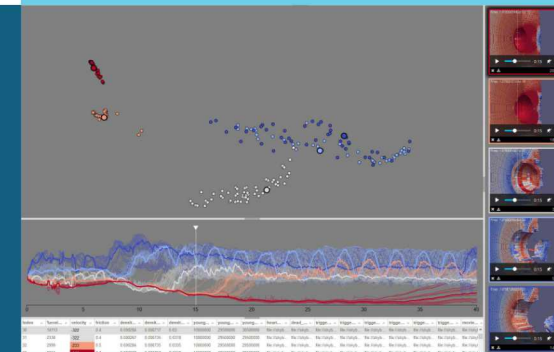
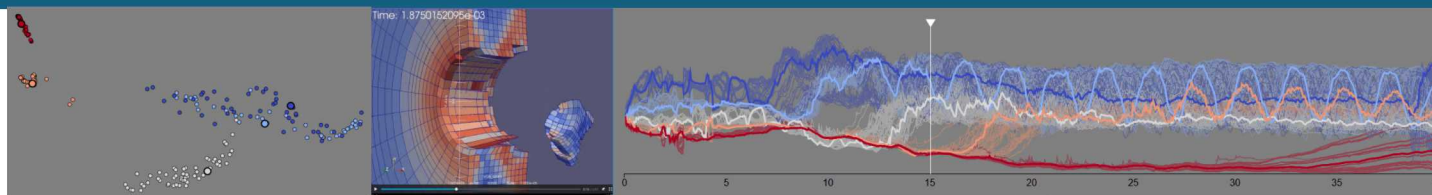


Slycat™ Overview and Update



PRESENTED BY

Patricia Crossno, Ph.D.

Collaborators: Jaxon Gittinger, Warren Hunt, Matthew Letter,
Shawn Martin, and Alex Sielicki

ALEGRA Winter Workshop, December 12, 2019

SAND 2019XXX-XX



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

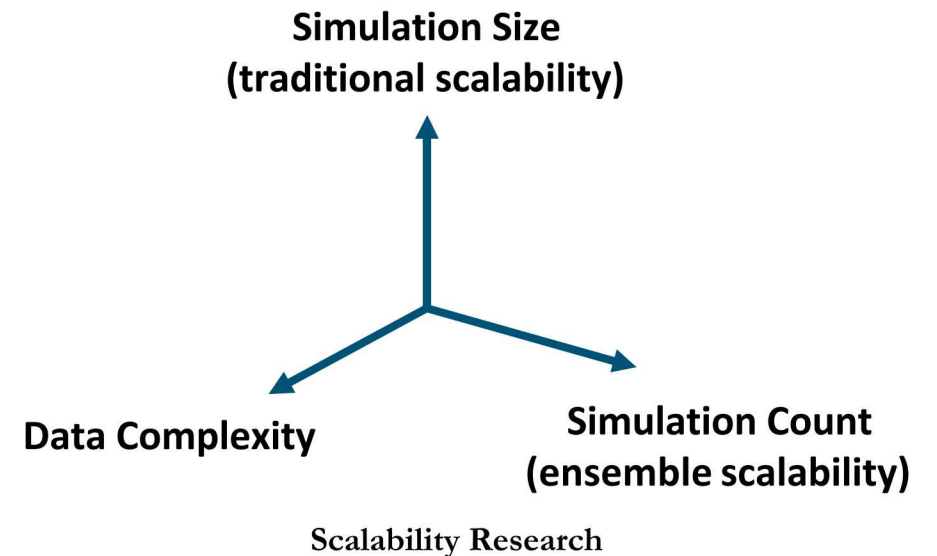
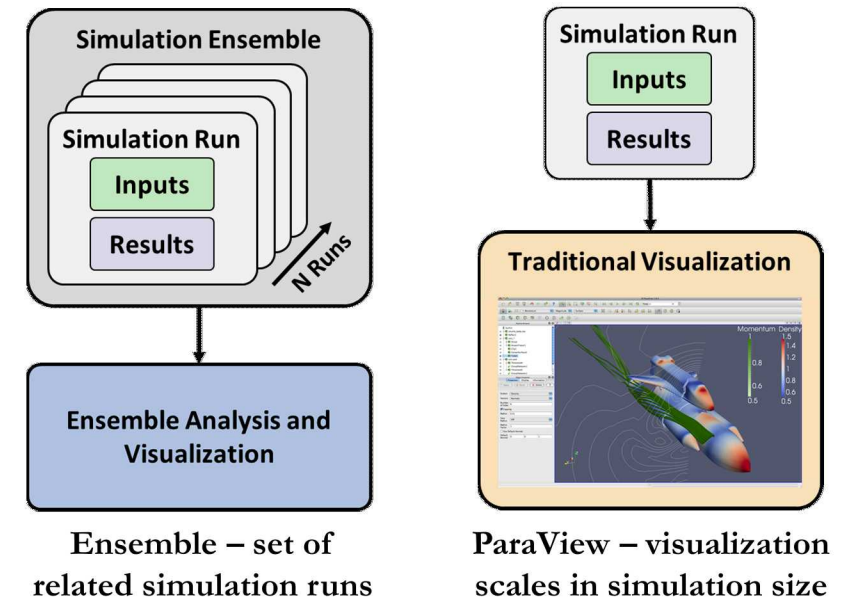
Some Important Questions Cannot Be Answered with Just One Simulation Run

Simulation ensembles are used for answering:

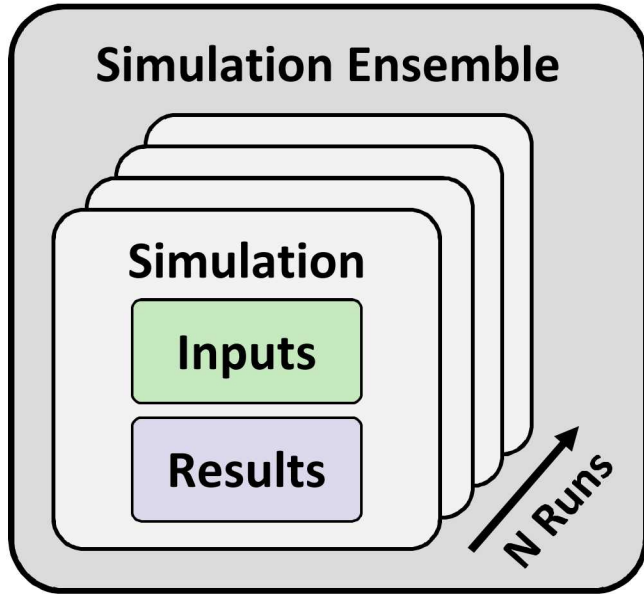
- What effect do small changes in input produce in the outputs? (Sensitivity Analysis)
- What range of output values are possible? (Model validation)
- How many different types of results are possible? (Categorization/Clustering)
- Which results are unique/anomalous? (Outliers)

Traditional tools not designed for ensembles

- Do not scale in number of runs
 - Designed to scale with problem size
 - Display limited to handful of results
- Not designed for ensemble results data
 - Designed for geometric data
 - Not designed for sets of multi-media outputs

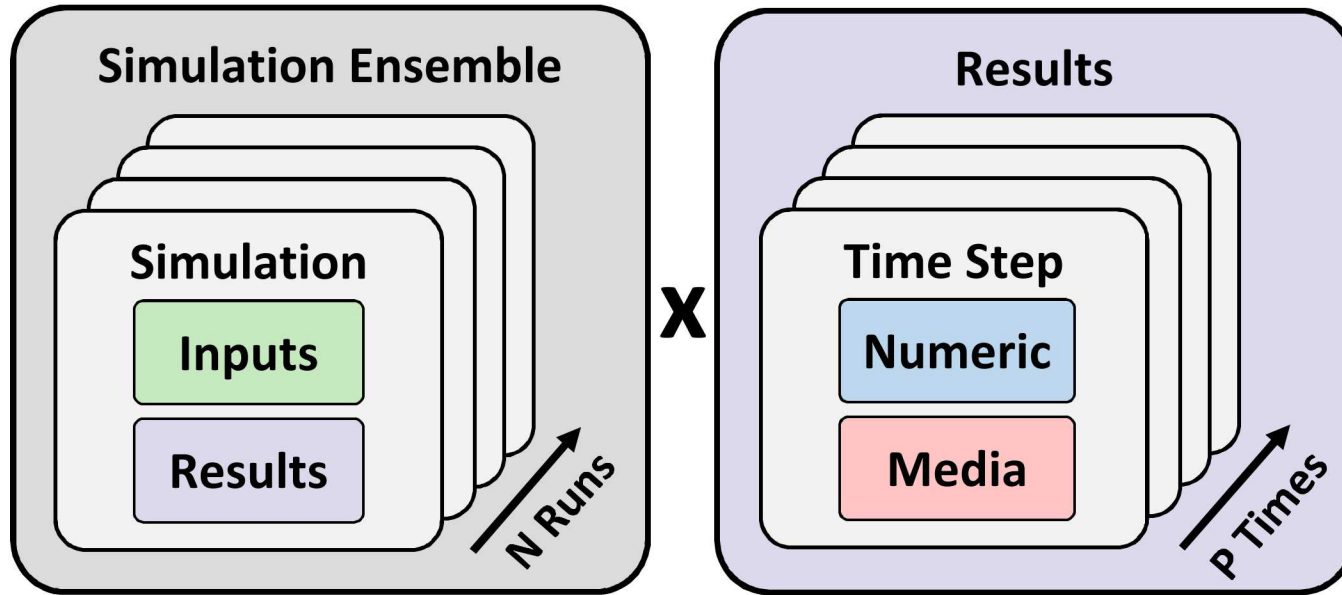


Ensemble Data is Challenging Because of Its Size & Complexity



500-1000 Runs
(Largest 50K)

