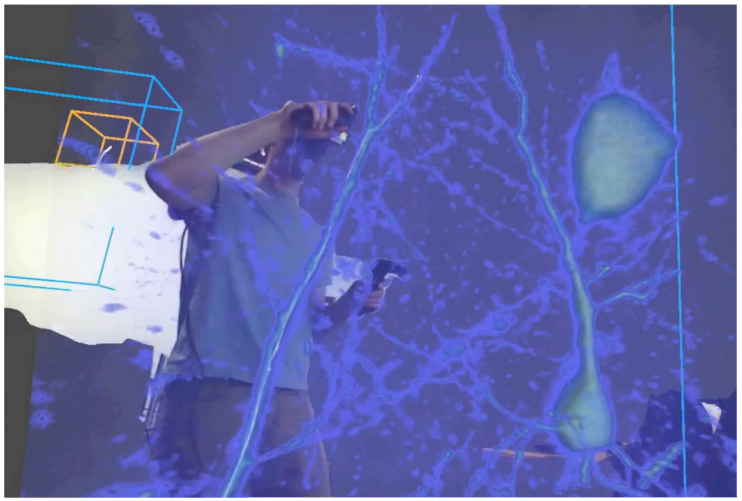
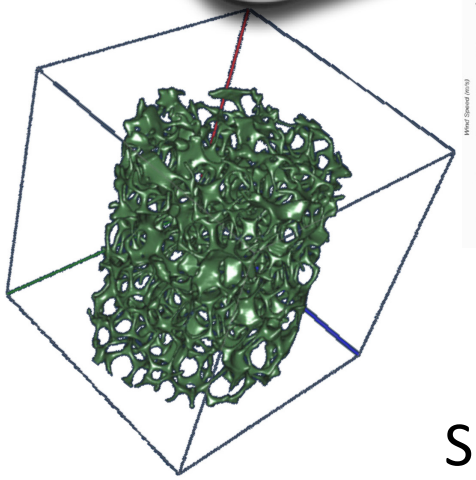
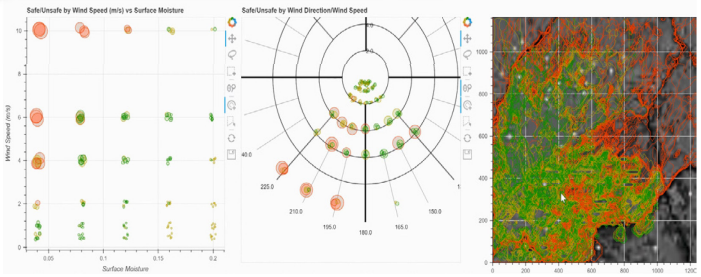
Advanced Modeling & Simulation (AMS) Seminar Series
NASA Ames Research Center, April 20th, 2023

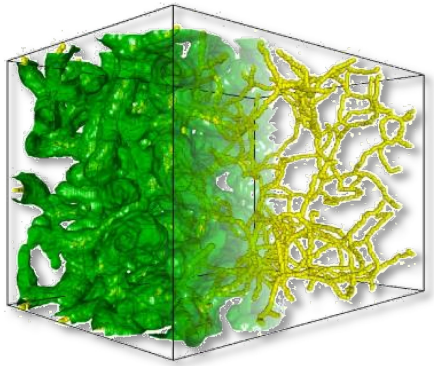
Big Data Analytics for Decision Making









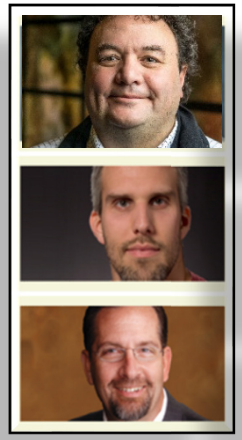
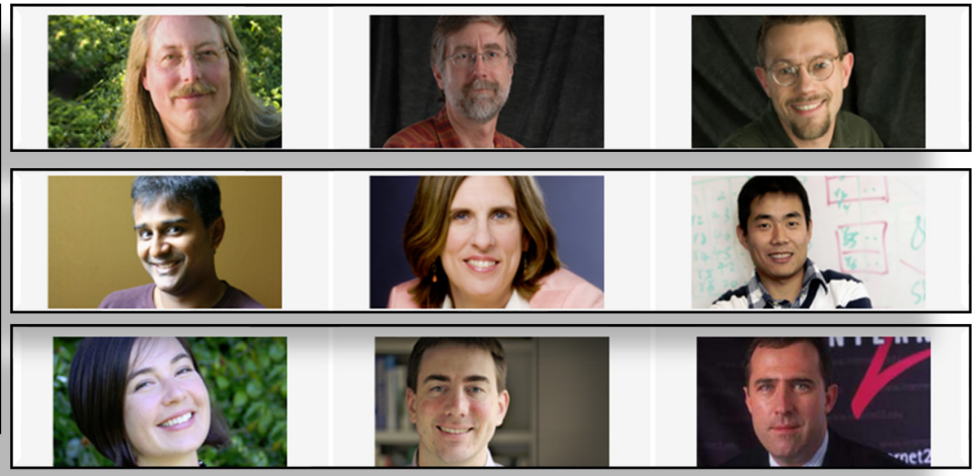


Valerio Pascucci

SCI Institute and School of Computing
Center for Extreme Data Management Analysis and Visualization

Center for Extreme Data Management, Analysis, and Visualization

- 10 Faculty + scientists, developers, students, …
- Primary partners: UU & PNNL
- Other partnerships: NSA, INL, LLNL, ANL, Battelle, ….
- Involvement in national Initiatives

SDAV Scalable Data Management, Analysis, and Visualization

EXACT

SUPER

THE UNIVERSITY OF UTAH Pacific Northwest NATIONAL LABORATORY

Lawrence Livermore National Laboratory

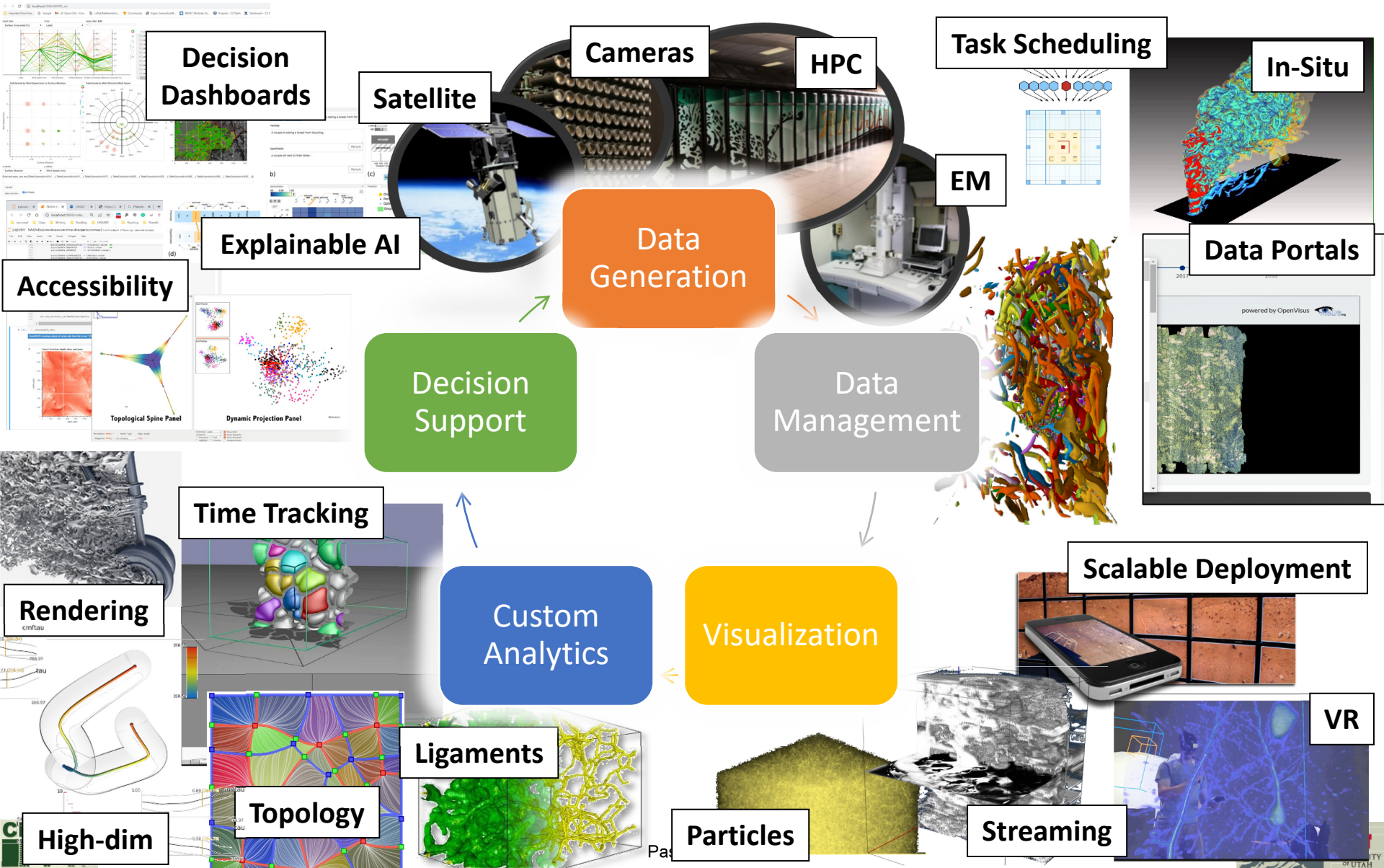
Sandia National Laboratories

INL Idaho National Laboratory



$1.6B NSA data center
(1.5 million-square-foot facility)

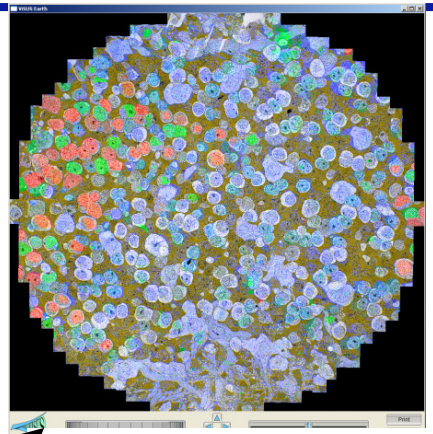
CEDMAV builds infrastructure and tools for scientific workflows



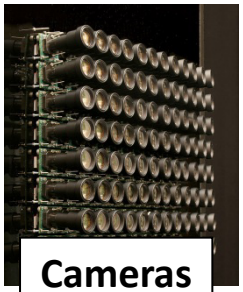
High Resolution Sensing Devices and Massively Parallel Simulation Generate Big Scientific Data


BlueGene/L


Satellite

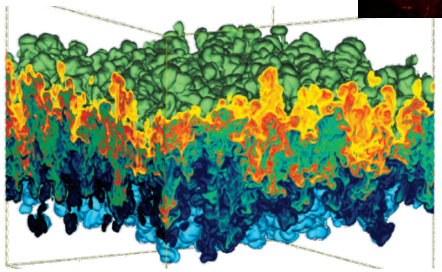

Retinal Connectome

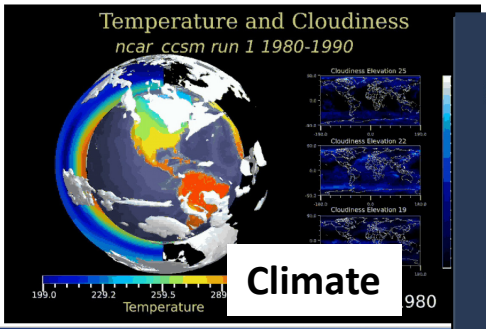

Wildfires


Cameras

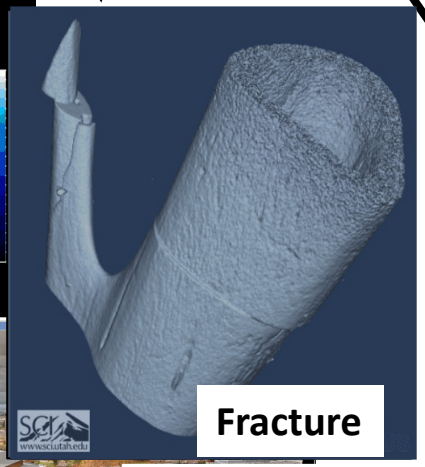

EM

Jaguar

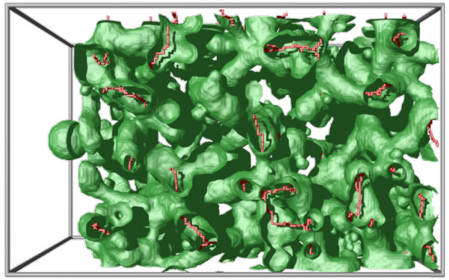

Hydrodynamic Inst.

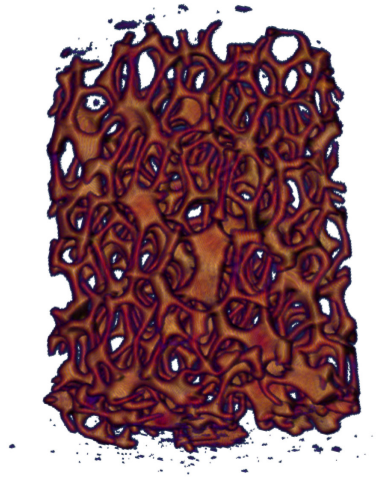


Climate


Fracture


Photography


Porous Materials


Metallic Foam

The National Science Data Fabric: Democratizing Data Access for Science and Society



National Science Data Democratization Consortium

SCI
www.sci.utah.edu

NSDDC

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Protocol Labs
Research

UNIVERSITY OF MICHIGAN

JOHNS HOPKINS UNIVERSITY

Spark · hadoop MapReduce · APACHE HBASE · Flink · H_2O · cloudera IMPALA

ALLUXIO

S3 · GlusterFS · HDFS · openstack · NFS · ceph

Petabytes of cloud storage for large scale scientific data (e.g.

SDSC
SAN DIEGO SUPERCOMPUTER CENTER









The National Science Data Fabric Catalog



Owen Koppe

An academia-industry collaboration to build a transdisciplinary, multi-federation catalog for real-time search over *75 petabytes* of data, *1.59 billion* records from 68 community repositories

Total size **75.31 PiB** Number of records **1.59 bilions**

Catalog * Limit 100 Show ☑ Catalog ☑ Bucket ☐ Extension

Catalog	Bucket	tot_size	num_files
aws-open-data	sra-pub-run-odp	13,616,090,404,273,786	17,381,287
ranch	AST20015	9,044,348,820,142,836	64,188
ranch	PHY20003	4,315,046,973,201,938	231,336
ranch	CTS20009	2,528,266,508,417,821	48,623
aws-open-data	deafrica-sentinel-2	2,389,902,124,666,594	45,886,993
aws-open-data	noaa-hrrr-bdp-pds	2,144,318,216,738,649	33,582,006
aws-open-data	noaa-gefs-pds	1,890,930,002,645,210	89,362,647
ranch	rctftacc	1,497,793,189,666,427	32,218
ranch	ATM20005	1,445,297,808,898,192	76,000
aws-open-data	human-pangenomics	1,305,993,292,339,773	183,868



intel
ge Technology
ud Storage
HY SEAL
l provides sustainable, immutable, and
rdable data storage.
DATA RETRIEVAL
Access data in hours vs days
COST EFFECTIVE
Up to 80% less than competitors
SECURE
Tamper proof and verifiable

Discovering data across scientific

The National Science Data Fabric Catalog



Aashish Panta

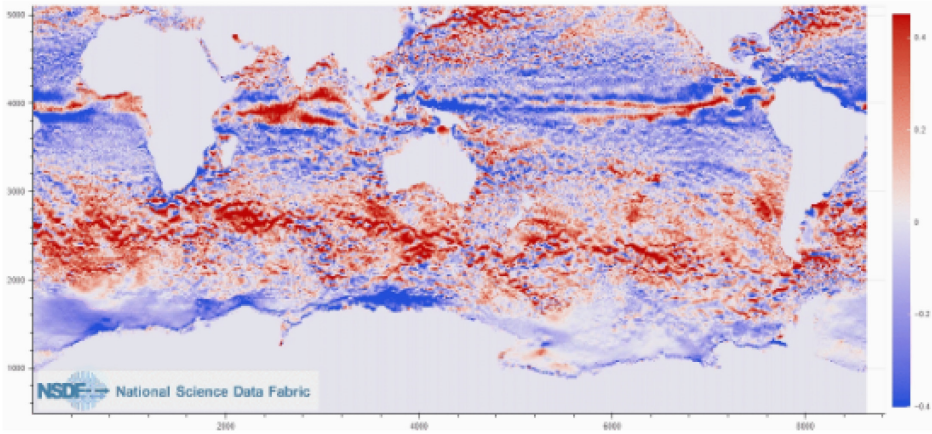
Catalog	E
aws-open-data	s
ranch	A
ranch	F
ranch	C
aws-open-data	d
aws-open-data	n
aws-open-data	n
ranch	r
ranch	A
aws-open-data	h

Discov

Seal Storage Technology and the University of Utah Enable Data Democratization for NASA Climate Data

Over 200TB of NASA climate data is now accessible through a Jupyter notebook due to a collaboration between Seal Storage Technology and the University of Utah's Scientific Computing and Imaging Institute (SCI), enabling researchers around the world to access the data with zero egress fees and visualize the data with ease.

Toronto, Ontario, and Salt Lake City, Utah, 12 April 2023: In an effort to democratize data, decentralized cloud storage provider Seal Storage Technology and the University of Utah's Scientific Computing and Imaging Insitute (SCI) announced today that over 200 terabytes (TB) of OpenViSUS IDX NASA atmospheric wind velocity and ocean current datasets are now available through a Jupyter notebook. A Jupyter notebook is a digital document that makes it easy to access and visualize data by combining text, code, and visualizations all in one place. This project is part of SCI's participation in The National Science Data Fabric (NSDF), a trans-disciplinary approach to integrated data delivery and access, democratizing data-driven scientific discovery through shared storage, networking, computing, and educational resources. "The democratization of data is essential to scientific progress, and we are thrilled to partner with the University of Utah to make this possible," said Seal Storage Technology COO Alex Altman. "Seal's decentralized cloud storage platform ensures that this valuable data is accessible to researchers around the world, without any



The Na



Aashish Panta

Catalog	E
aws-open-data	s
ranch	A
ranch	F
ranch	C
aws-open-data	d
aws-open-data	n
aws-open-data	n
ranch	r
ranch	A
aws-open-data	h

Discov

jupyter 10-bokeh-nasa-200tb (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

```
    view.setQuality(quality)
    view.setLogicToPixel([(0.0,1.0), (0.0,1.0), (0.0,30.0)])

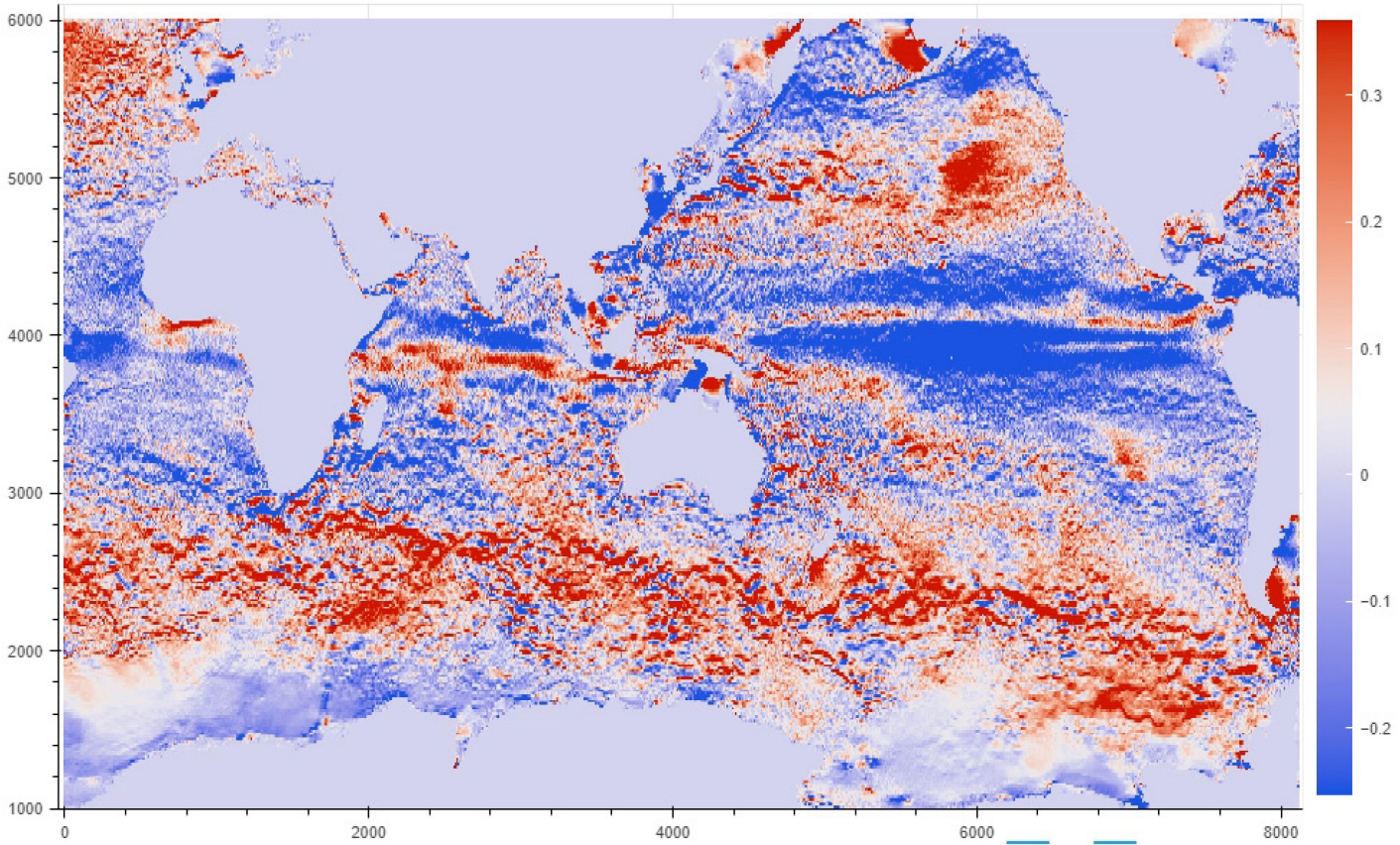
    view.children[0].setDirection(2)
    view.children[0].setOffset(0)

    doc.add_root(view.getBokehLayout(doc=doc))

ShowBokehApp(MyApp)
```

Palette: colorcet.cool | Time: 0 | Time delta: 2 | Field: u | Quality: 0 | Play | Play sec: 0.01

Direction: Z | Offset: 0



The Na

jupyter 10-bokeh-nasa-200tb (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run Code

neon National Ecological Observatory Network



Duong Hoang



NEON Distributed Sensor Networks

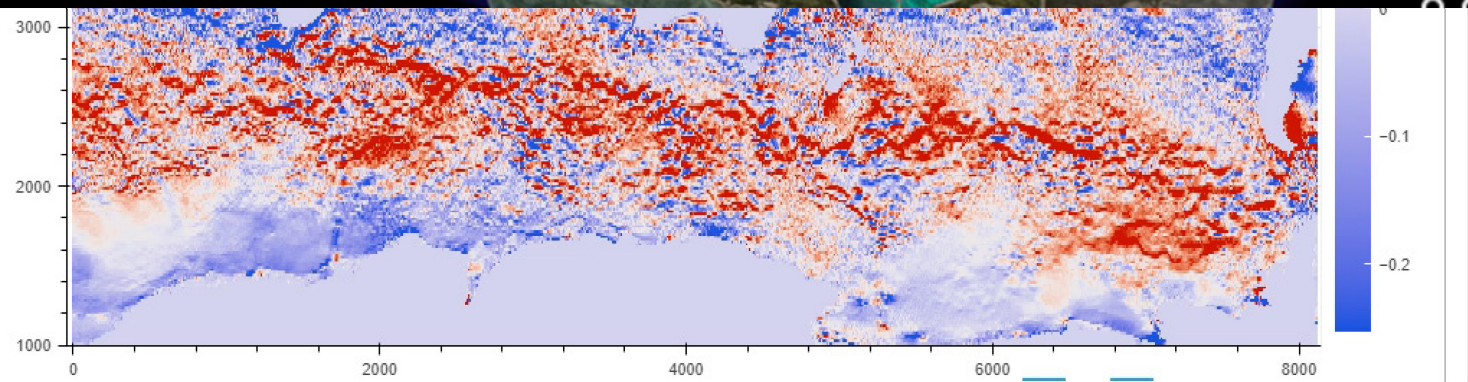
aws-open-data

ranch

ranch

aws-open-data

Discov



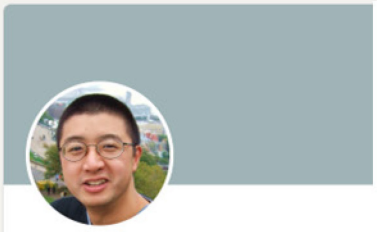
The Na

jupyter 10-bokeh-nasa-200tb (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run Code

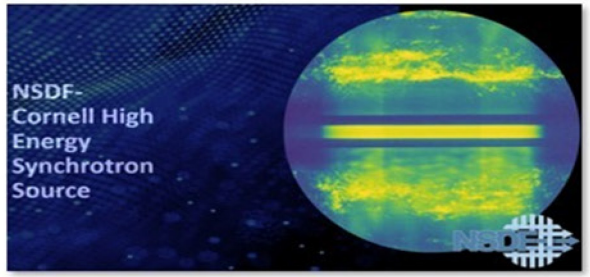


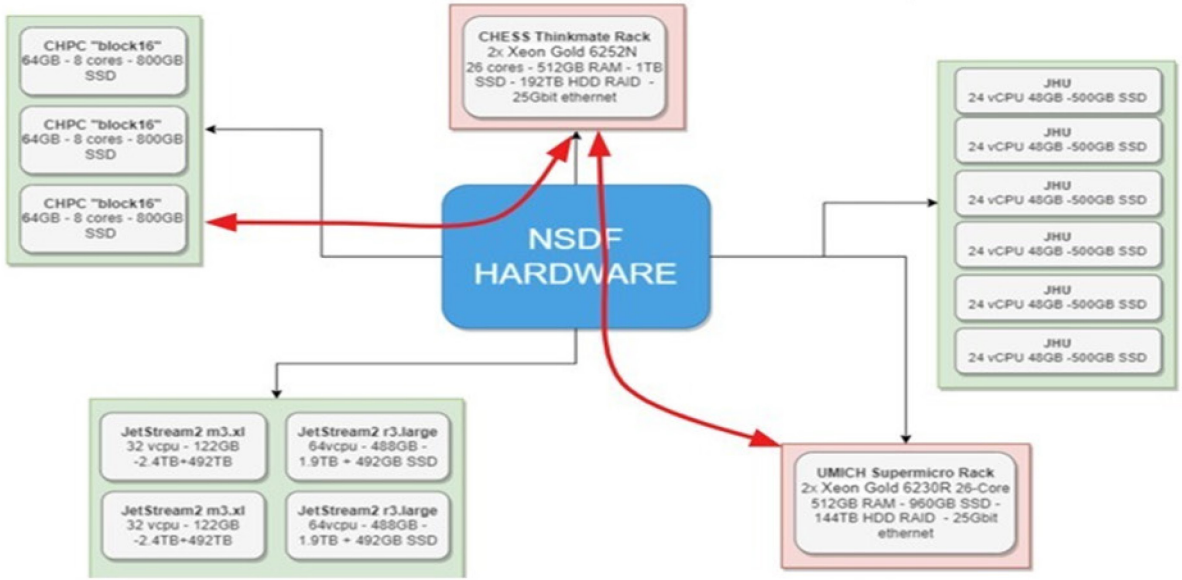


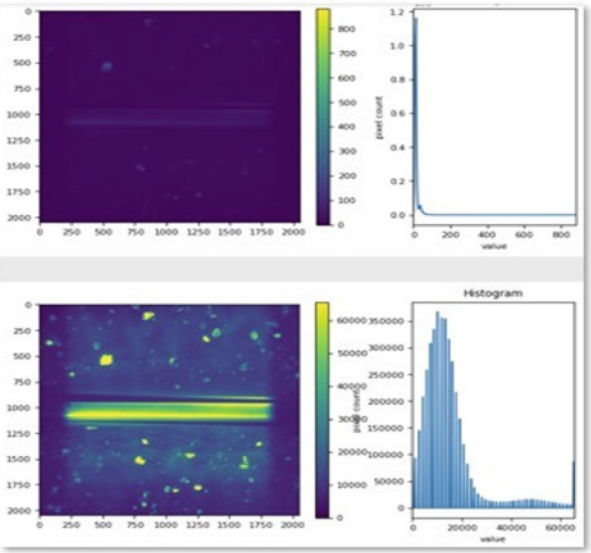
Werner Sun · 2nd
IT Director, Cornell Laboratory for Accelerator-Based S
Education at Cornell University
Ithaca, New York, United States · Contact info

Direct Data Access from Cornell High Energy Synchrotron Source (CHESS)

CHESS
CORNELL HIGH ENERGY SYNCHROTRON SOURCE



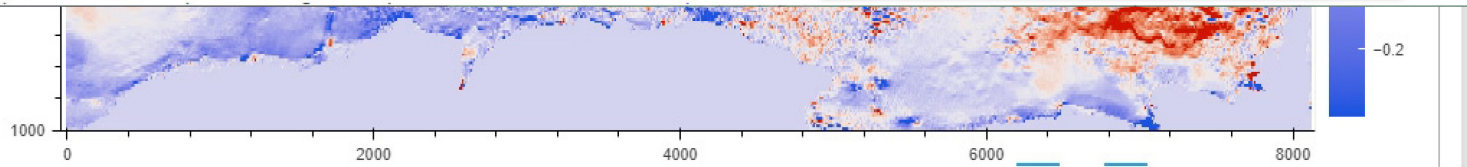




~1500 datasets
Uncompressed size 7.7TiB
Compressed size 2.4TiB

NSDF National Science Data Fabric

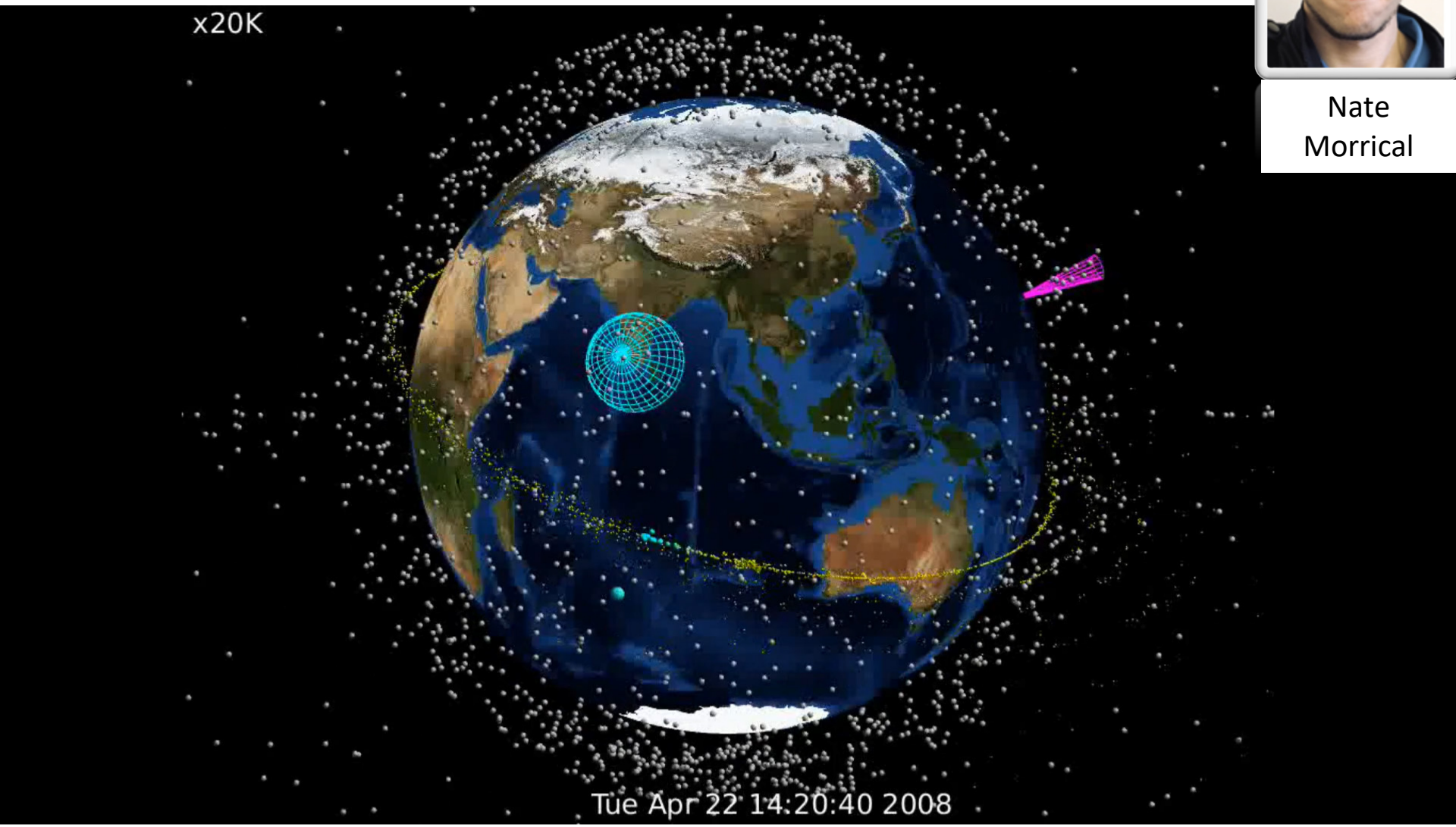
Discov



3D Visualization Provides Very Effective Intuition in Support of Decision Making

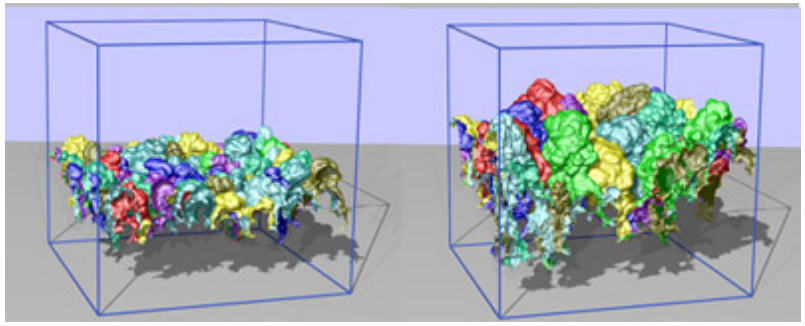


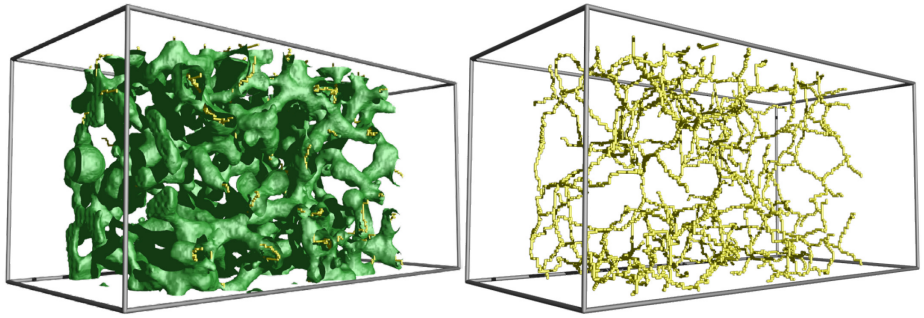
Nate Morrical



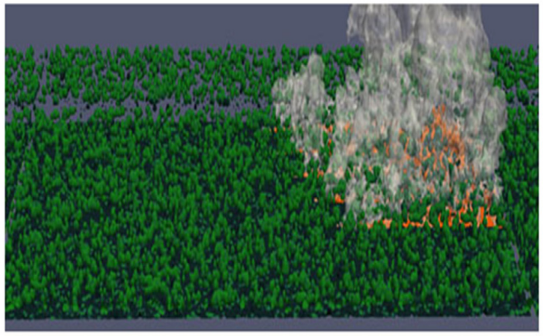
Complex Decisions Benefit from 3D Visualizations Combined Detailed Analytics

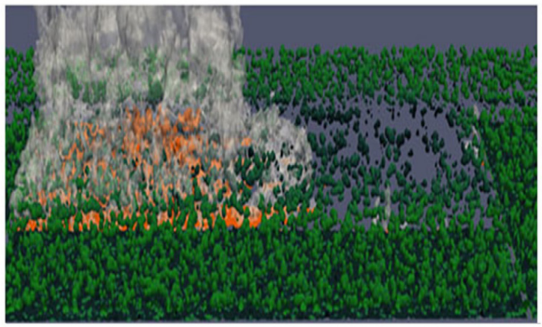
- Topological abstractions allowed to count features and express structural skeletons of scientific data spaces

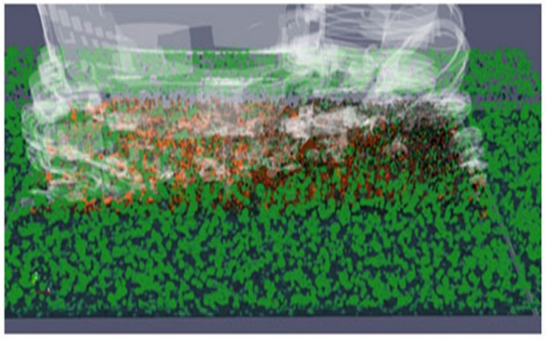




- Time tracking of models in time allows detailed analysis of dynamic phenomena







Explanations are Essential to Good Decision Making

A user may need to answer important questions such as:

How much can I trust a model/set of predictions?

Can I anticipate the outcome of a particular event?

Why did the model give me the result it did?

How stable is the prediction?

- w.r.t. perturbation of training, sample, model?

Could I have changed something to get a better prediction for the kinds of phenomena I am interested in?

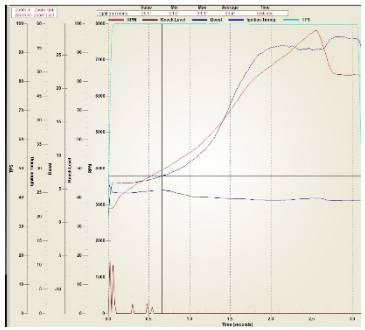
How would you make an automatic transmission “explainable”?

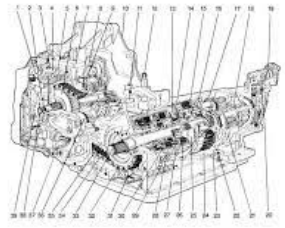


The user needs to *have an expectation of* what this will do to use/trust it



Interpretability approach: provide a view *inside* to see cause and effect







Explainability approach: provide the tools to let the user build a mental model



Explanations for Decision Making permeate science workflows, and users from diverse backgrounds with multiple roles

Domain Expert
Expert in application
Intuition about phenomena
Primary tool users

Expert Analyst
Knowledge in application
Intuition process->outcome
Adapts within tool context

Policy Maker
Not expert in application
Intuition about ML Models
Power to adapt tools

Needs to process data
- Quantity
- Speed

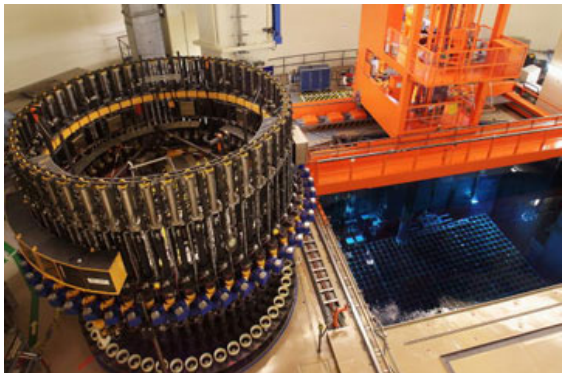
Success is
- Outcome
- Accuracy



New Clean Energy

Needs to manage process
- Identify problems
- Find solutions

Success is
- Domain Experts Happy
- Integrate tool in production



Manage Nuclear Reactors

Needs to develop tools
- Understand tradeoffs
- Robustness/stability

Success is
- Tools get used
- Improved performance/accuracy



Wildfire mitigation

Explanations for Decision Making permeate science workflows, and users from diverse backgrounds with multiple roles

Domain Expert
Expert in application
Intuition about phenomena
Primary tool users

Expert Analyst
Knowledge in application
Intuition process->outcome
Adapts within tool context

Policy Maker
Not expert in application
Intuition about ML Models
Power to adapt tools

Explainable Visualization Support Users for all Roles

Needs
- (
- S

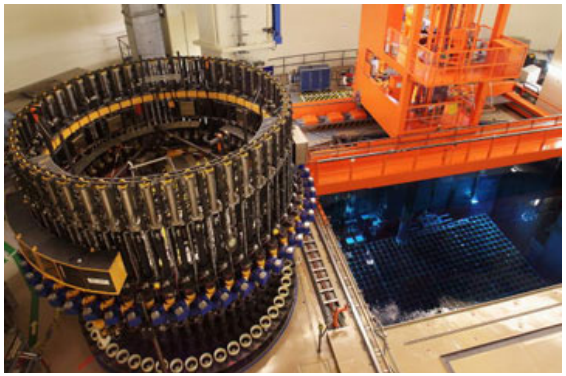
Succe
- (
- Accuracy

ffs
ty

performance/accuracy



New Clean Energy

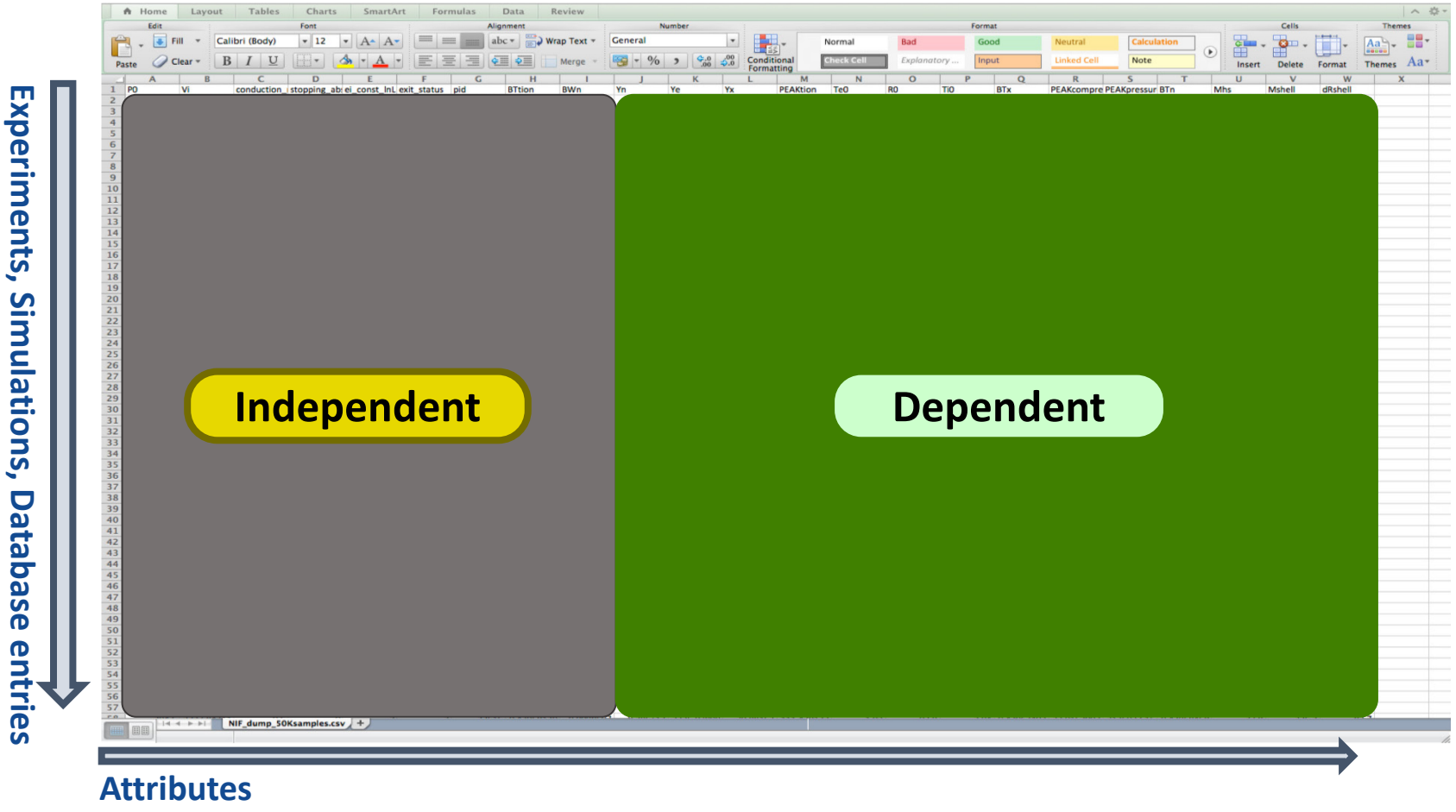


Manage Nuclear Reactors



Wildfire mitigation

The Core is to Understand Relationships in High-Dimensional Multivariate Data



Repeated analysis by selecting a “Domain of Interest” X and a “Figure of Merit” Y

$$F : X \rightarrow Y$$

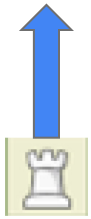
Scientists/Engineers/AI Optimize “the best” function for the task, given the input

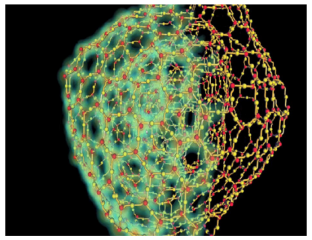
$f : X \rightarrow Y$

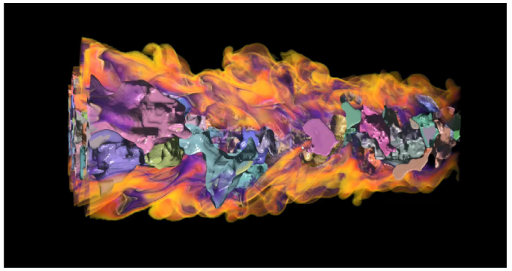
Maximize the quality of the outcome

f () → 

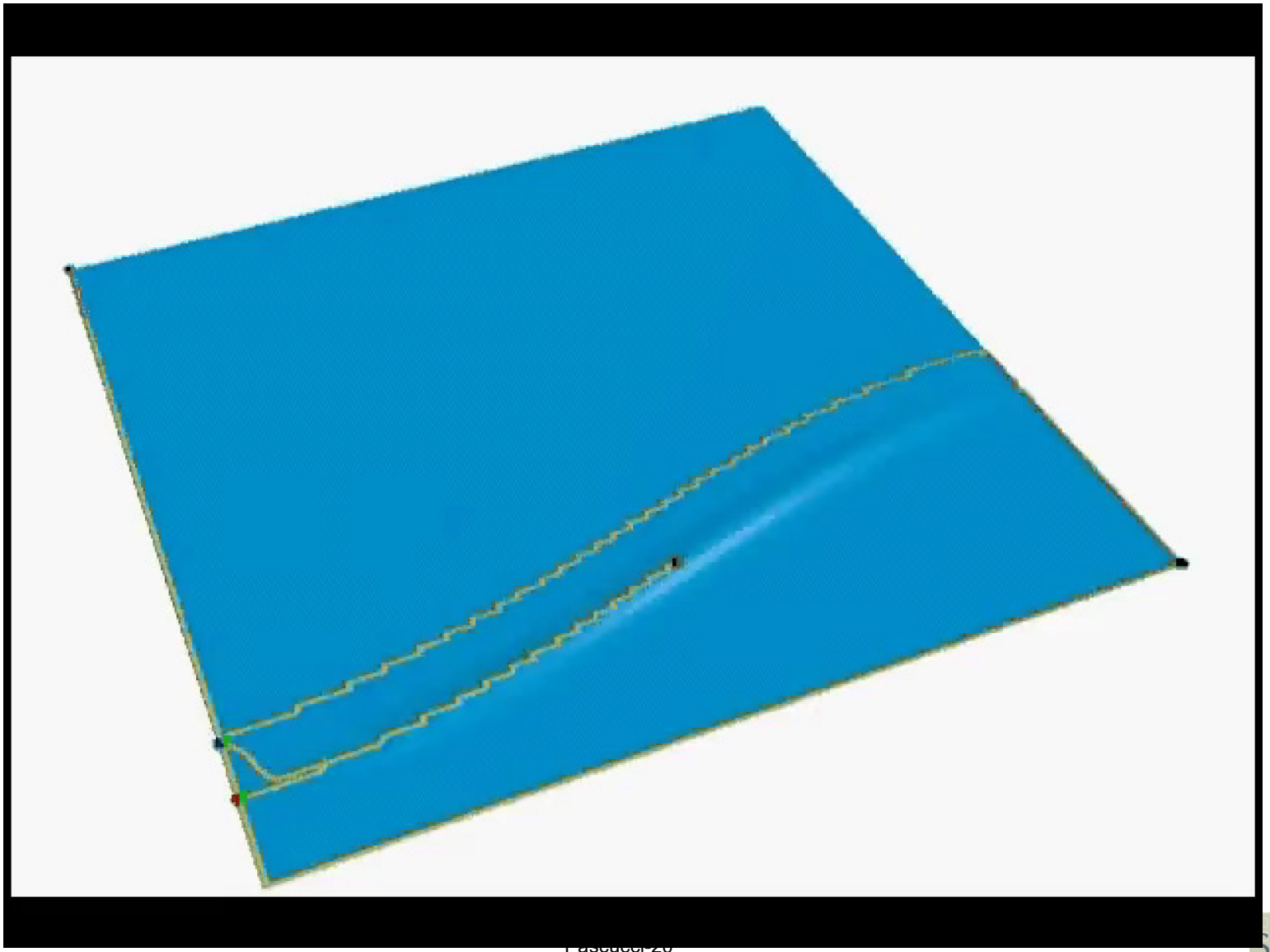
f () → cat

f () → 

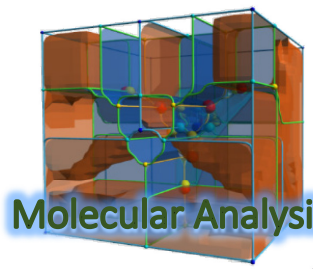
f () → Better Battery

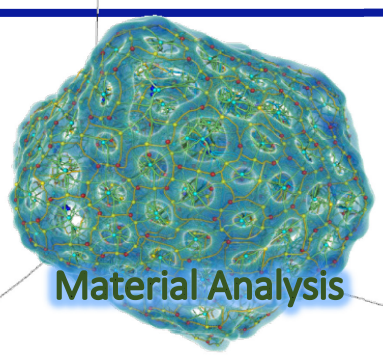
f () → Clean Energy

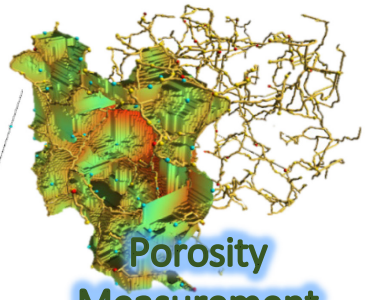
Topology is and Effective Language to Describe Abstractions of Features from Raw Data

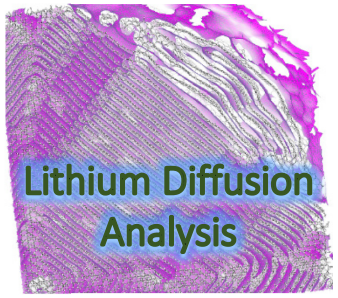


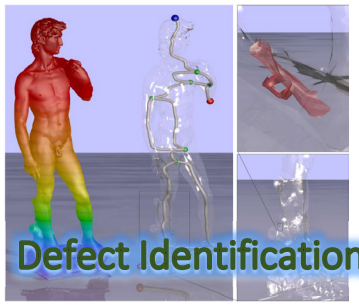
Topology Has Been Successful for Analysis and Visualization of Diverse Scientific Data

