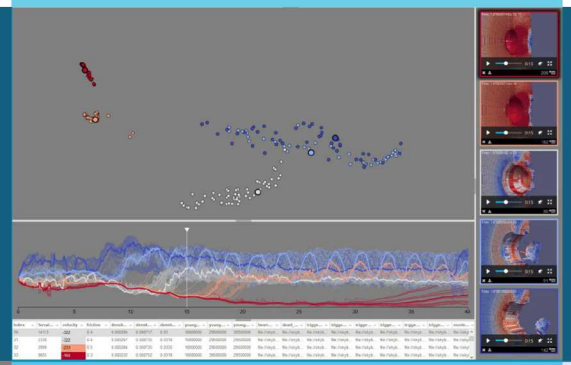
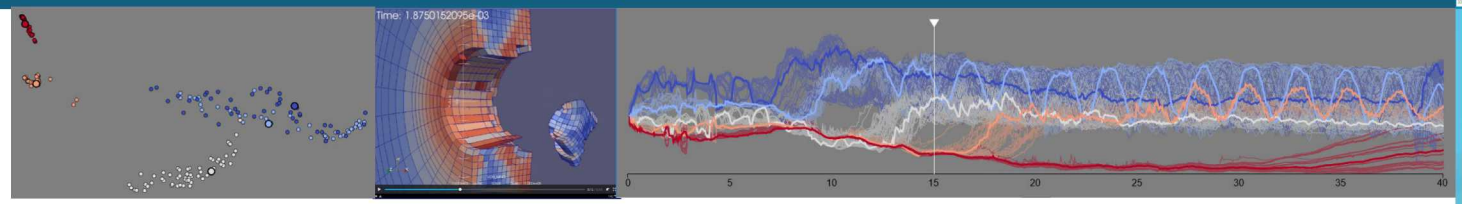
Sandia National Laboratories

SAND2019-9496PE

Slycat™ Ensemble Analysis and Visualization





PRESENTED BY

Patricia Crossno, Ph.D.

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

CIS External Review, August 26-29, 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.

Slycat™ Project Fills Sandia Mission Gap for Analysis & Visualization of Ensembles

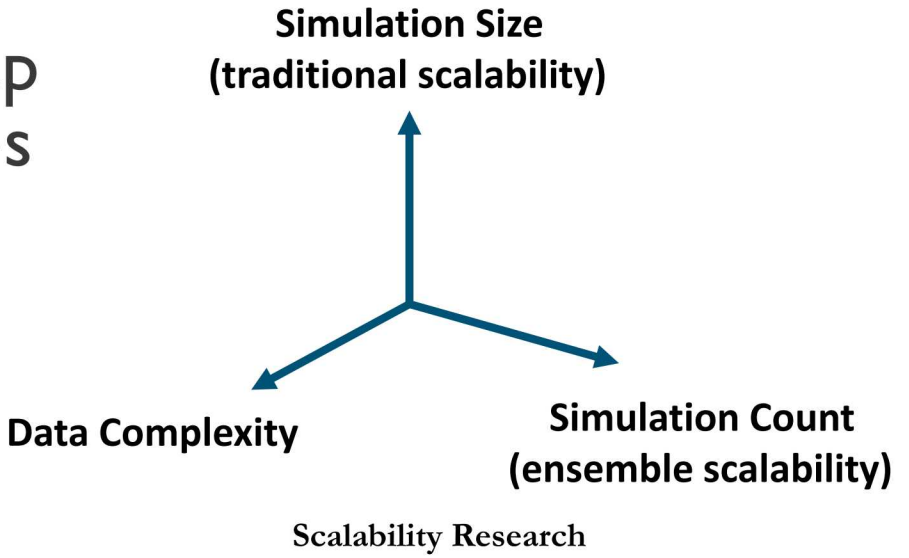
Slycat™ Project: 6 years, ~$1M/year (ASC), 6 people
- Open source https://github.com/sandialabs/slycat

Stockpile Stewardship Mission
- Uncertainty Quantification (UQ) critical
- Ensembles: UQ, model validation, sensitivity analysis, parameter studies

Origin: ParaText LDRD (text analysis)
- Vector space models => simulations
- **Understand content without viewing**

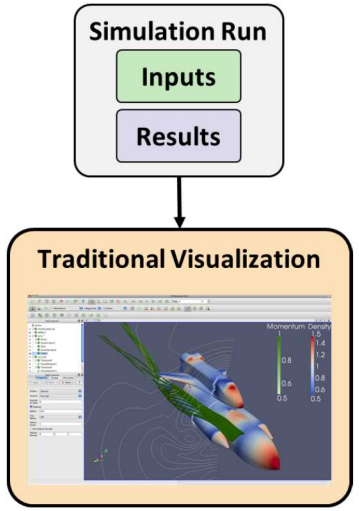
+Multimedia Data Challenges
- Created on Cluster
- View remotely*
- Map to input parameters*
- Compare*

Slycat™ Distinguishing Capabilities
- Scales in Simulation Count
- Media Support
- User Interface Design

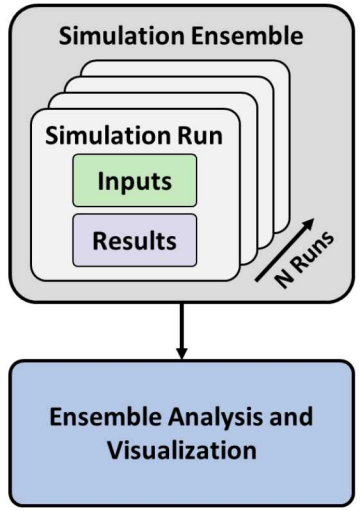


Scalability Research

Traditional tools not designed for ensembles



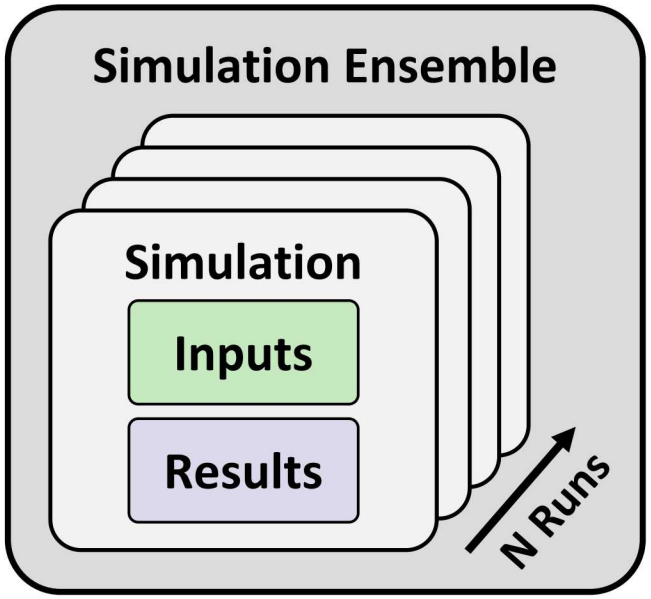
ParaView – visualization scales in simulation size



Ensemble – set of related simulation runs

* Previously, done manually (not scalable)

