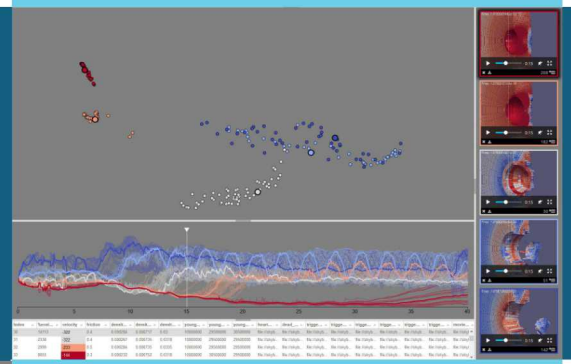
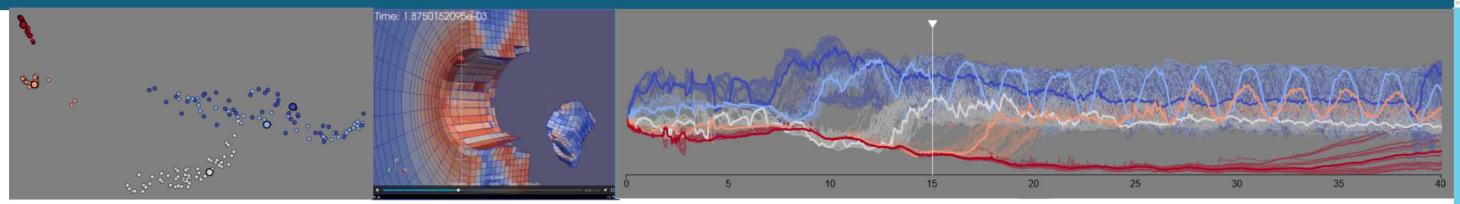
SAND2020-4086C

Slycat™ Ensemble Analytics





PRESENTED BY

Patricia Crossno, Ph.D.

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

DOE Computer Graphics Forum, April 2020

SAND xxxx PE

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.

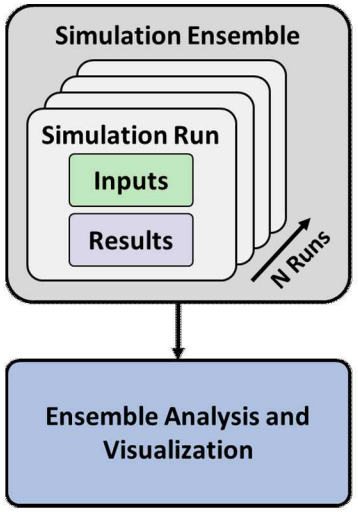
Some Important Questions Cannot Be Answered with Just One Simulation Run

Simulation ensembles are used for answering:

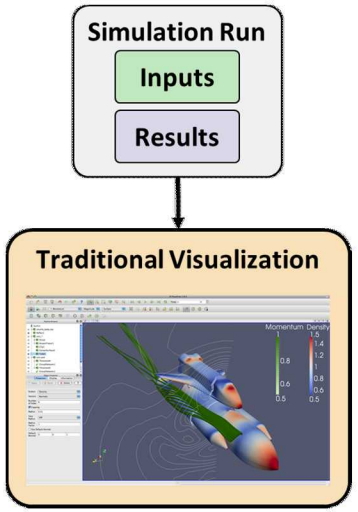
- What effect do small changes in input produce in the outputs? (Sensitivity Analysis)
- What range of output values are possible? (Model validation)
- How many different types of results are possible? (Categorization/Clustering)
- Which results are unique/anomalous? (Outliers)

Traditional tools not designed for ensembles

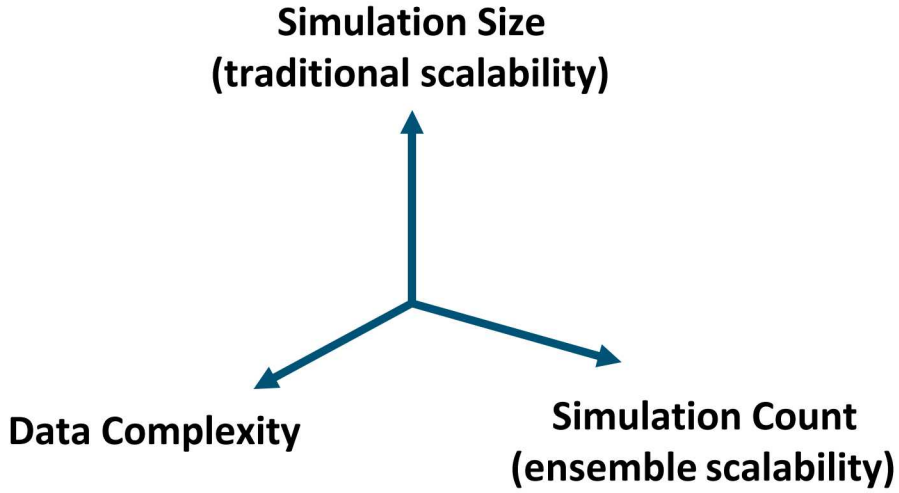
- Do not scale in number of runs
  - Designed to scale with problem size
  - Display limited to handful of results
- Not designed for ensemble results data
  - Designed for geometric data
  - Not designed for sets of multi-media outputs



Ensemble – set of related simulation runs

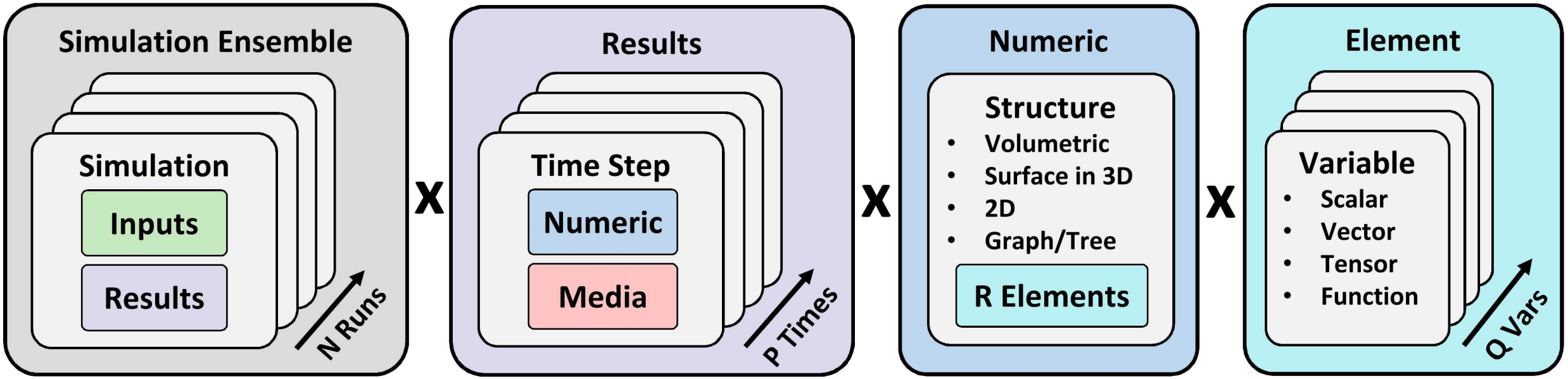


ParaView – visualization scales in simulation size



Scalability Research

Ensemble Data is Challenging Because of Its Size & Complexity



500-1000 Runs
(Largest 50K)

