HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

J Am Stat Assoc. 2022 ; 117(538): 561–573. doi:10.1080/01621459.2021.1962889.

A Cross-validated Ensemble Approach to Robust Hypothesis Testing of Continuous Nonlinear Interactions: Application to Nutrition-Environment Studies

Jeremiah Zhe Liu[*,1], **Wenying Deng**[1], **Jane Lee**[2,3], **Pi-i Debby Lin**[2], **Linda Valeri**[4], **David C. Christiani**[2,5], **David C. Bellinger**[2,3], **Robert O. Wright**[6], **Maitreyi M. Mazumdar**[2,3], **Brent A. Coull**[1,2]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[2]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[3]Department of Neurology, Boston Children's Hospital, Boston, MA, USA

[4]Department of Biostatistics, Columbia Mailman School of Public Health, New York, New York, USA

[5]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[6]Department of Environmental Medicine and Public Health, Icahn School of Medicine, New York, NY, USA

Abstract

Gene-environment and nutrition-environment studies often involve testing of high-dimensional interactions between two sets of variables, each having potentially complex nonlinear main effects on an outcome. Construction of a valid and powerful hypothesis test for such an interaction is challenging, due to the difficulty in constructing an efficient and unbiased estimator for the complex, nonlinear main effects. In this work we address this problem by proposing a Cross-validated Ensemble of Kernels (CVEK) that learns the space of appropriate functions for the main effects using a cross-validated ensemble approach. With a carefully chosen library of base kernels, CVEK flexibly estimates the form of the main-effect functions from the data, and encourages test power by guarding against over-fitting under the alternative. The method is motivated by a study on the interaction between metal exposures *in utero* and maternal nutrition on children's neurodevelopment in rural Bangladesh. The proposed tests identified evidence of an interaction between minerals and vitamins intake and arsenic and manganese exposures. Results suggest that the detrimental effects of these metals are most pronounced at low intake levels of the nutrients, suggesting nutritional interventions in pregnant women could mitigate the adverse impacts of *in utero* metal exposures on children's neurodevelopment.

* zhl112@mail.harvard.edu .

Keywords

Hypothesis Testing; Kernel Method; Ensemble Learning; Cross Validation; Nutrition-environment Interaction

1 Introduction

Investigation of the interplay between multiple lifestyle, biological, and environmental factors contributing to disease risk is a major goal in public health. Classic gene-environment and nutrition-environment studies focus primarily on the interaction between discrete factors (31; 8), or between discrete factors and the linear effect of a few continuous measurements (e.g. (28)). In recent years, however, recognizing the fact that populations are exposed to combinations of continuously-measured chemical and non-chemical factors that potentially have a nonlinear effect on outcome, there has been increasing interest in the best ways to statistically quantify the complex interplay of these continuous, nonlinear effects on health.

In this work we analyze data from a birth cohort study on the interaction between *in utero* exposure to a metal mixture and maternal nutrition intake during pregnancy on children's neurodevelopment in rural Bangladesh (13; 21). Bangladesh has been experiencing unparalleled levels of arsenic (As) other toxic metal poisoning through contaminated groundwater (33). Bangladesh also has rates of undernutrition that are among the highest in the world (42). A recent study (43) assessed the relationships between the arsenic (As), manganese (Mn), and lead (Pb) metal pollution mixture and infant neurodevelopment in Bangladesh, and has detected nonlinear, inverted-U shaped exposure–response relationships that differ among population subgroups. Its findings suggested a role of additional cultural/behavioral factors in affecting the impact of this metal mixture on children's health. One possible factor impacting these environmental effects is maternal nutrition during pregnancy. At vulnerable stages of fetal development, mother's overall nutrition intake may exacerbate adverse effects of chemical stressors. Specific nutrients may modify chemical effects because of their influence on the metabolism of the chemicals, on epigenetic programming in response to the chemicals, or through other mechanisms that vary by metal or by outcome. To answer this question, a companion nutritional study (26) was conducted to collect data on mother's nutrition intake during pregnancy, measuring the level of nutrition intake of 27 nutrients grouped in five nutrition categories (macronutrient, minerals, (pro-)vitamin As, vitamin Bs, and other vitamins), thereby providing an unique opportunity for researchers to quantitatively investigate the effect modification between nutrition intake during pregnancy and *in utero* metal exposures on infant development.

The Bangladesh study posed two challenges that are common in many modern data science applications: (1) high dimensionality of the interaction, as the interaction term contains second- and higher-order interactions between 27 nutrients and 3 metal exposures, and (2) the nonlinearity of the underlying exposure-response relationship, whose mathematical properties are unknown *a priori*. In such a scenario, linear-model based methods are known to suffer from misspecification of the main effects model for nutrients (that include nutrient-

nutrient interactions) and metals (including metal-metal interactions) even under the null of no nutrient-metal interactions, leading to inflated Type I error and reduced test power (40; 8; 46). To boost efficiency and incorporate nonlinearities in the exposure-response relationship, a recent line of research has focused on constructing interaction tests based on kernel machine regression (KMR) (36; 34). Building on the success of previous kernel testing literature (29; 47), these tests model the main-effect and interaction-effect functions as elements in reproducing kernel Hilbert spaces (RKHS) generated by pre-specified kernel functions, and build the hypothesis test by re-parametrizing the kernel machine regression as a linear mixed model (29). In this framework, the interaction term is an additional random effect term controlled by an univariate garrote parameter, on which one can construct a variance-component score test (27) for a test of the null hypothesis of no interaction. Successful applications of such tests include targeted gene effect identification in genetic pathway analysis (30), gene–gene interaction detection in genome-wide association study (24), and also in gene-environment interaction studies with discrete factor such as gender (4) and risk indicators of cardiovascular disease (12).

Applications of interaction tests involving sets of multiple continuous measurements with nonlinear effects, however, remain rare. The key challenge impeding the success of interaction tests in continuous settings lies in designing a proper kernel function for the multi-dimensional, nonlinear main-effect functions of unknown form. The kernel functions for the main effect terms need to generate a RKHS that is rich enough to contain the main-effect functions under the null, while at the same time be sufficiently structured to maintain power for detecting interactions. Earlier work (29; 30) approached this problem by selecting the kernel from an assumed parametric family (e.g. the Gaussian radial basis functions (RBF)) through maximum likelihood estimation, risking the specification of overly strong assumptions for these nonlinear functions. More recent approaches alleviate assumptions on the data-generation mechanism by incorporating multiple candidate kernels into the analysis, treating the kernel function as a weighted combination of candidate kernels, and learning kernel weights by maximizing various objective functions such as centered kernel alignment (10) or by an L_1-regularized model likelihood (37). However, designed primarily to maximize predictive accuracy, such procedures can be overly flexible under the alternative and potentially result in hypothesis tests with low power (49). Permutation tests are another popular approach for alleviating the issue of kernel misspecification (7; 49); however, constructing a permutation procedure for an interaction test is usually not possible in observational studies, since the gene-environment independence condition tends to not hold (6).

In this article, we propose a new approach to test for the interaction effect between groups of continuous features, each having potentially complex main effect functions relating outcome to that set of exposures. Built under the framework of kernel machine regression, we address the issue of kernel misspecification by deploying an ensemble of candidate kernels, and carefully design the ensemble strategy so that it minimizes the generalization error of the overall ensemble (11). Consequently, the proposed test automatically estimates the form of the kernel under the null from the data and guards against overfitting the interaction effect under the alternative, resulting in a powerful test that is robust under a wide range of data generation mechanisms. As we discuss in Section 3, such a strategy results in an estimator

that enjoys an oracle property for ensemble selection and good generalization performance in limited samples, thereby achieving a powerful null-model estimator especially suitable for hypothesis testing in epidemiology studies. We term our method the Cross-Validated Ensemble of Kernels (CVEK). In Section 4, we illustrate the robustness of our method by conducting simulation studies that evaluate the finite-sample performance (Type-I error and power) under a range of data-generating scenarios and compare the performance of the proposed approach with other popular interaction tests. Finally, in Section 5, we apply our method to data from the Bangladesh reproductive cohort study (13; 21) to investigate the interaction between mother's daily nutrient intake and *in-utero* exposure to an environmental metal mixture (As, Mn and Pb) on children's neurodevelopment.

2 Model and Inference

Assume we observe data from n independent subjects. For the i^{th} subject, let y_i be a continuous response, $\mathbf{x}_i$ be the set of p baseline covariates that can be entered into the model linearly, and $\mathbf{z}_i$ be the set of q continuous covariates that have a nonlinear effect on y_i. Furthermore, we assume that there exists a grouping structure among the $\mathbf{z}_i$ covariates such that $\mathbf{z}_i = \{\mathbf{z}_{1,i}, \mathbf{z}_{2,i}\}$, where the m^{th} group $z_{m,i} \in \mathbb{R}^{q_m}$ contains q_m covariates, $m = 1, 2$. We discuss the generalization to the case of more than two groups in $\mathbf{z}_i$ in Section 2.2.

We assume that the outcome y_i depends on covariates $\mathbf{x}_i$, $\mathbf{z}_i$ through the model:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i) + \epsilon_i \qquad \text{where } \epsilon_i \overset{iid}{\sim} N(0, \sigma^2), \tag{1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients for background covariates, $h(\mathbf{z}_i): \mathbb{R}^q \to \mathbb{R}$ is an unknown continuous function describing the effect of $\mathbf{z}_i$, and $\boldsymbol{\epsilon}_i$ is random noise that is independently and identically distributed as $N(0,\sigma^2)$. For identifiability purpose, h is assumed to be square-integrable and subject to the constraint $\int_{\mathbb{R}^q} h(\mathbf{z}) d\mathbf{z} = 0$.

Our main objective in this work is to test for the interaction between two chosen sets of covariates in $\mathbf{z}_i = \{\mathbf{z}_{1,i}\, \mathbf{z}_{2,i}\}$, while accounting for interactions within each covariate set. Without loss of generality, consider testing for the interaction between $\mathbf{z}_{1,i}$ and $\mathbf{z}_{2,i}$. Then our hypothesis is:

$$H_0: h \in \mathscr{H}_{12}^{\perp}, \tag{2}$$

where $\mathscr{H}_{12}$ is the functional space of "pure interaction" functions that contain only the interaction effect between $(\mathbf{z}_{1,i}, \mathbf{z}_{2,i})$. That is, under the null hypothesis, $h(\mathbf{z})$ may depend on the individual main effects of $\mathbf{z}_{1,i}$, $\mathbf{z}_{2,i}$, but does not depend on the interaction effect of the set pair $(\mathbf{z}_{1,i}, \mathbf{z}_{2,i})$.

