Nuclear Engineering and Design 380 (2021) 111308

Contents lists available at ScienceDirect

Nuclear Engineering and Design

journal homepage: www.elsevier.com/locate/nucengdes

Failure identification in a nuclear passive safety system by Monte Carlo simulation with adaptive Kriging

L. Puppo [a], N. Pedroni [a,*], A. Bersano [a], F. Di Maio [b], C. Bertani [a], E. Zio [b,c,d]

[a] Energy Department, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin 10129, Italy
[b] Energy Department, Politecnico di Milano, Via La Masa 34, Milano 20156, Italy
[c] MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France
[d] Department of Nuclear Engineering, College of Engineering, Kyung Hee University, South Korea

ARTICLE INFO

Keywords:
Nuclear power plant
Passive safety system
Critical failure region
Decay heat removal
Kriging
Adaptive sampling
Adaptive-Kriging Monte Carlo Sampling (AK-MCS)

ABSTRACT

Passive Safety Systems (PSSs) are increasingly employed in advanced Nuclear Power Plants (NPPs). Their safety performance is evaluated through computationally expensive Thermal-Hydraulic (T-H) simulations models and the identification of the operational conditions which lead to unsafe conditions (the so-called Critical failure Regions, CRs) may be challenging.

In the present paper, a computational framework is proposed to identify the CRs of a generic passive Decay Heat Removal (DHR) system of a NPP. A time-demanding Best-Estimate Thermal-Hydraulic (BE-TH) model of the system is used to train a fast-running metamodel embedded within an adaptive sampling technique of literature, namely Adaptive Kriging Monte Carlo Sampling (AK-MCS), so as to provide increased accuracy in proximity of the failure threshold and identify which input values lead the PSS to failure. To the best authors' knowledge this is the first time that the metamodel-based AK-MCS technique is applied for the identification of the CRs of a PSS of an NPP.

1. Introduction

In recent years, important efforts have been made for the design and development of Passive Safety Systems (PSSs) to increase the safety level of Nuclear Power Plants (NPPs). Innovative PSSs are being employed in new advanced nuclear reactors to provide the main safety functions, e.g., decay heat removal, reactivity control and fission product containment. Obviously, the conditions that lead PSSs to failure must be determined, analyzed and avoided (Jafari et al., 2003; Burgazzi, 2004). This implies considering the influence of many operational and environmental parameters on the PSS T-H performance and its dependence on natural driving forces (e.g. gravity, natural circulation etc.), and properly treating their uncertainties in a sufficiently wide range of *operational* conditions (Herer et al., 2019).

For this, the REPAS (Reliability Evaluation of Passive Safety Systems) methodology (Jafari et al., 2003; Pierro et al., 2009) has been developed to provide a structured procedure for conducting a T-H reliability assessment of PSSs accounting for uncertainties. However, in its original formulation REPAS does not explicitly include the detailed characterization of the PSS Critical failure Regions (CRs). This is necessary to define the configurations of critical operation for the system, i.e., those combination of values of PSS design and/or operational parameters which lead the PSS to fail providing its safety function. In mathematical terms, given the Input/Output (I/O) representation, $Y = f(\boldsymbol{X})$, of the PSS behaviour, a specific combination of input parameters values $\boldsymbol{x}$ is critical, if the resulting output value (e.g., the decay heat removed by the PSS) is lower (higher) than a predefined threshold, $y = f(\boldsymbol{x}) \leq (\geq) Y_{thres}$, representing the limit value for the system operation. These combinations define the CRs, whose determination is a challenging problem, which can be addressed with computational methods (Cadini et al., 2014; Picheny et al., 2010; Turati et al., 2017, 2018a, 2018b). In these methods, Best-Estimate Thermal-Hydraulic (BE-TH) models are not directly used to numerically compute the PSS response in the many accidental scenarios that need to be considered, because the computational cost for the high number of code runs required could become excessive. For this reason, advanced computational methods are being studied to reduce the cost of computation. On one side, fast-running surrogate metamodels can be adopted to mimic the behaviour of the computationally demanding, original T-H simulator and replace it in the analysis. On the other side, intelligent adaptive sampling strategies may be implemented to efficiently trace the CR boundary (i.e., the PSS limit

* Corresponding author.
E-mail address: nicola.pedroni@polito.it (N. Pedroni).

https://doi.org/10.1016/j.nucengdes.2021.111308
Received 13 December 2020; Received in revised form 3 May 2021; Accepted 21 May 2021
Available online 2 June 2021

Nomenclature

Acronyms

AK-MCS	Adaptive Kriging Monte Carlo Sampling
AV	Activation Valve
BE-TH	Best Estimate Thermal Hydraulic
CR	Critical (failure) Region
CV	Cross-Validation
DHR	Decay Heat Removal
DoE	Design of Experiment
E-HX	Emergency Heat Exchanger
FC	Failure Criterion
GA	Genetic Algorithm
GP	Gaussian Process
I/O	Input/Output
LHS	Latin Hypercube Sampling
MCS	Monte Carlo Sampling
MSIV	Main Steam Isolation Valve
NPP	Nuclear Power Plant
NRMSE	Normalized Root-Mean-Square Error
PCP	Parallel Coordinates Plot
PV	Pressure Vessel
PSS	Passive Safety System
QI	Quality Indicator
REPAS	Reliability Evaluation of Passive Safety Systems
RMSE	Root-Mean-Square Error
SBO	Station Black-Out
SRV	Safety Relief Valve
T-H	Thermal Hydraulic
TPI	Transient Performance Indicator

Symbols

A_{AV}	Activation Valve flow area
A_{MSIV}	Main Steam Isolation Valve flow area
β	Trend coefficients of Kriging approximation
$\widehat{\beta}$	Trend coefficients least square estimates
D	Domain
DEL_{AV}	Delay of Activation Valve opening
DEL_{MSIV}	Delay of Main Steam Isolation Valve closure
E_{ex}	Energy exchanged
$E_{ex,\%}$	Percentage of energy exchanged
$\in LOO_{abs}$	Absolute Leave-One-Out error
$\in LOO_{norm}$	Normalized Leave-One-Out error
f	Generic model function
H	Kriging metamodel Information matrix
h	Kriging metamodel trend
h	Arbitrary function of Kriging trend
i	Input combination index
K	Number of partitions of an I/O set
M	Problem dimensionality
$\mu_{\widehat{y}}$	Mean value of a metamodel prediction
$\mathscr{N}$	Normal Gaussian distribution
N_{cand}	Number of best candidates in AK-MCS procedure
N_{MCS}	Number of samples generated by MCS
N_{train}	Number of training samples
N_{val}	Number of validation samples
$NC\%$	Non-condensable gases percentage
n	Iteration number
n_{fin}	Final number of iterations
P	Number of arbitrary functions of Kriging trend
Pr	Probability
p_{max}	Maximum value of pressure
Q	Predictivity indicator
R	Correlation matrix
r	Vector of cross correlations between input vectors
r	element of the cross correlations vector
σ	Standard deviation
$\overline{\sigma}$	Average standard deviation
$\sigma_{\widehat{y}}$	Estimation error of a metamodel prediction
θ	Kriging approximation hyperparameters
U	U learning function
Var	Variance
X	Generic input
$\mathscr{X}$	Set of model input vectors
$\mathscr{X}^*$	Input vectors of the set of best candidates
$\mathscr{X}_{train}$	Set of training input vectors
$\mathscr{X}_{val}$	Set of validation input vectors
x	Model input vector
x	Model input parameter
Y	Generic output
Y_{thres}	Threshold output value
$\mathscr{Y}^*$	Outputs of the set of best candidates
$\mathscr{Y}_{train}$	Set of training outputs
$\mathscr{Y}_{val}$	Set of validation outputs
$\widehat{\mathscr{Y}}$	Set of metamodel predictions
$\widehat{\mathscr{Y}}_{val}$	Set of predictions of the validation outputs
y	Model output
$\widehat{y}$	Metamodel prediction output
$\overline{y}_{val}$	Average validation output value
Z	Unit variance stationary Gaussian Process of Kriging metamodel construction

surface bounding the CR), with minimum waste of computational time for samples far from the CR.

With respect to this latter point, the goal of adaptive sampling is, then, to find the best Design of Experiments (DoE) with the smallest number of samples (and, thus, of time-demanding simulations) (Garud et al., 2017). In general terms, samples are selected iteratively to fill the search domain (in this case, the PSS input parameters space) in such a way that any discontinuity or key feature are not missed (namely, good exploration), and, at the same time, the search is focused on regions that have been identified as potentially interesting because close to the CR (namely, good exploitation) (Crombecq et al., 2011b).

With respect to the former point, surrogate models, or metamodels, can accelerate the collection of experiments. The general idea consists in finding an approximation function that is constructed on multiple simulations at key points of the design space (training set) and on the analysis of the outcomes of such simulations (Crombecq et al., 2011a). This function manages not only to mimic the results of the samples in the DoE, but also to provide a good estimate of the (true) model output Y in correspondence of other input values of the domain. Gaussian Processes (GPs) have been extensively used for this purpose because they have been shown capable of reproducing numerous system responses (Ranjan et al., 2008), while providing the estimation confidence without adding further complexity. Kriging models add non-stationarity to the GPs (Turati et al., 2017) which is very useful if adaptive techniques are applied for exploitation.

An example of combination of adaptive sampling and Kriging metamodeling, known as AK-MCS (Echard et al., 2011), is here adopted and tailored to obtain the CRs of a PSS. The objective of the paper is, thus, to present the computational framework and show its feasibility, advantages and effectiveness for the PSS CRs exploration task. On the contrary, the

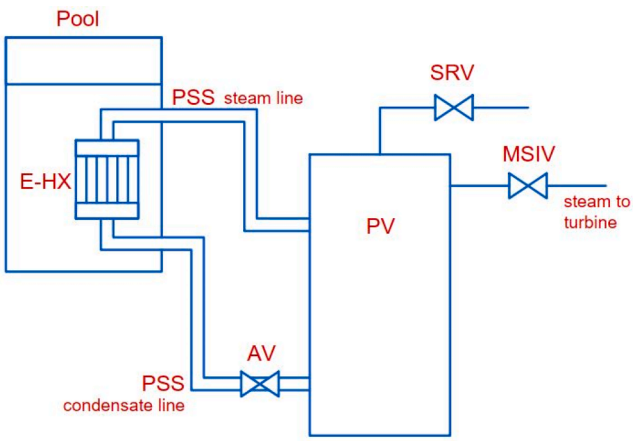

Fig. 1. PSS simplified sketch.

aim is not to carry out the complete reliability analysis of a specific PSS. The PSS considered is a Decay Heat Removal (DHR) system, based on an Emergency Heat eXchanger (E-HX), for removing the heat after the reactor shut down due to an accident initiation (specifically, a Station Black-Out (SBO) accident, in this work), whose operation is regulated by the simultaneous action of two valves. To the best authors' knowledge this is the first time that the metamodel-based AK-MCS technique is applied for the identification of the CRs of a PSS of an NPP. As a final remark, notice that the identification of CRs has a radically *different meaning* and *objective* with respect to classical Uncertainty Analysis (UA), even if the sampling techniques that can be adopted in the two tasks may be similar. The goal of the former is to identify and characterize the combinations of values of PSS design and/or operational *input* variables leading to functional failure, which is strictly related to PSS *thermal–hydraulic behavior*. The objective of the latter is to *propagate* the uncertainty affecting the *computer code* (e.g., its models, correlations, parameters, …) to the corresponding outputs of interest (Alcaro et al., 2021). In this work, we are not performing any uncertainty calculation and the parameters selected are related to the PSS design and operation and not to the code used to simulate it.

The paper is organized as follows: in Section 2, the case study is presented; Section 3 offers an exhaustive description of the AK-MCS algorithm tailored to the CR exploration, whereas in Section 4 the AK-MCS technique is applied to the PSS described in Section 2; in Section 5, some results are reported and, finally, in Section 6 some conclusions are drawn together with a perspective for future developments.

2. Case study

2.1. Description of the passive safety system and definition of its failure criteria

The PSS considered is a generic DHR system operating in natural circulation to remove the decay heat from a BWR reactor core in accidental conditions. The main components of the PSS are (see Fig. 1):

- an E-HX that condenses the steam produced in the Pressure Vessel (PV). It is composed by two cylindrical headers and a bundle of vertical straight pipes. The E-HX is submerged in a water pool;
- a steam line connecting the top of the PV to the inlet of the E-HX;
- a condensate line connecting the outlet of the E-HX back to the PV. On the pipe it is installed an Activation Valve (AV) that opens to trigger the PSS operation.

