interactions between individual metals: (Claus Henn et al., 2014) found increased lead toxicity in the presence of higher levels of manganese, arsenic, and cadmium. Secondly, the exposure-response relationship between metal mixture exposures and neurodevelopment may be complex, exhibiting both nonlinearity and non-additivity, which must be accounted for in the statistical model. For example, manganese has dual roles: it functions as an essential nutrient at low doses but as a neurotoxicant at high doses, resulting in an inverted-u relationship with neurodevelopment (Claus Henn et al., 2010). Lastly, the ideal method would flexibly capture complex

mixture-response relationships while also incorporating prior knowledge from previous studies or subject matter experts.

We propose a new Bayesian hierarchical model, termed Bayesian Varying Coefficient Kernel Machine Regression, to estimate cognitive growth trajectories of exposures to environmental mixtures, while flexibly capturing the exposure-response relationships of mixtures, incorporating prior knowledge, and accounting for non-linear and non-additive effects of individual exposures. Kernel machine regression (KMR) has been used substantially in the statistical genetics literature, where it is used primarily to test the significance of gene sets and predict risk for health outcomes (Cai et al., 2011; Maity and Lin, 2011; Liu et al., 2007). In the machine learning literature, KMR is often referred to as Gaussian Process regression, and is used for prediction and variable selection, through applications to chemometric calibrations and simulations of physical processes (Chen and Wang, 2010; Linkletter et al., 2006; Qian et al., 2008; Savitsky et al., 2011). Bobb et al. (2015) has developed Bayesian kernel machine regression methods for estimating health effects of complex mixtures and conducting variable selection for cross-sectional investigations. Liu et al. (2016) developed lagged kernel machine regression for estimating health effects of time-varying exposures to environmental mixtures. We extend BKMR methods from Bobb et al. (2015) to longitudinal outcomes case, using the modeling framework of Liu et al. (2016). Our method provides a novel conceptual framework for the estimation of cognitive trajectories associated with exposure to metal mixtures.

The paper is organized as follows: Section 3.2 provides a review of KMR; Section 3.3 introduces the statistical model BVCKMR; Section 3.4 describes the simulation studies for evaluating performance of BVCKMR; Section 3.5 addresses an application of the method to the PROGRESS dataset; and Section 3.6 provides discussion and concluding remarks.

3.2 Review of Bayesian kernel machine regression

We first review kernel machine regression as a framework for estimating the impact of a complex mixture, when both outcome and exposures are measured only at a single time point. Suppose we observe data from n subjects. We describe the model for a continuous, normally distributed outcome. For each subject $i = 1, \ldots, n$, kernel machine regression (KMR) relates the outcome (Y_i) to M components of the exposure mixture $\mathbf{z}_i = (z_{1i}, ..., z_{Mi})$ through a nonparametric function, $h(\cdot)$, while controlling for p relevant confounders $\mathbf{x}_i = (x_{1i}, ..., x_{pi})$. The model is

$$Y_i = h(z_{1i}, ..., z_{Mi}) + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{3.1}$$

where $\boldsymbol{\beta}$ represents the effects of the potential confounders, and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. $h(\cdot)$ can be estimated parametrically or non-parametrically. We employ a kernel representation for $h(\cdot)$ in order to accommodate the possibly complex exposure-response relationship.

The unknown function $h(\cdot)$ can be specified through basis functions or through a positive definite kernel function $K(\cdot, \cdot)$. Under regularity conditions, Mercer's theorem (Cristianini and Shawe-Taylor, 2000) showed that the kernel function $K(\cdot, \cdot)$ implicitly specifies a unique function space, H_k, that is spanned by a set of orthogonal basis functions. Thus, any function $h(\cdot) \in H_k$ can be represented through either a set of basis functions under the primal representation, or through a kernel function under the dual representation. The kernel function uses a similarity metric $K(\cdot, \cdot)$ to quantify the distance between the exposure profiles $\mathbf{z}_i$ between any two subjects in the study. For example, the Gaussian kernel quantifies similarity through the Euclidean distance; the polynomial kernel, through the inner product. Through specifying different kernels, one is able to control the complexity of the exposure-response function.

Liu et al. (2007) developed least-squares kernel machine semi-parametric regression for studying genetic pathway effects. The authors connect kernel machine methods with

linear mixed models, demonstrating that (1) can be expressed as the mixed model

$$y_i \sim N(h_i + \mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2) \tag{3.2}$$

$$\mathbf{h} = (h_1, ..., h_n)^T \sim N\left[\mathbf{0}, \tau \mathbf{K}(\cdot, \cdot)\right], \tag{3.3}$$

where **K** is a kernel matrix with *i,j* element $K(\mathbf{z}_i, \mathbf{z}_j)$.

3.3 Bayesian Varying Coefficient Kernel Machine Regression model formulation

Now suppose that exposures to a complex mixture are measured at a single time point, but we now have repeated measures on the outcome, with the goal of identifying the longitudinal health impact of exposure. For each subject $i = 1, \ldots, n$, with outcomes measured at ages $j = 1, \ldots, J$, we consider a random effects model. We relate the outcome (Y_{ij}) to M components of the exposure mixture $\mathbf{z}_i = (z_{1i}, ..., z_{Mi})$ through two non-parametric functions, $h_1(\cdot)$ and $h_2(\cdot)$. We again control for p relevant confounders $\mathbf{x}_i = (x_{1i}, ..., x_{pi})$. The Bayesian Varying Coefficient Kernel Machine Regression (BVCKMR) model is defined as:

$$Y_{ij} = h_1(z_{1i}, ..., z_{Mi}) + h_2(z_{1i}, ..., z_{Mi}) * \text{age}_{ij} + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{u}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}, \tag{3.4}$$

where $h_q(\cdot)$, $q = 1, 2$ can be estimated parametrically or non-parametrically, and represents the exposure-response functions for the exposures $\mathbf{z}_i$. For individual i, $h_{1,i}$ corresponds to the individual-specific baseline outcome level, while $h_{2,i}$ corresponds to the individual-specific age-related trajectory of the health outcome. We again employ a kernel representation for $h_q(\cdot)$ in order to accommodate the possibly complex exposure-response relationships. Also, $\boldsymbol{b}_i$ represents the random effects for subject i, where $b_{1,i}$ corresponds to the intercept and $b_{2,i}$ corresponds to the slope.

For simplicity, we write the random effects model as

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{W}_i \boldsymbol{h} + \mathbf{U}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i \tag{3.5}$$

for $i = 1, ..., n$ and repeated measures $j = 1, ..., J$, where $\boldsymbol{b}_i \sim N_2(0, \sigma^2\mathbf{D})$ and $\boldsymbol{\epsilon}_i \sim N_J(0, \sigma^2 I_J)$.

