Transient identification by clustering based on Integrated Deterministic and Probabilistic Safety Analysis outcomes

Francesco Di Maio, Matteo Vagnoli, Enrico Zio

TRANSIENT IDENTIFICATION BY CLUSTERING BASED ON INTEGRATED DETERMINISTIC AND PROBABILISTIC SAFETY ANALYSIS OUTCOMES

Francesco Di Maio[1], Matteo Vagnoli[1], Enrico Zio[1,2]

[1] *Energy Department, Politecnico di Milano*

Via Ponzio 34/3, 20133 Milano, Italy

francesco.dimaio@polimi.it

[2] *Chair on System Science and Energetic Challenge*

Foundation EDF – Electricite de France

Ecole Centrale, Paris, and Supelec, Paris, France

ABSTRACT

In this work, we present a transient identification approach that utilizes clustering for retrieving scenarios information from an Integrated Deterministic and Probabilistic Safety Analysis (IDPSA). The approach requires: i) creation of a database of scenarios by IDPSA; ii) scenario post-processing for clustering Prime Implicants (PIs), i.e., minimum combinations of failure events that are capable of leading the system into a fault state, and Near Misses, i.e., combinations of failure events that lead the system to a quasi-fault state; iii) on-line cluster assignment of an unknown developing scenario. In the step ii), we adopt a visual interactive method and risk-based clustering to identify PIs and Near Misses, respectively; in the on-line step iii), to assign a scenario to a cluster we consider the sequence of events in the scenario and evaluate the Hamming similarity to the sequences of the previously clustered scenarios. The feasibility of the analysis is shown with respect to the accidental scenarios of a dynamic Steam Generator (SG) of a NPP.

Keywords: Integrated Deterministic and Probabilistic Safety Analysis (IDPSA); Prime Implicants; Near Misses; on-line Clustering; Steam Generator.

1. INTRODUCTION

The safe operation of hazardous installations, such as Nuclear Power Plants (NPPs), depends on the capability of timely detecting possible accidental transients and promptly taking adequate actions to avoid catastrophic failures [Schirru et al., 2008]. Upon occurrence of an initiating failure event, it is important to predict whether the scenario that follows would lead to safe conditions or become an accidental scenario. In practice, this is done relying on the awareness of skilled operators who monitor and analyze recorded operational data of process variables, for early detection and diagnosis and, then, based on their own expert judgment follow the Emergency Operating Procedures (EOPs) and, if necessary, the Severe Accident Management Guidelines (SAMGs) to mitigate the scenario consequences. However, even for less dangerous accidental scenarios that do not lead to core damage but only to unplanned outage of production, it is sometimes difficult, if not impossible, for operators to promptly and accurately assess the plant and distinguish the occurring accidental scenario status simply by observing the large volume of operational data [Alaei et al., 2013]. For this reason, the decision process by the emergency management staff must be supported.

For such support, it is possible to devise automatic pattern recognition methods to predict the future evolution of a scenario initiated by a failure event. With this aim, we propose a novel method that combines post-processing of the outcomes of an Integrated Deterministic and Probabilistic Safety Analysis (IDPSA) and on-line clustering of data from the developing scenario.

We use Multiple-Valued Logic (MVL) theory for modeling the behavior of the system, accounting for the timing and order of occurrence of component failure events [Di Maio et al., 2015a].

Post-processing of the IDPSA results is performed for the: i) identification of the Prime Implicants (PI), i.e., those minimal sequences of failure events that are capable of leading the system into a fault state and cannot be covered by more general implicants [Quine, 1952], ii) identification of the Near Misses, i.e., those safe sequences of events that reach values of the safety parameters close to, but not exceeding, the corresponding acceptable thresholds [Zio et al., 2009].

In this work, we use a visual interactive method and a risk-based clustering method that have been shown effective for PI and Near Misses identification, respectively [Di Maio, 2014b; Di Maio et al., 2015].

For on-line identification of accidental transients, several methods have been presented in literature. Some of these are based on statistical techniques [Di Maio et al., 2013; Fink et al., 2015], which may have limitations with regards to the choice of parameters and difficulty in coping with noise in data

[Markou et al., 2003]; others, like neural networks and support vector machines [Basu et al., 1994; Palade et al., 2002; Widodo et al., 2007], require prior knowledge of the fault data set [Alaei et al., 2013]; and others are based on clustering by means of Euclidean metrics for measuring the similarity between transients [Schirru et al., 1999; Beringer et al., 2006; Collaghan et al., 2002] and fuzzy means [Zio et al., 2012; Baraldi et al., 2013].

In this paper, we develop an on-line clustering algorithm based on the Hamming distance [Hamming, 1950] to measure the similarity between developing transients and those obtained by IDPSA. At any instant of time, we compute the Hamming distance between the vector containing the event data of the developing accidental sequence with the vectors of the IDPSA post-processing scenarios, and identify the characteristics of the developing scenario as soon as any change in the trend of a process variable is detected. Finally, the developing transient is assigned to a cluster of safe scenarios, PIs, or, Near Misses, depending on its characteristics. In this way, we overcome the limitations of the methods already proposed in literature because i) the MVL approximation can be easily accommodated within a Hamming-based similarity definition (rather than using an Euclidean metric), ii) there is no need of additional efforts in tuning any parameter of the algorithm (as for the statistical techniques).

A case study is considered, regarding dynamic accidental scenarios occurring in the Steam Generator (SG) of a NPP [Aubry et al., 2012]. The paper is organized as follows. In Section 2, the SG model used to generate the scenarios for the dynamic reliability analysis is presented. In Section 3, a visual interactive method [Di Maio et al., 2015b] is applied for PIs identification, and, a risk-based Near Misses identification is performed. In Section 4, the on-line clustering method is introduced with reference to the case study considered. In Section 5, conclusions and remarks are given.

2. CASE STUDY

2.1 The U-Tube Steam Generator (UTSG) model

We consider a U-Tube Steam Generator (UTSG) (Fig. 1), part of the secondary circuit of a 900 MW Pressurized Water Reactor (PWR) [Aubry et al., 2012]. The improper control of the water level can be a major cause of this NPP unavailability [Kothare et al., 2000; Habibiyan et al., 2004]. The difficulties arises from non-minimum phase plant characteristics, i.e., plant strong inverse response behavior, particularly at low operating power, due to the so-called “swell and shrink” effects [Kothare et al., 2000].

The model and the parameters used serve the scope of mimicking the actual data of the real UTSG [Aubry et al., 2012]. A detailed model is, indeed, necessary for IDPSA because real data, necessary incomplete, would only partially cover the whole set of possible sequences of failure events and, therefore, endanger the identification of the set of PIs and Near Misses. Once the capability of the online identification clustering hereafter proposed is shown to be reliable with respect to the whole (simulated) set of accidental scenarios, we can be confident that its performance can be guaranteed on real (sparse) accidental scenarios, that, incidentally have already been classified by resorting to simulated scenarios.

The reactor coolant enters the UTSG at the bottom, moves upward and then downward in the inverted U-tubes, transferring heat to the secondary fluid before exiting at the bottom. The secondary fluid, the feedwater (Q_e), enters the UTSG at the top of the downcomer, through the space between the tube bundle wrapper and the SG shell. The value of Q_e is regulated by a system of valves: a low flow rate valve, used when the operating power (P_o) is smaller than 15% of nominal power (P_n), and a high flow rate valve when $P_o > 0.15\ P_n$ [Aubry et al., 2012]. In the secondary side of the tube bundle, water heats up, reaches saturation, starts boiling and turns into a two-phase mixture. The two-phase fluid moves up through the separator/riser section, where steam is separated from liquid water, and through the dryers, which ensure that the exiting steam (Q_v) is essentially dry. The separated water is recirculated back to the downcomer. The balance between the exiting Q_v and the incoming Q_e governs the change in the water level in the SG. Because of the two-phase nature, two types of water level measurements are considered, as shown in Fig. 1, each reflecting a different level concept: the Narrow Range Level (N_{rl}) is calculated by pressure difference between two points close to the water level and indicates the mixture level, whereas, the Wide Range Level (W_{rl}) is calculated by pressure difference between the two extremities of the SG (steam dome and bottom of the downcomer) and indicates the collapsed liquid level that is related with the mass of water in the SG.

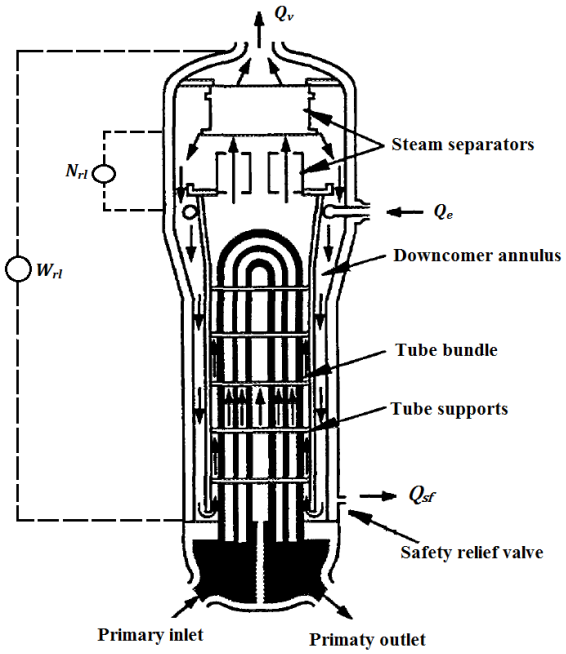


Fig. 1. *Schematic of the UTSG [IAEA-TECDOC-981, 1997]*

"Swell and shrink" phenomena are also modeled to reproduce the dynamic behavior of the SG: when Q_v increases, the steam pressure in the steam dome decreases and the two-phase fluid in the tube bundle expands causing N_{rl} to initially swell (i.e., rise), instead of decreasing as would have been expected by the mass balance; contrarily, if Q_v decreases or Q_e increases, a shrink effect occurs. A similar model has been presented in [Aubry et al., 2012].

The N_{rl} is governed by Q_e and Q_v across the tube bundle region of the SG as shown by the following transfer function:

$$N_{rl}(s) = \frac{1}{T_n s}\left(Q_{ef}(s) - Q_{GV}(s)\right) \tag{1}$$

where Q_{ef} is the flow-rate of the incoming water in the tube bundle, (Eq. (2)), Q_{GV} is the equivalent steam-water mixture flow-rate exiting the tube bundle region, (Eq. (3)), T_n is a time constant that accounts for the N_{rl} dynamics.