The PSS function is to remove the decay heat in accidental conditions. In the present case it has been considered an SBO with reactor shut down. The PSS operation should prevent the energy increase within the PV, which may lead to over-pressurization and over-heating of the various components. The reactor normally operates in steady state conditions at a pressure of about 70 bar, the Main Steam Isolation Valve (MSIV) is open and the steam produced in the PV is directed to the turbine.

During normal conditions the PSS is not active. The PSS steam line is initially filled with saturated steam at 70 bar, with possible presence of a certain amount of non-condensable gases. The PSS condensate line initially contains subcooled liquid assumed at 40 °C and 70 bar. The AV is initially closed, preventing the connection of the PSS condensate line with the PV. The pool is filled with water initially at 40 °C and the water level is above the E-HX upper header.

When the SBO accident occurs, the reactor is shut down and the MSIV closes. Simultaneously the AV opens, so that the vapor from the PV is directed into the PSS. Vapor condenses inside the E-HX due to the heat transferred to the pool and the condensate flows back to the PV through the condensate line.

A preliminary list of parameters possibly affecting the PSS operation has been identified based on *expert judgment*. The selected parameters are related only to the PSS thermal–hydraulic behavior and *not* to the *code* adopted to simulate it[1]. Considering that: (i) the aim of the paper is to show an exemplificative application of a metamodel-based methodology (AK-MCS) for the identification of the critical failure region; and (ii) parametrising and training a metamodel becomes hard or even intractable as the number of input parameters increases (see Section 3), *only a relatively small subset* of the identified parameters has been considered for this study. In particular, a simple *one-at-a-time sensitivity analysis* has been performed to select the parameters mostly affecting the PSS behavior. Five input parameters $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$have been identified as possibly relevant for the response of the PSS during the SBO accident:

1. *AV flow area* (A_{AV}): the opening of the AV triggers the PSS operation when the accident occurs.
2. *AV opening delay* (DEL_{AV}): the AV may open with a certain delay with respect to the beginning of the accidental sequence.
3. *MSIV residual flow area* (A_{MSIV}): the MSIV should close completely when the accident occurs, but some leakage may be present (i.e., normalized flow area > 0%).
4. *MSIV closure delay* (DEL_{MSIV}): the MSIV may close with a certain delay with respect to the beginning of the accidental sequence.
5. *Percentage of non-condensable gases* $(NC\%)$: in the PSS lines a certain amount of non-condensable gases may build up during the system operation. These gases tend to accumulate in the coldest regions of the system, where the vapor partial pressure is the lowest. Their quantity is expressed in terms of percentage of volume occupied by the non-condensable gases with respect to the total volume of the steam line.

To generate different combination of values of the five parameters,x_m $(m = 1, ..5)$, uniform probability distributions have been considered to span their ranges of variation and, thus, explore their possible combination of values in the search for the CRs. The parameter ranges have been selected based on *preliminary sensitivity calculations* driven by *expert judgment* as a compromise between two "competing objectives": on one side, they should be wide enough to allow a proper analysis and exploration of the failure domain (i.e., to contain a satisfactory amount of failure configurations); on the other side, they should

[1] For example, correction factors have been not included among the varying parameters, since they are mainly related to the uncertainty of the calculation. In a complete PSS analysis within a safety review process these two aspects should be obviously combined. Therefore, the identification of the *failure region* (related to the passive system *operation*) should be considered together with its *uncertainty* (related to the *computational tool*). However, this is outside the scope of the present paper.

Table 1
Ranges of variation of the input variables.

Input		Symbol	Range of variation
AV flow area	(%)	A_{AV}	$(0,100)$
AV opening delay	(s)	DEL_{AV}	$(0,720)$
MSIV residual flow area	(%)	A_{MSIV}	$(0,0.15)$
MSIV closure delay	(s)	DEL_{MSIV}	$(0,7200)$
Non-condensable gases percentage	(%)	$NC\%$	$(0,40)$

not be too large, to avoid sampling simulations points too far from the critical failure region itself. Table 1 lists the ranges of variation of the input parameters. For the AV and MSIV valves, the actuation time has not been considered since it is negligible with respect to the opening and closure delay.

The DHR system successful response to the accident is measured in terms of its heat removal function and, specifically, in relation to the amount of heat removed during the accidental transient, lasting about 8 h. If the heat is not removed adequately, the temperature and pressure may dangerously rise inside the PV and if the pressure increases beyond the Safety Relief Valve (SRV) set-point, considered at 75.5 bar, the valve opens to discharge the vapor inside the containment building. Two output parameters (Y_1, Y_2) are considered as Transient Performance Indicators (TPIs) (Pierro et al., 2009) to evaluate the PSS functional response:

1. *Energy exchanged* (E_{ex}): the total amount of energy removed during the transient;
2. *Maximum of PV pressure* (p_{max}): maximum value reached by the pressure evolution inside the PV during the transient.

Table 2 lists the values of the input and output parameters for the reference transient, i.e., the "reference conditions" of nominal functioning of the DHR system. The energy exchanged output is measured calculating the percentage ($E_{ex,\%}$) with respect to the value obtained in the reference conditions.

The reference conditions allow identifying two Failure Criteria (FC):

1. *Low heat removal*: if $E_{ex,\%} < 90\%$ (Pierro et al., 2009).
2. *Steam release in the containment* if $p_{max} > 75.5bar$ (i.e., pressure rise in the PV causes the SRV to open, which leads to vapor release in the containment of the NPP)

In the present paper, the exploration of the CRs is carried out only with respect to exchanged energy output E_{ex} and, thus, successful operation of the system is defined when $E_{ex,\%} > 90\%$; otherwise, the system fails to provide its function.

2.2. Description of the PSS model

A RELAP5-3D model of the PSS described in the previous Section 2.1 has been developed in cooperation by University of Pisa and Politecnico di Torino (Lanfredini et al., 2020). The RELAP5-3D model simulates the behaviour of the DHR system connected to a simplified reactor PV and is composed of two hydrodynamic regions: the primary side (with the PV, the E-HX and the pipe connections) and the pool side.

The PV is modelled using pipe and branch components, whereas its connections to the feedwater line and the steam supply line are represented by two time-dependent volumes. On the steam supply line, the MSIV is located and modelled as a servo-valve, whereas the SRV at the top of the PV is modelled as a trip valve. The E-HX is constituted by two headers represented by branch components and a pipe component for the heat exchanger tubes. Steam and condensate lines between the PV and E-HX are represented by a series of pipe components. On the condensate line, the AV is located and modelled as a motor valve. For what concerns the pool side, branch and pipe components laterally connected through crossflow junctions have been adopted.

A more detailed description of the RELAP5 model can be found in (Lanfredini et al., 2020). Some closure equations, relevant for the operation of the PSS (e.g., condensation heat transfer within the HX tubes), have been revised and correction factors have been applied to properly simulate the occurring phenomena (Bersano et al., 2020).

Each transient simulation with the RELAP5-3D code takes about 4.30 h on a PC with *CPU Intel Core i7-7500U CPU @ 2.70 GHz dual.*

3. Metamodel-based AK-MCS for CRs exploration

The AK-MCS iterative technique, introduced in (Echard et al., 2011) and further developed in (Turati et al., 2017), is here tailored specifically for the CRs characterization of the PSS. The RELAP5-3D simulations used for the metamodel construction are called training simulations and the corresponding I/O values constitute the metamodel I/O training set $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}$. The AK-MCS iterative procedure consists of the following steps, for each n-th iteration:

1. *Construction*: a Kriging metamodel is built with the available I/O training set $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}$ (see Appendix A for details). The first I/O training set used to construct the first Kriging is called $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}_{in}$; then, the set is progressively updated and enriched in the successive iterations.
2. *Generation of random input combinations*: a large number N_{MCS} of new input combinations $\mathscr{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{N_{MCS}})$ is generated by means of Latin Hypercube Sampling (LHS) (McKay et al., 1979), so as to efficiently span the input parameters space.
3. *Metamodel Evaluation*: the Kriging metamodel is used to evaluate the output values corresponding to the $\mathscr{X}$ combinations: $\widehat{\mathscr{Y}} = \left(\widehat{y}_1, \cdots, \widehat{y}_{N_{MCS}}\right)$.
4. *Convergence check*: Convergence of the metamodel construction is verified through an a priori defined convergence (e.g., a certain error metric) or stopping criterion (e.g., a limited computational budget, expressed in the form of a maximum number of BE-TH simulations).
5. *Selection*: if convergence criterion at step 4 is not satisfied, the best candidate subset $\mathscr{X}^* \subset \mathscr{X}$ of input combinations is added to the current training set by evaluating the corresponding output values $\mathscr{Y}^*$ through the long-running BE-TH model. The N_{cand} best candidates are selected on the basis of their learning function values. Among the several examples of learning functions provided in literature (Xiao et al., 2018), the U-function (Echard et al., 2011; Turati et al., 2017) is adopted:

$$U(\boldsymbol{x}) = \frac{|Y_{thres} - \mu_{\widehat{y}}(\boldsymbol{x})|}{\sigma_{\widehat{y}}(\boldsymbol{x})}. \tag{1}$$

The $U(\boldsymbol{x})$ value represents the distance, expressed relative to standard deviation, of the metamodel prediction (whose mean value is $\mu_{\widehat{y}}(\boldsymbol{x})$ and the related estimation error is $\sigma_{\widehat{y}}(\boldsymbol{x})$) from the contour of the CR,

Table 2
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

defined by Y_{thres}. The smaller is $U(\boldsymbol{x})$, the closer is the metamodel prediction to the failure threshold and the higher the interest in adding the observation corresponding to $\boldsymbol{x}$ to the current training set, since the main scope is to focus on the limit state and to increase the metamodel accuracy in that area. However, notice that the choice of $\mathscr{X}^*$ should *not* be made *only* among the N_{cand} combinations with the lowest U-function values. In fact, in this way the corresponding inputs could result too close to each other in their domain due to a high correlation function, bringing a small amount of information to the Kriging training process; some techniques (e.g., clustering) are proposed in literature to face this issue by evenly "spreading" the candidates along the limit state surface (Turati et al., 2017).

Once the new I/O relations $\{\mathscr{X}^*, \mathscr{Y}^*\}$ have been simulated with the original model and added to the training set, steps 1 to 5 are repeated until step 4 is verified.

The Kriging metamodel obtained at the end of the iterations must provide predictions of the output with satisfactory level of accuracy, especially in proximity of the CRs limit surfaces. A large number of new input combinations $\boldsymbol{x}$ (e.g., several thousands) can, then, be generated, again with LHS, and sent in input to the Kriging metamodel, and the critical ones, i.e., $\hat{y} = f(\boldsymbol{x}) \leq Y_{thres}$, are retained for characterizing the shape and number of the CRs. In mathematical terms, this corresponds to solving the inverse problem $\boldsymbol{x} = f^{-1}(\hat{y})$, with $\hat{y} \leq Y_{thres}$.Once this is done, a graphical representation of the CRs can be provided by high dimensional data visualization techniques (Zio and Bazzo, 2011, 2012), like scatter plots or Parallel Coordinates Plot (PCP) (Inselberg, 2009). In brief, scatter plots show the two-dimensional projections of the CRs over all the possible pairs of inputs. Instead, PCP allows representing all the input combinations belonging to the CRs in a unique plot: all the M input variables, normalized on their respective ranges, are reported on vertical axes and lined up horizontally; then, each input combination is represented by a line connecting in the horizontal direction 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.

