Reliability Engineering and System Safety 207 (2021) 107391

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Assessing variable activity for Bayesian regression trees

Akira Horiguchi *, Matthew T. Pratola, Thomas J. Santner

The Ohio State University, Cockins Hall, 1958 Neil Ave., Columbus, OH 43210, USA

ARTICLE INFO

Keywords:
Bayesian Additive Regression Trees
Global sensitivity analysis
Sobol´ indices
Nonparametric
Variable importance
Variable activity

ABSTRACT

Bayesian Additive Regression Trees (BART) are non-parametric models that can capture complex exogenous variable effects. In any regression problem, it is often of interest to learn which variables are most active. Variable activity in BART is usually measured by counting the number of times a tree splits for each variable. Such one-way counts have the advantage of fast computations. Despite their convenience, one-way counts have several issues. They are statistically unjustified, cannot distinguish between main effects and interaction effects, and become inflated when measuring interaction effects. An alternative method well-established in the literature is Sobol´ indices, a variance-based global sensitivity analysis technique. However, these indices often require Monte Carlo integration, which can be computationally expensive. This paper provides analytic expressions for Sobol´ indices for BART posterior samples. These expressions are easy to interpret and are computationally feasible. Furthermore, we will show a fascinating connection between first-order (main-effects) Sobol´ indices and one-way counts. We also introduce a novel ranking method, and use this to demonstrate that the proposed indices preserve the Sobol´-based rank order of variable importance. Finally, we compare these methods using analytic test functions and the En-ROADS climate impacts simulator.

1. Introduction

Bayesian Additive Regression Trees (BART) have become an increasingly popular tool for complex regression problems and as emulators of expensive computer simulations [1–3]. BART sidesteps the $O(n^3)$ matrix decompositions required by arguably the most popular statistical regression tool, Gaussian processes (GPs) [4]. These cubic matrix operations pose issues whose severity continues to grow in the era of big data. BART, like GPs, can capture complex exogenous variable effects without having to specify their functional forms.

To assess the activity of these exogenous input variables, BART offers a variable count heuristic proposed by Chipman et al. [1], which comes nearly for free once a BART model is fit. This method counts the number of times a variable is included in BART's trees as a split variable. For example, the tree in Fig. 1a splits on x_1 twice and on x_2 once. Using this heuristic, input x_1 would be considered to be twice as active as input x_2. The idea is that if many nodes in BART's trees split on a variable, then that variable is deemed important in predicting the response. To this day, count-based methods remain the most popular way of assessing input activity in BART. For example, Bleich et al. [5] also rely on these posterior inclusion proportions in their proposed variable selection methods.

But as Liu et al. [6] note, one-way counts are not theoretically well-understood. Furthermore, their ability to adequately capture even the order of input importance is suspect. Fig. 2 shows the variable counts of 1000 posterior samples from a BART model trained in data generated from the function $f(\mathbf{x}) = (x_1 - 0.5)(x_2 - 0.5) + 0.5(x_3 - 0.5)$ on the unit hypercube $[0,1]^3$. Marginally, variables x_1 and x_2 have zero effect on $f(\cdot)$, which makes variable x_3 marginally the most important input. But the variable counts in Fig. 2 show x_1 and x_2 to be more active than x_3. Thus, the individual marginal counts seem to conflate the interaction effect between x_1 and x_2 with their marginal effects.

To better assess input activity, we may instead use the variance-based global sensitivity analysis method introduced by Sobol´ [7]. He showed that if $f(\mathbf{x})$ is a real-valued, square-integrable function on $[0,1]^p$ then $f(\mathbf{x})$ can be decomposed into a sum

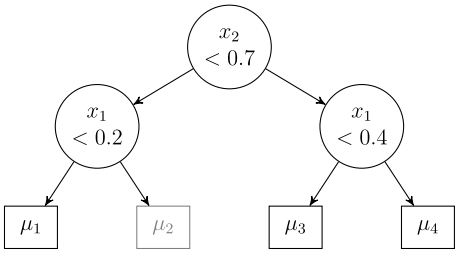
$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i=1}^{d}\sum_{j<i} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\ldots,p}(x_1, x_2, \ldots, x_p),$$

where each summand depends on a subset of $\mathbf{x}$. Assume that the relative frequency with which the inputs of $f(\mathbf{x})$ occur can be modeled by $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ where $X_1, \ldots, X_p \overset{iid}{\sim} U(0,1)$. Then if the variance of the ith term in the Sobol´ expansion which depends on x_i is large, then x_i is deemed important in predicting the response. Computing these variances and expectations requires Monte Carlo integration when $f(\mathbf{x})$ is not known in closed form, and hence becomes untractable as the number of inputs increases.

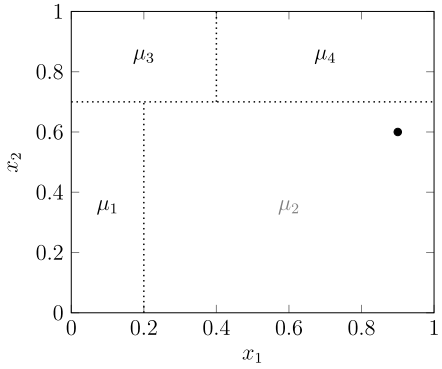
* Corresponding author.
E-mail addresses: horiguchi.6@osu.edu (A. Horiguchi), mpratola@stat.osu.edu (M.T. Pratola), santner.1@osu.edu (T.J. Santner).

https://doi.org/10.1016/j.ress.2020.107391

Available online 7 December 2020
0951-8320/© 2020 Elsevier Ltd. All rights reserved.



(a) Node view of the example tree. Input $\mathbf{x}^* = (0.9, 0.6)$ falls into the gray terminal node.



(b) Level-set view of the example tree. Input $\mathbf{x}^* = (0.9, 0.6)$ is shown as the black point.

Fig. 1. Two different views of the same example tree.

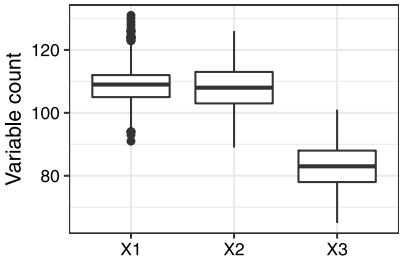


Fig. 2. One-way variable counts of 1000 posterior samples from a BART ensemble trained in data generated from the function $f(\mathbf{x}) = (x_1 - 0.5)(x_2 - 0.5) + 0.5(x_3 - 0.5)$ on the unit hypercube $[0, 1]^3$.

Sobol´ indices have been estimated or computed for various types of metamodels. These approaches can be divided into two groups depending on whether they assume the inputs are uncorrelated. For uncorrelated inputs, Chapter 7 of [4] provides an overview of GP-based Sobol´ indices and their formulae for GPs with certain mean and correlation structures [see also8–13]. Sudret [14] provide analytic expressions for polynomial-chaos-based Sobol´ indices, which reduces the computational burden to obtaining the desired polynomial-chaos coefficients. Finally, Gramacy et al. [15] and Gramacy and Taddy [16] suggest using Sobol´ indices for Dynamic Trees and Treed Gaussian Processes, which use integration approximations via Latin hypercube designs to compute these index estimates. For possibly correlated inputs, Da Veiga et al. [17] compute first-order sensitivity indices of local polynomial smoothers and provide theoretical asymptotic properties for the indices. Wei et al. [18], motivated by high-dimensional input spaces, show that the random-forest-based permutation variable importance measure converges in n to twice the unnormalized total-effects Sobol´ index. The methods from [17,18] avoid numerical integration, but do not compute interactions between specific input variables.

Our primary contribution is to use Sobol´ indices for BART model-based input activity. For a given BART MCMC draw, we derive analytic expressions that can be computed exactly for interactions of any order and do not require Monte Carlo integration, which can be expensive when the number of input variables is large. We furthermore establish a connection between first-order (main-effects) Sobol´ indices and one-way counts. Finally, we compare the methods using analytic test functions and demonstrate that Sobol´ indices applied to BART accurately capture true variable effects while remaining computationally attractive and easy to interpret. To perform this comparison, we consider both the estimation of the Sobol´ indices and evaluate the order-preserving sequence of active variables using a proposed novel rank-order statistic.

The rest of the paper is organized as follows. In Section 2, we review BART. In Section 3, we derive Sobol´ indices for BART and establish a connection between first-order Sobol´ indices and one-way counts. In Section 4, we provide computational details. In Section 5, we introduce our rank-order statistic, perform simulation studies, and apply Sobol´ indices to a BART-based emulator of the En-ROADS climate simulator. In Section 6, we conclude the paper with a discussion. Proofs of stated theorems can be found in Appendix A.

2. Review of BART

We wish to make inference on an unknown function $f : D \to \mathbb{R}$, where domain D is a p-dimensional subset of $\mathbb{R}^p$. We will assume for the rest of the text that domain D is a bounded hyperrectangle, i.e. $D = \prod_{j=1}^{p} I_D^j = \prod_{j=1}^{p} [a_D^j, b_D^j]$, where I_D^j is the jth marginal interval of D for $j = 1, \ldots, p$. We observe the data $\mathscr{D} := \{(y(\mathbf{x}_i), \mathbf{x}_i)\}_{i=1,\ldots,n}$, where each observation $y(\mathbf{x})$, based on predictor $\mathbf{x} = (x_1, \ldots, x_p) \in D$, is assumed to be a realization of the random variable

$$Y(\mathbf{x}) = f(\mathbf{x}) + \epsilon, \tag{1}$$

where $\epsilon \overset{iid}{\sim} N(0, \sigma^2)$.

To make inference about the unknown function $f(\cdot)$, we approximate it by a sum of m regression trees. That is, we make the approximation

$$f(\mathbf{x}) \approx \sum_{t=1}^{m} g(\mathbf{x}; \mathcal{T}_t, \mathcal{M}_t), \tag{2}$$

where each $g(\cdot; \mathcal{T}_t, \mathcal{M}_t) : D \to \mathbb{R}$ denotes a regression tree function and the parameters $\{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^{m}$ are given a prior distribution in BART's hierarchical Bayesian model structure. Each $g(\cdot; \mathcal{T}_t, \mathcal{M}_t)$ contributes a small portion to the total approximation of $f(\cdot)$. Hence, the expected response $\mathbb{E}\left[Y(\mathbf{x}) \mid \{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^{m}\right]$ at a given input $\mathbf{x}$ is the sum of each of the contributions $g(\mathbf{x}; \mathcal{T}_t, \mathcal{M}_t)$. The model in Eq. (2) is called a sum-of-trees model.

2.1. Single-tree model

To explain the sum-of-trees model, we will first set the number of trees $m = 1$ and describe the notation of the resulting single-tree model.

The single-tree model is the Bayesian implementation of the Classification and Regression Tree (CART) model as proposed in [19]. CART can be used for classification, but we assume for the paper that it is being applied to the regression setting as described in Eq. (1). The CART model provides a prediction of $f(\mathbf{x})$ at input point $\mathbf{x}$ given observed data $\mathscr{D}$. CART partitions the input space and fits a constant mean model in

each subregion to form the predictions. CART constructs the partition via a binary tree structure. To form the partitions, each internal node contains a boolean split rule. Starting at the root node, if an input point $\mathbf{x}$ satisfies the split rule, it will travel to the node's left child; otherwise $\mathbf{x}$ will travel to the right child. The input point $\mathbf{x}$ will continue to traverse through the tree in this way until it reaches a terminal node. This terminal node's parameter is the predicted value of $f(x)$.

Fig. 1 shows an illustrative example. Suppose the tree in Fig. 1a is used in a single-tree model to predict an output value for input $\mathbf{x}^* = (x_1^*, x_2^*) = (0.9, 0.6)$, where the input space D is the closed unit-square $[0,1]^2$. Starting at the root node in Fig. 1a, we see that $\mathbf{x}^*$ satisfies this split rule (i.e. $x_2^* < 0.7$), which moves $\mathbf{x}^*$ to the left child. We then see that $\mathbf{x}^*$ does not satisfy this split rule (i.e. $x_1^* \geq 0.2$), which moves $\mathbf{x}^*$ to the right child, which turns out to be a terminal node. Because we are using a single-tree model (i.e. there is exactly $m = 1$ tree), this mean parameter μ_2 becomes the predicted value for input $\mathbf{x}^*$. Fig. 1b shows the corresponding hyperrectangle view of the tree.

A tree's parameters can now be organized in the following manner. Let $\mathcal{T}$ denote the set of parameters associated with the tree's split rules (i.e. the split variable and cutpoint for each internal node) and topology. Let $\mathcal{M}$ denote the set $\{\mu_k\}$ of parameters associated with the tree's terminal nodes. The single-tree model is thus $f(\cdot) \approx g(\cdot; \mathcal{T}, \mathcal{M})$, where $f(\cdot)$, defined in Eq. (1), is the mean of the observed process. Here, we think of $g(\cdot; \mathcal{T}, \mathcal{M})$ being a function that assigns a value μ_k to input $\mathbf{x}$ according to the parameters in $\mathcal{T}$ and $\mathcal{M}$. Let $\mathbf{R}_k \subset D$ denote the hyperrectangle associated with the tree's terminal node that contains parameter μ_k. Then,

$$g(\cdot; \mathcal{T}, \mathcal{M}) = \sum_{k=1}^{|\mathcal{M}|} \mu_k \mathbf{1}_{\mathbf{R}_k}(\cdot). \tag{3}$$

We may further decompose each hyperrectangle $\mathbf{R}_k$ into the Cartesian product of its p marginal intervals $I_k^1, \ldots, I_k^p$ and hence write $\mathbf{1}_{\mathbf{R}_k}(\mathbf{x}) = \prod_{i=1}^p \mathbf{1}_{I_k^i}(x_i)$.

2.2. Sum-of-trees model

Now consider the sum-of-trees model in Eq. (2) for $m > 1$. If the parameter sets $\{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^m$ have been established, we will let the function

$$\mathcal{E}(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) := \sum_{t=1}^m g(\cdot; \mathcal{T}_t, \mathcal{M}_t) = \sum_{t=1}^m \sum_{k=1}^{|\mathcal{M}_t|} \mu_{tk} \mathbf{1}_{\mathbf{R}_{tk}}(\cdot) \tag{4}$$

denote the sum-of-trees approximation in Eq. (2). To streamline notation, we will refer to $\mathcal{E}$ as both the function $\mathcal{E}(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m)$ and as the collection $\{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^m$. Thus, we write $(\mathcal{T}, \mathcal{M}) \in \mathcal{E}$ if $(\mathcal{T}, \mathcal{M}) = (\mathcal{T}_t, \mathcal{M}_t)$ for some $t = 1, \ldots, m$.

2.3. BayesIan tree models

The sum-of-trees model is specified by the parameters $\{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^m$ and σ^2. Hence, a trained BART model will sample from the posterior distribution

$$\pi(\Theta \mid \mathcal{D}) \propto L(\Theta \mid \mathcal{D})\, \pi(\Theta), \tag{5}$$

where $\Theta = \{(\mathcal{T}_1, \mathcal{M}_1), (\mathcal{T}_2, \mathcal{M}_2), \ldots, (\mathcal{T}_m, \mathcal{M}_m), \sigma^2\}$ are the parameters, $\mathcal{D}$ is the observed data,

$$L(\Theta \mid \mathcal{D}) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y(\mathbf{x}_i) - \sum_{t=1}^m g(\mathbf{x}_i; \mathcal{T}_t, \mathcal{M}_t) \right)^2 \right)$$

is the likelihood, and $\pi(\Theta)$ is the prior.

Chipman et al. [1] specify the full prior $\pi(\Theta)$ by constraining it to satisfy independence conditions

$$\pi(\Theta) = \left[\prod_{t=1}^m \pi(\mathcal{M}_t \mid \mathcal{T}_t) \pi(\mathcal{T}_t) \right] \pi(\sigma^2), \tag{6}$$

and

$$\pi(\mathcal{M}_t \mid \mathcal{T}_t) = \prod_{k=1}^{|\mathcal{M}_t|} \pi(\mu_{tk} \mid \mathcal{T}_t) \tag{7}$$

for all $t = 1, \ldots, m$. In Eq. (6), the parameter sets $(\mathcal{T}_t, \mathcal{M}_t)$ and σ^2 are constrained to be mutually independent. In Eq. (7), the terminal node parameters of every tree are constrained to be independent. These independence conditions simplify the problem of specifying the full prior $\pi(\Theta)$ to specifying only the priors $\pi(\mathcal{T}_t)$, $\pi(\mu_{tk} \mid \mathcal{T}_t)$, and $\pi(\sigma^2)$. Forcing the priors $\pi(\mathcal{T}_t)$ and $\pi(\mu_{tk} \mid \mathcal{T}_t)$ to be identical for all $k = 1, \ldots, |\mathcal{M}_t|$ and $t = 1, \ldots, m$ further simplifies the prior specification problem. Furthermore, Chipman et al. [19] choose the three prior forms to simplify analysis and computation by taking advantage of known conjugacy pairs. In particular, they choose the $\pi(\mu_{tk} \mid \mathcal{T}_t)$ prior to be a conjugate Normal distribution. To configure the priors, Chipman et al. [1] recommend automatically specifying the relevant hyperparameters using data-driven methods.

