

Comparison of tree-based ensemble models for regression

Sangho Park^a, Chanmin Kim^{1,a}

^aDepartment of Statistics, Sungkyunkwan University, Korea

Abstract

When multiple classifications and regression trees are combined, tree-based ensemble models, such as random forest (RF) and Bayesian additive regression trees (BART), are produced. We compare the model structures and performances of various ensemble models for regression settings in this study. RF learns bootstrapped samples and selects a splitting variable from predictors gathered at each node. The BART model is specified as the sum of trees and is calculated using the Bayesian backfitting algorithm. Throughout the extensive simulation studies, the strengths and drawbacks of the two methods in the presence of missing data, high-dimensional data, or highly correlated data are investigated. In the presence of missing data, BART performs well in general, whereas RF provides adequate coverage. The BART outperforms in high dimensional, highly correlated data. However, in all of the scenarios considered, the RF has a shorter computation time. The performance of the two methods is also compared using two real data sets that represent the aforementioned situations, and the same conclusion is reached.

Keywords: Bayesian additive regression trees, random forest, missingness, high-dimensional data, multicollinearity

1. Introduction

A tree-based model does not require a parametric assumption or knowledge of the structure of the independent and response variables (Kuhn and Johnson, 2013) and can be fitted to non-linear or non-additive data structures (Kern *et al.*, 2019). The tree-based ensemble model incorporates several classification and regression trees (CART) that are the basis of the tree-based model to improve the predictive performance of tree-based models. Among others, the paper focuses on two ensemble models, random forest (RF) (Breiman, 2001) and Bayesian additive regression trees (BART) (Chipman *et al.*, 2010).

The CART of RF fit bootstrapped samples and randomly draw predictors at each node. The CART then chooses a split variable from the sampled predictors (Breiman, 2001). The CART's correlation is reduced because predictors are sampled at random. When compared to the original CART, RF improves predictive performance over a single CART by lowering prediction variance (Kuhn and Johnson, 2013). The RF calculates confidence intervals based on out-of-bag (OOB) prediction errors (Zhang *et al.*, 2019). The BART is a hybrid of several Bayesian CARTs proposed by Chipman *et al.* (1998). In each Bayesian CART, the prior and likelihood must be specified, and the posterior

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. NRF-2020R1F1A1A01048168)

¹ Corresponding author: Department of Statistics, SungKyunKwan University, 25-2 Seonggyunwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail: chanmin.kim@skku.edu

distribution can be inferred using the probabilistic search of Markov chain Monte Carlo (MCMC). The Bayesian backfitting algorithm is used to fit each Bayesian CART, and credible intervals are calculated using posterior samples (Chipman *et al.*, 2010).

As variations, methods for improving the performance of RF and BART in various data situations have been proposed. The RF can impute missing data using either the proximity matrix method or the missForest model (Stekhoven and Bühlmann, 2012). The missForest model fits observed data and imputes data based on predicted missing data (Stekhoven and Bühlmann, 2012). The BART's variant, bartMachine, on the other hand, employs missingness as a split criterion (Kapelner and Bleich, 2016).

The number of predictors in a high-dimensional situation is relatively large in comparison to the number of samples. The sparsity assumption states that the number of significant predictors is significantly less than the number of predictors (Bühlmann and Van De Geer, 2011). To address this situation, methods for understanding the relationship between the predictor and the response variable have been proposed. Breiman (2001) introduced variable importance as a variable selection method by permuting independent variables and computing changes in OOB error. When categorical and continuous variables coexist in a classification task, the importance of the variable is skewed (Strobl *et al.*, 2007). Linero (2018) proposed a BART variant (DART) that uses the Dirichlet distribution as a prior on selection probabilities in BART. The posterior inclusion probabilities (PIPs) can be used to determine significant variables in this method.

In a highly correlated and multidimensional situation, RF chooses a nonsignificant variable as a split variable with little effect on the response variable (Waldmann, 2016). In this case, MCMC in BART may be stuck in the posterior mode (Linero, 2018). We can expect poor model performance in both methods when high-dimensional data is highly correlated. Despite the excellent performance and numerous applications of RF and BART, there has been insufficient attention paid to the comparison of the two methods in the aforementioned special cases. As a result, in this paper, we compare the performance of RF and BART in such situations. This paper focuses on the following scenarios in BART and RF applications.

1. Missing data situation: There are one or more missing independent variables. We also consider three missing data mechanisms in this scenario: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).
2. High dimensional data situation: The number of independent variables (p) is relatively large compared to the number of observations (n). We look at three scenarios: $n < p$ (simulation data), $n = p$ (simulation data), and $n > p$ (simulation data, real data).
3. Highly correlated data situation: independent variables are highly correlated.

In the simulation study section, we use data-generating models from the existing literature to create the three scenarios mentioned above. These models were used to evaluate BART or RF or to compare the performance of the two methods in a high-dimensional data setting using only mean square error. We thoroughly evaluate the performance of two methods and their variants using a variety of metrics, including bias, mean square error, relative bias, coverage, and computation time. We also run the test using two real-world data sets in the same three scenarios.

The following is how this paper is structured. Section 2 provides a brief overview of RF and BART, and Section 3 compares the two methods in three different scenarios: missing data, high-dimensional predictors, and highly correlated data. Section 4 presents the simulation results. In Section 5, real-world data sets are used to compare model performance. Section 6 discusses findings and future research.

2. Random forest and Bayesian additive regression trees

2.1. Random forest

The RF model is made up of a series of CART. The CART divides data based on the criterion that maximizes the reduction of an impurity measure. The RF prediction value is the average of the CART prediction values. The CART learns bootstrapped samples at each node and chooses the split variable from sampled features. The MSE can be calculated from each CART that learns different subsets of data (Breiman *et al.*, 1984). It's also known as an OOB error (Breiman, 2001), which is formally defined as follows:

$$\text{MSE}_{\text{ooB}} = \frac{1}{n} \sum_i (Y_i - \hat{Y}_i^{\text{ooB}})^2,$$

where Y_i and \hat{Y}_i^{ooB} denotes the i th response and the corresponding OOB prediction value for each $i = 1, \dots, n$, respectively.

From the OOB prediction values, RF creates a confidence interval using the error distribution. The OOB prediction error D_i is defined as follows:

$$D_i = Y_i - \hat{Y}_i^{\text{ooB}}$$

for $i = 1, \dots, n$. When the numbers of CARTs and samples are large, D_i follows theoretical error distribution D that satisfies

$$\begin{aligned} 1 - \alpha &\approx P\left[D_{[n, \frac{\alpha}{2}]} \leq D \leq D_{[n, 1 - \frac{\alpha}{2}]}\right] \\ &= P\left[\hat{Y} + D_{[n, \frac{\alpha}{2}]} \leq Y \leq \hat{Y} + D_{[n, 1 - \frac{\alpha}{2}]}\right], \end{aligned}$$

where α denotes the miscoverage and $D_{[n, \gamma]}$ denotes the γ quartile of D_1, \dots, D_n (Zhang *et al.*, 2019).

The RF investigates the relationship between independent variables and a response variable using OOB error. The difference in OOB errors calculated by permuting each independent variable yields variable importance (VI) (Breiman, 2001), which can be used to understand the relationship (Janitza *et al.*, 2018). However, VI can be biased when a categorical variable has many categories and categorical and continuous variables coexist in the classification problem (Strobl *et al.*, 2007).

Missing data in RF can be imputed using the missForest method. To begin, the missForest method uses “strawman imputation,” which replaces the missing data with the median value in a continuous variable or the most frequent category in a categorical variable. The missForest employs RF prediction repeatedly to fill in missing data gaps after using strawman imputation until a specific criterion is met (Stekhoven and Bühlmann, 2012).

2.2. Bayesian additive regression trees

The BART is the ensemble model of trees. Each tree is a weak learner constrained by a priori regularization. The Bayesian backfitting algorithm (Chipman *et al.*, 2010) is used to fit the model, and the resulting posterior samples are used to quantify the uncertainty. The BART model described below

$$Y = \sum_{j=1}^m g_j(x; T_j, M_j) + \epsilon, \quad \epsilon \sim N(0, \sigma^2),$$

where g_j is the function that assigns $\mu_{ij} \in M_j$ to x , μ_{ij} 's ($\mu_{1j}, \dots, \mu_{bj} \in M_j$) are terminal node parameters, T_j is a binary tree, and σ is a standard deviation (SD). The regularization prior is specified as follows:

$$p((T_1, M_1), \dots, (T_m, M_m), \sigma) = \left[\prod_{i,j} p(\mu_{ij}|T_j)p(T_j) \right] p(\sigma),$$

where m is the number of trees and tree parameters (T_j, M_j) 's are independent to the SD and as well as each other. Tree parameter T_j consists of splitting variable and splitting value assignments. The probability of a non-terminal node $p_{\text{split}}(\eta) = \alpha(1 + d)^{-\beta}$ where (α, β) are set to $(0.95, 2)$ in general. Splitting variable and splitting value assignments are determined by discrete uniform distribution. The conditional expectation $E(Y|x)$ is the sum of $\mu_{ij} \in M_j$ that corresponds to x . It is suggested that y be rescaled to have $y_{\min} = -0.5$ and $y_{\max} = 0.5$. Furthermore, it is commonly assumed that $\mu_{ij} \sim N(0, \sigma_\mu^2)$ where $\sigma_\mu = 0.5/k \sqrt{m}$. If $k = 2$, then a prior probability that the interval (y_{\min}, y_{\max}) includes $E(Y|X)$ is 0.95. The SD σ follows the inverse chi-square distribution which is a conjugate prior: $\sigma^2 \sim \nu\lambda/\chi_\nu^2$ where ν and λ are hyperparameters. To tune these hyperparameters, let $p(\sigma < \hat{\sigma}) = q$ where $\hat{\sigma}$ is a residual SD. The default setting of (ν, q) is $(3, 0.90)$. Bayesian backfitting MCMC algorithm generates posterior samples via a series of Gibbs samplers. The posterior sample (T_j, M_j) is drawn from the corresponding full conditional based on $(T_{(j)}, M_{(j)}, \sigma)$, where $T_{(j)}$ is all trees except T_j and similarly for $M_{(j)}$. Each tree is fitted based on the j^{th} residual $R_j \equiv y - \sum_{k \neq j} g(x; T_k, M_k)$. By the conjugacy of M_j , we can easily obtain the distribution of T_j . In each MCMC iteration, two posterior sampling steps for M_j and T_j are iterated for all trees and then σ is sampled. To alternate the tree structure, growing, pruning, changing, and swapping are randomly proposed. Growing is the process of dividing a terminal node into two terminal nodes. Pruning is the process of combining two terminal nodes and converting the parent node into a terminal node. Changing is the act of reassigning a splitting rule in an internal node. Swapping is the process of changing the splitting rules in a parent-child pair. Variable selection is based on the posterior inclusion probability of splitting variables in the BART structure. In this paper, we construct a credible interval using the $100(\alpha/2)\%$ and $100(1 - \alpha/2)\%$ quantiles of posterior samples (Chipman *et al.*, 2010).

The DART (Linero, 2018), a variant of BART, can be used in high-dimensional or highly correlated data settings, with a Dirichlet prior producing the posterior for the following splitting probabilities (s_1, \dots, s_p) :

$$(s_1, \dots, s_p) \sim D\left(\frac{\alpha}{p} + m_1, \dots, \frac{\alpha}{p} + m_p\right),$$

where s_j is the probability of choosing the j^{th} variable as a splitting variable, p is the number of predictors, m_j is the number of the j^{th} splitting variable in the tree structures, and α follows the beta distribution as follows

$$\frac{\alpha}{\alpha + p} \sim \text{Beta}(a, b),$$

where $a = 0.5$ and $b = 1$ by default.

Another BART variant, bartMachine, handles missing data in MCMC steps by incorporating missingness as another useful piece of information when constructing the BART structure. During the grow or change steps, the following splitting rules are sampled uniformly. Let x_{ij}^* be a split point.

1. If x_{ij} is missing, send to the left; if it is not missing and $x_{ij} \leq x_{ij}^*$, send to the left, otherwise the right.
2. If x_{ij} is missing, send to the right; if it is not missing and $x_{ij} \leq x_{ij}^*$, send to the left, otherwise the right.
3. If x_{ij} is missing, send to the left; if it is not missing, send to the right (Kapelner and Bleich, 2016).

3. Random forest vs Bayesian additive regression trees

3.1. Missing data situation

There are three types of missing data situations: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Rubin, 1976). The MCAR refers to a situation in which the probability that an observation is missing is not correlated with either observed or missing values. The MAR refers to a situation in which the probability of an observation being missing is correlated with observed values but not with missing values. The MNAR denotes a situation in which the probability of an observation being missing is correlated with both observed and missing values. The BART and RF methods approach missing data inference differently. In MCMC steps, BART uses missing data as useful information. In contrast, RF imputes and fits imputed data.

Tang and Ishwaran (2017) demonstrated that when the missing proportion is moderate or low, RF imputation methods outperform strawman imputation. The RF imputation methods in MNAR are generally inferior to strawman imputation unless the correlation is high. In both the MAR and MNAR scenarios, Kapelner and Bleich (2015) consider MCAR, MAR, and MNAR settings, and their proposed BART method, *bartMachine*, outperforms *missForest*, a variant of RF. However, in the MCAR setting, *bartMachine* is inferior to *missForest*.

3.2. High-dimensional data situation

In this paper, we consider high-dimensional data with a large number of predictors. The number of significant variables, in this case, is significantly low in comparison to the total number of predictors. It is known as the sparsity assumption (Bühlmann and Van De Geer, 2011). It is known that RF performs well in high-dimensional settings because CART in RF eliminates outliers when data is bootstrapped and variable selection occurs at each node (Gunduz and Fokoué, 2015). The BART's predictive performance generally worsens as the number of predictors increases (Linero, 2018) since BART employs a uniform prior to select a splitting variable, resulting in high rejection rates during the MCMC process in the high-dimensional setting (Hernández *et al.*, 2018).

Under the sparsity assumption, it is critical to select significant predictors. The RF method makes extensive use of VI. for this purpose. Fox *et al.* (2017) used a VI-based backward elimination strategy. However, there is no significant change in predictive performance because RF is resistant to overfitting and can control nonsignificant predictors. The DART, a BART variant, can identify significant predictors and compute PIPs. It is well understood that as the number of trees decreases, the original BART includes more significant predictors (Chipman *et al.*, 2010).

3.3. Highly correlated data situation

Highly correlated data in this paper implies highly correlated independent variables (or predictors). In general, studies involving high-dimensional data have highly correlated data structures. In such cases, the likelihood of selecting highly correlated insignificant predictors in RF is high (Waldmann,

2016). The MCMC in BART may become stuck in a posterior mode when the predictor variables are highly correlated. Linero (2018) proposed DART, a BART with a sparsity-inducing Dirichlet prior to selection probabilities, to address this issue (Linero, 2018).

In the following simulation study section (Section 4), we consider scenarios with a large number of samples, some of which are missing or highly correlated. In addition, missing data and high dimensional data are highly correlated in the real data section (Section 5). We examine the performance of BART and RF methods in a variety of scenarios.

4. Simulation study

We compare the performance of two methods using several metrics: root-mean-square error (RMSE), 95% coverage, absolute bias, relative bias, and computation time. The number of simulated data set is $n_s = 200$. We fix the design matrix of train and test data. For each simulated data, the response variables are generated by the data generating models.

The RMSE of test unit RMSE_j ($j = 1, 2, \dots, n$) is defined as

$$\text{RMSE}_j = \sqrt{\frac{1}{n_s} \sum_{i=1}^{n_s} (Y_{ij} - \hat{Y}_{ij})^2},$$

where Y_{ij} is the j^{th} unit of the i^{th} simulated test data, and \hat{Y}_{ij} is the corresponding predicted value. The $\overline{\text{RMSE}}$ and SD (RMSE) denote the mean and standard deviation of RMSE_j 's, respectively. The coverage of the j^{th} test unit Cover_j is defined as

$$\text{Cover}_j = \frac{1}{n_s} \sum_{i=1}^{n_s} I(Y_{ij} \in C_{ij}),$$

where I is the indicator function that is 1 if it is true and 0 otherwise, and C_{ij} denotes an uncertainty interval of the j^{th} test unit for the i^{th} simulated test data. $\overline{\text{Cover}}$ and SD (Cover) denote the mean and standard deviation of Cover_j 's, respectively. The absolute bias of the j^{th} test unit Bias_j is defined as

$$\text{Bias}_j = \frac{1}{n_s} \sum_{i=1}^{n_s} |Y_{ij} - \hat{Y}_{ij}|.$$

$\overline{\text{Bias}}$ and SD (Bias) denote the mean and standard deviation of Bias_j 's, respectively. The relative bias of the j^{th} test unit Rbias_j is defined as

$$\text{Rbias}_j = \frac{1}{n_s} \sum_{i=1}^{n_s} \left| \frac{Y_{ij} - \hat{Y}_{ij}}{Y_{ij}} \times 100 \right|.$$

$\overline{\text{Rbias}}$ and SD (Rbias) denote the mean and standard deviation of Rbias_j 's, respectively. The posterior inclusion probability of the l^{th} variable, p_l , is defined as

$$p_l = \frac{1}{n_s \times n_{mc}} \sum_{i=1}^{n_s} \sum_{k=1}^{n_{mc}} I(\hat{M}_{ikl} > 0),$$

where \hat{M}_{ikl} is the number of the l^{th} variables in trees for the i^{th} simulated test data and the k^{th} MCMC iteration. In addition, n_{mc} denotes the number of MCMC iterations performed after burn-in. We consider the l^{th} variable as the significant variable if p_l is larger than 0.5. All simulations are run on MAC OS 16 cores 3.2GHz RAM 96GB.