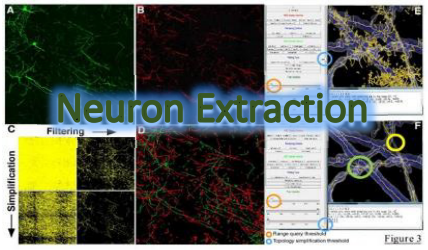

Molecular Analysis

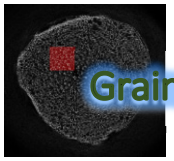

Material Analysis


Porosity Measurement

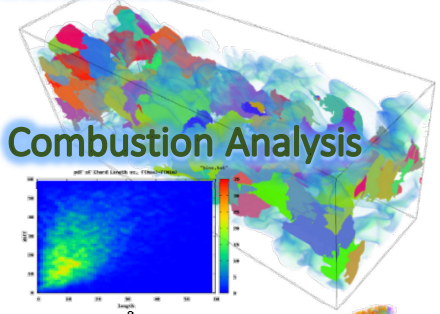

Lithium Diffusion Analysis

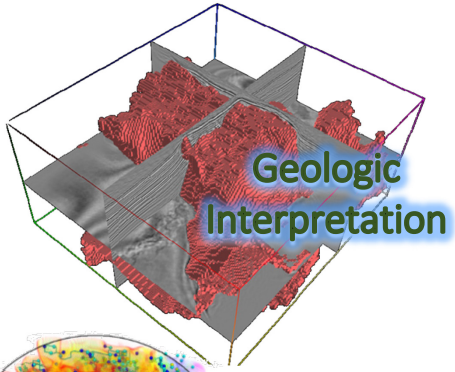

Defect Identification

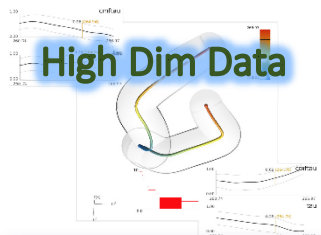

Neuron Extraction


Grain Analysis

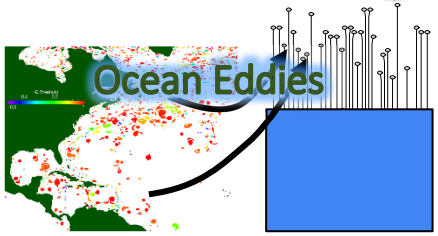


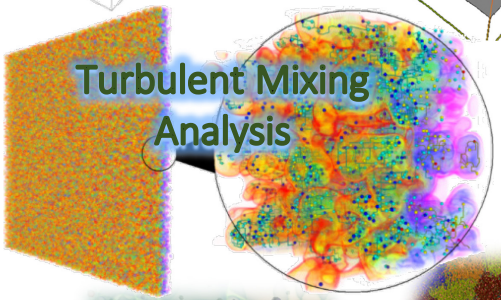

Combustion Analysis

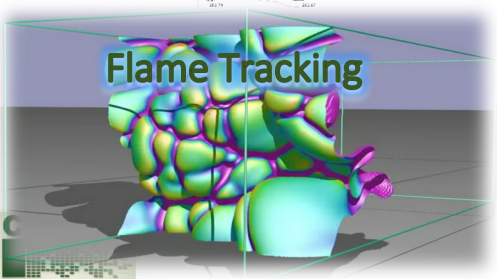

Geologic Interpretation

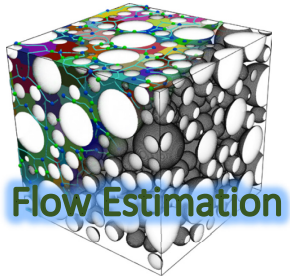

High Dim Data


Local Vorticity

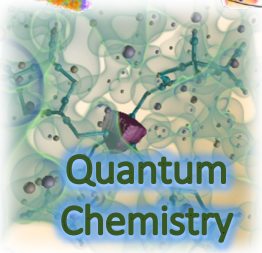

Ocean Eddies


Turbulent Mixing Analysis


Flame Tracking

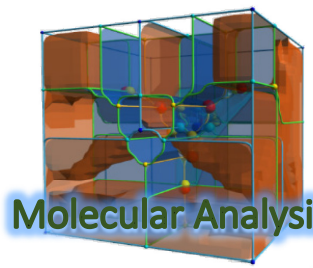

Flow Estimation

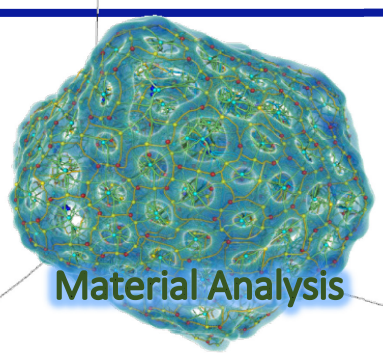

Explosives Performance

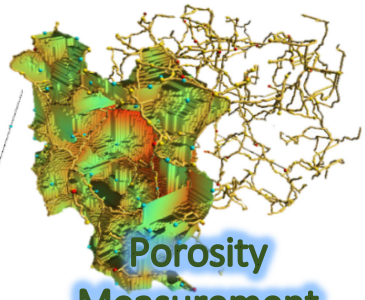

Quantum Chemistry

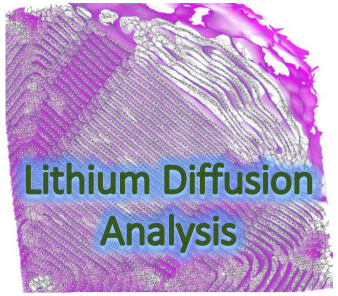

Osteosite Networks

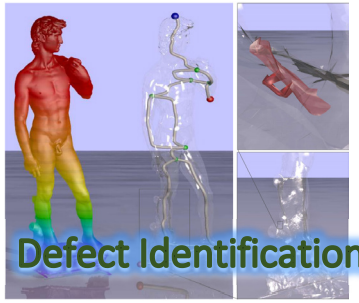
Topology Has Been Successful for Analysis and Visualization of Diverse Scientific Data

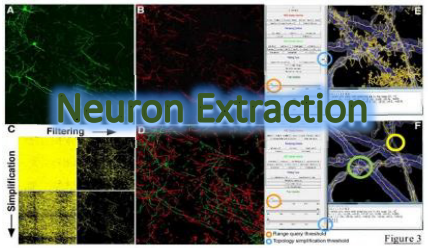

Molecular Analysis

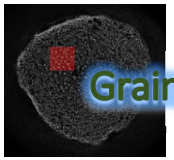

Material Analysis


Porosity Measurement

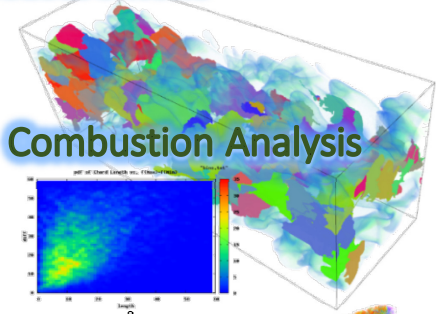

Lithium Diffusion Analysis

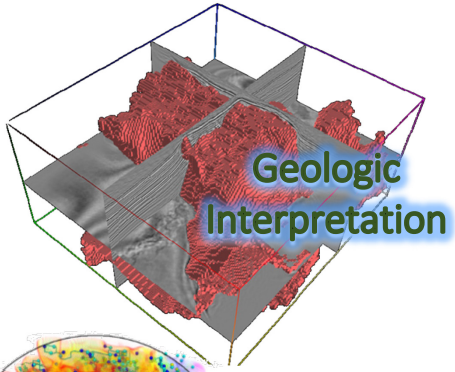

Defect Identification

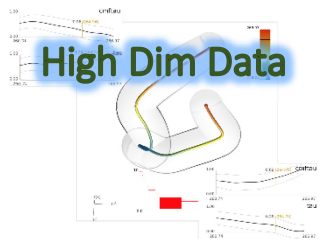

Neuron Extraction


Grain Analysis

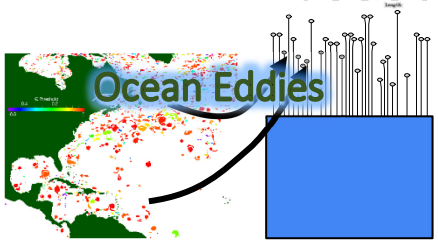


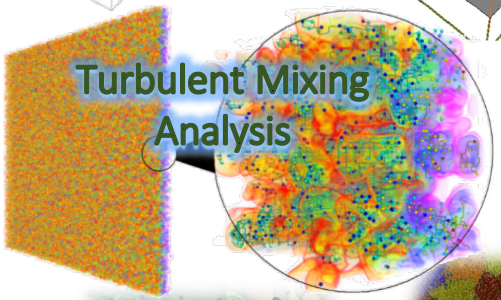

Combustion Analysis


Geologic Interpretation

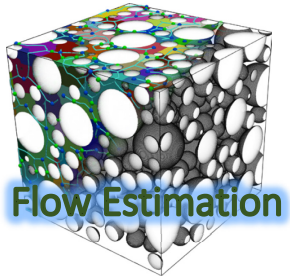

High Dim Data


Local Vorticity

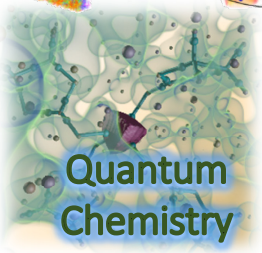

Ocean Eddies


Turbulent Mixing Analysis

Flame Tracking

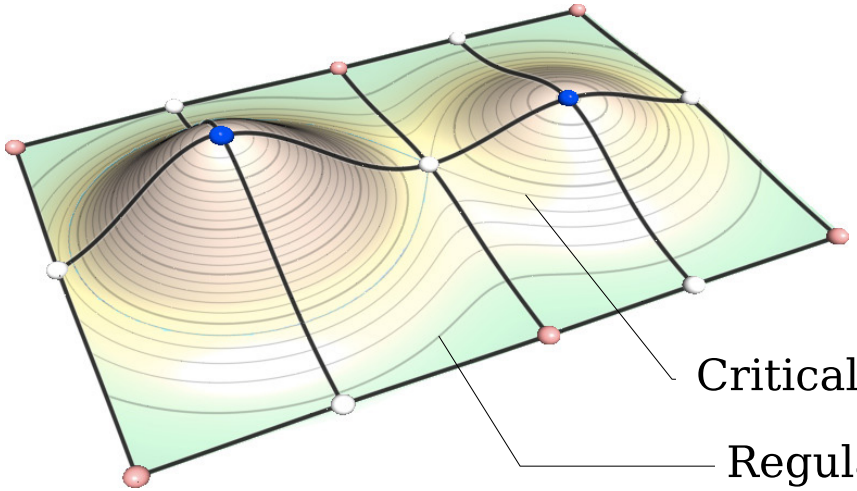

Flow Estimation

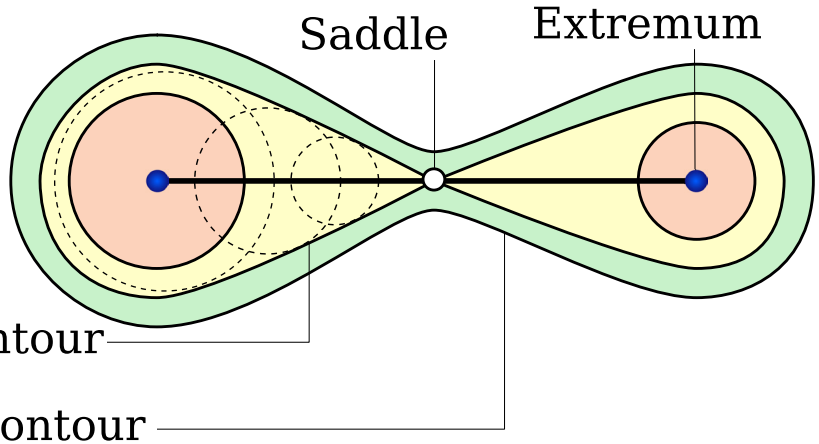

Explosives Performance

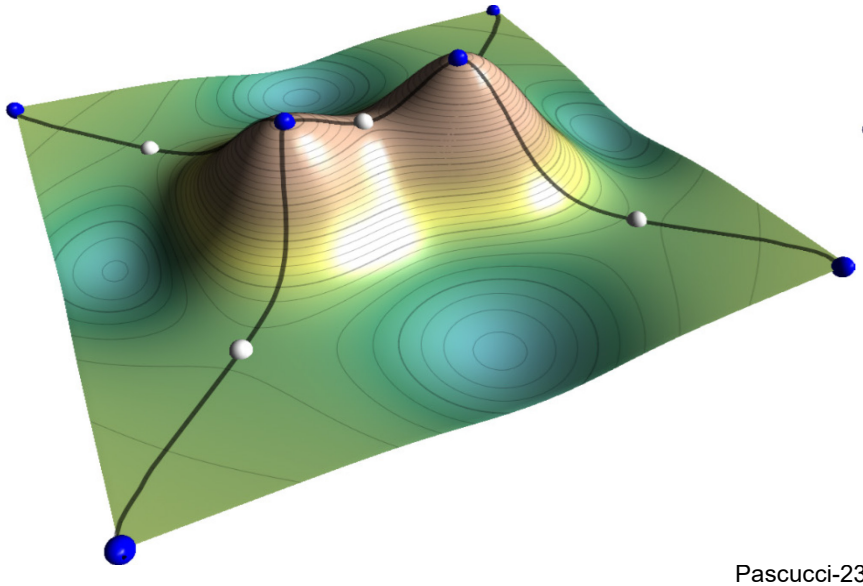

Quantum Chemistry

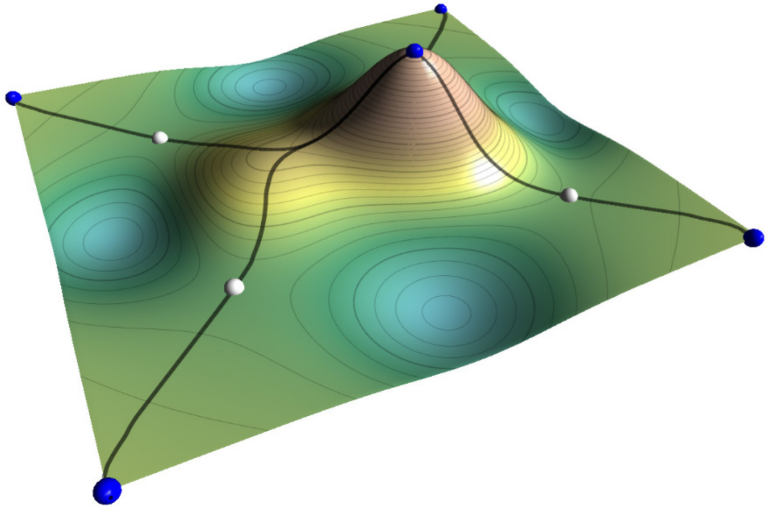

Osteosite Networks

Simplification Allows to Create a Multi-Scale Representations of the Data

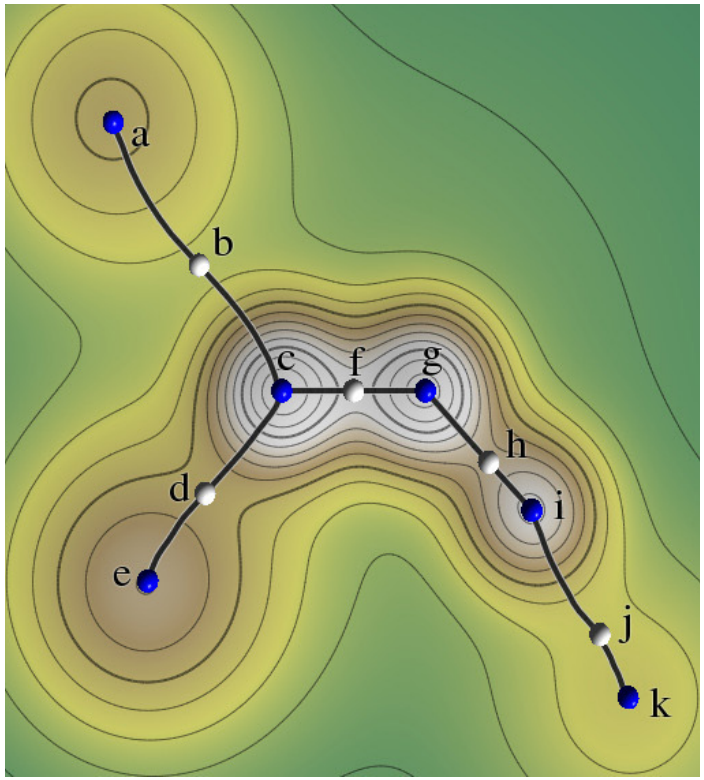


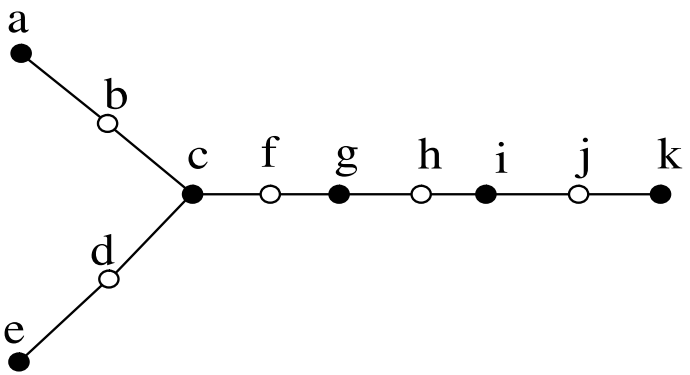


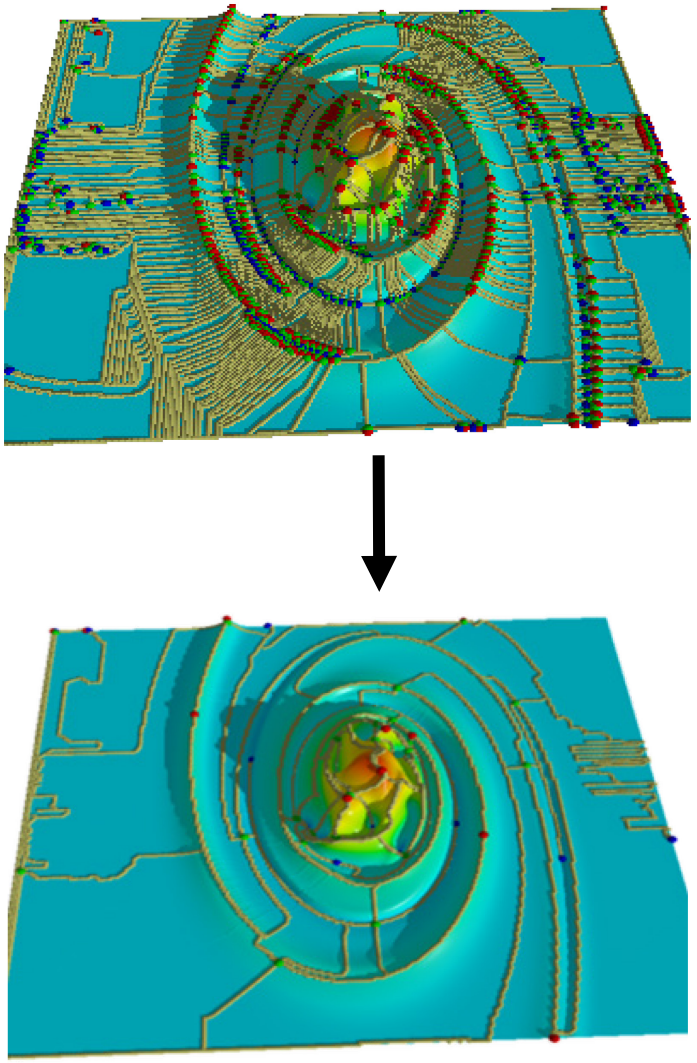




Cancellation Trees Identify Prominent Ridges as Part of the Simplification Process

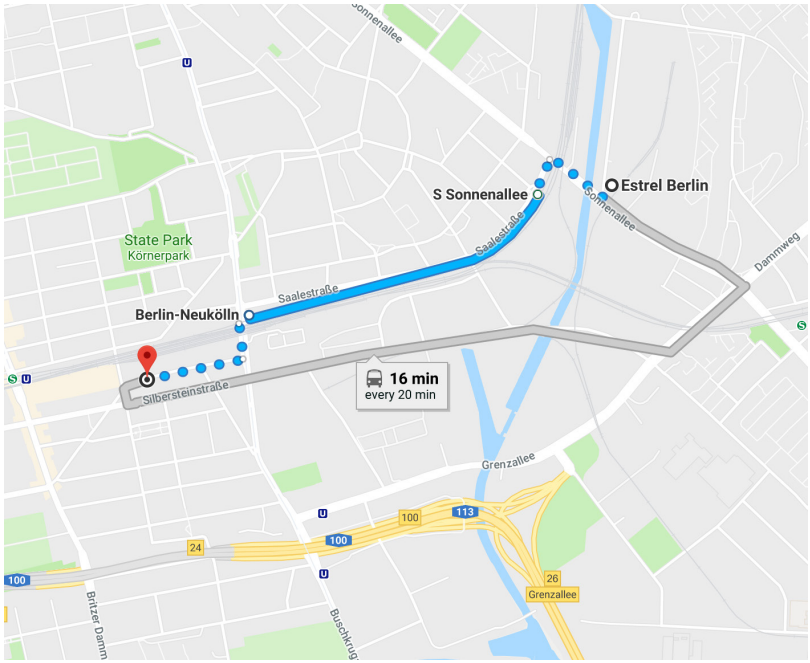






We need the right abstraction for the task



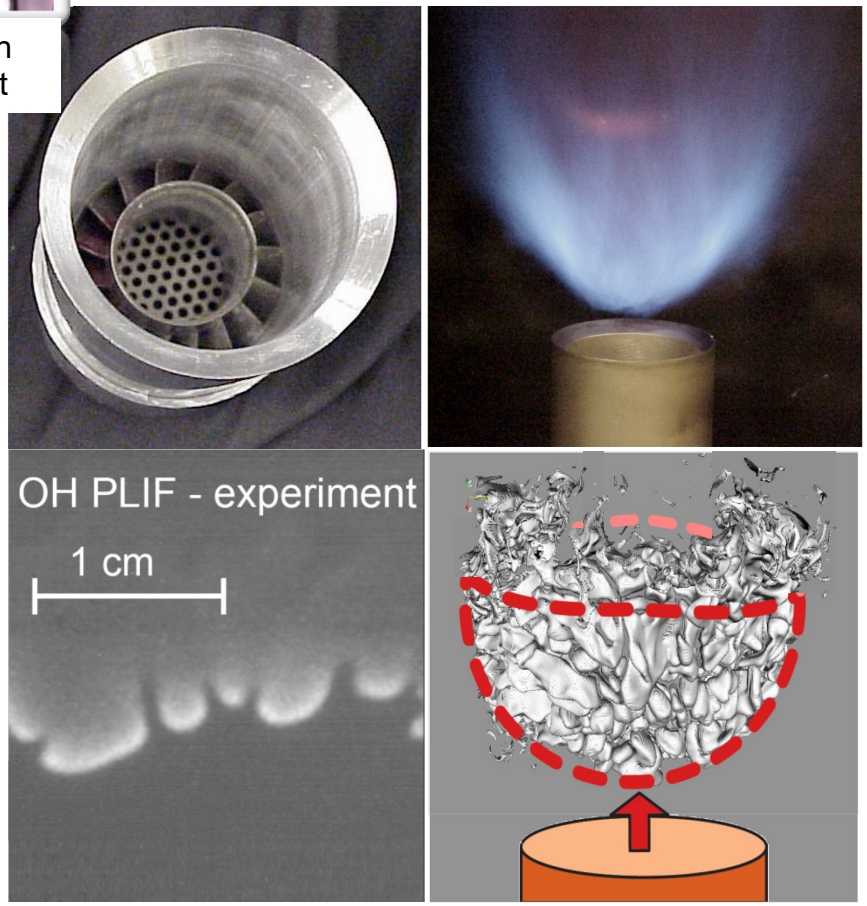


BURNING AND EXTINCTION REGION TRACKING



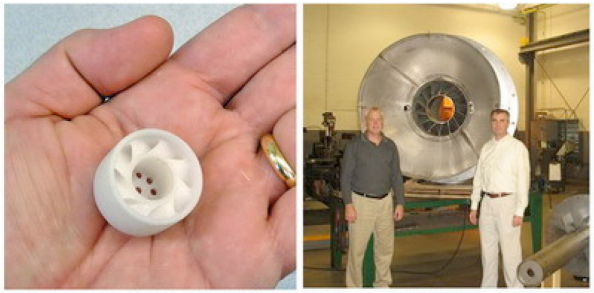
Aniketh Venkat

Understanding Turbulence for Low Emission, High Efficiency Combustion



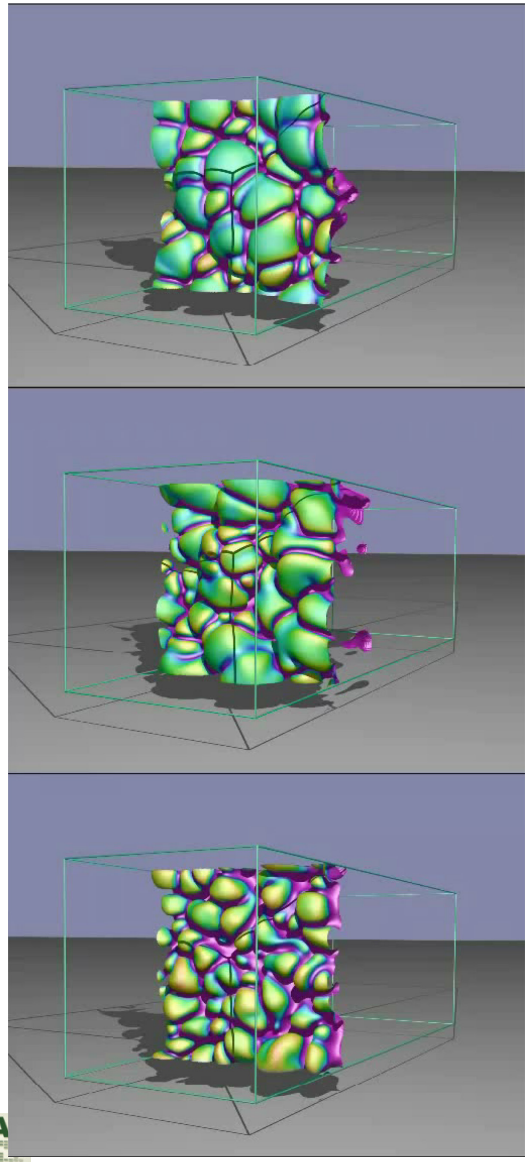
Experiment Simulation

- Lean premixed H_2 flames
- Low Swirl Combustion (LSC) Burners
- Low pollution in energy production
- High Efficiency in fuel consumption
- Scalable from residential to industrial use
- Each variable 3.9-4.5 TB



1" burner (5 kW, 17 KBtu/hr) 28" burner (44 MW, 150 MBtu/hr)

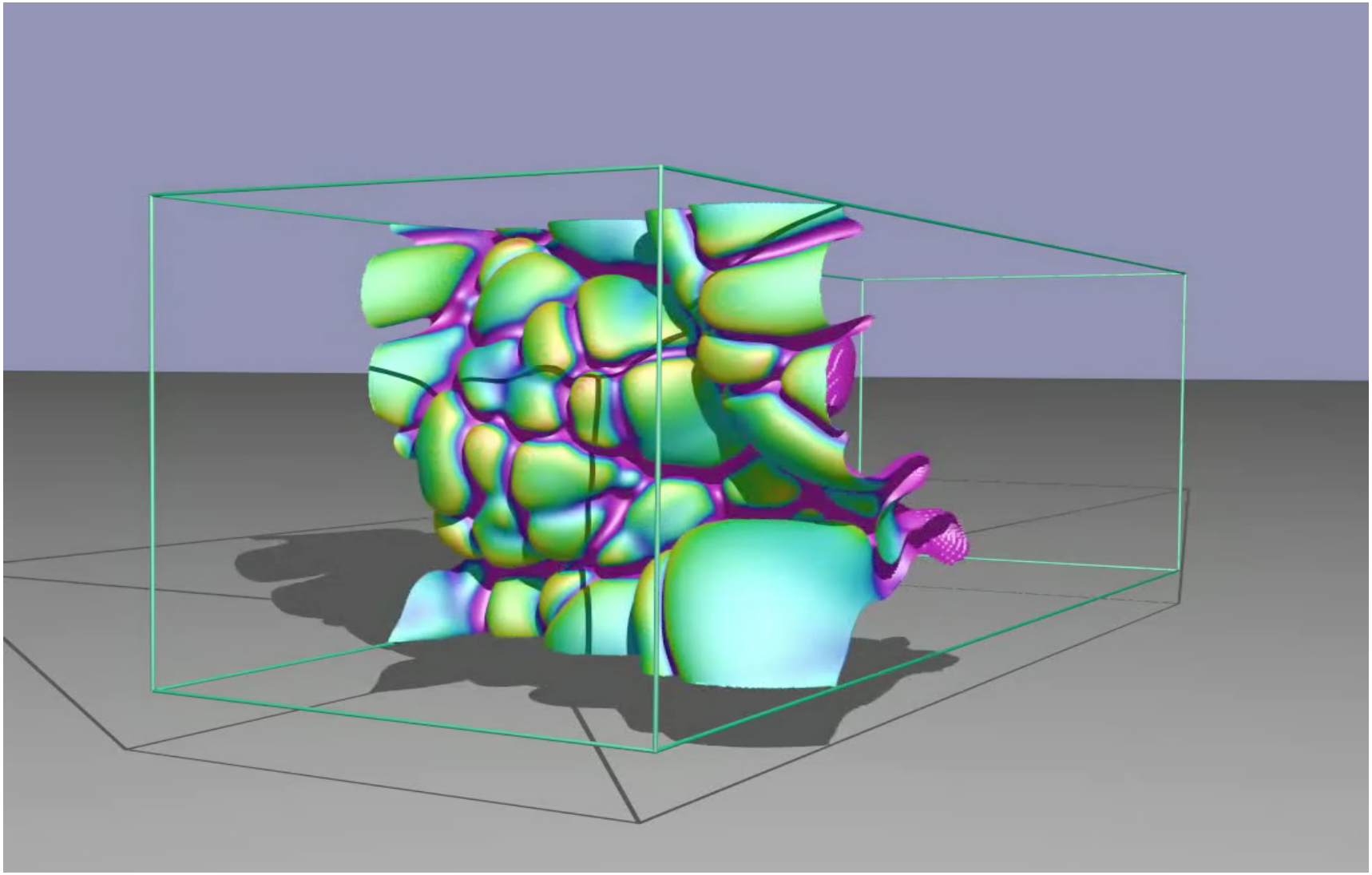
We Take on the Challenge of Developing a Quantitative Analysis Detecting Turbulence



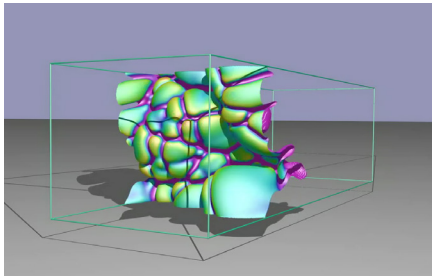
Understanding combustion processes over a broad range of burning conditions is an important problem for designing engines and power plants.

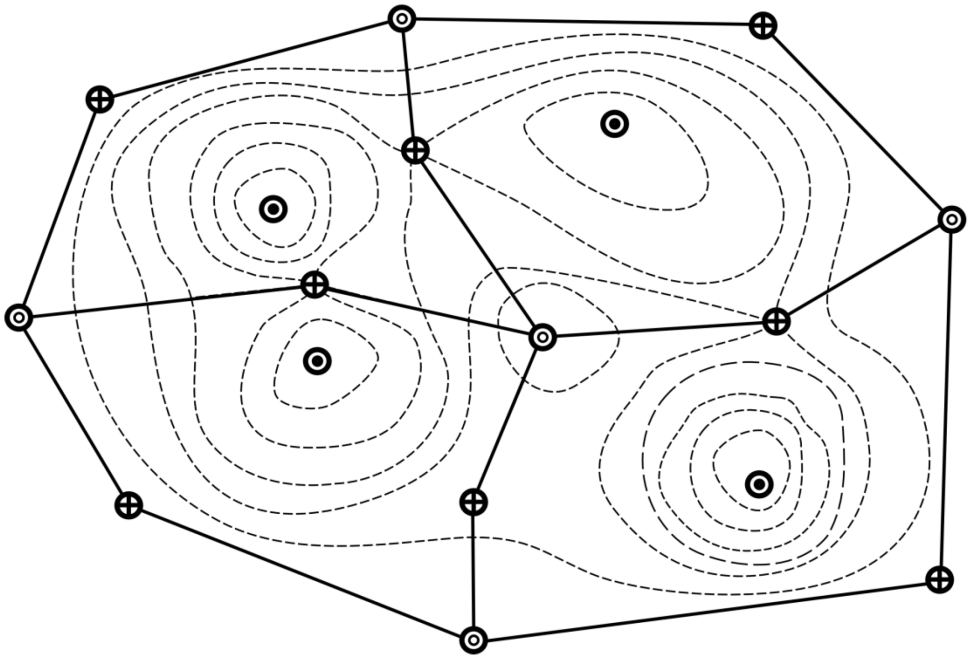
- **Simulation with AMR mesh.**
- **Simulations of lean premixed hydrogen flames with three degrees of turbulence.**
- **Can we identify precisely and track in time burning regions?**
- **Can we discriminate the degree of turbulence from a quantitative analysis?**

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



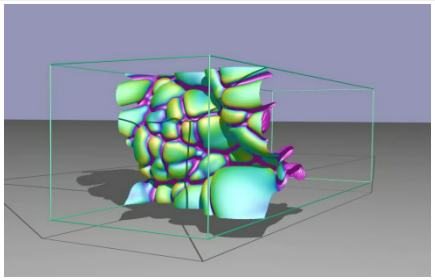


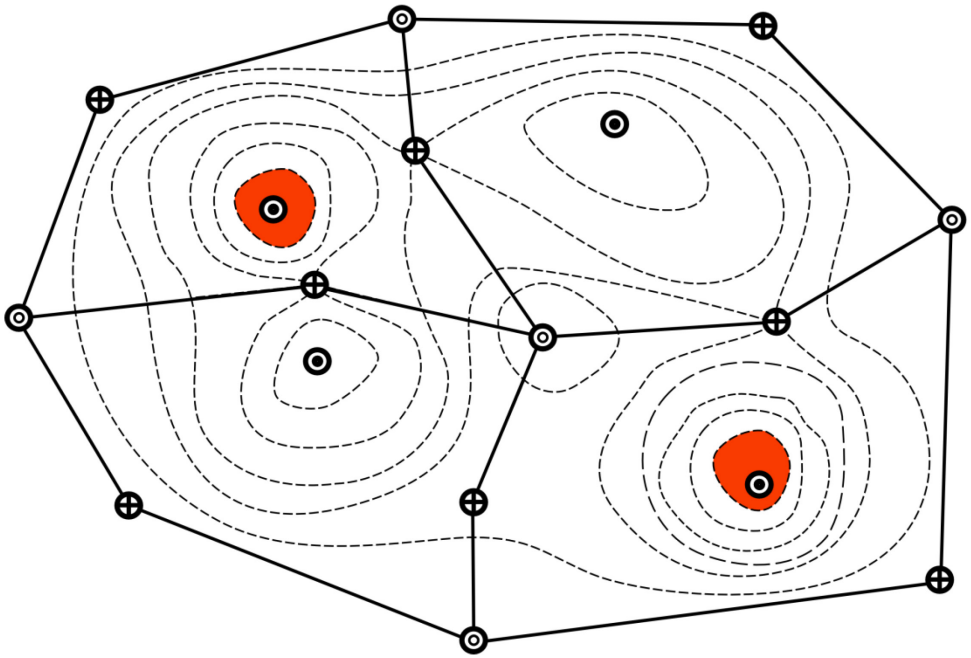


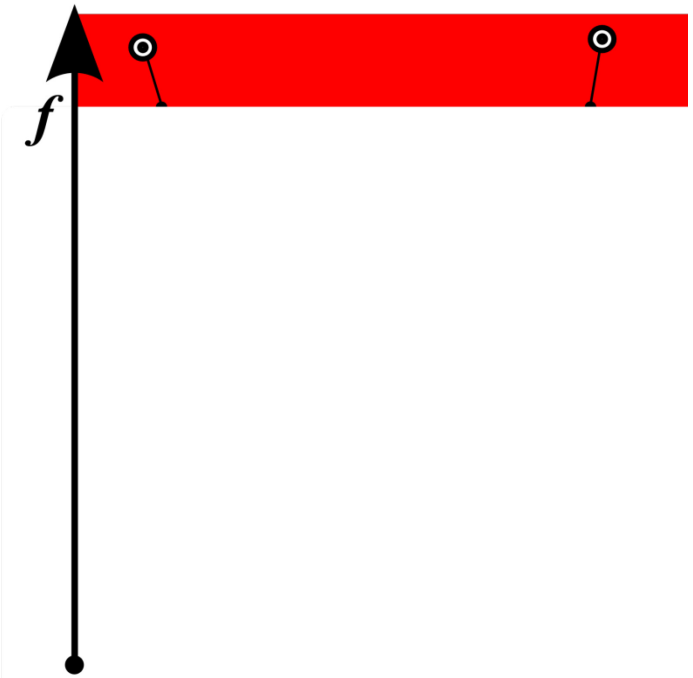
domain

hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



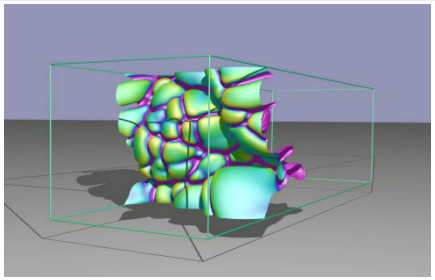


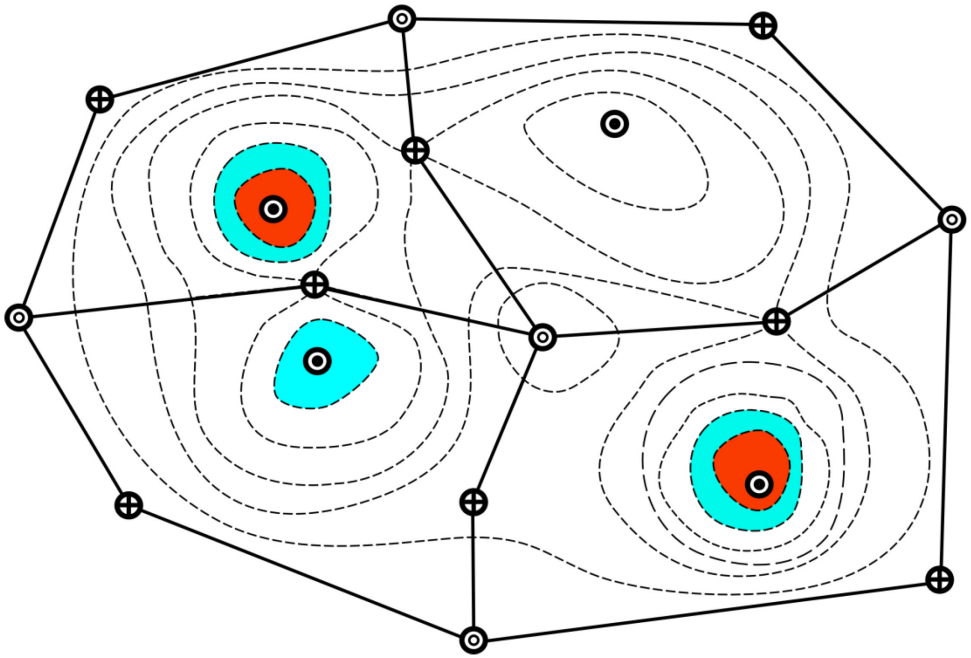


domain

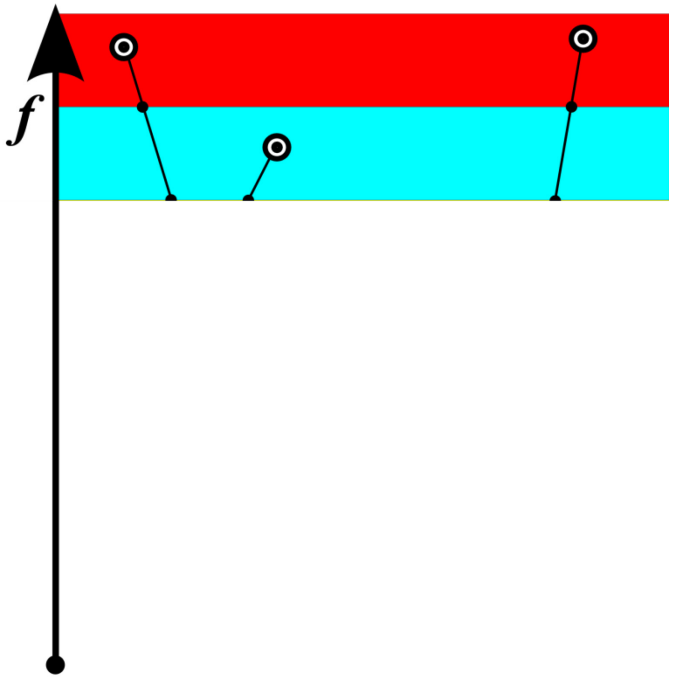
hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



