AN APPROACH BASED ON SUPPORT VECTOR MACHINES AND A K-D TREE SEARCH ALGORITHM FOR IDENTIFICATION OF THE FAILURE DOMAIN AND SAFEST OPERATING CONDITIONS IN NUCLEAR SYSTEMS

Francesco Di Maio[1], Alessandro Bandini[1], Enrico Zio[1,2], Alfonsi Andrea[3], Cristian Rabiti[3]

[1]Energy Department, Politecnico di Milano, Milan, Italy
[2] Chair System Science and the Energy Challenge, Fondation Electricite' de France (EDF), CentraleSupélec, Université Paris-Saclay, Chatenay-Malabry, France
[3]Idaho National Laboratory, Idaho Falls (ID), U.S.A.

Corresponding Author: Dr. Francesco Di Maio
Tel: +390223996372, Fax: +390223998566, Email: francesco.dimaio@polimi.it

Abstract

The safety of a Nuclear Power Plant (NPP) is verified by analyzing the system responses under normal and accidental conditions. This is done by resorting to a Best-Estimate (BE) Thermal-Hydraulic (TH) code, whose outcomes are compared to given safety thresholds enforced by regulation. This allows identifying the limit-state function that separates the failure domain from the safe domain.

In practice, the TH model response is affected by uncertainties (both epistemic and aleatory), which make the limit-state function and the failure domain probabilistic.

The present paper sets forth an innovative approach to identify the failure domain together with the safest plant operating conditions. The approach relies on the use of Reduced Order Models (ROMs) and K-D Tree.

The model failure boundary is approximated by Support Vector Machines (SVMs) and, then, projected onto the space of the controllable variables (i.e., the model inputs that can be manipulated by the plant operator, such as reactor control-rods position, feed-water flow-rate through the plant primary loops, accumulator water temperature and pressure, repair times, etc.). The farthest point from the failure boundary is, then, computed by means of a K-D Tree-based nearest neighbor algorithm; this point represents the combination of input values corresponding to the safest operating conditions.

The approach is shown to give satisfactory results with reference to one analytical example and one real case study regarding the Peak Cladding Temperature (PCT) reached in a Boiling Water Reactor (BWR) during a Station-Black-Out (SBO), simulated using RELAP5-3D.

Keywords: *Risk-Informed Safety Margins Characterization; Failure Boundary; Reduced-Order Models; Support Vector Machines; K-D Tree; Station Black Out Accident.*

1. INTRODUCTION

The Risk-Informed Safety Margins Characterization (RISMC) pathway of the Light Water Reactors Sustainability (LWRS) program of the U.S. Department of Energy (DOE) [DOE, 2009] aims at developing decision making methods and tools, for use in the process of licensing new nuclear technologies and evaluating existing Nuclear Power Plants (NPPs) for lifetime extension.

One key aspect is the safety assessment, which is performed based on the calculations by a Thermal-Hydraulic (TH) - neutronic code of the nuclear system response in normal and accidental conditions. Specific outputs are selected as safety-significant parameters and their calculated values are compared with some threshold values, in order to check that sufficient safety margins are kept during accident [Gavrilas et al., 2004].

Traditionally, this safety assessment procedure has been performed on a small set of Design Basis Accidents (DBAs) and under tight conservative assumptions (i.e., on the phenomena dynamics described, physical models implemented, etc.) to protect against the uncertainties in the model and its parameters.

In recent times, an extended and more realistic approach has been undertaken, including Beyond Design Basis Accidents (BDBAs) and relying on Best Estimate (BE) codes, in which more realistic assumptions are taken in the evaluation of the safety margins [Zio et al., 2010; Alvarenga et al., 2015]. Under this setting, an accurate and explicit treatment of the uncertainties is required, in order to provide confidence that plant safety margins are not actually reduced [Zio et al., 2008; Apostolakis, 1990; Schuëller et al., 2008].

Such uncertainty quantification has shifted the concept of safety margins to a probabilistic paradigm, whereby the code outcomes are treated as stochastic variables [Zio et al., 2008; Schuëller et al., 2008].

Mathematically, a BE-TH code for safety assessment may be seen as an ensemble of three elements: i) a set of equations coded to describe the system response ii) an *n*-dimensional input vector of stochastic variables $\overline{X} = \{X_1, X_2, \ldots, X_n\}$ and iii) an *o*-dimensional output vector of stochastic variables $\overline{Y} = \{Y_1, Y_2, \ldots, Y_o\}$. The input vector $\overline{X}$ consists of the model parameters and input variables that feed the coded equations to compute the model output vector $\overline{Y}$ that represents the systems response. In mathematical words, a BE-TH code can be seen as the multidimensional and non-linear operator m that maps the input vector $\overline{X}$ into the output vector $\overline{Y}$ [Bourinet et al., 2011]:

$$\overline{Y} = m\left(\overline{X}\right) \tag{1}$$

In general, uncertainties affecting the model outcome may be due to: inherent stochastic behavior of the process described by the model *m* (aleatory uncertainty), imperfect knowledge about the model input variables $\overline{X}$ and lack of information on the underlying physical phenomena (epistemic uncertainty) [Apostolakis, 1990; Möeller et al., 2008; Helton et al., 2011]. Then, mathematically, the input vector $\overline{X}$ is uncertain and, therefore, the output vector $\overline{Y}$ is uncertain as well, with stochastic realizations (in the following, upper case letters are used to identify stochastic variables and lower case letters are used to identify their realizations, as usual):

$$\overline{y} = \{y_1, y_2, \ldots, y_o\} = m\left(\overline{x}\right) = m(x_1, x_2, \ldots, x_n) \tag{2}$$

With reference to a plant accident scenario E_F (i.e., a sequence of events that can (or not) lead to system failure) and to a safety threshold γ_y of the vector $\overline{\gamma}_y$ of safety thresholds, each one of these not to be exceeded by the respective safety parameter $Y \in \overline{Y}$, a limit-state function G can be defined as:

$$G = G(\overline{X}, \gamma_y) = Y(\overline{X}) - \gamma_y \tag{3}$$

The model is in safe operating conditions when $G(\overline{X}, \gamma_y) < 0$ and fails when $G(\overline{X}, \gamma_y) > 0$. Then, $G(\overline{X}, \gamma_y)$ separates the input variables space $\Re^n$ in a safe domain, $S = \{\overline{X} : G(\overline{X}, \gamma_y) < 0\}$, and a failure domain, $F = \{\overline{X} : G(\overline{X}, \gamma_y) > 0\}$. The failure probability, i.e. the probability of occurrence of the plant accident scenario E_F is, then, given by:

$$P(E_F) = P(G(\overline{X}, \gamma_y) > 0) = \int_{G(\overline{X}, \gamma_y) > 0} f_{\overline{X}}(\overline{x}) d\overline{x} \tag{4}$$

where $f_{\overline{X}}(\overline{x})$ is the joint Probability Density Function (PDF) of the stochastic input vector $\overline{X}$ [Cadini et al., 2014]. The set of input values $\overline{X} : G(\overline{X}, \gamma_y) = 0$ defines the failure boundary ∂F

within the input space (i.e., $\Re^n$), for a given value of the safety threshold γ_y. Because the $P(E_F)$ values are low for high-reliable systems (such as NPPs) and the BE-TH models of these systems are computationally expensive, these latter can be replaced by Reduced Order Models (ROMs) to allow the estimation of ∂F within a reasonable computational time [Zio et al., 2008; Chakraborty et al., 2015]. Indeed, ROMs are designed to capture the dominant non-linear behavior of the BE-TH models based on a simplified mathematical representation [Lucia et al., 2004].

In this work, the model failure boundary ∂F is approximated by means of a Support Vector Machines (SVM)-based ROM [Basudhar et al., 2008; Cortes et al., 1995; Guyon et al., 1993] that is embedded in a K-D Tree-based nearest neighbors search algorithm [Bentley, 1975; Katayama et al., 2000; Maneewongvatana et al., 2001] to determine the farthest point from ∂F inside the input space $\Re^n$: this point represents the optimal combination of the model input values that results in the safest plant operating conditions (farthest from the failure boundary) with reference to some given safety requirements expressed by γ_y. The main advantage of adopting SVMs lies in their superior ability, with respect to other ROMs (such as Artificial Neural Networks (ANNs) and simple linear regression models), to define complex decision functions (i.e., hyper-planes) in a multidimensional space and exploit optimal separating functions in order to decompose multiple classes of data [Basudhar et al., 2008; Zio et al., 2012]. On the other hand, the selection of the K-D Tree algorithm as searching algorithm is motivated by the fact that it helps finding the nearest neighbors faster than other brute-force searching approaches [Maneewongvatana et al., 2001]. It is worth pointing out that the K-D Tree algorithm does not requires the SVM to be embedded, but, rather, this can be used as searching algorithm driven by any other ROM for the definition of ∂F.

Knowledge of the safest plant conditions offers practical benefits as $\overline{X}$ is comprised of two different types of inputs: *controllable* and *non-controllable* [Mohsine et al., 2010]. The former identify the levers under control of the plant operator, which can be manipulated to increase plant safety (i.e., reactor control-rods position, feed-water flow-rate through the plant primary loops, accumulator-water temperature and pressure, repair times, etc.), whereas the latter define the random parameters that may (adversely) affect the model response by increasing the likelihood of an accident (i.e., pipelines friction factors, temperature and pressure of the final heat-sink, break section equivalent diameter, failure times, etc.). In this respect, it should be pointed out that were γ_y not "a priori" known but, rather, obeying a probability distribution

$f_{\Gamma_y}(\gamma_y)$, the same input vector realization $\bar{x}$ might imply failure, $g(\bar{x},\gamma_y) > 0$, or success, $g(\bar{x},\gamma_y) < 0$ and, thus, the stochastic safety threshold Γ_y should be included in the *non-controllable* input variables subset [Banks et al., 2011] so that the input space $\Re^n$ is expanded into $\Re^{n+1}$.

Once the *controllable* variables are identified, we can project the failure boundary ∂F on the *controllable* variables space so as to draw "first principles" guidelines for counteracting the incipient plant failure that depends on the occurred accident and the *non-controllable* variables. The rest of this paper is organized as follows. Section 2 illustrates the application of SVMs for the failure boundary estimation. Section 3 shows the approach used to identify the system safest operating conditions. In Section 4, the proposed approach is applied to an analytical example used as proof of concept and in Section 5 it is tested on a Loss of Offsite Power (LOOP) case followed by a Station Black Out (SBO) accident in a Boiling Water Reactor (BWR), whose behavior is simulated by a RELAP5-3D BE-TH code. In Section 6 conclusions are drawn.

2. FAILURE BOUNDARY ESTIMATION

As already said, simulations for the safety assessment of NPPs are computationally expensive due to the small values of $P(E_F)$. As only limited computing resources are generally available, the investigation of an exhaustive set of simulation outcomes, accounting for all normal and accidental plant conditions, is impractical. For this reason, this work exploits a combination of two ROMs to minimize the computational time used to identify ∂F with sufficient accuracy (as later defined in terms of *persistence*):

i. a Physical ROM (P-ROM): a SVM regresses the physical model response of the BE-TH code (see Appendix A for more details on SVMs);
ii. a Boolean ROM (B-ROM): a SVM classifies the P-ROM outputs as belonging either to the safe or failure domain for the identification of ∂F. It is worth mentioning that the B-ROM is not built directly on the physical model responses of the BE-TH code but on the P-ROM responses as this allows speeding-up the B-ROM evaluations by making G smoother and easier to handle.