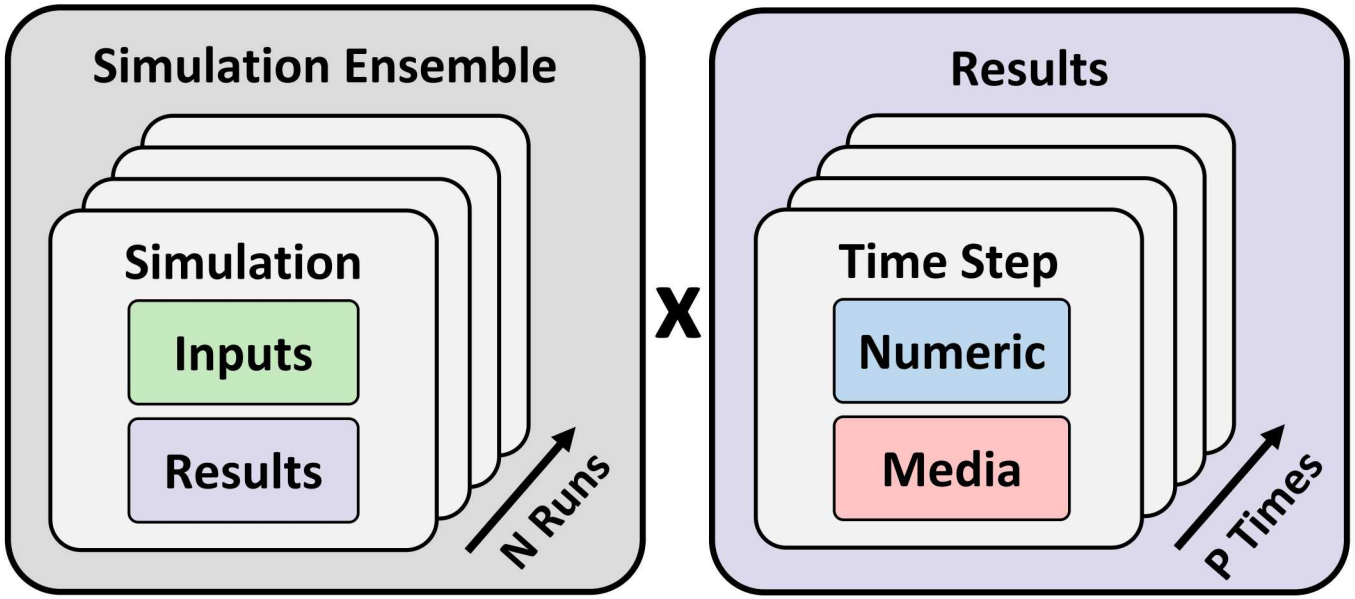
Ensemble Data is Challenging Because of Its Size & Complexity



500-1000 Runs
(Largest 50K)

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

Ensemble Data is Challenging Because of Its Size & Complexity



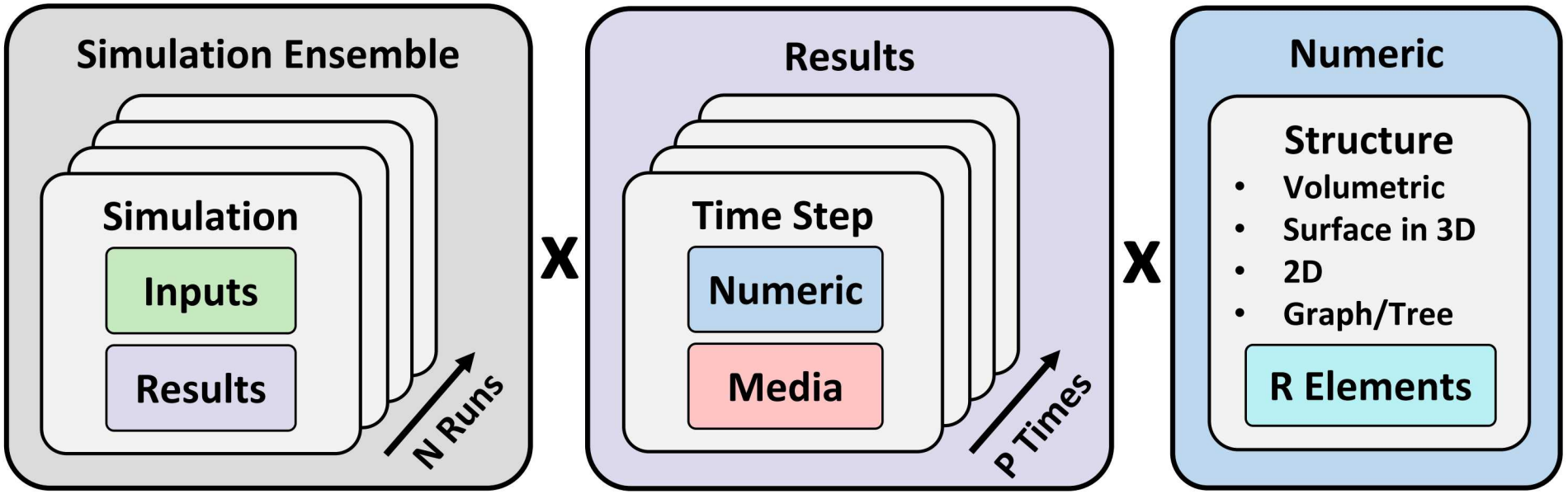
500-1000 Runs
(Largest 50K)

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

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

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

Ensemble Data is Challenging Because of Its Size & Complexity



500-1000 Runs
(Largest 50K)

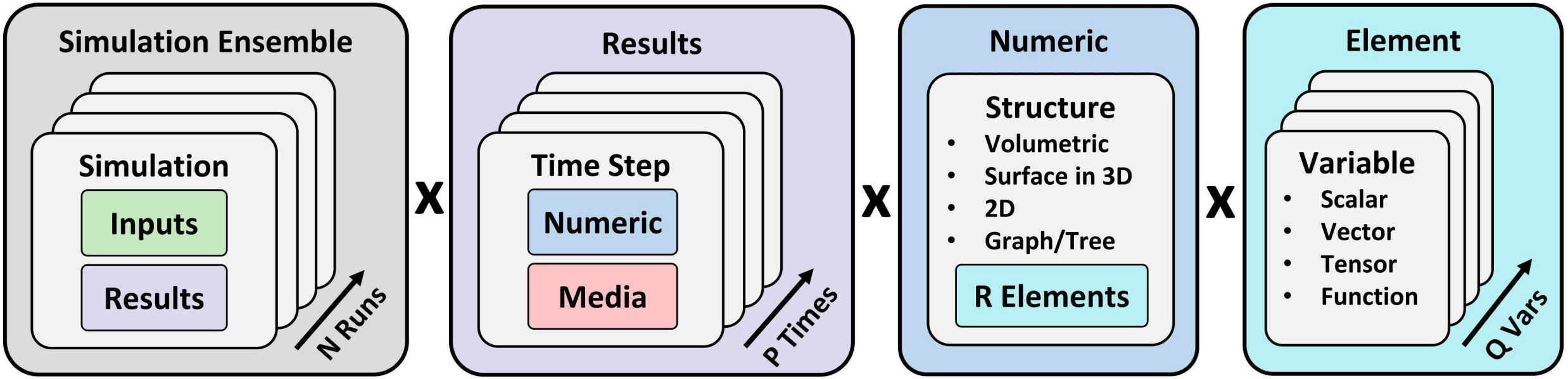
6 Histories (1M Timesteps)
1 Mesh (300 Timesteps)

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

1-5M Elements

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

Ensemble Data is Challenging Because of Its Size & Complexity



500-1000 Runs
(Largest 50K)

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

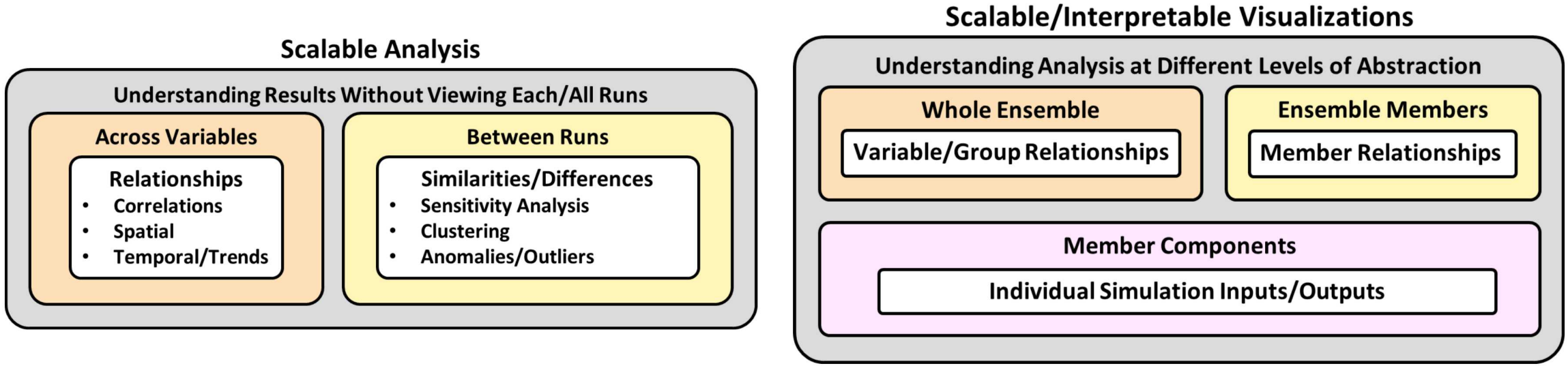
Media:
6-12 Movies (300 Frames)
24 Images
1 Pdf

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.)

