

Analyzing Dynamic Probabilistic Risk Assessment Data through Topology-Based Clustering

Dan Maljovec, Bei Wang, Valerio Pascucci

Scientific Computing and Imaging Institute, University of Utah
{maljovec,beiwang,pascucci}@sci.utah.edu

Peer-Timo Bremer

Lawrence Livermore National Laboratory
bremer5@llnl.gov

Diego Mandelli

Idaho National Laboratory
diego.mandelli@inl.gov

ABSTRACT

We investigate the use of a topology-based clustering technique on the data generated by dynamic event tree methodologies. The clustering technique we utilize focuses on a domain-partitioning algorithm based on topological structures known as the Morse-Smale complex, which partitions the data points into clusters based on their uniform gradient flow behavior. We perform both end state analysis and transient analysis to classify the set of nuclear scenarios. We demonstrate our methodology on a dataset generated for a sodium-cooled fast reactor during an aircraft crash scenario. The simulation tracks the temperature of the reactor as well as the time for a recovery team to fix the passive cooling system. Combined with clustering results obtained previously through mean shift methodology, we present the user with complementary views of the data that help illuminate key features that may be otherwise hidden using a single methodology. By clustering the data, the number of relevant test cases to be selected for further analysis can be drastically reduced by selecting a representative from each cluster. Identifying the similarities of simulations within a cluster can also aid in the drawing of important conclusions with respect to safety analysis.

Key Words: high-dimensional data analysis, computational topology, nuclear reactor safety analysis, visualization

1 Introduction

Dynamic Probabilistic Risk Assessment (DPRA) [1] methodologies couple system simulator codes (e.g., RELAP [2], MELCOR [3], MAACS [4]) with simulation controller codes (e.g., ADAPT [5], ADS [6], MCDET [7], RAVEN [8]). 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, the major challenges in using DPRA methodologies, especially those that employ dynamic event trees (DETs), are the heavier computational and memory requirements needed to search the uncertain parameter space systematically and to generate a large number of scenarios due to the time evolution of a large number of variables.

The analysis of such data is normally achieved by considering the end state of each simulation run (e.g., fuel melting temperature reached v.s. intact core) and observing the sequence and the timing of events that lead to that end state. When system complexity is very high (e.g., analysis of nuclear power plant accident scenarios), analysis of such large quantities of data may require advanced tools that evaluate impact of uncertainties and timing/sequence of events on system dynamics.

A first approach toward discovering these correlations from data generated by DPRA methodologies has been developed using Fuzzy classification [9] and clustering algorithms [10]. In particular, clustering algorithms partition the set of scenarios into clusters by identifying similarities based on certain criteria, allowing users to organize and interpret the trends in scenario evolution and risk contributors for each initial event [11]. The set of scenarios could be analyzed in two modes, either by *end state analysis* that classifies the scenarios into clusters based on their end state (e.g. final outcome) (e.g. [12]), or by *transient analysis* that considers the complete system dynamics (e.g. time evolution of scenarios) and identifies clusters having similar temporal behavior of the state variables (e.g. [10]). The work that is of most relevance to us utilizes a mode-seeking clustering method, the mean-shift algorithm [10]. The mean shift algorithm [13] is a non-parametric iterative procedure that assigns each data point to the average of data points in its neighborhood as the cluster center. The data points are therefore clustered based on observation density.

Contributions. In this paper, we present a software tool that provides the domain experts with an interactive analysis and environment for understanding the structures of high-dimensional nuclear simulation datasets. In particular, such a tool enables the end users to perform both *end state analysis* that apply to the end state of the scenarios, and *transient analysis* that apply to their time evolution data, to classify the set of nuclear scenarios. We focus on clustering algorithms based on topological structures such as the Morse-Smale complex, which partitions the data points into clusters based on their uniform gradient flow behavior. We compare our techniques with mean shift methodology on a dataset generated by a DET for a sodium-cooled fast reactor during an aircraft crashing scenario. We demonstrate that these techniques offer complementary views of the data that help illuminate key features that may be otherwise hidden using a single methodology.

2 Background

Our topology-based clustering algorithm relies heavily on the concepts of Morse-Smale complex, its approximation in high dimension and persistence simplification. We provide minimal technical background on these concepts with intuitive examples, to convey a basic understanding of these concepts for non-specialists. It is important to note, that topological clustering assumes we can treat our data as a scalar function, where an arbitrary number of inputs in the the domain space, map to scalar output values in the range space. The clustering we provide, defined in terms of gradient behavior with respect to the output value, maintains a coherent, non-overlapping segmentation of the domain space.

Morse-Smale Complex and Its Approximation. Our clustering methodology is based upon the topological structure know as the Morse-Smale complex [14–16], which is derived from the Morse theory [17]. Let $f : \mathbb{M} \rightarrow \mathbb{R}$ be a smooth function defined on a smooth manifold embedded in \mathbb{R}^n . A point $x \in \mathbb{M}$ is called *critical* if its gradient (e.g. a vector that points in the direction of the greatest increase of the function) $\nabla f(x) = 0$, otherwise it is *regular*. At any regular point x the gradient is well-defined and integrating it traces out an integral line. The function increases along the integral line, which begins at a local minimum and ends at a local maximum of f . The *ascending/descending* manifold of a critical point p is defined as all points whose integral lines start/end at p . The descending manifolds form the *Morse complex* of f while the ascending manifolds form the *Morse complex* of $-f$. The set of intersections of ascending and descending

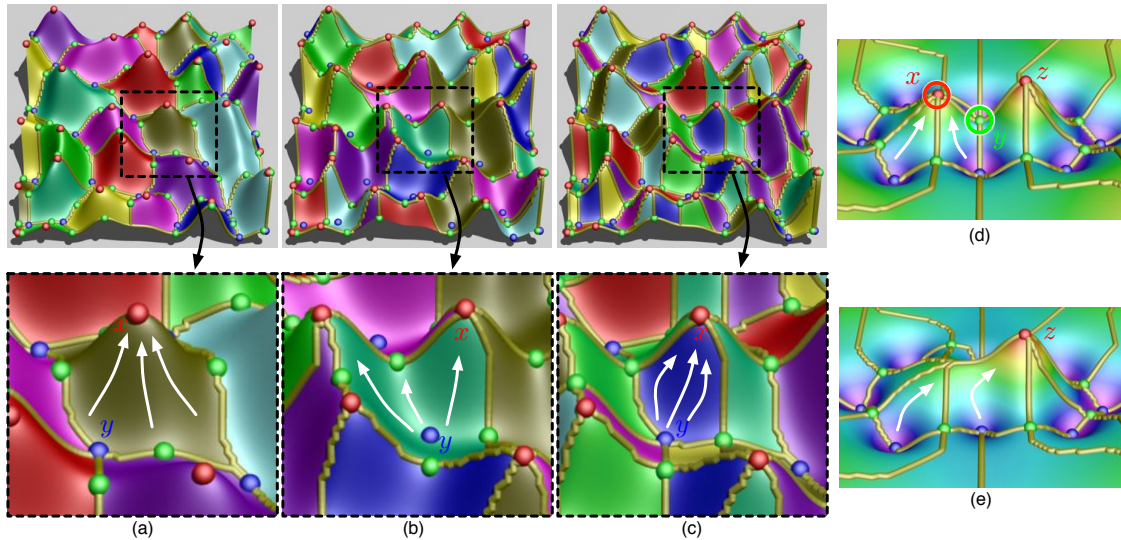


Figure 1. For a height function defined on a 2D domain (where maxima, minima and saddles are colored red, blue and green respectively): (a) For each descending manifold, the gradient flow (white arrow) ends at the same maxima; (b) For each ascending manifold, the gradient flow starts at the same minimum; (c) For each Morse-Smale crystal, the gradient flow begins and ends at the same maximum-minimum pair. To illustrate persistence simplification: In (d), the left peak at the maximum x is considered less important topologically than its nearby peak at maximum z , since x is lower. Therefore, at a certain scale, we would like to represent this feature as a single peak instead of two separate peaks, as shown in (e), by redirecting gradient flow (white arrow) that originally terminates at x to terminate at z . In this way, we simplify the function by removing (canceling) the local maximum x with its nearby saddle y .

manifolds creates the *Morse-Smale* complex of f . Each *crystal* of the Morse-Smale complex is a union of integral lines that all share the same origin and the same destination. In other words, all the points inside a single crystal have uniform gradient flow behavior. These crystals yield a decomposition into monotonic, non-overlapping regions of the domain. Figure 1(a)-(c) illustrate these concepts.