A consideration is in order with respect to the steps of the AK-MCS iterative procedure detailed above. In the present case of study (Section 2) the number of input variables selected by expert judgment is *quite small* (i.e., equal to 5), which allows: (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 training of the kriging surrogate model. However, parametrising and training a metamodel can become harder or even intractable as the number M of input parameters increases (in particular, when $M > 30$–100), 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, which is beyond the scope of this paper: see, e.g., (Gu and Berger, 2016). In general, it is very difficult (if not impossible) to provide a definitive statement about whether (and when) the AK-MCS methodology (or other metamodel-based techniques) may not be advantageous over a more traditional MCS or LHS, due to the size of the input/output parameter space at hand. This is due to several (competing) issues: (i) the effectiveness of the metamodel in mimicking the behavior of the original code *strongly* depends *also* on the properties of the underlying mathematical model $f(\boldsymbol{x})$ (e.g., its degree of complexity, nonlinearity, multimodality, discontinuity, etc.), which are unknown a priori; (ii) when we deal with long-running, detailed computer codes (like the one employed in this work), the reduction in the computational cost obtained thanks to the metamodel is always outstanding (typically of several orders of magnitude); (iii) in the presence of high-dimensional inputs, we can still resort to *dimensionality reduction* techniques before metamodel construction (step 1. above) (Turati et al., 2017, 2018a, 2018b; Lataniotis et al., 2020). In general terms, dimensionality reduction includes a number of strategies for identifying a lower-dimensional subspace of variables where it is possible to build a reduced and simplified, yet representative and understandable, model of the system behavior (Fodor, 2002; H. Liu and Motoda, 2012). From the point of view of the metamodel training and subsequent exploration of different system configurations, reducing the dimensionality of the state space to probe allows the definition of a more effective DoE. Three main strategies have been proposed in the literature: (i) feature selection, which aims at selecting a subset of the available variables and parameters input to the model (Guyon and Elisseeff, 2003); (ii) feature extraction, which aims at identifying a subset of "new" features created by means of transformations of the initial ones (Guyon and Elisseeff, 2006); and (iii) sensitivity analysis methods, which achieve the same final objective as feature selection, by ranking the factors according to their influence on the output of the model (Borgonovo and Plischke, 2016; Saltelli, 2008; Sudret, 2008). As highlighted above, in the present case none of these *structured* pre-processing steps was needed, except a *rough expert judgment-based* sensitivity analysis.

4. Application

The metamodel-based AK-MCS framework described in Section 3 has been applied for the characterization of CRs relative to the output variable "energy exchanged" (E_{ex}) of the PSS illustrated in Section 2. In the following Section, the relevant steps of the application are discussed.

4.1. I/O training set and metamodel construction

Training I/O combinations have been generated by simulations with varying values of each input x_m within its range (see Table 1). Unfortunately, no definite recommendations exist about the choice of the most suitable size of the training set (Liu, 2005; Liu et al., 2018). The criterion proposed for Kriging metamodels in (Loeppky et al., 2010) suggests a number of training combinations equal to about $10M$, where M is the dimensionality of the problem; hence, about 50 RELAP5-3D runs were necessary in this case. We proceeded, then, to build an initial I/O training set $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}_{in}$ by 64 RELAP5-3D simulations in correspondence of input values combinations generated as follows: some of them (in this case, 27) are "deterministically" selected by expert judgment to span the entire ranges of input variation, while exploring extreme parameter combinations (e.g., values of the input variables lying on the bounds of the corresponding intervals); the remaining (37) I/O patterns are randomly sampled by LHS to evenly cover the input domain.

The UQLab Software Framework for Uncertainty Quantification (Marelli and Sudret, 2014) has been used to fit the Kriging metamodel to the training set. UQLab provides straightforward parametrization of the Kriging (see Appendix A): constant, linear, polynomial, or arbitrary trends, associated to elliptic and separable correlation kernels, based on many possible one-dimensional distribution families (e.g., Exponential, Gaussian, Matérn, or user-defined). The metamodel hyperparameters can be estimated through the Cross-Validation (CV) or the Maximum-Likelihood (ML) methods, using different optimization techniques (local or global) (Lataniotis et al., 2019). The best Kriging setting for the specific case study has been established by testing different options with the CV procedure (see Appendix B). In particular, two Kriging features have been tested: the trend type and correlation function family, whereas the other features have been set to their default options defined in UQLab. The Kriging best setting has resulted to be:

- Trend type: *Linear*
- Family of correlation functions: *Matern-5_2*
- Type of correlation functions: *Ellipsoidal* (default)
- Estimation method: *CV* (default)
- Optimization method: *Genetic Algorithm (GA)* (default)

4.2. Adaptive procedure: AK-MCS

In the present section, the steps of the metamodel-based AK-MCS framework are illustrated in detail, following the structure introduced in Section 3, and tailored to the specific case study, in relation to the energy exchanged E_{ex} by the DHR during an SBO accidental transient.

1. *Construction*: a new Kriging metamodel is constructed at each n-th iteration using an I/O training set of increasing dimension, starting from the initial one $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}_{in}$ made by the 64 RELAP5-3D simulations. The Kriging features tailored on the initial training set, can be adjusted at each iteration of the adaptive procedure to improve the fit with the new training sets. The metamodel accuracy is improved specifically in proximity of the failure threshold (Y_{thres} = 90% of the energy exchanged during the reference transient).
2. *Generation of random input combinations:* $N_{MCS} = 10.000$ new input combinations, $\mathscr{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{N_{MCS}})$, are generated by LHS (see Table 1). The number of combinations N_{MCS} (=10.000) is *empirically* found to provide a satisfactory trade-off between thoroughness of PSS state space exploration and computational cost (associated to the metamodel evaluation) in the present case. On one side, the evaluation of the metamodel is *almost costless* with respect to that of the original code (i.e., it takes fractions of seconds with respect to several hours): this allows sampling a *very large* number N_{MCS} of configurations to exhaustively probe the PSS input space and perform a *reliable* selection of the N_{cand} best candidates to add to the current DoE (step 5 of the algorithm). On the other side, the effectiveness of the metamodel-based exploration depends *also* on the properties of the mathematical model $f(\boldsymbol{x})$ underlying the PSS state space (e.g., its degree of complexity, nonlinearity, multimodality, discontinuity, etc.). Since these properties are a priori unknown, an "optimal" value of N_{MCS} (if any) cannot be defined a priori. Rather, also N_{MCS} could be adaptively determined based on a set of convergence criteria, where N_{MCS} is progressively increased *only* if it is found to significantly improve the thoroughness of the exploration and to speed up the convergence of the overall methodology. Such refinement is not considered in the present study.
3. *Metamodel evaluation*: The sampled input combinations $\mathscr{X}$ are run through the metamodel to predict the corresponding output values of energy exchanged: $\widehat{\mathscr{Y}} = \left(\widehat{y}_1, \cdots, \widehat{y}_{N_{MCS}}\right)$.
4. *Convergence check*: a double convergence criterion is defined. On the one hand, the level of accuracy of the metamodel should be increased as much as possible; on the other hand, the computational cost of the successive iterations (and corresponding RELAP5-3D simulations) should be kept to a feasible level. Different criteria have been proposed in the open literature to adaptively enrich the DoE and check the convergence of the kriging metamodel training. In (Bichon et al., 2008; Echard et al., 2011), the Expected Feasibility Function (EFF) has been employed to quantify the *balance* trend between the search in the *vicinity* of the limit state and a more *global* search in the input space: when the maximum value of the EEF over the entire search space falls below a given threshold (e.g., EFF < 0.001 in (Bichon et al., 2008)), the algorithm is stopped. In (Echard et al., 2011) the the U-learning function (1) is introduced to increase the accuracy and precision of the kriging metamodel preferably *in proximity of the limit state*. The *smaller* is $U(\boldsymbol{x})$, the closer is the metamodel prediction to the failure threshold and the higher the interest in adding the observation corresponding to $\boldsymbol{x}$ to the current training set. In this view, when the minimum value of $U(\boldsymbol{x})$ over the search space exceeds a given threshold (e.g., $U(\boldsymbol{x}) > 2$ in (Cox and John, 1997; Echard et al., 2011)), the iterations are stopped. Since in some applications the EEF and the U-learning function (1) are found to exhibit a slow convergence to the critical region (Echard et al., 2011; Dubourg et al., 2013), other works of literature employ quantitative metrics based on a *cross-validation procedure* to check *both* metamodel accuracy *and* refinement convergence. Basically, the DoE $\{\mathscr{X}, \mathscr{Y}\}$ is split into a training subset $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}$ and a validation subset $\{\mathscr{X}_{val}, \mathscr{Y}_{val}\}$ such that $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\} \cap \{\mathscr{X}_{val}, \mathscr{Y}_{val}\} = \varnothing$ and $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\} \cup \{\mathscr{X}_{val}, \mathscr{Y}_{val}\} = \{\mathscr{X}, \mathscr{Y}\}$. The metamodel is then built using the training subset and validated by comparing the predicted values and the real values onto the validation subset. The *leave-one-out* (LOO) technique is a special case where the training subset is defined as $\{\mathscr{X}, \mathscr{Y}\}\backslash \boldsymbol{x}_i$. (Allen, 1971) uses an LOO assessment of the mean squared error referred to as the Predicted REsidual Sum of Squares (PRESS), whereas (Dubourg et al., 2013; Turati et al., 2017) employ an LOO-based correction factor to account for epistemic uncertainty in the failure probability estimates provided by kriging and to correspondingly check its accuracy and training convergence. Based on this critical review of literature, we propose an approach relying on *two* ("local" and "global") *validation sets*, which is *empirically* found to provide a satisfactory trade-off between (high) metamodel *accuracy*, (high) *speed of convergence* to the PSS critical region and (low) *computational cost* (i.e., number of calls to the long-running T-H simulation code). Details are provided in the following.

Accuracy is evaluated with respect to the N_{val}combinations of a *validation set* ($\mathscr{X}_{val}$): $\widehat{\mathscr{Y}}_{val} = \left(\widehat{y}_1, \cdots, \widehat{y}_{N_{val}}\right)$ different from the training set. The predicted output values $\widehat{\mathscr{Y}}_{val}$ are compared to the corresponding RELAP5-3D output values $\mathscr{Y}_{val}$ through the construction of some Quality Indicators (QIs). No definitive guidelines are found in literature about the correct size $N_{val}$of the validation set. (Martin and Simpson, 2005) suggests $N_{val} \gg N_{train}$, since a small N_{val}can be misleading in case validation points are taken, by chance, too close to training points where the metamodel is clearly more refined (Wu et al., 2018). However, this approach becomes extremely expensive in case of time-demanding simulators. (Iooss, 2009) proposes a "sequential validation design" to get to a meaningful validation, while keeping N_{val} small: validation is carried out gradually by adding validation samples in the unfilled regions of the input space to optimize the distance between the validation set and the training set. Here two validation sets are considered. The first one derived from 55 simulated transients, with output $E_{ex,\%} = 85 \div 95\%$, and it is used to verify the metamodel accuracy around the limit surface. The second one includes 138 I/O simulated transients, with $E_{ex,\%}$ values spreading all over the domain and it is employed to obtain an indication of the metamodel accuracy over the entire domain. The QIs used to quantify metamodel accuracy with respect to both validation sets are the well-known RMSE and two different predictivity indicators, Q_1, defined in (Iooss, 2009), and Q_2 presented by (Lataniotis et al., 2019):

$$RMSE = \sqrt{\sum_{i=1}^{Nval} \frac{\left(\widehat{y}_i - y_i\right)^2}{N_{val}}} \tag{2}$$

$$Q_1 = 1 - \frac{\sum_{i=1}^{Nval}\left(\widehat{y}_i - y_i\right)^2}{\sum_{i=1}^{Nval}\left(\overline{y}_{val} - y_i\right)^2} \tag{3}$$

$$Q_2 = \frac{N_{val} - 1}{N_{val}}\left(\frac{\sum_{i=1}^{Nval}\left(\widehat{y}_i - y_i\right)^2}{\sum_{i=1}^{Nval}\left(\overline{y}_{val} - y_i\right)^2}\right), \tag{4}$$

where y_i is the i-th output of $\{\mathscr{X}_{val}, \mathscr{Y}_{val}\}$, $\widehat{y}_i$is the corresponding metamodel prediction and $\overline{y}_{val}$ is the mean value of all the simulations output values in the validation set. RMSE and Q_2 should be as low as