The posterior in Eq. (5) can thus be sampled using the following Gibbs sampler:

1. Draw $\{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^m \mid \sigma^2, \mathcal{D}$.
2. Draw $\sigma^2 \mid \{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^m, \mathcal{D}$.

For Step 2, we can draw $\sigma^2 \mid \{(\mathcal{T}_t, \mathcal{M}_t)\}_{t=1}^m, \mathcal{D}$ by performing a simple conjugate Gibbs step. Step 1 itself will also be a Gibbs sampler that relies on being able to sample from the conditional distribution

$$\pi(\mathcal{T}_t, \mathcal{M}_t \mid \{(\mathcal{T}_\tau, \mathcal{M}_\tau)\}_{\tau \neq t}, \sigma^2, \mathcal{D}) \tag{8}$$

for all $t = 1, \ldots, m$. To sample from this conditional distribution, we simplify the likelihood by noting

$$L(\Theta \mid \mathcal{D}) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(r_t(\mathbf{x}_i) - g(\mathbf{x}_i; \mathcal{T}_t, \mathcal{M}_t) \right)^2 \right)$$

where $r_t(\mathbf{x}_i) := y(\mathbf{x}_i) - \sum_{\tau \neq t} g(\mathbf{x}_i; \mathcal{T}_\tau, \mathcal{M}_\tau)$. Therefore, the conditional distribution in Eq. (8) for any $t = 1, \ldots, m$ relies on $\{(\mathcal{T}_\tau, \mathcal{M}_\tau)\}_{\tau \neq t}$ and $\mathcal{D}$ only through $\mathbf{R}_t = \{(r_t(\mathbf{x}_i), \mathbf{x}_i)\}_{i=1,\ldots,n}$. Hence, the conditional distribution can be expressed as $\pi(\mathcal{T}_t, \mathcal{M}_t \mid \mathbf{R}_t, \sigma^2)$, where $\mathbf{R}_t$ plays the role of $\mathcal{D}$ in the single-tree version of Step 1 of the Gibbs sampler. Each draw from the conditional distribution in Eq. (8) for any $t = 1, \ldots, m$ is then reduced to two draws:

(a) Draw $\mathcal{T}_t \mid \sigma^2, \mathbf{R}_t$.
(b) Draw $\mathcal{M}_t \mid \mathcal{T}_t, \sigma^2, \mathbf{R}_t$.

3. Sobol´ indices

In Section 1, we introduced the idea from [7] that the variance of any real-valued function defined on and square-integrable in a unit-hypercube domain can be decomposed into a sum of variance terms. The Sobol´ [7] results apply when inputs $X_1, X_2, \ldots, X_p$ are continuous and mutually uncorrelated with finite interval supports. Thus, using Eq. (4), we can decompose the variance of a BART ensemble function into a sum of terms attributed to single inputs or to interactions between sets of inputs.

To develop our BART-based Sobol´ indices, we will require the following assumptions:

A.1 $X_1, \ldots, X_p$ are mutually uncorrelated;

A.2 X_i's density π_i is positive almost everywhere on the domain's ith margin;

A.3 Conditional on parameter sets $\{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m$, the relation $\mathcal{E}(\mathbf{x}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) = \mathcal{E}(\mathbf{x}^*; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m)$ holds if and only if input points $\mathbf{x}$ and $\mathbf{x}^*$ belong to the same set of m terminal nodes.

We use conditions A.1 and A.2 to extend the two original results from [7] and to derive Sobol´ indices for BART ensembles. Condition A.3 follows from each $\mu_{tk} | \mathcal{T}_t$ being conditionally Normal. If inputs $\mathbf{x}$ and

$\mathbf{x}^*$ belong to different terminal nodes in some tree $\mathcal{T}_\tau$, then their corresponding terminal node parameters almost surely have different values, which results in $g(\mathbf{x}; \mathcal{T}_\tau, \mathcal{M}_\tau) \neq g(\mathbf{x}^*; \mathcal{T}_\tau, \mathcal{M}_\tau)$. Therefore, condition A.3 is a reasonable assumption to make. We also note that condition A.3 is used only when relating Sobol´ indices to counts and does not affect the computation of Sobol´ indices for BART ensembles.

We can now state the desired generalized version of the variance decomposition described in [7]. For any random vector $\mathbf{X} = (X_1, \ldots, X_p)$ that satisfies conditions A.1 and A.2 on p-dimensional bounded hyperrectangle domain D and for any real-valued function f square-integrable on D, 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. That is,

$$\mathtt{Var}_{\mathbf{X}}\big(f(\mathbf{X})\big) = \sum_{i=1}^{p} V_i + \sum_{i=1}^{p}\sum_{i<j} V_{ij} + \cdots + V_{12\ldots p}, \tag{9}$$

where we recursively define for each variable index set $P \subseteq \{1, 2, \ldots, p\}$

$$V_P := \mathtt{Var}_{\mathbf{X}_P}\Big(\mathbb{E}_{\mathbf{X}_{-P}}[f(\mathbf{X}) \mid \mathbf{X}_P]\Big) - \sum_{Q \subset P} V_Q \tag{10}$$

where the sum is over all nonempty, proper subsets Q of P. In particular, the (unnormalized) first-order Sobol´ index V_i (i.e. V_P when $P = \{i\}$) is

$$V_i := \mathtt{Var}_{X_i}(\mathbb{E}_{\mathbf{X}_{-i}}[f(\mathbf{X}) \mid X_i])$$

for all $i = 1, \ldots, p$. Also, the (unnormalized) second-order Sobol´ index V_{ij} is

$$V_{ij} := \mathtt{Var}_{\mathbf{X}_{ij}}(\mathbb{E}_{\mathbf{X}_{-ij}}[f(\mathbf{X}) \mid \mathbf{X}_{ij}]) - V_i - V_j$$

for all $i \neq j$. Often, each V_P term is divided by the total variance $\mathtt{Var}_{\mathbf{X}}(f(\mathbf{X}))$ to produce normalized Sobol´ index

$$S_P := V_P / \mathtt{Var}_{\mathbf{X}}(f(\mathbf{X})). \tag{11}$$

By Eq. (9), the sum of all normalized indices equals unity. This allows comparison between indices of different functions.

3.1. Sobol´ indices applied to BART

Next, we apply this variance decomposition for general L_2 functions $f(\cdot)$ to BART ensemble functions $\mathcal{E}(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m)$. That is, we will compute the terms in the right hand side of Eq. (9) for BART ensembles.

The core terms to compute in Eq. (10) are the conditional expectation $\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]$ and its variance with respect to $\mathbf{X}_P$. By integrating both sides of Eq. (4), we obtain an analytic expression for the conditional expectation:

$$\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P] = \sum_{k \in B_\mathcal{E}} d_k^{-P} \mathbf{1}_{\mathbf{R}_k^P}(\mathbf{X}_P), \tag{12}$$

where the set $B_\mathcal{E}$ indexes the terminal nodes of ensemble $\mathcal{E}$, the $|P|$-dimensional hyperrectangle $\mathbf{R}_k^P$ is the projection of terminal node k's p-dimensional hyperrectangle $\mathbf{R}_k$ onto the dimensions in P, and $d_k^{-P} = \mu_k \mathbb{P}_{-P}(\mathbf{R}_k^{-P})$, where we introduce the notation $\mathbb{P}_P(\cdot) = \mathbb{P}_{\mathbf{X}_P}(\cdot) = \mathbb{P}(\mathbf{X}_P \in \cdot)$. Theorem 1 then provides an analytic expression for the variance of the conditional expectation.

Theorem 1. *For any random vector $\mathbf{X} = (X_1, \ldots, X_p)$ that satisfies conditions A.1 and A.2 on a p-dimensional bounded hyperrectangle domain D, the variance of the conditional expectation in Eq. (12) with respect to variable index set P is*

$$\mathtt{Var}_{\mathbf{X}_P}\Big(\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]\Big) = \sum_{k \in B_\mathcal{E}} \sum_{l \in B_\mathcal{E}} d_k^{-P} d_l^{-P} C_{k,l}^P \tag{13}$$

where $d_k^{-P} = \mu_k \mathbb{P}_{-P}(\mathbf{R}_k^{-P})$ and $C_{k,l}^P = \mathbb{P}_P(\mathbf{R}_k^P \cap \mathbf{R}_l^P) - \mathbb{P}_P(\mathbf{R}_k^P)\mathbb{P}_P(\mathbf{R}_l^P)$ (if $P = \{1, \ldots, p\}$, then $d_k^{-P} = \mu_k$). In particular, the (unnormalized) first-order Sobol´ index V_i is

$$V_i = \sum_{k \in B_\mathcal{E}} \sum_{l \in B_\mathcal{E}} d_k^{-i} d_l^{-i} C_{k,l}^i,$$

where $d_k^{-i} = \mu_k \prod_{j \neq i} \mathbb{P}_j(I_k^j)$ and $C_{k,l}^i = \mathbb{P}_i(I_k^i \cap I_l^i) - \mathbb{P}_i(I_k^i)\mathbb{P}_i(I_l^i)$.

3.2. How do counts and Sobol´ indices relate?

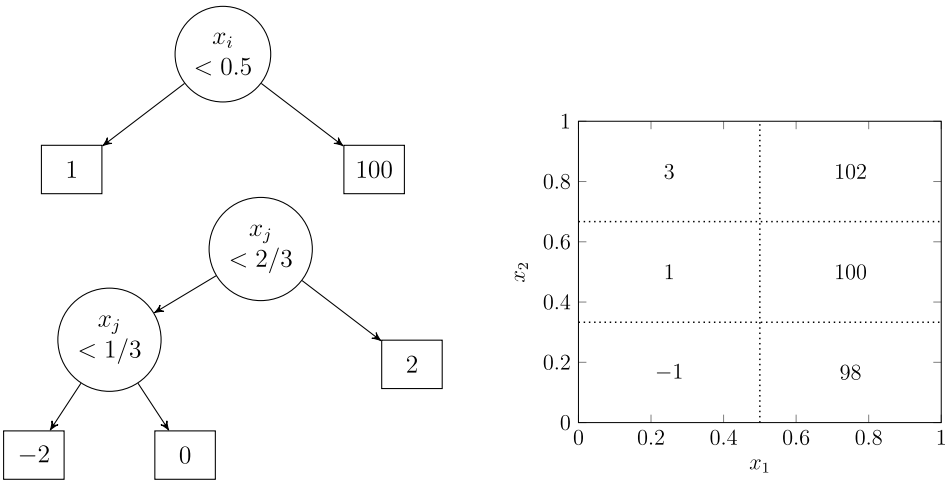
To this day, count-based methods remain the most popular ways of assessing input activity for BART. But as Liu et al. [6] note, they are not theoretically well-understood. We have seen in Fig. 2 a scenario in which the one-way count metric not only inaccurately measures input activity in the data-generating function f but also incorrectly ranks the variables in order of importance. Chipman et al. [1] and Bleich et al. [5] also detail scenarios that question how accurately one-way count metric assess input activity in the data-generating function and suggest ad-hoc work-arounds, such as fitting BART with small m to get an empirically better behaved estimate of input activity. But how do counts perform when assessing input activity in the BART ensemble itself? To answer this question, we turn to the example in Fig. 3a. The count metric will look at number of splits and conclude that variable x_j is twice as active than variable x_i. But if we look at the terminal node values of the ensemble, variable x_i is clearly more important than variable x_j in determining the ensemble's predicted value. If the count metric is not measuring variable importance in the ensemble, then what exactly does it measure? Theorem 2 answers this question.

Theorem 2. *Let $\mathcal{E}$ be a BART ensemble of m regression trees with parameters $\{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m$. Assume $\mathcal{E}$ satisfies assumptions A.1, A.2, and A.3, and fix $i \in \{1, \ldots, p\}$. Then the number of unique split rules in $\mathcal{E}$ that involve variable x_i equals the number of jumps in the piecewise-constant function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$.*

To see why Theorem 2 might be true, consider a BART ensemble $\mathcal{E}_0$ with m regression trees, where each tree is simply a terminal node. The ensemble $\mathcal{E}_0$, which predicts the same value for any input $\mathbf{x} \in D$, can turn into any m−tree BART ensemble $\mathcal{E}$ by undertaking an appropriate sequence of birth processes. Any birth process slices a terminal node's corresponding hyperrectangle into two smaller hyperrectangles according to some split rule. If we call this split rule "$x_i < c$", then this slice occurs on the $(p-1)$-dimensional hyperplane $x_i = c$ in D. The resulting "left" ("right") hyperrectangle gains a terminal node parameter μ_{left} (parameter μ_{right}), where $\mu_{left} \neq \mu_{right}$ through assumption A.3. If the split rule "$x_i < c$" did not already exist in the ensemble, then the piecewise-constant function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$ must have been constant at $x_i = c$ prior to the birth process, which means the birth process produces a jump in the piecewise-constant function at $x_i = c$. Meanwhile, no jumps are produced in any of the other piecewise-constant functions $\mathbb{E}_{\mathbf{X}_{-j}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_j = \cdot]$ (where $j \neq i$). Hence, under assumptions A.1, A.2, and A.3, each birth process that produces a unique split rule that involves variable x_i increments the number of jumps in the piecewise-constant function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$ by one.

Theorem 2 also provides a link between the one-way count metric and the theoretically more well-understood first-order Sobol´ index. Under the conditions of Theorem 2, the one-way count of variable x_i is the number of jumps in the conditional expectation function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$. Under the conditions of Theorem 1, the first-order Sobol´ index of variable x_i is the variance of the conditional expectation $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i]$. Thus, under certain conditions, both the one-way count and the first-order Sobol´ index of variable x_i are functions of the conditional expectation function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$.

Interestingly, the number of jumps and variance can each be viewed as a measure of variability. Under this lens, the one-way count metric can been seen as a more crude version of the first-order Sobol´ index. Theorem 3 describes how to "standardize" the conditional expectation function so that its variance becomes the number of jumps of the conditional expectation. We use the term standardize because many different conditional expectation functions can be transformed into the standardized conditional expectation function, but the standardized conditional expectation function cannot be transformed back into the original conditional expectation function.



(a) Ensemble $\mathcal{E}$ of two trees. (b) Level-set view of $\mathcal{E}$.

Fig. 3. Two different views of the same ensemble $\mathcal{E}$.

Theorem 3. *Let $\mathcal{E}$ be a BART ensemble satisfying A.1, A.2, and A.3. Recall that for all dimensions $i = 1, \dots, p$, the conditional expectation function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$ is piecewise constant and hence can be written as $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot] = \sum_{k^* \in B_{\mathcal{E}}^i} e_{k^*}^i \mathbf{1}_{I_{k^*}^i}(\cdot)$, where $B_{\mathcal{E}}^i$ indexes the intervals of this piecewise constant function. Suppose for any indices $k^*, l^* \in B_{\mathcal{E}}^i$ that $e_{k^*}^i = e_{l^*}^i$ implies $I_{k^*}^i = I_{l^*}^i$ (i.e. the piecewise constant function has distinct values in different input regions). Consider the following transformations to this conditional expectation function:*

1. *Center and scale $\{e_{k^*}^i : k^* \in B_{\mathcal{E}}^i\}$ so that the corrected sample variance[1] equals $|B_{\mathcal{E}}^i|$.*
2. *Assign equal probability mass $|B_{\mathcal{E}}^i|^{-1}$ to each $I_{k^*}^i$.*

Then the number of jumps in this transformed conditional expectation function equals its variance.

As with condition A.3, the added "$e_{k^*}^i = e_{l^*}^i$ implies $I_{k^*}^i = I_{l^*}^i$" assumption in Theorem 3 follows from each $\mu_k | \mathcal{T}$ being conditionally Normal. We may use reasoning similar to before to argue that this assumption is also reasonable to make.

4. Computational details

Given a L^2 function f, we wish to estimate its normalized Sobol´ indices S_P^f as defined in Eq. (11) for all variable index sets P. We do so by first training a BART model on data generated from Eq. (1) and drawing N samples $\Theta^{(1)}, \dots, \Theta^{(N)}$ from the resulting BART posterior in Eq. (5). For each variable index set P, we then compute

$$S_P^{\Theta^{(j)}} := \frac{\mathtt{Var}_{\mathbf{X}_P}(\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t^{(j)}, \mathcal{M}_t^{(j)}\}_{t=1}^m) \mid \mathbf{X}_P]) - \sum_{Q \in 2^P \setminus \{\emptyset, P\}} V_Q^{\Theta^{(j)}}}{\mathtt{Var}_{\mathbf{X}}(\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t^{(j)}, \mathcal{M}_t^{(j)}\}_{t=1}^m))}$$

for each posterior draw $\Theta^{(j)}$, where $j = 1, \dots, N$. We can then obtain a point estimate of S_P^f by approximating the integral $\int S_P^\Theta \, d\pi(\Theta \mid \mathcal{D})$ using the sample mean of $S_P^{\Theta^{(1)}}, \dots, S_P^{\Theta^{(N)}}$. That is, our point estimate of S_P^f is

$$\hat{S}_P^f = \frac{1}{N} \sum_{j=1}^N S_P^{\Theta^{(j)}}.$$

[1] The definition we use for the corrected sample variance of real numbers $x_1, \dots, x_n$ is $(n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$, where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ is the sample mean of $x_1, \dots, x_n$.

