A Framework based on Finite Mixture Models and Adaptive Kriging for Characterizing Non-Smooth and Multimodal Failure Regions in a Nuclear Passive Safety System

A Framework based on Finite Mixture Models and Adaptive Kriging for Characterizing Non-Smooth and Multimodal Failure Regions in a Nuclear Passive Safety System

L. Puppo[1], N. Pedroni[1]*, F. Di Maio[2], A. Bersano[1], C. Bertani[1], E. Zio[2,3,4]

[1] Energy Department, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italy

[2] Energy Department, Politecnico di Milano, Via La Masa 34, Milano, 20156, Italy

[3] MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

[4] Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, South Korea

* Corresponding author: nicola.pedroni@polito.it

Abstract

In the safety analyses of passive systems for nuclear energy applications, computationally demanding models can be substituted by fast-running surrogate models coupled with adaptive sampling techniques; for speeding up the exploration of the components and system state-space and the characterization of the conditions leading to failure (i.e., the system Critical failure Regions, CRs). However, in some cases of non-smoothness and multimodality of the state-space, the existing approaches do not suffice. In this paper, we propose a novel methodological framework, based on Finite Mixture Models (FMMs) and Adaptive Kriging (AK-MCS) for CRs characterization in case of non-smoothness and/or multimodality of the output. The framework contains three main steps: 1) dimensionality reduction through FMMs to tackle the output non-smoothness and multimodality, while focusing on its clusters defining the system failure; 2) adaptive training (AK-MCS) of the metamodel on the reduced space to mimic the time-demanding model and, finally, 3) use of the trained metamodel provide the output for new input combinations and retrieve information about the CRs.

The framework is applied to the case study of a generic Passive Safety System (PSS) for Decay Heat Removal (DHR) designed for advanced Nuclear Power Plants (NPPs). The PSS operation is modelled through a time-demanding Thermal-Hydraulic (T-H) model and the pressure selected for characterizing the PSS response to accidental conditions shows a strong non-smooth and multimodal behavior. A comparison with an alternative approach of literature relying on the use of Support Vector Classifier (SVC) to cluster the output domain is presented to support the framework as a valid approach in challenging CRs characterization.

Keywords: Critical Failure Region characterization; Dimensionality Reduction; Sensitivity Analysis; Finite Mixture Models (FMMs); Kriging; Adaptive Sampling; Adaptive-Kriging Monte Carlo Sampling (AK-MCS); Passive Safety System; Decay Heat Removal.

Symbols

A_{AV}	Activation Valve flow area	Q	Q-function in the FMM approximation
A_{MSIV}	Main Steam Isolation Valve flow area	Q	Predictivity indicator
α	Lagrange multiplier	R	Problem reduced dimensionality
b	Classifier bias parameter	σ	Standard deviation
C	Constant	$\sigma_{\hat{y}}$	Estimation error of a metamodel prediction
$D_{\boldsymbol{X}}$	Input domain	t	Mixture parameters estimate index
D_Y	Output domain	$\boldsymbol{\Theta}$	Set of mixture parameters
DEL_{AV}	Delay of Activation Valve opening	$\widehat{\boldsymbol{\Theta}}$	Estimate of the mixture parameters
DEL_{MSIV}	Delay of Main Steam Isolation Valve closure	Θ	Probability Density function parameters
d	Distance	$\widehat{\Theta}$	Probability Density function parameters estimate
E	Expected value	U	U learning function
E_{ex}	Energy exchanged	W	Conditional expectation of the set of FMMs labels
$E_{ex,\%}$	Percentage of energy exchanged	$\boldsymbol{w}$	Vector of hyperplane coefficients
f	Generic model function	w	Conditional expectation of the FMMs labels
Φ	Mapping function for the Support Vector Classifier Construction	$\boldsymbol{X}$	Generic input
		$\boldsymbol{X^R}$	Generic input vector of reduced dimensionality
H_{jm}	Hellinger distance	$\mathcal{X}$	Set of model input vectors
I	Encoding of information in the Minimum Message Length	$\mathcal{X}^*$	Input vectors of the set of best candidates
		$\mathcal{X}^{Krig}$	Set of input vectors to be evaluated with the Kriging metamodel
i	Input combination index		
j	Cluster index	$\mathcal{X}_{train}$	Set of training input vectors
k	Number of components in the FMMs approximation	$\mathcal{X}_{train}^{SVC}$	Set of training input vectors for the Support Vector Classifier
ker	Kernel		
ℓ	Classifier label	$\mathcal{X}_{val}$	Set of validation input vectors
M	Problem original dimensionality	$\boldsymbol{x}$	Model input vector
m	Input variable index	x	Model input parameter
μ	Mean value	Y	Generic output
$\mu_{\hat{y}}$	Mean value of a metamodel prediction	Y_{thres}	Threshold output value
N_{cand}	Number of best candidates in AK-MCS procedure	$\mathcal{Y}^*$	Outputs of the set of best candidates
N_{MCS}	Number of samples generated by Monte Carlo Sampling	$\mathcal{Y}_{train}$	Set of training outputs
		$\mathcal{Y}_{train}^{FMM}$	Set of training outputs for the FMMs approximation
N_{train}	Number of training samples	$\mathcal{Y}_{train}^{SVC}$	Set of training outputs for the Support Vector Classifier construction
N_{val}	Number of validation samples		
$NC\%$	Non-condensable gases percentage	$\mathcal{Y}_{val}$	Set of validation outputs
n	Iteration number	$\hat{\mathcal{Y}}$	Set of metamodel predictions
n_{fin}	Final number of iterations	$\hat{\mathcal{Y}}_{val}$	Set of predictions of the validation outputs
ξ	Slack term for penalization	y	Model output
p	Probability distribution	$\hat{y}$	Metamodel prediction output
p_{max}	Maximum value of pressure	$\hat{y}^{SVC}$	Support Vector Classifier prediction output
$\boldsymbol{\pi}$	Set of weights	$\bar{y}_{val}$	Average validation output value
π	Probability Density Function weight	$\mathcal{Z}$	Set of FMMs labels
$\hat{\pi}$	Weight estimate	$\boldsymbol{z}$	FMMs label vector
p_{max}	Maximum value of pressure	z	Component of the FMMs label vector

Acronyms

AE	AutoEncoders
AIC	Akaike Information Criterion
AK-MCS	Adaptive Kriging Monte Carlo Sampling
ALK	Active Learning Kriging
ASM	Active Subspace Method
AV	Activation Valve
BE-TH	Best Estimate Thermal Hydraulic
BIC	Bayes Information Criterion
CR	Critical (failure) Region
CV	Cross-Validation
DBSA	Distribution-Based Sensitivity Analysis
DHR	Decay Heat Removal
DS	Directional Sampling
EFF	Expected Feasibility Function
E-HX	Emergency Heat Exchanger
EM	Expectation Minimization
EMO	Evolutionary Multimodal Optimization
FC	Failure Criterion
FMM	Finite Mixture Model
FORM	First Order Reliability Method
GA	Genetic Algorithm
GP	Gaussian Process
I/O	Input/Output
IS	Importance Sampling
KDE	Kernel Density Estimation
LAR	Least Angle Regression
LHS	Latin Hypercube Sampling
LOO	Leave-One-Out
LS	Line Sampling
MAIS	Multimodal Adaptive Importance sampling
MCMC	Markov Chain Monte Carlo
MCS	Monte Carlo Sampling
MfEGRA	Multifidelity Efficient Global Reliability Analysis
ML	Maximum Likelihood
MML	Minimum Message Length
MSIV	Main Steam Isolation Valve
NPP	Nuclear Power Plant
PCA	Principal Component Analysis
PCC	Partial Correlation Coefficient
PCK	Polynomial Chaos Kriging
PCP	Parallel Coordinates Plot
PDF	Probability Density Function
PRESS	Predicted Residual Sum of Squares
PV	Pressure Vessel
PSS	Passive Safety System
QI	Quality Indicator
RBSA	Regression-Based Sensitivity Analysis
RMSE	Root-Mean-Square Error
SA	Sensitivity Analysis
SBO	Station Black-Out
SPLOM	Scatter PLOt Matrix
SRC	Standardized Regression Coefficient
SRV	Safety Relief Valve
SS	Subset Simulation
SVC	Support Vector Classifier
TCR	Truncated Candidate Region
T-H	Thermal Hydraulic
TPI	Transient Performance Indicator
VBSA	Variance-Based Sensitivity Analysis

1 Introduction

Let us assume a system behavior can be modelled with a mathematical Input/Output (I/O) representation $Y = f(\boldsymbol{X})$, where the input $\boldsymbol{X} \in D_{\boldsymbol{X}} \subset \mathbb{R}^M$ represents a given system operational configuration and whose output $Y \in D_{\boldsymbol{Y}} \subset \mathbb{R}$ reflects the system condition/state. For safety/reliability assessment, it is necessary to identify the critical combinations of inputs values (system design and/or operational parameters), which lead the system to failure. In mathematical terms, a specific combination of input parameters $\boldsymbol{x}$ is critical, if the resulting output value is higher (lower) than a predefined threshold, $y = f(\boldsymbol{x}) \geq (\leq) Y_{thres}$, representing the limit value for the system operation. These combinations define the so-called Critical failure Region (CR), i.e., $CR = \{\boldsymbol{x} \in D_{\boldsymbol{X}} \subset \mathbb{R}^M : y = f(\boldsymbol{x}) \geq (\leq) Y_{thres}\}$, whose identification and characterization can be addressed with computational methods: see, e.g., (Cadini et al. 2014; Picheny et al. 2010; Turati et al. 2017; Turati et al. 2018a and b). In these methods, the time-demanding models typically adopted to simulate the system behavior cannot be directly used to numerically

test the system under the many conditions that need to be considered, because the computational cost would be prohibitive for the high number of code runs required. Therefore, new advanced computational methods are being sought to reduce the cost of computation. On one side, fast-running metamodels may be exploited to mimic the behaviour of the time-demanding, original codes and replace them in the analysis. On the other side, adaptive sampling strategies may be adopted to intelligently trace the CR boundary (i.e., the system limit surface), with the minimum waste of computational time for drawing and simulating samples far from the CR.

One of these innovative techniques, known as AK-MCS (Echard et al. 2011), exploiting Kriging metamodeling coupled with adaptive sampling, has been proposed and widely applied for the CRs characterization of systems whose behavior has been assumed to have accommodating properties of regularity, such as continuity and smoothness (Turati et al. 2017; Turati et al. 2018a) . However, several engineering problems showing non-smooth and/or multimodal functional behavior can be found, e.g., in structural and mechanical engineering phenomena like snap-through, buckling or others (Missoum et al. 2002; Hrinda 2010; Boroson and Missoum 2017), challenging traditional smooth metamodels, like Kriging, and possibly lead to large estimation errors (Moustapha and Sudret 2019).

One possible approach developed in recent years to tackle *non-smoothness* and *multimodality*, proposes a clustering of the output domain (Basudhar et al. 2008). This allows separating the different output clusters or, even more generally, to distinguish the regions of different behaviors and to isolate potential discontinuities. For this purpose, a classifier (also called state-selecting model) represents a solution allowing to identify the output clusters (or states), which can be treated separately through different metamodel approximations. In particular, in (Moustapha and Sudret 2019), the "two-stage surrogate modelling" technique is proposed: after determining the domain partitions (e.g., by expert judgment or by an unsupervised clustering technique) and constructing a Support Vector Classifier (SVC) (Vapnik and Cortes 1995), a metamodel is trained for each partition considered interesting to explore. Then, a new input combination ($\boldsymbol{x}$) whose output needs to be predicted (and identified if critical or not) is, first, classified with the SVC (1st stage) and, then, evaluated by the metamodel specifically constructed for the region $\boldsymbol{x}$ belongs to (2nd stage). In the field of *reliability assessment* (resp., failure probability estimation), (Cadini et al. 2014) propose an algorithm combining the First Order Reliability Method (FORM) and an Adaptive Kriging-based Importance Sampling (AK-IS) strategy to deal with *multiple* failure regions characterized by low probability and by complex, non-linear limit states. In (Yang et al. 2018) a two-step algorithm is also developed: in the first step, Active Learning Kriging (ALK) is utilized to recognize the most probable (possibly *disconnected*) failure region(s) of the system; in the second, Kernel Density Estimation (KDE) is employed to build an instrumental density function for IS: the ALK metamodel is then iteratively updated by means of the training points thereby generated by IS. In (Razaaly and Congedo 2018) the objective of estimating small probabilities of *multimodal* failure regions is tackled by an efficient combination of AK-MCS, k-Means clustering algorithm, and Markov Chain Monte Carlo (MCMC) techniques. (Chaudhuri et al. 2021) introduce

the Multifidelity Efficient Global Reliability Analysis (MfEGRA) method, based on a two-stage adaptive sampling criterion that employs a multi-fidelity Gaussian process surrogate to leverage *multiple* information sources with different fidelities (which allows targeting also several, disconnected failure boundaries). (Yang and Cheng 2020) and (Yang et al. 2020) develop an active learning method combining Kriging metamodels (ALK) and Importance Sampling (IS) to analyze systems with very small failure probabilities and multiple failure regions: in particular, evolutionary algorithms from the field of *multimodal optimization* (Cheng et al. 2018) are used to find *all* the *local* and *global* most probable points on the (surrogate) failure boundaries at each iteration of the metamodel training process. (Zhao et al. 2021) presents an Adaptive Multi-Fidelity sparse Polynomial Chaos-Kriging (AMF-PCK) metamodeling for the *global* approximation of aerodynamic data, which proves useful for the efficient uncertainty analysis and optimization of expensive multimodal engineering problems. In this approach, low-fidelity computations are used to build the PCK model as a trend for the high-fidelity function and to capture the relative importance of sparse polynomial bases selected by Least Angle Regression (LAR). Then, high-fidelity model evaluations are employed to adaptively refine a scaling PCK model within an adaptive framework based on correction polynomial expansion-Gaussian process modeling. Finally, in (Zhang et al. 2021) the performance of AK-MCS in dealing with *multiple* failure regions of small probability is improved by combination with Directional Sampling (DS). As a remark, notice that the identification of multiple CRs, which is the task of interest in the present paper, is *different* from the estimation of the system failure probability, which is, instead, the task of the (advanced) techniques reviewed above. The goal in the former task is to identify and characterize the combinations of values of PSS design and/or operational *input* variables which lead to functional failure, that is strictly related to the PSS *thermal-hydraulic behavior*. The objective of the latter task is, instead, to *propagate* the *uncertainty* affecting the *computer code* (e.g., its models, correlations, parameters, …) onto the output of interest and estimating of the functional failure *likelihood*. In this work, we are not performing uncertainty propagation nor failure probability estimation.

A different approach consists in circumventing the dimensionality problem by means of feature selection (Guyon and Elisseeff 2003). Indeed, any metamodel, in general, greatly benefits from a dimensionality reduction (Verleysen and François, 2005; Auder et al., 2012; Gu and Berger, 2016; Turati et al., 2017, 2018a and b; Lataniotis et al., 2020). Moreover, if the analysis is restricted only to those input parameters significantly affecting the output clusters of interest (e.g., the clusters connected with the system failure), also the specific issue of output multimodality can be overcome (Moustapha and Sudret 2019). Feature selection techniques for dimensionality reduction usually rely on many computer simulations, which might become an issue, when the system model is time demanding. Alternatively, Sensitivity Analysis (SA) methods can be employed to achieve the same goal of feature selection by ranking the inputs in terms of their contribution to the model output (Sudret 2008; Borgonovo and Plischke 2016). Several SA techniques are available in literature and they can be subdivided into Local and Global methods (Saltelli et al. 2008), with the latter being more suitable for dimensionality reduction, since they quantify the contribution of each input to the variability of the output over the entire range of values of both the input and the output

(Di Maio et al. 2014). Global SA can be also divided into Regression-Based Sensitivity Analysis (RBSA) methods, also known as non-parametric techniques (Saltelli and Marivoet 1990), such as Standardized Regression Coefficients (SRCs) or Partial Correlation Coefficients (PCCs) (Saltelli et al. 1993); Variance-Based Sensitivity Analysis (VBSA) methods, such as Sobol' method (Saltelli and Sobol 1995; Archer et al. 1997; Nossent et al. 2011); and Distribution-Based Sensitivity Analysis (DBSA) methods, also known as moment-independent methods (Borgonovo and Plischke 2016), such as δ indicator (Borgonovo 2007), input saliency (Law et al. 2004), Hellinger distance (Gibbs and Su 2002) and Kullback-Leibler divergence (Gibbs and Su 2002). However, both the RBSA and VBSA methods, in general, suffer from the output function non-smoothness and/or multimodality (as explained in detail in Section 3.1). On the other hand, DBSA methods become suitable to overcome this issue (Borgonovo et al. 2012): for example, when based on Finite Mixture Models (FMMs), they provide a natural "clustering" of the output (e.g., subdividing the data in groups of large safety margin, low safety margin, failure) that can be used to calculate the SA indexes (Carlos et al. 2013; Di Maio et al. 2015). A synthetic comparison of different SA approaches is reported in Table 1.

Table 1: Comparison among different SA methods to tackle non-smoothness or multimodality

Method	Low cost	Non-smoothness	Multimodality
RBSA	YES	NO	NO
VBSA	NO	YES	NO
DBSA	NO	YES	YES

In particular, FMMs are a flexible and powerful modeling tool for univariate and multivariate data, providing a formal approach to unsupervised learning for statistical pattern recognition. Indeed, FMMs analyze a set of output variables (training set), each one assumed to be generated by a certain random model, i.e. a certain distribution of the mixture (also called component). Then, it infers the distributions parameters and identifies the distribution that originated each training output, leading to a clustering of the training output variables. Moreover, FMMs can be used in support of DBSA methods, aiming at identifying the most relevant input variables affecting the output clusters and, hence, performing a feature selection (Di Maio et al. 2015). Then, the choice of the most appropriate model space (i.e., the space generated by a linear combination of known distributions of a specific kind) and the extraction of the right number of components to approximate the output multimodal distribution remain the challenging tasks to be inferred. Different metrics, based on Maximum Likelihood (ML) estimation, have been developed in the past to guide the model space selection: Minimum Message Length (MML) (Wallace and Boulton 1968), Akaike Information Criterion (AIC) (Akaike 1974) and Bayes Information Criterion (BIC) (Schwarts 1978).