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{Y}_n \end{pmatrix} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ \vdots \\ Y_{n1} \\ \vdots \\ Y_{nJ} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \mathbf{U} = \begin{pmatrix} 1 & age_{11} & & & & & \\ \vdots & \vdots & & & & & \\ 1 & age_{1J} & & & & & \\ & & 1 & age_{21} & & & \\ & & \vdots & \vdots & & & \\ & & 1 & age_{2J} & & & \\ & & & & \ddots & & \\ & & & & & 1 & age_{n1} \\ & & & & & \vdots & \vdots \\ & & & & & 1 & age_{nJ} \end{pmatrix}$$

$$\boldsymbol{b} = \begin{pmatrix} b_{1,1} \\ b_{2,1} \\ b_{1,2} \\ b_{2,2} \\ \vdots \\ \vdots \\ b_{1,n} \\ b_{2,n} \end{pmatrix}, \mathbf{W} = \begin{pmatrix} 1 & 0 & \dots & 0 & age_{11} & 0 & \dots & \\ \vdots & & & & \vdots & & & \\ 1 & 0 & \dots & 0 & age_{1J} & 0 & \dots & \\ 0 & 1 & 0 & \dots & 0 & age_{21} & 0 & \dots \\ 0 & \vdots & & & & \vdots & & \\ 0 & 1 & 0 & \dots & 0 & age_{2J} & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots & 0 & age_{31} & 0 \\ 0 & 0 & \vdots & & & & \vdots & \ddots \end{pmatrix}, \boldsymbol{h} = \begin{pmatrix} h_{1,1} \\ h_{1,2} \\ \vdots \\ h_{1,n} \\ h_{2,1} \\ h_{2,2} \\ \vdots \\ h_{2,n} \end{pmatrix}$$

The conditional prior of $\boldsymbol{h}|\lambda_1^2$ is

$$\pi(\boldsymbol{h}|\lambda_1^2) \propto exp\left[-\lambda_1^2 \sum_{q=1}^{2} \|\boldsymbol{h}_q\|_{\mathbf{G}}\right], \tag{3.6}$$

where $\|\boldsymbol{h}_q\|_{\mathbf{G}} = (\boldsymbol{h}_{\boldsymbol{q}}^T \mathbf{G} \mathbf{h}_{\mathbf{q}})^{\mathbf{1/2}}$. We define $\mathbf{G} = \mathbf{K}^{-1}$, where $\mathbf{K}$ denotes the kernel matrix with *i,j* element $K(\mathbf{z}_i, \mathbf{z}_j)$. Depending on a particular application, a number of different kernel functions may be employed. Because previous literature suggests an inverted-*u* exposure-response relationship between metal mixture exposures and neurodevelopment, we choose $\mathbf{K}$ to be a quadratic kernel, such that $K(\mathbf{z, z'}) = (\mathbf{zz'} + 1)^2$.

The hierarchical model is represented as:

$$\mathbf{y}|\boldsymbol{h}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{b} \sim N_N(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{h} + \mathbf{U}\mathbf{b}, \sigma^{\mathbf{2}}\mathbf{I}_{\mathbf{N}}) \tag{3.7}$$

$$\boldsymbol{h}|\tau_1^2,\tau_2^2 \sim N_{2n}\left\{\mathbf{0}, \Sigma_h = \begin{pmatrix} \tau_1^2\mathbf{K} & 0 \\ 0 & \tau_2^2\mathbf{K} \end{pmatrix}\right\} \tag{3.8}$$

$$\tau_1^2,\tau_2^2 \overset{iid}{\sim} \text{gamma}(\frac{n+1}{2}, \frac{{\lambda_1}^2}{2}) \tag{3.9}$$

$$\lambda_1^2 \sim \text{gamma}(r,\delta) \tag{3.10}$$

$$\boldsymbol{b}|\sigma^2, D \sim N_{2n}(\mathbf{0}, \sigma^2(I_n \otimes D)) \tag{3.11}$$

$$D^{-1} \sim \text{Wishart}_q(v_0, C_0) \tag{3.12}$$

$$\sigma^2 \sim \text{inverse gamma}(a,\gamma) \tag{3.13}$$

$$\beta \sim \mathbf{1} \tag{3.14}$$

The form of Σ_h follows from representing the Laplace (double exponential) conditional prior of $\boldsymbol{h}|\lambda_1^2$ as a scale mixture of a normal distribution with an exponential mixing density (Andrews and Mallows, 1974). The diagonal blocks of size n x n arise due to the kernel structure placed on each $h_q(\cdot)$, $q = 1, 2$.

The joint density is:

$$f(\boldsymbol{\beta},\boldsymbol{h},\sigma^2,\boldsymbol{\tau}^2,\lambda_1^2,\boldsymbol{b},\boldsymbol{D}|\mathbf{y}) \propto \frac{1}{(\sigma^2)^{N/2}}\text{exp}\left[\frac{-1}{2\sigma^2}(\mathbf{y}-\mathbf{W}\boldsymbol{h}-\mathbf{X}\boldsymbol{\beta}-\mathbf{U}\mathbf{b})^{\mathbf{T}}(\mathbf{y}-\mathbf{W}\mathbf{h}-\mathbf{X}\boldsymbol{\beta}-\mathbf{U}\mathbf{b}\right] \times$$

$$\prod_{q=1}^{2}\left(\frac{1}{\tau_q^2}\right)^{n/2}\text{exp}\left(-\frac{\|\boldsymbol{h_q}\|_{\mathbf{G}}^2}{2\tau_q^2}\right)\frac{\left(\frac{\lambda_1^2}{2}\right)^{\frac{n+1}{2}}\left(\tau_q^2\right)^{\frac{n+1}{2}-1}}{\Gamma\left(\frac{n+1}{2}\right)}\text{exp}\left(-\lambda_1^2\tau_q^2/2\right) \times$$

$$\frac{\delta^r}{\Gamma(r)}(\lambda_1^2)^{r-1}\text{exp}\left(-\delta\lambda_1^2\right) \times \frac{1}{(\sigma^2)^n}|(I_n \otimes \boldsymbol{D})|^{-1/2}\text{exp}\left(-\frac{1}{2\sigma^2}\boldsymbol{b}^T(I_n \otimes \boldsymbol{D}^{-1})\boldsymbol{b}\right) \times \tag{3.15}$$

$$(\sigma_2)^{-a-1}\text{exp}\left(-\frac{\gamma}{\sigma^2}\right) \times |\boldsymbol{D}^{-1}|^{(v_0-3)/2}\text{exp}\left(-\frac{1}{2}tr(C_0^{-1}\boldsymbol{D}^{-1})\right) \tag{3.16}$$

We detail the full conditional distributions of the parameters:

