Shapley effect estimation in reliability-oriented sensitivity analysis with correlated inputs by importance sampling

Julien Demange-Chryst, François Bachoc, Jérôme Morio

International Journal for Uncertainty Quantification, x(x): 1–37 (2022)

SHAPLEY EFFECT ESTIMATION IN RELIABILITY-ORIENTED SENSITIVITY ANALYSIS WITH CORRELATED INPUTS BY IMPORTANCE SAMPLING

Julien Demange-Chryst,[1,2,*] François Bachoc,[1] & Jérôme Morio[2]

[1]*Institut de Mathématiques de Toulouse, UMR5219 CNRS, 31062 Toulouse, France*
[2]*ONERA/DTIS, Université de Toulouse, F-31055 Toulouse, France*

*Address all correspondence to: Julien Demange-Chryst, E-mail: julien.demange-chryst@onera.fr

Original Manuscript Submitted: 03/17/2022; Final Draft Received: mm/dd/yyyy

Reliability-oriented sensitivity analysis aims at combining both reliability and sensitivity analyses by quantifying the influence of each input variable of a numerical model on a quantity of interest related to its failure. In particular, target sensitivity analysis focuses on the occurrence of the failure, and more precisely aims to determine which inputs are more likely to lead to the failure of the system. The Shapley effects are quantitative global sensitivity indices which are able to deal with correlated input variables. They have been recently adapted to the target sensitivity analysis framework. In this article, we investigate two importance-sampling-based estimation schemes of these indices which are more efficient than the existing ones when the failure probability is small. Moreover, an extension to the case where only an i.i.d. input/output N-sample distributed according to the importance sampling auxiliary distribution is proposed. This extension allows to estimate the Shapley effects only with a data set distributed according to the importance sampling auxiliary distribution stemming from a reliability analysis without additional calls to the numerical model. In addition, we study theoretically the absence of bias of some estimators as well as the benefit of importance sampling. We also provide numerical guidelines and finally, realistic test cases show the practical interest of the proposed methods.

KEY WORDS: *Reliability-oriented sensitivity analysis, Target sensitivity analysis, Rare event estimation, Dependent inputs, Shapley effects, Importance sampling, Nearest-neighbour approximation.*

1. INTRODUCTION

More and more physical phenomenons and complex systems are numerically represented by black-box numerical models, which are often computationally expensive to evaluate, and whose complexity makes it impossible to study analytically. For safety and certification purposes, tracking the potential failures of a system is crucial, but it is not an option to do so experimentally with critical systems because it could lead to dramatic environmental, human or financial consequences. Numerical models enable to simulate the behaviour of a system far from nominal configurations.

The *reliability analysis* of a numerical model mainly consists in the estimation of its failure probability. The failure is often a rare event and thus has a small probability. The high computational cost of one evaluation of the numerical model (several minutes to several days CPU) and the low value of the probability make the usual quadrature methods [1] and Monte Carlo sampling [2] inappropriate to handle this problem, but various techniques reviewed in [3] have been developed to estimate more precisely a such probability at a limited computational cost, including importance sampling [4] for example.

Global sensitivity analysis (GSA) aims at studying the impact of the input variables of a numerical code on the behaviour of its output to provide a better understanding of the model. It can be carried out for various purposes such as fixing non influential input variables to nominal values or identifying the most influential ones to decrease

2152–5080/22/$35.00

their variability, see for example [5]. A deeper analysis of the failure of the system consists then in combining both reliability and sensitivity analyses by performing a GSA on a quantity of interest (QoI) characterizing the failure of the system. This new specific framework is called *reliability-oriented sensitivity analysis* (ROSA) and can be divided into two categories [6]:

- *target sensitivity analysis* (TSA) aims at determining the influence of each input variable on the occurrence of the failure of the system
- *conditional sensitivity analysis* (CSA) aims at performing a global sensitivity analysis of the numerical code restricted to the failure domain.

In the present article, we only focus on TSA. The well-known Sobol indices for variance-based GSA [7] have been adapted to TSA and first specific estimation schemes have been introduced in [8,9]. However, as in GSA, the interpretability of these indices requires the strong assumption of independent input variables. Several approaches have been investigated in GSA to adapt the Sobol indices to the case where the inputs are correlated [10] but recently, new GSA indices based on game theory [11] and which are more naturally able to deal with correlated inputs have been introduced: the *Shapley effects* [12]. Their adaptation to the TSA framework is recent [13] and the authors proposed first estimation schemes based on a classical Monte Carlo sampling according to the input distribution.

As illustrated in our numerical simulations in Section 4, the existing estimators of the Shapley effects for TSA based on a Monte Carlo sampling from [13] are not efficient when the failure probability is small because they require too many calls to the numerical code to be accurate. In this article, we introduce then new importance-sampling-based estimators of these indices which are able to deal more efficiently with a small failure probability. Moreover, we extend these new estimators to the case where only an i.i.d. input/output N-sample distributed according to the importance sampling auxiliary distribution is available, using the nearest neighbour approximation described in [14]. A major practical advantage is that our extended estimators enable to estimate efficiently the Shapley effects for TSA without additional calls to the function after the estimation of the failure probability by importance sampling. In addition, under the condition that the reliability analysis has been done efficiently, we show theoretically that the proposed estimators improve the estimation of the Shapley effects for TSA compared to the existing ones when the failure probability is getting smaller and finally, we give some numerical guidelines.

The remainder of this paper is organized as follows. First, Section 2 consists in a review on variance-based global sensitivity analysis, importance sampling and reliability-oriented sensitivity analysis. Then, Section 3 introduces and describes the proposed importance-sampling-based estimators of the Shapley effects for TSA. In addition, Section 4 illustrates the practical interest of the new estimators on numerical examples: the Gaussian linear case, a cantilever beam problem and a fire spread model. Finally, Section 5 concludes the present article and gives future research perspectives stemming from this work.

2. A REVIEW ON GLOBAL SENSITIVITY INDICES: DEFINITIONS, ESTIMATION SCHEMES AND ADAPTATION TO RELIABILITY

In this section, we recall the main principle of variance-based GSA and we describe very common existing sensitivity indices as well as some of their estimation schemes proposed in the literature. Next, after a brief reminder of importance sampling, we also review some tools from ROSA.

First of all, let us begin by introducing the notations that will be used throughout the paper. We let $\mathbf{X} = (X_1, \ldots, X_d)$ be the input random vector on the input domain $\mathbb{X} = \bigotimes_{i=1}^{d} \mathbb{X}_i \subseteq \mathbb{R}^d$ with joint PDF $f_{\mathbf{X}}$. Then, the black-box function is defined by:

$$\phi : \left| \begin{array}{ccc} \mathbb{X} & \longrightarrow & \mathbb{R} \\ \mathbf{x} & \longmapsto & y = \phi(\mathbf{x}). \end{array} \right. \tag{1}$$

No regularity hypothesis on ϕ is required but the random output $Y = \phi(\mathbf{X}) \in \mathbb{R}$ is supposed to be square integrable, i.e. $\mathbb{E}\left(Y^2\right) < +\infty$. Moreover, we let $\mathcal{P}(d) = \{u \subseteq [\![1, d]\!]\}$ denote all the subsets of $[\![1, d]\!] = \{1, \ldots, d\}$. Then, for any $u \in \mathcal{P}(d)$, let us write $-u = [\![1, d]\!] \backslash u$ for the complementary of the set u. In particular, for all $i \in [\![1, d]\!]$, $-i$ refers to the subset $[\![1, d]\!] \backslash \{i\}$. In addition, for any non-empty subset $u \in \mathcal{P}(d)$, letting $u = \{i_1, \ldots, i_r\}$ with $i_1 <$

$\cdots < i_r$, let $\mathbb{X}_u = \bigotimes_{j=1}^{r} \mathbb{X}_{i_j} \subseteq \mathbb{R}^r$ be the input domain of the random sub-vector $\mathbf{X}_u = (X_{i_1}, \ldots, X_{i_r}) = (X_i)_{i \in u}$. Furthermore, for $u \in \mathcal{P}(d) \backslash \{\varnothing, [\![1, d]\!]\}$, for any $\mathbf{x}_u \in \mathbb{X}_u$ and for any $\mathbf{x}_{-u} \in \mathbb{X}_{-u}$, $(\mathbf{x}_u, \mathbf{x}_{-u})$ represents the vector $\tilde{\mathbf{x}} \in \mathbb{X}$ such that $\tilde{\mathbf{x}}_u = \mathbf{x}_u$ and $\tilde{\mathbf{x}}_{-u} = \mathbf{x}_{-u}$. We also write $\phi(\mathbf{x}_u, \mathbf{x}_{-u}) = \phi(\tilde{\mathbf{x}})$. Finally, for any probability density $g : \mathbb{X} \longrightarrow \mathbb{R}_+$, we let $\mathbb{E}_g$ and $\mathbb{V}_g$ denote respectively the expectation and the variance operators of a random variable distributed according to the law of PDF g. When there is no ambiguity, we may also write $\mathbb{E}$ and $\mathbb{V}$ for $\mathbb{E}_{f_{\mathbf{X}}}$ and $\mathbb{V}_{f_{\mathbf{X}}}$.

2.1 From Sobol indices to Shapley effects

The Hoeffding functional decomposition [15] allows to represent a function defined on any subset of $\mathbb{R}^d$ as a sum of elementary functions. When considering an input measure with independent components, this decomposition is unique under orthogonality conditions stated by [7]. Then, in the sensitivity analysis framework with independent inputs in $\mathbf{X}$ and a square integrable random output on the form $Y = \phi(\mathbf{X})$, this decomposition leads to a unique functional decomposition of the variance of Y, also called ANOVA (ANalysis Of VAriance):

$$\mathbb{V}(Y) = \mathbb{V}(\phi(\mathbf{X})) = \sum_{u \in \mathcal{P}(d) \backslash \{\varnothing\}} \mathbb{V}(\phi_u(\mathbf{X}_u)), \tag{2}$$

where for $u \in \mathcal{P}(d) \backslash \{\varnothing\}$, $\phi_u(\mathbf{X}_u) = \mathbb{E}(\phi(\mathbf{X})|\mathbf{X}_u) + \sum_{v \subsetneq u} (-1)^{|u|-|v|} \mathbb{E}(\phi(\mathbf{X})|\mathbf{X}_v)$. Then, the well-known *first-order Sobol indices* [7] for GSA are obtained for all $i \in [\![1, d]\!]$ by:

$$S_i = \frac{\mathbb{V}[\mathbb{E}(\phi(\mathbf{X})|X_i)]}{\mathbb{V}(\phi(\mathbf{X}))} \in [0, 1]. \tag{3}$$

The index S_i quantifies the part of variance of the output explained only by the input X_i. From (2), it is also possible to define higher-order Sobol indices which take into account the interactions between the input variables in ϕ but their number increases exponentially with the dimension d and evaluating all of them becomes impossible. Thus, instead of computing higher-order indices, one typically prefers to consider another family of indices, the *total-order Sobol indices* introduced by [16], where each of them quantifies the part of variance of the output explained by the input X_i in interaction with any other group of variables. In practice, when d is large, the first-order and the total-order indices give satisfying information for the sensitivity analysis of the model.

When the inputs are correlated, the Hoeffding decomposition is no longer unique and even if it is still possible to define and compute Sobol indices, they don't allow to clearly identify the origin of the variability of the output anymore. To address this issue, by an analogy between game theory [11] and GSA, the author of [12] introduced new variance-based sensitivity indices which are able to deal with correlated inputs called *Shapley effects* or *Shapley values*, defined for all $i \in [\![1, d]\!]$ by:

$$\mathrm{Sh}_i = \frac{1}{\mathbb{V}(\phi(\mathbf{X}))} \frac{1}{d} \sum_{u \subseteq -i} \binom{d-1}{|u|}^{-1} (c(u \cup \{i\}) - c(u)), \tag{4}$$

with $c : \mathcal{P}(d) \longrightarrow \mathbb{R}$ a cost function which is specific to how input influence is measured. The d input variables are interpreted as players (from the game theory framework from [11]) and the author of [12] proposed to use as cost function the unormalized *closed Sobol indices*, that are defined for all $u \in \mathcal{P}(d)$ by:

$$\mathrm{VE}_u = \mathbb{V}[\mathbb{E}(\phi(\mathbf{X})|\mathbf{X}_u)]. \tag{5}$$

The increment $\left(\mathrm{VE}_{u \cup \{i\}} - \mathrm{VE}_u\right)$ quantifies the individual contribution of the variable X_i to the variance of the output in relation with the group of variables $u \subseteq -i$ taking into account both interaction and dependence. Moreover, the Shapley values using the alternative cost function $\mathrm{EV}_u = \mathbb{E}[\mathbb{V}(\phi(\mathbf{X})|\mathbf{X}_{-u})]$ are equal to the ones using VE_u [17], which thus provides an alternative way to compute them. We call VE_u and EV_u the *conditional indices*. Practical interest and theoretical properties of Shapley values for GSA have been widely studied since their introduction [17–19]. Two important properties allow for an easy interpretation of these values: they are all non negative and sum to one. Thus, they give a quantitative measure, as a percentage, of the influence of each input on the variability of the output taking into account both interaction and dependence between input variables.

Remark 1. *Remark that* $VE_{\varnothing} = EV_{[\![1,d]\!]} = 0$ *and that* $VE_{[\![1,d]\!]} = EV_{\varnothing} = \mathbb{V}(\phi(\mathbf{X}))$. *Thus, during the estimation process of the Shapley effects described in the remaining of the article, it will not be necessary to estimate the conditional indices for* $u \in \{\varnothing, [\![1,d]\!]\}$.

2.2 Shapley effect estimation schemes

Obtaining an accurate estimation of the Shapley effects at a reasonable cost is very challenging and is an active research topic. In the context of game theory, the authors of [20] presented a first algorithm to estimate the Shapley effects which was improved by [17] in sensitivity analysis by reducing the number of calls to the function ϕ. New approaches [14,21] and surrogate-model-based strategies [19,22,23] were explored to reduce even more the estimation cost of these indices while the articles [24,25] were focused on the estimation of the Shapley effects with independent groups of variable.

The estimation schemes of the Shapley effects considered in this paper can be divided into two parts:

1. estimation of the conditional indices VE_u or EV_u for some subsets $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$
2. an aggregation procedure which consists in computing all the Sh_i using the previous estimations of the conditional indices.

In the following, two sampling methods from the literature are presented for the estimation of the conditional indices, with for each of them, an extension to the case where only an i.i.d. sample distributed according to the input distribution and its corresponding output is available. Afterwards, we also present two aggregation procedures.

2.2.1 Estimation of EV_u by double Monte Carlo

In this sub-subsection and the following one, for any $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, assume that:

- we can evaluate the code ϕ in any point of $\mathbb{X}$
- it is possible to generate an i.i.d. sample from the distribution of $\mathbf{X}_u$
- for any $\mathbf{x}_u \in \mathbb{X}_u$, it is possible to generate an i.i.d. sample from the distribution of $\mathbf{X}_{-u}|\mathbf{X}_u = \mathbf{x}_u$.

These assumptions define the *given-model* framework. The two different cost functions VE_u and EV_u provide the same Shapley values, as mentioned above. However, the authors of [26] pointed out a natural double (or two-level) Monte Carlo estimator of VE_u but remarked that it is biased, whereas they suggested a natural double Monte Carlo estimator of EV_u which is unbiased. Hence, the authors of [17] chose to estimate EV_u instead of VE_u and then suggested the following double Monte Carlo estimator:

$$\widehat{\mathrm{EV}}_{u,\mathrm{MC}} = \frac{1}{N_u}\sum_{n=1}^{N_u}\frac{1}{N_I-1}\sum_{i=1}^{N_I}\left(\phi(\mathbf{X}_u^{(n,i)},\mathbf{X}_{-u}^{(n)}) - \overline{\phi(\mathbf{X}_{-u}^{(n)})}\right)^2, \tag{6}$$

where $(\mathbf{X}_{-u}^{(n)})_{n\in[\![1,N_u]\!]}$ is an i.i.d. sample from the distribution of $\mathbf{X}_{-u}$, where for all $n \in [\![1,N_u]\!]$, $(\mathbf{X}_u^{(n,i)})_{i\in[\![1,N_I]\!]}$ is an i.i.d. sample from the distribution of $\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}$ and where $\overline{\phi(\mathbf{X}_{-u}^{(n)})} = N_I^{-1}\sum_{j=1}^{N_I}\phi(\mathbf{X}_u^{(n,j)},\mathbf{X}_{-u}^{(n)})$. This estimator is composed of an inner loop of size N_I for the conditional variance, and of an outer loop of size N_u (which depends on u) for the expectation. It requires N_uN_I calls to ϕ and it is unbiased.

2.2.2 Estimation of VE_u by Pick-Freeze

The basics of the *Pick-Freeze* method were introduced in [7,16]. When the components of $\mathbf{X}$ are independent, it is possible to remove the expensive double loop in (6) by rewriting the conditional indices VE_u as a single expectation, with the interpretation of picking and freezing some input variables [7]. Recently, the Pick-Freeze method was generalized in [14] to the case where the inputs are correlated. The idea is to introduce a second random variable $\mathbf{X}^u = \left(\mathbf{X}_u, \mathbf{X}'_{-u}\right)$ with $\mathbf{X}'_{-u} \overset{d}{=} \mathbf{X}_{-u}|\mathbf{X}_u$ and $\mathbf{X}'_{-u} \perp\!\!\!\perp \mathbf{X}_{-u}|\mathbf{X}_u$, where $\perp\!\!\!\perp$ is the independence symbol, and to write:

$$\mathrm{VE}_u = \mathbb{V}\left[\mathbb{E}\left(\phi(\mathbf{X})|\mathbf{X}_u\right)\right] = \mathbb{E}\left[\phi(\mathbf{X})\phi(\mathbf{X}^u)\right] - \mathbb{E}\left[\phi(\mathbf{X})\right]^2. \tag{7}$$

The random variables $\mathbf{X}$ and $\mathbf{X}^u$ are correlated and have the same distribution. To obtain $\mathbf{X}^u$ from $\mathbf{X}$, the component according to u is frozen and the component according to $-u$ is chosen independently conditionally to $\mathbf{X}_u$. In order to estimate the conditional index based on (7), let first $\widehat{E}_{\phi,N}$ be the natural Monte Carlo estimator of $\mathbb{E}\left[\phi(\mathbf{X})\right]$ with a sample of size N from the distribution of $\mathbf{X}$. The following estimator of VE_u was then suggested by [14]:

$$\widehat{\mathrm{VE}}_{u,\mathrm{PF}} = \frac{1}{N_u}\sum_{n=1}^{N_u} \phi(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(n,1)})\phi(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(n,2)}) - \left(\widehat{E}_{\phi,N}\right)^2, \tag{8}$$

where $(\mathbf{X}_u^{(n)})_{n\in[\![1,N_u]\!]}$ is an i.i.d. sample from the distribution of $\mathbf{X}_u$ and where for all $n \in [\![1, N_u]\!]$, $(\mathbf{X}_{-u}^{(n,i)})_{i\in[\![1,2]\!]}$ are two independent random variables from the distribution of $\mathbf{X}_{-u}|\mathbf{X}_u = \mathbf{X}_u^{(n)}$. The inner loop of size N_I of the Monte Carlo estimator (6) is replaced by the product $\phi(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(n,1)})\phi(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(n,2)})$. This estimator requires $2N_u$ calls to ϕ and is unbiased.

