International Conference on Mathematics and Computational Methods Applied to Nuclear Science & Engineering (M&C 2013)
Sun Valley, Idaho, USA, May 5-9, 2013, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2013)

Exploration of High-Dimensional Scalar Function for Nuclear Reactor Safety Analysis and Visualization

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

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

Michael Pernice, Diego Mandelli, Robert Nourgaliev
Idaho National Laboratory
{michael.pernice,diego.mandelli,robert.nourgaliev}@inl.gov

ABSTRACT

The next generation of methodologies for nuclear reactor Probabilistic Risk Assessment (PRA) explicitly accounts for the time element in modeling the probabilistic system evolution and uses numerical simulation tools to account for possible dependencies between failure events. The Monte-Carlo (MC) and the Dynamic Event Tree (DET) approaches belong to this new class of dynamic PRA methodologies. A challenge of dynamic PRA algorithms is the large amount of data they produce which may be difficult to visualize and analyze in order to extract useful information. We present a software tool that is designed to address these goals. We model a large-scale nuclear simulation dataset as a high-dimensional scalar function defined over a discrete sample of the domain. First, we provide structural analysis of such 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. Our analysis is performed by 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 our software tool by highlighting its analysis and visualization capabilities, along with a use case involving dataset from a nuclear reactor safety simulation.

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

1 INTRODUCTION

Dynamic Probabilistic Risk Assessment (PRA) methodologies [19] couple numerical simulation tools and time-dependent stochastic models (i.e., probabilistic failure models or parameter uncertainties), to perform system safety analysis. Widely used dynamic PRA methodologies are based on Monte-Carlo [36] or Dynamic Event Tree algorithms [25]. The common underlying idea is to run a large number of simulations (by employing system simulators) where values of system stochastic parameters (e.g., timing of failure of a specific component or an uncertain parameter) are sampled from their own distribution at each run. This type of PRA analysis can be very time-consuming when a large number of stochastic parameters are considered and when large and complex system simulators are used. Moreover, a large volume of data is typically generated. Such large amounts of information can be difficult to organize for extracting useful

information. Furthermore, it is often not sufficient to merely calculate a quantitative value for the risk and its associated uncertainties. The development of risk insights that can increase system safety and improve system performance requires the interpretation of scenario evolutions and the principal characteristics of the events that contribute to the risk.

The need for software tools able to both analyze and visualize large amount of data generated by Dynamic PRA methodologies has been emerging only in recent years. A first step has been shown in [38, 39] using clustering-based algorithms which focus more on the analysis part than the visualization side.

In this paper, we present a software tool that provides scientists and domain experts with an interactive analysis and visualization environment for understanding the structures of high-dimensional nuclear simulation data. Our tool adapts and extends the innovative framework called HDViz first proposed by Gerber et. al. [26] in exploring high-dimensional scalar functions, and applies the underlying techniques to nuclear reactor safety analysis and visualization. Our tool includes a host of various analysis and visualization capabilities. We describe each of these capabilities on a modular basis, by explaining the underpinning theories and presenting usage cases. The software 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 segment have similar geometric and topological properties, and from these we can create compact statistical summaries of each segment. Such summaries are then presented to the user in an intuitive manner that highlights features of the dataset which are otherwise hidden in a global view of the data. 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 of the Morse-Smale complex, its approximation in high dimensions, persistence simplification, and visualizing high-dimensional data. We describe each analysis and visualization module in Section 3 applied to a nuclear reactor safety analysis data set.

2 PRELIMINARIES

Morse-Smale Complex. Several topological structures have been used in practice for analyzing the structure of a scalar function, such as Reeb graphs [44, 45], countour trees [6, 12, 34], and Morse-Smale complexes [7, 21, 27, 28]. 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 [7, 21, 28, 34, 44], and recent developments [12, 26] have enabled us to compute or approximate these structures in high dimensions.

The body of work presented in this paper relies heavily on the structure known as the Morse-Smale complex. The Morse-Smale complex is based on Morse theory [41, 42]. 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 [21]. 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.