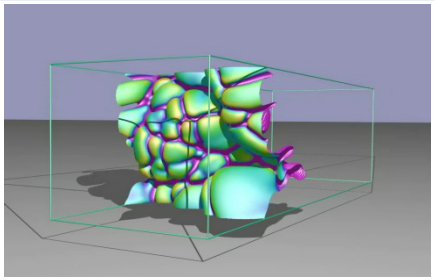


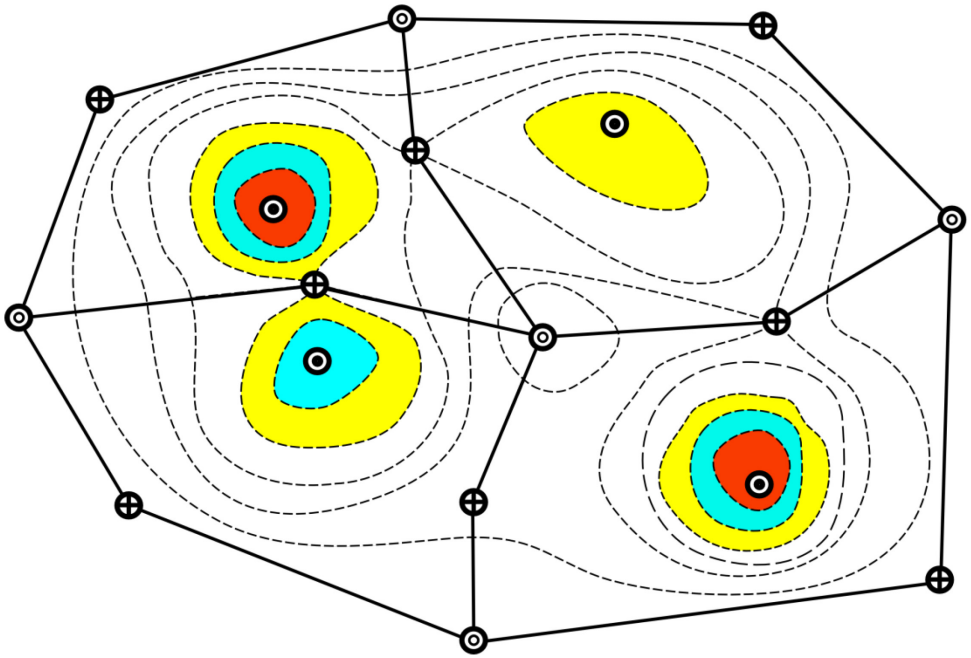
domain

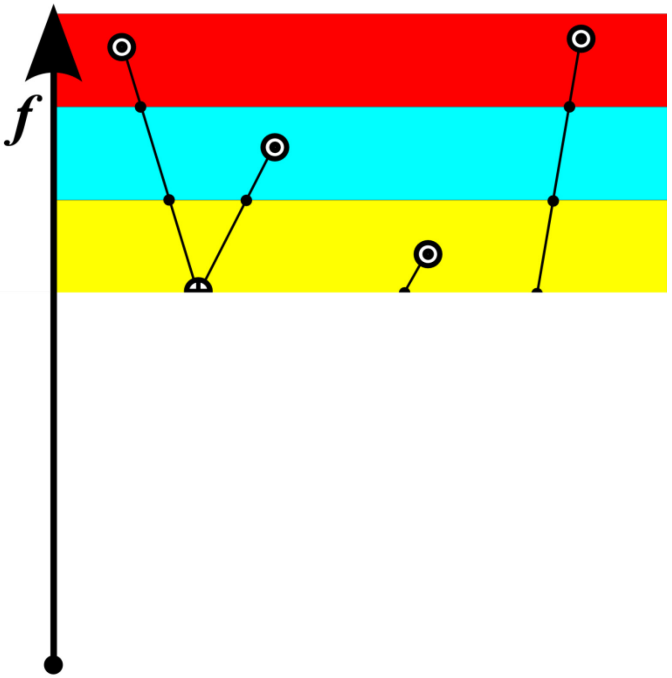


hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



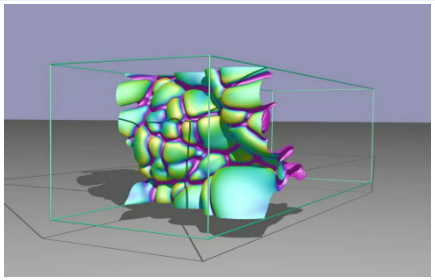


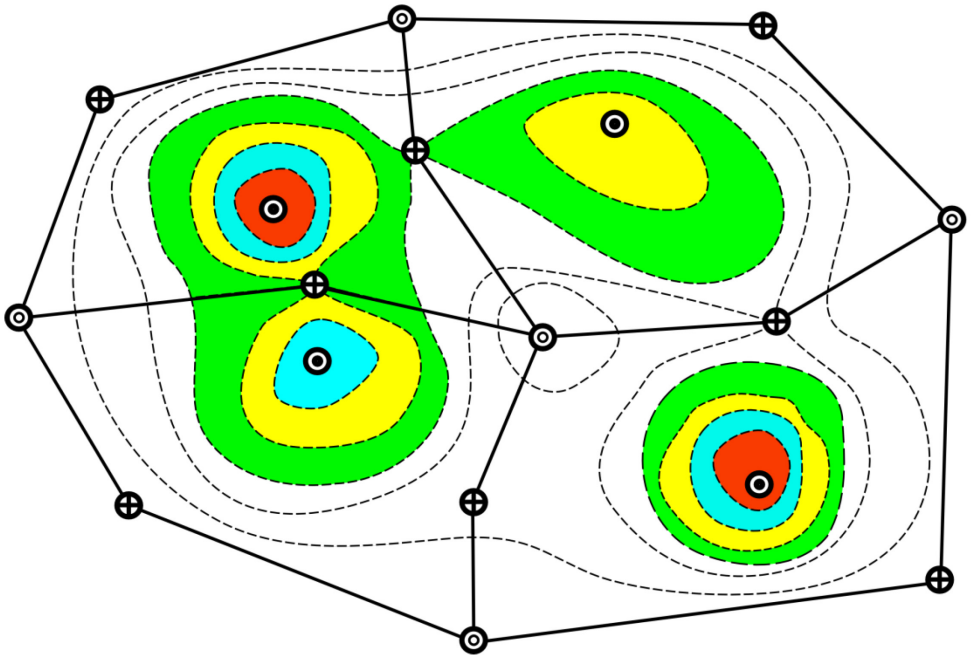


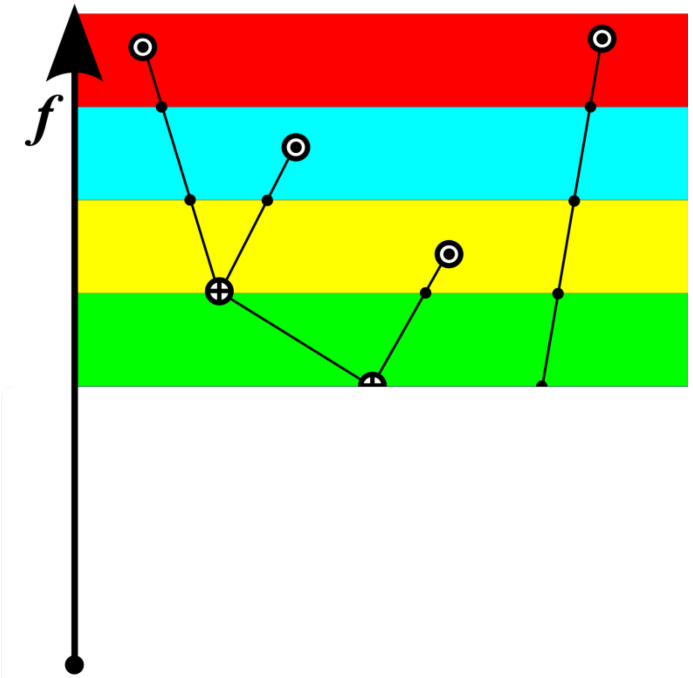
domain

hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



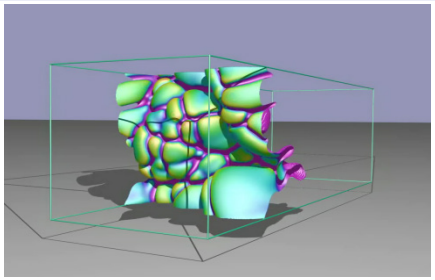


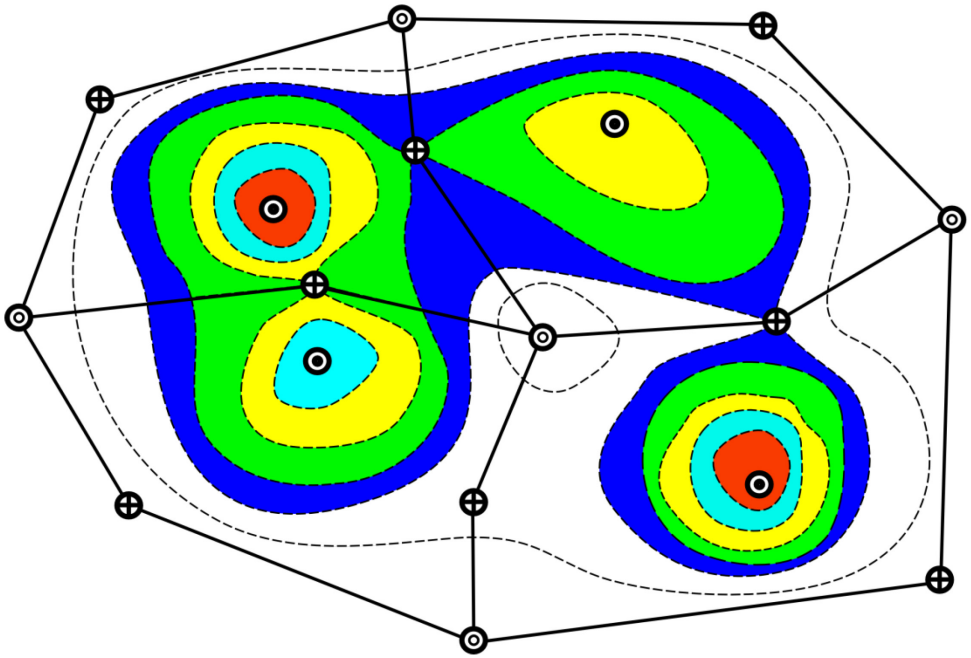


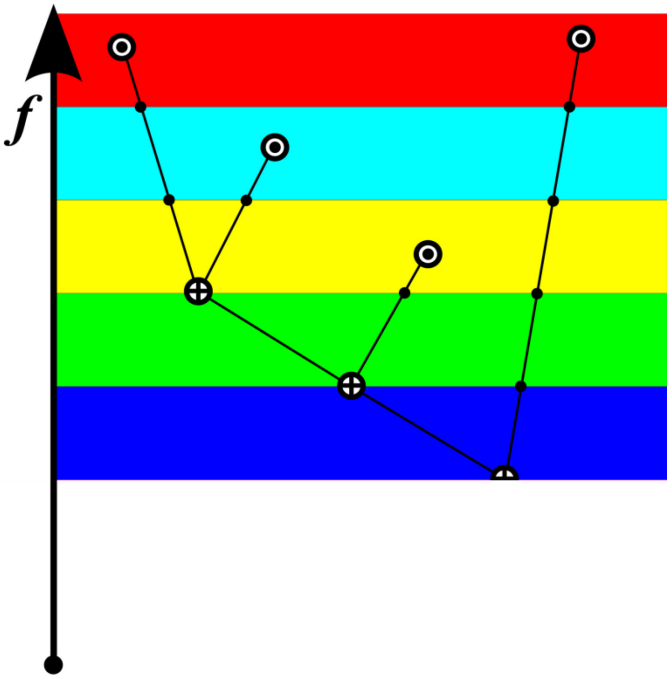
domain

hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



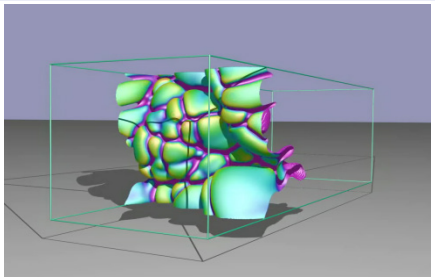


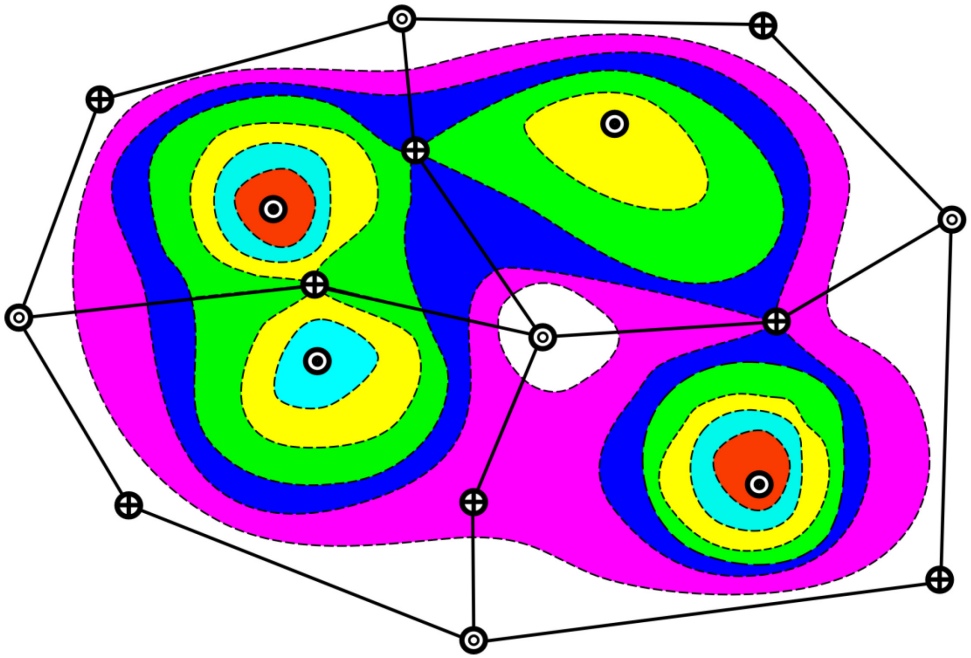


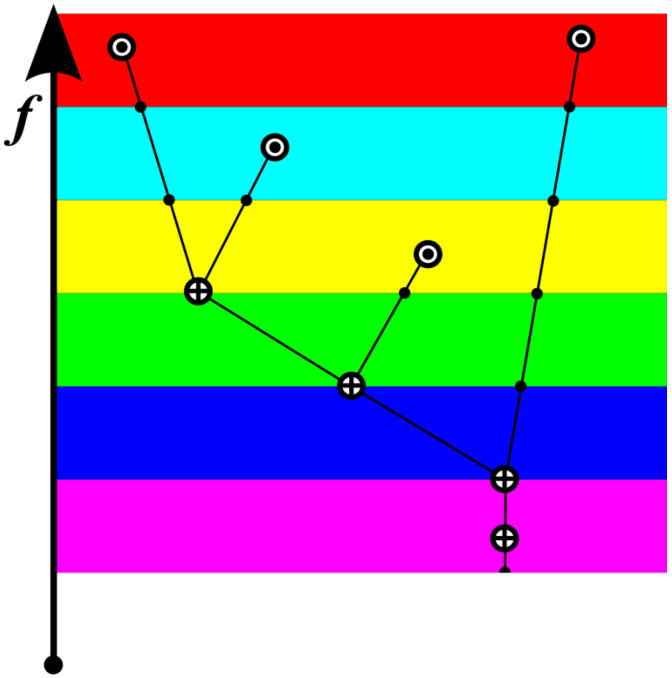
domain

hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



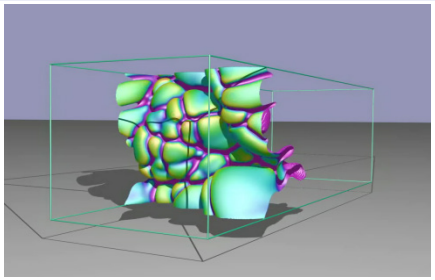


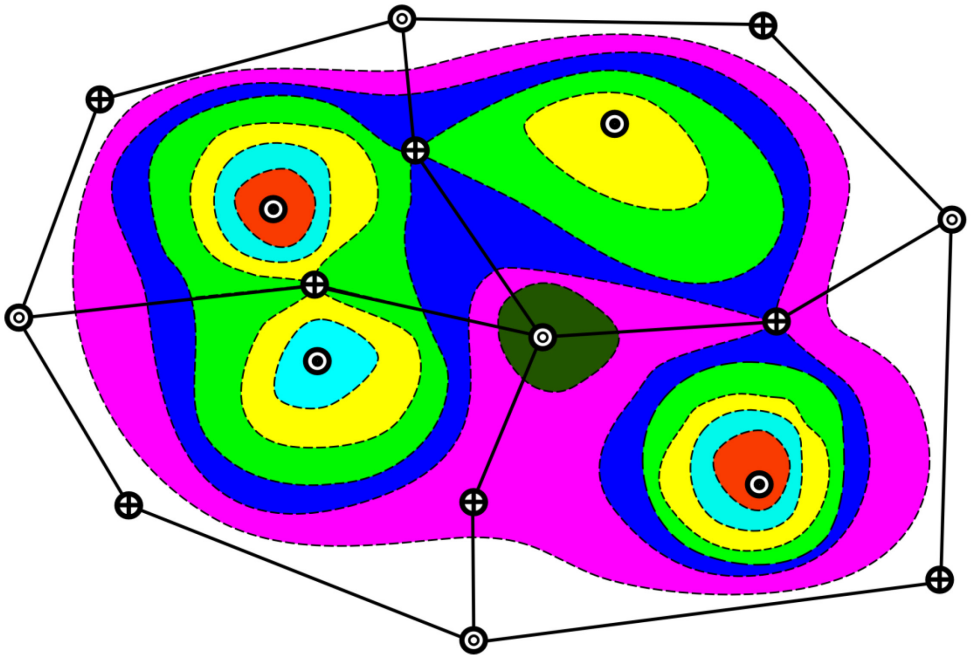


domain

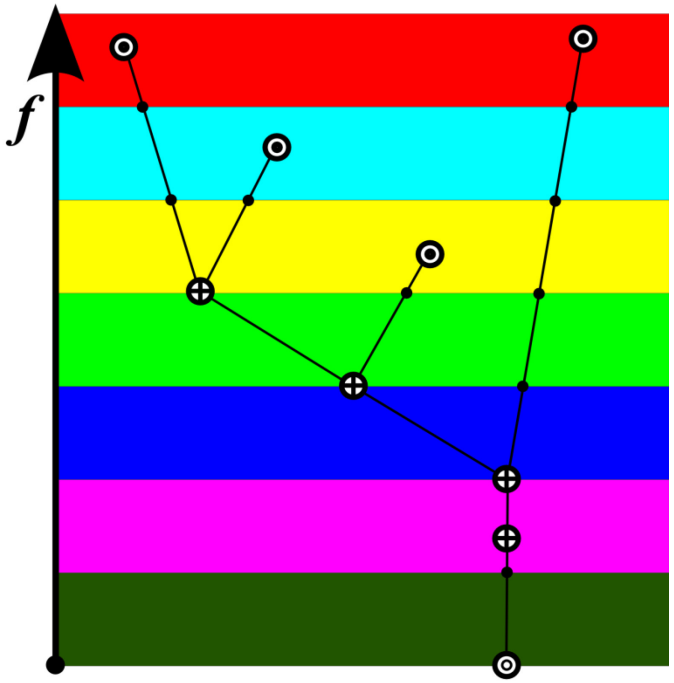
hierarchy

We Build a Reduced Topological Model of H2 Consumption on an Isotermal Surfaces



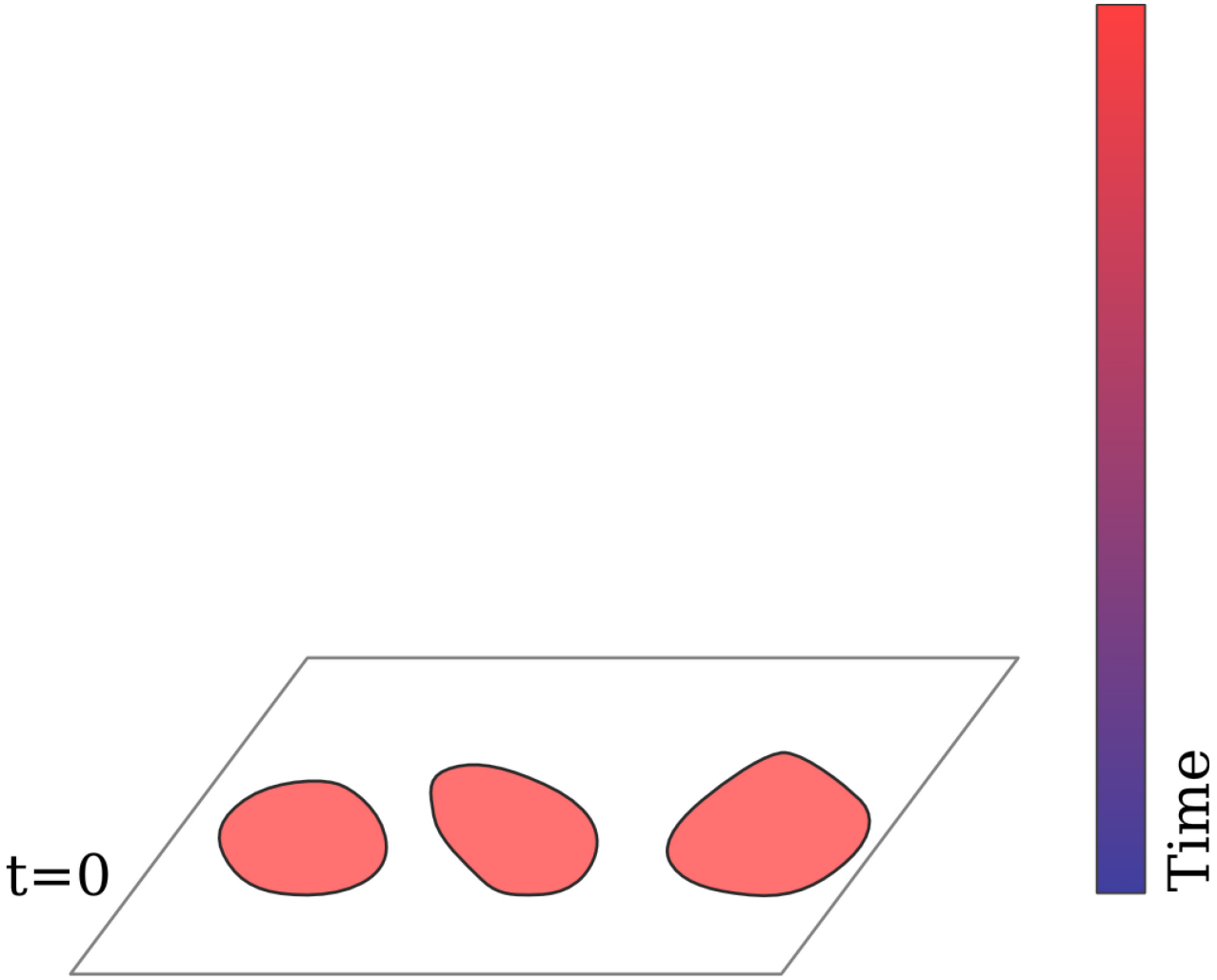


domain

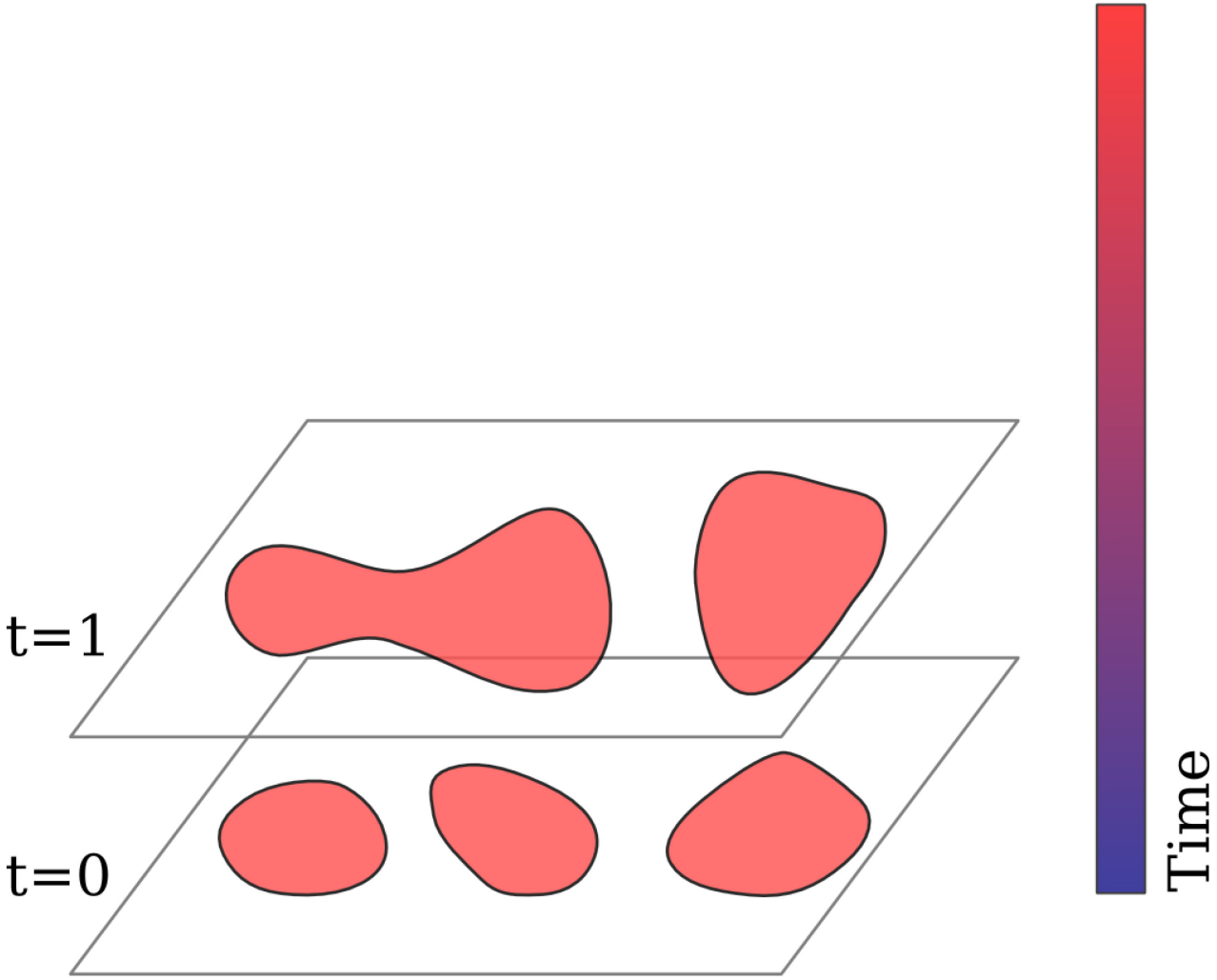


hierarchy

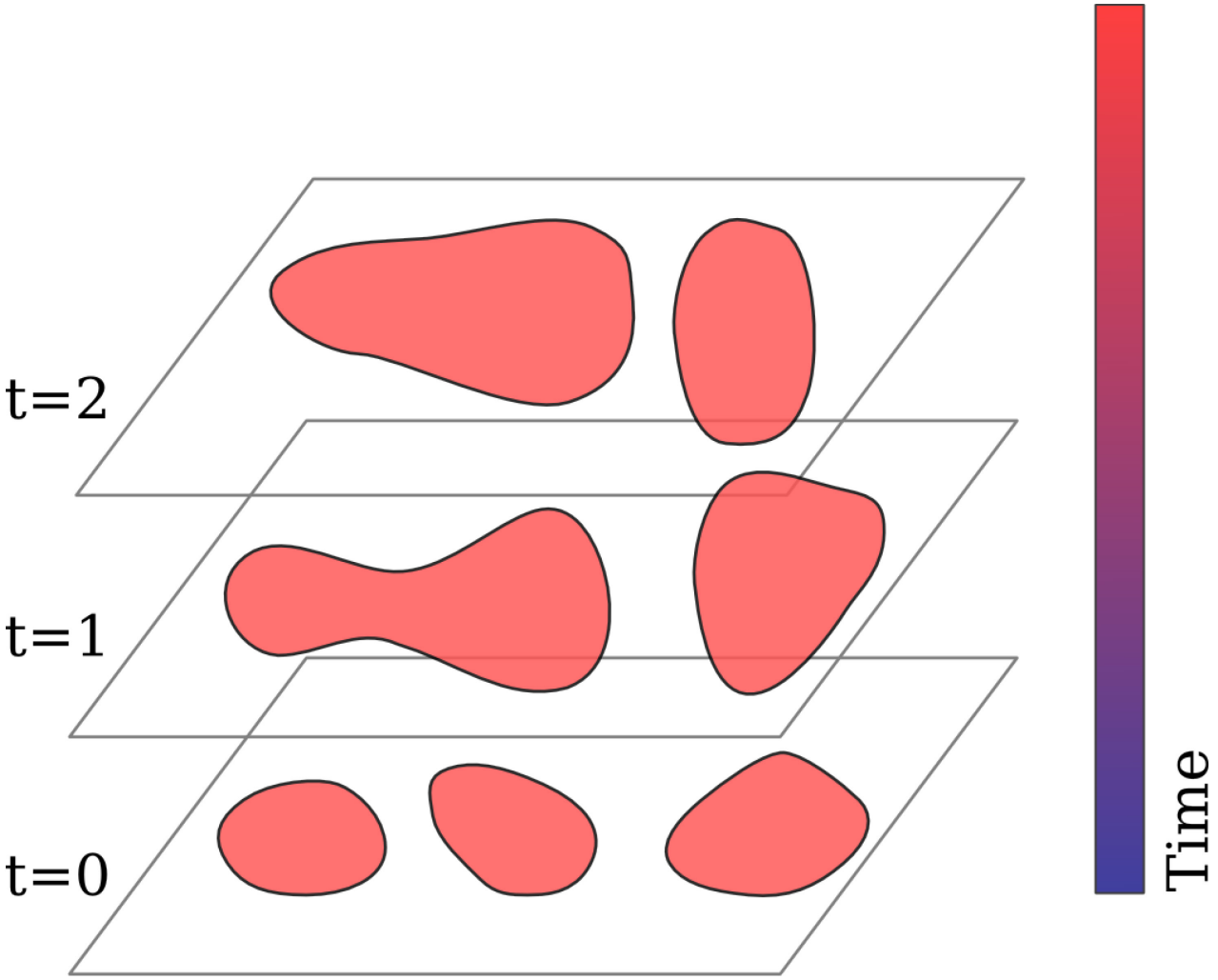
We track in time by interpolation in 4D and contraction to a Reeb Graph



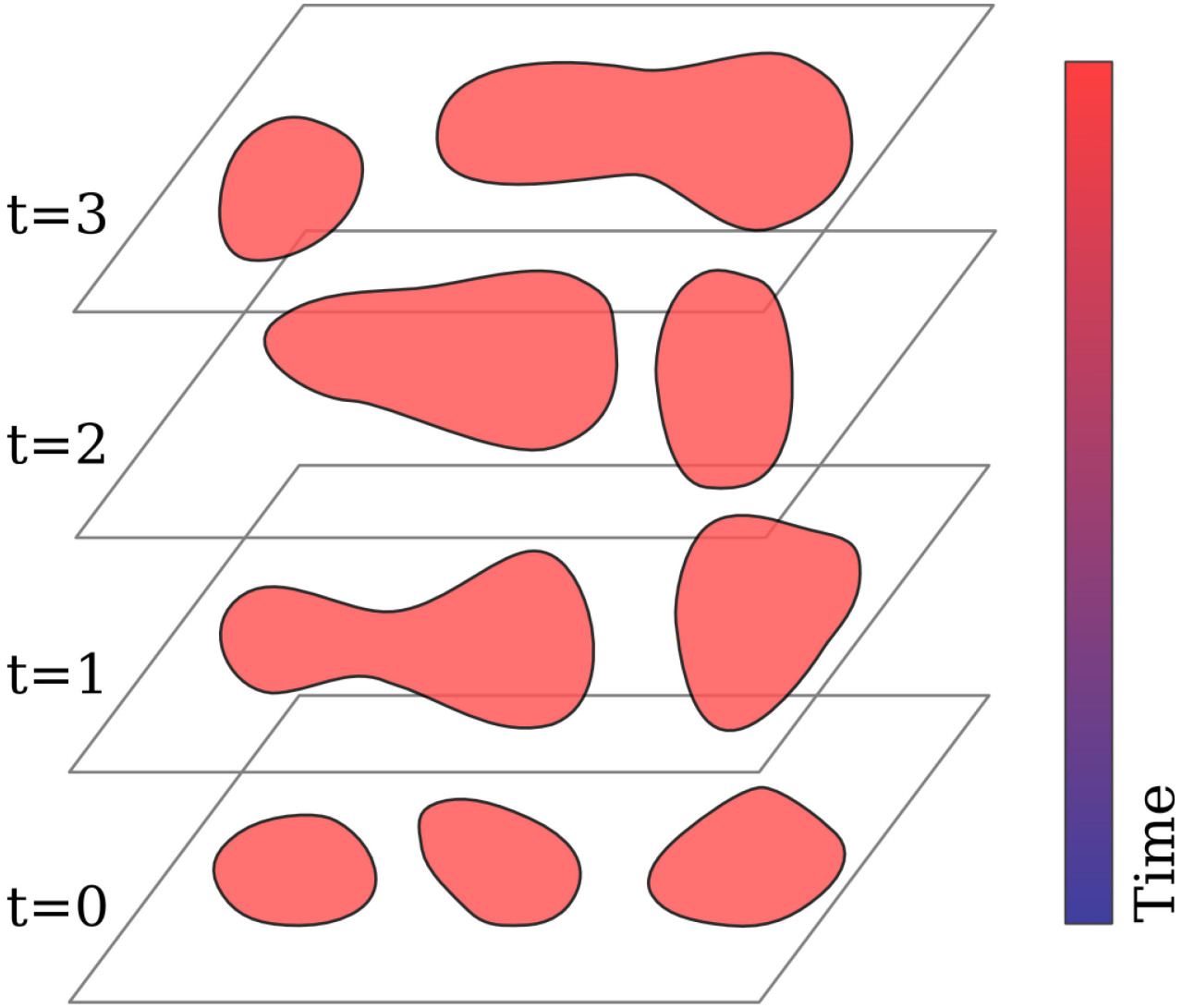
We track in time by interpolation in 4D and contraction to a Reeb Graph



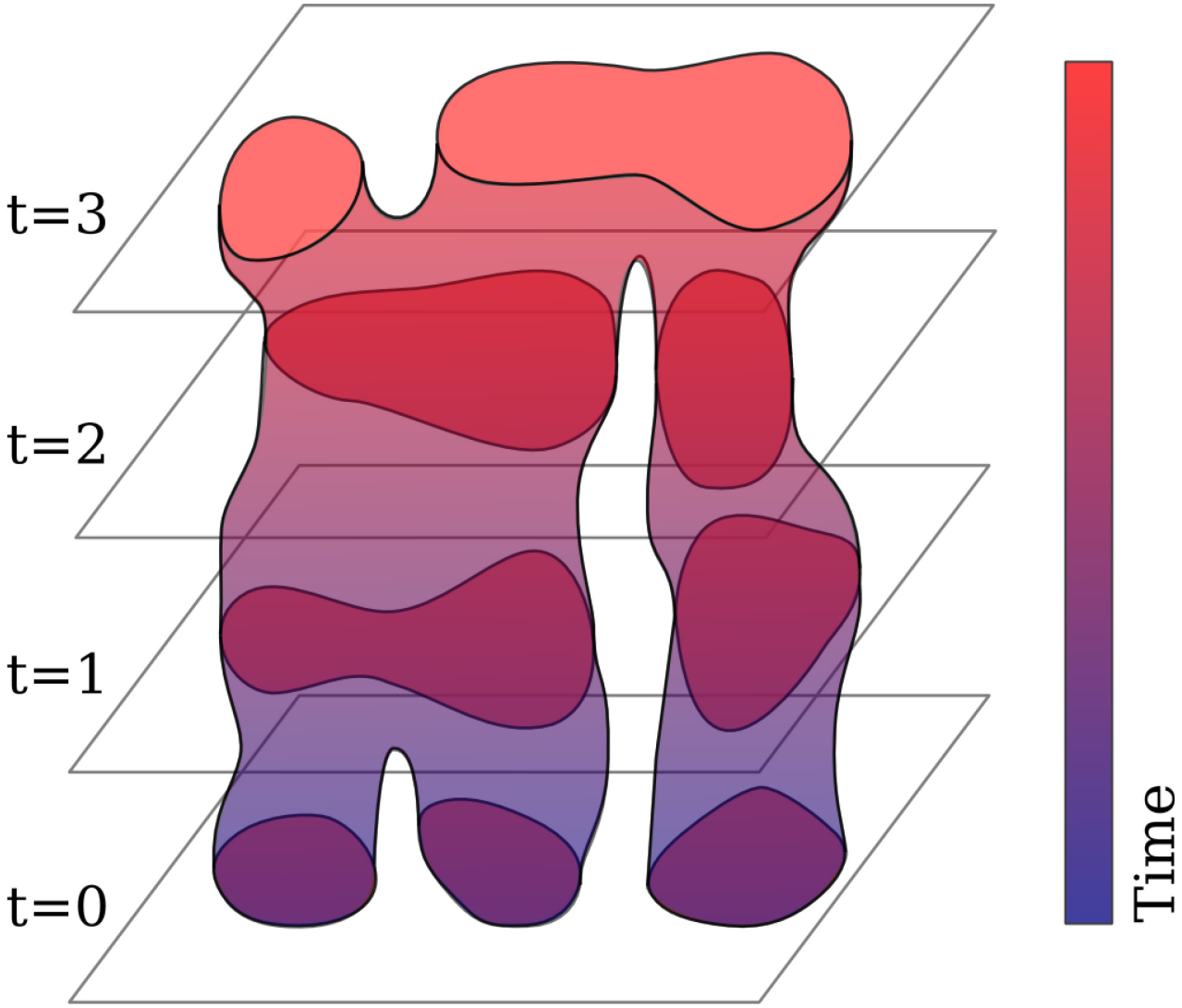
We track in time by interpolation in 4D and contraction to a Reeb Graph



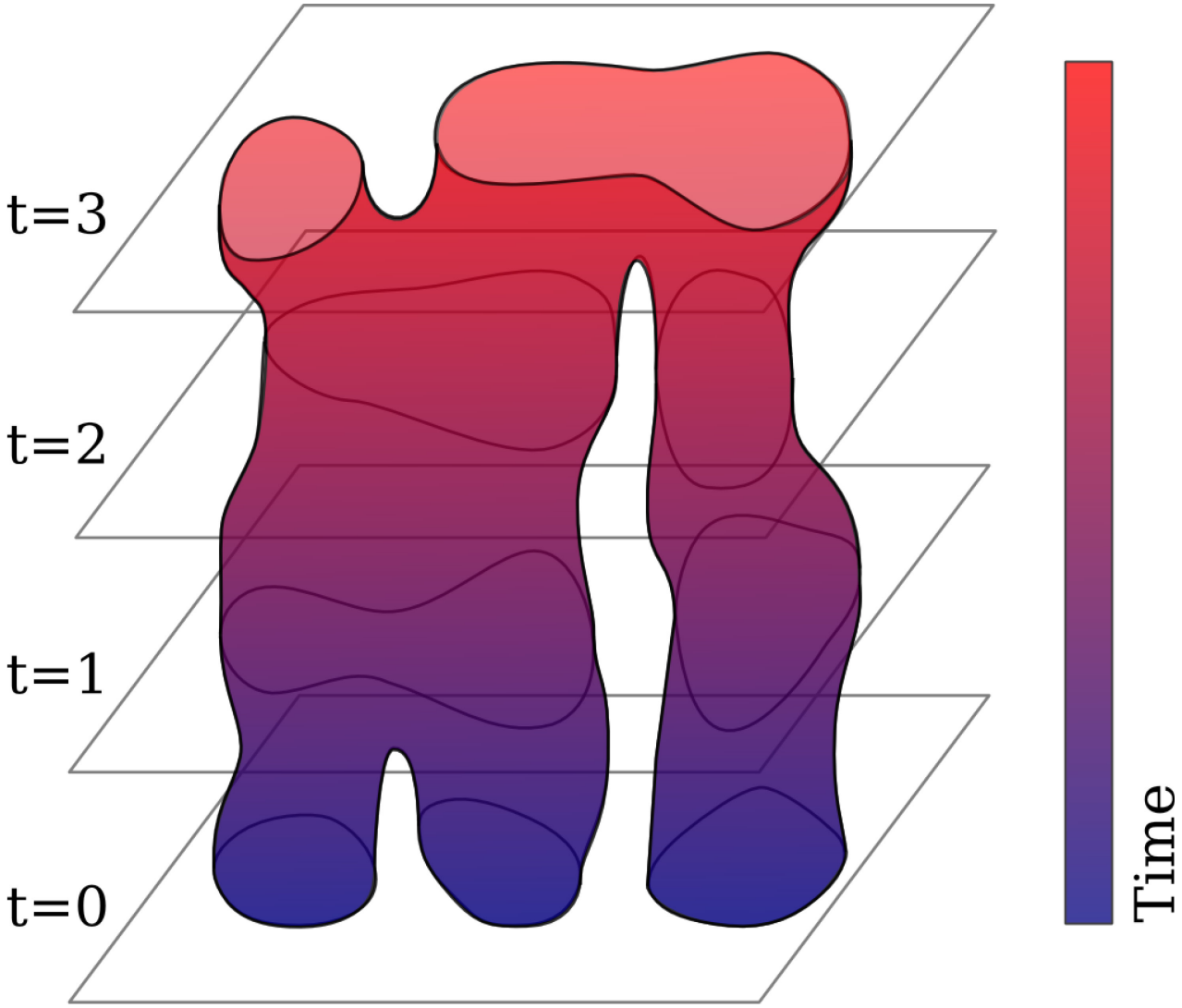
We track in time by interpolation in 4D and contraction to a Reeb Graph



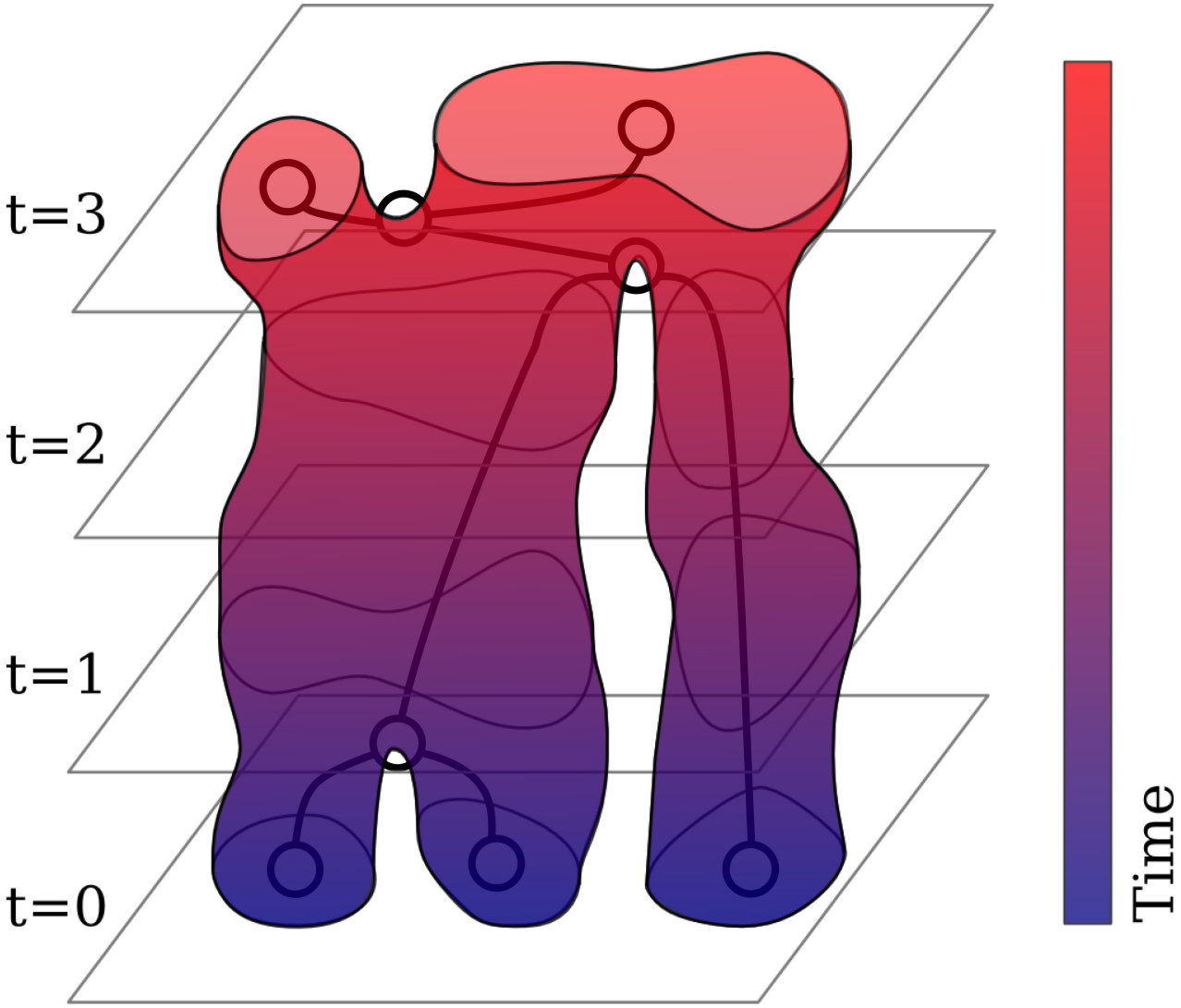
We track in time by interpolation in 4D and contraction to a Reeb Graph



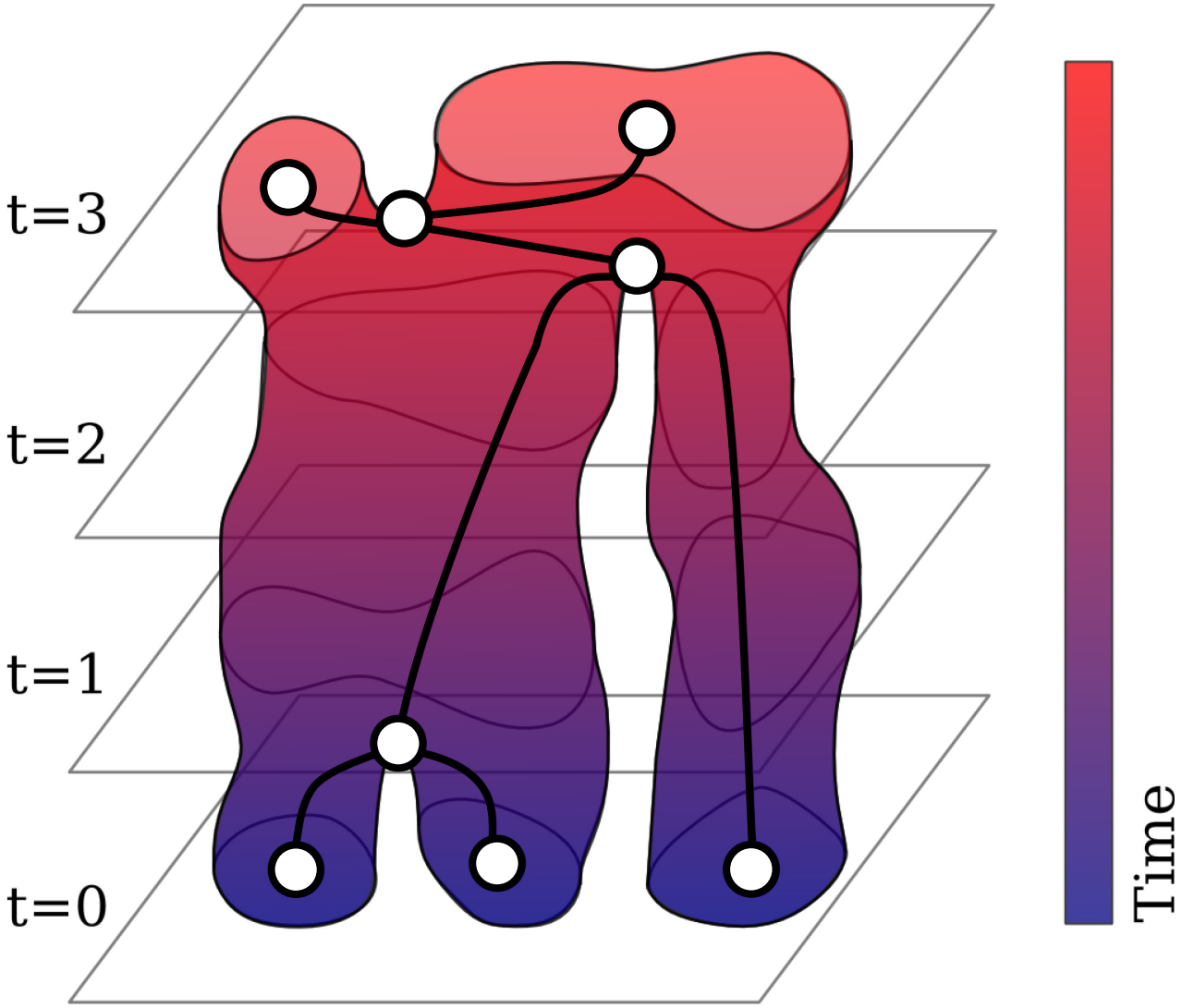
We track in time by interpolation in 4D and contraction to a Reeb Graph



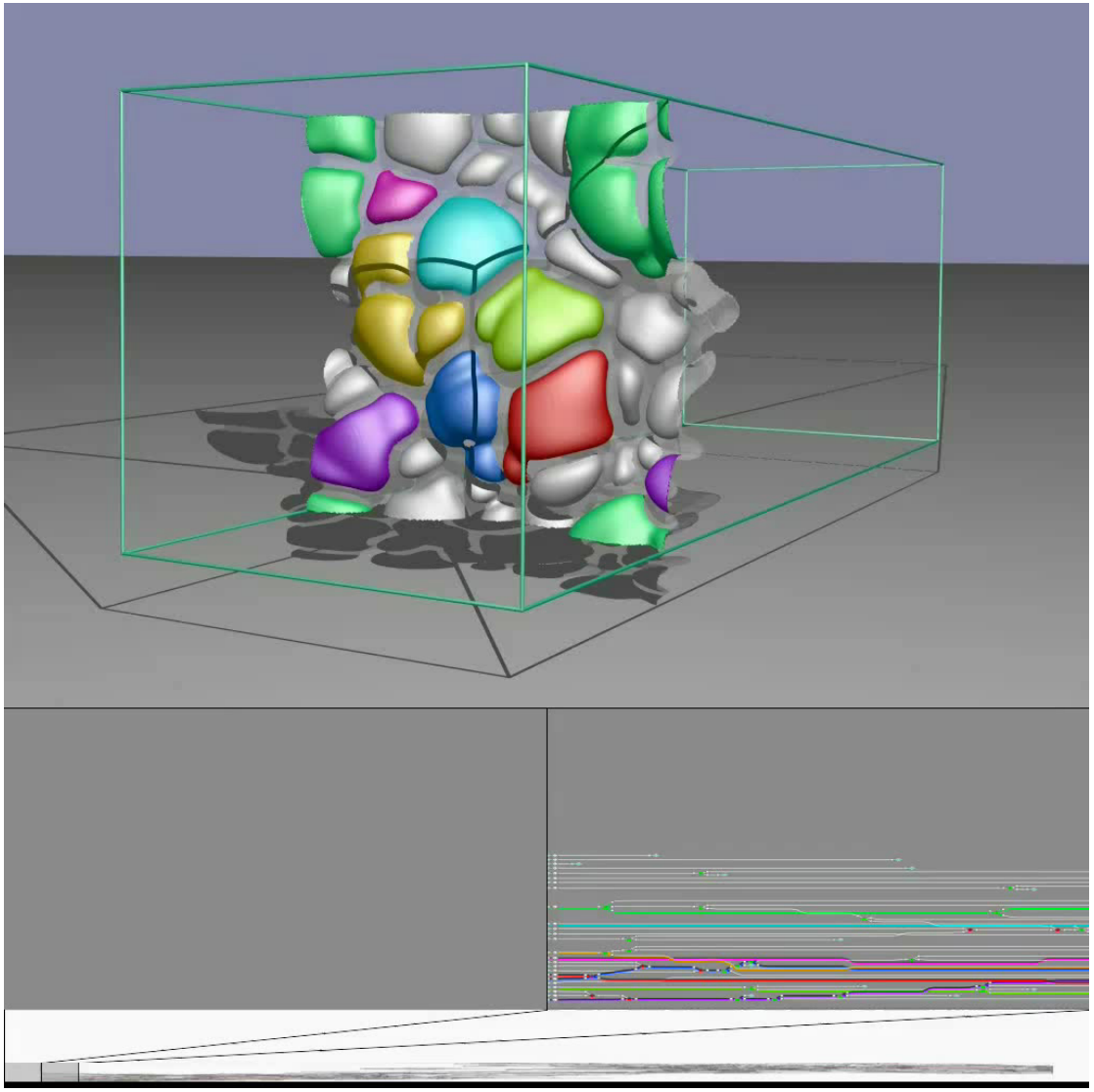
We track in time by interpolation in 4D and contraction to a Reeb Graph



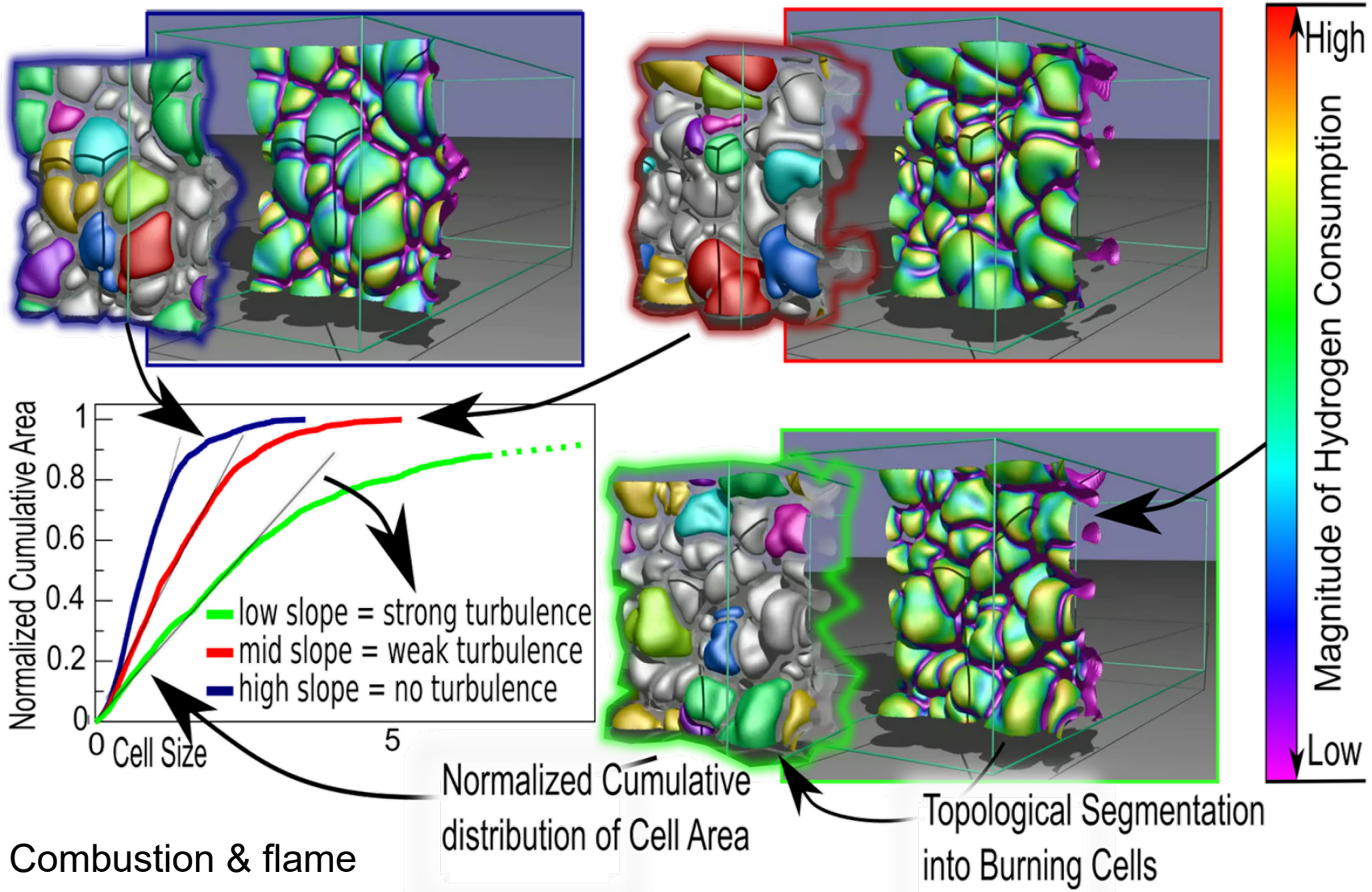
We track in time by interpolation in 4D and contraction to a Reeb Graph



Each Set of Parameters Results in a Robust Segmentation and Tracking of Burning Cells

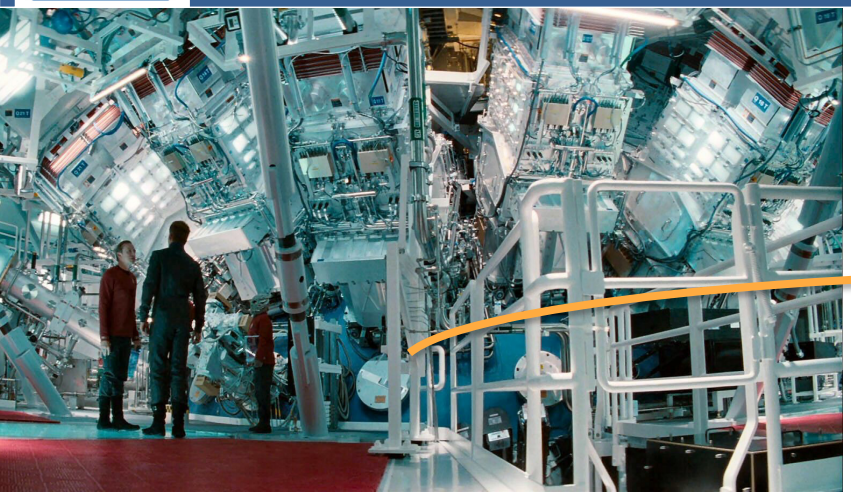


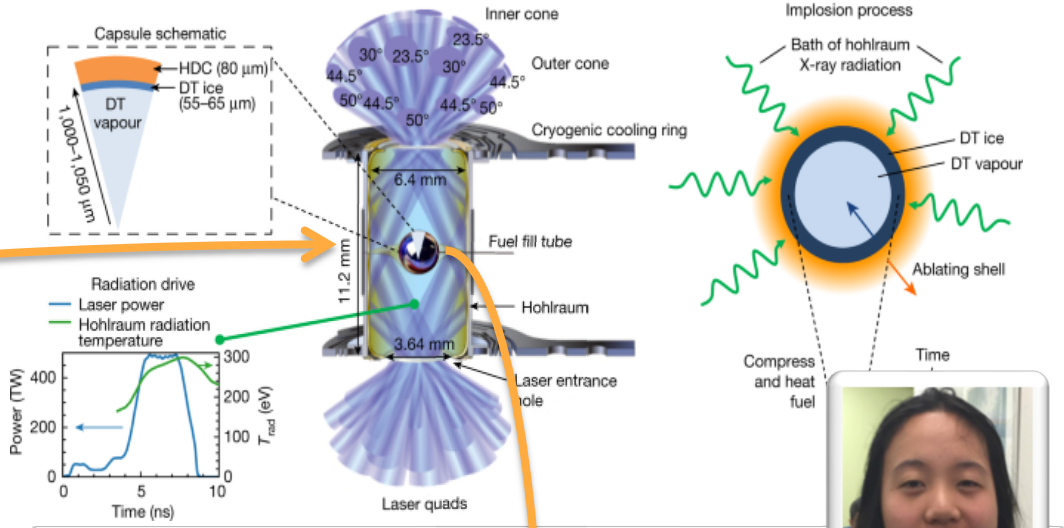
Topological Segmentation Allows to Quantify Turbulence as Slope of the Area Distributions



- Combustion & flame

LLNL Deployment for High-Resolution Metrology of Additive Manufacturing Parts, NIF Targets, Energetic Materials, …





Xuan Huang

- High-resolution CT scans provide insights into internal structures
  - Additive manufacturing
  - NIF targets: Voids and high Z inclusions that disturb the implosion
  - High energy materials: Material aging and performance characteristics
- These applications pose a number of new challenges
  - Massive collections of large (2Kx2Kx6K and above) scans exceeding local resources
  - Diverse stakeholders: CT technicians, scientists, decision-makers, etc.
  - Distributed teams and data collection
  - Data access restrictions

How do we explore a family of combustions to design new types of fuels?