Ensemble Data is Challenging Because of Its Size & Complexity

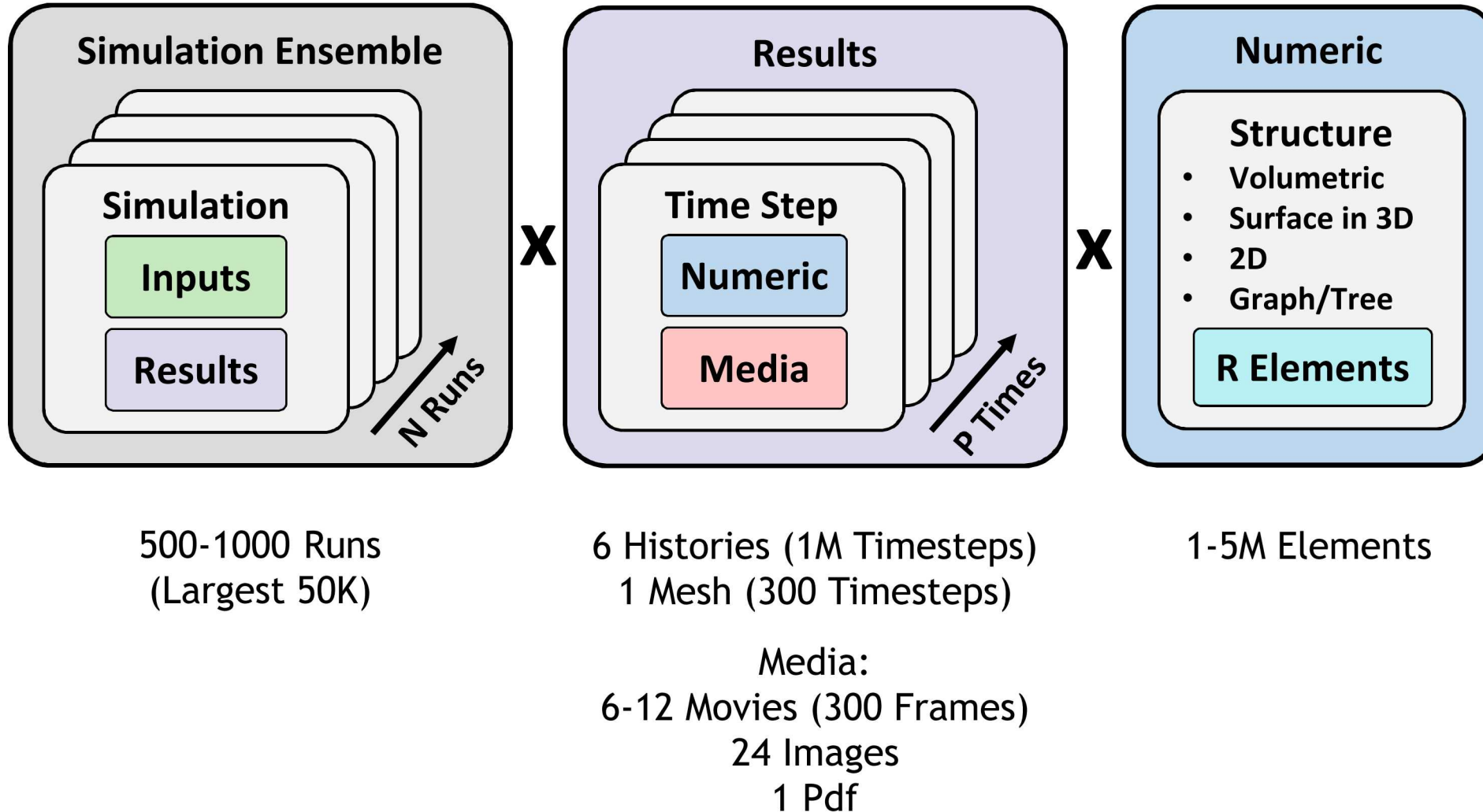


500-1000 Runs
(Largest 50K)

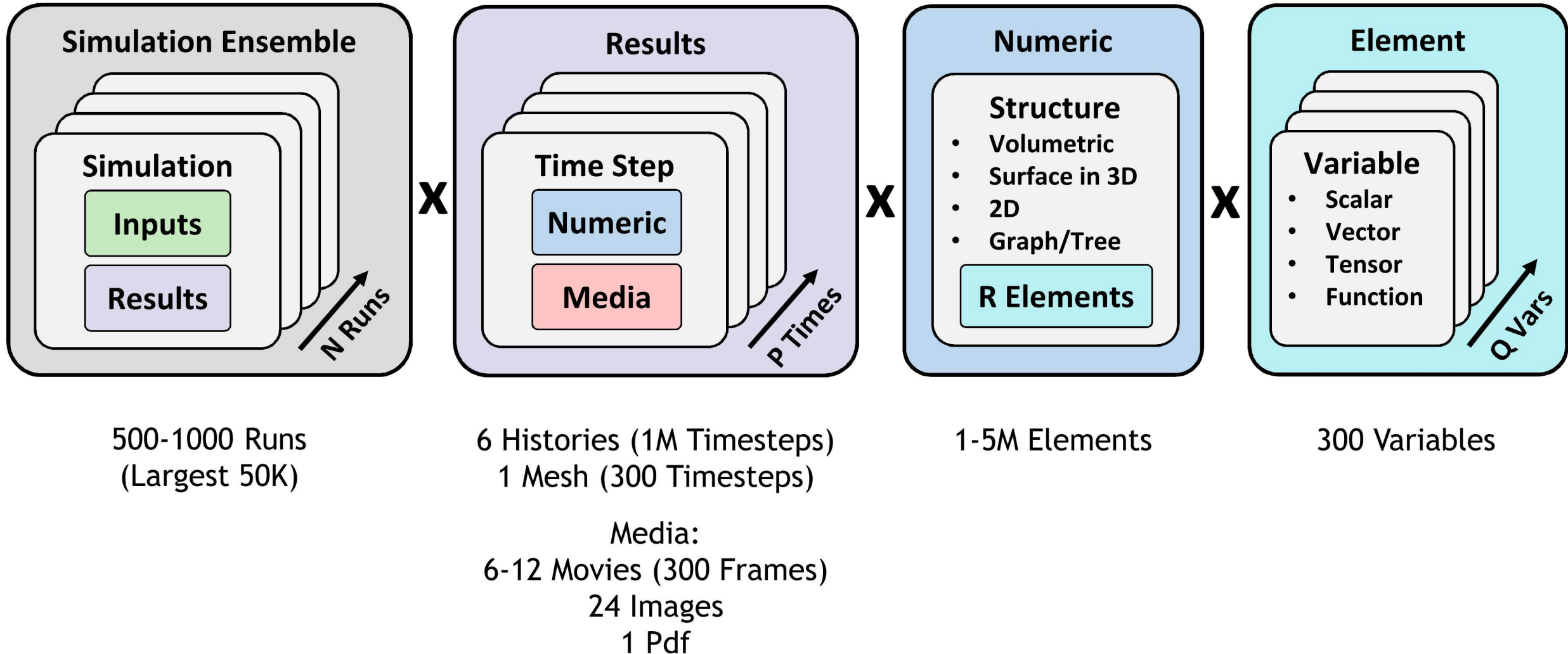
6 Histories (1M Timesteps)
1 Mesh (300 Timesteps)