Scalability and Interpretability are Challenges for Ensemble Analysis & Visualization



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™ Overview

Data Complexity

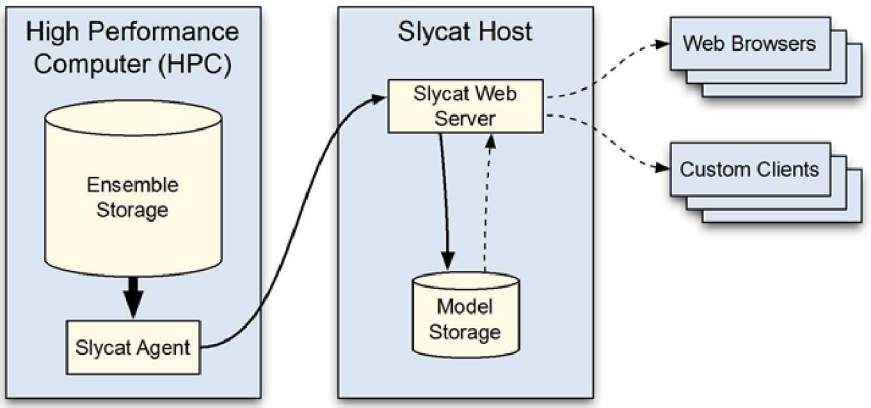
- Slycat™ models target specific result data types, encapsulating analysis results, visual representations, and interactions.

Scalability/Interpetability

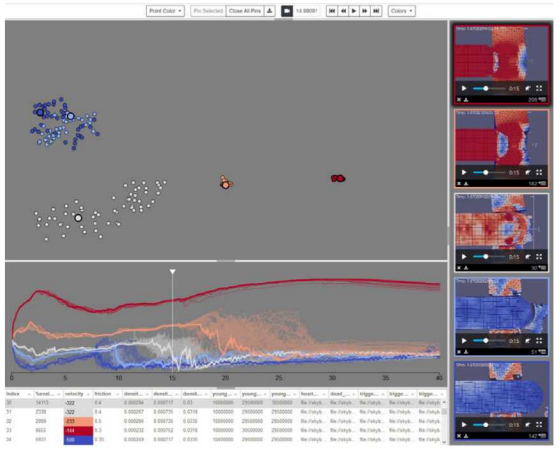
- Analysis done in parallel on ensemble data on HPC (Milestone 15K runs)
- Slycat™ server stores and manipulates reduced-scale models*
- Models provide multiple levels of abstraction: ensemble as a whole, individual runs, and raw data.

Usability

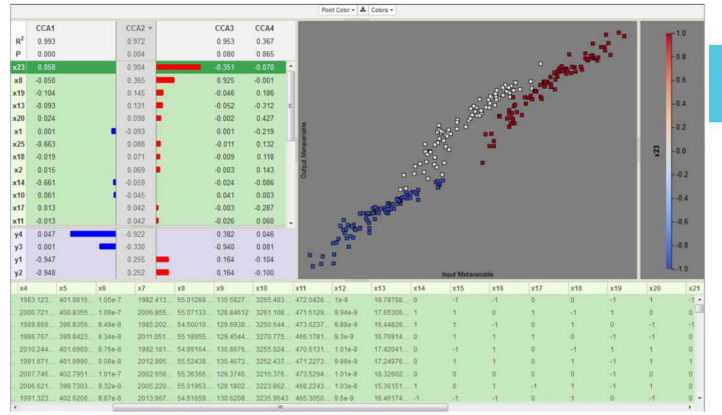
- Models visualized remotely, interactively, and collaboratively with desktop web-based interface



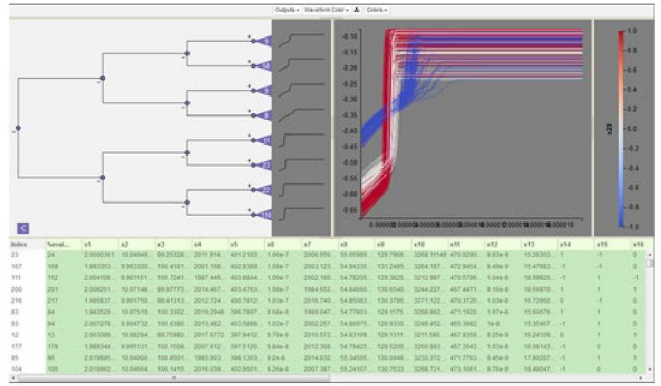
Slycat™ Architecture



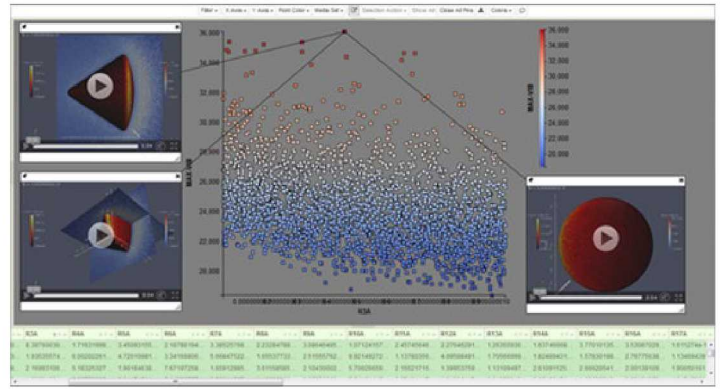
VideoSwarm Model



Canonical Correlation Analysis Model



Time Series Model

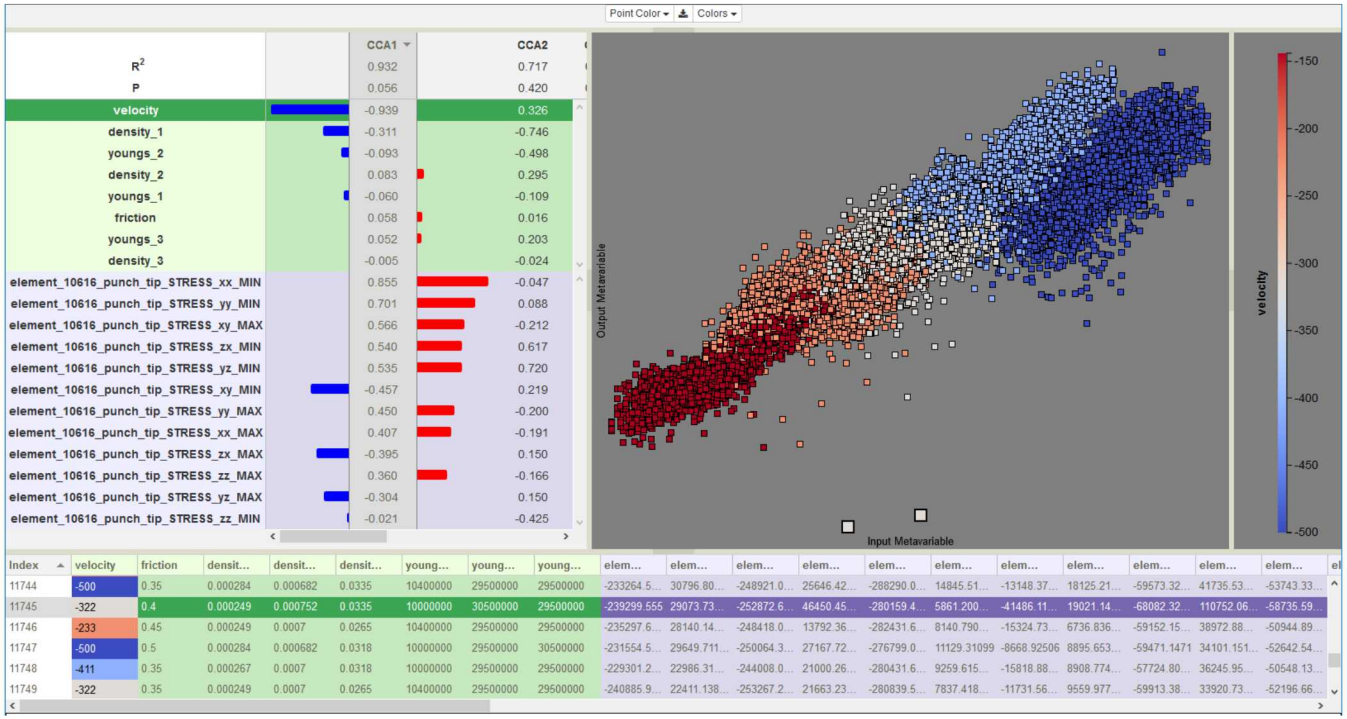


Parameter Space Model

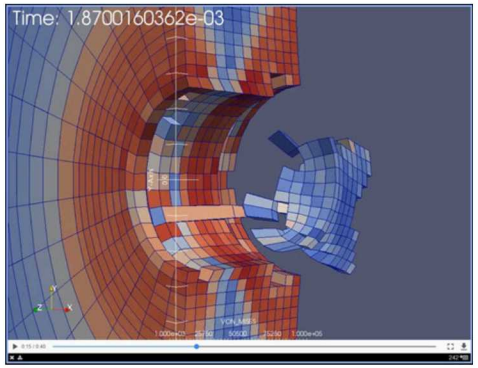
*Analytical models of results data, not physics models used by simulations

