INL/EXT-12-27523

Exploration of High-Dimensional Scalar Function for Nuclear Reactor Safety Analysis and Visualization: A User's Guide to TopoXG*

Dan Maljovec
Bei Wang
Valerio Pascucci
Peer-Timo Bremer
Michael Pernice
Diego Mandelli

October 2012

The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

INL/EXT-12-27523

Exploration of High-Dimensional Scalar Function for Nuclear Reactor Safety Analysis and Visualization: A User's Guide to TopoXG*

Dan Maljovec[1]
Bei Wang[1]
Valerio Pascucci[1]
Peer-Timo Bremer[2]
Michael Pernice
Diego Mandelli

[1]SCI Institute, University of Utah
[2]Lawrence Livermore National Laboratory

October 2012

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Through the INL LDRD Program
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Exploration of High-dimensional Scalar Function for Nuclear Reactor Safety Analysis and Visualization: A User's Guide to TopoXG*

Dan Maljovec, Bei Wang, Valerio Pascucci
maljovec@cs.utah.edu, {beiwang,pascucci@sci.utah.edu}
SCI Institute, University of Utah

Peer-Timo Bremer
bremer5@llnl.gov
Lawrence Livermore National Laboratory

Michael Pernice, Diego Mandelli
michael.pernice@inl.gov, *diego.mandelli@inl.gov*
Idaho National Laboratory

Abstract

Large-scale simulation datasets can be modeled as high-dimensional scalar functions defined over a discrete sample of the domain. The goals of our proposed research are two-fold. First, we would like to provide structural analysis of a function at multiple scales and provide insight into the relationship between the input parameters and the output. Second, we enable exploratory analysis for users, where we help the users to differentiate features from noise through multi-scale analysis on an interactive platform, based on domain knowledge and data characterization. TopoXG is a software package that is designed to address these goals. The unique contribution of TopoXG lies in exploiting the topological and geometric properties of the domain, building statistical models based on its topological segmentations and providing interactive visual interfaces to facilitate such explorations. We provide a user's guide to TopoXG, by highlighting its analysis and visualization capabilities, and giving several use cases involving datasets from nuclear reactor safety simulations.

1 Introduction

The TopoXG software package provides scientists and domain experts with an interactive analysis and visualization environment for understanding the structures of high-dimensional scalar field data. It is provided with a host of various analysis and visualization capabilities. This report describes each of these capabilities on a modular basis, by explaining the underpinning theories and presenting use cases. TopoXG is designed as an integrated infrastructure with different modules to be used in conjunction rather than a myriad of standalone components.

TopoXG segments the domain of a high-dimension function into regions of uniform gradient flow by decomposing the data based on its approximate Morse-Smale complex. Points belonging to a particular segmentation have similar geometric and topological properties, and from these we can create compact statistical summaries of each segmentation. Such summaries are then presented to the user in an intuitive manner that highlights features of the dataset which are otherwise hidden. In addition, the visual interfaces provided by the system are highly interactive and tightly integrated, providing users with the ability to explore various aspects of the datasets for both analysis and visualization purposes.

In Section 2, we review the technical background on the Morse-Smale complex and its approximation in high dimensions; persistence simplification and visualizing high-dimensional data. We describe each analysis and visualization module in Section 3 with intuitive examples. We then present the specific applications of the software in analyzing and visualizing data from nuclear reactor safety simulations in Section 4.

*SCI Technical Report: UUSCI-2012-005 Version 1

2 Preliminaries

Here we give some technical background for our work and describe how our current and future research goals are related to the state of the art.

Morse-Smale Complex. Several topological structures have been used in practice for analyzing the structure of a scalar function, such as Reeb graphs [38, 39], countour trees [6, 31, 12], and Morse-Smale complexes [19, 7, 24, 25]. Each of these structures provides an abstract view of the data that highlights the salient features within the function. Algorithms exist to compute these topological structures in low dimensions [38, 31, 19, 7, 25], and recent developments [12, 23] have enabled us to compute or approximate these structures in high dimensions.

The body of work presented here relies heavily on the structure known as the Morse-Smale complex. The Morse-Smale complex is based on Morse theory [36, 35]. Let $\mathbb{M}$ be a smooth manifold embedded in $\mathbb{R}^n$ without boundary and $f : \mathbb{M} \to \mathbb{R}$ be a smooth function with gradient ∇f. A point $x \in \mathbb{M}$ is called *critical* if $\nabla f(x) = 0$, otherwise it is *regular*. At any regular point x the gradient is well-defined and integrating it in both (ascending/descending) directions traces out an integral line, which is a maximal path whose tangent vectors agree with the gradient [19]. Each integral line begins and ends at critical points of f. Therefore all regular points can trace their ascending integral line to a local maximum. Similarly, tracing the descending integral line of a regular point will associate a point with a local minimum. The *unstable/stable manifolds* (or *ascending/descending manifolds*) of a critical point p are defined as all the points whose integral lines start/end at p. The set of intersections of unstable and stable manifolds creates the *Morse-Smale* complex of f. Each cell (*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 shows these individual partitions of a scalar function over a 2D surface.

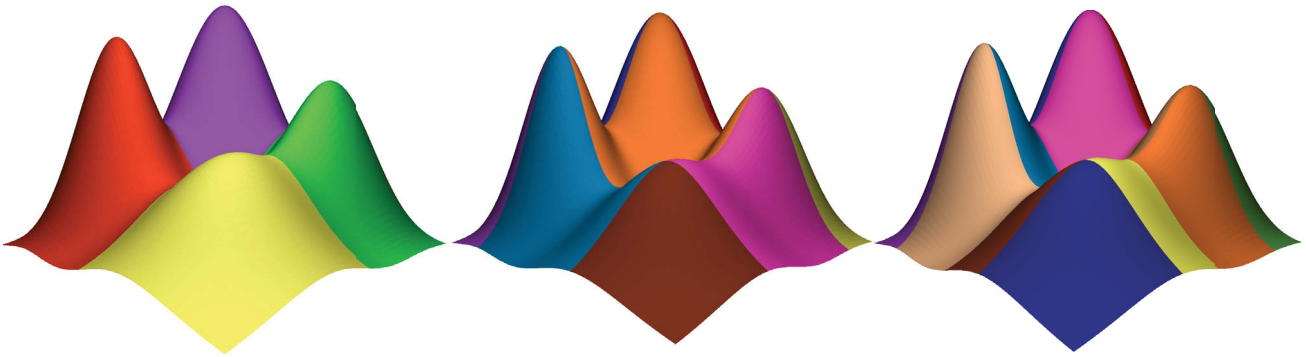

Figure 1: From left to right: unstable manifolds, stable manifolds, Morse-Smale complex. Each partition is colored by its corresponding gradient behavior, where points share the same color if: (Left) their ascending gradient flow end at the same local maximum; (Middle) their descending gradient flow end at the same local minimum and (Right) their gradient flow begin and end at the same maximum-minimum pair.

Approximating the Morse-Smale Complex in High Dimension. Suppose our input domain is a finite set of points $\mathbb{X}$ in $\mathbb{R}^n$ rather than a smooth n-manifold. To approximate the Morse-Smale complex in high dimension, our first task is to estimate the gradient at the input points $\mathbb{X}$ by employing a version of the quick shift algorithm [44]. First we compute a neighborhood graph such as the k-nearest neighbor graph of $\mathbb{X}$. At each point in $\mathbb{X}$, we choose the steepest ascending/descending edge to represent the gradient. With this gradient approximation, we can determine the local extrema by labeling all points with no neighbors of higher values as local maximum and all points with no neighbors of lower values as local minimum. We then label all points in $\mathbb{X}$ according to the local extrema at which its ascending and descending gradients terminate. 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 then serve as an approximation of the Morse-Smale complex [23]. One of our research goals is to understand how different neighborhood graphs (i.e. empty region graphs [8]) may impact our approximations, and what sampling conditions we should impose on the data to guarantee the approximation quality.

Persistence Simplification. We introduce the notion of *scale* for learning the structure of a function defined on a point cloud through the concept of *persistence*. Persistence studies the evolution of vectors in a sequence of vector spaces [14]. One main example of such a sequence comes from the homology groups of a sequence of sublevel sets of a real-valued function. Homology is an easily computed topological invariant, where homology features are components, tunnels, voids and high-dimensional "holes" of a space; a background is given in [37, 26]. Persistence provides a way of ranking the significance of the topological (homological) features in the sublevel sets of a real-valued function and is essential to achieve the robustness of our methods. The theory of persistence was first introduced in [21, 11], but borrows from the conventional notion of the saliency of watersheds in image segmentation. It has since been applied to a number of problems, including sensor networks [17], surface description and reconstruction [9], protein shapes [18] and images analysis [10]. In visualization, it has been used to simplify Morse-Smale complexes [7, 20, 24], Reeb graphs [16, 39] and contour trees [42, 13].

In real datasets, there is often noise which may manifest itself as small topological artifacts, which are 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 (complexity level) appropriate for the specified dataset, we introduce the notion of *persistence simplification* whereby less salient features are merged with neighboring, more significant features. In the case of the Morse-Smale complex, we assign a persistence to each critical point in the complex which, intuitively, describes the scale at which a critical point would disappear through simplification. An example of this simplification is shown in Figure 2, where at the finest level there are four maxima in the data, but as we increase the scale, neighboring topological features are merged where the upward gradient flow is directed to its more salient neighboring maxima. We thus build a filtration of segmentations where Morse-Smale crystals are merged based on the persistence value of their associated extrema.

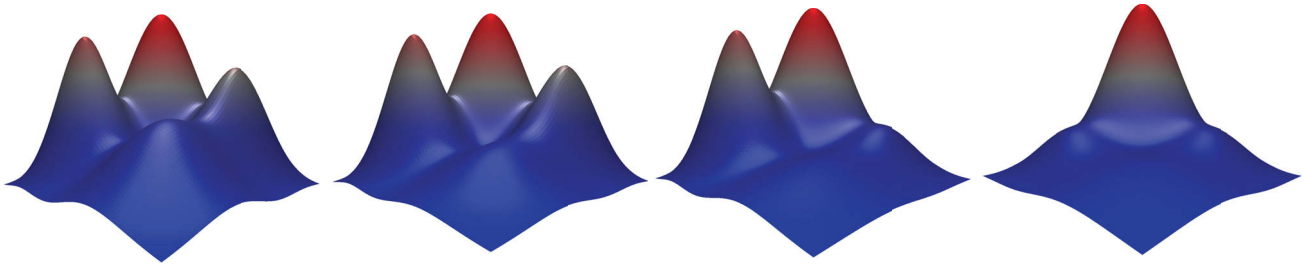

Figure 2: Progressive simplification of the critical points of a 2D function.