HIGH DIMENSIONAL DATA EXPLORATION

Intuition in 3D is Very Strong: Avoiding Misleading Visualization in 3D is Very Easy

Intuition in 3D is Very Strong: Avoiding Misleading Visualization in 3D is Very Easy



Intuition in 3D is Very Strong: Avoiding Misleading Visualization in 3D is Very Easy



Intuition in 3D is Very Strong: Avoiding Misleading Visualization in 3D is Very Easy

Intuition in 3D is Very Strong: Avoiding Misleading Visualization in 3D is Very Easy



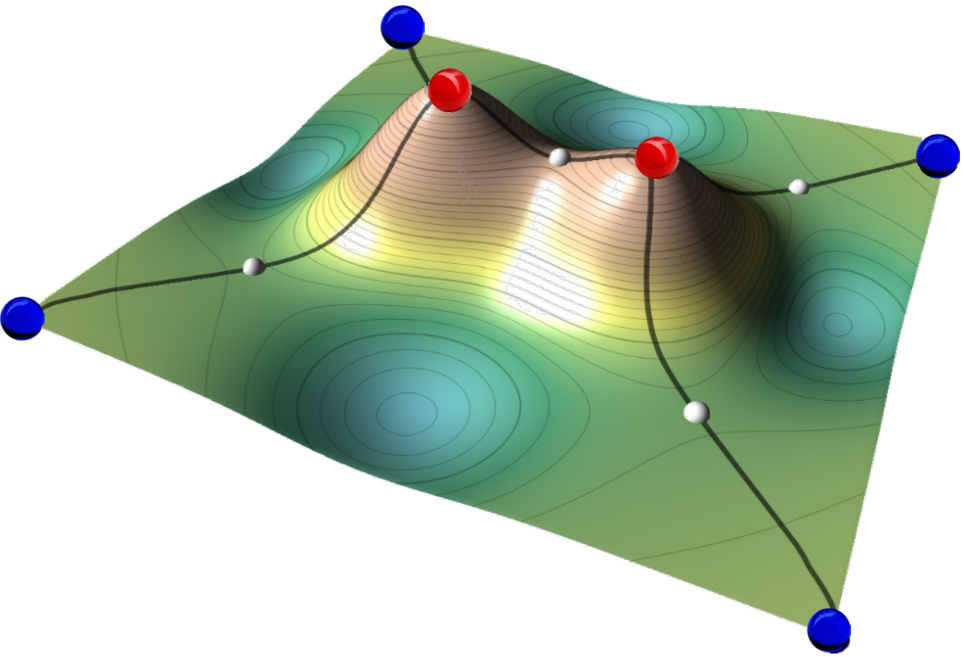
Intuition in 3D is Very Strong: Avoiding Misleading Visualization in 3D is Very Easy



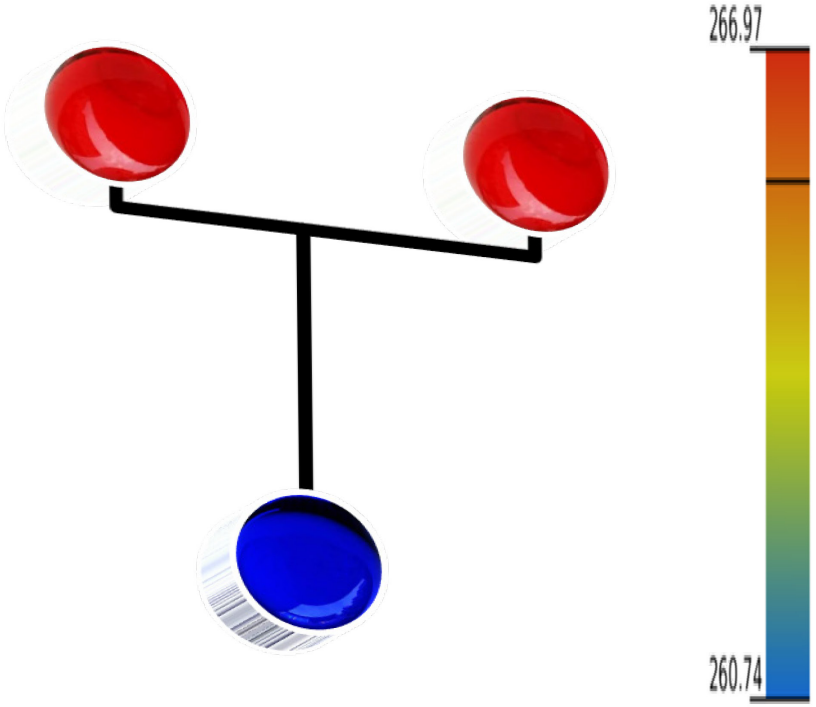
Lifting the Exploration from 2D/3D to High-Dimensional Complex Parameter Space

Geometric Domain + Merit Function

2D+1



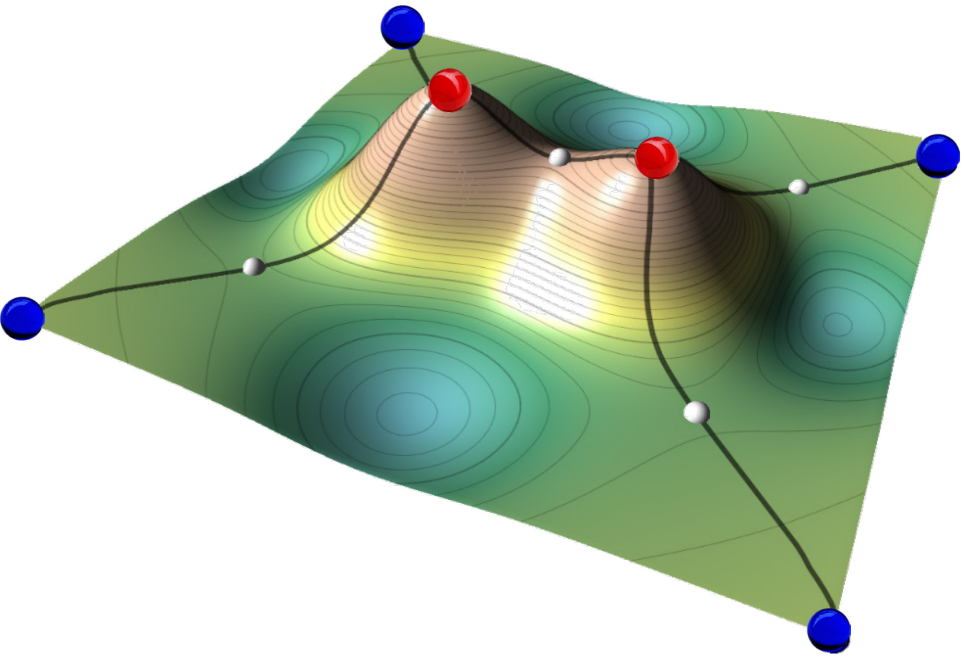
2 D+1



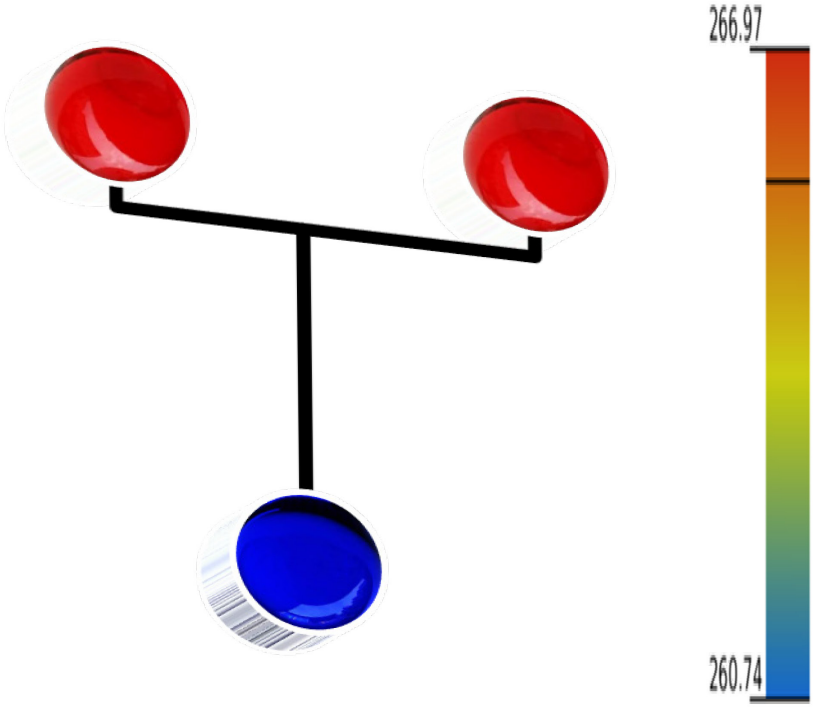
Lifting the Exploration from 2D/3D to High-Dimensional Complex Parameter Space

Geometric Domain + Merit Function

2D+1



21D+1

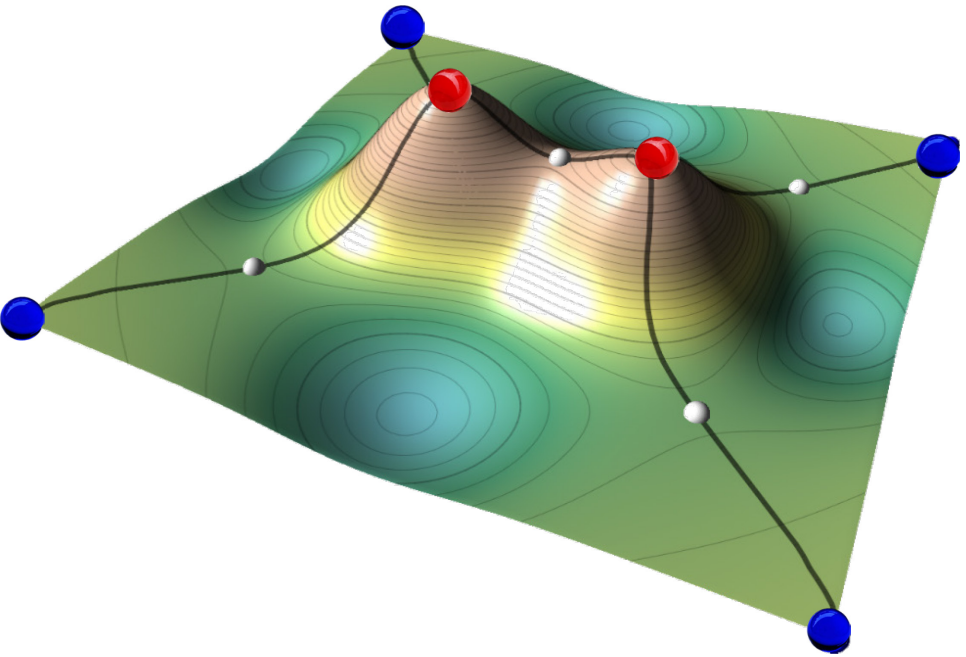


CEDMAV

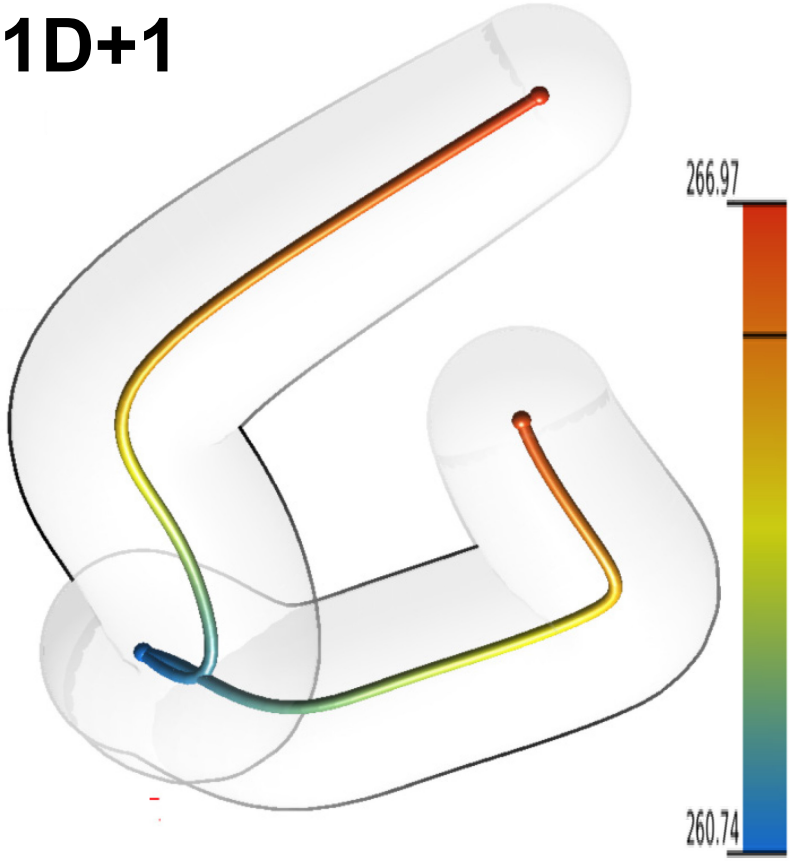
Lifting the Exploration from 2D/3D to High-Dimensional Complex Parameter Space

Geometric Domain + Merit Function

2D+1



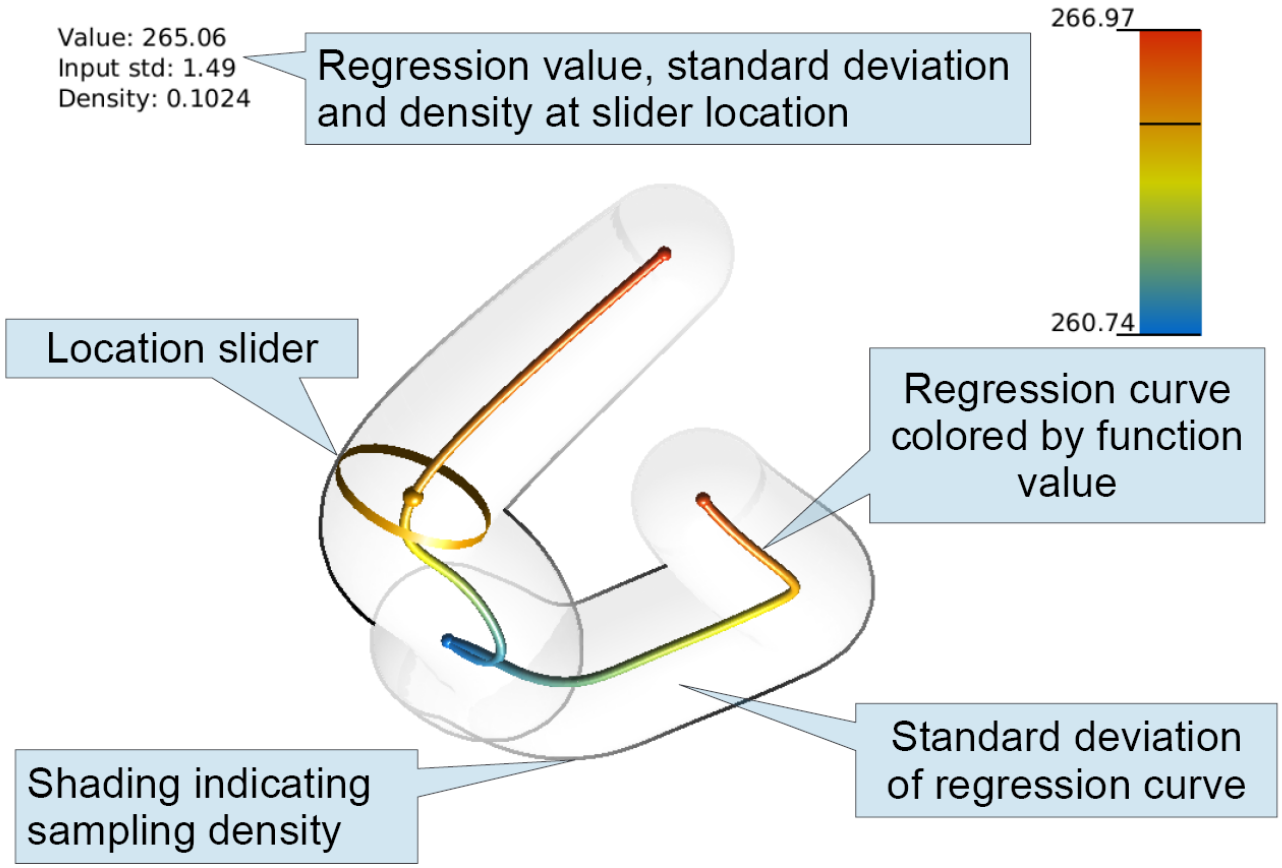
21D+1



CEDMAV

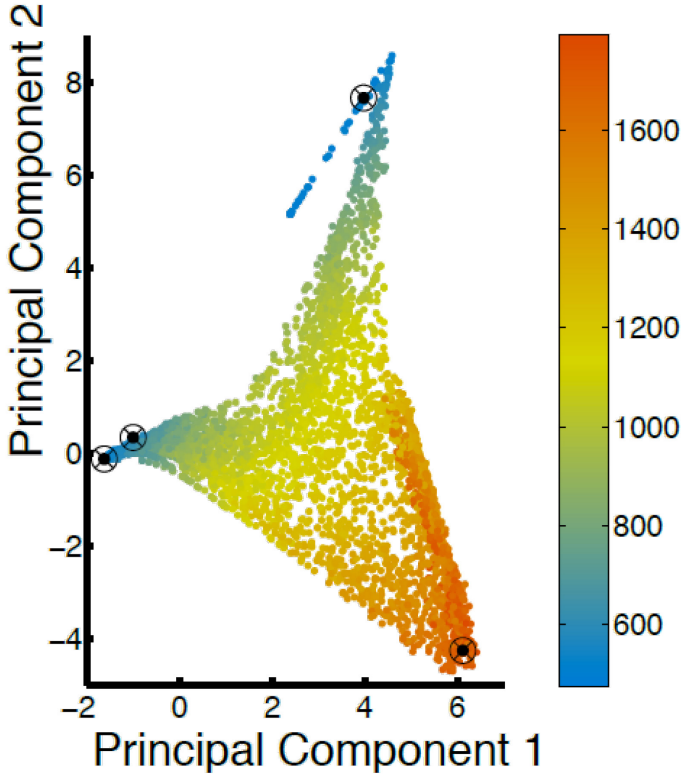
Exploration of High Dimensional Functions Data

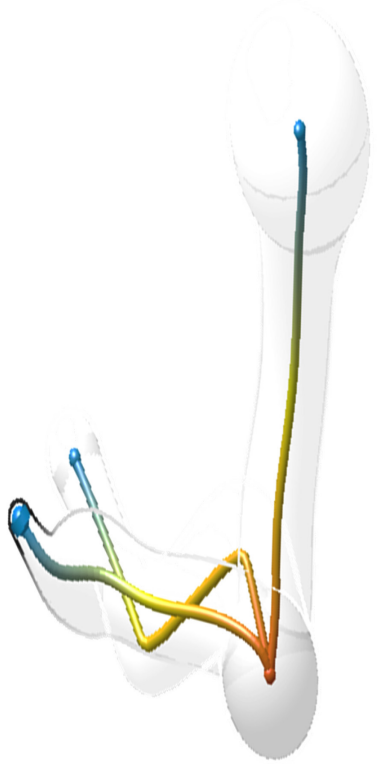
Integrated presentation of statistics and topology



Improve Techniques for Visualizing and Exploring High Dimensional Data

Reduce dimensionality
and then extract structure

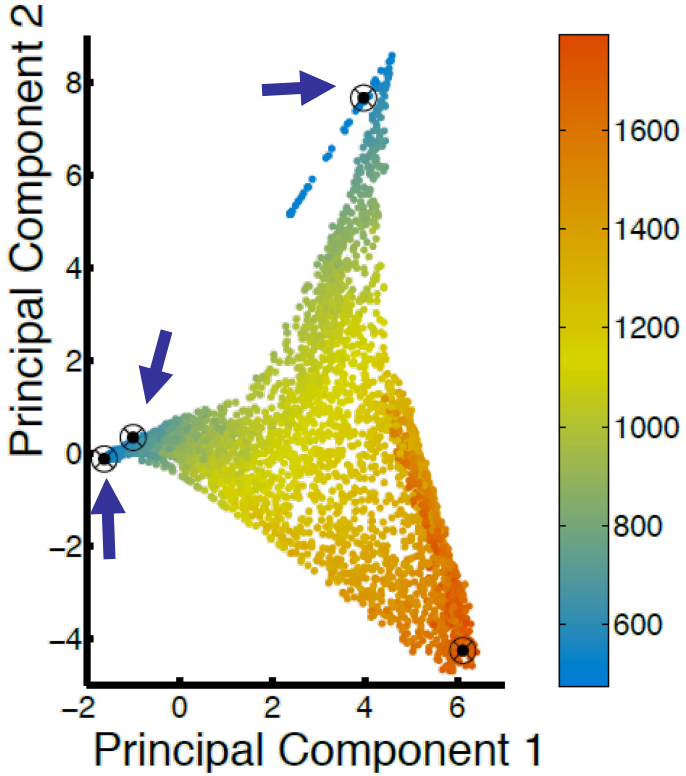


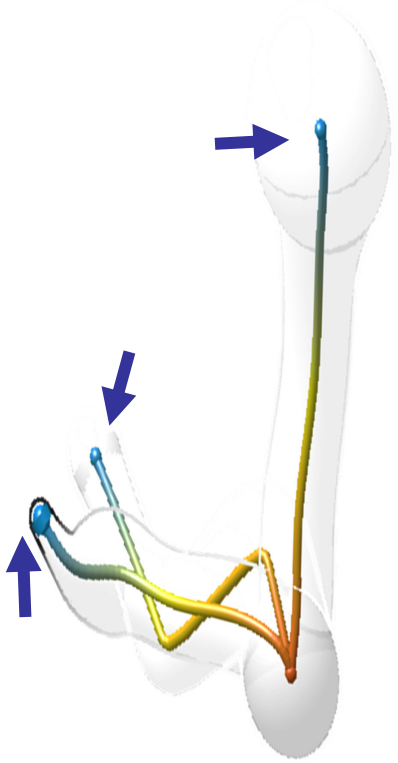


Gerber et al, 2010

Improve Techniques for Visualizing and Exploring High Dimensional Data

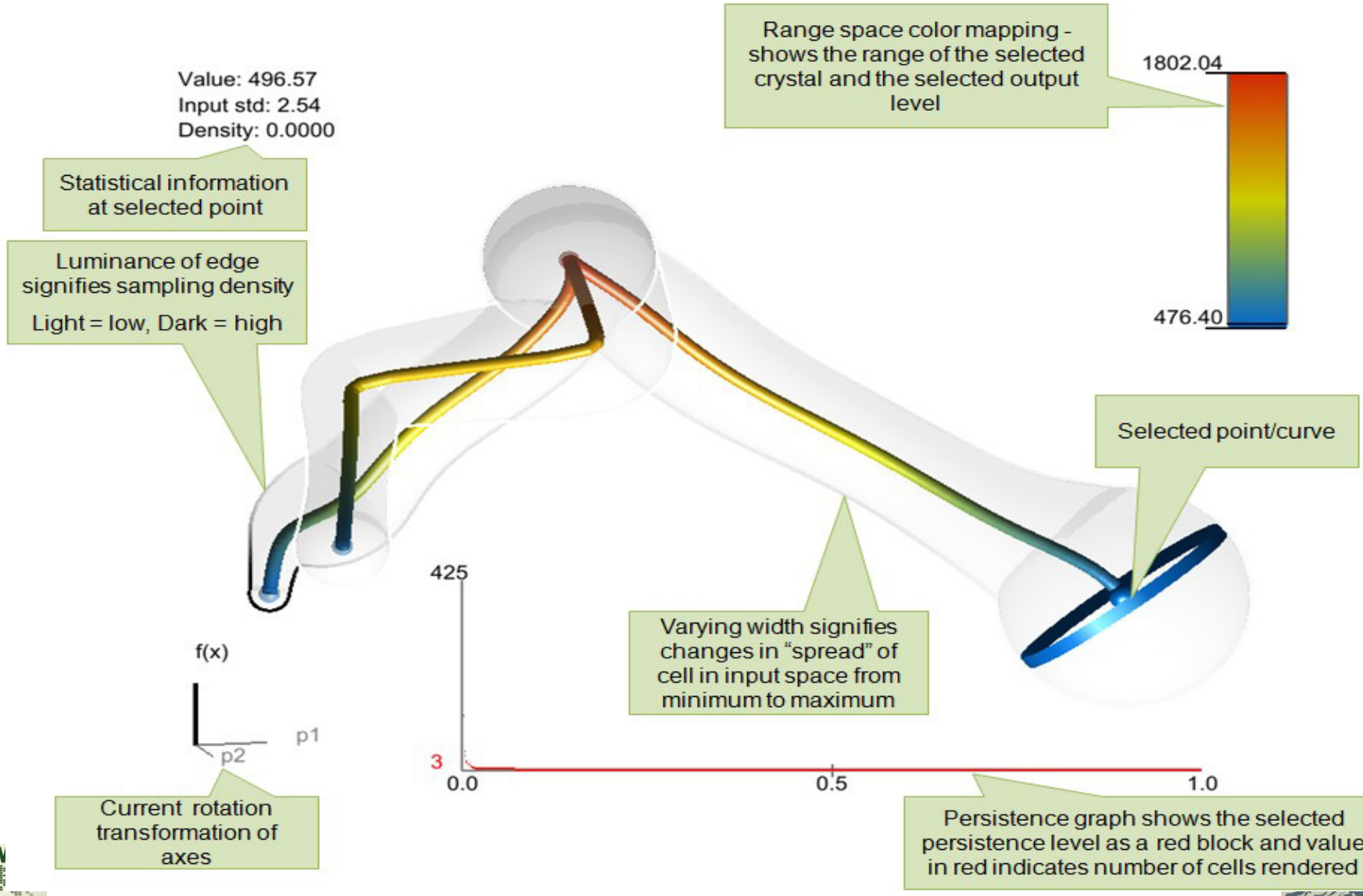
Reduce dimensionality
and then extract structure

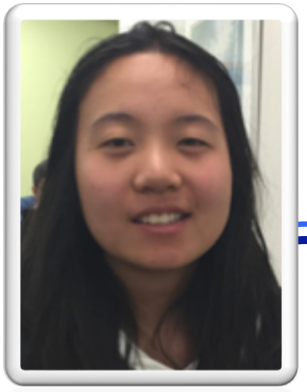




Gerber et al, 2010

Topological Summary





Xuan Huang

Interface: Statistical Summary

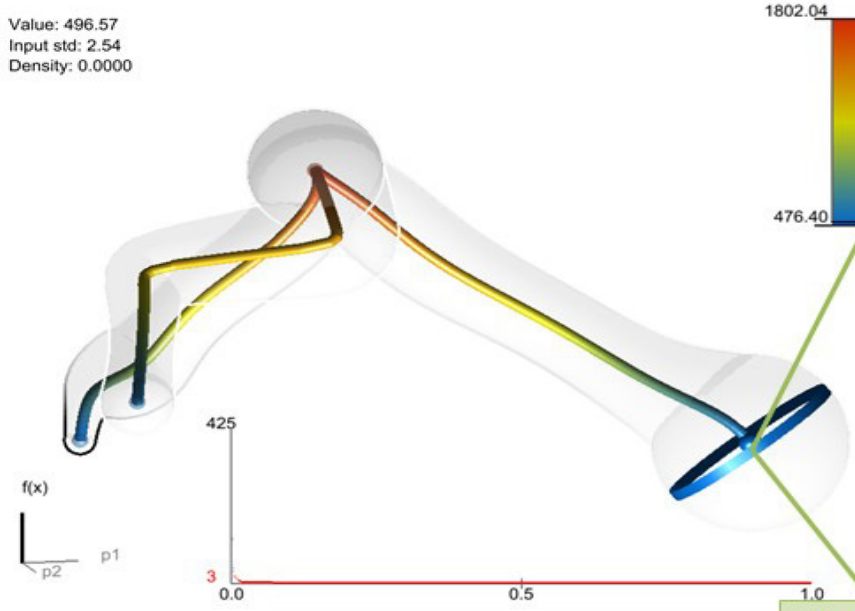
Named input coordinate

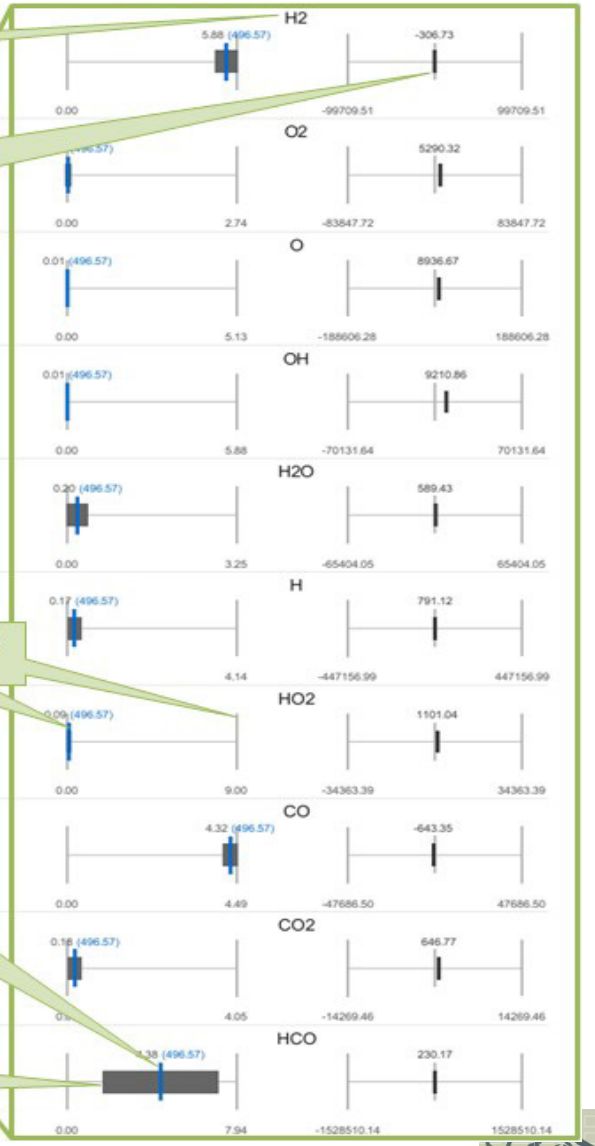
Encoded gradient information - the change in coordinate value with respect to the change in output value

Range of input coordinate

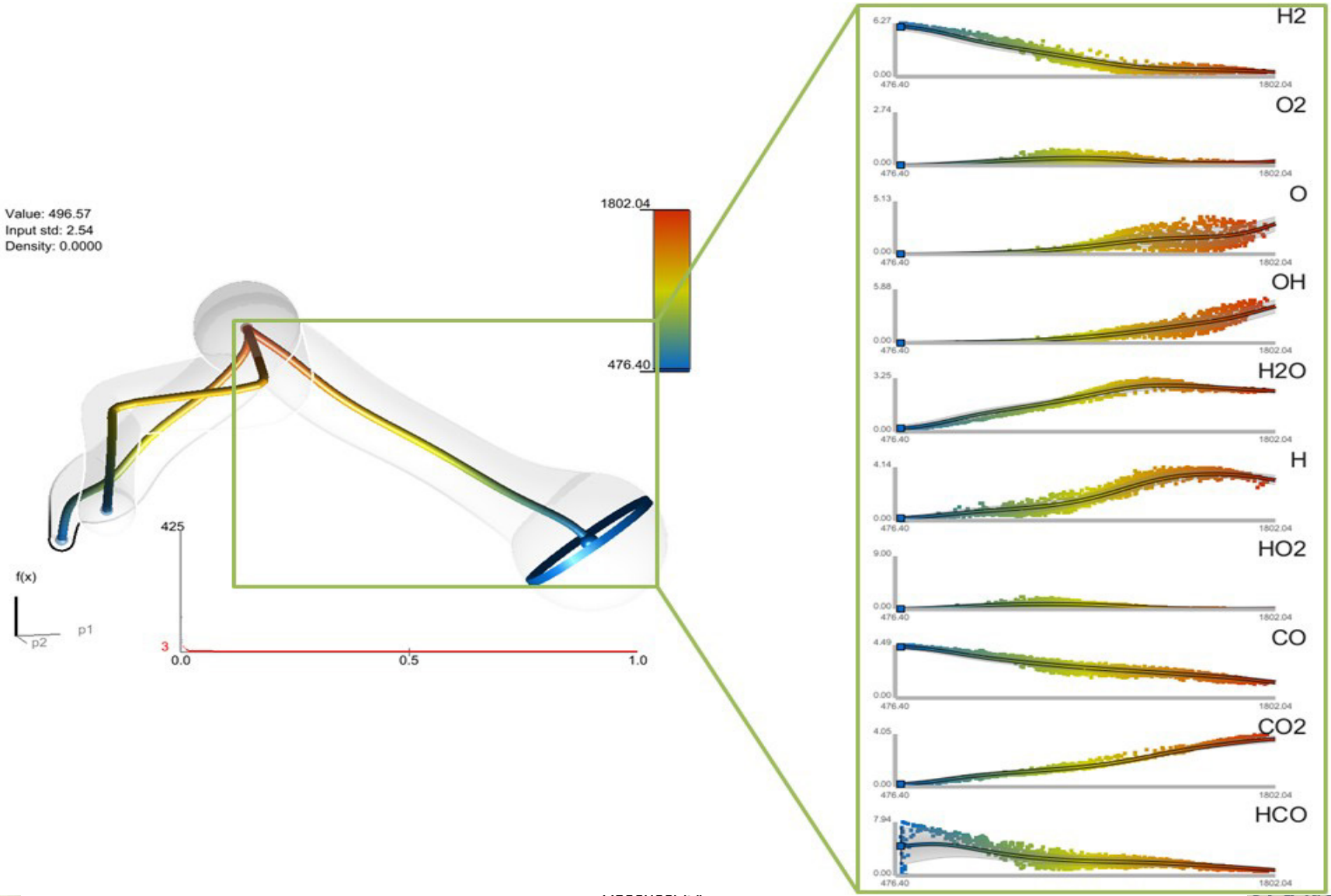
Mean of crystal data with respect to input coordinate

Standard deviation of crystal data with respect to input coordinate

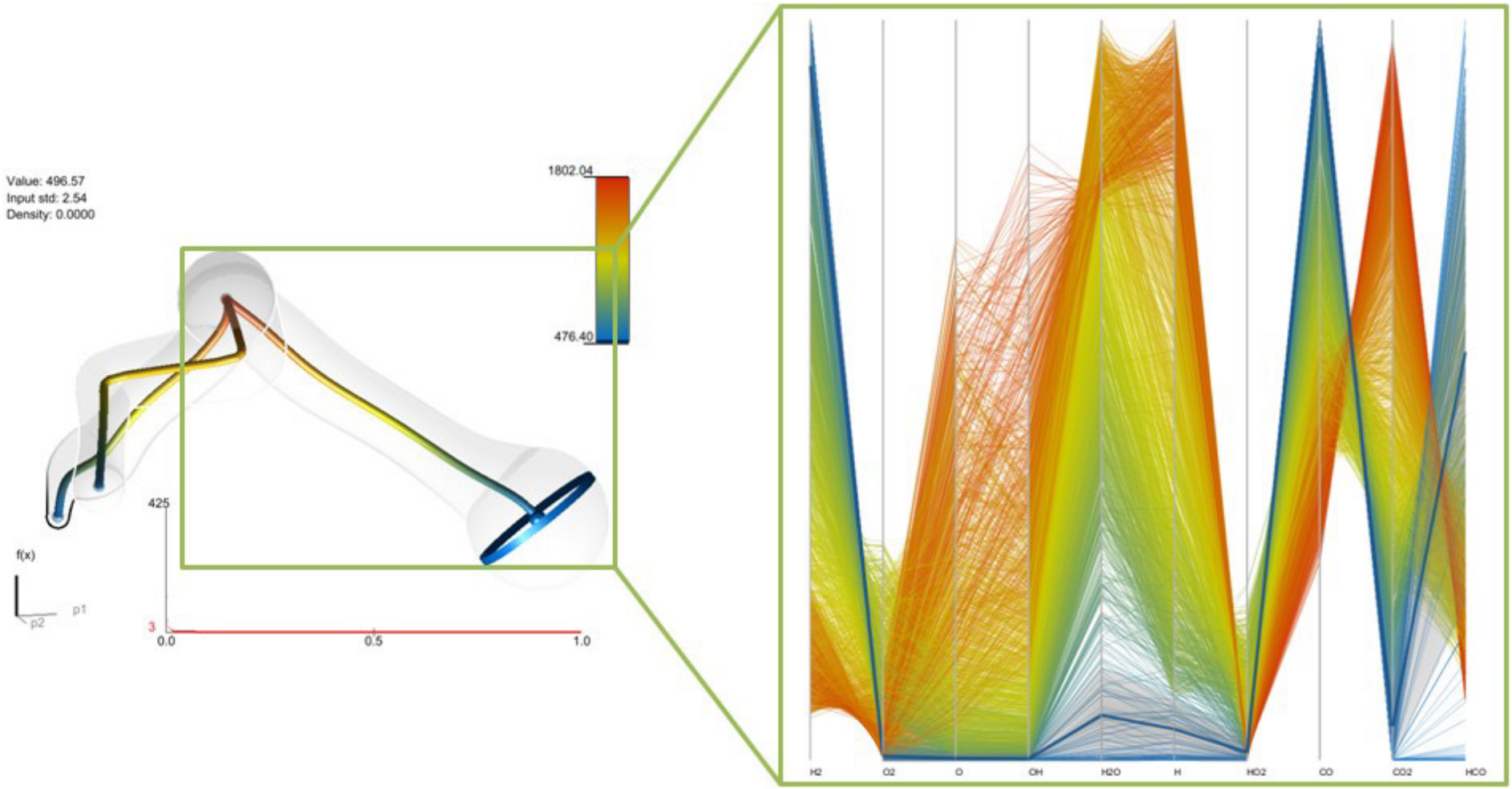




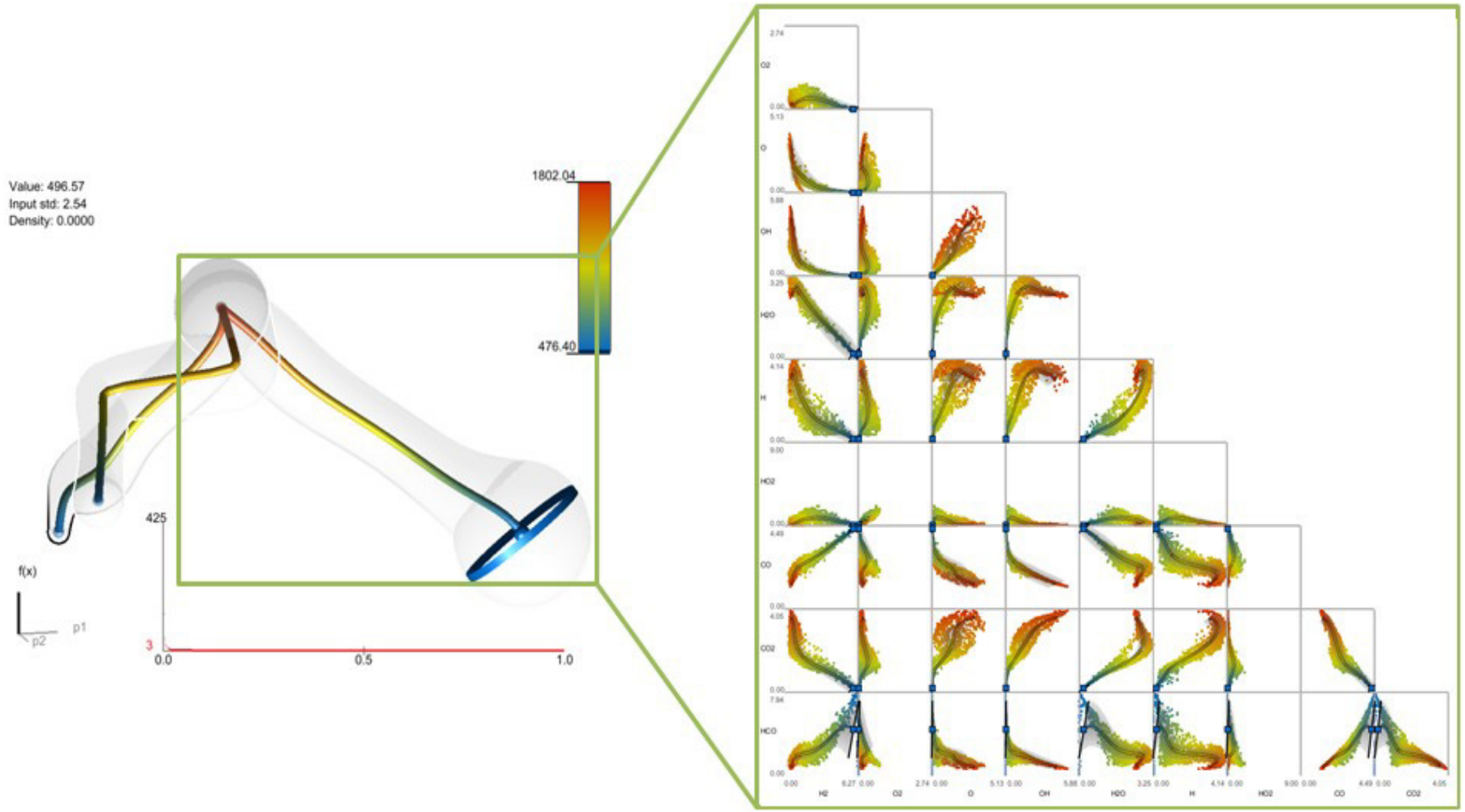
Inverse Coordinate Plots



Parallel Coordinates Plots



Pairwise Scatter Plots



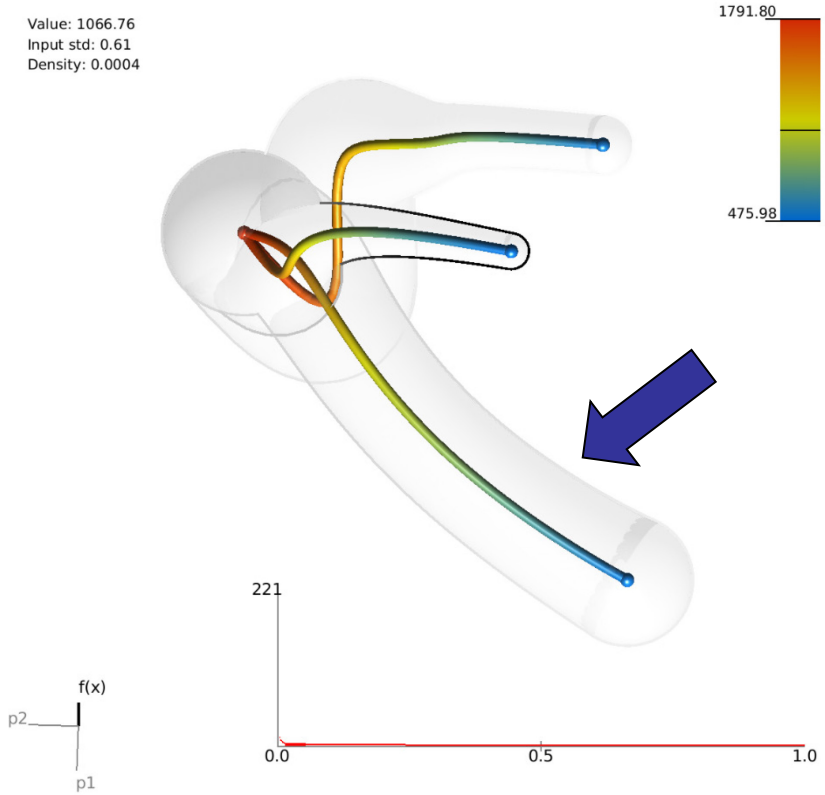
Analysis of Combustion Simulations

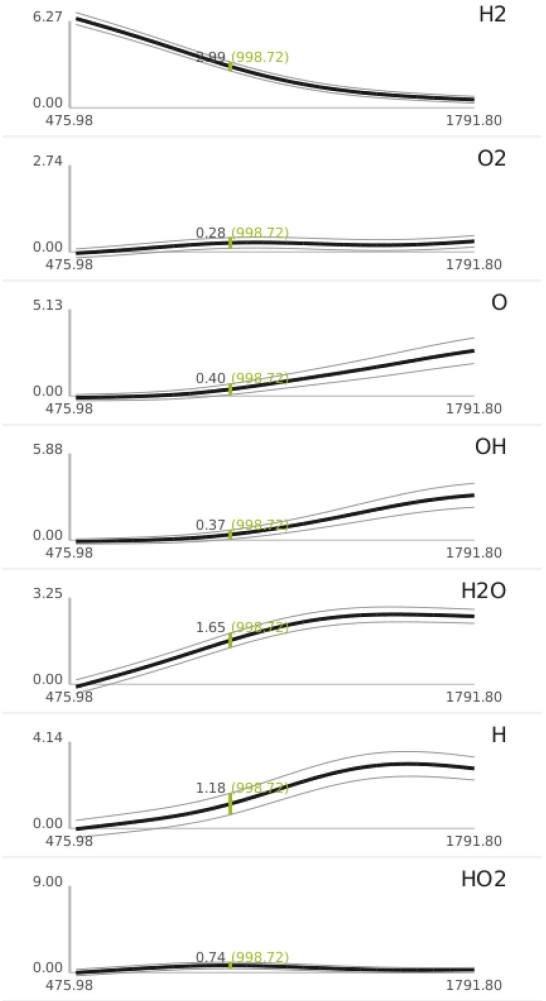
Combustion Simulation of Jet CO/H2-Air Flames

Input: Composition of 10 chemical species

Output: Temperature

The Framework Allows Detailed Visualization and Analysis of High Dimensional Functions

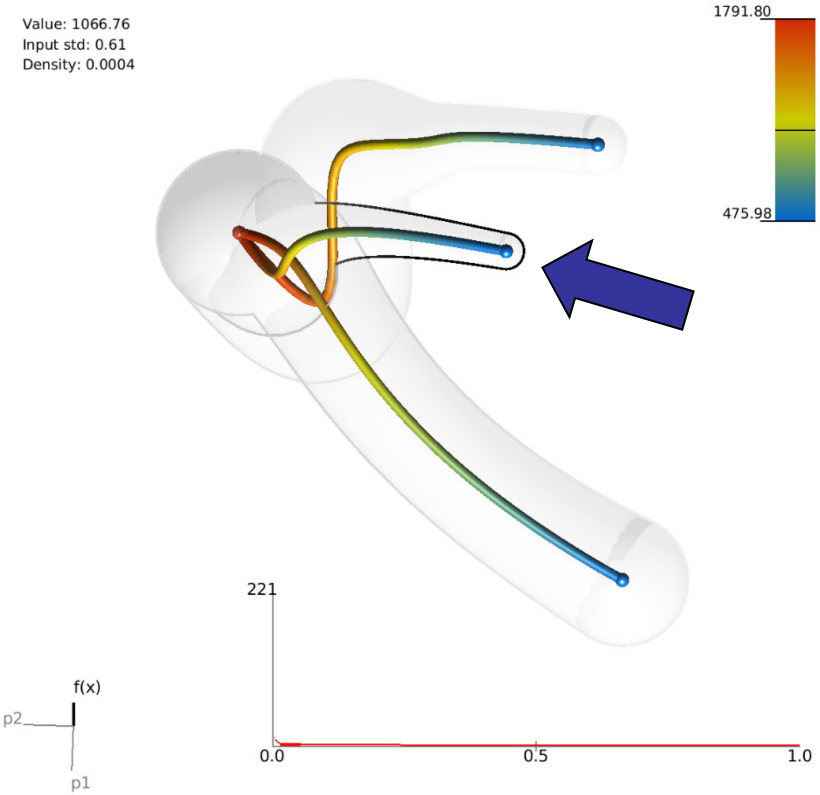


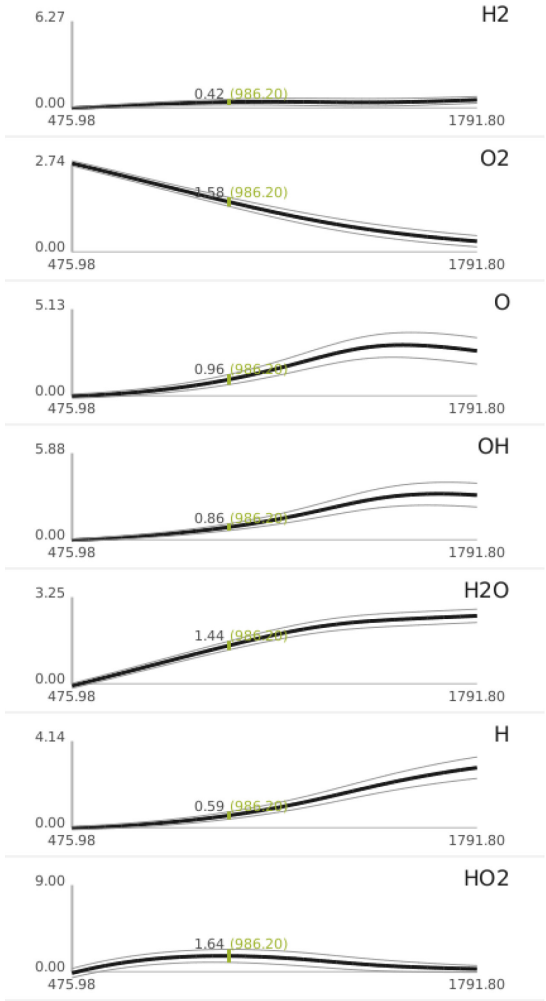


Pure fuel

10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

The Framework Allows Detailed Visualization and Analysis of High Dimensional Functions

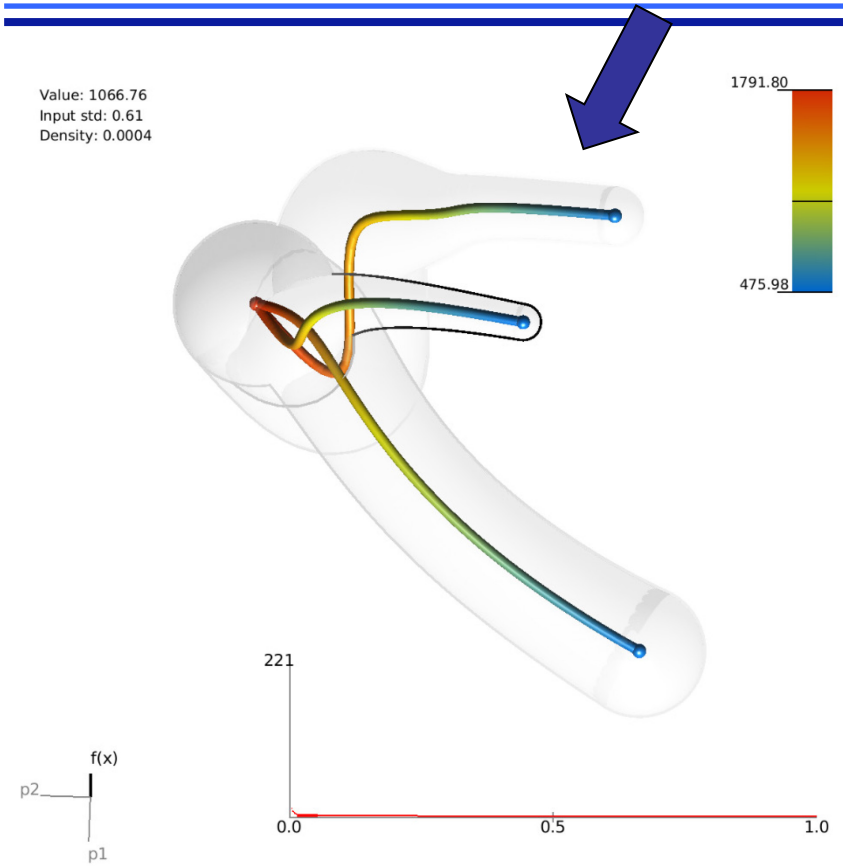


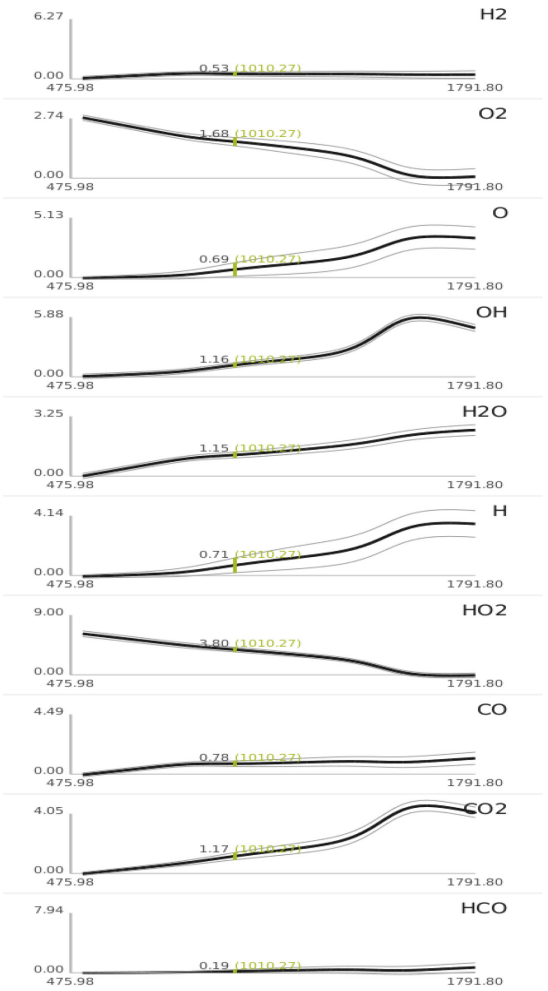


Pure oxidizer

10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

The Framework Allows Detailed Visualization and Analysis of High Dimensional Functions





Local extinction

10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

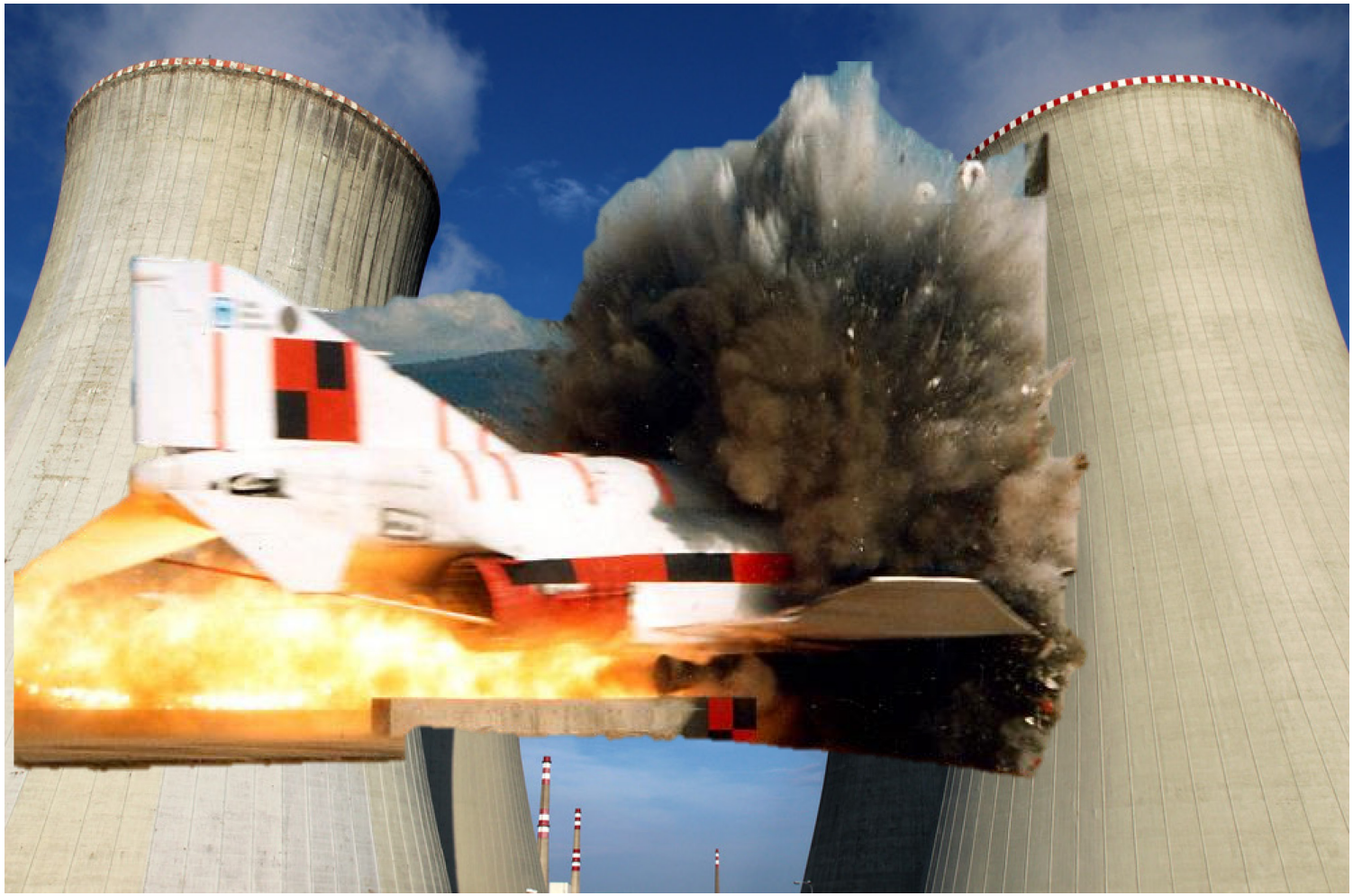
Nuclear Reactor Safety



Zhimin Li

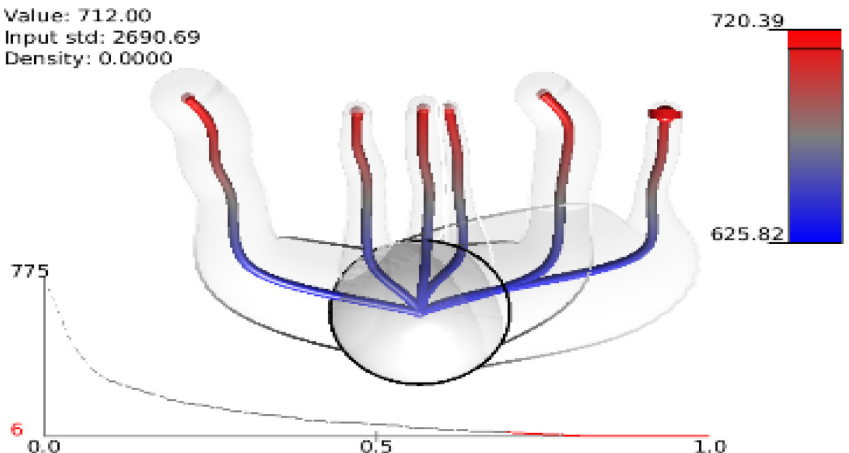


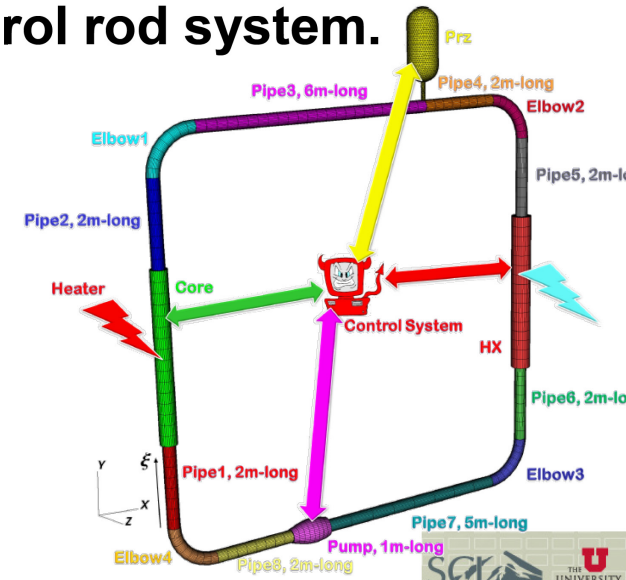
Nuclear Reactor Safety



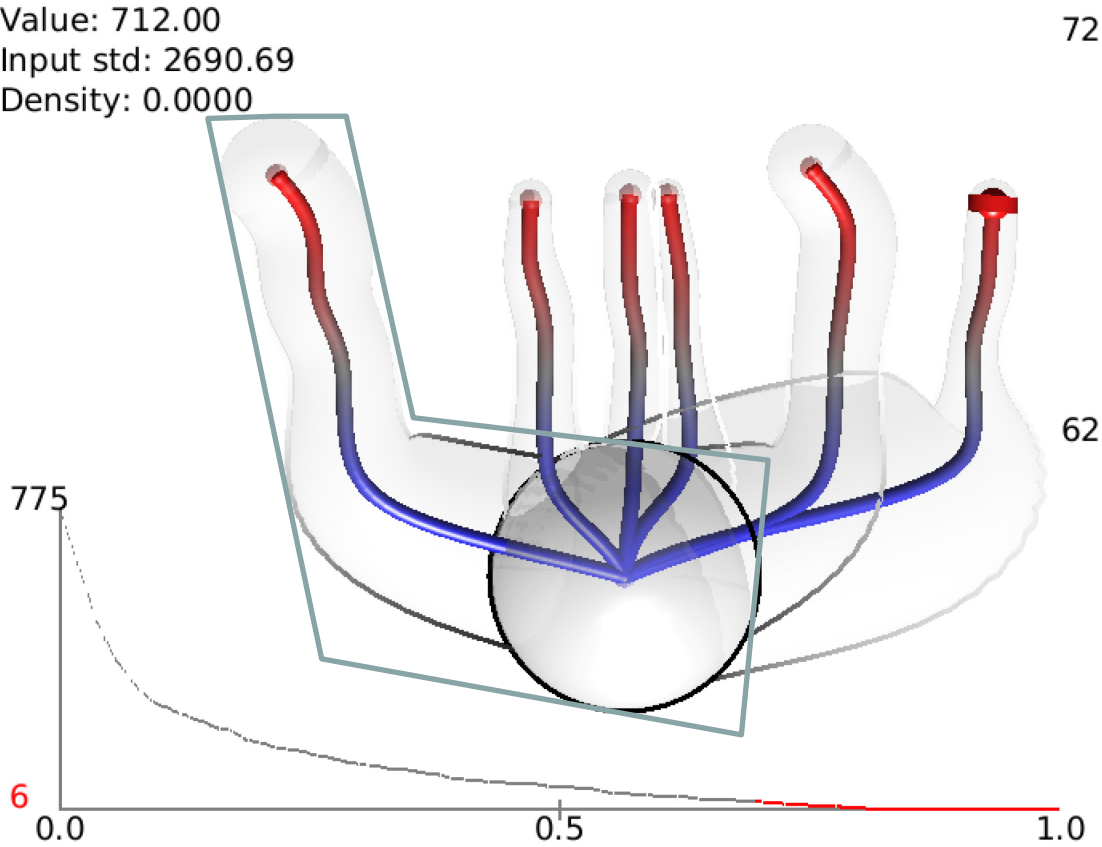
The SCRAM Event: Strategies to Avoid Catastrophic Consequences (Meltdown) for the Reactor

