Proceedings of ICAPP '10
San Diego, CA, USA, June 13-17, 2010
Paper 10293

Scenario Aggregation and Analysis via Mean-Shift Methodology

Diego Mandelli*, Alper Yilmaz, Kyle Metzroth, Tunc Aldemir, Richard Denning

The Ohio State University
Nuclear Engineering Program
201 West 19th Avenue, Columbus, OH 43210
**mandelli.1@osu.edu*

Abstract – *A new generation of dynamic methodologies is being developed for nuclear reactor probabilistic risk assessment (PRA) which explicitly account for the time element in modeling the probabilistic system evolution and use numerical simulation tools to account for possible dependencies between failure events. The dynamic event tree (DET) approach is one of these methodologies. One challenge with dynamic PRA methodologies is the large amount of data they produce which may be difficult to analyze without appropriate software tools. The concept of "data mining" is well known in the computer science community and several methodologies have been developed in order to extract useful information from a dataset with a large number of records. Using the dataset generated by the DET analysis of the reactor vessel auxiliary cooling system (RVACS) of an ABR-1000 for an aircraft crash recovery scenario and the Mean-Shift Methodology for data mining, it is shown how clusters of transients with common characteristics can be identified and classified.*

I. INTRODUCTION

A new generation of dynamic methodologies is being developed for nuclear reactor probabilistic risk assessment (PRA) which explicitly accounts for the time element in modeling the probabilistic system evolution and use numerical simulation tools to provide possible dependencies between failure events. A challenging aspect of these methodologies, such as the dynamic event tree (DET) methodology[1], is the large number of scenarios generated for a single initiating event. Such large amounts of information can be difficult to organize in order to extract useful information. In particular, as part of the PRA framework, it is important to identify the main scenarios that are the most significant risk contributors.

Each scenario obtained from DET analysis contains information on all the system components and the system process variables, such as the spatial and temporal distribution of pressure and temperature in the reactor coolant and the containment. In scenario aggregation, we are trying to accomplish two tasks:

- Identify the scenarios that have "similar" behaviors (i.e., identify the most evident clusters)
- Assign each scenario to a cluster (i.e., classification)

The objective of this paper is to illustrate an approach to group and classify a set of scenarios generated by a DET methodology. This approach is based on the Mean-Shift algorithm developed by Fukunaga and Hostetler[2]. We will show how to employ this methodology to a set of transient accident scenarios. As an application of this methodology, we will use the data generated by a previous study for the recovery of the reactor vessel auxiliary cooling system (RVACS) for an ABR-1000 reactor following an aircraft crash[3].

II. THE CLASSIFICATION PROBLEM

From a mathematical viewpoint, a clustering process attempts to search for a partition $\boldsymbol{C} = \{C_1, \dots, C_K\}$ of the set of patterns $\boldsymbol{X} = \{x_1, \dots, x_j, \dots, x_N\}$ where each pattern can be represented as a multi-dimensional vector $x_j = (x_{j1}, \dots, x_{j2}, \dots, x_{jd})$ and each component x_{ji} is said to be a feature (attribute, dimension or variable). The partition $\boldsymbol{C}$ of $\boldsymbol{X}$ is such that:

$$C_i \neq \emptyset \;\; \forall i = 1, \cdots, K$$
$$\bigcup_{n=1}^{K} C_i = \boldsymbol{X}$$
$$C_i \cap C_j = \emptyset \quad \forall i, j = 1, \cdots, K \quad and \;\; i \neq j$$

In our applications, we are dealing with transient scenarios and, thus, each x_i can be viewed as a trajectory distributed in the state space rather points in an n-dimensional space. However, we will show that we can sample all the scenarios at several time instants and consider the values of the variables of interest at these time instants as the generic measure x_{ji}.

III. METHODOLOGY PROPOSED

Mean-Shift Methodology (MSM) is a non-parametric iterative mode-seeking procedure that shifts each data point to the average of data points in its neighborhood. We adopt the MSM to set the modes as the cluster centers and to assign each point to one cluster center only. By cluster center we mean a region with high point density.

