Quantile oriented sensitivity analysis with random forest based on pinball loss

Ri Wang, Véronique Maume-Deschamps, Clémentine Prieur

Quantile oriented sensitivity analysis with random forest based on pinball loss

Ri Wang[(1)], Véronique Maume-Deschamps[(1)] and Clémentine Prieur[(2)]
[(1)]Universite Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, ICJ UMR5208, 69622 Villeurbanne, France.
[(2)]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000, Grenoble, France

Abstract

Global Sensitivity Analysis (GSA) is an important tool to better understand the behavior of black box models. Among the numerous methods for GSA, variance-based approaches have received much attention. Only a few papers focus on Quantile Oriented Sensitivity Analysis (QOSA), which can help in analysing the behavior of the response at different quantile levels. Moreover, existing QOSA estimation methods have flaws: bias when input variables are dependent, loss of accuracy and efficiency as input space dimension increases. In this paper, we propose a new estimation procedure of QOSA indices based on the notion of projected random forest, with the initial random forest built from a criterion designed for quantiles: the pinball loss also known as quantile loss.

Key words: Quantile Oriented Sensitivity Analysis, random forest, conditional distribution, conditional quantile, projected forest.

1 Introduction

Sensitivity Analysis (SA), as defined in Saltelli et al. (2004), is the study of how the uncertainty in the output of a system can be divided and allocated to different sources of uncertainty in its inputs. It is an invaluable tool to understand the behavior of numerical models, determining the most contributing input variables and ascertaining interaction effects within model (see, e.g., Da Veiga et al. (2021) for a recent review). Global Sensitivity Analysis (GSA) focuses on the impact of inputs on some output of interest, when varied over their whole domain (Iooss and Lemaître (2015)). Variance-based methods are well-established and widely used for GSA. In this context, Sobol' indices introduced in Sobol' (1993) are very popular.

In the present paper, we focus on Quantile Oriented Sensitivity Analysis (QOSA), a special case of indices based on contrast functions considered in Fort et al. (2016). QOSA indices are designed to capture the impact of inputs

at a given response quantile level. Browne et al. (2017); Maume-Deschamps and Niang (2018) proposed a kernel based estimator of first-order QOSA indices. Elie-Dit-Cosaque and Maume-Deschamps (2024) introduced an estimator based on random forest. Both approaches require tuning parameters and behave badly as dimension increases. In particular, computation burden of the approach in Elie-Dit-Cosaque and Maume-Deschamps (2024) increases dramatically with input dimension as their approach requires to build a different forest for each QOSA index, in other words, the estimation of the full set of first-order QOSA indices requires to build p different forests, with p the input space dimension.

In this paper, we introduce a new estimation procedure for QOSA indices. This estimation procedure is mainly based on three ingredients. First we leverage the notion of projected random forest (see Bénard et al. (2022a)) to estimate, for a given quantile level, QOSA indices of any order with only one initial forest. Secondly, to build the initial forest, we use a criterion based on the pinball loss, that Bhat et al. (2015) introduced to build a single quantile oriented decision tree. Finally, in order to avoid overfitting, we use out-of-bag (OOB) samples. Consistency results are proven.

The paper is organized as follows. In Section 2, we recall the definition of QOSA indices and background knowledge on random forests. Section 3 presents the projected random forest idea and consistency results for conditional distribution functions and conditional quantiles. Section 4 is devoted to consistency results for QOSA index estimators. Numerical study and application are shown in Section 5. Some supplemental materials are also provided in Appendix A (technical proofs).

2 General framework and tools

In this section, we recall the definition of quantile oriented sensitivity indices, the so-called QOSA indices, of any order. Then, we recall basics on random forests and introduce a new criterion to build quantile oriented random forests.

2.1 First-order QOSA indices

We are considering the usual framework in sensitivity analysis, that is $Y = m(\boldsymbol{X})$ with Y a scalar response, $\boldsymbol{X} = (X_1, ..., X_p) \in \mathcal{X} \subseteq \mathbb{R}^p$ a p-dimensional input vector. One could also consider $Y = m(\boldsymbol{X}) + \varepsilon$ with ε a random noise independent on $\boldsymbol{X}$, this setting is closer to statistical frameworks. The function $m(\cdot)$ is a deterministic model which could be unknown and / or computationally heavy (see Da Veiga et al. (2021) for a more detailed presentation of GSA). Consider input-output samples $\mathcal{D}_n = \{Y^i, \boldsymbol{X}^i\}_{i=1}^n, \boldsymbol{X}^i = (X_1^i, X_2^i, ..., X_p^i)$. For any subset $U \subseteq \{1, 2, ..., p\}$ and $\boldsymbol{x} \in \mathbb{R}^p$, define $\boldsymbol{x}_U =$

$(x_j, j \in U)$. We shall also consider $\mathcal{X}_U$ the trace of $\mathcal{X}$ in $\mathbb{R}^{|U|}$: $\mathcal{X}_U = \{\boldsymbol{x} \in \mathbb{R}^{|U|} \,/\, \exists\ \tilde{\boldsymbol{x}} \in \mathcal{X},\ \tilde{\boldsymbol{x}}_U = \boldsymbol{x}\}$. Our target is to quantify the impact of X_U on the output Y at different quantile levels. Let $\alpha \in (0,1)$, the QOSA index with respect to the j-th feature at level α, is defined as:

$$\begin{aligned} S_j^\alpha &= \frac{\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right] - \mathbb{E}\left[\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta) \mid X_j\right]\right]}{\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right]} \\ &= \frac{\mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha(Y)\right)\right] - \mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha\left(Y \mid X_j\right)\right)\right]}{\mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha(Y)\right)\right]} = 1 - \frac{\mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha\left(Y \mid X_j\right)\right)\right]}{\mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha(Y)\right)\right]} \end{aligned} \tag{2.1}$$

with $\psi_\alpha(y,\theta) = (y-\theta)(\alpha - \mathbb{I}_{\{y\le\theta\}})$ the pinball function (also known as check function, introduced in Koenker and Hallock (2001)). We denote by $q^\alpha(Y)$ the α-level quantile of Y and $q^\alpha\left(Y \mid X_j\right)$ the conditional α-level quantile of Y given X_j. QOSA indices are a particular case of sensitivity indices based on a contrast function, first introduced in Fort et al. (2016).

2.2 Higher order QOSA indices

Following Fort et al. (2016), we define for any $U \subseteq \{1,\dots,p\}$, higher order QOSA indices as:

$$S_U^\alpha = \frac{\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right] - \mathbb{E}\left[\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta) \mid X_U\right]\right]}{\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right]}\,. \tag{2.2}$$

In the particular case of second-order QOSA indices we get:

$$S_{ij}^\alpha = \frac{\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right] - \mathbb{E}\left[\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta) \mid X_i, X_j\right]\right]}{\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right]}\,.$$

With S_U^α, we measure the contribution on the output Y of the set of features indexed by U, from a quantile perspective. The aim of this paper is to estimate QOSA indices of any order by building a quantile oriented random forest and using its projection for subset U.
Remark that in Kala (2019), another definition for higher order QOSA indices is given. This last definition is similar to higher order Sobol' indices but since in for QOSA indices there is no ANOVA like decomposition, we think the above definition is more relevant.

2.3 Random forests

The idea of random forest proposed by Breiman (2001) is to build many randomized CART trees (Breiman et al. (1984)) at first and then average the estimation from different trees as the final estimation. Building different trees from the same dataset requires randomness in the tree growing process. Breiman (2001) proposed to introduce two sources of randomness in the tree construction. The first one is a subsampling step preliminary to each

tree construction. Like the bootstrap strategy from Efron (1992), Breiman (2001) randomly samples n observations with replacement from the original dataset $\mathcal{D}_n$ to construct a new dataset $\mathcal{D}_n^\star = \big((Y^{\star i}, \boldsymbol{X}^{\star i}), i = 1, \ldots, n\big)$, then grows the tree with $\mathcal{D}_n^\star$. The second randomness source appears during the growing tree process. Instead of optimizing the splitting criterion over all p variables, only part of the features is considered. Each tree is built node by node in a greedy fashion. For each node, the CART splitting criterion is optimized over a subset of features, $\mathcal{M}_{try} \subseteq \{1, 2, ..., p\}$, randomly selected, with the constraint that each feature has a positive probability to be chosen. For each node $A \subseteq \mathcal{X}$, the best split selects a variable X_{j_A} in $\mathcal{M}_{try}$ and a threshold z_A to maximize the CART splitting criterion, whose definition is recalled in (2.3). This splitting criterion measures the decrease of output variance between the parent node and the child nodes. The CART splitting criterion is defined as:

$$\begin{aligned} L_A^n(j,z) = & \frac{1}{N_n^\star(A)} \sum_{i=1}^{n} \left(Y^i - \bar{Y}_A\right)^2 \mathbb{I}_{\{\boldsymbol{X}^i \in A\}} - \\ & \frac{1}{N_n^\star(A)} \sum_{i=1}^{n} \left(Y^i - \bar{Y}_{A_L} \mathbb{I}_{\left\{X_j^i \leq z\right\}} - \bar{Y}_{A_R} \mathbb{I}_{\left\{X_j^i > z\right\}} \right)^2 \mathbb{I}_{\{\boldsymbol{X}^i \in A\}}, \end{aligned} \tag{2.3}$$

