Turati P., Pedroni N., & Zio E. (2017). Simulation-based exploration of high-dimensional system models for identifying unexpected events.

Reliability Engineering and System Safety.

http://dx.doi.org/10.1016/j.ress.2017.04.004

Simulation-based exploration of high-dimensional system models for identifying unexpected events

Pietro Turati[1], Nicola Pedroni[1], Enrico Zio[1,2,*]

[1] Chaire Systems Science and the Energy Challenge, Fondation Electricite' de France (EDF), Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay Grande voie des Vignes, 92290 Chatenay-Malabry, France

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

[*] Corresponding author: enrico.zio@polimi.it, enrico.zio@centralesupelec.fr

Abstract

Mathematical numerical models are increasingly employed to simulate system behavior and identify sequences of events or configurations of the system's design and operational parameters that can lead the system to extreme conditions (Critical Region, CR). However, when a numerical model is: i) computationally expensive, ii) high-dimensional, and iii) complex, these tasks become challenging.

In this paper, we propose an adaptive framework for efficiently tackling this problem: i) a dimensionality reduction technique is employed for identifying the factors and variables that most affect the system behavior; ii) a meta-model is sequentially trained to replace the computationally expensive model with a computationally cheap one; iii) an adaptive exploration algorithm based on Markov Chain Monte Carlo is introduced for exploring the system state space using the meta-model; iv) clustering and other techniques for the visualization of high dimensional data (e.g., parallel coordinates plot) are employed to summarize the retrieved information.

The method is employed to explore a power network model involving 20 inputs. The CRs are properly identified with a limited computational cost, compared to another exploration technique of literature (i.e., Latin Hypercube Sampling).

Keywords: Critical Region exploration; Unexpected Event; Polynomial Chaos Expansion; Kriging; Markov Chain Monte Carlo (MCMC); Clustering; Local Outlier Factor; Integrated Deterministic Probabilistic Safety Assessment (IDPSA).

Acronyms

AK-MCS	: Adaptive Kriging-Monte Carlo Simulation	LOF	: Local Outlier Factor
CR	: Critical Region	LOO	: Leave-One-Out
CRA	: Computational Risk Assessment	MCMC	: Markov Chain Monte Carlo
CSN	: Consejo de Seguridad Nuclear (Nuclear Safety Council)	MM	: Meta-Model
		MS	: Main Source
DC	: Direct Current	NSS	: Not Supplied Set
DEX	: Deep EXploration	PCE	: Polynomial Chaos Expansion
DOE	: Design Of Experiment	PCP	: Parallel Coordinates Plot
ENS	: Energy Not Served	QMC	: Quasi Monte Carlo
I/O	: Input/Output	RISMC	: Risk-Informed Safety Margin Characterization
IDPSA	: Integrated Deterministic Probabilistic Safety Assessment		
		SPLOM	: Scatter PLOt Matrix
INL	: Idaho National Laboratories	UCR	: Unexplored Critical Region
ISA	: Integrated Safety Assessment	UECR	: Unexplored Extreme Critical Region
kNN	: k-Nearest Neighbors		
LAR	: Least Angle Regression		
LHS	: Latin Hypercube Sampling		

1 Introduction

Unexpected, disruptive events are the real challenge of the residual risk from the operation of an engineering system, such as a nuclear power plant, an oil and gas plant, a power grid, etc. For example, unexpected events occurring in critical infrastructures, such as power distribution networks, can propagate their disrupting consequences to other connected systems and lead to cascading effects with huge economical and safety impacts (Kröger & Zio, 2011; Vaiman et al., 2012; Zio, 2016a). However, such consequences can be strongly mitigated if those events are - to some extent - known in advance (Pate-Cornell, 2002; Paté-Cornell, 2012). The importance of knowledge as a means for avoiding surprise and unexpected events is testified by the large number of works emerging on the subject in recent years (Terje Aven, 2015, 2016a; Terje Aven & Krohn, 2014; Cavalcante, Oriá, Sornette, Ott, & Gauthier, 2013; Sornette, 2009; Sornette, Maillart, & Kröger, 2013; Turati et al., 2016a; Zio, 2016b).

Modeling and simulation have long been advocated as ways to explore and understand system behavior. Their use has been steadily increasing with the complexity of the systems, which makes experimentation economically unsustainable and physically infeasible. Design-Of-the-Experiment (DOE) approaches have been proposed to test operating conditions in order to study the corresponding system response with respect to different criteria: safety, reliability, resilience, business continuity, etc. (Santner et al., 2003; Simpson et al., 2001). Particular interest is in the identification of those factors, parameters and variables

that can lead the system (actually the model that represents it) to critical conditions (Bier et al., 1999; Zio, 2016a).

In this paper, we focus on mathematical models that give an Input/Output (I/O) representation, i.e., $Y = f(\boldsymbol{X})$, of the system behavior. In this setting, a configuration of input values $\boldsymbol{x}$ is considered critical if the output y takes value above a predefined safety threshold, i.e., $y = f(\boldsymbol{x}) \geq Y_{thres}$, which represents a physical limit beyond which the system is unsafe. For example, if in a nuclear power plant the temperature of the fuel cladding exceeds a limit value, the plant state is considered critical. Nevertheless, in practical cases the numerical models employed are very complex and the function $f(\cdot)$ is not explicit, it is a sort of black-box function.

A possible approach to discover the configurations of factors that lead to undesired conditions, i.e. the Critical Regions (CRs), is to resort to a large number of model simulations and to a posteriori retrieve information regarding the shape and the number of regions (Santner et al., 2003). In this light, a huge research effort has been devoted to the selection of the input configurations to evaluate. In particular, several DOEs have been proposed with the aims of filling the system state space as uniformly as possible, such as Latin Hypercube Sampling (LHS) (Iman, 2008; Michael D McKay et al., 2000), which has a desirable projection property, or Sobol' sequences, which are Quasi Monte Carlo (QMC) sequences minimizing the discrepancy (Sobol, 1976), and others (V. C. P. Chen, K.-L. Tsui, R. R. Barton, & M. Meckesheimer, 2006). Even if these techniques are designed for filling high-dimensional spaces (Sobol et al., 2011), they are not a viable solution when the computational cost of a single model simulation is high and the number of calls to the model becomes the bottleneck of the analysis. For this reason, some iterative sampling techniques aimed at adaptively guiding the system towards the critical condition have been proposed (Cadini et al., 2014; J. H. Li et al., 2011; Maljovec, Wang, Pascucci, Bremer, & Mandelli, 2013; Picheny et al., 2010; Turati et al., 2016a). Still, there is a need for a general and efficient method capable of exploring models that are: *i)* computationally demanding; *ii)* high-dimensional; *iii)* complex, i.e., the CRs can be multiple, disconnected and involving different shapes (K H Kernstine, 2013).

Some approaches have been recently proposed in the nuclear community, which make use of sophisticated softwares that combine methodologies for risk assessment with very accurate thermo-hydraulic codes (Alfonsi et al., 2015; Alfonsi et al., 2016; Izquierdo et al., 2009), to perform the so-called Integrated Deterministic Probabilistic Safety Assessment (IDPSA) (Aldemir, 2013; Zio, 2014) or Computational Risk Assessment (CRA). We cite, for example, the Integrated Safety Assessment (ISA) (Queral et al., 2016) from the Spanish Consejo de Seguridad Nuclear (CSN) and the Risk-Informed Safety Margin Characterization (RISMC) from the Idaho National Lab (INL) (Smith et al., 2016). For example, the RISMC approach is built on four main steps in order to accurately define the safety margins of a complex system (Mandelli, Smith, et al., 2013): 1) sampling in the uncertainty space, 2) simulation of the computational expensive model to evaluate the system outputs, 3) analysis of the simulation results, 4) visualization of the

outcomes (Maljovec, Wang, Pascucci, Bremer, Pernice, et al., 2013) and information retrieval (Maljovec et al., 2016; Mandelli, Yilmaz, et al., 2013).

In this paper, we propose a new method for exploring a computational expensive model and retrieving information on the system CRs. The method is based on four main steps for iteratively exploring the system state space by exploiting the available I/O observations to guide the exploration towards the regions of interest. In short, the first step aims at reducing the dimensionality of the space to be explored by means of a Polynomial Chaos Expansion (PCE)-based sensitivity analysis (Ciriello et al., 2013; Sudret, 2008). The second step has the objective of reducing the computational cost associated to a simulation by training a meta-model, namely, Kriging (Cadini et al., 2014; Kleijnen, 2009). The third step is devoted to deeply exploring the state space to identify the CRs by means of a Markov Chain Monte Carlo (MCMC) algorithm (Andrieu & Thoms, 2008). Finally, the last step consists in retrieving information regarding the CRs, like the cardinality and the shape, through clustering by k-means and techniques for high-dimensional data visualization like Parallel Coordinates Plot (PCP) (Inselberg, 2009).

The main differences between our approach and those proposed in the above mentioned references, lie in the initial step for model reduction and in the techniques that are employed in the remaining steps, e.g. the sequential training of the Kriging meta-model and the MCMC-based exploration. Demonstration is given with regards to a representative, critical infrastructure made by a power network of 10 nodes with time-variant demands (Mena et al., 2014). The response of the network is analyzed with respect to different failure scenarios characterized by 20 factors, including the failure times and magnitudes. Results are compared to those of a LHS-based exploration.

The rest of the paper is organized as follows: in Section 2, the problem and the main challenges of the present work are set; in Section 3, the whole framework is described in detail; in Section 4, some metrics for evaluating the performance of an exploration algorithm are proposed; in Section 5, a case study concerning a power network is employed to show the effectiveness of the proposed method; finally in Section 6, some conclusions and prospective developments are given.

2 Critical Regions Exploration

Let us assume that a mathematical model $Y = f(\boldsymbol{X})$ of the system behavior is available, whose input $\boldsymbol{X} \in D_{\boldsymbol{X}} \subset \mathbb{R}^M$, represents a given system operational configuration and whose output $Y \in D_{\boldsymbol{Y}} \subset \mathbb{R}$ reflects the condition/state of the system. We define the conditions where $Y \geq Y_{thres}$ as "critical" and the corresponding configurations of factors, i.e. $CR = \{\boldsymbol{x} \in D_{\boldsymbol{X}} \subset \mathbb{R}^M : y = f(\boldsymbol{x}) \geq Y_{thres}\}$ as the Critical Region (CR). From a mathematical perspective, we are looking for the solution of the inverse problem $\boldsymbol{x} = f^{-1}(y)$, with $y \geq Y_{thres}$; however, this is not viable in the majority of the engineering systems where $f(\boldsymbol{x})$ is a function embedded in numerical codes which is: *i)* complex, *ii)* black-box *iii)* not invertible.

A solution is, then, to resort to a DOE for exploring the I/O relation by means of numerical simulations and, then, retrieve information concerning the CRs through a post-processing (Santner et al., 2003). However, this approach is hard to pursue when: *i)* the numerical model is computationally demanding, i.e., the number of calls to the model can be considered as *scarce resource*; *ii)* the input domain is high-dimensional, i.e., the dimensionality of the space to explore is relatively high (say, more than 10 inputs); *iii)* the model is complex, i.e., the CRs can be multiple, disconnected and involving several shapes (K H Kernstine, 2013).

The present paper addresses the above issues, proposing a self-adaptive algorithm for exploring the numerical model and retrieving information regarding the CR. We do not consider eventual probabilistic distributions associated to $\boldsymbol{X}$, focusing instead on its range of values (i.e., on its domain), in order to explore all possible configurations in the search for CRs. Hence, hereafter, without loss of generality, we assume that all input factors are standardized, e.g., $\boldsymbol{X} \in D_{\boldsymbol{X}} = [0,1]^M$ (Rosenblatt, 1952); likewise, a standardization can be applied to the output Y. This helps in designing a general, problem-independent algorithm and in removing effects related to the different orders of magnitudes possibly existing among the input factors.