Table 1: The result of absolute bias, RMSE, relative bias, coverage, and computation time in the missing data scenario

| Missing Prop. | Imputation | Prediction | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Time(sec) |
|---------------|------------|-------------|--------------|--------------|--------------|--------------|----------------|----------------|--------------|--------------|--------------|
| MCAR | 20% | bartMachine | 1.25* | 0.62* | 1.58* | 0.66* | 295.82 | 333.51 | 0.75 | 0.12* | 1.03 |
| | | missForest | 1.37 | 0.75 | 1.65 | 0.77 | 284.13* | 296.45* | 0.93* | 0.15 | 0.88* |
| | 40% | bartMachine | 1.40* | 0.68* | 1.76* | 0.73* | 346.89 | 512.73 | 0.74 | 0.12* | 0.76* |
| | | missForest | 1.50 | 0.79 | 1.83 | 0.83 | 314.92* | 383.65* | 0.93* | 0.12* | 0.83 |
| | 60% | bartMachine | 1.56* | 0.73* | 1.95* | 0.82* | 388.93 | 621.08 | 0.72 | 0.13 | 0.83 |
| | | missForest | 1.66 | 0.83 | 2.03 | 0.89 | 328.56* | 359.86* | 0.95* | 0.11* | 0.76* |
| MAR | 20% | bartMachine | 1.32* | 1.07* | 1.57* | 1.05* | 163.30 | 141.24 | 0.72 | 0.23* | 0.84 |
| | | missForest | 1.75 | 1.64 | 1.93 | 1.60 | 151.21* | 115.93* | 0.87* | 0.29 | 0.31* |
| | 40% | bartMachine | 1.42* | 1.10* | 1.67* | 1.08* | 161.88 | 146.99 | 0.70 | 0.26 | 0.63 |
| | | missForest | 1.64 | 1.37 | 1.89 | 1.35 | 151.10* | 122.54* | 0.90* | 0.23* | 0.27* |
| | 60% | bartMachine | 1.52* | 1.18* | 1.76* | 1.15* | 160.51 | 148.96 | 0.68 | 0.30 | 0.64 |
| | | missForest | 1.75 | 1.40 | 1.99 | 1.37 | 156.05* | 125.79* | 0.91* | 0.21* | 0.25* |
| MNAR | 20% | bartMachine | 1.35* | 1.11* | 1.64* | 1.11* | 265.81* | 467.02 | 0.74 | 0.19* | 0.99 |
| | | missForest | 1.58 | 1.44 | 1.83 | 1.42 | 287.00 | 379.93* | 0.94* | 0.19* | 0.37* |
| | 40% | bartMachine | 1.40* | 1.11* | 1.65* | 1.09* | 272.32* | 427.35* | 0.75 | 0.25 | 0.74 |
| | | missForest | 1.75 | 1.45 | 1.98 | 1.42 | 330.69 | 495.31 | 0.93* | 0.19* | 0.35* |
| | 60% | bartMachine | 1.43* | 1.19* | 1.66* | 1.16* | 279.52* | 404.70 | 0.75 | 0.27 | 0.76 |
| | | missForest | 1.74 | 1.41 | 1.98 | 1.38 | 303.24 | 394.56* | 0.94* | 0.18* | 0.41* |

* indicates the smallest values of Bias, RMS E, Rbias, SD(RMS E), SD(Cover), SD(Bias), SD(Rbias), and Time(sec) in each case. In addition, for Cover, * denotes the closest value to 0.95.

4.1. Missing data scenario

We consider the following data generating model from Kapelner and Bleich (2015):

$$Y = X_1 + X_2 + 2X_3 - X_1^2 + X_2^2 + X_1X_2 + \epsilon, \quad \epsilon \stackrel{iid}{\sim} N(0, \sigma_e^2) \quad (X_1, X_2, X_3) \stackrel{iid}{\sim} N_3 \left(0, \sigma_x^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \right),$$

where $\sigma_x^2 = 1$, $\rho_1 = 0.2$, $\rho_2 = 0.4$ and $\sigma_e^2 = 1$. For the missing data mechanism, we consider MCAR, MAR, and MNAR situations. In MCAR, X_1 , X_2 , and X_3 are missing with probability $\gamma = \{0.07, 0.16, 0.26\}$. In MAR, X_3 is missing ($M_3 = 1$) with the following probability

$$P(M_3 = 1 | X_1, X_2) = \Phi(\gamma_0 + \gamma_1 X_1 + \gamma_1 X_2^2),$$

where $\Phi(\cdot)$ denotes the standard normal CDF, $\gamma_0 = -3$ and $\gamma_1 = \{1.4, 2.7, 7.0\}$. In MNAR, X_2 is missing ($M_2 = 1$) with the following probability

$$P(M_2 = 1 | X_1, X_2) = \Phi(\gamma_0 + \gamma_1 X_1^2 + \gamma_1 X_2),$$

where $\gamma_0 = -3$ and $\gamma_1 = \{1.4, 2.7, 7.0\}$. The settings above correspond to 20%, 40%, and 60% of total incomplete cases, respectively. For BART, the R package `bartMachine` is used. For the RF method, we use two R packages `randomForest` and `missForest` (Kapelner and Bleich, 2016; Liaw and Wiener, 2002; Stekhoven and Bühlmann, 2012).

In the missing data scenario, the numbers of training and test data are 250 and 50, respectively, and the number of simulated data is 200. The total number of MCMC iterations is 1250, with the first 250 discarded as a burn-in. The number of trees is set to 50, and the k -value, which determines the prior probability that the conditional expectation $E(Y|X)$ will be within the response variable's range, is set to 2 (Kapelner and Bleich, 2016). To impute missing covariate values for the RF method, `missForest` from the R package `missForest` is used with 100 trees and a maximum number of data updates of 10 (Stekhoven and Bühlmann, 2012). To predict the test units, the package `randomForest` is used with

Table 2: The BART methods PIPs X_1, \dots, X_5 are true predictors. ‘# of Selected Variables’ denotes the number of variables with PIP greater than 0.5

| (n, p) | Method | X_1 | X_2 | X_3 | X_4 | X_5 | #of Selected Variables |
|------------|--------|-------|-------|-------|-------|-------|------------------------|
| (100, 100) | BART | 1.00 | 1.00 | 0.97 | 1.00 | 0.99 | 100 |
| | DART | 1.00 | 1.00 | 0.93 | 1.00 | 0.99 | 30 |
| (100, 200) | BART | 0.98 | 0.95 | 0.75 | 1.00 | 0.98 | 200 |
| | DART | 0.98 | 0.98 | 0.64 | 1.00 | 0.98 | 7 |
| (200, 100) | BART | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 100 |
| | DART | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 27 |
| (200, 200) | BART | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 200 |
| | DART | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 6 |

the number of trees being 500 and the number of sampled variables at each node being \sqrt{p} , where p is the number of predictors (Liaw and Wiener, 2002).

Table 1 displays the results of the missing data scenario. In the Appendix, we also present an comprehensive list of simulation results based on several BART and RF R packages with a variety of parameter settings (Tables A.1–A.3). In MCAR, MAR, and MNAR, the Bias and SD(Bias) of `bartMachine` have the smallest value in all missing proportions. The results for RMSE are similar to those for absolute bias. In the MCAR, MAR, and MNAR situations, the average of `missForest-randomForest`’s relative bias is the lowest, and the standard deviation of relative bias is also low. The conclusions drawn from relative bias and absolute bias are opposed. For a small test unit value close to 0, `bartMachine` generates a prediction value with a much higher absolute value than `randomForest`, resulting in an extreme relative bias for the unit (e.g., test value: -0.0019 , prediction value: 1.38 (`bartMachine`) vs. -0.03 (`randomForest`)). In terms of coverage, on the other hand, the results show that `missForest-randomForest` has values close to 0.95 while `bartMachine`’s values range from 0.68 to 0.75. The computation time for the `bartMachine` is slightly longer, but not significantly longer. The same conclusion can be drawn in the appendix for different parameter settings.

4.2. High-dimensional data scenario

We consider the following data generating model from Friedman (1991):

$$f(x) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \epsilon, \quad \epsilon \stackrel{iid}{\sim} N(0, 1),$$

where $x_j \stackrel{iid}{\sim} \text{Unif}(0, 1)$ for $j = 1, \dots, 5$. We consider the following four scenarios

$$(n, p) \in \{(100, 100), (100, 200), (200, 100), (200, 200)\},$$

where n is the number of samples and p is the number of predictors. The remaining predictors follow $\text{Unif}(0, 1)$. The number of test data units is set to 50 and we generate 200 simulated data sets.

For the BART and DART methods, the R package BART is used (Sparapani *et al.*, 2021). For RF, the R package ranger is used (Wright *et al.*, 2020) in this simulation study. We use the default settings in each package. The number of trees in BART and DART’s default settings is 200, and k equals 2 (Sparapani *et al.*, 2021). The total number of MCMC iterations is 1250, with a burn-in of 250. The number of trees in the default setting of ranger is 500, and the number of sampled variables at each node is set to \sqrt{p} (Wright *et al.*, 2020). In the Appendix, we also conduct the simulation study with other BART and RF implementations in R with different parameter settings (Tables A.4–A.11).

Table 3: Variable importance (VI) results from RF X_1, \dots, X_5 are true predictors. Each value represents the order of the VI value for $X_1 \sim X_5$ among p variables with the actual VI value enclosed in parenthesis.

| (n, p) | X_1 (VI) | X_2 (VI) | X_3 (VI) | X_4 (VI) | X_5 (VI) |
|------------|------------|------------|------------|------------|------------|
| (100, 100) | 3 (0.53) | 1 (2.98) | 4 (0.24) | 2 (2.23) | 5 (0.22) |
| (100, 200) | 3 (0.74) | 4 (0.32) | 6 (0.11) | 1 (2.71) | 2 (0.83) |
| (200, 100) | 2 (3.34) | 3 (2.97) | 5 (0.32) | 1 (4.69) | 4 (1.66) |
| (200, 200) | 1 (2.15) | 3 (1.19) | 5 (0.20) | 2 (1.30) | 4 (0.32) |

Table 4: The results of RMSE, coverage, absolute bias, and relative bias in the high dimensional data scenario

| (n, p) | Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Uncertainty | Time(s) |
|------------|--------|--------------|--------------|--------------|--------------|---------------|--------------|--------------|--------------|-------------|--------------|
| (100, 100) | BART | 2.54 | 2.09 | 2.78 | 1.78 | 12.98 | 5.13 | 0.92 | 0.20 | 11.11 | 1.09 |
| | DART | 1.94* | 1.70* | 2.19* | 1.37* | 13.00 | 5.06* | 0.95* | 0.16* | 9.51 | 1.08 |
| | ranger | 3.54 | 2.56 | 3.71 | 2.33 | 12.91* | 5.25 | 0.91 | 0.20 | 14.91 | 0.04* |
| (100, 200) | BART | 2.63 | 1.87 | 2.85 | 1.53 | 13.08 | 4.62 | 0.94 | 0.16 | 12.20 | 1.19 |
| | DART | 2.14* | 1.61* | 2.38* | 1.24* | 13.10 | 4.55* | 0.96 | 0.14* | 11.09 | 1.22 |
| | ranger | 3.17 | 2.42 | 3.34 | 2.20 | 13.06* | 4.72 | 0.94 | 0.15 | 16.73 | 0.04* |
| (200, 100) | BART | 1.91 | 1.56 | 2.18 | 1.16 | 14.19 | 5.07 | 0.95 | 0.14 | 10.15 | 1.39 |
| | DART | 1.58* | 1.26* | 1.84* | 0.83* | 14.19 | 5.03* | 0.95 | 0.10* | 8.44 | 1.41 |
| | ranger | 2.94 | 2.36 | 3.13 | 2.11 | 14.14* | 5.30 | 0.95 | 0.14 | 16.07 | 0.05* |
| (200, 200) | BART | 2.06 | 1.55 | 2.31 | 1.14 | 13.64 | 4.62 | 0.93 | 0.18* | 9.60 | 1.34 |
| | DART | 1.61* | 1.45* | 1.86* | 1.12* | 13.65 | 4.56* | 0.91 | 0.21 | 7.05 | 1.32 |
| | ranger | 3.25 | 2.30 | 3.43 | 2.05 | 13.61* | 4.83 | 0.94* | 0.18* | 15.32 | 0.06* |

* denotes the smallest values or values closest to 0.95 for Cover. The length of the uncertainty interval is denoted by the uncertainty.

Table 2 shows the posterior inclusion probabilities (PIPs) in BART and DART for the true predictors X_1, \dots, X_5 in the high-dimensional data scenario. In both BART and DART methods, the posterior probabilities of inclusion for $X_1 \sim X_5$ exceeds 0.5, but all remaining variables in BART have PIPs greater than 0.5, whereas DART gives PIPs greater than 0.5 only for a small set of predictors that includes the true predictors. However, as the number of trees increases, DART’s variable selection performance deteriorates. PIPs from other BART packages, `bartMachine` and `dbarts`, with different parameter settings are provided in Tables A.4–A.7 in the Appendix. Table 3 shows the order of variable importance values for $X_1 \sim X_5$ among all p variables, as well as their original VI values. In the majority of cases, the results show RF prioritizes true variables. The VIs from an additional RF package, `randomForest`, are provided in Tables A.8–A.11. The same conclusion can be drawn from those additional studies.

In Table 4, both BART methods outperform in terms of absolute bias and RMSE. In each case, DART, in particular, has the lowest bias and RMSE. In the high-dimensional case, this is to be expected because the method is designed to select a few important predictors with a sparsity-inducing prior in the presence of a large number of variables. In terms of relative bias, the `ranger` in RF exhibits the lowest `Rbias` in each case, implying that RF does not produce a relatively extreme predicted value for a small test unit value. The RF method has a faster computation time, but the BART also takes a reasonable computation time to run the model. In the Appendix, we also conduct the same simulation study with other BART and RF implementations in `R`, `darts`, `bartMachine`, and `randomForest`, with different parameter settings (Tables A.12–A.15).

4.3. Highly correlated data scenario

We consider the following data generating model from Friedman (1991):

$$f(x) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \epsilon, \quad \epsilon \stackrel{iid}{\sim} N(0, 1),$$

Table 5: In the highly correlated data scenario, the result of absolute bias, RMSE, coverage, relative bias, and computation time

| Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Time(s) |
|--------|--------------|--------------|--------------|--------------|---------------|---------------|--------------|--------------|--------------|
| BART | 4.52* | 4.3 | 4.77* | 4.07 | 20.76* | 22.70* | 0.96* | 0.17* | 2.63 |
| DART | 4.57 | 4.15* | 4.83 | 3.89* | 20.82 | 22.91 | 0.96* | 0.18 | 2.66 |
| ranger | 6.14 | 7.5 | 6.3 | 7.44 | 21.51 | 29.27 | 0.93 | 0.24 | 0.03* |

* indicates the smallest value in each category or the closest value to 0.95 in Cover.

where $x_i \sim N(\mu, \Sigma)$, $i = 1, \dots, 5$, $\mu = (0, 0, 0, 0, 0)$, and $\Sigma = (\Sigma)_{ij}$ with $\Sigma_{ij} = 0.9^{|i-j|}$ for $i, j \in \{1, 2, 3, 4, 5\}^{\otimes 2}$.

The numbers of training and test data are 250 and 50, respectively, with a total of five predictors. The number of MCMC iterations and burn-in is set to 1250 and 250, respectively. The number of trees in RF is set to 500 in the `ranger` package, and the number of sampled predictors at each node is set to \sqrt{p} , where p is the number of independent variables (Wright *et al.*, 2020). The BART methods (BART and DART) outperform in all metrics except computation time as shown in Table 5. The BART method slightly outperforms the DART method in terms of absolute bias, RMSE, and relative bias. We consider various settings (k , #num. of trees) with additional BART and RF implementations in R in the Appendix (Table A.16). The same conclusion can be drawn in that the BART methods outperform the RF methods in all metrics except computation time.

5. Real data test

In this chapter, we compare the BART and RF performance using real-world data representing three different scenarios.

5.1. Missing data

In the case of missing data, we use the 1985 auto imports database from the `bartMachine` package. The data set contains 201 automobiles (observations) with the response variable being the log selling price of the automobile. The range of price is from 8.54 to 10.72. There are 7 categorical predictors and 17 continuous/integer predictors, with missing data in one or more of the predictors in 41 automobiles. The parameter settings of BART (`bartMachine`) and RF (`missForest+randomForest`) methods are identical to those in Section 4.1. We divide the data set into training and test data using 5-fold cross validation. Table 6 shows that the BART method (`bartMachine`) performs slightly better in terms of bias, RMSE, and relative bias, while the RF method performs better in terms of coverage and computation time. These findings corroborate the findings of our simulation study. We look at other BART and RF implementations in R with different parameter settings in the Appendix (Table A.17). From there, we can reach the same conclusion.

5.2. High-dimensional and highly correlated data

We use blood-brain barrier data (Mente and Lombardo, 2005) from the `caret` package for high-dimensional and highly correlated data. It contains 208 samples for the log of a brain-blood ratio related to the drug (the response variable) and 134 independent variables (2 of them are removed since their values are constant except for one or two samples). With a SD of 0.78, the average brain-blood ratio is -0.02 . The majority of those independent variables have a moderate correlation, with some having a high correlation (> 0.8). The BART, DART, and RF(`ranger`) parameter settings are identical to those in Sections 4.2 and 4.3.

Table 6: The results for the missing real data

| Imputation | Prediction | Bias | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | Time |
|------------|--------------|------|------|----------|-------|-----------|-------|------|
| missForest | randomForest | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.95 | 1.03 |
| – | bartMachine | 0.10 | 0.08 | 0.13 | 1.09 | 0.86 | 0.82 | 1.08 |

Table 7: The results for the high dimensional real data

| | Bias | SD(Bias) | RMSE | Rbias | SD(Rbias) | cover | Uncertainty | Time |
|--------|------|----------|------|----------|-----------|-------|-------------|------|
| BART | 0.38 | 0.33 | 0.50 | 57553.37 | 616902.88 | 0.80 | 1.23 | 2.05 |
| DART | 0.39 | 0.34 | 0.52 | 56325.68 | 618025.92 | 0.79 | 1.24 | 2.56 |
| ranger | 0.40 | 0.35 | 0.53 | 73615.49 | 651519.24 | 0.95 | 2.08 | 0.05 |

Table 8: Some selected independent variables' PIPs in high-dimensional real data. The total number of variables with PIP greater than 0.5 is denoted by '# of Selected Var'

| | peoe.vsa.0.1 | peoe.vsa.4.1 | slogp.vsa0 | slogp.vsa8 | smr.vsa5 | clogp | prx | pol | wpcs | o.sp2 | # of Selected Var. |
|------|--------------|--------------|------------|------------|----------|-------|------|------|------|-------|--------------------|
| BART | 0.90 | 0.95 | 0.91 | 0.93 | 0.90 | 0.92 | 0.96 | 0.97 | 0.88 | 0.87 | 132 |
| DART | 0.68 | 0.81 | 0.73 | 0.75 | 0.70 | 0.81 | 0.78 | 0.92 | 0.66 | 0.65 | 58 |

Table 7 shows that both BART methods outperform in terms of bias, RMSE, relative bias, and coverage, while the RF method (`ranger`) outperforms in terms of computation time. These findings corroborate those found in the simulation study section. This is because some of the response variables are close to zero, the estimated relative biases in this analysis are enormous. In the Appendix, we present results from other BART and RF implementations in R with various parameter settings (Table A.18).

The PIPs of some of the independent variables used in both BART methods are shown in Table 8. Unlike the BART method, which considers all 132 independent variables to be significant, the DART method considers a smaller set of independent variables to be significant. The same analysis is carried out in the Appendix (Tables A.19 and A.20) with various BART and RF packages and parameter settings. From there, we arrive at the same conclusion. We have variable importance values of near-zero for each independent variable using the RF method (`ranger`). It only considers two variables (`tcsa` and `clogp`) to be significant.

6. Discussion

In this paper, we investigate the performance of BART and RF in three scenarios: missing data, high-dimensional data, and highly correlated data. The relative performance of BART and RF in these three scenarios is not well studied in the existing literature.