where for any $j \in \mathcal{M}_{try}$, for any $z \in \mathbb{R}$, the left child node A_L and the right child node A_R are defined as $A_L = \{\boldsymbol{x} \in A,\, x_j \leq z\}$, $A_R = \{\boldsymbol{x} \in A,\, x_j > z\}$, $N_n^\star(A) = \#\{i = 1, \ldots, n \;/\; \boldsymbol{X}^{\star i} \in A\}$ and $\bar{Y}_A$ is the empirical mean of Y on A on the bootstrap sample: $\bar{Y}_A = \dfrac{1}{N_n^\star(A)} \displaystyle\sum_{i=1}^{n} Y^{\star i} \mathbb{I}_{\{\boldsymbol{X}^{\star i} \in A\}}$. To build each tree of a forest composed with B trees, we choose randomly for each tree a bootstrap sample $\mathcal{D}_n^\star(b)$. This random choice is modeled by a random vector of indices Θ_b^1. We also select randomly splitting candidate features in each cell of the tree. This random selection is modeled by a random vector of indices Θ_b^2. The vectors $\Theta_b = \big(\Theta_b^1, \Theta_b^2\big)$, $b = 1, \ldots, B$, are sampled independently from each other and independently from the initial sample $\mathcal{D}_n$. Then, for a new query point $\boldsymbol{x} \in \mathcal{X}$, $m(\boldsymbol{x})$ is estimated from the b-th tree as:

$$m_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n) = \frac{1}{N_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)} \sum_{i=1}^{n} B_i(\Theta_b^1, \mathcal{D}_n) \mathbb{I}_{\{\boldsymbol{X}^i \in A_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)\}} Y^i \tag{2.4}$$

with $B_i(\Theta_b^1, \mathcal{D}_n)$ the number of occurrences of $\boldsymbol{X}^i$ in $\mathcal{D}_n^\star(b)$, $A_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)$ the leaf in which $\boldsymbol{x}$ falls and $N_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)$ the number of elements of $\mathcal{D}_n^\star(b)$ in $A_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)$. The estimators from different trees are aggregated as:

$$m_n^B(\boldsymbol{x}; \Theta_1, ..., \Theta_b, \mathcal{D}_n) = \frac{1}{B} \sum_{b=1}^{B} m_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n). \tag{2.5}$$

Equation (2.5) rewrites as:

$$m_n^B(\boldsymbol{x}; \Theta_1, ..., \Theta_b, \mathcal{D}_n) = \sum_{i=1}^{n} w_i(\boldsymbol{x}) Y^i, \tag{2.6}$$

with

$$w_i(\boldsymbol{x}) = \frac{1}{B} \sum_{b=1}^{B} \frac{B_i(\Theta_b^1, \mathcal{D}_n) \mathbb{I}_{\{\boldsymbol{X}^i \in A_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)\}}}{N_n(\boldsymbol{x}; \Theta_b, \mathcal{D}_n)}. \tag{2.7}$$

Meinshausen and Ridgeway (2006) proposed to replace Y with $\mathbb{I}_{\{Y \leq y\}}$ in (2.6) in order to estimate the conditional distribution:

$$\hat{F}_{Y|\boldsymbol{X}=\boldsymbol{x}}(y) = \sum_{i=1}^{n} w_i(\boldsymbol{x}) \mathbb{I}_{\{Y^i \leq y\}}.$$

Finally, the conditional quantile function can be estimated by the generalized inverse of $\hat{F}_{Y|\boldsymbol{X}=\boldsymbol{x}}(\cdot)$.

However CART-split criterion is sensitive to changes in conditional mean of Y given $\boldsymbol{X}$ and is not designed for changes in conditional quantiles. It is one of the reasons why alternative splitting criteria, more focused on quantiles, have been introduced. Bhat et al. (2015) uses the pinball loss to grow one quantile regression tree. They also propose an online update algorithm to improve the search efficiency. Athey et al. (2019) proposes a general framework, gradient tree, which could be applied for different regression tasks. In the quantile regression case, their splitting criterion is similar to Gini impurity for classification problem. Ćevid et al. (2022) consider the Maximum Mean Discrepancy (MMD), a distributional metric between left and right child node. They propose a fast approximation algorithm. Because the pinball loss is the quantile loss function, we generalize Bhat et al. (2015)'s splitting criterion to build a random forest. More precisely, we would like to maximize the quantity:

$$\begin{aligned} E[\psi_\theta(Y)|X \in A] - P[X \in A_L|X \in A] E[\psi_\alpha(Y, \theta_L)|X \in A_L] \\ - P[X \in A_R|X \in A] E[\psi_\alpha(Y, \theta_R)|X \in A_R], \end{aligned} \tag{2.8}$$

with $\psi_\alpha(y, \theta) = (y - \theta)(\alpha - \mathbb{I}_{\{y<\theta\}})$, $\alpha \in (0, 1)$, θ_L and θ_R the α-level quantile in left and right child node respectively. In practice, we optimize the empirical form of (2.8). Note that the first term in (2.8) can be ignored in the optimization because it does not depend on A_L and A_R. Thus we shall minimize the empirical counterpart of the last two terms in (2.8):

$$\frac{n_L}{n_P} \frac{1}{n_L} \sum_{X_i \in A_L} \psi_\alpha(Y^i, \hat{\theta}_L) + \frac{n_R}{n_P} \frac{1}{n_R} \sum_{X_i \in A_R} \psi_\alpha(Y^i, \hat{\theta}_R), \tag{2.9}$$

where n_L and n_R are the sample size in left and right child nodes respectively, $\hat{\theta}_L$ and $\hat{\theta}_R$ are the quantile estimation in each child node.

In the next two sections we explain how the projection algorithm introduced in Bénard et al. (2022a) can be applied on our forest to estimate the conditional distribution function of the output and conditional quantiles.

To conclude this section, let us show a simple tree structure and introduce some notation for the rest of the paper. As shown in Fig.1, there are

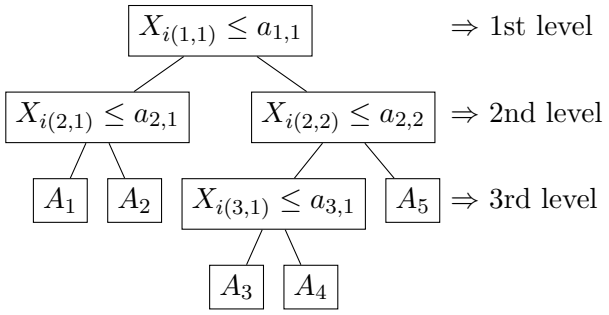

Figure 1: Decision tree structure

two kinds of nodes in a tree. Terminal leaf nodes are denoted by A_i. The other nodes are labeled by a pair of indices (l, k), where l is the tree level and k is the index of each node at l-level. We denote $i(l, k) \in \{1, 2, ..., p\}$ and $a_{l,k} \in \mathbb{R}$ the index of the feature and the corresponding threshold selected by optimizing the splitting criterion (2.9) at node (l, k).

3 Projected random forests

In this section, we present the notion of projected forest in Bénard et al. (2022a) which uses an idea from Lundberg and Lee (2017). The key idea is to build one random forest and then to project it adequately to estimate the indices S_U^α. This reduces computational burden in high dimension. The algorithm is available for all $U \subseteq \{1, 2, ..., p\}$. The details can be found in Algo. 1. The projected idea is as in Bénard et al. (2022a). The conditional quantile estimation is then similar to the random forest estimation using the projected leaves concept. So, we only display the core part of projecting in Algo. 1. The illustration in one dimension is shown in Fig. 2.

3.1 Estimation of conditional cumulative distribution functions

We shall need the following properties on tree construction. These construction hypotheses are also done in Athey et al. (2019) and Bénard et al. (2022a).

(P1) [Data sampling] Rather than using bootstrap sampling with replacement, we use subsampling with growing size s_n such that $s_n/n = \kappa$,

Algorithm 1: Projected Random Forest (Bénard et al. (2022a))

Inputs : A random forest, $\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_B$, fit with $\mathcal{D}_n$,
A variable subset $U \subseteq \{1, 2, ..., p\}$,
A query point $\boldsymbol{x} = (x_1, x_2, ..., x_p)$.

Outputs: Projected leafs: $\mathcal{A}(\boldsymbol{x}|\Theta_1, U), \mathcal{A}(\boldsymbol{x}|\Theta_2, U), ..., \mathcal{A}(\boldsymbol{x}|\Theta_B, U)$

for *all trees in the forest* **do**
  `/* Step 1: initialize variables */`
  initialize *nodes_level* as a list of nodes containing only the root node;
  initialize *nodes_child* as an empty list of child nodes;
  initialize *samples* as the list of observation indices of the full training data of the tree;
  for *all levels in the tree* **do**
    `/* Step 2: drop` x_U `to the next tree level with the relevant training samples */`
    for *all nodes in nodes_level* **do**
      if *the node splits on a variable in* U, **then**
        compute whether $\boldsymbol{x}_U$ falls in the left or right child node;
        append the child node to *nodes_child*;
        set *samples_child* as the observations in *samples* which satisfy the split
      else
        append both the left and right children nodes to *nodes_child*;
      end
      if *the size of samples_child is lower then min_node_size* **then**
        break the loop through the tree levels;
      else
        set *samples* = *samples_child*;
      end
    end
    set *nodes_level* = *nodes_child*;
  end
  $\mathcal{A}(\boldsymbol{x}|\Theta_b, U) = samples$;
end