Visualization in high dimensions. A common approach to visualize the high-dimensional data is based on projecting them onto one-, two- or three-dimensional subspaces and show labeled scatter plots or some interpolation of those densities. Common projection approaches focus on linear subspaces, such as those used by PCA and projection pursuit [22]. Slight variations of these methods look at sequences of projections onto different directions or multiple subspaces, such as Andrew's plots [1], parallel coordinates [29] and the grand tour approach [3]. The high-dimensional data is embedded onto the hyperbolic plane in [45], and 3D-sphere in [28]. Recently, specific tools have been developed to facilitate interactive visualization of high-dimensional data in conjunction with nonlinear dimensionality reduction techniques, such as VisuMap [32], VisHD [48] and Hyperbolic MDS [45]. Tools targeted towards specific applications in high-dimensional data visualization include design galleries [34] for exploring the parameter space of transfer functions in volume rendering, and Click and Expander [41] for identifying clusters in gene expression data. In the machine learning community, visualization of the low-dimensional embedding is a common strategy for quantifying the effectiveness of manifold learning approaches [5, 27, 40, 43]. Our work uses parallel coordinate plots, and adapts several high-dimensional projection schemes, including PCA, ISOMAP[43] and Hypervolume visualization [4]. We plan to expand our visualization capacities by designing more modules that employ a wide variety of projection and visualization techniques in high dimensions.

3 Analysis and Visualization Modules

Our software is an extension to HDViz [23]. We now describe each analysis and visualization module within our integrated system that are either part of the original capabilities provided by HDViz or part of the extension. The main modules are visual interfaces designed for analysis and visualization: topological summary, statistical summary, parallel coordinate plots, pairwise scatter plots, inverse coordinate plots and interactive projection. We demonstrate our infrastructure with a demo example from nuclear plan safety analysis.

3.1 Demo Example

The demo example we use is 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 [46, 47, 33]. 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 [46, 47, 33]. During a simulation, the plant is operating at 100% power when an airplane crashes into the plant, destroying three of the four towers. A recovery crew then arrives at the site and attempts to reestablish the capability of the reactor by restoring the damaged towers one by one. An ensemble of 610 simulations has been generated, and among which 132 cases are considered system failures when the reactor reaches a maximum temperature of 1000K before the end of simulation. These failure cases are extracted for our analysis.

The data from these 132 simulations is visualized as a four-dimensional scalar function with the time to reach the failure temperature being the output variable of interest. 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 $t_3 = 0$. We would like to understand how these input variables impact system dynamics [47, 33]. Understanding the structure of such a four-dimensional function may help domain scientists to make decisions regarding repair strategies and evacuation plans.

3.2 Topological Summary

The visual interface designed for topological summary is inherited and extended from the capabilities provided by HDViz. Considered as the main visual display of TopoXG, this interface summarizes each Morse-Smale crystal into a 1D curve in high-dimensional space which is then projected into a viewable 3D space. Each curve is encased in a transparent tube where the width of the tube represents the "spread" of the data at a particular scale, and the luminance of the tube encodes the density of data within each crystal. The interface encodes three steps to arrive at a 3D representation for analysis and visualization of the d-dimensional scalar function f defined on a point cloud $\mathbb{X}$ [23]: (1) *Morse-Smale Approximation*: Compute segmentation $\mathbb{X}_i$ and $\mathbb{Y}_i = f(\mathbb{X}_i)$ using a Morse-Smale complex approximation of f, where $\bigcup \mathbb{X}_i = \mathbb{X}$; (2) *Geometric Summaries:* Construct regression curves r_i as a geometric summary of each segmentation $\mathbb{X}_i$ and $\mathbb{Y}_i$; and (3) *Dimension Reduction:* Embed regression curves in 2D using a two-step dimension reduction approach. The third dimension is reserved for the output parameter. We give a high level description of this process, for details, see [23].

Morse-Smale Approximation. We first approximate the Morse-Smale crystals in high dimension using an approximate k-nearest neighbor graph [2] and the quick shift algorithm as detailed in Section 2. However, this is only one of the many common approximation schemes. One of the goals of this work is to understand the possibilities for approximating these complexes and to understand the implications of those approximations.

Geometric Summaries. For each crystal of the Morse-Smale complex, a geometric summary is constructed by an inverse regression. This yields a 1D curve in the d-dimensional domain of f. Formally, the input domain for each crystal C_i with samples $(\mathbb{X}_i, \mathbb{Y}_i)$ is summarized by a parametric curve r_i in $\mathbb{R}^d$. Modeling the curve by the conditional expectation $r_i(y) = E[\mathbb{X} \in C_i | \mathbb{Y}]$ yields a representation of the crystal C_i as the average of the level

sets $\{x : f(x) = y, x \in C_i\}$ within the partition. The conditional expectation is then estimated with locally linear regression [15]. We refer the reader to [23] for its derivation.

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. For visualization, we embed this graph into the plane preserving the spatial relation among extrema and the geometry of the partitions that connect them. It is important to point out that the goal of this dimension reduction is to provide an informative illustration of f rather than manifold learning of $\mathbb{X}$. As a first attempt, we compute the projection into the plane using a three-step approach: first, vertices are embedded using PCA or ISOMAP [43] on the corresponding point set; second, edges are embedded individually through their first two principle components; and third, the resulting two-dimensional curves are attached to the projected vertices through affine transformations. We plan to apply more general graph embedding techniques to provide representations suitable for visual exploration.

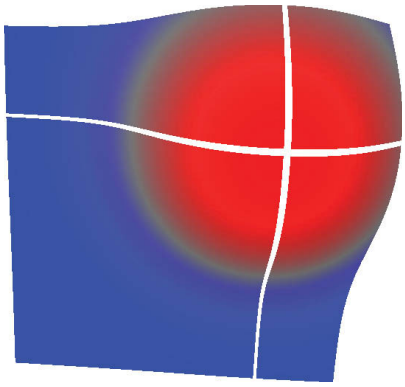

Figure 3: A simple 2D function is segmented into four Morse-Smale crystals.

Visual Components of the Framework. The visual components of the above framework are shown in Figure 4 for a simple 2D function with four Morse-Smale crystals (Figure 3). Users are given the flexibility to view the topological summary of a high-dimensional function by switching between PCA and ISOMAP projections, and using affine transformations to manipulate the projection directly on the screen. PCA finds and projects data into a space defined by principal components which are linearly uncorrelated directions of highest variance in the data. The ISOMAP differs by respecting geodesics of high dimension points which sometimes can more faithfully convey the geometry of the high-dimension manifold in lower dimensions.

In order to preserve the "width" of a crystal at a given scale, we compute the standard deviation with respect to each input parameter and also a single *average* standard deviation across all input parameters. The latter is direction-independent and can therefore be used as a generalized width of the data at a particular output level. The radius of the outer transparent tubes are defined by this direction-independent standard deviation. The last visual cue is the darkness of the edge of the transparent tube. Where the sampling density is high, the outline of the tube is drawn black and as the sampling density decreases, the luminance of this edge increases.

To enable multi-scale analysis, we use a modified version of the persistence diagram [21], referred to as the *persistence graph*. This is shown as a visual component at the bottom of Figure 4. 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 at its corresponding simplification level 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.

Figure 5 shows several examples of simple 2D surfaces represented using the techniques described in this section. Each function has one local maximum and a single local minimum surrounded by a flat area. Using an evenly

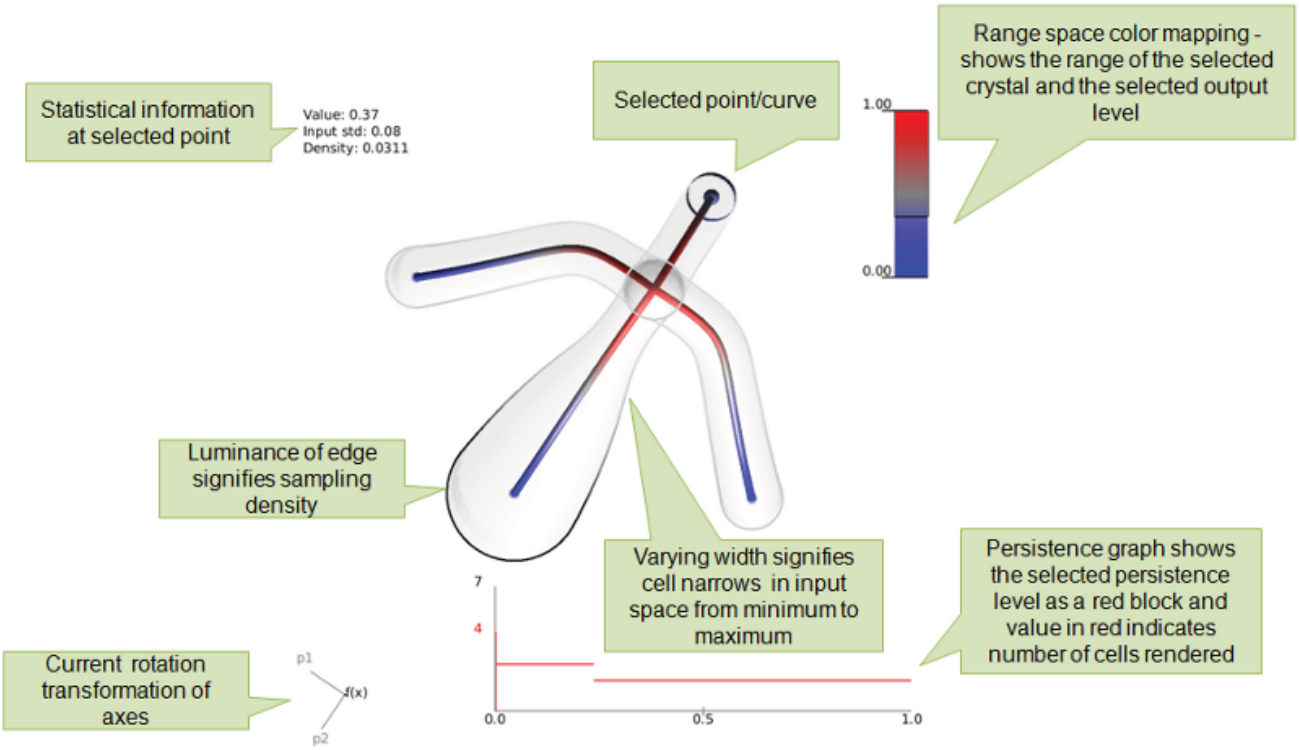

Figure 4: The topological summary visual interface of the simple 2D function shown in Figure 3.

sampled grid, the density of points available at each level is mapped to the color on the edge of the transparent tubes. The widths of the tubes vary with respect to the spread of the data at each level set. Note how the sharp spike in (a) is represented in the topological display. The steep peak covers a small area and the small width of the transparent tube demonstrates this in the area near the maximum. (b) has a wider area surrounding the peak and while its representing curve remains a single straight line, the tube surrounding it changes behavior to account for the width of each level set. The plateau function in (c) even maps the discontinuity in the curve by making a sudden jump from a low value to a high value. Note how the edge of the tube has high luminance in the middle section denoting a lack of data used to compute this section, whereas the most densely sampled region is at the maximum value which has a black outline.

