Analyzing Simulation-Based PRA Data through Topology-Based Clustering: a BWR Station Black Out Test Case

Dan Maljovec\(^a\)*, Shusen Liu\(^a\), Bei Wang\(^a\), Valerio Pascucci\(^a\), Peer-Timo Bremer\(^b\), Diego Mandelli\(^c\), and Curtis Smith\(^c\)

\(^a\)SCI Institute, University of Utah, Salt Lake City, USA
\(^b\)Lawrence Livermore National Laboratory, Livermore, USA
\(^c\)Idaho National Laboratory, Idaho Falls, USA

Abstract: Dynamic Probabilistic Risk Assessment (DPRA) methodologies couple system simulator codes (e.g., RELAP, MELCOR) with simulation controller codes (e.g., RAVEN, ADAPT). While system simulator codes accurately model system dynamics (deterministically), simulation controller codes introduce both deterministic (e.g., system control logic, operating procedures) and stochastic (e.g., component failures, parameter uncertainties) elements into the simulation. Typically, a DPRA is performed by: 1) sampling values of a set of parameters from the uncertainty space of interest (using the simulation controller codes), and 2) simulating the system behavior for that specific set of parameter values (using the system simulator codes). For complex systems, one of the major challenges in using DPRA methodologies is the large amount of information (i.e., large number of scenarios) generated. In this paper, we present a software tool that provides the domain experts with an interactive analysis and visualization environment for understanding the structures of high-dimensional nuclear simulation datasets. We compare classic clustering techniques to an algorithm based on the topological structure, Morse-Smale complex, which partitions the data points into clusters based on their uniform gradient flow behavior. For visualization, we rely on dimensionality reduction techniques that summarize the structure we compute on the high-dimensional input space. We describe our results for a nuclear simulation dataset that is part of the Risk Informed Safety Margin Characterization (RISMC) Boiling Water Reactor (BWR) station blackout (SBO) case study.

Keywords: PRA, computational topology, clustering, high-dimension analysis

* Corresponding author, maljovec@cs.utah.edu

1. INTRODUCTION

A recent trend in the nuclear power engineering field is the implementation of computationally heavy and time consuming algorithms and codes for both design and safety analysis. In particular, the new generation of system analysis codes aim to embrace several phenomena such as thermo-hydraulic, structural behavior, system dynamics and human behavior, as well as uncertainty quantification and sensitivity analyses associated with these phenomena. The use of dynamic probabilistic risk assessment (PRA) methodologies allows a systematic approach to uncertainty quantification.

Dynamic PRA (DPRA) methodologies account for possible coupling between triggered or stochastic events through explicit consideration of the time element in system evolution, often through the use of dynamic system models (simulators). They are usually needed when the system has more than one failure mode, control loops, and/or hardware/process/software/human interaction. Typically, a DPRA is performed by 1) sampling values of a set of parameters from the uncertainty space of interest (using the simulation controller codes), and 2) simulating the system behavior for that specific set of parameter values (using the system simulator codes).

Due to the intrinsic high level of details of such analyses, very large amounts of data are produced. Hence, the need for methodologies able to handle high volumes of data arises. In [9] we presented a...
methodology that aims to visualize high dimensional data through topological segmentation of multi-
dimensional surfaces. In our applications, such multi-dimensional spaces are determined by the set of \( n \) uncertain parameters \( x_1, x_2, \ldots, x_n \) while safety related outcomes \( f \), such as maximum core temperature, are considered for each simulation. Our topological tools aim to reconstruct the topological structure of \( f(x_1, x_2, \ldots, x_n) \) (i.e., the response surface) in this \( n \) dimensional space.

In this paper, we show some results for a nuclear simulation dataset using our analysis and visualization tool. The dataset is part of the Risk Informed Safety Margin Characterization (RISMC) Boiling Water Reactor (BWR) station blackout (SBO) case study [10]. We investigate the use of classic dimensionality reduction, hierarchical clustering, and a topology-based clustering technique on such a dataset. We believe such analysis and visualization present the user with views that could be considered complementary to the more standard clustering techniques of the data. Such an analysis could help illuminate key features that may be otherwise hidden using only traditional techniques.

