Estimating the small failure probability of a nuclear passive safety system by means of an efficient Adaptive Metamodel-Based Subset Importance Sampling method

Nicola Pedroni, Enrico Zio

Estimating the small failure probability of a nuclear passive safety system by means of an efficient Adaptive Metamodel-Based Subset Importance Sampling method

N. Pedroni & E. Zio
Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy – Fondation EdF, CentraleSupelec, France

E. Zio
Department of Energy, Politecnico di Milano, Italy

ABSTRACT: The assessment of the functional failure probability of a thermal-hydraulic (T-H) passive system can be done by Monte Carlo (MC) sampling of the uncertainties affecting the T-H system model and its parameters. The computational effort associated to this approach can be prohibitive because of the large number of lengthy T-H code simulations necessary for the accurate and precise quantification of the (typically small) failure probability. To overcome this issue, in the present paper we propose an Adaptive Metamodel-Based Subset Importance Sampling (AM-SIS) approach that originally and efficiently combines the powerful features of several advanced computational methods of literature: in particular, Subset Simulation (SS) and fast-running Artificial Neural Network (ANN) metamodels are coupled within an adaptive MC-based Importance Sampling (IS) scheme. The objective is to construct a fully nonparametric estimator of the ideal, zero-variance Importance Sampling Density (ISD) and iteratively refine it, in such a way that: (i) the accuracy and precision of the corresponding failure probability estimates are improved and (ii) the number of burdensome T-H code runs is reduced, along with the associated computational cost. The method is demonstrated on a case study of an emergency passive decay heat removal system of a Gas-cooled Fast Reactor (GFR). A thorough comparison is made with respect to several advanced MC methods of literature.

1 INTRODUCTION

Modern nuclear reactors make use of passive safety features, which do not need external input to operate and, thus, are expected to improve the safety of nuclear power plants because of simplicity and reduction of both human interactions and hardware failures (Hassija et al. 2014).

However, the aleatory and epistemic uncertainties involved in the operation and modeling of passive systems are usually larger than for active systems (USNRC 2009). Due to these uncertainties, the physical phenomena involved in the passive system functioning (e.g., natural circulation) might develop in such a way to lead the system to fail its function (e.g., decay heat removal) (Burgazzi 2014). In the analysis of such *functional failure* behavior, the passive system is modeled by a mechanistic Thermal-Hydraulic (T-H) code and the probability of failing to perform the required function is estimated based on a Monte Carlo (MC) sample of code runs which propagate the uncertainties in the model and numerical values of its parameters (Mezio et al. 2014).

In practice, the probability of functional failure of a passive system is very small (e.g., around 10^{-4} or less), so that a large number of samples is necessary for acceptable estimation accuracy. Given that the time required for each run of the detailed T-H system model code can be of the order of several hours, the MC-based procedure typically requires considerable computational efforts (Fong et al. 2009).

Two main classes of approaches are usually considered to tackle this issue (Zio & Pedroni, 2011). On one side, *fast-running* surrogate regression models (also called response surfaces or metamodels) can be ‘trained’ to reproduce the behavior of the long-running T-H model code and used in the functional failure analysis. Examples of surrogate metamodels include polynomial chaos expansions (Kersaudi et al. 2015), Artificial Neural Networks (ANNs) (Zio et al. 2010), Support Vector Machines (SVMs) (Hurtado 2007) and kriging (Bect et al. 2012).

On the other side, efficient Monte Carlo Simulation techniques can be employed to perform robust estimations with a limited number of input samples. Examples include Subset Simulation (SS) (Au and Wang 2014), Line Sampling (LS) (Valdebenito et al. 2010, Zio and Pedroni 2010) and splitting methods (Botev & Kroese 2012). However, one of the most popular advanced MCS techniques is that of Importance Sampling (IS), whereby an Importance Sampling Density (ISD) is chosen so as to force the rare failure event to occur more often. In this regard, it is known that there exists an optimal ISD so that the variance of the MC estimator is zero. Unfortunately, this optimal ISD is not implementable in practice, since its analytical expression depends on the unknown failure probability itself. With respect to that, techniques have been proposed to reduce some distances between the instrumental ISD and the optimal one: see, e.g., the Adaptive Kernel (AK) (Morio 2012), the Cross-Entropy (CE) (Botev & Kroese 2008), the Variance Minimization (VM) (Asmussen & Glynn 2007) and the Markov Chain Monte Carlo-Importance Sampling (MCMC-IS) (Botev et al. 2013) methods.

Finally, efficient *combinations* of advanced MCS methods with metamodeling can be found in (Echard et al. 2011 and 2013, Bourinet et al. 2011, Dubourg et al. 2013, Fauriat & Gayton 2014).

In the present paper, we propose a novel approach, namely, the Adaptive Metamodel-based Subset Importance Sampling (AM-SIS) method, which originally combines the powerful features of three existing techniques, i.e., MCMC-IS, SS and ANNs. The method consists of the following main steps: (1) an estimator of the optimal ISD is constructed in two stages: (a) the SS technique is adopted to generate a population of samples approximately distributed according to the optimal ISD. In order to reduce the computational effort associated to this step, the original T-H model code is replaced by an *adaptively refined* ANN; (b) the population thereby created is 'fitted' by means of a proper Probability Density Function (PDF) to obtain an estimator for the optimal ISD: in this paper, the *fully nonparametric* PDF proposed in (Botev et al. 2013) and based on the Gibbs Sampler is employed to this aim; (2) Importance Sampling (IS) is performed using the ISD estimator constructed at step (1).

The main contributions of the present work with respect to the reference paper by (Botev et al. 2013) are the following: (1) SS is employed for the construction of the quasi-optimal ISD; (2) ANNs are used to reduce the computational effort associated to the construction of the quasi-optimal ISD: such metamodels are trained according to a *sequential, iterative* algorithm that makes an intelligent use of the samples generated by SS in order to increase the ANN accuracy in proximity of the failure domain; (3) the performance of the AM-SIS method is assessed with a very small number of T-H code evaluations (e.g., of the order of few tens or hundreds): this is important for practical cases in which the T-H computer codes require several hours to run a single simulation; (4) the computational efficiency of AM-SIS is systematically compared to that of several other advanced simulation methods of literature.