The full conditional distribution of $\boldsymbol{h}$ is:

$$\boldsymbol{h}|\sigma^2,\boldsymbol{\tau},\boldsymbol{b},\boldsymbol{\beta},\mathbf{y} \sim \tag{3.17}$$

$$N_{2n}\left\{(\frac{1}{\sigma^2}\mathbf{W}^{\mathbf{T}}\mathbf{W}+\boldsymbol{\Sigma}_{\mathbf{h}}^{-\mathbf{1}})^{-\mathbf{1}}\frac{\mathbf{1}}{\boldsymbol{\sigma}^{\mathbf{2}}}\mathbf{W}^{\mathbf{T}}\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{W}\mathbf{b}),(\frac{\mathbf{1}}{\boldsymbol{\sigma}^{\mathbf{2}}}\mathbf{W}^{\mathbf{T}}\mathbf{W}+\boldsymbol{\Sigma}_{\mathbf{h}}^{-\mathbf{1}})^{-\mathbf{1}}\right\}$$

The full conditional of σ^2 is:

$$\sigma^2|\boldsymbol{h}, \boldsymbol{\tau}, \boldsymbol{\omega}, \boldsymbol{X}, \boldsymbol{\beta}, \mathbf{y} \sim \textit{Inverse Gamma}\left\{ N/2 + n + a, \right. \tag{3.18}$$

$$\left. \frac{(\mathbf{y} - \mathbf{W}\boldsymbol{h} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\mathbf{b})^{\mathbf{T}}(\mathbf{y} - \mathbf{W}\mathbf{h} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\mathbf{b}) + \mathbf{b}^{\mathbf{T}}(\mathbf{I_n} \otimes \mathbf{D})^{-\mathbf{1}}\mathbf{b} + \mathbf{2}\gamma}{2} \right\}$$

The full conditional of $\boldsymbol{\beta}$ is:

$$\boldsymbol{\beta}|\mathbf{y}, \sigma^2, \boldsymbol{h}, \boldsymbol{b} \sim N_p\left\{ (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{W}\boldsymbol{h} - \mathbf{U}\mathbf{b}), \sigma^{\mathbf{2}}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-\mathbf{1}} \right\} \tag{3.19}$$

The full conditional of $\boldsymbol{b}$ is:

$$\boldsymbol{b}|\mathbf{y}, \sigma^2, \boldsymbol{h}, \boldsymbol{\beta}, \boldsymbol{D} \sim N_{2n}\left\{ \mathbf{b}^*, \sigma^{\mathbf{2}}((\mathbf{I_n} \otimes \mathbf{D})^{-\mathbf{1}} + \mathbf{U}^{\mathbf{T}}\mathbf{U})^{-\mathbf{1}} \right\} \tag{3.20}$$

where

$$\mathbf{b}^* = ((\mathbf{I_n} \otimes \mathbf{D})^{-\mathbf{1}} + \mathbf{U}^{\mathbf{T}}\mathbf{U})^{-\mathbf{1}}(\mathbf{U}^{\mathbf{T}}\mathbf{U})\hat{\mathbf{b}}$$

and

$$\hat{\mathbf{b}} = (\mathbf{U}^{\mathbf{T}}\mathbf{U})^{-\mathbf{1}}\mathbf{U}^{\mathbf{T}}(\mathbf{y} - \mathbf{W}\mathbf{h} - \mathbf{X}\boldsymbol{\beta})$$

The full conditional of $\boldsymbol{\tau}^2$ is:

$$\frac{1}{\tau_q^2}|\boldsymbol{h}, \lambda_1^2 \sim \textit{Inverse Gaussian}\left\{ \left\{ \frac{\lambda_1^2}{||\boldsymbol{h}_q||_G^2} \right\}^{1/2}, \lambda_1^2 \right\} \tag{3.21}$$

The full conditional of λ_1^2 is:

$$\lambda_1^2|\boldsymbol{\tau}^2 \sim \textit{Gamma}\left\{ n + 1 + r, \sum_{q=1}^{2} \frac{\tau_q^2}{2} + \delta \right\} \tag{3.22}$$

Lastly, the full conditional of $\boldsymbol{D}^{-1}$ is:

$$\boldsymbol{D}^{-1}|\boldsymbol{b} \sim \textit{Wishart}_2\left\{ n + v_0, (C_0^{-1} + \frac{1}{\sigma^2}\mathbf{B}^T\mathbf{B})^{-1} \right\} \tag{3.23}$$

where

$$\mathbf{B} = (\mathbf{b}_1, ..., \mathbf{b}_n)$$

and

$$\mathbf{B}^T\mathbf{B} = \sum_{i=1}^{n} \mathbf{b}_i\mathbf{b}_i^T$$

3.1 MCMC sampler

To implement the Gibbs sampler, we use the following steps:

(0) Start with initial values $(\boldsymbol{\beta}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\tau}^{2(0)}, \boldsymbol{h}^{(0)}, \boldsymbol{D}^{-1(0)}, \sigma^{2(0)}, \lambda_1^{2(0)})$

(1) Generate $\boldsymbol{\tau}^{2(1)}$ from $[\boldsymbol{\tau}^2|\lambda_1^{2(0)}]$

(2) Generate $\boldsymbol{h}^{(1)}$ from $[\boldsymbol{h}|\boldsymbol{\tau}^{2(1)}, \lambda_1^{2(0)}]$

(3) Generate $\boldsymbol{b}^{(1)}$ from $[\boldsymbol{b}|\mathbf{y}, \boldsymbol{h}^{(1)}, \sigma^{2(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{D}^{-1(0)}]$

(4) Generate $\sigma^{2(1)}$ from $[\sigma^2|\mathbf{y}, \boldsymbol{h}^{(1)}, \boldsymbol{b}^{(1)}, \boldsymbol{\beta}^{(0)}]$

(5) Generate $\boldsymbol{\beta}^{(1)}$ from $[\boldsymbol{\beta}|\mathbf{y}, \sigma^{2(1)}, \boldsymbol{h}^{(1)}, \boldsymbol{b}^{(1)}]$

(6) Generate $\lambda_1^{2(1)}$ from $[\lambda_1^2|\boldsymbol{\tau}^{2(1)}]$

(7) Generate $\boldsymbol{D}^{-1(1)}$ from $[\boldsymbol{D}^{-1}|\boldsymbol{b}^{(1)}, \sigma^{2(1)}]$

Repeat steps (1) - (7) until we obtain R samples $(\boldsymbol{\beta}^{(r)}, \boldsymbol{b}^{(r)}, \boldsymbol{\tau}^{2(r)}, \boldsymbol{h}^{(r)}, \boldsymbol{D}^{-1(r)}, \sigma^{2(r)}, \lambda_1^{2(r)}), r = 1, ..., R.$ These samples will be samples from the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\tau}^2, \boldsymbol{h}, \boldsymbol{D}^{-1}, \sigma^2, \lambda_1^2|\mathbf{y})$.