2.2.3 Extension when only an i.i.d. sample is available

In this section, assume that the code ϕ is no longer available and that only an i.i.d. sample $\left(\mathbf{X}^{(n)}, \phi\left(\mathbf{X}^{(n)}\right)\right)_{n\in[\![1,N]\!]}$ with $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ from the distribution of $\mathbf{X}$ is available. This is the *given-data* framework as defined in [14]. The estimation of sensitivity indices in this framework was first explored in [27] but in restrictive cases, for example when $|u| = 1$.

The authors of [14] extended the previous estimators (6) and (8) to the given-data framework. The difficult point is that exact sampling from the conditional distributions $\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}$ for some $\mathbf{x}_{-u} \in \mathbb{X}_{-u}$ is no longer possible. The *nearest-neighbours approximation*, which is fully described in [14], allows to approximate these distributions with the available i.i.d. sample. To that end, for $v \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$ and $(l,i) \in [\![1,N]\!]^2$, let us write $k_N^v(l,i) \in [\![1,N]\!]$ for the index of the i-th nearest neighbour of the point $\mathbf{X}_v^{(l)}$ in the subspace $\mathbb{X}_v$ (according to the Euclidean distance) among $\left(\mathbf{X}_v^{(n)}\right)_{n\in[\![1,N]\!]}$. Moreover, $(k_N^v(l,i))_{i\in[\![1,N]\!]}$ are defined to be two by two distinct: if several points are at equal distance from $\mathbf{X}_v^{(l)}$ for some $l \in [\![1,N]\!]$, ties are broken arbitrarily.

Finally, the extended estimators of the conditional indices are given by:

$$\widehat{\mathrm{EV}}_{u,\mathrm{MC}}^{\mathrm{KNN}} = \frac{1}{N_u}\sum_{n=1}^{N_u}\frac{1}{N_I-1}\sum_{i=1}^{N_I}\left[\phi\left(\mathbf{X}^{\left(k_N^{-u}(s(n),i)\right)}\right) - \frac{1}{N_I}\sum_{j=1}^{N_I}\phi\left(\mathbf{X}^{\left(k_N^{-u}(s(n),j)\right)}\right)\right]^2, \tag{9}$$

and

$$\widehat{\mathrm{VE}}_{u,\mathrm{PF}}^{\mathrm{KNN}} = \frac{1}{N_u}\sum_{n=1}^{N_u}\phi\left(\mathbf{X}^{(k_N^u(s(n),1))}\right)\phi\left(\mathbf{X}^{(k_N^u(s(n),2))}\right) - \left(\widehat{E}_{\phi,N}\right)^2, \tag{10}$$

with $(s(n))_{n\in[\![1,N_u]\!]}$ a sample of uniformly distributed integers in $[\![1,N]\!]$. These estimators require no more calls to ϕ than those used to obtain the i.i.d. sample. The most costly step is the search of the nearest neighbours and under some assumptions, those given-data estimators are asymptotically consistent when N and N_u go to $+\infty$ [14].

2.2.4 Aggregation procedures

The final part of the estimation of the Shapley effects is the *aggregation procedure*. It consists in the use of the previous estimations of the conditional indices to deduce an estimation of the d Shapley values.

The following procedure is immediate and natural, and is called *subset procedure* in [14]:

1. estimate the conditional indices for all $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$ (see Remark 1)
2. for all $i \in [\![1,d]\!]$, estimate Sh_i with (4).

In the given-model framework, the computational cost to estimate all the Shapley values with this aggregation procedure is then $N_{\mathbb{V}} + (2^d - 2)N_I N_O$ with the double Monte Carlo method and $N_{\mathbb{V}} + 2(2^d - 2)N_O$ with the Pick-Freeze method, where $N_{\mathbb{V}}$ is the size of the sample used to estimate $\mathbb{V}(\phi(\mathbf{X}))$ and $N_O = N_u$ for all $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$ (notation of (6)). Note that $2^d - 2$ is the cardinal of $\mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$. Its computational cost prohibits its direct use when d increases, however, note that [14] also suggests using different values of N_u for $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$ to tackle larger values of d with a bounded computational budget.

The *random-permutation procedure* was introduced by [20] in the context of game theory and its computational cost was later reduced by [17]. The idea is to rewrite the Shapley effect Sh_i as an expectation over the set of all the permutations of $[\![1,d]\!]$, denoted as $\mathcal{S}(d)$:

$$\forall i \in [\![1,d]\!],\ \mathrm{Sh}_i = \frac{1}{\mathbb{V}(\phi(\mathbf{X}))}\mathbb{E}_\Pi\left(\mathrm{VE}_{P_i(\Pi)\cup\{i\}} - \mathrm{VE}_{P_i(\Pi)}\right), \tag{11}$$

where for $\pi \in \mathcal{S}(d)$, $P_i(\pi) = \left\{\pi(j)/j \in [\![1, \pi^{-1}(i) - 1]\!]\right\}$ and Π is a random variable uniformly distributed over $\mathcal{S}(d)$. The expectation is then estimated using an i.i.d. sample $(\pi_j)_{j\in[\![1,m]\!]}$ of permutations uniformly distributed over $\mathcal{S}(d)$ with $m \ll d!$. In the given-model framework, the computational cost of the improved algorithm to estimate all the Shapley values proposed by [17] is $N_{\mathbb{V}} + m(d-1)N_I N_O$ with the double Monte Carlo method and $N_{\mathbb{V}} + 2m(d-1)N_O$ with the Pick-Freeze method. The numerical experiments in [14] suggest that it has a higher variance than the subset procedure but its computational cost can be controlled with the parameter m.

2.3 Extension to reliability-oriented sensitivity analysis

Reliability analysis consists in the estimation of the failure probability $p_t = \mathbb{P}(\phi(\mathbf{X}) > t)$, for a fixed known threshold $t \in \mathbb{R}$. Classical Monte Carlo sampling is not adapted to this problem when p_t is getting smaller because its computational cost becomes too large to obtain an accurate estimation. Therefore, several techniques have been developed in order to estimate p_t more accurately: one can mentioned FORM/SORM methods [28,29], subset sampling [30] or line sampling [31] for example. Another method, importance sampling, is reviewed here before coming back to sensitivity analysis for reliability purpose.

2.3.1 Importance sampling

Importance sampling (IS) is a very usual variance-reduction technique which was introduced by [32] and applied for the first time in reliability analysis by [33]. In the case of the estimation of a failure probability $p_t = \mathbb{P}(\phi(\mathbf{X}) > t) = \mathbb{E}_{f_\mathbf{X}}(1(\phi(\mathbf{X}) > t))$, it consists in rewriting the expectation according to an auxiliary density $g : \mathbb{X} \longrightarrow \mathbb{R}_+$ as $\mathbb{E}_g(1(\phi(\mathbf{X}) > t)\, w^g(\mathbf{X}))$, where $w^g(\mathbf{x}) = f_\mathbf{X}(\mathbf{x})/g(\mathbf{x})$ is the *likelihood ratio*. To get an unbiased estimate, the support of g must contain the support of $\mathbf{x} \in \mathbb{X} \mapsto 1(\phi(\mathbf{x}) > t)\, f_\mathbf{X}(\mathbf{x})$. The corresponding estimator is then given by:

$$\widehat{p}_{t,N}^{IS} = \frac{1}{N}\sum_{n=1}^{N} 1\left(\phi(\mathbf{X}^{(n)}) > t\right) w^g\left(\mathbf{X}^{(n)}\right), \tag{12}$$

with $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ an i.i.d. sample distributed according to the IS auxiliary distribution g. It is consistent and unbiased, and it has zero-variance if and only if $g = g_{\text{opt}}$ with $\forall \mathbf{x} \in \mathbb{X}$, $g_{\text{opt}}(\mathbf{x}) \propto 1(\phi(\mathbf{x}) > t)\, f_\mathbf{X}(\mathbf{x})$, which is the

density of the input distribution with PDF $f_{\mathbf{X}}$ restricted to the failure domain [4]. This optimal density can not be considered in practice because the normalizing constant is p_t, which is the quantity to estimate, but many techniques exist to approach g_{opt} by a near-optimal auxiliary density: methods based on the design point [34], non-parametric methods [35] or the cross-entropy method [36,37].

2.3.2 Target sensitivity analysis

Reliability-oriented sensitivity analysis can be divided into two categories, regarding the goal considered [6,38]: *target sensitivity analysis* and *conditional sensitivity analysis*. In this article, only the first one will be examined. TSA combines both reliability and sensitivity analyses and aims at studying the influence of each input variable on the occurrence of the failure event. To that end, one can apply the previous variance-based global sensitivity indices described in Section 2.1 to the quantity of interest $1\left(\phi\left(\mathbf{X}\right)>t\right)$ instead of the output $Y=\phi\left(\mathbf{X}\right)$: these new indices are called *target indices* [6,13]. In particular, the definitions (3) to (5) are directly extended by replacing $\phi\left(\mathbf{X}\right)$ by $1\left(\phi\left(\mathbf{X}\right)>t\right)$. Techniques based on Monte Carlo sampling [8] and non-parametric estimation [9] have been proposed to estimate first-order and total-order target Sobol indices, whereas as far as we know, estimators of the target Shapley effects T-Sh$_i$ have only been recently suggested in [13] and are strongly inspired from those in [14]. The estimators from [13] are based on a classic Monte Carlo sampling from the input distribution of PDF $f_{\mathbf{X}}$ in both given-model and given-data frameworks. In the latter framework in [13], a reliability analysis is first performed and leads to the estimation of p_t with a Monte Carlo sampling. Then, the target Shapley effects are estimated from the available sample with the estimator in (9) combined with the random permutation procedure. However, numerical experiments in Section 4 highlight their limits when the failure probability is getting smaller: their variance increases and a large Monte Carlo sample is thus required to get an accurate estimation, and make them hardly applicable with a costly computer model ϕ. The main goal of this article is thus to remedy these shortcomings with importance sampling.

3. TARGET SHAPLEY EFFECTS ESTIMATION BY IMPORTANCE SAMPLING

In this section, we suggest new estimators of the d target Shapley effects T-Sh$_i$, defined as in (4) with $\phi\left(\mathbf{X}\right)$ replaced by $1\left(\phi\left(\mathbf{X}\right)>t\right)$, by importance sampling when the input variables are correlated. To that end, we propose four new estimators of the target conditional indices T-VE$_u$ and T-EV$_u$, defined as VE$_u$ and EV$_u$ with $\phi\left(\mathbf{X}\right)$ replaced by $1\left(\phi\left(\mathbf{X}\right)>t\right)$, based on those of the previous section:

1. an unbiased estimator of T-EV$_u$ by double Monte Carlo with importance sampling given-model
2. an estimator of T-EV$_u$ by double Monte Carlo with importance sampling given-data
3. an unbiased estimator of T-VE$_u$ by Pick-Freeze with importance sampling given-model
4. an estimator of T-VE$_u$ by Pick-Freeze with importance sampling given-data.

Recall that the given-model framework is described at the beginning of Section 2.2.1 and that the given-data framework is described in Section 2.2.3. Then, the two aggregation procedures described in Section 2.2.4 provide eight new estimators of the target Shapley effects which have a lower variance than the existing ones when the auxiliary density is adapted to the problem, as will be shown numerically in Section 4.

For any subset $u\in\mathcal{P}(d)\backslash\{\varnothing,[\![1,d]\!]\}$, for any $\mathbf{x}_u\in\mathbb{X}_u$, we let $f_{\mathbf{X}_u}$ and $f_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u}$ denote respectively the PDF of the marginal distribution of $\mathbf{X}_u$ and the PDF of the conditional distribution of $\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u$. In addition, let $g:\mathbb{X}\to\mathbb{R}_+$ be the PDF of the IS auxiliary distribution as in Section 2.3.1 and $g_{\mathbf{X}_u}$ and $g_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u}$ be the PDFs of its marginal and its conditional distributions. Moreover, in order to lighten the notations, for all $t\in\mathbb{R}$ and for all $\mathbf{x}\in\mathbb{X}$, let us write:

$$\psi_t\left(\mathbf{x}\right)=1\left(\phi\left(\mathbf{x}\right)>t\right)\text{ and }w_t^g\left(\mathbf{x}\right)=\psi_t\left(\mathbf{x}\right)\frac{f_{\mathbf{X}}\left(\mathbf{x}\right)}{g\left(\mathbf{x}\right)}. \tag{13}$$

Finally, in the following, the convention $0/0=0$ will be used, and in particular, this implies that for all $\mathbf{x}_{-u}\in\mathbb{X}_{-u}$, $f_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u}\left(\mathbf{x}_{-u}\right)=0$ for any $\mathbf{x}_u\in\mathbb{X}_u$ such that $f_{\mathbf{X}_u}\left(\mathbf{x}_u\right)=0$ (see also Remarks 3 and 4).

3.1 Estimation of T-EV$_u$ by double Monte Carlo with importance sampling

For the same reasons as in [17], for $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, we choose to estimate the target conditional index T-EV$_u$ by double Monte Carlo. However, contrary to the existing literature, in order to introduce importance sampling in the estimation process we write this target conditional index as:

$$\text{T-EV}_u = \mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{V}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)\right] = p_t - \mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right]. \tag{14}$$

The estimator $\widehat{p}^{\text{IS}}_{t,N}$ in (12) already provides an unbiased and convergent estimation of the failure probability p_t by importance sampling. The main problem is thus to estimate $\mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right]$ by importance sampling. Let us rewrite this quantity according to the IS auxiliary density g (see APPENDIX A):

$$\mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right] = \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X})\,\frac{f_{\mathbf{X}_u|\mathbf{X}_{-u}}(\mathbf{X}_u)}{g_{\mathbf{X}_u|\mathbf{X}_{-u}}(\mathbf{X}_u)}\middle|\,\mathbf{X}_{-u}\right)^2 \frac{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\right]. \tag{15}$$

However, in the rest of the article, we will not assume that it is possible to evaluate directly the conditional PDFs $f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}$ and $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}$ at any point of $\mathbb{X}_u$ for all $\mathbf{x}_{-u} \in \mathbb{X}_{-u}$, as it can be a restrictive hypothesis in practice, especially with non trivial input distributions. Therefore, using the definition of the conditional PDF, we choose to rewrite the ratio $f_{\mathbf{X}_u|\mathbf{X}_{-u}}/g_{\mathbf{X}_u|\mathbf{X}_{-u}}$ in the form $f_{\mathbf{X}}/f_{\mathbf{X}_{-u}} \times g_{\mathbf{X}_{-u}}/g_{\mathbf{X}}$. From this point, some calculations developed in APPENDIX A lead to the following lemma.

Lemma 1. *For any IS auxiliary density $g : \mathbb{X} \longrightarrow \mathbb{R}_+$, we have:*

$$\mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right] = \mathbb{E}_g\left[\mathbb{E}_g\left(w_t^g(\mathbf{X})\mid\mathbf{X}_{-u}\right)^2 \frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\right]. \tag{16}$$

Proof. See APPENDIX A. □

Because of the described transformation, one can notice that the likelihood ratios in (16) are not all in the classic form $f_{\mathbf{X}}/g$, there is a term in the form $g/f_{\mathbf{X}}$ in the outer expectation. From (16), it is thus possible to introduce double Monte Carlo estimators by importance sampling of the target conditional index T-EV$_u$ in both given-model and given-data frameworks.

3.1.1 Given-model framework with importance sampling

In the given-model framework with importance sampling, assume that:

- we can evaluate the code ϕ in any point of $\mathbb{X}$
- it is possible to generate an i.i.d. sample from the distribution of $g_{\mathbf{X}_{-u}}$
- for any $\mathbf{x}_{-u} \in \mathbb{X}_{-u}$, it is possible to generate an i.i.d. sample from the distribution of $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}$
- it is possible to evaluate $f_{\mathbf{X}}$ and g in any point of $\mathbb{X}$
- it is possible to evaluate $f_{\mathbf{X}_{-u}}$ and $g_{\mathbf{X}_{-u}}$ in any point of $\mathbb{X}_{-u}$.

The third hypothesis is reasonable because we are free to choose g such that the sampling from the conditional distributions is not problematic. Moreover, it is no longer necessary to evaluate the conditional PDFs of $f_{\mathbf{X}}$ and g, which is convenient as discussed above. The new writing (16) suggests then to introduce the following double Monte Carlo estimator by importance sampling:

$$\widehat{\text{T-EV}}^{\text{IS}}_{u,\text{MC}} = \widehat{p}^{\text{IS}}_{t,N} - \left(\frac{1}{N_u}\sum_{n=1}^{N_u}\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2 \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(n)}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(n)}\right)} - \widehat{E}^{IS}_{\text{bias},u}\right), \tag{17}$$

where $\left(\mathbf{X}_{-u}^{(n)}\right)_{n\in[\![1,N_u]\!]}$ is an i.i.d. sample from the distribution of $g_{\mathbf{X}_{-u}}$, where for all $n \in [\![1, N_u]\!]$, $\left(\mathbf{X}_{u}^{(n,k)}\right)_{k\in[\![1,N_I]\!]}$ is an i.i.d. sample from the distribution of $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{X}_{-u}^{(n)}}$ and

$$\overline{\psi_t^{\text{IS}}\left(\mathbf{X}_{-u}^{(n)}\right)} = \frac{1}{N_I}\sum_{k=1}^{N_I} w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right), \tag{18}$$

and where the term $\widehat{E}_{\text{bias},u}^{\text{IS}}$ is defined by:

$$\widehat{E}_{\text{bias},u}^{\text{IS}} = \frac{1}{N_u}\sum_{n=1}^{N_u}\frac{1}{N_I-1}\left[\frac{1}{N_I}\sum_{k=1}^{N_I} w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right)^2 - \left(\overline{\psi_t^{\text{IS}}\left(\mathbf{X}_{-u}^{(n)}\right)}\right)^2\right]\frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(n)}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(n)}\right)}. \tag{19}$$

As in [17], this estimator is composed of an inner loop of size N_I for the inner conditional expectation and an outer loop of size N_u (which depends on u) for the outer expectation. Moreover, it is clear that for all $n \in [\![1, N_u]\!]$, $\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}$ is an unbiased estimator of $\mathbb{E}_g\left(w_t^g(\mathbf{X}) \mid \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)$. However, $\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2$ is not an unbiased estimator of $\mathbb{E}_g\left(w_t^g(\mathbf{X}) \mid \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)^2$ because of the square which creates a bias. This bias from the inner loop spreads itself into the outer loop and finally the bias of the uncorrected double Monte Carlo estimator by importance sampling of T-EV$_u$ (obtained by removing $\widehat{E}_{\text{bias},u}^{IS}$ in (17)) is $\mathbb{E}_g\left[\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}|\mathbf{X}_{-u}\right) g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u}) \,/\, f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})\right]$, where for all $\mathbf{x}_{-u} \in \mathbb{X}_{-u}$:

$$\overline{\psi_t^{\text{IS}}\left(\mathbf{x}_{-u}\right)} = \frac{1}{N_I}\sum_{k=1}^{N_I} w_t^g\left(\mathbf{X}_u^{(k)}, \mathbf{x}_{-u}\right), \tag{20}$$

with $\left(\mathbf{X}_u^{(k)}\right)_{k\in[\![1,N_I]\!]}$ an i.i.d. sample from the distribution of $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}$. The term $\widehat{E}_{\text{bias},u}^{IS}$ in (19) is then an unbiased estimator of the latter bias and allows to correct it. This leads to the main result of this section:

Proposition 1. *$\widehat{\text{T-EV}}_{u,MC}^{IS}$ in* (17) *is an unbiased double Monte Carlo estimator by importance sampling of the conditional index T-EV$_u$, and requires $N_u N_I$ calls to the function ϕ in addition to those to estimate p_t.*

Proof. See APPENDIX B. □

3.1.2 Given-data framework with importance sampling

In the given-data framework with importance sampling, assume that:

- an i.i.d. input/output sample $\left(\mathbf{X}^{(n)}, \psi_t\left(\mathbf{X}^{(n)}\right)\right)_{n\in[\![1,N]\!]}$ with $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ distributed according to the IS auxiliary distribution g is available
- it is possible to evaluate $f_{\mathbf{X}}$ and g in any point of $\mathbb{X}$
- it is possible to evaluate $f_{\mathbf{X}_{-u}}$ and $g_{\mathbf{X}_{-u}}$ in any point of $\mathbb{X}_{-u}$.