The investigations are carried out with regards to a case study dealing with the functional failure analysis of a passive, natural convection-based decay heat removal system of a Gas-cooled Fast Reactor (GFR), modified from (Pagani et al. 2005).

2 FUNCTIONAL FAILURE ANALYSIS OF T-H PASSIVE SYSTEMS

The basic quantitative steps of the functional failure analysis of a T-H passive system are (Bassi & Marquès 2008):

1 Detailed modeling of the passive system response by means of a deterministic, best-estimate (typically long-running) T-H code.
2 Identification of the parameters/variables $\boldsymbol{x}$ = $\{x_1, x_2, \ldots, x_j, \ldots, x_{n_i}\}$, models and correlations (i.e., the inputs to the T-H code) which contribute to the uncertainty in the results, i.e., the outputs $\boldsymbol{y}$ = $\{y_1, y_2, \ldots, y_l, \ldots, y_{n_o}\}$, of the best estimate T-H calculations.
3 Propagation of the uncertainties through the deterministic, long-running T-H code in order to estimate the *functional failure probability* of the passive system. Let $Y(\boldsymbol{x})$ be a scalar variable indicator of the performance of the passive system (e.g., the fuel peak cladding temperature) and α_Y a threshold value defining the corresponding failure criterion (e.g., a limit value imposed by regulating authorities). Also, let us assume that the passive system operates as long as $Y(\boldsymbol{x}) < \alpha_Y$. The probability $P(F)$ of system failure can be expressed by:

$$P(F) = \iint \ldots \int I_F(\boldsymbol{x}) q(\boldsymbol{x}) d\boldsymbol{x} \tag{1}$$

where $q(\cdot)$ is the joint PDF representing the uncertainty in the parameters $\boldsymbol{x}$, F is the failure region (where $Y(\boldsymbol{x}) \geq \alpha_Y$) and $I_F(\cdot)$ is an indicator function such that $I_F(\boldsymbol{x}) = 1$, if $\boldsymbol{x} \in F$ and $I_F(\boldsymbol{x}) = 0$, otherwise.

Step 3 above relies on multiple (e.g., many thousands) evaluations of the T-H code for different combinations of system inputs; this can render the associated computing cost prohibitive, when the running time for each T-H code simulation takes several hours (which is often the case for T-H passive systems). This issue can be tackled in two effective ways: from one side, efficient MCS techniques (e.g., Importance Sampling-IS, see Section 3); from the other side, fast-running metamodels (e.g., Artificial Neural Networks-ANNs, see Section 4).

3 THE IMPORTANCE SAMPLING METHOD

The concept underlying Importance Sampling (IS) is to replace the original PDF $q(\boldsymbol{x})$ with an Importance Sampling Density (ISD) $g(\boldsymbol{x})$ chosen by the analyst so as to generate a large number of samples in the "important region" of the sample space, i.e. the failure region F (Dubourg et al. 2013). The failure probability can, then, be rewritten as:

$$P(F) = \int \left[\frac{I_F(\boldsymbol{x}) q(\boldsymbol{x})}{g(\boldsymbol{x})} \right] g(\boldsymbol{x}) d\boldsymbol{x} = E_g \left[\frac{I_F(\boldsymbol{x}) q(\boldsymbol{x})}{g(\boldsymbol{x})} \right] \tag{2}$$

and the corresponding MC estimator $\hat{P}(F)^{N_T}$ becomes

$$\hat{P}(F)^{N_T} = \frac{1}{N_T}\sum_{k=1}^{N_T}\frac{I_F(\boldsymbol{x}^k)q(\boldsymbol{x}^k)}{g(\boldsymbol{x}^k)}, \tag{3}$$

where $\{\boldsymbol{x}^k : k = 1, 2, ..., N_T\}$ are N_T independent and identically distributed (i.i.d.) samples drawn from the ISD $g(\boldsymbol{x})$.

In theory, by minimizing the variance of the MC estimator $\hat{P}(F)^{N_T}$ (3), it is possible to derive the expression of the optimal ISD $g^*(\boldsymbol{x})$ (Dubourg et al. 2013):

$$g^*(\boldsymbol{x}) = \frac{I_F(\boldsymbol{x})q(\boldsymbol{x})}{P(F)}. \tag{4}$$

It is well-known that this optimal ISD is of no practical use, as its definition requires the knowledge of the failure probability $P(F)$ itself. In this respect, several techniques have been developed in order to approximate the optimal ISD (4) or to at least find one giving small variance of the estimator (3): see the Introduction. In Section 5, we provide details about the approach here adopted to tackle this issue.