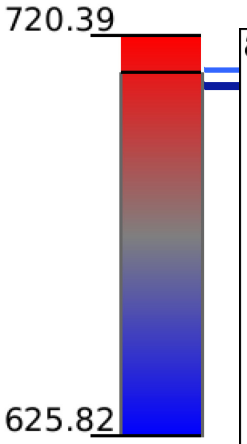
- 10,000 individual simulations of a SCRAM event.
- Input = 6 dimensions:
  - PumpTripPre - pressure in heat exchange pump causing SCRAM
  - PumpStopTime - the relaxation time (sec) of pump's phase-out
  - PumpPow - end power of the pump
  - SCRAMtemp - the maximum temperature (K) in the system
  - CRinject - control rod position at the end of SCRAM
  - CRtime - the relaxation time of the control rod system.
- Output = Peak Coolant Temperature (PCT)

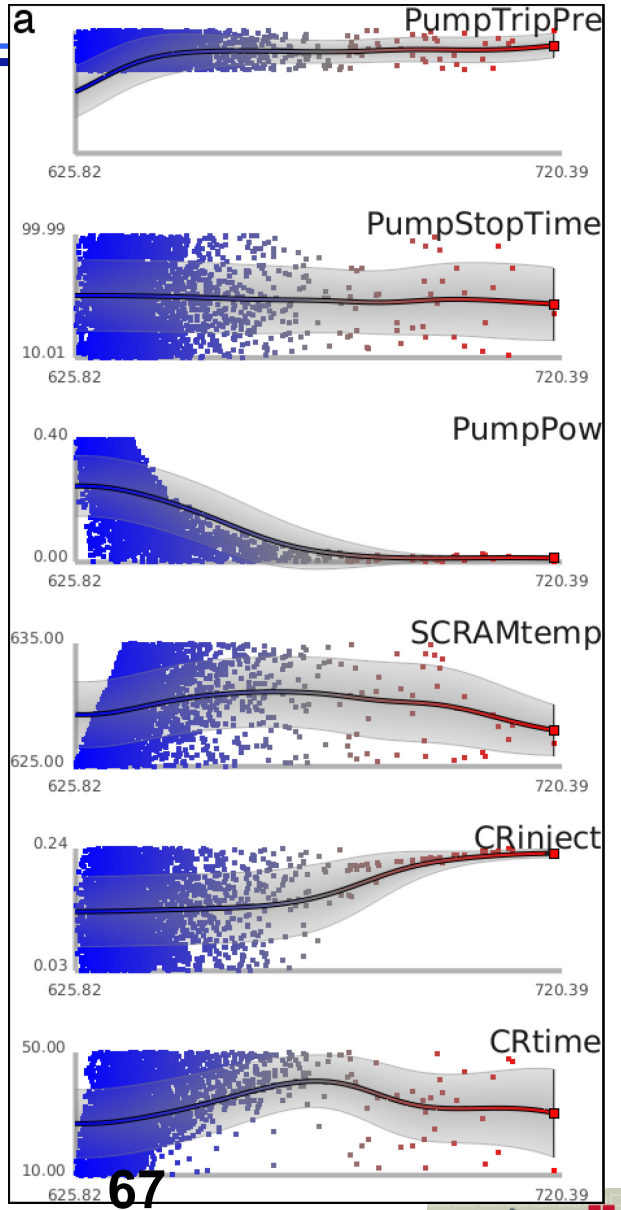




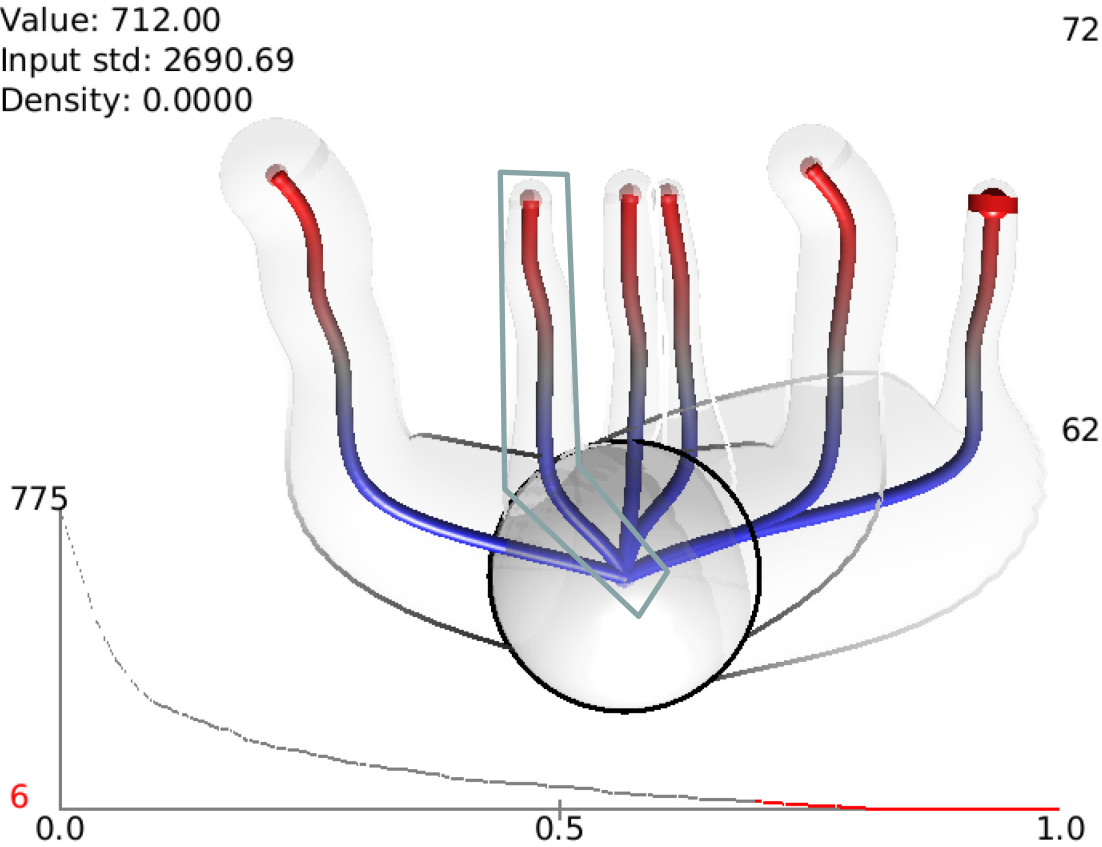
VR_2^+ Simulator

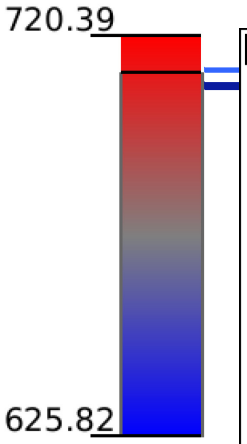


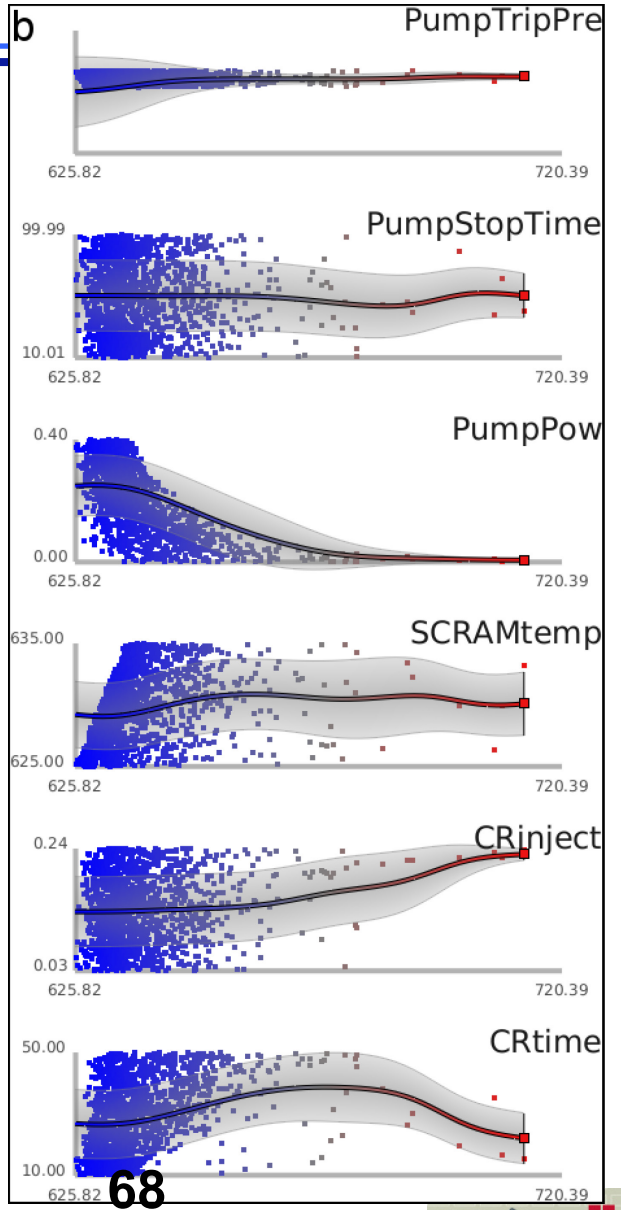




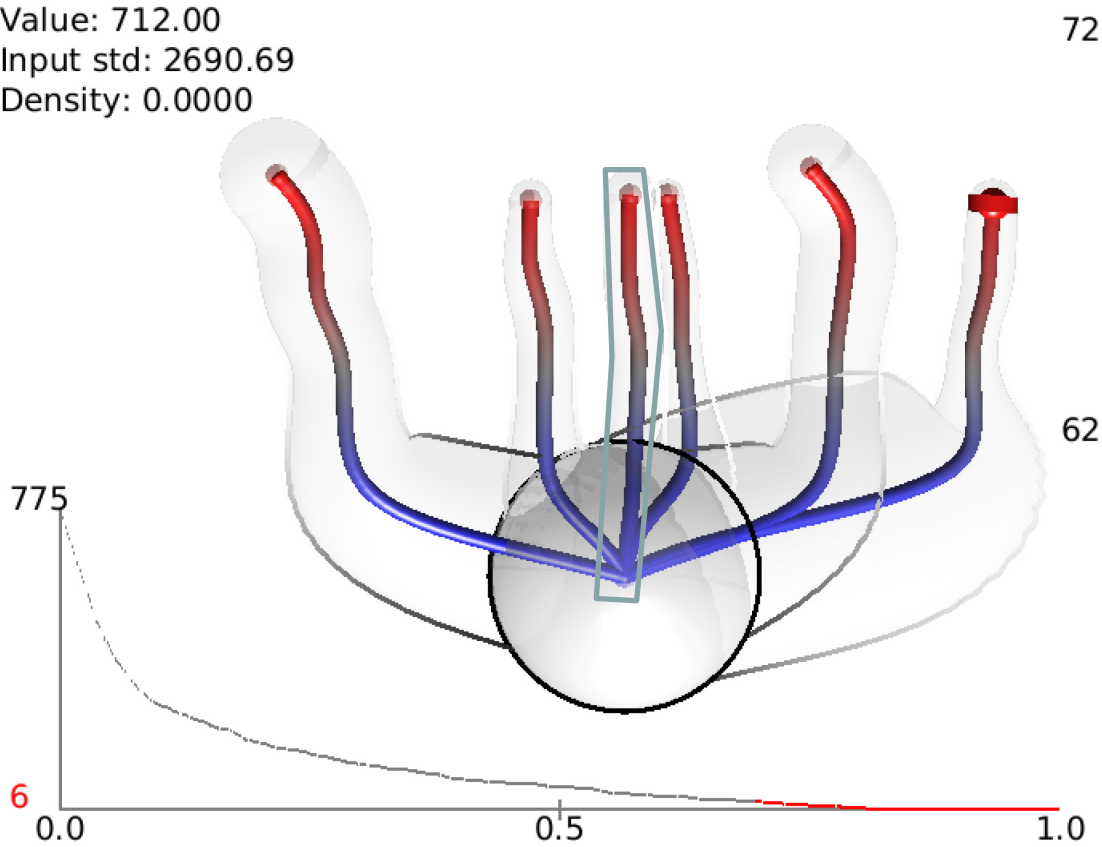
VR_2^+ Simulator

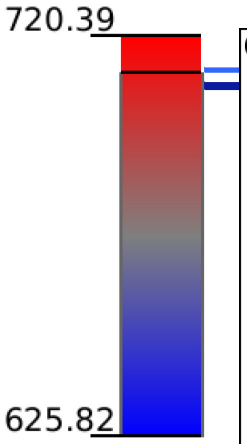


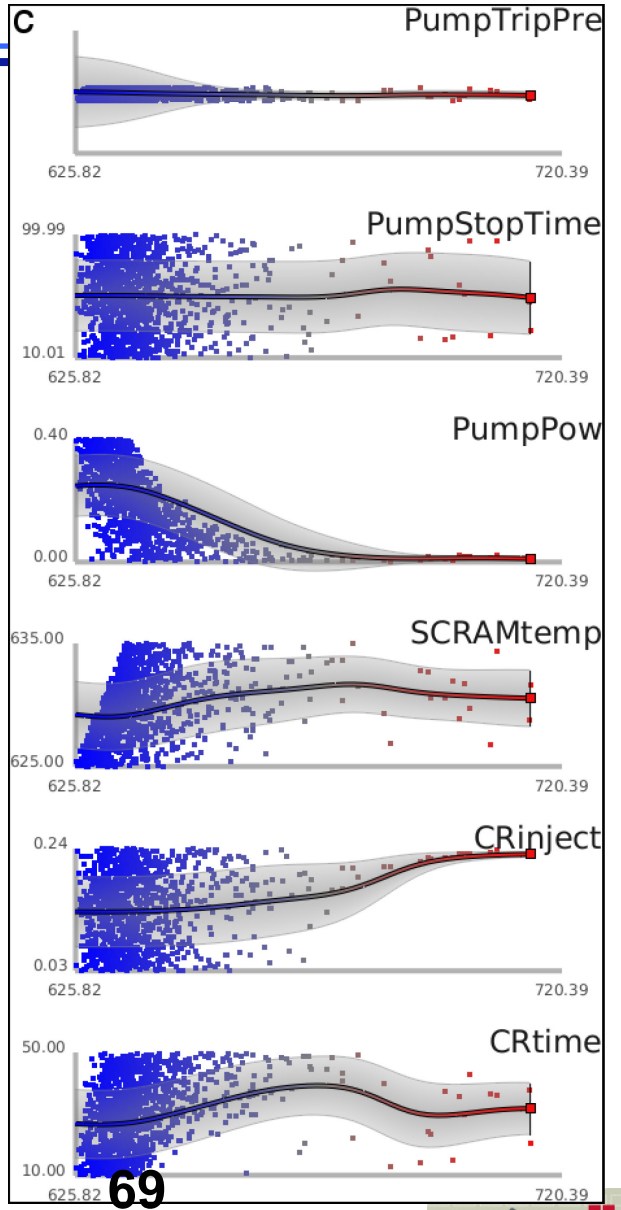




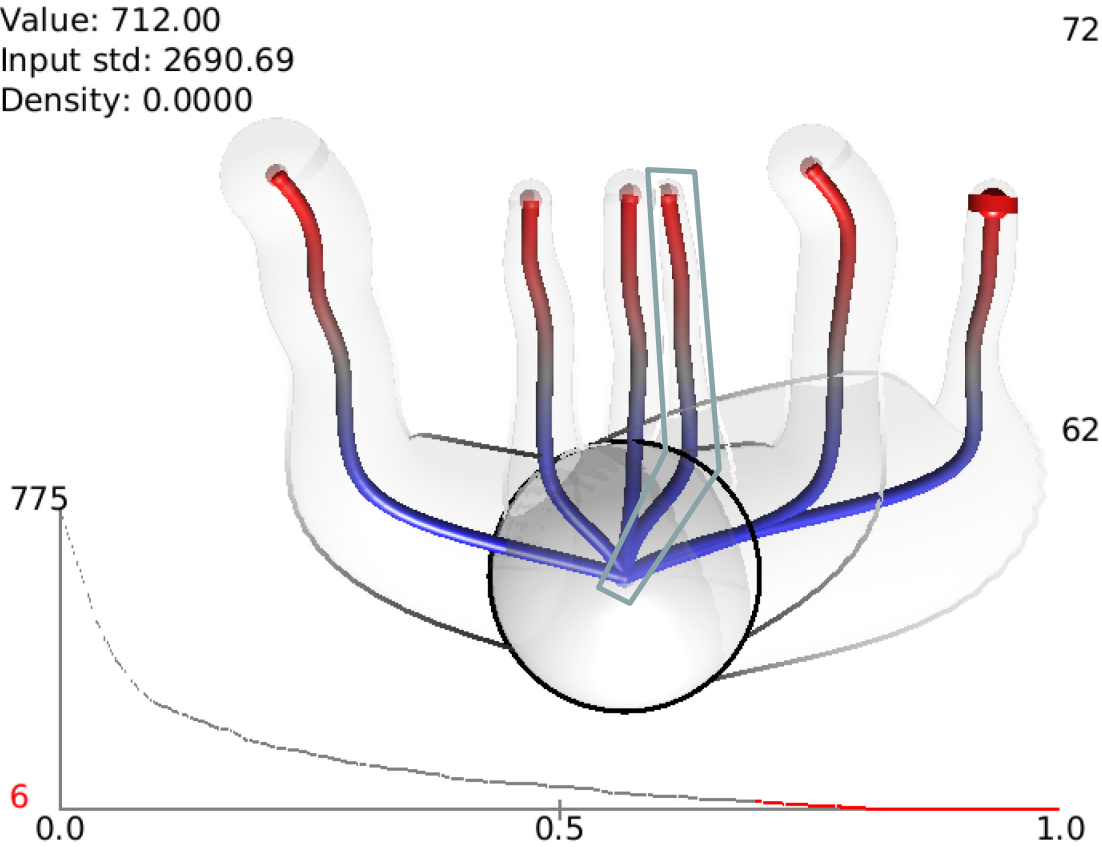
VR_2^+ Simulator

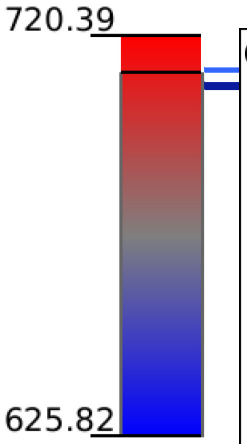


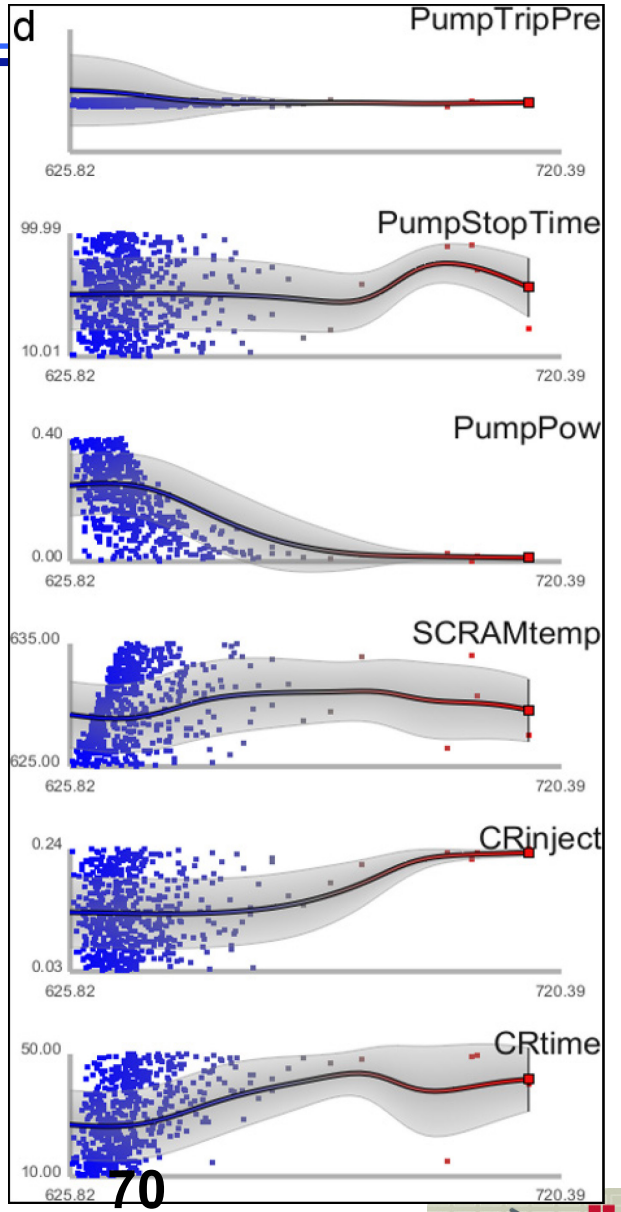




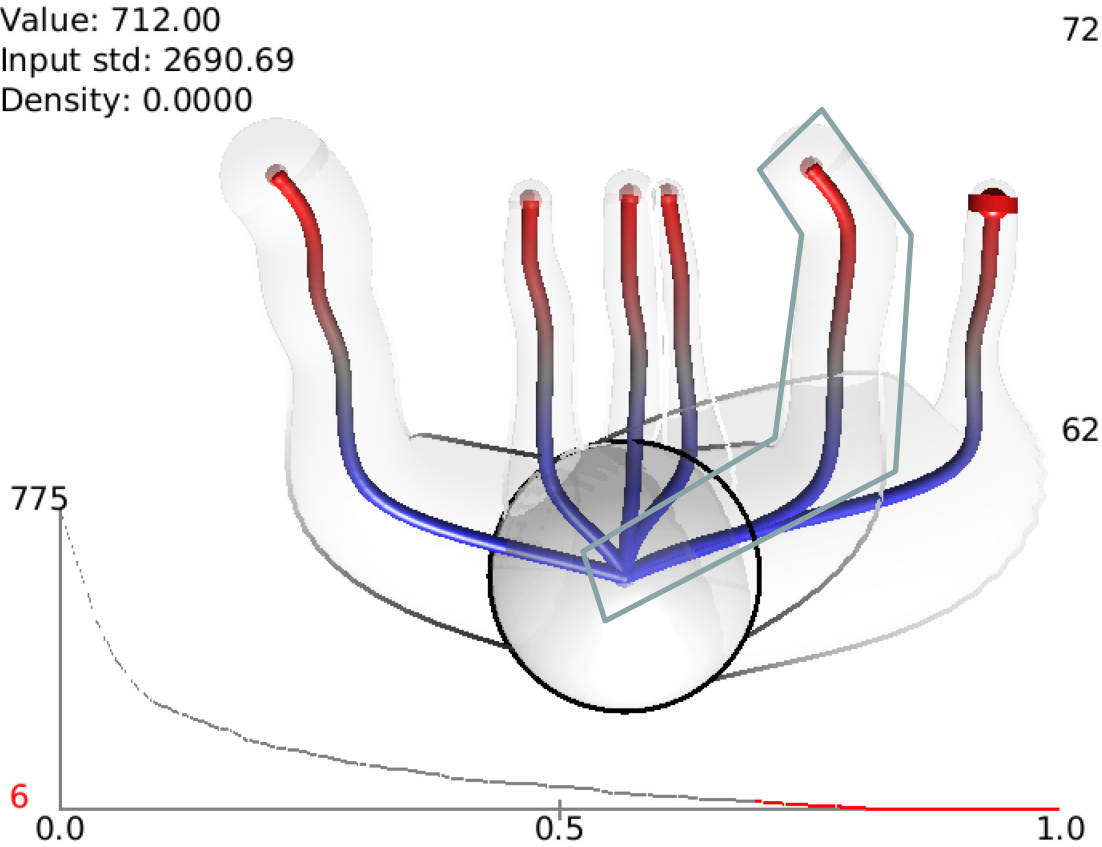
VR_2^+ Simulator

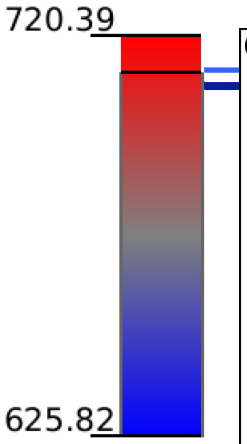


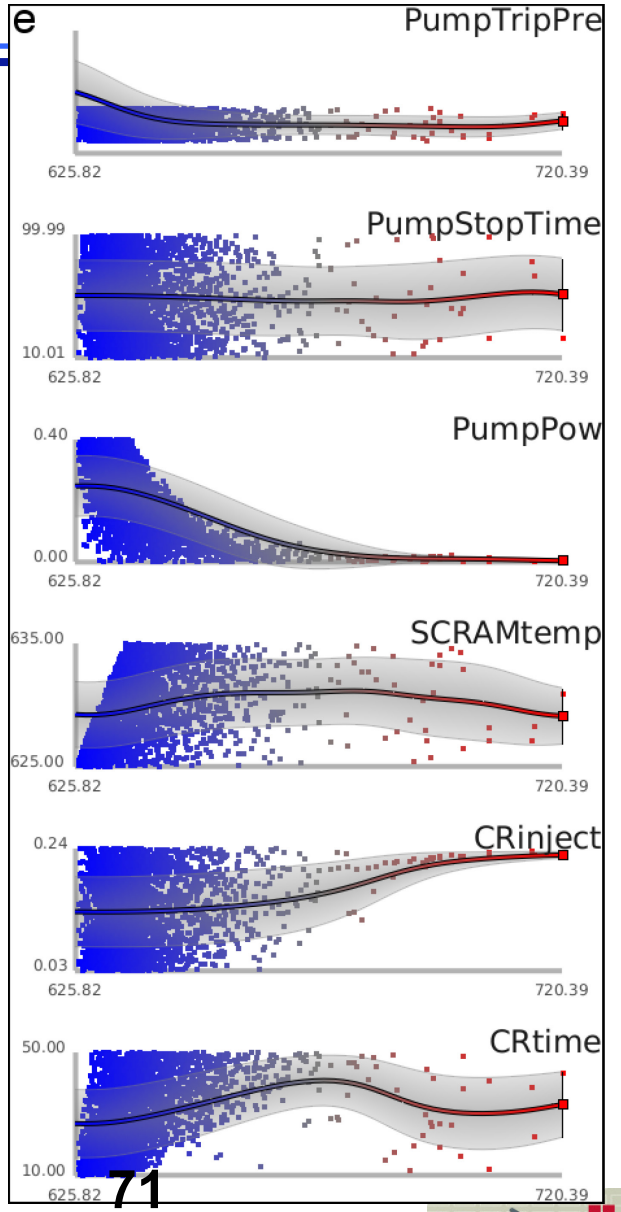




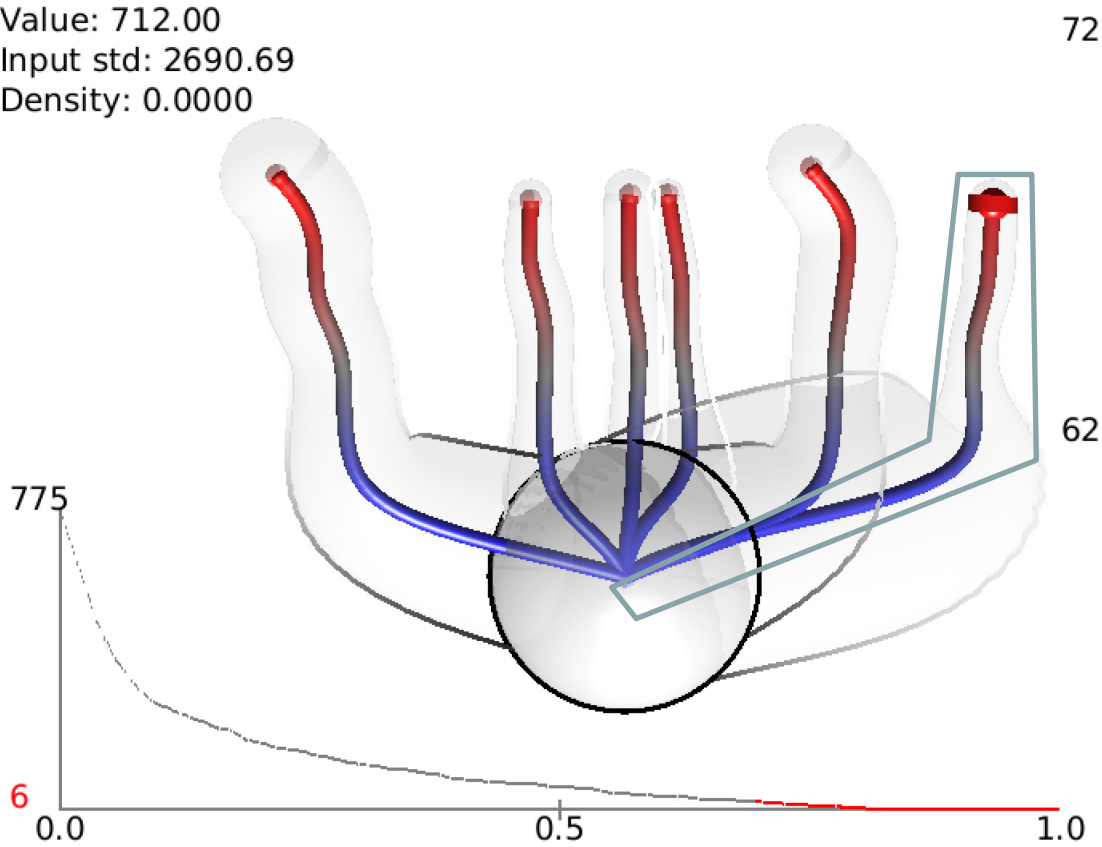
VR_2^+ Simulator

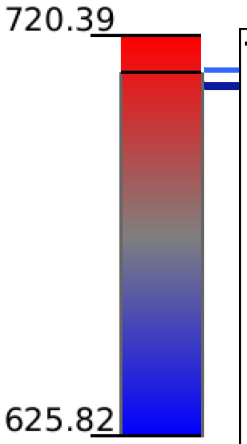


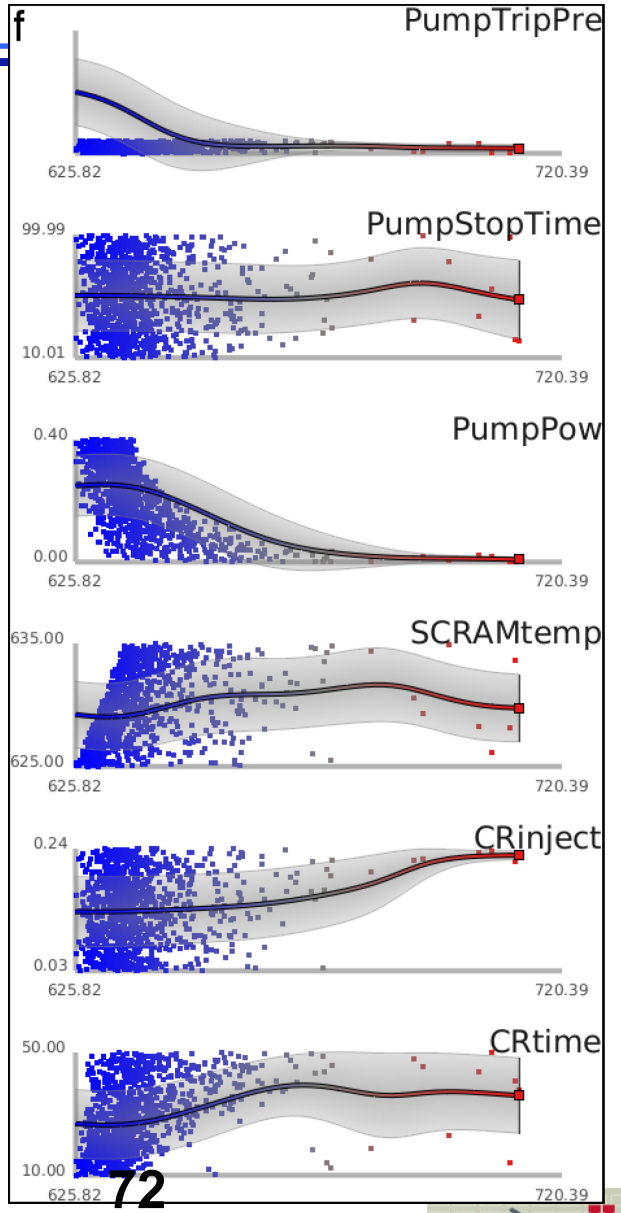




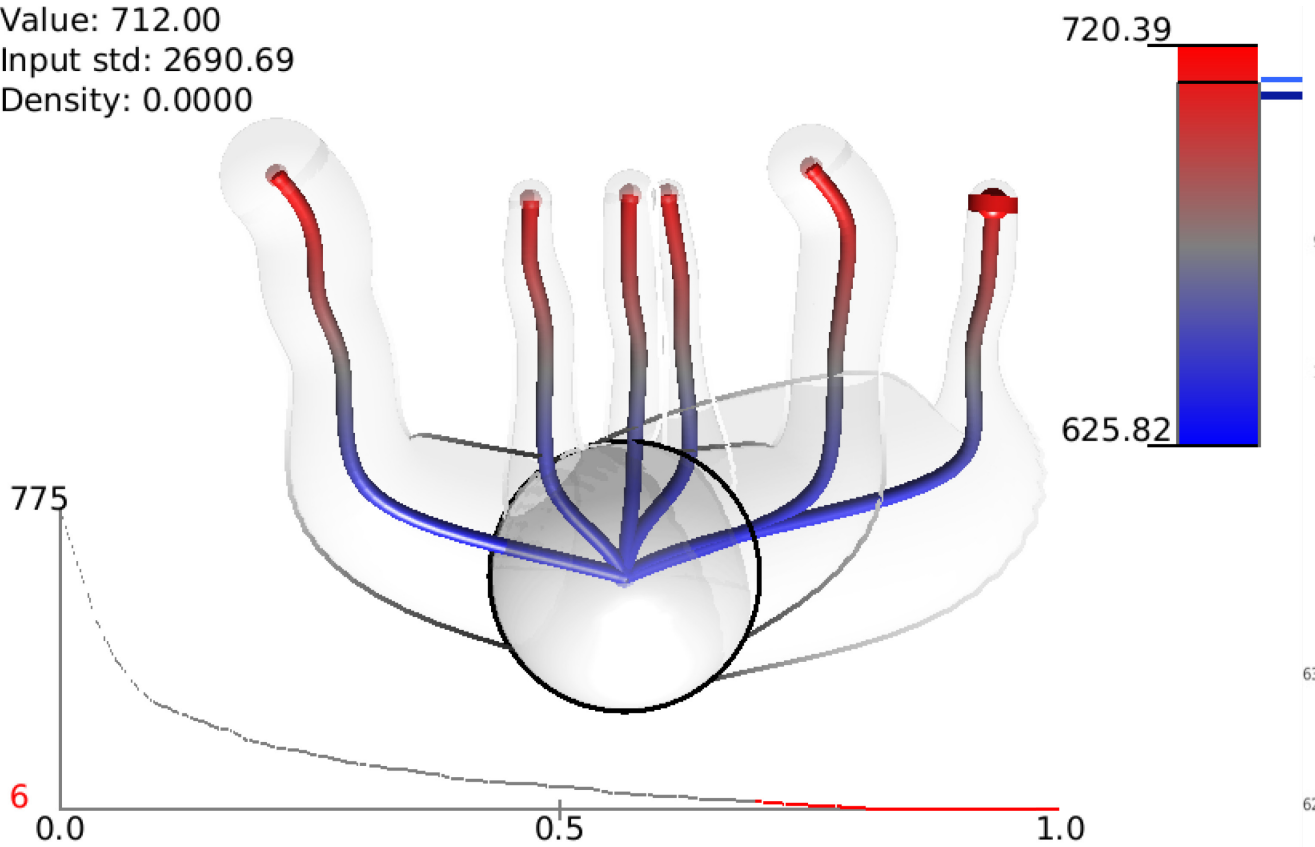
VR_2^+ Simulator

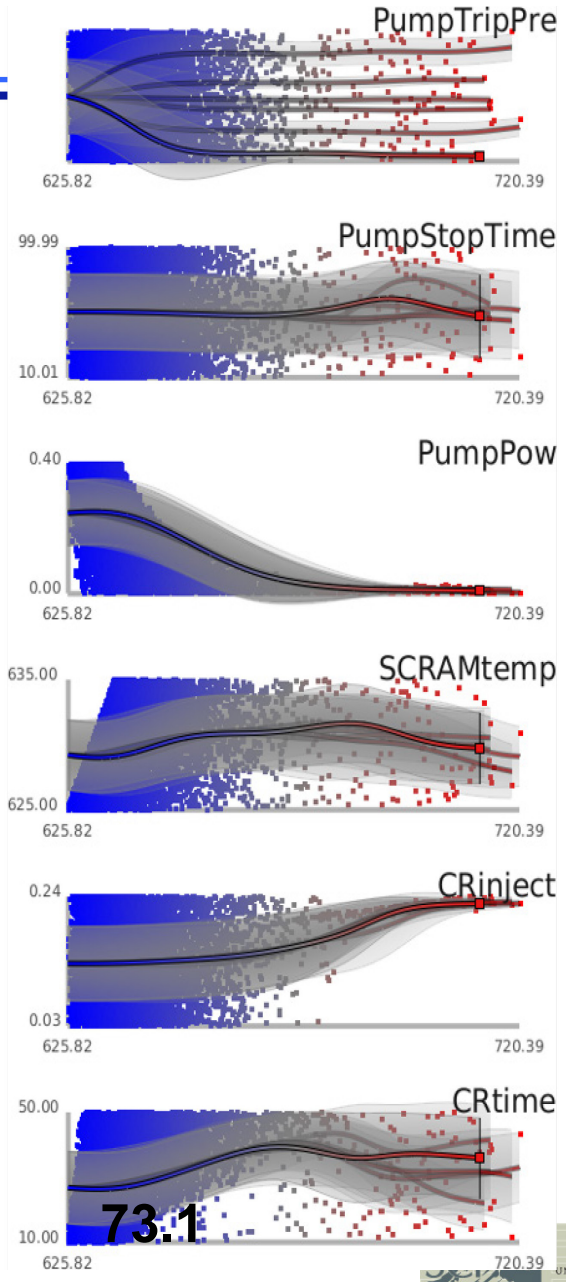






VR_2^+ Simulator





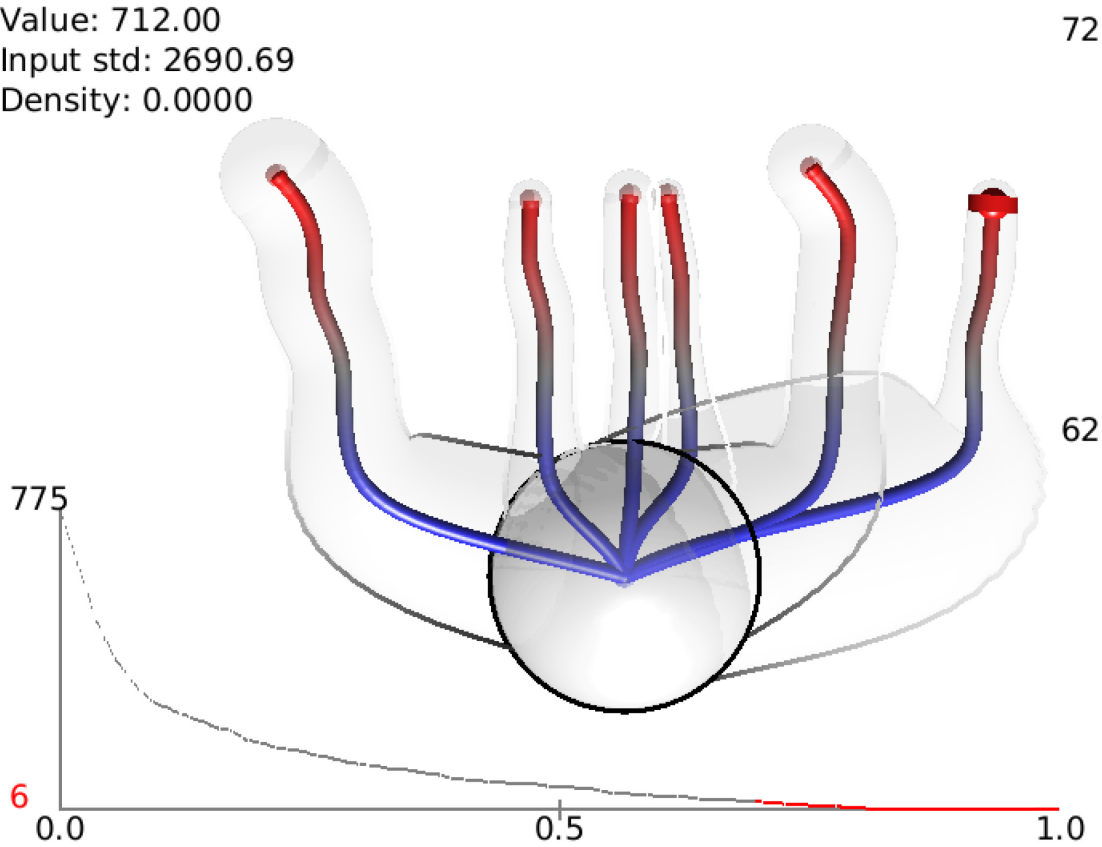
Pump trip pressure is the clearest distinguishing factor differentiating the high PCT (Peak Coolant Temperature) modes

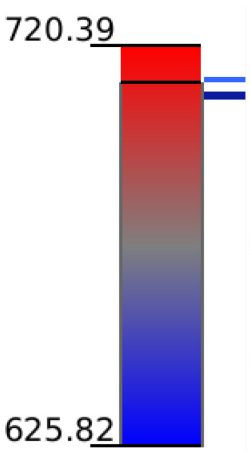
(*Necessary, but not sufficient conditions for high PCT)

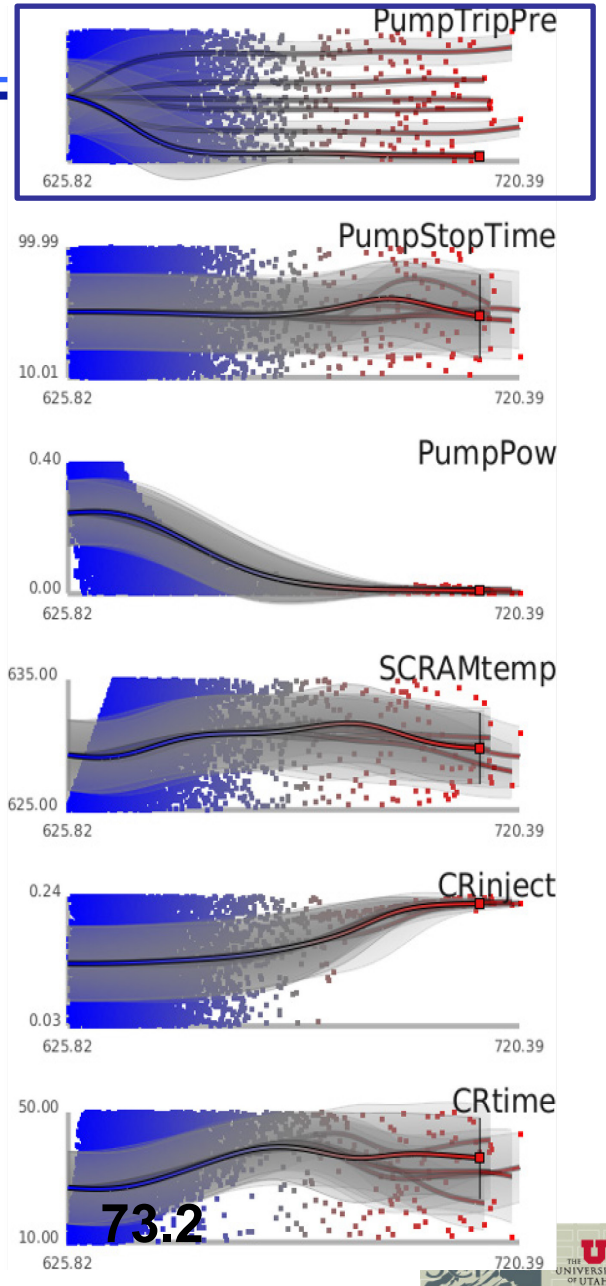
Invariants:

- Pump power
- Control rod injection position

VR_2^+ Simulator







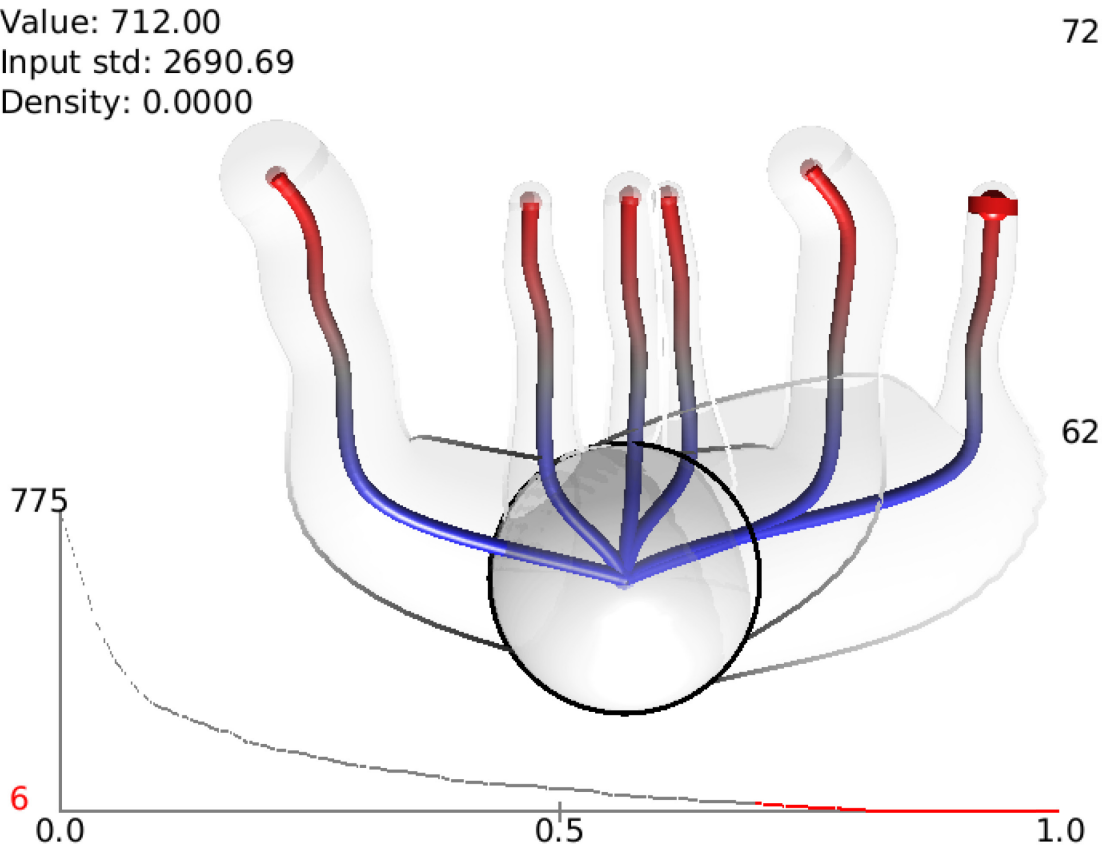
Pump trip pressure is the clearest distinguishing factor differentiating the high PCT (Peak Coolant Temperature) modes

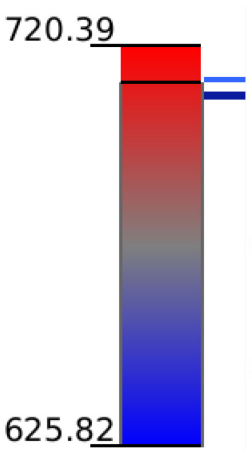
(*Necessary, but not sufficient conditions for high PCT)

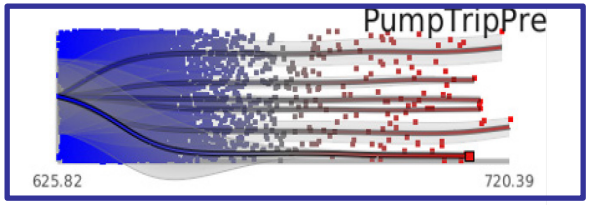
Invariants:

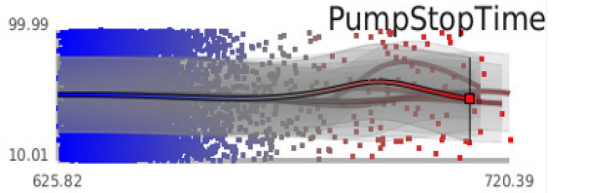
- Pump power
- Control rod injection position

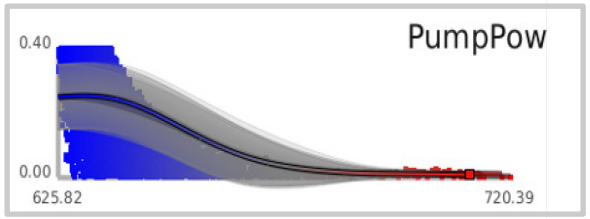
VR_2^+ Simulator

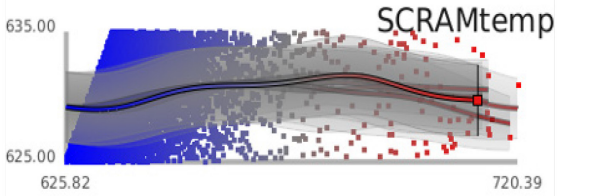


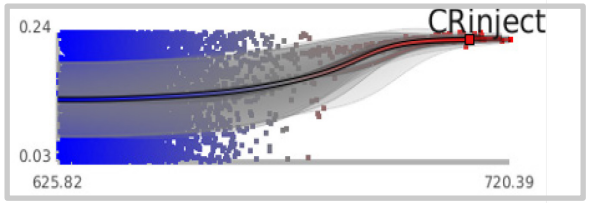


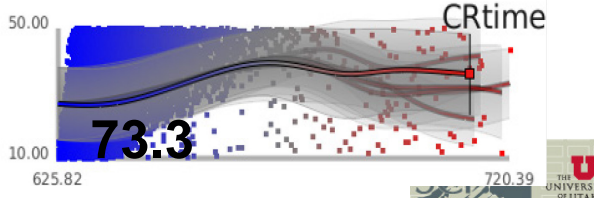












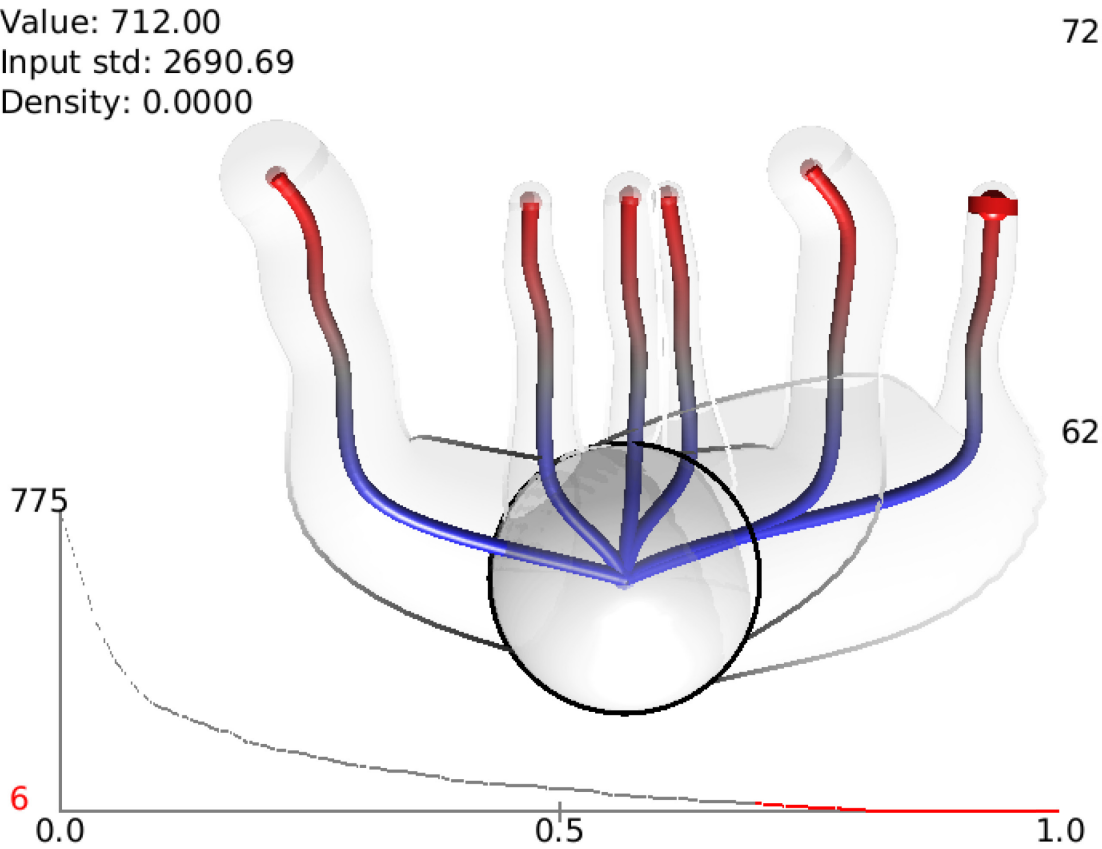
Pump trip pressure is the clearest distinguishing factor differentiating the high PCT (Peak Coolant Temperature) modes

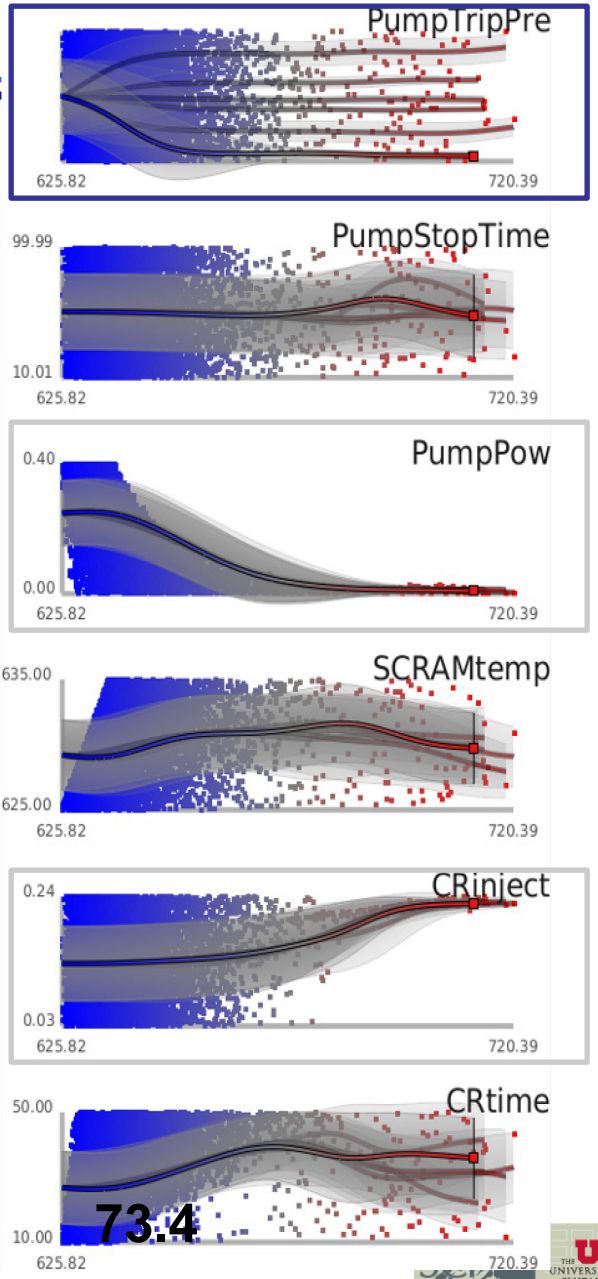
(*Necessary, but not sufficient conditions for high PCT)

Invariants:

- Pump power
- Control rod injection position

VR_2^+ Simulator





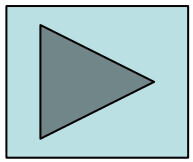
Pump trip pressure is the clearest distinguishing factor differentiating the high PCT (Peak Coolant Temperature) modes

(*Necessary, but not sufficient conditions for high PCT)

Invariants:

- Pump power
- Control rod injection position

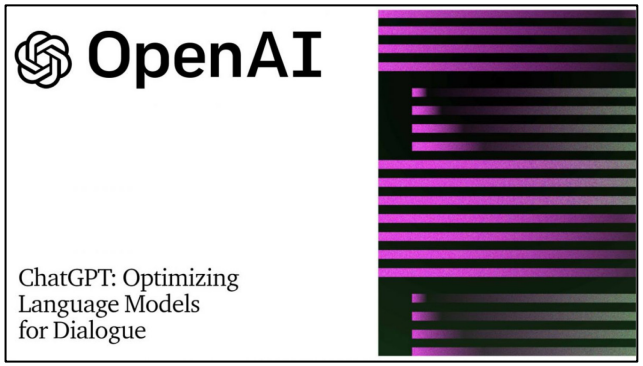
EXPLAINABILITY INCLUDING CATEGORICAL DATA FOR AI AND COMPLEX SYSTEMS

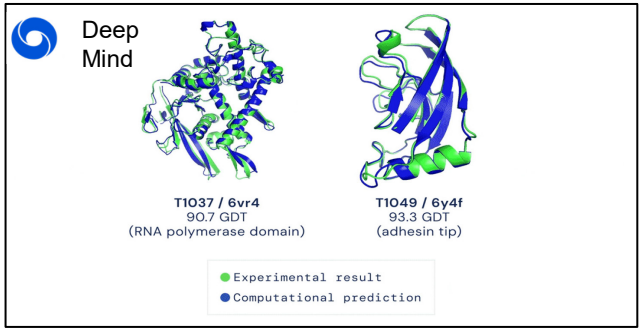


AI Applications and Explaining their Effectiveness

- ChatGPT
- AlphaGO
- Auto-pilot
- ImageNet
- …



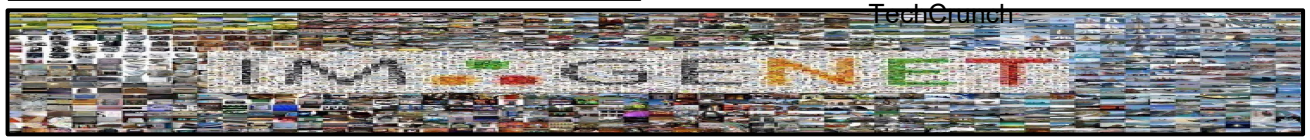








From TechCrunch



Explaining inference in natural language processing

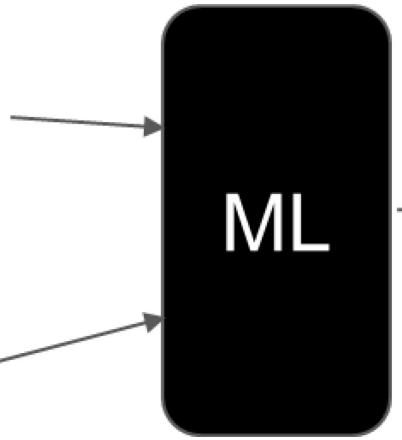
An illustration of natural language inference.

