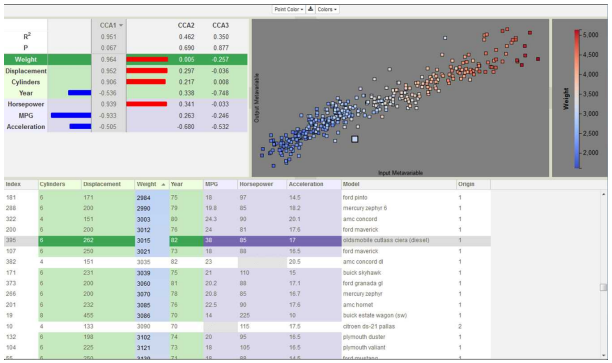
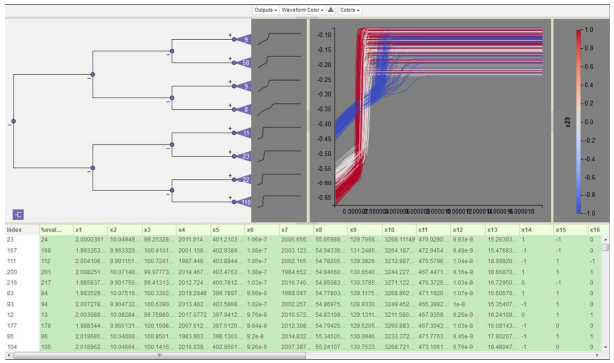
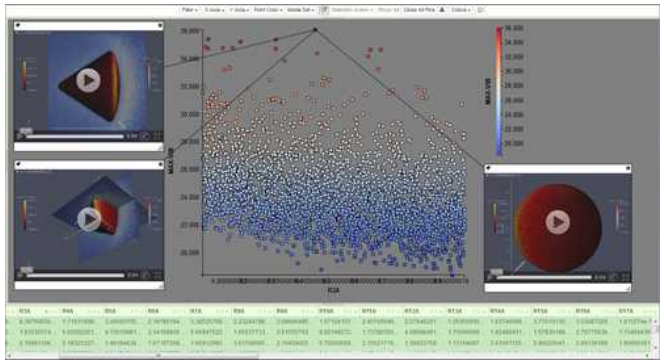
SAND2017-2767PE







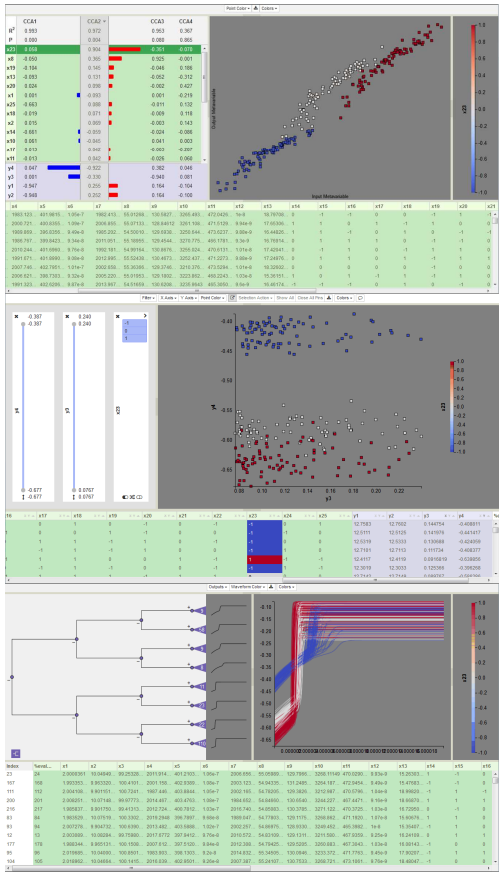
Slycat Tutorial

Patricia Crossno, PhD

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND2017-XXXX PE

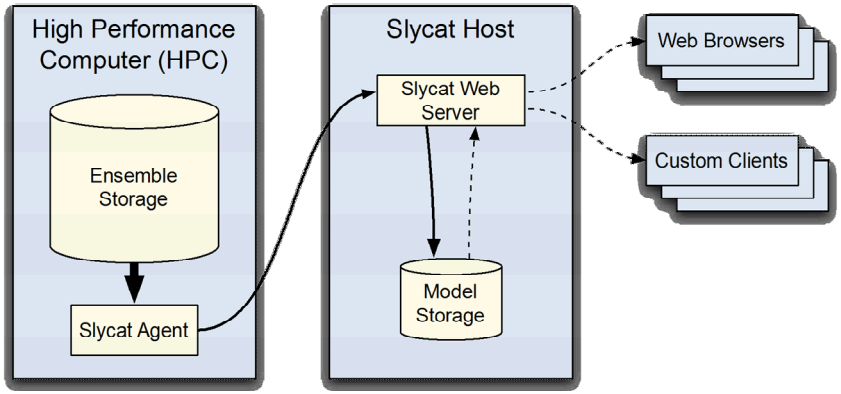
Tutorial Outline

- General Concepts
  - Slycat™ System
  - Projects
    - Navbars
    - Project-level Operations
  - Models
    - Data Tables
    - Abstractions & Encodings
    - Model-level Operations
- Models In Depth
  - Canonical Correlation Analysis
  - Parameter Space Model
  - Time Series Model