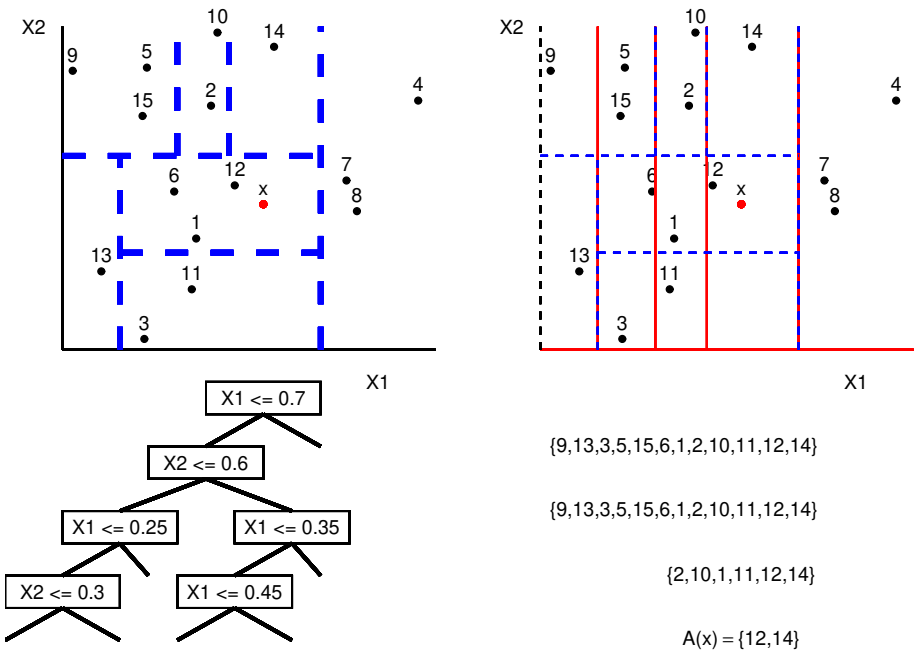

Figure 2: Projected algorithm illustration for $p = 2, U = \{1\}$. Left panel represents the structure of the original tree and partition in $\mathbb{R}^2$. Each dash line represents a node in the tree. Right panel is the projected leaf, which could be tracked by level.

where $\kappa \in (0,1)$ is a fixed number. For convenience, we reuse the notation $\mathcal{D}_n^*(b)$ for subsamples.

(P2) [γ-regularity] Each split leaves at least a fraction $\gamma \in (0, 0.5)$ of the available training sample on each side.

(P3) [Random-split] At each tree node, the number mtry of candidate variables drawn to optimize the split is set to mtry $= 1$ with a small probability $\pi > 0$ and for mtry $= 1$ and any $j = 1, \ldots, p$ the probability that the split occurs along feature X_j is set to π/p.

Let us consider the following estimator of the conditional distribution function $y \mapsto F_Y(y|\boldsymbol{X}_U) = \mathbb{P}(Y \leq y|\boldsymbol{X}_U)$: for $\boldsymbol{x} \in \mathcal{X}_U$, $U \subseteq \{1, \ldots, p\}$

$$w_i(\boldsymbol{x}|U) = \frac{1}{B}\sum_{b=1}^{B} \frac{\mathbb{I}_{\{\boldsymbol{X}^i \in A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), U)\}}}{N(\boldsymbol{x}; \mathcal{D}_n^*(b), \mathcal{D}_n^*(b), U)},$$

$$\hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}) \quad = \quad \textstyle\sum_{i=1}^{n} w_i(\boldsymbol{x}|U)\mathbb{I}_{(Y^i \leq y)},$$

where $A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), U)$ is the projected leaf containing $\boldsymbol{x}$ and $N(\boldsymbol{x}; \mathcal{D}_n^*(b), \mathcal{D}_n^*(b), U)$ is the number of elements in $\mathcal{D}_n^*(b)$ which belong to $A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), U)$. In what follows, we will refer to $\hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x})$ as the *standard estimator*. Also, for any $j \in \{1, \ldots, n\}$, we denote by $\hat{F}_{s_n,n-1}(y|\boldsymbol{X}_U^j)$ the standard estimator built with $\mathcal{D}_n \setminus (\boldsymbol{X}^j, Y^j)$ and with sample size a_n.

Assumption 1. *The covariate vector $X = (X_1, \ldots, X_p)$ admits a density over $\mathcal{X}$ bounded from below by a strictly positive constant.*

Assumption 2 (sample size in a projected leaf)**.** *For fixed $\beta > 1$ and $C > 0$, for any $U \subseteq \{1, 2, ..., p\}$, and $\boldsymbol{x}$ in $\mathcal{X}_U$, $N(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), \mathcal{D}_n^*(\Theta), U) \geq C\sqrt{n}(\ln n)^{\beta}$.*

The number of elements in a leaf is an hyper-parameter in the building of the forest, so it can be controlled.

Assumption 3. *Assumption on B (number of trees), $B = O(n^{\beta})$, for some fixed $\beta > 0$.*

Assumption 4. *For any $U \subseteq \{1, 2, ..., p\}$ and $\boldsymbol{x}$ in $\mathcal{X}_U$, the conditional distribution function $F_Y(\cdot|\boldsymbol{X}_U = \boldsymbol{x})$ is continuous and increasing, with the conditional density function $f_Y(\cdot|\mathbf{X}_U = \boldsymbol{x})$ continuous. For any $y \in \mathbb{R}$, $F_Y(y|\boldsymbol{X}_U = \cdot)$ is continuous.*

Note that properties (P1) to (P3) as far as Assumptions 1 to 4 are standard assumptions in the litterature (see, e.g., Bénard et al. (2022b); Elie-Dit-Cosaque and Maume-Deschamps (2022b)).

One of the main ingredient in the consistency proof of QOSA indices is Theorem 3.3 below. It uses as a technical tool, a dummy estimator constructed with an additional sample, whose consistency is stated in Lemma 3.1. Consider an additional sample $\mathcal{D}_n^\diamond = \big((Y^{\diamond i}, \boldsymbol{X}^{\diamond i}), i = 1, \ldots, n\big)$, independent of $\mathcal{D}_n$ with the $(Y^{\diamond i}, \boldsymbol{X}^{\diamond i})$'s independent and distributed as $(Y, \boldsymbol{X})$. The dummy estimator is defined as

$$\hat{F}^\diamond(y|\boldsymbol{X}_U = \boldsymbol{x}) = \sum_{i=1}^{n} w_i^\diamond(\boldsymbol{x}|U)\mathbb{I}_{(Y^{\diamond i} \leq y)},$$

where

$$w_i^\diamond(\boldsymbol{x}|U) = \frac{1}{B}\sum_{b=1}^{B} \frac{\mathbb{I}_{\{\boldsymbol{X}^{\diamond i} \in A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), U)\}}}{N(\boldsymbol{x}; \mathcal{D}_n^*(b), \mathcal{D}_n^\diamond, U)}, \tag{3.1}$$

$A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), U)$ is as above and $N(\boldsymbol{x}; \mathcal{D}_n^*(b), \mathcal{D}_n^\diamond, U)$ is the number of elements in $\mathcal{D}_n^\diamond$ which belong to $A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), U)$.
The following two lemmas are key ingredients in the proof of Theorem 3.3.

Lemma 3.1. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then for all* $U \subseteq \{1, 2, ..., p\}$,

$$\forall \boldsymbol{x} \in \mathcal{X}_U, \forall y \in \mathbb{R}, |\hat{F}^\diamond(y|\boldsymbol{X}_U = \boldsymbol{x}) - F(y|\boldsymbol{X}_U = \boldsymbol{x})| \xrightarrow[n\to\infty]{a.s.} 0.$$

Lemma 3.2. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then for all* $U \subseteq \{1, 2, ..., p\}$,

$$\forall \boldsymbol{x} \in \mathcal{X}_U, \forall y \in \mathbb{R}, |\hat{F}^\diamond(y|\boldsymbol{X}_U = \boldsymbol{x}) - \hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x})| \xrightarrow[n\to\infty]{a.s.} 0.$$

The proof of Lemmas 3.1 and 3.2 is the same as the one of Proposition 6.1 and Lemma 6.2 in Elie-Dit-Cosaque and Maume-Deschamps (2022b) using Corollary A.2 and Lemma A.3 in the Appendix.

