Statistica Sinica Preprint No: SS-2024-0318	
Title	Estimating Shapley Effects in Big-Data Emulation and Regression Settings using Bayesian Additive Regression Trees
Manuscript ID	SS-2024-0318
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0318
Complete List of Authors	Akira Horiguchi and Matthew T. Pratola
Corresponding Authors	Akira Horiguchi
E-mails	ahoriguchi@ucdavis.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

Statistica Sinica

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

Akira Horiguchi, Matthew T. Pratola

University of California Davis, Indiana University Bloomington

Abstract: Shapley effects are a particularly interpretable approach to assessing how a function depends on its various inputs. The existing literature contains various estimators for this class of sensitivity indices in the context of nonparametric regression where the function is observed with noise, but there does not seem to be an estimator that is computationally tractable for input dimensions in the hundreds scale. This article provides such an estimator that is computationally tractable on this scale. The estimator uses a metamodel-based approach by first fitting a Bayesian Additive Regression Trees model which is then used to compute Shapley-effect estimates. This article also establishes a theoretical guarantee of posterior consistency on a large function class for this Shapley-effect estimator. Finally, this paper explores the performance of these Shapley-effect estimators on four different test functions for various input dimensions, including $p = 500$.

Key words and phrases: Nonparametric, functional ANOVA, global sensitivity analysis, variable importance, surrogate model

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

1. Introduction

An important task in global sensitivity analysis is to measure how a real-valued function depends on its various inputs. A popular measure of variable importance is the class of Sobol´ indices (Sobol´, 1990), which decomposes the variance of outputs from a function into terms due to main effects for each input and interaction effects between the various inputs. To quantify the impact of any particular input dimension, either the *main-effect Sobol´ index* or the *total-effect Sobol´ index* can be used; the latter includes all interactions between the given input and any other input whereas the former excludes any such interaction. Straightforward interpretation of Sobol´ indices requires an orthogonal distribution on the inputs (Song et al., 2016). *Shapley effects* (Shapley, 1952; Song et al., 2016) form another class of variance-based global sensitivity indices that was first introduced in the context of game theory but has only recently been gaining traction in the statistics literature (Owen, 2014). Although the additional computation required to compute Shapley effects might render them unnecessary if the inputs are known to be independent, Shapley effects remain interpretable even if the inputs are correlated (Song et al., 2016) and hence are the more reasonable option in such a case.

If the function of interest is known and has a simple enough form, its

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

exact Shapley effects can sometimes be computed analytically, particularly when the required integrals can be computed easily. Otherwise, the Shapley effects can be estimated using values generated from the function. Many existing methods assume the function can be evaluated cheaply and without observation noise and indeed work well in such a scenario. Figure 1 shows various such Shapley-effect estimators (Song et al., 2016; Benoumechiara and Elie-Dit-Cosaque, 2019; Broto et al., 2020; Plischke et al., 2021; Goda, 2021) applied to n observations generated from a function (defined in the figure caption) evaluated on i.i.d. inputs drawn uniformly from the hypercube $[0, 1]^5$. When the function values are observed without noise, these methods track the g-function's true Shapley-effects very well. But when independent and identically distributed (i.i.d.) Gaussian noise with mean zero and moderate variance (defined in the figure caption) is added, these methods struggle to capture the true values even when the number of observations increases dramatically to compensate for the observation noise.

For noisy function observations, one can first estimate the function and then compute sensitivity indices of the estimated function as a post-processing step. One option is to fit a metamodel to the observations; the fitted metamodel then serves as the estimated function. (This approach is also useful in noisefree settings when the function can only be sparsely eval-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

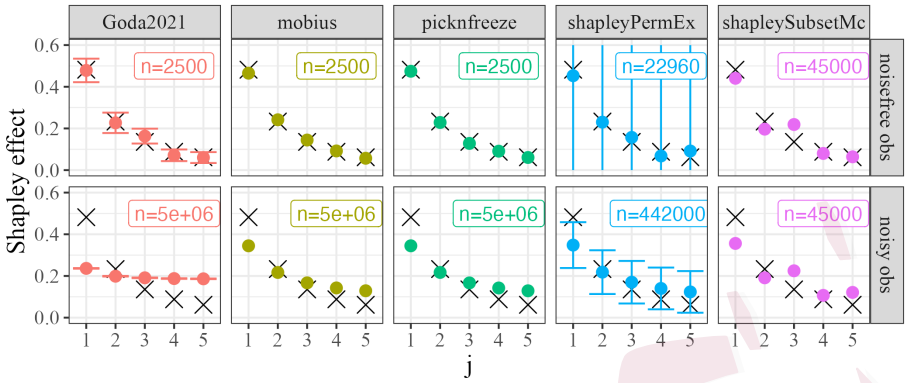

Figure 1: Shapley-effect estimates of various existing methods trained on data drawn from the Sobol´ g-function $f(\mathbf{x}) = \prod_{k=1}^{5} \frac{|4x_k-2|+(k-1)/2}{1+(k-1)/2}$. Crosses represent the true Shapley-effect values. Function values are evaluated at n i.i.d. inputs drawn uniformly from the hypercube $[0,1]^5$. In the top row, function values are observed without noise. In the bottom row, the observations are function values plus i.i.d. Gaussian noise with mean zero and variance 0.25×3.076, where 3.076 is the variance of the g-function under a uniform distribution on $[0,1]^5$. Each column represents an estimation method: "Goda2021" is from Goda (2021); "mobius" and "picknfreeze" are from Plischke et al. (2021); "shapleyPermEx" and "shapleySubsetMc" are from Iooss et al. (2023). Error bars represent approximate or exact 95% confidence intervals as implemented by the method which aim to capture the variability induced by the Monte Carlo approximation of expectations.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

uated and the fitted metamodel can be evaluated cheaply.) Popular metamodels include the Gaussian Process (GP), Bayesian multivariate adaptive regression splines (BMARS) (Denison et al., 1998), generalized polynomial chaos expansions (PCE) (Sudret, 2008), treed GPs Gramacy and Taddy (2010), dynamic trees (Gramacy et al., 2013), Gaussian radial basis function (Wu et al., 2016), artificial neural networks (Li et al., 2016), and deep GPs (Radaideh and Kozlowski, 2020). This paper makes its contributions using Bayesian Additive Regression Trees (BART) (Chipman et al., 2010) which is an increasingly popular tool for complex regression problems and as emulators of expensive computer simulations (Chipman et al., 2012; Gramacy and Haaland, 2016; Horiguchi et al., 2022). BART is a nonparametric sum-of-trees model embedded in a Bayesian inferential framework. Unlike many other metamodels, BART can easily incorporate categorical inputs, avoids strong parametric assumptions, and is relatively quick to fit even on a large number of observations. BART even has been shown to be resilient to the inclusion of inert inputs, particularly when the BART prior incorporates either the sparsity-inducing Dirichlet prior of Linero (2018) or the spike-and-tree prior of van der Pas and Ročková (2017); Liu et al. (2021). Furthermore, the Bayesian framework provides natural uncertainty quantification for both predictions and sensitivity-index estimates.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