The BART (`bartMachine`) performs well in the missing data situation regardless of the missing data mechanism (MCAR, MAR, MNAR). The BART, in particular, provides a lower bias and MSE estimate. If conducting inference is a concern, the RF may be a better choice because its coverage value is close to its nominal value of 0.95. When dealing with high-dimensional data, DART outperforms in terms of absolute bias, RMSE, and relative bias. The DART, a variant of BART with a sparsity-inducing Dirichlet prior on selection probabilities, was originally designed for high-dimensional data, and our results confirm that DART should be used instead of BART when high-dimensional predictors are available. As a result, our ultimate goal is to compare DART and RF, and we discover that DART continues to outperform RF in almost every metrics except computation time. The DART has a reasonable ability to select important variables when it comes to variable selection. In the simulation

studies, DART outperforms other methods in terms of uncertainty quantification, as evidenced by relatively narrow credible intervals and high coverage values. The BART methods (DART and BART) outperform in terms of bias, RMSE, relative bias, and coverage for highly correlated data. However, when compared to RF, computation time remains a bottleneck for BART.

We consider various BART and RF implementations (packages) in R with varying parameter settings in this paper (and the appendix). Because different algorithms are used, the outcomes differ slightly depending on the package and settings. For example, `bartMachine` employs the ‘grow,’ ‘prune,’ and ‘change’ tree alteration steps, whereas `dbarts` employs the ‘grow,’ ‘prune,’ ‘change,’ and one ‘swap’ tree alteration steps. However, as extensive simulation and real-world data studies have shown, the differences caused by using different algorithms are minor. To summarize, BART is the preferred method for precise estimation. If proper coverage is critical in analysis, RF may be the best option.

Acknowledgement

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. NRF-2020R1F1A1A01048168).

Appendix

Table A1: MCAR Scenario. The settings for (the num. of sampled variables, the num. of trees) are as follows: missRanger1 = (\sqrt{p} , 100), missRanger2 = (\sqrt{p} , 300), missForest1 = (\sqrt{p} , 100), missForest2 = (\sqrt{p} , 300), randomForest1 = (\sqrt{p} , 500), randomForest2 = (\sqrt{p} , 1,000), ranger1 = (\sqrt{p} , 500), ranger2 = (\sqrt{p} , 1,000). The settings for (k , the num. of trees) are as follows: bartMachine1 = (2, 50), bartMachine2 = (2, 100), bartMachine3 = (3, 50), bartMachine4 = (3, 100). ‘missForest1 + randomForest1’ and ‘bartMachine1’ correspond to the results in Table 1. The missRanger package uses the same algorithm as the missForest package.

| Imputation | Missing Prop. | Prediction | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Time | |
|-------------|---------------|---------------|---------------|----------|------|----------|--------|-----------|--------|-----------|------|------|
| missRanger1 | 20 | randomForest1 | 1.37 | 0.75 | 1.68 | 0.76 | 292.47 | 315.84 | 0.93 | 0.14 | 0.20 | |
| | | randomForest2 | 1.37 | 0.75 | 1.68 | 0.76 | 292.45 | 319.59 | 0.93 | 0.14 | 0.41 | |
| | | ranger1 | 1.37 | 0.75 | 1.68 | 0.75 | 293.15 | 330.19 | 0.93 | 0.14 | 0.08 | |
| | 40 | ranger2 | 1.37 | 0.75 | 1.68 | 0.76 | 294.77 | 327.67 | 0.93 | 0.14 | 0.10 | |
| | | randomForest1 | 1.51 | 0.77 | 1.86 | 0.80 | 290.75 | 286.50 | 0.94 | 0.11 | 0.20 | |
| | | randomForest2 | 1.51 | 0.77 | 1.86 | 0.80 | 286.36 | 281.95 | 0.94 | 0.11 | 0.41 | |
| | 60 | ranger1 | 1.51 | 0.77 | 1.86 | 0.80 | 287.18 | 280.83 | 0.94 | 0.11 | 0.07 | |
| | | ranger2 | 1.51 | 0.77 | 1.86 | 0.80 | 287.92 | 282.20 | 0.94 | 0.11 | 0.09 | |
| | | randomForest1 | 1.69 | 0.82 | 2.08 | 0.88 | 341.74 | 369.88 | 0.95 | 0.09 | 0.20 | |
| | missRanger2 | 20 | randomForest2 | 1.69 | 0.82 | 2.08 | 0.88 | 342.12 | 374.82 | 0.95 | 0.10 | 0.45 |
| | | | ranger1 | 1.69 | 0.82 | 2.08 | 0.88 | 343.54 | 379.30 | 0.95 | 0.09 | 0.08 |
| | | | ranger2 | 1.69 | 0.82 | 2.08 | 0.88 | 344.32 | 383.09 | 0.95 | 0.10 | 0.09 |
| missForest1 | 20 | randomForest1 | 1.37 | 0.75 | 1.68 | 0.76 | 296.39 | 311.03 | 0.93 | 0.15 | 0.22 | |
| | | randomForest2 | 1.37 | 0.75 | 1.68 | 0.76 | 299.46 | 337.02 | 0.93 | 0.15 | 0.42 | |
| | | ranger1 | 1.37 | 0.75 | 1.68 | 0.76 | 301.14 | 332.87 | 0.93 | 0.14 | 0.10 | |
| | 40 | ranger2 | 1.37 | 0.75 | 1.68 | 0.76 | 295.61 | 318.70 | 0.93 | 0.15 | 0.12 | |
| | | randomForest1 | 1.51 | 0.77 | 1.86 | 0.80 | 326.23 | 412.69 | 0.94 | 0.10 | 0.23 | |
| | | randomForest2 | 1.51 | 0.77 | 1.86 | 0.80 | 330.03 | 424.56 | 0.94 | 0.10 | 0.47 | |
| | 60 | ranger1 | 1.51 | 0.77 | 1.86 | 0.80 | 328.09 | 409.91 | 0.94 | 0.10 | 0.11 | |
| | | ranger2 | 1.51 | 0.77 | 1.86 | 0.80 | 327.31 | 412.54 | 0.94 | 0.10 | 0.12 | |
| | | randomForest1 | 1.69 | 0.82 | 2.09 | 0.88 | 335.43 | 363.89 | 0.95 | 0.09 | 0.22 | |
| | missForest2 | 20 | randomForest2 | 1.69 | 0.82 | 2.09 | 0.88 | 340.28 | 373.85 | 0.95 | 0.09 | 0.43 |
| | | | ranger1 | 1.69 | 0.82 | 2.09 | 0.88 | 334.62 | 353.01 | 0.95 | 0.09 | 0.10 |
| | | | ranger2 | 1.69 | 0.82 | 2.09 | 0.88 | 337.03 | 360.35 | 0.95 | 0.09 | 0.12 |
| 40 | | randomForest1 | 1.37 | 0.75 | 1.65 | 0.77 | 284.13 | 296.45 | 0.93 | 0.15 | 0.88 | |
| | | randomForest2 | 1.36 | 0.76 | 1.65 | 0.77 | 287.91 | 303.98 | 0.92 | 0.15 | 1.08 | |
| | | ranger1 | 1.36 | 0.76 | 1.66 | 0.77 | 286.45 | 302.08 | 0.93 | 0.15 | 0.75 | |
| 60 | | ranger2 | 1.36 | 0.76 | 1.65 | 0.77 | 290.33 | 310.91 | 0.92 | 0.15 | 0.77 | |
| | | randomForest1 | 1.50 | 0.79 | 1.83 | 0.83 | 314.92 | 383.65 | 0.93 | 0.12 | 0.83 | |
| | | randomForest2 | 1.49 | 0.79 | 1.82 | 0.83 | 317.63 | 401.93 | 0.93 | 0.12 | 1.03 | |
| - | | 20 | ranger1 | 1.49 | 0.79 | 1.83 | 0.83 | 316.70 | 391.06 | 0.93 | 0.12 | 0.69 |
| | | | ranger2 | 1.49 | 0.79 | 1.82 | 0.83 | 318.19 | 395.08 | 0.93 | 0.12 | 0.72 |
| | | | randomForest1 | 1.66 | 0.83 | 2.03 | 0.89 | 328.56 | 359.86 | 0.95 | 0.11 | 0.76 |
| | 40 | randomForest2 | 1.66 | 0.83 | 2.03 | 0.89 | 332.31 | 378.75 | 0.95 | 0.10 | 0.95 | |
| | | ranger1 | 1.66 | 0.83 | 2.03 | 0.89 | 332.96 | 379.26 | 0.95 | 0.10 | 0.62 | |
| | | ranger2 | 1.66 | 0.83 | 2.03 | 0.89 | 333.33 | 381.82 | 0.95 | 0.10 | 0.64 | |
| | 60 | randomForest1 | 1.36 | 0.75 | 1.66 | 0.77 | 290.71 | 309.75 | 0.93 | 0.15 | 2.32 | |
| | | randomForest2 | 1.36 | 0.75 | 1.66 | 0.77 | 285.41 | 294.87 | 0.93 | 0.15 | 2.52 | |
| | | ranger1 | 1.36 | 0.75 | 1.66 | 0.77 | 290.18 | 314.97 | 0.93 | 0.15 | 2.18 | |
| | - | 20 | ranger2 | 1.36 | 0.75 | 1.65 | 0.77 | 289.11 | 306.67 | 0.93 | 0.15 | 2.20 |
| | | | randomForest1 | 1.49 | 0.78 | 1.82 | 0.83 | 325.35 | 426.22 | 0.93 | 0.12 | 2.19 |
| | | | randomForest2 | 1.49 | 0.78 | 1.82 | 0.83 | 326.83 | 430.02 | 0.93 | 0.12 | 2.40 |
| 40 | | ranger1 | 1.49 | 0.78 | 1.82 | 0.83 | 319.52 | 395.76 | 0.93 | 0.12 | 2.07 | |
| | | ranger2 | 1.49 | 0.78 | 1.82 | 0.82 | 324.02 | 415.60 | 0.93 | 0.12 | 2.09 | |
| | | randomForest1 | 1.65 | 0.82 | 2.02 | 0.88 | 339.98 | 392.74 | 0.95 | 0.10 | 1.85 | |
| 60 | | randomForest2 | 1.65 | 0.81 | 2.02 | 0.88 | 334.12 | 375.02 | 0.95 | 0.10 | 2.05 | |
| | | ranger1 | 1.65 | 0.82 | 2.02 | 0.88 | 336.43 | 376.24 | 0.95 | 0.10 | 1.72 | |
| | | ranger2 | 1.65 | 0.82 | 2.02 | 0.88 | 339.78 | 398.22 | 0.95 | 0.10 | 1.73 | |
| - | | 20 | bartMachine1 | 1.25 | 0.62 | 1.58 | 0.66 | 295.82 | 333.51 | 0.75 | 0.12 | 1.03 |
| | | | bartMachine2 | 1.24 | 0.60 | 1.58 | 0.64 | 302.62 | 354.98 | 0.76 | 0.13 | 1.74 |
| | | | bartMachine3 | 1.24 | 0.65 | 1.56 | 0.68 | 302.81 | 378.93 | 0.73 | 0.14 | 0.78 |
| | 40 | bartMachine4 | 1.23 | 0.65 | 1.54 | 0.68 | 296.50 | 358.70 | 0.73 | 0.15 | 1.49 | |
| | | bartMachine1 | 1.40 | 0.68 | 1.76 | 0.73 | 346.89 | 512.73 | 0.74 | 0.12 | 0.76 | |
| | | bartMachine2 | 1.40 | 0.66 | 1.76 | 0.72 | 349.30 | 501.25 | 0.76 | 0.13 | 1.43 | |
| | 60 | bartMachine3 | 1.39 | 0.71 | 1.74 | 0.75 | 337.04 | 464.06 | 0.71 | 0.14 | 0.81 | |
| | | bartMachine4 | 1.38 | 0.71 | 1.73 | 0.75 | 341.96 | 487.48 | 0.72 | 0.15 | 1.55 | |
| | | bartMachine1 | 1.56 | 0.73 | 1.95 | 0.82 | 388.93 | 621.08 | 0.72 | 0.13 | 0.83 | |
| | - | 20 | bartMachine2 | 1.56 | 0.72 | 1.94 | 0.81 | 373.20 | 543.81 | 0.74 | 0.14 | 1.55 |
| | | | bartMachine3 | 1.56 | 0.76 | 1.93 | 0.84 | 382.52 | 609.50 | 0.69 | 0.14 | 0.92 |
| | | | bartMachine4 | 1.55 | 0.76 | 1.92 | 0.84 | 383.11 | 600.90 | 0.69 | 0.15 | 1.56 |

Table A2: MAR Scenario. The settings for (the num. of sampled variables, the num. of trees) are as follows: missRanger1 = (\sqrt{p} , 100), missRanger2 = (\sqrt{p} , 300), missForest1 = (\sqrt{p} , 100), missForest2 = (\sqrt{p} , 300), randomForest1 = (\sqrt{p} , 500), randomForest2 = (\sqrt{p} , 1,000), ranger1 = (\sqrt{p} , 500), ranger2 = (\sqrt{p} , 1,000). The settings for (k , the num. of trees) are as follows: bartMachine1 = (2, 50), bartMachine2 = (2, 100), bartMachine3 = (3, 50), bartMachine4 = (3, 100). ‘missForest1 + randomForest1’ and ‘bartMachine1’ correspond to the results in Table 1. The missRanger package uses the same algorithm as the missForest package.

| Imputation | Missing Prop. | Prediction | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Time | |
|-------------|---------------|---------------|---------------|----------|------|----------|--------|-----------|--------|-----------|------|------|
| missRanger1 | 20 | randomForest1 | 1.75 | 1.74 | 1.98 | 1.71 | 155.45 | 120.96 | 0.86 | 0.29 | 0.17 | |
| | | randomForest2 | 1.75 | 1.75 | 1.98 | 1.71 | 155.03 | 120.27 | 0.86 | 0.29 | 0.35 | |
| | | ranger1 | 1.74 | 1.74 | 1.98 | 1.71 | 155.95 | 122.69 | 0.86 | 0.29 | 0.07 | |
| | 40 | ranger2 | 1.75 | 1.74 | 1.98 | 1.71 | 155.19 | 121.98 | 0.86 | 0.29 | 0.08 | |
| | | randomForest1 | 1.68 | 1.43 | 1.93 | 1.42 | 154.62 | 120.78 | 0.90 | 0.24 | 0.16 | |
| | | randomForest2 | 1.67 | 1.42 | 1.93 | 1.42 | 155.62 | 121.60 | 0.90 | 0.24 | 0.34 | |
| | 60 | ranger1 | 1.67 | 1.42 | 1.93 | 1.41 | 155.73 | 121.50 | 0.90 | 0.23 | 0.06 | |
| | | ranger2 | 1.67 | 1.42 | 1.93 | 1.42 | 154.18 | 119.23 | 0.90 | 0.23 | 0.08 | |
| | | randomForest1 | 1.77 | 1.44 | 2.02 | 1.41 | 161.11 | 126.85 | 0.91 | 0.22 | 0.19 | |
| | | 20 | randomForest2 | 1.77 | 1.44 | 2.01 | 1.41 | 161.66 | 128.10 | 0.91 | 0.22 | 0.33 |
| | | | ranger1 | 1.77 | 1.44 | 2.02 | 1.41 | 161.96 | 127.25 | 0.91 | 0.22 | 0.05 |
| | | | ranger2 | 1.77 | 1.44 | 2.01 | 1.41 | 163.14 | 129.41 | 0.91 | 0.22 | 0.07 |
| missRanger2 | 20 | randomForest1 | 1.75 | 1.75 | 1.98 | 1.72 | 157.88 | 127.77 | 0.86 | 0.29 | 0.21 | |
| | | randomForest2 | 1.75 | 1.76 | 1.98 | 1.72 | 156.54 | 125.10 | 0.86 | 0.30 | 0.40 | |
| | | ranger1 | 1.75 | 1.76 | 1.98 | 1.72 | 155.5 | 123.16 | 0.86 | 0.29 | 0.11 | |
| | 40 | ranger2 | 1.75 | 1.76 | 1.98 | 1.72 | 156.83 | 125.39 | 0.86 | 0.30 | 0.12 | |
| | | randomForest1 | 1.66 | 1.42 | 1.92 | 1.41 | 152.58 | 118.63 | 0.90 | 0.23 | 0.19 | |
| | | randomForest2 | 1.66 | 1.42 | 1.92 | 1.41 | 154.08 | 119.38 | 0.90 | 0.23 | 0.37 | |
| | 60 | ranger1 | 1.66 | 1.41 | 1.92 | 1.40 | 155.44 | 121.20 | 0.90 | 0.23 | 0.08 | |
| | | ranger2 | 1.66 | 1.42 | 1.92 | 1.41 | 154.55 | 120.55 | 0.90 | 0.23 | 0.10 | |
| | | randomForest1 | 1.76 | 1.42 | 2.00 | 1.39 | 164.91 | 131.40 | 0.91 | 0.22 | 0.17 | |
| | | 20 | randomForest2 | 1.76 | 1.42 | 2.00 | 1.39 | 164.74 | 132.22 | 0.91 | 0.22 | 0.37 |
| | | | ranger1 | 1.76 | 1.42 | 2.00 | 1.39 | 163.66 | 130.58 | 0.91 | 0.22 | 0.07 |
| | | | ranger2 | 1.76 | 1.42 | 2.00 | 1.39 | 163.31 | 129.75 | 0.91 | 0.22 | 0.08 |
| missForest1 | 20 | randomForest1 | 1.75 | 1.64 | 1.93 | 1.60 | 151.21 | 115.93 | 0.87 | 0.29 | 0.31 | |
| | | randomForest2 | 1.70 | 1.64 | 1.93 | 1.60 | 150.52 | 115.40 | 0.87 | 0.29 | 0.49 | |
| | | ranger1 | 1.70 | 1.64 | 1.93 | 1.60 | 150.17 | 114.17 | 0.87 | 0.29 | 0.20 | |
| | 40 | ranger2 | 1.70 | 1.64 | 1.93 | 1.60 | 150.19 | 114.13 | 0.86 | 0.29 | 0.22 | |
| | | randomForest1 | 1.64 | 1.37 | 1.89 | 1.35 | 151.10 | 122.54 | 0.90 | 0.23 | 0.27 | |
| | | randomForest2 | 1.64 | 1.37 | 1.89 | 1.35 | 152.07 | 123.68 | 0.90 | 0.23 | 0.45 | |
| | 60 | ranger1 | 1.64 | 1.37 | 1.89 | 1.35 | 150.69 | 120.69 | 0.90 | 0.23 | 0.15 | |
| | | ranger2 | 1.64 | 1.36 | 1.89 | 1.35 | 152.51 | 123.94 | 0.90 | 0.23 | 0.17 | |
| | | randomForest1 | 1.75 | 1.40 | 1.99 | 1.37 | 156.05 | 125.79 | 0.91 | 0.21 | 0.25 | |
| | | 20 | randomForest2 | 1.75 | 1.40 | 1.99 | 1.37 | 157.04 | 125.88 | 0.91 | 0.21 | 0.42 |
| | | | ranger1 | 1.75 | 1.40 | 1.99 | 1.37 | 157.40 | 125.71 | 0.91 | 0.21 | 0.14 |
| | | | ranger2 | 1.75 | 1.40 | 1.99 | 1.37 | 157.40 | 126.53 | 0.91 | 0.21 | 0.14 |
| missForest2 | 20 | randomForest1 | 1.70 | 1.64 | 1.94 | 1.61 | 150.45 | 113.22 | 0.86 | 0.29 | 0.68 | |
| | | randomForest2 | 1.70 | 1.65 | 1.93 | 1.61 | 151.42 | 113.59 | 0.86 | 0.29 | 0.87 | |
| | | ranger1 | 1.70 | 1.64 | 1.93 | 1.61 | 150.27 | 113.02 | 0.86 | 0.29 | 0.56 | |
| | 40 | ranger2 | 1.70 | 1.65 | 1.93 | 1.61 | 151.45 | 114.17 | 0.86 | 0.29 | 0.58 | |
| | | randomForest1 | 1.64 | 1.37 | 1.90 | 1.36 | 150.64 | 119.16 | 0.90 | 0.23 | 0.54 | |
| | | randomForest2 | 1.64 | 1.38 | 1.90 | 1.36 | 150.92 | 118.72 | 0.90 | 0.23 | 0.73 | |
| | 60 | ranger1 | 1.64 | 1.37 | 1.90 | 1.36 | 151.40 | 118.86 | 0.90 | 0.23 | 0.42 | |
| | | ranger2 | 1.64 | 1.37 | 1.90 | 1.36 | 150.55 | 118.33 | 0.90 | 0.23 | 0.44 | |
| | | randomForest1 | 1.75 | 1.39 | 1.99 | 1.36 | 157.87 | 129.17 | 0.91 | 0.21 | 0.36 | |
| | | 20 | randomForest2 | 1.74 | 1.39 | 1.99 | 1.36 | 158.35 | 129.68 | 0.91 | 0.21 | 0.54 |
| | | | ranger1 | 1.74 | 1.39 | 1.99 | 1.36 | 159.33 | 129.46 | 0.91 | 0.21 | 0.24 |
| | | | ranger2 | 1.74 | 1.39 | 1.98 | 1.36 | 157.59 | 128.21 | 0.91 | 0.21 | 0.26 |
| - | 20 | bartMachine1 | 1.32 | 1.07 | 1.57 | 1.05 | 163.30 | 141.24 | 0.72 | 0.23 | 0.84 | |
| | | bartMachine2 | 1.32 | 1.06 | 1.56 | 1.04 | 165.76 | 143.47 | 0.73 | 0.24 | 1.52 | |
| | | bartMachine3 | 1.33 | 1.15 | 1.57 | 1.12 | 158.35 | 142.97 | 0.69 | 0.25 | 0.64 | |
| | 40 | bartMachine4 | 1.33 | 1.16 | 1.56 | 1.14 | 161.09 | 147.11 | 0.69 | 0.25 | 1.36 | |
| | | bartMachine1 | 1.42 | 1.10 | 1.67 | 1.08 | 161.88 | 146.99 | 0.70 | 0.26 | 0.63 | |
| | | bartMachine2 | 1.42 | 1.10 | 1.66 | 1.08 | 163.28 | 144.77 | 0.71 | 0.26 | 1.18 | |
| | 60 | bartMachine3 | 1.44 | 1.19 | 1.68 | 1.17 | 159.03 | 153.39 | 0.66 | 0.27 | 0.71 | |
| | | bartMachine4 | 1.44 | 1.20 | 1.68 | 1.18 | 161.34 | 156.11 | 0.66 | 0.27 | 1.39 | |
| | | bartMachine1 | 1.52 | 1.18 | 1.76 | 1.15 | 160.51 | 148.96 | 0.68 | 0.30 | 0.64 | |
| | | 20 | bartMachine2 | 1.52 | 1.18 | 1.76 | 1.14 | 164.53 | 152.72 | 0.69 | 0.30 | 1.25 |
| | | | bartMachine3 | 1.55 | 1.26 | 1.78 | 1.22 | 158.81 | 158.72 | 0.63 | 0.30 | 0.67 |
| | | | bartMachine4 | 1.56 | 1.27 | 1.79 | 1.23 | 162.81 | 163.70 | 0.63 | 0.30 | 1.70 |