4 EMPIRICAL REGRESSION MODELING

A metamodel is a regression function adopted for estimating the (possibly nonlinear) relationship between a vector of input variables $\boldsymbol{x} = \{x_1, x_2, ..., x_j, ..., x_{n_i}\}$ and a vector of output targets $\boldsymbol{y} = \{y_1, y_2, ..., y_l, ..., y_{n_o}\}$, on the basis of a *finite* (and possibly *reduced*) set of input/output data examples (i.e., patterns) $D_{TR} = \{(\boldsymbol{x}_p, \boldsymbol{y}_p), p = 1, 2, …, N_{TR}\}$, also referred to as Design Of Experiments (DOE). It can be assumed that the target vector $\boldsymbol{y}$ is related to the input vector $\boldsymbol{x}$ by an unknown nonlinear deterministic function $\boldsymbol{\mu}_y(\boldsymbol{x})$. In the present case of T-H passive system functional failure probability assessment the vector $\boldsymbol{x}$ contains the relevant uncertain system parameters/variables, the nonlinear deterministic function $\boldsymbol{\mu}_y(\boldsymbol{x})$ represents the complex T-H mechanistic model code and the vector $\boldsymbol{y}(\boldsymbol{x})$ contains the output variables of interest for the analysis. The objective is to estimate $\boldsymbol{\mu}_y(\boldsymbol{x})$ by means of a regression function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}^*)$ depending on a set of parameters $\boldsymbol{w}^*$ to be properly determined on the basis of the available data set D_{train}. The algorithm used to calibrate the set of parameters $\boldsymbol{w}^*$ is obviously dependent on the nature of the regression model adopted, but in general it aims at minimizing the mean (absolute or quadratic) error between the output targets of the original T-H code, $\boldsymbol{y}_p = \boldsymbol{\mu}_y(\boldsymbol{x})$, $p = 1, 2, ..., N_{TR}$, and the output vectors of the regression model, $\hat{\boldsymbol{y}}_p = \boldsymbol{f}(\boldsymbol{x}_p, \boldsymbol{w}^*)$, $p = 1, 2, ..., N_{TR}$. Once built, the regression model $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}^*)$ can be used as a simplified, quick-running surrogate of the original, long-running T-H model code for reducing the computational burden associated to the functional failure analysis of T-H passive systems.

In this work, three-layered feed-forward Artificial Neural Network (ANN) regression models trained by the error back-propagation algorithm are considered (Zio et al. 2010). They are composed of many computing units (called *neurons* or *nodes* and mathematically represented by sigmoidal basis functions) that are arranged in three *layers* (namely, the input, hidden and output layers) and interconnected by weighed connections (called *synapses*): the three layers contain n_i, n_h and n_o nodes, respectively. The interesting aspect of ANNs is that they have been demonstrated to be universal approximants of any *continuous* nonlinear function, i.e., in principle, of any nonlinear T-H code simulating the system of interest (Cybenko 1989).

In the following Section 5, we present an effective strategy, called Adaptive Metamodel-based Subset Importance Sampling (AM-SIS), that combines the powerful features of two advanced MCS methods (i.e., IS and SS) and a metamodel (i.e., an ANN) for reducing the computational efforts related to the assessment of the small functional failure probabilities of T-H passive systems.

5 THE ADAPTIVE METAMODEL-BASED SUBSET IMPORTANCE SAMPLING (AM-SIS) METHOD

The AM-SIS method here proposed consists of two main steps (Botev et al. 2013): (a) an estimator $\hat{g}^*(\boldsymbol{x})$ of the optimal ISD $g^*(\boldsymbol{x})$ (4) is constructed (Section 5.1); (b) Importance Sampling (IS) is performed, in which the estimator $\hat{g}^*(\boldsymbol{x})$ constructed at step a. above is used as an ISD to evaluate the functional failure probability $P(F)$ (1) (Section 5.2).

5.1 *Metamodel-based adaptive approximation of the optimal ISD by Subset Simulation*

The estimator $\hat{g}^*(\boldsymbol{x})$ of the optimal ISD $g^*(\boldsymbol{x})$ is, here, constructed in two stages: (a) the Subset Simulation (SS) technique (Au and Wang 2014) is adopted to generate a population $\{z_F^m: m = 1, 2, …, M\}$ of M samples *approximately* distributed according to the optimal ISD $g^*(\cdot)$ (4), i.e., $\{z_F^m: m = 1, 2, …, M\} \sim g^*(\cdot) = I_F(\cdot)q(\cdot)/P(F)$[1]. In order to reduce the computational effort associated to this step, the original long-running T-H model code is replaced by a fast-running surrogate ANN metamodel, properly constructed and *adaptively* refined by means of the samples *iteratively* generated by SS (Section 5.1.1);

[1] Notice that the change of name of the 'dummy input vector' from $\boldsymbol{x}$ to $\boldsymbol{z}$ has no 'conceptual reason' and is done only for notational simplicity and coherence with the description of the method reported in the following Sections.

(b) the population $\{z_F^m: m = 1, 2, \ldots, M\}$ is 'fitted' by means of a proper PDF to obtain the estimator $\hat{g}^*(\boldsymbol{x})$. The *fully nonparametric* PDF proposed in (Botev et al. 2013) and based on the Gibbs Sampler is here employed to this aim (Section 5.1.2).

5.1.1 *Generation of samples distributed as the optimal ISD by means of Subset Simulation and adaptively trained ANN metamodels*

As highlighted above, the construction of an estimator $\hat{g}^*(\cdot)$ for the optimal ISD $g^*(\cdot)$ (4) requires the generation of a population $\{z_F^m: m = 1, 2, \ldots, M\}$ of M samples *approximately* distributed according to the optimal ISD, i.e., $\{z_F^m: m = 1, 2, \ldots, M\} \sim g^*(\cdot) = I_F(\cdot)q(\cdot)/P(F)$. In this paper, the Subset Simulation (SS) technique is adopted to this aim. In synthesis, SS is an adaptive probabilistic simulation method originally developed for efficiently computing small failure probabilities in structural reliability problems. The idea underlying the SS method is to convert the simulation of an event (e.g., the rare failure event F) into a sequence of n_F simulations of intermediate conditional events corresponding to subsets (or subregions) F_i, $i = 1, 2, \ldots, n_F$, of the uncertain input parameter space. During simulation, the conditional samples (lying in the intermediate subsets or subregions F_i) are generated by Markov Chain Monte Carlo (MCMC): by so doing, the conditional samples gradually populate the successive intermediate subsets (or subregions) F_i up to the target (failure) region F (Au & Wang 2014).