3.2 Prediction of h_1 and h_2 at new exposure profiles

An important aim of environmental health studies is the characterization of the exposure-response surface. It is often of interest to predict health effects unobserved exposure profiles. Suppose we are interested in predicting the exposure-response relationship for new profiles of metal mixture exposures, $\mathbf{z}_{new} = (\mathbf{z}_{new,1}, ..., \mathbf{z}_{new,M})$, for n_{new} subjects, where $\boldsymbol{h}_{new} = (h_{1,n+1}, ..., h_{1,n+n_{new}}, h_{2,n+1}, ..., h_{2,n+n_{new}})^T$ represents the desired predictions. In order to estimate $\boldsymbol{h}_{new}$, we first re-arrange the $\boldsymbol{h}$ vector so that $\tilde{\boldsymbol{h}} = (h_{1,1}, ..., h_{1,n}, h_{2,1}, ..., h_{2,n}, h_{2,n+1}, ..., h_{2,n+n_{new}}, h_{1,n+1}, ..., h_{1,n+n_{new}})^T$. Because we have reordered the $\tilde{\boldsymbol{h}}$ vector, we need to similarly reorder the covariance matrix to correspond, and denote the reordered matrix by $\tilde{\Sigma}_h^{-1}$.

The joint distribution of observed and new exposure profiles is:

$$\begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{h}_{new} \end{pmatrix} \sim N\left\{ \mathbf{0}, \tilde{\Sigma}_h = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{12}^T & \tilde{\Sigma}_{22} \end{pmatrix} \right\} \tag{3.24}$$

where $\tilde{\Sigma}_{11}$ denotes the $2n$ x $2n$ matrix with $(i,j)^{th}$ element $K(\mathbf{z}_i, \mathbf{z}_j)$, $\tilde{\Sigma}_{12}$ denotes the n x n_{new} matrix with $(i, j_{new})^{th}$ element $K(\mathbf{z}_i, \mathbf{z}_{j_{new}})$, and $\tilde{\Sigma}_{22}$ denotes the n_{new} x n_{new} matrix

with $(i_{new}, j_{new})^{th}$ element $K(\mathbf{z}_{i_{new}}, \mathbf{z}_{j_{new}})$. It follows that the conditional posterior distribution of $\boldsymbol{h}_{new}$ is:

$$\boldsymbol{h^{new}}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\tau^2}, \sigma^2 \sim N_{n_{new}}\Bigg\{\tilde{\Sigma}_{12}^T\tilde{\Sigma}_{11}^{-1}\Big\{\frac{1}{\sigma^2}\mathbf{W}^T\mathbf{W} + \tilde{\Sigma}_{11}^{-1}\Big\}^{-1}\frac{1}{\sigma^2}\mathbf{W}^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\boldsymbol{b}), \tag{3.25}$$

$$\tilde{\Sigma}_{12}^T\tilde{\Sigma}_{11}^{-1}\Big\{\frac{1}{\sigma^2}\mathbf{W}^T\mathbf{W} + \tilde{\Sigma}_{11}^{-1}\Big\}^{-1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + \tilde{\Sigma}_{22} - \tilde{\Sigma}_{12}^T\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\Bigg\}$$

Due to the computational expense of a large amount of predictions - since it requires simulation from a high-dimensional multivariate normal distribution - the posterior mean and variance of $\boldsymbol{h}_{new}$ were found through their conditional posterior evaluated at the posterior mean of all other parameters.

3.4 Simulation studies

We conducted simulation studies to evaluate the performance of the proposed BVCKMR model for estimating $\boldsymbol{h}_q$, $q = 1, 2$. Our simulation study considered three scenarios in which different combinations of five toxicants exert an exposure-response effect. These scenarios were: (1) Linear, (2) Linear with interaction terms, and (3) Quadratic. We used the following model: $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \sum_t h_t(\mathbf{z}_{it}) + e_i$, where $e_i \sim N(0, 1)$, $\mathbf{x}_i = (x_{1i}, x_{2i})$ and $x_{1i} \sim N(10, 1)$ and $x_{2i} \sim Bernoulli(0.5)$. We simulated correlation between toxicants based on existing data on metal mixture exposures. The exposure-response functions $\boldsymbol{h_q}(z_m)$ were simulated as (1) Linear: $\boldsymbol{h}_1(\mathbf{z}) = 0.5 * z_1$ and $\boldsymbol{h}_2(\mathbf{z}) = 0.5 * z_1$; (2) Linear with interaction terms: $\boldsymbol{h}_1(\mathbf{z}) = 0.5 * (0.5z_1z_2 + z_1 + z_2)$ and $\boldsymbol{h}_2(\mathbf{z}) = 0.5 * (z_1 - z_2)$; (3) Quadratic: $\boldsymbol{h}_1(\mathbf{z}) = 0.5 * (z_1^2 - z_2^2 + 0.5z_1z_2 + z_1 + z_2)$ and $\boldsymbol{h}_2(\mathbf{z}) = 0.5 * (z_1^2 - z_2^2)$. In conducting the analysis, both exposure covariates and confounder variables are centered and scaled.

Table 3.1 presents the results of this simulation. For each simulated data set, to assess the performance of the model for the purposes of estimating the time-specific exposure-response function, we regressed the predicted $\widehat{\boldsymbol{h}}$ on $\boldsymbol{h}$ for each time point. We present the intercept, slope and R^2 of the regressions over 100 simulations. Good

estimation performance occurs when the intercept is close to zero, and the slope and R^2 are both close to one. We also present the root mean squared error (RMSE) and the coverage (the proportion of times the true $h_{i,t}$ is contained in the posterior credible interval).

Table 3.1 Simulation results, regression of $\widehat{\boldsymbol{h}}$ on h

Scenario	h function	Intercept	Slope	R^2	RMSE
Linear	1	0.001	1.011	0.740	0.309
	2	-0.005	1.017	0.900	0.174
Linear, interaction	1	0.001	1.006	0.880	0.309
	2	-0.005	1.011	0.935	0.175
Quadratic	1	0.001	1.003	0.946	0.309
	2	-0.005	0.998	0.969	0.175

Performance of estimated $h_q(\mathbf{z}_i)$ across 100 simulated datasets. RMSE denotes the root mean squared error of the $\hat{h}$ as compared to h. Coverage denotes the proportion of times that the true h falls within in the posterior credible interval of each time point.

The simulation results are further illustrated in Figure 3.1, in which we plot the posterior mean of the exposure-response surface for a grid of z_1 and z_2 values, at fixed values of z_3, z_4, z_5. For each simulation study, the upper panel contains predicted estimates of $h(z_1, z_2)$ using BVCKMR, while the lower panel contains the known true exposure-response relationship for the grid of points. The predicted exposures response surface was averaged across 100 datasets.