Table A3: MNAR Scenario. The settings for (the num. of sampled variables, the num. of trees) are as follows: missRanger1 = (\sqrt{p} , 100), missRanger2 = (\sqrt{p} , 300), missForest1 = (\sqrt{p} , 100), missForest2 = (\sqrt{p} , 300), randomForest1 = (\sqrt{p} , 500), randomForest2 = (\sqrt{p} , 1,000), ranger1 = (\sqrt{p} , 500), ranger2 = (\sqrt{p} , 1,000). The settings for (k , the num. of trees) are as follows: bartMachine1 = (2, 50), bartMachine2 = (2, 100), bartMachine3 = (3, 50), bartMachine4 = (3, 100). ‘missForest1 + randomForest1’ and ‘bartMachine1’ correspond to the results in Table 1. The missRanger package uses the same algorithm as the missForest package.

| Imputation | Missing Prop. | Prediction | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Time |
|-------------|---------------|---------------|------|----------|------|----------|--------|-----------|-------|-----------|------|
| missRanger1 | 20 | randomForest1 | 1.58 | 1.41 | 1.83 | 1.40 | 280.15 | 369.46 | 0.94 | 0.18 | 0.19 |
| | | randomForest2 | 1.58 | 1.41 | 1.82 | 1.40 | 281.93 | 371.75 | 0.94 | 0.18 | 0.40 |
| | | ranger1 | 1.58 | 1.41 | 1.82 | 1.40 | 284.43 | 374.11 | 0.94 | 0.18 | 0.07 |
| | 40 | ranger2 | 1.58 | 1.41 | 1.82 | 1.40 | 280.33 | 370.27 | 0.94 | 0.18 | 0.09 |
| | | randomForest1 | 1.75 | 1.45 | 1.99 | 1.42 | 306.71 | 455.97 | 0.93 | 0.19 | 0.19 |
| | | randomForest2 | 1.76 | 1.46 | 1.99 | 1.42 | 313.91 | 466.71 | 0.93 | 0.19 | 0.40 |
| | 60 | ranger1 | 1.76 | 1.46 | 1.99 | 1.42 | 316.94 | 468.58 | 0.93 | 0.19 | 0.07 |
| | | ranger2 | 1.75 | 1.46 | 1.99 | 1.42 | 315.47 | 475.83 | 0.93 | 0.19 | 0.09 |
| | | randomForest1 | 1.74 | 1.41 | 1.98 | 1.38 | 295.30 | 379.17 | 0.94 | 0.18 | 0.22 |
| | | randomForest2 | 1.73 | 1.41 | 1.98 | 1.38 | 291.13 | 376.30 | 0.94 | 0.18 | 0.37 |
| | | ranger1 | 1.74 | 1.41 | 1.98 | 1.38 | 295.78 | 383.58 | 0.93 | 0.19 | 0.05 |
| | | ranger2 | 1.74 | 1.41 | 1.98 | 1.38 | 297.53 | 387.47 | 0.94 | 0.18 | 0.07 |
| missRanger2 | 20 | randomForest1 | 1.58 | 1.42 | 1.82 | 1.40 | 281.51 | 368.14 | 0.94 | 0.18 | 0.22 |
| | | randomForest2 | 1.58 | 1.42 | 1.82 | 1.40 | 281.33 | 373.19 | 0.94 | 0.18 | 0.45 |
| | | ranger1 | 1.58 | 1.42 | 1.82 | 1.40 | 279.06 | 362.54 | 0.94 | 0.18 | 0.09 |
| | 40 | ranger2 | 1.58 | 1.42 | 1.82 | 1.40 | 277.23 | 364.19 | 0.94 | 0.18 | 0.11 |
| | | randomForest1 | 1.76 | 1.46 | 1.99 | 1.42 | 320.22 | 469.42 | 0.93 | 0.19 | 0.22 |
| | | randomForest2 | 1.76 | 1.45 | 1.99 | 1.42 | 317.45 | 469.51 | 0.93 | 0.19 | 0.41 |
| | 60 | ranger1 | 1.76 | 1.46 | 1.99 | 1.42 | 320.25 | 473.00 | 0.93 | 0.19 | 0.10 |
| | | ranger2 | 1.76 | 1.46 | 1.99 | 1.42 | 313.34 | 466.30 | 0.93 | 0.20 | 0.15 |
| | | randomForest1 | 1.73 | 1.41 | 1.97 | 1.37 | 300.10 | 387.31 | 0.94 | 0.18 | 0.18 |
| | | randomForest2 | 1.73 | 1.41 | 1.97 | 1.38 | 303.67 | 395.05 | 0.94 | 0.18 | 0.47 |
| | | ranger1 | 1.73 | 1.41 | 1.97 | 1.38 | 307.98 | 403.61 | 0.94 | 0.18 | 0.06 |
| | | ranger2 | 1.73 | 1.41 | 1.97 | 1.38 | 302.22 | 388.95 | 0.94 | 0.18 | 0.09 |
| missForest1 | 20 | randomForest1 | 1.58 | 1.44 | 1.83 | 1.42 | 287.00 | 379.93 | 0.94 | 0.19 | 0.37 |
| | | randomForest2 | 1.58 | 1.43 | 1.82 | 1.42 | 281.65 | 372.78 | 0.94 | 0.19 | 0.57 |
| | | ranger1 | 1.58 | 1.43 | 1.83 | 1.42 | 281.87 | 379.90 | 0.94 | 0.19 | 0.24 |
| | 40 | ranger2 | 1.58 | 1.43 | 1.83 | 1.42 | 287.76 | 386.17 | 0.94 | 0.19 | 0.26 |
| | | randomForest1 | 1.75 | 1.45 | 1.98 | 1.42 | 330.69 | 495.31 | 0.93 | 0.19 | 0.35 |
| | | randomForest2 | 1.75 | 1.45 | 1.99 | 1.42 | 333.73 | 502.08 | 0.93 | 0.19 | 0.55 |
| | 60 | ranger1 | 1.75 | 1.45 | 1.99 | 1.42 | 324.35 | 467.63 | 0.93 | 0.19 | 0.21 |
| | | ranger2 | 1.75 | 1.46 | 1.99 | 1.42 | 329.91 | 485.36 | 0.93 | 0.19 | 0.26 |
| | | randomForest1 | 1.74 | 1.41 | 1.98 | 1.38 | 303.24 | 394.56 | 0.94 | 0.18 | 0.41 |
| | | randomForest2 | 1.74 | 1.42 | 1.98 | 1.38 | 300.47 | 392.26 | 0.94 | 0.18 | 0.53 |
| | | ranger1 | 1.74 | 1.41 | 1.98 | 1.38 | 299.13 | 387.66 | 0.94 | 0.18 | 0.18 |
| | | ranger2 | 1.74 | 1.41 | 1.98 | 1.38 | 300.31 | 392.73 | 0.94 | 0.18 | 0.19 |
| missForest2 | 20 | randomForest1 | 1.58 | 1.44 | 1.82 | 1.42 | 283.36 | 380.45 | 0.94 | 0.18 | 0.72 |
| | | randomForest2 | 1.58 | 1.44 | 1.82 | 1.42 | 282.18 | 377.71 | 0.94 | 0.18 | 0.92 |
| | | ranger1 | 1.58 | 1.44 | 1.83 | 1.42 | 274.19 | 363.74 | 0.94 | 0.18 | 0.60 |
| | 40 | ranger2 | 1.58 | 1.44 | 1.82 | 1.42 | 281.07 | 377.70 | 0.94 | 0.18 | 0.62 |
| | | randomForest1 | 1.75 | 1.46 | 1.98 | 1.42 | 311.94 | 440.89 | 0.93 | 0.18 | 0.64 |
| | | randomForest2 | 1.75 | 1.45 | 1.98 | 1.42 | 305.38 | 433.67 | 0.93 | 0.18 | 0.83 |
| | 60 | ranger1 | 1.75 | 1.45 | 1.98 | 1.42 | 302.45 | 427.47 | 0.93 | 0.18 | 0.52 |
| | | ranger2 | 1.75 | 1.46 | 1.98 | 1.42 | 306.65 | 438.41 | 0.93 | 0.18 | 0.53 |
| | | randomForest1 | 1.73 | 1.42 | 1.97 | 1.38 | 298.67 | 387.49 | 0.94 | 0.18 | 0.53 |
| | | randomForest2 | 1.73 | 1.41 | 1.97 | 1.38 | 298.27 | 389.34 | 0.94 | 0.18 | 0.73 |
| | | ranger1 | 1.73 | 1.42 | 1.97 | 1.38 | 301.17 | 395.71 | 0.94 | 0.18 | 0.41 |
| | | ranger2 | 1.73 | 1.42 | 1.97 | 1.38 | 304.38 | 399.22 | 0.94 | 0.18 | 0.43 |
| - | 20 | bartMachine1 | 1.35 | 1.11 | 1.64 | 1.11 | 265.81 | 467.02 | 0.74 | 0.19 | 0.99 |
| | | bartMachine2 | 1.34 | 1.11 | 1.63 | 1.11 | 257.66 | 454.69 | 0.76 | 0.21 | 1.66 |
| | | bartMachine3 | 1.36 | 1.17 | 1.64 | 1.16 | 257.43 | 453.50 | 0.71 | 0.21 | 0.77 |
| | 40 | bartMachine4 | 1.37 | 1.21 | 1.63 | 1.20 | 265.03 | 457.66 | 0.71 | 0.22 | 1.45 |
| | | bartMachine1 | 1.40 | 1.11 | 1.65 | 1.09 | 272.32 | 427.35 | 0.75 | 0.25 | 0.74 |
| | | bartMachine2 | 1.41 | 1.14 | 1.65 | 1.11 | 268.78 | 415.46 | 0.76 | 0.26 | 1.37 |
| | 60 | bartMachine3 | 1.43 | 1.19 | 1.67 | 1.16 | 263.57 | 412.24 | 0.70 | 0.26 | 0.86 |
| | | bartMachine4 | 1.43 | 1.22 | 1.67 | 1.20 | 259.27 | 412.38 | 0.70 | 0.27 | 1.70 |
| | | bartMachine1 | 1.43 | 1.19 | 1.66 | 1.16 | 279.52 | 404.70 | 0.75 | 0.27 | 0.76 |
| | | bartMachine2 | 1.44 | 1.21 | 1.67 | 1.18 | 290.33 | 425.44 | 0.76 | 0.28 | 1.48 |
| | | bartMachine3 | 1.45 | 1.26 | 1.68 | 1.23 | 289.54 | 440.09 | 0.70 | 0.28 | 0.80 |
| | | bartMachine4 | 1.46 | 1.28 | 1.68 | 1.25 | 279.32 | 418.21 | 0.70 | 0.28 | 1.71 |

Table A4: High dimensional simulation data scenario $(n, p) = (100, 100)$. The table includes posterior probability of inclusion and the number of selected variables with posterior inclusion probability larger than 0.5. We consider various BART implementations in R: `bartMachine`, `dbarts`, BART from `BART`, DART from `BART` with a sparsity inducing prior. We also consider different settings of $(k, \text{the num. of trees})$: $1 = (2, 100)$, $2 = (3, 100)$, $3 = (2, 200)$, $4 = (3, 200)$, $5 = (2, 300)$, $6 = (3, 300)$. Note that BART3 and DART3 correspond to the results in main Table 2.

| | X_1 | X_2 | X_3 | X_4 | X_5 | # of Selected Variables |
|--------------|-------|-------|-------|-------|-------|-------------------------|
| bartMachine1 | 1.00 | 1.00 | 0.95 | 1.00 | 0.99 | 100 |
| bartMachine2 | 1.00 | 1.00 | 0.93 | 1.00 | 0.98 | 100 |
| bartMachine3 | 1.00 | 1.00 | 0.98 | 1.00 | 1.00 | 100 |
| bartMachine4 | 1.00 | 1.00 | 0.97 | 1.00 | 0.99 | 100 |
| bartMachine5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 100 |
| bartMachine6 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 100 |
| BART1 | 1.00 | 1.00 | 0.94 | 1.00 | 0.99 | 100 |
| BART2 | 1.00 | 1.00 | 0.89 | 1.00 | 0.97 | 100 |
| BART3 | 1.00 | 1.00 | 0.97 | 1.00 | 0.99 | 100 |
| BART4 | 0.99 | 1.00 | 0.94 | 1.00 | 0.97 | 100 |
| BART5 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 100 |
| BART6 | 1.00 | 1.00 | 0.97 | 1.00 | 0.98 | 100 |
| DART1 | 1.00 | 1.00 | 0.91 | 1.00 | 0.99 | 6 |
| DART2 | 1.00 | 1.00 | 0.85 | 1.00 | 0.98 | 6 |
| DART3 | 1.00 | 1.00 | 0.93 | 1.00 | 0.99 | 30 |
| DART4 | 1.00 | 1.00 | 0.87 | 1.00 | 0.97 | 22 |
| DART5 | 1.00 | 1.00 | 0.94 | 1.00 | 0.99 | 100 |
| DART6 | 1.00 | 1.00 | 0.88 | 1.00 | 0.96 | 100 |
| dbarts1 | 0.97 | 1.00 | 0.81 | 1.00 | 0.88 | 100 |
| dbarts2 | 0.91 | 0.99 | 0.79 | 0.99 | 0.83 | 100 |
| dbarts3 | 0.99 | 1.00 | 0.96 | 1.00 | 0.97 | 100 |
| dbarts4 | 0.97 | 0.99 | 0.95 | 0.99 | 0.96 | 100 |
| dbarts5 | 1.00 | 1.00 | 0.99 | 1.00 | 0.99 | 100 |
| dbarts6 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 100 |

Table A5: High dimensional simulation data scenario $(n, p) = (100, 200)$. The table includes posterior probability of inclusion and the number of selected variables with posterior inclusion probability larger than 0.5. We consider various BART implementations in R: `bartMachine`, `dbarts`, BART from `BART`, DART from `BART` with a sparsity inducing prior. We also consider different settings of $(k, \text{the num. of trees})$: $1 = (2, 100)$, $2 = (3, 100)$, $3 = (2, 200)$, $4 = (3, 200)$, $5 = (2, 300)$, $6 = (3, 300)$. Note that BART3 and DART3 correspond to the results in main Table 2.