The SS algorithm proceeds as follows. First, M vectors $\{z_0^m: m = 1, 2, \ldots, M\}$ are sampled by standard MCS, i.e., from the original PDF $q(\cdot)$. The corresponding values of the response variable $\{Y(z_0^m): m = 1, 2, \ldots, N\}$ are, then, computed and the first threshold value y_1 (identifying the first intermediate conditional event) is chosen as the $(1 - p_0)M^{\text{th}}$ value in the increasing list of values $\{Y(z_0^m): m = 1, 2, \ldots, M\}$ ($p_0 = 0.1$, in this paper). With this choice of y_1, there are now p_0M samples among $\{z_0^m: m = 1, 2, \ldots, M\}$ whose response $Y(z)$ lies in the intermediate subregion $F_1 = \{z: Y(z) > y_1\}$. Starting from each one of these samples, MCMC simulation is properly used to generate $(1 - p_0)M$ additional conditional samples in the intermediate subregion $F_1 = \{z: Y(z) > y_1\}$: by so doing, there are a total of M conditional samples $\{z_1^m: m = 1, 2, \ldots, M\} \in F_1$. Notice that the Markov chains are designed so that the target conditional distribution of the M samples $\{z_1^m: m = 1, 2, \ldots, M\}$ is $q(\cdot|F_1) = I_{F1}(\cdot)q(\cdot)/P(F_1)$. Then, the intermediate threshold value y_2 is chosen as the $(1 - p_0)M^{\text{th}}$ value in the ascending list of $\{Y(z_1^m): m = 1, 2, \ldots, M\}$ to define $F_2 = \{z: Y(z) > y_2\}$. The p_0M samples lying in F_2 function as 'seeds' for sampling $(1 - p_0)M$ additional conditional samples lying in F_2, for a total of M conditional samples $\{z_2^m: m = 1, 2, \ldots, M\} \in F_2$ (distributed as $q(\cdot|F_2) = I_{F2}(\cdot)q(\cdot)/P(F_2)$). This procedure is repeated until (at least) M samples $\{z_{nF}^m: m = 1, 2, \ldots, M\}$ are generated in the failure region $F_{nF} \equiv F = \{z: Y(z) > \alpha_Y\}$. It is demonstrated that following this procedure, the stationary density of the points $\{z_{nF}^m: m = 1, 2, \ldots, M\}$ is $q(\cdot|F) = I_F(\cdot)q(\cdot)/P(F)$, which coincides with the optimal ISD $g^*(\cdot)$ (4) (Au & Wang, 2014).

The additional main advantages of SS are the following: (i) no prior information about the failure region is required for its functioning; (ii) since the method relies on several Markov chains, it can identify also multiple, possibly disconnected failure regions. On the other hand, notice that a single run of the SS algorithm may require a large number N_c (e.g., hundreds or thousands) of evaluations of the long-running system model code, depending on the magnitude of $P(F)$ (i.e., approximately $N_c = M \cdot \log_{p0}[P(F)]$, with $M \geq 100$ for obtaining reliable estimates). Thus, in order to reduce the associated computational effort, the original T-H model code is replaced by a fast-running surrogate ANN, properly constructed and *adaptively* refined by means of the samples *iteratively* generated by SS itself (Bourinet et al. 2011). In synthesis, the idea is to build an initial ANN by means of a DOE generated by classical techniques (e.g., plain random sampling, Latin Hypercube Sampling, partial or full factorial designs, and so on). Then, SS is run using the constructed ANN (instead of the original T-H code) and samples are generated in the failure region F of interest. Some of the sampling points lying in proximity of F are added to the current DOE, which is then *enriched* in the areas close to the failure region. The enriched DOE is employed to *update* the ANN, which is, then, expected to be more *accurate* in reproducing the behavior of the original T-H code in proximity of the region F of interest. This process is *repeated* until the identified failure region F does not change much from one iteration to another, i.e., until it is not accurately and reliably "*located*". The proposed algorithm can be summarized as follows:

1. Build and train a first ANN regression model $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{w}^0)$ on the basis of a DOE $D_{TR}^0 = \{(\boldsymbol{x}_p^0, \boldsymbol{y}_p^0), p = 1, 2, \ldots, N_{TR}^0\}$ composed by N_{TR}^0 vectors selected by classical techniques. This allows obtaining a "first trial" fast-running surrogate of the T-H model code (see Section 4);
2. Run SS using the ANN regression model $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{w}^0)$ to obtain M samples $\{z_{nF}^{m,0}: m = 1, 2, \ldots, M\}$ approximately distributed as $g^*(\cdot) = q(\cdot|F) = I_F(\cdot)q(\cdot)/P(F)$ ($M = 1000$ in this paper). Notice that this operation is almost computationally costless, since it only involves computing by the fast-running ANN $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{w}^0)$ instead of by the original long-running T-H code $\boldsymbol{\mu}_{\boldsymbol{y}}(\boldsymbol{z})$;
3. Use $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{w}^0)$ to provide an estimate $\hat{P}(F)_{ANN}^{N_T,0}$ of the failure probability $P(F)$ (e.g., by FORM, direct MCS or other straightforward techniques).

As for step 2. above, this operation is almost computationally costless;

4 Set $t = 0$ and $\varepsilon_{P(F)} = 100\%$, where t is the iteration number and $\varepsilon_{P(F)}$ quantifies the relative change between two successive ANN-based estimates of $P(F)$ obtained during the adaptive and iterative process of metamodel refinement;

5 while $\varepsilon_{P(F)} > \varepsilon_{P(F)}{}^{th}$ (i.e., the relative variation between the ANN-based failure probability estimates at two successive iterations exceeds a predefined threshold $\varepsilon_{P(F)}{}^{th}$) *and* $N_{TR}{}^{t} < N_{TR}{}^{max}$ (i.e., the size of the DOE – in other words, the total number of T-H model runs – is below a maximal allowable value $N_{TR}{}^{max}$), perform the following steps in order to *adaptively* and *iteratively* refine the ANN metamodel in proximity of the failure region F and to produce samples approximately distributed as the optimal ISD $g^*(\cdot)$:

a. $t = t + 1$;

