An Approach to Grouping and Classification of Scenarios in Integrated Deterministic-Probabilistic Safety Analysis

Sergey Galushin[a*], Pavel Kudinov[a]
[a] Royal Institute of Technology (KTH), Stockholm, Sweden

Abstract: Integrated Deterministic Probabilistic Safety Assessment (IDPSA) methodologies aim to achieve completeness and consistency of the analysis. However, for the purpose of risk informed decision making it is often insufficient to merely calculate a quantitative value for the risk and its associated uncertainties. IDPSA combines deterministic model of a nuclear power plant with a method for exploration of the uncertainty space. Huge amount of data is generated usually in the process of such exploration. It is very difficult to "manually" process and extract from such data information that can be used by a decision maker for risk-informed characterization and eventually improvement of the system safety and performance. Such understanding requires an approach to the interpretation, grouping of similar scenario evolutions, and classification of the principal characteristics of the events that contribute to the risk. In this work we develop an approach for classification and characterization of failure domains (domains of uncertain parameters where critical system parameters exceed safety thresholds). The method is based on scenario grouping and clustering with application of decision trees for characterization of the influence of timing and order of the events.

Keywords: Dynamic PSA, Clustering, Classification Risk Informed Decision making, Decision Trees

INTRODUCTION

Dynamic methodologies in probabilistic safety assessment (PSA) employ system simulators codes with explicit consideration of time in the system evolution. This is necessary to account for the effects of process/hardware/software/human actions on the stochastic system behavior [1]. A review of IDPSA methodologies applied for safety assessment of nuclear power plants can be found in [1]. Integrated Deterministic Probabilistic Safety Assessment (IDPSA) methodologies aim to achieve completeness and consistency of the analysis. For decision making, however, it is often insufficient to merely calculate a quantitative value for the risk and its associated uncertainties [2]. IDPSA methodologies combine deterministic model of a nuclear power plant with a method for exploration of the uncertainty space which result in huge amount of data generated in the process [3].

For application of IDPSA methodologies one of the main problems is data post-processing and extraction of useful information that can be used by a decision maker for risk-informed characterization and eventually improvement of the system safety and performance. Such understanding requires an approach to the interpretation, grouping of similar scenario evolutions, and classification of the principal characteristics of the events that contribute to the risk.

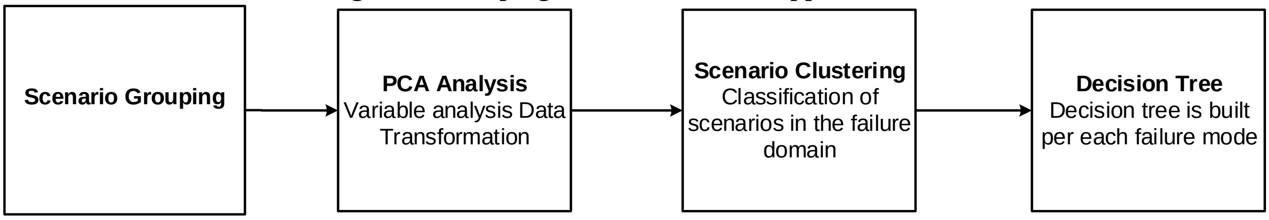
Several attempts to overcome this issue has been taken, particularly approaches to transient identification based on pattern classification by fuzzy C-means clustering [4], identification and classification of dynamic event tree scenarios via possibilistic clustering [5], probabilistic clustering for scenario analysis [6]. These methodologies use clustering tools for post-processing data to find patterns (scenarios that behave in a similar way are grouped into cluster/pattern) of system behaviors that lead to failure.

In this work we develop an approach for classification and characterization of failure domains (domains of uncertain parameters where critical system parameters exceed safety thresholds). The method is based on scenario grouping and clustering with application of decision trees for characterization of the influence of timing and order of the events. In this approach decision trees are constructed to represent failure domain as a set of leaf nodes and correspondent classification rules that lead to each node.