Slycat™ System

- Slycat™ is a web-based system for performing data analysis and visualization of potentially large quantities of remote, ensemble data.
- An ensemble is set of related simulation runs, each sampling a shared problem space.
- Slycat™ architecture



Project-based Structure

Sandia National Laboratories

- Projects contain models
  - Models encapsulate analysis results and data tables
  - Currently 3 model types (Correlation, Parameters, Time Series)
  - Multiple levels of abstraction
  - Interactive interfaces
- Access restricted to members
- Member roles:
  - Reader – view all data
  - Writer – create new models, modify models, view all data
  - Administrator – full read/write access to project (including deletion), add/subtract members

Getting Started

- Access Slycat™ through Firefox web browser
  - https://slycat.arl.hpc.mil/projects
- *Projects* list page Navbar

- *Models* list page within a project called *My Project*

- *Model* page for *My Model* within *My Project*

- Breadcrumbs for navigation -  return to *Projects*
- Functions in *Create*, *Edit*, *Info*, *Bookmarks*, and *Delete* differ relative to location

Projects

Sandia National Laboratories

- Project Creation
  - Name
  - Description
  - Default Administrator
- Project Edits
  - Membership
  - Cache
- Project Deletion
- Demonstration

Bookmarks

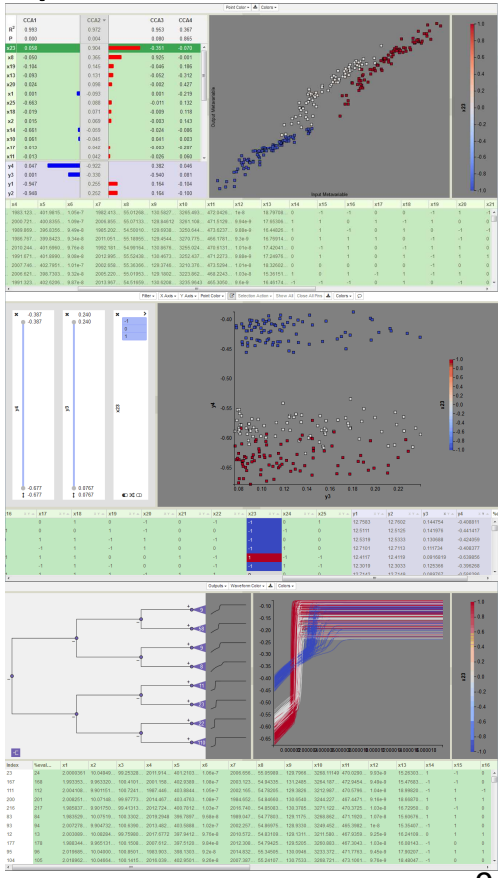
- Shared at project level (member only access)
- Store model state
  - Color-coding and selections
  - Model-specific information
    - Pinned images/movies and their sizes
    - Filters and their settings
    - Video settings (synch, shared time value)
    - Text Notes
  - Does NOT include layout information
- Slycat URL includes bookmark ID
  - Drag and drop into email
  - Back button in browser is NOT undo