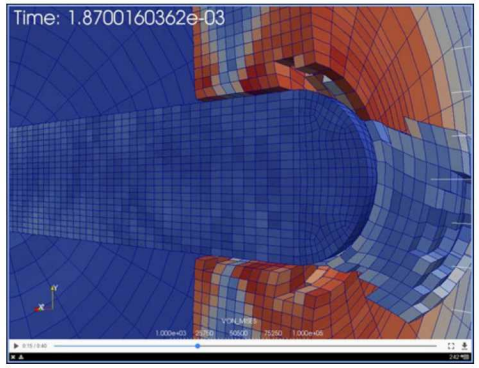
Identifying & Understanding Relationships Between Variables Is Important

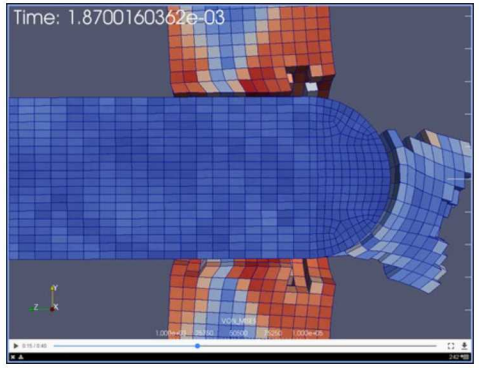
- **Algorithm:** Canonical Correlation Analysis (CCA) generalization of Principal Component Analysis (Hotelling 1936)
- **Analysis:** Correlations between 2 sets of variables, sensitivity analysis, anomaly detection
- **Data Type:** Tables of scalar values



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).







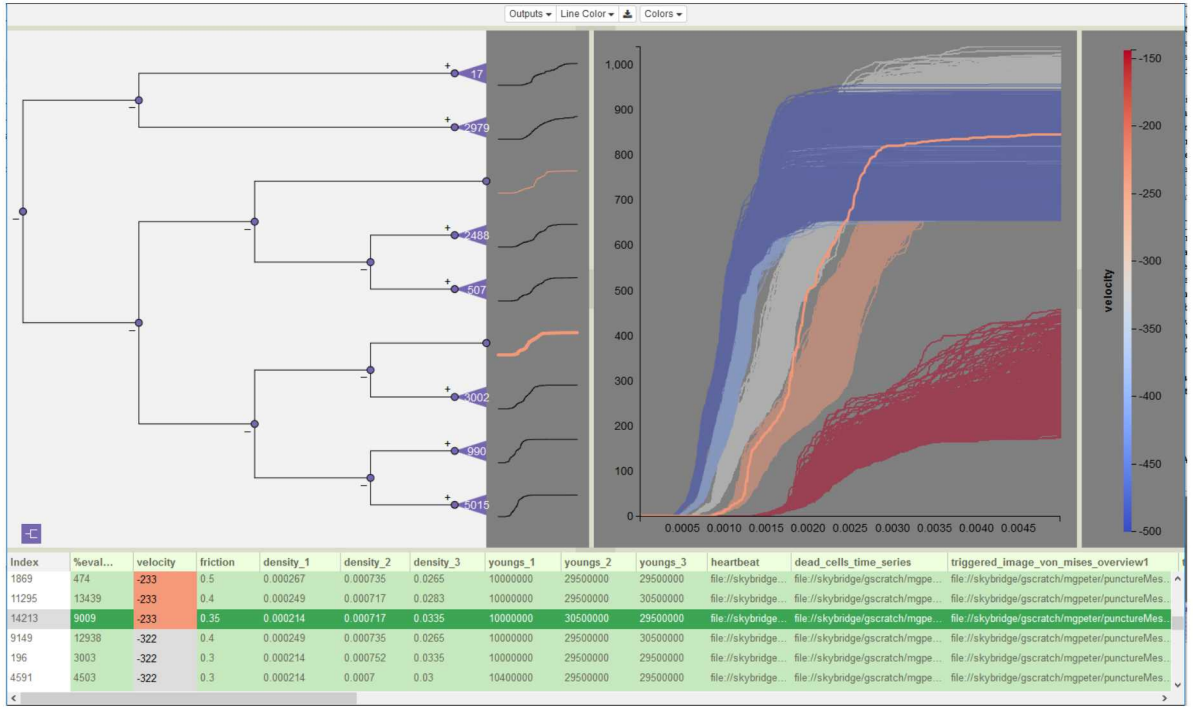
15K run ensemble for material fracturing problem, generated by Sierra/SolidMechanics and Catalyst (in situ image generation). Colored by Von Mises stress (red is high stress).

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

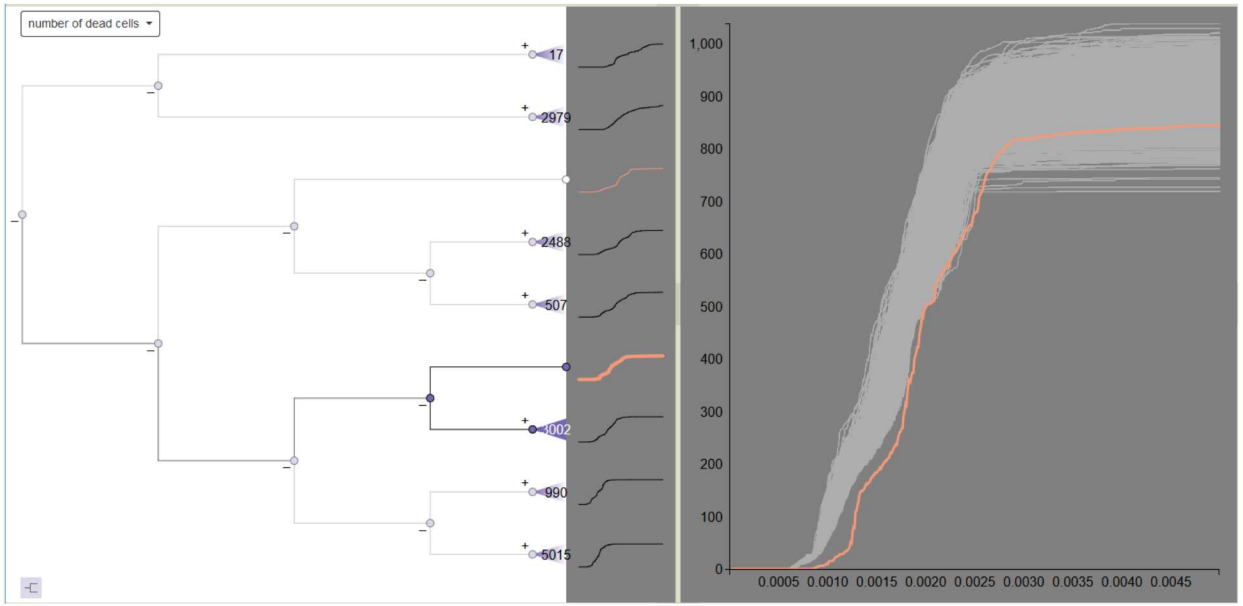
Time Series Clustering Reveals Relationships Between Inputs and Groups of Output Results

- **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.