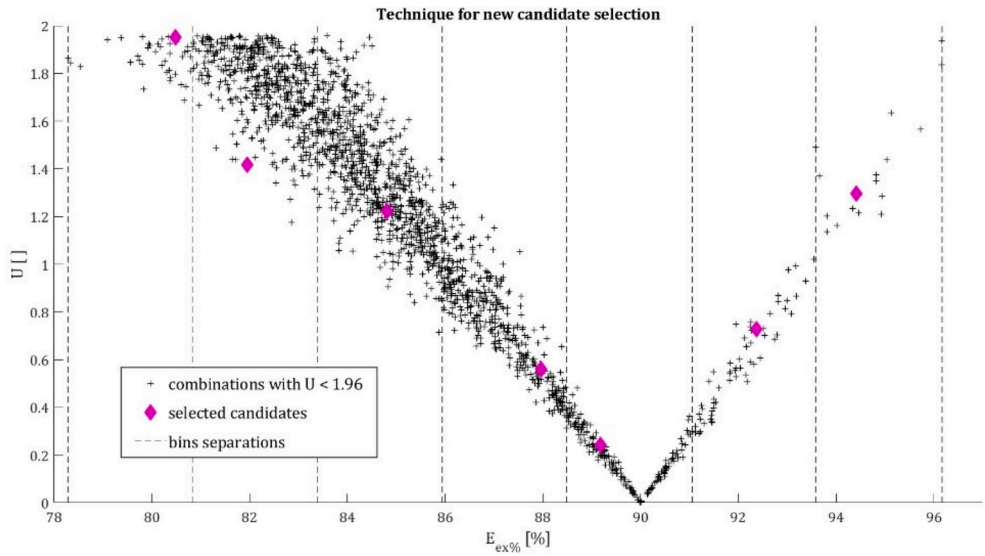

Fig. 2. Identification of new candidates (Step 5): random selection of combinations with U < 1.96 divided in equally-spaced bins.

possible, whereas Q_1 tends to 1 as the prediction accuracy increases. Notice that the RMSE has the same unit of measure of the physical quantity of interest ($E_{ex,\%}$) and, thus, it can be progressively compared to the exchanged energy output to understand whether the predictions are satisfactory. It can be also normalized (NRMSE) dividing it by $\bar{y}_{val}$. Q_1 and Q_2 have similar expressions and, differently from the RMSE, they take into account also the variability of the output in the validation set. The values of these QIs should be improved as much as possible through the successive iterations of the algorithm. In this work, the convergence (stopping) criterion related to the metamodel accuracy is considered satisfied when the NRMSE evaluated on the "local" validation set constructed around Y_{thres} becomes about 2%. On the other hand, if the metamodel quality is still unsatisfactory, the second convergence (stopping) criterion needs to be checked: further computational time is required to add a new algorithm iteration and to simulate new configurations by the BE-TH code in order to enrich the training set. The computational budget, i.e., the maximum number of simulations initially foreseen, has been here fixed to 100 RELAP5-3D simulations in addition to the initial ones (i.e., the 64 simulations belonging to $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}_{in}$). When the computational budget is completely run out, even if Kriging accuracy is still not satisfactory, the procedure stops. Notice that the number (100) of RELAP5-3D simulations to be added to the (64) initial ones has *not* been defined according to some *theoretical criterion*: this number is strongly problem- and model-dependent and must be obviously set by the analyst according to the computational time and power available.

5. *Selection:* if algorithm convergence has not been reached at step 4, new I/O simulations related to the so-called best candidates $\mathscr{X}^*$ are run and the input and output values added to the training set to refine the metamodel. The N_{cand} best candidates are selected among the $\mathscr{X}$ generated at step 2 according to their U-function values. Combinations with $U < 1.96$ are sorted in ascending order according to their predicted output value $\hat{y}$and, then, organized in N_{cand} equally-spaced *bins*. Then, one candidate is randomly picked from *each bin*. This procedure is applied to avoid selecting candidates "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 (see Section 3); hence, selecting the candidates only according to the N_{cand}lowest U values would cause them to be restricted in the same area of the domain, instead of spanning the whole input space. The selection procedure is illustrated in Fig. 2.

Table 3
AK-MCS results for $\mathbf{E_{ex}}$ output.

n	N_{train}	$\bar{\sigma}_1$	$\bar{\sigma}_2$	ϵLOO_{norm}	ϵLOO_{abs}
0	64	11.76	11.67	0.128	20.62
1	71	11.09	10.99	0.130	19.39
2	78	9.41	9.38	0.138	19.05
3	85	8.83	8.80	0.148	18.76
4	92	8.39	8.35	0.152	18.36
5	99	7.10	7.06	0.159	18.24
6	106	7.60	7.55	0.169	18.18
7	113	7.50	7.46	0.169	17.36
8	121	7.35	7.30	0.167	16.37
9	129	7.41	7.36	0.166	15.61
10	136	7.23	7.18	0.176	16.61
11	143	7.01	6.96	0.175	15.98
12	150	4.56	4.21	0.177	15.66
13	157	4.61	4.10	0.126	10.81
14	164	4.47	3.92	0.119	9.69

$N_{cand} = 7$ or 8 points have been chosen, as a satisfactory trade-off between computational cost, number of iterations of the algorithm and metamodel accuracy. Indeed, lower values of N_{cand} would require a larger number of algorithm iterations and training repetitions (i.e., higher computational cost) to obtain the same Kriging accuracy; also, an excessively small number of candidates implies a rougher exploration and "mapping" of the area close to the limit surface. On the other hand, limiting the number N_{cand} is useful, in particular in the first iterations when the metamodel is still inaccurate. Selecting many candidates according to its predictions may lead to a waste of computational time: actually, some candidates, simulated with the expensive BE-TH model, may later reveal to be not so useful for the scope of the analysis (e.g., they may lie far from the limit surface). In Fig. 2, the value $E_{ex,\%}$ is reported on the x-axis, whereas y-axis displays the corresponding U-function values. It is clear from the dashed vertical lines how the bins are constructed by subdividing the x-axis in segments of the same length. $\mathscr{Y}^*$ values are represented by diamonds, whereas all the samples with $U < 1.96$ are shown as crosses. The shape of the graph shows that the closer is a point to $Y_{thres} = 90\%$, the lower its U value is; this was easily foreseeable looking at equation (1).

Once the best candidates $\mathscr{X}^*$ have been selected and the corresponding I/O transients simulated with the BE-TH model to obtain the output $\mathscr{Y}^*$, the training set is enriched and steps 1 to 5 are repeated until convergence at step 4 is reached.

5. Results

5.1. Metamodel accuracy evaluation

The AK-MCS procedure has been stopped at iteration $n_{fin} = 14$, after enriching the initial training set $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}_{in}$ with the input and output values of 100 RELAP5-3D (i.e., maximum computational budget available). In Table 3 the salient aspects of each n-th iteration are reported.

In the 3rd and 4th columns of Table 3, two average standard deviations ($\overline{\sigma}$) are reported; they are calculated with respect to different metamodel outputs $\widehat{\mathscr{Y}}$: $\overline{\sigma}_1$ in column 3 is evaluated with respect to the outputs of all the combinations generated at step2 of the AK-MCS procedure, which are spread throughout the domain; instead, $\overline{\sigma}_2$ in column 4 is calculated using only the outputs of those combinations for which $U < 1.96$. Both the measures have been used in the successive iterations of the algorithm in order to check the Kriging gradual refinement, without resorting to the computationally expensive validation set. In particular, $\overline{\sigma}_1$ allows following the improvement in the metamodel accuracy over the entire domain, whereas $\overline{\sigma}_2$ is used to focus on the accuracy increase nearby the limit surface. As expected, the metamodel general improvement during the iterations makes both standard deviations decrease; however, $\overline{\sigma}_2$ diminishes more rapidly due to the nature of the AK-MCS algorithm, which adds new I/O data with outputs close to Y_{thres} (where U is lower), thus making the predictions more accurate in proximity of the limit surface than elsewhere in the domain (e.g., at iteration $n_{fin} = 14$, $\overline{\sigma}_1$ is equal to 4.47, whereas $\overline{\sigma}_2$ is equal to 3.92). The Kriging settings (Section 4.1) have been adjusted from iteration 12 onwards by changing the correlation function family from *Matérn 5/2* to *Exponential*, in order to improve the fit with the new, expanded training set; indeed, looking at the evolution of the two average standard deviations $\overline{\sigma}_1$ and $\overline{\sigma}_2$ up to that point, it can be noticed that the corresponding values were not decreasing anymore and the metamodel improvement seemed stuck.

The last two columns of Table 3 report the Leave-One-Out (LOO) error evolution with iterations: column 5 shows the LOO error directly returned by the UQLab tool, also called normalized LOO error ($\in LOO_{norm}$), whereas column 6 reports its absolute version ($\in LOO_{abs}$):

$$\in LOO_{norm} = \frac{1}{N_{train}}\left[\frac{\sum_{i=1}^{N_{train}}\left(y(\boldsymbol{x}_i) - \widehat{y}_{(-i)}(\boldsymbol{x}_i)\right)^2}{Var[\mathscr{Y}_{train}]}\right] \text{ and } \in LOO_{abs} = \in LOO_{norm} \cdot Var[\mathscr{Y}_{train}], \tag{5}$$

where $\widehat{y}_{(-i)}(\boldsymbol{x}_i)$ is the prediction made by the metamodel in correspondence of the i-th combination $\boldsymbol{x}_i \in \mathscr{X}_{train}$ and obtained using all the $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}$ pairs of values available, except $\{x_i, y_i\}$, and $Var[\mathscr{Y}_{train}]$ is the variance of the training outputs. The only difference is in the term $Var[\mathscr{Y}_{train}]$ representing the output variability in the training set. The LOO error is generally used (as the $\overline{\sigma}$'s previously introduced) to assess model accuracy when there is no availability of an external validation set due to its high computational cost; thus, the evolution of these two quantities has been followed since it gives an idea about the Kriging progressive refinement. A gradual decrease was expected, but what occurs in reality is that $\in LOO_{norm}$ initially rises and, then, sharply drops reaching its lowest value at $n_{fin} = 14$. This behaviour is justified by equation (5): the metamodel becomes progressively more refined, causing the numerator of $\in LOO_{norm}$ to decrease; however, at the same time, also $Var[\mathscr{Y}_{train}]$ at the denominator diminishes, because the I/O data are all selected with outputs close to Y_{thres} and hence the variability of $\mathscr{Y}_{train}$ reduces. On the other hand, $\in LOO_{abs}$, not containing the $Var[\mathscr{Y}_{train}]$ term, shows a more regular (decreasing) trend.

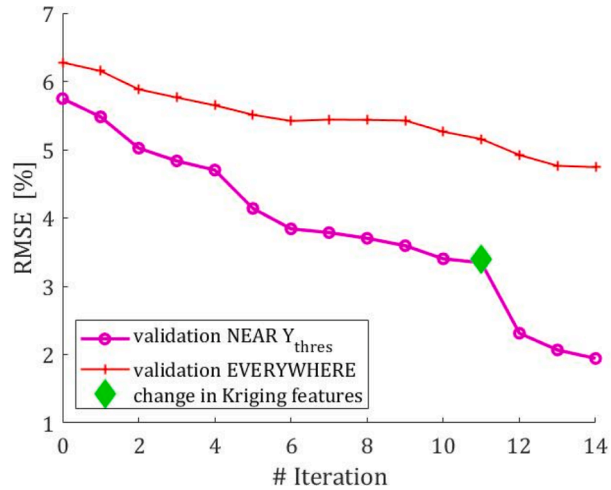

Fig. 3. **RMSE** evaluated with respect to 2 validation sets.

The best way to follow the Kriging accuracy improvement with iterations, if enough computational power is available, is to construct an external validation set and to evaluate the corresponding QIs by computing the Kriging predictions with respect to the validation data (as explained at step 4 of the AK-MCS procedure). For this case study, two validation sets have been constructed (see Section 4.2) and three QIs have been considered: RMSE, Q_1 and Q_2 (see equations (2), (3) and (4)). The QIs evolution is illustrated in Figs. 3 and 4. Note that the RMSE in Fig. 3 is expressed in percentage because it has the same unit of measure of the predicted output, i.e., the percentage energy exchanged ($E_{ex,\%}$); but, it should not be confused with the NRMSE.

The two curves in each plot (Figs. 3 and 4(a) and (b)) are referred to different validation sets, but they all show the same general trend: a decrease in RMSE and Q_2, and an increase towards 1 for Q_1, representing the improvement of the metamodel accuracy. The three curves associated to the validation set constructed near Y_{thres} display a faster improvement in accuracy because of the nature of the metamodel-based AK-MCS procedure, which gradually makes the metamodel more refined around the failure threshold. The diamond in correspondence of the 12th iteration symbolizes the change of Kriging setting. The two curves in Fig. 4(a) and (b) cross because of Q_1 and Q_2 mathematical expressions. For example, for what concerns Q_2, being the metamodel at the beginning still inaccurate, the numerator in equation (4) is small for both (local and global) validation sets; on the contrary, the denominator is obviously larger for the global validation set, with data spread all over the domain, than the local validation set. This is the reason why at the beginning the Q_2 value is lower (and hence better) when evaluated with respect to the most various validation set (curve with crosses in Fig. 4 (b)) than with the local validation set (curve with circles in Fig. 4(b)), unlike what is observed at the end of the AK-MCS procedure. The same behavior with respect to the two validation sets is observed also for Q_1, but in the opposite way (see Eq. (3)).

