Estimating Shapley Effects in Big-Data Emulation and Regression Settings using Bayesian Additive Regression Trees

University of California Davis, Indiana University Bloomington

Supplementary Material

S1 Table summarizing properties of various metamodels under nonparametric regression.

See Table 1.

	Consistency established?	Adapt to discontinuities in regression function?	UQ	Tractable to fit model for $p = 250$?	Analytical expression for Shapley effects or Sobol´ indices?	Available code to estimate Shapley effects?
BART	yes (Jeong and Rockova, 2023)	yes (Jeong and Rockova, 2023)	Bayesian	yes (Section 5)	yes (Horiguchi et al., 2021)	this paper (Pratola, 2023)
GP	yes	yes (Mohammadi et al., 2019)	Bayesian	no	yes for some covariance kernels	sensitivity R package (Iooss and Prieur, 2019)
PCE	no	no	bootstrap	no	yes (Sudret, 2008)	no
BMARS	no	no	Bayesian	yes (Francom et al., 2018)	yes (Francom et al., 2018)	no

Table 1: Properties of various metamodels under nonparametric regression.

S2 Review of posterior contraction

A posterior contraction rate quantifies how quickly a posterior distribution approaches the true parameter of the data's distribution. We use a simplified version of the definition from Ghosal and Van der Vaart (2017): for every $n \in \mathbb{N}$, let $X^{(n)}$ be an observation in a sample space $(\mathfrak{X}^{(n)}, \mathscr{X}^{(n)})$ with distribution $P_\theta^{(n)}$ indexed by θ belonging to a first countable topological space Θ. Given a prior Π_n on the Borel sets of Θ, let $\Pi_n(\cdot \mid X^{(n)})$ be (a fixed particular version of) the posterior distribution.

Definition 1 (Posterior contraction rate)**.** A sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ is a posterior contraction rate at the parameter θ_0 with respect to the semimetric d if $\Pi_n(\theta\colon d(\theta,\theta_0) \geq M_n\varepsilon_n \mid X^{(n)}) \to 0$ in $P_{\theta_0}^{(n)}$-probability, for every $M_n \to \infty$.

If there exists a constant $M > 0$ such that $\Pi_n(\theta\colon d(\theta,\theta_0) \geq M\varepsilon_n \mid X^{(n)}) \to 0$ in $P_{\theta_0}^{(n)}$-probability, then the sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfies the definition of posterior contraction rate. This will be relevant in interpreting Corollaries 3 and 4 in Section 3.

After reviewing the concept of contraction rates, we state for convenience the conditions made in the theorems of Jeong and Rockova (2023) that our contraction-rate results rely on. Because these conditions are not the focus of this paper, we leave discussion of the context behind these conditions to Jeong and Rockova (2023).

S3 Preliminaries for proofs

For any positive integer m, denote $[m] \coloneqq \{1,\dots,m\}$.

From the main text, we copy here the considered regression models with fixed and random design. The regression model with *fixed* design is

$$Y_i = f_0(\mathbf{x}_i) + \varepsilon_i, \qquad \varepsilon_i \sim N(0,\sigma_0^2), \qquad i = 1,\dots,n, \tag{S3.1}$$

where $\sigma_0^2 < \infty$ and each covariate $\mathbf{x}_i \in [0,1]^p$ is fixed. The regression model with *random* design is

$$Y_i = f_0(\mathbf{X}_i) + \varepsilon_i, \qquad \mathbf{X}_i \sim \pi, \qquad \varepsilon_i \sim N(0, \sigma_0^2), \qquad i = 1, \dots, n, \quad \text{(S3.2)}$$

where $\sigma_0^2 < \infty$, each $\mathbf{X}_i \in [0,1]^p$ is a p-dimensional random covariate, and π is a probability measure such that $\text{supp}(\pi) \subseteq [0,1]^p$.

Piecewise heterogeneous anisotropic functions Next we introduce the conditions of the theorems of Jeong and Rockova (2023) relevant to our work. The first set of conditions involves what values of f_0 and σ_0^2 are allowed for BART to contract around f_0. A common assumption for f_0 is isotropic smoothness, but this excludes the realistic scenario that f_0 is discontinuous and has different degrees of smoothness in different directions and regions. Jeong and Rockova (2023) introduce a new class of *piecewise heterogeneous anisotropic* functions whose domain is partitioned into many boxes (i.e. hyperrectangles), each of which has its own anisotropic smoothness with the same harmonic mean. First assume f_0 is d-sparse, i.e. there exists a function $h_0 \colon [0,1]^d \to \mathbb{R}$ and a subset $S_0 \subseteq [p]$ with $|S_0| = d$ such that $f_0(\mathbf{x}) = h_0(\mathbf{x}_{S_0})$ for any $\mathbf{x} \in [0,1]^p$. For any given box $\Xi \subseteq [0,1]^d$, smoothness parameter $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^T \in (0,1]^d$, and Hölder coefficient $\lambda < \infty$, an *anisotropic* $\boldsymbol{\alpha}$*-Hölder space* on Ξ is defined as

$$\mathcal{H}_\lambda^{\boldsymbol{\alpha},d}(\Xi) \coloneqq \Big\{ h \colon \Xi \to \mathbb{R}; |h(x) - h(y)| \le \lambda \sum_{j=1}^{d} |x_j - y_j|^{\alpha_j}, x, y \in \Xi \Big\}.$$

Though h_0 might have different anisotropic smoothness on different boxes, it is important to assume that all boxes have the same harmonic mean. Thus define the set $\mathcal{A}_{\bar{\alpha}}^{R,d}$ to be the set of R-tuples of smoothness parameters that have harmonic mean $\bar{\alpha} \in (0,1]$:

$$\mathcal{A}_{\bar{\alpha}}^{R,d} \coloneqq \Big\{(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_R) : \boldsymbol{\alpha}_r \in (0,1]^d, \bar{\alpha}^{-1} = p^{-1} \sum_{j=1}^{d} \alpha_{rj}^{-1}, r \in [R]\Big\}.$$

Given a partition $(\Xi_1, \ldots, \Xi_R)$ of $[0,1]^d$ with boxes $\Xi_r \subseteq [0,1]^d$ and a smoothness R-tuple $A_{\bar{\alpha}} \in \mathcal{A}_{\bar{\alpha}}^{R,d}$ for some $\bar{\alpha} \in (0,1]$, define a *piecewise heterogeneous anisotropic Hölder space* as

$$\mathcal{H}_{\lambda}^{A_{\bar{\alpha}},d}(\mathfrak{X}) \coloneqq \Big\{h \colon [0,1]^d \to \mathbb{R}; h|_{\Xi_r} \in \mathcal{H}_{\lambda}^{\boldsymbol{\alpha}_r,d}(\Xi_r), r \in [R]\Big\}.$$