Some metamodels struggle more than others with the two stages in the above approach, namely fitting the metamodel, then using the fitted metamodel to estimate the sensitivity indices. Regarding the first stage, many of these metamodel-based approaches struggle to fit if the number of inputs p and function evaluations n are not small. A GP has $O(n^3)$ computation time and struggles to fit for even $p = 10$. PCE has been fit for $p = 25$, but it has been noted that PCE struggles to fit for larger p (Sudret, 2008; Crestaux et al., 2009). BMARS works for $p = 200$ for Sobol' indices (Francom et al., 2018). Figure 7 of this paper provides an example where BART fits to a $p = 500$ scenario with $d = 250$ active variables. Regarding the second stage, if a metamodel is cheap to evaluate, then the fitted metamodel's Shapley effects can be estimated using Monte Carlo integration of the Shapley-effect integrals, as done in Algorithm 1 from Song et al. (2016) or a parallelized version of it (Zhang and Dimitrov, 2024). However, this will create another layer of approximation error that can be avoided if the metamodel allows for exact computation. On this front, BART (Horiguchi et al., 2021), BMARS (Francom et al., 2018), and PCE (Sudret, 2008) have closed-form expressions for Sobol´ indices (and thus for Shapley effects) that can be computed exactly once the metamodel is fit. Such expressions also exist for GPs with polynomial mean and either

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

a separable Gaussian, Bohman, or cubic correlation function (Oakley and O'Hagan, 2004; Chen et al., 2005, 2006; Marrel et al., 2009; Moon, 2010; Svenson et al., 2014; Santner et al., 2018). Table 1 in the Supplementary Material summarizes these metamodel properties.

To our knowledge, this article is the first to provide an estimator of a function's *Shapley effects* that is computationally tractable for a relatively large number of inputs and function evaluations, as well as theoretical guarantees of consistency in the context of nonparametric regression where the function is observed with noise. BART approximates a function by a piecewise-constant function whose exact Sobol´ indices are provided by Horiguchi et al. (2021) and can be easily computed (we will refer to these as "BART-based Sobol´ indices" for the rest of this article). Section 2 will show these closed-form expressions can also be used to compute BART-based Shapley effects, but because the number of expressions to compute increases dramatically, Section 3 discusses computationally friendly approximations. On the other hand, our contraction-rate results rely heavily on recent BART theory from Jeong and Rockova (2023), who introduce the large class of sparse piecewise heterogeneous anisotropic Hölder functions and show that over this function class, the contraction rate for Bayesian forests is optimal up to a logarithmic factor.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

This article is organized as follows. Section 2 reviews BART, Sobol´ indices, and Shapley effects. Section 3 provides our main theoretical posterior contraction results and discusses the computation of BART-based Shapley effects. Section 4 showcases their performance on numerical examples, including data from the En-Roads climate simulator (analogous discussion for BART-based Sobol´ indices can be found in Horiguchi et al. (2021)). Section 5 provides discussion on future work. Our results on posterior contraction for BART-based Sobol´ indices and Shapley effects, as well as proofs of these results, are included as Supplementary Material.

2. Review

Mirroring Jeong and Rockova (2023), this article considers regression settings with either a fixed or random design. The regression model with *fixed* design is

$$Y_i = f_0(\mathbf{x}_i) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma_0^2), \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\sigma_0^2 < \infty$ and each covariate $\mathbf{x}_i \in [0,1]^p$ is fixed. A fixed design would be assumed if, for example, the trees in BART are allowed to split only on observed covariate values (which was a specification used in the seminal BART paper (Chipman et al., 2010)) or on dyadic midpoints of the domain.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

The regression model with *random* design is

$$Y_i = f_0(\mathbf{X}_i) + \varepsilon_i, \qquad \mathbf{X}_i \sim \pi, \qquad \varepsilon_i \sim N(0, \sigma_0^2), \qquad i = 1, \ldots, n, \quad (2.2)$$

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 $\text{supp}(\pi) \subseteq [0,1]^p$. A random design would be assumed for estimation problems such as density estimation or regression/classification with random design. Our posterior contraction results deal separately with fixed or random designs.

2.1 BART

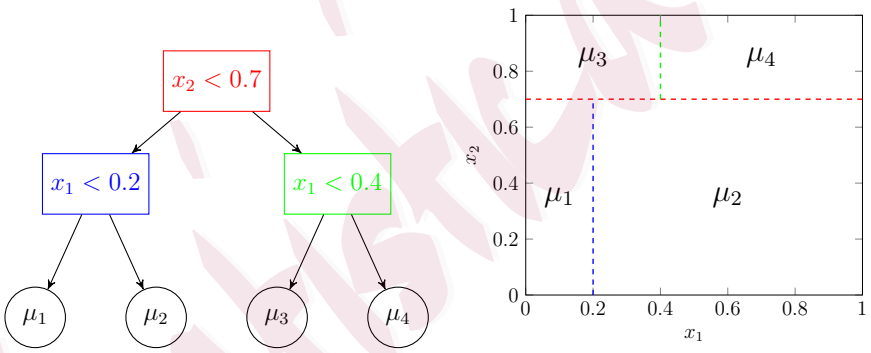

Figure 2: An example tree shown graphically (left) and as a piecewise-constant regression function (right) on the input space $[0,1]^2$.

In a regression setting in the form of either (2.1) and (2.2), a BART model approximates the unknown function f_0 by a sum of T regression

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

trees:

$$f_0(\cdot) \approx \sum_{t=1}^{T} g(\cdot; \Theta_t), \tag{2.3}$$

where each regression-tree function $g(\cdot; \Theta_t) \colon [0,1]^p \to \mathbb{R}$ is piecewise constant over the input space. Each parameter set Θ_t determines a partition of the input space $[0,1]^p$ into boxes (i.e. hyperrectangles) and the fitted response values assigned to each partition piece. The partition is induced by recursively applying binary splitting rules; Figure 2 shows an illustrative example. To regularize the model fit, the BART prior over the parameters $\{\Theta_t\}_{t=1}^T$ keeps the individual tree effects small, which causes each function $g(\cdot; \Theta_t)$ to contribute a small portion to the total approximation of f_0. The expected response $E\big[Y(\mathbf{x}) \mid \{\Theta_t\}_{t=1}^T\big]$ at a given input $\mathbf{x}$ is then the sum of each contribution $g(\mathbf{x}; \Theta_t)$.