Table 4 reports all the QI values at the last iteration, with the RMSEs computed in both the absolute and normalized forms:

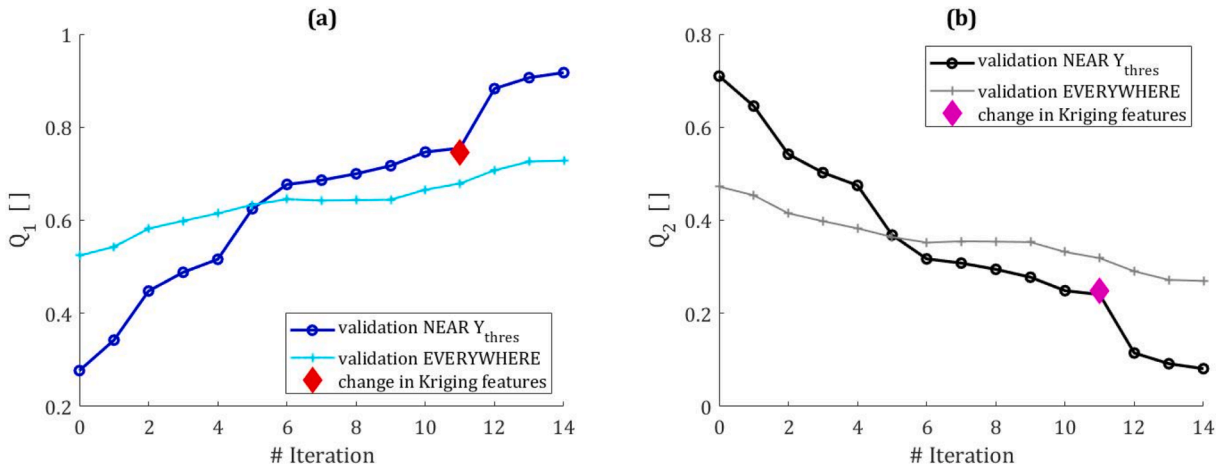

Fig. 4. $\mathbf{Q}_1$ (a) and $\mathbf{Q}_2$(b) predictivity indicators evaluated with respect to the two validation sets.

Table 4
QIs at the end of AK-MCS procedure (14th iteration).

Quality indicator (QI)	RMSE	NRMSE	Q_1	Q_2
Validation **EVERYWHERE**	4.74	5.95%	0.728	0.270
Validation **NEAR**Y_{thres}	1.94	2.24%	0.917	0.081

The final results are generally satisfying: a RMSE of 1.94 is acceptable if compared to $E_{ex,\%}$, which usually varies from 70% to 100% in the simulated transients. A value of 2.24% for the NRMSE is remarkable since it could be taken, at first instance, as a measure of the percentage error in prediction: an error around 2% near the limit surface is considered satisfactory. Actually, as highlighted at Step 4. of the algorithm in Section 4.2, the convergence (stopping) criterion related to the metamodel accuracy is considered satisfied when the NRMSE evaluated on the "local" validation set constructed around Y_{thres} becomes about 2%. Thus, in this case there is no need to include additional RELAP5-3D simulation points, since the associated (relevant) increase in the computational cost would *not* justify the (slight) improvement in the "local" metamodel accuracy (see the curve with circles in Fig. 3, reaching a sort of plateau).

5.2. Critical failure regions characterization

The Kriging metamodel obtained at the end of the AK-MCS iterations has been exploited to predict the outcomes of a large number (10.000) of new input combinations $\boldsymbol{x}$ in order to: (i) find the critical ones, with reference to the PSS function, i.e. $\hat{y} = f(\boldsymbol{x}) \leq Y_{thres}$; and (ii) retrieve information about the shape of the CRs of the PSS operation. Given that the input space dimensionality is $M = 5$, a graphical representation of the unique CR identified is provided through a series of scatter plots with paired inputs representing the CR two-dimensional projections (Fig. 5): green diamonds are used to indicate combinations leading to a safe operation (i.e., $E_{ex,\%} \geq 90\%$), whereas red crosses represent critical input values when the PSS fails its function.

The subplots in Fig. 5 allow catching the influence of the input parameters on the amount of energy exchanged (E_{ex}) by the PSS; in particular, each scatter plot provides information about the effect of the interaction between the two input parameters therein represented (whatever the values of the other parameters). The results show that only four of the five input parameters have significant influence on E_{ex}. Indeed, A_{AV} is not very relevant for driving the PSS response in terms of the energy exchanged: in fact, whatever its value, the DHR function may or may not be successfully accomplished (green diamonds or red crosses, indifferently). The interaction of *NC*% with the other input parameters (except A_{AV}) is shown in Fig. 5(b), (c) and (d). In particular, in all the subplots the maximum *NC*%value corresponding to a combination of functional success is around 30%, suggesting that PSS fails to provide its function whenever *NC*% > 30%, independently of the values of the other parameters. This is coherent with the underlying physics: the presence of non-condensable gases leads to a reduction in the heat transfer coefficient during condensation in the E-HX and, in fact, the higher *NC*%, the worse the impact on E_{ex}. According to Fig. 5(b), (c) and (d), the upper limit for *NC*% is generally reduced in case of interactions with other parameters that reduce the E_{ex} value, i.e., variations of DEL_{AV}, $DEL_{MSIV}$and A_{MSIV} from their reference values. The results are represented by triangle-shaped safe region (green diamonds). For example, a value $A_{MSIV} = 0.035\%$ represents a leakage in the MSIV that reduces the amount of steam directed into the PSS, thus lowering E_{ex}: in this situation, the maximum value of *NC*%, for which the PSS function can still be successfully accomplished, is about *NC*% = 15% (whatever the values taken by the other three parameters). Also, DEL_{AV} plays a central role in E_{ex} determination. Indeed, if AV opens with a certain delay, i.e., $DEL_{AV} > 0$, the whole heat transfer process is delayed and this impacts severely on E_{ex}, especially because the largest amount of energy is exchanged in the first part of the accidental transient. Moreover, if AV is not open, the PV pressure may rise, which causes the opening of the SRV and, hence, vapor discharging into the containment instead of condensing inside the PSS. Looking at subplots 5(b), (e), (h) and (i), the maximum DEL_{AV} for which the PSS function can still be accomplished is about 400 s: however, this upper limit is, in general, lowered in case of interactions with other parameters (as for input *NC*%). For example, when varied together with A_{MSIV}(see subplot 5(h)), again a triangle-shaped region of safe function is observed: e.g., if $A_{MSIV} = 0.025\%$, the maximum DEL_{AV} value for successful function is about 200 sec. For what concerns DEL_{MSIV}, the observed upper limit is about 4000 s. The MSIV closure is necessary to isolate the turbine side and start sending the vapor into the PSS for condensation; hence, if the closure is delayed, less vapor enters the PSS in the first part of the transient and E_{ex} is reduced. Whereas a priori the expected interaction of DEL_{MSIV} with other input parameters negatively affecting E_{ex} could have been a decrease in the value of DEL_{MSIV} leading to the PSS functional failure. Instead, what is observed from the simulations in case of interaction with, e.g., DEL_{AV} or A_{MSIV} is different (see subplots 5(i) and (j)). In such cases, the safe region is square-shaped. For example, focusing on the interaction between DEL_{MSIV} and DEL_{AV}, the upper limit of DEL_{MSIV} should, in general, decrease if $DEL_{AV} > 0s$, independently of the values of the other three parameters; instead, the upper limit remains about 4000 s (except in the extreme case where DEL_{AV} reaches its own upper limit causing PSS functional failure by itself). This behaviour (and the consequent square-shaped regions) is probably due to the influence of DEL_{MSIV} on E_{ex}, which is more significant than that of other input parameters. For example,

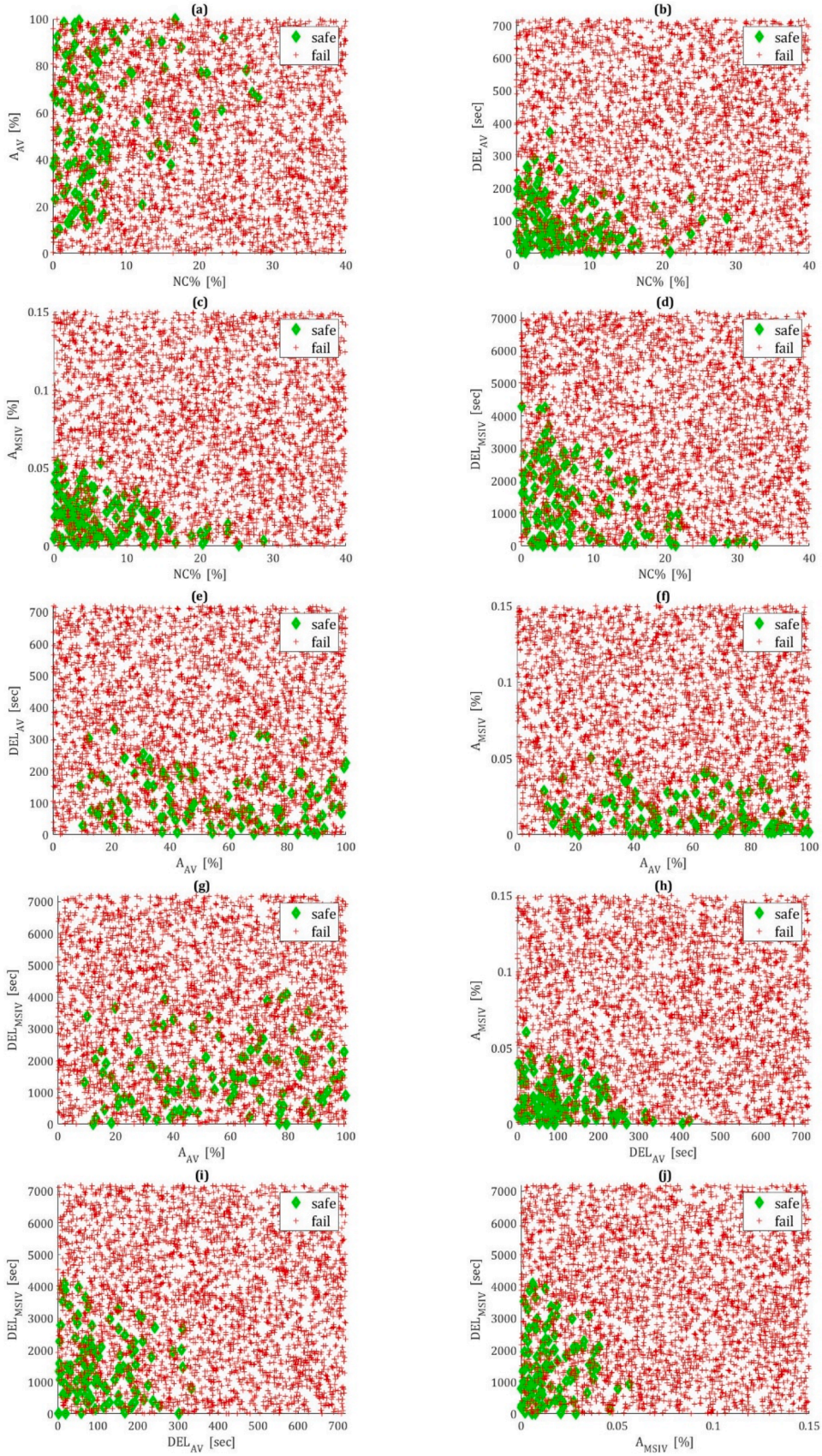

Fig. 5. Scatter plots of the PSS CR, obtained by AK-MCS with 164 RELAP5-3D simulations and 10,000 kriging metamodel evaluations.