We adopt an adaptive sampling algorithm [Rabiti et al., 2014a] for the approximation of the model failure boundary ∂F: i) N_R model responses are obtained from the original BE-TH simulations, ii) a P-ROM is built to capture the general BE-TH model behavior, iii) a few new input values are sampled and the BE-TH responses are predicted by the P-ROM, iv) a B-ROM is built to classify the P-ROM outputs as failure or success, iii) new samples are selected based on the B-ROM constructed, v) the B-ROM is iteratively updated based on the P-ROM responses to the new sampled points, until ∂F is identified. This iterative algorithm allows focusing samples on risk-sensitive regions of the input space so that the number of expensive trials needed to localize the boundary is reduced.

The strategy hereby described to estimate ∂F is implemented in the RAVEN code, within a project developed by the Idaho National Laboratory (INL) under the Nuclear Energy Advanced Modeling and Simulation (NEAMS) and Light Water Reactor Sustainability (LWRS) programs to provide software tools for the enforcement of the Risk Informed Safety Margins Characterization (RISMC) conceptual framework supported by the U.S. Department of Energy (DOE) [Rabiti et al., 2014b]. In more detail, the iterative algorithm is comprised of the following steps (without loss of generality, we consider a random safety threshold Γ_y and a single model output Y):

1. at the $\xi = 1$ iteration, a limited number n_0 of points $(x_1, x_2, \ldots, x_n, \gamma_y)^{(1)}, (x_1, x_2, \ldots, x_n, \gamma_y)^{(2)}, \ldots, (x_1, x_2, \ldots, x_n, \gamma_y)^{(n_0)}$ is sampled from the $\Re^{n+1}$ input space through a brute-force approach (i.e., Monte Carlo, grid, stratified sampling, etc.); The sampled n_0 points are, in principle, more than (and different from) the set of input values that have generated the N_R available BE-TH model responses;
2. at each ξ-th iteration, the P-ROM (previously trained on the N_R available BE-TH model responses) is employed to predict $\tilde{y}^{(1)}, \tilde{y}^{(2)}, \ldots, \tilde{y}^{(n_0)}$, which reproduce the BE-TH code responses $y^{(1)}, y^{(2)}, \ldots, y^{(n_0)}$ to the set of n_0 sampled points;
3. a Boolean function $z = z(x_1, x_2, \ldots, x_n, \gamma_y)$ is evaluated on each pair of points $(x_1, x_2, \ldots, x_n, \gamma_y, \tilde{y})^{(1)}, (x_1, x_2, \ldots, x_n, \gamma_y, \tilde{y})^{(2)}, \ldots, (x_1, x_2, \ldots, x_n, \gamma_y, \tilde{y})^{(n_0)}$:

$$z = z(x_1, x_2, \ldots, x_n, \gamma_y) = \begin{cases} 1, & G(x_1, x_2, \ldots, x_n, \gamma_y) < 0 \\ -1, & G(x_1, x_2, \ldots, x_n, \gamma_y) \geq 0 \end{cases} \quad (5)$$

4. a B-ROM is trained on the n_0 points and used to predict $\tilde{z}^{(1)}, \tilde{z}^{(2)}, \ldots, \tilde{z}^{(n_g)}$, e.g., the Boolean responses of $z(x_1, x_2, \ldots, x_n, \gamma_y)$ on a new set of n_g input values $(x_1, x_2, \ldots, x_n, \gamma_y)^{(1)}, (x_1, x_2, \ldots, x_n, \gamma_y)^{(2)}, \ldots, (x_1, x_2, \ldots, x_n, \gamma_y)^{(n_g)}$ that are sampled on a regular Cartesian grid in the input space;
5. the failure domain F is defined by the set of input values resulting in a B-ROM response $\tilde{z}(x_1, x_2, \ldots, x_n, \gamma_y) = -1$. This allows identifying the failure boundary ∂F as the set of input values $(x_1, x_2, \ldots, x_n, \gamma_y)^{(1)}_{\partial F}, (x_1, x_2, \ldots, x_n, \gamma_y)^{(2)}_{\partial F}, \ldots$ that determine the transition of $\tilde{z}(x_1, x_2, \ldots, x_n, \gamma_y)$ from *-1* to *+1*;
6. among points $(x_1, x_2, \ldots, x_n, \gamma_y)^{(1)}_{\partial F}, (x_1, x_2, \ldots, x_n, \gamma_y)^{(2)}_{\partial F}, \ldots,$ the farthest one from $(x_1, x_2, \ldots, x_n, \gamma_y)^{(1)}, (x_1, x_2, \ldots, x_n, \gamma_y)^{(2)}, \ldots, (x_1, x_2, \ldots, x_n, \gamma_y)^{(n_0)}$ is added to the n_0 training data and the algorithm is resumed at Step 2. By so doing, the B-ROM is retrained on a new point in the most risk-sensitive region of the input space (i.e., boundary between system safe and system failure), which is the farthest from the current training data;
7. a *persistence* value $\delta_{\xi}^{(\varphi)}$ is computed for each *φ*-th point of ∂F:

$$\delta_{\xi}^{(\varphi)} = \left| \tilde{z}_{\xi-1}^{(\varphi)} - \tilde{z}_{\xi}^{(\varphi)} \right| \tag{6}$$

   If all $\delta_{\xi}^{(\varphi)}$ are equal to *0* (i.e., any of the ∂F points have changed) for a pre-defined number of consecutive iterations: a) new input points are added to the training set to explore farther regions of the input space from ∂F and b) the process is resumed at Step 2.

When a pre-defined *persistence requirement* is met, the algorithm stops and ∂F is obtained as the set of input points of the B-ROM failure domain ($\tilde{z}(x_1, x_2, \ldots, x_n, \gamma_y) = -1$) and safe domain ($\tilde{z}(x_1, x_2, \ldots, x_n, \gamma_y) = 1$), that determine the transition of $\tilde{z}(x_1, x_2, \ldots, x_n, \gamma_y)$ from *-1* to *+1*.

3. SAFEST OPERATING CONDITIONS IDENTIFICATION

In the most general case, some model input variables are *controllable* (i.e., $X_1, X_2, \ldots, X_q$), while some others are not (i.e., $X_{q+1}, X_{q+2}, \ldots, X_n$). The *controllable* and *non-controllable* input spaces are $\Re^q$ and $\Re^{n-q+1}$, respectively, and the Γ_y has been included in the *non-controllable* input space as, without loss of generality, we consider the model safety threshold as a random variable.

The herein proposed approach for the safest operating conditions identification requires in input:

i. the set of *n+1*-dimensional points of ∂F (that can be estimated as in Section 2, by resorting to a P-ROM and a B-ROM, that in this case have been chosen to be SVMs for regression and classification, respectively);
ii. the distributions of the model input variables (i.e., $f_{X_1}(x_1), f_{X_2}(x_2), \ldots, f_{X_n}(x_n), f_{\Gamma_y}(\gamma_y)$).

In particular, the available information on ∂F (shown in Figure 1 for $n = 2$ *controllable* variables, where dots are safe points ($\tilde{z} = 1$) and stars are failure points ($\tilde{z} = -1$) is, then, manipulated within a K-D Tree algorithm [Bentley, 1975; Katayama et al., 2000; Maneewongvatana et al., 2001], as follows, for a $n+1=3$-dimensional problem with $q = 2$ *controllable* (X_1, X_2) and $n-q+1=1$ *non-controllable* (Γ_y) variables.

In general terms, the K-D Tree algorithm is a space-partitioning data structure for organizing points in a K-Dimensional space [Bentley, 1975]. The K-D tree is a binary tree structure which recursively partitions the input space along the axes that divide it into nested orthotropic regions into which data points are filed. This is done to address the computational inefficiencies of the brute-force Nearest Neighborhood approaches and to reduce the required number of distance calculations by efficiently encoding aggregate distance information for the sample (the basic idea is, indeed, that if point A is very distant from point B, and point B is very close to point C, then, A is distant from C without calculating the distance between A and C). The construction of a K-D tree is very fast: because partitioning is performed only along the axes, no D-dimensional distances need to be computed. Rather, when a (2-D) point is assigned to a node of the tree, the two coordinates are chosen, alternatively, and their medians are calculated to define horizontal or vertical lines, that, recursively, define areas containing other data that are classified in the left and right branches thereby departing that are, respectively, on the left and right of the point corresponding to that node in the 2-D space. Figure 2 shows an intuitive 2-D tree construction for the identification of the nearest safe point to any of the available safe conditions plotted in Figure 1 (i.e., only X_1 and X_2 are considered as input variables for the KD tree construction): the methodological generalization to a higher-dimensional problem is straightforward [Bentley, 1975].

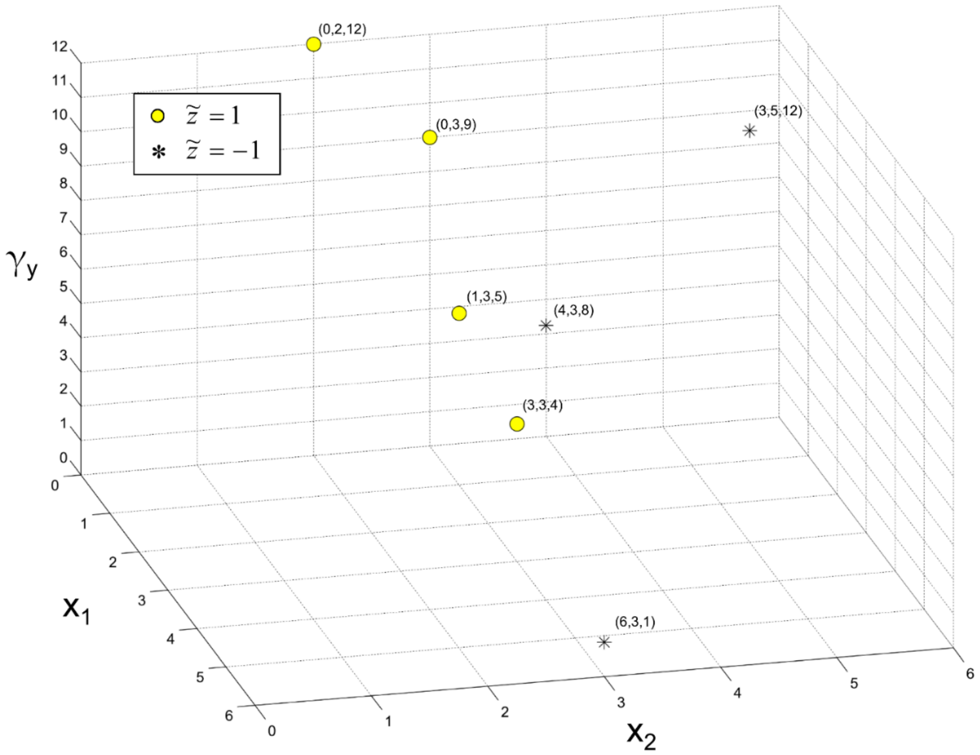


Figure 1: *failure boundary ∂F for $n = 2$ controllable input variables (dots are safe points ($\tilde{z} = 1$) and stars are failure points ($\tilde{z} = -1$)).*