Theorem 3.3. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then for all* $U \subseteq \{1, 2, ..., p\}$,

$$\forall \boldsymbol{x} \in \mathcal{X}_\mathcal{U}, \sup_{y\in\mathbb{R}} |\hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}) - F(y|\boldsymbol{X}_U = \boldsymbol{x})| \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof. The proof idea is similar to the one of Theorem 4.3 in Elie-Dit-Cosaque and Maume-Deschamps (2022b). The consistency result for the standard estimator is obtained by combining Lemmas 3.1 and 3.2 above:

$$\forall \boldsymbol{x} \in \mathcal{X}_U, \forall y \in \mathbb{R}, \hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}) \xrightarrow[n\to\infty]{a.s.} F(y|\boldsymbol{X}_U = \boldsymbol{x}).$$

Then, Dini's second theorem, leads to the uniform a.s. convergence under Assumption 4. □

Remark: For $U = \{1, 2, ..., p\}$, the projected random forest will return the original forest. So the consistency result of the original random forest is included in Theorem 3.3.
Next, we introduce the OOB estimator (Out Of Bag estimator). Recall that for each tree a subsample of size s_n is chosen, with respect to Θ_b. So that, all indices $i \in \{1, \dots, n\}$ will not participate in the tree construction. Let us consider the random set $\Lambda_{n,i}$,

$$\Lambda_{n,i} = \{b \in \{1, 2, ..., B\} : (Y^i, \boldsymbol{X}^i) \notin \mathcal{D}_n^*(\Theta_b)\},$$

it is the set of trees built without using the i-th element of $\mathcal{D}_n$. Then, for all $j = 1, \dots, n$, for all $U \subseteq \{1, \dots, p\}$, the OOB estimator is defined as

$$\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) = \sum_{i=1}^{n} w_i^{OOB}(\boldsymbol{X}^j|U)\mathbb{I}_{\{Y^i \leq y\}},$$

where $w_i^{OOB}(\boldsymbol{X}^j|U) = 0$ if $\Lambda_{n,i} = \emptyset$ and otherwise,

$$w_i^{OOB}(\boldsymbol{X}^j|U) = \frac{\mathbb{I}_{\{|\Lambda_{n,j}|>0\}}}{|\Lambda_{n,j}|} \sum_{b \in \Lambda_{n,j}} \frac{\mathbb{I}_{\{\boldsymbol{X}^i \in A(\boldsymbol{X}^j; \mathcal{D}_n^*(\Theta), U)\}}}{N(\boldsymbol{X}^j; \mathcal{D}_n^*(b), \mathcal{D}_n^*(b), U)}.$$

The L^2 consistency of the OOB estimator is stated in Theorem 3.5 below. It uses Lemma 3.4 which shows that the OOB error can be controlled by the standard one and which proof is postponed to Appendix A.

Lemma 3.4. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then for all $B \in \mathbb{N}^\star, j \in \{1, 2, ..., n\}$ and $U \subseteq \{1, 2, ..., p\}$, we have*

$$\sup_y \mathbb{E}[(\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2]$$

$$\leq \tfrac{2}{1-s_n/n} \sup_y \mathbb{E}[(\hat{F}_{B,s_n,n-1}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2] + O((s_n/n)^B)\,.$$

Theorem 3.5. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then $\forall j \in \{1, ..., n\}, \forall U \subseteq \{1, 2, ..., p\}$,*

$$\lim_{n \to \infty} \sup_y \mathbb{E}[(\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2] \to 0.$$

Proof. The proof follows directly from Lemma 3.4 which provides a relation between the OOB error and the standard error and Theorem 3.3. Remark that $\hat{F}_{B,s_n,n-1}(y|\boldsymbol{X}_U^j)$ and $F(y|\boldsymbol{X}_U^j))$ are distribution functions, so that their values are in $[0, 1]$ and the a.s. consistency from Theorem 3.3 leads to L^2 consistency. □

3.2 Conditional quantile estimation

In what follows, we are interested in the estimation of the conditional quantiles at level $\alpha \in (0,1)$ fixed:

$$q^\alpha(Y|\boldsymbol{X}_U) = \inf_{y\in\mathbb{R}}\{y,\ F_Y(y|\boldsymbol{X}_U) \geq \alpha\}.$$

We shall use the OOB estimator of the conditional distribution function to get an estimation of $q^\alpha(Y|\boldsymbol{X}_U)$:

$$\hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U = \boldsymbol{x}) = \inf_{y\in\mathbb{R}}\{y,\ \hat{F}^{OOB}(y|\boldsymbol{X}_U = \boldsymbol{x}) \geq \alpha\}.$$

It is easily noticed that

$$\hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U = \boldsymbol{x}) = \inf_{k=1,\dots,n}\{Y^k,\ \hat{F}^{OOB}(Y^k|\boldsymbol{X}_U = \boldsymbol{x}) \geq \alpha\}.$$

The standard estimator $\hat{Q}_{B,s_n,n}(\alpha|\boldsymbol{X}_U = \boldsymbol{x})$ of the conditional quantile is defined in the same way replacing $\hat{F}^{OOB}$ with $\hat{F}_{B,s_n,n}$.

In the following, in order to lighten the notation, for $\boldsymbol{x} \in \mathcal{X}_U$ we shall write $Q^\alpha(\boldsymbol{x})$ for $Q^\alpha(Y|\boldsymbol{X}_U = \boldsymbol{x})$ and $\hat{Q}_n(\alpha|\boldsymbol{x})$, for an estimator of $Q^\alpha(Y|\boldsymbol{X}_U = \boldsymbol{x})$.

Theorem 3.6. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then standard and OOB estimators of the conditional quantile are consistent. That is, if* $\boldsymbol{x} \in \mathcal{X}_U$,

$$\forall \alpha \in (0,1),\ \hat{Q}_{B,s_n,n}(\alpha|\boldsymbol{X}_U = \boldsymbol{x}) \xrightarrow[n\to\infty]{a.s.} Q(\alpha|\boldsymbol{X}_U = \boldsymbol{x}), \tag{3.2}$$

$$\forall j \in \{1,\dots,n\},\ \forall \alpha \in (0,1),\ \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^j) \xrightarrow[n\to\infty]{p} Q(\alpha|\boldsymbol{X}_U^j). \tag{3.3}$$

Proof. The proof of (3.2) follows from the result in Theorem 3.3. The one of (3.3) follows form the result in Theorem 3.5 which implies, $\forall y$, $\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) \xrightarrow[n\to\infty]{p} F(y|\boldsymbol{X}_U^j)$. □

4 QOSA index estimation

Let $U \subseteq \{1,\dots,p\}$, the first-order QOSA index estimation is obtained from (2.2) by plugin, namely $\hat{S}_U^\alpha = 1 - \dfrac{\hat{O}}{\hat{P}}$, with $\hat{O} = \frac{1}{n}\sum_{i=1}^n \psi_\alpha(Y^i, \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i))$ and $\hat{P} = \frac{1}{n}\sum_{i=1}^n \psi_\alpha(Y^i, \hat{q}^\alpha(Y))$.

Assumption 5. *(Y is bounded.)* $\exists C > 0$, *s.t.* $|Y| \leq C$ *almost surely.*

Theorem 4.1. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 5 are verified, then for all* $j \in \{1, 2, ..., n\}$ *and* $U \subseteq \{1, 2, ..., p\}$*, we have*

$$\lim_{n\to\infty} \mathbb{E}[(\hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^j) - Q(\alpha|\boldsymbol{X}_U^j))^2] \to 0\,.$$

Proof. Assumption 5 and (3.3) in Theorem 3.6 lead to the announced result □

Theorem 4.2. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 5 are verified, then the plugin QOSA estimator is consistent in probability:*

$$\hat{S}_U^\alpha \xrightarrow{p} S_U^\alpha.$$

Proof.

$$\begin{aligned}
\hat{S}_U^\alpha =& 1 - \frac{1}{\hat{P}}\frac{1}{n}\sum_{i=1}^{n}(Y^i - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i))(\alpha - \mathbb{I}_{\{Y^i \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i)\}}) \\
=& 1 - \frac{1}{\hat{P}}\frac{1}{n}\sum_{i=1}^{n}(Y^i - Q(\alpha|\boldsymbol{X}_U^i))(\alpha - \mathbb{I}_{\{Y^i \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i)\}}) \\
&+ \frac{1}{\hat{P}}\underbrace{\frac{1}{n}\sum_{i=1}^{n}(Q(\alpha|\boldsymbol{X}_U^i) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i))(\alpha - \mathbb{I}_{\{Y^i \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i)\}})}_{A} \\
=& 1 - \frac{1}{\hat{P}}\underbrace{\frac{1}{n}\sum_{i=1}^{n}(Y^i - Q(\alpha|\boldsymbol{X}_U^i))(\alpha - \mathbb{I}_{\{Y^i \le Q(\alpha|\boldsymbol{X}_U^i)\}})}_{D} \\
&+ \frac{1}{\hat{P}}\underbrace{\frac{1}{n}\sum_{i=1}^{n}(Y^i - Q(\alpha|\boldsymbol{X}_U^i)))(\mathbb{I}_{\{Y^i \le Q(\alpha|\boldsymbol{X}_U^i)\}} - \mathbb{I}_{\{Y^i \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i)\}})}_{B} + \frac{1}{\hat{P}}A\,.
\end{aligned}$$

The law of large numbers and the almost sure consistency of $\hat{q}_\alpha(Y)$ lead to, $\hat{P} \xrightarrow{a.s.} \mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha(Y)\right)\right]$ and $D \xrightarrow{a.s.} \mathbb{E}\left[\psi_\alpha\left(Y, q^\alpha\left(Y \mid \boldsymbol{X}_U\right)\right)\right]$. The consistency of $\hat{S}_U^\alpha$ is verified if both A and B terms go to 0 in probability. Now,

$$\begin{aligned}
\mathbb{E}[|A|] \le& \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}|(Q(\alpha|\boldsymbol{X}_U^i) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i))(\alpha - \mathbb{I}_{\{Y^i \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i)\}})|\right] \\
=& \mathbb{E}\left[|(Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1))(\alpha - \mathbb{I}_{\{Y \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)\}})|\right] \\
\le& \mathbb{E}|Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)|\,.
\end{aligned} \tag{4.1}$$

The second line is due to equidistribution of $(Y^i, \boldsymbol{X}^i)$. Using Eq. (3.3) in Theorem 3.6 and Assumption 5, the upper bound in (4.1) tends to zero so that by Markov's inequality, A tends to zero in probability. We next consider the B term.