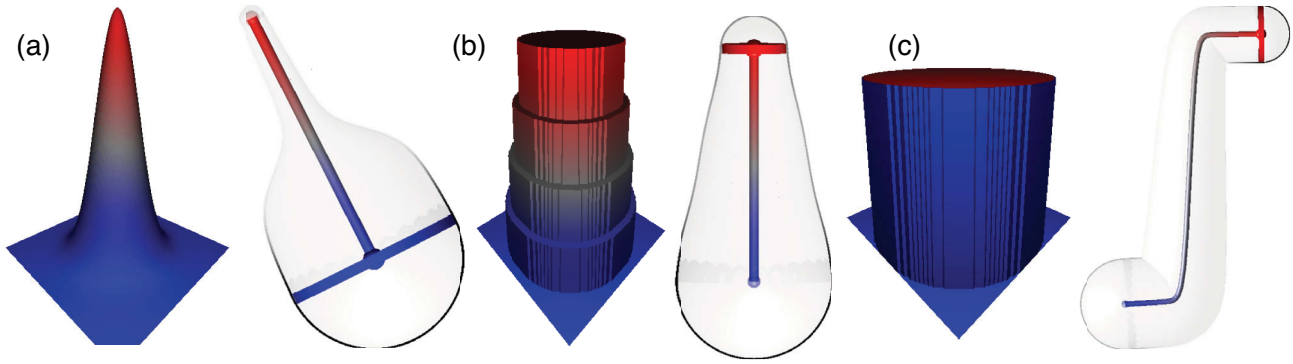

Figure 5: Three simple surfaces represented using geometric summary tubes.

We demonstrate how the topological summary visual interface allows the users to explore the data at multiple scales. We look at a dataset (initially shown in Figure 2) under several levels of persistence simplification in Figure 6. Here the leftmost image shows the full resolution of data with four local maxima, while each subsequent image

reduces the number of maxima by one until we are left with a single crystal describing the gradient flow from the global minimum to the global maximum. The numbers in red shown in the persistence graphs indicate the total number of crystals displayed, from left to right, as 8, 6, 4 and 1. Instead of giving the users a representation of the data at a fixed scale, we provide an interactive platform to help them differentiate features from noise through multi-scale analysis and to choose the appropriate scale based on domain knowledge and data characterization.

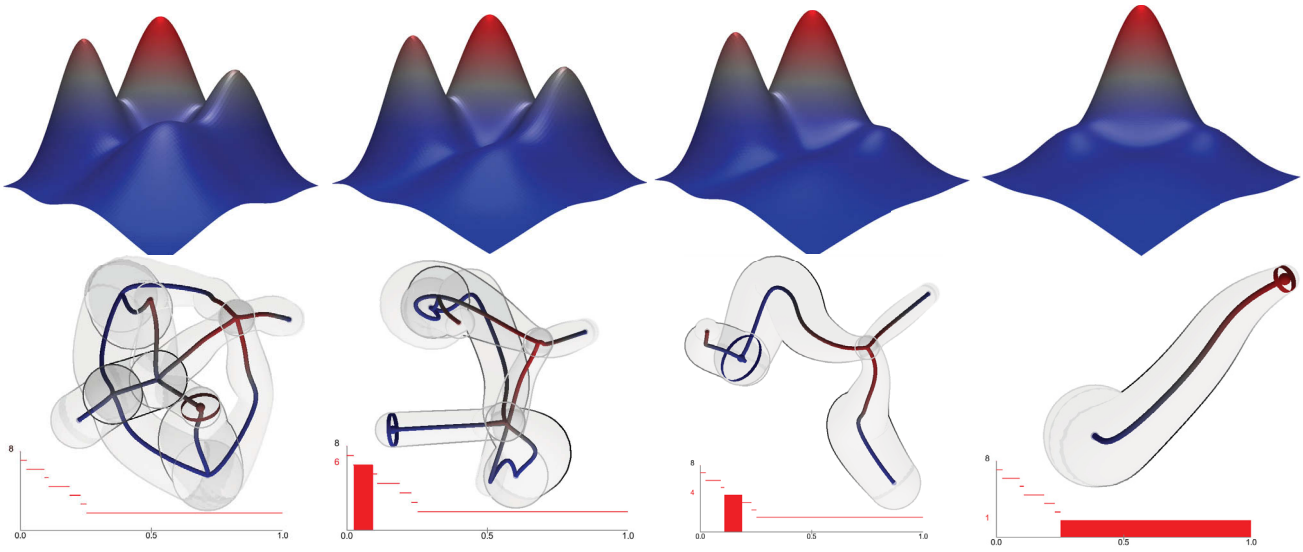

Figure 6: The topological summary of a simple dataset shown at various scales.

As a final example, we illustrate our analysis and visualization interface with our 4D demo example in Figure 7 under multiple scales. During the persistent simplification steps, the topology summaries consist of 3, 2 and 1 Morse-Smale crystals, respectively.

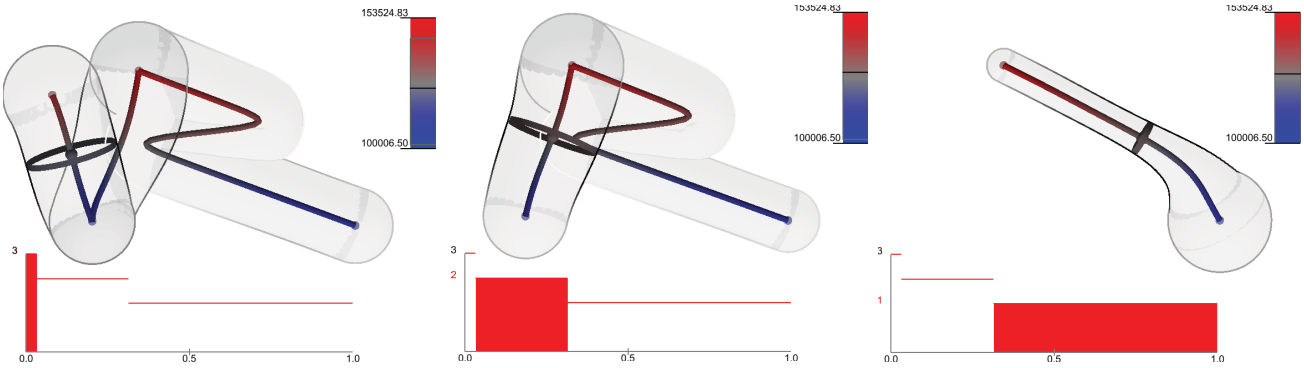

Figure 7: The topological summary of our demo dataset shown at various scales.

3.3 Statistical Summary

The visual interface shown in Figure 8 demonstrates statistical and geometric (i.e. gradient) information associated with the selected point in the topological summary interface. Each input dimension is viewed as an inverse function of the output parameter. In the left column of the statistical summary window, each horizontal axis describes the range of values of each input parameter and the coordinate mean and coordinate standard deviation associated with the selected point. The right column encodes the gradient information, that is, the change in the output with respect to the change in each input parameter.

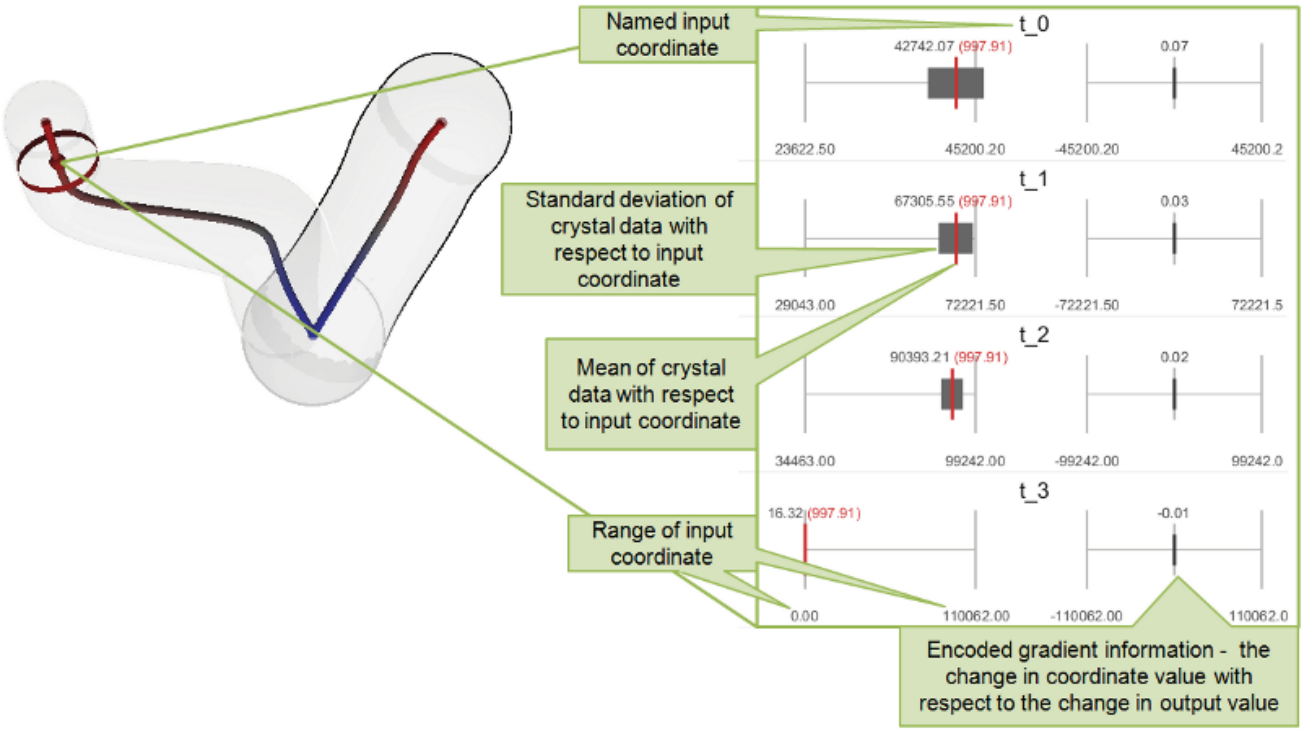

Figure 8: A snapshot of the statistical summary visual interface with highlighted visual components.

3.4 Parallel Coordinate Plots

We use parallel coordinate [30] plots to illustrate the correlations among input parameters in the high dimensional datasets. We showcase this visual interface using our demo example of a four dimensional function shown in the left column of Figure 9. In addition to allowing visual exploration of data points associated with each Morse-Smale crystal, we also enable exploration of data points clustered by stable manifold (union of Morse-Smale crystals sharing the same minima) and unstable manifold (union of Morse-Smale crystals sharing the same maxima). In the parallel coordinate plots, each dimension is plotted as a vertical axis. A single line connecting two parallel axes represents the two adjacent dimensions of a given data point. The variations among lines connecting the four dimensions (i.e. t_0, t_1, t_2 and t_3) indicate correlations among various dimensions of the datasets. In addition, the mean value (from the inverse regression) of each selected crystal (or stable/unstable manifold) is mapped as a bold line in the parallel coordinate plots, and a grey area surrounding it denotes one standard deviation off the mean. One important question associated with parallel coordinate plots is the ordering of the input parameters, we plan to use statistical techniques to infer the dependencies among these parameters to obtain the optimal order in the visual layout.