| | X_1 | X_2 | X_3 | X_4 | X_5 | # of Selected Variables |
|--------------|-------|-------|-------|-------|-------|-------------------------|
| bartMachine1 | 0.99 | 0.98 | 0.69 | 1.00 | 0.99 | 145 |
| bartMachine2 | 0.85 | 0.80 | 0.58 | 0.99 | 0.87 | 200 |
| bartMachine3 | 0.98 | 0.96 | 0.84 | 1.00 | 0.98 | 200 |
| bartMachine4 | 0.89 | 0.86 | 0.81 | 0.98 | 0.89 | 200 |
| bartMachine5 | 0.98 | 0.98 | 0.92 | 1.00 | 0.99 | 200 |
| bartMachine6 | 0.94 | 0.93 | 0.91 | 0.98 | 0.95 | 200 |
| BART1 | 0.98 | 0.97 | 0.60 | 1.00 | 0.97 | 20 |
| BART2 | 0.82 | 0.74 | 0.48 | 0.99 | 0.81 | 10 |
| BART3 | 0.98 | 0.95 | 0.75 | 1.00 | 0.98 | 200 |
| BART4 | 0.84 | 0.79 | 0.70 | 0.97 | 0.83 | 200 |
| BART5 | 0.98 | 0.96 | 0.86 | 1.00 | 0.97 | 200 |
| BART6 | 0.88 | 0.87 | 0.82 | 0.97 | 0.89 | 200 |
| DART1 | 0.98 | 0.97 | 0.56 | 1.00 | 0.96 | 5 |
| DART2 | 0.84 | 0.77 | 0.36 | 0.99 | 0.81 | 4 |
| DART3 | 0.98 | 0.98 | 0.64 | 1.00 | 0.98 | 7 |
| DART4 | 0.8 | 0.73 | 0.50 | 0.99 | 0.78 | 11 |
| DART5 | 0.97 | 0.96 | 0.73 | 1.00 | 0.96 | 190 |
| DART6 | 0.83 | 0.77 | 0.65 | 0.97 | 0.81 | 200 |
| dbarts1 | 0.78 | 0.69 | 0.55 | 0.99 | 0.76 | 91 |
| dbarts2 | 0.65 | 0.61 | 0.51 | 0.91 | 0.65 | 132 |
| dbarts3 | 0.87 | 0.84 | 0.78 | 0.98 | 0.87 | 200 |
| dbarts4 | 0.82 | 0.80 | 0.77 | 0.93 | 0.83 | 200 |
| dbarts5 | 0.93 | 0.92 | 0.89 | 0.99 | 0.93 | 200 |
| dbarts6 | 0.92 | 0.89 | 0.89 | 0.96 | 0.92 | 200 |

Table A6: High dimensional simulation data scenario $(n, p) = (200, 100)$. The table includes posterior probability of inclusion and the number of selected variables with posterior inclusion probability larger than 0.5. We consider various BART implementations in R: `bartMachine`, `dbarts`, BART from `BART`, DART from `BART` with a sparsity inducing prior. We also consider different settings of $(k, \text{the num. of trees})$: $1 = (2, 100)$, $2 = (3, 100)$, $3 = (2, 200)$, $4 = (3, 200)$, $5 = (2, 300)$, $6 = (3, 300)$. Note that BART3 and DART3 correspond to the results in main Table 2.

| | X_1 | X_2 | X_3 | X_4 | X_5 | # of Selected Variables |
|--------------|-------|-------|-------|-------|-------|-------------------------|
| bartMachine1 | 1 | 1 | 1 | 1 | 1 | 100 |
| bartMachine2 | 1 | 1 | 1 | 1 | 1 | 100 |
| bartMachine3 | 1 | 1 | 1 | 1 | 1 | 100 |
| bartMachine4 | 1 | 1 | 1 | 1 | 1 | 100 |
| bartMachine5 | 1 | 1 | 1 | 1 | 1 | 100 |
| bartMachine6 | 1 | 1 | 1 | 1 | 1 | 100 |
| BART1 | 1 | 1 | 1 | 1 | 1 | 100 |
| BART2 | 1 | 1 | 1 | 1 | 1 | 100 |
| BART3 | 1 | 1 | 1 | 1 | 1 | 100 |
| BART4 | 1 | 1 | 1 | 1 | 1 | 100 |
| BART5 | 1 | 1 | 1 | 1 | 1 | 100 |
| BART6 | 1 | 1 | 1 | 1 | 1 | 100 |
| DART1 | 1 | 1 | 1 | 1 | 1 | 6 |
| DART2 | 1 | 1 | 1 | 1 | 1 | 5 |
| DART3 | 1 | 1 | 1 | 1 | 1 | 27 |
| DART4 | 1 | 1 | 1 | 1 | 1 | 7 |
| DART5 | 1 | 1 | 1 | 1 | 1 | 100 |
| DART6 | 1 | 1 | 1 | 1 | 1 | 23 |
| dbarts1 | 1 | 1 | 1 | 1 | 1 | 100 |
| dbarts2 | 1 | 1 | 1 | 1 | 1 | 100 |
| dbarts3 | 1 | 1 | 1 | 1 | 1 | 100 |
| dbarts4 | 1 | 1 | 1 | 1 | 1 | 100 |
| dbarts5 | 1 | 1 | 1 | 1 | 1 | 100 |
| dbarts6 | 1 | 1 | 1 | 1 | 1 | 100 |

Table A7: High dimensional simulation data scenario $(n, p) = (200, 200)$. The table includes posterior probability of inclusion and the number of selected variables with posterior inclusion probability larger than 0.5. We consider various BART implementations in R: `bartMachine`, `dbarts`, BART from `BART`, DART from `BART` with a sparsity inducing prior. We also consider different settings of $(k, \text{the num. of trees})$: $1 = (2, 100)$, $2 = (3, 100)$, $3 = (2, 200)$, $4 = (3, 200)$, $5 = (2, 300)$, $6 = (3, 300)$. Note that BART3 and DART3 correspond to the results in main Table 2.

| | X_1 | X_2 | X_3 | X_4 | X_5 | # of Selected Variables |
|--------------|-------|-------|-------|-------|-------|-------------------------|
| bartMachine1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 48 |
| bartMachine2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 50 |
| bartMachine3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 200 |
| bartMachine4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 200 |
| bartMachine5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 200 |
| bartMachine6 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 200 |
| BART1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 17 |
| BART2 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 18 |
| BART3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 200 |
| BART4 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 200 |
| BART5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 200 |
| BART6 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 200 |
| DART1 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 5 |
| DART2 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 5 |
| DART3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 6 |
| DART4 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 6 |
| DART5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 8 |
| DART6 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 8 |
| dbarts1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 21 |
| dbarts2 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 30 |
| dbarts3 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 200 |
| dbarts4 | 1.00 | 1.00 | 0.93 | 1.00 | 0.99 | 200 |
| dbarts5 | 1.00 | 1.00 | 0.98 | 1.00 | 1.00 | 200 |
| dbarts6 | 1.00 | 1.00 | 0.95 | 1.00 | 0.99 | 200 |

Table A8: High dimensional simulation data scenario $(n, p) = (100, 100)$. The table includes the order of variable importance and variable importance value in parenthesis. We consider two RF implementation in R: `ranger` and `randomForest`. We also consider different settings of (# of sampled variables, # of trees): 1 = $(\sqrt{p}, 500)$, 2 = $(2\sqrt{p}, 500)$, 3 = $(\sqrt{p}, 1,000)$, 4 = $(2\sqrt{p}, 1,000)$, 5 = $(\sqrt{p}, 2,000)$, 6 = $(2\sqrt{p}, 2,000)$, 7 = $(\sqrt{p}, 10,000)$, 8 = $(2\sqrt{p}, 10,000)$. Note that `ranger1` corresponds to the results of RF in main Table 3.

| | X_1 (VI) | X_2 (VI) | X_3 (VI) | X_4 (VI) | X_5 (VI) |
|---------------|------------|------------|------------|------------|------------|
| ranger1 | 3 (0.53) | 1 (2.98) | 4 (0.24) | 2 (2.23) | 5 (0.22) |
| ranger2 | 3 (0.69) | 1 (4.75) | 4 (0.23) | 2 (3.45) | 5 (0.22) |
| ranger3 | 3 (0.52) | 1 (2.96) | 4 (0.24) | 2 (2.23) | 5 (0.22) |
| ranger4 | 3 (0.69) | 1 (4.79) | 4 (0.24) | 2 (3.46) | 5 (0.22) |
| ranger5 | 3 (0.53) | 1 (2.98) | 4 (0.24) | 2 (2.22) | 5 (0.22) |
| ranger6 | 3 (0.70) | 1 (4.77) | 4 (0.24) | 2 (3.43) | 5 (0.22) |
| ranger7 | 3 (0.53) | 1 (2.99) | 4 (0.24) | 2 (2.23) | 5 (0.21) |
| ranger8 | 3 (0.70) | 1 (4.78) | 4 (0.24) | 2 (3.44) | 5 (0.22) |
| randomForest1 | 3 (0.53) | 1 (3.02) | 4 (0.24) | 2 (2.20) | 5 (0.21) |
| randomForest2 | 3 (0.70) | 1 (4.75) | 4 (0.24) | 2 (3.41) | 5 (0.23) |
| randomForest3 | 3 (0.52) | 1 (2.97) | 4 (0.23) | 2 (2.24) | 5 (0.21) |
| randomForest4 | 3 (0.71) | 1 (4.79) | 4 (0.23) | 2 (3.45) | 5 (0.22) |
| randomForest5 | 3 (0.52) | 1 (2.99) | 4 (0.25) | 2 (2.23) | 5 (0.22) |
| randomForest6 | 3 (0.70) | 1 (4.76) | 4 (0.23) | 2 (3.43) | 5 (0.22) |
| randomForest7 | 3 (0.52) | 1 (2.97) | 4 (0.24) | 2 (2.23) | 5 (0.22) |
| randomForest8 | 3 (0.70) | 1 (4.77) | 4 (0.24) | 2 (3.44) | 5 (0.22) |

Table A9: High dimensional simulation data scenario $(n, p) = (100, 200)$. The table includes the order of variable importance and variable importance value in parenthesis. We consider two RF implementation in R: `ranger` and `randomForest`. We also consider different settings of (# of sampled variables, # of trees): 1 = $(\sqrt{p}, 500)$, 2 = $(2\sqrt{p}, 500)$, 3 = $(\sqrt{p}, 1,000)$, 4 = $(2\sqrt{p}, 1,000)$, 5 = $(\sqrt{p}, 2,000)$, 6 = $(2\sqrt{p}, 2,000)$, 7 = $(\sqrt{p}, 10,000)$, 8 = $(2\sqrt{p}, 10,000)$. Note that `ranger1` corresponds to the results of RF in main Table 3.

| | X_1 (VI) | X_2 (VI) | X_3 (VI) | X_4 (VI) | X_5 (VI) |
|---------------|------------|------------|------------|------------|------------|
| ranger1 | 3 (0.74) | 4 (0.32) | 6 (0.11) | 1 (2.71) | 2 (0.83) |
| ranger2 | 3 (0.94) | 4 (0.45) | 7 (0.09) | 1 (4.43) | 2 (1.19) |
| ranger3 | 3 (0.72) | 4 (0.31) | 6 (0.11) | 1 (2.74) | 2 (0.86) |
| ranger4 | 3 (0.92) | 4 (0.44) | 6 (0.10) | 1 (4.46) | 2 (1.17) |
| ranger5 | 3 (0.71) | 4 (0.33) | 6 (0.11) | 1 (2.77) | 2 (0.86) |
| ranger6 | 3 (0.91) | 4 (0.44) | 6 (0.10) | 1 (4.40) | 2 (1.18) |
| ranger7 | 3 (0.72) | 4 (0.32) | 6 (0.11) | 1 (2.72) | 2 (0.85) |
| ranger8 | 3 (0.93) | 4 (0.44) | 6 (0.10) | 1 (4.45) | 2 (1.17) |
| randomForest1 | 3 (0.70) | 4 (0.32) | 6 (0.11) | 1 (2.78) | 2 (0.83) |
| randomForest2 | 3 (0.98) | 4 (0.41) | 6 (0.10) | 1 (4.44) | 2 (1.17) |
| randomForest3 | 3 (0.69) | 4 (0.34) | 6 (0.10) | 1 (2.78) | 2 (0.88) |
| randomForest4 | 3 (0.91) | 4 (0.45) | 6 (0.10) | 1 (4.49) | 2 (1.14) |
| randomForest5 | 3 (0.72) | 4 (0.31) | 6 (0.11) | 1 (2.76) | 2 (0.84) |
| randomForest6 | 3 (0.92) | 4 (0.44) | 6 (0.10) | 1 (4.46) | 2 (1.17) |
| randomForest7 | 3 (0.72) | 4 (0.31) | 6 (0.11) | 1 (2.74) | 2 (0.86) |
| randomForest8 | 3 (0.93) | 4 (0.44) | 6 (0.10) | 1 (4.43) | 2 (1.17) |

Table A10: High dimensional simulation data scenario $(n, p) = (200, 100)$. The table includes the order of variable importance and variable importance value in parenthesis. We consider two RF implementation in R: `ranger` and `randomForest`. We also consider different settings of (# of sampled variables, # of trees): 1 = $(\sqrt{p}, 500)$, 2 = $(2\sqrt{p}, 500)$, 3 = $(\sqrt{p}, 1,000)$, 4 = $(2\sqrt{p}, 1,000)$, 5 = $(\sqrt{p}, 2,000)$, 6 = $(2\sqrt{p}, 2,000)$, 7 = $(\sqrt{p}, 10,000)$, 8 = $(2\sqrt{p}, 10,000)$. Note that `ranger1` corresponds to the results of RF in main Table 3.

| | X_1 (VI) | X_2 (VI) | X_3 (VI) | X_4 (VI) | X_5 (VI) |
|---------------|------------|------------|------------|------------|------------|
| ranger1 | 2 (3.34) | 3 (2.97) | 5 (0.32) | 1 (4.69) | 4 (1.66) |
| ranger2 | 2 (4.75) | 3 (4.12) | 5 (0.30) | 1 (7.10) | 4 (2.00) |
| ranger3 | 2 (3.29) | 3 (3.01) | 5 (0.31) | 1 (4.65) | 4 (1.67) |
| ranger4 | 2 (4.73) | 3 (4.15) | 5 (0.30) | 1 (7.12) | 4 (1.99) |
| ranger5 | 2 (3.27) | 3 (2.99) | 5 (0.31) | 1 (4.66) | 4 (1.67) |
| ranger6 | 2 (4.73) | 3 (4.14) | 5 (0.30) | 1 (7.11) | 4 (2.02) |
| ranger7 | 2 (3.29) | 3 (3.00) | 5 (0.31) | 1 (4.66) | 4 (1.67) |
| ranger8 | 2 (4.75) | 3 (4.14) | 5 (0.30) | 1 (7.13) | 4 (2.00) |
| randomForest1 | 2 (3.35) | 3 (3.00) | 5 (0.30) | 1 (4.62) | 4 (1.66) |
| randomForest2 | 2 (4.73) | 3 (4.14) | 5 (0.31) | 1 (7.12) | 4 (2.01) |
| randomForest3 | 2 (3.29) | 3 (3.01) | 5 (0.31) | 1 (4.66) | 4 (1.68) |
| randomForest4 | 2 (4.76) | 3 (4.15) | 5 (0.30) | 1 (7.13) | 4 (2.01) |
| randomForest5 | 2 (3.29) | 3 (2.99) | 5 (0.31) | 1 (4.63) | 4 (1.67) |
| randomForest6 | 2 (4.72) | 3 (4.13) | 5 (0.30) | 1 (7.16) | 4 (1.99) |
| randomForest7 | 2 (3.29) | 3 (2.99) | 5 (0.32) | 1 (4.66) | 4 (1.67) |
| randomForest8 | 2 (4.74) | 3 (4.14) | 5 (0.30) | 1 (7.13) | 4 (2.00) |

Table A11: High dimensional simulation data scenario $(n, p) = (200, 200)$. The table includes the order of variable importance and variable importance value in parenthesis. We consider two RF implementation in R: `ranger` and `randomForest`. We also consider different settings of (# of sampled variables, # of trees): 1 = $(\sqrt{p}, 500)$, 2 = $(2\sqrt{p}, 500)$, 3 = $(\sqrt{p}, 1,000)$, 4 = $(2\sqrt{p}, 1,000)$, 5 = $(\sqrt{p}, 2,000)$, 6 = $(2\sqrt{p}, 2,000)$, 7 = $(\sqrt{p}, 10,000)$, 8 = $(2\sqrt{p}, 10,000)$. Note that `ranger1` corresponds to the results of RF in main Table 3.

| | X_1 (VI) | X_2 (VI) | X_3 (VI) | X_4 (VI) | X_5 (VI) |
|---------------|------------|------------|------------|------------|------------|
| ranger1 | 1 (2.15) | 3 (1.19) | 5 (0.20) | 2 (1.30) | 4 (0.32) |
| ranger2 | 1 (3.51) | 3 (1.82) | 5 (0.23) | 2 (2.07) | 4 (0.41) |
| ranger3 | 1 (2.15) | 3 (1.18) | 5 (0.20) | 2 (1.29) | 4 (0.31) |
| ranger4 | 1 (3.51) | 3 (1.85) | 5 (0.23) | 2 (2.07) | 4 (0.41) |
| ranger5 | 1 (2.16) | 3 (1.17) | 5 (0.20) | 2 (1.30) | 4 (0.31) |
| ranger6 | 1 (3.50) | 3 (1.83) | 5 (0.23) | 2 (2.07) | 4 (0.40) |
| ranger7 | 1 (2.16) | 3 (1.18) | 5 (0.20) | 2 (1.30) | 4 (0.31) |
| ranger8 | 1 (3.52) | 3 (1.83) | 5 (0.23) | 2 (2.07) | 4 (0.41) |
| randomForest1 | 1 (2.18) | 3 (1.18) | 5 (0.20) | 2 (1.29) | 4 (0.31) |
| randomForest2 | 1 (3.49) | 3 (1.81) | 5 (0.22) | 2 (2.07) | 4 (0.39) |
| randomForest3 | 1 (2.16) | 3 (1.17) | 5 (0.20) | 2 (1.31) | 4 (0.30) |
| randomForest4 | 1 (3.53) | 3 (1.82) | 5 (0.22) | 2 (2.06) | 4 (0.41) |
| randomForest5 | 1 (2.16) | 3 (1.17) | 5 (0.20) | 2 (1.30) | 4 (0.31) |
| randomForest6 | 1 (3.52) | 3 (1.83) | 5 (0.23) | 2 (2.07) | 4 (0.41) |
| randomForest7 | 1 (2.16) | 3 (1.18) | 5 (0.20) | 2 (1.30) | 4 (0.31) |
| randomForest8 | 1 (3.52) | 3 (1.83) | 5 (0.23) | 2 (2.08) | 4 (0.41) |

Table A12: High dimensional simulation data result (n, p) is (100, 100) in high dimensional simulated data scenario. The settings for the BART methods (k , the num. of trees) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2, 300), 6 = (3, 300). The settings for the RF methods (the num. of sampled variables, the num. of trees) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000).

| Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Uncertainty | Time |
|---------------|------|----------|------|----------|-------|-----------|-------|-----------|-------------|------|
| ranger1 | 3.54 | 2.56 | 3.71 | 2.33 | 12.91 | 5.25 | 0.91 | 0.20 | 14.91 | 0.04 |
| ranger2 | 3.33 | 2.40 | 3.51 | 2.16 | 12.91 | 5.22 | 0.91 | 0.20 | 13.91 | 0.03 |
| ranger3 | 3.53 | 2.56 | 3.70 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.91 | 0.04 |
| ranger4 | 3.33 | 2.40 | 3.50 | 2.16 | 12.91 | 5.22 | 0.92 | 0.20 | 13.86 | 0.04 |
| ranger5 | 3.53 | 2.56 | 3.69 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.87 | 0.07 |
| ranger6 | 3.33 | 2.40 | 3.50 | 2.16 | 12.91 | 5.22 | 0.91 | 0.20 | 13.85 | 0.08 |
| ranger7 | 3.53 | 2.56 | 3.69 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.86 | 0.31 |
| ranger8 | 3.33 | 2.40 | 3.50 | 2.17 | 12.91 | 5.22 | 0.91 | 0.20 | 13.84 | 0.36 |
| randomForest1 | 3.54 | 2.56 | 3.70 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.91 | 0.18 |
| randomForest2 | 3.34 | 2.40 | 3.51 | 2.16 | 12.91 | 5.22 | 0.91 | 0.20 | 13.89 | 0.27 |
| randomForest3 | 3.53 | 2.56 | 3.70 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.89 | 0.35 |
| randomForest4 | 3.33 | 2.40 | 3.50 | 2.16 | 12.91 | 5.22 | 0.91 | 0.20 | 13.86 | 0.55 |
| randomForest5 | 3.53 | 2.56 | 3.69 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.88 | 0.70 |
| randomForest6 | 3.33 | 2.40 | 3.50 | 2.16 | 12.91 | 5.23 | 0.91 | 0.20 | 13.85 | 1.08 |
| randomForest7 | 3.53 | 2.56 | 3.69 | 2.34 | 12.91 | 5.25 | 0.91 | 0.20 | 14.86 | 3.48 |
| randomForest8 | 3.33 | 2.40 | 3.50 | 2.16 | 12.91 | 5.22 | 0.91 | 0.20 | 13.83 | 5.51 |
| bartMachine1 | 2.34 | 1.95 | 2.59 | 1.62 | 12.98 | 5.10 | 0.94 | 0.17 | 10.94 | 3.09 |
| bartMachine2 | 2.57 | 2.02 | 2.78 | 1.73 | 12.97 | 5.13 | 0.82 | 0.27 | 8.68 | 1.53 |
| bartMachine3 | 2.61 | 2.12 | 2.83 | 1.83 | 12.97 | 5.14 | 0.93 | 0.19 | 11.71 | 2.51 |
| bartMachine4 | 3.04 | 2.31 | 3.23 | 2.06 | 12.94 | 5.19 | 0.74 | 0.34 | 8.95 | 2.31 |
| bartMachine5 | 2.75 | 2.21 | 2.96 | 1.94 | 12.96 | 5.16 | 0.92 | 0.20 | 11.96 | 3.63 |
| bartMachine6 | 3.29 | 2.47 | 3.47 | 2.24 | 12.92 | 5.22 | 0.69 | 0.38 | 8.95 | 3.61 |
| BART1 | 2.29 | 1.93 | 2.55 | 1.59 | 12.99 | 5.10 | 0.92 | 0.18 | 10.41 | 0.68 |
| BART2 | 2.55 | 2.01 | 2.78 | 1.71 | 12.97 | 5.13 | 0.81 | 0.28 | 8.43 | 0.61 |
| BART3 | 2.54 | 2.09 | 2.78 | 1.78 | 12.98 | 5.13 | 0.92 | 0.20 | 11.11 | 1.09 |
| BART4 | 2.96 | 2.27 | 3.16 | 2.00 | 12.95 | 5.18 | 0.73 | 0.34 | 8.63 | 1.34 |
| BART5 | 2.69 | 2.17 | 2.91 | 1.89 | 12.97 | 5.15 | 0.91 | 0.21 | 11.36 | 1.57 |
| BART6 | 3.22 | 2.43 | 3.40 | 2.19 | 12.93 | 5.21 | 0.69 | 0.38 | 8.65 | 1.54 |
| DART1 | 1.72 | 1.56 | 1.98 | 1.23 | 13.01 | 5.03 | 0.93 | 0.16 | 8.13 | 0.63 |
| DART2 | 1.77 | 1.59 | 2.02 | 1.27 | 13.00 | 5.04 | 0.86 | 0.21 | 6.55 | 0.64 |
| DART3 | 1.94 | 1.70 | 2.19 | 1.37 | 13.00 | 5.06 | 0.95 | 0.16 | 9.51 | 1.08 |
| DART4 | 2.07 | 1.74 | 2.30 | 1.43 | 12.99 | 5.08 | 0.86 | 0.23 | 7.49 | 1.07 |
| DART5 | 2.12 | 1.81 | 2.37 | 1.50 | 12.99 | 5.08 | 0.94 | 0.16 | 10.27 | 1.53 |
| DART6 | 2.40 | 1.93 | 2.62 | 1.62 | 12.97 | 5.12 | 0.82 | 0.26 | 8.14 | 1.53 |
| dbarts1 | 2.91 | 2.20 | 3.10 | 1.94 | 12.95 | 5.18 | 0.90 | 0.21 | 11.88 | 0.72 |
| dbarts2 | 3.29 | 2.45 | 3.46 | 2.23 | 12.92 | 5.22 | 0.69 | 0.38 | 8.99 | 0.61 |
| dbarts3 | 3.16 | 2.39 | 3.33 | 2.15 | 12.93 | 5.21 | 0.89 | 0.23 | 12.41 | 1.20 |
| dbarts4 | 3.51 | 2.60 | 3.68 | 2.39 | 12.91 | 5.25 | 0.65 | 0.41 | 8.97 | 1.20 |
| dbarts5 | 3.27 | 2.47 | 3.44 | 2.25 | 12.93 | 5.22 | 0.87 | 0.25 | 12.54 | 1.76 |
| dbarts6 | 3.60 | 2.65 | 3.76 | 2.44 | 12.91 | 5.26 | 0.64 | 0.42 | 8.93 | 1.76 |

Table A13: High dimensional simulation data result (n,p) is (100, 200) in high dimensional simulated data scenario. The settings for the BART methods (k , the num. of tress) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2,300), 6 = (3, 300). The settings for the RF methods (the num. of sampled variables, the num. of trees) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000).

| Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Uncertainty | Time |
|---------------|------|----------|------|----------|-------|-----------|-------|-----------|-------------|------|
| ranger1 | 3.17 | 2.42 | 3.34 | 2.20 | 13.06 | 4.72 | 0.94 | 0.15 | 16.73 | 0.04 |
| ranger2 | 3.03 | 2.30 | 3.21 | 2.06 | 13.07 | 4.70 | 0.94 | 0.16 | 16.02 | 0.04 |
| ranger3 | 3.16 | 2.41 | 3.33 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.64 | 0.06 |
| ranger4 | 3.02 | 2.30 | 3.20 | 2.06 | 13.07 | 4.70 | 0.95 | 0.15 | 16.09 | 0.06 |
| ranger5 | 3.16 | 2.41 | 3.33 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.65 | 0.10 |
| ranger6 | 3.02 | 2.29 | 3.20 | 2.06 | 13.07 | 4.70 | 0.94 | 0.16 | 16.04 | 0.11 |
| ranger7 | 3.16 | 2.41 | 3.33 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.65 | 0.46 |
| ranger8 | 3.02 | 2.29 | 3.19 | 2.07 | 13.07 | 4.70 | 0.94 | 0.16 | 16.00 | 0.52 |
| randomForest1 | 3.17 | 2.42 | 3.34 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.72 | 0.24 |
| randomForest2 | 3.03 | 2.30 | 3.21 | 2.06 | 13.07 | 4.70 | 0.94 | 0.15 | 16.06 | 0.38 |
| randomForest3 | 3.16 | 2.41 | 3.33 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.64 | 0.46 |
| randomForest4 | 3.03 | 2.29 | 3.20 | 2.06 | 13.07 | 4.70 | 0.94 | 0.16 | 16.02 | 0.75 |
| randomForest5 | 3.16 | 2.41 | 3.33 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.66 | 0.92 |
| randomForest6 | 3.03 | 2.30 | 3.20 | 2.06 | 13.07 | 4.70 | 0.94 | 0.15 | 16.03 | 1.48 |
| randomForest7 | 3.16 | 2.41 | 3.33 | 2.20 | 13.06 | 4.72 | 0.94 | 0.16 | 16.64 | 4.51 |
| randomForest8 | 3.02 | 2.29 | 3.20 | 2.07 | 13.07 | 4.70 | 0.94 | 0.15 | 15.99 | 7.47 |
| bartMachine1 | 2.42 | 1.73 | 2.65 | 1.36 | 13.08 | 4.59 | 0.96 | 0.13 | 12.12 | 2.01 |
| bartMachine2 | 2.85 | 2.13 | 3.04 | 1.86 | 13.07 | 4.67 | 0.78 | 0.34 | 9.23 | 1.67 |
| bartMachine3 | 2.71 | 1.91 | 2.91 | 1.60 | 13.07 | 4.64 | 0.94 | 0.15 | 12.71 | 2.49 |
| bartMachine4 | 3.10 | 2.37 | 3.27 | 2.15 | 13.06 | 4.71 | 0.73 | 0.37 | 9.06 | 2.57 |
| bartMachine5 | 2.83 | 2.01 | 3.03 | 1.72 | 13.07 | 4.66 | 0.93 | 0.17 | 12.81 | 3.69 |
| bartMachine6 | 3.18 | 2.44 | 3.34 | 2.23 | 13.06 | 4.72 | 0.71 | 0.38 | 9.00 | 5.15 |
| BART1 | 2.42 | 1.75 | 2.65 | 1.38 | 13.09 | 4.59 | 0.95 | 0.14 | 11.89 | 0.79 |
| BART2 | 2.85 | 2.16 | 3.04 | 1.90 | 13.07 | 4.67 | 0.77 | 0.35 | 8.98 | 0.68 |
| BART3 | 2.63 | 1.87 | 2.85 | 1.53 | 13.08 | 4.62 | 0.94 | 0.16 | 12.20 | 1.19 |
| BART4 | 3.08 | 2.37 | 3.25 | 2.15 | 13.06 | 4.71 | 0.72 | 0.38 | 8.78 | 1.22 |
| BART5 | 2.77 | 1.97 | 2.97 | 1.67 | 13.07 | 4.65 | 0.93 | 0.17 | 12.36 | 2.14 |
| BART6 | 3.16 | 2.43 | 3.32 | 2.22 | 13.06 | 4.72 | 0.70 | 0.39 | 8.74 | 1.77 |
| DART1 | 1.93 | 1.50 | 2.17 | 1.12 | 13.12 | 4.51 | 0.95 | 0.14 | 9.90 | 0.71 |
| DART2 | 2.31 | 1.69 | 2.54 | 1.33 | 13.10 | 4.58 | 0.86 | 0.24 | 8.59 | 0.83 |
| DART3 | 2.14 | 1.61 | 2.38 | 1.24 | 13.10 | 4.55 | 0.96 | 0.14 | 11.09 | 1.22 |
| DART4 | 2.73 | 2.00 | 2.94 | 1.70 | 13.08 | 4.65 | 0.79 | 0.32 | 8.86 | 1.20 |
| DART5 | 2.38 | 1.70 | 2.60 | 1.35 | 13.09 | 4.58 | 0.96 | 0.13 | 11.79 | 1.71 |
| DART6 | 2.97 | 2.24 | 3.15 | 2.00 | 13.07 | 4.69 | 0.73 | 0.37 | 8.85 | 1.73 |
| dbarts1 | 2.88 | 2.14 | 3.07 | 1.88 | 13.07 | 4.67 | 0.90 | 0.21 | 12.43 | 0.65 |
| dbarts2 | 3.13 | 2.42 | 3.30 | 2.20 | 13.06 | 4.72 | 0.71 | 0.39 | 8.95 | 0.72 |
| dbarts3 | 3.02 | 2.29 | 3.20 | 2.06 | 13.06 | 4.70 | 0.89 | 0.23 | 12.74 | 1.25 |
| dbarts4 | 3.23 | 2.49 | 3.39 | 2.28 | 13.06 | 4.73 | 0.69 | 0.39 | 8.90 | 1.24 |
| dbarts5 | 3.08 | 2.33 | 3.25 | 2.10 | 13.06 | 4.70 | 0.88 | 0.24 | 12.79 | 1.87 |
| dbarts6 | 3.26 | 2.51 | 3.43 | 2.31 | 13.06 | 4.73 | 0.69 | 0.39 | 8.88 | 1.85 |

Table A14: High dimensional simulation data result (n, p) is (200, 100) in high dimensional simulated data scenario. The settings for the BART methods (k, the num. of tress) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2, 300), 6 = (3, 300). The settings for the RF methods (the num. of sampled variables, the num. of trees) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000).

| Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Uncertainty | Time |
|---------------|------|----------|------|----------|-------|-----------|-------|-----------|-------------|-------|
| ranger1 | 2.94 | 2.36 | 3.13 | 2.11 | 14.14 | 5.30 | 0.95 | 0.14 | 16.07 | 0.05 |
| ranger2 | 2.58 | 2.13 | 2.78 | 1.88 | 14.15 | 5.25 | 0.95 | 0.14 | 14.43 | 0.05 |
| ranger3 | 2.94 | 2.36 | 3.13 | 2.12 | 14.14 | 5.30 | 0.96 | 0.14 | 16.06 | 0.07 |
| ranger4 | 2.58 | 2.13 | 2.78 | 1.88 | 14.15 | 5.25 | 0.95 | 0.14 | 14.31 | 0.09 |
| ranger5 | 2.94 | 2.36 | 3.13 | 2.12 | 14.14 | 5.30 | 0.95 | 0.14 | 16.07 | 0.13 |
| ranger6 | 2.57 | 2.14 | 2.77 | 1.89 | 14.15 | 5.25 | 0.95 | 0.14 | 14.34 | 0.16 |
| ranger7 | 2.94 | 2.36 | 3.13 | 2.12 | 14.14 | 5.30 | 0.95 | 0.14 | 16.06 | 0.66 |
| ranger8 | 2.57 | 2.14 | 2.77 | 1.89 | 14.15 | 5.25 | 0.95 | 0.14 | 14.33 | 0.77 |
| randomForest1 | 2.94 | 2.36 | 3.14 | 2.12 | 14.14 | 5.30 | 0.96 | 0.13 | 16.12 | 0.51 |
| randomForest2 | 2.58 | 2.14 | 2.79 | 1.89 | 14.15 | 5.25 | 0.95 | 0.14 | 14.44 | 0.74 |
| randomForest3 | 2.94 | 2.36 | 3.14 | 2.12 | 14.14 | 5.30 | 0.96 | 0.13 | 16.07 | 1.00 |
| randomForest4 | 2.58 | 2.14 | 2.78 | 1.89 | 14.15 | 5.25 | 0.95 | 0.14 | 14.36 | 1.51 |
| randomForest5 | 2.94 | 2.36 | 3.13 | 2.12 | 14.14 | 5.30 | 0.95 | 0.14 | 16.08 | 2.03 |
| randomForest6 | 2.57 | 2.14 | 2.77 | 1.89 | 14.15 | 5.25 | 0.95 | 0.14 | 14.36 | 3.02 |
| randomForest7 | 2.94 | 2.36 | 3.13 | 2.12 | 14.14 | 5.30 | 0.95 | 0.14 | 16.05 | 9.83 |
| randomForest8 | 2.57 | 2.14 | 2.77 | 1.89 | 14.15 | 5.25 | 0.95 | 0.14 | 14.34 | 14.99 |
| bartMachine1 | 1.80 | 1.46 | 2.08 | 1.04 | 14.19 | 5.06 | 0.95 | 0.13 | 9.78 | 4.04 |
| bartMachine2 | 1.77 | 1.45 | 2.03 | 1.08 | 14.18 | 5.07 | 0.91 | 0.19 | 8.07 | 1.72 |
| bartMachine3 | 1.96 | 1.57 | 2.24 | 1.15 | 14.19 | 5.08 | 0.95 | 0.13 | 10.66 | 3.72 |
| bartMachine4 | 1.88 | 1.55 | 2.13 | 1.20 | 14.18 | 5.08 | 0.91 | 0.20 | 8.48 | 3.71 |
| bartMachine5 | 2.02 | 1.64 | 2.30 | 1.24 | 14.19 | 5.09 | 0.96 | 0.12 | 11.27 | 4.92 |
| bartMachine6 | 1.94 | 1.63 | 2.19 | 1.28 | 14.18 | 5.10 | 0.90 | 0.22 | 8.68 | 5.44 |
| BART1 | 1.75 | 1.44 | 2.02 | 1.04 | 14.19 | 5.06 | 0.94 | 0.14 | 9.44 | 0.94 |
| BART2 | 1.73 | 1.44 | 1.99 | 1.07 | 14.19 | 5.06 | 0.91 | 0.19 | 7.84 | 0.81 |
| BART3 | 1.91 | 1.56 | 2.18 | 1.16 | 14.19 | 5.07 | 0.95 | 0.14 | 10.15 | 1.39 |
| BART4 | 1.85 | 1.54 | 2.10 | 1.18 | 14.19 | 5.08 | 0.90 | 0.22 | 8.10 | 1.43 |
| BART5 | 1.98 | 1.61 | 2.25 | 1.21 | 14.19 | 5.08 | 0.95 | 0.14 | 10.58 | 1.95 |
| BART6 | 1.91 | 1.60 | 2.16 | 1.26 | 14.19 | 5.09 | 0.89 | 0.23 | 8.25 | 1.96 |
| DART1 | 1.30 | 1.02 | 1.56 | 0.54 | 14.20 | 5.00 | 0.95 | 0.08 | 6.95 | 0.79 |
| DART2 | 1.17 | 0.90 | 1.42 | 0.40 | 14.20 | 4.99 | 0.91 | 0.09 | 5.24 | 0.93 |
| DART3 | 1.58 | 1.26 | 1.84 | 0.83 | 14.19 | 5.03 | 0.95 | 0.10 | 8.44 | 1.41 |
| DART4 | 1.31 | 1.03 | 1.56 | 0.58 | 14.20 | 5.00 | 0.92 | 0.12 | 5.86 | 1.44 |
| DART5 | 1.67 | 1.34 | 1.94 | 0.91 | 14.19 | 5.04 | 0.96 | 0.11 | 9.14 | 2.35 |
| DART6 | 1.41 | 1.12 | 1.67 | 0.70 | 14.20 | 5.01 | 0.92 | 0.15 | 6.36 | 2.66 |
| dbarts1 | 1.94 | 1.53 | 2.29 | 0.94 | 14.18 | 5.05 | 0.89 | 0.16 | 8.19 | 0.66 |
| dbarts2 | 1.77 | 1.45 | 2.06 | 1.00 | 14.18 | 5.06 | 0.89 | 0.19 | 7.35 | 0.67 |
| dbarts3 | 1.94 | 1.58 | 2.25 | 1.12 | 14.19 | 5.07 | 0.95 | 0.13 | 10.20 | 1.28 |
| dbarts4 | 1.84 | 1.54 | 2.09 | 1.18 | 14.19 | 5.08 | 0.90 | 0.21 | 8.27 | 1.30 |
| dbarts5 | 1.99 | 1.61 | 2.27 | 1.20 | 14.19 | 5.08 | 0.96 | 0.12 | 11.01 | 1.96 |
| dbarts6 | 1.90 | 1.60 | 2.14 | 1.27 | 14.19 | 5.09 | 0.90 | 0.22 | 8.59 | 1.96 |