The incoming water flow-rate Q_{ef} is proportional to Q_e:

$$Q_{ef}(s) = \frac{1}{(1+T_h s)(1+\tau s)}\, Q_e(s) \tag{2}$$

where the lag $1/(1+\tau\, s)$ accounts for the feed-water valve dynamics and $1/(1+T_h s)$ accounts for the water mass transportation dynamics: their values are reported in Table 1.
The exiting steam-water mass Q_{GV} is proportional to Q_v:

$$Q_{GV}(\mathrm{s}) = \frac{(1-F_g\, T_g\, s)}{(1+T_g\, s)}\, Q_v(s) \tag{3}$$

where the first order lag $1/(1+T_g\, s)$ accounts for the elapsed time from the turbine steam demand and the increase of Q_{GV}, and the non-minimum phase term $(1-\ F_g\ T_g\ s)$ accounts for the two-phase swell and shrink effects.
Combining Eqs. (1), (2), and (3), N_{rl} is equal to:

$$N_{rl}(\mathrm{s}) = \frac{1}{T_n\, s}\left(\frac{Q_e(s)}{(1+T_h s)(1+\tau\, s)} - \frac{(1-\ F_g\, T_g\, s)}{(1+T_g\, s)}\, Q_v(s)\right) \tag{4}$$

and W_{rl}, i.e., the overall water mass in the steam generator, is:

$$W_{rl}(\mathrm{s}) = \frac{1}{T_{int}\, s}\big(Q_e(s) - Q_v(s)\big) \tag{5}$$

where T_{int} is a time constant that accounts for the W_{rl} dynamics.
We assume $y_1 = N_{rl}$ and $y_2 = W_{rl}$, and $u = Q_e$ and $d = Q_v$; the state space representation of the SG model is, thus:

$$\dot{x}(t) = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{T_n} \\ 0 & -\frac{1}{T_h} & 0 & -\frac{1}{T_n} \\ 0 & 0 & -\frac{1}{T_g} & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau} \end{pmatrix} \boldsymbol{x}(t) + \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{\tau} \end{pmatrix} \boldsymbol{u}(t) + \begin{pmatrix} -\frac{1}{T_n} \\ 0 \\ \frac{1+F_g}{T_n} \\ 0 \end{pmatrix} \boldsymbol{d}(t) \tag{6}$$

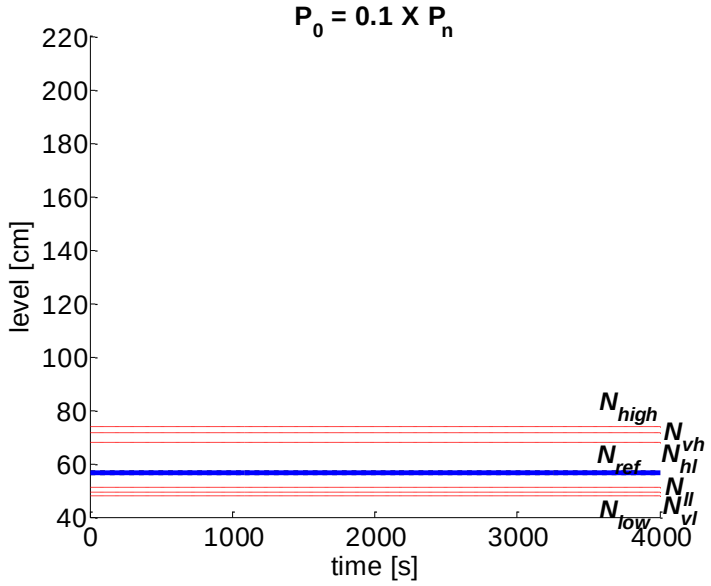
$$y(t) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ \frac{T_n}{T_{int}} & 0 & 0 & \frac{\tau}{T_{int}} \end{pmatrix} \boldsymbol{x}(t) \tag{7}$$

The values of the parameters $T_h, T_n, F_g, \tau, T_g, T_{int}$ change depending on the power P_o, as shown in Table 1.

Table 1. *Parameters of the UTSG model at different power levels* [Aubry et al., 2012]

P_o	$0.03 \times P_n$	$0.04 \times P_n$	$0.09 \times P_n$	$0.24 \times P_n$	$0.30 \times P_n$	$0.50 \times P_n$	P_n
T_n	36	56	63	44	40	40	40
F_g	13	18	10	4	4	4	4
T_h	170	56	30	10	8	5	5
τ	10	10	10	30	30	30	30
T_g	10	10	10	10	10	10	10
T_{int}	140	140	140	140	140	140	140

The goal of the system is to maintain the SG water level at a reference position (N_{ref}): the SG fails if the N_{rl} rises (falls) above (below) the threshold N_{high} (N_{low}), in which case automatic reactor or turbine trips are triggered. Indeed, if the N_{rl} exceeds N_{high}, the steam separator and dryer lose their functionality and excessive moisture is carried in Q_v, degrading the turbine blades profile and the turbine efficiency; if N_{rl} decreases below N_{low}, insufficient cooling capability of the primary fluid occurs. Similarly, the W_{rl}, is relevant for the cooling capability of the primary circuit [Kothare et al., 2000]. Pre-alarms are triggered when N_{rl} exceeds N_{hl} (N_{ll}) if a small deviation from N_{ref} occurs or when N_{rl} exceeds N_{vh} (N_{vl}), when the deviation is large. Set points of N_{ref} and of N_{rl} depend on P_o, as shown in Fig. 2, and, thus, also the alarms thresholds depend on P_o. The N_{rl} set point is low at low P_o, to partially account for the strong inverse response of N_{rl} [Kothare et al., 2000]; thus, the low level thresholds are more restrictive than the high level thresholds at low P_o.



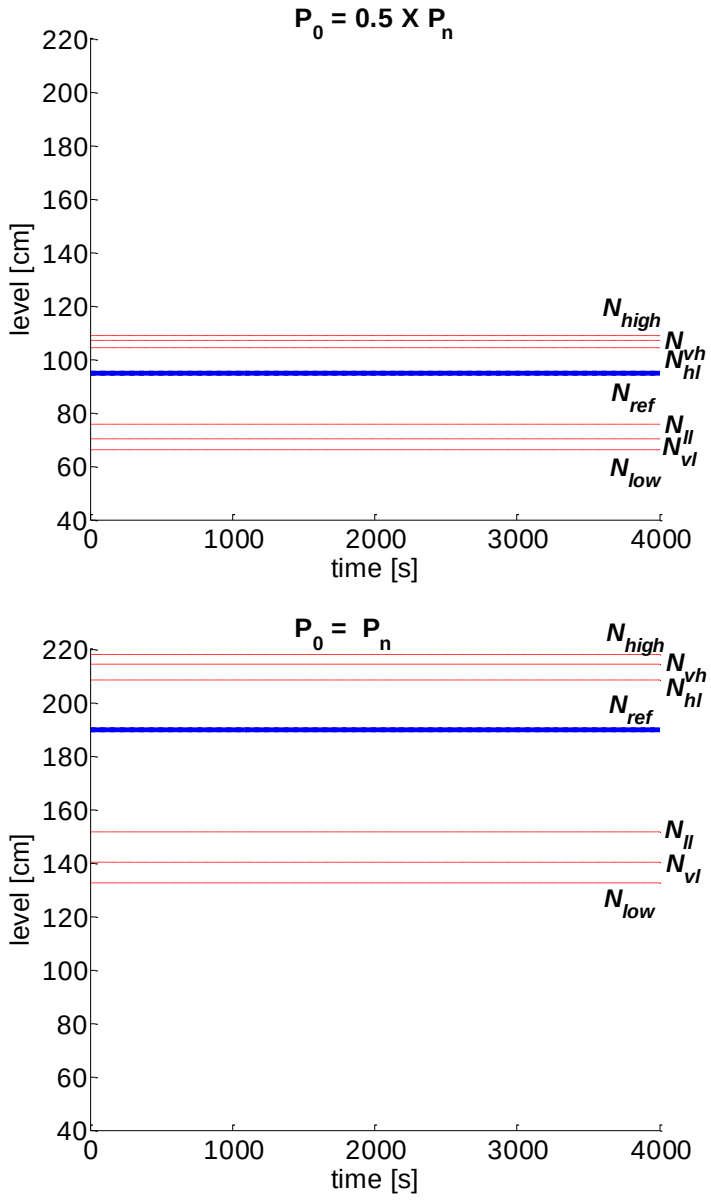


Fig. 2. *Set point for N_{rl} at different power rate P_o values.*

A dedicated model has been implemented in SIMULINK to simulate the dynamic response of the UTSG at different P_o values. Both feedforward and feedback digital control schemes have been adopted. The feedback controller is a PID that provides a flow rate Q_{pid} resulting from the residuals between N_{rl} and N_{ref}, whereas the feedforward controller operates a safety relief valve that is opened if and only if N_{rl} exceeds the N_{hl}, and removes a constant flow safety flow rate (Q_{sf}). The block diagram representing the SIMULINK model of the SG is shown in Fig. 3: the controlled variable is N_{rl}, whereas the control variable is Q_e.