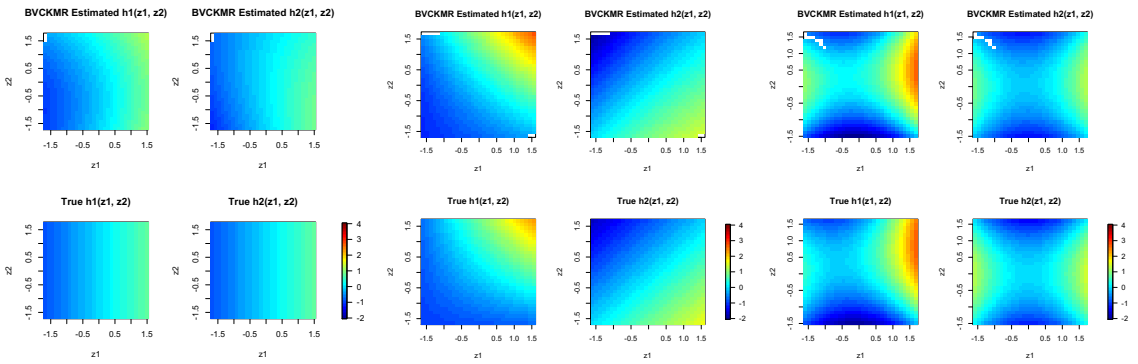

Figure 3.1: Comparison of BVCKMR estimated vs. true exposure-response surfaces for a simulation study. Three simulation scenarios were considered: Linear, quadratic and quadratic with interaction. Plot of the posterior mean of the exposure-response surface for a grid of z_1 and z_2, at fixed values of z_3, z_4, z_5. For each simulation study, the upper panel contains predicted estimates of $h(z_1, z_2)$ using BVCKMR, while the lower panel contains the known true exposure-response relationship for the grid of points. The predicted exposures response surface was averaged across 100 datasets.

3.5 Application

We applied the BVCKMR model to the PROGRESS dataset, in order to investigate early life cognitive trajectories associated with metal mixture exposures in mother's blood at birth. Exposures to a panel of 9 metals were considered (As, Cd, Co, Cr, Cs, Cu, Mn, Pb, Sb). We controlled for socioeconomic status (3 categories: low, middle, high), mother's hemoglobin during the second trimester of pregnancy, mother's educational level ($<$ high school, high school, $>$ high school), child gender, mother's WASI IQ, and Fenton's birthweight z-scores. In our analysis, metal exposure levels were logged, then centered and scaled. Confounder variables were also centered and scaled. The longitudinal outcome variable was Bayley composite cognition scores at four time points (6, 12, 18 and 24 months of age).

For preliminary analysis and a baseline for comparison to the BVCKMR model, we considered a linear mixed effects model with random intercept and random slope. Through this model, while no individual effects of metals were identified as significant,

we did identity several significant interactions. These included: a positive interaction between Co-Cd at baseline (p=0.04) and a positive Mn-Co interaction at baseline (p=0.03).

Using the BVCKMR model, we initially estimated the relative importance of each metal for both the baseline cognitive level and cognitive trajectories, as shown in Figure 3.2. Relative importance is quantified by the difference in the estimated main effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures. The results suggest of the positive effect of Co and the negative effects of Pb and Cd at baseline (h_1). They further suggest a positive effect of Co and Cu associated with cognitive trajectories (h_2), while Pb is negative associated with cognitive trajectories.

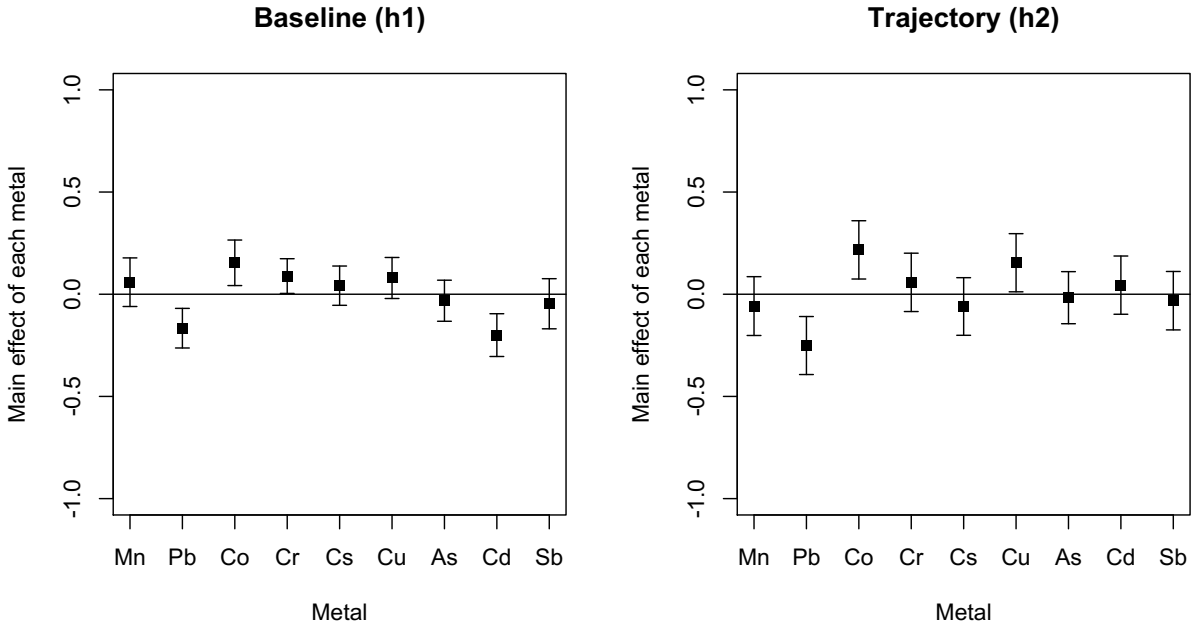

Figure 3.2: Estimated baseline and cognitive trajectory main effects applied to PROGRESS data. Plot of the relative importance of Co, Cs, As, Cd, Cr, Cu, Mn, Pb, and Sb. Relative importance is quantified by the difference in the estimated main effect of a single metal at high exposure (75th percentile) and low exposure (25th percentile), holding all other metals constant at median exposures.

Next, due to scientific interest regarding Mn-Pb co-exposures,we focus on those two metals when exploring the exposure-response relationship. Because the exposure response

surface is nine-dimensional, we use heat maps and cross-sectional plots to reduce dimensionality and graphically depict the exposure-response relationship. Figure 3.3 presents the plot of the predicted exposure-response surface for Mn and Pb, estimated using BVCKMR at the median of all other metal exposures. The shape of the surface for h2, cognitive trajectory, suggests an interaction effect. To explore this, in Figure 3.4 we plot the predicted cross-section of the exposure-response surface for Mn, at low and high Pb, holding all other metal exposures at their median. Comparing the top panel (low Pb) to the bottom panel (high Pb), we detect a possible suggestion of a Mn-Pb interaction, specifically effect modification in the presence of higher Pb levels for the cognitive trajectory. In the presence of higher Pb levels, there appears to be a stronger negative linear relationship between Mn exposures and cognitive trajectories. Lastly, as the cross-sectional graphs suggest the effects are mainly linear, a quadratic kernel appears to sufficiently capture the exposure-response relationship.