3 Proposed Exploration Framework

We propose a general framework for exploring the state space of a computational model of an engineering system. We firstly introduce the general idea of the framework and, then, we dedicate a specific subsection to each one of its four main steps. The driving idea is to iteratively run a (possibly small) number of model simulations, to retrieve knowledge from the available simulations and to guide the selection of new configurations towards the CRs. In short, the first step aims at identifying the factors that most affect the output of the model in order to limit the exploration only to the corresponding subspace (dimensionality reduction) resorting to PCE-based sensitivity analysis (Sudret, 2008). The second step aims at training a computationally cheap-to-run meta-model that accurately reproduces the response of the real model on the reduced space, with a particular attention to its ability to discriminate between the CRs and normal conditions, e.g., a Kriging meta-model (Kleijnen, 2009). The third step resorts to the meta-model for deeply exploring the reduced state space by means of MCMC, with the objective of visiting and, consequently, discovering those configurations of factors leading to critical outputs (Andrieu & Thoms, 2008). Finally, the last step employs clustering (e.g., k-means) and graphical representation techniques (e.g. PCP (Inselberg, 2009)) for retrieving information and describing the CRs found. Fig. 1 reports a flow diagram synthetizing the entire procedure, indicating as well the type of model used to run the simulations associated to each step.

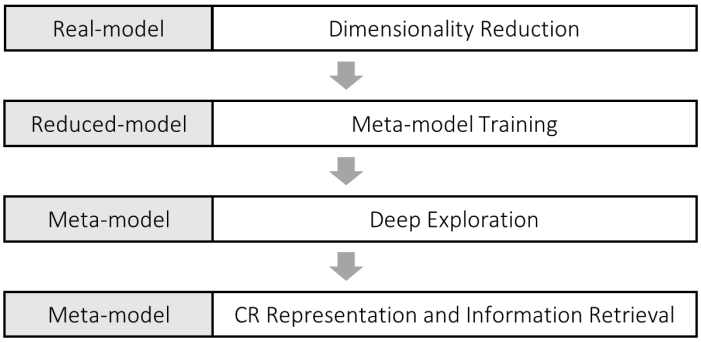

Fig. 1 Flow diagram of the entire framework.

3.1 Dimensionality Reduction

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 & Motoda, 2012). From the point of view of the exploration, reducing the dimensionality of the state space to explore allows the definition of a more effective DOE. Two 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 & Elisseeff, 2003), and *ii*) feature extraction, which aims at identifying a subset of "new" features created by means of transformations of the initial ones (Guyon & Elisseeff, 2006). Nevertheless, dimensionality reduction methods usually rely on a large set of I/O data examples that are not usually available, when the system model is computationally expensive.

In alternative, sensitivity analysis methods can be employed to achieve the same final objective as feature selection, by ranking the factors according to their influence on the output of the model (Borgonovo & Plischke, 2016; Saltelli, 2008; Sudret, 2008). In particular, to this aim, global order sensitivity indices are more appropriate than local sensitivity indices, because they provide a measure of how the inputs globally affect the output of the model, i.e., with respect to different configurations of the input factors. In this paper, we resort to the total order sensitivity index S_T (Homma & Saltelli, 1996; Sobol, 2001) that is a variance-based global sensitivity measure, assessing the expected fraction of the total variance of the output Y that is due to the variation of a specific input factor i and to its interactions with the others:

$$S_{Ti} = \frac{E_{\boldsymbol{X}_{\sim i}}\left[V_{\boldsymbol{X}_i}(Y|\boldsymbol{X}_{\sim i})\right]}{V(Y)}, \tag{1}$$

where $\boldsymbol{X_i}$ represents the *i*-th component of the input vector $\boldsymbol{X}$; $\boldsymbol{X}_{\sim i}$ represents the rest of the components of the vector $\boldsymbol{X}$ and $S_T \in [0,1]$. A large value of S_{Ti} indicates that the *i*-th factor heavily affects Y and, thus, should be kept in the reduced-model; on the contrary, a very low value of S_{Ti} indicates that

the *i*-th factor does not affect Y and, thus, it can be discarded or set to a constant value. Usually, a threshold $S_{thres} = 1/M$ is adopted to discriminate the important factors (Saltelli, 2008).

In order to limit the computational cost needed for assessing S_T, we resort to the PCE method (Sudret, 2008). In practice, by treating the input factors as stochastic variables described by a uniform distributions, the output function $Y = f(\boldsymbol{X})$ can be decomposed by means of a PC expansion:

$$Y = f(X_1, \dots, X_M) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \tilde{y}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(X_1, \dots, X_M), \tag{2}$$

where $\tilde{y}_{\alpha}$ is the coefficient associated to the multivariate Hilbertian basis $\psi_{\boldsymbol{\alpha}}(\cdot)$, orthonormal with respect to the multivariate uniform distribution ($\boldsymbol{X} \in D_{\boldsymbol{X}} = [0,1]^M$). Notice that $\psi_{\boldsymbol{\alpha}}(\cdot)$ is a multivariate Legendre polynomial, where the multi-index $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)$ indicates the order of the polynomials associated to each component of the vector $\boldsymbol{X}$. Thanks to the orthonormality of the Hilbertian basis, it can be shown that the variance of the output reads:

$$\sigma_Y^2 = V[f(\boldsymbol{X})] = V\left[\sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \tilde{y}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X})\right] = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \tilde{y}_{\boldsymbol{\alpha}}^2 E[\psi_{\boldsymbol{\alpha}}^2(\boldsymbol{X})] - \tilde{y}_0^2 = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \tilde{y}_{\boldsymbol{\alpha}}^2 - \tilde{y}_0^2 \tag{3}$$

where $\tilde{y}_0$ is the coefficient associated to the polynomial of order zero, i.e., that representing the expected value of Y. Similarly, the total order sensitivity indices can be computed as:

$$S_{Ti} = \frac{\sum_{\boldsymbol{u} \in \mathbb{N}_i^M} \tilde{y}_{\boldsymbol{u}}^2}{\sigma_Y^2}, \tag{4}$$

where $\mathbb{N}_i^M = \{\boldsymbol{u} \in \mathbb{N}^M \; s.t. \; u_i \neq 0\}$ is the subset of all the multi-indices corresponding to multivariate Legendre polynomials with non-zero degree associated to the *i*-th component, i.e., those polynomials that include the *i*-th component (Sudret, 2008). In this way, the numerator represents the contribution of the input *i* and its interactions with the other inputs to the total variance σ_Y^2. Nonetheless, the result in (4) requires to compute a countably infinite number of coefficients $\tilde{y}_{\boldsymbol{\alpha}}$, thus, in practice, the PCE in (2) is truncated to polynomials up to a given order p such that a sufficient approximation of the response function f is guaranteed:

$$Y = f(X_1, \dots, X_M) \approx \sum_{\boldsymbol{\alpha} \in A^{M,p}} \hat{y}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(X_1, \dots, X_M), \tag{5}$$

where $A^{M,p} \subset \mathbb{N}^M$ is the multi-index subset corresponding to polynomials having maximum order equal to p, i.e., $A^{M,p} = \{\boldsymbol{\alpha} \in \mathbb{N}^M \; s.t. \; |\boldsymbol{\alpha}| \leq p\}$ and $\hat{y}_{\boldsymbol{\alpha}}$ are the estimators of the corresponding coefficients, which are on the whole $P = \binom{M+p}{p}$. Similarly, the total order sensitivity indices can be approximated as:

$$S_{Ti} \approx \hat{S}_{Ti} = \frac{\sum_{\boldsymbol{u} \in U_i^M} \hat{y}_{\boldsymbol{u}}^2}{\sum_{\boldsymbol{\alpha} \in A^{M,p}} \hat{y}_{\boldsymbol{\alpha}}^2 - \hat{y}_{\boldsymbol{0}}^2}, \tag{6}$$

where $U_i^M \subset \mathbb{N}_i^M \cap A^{M,p}$ is the multi-index subset corresponding to polynomials having the *i*-th component larger than zero and maximum order equal to p. Thus, in practice, the computational cost for estimating S_T depends on the computational cost needed to estimate the coefficients $\tilde{y}_{\alpha}$, i.e., to approximate the function f (Sudret, 2008).

Regression methods have been shown to be a non-intrusive efficient tool for estimating the coefficient $\tilde{y}_{\alpha}$ by resorting to a set of I/O configurations (Sudret, 2008). In order to be trained, a regression model needs at least a number of samples N_{PCE} larger than the number of coefficients P, whose values could be obtained by means of QMC and LHS. However, the minimum number P of samples required increases with the dimension of the input space M and with the order of the polynomial p, making this approach unfeasible when simulations are computationally demanding, involve many inputs and have a non-smooth response function. Similar issues are shared also by the so-called Non-Intrusive Spectral Projection (NISP) (Crestaux et al., 2009).

For this reason, an adaptive sparse PCE representation, coupled with a Least Angle Regression (LAR), has been devised by (Blatman & Sudret, 2011) and it is here employed to detect the most significant polynomials. In practice, the PCE is built by adding, one at the time, the polynomial that most correlate with the residuals (i.e., the polynomial that better explains "part of the behavior" of the function f that is not yet captured by the polynomials already selected for the expansion), until a sufficient level of accuracy is reached. Consequently, only some of the P coefficients have to be estimated, thus limiting the number of I/O observations needed for accurately computing the corresponding coefficients. A further reduction in N_{PCE} can be achieved thanks to a recently proposed optimal DOE (Burnaev et al., 2016).

Moreover, since the objective of the current step is to identify those inputs that most affect the output, it is sufficient that the truncated PCE catches the global and general trend of the response function. Thus, in order to further limit the computational cost, the maximum order of the polynomial p can be fixed to a relatively low value, being aware that the lower p, the lower the capability of the PCE of reproducing local, possibly abrupt changes of the response function. All the analyses involving both the PCE approximation and the corresponding computation of the sensitivity indices, are conducted using the UQLab Toolbox for Matlab (Marelli & Sudret, 2014).

Once the important factors are identified, the remaining ones are either removed or fixed to constant values; nevertheless, in what follows, in both cases we will refer to the resulting model as "reduced-model".

3.2 Meta-modeling

The main objective of a meta-model is to reproduce the behavior of the real (typically long-running) system model with a less expensive computational model. The meta-model is trained by resorting to a typically limited number of I/O observations from the real reduced-model; on this basis, it should be capable of predicting the output values associated to input configurations that have not been explored yet. Since the

real model is assumed to be deterministic (i.e., simulations of the same input configuration lead to the same output), it is desirable that the meta-model predicts as well the exact output value in correspondence of the training configurations (i.e., those known with absolute certainty). In this respect, among the numerous methods available in the literature (Jin et al., 2001; Shan & Wang, 2010), we resort to Kriging (Kleijnen, 2009; Matheron, 1963), i.e., Gaussian process modeling. Actually, Kriging is capable of modeling local behaviors of the response function and of diversifying the levels of accuracy of the same model within different regions. For example, in our case, we are interested in a more refined model in the proximity of the CRs, whereas a rough one is sufficient for normal operating conditions. In practice, this can be obtained by concentrating the I/O training observations of the real model in the proximity of the CRs and of their limit surfaces. Sequential adaptive training strategies have been recently developed to this aim (Echard et al., 2011; Picheny et al., 2010). In what follows, the fundamental concepts of Kriging are recalled, with a focus on the adaptive strategy exploited for training the meta-model.