To extend a function from a sparse domain to the original domain $[0,1]^p$, for any nonempty subset $S \subseteq [p]$ define $W_S^p \colon \mathcal{C}(\mathbb{R}^{|S|}) \to \mathcal{C}(\mathbb{R}^p)$ as the map that extends $h \in \mathcal{C}(\mathbb{R}^{|S|})$ to the function $W_S^p h \colon \mathbf{x} \to h(\mathbf{x}_S)$ where $\mathbf{x} \in [0,1]^p$ and $\mathcal{C}(E)$ denotes the class of real-valued continuous functions defined on a Euclidean subspace E. With this definition, the space $\mathcal{H}_{\lambda}^{A_{\bar{\alpha}},d}(\mathfrak{X})$ from the preceding panel can be extended to the corresponding *d-sparse piecewise heterogeneous anisotropic Hölder space*

$$\Gamma_{\lambda}^{A_{\bar{\alpha}},d,p}(\mathfrak{X}) \coloneqq \bigcup_{S \subseteq [p] \colon |S| = d} W_S^p\Big(\mathcal{H}_{\lambda}^{A_{\bar{\alpha}},d}(\mathfrak{X})\Big).$$

With these definitions, we can now state the needed assumptions on the true f_0 and σ^2.

(A1) For $d > 0$, $\lambda > 0$, $R > 0$, $\mathfrak{X} = (\Xi_1, \ldots, \Xi_R)$, and $A_{\bar{\alpha}} \in \mathcal{A}_{\bar{\alpha}}^{R,d}$ with $\bar{\alpha} \in (0, 1]$, the true function satisfies $f_0 \in \Gamma_\lambda^{A_{\bar{\alpha}},d,p}(\mathfrak{X})$ or $f_0 \in \Gamma_\lambda^{A_{\bar{\alpha}},d,p}(\mathfrak{X}) \cap \mathcal{C}([0,1]^p)$.

(A2) It is assumed that d, p, λ, R, and $\bar{\alpha}$ satisfy $\epsilon_n \ll 1$, where

$$\epsilon_n \coloneqq \sqrt{\frac{d \log p}{n}} + (\lambda d)^{d/(2\bar{\alpha}+d)} \left(\frac{R \log n}{n}\right)^{\bar{\alpha}/(2\bar{\alpha}+d)}. \tag{S3.3}$$

(A3) The true function f_0 satisfies $\|f_0\|_\infty \lesssim \sqrt{\log n}$.

(A4) The true variance parameter satisfies $\sigma^2 \in [C_0^{-1}, C_0]$ for some sufficiently large $C_0 > 1$.

Split-net The second set of conditions (of the theorems of Jeong and Rockova (2023) relevant to our work) involves the split values c allowed in the binary split rules "$x_j < c$" of the regression trees. If a partition of $[0,1]^p$ can be created using the aforementioned tree-based procedure, call it a *flexible tree partition*. To restrict a flexible tree partition by a set of allowable split values in the binary split rules, for any integer b_n define a *split-net* $\mathcal{Z}$ to be a finite set of points in $[0,1]^p$ at which possible splits occur along coordinates. That is, the allowable split values for any input dimension $j \in [p]$ are the jth components of the points in the split-net. For a given split-net $\mathcal{Z}$, a flexible tree partition $(\Omega_1, \ldots, \Omega_K)$ of $[0,1]^p$ with boxes $\Omega_k \subseteq [0,1]^p$, $k \in [K]$, is called a *$\mathcal{Z}$-tree partition* if every split occurs at points in $\mathcal{Z}$.

A split net should be dense enough for a resulting partition to be close enough to the underlying partition $\mathfrak{X}^* = (\Xi_1^*, \ldots, \Xi_R^*)$ of the true function f_0. For any two box partitions $\mathfrak{Y}^1 = (\Psi_1^1, \ldots, \Psi_J^1)$ and $\mathfrak{Y}^2 = (\Psi_1^2, \ldots, \Psi_J^2)$ with the same number J of boxes, their closeness will be measured using the Hausdorff-type divergence

$$\Upsilon(\mathfrak{Y}^1, \mathfrak{Y}^2) := \min_{\tau \in \texttt{Perm}[J]} \max_{r \in [J]} \mathrm{Haus}(\Psi_r^1, \Psi_{\tau(r)}^2)$$

where $\texttt{Perm}[J]$ denotes the set of all permutations of $[J]$ and $\mathrm{Haus}(\cdot, \cdot)$ is the Hausdorff distance. For a subset $S \subseteq [p]$, a box partition of $[0,1]^p$ is called *S-chopped* if every box Ψ in the box partition satisfies $\max_{j \in S} \texttt{len}([\Psi]_j) < 1$ and $\min_{j \notin S} \texttt{len}([\Psi]_j) = 1$, where $[\Psi]_j$ denotes the interval created by projecting the box $[\Psi]$ onto the j-th principal axis. For a given subset $S \subseteq [p]$, consider an S-chopped partition $\mathfrak{Y}$ of $[0,1]^p$ with J boxes. For any given $c_n \geq 0$, a split-net $\mathcal{Z}_n$ is said to be *$(\mathfrak{Y}, c_n)$-dense* if there exists an S-chopped $\mathcal{Z}_n$-tree partition $\mathcal{T}_n$ of $[0,1]^p$ with J boxes such that $\Upsilon(\mathfrak{Y}, \mathcal{T}_n) \leq c_n$.