P / H	sentences	entail	contradict	neutral
premise	A boy ate an apple.	-	-	-
hypothesis	A kid ate fruit.	✓		
hypothesis	A boy ate a banana.		✓	
hypothesis	Tom ate an apple.			✓
hypothesis	A boy ate a Fuji apple.			✓

ML Task: determine if a hypothesis is an entailment, contradiction, or neutral w.r.t. a premise statement

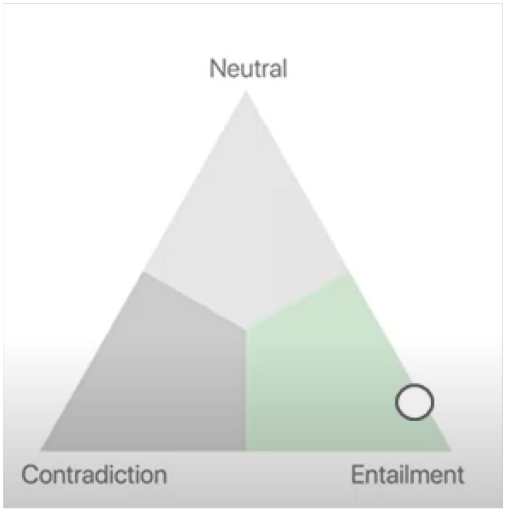
Premise(P):
A boy ate an apple

Hypothesis(H):
A kid ate a fruit



(Entail, Neutral, Contradict)

(0.9, 0.1, 0.0)



Questions a user may need to answer

How much can I trust a model/set of predictions?

Can I anticipate the outcome of a sample given a model?

Why did the model give me the result it did?

How stable is the prediction?

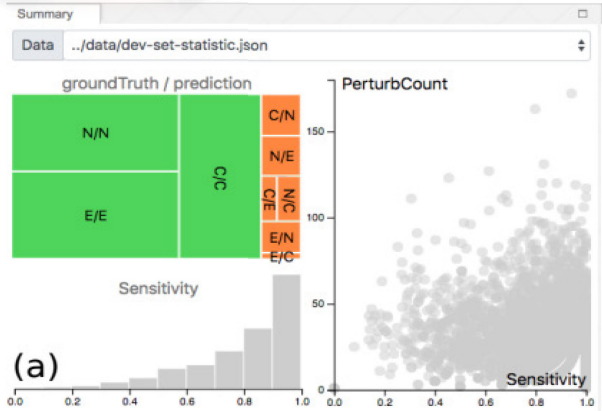
    w.r.t. perturbation of training, sample, model?

Could I have changed something to get a better prediction for the kinds of phenomena I am interested in?



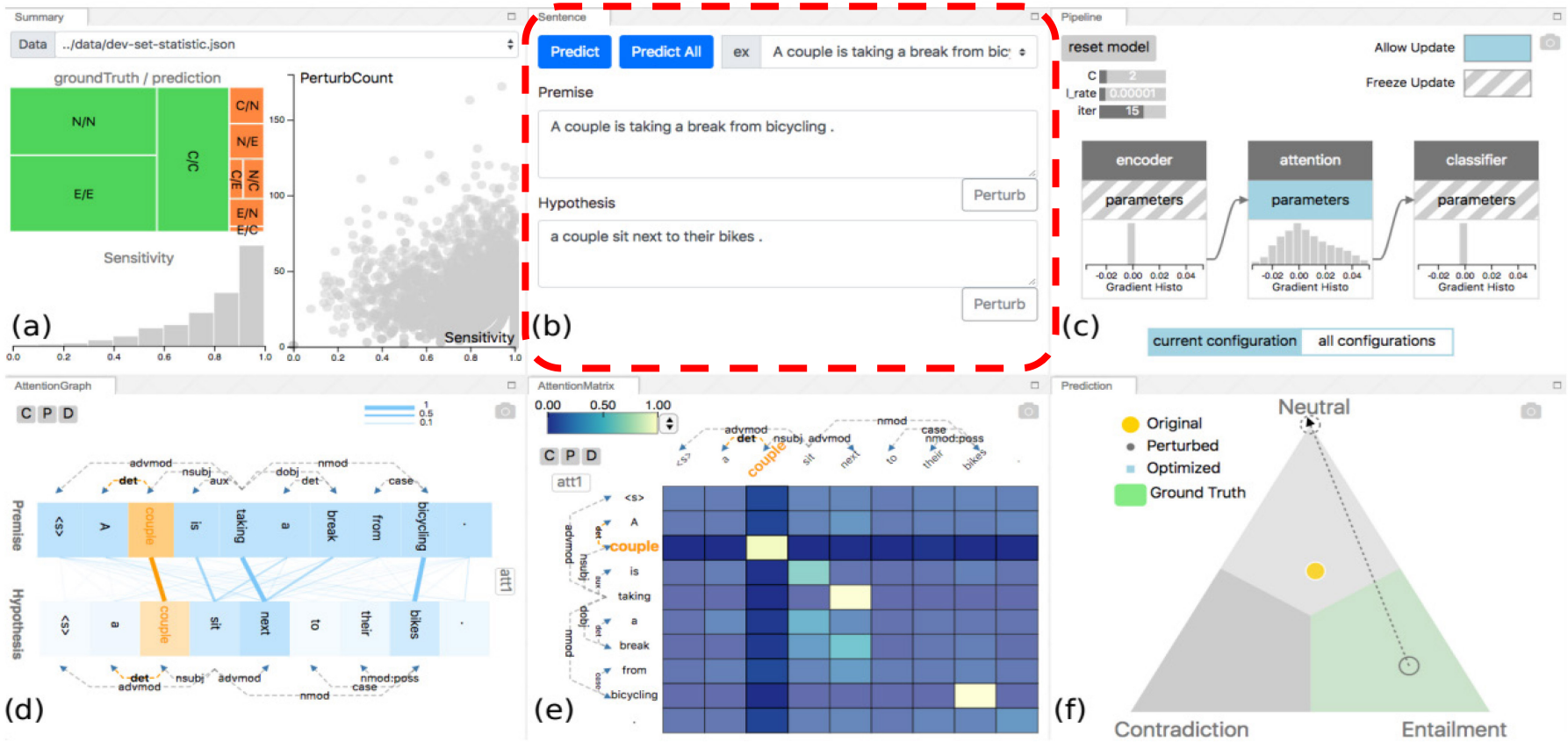
Zhimin Li

Interactive tool for understanding NLP Inference

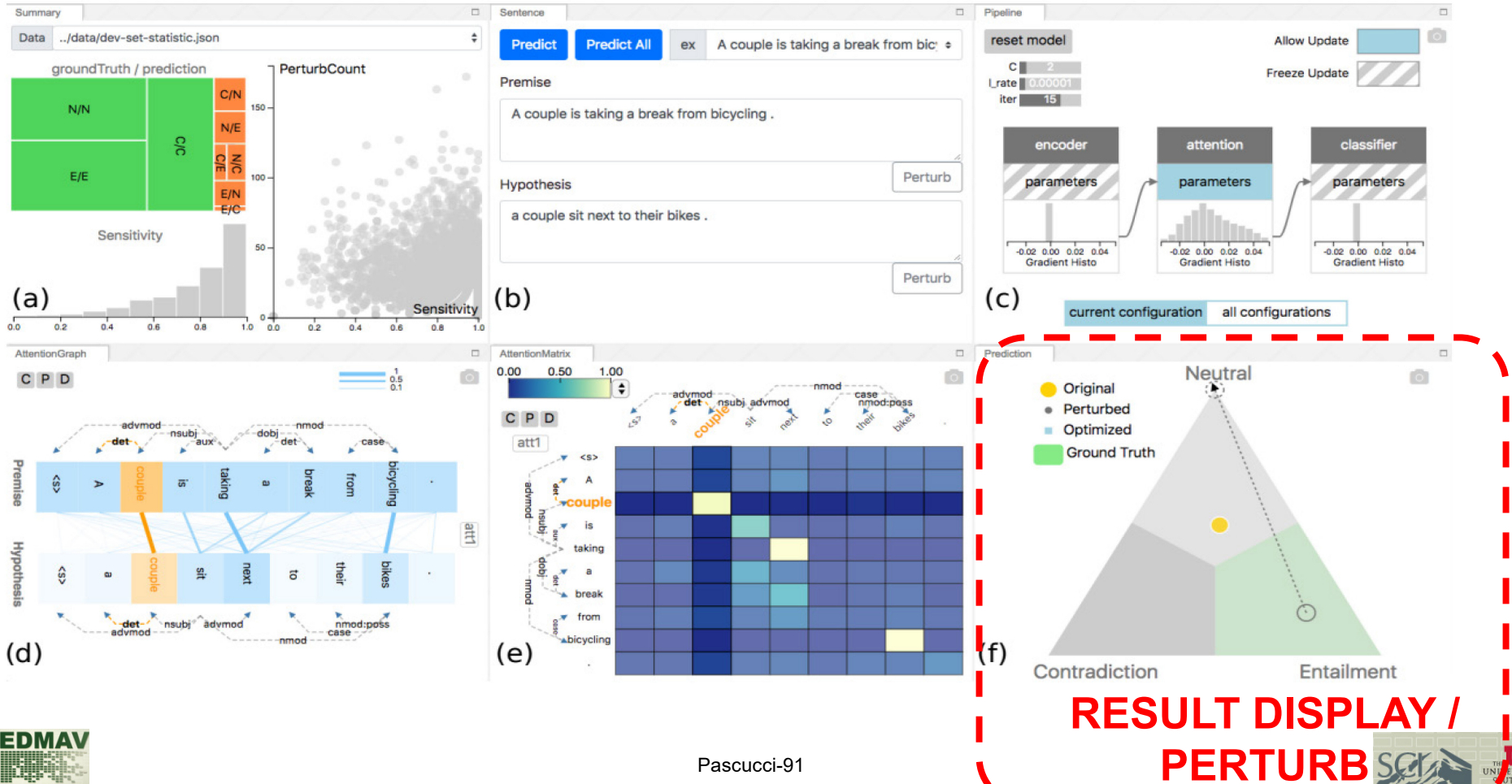


Interactive tool for understanding NLP Inference

INPUT DISPLAY

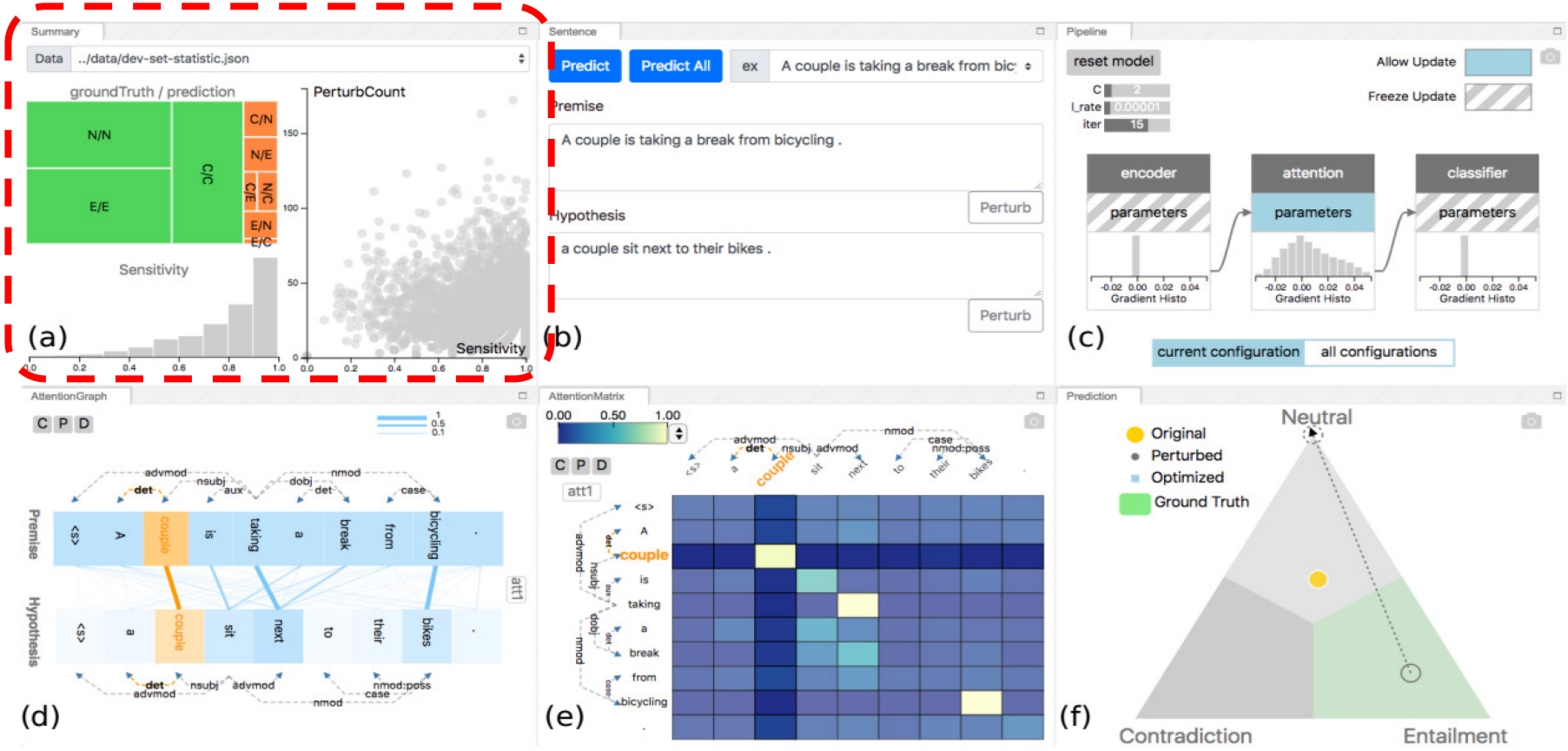


Interactive tool for understanding NLP Inference

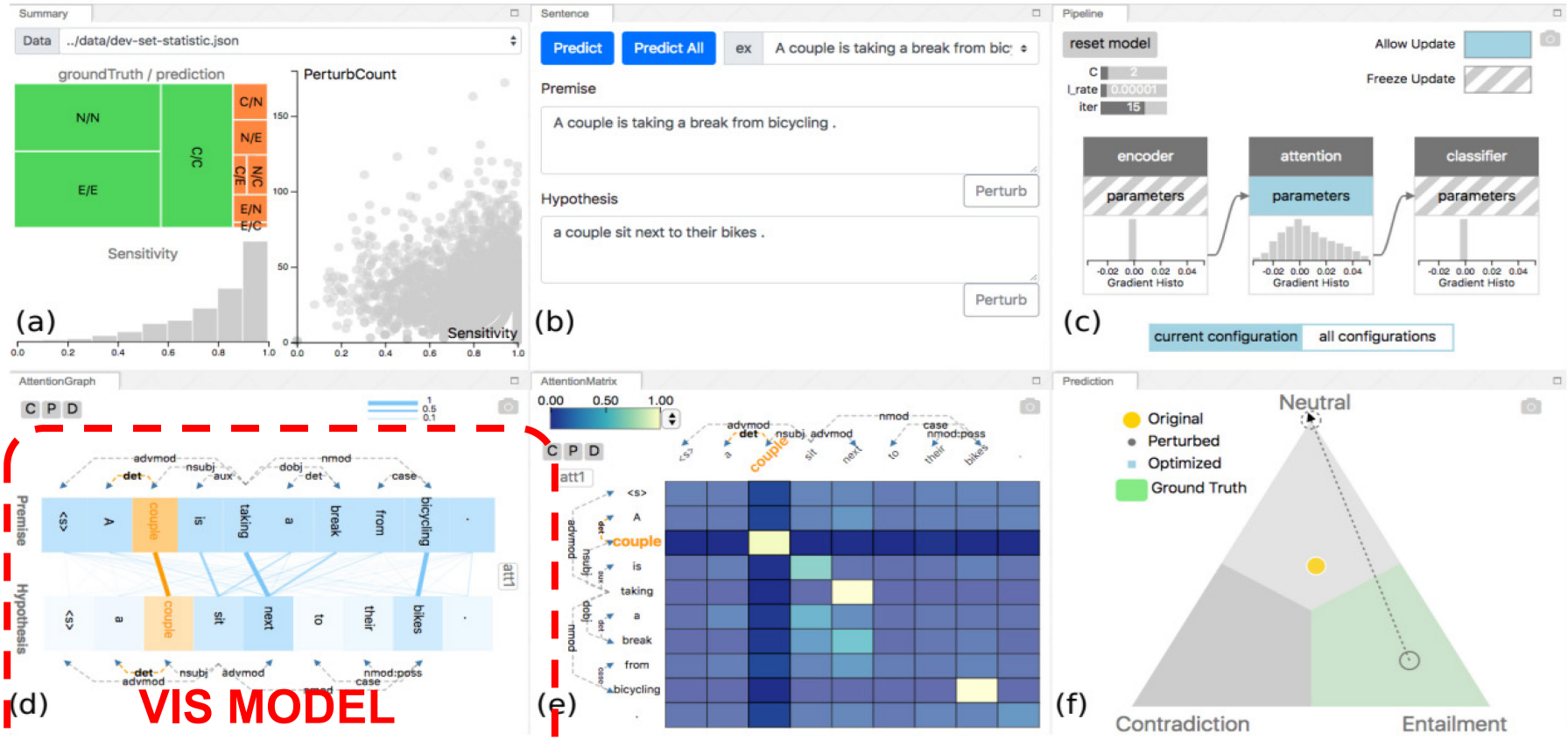


Interactive tool for understanding NLP Inference

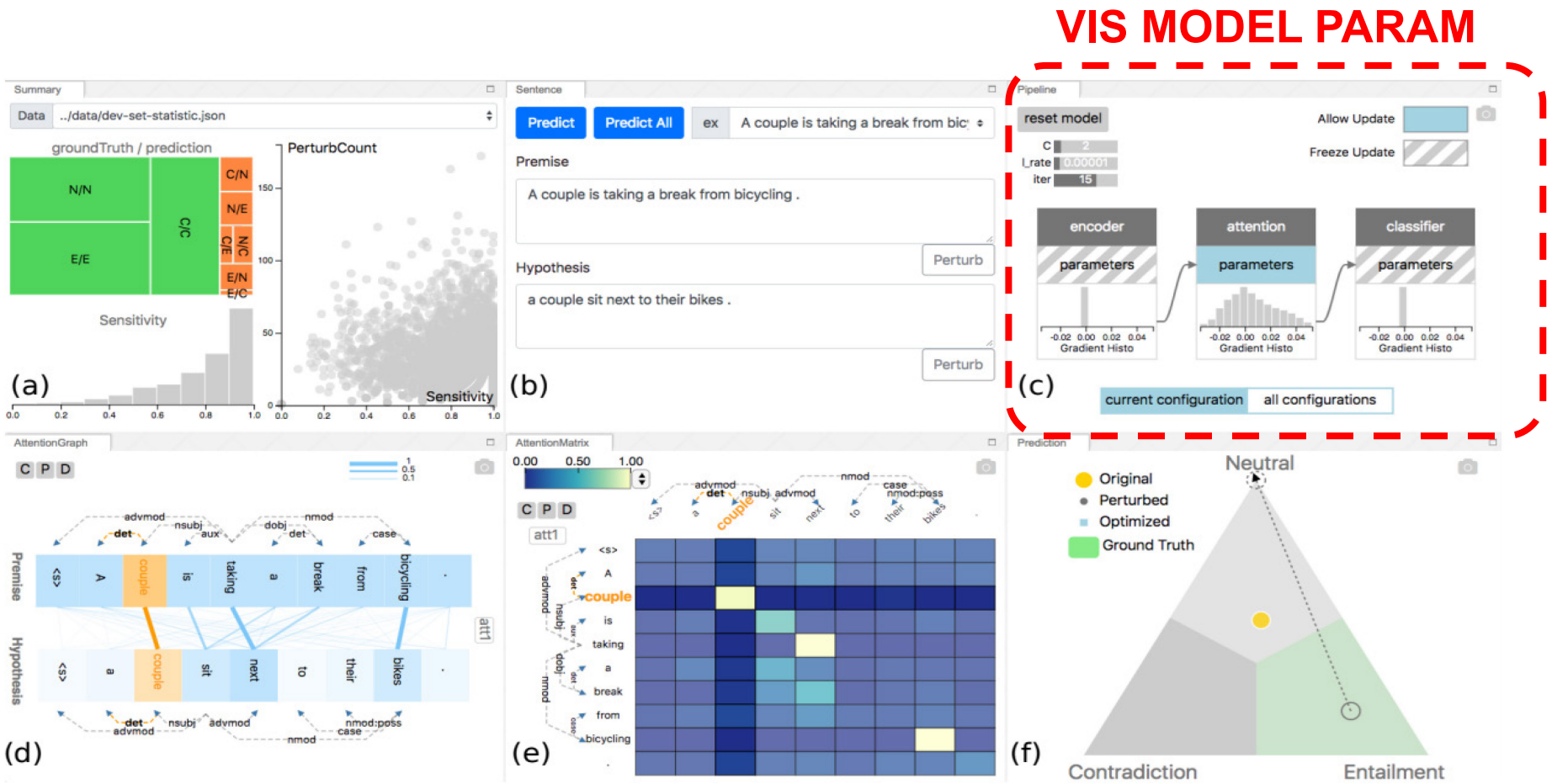
PERTURB TRAINING SET



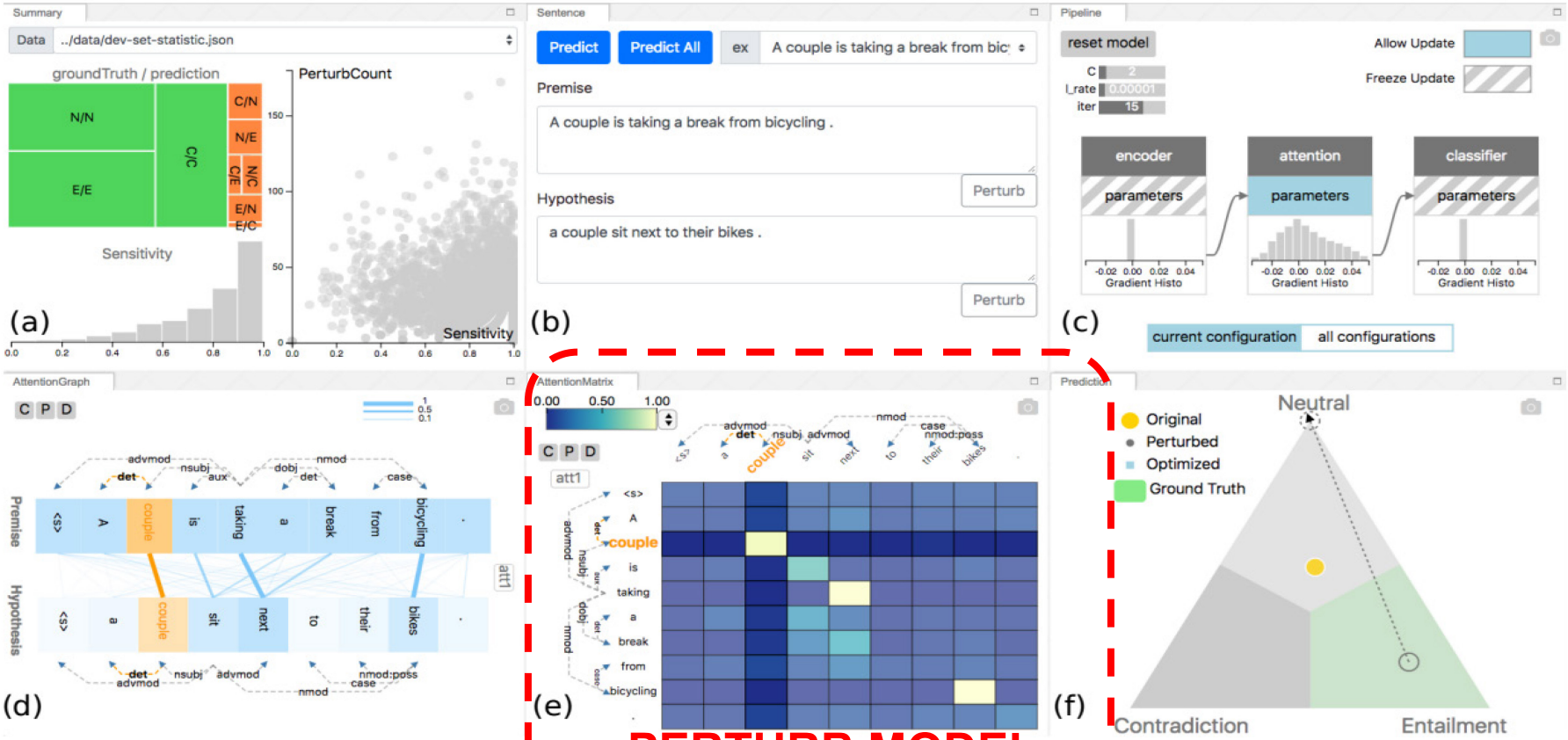
Interactive tool for understanding NLP Inference



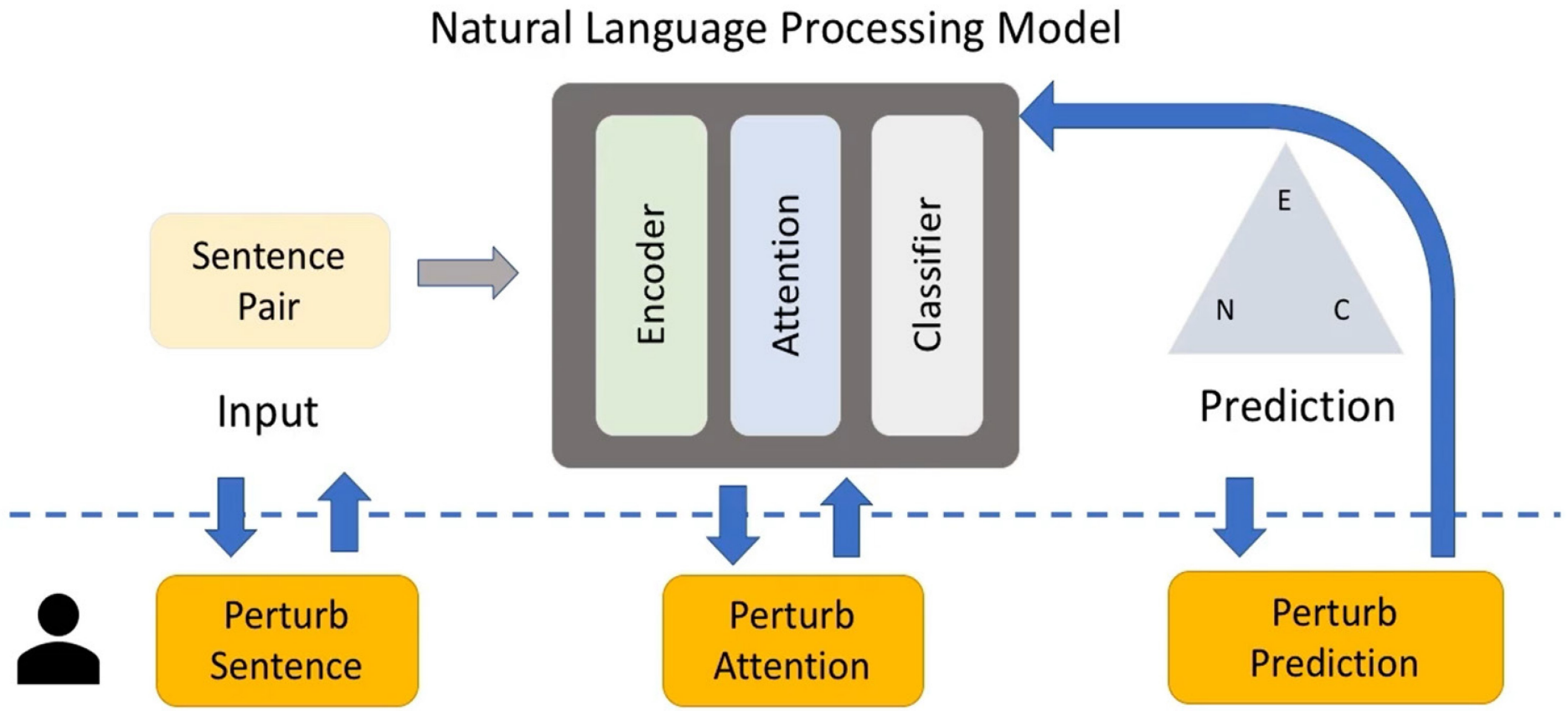
Interactive tool for understanding NLP Inference



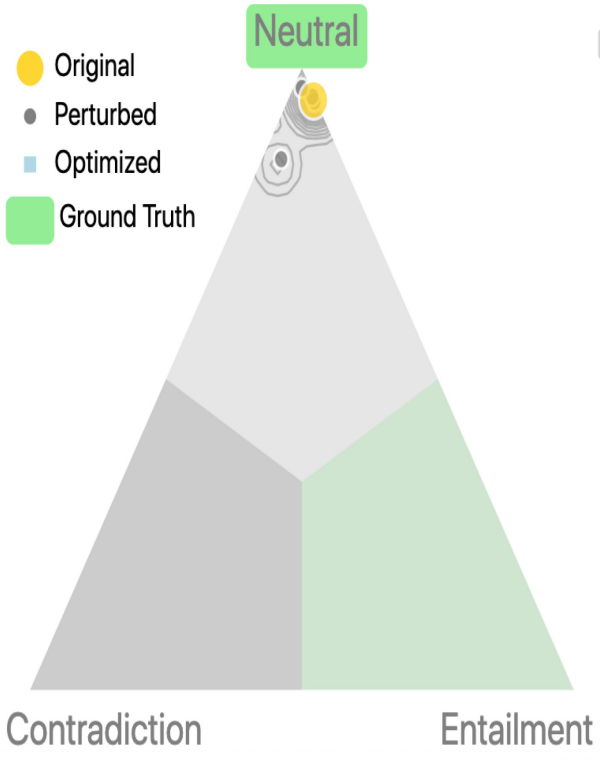
Interactive tool for understanding NLP Inference



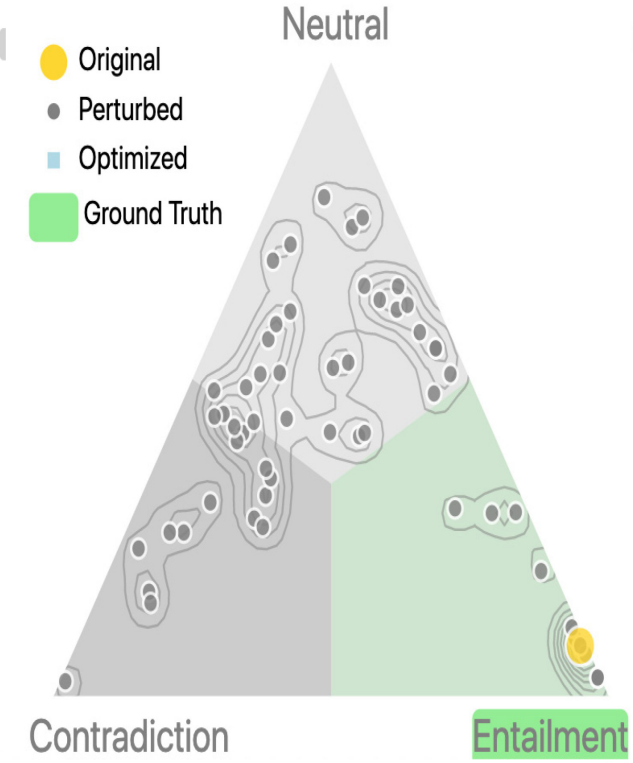
What can we manipulate intuitively to explain the ML for NLP?



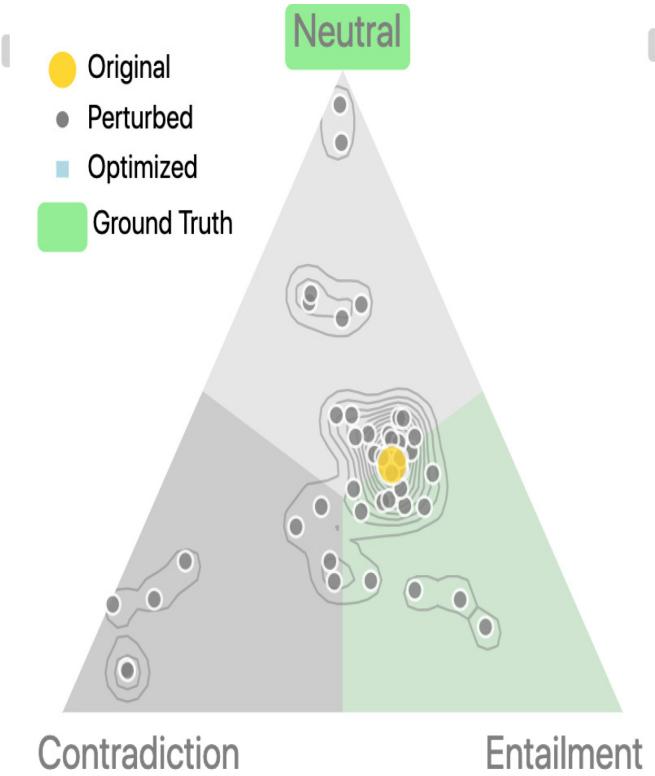
Perturbations Input - Different Modes of Failure After Applying Input Perturbations



Stable

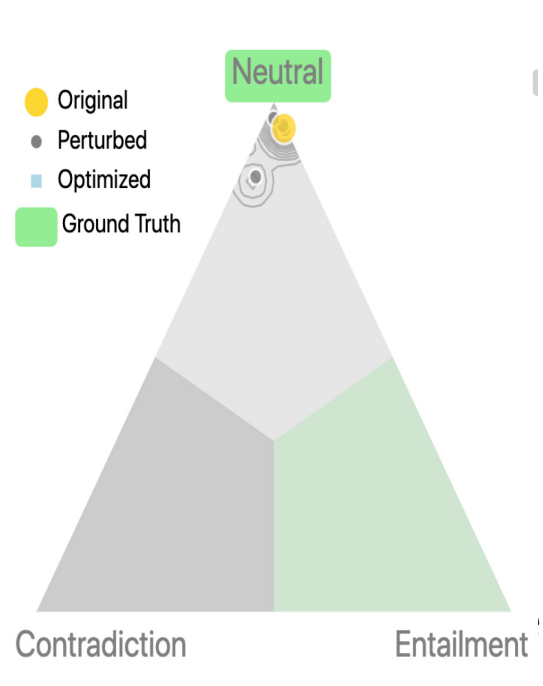


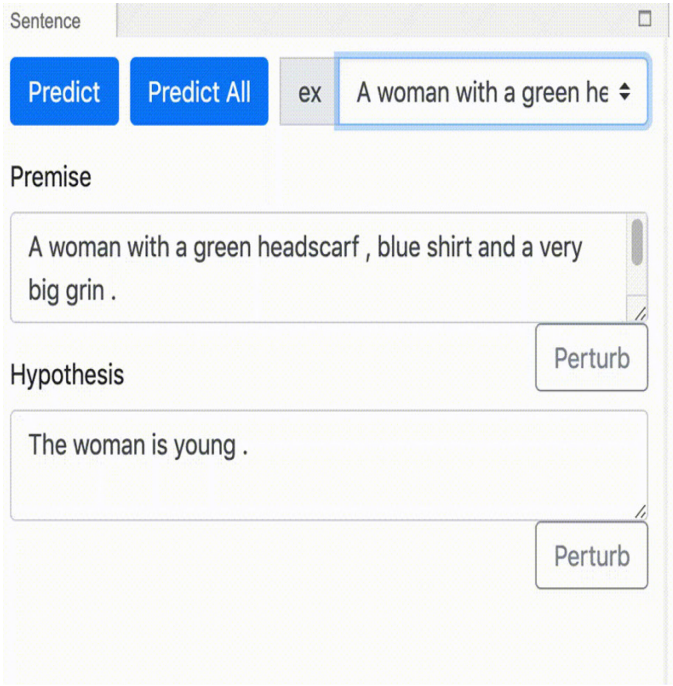
Truly Unstable

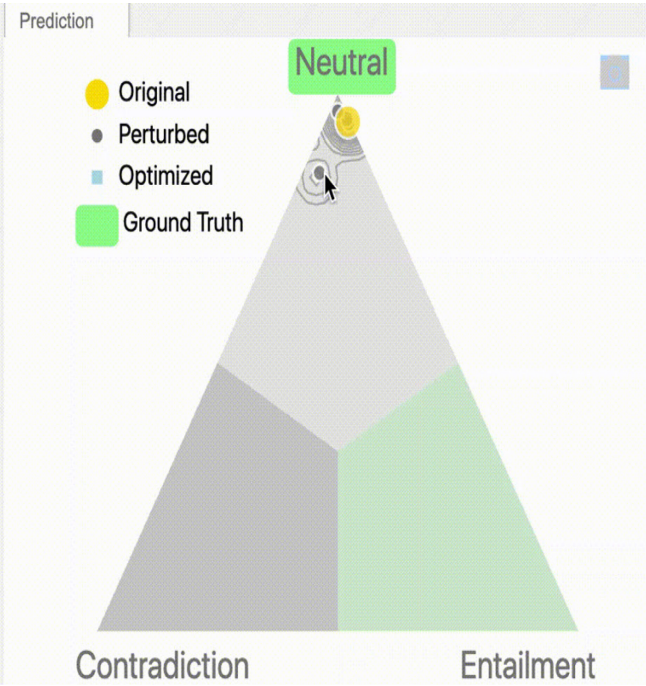


Boundary Condition

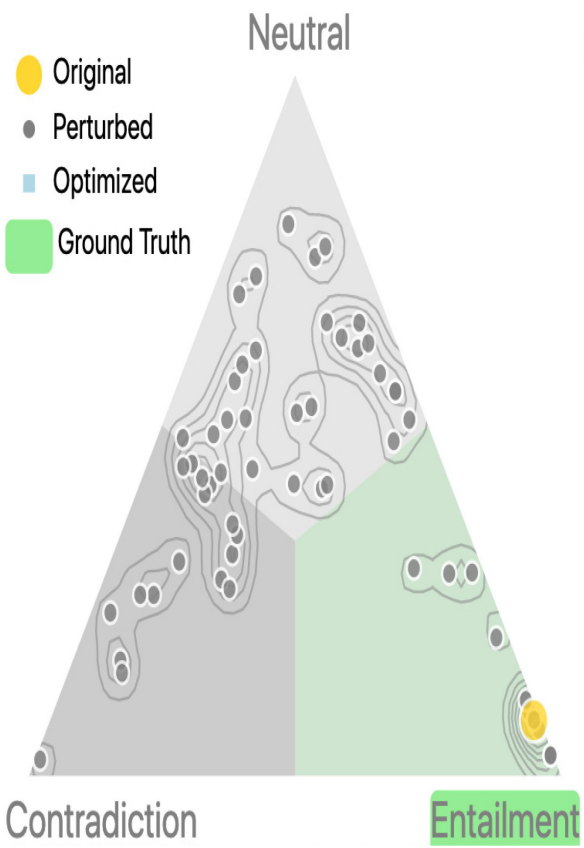
Perturbations Input - Different Modes of Failure After Applying Input Perturbations

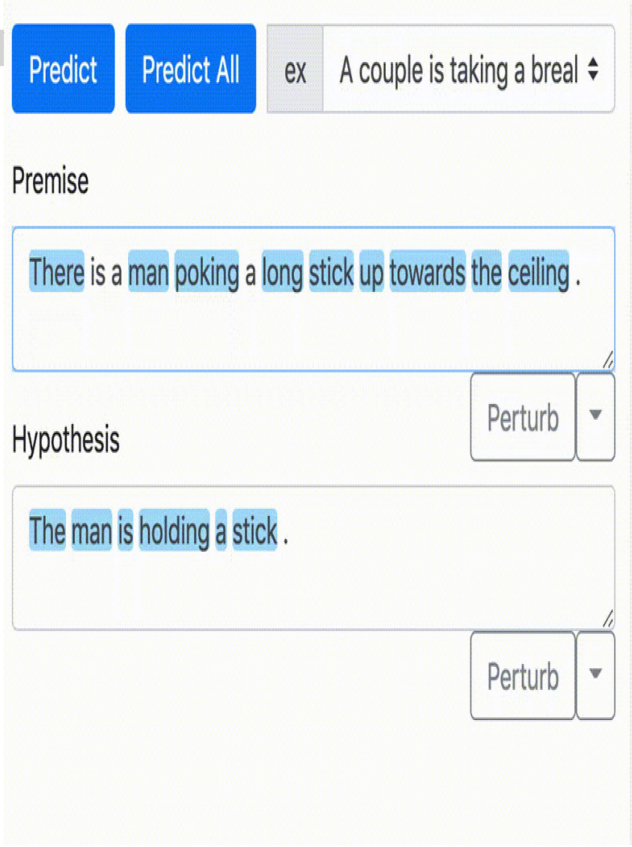


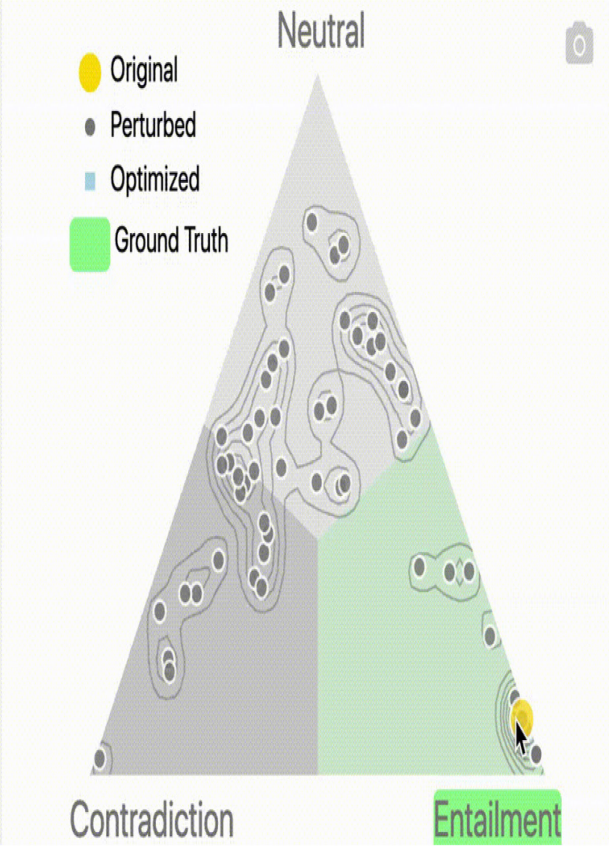




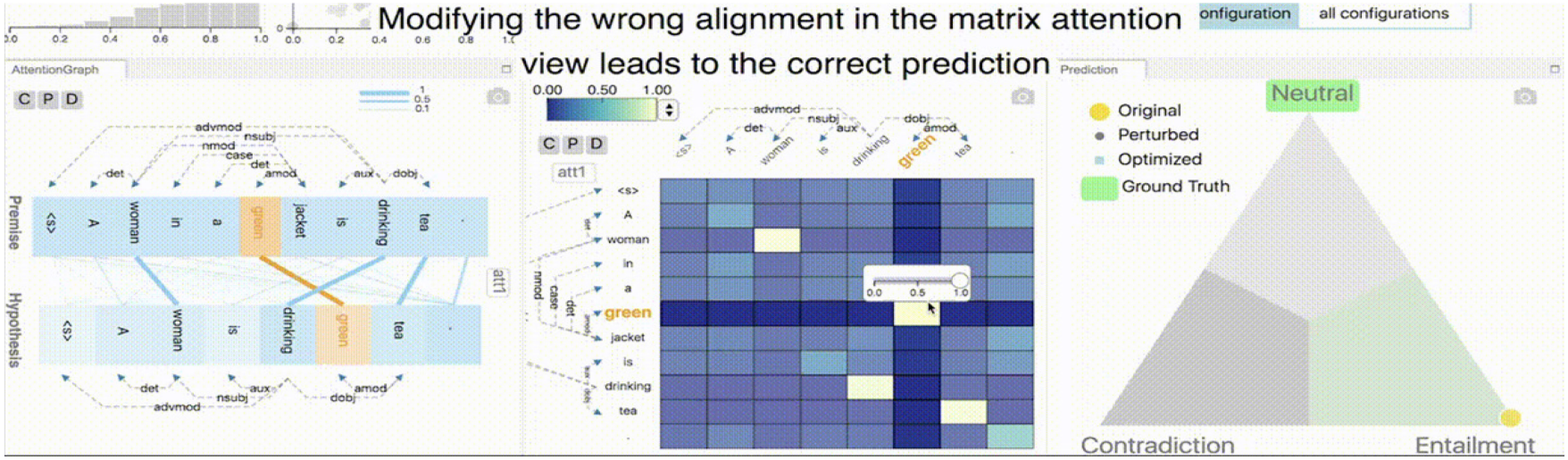
Perturbations Input - Different Modes of Failure After Applying Input Perturbations