At the core of these calculations is the variance term

$$\mathtt{Var}_{\mathbf{X}_P}\left(\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]\right)$$

which we showed can be computed exactly using Theorem 1. Furthermore, it turns out that possibly many, if not all, of the summands in Eq. (13) are zero. Theorem 4 below explains under what conditions a summand vanishes.

4.1. Unnormalized Sobol´ indices

A sensible goal in sensitivity analysis is to compute all first-order Sobol´ indices. According to Eq. (13), each unnormalized first-order index V_i for $\mathcal{E}$ is a sum of $|B_{\mathcal{E}}|^2$ terms. Hence, computing all p first-order indices requires calculating $p \times |B_{\mathcal{E}}|^2$ terms. However, we may take advantage of BART's additive structure to reduce the number of terms to compute.

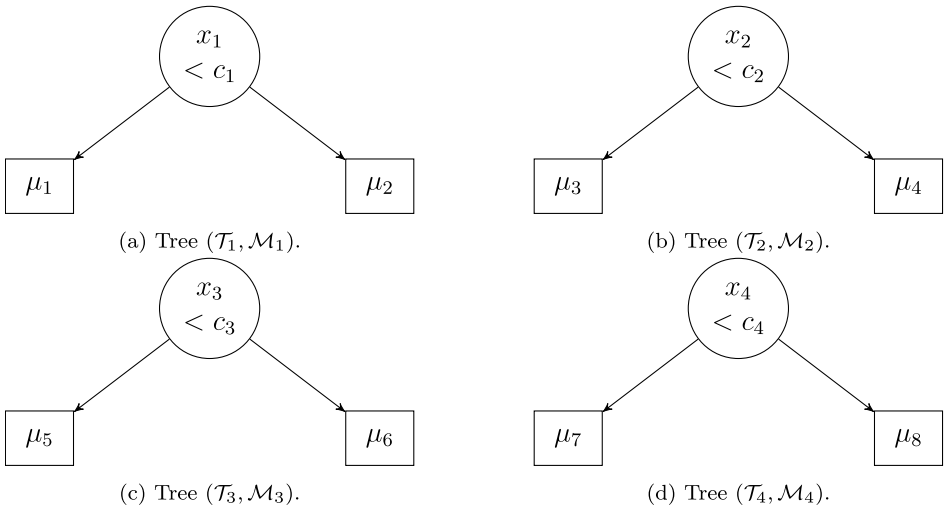
Consider the example ensemble $\mathcal{E}$ consisting only of the $m = 4$ trees in Fig. 4. Because ensemble $\mathcal{E}$ has $|B_{\mathcal{E}}| = 8$ terminal nodes, the unnormalized first-order Sobol´ index V_1 is a sum of $|B_{\mathcal{E}}|^2 = 64$ terms. However, only tree $(\mathcal{T}_1, \mathcal{M}_1)$ splits on variable x_1, which makes the conditional expectations $\mathbb{E}[g(\mathbf{X}; \mathcal{T}_2, \mathcal{M}_2) \mid X_1 = x_1]$, $\mathbb{E}[g(\mathbf{X}; \mathcal{T}_3, \mathcal{M}_3) \mid X_1 = x_1]$, and $\mathbb{E}[g(\mathbf{X}; \mathcal{T}_4, \mathcal{M}_4) \mid X_1 = x_1]$ constant in x_1. Thus,

$$\begin{aligned} V_1 &= \mathtt{Var}_{X_1}\left(\mathbb{E}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid X_1]\right) \\ &= \mathtt{Var}_{X_1}\left(\mathbb{E}\Big[\sum_{t=1}^m g(\mathbf{X}; \mathcal{T}_t, \mathcal{M}_t) \mid X_1\Big]\right) \\ &= \mathtt{Var}_{X_1}\left(\mathbb{E}[g(\mathbf{X}; \mathcal{T}_1, \mathcal{M}_1) \mid X_1]\right), \end{aligned}$$

which, according to Eq. (13), is a sum of only $|\mathcal{M}_1|^2 = 4$ terms. Using the same logic, each Sobol´ index V_2, V_3, and V_4 reduces from a sum of $|B_{\mathcal{E}}|^2 = 64$ terms to a sum of, respectively, $|\mathcal{M}_2|^2 = 4$, $|\mathcal{M}_3|^2 = 4$, and $|\mathcal{M}_4|^2 = 4$ terms. Hence, computing all four indices V_1, V_2, V_3, and V_4 reduces from a sum of $4 \times |B_{\mathcal{E}}|^2 = 256$ terms to a sum of $|\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + |\mathcal{M}_3|^2 + |\mathcal{M}_4|^2 = 16$ terms.

More generally, to compute Eq. (13) for arbitrary variable index set P, we may remove any tree that does not split on any variable in P. Furthermore, we may take advantage of the ensemble function's formulation in Eq. (4) to remove any node whose path to root node does not split on any variable in P. This statement is made precise in Theorem 4.

Theorem 4. *Let $\mathcal{E}(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) = \sum_{k \in B_{\mathcal{E}}} \mu_k \mathbf{1}_{\mathbf{R}_k}(\cdot)$ be the function of a BART ensemble $\mathcal{E}$. For any terminal node η_k, let $v(k)$ be the index set of all*



(a) Tree $(\mathcal{T}_1, \mathcal{M}_1)$. (b) Tree $(\mathcal{T}_2, \mathcal{M}_2)$.

(c) Tree $(\mathcal{T}_3, \mathcal{M}_3)$. (d) Tree $(\mathcal{T}_4, \mathcal{M}_4)$.

Fig. 4. Four trees.

split variables along the path to η_k's root node. For any variable index set P, let

$$\mathcal{E}_P(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) := \sum_{\substack{k \in B_\mathcal{E} \\ v(k) \cap P \neq \emptyset}} \mu_k \mathbf{1}_{\mathbf{R}_k}(\cdot)$$

be the ensemble function that results from removing from $B_\mathcal{E}$ any terminal node whose path to root node does not split on any variable in P. Then

$$\mathtt{Var}_{\mathbf{X}_P}(\mathbb{E}[\mathcal{E}(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]) = \mathtt{Var}_{\mathbf{X}_P}(\mathbb{E}[\mathcal{E}_P(\mathbf{X}; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]).$$

To get a better sense of how much computation Theorem 4 saves, consider again the goal of computing all p (unnormalized) first-order Sobol´ indices if an ensemble $\mathcal{E}$, but for a more realistic scenario: let $p = 10$ and suppose the ensemble $\mathcal{E}$ is such that $N_1 = N_2 = N_3 = N_4 = N_5 = \frac{1}{4}|B_\mathcal{E}|$ and $N_6 = N_7 = N_8 = N_9 = N_{10} = 0$, where N_i is the number of terminal nodes in $\mathcal{E}$ whose path to root node includes split variable x_i. Theorem 4 tells us for all $i = 1, \ldots, p$, the first-order index V_i is a sum of N_i^2 terms. Hence, computing all p first-order Sobol´ indices would require $\sum_{i=1}^p N_i^2 = 5 \times \frac{1}{16}|B_\mathcal{E}|^2$ terms to be computed, which is a 32-fold improvement over $p \times |B_\mathcal{E}|^2$, which is the number if terms to compute if Theorem 4 is not used.

4.2. Total-effects index

In sensitivity analysis, we also often wish to obtain some measure of interaction between the input variables. We can do so via the total-effects sensitivity index, which is defined to be the sum of all normalized sensitivity indices involving the input variable in question [20]. For example, if $p = 3$, then the total-effects index for input variable x_2 would be $T_2 = S_2 + S_{12} + S_{23} + S_{123}$. Hence, $T_2 - S_2$ provides a sense of the magnitude of all interactions involving variable x_2. However, in order to compute all p total-effects sensitivity indices, this formulation requires computing all $2^p - 1$ normalized sensitivity indices S_P. Fortunately, the total-effects index is equivalent to the following expression:

$$T_i = 1 - S_{-i},$$

where $S_{-i} = S_{\{1,\ldots,i-1,i+1,\ldots,p\}}$, and $\sum_{i=1}^p T_i \geq 1$ with equality only if the model is purely additive. Hence, we only need to compute p of these variance expressions in order to compute all p total-effects sensitivity indices.

5. Applications

5.1. Simulation study

Simulation settings discussion.

Given data generated from Eq. (1), where the true Sobol´ index values for $f(\mathbf{x})$ are known, this section identifies the number of inputs, the sample size, and the magnitude of the measurement error standard deviation which answer the following questions:

- **Q.1** What is the bias of BART-based Sobol´ indices for estimating the Sobol´ indices of $f(\cdot)$ when $f(\mathbf{x})$ is measured with error?
- **Q.2** How does this bias compare to the bias of estimated Sobol´ indices obtained using alternative nonparametric prediction models?
- **Q.3** How close are the first-order rankings provided by BART-based Sobol´ indices to the first-order ranking provided by the Sobol´ indices of $f(\cdot)$?
- **Q.4** How close are the first-order rankings provided by one-way BART counts to the first-order ranking provided by the Sobol´ indices of $f(\cdot)$?

For questions **Q.1**, **Q.3**, and **Q.4**, and for each simulation setting below, we generate a n−point maximin LHD sample on $[0,1]^p$ [21]. At each input point $\mathbf{x}$ of the LHD sample, we generate response values for a given data-generating function $f(\cdot): [0,1]^p \to \mathbb{R}$ from Eq. (1) where $\epsilon \sim N(0, \sigma^2)$. We generate 500 data sets for each possible combination of the following different parameter settings: $p/p_0 \in \{1,2,3\}$, $n \in \{10p, 50p\}$, and $\sigma^2 \in \{0.1\mathtt{Var}(f(\mathbf{X})), 0.25\mathtt{Var}(f(\mathbf{X}))\}$, where $(p - p_0)$ is the number of inert variables in $f(\cdot)$ and $\mathbf{X} = (X_1, \ldots, X_p)$, where each $X_i \overset{iid}{\sim} U(0,1)$. For each $f(\cdot)$, the variance $\mathtt{Var}(f(\mathbf{X}))$ is calculated analytically where possible, otherwise numerical integral approximations are used. To each of these 500 data sets, we fit a BART model using the default parameter settings of the Open Bayesian Trees (OpenBT) project found at https://bitbucket.org/mpratola/openbt/ [22].

The five data-generation functions to be examined are:

1. From [23], the data-generating function is defined as

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

   This function, used in [1] and many other BART papers for variable activity and selection, is a challenging mix of interactions and nonlinearities. Here, only $p_0 = 5$ variables influence the response. This function's first-order Sobol´ indices and total-effects Sobol´ indices have the same ordering. That is, $S_4^f >$

$S_1^f = S_2^f > S_3^f > S_5^f$ and $T_4^f > T_1^f = T_2^f > T_3^f > T_5^f$. Also, $\texttt{Var}(f(\mathbf{X})) \approx 23.8$.

2. We modify the Friedman function to create the data-generating function defined as

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

For this modified version, the first-order Sobol´ indices for variables 1 and 2 are zero, which changes the order of the first-order Sobol´ indices while maintaining the total-effects order (see Table 1). Also, $\texttt{Var}(f(\mathbf{X})) \approx 19.0$.

3. The $g-$function from [24] with p_0 inputs is defined to be

$$f(x_1, \ldots, x_{p_0}) = \prod_{k=1}^{p_0} \frac{|4x_k - 2| + c_k}{1 + c_k},$$

where $\mathbf{c} = (c_1, \ldots, c_{p_0})$ has nonnegative components. This function is a product of univariate functions, which presents a greater challenge to BART than sums of univariate or bivariate functions provide. Here we use the coefficient values $c_k = (k-1)/2$ for $k = 1, \ldots, p_0$ suggested by Crestaux et al. [25]. We also use $p_0 = 5$ active variables which gives us $\texttt{Var}(f(\mathbf{X})) \approx 3.076$.

4. The Bratley function [26,27] with p_0 inputs is defined to be

$$\begin{aligned} f(\mathbf{x}) &:= \sum_{i=1}^{p_0} (-1)^i \prod_{j=1}^{i} x_j \\ &= -x_1 + x_1x_2 - x_1x_2x_3 + x_1x_2x_3x_4 - x_1x_2x_3x_4x_5 + \cdots \end{aligned}$$

This function is a sum of products of inputs, which provides an even greater challenge to BART than the $g-$function provides. Furthermore, this function produces nonzero Sobol' indices for all (non-null) variable index sets. Again, we use $p_0 = 5$ active variables which gives us $\texttt{Var}(f(\mathbf{X})) \approx 0.057$.

5. The function inspired by Morris et al. [28] is defined for $p_0 = 5$ as

$$\begin{aligned} f(\mathbf{x}) &:= \alpha \sum_{i=1}^{p_0} x_i + \beta \sum_{i=1}^{p_0-1} x_i \sum_{j=i+1}^{p_0} x_j \\ &= \alpha x_1 + \beta(x_1x_2 + x_1x_3 + x_1x_4 + x_1x_5) \\ &\quad + \alpha x_2 + \beta(x_2x_3 + x_2x_4 + x_2x_5) \\ &\quad + \alpha x_3 + \beta(x_3x_4 + x_3x_5) \\ &\quad + \alpha x_4 + \beta(x_4x_5) \\ &\quad + \alpha x_5 \end{aligned}$$

where $\alpha = \sqrt{12} - 6\sqrt{0.1(p_0 - 1)} \approx -0.331$ and $\beta = \frac{12}{\sqrt{10(p_0-1)}} \approx 1.897$ are chosen so that $S_1 = S_2 = \cdots = S_5 \approx 0.05$ and total-effects indices $T_1 = T_2 = \cdots = T_5 \approx 0.35$. Here, $\texttt{Var}(f(\mathbf{X})) \approx 5.25$. See also www.sfu.ca/~ssurjano/morretal06.html.

Hence, we consider $(3 \times 2 \times 2) \times 5 = 60$ possible combinations of (p, n, σ^2) parameter settings and data-generating functions. We will call these the 60 simulation scenarios.

For question **Q.2**, Treed Gaussian Processes (TGP) provide a benchmark to compare exact BART-based Sobol´ indices against [29,30]. However, there exists no literature on the exact computation of TGP-based Sobol´ indices. Hence, we estimate first-order and total-effects TGP-based Sobol´ indices using the `tgp::sens()` R function [16,31]. This function uses Monte Carlo integration whose error depends on its Latin hypercube design (LHD) sampling scheme. The $p + 2$ required LHD samples are randomly drawn for every MCMC iteration of the TGP model fitting, which propagates the Monte Carlo integration error into the posterior variability of the indices. To minimize this error, we use $100p-$point LHD samples.

5.2. Performance metrics

We evaluate our results in terms of two metrics: the L_1 performance metric and a rank-based performance metric.

L_1 performance metric

To answer question **Q.1** posed at the beginning of the section, we will first estimate the expectation of the L_1 distance $d_{L_1}(\cdot,\cdot)$ between BART-based Sobol´ indices and the true Sobol´ indices with respect to the BART posterior $\pi(\Theta \mid \mathcal{D})$ from Eq. (5). For example, if we are assessing the bias of BART-based first-order Sobol´ indices for a given number of inputs, sample size, and magnitude of the measurement error standard deviation (i.e for a given (p, n, σ^2)), we will estimate the expectation

$$\int d_{L_1}(\boldsymbol{S}^{\mathcal{E}}, \boldsymbol{S}^f)\, d\pi(\Theta \mid \mathcal{D}) \approx \frac{1}{1000} \sum_{i=1}^{1000} d_{L_1}(\boldsymbol{S}^{\mathcal{E}^{(i)}}, \boldsymbol{S}^f) \tag{14}$$

using 1000 posterior samples $\{(\Theta^{(i)} \mid \mathcal{D})\}_{i=1}^{1000}$, where the vectors $\boldsymbol{S}^{\mathcal{E}} = (S_1^{\mathcal{E}}, S_2^{\mathcal{E}}, \ldots, S_p^{\mathcal{E}})$ and $\boldsymbol{S}^f = (S_1^f, S_2^f, \ldots, S_p^f)$ contain all first-order Sobol´ indices of, respectively, BART ensemble function $\mathcal{E}(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m)$ and data-generating function $f(\cdot)$. Here, $\mathcal{E}(\cdot; \{\mathcal{T}_t, \mathcal{M}_t\}_{t=1}^m)$ is the BART ensemble function that results from posterior sample $(\Theta \mid \mathcal{D})$ while each $\mathcal{E}^{(i)}(\cdot)$ is similarly the BART ensemble function that results from the ith posterior sample $(\Theta^{(i)} \mid \mathcal{D})$. We will make similar estimates for two-way and total-effects Sobol´ index calculations. Finally, we average over replicated data sets $\mathcal{D}$ to arrive at our overall estimated L_1 distance. In the example above, we will generate 500 values of the expected L_1 distance estimate. The sample mean and standard deviation of these 500 estimates are shown in Table 2 (the numerical results are discussed in Section 5.3).