1. CLASSIFICATION APPROACH

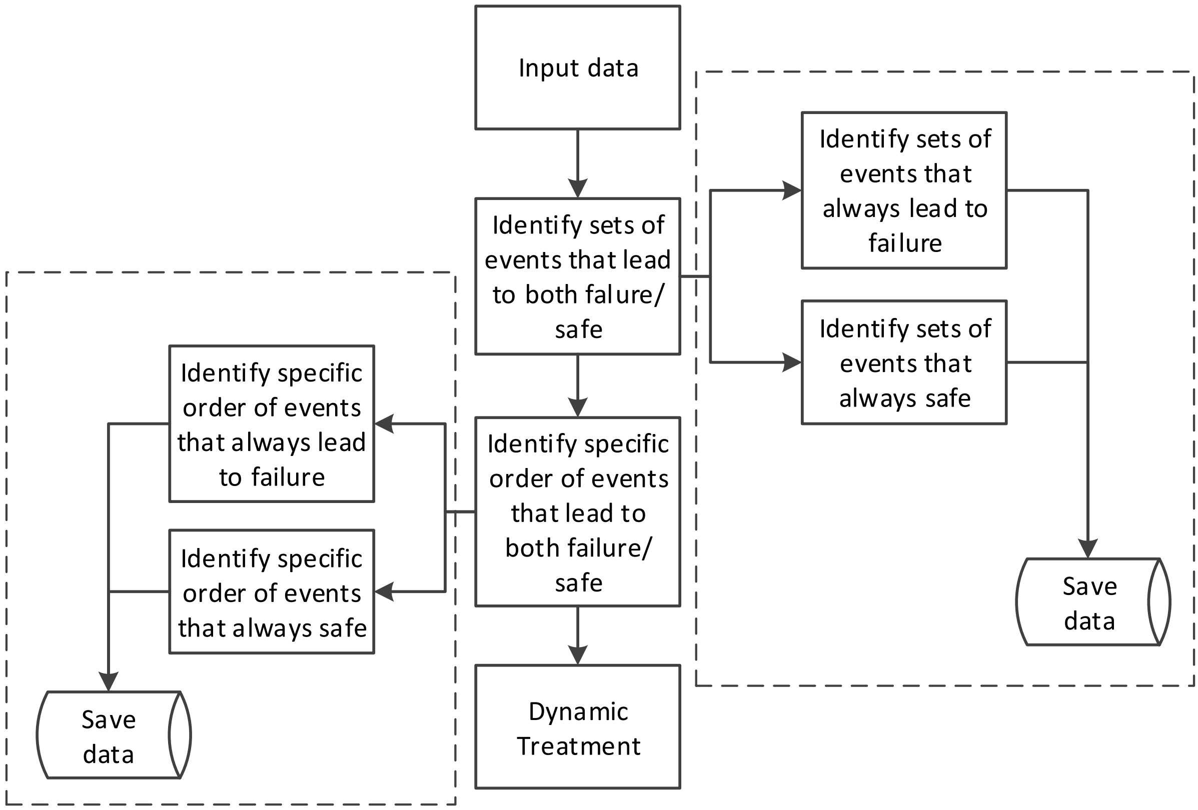
One of the major issues in using dynamic methodologies is a large number of scenarios that can be produced by a single initiating event. For decision making, it is often insufficient to merely calculate a quantitative value for the risk and its associated uncertainties, the extraction of useful information which can be appreciated and handled by a decision maker is a challenge. The development of risk insights that can improve system safety and performance requires the interpretation of scenario evolutions and the principal characteristics of the events that contribute to the risk.
The approach used to resolve this problem is based on decision tree built using clustering results data, to explain cluster structure attending to the values of uncertain parameters.

Figure 1. Grouping and Classification approach



The main steps of this approach are briefly explained below. Firstly, the scenario grouping is performed. The main idea of this step is to focus the analysis on the sequences intractable in classical PSA. Thus scenarios where the order and timing of events are not important are grouped first an excluded from further considerations as directly amenable to PSA analysis. Then scenarios where the order of events is important but not their timing are grouped. Remaining group of scenarios contains sequences where the outcome depends on the order and timing of the events.

Figure 2. Scenario grouping algorithm



Next, a Principal Component Analysis (PCA) is carried out. PCA is a technique for revealing the relationships between variables in a data set by identifying and quantifying a group of principal components which have the largest influence on the system response [7]. Then, based on the PCA results the clustering analysis is performed using adaptive mesh refinement (AMR) method. In the final step a decision tree is built for each failure mode using clustering results data [8]. Decision tree is used for data representation that explains failure domain-cluster structure. Decision tree classification algorithm performs orthogonal partitioning of the search space using data impurity measure as a splitting criterion [7, 9]. The main purpose of decision tree application is to present data in easy to interpret and transparent way to a decision maker to support decision making process. Finally, information of the leaf nodes is used for failure domain probability calculation.

2. APPLICATION OF DECISION TREES

A decision tree is a classification and data-mining tool for extraction of useful information contained in large data sets. It can be used to help in decision-making process. Decision Tree is a flow-chart like structure in which internal nodes represent test on an attribute, each branch represents outcome of test and each leaf node represents class label (decision taken after computing all attributes). A path from root to leaf represents classification rules. Decision threes also can be used as a powerful visual and analytical decision support tool, especially in case of multidimensional data, where it is impossible to visualize results in the original space. Decision tree can be constructed using different data impurity measures (e.g. Gini impurity measure, information gain measure) to select the best split among the candidate attributes at each step while growing the tree [10].