b. select N_{add} points $\{z_p{}^{add}$: $p = 1, 2, \ldots, N_{add}\}$ among *all* those generated by SS at iteration $(t - 1)$: the most consistent fraction of these points (say, $K = 0.5$) is chosen around the limit state in order to better characterize the behaviour of the T-H passive system in proximity of the failure region F; the remaining vectors are "distributed" around the boundaries of the other intermediate conditional regions F_i identified by SS. This allows a robust "anchoring" of the ANN in the key areas of the uncertain input space "visited" by SS;

c. join the N_{add} new points to the current DOE upon N_{add} new evaluations of the passive system performance function by means of the *original* T-H code $\boldsymbol{\mu}_y(\boldsymbol{z})$: thus, $D_{TR}^{t} = D_{TR}^{t-1} \cup D_{TR}^{add}$, where $D_{TR}{}^{add} = \{(\boldsymbol{z}_p{}^{add}, \boldsymbol{y}_p{}^{add} = \boldsymbol{\mu}_y(\boldsymbol{z}_p{}^{add})), p = 1, 2, \ldots, N_{add}\}$;

d. build and train an ANN regression model $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}^t)$ using the updated DOE D_{TR}^{t};

e. run SS using the ANN regression model $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}^t)$ to obtain M samples $\{\boldsymbol{z}_{nF}{}^{m,t}$: $m = 1, 2, \ldots, M\}$ approximately distributed as $g^*(\cdot)$;

f. provide an ANN-based estimate $\hat{P}(F)_{ANN}^{N_T,t}$ of the failure probability $P(F)$ and compute $\varepsilon_{P(F)}$ as $\left|\hat{P}(F)_{ANN}^{N_T,t} - \hat{P}(F)_{ANN}^{N_T,t-1}\right| / \hat{P}(F)_{ANN}^{N_T,t-1}$;

6 the failure region F of interest is accurately and robustly "located" and the ANN metamodel is sufficiently refined in its proximity. Thus, retain the samples $\{\boldsymbol{z}_{nF}{}^{m,t}$: $m = 1, 2, \ldots, M\}$ generated at the last iteration t, as if they were the "optimal ones" $\{\boldsymbol{z}_F{}^{m}$: $m = 1, 2, \ldots, M\}$ distributed according to the *real* zero-variance ISD $g^*(\cdot)$.

5.1.2 *Construction of an estimator of the optimal ISD by means of a fully nonparametric PDF*

The adaptive ANN-based algorithm described above (relying on SS) allows sampling states approximately from the optimal zero-variance ISD. Botev et al. (2013) propose to 'fit' those points by means of the one step ahead transition density $\kappa(\boldsymbol{x}|\boldsymbol{z})$ of a Markov chain, when the chain is in state $\boldsymbol{z}_F^m$, and average over the M states $\{\boldsymbol{z}_F{}^{m}$: $m = 1, 2, \ldots, M\}$ (obviously, the *target* stationary distribution of the selected transition density $\kappa(\boldsymbol{x}|\boldsymbol{z})$ need to be the optimal ISD $g^*(\boldsymbol{x})$). In other words, regardless of how the transition density $\kappa(\boldsymbol{x}|\boldsymbol{z})$ is constructed, the optimal ISD $g^*(\boldsymbol{x})$ is approximated by $\hat{g}^*(\boldsymbol{x})$ (5):

$$\hat{g}^*(\boldsymbol{x}) = \frac{1}{M}\sum_{m=1}^{M}\kappa\left(\boldsymbol{x}|\boldsymbol{z}_F^m\right). \tag{5}$$

Following the suggestion of Botev et al. (2013), in this paper the transition density $\kappa\left(\boldsymbol{x}|\boldsymbol{z}_F^m\right)$ adopted by the very well-known Gibbs Sampler is considered (see the cited reference for technical details).

The advantage of the ISD (5) is that we are no longer restricted by parametric models, such as those used in the Cross Entropy (CE) or Variance Minimization (VM) methods. On the other hand, for ISD (5) in the rare-event setting it is not guaranteed that the positivity condition holds (i.e., that $\hat{g}^*(\boldsymbol{x}) > 0$ whenever $I_F(\boldsymbol{x})q(\boldsymbol{x}) > 0$). In such a case, we modify $\hat{g}^*(\boldsymbol{x})$ to make sure that it is nonzero over the entire support of $g^*(\boldsymbol{x})$. One way of doing this is to take a mixture of $\hat{g}^*(\boldsymbol{x})$ with the original density $q(\boldsymbol{x})$, i.e.:

$$\hat{g}_w^*(\boldsymbol{x}) = w \cdot q(\boldsymbol{x}) + (1-w)\cdot \hat{g}^*(\boldsymbol{x}), \tag{6}$$

where w is an arbitrary weight between 0 and 1 (in this paper, $w = 0.01$).

5.2 *Importance sampling from the estimator of the optimal ISD*

The procedure for drawing N_T samples from the estimator $\hat{g}_w^*(\boldsymbol{x})$ (6) of the optimal ISD $g^*(\boldsymbol{x})$ is based on the so-called 'composition method':

1 Set the sample index $k = 1$;

2 Generate a uniform random number u in $[0, 1)$;

3 If $u < w$, then generate a random sample $\boldsymbol{x}^k$ from the original PDF $q(\boldsymbol{x})$; otherwise, sample from $\hat{g}_w^*(\boldsymbol{x})$; in particular: (a) generate the integer random number m over the set $\{1, 2, \ldots, M\}$; (b) sample $\boldsymbol{x}^k$ from the corresponding transition density $\kappa\left(\boldsymbol{x}|\boldsymbol{z}_F^m\right)$ in (5);

4 If $k < N_T$, then set $k = k + 1$ and return to step 2. above; otherwise, go to step 5. below;

5 Referring to (3), calculate the AM-SIS estimator $\hat{P}(F)^{N_T}$ of the failure probability $P(F)$ employing (6) as instrumental ISD.

6 CASE STUDY DESCRIPTION

The case study considered in this work concerns the natural convection cooling in a 600-MW Gas-cooled Fast Reactor (GFR) under a post-Loss Of Coolant Accident (LOCA) condition (Pagani et al. 2005).