Saved Bookmarks

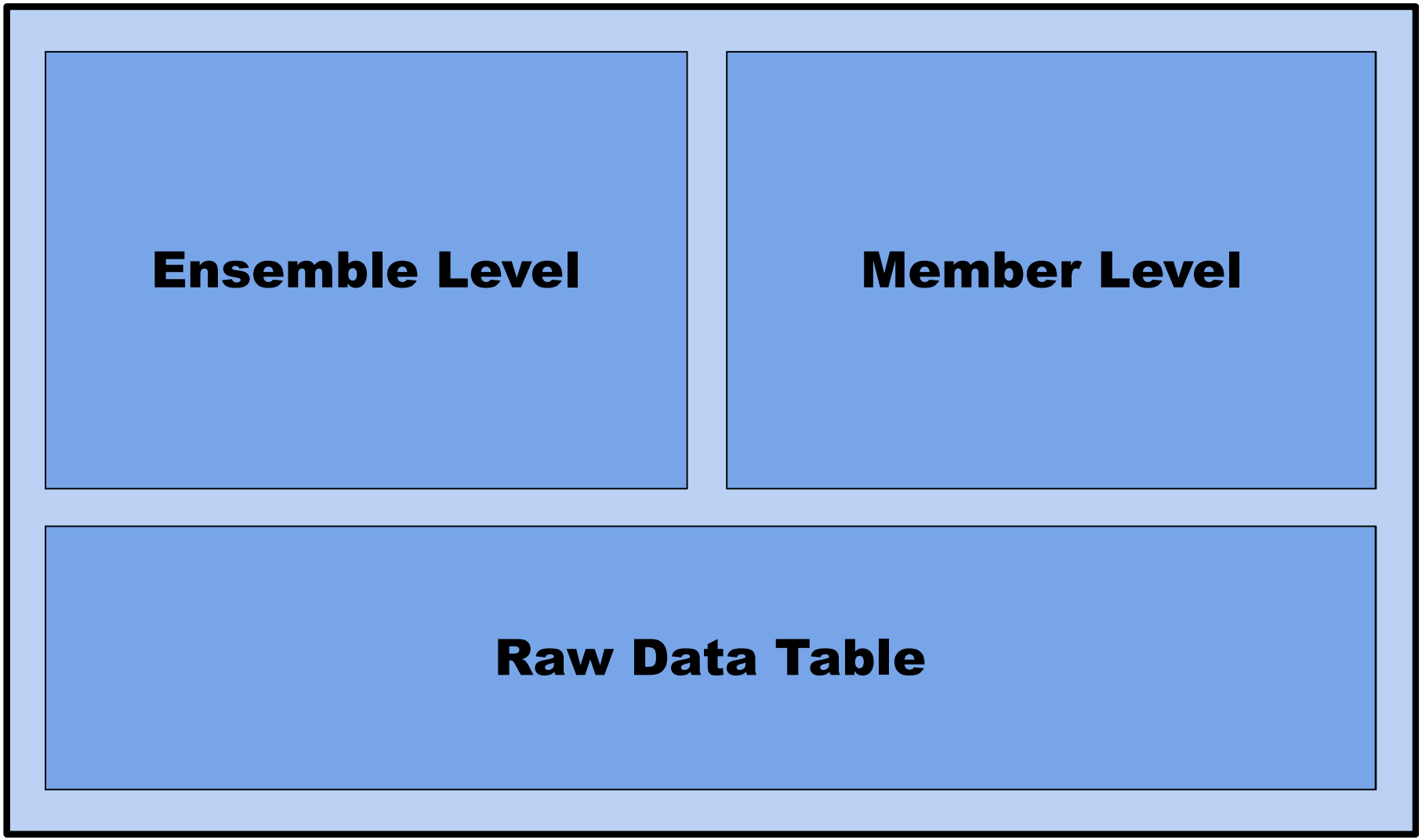
- Creation
- Retrieval
- Editing
- Deletion
- Demonstration in pre-existing model