Although the main purpose of decision tree in this approach is to present data in easy to interpret and transparent way to a decision maker to facilitate a decision making process, decision tree also can be used as a predictive model which maps observations about an item to conclusions about the item's target value [10].

3. PROBABILITY ESTIMATION USING DECISION TREES

Given an uncertainty space split into non-overlapping cells (grids) the failure domain is represented by agglomerations (clusters) of these cells. Therefore, the probability of being in the failure domain is related to the volume of the failure domain $p \sim V_{failure}$, where $V_{failure}$ - total volume of cells in a cluster.

When growing decision tree the failure domain becomes represented by final nodes in the tree and correspondent classification rules, so taking into account cells probabilities as average probability values of scenarios contained in correspondent cell:

$$\overline{p}_k = \frac{\sum_{i=1}^{Nscen} p_i}{Nscen} \tag{1}$$

the probability of being in a failure mode *i* is as follows:

$$P_i = \sum_{j=1}^{N} \sum_{k=1}^{M_j} \overline{p}_k \xi^n \tag{2}$$

where n – dimensionality, ξ^n - cell volume, V_{total} - total volume, $\overline{p}_k$ - is average probability of scenarios contained in cell k, M_j - cells contained in the final failure node (leaf) j and N – total amount of failure nodes (leafs). Depending on the values $\bar{p}_k$ it is possible to assign weights per each cell when building a tree, so the scenarios (cells) with higher probability are likely to be classified into the same final node.

4. APPLICATION EXAMPLE

As an example case to illustrate the presented approach we chose a benchmark exercise performed in the work frame of the SARNET research network [11]. The exercise is based on a hypothetical transient of a typical French 900 MWe PWR (3 loops, with Passive Autocatalytic Recombiners – PAR). The transient description:

- Loss of Coolant accident (LOCA) with a 3’’ break size on cold leg of RCS (INI – initiation event),
- The safety injection system and Containment Heat Removal System (CHRS or spray system) are not available until the beginning of core dewatering,
- The steam generators are available but not used by the operators,
- No water injection occurs before core dewatering (SIS – water injection event),
- The reactor is operating at nominal power before the initiating event.
- The calculated core dewatering occurs at 4080 s (1h08mn). The vessel rupture occurs at 14220 s (3h57min) if no action is undertaken.

During the core degradation phase, the situation is supposed to be as follows:

- A water injection (SIS) means is available (with an “average” flow rate) and can be used by the operators,
- The spray system (CHRS) is available and can be used by the operators,
- water injection after the beginning of clad oxidation causes an increase of the hydrogen flow rate towards containment,
- Hydrogen combustions (IGNI) can occur if the containment gas mixture is flammable; recombiners, because of their high temperature, can initiate a combustion ; such combustions can be total (all the hydrogen in the containment is burnt) or not.

For the test case, to make it simple enough but yet having a physical meaning we took only containment pressure of P_{Lim} = 0.3MPa threshold as a failure criterion. Using Monte Carlo sampling over 443200 scenarios have been generated for INI(initiating event) + all possible combinations of SIS, CHRS, IGNI).

5. RESULTS

Performing grouping analysis we identified the following possible sequences: INI SIS; INI SIS CHRS; INI SIS IGNI; INI SIS CHRS IGNI; INI CHRS; INI CHRS SIS; INI CHRS IGNI; INI CHRS SIS IGNI; INI CHRS IGNI SIS;

Analyzing classification results it has been found that sets [INI,CHRS], [INI,SIS] and [INI,CHRS,SIS],[INI,SIS,CHRS] do not cause containment over pressurization when not followed by hydrogen ignition event (IGNI).

Sequences [INI CHRS IGNI], [INI CHRS IGNI SIS] – also do not generate the pressure spike big enough to cause containment failure.

In sequences [INI SIS IGNI], [INI SIS CHRS IGNI] and [INI CHRS SIS IGNI] the outcome depends on the time when safety systems have been actuated and stochastic event (IGNI) have occurred.

In figures below we illustrate an example of application of clustering analysis and decision trees for the sequences that require dynamic treatment.