To approximate the Morse-Smale complex of a high-dimensional scalar function f defined on a finite set of points \mathbb{X} in \mathbb{R}^n , we would need to estimate the gradient at each input point. This is done by imposing a neighborhood graph that connects points in \mathbb{X} with edges that approximates its underlying structure. At each point in \mathbb{X} , we choose the steepest ascending edge to represent the gradient. With this gradient approximation, we can determine the extrema by labeling all points with no neighbors of higher values as maximum and all points with no neighbors of lower values as minimum. We then label all points in \mathbb{X} according to the local extrema at which its gradient begins and terminates. Subsequently, we collect all vertices with the same pair of labels into crystals and add the extrema to all crystals that share the corresponding label. These crystals form an approximation of the Morse-Smale complex [18]. The key component in our proposed work is that such Morse-Smale crystals form the clustering of \mathbb{X} . In fact, such type of clustering based on topological structure is not entirely new. As pointed out in [18], the Morse complex can be thought of as a variant of mean shift clustering in which the kernel density estimation is replaced with function values of f .

While many neighborhood graphs are possible [19,20], we use a *relaxed Gabriel graph* [21] here which has certain desirable properties (such as relatively dense connectivities). The Gabriel graph [22] is a specific type of β -skeleton graph ($\beta = 1$) all of which belong to a larger class of graphs known as empty region graphs. As such, these class of graphs are comprised of all edges that pass an empty region test. For the Gabriel graph, this entails circumscribing a proposed edge between points p and q with a d -dimensional ball

with a diameter equal to the length of the edge \overline{pq} . As long as no other points lie within the circumscribing ball, the edge is added to the graph. We can relax this property by further prescribing that only neighbors of p and q are considered in the empty region test. Thus, if r is not a neighbor of p or q , but lies within the circumscribing region, then the edge \overline{pq} may be added. In order for this process to work correctly, we begin at a given point and add edges in increasing order of distance.

Persistence Simplification. In real datasets, there is often noise which may manifest itself as small topological artifacts, either spurious extrema that may not truly exist in the data or small features the user does not deem relevant. To account for this and allow the user to select a *scale* appropriate for the specified dataset, we use the notion of *persistence simplification* [23,24], whereby less salient features are merged with neighboring, more significant features. We illustrate the persistence simplification procedure intuitively through the example in Figure 1(d)-(e).

3 Analysis Methodology and Results

Analysis methodology overview. In this paper, we apply clustering algorithms based on Morse-Smale complex to nuclear datasets generated by DETs. We model such a dataset as a high-dimensional scalar function f defined on a finite set of points \mathbb{X} in \mathbb{R}^n . We partition the points in \mathbb{X} based on their function values and gradient behavior with respect to the approximated Morse-Smale complex. That is, points belong to the same cluster if they have uniform gradient flow behavior.

Demo example. To illustrate the proposed clustering methodology, we use a dataset generated for the analysis of recovery from an aircraft crash into an RVACS of a conceptual design for a sodium-cooled fast reactor [10,25–27]. 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 four cooling towers. The Analysis of Dynamic Accident Progression Trees (ADAPT) tool [5] is used as the DET generator while the system dynamics is modeled using RELAP5 [2]. A typical scenario is the following: the plant is operating at 100% power when an airplane crashes into the plant, destroying three of the four towers and, thus, the reactor core cooling capabilities are disabled. A recovery crew then arrives at the site and attempts to reestablish the cooling of the reactor by restoring the damaged towers one by one. An ensemble of 609 transient simulations has been generated*, and among which 132 scenarios are considered system failures when the reactor reaches a maximum temperature of 1000K before the end of simulation, which is fixed to be 2×10^5 seconds†. The rest of the 477 scenarios are considered simulation completions. Each simulation includes information regarding: (a) Time profile of core temperature; (b) Crew arrival time; and (c) Recovery time of tower 1, 2 and 3.

Data representation for end state analysis. For the end state analysis, the above data is represented as various 4D scalar functions. We employ two types of representations: 4D scalar function based on absolute timing and relative timing, respectively.

For the 4D absolute timing analysis, the four input parameters represent the time for the crew to arrive at the plant (t_0), and the time for them to recover the first (t_1), second (t_2) and third (t_3) tower, respectively. In certain cases, the third tower has not been recovered before the end of the simulation, so $t_3 = 0$. The output variables considered are: (1) MT: maximum temperature reached in the simulation; (2) TTF: time to reach

*One transient simulation out of 610 simulations is considered as an outlier due to discrepancies among crew arrival times.

†Scenarios that reach 1000K before the end of simulation have been extended in time with the last value simulated.

Label	Description	Label	Description	Label	Description
A.1	4D-MT-all-2C	B.1	4D-Delta-MT-all-5C	C.1	100D-MT-all-4C
A.2	4D-MT-completions-4C	B.2	4D-Delta-MT-completions-5C	C.2	100D-MT-completions-3C
A.3	4D-TTF-all-2C	B.3	4D-Delta-TTF-all-4C	C.3	100D-TTF-all-3C
A.4	4D-TTF-failures-3C	B.4	4D-Delta-TTF-failures-4C	C.4	100D-TTF-failures-3C
A.5	4D-TTM-all-3C	B.5	4D-Delta-TTM-all-4C	C.5	100D-TTM-all-3C
A.6	4D-TTM-completions-3C	B.6	4D-Delta-TTM-completions-4C	C.6	100D-TTM-completions-3C

Table I. Various 4D and 100D functions under analysis. A: End state absolute timing analysis. B: End state relative timing analysis. C: Transient analysis.

failure temperature[‡]; (3) TTM: time to reach maximum temperature. In other words, in the Morse-Smale complex formulation, the finite set of points \mathbb{X} is embedded in \mathbb{R}^4 with coordinates coming from the 4 input parameters, and the scalar function f defined on \mathbb{X} obtains its values as one of the three figures of merit (e.g. MT, TTF or TTM).

For the 4D relative timing analysis, the four input parameters represent the relative time between recovering procedures, that is, the time for the crew to arrive at the plant from the time when the accident occurs ($t'_0 = t_0 - 0 = t_0$), the time between recovering the first, second and third tower respectively, that is, $t'_1 = t_1 - t_0$, $t'_2 = t_2 - t_1$, $t'_3 = t_3 - t_2$. The output variable of interest is designed to be MT, TTF or TTM.

Data representation for transient analysis. For the transient analysis, since only a single state variable (e.g. the core temperature) is under consideration, we model the temporal evolution of each scenario as a d -dimensional point, where d is the number of uniform temperature samples obtained along its time profile. We set $d = 100$. Therefore, the above data is modeled as various 100D scalar functions. Again, the output variable of interest is designed to be MT, TTF or TTM.

Cases under analysis. A further analysis of these various 4D and 100D functions based on all scenarios, system failure scenarios and simulation completion scenarios leads to the detailed clustering results for the following cases summarized in Table I. The shorthand description of each case encodes the dimension, output variable of interest, scenarios included and number of crystals obtained. For example, case (A.1) *4D-MT-all-2C* means a 4D function under the end state absolute timing analysis, where the output variable is MT, and the obtained result includes 2 crystals. Similarly, case (B.4) *4D-Delta-TTF-failures-4C* means a 4D function under the end state relative timing analysis for the system failure cases, where the output variable is TTF and the result includes 3 crystals; case (C.6) *100D-TTM-completions-3C* means a 100D function for the simulation completion cases under the 4 crystal setting, where the output variable of interest is TTM.