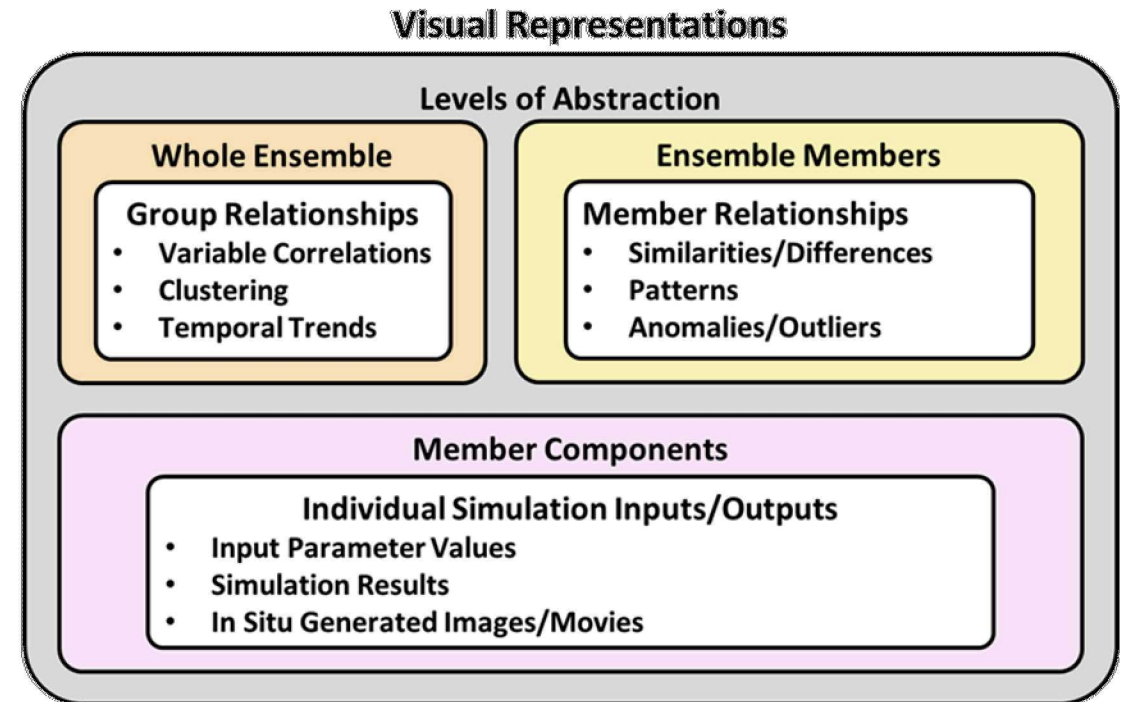
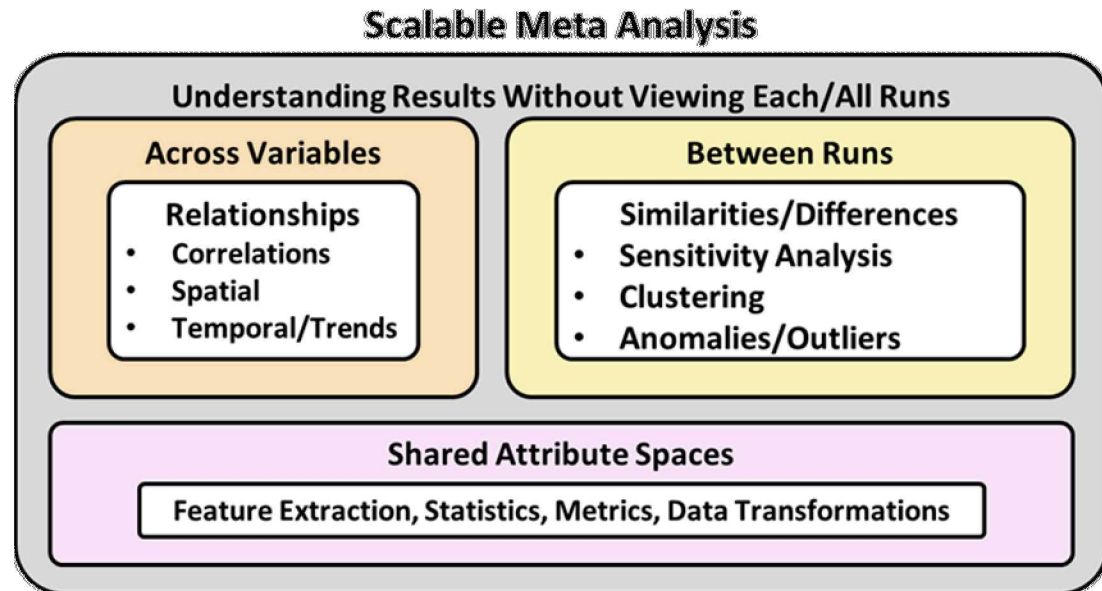
Media:
6-12 Movies (300 Frames)
24 Images
1 Pdf

Ensemble Data is Challenging Because of Its Size & Complexity



Ensemble Data is Challenging Because of Its Size & Complexity





Goal: A scalable, extensible analysis & visualization system for understanding ensemble data at multiple levels of abstraction: the ensemble as a whole, runs/sets of runs, result components.

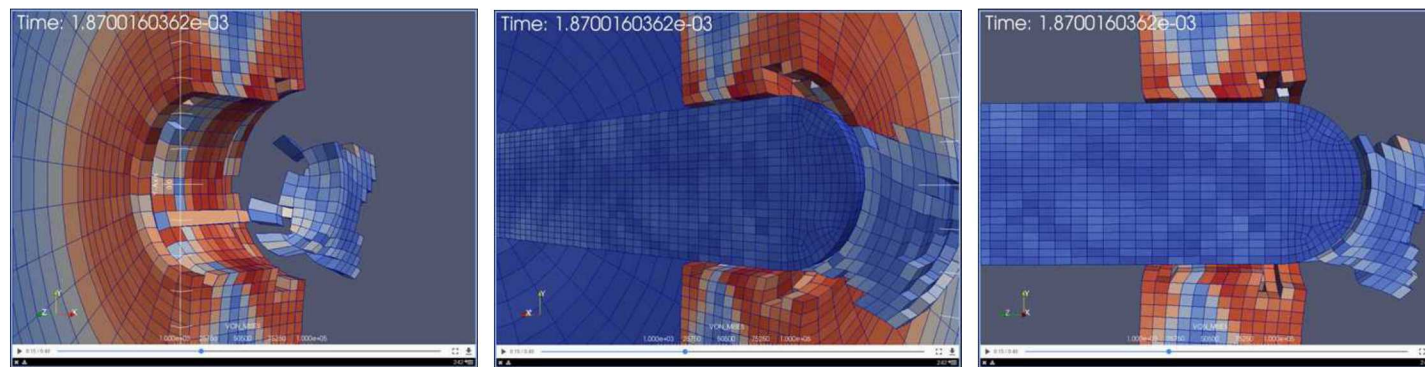
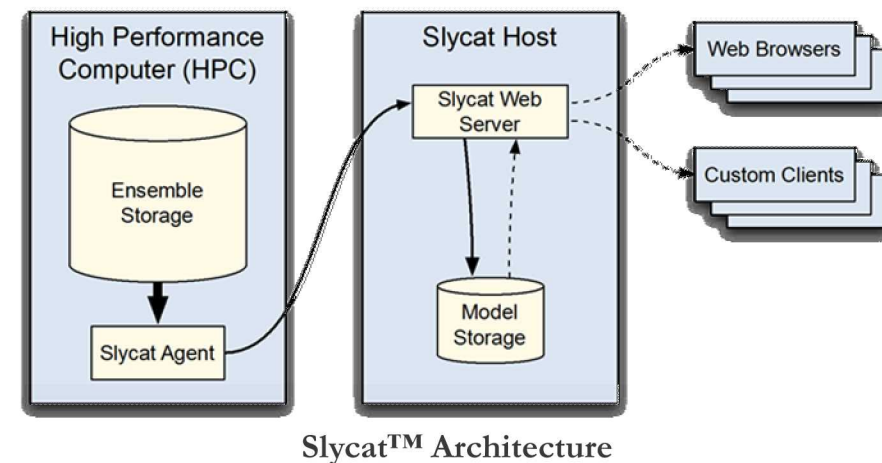
Slycat™ Architecture Targets Ensemble Size Scalability

Scalable Meta Analysis

- Analysis done in parallel on ensemble data on HPC
- Slycat™ server stores and manipulates reduced-scale models*
- Models are visualized remotely, interactively, and collaboratively with desktop web-based interface

Milestone Data Examples

- Milestone demonstrated scalability to 15K runs in web interface
- Generated using Sierra/SolidMechanics and Catalyst
- Material fracturing problem, punch impacting plate



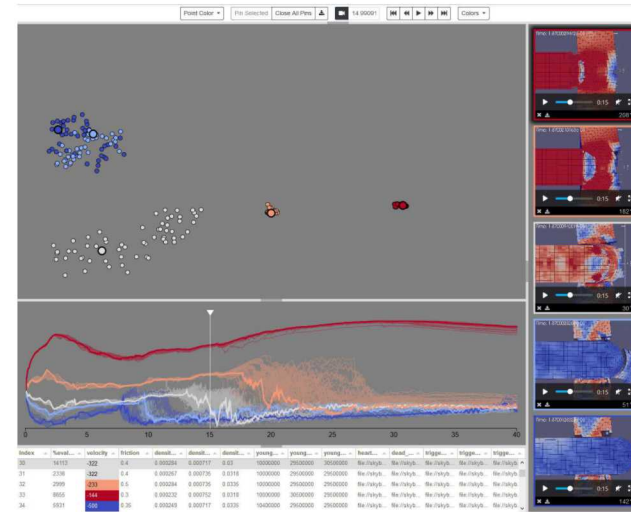
Three videos showing the simulation from different viewpoints, each consisting of 1000 frames.
All share identical colormaps, encoding Von Mises stress (red is high stress).