Figure 3. Scenario Grouping

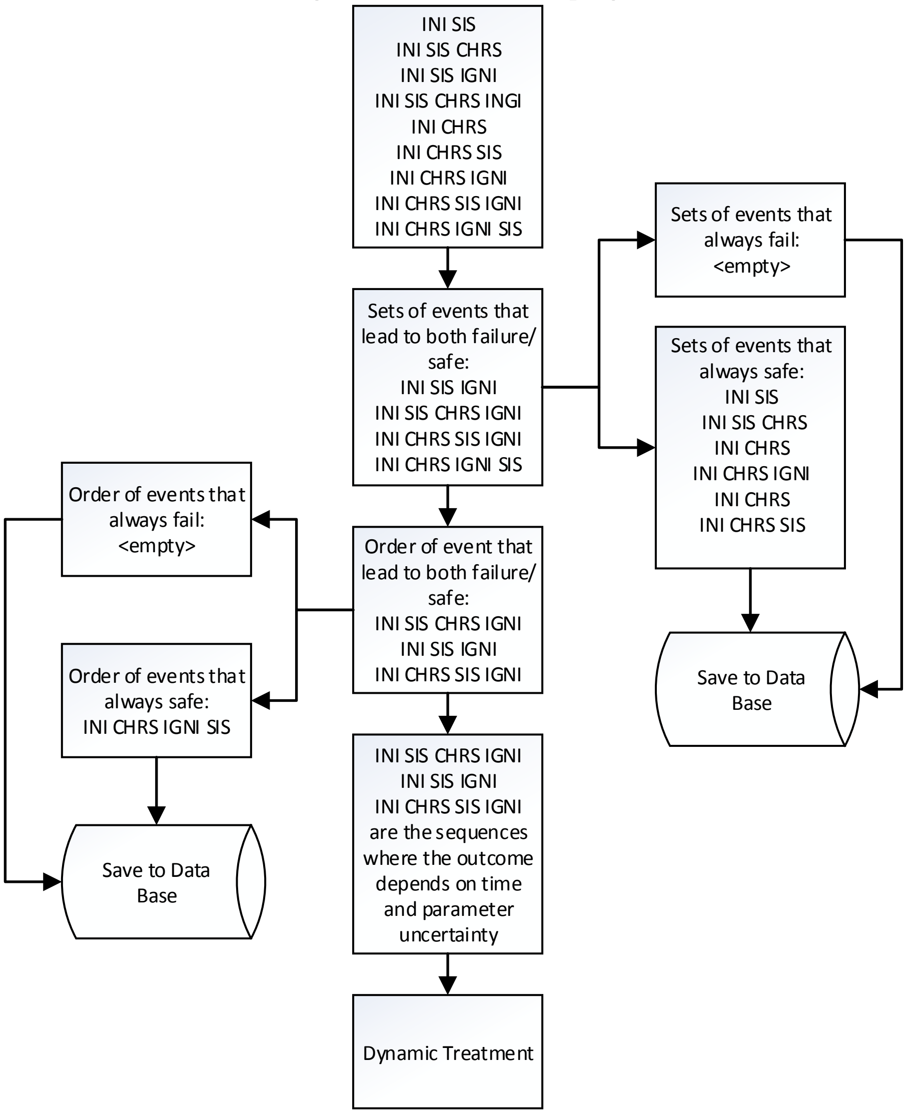


Figure 4. Clustering analysis results for sequence [INI SIS IGNI], no refinement.