Contrary to the given-model framework, here the black-box model ϕ is no longer available and it is not possible to generate any additional sample from any distribution. Typically, in the context of ROSA, the i.i.d. sample from g is already obtained from a previous importance-sampling-based estimation of p_t (see Section 3.3). The extension of the double Monte Carlo estimator of T-EV$_u$ by importance sampling (17) to the given-data framework is based on the nearest-neighbours method introduced in [14] and briefly described in Section 2. Here, for $i \in [\![1, N]\!]$, the exact

sampling from the conditional auxiliary PDF $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{X}_{-u}^{(i)}}$ is approximated by the N_I nearest neighbours of $\mathbf{X}_{-u}^{(i)}$ among the sample $\left(\mathbf{X}_{-u}^{(n)}\right)_{n\in[\![1,N]\!]}$. Then, the extended estimator is:

$$\widehat{\text{T-EV}}_{u,\text{MC}}^{\text{IS,KNN}} = \widehat{p}_{t,N}^{\text{IS}} - \left(\frac{1}{N_u} \sum_{n=1}^{N_u} \left(\overline{\psi_{t,u}^{\text{IS,KNN}}(\mathbf{X}^{(s(n))})} \right)^2 \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(s(n))}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(s(n))}\right)} - \widehat{E}_{\text{bias},u}^{\text{IS,KNN}} \right), \tag{21}$$

where $(s(n))_{n\in[\![1,N_u]\!]}$ is a sample of uniformly distributed integers in $[\![1,N]\!]$, where for all $n \in [\![1,N_u]\!]$,

$$\overline{\psi_{t,u}^{\text{IS,KNN}}\left(\mathbf{X}^{(s(n))}\right)} = \frac{1}{N_I} \sum_{k=1}^{N_I} w_t^g \left(\mathbf{X}^{(k_N^{-u}(s(n),k))} \right), \tag{22}$$

with k_N^{-u} defined in Section 2.2.3 and where the term $\widehat{E}_{\text{bias},u}^{\text{IS,KNN}}$ is defined by:

$$\widehat{E}_{\text{bias},u}^{\text{IS,KNN}} = \frac{1}{N_u} \sum_{n=1}^{N_u} \frac{1}{N_I - 1} \left[\frac{1}{N_I} \sum_{k=1}^{N_I} w_t^g \left(\mathbf{X}^{(k_N^{-u}(s(n),k))} \right)^2 - \left(\overline{\psi_{t,u}^{\text{IS,KNN}}\left(\mathbf{X}^{(s(n))}\right)} \right)^2 \right] \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(s(n))}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(s(n))}\right)}. \tag{23}$$

It does not require any additional call to ϕ to those from the already available i.i.d. sample. In the same way as the existing given-data estimators in GSA and ROSA (see Section 2), the costly part of this algorithm is the computation of the nearest neighbours at each step. Moreover, both estimators (17) and (21) have the same general structure. In (17), the bias created by the square in the inner loop is corrected by the estimator (19). This structure is kept in (21) and (23). Nevertheless, for given finite values of N_u and N_I, the given-data estimator (21) still has a bias caused by the nearest neighbour approximation, and as far as we know, no article has proposed any theoretical study of this bias.

Remark 2. *Let us provide some additional motivations for the given-data with importance sampling framework. First, we assumed here that the computer model is no longer available and that we can only use the available sample to estimate the target Shapley effects. One can notice that it might be possible to get around this problem by fitting a surrogate model, such as a Gaussian process for example, with the available sample and then coming back to the given-model framework. However, this approach has some drawbacks that a practitioner may prefer to avoid. In fact, when N is large, it is not straightforward to fit efficiently a surrogate model, even more when the input dimension increases. Furthermore, even if the given-model framework with a surrogate model does not suffer from the nearest-neighbour approximation, the surrogate model introduces another approximation error which is hard to quantify and is added to the estimation error of each index. This error is not necessarily lower than the error due to the nearest-neighbour approximation when the dimension increases.*

Second, we assumed that it is not possible to draw additional samples from any distribution. In the given-model framework, it is necessary to be able to draw samples according to the conditional distributions of g. In practice, most of the time, Gaussian auxiliary distributions are chosen and it is then straightforward to sample from their conditional distributions. However, in some cases, auxiliary distributions belonging to other parametric families could be more relevant but do not satisfy this criterion.

3.2 Estimation of T-VE$_u$ by Pick-Freeze with importance sampling

For any subset $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, we estimate the target conditional index T-VE$_u$ by Pick-Freeze and with importance sampling. In ROSA with correlated inputs, we recall the fundamental equation of Pick-Freeze (7):

$$\text{T-VE}_u = \mathbb{V}_{f_\mathbf{X}}\left[\mathbb{E}_{f_\mathbf{X}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] = \mathbb{E}_{f_\mathbf{X}}\left[\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right] - p_t^2, \tag{24}$$

where $\mathbf{X}^u = \left(\mathbf{X}_u, \mathbf{X}'_{-u}\right)$ with $\mathbf{X}'_{-u} \overset{d}{=} \mathbf{X}_{-u}|\mathbf{X}_u$ and $\mathbf{X}'_{-u} \perp\!\!\!\perp \mathbf{X}_{-u}|\mathbf{X}_u$. For $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ an i.i.d. sample distributed according to the IS auxiliary distribution g, the estimator

$$\widehat{p}_{t,N}^{\text{IS,ub}} = \left(\widehat{p}_{t,N}^{\text{IS}}\right)^2 - \frac{1}{N-1}\left[\frac{1}{N}\sum_{n=1}^{N} w_t^g\left(\mathbf{X}^{(n)}\right)^2 - \left(\widehat{p}_{t,N}^{\text{IS}}\right)^2\right] \tag{25}$$

is easily shown (see APPENDIX C) to be an unbiased estimator by importance sampling of p_t^2. The main problem is then to estimate the expectation $\mathbb{E}_{f_{\mathbf{X}}}\left[\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right]$ by importance sampling. Let us rewrite it according to the IS auxiliary density g:

Lemma 2. *For any IS auxiliary density $g : \mathbb{X} \longrightarrow \mathbb{R}_+$, we have:*

$$\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right) = \mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right) w_t^g\left(\mathbf{X}^u\right) \frac{g_{\mathbf{X}_u}(\mathbf{X}_u)}{f_{\mathbf{X}_u}(\mathbf{X}_u)}\right). \tag{26}$$

Proof. See APPENDIX D. □

The term in the expectation is a function of the three random variables $\mathbf{X}_u$, $\mathbf{X}_{-u}$ and $\mathbf{X}'_{-u}$ with the correlation structure described above. From (26), it is then possible to propose Pick-Freeze estimators by importance sampling of the target conditional index T-VE$_u$ in both given-model and given-data frameworks.

3.2.1 Given-model framework with importance sampling

Here, we make the assumptions of the given-model with importance sampling framework defined in Section 3.1.1. The writing (26) suggests then to introduce the following Pick-Freeze estimator of T-VE$_u$ by importance sampling:

$$\widehat{\text{T-VE}}_{u,\text{PF}}^{\text{IS}} = \frac{1}{N_u}\sum_{n=1}^{N_u} w_t^g\left(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(n,1)}\right) w_t^g\left(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(n,2)}\right) \frac{g_{\mathbf{X}_u}(\mathbf{X}_u^{(n)})}{f_{\mathbf{X}_u}(\mathbf{X}_u^{(n)})} - \widehat{p}_{t,N}^{\text{IS,ub}}, \tag{27}$$

where $(\mathbf{X}_u^{(n)})_{n\in[\![1,N_u]\!]}$ is an i.i.d. sample from the distribution of $g_{\mathbf{X}_u}$ and where for all $n \in [\![1, N_u]\!]$, $(\mathbf{X}_{-u}^{(n,k)})_{k\in[\![1,2]\!]}$ are two independent random variables from the distribution of $g_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{X}_u^{(n)}}$. This then leads to the main result of this sub-section:

Proposition 2. *$\widehat{\text{T-VE}}_{u,PF}^{IS}$ in (27) is an unbiased estimator of T-VE$_u$ and it requires $2N_u$ calls to the function ϕ in addition to those to estimate p_t^2.*

Proof. The first term in the estimator (27) is an unbiased estimator of the expectation $\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right) w_t^g\left(\mathbf{X}^u\right) \frac{g_{\mathbf{X}_u}(\mathbf{X}_u)}{f_{\mathbf{X}_u}(\mathbf{X}_u)}\right)$ since it is its empirical mean, and $\widehat{p}_{t,N}^{\text{IS,ub}}$ in (25) is an unbiased estimator of p_t^2. The linearity of the expectation concludes the proof. □

3.2.2 Given-data framework with importance sampling

We make here the same assumptions of the given-data with importance sampling framework defined in Section 3.1.2. In particular, an i.i.d. input/output sample $\left(\mathbf{X}^{(n)}, \psi_t\left(\mathbf{X}^{(n)}\right)\right)_{n\in[\![1,N]\!]}$ with $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ distributed according to the IS auxiliary distribution g is available. Once more, we will use the nearest-neighbour approximation to extend the Pick-Freeze estimator by importance sampling (27) of T-VE$_u$ to the given-data framework and the corresponding estimator is:

$$\widehat{\text{T-VE}}_{u,\text{PF}}^{\text{IS,KNN}} = \frac{1}{N_u}\sum_{n=1}^{N_u} w_t^g\left(\mathbf{X}^{(k_N^u(s(n),1))}\right) w_t^g\left(\mathbf{X}^{(k_N^u(s(n),2))}\right) \frac{g_{\mathbf{X}_u}(\mathbf{X}_u^{(s(n))})}{f_{\mathbf{X}_u}(\mathbf{X}_u^{(s(n))})} - \widehat{p}_{t,N}^{\text{IS,ub}}, \tag{28}$$

where $(s(n))_{n\in[\![1,N_u]\!]}$ is a sample of uniformly distributed integers in $[\![1,N]\!]$ and where k_N^u is defined in Section 2.2.3. This estimator does not require any additional call to the function ϕ to those from the already available i.i.d. sample and more generally, all the remarks previously made in Section 3.1.2 about the double Monte Carlo given-data estimator by importance sampling (21) of T-EV$_u$ are still valid.

Remark 3. *In all the above estimators* (17) (21) (27) *and* (28)*, for $v \in \{u, -u\}$, it can be possible to draw a sample $\mathbf{X}_v^{(i)}$ according to the IS auxiliary distribution $g_{\mathbf{X}_v}$ such that $f_{\mathbf{X}_v}\left(\mathbf{X}_v^{(i)}\right) = 0$ which is at the denominator. However, this implies that for all $\mathbf{x}_{-v} \in \mathbb{X}_{-v}$, $f_{\mathbf{X}}\left(\mathbf{X}_v^{(i)}, \mathbf{x}_{-v}\right) = 0$ and then $w_t^g\left(\mathbf{X}_v^{(i)}, \mathbf{x}_{-v}\right) = 0$. Thus, the convention $0/0 = 0$ adopted at the beginning of the section allows to set as 0 the term corresponding to this sample in all the previous estimators.*

3.3 From reliability analysis to reliability-oriented sensitivity analysis

In practice, the ROSA of a complex system always comes after the reliability analysis, i.e. the estimation of the failure probability. When the latter has been done with importance sampling, we already have at our disposal a sub-optimal auxiliary PDF g close to the optimal one $g_{\text{opt}}(\mathbf{x}) \propto 1(\phi(\mathbf{x}) > t) f_{\mathbf{X}}(\mathbf{x})$ (see Section 2.3.1), as well as an i.i.d. N-sample distributed according to it. In order to check if it is beneficial to re-use the available data from the reliability analysis to estimate the target conditional indices and thus the target Shapley values, let us consider the variance of the previous new estimators when the auxiliary density is g_{opt}. Since it is based on a double loop, it may be complicated to write the variance of the double Monte-Carlo estimator (17) in closed form, but this is feasible for the Pick-Freeze estimator (27) since it is an empirical mean. Considering N_u as fixed, we have (see APPENDIX E.1):

$$\mathbb{V}_{g_{\text{opt}}}\left(\widehat{\text{T-VE}}_{u,\text{PF}}^{\text{IS}}\right) \leq \frac{p_t^2}{N_u} \propto p_t^2. \tag{29}$$

This result does not prove that g_{opt} is the optimal IS auxiliary distribution to estimate the target closed Sobol index T-VE$_u$ from (27), but it proves that using g_{opt} as the IS auxiliary density improves its estimation in comparison to its existing estimators without importance sampling in the regime where $p_t \to 0$. Indeed, letting $\widehat{\text{T-VE}}_{u,\text{PF}}$ be the unbiased estimator by Pick-Freeze without importance sampling of T-VE$_u$, defined as in (8) with ϕ replaced by ψ_t, we have (see APPENDIX E.2):

$$\mathbb{V}_{f_{\mathbf{X}}}\left(\widehat{\text{T-VE}}_{u,\text{PF}}\right) \geq \frac{1}{N_u}\left(\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] + p_t^2\right)\left(1 - \left(\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] + p_t^2\right)\right). \tag{30}$$

Thus, at best, when ϕ depends only on $\mathbf{X}_{-u}$, we have $\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] = 0$ and then $\mathbb{V}_{f_{\mathbf{X}}}\left(\widehat{\text{T-VE}}_{u,\text{PF}}\right) \geq p_t^2\left(1-p_t^2\right)/N_u \propto p_t^2\left(1-p_t^2\right) \underset{p_t\to 0}{\sim} p_t^2$. At worst, when ϕ depends only on $\mathbf{X}_u$, we have $\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] = p_t(1-p_t)$ and then $\mathbb{V}_{f_{\mathbf{X}}}\left(\widehat{\text{T-VE}}_{u,\text{PF}}\right) \geq p_t(1-p_t)/N_u \propto p_t(1-p_t) \underset{p_t\to 0}{\sim} p_t$. Since p_t is the probability of a rare event, $p_t \ll 1$ and so $p_t^2 \ll p_t$, thus using g_{opt} as the IS auxiliary density improves the estimation of all the target closed Sobol indices in the regime where $p_t \longrightarrow 0$, and then of the d target Shapley effects.

However, in practice, the available sample is not drawn exactly from g_{opt} but from an IS auxiliary distribution g close to g_{opt}, but we can still expect a significant improvement in the estimation of the target closed Sobol indices. Hence, this theoretical analysis supports the intuition that it is beneficial to re-use the available data from the reliability analysis to estimate the target Shapley effects.

Remark 4. *The input domain $\mathbb{X}$ is not necessarily equal to $\mathbb{R}^d$. Nevertheless, it can be practically convenient to use an IS auxiliary distribution supported on $\mathbb{R}^d$ (or more generally on a subset of $\mathbb{R}^d$ strictly greater than $\mathbb{X}$) such as a normal distribution for example. To that end, the solution adopted in this article consists in extending the input domain of ϕ and $f_{\mathbf{X}}$ on $\mathbb{R}^d$ by setting $\phi(\mathbf{x}) = 0$ and $f_{\mathbf{X}}(\mathbf{x}) = 0$ for any $\mathbf{x} \in \mathbb{R}^d\backslash\mathbb{X}$. The convention $0/0 = 0$ allows then to ignore the samples drawn by g on $\mathbb{R}^d\backslash\mathbb{X}$ during the estimation process (see Remark 3). However, if the reliability analysis has been done efficiently, the IS auxiliary distribution g should be close to $g_{opt}(\mathbf{x}) \propto 1(\phi(\mathbf{x}) > t) f_{\mathbf{X}}(\mathbf{x})$ and therefore very few samples should be drawn in $\mathbb{R}^d\backslash\mathbb{X}$.*

4. NUMERICAL APPLICATIONS

In order to illustrate the practical interest of the previous efforts, this section aims to evaluate numerically the performances of the suggested estimators of the target Shapley effects on various test functions with correlated input variables and to compare them to the performances of the existing estimators. The code to reproduce the numerical experiments is publicly available at: `https://github.com/Julien6431/Target-Shapley-effects.git`

In the following examples, as explained in Section 3.3, we will consider that the reliability analysis has already been done, i.e. that an IS auxiliary density g close to g_{opt} has been determined. The present article does not aim to compare different importance sampling techniques, therefore we choose here to compute the IS auxiliary PDF g only by adaptive importance sampling with the cross-entropy algorithm [37], from one of the two following IS parametric families: the Gaussian distributions (single Gaussian, IS-SG) [37] and the Gaussian mixture (IS-GM) distributions [39].

Moreover, the dimension of the following problems will be low or moderate, therefore only the subset aggregation procedure will be used, and we adopt then the following numerical parameters:

- $N_{tot} = 2 \times 10^4$ which represents the total number of calls to ϕ
- $N_I = 3$ in Sections 4.1 and 4.2 as suggested in [14,17], $N_I = 2$ in Section 4.3, the size of the inner loop in the double Monte Carlo estimator
- $N_{\mathbb{V}} = 10^4$ the size of the sample to estimate $\mathbb{V}\left(\psi_t\left(\mathbf{X}\right)\right)$ in the given-model framework
- in the given-model framework, for $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, $N_u = N_O = \left\lfloor \left(N_{tot} - N_{\mathbb{V}}\right)\left(N_I(2^d-2)\right)^{-1} \right\rfloor$ with the double Monte Carlo estimators and $N_u = N_O = \left\lfloor \left(N_{tot} - N_{\mathbb{V}}\right)\left(2(2^d-2)\right)^{-1} \right\rfloor$ with the Pick-Freeze estimators, in order to reach N_{tot} calls or less to ϕ (according to both expressions of N_{tot} given in Section 2.2.4)
- in the given-data framework, for all $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, $N_u = N_O = 10^3$
- $n_{rep} = 200$ realisations of each estimator to represent the results as boxplots.

For different reasons described in more details in APPENDIX G, the preprocessing procedure presented in APPENDIX G.1 will be applied as soon as a given-data estimator will be used. It is based on Theorem 9 stated by [18].

In the following, results will be presented as boxplots. Let us define first the acronyms that will be used in the legends:

- *IS-SG* refers to estimators with importance sampling using an auxiliary distribution in the single Gaussian family determined by the cross-entropy algorithm,
- *IS-GM* refers to estimators with importance sampling using an auxiliary distribution in the Gaussian mixture family determined by the cross-entropy algorithm.

Then, each box extends from the first to the third quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data, and flier points are those past the end of the whiskers.

4.1 Gaussian linear case

First, let us consider the simple Gaussian linear example. For $d \geq 2$ and a vector $\boldsymbol{\beta} \in \mathbb{R}^d\backslash\{0\}$, let us define the d-dimensional linear function $\phi_{\boldsymbol{\beta}}$ by:

$$\phi_{\boldsymbol{\beta}} : \left| \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \mathbb{R} \\ \mathbf{x} & \longmapsto & \boldsymbol{\beta}^\top \mathbf{x}. \end{array} \right. \tag{31}$$

Then, the input vector $\mathbf{X}$ is assumed normally distributed with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma} \in \mathcal{M}_d\left(\mathbb{R}\right)$ which is symmetric positive-definite, i.e. $\mathbf{X} \sim \mathcal{N}_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$. Moreover the covariance matrix $\boldsymbol{\Sigma}$ is here considered non-diagonal in order to include dependence between the input variables. Finally, since the failure domain is here clearly located in one region of $\mathbb{R}^d$, only Gaussian auxiliary IS distributions will be considered.