3.5 Pairwise Scatter Plots

Similar to the parallel coordinate plots, we also demonstrate the 2D pairwise scatter plots for our demo example (as shown in the right column of Figure 9), for data points associated with each Morse-Smale crystal (or stable/unstable manifold). Here, the user is presented with a series of 2D scatter plots that display each input parameter against every other input parameter, where the colors of points are mapped to the values of the output. The regression curve of a selected crystal is also projected onto the scatter plot. The transparent grey tube surrounding it represents one standard deviation. However, as dimension of the dataset increases, pairwise scatter plots will eventually become a dense visual clutter, rendering the interface ineffective. We are currently investigating ways to reduce such a visual clutter.

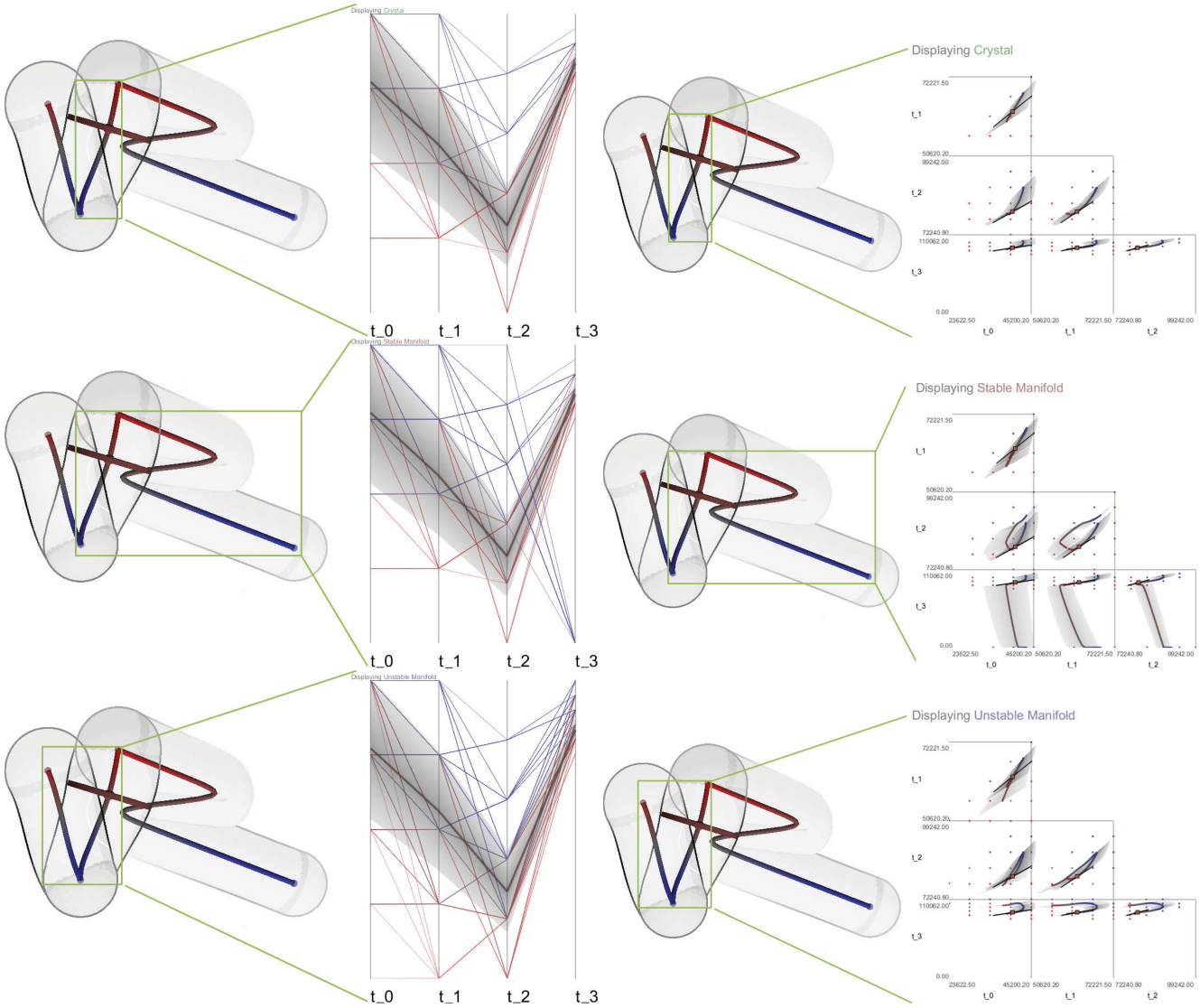

Figure 9: Left column: Parallel coordinates plots. Right column: Pairwise scatter plots. From top to bottom: explorations of data points associated with Morse-Smale crystals (top), stable manifold (middle) and unstable manifold (bottom).

3.6 Inverse Coordinate Plots

In the inverse coordinate plots, each input parameter is considered as a one-dimensional function of the output variable. Again, this visual interface is shown in Figure 10 for our four-dimensional demo example, for data points associated with selected Morse-Smale crystal in the topological summary interface (or stable/unstable manifold). For a selected crystal, its regression curve is drawn in the inverse coordinate plots with a grey tube representing coordinates within one standard deviation.

3.7 Interactive Projection

The interactive projection visual interface (shown in Figure 11) is similar to the topological summary, except that instead of using computed projections into screen space, like PCA and ISOMAP, the data remains in high-dimension and a graphical user interface is used to define how the data is projected into the screen space. We discuss the theory behind this interface and its implementation in TopoXG.

As the number of dimensions increases for a high-dimensional function defined on the sampled point, selecting the interesting projection dimensions while interacting with the high-dimensional function could become counter-

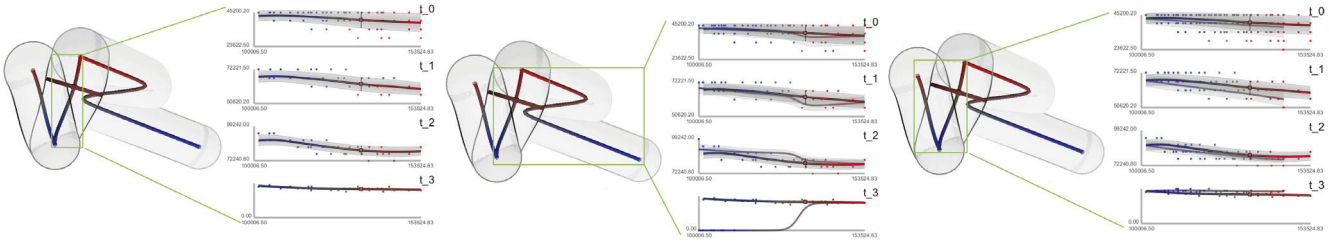

Figure 10: Inverse coordinate plots. From left to right, explorations of data points associated with selected Morse-Smale crystal, stable manifold and unstable manifold.

intuitive. We design an interactive visualization interface that provides simple and fully explanatory pictures that give comprehensive insights into the global structure of the high-dimensional function. Based on hypervolume visualization techniques developed in [4], our basic idea is a generalization of direct parallel projection methods. We first create an independent viewing system that scales with the number of dimensions where the user is allowed to manipulate how each axis is projected. We then apply these manipulations to project the geometric summaries from the functional space to the screen space. A possible extension would be to create an automated process for finding "ideal" projections, where a user would select multiple points that they would like to manipulate with bounded distance metric in the projection, and the system would solve an optimization problem that produces the desired projection. We believe such a system would help the user to obtain multiple "local" views of the function by exploring combinations/correlations of variables in a systematic way, therefore eventually forming a "global" view of the data [4].

To implement such a technique in TopoXG, consider a point $x \in \mathbb{R}^n$ to be projected onto $\mathbb{R}^2$ using a $2 \times n$ matrix $\mathbf{L}$. Let $\mathbf{e_i}$ be the unit vector of the i-th dimension. Its projection onto the 2D image space is obtained by multiplying $\mathbf{L}$:

$$\mathbf{v_i} = \mathbf{L}\mathbf{e_i} = \begin{bmatrix} l_1^x & \cdots & l_n^x \\ l_1^y & \cdots & l_n^y \end{bmatrix} \begin{bmatrix} e_1^i \\ \vdots \\ e_n^i \end{bmatrix}$$

Since each unit vector has only one non-zero component, this multiplication directly extracts the i^{th} column of $\mathbf{L}$. We therefore design an interface where the user manipulates in 2D the direction and magnitude of each projected vector $\mathbf{v_i}$, that is, the user decides how each unit vector is projected on to the 2D image space. The combined projected vectors $\mathbf{v_i}$ are then used to construct the projection matrix $\mathbf{L}$.

Users are provided with a wheel of labeled axes which they can manipulate by stretching, contracting, and rotating. The geometry of the inverse regression curves therefore is meaningfully preserved. With a single manipulated projection displayed at a time, the interactive nature of such a tool provides the user with intuition that is otherwise lost in a PCA or ISOMAP projection. Currently this system only supports a view of the geometric summaries, but a possible extension is to create a full hypervolume visualization of the raw high-dimensional data points as proposed in [4]. Efficiency of such a visualization system relies on a hierarchical representation of the high-dimensional function for multi-resolution display and generalized object space splatting combined with texture-mapped graphics hardware acceleration [4].

4 Applications in Nuclear Safety Analysis

Here we present the specific applications of TopoXG in the analysis and visualization of a 6-dimensional and a 21-dimensional datasets extracted from a VR_2^+ nuclear reactor simulator. Each dataset is an ensemble of 10000 simulation trials. The output variable is the peak coolant temperature (PCT) in both cases. The domain scientists are interested in what combination of conditions (in the form of input parameters) can cause potential reactor failure (i.e. nuclear meltdown witnessed by PCT exceeding a threshold value).