A GFR decay heat removal configuration is shown schematically in Figure 1; in the case of a LOCA, the long-term heat removal is ensured by natural circulation in a given number N_{loops} of identical and parallel loops. Only one of the N_{loops} loops is reported for clarity of the picture: the path of the cooling helium gas is indicated by the black arrows.

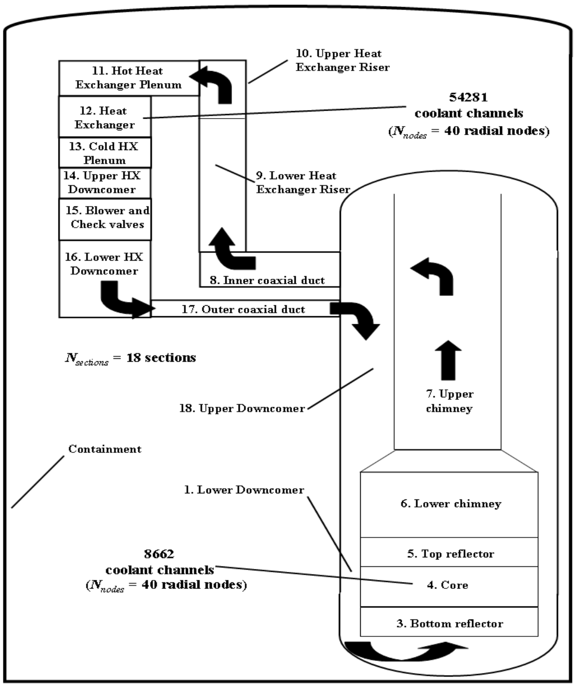

Figure 1. Schematic representation of one loop of the 600-MW GFR passive decay heat removal system (Pagani et al. 2005).

In the present analysis, the average core power to be removed is assumed to be 18.7MW: to guarantee natural circulation cooling at this power level, a nominal pressure of 1650kPa in the loops is required. Finally, the secondary side of the heat exchanger (i.e., item 12 in Figure 1) is assumed to have a nominal wall temperature of 90°C.

Only epistemic uncertainties are considered. Parameter uncertainties are associated to the reactor power level (x_1), the pressure in the loops (x_2) and the cooler wall temperature (x_3); model uncertainties are associated to the correlations used to calculate the Nusselt numbers (x_4, x_5 and x_6) and friction factors (x_7, x_8 and x_9) in the forced, mixed and free convection regimes, respectively. These uncertainties leads to the definition of a vector $\boldsymbol{x}$ of nine uncertain model inputs $\boldsymbol{x} = \{x_j: j = 1, 2, ..., 9\}$, assumed described by normal distributions of known means $\boldsymbol{\mu}$ = [18.7MW, 1650kPa, 90°C, 1, 1, 1, 1, 1, 1] and standard deviations $\boldsymbol{\sigma}$ (% of $\boldsymbol{\mu}$) = [1.5%, 7.5%, 5%, 5%, 15%, 7.5%, 1%, 10%, 1.5%].

The passive decay heat removal system is considered failed when the temperature of the coolant helium leaving the core (item 4 in Figure 1) exceeds either 1200°C in the hot channel (model output $y_1(\boldsymbol{x})$) or 850°C in the average channel (model output $y_2(\boldsymbol{x})$). The corresponding true value of the functional failure probability $P(F)$ is $1.8089 \cdot 10^{-7}$.

7 RESULTS

The objective is the estimation of the *very small* functional failure probability of the T-H passive system described in Section 6, by means of the AM-SIS algorithm. The benefits coming from the use of the proposed technique are shown by means of a comparison between the estimation *accuracies* and *precisions* of the following methods: (i) standard MCS; (ii) Latin Hypercube Sampling (LHS); (iii) classical IS, where the ISD is centered on the “design point” of the problem identified in the standard Normal space (the approach is hereafter referred to as IS-Des for brevity); (iv) Adaptive Kernel-based Importance Sampling (AK-IS), where an estimator of the optimal ISD is constructed as a weighed superposition of optimized Gaussian kernels centered on failure samples generated by a Markov chain; (v) Line Sampling (LS), where lines are used to probe the failure domain of interest F: an “important direction” is determined to point towards F and a number of conditional “one-dimensional” problems are solved along such direction; (vi) the proposed AM-SIS method.

The AM-SIS algorithm is run with the following set of parameters. N_{TR}^0 and N_{add} are chosen equal to three times the dimension n_i of the input vector $\boldsymbol{x}$ (i.e., n_i = 9 in this case) (Roussouly et al. 2013). Parameter $\varepsilon_{P(F)}{}^{th}$ is used to check the convergence of the ANN-based failure probability estimates $\hat{P}(F)_{ANN}^{N_T}$ during the ANN refinement step. As verified by the authors but not shown here for brevity, the value $\varepsilon_{P(F)}{}^{th} = 0.01$ represents a satisfactory trade-off between ANN accuracy and acceptable computational effort. Finally, parameter K is set to 0.5. In this configuration the SS-based ANN refinement phase converges after t = 14 iterations and 405 calls to the original, long-running T-H model code.

In order to properly represent the randomness of the probabilistic simulation methods adopted and provide a statistically meaningful comparison between their performances, 10000 independent runs of each method have been carried out and in each simulation an estimate of the failure probability $P(F)$ has been computed (in practice, an empirical distribution of the estimator $\hat{P}(F)^{N_T}$ is constructed). The following performance indices have been, then, considered: (i) the expected value $E[\hat{P}(F)^{N_T}]$ of the failure probability estimator; (ii) its standard deviation $\sigma[\hat{P}(F)^{N_T}]$ and (iii) the well-known Figure Of Merit (FOM) defined as FOM = $1/\sigma^2[\hat{P}(F)^{N_T}]\cdot t_{comp}$, where t_{comp} is the time required by the method. Since $\sigma^2[\hat{P}(F)^{N_T}] \propto N_T$ and approximately $t_{comp} \propto N_T$, the FOM is almost independent of N_T. Obviously, the higher the FOM, the higher the computational efficiency of the approach.