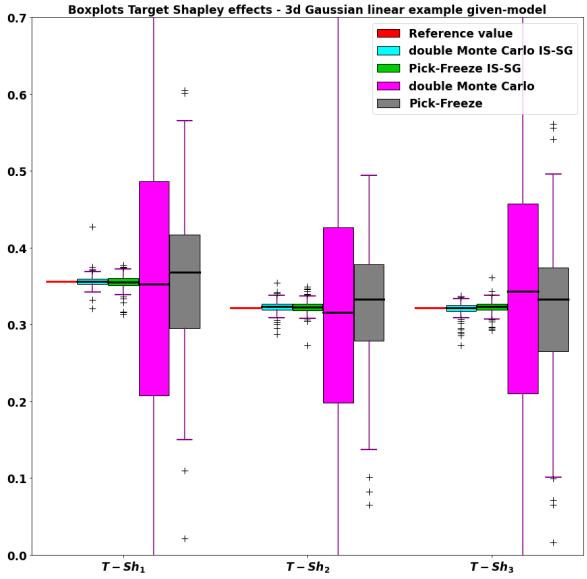

FIG. 1: Estimation of the target Shapley effects in the 3-dimensional Gaussian linear example in the given-model framework.

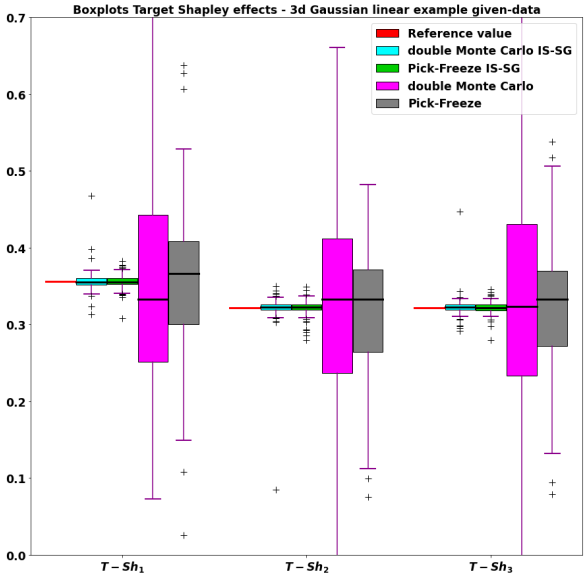

FIG. 2: Estimation of the target Shapley effects in the 3-dimensional Gaussian linear example in the given-data framework.

The parameters of the toy case are specified by $\boldsymbol{\beta} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^\top$, $\boldsymbol{\mu} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^\top$, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.3 \\ 0 & -0.3 & 1 \end{pmatrix}$ and $t = 4$. The failure threshold is set at $t = 4$ such that the failure probability is $p_t^{\boldsymbol{\beta}} \approx 4.9 \times 10^{-3}$. Reference values of the target Shapley effects will be computed with (F.5), (F.4) and the definition in (4). The performances of the estimators with and without importance sampling in both given-model and given-data frameworks are compared graphically respectively in Figure 1 and Figure 2 (recall that the acronyms are defined above). As expected, in both frameworks, when the IS auxiliary density is adapted to the problem, the estimators with importance sampling give a much better estimation of the target Shapley effects when the failure probability is small. Despite few outliers, the corresponding boxplots have a much smaller stretch and they are centered on the reference values. This first example highlights then that importance sampling has a huge positive impact on the quality of the estimation of the target Shapley effects with a low failure probability.

4.2 Cantilever beam

4.2.1 Presentation of the model

The second example is a real structure engineering problem which is presented in [40,41]. Consider a rectangular cantilever beam structure. The dimensional parameters of the beam are denoted l_X, l_Y and L. The elastic modulus of the structure is represented by E. Two random forces F_X and F_Y are exerted on the tip of the section. The goal function is then the maximum vertical displacement of the tip section, which can be given analytically according to the previous parameters by:

$$\phi\left(F_X, F_Y, E, l_X, l_Y, L\right) = \frac{4L^3}{E l_X l_Y} \sqrt{\left(\frac{F_X}{l_X^2}\right)^2 + \left(\frac{F_Y}{l_Y^2}\right)^2}. \tag{32}$$

The maximum vertical displacement allowed is $t = 0.066$ m, which is then the failure threshold of the reliability problem.

	Input variable	Distribution	Mean	Coefficient of variation
1	F_X	LogNormal	556.8 N	0.08
2	F_Y	LogNormal	453.6 N	0.08
3	E	LogNormal	200.10^9 Pa	0.06
4	l_X	Normal	0.062 m	0.1
5	l_Y	Normal	0.0987 m	0.1
6	L	Normal	4.29 m	0.1

TABLE 1: Distributions of each input variable of the cantilever beam example.

$\text{T-Sh}_1^{\text{ref}}$	$\text{T-Sh}_2^{\text{ref}}$	$\text{T-Sh}_3^{\text{ref}}$	$\text{T-Sh}_4^{\text{ref}}$	$\text{T-Sh}_5^{\text{ref}}$	$\text{T-Sh}_6^{\text{ref}}$
0.146	0.001	0.103	0.282	0.254	0.214

TABLE 2: Reference values of the target Shapley effects in the cantilever beam problem.

The distributions of each input variable are listed in Table 1. Moreover, the dimensional variables are linearly dependent through the following Pearson correlation coefficients:

$$\rho_{l_X,l_Y} = -0.55 \text{ and } \rho_{L,l_X} = \rho_{L,l_Y} = 0.45. \tag{33}$$

We do not know anything about the form of the failure domain, therefore both Gaussian and Gaussian mixture IS auxiliary distributions will be used in order to evaluate the influence of the IS auxiliary density on the estimation of the target Shapley effects with importance sampling. First, we will compare the performances of the given-data estimators, according to the framework described in Section 3.3. Then, in order to evaluate graphically the error due to the nearest neighbour approximation, we will evaluate the performances of the given-model estimators with importance sampling with a Gaussian IS auxiliary distribution. Since the input vector $\mathbf{X} = (F_X, F_Y, E, l_X, l_Y, L)^\top$ is not Gaussian, we will not generate samples according to the input conditional distributions and thus we will not evaluate the performances of the existing given-model estimators without importance sampling.

4.2.2 Procedure to obtain the reference values

A Monte Carlo estimation with a sample of size $N = 10^6$ gives a reference value for the failure probability: $p_t^{\text{ref}} \approx 1.5 \times 10^{-2}$. Next, one can remark that a random variable following a LogNormal distribution can be written as a bijective transformation of a Gaussian random variable. Then, thanks to Theorem 9, we decide to transform the input random variable $\mathbf{X} = (F_X, F_Y, E, l_X, l_Y, L)$ into a 6-dimensional Gaussian random vector with the correct mean vector and covariance matrix. At last, reference values of the target Shapley effects are computed using the double Monte Carlo existing estimator in (6) combined with the non-given data sampling procedure described in APPENDIX I with the input distribution and finally, with $N = 10^6$, $N_O = 10^6$ and $N_I = 3$, we obtain the reference values presented in Table 2.

4.2.3 Numerical results

The results of the ROSA of the cantilever beam problem in the given-data and given-model frameworks are respectively given in Figures 3 and 4. Once more, when the IS auxiliary density is adapted to the problem, the estimators with importance sampling give a better estimation of the target Shapley effects than the existing estimators for the same reasons as in the previous example. In addition, unreported numerical simulations provide that the failure domain is here located in one region of the input space which explains why both Gaussian and Gaussian mixture IS auxiliary densities provide similar performances. Both figures show that the dimensional parameters of the beam are the most influential inputs on the occurrence of the failure and show that the estimators with importance sampling

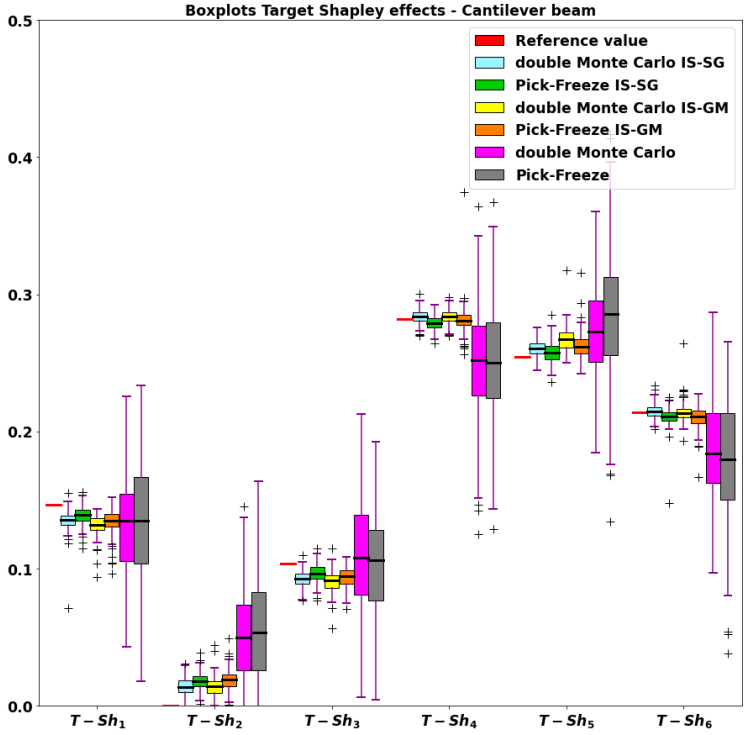

FIG. 3: Estimation of the target Shapley effects in the cantilever beam example, in the given-data framework and with the preprocessing described in APPENDIX G.1.

provide the same hierarchy of importance of the inputs as the reference values of the target Shapley effects whereas the existing estimators without importance sampling switch the importance of T-Sh$_4$ and T-Sh$_5$.

In addition, one can remark that the dispersion of each given-model boxplot in Figure 4 is bigger than the dispersion of each given-data boxplot in Figure 3. In fact, in both cases, the maximal number of calls to ϕ allowed is $N_{tot} = 2 \times 10^4$. In the given-model framework, it is necessary to choose the parameters N_I, N_O and $N_{\mathbb{V}}$ such that we exactly reach N_{tot} calls to the function, whereas in the given-data framework, we are free to chose N_I and N_O as we want because we already have the N_{tot}-sample at our disposal. In both frameworks, the value of N_I is already fixed. As a consequence, in the given-model framework, the value of N_O must be equal to the one given in the introduction of Section 4, which is always lower than 10^2 in each numerical example here, whereas in the given-data framework, we choose $N_O = 10^3$. This gap between the value of N_O in both frameworks explains the larger dispersion observed in the given-model algorithms.

Moreover, Figure 3 illustrates a problem already mentioned in Section 3.1.2. On some indices in Figure 3, there is a gap between the boxplot median and the reference value whereas the boxplots of the given-model estimators with importance sampling presented in Figure 4 are centered on the reference values. Indeed, when the dimension increases, distances between points tend to become larger and thus the nearest neighbour approximations of the conditional distributions are getting less accurate. This phenomenon might create a bias in the estimation of the target Shapley effects with the given-data estimators when the dimension increases. However, one can remark on Figure 3 that importance sampling seems to reduce this error. Indeed, without importance sampling, the points of interest, the failure points, are in the tail of the distribution, where the concentration of points is small and thus where the distances between points are larger than on average, which is not the case with importance sampling when the auxiliary distribution is adapted to the problem. This is another advantage of using importance sampling to estimate the target Shapley effects. Finally, as seen in Figure G.8 in APPENDIX G, the preprocessing introduced and described in APPENDIX G.1 seems as well to reduce significantly the error.

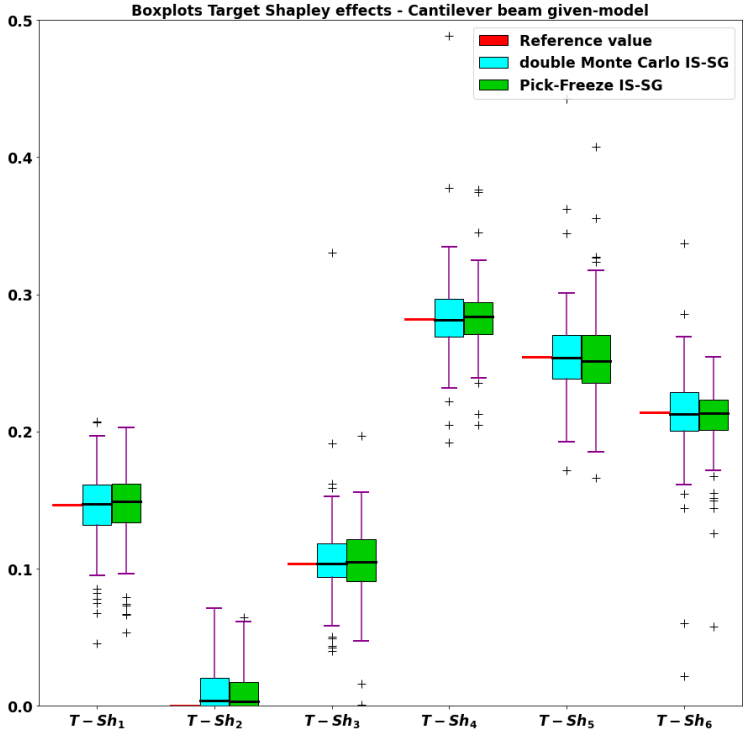

FIG. 4: Estimation of the target Shapley effects in the cantilever beam example, in the given-model framework.

4.3 Fire spread

4.3.1 Presentation of the model

The Rothermel's model introduced in [42] is a semi-physical model which aims at modeling the spread of forest fires. It has two main outputs: the rate of spread of a point in the fire front (R given in $\mathrm{cm} \cdot \mathrm{s}^{-1}$) and the reaction intensity (I_R given in $\mathrm{kW} \cdot \mathrm{m}^{-2} \cdot \mathrm{s}^{-1}$). The initial equations from [42] have been modified several times since their introduction and in the present study, we adopt the point of view of [17,43]. In the same way, we take into account the modifications of [44] on the net fuel loading and the optimum reaction velocity and the modifications of [45] on the moisture damping coefficient and the heat preignition. In addition, we adopt as well the modifications of the marginal input distributions introduced and explained in [17].

The model considered in the present article has 10 input variables grouped in the random input vector $\mathbf{X} = (\delta, \sigma, h, \rho_p, m_l, m_d, S_T, U, \tan\varphi, P)$ and whose physical meanings as well as their marginal distributions are given in Table 3. Moreover, real observations in [46] highlight a negative correlation between the moisture content of dead fuel m_d and the wind speed U, i.e. the windier it is, the less moisture the dead fuels contain. We suppose here that the correlation is strong and that it is specified by the following Pearson correlation coefficient:

$$\rho_{m_d,U} = -0.8. \tag{34}$$

The joint distribution of (m_d, U) is then based on a Gaussian copula. Finally, in order to ensure the physical consistency of the model, the following rules are adopted:

- all negative values of any input variable are rejected
- all values of S_T and P over 1 are rejected

	Input variable	Symbol and unit	Distribution
1	Fuel depth	δ (cm)	LogN $(2.19, 0.517)$
2	Fuel particle area-to-volume ratio	σ (cm^{-1})	LogN $(3.31, 0.294)$
3	Fuel particle low heat content	h ($\mathrm{Kcal} \cdot \mathrm{kg}^{-1}$)	LogN $(8.48, 0.063)$
4	Oven-dry particle density	ρ_p ($\mathrm{D} \cdot \mathrm{W} \cdot \mathrm{g} \cdot \mathrm{cm}^{-3}$)	LogN $(-0.592, 0.219)$
5	Moisture content of the live fuel	m_l ($\mathrm{H_2OgD} \cdot \mathrm{W} \cdot \mathrm{g}^{-1}$)	N $(1.18, 0.377)$
6	Moisture content of the dead fuel	m_d ($\mathrm{H_2OgD} \cdot \mathrm{W} \cdot \mathrm{g}^{-1}$)	N $(0.19, 0.047)$
7	Fuel particle total mineral content	S_T ($\mathrm{MIN} \cdot \mathrm{gD} \cdot \mathrm{W} \cdot \mathrm{g}^{-1}$)	N $(0.049, 0.011)$
8	Wind speed at midflame height	U ($\mathrm{km} \cdot \mathrm{h}^{-1}$)	6.9LogN $(1.0174, 0.5569)$
9	Slope	$\tan \varphi$	N $(0.38, 0.186)$
10	Dead fuel loading to total fuel loading	P	LogN $(-2.19, 0.64)$

TABLE 3: Distributions of each input variable of the fire spread example. D.W.: dry weight - MIN: mineral weight - N (μ, σ): the 1-dimensional normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$ - aLogN (μ, σ): the distribution of $a \exp(A)$ with A a 1-dimensional normal random variable of mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$.

$\text{T-Sh}_1^{\text{ref}}$	0.152	$\text{T-Sh}_2^{\text{ref}}$	0.247	$\text{T-Sh}_3^{\text{ref}}$	0.011	$\text{T-Sh}_4^{\text{ref}}$	0.003	$\text{T-Sh}_5^{\text{ref}}$	0.162
$\text{T-Sh}_6^{\text{ref}}$	0.145	$\text{T-Sh}_7^{\text{ref}}$	0.016	$\text{T-Sh}_8^{\text{ref}}$	0.182	$\text{T-Sh}_9^{\text{ref}}$	0.009	$\text{T-Sh}_{10}^{\text{ref}}$	0.073

TABLE 4: Reference values of the target Shapley effects in the fire spread problem.

- all values of m_d lower than $3/0.6$ are rejected since this value is the smallest possible surface area to volume ratio for fuels with a diameter less than 6mm.

This means that the input distribution is given by the distribution defined by Table 3 and (34), conditioned to the fact that none of the above rules lead to a rejection. In practice, we use truncated distributions in order to handle the inputs numerically. To sum up, given the random input vector $\mathbf{X} = (\delta, \sigma, h, \rho_p, m_l, m_d, S_T,\ U, \tan\varphi, P)$, the rate of spread is obtained through the system of equations in APPENDIX H. Finally, the critical threshold of the rate of spread is arbitrarily set to $t = 60\ \mathrm{cm} \cdot \mathrm{s}^{-1}$.

4.3.2 Reference values

A Monte Carlo estimation with a sample of size $N = 10^7$ gives a reference value for the failure probability: $p_t^{\text{ref}} \approx 1.4 \times 10^{-4}$. Next, since the random input vector $\mathbf{X} = (\delta, \sigma, h, \rho_p, m_l, m_d, S_T, U, \tan\varphi, P)$ is composed of normal and LogNormal random variables, as in the previous example, we transform it into a 10-dimensional Gaussian random vector with the correct mean vector and covariance matrix. At last, reference values of the target Shapley effects are computed using the double Monte Carlo existing estimator in (6) combined with the non-given data sampling procedure described in APPENDIX I with the input distribution and finally, with $N = 10^7$, $N_O = 10^6$ and $N_I = 3$, we obtain the reference values presented in Table 4.

4.3.3 Numerical results

The results of the ROSA of the fire spread problem in the given-data framework with $N_{tot} = 2 \times 10^4$ are given in Figure 5. First, since the failure probability is around $p_t^{\text{ref}} \sim 10^{-4}$, the existing estimators without importance sampling with samples of size $N_{\text{tot}} = 2 \times 10^4$ drawn according to the input distribution return a value very close to 0 almost every time because there are too few failure points. Thus, for the sake of conciseness, it is not worthy to show their boxplots. Second, unreported numerical simulations provide that the failure domain is once more located in one region of the input space and explain why both Gaussian and Gaussian mixture IS auxiliary densities provide similar performances.