Slycat Models

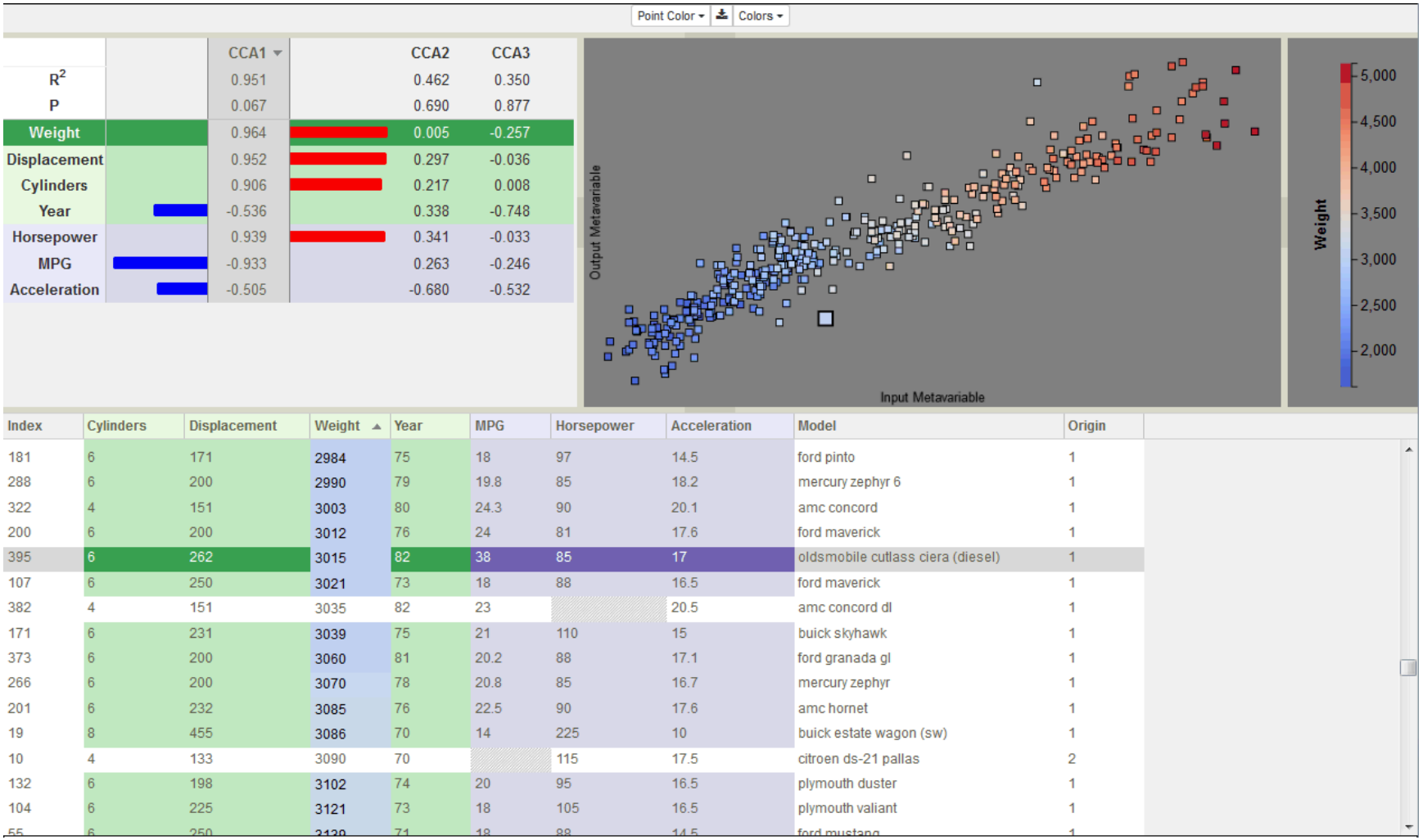
- Models of result data, not physics models used by simulations
- Different models provide complementary perspectives
- Each model has multiple levels of abstraction
- Main model types
  - Canonical Correlation Analysis (CCA)
    - Correlations between two sets of variables
    - Sensitivity analysis, anomaly detection
  - Parameter Space
    - Visual exploration, filtering, image/video retrieval
    - Parameter studies, multi-objective optimization
  - Time Series Clustering
    - Time series similarity, shape filtering
    - Map output variability to inputs, find outliers