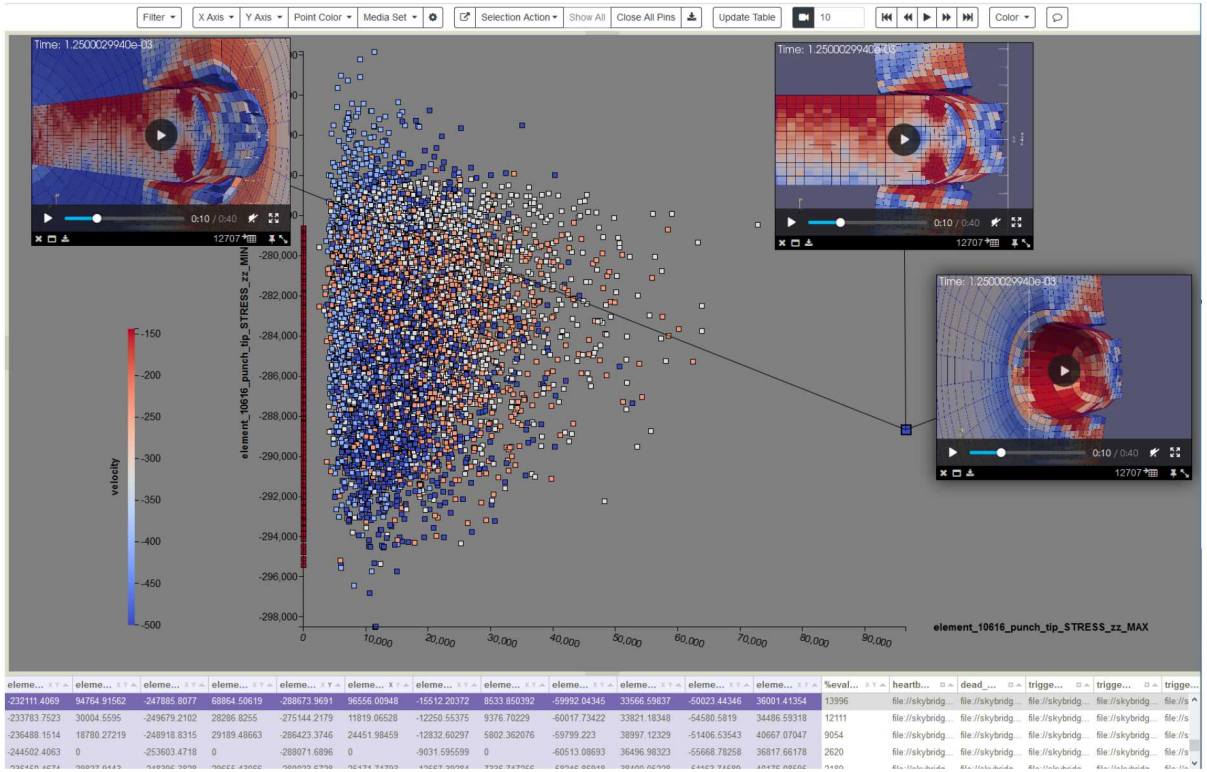


Dendrogram interactively controls plot visibility to compare selected anomalous run with cluster.

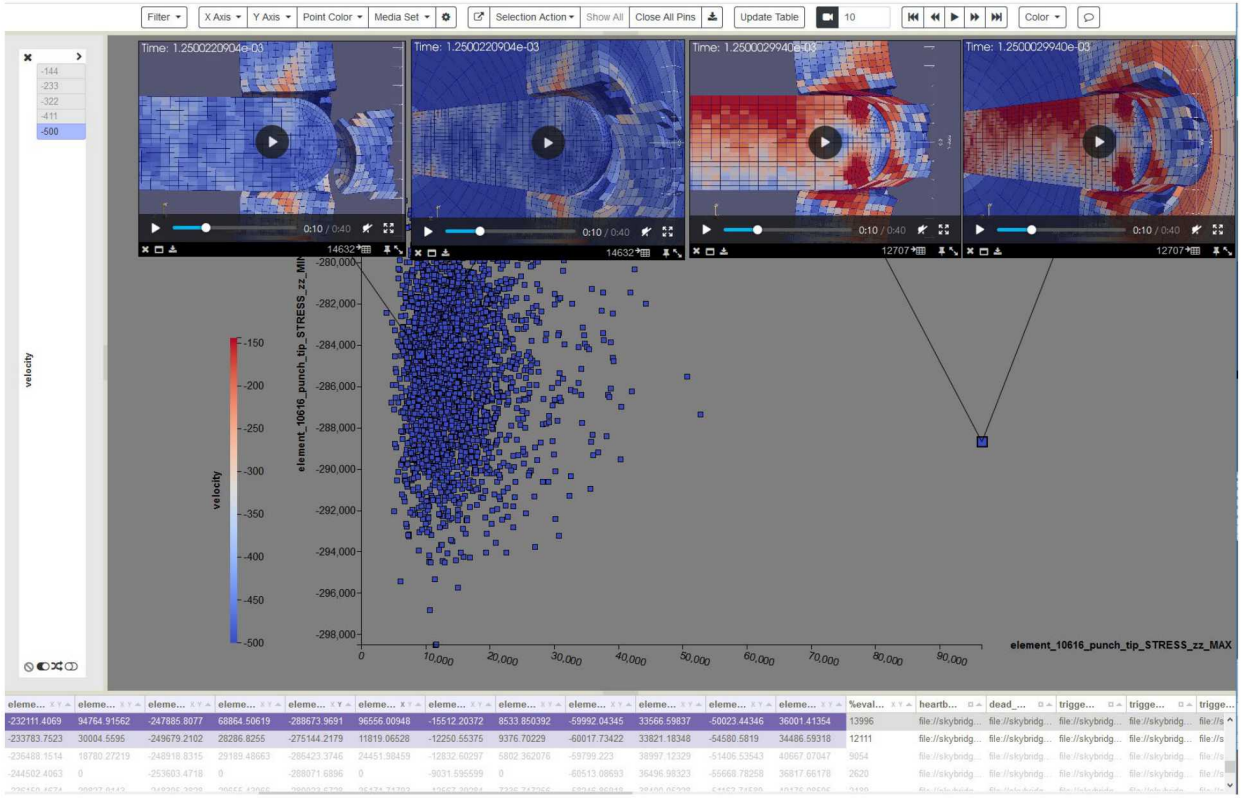
S. Singh, S. Zhang, W. A. Pruett, and R. Hester, "Ensemble Traces: Interactive Visualization of Ensemble Multivariate Time Series Data," in *Electronic Imagining, Visualization and Data Analysis*, pp. 1-9(9). 2016

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.

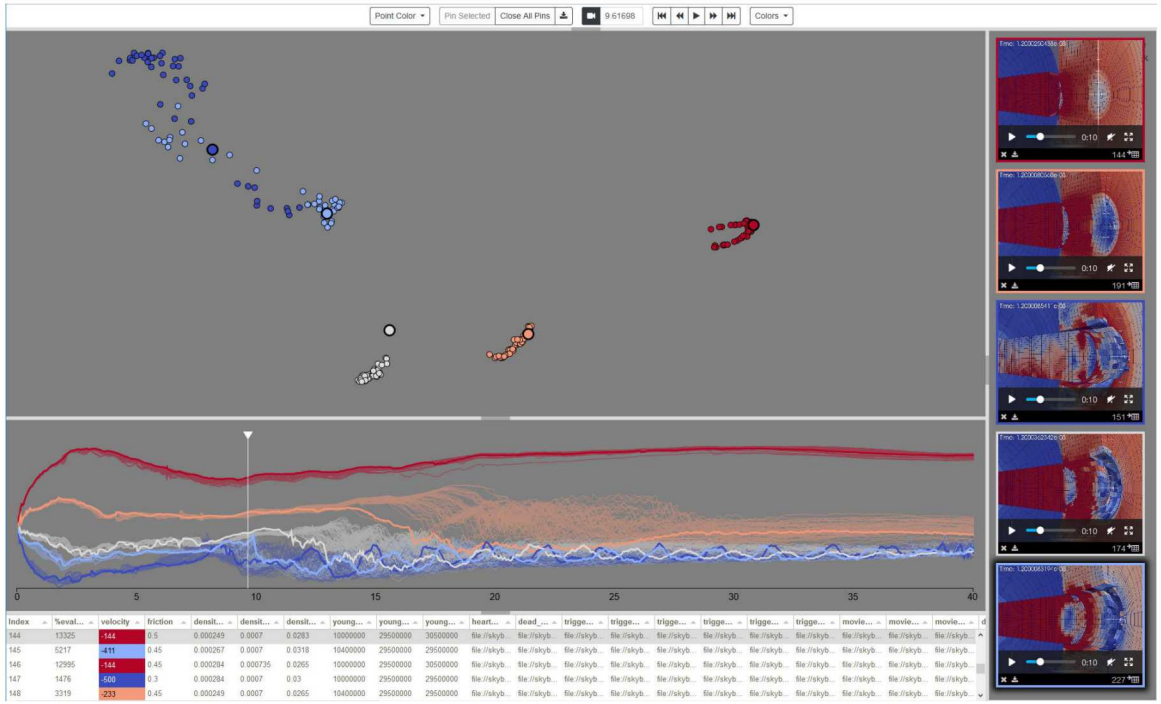


Scatterplot filtered to runs with same velocity value as outlier. Movies from a typical run compared with outlier's movies.