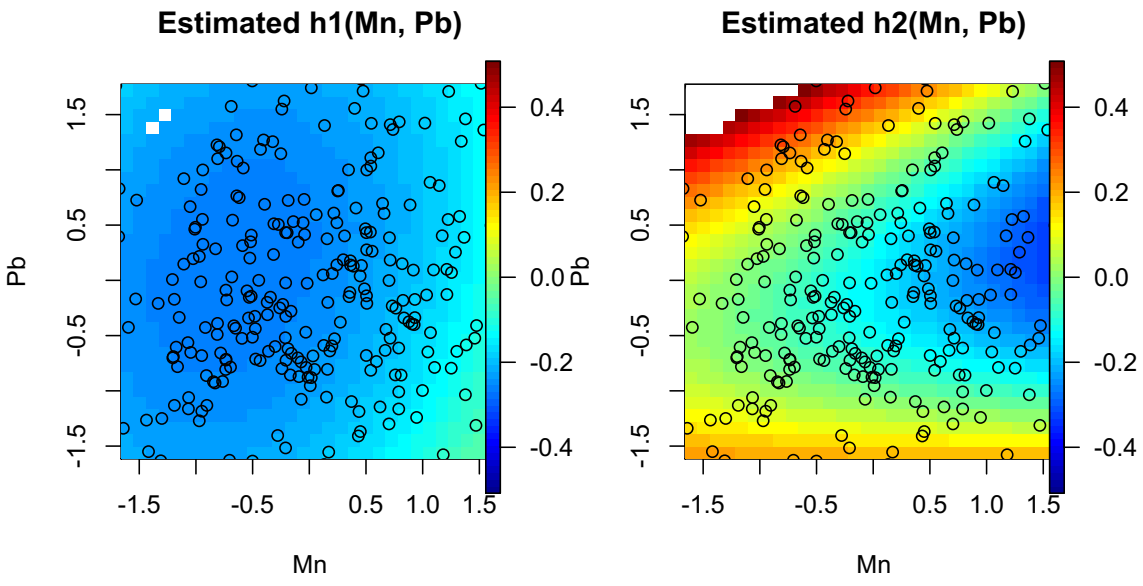

Figure 3.3: Estimated Mn-Pb exposure response functions for baseline (h1) and cognitive trajectories (h2) applied to PROGRESS data. Plot of the estimated posterior mean of the exposure-response surface for Mn and Pb, at the median of As, Cd, Cr, Cu, Co, Cs and Sb.

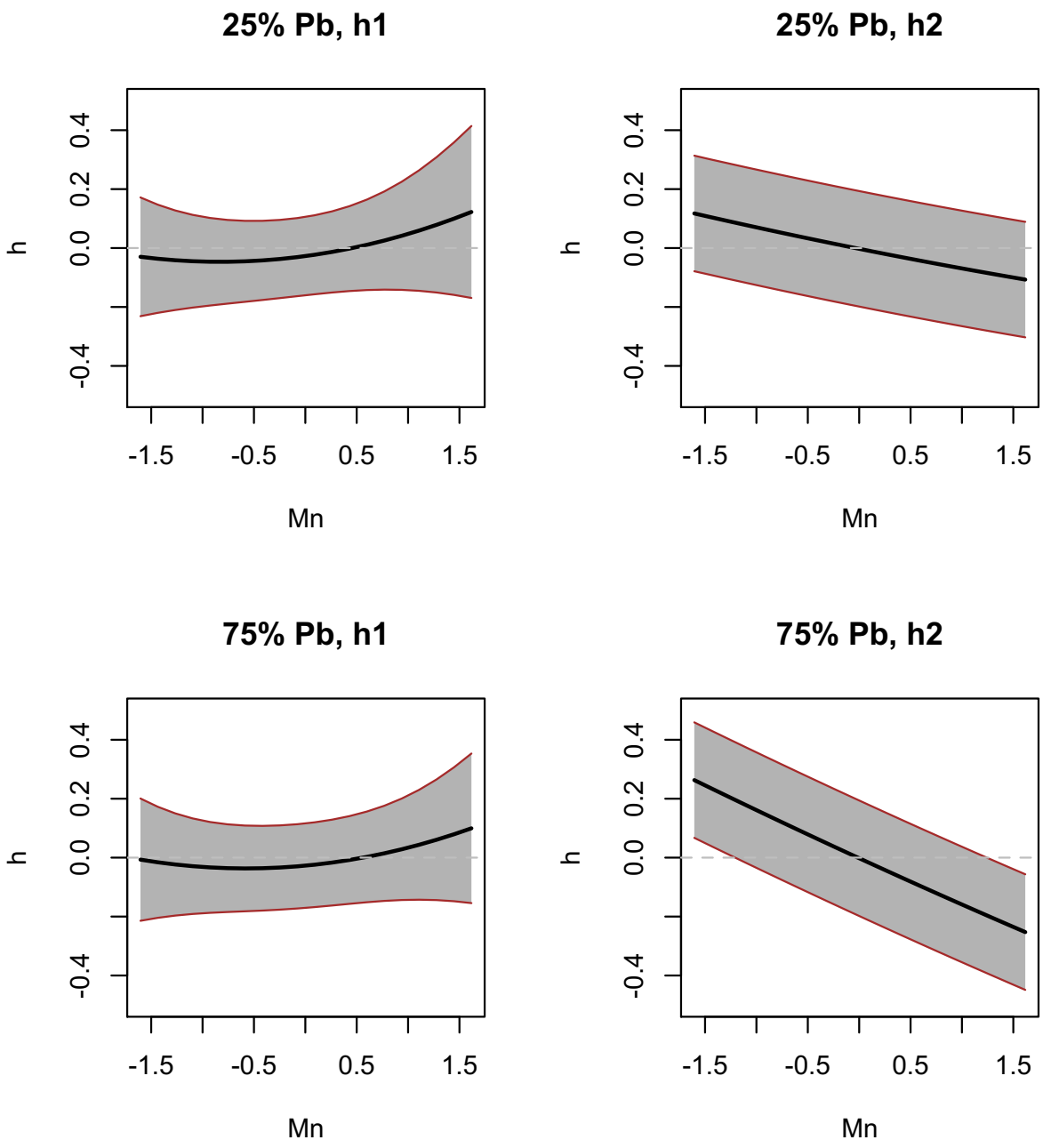

Figure 3.4: Estimated baseline (h1) and cognitive trajectory (h2) exposure-response functions for Mn at low and high Pb levels applied to PROGRESS data. Plot of the cross-section of the estimated exposure-response surface for Mn, at low Pb exposure of 25th percentile (top panel) and high Pb exposure of 75th percentile (bottom panel), holding As, Cd, Cr, Cu, Co, Cs and Sb constant at median exposures.