A split net should also be regular enough (defined below) for a tree partition to capture local features of f_0 on each box. Assume the underlying partition $\mathfrak{X}^*$ can be approximated well by an $S(\mathfrak{X}^*)$-chopped $\mathcal{Z}$-tree partition $(\Omega_1^*, \ldots, \Omega_R^*) := \arg\min_{\mathcal{T} \in \mathscr{T}_{S(\mathfrak{X}^*), R, \mathcal{Z}}} \Upsilon(\mathfrak{X}^*, \mathcal{T})$. In each box Ω_r^*, the idea is to allow splits to occur more often along the input dimensions with less smoothness. Given a split-net $\mathcal{Z}$ and splitting coordinate j, define the *midpoint-split* of a box Ψ as the bisection of Ψ along coordinate j at the $\lceil \tilde{b}_j(\mathcal{Z}, \Psi)/2 \rceil$th split-candidate in $[\mathcal{Z}]_j \cap \mathtt{int}([\Psi]_j)$, where $\tilde{b}_j(\mathcal{Z}, \Psi)$ is the cardinality of $[\mathcal{Z}]_j \cap \mathtt{int}([\Psi]_j)$. Given a smoothness vector $\boldsymbol{\alpha} \in (0,1]^d$, box $\Psi \subseteq [0,1]^p$, split-net $\mathcal{Z}$, integer $L > 0$, and index set $S = \{s_1, \ldots, s_d\} \subseteq [p]$, define the *anisotropic k-d tree* $AKD(\Psi; \mathcal{Z}, \boldsymbol{\alpha}, L, S)$ as the iterative splitting procedure that partitions Ψ into disjoint boxes $\Omega_1^\circ, \ldots, \Omega_{2^{L^\circ}}^\circ$ as follows:

1. Set $\Omega_1^\circ = \Psi$ and set counter $l_j = 0$ for each $j \in [d]$.
2. Let $L^\circ = \sum_{j=1}^d l_j$ for the current counters. For splits at iteration $1+L^\circ$, choose $j' = \min\{\arg\min_j l_j \alpha_j\}$. Midpoint-split all boxes $\Omega_1^\circ, \ldots, \Omega_{2^{L^\circ}}^\circ$ with the given $\mathcal{Z}$ and splitting coordinate $s_{j'}$. Relabel the generated new boxes as $\Omega_1^\circ, \ldots, \Omega_{2^{1+L^\circ}}^\circ$, and then increment $l_{j'}$ by one.
3. Repeat step 2 until either the updated L° equals L or the midpoint-split is no longer available. Return counters $l_1, \ldots, l_d$ and boxes $\Omega_1^\circ, \ldots, \Omega_{2^{L^\circ}}^\circ$.

For a given box $\Psi \subseteq [0,1]^p$, smoothness vector $\boldsymbol{\alpha} \in (0,1]^d$, integer $L > 0$, and index set $S = \{s_1, \ldots, s_d\} \subseteq [p]$, a split-net $\mathcal{Z}$ is called $(\Psi, \boldsymbol{\alpha}, L, S)$-*regular* if the counters and boxes returned by $AKD(\Psi; \mathcal{Z}, \boldsymbol{\alpha}, L, S)$ satisfy $L^\circ = L$ and $\max_k \mathtt{len}([\Omega_k^\circ]_{s_j}) \lesssim \mathtt{len}([\Psi]_{s_j}) 2^{-l_j}$ for every $j \in [d]$.

With these definitions, we can now state the needed assumptions on the sequence $\{\mathcal{Z}_n\}_{n=1}^\infty$ of split-nets.

(A5) Each split-net $\mathcal{Z}_n$ satisfies $\max_{1\le j\le p}\log b_j(\mathcal{Z}_n) \lesssim \log n$, where $b_j(\mathcal{Z}_n)$ is the cardinality of the set $\{z_j : (z_1,\dots,z_p)\in\mathcal{Z}_n\}$.

(A6) Each split-net $\mathcal{Z}_n$ is suitably dense and regular to construct a $\mathcal{Z}_n$-tree partition $\hat{\mathcal{T}}$ such that there exists a simple function $\hat{f}_0 \in \mathcal{F}_{\hat{\mathcal{T}}}$ satisfying $\|f_0 - \hat{f}_0\|_n \lesssim \bar{\epsilon}_n$, where

$$\bar{\epsilon}_n \coloneqq (\lambda d)^{d/(2\bar{\alpha}+d)}\{(R\log n)/n\}^{\bar{\alpha}/(2\bar{\alpha}+d)}, \tag{S3.4}$$

the empirical L_2-norm $\|\cdot\|_n$ is defined as $\|f\|_n^2 = n^{-1}\sum_{i=1}^n |f(\mathbf{x}_i)|^2$, and $\mathcal{F}_{\hat{\mathcal{T}}}$ is the set of functions on $[0,1]^p$ that are constant on each piece of the partition $\hat{\mathcal{T}}$.

(A7) Each $\mathcal{Z}_n$-tree partition $(\Omega_1^*,\dots,\Omega_R^*)$ approximating the underlying partition $\mathfrak{X}^*$ for the true function f_0 satisfies $\max_{r\in[R]} \texttt{depth}(\Omega_r^*) \lesssim \log n$, where `depth` means the depth of a node (i.e. number of nodes in the path from that node to the root node).

Finally, we state the required prior specification.

(P1) Each tree partition in the ensemble is independently assigned a tree prior with Dirichlet sparsity from Linero (2018). This sparse Dirichlet prior places a Dirichlet prior on the proportion vector used to select the splitting coordinate j during the creation of a split rule.

(P2) The step-heights of the regression-tree functions are each assigned a normal prior with mean zero and covariance matrix whose eigenvalues are bounded below and above.

(P3) The variance parameter σ^2 is assigned an inverse gamma prior.

Jeong and Rockova (2023) make the above assumptions and prior specification for their contraction-rate results in the fixed design setting (S3.1). For their contraction-rate results in the random design setting (S3.2), a few of the above assumptions and prior specifications are replaced by the following:

(A3∗) The true function f_0 satisfies $\|f_0\|_\infty \leq C_0^*$ for some sufficiently large $C_0^* > 0$.

(A6∗) The split-net $\mathcal{Z}$ is suitably dense and regular to construct a $\mathcal{Z}$-tree partition $\hat{\mathcal{T}}$ such that there exists $\hat{f}_0 \in \mathcal{F}_{\hat{\mathcal{T}}}$ satisfying $\|f_0 - \hat{f}_0\| \lesssim \bar{\epsilon}_n$ where $\bar{\epsilon}_n$ is given by (S3.4).

(P2∗) A prior on the compact support $[-\bar{C}_1, \bar{C}_1]$ is assigned to the step-heights of the regression-tree functions for some $\bar{C}_1 > C_0^*$.

(P3∗) A prior on the compact support $[\bar{C}_2^{-1}, \bar{C}_2]$ is assigned to the variance parameter σ^2 for some $\bar{C}_2 > C_0$.

S4 Posterior asymptotics

This section establishes our contraction-rate results (Corollaries 3 and 4) for estimators of Sobol´ indices and Shapley effects under either the fixed design (S3.1) or the random design (S3.2). Our proofs rely on these sensitivity indices having a property (defined in Lemma 2 below) similar to but slightly less restrictive than Lipschitz continuity. However, the tasks of proving this property for all of these sensitivity indices are very similar to each other. Because these indices are linear combinations of the functional $c_{P,\pi}$ defined in (S5.5), we can use Lemma 2 to reduce the above tasks to the single task of proving this property for $c_{P,\pi}$.