Table A15: High dimensional simulation data result (n,p) is (200, 200) in high dimensional simulated data scenario. The settings for the BART methods (k , the num. of tress) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2, 300), 6 = (3, 300). The settings for the RF methods (the num. of sampled variables, the num. of trees) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000).

| Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Uncertainty | Time |
|---------------|------|----------|------|----------|-------|-----------|-------|-----------|-------------|-------|
| ranger1 | 3.25 | 2.30 | 3.43 | 2.05 | 13.61 | 4.83 | 0.94 | 0.18 | 15.32 | 0.06 |
| ranger2 | 3.01 | 2.13 | 3.19 | 1.86 | 13.61 | 4.80 | 0.95 | 0.17 | 14.47 | 0.07 |
| ranger3 | 3.25 | 2.30 | 3.43 | 2.06 | 13.61 | 4.83 | 0.94 | 0.18 | 15.34 | 0.11 |
| ranger4 | 3.00 | 2.12 | 3.18 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.43 | 1.00 |
| ranger5 | 3.25 | 2.30 | 3.42 | 2.06 | 13.61 | 4.83 | 0.94 | 0.18 | 15.32 | 0.20 |
| ranger6 | 3.01 | 2.13 | 3.18 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.43 | 0.23 |
| ranger7 | 3.25 | 2.30 | 3.42 | 2.06 | 13.61 | 4.83 | 0.94 | 0.18 | 15.30 | 0.98 |
| ranger8 | 3.00 | 2.12 | 3.18 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.41 | 1.14 |
| randomForest1 | 3.25 | 2.30 | 3.43 | 2.05 | 13.61 | 4.83 | 0.94 | 0.18 | 15.35 | 0.60 |
| randomForest2 | 3.01 | 2.13 | 3.19 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.47 | 0.94 |
| randomForest3 | 3.25 | 2.30 | 3.42 | 2.05 | 13.61 | 4.83 | 0.94 | 0.18 | 15.33 | 1.20 |
| randomForest4 | 3.01 | 2.13 | 3.19 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.44 | 1.88 |
| randomForest5 | 3.25 | 2.30 | 3.42 | 2.06 | 13.61 | 4.83 | 0.94 | 0.18 | 15.31 | 2.43 |
| randomForest6 | 3.00 | 2.12 | 3.18 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.43 | 3.76 |
| randomForest7 | 3.25 | 2.30 | 3.42 | 2.06 | 13.61 | 4.83 | 0.94 | 0.18 | 15.31 | 11.97 |
| randomForest8 | 3.00 | 2.12 | 3.18 | 1.87 | 13.61 | 4.80 | 0.95 | 0.17 | 14.41 | 18.67 |
| bartMachine1 | 1.86 | 1.47 | 2.12 | 1.08 | 13.64 | 4.59 | 0.93 | 0.18 | 9.01 | 4.05 |
| bartMachine2 | 1.86 | 1.48 | 2.11 | 1.11 | 13.64 | 4.61 | 0.88 | 0.22 | 7.40 | 2.25 |
| bartMachine3 | 2.03 | 1.52 | 2.29 | 1.10 | 13.64 | 4.62 | 0.94 | 0.15 | 9.97 | 3.77 |
| bartMachine4 | 2.06 | 1.54 | 2.29 | 1.17 | 13.63 | 4.64 | 0.88 | 0.23 | 7.89 | 3.67 |
| bartMachine5 | 2.18 | 1.58 | 2.43 | 1.18 | 13.63 | 4.65 | 0.95 | 0.14 | 10.61 | 5.03 |
| bartMachine6 | 2.26 | 1.64 | 2.48 | 1.30 | 13.63 | 4.68 | 0.86 | 0.26 | 8.18 | 3.53 |
| BART1 | 1.93 | 1.50 | 2.19 | 1.11 | 13.64 | 4.61 | 0.93 | 0.18 | 9.14 | 0.88 |
| BART2 | 1.97 | 1.53 | 2.21 | 1.16 | 13.64 | 4.63 | 0.88 | 0.23 | 7.60 | 1.08 |
| BART3 | 2.06 | 1.55 | 2.31 | 1.14 | 13.64 | 4.62 | 0.93 | 0.18 | 9.60 | 1.34 |
| BART4 | 2.11 | 1.58 | 2.34 | 1.21 | 13.64 | 4.65 | 0.86 | 0.25 | 7.70 | 1.32 |
| BART5 | 2.17 | 1.59 | 2.43 | 1.19 | 13.64 | 4.64 | 0.93 | 0.16 | 10.05 | 1.93 |
| BART6 | 2.27 | 1.66 | 2.5 | 1.32 | 13.63 | 4.68 | 0.84 | 0.28 | 7.87 | 1.93 |
| DART1 | 1.47 | 1.36 | 1.72 | 1.04 | 13.65 | 4.55 | 0.89 | 0.21 | 6.10 | 0.89 |
| DART2 | 1.42 | 1.35 | 1.66 | 1.05 | 13.65 | 4.55 | 0.84 | 0.23 | 4.93 | 0.77 |
| DART3 | 1.61 | 1.45 | 1.86 | 1.12 | 13.65 | 4.56 | 0.91 | 0.21 | 7.05 | 1.32 |
| DART4 | 1.53 | 1.42 | 1.76 | 1.12 | 13.65 | 4.55 | 0.85 | 0.23 | 5.35 | 1.34 |
| DART5 | 1.69 | 1.49 | 1.94 | 1.15 | 13.65 | 4.57 | 0.92 | 0.20 | 7.86 | 1.92 |
| DART6 | 1.61 | 1.46 | 1.85 | 1.16 | 13.64 | 4.57 | 0.86 | 0.24 | 5.88 | 2.03 |
| dbarts1 | 1.94 | 1.53 | 2.19 | 1.16 | 13.64 | 4.62 | 0.93 | 0.19 | 9.10 | 0.65 |
| dbarts2 | 2.07 | 1.57 | 2.30 | 1.23 | 13.64 | 4.66 | 0.86 | 0.26 | 7.58 | 0.64 |
| dbarts3 | 2.22 | 1.64 | 2.44 | 1.30 | 13.63 | 4.68 | 0.94 | 0.17 | 10.63 | 1.26 |
| dbarts4 | 2.42 | 1.76 | 2.62 | 1.45 | 13.62 | 4.72 | 0.83 | 0.29 | 8.20 | 1.27 |
| dbarts5 | 2.40 | 1.75 | 2.61 | 1.42 | 13.63 | 4.71 | 0.93 | 0.18 | 11.18 | 1.89 |
| dbarts6 | 2.69 | 1.93 | 2.88 | 1.65 | 13.61 | 4.76 | 0.80 | 0.32 | 8.40 | 1.88 |

Table A16: Highly correlated simulation data result The settings for BART methods (k , the num. of tress) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2, 300), 6 = (3, 300). The settings for RF methods (the num. of sampled variables, the num. of tress) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000). Note that ranger1, BART3 and DART3 correspond to the results in main Table 5.

| Method | Bias | SD(Bias) | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | SD(Cover) | Time |
|---------------|------|----------|------|----------|-------|-----------|-------|-----------|------|
| ranger1 | 6.14 | 7.50 | 6.30 | 7.44 | 21.51 | 29.27 | 0.93 | 0.24 | 0.03 |
| ranger2 | 5.72 | 5.76 | 5.89 | 5.64 | 21.23 | 27.12 | 0.94 | 0.24 | 0.03 |
| ranger3 | 6.13 | 7.51 | 6.28 | 7.45 | 21.49 | 28.99 | 0.93 | 0.24 | 0.05 |
| ranger4 | 5.72 | 5.77 | 5.88 | 5.66 | 21.22 | 27.21 | 0.94 | 0.24 | 0.05 |
| ranger5 | 6.13 | 7.49 | 6.27 | 7.44 | 21.49 | 28.84 | 0.93 | 0.24 | 0.09 |
| ranger6 | 5.71 | 5.75 | 5.87 | 5.64 | 21.22 | 27.14 | 0.94 | 0.24 | 0.10 |
| ranger7 | 6.12 | 7.50 | 6.27 | 7.45 | 21.49 | 28.89 | 0.93 | 0.24 | 0.44 |
| ranger8 | 5.72 | 5.75 | 5.88 | 5.65 | 21.22 | 27.06 | 0.94 | 0.24 | 0.47 |
| randomForest1 | 6.15 | 7.54 | 6.31 | 7.48 | 21.50 | 29.01 | 0.93 | 0.24 | 0.22 |
| randomForest2 | 5.73 | 5.78 | 5.90 | 5.67 | 21.22 | 27.00 | 0.94 | 0.24 | 0.33 |
| randomForest3 | 6.13 | 7.49 | 6.29 | 7.44 | 21.50 | 29.04 | 0.93 | 0.24 | 0.44 |
| randomForest4 | 5.70 | 5.74 | 5.87 | 5.62 | 21.21 | 26.98 | 0.94 | 0.24 | 0.66 |
| randomForest5 | 6.13 | 7.51 | 6.28 | 7.46 | 21.48 | 28.59 | 0.93 | 0.24 | 0.90 |
| randomForest6 | 5.73 | 5.77 | 5.89 | 5.66 | 21.22 | 27.10 | 0.94 | 0.24 | 1.39 |
| randomForest7 | 6.12 | 7.50 | 6.27 | 7.45 | 21.49 | 28.85 | 0.93 | 0.24 | 4.90 |
| randomForest8 | 5.72 | 5.76 | 5.88 | 5.65 | 21.22 | 27.14 | 0.94 | 0.24 | 6.87 |
| bartMachine1 | 3.29 | 4.08 | 3.62 | 3.83 | 20.66 | 22.04 | 0.97 | 0.16 | 2.81 |
| bartMachine2 | 3.28 | 4.01 | 3.59 | 3.77 | 20.68 | 22.42 | 0.96 | 0.16 | 1.37 |
| bartMachine3 | 3.42 | 3.95 | 3.78 | 3.64 | 20.82 | 22.68 | 0.97 | 0.15 | 2.27 |
| bartMachine4 | 3.33 | 3.89 | 3.64 | 3.64 | 20.79 | 22.67 | 0.97 | 0.17 | 2.28 |
| bartMachine5 | 3.65 | 3.92 | 4.04 | 3.55 | 20.93 | 22.95 | 0.98 | 0.13 | 3.15 |
| bartMachine6 | 3.51 | 3.89 | 3.83 | 3.61 | 20.86 | 22.61 | 0.97 | 0.16 | 3.30 |
| BART1 | 4.92 | 4.76 | 5.16 | 4.53 | 20.75 | 23.11 | 0.95 | 0.18 | 1.49 |
| BART2 | 5.01 | 5.05 | 5.23 | 4.86 | 20.79 | 23.43 | 0.93 | 0.22 | 0.88 |
| BART3 | 4.52 | 4.30 | 4.77 | 4.07 | 20.76 | 22.70 | 0.96 | 0.17 | 2.63 |
| BART4 | 4.68 | 4.68 | 4.89 | 4.51 | 20.80 | 23.27 | 0.95 | 0.20 | 2.68 |
| BART5 | 4.49 | 4.16 | 4.73 | 3.91 | 20.80 | 23.12 | 0.97 | 0.16 | 3.91 |
| BART6 | 4.66 | 4.56 | 4.86 | 4.39 | 20.81 | 23.43 | 0.95 | 0.20 | 4.44 |
| DART1 | 5.01 | 4.53 | 5.25 | 4.30 | 20.78 | 23.08 | 0.95 | 0.19 | 1.43 |
| DART2 | 5.20 | 4.58 | 5.41 | 4.37 | 20.93 | 23.88 | 0.93 | 0.23 | 1.44 |
| DART3 | 4.57 | 4.15 | 4.83 | 3.89 | 20.82 | 22.91 | 0.96 | 0.18 | 2.66 |
| DART4 | 4.72 | 4.25 | 4.93 | 4.05 | 20.87 | 23.77 | 0.95 | 0.21 | 2.71 |
| DART5 | 4.48 | 4.12 | 4.73 | 3.87 | 20.79 | 22.78 | 0.96 | 0.18 | 3.91 |
| DART6 | 4.62 | 4.27 | 4.83 | 4.08 | 20.89 | 23.68 | 0.95 | 0.20 | 4.41 |
| dbarts1 | 4.44 | 5.12 | 5.12 | 4.48 | 20.85 | 22.80 | 0.81 | 0.21 | 0.79 |
| dbarts2 | 4.28 | 4.95 | 4.90 | 4.39 | 20.73 | 21.66 | 0.81 | 0.22 | 0.83 |
| dbarts3 | 4.12 | 4.81 | 4.72 | 4.27 | 20.65 | 21.44 | 0.88 | 0.18 | 1.57 |
| dbarts4 | 3.90 | 4.73 | 4.39 | 4.33 | 20.65 | 21.81 | 0.88 | 0.22 | 1.60 |
| dbarts5 | 4.02 | 4.76 | 4.57 | 4.28 | 20.68 | 21.76 | 0.92 | 0.17 | 2.41 |
| dbarts6 | 3.80 | 4.61 | 4.23 | 4.26 | 20.69 | 22.07 | 0.91 | 0.20 | 2.44 |

Table A17: Missing real data. The settings for (the num. of sampled variables, the num. of trees) are as follows: missRanger1 = (\sqrt{p} , 100), missRanger2 = (\sqrt{p} , 300), missForest1 = (\sqrt{p} , 100), missForest2 = (\sqrt{p} , 300), randomForest1 = (\sqrt{p} , 500), randomForest2 = (\sqrt{p} , 1,000), ranger1 = (\sqrt{p} , 500), ranger2 = (\sqrt{p} , 1,000). The settings for (k , the num. of trees) are as follows: bartMachine1 = (2, 50), bartMachine2 = (2, 100), bartMachine3 = (3, 50), bartMachine4 = (3,100). Note that ‘bartMachine1’ and ‘missForest1 + randomForest1’ correspond to the results in main Table 6.

| Imputation | Prediction | Bias | RMSE | SD(RMSE) | Rbias | SD(Rbias) | Cover | Time |
|-------------|---------------|------|------|----------|-------|-----------|-------|------|
| missRanger1 | randomForest1 | 0.11 | 0.09 | 0.14 | 1.14 | 0.92 | 0.93 | 0.48 |
| | randomForest2 | 0.11 | 0.09 | 0.14 | 1.14 | 0.91 | 0.95 | 0.79 |
| | ranger1 | 0.11 | 0.09 | 0.14 | 1.16 | 0.93 | 0.95 | 0.34 |
| | ranger2 | 0.11 | 0.09 | 0.14 | 1.14 | 0.92 | 0.96 | 0.36 |
| missRanger2 | randomForest1 | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.96 | 0.81 |
| | randomForest2 | 0.11 | 0.09 | 0.14 | 1.13 | 0.91 | 0.95 | 1.19 |
| | ranger1 | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.94 | 0.71 |
| | ranger2 | 0.11 | 0.09 | 0.14 | 1.14 | 0.93 | 0.95 | 0.84 |
| missForest1 | randomForest1 | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.95 | 1.03 |
| | randomForest2 | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.96 | 1.50 |
| | ranger1 | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.94 | 0.90 |
| | ranger2 | 0.11 | 0.09 | 0.14 | 1.15 | 0.91 | 0.96 | 0.92 |
| missForest2 | randomForest1 | 0.11 | 0.09 | 0.14 | 1.14 | 0.91 | 0.96 | 2.87 |
| | randomForest2 | 0.11 | 0.09 | 0.14 | 1.14 | 0.91 | 0.96 | 3.10 |
| | ranger1 | 0.11 | 0.09 | 0.14 | 1.15 | 0.92 | 0.93 | 2.72 |
| | ranger2 | 0.11 | 0.09 | 0.14 | 1.14 | 0.92 | 0.96 | 2.78 |
| - | bartMachine1 | 0.10 | 0.08 | 0.13 | 1.09 | 0.86 | 0.82 | 1.08 |
| | bartMachine2 | 0.10 | 0.08 | 0.13 | 1.07 | 0.85 | 0.80 | 1.81 |
| | bartMachine3 | 0.11 | 0.08 | 0.13 | 1.12 | 0.87 | 0.78 | 0.92 |
| | bartMachine4 | 0.10 | 0.08 | 0.13 | 1.10 | 0.84 | 0.78 | 1.29 |

Table A18: High dimensional real data. The settings for BART methods (k , the num. of tress) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2, 300), 6 = (3, 300). The settings for RF methods (the num. of sampled variables, the num. of tress) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000).