In the present paper, we investigate a novel framework that employs FMMs for the selection of relevant features to be used as inputs to AK-MCS for the CRs characterization of a generic Passive Safety System (PSS) of a Nuclear Power Plant (NPP), based on an Emergency Heat eXchanger (E-HX) designed for the Decay Heat Removal (DHR) after the reactor shut down due to an accident initiation (e.g., Station Black-Out (SBO)).

The application of the proposed framework to a PSS of a NPP is motivated by the growing interest in PSSs employed in advanced NPPs to provide the main safety functions, e.g., reactivity control, decay heat removal and fission product containment, and the need of determining the conditions leading them to failure for safety analysis (Herer et al. 2019). This leads to the necessity of developing methods for CRs characterization, within a more general reliability assessment, to identify the limits of the safe operation of the system (Picheny et al. 2010; Cadini et al. 2014; Turati et al. 2017; Turati et al. 2018a and b; Zio and Pedroni, 2009 and 2011; Zio et al., 2010; Pedroni and Zio, 2017).

In all these works the underlying system function to approximate has, in general, smoothness and unimodality properties. This is also the case in a previous work by the authors (Puppo et al., 2021), where the amount of energy exchanged by a PSS during an accidental transient (E_{ex}) is used to evaluate the success of the PSS intervention. In that case, the AK-MCS technique has proved its capability of successfully replacing the original BE-TH code to model the system response with increasing accuracy in proximity of Y_{thres}. However, in the case here considered, we tackle the problem of the PSS output measuring the maximum pressure value (p_{max}) reached in the Pressure Vessel (PV), which shows a strong *non-smooth* and *multimodal* behaviour (see Fig. 2). Thus, poor results would be obtained if traditional AK-MCS procedure were applied in this case. To address this problem, we develop a novel framework, inspired by that of (Turati et al. 2017), which comprises three steps: i) "dimensionality reduction" carried out through a DBSA method supported by FMMs, to tackle the output non-smoothness and multimodality; ii) "iterative metamodel training" based on AK-MCS, to substitute the computationally expensive model simulations on the reduced input space by means of an accurate Kriging metamodel; iii) "CRs representation and information retrieval" for evaluating a large number of new input combinations with the Kriging model obtained at the previous step to retrieve useful information about the CRs and graphically represent them. The benefit of the proposed framework is twofold: i) speeding up the calculation with respect to the use of the Best-Estimate Thermal-Hydraulic (BE-TH) code available for the analysis of the PSS behaviour, and ii) overcoming the issue of the *non-smoothness* and *multimodality* of the PSS state-space. A comparison (in terms of estimation accuracy and computational cost) with an alternative state-of-the-art approach of *different nature*, i.e., not relying on FMMs-based DBSA but on an SVC to cluster the output domain (Moustapha and Sudret 2019), is presented to show that the proposed framework is valuable for challenging CRs characterization.

The rest of the paper is organized as follows: in Section 2 the case study is briefly presented with a focus on the pressure output distribution in the state-space; Section 3 offers an exhaustive description of the novel framework for CRs exploration in case of output multimodality, whereas, in Section 4, the framework is applied to the PSS described in Section 2 to prove its effectiveness; in Section 5, a comparison with the strategy of clustering the output domain with a classifier is carried out and, finally, in Section 6 some conclusions are drawn.

2 Case Study

The generic PSS considered is a DHR system based on natural circulation and we consider its operation in case of reactor shutdown during a SBO accident, to prevent over-pressurization and over-heating in the PV. A schematic view of the PSS is shown in Fig. 1.

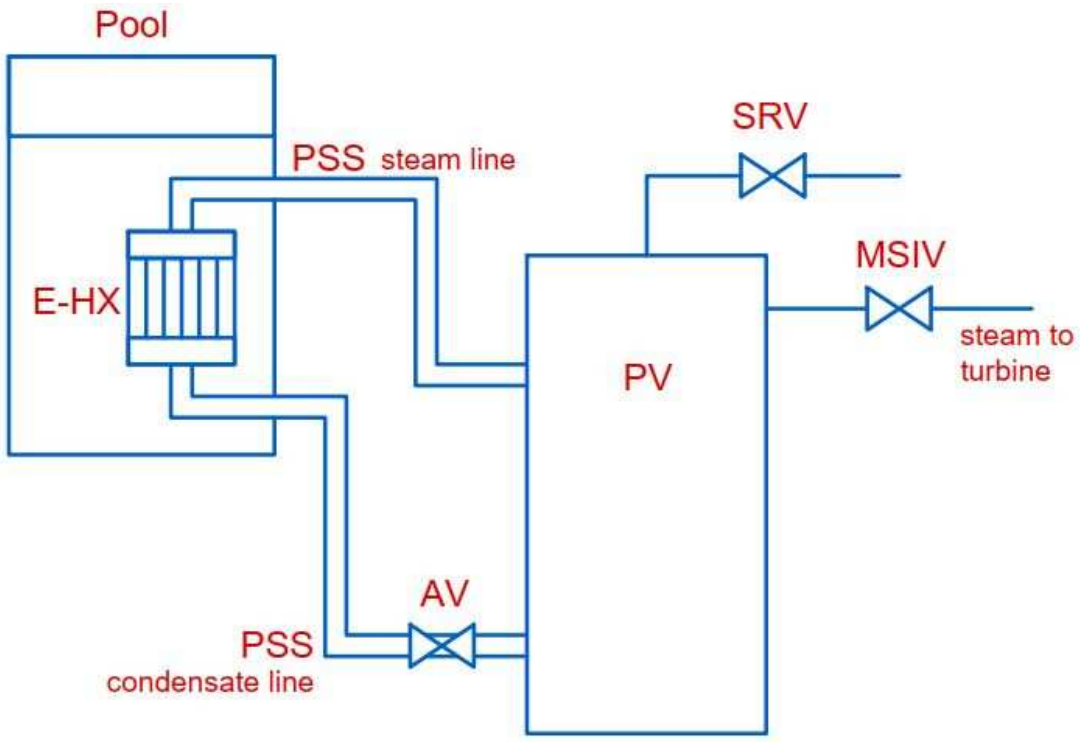

Figure 1: PSS system simplified sketch

At the beginning of the SBO accident, the steam produced in the PV (initially operating in steady state conditions at around 70 bar) is no longer sent to the steam turbine, due to the simultaneous closure of the Main Steam Isolation Valve (MSIV) and opening of the Activation Valve (AV), but instead it is directed to the E-HX through the PSS steam line. The steam is condensed inside the E-HX, which is completely submerged in a pool, and flows back to the PV through the PSS condensate line. For further details about the PSS components and operation see (Lanfredini et al. 2020)

For the reliability analysis of the PSS, five input parameters $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$, with x that is the generic m-th input parameter ($m = 1, \dots 5$), have been identified by the authors through expert judgement as most influential to the system response during SBO accident (Table 2). Uniform probability distributions have been considered to span their ranges of variation and, thus, explore the possible combinations of values in the search for those of the CRs. The corresponding interval bounds have been chosen based on *rough sensitivity calculations* driven by *expert judgment* to obtain a satisfactory balance between two "competing objectives": on one side, the ranges should be large enough to allow a detailed analysis and deep exploration of the failure regions (i.e., to include a relevant number of combinations to failure); on the other side, they should not be too wide, to avoid wasting time in searching way far from the CR.

Table 2: Range of variation of the inputs

Input		Symbol	Range of variation
AV flow area	(%)	A_{AV}	$0 \div 100$
AV opening delay	(sec)	DEL_{AV}	$0 \div 720$
MSIV residual flow area	(%)	A_{MSIV}	$0 \div 0.15$
MSIV closure delay	(sec)	DEL_{MSIV}	$0 \div 7200$
Non-condensable gases percentage in the PSS steam line	(%)	$NC\%$	$0 \div 40$

The PSS response is measured in terms of the amount of decay heat removed during an accidental transient lasting about 8h. If the heat is not removed adequately, temperature and pressure may dangerously rise inside the PV and if the pressure increases beyond the Safety Relief Valve (SRV) set-point assumed at 75.5 bar, the SRV opens to discharge the vapor inside the NPP containment building (not simulated in the model). Two output parameters (Y_1, Y_2), are considered as Transient Performance Indicators (TPIs) (Pierro et al. 2009) to evaluate the PSS functional response, where Y_1 is the total amount of energy removed by the PSS (E_{ex}), and Y_2 is the maximum value reached by the pressure evolution inside the PV (p_{max}).

Table 3 lists the values of the input and output parameters for the reference transients, i.e., the "reference conditions" of nominal functioning of the PSS (note that the total energy exchanged is indicated as percentage $E_{ex,\%}$ with respect to the value obtained in reference conditions).

Table 3: I/O reference conditions

Variable symbol	A_{AV}	DEL_{AV}	A_{MSIV}	DEL_{MSIV}	$NC\%$	$E_{ex,\%}$	p_{max}
Reference Value	100%	0 sec	0.00 %	0 sec	0 %	100%	70.0 bar

In reference conditions, the functions that the system needs to provide are: 1) to ensure $E_{ex,\%} > 90\%$, and 2) to keep p_{max} below 75.5 bar. Therefore, two Failure Criteria (FC) are identified: 1) "Low heat removal", if $E_{ex,\%} < 90\%$ (Pierro et al. 2009); 2) "Steam release in the containment", if $p_{max} > 75.5\ bar$ (i.e., the pressure increase in the PV causes the SRV to open, which leads to vapor release in the NPP containment). In (Puppo et al. 2021), the authors have proposed a metamodel-based AK-MCS framework for the characterization of the CRs for $E_{ex,\%}$ output, with respect to the FC "Low heat removal"; in this present paper instead, the analysis of the CRs related to p_{max}, that shows a non-smooth and multimodal behaviour, has required the development of a suitable exploration framework. In this case, the successful operation of the PSS will be defined when $p_{max} < 75.5\ bar$; otherwise, the system fails providing its function.

A RELAP5-3D model of the PSS has been developed in cooperation by University of Pisa and Politecnico di Torino to simulate the generic PSS connected to a simplified reactor pressure vessel (Lanfredini et al.

2020). Each transient simulation takes about 4.30h on a PC with *CPU Intel Core i7-7500U CPU @ 2.70GHz dual core.*

The Probability Density Function (PDF) of p_{max} is illustrated in Fig. 2, based on a collection of the outcomes of 200 RELAP5-3D simulations. At least two output modes can be identified and hence two corresponding clusters are defined: a first cluster with low pressure values (70.0 bar), which is associated to the majority of the simulations collected; if the decay heat was correctly removed, the pressure should never increase during the accidental transient and hence p_{max} coincides with the pressure value at the beginning of the transient, i.e., $p_{max} = 70.0\ bar$. A second cluster is concentrated around $Y_{thres} = 75.5\ bar$; if the MSIV closes before the AV opening, the decay heat cannot be removed through the E-HX and the vapor builds up in the PV, causing the pressure to rise. In this case, a quite short time interval is sufficient, in which the PV remains without outlets for vapor discharge to cause a sharp pressure increase towards Y_{thres}, with consequent SRV opening. Finally, very few points fall in the middle region showing intermediate values of pressure.

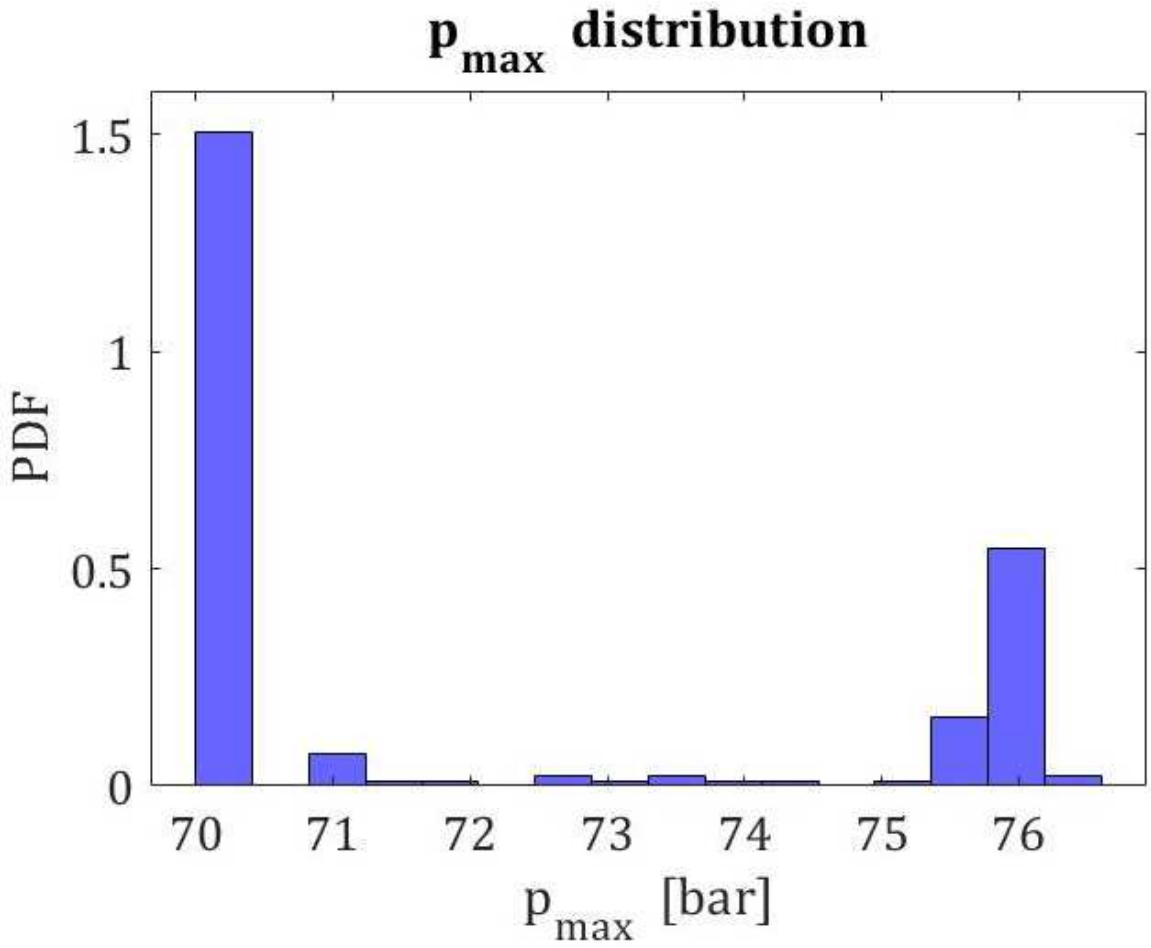

Figure 2: p_{max} multimodal distribution

3 The Proposed Exploration Framework

We propose a novel framework, inspired by (Turati et al. 2017), for exploring the state-space of a system, for which a time-demanding computational model is available and the output is a non-smooth and multimodal function of the input. Firstly, the general idea and purpose of the framework is introduced and, then, the details of the steps concerning its implementation are expanded into the following subsections (3.1, 3.2 and 3.3).

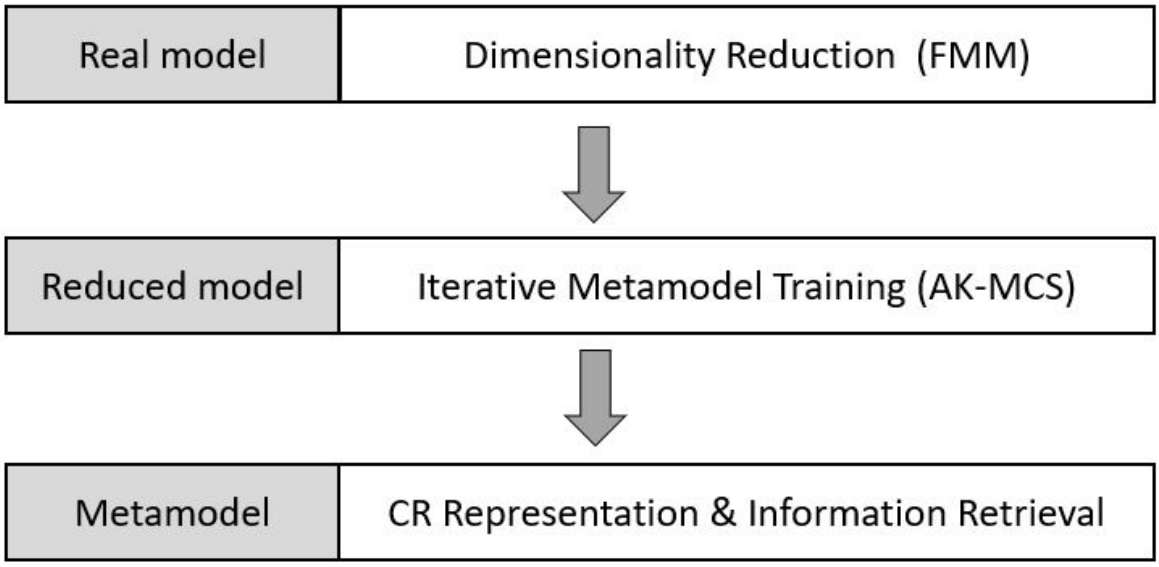

Figure 3: Flow diagram of the proposed framework

The main goal is to iteratively run a (possibly low) number of real model simulations to construct an accurate metamodel not suffering for the output non-smoothness and multimodality. Then, the metamodel is exploited to predict the outputs values for a large number of input values combinations, which are then manipulated to retrieve information about the CRs characteristics. In short, the first step, i.e., "dimensionality reduction", aims at identifying the input parameters most affecting the output distribution, specifically those related to the output clusters in correspondence of the failure threshold and, thus, related to the CRs. For this we resort to a DBSA method supported by FMMs technique. The second step, i.e., "iterative metamodel training" aims at iteratively constructing an accurate and fast-running metamodel to use for simulation on the reduced input space in place of the real model, with specific attention to the boundary of the CRs (limit surface). The metamodel accuracy is verified (e.g., exploiting a validation set) and, then, in the third step, i.e., "CR representation & information retrieval", the metamodel is employed to generate the output values for a large number (several thousands) of new input combinations, which are manipulated to retrieve information about the CRs, like their number and shape, and, finally, to graphically

represent them by exploiting high-dimensional data visualization techniques, like scatter plots or Parallel Coordinates Plot (PCP).