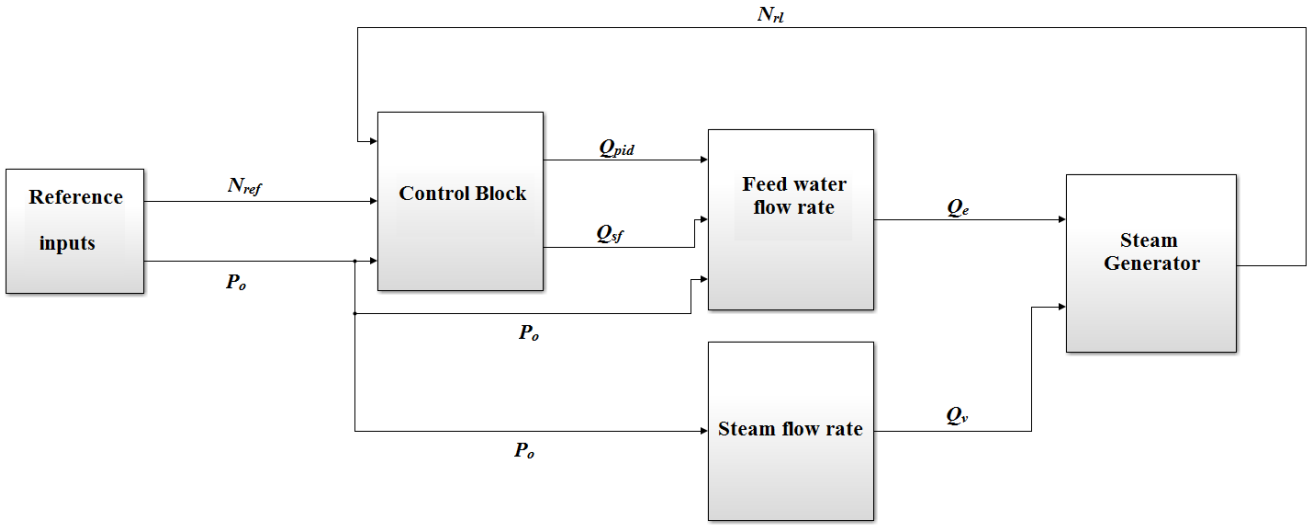


Fig. 3. *Block diagram representing the SIMULINK model of the SG.*

2.2 The set of possible failures

The set of multiple component failures that can occur during the system life are shown in Fig. 4:

1. The outlet steam valve can fail stuck at a random time in [0, 4000] (s), for example, due to corrosion, cracking and stress [IAEA-TECDOC-981, 1997], in three different positions: i) closed; ii) stuck open at 50% of the nominal Q_v that should be provided at P_o; iii) stuck open at 150% of the nominal Q_v that should be provided at P_o.
2. The communication between the sensor that monitors N_{rl} and the PID controller can fail at random times in [0, 4000] (s), in which case the PID is provided with the same input value of the previous time step, that would affect the actuation of the safety functions [Kang et al., 2007].
3. The safety relief valve can fail stuck at a random time in [0, 4000] (s), at a uniform random value Q_{sf} in the range [0.5, 50.5] (kg/s), for example, due to corrosion, cracking and stress [IAEA-TECDOC-981, 1997].
4. The PID controller can fail stuck at random times in [0, 4000] (s), providing a uniform random flow rate Q_{pid} belonging to [-18, 18] % of the nominal Q_e that should be provided at P_o, that would affect the actuation of the safety functions [Kang et al., 2007].

It is worth noticing that in the UTSG there are two PID controllers and, thus, two communications between the sensors measuring N_{rl} and the PIDs (one for high power feedback control and the other for low power feedback control). The selective action of the PIDs depending on P_o hides some of the

failures. For example, if the power profile of the scenario under investigation is a ramp, both PIDs are called in operation: if anyone (or both) is (are) failed, their fault state is detectable. On the contrary, if we consider scenarios with constant power profile, e.g., low power rate (P_o < 15% P_n), the occurrence of a high power feedback control failure cannot be detected, and, thus, the fault remains hidden.

The choice of a mission time (T_{miss}) equal to 4000 (s) has been made, because it is a long enough interval of time to allow the complete development also of slow dynamic accident scenarios [Di Maio et al., 2015].

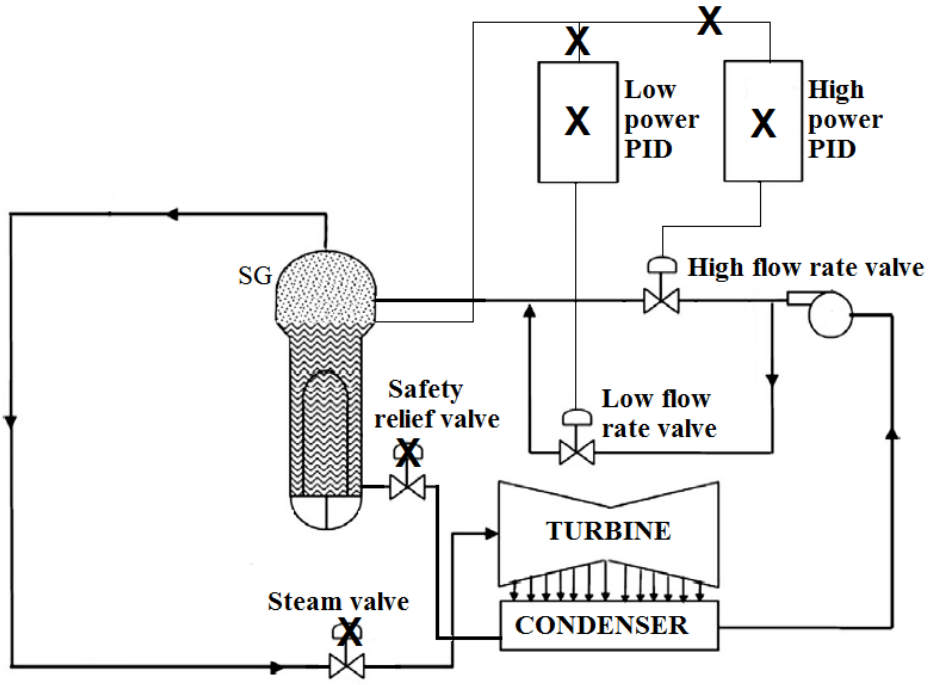


Fig. 4. *Sketch of the failures (**X**) that can be injected into the system*

3. Post-processing analysis for PIs and Near Misses identification

We adopt a computational framework based on Multiple Value Logic (MVL) [Garibba et al., 1985; Di Maio et al., 2014], for describing the components failure events in terms of their (discrete) times of occurrence and (discrete) magnitudes. The discretization of the time and magnitudes values is as follows:

- time discretization: we use the label t=1, t=2, t=3 and t=4, for failures occurring in the intervals [0, 1000] (s), [1001, 2000] (s), [2001, 3000] (s), [3001, 4000] (s), respectively; if the label t=0, the component does not fail within the time of the whole scenario, T_{miss}.
- Magnitude discretization:
  - the steam valve failure magnitude is indicated as 1, 2 or 3 for failure states corresponding to stuck at 0%, stuck at 50% and stuck at 150% of the Q_e value that should be provided at P_o, respectively; if the steam valve magnitude is indicated as 0, the component does not fail in T_{miss};
  - the safety relief valve fails with magnitude indicated as 1, 2, 3 and 4, if it is stuck between [0.5, 12.6] (kg/s), (12.6, 25.27] (kg/s), (25.27, 37.91] (kg/s) and (37.91, 50.5] (kg/s), respectively; if the safety relief valve magnitude is indicated as 0, the component does not fail in T_{miss};
  - the communication between the sensor measuring N_{rl} and the PID controller is labelled 0 if the communication works, 1 otherwise;
  - the PID controller failure magnitude range is discretized into 8 equally spaced magnitude intervals, labelled from 1 to 8, representative of failure states corresponding to discrete intervals of output value belonging to [-18,18]% of the Q_e value that should be provided at P_o; if the PID controller magnitude is labelled as 0, the component does not fail in T_{miss}.

The values of time, magnitude and order of failure occurrence for each component are included into a sequence vector that represents a scenario. It is worth mentioning that a finer MVL discretization (for example, by using a larger number of time intervals) would improve the adherence of the model to reality at the expenses of i) a larger computational burden for the MVL discretization, ii) a further increased number of scenarios to be post-processed, and iii) a larger complexity of the on-line clustering algorithm based on the Hamming distance to be built, without any actual benefit in improving the knowledge of the system behavior, as we shall see in what follows. The MVL approximation here undertaken can, thus, be considered a trade-off framework between computational costs and capability of timely characterization of the developing scenario.

A Monte Carlo-driven fault injection engine is used to sample combinations of discrete times and discrete magnitudes of components failure occurrences. The post-processing analysis of the sampled sequence vectors amounts to: i) the identification of Prime Implicants (PI) (Section 3.1) and ii) the identification of Near Misses (Section 3.2).

For the identification of the Near Misses, i.e., sequences of events that are similar to those accidental sequences leading the system into fault conditions, with the exception of a characteristic which is missing or is slightly different (e.g., sequence time lag, different failure magnitude, involved components) [Saleh et al., 2013], we adopt a risk-based clustering method [Di Maio et al., 2015c] that accounts for the order and timing of the events occurring along an accident sequence, and the magnitude of the process variables at the time of event occurrence [Aldemir et al., 2008; Di Maio et al., 2015c].

3.1 Prime implicants identification

A PI is a minimal set of variables that represents a minimal combination of accident component failures necessary for system failure and cannot be covered by a more reduced implicant [Quine, 1952; Di Maio et al., 2015a].

The PIs identification among the whole set of 100509 possible scenarios obtained by MVL approximation of the SG real behavior, is here performed with visual interactive method presented in [Di Maio et al., 2015b]. Since PIs are those scenarios with as few as possible events that are capable of leading the system into a failure state [Rocco et al., 2004], we select as most important feature for the PIs identification the literal cost of the sequence vector (i.e., the number of components whose behavior is specified in the accident sequence). The accident sequences associated with the lowest literal cost are selected and stored as PIs. In fact, these are the most reduced sequences (i.e., with least number of events) that cannot be covered by any other implicant, and thus, these are PIs by definition. The selected PIs, and implicants covered by selected PIs, are deleted from the set of implicants and, then, we repeat the procedure for the remaining implicants until all implicants are covered. By so doing, we identify 1255 PIs for the high level failure mode, which cover 36128 minterms.

3.2 Near Misses identification

Once the (1255) PIs for the SG high level failure mode have been identified as explained in Section 3.1, these are removed from the whole set of possible scenarios: the set of safe scenarios consists of 64381 sequence vectors. The Near Misses search is performed with the clustering method presented in [Di Maio et al., 2015c], where a risk-based characterization of the safe scenarios is done in terms of: i) the probability $p(t)$, that at time t the developing scenario can lead the system into an accidental scenario, ii) the consequence $c(t)$, that at time t the developing scenario is predicted to cause to the system, and iii) the overall risk r(t), that we synthetically compute as $r(t) = p(t) \times c(t)$. This is a

consolidated of risk definition that entails, for a given accidental scenario, determining $r(t)$ on the basis of how likely the scenario is ($p(t)$), and what are its consequences ($c(t)$) [NUREG-75/014, 1975]. In such definition, one is neglecting other aspects like the ease of accident detection [Zhang et al., 2009; Garaniya et al., 2015], or the maintenance costs and the time to repair the failed components [Haddara et al., 2004; Krishnasamy et al., 2005].