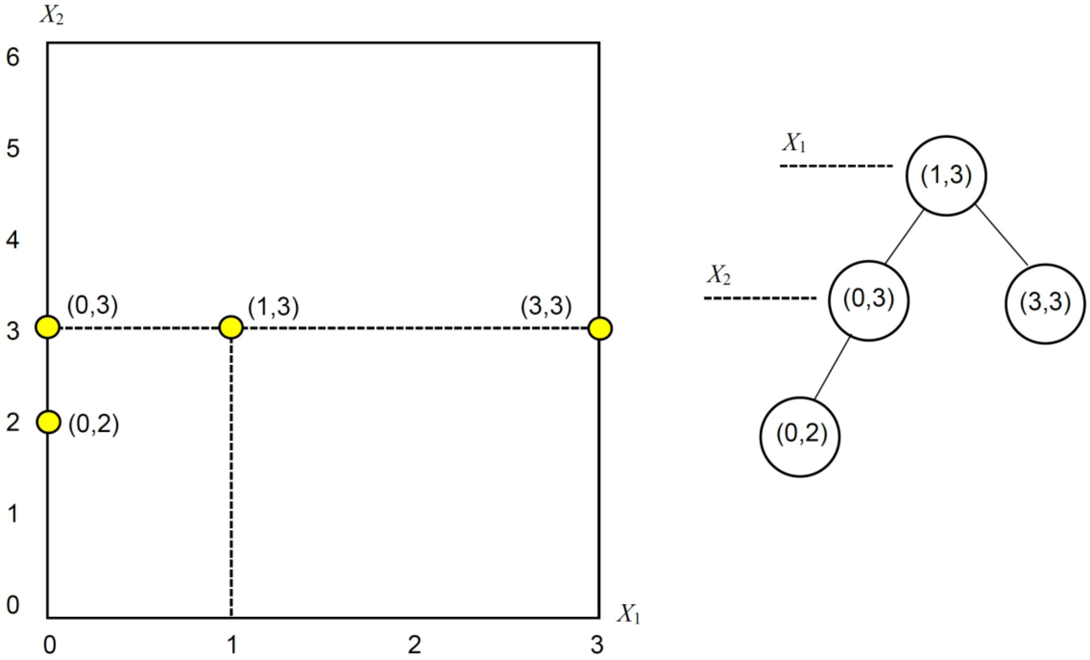


Figure 2: *KD tree construction for the identification of the nearest safe point to any of the available safe conditions plotted in Figure 1.*

The proposed approach can be summarized as follows:

1. One set of values (i.e., $x_1^{(1)}, x_1^{(2)}, \ldots, x_2^{(1)}, x_2^{(2)}, \ldots, x_n^{(1)}, x_n^{(2)}, \ldots, \gamma_y^{(1)}, \gamma_y^{(2)}, \ldots$) is sampled for each input variable from its PDF (i.e., $f_{X_1}(x_1), f_{X_2}(x_2), \ldots, f_{X_n}(x_n), f_{\Gamma_y}(\gamma_y)$);
2. the sampled values of the *controllable* variables (i.e., $x_1^{(1)}, x_1^{(2)}, \ldots, x_2^{(1)}, x_2^{(2)}, \ldots, x_q^{(1)}, x_q^{(2)}, \ldots$) are used to build a *q*-dimensional grid (hereafter called *controllable grid*), whereas the sampled values of the *non-controllable* variables (i.e., $x_{q+1}^{(1)}, x_{q+1}^{(2)}, \ldots, x_{q+2}^{(1)}, x_{q+2}^{(2)}, \ldots, x_n^{(1)}, x_n^{(2)}, \ldots, \gamma_y^{(1)}, \gamma_y^{(2)}, \ldots$) are used to build a *n-q+1*-dimensional grid (hereafter called *non-controllable grid*) (shown in Figure 3 for $n = 2$ and $q = 2$);

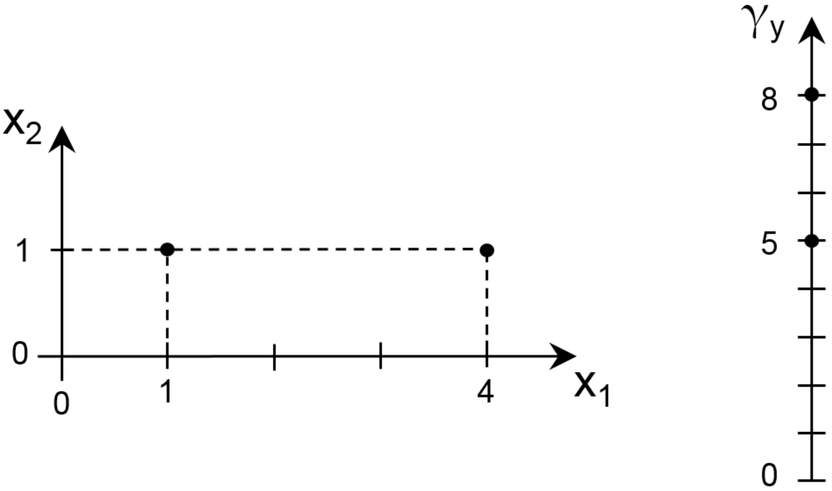


Figure 3: *on the left side, a* $q = 2$*-dimensional controllable grid; on the right side, a* $n - q + 1 = 1$*-dimensional non-controllable grid.*

3. an exhaustive list of pairwise combinations $(x_1, x_2, \ldots, x_q, x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)$ of the *controllable* and *non-controllable* coordinates is built;
4. for each point $(x_1, x_2, \ldots, x_q, (x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)^{\psi})$ belonging to the set of entries $(x_1, x_2, \ldots, x_q, x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)$, which is defined by the same ψ-th set of *non-controllable* variables $(x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)^{(\psi)}$, a K-D Tree-based nearest neighbor algorithm is employed to identify the closest point $(x_1, x_2, \ldots, x_q, x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)_{\partial F}$ of ∂F (Figure 4) for which $\tilde{z}((x_1, x_2, \ldots, x_q, x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)_{\partial F}) = 1$.
5. the projection d on the *controllable* input space (i.e., $\Re^q$) of the Euclidean distance between $(x_1, x_2, \ldots, x_q, x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)$ and $(x_1, x_2, \ldots, x_q, x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)_{\partial F}$ is

computed (Figure 5);

6. the farthest point $\bar{x}_{\psi}^{*} \equiv (x_1^*, x_2^*, \ldots, x_q^* \,|\, (x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)^{(\psi)})$ in the safe domain S from the ψ-th projection of ∂F on $\Re^q$ is identified, which is the point such that $d_{\psi} = \max\{d\}$ In other words, $\bar{x}_{\psi}^{*}$ is the safest point of the *controllable* input space for the ψ-th set of *non-controllable* values $(x_{q+1}, x_{q+2}, \ldots, x_n, \gamma_y)^{(\psi)}$;
7. to each $\bar{x}_{\psi}^{*}$ one probability value P_{ψ}^{*} is associated, which is computed as the product of all the *non-controllable* variables marginal densities:

$$P_{\psi}^{*} = P(X_{q+1} = x_{q+1}^{(\psi)}) \cdot P(X_{q+2} = x_{q+2}^{(\psi)}) \cdot \ldots \cdot P(X_n = x_n^{(\psi)}) \cdot P(\Gamma_y = \gamma_y^{(\psi)}) \tag{7}$$

8. the *absolute* safest position $\bar{x}^{*}$ can be computed as one of the following quantities:
   i. mean:

$$\bar{x}^{*} = \sum_{\psi} P_{\psi}^{*} \cdot \bar{x}_{\psi}^{*} \tag{8}$$

   ii. median:

$$\bar{x}^{*} \equiv \bar{x}_{\psi}^{*} : P(X_{q+1} < x_{q+1}^{(\psi)}, X_{q+2} < x_{q+2}^{(\psi)}, \ldots, X_n < x_n^{(\psi)}, \Gamma_y < \gamma_y^{(\psi)}) = 0.5 \tag{9}$$

   iii. α-th percentile:

$$\bar{x}^{*} \equiv \bar{x}_{\psi}^{*} : P(X_{q+1} < x_{q+1}^{(\psi)}, X_{q+2} < x_{q+2}^{(\psi)}, \ldots, X_n < x_n^{(\psi)}, \Gamma_y < \gamma_y^{(\psi)}) = \frac{\alpha}{100} \tag{10}$$

As it is easy to see, both mean and median of the $\bar{x}_{\psi}^{*}$ population are solutions based on the most probable behavior of the *non-controllable* variables, whereas the α-th percentile defines a more or less risk-oriented solution depending on α and on what *non-controllable* variables are actually considered.

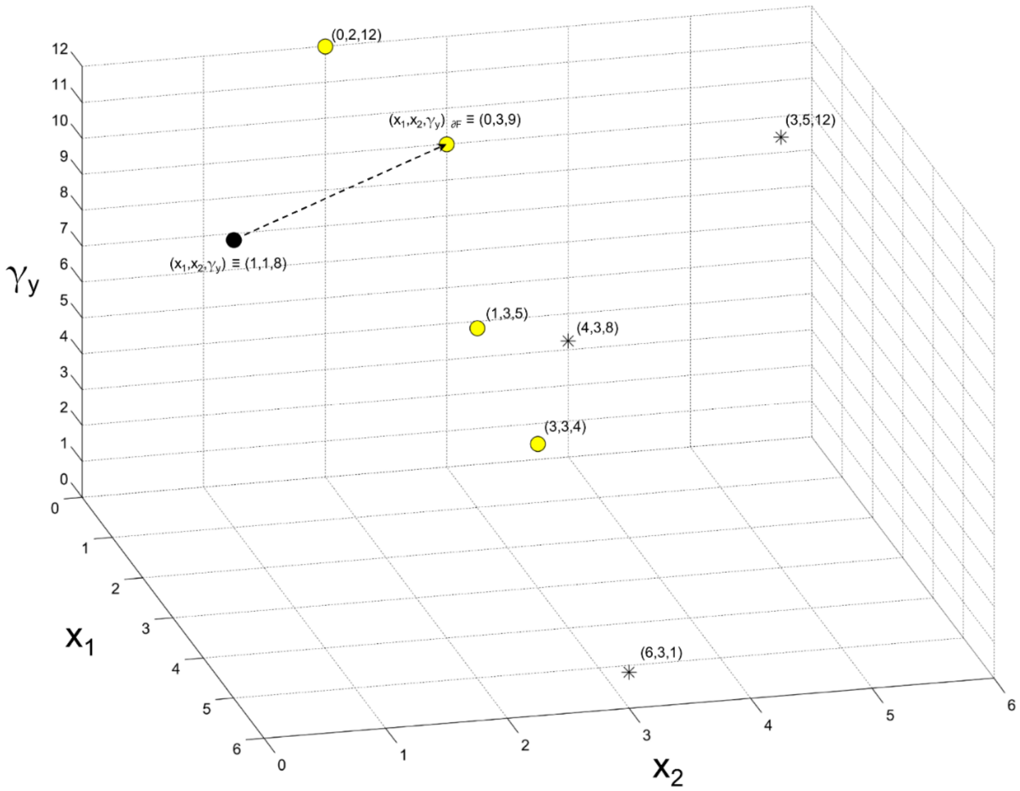


Figure 4: *identification of the nearest neighbor of entry point* $(x_1, x_2, \gamma_y) \equiv (1,1,8)$.