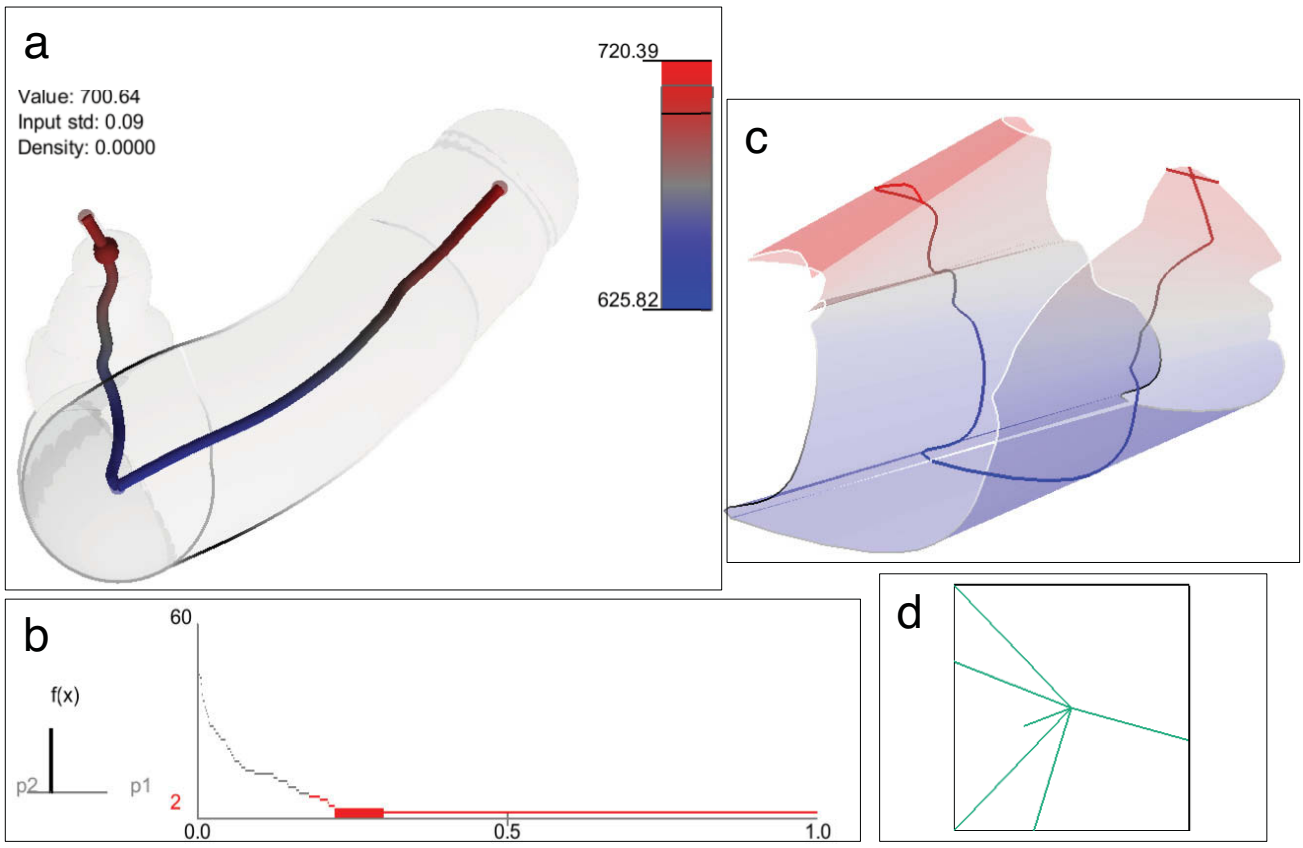

Figure 11: Interactive Projection Visual Interface. The data shown has a 6-dimensional input space. (a)-(b) represents the 2-step PCA topological summary detailed in section 3.2. The data is segmented into two crystals shown in (a) based on a chosen scale in (b). (c) shows the projection of the high-dimensional inverse regression curves and their associated standard deviation tubes, based on the mapping shown in the square in (d). In (d), each input dimension is mapped to a single green line segment, which the user could manipulate by stretching and rotation, to emphasize or diminish the effect of a particular dimension on the projection.

4.1 6-Dimensional Nuclear Safety Simulation Data

The 6-dimensional VR_2^+ dataset contains 10000 sample points (each corresponds to a simulation run). The input space is defined by the parameters listed below:

- $P_{\mathrm{Pump}}^{\mathrm{Trip}}$: Minimum pressure in the pump (PumpTripPre)
- τ_{Pump}: Relaxation time of pump's phase-out (PumpPow)
- Pump Stop Time: End power of the pump (PumpStopTime)
- $T_{\mathrm{max}}^{\mathrm{SCRAM}}$: Maximum temperature in the system causing the SCRAM (SCRAMtemp)
- $\mathrm{CRP}_{\mathrm{end}}$: Control rod position at the end of SCRAM (CRinject)
- τ_{CRP}: Relaxation time of the control rod system (CRtime)

The topological summary of the dataset is shown in Figure 12. The persistence graph at the bottom of Figure 12 indicates that the dataset is highly complex at the finest scale, with 775 Morse-Smale crystals. We further simplify such a complex structure by increasing the persistence threshold and arrive at a certain simplification with only six of the most significant (persistent) crystals. The data is characterized by a single global minimum with high sampling density. The widths of all the crystals at the minimum is expansive compared to the widths at the maxima. We could infer from this analysis result that most of the data points (simulations) represent lower PCTs and the conditions to reach these lower PCTs varies widely in the domain space. On the other hand, the maximum

PCTs are reached at more specific input ranges which is made clear through the narrow widths of the tubes (of one standard deviation) surrounding the red maxima. Two of the crystals, a and e, have wider tubes surrounding their local maxima, signifying that their corresponding data points occupy a larger area of the input space. The local maxima associated with these two crystals may be worth investigating further as they can be reached within a less-restricted range of input parameter settings.

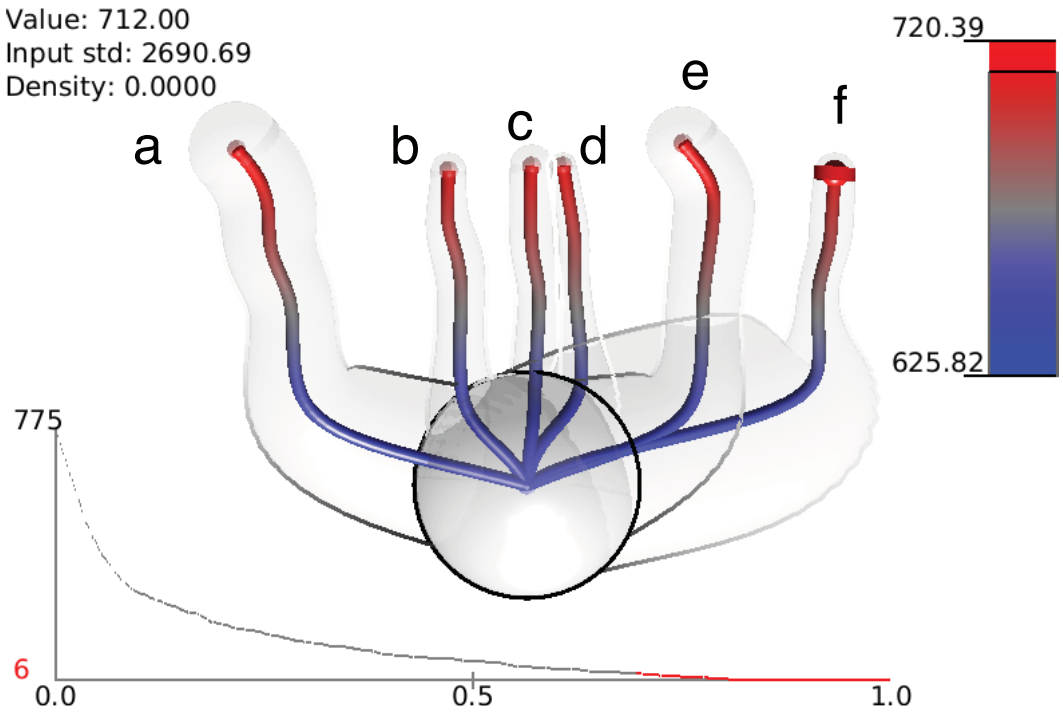

Figure 12: The topological summary visual interface of the VR_2^+ 6D dataset shown with six separate Morse-Smale crystals that share a single, global minimum, at a certain scale.

To investigate how various input parameters vary with respect to the output, we shift our attention to other visual interfaces for each selected Morse-Smale crystal: the inverse coordinate plots (Figure 13), the statistical summary visual interface (Figure 14), and the parallel coordinate plots (Figure 15).

From these visual interfaces, it is evident that, among the six crystals, the only input that changes significantly is the PumpTripPre parameter. For example, let us focus on the first vertical axes in Figure 15, which represent the PumpTripPre parameter. Across the six crystals, the PumpTripPre parameter shows different clustering behavior along the first vertical axes. In crystal a (top leftmost plot), PumpTripPre values are concentrated on the upper range while in crystal f (bottom rightmost plot), their values are clustered in the lower range. From the parallel coordinate plots, one could conclude that various levels of PumpTripPre can result in high temperatures. In addition, we could also combine the inverse coordinate plots of all Morse-Smale crystals together into one stable manifold and visualize the combined plot in Figure 16. We see the top image where each of the five regression curves is readily distinguishable with respect to different ranges of values for PumpTripPre, whereas in the lower plots they vary little toward the left and only slightly at the right.

On the other hand, parameters such as PumpPow and CRinject result in high temperatures only at specific levels. This conclusion is based upon the tight configuration of points resulting in high PCT evidenced in Figure 16 and it is also supported by the consistent locations of the mean values and low standard deviations of these parameters in Figure 14.

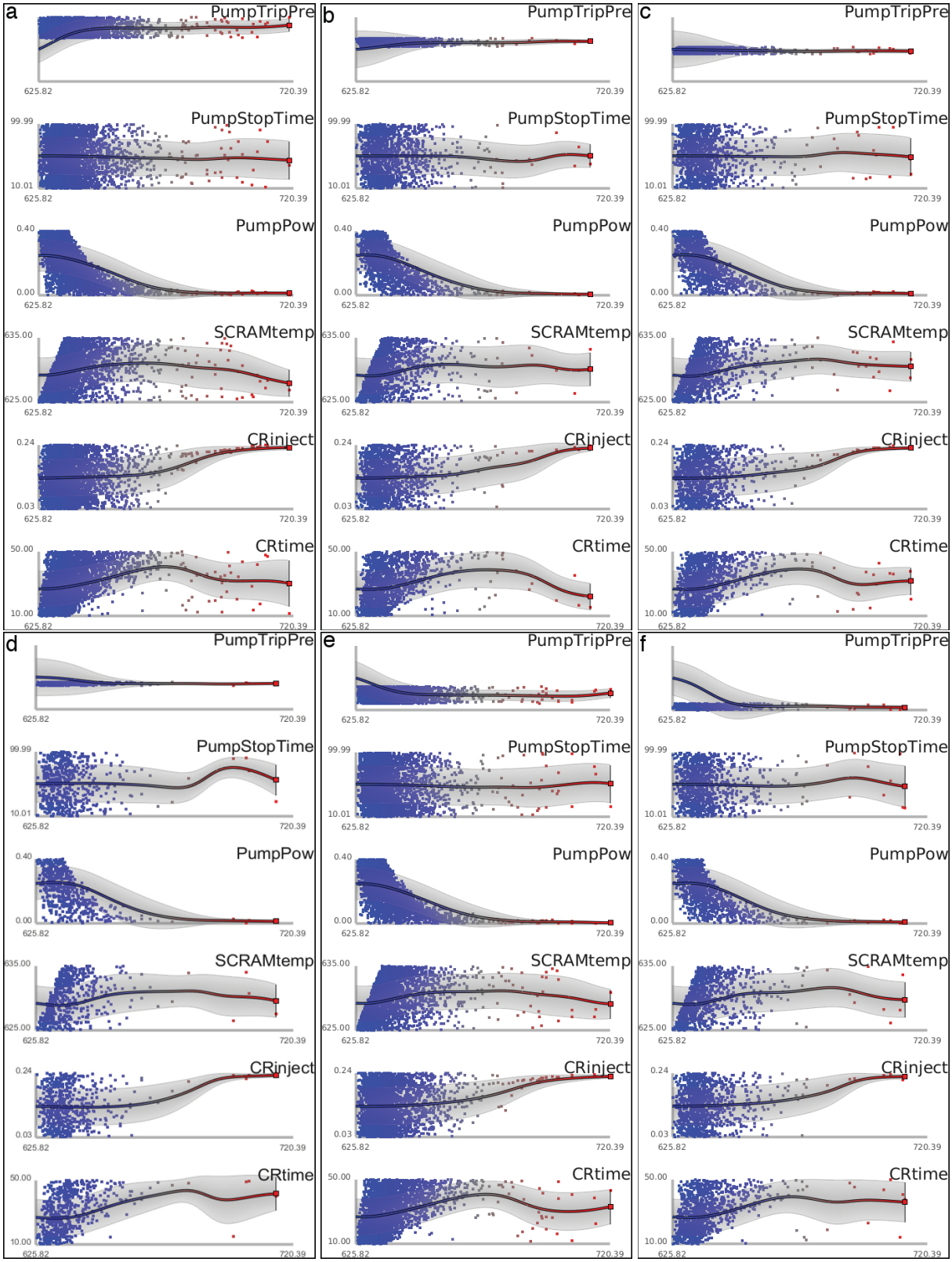

Figure 13: Inverse coordinate plots for the VR_2^+ 6D dataset. From left to right, from top to bottom, the inverse coordinate plots for crystals a through f, respectively.