The identification of the Near Misses is treated as an unsupervised classification problem and addressed by clustering, where i) the number of clusters is unknown and ii) the features that enable the best clustering according to the risk-based characteristic profiles of $p(t)$, $c(t)$ and $r(t)$ of the accidental scenarios are unknown. Thus, we resort to a wrapper framework [Kohavi et al., 1997; Baraldi et al., 2012], whereby a Modified Binary Differential Evolution (MBDE) search engine [Wang et al., 2010; Di Maio et al., 2013] searches candidate groups of features sets that are fed to a K-means clustering algorithm [MacQueen, 1967]; eventually, the wrapper evolves so that among these candidate groups, the group retained is that which makes the K-means clustering algorithm perform best (most compact and separate clusters). The search proceeds iteratively until the Calinski-Harabasz (CH) index [Calinski et al., 1974], which accounts for the ratio of the overall between-cluster variance (separation) and the overall within-cluster variance (compactness), is maximised and the number of clusters K is fixed.

The optimal features selection provides as best features: the standard deviation of $c(t)$, the standard deviation of $r(t)$ and the root mean square of $r(t)$; the best performance is obtained with CH=9.35e+04 and K= 5.

The K= 5 obtained clusters of the safe scenarios are shown in Fig. 5, with reference to the features of mean risk (μ_{risk}) and time elapsed from the instant t_{risk} at which $r(t)$ starts to deviate from zero, i.e., the time interval during which the system is exposed to risk. The rationale behind this choice is that the larger μ_{risk} and the longer t_{risk}, the more dangerous the scenarios. In Fig. 5, clusters 3, 4, 5 (triangles, crosses and squares, respectively) are well separated: it is possible to distinguish the scenarios having the lowest risk level from the scenarios having low risk level, and, thus, the highest risk scenarios are well separated from the lower risk scenarios. The 332 circles in Fig. 5 can, thus, be considered the Near Misses scenarios, i.e., scenarios that incidentally keep the system into safe state, although in endangered and insecure, operational conditions. Looking for the minimum conditions, i.e., minimum μ_{risk} and minimum t_{risk}, that lead the system into a quasi-fault state, we can find the most similar characteristics among Near Misses in terms of their Multiple Value sequences, i.e., order and timing of event occurrences and deterministic process variables values:

– the failure of the communication between the sensor monitoring the N_{rl} and the PID controller;
– the failure of the PID controller with magnitude belonging to [-5, -1] % of the Q_e value that should be provided at P_o, i.e., magnitude equal to 4 in MVL framework, and it is the first accident occurring along the sequence of events in over 85% of the Near Misses scenarios.

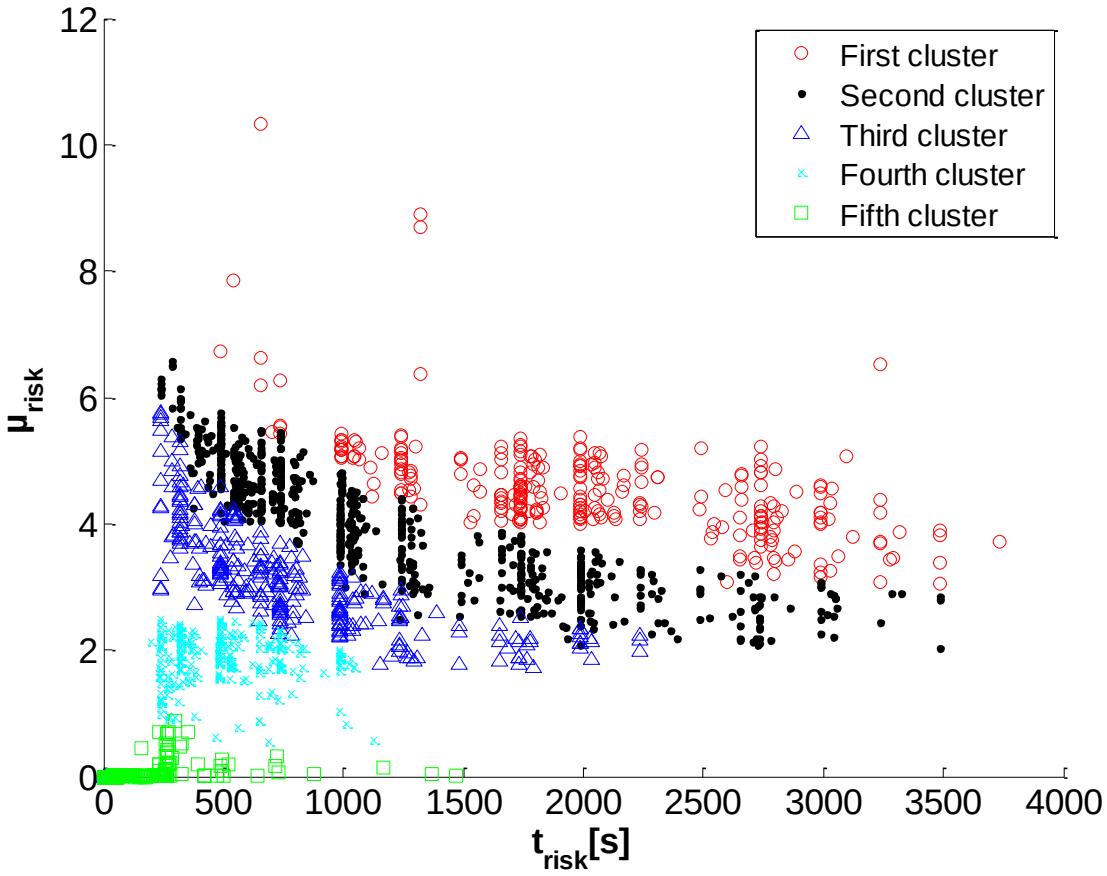


Fig.5. *Near Misses identification clustering results*

4. On-line scenario clustering

4.1 The method

As a result of the previous steps of the analysis, we have unambiguously assigned each one of the sampled scenarios into three classes: Safe scenarios, PI scenarios, Near Misses scenarios. This

database of labelled scenarios is exploited for on-line identification of an unknown developing scenario on the basis of its developing event sequence vector and on the information carried by the monitored process variables. In our case, the process variable considered is the water level N_{rl}: if the residual between N_{rl} and N_{ref}, $r = N_{rl} - N_{ref}$ differs from 0 and the Hamming distance [Hamming, 1950] between the developing sequence vector and each of the sequence vectors belonging either to the PIs cluster or to the Near Misses cluster is low, an alarm is triggered. The Hamming distance [Hamming, 1950] is equal to the number of digits that must be changed in a vector in order to obtain a different vector [Popa et al., 2010]. An example of Hamming distance computation is shown in Fig. 6: the developing sequence vector is compared with a vector belonging to the Near Misses cluster (Fig. 6, left) and with a vector belonging to the PI cluster (Fig. 6, right). In this case, the Hamming distance between the developing sequence vector and the PI sequence vector is equal to zero, whereas the Hamming distance from the vector belonging to the Near Misses cluster is equal to 12. Thus, the developing scenario, which is caused by the failure of the safety relief valve in the first time interval with magnitude equal to 2 followed by the failure of the communication between the sensor measuring N_{rl} and the PID controller, is identified as a developing PI scenario and the failure alarm is triggered.

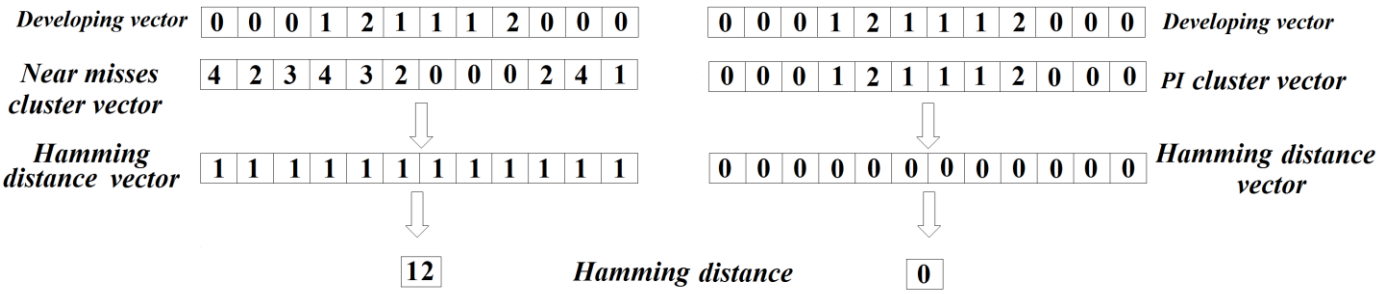


Fig.6. *Hamming distance between a developing sequence vector and a PI- classified sequence vector.*

Practically, at the generic time t at which $r > 0$, the developing scenario is assigned to a specific cluster according to the following criteria:

1. It belongs to the PIs cluster if the Hamming distance between the developing sequence vector at t and a sequence vector of PI cluster is the smallest among all the 1255 of PI and 332 of Near Misses calculated distances.

2. It belongs to the Near Misses cluster if the Hamming distance between the developing sequence vector at t and a sequence vector of Near Misses cluster is the smallest among all the 1255 of PI and 332 of Near Misses calculated distances.
3. It belongs to the safe cluster, otherwise.

To verify the method, we generate a set of 1000 new developing scenarios by injecting component faults at random times and of random magnitudes via Monte Carlo sampling. The performance of the proposed method of scenario clustering is compared with the simplest process-variable, threshold-based classification method frequently used in industrial practice [Tompkins et al., 1985; Friston et al., 1996; Thomas et al., 2010]. That assigns the developing scenario to a cluster depending on the following criteria looking exclusively at the process variable value (while not considering any information of the scenario sequence vector):

- PI, if N_{rl} exceeds N_{high} (faulty state);
- Near Misses, if $N_{vh} < N_{rl} < N_{high}$ (quasi-fault state).

It is worth noticing that the exceedance of N_{vh} is neither a sufficient nor a necessary condition to define a scenario as Near Miss [Di Maio et al., 2015c], because timing and speed of the water level changes have been shown to play a key role for the Near Misses characterization of dynamic systems.

The performance of the on-line classification will be measured by:

- The number of false positives, i.e., those scenarios classified as PIs or Near Misses, when they actually are not;
- The number of false negatives, i.e., those scenarios classified as Safe, when they are actually either PIs or Near Misses scenarios;
- The value of Grace Time (GT), i.e., the time between the alarm triggering and the threshold exceedance of N_{high} and N_{vh} for PIs and Near Misses, respectively.
- Percentage of Accident Progression (PAP), i.e., the percentage of time between the IE (at which PAP=0%) and the time at which N_{rl} is equal to N_{vh} or N_{high} (at which PAP=100%) for PIs and Near Misses, respectively.