Kriging is 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, in our case, D_X is the domain of validity of the meta-model and $M' < M$ is the dimensionality of the reduced-model. In practice, Kriging is a linear regression model where the residuals are correlated by means of a Gaussian process, instead of being independent:

$$Y = f(\boldsymbol{X}) = N\big(\boldsymbol{h}(\boldsymbol{X})^T\boldsymbol{\beta}, \sigma^2 Z(\boldsymbol{X})\big), \tag{7}$$

where $\boldsymbol{h}(\boldsymbol{X})^T\boldsymbol{\beta}$ represents the mean value, also known as trend, which is a general linear regression model (e.g., $\boldsymbol{h}(\boldsymbol{X})$ can involve polynomial terms and it reflects the prior knowledge about the model), σ^2 is the variance of the Gaussian process, and $Z(\boldsymbol{X})$ is a zero mean, unit variance stationary Gaussian process whose underlying correlation function is represented by $R(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$. The correlation function is characterized by the parameters $\boldsymbol{\theta}$, which are function-specific and, it depends typically on the distance of the two vectors $\boldsymbol{x}, \boldsymbol{x}'$: the closer they are, the higher their correlation. Due to the Gaussian process hypothesis, 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 $\boldsymbol{y} \in \mathbb{R}^{N_{Krig}}$ reads:

$$\begin{bmatrix} Y(\boldsymbol{x}) \\ \boldsymbol{y} \end{bmatrix} \sim N_{N_{Krig}+1}\left(\begin{bmatrix} \boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{\beta} \\ \boldsymbol{H}\boldsymbol{\beta} \end{bmatrix}; \sigma^2 \begin{bmatrix} 1 & \boldsymbol{r}^T(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x}) & \boldsymbol{R} \end{bmatrix}\right). \tag{8}$$

In detail, $\boldsymbol{H}$ is the information matrix of $\boldsymbol{y}$ where each row represents the regressors associated to the corresponding observation $\boldsymbol{x}^{(i)}$ (i.e., $\boldsymbol{H}_i = \boldsymbol{h}\big(\boldsymbol{x}^{(i)}\big), i = 1, \dots, N_{Krig}$); $\boldsymbol{\beta}$, $\boldsymbol{h}(\boldsymbol{x})$ and σ^2 are defined as above; $\boldsymbol{R}$ is the correlation matrix (i.e., $\boldsymbol{R}_{ij} = R\big(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}; \boldsymbol{\theta}\big),\ i, j = 1, \dots, N_{Krig}$) and $\boldsymbol{r}(\boldsymbol{x})$ is the vector of the correlation between $\boldsymbol{x}$ and the other vector, (i.e., $\boldsymbol{r}(\boldsymbol{x}) = R\big(\boldsymbol{x}, \boldsymbol{x}^{(i)}; \boldsymbol{\theta}\big), i = 1, \dots, N_{Krig}$).

Assuming that $\boldsymbol{y} = (y_1, \dots, y_{N_{Krig}})$ is an experimental design with associated information matrix $\boldsymbol{H}$ and correlation matrix $\boldsymbol{R}$, then the prediction of the output $\hat{Y}$ for a given configuration $\boldsymbol{x}$ is given by:

$$\hat{Y}(\boldsymbol{x})|\boldsymbol{y}, \sigma^{\boldsymbol{2}}, \boldsymbol{\theta} \sim N\left(\mu_{\hat{Y}}; \sigma_{\hat{Y}}^2\right), \tag{9}$$

where

$$\mu_{\hat{Y}}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta} + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R^{-1}} (\boldsymbol{y} - \boldsymbol{H\beta}), \tag{10}$$

$$\sigma_{\hat{Y}}^2(\boldsymbol{x}) = \sigma^2 (1 - \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R^{-1}} \boldsymbol{r}(\boldsymbol{x})^T) + (\boldsymbol{h}(\boldsymbol{x})^T - \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{R^{-1}} \boldsymbol{H})(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{R^{-1}} \boldsymbol{H})^{-1} (\boldsymbol{h}(\boldsymbol{x})^T - \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{R^{-1}} \boldsymbol{H})^T \tag{11}$$

with the regression coefficients estimated by $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{R^{-1}} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{R^{-1}} \boldsymbol{y}$.

One of the main advantages of this formulation is that a confidence interval can be associated to each prediction $\hat{Y}(\boldsymbol{x})$. This can be used for assessing the accuracy and precision of the meta-model: the smaller the confidence interval, the more precise the model prediction for the corresponding configuration.

Since in this paper the focus is on the CR, the accuracy of the meta-model should be higher in the proximity of the CR. In particular, it is important that the meta-model is capable of discriminating the CR from the normal conditions; thus, instead of using an a-priori fixed DOE, a sequential one (where the experimental observations are iteratively and adaptively added to increase the accuracy of the meta-model around the regions of interest) is preferable. The Adaptive Kriging Monte Carlo Simulation (AK-MCS) (Echard et al., 2011) is here employed to this aim. In the AK-MCS, an initial Kriging model is trained with a small set of I/O observations, e.g., sampled according to LHS scheme; then, the algorithm proceeds iteratively according to the following steps: *i)* randomly sample a large set of input configurations $\mathcal{X} = \left(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N_{MCS})}\right)$, e.g., by means of LHS; *ii)* evaluate the associated responses using the Kriging meta-model $\hat{\mathcal{Y}} = \left(\hat{\boldsymbol{y}}_1, \dots, \hat{\boldsymbol{y}}_{N_{MCS}}\right)$; *iii)* check if a convergence criterion has been reached: if so, the meta-model is sufficiently accurate; otherwise, *iv)* select, according to a predefined learning function/criterion, the best candidate subset $\mathcal{X}^* \subset \mathcal{X}$ to add to the current DOE and evaluate the corresponding real model output $\mathcal{Y}^*$; *v)* retrain a new Kriging meta-model by adding the $\{\mathcal{X}^*, \mathcal{Y}^*\}$ to the training set and go back to step *i)*.

As learning function (step *iv* above), we consider the so-called *U*-function, which is based on the concept of misclassification (Echard et al., 2011):

$$U(\boldsymbol{x}) = \frac{|Y_{thres} - \mu_{\hat{Y}}(\boldsymbol{x})|}{\sigma_{\hat{Y}}(\boldsymbol{x})}. \tag{12}$$

In practice, $U(\boldsymbol{x})$ represents the distance in terms of standard deviations of the meta-model prediction from the limit state Y_{thres}. The smaller the value, the closer the prediction is to the limit state and, thus, the higher the interest in adding the corresponding I/O observation to the training set, because it reduces the prediction uncertainty regarding configurations "close" to the limit surface (in a probabilistic sense). Theoretically, the best DOE is obtained by adding at each iteration only one best candidate configuration. However, this increases the computational cost related to the training of the meta-model, which can be significant when a large number of I/O configurations are used and/or when many parameters have to be estimated in reason of the high dimensionality.

To overcome this problem, a larger number of I/O configurations can be added at the same time to the training set. Due to the correlation function, prediction points that are close share similar prediction values and misclassification probabilities; thus, it is likely that in the best candidate set, there are configurations having similar input factors values. However, evaluating the real model with respect to similar configurations increases the computational cost without adding the desired amount of knowledge to the meta-model. To this aim, clustering techniques are here employed to select, among the best candidate set, the most representative configurations before evaluating the corresponding real model output (Schöbi et al., 2016). An alternative method for optimally adding multiple observations to the training set has been recently proposed in (Chevalier et al., 2014).

As a stopping criterion (step *iii* above), we resort to the leave-one-out estimate of the correction factor $\hat{\alpha}_{corr\ LOO}$ (Dubourg et al., 2013):

$$\hat{\alpha}_{corr\ LOO} = \frac{1}{N_{Krig}} \sum_{n=1}^{N_{Krig}} \frac{\mathbb{I}_{f(\boldsymbol{x}^{(n)}) \geq Y_{thres}}(\boldsymbol{x}^{(n)})}{P\left(\hat{Y}_{DOE\setminus \boldsymbol{x}^{(n)}}(\boldsymbol{x}^{(n)}) \geq Y_{thres}\right)}, \tag{13}$$

where $\hat{Y}_{DOE\setminus \boldsymbol{x}^{(n)}}(\boldsymbol{x}^{(n)})$ is the prediction of the output associated to the factors $\boldsymbol{x}^{(n)}$, obtained with a Kriging model having as training set all the I/O observations except $\{\boldsymbol{x}^{(n)}, y_n\}$. This verifies that the probabilistic discriminating function (i.e., the prediction) converges towards the real discriminating function (i.e., the real limit surface). In practice, a value of $\hat{\alpha}_{corr\ LOO}$ close to 1 indicates a satisfactory approximation of the real model, whereas very small or very large values indicate an inaccurate approximation. It must be noticed that, since the estimation is based on a LOO cross-validation, a minimum number of initial I/O observations, (e.g. 30 (Dubourg et al., 2013)), has to be provided to guarantee accurate estimates. On the other side, a maximum number of iterations can be set, in order to limit the number of calls to the real model.

For building the meta-model, we resort to the UQLab Toolbox for Matlab (Marelli & Sudret, 2014), whereas the sequential training has been developed by the authors.

3.3 Deep Exploration

The aim of the Deep EXploration (DEX) phase is to explore the system state space by resorting to the meta-model, instead of computational expensive real one, to discover (possibly unknown) configurations leading to CRs. To achieve this goal, DEX starts from the "available knowledge" (in particular, from the set of critical configurations visited during the training of the meta-model) and exploits the following MCMC-based algorithm for adaptively exploring the state space by means of the meta-model.

The algorithm proceeds as follows:

1. Use a clustering algorithm for grouping the configurations $\mathcal{X}_{exp}^{CR}$ belonging to CRs. We resort to a k-means algorithm and to an ensemble of metrics to select the number of clusters *K* that provides the best description of the available data: see (Charrad, Ghazzali, Boiteau, Niknafs, & Charrad, 2014) for details. Then, check the stopping criterion: if it is satisfied, stop the DEX algorithm, otherwise, move to step 2.

2. Estimate for each cluster the associated covariance matrix Σ_k.

3. Distribute *C* MCMC chains among the identified clusters and sample the corresponding starting points. The simplest way to distribute the Markov chain is to randomly assign each chain to a cluster. However, this can lead to over-explore some regions and under-explore others. To tackle this problem, a new strategy for optimizing the distribution of the Markov chains have been here implemented. For each cluster k, its density δ_k is approximated by:

$$\delta_k = \frac{N_k}{V_k} = \frac{N_k}{\sqrt{\det(S_k)}}, \tag{14}$$

where N_k is the number of configurations in the *k*-th cluster whose volume V_k is approximated by the square root of the determinant of the corresponding covariance matrix, i.e., the square root of the product of the corresponding eigenvalues. In practice, a high value of density indicates that the CR has been deeply explored, whereas a small value corresponds to a roughly explored CR that needs additional exploration. Then, *C* Markov chains are distributed with the aim of maximizing the minimum density among clusters, which can be expressed by the following equations:

$$\max\left(\min_{k\in K} \delta_k + \frac{c_k \cdot N_{chain}}{V_k}\right) \tag{15}$$

$$\sum_{k\in K} c_k = C, \qquad c_k \in \mathbb{N}_0 \tag{16}$$

where c_k is a non-negative integer representing the number of Markov chains associated to the *k*-th cluster and N_{chain} is the length of each chain. In practice, the second term in the sum (15) represents the expected increase of density provided by running a Markov chain within the *k*-th cluster.

4. Sample a large number of factor configurations $N_{MH} = C \cdot N_{chain}$ by the Metropolis-Hasting algorithm: usually, the larger the dimensionality of the state space, the larger should be the number of configurations to sample (Chib & Greenberg, 1995; Robert & Casella, 2004). Multivariate Gaussian distributions with covariance matrices estimated at step 2 are employed as proposal distributions and a uniform distribution on the CRs support is considered as the target one. Adaptive MCMC algorithms can be employed to further increase the flexibility of the algorithm (Andrieu & Thoms, 2008; Roberts & Rosenthal, 2009).

5. Add the configurations belonging to CRs to $\mathcal{X}_{exp}^{CR}$ and go back to step 1.

The stopping criterion requests that the clustering algorithm at step 1 identifies the same optimal number of clusters for three consecutive iterations. This implies that the DEX algorithm is filling with extra configurations the same number of CRs, without adding extra knowledge. Alternative and/or additional criteria could be used: *i)* a maximum number of iterations N_{exp} can be set in order to control the computational cost associated to the clustering step; *ii)* a minimum value of density δ_k can be set to guarantee a desired level of accuracy in the meta-model exploration. A flow diagram of the entire procedure is reported in Fig. 2.