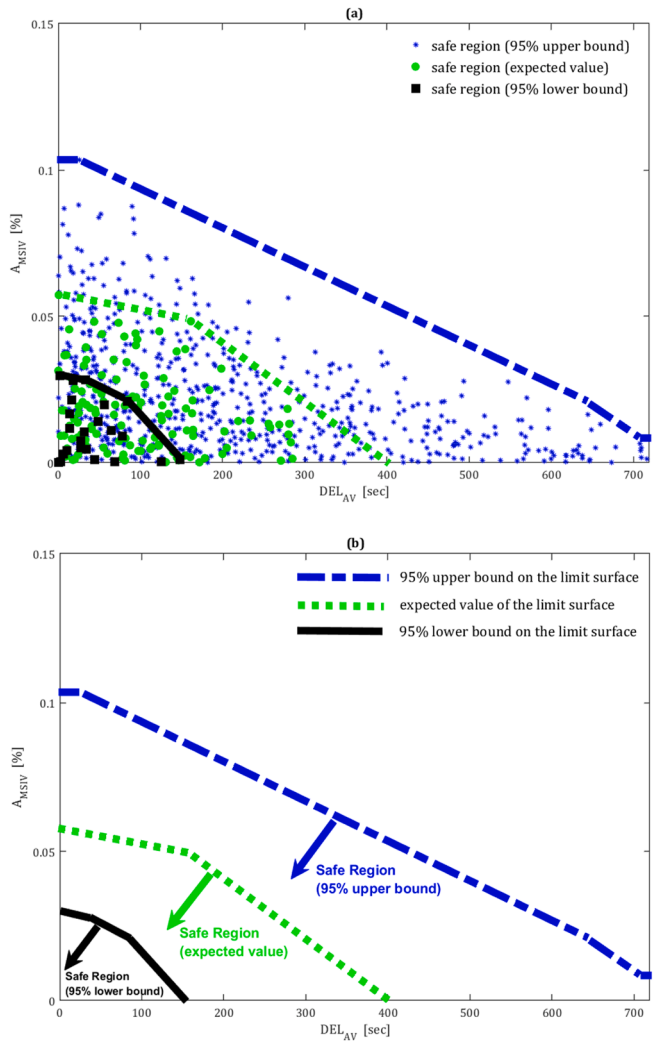

Fig. 6. a) two-dimensional projection (in the plane DEL_{AV}-A_{MSIV}) of the safe region (green diamonds) together with the corresponding lower (black squares) and upper (blue asterisks) 95% confidence bounds, resulting from 10,000 kriging metamodel evaluations; b) convex hulls enveloping the safe configurations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

considering the interaction between DEL_{MSIV} and A_{MSIV}, the possible presence of a certain leakage in the MSIV after its closure (i.e., $A_{MSIV} > 0\%$) is less relevant in terms of contribution to the amount of energy exchanged, if it occurs in case of a significant delay in the MISV closure, which is much more influential[2].

Relevant insights about the safe and failure regions can be also drawn by exploiting the intrinsic properties of the kriging metamodel. As highlighted in Appendix A, one of the main advantages of this technique is that – due to the Gaussian assumption – a standard deviation $\sigma^2_{\widehat{y}}(\boldsymbol{x})$ (and, thus, a Confidence Interval-CI) can be associated to each prediction $\widehat{y}(\boldsymbol{x})$: hence, it can be exploited for assessing the accuracy and

[2] A consideration is in order with respect to the scatterplots reported in Fig. 5. On one side, it is clear there is a *well-defined* and *limited* safe region (green diamonds) in each bi-variate comparison; on the other side, it may seem that the distribution of failed states (red crosses) exists across the *entire domain* of all the uncertain variables (actually, in each subplot many safe and failed configurations appear to *overlap*). However, notice that this is only a visualization issue, due to the fact that each subplot represents a *two-dimensional projection* of the *five-dimensional* safe and failed regions.

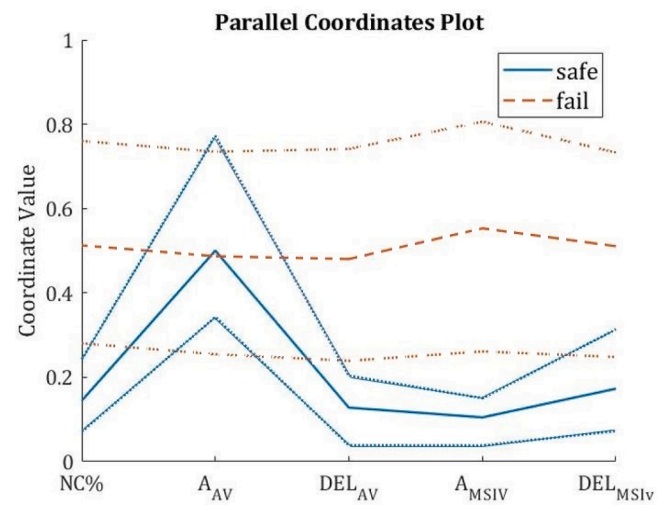

Fig. 7. PCP with 0.25 quantiles (quartiles), obtained by AK-MCS with 164 RELAP5-3D simulations and 10,000 kriging metamodel evaluations.

precision of the metamodel in predicting a new PSS configuration. The lower and the upper bounds of the (α·100)% CI for prediction $\widehat{y}(\boldsymbol{x})$ read $\widehat{y}(\boldsymbol{x}) - k(\alpha)\cdot\sigma_{\widehat{y}}(\boldsymbol{x})$ and $\widehat{y}(\boldsymbol{x}) + k(\alpha)\cdot\sigma_{\widehat{y}}(\boldsymbol{x})$, respectively: $k(\alpha)$ sets the confidence level as $k(\alpha) = \Phi^{-1}[(1+\alpha)/2]$, where $\Phi^{-1}[\cdot]$ is the inverse cumulative distribution function of the standard Normal distribution. Analogously, we can identify a "confidence interval" on the safe and critical failure regions. For example, the lower and upper (α·100)% confidence bounds on the safe region are defined as $\left\{\boldsymbol{x} : \widehat{y}(\boldsymbol{x}) - k(\alpha)\cdot\sigma_{\widehat{y}}(\boldsymbol{x}) \geq Y_{thres}\right\} = \left\{\boldsymbol{x} : \widehat{E_{ex,\%}}(\boldsymbol{x}) - k(\alpha)\cdot\sigma_{\widehat{E_{ex,\%}}}(\boldsymbol{x}) \geq E^{thres}_{ex,\%}\right\}$ and $\left\{\boldsymbol{x} : \widehat{y}(\boldsymbol{x}) + k(\alpha)\cdot\sigma_{\widehat{y}}(\boldsymbol{x}) \geq Y_{thres}\right\} = \left\{\boldsymbol{x} : \widehat{E_{ex,\%}}(\boldsymbol{x}) + k(\alpha)\cdot\sigma_{\widehat{E_{ex,\%}}}(\boldsymbol{x}) \geq E^{thres}_{ex,\%}\right\}$, respectively. By way of example and only for illustration purposes, Fig. 6(a) shows the two-dimensional projection (in the plane DEL_{AV}-A_{MSIV}) of the safe region (green diamonds) together with the corresponding lower (black squares) and upper (blue asterisks) 95% confidence bounds, resulting from 10,000 kriging evaluations; as a guide to the eye, the convex hulls enveloping the safe configurations sampled are also represented as green dotted, black solid and blue dot-dashed lines, respectively. In Fig. 6(b) only the convex hulls are reported for the sake of clarity. It is worth noting that the lower and upper bounds can be interpreted as the least and most conservative estimates of the safe domain, respectively. In other words, we are able to quantify a sort of licensing-defined (α·100)% tolerance (or (α·100)% CI limit) on the input parameter uncertain values that still result in a safe PSS operation.

Other interesting conclusions about the CR can be inferred from the PCP (Fig. 7). PCP allows displaying in a unique plot all the five input values corresponding to each combination; each of the five vertical axes reports the values of one input parameter (normalized on its range) and, hence, one input combination is represented by a line in the horizontal direction connecting the corresponding input values on the different axes.

In particular, Fig. 7 shows the quantile representation of all the predicted combinations together: the solid blue lines are representative of the PSS safe operation, whereas the dashed orange lines represent the input combinations leading the PSS to fail its function. In particular, in both cases (solid or dashed lines) the line in the middle stands for an "average" combination (average safe combination or average failure combination), whereas the other two external lines are its 0.25 upper and lower quantiles (also called quartiles).

The contribution of A_{AV} to safe function immediately stands out: the normalized value associated to this curve is about the same as the one observed for functional failure. This confirms the scarce importance of

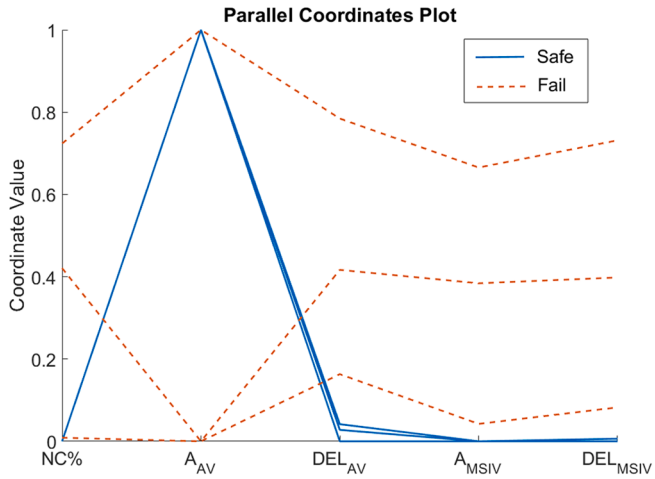

Fig. 8. PCP with 0.25 quantiles (quartiles), revealed by the initial set $\{\mathscr{X}_{\mathbf{train}}, \mathscr{Y}_{\mathbf{train}}\}_{\mathbf{in}}$ of 64 RELAP5-3D simulations.

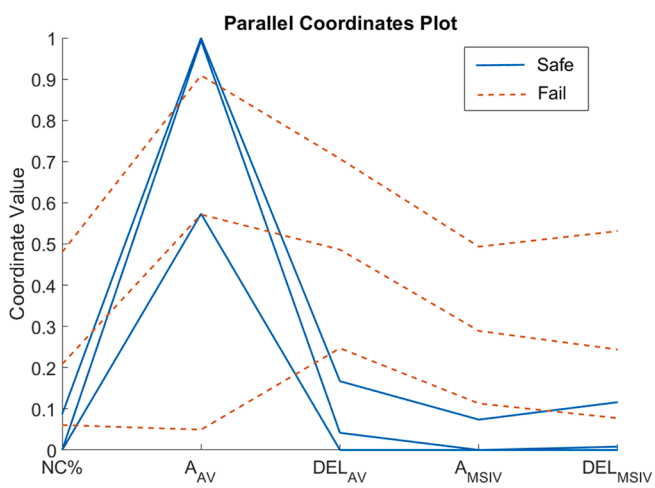

Fig. 9. PCP with 0.25 quantiles (quartiles), revealed by the final set of 164 RELAP5-3D samples (generated by the adaptive AK-MCS search).

A_{AV} in the PSS function, as already seen from Fig. 5, but in a more quantitatively way. Moreover, the solid blue lines are mainly located in the lower part of the graph, close to the 10% of the range of variation of each input (except for A_{AV}, as previously mentioned); on the contrary, the dashed orange lines occupy quite a large portion (in particular, the middle part) of these intervals, meaning that the majority of the input combinations leads to PSS functional failure. The same result can be deduced from Fig. 5, where the red crosses are far more abundant than the green diamonds. A word of caution is in order in this respect. This result does not mean that the PSS analysed is more prone to fail than to succeed, since this type of conclusion should be supported by a *probabilistic analysis* of the occurrence of the input combinations, which is not carried out in this work. Actually, the probability estimated for the event that the PSS fails its function strongly depends on: (i) the characteristics of the system itself, and (ii) the (data and/or expert-based) probability distributions of the PSS input parameters. In this paper, as mentioned in the Introduction and Section 2.1, no realistic probability distributions are assigned to the PSS parameters, since the objective is not to carry out a reliability analysis of the PSS, but to describe how the metamodel-based AK-MCS procedure can be exploited for critical regions characterization.