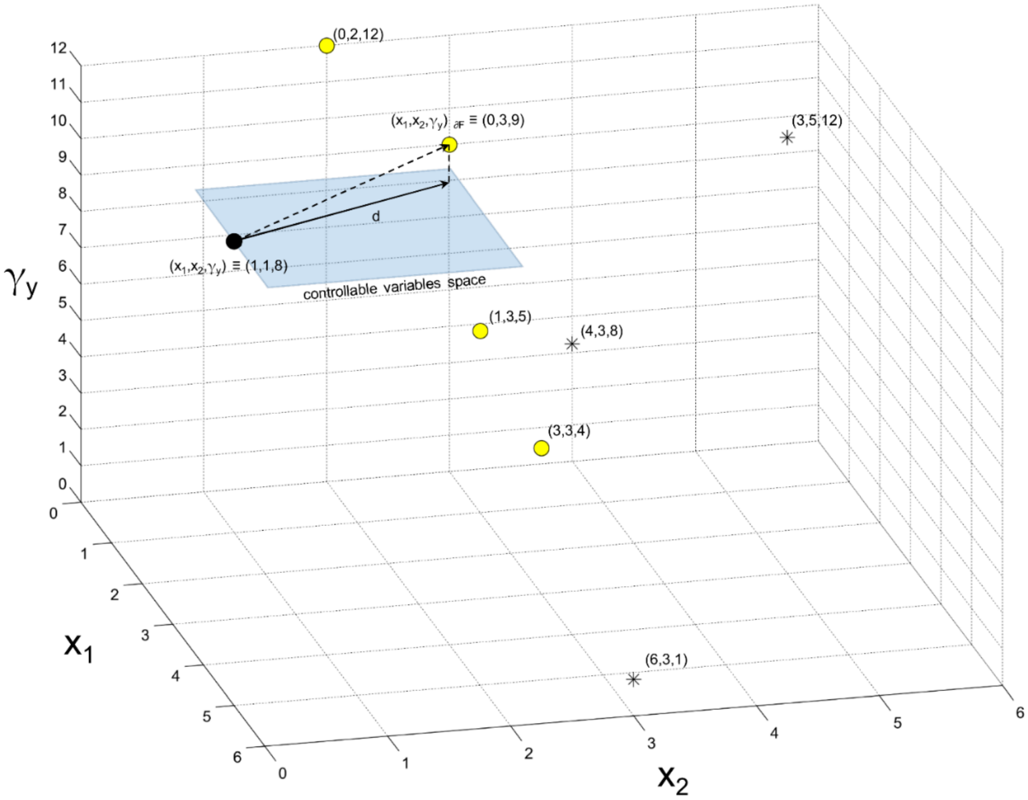


Figure 5: *projected Euclidean distance* d *between entry point* $(x_1, x_2, \gamma_y) \equiv (1,1,8)$ *and its nearest neighbor* $(x_1, x_2, \gamma_y)_{\partial F} \equiv (0,3,9)$ *on the controllable variables space.*

4. PROOF OF CONCEPT USING AN ANALYTICAL EXAMPLE

4.1 Analytical Model Description

The proposed approach is tested on an analytical model m, whose mathematical expression is given as:

$$Y = m(X_1, X_2) = 8X_1^2 - (X_1 - 3)\cdot(X_2 + 2)^2 - (X_1 - 1)^3 \qquad (11)$$

where inputs X_j ($j = 1,2$) are independent random variables obeying two truncated normal distributions: $X_1 \in [-10{,}10] \sim N_1(2{,}4)$, $X_2 \in [-10{,}10] \sim N_1(0{,}6.25)$. The model limit-state function G can be written as:

$$G = G(\overline{X}, \Gamma_y) = Y - \Gamma_y = 8X_1^2 - (X_1 - 3)\cdot(X_2 + 2)^2 - (X_1 - 1)^3 - \Gamma_y \qquad (12)$$

where the model safety threshold is distributed as a truncated normal variable $\Gamma_y \in [-500{,}2500] \sim N_3(500{,}2500)$ and the model failure boundary is $\partial F = \{(X_1, X_2, \Gamma_y) : G(X_1, X_2, \Gamma_y) = 0\}$.

4.2 Failure Boundary Estimation

The methodological steps described in Section 2 have been applied to model m to obtain the estimate $\partial\tilde{F}$ of the failure boundary, where:

1. an initial training set of $n_0 = 11025$ input points $(x_1, x_2, \gamma_y)^{(1)}, (x_1, x_2, \gamma_y)^{(2)}, \ldots, (x_1, x_2, \gamma_y)^{(n_0)}$ is sampled on a regular Cartesian grid $[-10:1:10]\times[-10:1:10]\times[-500:125:2500]$;
2. then, a P-ROM should be trained to reproduce the model m responses $y^{(1)}, y^{(2)}, \ldots, y^{(n_0)}$ to the input set of points $(x_1, x_2, \gamma_y)^{(1)}, (x_1, x_2, \gamma_y)^{(2)}, \ldots, (x_1, x_2, \gamma_y)^{(n_0)}$. However, in this particular analytical example considered, the model m of Eq. (11) and the

corresponding limit state function *G* of Eq. (12) are known, so that we resort directly to Eq. (11) to compute $y^{(1)}, y^{(2)}, \dots, y^{(n_0)}$, instead of training the P-ROM; in other words, simulation data are directly used through Eq. (12) with a sampled safety limit γ_y to identify the set of inputs $(x_1, x_2,..,x_n| \gamma_y)$ that are on the limit surface;

3. a B-ROM, i.e., an SVM-Classifier is trained on the available set of non-linearly separable data with: i) a Gaussian kernel $K_e(\bar{\tau}_\nu, \bar{\tau}) = \exp\left(\frac{\left\|\bar{\tau}_\nu - \bar{\tau}\right\|^2}{\beta}\right)$, where $\bar{\tau}_\nu = (x_1, x_2, \dots, x_q, x_{q+1}, x_{q+2}, \dots, x_n, \gamma_y)^{(\nu)}$, $0 \le \nu \le n_0$, is one of the n_0 training points and $\bar{\tau}$ is the test point to be classified as belonging to the failure or safe domain; ii) a large value of parameter $\beta = 10$ (to assign high influence to each training point $\bar{\tau}_\omega$); iii) a relatively low value of the misclassification cost $C = 10$ (which ensures smoothness of the decision function [Maneewongvatana et al., 2001; Basudhar et al., 2008; Cortes et al., 1995; Guyon et al., 1993]) (see Appendix A for more details on SVMs);
4. the *persistence requirement* on $\delta_\xi^{(\varphi)}$ is set equal to *30*.

The B-ROM estimates the failure boundary $\partial\tilde{F}$ as shown in Figure 6 (where, for clarity, only points $(x_1, x_2, \gamma_y) \in \partial\tilde{F} : \tilde{z}(x_1, x_2, \gamma_y) = 1$ are shown): it is clear that $\partial\tilde{F}$ (dots) well approximates the actual failure boundary ∂F (continuous grid).

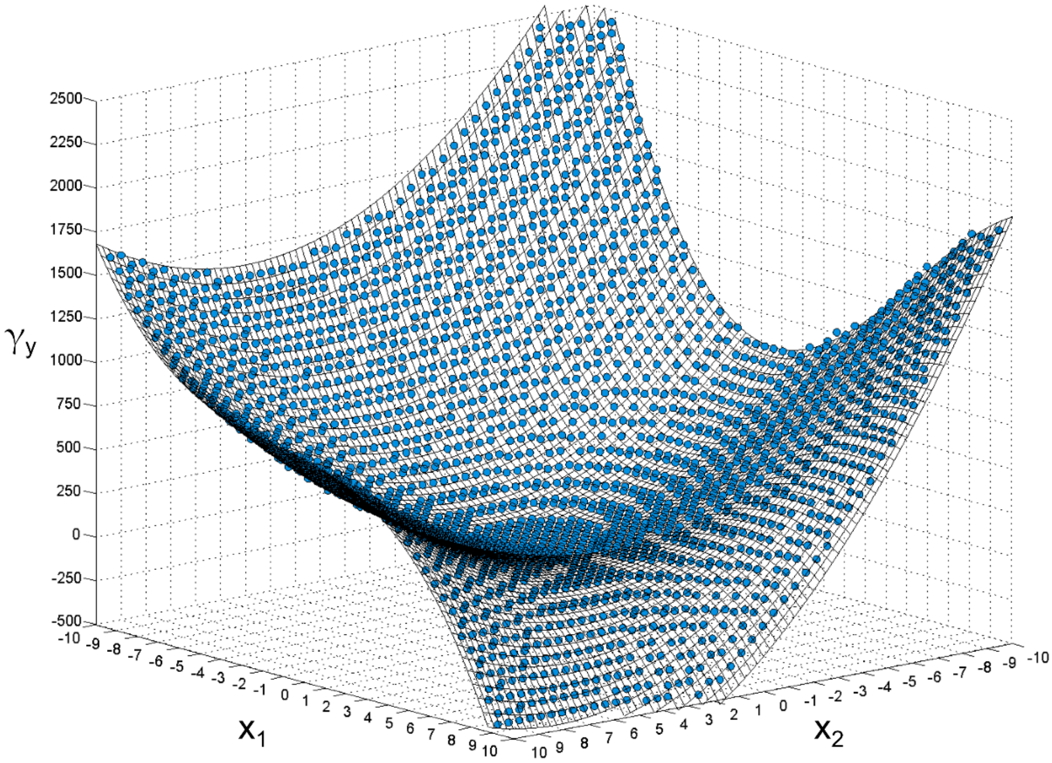


Figure 6*: plot of the estimated $\partial\tilde{F}$ (i.e., dots) and of the actual failure boundary ∂F (i.e., continuous grid) for the analytical model h considered.*

For the case of interest, inputs X_j ($j=1,2$) are considered the model *controllable* variables and Γ_y is the only *non-controllable* variable. Accordingly, the *controllable* and *non-controllable* input spaces are $[-10,10]\times[-10,10]$ and $[-500,2500]$, respectively.

The model *absolute* safest operating conditions $\bar{x}^*$ will be given as pairwise combinations of X_1 and X_2 values (i.e., points in the *controllable space* $[-10,10]\times[-10,10]$), while $f_{\Gamma_y}(\gamma_y)$ will be exploited to assign a P_ψ^* probability value to each *relative* safest operating condition $\bar{x}_\psi^*$ (as shown in detail in Sections 3, Steps from 2 to 7).

4.3 Safest Operating Conditions Identification

In order to identify the safest operating conditions $\bar{x}^*$ of the system whose behavior is modeled by m, the approach proposed in Section 3 has been enforced on the failure boundary $\partial\tilde{F}$ estimated in Section 4.2.