Table 1 reports the values of these indicators obtained by simulation methods i)-vi). The number of T-H code runs required by each method is also reported: actually, in the presence of burdensome system model codes (which is often the case for passive safety systems), the total number of code simulations is the *critical* parameter which determines the overall computational cost (i.e., t_{comp}) associated to the method. In detail, $N_{c,P(F)}$ is the number of code runs used by the algorithm *only* to *estimate* the failure probability $P(F)$. On the contrary, $N_{c,aux}$ is the number of *additional* (*auxiliary*) code runs required to "*set up*" the method: for example, for IS-Des, AK-IS and AM-SIS, $N_{c,aux}$ code runs are used to build the corresponding Importance Sampling Densities (ISDs); instead, for LS, $N_{c,aux}$ code runs are used to identify the LS "important direction". In this respect, two considerations need to be made. First, in order for the comparison to be fair, the same number of additional code runs $N_{c,aux}$ has been used for the IS-Des (iii), AK-IS (iv), LS (v) and AM-SIS (vi) methods: this has been achieved by using the *same* ANN metamodel (trained by the adaptive SS-based strategy of Section 5.1) in the "set up stage" of *all* the above mentioned techniques. Secondly, notice that for both the LS and the AM-SIS approaches $N_{c,P(F)} = 2\cdot N_T$ T-H analyses are needed to *estimate* the failure probability $P(F)$ with N_T random samples. This is due to the fact that both methods require the computation of N_T conditional "one-dimensional" failure probability estimates by means of a linear or quadratic interpolation of the T-H passive system performance function $Y(\boldsymbol{x})$: such procedure implies *two* or *three* evaluations of the system model code for *each* random sample drawn. See (Zio and Pedroni 2010) for technical details on this issue.

The performance of AM-SIS is verified by comparison with methods i)–v) in the practically relevant case where the *number* of random *samples* drawn (and, consequently, of *system model evaluations*) is quite *small*: this may be mandatory in practical applications of computer codes requiring several hours to run a single simulation. In particular, we consider here the situation where the number $N_{c,P(F)}$ of T-H code runs "allowed" for estimating the *small* failure probability is set to 20. It is straightforward to note that using $N_T = 20$ simulations no failure samples are obviously generated by standard MCS and LHS, so that the expected value of the corresponding failure probability estimator is equal to 0. This result is reasonable: in fact, the estimation of failure probabilities near 10^{-7} by means of standard MCS and LHS with $N_T = 20$ samples is not efficient, since on average $20\cdot 10^{-7} \sim 0$ failure samples are available in the failure region of interest. On the contrary, the use of ISDs (for IS-Des, AK-IS and AM-SIS) or of preferential lines (for LS) to probe the failure domain F of interest allows *accurate* estimations even with very few samples (and T-H model evaluations): actually, the expected values of the failure probability estimators are very close to the true value $P(F) = 1.8089\cdot 10^{-7}$ for all the mentioned techniques. It can be also seen that AM-SIS provides *more precise* estimates than all the other methods: actually, the standard deviation is about 2 to 10 times lower than that of the other advanced approaches iii)-v), whereas it is about 3 *orders of magnitude* lower than that of the classical sampling schemes i)-ii). Also, the overall computational efficiency of AM-SIS is significantly *higher*: actually, the FOM is 2 to 42 times larger than that of methods iii)-v), whereas it is 6 *orders of magnitude* larger than that of methods i)-ii).

In summary, the results obtained confirm the possibility of achieving accurate and precise estimates of small failure probabilities by the AM-SIS method with a very low number N_T of samples and, in general, with a very low number N_c of calls to the original long-running T-H code: in this latter example, a failure probability of the order of 10^{-7} has been accurately and precisely estimated by resorting to a total of $N_c = N_{c,aux} + N_{c,P(F)} = 405 + 20 = 425$ code runs.

8 CONCLUSIONS

In this work, we have introduced the new Adaptive Metamodel-based Subset Importance Sampling (AM-SIS) method for efficiently estimating very low probability values ($P(F) \approx 10^{-7}$). We have demonstrated it in the functional failure analysis of a natural convection-based passive cooling system of a GFR, under a post-LOCA condition. The performance of AM-SIS has been compared with other advanced simulation methods in a case where the number of random samples drawn is quite *small*. The results have shown that AM-SIS outperforms all the other approaches in terms of (i) *accuracy* and *precision* of the failure probability estimates and (ii) *reduced* overall *computational burden*: in summary, only *few hundreds* of code runs (i.e., around 400) were necessary for 'completing' the two phases of ISD construction and of failure probability evaluation. The outstanding performances of AM-SIS reported in this paper allow drawing a strong, positive conclusion regarding the actual feasibility of application of the method to the *realistic*, *nonlinear* and *non-monotonous* cases of practical interest in the reliability analysis of passive systems.

Table 1. Values of the performance indicators obtained with $N_{c,P(F)}$ = 20 T-H code runs by methods i)-vi) in the estimation of the functional failure probability $P(F)$ of the passive system considered

Functional failure probability ("True" value, $P(F) = 1.8089 \cdot 10^{-7}$)					
			Performance indicators ($N_{c,P(F)}$ = 20; S = 10000 runs)		
Method	N_T	$N_{c,aux}$	$E[\hat{P}(F)^{N_T}]$	$\sigma[\hat{P}(F)^{N_T}]$	FOM
Standard MCS (i)	$N_{c,P(F)} = 20$	0	0	$1.2185 \cdot 10^{-4}$	$6.7352 \cdot 10^{5}$
LHS (ii)	$N_{c,P(F)} = 20$	0	0	$1.1352 \cdot 10^{-4}$	$7.7599 \cdot 10^{5}$
IS-Des (iii)	$N_{c,P(F)} = 20$	405	$1.7641 \cdot 10^{-7}$	$5.6585 \cdot 10^{-7}$	$3.1232 \cdot 10^{10}$
AK-IS (iv)	$N_{c,P(F)} = 20$	405	$1.8097 \cdot 10^{-7}$	$2.7310 \cdot 10^{-7}$	$1.3408 \cdot 10^{11}$
Optimized LS (v)	$N_{c,P(F)}/2 = 10$	405	$1.8155 \cdot 10^{-7}$	$1.2377 \cdot 10^{-7}$	$6.5278 \cdot 10^{11}$
AM-SIS (vi)	$N_{c,P(F)}/2 = 10$	405	$1.8088 \cdot 10^{-7}$	$5.8888 \cdot 10^{-8}$	$1.3108 \cdot 10^{12}$