Objective. We would like to understand how these input variables impact system dynamics [26, 27]. Understanding the structure of such 4D and 100D functions may help domain scientists to make decisions regarding repair strategies and evacuation plans.

3.1 Visual Interface

We review and illustrate our visual interface in Figure 2 through the analysis of a 4D function under case (A.1) described above. Such visual interface is inherited and extended from the capabilities provided by HD-Viz [18] and has been previously employed for nuclear reactor analysis and visualization [28]. As shown

[‡]For scenarios where failure temperature is never reached, this is set to 2×10^5 .

in Figure 2 (a), the topological summary interface summarizes each Morse-Smale crystal into a 1D curve in high-dimensional space which is then projected onto a viewable 3D space. The interface encodes three steps (detailed in [18]), to arrive at a 3D representation for analysis and visualization of the d -dimensional scalar function f , defined on a set of sampled points \mathbb{X} . (1) *Morse-Smale approximation*: We approximate the Morse-Smale crystals in high dimension using a relaxed Gabriel graph. (2) *Geometric summaries*: For each crystal of the Morse-Smale complex, since each point has similar approximated monotonic gradient behavior, a geometric summary is constructed by an inverse regression, yielding a 1D curve in the d -dimensional domain of f . Intuitively, the value of the parametrized curve at a given location yields a representation of the crystal as the average of the function values of level sets within the crystal. (3) *Dimension reduction*: The set of regression curves can be represented by a graph embedded in \mathbb{R}^d with each edge corresponding to a curve and vertices corresponding to extremal points. Such a graph is then embedded into 2D preserving the spatial relation among the extrema and the geometry of the crystals that connect them using PCA or ISOMAP [29], while the third dimension is reserved for the output parameter. A simple example of the above three step process is shown for a simple 2D height function in Figure 3.

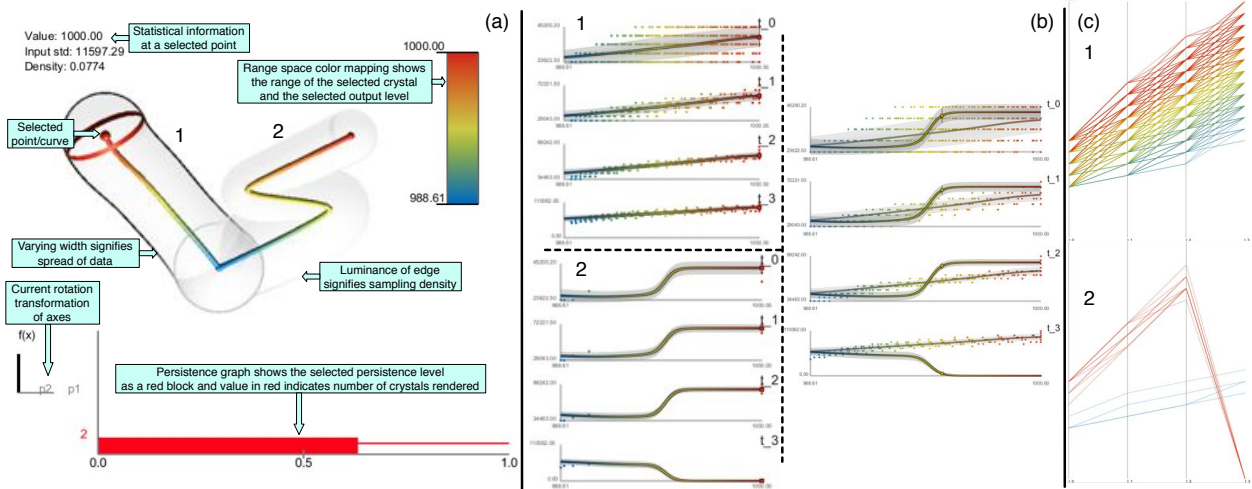


Figure 2. Visual interface illustrated through analysis of 4D function described in (A.1). (a) The topological summary visual interface. (b) Left: Inverse coordinate plots for crystals 1 and 2 in (a); (b) Right: combined inverse coordinate plots for crystals sharing the same minima; (c) Parallel coordinate plots for crystals 1 and 2 in (a).

Various visual components are described in Figure 2 (a), three of which are of most relevance to us in the current context: the crystal projection, the persistence graph and the range space color mapping. To enable multi-scale analysis, we use a modified version of the persistence diagram [23], referred to as the *persistence graph*. It shows the number of Morse-Smale crystals (y -axis) as a function of scale (i.e. x -axis, persistence threshold normalized by the range of the dataset). A selected scale is drawn with a red box and a corresponding number of Morse-Smale crystals is displayed along the y -axis in red. Stable features are considered as those that exist over a large range of scales (i.e. a sequence of persistence simplification with increasing scales), which correspond to long horizontal lines in the persistence graph.

In the inverse coordinate plots, each input parameter (e.g. t_0) is considered as a 1D function of the output variable (e.g. MT). This is shown in Figure 2(b), where the 1D regression curve as well as the data points are colored according to range space values. Furthermore, we could create combined inverse coordinate plots by superimposing those from crystals that belong to the same ascending/descending manifold. This

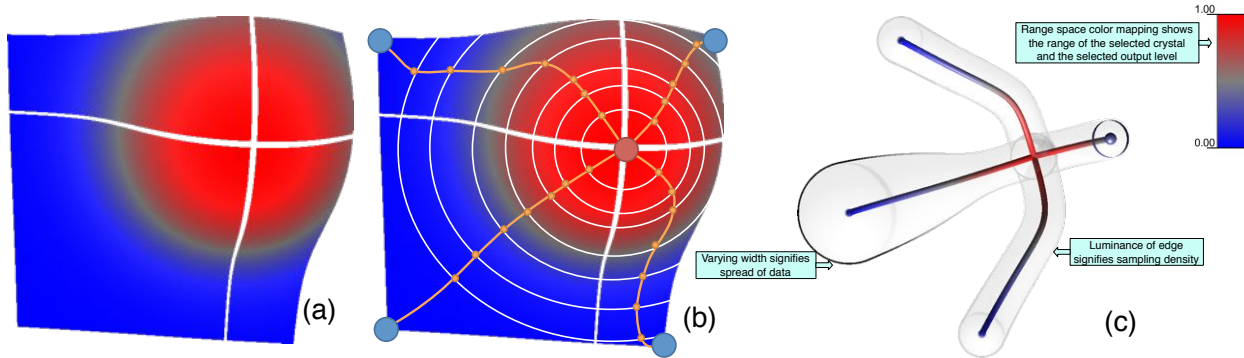


Figure 3. Key steps in generating the topological summary of a simple 2D height function. (a) Morse-Smale crystal approximations that decompose the domain into 4 crystals. (b) Each crystal obtains a 1D geometric summary which encodes the average of function values at level sets within the crystal. (c) The collection of 1D curves are then projected to a viewable 3D space preserving the spatial relations among the extrema and the geometry of the crystals. Several important visual components are described here as well.

enables further differentiations among various dimensions across crystals. We further use parallel coordinate plots [30] to illustrate the correlations among input parameters in the high dimensional datasets for each crystal, as shown in Figure 2 (c), where the curves are colored based on range space values as well.