Multiple Levels of Abstraction



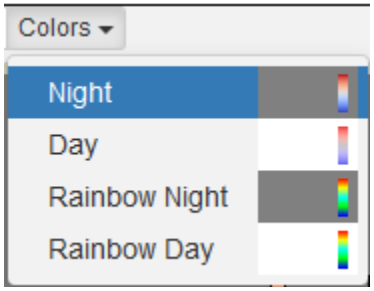
View Panels

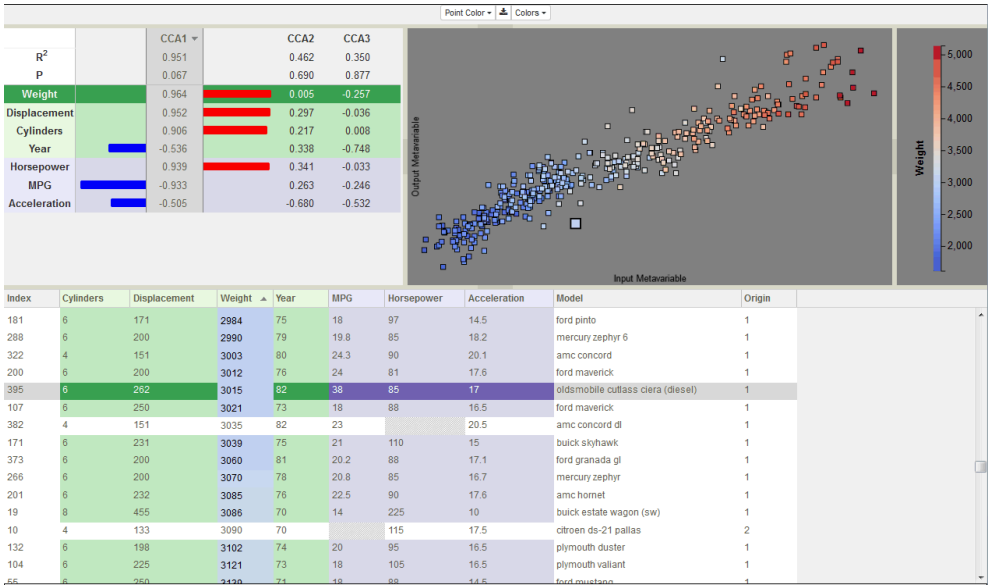


Index	Cylinders	Displacement	Weight	Year	MPG	Horsepower	Acceleration	Model	Origin
181	6	171	2984	75	18	97	14.5	ford pinto	1
288	6	200	2990	79	19.8	85	18.2	mercury zephyr 6	1
322	4	151	3003	80	24.3	90	20.1	amc concord	1
200	6	200	3012	76	24	81	17.6	ford maverick	1
395	6	262	3015	82	38	85	17	oldsmobile cutlass ciera (diesel)	1
107	6	250	3021	73	18	88	16.5	ford maverick	1
382	4	151	3035	82	23		20.5	amc concord dl	1
171	6	231	3039	75	21	110	15	buick skyhawk	1
373	6	200	3060	81	20.2	88	17.1	ford granada gl	1
266	6	200	3070	78	20.8	85	16.7	mercury zephyr	1
201	6	232	3085	76	22.5	90	17.6	amc hornet	1
19	8	455	3086	70	14	225	10	buick estate wagon (sw)	1
10	4	133	3090	70		115	17.5	citroen ds-21 pallas	2
132	6	198	3102	74	20	95	16.5	plymouth duster	1
104	6	225	3121	73	18	105	16.5	plymouth valiant	1
55	6	250	3139	71	18	88	14.5	ford mustang	1

Colors in Models

- Color Themes
  - Night
  - Day
  - Rainbow Night
  - Rainbow Day
- Color-coding
  - Input variables – green
  - Output variables – purple
  - Undefined/neither – white
- Red = High Values
- Blue = Low Values
- Legends





Data Tables

- Data table is core of all model types
  - dakota_tabular.dat
  - CSV
- Rows = simulation runs
- Columns = variables
- Media variable URIs point to images/movies/time series/STL
  file://machine_name/absolute_directory_path/filename.ext
- Movies must be generated using ffmpeg and h264 codec
- Each time series is also a data table
  - CSV
  - Xyce generated .prn files
- Data tables can be downloaded from models 

Model Creation

- Wizards
- Data Location
  - Local
  - Remote
  - Split (local table, remote media)
- Processing
  - Serial – CCA and Parameter Space
  - Parallel – Time Series (all data on cluster)

Canonical Correlation Analysis

Sandia National Laboratories

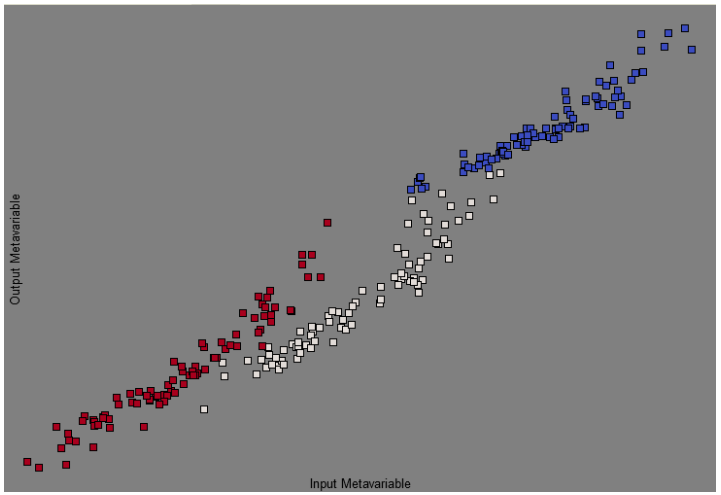
- Correlation between 2 multivariate data sets
- Generalization of PCA developed by Hotelling in 1936
- CCA Data Requirements
  - Row count > min (#input vars, #output vars)
  - Must have at least 1 input and 1 output var
  - No missing data, Inf, NAN, or NULL values (we remove rows)
  - No columns with constant values
  - Only numeric variables (no strings) -> ordered variables can be converted to numeric values
- Demonstrate CCA Wizard

CCA Visualization – Ensemble Level

- Correlation View
  - CCA Components
    - R^2, P
    - Column selection expansion
    - Column selection linking
  - Variables
    - Bar-chart, sorting
    - Color-coding
    - Row selection linking

	CCA1	CCA2	CCA3	CCA4
R^2	0.993	0.972	0.953	0.367
P	0.000	0.004	0.080	0.865
x23	-0.058	-0.904	-0.351	-0.070
x8	0.050	-0.365	0.925	-0.001
x19	0.104	-0.145	-0.046	0.186
x13	0.093	-0.131	-0.052	-0.312
x20	-0.024	-0.098	-0.002	0.427
x1	-0.001	0.093	0.001	-0.219
x25	0.663	-0.088	-0.011	0.132
x18	0.019	-0.071	-0.009	0.118
x2	-0.015	-0.069	-0.003	0.143
x14	0.661	0.059	-0.024	-0.086
x10	-0.061	0.045	0.041	0.003
x17	-0.013	-0.042	-0.003	-0.287
x11	0.013	-0.042	-0.026	0.060
x22	0.268	0.041	-0.001	0.105
y4	-0.047	0.922	0.382	0.046
y3	-0.001	0.330	-0.940	0.081
y1	0.947	-0.255	0.164	-0.104
y2	0.948	-0.252	0.164	-0.100