4.2 21-Dimensional Nuclear Safety Simulation Data

We perform a similar analysis on a more complicated simulation involving 21 input parameters where the output variable is again peak coolant temperature. As stated previously, the dataset consists of 10,000 individual simula-

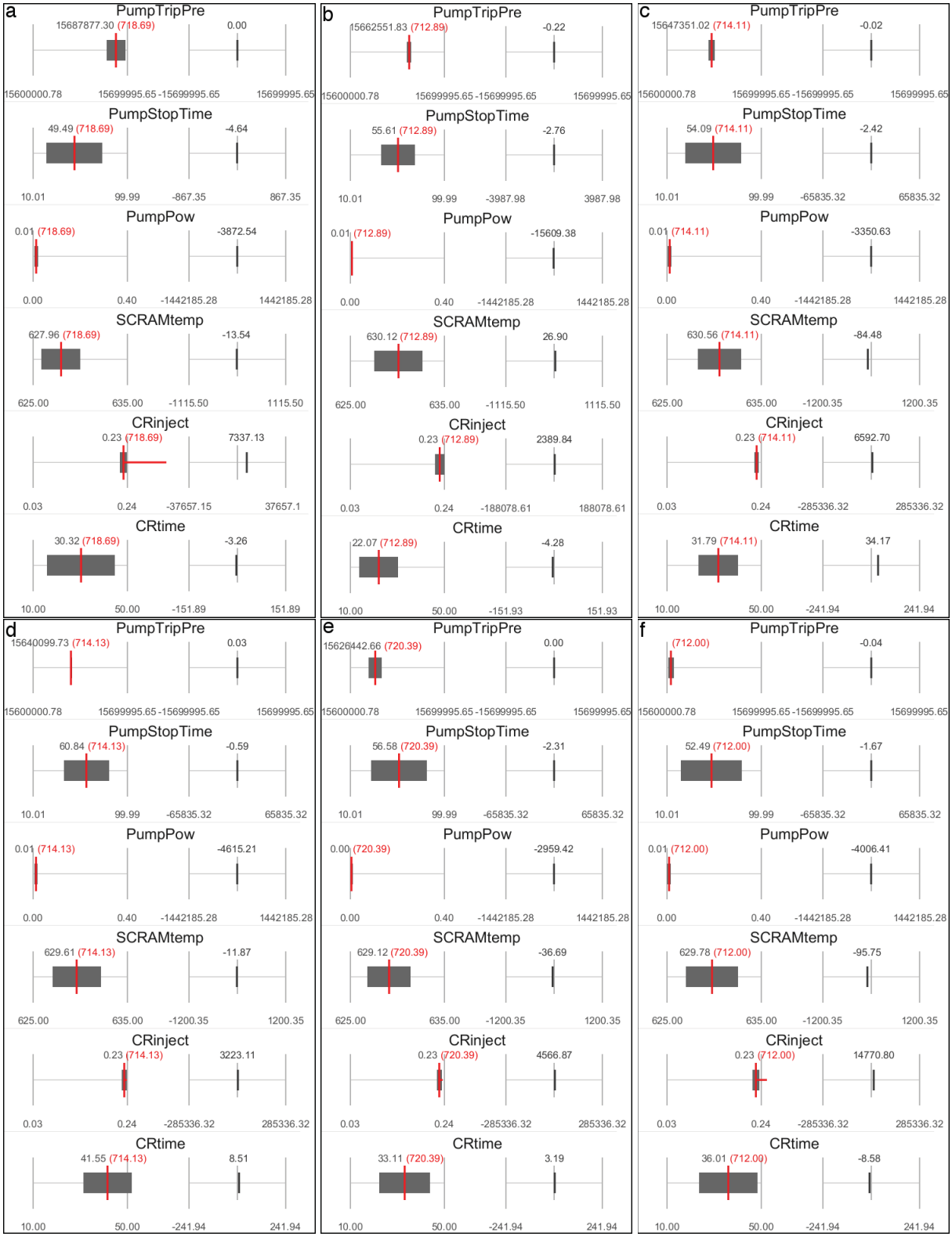

Figure 14: Statistical summary visual interfaces for the VR_2^+ 6D dataset. They are ordered from left to right, from top to bottom, for crystals a through f, respectively.

tions where the parameters include the parameters of the 6-dimension dataset in addition to 15 others. As shown in the persistence graph of the topological summary visual interface (Figure 17), the dataset at its finest scale con-

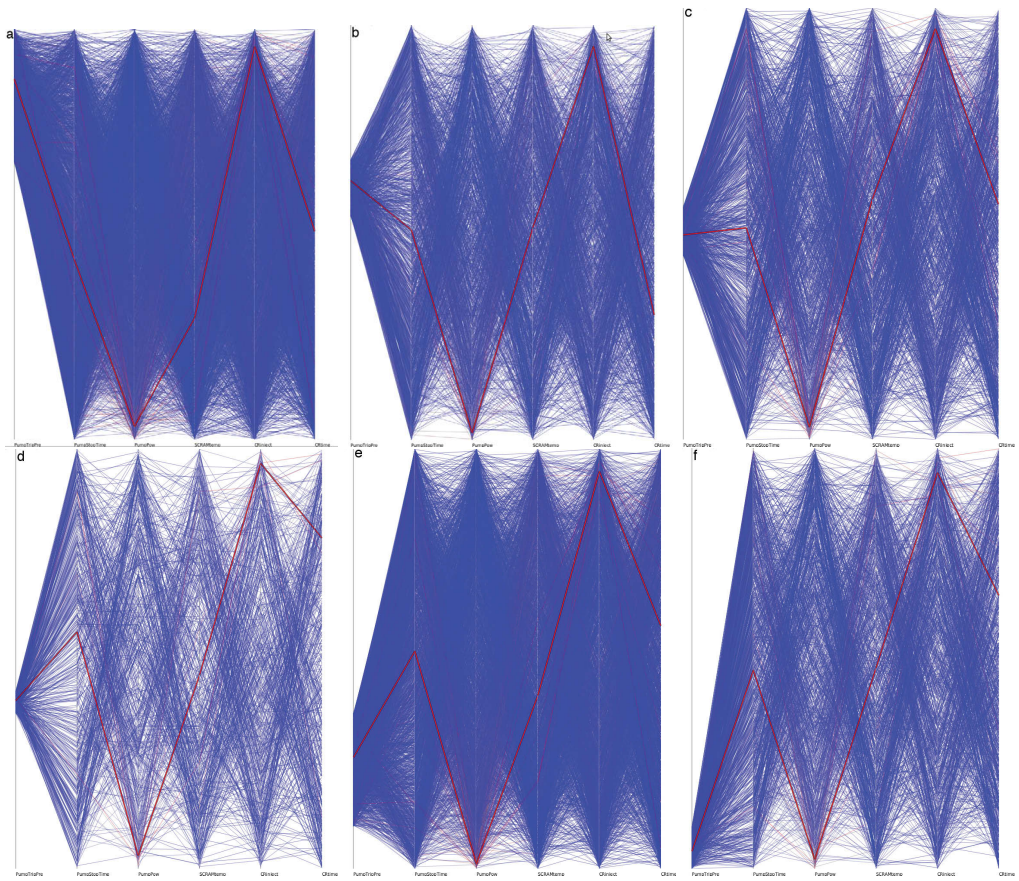

Figure 15: Parallel coordinate plots for the VR_2^+ 6D dataset. They are ordered from left to right, from top to bottom, for crystals a through f, respectively.

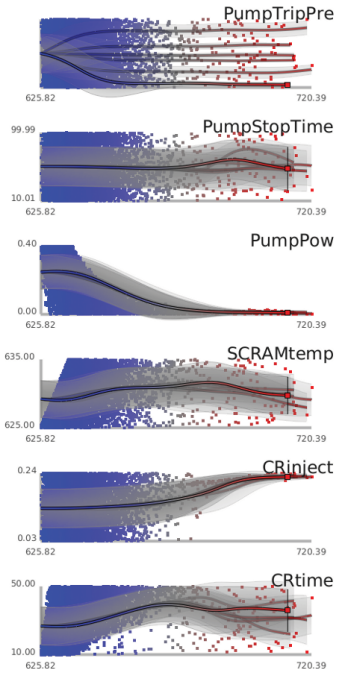

Figure 16: VR_2^+ 6D dataset: Inverse coordinate plots of all Morse-Smale crystals combined (into one single stable manifold).

sists of a large number of unstable features (732 Morse-Smale crystals in total). We then simplify such a complex structure by increasing the persistence threshold and arrive at a relatively stable structure with five Morse-Smale crystals.