For illustration purposes, the safe and critical regions identified by the AK-MCS method with 10,000 kriging metamodel evaluations (Fig. 7) are compared to those revealed by: (i) the initial set $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}_{in}$ of 64 RELAP5-3D simulations, which serves as a reference baseline result (Fig. 8); (ii) the final set of 164 RELAP5-3D samples (generated by the adaptive AK-MCS search), as a point of comparison to the additional refinements provided by the extensive sampling capability of the kriging metamodel (Fig. 9). Comparing the PCP of Figs. 7 and 8 the following considerations can be done. Using only the initial 64 RELAP5-3D simulations (Fig. 8), the solid blue lines (that represent the input combinations leading the PSS to safe operation) overlap almost completely (i.e., the "average" safe combination of inputs coincides with the 0.25 upper and lower quantiles). This means that the initial 64 samples are *not* sufficient to thoroughly explore the state space of the PSS and, thus, they cannot properly identify the safe (resp., critical) *region* and cannot precisely characterize its *boundaries* and *width*. Rather, only few sparse safe configurations (in this case, 27) are identified, most of them lying on the bounds of the uncertainty ranges of the input variables: in fact, all the solid blue lines in Fig. 8 are located in the lower (resp., upper) part of the graph, very close to the 0% (resp., 100%) of the range of variation of each input. This is a consequence of the "deterministic" (expert-based) selection of *some* of the 64 initial input configurations, as detailed in Section 4.1; instead, most of the other configurations (randomly selected by LHS) fall in the failure region. In addition, it can be noticed that the contribution of A_{AV} to safe function suggested by Fig. 8 is completely *different* from that resulting in Fig. 7. In particular, the normalized value associated to this (solid blue) curve is *much larger* than the (orange dashed) one observed for functional failure. This would imply a significant importance of A_{AV} in the PSS function, which contradicts the results of the detailed exploration carried out by the adaptively trained kriging metamodel (Fig. 7). Finally, it is worth noting that for some input variables (in particular, *NC*%, A_{AV} and DEL_{AV}) the distance between the 0.25 upper and lower quantiles of the critical failure ranges (orange dashed lines) is larger in Fig. 8 than in Fig. 7: this means that in the present case, relying only on 64 RELAP5-3D simulations (selected in a mixed deterministic and probabilistic fashion) leads to an *overestimation* of the size of the critical failure region. These considerations and results call for a deeper and more systematic exploration of the PSS state space, to precisely discriminate between safe and failed configurations and to provide robust information to designers and operators.

In this view, the PCP constructed on the final set of 164 RELAP5-3D samples (Fig. 9) provide *more reliable* indications. For example, the location of the safe regions for input variables *NC*%, DEL_{AV}, A_{MSIV} and DEL_{MSIV} (blue solid lines) is *similar* to the reference one represented in Fig. 7 (i.e., close to the 0–10% of the range of variation). This is obviously due to the fact that the additional 100 RELAP5-3D samples are a result of the metamodel-aided intelligent adaptive search carried out by the AK-MCS methodology itself. However, it is also worth noting that the size of the corresponding safe regions (i.e., the distance between the 0.25 upper and lower quantiles) is *underestimated*. In addition, the importance of A_{AV} in the PSS function is still overestimated with respect to the results of the detailed exploration carried out by the adaptively trained kriging metamodel (Fig. 7): actually, the normalized value associated to the corresponding "average" (solid blue) curve is still *larger* than the (orange dashed) one observed for functional failure. Finally, also the location and size of the critical failure region of some input variables (in particular, *NC* %, A_{AV}, A_{MSIV} and DEL_{MSIV}) are different from the reference ones (Fig. 7). For example, the orange dashed lines of *NC*%, A_{MSIV} and DEL_{MSIV} are in the lower part of the corresponding range of variation (instead of the middle part as reported in Fig. 7) and the corresponding distance between the 0.25 upper and lower quantiles is slightly underestimated. In conclusion, the comparison between these three levels of parameter space exploration highlights the need for an accurate and thorough, yet computationally feasible, analysis of the PSS safe and failed configurations, which strengthens the argument for using a methodology like AK-MCS for PSS design and operation and possibly licensing applications.

As a conclusive remark, notice that thanks to the use of the proposed metamodel-based technique, the computational cost associated to the analysis above has been reduced by 3–4 *orders of magnitude*.

6. Conclusions

The adoption of PSS is a promising way to increase the safety of an NPP. However, the operational experience with PSS is lower than with active systems: thus, a more thorough analysis of the possibly wide range of PSS operating conditions and a detailed exploration of the critical combinations that may lead to PSS functional failures are mandatory. A structured procedure for the reliability assessment of T-H PSSs (namely, REPAS) has been developed in the past (Jafari et al., 2003; Pierro et al., 2009): however, it does not explicitly include an important step, i.e., the detailed characterization of the PSS Critical Regions (CRs), which is necessary to identify the combinations of critical operation of the system (i.e., those combinations leading the PSS to fail providing its function). The identification of the states that lead a PSS to safety–critical conditions provides relevant insights for improving system safety, since it allows prevention and preparation. In particular, in the *design* phase, such information may be used, e.g., to implement proper modifications aimed at "reducing the size" of the critical regions to the extent possible. In the *operational* phase, a detailed "mapping" of the PSS state-space allows in principle to characterize and classify, in a *timely* manner, new (developing) scenarios as 'safe' or 'faulty'. In this view, the CR characterization process can serve as a *basis* for the identification of *critical system components* that are more likely to lead the PSS into functional failure and for the consequent *prioritization* of *inspection/maintenance actions* on such relevant components. To these aims, computational models and simulators are frequently employed for studying the PSS response under different conditions. This is challenging because the simulation models are *black-box*, *dynamic* and *computationally demanding*.

To tackle these issues, we have applied an intelligent exploration framework, namely the metamodel-based AK-MCS, combining a fast-running Kriging metamodel and adaptive sampling (Echard et al., 2011; Turati et al., 2017). The AK-MCS procedure automatically refines the Kriging in proximity of the PSS limit surface to predict – at an acceptable computational cost – those physical input combinations leading the PSS to functional failure (and hence to trace the boundary between safe and failed behavior). To the best authors' knowledge this is the first time that the AK-MCS technique is applied for the identification of the CRs of a PSS of an NPP.

The proposed methodology has been applied to a generic PSS, the DHR, to analyze the event of "Low heat removal" (i.e., energy exchanged $E_{ex,\%} < 90\%$). The metamodel-based AK-MCS has been shown capable of accurately identifying the combinations leading to functional failure by resorting to a limited number (i.e., few hundreds) of computationally expensive BE TH code runs. The CR and safe region have been visualized by scatter plots and PCPs. In the case study considered, it has been shown that one of the five input variables initially chosen to describe the PSS response, A_{AV} (i.e., the flow area of the Activation Valve that opens to trigger the PSS operation), does not play a significant role in the determination of the PSS energy output. The other four parameters, as expected, affect the amount of energy exchanged by the PSS: for example, a certain percentage of non-condensable gases (*NC*%) inside the PSS steam line causes a deterioration of the heat transfer coefficient inside the E-HX; moreover, a possible delay or leakage in the Main Steam Isolation Valve closure (i.e., DEL_{MSIV} and A_{MSIV} respectively), as a delay in the Activation Valve opening (i.e., DEL_{AV}), negatively affect $E_{ex,\%}$. Most of the input combinations explored by Kriging have been found to lead to PSS failure: this suggests that, in a hypothetical design phase, the variations of such relevant inputs should be limited only for a small portion of the explored ranges. Moreover, as expected, it has been shown that in many cases the PSS functional failure is favored by the "*combined action*" of input parameters varying together (with respect to one-at-a-time variations). Thanks to the use of the proposed metamodel-based technique, the computational cost associated to the analysis has been reduced by 3–4 *orders of magnitude*.

One of the advantages of the proposed method is that – due to the Gaussian assumption – a standard deviation (and, thus, a Confidence Interval-CI) can be associated to each kriging evaluation. This property has been exploited for assessing the accuracy and precision of the metamodel in predicting a new PSS configuration. Most important, this has allowed defining a "confidence interval" even for the safe (resp., critical failure) regions identified. The corresponding lower and upper bounds can be interpreted as the least and most conservative estimates of the safe (resp., failure) domain, respectively. In other words, we have been able to quantify a sort of licensing-defined tolerance (confidence) on the input parameter uncertain values that result in a safe PSS operation. On the other hand, the method inherits the limits of the technique employed: in particular, traditional Kriging metamodels require accommodating properties like continuity and smoothness of the approximated function. Thus, large prediction errors may result when AK-MCS is used to approximate non-smooth or multimodal output distributions. In such cases, different approaches (e.g., multiple metamodels trained by clustered training sets, Finite Mixture Models, …) could be adopted, which is the object of future research.

Finally, it is worth providing a closing comment on the relevance and usefulness of the approach here proposed, by highlighting its possible role within the safety and risk analyses classically carried out for nuclear systems. As mentioned before, the main purpose of the procedure is to identify the combinations of input configurations (i.e., sequences of *events* or *component states* or *design variable values*) that lead the PSS to critical conditions (i.e., to *functional failure*). However, a complete picture of the risk associated to the PSS also requires the assessment of the *probabilities* of these critical, dangerous conditions (which is in general accomplished by *quantifying* the uncertainties in the system behavior, *modeling* by proper probability distributions and *propagating* them through the model code). Since this task was beyond the scopes of the present paper, it has not been addressed. However, research is planned to *embed* the adaptive exploration scheme here presented within advanced Monte Carlo Simulation (MCS) approaches for the *efficient* estimation of the functional failure probability of PSSs, with particular attention to those situations where the failure region is very small and far from the nominal design (Villén-Altamirano and Villén-Altamirano, 2011; Pedroni and Zio, 2017; Schöbi et al., 2017; Turati et al., 2016; Yang and Cheng, 2020; Yang et al., 2020).

Table 5
Average CV error in trend type estimation.

Trend type	Ordinary	Linear	Quadratic
Average CV error [%]	8.82	5.13	9.53

Table 6
Average CV error in correlation function family estimation.

Corr. Function family	Exponential	Gaussian	Matérn 3/2	Matérn 5/2
Average CV error [%]	5.58	5.31	5.55	4.99

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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 thank the three anonymous referees for their constructive comments that significantly helped improving the paper.

Appendix A. – Kriging metamodels

Kriging is defined as a stochastic interpolation algorithm, which assumes that the model output $y = f(\boldsymbol{X})$ is the realization of a Gaussian process indexed by $\boldsymbol{X} \in D_X \subset \mathbb{R}^M$, where $D_X$is the metamodel domain of validity and M is the dimensionality of the problem (Turati et al., 2017):

$$y = f(\boldsymbol{X}) = \mathscr{N}\left(\boldsymbol{h}(\boldsymbol{X})^T\boldsymbol{\beta}, \sigma^2 Z(\boldsymbol{X})\right), \tag{6}$$

where the first term, $\boldsymbol{h}(\boldsymbol{X})^T\boldsymbol{\beta}$, is the mean value of the Gaussian process, also known as trend, consisting of Parbitrary functions $\{h_j; j = 1, \cdots P\}$ and the corresponding coefficients $\{\beta_j; j = 1, \cdots P\}$, and the second term consists of the variance of the Gaussian process σ^2 and a zero mean, unit variance stationary Gaussian process $Z(\boldsymbol{X})$; the correlation function underlying $Z(\boldsymbol{X})$is represented by $R(\boldsymbol{x}, \boldsymbol{x}^{'}; \boldsymbol{\theta})$, where $R$is the correlation matrix (given a certain correlation function family) and $\boldsymbol{\theta}$ its hyperparameters. In particular, $R(\boldsymbol{x}, \boldsymbol{x}^{'}; \boldsymbol{\theta})$ describes the correlation between two vectors $\boldsymbol{x}, \boldsymbol{x}^{'}$: the closer they are the higher their correlation is. The Gaussian process assumption states that every set of realizations of the model output can be described by a Gaussian vector, whose relation between a single realization $y(\boldsymbol{x})$ and the rest of the set $\mathscr{Y}_{train} \in \mathbb{R}^{N_{train}}$ follows a Gaussian distribution defined by:

$$\begin{bmatrix} y(\boldsymbol{x}) \\ \mathscr{Y}_{train} \end{bmatrix} \mathscr{N}_{N_{train}+1}\left(\begin{bmatrix} h(\boldsymbol{x})^T\beta \\ H\beta \end{bmatrix}; \sigma^2 \begin{bmatrix} 1 & \boldsymbol{r}^{\boldsymbol{T}}(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x}) & \boldsymbol{R} \end{bmatrix}\right) \tag{7}$$

In detail, H is the information matrix of the Kriging metamodel trend and in each row there are the regressors related to the corresponding observation $\boldsymbol{x}_i$ (i.e., $H_i = \boldsymbol{h}(\boldsymbol{x}_i), i = 1, \cdots, N_{train}$); R is the correlation matrix with elements $R_{ij} = R(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\theta}), i, j = 1, \cdots, N_{train}$, and $\boldsymbol{r}(\boldsymbol{x})$ is the vector of cross correlations between $\boldsymbol{x}$ and each of the other vectors whose elements reads as $r_i = R(\boldsymbol{x}, \boldsymbol{x}_i; \boldsymbol{\theta}), i = 1, \cdots, N_{train}$).