*Analytical models of results data, not physics models used by simulations

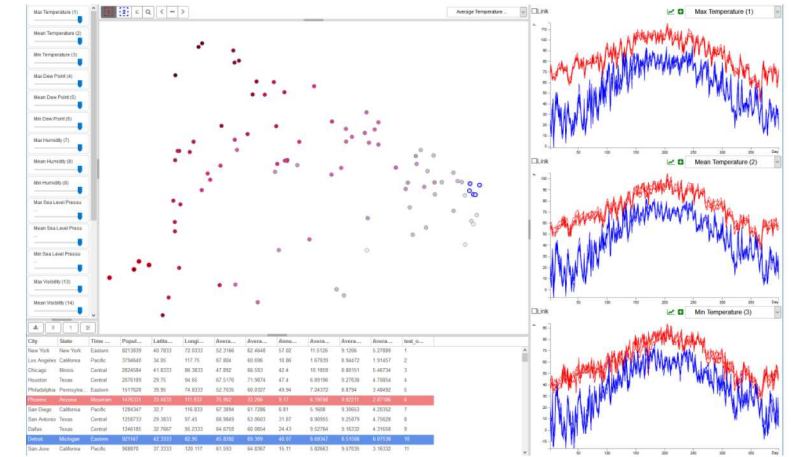
9 Slycat™ Models Provide Multiple Perspectives

Visual Representations

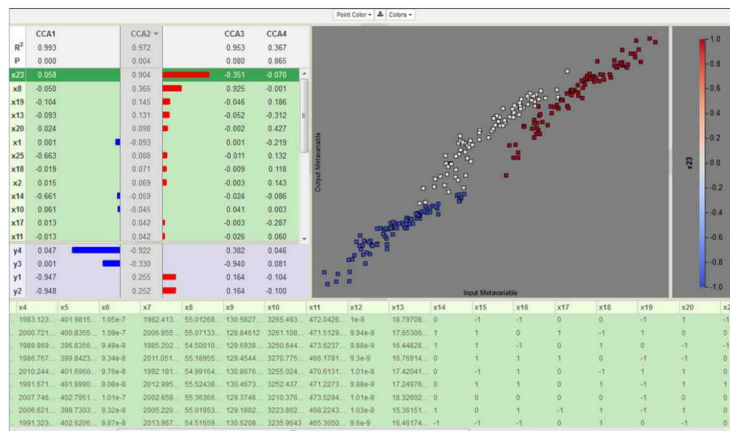
- Slycat™ models target specific result data types, encapsulating analysis results, visual representations, and interactions.
- Models provide multiple levels of abstraction: the ensemble as a whole, individual runs, and raw data.
- Models support multi-media results.



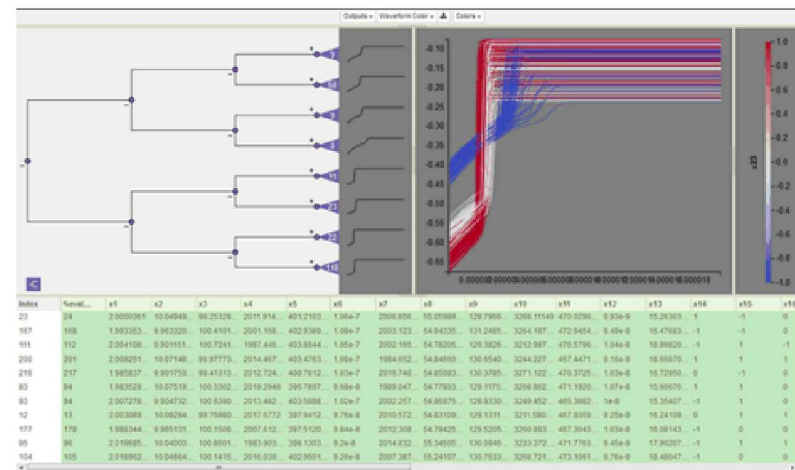
VideoSwarm Model



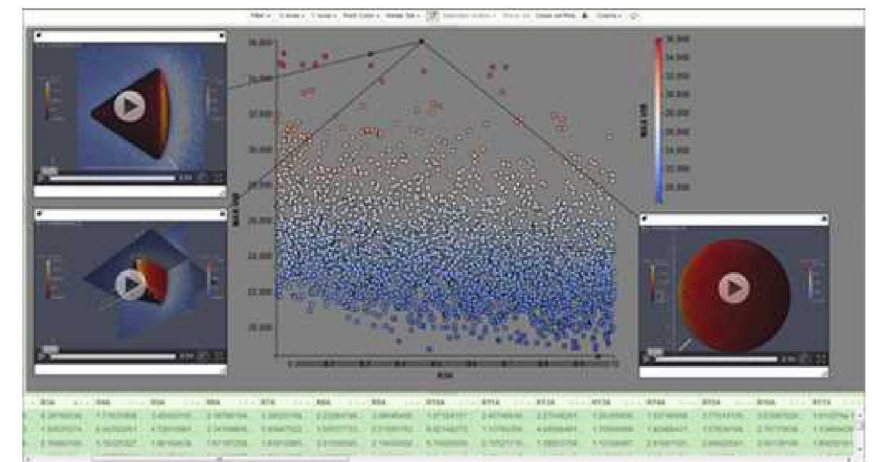
Dial-A-Cluster Model



Canonical Correlation Analysis Model



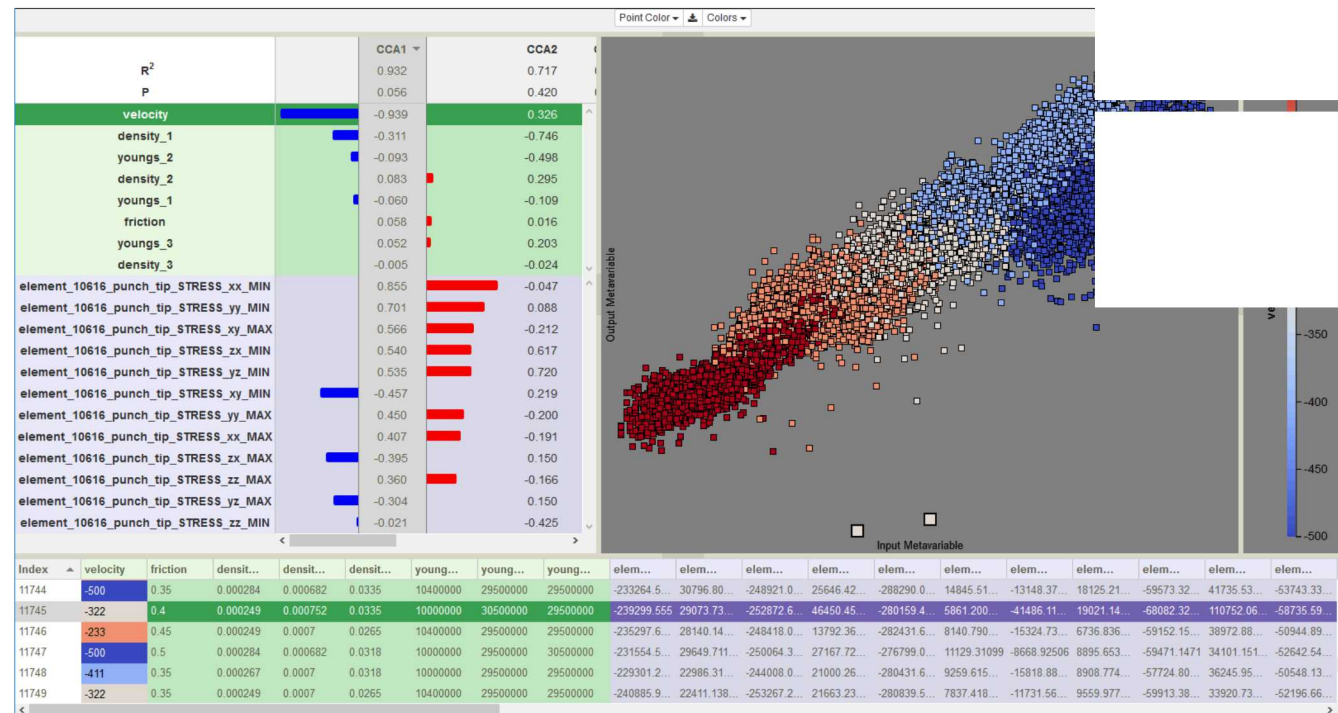
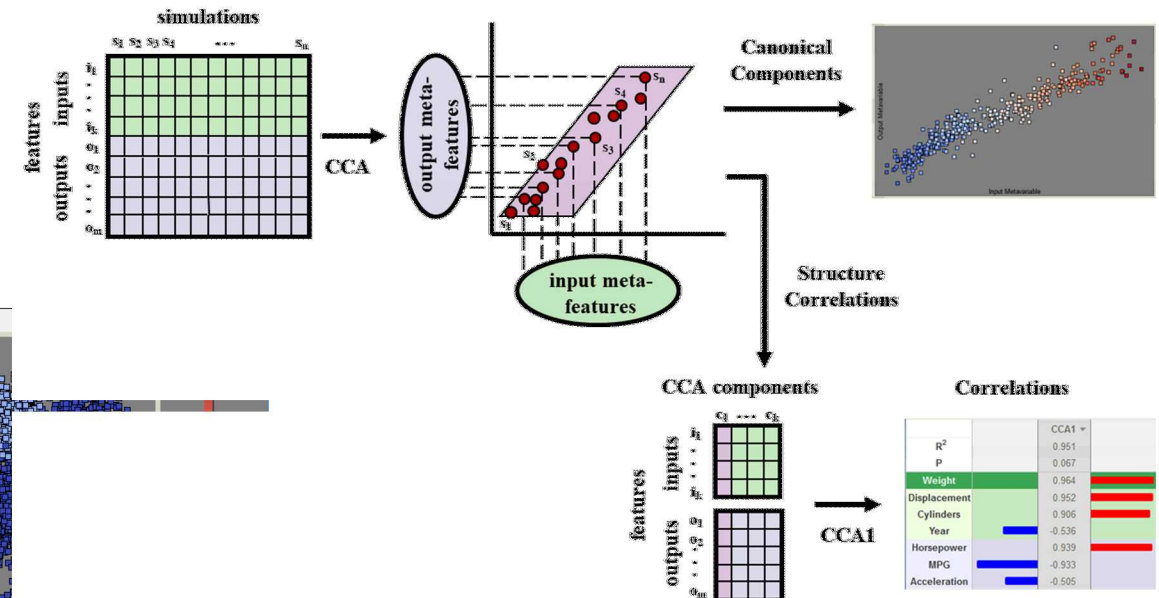
Time Series Model