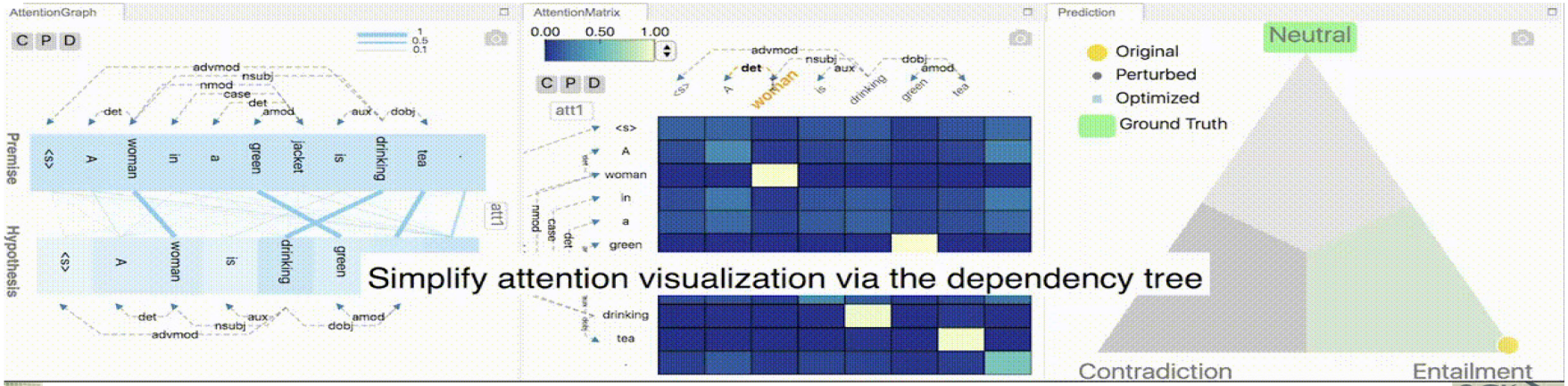






Perturb Attention - Edit the attention to correct a prediction



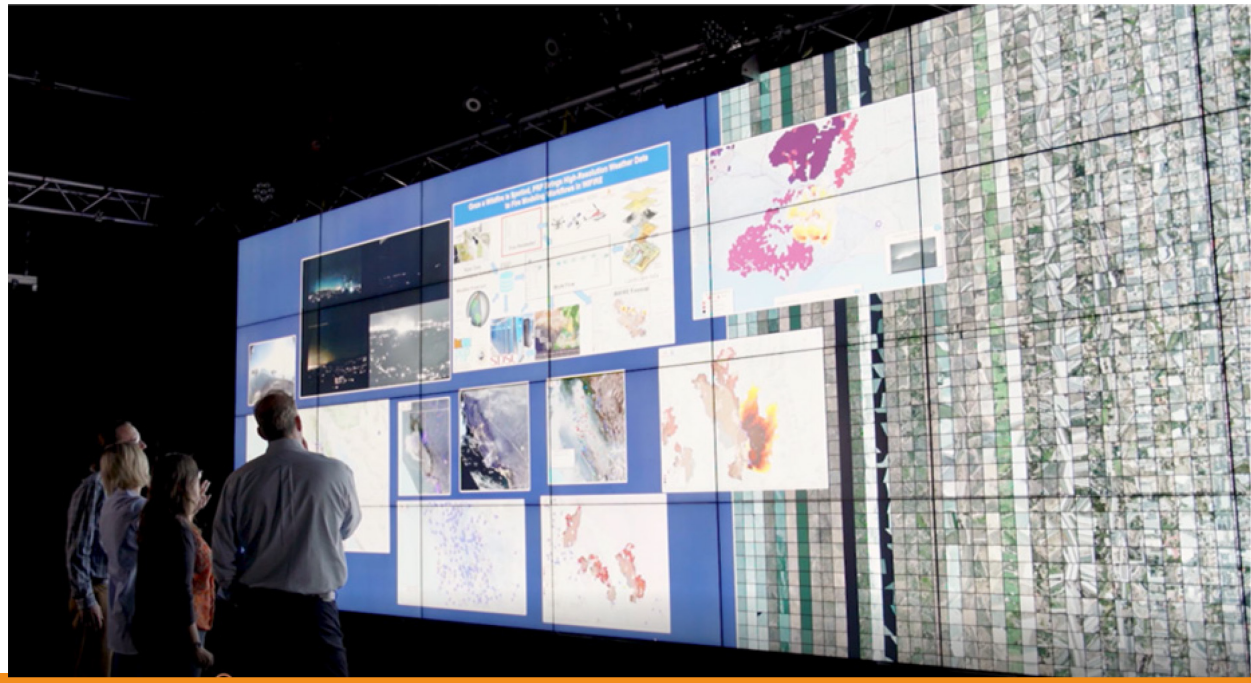


EXPLAINABILITY INCLUDING CATEGORICAL DATA FOR AI AND COMPLEX SYSTEMS

WIFIRE: TOWARDS MAKING WILDFIRE MANAGEMENT EXPLAINABLE

WIFIRE project – Tools for fire management community

- Led by UC San Diego's WIFIRE Lab
- Develop science-driven technologies
- Deployed with the fire management community



BurnPro3D is a decision-making tool to plan safe and effective prescribed burns

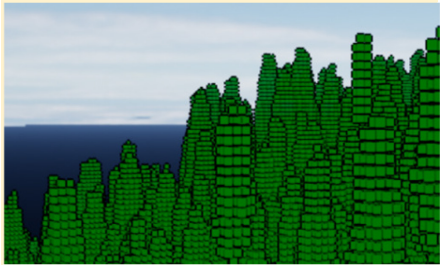
- Users create scenarios
- Combines LIDAR inputs and Geo data
- Simulation ensembles for burn scenarios
- Variations in:
  - Environmental conditions
  - Fuels data
  - Ignition Patterns
- Advanced metrics for evaluation
- ML proxies
- Landscape level optimization



Four Pillars of BurnPro3D

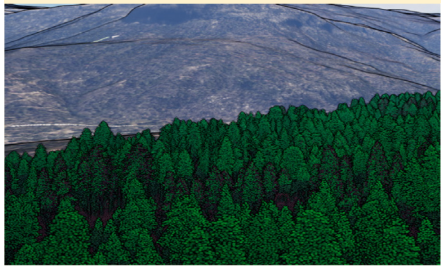
FastFuels

3D Fuels for Fire Models



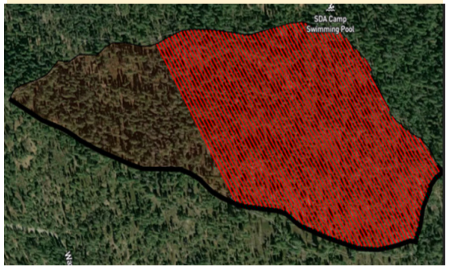
TrueTrees

Lidar-Derived Tree Positions and Heights



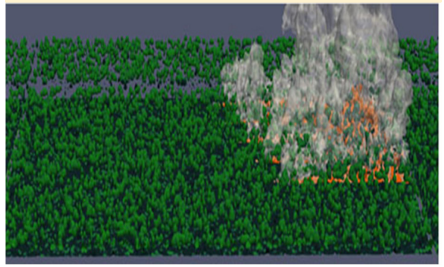
DripTorch

Custom Ignition Patterns



QUIC-Fire

Prescribed Fire Simulator

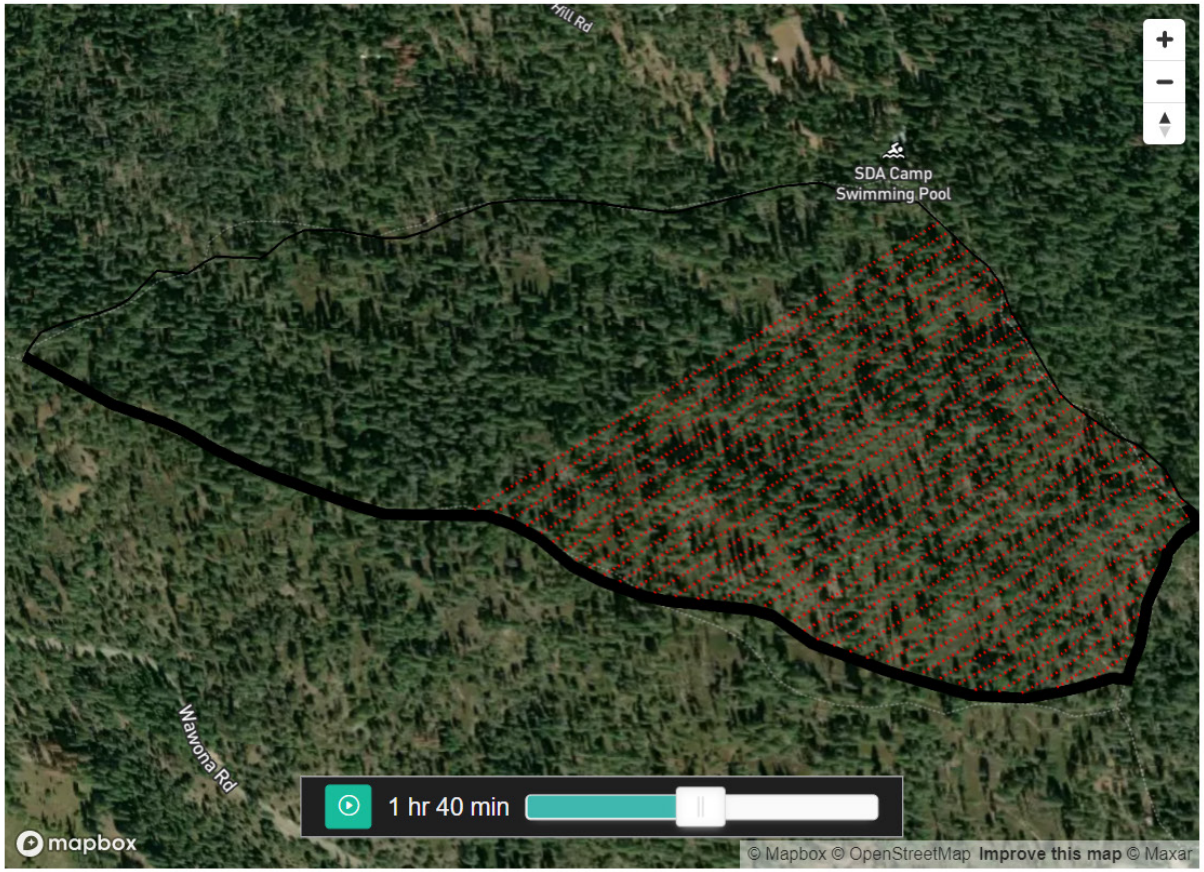


Machine models augment human interpretation: reduce human labeling and recommendations

- Replicate a domain expert in labeling data
- Asses is a prescribed burn is safe

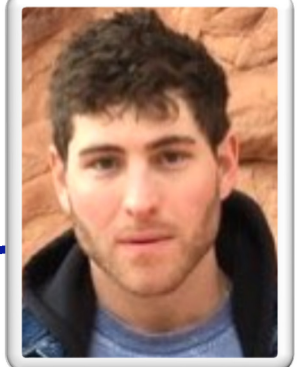
Name	Strip and Dash
Description	Firing @ 330
Firing Technique	Strip
Ignition Type	Dashes
Dash Length (m)	2
Gap Length (m)	4
Number of Igniters	5
Igniter Velocity (mph)	2.25
Stagger Distance (m)	10
Depth Distance (m)	15
Pass Distance (m)	15
Side of Wind to Start Ignition	Left
Blackline Width (m)	10
Control Line Width (m)	2

Total ignition time: 3 h 22 m

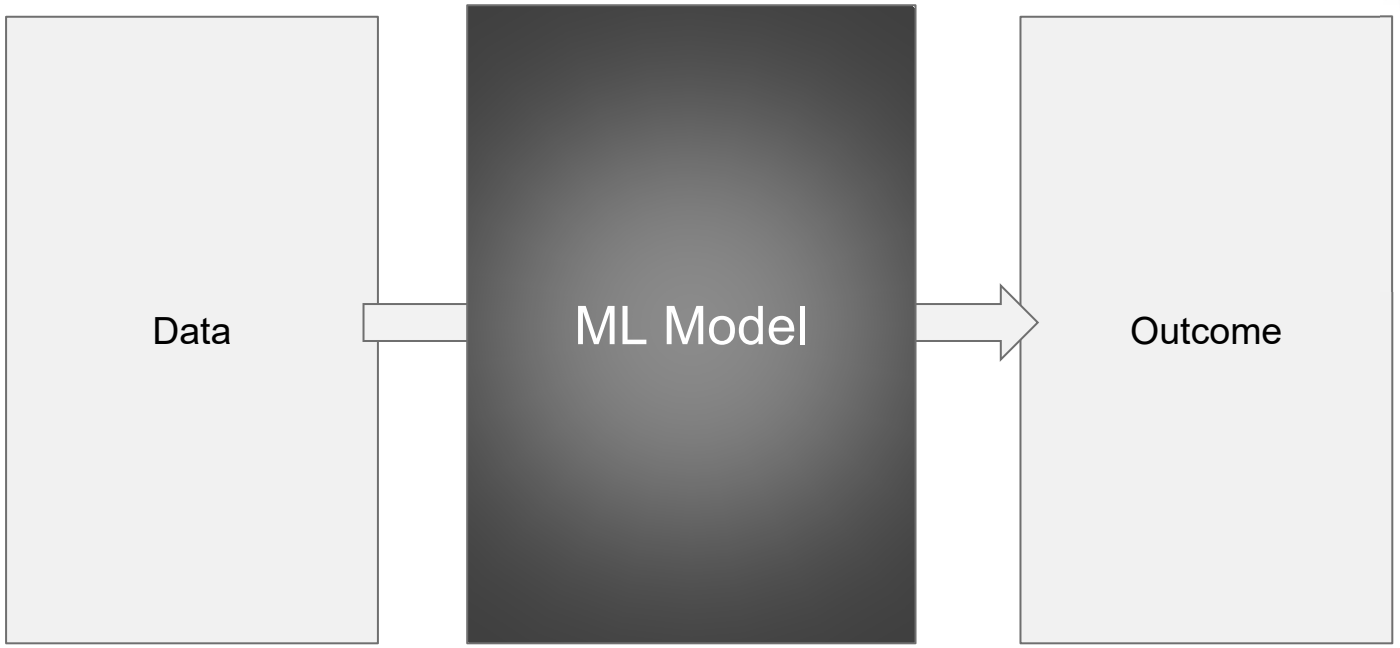


Level 0: Explaining a trained black box model

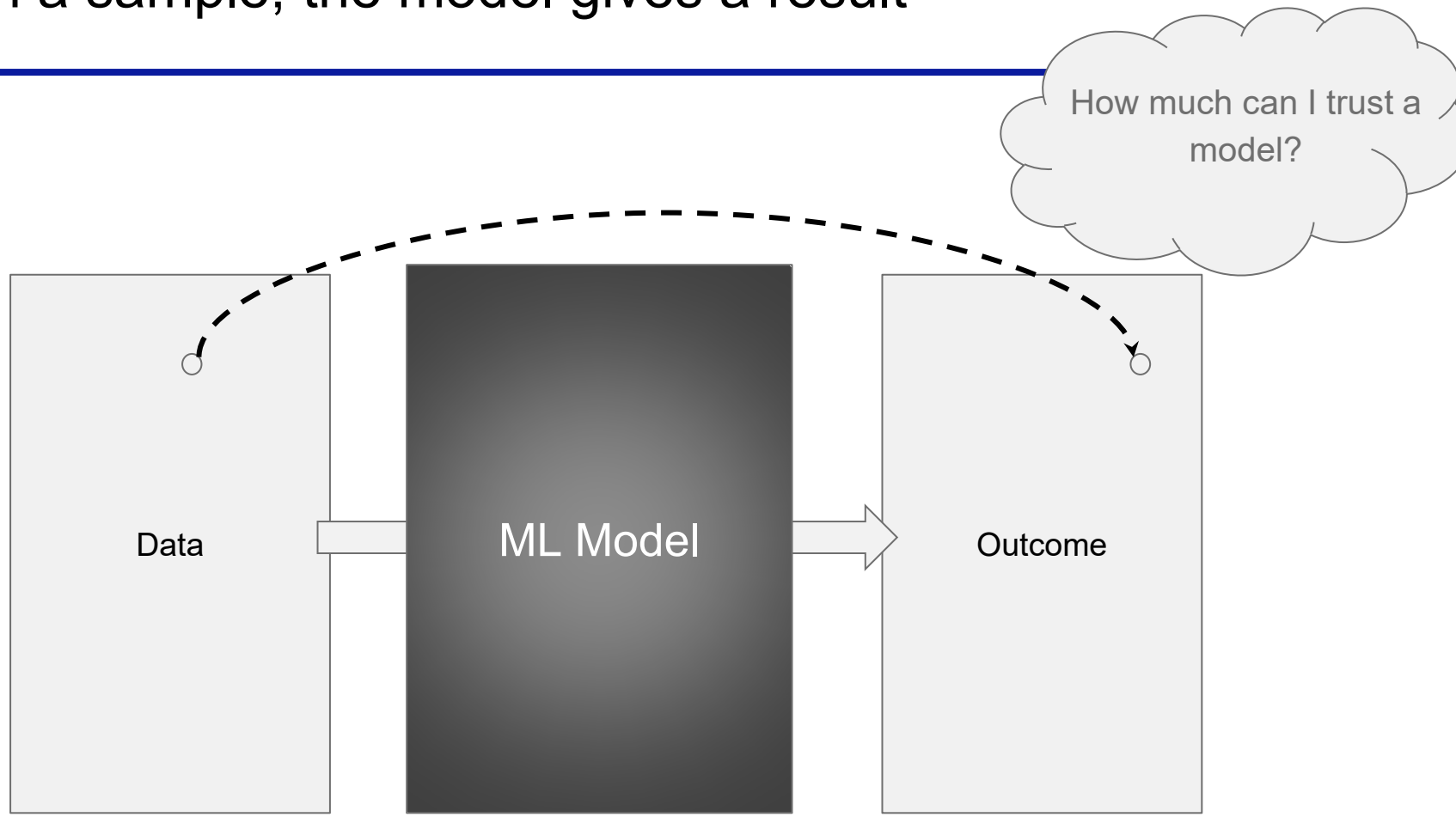
ML can be an opaque process to a user



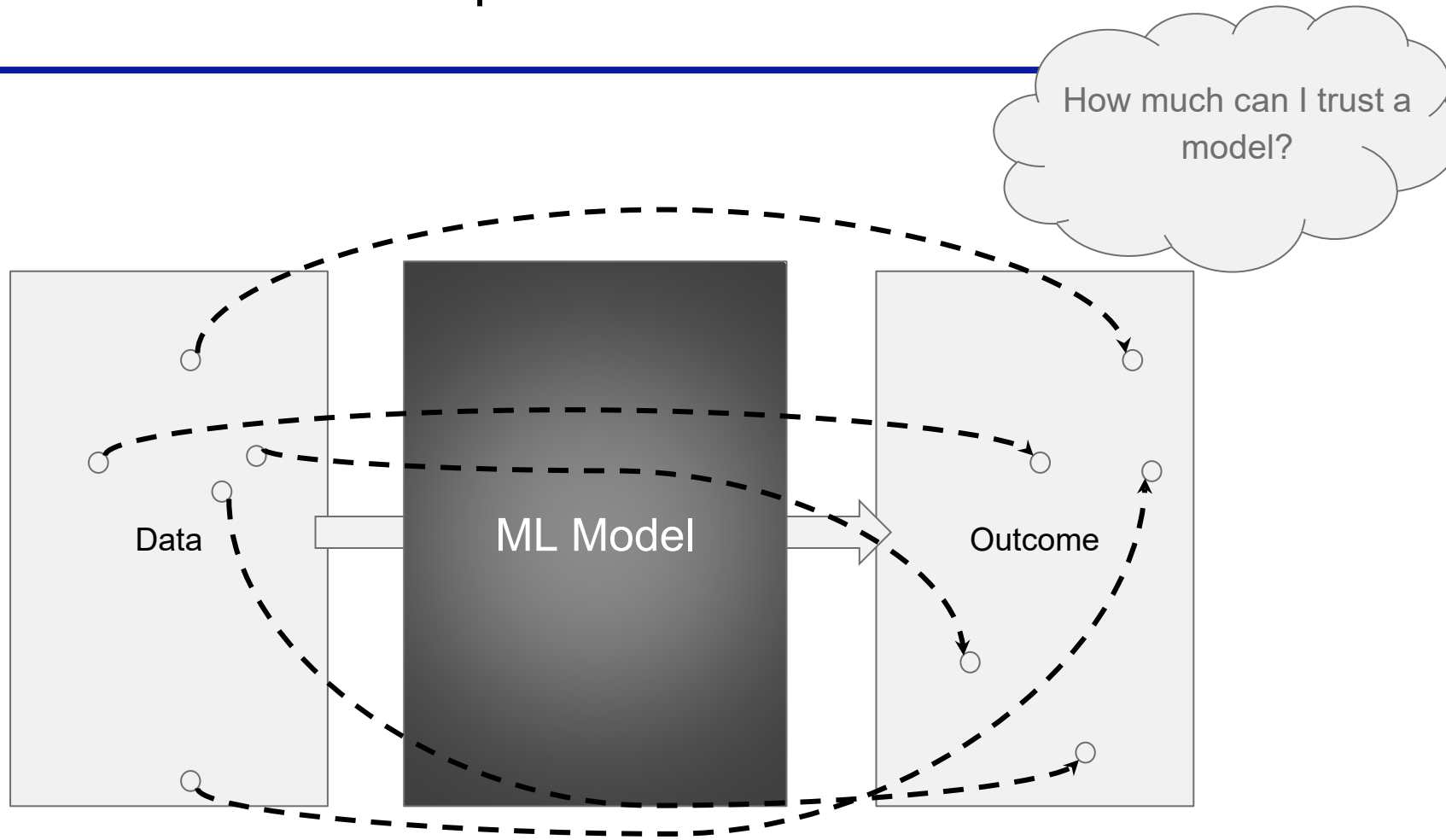
Sam Leventhal



Given a sample, the model gives a result



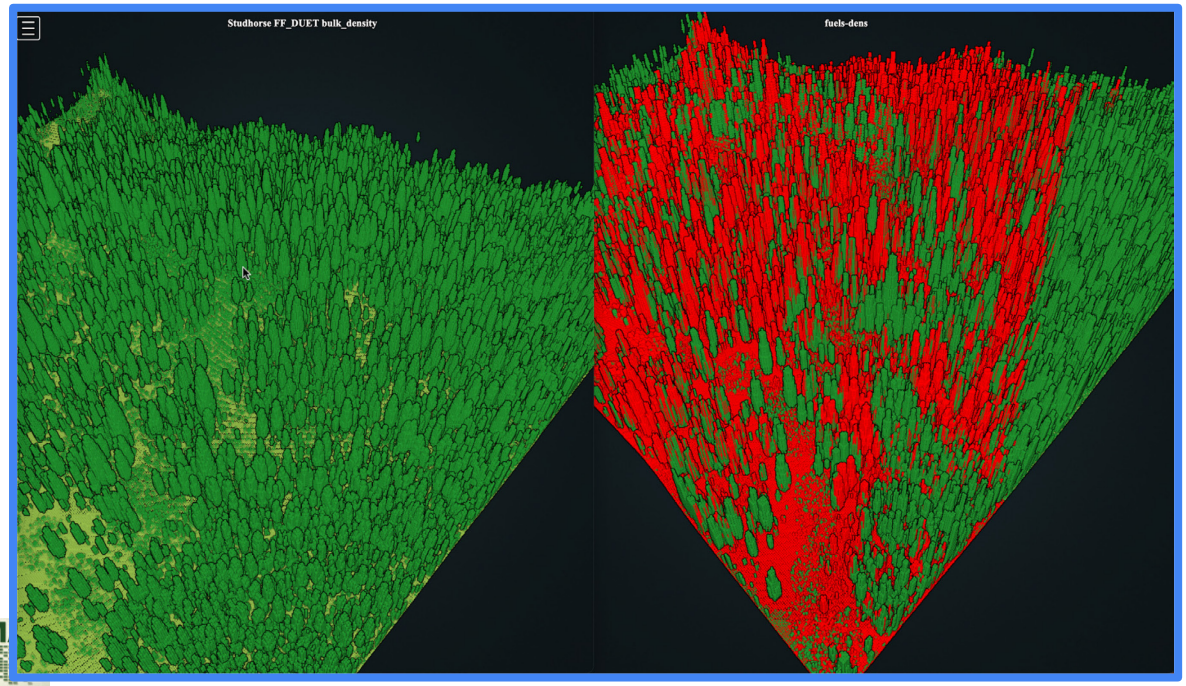
Co-visualization of samples and outcomes



Co-visualize results on top of samples

sample->outcome outcome->sample

<u>Visualization uses:</u>
Context to show outliers
Turn on-off outcomes



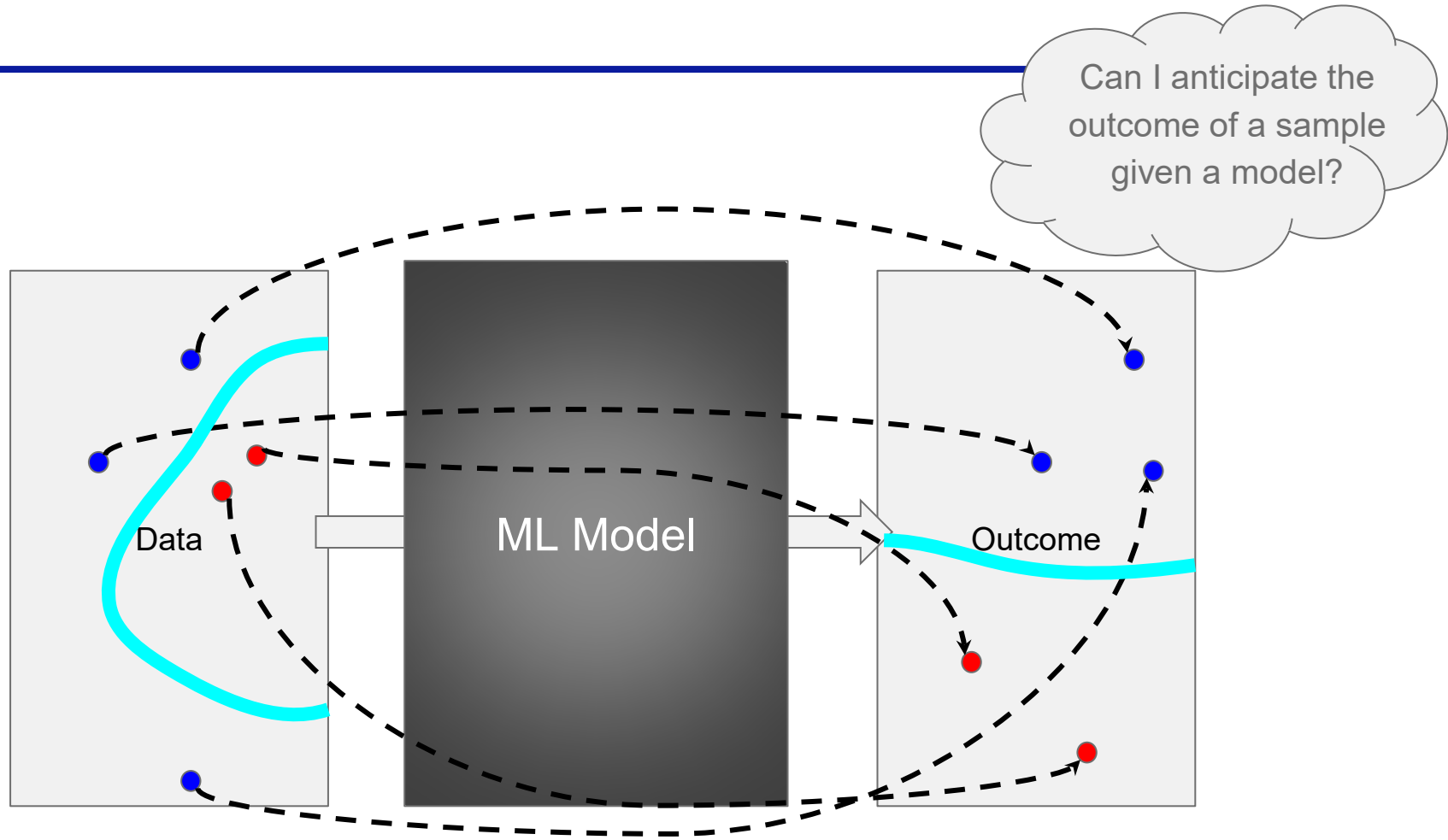
How much can I trust a model?

<u>Added to Vis Tool:</u>

Capability to load various outcomes

Color by class/prediction value

Co-visualization of samples and outcomes



Co-visualize results on top of samples

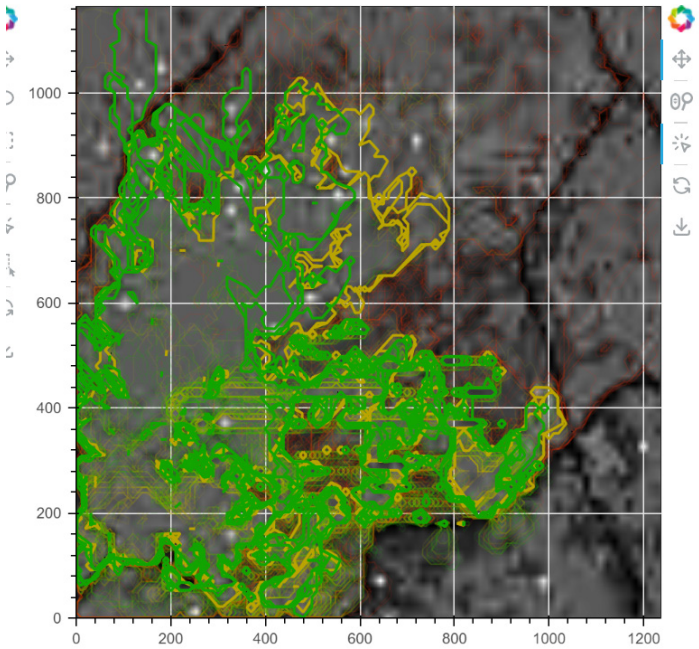
Can I anticipate the outcome of a sample given a model?

sample->outcome

outcome->sample

<u>Visualization uses:</u>
Context to show outliers

Turn on-off outcomes



Co-visualize results on top of samples, and high dimensional exploration.

Can I anticipate the outcome of a sample given a model?

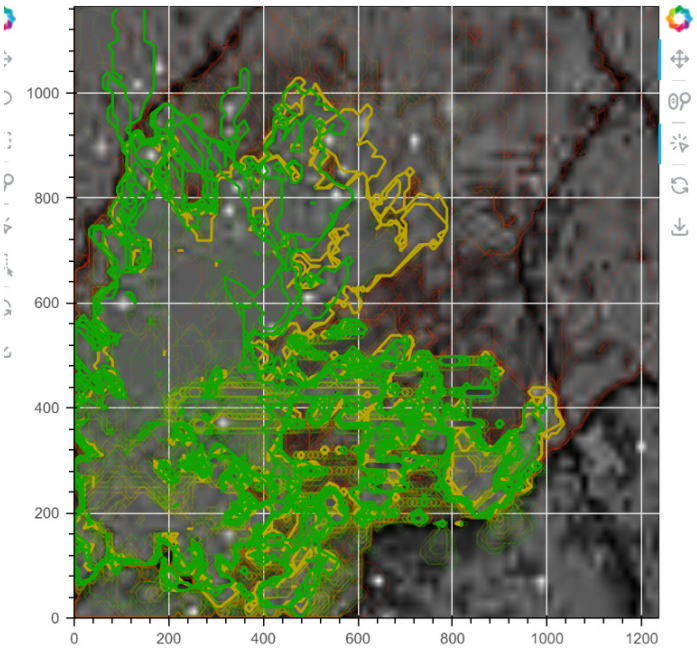
sample->outcome

outcome->sample

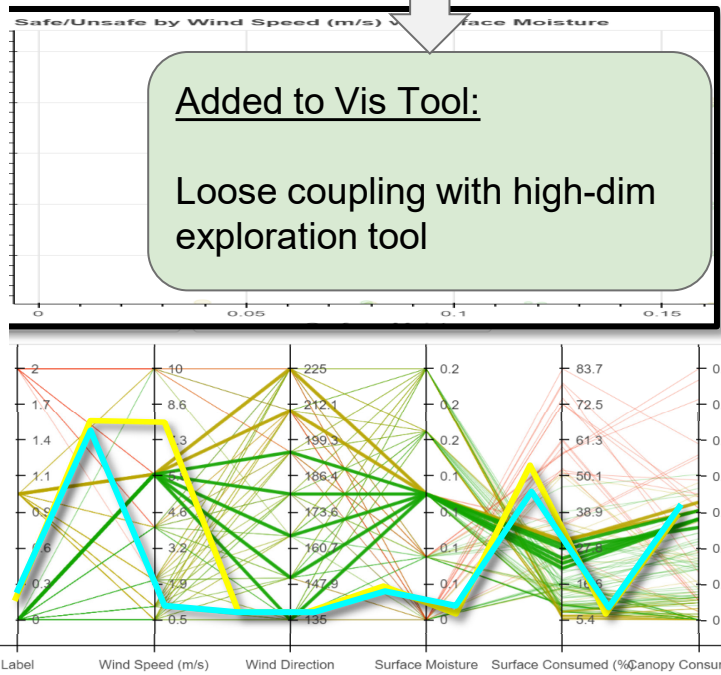
<u>Visualization uses:</u>
Context to show outliers

Turn on-off outcomes

Brushing samples for deeper dive



Regression: PCA, T-SNE



<u>Added to Vis Tool:</u>

Loose coupling with high-dim exploration tool

Parallel coordinates

Co-visualize results on top of samples, and high dimensional exploration.

Can I anticipate the outcome of a sample given a model?

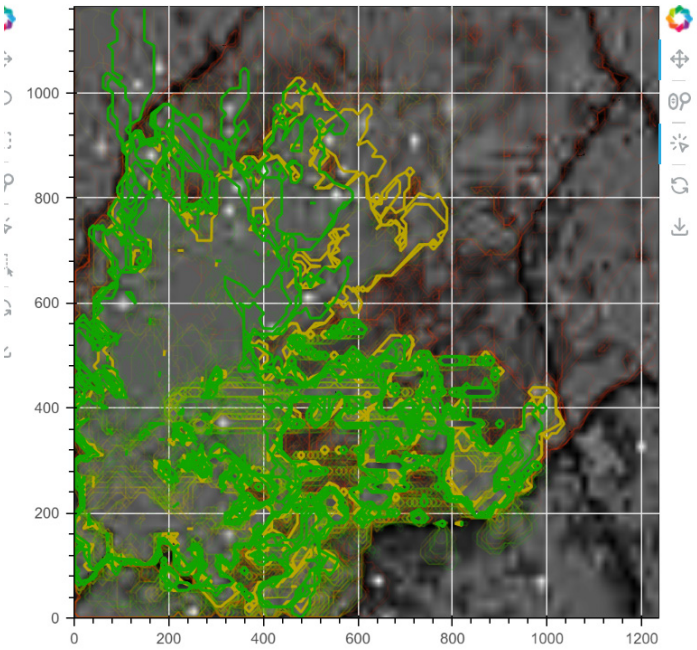
sample->outcome

outcome->sample

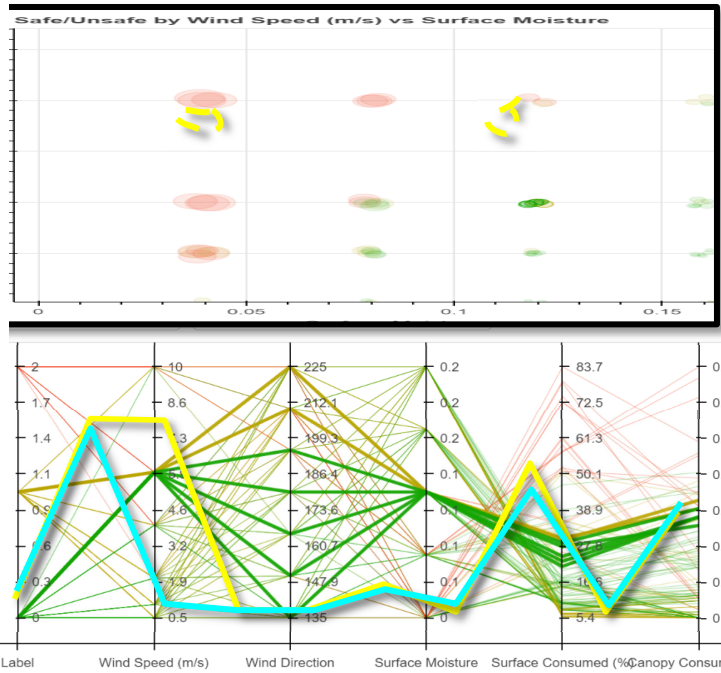
<u>Visualization uses:</u>
Context to show outliers

Turn on-off outcomes

Brushing samples for deeper dive



Regression: PCA, T-SNE



Parallel coordinates

Co-visualize results on top of samples, and high dimensional exploration.

Why did the model give me the result it did?

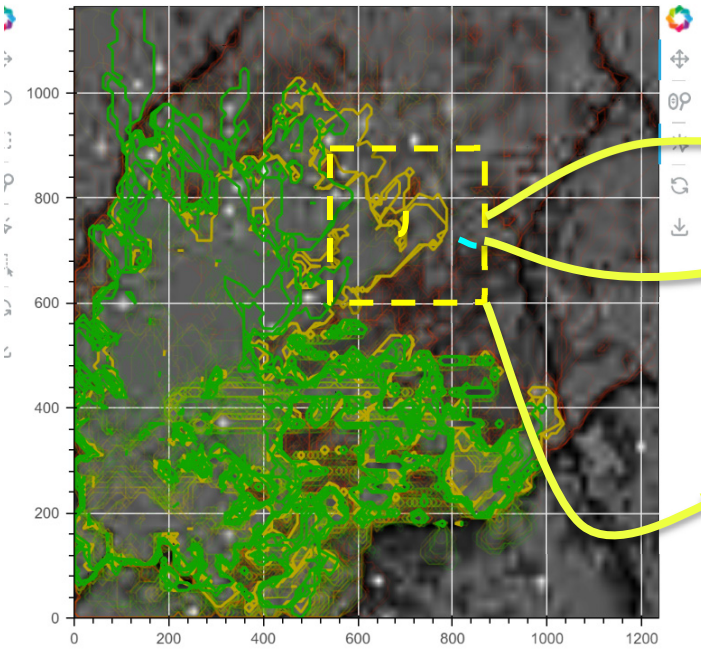
sample->outcome

outcome->sample

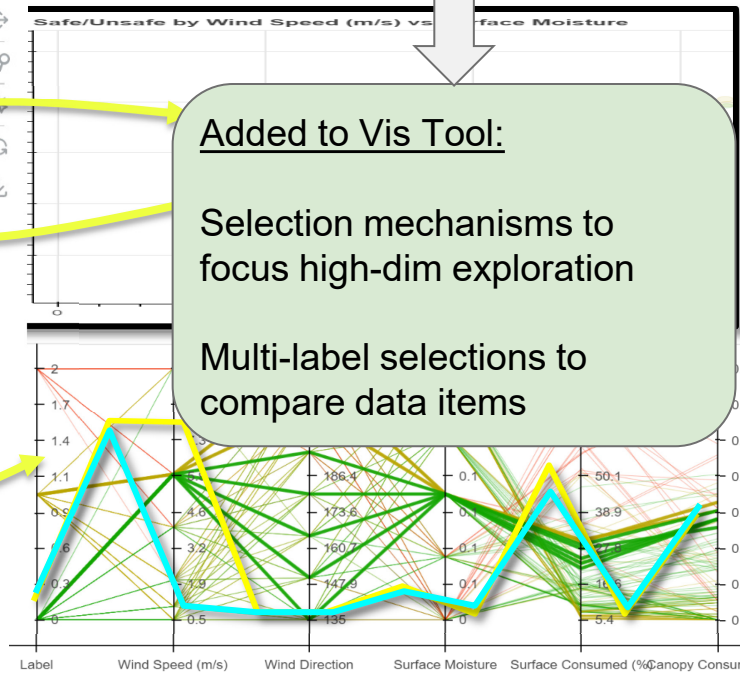
<u>Visualization uses:</u>
Context to show outliers

Turn on-off outcomes

Brushing samples for deeper dive



Regression: PCA, T-S E



<u>Added to Vis Tool:</u>

Selection mechanisms to focus high-dim exploration

Multi-label selections to compare data items

Parallel coordinates

Co-visualize results on top of samples, and high dimensional exploration.

Why did the model give me the result it did?

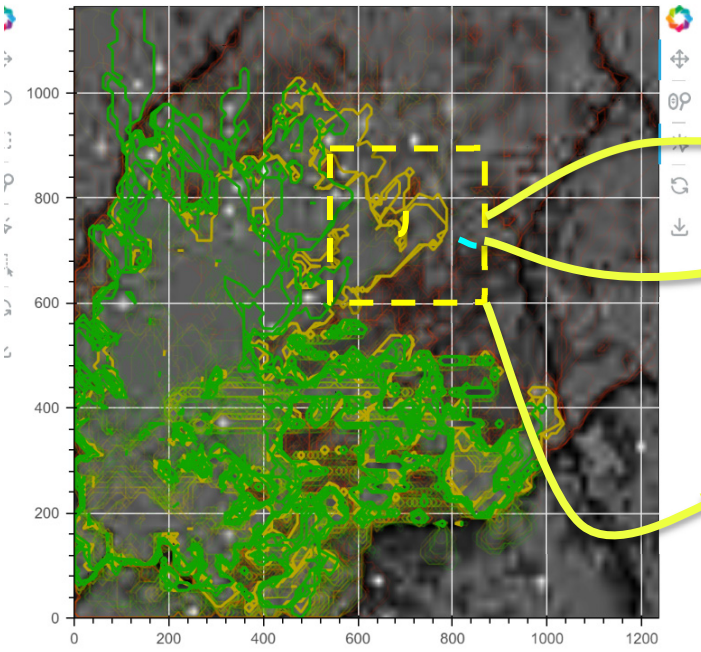
sample->outcome

outcome->sample

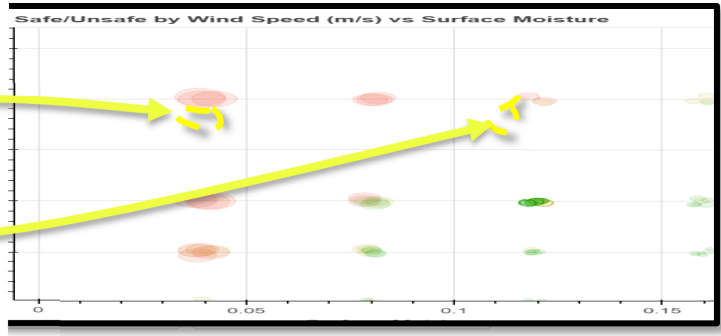
<u>Visualization uses:</u>
Context to show outliers

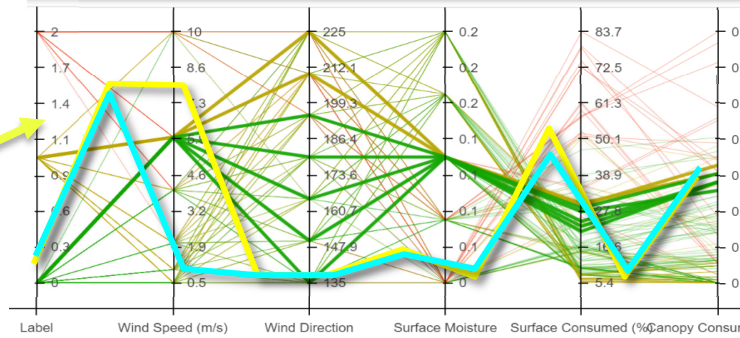
Turn on-off outcomes

Brushing samples for deeper dive



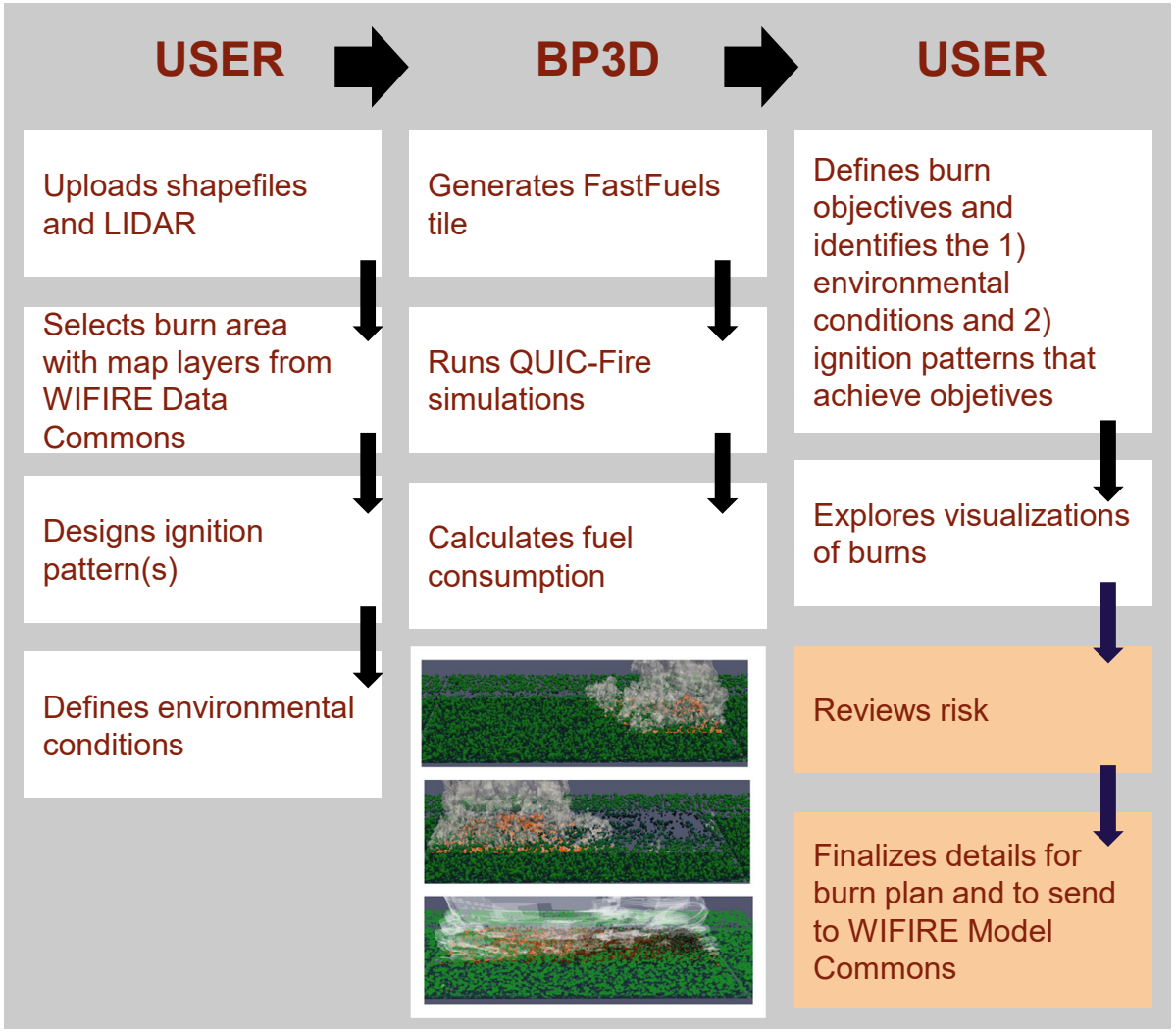
Regression: PCA, T-SNE





Parallel coordinates

User Flow - Scenario Creation/Analysis



Inputs: Map Layers, LIDAR, burn unit shapefiles, ignition patterns

×

Variations: Environmental conditions including wind speed, direction, moisture levels

=

Outputs: 1000's of simulated burns, 10s Tb time-varying 3d arrays

Derived: computed metrics, labeled outputs, AI predicted labels

Tools are needed to support decision-making across stakeholder roles

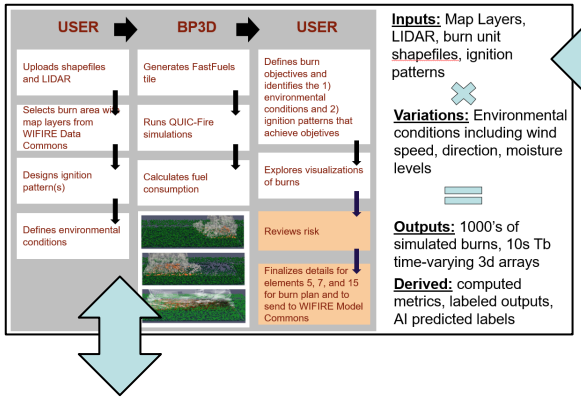
BP3D Developers:

Did the simulations produce expected outputs?

Did hyperparameters have intended effects?

Are there systematic instabilities?

Are outliers “real” or “bug”?



Burn Bosses:

What environmental conditions lead to safe/unsafe predictions?

Can we meet objectives without excess risk?

Are there unexpected hazards (slop over, smoke)?

Day of burn: how risky are the current conditions?

Fire Managers:

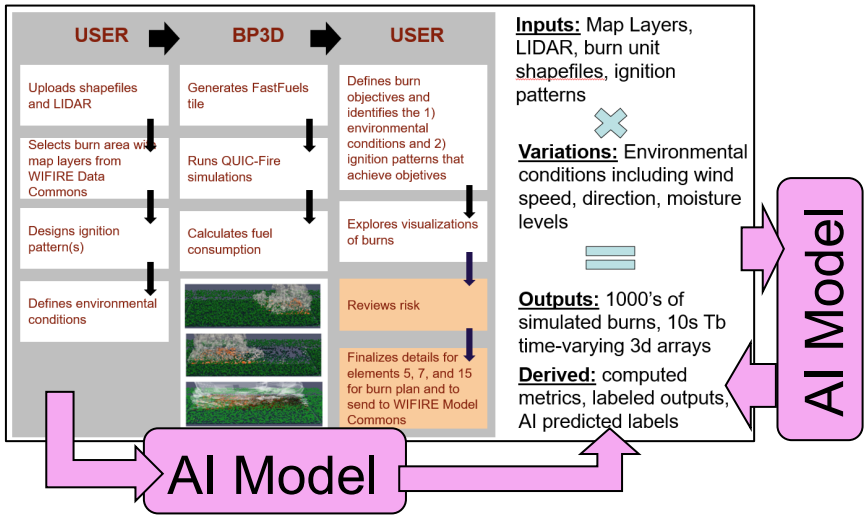
Which burn units should I prioritize this year?

How would changes in available mitigation resources affect fire risks?

How can I make fair decisions?

How do we balance area, habitat, and property risks?