3.1 Dimensionality Reduction

The purpose of dimensionality reduction is to find a lower-dimensional subspace of variables, i.e., $\boldsymbol{X}^R \epsilon D_{\boldsymbol{X}^R} \subset \mathbb{R}^R$ (where $R < M$ is the reduced dimensionality of the problem), to build a reduced model still capable of correctly representing the system behavior (Fodor, 2002; Guyon and Elisseeff 2003; Liu and Motoda, 2012). In several research fields involving *data-driven modeling*, it has been shown that the use of many input variables/parameters often degrades the performance of empirical (regression) models (Verikas and Bacauskiene, 2002; Baraldi et al., 2009; Benkedjouh et al., 2013; Bolón-Canedo et al., 2015; Hu et al., 2017). In general, this is due at least to three reasons: i) irrelevant, non-informative variables result in an empirical model which is not robust; ii) when the empirical model handles many parameters, a large number of observation data is necessary to properly span the high-dimensional input space for accurate multivariable interpolation; and iii) many input features unnecessarily increase the complexity of the data-driven (regression) model. From the point of view of exploring the state-space for CRs characterization (which is of interest in the present paper), reducing the dimensionality allows tackling two issues. First, a *more effective* I/O training set can be defined to construct a more accurate metamodel (Verleysen and François, 2005; Auder et al., 2012; Gu and Berger, 2016; Turati et al., 2017, 2018a and b; Lataniotis et al., 2020). In previous works, some of the authors have already verified the effectiveness of dimensionality reduction for improved metamodel training, in the presence of relatively high-dimensional input spaces (i.e., $M \geq 20$). For example, in (Turati et al., 2017 and 2018b) a power network model involving $M = 20$ inputs has been reduced to $R = 4$ for effectively identifying the electrical feeders' failure times and magnitudes leading the system to the most critical state, i.e., the one with the largest quantity of energy not supplied to the consumers. Also, in (Turati et al., 2018a) the long-running model of a next-generation, lead-cooled fast nuclear reactor involving $M = 32$ inputs (28 parameters related to system neutronics and physics and 4 parameters associated to components' mechanical failures) has been reduced to, again, $R = 4$, in order to precisely bound the regions of reactor safe operation at an affordable computational cost. Second, in case of *non-smooth* and/or *multimodal* output specific attention must be given to the input variables mostly contributing to the definition of the *output clusters* corresponding to system failure conditions: by so doing, also the specific problem of output multimodality can be overcome (Moustapha and Sudret 2019). In this paper, we are particularly concerned with this latter issue.

Several examples are available in literature on how to carry out a dimensionality reduction: in particular, three classes of strategies have been proposed. *Feature selection* aims at optimally identifying a *subset* of the available model input variables and parameters, *most representative* for capturing and describing the overall system behaviour (Guyon and Elisseeff, 2003; Saeys et al., 2007). A feature selection technique can be seen as the combination of a (possibly burdensome) search algorithm for proposing multiple diverse

feature subsets, along with an evaluation metric (e.g., a regression error), which scores the different feature subsets with respect to their representativeness (Dy and Brodley, 2004; Zhang et al., 2015). Instead, *feature extraction* aims at identifying a set of "*new*" features (i.e., new input parameters or variables), generated by *transformations* of the initial ones (in other words, generating a *new*, *lower-dimensional* input subspace as a linear or nonlinear function of the original one) (Guyon and Elisseeff, 2006). Some of the most effective and widely used feature extraction techniques are Principal Component Analysis (PCA) (Jolliffe, 2002; Higdon et al., 2013; Van Der Maaten et al., 2009; Wu et al., 2018; Nagel et al., 2020), the Active Subspace Method (ASM) (Constantine, 2015; Erdal and Cirpka, 2019) and AutoEncoders (AEs) (Holden et al., 2006; Wang et al., 2016; Monisha et al., 2019; Roma et al., 2021). Finally, *Sensitivity Analysis* (SA) methods have the same objective as feature selection, but they achieve it by ranking the input parameters and variables according to their *influence* on the *outputs* of the model (Borgonovo & Plischke, 2016; Saltelli et al., 2008; Sudret, 2008). In this paper, SA-based techniques are chosen for two reasons: i) as for feature selection, they retain a subset of the *original* input parameters/variables (without performing any transformation on them), which allows a more direct physical interpretation of the PSS critical regions (coherently with the main scope of the work) and ii) differently from feature selection approaches, they do *not* require the solution of typically burdensome *optimization problems* to search for the best and most representative subset of inputs (which is a relevant issue in the presence of long-running codes). Among the SA techniques, it is possible to identify two families: Local and Global. The Local approach to SA considers small variations of each input parameter around its nominal value, whereas Global SA allows to quantify the contribution of an input to the variability of the output, computed over the entire range of both the input and the output (Saltelli et al. 2008). Global SA offers higher capabilities than Local SA, especially when model responses are not regular (e.g., non-linear and non-monotonic), but at a higher computational cost (Di Maio et al. 2014). Global SA methods can be divided into three categories (Borgonovo and Plischke 2016): 1) RBSA methods, 2) VBSA methods and 3) DBSA methods. RBSA or non-parametric methods (Saltelli and Marivoet 1990) exploit regression techniques to fit a regression model on a set of I/O relations and to use the regression coefficients as indices of sensitivity. RBSA methods are typically the simplest ones, also associated to the lowest computational cost, but their performance strongly depends on the output form which is often required to be linear. Indeed, if the regression model does not fit the underlying I/O relationships (e.g., in case of non-smoothness), the SA performs poorly. VBSA methods (Saltelli et al. 2010) quantify the contribution of each input parameter (first-order effect) and each possible two- or high-order interaction among multiple parameters to the total output variance. The ratio of such contribution to the total variance is taken as sensitivity coefficient (Razavi and Gupta 2015). VBSA methods are the most widespread, because they do not introduce any hypothesis on the model since they do not carry out any approximation of it. Anyway, VBSA methods are unable to distinguish between output structures (i.e., how the output values are organized in the state-space) with identical global variances, but different distributions and spatial organizations (Razavi and Gupta 2015). Thus, they may suffer for output multimodality since, by definition, the calculation of variance in case of a multimodal variable is not trivial. On the other side, DBSA or moment-

independent methods (e.g., Hellinger distance, Kullback-Liebler divergence etc.) (Di Maio et al. 2014) rank the input variables most affecting the *entire* output distribution and they may overcome the issue of non-smoothness and multimodality, if the output distribution is properly approximated, despite its irregular form, by means of FMMs technique (Di Maio et al. 2015). FMMs are classically implemented for pattern recognition to approximate the output distribution, even in case of multimodality, by identifying the output clusters (corresponding to the different output modes) and, hence, representing the output as a linear combination of known distributions, also called components (e.g., Gaussian, Exponential, etc.). Anyway, FMMs can be also adopted as a support for SA: indeed, the output clustering is mapped to the input space and, in the end, the contribution of each input to the clustering of the output is ranked according to the different DBSA methods.

FMMs application for SA entails following at the beginning the same procedure adopted in case of the more general pattern recognition: the primary goal is to find the appropriate type and number of components (k) to approximate the output distribution, given a set of I/O relations (see Appendix A). The best k is historically determined through the application of the Expectation Minimization (EM) algorithm (Dempster et al. 1977). However, classical EM presents several drawbacks: it is a local method, thus, it is sensitive to initialization and, for certain kinds of mixtures, it may converge toward an estimate at the boundary of the parameter space where the likelihood is unbounded (Figueiredo and Jain 2002). For the case study of p_{max}, we propose the SNOB algorithm, introduced for the first time in (Wallace 1968) and, then, updated through the years and implemented in MATALAB by (Statovic 2020). It exploits the MML inference criterion:

$$I(\boldsymbol{\Theta}|\mathcal{Y}_{train}^{FMM}) = I(k) + I(\boldsymbol{\pi}) + \sum_{j=1}^{k} I(\Theta j) + I(\mathcal{Y}_{train}^{FMM}|\boldsymbol{\Theta}), \quad (1)$$

where $\mathcal{Y}_{train}^{FMM} = \{y_1, \dots, y_n\}$ are the output values of the transients simulated and $\boldsymbol{\Theta} = \{\pi_1, \dots, \pi_k, \Theta_1, \dots, \Theta_k\}$ are the mixture parameters (π_j and $p(y|\Theta_j)$ are the weight and PDF of the j-th component, respectively). The output approximation is encoded in a message, which comprises all its terms. The lower is the encoding of this information, i.e., $I\left(\boldsymbol{\Theta}\middle|\mathcal{Y}_{train}^{FMM}\right)$, the lower is the message length and, hence, the more accurate is the output distribution approximation with that mixture of components (Kasarapu and Allison 2015). In particular, $I(k)$ represents the encoding of the number of components (k), $I(\boldsymbol{\pi})$ the encoding of the weights ($\boldsymbol{\pi}$), $\sum_{j=1}^{k} I(\Theta j)$ the encoding of the component parameters (Θ_j) and $I\left(\mathcal{Y}_{train}^{FMM}\middle|\boldsymbol{\Theta}\right)$ the encoding of the data. All these terms are logarithmic and in the most favorable situations they could assume negative values. To sum up, the MML criterion (1) is a trade-off between the complexity of the model and the goodness of fit (Olvier et al. 1996); indeed, when a new component is added, the encoding of the new component parameters increases the message length, whereas the term $I\left(\mathcal{Y}_{train}^{FMM}\middle|\boldsymbol{\Theta}\right)$ reduces it due to the improved fit quality.

The SNOB algorithm allows the user to choose among several types of distributions (i.e., model space), e.g., Gaussian, Weibull, Exponential etc. The algorithm automatically finds the best k according to the distribution types and provides in output the MML metric that can be used to justify the model space selection. The solution associated to the lowest MML value is the most accurate for the case study.

Once the parameters of the mixture of models are known, the output distribution is completely characterized: some of the clusters obtained may be representative of safe conditions, whereas others represent failure conditions. For Global SA, the focus is shifted to the input space and the output clustering is exploited to cluster also the inputs. The PDFs of each input variable (x_m) with the conditioning on the different j-th clusters are constructed, i.e., $p(x_m|\,\Theta_{jm})$, and, then, the difference between $p(x_m|\,\Theta_{jm})$ and the input common distribution, i.e., $p(x_m)$ is measured according to one of the DBSA methods introduced before (e.g., Hellinger distance, Kullback-Liebler divergence). These measures allow ranking the input variables contribution to the different output clusters, with special attention to the clusters of interest, e.g., those related to the failure of the system, and, finally, the most important inputs are selected.

3.2 Iterative Metamodel Training (AK-MCS)

After reducing the number of input parameters through the dimensionality reduction, a surrogate metamodel is constructed to approximate the real model I/O relationships on the reduced input space, i.e., $Y = f(\boldsymbol{X}^R)$, where $\boldsymbol{X}^R \in D_{\boldsymbol{X}^R} \subset \mathbb{R}^R$ ($R < M$ is the dimensionality of the reduced space). Among the several options available in literature (Jin et al. 2001), we resort to Gaussian Processes (GPs) and particularly to one specific category of GPs: the Kriging metamodels (Kleijnen 2009). Kriging metamodels can fit numerous response functions without adding further complexity and they are non-stationary, which is useful for the specific aim of the present work of characterizing CRs, because the metamodel can be refined in proximity of the CR limit surface. This can be achieved by training the Kriging with simulations whose outputs are concentrated nearby the limit surface and, indeed, adaptive training strategies have been recently developed to this aim. In the present paper, the metamodel-based AK-MCS framework developed in (Turati et al. 2017) is followed. A Kriging metamodel is initially built according to a small I/O training set, i.e., $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}_{in}$, whose simulations have been generated by Latin Hypercube Sampling (LHS) (McKay et al. 1979). Then, the metamodel refinement is carried out through the AK-MCS iterative procedure, which consists of the following steps (Puppo et al. 2021), for each n-th iteration (the algorithm is also sketched in Fig. 4 for the sake of clarity):

1. Construction: a Kriging metamodel is constructed with the available I/O training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ defined on the reduced space.
2. Generation of random input combinations: a large number N_{MCS} of new input configurations $\mathcal{X} = \left(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_{MCS}}\right)$ is generated by means of LHS, so as to evenly span the input space.

3. Metamodel Evaluation: the Kriging metamodel is exploited to estimate the output corresponding to the $\mathcal{X}$ input combinations: $\hat{\mathcal{Y}} = (\hat{y}_1, \dots, \hat{y}_{N_{MCS}})$.
4. Convergence check: convergence of the metamodel refinement process can be verified up to an a priori-defined convergence (e.g., a certain error metric) or stopping criterion (e.g., a limit on the computational budget, expressed in terms of a maximum number of BE-TH code simulations). Several criteria have been introduced to adaptively verify the convergence of the kriging training process. (Bichon et al., 2008; Echard et al., 2011) have introduced the Expected Feasibility Function (EFF) as a quantitative indicator of the *trade-off* between a detailed, refined search in *proximity* of the failure threshold and a more thorough, *global* exploration of the overall system state-space: the iterations are stopped when the largest value the EEF over the input space falls below a predefined limit (e.g., EFF < 0.001 in (Bichon et al., 2008)). (Echard et al., 2011) adopts the the U-learning function (2) to improve the modeling performance of Kriging preferably *across the failure surface*. The *smaller* $U(\boldsymbol{x})$, the higher the metamodel accuracy and precision in the region close to the limit state (correspondingly, the higher the advantage in including the simulation result corresponding to $\boldsymbol{x}$ in the current DoE). In this respect, when the smallest value of $U(\boldsymbol{x})$ over the input space exceeds a predefined threshold (e.g., $U(\boldsymbol{x})$ > 2 in (Cox and John, 1997; Echard et al., 2011)), the algorithm stops. However, it has been demonstrated that in several contexts, the EEF and the U-learning function (2) converge slowly to the failure region (Echard et al., 2011; Dubourg et al., 2013). Thus, other metrics have been introduced, relying on *cross-validation* to quantify *both* the kriging modelling performance *and* the convergence rate of the adaptive training process. In practice, the entire DoE $\{\mathcal{X}, \mathcal{Y}\}$ is divided into a training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ and a validation set $\{\mathcal{X}_{val}, \mathcal{Y}_{val}\}$ with the following properties: $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\} \cap \{\mathcal{X}_{val}, \mathcal{Y}_{val}\} = \emptyset$ and $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\} \cup \{\mathcal{X}_{val}, \mathcal{Y}_{val}\} = \{\mathcal{X}, \mathcal{Y}\}$. The kriging regression model is, then, built using the training subset and its prediction capabilities are quantified using the validation set. The *leave-one-out* (LOO) algorithm is a particular case, in which the training set is selected as $\{\mathcal{X}, \mathcal{Y}\} \backslash \boldsymbol{x}_i$. In (Allen, 1971) the mean squared error is estimated by a LOO approach and is termed Predicted REsidual Sum of Squares (PRESS). Instead, in (Dubourg et al., 2013; Turati et al., 2017) an error factor is computed by LOO cross-validation to assess the uncertainty affecting the failure probability values produced by kriging, and to quantify its prediction performance and convergence rate of the adaptive training procedure.

   In this paper, we use *only* the *computational cost* as a *stopping criterion* (i.e., we set a *maximum* number of simulations foreseen for the metamodel training): this choice is motivated by the significant computational effort typically associated to the dimensionality reduction phase carried out before this step (Section 3.1) and by the relevant amount of time needed to carry out a single transient simulation in the present application (i.e., around 4.3h on average). Nevertheless, it is important to notice that, even if a rigorous convergence/stopping criterion is *not* used, the evolution of the metamodel accuracy with the iterations is still checked by means of an a priori-defined (typically *small-sized*) validation set $\{\mathcal{X}_{val}, \mathcal{Y}_{val}\}$, made by N_{val} I/O relations: this allows to have at

least a rough idea of the kriging performance during and at the end of the training process. Further (numerical) implementation details are reported in Section 4, devoted to the application results.

5. Selection: if convergence at step 4 is not verified, the best candidate subset $\mathcal{X}^* \subset \mathcal{X}$ of input combinations is added to the Kriging training set by evaluating the corresponding outputs $\mathcal{Y}^*$ with the long-running model. The N_{cand} best candidates are selected among $\mathcal{X}$ according to their learning function values. U-function (Echard et al. 2011; Turati et al. 2017) has been chosen for the analysis among the several options available in literature (Xiao et al. 2018):

$$U(\boldsymbol{x}) = \frac{|Y_{thres} - \mu_{\hat{y}}(\boldsymbol{x})|}{\sigma_{\hat{y}}(\boldsymbol{x})}, \tag{2}$$

where $U(\boldsymbol{x})$ measures the distance, expressed in terms of metamodel standard deviation $\sigma_{\hat{y}}(\boldsymbol{x})$ between the mean value of the metamodel prediction $\mu_{\hat{y}}(\boldsymbol{x})$, corresponding to $\boldsymbol{x}$ and the failure threshold Y_{thres}. In general, the smaller is the U-function value, the closer is the predicted output to the limit surface and, hence, the higher the interest in adding that point to $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$. The best candidate inputs need to be spread over the domain, but it might occur that, due to the correlation, the points with the lowest U values result to be all restricted to the same portion of the input domain, providing a small amount of information when added to $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$. Some techniques can be implemented to overcome this problem: e.g., in (Turati et al. 2017) it is proposed to cluster the input domain to evenly "spread" the candidates over it.