4.2 Results

4.2.1 PI on-line identification results

Fig. 7 shows two examples of PI scenario identification. In Fig. 7 (a), IE is caused by the failure of the safety relief valve with magnitude equal to 1 at 250 (s) and, then, the failure of the communication between the sensor measuring N_{rl} and the PID controller occurs at 750 (s). The risk-based clustering method identifies the developing scenario as PI at 751 (s) (Fig. 7 (a), circles line), whereas the threshold-based clustering method identifies the developing scenario at 3215 (s) (Fig. 7 (a), squares line). This shows that the proposed method is able to early identify the developing PI scenario, when the GT is still 2464 (s), differently from the threshold-based method which identifies the scenario with a GT equal to 0 (s). Similarly, Fig. 7 (b) shows the scenario developing from the IE of a PID controller failure with magnitude equal to 2 at 500 (s): this failure leads very quickly to N_{high} exceedance, as promptly captured by the risk-based clustering method that identifies the developing scenario as soon as N_{rl} departs from N_{ref}, at 501 (s) (Fig. 7 (b), circles line) with a GT equal to 48 (s), whereas the threshold-based classification algorithm (Fig. 7 (b), squares line) can only identify the developing scenario as PI when N_{rl} exceeds N_{high}.

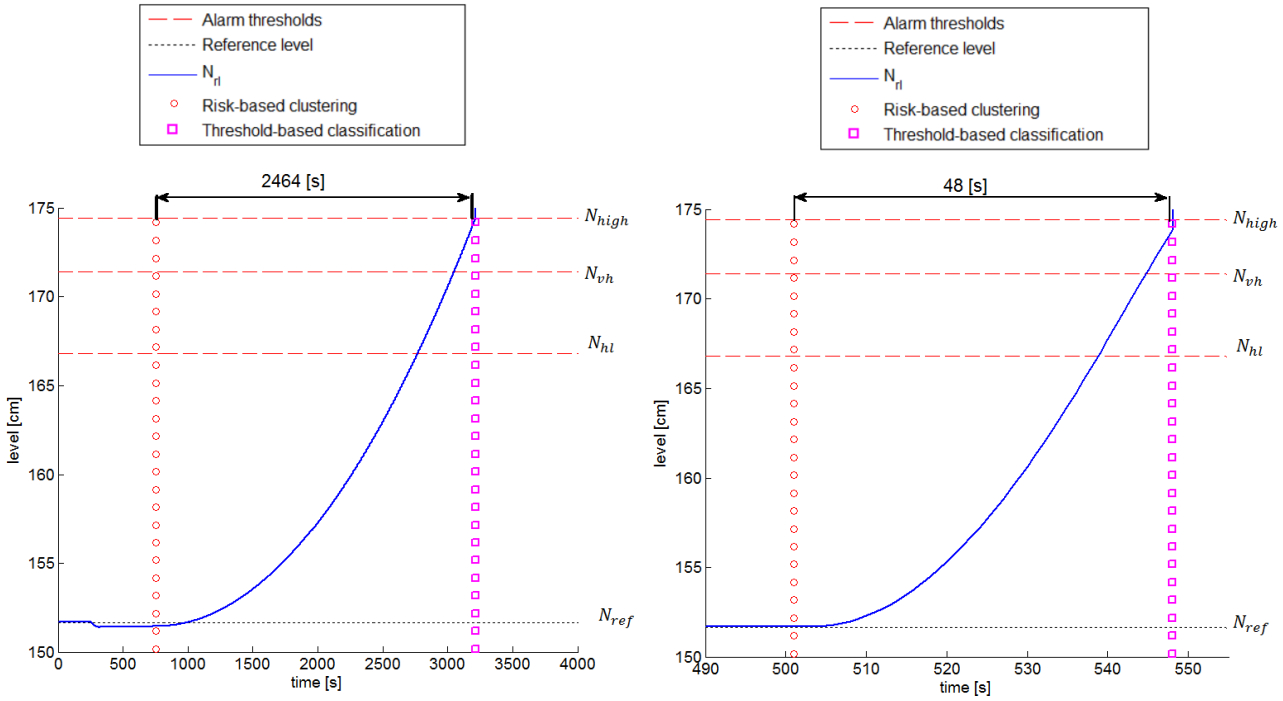


Fig.7. *Example of PI transient identification by risk-based clustering and thresholds-based classification.*

For 517 randomly extracted PIs scenarios (see Section 3), false positive/negative ratios, and GT mean values are given in Table 2.

Table 2. *Results of on-line PIs identification*

	Risk-based on-line clustering algorithm	Threshold-based classification algorithm
False Positive	10 (1%)	0
False Negative	0	0
GT mean values [s]	92	0
PAP mean value [%]	49	100

The risk-based on-line clustering method is superior in the anticipation of the alarm, with respect to the threshold classification method that triggers it only when the process variable already exceeds the threshold of faulty condition (GT=0).

However, the risk based-clustering records 1% of misclassified transients, triggering them as PI scenarios even though they are not. An example of false positive scenario is shown in Fig. 8. The misleading classification of the developing scenario as PI is due to the values of the developing sequence vector at time 502 (s), as shown in Table 3. At 502 (s) the developing scenario (Fig. 8, solid line) has a sequence vector equal to that of a PI sequence vector (Table 3, second and first rows, respectively) and the Hamming distance is, thus, equal to 0 (Table 3, third row): the developing scenario is classified as PI and an alarm is triggered (Fig. 8, circles line). If no other component failure occurred, the developing N_{rl} would have followed the trend of the PI previously matched scenario (Fig. 8, dashed-dotted line). Actually, another component failure occurs at 833 (s) that modifies the developing sequence vector into a non-PI vector (Table 3, fourth row), bringing N_{rl} to a different end state (Table 3, fifth row). When a false positive PI is recorded, the failure alarm is triggered, and thus, this can lead operators to take the wrong corrective decision; the risk-based clustering method is more conservative than threshold-based classification. However, the coupling of risk-based clustering algorithm and the threshold-based classification with threshold smaller than the failure ones, can give the optimal solution for monitoring and predicting the system evolution.

Table 3. *Comparison between developing sequence vector and PI sequence vector*

PI sequence vector	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**2**	**0**	**0**	**0**
Developing sequence vector at 502 (s)	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**2**	**0**	**0**	**0**
Hamming distance vector	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Developing sequence vector at 833 (s)	**1**	**2**	**3**	**1**	**1**	**1**	**1**	**1**	**2**	**0**	**0**	**0**
Hamming distance vector	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

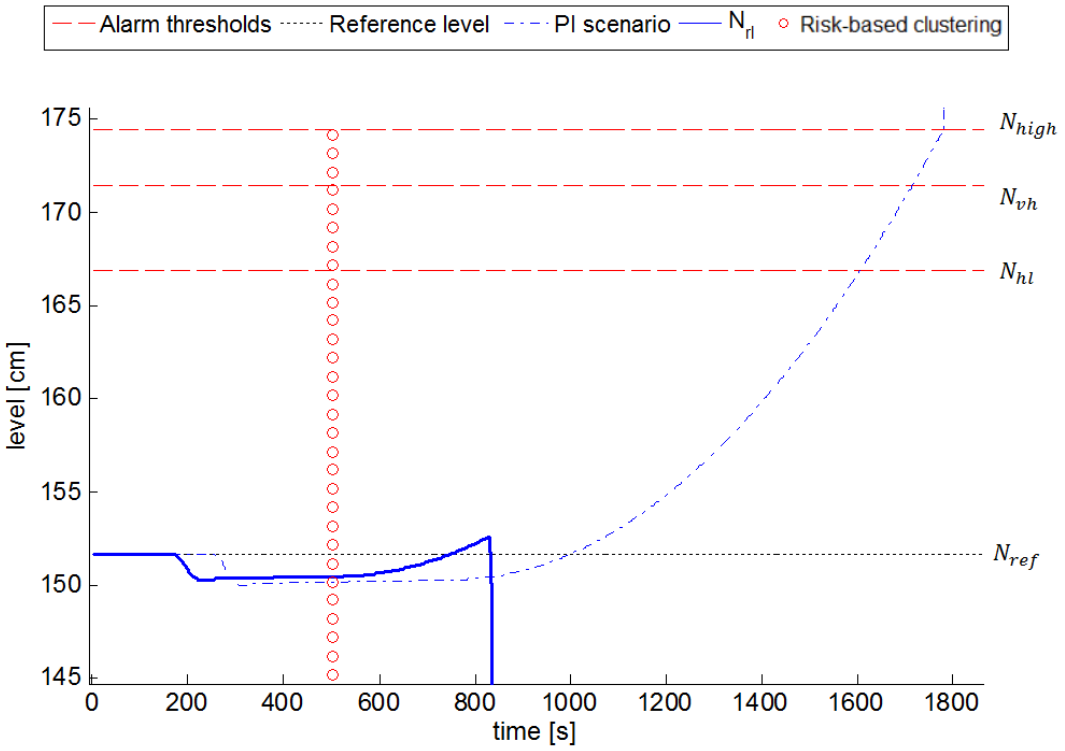


Fig.8. *Example of false positive PI scenario.*

Fig. 9 shows the distribution of GTs for the 517 transients: most of the GTs are within 200 [s], i.e., in the 88% of the developing PI scenarios tested, the operators have at least 200 (s) to counteract the occurring accidental scenario. In 11% of the tested PI transients, the operators have at least 600 (s) to take corrective actions, whereas in 1% of the tested developing PI scenarios they have more than 1000 (s), i.e., the NPPs operators know more than 16 minutes in advance that the developing N_{rl} will exceed N_{high}.

With respect to the percentage of PAP indicator (equal to 0% at the instant when IE occurs and to 100% when $N_{rl} = N_{high}$), on average, PIs are correctly classified with PAP=49% (Fig. 10) and 29% of them are identified with PAP<4%. Instead, obviously, the threshold-based classification method classifies the developing scenario as PI or Near Misses only when it is fully developed, i.e., at 100% of accident progression.