Starting from a generic point (i.e., point s_A in Fig. 1), the algorithm associates a hyper-dimensional sphere centered at that point. The radius of this sphere is called the bandwidth *BW*. The idea is to consider all the points that are inside this sphere and determine the center of mass of these points (point $m(s_A)$ in Fig. 1). In particular, given a point $s \in S$ in an n-dimensional Euclidean space, the center of mass of s_A is simply:

$$m(s_A) = \frac{\sum_{s\in S} K(s-s_A)s}{\sum_{s\in S} K(s-s_A)} \qquad (1)$$

where $K(x)$ is often referred to as the Kernel.

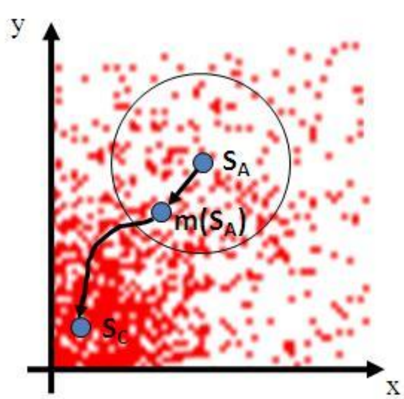

Fig. 1 Graphical representation of the Mean-Shift algorithm

The purpose of this function is its ability to assign different weights to different points during the estimation of the center of mass. Several Kernels can be used as illustrated in Ref. 4. If, for example, we use a 2-D flat kernel (see Fig. 2)

$$K(\vec{x}) = \begin{cases} 1 & \text{if } \|\vec{x}\| \le \text{BW} \\ 0 & \text{if } \|\vec{x}\| > \text{BW}, \end{cases} \qquad (2)$$

then the center of mass simply becomes the mean of the points within the disc shown in the upper plane of Fig. 2. In this work, we have used the biweight kernel[4] (see Fig. 3):

$$K(\vec{x}) = \begin{cases} (1-\|\vec{x}\|^2)^2 & \text{if } \|\vec{x}\| \le \text{BW} \\ 0 & \text{if } \|\vec{x}\| > \text{BW} \end{cases} \qquad (3)$$

After estimating the center of mass, the MSM determines the calculated position $m(s_A)$ (see Fig. 1), and repeats the calculation for the center of mass for the points included in the ball having an identical value of radius (i.e., *BW*) but now centered on $m(s_A)$.

This operation converges to the mode of the data distribution when the distance between the new center of mass and the old one is below a fixed threshold (i.e., s_C in Fig. 1 is reached). When this condition is reached:

- point s_C is considered the center of a cluster, and
- the original point s_A is uniquely assigned to the cluster centered by point s_C.

This procedure is repeated for all the points $s \in S$ to estimate:

- the center of all the clusters in *S*, and,
- the cluster to which each point belongs (each point belongs to one cluster only).

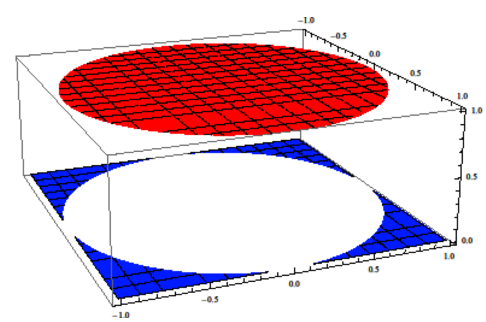

Fig. 2. Graphical representation of the 2-D flat kernel with *BW*=1

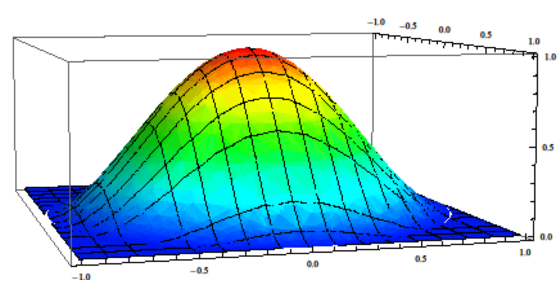

Fig. 3. Graphical representation of the 2-D biweight kernel with *BW*=1

An important component in clustering is selection of distance measure between the points. Considering that the set of points $s \in S$ lies in a metric space, the choice of distance should satisfy the metric properties. In Table 1, we list commonly used metrics from among which we use the Euclidean distance as our distance measure.