$$\begin{aligned}
\mathbb{E}[|B|] =& \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(Y^i - Q(\alpha|\boldsymbol{X}_U^i))(\mathbb{I}_{\{Y^i \le Q(\alpha|\boldsymbol{X}_U^i)\}} - \mathbb{I}_{\{Y^i \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^i)\}})\right] \\
\le& \mathbb{E}\left[|(Y - Q(\alpha|\boldsymbol{X}_U^1))(\mathbb{I}_{\{Y \le Q(\alpha|\boldsymbol{X}_U^1)\}} - \mathbb{I}_{\{Y \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)\}})|\right] \\
\le& C\mathbb{E}|\mathbb{I}_{\{Y \le Q(\alpha|\boldsymbol{X}_U^1)\}} - \mathbb{I}_{\{Y \le \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)\}}| = C\mathbb{P}(Y \in (Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)) \\
=& C\mathbb{P}(Y \in (Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1), |Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)| \le \varepsilon) \\
&+ C\mathbb{P}(Y \in (Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1), |Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)| > \varepsilon) \\
\le& C\underbrace{\mathbb{P}(Y \in (Q(\alpha|\boldsymbol{X}_U^1) - \varepsilon, Q(\alpha|\boldsymbol{X}_U^1) + \varepsilon))}_{T_1} + C\underbrace{\mathbb{P}(|Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)| > \varepsilon)}_{T_2}
\end{aligned}$$

Under Assumption 4, $\lim_{\varepsilon \to 0} T_1 = \mathbb{P}(Y = Q(\alpha|\boldsymbol{X}_U)) = 0$. By Chebyshev's inequality, $T_2 \le \frac{1}{\varepsilon^2}\mathbb{E}\left[|Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)|^2\right]$. With Theorem 4.1, for any $\delta > 0$, we can find n_0 such that $\forall n > n_0, \mathbb{E}\left[|Q(\alpha|\boldsymbol{X}_U^1) - \hat{Q}_n^{OOB}(\alpha|\boldsymbol{X}_U^1)|^2\right] < \delta\varepsilon^2$, which means $T_2 < \delta$. Finally $\mathbb{E}[|B|]$ goes to 0. Then B goes to 0 in probability by Markov's inequality. This achieves the proof of Theorem 4.2, that is the convergence in probability of $\hat{S}_U^\alpha$.

□

5 Numerical study and application

This section is devoted to numerical experiments. In order to evaluate the performance of our estimation procedure, we first consider models of the form $Y = m(\boldsymbol{X})$ for specific m and input probability distribution $\mathbb{P}_{\boldsymbol{X}}$ for which the theoretical value of QOSA indices can be computed analytically. We first validate on simulated data our estimation procedure of first-order QOSA indices in Section 5.1, comparing to the state of the art. Then in Section 5.2 we focus on higher order QOSA index estimation for which, to our knowledge, no estimation procedure was proposed untill now. Finally in Section 5.3 we apply our estimation methodology to an environmental real dataset.

5.1 First-order QOSA index estimation

In this section, we aim at comparing our pinball estimation procedure for first-order QOSA indices with two alternative procedures from the state of the art:

(i) the kernel-based method (Browne et al. (2017); Maume-Deschamps and Niang (2018)) available in the `R` package `sensitivity`, with the default setting, that is Gaussian kernel and the bandwidth selected following Wand et al. (1994);

(ii) the estimator introduced in (Elie-Dit-Cosaque and Maume-Deschamps, 2022b, Section 4.2.1) based on CART random forests constructed with only one variable, available in the `Python` package `qosa`, with the number of trees set to 100 and *min_node_size* optimized by cross-validation.

In the following we denote by *pinball* the new estimation procedure we introduced in this paper, and by *Kernel*, respectively *CART*, the alternative procedure described in (i), respectively in (ii). Concerning our approach based on projected pinball forest, for our experiments on simulated data, we fixed the number of trees to 100 and *mtry* parameter to p, the input space dimension. The γ parameter in Assumption (P2) was set to 0.2, and the π parameter in Assumption (P3) to 0.1. Finally, we set $min\_node\_size = 10$ for all experiments. One advantage of our approach is that results are good without any specific tuning of this last parameter. We also implemented the following additional rule to limit the number of splits: namely, we skip the split if $sd_l/\bar{l} < 0.03$, where $\bar{l}$, resp. sd_l, denotes the empirical mean, resp. standard deviation of $(l_i)_{i=1,\dots,n}$, with l_i the loss function at each sample point. Concerning *Kernel* and *CART*, their implementation is based on a splitting of the original input-output sample $\mathcal{D}$, using the first part (typically two third of the samples) to learn the conditional distribution and the second one (typically the remaining third of the samples) to estimate QOSA index by plugin. To be fair in the comparison, we implemented all the procedures with n samples (n varied from 500 to 2000). We present the comparison for the three examples. Note that in the second and third examples, Assumption 5 is violated. The results we obtain show that our procedure is robust to this assumption.

Example 1: As first example in this section, we consider $Y = X_1 + X_2, X_1, X_2 \sim \mathcal{U}(0,1)$, independent. The analytical values of first-order QOSA indices are:

$$S_1^\alpha = S_2^\alpha = \begin{cases} 1 - \frac{\alpha/2 - \alpha^2/2}{\alpha - (\sqrt{2\alpha})^3/3} & \text{if} \quad \alpha \geq 1/2 \\ 1 - \frac{\alpha/2 - \alpha^2/2}{\alpha - t_\alpha^2 + t_\alpha^3/3 + 1/3} & \text{if} \quad \alpha < 1/2 \end{cases}$$

with $t_\alpha = 2 - \sqrt{2(1-\alpha)}$. In the experiments, we add a dummy variable X_3, that is Y is independent on X_3 which does not appear in the model, and test the ability of each method to detect it. We compute the root mean

squared error (RMSE) with respect to i-th feature as:

$$RMSE_i^\alpha = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{S}_i^{\alpha,m} - S_i^\alpha)^2},$$

from $M = 100$ repetitions.

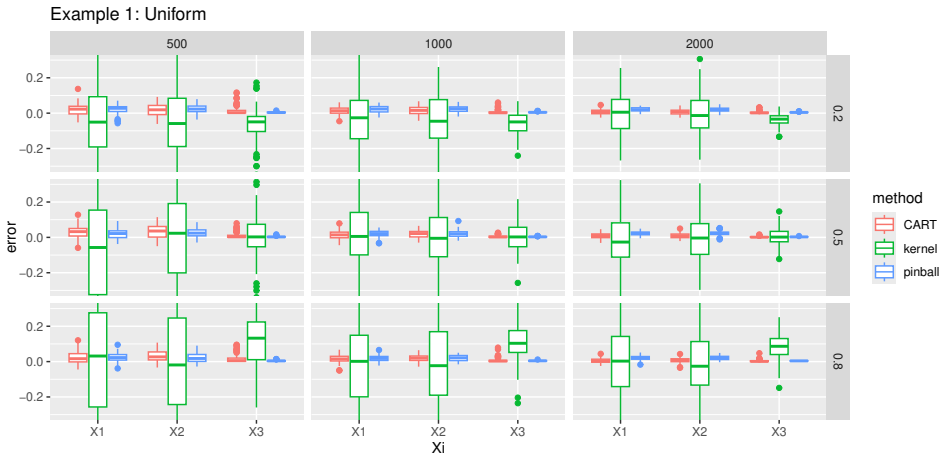

Figure 3: Estimation error for Example 1, at different quantile levels (0.2, 0.5, 0.8 and for different sample size (500, 1000, 2000).

For this example with bounded inputs, the *Kernel* procedure has poor performance. In particular, it exhibits a large variance. It is well-known that kernel-based estimation procedures are prone to boundary issues. For this example, both *CART* and *pinball* show good performances. Note that the main advantage of *pinball* over *CART* is that only one forest has to be fitted and is then projected to estimate each first-order Sobol' index.

		CART			Kernel			pinball		
α	n	X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
0.2	truth	0.308	0.308	0	0.308	0.308	0	0.308	0.308	0
	500	0.038	0.038	0.027	0.234	0.234	0.115	**0.033**	**0.034**	**0.005**
	1000	**0.024**	**0.028**	0.013	0.177	0.168	0.088	0.029	0.03	**0.005**
	2000	**0.016**	**0.017**	0.007	0.111	0.117	0.05	0.024	0.024	**0.005**
0.5	truth	0.25	0.25	0	0.25	0.25	0	0.25	0.25	0
	500	0.045	0.048	0.019	0.469	0.382	0.178	**0.032**	**0.035**	**0.004**
	1000	**0.028**	0.03	0.008	0.199	0.235	0.078	**0.028**	**0.027**	**0.004**
	2000	**0.017**	**0.018**	0.004	0.14	0.15	0.047	0.025	0.026	**0.003**
0.8	truth	0.308	0.308	0	0.308	0.308	0	0.308	0.308	0
	500	0.041	0.043	0.028	1.72	2.623	0.487	**0.033**	**0.033**	**0.005**
	1000	0.027	0.027	0.014	0.451	0.411	0.212	**0.025**	**0.026**	**0.005**
	2000	**0.014**	**0.016**	0.007	0.286	0.219	0.11	0.024	0.025	**0.005**

Table 1: RMSE for Example 1

Example 2: This is one of the examples in Fort et al. (2016). $Y = X_1 - X_2, X_i, i = 1, 2, 3$, are independent and $X_i \sim Exp(1)$. The analytical value of first-order QOSA indices for this model is:

$$S_1^\alpha = \begin{cases} \frac{(1-\alpha)(1-\log(2(1-\alpha)))+\alpha\log(\alpha)}{(1-\alpha)(1-\log(2(1-\alpha)))} & \text{if} \quad \alpha \geq 1/2 \\ \frac{\alpha(1-\log(2\alpha))+\alpha\log(\alpha)}{\alpha(1-\log(2\alpha))} & \text{if} \quad \alpha < 1/2 \end{cases},$$

$$S_2^\alpha = \begin{cases} \frac{(1-\alpha)(1-\log(2(1-\alpha)))+(1-\alpha)\log(1-\alpha)}{(1-\alpha)(1-\log(2(1-\alpha)))} & \text{if} \quad \alpha \geq 1/2 \\ \frac{\alpha(1-\log(2\alpha))+(1-\alpha)\log(1-\alpha)}{\alpha(1-\log(2\alpha))} & \text{if} \quad \alpha < 1/2 \end{cases},$$

and X_3 is a dummy variable.