Once the best candidates have been selected and sent in input to the real model which evaluates the corresponding output, $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ is updated and steps 1 to 5 are repeated until step 4 is verified.

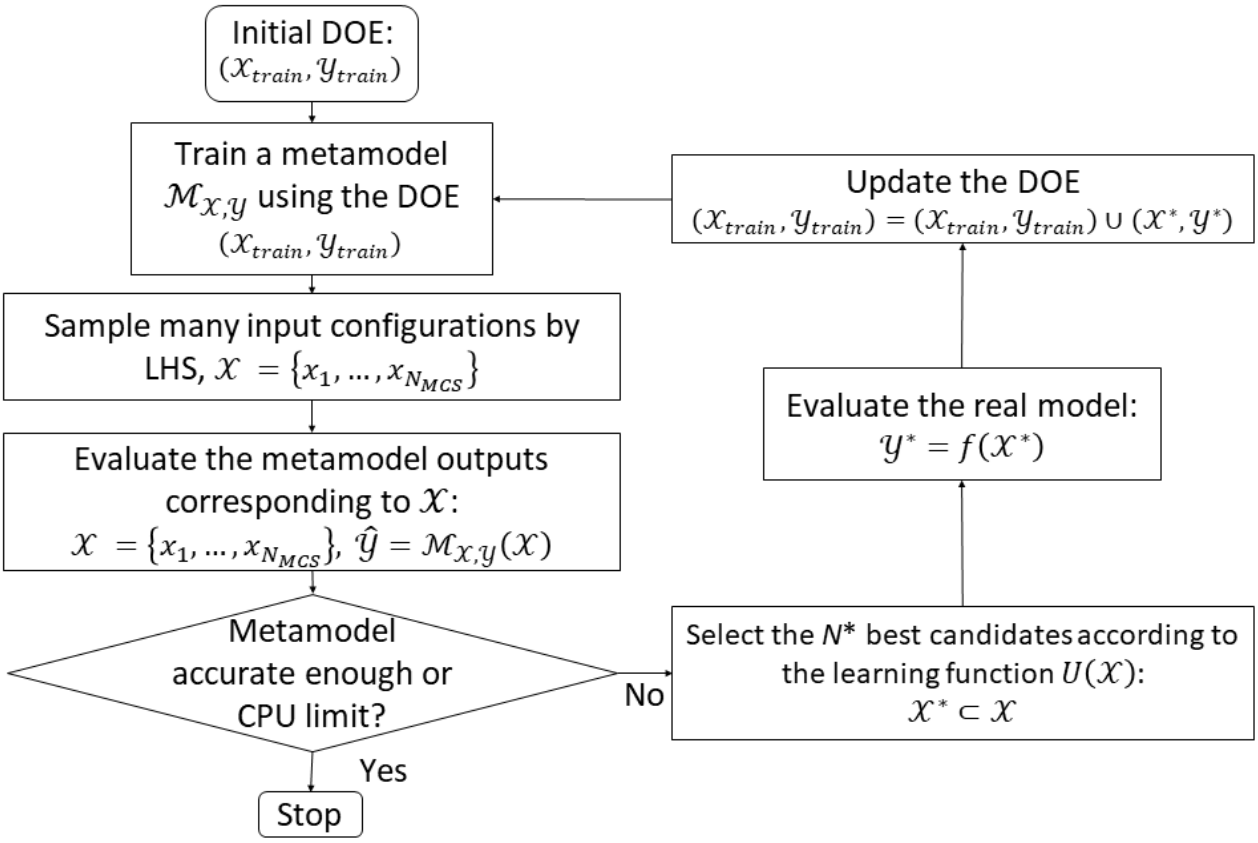

Figure 4: Flow diagram of the iterative metamodel training algorithm (AK-MCS)

A large amount of works has been devoted, in the last few years, to the intelligent, *iterative* improvement of the AK-MCS algorithm. Only few of the most relevant techniques are listed hereafter: the Adaptive Kriging-Importance Sampling (AK-IS) (Echard et al. 2013); the Meta Adaptive Kriging-Importance Sampling^2 (MetaAK-IS^2) (Cadini et al. 2014), which combines AK-IS and Meta-IS (Dubourg et al. 2013); the Active learning and Kriging-based SYStem reliability method (AK-SYS) (Fauriat and Gayton 2014); the Adaptive Kriging-Line Sampling (AK-LS) (Lu et al. 2015); the AK-Subset Simulation (AK-SS) (Huang et al. 2016) and the AK-Subset Simulation-Importance Sampling (AK-SS-IS) (Tong et al. 2015); the Polynomial Chaos Kriging (PCK) (Schöbi et al. 2017); the AK-MCSi algorithm, employing sequential MCS and multipoint enrichment techniques to allow parallelization (Lelievre et al. 2018); the Active Learning Kriging-Kernel Density Estimation-Importance Sampling (ALK-KDE-IS) (Yang et al. 2018); the ALK-Evolutionary Multimodal Optimization-Importance Sampling (ALK-EMO-IS) (Yang and Cheng 2020) and the ALK-Multimodal Adaptive Importance Sampling-Truncated Candidate Region (ALK-MAIS-TCR) (Yang et al. 2020); the Multifidelity Efficient Global Reliability Analysis (MfEGRA) method (Chaudhuri et al. 2021); the Adaptive Multi-Fidelity sparse Polynomial Chaos Kriging (AMF-PCK) technique (Zhao et al. 2021) and the Adaptive Kriging-based Directional Sampling (AK-DS) (Zhang et al. 2021). The objective of all the methods listed above is the *efficient propagation of uncertainties, physically* described by different probability distributions, through the (expensive) system models, for the *accurate* and *precise estimation* of (small) failure probabilities. On the

contrary, as already highlighted in the Introduction, in this work we are *not* performing any uncertainty propagation nor probability estimation, but rather we carry out an *inverse* analysis for identifying and characterizing – in a *multimodal* landscape – the combinations of values of the design and/or operational *input* variables driving a particular type of nuclear safety system to failure, i.e., the so-called failure region).

3.3 CR Representation & Information Retrieval

The Kriging metamodel obtained at the end of the procedure described in Section 3.2 must provide predictions of the output with satisfactory level of accuracy, especially in proximity of the CRs limit surfaces; this can be verified with an external validation set. Thus, a large number of new input combinations $\boldsymbol{x}$ (e.g., several thousands) is, then, generated, by LHS and provided in input to the metamodel; the critical ones, i.e., $\hat{y} = f(\boldsymbol{x}) \leq Y_{thres}$, are exploited for characterizing the shape and cardinality of the CRs (Puppo et al. 2021). In mathematical terms, this is equivalent to solving the inverse problem $\boldsymbol{x} = f^{-1}(\hat{y})$, with $\hat{y} \leq Y_{thres}$. Once this is done, CRs can be graphically represented by means of high-dimensional data visualization techniques, like scatter plots or Parallel Coordinates Plot (PCP).

In brief, scatter plots show the two-dimensional projections of the CRs over all possible pairs of inputs (this is useful to visualize the shape of the CRs). Moreover, in case of many input parameters involved, multiple scatter plots can be collected together in the so-called Scatter PLOt Matrix (SPLOM), providing a more complete view (Sedlmair et al. 2013).

On the other hand, PCP (Inselberg 2009) allows representing all the critical input combinations in a unique plot: all the M input variables (coordinates), normalized on their respective ranges, are reported on vertical axes and lined up horizontally; then, each input combination is represented by a horizontal line connecting the corresponding input variables values on the vertical axes. In this way, the analyst is provided with exemplary patterns of typical critical conditions for the system operation.

4 Application

The framework illustrated in Section 3 has been applied for the characterization of the CRs of the PSS introduced in Section 2. In the following Sections, the relevant steps of this application are illustrated in detail with reference to the characterization of the CRs relative to the multimodal output variable "maximum pressure value inside the PV" (p_{max}).

4.1 Dimensionality Reduction

With the aim of defining the I/O training set to construct an accurate metamodel for the approximation of the PSS response with respect to p_{max} output, the input vector dimensionality has been reduced from

M to R ($R < M = 5$); hence, a reduced model dealing with the reduced input vector $\boldsymbol{X}^R \in D_{\boldsymbol{X}^R} \subset \mathbb{R}^R$ can be constructed. The DBSA method supported by FMMs technique has been implemented to tackle p_{max} non-smoothness and multimodality (see Fig. 2) by identifying the different output clusters and, finally, selecting the most relevant inputs contributing to the output distribution. In particular, the analysis has been restricted only to those input parameters significantly affecting the output clusters connected with the critical conditions, i.e., those with p_{max} around $75.5\ bar$.

A total of 200 RELAP5-3D simulated transients have been used for the FMMs development (see Appendix A) with the SNOB algorithm, introduced in Section 3.1. The SNOB algorithm is based on the EM and selects the best number of components (k), guided by the MML criterion (see equation (1)).

The goal of the FMMs application is not to approximate the p_{max} distribution in the best way possible, whatever the number of components, but to obtain a good fit while ensuring that the k components reproduce the underlying physics of the problem. The SNOB algorithm gives the optimal fitting of the p_{max} multimodal distribution with $k = 3$ Gaussian distributions (whose characteristic parameters, i.e., mean value μ and standard deviation σ are reported in Table 4). In order to rank the most relevant inputs by means of one of the DBSA methods introduced in Section 3.1, firstly, it is necessary to assign each output variable in the set of 200 RELAP5-3D simulations to the cluster that generated it. In particular, it is assumed that the sample y_i belongs to the j-th cluster, if it returns the highest probability value when substituted into the PDF expression of that cluster.

Table 4: FMMs components parameters

Cluster name		μ [bar]	σ [bar]
Low-pressure	(green)	70.0	1E-3
Medium-pressure	(orange)	72.6	2.48
High-pressure	(red)	75.9	0.04

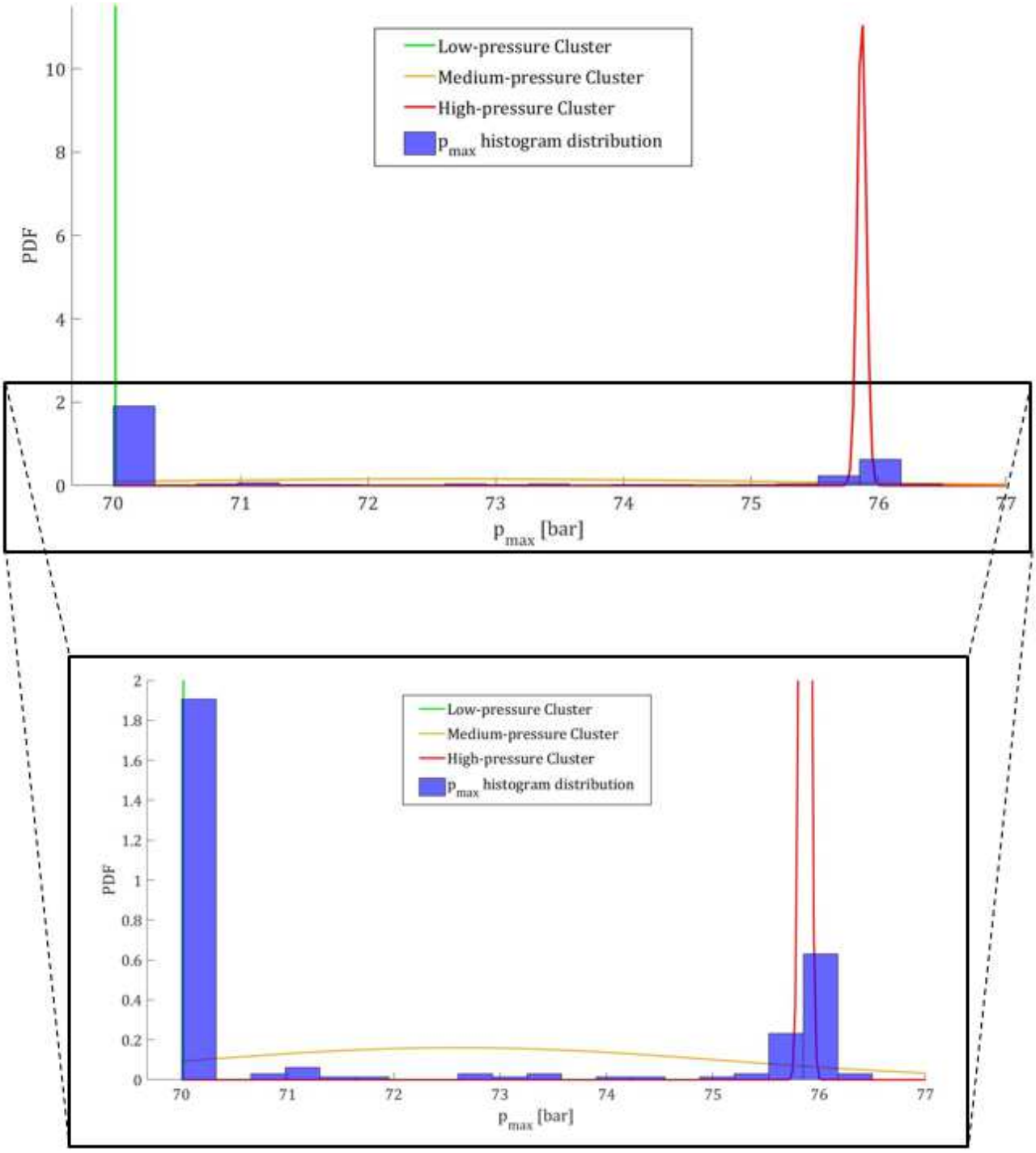

Figure 5: $\boldsymbol{p_{max}}$ clustering according to $\boldsymbol{k = 3}$ Gaussian distributions

The three Gaussian distributions and their related output clusters are reported in Fig. 5. A "low-pressure" cluster on the left is associated with the system safe conditions (in green in Fig. 5) and is approximated by a Dirac's delta distribution. It represents the p_{max} concentration around 70.0 bar, corresponding to all those transients in which the decay heat is correctly removed by the PSS and the pressure never increases

(122 simulations out of 200). Thus, in these simulations, p_{max} is always equal to the pressure value at the beginning of the transient, i.e., 70.0 bar. The remaining 78 outputs are almost equally split among the "medium-pressure" cluster in the middle (safe conditions, but with lower safety margin) and the "high-pressure" cluster on the right (critical conditions). They are associated to two Gaussian distributions (respectively orange and red in Fig. 5), with the second that is more peaked. Both clusters include transients in which the pressure initially increases beyond 70.0 bar, due to the AV delayed opening with respect to the MSIV closure: this causes the PV to remain without vapor outlets. The only difference is that p_{max} values in the "high-pressure" cluster reach $Y_{thres} = 75.5\ bar$, causing the SRV opening, whereas in most of the transients assigned to the "medium-pressure" cluster the pressure increases without reaching Y_{thres}.

The output clustering performed is exploited to identify those input variables that most affect the output clusters (DBSA) by constructing the PDF of each input x_m conditioned on each j-th cluster, i.e., $p(x_m|\,\Theta_{jm})$. In particular, the conditional PDFs are constructed by assigning the input variables belonging to the set of 200 RELAP5-3D simulations to the same cluster of the associated outputs; then, the PDF $p(x_m|\,\Theta_{jm})$ is created using only the x_m inputs assigned to the j-th cluster. In this way, it is possible to measure the difference between $p(x_m|\,\Theta_{jm})$ and the original (unconditional) input distribution of x_m, i.e., $p(x_m)$, and to use this difference to rank x_m. For the case study, the Hellinger distance method for SA (Gibbs and Su 2002; Di Maio et al. 2014) is adopted:

$$H_{jm} = \left[\frac{1}{2}\int\left(\sqrt{p(x_m)} - \sqrt{p(x_m|\,\Theta_{jm})}\right)^2 dx_m\right]^{1/2}, \tag{3}$$

with H_{jm} that needs to satisfy the inequality $0 \le H_{jm} \le 1$. The quantity H_{jm} represents the importance of the m-th input in affecting the j-th cluster of the output distribution: the higher the H_{jm} value with respect to the one of the other input parameters, the greater the relative importance of x_m.

For the analysis of p_{max}, special attention is paid to the "high-pressure" cluster, since it is the one connected with the failure of the PSS function (critical conditions). Hence, for each input parameter, the corresponding H_{jm} value referred to this cluster (i.e. with $j = 3$) is exploited as a sensitivity index. Fig. 6 reports a comparison between the H_{3m} values calculated for each of the five input parameters.

As it can be deduced from Fig. 6, the two valves operation delays, i.e., DEL_{AV} and DEL_{MISV}, mostly affect p_{max} "high-pressure" cluster and hence they are more likely to generate those scenarios in which the pressure increases towards Y_{thres}, with the consequent SRV opening. Therefore, the problem dimensionality is reduced from $M = 5$ to $R = 2$, and a reduced model (dealing with a reduced input vector) is obtained: i.e., $f(\boldsymbol{X}^R) = Y$, with $\boldsymbol{X}^R \in D_{\boldsymbol{X}^R} \subset \mathbb{R}^R$ and Y still equal to p_{max}.