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

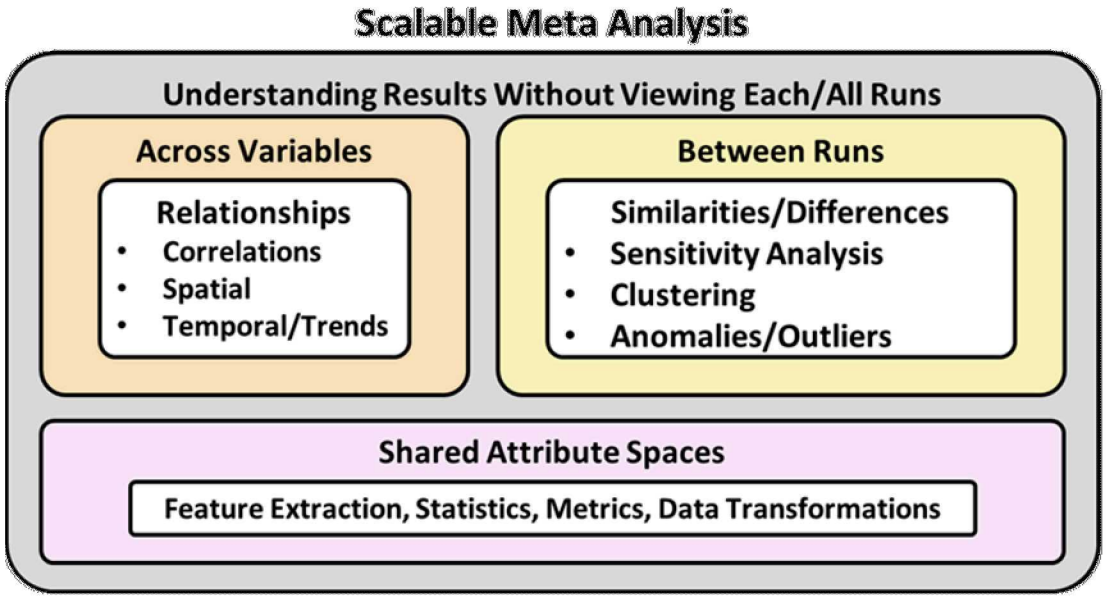
Media:
6-12 Movies (300 Frames)
24 Images
1 Pdf

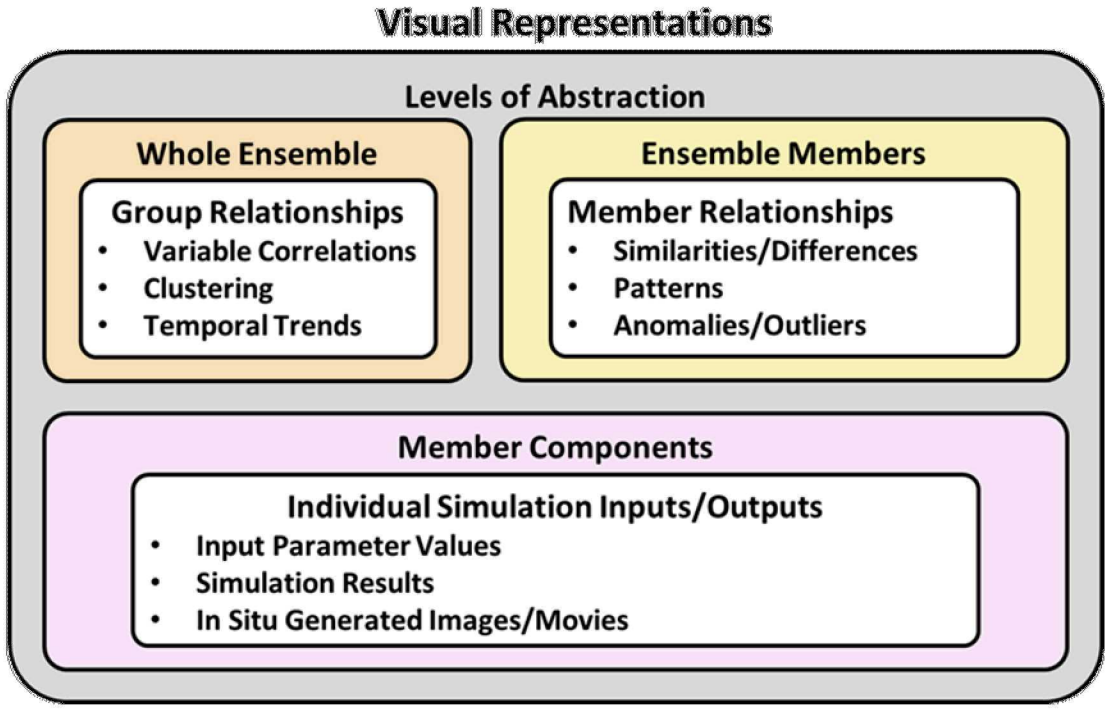
1-5M Elements

300 Variables

P. Crossno, "Challenges in Visual Analysis of Ensembles," in IEEE Computer Graphics and Applications, 38/2, pp. 122-131, Mar./Apr. 2018.)

Ensembles Demand Supporting Analysis & Visualization Capabilities





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.

P. Crossno, "Challenges in Visual Analysis of Ensembles," in IEEE Computer Graphics and Applications, 38/2, pp. 122-131, Mar./Apr. 2018.)

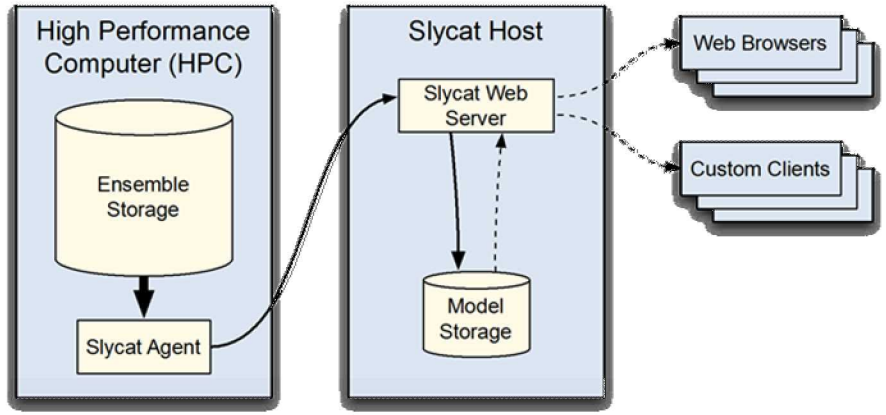
Slycat™ Architecture Targets Ensemble Size Scalability