To answer question **Q.2**, we follow a methodology similar to that for answering question **Q.1**. We highlight three key differences. First, we compute TGP-based Sobol´ indices, which are approximated for any given posterior sample of the trained TGP. Second, we estimate TGP-based Sobol´ indices for only 40 data sets due to the substantial added computational demands of the required integral approximations. Table 2 reports the sample mean and standard of these 40 estimates. Third, we do not compute TGP-based second-order Sobol´ indices (i.e. $\hat{S}_{ij}^{TGP}$) because the `tgp::sens()` R function does not easily lend itself to such calculations. Hence, we rely on TGP-based total-indices to capture input variable interactions. Regarding the latter two key differences, we emphasize that TGP serves merely as a benchmark and not as the focus of this paper. Hence, we use TGP as only one possible popular alternative to BART-based Sobol´ indices.

Rank-based performance metric

To answer questions **Q.3** and **Q.4**, we replace the L_1 distance $d_{L_1}(\cdot,\cdot)$ in Eq. (14) with a discrepancy measure $d_r(\cdot,\cdot)$, to be defined in Eq. (15). This allows a more interpretable comparison between the performances of one-way BART counts and BART-based Sobol' indices. Table 3 shows the sample mean and standard deviation of these 500 estimates.

As an example, we will rank the Friedman function's normalized first-order Sobol´ index values $(S_1, S_2, S_3, S_4, S_5) = (.197, .197, .093, .350, .087)$ shown in Table 1 as $(2, 2, 4, 1, 5)$, where the most active variable (i.e. variable 4) is assigned ranking number 1 and the least active variable (i.e. variable 5) is assigned ranking number 5. Variables 1 and 2 are equally active, so we will adopt the convention used in many sports competitions of assigning the minimum rank to the two variables and then leaving a gap in the ranking numbers so that the positions of all variables less active than variables 1 and 2 are unaffected.

Several options exist for comparing two rankings. Kendall [32] introduces a distance that, when ties in rankings are not allowed, is the graphical distance between two vertices in the well-studied permutation polytope that represents all possible rankings of p objects [33]. Emond and Mason [34] point out that when ties are allowed, Kendall's "distance" violates the triangle inequality and hence is no longer a true metric. They advocate the distance defined by Kemeny and Snell [35], which equals Kendall's distance when ties are not allowed, but remains a metric when ties are allowed.

Table 1
S_i^f, T_i^f, and $T_i^f - S_i^f$ for various data-generating functions f and variable indices i.

i	Friedman			Modified Friedman			$g-$function			Bratley			Morris		
	S_i^f	T_i^f	$T_i^f - S_i^f$	S_i^f	T_i^f	$T_i^f - S_i^f$	S_i^f	T_i^f	$T_i^f - S_i^f$	S_i^f	T_i^f	$T_i^f - S_i^f$	S_i^f	T_i^f	$T_i^f - S_i^f$
1	0.197	0.274	0.077	0	0.335	0.335	0.433	0.701	0.268	0.688	0.766	0.078	0.190	0.210	0.019
2	0.197	0.274	0.077	0	0.335	0.335	0.108	0.284	0.176	0.142	0.220	0.078	0.190	0.210	0.019
3	0.093	0.093	0	0.117	0.117	0	0.048	0.135	0.087	0.051	0.099	0.048	0.190	0.210	0.019
4	0.350	0.350	0	0.438	0.438	0	0.027	0.078	0.051	0.006	0.018	0.012	0.190	0.210	0.019
5	0.087	0.087	0	0.110	0.110	0	0.017	0.050	0.033	0.006	0.018	0.012	0.190	0.210	0.019

Table 2
Estimates of the expected L_1 distance between BART-based Sobol´ indices and true Sobol´ indices when $f(\mathbf{x})$ is measured with error. Each block of four scenarios is ordered roughly in decreasing order of "signal-to-noise".

	(p, n, σ^2)	BART: mean (sd) of 500 replicates			TGP: mean (sd) of 40 replicates	
		S_i^f vs. $S_i^{\mathcal{E}}$	S_{ij}^f vs. $S_{ij}^{\mathcal{E}}$	T_i^f vs. $T_i^{\mathcal{E}}$	S_i^f vs. $\hat{S}_i^{TGP}$	T_i^f vs. $\hat{T}_i^{TGP}$
Friedman function	(5, 50p, 0.10)	0.072 (0.021)	0.067 (0.006)	0.137 (0.037)	0.198 (0.007)	0.444 (0.030)
	(5, 50p, 0.25)	0.099 (0.034)	0.082 (0.005)	0.174 (0.047)	0.269 (0.019)	0.938 (0.108)
	(5, 10p, 0.10)	0.184 (0.065)	0.089 (0.001)	0.267 (0.095)	0.289 (0.048)	0.886 (0.218)
	(5, 10p, 0.25)	0.206 (0.077)	0.091 (0.002)	0.289 (0.095)	0.339 (0.058)	1.272 (0.275)
	(10, 50p, 0.10)	0.074 (0.018)	0.050 (0.006)	0.137 (0.024)	1.030 (0.009)	1.100 (0.027)
	(10, 50p, 0.25)	0.120 (0.026)	0.075 (0.008)	0.219 (0.034)	1.122 (0.017)	1.497 (0.071)
	(10, 10p, 0.10)	0.257 (0.051)	0.098 (0.002)	0.403 (0.059)	1.116 (0.043)	1.410 (0.179)
	(10, 10p, 0.25)	0.321 (0.061)	0.102 (0.003)	0.465 (0.068)	1.189 (0.044)	2.025 (0.352)
	(15, 50p, 0.10)	0.068 (0.014)	0.037 (0.006)	0.119 (0.018)	1.903 (0.009)	1.693 (0.013)
	(15, 50p, 0.25)	0.123 (0.021)	0.064 (0.008)	0.212 (0.025)	2.008 (0.015)	2.009 (0.082)
	(15, 10p, 0.10)	0.269 (0.045)	0.100 (0.002)	0.434 (0.050)	1.992 (0.039)	1.998 (0.227)
	(15, 10p, 0.25)	0.349 (0.055)	0.105 (0.003)	0.515 (0.059)	2.074 (0.037)	3.081 (0.525)
Mod. Friedman function	(5, 50p, 0.10)	0.080 (0.019)	0.093 (0.019)	0.166 (0.052)	0.167 (0.013)	0.409 (0.047)
	(5, 50p, 0.25)	0.118 (0.028)	0.136 (0.026)	0.241 (0.070)	0.225 (0.023)	0.862 (0.080)
	(5, 10p, 0.10)	0.370 (0.047)	0.355 (0.004)	0.913 (0.051)	0.207 (0.031)	0.651 (0.168)
	(5, 10p, 0.25)	0.343 (0.049)	0.352 (0.002)	0.682 (0.067)	0.288 (0.058)	1.220 (0.234)
	(10, 50p, 0.10)	0.079 (0.016)	0.057 (0.014)	0.134 (0.027)	0.792 (0.011)	1.225 (0.019)
	(10, 50p, 0.25)	0.128 (0.025)	0.101 (0.021)	0.234 (0.042)	0.864 (0.017)	1.514 (0.041)
	(10, 10p, 0.10)	0.336 (0.035)	0.350 (0.002)	0.717 (0.054)	0.856 (0.039)	1.480 (0.117)
	(10, 10p, 0.25)	0.400 (0.057)	0.362 (0.003)	0.931 (0.055)	0.909 (0.041)	1.921 (0.261)
	(15, 50p, 0.10)	0.072 (0.013)	0.044 (0.011)	0.121 (0.020)	1.444 (0.012)	1.994 (0.014)
	(15, 50p, 0.25)	0.130 (0.020)	0.084 (0.017)	0.227 (0.030)	1.523 (0.016)	2.179 (0.029)
	(15, 10p, 0.10)	0.335 (0.046)	0.314 (0.017)	0.853 (0.070)	1.506 (0.033)	2.256 (0.168)
	(15, 10p, 0.25)	0.404 (0.056)	0.348 (0.012)	0.966 (0.063)	1.562 (0.038)	3.002 (0.431)
$g-$function	(5, 50p, 0.10)	0.374 (0.061)	0.272 (0.005)	0.448 (0.069)	0.320 (0.039)	1.640 (0.140)
	(5, 50p, 0.25)	0.432 (0.092)	0.269 (0.005)	0.530 (0.090)	0.410 (0.042)	2.355 (0.185)
	(5, 10p, 0.10)	0.643 (0.108)	0.280 (0.002)	0.747 (0.112)	0.458 (0.103)	2.217 (0.366)
	(5, 10p, 0.25)	0.697 (0.112)	0.279 (0.002)	0.815 (0.116)	0.539 (0.097)	2.571 (0.330)
	(10, 50p, 0.10)	0.374 (0.057)	0.326 (0.007)	0.570 (0.050)	0.911 (0.022)	3.407 (0.190)
	(10, 50p, 0.25)	0.446 (0.077)	0.341 (0.007)	0.707 (0.066)	1.008 (0.022)	4.527 (0.207)
	(10, 10p, 0.10)	0.688 (0.094)	0.333 (0.003)	0.997 (0.092)	1.052 (0.064)	4.436 (0.428)
	(10, 10p, 0.25)	0.786 (0.105)	0.336 (0.003)	1.130 (0.106)	1.095 (0.053)	5.039 (0.512)
	(15, 50p, 0.10)	0.367 (0.048)	0.325 (0.012)	0.584 (0.046)	1.635 (0.020)	6.983 (0.267)
	(15, 50p, 0.25)	0.444 (0.064)	0.356 (0.008)	0.756 (0.052)	1.527 (0.020)	5.446 (0.221)
	(15, 10p, 0.10)	0.650 (0.089)	0.348 (0.005)	1.050 (0.086)	1.680 (0.052)	7.474 (0.794)
	(15, 10p, 0.25)	0.802 (0.092)	0.354 (0.005)	1.240 (0.099)	1.738 (0.035)	8.680 (0.815)
Bratley function	(5, 50p, 0.10)	0.070 (0.020)	0.043 (0.007)	0.099 (0.016)	0.180 (0.017)	0.527 (0.039)
	(5, 50p, 0.25)	0.106 (0.037)	0.050 (0.009)	0.146 (0.024)	0.255 (0.028)	0.962 (0.087)
	(5, 10p, 0.10)	0.221 (0.078)	0.081 (0.002)	0.276 (0.049)	0.219 (0.051)	0.679 (0.100)
	(5, 10p, 0.25)	0.293 (0.098)	0.079 (0.002)	0.343 (0.068)	0.288 (0.052)	1.074 (0.200)
	(10, 50p, 0.10)	0.068 (0.017)	0.041 (0.007)	0.122 (0.012)	1.027 (0.015)	1.373 (0.032)
	(10, 50p, 0.25)	0.125 (0.030)	0.058 (0.009)	0.200 (0.021)	1.112 (0.017)	1.954 (0.087)
	(10, 10p, 0.10)	0.231 (0.054)	0.103 (0.001)	0.376 (0.035)	1.039 (0.033)	1.562 (0.145)
	(10, 10p, 0.25)	0.331 (0.071)	0.106 (0.002)	0.472 (0.051)	1.125 (0.050)	2.179 (0.283)
	(15, 50p, 0.10)	0.065 (0.015)	0.033 (0.006)	0.117 (0.010)	1.887 (0.009)	2.234 (0.038)
	(15, 50p, 0.25)	0.125 (0.021)	0.054 (0.008)	0.205 (0.017)	1.974 (0.017)	2.990 (0.122)
	(15, 10p, 0.10)	0.223 (0.042)	0.107 (0.002)	0.394 (0.029)	1.905 (0.035)	2.511 (0.166)
	(15, 10p, 0.25)	0.326 (0.061)	0.113 (0.002)	0.504 (0.052)	1.982 (0.024)	3.344 (0.253)

(*continued on next page*)

Unfortunately, the Kemeny–Snell (KS) distance is likely to artificially inflate when the data-generating function has either more than two inert variables or has equally-active non-inert variables. In the former scenario, a fitted BART model is unlikely to entirely shrink all of its input activity measures of the inert variables. In this case, the KS distance will be inflated by the fitted BART model assigning small but positive effects to the inert variables. In the latter scenario, a fitted BART model is unlikely to perfectly match its input activity measures of the equally-active non-inert variables. In this case, the fitted BART model could be incorrectly "punished" for even the slightest

Table 2 (*continued*).

	$(p,\ n,\ \sigma^2)$	BART: mean (sd) of 500 replicates			TGP: mean (sd) of 40 replicates	
		S_i^f vs. $S_i^{\mathcal{E}}$	S_{ij}^f vs. $S_{ij}^{\mathcal{E}}$	T_i^f vs. $T_i^{\mathcal{E}}$	S_i^f vs. $\hat{S}_i^{TGP}$	T_i^f vs. $\hat{T}_i^{TGP}$
Morris function	(5, 50p, 0.10)	0.063 (0.023)	0.017 (0.003)	0.064 (0.025)	0.439 (0.030)	0.468 (0.110)
	(5, 50p, 0.25)	0.092 (0.035)	0.018 (0.003)	0.093 (0.036)	0.283 (0.028)	1.238 (0.163)
	(5, 10p, 0.10)	0.137 (0.048)	0.033 (0.002)	0.140 (0.049)	0.463 (0.054)	0.517 (0.184)
	(5, 10p, 0.25)	0.188 (0.070)	0.031 (0.003)	0.191 (0.072)	0.343 (0.063)	1.152 (0.315)
	(10, 50p, 0.10)	0.079 (0.017)	0.029 (0.004)	0.103 (0.016)	0.627 (0.019)	0.719 (0.011)
	(10, 50p, 0.25)	0.130 (0.024)	0.044 (0.004)	0.167 (0.023)	0.433 (0.024)	2.072 (0.271)
	(10, 10p, 0.10)	0.220 (0.039)	0.058 (0.002)	0.284 (0.034)	0.614 (0.041)	0.785 (0.114)
	(10, 10p, 0.25)	0.295 (0.056)	0.060 (0.002)	0.360 (0.051)	0.457 (0.052)	1.973 (0.511)
	(15, 50p, 0.10)	0.076 (0.013)	0.026 (0.003)	0.098 (0.013)	0.830 (0.018)	0.958 (0.090)
	(15, 50p, 0.25)	0.135 (0.018)	0.044 (0.005)	0.180 (0.019)	0.632 (0.021)	2.921 (0.353)
	(15, 10p, 0.10)	0.239 (0.034)	0.064 (0.002)	0.326 (0.031)	0.805 (0.038)	0.990 (0.135)
	(15, 10p, 0.25)	0.326 (0.043)	0.069 (0.003)	0.418 (0.042)	0.622 (0.043)	2.775 (0.679)

Table 3
Estimates of the expected d_r discrepancy between BART-based Sobol´ index rankings and true Sobol´ index rankings when $f(\mathbf{x})$ is measured with error. Each block of four scenarios is ordered roughly in decreasing order of "signal-to-noise".