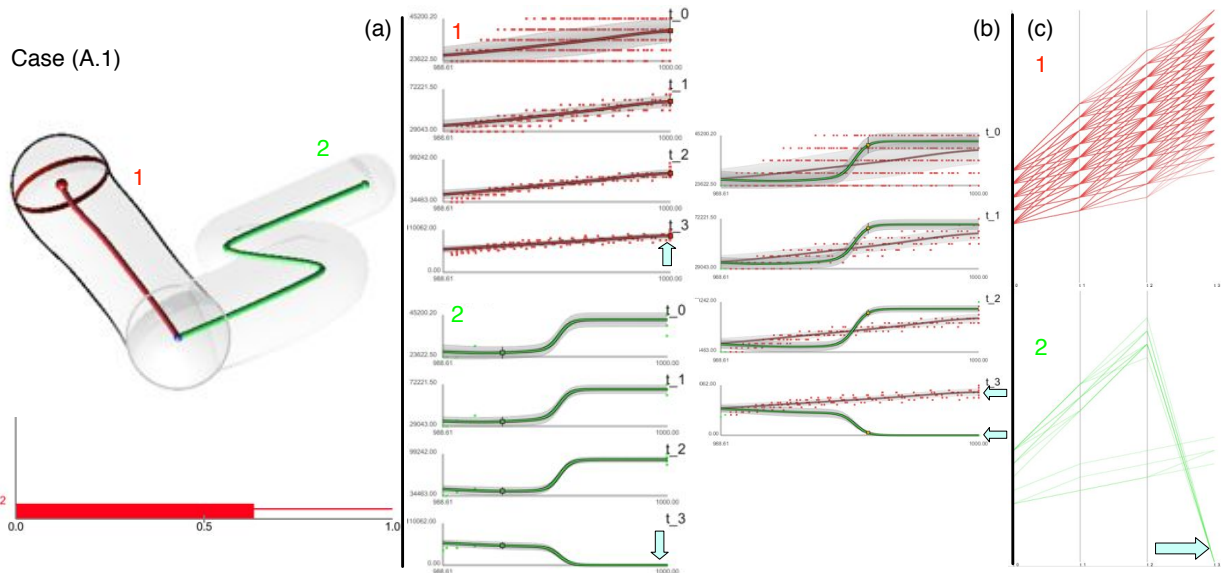


Figure 4. Visual interface highlighting clustering structure for 4D function described in (A.1). (a) The topological summary visual interface. (b) Inverse coordinate plots for both crystals individually and combined. (c) Parallel coordinate plots.

To highlight clustering structures obtained from the topological segmentation, we change our color map of the above visual interface where the 1D regression curve as well as data points are colored based on clustering membership, as shown in Figure 4.

3.2 End State Analysis

For this section, we showcase some insights obtained from various analysis cases. These insights and their implied conclusions are only a fraction of possible outcomes from our analysis and visualization framework, and a lot of these observations demand further investigations. Nevertheless, they give us some initial understanding of the nuclear dataset analyzed, and offer complementary views compared to results obtained with mean-shift methodology [10]. For each of the following cases, the number of crystals (i.e. number of clusters) are chosen based on the persistence graph. In the situations where several possible clusterings with different cluster numbers give comparable results, we choose one of the clusterings arbitrarily. We use the word cluster and crystal to mean the same thing. Important observations/insights are obtained from the part of the visual interface highlighted by arrows in cyan.

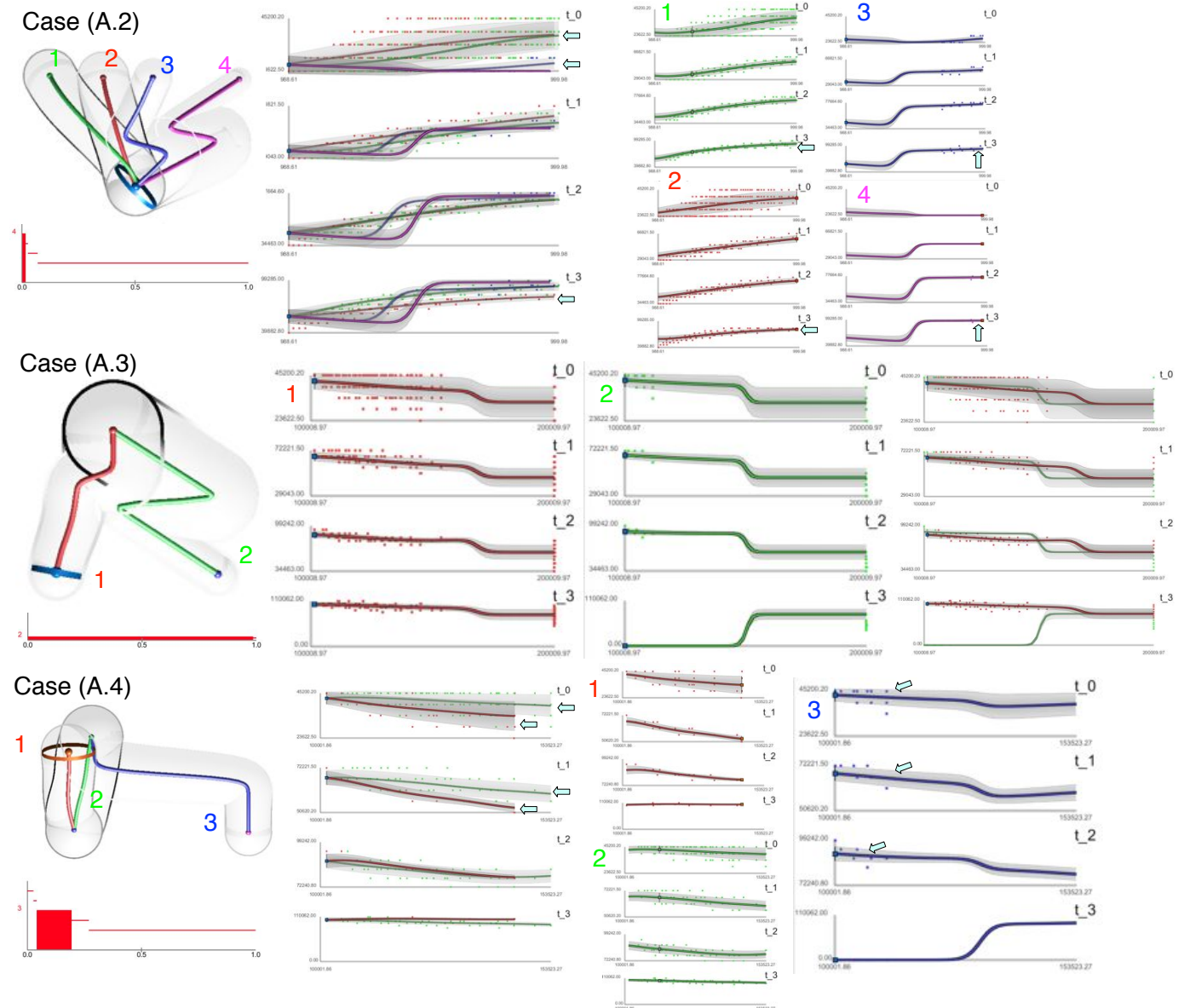


Figure 5. End state analysis, cases (A.2) through (A.4).