3.6 Discussion

We have developed a flexible Bayesian hierarchical modeling framework to simultaneously analyze data on exposures to environmental mixtures, and study their effects on baseline cognitive and cognitive trajectories. We applied the BVCKMR inference procedure to analyze a prospective cohort study of children's environmental health in Mexico City, where we found evidence of interaction and effect modification between Mn-Pb joint exposures. Specifically, we found that the negative association between Mn exposure and cognitive trajectories was heightened in the presence of high Pb levels. The negative association between Mn and cognition after birth has been previously reported (Claus Henn et al., 2010), and as Pb is a neurotoxicant, these results illustrate that Pb co-exposures can be an effect modifier for Mn exposures in relation to cognition.

A key contribution of this work is developing a Bayesian kernel machine regression framework that accounts for longitudinal outcomes to determine exposure effects that contribute to baseline cognition as well as cognitive changes across time. The flexible Bayesian representation allows for the investigation of complex exposure-response relationships accounts for correlation of mixture components over time while exploring for the possibility of non-linear and non-additive effects of individual exposures. By studying the complex relationship between environmental exposures and cognitive outcomes, this work allows for more precise identification and estimation environmental mixture exposures. To our knowledge, this is the first Bayesian supervised learning approach for investigating the effects of multi-pollutant mixtures on cognitive trajectories.

A potential extension of this work is to develop characterizations of the joint effects of a higher-dimension of metals. We currently focus on the health effects of two metals due to the straightforward visualization of this relationship through two-dimensional heat maps and cross-sectional plots. However, it may be of scientific interest to jointly model three or more metals together, but this type of analysis is limited by the difficulty of visualization. One potential approach is to compute a surface of differences between two

joint interaction models, thus detailing the effects of one metal in connection with the others. For example, one could first predict the exposure response surface between two metals, held at the 75th percentile of exposure for the third metal and the median of all other metals, and then predict another exposure response surface for the two metals of interest, this time held at the 25th percentile of exposure for the third metal. The surface of differences would be the difference between these two exposure-response surfaces.

References

ANDREWS, D. F. and MALLOWS, C. L. (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society, Series B (Methodological)* **36** 99–102.

BELLINGER, D. C. (2008). Very low lead exposures and children's neurodevelopment. *Current Opinion in Pediatrics* **20(2)** 172–7.

BILLIONNET, C., SHERRILL, D. and ANNESI-MAESANO, I. (2012). Estimating the health effects of exposure to multi-pollutant mixture. *Annals of epidemiology* **22** 126–41.

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.

BOBB, J. F., VALERI, L., CLAUS HENN, B., CHRISTIANI, D. C., WRIGHT, D. 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.

CAI, T., TONINI, G. and LIN, X. (2011). Kernel machine approach to testing the significance of multiple genetic markers for risk prediction. *Biometrics* **67** 975–86.

CANFIELD, R. L., HENDERSON, C. R., CORY-SLECHTA, D. A., COX, C., JUSKO, T. A. and LANPHEAR, B. P. (2003). Intellectual impairment in children with blood lead concentrations below 10 μg per deciliter. *New England Journal of Medicine* **348** 1517 – 1526.

CARLIN, D. J., RIDER, C. V., WOYCHICK, R. and BIRNBAUM, L. S. (2013). Unraveling the Health Effects of Environmental Mixtures: An NIEHS Priority. *Environmental Health Perspectives* **121** A6–A8.

CHEN, T. and WANG, B. (2010). Bayesian variable selection for Gaussian process re-

gression: Application to chemometric calibration of spectrometers. *Neurocomputing* **73** 2718–2726.

CLAUS HENN, B., COULL, B. A. and WRIGHT, R. O. (2014). Chemical Mixtures and Children's Health. *Current Opinion in Pediatrics* **26(2)** 223–229.

CLAUS HENN, B., ETTINGER, A. S., SCHWARTZ, J., TELLEZ-ROJO, M. M., LAMADRID-FIGUEROA, H., HERNANDEZ-AVILA, M., SCHNAAS, L., AMARASIRIWARDENA, C., BELLINGER, D. C., HU, H. and WRIGHT, R. O. (2010). Early postnatal blood manganese levels and children's neurodevelopment. *Epidemiology* **21(4)** 433–439.

CRISTIANINI, N. and SHAWE-TAYLOR, J. (2000). *An introduction to support vector machines*. Cambridge University Press.

DARROW, L. A., KLEIN, M., STRICKLAND, M. J., MULHOLLAND, J. A. and TOLBERT, P. E. (2011). Ambient Air Pollution and Birth Weight in Full-Term Infants in Atlanta, 1994-2004. *Environmental Health Perspectives* **119** 731–737.

DE VOCHT, F., CHERRY, N. and WAKEFIELD, J. (2012). A Bayesian mixture modeling approach for assessing the effects of correlated exposures in case-control studies. *Journal of exposure science & environmental epidemiology* **22** 352–60.

DIEZ, D. M., DOMINICI, F., ZARUBIAK, D. and LEVY, J. I. (2012). Statistical approaches for identifying air pollutant mixtures associated with aircraft departures at Los Angeles International Airport. *Environmental science & technology* **46** 8229–35.

FAES, C., ORMEROD, J. T. and WAND, M. P. (2011). Variational Bayesian Inference for Parametric and Nonparametric Regression With Missing Data. *Journal of the American Statistical Association* **106** 959–971.

GENNINGS, C., CARRICO, C., FACTOR-LITVAK, P., KRIGBAUM, N., CIRILLO, P. M. and COHN, B. A. (2013). A Cohort study evaluation of maternal PCB exposure related to time to pregnancy in daughters. *Environmental Health* **12** 66.

GOLDSMITH, J., WAND, M. P. and CRAINICEANU, C. (2011). Functional regression via variational Bayes. *Electronic Journal of Statistics* **5** 572–602.

GONZALEZ-ALZAGA, B., HERNANDEZ, A. F., RODRIGUEZ-BARRANCO, M., GOMEZ, I., AUILAR-GARDUNO, C., LOPEZ-FLORES, I., PARRON, T. and LACASANA, M. (2015). Pre- and postnatal exposures to pesticides and neurodevelopment effects in children living in agricultural communities from south-eastern spain. *Environment International* **85** 229–237.

HALL, P., PHAM, T., WAND, M. P. and WANG, S. S. J. (2011). Asymptotic normality and valid inference for Gaussian variational approximation. *The Annals of Statistics* **39** 2502–2532.

HEATON, M. J. and PENG, R. D. (2013). Extending distributed lag models to higher degrees. *Biostatistics* **15(2)** 398–412.