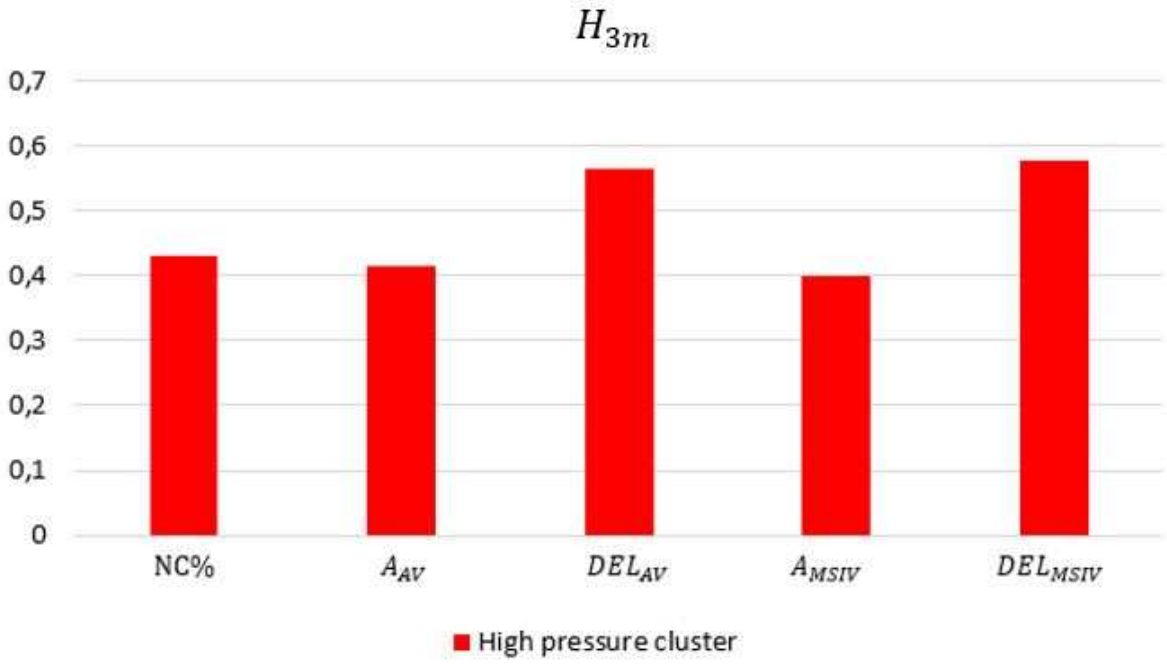

Figure 6: Hellinger distance for each input parameter ($\boldsymbol{x_m}$) evaluated with respect to the high-pressure cluster

4.2 Iterative Metamodel Training (AK-MCS)

After the dimensionality reduction previously presented, the input parameters used to model the generic PSS behaviour with respect to p_{max} are only DEL_{AV} and DEL_{MSIV}; thus, a Kriging metamodel has been built to mimic the RELAP5-3D model I/O relationships on a reduced space of dimensionality $R = 2$.

For the purpose of CRs exploration, the fact that p_{max} can approach $Y_{thres} = 75.5\ bar$ only if $DEL_{AV} > DEL_{MSIV}$, with a quite significant interval of time between the two valves actions, has led to adjust the range of variation of DEL_{MSIV} from $DEL_{MSIV} = 0 \div 7200\ sec$ to $DEL_{MSIV} = 0 \div 480\ sec$: this has allowed to be coherent with $DEL_{AV} = 0 \div 720\ sec$ (see Table 2) and to avoid sampling far from the limit surface.

Following the criterion proposed in (Loeppky et al. 2010), who suggests a number of training simulations $N_{train} \geq 10R$, the Kriging metamodel has been initially constructed with an I/O training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}_{in}$ of 25 RELAP5-3D runs (obtained in correspondence of input values generated by LHS). In particular, the construction has been performed by means of the UQLab Software Framework for Uncertainty Quantification (Marelli and Sudret 2014). UQLab provides a straightforward parametrization of the Kriging (Lataniotis et al. 2019): constant, linear, polynomial, or arbitrary trends, related to elliptic and separable correlation kernels, based on many possible one-dimensional distribution families (e.g., Gaussian,

Exponential, Matérn, or user-defined). The hyperparameters can be estimated through the Cross-Validation (CV) or the ML methods using different optimization techniques (local or global). The best Kriging setting for the specific study of p_{max} output has been established by testing different Kriging features with the CV procedure. In particular, the Kriging best setting has resulted to be:

- Trend type: *Ordinary*
- Family of correlation functions: *Exponential*
- Type of correlation functions: *Ellipsoidal*
- Estimation method: *CV*
- Optimization method: *Genetic Algorithm (GA)*

Then, the Kriging metamodel has been adaptively refined with a focus on the CR limit surface by enriching $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ within the AK-MCS framework introduced in Section 3.2. The AK-MCS procedure has been here tailored to the specific case study of the PSS introduced in Section 2, in relation to the pressure output evolution during a SBO accident. The details of the steps concerning the AK-MCS application are reported in what follows, for each n-th iteration:

1. Construction: a Kriging metamodel is constructed with the available I/O training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$, which increases its size with the iterations. The metamodel accuracy is improved specifically near $Y_{thres} = 75.5\ bar$.
2. Generation of random input combinations: $N_{MCS} = 10.000$ new input combinations $\mathcal{X} = \left(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_{MCS}}\right)$ (of reduced dimensionality $R = 2$) are sampled with LHS (see the input ranges defined at the beginning of this Section).
3. Metamodel evaluation: the sampled input combinations $\mathcal{X}$ are run through the Kriging metamodel to predict the corresponding output values (i.e., maximum vessel pressure): $\hat{\mathcal{Y}} = \left(\hat{y}_1, \dots, \hat{y}_{N_{MCS}}\right)$.
4. Convergence check: a convergence or stopping criterion regarding the computational cost has been defined. The maximum number of simulations foreseen for the metamodel training has been set to 100, due to the significant computational cost associated to the dimensionality reduction procedure carried out before (200 RELAP5-3D simulations required). Thus, considering that $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}_{in}$ is constituted by 25 simulations, only 75 simulations can be iteratively added during the AK-MCS procedure; when the size of $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ reaches its maximum value, the procedure stops.
5. Selection: if the convergence criterion at step 4 is not verified, new I/O simulations related to the so-called best candidate subset, i.e., $\mathcal{X}^* \subset \mathcal{X}$, are conducted, and the corresponding inputs and outputs $\{\mathcal{X}^*, \mathcal{Y}^*\}$ are added to $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ to refine the metamodel. The N_{cand} best candidates $\mathcal{X}^*$ are randomly selected among the $\mathcal{X}$ combinations according to their U-function values (see equation (2)), in order to choose them close to Y_{thres}. They should present a low U value, but at the same time not be "clustered" in the same area of the input space (i.e., too similar to each other). Actually, combinations that are close in the input space share similar U values and, hence, the candidates should be selected not only according to the N_{cand} lowest U values, because they would

all be restricted in the same area of the domain, instead of spanning the whole input space. Thus, $N_{cand} = 7$ or 8 candidates are added at each n-th iteration, according to the same rationale presented in (Puppo et al., 2021). Once $\mathcal{X}^*$combinations have been selected and the corresponding RELAP5-3D transients simulated with the RELAP5-3D code to obtain the output $\mathcal{Y}^*$, $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ is enriched and steps 1 to 5 are repeated until convergence at step 4 is verified.

The AK-MCS procedure has been stopped at iteration $n_{fin} = 10$, when the maximum number (100) of RELAP5-3D simulations allowed for the construction of the training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ has been reached. The evolution of the metamodel accuracy with the iterations has been followed through the introduction of an a priori-defined validation set $\{\mathcal{X}_{val}, \mathcal{Y}_{val}\}$, made by N_{val} I/O relations. Some recommendations about how to determine the best N_{val} can be found in (Martin and Simpson 2005; Iooss 2009; Wu et al. 2018), but no definitive guidelines are available. Considering the available computational budget, a validation set of 50 RELAP5-3D simulations with the outputs mainly distributed around Y_{thres} has been constructed to measure the accuracy increase, especially in proximity of the limit surface. The metamodel has been used to predict the outcomes $\hat{\mathcal{Y}}_{val} = (\hat{y}_1, \dots, \hat{y}_{N_{val}})$ corresponding to the N_{val} input combinations of the validation set ($\mathcal{X}_{val}$); then, the accuracy has been quantified through Quality Indicators (QIs), comparing $\hat{\mathcal{Y}}_{val}$ to the real outputs evaluated with the RELAP5-3D model. The closer the Kriging prediction to the RELAP5-3D output, the better the QI value calculated and the higher the accuracy. The QIs adopted are the following: the well-known Root-Mean-Square Error (RMSE) and two different predictivity indicators, namely respectively Q_1, defined in (Iooss 2009), and Q_2 presented by (Lataniotis et al. 2019):

$$RMSE = \sqrt{\sum_{i=1}^{Nval} \frac{(\hat{y}_i - y_i)^2}{N}}, \tag{4}$$

$$Q_1 = 1 - \frac{\sum_{i=1}^{Nval}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{Nval}(\bar{y}_{val} - y_i)^2}, \tag{5}$$

$$Q_2 = \frac{N_{val} - 1}{N_{val}}\left(\frac{\sum_{i=1}^{Nval}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{Nval}(\bar{y}_{val} - y_i)^2}\right), \tag{6}$$

where y_i is the i-th output of the set $\{\mathcal{X}_{val}, \mathcal{Y}_{val}\}$, $\hat{y}_i$ is the corresponding Kriging prediction and $\bar{y}_{val}$ is the mean value of all the validation outputs evaluated with the RELAP5-3D model. RMSE and Q_2 should be as low as possible, whereas Q_1 tends to 1 as the prediction accuracy increases. The RMSE has the same unit of measure of the physical quantity of interest (p_{max}) and, hence, it can be directly compared to the maximum pressure to understand whether the level of accuracy is satisfactory. It can be also normalized (NRMSE) dividing it by $\bar{y}_{val}$. Q_1 and Q_2 have similar expressions and, differently from $RMSE$, they also account for the variability of y_i in the set.

The progressive increase in accuracy is shown by the trends of the three QIs illustrated in Fig. 7. All the QIs considered show a significant improvement at the beginning; then, in the successive iterations, the

relative improvement becomes negligible. Also for this reason, stopping the AK-MCS procedure at iteration $n_{fin} = 10$ represents a reasonable choice.

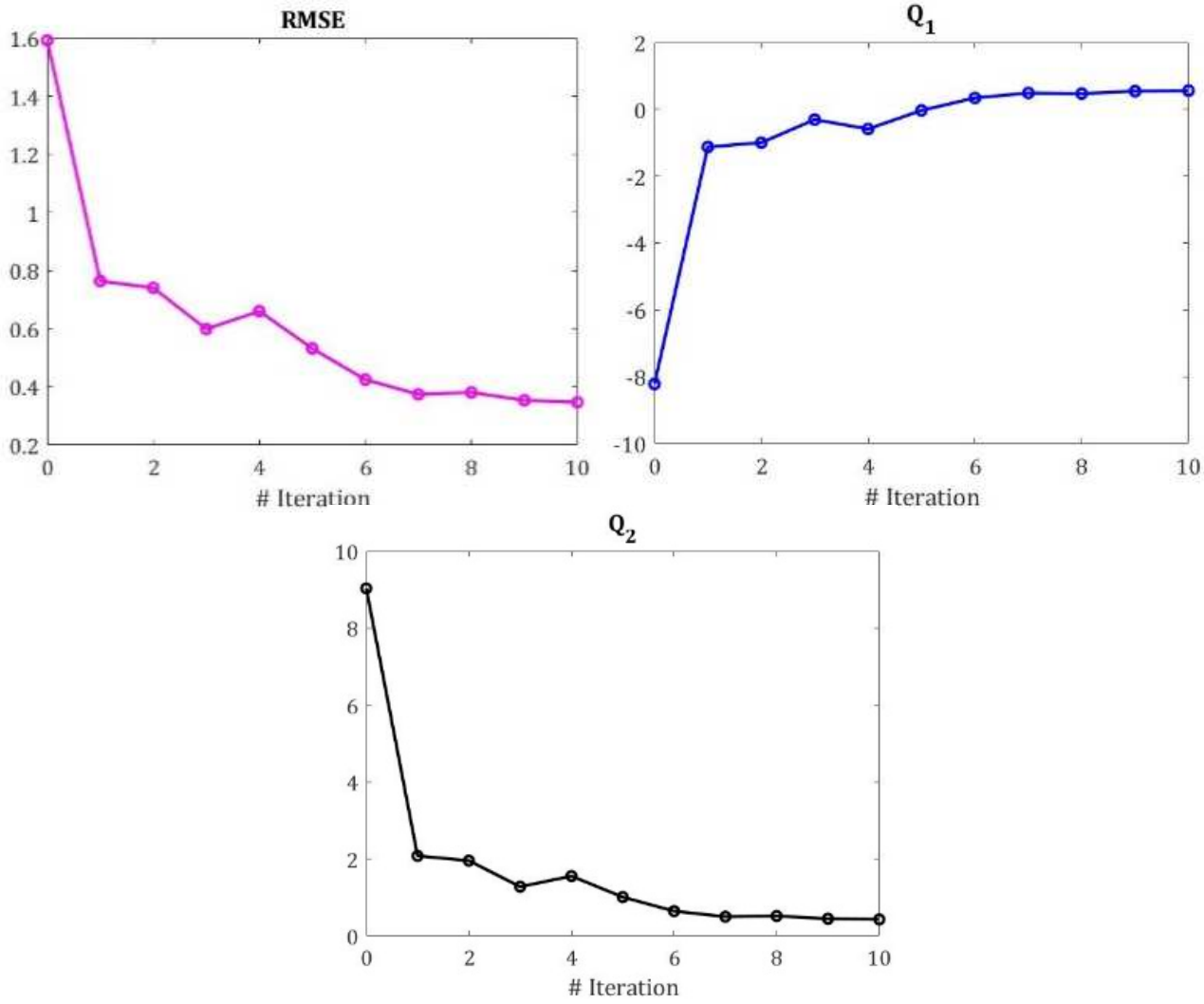

Figure 7: QIs evaluated with respect to a given validation set

The three QI values at the end of the AK-MCS procedure are reported in Table 5. The RMSE at the last iteration is satisfactory, indeed $RMSE = 0.35\ bar$ is really low if compared to the p_{max} values in the simulated transients ($p_{max} = 70.0 \div 76.5\ bar$). Moreover, a final $NRMSE = 0.46\%$ is remarkable, since it can be taken, in the first instance, as a measure of the percentage error of the Kriging predictions. For what concerns Q_1 and Q_2, they show a significant improvement during the successive iterations, but their final values are not so satisfactory, especially for the final Q_1 which lies far from 1. This is probably due to the very low variability of the validation set chosen for the analysis: indeed, most of the p_{max} values of $\mathcal{Y}_{val}$ are spread on a range of only 2 bar around $Y_{thres} = 75.5\ bar$.

Table 5: QIs values at the end of the AK-MCS procedure

Quality indicator	**RMSE** [bar]	**NRMSE** [%]	Q_1	Q_2
Final value	0.35	0.46%	0.56	0.43

4.3 CR Representation & Information Retrieval

The Kriging metamodel obtained at the end of the AK-MCS procedure has been demonstrated to present a high accuracy, especially in proximity of Y_{thres}: thus, it can be used for CR characterization instead of the more time-demanding RELAP5-3D model. For this purpose, 10.000 new input combinations have been generated by LHS and, then, predicted with the metamodel to: (i) find the critical ones, i.e., $\hat{y} = f(\boldsymbol{x}) \geq Y_{thres}$; and (ii) retrieve useful information about the CRs (like their number and shape).

One single CR has been identified (see Fig. 8); moreover, given that the analysis has been restricted only to two parameters after dimensionality reduction, no high-dimensional data visualization techniques like SPLOM or PCP (see Section 3.3) were needed. The CR has been represented in the two-dimensional input space through a scatter plot, in which green diamonds indicate combinations leading to safe operation (p_{max} is kept $< 75.5\ bar$), whereas red crosses represent the critical input combinations of PSS functional failure.

A triangle-shaped CR has been identified, showing the direct influence of both DEL_{AV} and DEL_{MSIV} on the FC "Steam release in the containment"; indeed, it is evident that p_{max} may exceed 75.5 bar *only when* the MSIV closes before the opening of the AV, i.e., when $DEL_{MSIV} < DEL_{AV}$ (as introduced in Section 2). This occurs because the PV remains without vapor discharge outlets and, hence, the vapor builds up causing the PV over-pressurization. Also, Fig. 8 shows that not always $DEL_{MSIV} < DEL_{AV}$ leads the PSS to fail its function: e.g., even if the MSIV is supposed in its reference conditions (i.e., $DEL_{MSIV} = 0\ sec$), if $DEL_{AV} < 50\ sec$, p_{max} remains below Y_{tresh}. In general, the higher DEL_{MSIV}, the lower the chances to lead to functional failure: eventually, if $DEL_{MSIV} > 380\ sec$ failure is never reached, whatever the value assumed by DEL_{AV}.

A word of caution is in order with respect to the results included in Fig. 8. The comparatively large size of the failure region (red crosses) with respect to the safe one (green diamond) does *not* necessarily mean that the PSS under analysis is "prone" to failure, since such type of conclusion can only be based on the quantitative assessment of the *probability of occurrence* of the corresponding input combinations (which is not performed in the present paper). In fact, the probability of functional failure of the PSS is strongly dependent on: (i) the structure and characteristics of the system itself, and (ii) the (data- and/or expert-based) probability density functions of the PSS input variables. In this work, as mentioned in Section 2, the PSS parameters are *not* described by realistic probability distributions, since the objective is not to perform a reliability assessment of the PSS, but to show how the FMMs-based adaptive Kriging procedure can be exploited for the thorough exploration, identification and characterization of multi-modal critical regions.