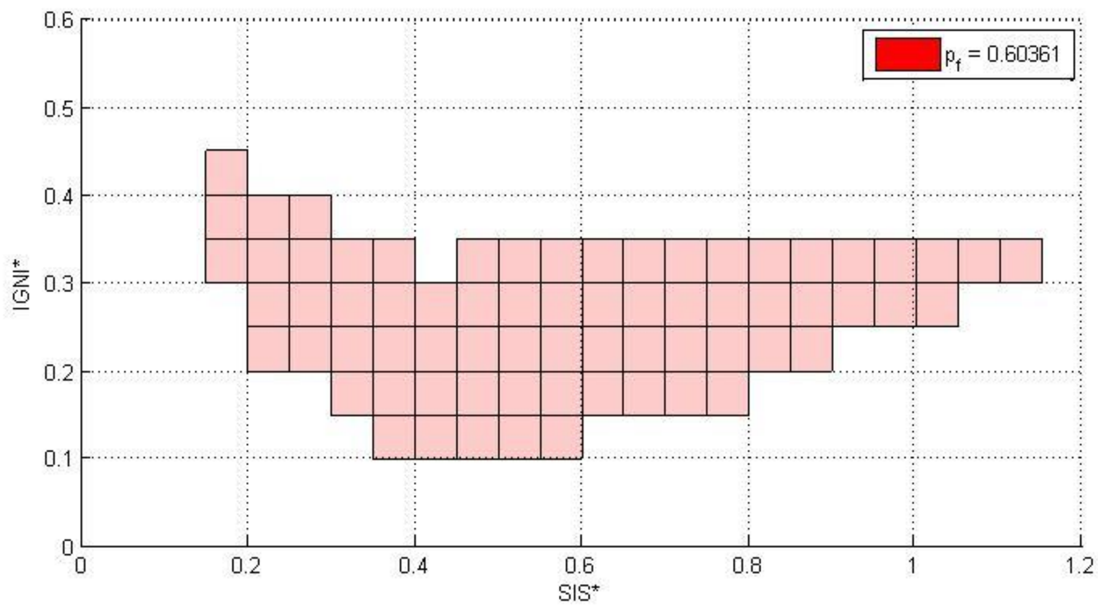
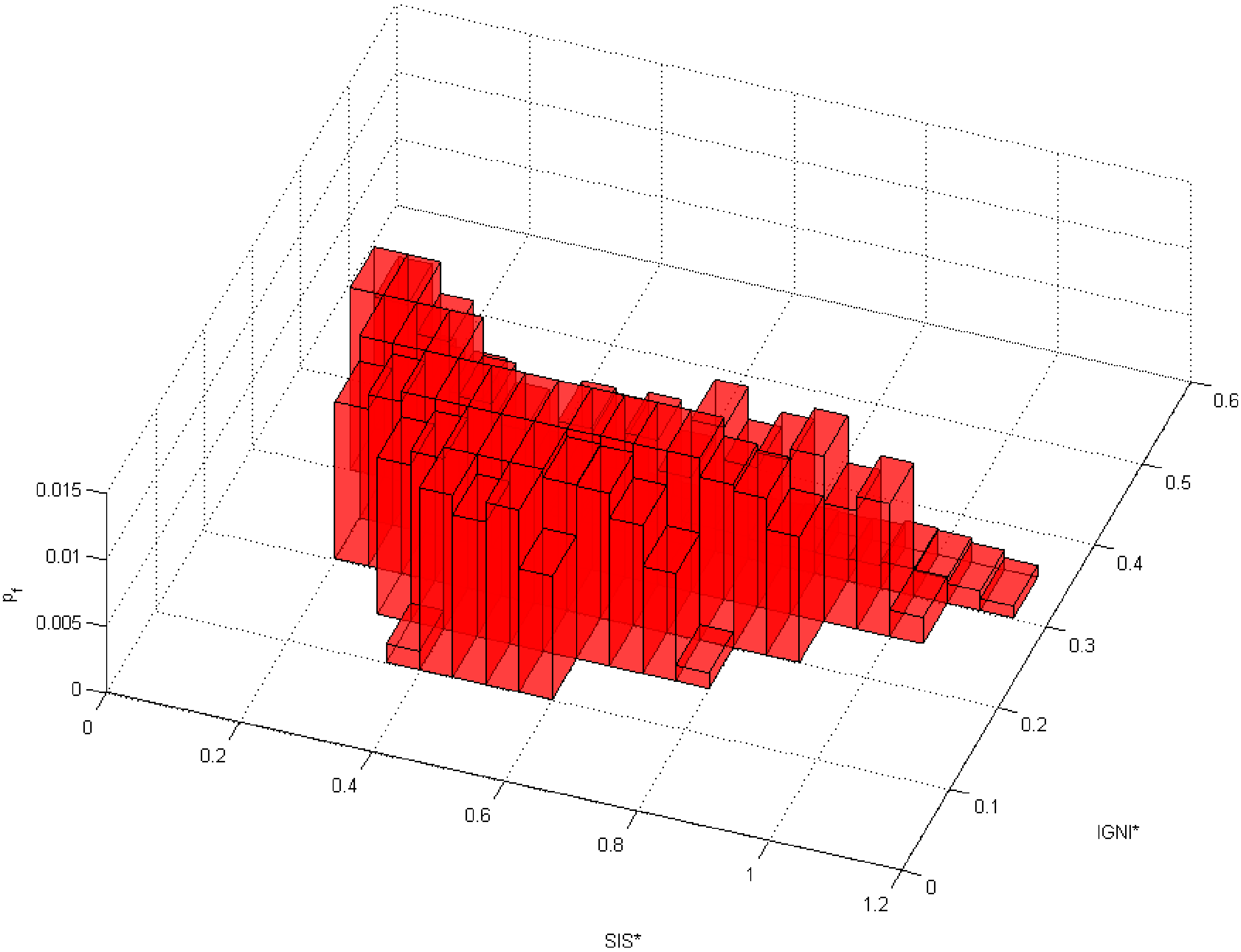


Table 1. Containment failure probabilities

Sequence	$p(P>P_{Lim})$
[INI,SIS,IGNI]	0.51379
[INI,SIS,CHRS,IGNI]	0.07221
[INI,CHRS,SIS,IGNI]	0.00189

illustrates results of clustering analysis for the sequence [INI SIS IGNI] with static grid, cells that contain failure scenarios inside are grouped into cluster that represent failure domain. For each cell in the cluster the algorithm calculates correspondent probability, see figure below.