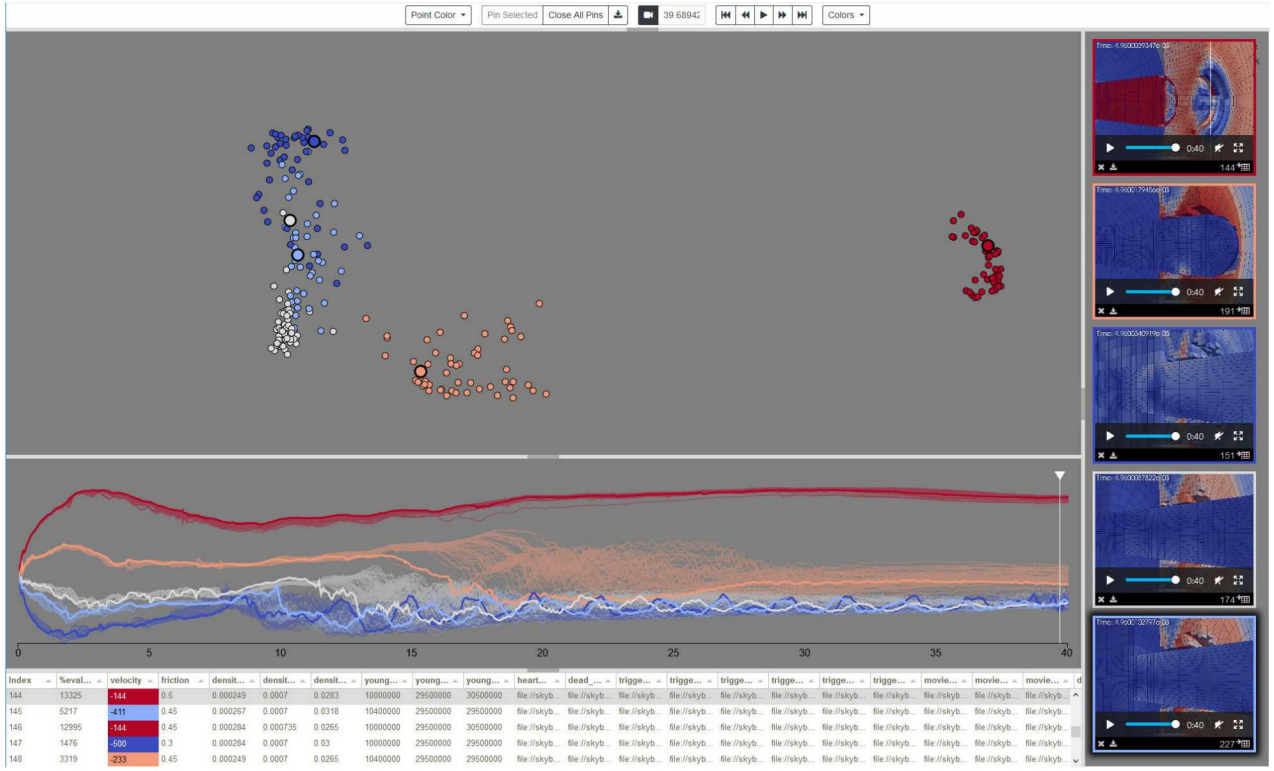
M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer and T. Möller, "Visual Parameter Space Analysis: A Conceptual Framework," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2161-2170, 31 Dec. 2014

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.

We Have Had Significant Impact on Ensemble Simulation Studies Both Within and External to Sandia

Russ Teeter (Component Science & Mechanics) on the **B61-12** Trajectory Strong Link model in abnormal shock environments:

> **"[Slycat's] type of access to model results is unprecedented and allows designers a much more intuitive way to understand complex results as well as find solutions to design problems."**

Similar comments in Supplemental Slides from:

George Orient (V&V, UQ, Credibility Processes) on shock physics analysis supporting **W80-4**

Dallas Johnson (Electrical Surety Analysis) for development, debug, & analysis of **W88 Alt** 370 FSS

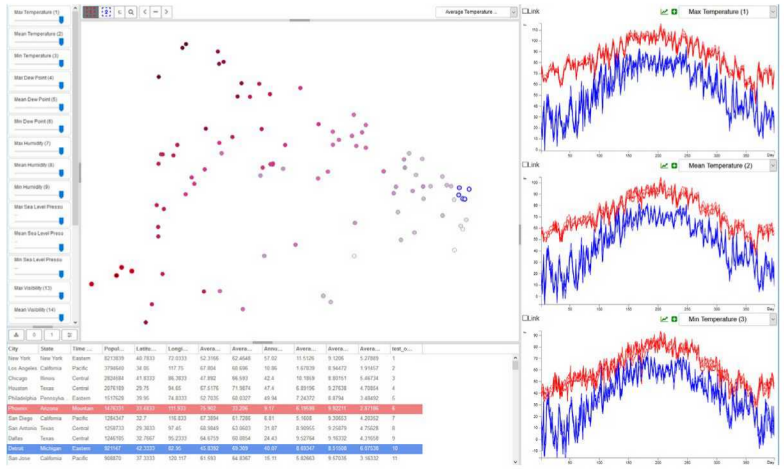
Peter Yeh (Computational Shock Physics) for **NNSA NA-80**

Jeremy Little (Monitor Intg & SWTB Cmpnt ENG) for test data

Switchtubes Product Realization Team:

> "We have just incorporated it into our acceptance process as a requirement to review data in the Dial-A-Cluster Model prior to stocking Switchtubes. Our department has over 6TB of waveform data that would only be referenced after acceptance if a failure was observed. Slycat has given us the ability to look at large amounts of data, without the need to look at each individual waveform."

Brady Aydelotte (**Army Research Lab**) for material modeling study comparing 2D and 3D simulations results with experimental data.



Dial-A-Cluster model: plug-in developed by Neutron Generator Group

Non-Sandia Scalable Ensemble Analysis Research Involving 1000+ Simulations Does Not Meet Our Mission Needs

Paper	Runs	SW Availability	Tightly Coupled	Media Support	Domain Examples
W. Berger (2011) VRVis, Austria	1000	License	No	No	Auto Design
A. Pretorius (2011) Univ. Leeds, UK	1024	No	No	Yes	Biomedical Image Processing
D. Coffey (2013) U. Minnesota, USA	1500	No	Yes	No	Medical Device Design
M. Luboschik (2014) U. Rostock, Germany	4096 - each is 50 reps avg'd	No - Research	Yes (steering)	No	Biology
Z. Konyha (2006) VRVis, Austria	4375	No	Yes	No	Fuel Injection
K. Matković (2018) VRVis, Austria	4373	No – ComVis Research	Yes (steering)	No	Diesel Engine Injector
P. Crossno (2018) SNL, USA	**15000**	**Yes**	**No**	**Yes**	**Physics, Electrical Circuit Design**