We take the penalized likelihood approach to estimate parameters $(\boldsymbol{\beta}, h)$. Namely, we first specify $\mathscr{H}$ the candidate space and λ the penalty parameter, then estimate parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}, \hat{h}\right)$ by minimizing the penalized negative log likelihood:

$$\left(\boldsymbol{\beta}, \widehat{h}\right) = \underset{\boldsymbol{\beta} \in \mathbb{R}, h \in \mathscr{H}}{argmin} L_\lambda(\boldsymbol{\beta}, h), \quad \text{where} \quad L_\lambda(\boldsymbol{\beta}, h) = \sum_{i=1}^{n} \|y_i - \mathbf{x}_i\boldsymbol{\beta} - h(z_i)\|^2 + \lambda\|h\|_{\mathscr{H}}^2. \tag{3}$$

We model $\mathscr{H}$ using Kernel Machine Regression (KMR) (36). Specifically, we assume $\mathscr{H}$ to be a Reproducing Kernel Hilbert Space (RKHS) generated by a positive-definite kernel function $k(\mathbf{z}_i, \mathbf{z}_i')$, such that any $h \in \mathscr{H}$ can be expressed in terms the kernel function as $h(\mathbf{z}_i) = \langle h, k(\mathbf{z}_i, .)\rangle_{\mathscr{H}}$. Then by the Representer theorem (5), if we define $\mathbf{y}_{n\times 1} = [y_1, \ldots, y_n]^T$, $\mathbf{X}_{n\times p} = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T\right]^T$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$ and also denote $\mathbf{K}_{n\times n}$ the kernel matrix with its $(i, j)^{th}$ element to be $\delta = I(T \le C)$, then (3) can be re-written as

$$(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \underset{\boldsymbol{\beta} \in \mathbb{R}, \boldsymbol{\alpha} \in \mathbb{R}^n}{argmin} L_\lambda(\boldsymbol{\beta}, \boldsymbol{\alpha}), \quad \text{where } L_\lambda(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{K}\boldsymbol{\alpha}\|^2 + \lambda\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha} \tag{4}$$

Furthermore, if we define $\tau = \frac{\sigma^2}{\lambda}$, $\mathbf{h} = [h(\mathbf{z}_1), \ldots, h(\mathbf{z}_n)]^\top$ can arise exactly from a linear mixed model (LMM) (29)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{h} + \epsilon \quad \text{where} \quad \mathbf{h} \sim N(0, \tau\mathbf{K}) \quad \epsilon \sim N\left(0, \sigma^2\mathbf{I}\right). \tag{5}$$

2.1 A Variance Component Test for Kernel Interaction

Under the LMM formulation of the kernel machine regression model in (5), Maity and Lin (30) built a general test for the hypothesis $H_0: h \in \mathscr{H}_0$ by assuming that h lies in a RKHS generated by a *garrote kernel function* $k_\delta(\mathbf{z}, \mathbf{z}')$, which is constructed by attaching an extra *garrote parameter* δ to a regular kernel function. When $\delta = 0$, the garrote kernel function $k_0(\mathbf{z}, \mathbf{z}') = k_\delta(\mathbf{z}, \mathbf{z})|_{\delta=0}$ generates exactly $\mathscr{H}_0$ the space of functions under the null hypothesis. The authors further proposed a REML-based variance component score test for H_0.

In order to adapt the above approach to the hypothesis for interaction $H_0: h \in \mathscr{H}_{12}{}^{\perp}$, we construct the garrote kernel function $K_\delta(\mathbf{z}, \mathbf{z}')$ by building its corresponding RKHS for the main-effect and interaction spaces using the tensor-product construction (15; 16). Briefly, for the two sets of covariates $\mathbf{z}_m \in \mathbb{R}^{q_m}$, where $m = 1, 2$, let μ_m be the probability measure of $\mathbf{z}_m$ on $k \in \mathscr{D}$, let $\mathbf{1}_m = \left\{f: \mathbb{R}^{q_m} \to \mathbb{R} \mid f \propto 1\right\}$ be the RKHS of constant functions with kernel function $P_{max,k}$, and let $\mathscr{H}_m$ be the RKHS of centered and square-integrable functions on $\mathbf{z}_m$ (i.e. $\int h(\mathbf{z}_m)d\mu(\mathbf{z}_m) = 0$ and $\int h^2(\mathbf{z}_m)d\mu(\mathbf{z}_m) < \infty$). Now consider the space $\mathbf{1}_m \oplus \mathscr{H}_m$. Any function h in this space can be decomposed as $h = P_c h + (h - P_c h)$ with a constant component $P_c h = \int h(\mathbf{z}_m)d\mu(\mathbf{z}_m) \in \mathbf{1}_m$ and a centered non-constant component $(h - P_c h) \in \mathscr{H}_m$. As a result, the tensor product space $\mathscr{H} = \{\mathbf{1}_1 \oplus \mathscr{H}_1\} \otimes \{\mathbf{1}_2 \oplus \mathscr{H}_2\}$ adopts the following orthogonal decomposition:

$$\begin{aligned}\mathcal{H} &= \{\mathbf{1}_1 \otimes \mathbf{1}_2\} \oplus \{\mathcal{H}_1 \otimes \mathbf{1}_2\} \oplus \{\mathbf{1}_1 \otimes \mathcal{H}_2\} \oplus \{\mathcal{H}_1 \otimes \mathcal{H}_2\} \\ &= \{\mathbf{1}\} \oplus \{\mathcal{H}_1 \oplus \mathcal{H}_2\} \oplus \{\mathcal{H}_1 \otimes \mathcal{H}_2\},\end{aligned}$$

where **1** is the space of constant functions with support on $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2\}$ (24). This decomposition is shown to be unique under mild regularity conditions (i.e., the probability measures μ_m's are absolutely continuous and bounded away from zero and infinity) (39; 20). In the equation above, $\mathcal{H}_0 = \mathcal{H}_1 \oplus \mathcal{H}_2$ is the space of main-effect functions that does not contain the $(\mathbf{z}_{1,i}, \mathbf{z}_{2,i})$ interaction, and $\mathcal{H}_{12}{}^{\perp} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is the space of "pure interaction" function whose elements describe only the interaction effect between $(\mathbf{z}_{1,i}, \mathbf{z}_{2,i})$ and are orthogonal to the main-effect functions. Consequently, we can construct the garrote kernel function for the tensor product space $\mathcal{H}$ as

$$k_\delta(\mathbf{z}, \mathbf{z}') = k_0(\mathbf{z}, \mathbf{z}') + \delta^* k_{12}(\mathbf{z}, \mathbf{z}') \tag{6}$$

where $k_0(\mathbf{z}, \mathbf{z}') = k_1(\mathbf{z}, \mathbf{z}') + k_2(\mathbf{z}, \mathbf{z}')$ is the kernel function for $\mathcal{H}_0 = \mathcal{H}_1 \oplus \mathcal{H}_2$ that corresponds to the null hypothesis of no interaction, and $k_{12}(\mathbf{z}, \mathbf{z}') = k_1(\mathbf{z}, \mathbf{z}')^* k_2(\mathbf{z}, \mathbf{z}')$ is the kernel function for space of interaction-effect functions $\mathcal{H}_{12}^{\perp} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Finally, notice that k_δ does not include the kernel function for the space of the constant functions **1** since this is already modeled by the intercept term.

Under the above form of the garrote kernel function, the derivative of the kernel function with respect to the garrote parameter is $\frac{\partial}{\partial \delta} k_\delta(\mathbf{z}, \mathbf{z}') = k_{12}(\mathbf{z}, \mathbf{z}')$, i.e., the kernel function that corresponds to $\mathcal{H}_{12}^{\perp}$. Therefore given n data points $\{y_i, \mathbf{x}_i, \mathbf{z}_i\}_{i=1}^n$, the derivative kernel matrix $\partial \mathbf{K}_0$ under the null is simply the $n \times n$ kernel matrix $\mathbf{K}_{12}$ whose $(i, j)^{th}$ element is $k_{12}(\mathbf{z}_i, \mathbf{z}_j)$, and the score test statistic is:

$$\begin{aligned}\hat{T}_0 &= \hat{\tau}^* (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathrm{V}_0^{-1} \partial \mathbf{K}_0 \mathrm{V}_0^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \hat{\tau}^* (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathrm{V}_0^{-1} \mathbf{K}_{12} \mathrm{V}_0^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).\end{aligned} \tag{7}$$

where $\mathrm{V}_0 = \hat{\sigma}^2 \mathbf{I}_{n \times n} + \lambda \mathbf{K}_0$ is the marginal covariance matrix of $\mathbf{y}_{n \times 1}$, $\mathbf{K_0}$ is the $n \times n$ kernel matrix whose $(i, j)^{th}$ element is $k_0(\mathbf{z}_i, \mathbf{z}_j)$, and $\left(\hat{\beta}, \hat{\sigma}^2, \hat{\tau}\right)$ are the model parameters estimated under the null hypothesis (30). The null distribution of $\hat{T}_0$ is a mixture of chi-squares that can be approximated using a scaled chi-square distribution $\kappa \chi_\nu^2$ using either Satterthwaite-Welch method (48) or other higher-moment approximations (3).

2.2 Generalization to Multiple Groups with Nuisance Interaction

Our description so far assumes there exists no nuisance interaction terms in the model $y = \mathbf{x}^T\boldsymbol{\beta} + h(\mathbf{z}) + \epsilon$. However, in more realistic scenario, $\mathbf{z}$ usually exhibits complex hierarchical structure subsuming multiple groups, and it is often of interest to test only for the interaction between two small subgroups of $\mathbf{z}$, leaving other interactions as nuisance effect to be accounted for by the null model. For example, consider the case of nutrition-

environment interaction in Bangladesh birth cohort, $\mathbf{z}_i$ is the 30 × 1 vector of during-pregnancy exposure to 27 nutrients and 3 metal pollutants, corresponding the grouping structure $\mathbf{z}_i = \{\mathbf{z}_{\text{metal}}, \mathbf{z}_{\text{nutr}}\}$, where $\mathbf{z}_{\text{nutr}}$ is further divided into $\mathbf{z}_{\text{nutr}} = \{\mathbf{z}_{\text{macro}}, \mathbf{z}_{\text{mineral}}, \mathbf{z}_{\text{vitA}}, \mathbf{z}_{\text{vitB}}, \mathbf{z}_{\text{vitO}}\}$. Therefore, when testing for the interaction between metal mixture exposures and a specific nutrient group of interest, care should be given to formulate $h(\mathbf{z}_i)$ such that it not only explicitly characterizes the interaction of interest, but also account for all nuisance interactions among other $\mathbf{z}_i$ subgroups.