The *controllable* and *non-controllable grids* are built on a Cartesian grid (i.e.,

$[-10:0.5:10]\times[-10:0.5:10]\times[-500:125:2500]$). For each sampled value of Γ_y one projection of $\partial\tilde{F}$ is obtained on the *controllable* input space $[-10,10]\times[-10,10]$ and one *relative* safest position $\bar{x}_\psi^*$ is computed.

Figure 7 shows that: i) the ψ-th projection $\tilde{F}_\psi$ of the estimated failure domain $\tilde{F}$ changes in size and shape with Γ_y (i.e., as Γ_y increases $\tilde{F}_\psi$ decreases as shown in Figure 8), ii) $\bar{x}_\psi^*$ changes its position in the *controllable* input space as $\tilde{F}_\psi$ changes (Figure 8), iii) different statistical quantities (i.e., mean, median, *20*-th and *80*-th percentiles) derived from the population of $\bar{x}_\psi^*$ (the *relative* safest operating conditions for each ψ-th set of *non-controllable* variables) and their associated P_ψ^* values result in different positions for $\bar{x}^*$ (the safest operating conditions) in the *controllable* input space.

In particular, it is worthwhile considering that when the system is operated under very stressful conditions (i.e., system response Y is allowed to approach Γ_y upper limit) as Γ_y is set equal to its *99*-th percentile, for instance, the *relative* safest point $\bar{x}_\psi^*$ might actually be localized within a failure region defined for conservative conditions, i.e., for small values of Γ_y (Figure 7).

For this reason, different strategies should be defined to help the plant operator decide on what safest operating condition $\bar{x}^*$ to select depending his/her attitude towards risk:

i. *risk-averse*: $\bar{x}^*$ is chosen as the *20*-th percentile of the $\bar{x}_\psi^*$ population, which amounts to the safest operating conditions when the system is working under very conservative hypothesis (Y is kept far below the upper limit of Γ_y);

ii. *risk-prone*: $\bar{x}^*$ is the *80*-th percentile of the $\bar{x}_\psi^*$ population, which is the safest operating conditions determined for the system functioning in extreme conditions (Y is allowed getting close to Γ_y upper limit);

iii. mean/median: both strategies aim at identifying $\bar{x}^*$ based on the average behavior of Γ_y.

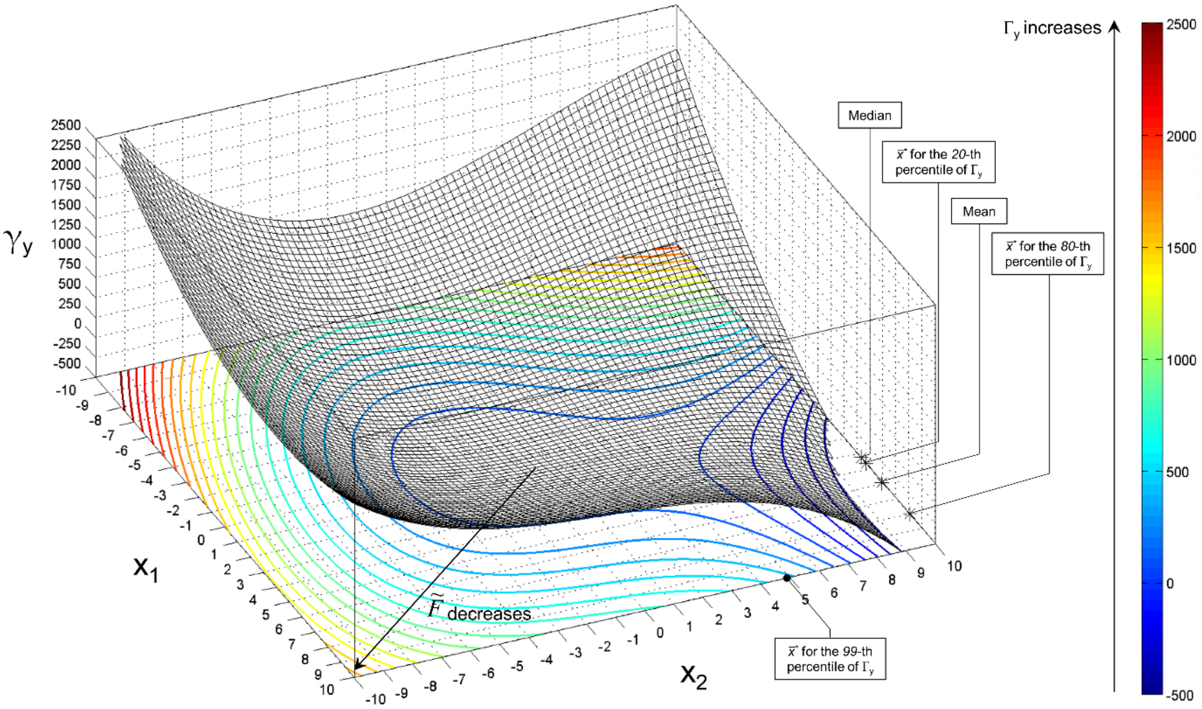


Figure 7: *projections ∂F_ψ of ∂F on the controllable input space and computed absolute safest points $\bar{x}^*$. Notice that ∂F_ψ contours are plotted in continuous lines even if they are only known pointwise, as shown in Fig. 6.*

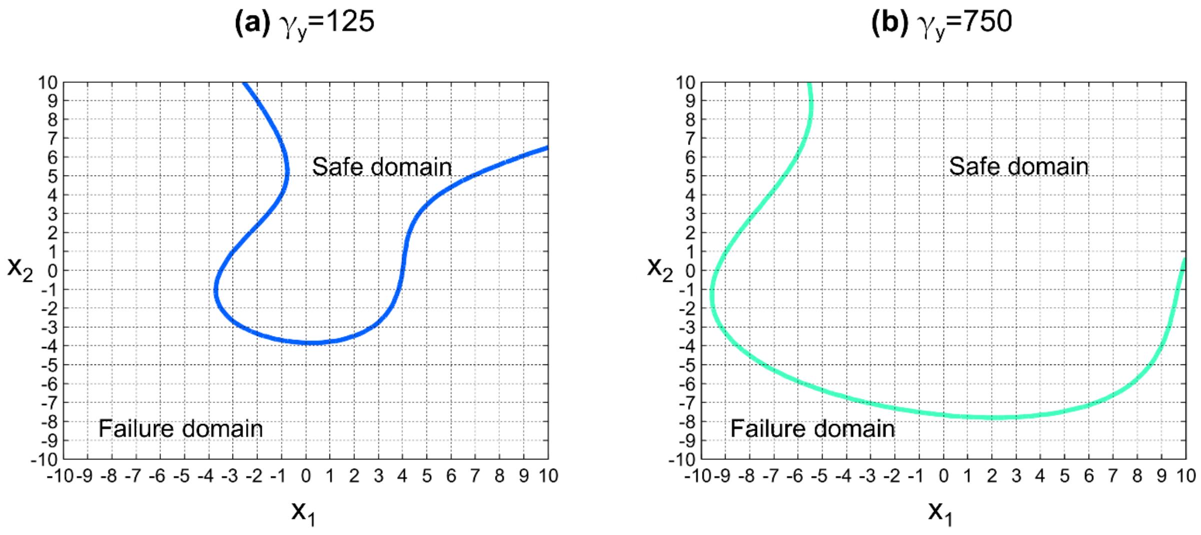


Figure 8: *projections of the failure boundary ∂F on the controllable input space for: (a) $\gamma_y = 125$, (b) $\gamma_y = 750$. Notice that ∂F_ψ contours are plotted in continuous lines even if they are only known pointwise, as shown in Fig. 6.*

As a final remark on the analytical case considered in this Section, it is noted that the 2D cubit test function used is sufficiently non linear to provide a limit surface that has variable features that could be used to benchmark the K-D tree algorithm. For instance, a misclassification cost factor *C*=10 has been selected to ensure the smoothness of the decision function of the B-ROM.

To test the smoothness (or lack of overfitting) of the B-ROM, a random noise variable can be added to Eq. (11) and the analysis can be performed on several training sets with varying magnitudes of the noise variance. If the B-ROM is robust, the limit surface shown in Figure 6 should be insensitive to random noise. However, the demonstration that the training of the B-ROM is robust to random noise and that overfitting can be avoided is not the scope of the work, while it can be found in the literature [Xu et al., 2009].

5. CASE STUDY

5.1 Nuclear Power Plant and Accident Scenario Description

The NPP considered for testing the proposed approach is a Boiling Water Reactor (BWR) with a Mark I containment (Figure 9a). The BWR dynamics has been modeled by the RELAP5-3D code based on the plant nodalization shown in Figure 9b.

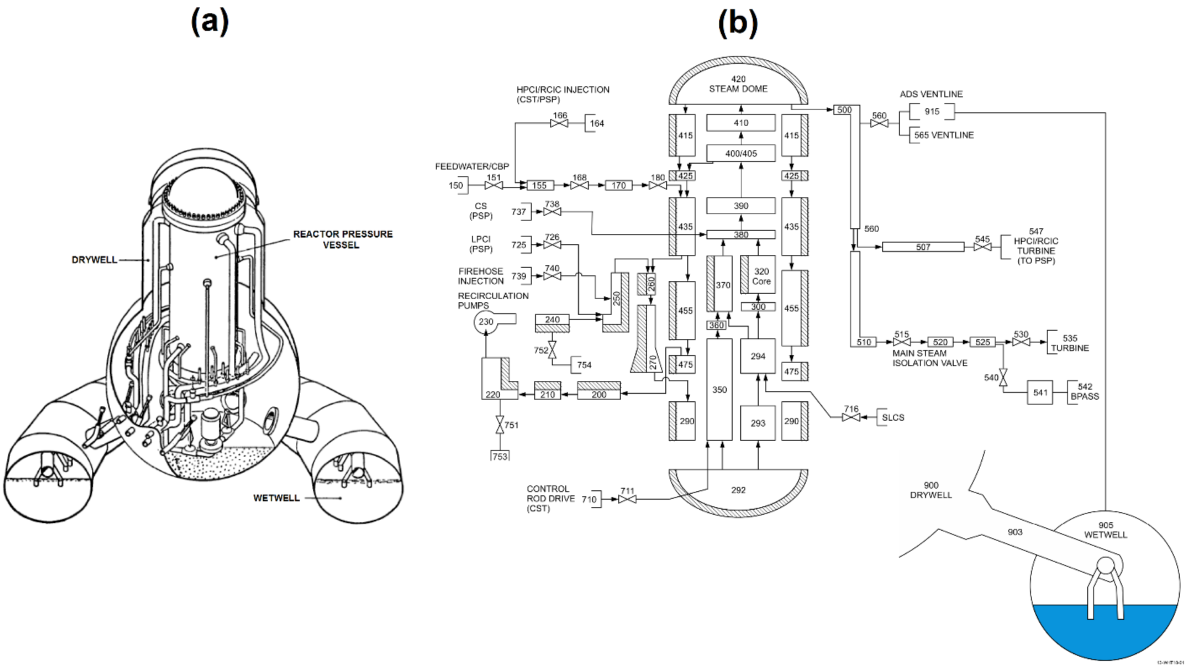


Figure 9: *(a) overview and (b) RELAP5-3D nodalization of the BWR NPP with Mark I containment considered.*

The BWR Mark I primary containment includes the following main components [Mandelli et al., 2013]:

1. a Drywell (DW) comprised of the Reactor Pressure Vessel (RPV) and of a pressurized vessel containing the reactor core and the circulation pumps; within the RPV, a water

level control system includes:

    a. a Reactor Core Isolation Cooling (RCIC) provides high-pressure injection water from the Condensate Storage Tank (CST) to the RPV;
    b. a High Pressure Coolant Injection (HPCI) that is similar to the RCIC but allows for larger water flow rates;

2. a Wetwell (WW) that is a torus-shaped container filled with water that is used as ultimate heat sink;
3. Reactor Circulation Pumps (RCPs).
4. Safety Relief Valves (SRVs) to depressurize the RPV;
5. an Automatic Depressurization System (ADS) that consists in a separate set of relief valves.