1.1. BWR System

The system considered in this test case is a generic BWR power plant with Mark I containment. The three main structures are: 1) the Reactor Pressure Vessel (RPV), a pressurized vessel that contains the reactor core, 2) the primary containment including the Drywell (DW) that houses the RPV and circulation pumps, and 3) the Pressure Suppression Pool (PSP) also known as wetwell. The wetwell is a large torus shaped container that contains a large amount of water (almost 1 M gallons of fresh water) and is used in specific situations as an ultimate heat sink. While the original BWR Mark I includes a large number of systems, for the scope of this report and for the test case considered, we will consider a smaller subset of systems that include the RPV level control systems, the RPV pressure control systems, the cooling water inventory, and the AC power system which consists of two power grids, emergency Diesel Generators (DGs), and battery systems (for instrumentation and control systems).

The RPV level control systems provide manual and automatic control of the water level within the RPV and consists of two components, the Reactor Core Isolation Cooling (RCIC) and the High Pressure Core Injection (HPCI). The RCIC provides high-pressure injection of water from the CST to the RPV. Water flow is provided by a turbine-driven pump that takes steam from the main steam line and discharges it to the suppression pool. The HPCI is similar, but allows much greater water flow rates.

The RPV pressure control systems provide manual and automatic control of the RPV internal pressure and consists of a set of Safety Relief Valves (SRV), safety valves, and the Automatic Depressurization System (ADS). The SRVs are DC-powered valves that control and limit the RPV pressure, and the ADS is a separate set of relief valves that are employed in order to depressurize the RPV.

The cooling water inventory includes the Condensate Storage Tank (CST), the PSP, and the firewater system. The CST in the considered plant is a 375 Kgal fresh water reservoir that can be used to cool the reactor. The PSP contains a large amount of fresh water that is relied upon as an ultimate heat sink when AC power is lost. Water from the firewater system can be injected into the RPV when other water injection systems are disabled and when the RPV is depressurized.

1.2. SBO Scenario

The analysis considered is a BWR Mk. I system during a Loss of Offsite Power (LOOP) event followed by loss of the diesel generators (DGs), i.e. station blackout (SBO). In more details, at time \( t = 0 \), LOOP condition occurs due to an external event (i.e. the offsite power lines are damaged). The LOOP alarm triggers the following events: a successful scram of the reactor; MSIVs are successfully closed, isolating the primary containment from the turbine building; emergency DGs successfully start keeping the AC
power busses energized; DC systems (i.e., batteries) are functional; and the decay heat generated by the core is removed from the pressure vessel through the residual heat removal system.

Next the SBO condition occurs due to internal failure which results in the failure of the DGs. As a result, removal of decay heat is impeded. Reactor operators start the SBO emergency procedures and perform: RPV level control using RCIC or HPCI, RPV pressure control using SRVs, and Containment monitoring (both drywell and PSP). At this point, plant staff start recovery operations to bring back on-line the DGs while the recovery of the off-site power grid is underway as well. Due to heavy usage, battery power can deplete. When this happens, all remaining control systems are offline causing the reactor core to heat until maximum temperature limit for the clad is reached: core damage (CD) condition occurs. If DC power is still available and one of three conditions is met, then the reactor operators activate the ADS in order to depressurize the RPV and allow firewater injection, if available. The aforementioned conditions are Failure of both RCIC and HPCI, HCTL limits reached, or the RPV water level becomes too low. When AC power is recovered, through successful restart/repair of DGs or off-site power, RHR can be employed to keep the reactor core cool.

2. BACKGROUND

2.1. Traditional High Dimensional Data Analysis Approaches

Dimensionality reduction (DR) and clustering analysis are widely applied techniques for analyzing high-dimension data. They provide very different approaches for understanding the structure of the high-dimensional space. In part of this study, we employed a visualization system that utilizing both dimension reduction and clustering where dimension reduction builds a visual mapping allowing the clustering result to be shown in 2D. Therefore allowing instant and intuitive visual analysis.