More specifically, assume $\mathbf{z}_i = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$, when testing for the interaction between $\mathbf{z}_1$ and $\mathbf{z}_2$, the *nuisance interactions* between $\mathbf{z}_1$ and $\mathbf{z}_3$, as well as between $\mathbf{z}_2$ and $\mathbf{z}_3$, should also be included in the null model. To this end, following the tensor-product construction in Section 2.1 and under the same regularity conditions, $h(\mathbf{z})$ adopts an unique orthogonal decomposition:

$$h(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) = [h_1(\mathbf{z}_1) + h_2(\mathbf{z}_2) + h_3(\mathbf{z}_3)] + [h_{12}(\mathbf{z}_1, \mathbf{z}_2) + h_{13}(\mathbf{z}_1, \mathbf{z}_3) + h_{23}(\mathbf{z}_2, \mathbf{z}_3)] + h_{123}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3), \tag{8}$$

where for $\mathscr{H}_m$ the RKHS of centered and square-integrable functions, h_m are the main-effect functions such that $h_m \in \mathscr{H}_m$, and $h_{m_1 m_2}$ and $h_{m_1 m_2 m_3}$ are the higher-order interaction functions that belong to $\mathscr{H}_{m_1 m_2} = \mathscr{H}_{m_1} \otimes \mathscr{H}_{m_2}$ and $\mathscr{H}_{m_1 m_2 m_3} = \mathscr{H}_{m_1} \otimes \mathscr{H}_{m_2} \otimes \mathscr{H}_{m_3}$, respectively. Under such construction, the null hypothesis of no interaction between $\mathbf{z}_1$ and $\mathbf{z}_2$ corresponds to h_{12} and h_{123} equaling zero, i.e.

$$H_0: \quad h = h_1 + h_2 + h_3 + h_{13} + h_{23}$$

$$H_a: \quad h = h_1 + h_2 + h_3 + h_{13} + h_{23} + h_{12} + h_{123},$$

and the corresponding garrote kernel for the null hypothesis is $k_\delta(\mathbf{z}, \mathbf{z}') = k_0(\mathbf{z}, \mathbf{z}') + \delta * k_a(\mathbf{z}, \mathbf{z}')$, where $k_0 = k_1 + k_2 + k_3 + k_{13} + k_{23}$ and $k_a = k_{12} + k_{123}$. Here k_m is the reproducing kernels for the main-effect space H_m, and the higher-order interaction kernels are constructed as $\forall(i, j), k_{ij} = k_i * k_j$ and $k_{123} = k_1 * k_2 * k_3$ similar to Section 2.1. Consequently, denoting $\mathbf{K}_m$ as the $n \times n$ kernel matrix corresponding to k_m, the null kernel matrix $\mathbf{K}_0$ and the interaction kernel matrix $\mathbf{K}_{12}$ are $n \times n$ matrices that are computed as:

$$2\mathbf{K}_0 = \mathbf{K}_1 + \mathbf{K}_2 + \mathbf{K}_3 + \mathbf{K}_1 \circ \mathbf{K}_2 + \mathbf{K}_2 \circ \mathbf{K}_3$$

$$\mathbf{K}_{12} = \mathbf{K}_1 \circ \mathbf{K}_2 + \mathbf{K}_1 \circ \mathbf{K}_2 \circ \mathbf{K}_3 .$$

where ° indicates the Hadamard (i.e., element-wise) product. As a result, the test statistic can be constructed as in (7).

3 Robust Effect Estimation using Cross-validated Ensemble

We motivate the importance of robust null model estimation by considering the possible impact of a misspecified null kernel function k_0 on the performance of the resulting hypothesis test. Specifically, we express the test statistic $\hat{T}_0$ in (7) in terms of the model residual $\boldsymbol{\epsilon} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{h}$:

$$\hat{T}_0 \propto \epsilon^T \mathbf{K}_{12} \epsilon, \tag{9}$$

where we have used the fact $\mathrm{V}_0^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \left(\hat{\sigma}^2\right)^{-1}(\epsilon)$ (17). Therefore, the test statistic $\hat{T}_0$ is a scaled quadratic-form statistic that is a function of the model residual. If k_0 is too restrictive, model estimates will underfit the data under the null hypothesis, introducing extraneous correlation among the $\hat{\epsilon}_i$'s that yield inflated $\hat{T}_0$ values and deflated p-values under the null. Therefore, this approach will yield an invalid test having inflated Type I error. On the other hand, if k_0 is too flexible, model estimates will likely overfit the data in small samples, producing underestimated residuals, which leads to underestimated test statistics and overestimated p-values. Accordingly, the resulting test will have low power.

The above observations motivate a kernel estimation strategy that is flexible in that it does not underfit under the null, yet stable so that it does not overfit under the alternative. To this end, we propose estimating h using the convex ensemble of a library of fixed base kernels $\{k_d\}_{d=1}^D$:

$$\hat{h}(\mathbf{x}) = \sum_{d=1}^{D} u_d \hat{h}_d(\mathbf{x}) \qquad \mathbf{u} \in \Delta = \left\{ \mathbf{u} \middle| \mathbf{u} \geq 0, \mathbf{1}^T \mathbf{u} = 1 \right\}, \tag{10}$$

where $\hat{h}_d$ is the kernel predictor generated by d^{th} base kernel k_d. In order to maximize model stability, we divide data $\mathscr{D} = \{(y_i, \mathbf{x}_i, \mathbf{z}_i)\}_{i=1}^n$ into a training and validation set $\mathscr{D}_{\text{train}}$, $\mathscr{D}_{\text{valid}}$ and estimate the ensemble weights $\mathbf{u}$ to minimize the overall cross-validation error of $\hat{h}$ on $\mathscr{D}_{\text{valid}}$. We term this method the *Cross-Validated Ensemble of Kernels* (CVEK). The exact algorithm proceeds in three stages as follows (see Algorithm 1 for summary):

Stage 1: Candidate Kernel Predictors

For each basis kernel in the library $\{k_d\}_{d=1}^D$, we first standardize the kernel matrix by its trace $\mathbf{K}_d = \mathbf{K}_d / tr(\mathbf{K}_d)$, and then estimate the prediction based on each kernel as $\mathbf{h}_{d,\hat{\lambda}_d} = \mathbf{K}_d\left(\mathbf{K}_d + \hat{\lambda}_d \mathbf{I}\right)^{-1}\mathbf{y}$, $d \in \{1, \ldots, D\}$ where the tuning parameter $\hat{\lambda}_d$ is selected by minimizing the k–fold cross-validation error on $\mathscr{D}_{\text{train}}$, and compute the validation cross-validation error on $\mathscr{D}_{\text{valid}}$ for the d^{th} kernel as $\hat{\epsilon}_d = CV\left(\hat{\lambda}_d | \mathbf{K}_d\right)$.

Stage 2: Cross-validated Ensemble

Using the estimated validation cross-validation errors $\{\hat{\epsilon}_d\}_{d=1}^{D}$, estimate the ensemble weights $\mathbf{u} = \{u_d\}_{d=1}^{D}$ by minimizing the overall cross-validation error $\epsilon_{\mathbf{u}} = \sum_{d=1}^{D} u_d \hat{\epsilon}_d$:

$$\mathbf{u} = \underset{\mathbf{u} \in \Delta}{argmin} \left\| \sum_{d=1}^{D} u_d \hat{\epsilon}_d \right\|^2, \qquad \text{where } \Delta = \left\{ \mathbf{u} \middle| \mathbf{u} \geq 0, \mathbf{1}^T \mathbf{u} = 1 \right\},$$

and produce the final ensemble prediction $\mathbf{h} = \sum_{d=1}^{D} \hat{u}_d \mathbf{h}_d = \sum_{d=1}^{D} \hat{u}_d \mathrm{A}_{d, \hat{\lambda}_d} \mathbf{y} = \mathrm{A}\mathrm{y}$, where $\mathrm{A} = \sum_{d=1}^{D} \hat{u}_d \mathrm{A}_{d, \hat{\lambda}_d}$ is the ensemble hat matrix.

Stage 3: Ensemble Kernel Matrix

Using the ensemble hat matrix $\mathbf{A}$, estimate the ensemble kernel matrix $\mathbf{K}$ by solving $\mathbf{K}(\mathbf{K} + \lambda\mathbf{I})^{-1} = \mathbf{A}$. Specifically, if we denote $\mathbf{U}_A$ and $\{\delta_{A,k}\}_{k=1}^{n}$ as the eigenvector and eigenvalues of $\mathbf{A}$, respectively, then the ensemble kernel matrix $\mathbf{K}$ adopts the form:

$$\mathbf{K} = \lambda_{\mathbf{K}} * \left[\mathbf{U}_A \mathrm{diag}\left(\frac{\delta_{A,k}}{1 - \delta_{A,k}} \right) \mathbf{U}_A^T \right], \tag{11}$$

where we recommended setting $\lambda_{\mathbf{K}} = min\left(1, \left(\sum_{k=1}^{n} \frac{\delta_{A,k}}{1 - \delta_{A,k}} \right)^{-1} \right)$ (see Supplementary Section A).

We remind readers that the CVEK's ensemble form (Stage 2) belongs to the general class of model aggregation method known as *convex aggregation* (41), whose oracle property in model selection has been established both asymptotically and in finite-sample (44; 23). It can be also considered as a special case of *ensemble of kernel predictors* (EKP) (9), whose generalization behavior is well characterized in terms of the rate of eigenvalue decay of the base kernels. Consequently, under the null hypothesis, with a diverse set of base kernels, the CVEK ensemble converges in $O\left(\frac{1}{n}\right)$ rate to the "oracle ensemble" made by an oracle that has access to infinite amount of validation data, thereby resulting in correct Type I error by mitigating null model misspecification. Under the alternative, by setting the diverse kernel library to be a mix of parametric kernels (linear, polynomial) and smooth kernels of exponential eigendecay rate (e.g. a collection of Gaussian RBF kernel with different fixed spatial smoothness parameters), CVEK converges to its asymptotic counterpart in $O\left(\frac{1}{n}\right)$ rate if the data-generation function is indeed parametric, and in the "near-parametric" rate of $O\left(\frac{log(n)}{n}\right)$ if the data-generation function is complex and nonlinear, thereby encouraging good test power by not overfitting the interaction effect due to fast generalization rate. The resulting ensemble kernel is therefore a strong candidate for a null model estimator that is suitable for hypothesis testing. We refer readers to Supplementary Section B for detailed discussion.