Figure 5 shows the practical interest of the previous efforts on a real semi-physical model. Indeed, when the IS auxiliary density is adapted to the problem, the performances of the suggested estimators with importance sampling are satisfying because they have a low variance and a moderate bias. Note nevertheless that the hierarchy of importance of the inputs is not exactly the same for the reference values of the target Shapley effects. In contrast, the existing estimators can not give meaningful results as explained above. Comparing both results presented in Figure 5 and in [17], one can remark that on the fire spread example, the five most influential inputs on the variability of the output R and on the variability of the random variable $1\,(R > 60)$ are the same: the fuel depth δ, the fuel particle area-to-volume ratio σ, the moisture contents of the live and dead fuel m_l and m_d, and the wind speed at midflame height U.

Moreover, this test case highlights the importance of the choice of the size of the inner loop N_I for the double Monte Carlo estimator in the given-data framework when the dimension is getting higher. Indeed, as in the previous examples, we first applied the double Monte Carlo estimator with the value $N_I = 3$ as suggested in [14,17]. However, the estimations were inaccurate: the variance of each estimator was extremely high and the estimation of each target Shapley effect was very often over 10^{40} in absolute value whereas the effect should theoretically lie between 0 and 1. Unreported numerical tests show that this phenomenon is getting even worse when N_I increases. Recalling that N_I is the number of nearest neighbours to find in the given-data framework, the origin of this problem seems to be once more the nearest neighbour approximation, which is getting less accurate when the dimension increases. With importance sampling, the error does not only come from the gap between the target value $\psi_t\left(\mathbf{X}_u^{(n_0)}, \mathbf{X}_{-u}^{(k_N^u(n_0,k))}\right)$ and its approximation $\psi_t\left(\mathbf{X}^{(k_N^u(n_0,k))}\right)$ (for $n_0 \in [\![1, N]\!]$ and $k \in [\![1, N_I]\!]$) but also from the gap between the likelihood ratios evaluated in both points $f_{\mathbf{X}}\left(\mathbf{X}^{(k_N^u(n_0,k))}\right) / g\left(\mathbf{X}^{(k_N^u(n_0,k))}\right)$ and $f_{\mathbf{X}}\left(\mathbf{X}_u^{(n_0)}, \mathbf{X}_{-u}^{(k_N^u(n_0,k))}\right) \Big/ g\left(\mathbf{X}_u^{(n_0)}, \mathbf{X}_{-u}^{(k_N^u(n_0,k))}\right)$, which might become extremely large when the approximation is not accurate. To decrease this error, we hence decided to reduce the size of the inner loop to $N_I = 2$ such that the double Monte Carlo algorithm has to find as many neighbours as in the Pick-Freeze algorithm, and the corresponding results presented in Figure 5 are much more satisfying.

In order to evaluate the error due to the nearest neighbour approximation, Figure 6 represents the estimated ROSA indices for the fire spread problem in the given-model framework with $N_{tot} = 2 \times 10^4$ and the numerical parameters from the beginning of Section 4. In contrast to the given-data results presented in Figure 5, the reference value of each target Shapley effect is included in its corresponding boxplots and the boxplot medians are close to the reference value. The dispersion of the boxplots is nevertheless relatively large. Figure 7 provides a deeper analysis and represents graphically the performances of the given-data and given-model estimators with importance sampling with samples of size $N_{tot} = 10^5$, with $N_{\mathbb{V}} = 10^4$ and N_O as specified in the beginning of Section 4 using the new value of N_{tot} in the given-model framework. The given-model estimators provide the same hierarchy of importance of the inputs as the reference values and for each index, the boxplot medians are almost centered on the reference value, with moderate dispersion. In contrast, even if the bias of the given-data estimators seems to be smaller with larger samples, there is still a gap between the boxplot medians and the reference values, and the hierarchy of importance of the inputs provided is not exactly the same as the reference values. These observations reaffirm on an example in dimension 10 the effect of the nearest neighbour approximation on the given-data estimators for the estimation of the target Shapley effects.

Remark 5. *In all the figures, one can remark that there is no benefit to use the Pick-Freeze estimators instead of the double Monte Carlo estimators. This fact can be counter-intuitive knowing what happens in the independent case: the authors of [47] proved that some Pick-Freeze estimators are asymptotically efficient. A similar remark has already been made in [14], which can also be applied to our case. They highlighted first that the authors of [47] estimate the variance of the output Y in their procedure to estimate each Sobol indices, and second that the double Monte Carlo estimators are based on different observations from the Pick-Freeze estimators.*

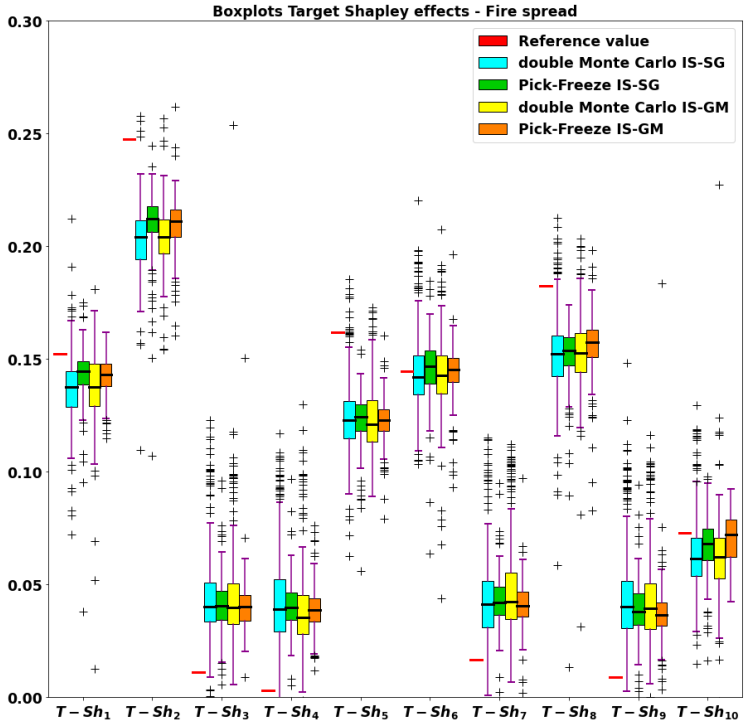

FIG. 5: Estimation of the target Shapley effects in the fire spread example, in the given-data framework, with the preprocessing described in APPENDIX G.1 and with $N_I = 2$.

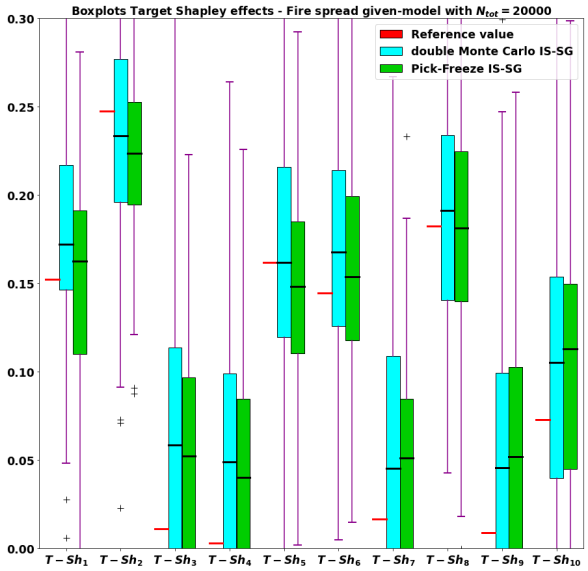

FIG. 6: Estimation of the target Shapley effects in the fire spread example, in the given-model framework with $N_{tot} = 2 \times 10^4$.

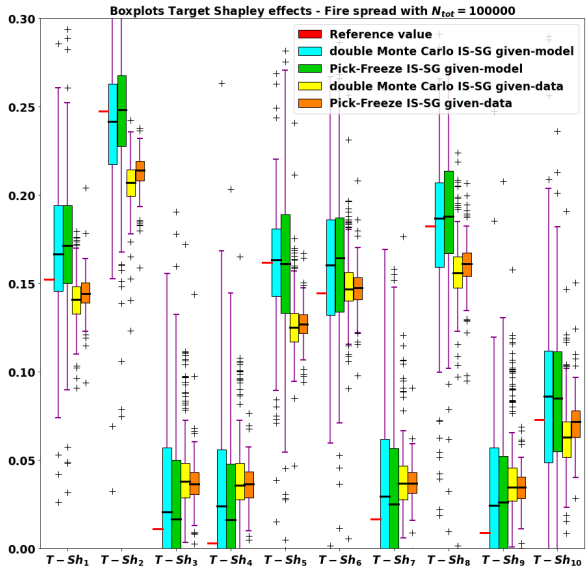

FIG. 7: Estimation of the target Shapley effects in the fire spread example, in both given-model and given-data frameworks with $N_{tot} = 10^5$.

5. CONCLUSION

In the present article, we are interested in the estimation of the target Shapley effects, whose goal is to quantify the influence of each input variable on the occurrence of the failure of the system and which are able to handle correlated inputs. We suggest new importance-sampling-based estimators of the target Shapley effects, extending the previous works of [13,14], which are more efficient than the existing ones when the failure probability is low. Moreover, we also introduce less expensive importance-sampling-based estimators requiring only an i.i.d. input/output N-sample distributed according to the IS auxiliary distribution, which enable to estimate efficiently the target Shapley effects without additional calls to the function ϕ after the estimation of the failure probability by importance sampling. In addition, we show theoretically that using the optimal IS auxiliary distribution for estimating a failure probability by importance sampling as the IS auxiliary distribution in the Pick-Freeze estimator improves the estimation of the target Shapley effects in comparison to the existing Pick-Freeze estimators. This result has a massive practical advantage because it justifies that it is beneficial to reuse the available sample from the reliability analysis to estimate the target Shapley effects by importance sampling. Finally, we illustrate and discuss the practical interest of the proposed estimators on the Gaussian linear case and on two real physical examples, all involving correlated inputs.

The main perspective for improvement of the suggested approach is to make the proposed estimators more robust faced with the dimension. Indeed, as explained and illustrated in the article, the nearest neighbour approximation creates an error which is getting larger when the dimension increases. The preprocessing procedure introduced in APPENDIX G.1 seems to reduce this error, but potentially not enough so in higher dimension. Note also that in the highest dimension considered here, $d = 10$ in Figure 5, the performances provided by the nearest neighbour approximation are not as good as in the other settings. This classic problem in the analysis of a complex system is called the *curse of dimensionality*. Nevertheless, new approaches based on projected random forests [23] could improve the estimation of the target Shapley effects when the dimension increases. It might be possible to adapt the proposed method to our framework by taking into account the weights from importance sampling in the construction of the random forests. One can also mention recent projection methods [48] to reduce the dimension of the problem.

Finally, it could be interesting to inspect methods to estimate the target Shapley effects efficiently while building a surrogate-model with importance sampling, in the same way as the method presented in [49]. At last, subset simulation could be used to estimate efficiently the target Shapley effects when the failure probability is very small instead of importance sampling. It may be done in low dimension by extending the work presented in [9] which aims to estimate the first and total orders target Sobol indices with a failure sample obtained by subset simulation.

ACKNOWLEDGMENTS

The first author is enrolled in a Ph.D. program co-funded by *ONERA – The French Aerospace Lab* and *Toulouse III - Paul Sabatier University*. Their financial supports are gratefully acknowledged.

APPENDIX A. PROOF OF LEMMA 1

First of all, let us remark that the convention $0/0 = 0$ introduced and adopted at the beginning of Section 3 prevents the following proofs from any possible problem caused by a division by 0 or caused by a non-definition of any conditional PDF.

For any IS auxiliary density $g : \mathbb{X} \longrightarrow \mathbb{R}_+$, we have:

$$\begin{aligned}
\mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right] &= \int_{\mathbb{X}_{-u}} \mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}=\mathbf{x}_{-u}\right)^2 f_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right) d\mathbf{x}_{-u} \\
&= \int_{\mathbb{X}_{-u}} \left(\int_{\mathbb{X}_u} \psi_t\left(\mathbf{x}_u,\mathbf{x}_{-u}\right) f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}\left(\mathbf{x}_u\right) d\mathbf{x}_u\right)^2 f_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right) d\mathbf{x}_{-u} \\
&= \int_{\mathbb{X}_{-u}} \left(\int_{\mathbb{X}_u} \psi_t\left(\mathbf{x}_u,\mathbf{x}_{-u}\right) \frac{f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}\left(\mathbf{x}_u\right)}{g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}\left(\mathbf{x}_u\right)} g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}\left(\mathbf{x}_u\right) d\mathbf{x}_u\right)^2 \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right)} g_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right) d\mathbf{x}_{-u} \\
&= \int_{\mathbb{X}_{-u}} \left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}_u|\mathbf{X}_{-u}}\left(\mathbf{X}_u\right)}{g_{\mathbf{X}_u|\mathbf{X}_{-u}}\left(\mathbf{X}_u\right)} \middle| \mathbf{X}_{-u}=\mathbf{x}_{-u}\right)\right]^2 \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right)} g_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right) d\mathbf{x}_{-u} \\
&= \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}_u|\mathbf{X}_{-u}}\left(\mathbf{X}_u\right)}{g_{\mathbf{X}_u|\mathbf{X}_{-u}}\left(\mathbf{X}_u\right)} \middle| \mathbf{X}_{-u}\right)^2 \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right].
\end{aligned}$$

Then, using the definition of the conditional PDF, remark that:

$$\forall\left(\mathbf{x}_u,\mathbf{x}_{-u}\right) \in \mathbb{X}_u \times \mathbb{X}_{-u}, \begin{cases} f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}\left(\mathbf{x}_u\right) &= \dfrac{f_{\mathbf{X}}\left(\mathbf{x}_u,\mathbf{x}_{-u}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right)} \\ g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}\left(\mathbf{x}_u\right) &= \dfrac{g\left(\mathbf{x}_u,\mathbf{x}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{x}_{-u}\right)}. \end{cases}$$

By replacing in the above, we have:

$$\begin{aligned}
\mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right] &= \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}_u|\mathbf{X}_{-u}}\left(\mathbf{X}_u\right)}{g_{\mathbf{X}_u|\mathbf{X}_{-u}}\left(\mathbf{X}_u\right)} \middle| \mathbf{X}_{-u}\right)^2 \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right] \\
&= \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}}\left(\mathbf{X}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)} \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{g\left(\mathbf{X}\right)} \middle| \mathbf{X}_{-u}\right)^2 \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right] \\
&= \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}}\left(\mathbf{X}\right)}{g\left(\mathbf{X}\right)} \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)} \middle| \mathbf{X}_{-u}\right)^2 \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right] \\
&= \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}}\left(\mathbf{X}\right)}{g\left(\mathbf{X}\right)} \middle| \mathbf{X}_{-u}\right)^2 \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)^2}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)^2} \frac{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right] \\
&= \mathbb{E}_g\left[\mathbb{E}_g\left(\psi_t(\mathbf{X}) \frac{f_{\mathbf{X}}\left(\mathbf{X}\right)}{g\left(\mathbf{X}\right)} \middle| \mathbf{X}_{-u}\right)^2 \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right] \\
&= \mathbb{E}_g\left[\mathbb{E}_g\left(w_t^g(\mathbf{X}) \mid \mathbf{X}_{-u}\right)^2 \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)}\right].
\end{aligned}$$

That concludes the proof of Lemma 1. □

APPENDIX B. PROOF OF PROPOSITION 1

First of all, recall that for $n \in [\![1, N_u]\!]$, the estimator $\overline{\psi_t^{\mathrm{IS}}(\mathbf{X}_{-u}^{(n)})}$ in (18) is given by:

$$\overline{\psi_t^{\mathrm{IS}}(\mathbf{X}_{-u}^{(n)})} = \frac{1}{N_I} \sum_{k=1}^{N_I} w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right). \tag{B.1}$$

Then, let us write $\widehat{E}^{\text{IS}}_{u,\text{MC}}$ for the uncorrected double Monte Carlo estimator of the term $\mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right] = \mathbb{E}_g\left[\mathbb{E}_g\left(w_t^g(\mathbf{X})\,|\,\mathbf{X}_{-u}\right)^2 g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)/f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)\right]$, obtained by removing $\widehat{E}^{\text{IS}}_{\text{bias},u}$ in (17):

$$\widehat{E}^{\text{IS}}_{u,\text{MC}} = \frac{1}{N_u}\sum_{n=1}^{N_u}\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2 \frac{g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(n)}\right)}{f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}^{(n)}\right)}. \tag{B.2}$$

Then, this proof can be divided into two steps:

1. compute the bias of the estimator $\widehat{E}^{\text{IS}}_{u,\text{MC}}$ in (B.2)

2. propose an unbiased estimator of the latter bias in order to correct it.

APPENDIX B.1 Bias in the inner loop

First, for a given sample $\mathbf{X}_{-u}^{(n)} \in \mathbb{X}_{-u}$, let us compute the bias of the estimator $\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2$ in (18):

$$\begin{aligned}
&\mathbb{E}_g\left(\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2 \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right) = \mathbb{E}_g\left(\left(\frac{1}{N_I}\sum_{k=1}^{N_I} w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right)\right)^2 \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)\\
&= \frac{1}{N_I^2}\sum_{k=1}^{N_I}\mathbb{E}_g\left(w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right)^2 \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)\\
&\quad + \frac{1}{N_I^2}\sum_{1\le i\neq j\le N_I}\mathbb{E}_g\left(w_t^g\left(\mathbf{X}_u^{(n,i)}, \mathbf{X}_{-u}^{(n)}\right) \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)\times\mathbb{E}_g\left(w_t^g\left(\mathbf{X}_u^{(n,j)}, \mathbf{X}_{-u}^{(n)}\right) \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)\\
&= \frac{1}{N_I}\mathbb{E}_g\left[w_t^g\left(\mathbf{X}_u, \mathbf{X}_{-u}^{(n)}\right)^2 \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right] + \frac{N_I-1}{N_I}\mathbb{E}_g\left[w_t^g\left(\mathbf{X}_u, \mathbf{X}_{-u}^{(n)}\right) \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right]^2\\
&= \mathbb{E}_g\left[w_t^g\left(\mathbf{X}_u, \mathbf{X}_{-u}^{(n)}\right) \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right]^2\\
&\quad + \frac{1}{N_I}\mathbb{E}_g\left[w_t^g\left(\mathbf{X}_u, \mathbf{X}_{-u}^{(n)}\right)^2 \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right] - \frac{1}{N_I}\mathbb{E}_g\left[w_t^g\left(\mathbf{X}_u, \mathbf{X}_{-u}^{(n)}\right) \middle| \mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right]^2\\
&= \mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)|\,\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)^2 + \frac{1}{N_I}\mathbb{V}_g\left(w_t^g\left(\mathbf{X}\right)|\,\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)\\
&= \mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)|\,\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)^2 + \mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}|\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right).
\end{aligned}$$

The bias of the estimator $\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2$ is thus $\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}|\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)$.