Scalable Meta Analysis

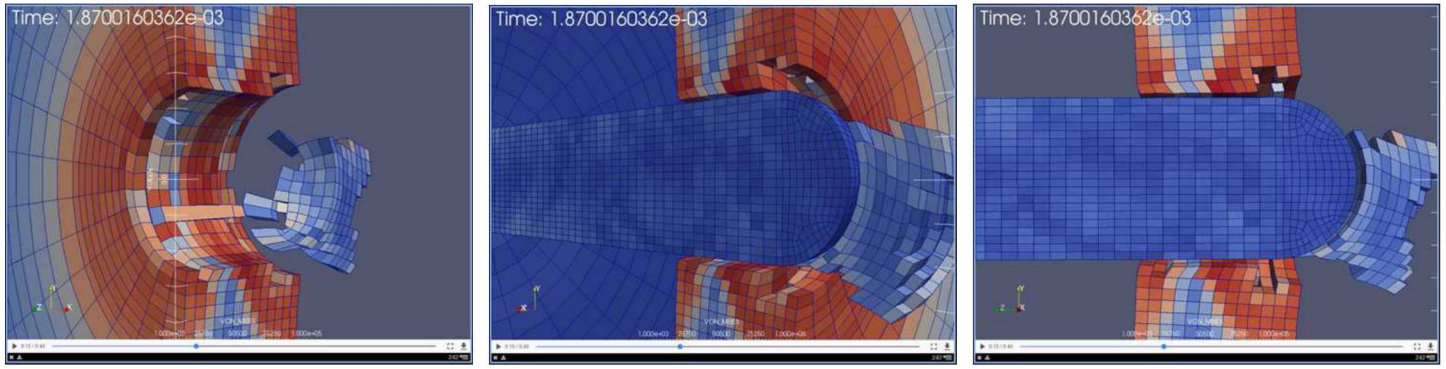
- Analysis done in parallel on ensemble data on HPC
- Slycat™ server stores and manipulates reduced-scale models*
- Models are visualized remotely, interactively, and collaboratively with desktop web-based interface

Milestone Data Examples

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



Slycat™ Architecture



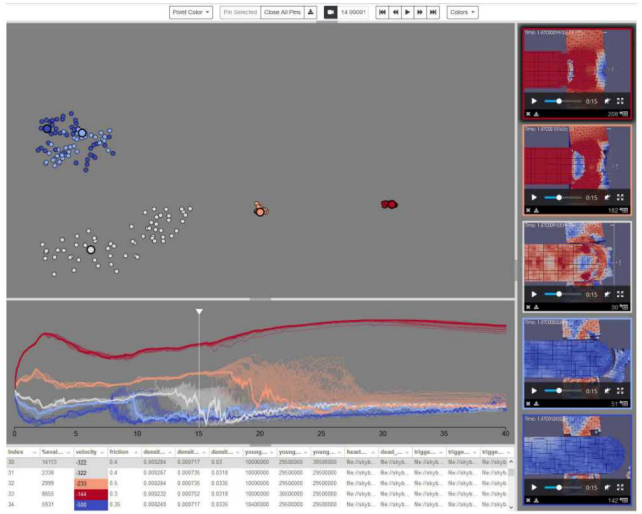
Three videos showing the simulation from different viewpoints, each consisting of 1000 frames. All share identical colormaps, encoding Von Mises stress (red is high stress).

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

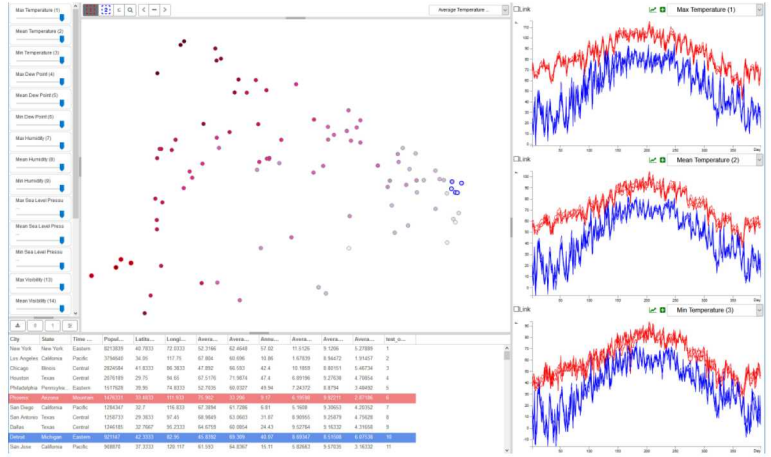
Slycat™ Models Provide Multiple Perspectives

Visual Representations

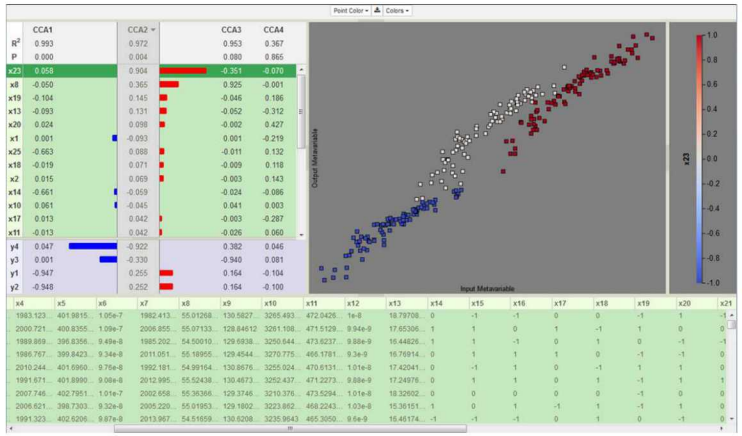
- Slycat™ models target specific result data types, encapsulating analysis results, visual representations, and interactions.
- Models provide multiple levels of abstraction: the ensemble as a whole, individual runs, and raw data.
- Models support multi-media results.