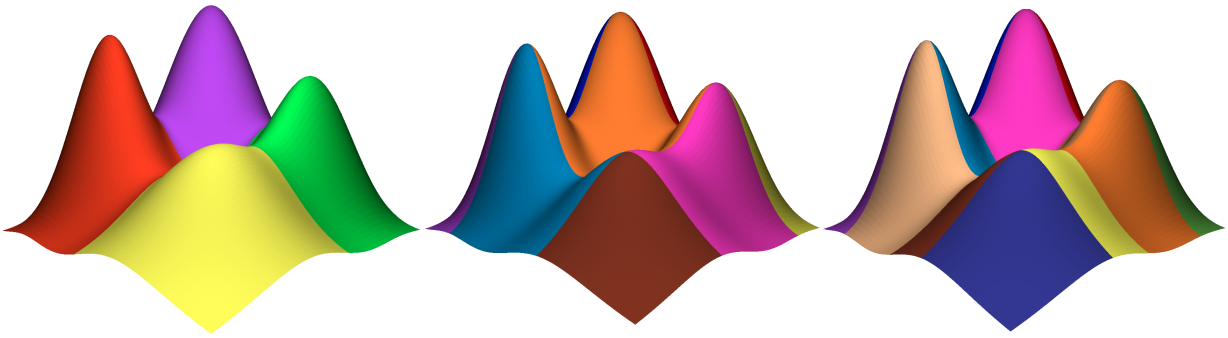

Figure 1. From left to right: unstable manifolds, stable manifolds, Morse-Smale complex. Each partition is colored by its corresponding gradient behavior, where points share the same color if: (Left) their ascending gradient flow end at the same local maximum; (Middle) their descending gradient flow 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 [50]. 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 [26]. Some research is underway [17, 37] 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 also 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 [29, 43]. 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 [11, 23], 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 [18], surface description and reconstruction [9], protein shapes [20] and images analysis [10]. In visualization, it has been used to simplify Morse-Smale complexes [7, 22, 27], Reeb graphs [16, 45] and contour trees [13, 48].

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 (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.

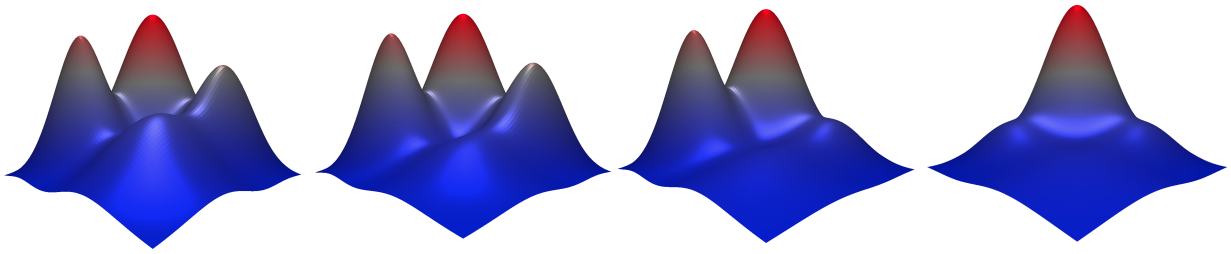

Figure 2. Progressive simplification of the critical points of a 2D function.

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 [24]. Slight variations of these methods look at sequences of projections onto different directions or multiple subspaces, such as Andrew's plots [1], parallel coordinates [32] and the grand tour approach [3]. The high-dimensional data is embedded onto the hyperbolic plane in [51], and 3D-sphere in [31]. Recently, specific tools have been developed to facilitate interactive visualization of high-dimensional data in conjunction with nonlinear dimensionality reduction techniques, such as VisuMap [35], VisHD [52] and Hyperbolic MDS [51]. Tools targeted towards specific applications in high-dimensional data visualization include design galleries [40] for exploring the parameter space of transfer functions in volume rendering, and Click and Expander [47] 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, 30, 46, 49]. Our work uses parallel coordinate plots, and adapts several high-dimensional projection schemes, including PCA, ISOMAP [49] 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

In this section, we describe each analysis and visualization module within our integrated system that are either part of the original capabilities provided by HDViz [26] or part of our extension. The modules included in the software and described below are: topological summary, statistical summary, parallel coordinate plot, pairwise scatter plots, inverse coordinate plots and interactive projection. We demonstrate our infrastructure with an example dataset from nuclear plan safety analysis.