In the context of metamodeling, the interest is to predict a new point response, hence, given an experimental design or training set, i.e., $\{\mathscr{X}_{train}, \mathscr{Y}_{train}\}$, with $\mathscr{Y}_{train} = (y_1, \cdots, y_{N_{train}})$ and with an associated information matrix H and correlation matrix R, the prediction of the output, i.e., $\widehat{y}$ for a given input configuration $\boldsymbol{x}$ is given by:

$$\widehat{y}(\boldsymbol{x}) | \boldsymbol{y}, \sigma^2, \boldsymbol{\theta}\ \mathscr{N}\left(\mu_{\widehat{y}}; \sigma^2_{\widehat{y}}\right), \tag{8}$$

where $\mu_{\widehat{y}}(\boldsymbol{x})$ and $\sigma^2_{\widehat{y}}(\boldsymbol{x})$ are respectively the mean value and the variance of the Gaussian random variate $\widehat{y}$, defined by:

$$\mu_{\widehat{y}}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{\beta} + \boldsymbol{r}(\boldsymbol{x})^T R^{-1}(\mathscr{Y}_{train} - H\boldsymbol{\beta}), \tag{9}$$

$$\sigma^2_{\widehat{y}}(\boldsymbol{x}) = \sigma^2\left(1 - \boldsymbol{r}(\boldsymbol{x})^T R^{-1}\boldsymbol{r}(\boldsymbol{x})^T\right) + \left(\boldsymbol{h}(\boldsymbol{x})^T - \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}} R^{-1} H\right)\left(H^{\mathrm{T}} R^{-1} H\right)^{-1}\left(\boldsymbol{h}(\boldsymbol{x})^T - \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}} R^{-1} H\right)^T \tag{10}$$

And the least square estimates of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \left(H^{\mathrm{T}} R^{-1} H\right)^{-1} H^T R^{-1}\mathscr{Y}_{train} \tag{11}$$

An important property of Kriging predictor is that is an exact interpolator, i.e., the prediction variance at experimental design points collapses to zero. Another main advantage of this approach is that a confidence interval is returned together with each prediction $\widehat{y}(\boldsymbol{x})$ and, hence, it can be exploited for assessing he accuracy and precision of the metamodel in predicting a new configuration.

Appendix B. – Best Kriging setting calculations

The UQLab software allows to tune different Kriging features accordingly to the properties of the case study. For the specific application of the Kriging metamodel used to mimic the RELAP5-3D model simulating the response of the DHR system, different options have been tested through the CV procedure. The initial I/O training set (made by 64 RELAP5-3D simulations) has been split into K partitions (with $K = 4$) of the same size: a metamodel has been trained on $K-1$ partitions and the CV error (which is actually a RMSE error) has been evaluated by comparing the metamodel predictions corresponding to the input combinations of the left out partition k with the true model outputs. The process has been repeated for $k = 1, 2, \cdots K$ and the CV error has been averaged. Then, the whole procedure has been repeated to calculate the average CV error with other Kriging options and, finally, the best option of a certain kind (e.g., the best trend option) has been selected according to the lowest CV error. In particular, two Kriging features have been tested: the trend type and correlation function family (see Tables 5 and 6), whereas the other features have been set to their default options defined in UQLab.

Note that the average CV error used to rank the different options is expressed in percentage because it has the same unit of measure of the predicted output, i.e., the percentage of energy exchanged ($E_{ex,\%}$). The best trend type has been evaluated with all the other Kriging features set to their default options defined in UQLab (the correlation function family set by default is *Matérn 5/2*). The same default options have been used for the estimation of the best correlation function family, with the trend type set to *Linear* (optimal setting found at the previous step).

References

Alcaro, F., Bersano, A., Bertani, C., Mascari, F., 2021. BEPU analysis of a passive decay heat removal system with RELAP5/MOD3.3 and RELAP5-3D. Prog. Nucl. Energy 136, 103724. https://doi.org/10.1016/j.pnucene.2021.103724.

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

Bersano, A., Bertani, C., Falcone, N., Salve, M. de, Mascari, F., Meloni, P., 2020. Qualification of RELAP5-3D code against the in-pool passive energy removal system PERSEO data. In : Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference, 2020.

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 J. 46 (10), 2459–2468. https://doi.org/10.2514/1.34321.

Borgonovo, E., Plischke, E., 2016. Sensitivity analysis: a review of recent advances. Eur. J. Oper. Res. 248 (3), 869–887. https://doi.org/10.1016/j.ejor.2015.06.032.

Burgazzi, L., 2004. Evaluation of uncertainties related to passive systems performance. Nucl. Eng. Des. 230 (1–3), 93–106. https://doi.org/10.1016/j.nucengdes.2003.10.011.

Cadini, F., Santos, F., Zio, E., 2014. An improved adaptive kriging-based importance technique for sampling multiple failure regions of low probability. Reliab. Eng. Syst. Saf. 131, 109–117. https://doi.org/10.1016/j.ress.2014.06.023.

Crombecq, K., Laermans, E., Dhaene, T., 2011a. Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. Eur. J. Oper. Res. 214 (3), 683–696. https://doi.org/10.1016/j.ejor.2011.05.032.

Crombecq, K., Gorissen, D., Deschrijver, D., Dhaene, T., 2011b. A novel hybrid sequential design strategy for global surrogate modeling of computer experiments. SIAM J. Sci. Comput. 33 (4), 1948–1974. https://doi.org/10.1137/090761811.

Cox, D.D., John, S., 1997. SDO: a statistical method for global optimization. In: Alexandrov, M.N., Hussaini, M.Y. (Eds.), Multidisciplinary Design Optimization: State-of-the-art. Philadelphia: Siam, pp. 315–29.

Dubourg, V., Sudret, B., Deheeger, F., 2013. Metamodel-based importance sampling for structural reliability analysis. Probab. Eng. Mech. 33, 47–57. https://doi.org/10.1016/j.probengmech.2013.02.002.

Echard, B., Gayton, N., Lemaire, M., 2011. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo Simulation. Struct. Saf. 33 (2), 145–154. https://doi.org/10.1016/j.strusafe.2011.01.002.

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

Garud, S.S., Karimi, I.A., Kraft, M., 2017. Design of computer experiments: a review. Comput. Chem. Eng. 106, 71–95. https://doi.org/10.1016/j.compchemeng.2017.05.010.

Gu, M., Berger, J.O., 2016. Parallel partial gaussian process emulation for computer models with massive output. Ann. Appl. Stat 10 (3), 1317–1347.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3 (Mar), 1157–1182.

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

Herer, C., Dimitrov, B., Evrard, J.M., Lejosne, A., Wattelle, E., 2019. IRSN activities related to passive safety systems assessment. ICAPP 2019 - International Congress on Advances in Nuclear Power Plants.

Inselberg, A., 2009. Parallel Coordinates. Springer International Publishing, Visual Multidimensional Geometry and its Application.

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

Jafari, J., D'Auria, F., Kazeminejad, H., Davilu, H., 2003. Reliability evaluation of a natural circulation system. Nucl. Eng. Des. 224 (1), 79–104. https://doi.org/10.1016/S0029-5493(03)00105-5.

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., 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.

Lataniotis, C., Marelli, S., Sudret, B., 2020. Extending classical surrogate modelling to high dimensions through supervised dimensionality reduction: a data-driven approach. Int. J. Uncertainty Quantification 10 (1), 55–82. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020031935.

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

Liu, H., Ong, Y.-S., Cai, J., 2018. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Struct. Multidisc. Optim. 57 (1), 393–416. https://doi.org/10.1007/s00158-017-1739-8.

Liu, Longjun, 2005. Could Enough Samples be more Important than Better Designs for Computer Experiments? 2005, pp. 107–115. DOI: 10.1109/ANSS.2005.17.

Loeppky, J.L., Moore, L.M., Williams, B.J., 2010. Batch sequential designs for computer experiments. J. Stat. Plann. Inference 140 (6), 1452–1464. https://doi.org/10.1016/j.jspi.2009.12.004.

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, J.D., Simpson, T.W., 2005. Use of Kriging models to approximate deterministic computer models. AIAA J. 43 (4), 853–863. https://doi.org/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.

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, F., Araneo, D., Galassi, G., D'Auria, F., 2009. Application of REPAS methodology to assess the reliability of passive safety systems. Sci. Technol. Nucl. Install. 2009, 1–18. https://doi.org/10.1155/2009/768947.

Pedroni, N., Zio, E., 2017. An adaptive metamodel-based subset importance sampling approach for the assessment of the functional failure probability of a thermal-hydraulic passive system. Appl. Math. Model. 48, 269–288. https://doi.org/10.1016/j.apm.2017.04.003.

Ranjan, P., Bingham, D., Michailidis, G., 2008. Sequential experiment design for contour estimation from complex computer codes. Technometrics 50 (4), 527–541. https://doi.org/10.1198/004017008000000541.

Saltelli, A., 2008. Global Sensitivity Analysis: The Primer. John Wiley, Chichester, England; Hoboken, NJ.

Schöbi, R., Sudret, B., Marelli, S., 2017. Rare Event Estimation Using Polynomial-Chaos Kriging. ASCE-ASME J. Risk Uncertainty Eng. Syst. Part A: Civil Eng. 3 (2), D4016002.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliab. Eng. Syst. Saf. 93 (7), 964–979.

Turati, P., Cammi, A., Lorenzi, S., Pedroni, N., Zio, E., 2018a. Adaptive simulation for failure identification in the Advanced Lead Fast Reactor European Demonstrator. Prog. Nucl. Energy 103, 176–190. https://doi.org/10.1016/j.pnucene.2017.11.013.

Turati, P., Pedroni, N., Zio, E., 2016. Advanced RESTART method for the estimation of the probability of failure of highly reliable hybrid dynamic systems. Reliab. Eng. Syst. Saf. 154, 117–126. https://doi.org/10.1016/j.ress.2016.04.020.

Turati, P., Pedroni, N., Zio, E., 2017. Simulation-based exploration of high-dimensional system models for identifying unexpected events. Reliab. Eng. Syst. Saf. 165, 317–330. https://doi.org/10.1016/j.ress.2017.04.004.

Turati, P., Pedroni, N., Zio, E., 2018b. In: Knowledge in Risk Assessment and Management. John Wiley & Sons, Ltd, Chichester, UK, pp. 165–219. https://doi.org/10.1002/9781119317906.ch8.

Verleysen, M., François, D., 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.

Villén-Altamirano, M., Villén-Altamirano, J., 2011. The rare event simulation method RESTART: efficiency analysis and guidelines for its application. Vol. 5233. In: Kouvatsos, D.D. (Eds.), Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (pp. 509–547), Springer, Berlin, Heidelberg.

Wu, X.u., Kozlowski, T., Meidani, H., Shirvan, K., 2018. Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory. Nucl. Eng. Des. 335, 339–355. https://doi.org/10.1016/j.nucengdes.2018.06.004.

Xiao, N.-C., Zuo, M.J., Zhou, C., 2018. A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis. Reliab. Eng. Syst. Saf. 169, 330–338. https://doi.org/10.1016/j.ress.2017.09.008.

Yang, X., Cheng, X., 2020. Active learning method combining Kriging model and multimodal-optimization-based importance sampling for the estimation of small failure probability. Int. J. Numer. Meth. Eng. 121 (21), 4843–4864. https://doi.org/10.1002/nme.v121.2110.1002/nme.6495.

Yang, X., Cheng, X., Wang, T., Mi, C., 2020. System reliability analysis with small failure probability based on active learning Kriging model and multimodal adaptive importance sampling. Struct. Multidiscip. Optim. 62 (2), 581–596.

Zio, E., Bazzo, R., 2011. Level Diagrams analysis of Pareto Front for multiobjective system redundancy allocation. Reliab. Eng. Syst. Saf. 96 (5), 569–580.

Zio, E., Bazzo, R., 2012. Multiobjective reliability allocation in multi-state systems: decision making by visualization and analysis of pareto fronts and sets. Springer Ser. Reliab. Eng. 51, 195–208.