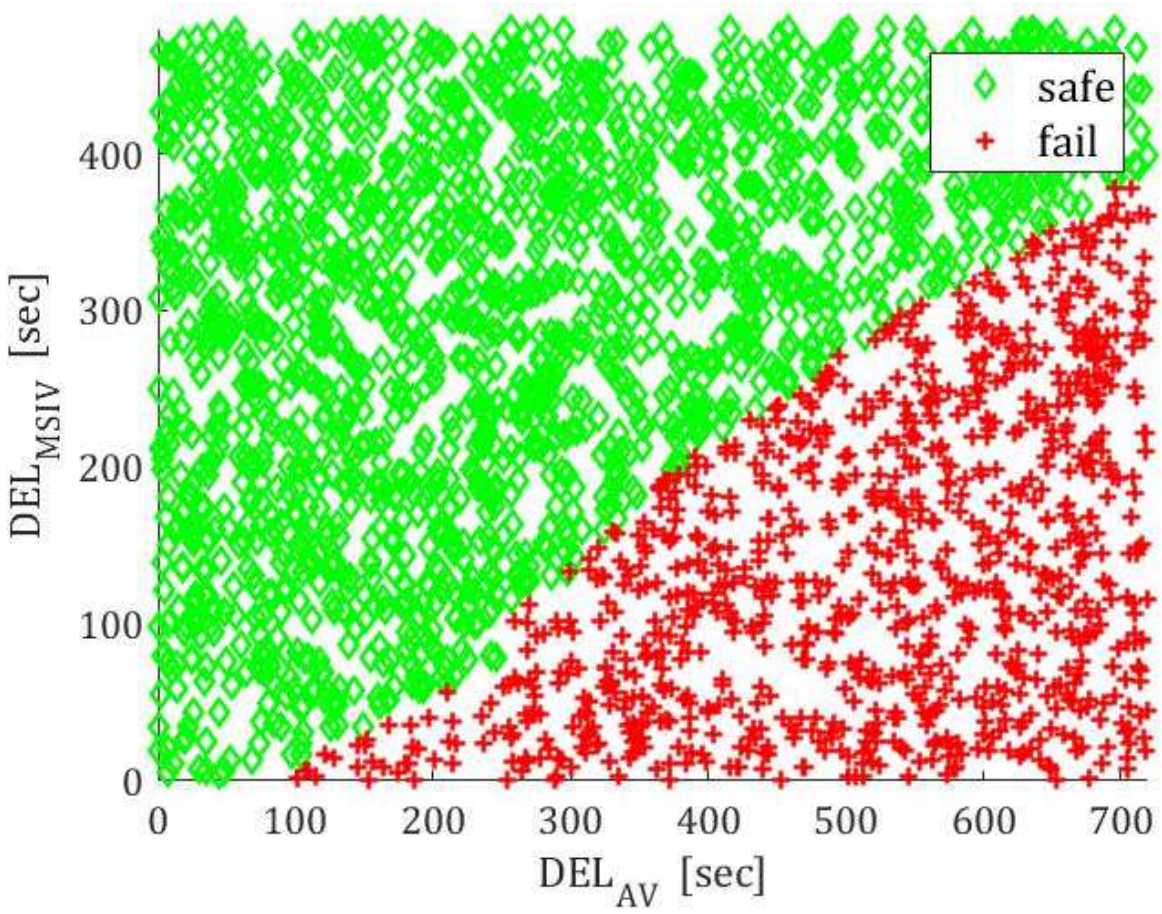

Figure 8: CR for $\boldsymbol{p_{max}}$ output

5 Comparison with the Results obtained with SVC + AK-MCS

An alternative approach to tackle the output non-smoothness and multimodality is represented by the use of a classifier. Given an I/O training set, the training outputs are clustered (e.g., by expert judgement) to separate the regions of different output behaviours (also called partitions). The same clusters are identified also in the input domain by simply assigning the training inputs to the same clusters of the corresponding output values. In this way, a classifier can be built according to these clustered I/O relations and, then, a new input combination $\boldsymbol{x}$ can be classified to one of the different input domain partitions. Multiple metamodels can be fitted to each cluster in the input and output space to obtain a better approximation (rather than constructing a unique metamodel for the whole space). Thus, the new combination $\boldsymbol{x}$ is predicted according to the specific metamodel developed for the partition $\boldsymbol{x}$ belongs to.

We have here applied this approach to the PSS presented in Section 2, with respect to p_{max} output, by resorting to one of the most popular classifiers, i.e., SVC (Vapnik and Cortes 1995). The results have been

compared to the ones obtained with the framework of Section 4. In particular, a hard SVC (i.e., where one input combination cannot belong to different domain partitions) has been trained within the "two stage-surrogate modelling" technique introduced in (Moustapha and Sudret 2019). A new input combination $\boldsymbol{x}$ whose output needed to be predicted (and identified as critical or not) has been, first, classified with the hard SVC (1st stage); then, the corresponding output has been predicted with the metamodel specifically built for the partition (cluster) which $\boldsymbol{x}$ is classified to (2nd stage).

At first, two output domain partitions have been identified according to expert judgement: a "low-pressure" region corresponding to $p_{max} = 70.0\ bar$ (which occurs in most of the transients simulated, see Section 2), and a "high-pressure" region with $p_{max} > 70.0\ bar$, representing those transients in which the pressure rises. Thus, a binary classification results: i.e., given a certain input combination $\boldsymbol{x}$, the corresponding label assigned by the classifier is $\ell_i = \{-1, +1\}$, with the $\ell_i = -1$ that is associated to the low-pressure region and $\ell_i = 1$ that represents the high-pressure region.

The SVC has been constructed (see Appendix B) according to the two output domain regions identified, thanks to an I/O training set and their corresponding labels $\{\mathcal{X}_{train}^{SVC}, \mathcal{Y}_{train}^{SVC}\}$. The same training set of 200 RELAP5-3D simulations adopted for the FMM-based approach (see Section 4.1) has been used (122 with $\ell_i = -1$ and 78 with $\ell_i = 1$). Indeed, the criterion introduced in (Basudhar et al. 2008), which proposes convergence points instead of a far more expensive validation set to quantify the accuracy of the SVC, has proven that at least 180 simulations were necessary to construct an initial, but sufficiently accurate SVC for the case study. Then, a Kriging metamodel has been built to predict p_{max} only in the "high-pressure" region, because it was not worth exploring also the "low-pressure" region with constant p_{max} (70.0 bar). Then, a metamodel has been constructed with an I/O training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ made by the same I/O relations collected for the SVC training, but taking only those classified as belonging to the high-pressure region (78 simulations out of 200). In this case, no dimensionality reduction has been carried out and, hence, the metamodel is used to mimic the RELAP5-3D model on the original input space of dimensionality $M = 5$, i.e., $f(\boldsymbol{X}) = Y$, with $\boldsymbol{X} \in D_{\boldsymbol{X}^M} \subset \mathbb{R}^M$. Thus, $\mathcal{X}_{train}$ is a set of five-dimensional input combinations.

The Kriging metamodel has been adaptively refined in proximity of $Y_{thres} = 75.5\ bar$ with a sort of AK-MCS procedure (see Section 3.2), conveniently adjusted to be coupled with SVC. At each n-th iteration, $N_{MCS} = 100.000$ new input combinations $\mathcal{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_{MCS}})$ have been generated by LHS and classified by the SVC according to the two regions identified (1st stage). Only the combinations classified as belonging to the high-pressure region, i.e., $\mathcal{X}^{Krig} \subset \mathcal{X}$, have been then evaluated with the Kriging metamodel to find the corresponding outputs (2nd stage). The most interesting input combinations among $\mathcal{X}^{Krig}$, in terms of learning function U value (7-8 candidates at each iteration), have been selected for simulation with the RELAP5-3D model and added to $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ for the metamodel refinement. This procedure is repeated until the level of accuracy of Kriging predictions becomes satisfactory. The I/O relations simulated at each iteration to enrich the metamodel training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}$ have been labelled and exploited to enrich also the classifier training set $\{\mathcal{X}_{train}^{SVC}, \mathcal{Y}_{train}^{SVC}\}$. This procedure is called SVC+AK-MCS, hereafter.

The idea is to exploit the same number of RELAP5-3D simulations, i.e., the same computational budget, as the one used for the novel exploration framework implemented in Section 4.2 (FMM+AK-MCS), to refine both the Kriging metamodel and the SVC within SVC+AK-MCS framework, with the aim of fairly comparing the final Kriging accuracy. The initial metamodel training set $\{\mathcal{X}_{train}, \mathcal{Y}_{train}\}_{in}$ has been adaptively enriched together with $\{\mathcal{X}_{train}^{SVC}, \mathcal{Y}_{train}^{SVC}\}$, up to the limit of the available 300 simulations (the same limit as FMM+AK-MCS). Thus, starting from $\{\mathcal{X}_{train}^{SVC}, \mathcal{Y}_{train}^{SVC}\}$ made by 200 I/O samples (the same used for the FMMs application), 100 simulations have been added (only 82 of them could have been used for the Kriging training) and the Kriging training set size has been simultaneously increased from 78 to 160. The Kriging accuracy has been quantified with respect to a validation set of the same size of the one used in Section 4.2 (i.e., 50 I/O relations). Again, the validation set is constituted by samples mainly distributed around Y_{thres}, to verify the metamodel accuracy improvement with specific attention to the area close to the limit surface. Table 6 reports the values of three QIs (RMSE, Q_1 and Q_2) computed on this validation set according to the Kriging metamodel obtained at the end of SVC+AK-MCS procedure.

Table 6: QIs at the end of SVC+AK-MCS iterative procedure

Quality indicator	**RMSE** [bar]	**NRMSE** [%]	Q_1	Q_2
Final value	0.85	1.12%	0.16	0.82

All the QIs values are worse than those obtained by the FMM+AK-MCS framework (see Table 5). For example, the RMSE and NRMSE are more than twice larger, and Q_1 is even 3.5 times lower, meaning that the accuracy of the Kriging metamodel at the end of the SVC+AK-MCS procedure is lower. Indeed, Fig. 9 shows how the adaptive exploration framework applied in Section 4, based on FMM+AK-MCS, outperforms the SVC+AK-MCS procedure in terms of $Q_1$and Q_2, after six iterations (i.e., with 270 simulations rather than 300) and, for what concerns the RMSE, after one iteration (i.e., with 233 simulations rather than 300).

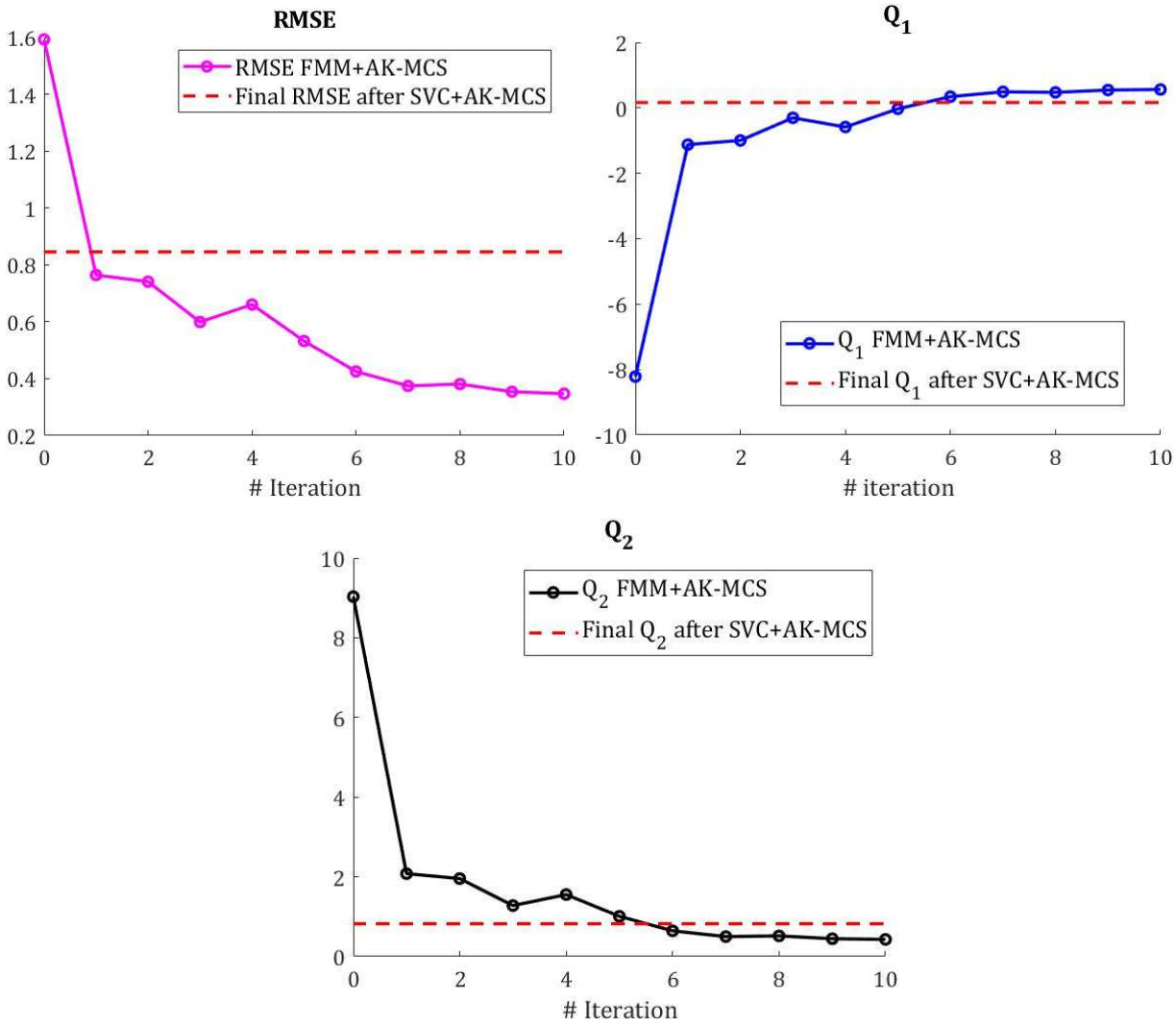

Figure 9: QIs evolution in FMM+AK-MCS strategy compared with the QIs values at the end of SVC+AK-MCS

A final consideration is in order with respect to the results obtained by the proposed methodology, in particular with reference to the *dimensionality reduction* step carried out above (Sections 3.1 and 4.1). As already mentioned, parameterizing, and training a metamodel becomes hard or even intractable as the number M of input parameters increases (in particular, when $M > 20$), a well-known problem often referred to as *curse of dimensionality* (see, e.g., (Verleysen and François, 2005; Lataniotis et al., 2020)); similar challenges arise in the presence of high-dimensional model outputs (Auder et al., 2012; Gu and Berger, 2016). However, in the case study here considered (Section 2), the number of input variables selected by expert judgment is *quite small* (i.e., equal to 5), which allows in principle: i) the construction of a relatively

small-sized DoE still able to evenly cover the entire input space; and ii) a satisfactorily accurate, precise, and fast (iterative) training of the kriging surrogate model (Sections 3.2 and 4.2). In light of this, the dimensionality reduction step may not seem essential here. However, the improved performance of the DBSA-based AK-MCS supported by FMMs (employing a reduced input space of size $R = 2 < M = 5$) with respect to the SVC + AK-MCS (employing the full input space of size $M = 5$) demonstrates the advantage of the dimensionality reduction also in this case (see Figure 7 and Figure 9, respectively): this is particularly true when the analyst needs to *approximate non-smooth* and *multimodal* distributions by metamodels and, then, *restrict* the state-space to the input parameters affecting *only* the *output clusters* connected with *system failure* (which is of interest in the present application).

6 Conclusions

There is a growing interest in PSSs applications to increase the safety level of advanced NPPs: in this light, the CRs characterization of PSSs becomes of paramount importance to discover the combinations of factors leading them to critical conditions. The adoption of innovative computational methods, like fast-running surrogate metamodels coupled with adaptive sampling techniques, represents a promising way to replace computationally demanding models and speed up the exploration of components and systems state-spaces, especially for the characterization of their CRs. However, a significant issue may be represented by the irregularity of the state-space, e.g., in case of non-smoothness and/or multimodality of the system response. To this purpose, we have developed a novel adaptive exploration framework, based on FMM and AK-MCS, capable of tackling the state-space *non-smoothness* and *multimodality*, while searching for the system CRs.

The proposed framework consists of three steps: i) "dimensionality reduction", relying on a DBSA method (specifically, Hellinger distance in the present work), supported by FMMs technique to approximate the non-smooth and multimodal output distribution and, then, restrict the analysis only to the input parameters affecting the output clusters connected with system failure; 2) "Iterative metamodel training", based on the AK-MCS technique for the construction of an accurate Kriging metamodel to replace the typically long-running system model codes and predict the system response on a space of reduced dimensionality. The metamodel is trained with a possibly small number (e.g., few hundreds) of time-demanding code runs; 3) "CR representation and information retrieval", using the Kriging metamodel obtained at the previous step to predict a large number of new input combinations and retrieve useful information about the system CRs. The CRs can be, then, visualized by exploiting high-dimensional data visualization techniques (specifically, scatter plots in the present work).

The framework has been applied to the exploration of the CRs of a generic PSS of an NPP, designed for DHR in case of reactor shut down (due to a SBO accident, in this work) in order to provide limits for the system safe operation. In particular, the DHR system here considered is modelled through a time-

demanding BE-TH code (RELAP5-3D model) and the success of its operation has been analyzed with respect to one output of interest, i.e., the maximum value of pressure reached inside the reactor PV (p_{max}).