Though the right hand side of (2.3) is piecewise constant, Jeong and Rockova (2023) shows that under certain conditions, BART can approximate the unknown function f_0 (which itself need not be piecewise constant) arbitrarily closely with attractive posterior contraction rates. For space consideration, the Supplementary Materials will describe the types of functions that BART can capture and the conditions made in the theorems of Jeong and Rockova (2023) that our contraction-rate results rely on.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

2.2 Sobol´ indices

Let $L^2 \equiv L^2([0,1]^p)$ denote the space of real-valued, square-integrable functions on the hypercube $[0,1]^p$. Sobol´ (1990, 1993) shows that if the random variable $\mathbf{X}$ follows an orthogonal distribution whose support is $[0,1]^p$ and if $f \in L^2$, then the variance of $f(\mathbf{X})$ can be decomposed into a sum of terms attributed to single inputs or to interactions between sets of inputs:

$$\mathrm{Var}\{f(\mathbf{X})\} = \sum_{j=1}^{p} V_j + \sum_{j=1}^{p} \sum_{k<j} V_{jk} + \cdots + V_{1,2,\ldots,p} \tag{2.4}$$

where we recursively define for each variable index set $P \subseteq [p]$

$$V_P \coloneqq \mathrm{Var}[E\{f(\mathbf{X}) \mid \mathbf{X}_P\}] - \sum_{Q \subset P} V_Q$$

where we set $V_\emptyset = 0$ and the relation $\subset$ denotes a strict subset. For any variable index $j \in [p]$, the term $V_{\{j\}} = V_j$ is known as the jth (unnormalized) first-order (or main-effect) Sobol´ index, and the sum $T_j = \sum_{P \subseteq ([p] \setminus \{j\})} V_{P \cup \{j\}}$ is known as the jth (unnormalized) total-effect Sobol´ index. We note that $T_j \geq V_j \geq 0$ for all $j \in [p]$.

The V_P terms in (2.4) are often divided by the total variance to produce the normalized terms $V_P/[\mathrm{Var}\{f(\mathbf{X})\}]$, which have the nice interpretation of being the proportion of the total variance attributed to the interaction between the variables whose indices are in the index set P. If P is the singleton $\{j\}$, then the normalized term $V_j/[\mathrm{Var}\{f(\mathbf{X})\}]$ can be interpreted

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

as the proportion of the total variance attributed to variable j by itself. Despite this nice interpretation, the remainder of the article will assume that such indices are unnormalized unless otherwise stated.

To see why these indices' interpretation requires $\mathbf{X}$ to follow an orthogonal distribution, we extend the definition of V_P by allowing $\mathbf{X}$ to follow a possibly non-orthogonal distribution π whose support is $[0,1]^p$. We first define the functional $c_{P,\pi}\colon L^2 \to \mathbb{R}$ as

$$c_{P,\pi}(f) = \mathrm{Var}_\pi[E_\pi\{f(\mathbf{X}) \mid \mathbf{X}_P\}] \tag{2.5}$$

for any $f \in L^2$. Then the generalized V_P under the distribution π is recursively defined as

$$V_{P,\pi}(f) \coloneqq c_{P,\pi}(f) - \sum_{Q \subset P} V_{Q,\pi}(f),$$

where again we set $V_{\emptyset,\pi}(f) = 0$. Similarly, we define the generalized jth total-effect term:

$$T_{j,\pi}(f) = \sum_{P \subseteq ([p] \setminus \{j\})} V_{P \cup \{j\},\pi}(f)$$

where $\subseteq$ denotes a subset that is not necessarily strict. Recall that if π is orthogonal and $f \in L^2$, then $T_{j,\pi}(f) \geq V_{j,\pi}(f) \geq 0$ for all $j \in [p]$ and the variance decomposition (2.4) (where orthogonality implies $V_P = V_{P,\pi}(f)$ for all $P \subseteq [p]$) holds. However, Theorem 2 of Song et al. (2016) asserts the existence of a non-orthogonal distribution π and a function $f \in L^2$ such

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

that $\sum_{j=1}^{p} V_{j,\pi}(f) > \mathrm{Var}_{\pi}\{f(\mathbf{X})\} > \sum_{j=1}^{p} T_{j,\pi}(f)$. In such a case, these Sobol´ indices can no longer be interpreted as in the orthogonal case.

2.3 Shapley effects

One way to measure variable activity, regardless of dependence among inputs, are the Shapley effects defined by Song et al. (2016) as the Shapley values in Owen (2014) using the functional (2.5) as the "value" or "cost." For $j \in [p]$ the jth Shapley effect is defined as

$$S_{j,\pi}(f) = (p!)^{-1} \sum_{P \subseteq ([p] \setminus \{j\})} (p - |P| - 1)!\,|P|!\,\left\{c_{P\cup\{j\},\pi}(f) - c_{P,\pi}(f)\right\}, \quad (2.6)$$

which has the desirable property $\sum_{j=1}^{p} S_{j,\pi}(f) = \mathrm{Var}_{\pi}\{f(\mathbf{X})\}$ for any distribution π (possibly nonorthogonal) whose support is $[0,1]^p$. Hence, the jth (normalized) Shapley effect can be nicely interpreted as the contribution of input j to the total output variance. Furthermore, if π is orthogonal, then

$$V_{j,\pi}(f) \leq S_{j,\pi}(f) \leq T_{j,\pi}(f) \quad (2.7)$$

for any $f \in L^2$ and $j \in [p]$ (Owen, 2014, Section 3), i.e. the jth Shapley effect is bounded between the jth main-effect and total-effect Sobol´ index.

Calculating (2.6) can be prohibitively costly due to it being a sum of values (2.5) over all subsets of a set $[p] \setminus \{j\}$. Its computational tractability will be discussed in Section 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

3. Main results and computation of Shapley effects

This section will address theoretical support and computation of Shapley effects using a BART metamodel. The metamodel-based approach in estimating Shapley effects has two approximation layers: how well the metamodel approximates (functionals of) the underlying regression function f_0, and how well the Shapley-effect estimates approximate the Shapley effects of the metamodel function.

3.1 Consistency Result

For the second layer, we establish posterior consistency for our BART-based Shapley effects using (first-layer) posterior consistency for BART from Jeong and Rockova (2023). The required theoretical results, fully developed in the Supplementary Material, characterize the posterior contraction as the dataset size $n \to \infty$. The contraction rate quantifies how quickly the posterior distribution approaches the underlying function's true Shapley effects. In particular, for random designs, we have the following.

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 in Section S7, there exist positive constants $L_{V,\pi,|P|}$, $L_{T,\pi}$, and L_S such that as*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

$n \to \infty$ for ϵ_n in Eq. (S5.4) in Section S7,

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

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

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

The supplement contains a similar result for fixed designs, as well as proofs for all theoretical results.

3.2 Shapley Effect Computation

The remainder of this section will address the computation of Shapley effects, and how well the Shapley-effect estimates approximate the Shapley effects of the metamodel function. Since BART is a Bayesian metamodel, our focus is to address the computational aspects when f_0 is approximated by n_{draw} posterior draws $\hat{f}^{(1)},\ldots,\hat{f}^{(n_{draw})}$ of the fitted metamodel.

For each input $j \in [p]$, we can construct a posterior distribution for the jth Shapley effect $S_{j,\pi}(f_0)$ of f_0 using the n_{draw} values

$$S_{j,\pi}(\hat{f}^{(i)}) = \sum_{P\subseteq([p]\setminus\{j\})} \frac{(p-|P|-1)!\,|P|!}{p!}\Big[c_{P\cup\{j\},\pi}(\hat{f}^{(i)}) - c_{P,\pi}(\hat{f}^{(i)})\Big], \quad (3.8)$$

for $i = 1,\ldots,n_{draw}$. We can use the sample mean of $\{S_{j,\pi}(\hat{f}^{(i)})\}_{i=1}^{n_{draw}}$ as a point estimate for $S_{j,\pi}(f_0)$, and use the end points of the middle 95% values as a 95% credible interval for $S_{j,\pi}(f_0)$. For each $\hat{f}^{(i)}$, computing its

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

p Shapley effects would require computing the cost function (2.5) for 2^p subsets of the set $[p]$. The exponential increase in p is undesirable, but also the calculation of even a single cost function might be computationally intractable if p is large enough.