4 Numeric Studies

We evaluate the finite-sample performance of the proposed interaction test in a simulation study that mimics a small-sample nutrition-environment interaction study. We generate the fixed-effect covariates $\mathbf{x}_i \in \mathbb{R}^{p_x}$ and the input features $(\mathbf{z}_{i,1}, \mathbf{z}_{i,2}) \in \mathbb{R}^{p1} \times \mathbb{R}^{p2}$ independently from a standard multivariate Gaussian distribution. Here, $(\mathbf{z}_{i,1}, \mathbf{z}_{i,2})$ reflects each subject's level of exposure to p_1 environmental pollutants and the levels of a subject's intake of p_2 nutrients during the study. We generate the outcome y_i as:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + h_1(\mathbf{z}_{i,1}) + h_2(\mathbf{z}_{i,2}) + \delta * h_{12}(\mathbf{z}_{i,1}, \mathbf{z}_{i,2}) + \epsilon_i, \tag{12}$$

where the fixed-effect coefficients $\boldsymbol{\beta}_{p_x \times 1}$ is sampled from a standard Gaussian distribution. The nonlinear functions h_1, h_2, h_{12} are sampled from RKHSs $\mathscr{H}_1$, $\mathscr{H}_2$ and $\mathscr{H}_1 \otimes \mathscr{H}_2$, generated using a ground-truth kernel k_{true}. We standardize all sampled functions to have unit norm, so that δ represents the strength of interaction relative to the main effect. For the main section of the numeric study, we consider sample size $n = 200$, data dimension $p_x = 5$, $p_1 = p_2 = 3$ and $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2 = 0.1)$. For each simulation scenario, we first generate data using δ and and k_{true}, and then use a k_{model} to estimate the null model and obtain p-value using the proposed test. We repeat each scenario 200 times, and evaluate the test performance using the empirical probability $\hat{P}(p \leq 0.05)$.

In this study, we vary k_{true} to produce data-generating functions $h_\delta(\mathbf{z}_{i,1}, \mathbf{z}_{i,2})$ with different smoothness and complexity properties, and vary k_{model} to reflect different common modeling strategies for the null model in addition to using CVEK. We then evaluate how these two aspects impact the hypothesis test's Type I error and power. More specifically, we sample the data-generating function using k_{true} from Matérn kernel family (34):

$$k(\mathbf{r}|\nu, \sigma) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\sqrt{2\nu}\sigma\|\mathbf{r}\|)^{\nu} K_\nu(\sqrt{2\nu}\sigma\|\mathbf{r}\|), \quad \text{where} \quad \mathbf{r} = \mathbf{x} - \mathbf{x}',$$

with two non-negative hyperparameters (ν, σ). For a function h sampled using a Matérn kernel, ν determines the function's smoothness (i.e. degree of mean-square differentiability), and σ determines the function's complexity in terms of spectral frequency (34).

In this work, we vary $\nu \in \left\{\frac{3}{2}, \frac{5}{2}, \infty\right\}$ to generate once-, twice, and infinitely-differentiable functions, and vary $\sigma \in \{0.5, 1, 1.5\}$ to generate functions with varying degree of complexity.

We consider 12 k_{model} 's that are grouped into five model families (See Table 1 for a complete summary): (1) **Polynomial Kernels** that is equivalent to polynomial ridge regression. In this work, we use the **linear** kernel $k_{linear}(\mathbf{x}, \mathbf{x}'|p) = \mathbf{x}^T\mathbf{x}'$ and **quadratic** kernel $k_{\text{quad}}(\mathbf{x}, \mathbf{x}'|p) = \left(1 + \mathbf{x}^T\mathbf{x}'\right)^2$. (2) **Gaussian RBF Kernels**: $k_{RBF}(\mathbf{x}, \mathbf{x}'|\sigma) = exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2/\sigma^2\right)$ is a general-purpose kernel family that generates nonlinear, but very smooth (infinitely differentiable), functions. Under this kernel, we consider two hyperparameter selection strategies commonly seen in application: **RBF-Median** where we set σ to the sample

median of $\left\{\|\mathbf{x}_i - \mathbf{x}_j\|\right\}_{i \neq j}$, and **RBF-MLE**, which estimates σ by maximizing the model likelihood. (3) **Matérn** and (4) **Neural Network Kernels** are two flexible kernel families both containing a rich space of candidate functions. For Matérn kernel, we use **Matern 1/2**, **Matern 3/2** and **Matern 5/2**, corresponding to flexible models that is capable of approximating non-differentiable, once-differentiable, and twice-differentiable functions. Neural network kernels (34), on the other hand, represent a 1-layer Bayesian neural network with σ being the prior variance on the hidden weights, and it is theoretically capable of approximate arbitrary continuous functions on the compact domain (19). In this work, we let **NN 0.1**, **NN 1** and **NN 10** denote Bayesian neural networks with different prior constraints $\sigma \in \{0.1,1,10\}$. Finally, we evaluate the performance of the (5) **Cross-validated Kernel Ensemble** estimator we propose here. Specifically, we consider a CVEK estimator based on a Gaussian RBF kernel with $log(\sigma) \in \{-2, -1, 0,1, 2\}$, which we label **CVEK-RBF**. Furthermore, to evaluate the consequence of more flexible kernel families on ensemble behavior, we also consider **CVEK-NN**, which is a ensemble of neural network kernels with $\sigma \in \{0.1,1,10,50\}$).

The results of our primary simulation scenario ($n = 200$ and $p_1 = p_2 = 3$) are presented graphically in Figure 1. We first observe that for reasonably specified values of k_{model}, the proposed hypothesis test with CVEK estimator always has the correct Type I error and reasonable power. We also observe that the complexity of the data-generating function h_δ (12) plays a role in test performance, in the sense that the power of the hypothesis tests increases as the Matérn k_{true} 's complex parameter σ becomes larger, which corresponds to functions that put more weight on the simpler, slow-varying eigenfunctions in Bochner's spectral decomposition (34).

There exist clear differences in test performance between different model families. In general, polynomial models (**linear** and **quadratic**) appear to be too restrictive and underfit the data under both the null and the alternative, producing inflated Type I error and diminished power. On the other hand, lower-order Matérn kernels (**Matérn 1/2** and **Matérn 3/2**, dark blue lines) appear to be too flexible, due to their slow eigenvalue decay. Whenever data are generated from similarly or smoother k_{true}, **Matérn 1/2** and **3/2** overfit the data and produce deflated Type I error and severely diminished power, even if the hyperparameter σ is fixed at its true value. Comparatively, Gaussian RBF works well for a wider range of k_{true} 's but only if the hyperparameter σ is selected carefully. Specifically, **RBF-Median** (black dashed line) works generally well, despite being slightly conservative (i.e. lower power) when the data-generation function is smooth and of low complexity. **RBF-MLE** (black solid line), on the other hand, tends to overfit the data and exhibits weak power especially in higher dimension and for complex data generation functions (45). Neural Network kernels also perform well for a wide range of k_{true} and with the Type I error more robust to the specification of hyperparameters. Finally, the two ensemble estimators **CVEK-RBF** and **CVEK-NN** perform as well or better than the non-ensemble approaches for all k_{true} 's, despite being slightly conservative under the null.

To understand how the performance of CVEK depends on the dimension of the inputs, we conduct additional studies for $p_1 = p_2 = 6$ and $p_1 = p_2 = 10$, and report the results in Figure

C.1–C.2 in the Supplementary Material (Section C). Briefly, there is a clear effect of the data dimension and the complexity of the data generation mechanism on the test power. As the dimension $p_1 = p_2 = p$ increases, we observe a consistent pattern of power degradation in test performance that strongly depends on the complexity of the data-generating function h_δ. For example, consider the test power of **CVEK-NN** at an interaction strength $\delta = 1$ for k_{true} = Gaussian RBF. As data dimension increases from $p = 3$ (Figure 1) to $p = 10$ (Figure C.2), test power degrades to as low as 0.2 for highly complex data-generating functions ($\sigma = 0.5$), around 0.6 for moderately complex data-generating functions $\sigma = 1$, yet remains close 1.0 for smooth data-generating functions ($\sigma = 1.5$). Comparing different model families in high dimensions ($p = 10$; Figure C.1–C.2), we find that the polynomial models either have difficulty maintaining the correct Type I error, or have weak power under the alternative. Among the nonlinear models, the **RBF-MLE** model's test power degrades particularly quickly due to overfitting. Interestingly, we also observe that the hypothesis test based on NN-based models (e.g., **CVEK-NN**) becomes more powerful than the RBF-based models in higher data dimension, which is consistent with the recent theoretical observations on the effectiveness of neural network models in high-dimensional scenarios (1).

5 Nutrition-Environment Analysis for Child Neurodevelopment in Bangladesh Birth Cohort

We use the proposed methods to test for nutrition-environment interactions for the child neurodevelopment in the Bangladesh birth cohort study. Section 5.1 presents a detailed description of the Bangladesh birth cohort study, and the Section 5.2 present the analysis. Our aim is to detect whether mother's nutrient intake during pregnancy modifies the effect of metal mixture exposures on children's early-stage fine motor BSID-III scores in the district of Pabna (n = 351).

5.1 Study Background

The Bangladesh Reproductive Cohort Study (Project Jeebon) was initiated in 2008 to investigate the effects of prenatal and early childhood exposure to As, Mn and Pb on early childhood development. During 2008–2011, pregnant female participants (with gestational age < 16 weeks) were recruited from two rural health clinics operated by the Dhaka Community Hospital Trust (DCH) in the Sirajdikhan and Pabna Sadar upazilas of Bangladesh. During 2008–2013, data were collected at five time points spanning the entire perinatal and early childhood period, including: initial clinic visit (gestational age < 16 weeks, Visit 1); pre-delivery clinic visit (gestational age = 28 weeks, Visit 2), time of delivery (Visit 3), post-delivery clinic visit (infant age less than 1 month, Visit 4), and a postnatal follow-up visit (infant age between 20–40 weeks, Visit 5). Our central hypothesis is that children born from mother who had lower nutrient intake will be the most susceptible to adverse effects of metal exposures.

Detailed procedures for data collection and measurement protocols have been documented previously (13; 21; 43). Briefly, background information on parent's demographic status, including age, education, smoking history and socioeconomic status were collected through structured questionnaires at the two clinic visits during pregnancy (Visits 1–2). Information

on infant's biometric measurements, including sex, birth weight, length, head circumference, birth order and gestational age, were recorded at birth. Information on maternal medical history, maternal depression status (in Edinburgh Depression scale), maternal IQ (assessed using the Raven's Progressive Indices (35)) were measured during the pregnancy visits (Visits 1–2), and an infant's early childhood development, medical history, and quality of home environment (in terms of emotional, social, and cognitive stimulation, measured by Home Observation Measurement of Environment (HOME) instrument score (2) were measured during the follow-up visits (Visits 4–5), respectively.