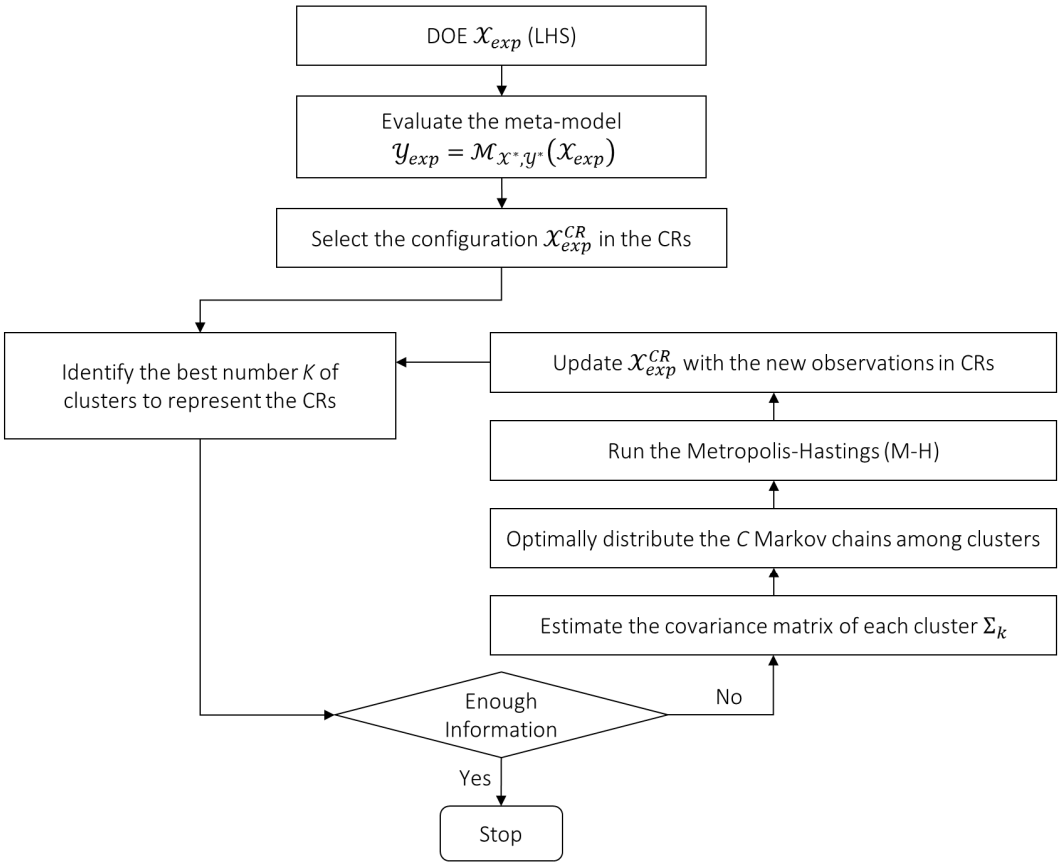

Fig. 2 Flow diagram of the deep exploration method.

3.4 Representation and Information Retrieval

At the end of the deep exploration step a number of critical configurations are available in the form of a list of vectors, whose physical interpretation (and visualization) may not be trivial. This is even more apparent when the dimensionality of the state space is larger than 3-4, since the most common representation techniques cannot be used (S. Liu et al., 2017). For this reason, a post-processing of the available critical configurations is needed for the retrieval of useful information.

Firstly, the critical configurations are grouped according to the optimal number of clusters obtained from the last iteration of the deep exploration algorithm (see Section 3.3). This allows the analyst to understand how many CRs characterize the system. Secondly, two techniques for high-dimensional data visualization are employed for their representation. These are the ScatterPLOt Matrix (SPLOM) (Hartigan, 1975) and the PCP (Inselberg, 2009), which help in retrieving complementary information about the CRs, such as their shapes and the corresponding input values in a unique, "readable", graphical representation.

In particular, SPLOM represents the two-dimensional projections of the CRs over all the possible pairs of coordinates, which aids the analyst in identifying the shapes of the CRs. Interactive scatter plots have been recently proposed to allow exploring interactively not only the projection on the two-dimensional spaces defined by the main coordinates, but also every two-dimensional projection of the available data (Cook & Swayne, 2007).

On the other side, in PCP each of the Mcoordinates is represented by a vertical axis and a point (i.e., an input vector) in the M-dimensional space is represented by a line in the PCP. For the sake of clarity, assuming to have a 5-dimensional point $\boldsymbol{x} = (2\ 4\ 1\ 5\ 1)$, its representation in PCP is given by Fig. 3. The advantage of PCP is that it represents all the coordinates in a single plot and, by so doing, it provides the analyst with information on the range of values defining the CRs and helps in distinguishing possible patterns characterizing the different clusters.

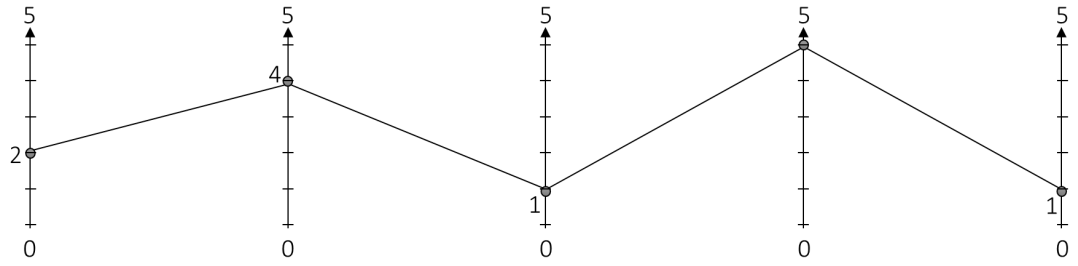

Fig. 3 PCP of a 5-dimensional point.

4 Exploration Assessment

Assuming that the real limit function representing the configurations in the CRs is available, the objective of the assessment phase is to measure how satisfactorily the exploration method has identified the configurations leading to critical conditions. Only for illustrative purposes, Fig. 4 left shows the output of an accurate exploration of a two-dimensional space, where the real CR (shadowed) is sufficiently covered by the configurations selected by the explorative method (circles); on the contrary, Fig. 4 right shows an incomplete exploration where a fraction of a CR is identified, but not entirely covered, and another CR is not even explored.

Quantitative metrics are here introduced to assess the quality of the exploration: in particular, the population of critical configurations visited by the proposed methodology $\mathcal{X}_{exp}^{CR}$ (cyrcles) is compared to a uniformly distributed population of samples belonging to the real CRs $\mathcal{X}_{real}^{CR}$ (crosses), according to a distance-based criterion.

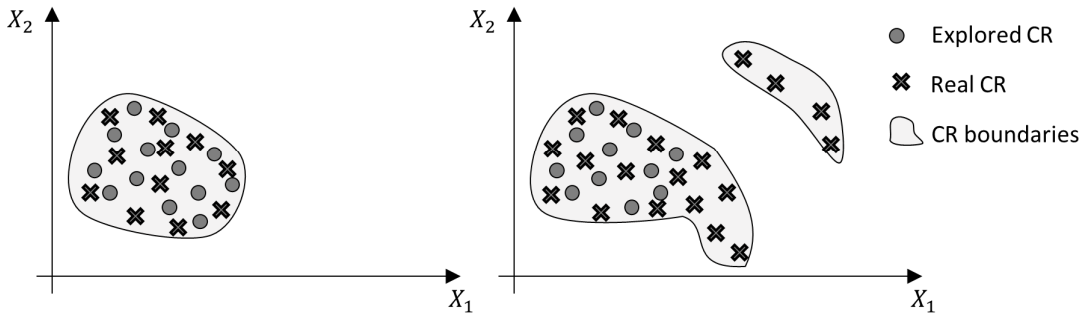

Fig. 4 Representation of an accurate CR exploration (left) and of an incomplete CR exploration (right).

A one-vs-all version of the Local Outlier Factor (LOF) is employed to this aim, where each configuration in the real CRs is compared to the whole population of critical configurations obtained by the exploration method. For the sake of completeness, LOF is a density-based outlier detection method capable of measuring how isolated is a sample from the rest of a given population of interest (Breunig et al., 2000). In our case, the more isolated a real CR configuration is from the explored ones, the higher the probability that it belongs to an unexplored CR.

The definition of the LOF relies on the concept of reachability distance between points $\boldsymbol{x}$ and $\boldsymbol{o}$:

$$d_{reach}(\boldsymbol{x},\boldsymbol{o}) = \max\big(d_{kNN}(\boldsymbol{o}), d(\boldsymbol{x},\boldsymbol{o})\big), \tag{17}$$

where $d(\cdot,\cdot)$ is a generic distance and $d_{kNN}(\boldsymbol{o})$ is the distance of the *k* Near Neighbor (*kNN*) of $\boldsymbol{o}$. In this paper, the Euclidean distance is employed; however, the Manhattan or even lower order L^p distances can be preferable in high dimensionality (Aggarwal et al., 2001). Then, the local reachability distance, which measures how close is the configuration $\boldsymbol{x}$ to its *kNNs,* can be defined as:

$$lrd_k(\boldsymbol{x}) = \frac{k}{\sum_{\boldsymbol{o}\in kNN(\boldsymbol{x})} d_{reach}(\boldsymbol{x},\boldsymbol{o})}. \tag{18}$$

In this light, the LOF of a configuration $\boldsymbol{x}$ is defined as:

$$LOF(\boldsymbol{x}) = \frac{1}{k}\sum_{\boldsymbol{o}\in kNN(\boldsymbol{x})} \frac{lrd_k(\boldsymbol{o})}{lrd_k(\boldsymbol{x})}, \tag{19}$$

where the parameter k has to be set by the analyst (and it is not related to the number of clusters K identified in Section 3.3).

In general, a value of $LOF(\boldsymbol{x}) \approx 1$ indicates that the configuration $\boldsymbol{x}$ is well represented by the rest of the configurations, whereas a value of $LOF(\boldsymbol{x}) \gg 1$ indicates that the configuration $\boldsymbol{x}$ is isolated. In order to

have a reference value for detecting a critical configuration as unexplored, the LOF is evaluated for all critical configurations $\boldsymbol{x} \in \mathcal{X}_{exp}^{CR}$ (namely, LOF_{exp}). Likewise, LOF_{real} represents the random variables corresponding to the one-vs-all evaluations of the configurations $\boldsymbol{x} \in \mathcal{X}_{real}^{CR}$. A configuration $\boldsymbol{x} \in \mathcal{X}_{real}^{CR}$ is considered "unexplored", if $LOF(\boldsymbol{x}) > \overline{LOF}_{exp}$ where:

$$\overline{LOF}_{exp} = \max_{\boldsymbol{x} \in \mathcal{X}_{exp}^{CR}} LOF(\boldsymbol{x}) \tag{20}$$

is the LOF corresponding to the most isolated configuration explored.

The following distance-based statistics have been considered to synthetize the overall performance of the exploration method:

1. Expected LOF:

$$\mu_{LOF}^{real} = E[LOF_{real}] \tag{21}$$

   A value of $\mu_{LOF}^{real} \gg 1$ indicates that some CRs are probably unexplored.

2. Unexplored Critical Region (UCR):

$$UCR = \frac{\#\left(LOF_{real} > \overline{LOF}_{exp}\right)}{\#\mathcal{X}_{real}^{CR}} \tag{22}$$

   which is the ratio between the number of real critical configurations identified as unexplored and the cardinality of $\mathcal{X}_{real}^{CR}$. In practice, it represents the "fraction" of CR that have not been explored by the method.