**Dimensionality Reduction.** Dimensionality reduction techniques \[2\], such as Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS), are common tools for analyzing high-dimension data. These techniques, which construct low-dimensional representations of data, are typically geometrically motivated, computationally efficient, and approximately preserve structural properties of the data. Under most situations, directly visualizing a high-dimensional space is not possible, but we still would hope to obtain a certain level of intuition regarding the structure in high-dimensional space. In order to generate a meaningful visual mapping for visualization purposes, we usually reduce the original high-dimension space into 2D or 3D space that can be direct visualized. Even though a number of DR methods are integrated into our system, we utilized PCA in our study for it’s simplicity and computational efficiency. The limitation regarding these techniques is two-fold. First, for linear dimension reduction, the reduced axis is a linear combination of the original dimensions, which is hard to interpret. Second, the structural information in high-dimensional space will suffer loss during the dimensionality reduction process, and it is even more difficult to detect the degree of information loss. By combining the clustering result with dimensionality reduction, we provide additional structural information in the embedded space, therefore mitigating some of the shortcomings of dimensionality reduction.

**Clustering Analysis.** Clustering analysis is another approach for understanding the high-dimensional structure of the data, it groups the dataset in such a way that points in the same cluster share more similarity to each other than to those in other clusters. Clustering usually depends on a heuristic rather than a “correctness guaranteed” algorithm and there are numerous criteria for defining what is a cluster therefore there are a large number of clustering methods. The various criteria most often used include density, distribution, centroidal distance, and connectivity etc. In our case, we adopted an agglomerative hierarchical clustering \[3\] (average linkage). Agglomerative hierarchical clustering is based on the connectivity of points. The core idea is that points should be more related to nearby points than points that are farther away. Starting from individual points as their own clusters, this method builds a dendrogram from the bottom up, merging and connecting clusters with nearby points or clusters. With hierarchical clustering we do not need to specify the number of clusters we are looking for, instead we can interactively expand or collapse different levels of clustering in the hierarchy.
2.2. Topological Based Analysis Approach

The Approximate Morse-Smale Complex. We consider an alternative method for analyzing high-dimensional data that models the simulation data as a scalar function $f$ defined on a finite set of points $X$ in $\mathbb{R}^n$ with scalar output, $Y$ in $\mathbb{R}$. In this way, we can then summarize the structure of the data with respect to the topology imposed by a variable of interest that is classified as the output of our constructed function. Various topological constructs and approximations exist that can be used to impose a structure on the arbitrary dimension, functional data, such as the contour tree [1], the reeb graph [12], and the Morse-Smale complex [5]. This paper makes use of the approximated Morse-Smale technique first established in [5] to present a hierarchical partitioning of the data based on estimated gradient flow behavior of the sampled data. In this way, we can still present the data in the same manner as the hierarchical clustering, but also take advantage of the scalar function property to offer alternative visualizations. A summary of this technique follows.

We partition the points in $X$ based on their function values and gradient behavior using an approximated hierarchical Morse-Smale complex. That is, at the finest level of detail, points belong to the same cluster if they share gradient flow to the same local maximum and local minimum. Gradient flow is estimated by imposing a neighborhood graph on the data and using the steepest ascending/descending neighbor of a point to represent the gradient at each point sample. In this way, each point can be traced to a local maximum or minimum.

Hierarchical Simplification. We can then merge clusters based on the persistence of their corresponding local extrema. We define persistence for an extremum as the minimum difference in function value between the extremum and its neighboring saddle points. We assume every saddle point is a simple saddle, thus every saddle will pair two local maxima or two local minima, and the merging of clusters is unambiguous as the adjusted gradient will be simulated to flow to the higher persistence extremum after two clusters are merged. An intuitive two-dimensional example is shown in Figure 1 where local maxima of various sizes are merged, in order, with a neighboring saddle point redirecting upward flow toward the maximum also sharing the same saddle point. The grey lines simulate this redirected flow and the circles represent the saddle locations. At each refinement, we simulate the new gradient flow with a simplified surface model in the figure.

![Figure 1](image.png)

**Figure 1:** A simulation of persistence simplification showing how the local maxima of this two-dimensional function are hierarchically merged with a neighboring saddle redirecting upward flow to a neighboring maximum that shares the saddle point.