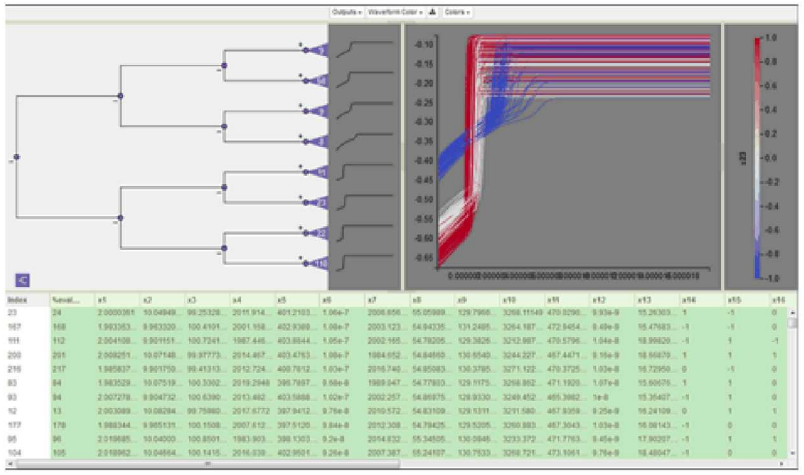
VideoSwarm Model



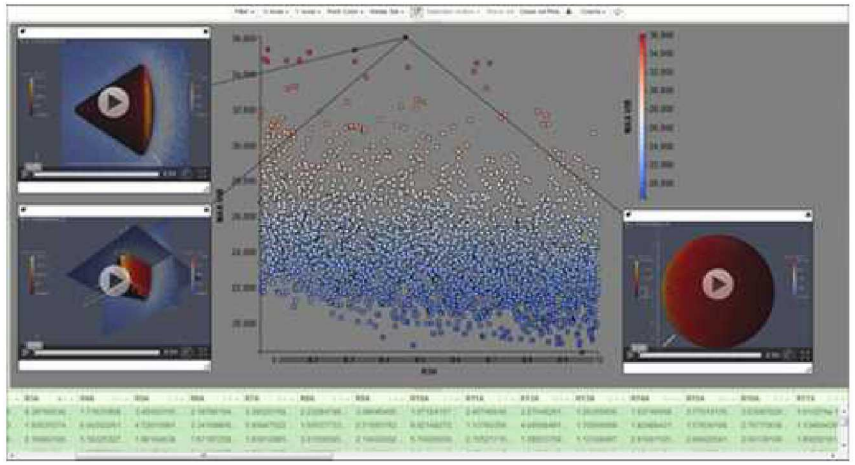
Dial-A-Cluster Model



Canonical Correlation Analysis Model



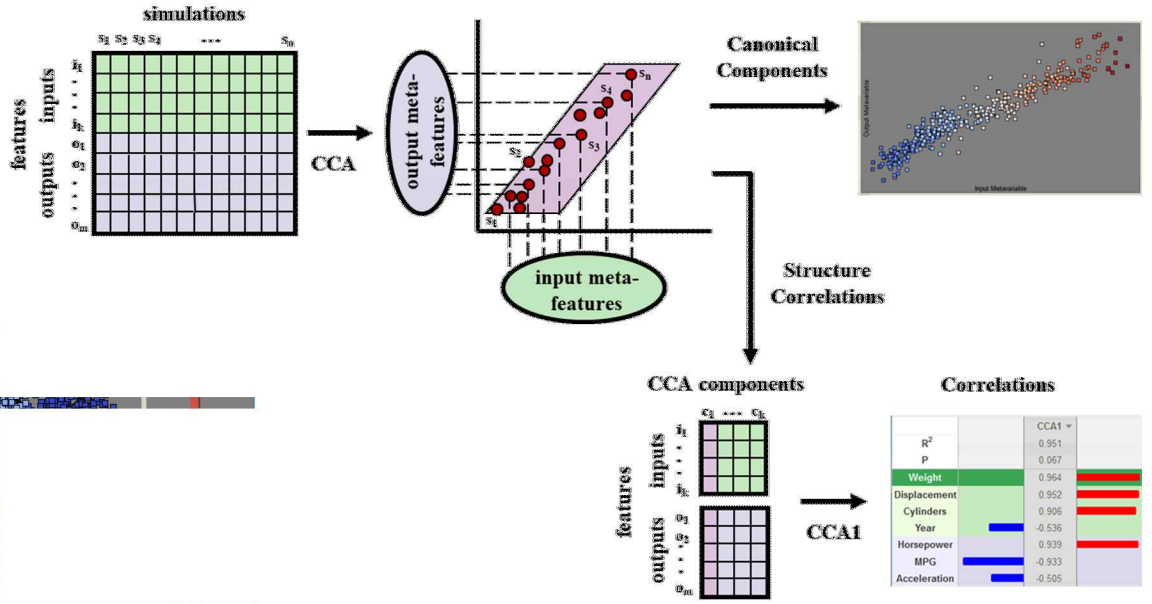
Time Series Model



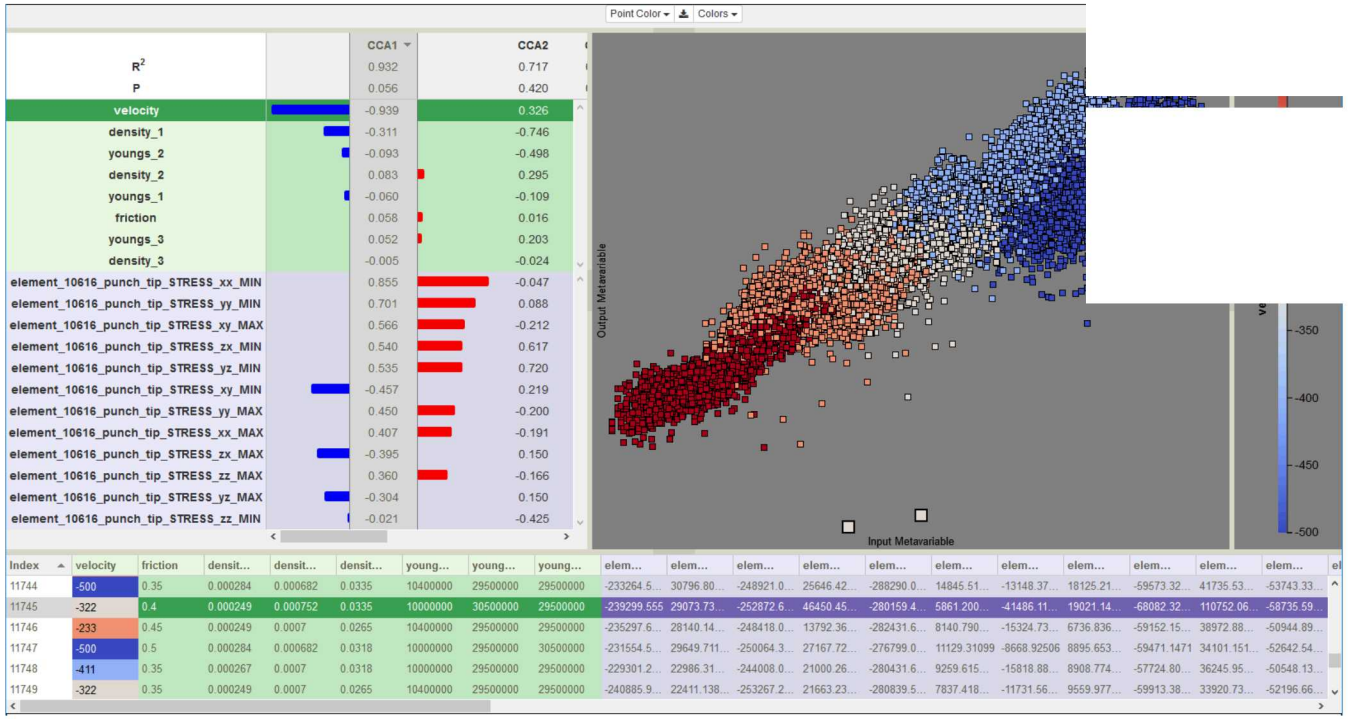
Parameter Space Model

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



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



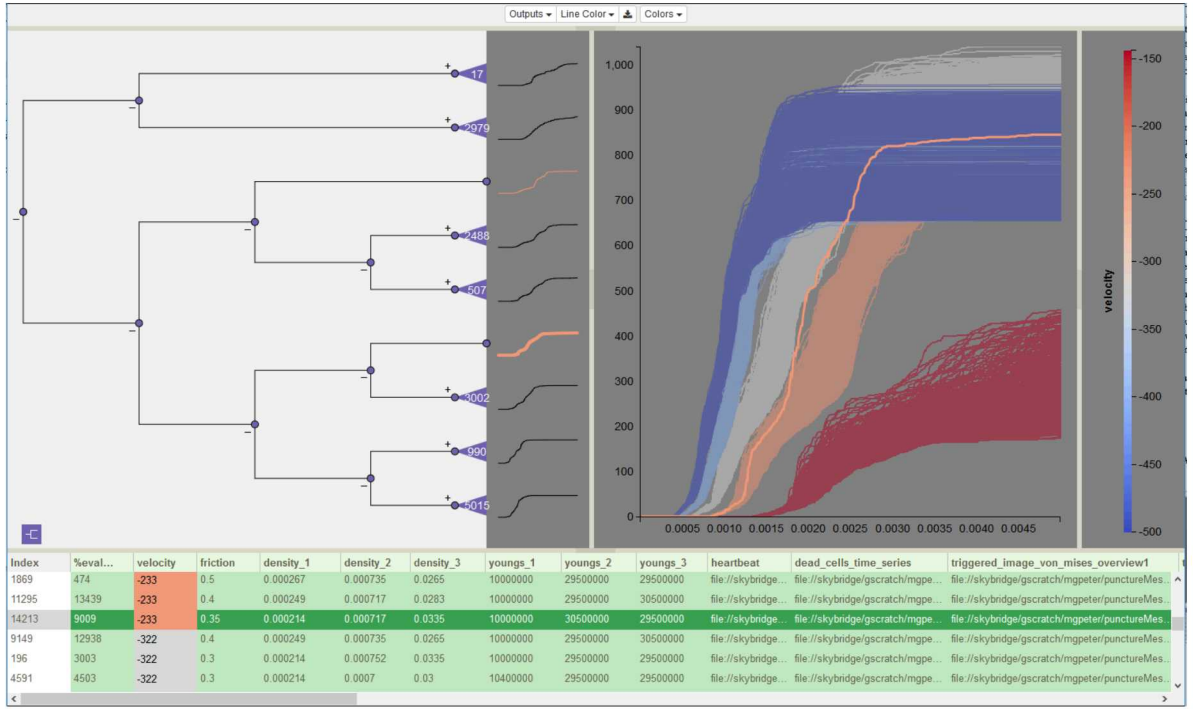
Bar chart shows velocity input parameter is highly correlated with output stress metrics (15K runs). Scatterplot color-coded by velocity (5 values due to Latin Hypercube sampling).