	$(p,\ n,\ \sigma^2)$	$S_i^{\mathcal{E}}$ vs. S_i^f	Count vs. S_i^f	$T_i^{\mathcal{E}}$ vs. T_i^f	Count vs. T_i^f	$S_{ij}^{\mathcal{E}}$ vs. S_{ij}^f
Friedman function	Max value	20	20	20	20	20
	(5, 50p, 0.10)	1.048 (1.000)	5.420 (2.195)	1.040 (1.000)	5.420 (2.195)	0.000 (0.000)
	(5, 50p, 0.25)	1.152 (1.005)	6.404 (3.186)	1.144 (1.015)	6.404 (3.186)	0.008 (0.126)
	(5, 10p, 0.10)	2.540 (1.484)	8.620 (3.011)	2.536 (1.488)	8.620 (3.011)	8.716 (5.690)
	(5, 10p, 0.25)	2.816 (1.892)	8.916 (3.086)	2.812 (1.888)	8.916 (3.086)	9.848 (5.696)
	Max value	70	70	70	70	90
	(10, 50p, 0.10)	0.976 (1.001)	4.984 (1.430)	0.960 (1.000)	4.984 (1.430)	0.000 (0.000)
	(10, 50p, 0.25)	1.136 (0.992)	6.144 (2.776)	1.108 (0.995)	6.144 (2.776)	0.000 (0.000)
	(10, 10p, 0.10)	2.020 (0.900)	18.608 (8.795)	2.016 (0.904)	18.608 (8.795)	4.820 (9.915)
	(10, 10p, 0.25)	2.412 (1.591)	23.844 (10.447)	2.424 (1.615)	23.844 (10.447)	12.772 (16.712)
	Max value	120	120	120	120	210
	(15, 50p, 0.10)	0.792 (0.979)	4.432 (0.862)	0.768 (0.974)	4.432 (0.862)	0.000 (0.000)
	(15, 50p, 0.25)	1.008 (1.001)	5.172 (1.919)	1.008 (1.001)	5.172 (1.919)	0.000 (0.000)
	(15, 10p, 0.10)	1.780 (0.754)	8.504 (5.658)	1.772 (0.762)	8.504 (5.658)	0.376 (1.647)
	(15, 10p, 0.25)	1.988 (1.093)	15.456 (10.630)	1.988 (1.107)	15.456 (10.630)	3.760 (9.419)
Mod. Friedman function	Max value	20	20	20	20	20
	(5, 50p, 0.10)	1.024 (1.001)	13.292 (1.359)	1.004 (1.001)	5.300 (1.363)	0.000 (0.000)
	(5, 50p, 0.25)	1.004 (1.001)	13.728 (1.761)	1.024 (1.001)	5.752 (1.791)	0.000 (0.000)
	(5, 10p, 0.10)	1.864 (0.789)	6.796 (5.295)	13.896 (4.901)	27.528 (8.823)	0.000 (0.000)
	(5, 10p, 0.25)	2.020 (1.211)	3.960 (2.851)	9.280 (1.257)	9.360 (2.678)	0.136 (0.656)
	Max value	70	70	70	70	90
	(10, 50p, 0.10)	0.856 (0.991)	12.736 (1.098)	0.860 (0.991)	4.736 (1.098)	0.000 (0.000)
	(10, 50p, 0.25)	1.024 (1.001)	13.940 (2.078)	1.020 (1.001)	5.940 (2.078)	0.000 (0.000)
	(10, 10p, 0.10)	1.888 (0.832)	3.840 (2.521)	9.648 (0.842)	9.392 (2.665)	0.000 (0.000)
	(10, 10p, 0.25)	1.912 (1.390)	8.552 (5.685)	17.740 (5.670)	31.520 (8.478)	0.000 (0.000)
	Max value	120	120	120	120	210
	(15, 50p, 0.10)	0.696 (0.954)	12.500 (0.867)	0.680 (0.948)	4.500 (0.867)	0.000 (0.000)
	(15, 50p, 0.25)	0.976 (1.001)	12.940 (1.348)	0.952 (1.000)	4.948 (1.354)	0.000 (0.000)
	(15, 10p, 0.10)	1.724 (0.690)	5.292 (4.161)	8.096 (2.377)	12.596 (10.107)	0.000 (0.000)
	(15, 10p, 0.25)	1.792 (0.958)	5.716 (5.927)	10.928 (4.388)	29.084 (15.196)	0.000 (0.000)
g−function	Max value	20	20	20	20	90
	(5, 50p, 0.10)	1.068 (1.311)	5.768 (3.576)	1.092 (1.328)	5.768 (3.576)	35.640 (10.136)
	(5, 50p, 0.25)	1.920 (1.757)	8.492 (4.441)	1.968 (1.759)	8.492 (4.441)	42.168 (12.003)
	(5, 10p, 0.10)	4.636 (3.029)	12.864 (4.367)	4.648 (3.020)	12.864 (4.367)	52.236 (13.342)
	(5, 10p, 0.25)	6.160 (3.358)	13.576 (4.197)	6.204 (3.376)	13.576 (4.197)	54.632 (13.402)
	Max value	70	70	70	70	790
	(10, 50p, 0.10)	0.652 (0.996)	4.188 (3.201)	0.640 (0.984)	4.188 (3.201)	175.484 (52.348)
	(10, 50p, 0.25)	1.392 (1.591)	8.816 (5.697)	1.444 (1.647)	8.816 (5.697)	252.916 (64.184)
	(10, 10p, 0.10)	6.388 (4.313)	27.040 (12.884)	6.428 (4.286)	27.040 (12.884)	338.148 (87.447)
	(10, 10p, 0.25)	10.500 (6.133)	35.024 (12.362)	10.668 (6.162)	35.024 (12.362)	377.112 (88.710)
	Max value	120	120	120	120	1990
	(15, 50p, 0.10)	0.408 (0.817)	2.088 (2.360)	0.400 (0.811)	2.088 (2.360)	260.228 (98.276)
	(15, 50p, 0.25)	0.864 (1.177)	6.120 (4.861)	0.928 (1.209)	6.120 (4.861)	456.692 (133.275)
	(15, 10p, 0.10)	4.780 (4.517)	18.124 (10.946)	4.808 (4.544)	18.124 (10.946)	578.464 (172.922)
	(15, 10p, 0.25)	10.908 (7.537)	34.672 (18.045)	11.016 (7.569)	34.672 (18.045)	733.748 (211.209)

(*continued on next page*)

discrepancy between the variable-activity measures of two equally-active variables. To resolve this issue, we create a discrepancy measure based on the multi-stage discordance measures discussed in [36].

To compute the discordance measures between two rankings ρ_f and $\rho_{\mathcal{E}}$, where in our variable activity setting ρ_f represents the true input activity and $\rho_{\mathcal{E}}$ represents the input activity of our fitted BART model, Fligner and Verducci [36] assume that neither ranking has any

Table 3 (*continued*).

	(p, n, σ^2)	$S_i^{\mathcal{E}}$ vs. S_i^f	Count vs. S_i^f	$T_i^{\mathcal{E}}$ vs. T_i^f	Count vs. T_i^f	$S_{ij}^{\mathcal{E}}$ vs. S_{ij}^f
Bratley function	Max value	20	20	20	20	90
	(5, 50p, 0.10)	0.000 (0.000)	0.656 (1.097)	0.000 (0.000)	0.656 (1.097)	0.124 (0.627)
	(5, 50p, 0.25)	0.008 (0.126)	2.004 (2.188)	0.000 (0.000)	2.004 (2.188)	1.056 (1.991)
	(5, 10p, 0.10)	0.280 (0.771)	9.132 (4.824)	0.280 (0.771)	9.132 (4.824)	11.872 (8.695)
	(5, 10p, 0.25)	0.716 (1.231)	9.932 (4.746)	0.716 (1.237)	9.932 (4.746)	15.480 (9.515)
	Max value	70	70	70	70	790
	(10, 50p, 0.10)	0.084 (0.553)	2.124 (2.933)	0.004 (0.089)	2.124 (2.933)	28.348 (26.519)
	(10, 50p, 0.25)	1.236 (2.459)	4.884 (4.212)	0.580 (1.495)	4.884 (4.212)	79.328 (45.473)
	(10, 10p, 0.10)	5.352 (4.851)	20.588 (8.920)	5.280 (4.784)	20.588 (8.920)	197.556 (60.194)
	(10, 10p, 0.25)	7.352 (4.991)	25.356 (9.913)	7.284 (4.890)	25.356 (9.913)	237.864 (69.558)
	Max value	120	120	120	120	1990
	(15, 50p, 0.10)	0.020 (0.236)	0.948 (1.905)	0.000 (0.000)	0.948 (1.905)	22.344 (31.235)
	(15, 50p, 0.25)	1.216 (2.566)	6.600 (6.780)	0.464 (1.517)	6.600 (6.780)	126.056 (91.807)
	(15, 10p, 0.10)	7.052 (7.614)	24.312 (12.077)	7.008 (7.603)	24.312 (12.077)	402.292 (128.257)
	(15, 10p, 0.25)	11.728 (9.074)	30.804 (12.989)	11.636 (9.031)	30.804 (12.989)	483.504 (146.360)
Morris function	Max value	20	20	20	20	90
	(5, 50p, 0.10)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	(5, 50p, 0.25)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	(5, 10p, 0.10)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	(5, 10p, 0.25)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Max value	70	70	70	70	790
	(10, 50p, 0.10)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	1.396 (3.844)
	(10, 50p, 0.25)	0.000 (0.000)	0.240 (0.951)	0.000 (0.000)	0.240 (0.951)	25.252 (24.050)
	(10, 10p, 0.10)	0.000 (0.000)	11.880 (7.979)	0.000 (0.000)	11.880 (7.979)	129.608 (65.530)
	(10, 10p, 0.25)	0.000 (0.000)	15.144 (9.359)	0.000 (0.000)	15.144 (9.359)	191.964 (78.236)
	Max value	120	120	120	120	1990
	(15, 50p, 0.10)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.304 (2.605)
	(15, 50p, 0.25)	0.000 (0.000)	0.008 (0.126)	0.000 (0.000)	0.008 (0.126)	19.228 (24.255)
	(15, 10p, 0.10)	0.000 (0.000)	2.800 (4.816)	0.000 (0.000)	2.800 (4.816)	137.684 (91.270)
	(15, 10p, 0.25)	0.000 (0.000)	8.124 (8.533)	0.000 (0.000)	8.124 (8.533)	269.444 (125.489)

ties. As an example, suppose $\rho_f = (4, 3, 1, 2)$ and $\rho_{\mathcal{E}} = (3, 1, 2, 4)$. The discordances $W_1, W_2, \ldots, W_4$ will be computed sequentially. To compute discordance W_1, we see that variable 3 is the most active in ρ_f. Since variable 3 is the second most active in $\rho_{\mathcal{E}}$, we set $W_1 = 2 - 1 = 1$. We then remove variable 3 from consideration to compute W_2, W_3, and W_4. To compute discordance W_2, we see that variable 4 is the most active of the remaining variables (1, 2, and 4) in ρ_f. Since variable 4 is the third most active of the remaining variables (1, 2, and 4) in $\rho_{\mathcal{E}}$, we set $W_2 = 3 - 1 = 2$. We then remove variable 4 from consideration to compute discordances W_3 and W_4. To compute W_3, we see that variable 2 is the most active of the remaining variables (1 and 2) in ρ_f. Since variable 2 is the most active of the remaining variables (1 and 2) in $\rho_{\mathcal{E}}$, we set $W_3 = 1 - 1 = 0$. We then remove variable 2 from consideration to compute W_4. Since only one variable remains, we set $W_4 = 0$. Hence, the discordances in this example are $(W_1, W_2, W_3, W_4) = (1, 2, 0, 0)$.

More generally (but still assuming neither ranking has any ties), suppose we have already computed discordances $W_1, \ldots, W_{k-1}$ for some $k = 1, \ldots, q$, where q is the number of elements in vector ρ_f, and wish to compute W_k. Note that q is not necessarily p (e.g. $q = \binom{p}{2}$ when the rankings represent two-way interactions). Thus, we have removed $k - 1$ of the q items (e.g. variables, variable pairs, variable triplets) from consideration. If item i is the most active in ranking ρ_f among the remaining considered items, we then find j, where item i is the jth most active in ranking $\rho_{\mathcal{E}}$ among the remaining considered items. The value W_k is then set to be $j - 1$.

Now suppose both ranking ρ_f and ranking $\rho_{\mathcal{E}}$ are allowed to have ties. As mentioned earlier, we will adopt the "standard competition" ranking convention of, for each set of items tied with each other, assigning the minimum rank to the tied items and then leaving a gap in the ranking numbers so that the positions of all items less active than the tied items are unaffected. For example, first-order Sobol´ index values $(0.1, 0.1, 0.2, 0.2, 0.35, 0.05)$ would be ranked $(4, 4, 2, 2, 1, 6)$. Suppose we have already computed $W_1, \ldots, W_{k-1}$ for some $k = 1, \ldots, q$ and wish to compute W_k. If $u \geq 1$ items $i_1, \ldots, i_u$ are tied for most active in ranking ρ_f among the remaining considered items, we can find $j_1, \ldots, j_u$, where item i_l is the j_lth most active item in ranking $\rho_{\mathcal{E}}$ among the remaining considered items. The value W_k is then set to be $\min_{l=1,\ldots,u} j_l - 1$. If $argmin_{l=1,\ldots,u} j_l$ has more than one value, then we pick any one (it does not matter which) of the corresponding items i_l to remove from consideration. Once an item is removed from consideration, the value W_{k+1} can then be computed (if $k < p$). Note that if $u = 1$, this reduces to the "no-ties" case.

We can now define our discrepancy measure between rankings ρ_f and $\rho_{\mathcal{E}}$:

$$d_r(\rho_f, \rho_{\mathcal{E}}) = 2 \sum_{k=1}^{q} W_k, \tag{15}$$

where discordances $W_1, W_2, \ldots, W_q$ are computed as described in the previous paragraph. This measure has three particularly desirable properties. First, it equals Kendall's distance (and hence the KS distance) when ties are not allowed. Second, it does not inflate as the number of data-generating function f's inert variables increases. In particular, discordance $W_k = 0$ for all $k > q_0$, where q_0 is the number of active items (i.e. items with non-zero input activity measure) in f. Hence, the discrepancy measure is invariant to the number of inert variables. Third, it does not inflate when f has equally-active non-inert items. If f has a set of equally-active non-inert items, then the discrepancy measure will not inflate as long as the equally-active items in the set are consecutively ranked. These three properties can be stated as Theorems 5, 6, and 7 whose proofs are in the Appendix.

Theorem 5. *If rankings $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ each have no ties, then the Kemeny–Snell distance between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ equals the discrepancy measure $d_r(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in Eq. (15).*

Theorem 6. *Consider the discrepancy $d_r(\rho_f, \rho_{\mathcal{E}})$ between rankings ρ_f and $\rho_{\mathcal{E}}$. Then the discordance $W_k = 0$ for all $k > q_0$, where q_0 is the number of active items in f.*

Theorem 7. *Consider the discrepancy $d_r(\rho_f, \rho_{\mathcal{E}})$ between rankings ρ_f and $\rho_{\mathcal{E}}$. Suppose $u \geq 1$ items $i_j, \ldots, i_{j+u-1}$ (and no other items) have ranking number j in ρ_f and ranking numbers $r_j, \ldots, r_{j+u-1}$ in $\rho_{\mathcal{E}}$. Then all $|\rho_f|$*

discordances are invariant to choice of permutation ϕ of set $\{r_j, \ldots, r_{j+u-1}\}$ of ranking numbers.

5.3. Simulation results

To answer question **Q.2**, we find in Table 2 that when the underlying data-generating function is not known, our BART-based Sobol´ indices should always be preferred over TGP-based Sobol´ indices. For all five data-generating functions, and for $p = 10$ or 15, if we compare the BART and TGP "S_i" columns to each other and the BART and TGP "T_i" columns to each other, the BART-based L_1 distances are uniformly lower than the TGP-based L_1 distances, often by a dramatic amount. For $p = 5$, this observation does not hold uniformly (e.g. for the (function, p, n, σ^2) = (Mod. Friedman, $5, 10p, 0.10$) row, the TGP-based total-effects L_1 distance is lower than that of BART), but is still valid in most of these cases.

To answer question **Q.1**, we find in the BART columns of Table 2 that the bias of BART-based Sobol´ indices is largest with the multiplicative $g-$function and increases with increasing noise and decreasing sample size.

For the five data-generating functions, the L_1 distances for all five data-generating functions and all three Sobol´ index measures tend to increase with each of increasing noise and decreasing sample size. That is, for a given number p of variables, the BART-based Sobol´ indices perform better as "signal-to-noise" ratio increases.

Interestingly, for the Friedman, Modified Friedman, $g-$, and Bratley functions, and for each set of four scenarios, the performance difference between data scenarios ($n = 50p, \sigma^2 = 0.10\mathtt{Var}(f(\mathbf{X}))$) and ($n = 50p, \sigma^2 = 0.25\mathtt{Var}(f(\mathbf{X}))$) is much smaller than the performance differences between $(50p, 0.25)$ and $(10p, 0.10)$ and between $(10p, 0.10)$ and $(10p, 0.25)$. We might infer that the $n = 50p$ scenario saturates the data with enough signal for modest noise increases to not make much of a performance difference, but the change from $50p$ to $10p$ makes the signal so scarce that modest noise increases does make a performance difference. These differences seem to be more evenly spread out for the Morris function.

Recall from Table 1 that for the modified Friedman function, first-order indices $S_1^f = S_2^f = 0$ while total-effects indices $T_1^f = T_2^f = 0.335$. That is, variables x_1 and x_2 interact strongly with other inputs but are not important on their own. We see in Table 2 that in the ($p = 5, n = 50p, \sigma^2 = 0.10\mathtt{Var}(f(\mathbf{X}))$) and ($p = 5, n = 50p, \sigma^2 = 0.25\mathtt{Var}(f(\mathbf{X}))$) scenarios, BART captures both the first-order and total-effects indices of the modified Friedman function about as well as it captures the same indices of the original Friedman function. This implies that with enough signal, BART is able to tell if an input is important on its own or if it merely interacts strongly with other inputs.

Also perhaps unsurprisingly, BART performs worse with the multiplicative $g-$function than it does with the four other data-generating functions. The Friedman function, modified Friedman function, and Morris function each is a sum of either univariate or bivariate functions which BART's additive structure can presumably capture well. The Bratley function is a sum of five simple terms with two of them being either univariate or bivariate. On the other hand, our $g-$function is a product of five univariate functions. If we note that the log of our $g-$function is also a sum of univariate functions, we might expect BART to perform better if we took the log of the $g-$function response data.

Finally, BART tends to capture total-effects indices less accurately than it does first-order indices. Interestingly, this observation holds even for the high-signal scenarios. To answer questions **Q.3** and **Q.4**, we find in Table 3 that the first-order rankings provided by BART-based Sobol´ indices are uniformly more accurate than those provided by one-way BART counts. First, the "$S_i^{\mathcal{E}}$ vs. S_i^f" column implies our BART-based first-order Sobol´ indices perform incredibly well at predicting the correct order of the true first-order Sobol´ indices across all data scenarios for the original Friedman, modified Friedman, and Morris functions, and across the $n = 50p$ scenarios for the $g-$ and Bratley functions. For all Morris-function scenarios, all 500 sets of BART-based first-order Sobol´ indices correctly rank the first $p_0 = 5$ input variables as more active than any of the other $p - p_0$ input variables.