Visualization Summary. We further the analysis by computing a summary curve via the following three step process: (1) perform inverse regression with each cluster’s data, (2) project the curves embedded in $\mathbb{R}^n$ to a two-dimensional viewing window using PCA [6], and (3) align the curves to meet at there shared extrema to maintain the coherency of the topological structure extracted. The resulting topological skeleton visualization encodes the projected average of each levelset of a cluster, a direction-independent estimate of the spread in domain space at each levelset presented as a transparent halo, and the sampling
Figure 2: An abstract view of our visualization technique on a 2D smooth surface: The surface is first segmented into areas of uniform gradient flow (a), then each levelset (modeled by a white line) is averaged to a single point and consecutive levelsets are connected to form a curve per segment (orange curves), and finally our result is shown with labels for each visualization cue.

density at each levelset of a cluster encoded as the darkness of the halo. Figure 2 details each of these visualization components given a 2D example surface.

As a supplementary view, we project the data colored by its segment as well as the high-dimensional curve summary onto a set of stacked two-dimensional scatter plots. The horizontal axis is the output dimension which is aligned on all plots, and the separate vertical axes of each plot are the individual input dimensions. The curve summary gives the average location at each levelset and each curve now uses a dimension-specific standard deviation for the width of the transparent halo. Furthermore, the color of the halo is held constant as the sampling density is better encoded by the overlaying of the data points themselves. In this way, we can begin to distinguish with respect to each individual dimension how the segments differ. Figure 3 shows a labeled version of this interface. The topological skeleton makes it more clear at a glance that there are four distinct segments, yet gives no underlying information about the sampled data in each cluster. This is why we believe that the combination of these techniques represents the best view of the data. More details of the visualization pipeline as well as additional views using the same input data format are described in prior works [4], [8], [9].

3. DATASET

An ensemble of 4997 transient simulations have been generated using classical Monte-Carlo sampling of seven input parameters. Out of 4997, 833 scenarios resulted in system failure (the core temperature reached the clad failure temperature set to 2200 F); the rest of the 4164 scenarios ended up in system success (AC power is recovered or the firewater becomes available if the RPV is depressurized). Each simulation includes information regarding the timing of various recovery attempts (e.g. cooling recovery, fire water, etc.) and component failures (e.g. battery life exhausted, a safety relief valve gets stuck open, etc.). The 7 input parameters* are the following:

- **FailureTimeDG**: Failure time of the DGs corresponding to the time of SBO event.
- **ACPowerRecoveryTime**: min{Recovery time of DGs, Off-site power recovery time}. The minimum of these two times will determine when the simulation is considered recovered.

*These input parameters are chosen to be analyzed as they are the only uncertain parameters under consideration in our current context.
Figure 3: Using a point sampling of the same function in Figure 2, we show the topological summary on the left and the inverse coordinate plot on the right where data is colored by the segment to which it belongs rather than a scalar colormap of the function values. Note the horizontal axes in the plots on the left are the output variable, whereas each plot uses a different input, x and y, for the vertical axes.

- **SRVStuckOpenTime**: The time when an SRV is stuck in the open position.
- **CoolingFailtoRunTime**: $\max\{\text{HPCI failure time, RCIC failure time}\}$. As long as one of these systems is functioning, the reactor is being actively cooled, so it is important to understand when both are lost. Thus, we take the maximum of these two times.
- **ADSactivationTimeDelay**: The time when the operator manually depressurizes the RPV by activating the ADS system. This parameter actually measures the time delay from the HCTL event, not the time from 0 to when ADS is activated.
- **firewaterTime**: As an emergency action, when RPV pressure is below 150 psi, plant staff can connect the firewater system to the RPV in order to cool the core and maintain an adequate water level.
- **extendedECCSOperation**: Battery life combined with Extended ECCS operation. That is, operators may extend RCIC/HPCI and SRV control even after the batteries have been depleted. They manually control RCIC/HPCI by acting on the steam inlet valve of the turbine and/or supply DC power to the SRVs through spare batteries.