3.3 Sum over subsets

We first tackle the exponential increase in p. If the inputs are orthogonal, then (3.8) is bounded between the main-effect and total-effect Sobol´ indices, so we can avoid computing (3.8) entirely by crudely estimating it with, for example, the mean of the two Sobol´ indices. If the inputs are not assumed to be orthogonal, we can reduce the increase from exponential to linear by using the following random-subset approach. Rather than compute the cost difference in (3.8) for all 2^{p-1} subsets, we instead compute the cost difference for only a small number m of subsets that are randomly created by including each $j' \in ([p] \setminus \{j\})$ with probability 0.5. (We could incorporate prior information about which inputs j' are important by increasing or decreasing the probability of including any particular j'. We could also incorporate prior information about interactions between groups of inputs by increasing the probability that they appear together in a subset and decreasing the probability that they appear separately. We save

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

further exploration for future work.) Hence any subset $P \subseteq ([p] \setminus \{j\})$ is chosen with probability $|P|!(p - (|P| + 1))!/(p!)$. Under this approach, we approximate (3.8) by

$$S_{j,\pi}(\hat{f}^{(i)}) \approx m^{-1} \sum_{l=1}^{m} \left\{ c_{P_l^{(i)} \cup \{j\},\pi}(\hat{f}^{(i)}) - c_{P_l^{(i)},\pi}(\hat{f}^{(i)}) \right\}, \tag{3.9}$$

where $P_l^{(i)}$ is the lth of m randomly drawn subsets of $([p] \setminus \{j\})$ for the ith posterior draw. Hence this approach reduces the number of cost-function calculations from $n_{draw} \times 2^p$ to $n_{draw} \times p \times (2m)$.

(As an aside, this approach is equivalent to how subsets are chosen in the random-permutation scheme of Castro et al. (2009): under this scheme, any subset $P \subseteq ([p] \setminus \{j\})$ can be obtained by any permutation of $[p]$ whose first $|P|$ elements are the elements in P and whose $(|P| + 1)$th element is j. Because there are $|P|!(p - (|P| + 1))!$ such permutations of $[p]$ and each permutation of $[p]$ is drawn with equal probability $(p!)^{-1}$, the subset P is selected with probability $|P|!(p - (|P| + 1))!/(p!)$. Song et al. (2016); van Campen et al. (2018); Yang et al. (2024) provide improvements and modifications on this "simple random sampling" of permutations.)

What value of m should be used for (3.9)? Both its computational cost and its fidelity to (3.8) increase with $m \times n_{draw}$. We argue that if $n_{draw} \gg 1$ (a requirement for any decent posterior summary of the surrogate model), then $m = 1$ random subset will suffice. Because the randomness from the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

original MCMC mechanism is independent of how the random subsets are chosen, (3.9) is just a noisier version of (3.8). Our approach does not inflate the correlation between any two MCMC draws. Also, it does not introduce any additional bias, seeing as each random subset P is i.i.d. and is drawn with probability exactly equal to the weight $(p - |P| - 1)!|P|!/(p!)$ in the Shapley-effect expression (2.6). Now we consider the additional variability induced by the random subsets. For each $j \in [p]$, we aim to approximate the cumulative distribution function (CDF) $F_{n,j}$ of the posterior distribution of the jth Shapley effect by an empirical CDF comprised of n_{draw} MCMC draws. (Here the sample size n is fixed and finite, but if the metamodel has posterior consistency, then $\lim_{n\to\infty} F_{n,j}$ would be a step function consisting of a single jump located at the true jth Shapley effect of the underlying regression function f_0.) Generally speaking, MCMC draws are thinned in order to be approximately i.i.d. from the limiting distribution. If we assume i.i.d. MCMC draws, Donsker's theorem (Donsker, 1952) tells us that as $n_{draw} \to \infty$, both empirical CDFs — one using random subsets, the other using exact subsets — will converge to the target CDF $F_{n,j}$ at the same rate $\mathcal{O}(n_{draw}^{-1/2})$, regardless of how many random subsets are used. Thus, we do not need to increase the number of random subsets in order to reduce the additional variability, since this variability will already shrink to zero

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

as $n_{draw} \to \infty$. Hence we use $m = 1$ for all experiments in this paper.

3.4 Cost calculation

We now consider how each calculation of (2.5) is affected by which metamodel is used. For any metamodel that can be evaluated cheaply, Algorithm 1 of Song et al. (2016) can be used to approximate the integrals in (2.5) for the metamodel by first sampling from the input distribution many times and then evaluating the metamodel on the many generated inputs. However, keeping the resulting integral approximation error small will likely require the number of random permutations and hence the computation time for each integral to grow at least linearly in p (Tang, 2024), and thus for all p inputs the computation time to grow at least quadratically in p. Additionally, approximating (2.5) in this way produces two inference issues illustrated in the independent-inputs example in Section 4.1. First, even though the exact cost difference in (2.6) is nonnegative for independent inputs by definition, the *estimated* cost difference can be negative and hence often produces negative Shapley-effect values for inert inputs (see e.g., the Morris function GP estimates in Section 4.1), which creates interpretability issues. Second, the variability due to approximating (2.5) is much larger than the variability due to the random subsets/permutations

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

(see Section 4.1 for more detail on this point).

For these reasons, it is desirable to compute (2.5) exactly, which can be done for some metamodels. For BART, a closed-form expression for (2.5) can be found using Theorem 1 of Horiguchi et al. (2021). For Bayesian MARS, Francom et al. (2018) provides a closed-form expression for estimating Sobol´ indices and contains a numerical example with $p = 200$. For a GP, a closed-form expression can be found for certain correlation functions, but these functions are typically restrictive i.e., assume stationarity and isotropy. Exact computation of each of these expressions is easy when the inputs are independent, but otherwise is challenging, in which case we will resort to using Algorithm 1 of Song et al. (2016).

4. Numerical examples

This section explores the numerical performance of BART-based Shapley effects. (BART-based Sobol´ indices are evaluated in detail in Horiguchi et al. (2021) and Horiguchi (2020) and hence are not evaluated in this paper.) Details of the test functions used in this section are listed in Section S6 in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

4.1 Exact vs Monte Carlo cost calculation

This section explores the computational and accuracy differences between computing the cost exactly as in (3.8) and estimating the cost using Algorithm 1 of Song et al. (2016) when the inputs are independent. For our first set of experiments, we create a dataset with $n = 50p$ observations and noise variance $\sigma_0^2 = 0.25\text{Var}\{f(\mathbf{X})\}$ from (2.1) for each test function f and each $p \in \{5, 50, 200\}$. To each dataset, we fit a BART model with $n_{draw} = 1000$ posterior draws and 200 trees with code from Pratola (2023). For comparison, we also fit a Gaussian process (GP) model and estimate the Shapley effects of the fitted GP mean model using Algorithm 1 of Song et al. (2016) as implemented in `shapleyPermRand`, which was the only function in the `sensitivity` R package (Iooss et al., 2023) that we found could fit to our $p = 50$ data sets in a reasonable amount of time. Parameter specifications are in the caption of Figure 3. For the $p = 200$ cases, we could not read the large GP-model file sizes into R and hence do not include these results.