Second, the "$T_i^{\mathcal{E}}$ vs. T_i^f" column implies our BART-based total-effects Sobol´ indices also perform very well at predicting the correct order of the true total-effects Sobol´ indices across all data scenarios for the original Friedman and Morris functions, and across the $n = 50p$ scenarios for the modified Friedman, $g-$, and Bratley functions. Again, for all Morris-function scenarios, all 500 sets of BART-based total-effects Sobol´ indices correctly rank the first $p_0 = 5$ input variables as more active than any of the other $p - p_0$ input variables.

Finally, for each row, the "$S_i^{\mathcal{E}}$ vs. S_i^f" expected discrepancy estimate is the same[2] or lower than the "Count vs. S_i^f" expected discrepancy estimate. Similarly, the "$T_i^{\mathcal{E}}$ vs. T_i^f" expected discrepancy estimate is the same or lower than the "Count vs. T_i^f" expected discrepancy estimate for all rows. These observations imply that our BART-based Sobol´ indices outperform one-way BART counts across the board when predicting the correct order of the true first-order Sobol´ indices and of the true total-effects Sobol´ indices. Hence, our BART-based first-order and total-effects Sobol´ indices should always be preferred over one-way counts.

We conclude that our BART-based first-order Sobol´ indices can accurately predict the raw values of first-order Sobol´ indices of additive data-generating functions. Also, our BART-based first-order and total-effects Sobol´ indices can accurately predict the rankings of, respectively, first-order Sobol´ indices and total-effects Sobol´ indices of both additive and multiplicative data-generating functions. Finally, our BART-based first-order and total-effects Sobol´ indices outperform one-way BART counts for all three data-generating functions.

5.4. Application to the En-ROADS climate simulator

We compute Sobol´ indices for a BART model trained on data generated from the En-ROADS climate simulator [37]. This simulator is a mathematical model of how global temperature and carbon emissions, among other factors, are influenced by changes in energy, land use, consumption, agriculture, and other policies. It is designed to be easily used by policymakers, educators, and the general public. The model, an ordinary differential equation solved by Euler integration, synthesizes what its developers consider to be the best available climate science. This simulator is available from the Climate Interactive website.

For this paper, we looked specifically at how the average global temperature increase by 2100 from pre-industrial levels is influenced by the 18 "top-level" input variables shown when the En-ROADS climate simulator is first loaded on to a web browser. We explored a subset of 11 variables as summarized in Fig. 5 and left the remaining 7 variables at their default settings. Based on an initial exploratory analysis, we found these 7 variables to be either redundant (coal, bioenergy, nuclear, electrification of buildings and industry, deforestation), impractical or unethical to control (population growth), or too discrete to treat as a continuous variable (technological carbon removal). Each input variable is bounded by a minimum and maximum value. We found a maximin LHD of $10 \times 11 = 110$ points on $[0, 1]^{11}$ and scaled it so that the design space contained the range of possible values. However, the simulator rounds values entered into its text fields, effectively rounding each design point to the nearest point on the induced 11-dimensional grid. We then manually obtained response values for each design point. The simulator also rounded the response values to the nearest first decimal place. Because a 0.1 °F difference is smaller than a 0.1 °C difference, we used Fahrenheit values. We then rescaled this "rounded" maximin LHD design back onto $[0, 1]^{11}$ to which we trained a BART model with the default parameter settings of, in particular,

[2] We say two estimates $est_1(se_1)$ and $est_2(se_2)$ are the same if the intervals produced by the estimate plus or minus the shown standard error overlap (i.e. if interval $(est_1 - se_1, est_1 + se_1)$ and interval $(est_2 - se_2, est_2 + se_2)$ overlap).

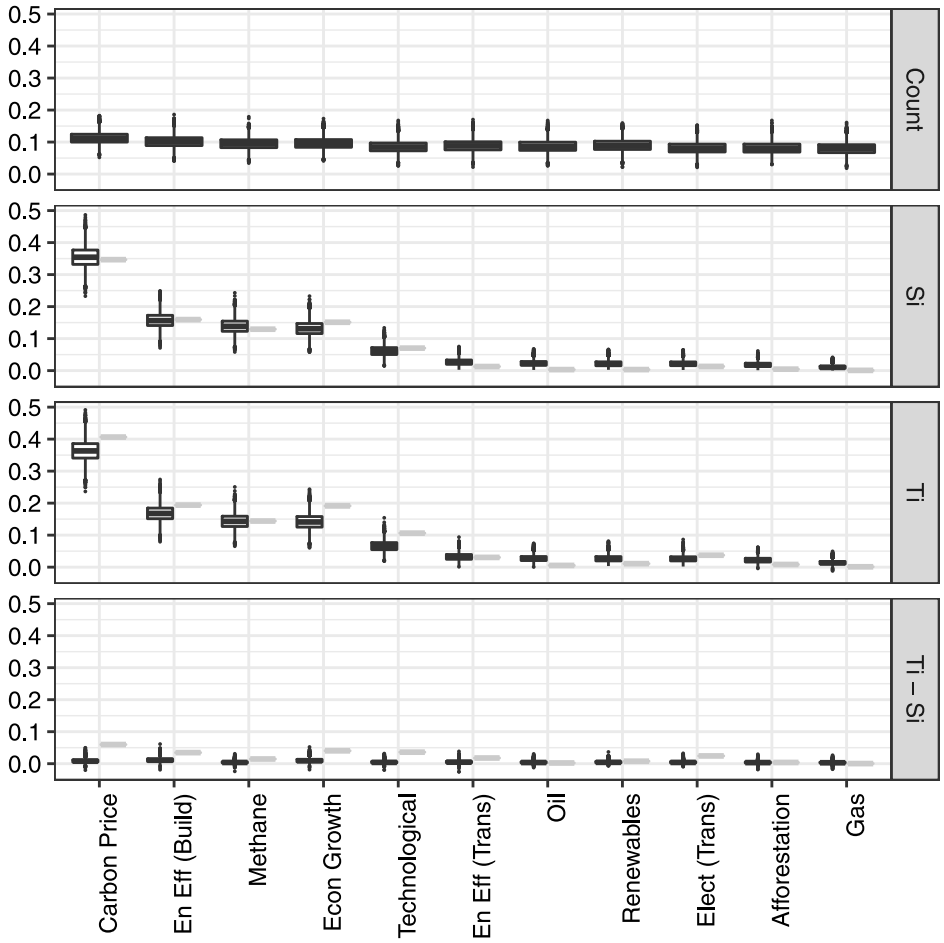


Fig. 5. Variable activity measures of BART and GP models trained on data from En-ROADS climate simulator. Variable counts (top panel), BART-based first-order Sobol´ indices (second panel), BART-based total-effects Sobol´ indices (third panel), and the difference between total-effects and first-order (bottom panel) are shown. Variable activity measures of the 10,000 ensembles corresponding to posterior samples of the trained BART model are shown in black. Point estimates of Sobol´ indices of the trained GP model based on the same data are shown in gray.

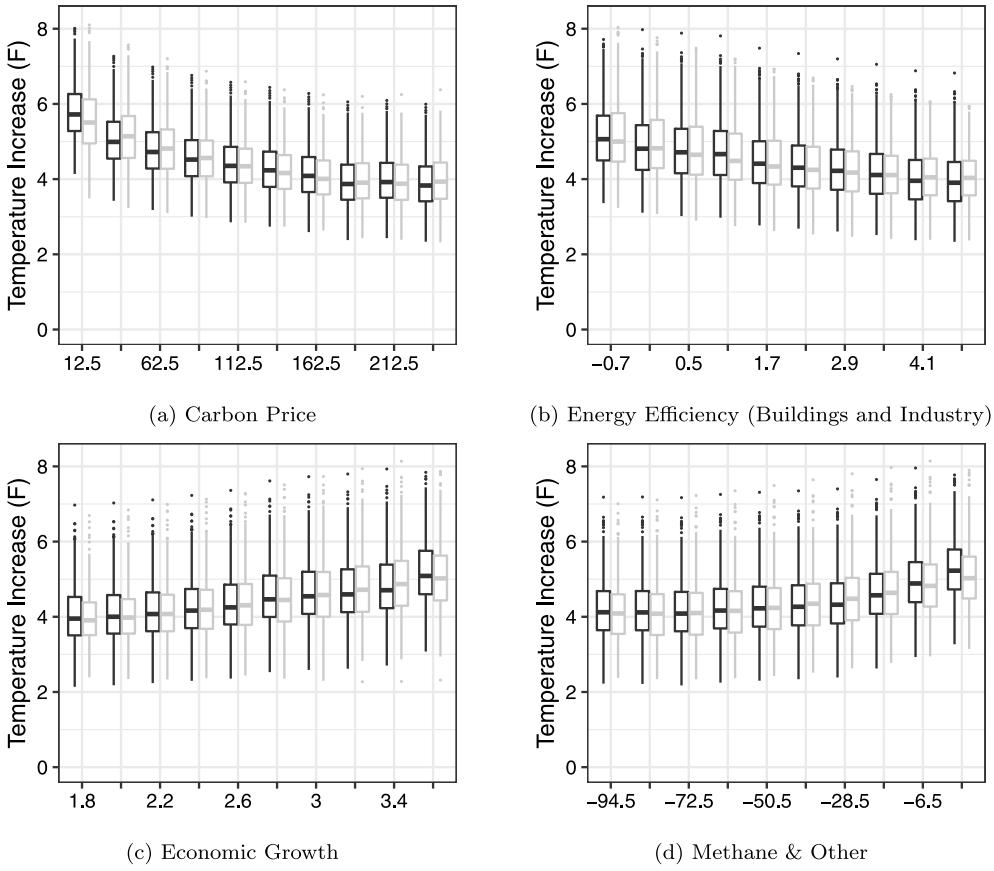
10,000 posterior samples from the distribution in Eq. (5) and 200 trees. For this climate application, we use the BART R package [38].

We computed the first-order, second-order, and total-effects Sobol´ indices of the BART model trained on our collected climate simulator data. Because main effects account for more than 96% of the BART model's total variance, we do not show two-way Sobol´ indices. By taking the mean first-order Sobol´ indices of the 11 input variables over the 10,000 posterior samples, we see in Fig. 5 that carbon price accounts for 35.1% of the BART model's total variance, which is twice as much as the next largest impacts of energy efficiency of buildings and industry at 16.1%, methane and other (which includes nitrous oxide and fluorinated gases) at 14.4%, and economic growth at 13.1%. The total-effects Sobol´ indices imply a similar conclusion. Variable counts fail to provide evidence of such large differences in impacts of the input variables.

For comparison, we also estimated sensitivity indices and created range plots (see [4], Chapter 7) by fitting a constant mean Gaussian process (GP) model (i.e. kriging model; see [39]) to the training data. In Fig. 5, we see that the Sobol´ indices of the trained BART model match those of the trained kriging model quite well. We also see that the one-way counts of the trained BART model poorly matches both the first-order and total-effects Sobol´ indices of the trained kriging model, which supports the hypothesis that the variable count heuristic is not a meaningful input activity measure. In Fig. 6, we show range plots of the four most active input variables. For each input, we approximated the marginal response at each of 10 equally spaced points by varying the remaining inputs using a 2^9-point Sobol´ sequence design [40]. In the left two plots (carbon price and energy efficiency of buildings and industry), the response and its slope decrease with increasing input values. In the right two plots (economic growth and methane & other), the response and its slope increase with increasing input values. Hence, all four of these input variables seem to marginally have diminishing effects on future temperature increase.

To assess prediction accuracy, we predicted temperature increases from an out-of-sample test set of 37 samples, which are chosen manually to achieve a wide range of true temperature increase values. The mean-squared prediction error of the mean BART and GP predictions at the 37 out-of-sample points are, respectively, 0.333 and 0.968. Fig. 7a shows that the BART model accurately predicts temperature increase at, roughly, mid-range values of $f(\mathbf{x}) \in [2.5\ °\text{F}, 7.5\ °\text{F}]$. Outside of this range, the BART model tends to underpredict temperature increase. We suspect this underprediction issue at the upper range is due to the training points having a maximum global temperature increase value of 7.8 °F and hence can be fixed by adding more training samples with extreme response values, which can be done by using combinations of extreme values of the four most active input variables. The underprediction might also be fixed by increasing the prior variance of BART's terminal node parameters as discussed in [2]. However, for this paper we use the default parameter settings in the BART R package [38]. We also see in Fig. 7b that the σ samples appear to be stationary, which implies MCMC convergence.

We conclude that in this no-noise application with $p = 11$ predictors, a sample size of $n = 10p$ suffices for a trained BART model to adequately capture input importance through its Sobol´ indices. Variable counts,



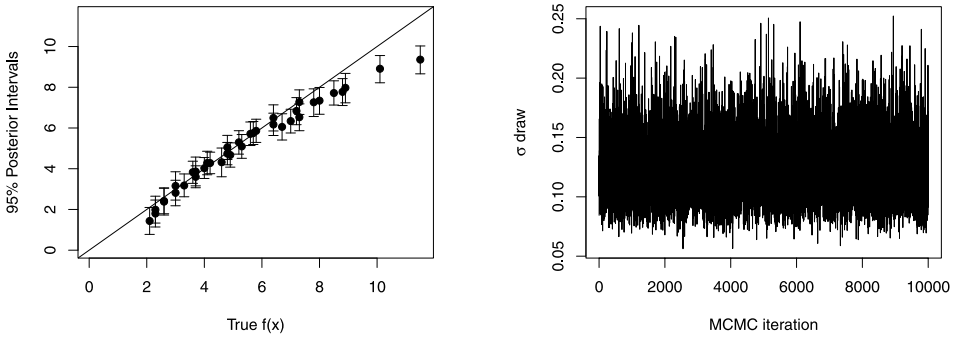
(a) Carbon Price

(b) Energy Efficiency (Buildings and Industry)

(c) Economic Growth

(d) Methane & Other

Fig. 6. Range plots of the four most active input variables. The trained BART model is shown in black while the trained GP model is shown in gray.



(a) Posterior prediction intervals (PPIs) for 37 out-of-sample test values of $f(\mathbf{x})$. For each test input point $\mathbf{x}$, error bars reflect the 0.025 and 0.975 quantiles of the 10,000 posterior predictions of $f(\mathbf{x})$.

(b) Trace plot for posterior draws of σ.

Fig. 7. Diagnostic plots of BART model trained on data from En-ROADS climate simulator.

on the other hand, do not provide enough evidence to convincingly order variables in terms of their importance.

6. Summary and discussion

This paper has provided analytic expressions, explicit interpretations, and computational algorithms for determining Sobol´ indices for BART models. The indices are computed exactly and avoid Monte Carlo approximations. We showed the relationship between Sobol´ indices for BART models and sensitivity indices obtained from one-way counts, which are the predominant way of assessing input activity in BART (see [5,41] among others). Theorem 2 showed that under certain conditions, both the one-way count and the first-order Sobol´ index of variable x_i are functions of the conditional expectation function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m)|X_i=\cdot]$. We then quantify the properties of Sobol´ indices estimated from the BART model for five different analytic functions. First the bias and the uncertainty of the BART-based Sobol´ indices for estimating the true Sobol´ indices for the underlying $f(\cdot)$ are estimated. We find that the bias of the BART-based Sobol´ indices is largest with the multiplicative $g-$function than with any of the additive test functions. Then the rankings of variable activity as measured by the BART-based Sobol´ indices are compared with those provided by one-way counts. To make the second comparisons, we proposed a rank discrepancy d_r to better suit the problem of comparing input

activity assessments. We find that the first-order and total-effects BART-based Sobol´ indices empirically outperform one-way BART counts at capturing, respectively, a function's first-order and total-effects Sobol´ indices.

Finally, we applied our BART-based Sobol´ indices to data generated by the En-ROADS climate simulator to explore how to best reduce future global temperature increases. In particular we note that 32 of the 37 input values in Fig. 7a and 109 of the 110 training inputs result in future global temperature increases above 1.5 °C (2.7 °F), which is the agreed upon upper limit of average global temperature increase above pre-industrial levels set by the 2016 Paris Agreement under the United Nations Framework Convention on Climate Change [42]. In words, the vast majority of policy scenarios described in Section 5 will result in global temperature increases of at least 1.5 °C by the year 2100. Indeed, a 2018 report from the Intergovernmental Panel on Climate Change claims that this temperature increase will likely reach 1.5 °C between 2030 and 2052 if it increases at its current rate [43]. The IPCC report also details the global impact of a 1.5 °C increase. Given these drastic predictions, it is imperative to identify the most impactful factors in minimizing this temperature increase. To achieve a temperature increase below 1.5 °C by 2100, Figs. 5 and 6 suggest maximizing carbon price and the energy efficiency of buildings and industry while minimizing economic growth and the use of methane and other gases (which includes nitrous oxide and fluorinated gases).

This research suggests additional statistical investigations. Linero [41] shows empirically that when a Dirichlet prior is used to generate variable selection probabilities for tree nodes, the posterior probability of an arbitrary inert variable being included in a BART model can drastically shrink. If inert variables are simply not used in a BART model's split rules, then Theorem 4 tells us that computing all Sobol´ indices up to some order will require fewer calculations. Furthermore, excluding inert variables might also improve the accuracy or efficiency of our BART-based Sobol´ indices. These observations suggest comparing the accuracy of our BART-based Sobol´ indices using a Dirichlet prior with those obtained from the default prior as well as the effect of increasing sample sizes.