3. Unexplored Extreme Critical Region (UECR):

$$UECR_{\gamma\%} = UCR_{\gamma\%} | \mathcal{X}_{real}^{ECR} = \frac{\#\left(LOF_{real} > \overline{LOF}_{exp} | \mathcal{X}_{real}^{ECR}\right)}{\#\mathcal{X}_{real}^{ECR}} \tag{23}$$

   where $\mathcal{X}_{real}^{ECR} \subset \mathcal{X}_{real}^{CR}$ is the subset of critical configurations leading to the most "extreme" outputs. In particular, $\gamma \in [0,100]\%$ is the quantile used to characterize the extreme outputs: letting $\gamma = 0.9$, then a critical configuration is considered "extreme" if its output is larger than the output of 90% of the population. This metric allows the analyst to understand whether the method has discovered the CRs leading to the most critical outputs.

4. Conditional Expected LOF:

$$\mu_{LOF|UCR} = E\left[\frac{LOF_{real}}{\overline{LOF}_{exp}} | LOF_{real} > \overline{LOF}_{exp}\right] \tag{24}$$

   that indicates how much isolated are on average the unexplored critical configurations with respect to the most isolated critical configuration explored. In practice, values of $\mu_{LOF|UCR} \gg 1$ indicate the presence of critical configurations that are very isolated from the explored CRs and, thus, warn the analyst on the presence of CRs disconnected from those already identified.

5 Case Study

In this section, a power distribution network is analyzed, under the Direct Current (DC) approximation (Purchala et al., 2005), in order to explore and discover possible critical scenarios (Mena et al., 2014). The network, represented in Fig. 5, is composed of 10 feeders transporting energy from a unique Main Source *MS* to 8 demanding nodes (consumers) characterized by different daily load profiles. The simplified model considers only the active power flows, neglecting power losses, and assumes a constant value of the voltage throughout the network. The main source is assumed to be always capable of satisfying any demand and no constraints on the feeders capacity are defined.

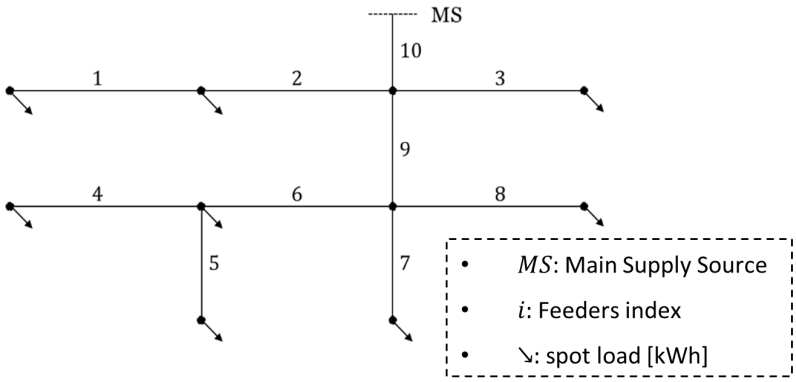

Fig. 5 Power network configuration.

The load profiles L_j assume different shapes according to the corresponding type of consumers associated. These include residential consumers and offices whose per unit (p.u.) daily spot load profiles are reported in Fig. 6. In detail, the daily load L_j of a demanding node is given by:

$$L_j(t) = r_j R(t) + o_j O(t) \tag{25}$$

where $R(t)$ and $O(t)$ are the p.u. daily loads, whereas r_j and o_j are the corresponding average loads for the residential consumers and office, respectively (Jardini et al., 2000). The values of the average loads used in this paper are reported in Table I. Uncertainty and seasonality effects on the average loads can be easily embedded into the model. Nevertheless, since the focus of the study is on the exploration of the daily profiles to verify the impact of feeder failures, they are not taken into account in this paper.

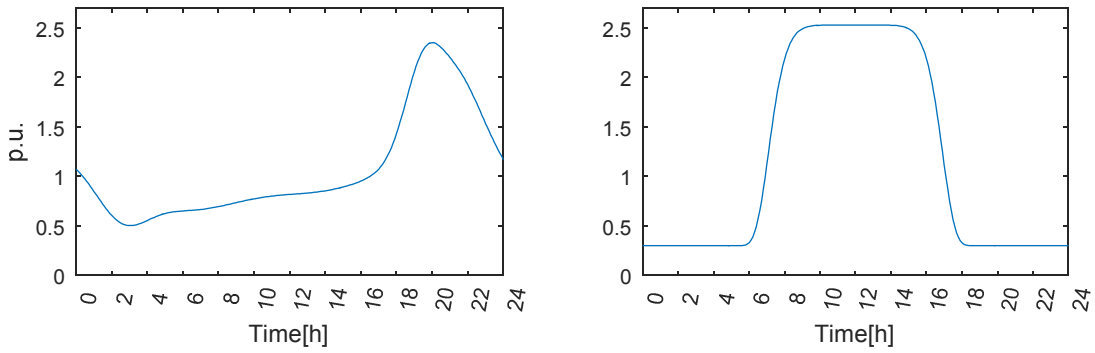

Fig. 6 Power per unit load profiles for a residential consumer (left) and for a commercial office (right).

Table I Average load values for the 8 nodes of the network in kW.

NODE	1	2	3	4	5	6	7	8
r	0	0	0	1	1	5	5	5
o	5	5	100	0	0	0	0	0

We assume that each feeder i can independently fail only once within the 24 hour, at a random time $T_i \in [0,24)$ and with associated magnitude of the failure F_i. When the *i*-th feeder fails, no power can flow through it for a time proportional to the magnitude of the failure: for example, $F_i = 0.5$ means that the feeder is out of service for half an hour. In this view, $\boldsymbol{X} = [T_1, \dots, T_{10}, F_1, \dots, F_{10}]$ is the M-dimensional vector of the inputs to the model and represents a given failure configuration.

The electrical Energy Not Served (ENS) to the consumers is considered as output of the model and it is defined in this case as:

$$ENS(\boldsymbol{X}) = \int_0^{24} \sum_{j=1}^{8} \mathbf{1}_{NSS(t)}(j) \cdot L_j(t)\, dt, \tag{26}$$

where $NSS(t)$ indicates the Not Supplied Set at time t, i.e., the set of nodes that are not served at time t and $\mathbf{1}$ is the indicator function which takes value 1 if $j \in NSS(t)$, and 0 otherwise. Moreover, ENS is used to discriminate the critical conditions, i.e., a value of $ENS(\boldsymbol{X}) \geq ENS_{thres}$ implies that the failure configuration $\boldsymbol{X}$ is critical; otherwise, $\boldsymbol{X}$ is considered as "normal". The value of ENS_{thres} is set equal to 500 kWh, in order to focus the attention on critical events.

5.1 Dimensionality Reduction

For the dimensionality reduction step, we resort to PCE, where the maximum degree of the polynomials is fixed to 5 in order to reduce the computational cost and focus the attention on the main trend of the model.

The coefficients of the PCE are estimated by Least Angle Regression on the basis of a DOE of 500 samples obtained with a QMC Sobol' sequence. Fig. 7 shows that there is a huge difference between the total order indices S_T of the inputs: those associated to feeders 3 and 10 (i.e., T_3, T_{10}, F_3, F_{10}) take values larger than 0.2, whereas the others take values lower than 0.05. This is in accordance with the fact that feeders 3 and 10 are the only two that can affect the energy supplied to the most demanding consumer.

In this light, the dimensionality of the reduced-model is set to 4 with $\boldsymbol{X}^* = (T_3, T_{10}, F_3, F_{10})$, and the rest of the factors are set to randomly fixed values, since they are expected to have no effect on the output. Notice that alternatively, they can be set to values expected to lead to the worst outputs, i.e., the magnitudes of the failures at their maximum values and the corresponding failure times within the time window of high load: this would lead to higher values of ENS and, thus, to a more conservative exploration.

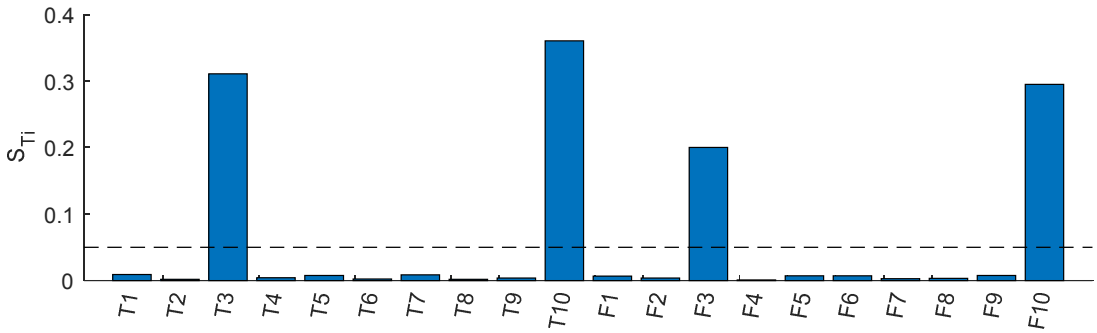

Fig. 7 Sobol' total order indices for the 20 input factors.

Finally, it must be observed that two main elements contribute to the importance of a feeder in the network: i) its topological position and ii) the demand of the nodes that rely on it. For example, if the highest demand were associated to node 7 instead of node 3, feeder 10 would still remain the most important (since its failure prevents all nodes to be served), whereas we would expect feeders 9 and 7 being more important than feeder 3. In this view, it must be observed that the results are conditioned to the average demands reported in Table I.

5.2 Meta-modeling

For training the meta-model we resort to an ordinary kriging, i.e., the trend is assumed to be unknown but constant, which allows the Gaussian process to completely adapt to the training data. An ellipsoidal anisotropic correlation function is used to take into account possible different behaviors of the response function with respect to different factors: in particular, we resort to the 3/2 Matérn one (Abramowitz & Stegun, 1964; Rasmussen & Williams, 2006):

$$h(x,x';\,\boldsymbol{\theta}) = \sqrt{\sum_{m\in M'}\left(\frac{x_m - x'_m}{\theta_m}\right)^2}$$
$$R\left(h, v=\frac{3}{2}\right) = \left(1+h\sqrt{3}\,\right)\cdot e^{-h\sqrt{3}} \tag{27}$$

where v is the shape parameter and $\boldsymbol{\theta}$ the scale one.

Given the dimensionality of the reduced model, 100 configurations sampled with a Sobol' QMC and the corresponding ENS are used for initializing the meta-model. Then, through the iterative AK-MCS introduced in Section 3.2, 10000 configurations are sampled by means of LHS and a maximum of 50 candidate configurations are evaluated and added to the DOE $\{\boldsymbol{X}_{krig}, \boldsymbol{Y}_{krig}\}$ at each step. Only configurations having a value of the U-function lower than 4 are eligible as candidates. Actually, $U(\boldsymbol{x}) > 4$ indicates that the corresponding configuration is, in a probabilistic view, very distant from the critical threshold. A maximum number of 1000 I/O observations for training the meta-model is set in order to limit the maximum computational effort. Fig. 8 shows the projection on the two-dimensional subspace $[T_3, T_{10}]$ of the configurations used to train the meta-model: on the left panel, we report the initial 100 samples used for the initialization, whereas on the right, the 900 sample added iteratively by the AK-MCS are shown. It is worth noticing how the adaptive DOE distributes the observations differently in the different portions of the input domain (i.e., a significantly higher density in the CRs).

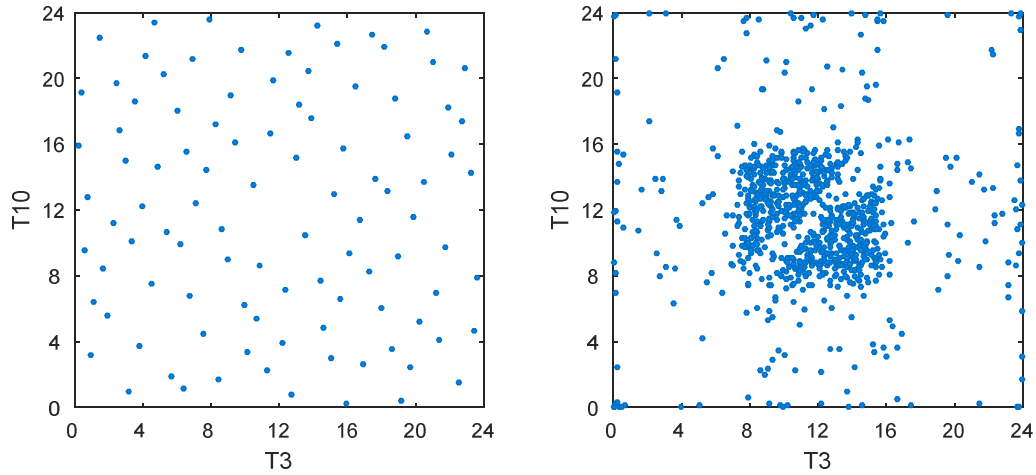

Fig. 8 Projection of the DOE used for training the meta-model. The figure on the left shows the initial 100 Sobol' QMC samples, whereas on the right those added by the AK-MCS.