Lemma 1. *Suppose the following relationship is true for all indices k in a finite set $\mathcal{A}$: given two metric spaces X and X_0 with the same metric d_X, there exists a constant $C > 0$ such that, for all $(x, x_0) \in X \times X_0$, the function $\phi_k \colon X \cup X_0 \to \mathbb{R}$ satisfies*

$$|\phi_k(x) - \phi_k(x_0)| \leq C d_X(x, x_0).$$

Then any set $\{a_k\}_{k \in \mathcal{A}}$ of real numbers satisfies

$$\Big| \sum_{k \in \mathcal{A}} a_k \phi_k(x) - \sum_{k \in \mathcal{A}} a_k \phi_k(x_0) \Big| \leq C^* d_X(x, x_0).$$

where $C^* = C \sum_{k \in \mathcal{A}_+} |a_k|$ and $\mathcal{A}_+ \coloneqq \{k \in \mathcal{A} \colon \big|\phi_k(x) - \phi_k(x_0)\big| > 0\}$.

S4.1 Nonparametric regression with random design

This section assumes the random-design regression setting (S3.2); all expectations in this section are with respect to the probability measure π in (S3.2).

Theorem 1. *Assume (A3*). If $f \in L^2([0,1]^p)$ shares the same bound C_0^* from (A3*), then for any subset $P \subseteq [p]$ and distribution π with support $[0,1]^p$ we have*

$$|c_{P,\pi}(f) - c_{P,\pi}(f_0)| \leq 4C_0^* \|f - f_0\|_{2,\pi}$$

for the functional $c_{P,\pi}$ defined in (S5.5).

Corollary 1. *Under the assumptions of Theorem 4 of Jeong and Rockova (2023) – Assumptions (A1), (A2), (A3*), (A4), (A5), (A6*), and (A7), and the prior assigned through (P1), (P2*), and (P3*) – and Theorem 3 above, there exist positive constants $L_{V,\pi,|P|}$, $L_{T,\pi}$, and L_S such that as $n \to \infty$ for ϵ_n in* (S5.6),

$$E_0\Pi\Big\{(f,\sigma^2)\colon |V_{P,\pi}(f) - V_{P,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{V,\pi,|P|}\epsilon_n \Big| Y_1,\dots,Y_n\Big\} \to 0,$$

$$E_0\Pi\Big\{(f,\sigma^2)\colon |T_{j,\pi}(f) - T_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{T,\pi}\epsilon_n \Big| Y_1,\dots,Y_n\Big\} \to 0,$$

$$\text{and } E_0\Pi\Big\{(f,\sigma^2)\colon |S_{j,\pi}(f) - S_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_S\epsilon_n \Big| Y_1,\dots,Y_n\Big\} \to 0.$$

S4.2 Nonparametric regression with fixed design

This section assumes the fixed-design regression setting (S3.1); all expectations in this section are with respect to the probability measure $P_{\mathscr{X}}(\cdot) = n^{-1}\sum_{\mathbf{x}\in\mathscr{X}}\delta_{\mathbf{x}}(\cdot)$ where $\mathscr{X}$ is the set of the fixed covariates assumed in (S3.1).

Theorem 2. *Assume (A3). If $f \in L^2([0,1]^p)$ shares the same bound $\sqrt{\log n}$ from (A3), then for any subset $P \subseteq [p]$ and distribution π with support $[0,1]^p$ we have*

$$|c_{P,P_{\mathscr{X}}}(f) - c_{P,P_{\mathscr{X}}}(f_0)| \lesssim 4\sqrt{\log n}\|f - f_0\|_{2,P_{\mathscr{X}}}$$

where the empirical L_2-norm $\|\cdot\|_{2,P_{\mathscr{X}}}$ is defined as $\|f\|_{2,P_{\mathscr{X}}}^2 = n^{-1}\sum_{\mathbf{x}\in\mathscr{X}}|f(\mathbf{x})|^2$.

Corollary 2. *Under the assumptions of Theorem 2 of Jeong and Rockova (2023) – Assumptions (A1), (A2), (A3), (A4), (A5), (A6), and (A7), and the prior assigned through (P1), (P2), and (P3) – and Theorem 4 above, there exist positive constants $L_{V,\pi,|P|}$, $L_{T,\pi}$, and L_S such that as $n \to \infty$ for ϵ_n in* (S5.6),

$$E_0\Pi\Big\{(f,\sigma^2)\colon |V_{P,\pi}(f) - V_{P,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{V,\pi,|P|}\epsilon_n\sqrt{\log n} \mid Y_1,\dots,Y_n\Big\} \to 0,$$

$$E_0\Pi\Big\{(f,\sigma^2)\colon |T_{j,\pi}(f) - T_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{T,\pi}\epsilon_n\sqrt{\log n} \mid Y_1,\dots,Y_n\Big\} \to 0,$$

$$\textit{and } E_0\Pi\Big\{(f,\sigma^2)\colon |S_{j,\pi}(f) - S_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_S\epsilon_n\sqrt{\log n} \mid Y_1,\dots,Y_n\Big\} \to 0.$$

S5 Proofs of results in main text

For convenience, here we replicate the theorems, lemmas, and relevant equations from the main text.

S5.1 Relevant quantities

$$c_{P,\pi}(f) = \mathrm{Var}_\pi[E_\pi\{f(\mathbf{X}) \mid \mathbf{X}_P\}] = E_\pi([E_\pi\{f(\mathbf{X}) \mid \mathbf{X}_P\}]^2) - [E_\pi\{f(\mathbf{X})\}]^2. \tag{S5.5}$$

$$\epsilon_n := \sqrt{\frac{d\log p}{n}} + (\lambda d)^{d/(2\bar{\alpha}+d)}\left(\frac{R\log n}{n}\right)^{\bar{\alpha}/(2\bar{\alpha}+d)}. \tag{S5.6}$$

$$S_{j,\pi}(f) = (p!)^{-1} \sum_{P\subseteq([p]\setminus\{j\})} (p-|P|-1)!\,|P|!\left\{c_{P\cup\{j\},\pi}(f) - c_{P,\pi}(f)\right\}. \tag{S5.7}$$

S5.2 Theorems and lemmas

Lemma 2. *Suppose the following relationship is true for all indices k in a finite set $\mathcal{A}$: given two metric spaces X and X_0 with the same metric d_X, there exists a constant $C > 0$ such that, for all $(x, x_0) \in X \times X_0$, the function $\phi_k \colon X \cup X_0 \to \mathbb{R}$ satisfies*

$$|\phi_k(x) - \phi_k(x_0)| \leq C d_X(x, x_0).$$

Then any set $\{a_k\}_{k\in\mathcal{A}}$ *of real numbers satisfies*

$$\Big|\sum_{k\in\mathcal{A}} a_k\phi_k(x) - \sum_{k\in\mathcal{A}} a_k\phi_k(x_0)\Big| \leq C^* d_X(x, x_0).$$

where $C^* = C\sum_{k\in\mathcal{A}_+}|a_k|$ *and* $\mathcal{A}_+ := \{k \in \mathcal{A}\colon \big|\phi_k(x) - \phi_k(x_0)\big| > 0\}$.