TABLE 1

Comparison of different metrics

Measures	Forms
Minkowski distance	$d(x,y) = (\sum_{k=1}^{d} \lvert x_k - y_k \rvert^n)^{\frac{1}{n}}$
Euclidean distance	$d(x,y) = (\sum_{k=1}^{d} \lvert x_k - y_k \rvert^2)^{\frac{1}{2}}$
Taxicab distance	$d(x,y) = \sum_{k=1}^{d} \lvert x_k - y_k \rvert$
Supremum distance	$d(x,y) = max_k \lvert x_k - y_k \rvert$

Aside from the choice of the distance measure, the geometry of the data points plays a vital role in the "point to cluster" decision process. Figure 4 illustrates the results of MSM when it is applied to a set of points distributed normally along two rings. As shown in Fig. 4, the data lying on two rings are clustered into two clusters denoted by red and green colors.

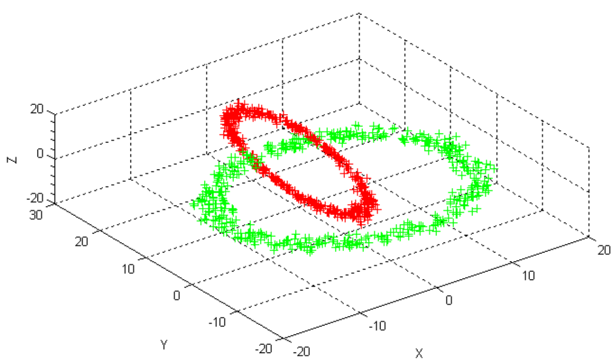

Fig. 4. MSM applied to a set of data distributed over two rings. Different colors denote different clusters generated by the MSM.

IV. DEMONSTRATION CASE

As mentioned in the Section 1, the system that is analyzed in this work is the RVACS of ABR-1000 reactor[3], schematically shown in Fig. 5. The RVACS is a passive decay-heat removal system that removes heat by natural circulation of air in the gap between the vessel and a duct surrounding the vessel. With this system, the reactor decay heat is released to the atmosphere through 4 towers.

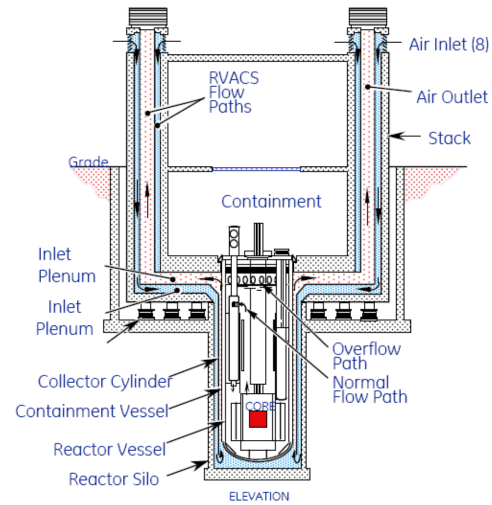

Fig. 5. RVACS system applied to ABR-1000

ADAPT (Analysis of Dynamic Accident Progression Trees)[1] is used here as the DET generator tool while the system dynamics are modeled using RELAP5[5]. At time zero with the plant operating at 100% power, an aircraft crashes into the plant. Three of the four towers are assumed to be destroyed, producing debris that blocks the air passages (hence, impeding the possibility to remove the decay heat). The reactor trips, offsite power is lost, the pump trips and coasts down.

A recovery crew and heavy equipment are used to remove the debris. Figure 6 illustrates the strategy that is followed by the crew in reestablishing the capability of the RVACS to remove the decay heat. Several assumptions have been made for the purpose of the analysis:

- A tower is assumed to have no heat removal capacity until the rubble has been removed. At that point it is assumed to regain full capacity.
- There is a one hour period following the crash in which a fire is being extinguished.
- There is a uniform probability of work being initiated between one and nine hours after the crash.
- The workers remove debris from one tower at a time.
- After work begins on a tower there is a minimum time of two hours to recover the tower.
- There is a uniform distribution of recovery between two and ten hours. The team then moves on to the next tower.
- The difficulty of recovering each tower is assumed to be independent of the other towers.