We first compare the GP estimates to the BART estimates. Figure 3 shows the Shapley-effect confidence intervals of the GP approach as computed by the `sensitivity` package, and Figure 4 shows our Shapley-effect credible intervals using BART as defined in (3.9). The GP model seems to capture the large Shapley effects better than the BART model does,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

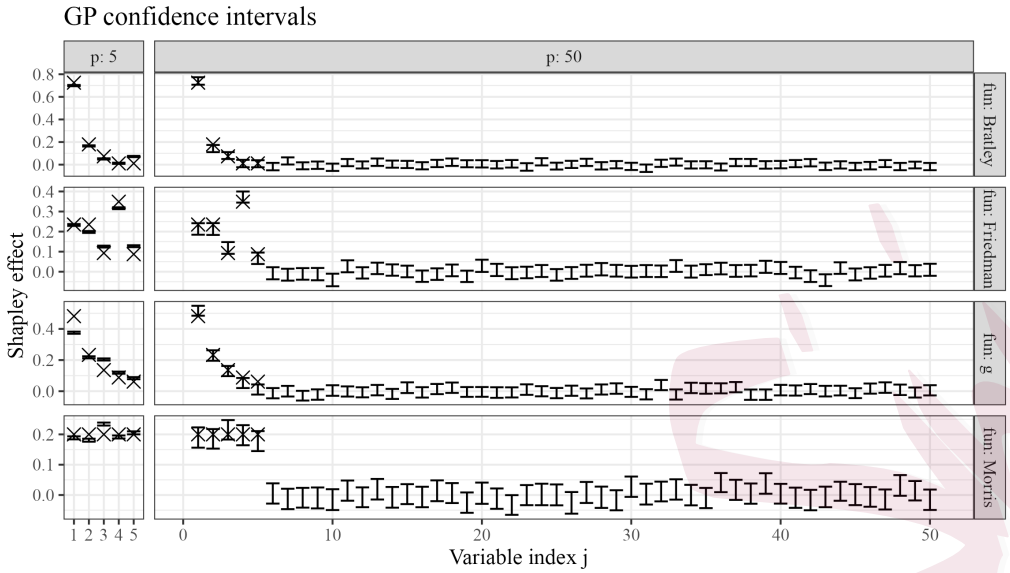

Figure 3: 95% confidence intervals (computed by `sensitivity`) for the Shapley-effect estimates from a GP fit to $n = 50p$ observations with $d = 5$ active variables. Crosses indicate a function's true Shapley effects. The `shapleyPermRand` approach samples $N_V + m(p-1)N_O N_I$ inputs to estimate expectation. Per Song et al. (2016), we set $N_V = 10^5$ samples to estimate the total variance, and $N_O = 1$ and $N_I = 3$ to estimate the outer and inner expectations, respectively. For $p = 5$ we use $m = 10^5$ random permutations; for $p = 50$ we reduce this to $m = 3 \times 10^3$ to avoid numerical overflow.

which might be explained by the fact that the data-generating functions are all continuously differentiable and thus are well suited for GPs. However, the GP model also seems to have more trouble setting the inactive

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

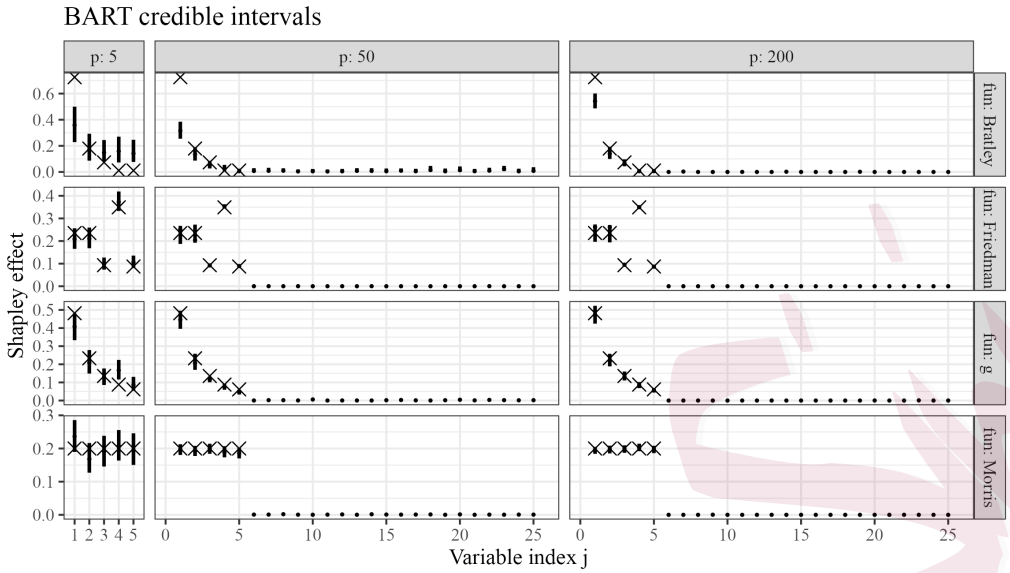

Figure 4: 95% credible intervals (as in (3.9)) over 1000 posterior draws for the Shapley-effect estimates from a BART model fit to $n = 50p$ observations with $d = 5$ active variables. Crosses indicate a function's true Shapley effects. For $p = 50$ and $p = 200$, only the first 25 input variables are shown for space considerations.

variables to have zero estimated Shapley effect; indeed, for the g-function with $p = 50$, the confidence intervals for many of the inactive variables are higher than the interval for the active variable $j = 5$. Furthermore, the GP confidence intervals for all *inactive* variables cover negative values (as computed by the `sensitivity` package), even though Shapley effects are nonnegative by definition. In contrast, the BART credible intervals never

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

cover negative values. Finally, the GP confidence intervals are wider than the BART credible intervals for the inert inputs, which is surprising since both intervals capture the variability due to the Shapley-effect estimation of the metamodel, but the BART intervals also capture the metamodel's posterior uncertainty (i.e., how accurately the metamodel fits the true regression function), whereas the GP intervals do not. (We emphasize that this stems from the different methods of estimating the Shapley effects of the respective metamodels — see Section 3.4 — rather than from the difference between GP and BART.) Hence we can conclude that in this scenario, the $m = 1$ subset approximation in Section 3.3 produces negligible errors, especially compared to the approximation error from the estimation of (2.5) even with 3000 random permutations.

Theoretically, we could increase the number of random permutations in order to reduce the Monte Carlo error of estimating (2.5) to an arbitrarily small amount, but as discussed in Section 3.4, maintaining a small approximation error for computing the cost function (2.5) for all p inputs will likely require the computation time to grow at least quadratically in p. But how does the exact computation of (2.5), which we recall is currently only implemented for BART, scale with increasing p? For two regression functions, $p = 3, \ldots, 10$, and fixed $n = 500$, Figure 5 shows that the com-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

putation time of training a BART model appears to be constant in p. The figure also shows that computing the Shapley effects of the fitted BART model using the exact (2.5) appears to grow faster than quadratic in p. This implies that the exact-cost approach scales better in p than the Monte Carlo approach if the number of random permutations increases in order to maintain a small Monte Carlo approximation error from estimating (2.5). (Because the computation times for the exact-cost approach is specific to BART, the behavior may not generalize if the exact-cost approach is implemented for other metamodels.)