The analysis of the PSS CRs relative to the FC "Steam release in the containment" (i.e., $p_{max} > 75.5\ bar$) has required the application of the FMM-based exploration framework, due to the strong non-smooth and multimodal distribution of the pressure output. The FMMs technique has been shown capable of approximating p_{max} distribution by identifying three different clusters, associated to three different kinds of responses with respect to the failure limit of 75.5 bar. Also, the Hellinger distance method for SA has been exploited to select the input parameters most affecting the output cluster associated to critical conditions. By so doing, the analysis has been restricted to two relevant input parameters out of the five ones initially identified: DEL_{AV} (i.e., the delay of Activation Valve opening) and DEL_{MSIV} (i.e., the delay of Main Steam Isolation Valve closure). Then, the AK-MCS technique has allowed the adaptive construction of an accurate Kriging metamodel (with increased accuracy nearby the $Y_{thres} = 75.5\ bar$) to replace the time-demanding RELAP5-3D model on the reduced input space (two-dimensional), by resorting to a limited number of simulations (specifically, 300 in this work). Thanks to dimensionality reduction, the Kriging metamodel has managed to correctly predict the output p_{max}, despite the non-smoothness and multimodality of its distribution (e.g., $NRMSE < 0.5\%$ when evaluated with respect to a validation set constructed around Y_{thres}).

A comparison with an alternative state-of-the-art approach to tackle the *non-smoothness* and *multimodality* of a system response (not relying on FMMs-based DBSA) has been carried out. Output domain regions with different behaviors have been identified and, then, both the output and input space have been partitioned. An SVC has been trained and coupled with the AK-MCS technique, within the innovative "two-stage surrogate modelling" strategy proposed in (Moustapha and Sudret 2019). First, a new input combination is assigned to the correct domain partition, then, the corresponding output is predicted to identify if it is critical or not. The results, in terms of metamodel accuracy, have been compared with those obtained by the FMM-based exploration framework proposed in this work, considering the same computational budget (i.e., same number of RELAP5-3D simulations). The strategy adopting an initial dimensionality reduction based on a DBSA method supported by FMMs outperformed the one relying on SVC. This represents a strong statement in support of dimensionality reduction techniques when dealing with the *metamodel-based* exploration of abrupt, irregular, and disconnected state-spaces.

Also, it is worth acknowledging that the proposed framework inherits the intrinsic limitations of the techniques employed. Actually, if the number of parameters identified after the dimensionality reduction is not sufficiently low to be managed by a Kriging metamodel, which suffers high-dimensionality and irregular output behavior, the success of the entire framework may be compromised.

Finally, a closing remark is due with respect to the importance and usefulness of the framework here developed, by discussing its possible applicability within the reliability and risk analyses traditionally performed for nuclear systems and components. As already said, the main objective of the proposed

framework is to thoroughly explore, find and characterize those input configurations (i.e., combinations of *phenomenological events* and/or *components failure modes* and/or *design parameters values*) which drive the PSS to critical states (i.e., to *fail its function*). Instead, for a proper evaluation of the risk of failure of the PSS, we would need to assess the *likelihood* of such hazardous and severe configurations (which is typically obtained by *representing* the uncertainties in the PSS system *behavior* and *modeling* by probability density functions and *propagating* them through the deterministic T-H code). This is beyond the scope of the present study. Yet, research is envisaged to *combine* the FMMs- and Kriging-based iterative exploration framework here proposed with state-of-the-art stochastic simulation techniques for the *accurate* and *precise* evaluation of the PSSs functional failure probability (e.g., Importance sampling-IS, Markov Chain Monte Carlo-MCMC, Subset Simulation-SS, Line Sampling-LS), with emphasis on: i) abrupt, multi-modal, possibly disconnected system state-spaces to be probed (like the one of interest in the present article) and ii) those challenging cases where the size of the critical region is quite small and its location is far from the nominal design (Schöbi et al., 2017; Yang et al. 2018; Yang and Cheng, 2020; Yang et al., 2020; Chaudhuri et al. 2021; Zhao et al., 2021; Zhang et al. 2021).

Acknowledgments

The authors express their deep gratitude to Prof. F. D'Auria (University of Pisa-UNIPI, Pisa, Italy; email: f.dauria@ing.unipi.it) and to Dr. M. Lanfredini (University of Pisa-UNIPI, Pisa, Italy; email: m.lanfredini@dimnp.unipi.it) for contributing to this work by providing the RELAP5-3D model of the passive safety system and the corresponding nodalization and input data employed in the code. The authors also thank the two anonymous referees for their constructive comments that significantly helped improving the paper.

Appendix A – Finite Mixture Models (FMMs)

Here is provided a description of FMMs construction through the classical EM Algorithm (Figueiredo and Jain 2002). Let assume a set of n output variables $\mathcal{Y}_{train}^{FMM} = \{y_1, \dots, y_n\}$, the generic y_i is said to follow a k-component finite mixture distributions if its PDF can be written as:

$$p(y_i|\boldsymbol{\Theta}) = \sum_{j=1}^{k} \pi_j p_j(y_i|\Theta_j), \tag{7}$$

where $\{\pi_j, j = 1, .., k\}$ are the mixing parameters or weights, Θ_j are the parameters of each j-th component and $\boldsymbol{\Theta} = \{\pi_1, \dots \pi_k, \Theta_1, \dots, \Theta_k\}$ is the complete set of mixture parameters; being probabilities, π_j must satisfy:

$$\sum_{j=1}^{k} \pi_j = 1. \tag{8}$$

Considering the set of samples $\mathcal{Y}_{train}^{FMM}$, the log-likelihood corresponding to a k-component mixture is:

$$\log p(\mathcal{Y}_{train}^{FMM}|\boldsymbol{\Theta}) = \log \prod_{i=1}^{n} p(y_i|\boldsymbol{\Theta}) = \sum_{i=1}^{n} \log \sum_{j=1}^{k} \pi_j p(y_i|\Theta_j), \tag{9}$$

and the related ML estimate reads:

$$\widehat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \{ \log p(\mathcal{Y}_{train}|\boldsymbol{\Theta}) \}. \tag{10}$$

$\widehat{\boldsymbol{\Theta}}$ cannot be found analytically since it implies to solve a non-linear equations system. Hence, the solution is provided through the application of EM Algorithm which interprets $\mathcal{Y}_{train}^{FMM}$ as a set of incomplete data. The "missing part" is represented by a set of labels, i.e., $\mathcal{Z} = \{\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(n)}\}$, associated to the y_i values numbered n, where each i-th label is a binary vector, i.e., $\boldsymbol{z}^{(i)} = \{z_1^{(i)}, \dots, z_k^{(i)}\}$, whose components are all zeros except for $z_j^{(i)} = 1$, i.e., the $\boldsymbol{z}^{(i)}$ component associated to the j-th distribution of the mixture that has generated y_i. Now, the complete log-likelihood for the estimation of $\widehat{\boldsymbol{\Theta}}$ can be written as:

$$\log p(\mathcal{Y}_{train}^{FMM},\ \mathcal{Z}|\boldsymbol{\Theta}) = \sum_{i=1}^{n} \sum_{j=1}^{k} z_j^{(i)} \log\left[\, \pi_j p(y_i|\Theta_j)\right]. \tag{11}$$

The EM Algorithm provides a sequence of estimates $\{\widehat{\boldsymbol{\Theta}}(t)\ with\ t = 0, 1, 2 \dots\}$ through the alternate realization of two steps, until some convergence criterion is satisfied:

- **E-step:** given the $\mathcal{Y}_{train}^{FMM}$ estimate through the current $\widehat{\boldsymbol{\Theta}}(t)$, and considering that $\log p\left(\mathcal{Y}_{train}^{FMM},\ \mathcal{Z}\middle|\boldsymbol{\Theta}\right)$ is linear with respect to $\mathcal{Z}$, the conditional expectation of the log-likelihood is computed through the construction of the so-called $\mathcal{Q}$-function by simply evaluating the conditional expectation, i.e., $W \equiv E\left[\, \mathcal{Z} \mid \mathcal{Y}_{train}^{FMM}, \widehat{\boldsymbol{\Theta}}(t)\right]$, and plugging it into $\log p\left(\mathcal{Y}_{train}^{FMM},\ \mathcal{Z}\middle|\boldsymbol{\Theta}\right)$:

  $$\mathcal{Q}\left(\Theta, \widehat{\boldsymbol{\Theta}}(t)\right) \equiv\ E\left[\, \log p(\mathcal{Y}_{train}^{FMM},\ \mathcal{Z}|\boldsymbol{\Theta}) \mid \mathcal{Y}_{train}^{FMM}, \widehat{\boldsymbol{\Theta}}(t)\right] = \log p(\mathcal{Y}_{train}^{FMM}, W|\boldsymbol{\Theta}). \tag{12}$$

  Knowing that $z_j^{(i)}$ coefficients are of binary kind, Bayes law can be expoited to calculate their conditional expectation:

  $$w_j^{(i)} \equiv\ E\left[\, z_j^{(i)} \mid \mathcal{Y}_{train}^{FMM}, \widehat{\boldsymbol{\Theta}}(t)\right] = \Pr\left[\, z_j^{(i)} = 1 \mid y_i, \widehat{\boldsymbol{\Theta}}(t)\right] = \frac{\hat{\pi}_j(t) p\left(y_i\middle|\widehat{\Theta}_j(t)\right)}{\sum_{m=1}^{k} \hat{\pi}_m(t) p\left(y_i\middle|\widehat{\Theta}_m(t)\right)}, \tag{13}$$

  where π_j and $w_j^{(i)}$ are the respectivey the a priori probability and a posteriori probability, after observing y_i, that $z_j^{(i)} = 1$.
- **M-step:** the mixture parameters are updated, under the constraints introduced by (8), according to:

$$\widehat{\boldsymbol{\Theta}}(t+1) = \arg\max_{\boldsymbol{\Theta}} \left\{ Q\left(\Theta, \widehat{\boldsymbol{\Theta}}(t)\right)\right\} \tag{14}$$

Appendix B – Support Vector Classifiers (SVC)

Here is provided a description of SVC construction process in case of binary classification, i.e., when only two classes have been identified (Moustapha et al. 2019). Let us assume N_{train} training input combinations of dimension M in the form $\mathcal{X}_{train}^{SVC} = \{\boldsymbol{x}_i \in \mathbb{R}^M,\ i = 1, \dots N_{train}\}$ and the corresponding labels $\mathcal{Y}_{train}^{SVC} = \left\{y_i^{SVC} = \ell_i = \{-1, +1\},\ i = 1, \dots N_{train}\right\}$ indicating the class of each combination. SVC classification is carried out according to the separating hyperplane that maximizes its distance (also known as margin) from the closest training combinations. The separating hyperplane can be defined as:

$$\{\boldsymbol{x} \in \mathbb{R}^M :\ \boldsymbol{w}^T\boldsymbol{x} + b\}, \tag{15}$$

where $\boldsymbol{w}$ is the vector of hyperplane coefficients and b is the bias. The perpendicular distance of any input combination from this hyperplane is:

$$d(\boldsymbol{x}_i) = \frac{|\boldsymbol{w}^T\boldsymbol{x}_i + b|}{\|\boldsymbol{w}\|}. \tag{16}$$

It turns out that maximizing the margin corresponds to the minimization of the norm of $\boldsymbol{w}$ under some constraints. Therefore, determining the separating hyperplane reduces to the following optimization problem:

$$\min_{w}\ \frac{1}{2}\|w\|^2, \quad \text{subject to} \quad y_i^{SVC}(\boldsymbol{w}^T\boldsymbol{x}_i + b) - 1 \geq 0, \quad i = \{1, \dots N_{train}\}, \tag{17}$$

where the constraints ensure that no samples can lie inside the area covered by the margin. The optimization problem is convex and it can be solved by introducing the Lagrange multipliers. After some algebra, the final optimization problem becomes:

$$\begin{gathered}\min_{\alpha}\ -\frac{1}{2}\sum_{i=1}^{N_{train}}\sum_{j=1}^{N_{train}} \alpha_i\alpha_j y_i^{SVC} y_j^{SVC} \boldsymbol{x}_i^T\boldsymbol{x}_j + \sum_{i=1}^{N_{train}} \alpha_i\,, \\ \text{subject to}\quad \textstyle\sum_{i=1}^{N_{train}} \alpha_i y_i^{SVC} = 0, \quad \alpha_i \geq 0, \quad i = \{1, \dots N_{train}\}.\end{gathered} \tag{18}$$

After finding the Lagrange multipliers $\{\alpha_i, i = 1, \dots N_{train}\}$ and the bias b, the SVC classification of a new configuration can be expressed in terms of training input combinations:

$$\hat{y}^{SVC}(\boldsymbol{x}_i) = \ell(\boldsymbol{x}_i) = sign\left(\sum_{i=1}^{N_{train}} \alpha_i y_i \boldsymbol{x}_i^T\,\boldsymbol{x} + b\right). \tag{19}$$

In some situations the optimization problem becomes unfeasible. A new solution is provided by allowing misclassifications, i.e. by relaxing the inequality constraints through the introduction of the so-called slack terms ξ_i, which measures the distance of the misclassified sample from its actual class. A penalized objective

function is obtained in which the slack terms are minimized. Two final expressions are obtained according to the type of penalization:

- Linear penalization

$$\min_{\alpha} \; -\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N_{train}}\xi_i \quad \text{subject to} \quad y_i^{SVC}(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i = \{1, \dots N_{train}\}. \tag{20}$$

- Quadratic penalization

$$\min_{\alpha} \; -\frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{C}{2}\sum_{i=1}^{N_{train}}\xi_i^2 \quad \text{subject to} \quad y_i^{SVC}(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i = \{1, \dots N_{train}\}. \tag{21}$$

In the case where the data are not linearly separable, the training combinations are mapped into a higher dimensional space referred to as feature space and, therefore, the construction of the optimal separating hyperplane is shifted to this new space. A new classification formula is given by the sign of the following expression:

$$\boldsymbol{w}^T\Phi(\boldsymbol{x}) + b = \sum_{i=1}^{N_{train}} \alpha_i y_i^{\mathrm{SVC}}\Phi(\boldsymbol{x_i})^T\Phi(\boldsymbol{x}) + b, \tag{22}$$

where $\Phi(\bullet)$ is the mapping function and hence the components of $\boldsymbol{x}$ in the feature space are $(\Phi_1(\boldsymbol{x}), \dots \Phi_M(\boldsymbol{x}))$. The expression in equation (22) shows how, if one is able to calculate the inner product of the two vector images in the feature space, i.e., $\Phi(\boldsymbol{x_i})^T\Phi(\boldsymbol{x})$, no further cumbersome operations need to be carried out in that space. This operation is named "kernel trick" since it is conducted thanks to kernel functions. Several examples of kernel functions are available in literature (e.g., Polynomial, Gaussian, Exponential etc.). Once the kernel function k has been chosen, the final classification reads:

$$\hat{y}^{SVC}(\boldsymbol{x}_i) = \ell(\boldsymbol{x}_i) = sign\left(\sum_{i=1}^{N_{train}} \alpha_i y_i \, ker(\boldsymbol{x_i}, \boldsymbol{x}) + b\right). \tag{23}$$

References

Akaike, H. (1974): A New Look at the Statistical Model Identification. In *IEEE Trans. Automat. Contr.* 19 (6), pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

Allen, D (1971). The prediction sum of squares as a criterion for selecting prediction variables. Technical Report 23, Department of Statistics, University of Kentucky.

Archer, G. E. B.; Saltelli, A.; Sobol, I. M. (1997): Sensitivity measures, anova-like Techniques and the use of bootstrap. In *Journal of Statistical Computation and Simulation* 58 (2), pp. 99–120. DOI: 10.1080/00949659708811825.

Auder, B., De Crecy, A., Iooss, B., Marques, M. (2012). Screening and metamodeling of computer experiments with functional outputs. Application to thermal-hydraulic computations. Reliability Engineering and System Safety 107, 122-131.

Baraldi, P., Pedroni, N., Zio, E. (2009). Application of a Niched Pareto Genetic Algorithm for Selecting Features for Nuclear Transients Classification, International Journal of Intelligent Systems, Volume 24, Issue 2, pp. 118-151, DOI: 10.1002/int.20328.

Basudhar, Anirban; Missoum, Samy; Harrison Sanchez, Antonio (2008): Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains. In *Probabilistic Engineering Mechanics* 23 (1), pp. 1–11. DOI: 10.1016/j.probengmech.2007.08.004.

Benkedjouh, T., Medjaher, K., Zerhouni, N., Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression, Eng. Appl Artif. Intell. 26, 1751–1760. http://dx.doi.org/10.1016/j.engappai.2013.02.006.

Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., & McFarland, J. M. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. AIAA Journal, 46(10), 2459-2468. doi:10.2514/1.34321.

Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A. (2015) Recent advances and emerging challenges of feature selection in the context of big data, Knowl.-Based Syst. 86, 33–45. http://dx.doi.org/10.1016/j.knosys.2015.05.014.

Borgonovo, E. (2007): A new uncertainty importance measure. In *Reliability Engineering & System Safety* 92 (6), pp. 771–784. DOI: 10.1016/j.ress.2006.04.015.

Borgonovo, E.; Castaings, W.; Tarantola, S. (2012): Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. In *Environmental Modelling & Software* 34, pp. 105–115. DOI: 10.1016/j.envsoft.2011.06.006.

Borgonovo, Emanuele; Plischke, Elmar (2016): Sensitivity analysis: A review of recent advances. In *European Journal of Operational Research* 248 (3), pp. 869–887. DOI: 10.1016/j.ejor.2015.06.032.

Boroson, Ethan; Missoum, Samy (2017): Stochastic optimization of nonlinear energy sinks. In *Struct Multidisc Optim* 55 (2), pp. 633–646. DOI: 10.1007/s00158-016-1526-y.

Cadini, F.; Santos, F.; Zio, E. (2014): An improved adaptive kriging-based importance technique for sampling multiple failure regions of low probability. In *Reliability Engineering & System Safety* 131, pp. 109–117. DOI: 10.1016/j.ress.2014.06.023.

Carlos, S.; Sánchez, A.; Ginestar, D.; Martorell, S. (2013): Using finite mixture models in thermal-hydraulics system code uncertainty analysis. In *Nuclear Engineering and Design* 262, pp. 306–318. DOI: 10.1016/j.nucengdes.2013.04.030.