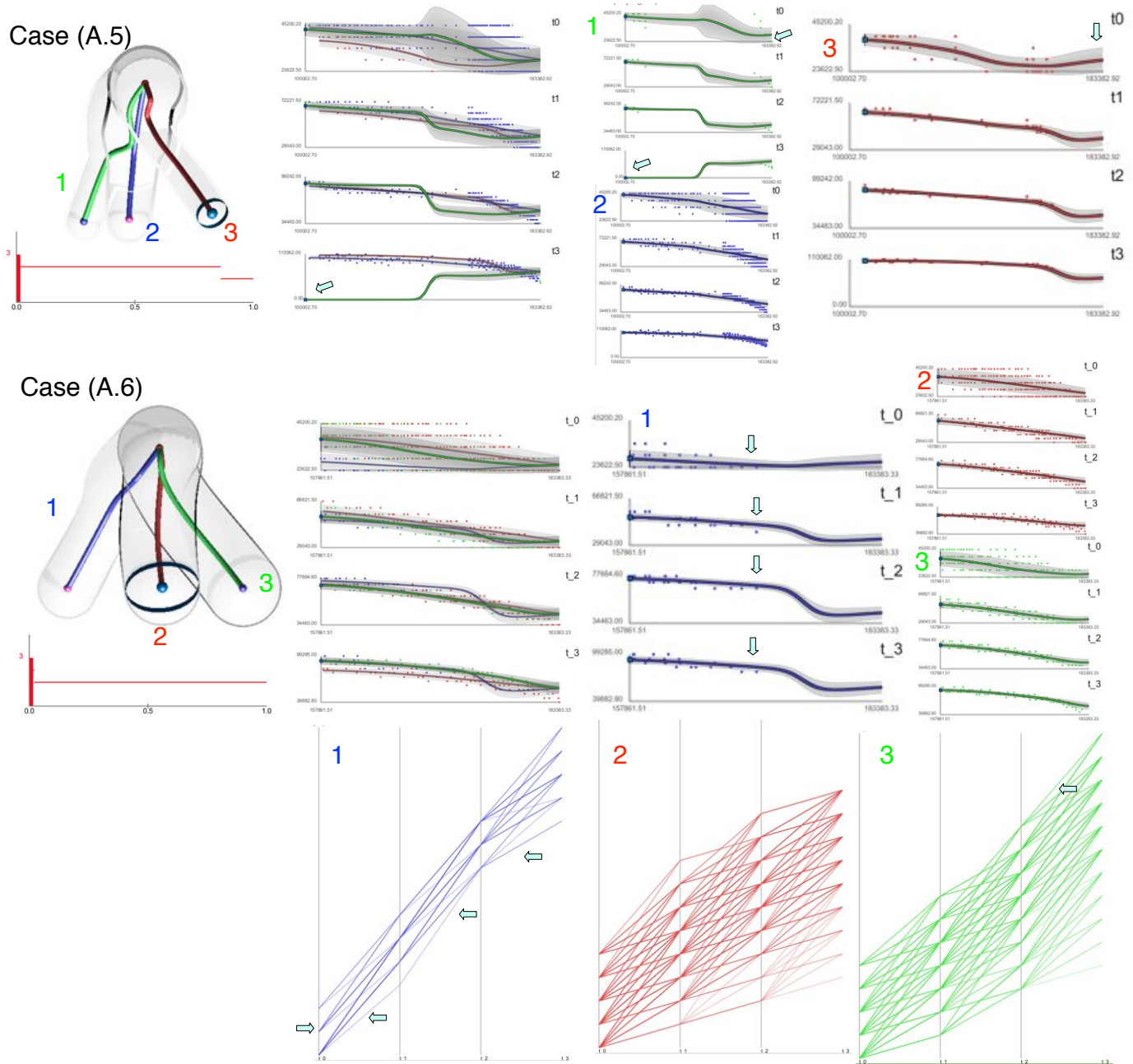


Figure 6. End state analysis, cases (A.5) and (A.6).

4D Absolute Timing Analysis. The results are shown in Figure 4 for case (A.1) and Figure 5 for cases (A.2) through (A.4) and Figure 6 for cases (A.5) and (A.6). We now describe each case in detail. For (A.1) **4D-MT-all-2C**, both crystals have maximum temperature of 1000K. Crystal 1 (red) contains scenarios where $t_3 \neq 0$, however, in these scenarios, the maximum temperature reaches the failure temperature of 1000K. This means that even though the crew managed to recover all three towers, since the recovery time for all towers is high, the core temperature keeps rising reaching system failure anyway. Crystal 2 (green) contains all scenarios where $t_3 = 0$, indicating all scenarios where the 3rd tower is not recovered and the failure temperature is reached. For (A.2) **4D-MT-completions-4C**, we only analyze simulation completion cases.

The four local maximum have temperatures 999.98K, 999.79K, 999.80K and 999.61K respectively. Crystal 1 (green) and 2 (red) have comparable range of values for t_0 , t_1 and t_2 , however crystal 1 contains data points with higher t_3 values compared to crystal 2, this indicates a later recover time for tower 3. Or in other words, recovery time for tower 3 differentiates these two crystals. Crystal 3 (blue) and 4 (purple) contain scenarios characterized by small values of t_0 and high values of t_1 , t_2 and t_3 , i.e., early crew arrival time and late recovery of the 3 towers. When looking at the combined inverse regression plots based on the ascending manifold, we see that the crystals 1 and 2 contain data points with higher t_0 , compared to the other two crystals, indicating late crew arrival time. For **(A.3) 4D-TTF-all-2C**, crystal 2 (green) contains all cases where $t_3 = 0$. For **(A.4) 4D-TTF-failures-3C**, we exclude non-failure scenarios. The two local maxima correspond to approximately 1.45×10^5 seconds and 1.54×10^5 seconds. The two local minima are roughly 1.01×10^5 seconds and 1.00×10^5 seconds. Crystals 1 (red) and 2 (green) both contain significant number of scenarios. However, they do not share the same local maxima, that is, time to reach failure in their local regions. Based on the combined inverse coordinates plot of crystals on the ascending manifold, they can also be differentiated by variations in the general trends of t_0 and t_1 . In particular, crystal 3 (blue) contains scenarios that lead to an early time to reach failure characterized by late crew arrival and late recovery of the first 2 towers. In **(A.5) 4D-TTM-all-3C**, the three local minimum corresponding to roughly 1.00×10^5 , 1.01×10^5 and 1.05×10^5 seconds. Crystal 1 (green) contains all scenarios where $t_3 = 0$, and some scenarios where the crew arrived at the site early (where t_0 is small) and the temperature peaked (reached its maximum) late. Both crystals 1 and 2 (blue) have some scenarios of the crew arriving early and the temperature peaking late, while crystal 3 (red) has none of the extreme cases. Finally, in **(A.6) 4D-TTM-completions-3C**, the three local minimum are very close in value, roughly around 1.58×10^5 seconds. Crystal 1 (blue) contains purely scenarios that reach maximum temperature early, with early crew arrival time and a large gap between time to fix the towers (see the large slope in its corresponding parallel coordinate plot). Between crystal 1 and crystal 3 (green), v.s. crystal 2 (red), the former two have higher gap between t_2 and t_3 (large slope in the parallel coordinate plots), meaning a big delay in recovering tower 3. Again, late crew arrival and late tower recovery times cause these scenarios to reach the maximum temperature early; while early crew arrival and early tower recovery times result in reaching maximum temperature late.

4D Relative Timing Analysis. The results are shown in Figure 7 with detailed descriptions below. We omit cases (B.2), (B.4) and (B.6) since they do not lead to drastically different conclusions. For **(B.1) 4D-Delta-MT-all-5C**, the two local maxima correspond to 991.66K and 1000K. The four local minima are valued at 990.48K, 990.5K, 988.61K and 990.15 respectively. Crystal 1(orange) contains only a single data point. However it is persistent across multiple scales up to two crystals. It also reaches the lower local maxima temperature of 991.66K, and corresponds to the largest t'_3 value (according to the parallel coordinate plot), indicating a very late recovering time for tower 3. Both crystal 3 (purple) and 4 (green) contain some cases where $t'_3 = 0$. Crystal 3 contains late crew arrival (e.g. large t'_0) scenarios and some early tower three fixing time (small t'_3). Crystal 2 (red) contains late t'_3 while crystal 5 (blue) contains late t'_2 and moderate t'_3 scenarios. In **(B.3) 4D-Delta-TTF-all-4C** and **(B.5) 4D-Delta-TTM-all-4C**, both corresponding crystal 1 (blue) contain all cases where $t'_3 = 0$ that is $t_3 = 0$, meaning no recovering of tower 3. These two cases have extremely similar clustering structure, which may be interesting to further investigate.