The scenario under analysis is a Loss Of Offsite Power (LOOP) followed by the Diesel Generators (DGs) failure, that initiates a Station Black Out (SBO) accident.

In more detail [Mandelli et al., 2013; Mandelli, 2014], LOOP condition occurs due to some Power Grid (PG)-related *external* failure; the recovery of the PG is started and LOOP emergency counteractions are undertaken by the plant operators as follows:

1. the reactor is scrammed and put in sub-critical conditions through full insertion of the control rods in the reactor core;
2. DGs are successfully started so that emergency Alternate Current (AC) power is available;
3. core decay heat is removed by the AC-powered Residual Heat Removal (RHR) system.

SBO condition occurs due to *internal* DGs failure, which renders not possible the removal of decay heat by the RHR; the SBO emergency procedure is immediately enforced by the plant operators as follows [Mandelli et al., 2013]:

1. batteries are activated so that emergency Direct Current (DC) power is available;
2. RPV water level is controlled by RCIC or HPCI;
3. RPV pressure is controlled by SRVs;
4. primary containment is monitored;
5. ADS is activated only if one of the following conditions is reached:
    a. both RCIC and HPCI are disabled;

b. Heat Capacity Temperature Limits (HCTL) are crossed;

c. RPV is depressurized;

6. Firewater (FW) injection is activated only if both the following conditions are fulfilled:

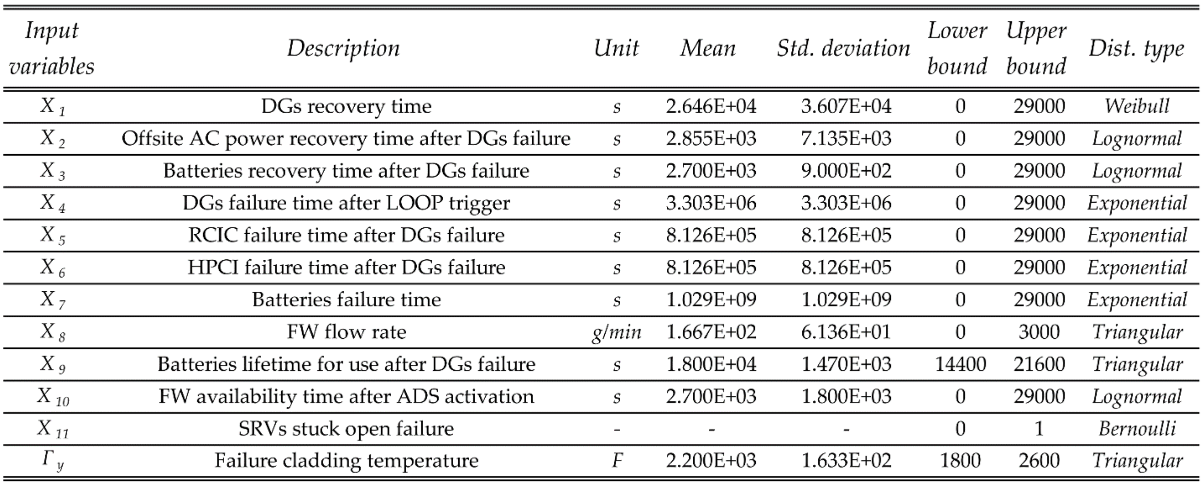
d. all other injection systems are disabled;

e. RPV pressure is below *100* [psi].

Furthermore, we assume that batteries can fail due to the running out of stored power or to *external* failure and, thus, DC power is unavailable. In this case, all control systems are offline causing the reactor core to heat and the PCT to rise. Hence, the DC power recovery process has to be triggered so that if the HPCI or RCIC turbine did not flood during the DC power failure and does not fail on demand, HPCI and RCIC resume normal operations.

The available data consists in $N_R = 10000$ RELAP5-3D code runs that simulate the NPP thermal-hydraulic behavior during LOOP followed by SBO. For each BE-TH code simulation i) *11* input variables (i.e., X_j, $j = 1,2,\dots,11$) and the threshold Γ_y are sampled from their respective PDFs ($f_{X_1}(x_1), f_{X_2}(x_2), \dots, f_{X_{11}}(x_{11}), f_{\Gamma_y}(\gamma_y)$) listed in Table 1, ii) the maximum PCT (i.e., Y) during the SBO transient is computed as safety parameter before [Mandelli, 2014]:

1. Y reaches Γ_y;
2. AC power is recovered by PG or DG resumption;
3. enough core cooling through FW is supplied.



Input variables	*Description*	*Unit*	*Mean*	*Std. deviation*	*Lower bound*	*Upper bound*	*Dist. type*
X_1	DGs recovery time	*s*	2.646E+04	3.607E+04	0	29000	*Weibull*
X_2	Offsite AC power recovery time after DGs failure	*s*	2.855E+03	7.135E+03	0	29000	*Lognormal*
X_3	Batteries recovery time after DGs failure	*s*	2.700E+03	9.000E+02	0	29000	*Lognormal*
X_4	DGs failure time after LOOP trigger	*s*	3.303E+06	3.303E+06	0	29000	*Exponential*
X_5	RCIC failure time after DGs failure	*s*	8.126E+05	8.126E+05	0	29000	*Exponential*
X_6	HPCI failure time after DGs failure	*s*	8.126E+05	8.126E+05	0	29000	*Exponential*
X_7	Batteries failure time	*s*	1.029E+09	1.029E+09	0	29000	*Exponential*
X_8	FW flow rate	*g/min*	1.667E+02	6.136E+01	0	3000	*Triangular*
X_9	Batteries lifetime for use after DGs failure	*s*	1.800E+04	1.470E+03	14400	21600	*Triangular*
X_{10}	FW availability time after ADS activation	*s*	2.700E+03	1.800E+03	0	29000	*Lognormal*
X_{11}	SRVs stuck open failure	-	-	-	0	1	*Bernoulli*
Γ_y	Failure cladding temperature	*F*	2.200E+03	1.633E+02	1800	2600	*Triangular*

Table 1*: input variables list with their associated probability distributions as in [Mandelli, 2014].*

Following [Mandelli, 2014; Sherry et al., 2012], we assume Γ_y to be uncertain and characterized by a triangular probability distribution having: i) a lower limit of *1800* [*F*], ii) an upper limit equal to the Urbanic-Heidrick transition temperature of *2600* [*F*] [Urbanic et al., 1981], iii) the Code of Federal Regulations (CFR) temperature limit of *2200* [*F*] as mode (Figure 10).

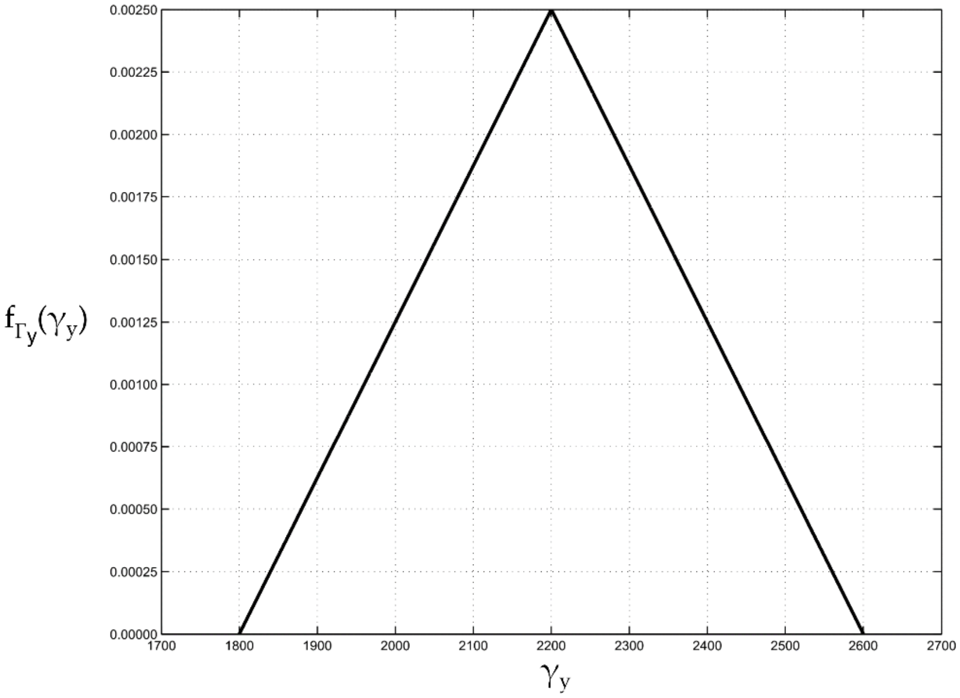


Figure 10: *PDF of the cladding failure temperature* $f_{\Gamma_y}(\gamma_y)$.

Furthermore, a preliminary sensitivity analysis aimed at quantifying the contribution of $X_1, X_2, \ldots, X_{11}$ in affecting the uncertainty of model output Y has been performed based on Sample Pearson Correlation Coefficients (SPCCs) η_j between the *j*-th model input X_j and Y [Stigler, 1989]:

$$\eta_i = \frac{\sum_{\omega=1}^{N_R}(X_{j,\omega} - \overline{X}_j)(Y_\omega - \overline{Y})}{\sqrt{\sum_{\omega=1}^{N_R}(X_{j,\omega} - \overline{X}_i)^2}\sqrt{\sum_{\omega=1}^{N_R}(Y_\omega - \overline{Y})^2}} \qquad (13)$$

where $X_{\omega,j}$ and Y_{ω} ($\omega = 1,2,\ldots,N_R$) are the ω-th sample of X_j and the ω -th computed value of Y, respectively, while:

$$\overline{X}_j = \frac{1}{N_R} \sum_{\omega=1}^{N_R} X_{j,\omega} \tag{14}$$

$$\overline{Y} = \frac{1}{N_R} \sum_{\omega=1}^{N_R} Y_{\omega} \tag{15}$$

According to the sensitivity analysis results shown in Table 2, the three most relevant input variables are found to be "DGs recovery time" (X_1), "Offsite AC power recovery time after DGs failure" (X_2) and "Battery failure time" (X_7). It is worth mentioning that, among the remaining 8 input variables, the "RCIC and HPCI failure times after DG failure", X_5 and X_6 are considered negligible as compared to X_1, X_2 and X_7 even if expected to affect the PCT because the RCIC and the HPCI are two steam-driven systems that inject water to the RPV. Thus, they would be kept constant to their mean values in case additional BE-TH code simulations were run to retrain and improve the P-ROM predictive accuracy. Based on this sensitivity analysis results, the P-ROM and B-ROM are built only on the selected set of input variables rather than on the whole set of 11 input variables listed in Table 1, so as to reduce the burned of additional BE-TH code runs for improving the P-ROM predictive accuracy (reduction of the dimensionality of the input deck of the code, from 11 to 3 input variables). Alternatively, a P-ROM function of all 11 sampled inputs could be obtained and used to generate the 4D data used to train the B-ROM with the remaining 8 variables held at mean values or limiting values, but, in this latter case, any additional BE-TH code run should be fed with an input deck of 11 inputs variables, sampled from the respective distributions, that would increase the computational demand with respect to the use of a P-ROM as a function of 3 sampled inputs.