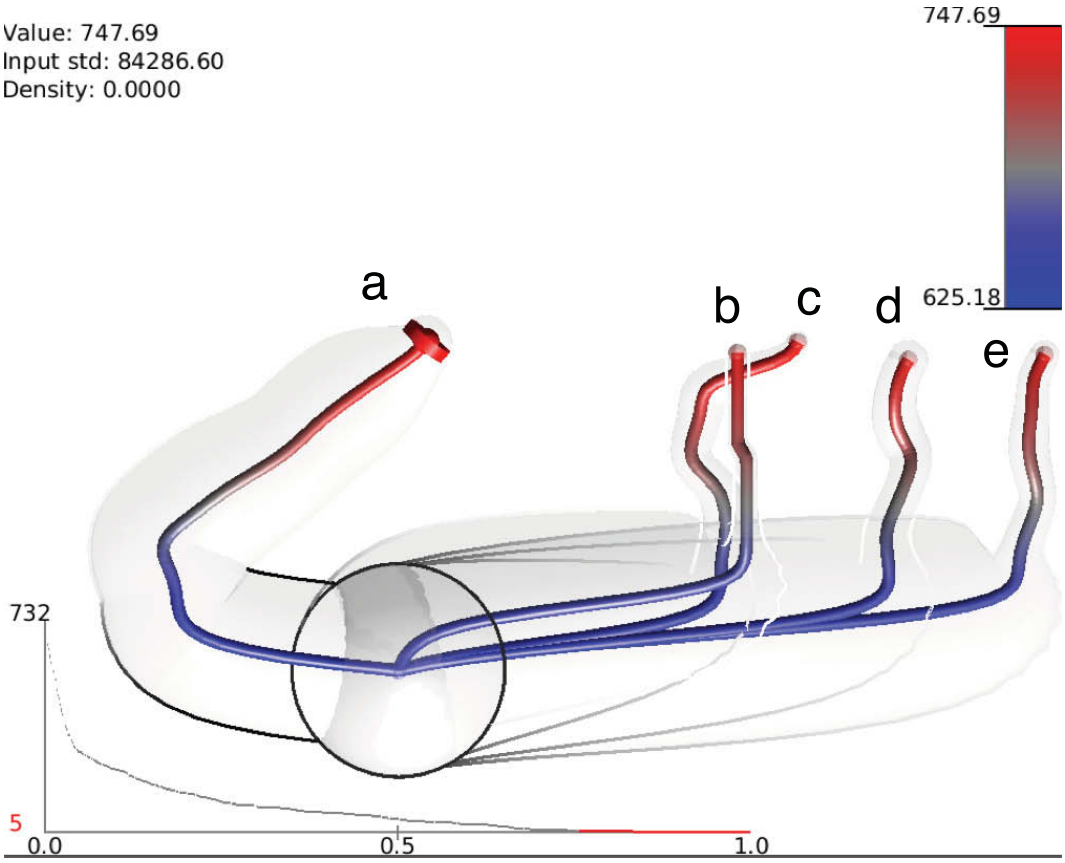

Figure 17: The topological summary interface for the VR_2^+ 21D dataset: at a certain scale, there are five Morse-Smale crystals sharing one global minimum.

Based on its topological summary (Figure 17), it is clear that the overall simplified structure of the simulation space is similar to that of the 6D dataset. That is, all Morse-Smale crystals share a single global minimum and the data points that reside at or near such a minimum constitute a large percentage of the total sampled points. The tubes near each maxima are relatively narrow, representing a small number of input parameter settings that result in the production of high PCT values.

Selecting a specific crystal in the topological summary shows us where in the range space the crystal is most heavily sampled, but this interface tells us little about how much data is associated with an entire crystal. On the other hand, the parallel coordinate plots (Figure 18) readily demonstrate that the crystal b contains much fewer data points than the other crystals. Furthermore, we note that each crystal targets separate ranges of values for the Qcore parameter (12-th axe from the left) which is not clearly evident in the other visual interfaces.

In addition, again we could also combine the inverse coordinate plots of all Morse-Smale crystals together into one stable manifold and visualize the combined plot in Figure 19. The visual interface shows that PumpPow and CRinject remain the dominating factors in terms of reaching a high PCT, that is, only a high CRinject value and a low PumpPow value will yield maximum PCTs for all crystals.

Finally, we can also construct the pairwise scatter plots (Figure 20). As discussed before, as the dimension of the dataset increases, such plots eventually become a dense visual cluster, causing the interface to be ineffective. We are investigating other ways to present these pairwise relations among input parameters, such as focused or sub-sampled views.

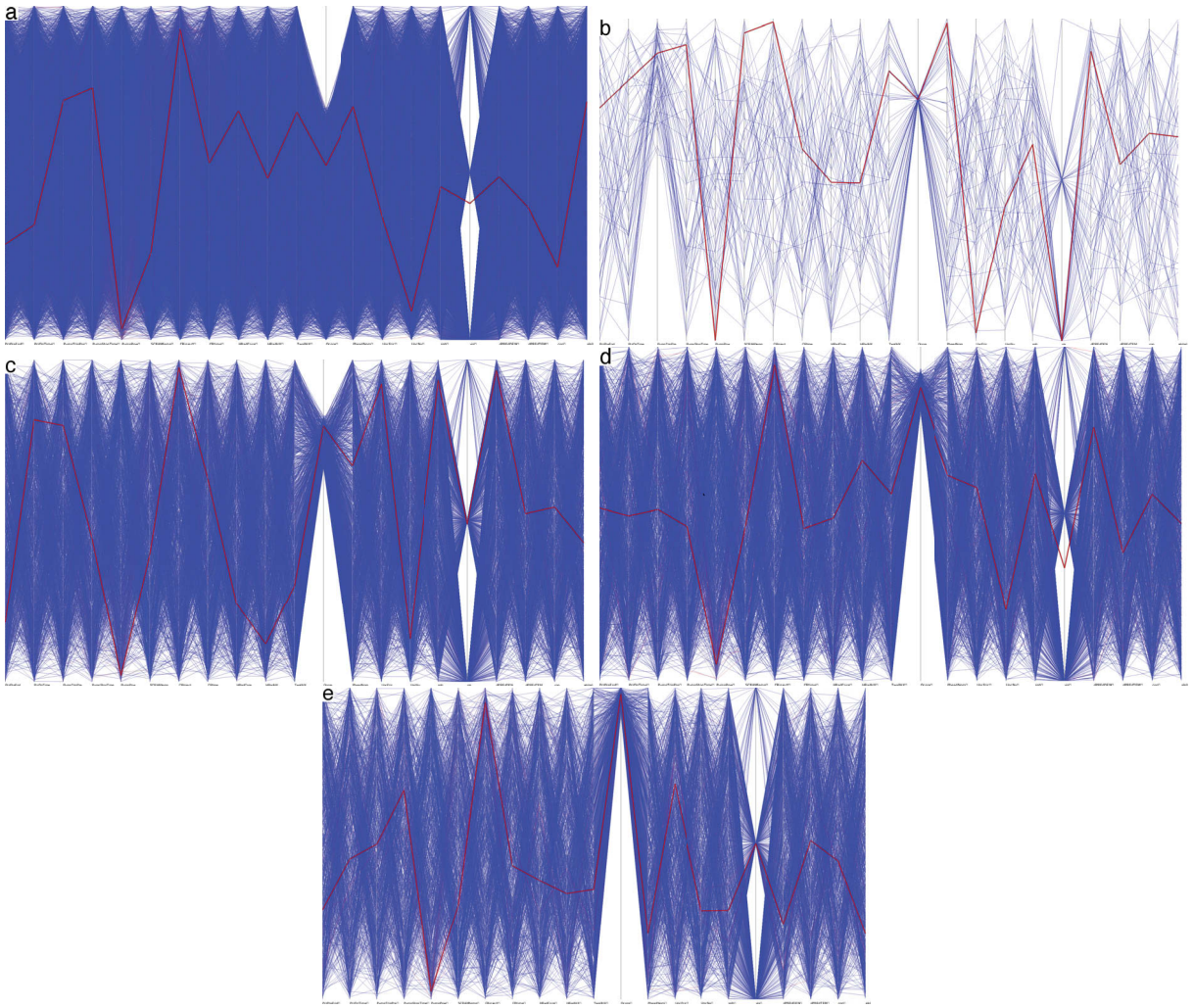

Figure 18: The parallel coordinate plots of the VR_2^+ 21D dataset. They are ordered from left to right, from top to bottom, for crystals a through e.

5 Conclusions

The main purpose of this paper is to provide a user's guide to TopoXG, a software package that is designed for the analysis and visualization of high-dimensional scalar functions. We describe each analysis and visualization module in the system, by explaining the theories and techniques based on a simple 4-dimensional demo example for the analysis of recovery from an aircraft crash into an RVACS of a conceptual design for a sodium-cooled fast reactor . We then present the specific applications of the software in two datasets, a 6-dimensional and a 21-dimensional dataset, for nuclear reactor safety analysis and visualization. We have obtained some intuitive understanding of the structures of these high-dimensional functions, although a more comprehensive interpretation of our results depends on extensive testing and explorations from our collaborators. We expect to work closely with the domain scientists and obtain feedbacks from the end users regarding (a) the interpretation of the testing datasets using our software, (b) the limitations and potential improvements of current analysis techniques, and (c) the potential improvements of the visualization interfaces in terms of usability and interactivity.

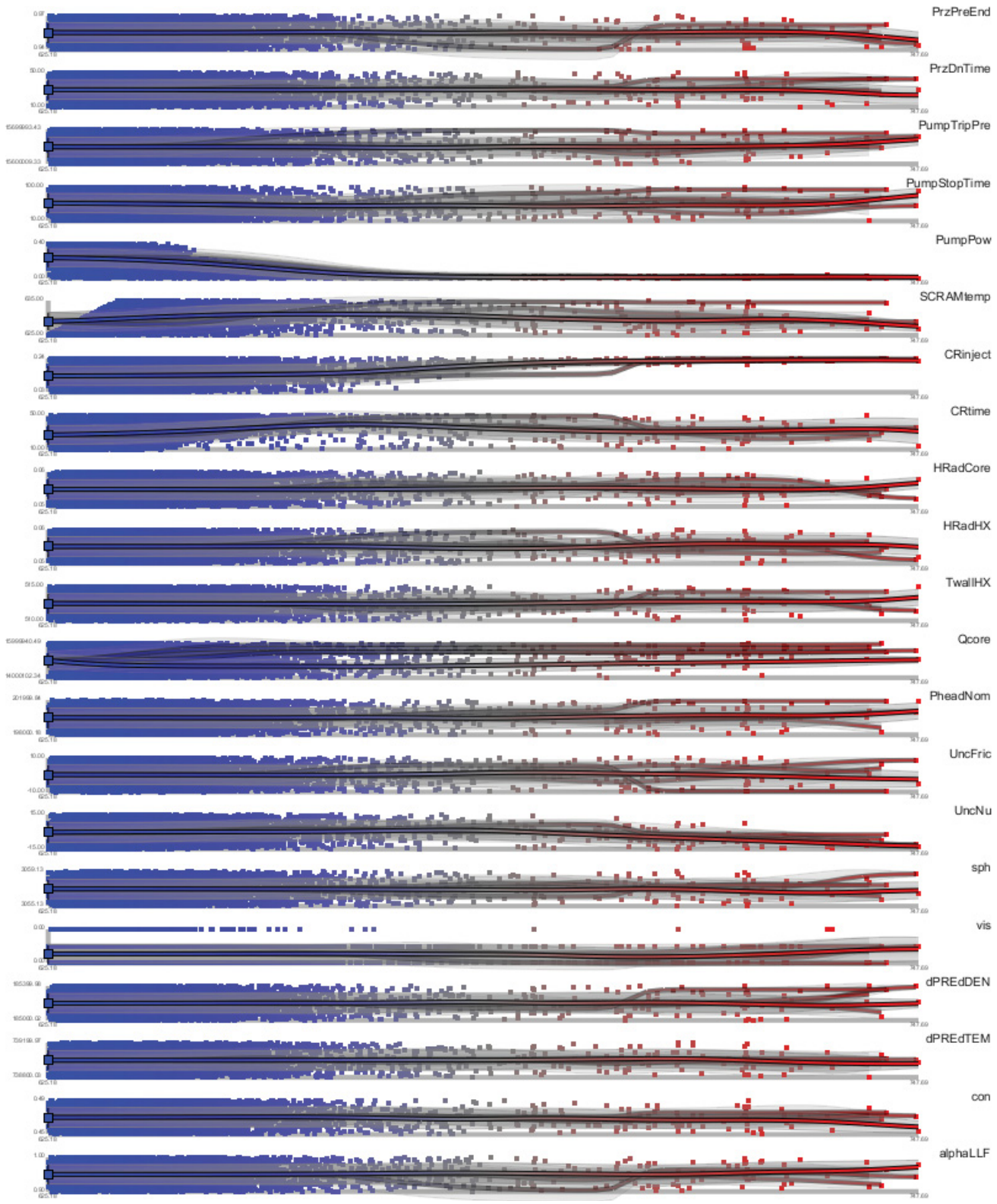

Figure 19: VR_2^+ 21D dataset: Inverse coordinates plots for all Morse-Smale crystals combined (into one stable manifold).