5.3 Deep Exploration

From the Kriging DOE, 169 configurations are identified as critical. In order to deeply explore the CR, 5 iterations of the method proposed in Section 3.3 are run with 5 Markov chains and a maximum number of samples equal to 5000. Fig. 9 shows the projections on the two-dimensional subspace $[T_3, T_{10}]$ of the

configurations belonging to the CR. The left panel reports the configurations available from the meta-model DOE, whereas that on the right contains those obtained as a result of the deep exploration (~3000 configurations). It is worth noticing that the deep exploration allows better highlighting the boundaries of the CRs and, thus, to better retrieve their shapes and characteristics. This is even more apparent in high-dimensional spaces. Only one projection of the CRs configurations is here reported for brevity; nevertheless, a detailed analysis is given in the following sections.

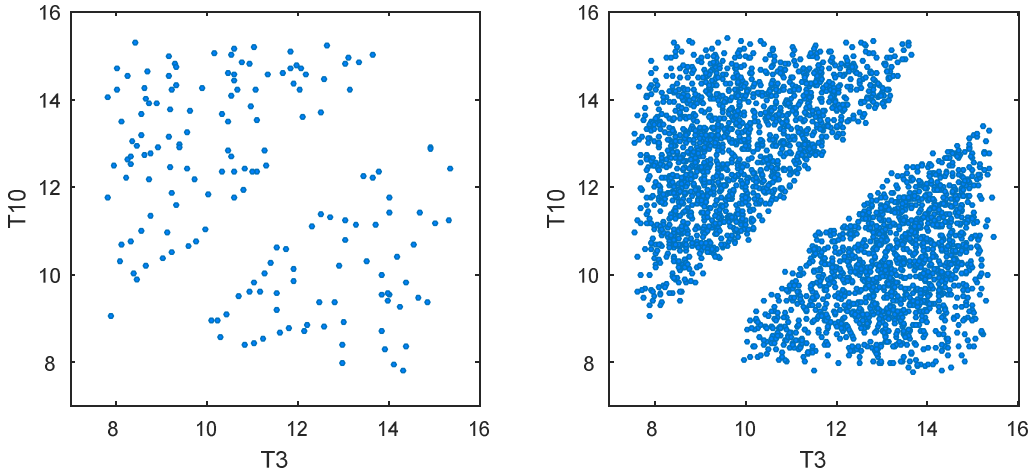

Fig. 9 Two-dimensional projections of the observations belonging to the CRs: those available from the DOE of the meta-model (left) and those obtained with the deep exploration step (right).

5.4 Representation and Information Retrieval

A sequence of k-means clusterings with different cluster cardinality (from K=1 to 10) is applied to the critical configurations for identifying the more representative number of separate CRs. Several cluster validity indices (e.g., Hubert statistic, Dunn, Silhouette, Davies and Bouldin, Calinski and Harabasz indices, etc.) have been computed to this aim; however since this analysis goes beyond the scope of the present paper, the reader is referred to (Arbelaitz et al., 2013; Charrad, Ghazzali, Boiteau, & Niknafs, 2014) for details. Two clusters have been identified and the corresponding PCP is reported in Fig. 10. For the sake of clarity, the envelopes of the parallel coordinates representing the two clusters (i.e., the ranges of values characterizing the clusters) are shown in Fig. 11. By observing these ranges, it is also possible to have an idea of the dimension of the CRs. In this case, for example, they occupy respectively around the (30%, 30%, 20%, 20%) of the entire range of the four important factors T_3, T_{10}, F_3 and F_{10}, which corresponds to ~0.36% of the entire input domain. The CRs are characterized by failures occurring during the central hours of the day (between 8-15) and with a failure magnitude above the 0.8, i.e. the feeders are out of order for at least 48min each. In addition, it is worth noticing that the two clusters show different behaviors on the two axes corresponding to the failure times, i.e., T_3 and T_{10}.

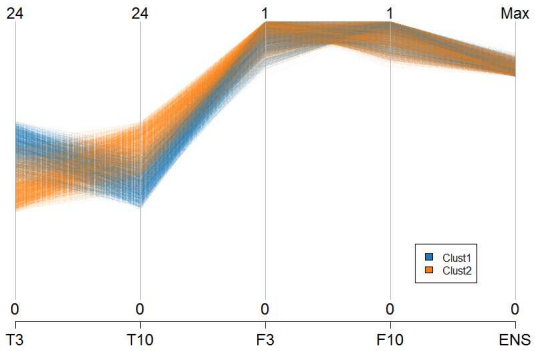

Fig. 10 PCP of the two CRs identified.

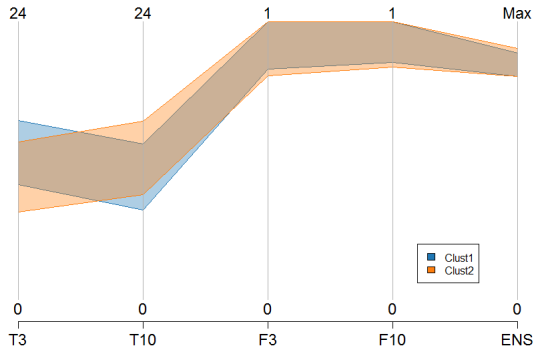

Fig. 11 Envelopes of the PCP representing the factors ranges.

For this reason, the corresponding SPOLM is given in Fig. 12, where the "envelopes" identified on the PCP are represented in the panels above the diagonal by means of shadowed rectangles. It can be observed that the two clusters are recognizable and well separated on the subspace defined by $[T_3, T_{10}]$: cluster 1 is characterized by an initial failure of feeder 10 followed by a failure of feeder 3 with a delay of at least one hour, whereas cluster 2 is characterized by the inverse sequence, still with a delay of at least one hour between failures. Indeed, if both failures happen at the same time, the ENS associated to node 3 is the same as if only one of the two failures had happened, because both feeders are put under repair at the same time and, thus, the total time of energy not supplied to user 3 is "just" one hour.

Concerning the subspace defined by $[F_3, F_{10}]$, it must be noticed that there is no difference between the two clusters. However, the triangular shape of the region shows that the sum of the two failure magnitudes must be at least equal to 1.80, i.e., the consumer at node 3 is not served for at least 1h:48m.

Finally, although the two-dimensional projections of the PCP envelopes overestimate the regions of the associated CRs, they provide a synthetic representation, which can be useful as first approximation of the CRs.

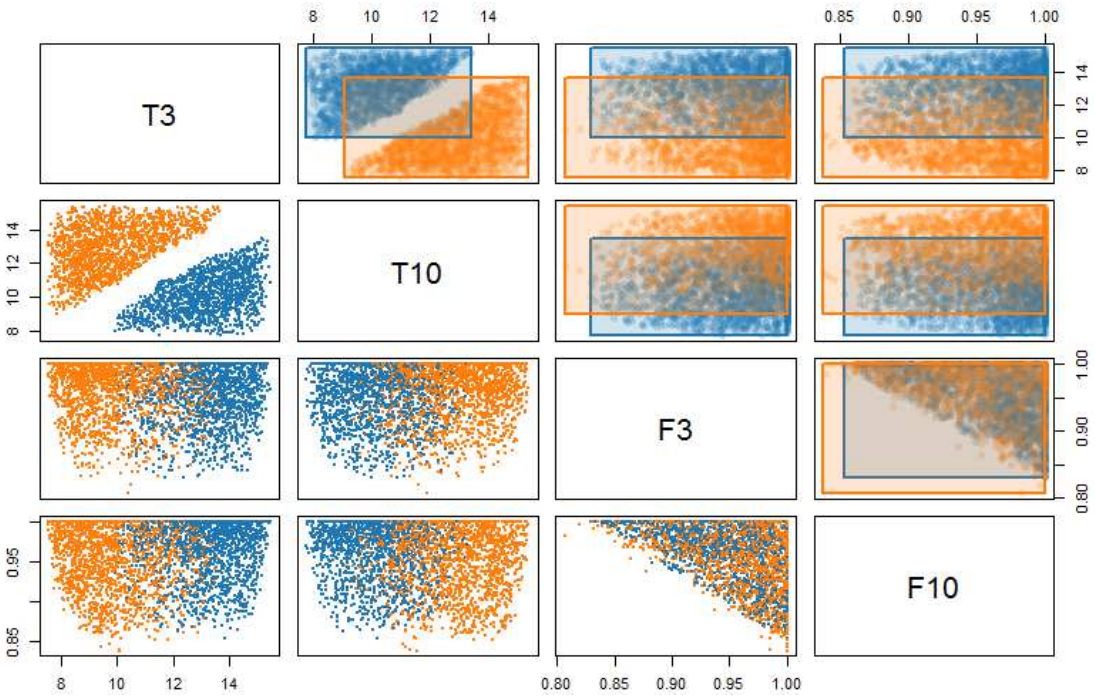

Fig. 12 SPLOM of the two CRs discovered by the exploration algorithm. Above the diagonal, the projections of the PCP envelopes are depicted by means of shadowed rectangles.

5.5 Performance Assessment

In order to have a representative picture of the real CRs, a large number of configurations involving all 20 input factors of the model have been sampled by means of LHS and the corresponding output has been evaluated. Moreover, the outputs of the reduced model involving the projections of the 20 factors on the 4-dimensional space defined by $[T_3, T_{10}, F_3, F_{10}]$ have been evaluated as representative of the ideal "target", meta-model representation. The number of calls to the expensive model and/or to the cheap one (i.e., the meta-model) is given in Table II for each exploration strategy.

Table II Number of calls made to the computationally cheap and/or expensive model for the different exploration strategies.

COPUTATIONAL COST	META-MODEL	REDUCED-MODEL	REAL-MODEL
CHEAP	~200000	0	0
EXPENSIVE	1500	100000	100000

Among the large number of configurations sampled, those leading to critical values of ENS are selected and the corresponding LOF evaluated to verify to what extent the CRs discovered by the meta-model are similar to those found by the reduced and real-models (see Section 4). The values of the associated statistics are given in Table III. The CRs of the meta-model are used as the reference set, thus, only the corresponding

expected value of the LOF can be evaluated. By looking at the results obtained for the reduced-model, it must be observed that all the statistics assume low values: the average value of LOF is very close to that of the meta-model; the percentage of CR that remains unexplored is only 3%, and the associated conditional value is still very low (i.e., 1.08), which means that the unexplored CRs are very close to the boundaries of the CRs identified by the meta-model. In this light, it can be stated that the meta-model exploration has accurately explored and discovered the CRs associated to the reduced model.

On the other side, with respect to the real model, the average LOF takes a large value compared to the meta-model, suggesting that a part of the CRs remains unexplored. This is confirmed by the percentage of unexplored CR. However, it must be noticed that the percentage of unexplored extreme CR is very low, i.e., the meta-model exploration has been able to identify the configurations leading to the most critical outputs. Finally, the conditional expected value $\mu_{LOF|UCR}$ takes a value that is not very large, suggesting that the unexplored portion of CRs is likely to be close to the boundaries.

Table III Local Outlier Factor (LOF) based statistics for the different exploration strategies.