Furthermore, despite the reduced dimension of the training space (from 11 to 3 input variables) and that the available $N_R = 10000$ RELAP5-3D code runs are not necessary for the training of the surrogate models with sufficient predictive accuracy, we carry on the analysis with all the N_R code runs. The reason is twofold: a very large data set can be very useful for benchmarking purposes and training the ROMs on a large dataset (when available) can overcome the presence of spurious results in the dataset due to RELAP5 numerical instability that provides unphysical results in some flow regimes.

Input variables	Pearson coefficient
X_7	0.331
X_2	0.300
X_1	0.148
X_6	0.096
X_{11}	0.063
X_3	0.061
X_{10}	0.046
X_4	0.023
X_9	0.018
X_5	0.016
X_8	0.010

Table 2: *Sensitivity analysis results.*

It is worthwhile pointing out that the plant emergency staff is free to decide when to start the failed DGs (i.e., X_1) and the offsite PG repair (i.e., X_2), but it is impossible for them to know when the DC batteries may fail (i.e., X_7) and which cladding temperature may cause the system failure (i.e., Γ_y). Because of this, variables X_7 and Γ_y shall be considered *non-controllable* inputs that may adversely impact on the NPP safety, whereas inputs X_1 and X_2 should be identified with the model *controllable* variables. As a result, the *controllable* and *non-controllable* input spaces are $[0{,}29000]\times[0{,}29000]$ and $[0, 29000]\times[1800, 2600]$ (whose units are given in Table 1, 3rd column), respectively.

Similar to the previous case solved in Section 4, the NPP safest operating conditions $\bar{x}^*$ will be identified on the *controllable* input space $[0{,}29000]\times[0{,}29000]$ as the safest pairwise combination of variables X_1 and X_2. On the other hand, the joint PDF $f_{X_7,\Gamma_y}(x_7,\gamma_y)$ of variables X_7 and Γ_y will be used to compute a probability value P_ψ^* for each *relative* safest operating condition $\bar{x}_\psi^*$.

5.2 Failure Boundary Estimation

The failure boundary estimation $\partial\tilde{F}$ is obtained as detailed in Section 2:

1. the training set of $n_0 = 11979$ sampled points $(x_1, x_2, x_7, \gamma_y)^{(1)}, (x_1, x_2, x_7, \gamma_y)^{(2)}, \ldots, (x_1, x_2, x_7, \gamma_y)^{(n_0)}$ is obtained from a regular grid $[0:2900:29000] \times [0:2900:29000] \times [0:2900:29000] \times [1800:100:2600]$;
2. the P-ROM predicting the model responses $\tilde{y}^{(1)}, \tilde{y}^{(2)}, \ldots, \tilde{y}^{(n_0)}$ is a SVM-Regression trained on the N_R values of variables X_1, X_2, X_7 and Y with: i) a Gaussian kernel, ii) $\beta = 25$ and iii) $C = 1500$ (see Section 4.1 and Appendix for more details);
3. the B-ROM is a SVM-Classifier trained on a set of non-linearly separable data with the same specifications as the P-ROM;
4. the persistence requirement on $\delta_{\xi}^{(\varphi)}$ is set equal to *30*.

5.3 Safest Operating Conditions Identification

The NPP safest operating conditions $\bar{x}^*$ are found as in Section 3 on the failure boundary approximation $\partial\tilde{F}$ determined in Section 5.2, where:

1. the *controllable* and *non-controllable grids* are built based on a regular grid $[0:1000:29000] \times [0:1000:29000] \times [0:1000:29000] \times [1800:25:2600]$;
2. for each sampled value of Γ_y one relative safest position $\bar{x}_{\psi}^*$ is computed and localized on the *controllable* inputs space $[0{,}29000] \times [0{,}29000]$;
3. from the population of $\bar{x}_{\psi}^*$ operating conditions and their associated P_{ψ}^* values, a manifold of *absolute* safest conditions $\bar{x}^*$ is obtained:
   i. the *20*-th percentile of the population is selected as *risk-averse* solution: $\bar{x}^*$ is identified with $\bar{x}_{\psi}^* \equiv (0{,}0)$ that is found for the ψ-th combination $(x_7, \gamma_y)^{(\psi)} \equiv (6000{,}2100)$ of the *non-controllable* variables; this means that AC power has to be conservatively resumed right after the DGs failure. The maximum PCT that is reached for this risk-averse emergency strategy (equal to *1011* [F]) is plotted with a star in Fig. 11, where the maximum PCT values reached for all the $N_R = 10000$ are plotted sorted in ascending order (with continuous bold line), together with the uncertain failure threshold Γ_y (shaded

area).

ii. the *80*-th percentile of the population is selected as *risk-prone* solution: it corresponds to the *relative* safest conditions $\bar{x}_{\psi}^{*} \equiv (1000, 2200)$ obtained for $(x_7, \gamma_y)^{(\psi)} \equiv (23000, 2275)$, which implies that AC power can be resumed even *1000* [s] after the DGs failure; the maximum PCT that is reached for this risk-prone emergency strategy (equal to *2208* [F]) is plotted with a square in Fig. 11. It becomes clear that in in this case the plant operator accepts a greater risk, being the actual value of Γ_y “a priori” unknown

iii. the median of the population is a conservative solution given by $\bar{x}_{\psi}^{*} \equiv (0,0)$.

For the case study here considered, Figures similar to Figures 7 and 8 are not provided: the equivalent of Figure 7 would be a 4D input plot (X_1, X_2, X_7 and PCT), with a failure boundary hardly visible, whereas the equivalent of Figure 8 would be a 3D input plot with a collection of points representing the failure boundary for a given PCT, that, on the other hand, would be not informative if not supported by the overall picture of the failure boundaries at different PCT values, as it is for the analytic case study (Figures 7 and 8 are, indeed, jointly described in Section 4.3).

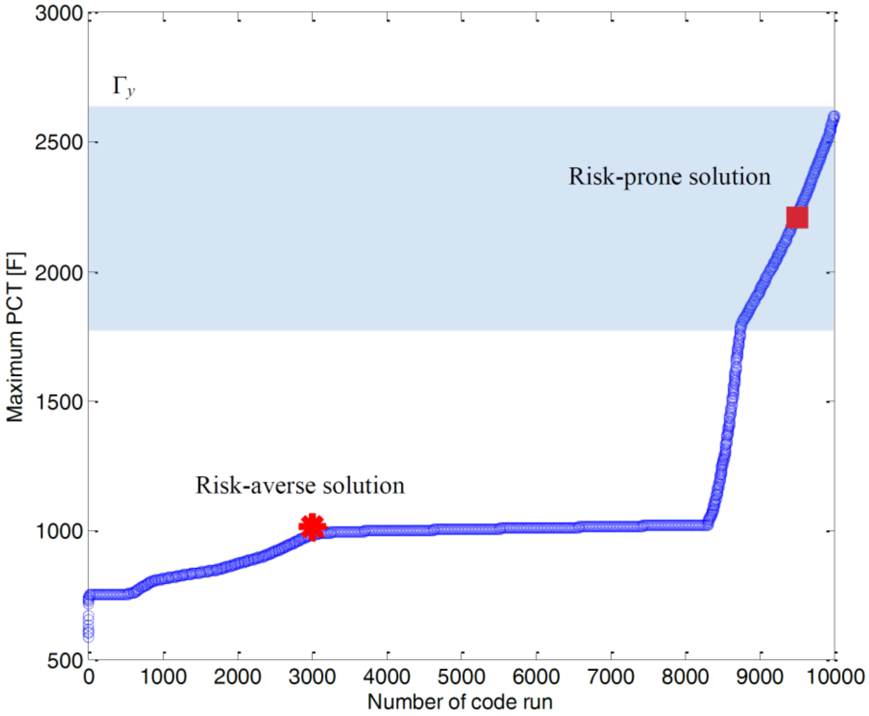


Figure 11*: Maximum PCT values reached for the* $N_R = 10000$ *RELAP runs. The maximum PCT that is reached for the risk-averse emergency strategy is plotted with a star, whereas the maximum PCT that is reached for the risk-prone emergency strategy is plotted with a square. Shaded area is the uncertain failure temperature.*

One can also consider that the estimated failure domain $\tilde{F}$ might not be a compact space and, then, larger percentile values of the $\bar{x}_{\xi}^{*}$ population might not necessarily correspond to larger values of x_1^* and x_2^*, as in this particular case.

As a final remark on the case study considered in this Section, it is noted that a possible demonstration of the efficiency of the adaptive sampling algorithm could be sought by the following methodology:

1) Initialize the training with 50-100 RELAP5 simulations through LHS, grid or stratified sampling (any space-filling experimental design). A total of 50-100 SBO simulations using a system TH code is a practical limit for the number of simulations that can be performed on any modern desktop computer in about one day and for the output to be easily verified by the analyst.

2) Implement the adaptive sampling strategy using batches of 25-50 simulations per batch. The existing 10000 data set is a good benchmark resource to compare with the evolving limit surface.

3) Both the persistence requirement Eq. (6) and an engineering judgment should be used to determine when the algorithm is to be stopped. If the P-ROM or B-ROM is suffering from overfitting, the limit surface may be identified but the algorithm will not converge.

6. CONCLUSIONS

The RISMC program, sustained by the U.S. DOE, is engaged in the development of new methods and tools to support effective decision making on NPPs life extension and licensing of new nuclear technologies.

The present paper represents a contribution to this vast and ambitious program, as it sets forth an adaptive sampling algorithm that embeds a support vector machine (SVM) for multivariate regression, a SVM for classification, and a K-D tree search algorithm for nearest neighbor search in multi-dimensional space to identify the NPP safest operating conditions in the subspace of controllable variables as a function of distance from a limit surface under aleatory and epistemic uncertainties. The partitioning of the model inputs into two subspaces of controllable and non-controllable variables allows, indeed, the non-controllable variables to be treated probabilistically and the safest operating conditions to be defined as a function of risk, being the ultimate goal of the analysis to guarantee the required plant safety margins in accident scenarios. We demonstrate that multivariate parameter sampling, surrogate/reduced order/regression models, classification models and limit surface can yield practical, useful results for the NPPs safety assessment.