All the above time-related parameters are measured from the time of the SBO event (in seconds), which is the FailureTimeDG, with the exception of the ADSactivationTimeDelay which is measured from the time of the HCTL event. The output variables obtained from the simulations are the **maxCladTemp**, which is the maximum clad temperature reached during the entire course of the simulation, and the **EndSimulationTime** which for failure cases represents the time to reach the failure temperature of 2200 F. We study the topology of the data with respect to each of these outputs in isolation.

The above data is pre-processed with a standardization process. We employ the Z-score scaling which is a data standardization process whereby values $V$ of each dimension are recomputed as $V - \text{mean}(V)/\text{std}(V)$; therefore all input parameters have the same mean (0) and standard deviation (1) but may vary in ranges. For the Morse-Smale segmentation, the output variable is not scaled, but all input dimensions are scaled using this method. Although more robust approaches are available [7], the above technique is often sufficient.
4. RESULTS

In this section, we will provide analysis for two cases using the seven dimensional input data. In the first case, sections 4.1.1 and 4.2.1, we consider all 4997 simulations and use the maximum clad temperature as the output variable. In this analysis, all failure cases will have the same output value of 2200. Our second case, detailed in sections 4.1.2 and 4.2.2, looks at only the 833 failure scenarios and since the maximum clad temperature does not vary for these cases, we instead look at the end simulation time as the output variable. In order to give the reader a comprehensive picture, and provide comparison for our proposed topology-based analysis. We first demonstrate traditional dimensionality reduction and clustering analysis, followed by results from the topological analysis. Afterwards, we will summarize the benefits and drawbacks inherent in each approach.

4.1. Dimension Reduction & Clustering Combined Analysis

As shown in Figure 5, our analysis framework allows the display of clusters in the embedded space as well as displaying an individual cluster and its statistical summary (shown as a “box” next to the cluster). In each cluster statistical summary, one bar corresponds to each of the eight variables (seven inputs plus one output). For each bar, we use a yellow span to denote the min-max range and a red marker to indicate the mean value for a particular cluster. With this summary, we can quickly compare and investigate the defining characteristics of each cluster at a glance.

We start our analysis of the data by applying PCA, reducing the original eight dimensions to two dimensions. Once we obtained the reduced dimension coordinates, we plot the points in visual space for direct visual analysis. In order to study each dimension’s distribution to the reduced dimensional structure in two dimensions, we color the points according to each dimension, as shown in Figure 4 and Figure 6. This give us a direct view for each dimension, however, the structure in high-dimension is defined by all the dimensions. In the next two sections, we will look at each of the scenarios.

![Figure 4](image.png)

**Figure 4:** PCA dimension reduction results for the 7+1D full dataset (colored by the labeled dimension), only dimensions that show relatively strong correlation patterns are displayed. The corresponding colormap is shown in the bottom.

4.1.1. 7D Maximum Clad Temperature - All Cases

Figure 4 shows the PCA projection of all 4997 scenarios. Each of the plots shows the same data points colored by a different dimension. ADSActivationTimeDelay and FailureTimeDG are omitted as they do not exhibit any visual correlation with the projection. The colormap of the values is shown underneath where the red values represent low parameter value and the blue values represent high values.
In Figure 5(b), there appears to be few data points with a moderate maxCladTemp as the upper portion of the data is dominated by success cases characterized by low MaxCladTemp values, and the bottom half of the data consists of mostly failure cases. As with MaxCladTemp, ACPowerRecoveryTime is also correlated with the vertical dimension, although we see the range of values appear in both halves, so the claim cannot be made that a low AC power recovery time will yield a low maximum clad temperature. Furthermore, there is a cluster of relatively high ExtendedECCSOperation times that appears near the top of the projection where success cases dominate, so one could conclude that a long extended ECCS operation time is a main contributing factor for stable system recovery, although again not a sufficient condition, as there are some higher-valued ExtendedECCSOperation values appearing in the lower half of the data. The remaining three parameters are orthogonal to the output dimension in this projection and one conclusion to draw from this projection is that they will have a lesser impact on the outcome of the simulation. It is important to note that a vertical or horizontal pattern corresponds to the variance of the dimension. That is, a larger variance corresponds to a more noticeable pattern, this is most likely due to the fact PCA is optimized for capturing maximum variance.