REFERENCES

Asmussen, S. & Glynn, P.W. 2007. *Stochastic Simulation: Algorithms and Analysis: Algorithms and Analysis, Vol. 57*. New York: Springer-Verlag.

Au, S.K. & Wang, Y. 2014. *Engineering Risk Assessment with Subset Simulation*. Chichester, UK: John Wiley & Sons.

Bassi, C. & Marquès, M. 2008. Reliability assessment of 2400 MWth gas-cooled fast reactor natural circulation decay heat removal in pressurized situations. *Science and Technology of Nuclear Installations: Special Issue "Natural Circulation in Nuclear Reactor Systems"*, Hindawi Publishing Corporation, Paper 87376.

Bect, J., Ginsbourger, D., Li, L., Picheny, V. & Vazquez E. 2012. Sequential design of computer experiments for the estimation of a probability of failure. *Stat Comput* 22: 773–93.

Botev, Z.I. & Kroese, D.P. 2008. An efficient algorithm for rare-event probability estimation, combinatorial optimization, and counting. *Methodology and Computing in Applied Probability* 10(4): 471-505.

Botev, Z.I. & Kroese, D.P. 2012. Efficient Monte Carlo simulation via the generalized splitting method. *Statistics and Computing* 22(1): 1-16.

Botev, Z.I., L'Ecuyer, P. & Tuffin, B. 2013. Markov chain importance sampling with applications to rare event probability estimation. *Statistics and Computing* 23(2): 271-285.

Bourinet, J.-M., Deheeger, F. & Lemaire, M. 2011. Assessing small failure probabilities by combined subset simulation and Support Vector Machines. *Struct. Saf.* 33: 343–353.

Burgazzi, L. 2014. Reliability Study of a Special Decay Heat Removal System of a Gas-Cooled Fast Reactor Demonstrator. *Nuclear Engineering and Design* 280: 473-480.

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals Systems* 2: 303-314.

Dubourg, V., Sudret, B. & Deheeger, F. 2013. Metamodel-based importance samplingfor structural reliability analysis. *Probab. Eng. Mech.* 33: 47–57.

Echard, B., Gayton, N. & Lemaire, M. 2011. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. *Struct. Saf.* 33: 145–154.

Echard, B., Gayton, N., Lemaire, M. & Relun, N. 2013. A combined Importance Sampling and Kriging reliability method for small failure probabilities with time-demanding numerical methods. *Reliab. Eng. Syst. Saf.* 111: 232–240.

Fauriat, W. & Gayton, N. 2014. AK-SYS: an adaptation of the AK-MCS method for system reliability. *Reliab Eng Syst Saf* 123: 137–44.

Fong, C.J., Apostolakis, G.E., Langewish, D.R., Hejzlar, P., Todreas, N.E. & Driscoll, M.J. 2009. Reliability analysis of a passive cooling system using a response surface with an application to the flexible conversion ratio reactor. *Nuclear Engineering and Design* 239(12): 2660-2671.

Hassija, V., Kumar, C.S. & Velusamy, K. 2014. Markov analysis for time dependent success criteria of passive decay heat removal system. *Annals of Nuclear Energy* 72: 298–310.

Hurtado, J.E. 2007. Filtered importance sampling with support vector margin: a powerful method for structural reliability analysis. *Structural Safety* 29: 2-15.

Kersaudy, P., Sudret, B., Varsier, N., Picon, O., Wiart, J. 2015. A new surrogate modeling technique combining Kriging and polynomial chaos expansions – Application to uncertainty analysis in computational dosimetry. *Journal of Computational Physics* 286: 103–117.

Mezio, F. Grinberg, M., Lorenzo, G. & Giménez, M. 2014. Integration of the functional reliability of two passive safety systems to mitigate a SBLOCA+BO in a CAREM-like reactor PSA. *Nuclear Engineering and Design* 270: 109–118.

Morio, J. 2012. Extreme quantile estimation with nonparametric adaptive importance sampling. *Simul Model Pract Theory* 27: 76–89.

Pagani, L., Apostolakis, G.E. & Hejzlar, P. 2005. The impact of uncertainties on the performance of passive systems. *Nuclear Technology* 149: 129-140.

Roussouly, N., Petitjean, F. & Salaun, M. 2013. A new adaptive response surface method for reliability analysis. *Probabilistic Engineering Mechanics* 32: 103–115.

USNRC 2009. Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making. NUREG-1855, US Nuclear Regulatory Commission, Washington, DC.

Valdebenito, M.A., Pradlwarter, H.J. & Schueller, G.I. 2010. The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities. *Structural Safety* 32(2): 101-111.

Zio, E., Apostolakis, G.E. & Pedroni, N. 2010. Quantitative functional failure analysis of a thermal–hydraulic passive system by means of bootstrapped Artificial Neural Networks. *Annals of Nuclear Energy* 37(5): 639-649.

Zio, E. & Pedroni, N. 2010. An optimized Line Sampling method for the estimation of the failure probability of nuclear passive systems. *Reliability Engineering and System Safety* 95(12): 1300-1313.

Zio, E. & Pedroni, N. 2011. How to effectively compute the reliability of a thermal-hydraulic passive system. *Nuclear Engineering and Design* 241(1): 310-327.