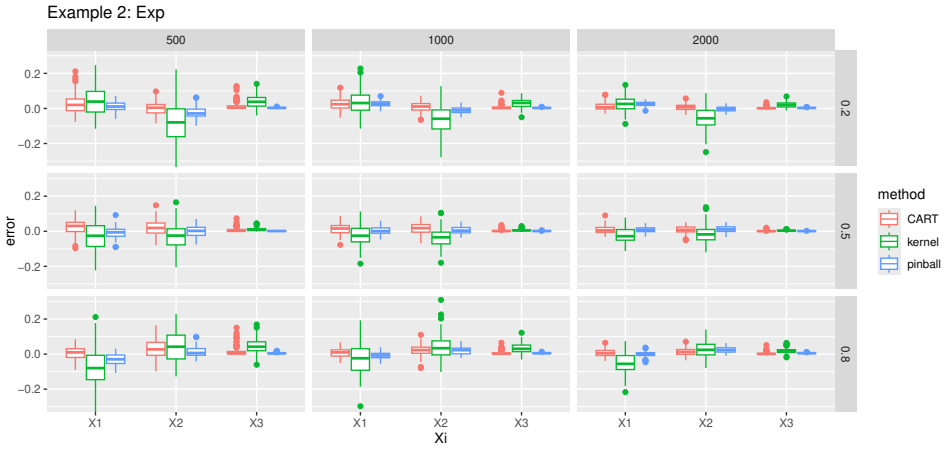

Figure 4: Estimation error for Example 2 at different quantile levels and for different sample size. The first line is for $\alpha = 0.2$, the second line is for $\alpha = 0.5$ and the third line is for $\alpha = 0.8$.

From the boxplots presented in Fig.4, the random forest based methods (*CART* and *pinball*) have smaller variance than *Kernel* method. From Tab. 2, we conclude that for this example our method is the best in most cases, especially to detect the dummy variable X_3.

		CART			Kernel			pinball		
α	n	X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
0.2	truth	0.16	0.534	0	0.16	0.534	0	0.16	0.534	0
	500	0.061	**0.036**	0.027	0.102	0.147	0.055	**0.029**	0.04	**0.004**
	1000	0.042	0.031	0.013	0.078	0.109	0.038	**0.032**	**0.021**	**0.004**
	2000	**0.025**	0.02	0.007	0.051	0.082	0.026	0.029	**0.015**	**0.004**
0.5	truth	0.307	0.307	0	0.307	0.307	0	0.307	0.307	0
	500	0.05	0.049	0.015	0.083	0.074	0.015	**0.032**	**0.03**	**0.002**
	1000	0.033	0.034	0.006	0.063	0.065	0.008	**0.023**	**0.021**	**0.002**
	2000	0.025	0.023	0.004	0.046	0.051	0.005	**0.017**	**0.02**	**0.001**
0.8	truth	0.534	0.16	0	0.534	0.16	0	0.534	0.16	0
	500	**0.036**	0.057	0.033	0.132	0.102	0.059	0.043	**0.03**	**0.006**
	1000	0.027	0.039	0.013	0.088	0.084	0.043	**0.021**	**0.029**	**0.006**
	2000	0.022	**0.024**	0.008	0.08	0.053	0.024	**0.015**	0.029	**0.006**

Table 2: RMSE for Example 2

Example 3: This last example is one of the examples in Elie-Dit-Cosaque and Maume-Deschamps (2022a). $Y = X_1 + X_2$, with (X_1, X_2) a gaussian vector, $X_1 \sim \mathcal{N}(0,1)$, $X_2 \sim \mathcal{N}(0,2)$ and correlation ρ. The analytical values of first-order QOSA indices are:

$$S_1^\alpha = 1 - \frac{\sigma_2\sqrt{1-\rho^2}}{\sigma_Y}\,,\ S_2^\alpha = 1 - \frac{\sigma_1\sqrt{1-\rho^2}}{\sigma_Y}.$$

We also add a dummy variable X_3.

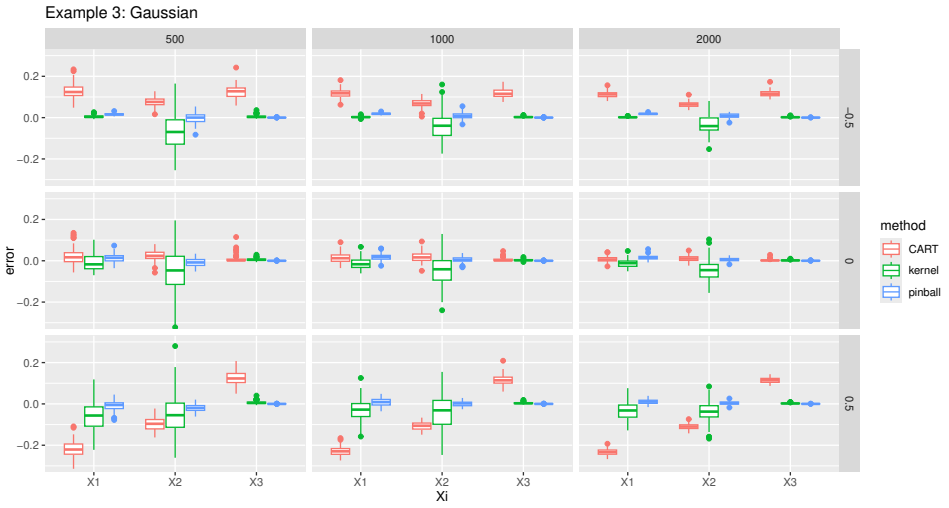

Figure 5: Estimation error for Example 3 for different correlation coefficients, $\rho = -0.5$ (top line), $\rho = 0$ (middle line), $\rho = 0.5$ (bottom line).

From the boxplots in Fig.5, the performance of the *CART* method decreases when dependence between variables increases. Its estimation accuracy for the dummy variable X_3 is poor, in comparison with both *Kernel* and *pinball* methods. The *pinball* approach we introduced reduces significantly the bias in most cases.

		CART			Kernel			pinball		
ρ	n	X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
-0.5	truth	0	0.5	0	0	0.5	0	0	0.5	0
	500	13.255	8.075	12.881	**0.745**	10.939	0.833	1.662	**2.264**	**0.04**
	1000	11.971	7.178	11.919	**0.434**	7.596	0.428	1.91	**1.897**	**0.043**
	2000	11.367	6.445	11.703	**0.259**	5.589	0.28	1.874	**1.351**	**0.032**
0	truth	0.106	0.553	0	0.106	0.553	0	0.106	0.553	0
	500	4.324	3.61	1.722	3.91	10.558	0.814	**2.375**	**2.111**	**0.025**
	1000	2.673	2.963	0.969	2.949	8.41	0.438	**2.244**	**1.417**	**0.028**
	2000	**1.473**	1.744	0.458	2.166	6.757	0.281	1.822	**1.019**	**0.035**
0.5	truth	0.345	0.673	0	0.345	0.673	0	0.345	0.673	0
	500	22.271	10.304	13.071	8.669	11.101	0.956	**2.646**	**2.696**	**0.022**
	1000	23.038	10.948	11.876	5.697	8.925	0.532	**2.049**	**1.16**	**0.025**
	2000	23.346	11.026	11.523	5.177	6.141	0.328	**1.525**	**0.829**	**0.027**

Table 3: RMSE for Example 3, scaled by 10^{-2}

5.2 Higher order QOSA index estimation

In this section, we aim to validate the estimation of Higher order QOSA indices defined in Section 2.2. Though Kernel and CART methods could be extended for this task, they are not implemented in respective packages. So we focus on results for our pinball projected forest method.

Example 4: In this section, we consider a classical linear model $Y = \boldsymbol{\beta}^\intercal \boldsymbol{X}$, where $\boldsymbol{\beta} = (1,1,1)^\intercal$ and $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (0,0,0)^\intercal$,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}, -1 \leq \rho \leq 1, \sigma_1 = \sigma_2 = 1, \sigma_3 = 2.$$

The analytical value of order two QOSA indices for different sets U are:

$$S^\alpha_{1,2} = 1 - \frac{\sigma_3\sqrt{1-\rho^2}}{\sqrt{\boldsymbol{\beta}^\intercal\boldsymbol{\Sigma}\boldsymbol{\beta}}}, S^\alpha_{1,3} = 1 - \frac{\sigma_2\sqrt{1-\rho^2}}{\sqrt{\boldsymbol{\beta}^\intercal\boldsymbol{\Sigma}\boldsymbol{\beta}}}, S^\alpha_{2,3} = 1 - \frac{\sigma_1}{\sqrt{\boldsymbol{\beta}^\intercal\boldsymbol{\Sigma}\boldsymbol{\beta}}}.$$

From Tab. 4, the error decreases as n increases, as expected.

n	ρ	$\{X_1, X_2\}$	$\{X_1, X_3\}$	$\{X_2, X_3\}$
	-0.5	0.027	0.047	0.068
500	0	0.022	0.053	0.053
	0.5	0.057	0.066	0.039
	-0.5	0.022	0.021	0.031
1000	0	0.015	0.026	0.024
	0.5	0.027	0.033	0.017
	-0.5	0.022	0.011	0.015
2000	0	0.017	0.011	0.011
	0.5	0.015	0.016	0.007

Table 4: RMSE for Example 4

5.3 Application to a real dataset

In this section, we present the results obtained by applying our methodology to analyse MOCAGE data. This dataset was proposed and studied in Besse et al. (2007). The description is shown in Tab. 5. It contains 1041 observations with 10 variables (2 categorical and 8 continuous). The target is to predict O3obs (Observed ozone concentration) with the other 9 variables. The summary distribution of all continuous variables can be seen in Fig. 6.