CCA Visualization – Member Level



- Simulation View
  - Points = Runs
  - Axes = Metavariables
  - Perfect correlation = diagonal line
  - Anomalies = distant from diagonal
- Color-coding
  - 3 ways to set:
    - *Point Color* drop down
    - Row (variable) selection in bar chart
    - Column (variable) selection in table
  - Colors determined by Color Theme
- Selections linked to table
  - 2 ways to set: click (single) or rubberband (multiple)

CCA Visualization – Data Level

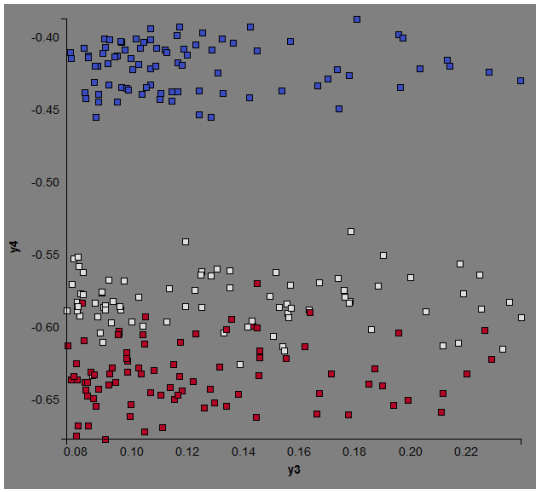
x20	x21	x22	x23	x24	x25	y1	y2	y3	y4
1	-1	-1	0	-1	0	12.5644	12.5651	0.112325	-0.596289
0	0	0	1	-1	0	12.8128	12.813	0.107739	-0.629639
-1	-1	1	0	0	-1	12.5552	12.5559	0.21766	-0.556342
-1	0	0	-1	-1	-1	12.5218	12.5231	0.0963738	-0.40295
1	0	1	1	-1	-1	12.538	12.5386	0.113548	-0.641729

- Table View
  - Linked Variable/Column Selection (member color coding)
    - Scatterplot
    - Bar chart
  - Linked Member/Row Selection (highlights)
    - Scatterplot
  - Column Sort (triangle orientation = ascending/descending)

Parameter Space

- No analysis component
- Interactive data exploration
  - Filtering
  - Remote media viewing
  - Group interactions
  - Video synchronization
- Demonstrate Parameter Space Wizard
  - Categorical Variables
  - Editable Variables

PS Visualization – Member Level



- Scatterplot
  - Points = Runs
  - Axes – 2 ways to set:
    - *X* and *Y* drop downs
    - *X* and *Y* icons in table column headers
- Color-coding
  - 3 ways to set:
    - *Point Color* drop down
    - Row (variable) selection in bar chart
    - Column (variable) selection in table
  - Colors determined by Color Theme
- Selections linked to table
  - 2 ways to set: click (single) or rubberband (multiple)

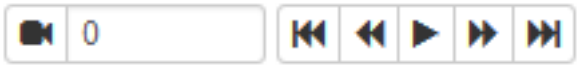
PS Visualization – Selections

- Selections
  - 2 ways to set: click (single) or rubberband (multiple)
  - Linked to table rows (highlight)
  - Linked to *Selection Action*
- *Selection Action* button
  - Drop down actions are data-based
  - *Editable* variables can be set
  - Group hide/show selected points
  - Group