Figure 5. Containment failure probability distribution for sequence [INI SIS IGNI]



Different values of probabilities in the different parts of the failure domain are caused by different H_2 concentrations, that have, according to benchmark specifications (see reference [11]), different probabilities distribution for time delays for IGNI event to occur, see figure below.

Figure 6. H2 molar fraction (%), H2 inflammability and ignition limits (%).

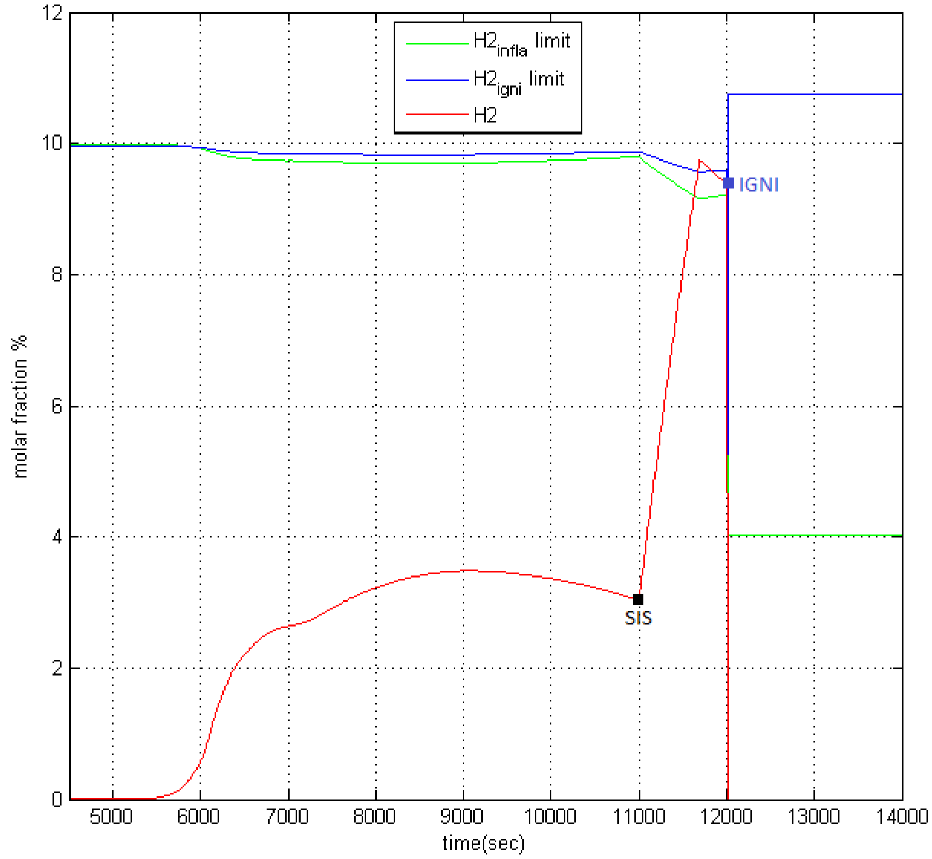
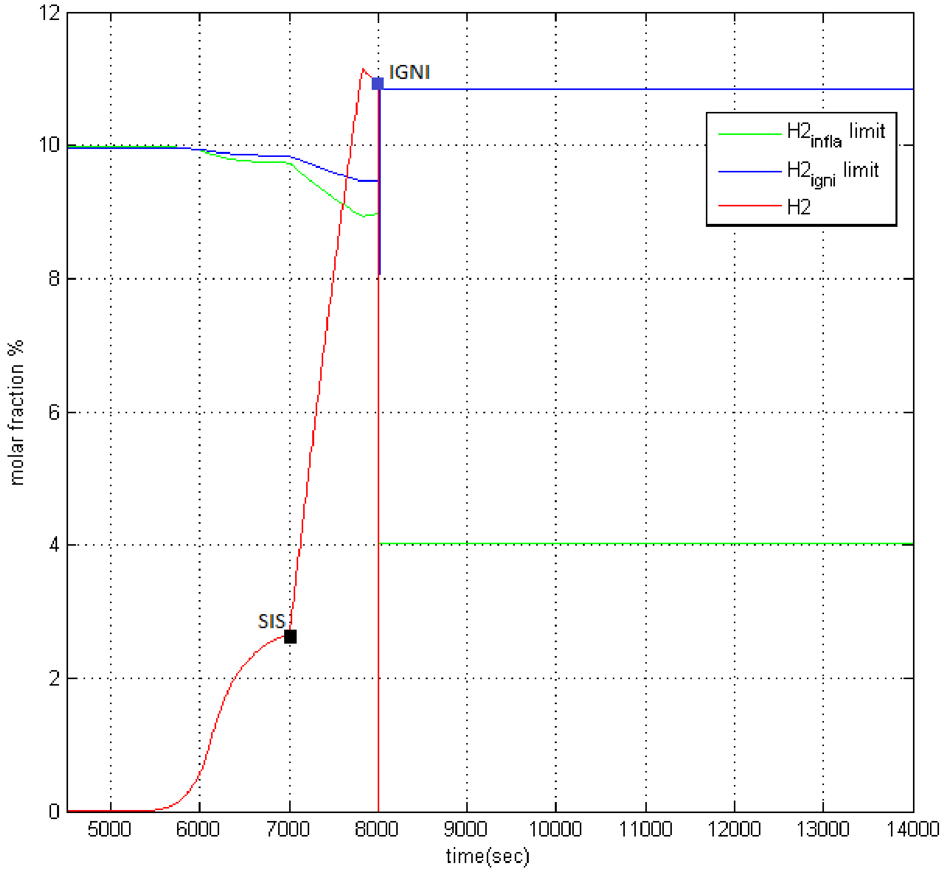
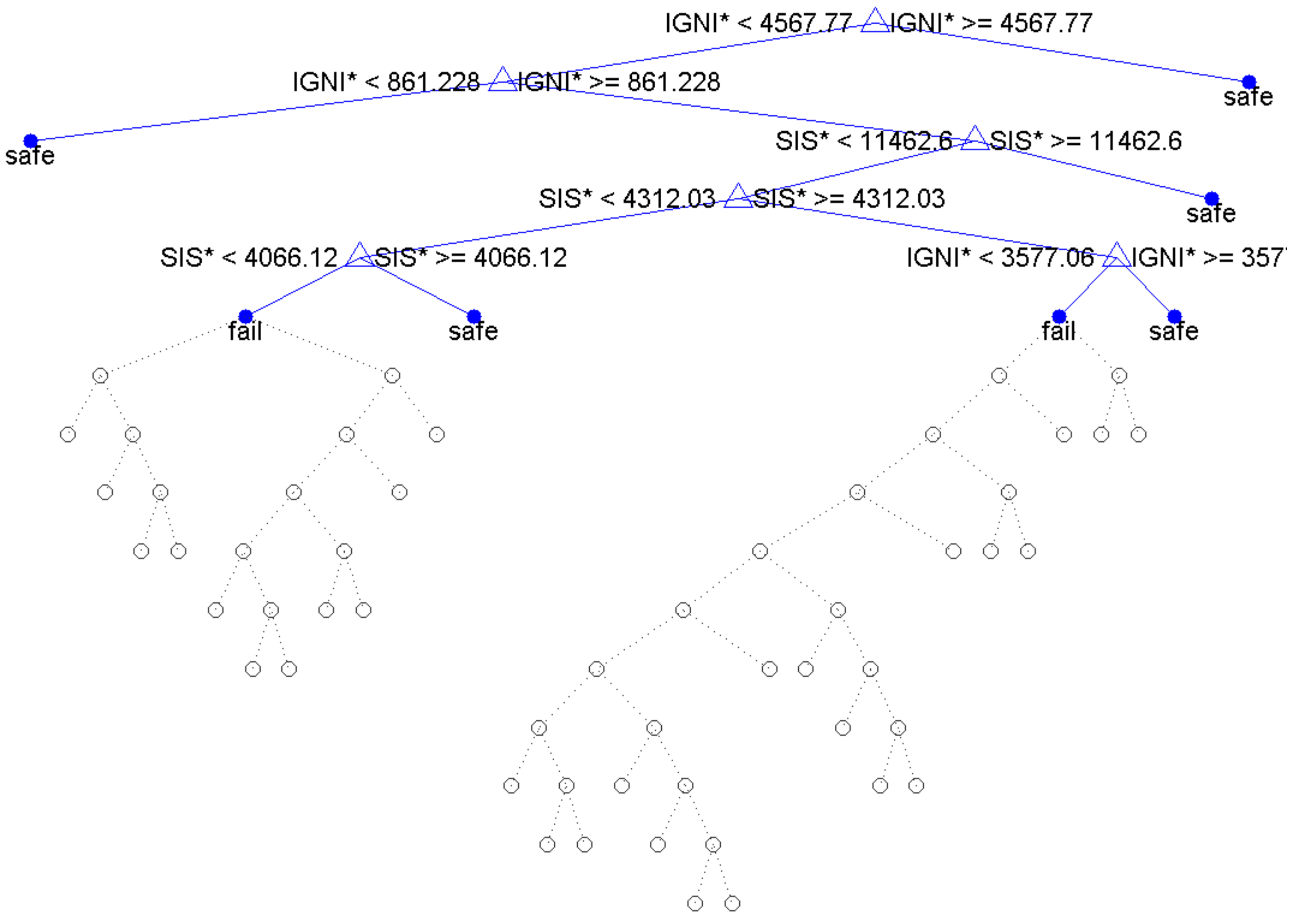