The proposed approach has been applied to two case studies, i.e., a proof of concept using an analytical example and a Loss of Offsite Power (LOOP) followed by a Station Black Out (SBO) accident occurring in a Boiling Water Reactor with Mark I containment. In particular, the SBO study involved a 3D SVM for regression, a 4D SVM for classification, and a 4D/2D K-D tree. Coherently with the RISMC main objective, as a result of the application of this suite of algorithms, various options are presented to the analyst to set the NPP in the safest operating conditions: risk-averse, risk-prone decisions have been defined and illustrated, together with a strategy that strikes a balance between these two extremes by identifying the plant "mean" safest operating conditions. Even if practically viable for this case, the extension of the presented approach to higher-dimensional problems should be further investigated from the computational point of view, the K-D tree suffering a curse of dimensionality when dealing with D larger than 20. In any case, as here proposed, a prior sensitivity analysis aimed at reducing the multi-dimensional controllable variable space would tackle the computational problem without the need to resort to other searching algorithms.

ACKNOWLEDGEMENT

The authors would like to thank all the reviewers for their valuable comments to improve the quality of this paper.

APPENDIX

SVMs are a set of supervised learning methods that can be used for: i) classification, ii) regression, iii) outliers or novelty detection [Basudhar et al., 2008; Cortes et al., 1995; Guyon et al., 1993].

When a set of N_τ training points $\bar{\tau}_\nu$ $(0 \le \nu \le N_\tau)$ in a multi-dimensional space is given and each point is associated with one of two classes characterized by a value $z_\nu = \pm 1$, the SVM algorithm finds the boundary (i.e., decision function) that optimally separates the training data into the two classes [Basudhar et al., 2008].

In the case of linear decision functions (i.e., the training data is linearly separable), the idea is to maximize the margin between two parallel hyper-planes that separate the training data. The pair of hyper-planes is required to pass at least through one of the training points $\bar{\tau}_\nu$ of each class (i.e., support vectors) and no points are admitted inside the margin. The optimization problem whose solution determines the optimal pair of hyper-planes is [Basudhar et al., 2008]:

$$\min_{\bar{w},b} \frac{1}{2}\|\bar{w}\|^2 \tag{1A}$$

$$\text{subject to } z_\nu(\bar{w}\cdot\bar{\tau}_\nu + b) - 1 \ge 0,\ 0 \le \nu \le N$$

where $\bar{w}$ is the vector of the hyper-plane coefficients, b the bias and $\frac{2}{\|\bar{w}\|^2}$ the distance between the two hyper-planes. Problem (1A) is a Quadratic Programming (QP) problem that can be solved through the method of Lagrange multipliers λ_ν.

For the case of non-linear decision functions, non-negative slack variables ζ_ν are introduced [Basudhar et al., 2008; Cortes et al., 1995]:

$$\min_{\bar{w},b,\bar{\zeta}} \frac{1}{2}\|\bar{w}\|^2 + C\sum_{\nu=1}^{N_\tau}\zeta_\nu \tag{2A}$$

$$\text{subject to } z_\nu(\bar{w}\cdot\bar{\tau}_\nu + b) - 1 \ge -\zeta_\nu,\ 0 \le \nu \le N_\tau$$

The original set of variables can be mapped to a higher dimensional space, where the classification of any test point $\bar{\tau}$ is obtained by the sign of function:

$$s = b + \sum_{\nu=1}^{N_\tau} \lambda_\nu \cdot z_\nu \cdot K_e(\bar{\tau}_\nu, \bar{\tau}) \tag{3A}$$

Where $K_e(\bar{\tau}_\nu, \bar{\tau})$ is a kernel function. Common types of kernel functions used with SVM are: Gaussian, polynomial kernels, multilayer perceptrons, Fourier series and splines [Basudhar et al., 2008; Guyon et al., 1993].

REFERENCES

[Alvarenga et al., 2015] M. A. B. Alvarenga and P. F. Frutuoso e Melo. Including Severe Accidents in the Design Basis of Nuclear Power Plants: an Organizational Factors Perspective after the Fukushima Accident. In *Annals of Nuclear Energy*, Vol. 79, pp. 68-77. Elsevier, May 2015.

[Apostolakis, 1990] G. E. Apostolakis. The Concept of Probability in Safety Assessments of Technological Systems. In *Science*, Vol. 250, pp. 1359-1364. December 1990.

[Banks et al., 2011] H. T. Banks and H. Shuhua. Propagation of Uncertainty in Dynamical System. In *Proceedings 43rd ISCIE International Symposium on Stochastic Systems Theory and Its Applications*. October 2011.

[Basudhar et al., 2008] A. Basudhar, S. Missoum and A. H. Sanchez. Limit State Function Identification Using Support Vector Machines for Discontinuous Responses and Disjoint Failure Domains. In *Probabilistic Engineering Mechanics*, Vol. 23, pp. 1-11. Elsevier, 2008.

[Bentley, 1975] J. L. Bentley. Multidimensional Binary Search Trees Used for Associative Searching. In *Communications of the Association for Computing Machinery*, Vol. 18, pp. 509-517. September, 1975.

[Bourinet et al., 2011] J. M. Bourinet, F. Deheeger and M. Lemaire. Assessing Small Failure Probabilities by Combined Subset Simulation and Support Vector Machines. In *Structural Safety*, Vol. 33, pp. 343-353. Elsevier, September 2011.

[Cadini et al., 2014] F. Cadini, F. Santos and E. Zio. An Improved Adaptive Kriging-Based Importance Technique for Sampling Multiple Failure Regions of Low Probability. In *Reliability Engineering & System Safety*, Vol. 131, pp. 109-117. Elsevier, November 2014.

[Chakraborty et al., 2015] S. Chakraborty and R. Chowdhury. A Semi-Analytical Framework for Structural Reliability Analysis. In *Computer Methods in Applied Mechanics* *and Engineering*, Vol. 289, pp. 475-497. Elsevier, February 2015.

[Cortes et al., 1995] C. Cortes and V. Vapnik. Support Vector Networks. In *Machine Learning*, Vol. 20, pp. 273-297. Kluwer Academic Publishers, 1995.

[DOE, 2009] VV. AA. Light Water Reactor Sustainability Research and Development Program Plan. Fiscal Year 2009-2013. DOE Office of Nuclear Energy, 2009.

[Gavrilas et al., 2004] M. Gavrilas, J. Meyer, B. Youngblood, D. Prelewicz and R. Beaton. A Generalized Framework for Assessment of Safety Margins in Nuclear Power Plants. In *Proceedings to BE 2004: International Meeting on Updates in Best Estimate Methods in Nuclear Installations Safety Analysis*, pp. 28-36. November, 2004.

[Guyon et al., 1993] I. Guyon, B. Boser and V. Vapnik. Automatic Capacity Tuning of Very Large VC-Dimension Classifier. In *Advances in Neural Information Processing*, pp. 147-155. Morgan Kaufmann, 1993.

[Helton et al., 2011] J. C. Helton and J. D. Johnson. Quantification of Margins and Uncertainties: Alternative Representations of Epistemic Uncertainty. In *Reliability Engineering & System Safety*, Vol. 96, pp. 1034-1052. Elsevier, September 2011.

[Katayama et al., 2000] N. Katayama and S. Satoh. Experimental Evaluation of Disk-Based Data Structures for Nearest Neighbor Searching. In *DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, Vol. 59, pp. 87-104. American Mathematical Society, 2000.

[Lucia et al., 2004] D. J. Lucia, P. S. Beran and W. A. Silva. Reduced-Order Modeling: New Approaches for Computational Physics. In *Progress in Aerospace Sciences*, Vol. 40, pp. 51-117. Elsevier, February 2004.

[Mandelli et al., 2013] D. Mandelli, C. Smith, T. Riley, J. Schroeder, C. Rabiti, A. Alfonsi, J. Nielsen, D. Maljovec, B. Wang and V. Pascucci. Support and Modeling for the Boiling Water Reactor Station Black Out Case Study using RELAP and RAVEN. *Idaho National Laboratory Technical Report*. 2013.

[Mandelli, 2014] D. Mandelli. Simulation of a SBO accident for the BWR Mark I Nuclear Power Plant Model by the RELAP5-3D Code. Private Communication, 2014.

[Maneewongvatana et al., 2001] S. Maneewongvatana and D. M. Mount. On the Efficiency of Nearest Neighbor Searching with Data Clustered in Lower Dimensions. In *Proceedings of the International Conference on Computational Sciences*, pp. 842-851. Springer-Verlag, 2001.

[Mohsine et al., 2010] A. Mohsine and A. El Hami. A Robust Study of Reliability-Based Optimization Methods under Eigen-Frequency. In *Computer Methods in Applied Mechanics and Engineering*, Vol. 199, pp. 1006-1018. Elsevier, 2010.

[Möller et al., 2008] B. Möller and M. Beer. Engineering Computation under Uncertainty –Capabilities of Non-Traditional Models. In *Computer & Structures*, Vol. 86, pp. 1024-1041. Elsevier, May 2008.

[Rabiti et al., 2014a] C. Rabiti, A. Alfonsi, J. Cogliati, D. Mandelli and R. Kinoshita. Introduction of Supervised Learning Capabilities of the RAVEN Code for Limit Surface Analysis. In *Transactions of the American Nuclear Society*, Vol. 110, pp. 355-358. June, 2014.

[Rabiti et al., 2014b] C. Rabiti, A. Alfonsi, J. Cogliati, D. Mandelli and R. Kinoshita. RAVEN, a New Software for Dynamic Risk Analysis. In *Probabilistic Safety Assessment and Management Conference*. June 2014.

[Schuëller et al., 2008] G. I. Schuëller and H. A. Jensen. Computational Methods in Optimization Considering Uncertainties – An Overview. In *Computer Methods in Applied Mechanics and Engineering*, Vol. 198, pp. 2-13. Elsevier, May 2008.

[Sherry et al., 2012] R. Sherry and J. Gabor. Pilot Application of Risk Informed Safety Margins to Support Nuclear Plant Long Term Operation Decisions – Impacts on Safety Margins of Power Uprates for Loss of Main Feed-Water Events. EPRI, December 2012.

[Stigler, 1989] S. M. Stigler. Francis Galton’s Account of the Invention of Correlation.

In *Statistical Science*, Vol. 4, pp. 73-79. 1989.

[Urbanic et al., 1981] V. F. Urbanic and T. R. Heidrick. High Temperature Oxidation of Zircaloy-2 and Zircaloy-4 in Steam. In *Journal of Nuclear Materials*, Vol. 75, pp. 251. Elsevier, August 1981.

[Xu et al., 2009] H. Xu, C. Caramanis, S. Mannor, Robustness and Regularization of Support Vector Machines, *Journal of Machine Learning Research*, Vol. 10, 1485-1510, 2009.

[Zio et al., 2008] E. Zio and F. Di Maio. Bootstrap and Order Statistics for Quantifying Thermal-Hydraulic Code Uncertainties in the Estimation Of Safety Margins. In *Science and Technology of Nuclear Installations*, Vol. 2008, Article ID 340164. Hindawi Publishing Corporation, 2008.

[Zio et al., 2010] E. Zio, F. Di Maio and J. Tong. Safety Margins Confidence Estimation for a Passive Residual Heat Removal System. In *Reliability Engineering & System Safety*, Vol. 95, pp. 828-836. Elsevier, August 2010.

[Zio et al., 2012] E. Zio, F. Di Maio, Fatigue Crack Growth Estimation by Relevance Vector Machines, Expert Systems with Applications, Volume 39, pp. Volume 39, Issue 12, Pages 10681–10692, 2012.