3.1 Example Dataset

We use a 6-dimensional dataset as an example for demonstration purposes. The data is extracted from a VR_2^+ nuclear reactor simulator and represents an ensemble of 10000 simulation trials where a SCRAM is simulated due to a failure in the system. A SCRAM event is when the control rods of the reactor are inserted into the core in order to prevent overheating of the reactor core. The output variable is the peak coolant temperature (PCT), measured in Kelvin. 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). The input space is defined by six parameters:

- **PumpTripPre** - the minimum pressure (MPa) in the heat exchange pump causing the SCRAM to trip
- **PumpStopTime** - the relaxation time (sec) of pump's phase-out
- **PumpPow** - end power of the pump
- **SCRAMtemp** - the maximum temperature in the system causing the SCRAM to trip
- **CRinject** - the control rod position at the end of SCRAM
- **CRtime** - the relaxation time (sec) of the control rod system.

3.2 Topological Summary

The visual interface designed for topological summary is inherited and extended from the capabilities provided by HDViz [26]. Considered as the main visual display of our software, this interface summarizes each Morse-Smale crystal into a 1D curve in high-dimensional space which is then projected onto 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 points 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, $\mathbb{X}$ [26]: (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 segment $\mathbb{X}_i$ 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 [26].

Morse-Smale Approximation. We 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 just one of many approximation schemes. Some of our on-going work [37] focuses on understanding the possibilities for approximating these complexes and 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], see [26] for its detailed 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 [49] 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.

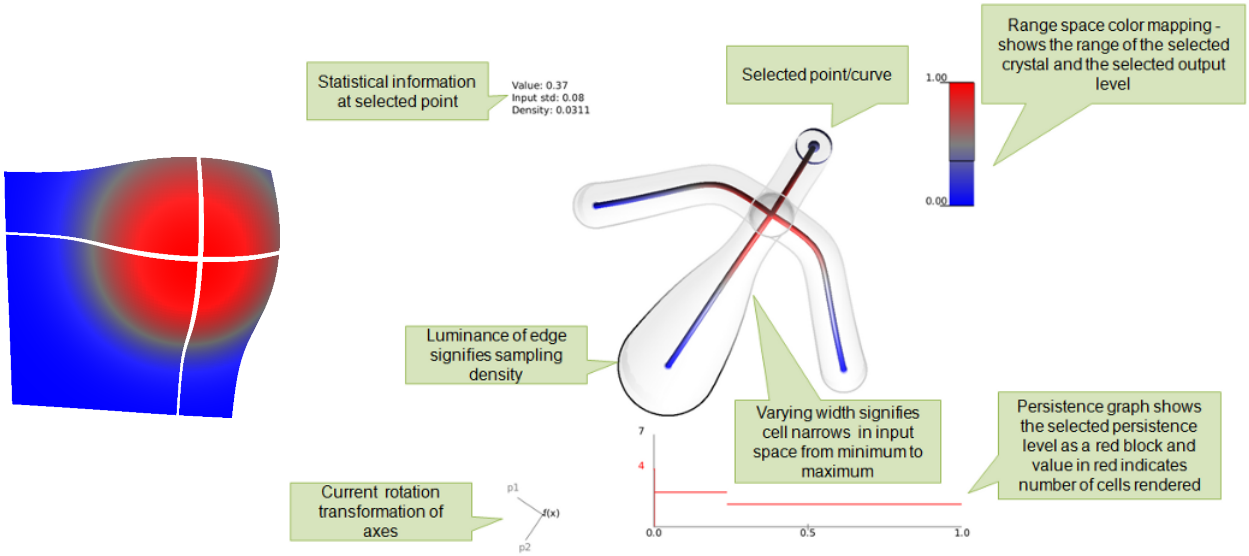

Figure 3. Left: a simple 2D surface decomposed into its Morse-Smale cells. Right: The topological summary visual interface of the simple 2D function shown in the left image.