3.3 Transient Analysis Results

For transient analysis results, we first focus on case **(C.1) 100D-MT-all-4C**, illustrated in Figure 8. As shown in (a), the analysis results in 4 clusters, and the 4th cluster contains most of the scenarios in (d). As for MT profiles in (e), cluster 2 contains scenarios with the lowest MT, while cluster 3 contains some of

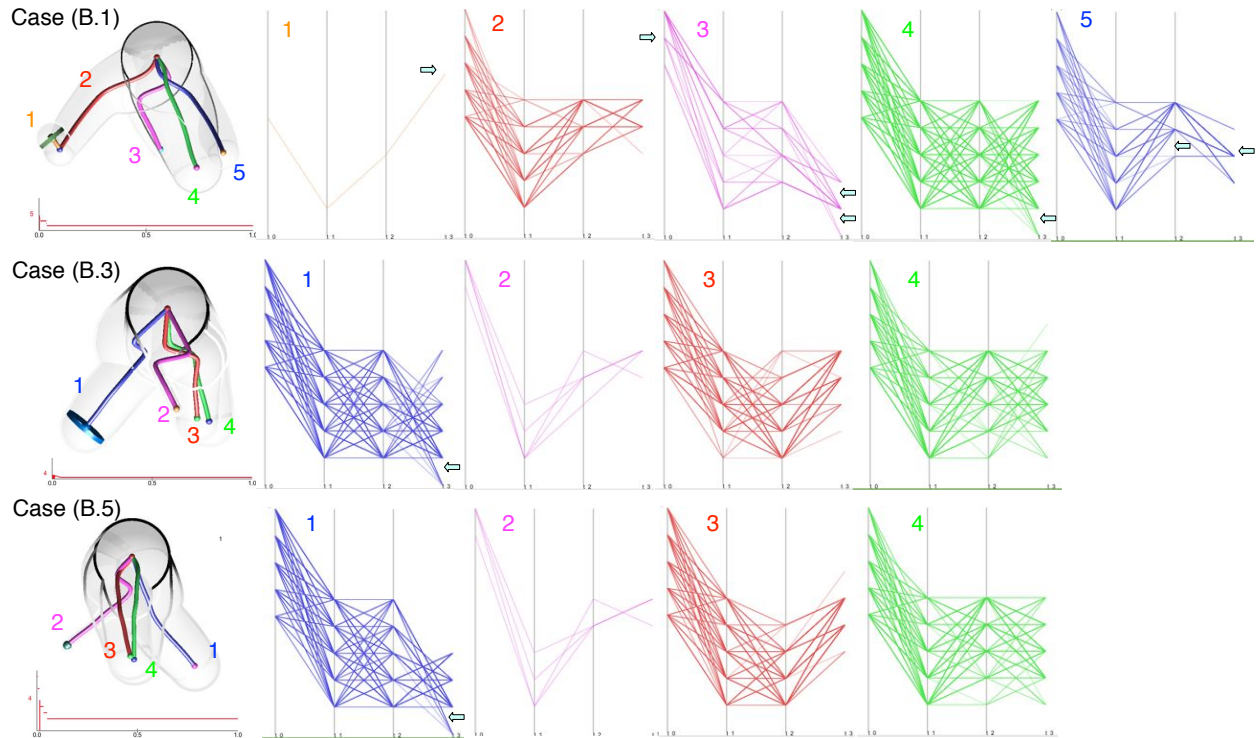


Figure 7. End state analysis, cases (B.1), (B.3) and (B.5).

those with the highest MT. For each cluster, the distributions of crew arrival time and the recover time of each tower are shown in (f). Note that for cluster 3 and 4, part of the distribution for t_3 is concentrated around 0 which correspond to the failure scenarios where the crew was not able to recover tower 3 on time (e.g. $t_3 = 0$). Cluster 3 is characterized by a fixed range of crew arrival time (t_0) and an extreme late crew recover time of tower 3 (t_3), this explains the high MT obtained by those scenarios.

Transient analysis results for (C.2) and (C.3) are shown in Figure 9. For (C.2) **100D-MT-completion-3C**, among all completion scenarios, cluster 2 contains those with the lowest MT, which is also characterized by very early crew arrival time and relatively early tower 3 recover time, leading to low MT. On the contrary, cluster 1 has late crew arrival time and late tower 3 recover time, leading to scenarios with some of the highest MT. In (C.3) **100D-TTF-all-2C**, cluster 3 contains those with the highest TTF, while cluster 1 contains those with the lowest TTF. Scenarios in cluster 1 reach failure temperature earliest because they are characterized by late crew arrival time and late recovering time for all towers.

For completeness, the remaining transient analysis results are shown in Figure 10. As we could observe from these figures, for cases (C.4), (C.5) and (C.6), there are clusters with distinctive patterns for the distributions of the variables t_0 , t_1 , t_2 and t_3 , differentiating them from one another.

Complementary Views. Now we could combine the above transient analysis results based on topological clustering, with previous results obtained through mean shift methodology ([10], Figure 19 and Figure 20). We argue that both methodologies offer complementary views of the nuclear simulation dataset. Further investigations are currently underway to understand how such views help the end users in the drawing of important conclusions with respect to safety analysis.

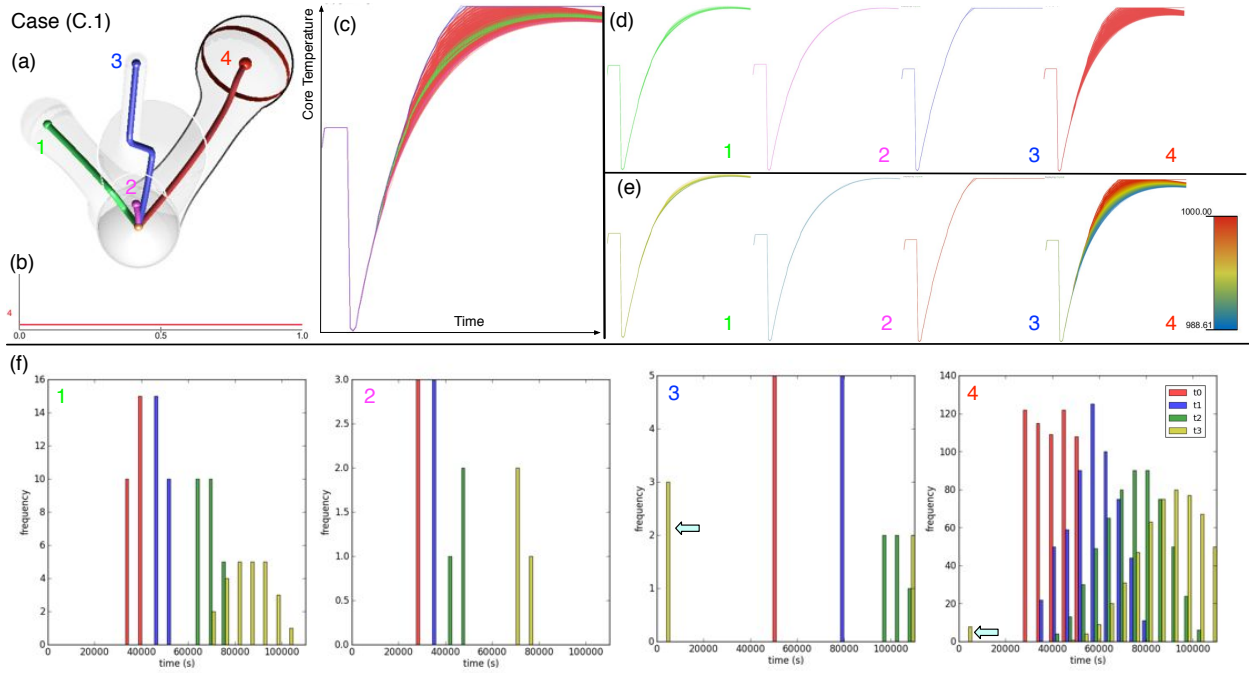


Figure 8. Transient analysis for case (C.1). (a) Topological summary with 4 crystals, corresponding to 4 clusters colored accordingly. (b) Persistence graph under the 4 crystal setting. (c) Parallel coordinate plot combining all 4 clusters, colored by cluster labels, showcasing time-varying profiles of all scenarios: x -axis corresponds to 100 values sampled along time, y -axis indicates core temperature. Parallel coordinate plots corresponding to each individual cluster, colored by cluster labels in (d) and by function values in (e). (f) Distribution of crew arrival time t_0 (red) and the recovery time of tower 1 (blue), 2 (green) and 3 (yellow) for each cluster.