In Figure 5 on the left, we have colored the 2D reduced dimension point cloud according to a hierarchical clustering. We select a level in the hierarchy composed of seven clusters in total, two very small clusters contain possibly outliers in the dataset. Each individual cluster along with its statistics are shown on the right. Starting from the root node of the clustering hierarchy, an interactive exploration shows the whole dataset is split into two parts. These parts are the upper successful case cluster noted earlier (the upper four clusters belong to this cluster in the hierarchy) and the lower failure dominated cluster (the lower three clusters belong to this cluster in the hierarchy). We have verified that the upper four clusters contain exclusively success cases, while the lower clusters contain all of the failure cases and a select number of successes.

Among the bottom three clusters, containing mostly failure case, we can see from the cluster statistics that the green and pink clusters differ mostly in ExtendedECCSOperation and CoolingFailToRunTime. The green cluster is concentrated with data points exhibiting lower ExtendedECCSOperation and higher CoolingFailToRunTime compared to the pink cluster. Focusing on the success cases, the blue and cyan clusters share a lot of similarity in their cluster summary statistics. Here ACPowerRecoveryTime seems to be the main contributing factor for distinguishing these two clusters.
Figure 6: PCA dimension reduction results for the 7+1D failure case dataset (colored by the labeled dimension), only dimensions that show relatively strong correlation patterns are displayed. The corresponding colormap is again shown at the bottom.

Figure 7: On the left, we show the clustering result for the failure cases embedded in the projection view. On the right, we provide a separate illustration of each individual clusters and their summary statistics (box next to each cluster containing cluster min, max, mean for each dimension).

4.1.2. 7D End Simulation Time - Failure Cases

Figure [6] shows the PCA projection of the failure cases colored by five different dimensions using the same colormap as before. Again, the SRVstuckOpenTime and firewaterTime are correlated with the horizontal axis whereas the extendedECCSOperation is correlated vertically. We can note that very few points exist with a high EndSimulationTime, and so it is more difficult to draw conclusions with respect to the output variable from this view.

In Figure [7], 5 clusters are presented, one of which is a small cluster containing a possible outlier. In this focused analysis on failure cases, without the interference from the dominating dimension MaxCladTemp, we observe different separations of clusters. Looking at the green and red clusters, they both consists of points with low simulationEndTime. The differentiating factors here are the CoolingFailureToRunTime and ExtendedECCSOperation.
4.2. Topological Analysis

4.2.1. 7D Maximum Clad Temperature - All Cases

Figures 8 and 9 summarize the results of this analysis which focuses on a level consisting of four segments, three of which share the global maximum, and one that consists of cases exhibiting low MaxCladTemp values denoting that only success cases are incorporated in it. Here we will focus on the conditions that lead to distinct local minima, or the different parameters that yield a stable success scenario. For this, we will focus on the behavior of the summary curves occurring at the left side of the rightmost image of Figure 8.

The pink cluster exhibits a fast ACPowerRecoveryTime, a slow firewaterTime, and a fast ExtendedECC-SOperation time, whereas the minimum associated with the blue cluster has a slow ACPowerRecoveryTime, a very fast firewaterTime, a faster ADSActivationTimeDelay, and a slow ExtendedECCSoperation time. The third local minima, shared by the green and cyan clusters, has a moderate firewaterTime paired with a fast ACPowerRecoveryTime and a longer ExtendedECCSoperation time. The parameters that seem to matter least are the FailureTimeDG, CoolingFailToRunTime, and the SRVStuckOpenTime. This last observation seems well-aligned with the observations of section 4.1.1 where we saw no correlation between the maxCladTemp and the FailureTimeDG, and that the CoolingFailToRunTime and SRVStuckOpenTime were orthogonal to the maxCladTemp in the PCA projections. The new information provided here is that the firewaterTime setting does play a role in differentiating flow to distinct local minima.

Thus, from a safety point of view, we can observe that in order to assure a low value of maximum clad temperature the high pressure injection system needs to be available for a long time for all three clusters. Interestingly, note that the failure time of the DGs (i.e., initial time of the SBO condition) does not play a relevant role into guaranteeing a low value of max clad temperature. Note also that (by looking at the pink cluster in Figure 8) an early AC recovery time guarantees system success even for early values of SRVstuckOpenTime. ExtendedECCSoperation time and late firewaterTime. Hence, even in case of an early RPV depressurization (i.e., SRV stuck open), the core heating rate is slow enough that an early AC recovery time guarantees low values of max core temperature.