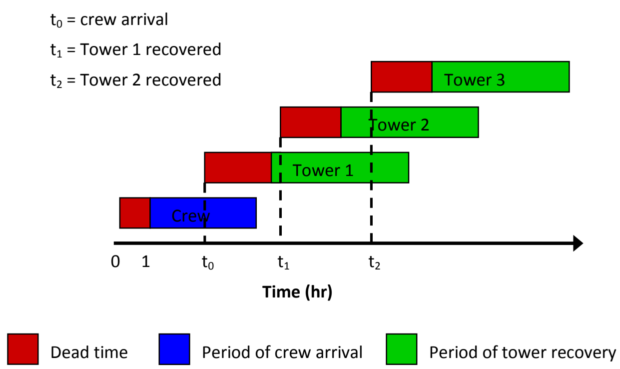

Fig. 6. Crew recovery strategy for the aircraft crash scenario[6]

For an actual application of dynamic event trees to the development of optimal accident management strategies, a substantial effort would be required to determine the durations and probability density functions employed in the analysis. The uncertainties in crew arrival time and tower recovery are treated as aleatory (stochastic) in nature and represented in the analysis by cumulative distribution functions. Branching occurs at the times associated with the probabilities 0.001, 0.25, 0.5, 0.75, and 1.0 (upper limit up recovery) on the cumulative distribution function. When the trigger time is reached, a branching occurs in which one branch represents tower recovery and the other branch represents non-recovery. The latter branch then continues until the next branching point is reached on the cumulative distribution function. ADAPT keeps track of the scenario probability for each pathway through the growing tree.

Only one Top Event is considered: temperature T of the core reaches the limit of 1000 K, associated with clad failure by eutectic formation. Figure 7 plots the temporal behavior of the temperature of the core for all the scenarios generated by ADAPT (about 500 different transient runs).

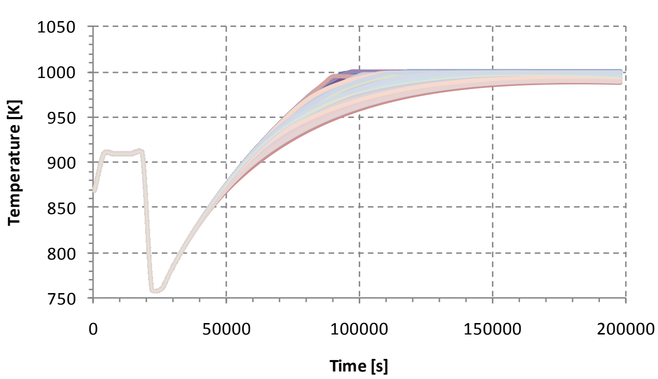

Fig. 7. Graphical representation of the scenarios generated by ADAPT for the aircraft crash scenario

Mission time for this system analysis has been fixed to 200,000 s beyond which time clad failure has either occurred or the temperature has peaked and is declining.

V. RESULTS

As mentioned in Section II, each transient is represented as a vector in an *n*-dimensional space. Each component of this vector corresponds to the value of the variables of interest sampled at a specific time instant[7]. For this example, we consider only the temperature T of the core as the variable of interest sampled in 56 time instants. Thus each scenario x_i can be represented as a vector in a 56-dimensional space as following:

$$x_i = [T(0), T(1), \dots, T(56)] \quad (4)$$

When dealing with DET methodologies several issues may arise at this point:

1. Branching might occur at different time instants depending when the branching rules are met during a transient evolution.
2. When the failure criterion (fuel temperature reaches 1000 K) or the mission time is reached, the simulation stops. Hence the length of the scenarios might vary.

The first issue is resolved by adding a special script inside ADAPT to print results of each scenarios at specific instants. The second issue was resolved for this example by assuming that for those scenarios in which the failure temperature of 1000 K was achieved the temperature is assumed to remain fixed until the end of the mission time.

Figures 8 and 9 show the cluster centers obtained for two different values of bandwidth BW (40 and 30 respectively). A cluster center can be seen as the representative scenario for a subset of scenarios (i.e., a cluster of scenarios) where the size of the cluster depends on the value of the bandwidth. In both Figures 8 and 9, the numbers in the legend indicate the number of scenarios that are part of each specific cluster. In this analysis, more than 80% of the scenarios lead to fuel failure. For the broader bandwidth, the clusters appear to represent the two key differences obtained in the analysis, events that lead to fuel failure and events that do not lead to fuel failure. However, a large number of failure scenarios are mapped into what could be interpreted as the non-failure group. For the narrower bandwidth, there is better discrimination between the groups.