Chaudhuri, A., Marques, A.N., Willcox, K. (2021): mfEGRA: Multifidelity efficient global reliability analysis through active learning for failure boundary location. Structural and Multidisciplinary Optimization, https://doi.org/10.1007/s00158-021-02892-5.

Cox DD, John S. (1997) SDO: a statistical method for global optimization. In: Alexandrov MN, Hussaini MY, editors. Multidisciplinary design optimization: state-of-the-art. Philadelphia: Siam; 1997. p. 315–29.

Constantine, P.G. (2015). Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (USA), ISBN: 978-1-61197-385-3.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. (1977): Maximum Likelihood from Incomplete Data via the EM Algorithm.

Di Maio, Francesco; Nicola, Giancarlo; Zio, Enrico; Yu, Yu (2014): Ensemble-based sensitivity analysis of a Best Estimate Thermal Hydraulics model: Application to a Passive Containment Cooling System of an AP1000 Nuclear Power Plant. In *Annals of Nuclear Energy* 73, pp. 200–210. DOI: 10.1016/j.anucene.2014.06.043.

Di Maio, Francesco; Nicola, Giancarlo; Zio, Enrico; Yu, Yu (2015): Finite mixture models for sensitivity analysis of thermal hydraulic codes for passive safety systems analysis. In *Nuclear Engineering and Design* 289, pp. 144–154. DOI: 10.1016/j.nucengdes.2015.04.035.

Dubourg, V., Sudret, B., & Deheeger, F. (2013). Metamodel-based importance sampling for structural reliability analysis. Probabilistic Engineering Mechanics, 33, 47-57. doi:http://dx.doi.org/10.1016/j.probengmech.2013.02.002.

Dy, J.G.; Brodley, C.E. (2004) Feature selection for unsupervised learning, J. Mach. Learn Res 5, 845–889.

Echard, B.; Gayton, N.; Lemaire, M. (2011): AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. In *Structural Safety* 33 (2), pp. 145–154. DOI: 10.1016/j.strusafe.2011.01.002.

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 models. Reliab Eng Syst Saf 111: 232–240.

Erdal, D., Cirpka, O.A. (2019). Global sensitivity analysis and adaptive stochastic sampling of a subsurface-flow model using active subspaces. Hydrol. Earth Syst. Sci., 23, 3787–3805, https://doi.org/10.5194/hess-23-3787-2019.

Fauriat W, Gayton N (2014) AK-SYS: an adaptation of the AK-MCS method for system reliability. Reliab Eng Syst Saf 123:137–144.

Figueiredo, M.A.T.; Jain, A. K. (2002): Unsupervised learning of finite mixture models. In *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (3), pp. 381–396. DOI: 10.1109/34.990138.

Fodor, I. K. (2002). A Survey of Dimension Reduction Techniques. Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, 9, 1-18.

Gibbs, Alison; Su, Francis Edward (2002): On Choosing and Bounding Probability Metrics.

Gu, M. and Berger, J.O. (2016): Parallel Partial Gaussian Process Emulation for Computer Models with Massive Output, Annals Appl. Stat, 10(3):1317-1347.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

Guyon, I., & Elisseeff, A. (2006). An Introduction to Feature Extraction. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds.), Feature Extraction: Foundations and Applications (pp. 1-25). Berlin, Heidelberg: Springer Berlin Heidelberg.

Herer, C.; Dimitrov, B.; Evrard, J. M.; Lejosne, A.; Wattelle, E. (2019): IRSN Activities related to Passive Safety Systems Assessment. In : ICAPP 2019 - International Congress on Advances in Nuclear Power Plants.

Higdon D, Geelhood K, Williams B, Unal C. (2013). Calibration of tuning parameters in the FRAPCON model. Ann Nucl Energy;52:95–102. https://doi.org/10.1016/j.anucene.2012.06.018.

Hrinda, G. A. (2010): Snap-through instability patterns in truss structures. In : Proceedings of the 51st AIAA/ASME/ASCE/AHS/ASC Dynamics, and Material Conference.

Holden AJ, Robbins DJ, Stewart WJ, Smith DR, Schultz S, Wegener M, et al. (2006). Reducing the Dimensionality of Data with Neural Networks; 313:504–7.

Hu, Y., Baraldi, P., Di Maio, F., Zio, E. (2017). A Systematic Semi-Supervised Self-adaptable Fault Diagnostics approach in an evolving environment. Mechanical Systems and Signal Processing 88, 413–427.

Huang, X.; Chen, J.; Zhu, H. (2016). Assessing small failure probabilities by AK–SS: An active learning method combining Kriging and Subset Simulation, Struct. Saf. 59, 86–95.

Inselberg, Alfred (2009): Parallel Coordinates. Visual Multidimensional Geometry and its Application: Springer International Publishing.

Iooss, Bertrand (2009): Numerical Study of the Metamodel Validation Process, 2009.

Jin, R.; Chen, W.; Simpson, T. W. (2001): Comparative studies of metamodelling techniques under multiple modelling criteria 2001.

Jolliffe, I.T., 2002. Principal Component Analysis, 2nd ed. Springer-Verlag New York.

Kasarapu, Parthan; Allison, Lloyd (2015): Minimum message length estimation of mixtures of multivariate Gaussian and von Mises-Fisher distributions. In *Mach Learn* 100 (2-3), pp. 333–378. DOI: 10.1007/s10994-015-5493-0.

Kleijnen, Jack P.C. (2009): Kriging metamodeling in simulation: A review. In *European Journal of Operational Research* 192 (3), pp. 707–716. DOI: 10.1016/j.ejor.2007.10.013.

Lanfredini, M.; Bersano, A.; D'Auria, F. (2020): A Demonstrative Application of a Methodology for Thermal-Hydraulics Passive Systems Reliability Assessment - Extreme Cases Analysis. In : Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference, 2020.

Lataniotis, C.; Marelli, S.; Sudret, B. (2020): Extending classical surrogate modelling to high dimensions through supervised dimensionality reduction: a data-driven approach. International Journal for Uncertainty Quantification, Volume 10, Issue 1, pp. 55-82; DOI: 10.1615/Int.J.UncertaintyQuantification.2020031935.

Lataniotis, C.; Wicaksono, D.; Marelli, S.; Sudret, B. (2019): UQLab user manual – Kriging (Gaussian process modeling). Report # UQLab-V1.3-105, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland 2019.

Law, M.H.C.; Figueiredo, M.A.T.; Jain, A. K. (2004): Simultaneous feature selection and clustering using mixture models. In *IEEE Trans. Pattern Anal. Machine Intell.* 26 (9), pp. 1154–1166. DOI: 10.1109/TPAMI.2004.71.

Lelièvre, N.; Beaurepaire, P.; Mattrand, C.; Gayton, N. (2018) AK-MCSi: A Kriging-based method to deal with small failure probabilities and time-consuming models, Struct. Saf. 73, 1–11.

Liu, H., & Motoda, H. (2012). Feature selection for knowledge discovery and data mining (Vol. 454): Springer Science & Business Media.

Loeppky, Jason L.; Moore, Leslie M.; Williams, Brian J. (2010): Batch sequential designs for computer experiments. In *Journal of Statistical Planning and Inference* 140 (6), pp. 1452–1464. DOI: 10.1016/j.jspi.2009.12.004.

Lu, Z.Y.; Lu, Z.Z.; Wang, P. (2015). A new learning function for Kriging and its applications to solve reliability problems in engineering, Comput. Math. Appl. 70, 1182–1197.

Marelli, Stefano; Sudret, Bruno (2014): UQLab: A Framework for Uncertainty Quantification in Matlab. In : 2nd International Conference on Vulnerability, Risk Analysis and Management (ICVRAM), Liverpool, United Kingdom, 2014.

Martin, Jay D.; Simpson, Timothy W. (2005): Use of Kriging Models to Approximate Deterministic Computer Models. In *AIAA Journal* 43 (4), pp. 853–863. DOI: 10.2514/1.8650.

McKay, M. D.; Beckham, R. J.; Conover, W. J. (1979): A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code 1979.

Missoum, S.; Gürdal, Z.; Gu, W. (2002): Optimization of nonlinear trusses using a displacement-based approach. In *Struct Multidisc Optim* 23 (3), pp. 214–221. DOI: 10.1007/s00158-002-0179-1.

Monisha R, Mrinalini R, Britto MN, Ramakrishnan R, Rajinikanth V. (2019). Smart Intelligent Computing and Applications. vol. 104. https://doi.org/10.1007/978-981-13-1921-1.

Moustapha, M.; Lataniotis, C.; Marelli, S.; Sudret, B. (2019): UQLab user manual – Support vector machines for classification. Report # UQLab-V1.3-112, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland.

Moustapha, Maliki; Sudret, Bruno (2019): A Two-stage Surrogate Modeling Approach for the Approximation of Models with Non-smooth Outputs. In : UNCECOMP 2019 - 3rd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Science and Engineering, pp. 357–366.

Nagel JB, Rieckermann J, Sudret B (2020). Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation. Reliab Eng Syst Saf;195:106737. https://doi.org/10.1016/j.ress.2019.106737.

Nossent, Jiri; Elsen, Pieter; Bauwens, Willy (2011): Sobol' sensitivity analysis of a complex environmental model. In *Environmental Modelling & Software* 26 (12), pp. 1515–1525. DOI: 10.1016/j.envsoft.2011.08.010.

Olvier, Jonathan; Baxter, Rohan; Wallace, Chris (1996): Unsupervised Learning using MML.

Pedroni, Nicola; Zio, Enrico (2017): An Adaptive Metamodel-Based Subset Importance Sampling approach for the assessment of the functional failure probability of a thermal-hydraulic passive system. *Applied Mathematical Modelling* 48, pp. 269-288.

Picheny, Victor; Ginsbourger, David; Routsant, Olivier; Haftka, Raphael T.; Kim, Nam-Ho (2010): Adaptive Designs of Experiments for Accurate Approximation of a Target Region of target region 2010.

Pierro, Franco; Araneo, Dino; Galassi, Giorgio; D'Auria, Francesco (2009): Application of REPAS Methodology to Assess the Reliability of Passive Safety Systems. In *Science and Technology of Nuclear Installations* 2009, pp. 1–18. DOI: 10.1155/2009/768947.

Puppo, L., Pedroni, N., Bersano, A., Di Maio, F., Bertani, C., Zio, E. (2021). Failure identification in a nuclear passive safety system by Monte Carlo simulation with adaptive Kriging. Nuclear Engineering and Design, Volume 380, 111308.

Razaaly N, Congedo, PM (2018) Novel algorithm using active metamodel learning and importance sampling: application to multiple failure regions of low probability. J Comput Phys 368:92–114

Razavi, Saman; Gupta, Hoshin V. (2015): What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in Earth and Environmental systems models. In *Water Resour. Res.* 51 (5), pp. 3070–3092. DOI: 10.1002/2014WR016527.

Roma, G., Di Maio, F., Bersano, A., Pedroni, N., Bertani, C., Mascari, F., Zio, E. (2021). A Bayesian framework of inverse uncertainty quantification with principal component analysis and Kriging for the reliability analysis of passive safety systems. Nuclear Engineering and Design, Volume 379, 111230. DOI: https://doi.org/10.1016/j.nucengdes.2021.111230.

Saeys, Y., Inza, I., Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics, Bioinformatics 23, 2507–2517. http://dx.doi.org/10.1093/bioinformatics/btm344.

Saltelli, A.; Andres, H.; Homma, T. (1993): Sensitivity Analysis of Model Output. An Investigation of New Techniques.

Saltelli, A.; Marivoet, J. (1990): Non-parametric statistics in sensitivity analysis for model output: A comparison of selected techniques. In *Reliability Engineering and System Safety*.

Saltelli, A.; Sobol, I. M. (1995): About the use of rank transformation in sensitivity analysis of model output.

Saltelli, Andrea; Annoni, Paola; Azzini, Ivano; Campolongo, Francesca; Ratto, Marco; Tarantola, Stefano (2010): Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. In *Computer Physics Communications* 181 (2), pp. 259–270. DOI: 10.1016/j.cpc.2009.09.018.

Saltelli, Andrea; Ratto, Marco; Andres, Terry; Campolongo, Francesca; Cariboni, Jessica; Gatelli, Debora et al. (2008): Global Sensitiviy Analysis. The Primer: John Wiley & Sons.

Schöbi, R., Sudret, B., & Marelli, S. (2017). Rare Event Estimation Using Polynomial-Chaos Kriging. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 3(2), D4016002.

Schwarts, Gideon (1978): Estimating the Dimension of a Model.

Sedlmair, M.; Munzner, T.; Tory, M. (2013): Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices.

Statovic (2020): Flexible mixture models for automatic clustering. Available online at https://it.mathworks.com/matlabcentral/fileexchange/72310-flexible-mixture-models-for-automatic-clustering.

Sudret, Bruno (2008): Global sensitivity analysis using polynomial chaos expansions. In *Reliability Engineering & System Safety* 93 (7), pp. 964–979. DOI: 10.1016/j.ress.2007.04.002.

Tong, C.; Sun, Z.; Zhao, Q.; Wang, Q.; Wang, S. (2015) A hybrid algorithm for reliability analysis combining Kriging and subset simulation importance sampling, J. Mech. Sci. Technol. 29, 3183–3193.

Turati, Pietro; Cammi, Antonio; Lorenzi, Stefano; Pedroni, Nicola; Zio, Enrico (2018a): Adaptive simulation for failure identification in the Advanced Lead Fast Reactor European Demonstrator. In *Progress in Nuclear Energy* 103, pp. 176–190. DOI: 10.1016/j.pnucene.2017.11.013.

Turati, Pietro; Pedroni, Nicola; Zio, Enrico (2017): Simulation-based exploration of high-dimensional system models for identifying unexpected events. In *Reliability Engineering & System Safety* 165, pp. 317–330. DOI: 10.1016/j.ress.2017.04.004.

Turati, Pietro; Pedroni, Nicola; Zio, Enrico (2018b): Knowledge-driven System Simulation for Scenario Analysis in Risk Assessment. In: T. Aven, E. Zio (Eds.), Knowledge in Risk Assessment and Management, First Edition, pp. 165-220, John Wiley & Sons Ltd, 2018.

Vapnik, Vladimir; Cortes, Corinna (1995): Support-Vector Networks.

Verikas, A., Bacauskiene, M., (2002). Feature selection with neural networks, Pattern Recognit. Lett. 23, 1323–1335. http://dx.doi.org/10.1016/S0167-8655(02)00081-8.

Verleysen, M. and D. François (2005). The curse of dimensionality in data mining and time series prediction. In J. Cabestany, A. Prieto, and F. Sandoval (Eds.), Computational Intelligence and Bioinspired Systems, Volume 3512 of Lecture Notes in Computer Science, pp. 758–770. Springer Berlin Heidelberg

Wallace, C. S.; Boulton, D. M. (1968): An Information Measure for Classification.

Wang Y, Yao H, Zhao S (2016). Auto-encoder based dimensionality reduction. Neurocomputing; 184:232–42. https://doi.org/10.1016/j.neucom.2015.08.104.

Wu, Xu; Kozlowski, Tomasz; Meidani, Hadi; Shirvan, Koroush (2018): Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory. In *Nuclear Engineering and Design* 335, pp. 339–355. DOI: 10.1016/j.nucengdes.2018.06.004.

Xiao, Ning-Cong; Zuo, Ming J.; Zhou, Chengning (2018): A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis. In *Reliability Engineering & System Safety* 169, pp. 330–338. DOI: 10.1016/j.ress.2017.09.008.

Yang X, Liu Y,Mi C et al (2018) Active learning Kriging model combining with kernel-density-estimation-based importance sampling method for the estimation of low failure probability. J Mech Des 140:051402

Yang, Xufeng; Cheng, Xin (2020): Active learning method combining Kriging model and multimodal-optimization-based importance sampling for the estimation of small failure probability. In *International Journal for Numerical Methods in Engineering* 121 (21), pp. 4843-4864. Doi: 10.1002/nme.6495.

Yang, Xufeng; Cheng, Xin; Wang, Tai; Mi, Caiying (2020): System reliability analysis with small failure probability based on active learning Kriging model and multimodal adaptive importance sampling. In *Structural and Multidisciplinary Optimization* 62, pp. 581–596.

Zhao, H., Gao, Z., Xu, F., Xia, L. (2021): Adaptive multi-fidelity sparse polynomial chaos-Kriging metamodeling for global approximation of aerodynamic data. Structural and Multidisciplinary Optimization, https://doi.org/10.1007/s00158-021-02895-2.

Zhang, Z., Chen, H., Xu, Y., Zhong, J., Lu, N., Chen, S. (2015). Multisensor-based real-time quality monitoring by means of feature extraction, selection and modeling for Al alloy in arc welding, Mech. Syst. Signal Process 60–61, 151–165. http://dx.doi.org/10.1016/j.ymssp.2014.12.021.

Zhang, X., Lu, Z., Cheng, K. (2021): AK-DS: An adaptive Kriging-based directional sampling method for reliability analysis. Mechanical Systems and Signal Processing 156, 107610.

Zio, Enrico; Apostolakis, George E., Pedroni, Nicola (2010). Quantitative functional failure analysis of a thermal-hydraulic passive system by means of bootstrapped Artificial Neural Networks. *Annals of Nuclear Energy* 37(5), pp. 639-649.

Zio, Enrico; Pedroni, Nicola (2009). Functional failure analysis of a thermal-hydraulic passive system by means of Line Sampling. *Reliability Engineering and System Safety* 94(11), pp. 1764-1781.

Zio, Enrico; Pedroni, Nicola (2011). How to effectively compute the reliability of a thermal-hydraulic nuclear passive system. *Nuclear Engineering and Design* 241(1), pp. 310-327.