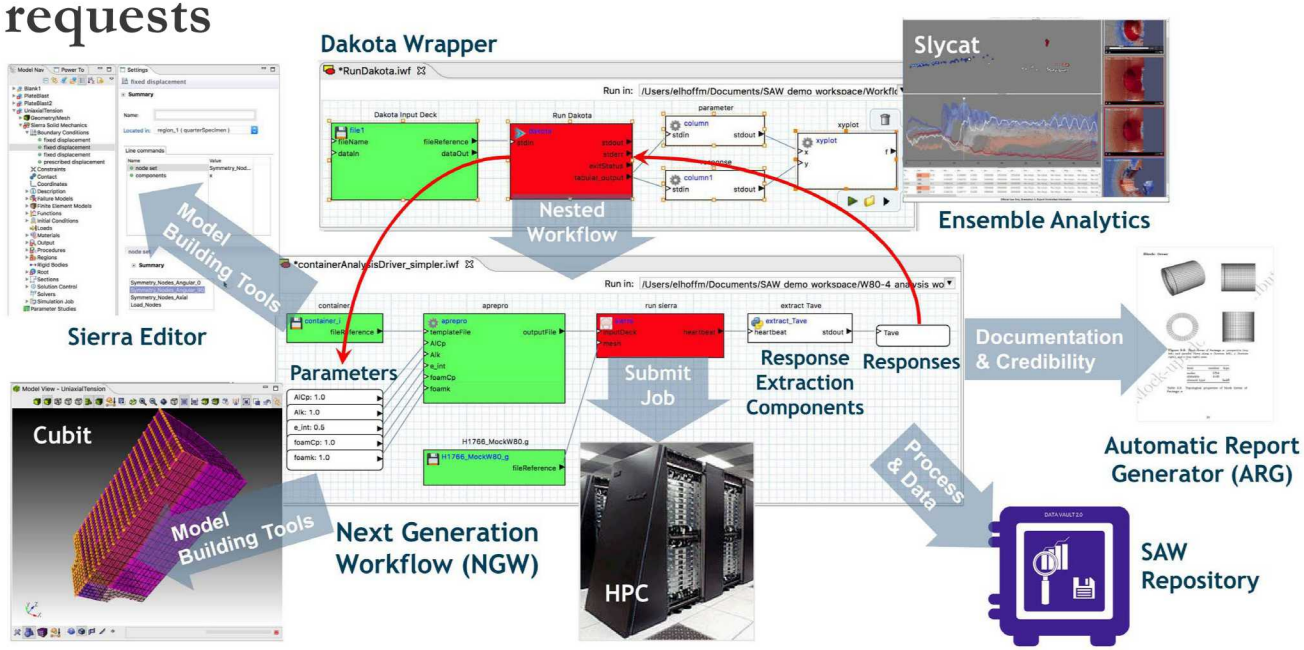
Expanding Use and Importance of Ensembles Drives Our Work in Data Visualization and Analysis

Long-term research needs

- Ensemble Analysis increasingly important in Nuclear Deterrence mission
  - Reduced design cycle initiative (12-year to 3-year)
  - Ensembles of reduced-order models (Patrick Blonigan talk)
- Experimental data analysis
- Machine Learning – exploring tensor methods, reduced-order models

Near-term plans in response to customer requests

- Integrated Workflow (Sandia Analysis Workbench, Cubit, Sierra, Dakota, Slycat™)
- Expand results data types
  - Viewer for surface mesh data
  - Synchronized interaction in multiple viewers
  - Web-based temporal surface mesh data management
- Add analysis methods
  - Comparison metrics for surfaces
  - Labeling support for machine learning



Integrated Workflow (IWF) to provide unsupervised Slycat™ model creation.

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.

More Information

Selected Publications

- P. Crossno, "Challenges in Visual Analysis of Ensembles," in *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 122-131, Mar./Apr. 2018. doi: 10.1109/MCG.2018.021951640
- 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,* J. Bennet, F. Vivodtzev, and V. Pascucci Eds., Springer-Verlag, 2015. doi: 10.1007/978-3-662-44900-4_16
- S. Martin, M. A. Sielicki, J. Gittinger, M. Letter, W. L. Hunt, P. J. Crossno, "VideoSwarm: Analyzing Video Ensembles", in *IS&T Electronic Imaging 2019,* pp. 685-1 – 685-11. doi: 10.2352/ISSN.2470-1173.2019.1.VDA-685

Benchmark References

- W. Berger, H. Piringer, P. Filzmoser, and E. Groller. "Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction." *Computer Graphics Forum*, 30(3):911 – 920, 2011. doi: 10.1111/j.1467-8659.2011.01940.x
- A. Pretorius, M.-A. Bray, A. Carpenter, and R. Ruddle. "Visualization of parameter space for image analysis." *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2402–2411, 2011. doi: 10.1109/TVCG.2011.253
- D. Coffey, C.-L. Lin, A. Erdman, and D. Keefe. "Design by dragging: An interface for creative forward and inverse design with simulation ensembles." *IEEE Trans. on Visualization and Computer Graphics*,19(12):2783–2791, 2013. doi: 10.1109/TVCG.2013.147
- M. Luboschik, S. Rybacki, F. Haack, and H.-J. Schulz. "Supporting the integrated visual analysis of input parameters and simulation trajectories." *Computers & Graphics*, 39(0):37 – 47, 2014. doi: 10.1016/j.cag.2013.09.004
- Z. Konyha, K. Matkovic, D. Gracanin, M. Jelovic, and H. Hauser. "Interactive visual analysis of families of function graphs." *IEEE Trans. on Visualization and Computer Graphics*, 12(6):1373–1385, 2006. doi: 10.1109/TVCG.2006.99
- K. Matković, D. Gračanin, and H. Hauser. "Visual analytics for simulation ensembles." In *Proceedings of the 2018 Winter Simulation Conference* (WSC '18), pp. 321-335, 2018. doi: 10.1109/wsc.2018.8632312