| Method | Bias | SD(Bias) | RMSE | Rbias | SD(Rbias) | Cover | Uncertainty | Time |
|---------------|------|----------|------|----------|-----------|-------|-------------|-------|
| ranger1 | 0.40 | 0.35 | 0.53 | 73615.49 | 651519.24 | 0.95 | 2.08 | 0.05 |
| ranger2 | 0.40 | 0.35 | 0.53 | 73769.63 | 672446.82 | 0.94 | 2.05 | 0.05 |
| ranger3 | 0.40 | 0.35 | 0.53 | 75690.78 | 702101.48 | 0.93 | 2.04 | 0.08 |
| ranger4 | 0.40 | 0.35 | 0.53 | 74858.40 | 701579.92 | 0.93 | 2.05 | 0.09 |
| ranger5 | 0.40 | 0.35 | 0.53 | 75510.63 | 699750.82 | 0.93 | 2.04 | 0.15 |
| ranger6 | 0.40 | 0.35 | 0.53 | 75861.80 | 717969.86 | 0.93 | 2.06 | 0.16 |
| ranger7 | 0.40 | 0.35 | 0.53 | 72410.68 | 669895.55 | 0.93 | 2.05 | 0.72 |
| ranger8 | 0.40 | 0.35 | 0.53 | 74013.96 | 698550.44 | 0.93 | 2.05 | 0.80 |
| randomForest1 | 0.40 | 0.35 | 0.53 | 70831.52 | 659808.60 | 0.92 | 2.06 | 0.42 |
| randomForest2 | 0.40 | 0.35 | 0.53 | 72916.64 | 680479.42 | 0.94 | 2.09 | 0.72 |
| randomForest3 | 0.40 | 0.35 | 0.53 | 75296.51 | 697135.50 | 0.93 | 2.03 | 0.69 |
| randomForest4 | 0.40 | 0.35 | 0.53 | 78485.15 | 736856.48 | 0.93 | 2.07 | 1.26 |
| randomForest5 | 0.40 | 0.35 | 0.53 | 73732.11 | 688131.08 | 0.93 | 2.05 | 1.49 |
| randomForest6 | 0.40 | 0.35 | 0.53 | 72183.42 | 682200.14 | 0.93 | 2.07 | 2.26 |
| randomForest7 | 0.40 | 0.35 | 0.53 | 74152.46 | 684353.14 | 0.93 | 2.05 | 7.96 |
| randomForest8 | 0.40 | 0.35 | 0.53 | 75435.04 | 703059.30 | 0.94 | 2.05 | 10.58 |
| bartMachine1 | 0.38 | 0.34 | 0.51 | 63852.88 | 633243.28 | 0.83 | 1.37 | 3.25 |
| bartMachine2 | 0.39 | 0.34 | 0.52 | 62758.59 | 546336.30 | 0.74 | 1.09 | 2.08 |
| bartMachine3 | 0.37 | 0.33 | 0.50 | 66793.59 | 613913.42 | 0.84 | 1.35 | 3.93 |
| bartMachine4 | 0.39 | 0.34 | 0.51 | 63495.24 | 608996.38 | 0.75 | 1.06 | 3.19 |
| bartMachine5 | 0.38 | 0.33 | 0.50 | 64036.63 | 636505.89 | 0.82 | 1.33 | 3.97 |
| bartMachine6 | 0.39 | 0.34 | 0.51 | 63819.95 | 607013.97 | 0.75 | 1.05 | 3.87 |
| BART1 | 0.38 | 0.34 | 0.51 | 54681.65 | 658966.98 | 0.80 | 1.26 | 1.17 |
| BART2 | 0.40 | 0.34 | 0.53 | 56348.96 | 556497.26 | 0.70 | 1.02 | 1.48 |
| BART3 | 0.38 | 0.33 | 0.50 | 57553.37 | 616902.88 | 0.80 | 1.23 | 2.05 |
| BART4 | 0.40 | 0.34 | 0.52 | 60854.74 | 596433.40 | 0.71 | 0.99 | 2.02 |
| BART5 | 0.38 | 0.33 | 0.50 | 56636.10 | 606572.35 | 0.81 | 1.21 | 2.99 |
| BART6 | 0.40 | 0.35 | 0.53 | 61133.50 | 586680.71 | 0.70 | 0.97 | 2.99 |
| DART1 | 0.39 | 0.34 | 0.52 | 56996.00 | 644438.34 | 0.80 | 1.26 | 1.19 |
| DART2 | 0.40 | 0.34 | 0.53 | 56351.56 | 523665.58 | 0.73 | 1.03 | 1.31 |
| DART3 | 0.39 | 0.34 | 0.52 | 56325.68 | 618025.92 | 0.79 | 1.24 | 2.56 |
| DART4 | 0.41 | 0.35 | 0.53 | 52813.93 | 510855.22 | 0.69 | 1.00 | 2.27 |
| DART5 | 0.39 | 0.34 | 0.52 | 58064.41 | 655113.30 | 0.79 | 1.23 | 4.02 |
| DART6 | 0.41 | 0.35 | 0.53 | 57515.88 | 543595.98 | 0.70 | 0.99 | 3.25 |
| dbarts1 | 0.38 | 0.34 | 0.51 | 67785.43 | 656038.43 | 0.79 | 1.23 | 0.71 |
| dbarts2 | 0.40 | 0.35 | 0.53 | 60742.01 | 582578.44 | 0.72 | 0.99 | 0.62 |
| dbarts3 | 0.38 | 0.34 | 0.51 | 63286.26 | 588439.25 | 0.80 | 1.21 | 1.30 |
| dbarts4 | 0.40 | 0.35 | 0.53 | 63116.31 | 596129.89 | 0.70 | 0.95 | 1.25 |
| dbarts5 | 0.38 | 0.34 | 0.50 | 59622.70 | 619096.33 | 0.78 | 1.18 | 1.74 |
| dbarts6 | 0.40 | 0.35 | 0.53 | 64168.58 | 603980.38 | 0.69 | 0.94 | 1.79 |

Table A19: The posterior inclusion probabilities (PIPs) for the high dimensional real data. ‘# of Sel.Var.’ denotes the number of variables with PIPs larger than 0.5. The table includes PIPs for the selected variables based on the results of DART1. The settings for BART methods (k , the num. of trees) are as follows: 1 = (2, 100), 2 = (3, 100), 3 = (2, 200), 4 = (3, 200), 5 = (2, 300), 6 = (3, 300). BART3 and DART3 correspond to the results in main Table 8.

| | X14 | X15 | X18 | X30 | X38 | X45 | X53 | X55 | X61 | X63 | X70 | X73 | X106 | X133 | # of Sel.Var. |
|--------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---------------|
| bartMachine1 | 0.71 | 0.70 | 0.95 | 0.69 | 0.76 | 0.75 | 0.77 | 0.89 | 0.87 | 0.89 | 0.73 | 0.75 | 0.79 | 0.71 | 132 |
| bartMachine2 | 0.69 | 0.71 | 0.87 | 0.73 | 0.80 | 0.74 | 0.73 | 0.90 | 0.83 | 0.89 | 0.72 | 0.73 | 0.79 | 0.75 | 132 |
| bartMachine3 | 0.93 | 0.90 | 0.97 | 0.92 | 0.93 | 0.92 | 0.90 | 0.97 | 0.96 | 0.96 | 0.91 | 0.93 | 0.95 | 0.91 | 132 |
| bartMachine4 | 0.92 | 0.92 | 0.95 | 0.92 | 0.94 | 0.92 | 0.89 | 0.97 | 0.94 | 0.96 | 0.91 | 0.91 | 0.92 | 0.91 | 132 |
| bartMachine5 | 0.98 | 0.96 | 0.99 | 0.97 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.97 | 0.97 | 0.98 | 0.97 | 132 |
| bartMachine6 | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.97 | 0.98 | 0.98 | 0.97 | 132 |
| BART1 | 0.70 | 0.66 | 0.84 | 0.59 | 0.71 | 0.73 | 0.61 | 0.73 | 0.83 | 0.90 | 0.66 | 0.63 | 0.73 | 0.63 | 97 |
| BART2 | 0.61 | 0.72 | 0.81 | 0.71 | 0.81 | 0.68 | 0.65 | 0.75 | 0.80 | 0.91 | 0.68 | 0.59 | 0.69 | 0.66 | 96 |
| BART3 | 0.90 | 0.87 | 0.95 | 0.91 | 0.93 | 0.9 | 0.86 | 0.92 | 0.96 | 0.97 | 0.86 | 0.85 | 0.88 | 0.87 | 132 |
| BART4 | 0.87 | 0.87 | 0.89 | 0.91 | 0.95 | 0.89 | 0.88 | 0.90 | 0.91 | 0.97 | 0.84 | 0.82 | 0.86 | 0.90 | 131 |
| BART5 | 0.96 | 0.95 | 0.96 | 0.94 | 0.97 | 0.97 | 0.95 | 0.96 | 0.98 | 0.99 | 0.95 | 0.95 | 0.92 | 0.94 | 132 |
| BART6 | 0.94 | 0.96 | 0.95 | 0.97 | 0.98 | 0.97 | 0.94 | 0.94 | 0.97 | 0.99 | 0.95 | 0.92 | 0.92 | 0.96 | 132 |
| DART1 | 0.53 | 0.53 | 0.77 | 0.57 | 0.57 | 0.65 | 0.51 | 0.63 | 0.75 | 0.85 | 0.56 | 0.51 | 0.61 | 0.51 | 14 |
| DART2 | 0.46 | 0.49 | 0.59 | 0.56 | 0.66 | 0.51 | 0.46 | 0.6 | 0.65 | 0.87 | 0.57 | 0.46 | 0.44 | 0.50 | 11 |
| DART3 | 0.68 | 0.58 | 0.81 | 0.73 | 0.75 | 0.70 | 0.61 | 0.81 | 0.78 | 0.92 | 0.71 | 0.55 | 0.66 | 0.65 | 58 |
| DART4 | 0.67 | 0.65 | 0.62 | 0.71 | 0.75 | 0.68 | 0.55 | 0.59 | 0.80 | 0.92 | 0.61 | 0.47 | 0.53 | 0.58 | 41 |
| DART5 | 0.78 | 0.72 | 0.76 | 0.68 | 0.77 | 0.82 | 0.74 | 0.77 | 0.82 | 0.98 | 0.75 | 0.59 | 0.64 | 0.76 | 70 |
| DART6 | 0.65 | 0.75 | 0.72 | 0.75 | 0.86 | 0.71 | 0.67 | 0.62 | 0.82 | 0.95 | 0.66 | 0.53 | 0.50 | 0.71 | 55 |
| dbarts1 | 0.67 | 0.64 | 0.87 | 0.68 | 0.67 | 0.69 | 0.68 | 0.85 | 0.83 | 0.87 | 0.68 | 0.67 | 0.77 | 0.62 | 132 |
| dbarts2 | 0.62 | 0.64 | 0.81 | 0.66 | 0.73 | 0.67 | 0.70 | 0.81 | 0.84 | 0.85 | 0.67 | 0.69 | 0.75 | 0.63 | 132 |
| dbarts3 | 0.89 | 0.87 | 0.95 | 0.87 | 0.91 | 0.86 | 0.92 | 0.96 | 0.94 | 0.97 | 0.88 | 0.91 | 0.92 | 0.85 | 132 |
| dbarts4 | 0.86 | 0.87 | 0.92 | 0.9 | 0.91 | 0.88 | 0.91 | 0.94 | 0.91 | 0.95 | 0.88 | 0.87 | 0.90 | 0.85 | 132 |
| dbarts5 | 0.97 | 0.95 | 0.98 | 0.96 | 0.97 | 0.96 | 0.95 | 0.98 | 0.97 | 0.99 | 0.95 | 0.96 | 0.97 | 0.94 | 132 |
| dbarts6 | 0.96 | 0.95 | 0.97 | 0.96 | 0.97 | 0.96 | 0.95 | 0.98 | 0.97 | 0.98 | 0.96 | 0.96 | 0.97 | 0.95 | 132 |

Table A20: The variable importance values for the high dimensional real data. Each value denotes the order of variable importance value and actual variable importance value in parenthesis. The list of the variables in this table follows Table 19. The settings for RF methods (the num. of sampled variables, the num. of trees) are as follows: 1 = (\sqrt{p} , 500), 2 = ($2\sqrt{p}$, 500), 3 = (\sqrt{p} , 1,000), 4 = ($2\sqrt{p}$, 1,000), 5 = (\sqrt{p} , 2,000), 6 = ($2\sqrt{p}$, 2,000), 7 = (\sqrt{p} , 10,000), 8 = ($2\sqrt{p}$, 10,000).

| | X14 (VI) | X15 (VI) | X18 (VI) | X30 (VI) | X38 (VI) | X45 (VI) | X53 (VI) |
|---------------|------------|-----------|------------|-----------|-----------|------------|-----------|
| ranger1 | 118 (0.00) | 67 (0.00) | 103 (0.00) | 46 (0.00) | 86 (0.00) | 43 (0.01) | 73 (0.00) |
| ranger2 | 98 (0.00) | 55 (0.00) | 106 (0.00) | 34 (0.01) | 78 (0.00) | 38 (0.01) | 63 (0.00) |
| ranger3 | 106 (0.00) | 63 (0.00) | 104 (0.00) | 53 (0.00) | 89 (0.00) | 43 (0.00) | 83 (0.00) |
| ranger4 | 102 (0.00) | 49 (0.00) | 103 (0.00) | 38 (0.01) | 86 (0.00) | 43 (0.00) | 76 (0.00) |
| ranger5 | 106 (0.00) | 63 (0.00) | 102 (0.00) | 44 (0.00) | 95 (0.00) | 46 (0.00) | 87 (0.00) |
| ranger6 | 109 (0.00) | 50 (0.00) | 102 (0.00) | 45 (0.00) | 68 (0.00) | 42 (0.00) | 79 (0.00) |
| ranger7 | 109 (0.00) | 59 (0.00) | 105 (0.00) | 48 (0.00) | 89 (0.00) | 41 (0.01) | 85 (0.00) |
| ranger8 | 101 (0.00) | 47 (0.00) | 104 (0.00) | 38 (0.01) | 82 (0.00) | 41 (0.00) | 71 (0.00) |
| randomForest1 | 103 (0.00) | 48 (0.00) | 106 (0.00) | 52 (0.00) | 81 (0.00) | 39 (0.01) | 75 (0.00) |
| randomForest2 | 113 (0.00) | 52 (0.00) | 100 (0.00) | 33 (0.01) | 90 (0.00) | 50 (0.00) | 75 (0.00) |
| randomForest3 | 110 (0.00) | 51 (0.00) | 105 (0.00) | 53 (0.00) | 94 (0.00) | 37 (0.01) | 84 (0.00) |
| randomForest4 | 94 (0.00) | 46 (0.00) | 107 (0.00) | 31 (0.01) | 71 (0.00) | 42 (0.00) | 74 (0.00) |
| randomForest5 | 108 (0.00) | 47 (0.00) | 105 (0.00) | 48 (0.00) | 88 (0.00) | 36 (0.01) | 76 (0.00) |
| randomForest6 | 99 (0.00) | 49 (0.00) | 102 (0.00) | 46 (0.00) | 80 (0.00) | 37 (0.01) | 71 (0.00) |
| randomForest7 | 111 (0.00) | 59 (0.00) | 104 (0.00) | 49 (0.00) | 86 (0.00) | 39 (0.01) | 80 (0.00) |
| randomForest8 | 104 (0.00) | 48 (0.00) | 102 (0.00) | 40 (0.00) | 82 (0.00) | 41 (0.00) | 73 (0.00) |
| | X55 (VI) | X61 (VI) | X63 (VI) | X70 (VI) | X73 (VI) | X106 (VI) | X133 (VI) |
| ranger1 | 12 (0.02) | 26 (0.01) | 78 (0.00) | 6 (0.02) | 59 (0.00) | 95 (0.00) | 21 (0.01) |
| ranger2 | 13 (0.02) | 21 (0.01) | 96 (0.00) | 6 (0.02) | 42 (0.01) | 76 (0.00) | 19 (0.01) |
| ranger3 | 14 (0.02) | 23 (0.01) | 96 (0.00) | 6 (0.02) | 56 (0.00) | 79 (0.00) | 22 (0.01) |
| ranger4 | 13 (0.02) | 21 (0.01) | 95 (0.00) | 6 (0.03) | 40 (0.00) | 80 (0.00) | 19 (0.01) |
| ranger5 | 13 (0.02) | 20 (0.01) | 88 (0.00) | 6 (0.02) | 53 (0.00) | 79 (0.00) | 22 (0.01) |
| ranger6 | 13 (0.02) | 21 (0.01) | 92 (0.00) | 6 (0.03) | 47 (0.00) | 82 (0.00) | 19 (0.01) |
| ranger7 | 13 (0.02) | 21 (0.01) | 88 (0.00) | 6 (0.02) | 55 (0.00) | 82 (0.00) | 22 (0.01) |
| ranger8 | 13 (0.02) | 21 (0.01) | 91 (0.00) | 6 (0.03) | 48 (0.00) | 78 (0.00) | 19 (0.01) |
| randomForest1 | 15 (0.01) | 23 (0.01) | 98 (0.00) | 7 (0.02) | 59 (0.00) | 76 (0.00) | 18 (0.01) |
| randomForest2 | 10 (0.02) | 22 (0.01) | 81 (0.00) | 6 (0.03) | 36 (0.01) | 106 (0.00) | 15 (0.01) |
| randomForest3 | 13 (0.02) | 22 (0.01) | 87 (0.00) | 8 (0.02) | 46 (0.00) | 83 (0.00) | 19 (0.01) |
| randomForest4 | 13 (0.02) | 22 (0.01) | 90 (0.00) | 6 (0.03) | 43 (0.00) | 78 (0.00) | 18 (0.01) |
| randomForest5 | 13 (0.02) | 20 (0.01) | 100 (0.00) | 6 (0.02) | 57 (0.00) | 83 (0.00) | 23 (0.01) |
| randomForest6 | 12 (0.02) | 20 (0.01) | 87 (0.00) | 6 (0.03) | 45 (0.00) | 84 (0.00) | 18 (0.01) |
| randomForest7 | 13 (0.02) | 20 (0.01) | 91 (0.00) | 6 (0.02) | 50 (0.00) | 81 (0.00) | 22 (0.01) |
| randomForest8 | 13 (0.02) | 21 (0.01) | 87 (0.00) | 6 (0.03) | 47 (0.00) | 80 (0.00) | 19 (0.01) |

References

- Breiman L (2001). Random forests, *Machine Learning*, **45**, 5–32.
- Breiman L, Friedman JH, Olshen R, and Stong CJ (1984). *Classification and Regression Trees*, Routledge, New York.
- Bühlmann P and Van De Geer S (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Science & Business Media, New York.
- Chipman HA, George EI, and McCulloch RE (1998). Bayesian CART model search, *Journal of the American Statistical Association*, **93**, 935–948.
- Chipman HA, George EI, and McCulloch RE (2010). BART: Bayesian additive regression trees, *The Annals of Applied Statistics*, **4**, 266–298.
- Fox EW, Hill RA, Leibowitz SG, Olsen AR, Thornbrugh DJ, and Weber MH (2017). Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology,

- Environmental Monitoring and Assessment*, **189**, 1–20.
- Friedman JH (1991). Multivariate adaptive regression splines, *The Annals of Statistics*, **19**, 1–141.
- Gunduz N and Fokoué E (2015). Robust classification of high dimension low sample size data, *arXiv:1501.00592*.
- Hernández B, Raftery AE, Pennington SR, and Parnell AC (2018). Bayesian additive regression trees using Bayesian model averaging, *Statistics and Computing*, **28**, 869–890.
- Janitza S, Celik E, and Boulesteix AL (2018). A computationally fast variable importance test for random forests for high-dimensional data, *Advances in Data Analysis and Classification*, **12**, 885–915.
- Kapelner A and Bleich J (2015). Prediction with missing data via Bayesian additive regression trees, *Canadian Journal of Statistics*, **43**, 224–239.
- Kapelner A and Bleich J (2016). bartMachine: Machine learning with Bayesian additive regression trees, *Journal of Statistical Software*, **70**, 1–40.
- Kern C, Klausch T, and Kreuter F (2019). Tree-based machine learning methods for survey research, *Survey Research Method*, **13**, 73–93.
- Kuhn M and Johnson K (2013). Applied predictive modeling, Springer, New York.
- Liaw A and Wiener M (2002). Classification and regression by randomForest, *R News*, **2**, 18–22.
- Linero AR (2018). Bayesian regression trees for high-dimensional prediction and variable selection, *Journal of the American Statistical Association*, **113**, 626–636.
- Rubin DB (1976). Inference and missing data, *Biometrika*, **63**, 581–592.
- Sparapani R, Spanbauer C, and McCulloch R (2021). Nonparametric machine learning and efficient computation with Bayesian additive regression trees: the BART R package, *Journal of Statistical Software*, **97**, 1–66.
- Stekhoven DJ and Bühlmann P (2012). MissForest—non-parametric missing value imputation for mixed-type data, *Bioinformatics*, **28**, 112–118.
- Strobl C, Boulesteix AL, Zeileis A, and Hothorn T (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics*, **8**, 1–21.
- Tang F and Ishwaran H (2017). Random forest missing data algorithms, *Statistical Analysis and Data Mining*, **10**, 363–377.
- Waldmann P (2016). Genome-wide prediction using Bayesian additive regression trees, *Genetics Selection Evolution*, **48**, 1–12.
- Wright MN, Wager S, and Probst P (2020). Ranger: A fast implementation of random forests, R package version 0.12.1.
- Zhang H, Zimmerman J, Nettleton D, and Nordman DJ (2019). Random forest prediction intervals, *The American Statistician*, **74**, 392–406.