Theorem 3. *Assume (A3*). If* $f \in L^2([0,1]^p)$ *shares the same bound* C_0^* *from (A3*), then for any subset* $P \subseteq [p]$ *and distribution* π *with support* $[0,1]^p$ *we have*

$$|c_{P,\pi}(f) - c_{P,\pi}(f_0)| \leq 4C_0^*\|f - f_0\|_{2,\pi}$$

for the functional $c_{P,\pi}$ *defined in* (S5.5).

Corollary 3. *Under the assumptions of Theorem 4 of Jeong and Rockova (2023) – Assumptions (A1), (A2), (A3∗), (A4), (A5), (A6∗), and (A7), and the prior assigned through (P1), (P2∗), and (P3∗) – and Theorem 3 above, there exist positive constants* $L_{V,\pi,|P|}$, $L_{T,\pi}$, *and* L_S *such that as* $n \to \infty$ *for* ϵ_n *in* (S5.6),

$$E_0\Pi\Big\{(f,\sigma^2)\colon |V_{P,\pi}(f) - V_{P,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{V,\pi,|P|}\epsilon_n\Big|Y_1,\dots,Y_n\Big\},$$
$$E_0\Pi\Big\{(f,\sigma^2)\colon |T_{j,\pi}(f) - T_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{T,\pi}\epsilon_n\Big|Y_1,\dots,Y_n\Big\},$$
$$\text{and } E_0\Pi\Big\{(f,\sigma^2)\colon |S_{j,\pi}(f) - S_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_S\epsilon_n\Big|Y_1,\dots,Y_n\Big\}$$

each shrink to zero.

Theorem 4. *Assume (A3). If* $f \in L^2([0,1]^p)$ *shares the same bound* $\sqrt{\log n}$ *from (A3), then for any subset* $P \subseteq [p]$ *and distribution* π *with support* $[0,1]^p$ *we have*

$$|c_{P,P_{\mathscr{X}}}(f) - c_{P,P_{\mathscr{X}}}(f_0)| \lesssim 4\sqrt{\log n}\|f - f_0\|_{2,P_{\mathscr{X}}}$$

where the empirical L_2*-norm* $\|\cdot\|_{2,P_{\mathscr{X}}}$ *is defined as* $\|f\|_{2,P_{\mathscr{X}}}^2 = n^{-1}\sum_{\mathbf{x}\in\mathscr{X}} |f(\mathbf{x})|^2$.

Corollary 4. *Under the assumptions of Theorem 2 of Jeong and Rockova (2023) – Assumptions (A1), (A2), (A3), (A4), (A5), (A6), and (A7), and the prior assigned through (P1), (P2), and (P3) – and Theorem 4 above, there exist positive constants* $L_{V,\pi,|P|}$, $L_{T,\pi}$, *and* L_S *such that as* $n \to \infty$ *for* ϵ_n *in* (S5.6),

$$E_0\Pi\Big\{(f,\sigma^2)\colon |V_{P,\pi}(f) - V_{P,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{V,\pi,|P|}\epsilon_n\sqrt{\log n} \mid Y_1,\ldots,Y_n\Big\},$$

$$E_0\Pi\Big\{(f,\sigma^2)\colon |T_{j,\pi}(f) - T_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_{T,\pi}\epsilon_n\sqrt{\log n} \mid Y_1,\ldots,Y_n\Big\},$$

$$\textit{and } E_0\Pi\Big\{(f,\sigma^2)\colon |S_{j,\pi}(f) - S_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > L_S\epsilon_n\sqrt{\log n} \mid Y_1,\ldots,Y_n\Big\}$$

each shrink to zero.

S5.3 Proofs

Proof of Lemma 2. We have

$$\Big|\sum_{k\in\mathcal{A}} a_k\phi_k(x) - \sum_{k\in\mathcal{A}} a_k\phi_k(x_0)\Big| \le \sum_{k\in\mathcal{A}} |a_k|\,\big|\phi_k(x) - \phi_k(x_0)\big| \le \sum_{k\in\mathcal{A}_+} |a_k|\, C d_X(x,x_0)$$

where the right-most sum in the preceding panel is exactly $C^* d_X(x,x_0)$. $\square$

Proof of Theorem 3. Note that

$$\begin{aligned}\Big|c_{P,\pi}(f) - c_{P,\pi}(f_0)\Big| &= \Big|\Big(E\big[(E[f(\mathbf{X}) \mid \mathbf{X}_P])^2 - (E[f_0(\mathbf{X}) \mid \mathbf{X}_P])^2\big]\Big) \\ &\qquad - \Big([Ef(\mathbf{X})]^2 - [Ef_0(\mathbf{X})]^2\Big)\Big| \\ &\leq E\Big|(E[f(\mathbf{X}) \mid \mathbf{X}_P])^2 - (E[f_0(\mathbf{X}) \mid \mathbf{X}_P])^2\Big| \\ &\qquad + \Big|[E\{f(\mathbf{X})\}]^2 - [E\{f_0(\mathbf{X})\}]^2\Big|.\end{aligned}$$

From the assumption that f and f_0 are bounded in supremum norm by C_0^*, we get

$$\begin{aligned}&\Big|[E\{f(\mathbf{X})\}]^2 - [E\{f_0(\mathbf{X})\}]^2\Big| \\ &\quad = \Big|\big[E\{f(\mathbf{X})\} + E\{f_0(\mathbf{X})\}\big]\big[E\{f(\mathbf{X})\} - E\{f_0(\mathbf{X})\}\big]\Big| \\ &\quad \leq 2C_0^* E|f(\mathbf{X}) - f_0(\mathbf{X})|.\end{aligned}$$

We can similarly deduce for any $\mathbf{X}_P$ that

$$\Big|(E[f(\mathbf{X}) \mid \mathbf{X}_P])^2 - (E[f_0(\mathbf{X}) \mid \mathbf{X}_P])^2\Big| \leq 2C_0^*\ E[|f(\mathbf{X}) - f_0(\mathbf{X})| \mid \mathbf{X}_P].$$