Visual Components of the Framework. The visual components of the above framework are shown in Figure 3 for a simple 2D function with four Morse-Smale crystals. 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 [23], referred to as the *persistence graph*. This is shown as a visual component at the bottom of Figure 3. 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.

Figure 4 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 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.

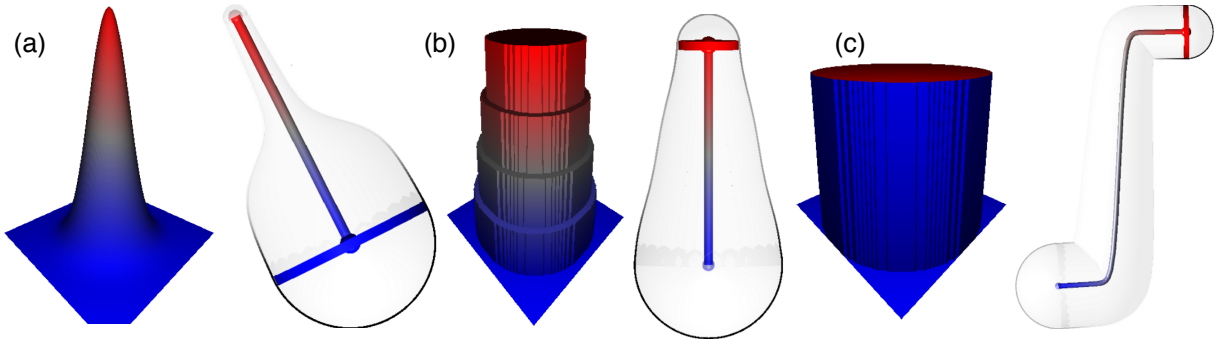

Figure 4. Three simple surfaces represented using geometric summary tubes.

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 5. 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.

As a final example, we illustrate our analysis and visualization interface with our 6D demo example in Figure 6 under multiple scales. This dataset (described in Section 3.1) contains 753 individual crystals at the finest level, though due to low persistence and visual clutter caused by these crystals, we restrict our analysis to only a handful of the most salient crystals. During the persistence simplification steps, the topological summaries consist of 6, 3 and 1 Morse-Smale crystal(s), respectively.

At each of the shown persistence level, the dataset 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,

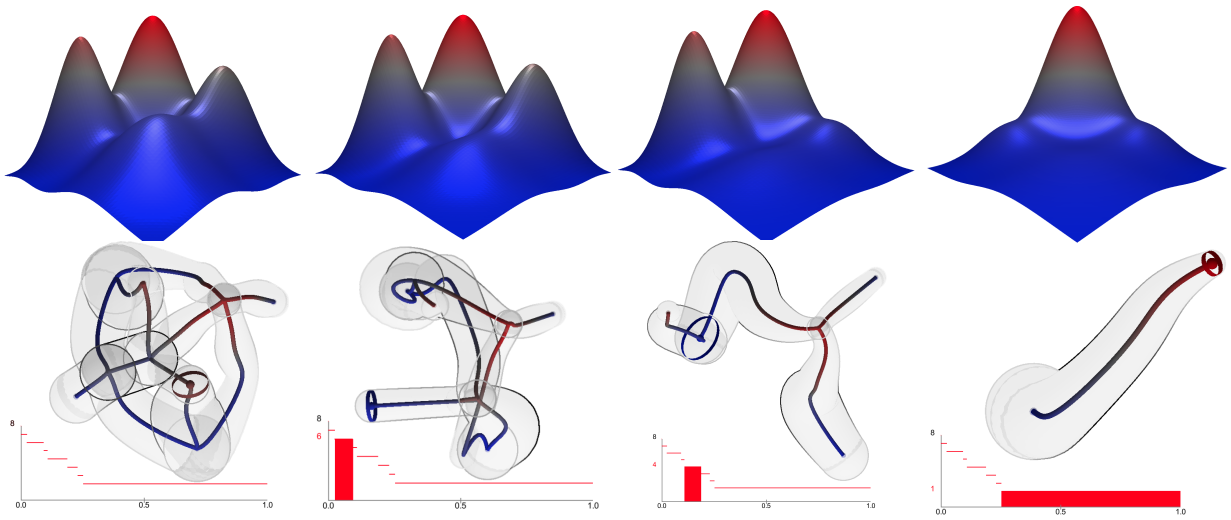

Figure 5. The topological summary of a simple dataset shown at various scales.