Parameter Space Model

Identifying & Understanding Relationships Between Variables Is Important

- **Algorithm:** Canonical Correlation Analysis (CCA) generalization of Principal Component Analysis (Hotelling 1936)
- **Analysis:** Correlations between 2 sets of variables, sensitivity analysis, anomaly detection
- **Data Type:** Tables of scalar values



Bar chart shows velocity input parameter is highly correlated with output stress metrics (15K runs). Scatterplot color-coded by velocity (5 values due to Latin Hypercube sampling).

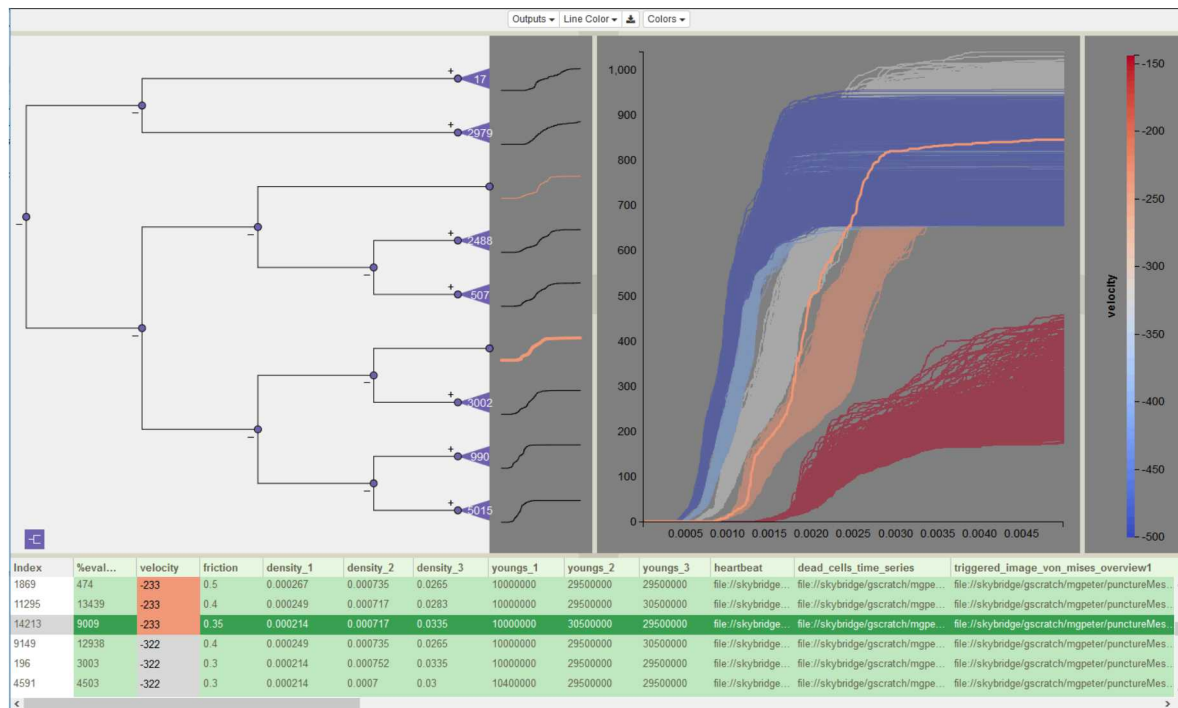
CCA results used in scatterplot (relationships between runs) and bar chart (relationships for ensemble-as-a-whole)

H. Hotelling "Relations between two sets of variates," *Biometrika* 28(3/4): pp. 321-377, 1936

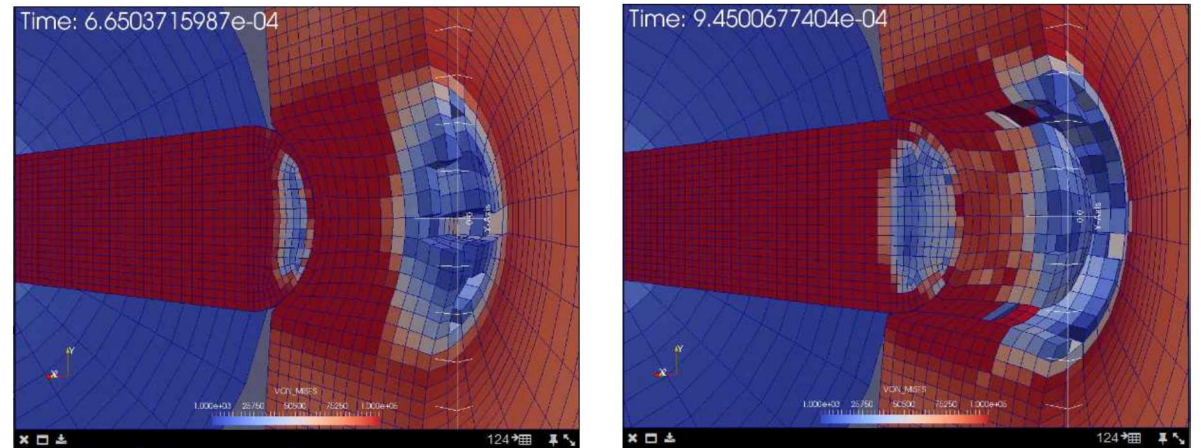
P. J. Crossno, T. M. Shead, M. A. Sielicki, W. L. Hunt, S. Martin, and M.-Y. Hsieh, "Slycat Ensemble Analysis of Electrical Circuit Simulations," *Topological and Statistical Methods for Complex Data*, Springer-Verlag, 2015

Time Series Clustering Reveals Relationships Between Inputs and Groups of Output Results

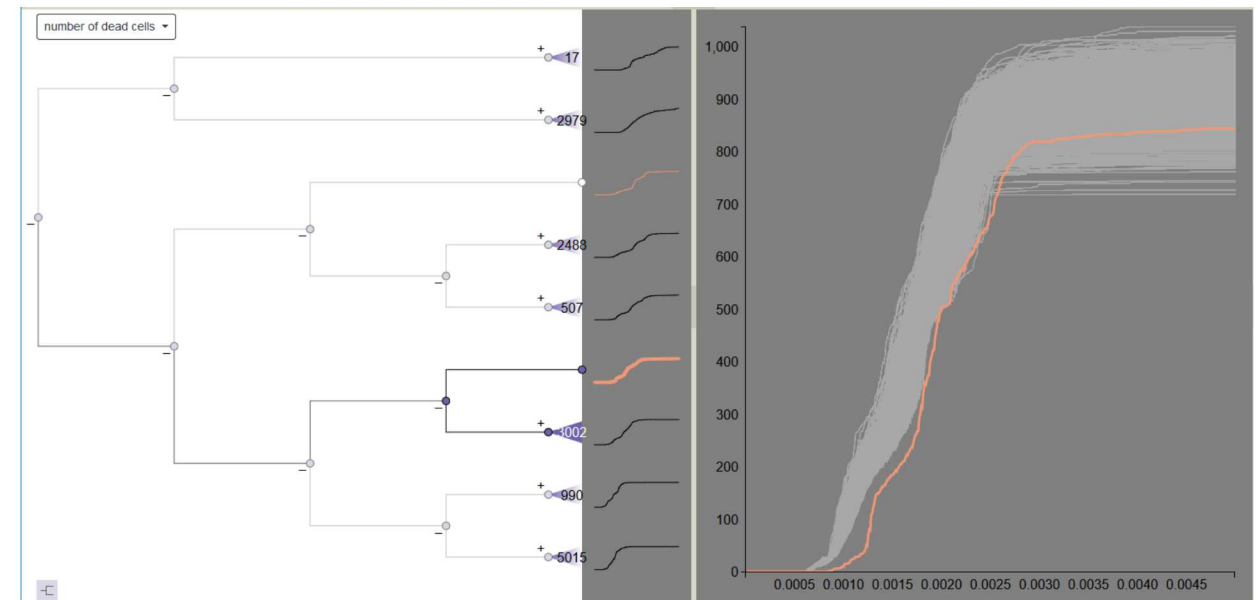
- **Algorithm:** Agglomerative hierarchical clustering
- **Analysis:** Time series similarity, shape filtering, map output variability to inputs, find outliers
- **Data Type:** time series data, vector data



Dendrogram clusters similar time series (ensemble-as-a-whole); plot reveals outliers (relationships between runs) for 15K runs showing cell deaths over time colored by input velocity.