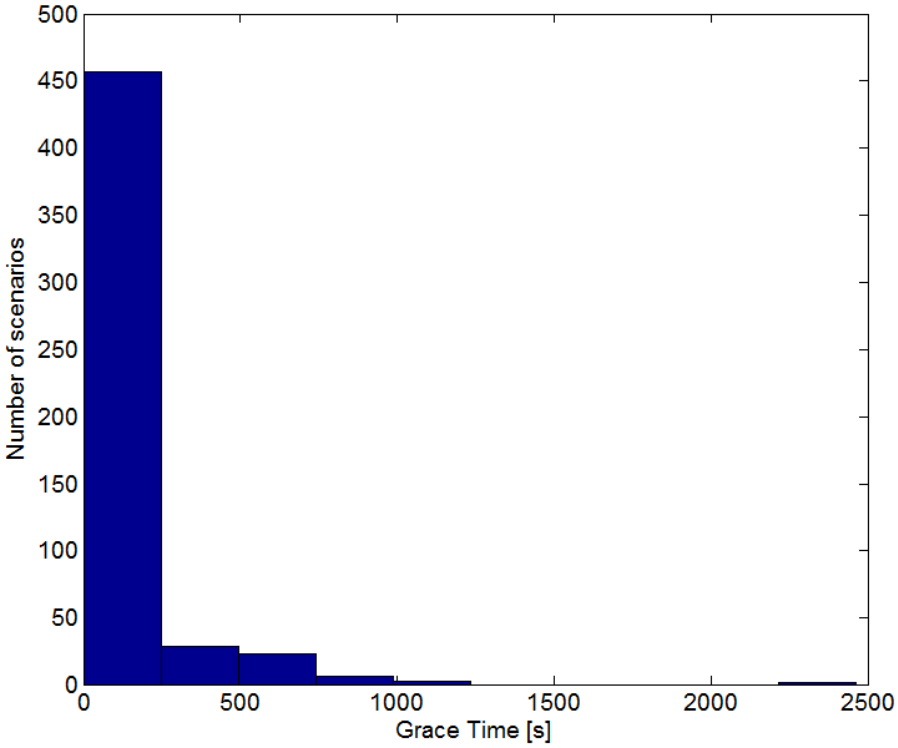


Fig.9. *Histogram of GT values for PI identification.*

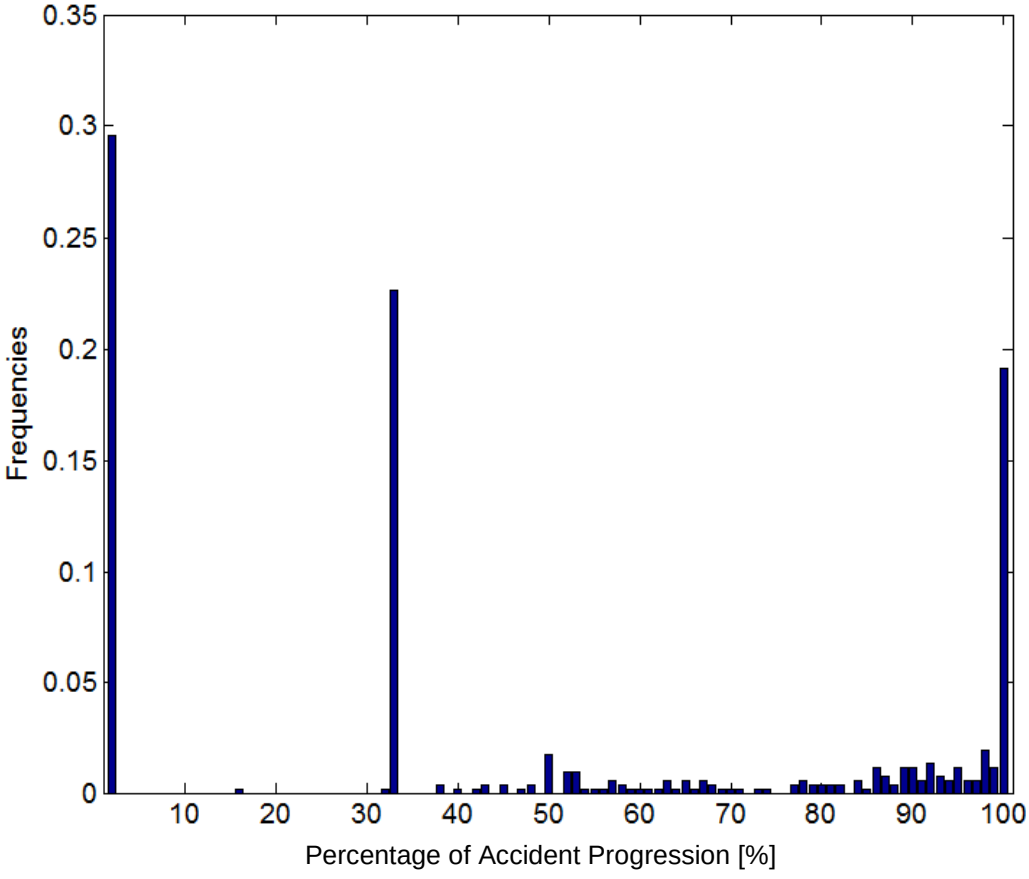
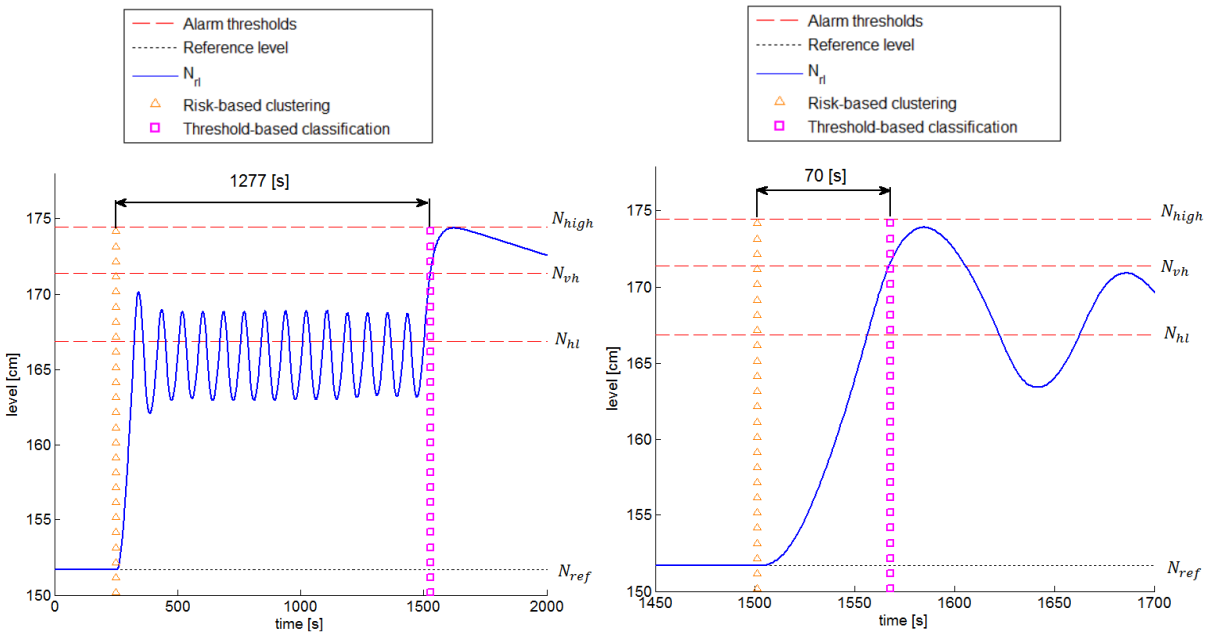


Fig.10. *Distribution of the percentage of accident progression.*

4.2.2 Near Misses on-line identification results

Fig. 11 shows two examples of Near Misses identification. In Fig. 11 (a), IE is the failure of the PID controller with magnitude equal to 4 at 250 (s). This IE failure is common in the Near Misses identified sequence vectors (see Section 3) and so the proposed risk-based clustering algorithm is able to recognize the developing scenario as Near Miss immediately at 251 (s) (Fig. 11 (a), triangles line) with GT equal to 1277 (s). On the other hand, the threshold-based classification method is only able to classify the developing scenario at 1528 (s), when the safety relief valve fails, leading N_{rl} to exceed N_{vh} (Fig. 11 (a), squares line).

Instead, Fig. 11 (b) shows a PID controller failure with magnitude equal to 3 at 1500 (s), which brings N_{rl} to quickly exceed N_{vh}: in this case, the risk-based clustering method classifies the developing scenario as Near Miss as soon as N_{rl} departs from N_{ref}, at 1501 (s) (Fig. 11 (b), triangles line) with a GT equal to 70 (s), whereas the threshold-based classification algorithm has to wait until N_{rl} exceeds N_{vh} (Fig. 11 (b), squares line).



(a) ***(b)***

Fig.11. *Example of Near Miss scenario classification by risk-based clustering and threshold-based classification.*

For the 179 randomly sampled Near Miss scenarios, the false positive and negative ratios, together with the mean value of the GT are given in Table 4.

Table 4. *Results of on-line Near Misses identification*

	Risk-based on-line clustering algorithm	Threshold-based classification algorithm
False Positive	100 (10%)	12 (1.2%)
False Negative	0	99 (9.9%)
GT mean value [s]	129	0
PAP mean value [%]	13	100

As for PIs identification, the risk-based clustering method allows triggering the alarm as soon as the developing sequence vector becomes similar to anyone Near Miss in the labelled scenario database constructed in the previous post-processing analysis allowing mean values of GT equal to 129 (s) (Table 4). False positive scenarios by the proposed risk-based clustering method are caused by the similarity between the values of the developing sequence vector and those belonging to the Near Misses cluster; on the contrary, in the threshold-based classification method they are due to the fact that the exceedance of N_{vh} is neither a sufficient nor a necessary condition to label a scenario as Near Miss [Di Maio et al., 2015c]. Also, the threshold-based classification method counts 99 scenarios mistakenly classified as safe scenarios. Fig. 12 shows an example of such false negative scenario: the PID controller fails at 750 (s) with magnitude equal to 3, so, N_{rl} departs from N_{ref}; however, N_{rl} does not exceed N_{vh} (Fig. 12, solid line) and, thus, the threshold-based classification method is not able to classify the developing scenario (Fig. 12, squares line). On the contrary, the proposed risk-based clustering method comparing the developing sequence vector with those belonging to the Near Misses cluster by Hamming Distance, is able to identify the developing scenario as a Near Miss at 501 (s) (Fig. 12, triangles line).