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.

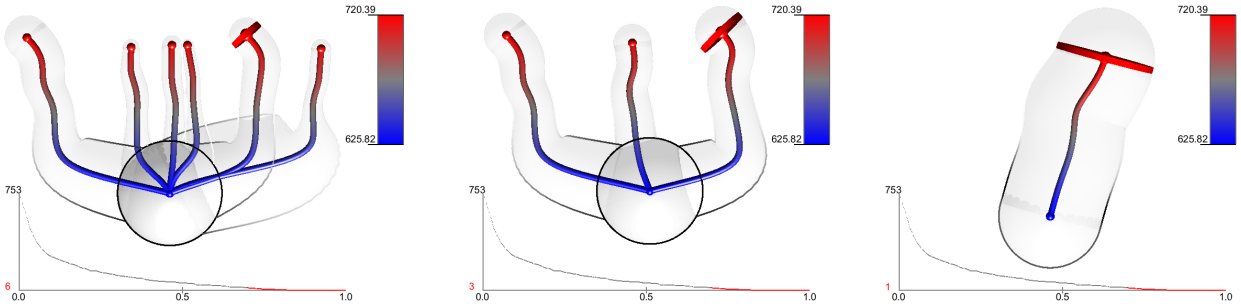

Figure 6. The topological summary of our demo dataset shown at various scales.

3.3 Statistical Summary

The visual interface shown in Figure 7 demonstrates statistical and geometric (i.e. gradient) information associated with the selected point in the topological summary interface. Each input dimension, or coordinate, 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. In Figure 7, we can see that three parameters have quite large standard deviations, PumpStopTime, SCRAMtemp, and CRtime, whereas the other parameters vary much less. The topological summary interface on the left of Figure 7 is a global summary of the data that presents the user with a transparent tube of moderate width that approximate such kind of detail. The statistical summary and the remaining display methods are integrated within the topological summary to provide deeper meanings to the analyst.

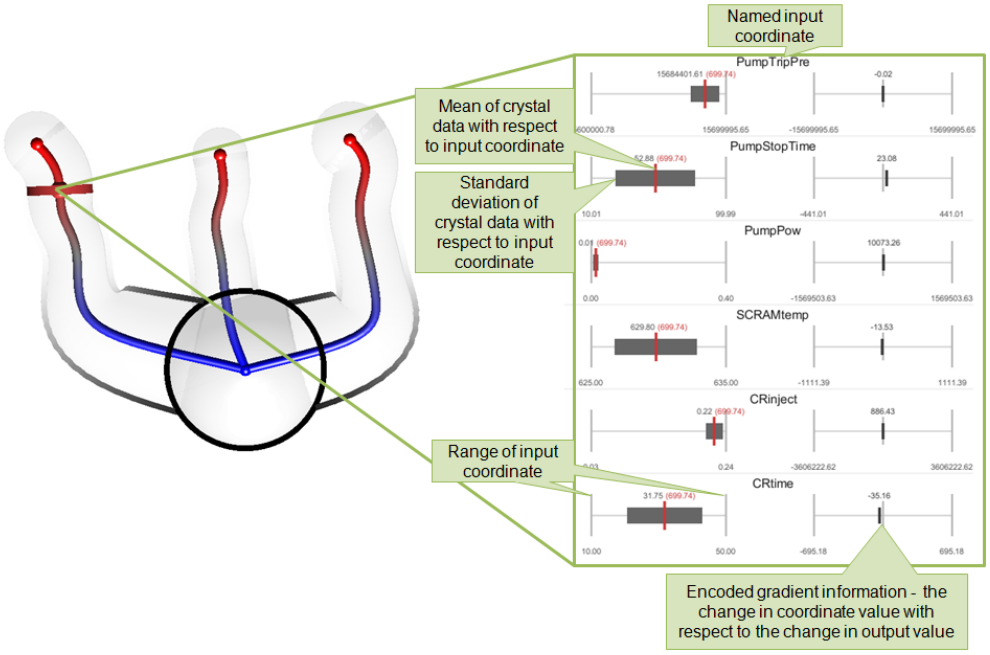

Figure 7. A snapshot of the statistical summary visual interface with highlighted visual components.