For comparison, we can use the same projection view used in Section 4.1 to view the topological clustering result, as shown in Figure 9. This view gives less information about the evolution of the data with respect to the output variable, yet provides a more compact analysis with respect to each input dimension with respect to the identified clusters.

4.2.2. 7D End Simulation Time - Failure Cases

In this dataset, we consider only failure cases and use the end simulation time (EST), how quickly the system failure occurred, as the output value. Results are shown in Figure 10 and Figure 11. From the left image in Figure 10, we note that there is a global fast failure case marked by a time of 434.82 seconds. All failure cases can be traced down to this single, global minimum case, but from it arise four distinct local maxima. One interpretation is to look at the local maxima as possible near successes since they represent the locally latest times for failure. In other words, the temperature growth of these scenarios was slower and thus allowed more time to be contained.

From a safety point of view, we are interested in understanding under which conditions we have a late core damage event. By looking at the green cluster of Figure 10 as expected, a driving factor to reach a late core damage is a high value of ECCS operation. This implies that is preferable to keep the RPV
Figure 8: The topological skeleton of all 4997 cases are shown on the left. The right two images show the inverse coordinate plots of the same data where points and regression curves are colored by cluster (crystal) memberships.

Figure 9: The left image depicts the topological clustering result in the projection view of all 4997 simulations, and the right images show each of the individual clusters and a statistical summary with respect to each dimension.

Pressurized as long as possible and maintain high pressure cooling instead of activating the ADS system and obtaining cooling through the FW system. Also note that this occurs for late values of AC recovery time.

By looking at the purple cluster of Figure 10, we also notice that a late core damage is reached for high value of FailureTimeDG since a large quantity of heat has been discharged before reaching SBO condition. On the other hand (see red cluster of Figure 10), a late core damage occurs when a small quantity of heat has been rejected from the core following reactor scram (low value of FailureTimeDG) and late failure of the high pressure core cooling system. In summary, note that for all clusters, a late failure of the high pressure core cooling system and a late AC-Power-Recovery-Time are always needed in order to guarantee a late core damage condition. However, FailureTimeDG plays a role when coupled with the firewaterTime.
As before, we project the data colored by the topological segmentation into a two dimension view using PCA and display a summary of each cluster in Figure II. Here again, we are able to see how the clusters differ, but what is hidden is at what output level they vary. For example, looking at ACPowerRecoveryTime in Figure II, we see that each cluster varies in the range and mean, however the plots in Figure fig:7D-EST-Failure-4C reveal that these different settings are not apparent at the local maxima of these curves, and in fact this dimension shows the summary curves overlapping significantly over their spans.

Figure 11: 7D-EST-Failure-4C: One the left is the topological clustering result in the projection view, and the right is a illustration each of the individual clusters and their statistic in the box next to each cluster

5. CONCLUSION

In this paper, we investigated the use of both classic and topology-based clustering techniques in conjunction with dimensionality reduction on data generated by DPRA methodologies. The scope is to provide the user a set of tools able to analyze large complex datasets and obtain insights about system response under the simulated accident scenario. We presented two methodologies based on classical hierarchical clustering and topology-based algorithms. We applied both methodologies to a dataset generated within
the RISMC project. Such dataset simulates the response of a BWR system during a SBO accident scenario. We performed a series of simulations where, for each simulation run, we randomly changed timing and sequencing of events. The scope of the analysis is to identify how timing/sequencing of events affects the maximum core temperature.

We saw that a standard hierarchical clustering was adept at distinguishing failure cases from success cases and grouping points that had similar parameter settings, however the topological segmentation presented data in a different manner allowing the users to see how one proceeds, in parameter space, from a success scenario to a failure scenario or vice versa. In this way, the topology-based method often had data points that spanned the range of values for each dimension, but we saw how the range of values in each input dimension shifted at different levelsets captured by the summary curves overlaid on the inverse coordinate plots.

References