References

[1] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 28(1):125–136, 1972.

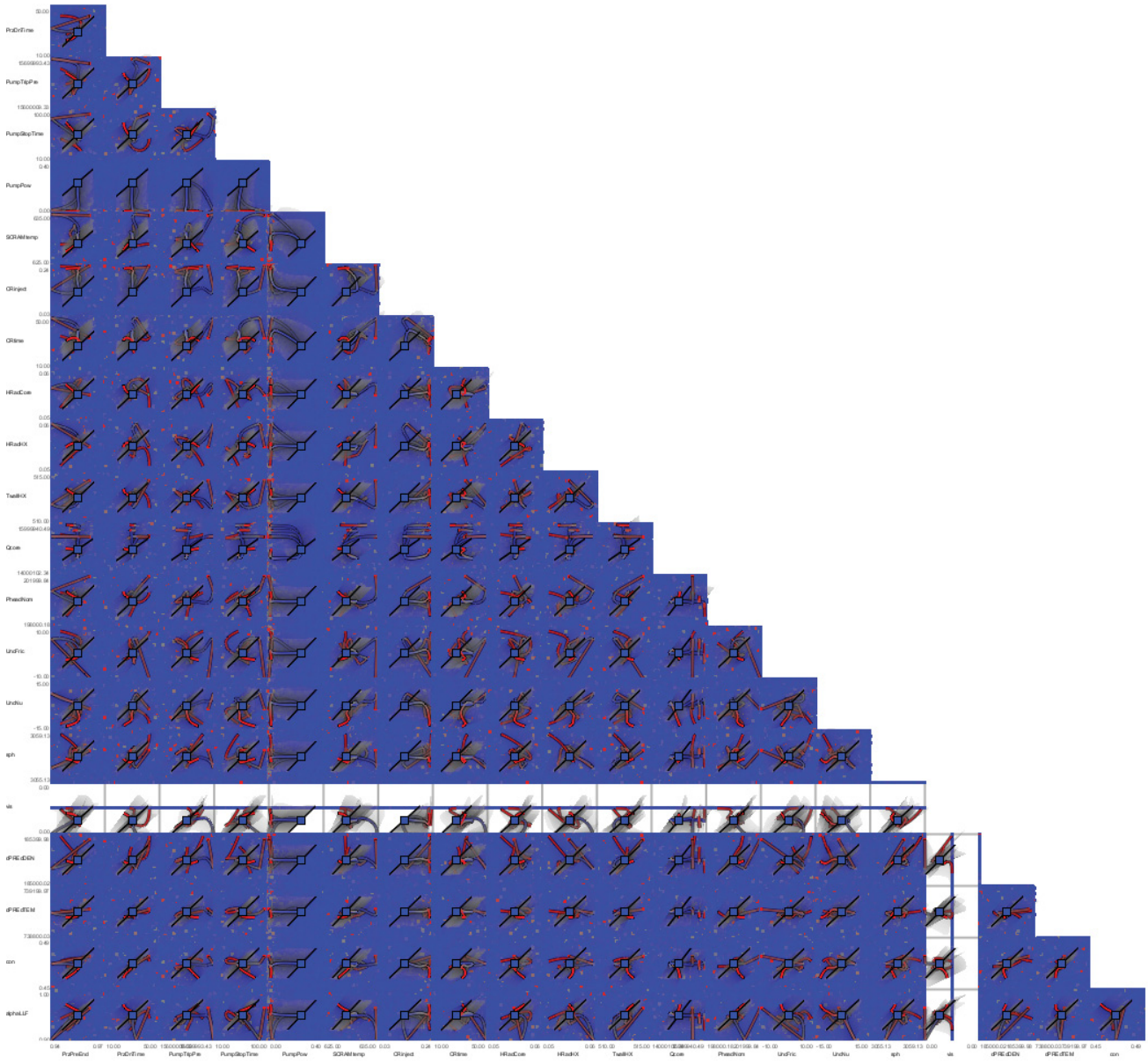

Figure 20: The pairwise scatter plots of the VR_2^+ 21D dataset.

[2] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, 45:891–923, 1998.

[3] D. Asimov. The grand tour: a tool for viewing multidimensional data. *SIAM Journal on Scientific and Statistical Computing*, 6:128–143, 1985.

[4] C. L. Bajaj, V. Pascucci, G. Rabbiolo, and D. Schikorc. Hypervolume visualization: a challenge in simplicity. *Proceedings IEEE Symposium on Volume Sisualization*, pages 95–102, 1998.

[5] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2003.

[6] R. L. Boyell and H. Ruston. Hybrid techniques for real-time radar simulation. In *Proceedings Fall Joint Computer Conference*, pages 445–458, 1963.

[7] P.-T. Bremer, H. Edelsbrunner, B. Hamann, and V. Pascucci. A topological hierarchy for functions on triangulated surfaces. *IEEE Transactions of Visualization and Computer Graphics*, 10:385–396, 2004.

[8] J. Cardinal, S. Collette, and S. Langerman. Empty region graphs. *Computational Geometry: Theory and Applications*, 42:183–195, 2009.

[9] E. Carlsson, G. Carlsson, and V. de Silva. An algebraic topological method for feature identification. *International Journal of Computational Geometry and Applications*, 16:291–314, 2003.

[10] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *International Journal of Computer Vision*, 76:1–12, 2008.

[11] G. Carlsson, A. J. Zomorodian, A. Collins, and L. J. Guibas. Persistence barcodes for shapes. In *Proceedings Eurographs/ACM SIGGRAPH Symposium on Geometry Processing*, pages 124–135, 2004.

[12] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Proceedings 11th Annual ACM-SIAM Symposium on Discrete algorithms*, pages 918–926, 2000.

[13] H. Carr, J. Snoeyink, and M. van de Panne. Simplifying flexible isosurfaces using local geometric measures. In *Proceedings 15th IEEE Visualization*, pages 497–504, 2004.

[14] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings 25th Annual Symposium on Computational Geometry*, pages 237–246, 2009.

[15] W. S. Cleveland. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician*, 35:54, 1981.

[16] K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Loops in reeb graphs of 2-manifolds. In *Proceedings 19th Annual Symposium on Computational Geometry*, pages 344–350, 2003.

[17] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic and Geometric Topology*, 7:339–358, 2007.

[18] H. Edelsbrunner and J. Harer. Persistent homology - a survey. *Contemporary Mathematics*, 453:257–282, 2008.

[19] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical morse complexes for piecewise linear 2-manifolds. In *Proceedings 17th Annual Symposium on Computational Geometry*, pages 70–79, 2001.

[20] H. Edelsbrunner, J. Harer, and A. J. Zomorodian. Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Discrete and Computational Geometry*, 30(87-107), 2003.

[21] H. Edelsbrunner, D. Letscher, and A. J. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

[22] J. Friedman and J. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23:881 – 890, 1974.

[23] S. Gerber, P.-T. Bremer, V. Pascucci, and R. Whitaker. Visual exploration of high dimensional scalar functions. *IEEE Transactions on Visualization and Computer Graphics*, 16:1271–1280, 2010.

[24] A. Gyulassy and V. Natarajan. Topology-based simplification for feature extraction from 3d scalar fields. In *Proceedings IEEE Visualization*, pages 535–542, 2005.

[25] 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*, 13:1440–1447, 2007.

[26] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[27] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313:504 – 507, 2006.

[28] F. Höppner and F. Klawonn. Visualising clusters in high-dimensional data sets by intersecting spheres. In *Proceedings International Symposium on Evolving Fuzzy Systems*, pages 106–111, 2006.

[29] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, August 1985.

[30] A. Inselberg. *Parallel Coordinates: VISUAL Multidimensional Geometry and its Applications.* Springer, 2009.

[31] T. Itoh and K. Koyamada. Automatic isosurface propagation using an extrema graph and sorted boundary cell lists. *IEEE Transactions on Visualization and Computer Graphics*, 1:319–327, 1995.

[32] J. X. Li. Visualization of high-dimensional data with relational perspective map. *Information Visualization*, 3(1):49–59, 2004.

[33] 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.

[34] J. Marks, B. Andalman, P. A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber. Design galleries: a general approach to setting parameters for computer graphics and animation. In *Proceedings 24th Annual Conference on Computer Graphics and Interactive Techniques*, pages 389–400, 1997.

[35] J. Milnor. *Morse Theory*. Princeton University Press, New Jersey, NY, USA, 1963.

[36] M. Morse. Relations between the critical points of a real function of n independent variables. *Transactions of the American Mathematical Society*, 27:345–396, 1925.

[37] J. R. Munkres. *Elements of algebraic topology*. Addison-Wesley, Redwood City, CA, USA, 1984.

[38] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: Simplicity and speed. *ACM Transactions on Graphics*, 26:58.1–58.9, 2007.

[39] G. Reeb. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Comptes Rendus de lÁcadémie des Sciences*, 222:847–849, 1946.

[40] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[41] R. Sharan, A. Maron-Katz, and R. Shamir. Click and expander: a system for clustering and visualizing gene expression data. *Bioinformatics*, 19(14):1787—1799, 2003.

[42] S. Takahashi and Y. Takeshima. Topological volume skeletonization using adaptive tetrahedralization. In *Proceedings Geometric Modeling and Processing*, pages 227–236, 2004.

[43] J. B. Tenenbaum, V. D. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[44] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *Proceedings European Conference on Computer Vision*, pages 705–718, 2008.

[45] J. A. Walter and H. Ritter. On interactive visualization of high-dimensional data using the hyperbolic plane. In *Proceedings 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 123–132, 2002.

[46] 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.

[47] 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*, 100:461–462, 2009.

[48] C.-C. Yang, C.-C. Chiang, Y.-P. Hung, and G. C. Lee. Visualization for high-dimensional data: VisHD. *Proceedings 9th International Conference on Information Visualization*, pages 692–696, 2005.