3.4 Parallel Coordinate Plots

We use parallel coordinate [33] plots to illustrate the correlations among input parameters in the high dimensional datasets. We showcase this visual interface using our example 6D function in the left image of Figure 8. 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 six dimensions (i.e. PumpTripPre, PumpPow, PumpStopTime, SCRAMtemp, CRinject, and CRtime) indicate correlations among various dimensions of the datasets. In addition, the mean value (from the inverse regression) of the selected crystal 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.

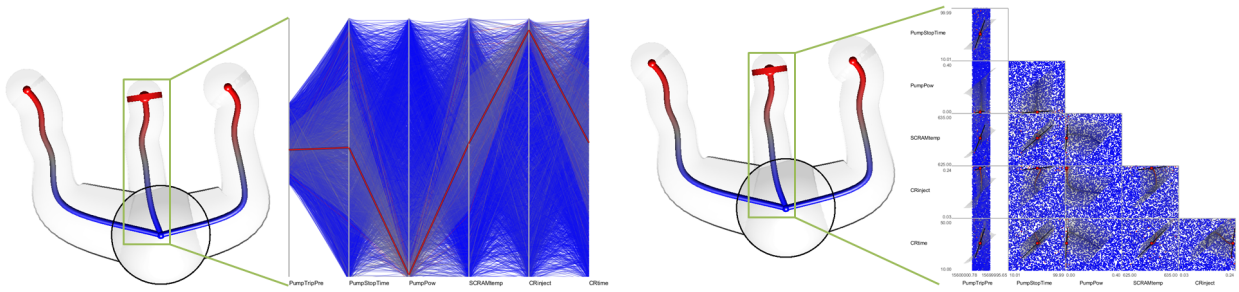

Figure 8. Left: Parallel coordinates plot. Right: Pairwise scatter plots.

3.5 Pairwise Scatter Plots

Similar to the parallel coordinate plots, we also demonstrate the 2D pairwise scatter plots on our 6D example (right image of Figure 8), for data points associated with each Morse-Smale crystal. Here, the user is presented with a series of 2D scatter plots that display each input parameter against every other input parameter, where the color of a point is mapped to the value of its 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.

3.6 Inverse Coordinate Plots

In the inverse coordinate plots, each input parameter is considered as a one-dimensional function of the output variable. This visual interface is shown in Figure 9 for our 6D example. On the left, at the persistence level with three crystals, the interface displays the inverse coordinate plot for data points associated with the selected Morse-Smale crystal in the topological summary interface. On the right, at the persistence level with six crystals, the interface shows combined inverse coordinate plots associated with all the crystals. For the selected crystal(s), the regression curve is drawn in the inverse coordinate plots with a grey tube representing the parameter-specific standard deviation.

From the right image in Figure 9, we see the top set of axes where each of the six 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 and it is also supported by the consistent locations of the mean values across all crystals and the low standard deviations of these parameters.