For the implementation, we used 200 trees with $mtry = 4$, $\gamma = 0.2$ and $min\_node\_size = 10$. Also, there are 2 categorical variables in this dataset. The usual way to handle categorical variables is to transform them to continuous variables or transform them to dummy binary variables. We follow the dummy binary solution but with a modification. The STATION variable has 5 modalities, corresponding to 5 geographical sites, it is recoded into 5 binary dummy variables. We bind them together is the sense that if one of these 5 variables is selected in $\mathcal{M}try$, we add the left 4 other dummy variables into $\mathcal{M}try$.

The normalised QOSA index estimation at different quantile levels is shown in Fig. 7. In Besse et al. (2007); Broto et al. (2020), the variables selected as most influential are MOCAGE and TEMPE, then STATION and NO2. These two works focus on the impact of the different variables around the mean of O3obs (with GLM resp. Shapley values), while we are interested in the impact around quantiles at different levels. From Fig. 7, MOCAGE and TEMPE also are the most important variables, but the ranking is different depending on quantile levels. Also, RMH2O is the third important variable for high quantile levels (0.9, 0.8, 0.7) and STATION is more important at lowest quantile levels. This is consistent with the results in Elie-Dit-Cosaque and Maume-Deschamps (2024). This example shows that QOSA indices give different informations than variance based sensitivity analysis.

Table 5: Summary of MOCAGE data

Variable name	Type	Summary
O3obs	Continous	Observed ozone concentration (Response)
JOUR	Binary	holiday = 0, non-holiday = 1 (holiday: 724, non-holiday: 317)
MOCAGE	Continuous	Ozone concentration predicted by a fluid mechanics model[a]
TEMPE	Continuous	Officially predicted temperatures
RMH2O	Continuous	Humidity ratio
NO2	Continuous	Nitrogen dioxide concentration
NO	Continuous	Nitric oxide concentration
STATION	Categorical	5 different sites (Aix:199, Als:222, Cad:202, Pla:208, Ram:210)[b]
VentMOD	Continuous	Wind force
VentANG	Continuous	Wind direction

[a]Large Scale Atmospherical Chemestrial Model: MOCAGE (Modèle de Chimie Atmosphérique à Grande Echelle)

[b]Aix=Aix-en-Provence, Ram=Rambouillet, Als=Munchhausen, Cad=Cadarache, Pla=Plan-de-Cuques

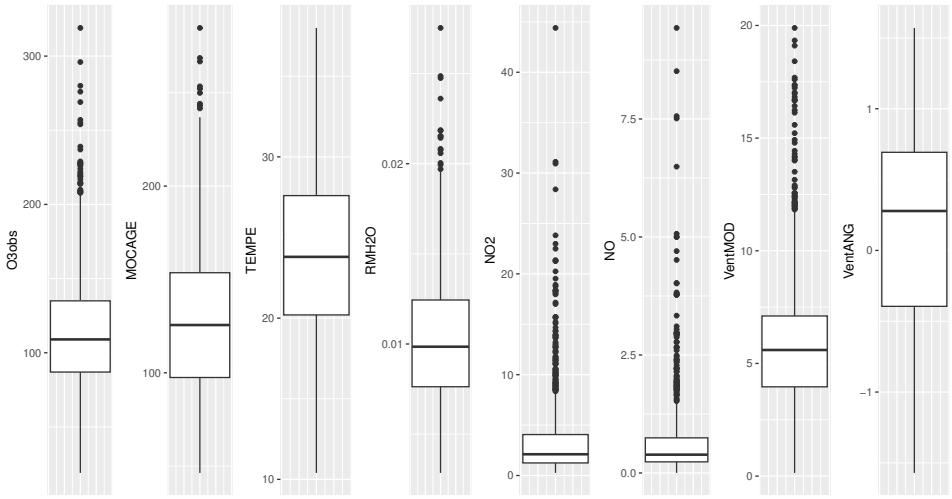

Figure 6: Summary of MOCAGE data

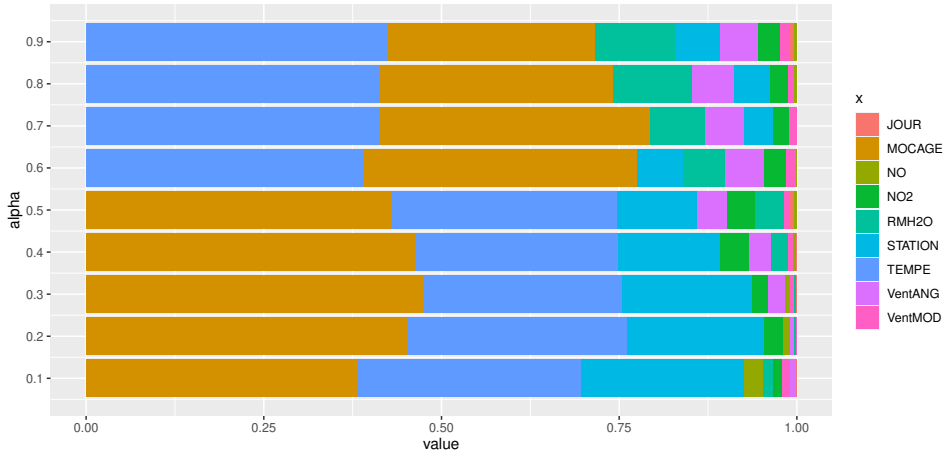

Figure 7: First-order QOSA indices at different quantile levels

References

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Bénard, C., Biau, G., Da Veiga, S., and Scornet, E. (2022a). Shaff: Fast and consistent shapley effect estimates via random forests. In *International Conference on Artificial Intelligence and Statistics*, pages 5563–5582. PMLR.

Bénard, C., Da Veiga, S., and Scornet, E. (2022b). Mean decrease accuracy for random forests: inconsistency, and a practical solution via the sobol-mda. *Biometrika*, 109(4):881–900.

Besse, P., Milhem, H., Mestre, O., Dufour, A., and Peuch, V.-H. (2007). Comparaison de techniques de « data mining » pour l'adaptation statistique des prévisions d'ozone du modèle de chimie-transport mocage. *Pollution atmosphérique*.

Bhat, H. S., Kumar, N., and Vaz, G. J. (2015). Towards scalable quantile regression trees. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 53–60. IEEE.

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

Breiman, L., Friedman, J., Olshen, R.A., and Stone, C. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC.

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

Browne, T., Fort, J.-C., Iooss, B., and Le Gratiet, L. (2017). Estimate of quantile-oriented sensitivity indices.

Ćevid, D., Michel, L., Näf, J., Meinshausen, N., and Bühlmann, P. (2022). Distributional random forests: Heterogeneity adjustment and multivariate distributional regression. *Journal of Machine Learning Research*, 23(333):1–79.

Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021). *Basics and trends in sensitivity analysis: Theory and practice in R*. SIAM.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer.

Elie-Dit-Cosaque, K. and Maume-Deschamps, V. (2022a). Goal-oriented shapley effects with special attention to the quantile-oriented case. *SIAM/ASA Journal on Uncertainty Quantification*, 10(3):1037–1069.

Elie-Dit-Cosaque, K. and Maume-Deschamps, V. (2022b). Random forest estimation of conditional distribution functions and conditional quantiles. *Electronic Journal of Statistics*, 16(2):6553–6583.

Elie-Dit-Cosaque, K. and Maume-Deschamps, V. (2024). Random forest based quantile-oriented sensitivity analysis indices estimation. *Computational Statistics*, pages 1–31.

Fort, J.-C., Klein, T., and Rachdi, N. (2016). New sensitivity analysis subordinated to a contrast. *Communications in Statistics-Theory and Methods*, 45(15):4349–4364.

Iooss, B. and Lemaître, P. (2015). A review on global sensitivity analysis methods. In *Uncertainty management in simulation-optimization of complex systems*, pages 101–122. Springer.

Kala, Z. (2019). Quantile-oriented global sensitivity analysis of design resistance. *Journal of Civil Engineering and Management*, 25(4):297–305.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Maume-Deschamps, V. and Niang, I. (2018). Estimation of quantile oriented sensitivity indices. *Statistics & Probability Letters*, 134:122–127.

Meinshausen, N. and Ridgeway, G. (2006). Quantile regression forests. *Journal of machine learning research*, 7(6).

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., et al. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*, volume 1. Wiley Online Library.

Sobol', I. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.*, 1:407.

Wand, M. P., Jones, M. C., et al. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, 9(2):97–116.

A Technical results

This appendix contains some technical results that are needed for the proofs.