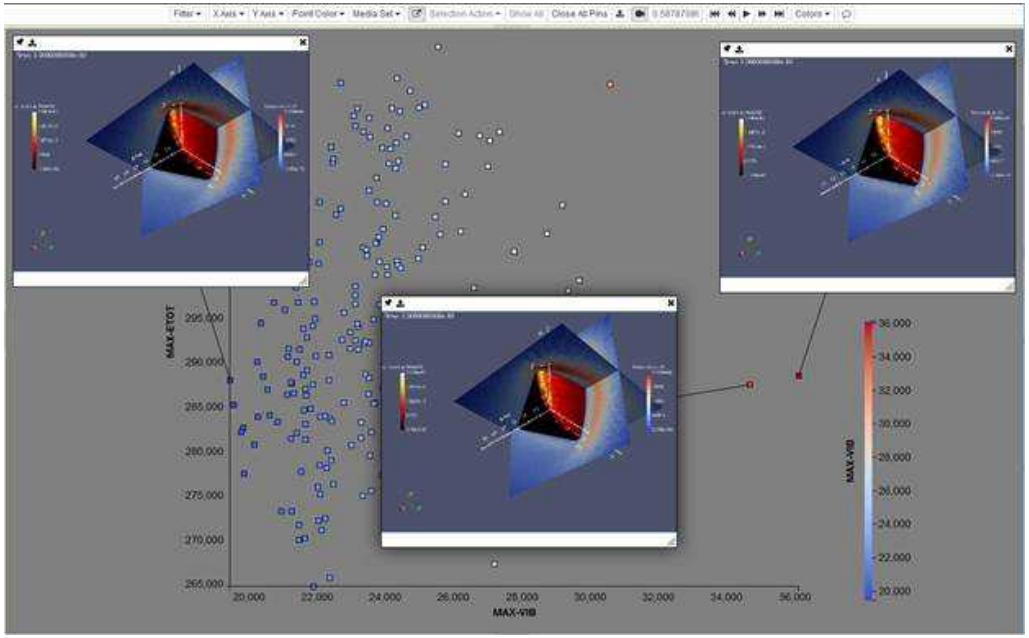
PS Visualization – Media

- Media Retrieval (only if media column in table)
  - Media set in two ways:
    - *Media Set* drop down
    - Data table box icon
  - Hover – temporary single member view
  - Each media type has different viewer
- Pinned Images
  - Hover + viewer movement – fixed single full view
  - Hover + pin icon – fixed single reduced view
  - Selection + selection action – pinned group

PS Visualization – Video Synch

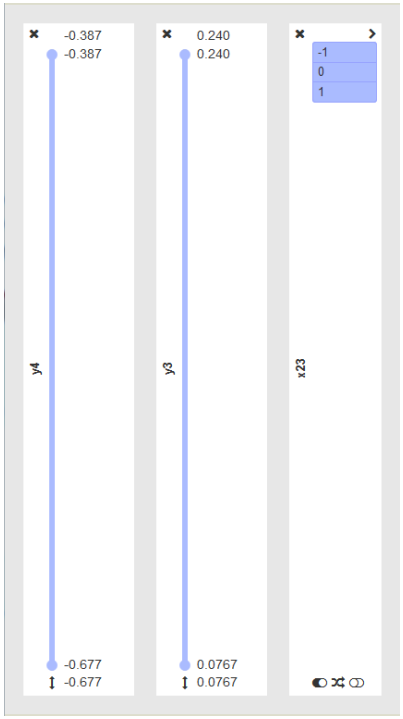


- Video Controls (only when pinned videos)
  - Synch button
  - Shared video time (seconds) – not simulation time
  - Video start
  - Step back 1 frame
  - Play/Pause /
  - Step forward 1 frame
  - Video end
- Current video
  - shadow behind
  - non-synch interaction



PS Visualization– Ensemble Level

- Filters
  - Continuous
    - Slider min/max
    - Range min/max
    - Slider region (blue = values displayed)
    - Invert icon (middle vs. ends)
  - Categorical
    - Button per value
    - Blue = displayed
    - Icons
      - All on
      - Invert button state
      - All off



PS Visualization– Data Level

x20	x21	x22	x23	x24	x25	y1	y2	y3	y4
1	-1	-1	0	-1	0	12.5644	12.5651	0.112325	-0.596289
0	0	0	1	-1	0	12.8128	12.813	0.107739	-0.629639
-1	-1	1	0	0	-1	12.5552	12.5559	0.21766	-0.556342
-1	0	0	-1	-1	-1	12.5218	12.5231	0.0963738	-0.40295
1	0	1	1	-1	-1	12.538	12.5386	0.113548	-0.641729

- Table View
  - Linked Variable/Column Selection (member color coding)
    - Scatterplot
    - Bar chart
  - Linked Member/Row Selection (highlights)
    - Scatterplot
  - X/Y axis
  - Column Sort (triangle orientation = ascending/descending)
  - Media Selection (box icon)

Time Series Model

- Clusters time series by shape and time
  - Same shape at different time NOT similar
  - Similar shape with different amplitude NOT similar
- Agglomerative clustering builds similarity tree (dendrogram)
- 4 Cluster Linkage Measures
  - single: Nearest Point Algorithm
  - complete: Farthest Point Algorithm
  - average: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) Algorithm
  - weighted: Weighted Pair Group Method with Arithmetic Mean (WPGMA) Algorithm
- Cluster Distance Metric - Euclidean

Time Series Model

- Time series data expectations
  - Start/end times must match
  - Sample count and frequency need NOT match
- Binning
  - Bin count = samples compared in distance calculation
  - Bin value = average of samples in bin's time interval
  - More bins
    - greater accuracy
    - longer calculation & slower rendering
  - Less bins
    - loss of features, damping of spikes
    - shorter calculation & faster rendering
  - 500-1000 typical (must be << original sample count)