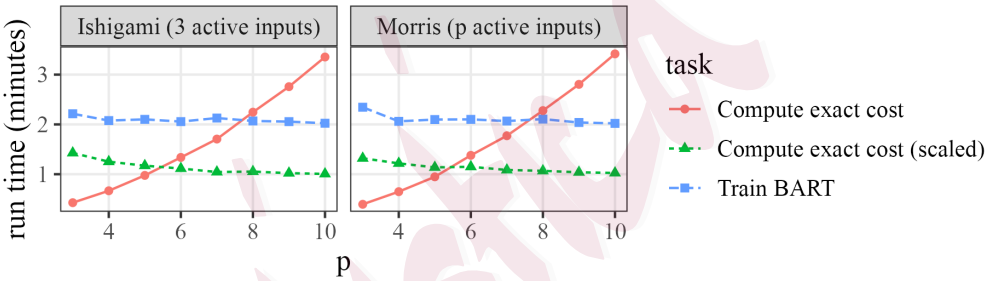

Figure 5: For $n = 500$ and various p, the figure shows the run times of training the BART model (1000 posterior draws, 9000 burn-in draws, M1-chip 4-core laptop) and of computing the Shapley effects of the fitted BART model using the exact (2.5). "Compute exact cost (scaled)" indicates the latter run time divided by p^2 (and multiplied by 30 for better visual comparison) to get an upper bound on how it scales with p for fixed n.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

Similarly, Figure 6 explores how these calculations scale with increasing n for fixed $p = 3$. The BART training time appears to grow sublinearly in n (a little faster than rate $\sqrt{n}$), and the Shapley-effect calculation time using exact computation of (2.5) appears to grow more slowly than $\log\{\log(n)\}$. This justifies the use of BART if n increases with p. (Section 5 of Pratola et al. (2014) studies the scalability of a parallel BART MCMC algorithm.)

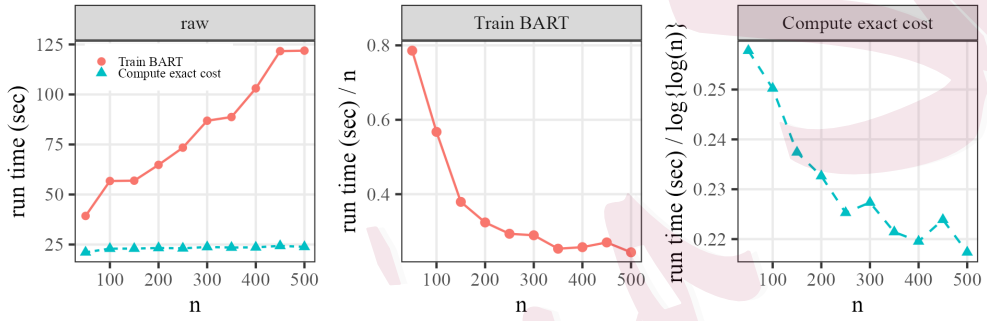

Figure 6: Left: run times of training a BART model (1000 posterior draws, 9000 burn-in draws, M1-chip 4-core laptop) and of computing the Shapley effects of the fitted BART model using the exact (2.5) for $p = 3$ and various n. Center and right panels scale these run times to get an upper bound on how the two run times scale with n for fixed $p = 3$.

We further examine the BART results in Figure 4. For the Friedman and Morris functions, the true Shapley effects are contained in the credible intervals and are often near the center of the intervals. For the g-function,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

the $p = 5$ scenario shows the credible intervals struggling a bit to capture the true Shapley effects, but the $p = 200$ scenarios show better performance from the intervals. This $p = 200$ result becomes even more notable if we consider the fitted BART models do not use (P1)'s tree prior with Dirichlet sparsity from Linero (2018), and that the g function is purely a product of univariate functions. For the Bratley function, the intervals struggle quite a bit to capture the true Shapley effects. For this challenging Bratley function, we next explore what parameters or priors should be changed to improve the Shapley-effect estimates. Of the three directions we explored – increasing the number of trees to 300, weakening the tree-depth prior to encourage higher order interactions, and increasing n – only the third (with 200 trees, the same tree-depth prior as in the first set of explorations, and $p = 5$) yielded estimates closer to the true Shapley effects. This provides assurance that for these more challenging functions, the estimates can be close to the true Shapley effects if n is large enough without having to change any other parameters or priors.

Finally, we explore the performance of BART-based Shapley effects for $d = 250$ active input variables and input dimension $p = 500$, which is a regime that bottlenecks most other methods. (We omit GP results here due to not being able to compute GP-based Shapley-effect estimates.) For

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

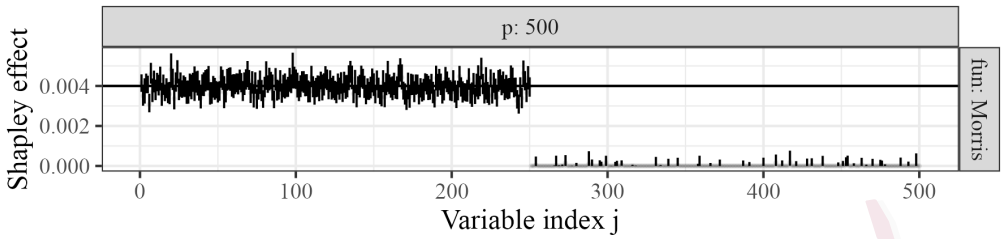

Figure 7: 95% credible intervals (as computed in (3.9)) over 300 posterior draws for the Shapley-effect estimates from a BART model fit to $n = 50p$ observations. Horizontal line corresponds to the function's true Shapley effects of the $d = 250$ active variables.

ease, we use the Morris function since its Shapley effects are $1/d$ for the d active variables. Figure 7 shows that BART clearly distinguishes between the first 250 inputs (these intervals are centered around $1/d$) and the second 250 inputs (these intervals are centered around zero) for such a large p.

4.2 Robustness of Shapley-effect estimates to input correlation

Now we explore how sensitive the estimated Shapley effects are to varying degrees of correlation. When the inputs are not independent, our proposed BART-based estimator requires being able to compute the input probability measure of various subsets of the input space $[0, 1]^p$ (this point is discussed further in Section 5). Hence, here we compute the cost function (2.5) using Monte Carlo integration as implemented by the `sensitivity` R package.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

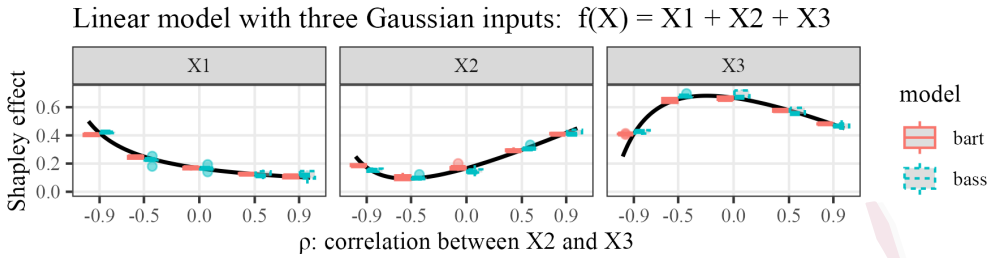

Figure 8: Shapley effects estimated by BART and BMARS (as implemented in the `BASS` R package, Francom and Sansó, 2020). The boxplot variability comes from 5 replicates of the process described in Section 4.2. The solid curves represent the function's true Shapley effects.