APPENDIX B.2 Bias of the outer loop

Second, let us derive the bias of the uncorrected double Monte Carlo estimator $\widehat{E}^{\text{IS}}_{u,\text{MC}}$ in (B.2) of the double expectation $\mathbb{E}_g\left[\mathbb{E}_g\left(w_t^g(\mathbf{X})\mid\mathbf{X}_{-u}\right)^2 g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)/f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)\right]$ composed of both inner and outer loops:

$$
\begin{aligned}
\mathbb{E}_g\left(\widehat{E}^{\text{IS}}_{u,\text{MC}}\right) &= \frac{1}{N_u}\sum_{n=1}^{N_u}\mathbb{E}_g\left(\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u}^{(n)})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u}^{(n)})}\right)\\
&= \frac{1}{N_u}\sum_{n=1}^{N_u}\mathbb{E}_g\left[\underbrace{\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u}^{(n)})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u}^{(n)})}\mathbb{E}_g\left(\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}\right)^2\middle|\,\mathbf{X}_{-u}=\mathbf{X}_{-u}^{(n)}\right)}_{\text{function of }\mathbf{X}_{-u}^{(n)}}\right]\\
&= \mathbb{E}_g\left[\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\mathbb{E}_g\left(\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}\right)^2\middle|\,\mathbf{X}_{-u}\right)\right]\qquad\text{where }\overline{\psi_t^{\text{IS}}}\text{ is as in (20)}\\
&= \mathbb{E}_g\left[\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\left(\mathbb{E}_g\left(w_t^g(\mathbf{X})|\,\mathbf{X}_{-u}\right)^2+\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}|\mathbf{X}_{-u}\right)\right)\right]\\
&= \mathbb{E}_g\left[\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\mathbb{E}_g\left(w_t^g(\mathbf{X})|\,\mathbf{X}_{-u}\right)^2\right]+\mathbb{E}_g\left[\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}|\mathbf{X}_{-u}\right)\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\right]\\
&= \mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_{-u}\right)^2\right]+\mathbb{E}_g\left[\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}|\mathbf{X}_{-u}\right)\frac{g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})}\right]\\
&\qquad\qquad\text{thanks to Lemma 1.}
\end{aligned}
$$

Therefore, the bias of $\widehat{E}^{\text{IS}}_{u,\text{MC}}$ is $\mathbb{E}_g\left[\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}|\mathbf{X}_{-u}\right)g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})/f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})\right]$. The problem is now to estimate it in order to propose an unbiased estimator of T-EV$_u$ by double Monte Carlo with importance sampling.

APPENDIX B.3 Estimation of the bias

To estimate the previous bias, let us before prove the following lemma.

Lemma 3. *Let $(Z_n)_{n\in[\![1,N]\!]}$ be a sequence of independent and identically distributed random variables such that $\mathbb{E}\left(Z_1^2\right)<+\infty$. Let us consider the empirical estimator $\widehat{Z}_N=N^{-1}\sum_{n=1}^N Z_n$ of the mean value of Z_1. Then:*

$$
\widehat{V}_Z=\frac{1}{N-1}\left[\frac{1}{N}\sum_{n=1}^N Z_n^2-\widehat{Z}_N^2\right] \tag{B.3}
$$

is an unbiased estimator of $\mathbb{V}\left(\widehat{Z}_N\right)$.

Proof. Let us compute the expectation of the estimator $\widehat{V}_Z$:

$$\begin{aligned}
\mathbb{E}\left(\widehat{V}_Z\right) &= \mathbb{E}\left(\frac{1}{N-1}\left[\frac{1}{N}\sum_{n=1}^{N} Z_n^2 - \widehat{Z}_N^2\right]\right) \\
&= \frac{1}{N-1}\left[\mathbb{E}\left(Z^2\right) - \mathbb{E}\left(\widehat{Z}_N^2\right)\right] \\
&= \frac{1}{N-1}\left[\mathbb{E}\left(Z^2\right) - \mathbb{E}\left(\frac{1}{N^2}\sum_{n=1}^{N} Z_n^2 + \frac{1}{N^2}\sum_{1\leq i\neq j\leq N} Z_i Z_j\right)\right] \\
&= \frac{1}{N-1}\left[\mathbb{E}\left(Z^2\right) - \frac{1}{N}\mathbb{E}\left(Z^2\right) - \frac{N-1}{N}\mathbb{E}\left(Z\right)^2\right] \\
&= \frac{1}{N-1}\left[\frac{N-1}{N}\left(\mathbb{E}\left(Z^2\right) - \mathbb{E}\left(Z\right)^2\right)\right] \\
&= \frac{N-1}{N(N-1)}\mathbb{V}(Z) \\
&= \frac{1}{N}\mathbb{V}(Z) \\
&= \mathbb{V}(\widehat{Z}_N).
\end{aligned}$$

That concludes the proof of this lemma. □

By applying the previous lemma to the i.i.d. sequence $\left(w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right)\right)_{k\in[\![1,N_I]\!]}$ for all $n \in [\![1, N_u]\!]$, the estimator $(N_I-1)^{-1}\left[N_I^{-1}\sum_{k=1}^{N_I} w_t^g\left(\mathbf{X}_u^{(n,k)}, \mathbf{X}_{-u}^{(n)}\right)^2 - \left(\overline{\psi_t^{\text{IS}}\left(\mathbf{X}_{-u}^{(n)}\right)}\right)^2\right]$ estimates without bias the conditional variance $\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u}^{(n)})}|\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}\right)$. Then, since the outer loop is only an empirical mean, $\widehat{E}_{\text{bias},u}^{\text{IS}}$ in (19) is therefore an unbiased estimator of the bias $\mathbb{E}_g\left[\mathbb{V}_g\left(\overline{\psi_t^{\text{IS}}(\mathbf{X}_{-u})}|\mathbf{X}_{-u}\right) g_{\mathbf{X}_{-u}}(\mathbf{X}_{-u}) \,/\, f_{\mathbf{X}_{-u}}(\mathbf{X}_{-u})\right]$ and allows to correct the bias created by the square in the inner loop. Finally, by the linearity of the expectation, $\widehat{E}_{u,\text{MC}}^{\text{IS}} - \widehat{E}_{\text{bias},u}^{\text{IS}}$ is an unbiased estimator of $\mathbb{E}_g\left[\mathbb{E}_g\left(w_t^g(\mathbf{X}) \mid \mathbf{X}_{-u}\right)^2 g_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right) / f_{\mathbf{X}_{-u}}\left(\mathbf{X}_{-u}\right)\right]$, and thus $\widehat{\text{T-EV}}_{u,\text{MC}}^{\text{IS}}$ in (17) is an unbiased estimator by double Monte Carlo with importance sampling of T-EV$_u$. That concludes the proof of Proposition 1. □

APPENDIX C. PROOF THAT THE ESTIMATOR IN EQ. (25) IS AN UNBIASED ESTIMATOR OF P_T^2

Let us compute the bias of the estimator $\left(\widehat{p}_{t,N}^{\mathrm{IS}}\right)^2$ in (25):

$$\begin{aligned}
\mathbb{E}_g\left(\left(\widehat{p}_{t,N}^{\mathrm{IS}}\right)^2\right) &= \mathbb{E}_g\left(\left(\frac{1}{N}\sum_{n=1}^{N} w_t^g\left(\mathbf{X}^{(n)}\right)\right)^2\right) \\
&= \frac{1}{N^2}\mathbb{E}_g\left(\sum_{n=1}^{N} w_t^g\left(\mathbf{X}^{(n)}\right)^2\right) + \frac{1}{N^2}\mathbb{E}_g\left(\sum_{1\le i\neq j\le N} w_t^g\left(\mathbf{X}^{(i)}\right) w_t^g\left(\mathbf{X}^{(j)}\right)\right) \\
&= \frac{N}{N^2}\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)^2\right) + \frac{N(N-1)}{N^2}\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)\right)\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)\right) \\
&= \frac{1}{N}\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)^2\right) + \frac{N-1}{N}\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)\right)^2 \\
&= \mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)\right)^2 + \frac{1}{N}\left(\mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)^2\right) - \mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)\right)^2\right) \\
&= \mathbb{E}_g\left(w_t^g\left(\mathbf{X}\right)\right)^2 + \frac{1}{N}\mathbb{V}_g\left(w_t^g\left(\mathbf{X}\right)\right) \\
&= p_t^2 + \mathbb{V}_g\left(\widehat{p}_{t,N}^{\mathrm{IS}}\right).
\end{aligned}$$

Then, Lemma 3 justifies that $(N-1)^{-1}\left[N^{-1}\sum_{n=1}^{N} w_t^g\left(\mathbf{X}^{(n)}\right)^2 - \left(\widehat{p}_{t,N}^{\mathrm{IS}}\right)^2\right]$ is an unbiased estimator of $\mathbb{V}_g\left(\widehat{p}_{t,N}^{\mathrm{IS}}\right)$. Therefore, by the linearity of the expectation, $\widehat{p}_{t,N}^{\mathrm{IS,ub}}$ in (25) is an unbiased estimator of p_t^2. □

APPENDIX D. PROOF OF LEMMA 2

To begin with, let us prove the following lemma.

Lemma 4. *The PDF of the joint distribution of the random vector* $\left(\mathbf{X}_u, \mathbf{X}_{-u}, \mathbf{X}'_{-u}\right)$ *satisfies for all* $\mathbf{x}_u, \mathbf{x}_{-u}, \mathbf{x}'_{-u} \in \mathbb{X}_u \times \mathbb{X}_{-u} \times \mathbb{X}_{-u}$*:*

$$f_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u},\mathbf{x}'_{-u}) = f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u})\frac{f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}'_{-u})}{f_{\mathbf{X}_u}(\mathbf{x}_u)}. \tag{D.1}$$

Proof. Let $\mathbf{x}_u, \mathbf{x}_{-u}, \mathbf{x}'_{-u} \in \mathbb{X}_u \times \mathbb{X}_{-u} \times \mathbb{X}_{-u}$. Then, regardless whether $\mathbf{x}_u$ is in the support of $f_{\mathbf{X}_u}$ or not, the convention $0/0 = 0$ allows to write:

$$\begin{aligned}
f_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u},\mathbf{x}'_{-u}) &= f_{\mathbf{X}_u}(\mathbf{x}_u) f_{\mathbf{X}_{-u},\mathbf{X}'_{-u}|\mathbf{X_u}=\mathbf{x}_u}(\mathbf{x}_{-u},\mathbf{x}'_{-u}) \\
&= f_{\mathbf{X}_u}(\mathbf{x}_u) f_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u}(\mathbf{x}_{-u}) f_{\mathbf{X}'_{-u}|\mathbf{X}_u=\mathbf{x}_u}(\mathbf{x}'_{-u}) \\
&= f_{\mathbf{X}_u}(\mathbf{x}_u)\frac{f_{\mathbf{X}_u,\mathbf{X}_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u})}{f_{\mathbf{X}_u}(\mathbf{x}_u)}\frac{f_{\mathbf{X}_u,\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}'_{-u})}{f_{\mathbf{X}_u}(\mathbf{x}_u)} \\
&= f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u})\frac{f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}'_{-u})}{f_{\mathbf{X}_u}(\mathbf{x}_u)}.
\end{aligned}$$

That concludes the proof of the lemma. □

Then, by remarking that the term in the expectation $\mathbb{E}_f\left(\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right)$ is a function of the three random variables $\mathbf{X}_u$, $\mathbf{X}_{-u}$ and $\mathbf{X}'_{-u}$ with the correlation structure described above, we have:

$$
\begin{aligned}
&\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right) = \mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t\left(\mathbf{X}_u,\mathbf{X}_{-u}\right)\psi_t\left(\mathbf{X}_u,\mathbf{X}'_{-u}\right)\right)\\
&= \int_{\mathbb{X}_u}\int_{\mathbb{X}_{-u}}\int_{\mathbb{X}_{-u}} \psi_t(\mathbf{x}_u,\mathbf{x}_{-u})\psi_t(\mathbf{x}_u,\mathbf{x}'_{-u}) f_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u},\mathbf{x}'_{-u}) d\mathbf{x}'_{-u} d\mathbf{x}_{-u} d\mathbf{x}_u\\
&= \int_{\mathbb{X}_u}\int_{\mathbb{X}_{-u}}\int_{\mathbb{X}_{-u}} \psi_t(\mathbf{x}_u,\mathbf{x}_{-u})\psi_t(\mathbf{x}_u,\mathbf{x}'_{-u}) \frac{f_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u},\mathbf{x}'_{-u})}{g_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u},\mathbf{x}'_{-u})}\\
&\qquad\qquad g_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{x}_u,\mathbf{x}_{-u},\mathbf{x}'_{-u}) d\mathbf{x}'_{-u} d\mathbf{x}_{-u} d\mathbf{x}_u\\
&= \mathbb{E}_g\left(\psi_t\left(\mathbf{X}_u,\mathbf{X}_{-u}\right)\psi_t\left(\mathbf{X}_u,\mathbf{X}'_{-u}\right)\frac{f_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u})}{g_{\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u}}(\mathbf{X}_u,\mathbf{X}_{-u},\mathbf{X}'_{-u})}\right)\\
&= \mathbb{E}_g\left(\psi_t\left(\mathbf{X}_u,\mathbf{X}_{-u}\right)\psi_t\left(\mathbf{X}_u,\mathbf{X}'_{-u}\right)\frac{f_{\mathbf{X}}(\mathbf{X}_u,\mathbf{X}_{-u})\dfrac{f_{\mathbf{X}}(\mathbf{X}_u,\mathbf{X}'_{-u})}{f_{\mathbf{X}_u}(\mathbf{X}_u)}}{g(\mathbf{X}_u,\mathbf{X}_{-u})\dfrac{g(\mathbf{X}_u,\mathbf{X}'_{-u})}{g_{\mathbf{X}_u}(\mathbf{X}_u)}}\right)\\
&= \mathbb{E}_g\left(\psi_t\left(\mathbf{X}\right)\psi_t\left(\mathbf{X}^u\right)\frac{f_{\mathbf{X}}(\mathbf{X})\dfrac{f_{\mathbf{X}}(\mathbf{X}^u)}{f_{\mathbf{X}_u}(\mathbf{X}_u)}}{g(\mathbf{X})\dfrac{g(\mathbf{X}^u)}{g_{\mathbf{X}_u}(\mathbf{X}_u)}}\right)\\
&= \mathbb{E}_g\left(\psi_t\left(\mathbf{X}\right)\psi_t\left(\mathbf{X}^u\right)\frac{f_{\mathbf{X}}(\mathbf{X})f_{\mathbf{X}}(\mathbf{X}^u)g_{\mathbf{X}_u}(\mathbf{X}_u)}{g(\mathbf{X})g(\mathbf{X}^u)f_{\mathbf{X}_u}(\mathbf{X}_u)}\right).
\end{aligned}
$$

That concludes the proof of the Lemma 2. □

APPENDIX E. PROOFS OF INEQUALITIES OF SECTION 3.3

APPENDIX E.1 Proof of inequality (29)

First, recall that g_{opt} defined for all $\mathbf{x} \in \mathbb{X}$ by $g_{\text{opt}}(\mathbf{x}) = p_t^{-1}\psi_t(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})$ is the optimal IS auxiliary density to estimate the failure probability by importance sampling, and its marginal PDF according to $\mathbf{X}_u$ is given by:

$$
\forall \mathbf{x}_u \in \mathbb{X}_u,\ g_{\text{opt}_{\mathbf{X}_u}}(\mathbf{x}_u) = \frac{1}{p_t}\int_{\mathbb{X}_{-u}} f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u})\,\psi_t(\mathbf{x}_u,\mathbf{x}_{-u})\,d\mathbf{x}_{-u}. \tag{E.1}
$$

Therefore, when the considered IS auxiliary distribution is g_{opt}, the variance of the Pick-Freeze given-model estimator with importance sampling $\widehat{\text{T-VE}}^{\text{IS}}_{u,\text{PF}}$ in (27) of T-VE$_u$ satisfies:

$$
\begin{aligned}
\mathbb{V}_{g_{\text{opt}}}\left(\widehat{\text{T-VE}}^{\text{IS}}_{u,\text{PF}}\right) &= \mathbb{V}_{g_{\text{opt}}}\left(\frac{1}{N_u}\sum_{n=1}^{N_u} w_t^{g_{\text{opt}}}\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right) w_t^{g_{\text{opt}}}\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right)\frac{g_{\text{opt}_{\mathbf{X}_u}}(\mathbf{X}_u^{(n)})}{f_{\mathbf{X}_u}(\mathbf{X}_u^{(n)})} - \underbrace{\widehat{p}^{\text{IS,ub}}_{t,N}}_{p_t}\right)\\
&= \mathbb{V}_{g_{\text{opt}}}\left(\frac{1}{N_u}\sum_{n=1}^{N_u} w_t^{g_{\text{opt}}}\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right) w_t^{g_{\text{opt}}}\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right)\frac{g_{\text{opt}_{\mathbf{X}_u}}(\mathbf{X}_u^{(n)})}{f_{\mathbf{X}_u}(\mathbf{X}_u^{(n)})}\right)\\
&= \frac{1}{N_u}\mathbb{V}_{g_{\text{opt}}}\left(w_t^{g_{\text{opt}}}(\mathbf{X})\, w_t^{g_{\text{opt}}}(\mathbf{X}^{\mathbf{u}})\,\frac{g_{\text{opt}_{\mathbf{X}_u}}(\mathbf{X}_u)}{f_{\mathbf{X}_u}(\mathbf{X}_u)}\right)\\
&= \frac{1}{N_u}\mathbb{V}_{g_{\text{opt}}}\left(\psi_t(\mathbf{X})\,\psi_t(\mathbf{X}^u)\,\frac{f_{\mathbf{X}}(\mathbf{X})f_{\mathbf{X}}(\mathbf{X}^u)g_{\text{opt}_{\mathbf{X}_u}}(\mathbf{X}_u)}{g_{\text{opt}}(\mathbf{X})g_{\text{opt}}(\mathbf{X}^u)f_{\mathbf{X}_u}(\mathbf{X}_u)}\right)\\
&= \frac{1}{N_u}\mathbb{V}_{g_{\text{opt}}}\left(p_t^2 \times \frac{1}{p_t}\frac{\int_{\mathbb{X}_{-u}} f_{\mathbf{X}}(\mathbf{X}_u,\mathbf{x}_{-u})\psi_t(\mathbf{X}_u,\mathbf{x}_{-u})\,d\mathbf{x}_{-u}}{f_{\mathbf{X}_u}(\mathbf{X}_u)}\right)\\
&\qquad\qquad \text{by integrating the exact expressions of } g_{\text{opt}} \text{ and } g_{\text{opt}_{\mathbf{X}_u}} \text{ given above}\\
&= \frac{1}{N_u}\mathbb{V}_{g_{\text{opt}}}\left(p_t\int_{\mathbb{X}_{-u}}\frac{f_{\mathbf{X}}(\mathbf{X}_u,\mathbf{x}_{-u})}{f_{\mathbf{X}_u}(\mathbf{X}_u)}\psi_t(\mathbf{X}_u,\mathbf{x}_{-u})\,d\mathbf{x}_{-u}\right)\\
&= \frac{p_t^2}{N_u}\mathbb{V}_{g_{\text{opt}}}\left(\int_{\mathbb{X}_{-u}} f_{\mathbf{X}_{-u}|\mathbf{X}_u}(\mathbf{x}_{-u})\psi_t(\mathbf{X}_u,\mathbf{x}_{-u})\,d\mathbf{x}_{-u}\right)\\
&= \frac{p_t^2}{N_u}\mathbb{V}_{g_{\text{opt}}}\left[\mathbb{E}_f\left(\psi_t(\mathbf{X})\,|\mathbf{X}_u\right)\right]\\
&\leq \frac{p_t^2}{N_u}.
\end{aligned}
$$

That concludes the proof of inequality (29). □

APPENDIX E.2 Proof of inequality (30)

The estimator $\widehat{\text{T-VE}}_{u,\text{PF}}$ by Pick-Freeze given-model without importance sampling of T-VE$_u$ is given by:

$$
\widehat{\text{T-VE}}_{u,\text{PF}} = \frac{1}{N_u}\sum_{n=1}^{N_u}\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right)\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right) - \widehat{p}^2_{t,N}, \tag{E.2}
$$

where $\widehat{p}_{t,N}$ is the empirical Monte Carlo estimator of p_t and where the required samples are drawn according to the input distribution $f_{\mathbf{X}}$. In the given-model framework, independent samples are used to estimate the Pick-Freeze

expectation $\mathbb{E}_{f_{\mathbf{X}}}\left[\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right]$ and the square failure probability p_t^2. Therefore, the variance of $\widehat{\text{T-VE}}_{u,\text{PF}}$ satisfies:

$$\begin{aligned}
\mathbb{V}_{f_{\mathbf{X}}}\left(\widehat{\text{T-VE}}_{u,\text{PF}}\right) &= \mathbb{V}_{f_{\mathbf{X}}}\left(\frac{1}{N_u}\sum_{n=1}^{N_u}\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right)\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right) - \widehat{p}_{t,N}^2\right) \\
&= \mathbb{V}_{f_{\mathbf{X}}}\left(\frac{1}{N_u}\sum_{n=1}^{N_u}\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right)\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right)\right) + \mathbb{V}_{f_{\mathbf{X}}}\left(\widehat{p}_{t,N}^2\right) \\
&\geq \mathbb{V}_{f_{\mathbf{X}}}\left(\frac{1}{N_u}\sum_{n=1}^{N_u}\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right)\psi_t\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right)\right) \\
&= \frac{1}{N_u}\mathbb{V}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right) \\
&= \frac{1}{N_u}\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right) - \frac{1}{N_u}\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})\psi_t(\mathbf{X}^u)\right)^2 \\
&= \frac{1}{N_u}\left(\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] + p_t^2\right) - \frac{1}{N_u}\left(\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] + p_t^2\right)^2 \\
&\qquad\qquad \text{thanks to (24)} \\
&= \frac{1}{N_u}\left(\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] + p_t^2\right)\left(1 - \left(\mathbb{V}_{f_{\mathbf{X}}}\left[\mathbb{E}_{f_{\mathbf{X}}}\left(\psi_t(\mathbf{X})|\mathbf{X}_u\right)\right] + p_t^2\right)\right).
\end{aligned}$$

That concludes the proof of inequality (30). □

APPENDIX F. THEORETICAL VALUES OF THE TARGET SHAPLEY EFFECTS IN THE GAUSSIAN LINEAR FRAMEWORK

Let us consider the Gaussian linear framework introduced in Section 4.1, and a failure threshold $t \in \mathbb{R}$. Recall that the input covariance matrix is symmetric positive-definite and that $\boldsymbol{\beta} \neq 0$, thus we have $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} > 0$.