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. Hseih, "Slycat Ensemble Analysis of Electrical Circuit Simulations," *Topological and Statistical Methods for Complex Data*, Springer-Verlag, 2015

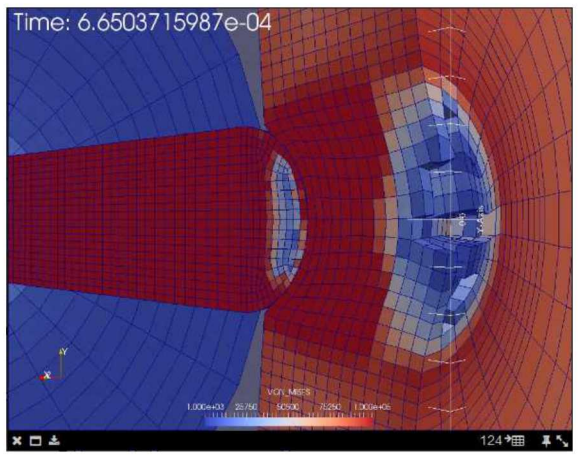
Time Series Clustering Reveals Relationships Between Inputs and Groups of Output Results

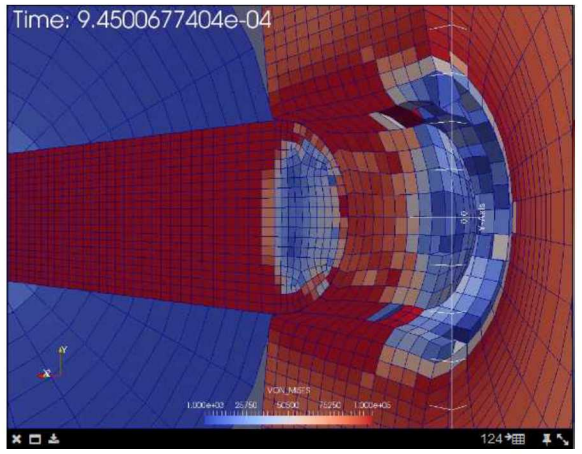
- **Algorithm:** Agglomerative hierarchical clustering
- **Analysis:** Time series similarity, shape filtering, map output variability to inputs, find outliers
- **Data Type:** time series data, vector data