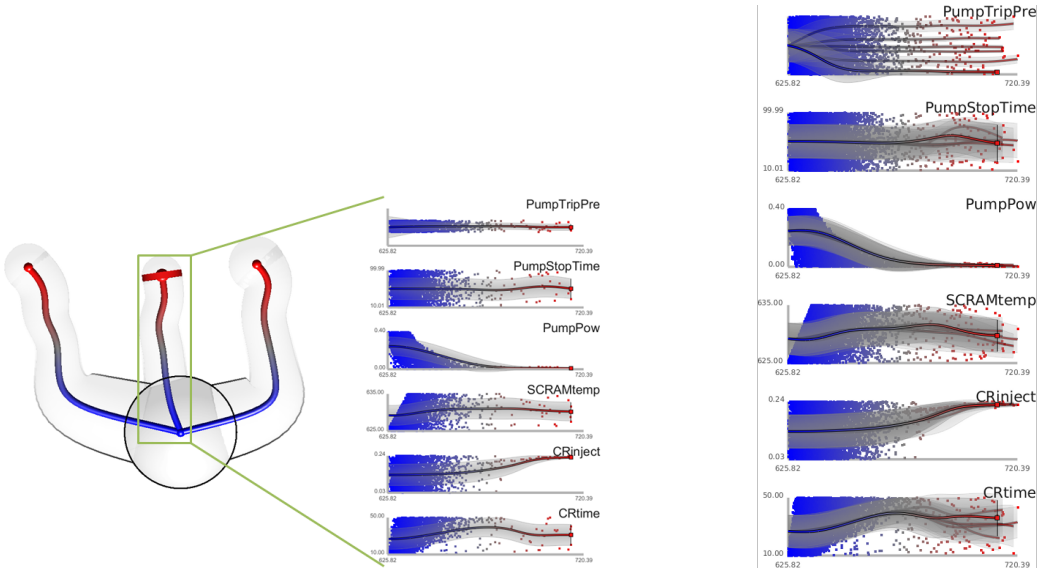

Figure 9. Left: Inverse coordinate plots shown for the highlighted crystal for the three crystal case. Right: Inverse coordinate plots shown for the six crystal case of the same 6D data.

3.7 Interactive Projection

The interactive projection visual interface (shown in Figure 10) 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 onto the screen space. We discuss the theory behind this interface and its implementation in the presented software.

As the number of dimensions increases for a high-dimensional function defined on the sampled point set, selecting the interesting projection dimensions while interacting with the high-dimensional function could become counter-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 [4].

To implement such a technique, 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, $\mathbf{v_i}$. In other words, 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].

4 CONCLUSIONS

In this paper, we present a software tool suitable to analyze and visualize datasets generated by safety analysis codes. In particular, our software could be interfaced with Dynamic PRA algorithms. In such a configuration, a large number of runs of the safety analysis code are performed, where during each run, initial conditions and timing/sequencing of events are changed accordingly to their statistical distribution. For large complex systems, this type of code generates a large amount of data, modeled as high-dimensional scalar functions. Our software is designed to assist the user in the process of extracting useful information from such functions. We describe each analysis and visualization module in the system, by explaining the theories and techniques based on a specific application of the software in a 6-dimensional