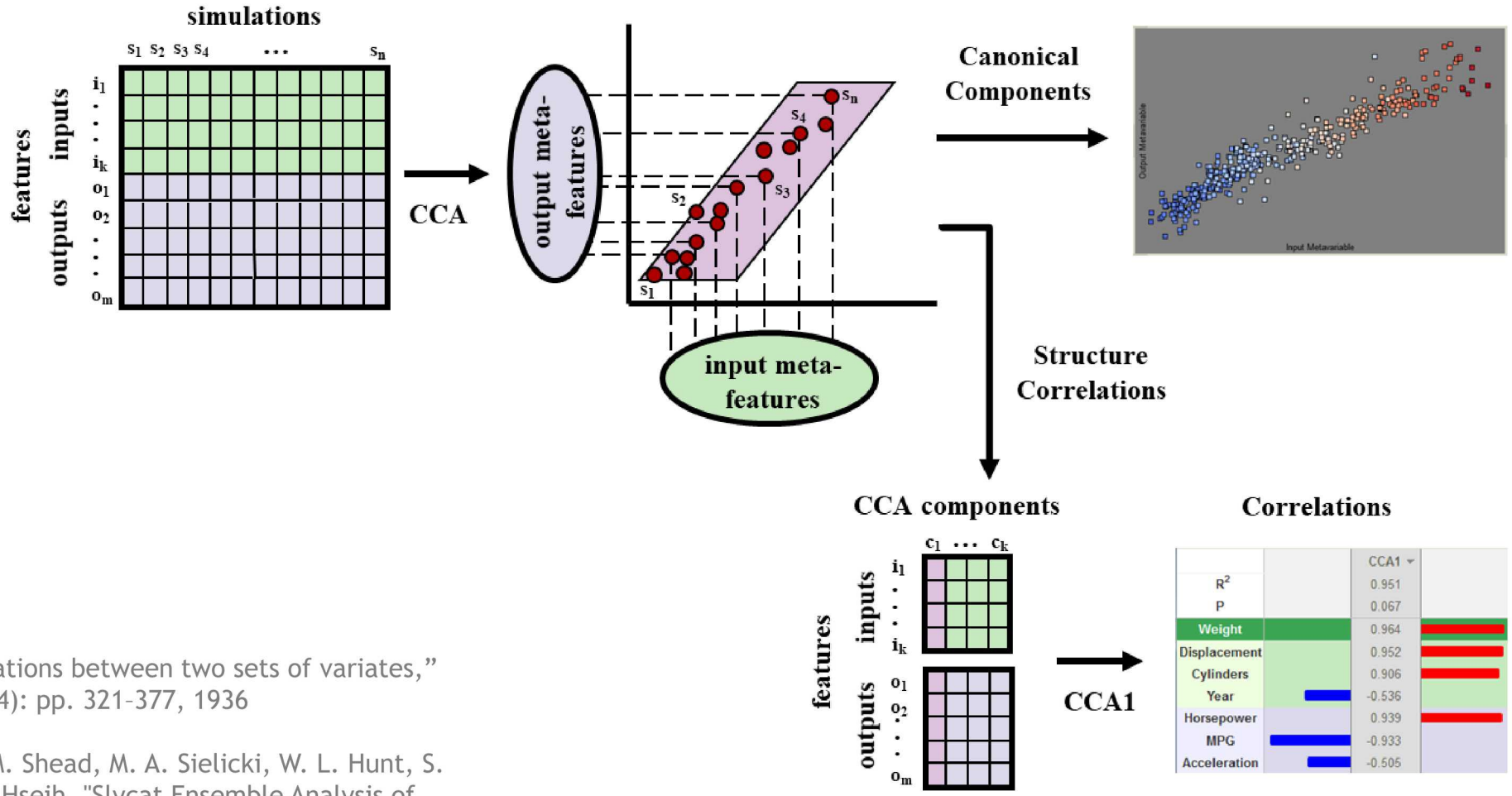
Additional References

- J. Wang, S. Hazarika, C. Li and H. Shen, "Visualization and Visual Analysis of Ensemble Data: A Survey," in *IEEE Transactions on Visualization and Computer Graphics,* July 2018. doi: 10.1109/TVCG.2018.2853721
- H. Hotelling "Relations between two sets of variates," Biometrika 28(3/4): pp. 321–377, 1936. doi: 10.2307/2333955
- S. Singh, S. Zhang, W. A. Pruett, and R. Hester, "Ensemble Traces: Interactive Visualization of Ensemble Multivariate Time Series Data," in *Electronic Imagining, Visualization and Data Analysis*, pp. 1-9(9). 2016. doi: 10.2352/ISSN.2470-1173.2016.1.VDA-505
- M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer and T. Möller, "Visual Parameter Space Analysis: A Conceptual Framework," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2161-2170, 31 Dec. 2014. doi: 10.1109/TVCG.2014.2346321
- 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. doi: 10.1109/TVCG.2015.2467553

Contact: Patricia Crossno: pjcross@sandia.gov

Supplemental Slides

How We Use Canonical Correlation Analysis Results to Drive Our CCA Model Views

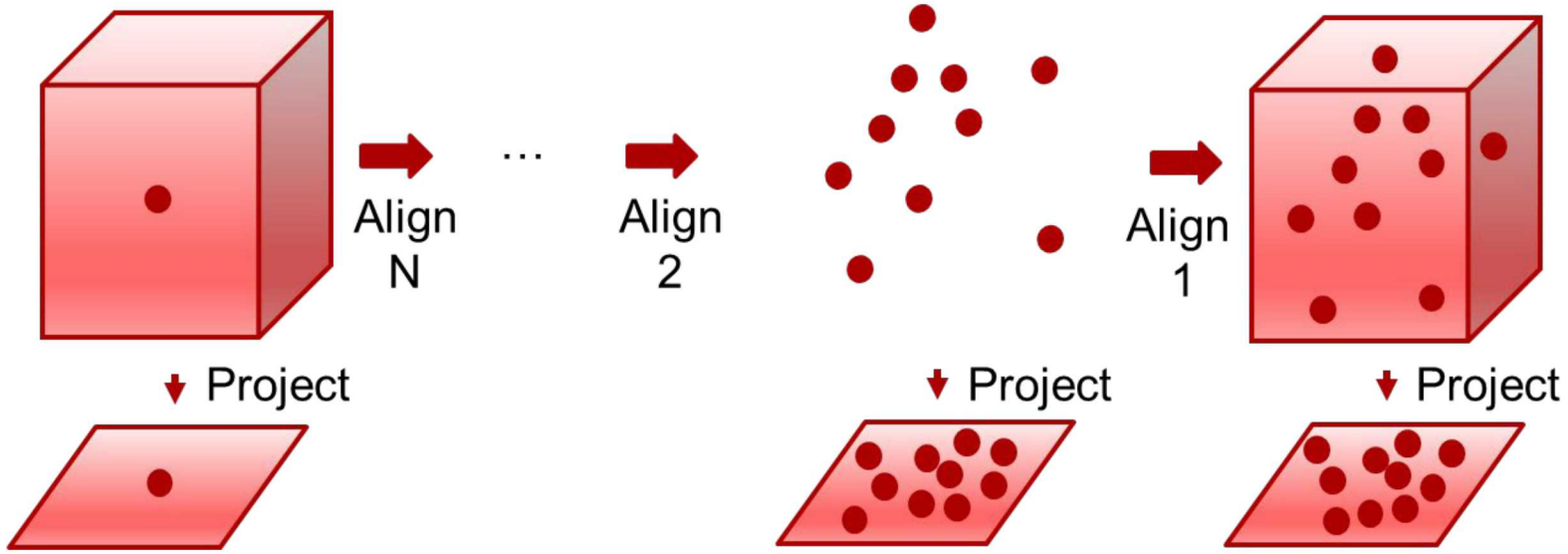


CCA results used in scatterplot (relationships between runs) and bar chart (relationships for ensemble-as-a-whole)

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

How VideoSwarm Does Temporal Alignment in High-Dimensional Space



MDS-derived coordinate axes differ between frames. VideoSwarm aligns coordinates in high dimensional space prior to projection, using the last frame as the basis (since in the first frame all points/runs typically are identical, hence co-located). Alignment proceeds by stepping through the frames in reverse order.

Shawn Martin, Milosz A. Sielicki, Jaxon Gittinger, Matthew Letter, Warren L. Hunt, and Patricia J. Crossno. "VideoSwarm: Analyzing Video Ensembles", *S&T International Symposium on Electronic Imaging 2019, Visualization and Data Analysis (VDA)* 2019.

Expanded User Feedback on Slycat™ Impact

Russ Teeter (Component Science & Mechanics)

Slycat was employed to visualize the vast amounts of simulation data generated from the Trajectory Strong Link (TSL) finite element model. The TSL model is used to understand the response of the various structures inside the TSL when it is subjected to its harsh abnormal shock environments. Each simulation (~100's) run in Sierra/SM produced 100's of QoI's including data points, time history plots, images, and videos. Using Slycat to organize and filter the data, designers were able to quickly drill down into simulations of interest, isolating worst case events, and understand responses which produced stressing situations. This **type of access to model results is unprecedented and allows designers a much more intuitive way to understand complex results as well as find solutions to design problems.**

George Orient (V&V, UQ, Credibility Processes)

Full system Solid Mechanics model is being used to examine impact scenarios to define LoI (Loss of Isolation) boundary in the velocity-impact angle mission space. LoI is determined by a hybrid quantitative/SME judgment based assessment of the evaluated UQ ensemble of models. The analysis workflow executed by Dakota creates a set of videos, x-y plots and parametric data for each instance. Slycat is an essential communication tool within the mod/sim team and with customers to collectively understand the behavior of the system to make credible statement about nuclear safety. The Reaper tool serves as a gateway to efficiently convert artifacts from the Dakota study to information Slycat can use. In addition to supporting decisions in a team environment these two tools are convenient to monitor a UQ study while it is being executed (total study run time: 14 days, generated data: 12 terabytes, 12 hrs., 3200 processors/Sierra SM instance)

Peter Yeh (Computational Shock Physics)

Slycat is an important tool not only for analysts but experimentalists as well. The centralized location and intuitive interface makes it easy to organize test data and rapidly summarize results to customers.

Expanded User Feedback on Slycat™ Impact

Dallas Johnson (Electrical Surety Analysis)

I have used Slycat for development, debug, and analysis of the W88 Alt 370 FSS model. Slycat has been a useful tool for understanding the simulation results, and how the components in the model affect its outputs. I was able to use Dakota and Slycat together to show my customers which components were strongly correlated to the outputs of interest, and provide visualization of how they affect the outputs. Slycat has also been an effective tool for debugging models that are problematic in the top level simulations, I used Slycat to debug simulations that had failed, and use its plotting tools to identify a component that was not working correctly in the top level simulation. This would have been difficult and time consuming to do otherwise.

Michael Haas (Mission Algorithms R&S)

Slycat has been instrumental in expediting the development, evaluation and delivery of a new multivariate analysis tool designed for use by NW component engineers and technicians. The analysis tool is a lightweight system for analyzing multivariate time series data and is called Dial-A-Cluster (DAC). The DAC analysis tool, running as a Slycat plug-in, is currently being used by select NW component engineers to enhance their data-based decision capabilities.

Jeremy Little (Monitor Intg & SWTB Cmpnt ENG)

Slycat has been useful. **We have just incorporated it into our acceptance process as a requirement to review data in the Dial-A-Cluster Model prior to stocking Switchtubes. Our department has over 6TB of waveform data that would only be referenced after acceptance if a failure was observed. Slycat has given us the ability to look at large amounts of data, without the need to look at each individual waveform.** The hope is that we will be able to catch any failures in advance of them happening, as well as catching any failures that the PT tester algorithm might miss.