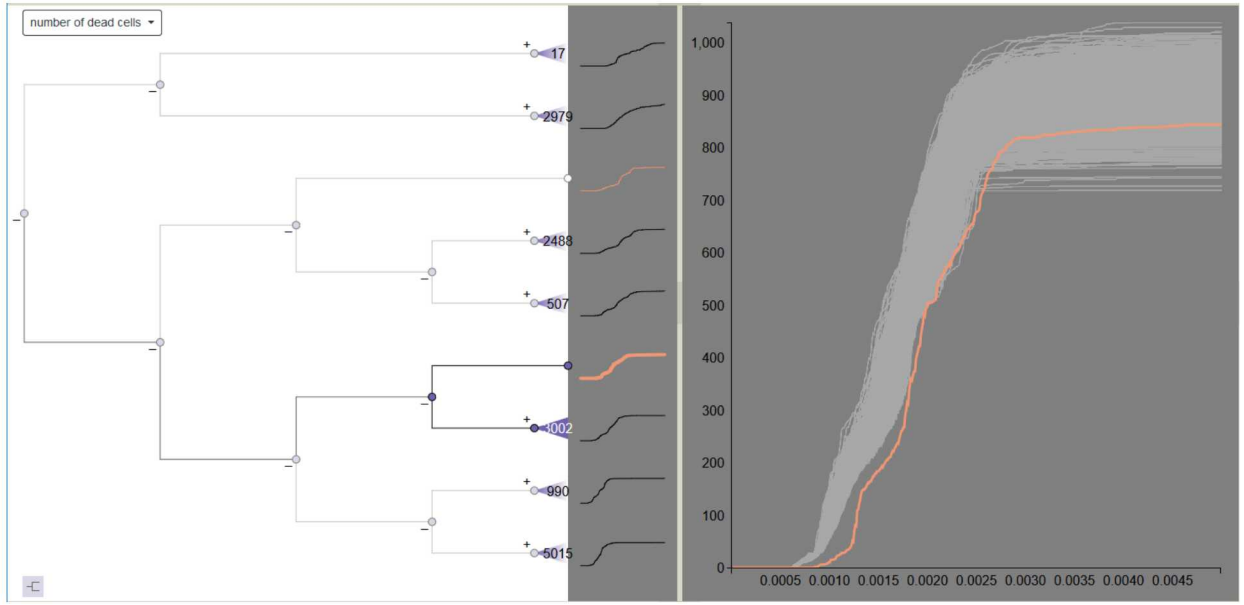
Index	%eval...	velocity	friction	density_1	density_2	density_3	youngs_1	youngs_2	youngs_3	heartbeat	dead_cells_time_series	triggered_image_von_mises_overview1
1869	474	-233	0.5	0.000267	0.000735	0.0265	10000000	29500000	29500000	file://skybridge...	file://skybridge/gscratch/mgpe...	file://skybridge/gscratch/mgpeter/punctureMes...
11295	13439	-233	0.4	0.000249	0.000717	0.0283	10000000	29500000	30500000	file://skybridge...	file://skybridge/gscratch/mgpe...	file://skybridge/gscratch/mgpeter/punctureMes...
14213	9009	-233	0.35	0.000214	0.000717	0.0335	10000000	30500000	29500000	file://skybridge...	file://skybridge/gscratch/mgpe...	file://skybridge/gscratch/mgpeter/punctureMes...
9149	12938	-322	0.4	0.000249	0.000735	0.0265	10000000	29500000	30500000	file://skybridge...	file://skybridge/gscratch/mgpe...	file://skybridge/gscratch/mgpeter/punctureMes...
196	3003	-322	0.3	0.000214	0.000752	0.0335	10000000	29500000	29500000	file://skybridge...	file://skybridge/gscratch/mgpe...	file://skybridge/gscratch/mgpeter/punctureMes...
4591	4503	-322	0.3	0.000214	0.0007	0.03	10400000	29500000	29500000	file://skybridge...	file://skybridge/gscratch/mgpe...	file://skybridge/gscratch/mgpeter/punctureMes...

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.





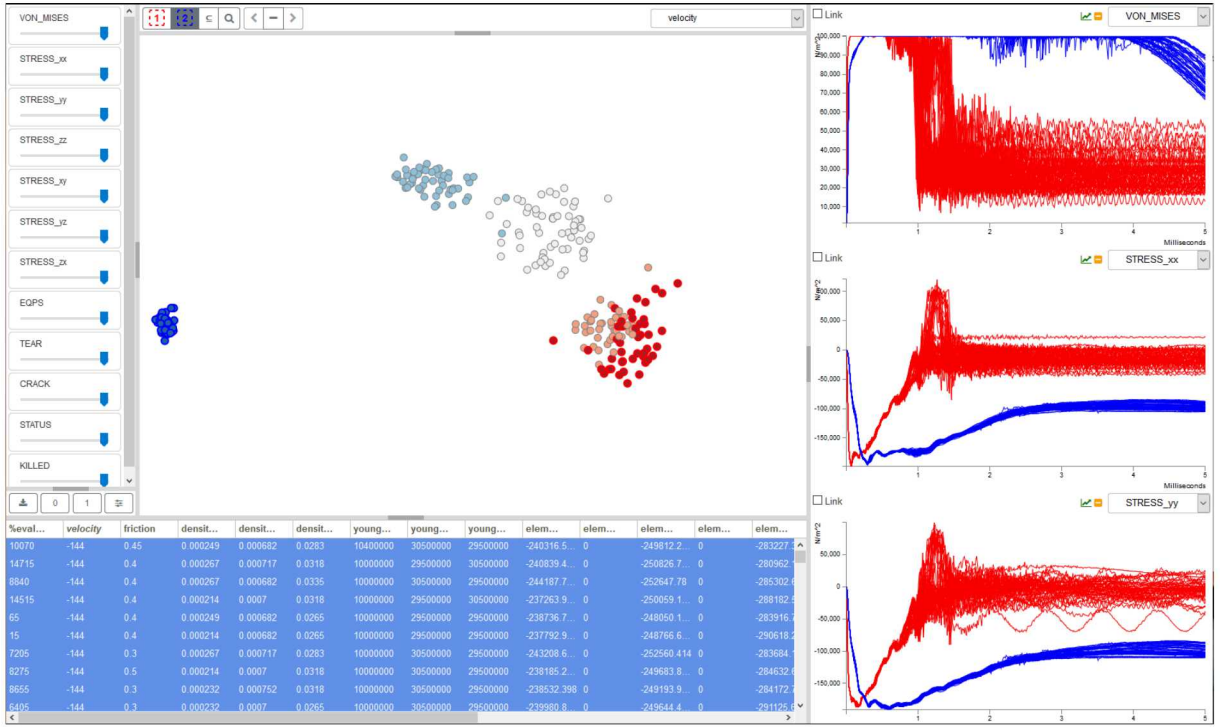
Cells die over time, allowing a plug of cells to be pushed out from the plate.



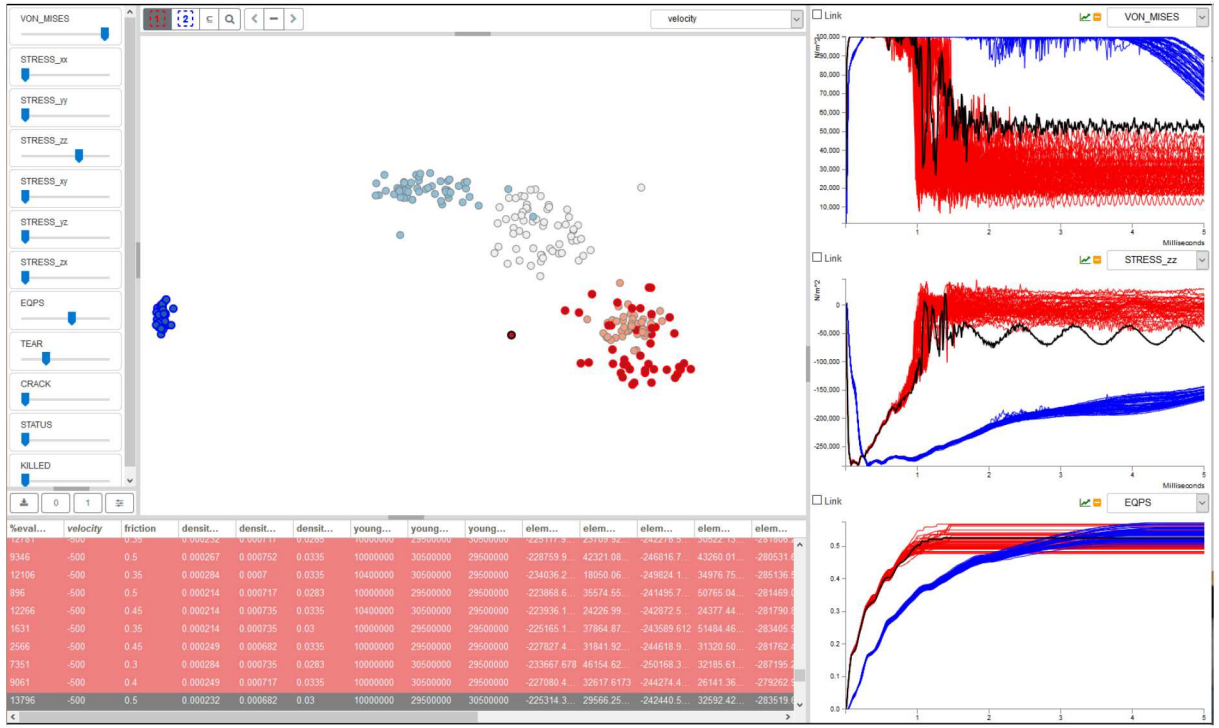
Dendrogram interactively controls plot visibility to compare selected run with cluster.