Each infant's exposure to multiple metals As, Mn and Pb (concentrations in $\mu g / dL$) during pregnancy were measured using blood samples from infant's umbilical cord venous blood collected at the time of the birth. Mother's overall nutrition intake status during pregnancy was measured for 27 nutrients derived from semi-quantitative Food Frequency Questionnaires (FFQs) specially adapted to Bangladeshi diet (25) at both the pre- and post-delivery visits (Visit 2 and 4). This instrument derives data on these 27 nutrients from measures of the consumption frequency (amount per week) of 42 food items during the 12-month period preceding delivery. The nutrients measured can be grouped into 5 categories including macro-nutrients (5 nutrients: protein, fat, carbohydrate, dietary fiber and ash), minerals (7 nutrients: calcium, magnesium, phosphorus, potassium, sodium, zinc and copper), vitamin A and provitamin As (6 nutrients: vitamin A, retinol, beta-carotene equivalents, alpha-carotene, beta-carotene, and cryptoxanthin), vitamin B (5 nutrients: thiamin (B1), riboflavin (B2), niacin (B3), vitamin B6 and folate (B9)), and other vitamins (3 nutrients: vitamin C (i.e. L-ascorbic acid), vitamin D, and vitamin E). Finally, infant's neurodevelopmental outcomes were assessed at 20–40 months of age (Visit 5) using a translated and culturally-adapted version of the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) including five cognitive domains: cognitive, receptive language, expressive language, fine motor and gross motor.

We compare our method with three existing approaches for testing high-dimensional interaction. The (1) *Interaction Sequence Kernel Association Test* (iSKAT)(28) is a baseline approach that assumes linear relationship between exposures and outcome. It estimates the null model using ridge linear regression and corresponds to the **linear** model in simulation. (2) The *Gaussian Kernel Machine* test (GKM) (30) estimates the null model using kernel machine regression with Gaussian RBF kernels and tunes the kernel hyperparameter by maximizing REML. It correspondes to the **RBF-MLE** model in simulation. Finally, the (3) GE-spline test (18) which uses the generalized additive regression to model the nonlinear effect of environmental exposures using spline sieves. It can be considered as a special case of kernel machine regression with the kernel matrix constructed adaptively using spline basis functions (22). In order to visualize the identified interaction and thereby provide interpretable findings, we graphically summarize the multivariate interaction effects by examining the joint exposure-response surface between the principal components of the pollutant mixture and those for each nutrient group.

5.2 Data Analysis

In this nutrition-environment interaction study, our interest concentrates on the interaction between the mixture of As, Mn and Pb, and five major nutrient groups: *macro-nutrient*, *mineral*, *vitamin A*, *vitamin B* and the other vitamins (denoted as *vitamin, other*). For each of the five nutrient groups, we test for the overall interaction between the selected group and the joint effect of the As, Pb, Mn mixture. We adjust for parent's demographic status (age, education, smoking history), infant's biometric measurements at birth (sex, birth weight, length, head circumference, birth order and gestational age), and quality of early-childhood home environment (HOME score, maternal depression scale, maternal IQ).

5.2.1 Nutrient - Mixture Interactions—Table 2 presents p–values for the interaction between the overall metal mixture and each of the five nutrient groups. We conducted the proposed test using two types of CVEK models: an ensemble of seven RBF kernels with bandwidth parameter set between $log(\sigma^2) \in \{-3, -2, -1, 0, 1, 2, 3\}$ (denoted as CVEK-RBF), and an ensemble of seven neural network kernels with prior parameters set between $log(\sigma^2) \in \{-3, -2, -1, 0, 1, 2, 3\}$ (denoted as CVEK-NN). We compared the results of each to those generated by iSKAT, GKM, and GE-Spline. As shown in the table, most tests yielded strong evidence of interaction ($p < 0.05$) for the *vitamin A* and the *vitamin, other* groups, as well as weak-to-moderate evidence of interaction ($p \lessapprox 0.1$) for the *mineral* and the *vitamin B* groups. There was no evidence of an interaction between metal exposures and macro-nutrients.

Comparing the performance across different tests, we observed similar patterns for p-values for CVEK-NN and CVEK-RBF, suggesting robustness in test performance with respect to the choice of the family of the base kernels. We also observed higher values of p-values for the iSKAT (linear kernel) and GKM, the latter of which used a single RBF kernel with REML-based hyperparameter tuning. The statistical conclusions from these two tests are similar to those from the CVEK tests for *vitamin A* and *vitamin, other* groups of nutrients (at the significance level of 0.05). However, they are less powerful in detecting the interaction for the *mineral* and the *vitamin B* groups. This is consistent with our observation in Section 4 that, when the true effect is smooth, nonlinear and exhibits a moderate level of complexity (a scenario that is likely to hold for the effect of environmental exposures, see Figure 1 (h)), the hypothesis test based on GKM is slightly more powerful than that based on the iSKAT but is less powerful than the CVEK-based test. This reduction in power is possibly due to the overly strong smoothness assumption imposed by these two models. Finally, we notice that the performance of the test from the GE-spline model appears sub-optimal when compared to that of the other methods. GE-spline produced much higher p–values for all nutrient groups, failing to detect the interaction for the A vitamins and the other vitamins. We hypothesize that the observed instability of GE-spline is likely caused by the lack of fit of the null model, due to the difficulty in estimating multivariate splines in high dimensions.

5.2.2 Visualization of Exposure-Response Surface—To better understand the nature of the multivariate interactions between the environmental exposures and nutrition, in Figure 2 and 3, we visualize the fitted exposure-response surface relating the mean normalized fine motor BSID-III score and the principal components (PCs) of the pollution

mixture and of the nutrient groups. Every panel in Figure 2 and 3 depicts the joint effect of a pollutant PC and a nutrient PC for a selected nutrient group on the fine motor score, holding all the other PCs at their median. For each joint-effect term, the strength of evidence of an interaction between metal exposure and nutrition is driven by the ”importance” of the corresponding PCs, i.e. the amount of variation the corresponding PCs explain in their respective pollutant/nutrient group. For example, in Figure 2, the pollutant PCs account for 42.6%, 37.3% and 20.1% of the total variation in pollutant mixture, and the nutrient PCs account for 63.5%, 28.5% and 7.4% of the total variation in the macronutrient group. Consequently, the strength of the signal of the interaction between the 1^{st} PCs (e.g. Figure 2 (a)) in the overall interaction is expectedly much stronger than that between the 3^{rd} PCs (e.g. Figure 2 (i)). This explains the lack of significant evidence of overall interaction for the macronutrients in Table 2, since the joint effect between the PCs accounting for more variance (e.g. Figure 2 (a),(b) and (d)) do not display strong evidence of interaction. In comparison, for the other four nutrient groups, evidence of interaction can be observed between at least two nutrient-pollutant PC pairs among their leading PCs (Figure 3), thereby suggesting evidence of overall interaction between nutrients and the pollutant mixture, and consequently providing additional evidence for the findings from the CVEK tests in Table 2. Finally, we observe that across all nutrient groups, the nutrient PCs interacts the most often with the 1^{st} pollutant PC, which is strongly associated with As, and also with the 3^{rd} pollutant PC, which is strongly associated with Mn, suggesting that As and Mn are the two main pollutants driving the overall interaction. Furthermore, the pattern of interaction between nutrient and pollutant PCs are observed to be similar across nutrient groups: at lower levels of nutrients (x-axis), higher levels of metal exposure (y-axis) is associated with lower neurodevelopment scores. At intermediate or high levels of the nutrient, however, this negative association either disappears (see, e.g. Figure 3 (c), (d), (h)) or even becomes positive (see, e.g. Figure 3 (a), (b), (f)).

6 Discussion

Under the framework of kernel machine regression, we have developed a hypothesis testing procedure for detecting nonlinear interactions between groups of continuous covariates. In this context, we identified the unique challenge of possible kernel misspecification for the main-effect terms in the model, and illustrated the negative consequences of misspecified main effect kernels both in terms of Type I error and power. Specifically, we showed that an overly smooth model, even when including all causal covariates, can still underfit the data under the null and thereby produce inflated Type I error rates. On the other hand, an overly flexible model tends to overfit the data under both the null and the alternative, resulting in deflated Type I error and weak power. While these observations motivate careful selection of the form of the main effect kernels, we also observe that choice of regularization parameters via a likelihood-based model selection strategy (for example, estimating the bandwidth parameter in a Gaussian RBF kernel via REML (30)) can also over-smooth the main-effect terms under the null. This situation appears to be especially severe in limited sample sizes and for misspecified kernel functions (Figure 1 (a)–(c)). Our work addresses this challenge by estimating the main-effect model using a flexible ensemble of carefully selected base kernels, which we term Cross-validated Ensemble of Kernels (CVEK), coupled with a

hyperparameter selection strategy based on cross validation. This approach avoids kernel misspecification under the null and mitigates overfitting under the alternative, resulting in tests that are powerful yet maintain nominal Type I error rates. We validated the approach through extensive numerical studies. Under a wide variety of data-generation mechanisms, CVEK consistently produced correct Type I error and reasonable power.

We applied the proposed method to estimate nutrition-environment interactions between exposure to a metal mixture and multiple nutrient groups on neurodevelopment in Bangladeshi children. Challenges presented by the analysis included the presence of nonlinear within-group interactions within the effect of the metal mixture, the high-dimensionality for the between-group interaction terms ($d_{N\times E} \geq 9$), and the limited sample size ($n = 351$).

The proposed test identified evidence of interaction between the metal mixture and four nutrient groups, and we observed differences between the CVEK-based results and those from existing approaches for the mineral group. Visualization of bivariate exposure-response surfaces based on nutrient and metal PCs allowed us to visualize the direction of these interactions. The application is important in that identification of nutritional factors that can effectively mitigate the impact of adverse effects of environmental exposures can inform recommendations for pregnant women to improve the health of children across the lifespan.