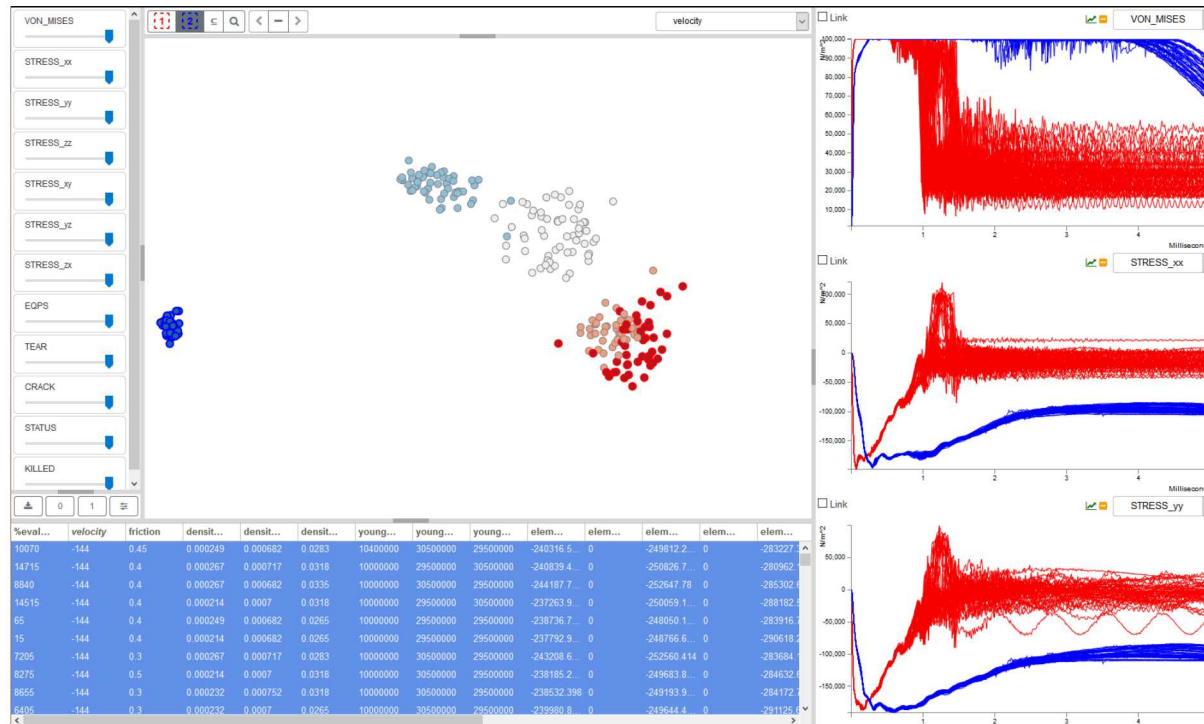
Cells die over time, allowing a plug of cells to be pushed out from the plate.



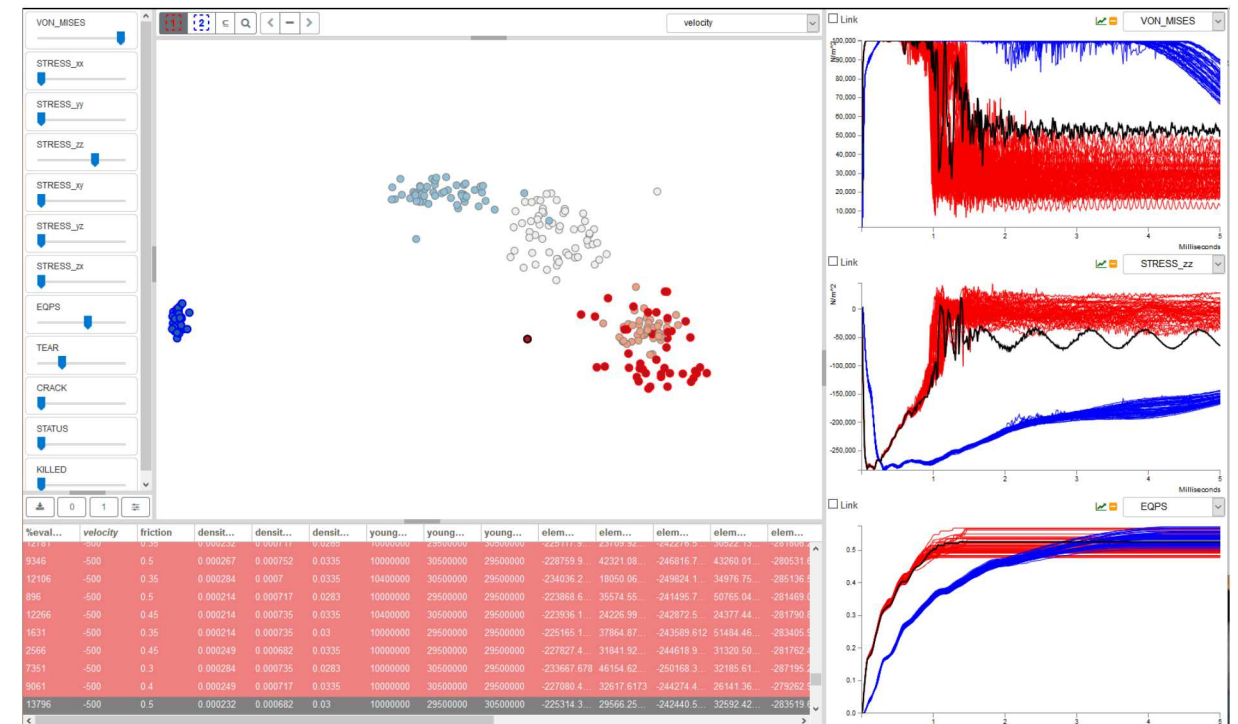
Dendrogram interactively controls plot visibility to compare selected run with cluster.

Dial-A-Cluster Combines Multiple Time Series Variables Into a Single Projection Space

- **Algorithm:** Multi-dimensional scaling (MDS)
- **Analysis:** Multivariate time series clustering, temporal comparison of sets
- **Data Type:** time series data, vector data