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.

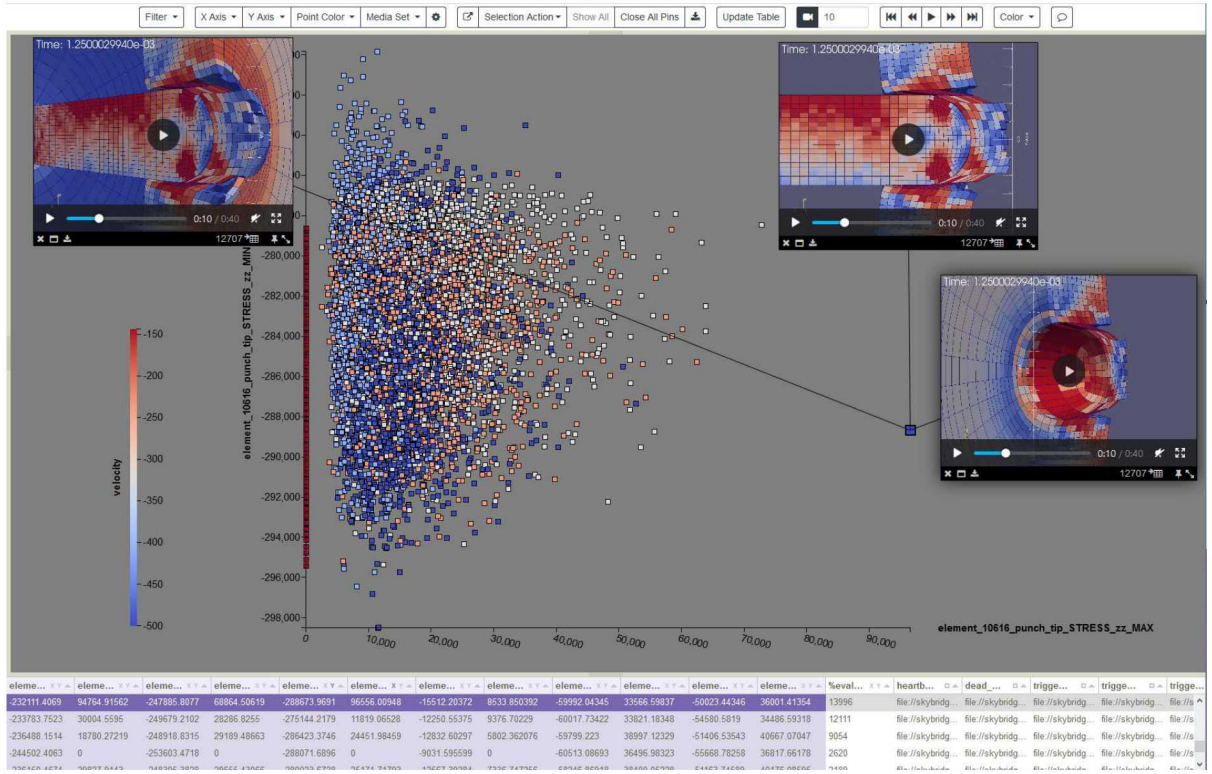


Time series for element 5768, the dimple element on the plate, after optimizing slider values against input velocity values. The 3 most significant temporal variables relative to input velocity are shown in the plots.

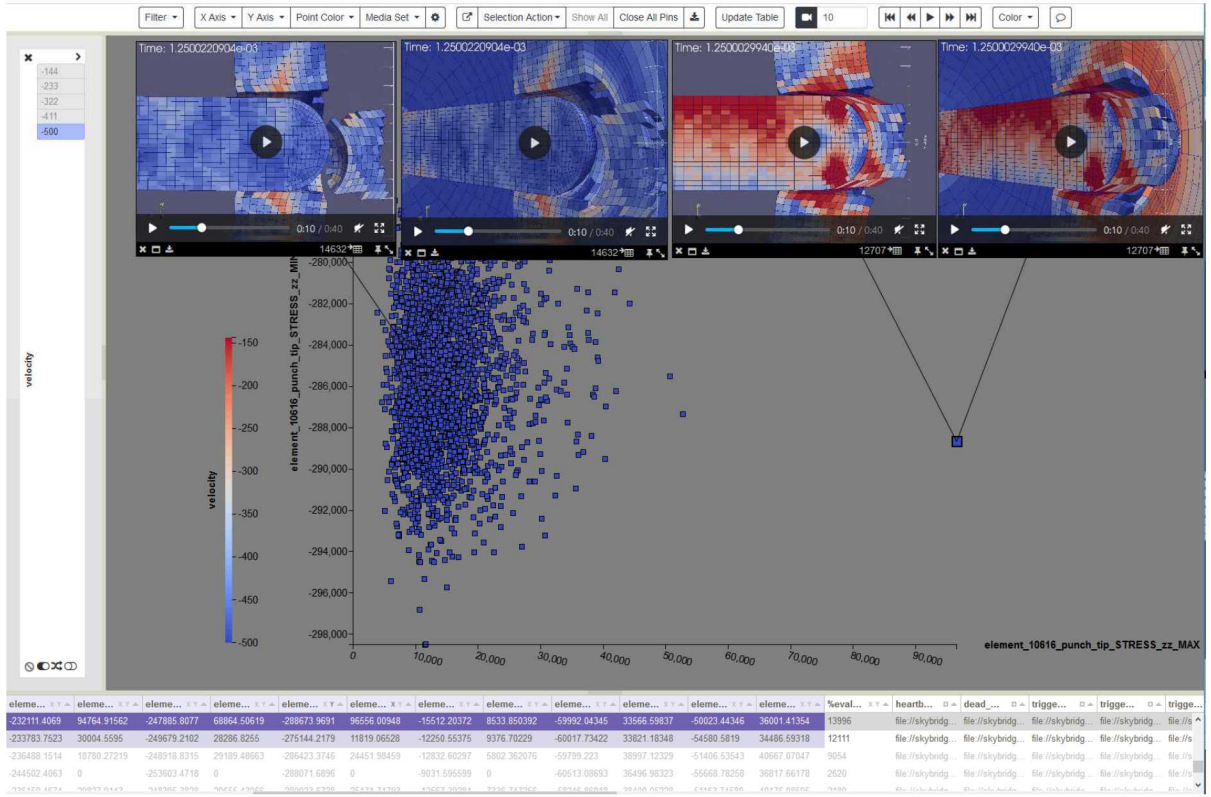
Martin, S., and Quach, T.-T., "Interactive Visualization of Multivariate Time Series Data," in *Human Computer Interaction International (HCII), Foundations of Augmented Cognition*: 322-332. 2016