An important extension of the proposed method would be to incorporate variable selection methods (e.g., shrinkage estimators) to further improve the effectiveness of the proposed approach in the high-dimensional settings. As shown in the numerical studies presented in Section 4, for complex and non-smooth data-generation functions (e.g., k_{true} = Matérn 3/2 or $\sigma = 0.5$ in Figure 1), the proposed method may have weak power even in moderate data dimension. Consequently, it is desirable to identify methods that prune out the effect of irrelevant main effect and nuisance interactions during estimation and inference, rather than fitting all possible nuisance terms. To this end, one particular interesting direction is to combine the sparseness-inducing penalties (e.g., L_1 penalty on the nuisance-effect functions h_1, h_2, h_3, h_{13}, etc in (8)) with a post-selection inference procedure that handles high-dimensional nuisance parameters. One possible avenue to pursue is the de-correlated score test, where the test statistic is made to be orthogonal to the score statistic of the high-dimensional nuisance parameters (32).

A second important extension is to improve the method's robustness to the non-normality of the residual distributions. Although the kernel ensemble approach is designed to be robust against the mis-specification of the mean functions, the inference procedure employed in this work is consistent with that of the classic variance component test. Under heavier-tailed distributions, the variance component test is known to overestimate the spread of its null distribution, leading to an overly conservative test with weak power (38). To verify this empirically, we conducted a simulation study in which the residual follow a t-distribution with degrees of freedom equal to either 5 and 10. (Supplementary Figure C.3–C.4). These results showed that the test is able to maintain Type I error, but has extremely weak power. Consequently, it is of great practical interest to explore combining kernel ensemble methods

with a hypothesis testing procedure based on flexible distributional assumptions, thereby improving the method's performance in non-Gaussian scenarios.

Finally, the ensemble weights $\{u_d\}_{d=1}^{D}$ (see (10)) in CVEK were estimated to maximize the estimator's cross-validation stability. The optimality of such method in terms of the power of the hypothesis test has not been fully investigated. It is desirable to develop an optimal estimation procedure for the ensemble weights $\{u_d\}_{d=1}^{D}$ that maximizes the power of the hypothesis test, in a manner similar to (14). Given such a procedure, it is also of theoretical interest to compare the difference between the ensemble weights generated by maximizing cross-validation stability to those generated by maximizing the power of the test in both finite samples and asymptotically.

Algorithm 1 Cross Validated Ensemble of Kernels (CVEK)

procedure CVEK

Input: A library of kernels $\{k_d\}_{d=1}^{D}$, Data $(\mathbf{y}, \mathbf{X}, \mathbf{x})$

Output: Ensemble Kernel Matrix $\mathbf{K}$

Stage 1: Estimate λ and CV error for each kernel

for $d = 1$ to D **do**

$\mathbf{K}_d = \mathbf{K}_d / tr(\mathbf{K}_d)$

$\hat{\lambda}_d = argmin\, \mathrm{CV}(\lambda | \mathbf{K}_d)$

$\hat{\epsilon}_d = \mathrm{CV}\left(\hat{\lambda}_d \middle| \mathbf{K}_d\right)$

end for

Stage 2: Estimate ensemble weights $\mathbf{u}_{D\times 1} = \{u_1, \ldots, u_D\}$

$$\mathbf{u} = \underset{\mathbf{u} \in \Delta}{argmin} \left\| \sum_{d=1}^{D} u_d \hat{\epsilon}_d \right\|^2 \quad \text{where } \Delta = \left\{ \mathbf{u} \middle| \mathbf{u} \geq 0, \|\mathbf{u}\|_2^2 = 1 \right\}$$

Stage 3: Assemble the ensemble kernel matrix $\mathbf{K}_{ens}$

$$\mathbf{A} = \sum_{d=1}^{D} \hat{\mu}_d \mathbf{A}_{\hat{\lambda}_d, k_d}$$

$\mathbf{U}_A, \boldsymbol{\delta}_A = \mathrm{spectral\_decomp}(\mathbf{A})$

$$\lambda_{\mathbf{K}} = min\left(1, \left(\sum_{k=1}^{n} \frac{\delta_{A,k}}{1-\delta_{A,k}}\right)^{-1}, min\left(\{\hat{\lambda}_d\}_{d=1}^{D}\right)\right)$$

$$\mathbf{K} = \lambda_{\mathbf{K}} * \mathbf{U}_A\, diag\left(\frac{\delta_{A,k}}{1-\delta_{A,k}}\right) \mathbf{U}_A^T$$

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants ES007142, ES016454, ES000002, ES014930, ES013744, ES017437, ES015533, ES022585 from the National Institutes of Health.

References

[1]. Bach Francis. Breaking the Curse of Dimensionality with Convex Neural Networks. Journal of Machine Learning Research, 18(19):1–53, 2017.

[2]. Black Maureen M., Baqui Abdullah H., Zaman K, Persson Lars Ake, Arifeen Shams El, Le Katherine, McNary Scot W., Parveen Monowara, Hamadani Jena D., and Black Robert E. Iron and zinc supplementation promote motor development and exploratory behavior among Bangladeshi infants. The American Journal of Clinical Nutrition, 80(4):903–910, October 2004. [PubMed: 15447897]

[3]. Bodenham Dean A. and Adams Niall M. A comparison of efficient approximations for a weighted sum of chi-squared random variables. Statistics and Computing, 26(4):917–928, July 2016.

[4]. Broadaway K. Alaine, Duncan Richard, Conneely Karen N., Almli Lynn M., Bradley Bekh, Ressler Kerry J., and Epstein Michael P. Kernel Approach for Modeling Interaction Effects in Genetic Association Studies of Complex Quantitative Traits. Genetic epidemiology, 39(5):366–375, July 2015. [PubMed: 25885490]

[5]. Burges Christopher J. C. Advances in Kernel Methods. pages 89–116. MIT Press, Cambridge, MA, USA, 1999.

[6]. Bůžková Petra, Lumley Thomas, and Rice Kenneth. Permutation and parametric bootstrap tests for gene–gene and gene–environment interactions. Annals of human genetics, 75(1):36–45, January 2011. [PubMed: 20384625]

[7]. Cai Tianxi, Tonini Giulia, and Lin Xihong. Kernel Machine Approach to Testing the Significance of Multiple Genetic Markers for Risk Prediction. Biometrics, 67(3):975–986, September 2011. [PubMed: 21281275]

[8]. Cornelis Marilyn C., Tchetgen Eric J. Tchetgen, Liang Liming, Qi Lu, Chatterjee Nilanjan, Hu Frank B., and Kraft Peter. Gene-Environment Interactions in Genome-Wide Association Studies: A Comparative Study of Tests Applied to Empirical Studies of Type 2 Diabetes. American Journal of Epidemiology, 175(3):191–202, February 2012. [PubMed: 22199026]

[9]. Cortes Corinna, Mohri Mehryar, and Rostamizadeh Afshin. Ensembles of Kernel Predictors. In Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11), pages 145–152, Corvallis, Oregon, 2011. AUAI Press.

[10]. Cortes Corinna, Mohri Mehryar, and Rostamizadeh Afshin. Algorithms for Learning Kernels Based on Centered Alignment. Journal of Machine Learning Research, 13(Mar):795–828, 2012.

[11]. Elisseeff A and Pontil M. Leave-one-out Error and Stability of Learning Algorithms with Applications. In Suykens J, Horvath G, Basu S, Micchelli C, and Vandewalle J, editors, Learning Theory and Practice. IOS Press, 2002.

[12]. Ge Tian, Nichols Thomas E., Ghosh Debashis, Mormino Elizabeth C., Smoller Jordan W., and Sabuncu Mert R. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. NeuroImage, 109:505–514, April 2015. [PubMed: 25600633]

[13]. Gleason Kelsey, Shine James P., Shobnam Nadia, Rokoff Lisa B., Suchanda Hafiza Sultana, Hasan Md Ibne, Mostofa Md, Amarasiriwardena Chitra, Quamruzzaman Quazi, Rahman Mahmuder, Kile Molly, Bellinger David, Christiani David, Wright Robert O., and Mazumdar Maitreyi. Contaminated Turmeric Is a Potential Source of Lead Exposure for Children in Rural Bangladesh. Journal of Environmental and Public Health, 2014, August 2014.

[14]. Gretton Arthur, Sejdinovic Dino, Strathmann Heiko, Balakrishnan Sivaraman, Pontil Massimiliano, Fukumizu Kenji, and Sriperumbudur Bharath K. Optimal kernel choice for large-scale two-sample tests. In Pereira F, Burges CJC, Bottou L, and Weinberger KQ, editors, Advances in Neural Information Processing Systems 25, pages 1205–1213. Curran Associates, Inc., 2012.

[15]. Gu Chong. Smoothing Spline ANOVA Models. Springer Science & Business Media, January 2013.

[16]. Gu Chong and Wahba Grace. Smoothing Spline ANOVA with Component-Wise Bayesian "Confidence Intervals". Journal of Computational and Graphical Statistics, 2(1):97–117, March 1993. Publisher: Taylor & Francis.

[17]. Harville David A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistical Association, 72(358):320–338, 1977.

[18]. He Zihuai, Zhang Min, Lee Seunggeun, Smith Jennifer A., Kardia Sharon L. R., Roux Ana V. Diez, and Mukherjee Bhramar. Set-Based Tests for Gene-Environment Interaction in Longitudinal Studies. Journal of the American Statistical Association, 0(ja):0–0, December 2016.

[19]. Hornik Kurt. Approximation capabilities of multilayer feedforward networks. Neural Networks, 4(2):251–257, 1991.

[20]. Huang Jianhua Z. Functional ANOVA Models for Generalized Regression. Journal of Multivariate Analysis, 67(1):49–71, October 1998.

[21]. Kile Molly L., Rodrigues Ema G., Mazumdar Maitreyi, Dobson Christine B., Diao Nancy, Golam Mostofa, Quamruzzaman Quazi, Rahman Mahmudar, and Christiani David C. A prospective cohort study of the association between drinking water arsenic exposure and self-reported maternal health symptoms during pregnancy in Bangladesh. Environmental Health, 13:29, April 2014. [PubMed: 24735908]

[22]. Kimeldorf George S. and Wahba Grace. A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. The Annals of Mathematical Statistics, 41(2):495–502, April 1970.

[23]. Lecué Guillaume and Mitchell Charles. Oracle inequalities for cross-validation type procedures. Electronic Journal of Statistics, 6:1803–1837, 2012.

[24]. Li Shaoyu and Cui Yuehua. Gene-centric gene–gene interaction: A model-based kernel machine method. The Annals of Applied Statistics, 6(3):1134–1161, September 2012.

[25]. Lin Pi-I. D., Bromage Sabri, Mostofa Md. Golam, Allen Joseph, Oken Emily, Kile Molly L., and Christiani David C. Associations between Diet and Toenail Arsenic Concentration among Pregnant Women in Bangladesh: A Prospective Study. Nutrients, 9(4), April 2017.