Then

$$\begin{aligned}&|c_{P,\pi}(f) - c_{P,\pi}(f_0)| \\ &\leq E\big[2C_0^* E[|f(\mathbf{X}) - f_0(\mathbf{X})| \mid \mathbf{X}_P]\big] + 2C_0^* E|f(\mathbf{X}) - f_0(\mathbf{X})| \\ &= 4C_0^* E|f(\mathbf{X}) - f_0(\mathbf{X})|.\end{aligned}$$

To finish, Jensen's inequality implies $E|f(\mathbf{X}) - f_0(\mathbf{X})| \leq \|f - f_0\|_{2,\pi}$. $\square$

Proof of Corollary 3. Below is the proof just for the jth (where $j \in [p]$) total-effect Sobol´ index. The same argument can be followed to obtain the corresponding results for any main-effect Sobol´ index and any Shapley effect after making the appropriate substitutions for the a_P below. Lemma 2 and Theorem 3 together imply

$$|T_{j,\pi}(f) - T_{j,\pi}(f_0)| \leq D_{T,\pi} \|f - f_0\|_{2,\pi}.$$

where $D_{T,\pi} \leq \max\{1, 4C_0^* \sum_{P \in [p]} |a_{P,\pi}|\}$ and the real values a_P are the coefficients corresponding to $T_{j,\pi}$ expressed as a linear combination of $c_{P,\pi}$. (Theorem 5 provides upper bounds for the sum $\sum_{P \in [p]} |a_{P,\pi}|$.) For any constant $\delta > 0$, define the two sets

$$A_\delta \coloneqq \{(f, \sigma^2)\colon |T_{j,\pi}(f) - T_{j,\pi}(f_0)| + |\sigma^2 - \sigma_0^2| > \delta\}$$

$$B_\delta \coloneqq \{(f, \sigma^2)\colon D_{T,\pi}|T_{j,\pi}(f) - T_{j,\pi}(f_0)| + D_{T,\pi}|\sigma^2 - \sigma_0^2| > \delta\}.$$

Because $D_{T,\pi} \geq 1$, we have $A_\delta \subseteq B_\delta$ for all $\delta > 0$. Let $\mathscr{D}_n \coloneqq \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. By Theorem 4 of Jeong and Rockova (2023), there exists a constant $M > 0$ such that $\mathbb{E}_0\Pi(B_{L_{T,\pi}\epsilon_n} \mid \mathscr{D}_n) \to 0$ as $n \to \infty$, where $L_{T,\pi} = D_{T,\pi}M$. Because $A_{L_{T,\pi}\epsilon_n} \subseteq B_{L_{T,\pi}\epsilon_n}$ for all n, we have $\mathbb{E}_0\Pi(A_{L_{T,\pi}\epsilon_n} \mid \mathscr{D}_n) \to 0$ as $n \to \infty$. $\square$

The proofs of Theorem 4 and Corollary 4 can be obtained by replacing the random-design bound C_0^* with $\sqrt{\log n}$ and the distribution π with the probability measure $P_{\mathscr{X}}$.

Regarding the constant $D_{T,\pi}$ (and the corresponding constants $D_{V,\pi,|P|}$ and $D_{T,\pi}$) in the proof of Corollary 3, the sum $\sum_{P\in[p]} |a_{P,\pi}|$ seems to grow exponentially in p. Theorem 5 below states that this exponential dependence on p holds really only for a total-effect Sobol´ index (although the sum for a Sobol´ index V_P is $2^{|P|}-1$, in practice such indices are computed only for $|P| \leq 3$). However, p is often much larger than the order of the highest-order interaction in the true function. If the input distribution π is orthogonal (which is needed for a Sobol´ index to be interpretable), if the true function does not contain interactions of order larger than $q \leq p$, and if the BART posterior assigns zero probability to functions with interactions of order larger than q (this third assumption is not unreasonable for even moderately large q, given that BART's prior discourages deep trees and a tree's regression function cannot have interactions of order larger than the tree's depth), then the sum's dependence on p for the total-effect index reduces to an exponential dependence on q, which is often quite small. (We can further reduce this dependence on q if π is orthogonal by omitting Sobol´ index terms for subsets containing inert variables in a similar fashion as described in the proof of Theorem 5.)

Theorem 5. *Upper bounds for D_S, $D_{V,\pi,|P|}$, and $D_{T,\pi}$ in the proof of Corollary 3 are, respectively,* $\max\{1, 8C_0^*\}$, $\max\{1, 4C_0^*(2^{|P|}-1)\}$, *and* $\max\{1, 4C_0^* \sum_{i=0}^{p-1}(p-$

1)!/\{(p-1-i)!i!\}(2^{i+1}-1)\}$. *If* π *is orthogonal and neither* f *nor* f_0 *in Corollary 3 contain interactions of order larger than* $q < p$*, then the preceding upper bounds for* $D_{V,\pi,|P|}$ *and* $D_{T,\pi}$ *can be reduced to, respectively,* $\max\{1, 4C_0^* \sum_{i=1}^{\min\{q,|P|\}} P!/\{(P-i)!i!\}\}$ *and* $\max\{1, 4C_0^* \sum_{i=0}^{q-1} (q-1)!/\{(q-1-i)!i!\}(2^{i+1}-1)\}$.

Proof of Theorem 5. For the Shapley effect bound, we note that

$$\frac{1}{p} \sum_{P \subseteq ([p] \setminus \{j\})} \frac{(p-1-|P|)!|P|!}{(p-1)!} = \frac{1}{p} \sum_{i=0}^{p-1} 1 = 1.$$

This with (S5.7) and Theorem 3 together imply

$$\begin{aligned}
\Big| S_{j,\pi}(f) - S_{j,\pi}(f_0) \Big| &\le \left| \frac{1}{p} \sum_{P \subseteq ([p] \setminus \{j\})} \frac{(p-1-|P|)!|P|!}{(p-1)!} \left[c_{P\cup\{j\},\pi}(f) - c_{P\cup\{j\},\pi}(f_0) \right] \right| \\
&\quad + \left| \frac{1}{p} \sum_{P \subseteq ([p] \setminus \{j\})} \frac{(p-1-|P|)!|P|!}{(p-1)!} \left[c_{P,\pi}(f) - c_{P,\pi}(f_0) \right] \right| \\
&\le 4C_0^* \|f - f_0\|_{2,\pi} + 4C_0^* \|f - f_0\|_{2,\pi}.
\end{aligned}$$

As defined in Section 2.2, a Sobol´ index $V_{P,\pi}(f)$ is a linear combination of costs (S5.5) over all nonempty subsets of P, where each coefficient in the linear combination is either 1 or -1. Since P has $\sum_{i=1}^{|P|} |P|!/\{(|P|-i)!i!\} = 2^{|P|} - 1$ many nonempty subsets, we can use Theorem 3 to get

$$\Big| V_{P,\pi}(f) - V_{P,\pi}(f_0) \Big| \le (2^{|P|} - 1) 4C_0^* \|f - f_0\|_{2,\pi}.$$

If π is orthogonal and there are no interactions of order larger than q, then the Sobol´ indices for the subsets of P containing more than q elements are zero, and hence we can omit those Sobol´ indices from $V_{P,\pi}(f) - V_{P,\pi}(f_0)$. Since P has $\sum_{i=1}^{\min\{q,|P|\}} |P|!/\{(|P|-i)!i!\}$ many nonempty subsets containing at most q elements, the desired result follows.