4 Conclusions

In this paper, we investigate the use of a topology-based clustering technique on the data generated by dynamic event tree methodologies. Our clustering technique is inspired by Morse-Smale complex which partitions the data based on their uniform gradient behavior. We demonstrate our analysis technique and visual interface on a dataset generated for a sodium-cooled fast reactor during an aircraft crash scenario, and describe some interesting insights obtained from its end state and transient analysis. Combining with previously obtained mean-shift clustering results, and with further involvement by domain scientists, we expect to obtain in-depth understandings of such datasets.

In comparison to mean-shift clustering, each cluster has a representative center, but in an optimization or goal-seeking setting, a maximum/minimum representative of each cluster could be more beneficial to domain experts which is something that is readily available and inherent to topological clustering. It is also important to note that new topology-based clustering techniques could be inspired by the various topological structures such as contour trees and Reeb graphs. However understanding whether and how such techniques could produce meaningful results for analyzing DPRA-related nuclear datasets remains a challenge.

Analyzing Dynamic Probabilistic Risk Assessment Data through Topology-Based Clustering

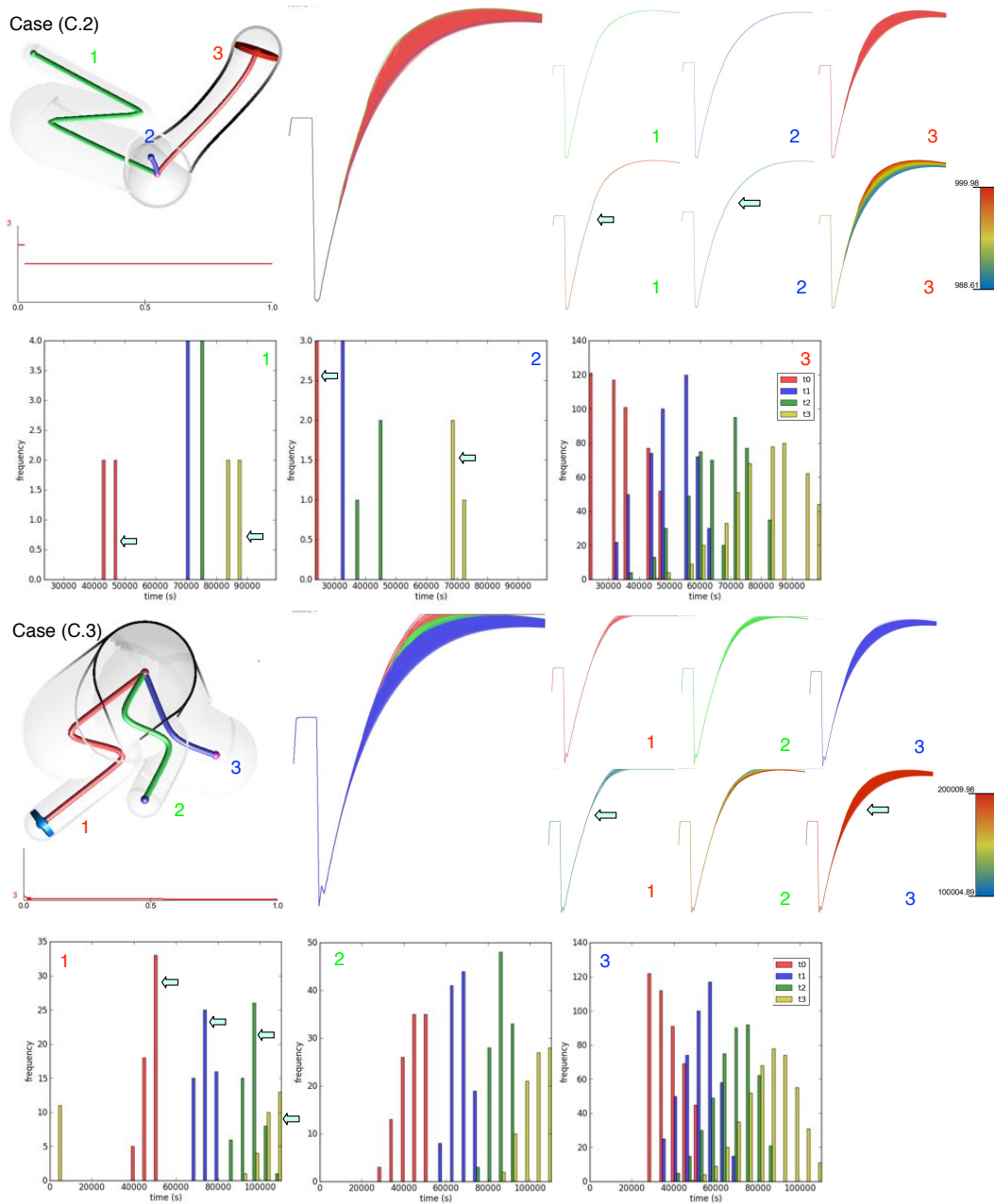


Figure 9. Transient analysis, cases (C.2) and (C.3).

REFERENCES

- [1] N. Siu, "Risk assessment for dynamic systems: an overview," *Reliability Engineering and System Safety*, vol. 43, no. 1, pp. 43–73, 1994.
- [2] RELAP5-3D Code Development Team, "Relap5-3d code manual," 2005.
- [3] R. O. Gauntt, *MELCOR Computer Code Manual, Version 1.8.5, Vol. 2, Rev. 2*. Sandia National Laboratories, NUREG/CR-6119.