Figure 7. H2 molar fraction (%), H2 inflammability and ignition limits (%).



Using clustering results data we build a decision tree, to represent failure domain structure. In this work, to illustrate presented approach, we use only two uncertain parameters, however, the main advantage of using decision trees is the ability to represent complex failure domains with 4 or more uncertain parameters, when it is difficult to visualize results using other methods. Decision tree complexity depends on the shape of the failure domain and level of details (initial grid and refinement step). However, it is possible to prune decision tree, so the complexity and precision are both at the acceptable levels.

Figure 8. Decision tree fitted into clustering results data for the sequence [INI SIS IGNI] with pruning.



6. DISCUSSION

In this work we present an approach for grouping and classification of typical “failure/safe” scenarios identified using IDPSA methods. This approach allows to classify scenarios that are directly amenable in classical PSA and scenarios where order of events, timing and parameter uncertainty affect system evolution and determine violation of safety criteria.

We use grid based clustering with AMR and decision trees for failure domain characterization. Clustering analysis is used to represent failure domain as a finite set of representative scenarios (cell centers) grouped into clusters based on failure criteria. Decision trees are used to visualize failure domain structure as a set of classification rules that lead to the leaf nodes that represent the outcome for each set of classification rules. Decision trees can be applied to the cases where 4 or more uncertain parameters are included in the analysis and it is difficult to visualize results using 3D space.

This approach can serve as a tool for establishing connection between classical PSA methodologies and advanced Risk-Informed approaches to facilitate Decision Making. It can provide useful insights into likelihood of various accident scenarios, accident progression and can be used for development of understanding, managing and mitigation of complex plant behaviour including severe accident scenarios.

Acknowledgements

This study was supported by the Swedish Radiation Safety Authority (SSM). Authors are grateful to Dr. Wiktor Frid (SSM) for very useful discussions.

References

1. Aldemir, T., *A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants.* Annals of Nuclear Energy, 2013. **52**: p. 113-124.
2. Hess, S., *Framework for Risk-Informed Safety Margin Characterization, EPRI Report Number 1019206*. 2009, EPRI: Palo Alto, CA.
3. P.E. Labeau, C. Smidts, and S. Swaminathan, *Dynamic reliability: towards an integrated platform for probabilistic risk assessment.* Reliability Engineering and System Safety, 2000. **68**(3): p. 219 - 254.
4. E. Zio and P. Baraldi, *Identification of nuclear transients via optimized fuzzy clustering.* Annals of Nuclear Energy, 2005. **32**: p. 1068–1080.
5. Mercurio, D., et al., *Identification and classification of dynamic event tree scenarios via possibilistic clustering: application to a steam generator tube rupture event.* Accident Analysis and Prevention, 2009. **41**(6): p. 1180-91.
6. D. Mandelli, *Scenario Clustering and Dynamic PRA*, in *Nuclear Engineering Department*. 2011, The Ohio State University.
7. Tuffery, S.p., *Data mining and statistics for decision making*. Wiley series in computational statistics. 2011, Chichester, West Sussex ; Hoboken, NJ.: Wiley. xxiv, 689 p.
8. Ilango and V. Mohan, *A Survey of Grid Based Clustering Algorithms.* International Journal of Engineering Science and Technology, 2010. **Vol. 2(8)**.
9. Mitchell, T.M., *Machine Learning*. McGraw-Hill series in computer science. 1997, New York: McGraw-Hill. xvii, 414 p.
10. T.M. Mitchell, *Machine Learning*. 1 ed. 1997, New York, NY, USA: McGraw-Hill, Inc.
11. Raimond, E., *SARNET workpackage 5.3 – Level 2 PSA Specification of a benchmark exercise relative to hydrogen combustion for application of dynamic reliability methods, IRNS/DSR/SAGR/FT.2005-154*, in *Network of Excellence for a Sustainable Integration of European Research on Severe Accident Phenomenology*. 2005, IRNS.