[26]. Lin Pi-I. D., Bromage Sabri, Mostofa Md Golam, Allen Joseph, Oken Emily, Kile Molly L., and Christiani David C. Validation of a Dish-Based Semiquantitative Food Questionnaire in Rural Bangladesh. Nutrients, 9(1), January 2017.

[27]. Lin Xihong. Variance component testing in generalised linear models with random effects. Biometrika, 84(2):309–326, June 1997.

[28]. Lin Xinyi, Lee Seunggeun, Wu Michael C., Wang Chaolong, Chen Han, Li Zilin, and Lin Xihong. Test for rare variants by environment interactions in sequencing association studies. Biometrics, 72(1):156–164, March 2016. [PubMed: 26229047]

[29]. Liu Dawei, Lin Xihong, and Ghosh Debashis. Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models. Biometrics, 63(4):1079–1088, December 2007. [PubMed: 18078480]

[30]. Maity Arnab and Lin Xihong. Powerful tests for detecting a gene effect in the presence of possible gene-gene interactions using garrote kernel machines. Biometrics, 67(4):1271–1284, December 2011. [PubMed: 21504419]

[31]. Manuck Stephen B. and McCaffery Jeanne M. Gene-environment interaction. Annual Review of Psychology, 65:41–70, 2014.

[32]. Ning Yang and Liu Han. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. Annals of Statistics, 45(1):158–195, February 2017. Publisher: Institute of Mathematical Statistics.

[33]. Raessler M The Arsenic Contamination of Drinking and Groundwaters in Bangladesh: Featuring Biogeochemical Aspects and Implications on Public Health. Archives of environmental contamination and toxicology, 75(1):1–7, July 2018. [PubMed: 29520432]

[34]. Rasmussen Carl Edward and Williams Christopher K. I. Gaussian Processes for Machine Learning. University Press Group Limited, January 2006.

[35]. Raven John, Raven JC, and Court John Hugh. Manual for Raven's progressive matrices and vocabulary scales. Oxford: Oxford Psychologists, 1998 ed edition, 1998.

[36]. Bernhard Schölkopf J. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, January 2002.

[37]. Seoane José A., Day Ian N. M., Gaunt Tom R., and Campbell Colin. A pathway-based data integration framework for prediction of disease progression. Bioinformatics, 30(6):838–845, March 2014. [PubMed: 24162466]

[38]. Staudenmayer J, Lake EE, and Wand MP Robustness for general design mixed models using the t-distribution:. Statistical Modelling, October 2009.

[39]. Stone Charles J. The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation. Annals of Statistics, 22(1):118–171, March 1994. Publisher: Institute of Mathematical Statistics.

[40]. Tchetgen Eric J. Tchetgen and Kraft Peter. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. Epidemiology (Cambridge, Mass.), 22(2):257–261, March 2011.

[41]. Tsybakov Alexandre B. Optimal Rates of Aggregation. In Learning Theory and Kernel Machines, Lecture Notes in Computer Science, pages 303–313. Springer, Berlin, Heidelberg, 2003.

[42]. UNICEF. 2016 Global Nutrition Report. Technical report, June 2016.

[43]. Valeri Linda, Mazumdar Maitreyi M., Bobb Jennifer F., Henn Birgit Claus, Rodrigues Ema, Sharif Omar I. A., Kile Molly L., Quamruzzaman Quazi, Afroz Sakila, Golam Mostafa, Amarasiriwardena Citra, Bellinger David C., Christiani David C., Coull Brent A., and Wright Robert O. The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20–40 Months of Age: Evidence from Rural Bangladesh. Environmental Health Perspectives, 125(6):067015, June 2017. [PubMed: 28669934]

[44]. van der Laan Mark and Dudoit Sandrine. Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples. U.C. Berkeley Division of Biostatistics Working Paper Series, November 2003.

[45]. Wahba Grace. Spline Models for Observational Data. SIAM, September 1990.

[46]. Wu Michael C., Maity Arnab, Lee Seunggeun, Simmons Elizabeth M., Harmon Quaker E., Lin Xinyi, Engel Stephanie M., Molldrem Jeffrey J., and Armistead Paul M. Kernel Machine SNP-set Testing under Multiple Candidate Kernels. Genetic epidemiology, 37(3):267–275, April 2013. [PubMed: 23471868]

[47]. Wu Michael C., Lee Seunggeun, Cai Tianxi, Li Yun, Boehnke Michael, and Lin Xihong. Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. American Journal of Human Genetics, 89(1):82–93, July 2011. [PubMed: 21737059]

[48]. Zhang Daowen and Lin Xihong. Hypothesis testing in semiparametric additive mixed models. Biostatistics (Oxford, England), 4(1):57–74, January 2003.

[49]. Zhao Ni, Chen Jun, Carroll Ian M., Ringel-Kulka Tamar, Epstein Michael P., Zhou Hua, Zhou Jin J., Ringel Yehuda, Li Hongzhe, and Wu Michael C. Testing in Microbiome-Profiling Studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test. American Journal of Human Genetics, 96(5):797–807, May 2015. [PubMed: 25957468]

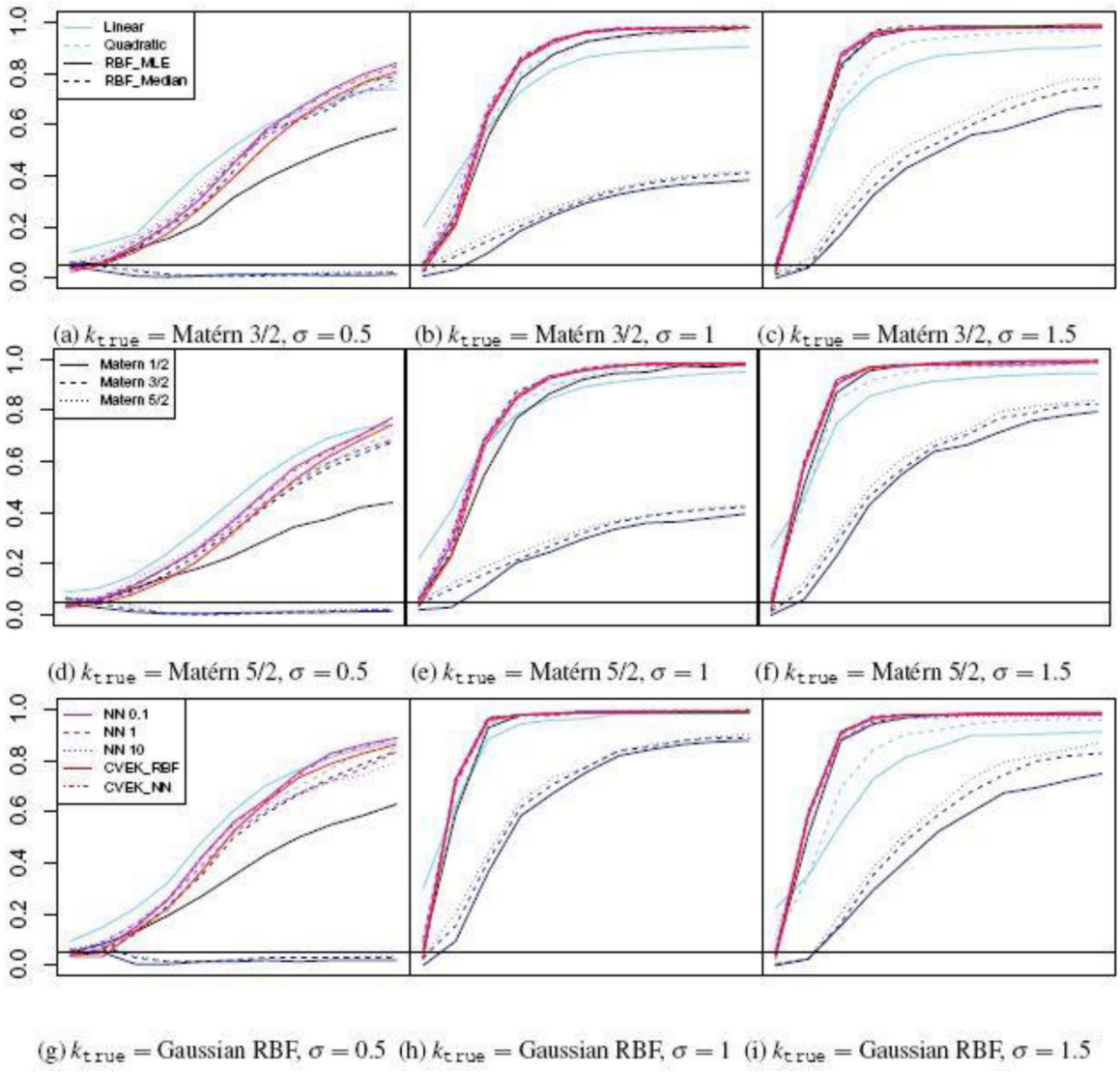
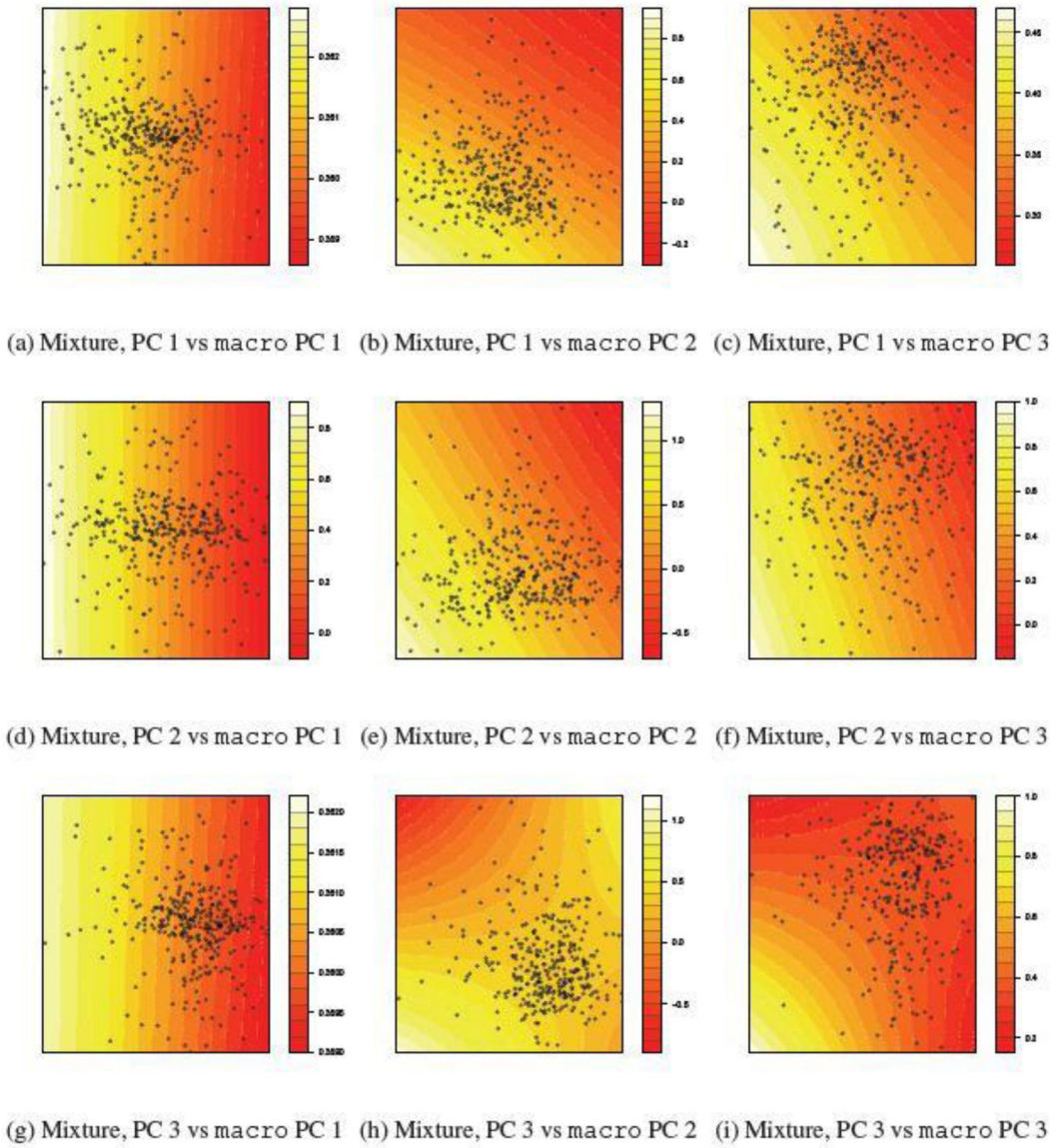