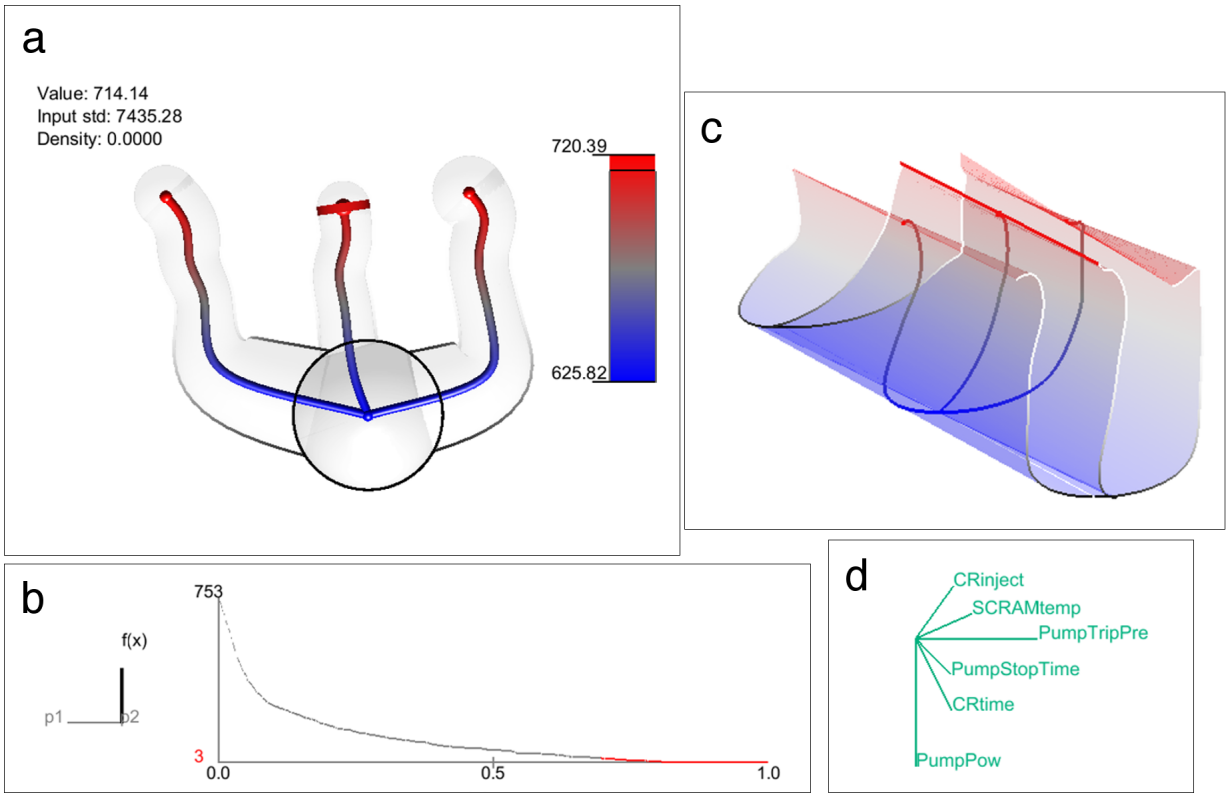

Figure 10. Interactive Projection Visual Interface for the 6D demo example. (a)-(b) represents the 2-step PCA topological summary detailed in section 3.2. The data is segmented into three crystals shown in (a) based on a chosen scale in (b). (c) shows the projection of the high-dimensional inverse regression 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 can manipulate by stretching and rotation, to emphasize or diminish the effect of a particular dimension on the projection.

demo example involving nuclear safety simulation. Such an example is based on code performed on 10000 simulations of a simplified PWR system for a SCRAM scenario. Our software allows us to identify and visualize the correlations among six parameters and the maximum coolant temperature. We first perform topological analysis to identify the underlying topological structure of the dataset. Statistical information is then summarized and linked to the topological structures extracted from the data, allowing the users to identify correlations between timing/sequencing of events and simulation outcome such as core temperature. We have obtained some intuitive understanding of the structures of such a high-dimensional function, 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.

REFERENCES

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

[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] C. D. Correa and P. Lindstrom. Towards robust topology of sparsely sampled data. *IEEE Transactions on Visualization and Computer Graphics*, 17:1852–1861, 2011.

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

[19] J. Devooght. Dynamic reliability. *Advances in Nuclear Science and Technology*, 25:215–278, 1997.

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

[21] 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.

[22] 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.

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

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

[25] E. M. G. Cojazzi, J. M. Izquierdo and M. S. Perea. The reliability and safety assessment of protection systems by the use of dynamic event trees: the dylam-treta package. In *In Proc. XVIII Annual Meeting Spanish Nucl. Soc.*, 1992.

[26] 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.

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

[28] 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.

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

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

[31] 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.

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

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

[34] 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.

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

[36] E. Z. M. Marseguerra, J. Devooght, and P. Labeau. A concept paper on dynamic reliability via monte carlo simulation. *Mathematics and Computers in Simulation*, 47:371–382, 1998.

[37] 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. TopoInVis, accepted, 2013.

[38] D. Mandelli, A. Yilmaz, and T. Aldemir. Data processing methodologies applied to dynamic PRA: an overview. In *Proceeding of Probabilistic Safety Analysis (PSA)*, 2011.

[39] D. Mandelli, A. Yilmaz, and T. Aldemir. Scenario analysis and pra: Overview and lessons learned. In *Proceedings of European Safety and Reliability Conference (ESREL 2011), Troyes (France)*, 2011.

[40] 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.

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

[42] 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.

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

[44] 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.

[45] 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.

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

[47] 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.

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

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

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

[51] 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.

[52] 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.