METRIC	META-MODEL	REDUCED-MODEL	REAL-MODEL
$\boldsymbol{\mu_{LOF}}$	1.02	1.03	2.66
$\boldsymbol{UCR}$	-	3%	72%
$\boldsymbol{UECR_{90\%}}$	-	0%	7%
$\boldsymbol{\mu_{LOF|UCR}}$	-	1.08	2.20

In order to visualize the results, we resort to a SPLOM where the CRs identified by the meta-model exploration are depicted by light circles and the configuration belonging to the CRs associated to the real model are depicted by crosses and squares according to their values of LOF. In particular, in accordance with Section 4: those configurations having $LOF \leq \overline{LOF}_{exp}$ (see Eq. (20)) are defined as identified CR (crosses), whereas those having $\overline{LOF}_{exp} < LOF$ are defined as undiscovered CR (squares). It must be noticed that there is not a significant difference between the Meta-Model (MM)-based and the real-model-based exploration in the subspace characterized by the failure times $[T_3, T_{10}]$. On the contrary, there is a significant difference in the failure magnitude subspace $[F_3, F_{10}]$: according to the real model, it is enough that the sum of the magnitudes is larger than ~1.60. This means that the real model can reach a critical condition even if the consumer at node 3 is not served for at least 1h:36m. Indeed, the rest of the ENS needed to reach the critical threshold can come from the failures of the feeders discarded during the dimensionality reduction step. Finally, by looking at the last column of Fig. 13, it can be seen that the largest values of ENS, i.e., the most critical ones, are correctly discovered by our methodology (crosses).

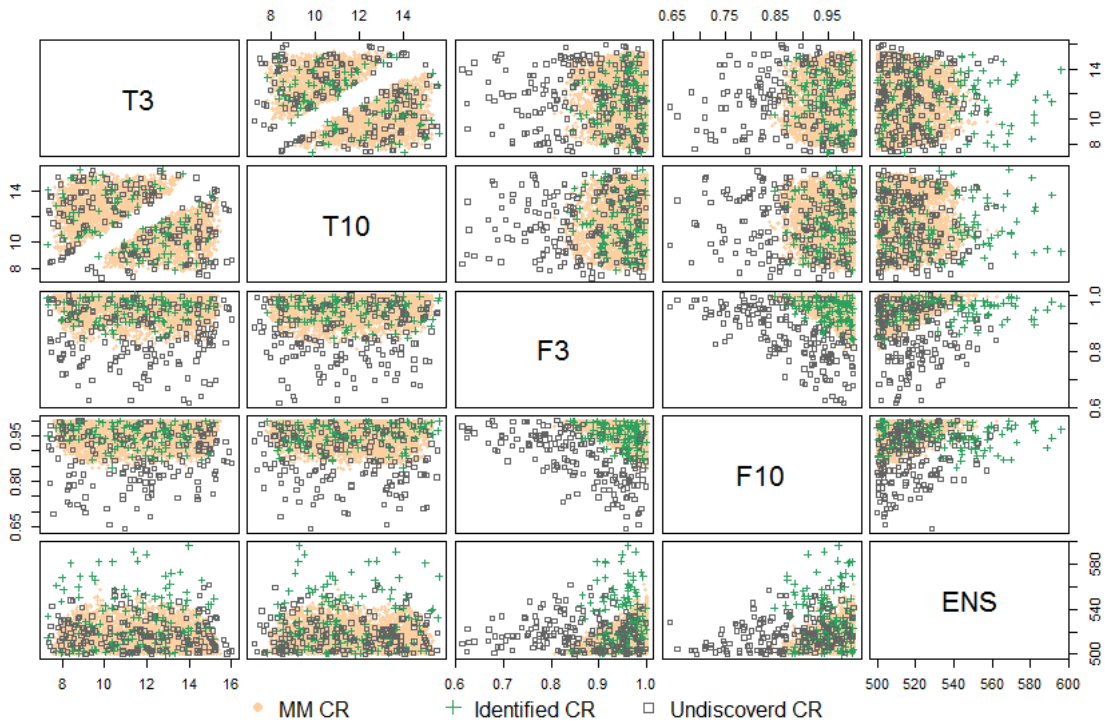

Fig. 13 SPLOM of the CRs discovered by the Meta-Model (MM) exploration (light circles. The CR of the real model are depicted with different symbols whether identified (cross) or not (square).

A sort of sensitivity analysis to the model parameters has also been conducted to verify the performance of the proposed methodology when the impacts of the discarded factors is very low, i.e., when the reduced-model is likely to represent the real model. To this aim, all the loads except that of node 3 have been reduced of a factor 10 (the corresponding values are reported in Table IV). In order to assure the presence of a CR despite the loading reduction, the threshold ENS_{thres} has been set equal to 475 kWh, i.e. 5% lower than the initial one. All the analyses have been run with the same settings and with the same number of calls to the model as in the initial case.

Table IV Average load values for the 8 nodes of the network in kW.

NODE	1	2	3	4	5	6	7	8
r	0	0	0	0.1	0.1	0.5	0.5	0.5
o	0.5	0.5	100	0	0	0	0	0

Table V reports the result of the statistics associated to the LOF for the reduced and the real model-based exploration. The average value of the LOF is for all types of exploration very close to 1, indicating that it is likely that all CRs have been discovered. This is confirmed by the percentage of unexplored CR, which is null for both models. The value of $\boldsymbol{\mu_{LOF|UCR}}$ is not reported, since no configuration has been identified as unexplored.

Table V Local Outlier Factor (LOF)-based statistics for the different exploration strategies.

METRIC	META-MODEL	REDUCED-MODEL	REAL-MODEL
μ_{LOF}	1,02	1,01	1,07
UCR	-	0	0
$UECR_{90\%}$	-	0	0

Fig. 14 shows that all critical configurations discovered by means of the real model-based exploration lay inside or at the boundaries of the CRs discovered by our methodology (dark crosses). These results demonstrate how the proposed methodology is capable of identifying the CRs resorting to a limited number of calls to the real model: in this case, two orders of magnitude lower than the exploration based on the real-model.

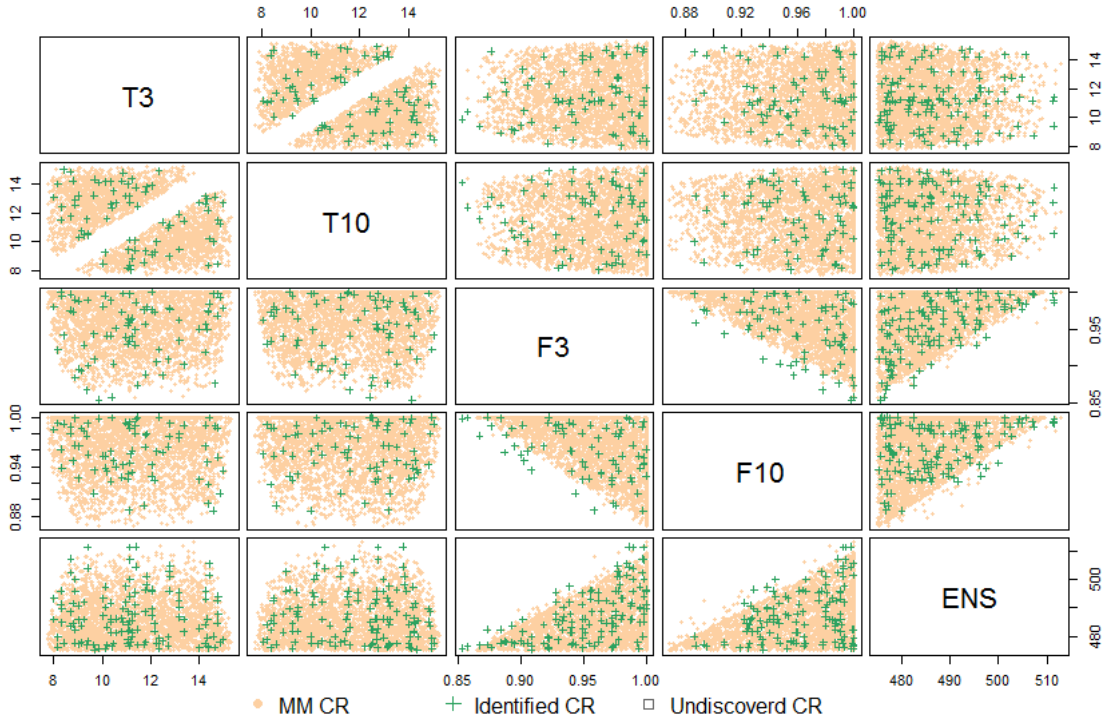

Fig. 14 SPLOM of the CRs discovered by the Meta-Model (MM) exploration (light circles. The CRs of the real model are depicted with different symbols, whether identified (cross) or not (square).

6 Conclusions

Knowing the possible behavior of a safety-critical, engineered system under different setting of the influencing (operational and environmental) factors is of paramount importance for improving the safety of the system. In particular, discovering configurations of factors that lead the system to critical, dangerous

conditions allows prevention and preparation. To this aim, in this paper, we have proposed an adaptive exploration framework for discovering and characterizing the Critical Regions (CRs) of a system, whose numerical model is: i) computational expensive, ii) high-dimensional, iii) complex.

The proposed methodology resorts to: *i*) a dimensionality reduction technique relying on a PCE-based sensitivity analysis for identifying the subspace that most characterizes the system behavior; *ii*) a meta-model namely Kriging for reproducing the system behavior while reducing the computational cost to run a simulation; *iii*) an adaptive exploration technique employing an adaptive MCMC algorithm for collecting information about the CRs; *iv*) clustering (i.e., k-means) and high-dimensional data visualization methods (i.e., PCP and SPLOM), for defining and characterizing the shape of the CRs of the system.

For exemplification, the method has been used for exploring the response of a power network model to several types of accidents involving up to 20 input factors. The method has been shown capable of effectively identifying the failure times and magnitudes leading the system to the most critical state, i.e., the one with the largest quantity of energy not supplied, indicating two specific failures sequences as the most dangerous ones.

The main benefit of the proposed method is the capability of exploring and retrieving useful information by resorting to a limited number of calls to the real (computationally expensive) model. This has been demonstrated by comparison with another exploration strategy based on LHS showing a computational saving of two orders of magnitudes. Furthermore, the framework has turned out to be very versatile by allowing the analyst to possibly select only some of the steps proposed and focusing on the problem requirements: for example, if the real model is not high-dimensional, the dimensionality reduction step can be avoided.

Although the method has been designed for and proved on a model that has a scalar output, it can also be applied to models having a functional output, such as the time varying trajectories of a dynamic system (e.g., the core temperature of a nuclear reactor in accidental conditions). In this case, an additional step for extracting some (possibly few) scalar features characterizing the trajectories would be needed at the beginning of the analysis.

Finally, the proposed method inherits the limitations of the techniques employed. In particular, the results of the PCE-based sensitivity depends on the capability of the PCE of capturing the global behavior of the response function, and the Kriging performance usually decreases with high-dimensional inputs. In this light, the method is expected to be efficient as long as the number of important variables identified by the dimensionality reduction step is sufficiently low to be managed by a Kriging meta-model.

7 References

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables* (Vol. 55): Courier Corporation.

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). *On the surprising behavior of distance metrics in high dimensional space*: Springer.

Aldemir, T. (2013). A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants. *Annals of Nuclear Energy, 52*, 113-124. doi:DOI 10.1016/j.anucene.2012.08.001

Alfonsi, A., Rabiti, C., Mandelli, D., Cogliati, J., Kinoshita, R., & Naviglio, A. (2015). *RAVEN and dynamic probabilistic risk assessment: Software overview.* Paper presented at the Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference, ESREL 2014.

Alfonsi, A., Rabiti, C., Mandelli, D., Cogliati, J. J., Wang, C., Maljovec, D. P., . . . Smith, C. L. (2016). *RAVEN Theory Manual*. Retrieved from Idaho Falls, ID:

Andrieu, C., & Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing, 18*(4), 343-373. doi:DOI 10.1007/s11222-008-9110-y

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition, 46*(1), 243-256.