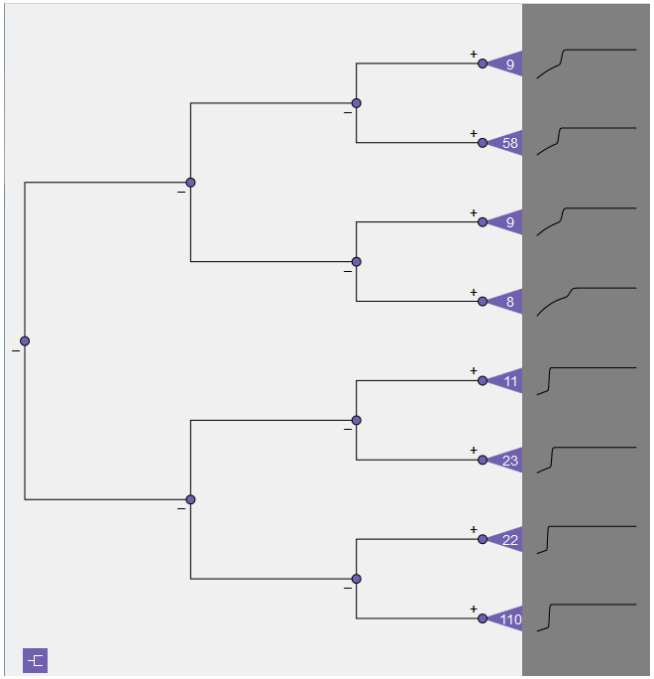
Time Series Wizard

Sandia National Laboratories

- 3 Data Table Formats
  - Xyce - dakota_tabular.dat
  - CSV
  - HDF5 (Slycat intermediate format only)
- 2 Time Series Data Formats
  - .prn files (corresponding to Xyce table)
  - CSV files
- Requires parallel processing
  - All data MUST be stored on cluster
  - You MUST have cluster accounts
  - Batch Setup & asynchronous completion
  - Job status checking
- Demonstration of wizard

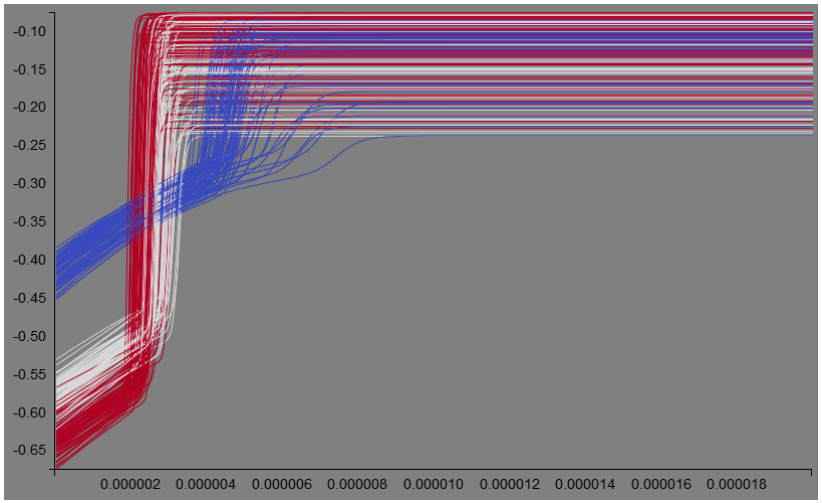
TS Visualization – Ensemble Level

- Dendrogram
  - Select time series variable (dendrogram)
  - Node selection restricts visible tree
    - Control click adds/subtracts nodes
    - Line plot & data table = only visible tree
  - +/- icons expand/contract subtree detail
  - Subtree triangles
    - Show leaf count
    - CAUTION: Clicking expands tree to leaf level
  - Sparklines
    - Exemplar shape for subtree
    - Colored black unless leaf, leaf colored by line color
    - Highlighted sparkline = selection in subtree
    - Sparkline selection = select subtree members in line plot & table rows
  - Dendrogram order icon for sorting data table



TS Visualization – Member Level

- Line Plot
  - Lines = runs
  - X axis = time
  - Y axis = amplitude of series var
  - Color-coded by selected variable
  - Limited to visible dendrogram
  - Hover highlights lines
  - Clicking selects lines
    - Linked to row select in data table
    - Linked to sparkline highlight in dendrogram subtree



TS Visualization – Data Level

x20	x21	x22	x23	x24	x25	y1	y2	y3	y4
1	-1	-1	0	-1	0	12.5644	12.5651	0.112325	-0.596289
0	0	0	1	-1	0	12.8128	12.813	0.107739	-0.629639
-1	-1	1	0	0	-1	12.5552	12.5559	0.21766	-0.556342
-1	0	0	-1	-1	-1	12.5218	12.5231	0.0963738	-0.40295
1	0	1	1	-1	-1	12.538	12.5386	0.113548	-0.641729

- Table View
  - Linked Variable/Column Selection (member color coding)
    - Scatterplot
    - Bar chart
  - Linked Member/Row Selection (highlights)
    - Scatterplot
  - Column order
    - Sorted value order (triangle icon = ascending/descending)
    - Dendrogram order (icon in dendrogram)

Questions?

Sandia National Laboratories