Parameter Space Exploration Guides Media Viewing

- **Algorithm:** Visual pattern recognition by analyst
- **Analysis:** Parameter studies, multi-objective optimization, visual exploration, filtering, media retrieval and comparison
- **Data Types:** tables of scalar values, media (images, video, pdf, stl)



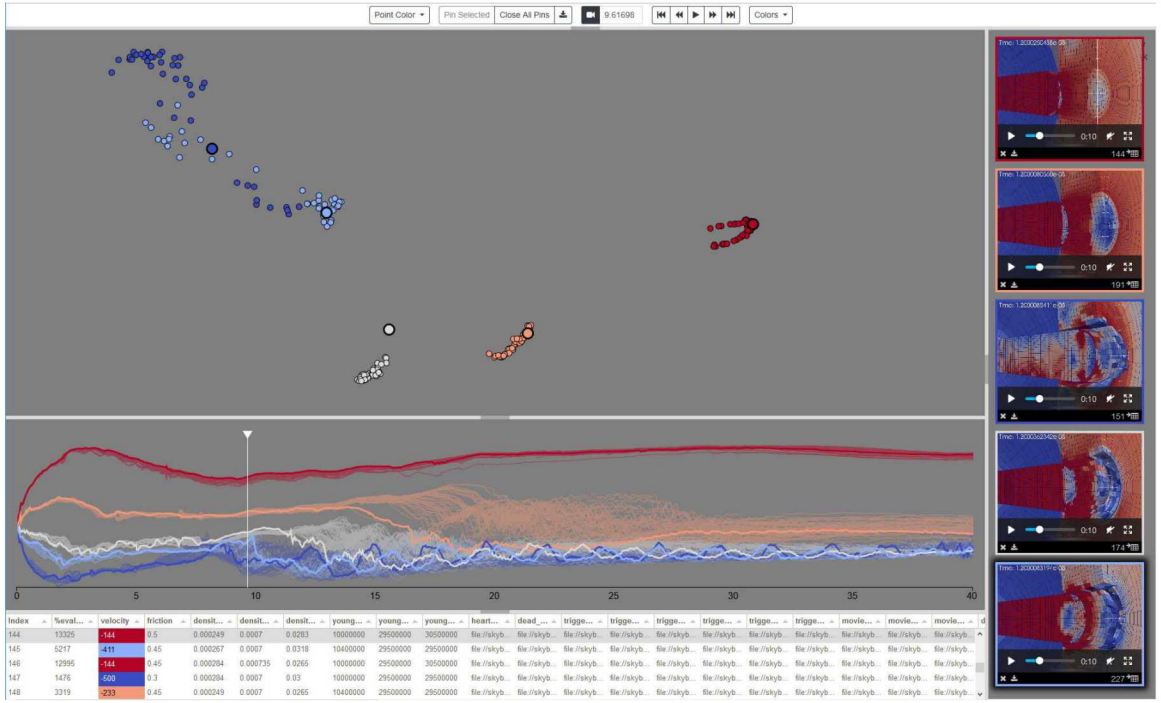
Stress outlier in 15K run ensemble guides retrieval of 3 different movie views, which can all be synchronously played.



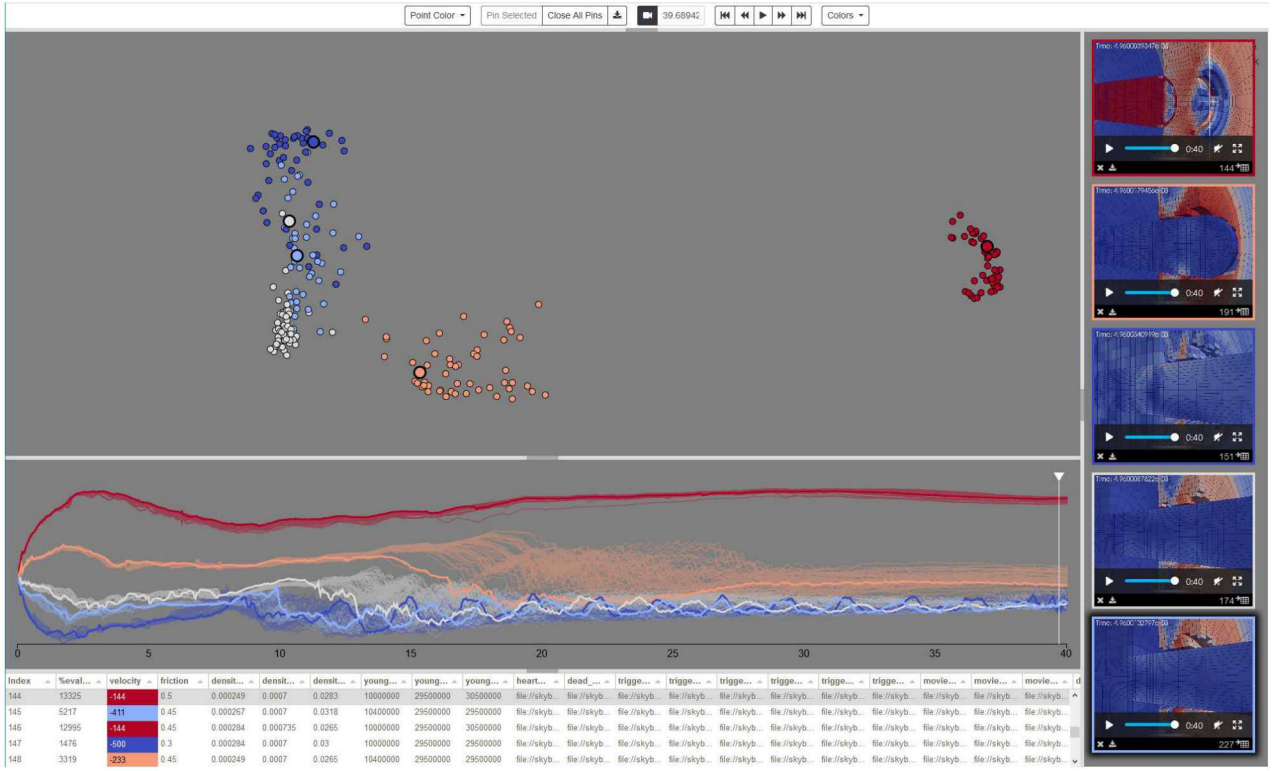
Scatterplot filtered to runs with same velocity value as outlier. Movies from a typical run compared with outlier's movies.

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.