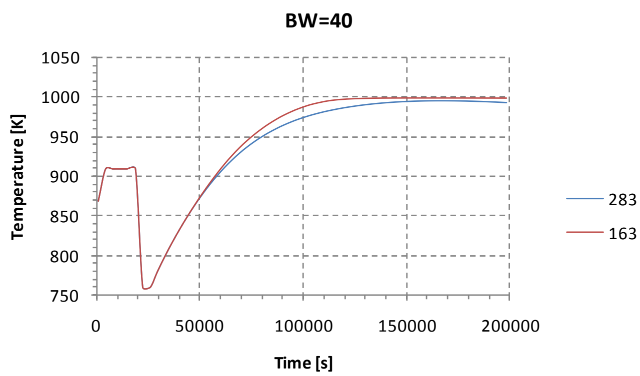

Fig. 8. Cluster centers for *BW*=40

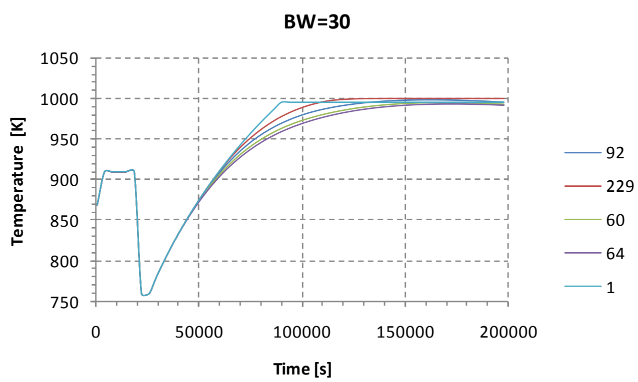

Fig. 9. Cluster centers for *BW*=30

VI. CONCLUSIONS

This paper illustrates a practical methodology to analyze the set of transients generated by the DET approach. We have illustrated how it is possible to group the scenarios in clusters and to assign uniquely each scenario to one cluster. We also demonstrated the application of Mean-Shift analysis to a set of scenarios generated by the DET analysis of the reactor RVACS of an ABR-1000 for an aircraft crash recovery scenario. ADAPT was used as a DET generator tool while the system dynamics are modeled using RELAP5.

VII. ACKNOWLEDGMENTS

"This work is a product of the project "Risk-Informed Balancing of Safety, Non-Proliferation, and Economics for the Sodium-Cooled Fast Reactor (SFR)" supported by the US Department of Energy under a NERI contract (DE-FG07-07ID14888). The views presented here are those of the authors and do not necessarily represent the views of the US Department of Energy."

REFERENCES

1. B. RUTT, U. Catalyurek, A. Hakobyan, K. Metzroth, T. Aldemir, R. Denning, S. Dunagan, D. Kunsman, "Distributed Dynamic Event Tree Generation for Reliability and Risk Assessment", 1-4244-0420-7/06, *IEEE* (2006).

2. K. FUKUNAGA, L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition", *IEEE Transactions on Information Theory*, **21**, 32-40 (1975).

3. R. WINNINGHAM, K. Metzroth, T. Aldemir, R. Denning, "Aircraft Crash Recovery Scenario Dynamic Event Tree Analysis of the RVACS Passive Decay Heat System Employing the ADAPT tool with RELAP5-3D", *Proceedings of American Nuclear Society*, (2009).

4. Y. CHENG, "Mean Shift, Mode Seeking, and Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, **17**, no. 8, (1995).

5. The RELAP5-3D Code Development Team, RELAP5-3D Code Manual, Volume I: Code Structure, System Models, and Solution Methods, INEEL-EXT-98-00834, June 2005.

6. R. D. WINNINGHAM, K. Metzroth, T. Aldemir, R, Denning, "Passive Heat Removal System Recovery following an Aircraft Crash using Dynamic Event Tree Analysis", *Trans. Am. Nucl. Soc.*, **100**, 461 - 462 (June 2009)

7. D. MANDELLI, T. Aldemir, A. Yilmaz, "Scenario Aggregation in Dynamic PRA Uncertainty Quantification", under review for potential publication in *Trans. Am. Nucl. Soc.*, 102, (June 2010).