It can be difficult to analytically compute Shapley effects for even moderately complicated functions. Here we use the regression function $f(\mathbf{x}) = x_1 + x_2 + x_3$, where the inputs have a trivariate Gaussian distribution with mean zero and covariance matrix Σ with entries $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{33} = 4$, $\Sigma_{23} = \Sigma_{32} = 2\rho$ (where $-1 < \rho < 1$), and zero for all other entries. Thus, the input distribution is parameterized by the correlation ρ between input variables X_2 and X_3. For each $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$, we generate $n = 1000$ input values according to ρ, and then evaluate the regression function with additive Gaussian noise whose standard deviation is 0.1 times the standard deviation of the regression function under ρ. We then fit a BART model and a Bayesian MARS (BMARS) model to these observa-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

tions before using the `shapleyPermEx()` function from the `sensitivity` R package to estimate the function's Shapley effects. (The parameters we use are `Nv=1000, No=100, Ni=3`.) We repeat this process five times.

Figure 8 shows the estimated Shapley effects for these five correlation values. We see that both BART and BMARS seem to recover the true Shapley effects even for large correlations between inputs 2 and 3. There also does not seem to be any systematic performance difference between the two metamodels. Hence, if we extrapolate these results to larger input dimensions and account for the curse of dimensionality of the Monte Carlo approach to computing the cost function, we believe that the resulting large variability from the MC approach is likely to overshadow any performance difference due to the chosen metamodel. This further motivates the task of computing the cost function exactly under dependent inputs (see Section 5).

4.3 Application to climate simulator

Here we estimate Shapley effects from data generated from the En-ROADS climate simulator (Climate Interactive et al., 2020). This simulator is a mathematical model of how global temperature is influenced by changes in energy, land use, consumption, agriculture, and other factors. It is designed to be easily used by the general public. The model is an ordinary differential

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

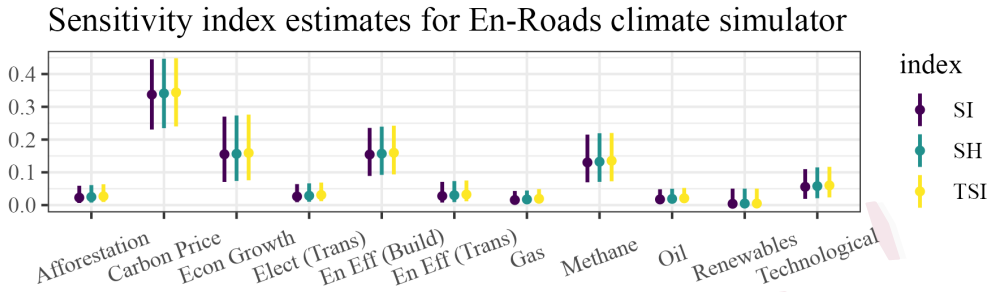

Figure 9: 95% credible intervals (as computed in (3.9)) for (normalized) first-order Sobol´ indices (SI), Shapley effects (SH), and total-effect Sobol´ indices (TSI) over 1000 posterior draws from a BART model fit to climate simulator data (Climate Interactive et al., 2020).

equation solved by Euler integration and synthesizes the important drivers of climate in a computationally efficient and easy-to-use web interface.

The data consists of $n = 110$ observations with $p = 11$ inputs and was collected using the scheme described in Horiguchi et al. (2021). To this data we fit a BART model and compute Shapley-effect estimates using the implementation in Pratola (2023) with 1000 posterior draws, 200 trees, and the remaining default parameter settings.

Figure 9 shows the estimates for the first-order Sobol´ index, Shapley effect, and total-effect Sobol´ index of the 11 inputs. For each input, the relationship (2.7) between the three indices is shown. As expected given this relationship and the analysis in Horiguchi et al. (2021), the small differences

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

between the Shapley-effect estimates and the two Sobol´-index estimates indicate small interaction effects between any group of inputs. Hence, takeaways about the impact of each input are the same as discussed in Horiguchi et al. (2021). In particular, the four most impactful inputs seem to be carbon price, energy efficiency of buildings, methane, and economic growth.

5. Discussion

This article establishes posterior contraction rates for Sobol´-index and Shapley-effect estimators computed using BART. The proofs of our contraction rates required proving a property similar to Lipschitz continuity for Sobol´ indices and Shapley effects before using recent contraction-rate results that apply to function spaces with heterogeneous smoothness and sparsity in high dimensions and to fixed and random designs. This article also illustrates the computational tractability and performance of BART-based Shapley effects on four different test functions under orthogonal inputs and $p = 500$. Code to fit BART models and compute Sobol´ index and Shapley effect estimates is found in Pratola (2023).

Our theoretical consistency results apply to input distributions that are not orthogonal, and thus uncertainty quantification of the Shapley effects would maintain its validity under such distributions. However, to imple-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

ment our approach under such distributions, we would need to be able to compute the cost function. Specifically, the input distribution would affect the values of the probability measure of the boxes that the BART ensemble partitions the input space into. Under independent inputs, the probability measure is simply the volume (i.e., the Lebesgue measure) of the boxes, which is what we currently have implemented. Current Monte Carlo methods of approximating the cost function require being able to sample from the input distribution. It is more useful (and more challenging) to learn the input distribution based on the observed covariates, and then use this learned distribution to estimate the cost function. For our BART-based approach, this can be possibly achieved by replacing the volume of each hyperrectangle used to compute BART-based Sobol´ indices and Shapley effects with the proportion of observations that fall in each hyperrectangle. Another possible approach would be to incorporate a tree-based density estimation method suitable for higher-dimensional spaces, such as Awaya and Ma (2024a,b). We will reserve this exploration as future work.

Supplementary Material

The online Supplementary Material contains a summary table of metamodel properties, a review of posterior contraction theory, and the preliminaries

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

required to establish the posterior asymptotic results. It further presents the statements and proofs of these asymptotic results, along with detailed proofs of results from the main text. Definitions of the functions used in the experiments described in Section 5 of the main paper are provided, together with additional experiments examining how the metamodels scale with input dimensionality.

Acknowledgements

Akira Horiguchi would like to acknowledge Miheer Dewaskar for insightful discussions. The work of Matthew T. Pratola was supported in part by the National Science Foundation under Agreements DMS-1916231 and OAC-2004601, and in part by the Office of Sponsored Research (OSR) at the King Abdullah University of Science and Technology (KAUST) under Award No. OSR-2018-CRG7-3800.3.