Aven, T. (2015). Implications of black swans to the foundations and practice of risk assessment and management. *Reliability Engineering & System Safety, 134*, 83-91. doi:http://dx.doi.org/10.1016/j.ress.2014.10.004

Aven, T. (2016a). Ignoring scenarios in risk assessments: Understanding the issue and improving current practice. *Reliability Engineering & System Safety, 145*, 215-220. doi:http://dx.doi.org/10.1016/j.ress.2015.08.012

Aven, T., & Krohn, B. S. (2014). A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety, 121*, 1-10. doi:http://dx.doi.org/10.1016/j.ress.2013.07.005

Bier, V. M., Haimes, Y. Y., Lambert, J. H., Matalas, N. C., & Zimmerman, R. (1999). A survey of approaches for assessing and managing the risk of extremes. *Risk Analysis, 19*(1), 83-94.

Blatman, G., & Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics, 230*(6), 2345-2367.

Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research, 248*(3), 869-887. doi:http://dx.doi.org/10.1016/j.ejor.2015.06.032

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). *LOF: identifying density-based local outliers.* Paper presented at the ACM sigmod record.

Burnaev, E., Panin, I., & Sudret, B. (2016). Effective Design for Sobol Indices Estimation Based on Polynomial Chaos Expansions *Conformal and Probabilistic Prediction with Applications* (pp. 165-184): Springer International Publishing.

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

Cavalcante, H. L. D. S., Oriá, M., Sornette, D., Ott, E., & Gauthier, D. J. (2013). Predictability and Suppression of Extreme Events in a Chaotic System. *Physical Review Letters, 111*(19), 198701.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *2014, 61*(6), 36. doi:10.18637/jss.v061.i06

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., & Charrad, M. M. (2014). Package 'NbClust'. *J. Stat. Soft, 61*, 1-36.

Chen, V. C. P., Tsui, K.-L., Barton, R. R., & Meckesheimer, M. (2006). A review on design, modeling and applications of computer experiments. *IIE Transactions, 38*(4), 273-291. doi:10.1080/07408170500232495

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., & Richet, Y. (2014). Fast Parallel Kriging-Based Stepwise Uncertainty Reduction With Application to the Identification of an Excursion Set. *Technometrics, 56*(4), 455-465. doi:10.1080/00401706.2013.860918

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *American Statistician, 49*(4), 327-335. doi:Doi 10.2307/2684568

Ciriello, V., Di Federico, V., Riva, M., Cadini, F., De Sanctis, J., Zio, E., & Guadagnini, A. (2013). Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly

heterogeneous aquifer. *Stochastic Environmental Research and Risk Assessment, 27*(4), 945-954. doi:10.1007/s00477-012-0616-7

Cook, D., & Swayne, D. F. (2007). *Interactive and dynamic graphics for data analysis: with R and GGobi*: Springer Science & Business Media.

Crestaux, T., Le Maître, O., & Martinez, J. M. (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering and System Safety, 94*(7), 1161-1172. doi:10.1016/j.ress.2008.10.008

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

Echard, B., Gayton, N., & Lemaire, M. (2011). AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety, 33*(2), 145-154. doi:http://dx.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.

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

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

Hartigan, J. A. (1975). Printer graphics for clustering. *Journal of Statistical Computation and Simulation, 4*(3), 187-213. doi:10.1080/00949657508810123

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety, 52*(1), 1-17. doi:Doi 10.1016/0951-8320(96)00002-6

Iman, R. L. (2008). Latin Hypercube Sampling *Encyclopedia of Quantitative Risk Analysis and Assessment*: John Wiley & Sons, Ltd.

Inselberg, A. (2009). *Parallel coordinates*: Springer.

Izquierdo, J. M., Hortal, J., Sánchez, M., Meléndez, E., Herrero, R., Gil, J., . . . Rodríguez, G. (2009). *SCAIS (Simulation code system for integrated safety assessment): Current status and applications.* Paper presented at the Safety, Reliability and Risk Analysis: Theory, Methods and Applications - Proceedings of the Joint ESREL and SRA-Europe Conference.

Jardini, J. A., Tahan, C., Gouvea, M., Ahn, S. U., & Figueiredo, F. (2000). Daily load profiles for residential, commercial and industrial low voltage consumers. *Power Delivery, IEEE Transactions on, 15*(1), 375-380.

Jin, R., Chen, W., & Simpson, T. W. (2001). Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization, 23*(1), 1-13. doi:10.1007/s00158-001-0160-4

Kernstine, K. H. (2013). Inadequacies of Traditional Exploration Methods in Systems-of-Systems Simulations. *IEEE Systems Journal, 7*(4), 528-536. doi:10.1109/JSYST.2013.2252864

Kleijnen, J. P. C. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research, 192*(3), 707-716. doi:http://dx.doi.org/10.1016/j.ejor.2007.10.013

Kröger, W., & Zio, E. (2011). *Vulnerable systems*: Springer Science & Business Media.

Li, J. H., Kang, R., Mosleh, A., & Pan, X. (2011). Simulation-based automatic generation of risk scenarios. *Journal of Systems Engineering and Electronics, 22*(3), 437-444. doi:DOI 10.3969/j.issn.1004-4132.2011.03.011

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

Liu, S., Maljovec, D., Wang, B., Bremer, P. T., & Pascucci, V. (2017). Visualizing High-Dimensional Data: Advances in the Past Decade. *IEEE Transactions on Visualization and Computer Graphics, 23*(3), 1249-1268. doi:10.1109/TVCG.2016.2640960

Maljovec, D., Liu, S., Wang, B., Mandelli, D., Bremer, P. T., Pascucci, V., & Smith, C. (2016). Analyzing simulation-based PRA data through traditional and topological clustering: A BWR station blackout case study. *Reliability Engineering & System Safety, 145*, 262-276. doi:http://dx.doi.org/10.1016/j.ress.2015.07.001

Maljovec, D., Wang, B., Pascucci, V., Bremer, P. T., & Mandelli, D. (2013). *Adaptive sampling algorithms for probabilistic risk assessment of nuclear simulations.* Paper presented at the International Topical Meeting on Probabilistic Safety Assessment and Analysis 2013, PSA 2013.

Maljovec, D., Wang, B., Pascucci, V., Bremer, P. T., Pernice, M., Mandelli, D., & Nourgaliev, R. (2013). *Exploration of high-dimensional scalar function for nuclear reactor safety analysis and visualization.* Paper presented at the International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering, M and C 2013.

Mandelli, D., Smith, C., Rabiti, C., Alfonsi, A., Youngblood, R., Pascucci, V., . . . Zamalieva, D. (2013). *Dynamic PRA: An overview of new algorithms to generate, analyze and visualize data.* Paper presented at the Transactions of the American Nuclear Society.

Mandelli, D., Yilmaz, A., Aldemir, T., Metzroth, K., & Denning, R. (2013). Scenario clustering and dynamic probabilistic risk assessment. *Reliability Engineering & System Safety, 115*, 146-160. doi:http://dx.doi.org/10.1016/j.ress.2013.02.013

Marelli, S., & Sudret, B. (2014). *UQLab: a framework for uncertainty quantification in MATLAB.* Paper presented at the International Conference on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool (United Kingdom).

Matheron, G. (1963). Principles of geostatistics. *Economic geology, 58*(8), 1246-1266.

McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics, 42*(1), 55-61.

Mena, R., Hennebel, M., Li, Y.-F., Ruiz, C., & Zio, E. (2014). A risk-based simulation and multi-objective optimization framework for the integration of distributed renewable generation and storage. *Renewable and Sustainable Energy Reviews, 37*, 778-793.

Pate-Cornell, E. (2002). Finding and fixing systems weaknesses: probabilistic methods and applications of engineering risk analysis. *Risk Analysis, 22*(2), 319-334.

Paté-Cornell, E. (2012). On "Black Swans" and "Perfect Storms": risk analysis and management when statistics are not enough. *Risk Analysis, 32*(11), 1823-1833.

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., & Kim, N.-H. (2010). Adaptive Designs of Experiments for Accurate Approximation of a Target Region. *Journal of Mechanical Design, 132*(7), 071008-071008. doi:10.1115/1.4001873

Purchala, K., Meeus, L., Van Dommelen, D., & Belmans, R. (2005). *Usefulness of DC power flow for active power flow analysis.* Paper presented at the 2005 IEEE Power Engineering Society General Meeting.

Queral, C., Mena-Rosell, L., Jimenez, G., Sanchez-Perea, M., Gomez-Magan, J., & Hortal, J. (2016). Verification of SAMGs in SBO sequences with Seal LOCA. Multiple damage domains. *Annals of Nuclear Energy, 98*, 90-111. doi:http://dx.doi.org/10.1016/j.anucene.2016.07.021

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*: the MIT Press.

Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer.

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics, 18*(2), 349-367. doi:DOI 10.1198/jcgs.2009.06134

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics, 23*(3), 470-472.

Saltelli, A. (2008). *Global sensitivity analysis : the primer*. Chichester, England ; Hoboken, NJ: John Wiley.

Santner, T. J., Williams, B. J., & Notz, W. (2003). *The Design and analysis of computer experiments*. New York: Springer.

Schöbi, R., Sudret, B., & Marelli, S. (2016). Rare Event Estimation Using Polynomial-Chaos Kriging. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, D4016002.

Shan, S., & Wang, G. G. (2010). Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Structural and Multidisciplinary Optimization, 41*(2), 219-241. doi:10.1007/s00158-009-0420-2

Simpson, T. W., Poplinski, J. D., Koch, N. P., & Allen, J. K. (2001). Metamodels for Computer-based Engineering Design: Survey and recommendations. *Engineering with Computers, 17*(2), 129-150. doi:10.1007/pl00007198

Smith, C., Rabiti, C., Martineau, R., & Szilard, R. (2016). *Risk Informed Safety Margins Characterization (RISMC) Pathway Technical Program Plan*. Retrieved from Idaho Falls, ID:

Sobol. (1976). Uniformly distributed sequences with an additional uniform property. *USSR Computational mathematics and mathematical physics, 16*(5), 236-242.

Sobol. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation, 55*(1-3), 271-280. doi:Doi 10.1016/S0378-4754(00)00270-6

Sobol, Asotsky, D., Kreinin, A., & Kucherenko, S. (2011). Construction and Comparison of High ‐ Dimensional Sobol'Generators. *Wilmott, 2011*(56), 64-79.

Sornette, D. (2009). Dragon-kings, black swans and the prediction of crises. *Swiss Finance Institute Research Paper*(09-36).

Sornette, D., Maillart, T., & Kröger, W. (2013). Exploring the limits of safety analysis in complex technological systems. *International Journal of Disaster Risk Reduction, 6*, 59-66. doi:http://dx.doi.org/10.1016/j.ijdrr.2013.04.002

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety, 93*(7), 964-979.

Turati, P., Pedroni, N., & Zio, E. (2016a). An Adaptive Simulation Framework for the Exploration of Extreme and Unexpected Events in Dynamic Engineered Systems. *Risk Analysis, 37*(1), 147-159. doi:10.1111/risa.12593

Vaiman, M., Bell, K., Chen, Y., Chowdhury, B., Dobson, I., Hines, P., . . . Zhang, P. (2012). Risk Assessment of Cascading Outages: Methodologies and Challenges. *IEEE Transactions on Power Systems, 27*(2), 631-641. doi:10.1109/TPWRS.2011.2177868

Zio, E. (2014). Integrated deterministic and probabilistic safety assessment: Concepts, challenges, research directions. *Nuclear Engineering and Design, 280*, 413-419. doi:DOI 10.1016/j.nucengdes.2014.09.004

Zio, E. (2016a). Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety, 152*, 137-150. doi:10.1016/j.ress.2016.02.009

Zio, E. (2016b). Some Challenges and Opportunities in Reliability Engineering. *IEEE Transactions on Reliability, 65*(4), 1769-1782. doi:10.1109/TR.2016.2591504