Moreover, let us recall the following theorem:

Theorem 6. *For $k \geq 1$, if $\mathbf{A} \in \mathcal{M}_{k,d}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^k$ and $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then:*

$$\mathbf{A}\mathbf{X} + \mathbf{b} \sim \mathcal{N}_k\left(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\right). \tag{F.1}$$

APPENDIX F.1 Theoretical value of the failure probability

Theorem 7. *The failure probability is given by:*

$$p_t^{\boldsymbol{\beta}} = 1 - \Phi\left(\frac{t - \boldsymbol{\beta}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}}\right). \tag{F.2}$$

Proof. In the Gaussian linear framework, the failure probability satisfies:

$$p_t^{\boldsymbol{\beta}} = \mathbb{P}\left(\phi_{\boldsymbol{\beta}}(\mathbf{X}) > t\right) = \mathbb{P}\left(\boldsymbol{\beta}^\top \mathbf{X} > t\right) = \mathbb{P}\left(\frac{\boldsymbol{\beta}^\top \mathbf{X} - \boldsymbol{\beta}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}} > \frac{t - \boldsymbol{\beta}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}}\right), \tag{F.3}$$

with $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} > 0$. Then, Theorem 6 provides that $\left(\boldsymbol{\beta}^\top \mathbf{X} - \boldsymbol{\beta}^\top \boldsymbol{\mu}\right) \Big/ \sqrt{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}} \sim \mathcal{N}_1(0,1)$. Finally, we have:

$$p_t^{\boldsymbol{\beta}} = 1 - \Phi\left(\frac{t - \boldsymbol{\beta}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}}\right), \tag{F.4}$$

where Φ is the CDF of the 1-dimensional standard Normal distribution. □

APPENDIX F.2 Theoretical values of the target closed Sobol indices

Theorem 8. *For $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, the target closed Sobol index is given by:*

$$\textit{T-VE}_u = \begin{cases} \mathbb{V}\left[\Phi\left(\dfrac{t-\boldsymbol{\beta}_u^\top \mathbf{X}_u - \boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\mu}_{-u}+\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\left(\mathbf{X}_u-\boldsymbol{\mu}_u\right)\right)}{\sqrt{\boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\Sigma}_{-u,-u}-\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\boldsymbol{\Sigma}_{u,-u}\right)\boldsymbol{\beta}_{-u}}}\right)\right] & \textit{if } \boldsymbol{\beta}_{-u} \neq 0 \\ p_t^{\boldsymbol{\beta}}\left(1-p_t^{\boldsymbol{\beta}}\right) & \textit{else,} \end{cases} \tag{F.5}$$

where for $u_1, u_2 \in \mathcal{P}(d)$, $\boldsymbol{\Sigma}_{u_1,u_2} = (\Sigma_{i,j})_{i\in u_1, j\in u_2}$. At last, using the definition in (4)*, one can derive the theoretical values of the target Shapley effects in the Gaussian linear framework.*

Proof. For any subset $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, let us compute the theoretical value of T-VE$_u$ in the Gaussian linear framework.

- Case 1: $\boldsymbol{\beta}_{-u} \neq 0$

In that case, we have:

$$\begin{aligned}\mathbb{E}\left(1\left(\phi_{\boldsymbol{\beta}}\left(\mathbf{X}\right)>t\right)|\mathbf{X}_u=\mathbf{x}_u\right) &= \mathbb{P}\left(\phi_{\boldsymbol{\beta}}\left(\mathbf{X}\right)>t|\mathbf{X}_u=\mathbf{x}_u\right)\\ &= \mathbb{P}\left(\boldsymbol{\beta}^\top\mathbf{X}>t|\mathbf{X}_u=\mathbf{x}_u\right)\\ &= \mathbb{P}\left(\boldsymbol{\beta}_{-u}^\top\mathbf{X}_{-u}>t-\boldsymbol{\beta}_u^\top\mathbf{x}_u|\mathbf{X}_u=\mathbf{x}_u\right).\end{aligned}$$

Then, recall that the conditional normal distribution satisfies:

$$\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u \sim \mathcal{N}_{|-u|}\left(\boldsymbol{\mu}_{-u}+\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\left(\mathbf{x}_u-\boldsymbol{\mu}_u\right), \boldsymbol{\Sigma}_{-u,-u}-\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\boldsymbol{\Sigma}_{u,-u}\right). \tag{F.6}$$

Therefore, thanks to Theorem 6, the conditional distribution of $\boldsymbol{\beta}_{-u}^\top\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u$ is given by:

$$\begin{aligned}&\boldsymbol{\beta}_{-u}^\top\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u \sim \\ &\qquad \mathcal{N}_1\left[\boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\mu}_{-u}+\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\left(\mathbf{x}_u-\boldsymbol{\mu}_u\right)\right), \boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\Sigma}_{-u,-u}-\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\boldsymbol{\Sigma}_{u,-u}\right)\boldsymbol{\beta}_{-u}\right].\end{aligned} \tag{F.7}$$

Finally, we have:

$$\mathbb{E}\left(1\left(\phi_{\boldsymbol{\beta}}\left(\mathbf{X}\right)>t\right)|\mathbf{X}_u=\mathbf{x}_u\right) = 1-\Phi\left(\frac{t-\boldsymbol{\beta}_u^\top\mathbf{x}_u-\boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\mu}_{-u}+\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\left(\mathbf{x}_u-\boldsymbol{\mu}_u\right)\right)}{\sqrt{\boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\Sigma}_{-u,-u}-\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\boldsymbol{\Sigma}_{u,-u}\right)\boldsymbol{\beta}_{-u}}}\right), \tag{F.8}$$

an so:

$$\text{T-VE}_u = \mathbb{V}\left[\Phi\left(\frac{t-\boldsymbol{\beta}_u^\top\mathbf{X}_u-\boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\mu}_{-u}+\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\left(\mathbf{X}_u-\boldsymbol{\mu}_u\right)\right)}{\sqrt{\boldsymbol{\beta}_{-u}^\top\left(\boldsymbol{\Sigma}_{-u,-u}-\boldsymbol{\Sigma}_{-u,u}\boldsymbol{\Sigma}_{u,u}^{-1}\boldsymbol{\Sigma}_{u,-u}\right)\boldsymbol{\beta}_{-u}}}\right)\right]. \tag{F.9}$$

- Case 2: $\boldsymbol{\beta}_{-u} = 0$

In that case, the linear function $\phi_{\boldsymbol{\beta}}$ depends only on $\mathbf{x}_u$. Then:

$$\mathbb{E}\left(1\left(\phi_{\boldsymbol{\beta}}(\mathbf{X})>t\right)|\mathbf{X}_u\right) = 1\left(\phi_{\boldsymbol{\beta}}(\mathbf{X})>t\right), \tag{F.10}$$

and finally:

$$\text{T-VE}_u = \mathbb{V}\left[\mathbb{E}\left(1\left(\phi_{\boldsymbol{\beta}}(\mathbf{X})>t\right)|\mathbf{X}_u\right)\right] = \mathbb{V}\left(1\left(\phi_{\boldsymbol{\beta}}(\mathbf{X})>t\right)\right) = p_t^{\boldsymbol{\beta}}\left(1-p_t^{\boldsymbol{\beta}}\right). \tag{F.11}$$

To sum up, we have just proved the required result in (F.5) . □

APPENDIX G. PREPROCESSING PROCEDURE FOR THE GIVEN-DATA ESTIMATORS

APPENDIX G.1 Presentation of the procedure

The preprocessing procedure presented below will be applied as soon as a given-data estimator will be used, and it is based on the following theorem stated by [18].

Theorem 9. *Consider a family of uni-dimensional bijective transformations* $(\tau_i)_{i\in[\![1,d]\!]}$*. Let us define the following function:*

$$\widetilde{\phi}: \left| \begin{array}{rcl} \bigotimes_{i=1}^{d} \tau_i(\mathbb{X}_i) & \longrightarrow & \mathbb{R} \\ \mathbf{z} & \longmapsto & \phi\left(\tau_1^{-1}(z_1), \ldots, \tau_d^{-1}(z_d)\right), \end{array} \right. \tag{G.1}$$

its random input vector $\mathbf{Z} = (\tau_i(X_i))_{i\in[\![1,d]\!]}$ *and let us write* $\left(\widetilde{Sh}_i\right)_{i\in[\![1,d]\!]}$ *its Shapley effects. Recalling that* $(Sh_i)_{i\in[\![1,d]\!]}$ *are the Shapley effects of* ϕ*, then:*

$$\forall i \in [\![1,d]\!],\ Sh_i = \widetilde{Sh}_i. \tag{G.2}$$

In other words, this theorem shows that the Shapley effects are unchanged when bijective transformations are applied on each input variable. Practically, the preprocessing consists in applying the following procedure:

1. choose a sampling distribution $h \in \{f_{\mathbf{X}}, g\}$
2. if it is possible, for all $i \in [\![1,d]\!]$, compute the exact values of $\mu_i^{(h)} = \mathbb{E}_h(X_i)$ and $\left(\sigma_i^{(h)}\right)^2 = \mathbb{V}_h(X_i)$, else estimate them
3. for all $i \in [\![1,d]\!]$, define the linear bijective transformations by:

$$\forall x_i \in \mathbb{R},\ \tau_i(x_i) = \frac{x_i - \mu_i^{(h)}}{\sigma_i^{(h)}} \tag{G.3}$$

4. from a sample $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ drawn according to h, build the transformed sample $\left(\mathbf{Z}^{(n)}\right)_{n\in[\![1,N]\!]}$ defined for all $n \in [\![1,N]\!]$ by $\mathbf{Z}^{(n)} = \left(\tau_i\left(X_i^{(n)}\right)\right)_{i\in[\![1,d]\!]}$
5. apply the previous given-data estimators proposed in this article to the new sample $\left(\mathbf{Z}^{(n)}\right)_{n\in[\![1,N]\!]}$.

Hence, Theorem 9 applied to $\psi_t : \mathbf{x} \in \mathbb{X} \mapsto 1(\phi(\mathbf{x}) > t)$ with the family of linear bijective transformations $(\tau_i)_{i\in[\![1,d]\!]}$ defined in (G.3) justifies that the described preprocessing procedure should theoretically provide the expected target Shapley effects. Moreover, the transformations $(\tau_i)_{i\in[\![1,d]\!]}$ defined in (G.3) only consist here in a standardisation of the input sample $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$ drawn according to $h \in \{f_{\mathbf{X}}, g\}$, i.e. a re-scaling by $\sqrt{\mathbb{V}_h(X_i)}$ and a shifting by $\mathbb{E}_h(X_i)$ of each component of each point. In particular, the nearest-neighbour search is performed among the new sample $\left(\mathbf{Z}^{(n)}\right)_{n\in[\![1,N]\!]}$ in which distances between points are expected to be more homogeneous.

APPENDIX G.2 Motivations and practical interest

In order to motivate the introduction of the preprocessing procedure presented in APPENDIX G.1, let us reconsider the cantilever beam example of Section 4.2. If we estimate the target Shapley effects of this problem without applying the preprocessing procedure, we obtain the results presented in Figure G.8. The existing given-data estimators without importance sampling, especially those of the second, fourth and sixth indices, seem to be badly biased. This phenomenon can perhaps be explained by the huge scale difference between the third variable, the elastic modulus E, and the others. When 3 is in a subset $u \in \mathcal{P}(6)\backslash\{\varnothing, [\![1,6]\!]\}$, the distance in the subspace $\mathbb{X}_u$ is approximately equal to the distance in $\mathbb{X}_{\{3\}}$, which makes the nearest neighbour approximation of a conditional distribution given

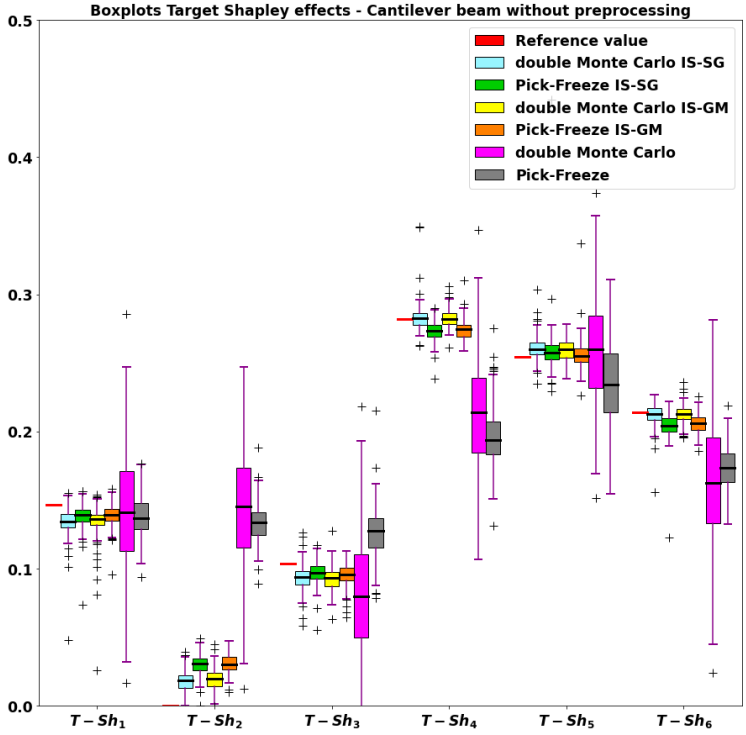

FIG. G.8: Estimation of the target Shapley effects in the cantilever beam example, in the given-data framework and without the preprocessing described in APPENDIX G.1.

some $\mathbf{x}_u \in \mathbb{X}_u$ very inaccurate. This phenomenon is getting worse without importance sampling because the points of interest, the failure points, are in the tail of the distribution, where the concentration of points is small and thus where the distances between points are even larger.

Consequently, the preprocessing procedure described in APPENDIX G.1 aims to restructure the available sample such that each component has the same scale in order to decrease the error due to the nearest neighbour approximation. By definition, for a given $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, this error mainly comes from the gap between $\psi_t\left(\mathbf{X}_u^{(n)}, \mathbf{X}_{-u}^{(k_N^u(n,2))}\right)$, the target value on the subspace $\left\{\mathbf{x} \in \mathbb{R}^d / \mathbf{x}_u = \mathbf{X}_u^{(n)}\right\}$, and $\psi_t\left(\mathbf{X}^{(k_N^u(n,2))}\right)$ which is its approximation, for all $n \in [\![1, N_u]\!]$. The restructuring caused by the linear bijective transformations defined in (G.3) aims then at reducing this error. Indeed, the nearest neighbour of some points of the new sample $\left(\mathbf{Z}^{(n)}\right)_{n\in[\![1,N]\!]}$ might not be the same as in the starting sample $\left(\mathbf{X}^{(n)}\right)_{n\in[\![1,N]\!]}$. Figure G.9 illustrates this phenomenon. More precisely, for all $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, there can exist some $n_0 \in [\![1, N]\!]$ such that $k_{\mathbf{x}}^u(n_0,2) \neq k_{\mathbf{z}}^u(n_0,2)$, where $k_{\mathbf{x}}^u(n_0,2)$ and $k_{\mathbf{z}}^u(n_0,2)$ represent the indices of the second nearest neighbour of the point n_0 respectively in the $\mathbf{x}$-space and in the $\mathbf{z}$-space, and thus $\psi_t\left(\mathbf{X}^{(k_{\mathbf{x}}^u(n_0,2))}\right) \neq \widetilde{\psi_t}\left(\mathbf{Z}^{(k_{\mathbf{z}}^u(n_0,2))}\right)$. At last, we expect that most of the time, this change leads to a reduction of the error, i.e. $\left|\widetilde{\psi_t}\left(\mathbf{Z}^{(k_{\mathbf{z}}^u(n_0,2))}\right) - \widetilde{\psi_t}\left(\mathbf{Z}_u^{(n_0)}, \mathbf{Z}_{-u}^{(k_{\mathbf{z}}^u(n_0,2))}\right)\right| \leq \left|\psi_t\left(\mathbf{X}^{(k_{\mathbf{x}}^u(n_0,2))}\right) - \psi_t\left(\mathbf{X}_u^{(n_0)}, \mathbf{X}_{-u}^{(k_{\mathbf{x}}^u(n_0,2))}\right)\right|$. A more advanced theoretical study is required to better understand this preprocessing procedure and potentially to improve it, but it seems beneficial in our examples, especially on the cantilever beam example. At last, note that in Section 4.1, we apply the preprocessing procedure with the given-data estimators. However, we would obtain almost exactly the same results without having applied it because in those examples, the transformations have a very mild impact.