As defined in Section 2.2, a total-effects Sobol´ index $T_{j,\pi}(f)$ is the sum of $V_{P,\pi}(f)$ over all subsets $P \subseteq [p]$ containing j. Using the above result, we get

$$\begin{aligned}
\left|T_{j,\pi}(f) - T_{j,\pi}(f_0)\right| &= \left|\sum_{P\subseteq([p]\setminus\{j\})} V_{P\cup\{j\},\pi}(f) - V_{P\cup\{j\},\pi}(f_0)\right| \\
&\leq \sum_{P\subseteq([p]\setminus\{j\})} (2^{|P|+1} - 1)4C_0^*\|f - f_0\|_{2,\pi} \\
&\leq \sum_{i=0}^{p-1} \frac{(p-1)!}{(p-1-i)!i!}(2^{i+1} - 1)4C_0^*\|f - f_0\|_{2,\pi}.
\end{aligned}$$

If π is orthogonal and there are no interactions of order larger than q, then the remaining result follows if we again omit from each sum the subsets of P containing more than q elements. □

S6 Functions used in simulation studies in Section 5

Table 2 contains the variances, Sobol´ indices, and Shapley values for each test function.

1. The "Friedman" function (Friedman, 1991) is defined as

$$f(\mathbf{x}) := 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5.$$

2. The "Morris" function inspired by Morris et al. (2006) is defined as

$$f(\mathbf{x}) \coloneqq \alpha \sum_{i=1}^{d} x_i + \beta \sum_{i=1}^{d-1} x_i \sum_{j=i+1}^{d} x_j$$

where $\alpha = \sqrt{12} - 6\sqrt{0.1(d-1)} \approx -0.331$ and $\beta = \frac{12}{\sqrt{10(d-1)}} \approx 1.897$ are chosen.

3. The "Bratley" function (Bratley et al., 1992; Kucherenko et al., 2011) is defined as

$$f(\mathbf{x}) \coloneqq \sum_{i=1}^{d} (-1)^i \prod_{j=1}^{i} x_j = -x_1 + x_1x_2 - x_1x_2x_3 + x_1x_2x_3x_4 - x_1x_2x_3x_4x_5.$$

4. The "g-function" from Saltelli and Sobol´ (1995) is defined as

$$f(\mathbf{x}) \coloneqq \prod_{k=1}^{d} \frac{|4x_k - 2| + c_k}{1 + c_k},$$

where we use $c_k = (k-1)/2$ for $k = 1, \ldots, d$ suggested by Crestaux et al. (2009).

S7 How do metamodels scale with input dimension?

Here we explore how various metamodels scale with input dimension p. Sparse variational GPs are known for being scalable in sample size n, but as explained in the abstract of Burt et al. (2020), to make the KL-divergence between the approximate model and the exact posterior arbitrarily small for

	Friedman			Morris			Bratley			$g-$function		
	Var: 23.8			Var: 5.25			Var: 0.057			Var: 3.076		
j	V_j^*	T_j^*	S_j^*	V_j^*	T_j^*	S_j^*	V_j^*	T_j^*	S_j^*	V_j^*	T_j^*	S_j^*
1	0.197	0.274	0.235	0.190	0.210	0.2	0.688	0.766	0.725	0.411	0.558	0.482
2	0.197	0.274	0.235	0.190	0.210	0.2	0.142	0.220	0.179	0.183	0.288	0.233
3	0.093	0.093	0.093	0.190	0.210	0.2	0.051	0.099	0.073	0.103	0.172	0.135
4	0.350	0.350	0.350	0.190	0.210	0.2	0.006	0.018	0.011	0.066	0.113	0.088
5	0.087	0.087	0.087	0.190	0.210	0.2	0.006	0.018	0.011	0.046	0.080	0.062

Table 2: Normalized main-effects $V_j^* = V_j^*(f)$, total-effects $T_j^* = T_j^*(f)$, and Shapley effects $S_j^* = S_j^*(f)$ for various functions f and variable indices $j \in [5]$ under orthogonal inputs.

a Gaussian-noise regression model with $M << N$ inducing points, squared-exponential kernel, and p-dimensional Gaussian distributed covariates, an overall computational cost of $O(N(\log N)^{2p}(\log\log N)^2)$, which is exponential in the input dimension p, is required.

Deep GPs (Damianou and Lawrence, 2013; Sauer et al., 2023) are meant to be scalable in sample size, but not necessarily in total input dimension p. Figure 1 shows that for a single-layer deep GP (which is essentially a typical GP) with anisotropic lengthscales, its training time increases strongly with p because a different parameter is needed for each input dimension. As discussed in our theoretical results, our approach adapts to anisotropy and hence this figure shows only models that also allow for anisotropy. However, we also found in this same study that for a GP with isotropic lengthscales, its training time is roughly constant with p because the same parameter is used for all input dimensions; hence such a GP could be computationally useful for isotropic regression functions with high-dimensional input variables.