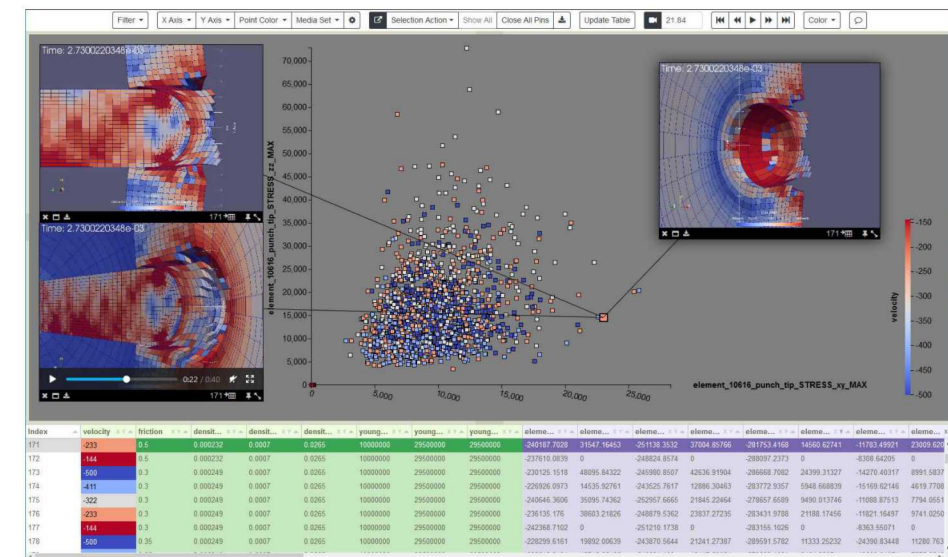
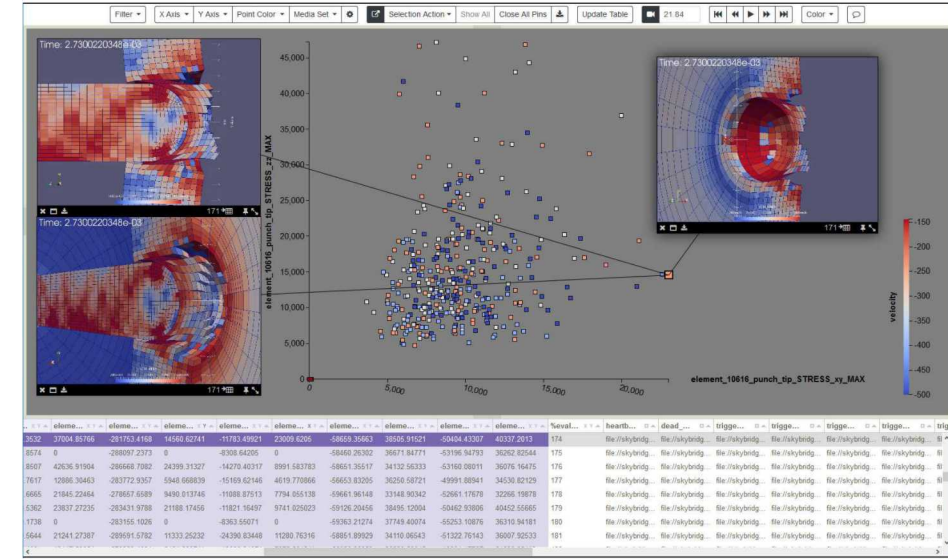
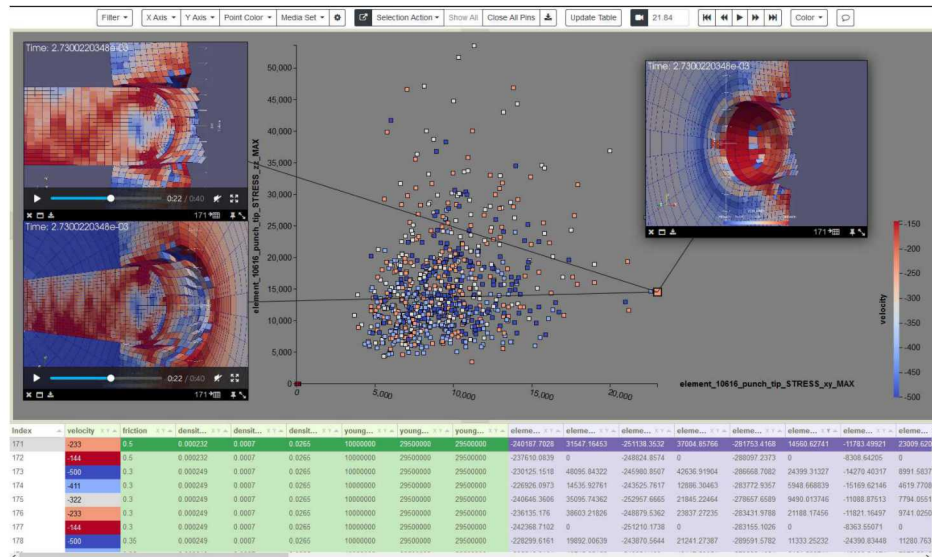
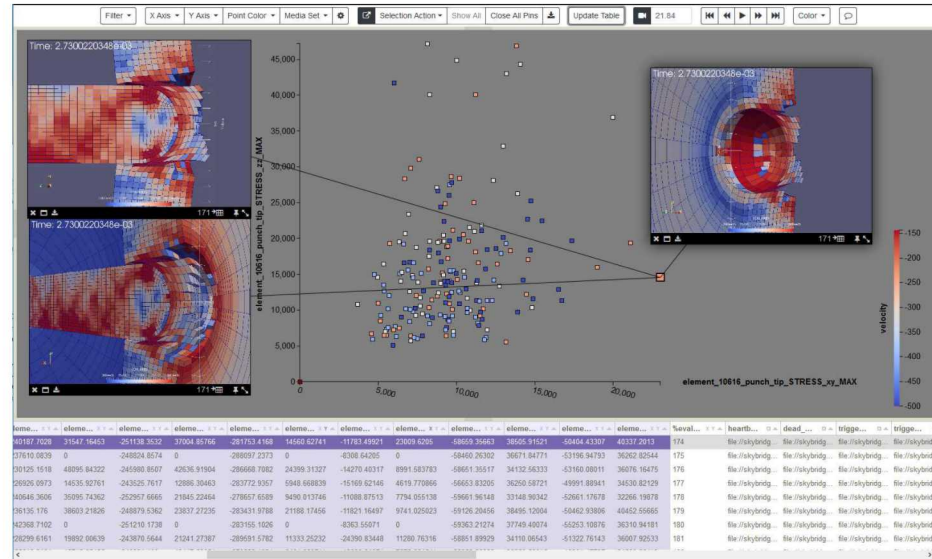
Time series for element 5768, the dimple element on the plate, prior to optimizing slider values against input velocity values.



Time series for element 5768, the dimple element on the plate, after optimizing slider values against input velocity values. The 3 most significant temporal variables relative to input velocity are shown in the plots.

Martin, S., and Quach, T.-T., "Interactive Visualization of Multivariate Time Series Data," in *Human Computer Interaction International (HCII), Foundations of Augmented Cognition*: 322-332. 2016

New Iterative Ensemble Analysis in Parameter Space Model

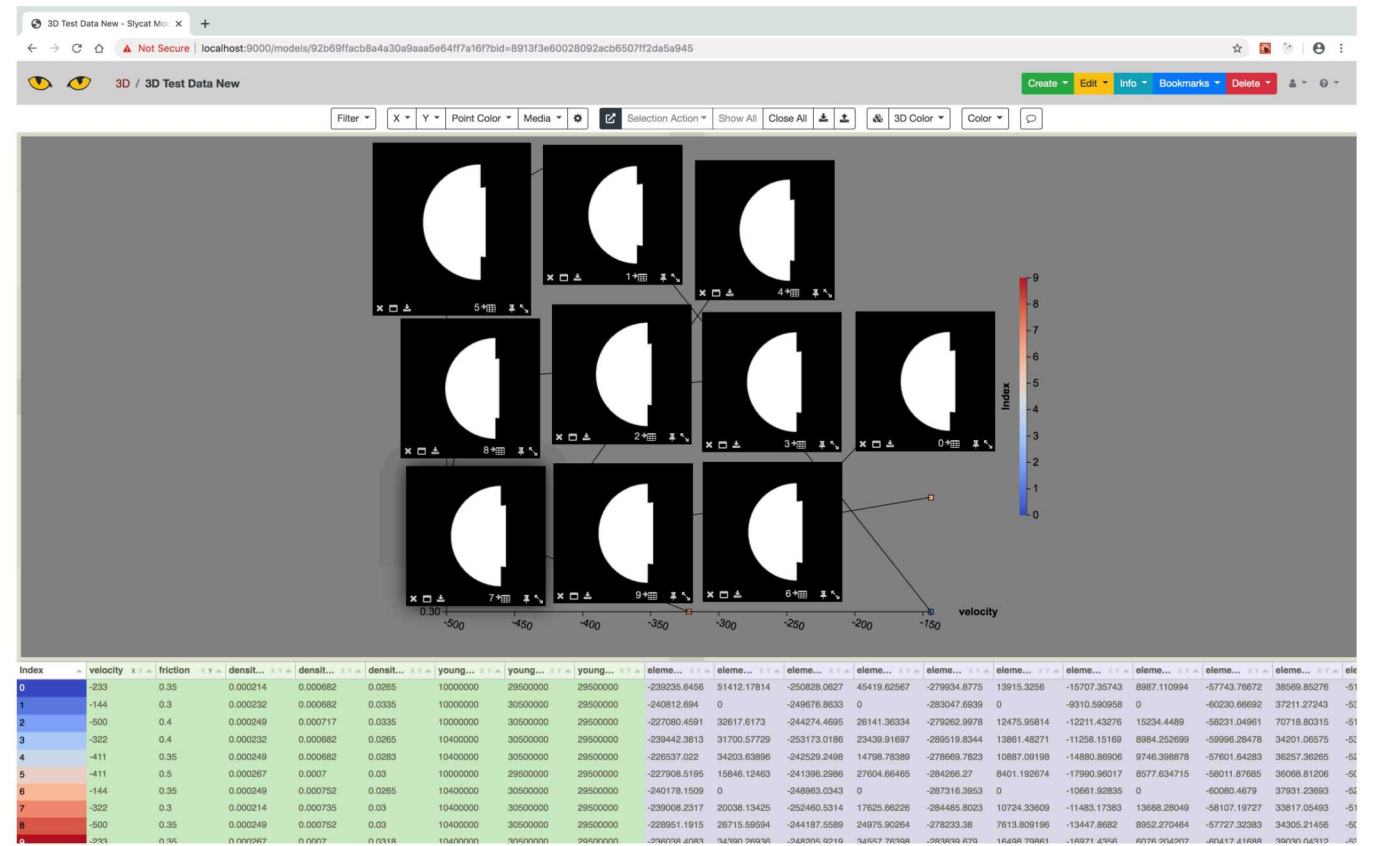
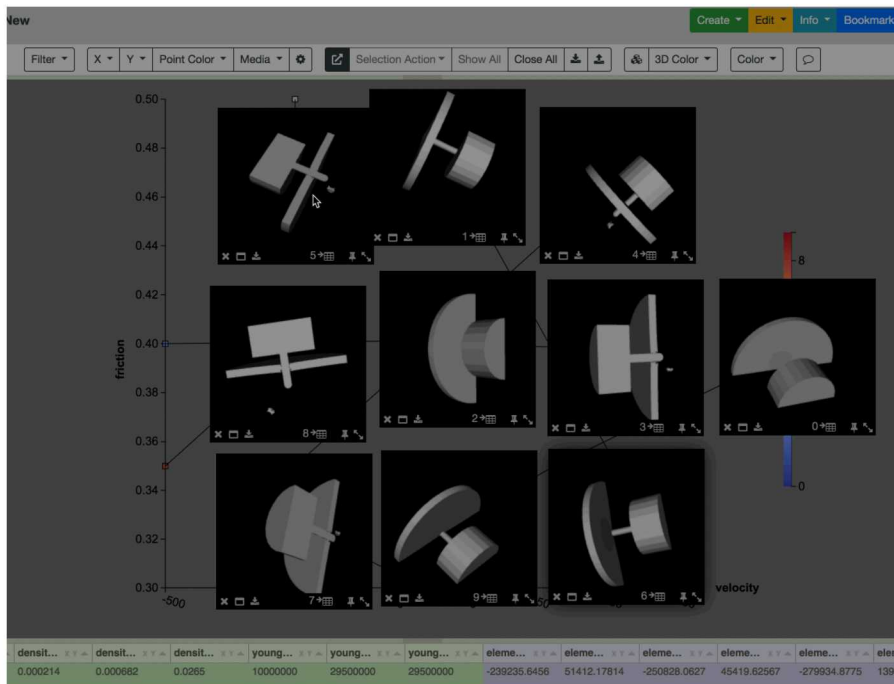