Lemma A.1. *Assume the forest construction satisfies properties (P1-3) and that Assumption 1 is satisfied, then for any $U \subseteq \{1, \dots, p\}$, any $j \in U$, the diameter of leaves goes to 0 almost surely:*

$$\forall \mathbf{x} \in \mathcal{X}_U,\ \operatorname{diam}_j(A(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U)) \xrightarrow[n\to\infty]{a.s.} 0,$$

where $\operatorname{diam}_j(A) = \sup_{\boldsymbol{z}, \boldsymbol{z}' \in A} |\boldsymbol{z}_j - \boldsymbol{z}'_j|$.

Proof. The proof follows partially the proof of Lemma 2 in Meinshausen and Ridgeway (2006) (see also Lemma 5 in Bénard et al. (2022b)). Let $j \in \{1, \dots, p\}$. We denote by $(C_k)_k$ the set of nodes from the root node to the leaf node containing $\boldsymbol{x}$, and by $S(\boldsymbol{x}, j)$ the number of such nodes for which the splitting is done along the j-th feature. We define the following event:

$$A = \{\exists \text{ infinitely many indices } k \text{ in } S(\boldsymbol{x}, j) \text{ for which, mtry } = 1 \text{ and } \Gamma = j\},$$

where Γ (or $\Gamma(C_k)$ if we need to emphasize the dependency on C_k) is the random variable from (P3), uniform on $\{1, 2, ..., p\}$ when mtry equals 1. It is easily seen that

$$A \subseteq \{S(\boldsymbol{x}, j) \to +\infty\}. \tag{A.1}$$

Let $\text{mtry}(C_k)$ denote the value of mtry on the cell C_k. We have:

$$A = \{\forall\, N, \exists k \geq N \text{ with } \text{mtry}(C_k) = 1 \text{ and } \Gamma(C_k) = j\},$$

that is

$$A = \bigcap_N \bigcup_{k \geq N} \{\text{mtry}(C_k) = 1, \Gamma(C_k) = j\}.$$

Thus, by the Kolmogorov's zero–one law, event A has probability 0 or 1. For any fixed N and $k_0 \geq N$, we have from (P3):

$$\mathbb{P}(\{\text{mtry}(C_{k_0}) = 1, \Gamma(C_{k_0}) = j\}) = \pi/p,$$

which implies $\mathbb{P}(A) \geq \pi/p$ and thus $\mathbb{P}(A) = 1$. Together with (A.1), it yields

$$S(\boldsymbol{x}, j) \xrightarrow[h \to \infty]{a.s.} \infty.$$

With the notation

$$A(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U)) = \prod_{j=1}^{p} A^{(j)}(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U))$$

and using (P1) and (P2), we get:

$$N^{(j)}(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U)) \leq s_n (1 - \gamma)^{S(\boldsymbol{x}, j)}$$

with $N^{(m)}(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U))$ the number of observations whose j-th coordinate belongs to $A^{(j)}(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U))$. Then, from Assumption 1 and following the end of the proof of (Bénard et al., 2022b, Lemma 5), we get the result, that is $\text{diam}_j(A(\boldsymbol{x}; \mathcal{D}_n^*(\Theta))) \xrightarrow[n \to +\infty]{a.s.} 0$. □

Then Corollary A.2 follows straightforwardly.

Corollary A.2. *Assume that Assumptions of Lemma A.1 are verified and that for any $y \in \mathbb{R}$, $F_Y(y|\boldsymbol{X_U} = \cdot)$ is continuous, then for each tree $b = 1, 2, ..., B$, the variation of the conditional cumulative distribution function in leaf goes to 0:*

$$\forall \boldsymbol{x} \in \mathcal{X}_U, \forall y \in \mathbb{R}, \sup_{\boldsymbol{z} \in A(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), U)} |F(y|\boldsymbol{X}_U = \boldsymbol{z}) - F(y|\boldsymbol{X}_U = \boldsymbol{x})| \xrightarrow[n\to\infty]{a.s.} 0.$$

The number of elements of $\mathcal{D}_n^\star(\Theta_b)$, *resp.* $\mathcal{D}_n^\diamond(\Theta_b)$, in a leaf are compared in the following Lemma A.3 below.

Lemma A.3. *For any $\varepsilon > 0$, we have*

$$\mathbb{P}(|N(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), \mathcal{D}_{s_n}^\diamond, U) - N(\boldsymbol{x}; \mathcal{D}_n^*(\Theta), \mathcal{D}_n^*(\Theta), U))| > \varepsilon) \leq 24(s_n + 1)^{2p} e^{-\varepsilon^2/(188 s_n)}.$$

Proof. The proof is similar to the one of Lemma 6.3 in Elie-Dit-Cosaque and Maume-Deschamps (2022b), where we use subsampling instead of bootstrap. □

We finish this appendix by sketching the proof of Lemma 3.4 which is used to obtain the consistency of the OOB estimator.

Lemma A.4 (Lemma 4 in Bénard et al. (2022b))**.** *Assume the forest construction satisfies property (P1). Consider $\delta_{B,n}$ and $\gamma_{B,n}$ as follows:*

$$\delta_{B,n} = B^2 \mathbb{E}\left[\frac{1}{|\Lambda_{n,i}|^2} \middle| 1, 2 \in \Lambda_{n,i}\right] \mathbb{P}(1, 2 \in \Lambda_{n,i}),$$

$$\gamma_{B,n} = B^2 \mathbb{E}\left[\frac{1}{|\Lambda_{n,i}|^2} \middle| 1 \in \Lambda_{n,i}\right] \mathbb{P}(1 \in \Lambda_{n,i}).$$

Then, for all $B \in \mathbb{N} \setminus \{0, 1\}$, we have

$$\delta_{B,n} \leq 1, \delta_{B,n} \leq \gamma_{B,n} \leq \frac{2}{1 - s_n/n},$$

and for a fixed sample size n, $1 - \delta_{B,n} = O(B^{-1})$.

Lemma 3.4. *Assume that the forest construction properties (P1-3) and Assumptions 1 to 4 are verified, then for all $B \in \mathbb{N}^\star, j \in \{1, 2, ..., n\}$ and $U \subseteq \{1, 2, ..., p\}$, we have*

$$\sup_y \mathbb{E}[(\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2]$$

$$\leq \tfrac{2}{1 - s_n/n} \sup_y \mathbb{E}[(\hat{F}_{B,s_n,n-1}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2] + O((s_n/n)^B).$$

Proof. We follow the proof of Lemma 2 in Bénard et al. (2022b). Consider $\boldsymbol{x} \in \mathcal{X}_U$. Recall the definition of standard estimator:

$$w_i(\boldsymbol{x}|U) = \frac{1}{B}\sum_{b=1}^{B} \frac{\mathbb{I}_{\{\boldsymbol{X}^i \in A_n(\boldsymbol{x}; \mathcal{D}_n^*(b), \mathcal{D}_n^*(b), U)\}}}{N(\boldsymbol{x}; \mathcal{D}_n^*(b), \mathcal{D}_n^*(b), U)}, \tag{A.2}$$

$$\hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}) = \sum_{i=1}^{n} w_i(\boldsymbol{x}|U)\mathbb{I}_{(Y^i \leq y)}.$$

Now we define $\hat{F}_{s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}; \Theta_l)$ from the following equation:

$$\hat{F}_{B,s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}) = \frac{1}{B}\sum_{l=1}^{B} \hat{F}_{s_n,n}(y|\boldsymbol{X}_U = \boldsymbol{x}; \Theta_l).$$

Then, for fixed y,

$$\begin{aligned}\mathbb{E}[(\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2] \quad &= \mathbb{P}(|\Lambda_{n,j}| > 0)\mathbb{E}[(\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2 \big| |\Lambda_{n,j}| > 0] \\ &\quad + \mathbb{P}(|\Lambda_{n,j}| = 0)\mathbb{E}[(F(y|\boldsymbol{X}_U^j))^2 \big| |\Lambda_{n,j}| = 0].\end{aligned}$$

As $\mathbb{P}(|\Lambda_{n,j}| = 0) = (s_n/n)^B$, we have

$$\mathbb{P}(|\Lambda_{n,j}| = 0)\mathbb{E}[(F(y|\boldsymbol{X}_U^j))^2 | |\Lambda_{n,j}| = 0] \leq \left(\frac{s_n}{n}\right)^B$$

Moreover, with the notation in Lemma A.4,

$$\begin{aligned}&\mathbb{E}[(\hat{F}^{OOB}(y|\boldsymbol{X}_U^j) - F(y|\boldsymbol{X}_U^j))^2 \big| |\Lambda_{n,j}| > 0]\mathbb{P}(|\Lambda_{n,j}| > 0) \\ =&\delta_{B,n}\mathbb{E}\left[\frac{1}{B^2}\sum_{l,l'=1}^{B}\left(\hat{F}_{s_n,n-1}(y|\boldsymbol{X}_U^j;\Theta_l) - F(y|\boldsymbol{X}_U^j)\right)\left(\hat{F}_{s_n,n-1}(y|\boldsymbol{X}_U^j;\Theta_{l'}) - F(y|\boldsymbol{X}_U^j)\right)\right] \\ &+ (\gamma_{B,n} - \delta_{B,n})\mathbb{E}\left[\frac{1}{B^2}\sum_{l=1}^{B}\left(\hat{F}_{s_n,n-1}(y|\boldsymbol{X}_U^j;\Theta_l) - F(y|\boldsymbol{X}_U^j)\right)^2\right].\end{aligned}$$

The rest of the proof follows line by line the proof of Lemma 2 in Bénard et al. (2022b). □