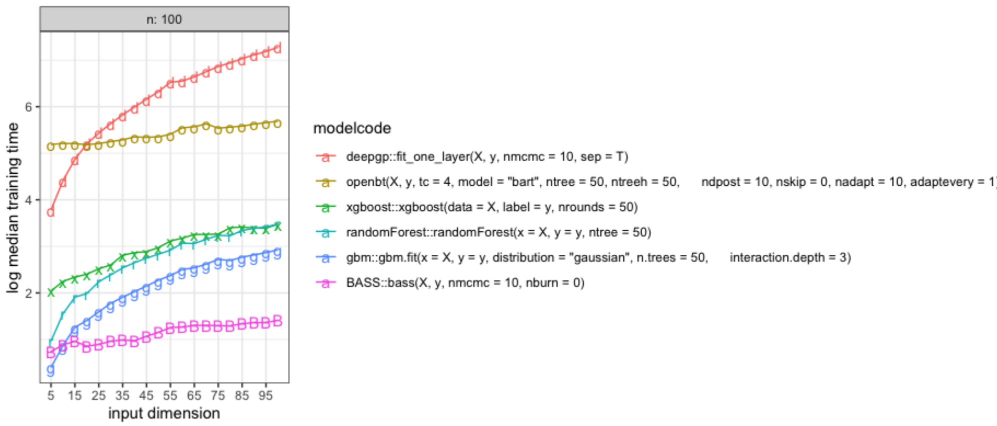

Figure 1: Average training time (over 10 evaluations per input dimension) of metamodel implementations on an M1-chip 4-core laptop for $n = 100$.

Figure 2 shows the training time for the `OpenBT` implementation of BART Pratola (2023), the `randomForest` package (Liaw and Wiener, 2002) implementation of random forests (Breiman, 2001), the `xgboost` package (Chen et al., 2024) implementation of Extreme Gradient Boosting, the `gbm` package (Ridgeway and Developers, 2024) implementation of Generalized Boosted Regression Models, and the `BASS` package (Francom and Sansó, 2020) implementation of Bayesian MARS (Francom et al., 2018). We see that as p increases, `randomForest`, `xgboost`, and `gbm` seem to increase at roughly the same rate, and that this rate is larger than the rate for either `openBT` or `BASS`. The training time for `openBT` seems to have a larger start-up time, but based on Figure 2 and Figure 1, the training time for `openBT` seems to grow more slowly in n than do the other ensemble methods. Interestingly, `BASS` has a small training time for both sample sizes, and seems to grow slowly in p.

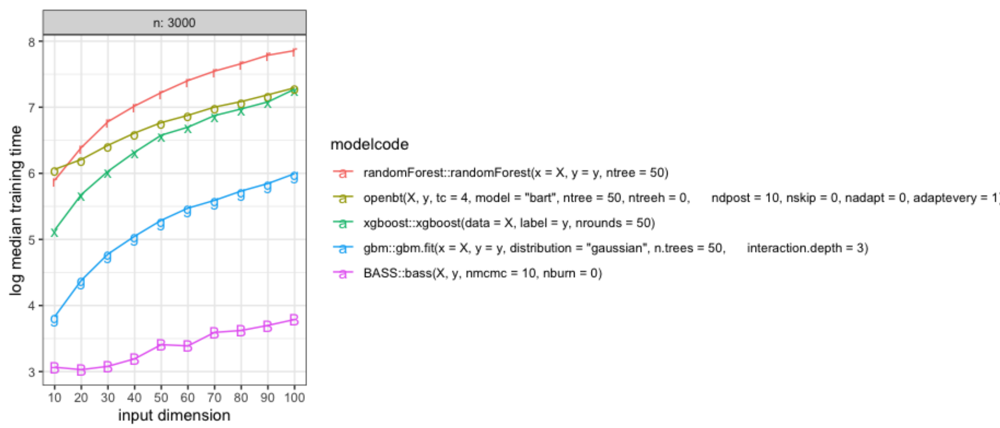

Figure 2: Average training time (over 10 evaluations per input dimension) of metamodel implementations on an M1-chip 4-core laptop for $n = 3000$.

References

Bratley, P., B. L. Fox, and H. Niederreiter (1992). Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation (TOMACS) 2*(3), 195–213.

Breiman, L. (2001). Random forests. *Machine Learning 45*, 5–32.

Burt, D. R., C. E. Rasmussen, and M. Van Der Wilk (2020). Convergence of sparse variational inference in gaussian processes regression. *Journal of Machine Learning Research 21*(131), 1–63.

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, Y. Li, and J. Yuan (2024). *xgboost: Extreme Gradient Boosting*. R package version 1.7.8.1.

Crestaux, T., O. L. Maître, and J.-M. Martinez (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety 94*(7), 1161 – 1172. Special Issue on Sensitivity Analysis.

Damianou, A. and N. D. Lawrence (2013). Deep gaussian processes. In *Artificial intelligence and statistics*, pp. 207–215. PMLR.

Francom, D. and B. Sansó (2020). BASS: An R package for fitting and performing sensitivity analysis of Bayesian adaptive spline surfaces. *Journal of Statistical Software 94*(8), 1–36.

Francom, D., B. Sansó, A. Kupresanin, and G. Johannesson (2018). Sensitivity analysis and emulation for functional data using bayesian adaptive splines. *Statistica Sinica 28*(2), 791–816.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics 19*(1), 1–67.

Ghosal, S. and A. Van der Vaart (2017). *Fundamentals of nonparametric Bayesian inference*, Volume 44. Cambridge University Press.

Horiguchi, A., M. T. Pratola, and T. J. Santner (2021). Assessing variable activity for bayesian regression trees. *Reliability Engineering & System Safety 207*, 107391.

Iooss, B. and C. Prieur (2019). Shapley effects for sensitivity analysis with correlated inputs: comparisons with sobol´ indices, numerical estimation and applications. *International Journal for Uncertainty Quantification 9*(5).

Jeong, S. and V. Rockova (2023). The art of bart: Minimax optimality over nonhomogeneous smoothness in high dimension. *Journal of Machine Learning Research 24*(337), 1–65.

Kucherenko, S., B. Feil, N. Shah, and W. Mauntz (2011). The identification of model effective dimensions using global sensitivity analysis. *Reliability Engineering & System Safety 96*(4), 440–449.

Liaw, A. and M. Wiener (2002). Classification and regression by randomforest. *R News 2*(3), 18–22.

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

Mohammadi, H., P. Challenor, M. Goodfellow, and D. Williamson (2019). Emulating computer models with step-discontinuous outputs using gaussian processes. *arXiv preprint arXiv:1903.02071*.

Morris, M. D., L. M. Moore, and M. D. McKay (2006). Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs. *Journal of Statistical Planning and Inference 136*(9), 3203–3220.

Pratola, M. T. (2023). Open Bayesian trees. Accessed: 2023-04-03.

Ridgeway, G. and G. Developers (2024). *gbm: Generalized Boosted Regression Models*. R package version 2.2.2.

Saltelli, A. and I. M. Sobol´ (1995). About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering & System Safety 50*(3), 225 – 239.

Sauer, A., R. B. Gramacy, and D. Higdon (2023). Active learning for deep gaussian process surrogates. *Technometrics 65*(1), 4–18.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety 93*(7), 964–979.