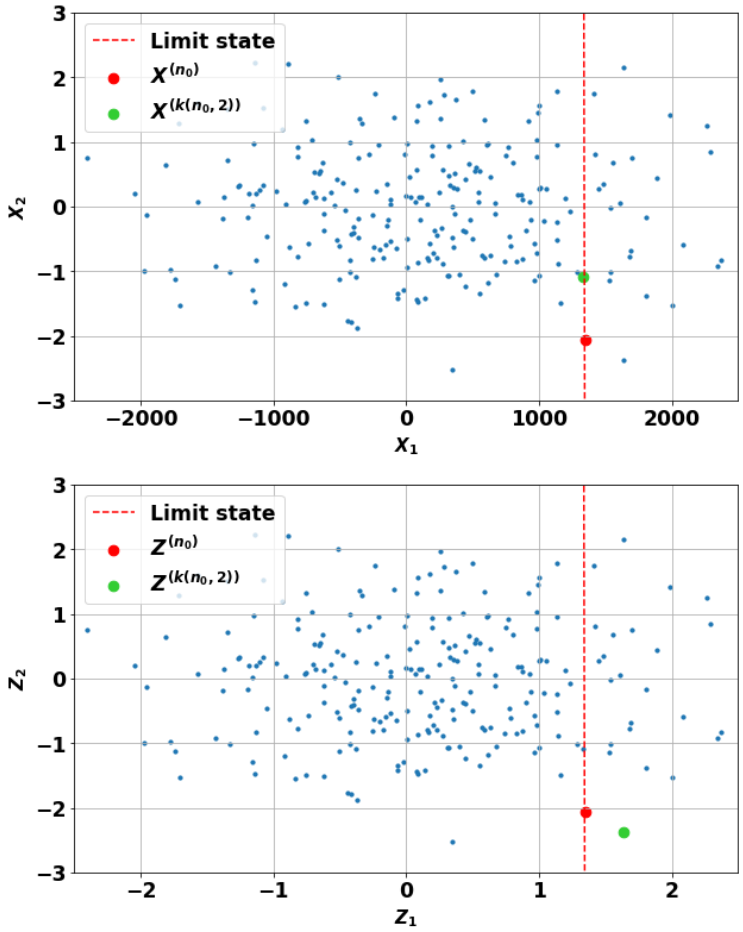

FIG. G.9: The upper figure represents in the $\mathbf{x}$-space a sample of $N = 200$ points drawn according to a zero-mean bi-dimensional normal distribution with independent components and such that $\mathbb{V}(X_1) = 10^9$ and $\mathbb{V}(X_2) = 1$. The lower figure represents in the $\mathbf{z}$-space the same sample after having applied on each component the transformations defined in (G.3). On both figures, the dotted red line represents the limit state of the linear function $\phi_{(1\ 1)}(x_1, x_2) = x_1 + x_2$ defined by the failure threshold $t = 1350$, the red point represents the point $n_0 \in [\![1, 200]\!]$ and the green point represents its nearest neighbour among the corresponding sample with the 2-dimensional distance. In both cases, the red point is in the failure domain. Moreover, in the $\mathbf{x}$-space (upper figure), the green point is in the safe domain whereas in the $\mathbf{z}$-space (lower figure), the green point is in the failure domain.

APPENDIX H. EQUATIONS OF THE FIRE SPREAD MODEL IN SECTION 4.3

Given the random input vector $\mathbf{X} = (\delta, \sigma, h, \rho_p, m_l, m_d, S_T, U, \tan\varphi, P)$, the rate of spread is obtained through the following system of equations:

$$R = \frac{I_R \xi (1 + \phi_W + \phi_S)}{\rho_b \epsilon Q_{ig}} \quad \text{rate of fire-spread, ft} \cdot \text{min}^{-1}$$

where

$$w_0 = \frac{4.8}{4.8824} \times \frac{1}{1 + \exp\left[(15 - \delta)/3.5\right]} \quad \text{fuel loading, kg} \cdot \text{m}^{-2}$$

$$\Gamma_{\max} = \frac{\sigma^{1.5}}{495 + 0.0594\sigma^{1.5}} \quad \text{maximum reaction velocity, min}^{-1}$$

$$\beta_{\text{op}} = 3.348\sigma^{-0.8189} \quad \text{optimum packing ratio}$$

$$A = 133\sigma^{-0.7913}$$

$$\theta^* = \frac{301.4 - 305.87(m_l - m_d) + 2260 m_d}{2260 m_l}$$

$$\theta = \min\left(1, \max\left(0, \theta^*\right)\right)$$

$$\mu_M = \exp\left[-7.3 P m_d - (7.3\theta + 2.13)(1 - P) m_l\right] \quad \text{moisture damping coefficient}$$

$$\mu_S = 0.174 S_T^{-0.19} \quad \text{mineral damping coefficient}$$

$$C = 7.47 \exp\left[-0.133\sigma^{0.55}\right]$$

$$B = 0.02526\sigma^{0.54}$$

$$E = 0.715 \exp\left[-3.59 \times 10^{-4}\sigma\right]$$

$$w_n = w_0(1 - S_T) \quad \text{net fuel loading, lb} \cdot \text{ft}^{-2}$$

$$\rho_b = \frac{w_0}{\delta} \quad \text{ovendry bulk density, lb} \cdot \text{ft}^{-3}$$

$$\epsilon = \exp\left[\frac{-138}{\sigma}\right] \quad \text{effective heating number}$$

$$Q_{ig} = 130.87 + 1054.43 m_d \quad \text{heat of preignition, Btu} \cdot \text{lb}^{-1}$$

$$\beta = \frac{\rho_b}{\rho_p} \quad \text{packing ratio}$$

$$\Gamma = \Gamma_{\max} \left(\frac{\beta}{\beta_{\text{op}}}\right)^A \exp\left[A\left(1 - \frac{\beta}{\beta_{\text{op}}}\right)\right] \quad \text{optimum reaction velocity, min}^{-1}$$

$$\xi = \frac{\exp\left[\left(0.792 + 0.681\sigma^{0.5}\right)(\beta + 0.1)\right]}{192 + 0.2595\sigma} \quad \text{propagating flux ratio}$$

$$\phi_W = C U^B \left(\frac{\beta}{\beta_{\text{op}}}\right)^{-E} \quad \text{wind coefficient}$$

$$\phi_S = 5.275\beta^{-0.3} (\tan\varphi)^2 \quad \text{slope factor}$$

$$I_R = \Gamma w_n h \mu_M \mu_S \quad \text{reaction intensity, Btu} \cdot \text{ft}^{-2} \cdot \text{min}^{-1}.$$

Note that the expression of the fuel loading w_0 according to the fuel depth δ is conjectured from the data analysis performed in [43]. In addition, it is important to remark that almost all the above equations, which mainly come from [42], are given in imperial units whereas the inputs are specified in metric units in Table 3. In order to have consistent results, it is thus necessary to convert the input variables into the imperial units at the beginning of the numerical calculus and to convert the output into $\text{cm} \cdot \text{s}^{-1}$ at the end.

APPENDIX I. COST-REDUCTION ESTIMATION PROCEDURE IN THE GIVEN-MODEL FRAMEWORK

In practice, the ROSA of a complex system always comes after the reliability analysis, i.e. the estimation of the failure probability. The given-model estimators (17) and (27) of the target conditional indices and thus the corresponding target Shapley effect estimators have a high computational cost, and the reliability analysis provides an i.i.d. input/output N-sample $\left(\mathbf{X}^{(n)}, \psi_t\left(\mathbf{X}^{(n)}\right)\right)_{n \in [\![1,N]\!]}$ with $\left(\mathbf{X}^{(n)}\right)_{n \in [\![1,N]\!]}$ distributed according to the IS auxiliary

density g. We present here a new procedure which re-uses the available sample in order to reduce the computational cost of the previous given-model estimators of the target conditional indices T-VE$_u$ and T-EV$_u$ and thus of the target Shapley effects. Remark first that estimators by importance sampling of p_t and $\mathbb{V}_{f_{\mathbf{X}}}\left(1\left(\phi\left(\mathbf{X}\right)>t\right)\right)$ can be easily computed with only the available sample and so do not require additional calls to ϕ.

APPENDIX I.1 Double Monte Carlo procedure

With the double Monte Carlo method in (17), for $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, set the parameters N_u and N_I and consider a sequence $(s(n))_{n\in[\![1,N_u]\!]}$ of uniformly distributed integers in $[\![1,N]\!]$ and apply the following scheme:

1. for $n \in [\![1,N_u]\!]$, draw an i.i.d. sample $\left(\widetilde{\mathbf{X}}_u^{(s(n),2)},\ldots,\widetilde{\mathbf{X}}_u^{(s(n),N_I)}\right)$ distributed according to the conditional distribution of $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{X}_{-u}^{(s(n))}}$
2. for $k \in [\![2,N_I]\!]$, compute $w_t^g\left(\widetilde{\mathbf{X}}_u^{(s(n),k)},\mathbf{X}_{-u}^{(s(n))}\right)$ and use the value of $w_t^g\left(\mathbf{X}^{(s(n))}\right)$ for the term corresponding to $k=1$
3. compute the estimators (19) and then (17).

This procedure requires $N_u\left(N_I-1\right)$ additional calls to ϕ to those from the reliability analysis to estimate the target conditional index T-EV$_u$ by double Monte Carlo with importance sampling. Typically, the authors of [17] suggest to use $N_I=3$ for numerical purposes, thus the proposed procedure is interesting because it does not require too many additional calls to the code ϕ.

APPENDIX I.2 Pick-Freeze procedure

With the Pick-Freeze method in (27), for any subset $u \in \mathcal{P}(d)\backslash\{\varnothing, [\![1,d]\!]\}$, set the parameter N_u, consider a sequence $(s(n))_{n\in[\![1,N_u]\!]}$ of uniformly distributed integers in $[\![1,N]\!]$ and apply the following scheme:

1. for $n \in [\![1,N_u]\!]$, draw a random variable $\widetilde{\mathbf{X}}_{-u}^{(s(n),2)}$ from the conditional distribution of $g_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{X}_u^{(s(n))}}$
2. compute $w_t^g\left(\mathbf{X}_u^{(s(n))},\widetilde{\mathbf{X}}_{-u}^{(s(n),2)}\right)$ and use the value of $w_t^g\left(\mathbf{X}^{(s(n))}\right)$ for the first term in the Pick-Freeze product
3. compute the estimator (27).

This procedure requires N_u additional calls to ϕ to those from the reliability analysis to estimate the target conditional index T-VE$_u$ by Pick-Freeze with importance sampling.

APPENDIX I.3 Cost reduction provided

In the above cost-reduction procedure with the double Monte Carlo (resp. Pick-Freeze) method, the sampling from the marginal distribution $g_{\mathbf{X}_{-u}}$ (resp. $g_{\mathbf{X}_u}$) is replaced by picking a random sub-sample among the sample $\left(\mathbf{X}_{-u}^{(n)}\right)_{n\in[\![1,N]\!]}$ (resp. $\left(\mathbf{X}_u^{(n)}\right)_{n\in[\![1,N]\!]}$) through the random sequence $(s(n))_{n\in[\![1,N_u]\!]}$. Then, for any $n \in [\![1,N_u]\!]$, a sample of size N_I-1 (resp. 1) is drawn according to the conditional distribution $g_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{X}_{-u}^{(s(n))}}$ (resp. $g_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{X}_u^{(s(n))}}$) and the missing point is set to $\mathbf{X}_u^{(s(n))}$ (resp. $\mathbf{X}_{-u}^{(s(n))}$) such that we obtain an i.i.d. sample of size N_I (resp. 2) distributed according to the corresponding conditional distribution. The latter missing point belongs to the available sample and does not require to be evaluated and so allows to save one call to ϕ. Eventually, given the data from the reliability analysis, this new procedure allows to save N_u calls to ϕ to estimate each target conditional index and so allows to save $N_{\mathbb{V}}+m(d-1)N_O$ calls to ϕ to estimate the d target Shapley effects with the random permutation aggregation procedure and $N_{\mathbb{V}}+(2^d-2)N_O$ calls with the subset aggregation procedure, where $N_{\mathbb{V}}$ and N_O are defined in Section 2.2.4.

REFERENCES

1. Davis, P.J. and Rabinowitz, P., *Methods of numerical integration*, Courier Corporation, 2007.
2. Rubinstein, R.Y. and Kroese, D.P., *Simulation and the Monte Carlo method*, Vol. 10, John Wiley & Sons, 2016.
3. Morio, J. and Balesdent, M., *Estimation of rare event probabilities in complex aerospace and other systems: a practical approach*, Woodhead publishing, 2015.
4. Bucklew, J., *Introduction to rare event simulation*, Springer Science & Business Media, 2004.
5. Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M., *Sensitivity analysis in practice: a guide to assessing scientific models*, Vol. 1, Wiley Online Library, 2004.
6. Marrel, A. and Chabridon, V., Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor, *Reliability Engineering & System Safety*, 214:107711, 2021.
7. Sobol, I.M., Sensitivity analysis for non-linear mathematical models, *Mathematical modelling and computational experiment*, 1:407–414, 1993.
8. Wei, P., Lu, Z., Hao, W., Feng, J., and Wang, B., Efficient sampling methods for global reliability sensitivity analysis, *Computer Physics Communications*, 183(8):1728–1743, 2012.
9. Perrin, G. and Defaux, G., Efficient evaluation of reliability-oriented sensitivity indices, *Journal of Scientific Computing*, 79(3):1433–1455, 2019.
10. Chastaing, G., Gamboa, F., and Prieur, C., Generalized Hoeffding-Sobol decomposition for dependent variables-application to sensitivity analysis, *Electronic Journal of Statistics*, 6:2420–2448, 2012.
11. Shapley, L.S., A value for n-person games, *Contributions to the Theory of Games*, (28):307–317, 1953.
12. Owen, A.B., Sobol'indices and Shapley value, *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251, 2014.
13. Il Idrissi, M., Chabridon, V., and Iooss, B., Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs, *Environmental Modelling & Software*, 143:105115, 2021.
14. Broto, B., Bachoc, F., and Depecker, M., Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution, *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):693–716, 2020.
15. Hoeffding, W., A class of statistics with asymptotically Normal distribution, *The Annals of Mathematical Statistics*, 19(3):293 – 325, 1948.
16. Homma, T. and Saltelli, A., Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering & System Safety*, 52(1):1–17, 1996.
17. Song, E., Nelson, B.L., and Staum, J., Shapley effects for global sensitivity analysis: Theory and computation, *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083, 2016.
18. Owen, A.B. and Prieur, C., On Shapley value for measuring importance of dependent inputs, *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):986–1002, 2017.
19. Iooss, B. and Prieur, C., Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol'indices, numerical estimation and applications, *International Journal for Uncertainty Quantification*, 9(5), 2019.
20. Castro, J., Gómez, D., and Tejada, J., Polynomial calculation of the Shapley value based on sampling, *Computers & Operations Research*, 36(5):1726–1730, 2009.
21. Plischke, E., Rabitti, G., and Borgonovo, E., Computing Shapley effects for sensitivity analysis, *SIAM/ASA Journal on Uncertainty Quantification*, 9(4):1411–1437, 2021.
22. Benoumechiara, N. and Elie-Dit-Cosaque, K., Shapley effects for sensitivity analysis with dependent inputs: bootstrap and Kriging-based algorithms, *ESAIM: Proceedings and Surveys*, 65:266–293, 2019.
23. Bénard, C., Biau, G., Da Veiga, S., and Scornet, E., Shaff: Fast and consistent shapley effect estimates via random forests, In *International Conference on Artificial Intelligence and Statistics*, pp. 5563–5582. PMLR, 2022.
24. Broto, B., Bachoc, F., Clouvel, L., and Martinez, J.M., Block-diagonal covariance estimation and application to the Shapley effects in sensitivity analysis, *arXiv preprint arXiv:1907.12780*, 2019.
25. Broto, B., Bachoc, F., Depecker, M., and Martinez, J.M., Sensitivity indices for independent groups of variables, *Mathematics and Computers in Simulation*, 163:19–31, 2019.

26. Sun, Y., Apley, D.W., and Staum, J., Efficient nested simulation for estimating the variance of a conditional expectation, *Operations research*, 59(4):998–1007, 2011.
27. Da Veiga, S. and Gamboa, F., Efficient estimation of sensitivity indices, *Journal of Nonparametric Statistics*, 25(3):573–595, 2013.
28. Hasofer, A.M. and Lind, N.C., Exact and invariant second-moment code format, *Journal of the Engineering Mechanics division*, 100(1):111–121, 1974.
29. Breitung, K., Asymptotic approximations for multinormal integrals, *Journal of Engineering Mechanics*, 110(3):357–366, 1984.
30. Cérou, F., Del Moral, P., Furon, T., and Guyader, A., Sequential Monte Carlo for rare event estimation, *Statistics and Computing*, 22(3):795–908, 2012.
31. Koutsourelakis, P.S., Pradlwarter, H., and Schuëller, G., Reliability of structures in high dimensions, part I: algorithms and applications, *Probabilistic Engineering Mechanics*, 19(4):409–417, 2004.
32. Kahn, H. and Harris, T.E., Estimation of particle transmission by random sampling, *National Bureau of Standards applied mathematics series*, 12:27–30, 1951.
33. Shinozuka, M., Basic Analysis of Structural Safety, *Journal of Structural Engineering-asce*, 109:721–740, 1983.
34. Harbitz, A., Efficient and accurate probability of failure calculation by the use of importance sampling technique, In *Proc. of ICASP*, Vol. 4, pp. 825–836, 1983.
35. Zhang, P., Nonparametric importance sampling, *Journal of the American Statistical Association*, 91(435):1245–1253, 1996.
36. De Boer, P.T., Kroese, D.P., Mannor, S., and Rubinstein, R.Y., A tutorial on the cross-entropy method, *Annals of operations research*, 134(1):19–67, 2005.
37. Rubinstein, R.Y. and Kroese, D.P., *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*, Springer Science & Business Media, 2013.
38. Raguet, H. and Marrel, A., Target and conditional sensitivity analysis with emphasis on dependence measures, *arXiv preprint arXiv:1801.10047*, 2018.
39. Geyer, S., Papaioannou, I., and Straub, D., Cross entropy-based importance sampling using Gaussian densities revisited, *Structural Safety*, 76:15–27, 2019.
40. Zhou, C., Lu, Z., Zhang, L., and Hu, J., Moment independent sensitivity analysis with correlations, *Applied Mathematical Modelling*, 38(19-20):4885–4896, 2014.
41. Li, B., Zhang, L., Zhu, X., Yu, X., and Ma, X., Reliability analysis based on a novel density estimation method for structures with correlations, *Chinese Journal of Aeronautics*, 30(3):1021–1030, 2017.
42. Rothermel, R.C., *A mathematical model for predicting fire spread in wildland fuels*, Vol. 115, Intermountain Forest & Range Experiment Station, Forest Service, US ..., 1972.
43. Salvador, R., Pinol, J., Tarantola, S., and Pla, E., Global sensitivity analysis and scale effects of a fire propagation model used over Mediterranean shrublands, *Ecological Modelling*, 136(2-3):175–189, 2001.
44. Albini, F.A., *Estimating wildfire behavior and effects*, Vol. 30, Department of Agriculture, Forest Service, Intermountain Forest and Range ..., 1976.
45. Catchpole, E.A. and Catchpole, W.R., Modelling moisture damping for fire spread in a mixture of live and dead fuels, *International Journal of Wildland Fire*, 1:101–106, 1991.
46. Clark, R., Hope, A., Tarantola, S., Gatelli, D., Dennison, P.E., and Moritz, M.A., Sensitivity analysis of a fire spread model in a chaparral landscape, *Fire Ecology*, 4(1):1–13, 2008.
47. Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C., Asymptotic normality and efficiency of two Sobol index estimators, *ESAIM: Probability and Statistics*, 18:342–364, 2014.
48. Zahm, O., Cui, T., Law, K., Spantini, A., and Marzouk, Y., Certified dimension reduction in nonlinear Bayesian inverse problems, *arXiv preprint arXiv:1807.03712*, 2018.
49. Echard, B., Gayton, N., Lemaire, M., and Relun, N., A combined importance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models, *Reliability Engineering & System Safety*, 111:232–240, 2013.