References

Awaya, N. and L. Ma (2024a). Hidden Parkov Pólya trees for high-dimensional distributions. *Journal of the American Statistical Association 119*(545), 189–201.

Awaya, N. and L. Ma (2024b). Unsupervised tree boosting for learning probability distributions. *Journal of Machine Learning Research 25*(198), 1–52.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

REFERENCES

Benoumechiara, N. and K. Elie-Dit-Cosaque (2019). Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: Proceedings and Surveys 65*, 266–293.

Broto, B., F. Bachoc, and M. Depecker (2020). Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. *SIAM/ASA Journal on Uncertainty Quantification 8*(2), 693–716.

Castro, J., D. Gómez, and J. Tejada (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & operations research 36*(5), 1726–1730.

Chen, W., R. Jin, and A. Sudjianto (2005). Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. *Journal of mechanical design 127*(5), 875–886.

Chen, W., R. Jin, and A. Sudjianto (2006). Analytical global sensitivity analysis and uncertainty propagation for robust design. *Journal of quality technology 38*(4), 333–348.

Chipman, H., P. Ranjan, and W. Wang (2012). Sequential design for computer experiments with a flexible Bayesian additive model. *Canadian Journal of Statistics 40*(4), 663–678.

Chipman, H. A., E. I. George, and R. E. McCulloch (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics 4*(1), 266–298.

Climate Interactive, Ventana Systems, Todd Fincannon, UML Climate Change Initiative, and MIT Sloan (2020). En_ROADS climate change solutions simulator. `https://en-roads.climateinteractive.org/scenario.html?v=2.7.15`. Accessed: 2020-04-03.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

REFERENCES

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.

Denison, D. G., B. K. Mallick, and A. F. Smith (1998). Bayesian MARS. *Statistics and Computing 8*, 337–346.

Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics 23*(2), 277–281.

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.

Goda, T. (2021). A simple algorithm for global sensitivity analysis with Shapley effects. *Reliability Engineering & System Safety 213*, 107702.

Gramacy, R. B. and B. Haaland (2016). Speeding up neighborhood search in local Gaussian process prediction. *Technometrics 58*(3), 294–303.

Gramacy, R. B. and M. Taddy (2010). Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models. *Journal of Statistical Software 33*(6), 1–48.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

REFERENCES

Gramacy, R. B., M. Taddy, S. M. Wild, et al. (2013). Variable selection and sensitivity analysis using dynamic trees, with an application to computer code performance tuning. *The Annals of Applied Statistics 7*(1), 51–80.

Horiguchi, A. (2020). *Bayesian Additive Regression Trees: Sensitivity Analysis and Multiobjective Optimization*. Ph. D. thesis, The Ohio State University. `http://rave.ohiolink.edu/etdc/view?acc_num=osu1606841319315633`.

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

Horiguchi, A., T. J. Santner, Y. Sun, and M. T. Pratola (2022). Using BART to perform pareto optimization and quantify its uncertainties. *Technometrics 64*(4), 1–11.

Iooss, B., S. D. Veiga, A. Janon, and G. Pujol (2023). *sensitivity: Global Sensitivity Analysis of Model Outputs*. R package version 1.28.1.

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.

Li, S., B. Yang, and F. Qi (2016). Accelerate global sensitivity analysis using artificial neural network algorithm: Case studies for combustion kinetic model. *Combustion and Flame 168*, 53–64.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

REFERENCES

Liu, Y., V. Ročková, and Y. Wang (2021, 04). Variable selection with ABC Bayesian forests. *Journal of the Royal Statistical Society Series B: Statistical Methodology 83*(3), 453–481.

Marrel, A., B. Iooss, B. Laurent, and O. Roustant (2009). Calculations of Sobol´ indices for the Gaussian process metamodel. *Reliability Engineering & System Safety 94*(3), 742–751.

Moon, H. (2010). *Design and analysis of computer experiments for screening input variables*. Ph. D. thesis, The Ohio State University. `http://rave.ohiolink.edu/etdc/view?acc_num=osu1275422248`.

Oakley, J. E. and A. O'Hagan (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(3), 751–769.

Owen, A. B. (2014). Sobol´ indices and Shapley value. *SIAM/ASA Journal on Uncertainty Quantification 2*(1), 245–251.

Plischke, E., G. Rabitti, and E. Borgonovo (2021). Computing Shapley effects for sensitivity analysis. *SIAM/ASA Journal on Uncertainty Quantification 9*(4), 1411–1437.

Pratola, M. T. (2023, 4). Open Bayesian trees. `https://bitbucket.org/mpratola/openbt/src/master/`. Accessed: 2023-04-03.

Pratola, M. T., H. A. Chipman, J. R. Gattiker, D. M. Higdon, R. McCulloch, and W. N. Rust (2014). Parallel Bayesian additive regression trees. *Journal of Computational and Graphical Statistics 23*(3), 830–852.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

REFERENCES

Radaideh, M. I. and T. Kozlowski (2020). Surrogate modeling of advanced computer simulations using deep Gaussian processes. *Reliability Engineering & System Safety 195*, 106731.

Santner, T. J., B. J. Williams, and W. I. Notz (2018). *The Design and Analysis of Computer Experiments, Second Edition*. Springer-Verlag.

Shapley, L. S. (1952). A value for n-person games. Technical report, The RAND Corporation. `https://www.rand.org/content/dam/rand/pubs/papers/2021/P295.pdf`.

Sobol´, I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe modelirovanie 2*(1), 112–118.

Sobol´, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *MMCE 1*(4), 407–414.

Song, E., B. L. Nelson, and J. Staum (2016). Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification 4*(1), 1060–1083.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety 93*(7), 964–979.

Svenson, J., T. Santner, A. Dean, and H. Moon (2014). Estimating sensitivity indices based on Gaussian process metamodels with compactly supported correlation functions. *Journal of Statistical Planning and Inference 144*, 160–172.

Tang, Y. (2024). A note on Monte Carlo integration in high dimensions. *The American Statis-*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

tician 78(3), 290–296.

van Campen, T., H. Hamers, B. Husslage, and R. Lindelauf (2018). A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack. *Social Network Analysis and Mining 8*, 1–12.

van der Pas, S. and V. Ročková (2017). Bayesian dyadic trees and histograms for regression. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 2089–2099. Curran Associates, Inc.

Wu, Z., D. Wang, P. Okolo, F. Hu, and W. Zhang (2016). Global sensitivity analysis using a Gaussian radial basis function metamodel. *Reliability Engineering & System Safety 154*, 171–179.

Yang, L., Y. Zhou, H. Fu, M.-Q. Liu, and W. Zheng (2024). Fast approximation of the Shapley values based on order-of-addition experimental designs. *Journal of the American Statistical Association 119*(547), 2294–2304.

Zhang, X. and N. Dimitrov (2024). Variable importance analysis of wind turbine extreme responses with Shapley value explanation. *Renewable Energy 232*, 121049.

Department of Statistics, University of California, Davis

E-mail: ahoriguchi@ucdavis.edu

Department of Statistics, Indiana University Bloomington

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0318

E-mail: mpratola@iu.edu