Fig. 13 shows the distribution of GTs for the 179 Near Misses: 92% of these scenarios are identified by the risk-based clustering algorithm with GT within [1, 277] (s), leaving operators with up to 277 (s) to take corrective actions for the safety of the system. In 6%, GT belongs to (277, 913] (s), i.e., the developing Near Miss scenario is identified 913 (s) earlier than N_{rl} exceeds N_{vh}. Finally, 2% of the Near Miss scenarios are identified 1000 (s) earlier than the threshold-based classification method identification by threshold exceedance.

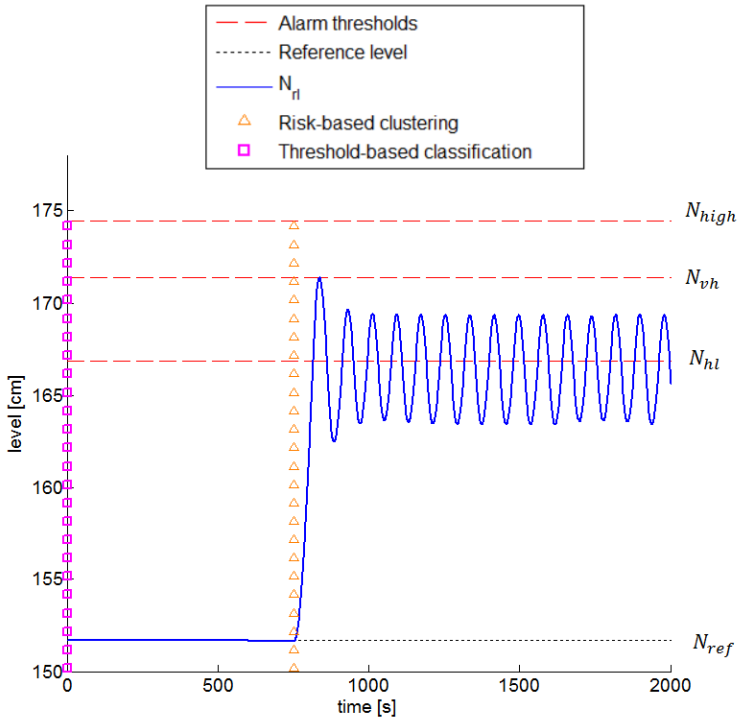


Fig.12. *Threshold-based classification false negative scenario.*

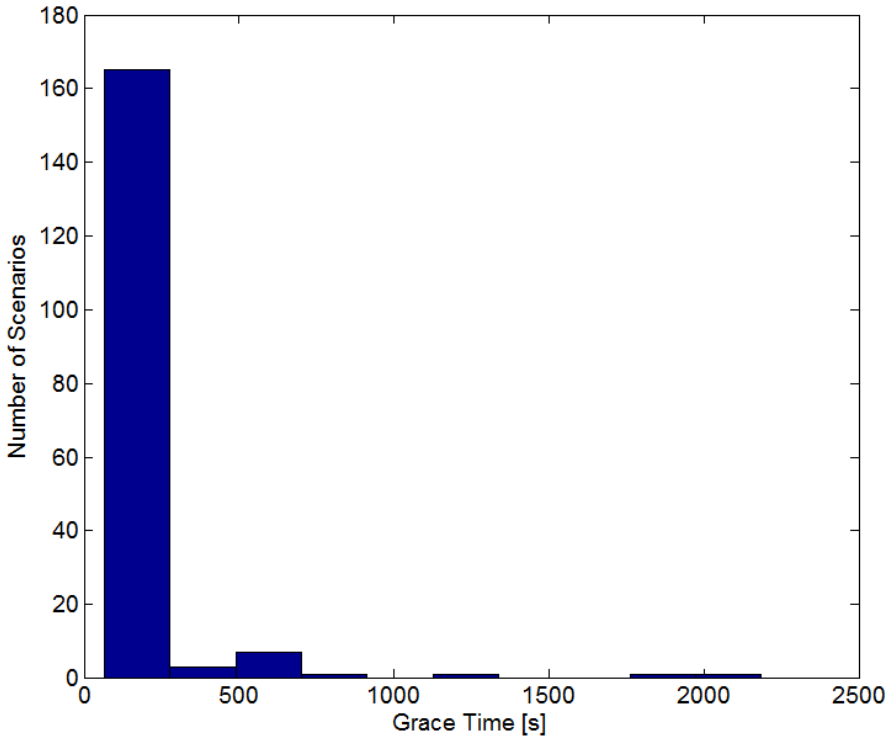


Fig.13. *Distribution of GT for Near Misses identification.*

The capability of the proposed method to effectively trigger the Near Misses alarm at an early stage of the dangerous progression is shown in Fig. 14. Note that for the on-line classification of Near Misses, PAP=100% with $N_{rl} = N_{vh}$. Fig. 14 shows that, on average, Near Misses are correctly classified with PAP=13%. Furthermore, it is worth noticing that 80% of the Near Misses scenarios are identified with PAP<2%.

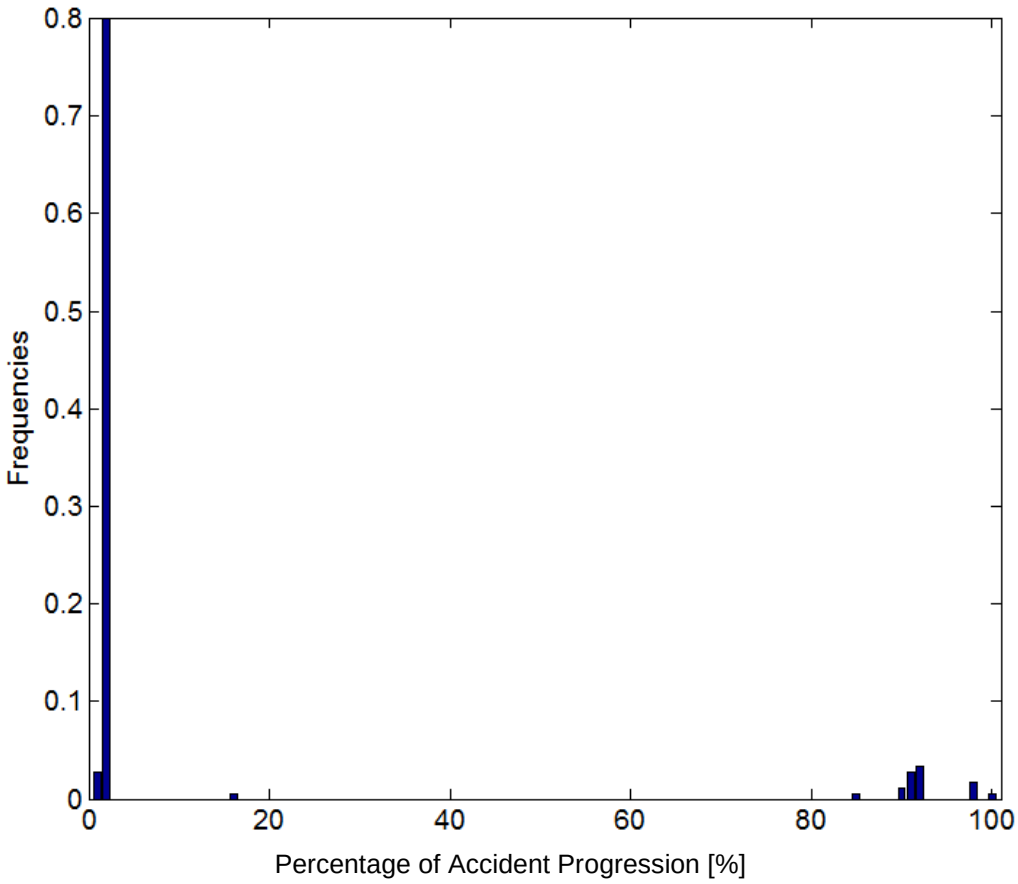


Fig.14. *Distribution of the percentage of accident progression.*

5. CONCLUSIONS

In this paper, a novel approach has been proposed for on-line identification of transients. An IDPSA of a steam generator of a NPP has been developed and, thus, the post-processing of its outcomes has been carried out for creating a database of PI and Near Misses scenarios. Indeed, it is worth pointing out that, due to the intrinsic incompleteness of real data (that only record historically occurred scenarios), the efforts to develop a simulation model of the system to be analyzed is indispensable to cover the whole set of possible sequences of failure events that would occur in a real system.

The off line characterization of the class of PI transients has been solved with a visual interactive method and the Near Misses identification with a risk-based clustering method. The IDPSA has led to an exhaustive and complete exploration of the scenario space and coverage of undesired events, with the consistent treatment of the different sources of uncertainty involved in the analysis, both aleatory and epistemic. A novel on-line clustering analysis has, then, been presented for the identification and prediction of accident progression when an initiating failure event occurs. The on-line transient identification is based on the sequence of events that compose the scenario as it develops, which is compared by Hamming distance with the sequence vectors of the IDPSA scenarios database, monitoring the whole set of the process variables together with the controlled variable and, thus, the proposed algorithm is robust to cope with noise in data.

The results obtained for a UTSG of a NPP have shown that for both PIs and Near Misses identification the proposed risk-based on-line clustering method is superior in the anticipation of the alarm, with respect to traditional threshold-based classification algorithms. Despite that the risk based-clustering has recorded 1% of misclassified transients as false positive scenarios (for PIs) and 10% (for Near Misses), this performance is comparable to that of the threshold-based classification algorithm, that rates to 9.9% the ratio of false negative scenarios wrongly classified.

As a final remark it is worth pointing out that, even if a detailed USTG model as adherent as possible to reality has been here used to simulate the whole set of accidental scenarios, in future the benchmark of these results with those obtained with finer MVL discretization will be the focus of the research, to verify the performance of the proposed method on more realistic case studies and improve it to achieve lower misclassification rates. Nevertheless, we can henceforth be confident that the performance of the proposed method can be guaranteed on real accidental scenarios thanks to the confidence we have that the considered simulated UTSG adheres to the real component of a NPP and that the proposed method properly balances false positives/negatives with correct classifications. In conclusion, we can claim the reliability of the proposed approach for practical use.

References

[Alaei et al., 2013] Alaei, H.K., Salahshoor, K., Alaei, H.K.,"*A new integrated on-line fuzzy clustering and segmentation methodology with adaptive PCA approach for process monitoring and fault detection and diagnosis*", Soft Computing, 17 (3), pp. 345-362, 2013.