Fig. 1.

Estimated $\hat{P}(p < 0.05)$ (y-axis) as a function of Interaction Strength $\delta \in [0,1]$ (x-axis) for $n = 200$ and $p_1 = p_2 = 3$. **Sky Blue**: Linear (Solid) and Quadratic (Dashed) Kernels, **Black**: RBF-Median (Solid) and RBF-MLE (Dashed), **Dark Blue**: Matérn Kernels with $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, **Purple**: Neural Network Kernels with $\sigma = 0.1,1,10$, **Red**: CVEK based on RBF (Solid) and Neural Networks (Dashed). Horizontal line marks the test's significance level (0.05). When $\delta = 0$, $\hat{P}$ should be below this line.



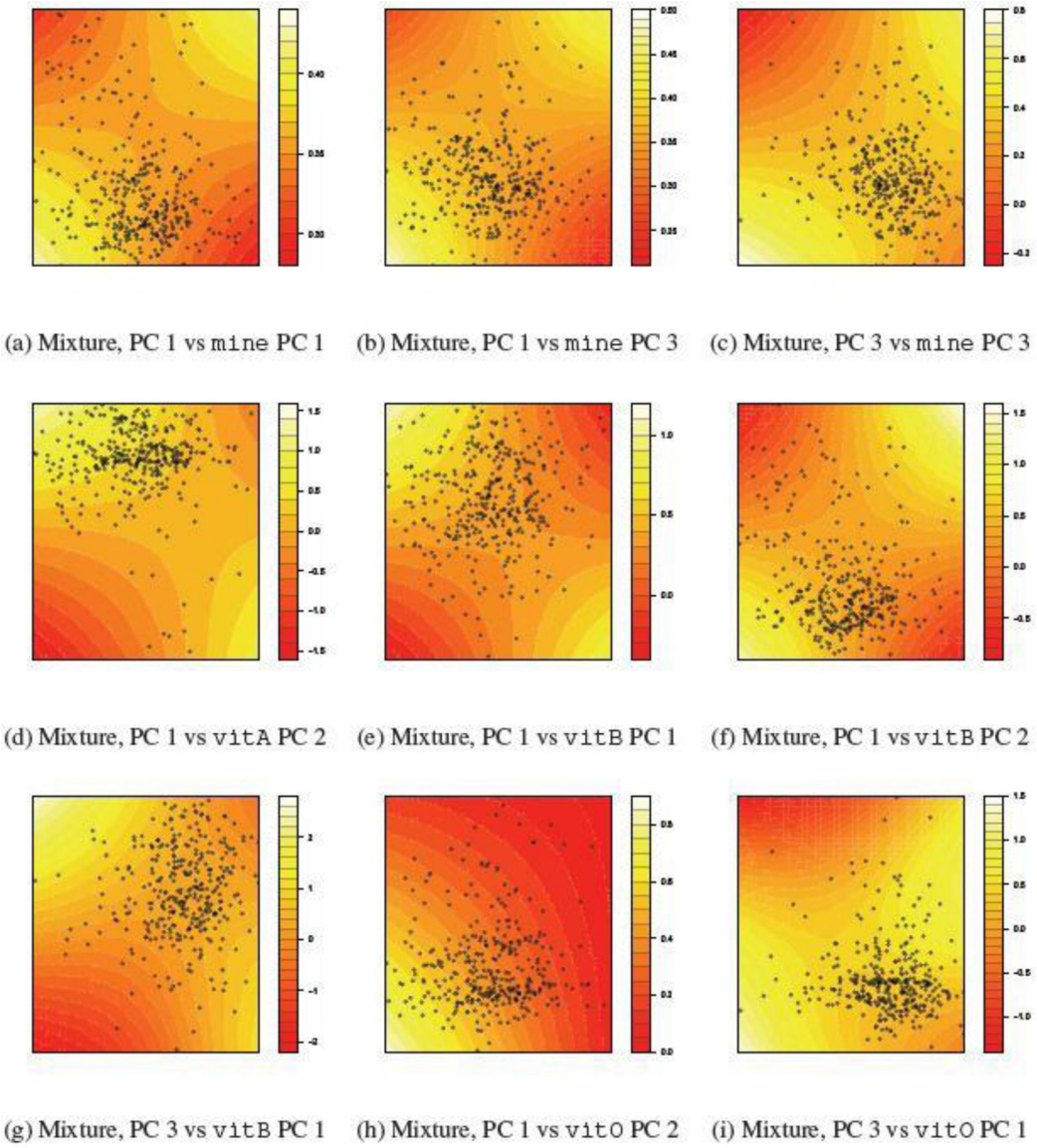
(a) Mixture, PC 1 vs macro PC 1 (b) Mixture, PC 1 vs macro PC 2 (c) Mixture, PC 1 vs macro PC 3

(d) Mixture, PC 2 vs macro PC 1 (e) Mixture, PC 2 vs macro PC 2 (f) Mixture, PC 2 vs macro PC 3

(g) Mixture, PC 3 vs macro PC 1 (h) Mixture, PC 3 vs macro PC 2 (i) Mixture, PC 3 vs macro PC 3

Fig. 2.

Interaction between joint mixture and macronutrient by principal components The top 3 PCs for pollutants accounts for 42.60%, 37.34%, 20.05% of total variation, The top 3 PCs for macro accounts for 63.54%, 28.46% and 7.36% of total variation.



(a) Mixture, PC 1 vs mine PC 1 (b) Mixture, PC 1 vs mine PC 3 (c) Mixture, PC 3 vs mine PC 3

(d) Mixture, PC 1 vs vitA PC 2 (e) Mixture, PC 1 vs vitB PC 1 (f) Mixture, PC 1 vs vitB PC 2

(g) Mixture, PC 3 vs vitB PC 1 (h) Mixture, PC 1 vs vitO PC 2 (i) Mixture, PC 3 vs vitO PC 1

Fig. 3.

Interactions between joint mixture and selected principal components in other four nutrition groups (i.e. Mineral, Vitamin A, Vitamin B and Other Vitamins).

Table 1

List of k_{model}'s considered in the numeric study

Kernel Family	Kernel Function	Model Name	Parameter Value
Polynomial	$\left(1+\mathbf{x}^T\mathbf{x}'\right)^d$	Linear	$d=1$
		Quadratic	$d=2$
Gaussian RBF	$exp\left(-\|\mathbf{x}\text{-}\mathbf{x}'\|^2/\sigma^2\right)$	RBF-MLE	σ = argmax(ML(σ))
		RBF-Median	σ = median($\{\|\mathbf{x}_i-\mathbf{x}_j\|\}_{i\neq j}$)
Matérn	$\frac{2^{1-\nu}}{\Gamma(\upsilon)}\left(\sqrt{2\upsilon}\|\mathbf{r}\|\right)^{\upsilon}K_{\upsilon}\left(\sqrt{2\upsilon}\sigma\|\mathbf{r}\|\right)$	Matérn 1/2	$\nu=1/2$
		Matérn 3/2	$\nu=3/2$
		Matérn 5/2	$\nu=5/2$
Neural Network	$\frac{2}{\pi}*sin^{-1}\left(\frac{2\sigma\mathbf{x}^T\mathbf{x}'}{\sqrt{\left(1+2\sigma\mathbf{x}^T\mathbf{x}\right)\left(1+2\sigma\mathbf{x}'^T\mathbf{x}'\right)}}\right)$	NN 0.1	$\sigma=0.1$
		NN 1	$\sigma=1$
		NN 10	$\sigma=10$
CVEK	$\mathrm{K}=\lambda_{\mathbf{K}}*\left[\mathbf{U}_A diag\left(\frac{\delta_{A,k}}{1-\delta_{A,k}}\right)\mathbf{U}_A^T\right]$	CVEK-RBF	$log(\sigma)\in\{-2,-1,0,1,2\}$
		CVEK-NN	$\sigma\in\{0.1,1,10,50\}$

Table 2

$p-value$ for Nutrient - Environment interaction test with joint As, Pb, Mn mixture

Model	Nutrient Group				
	macro	**mineral**	**vitamin A**	**vitamin B**	**vitamin, other**
CVEK-NN	0.1442	0.0456	0.0135	0.0672	0.0315
CVEK-RBF	0.1257	0.0270	0.0124	0.0541	0.0288
iSKAT	0.2530	0.0905	0.0442	0.1299	0.0459
GKM	0.2081	0.0707	0.0297	0.1075	0.0391
GE-spline	0.1167	0.2080	0.0745	0.2562	0.2133