We Are Currently Developing a Coordinated Surface Mesh Viewer for Parameter Space Models

Still in development, not available on servers

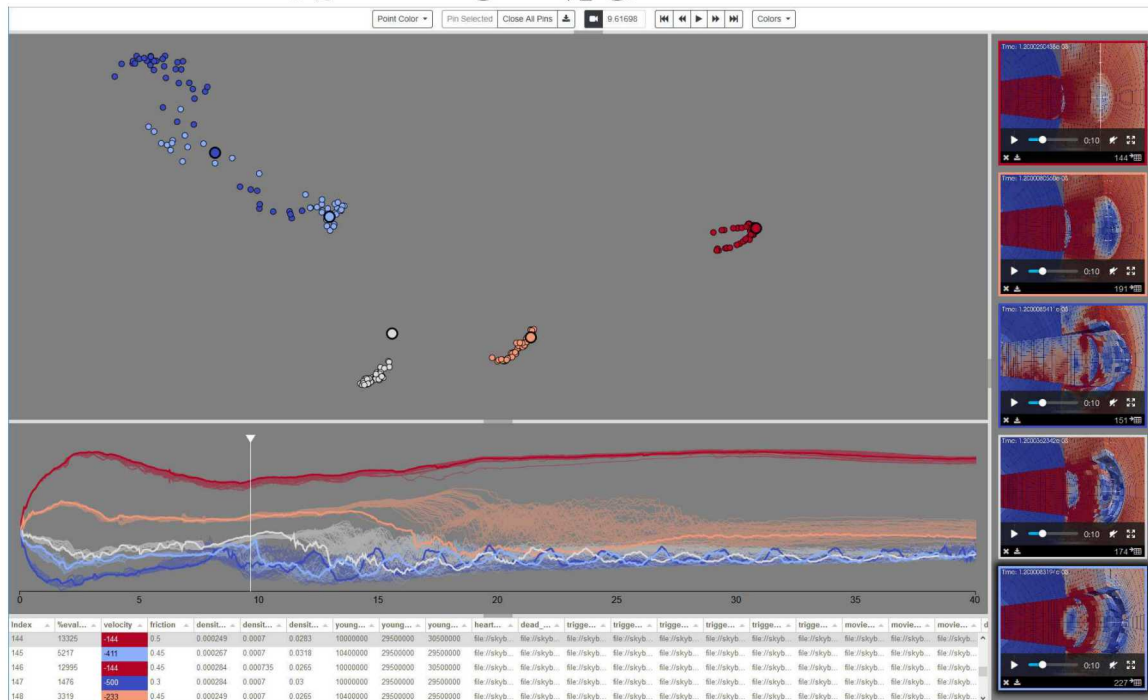
- Surface mesh geometry, no volumes
- Selectable variables for surface color
- Similar to videos, synchronized viewing

Operates on geometry in Kitware's **vtp** format

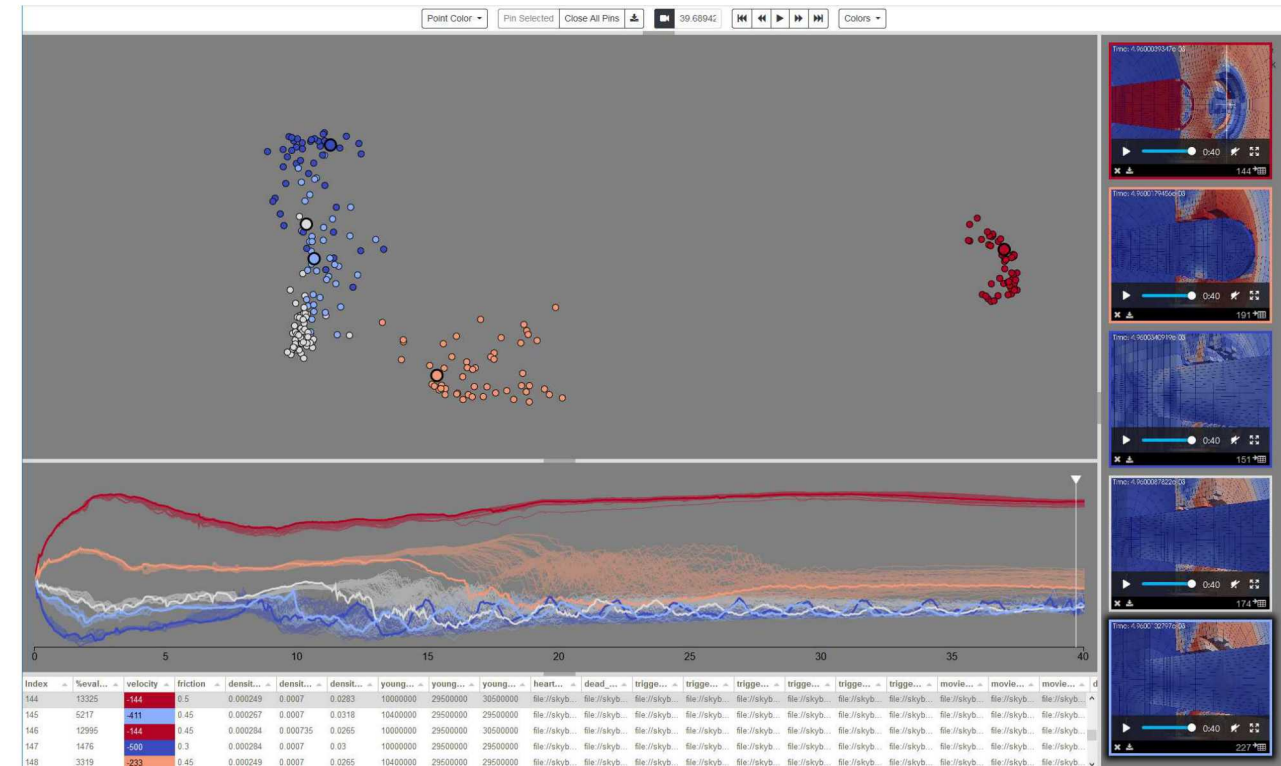


Video Comparison Alone is Insufficient, We Need Temporal Video Clustering

- **Algorithm:** Multi-dimensional scaling (MDS), aligned temporal evolution (see our paper)
- **Analysis:** Identify clusters, reveal temporal behaviors, map output variability to inputs, find outliers, compare videos
- **Data Type:** images (jpg), videos



Video similarity: current frame (scatterplot), over entire simulation (graph), plus synchronized playback for selected runs.



Slowest input velocity runs (red) are separate in graph because they all fail to penetrate the plate.

S. Martin, M. A. Sielicki, J. Gittinger, M. Letter, W. L. Hunt, and P. J. Crossno. "VideoSwarm: Analyzing Video Ensembles", *S&T International Symposium on Electronic Imaging 2019, Visualization and Data Analysis (VDA) 2019*.

D. Jäckle, F. Fischer, T. Schreck and D. A. Keim, "Temporal MDS Plots for Analysis of Multivariate Data," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 141-150, 31 Jan. 2016. Uses sliding window approach.

Slycat™ Addresses Multiple Ensemble Analysis and Visualization Challenges to Successfully Meet Our Goal

- **Data Complexity**
 - Models tailored to address specific results types
 - Multiple model types with complimentary perspectives on the data
 - Remote access to in situ generated media (images, movies)
- **Scalability**
 - Parallel analysis on HPC, visualize analysis artifacts remotely
 - Visual representations with multiple levels of abstraction
 - Drill-down to result data of interest
- **Usability**
 - User Interface Design and Interactivity (key to user adoption)
 - Group operations on sets of ensemble members
 - Extensible plug-in architecture

Goal: A scalable, extensible analysis & visualization system for understanding ensemble data at multiple levels of abstraction: the ensemble as a whole, runs/sets of runs, result components.