[Aldemir et al., 2008] Aldemir, T., Guarro, S., Kirschenbaum, J., Mandelli, D., Mangan, L.A., Bucci, P., Yau, M., Johnson, B., Elks, C., Ekici, E., Stovsky, M.P., Miller, D.W., Sun, X., Arndt, S.A., *"A*

Benchmark Implementation of Two Dynamic Methodologies for the Reliability Modeling of Digital Instrumentation and Control Systems". NUREG-CR Report Draft, 2008.

[Aubry et al., 2012] Aubry J. F., Babykina G., Barros A., Brinzei N., Deleuze G., De Saporta B., Dufour F., Langeron Y., Zhang H., *"Project APPRODYN: APPROches de la fiabilité DYNamique pour modéliser des systèmes critiques"*, Technical report, collaboration CRAN, EDF R&D, INRIACQFD, UTT-ICD, 2012.

[Baraldi et al., 2012] Baraldi P., Di Maio F., Pappaglione L., Zio E., Seraoui R., *"Condition Monitoring of Electrical Power Plant Components During Operational Transients"*, Proceedings of the Institution of Mechanical Engineers, Part O, Journal of Risk and Reliability, 226(6) 568–583, 2012.

[Baraldi et al., 2013] Baraldi P., Di Maio F., Zio E., "Unsupervised Clustering for Fault Diagnosis in Nuclear Power Plant Components", International Journal of Computational Intelligence Systems, 6 (4), pp.764-777, 2013.

[Basu et al., 1994] Basu, A., Bartlett, E.B., "*Detecting faults in a nuclear power plant by using a dynamic node architecture artificial neural network*", Nuclear Science and Engineering 116, 313–325, 1994.

[Beringer et al., 2006] Beringer J, Hullermeier E, "*Online clustering of parallel data streams*", Data Knowledge and Engineering 58(2),180–204, 2006.

[Calinski et al., 1974] Calinski, T., and J. Harabasz., "*A dendrite method for cluster analysis*." Communications in Statistics. Vol. 3, No. 1, pp. 1–27, 1974.

[Collaghan et al., 2002] Collaghan LO, Mishra N, Meyerson A, "*Streming-data algorithms for high-quality clustering*", Proceedings of IEEE international conference on data engineering, 2002.

[Di Maio et al., 2013] F. Di Maio, P. Baraldi, E. Zio, R. Seraoui, "*Fault Detection in Nuclear Power Plants Components by a Combination of Statistical Methods*", IEEE Transaction on Reliability, 62 (4) , pp. 833-845, 2013.

[Di Maio et al., 2014] F. Di Maio, S. Baronchelli, E. Zio, "*Hierarchical Differential Evolution for Minimal Cut Sets Identification: Application to Nuclear Safety Systems*", European Journal of Operational Research, Volume 238, Issue 2, Pages 645–652, 2014.

[Di Maio et al., 2015a] F. Di Maio, S. Baronchelli, E. Zio, "*A Computational framework for Prime Implicants Identification in non-coherent Dynamic Systems*", Risk Analysis, 35 (1), pp. 142-156, 2015.

[Di Maio et al., 2015b] Di Maio F., Baronchelli S., Zio E., *"A visual interactive method for prime implicants identification"*, IEEE Transactions on Reliability, 64 (2), art. no. 6969120, pp. 539-549., 2015.

[Di Maio et al., 2015c] Di Maio F., Vagnoli M., Zio E., "Risk-Based clustering for Near Misses identification in Integrated Deterministic and Probabilistic Safety Analysis", Science and Technology of Nuclear Installations, Article ID 693891.

[Fink et al., 2015] Fink, O., Zio, E., Weidmann, U., "*Novelty detection by multivariate kernel density estimation and growing neural gas algorithm*", Mechanical Systems and Signal Processing, 50-51, pp. 427-436, 2015.

[Friston et al., 1996] Friston, K.J., Holmes, A., Poline, J.-B., Price, C.J., Frith, C.D., "*Detecting activations in pet and fMRI: Levels of inference and power NeuroImage*", 4 (3), pp. 223-235, 1996.

[Garaniya et al., 2015] Yu, H., Khan, F., Garaniya, V., "*Risk-based fault detection using Self-Organizing Map*", Reliability Engineering & System Safety, Volume 139, Pages 82-96, 2015.

[Garibba et al., 1985] Garibba S., Guagnini E., Mussio P., "*Multiple-Valued Logic Trees: Meaning and Prime Implicants"*, IEEE Transactions on Reliability, Volume R-34, No. 5, pp.463-472, 1985.

[Habibiyan et al., 2004] Habibiyan, H., Setayeshi, S., Arab-Alibeik, H. "*A fuzzy-gain-scheduled neural controller for nuclear steam generators*", Annals of Nuclear Energy, 31 (15), pp. 1765-1781, 2004.

[Haddara et al., 2004] Khan, F.I., Haddara, M.R., "*Risk-based maintenance of ethylene oxide production facilities*", Journal of Hazardous Materials, 108 (3), pp. 147-159, 2004.

[Hamming, 1950] Hamming, R.W., "*Error detecting and error correcting codes*", Bell Syst. Tech. J.2, 147–160, 1950.

[Kang et al., 2007] Kang, H.G., Jang, S.-C., "Plant risk effect analysis focusing on digital I&C equipment failures", Journal of Nuclear Science and Technology, 44 (4), pp. 590-596, 2007.

[Khakzad et al., 2012] Khakzad, N., Khan, F., Amyotte, P. "Dynamic risk analysis using bow-tie approach", Reliability Engineering and System Safety, 104, pp. 36-44, 2012.

[Kohavi et al., 1997] Kohavi, R., John, G.H.,"Wrappers for feature subset selection", Artificial Intelligence, 97 (1-2), pp. 273-324, 1997.

[Kothare et al., 2000] Kothare, M.V., Mettler, B., Morari, M., Bendotti, P., Falinower, C.-M., "*Level control in the steam generator of a nuclear power plant*", IEEE Transactions on Control Systems Technology, 8 (1), pp. 55-69, 2000.

[Krishnasamy et al., 2005] Krishnasamy, L., Khan, F., Haddara, M., "*Development of a risk-based maintenance (RBM) strategy for a power-generating plant*", Journal of Loss Prevention in the Process Industries, 18 (2), pp. 69-81, 2005.

[IAEA-TECDOC-981, 1997] "*Assessment and management of ageing of major nuclear power plant component important to safety: Steam*", IAEA, IAEA-TECDOC-98, Vienna, ISSN 1011-4289, 1997.

[MacQueen, 1967] MacQueen J., "*Some methods for classification and analysis of multivariate observations*" *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[Markou et al., 2003] Markou, M., Singh, S., "*Novelty detection: A review - Part 1: Statistical approaches*",Signal Processing, 83 (12), pp. 2481-2497, 2003.

[NUREG-75/014, 1975] "*Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants [NUREG-75/014 (WASH-1400)]*", U.S. Nuclear Regulatory Commission, Washington, DC, 1975.

[Palade et al., 2002] Palade V, Patton RJ, Uppal FJ, Quevedo J, Daley S, "*Fault diagnosis of an industrial gaz turbine using neuro-fuzzy methods*", IFAC World Congress, IFAC'02, Barcelona, 2002.

[Popa et al., 2010] Popa, A., McDowell, J.J. The effect of Hamming distances in a computational model of selection by consequences", Behavioural Processes, 84 (1), pp. 428-434, 2010.

[Quine, 1952] Quine W.V., "*The problem of simplifying truth functions*", Am. Math. Monthly, Volume 59, 521-531, 1952.

[Rocco et al., 2004] Rocco C.M., Muselli M., "*A Machine Learning Algorithm to Estimate Minimal Cut and Path Sets from a Monte Carlo Simulation*", Proceedings Probabilistic Safety Assessment and Management PSAM7/ESREL'04, 2008.

[Saleh et al., 2013] Saleh, J.H., Saltmarsh, E.A., Favarò, F.M., Brevault, L., "*Accident precursors, near misses, and warning signs: Critical review and formal definitions within the framework of Discrete Event Systems*", Reliability Engineering and System Safety, 114 (1), pp. 148-154, 2013.

[Schirru et al., 1999] Schirru, R., Martinez, A.S., Pereira, C.M.N.A., Domingos, R.P., Machado, M.D., and Machado, L., "*Intelligent Soft Computing in Nuclear Energy in Brazil*", Progress in Nuclear Energy 35, 367–391, 1999.

[Schirru et al., 2008] Carlos Canedo Medeiros, J.A., Schirru, R. "*Identification of nuclear power plant transients using the Particle Swarm Optimization algorithm*", Annals of Nuclear Energy, 35 (4), pp. 576-582, 2008.

[Thomas et al., 2010] Thomas, B., Raju, G., "*A fuzzy threshold based unsupervised clustering algorithm for natural data exploration*",ICNIT 2010, pp. 473-477, 2010.

[Tompkins et al., 1985] Pan, Jiapu, Tompkins, Willis J., "*Real-time QRS detection algorithm*", IEEE Transactions on Biomedical Engineering, BME-32 (3), pp. 230-236, 1985.

[Zhang et al., 2009] Zhang, Y., Guo, J., Xiao, N., Zhao, Y., "*Risk-based safety management on railway system*" Proceedings of the 2nd International Conference on Transportation Engineering, ICTE 2009, pp. 2839-2844, 2009.

[Zio et al., 2009] Zio E., Di Maio F., "*Processing Dynamic Scenarios from a Reliability Analysis of a Nuclear Power Plant Digital Instrumentation and Control System*", Annals of Nuclear Energy, Volume 36, pp.1386-1399, 2009.

[Zio et al., 2012] E. Zio, F. Di Maio, "*Fault Diagnosis and Failure Mode Estimation by a Data-Driven Fuzzy Similarity Approach*", International Journal of Performability Engineering, Vol. 8, No.1, pp. 49-66, 2012.

[Wang et al., 2010] Wang L., Fu X., Menhas M.I., "*A Modified Binary Differential Evolution Algorithm*", Life Modelling and Intelligent Computing, Lecture Notes in Computer Science, Volume 6329/2010, 2010.

[Widodo et al., 2007] Widodo A, Yang BS, "*Support vector machine in machine condition monitoring and fault diagnosis*", Mech Syst Fault Diagn 21:2560–2574, 2007.