As has been noted, Bleich et al. [5] and Linero [41], among others, use variable counts in their variable selection methods for BART models. We have seen in the En-ROADS climate simulator example that a trained BART model better captures input activity through its Sobol´ indices rather than through one-way counts. This example suggests that additional research is needed to study the specificity and sensitivity in selecting active inputs for the two methods in order definitively draw this conclusion.

We conclude with two important data/model extensions. Our derivation of the Sobol´ index calculations assumed that the input variables are uncorrelated. This is not true in many applications. For dependent input variables it will be very useful to derive analytic expressions for BART-based Sobol´ indices. Here ideas from [44–46], who all discuss various ways to estimate Sobol´ sensitivity indices in the dependent input variable case will be of use. Finally, we note that most of the results in the paper should extend to other tree ensemble methods, such as the random forest method described in [47].

Implementation of BART-based first-order, second-order, and total-effects Sobol´ indices can be found in the Open Bayesian Trees (OpenBT) project at https://bitbucket.org/mpratola/openbt/ [22].

CRediT authorship contribution statement

Akira Horiguchi: Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Matthew T. Pratola:** Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Supervision. **Thomas J. Santner:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Professor Joseph Verducci for introducing A.H. to ranking models. We also thank the two referees and an Associate Editor for their comments, which have improved this paper.

Funding

This work was supported by the Graduate School at The Ohio State University, USA; the National Science Foundation, USA [Agreements DMS-1916231, DMS-0806134, DMS-1310294]; the King Abdullah University of Science and Technology (KAUST), Saudi Arabia Office of Sponsored Research (OSR) [Award No. OSR-2018-CRG7-3800.3]; and the Isaac Newton Institute for Mathematical Sciences, United Kingdom.

Appendix A. Proofs of theorems

Proof of Theorem 1. According to Eq. (10), the first step to computing the Sobol´ index for tree function $g(\cdot;\mathcal{T},\mathcal{M})$ and variable index set $P \subset \{1,2,\ldots,p\}$ is to compute the conditional expectation $\mathbb{E}_{\mathbf{X}_{-P}}[g(\mathbf{X};\mathcal{T},\mathcal{M}) \mid \mathbf{X}_P]$. By taking the appropriate conditional expectation of both sides of Eq. (3), we get

$$\mathbb{E}_{\mathbf{X}_{-P}}[g(\mathbf{X};\mathcal{T},\mathcal{M}) \mid \mathbf{X}_P] = \sum_{k=1}^{|\mathcal{M}|} d_k^{-P} \mathbf{1}_{\mathbf{R}_k^P}(\mathbf{X}_P), \tag{A.1}$$

where hyperrectangle $\mathbf{R}_k^P = \prod_{i\in P} I_k^i$, and coefficients $d_k^{-P} = \mu_k \mathbb{P}_{-P}(\mathbf{R}_k^{-P})$. Due to Assumption A.1, the coefficient expression simplifies to

$$d_k^{-P} = \mu_k \prod_{j\notin P} \mathbb{P}_j(I_k^j). \tag{A.2}$$

According to Eq. (10), the first step to computing the Sobol´ index for ensemble function $\mathcal{E}(\cdot;\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m)$ and variable index set $P \subset \{1,2,\ldots,p\}$ is to compute the conditional expectation $\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]$. By taking the appropriate conditional expectation of both sides of Eq. (4), using linearity of expectations, and plugging in Eq. (A.1), we get

$$\begin{aligned}\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P] &= \sum_{t=1}^m \mathbb{E}_{\mathbf{X}_{-P}}[g(\mathbf{X};\mathcal{T}_t,\mathcal{M}_t) \mid \mathbf{X}_P] \\ &= \sum_{t=1}^m \sum_{k=1}^{|\mathcal{M}_t|} d_{tk}^P \mathbf{1}_{\mathbf{R}_{tk}^P}(\mathbf{X}_P).\end{aligned}$$

It is more convenient to view this conditional expectation as a single sum over the ensemble's terminal nodes rather than as a double sum as shown above. Hence, we can express this conditional expectation as

$$\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P] = \sum_{k\in B_\mathcal{E}} d_k^{-P} \mathbf{1}_{\mathbf{R}_k^P}(\mathbf{X}_P), \tag{A.3}$$

where $B_\mathcal{E} = \cup_{t=1}^m B_{\mathcal{T}_t}$ is the index set over the terminal nodes of the trees in ensemble $\mathcal{E}$.

Finally we are able to compute the variance terms in Eq. (10) for general variable index set $P \subset \{1,2,\ldots,p\}$. First, we compute $\mathtt{Var}_{\mathbf{X}_P}\left(\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]\right)$. Into this term we plug in Eq. (A.3), apply the general result $\mathtt{Var}(U) = \mathtt{Cov}(U,U)$ for generic random variable U, and use the bilinearity property of covariance to get

$$\begin{aligned}&\mathtt{Var}_{\mathbf{X}_P}\left(\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]\right) \\ &= \sum_{k\in B_\mathcal{E}} \sum_{l\in B_\mathcal{E}} d_k^{-P} d_l^{-P} \mathtt{Cov}_{\mathbf{X}_P}\left(\mathbf{1}_{\mathbf{R}_k^P}(\mathbf{X}_P), \mathbf{1}_{\mathbf{R}_l^P}(\mathbf{X}_P)\right),\end{aligned}$$

where the coefficients d_k^{-P} and d_l^{-P} are defined in Eq. (A.2). To each covariance term, which we will denote as $C_{k,l}^P$, we can apply the elementary covariance result $\mathtt{Cov}(U,V) = \mathbb{E}UV - \mathbb{E}U\mathbb{E}V$ for generic random variables U and V to get

$$C_{k,l}^P = \mathbb{P}_P(\mathbf{R}_k^P \cap \mathbf{R}_l^P) - \mathbb{P}_P(\mathbf{R}_k^P)\mathbb{P}_P(\mathbf{R}_l^P).$$

Thus,

$$\mathtt{Var}_{\mathbf{X}_P}\left(\mathbb{E}_{\mathbf{X}_{-P}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]\right) = \sum_{k\in B_\mathcal{E}}\sum_{l\in B_\mathcal{E}} d_k^{-P} d_l^{-P} C_{k,l}^P.$$

In particular, when $P = \{i\}$, then

$$\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid X_i] = \sum_{k\in B_\mathcal{E}} d_k^{-i}\mathbf{1}_{I_k^i}(X_i),$$

where $d_k^{-i} = \mu_k \prod_{j\neq i}\mathbb{P}_j(I_k^j)$. The first-order Sobol´ index in Eq. (10) then becomes

$$V_i = \sum_{k\in B_\mathcal{E}}\sum_{l\in B_\mathcal{E}} d_k^{-i} d_l^{-i} C_{k,l}^i,$$

where

$$C_{k,l}^i = \mathbb{P}_i(I_k^i \cap I_l^i) - \mathbb{P}_i(I_k^i)\mathbb{P}_i(I_l^i).$$

Proof of Theorem 2. Consider a BART ensemble $\mathcal{E}_0$ with m regression trees, where each tree is simply a terminal node with one terminal node parameter. Thus, the ensemble $\mathcal{E}_0$ predicts the same value for any input $\mathbf{x} \in D$ and is hence a constant-mean model. Then any BART ensemble $\mathcal{E}$ with m regression trees can be thought of as $\mathcal{E}_0$ having undergone a sequence of birth processes. Any birth process slices a terminal node's corresponding hyperrectangle into two smaller hyperrectangles according to some split rule. If we call this split rule "$x_i < c$", then this slice occurs on the $(p-1)$-dimensional hyperplane $x_i = c$ in D. The resulting "left" hyperrectangle gains a terminal node parameter μ_{left} while the resulting "right" hyperrectangle gains a terminal node parameter μ_{right}. Thus, if prior to the birth process the piecewise-constant function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$ is constant at $x_i = c$ (i.e. the split rule "$x_i < c$" does not already exist in the ensemble) and $\mu_{left} \neq \mu_{right}$ (which is true almost surely but is also ensured through assumption A.3), then the birth process produces a jump in the piecewise-constant function at $x_i = c$. Meanwhile, the birth process does not produce a jump in any of the other piecewise-constant functions $\mathbb{E}_{\mathbf{X}_{-j}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid X_j = \cdot]$ (for $j \neq i$). Hence, under the mentioned conditions, each birth process that produces a unique split rule that involves variable x_i increments the number of jumps in the piecewise-constant function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$ by one.

Proof of Theorem 3. We have the following transformations to the conditional expectation function $\mathbb{E}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid X_i = \cdot]$:

1. First, center and scale the $e_{k^*}^i$. Let

   $$\tilde{e}^i{}_{k^*} = \sqrt{|B_\mathcal{E}^i|}\frac{e_{k^*}^i - \bar{e}_\cdot^i}{s},$$

   where $\bar{e}_\cdot^i = |B_\mathcal{E}^i|^{-1}\sum_{k^*\in B_\mathcal{E}^i} e_{k^*}^i$ and $s^2 = [\sum_{k^*\in B_\mathcal{E}^i}(e_{k^*}^i - \bar{e}_\cdot^i)^2]/(|B_\mathcal{E}^i| - 1)$ is the corrected sample variance of the $\{e_{k^*}^i\}$. Note that for any two indices $k^*, l^* \in B_\mathcal{E}^i$, the relation $\tilde{e}^i{}_{k^*} = \tilde{e}^i{}_{l^*}$ holds if and only if $e_{k^*}^i = e_{l^*}^i$.
2. Second, assign equal probability mass $|B_\mathcal{E}^i|^{-1}$ to each $I_{k^*}^i$. Introduce new intervals $\tilde{I}^i{}_{k^*}$ by shifting and scaling $I_{k^*}^i$ so that
   (a) $\{\tilde{I}^i{}_{k^*}\}_{k^*\in B_\mathcal{E}^i}$ still partitions I_D^i into exactly $|B_\mathcal{E}^i|$ sets, and
   (b) $\mathbb{P}_i(\tilde{I}^i{}_{k^*}) = |B_\mathcal{E}^i|^{-1}$ for all $k^* \in B_\mathcal{E}^i$.

Now define $\tilde{h}_\mathcal{E}^i(X_i) := \tilde{\mathbb{E}}_{\mathbf{X}_{-i}}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid X_i] = \sum_{k^*\in B_\mathcal{E}^i}\tilde{e}^i{}_{k^*}\mathbf{1}_{\tilde{I}^i{}_{k^*}}(\boldsymbol{X}_i)$. Using previous definitions, we have

$$\mathtt{Var}_{X_i}\left(\tilde{h}_\mathcal{E}^i(X_i)\right) = \mathtt{Var}_{X_i}\left(\sum_{k^*\in B_\mathcal{E}^i}\tilde{e}^i{}_{k^*}\mathbf{1}_{\tilde{I}^i{}_{k^*}}(X_i)\right)$$

$$= \sum_{k^*\in B_\mathcal{E}^i}\sum_{l^*\in B_\mathcal{E}^i}\tilde{e}^i{}_{k^*}\tilde{e}^i{}_{l^*}\left[\mathbb{P}_i(\tilde{I}^i{}_{k^*}\cap\tilde{I}^i{}_{l^*}) - \mathbb{P}_i(\tilde{I}^i{}_{k^*})\mathbb{P}_i(\tilde{I}^i{}_{l^*})\right].$$

Recall that the intervals $\tilde{I}^i{}_{k^*}$ still partition the original domain I_D^i. So if $k^* \neq l^*$, then $\tilde{I}^i{}_{k^*}\cap\tilde{I}^i{}_{l^*} = \emptyset$ and hence $\mathbb{P}_i(\tilde{I}^i{}_{k^*}\cap\tilde{I}^i{}_{l^*}) = 0$. Thus,

$$\sum_{k^*\in B_\mathcal{E}^i}\sum_{l^*\in B_\mathcal{E}^i}\tilde{e}^i{}_{k^*}\tilde{e}^i{}_{l^*}\mathbb{P}_i(\tilde{I}^i{}_{k^*}\cap\tilde{I}^i{}_{l^*}) = \sum_{k^*\in B_\mathcal{E}^i}(\tilde{e}^i{}_{k^*})^2\mathbb{P}_i(\tilde{I}^i{}_{k^*}).$$

Since each interval $\tilde{I}^i{}_{k^*}$ has equal probability mass, each $\mathbb{P}_i(\tilde{I}^i{}_{k^*})$ becomes simply $|B_\mathcal{E}^i|^{-1}$. So then

$$\mathtt{Var}_{X_i}\left(\tilde{h}_\mathcal{E}^i(X_i)\right) = |B_\mathcal{E}^i|^{-1}\sum_{k^*\in B_\mathcal{E}^i}(\tilde{e}^i{}_{k^*})^2 - |B_\mathcal{E}^i|^{-2}\sum_{k^*\in B_\mathcal{E}^i}\sum_{l^*\in B_\mathcal{E}^i}\tilde{e}^i{}_{k^*}\tilde{e}^i{}_{l^*}.$$

Note that the coefficients $\tilde{e^i_{k^*}}$ are centered so that the sum $\sum_{k^*\in B_\mathcal{E}^i}\tilde{e^i_{k^*}}$ (and hence the double sum term in the preceding equation) equals zero. Also note that the $\tilde{e^i_{k^*}}$ are scaled so that $\sum_{k^*\in B_\mathcal{E}^i}(\tilde{e^i_{k^*}})^2 = |B_\mathcal{E}^i|(|B_\mathcal{E}^i| - 1)$. We then have

$$\mathtt{Var}_{X_i}\left(\tilde{h}_\mathcal{E}^i(X_i)\right) = |B_\mathcal{E}^i| - 1.$$

Now let $C_\mathcal{E}^i$ be the set of all unique cutpoints involved in any split rule in $\mathcal{E}$ that includes variable x_i. Recall that the set $B_\mathcal{E}^i$ indexes a set of intervals that partition the domain's ith margin, i.e. the set $B_\mathcal{E}^i$ indexes the set of intervals $I_{k^*}^i = [\gamma_1,\gamma_2)$ (or $[\gamma_1,\gamma_2]$ if $\gamma_2 = b_D^i$), where γ_1 and γ_2 are any two consecutive (in value) points in $C_\mathcal{E}^i \cup \{a_D^i, b_D^i\}$. Between any two such partitioning intervals must be a cutpoint in $C_\mathcal{E}^i$. Furthermore, by the assumption that $e_{k^*}^i = e_{l^*}^i$ implies $I_{k^*}^i = I_{l^*}^i$ for any indices $k^*, l^* \in B_\mathcal{E}^i$, all cutpoints in $C_\mathcal{E}^i$ must lie between two such partitioning intervals. That is, no cutpoint in $C_\mathcal{E}^i$ can lie in the interior of any such partitioning interval. Thus, $|B_\mathcal{E}^i| - 1 = |C_\mathcal{E}^i|$, which is simply the number of unique split rules in $\mathcal{E}$ that include variable x_i and hence, by Theorem 2, equals the number of jumps in the original conditional expectation function, which equals the number of jumps in the transformed conditional expectation function.

Proof of Theorem 4. For any terminal node $k \in B_\mathcal{E}$ where $v(k)\cap P = \emptyset$, the random quantity $\mathbb{E}[\mu_k\mathbf{1}_{\mathbf{R}_k}(\mathbf{X}) \mid \mathbf{X}_P]$ is constant almost surely. Therefore,

$$\begin{aligned}&\mathtt{Var}_{\mathbf{X}_P}(\mathbb{E}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) - \mu_k\mathbf{1}_{\mathbf{R}_k}(\mathbf{X}) \mid \mathbf{X}_P])\\ &= \mathtt{Var}_{\mathbf{X}_P}(\mathbb{E}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P] - \mathbb{E}[\mu_k\mathbf{1}_{\mathbf{R}_k}(\mathbf{X}) \mid \mathbf{X}_P])\\ &= \mathtt{Var}_{\mathbf{X}_P}(\mathbb{E}[\mathcal{E}(\mathbf{X};\{\mathcal{T}_t,\mathcal{M}_t\}_{t=1}^m) \mid \mathbf{X}_P]).\end{aligned}$$

The result of the theorem will follow by applying this argument to all such terminal nodes.