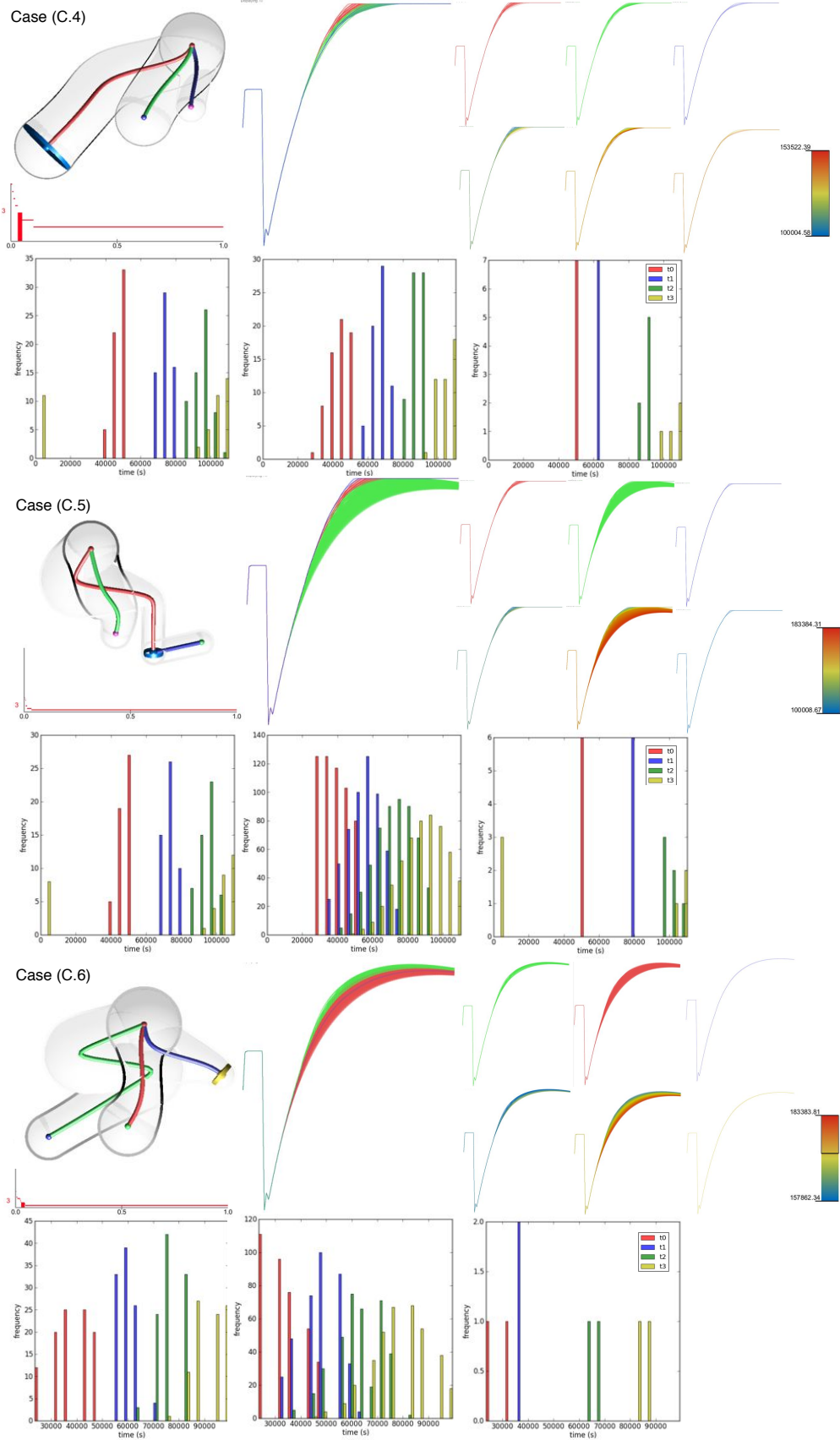


Figure 10. Transient analysis, cases (C.4) to (C.6).

- [4] D. I. Chanin, H. N. Jow, and J. A. Rollstin, *MELCOR Accident Consequence Code System(MACCS)*. Sandia National Laboratories, NUREG/CR-4691, Vols. 1-3, SAND86-1562, 1990.
- [5] A. Hakobyan, T. Aldemir, R. Denning, S. Dunagan, D. Kunsman, B. Rutt, and U. Catalyurek, “Dynamic generation of accident progression event trees,” *Nuclear Engineering and Design*, vol. 238, no. 12, pp. 3457–3467, 2008.
- [6] K. S. Hsueh and A. Mosleh, “The development and application of the accident dynamic simulator for dynamic probabilistic risk assessment of nuclear power plants,” *Reliability Engineering and System Safety*, vol. 52, pp. 297–314, 1996.
- [7] E. Hofer, M. Kloos, B. Krzykacz-Hausmann, J. Peschke, and M. Woltereck, “An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties,” *Reliability Engineering and System Safety*, vol. 77, pp. 229–238, 2002.
- [8] C. Rabiti, A. Alfonsi, D. Mandelli, J. Cogliati, and R. Martineau, “Raven as control logic and probabilistic risk assessment driver for relap-7,” in *Proceeding of American Nuclear Society (ANS), San Diego (CA)*, vol. 107, pp. 333–335, 2012.
- [9] F. D. Maio, M. Stasi, E. Zio, D. Mandelli, and T. Aldemir, “Identification of faults in a level control dynamic system,” in *Proceedings of NPIC-HMIT 2009, Knoxville (TN)*, 2009.
- [10] D. Mandelli, A. Yilmaz, T. Aldemir, K. Metzroth, and R. Denning, “Scenario clustering and dynamic probabilistic risk assessment,” *Reliability Engineering & System Safety*, vol. 115, pp. 146 – 160, 2013.
- [11] E. Zio, “Reliability engineering: Old problems and new challenges,” *Reliability Engineering and System Safety*, vol. 94, no. 2, pp. 125–141, 2009.
- [12] E. Zio and F. D. Maio, “Processing dynamic scenarios from a reliability analysis of a nuclear power plant digital instrumentation and control system,” *Annals of Nuclear Energy*, vol. 36, pp. 1386–1399, 2009.
- [13] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.
- [14] H. Edelsbrunner, J. Harer, and A. J. Zomorodian, “Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds,” *Discrete and Computational Geometry*, vol. 30, pp. 87–107, 2003.
- [15] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci, “Morse-Smale complexes for piecewise linear 3-manifolds,” *Proceedings 19th ACM Symposium on Computational Geometry*, pp. 361–370, 2003.
- [16] A. Gyulassy, V. Natarajan, V. Pascucci, and B. Hamann, “Efficient computation of Morse-Smale complexes for three-dimensional scalar functions,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, pp. 1440–1447, 2007.
- [17] J. Milnor, *Morse Theory*. New Jersey, NY, USA: Princeton University Press, 1963.
- [18] S. Gerber, P.-T. Bremer, V. Pascucci, and R. Whitaker, “Visual exploration of high dimensional scalar functions,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, pp. 1271–1280, 2010.
- [19] C. D. Correa and P. Lindstrom, “Towards robust topology of sparsely sampled data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, pp. 1852–1861, 2011.

- [20] D. Maljovec, A. Saha, P. Lindstrom, P.-T. Bremer, B. Wang, C. Correa, and V. Pascucci, “A comparative study of morse complex approximation using different neighborhood graphs.” Workshop on Topological Methods in Data Analysis and Visualization (accepted), 2013.
- [21] P. Bose, J. Cardinal, S. Collette, E. D. Demaine, B. Palop, P. Taslakian, and N. Zeh, “Relaxed gabriel graphs,” *Proceedings Canadian Conference on Computational Geometry*, 2009.
- [22] K. Gabriel and R. Sokal, “A new statistical approach to geographic variation analysis,” *Systematic Zoology*, vol. 18, pp. 259–278, 1969.
- [23] H. Edelsbrunner, D. Letscher, and A. J. Zomorodian, “Topological persistence and simplification,” *Discrete and Computational Geometry*, vol. 28, pp. 511–533, 2002.
- [24] H. Edelsbrunner and J. Harer, “Persistent homology - a survey,” *Contemporary Mathematics*, vol. 453, pp. 257–282, 2008.
- [25] R. Winningham, K. Metzroth, T. Aldemir, and 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 American Nuclear Society*, 2009.
- [26] R. Winningham, K. Metzroth, T. Aldemir, and R. Denning, “Passive heat removal system recovery following an aircraft crash using dynamic event tree analysis,” *Proceedings American Nuclear Society*, vol. 100, pp. 461–462, 2009.
- [27] D. Mandelli, A. Yilmaz, K. Metzroth, T. Aldemir, and R. Denning, “Scenario aggregation and analysis via mean-shift methodology,” *Proceedings International Congress on Advances in Nuclear Power Plants*, 2010.
- [28] D. Maljovec, B. Wang, V. Pascucci, P.-T. Bremer, M. Pernice, D. Mandelli, and R. Nourgaliev, “Exploration of high-dimensional scalar function for nuclear reactor safety analysis and visualization,” *International Conference on Mathematics and Computational Methods Applied to Nuclear Science & Engineering*, 2013.
- [29] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, pp. 2319–2323, 2000.
- [30] A. Inselberg, *Parallel Coordinates: Visual Multidimensional Geometry and its Applications*. Springer, 2009.