HERRING, A. H. (2010). Nonparametric bayes shrinkage for assessing exposures to mixtures subject to limits of detection. *Epidemiology* **21** S71–6.

HSU, L. H., CHIU, L. H., COULL, B. A., KLOOG, I., SCHWARTZ, J., LEE, A., WRIGHT, A. and WRIGHT, R. J. (2015). Prenatal Particulate Air Pollution and Asthma Onset in Urban Children: Identifying Sensitive Windows and Sex Differences. *American Journal of Respiratory and Critical Care Medicine* **DOI: 10.1164/rccm.201504-0658OC**.

HUANG, J., BREHENY, P. and MA, S. (2012). A selective review of group selection in high-dimensional models. *Statistical Science)* **27(4)** 481–499.

KALKBRENNER, A. E., SCHMIDT, R. J. and PENLESKY, A. C. (2014). Environmental chemical exposures and autism spectrum disorders: A review of the epidemiological evidence. *Current Problems in Pediatric and Adolescent Health Care* **44** 277–318.

KIM, Y., HA, E. H., PARK, H., HA, M., KIM, Y., HONG, Y. C., KIM, E. J. and KIM, B. N. (2013). Prenatal lead and cadmium co-exposure and infant neurodevelopment at 6 months of age: The Mothers and Children's Environmental Health (MOCEH) study. *NeuroToxicology* **35** 15–22.

KORTENKAMP, A., FAUST, M., SCHOLZE, M. and BACKHAUS, T. (2007). Low-Level Exposure to Multiple Mixtures: Reason for Human Health Concerns? *Environmental Health Perspectives* **115 Suppl 1** 106–114.

KYUNG, M., GILL, J., GHOSH, M. and CASELLA, G. (2010). Penalized regression, standard errors and Bayesian lassos. *Bayesian Analysis)* **5(2)** 369–412.

LINKLETTER, C., BINGHAM, D., HENGARTNER, N., HIGDON, D. and YE, K. Q. (2006). Variable Selection for Gaussian Process Models in Computer Experiments. *Technometrics* **48** 478–490.

LIU, D., LIN, X. and GHOSH, D. (2007). Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics* **63** 1079–88.

LIU, S. H., BOBB, J., LEE, K. H., GENNINGS, C., CLAUS HENN, B., WRIGHT, R. O., SCHNAAS, L., TELLEZ-ROJO, M., ARORA, M. and COULL, B. A. (2016). Lagged kernel machine regression for identifying time windows of susceptibility to exposures of complex metal mixtures. *In preparation* .

MAITY, A. and LIN, X. (2011). Powerful tests for detecting a gene effect in the presence of possible gene-gene interactions using garrote kernel machines. *Biometrics* **67** 1271–84.

MARQUES, R. C., BERNARDI, J. V., DOREA, J. G., MOREIRA, M. and MALM, O. (2014). Perinatal multiple exposure to neurotoxic (lead, methylmercury, ethylmercury and aluminum) substances and neurodevelopment at six and 24 months of age. *Environmental Pollution* **187** 130–135.

MENICTAS, M. and WAND, M. (2013). Variational inference for marginal longitudinal semiparametric regression. *Stat* **2** 61–71.

ORMEROD, J. T. (2011). Grid based variational approximations. *Computational Statistics and Data Analysis* **55** 45–56.

ORMEROD, J. T. and WAND, M. P. (2010). Explaining Variational Approximations. *The American Statistician* **64** 140–153.

PARK, T. and CASELLA, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103(482)** 681–86.

PHAM, T. H., ORMEROD, J. T. and WAND, M. (2013). Mean field variational Bayesian inference for nonparametric regression with measurement error. *Computational Statistics & Data Analysis* **68** 375–387.

QIAN, P., WU, H. and WU, C. (2008). Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors. *Technometrics* **50** 383–96.

RAUH, V. A., GARFINKEL, R., PERERA, F. P., ANDREWS, H. F., HOEPNER, L., BARR, D. B., WHITEHEAD, R., TANG, D. and WHYATT, R. W. (2006). Impact of prenatal chlorpyrifos exposure on neurodevelopment in the first 3 years of life among inner-city children. *Pediatrics* **118** e1845–e1859.

ROBERTS, S. and MARTIN, M. A. (2006). The Use of Supervised Principal Components in Assessing Multiple Pollutant Effects. *Environmental Health Perspectives* **114** 1877–1882.

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* **71, Part 2** 319–392.

SAVITSKY, T., VANNUCCI, M. and SHA, N. (2011). Variable Selection for Nonparametric Gaussian Process Priors: Models and Computational Strategies. *Statistical science : a review journal of the Institute of Mathematical Statistics* **26** 130–149.

STILES, J. and JERNIGAN, T. (2010). The Basics of Brain Development. *Neuropsychology Review* **20(4)** 327–348.

TAU, G. Z. and PETERSON, B. S. (2010). Normal development of brain circuits. *Neuropsychopharmacology* **35(1)** 127–168.

TELLEZ-ROJO, M. M., BELLINGER, D. C., ARROYO-QUIROZ, C., LAMADRID-FIGUEROA, H., MERCADO-GARCIA, A., SCHNAAS-ARRIETA, L., WRIGHT, R. O., HERNANDEZ-AVILA, M. and HU, H. (2006). Longitudinal associations between blood lead concentrations lower than 10 microg/dL and neurobehavioral development in environmentally exposed children in Mexico City. *Pediatrics* **118(2)** e323–30.

WAND, M. (2014). Fully simplified multivariate normal updates in non-conjugate variational message passing. *Journal of Machine Learning Research* **15** 1351–1369.

WAND, M. P. and ORMEROD, J. T. (2012). Continued fraction enhancement of Bayesian computing. *Stat* **1** 31–41.

WAND, M. P., ORMEROD, J. T., PADOAN, S. A. and FR, R. (2011). Mean Field Variational Bayes for Elaborate Distributions. *Bayesian Analysis* **6** 847–900.

WANG, B. and TITTERINGTON, D. M. (2005). Inadequacy of interval estimates corresponding to variational Bayesian approximations. *Proc. 10th Int. Wrkshp Artificial Intelligence and Statistics* 373–380.

WARREN, J., M, F., A, H. and LANGLOIS, P. (2012). Spatial-Temporal Modeling of the Association between Air Pollution Exposure and Preterm Birth: Identifying Critical Windows of Exposure. *Biometrics* **68(4)** 1157–1167.

WARREN, J., M, F., A, H. and LANGLOIS, P. (2013). Air Pollution Metric Analysis While Determining Susceptible Periods of Pregnancy for Low Birth Weight. *Obstetrics and Gynecology* **2013** 1–9.

WINN, J. and BISHOP, C. M. (2005). Variational message passing. *Journal of Machine Learning Research* **6** 661–694.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B (Methodological)* **68(1)** 49–67.