Proof of Theorem 5. In general, the Kemeny–Snell (KS) distance between rankings $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_p)$ and $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)$ is defined to be

$$d_{KS}(\alpha,\beta) = \frac{1}{2}\sum_{i=1}^p\sum_{j=1}^p |A_{ij} - B_{ij}|$$

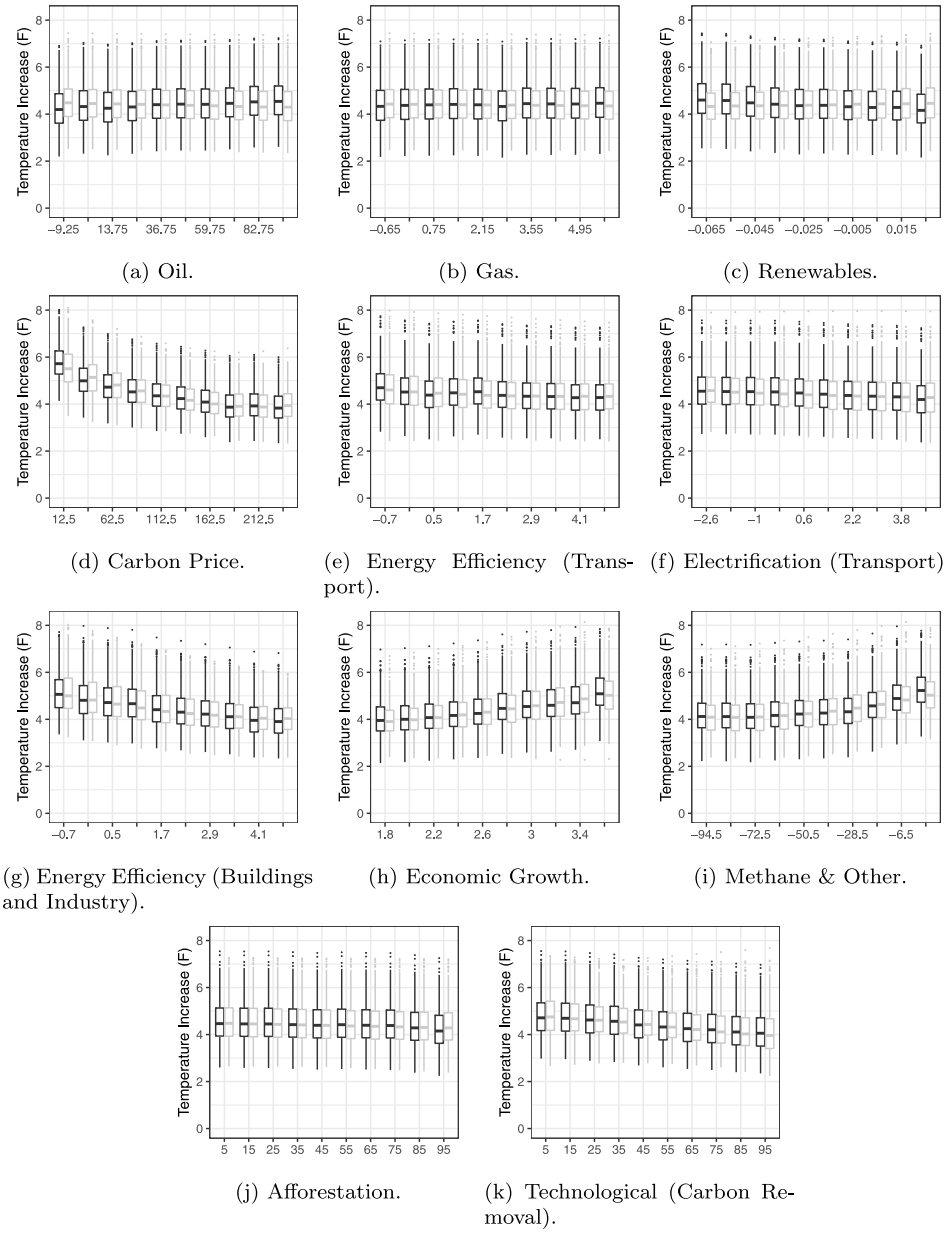
where

$$A_{ij} = \begin{cases}1 & \boldsymbol{\alpha} \text{ prefers object } i \text{ to object } j\\ -1 & \boldsymbol{\alpha} \text{ prefers object } j \text{ to object } i\\ 0 & \boldsymbol{\alpha} \text{ prefers objects } i \text{ and } j \text{ equally}\end{cases}$$

and B_{ij} is similarly defined for ranking $\boldsymbol{\beta}$.

For the rest of the proof, we will assume that rankings $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ each have no ties. We will also take ranking $\boldsymbol{\alpha}$ to be the reference vector and hence will, without loss of generality, assume $\boldsymbol{\alpha} = (1,2,\ldots,p)$. We will also refer to the sum in Eq. (15) as the discrepancy d_r. Finally, we will prove desired equality via induction.

We first note that these assumptions greatly simplify the KS distance. If we think of the values A_{ij} (similarly B_{ij}) as constituting a $p\times p$ matrix A (similarly B) whose ij entry is A_{ij} (similarly B_{ij}), then



(a) Oil. (b) Gas. (c) Renewables.

(d) Carbon Price. (e) Energy Efficiency (Transport). (f) Electrification (Transport).

(g) Energy Efficiency (Buildings and Industry). (h) Economic Growth. (i) Methane & Other.

(j) Afforestation. (k) Technological (Carbon Removal).

Fig. B.8. Range plots for each of the 11 input variables.

both matrices A and B are antisymmetric, which implies $|A_{ij} - B_{ij}| = |A_{ji} - B_{ji}|$ for all $i, j = 1, \dots, p$ and $A_{ij} = B_{ij} = 0$ if $i = j$. Therefore, we may reformulate the KS distance as

$$d_{KS}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i<j} |A_{ij} - B_{ij}|.$$

We now proceed with the proof by induction. Suppose $p = 2$. Half the KS distance is then $\frac{1}{2}|A_{12} - B_{12}|$, where $A_{12} = 1$, while the AH distance becomes W_1 (since $W_p = 0$ by default). One of two cases may occur. If $\beta_1 < \beta_2$, then $\beta_1 = 1$ and $\beta_2 = 2$. In this case, both the AH distance and half the KS distance are zero. If $\beta_1 > \beta_2$, then $\beta_1 = 2$ and $\beta_2 = 1$. In this case, both the AH distance and half the KS distance are unity. We note that values β_1 and β_2 must be distinct due to ranking $\boldsymbol{\beta}$ having no ties. Thus, the induction hypothesis holds for $p = 2$.

Now suppose the induction hypothesis holds for arbitrary $p - 1 \geq 3$. The KS distance can be decomposed into

$$d_{KS}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = d_{KS}(\boldsymbol{\alpha}_{-1}, \boldsymbol{\beta}_{-1}) + \sum_{j=2}^{p} |A_{1j} - B_{1j}|,$$

where we define $\boldsymbol{\alpha}_{-1} := (\alpha_2, \dots, \alpha_p)$ and $\boldsymbol{\beta}_{-1}$ similarly for ranking $\boldsymbol{\beta}$. The discrepancy d_r, due to its stagewise nature, can also be decomposed:

$$d_{AH}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = W_1 + d_{AH}(\boldsymbol{\alpha}_{-1}, \boldsymbol{\beta}_{-1}),$$

where $W_1 = \beta_1 - 1$ by default. By assumption, half the KS distance between $\boldsymbol{\alpha}_{-1}$ and $\boldsymbol{\beta}_{-1}$ equals the discrepancy d_r between the same two quantities. Hence, we need only prove that $\frac{1}{2}\sum_{j=2}^{p} |A_{1j} - B_{1j}| = \beta_1 - 1$ to complete the proof.

First, we note that $A_{1j} = 1$ for all $j > 1$ and, since B_{1j} is either 1 or -1, the quantity $A_{1j} - B_{1j}$ is nonnegative. Thus, $|A_{1j} - B_{1j}| = 1 - B_{1j}$ for all $j > 1$. But B_{1j} is simply $\mathbf{1}_{\beta_1 < \beta_j} - \mathbf{1}_{\beta_1 > \beta_j}$. Hence, $\sum_{j=2}^{p} B_{1j} = (p - \beta_1) - (\beta_1 - 1) = p - 2\beta_1 + 1$. Therefore, $\frac{1}{2}\sum_{j=2}^{p} |A_{1j} - B_{1j}| = \beta_1 - 1$.

Proof of Theorem 6. Suppose we have computed discordances $W_1, \dots, W_{k-1}$ for some $k > q_0$ and wish to compute discordance W_k. Then the remaining considered items, each having input activity measure values of zero, all have ranking number 1 in ranking ρ_f. Since at least one remaining considered item has ranking number 1 in ranking $\rho_{\mathcal{E}}$, we get $W_k = 1 - 1 = 0$.

Proof of Theorem 7. We will partition the discordances into three sets: $\{W_1, \ldots, W_{j-1}\}$, $\{W_j \ldots, W_{j+u-1}\}$, and $\{W_{j+u}, \ldots, W_{|\rho_f|}\}$. After letting, for all $k = 1, \ldots, j-1, j+u, \ldots, |\rho_f|$, item i_k be the item removed from consideration after computing W_k but (if $k < |\rho_f|$) before computing W_{k+1}, we will then prove the desired invariance to permutation ϕ for the three sets of discordances.

First, consider discordances $W_1, \ldots, W_{j-1}$. These discordances depend only on the ranking numbers of items $i_1, \ldots, i_{j-1}$ in ρ_f and in $\rho_{\mathcal{E}}$. Because these ranking numbers are invariant to choice of permutation ϕ, these discordances are also invariant to ϕ.

Now consider discordances $W_{j+u}, \ldots, W_{|\rho_f|}$. Similar to the previous set of discordances, these discordances depend only on the ranking numbers of items $i_{j+u}, \ldots, i_{|\rho_f|}$ in ρ_f and in $\rho_{\mathcal{E}}$. Because these ranking numbers are invariant to choice of permutation ϕ, these discordances are also invariant to ϕ.

Finally, consider discordances $W_j \ldots, W_{j+u-1}$. Because items $i_{j+1}, \ldots, i_{j+u}$ (and no other items) have ranking number j in ρ_f, these discordance values are $W_k = r_{(k)} - j + 1$ for $k = j, \ldots, j+u-1$, where $r_{(j)}, \ldots, r_{(j+u-1)}$ are the order statistics of ranking numbers $r_j, \ldots, r_{j+u-1}$. Because order statistics are invariant to permutations of the statistic values, these discordances are invariant to permutation ϕ.

Appendix B. Figures for En-ROADS climate simulator

See Fig. B.8.

References

[1] Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. Ann Appl Stat 2010;4:266–98.
[2] Chipman H, Ranjan P, Wang W. Sequential design for computer experiments with a flexible Bayesian additive model. Canad J Statist 2012;40(4):663–78.
[3] Gramacy RB, Haaland B. Speeding up neighborhood search in local Gaussian process prediction. Technometrics 2016;58(3):294–303.
[4] Santner TJ, Williams BJ, Notz WI. The design and analysis of computer experiments. Second edn.. Springer-Verlag; 2018, p. 436.
[5] Bleich J, Kapelner A, George EI, Jensen ST. Variable selection for BART: An application to gene regulation. Ann Appl Stat 2014;8(3):1750–81. http://dx.doi.org/10.1214/14-AOAS755.
[6] Liu Y, Ročková V, Wang Y. Abc variable selection with Bayesian forests. 2018, arXiv preprint arXiv:1806.02304.
[7] Sobol´ IM. Sensitivity estimates for nonlinear mathematical models. MMCE 1993;1(4):407–14.
[8] Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc Ser B Stat Methodol 2004;66(3):751–69.
[9] Chen W, Jin R, Sudjianto A. Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. 2005.
[10] Chen W, Jin R, Sudjianto A. Analytical global sensitivity analysis and uncertainty propagation for robust design. J Qual Technol 2006;38(4):333–48.
[11] Marrel A, Iooss B, Laurent B, Roustant O. Calculations of Sobol´ indices for the Gaussian process metamodel. Reliab Eng Syst Saf 2009;94(3):742–51.
[12] Moon H. Design and analysis of computer experiments for screening input variables (Ph.D. thesis), The Ohio State University; 2010.
[13] Svenson J, Santner T, Dean A, Moon H. Estimating sensitivity indices based on Gaussian process metamodels with compactly supported correlation functions. J Statist Plann Inference 2014;144:160–72.
[14] Sudret B. Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst saf 2008;93(7):964–79.
[15] Gramacy RB, Taddy M, Wild SM, et al. Variable selection and sensitivity analysis using dynamic trees, with an application to computer code performance tuning. Ann Appl Stat 2013;7(1):51–80.
[16] Gramacy RB, Taddy M. Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models. J Stat Softw 2010;33(6):1–48, URL: http://www.jstatsoft.org/v33/i06/.
[17] Da Veiga S, Wahl F, Gamboa F. Local polynomial estimation for sensitivity analysis on models with correlated inputs. Technometrics 2009;51(4):452–63.
[18] Wei P, Lu Z, Song J. A comprehensive comparison of two variable importance analysis techniques in high dimensions: Application to an environmental multi-indicators system. Environ Model Softw 2015;70:178–90.
[19] Chipman HA, George EI, McCulloch RE. Bayesian CART model search. J Amer Statist Assoc 1998;93:935–48.
[20] Saltelli A, Chan K, Scott M, et al. Sensitivity analysis. Probability and statistics series. New York: John and Wiley & Sons; 2000.
[21] Carnell R. Lhs: Latin hypercube samples. 2019, URL: https://CRAN.R-project.org/package=lhs, R package version 1.0.1.
[22] Pratola MT. Open Bayesian trees. 2020, URL: https://bitbucket.org/mpratola/openbt/src/master/. Accessed: 2020-08-29.
[23] Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991;19(1):1–67.
[24] Saltelli A, Sobol´ IM. About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Saf 1995;50(3):225–39. http://dx.doi.org/10.1016/0951-8320(95)00099-2, URL: http://www.sciencedirect.com/science/article/pii/0951832095000992.
[25] Crestaux T, Maître OL, Martinez J-M. Polynomial chaos expansion for sensitivity analysis. Reliab Eng Syst Saf 2009;94(7):1161–72. http://dx.doi.org/10.1016/j.ress.2008.10.008, URL: http://www.sciencedirect.com/science/article/pii/S0951832008002561. Special Issue on Sensitivity Analysis.
[26] Bratley P, Fox BL, Niederreiter H. Implementation and tests of low-discrepancy sequences. ACM Trans Model Comput Simul (TOMACS) 1992;2(3):195–213.
[27] Kucherenko S, Feil B, Shah N, Mauntz W. The identification of model effective dimensions using global sensitivity analysis. Reliab Eng Syst Saf 2011;96(4):440–9.
[28] Morris MD, Moore LM, McKay MD. Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs. J Statist Plann Inference 2006;136(9):3203–20.
[29] Gramacy RB. Bayesian treed Gaussian process models. Santa Cruz: University of California; 2005.
[30] Gramacy RB, Lee HKH. Bayesian treed Gaussian process models with an application to computer modeling. J Amer Statist Assoc 2008;103(483):1119–30.
[31] Gramacy RB. Tgp: An r package for Bayesian nonstationary, semiparametric nonlinear regression and design by treed Gaussian process models. J Stat Softw 2007;19(9):1–46, URL: http://www.jstatsoft.org/v19/i09/.
[32] Kendall MG. Rank correlation methods. Griffin; 1948.
[33] Heiser WJ, D'Ambrosio A. Clustering and prediction of rankings within a kemeny distance framework. In: Algorithms from and for Nature and Life. Springer; 2013, p. 19–31.
[34] Emond EJ, Mason DW. A new rank correlation coefficient with application to the consensus ranking problem. J Multi-Criteria Decis Anal 2002;11(1):17–28.
[35] Kemeny JG, Snell L. Preference ranking: an axiomatic approach. Math Models Soc Sci 1962;9–23.
[36] Fligner MA, Verducci JS. Multistage ranking models. J Amer Stat Assoc 1988;83(403):892–901.
[37] Climate Interactive, Ventana Systems, Todd Fincannon, UML Climate Change Initiative, MIT Sloan. En-ROADS climate change solutions simulator. 2020, URL: https://en-roads.climateinteractive.org/scenario.html?v=2.7.15. Accessed: 2020-04-03.
[38] McCulloch R, Sparapani R, Gramacy R, Spanbauer C, Pratola M. BART: Bayesian additive regression trees. 2019, URL: https://CRAN.R-project.org/package=BART. R package version 2.6.
[39] Cressie N. Statistics for spatial data. John Wiley & Sons; 2015.
[40] Kuo F, Joe S. Sobolsequence: Sobol sequences with better two-dimensional projections. 2017, URL: https://CRAN.R-project.org/package=SobolSequence. R package version 1.0.
[41] Linero AR. Bayesian regression trees for high-dimensional prediction and variable selection. J Amer Statist Assoc 2018;113(522):626–36. http://dx.doi.org/10.1080/01621459.2016.1264957, URL: arXiv:https://doi.org/10.1080/01621459.2016.1264957.
[42] UNFCCC. Adoption of the Paris agreement. 2015, URL: https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.
[43] IPCC. Summary for policymakers. In: Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, Connors S, Matthews JBR, Chen Y, Zhou X, Gomis MI, Lonnoy E, Maycock T, Tignor M, Waterfield T, editors. Global Warming of 1.5C. An IPCC Special Report on the impacts of global warming of 1.5C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Geneva, Switzerland: World Meteorological Organization; 2018, URL: https://www.ipcc.ch/sr15/.
[44] Kucherenko S, Tarantola S, Annoni P. Estimation of global sensitivity indices for models with dependent variables. Comput Phys Commun 2012;183(4):937–46.
[45] Mara TA, Tarantola S. Variance-based sensitivity indices for models with dependent inputs. Reliab Eng Syst Saf 2012;107:115–21.
[46] Glen G, Isaacs K. Estimating Sobol´ sensitivity indices using correlations. Environ Model Softw 2012;37:157–66.
[47] Breiman L. Random forests. Mach Learn 2001;45:5–32.