Next Steps: Add interface-driven counterfactual creation/evaluation/vis

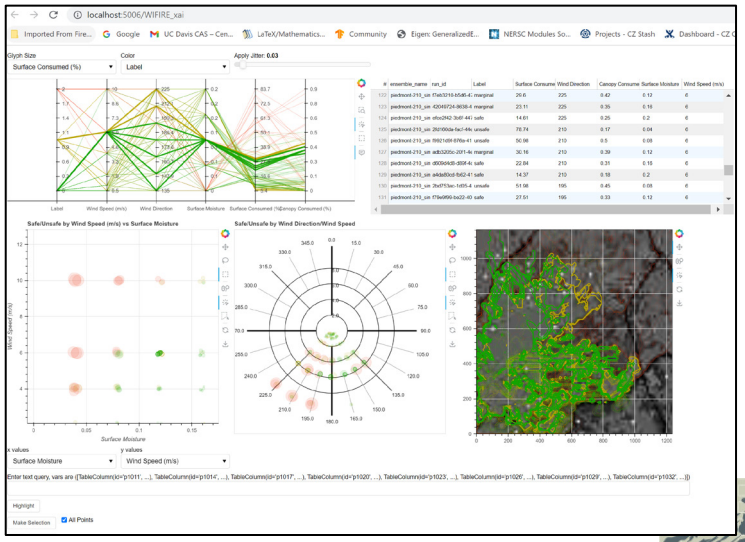


Principles of Explainability to apply:

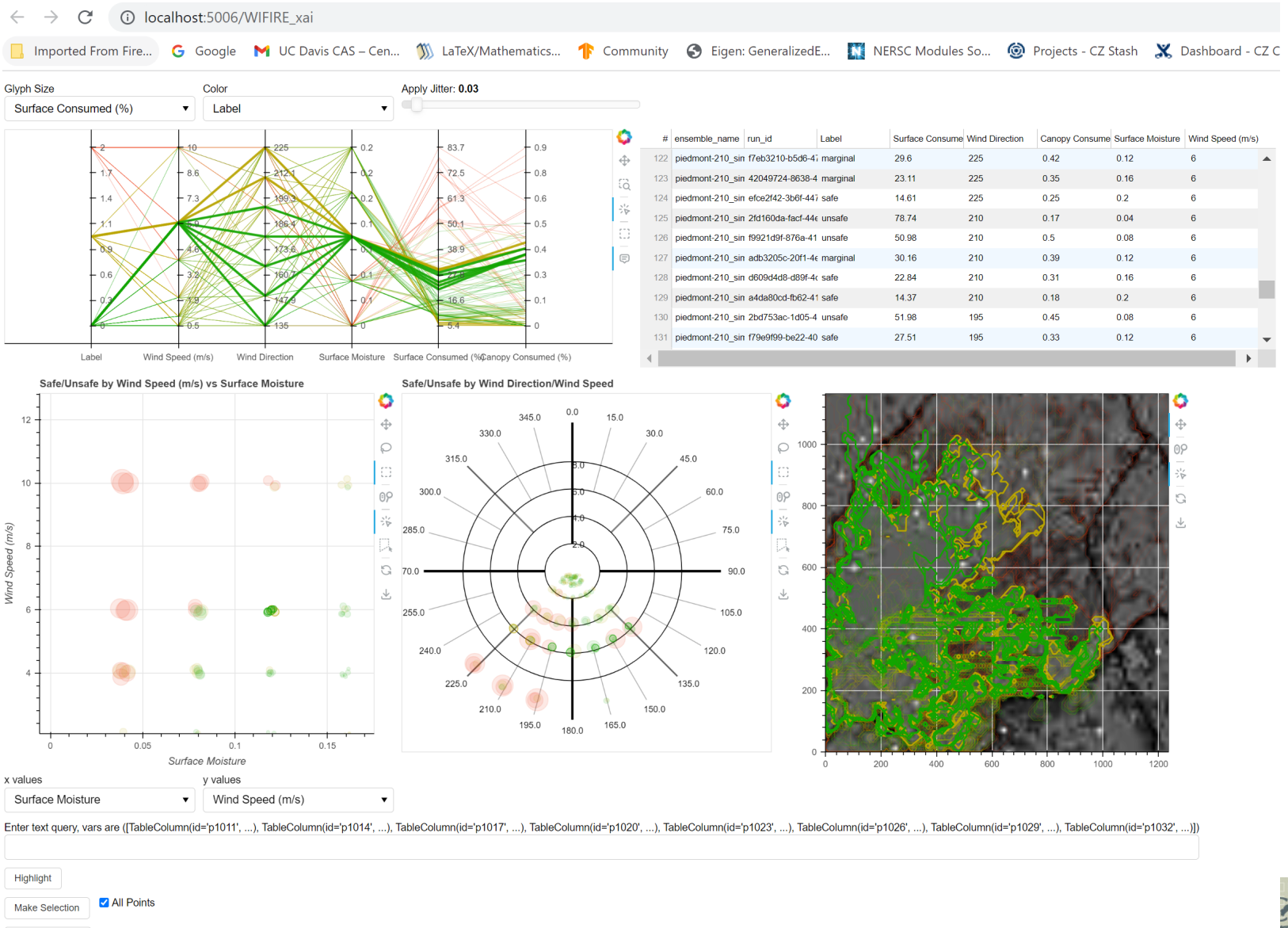
- *co-visualization* of ground truths and predictions throughout interface, and between models (glyph shapes)
- *Enable what-ifs:* draw polygons to "fill" with perturbations, select runs and auto-generate perturbations

AI Models are coming online:

- PGML to speed up fire sim
- Regression models to automate output labeling

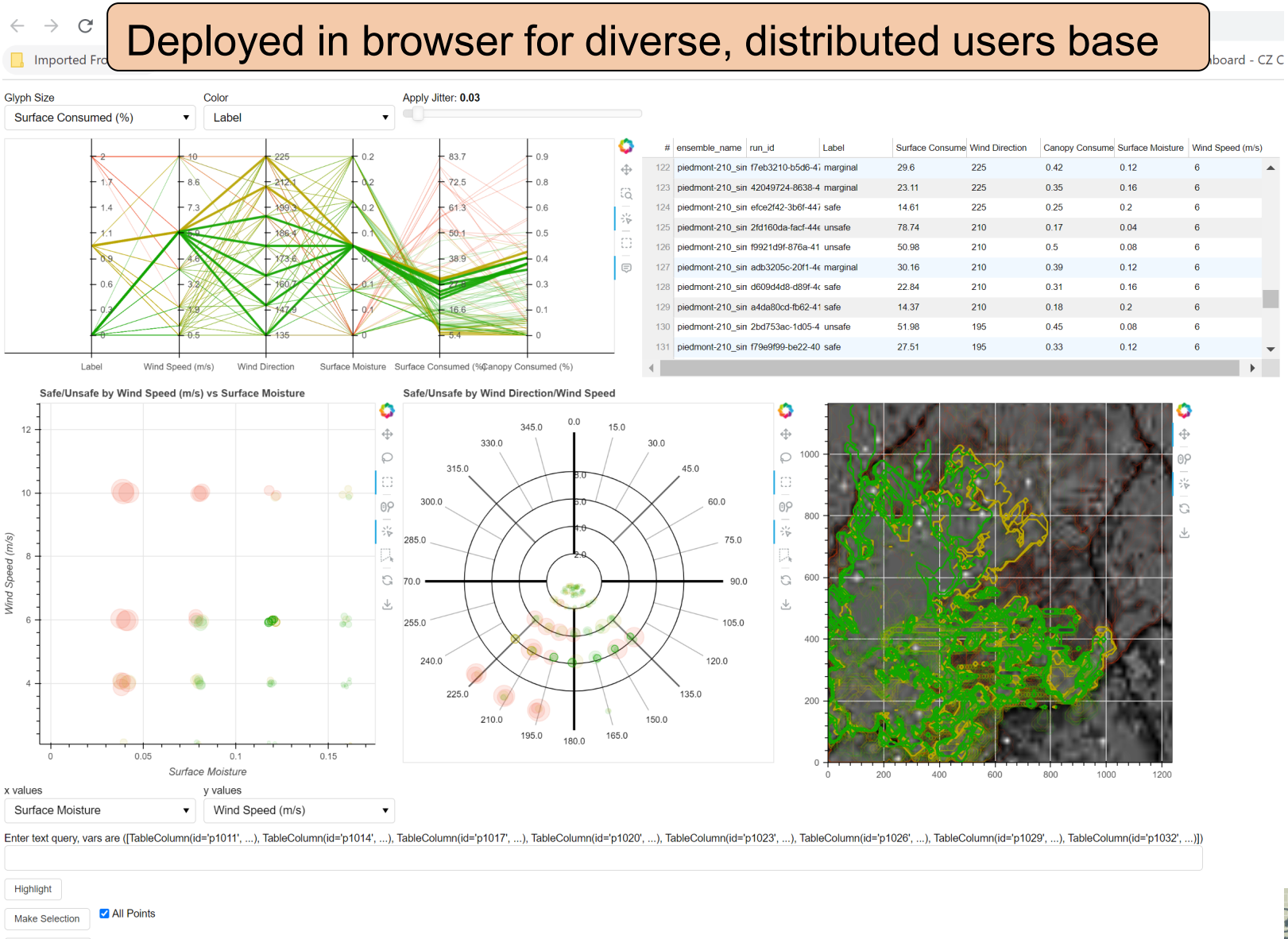


We support developers and burn bosses through an interactive interface

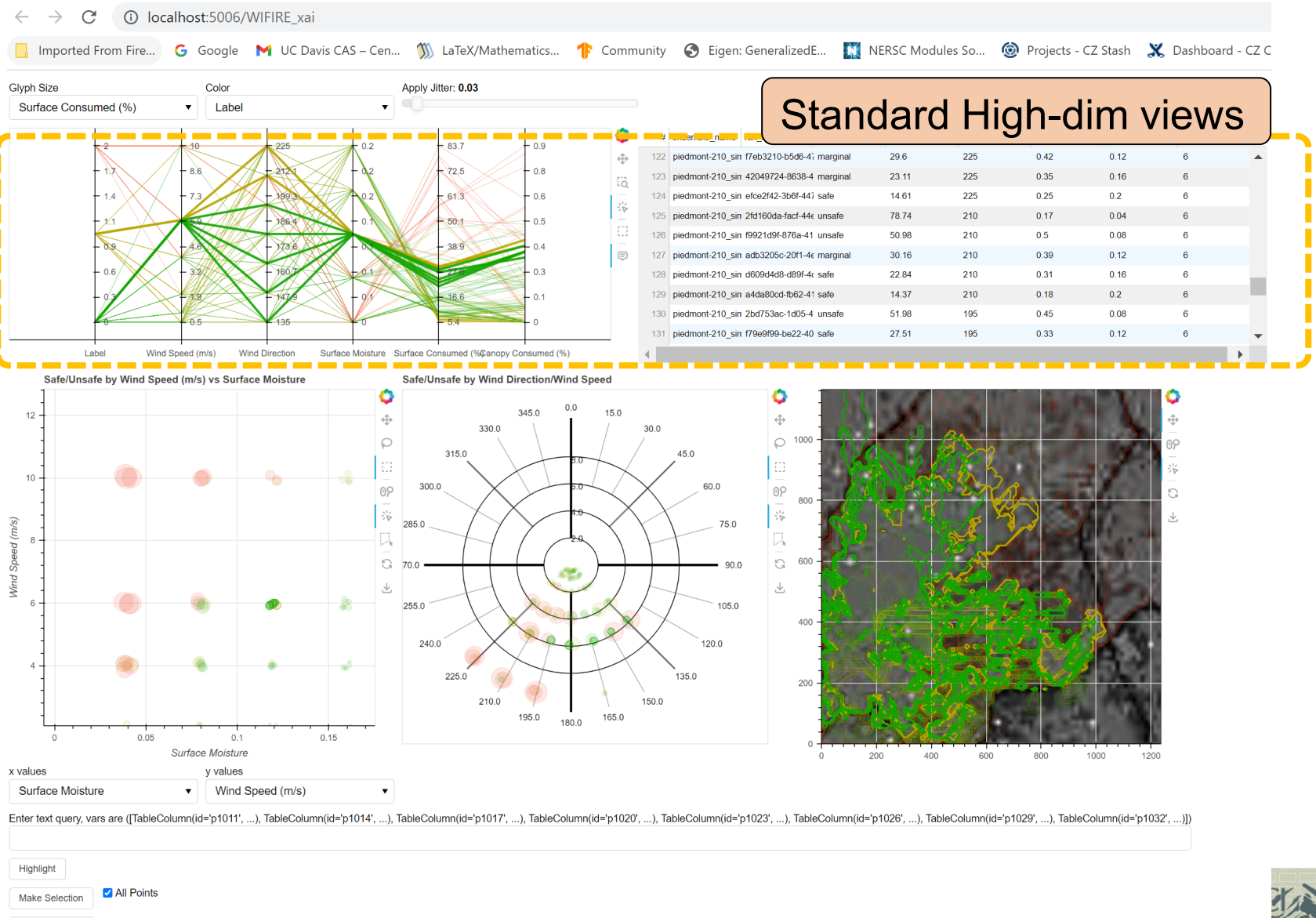


We support developers and burn bosses through an interactive interface

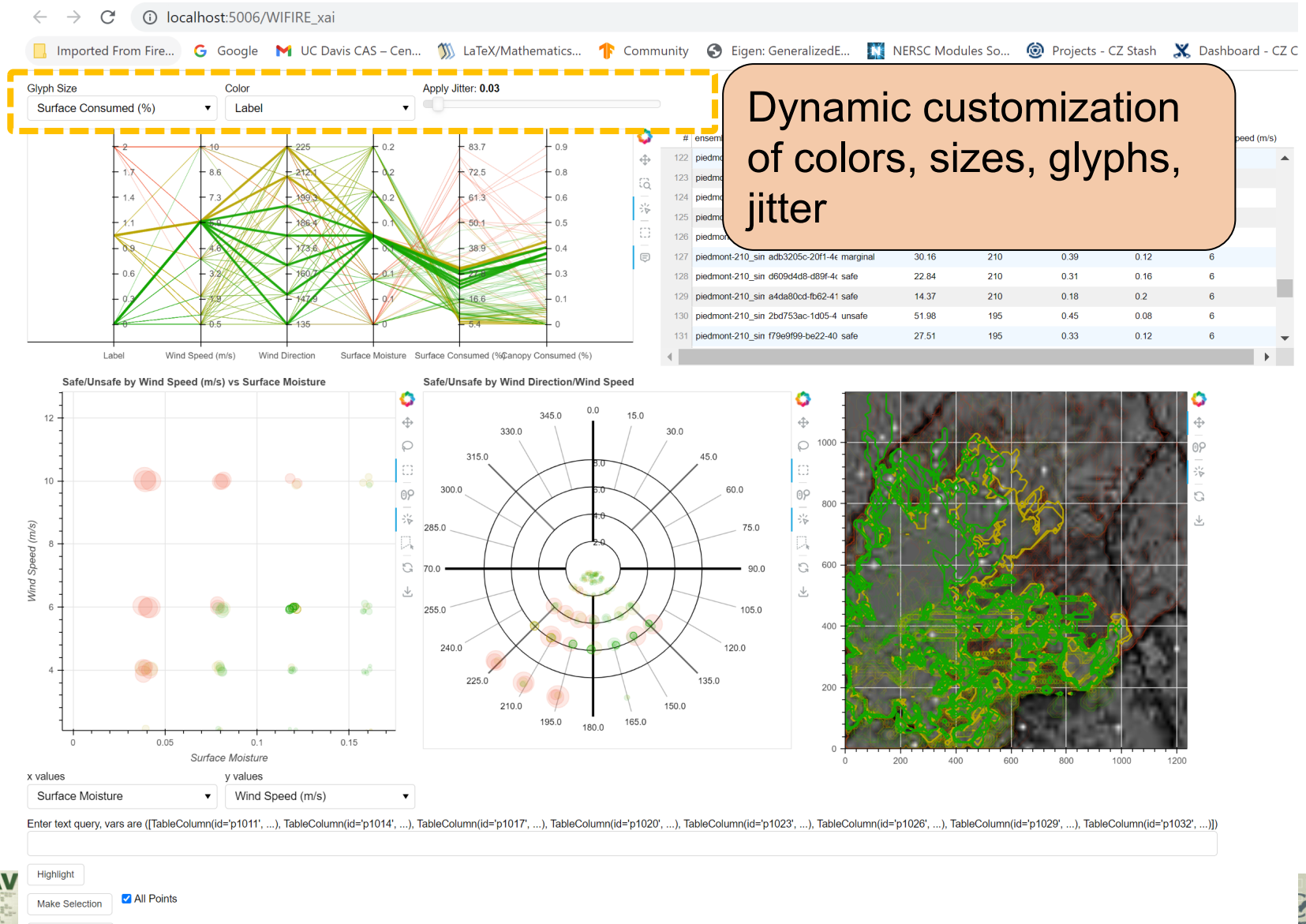
Deployed in browser for diverse, distributed users base



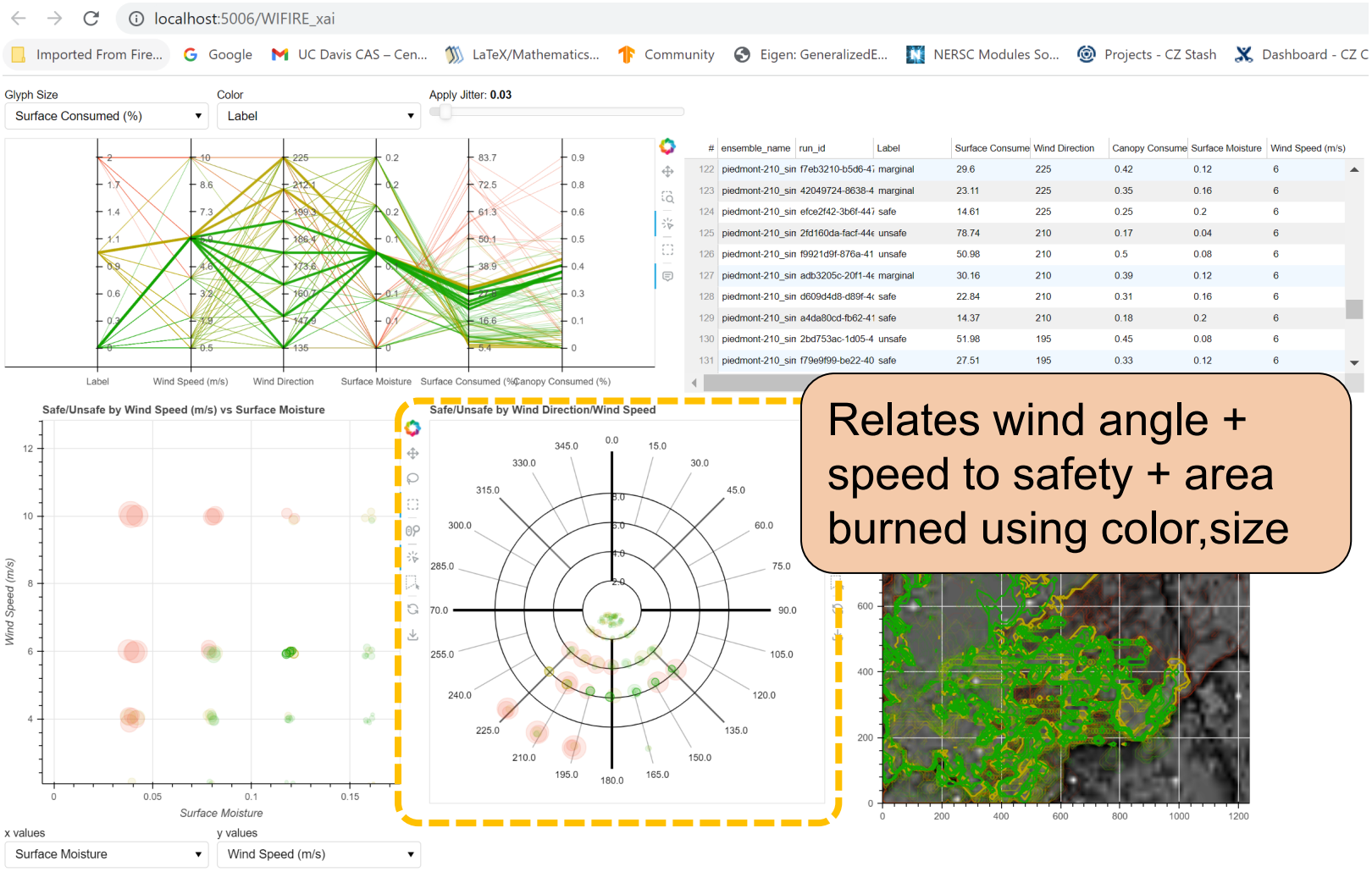
We support developers and burn bosses through an interactive interface



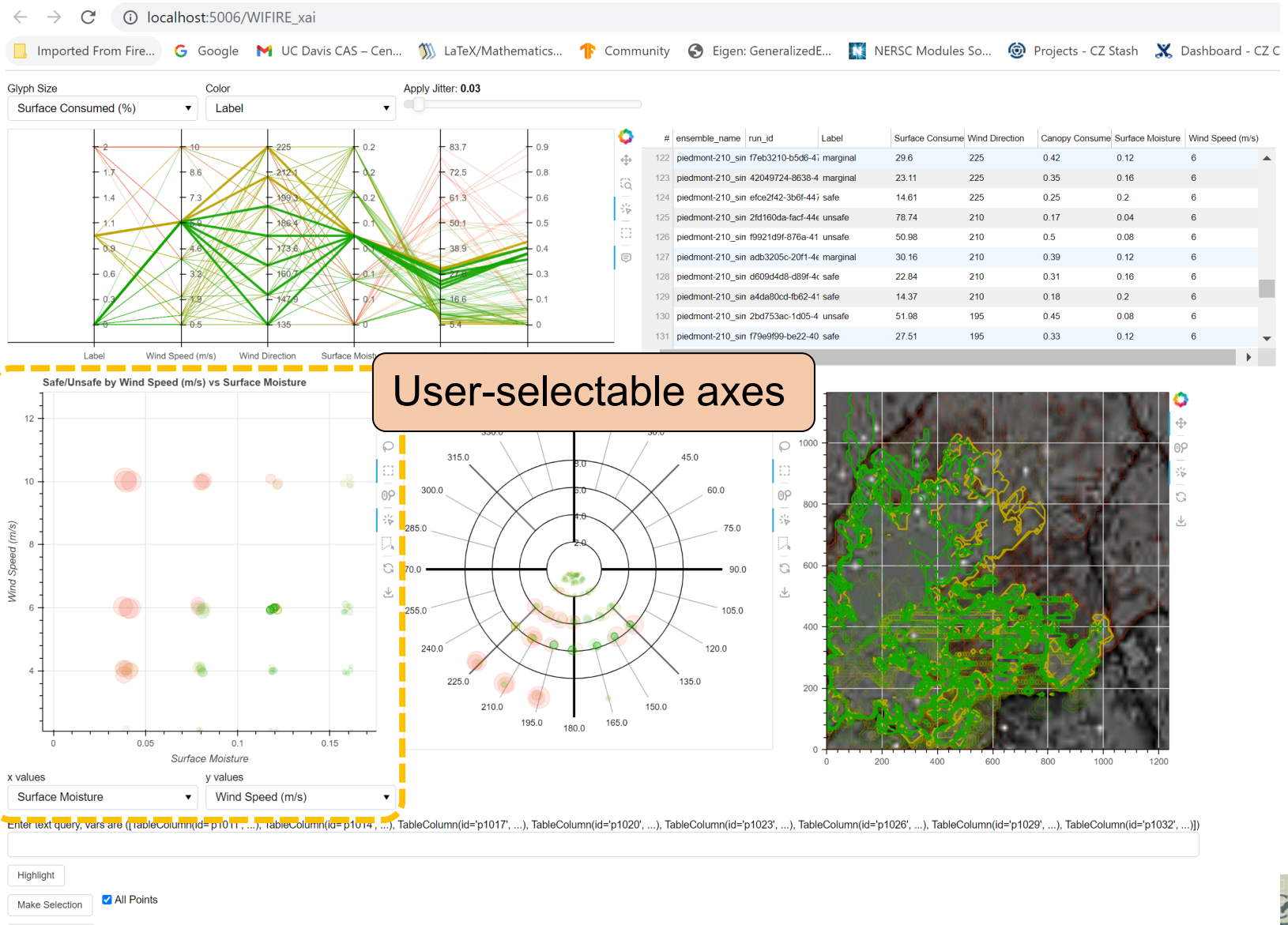
We support developers and burn bosses through an interactive interface



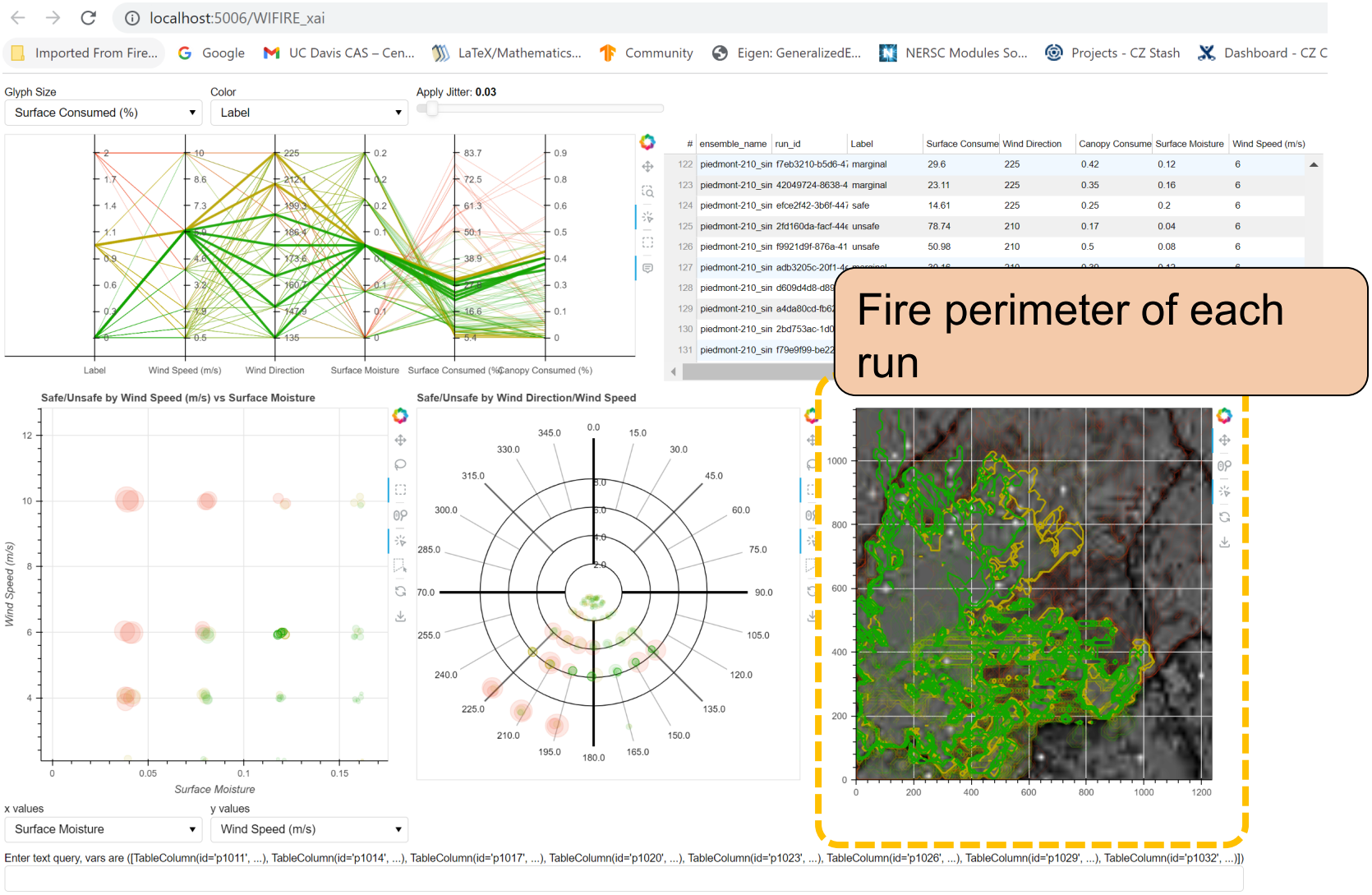
We support developers and burn bosses through an interactive interface



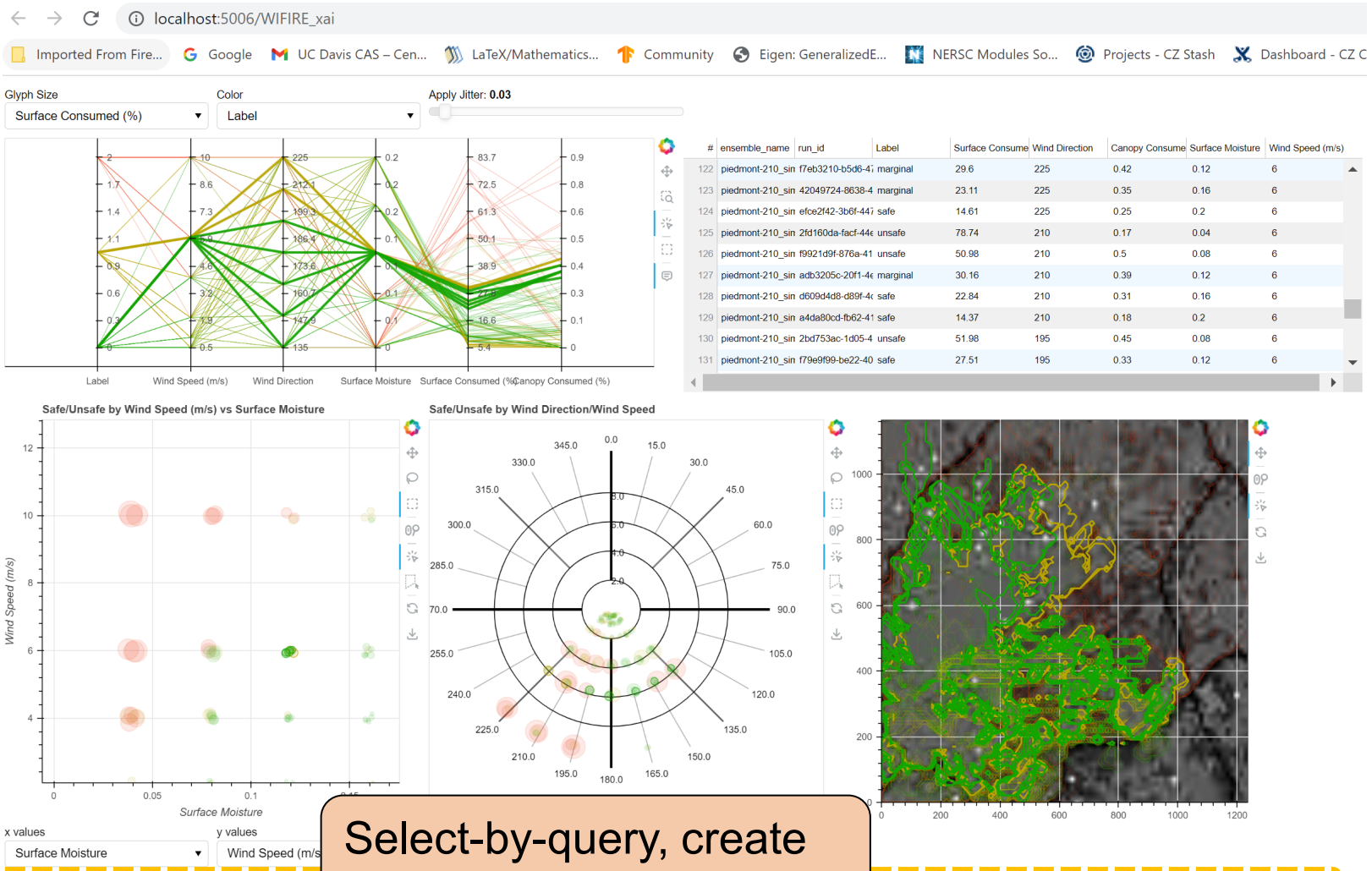
We support developers and burn bosses through an interactive interface



We support developers and burn bosses through an interactive interface



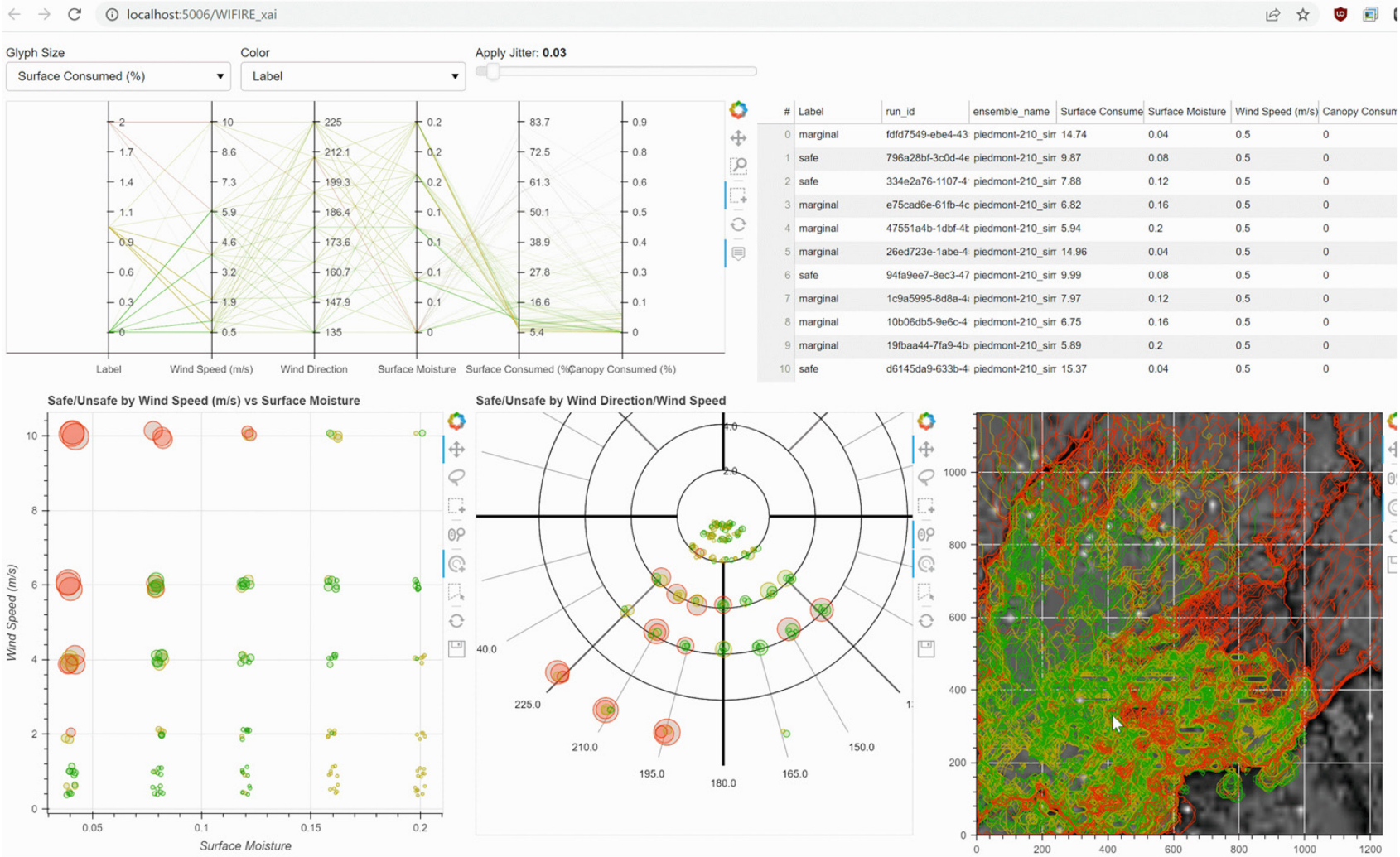
We support developers and burn bosses through an interactive interface



Select-by-query, create subsets of interest

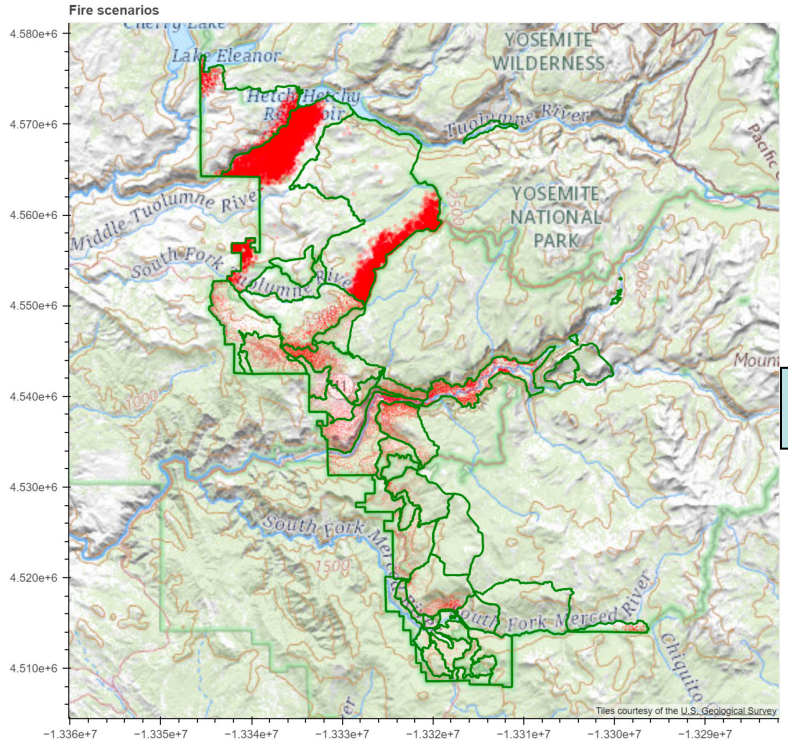
We support developers and burn bosses through an interactive interface

Key Design Strategy: Enable interaction, use natural visualizations, flexibility to explore



Fire simulations provide landscape-level fire risks

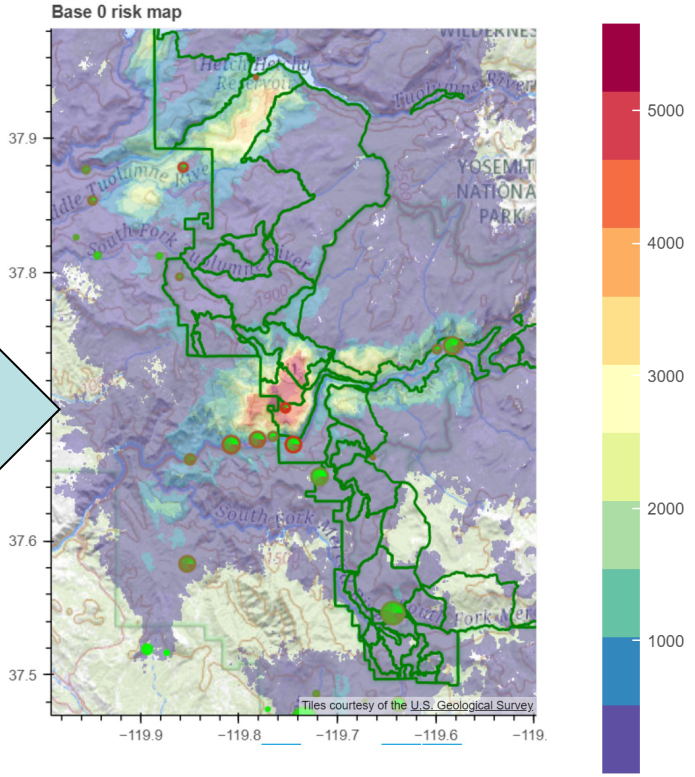
700K Generated Fire Scenarios (lightning, human causes, etc.)



Each dot is a fire, area is simulated fire area

Fire rasters aggregated

Risk map



How can we help fire managers pick which units to burn?

Separate landscape-level optimization dashboard

What units should be burned each year to minimize overall fire risk?

- **Given**: candidates for prescribed burn (based on topology, breaks, barriers) and associated operational costs/resource needs

- **Select** a set of burn areas **within a burn-management budget** such that landscape-level effect results in

- **Maximizing reduction** in potential **wildfire** damages such as
  - **area** burned
  - cost/number of **structures** damaged
  - **habitat** loss (critical species)

  - **CO2** emissions

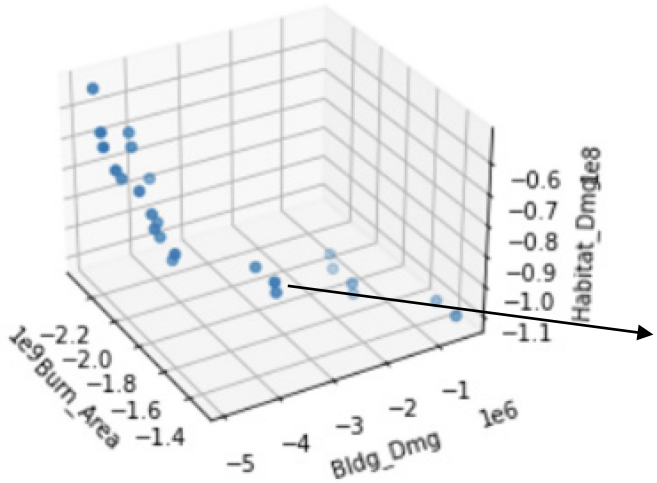
etc.

Genetic algorithm for multi-objective optimization yields Pareto-frontier

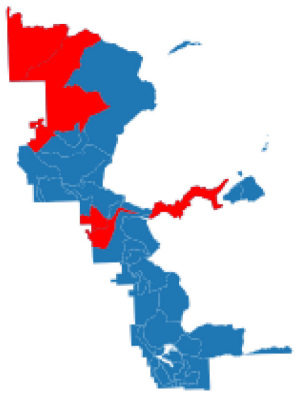
- Each solution is set of 5 burn units

- Each solution is not “dominated” by any other solution, e.g. not “worse” in every dimension

A set of solutions on Pareto Front



Specific Plan (red)



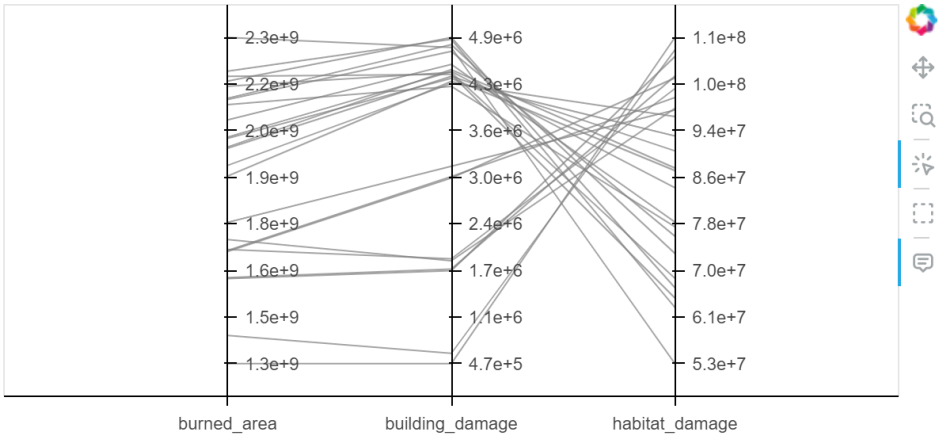
Making sense of landscape optimization

Landscape Optimization: Burn Units and Pareto Frontier

Add/Move constraints on parallel coordinates plot (draw box, drag) to constrain probability computation. Double-click box to remove, or double-click background to clear all constraints

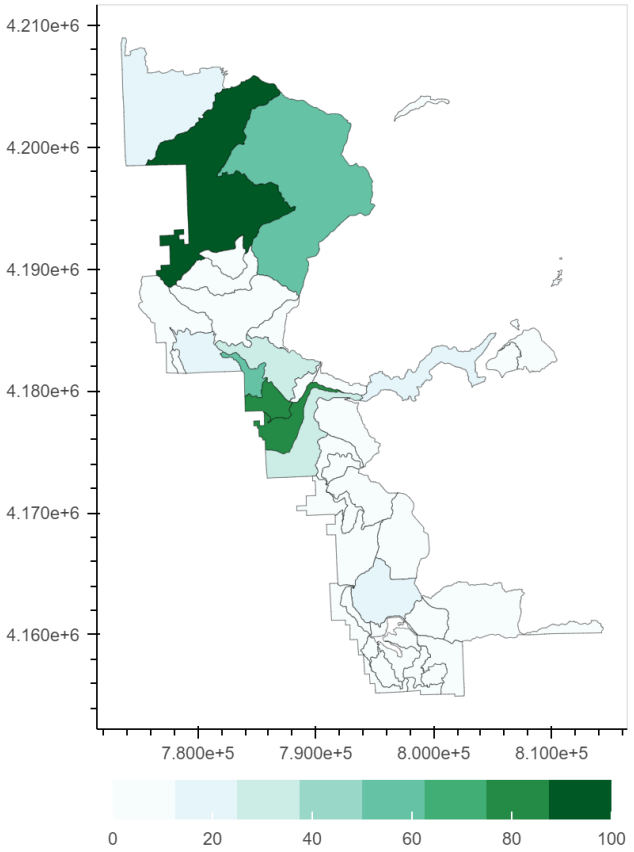
Click-select on map (hold SHIFT for multiple) to restrict view of solutions to those that include selected regions. Click background to clear region constriants.

Pareto Frontier (each line is one solution)



#	burned_area	building_damage	habitat_damage	ids
0	1989772154	4369425	90874449	13,14,21,27,28
1	2021041248	4343495	87386636	13,14,27,28,29
2	2147710446	4801440	72692578	13,23,24,27,28
3	2232898412	4872835	63119445	13,24,27,28,29
4	1902170952	4300050	93533147	8,13,23,27,28
5	1667302455	3007825	100344179	13,14,20,23,28
6	1664282068	2987420	103832024	8,13,14,23,28
7	2143563352	4709795	75735030	13,14,24,27,28
8	1993919248	4461070	87831997	13,21,23,27,28
9	1398802725	606205	108795771	8,13,14,21,23

The percent of selected solutions each burn unit appea



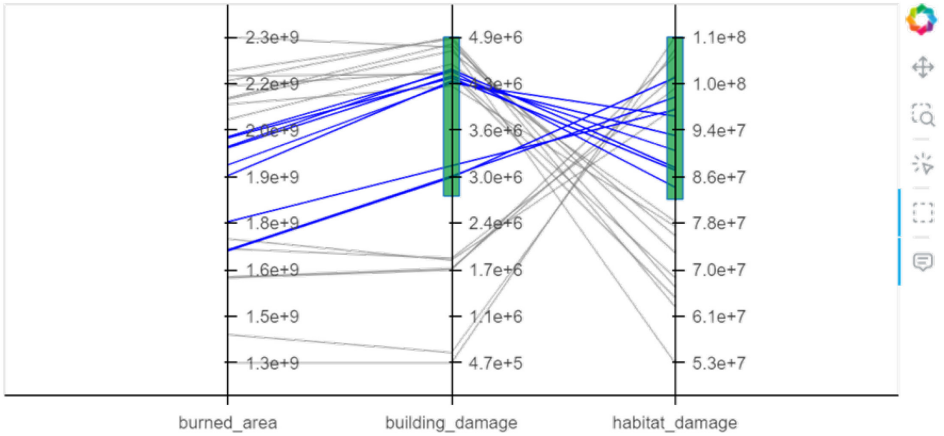
Making sense of landscape optimization

Landscape Optimization: Burn Units and Pareto Frontier

Add/Move constraints on parallel coordinates plot (draw box, drag) to constrain probability computation. Double-click box to remove, or double-click background to clear all constraints

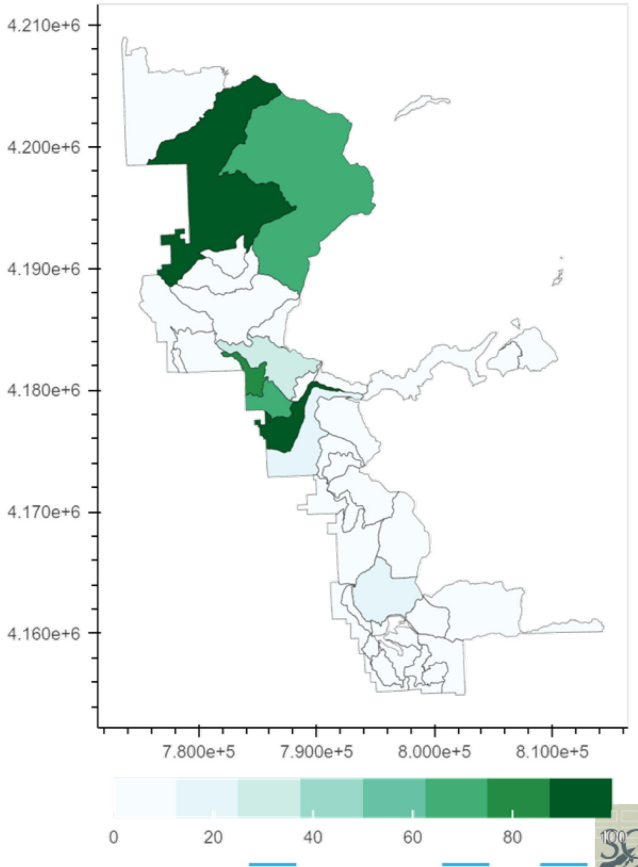
Click-select on map (hold SHIFT for multiple) to restrict view of solutions to those that include selected regions. Click background to clear region constriants.

Pareto Frontier (each line is one solution)



#	burned_area	building_damage	habitat_damage	ids
0	1989772154	4369425	90874449	13,14,21,27,28
1	2021041248	4343495	87386636	13,14,27,28,29
2	2147710446	4801440	72692578	13,23,24,27,28
3	2232898412	4872835	63119445	13,24,27,28,29
4	1902170952	4300050	93533147	8,13,23,27,28
5	1667302455	3007825	100344179	13,14,20,23,28
6	1664282068	2987420	103832024	8,13,14,23,28
7	2143563352	4709795	75735030	13,14,24,27,28
8	1993919248	4461070	87831997	13,21,23,27,28
9	1398802725	606205	108795771	8,13,14,21,23

The percent of selected solutions each burn unit appea



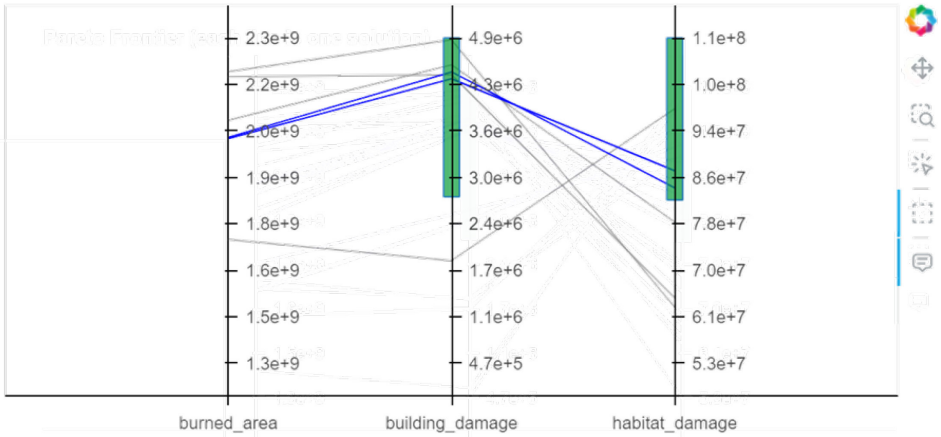
Making sense of landscape optimization

Landscape Optimization: Burn Units and Pareto Frontier

Add/Move constraints on parallel coordinates plot (draw box, drag) to constrain probability computation. Double-click box to remove, or double-click background to clear all constraints

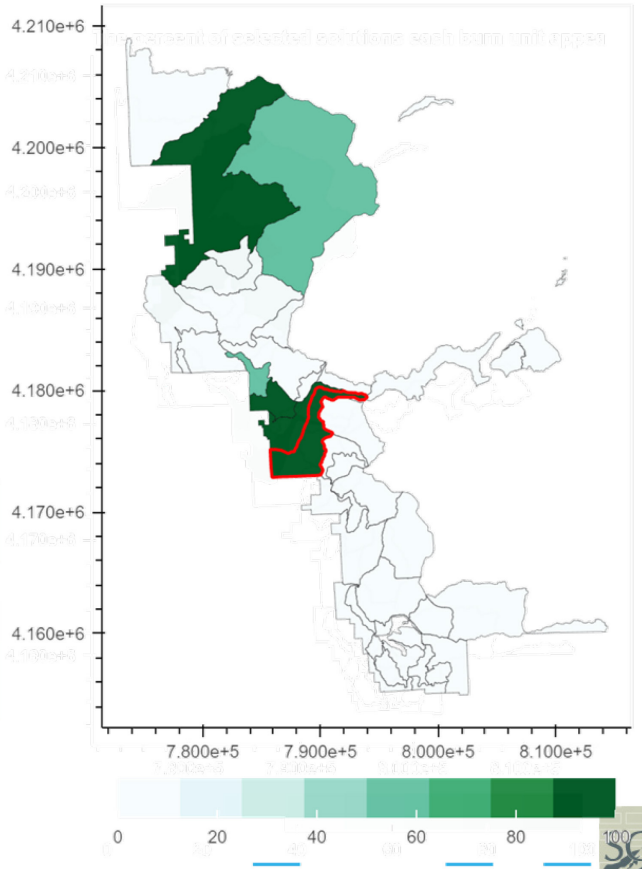
Click-select on map (hold SHIFT for multiple) to restrict view of solutions to those that include selected regions. Click background to clear region constriants.

Pareto Frontier (each line is one solution)



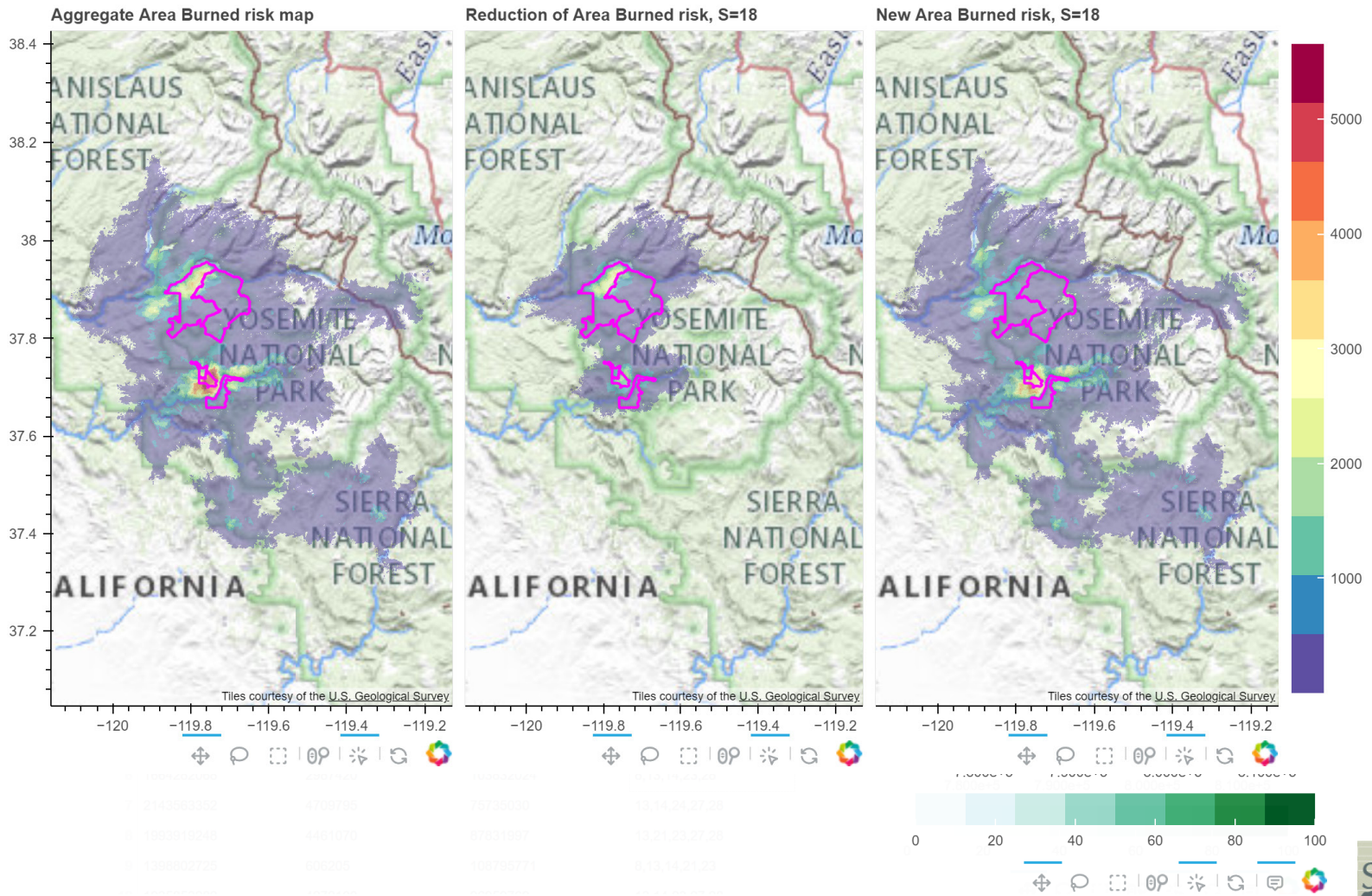
#	burned_area	building_damage	habitat_damage	ids
1	2021041248	4343495	87386636	13,14,27,28,29
3	2232898412	4872835	63119445	13,24,27,28,29
11	2025188342	4435140	84344184	13,23,27,28,29
14	2079107214	4532465	78258864	13,21,27,28,29
15	2217046573	4396800	64790023	12,13,27,28,29
18	1701643033	1864955	98435703	13,14,23,27,29

The percent of selected solutions each burn unit appea



Making sense of landscape optimization

Risk reduction for a selected solution



Landscape Optimization for WIFI

localhost:5006/mockup_vis

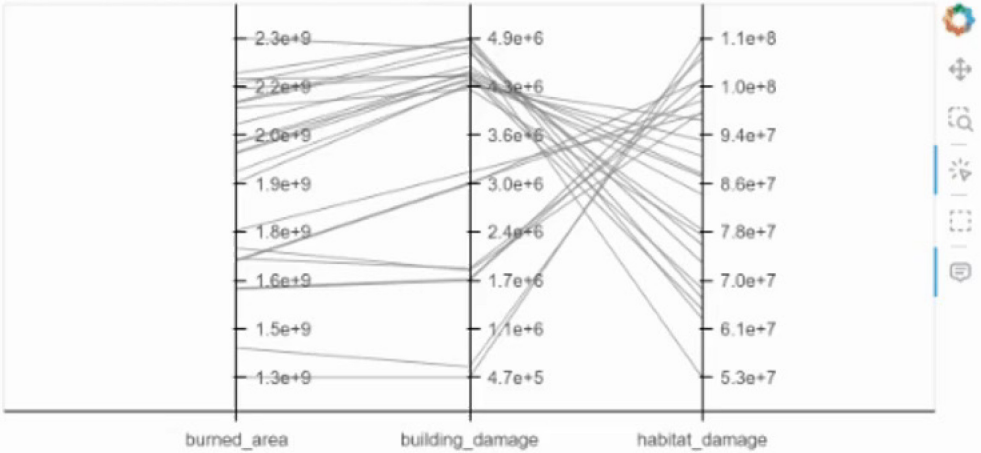
Update

Landscape Optimization: Burn Units and Pareto Frontier

Add/Move constraints on parallel coordinates plot (draw box, drag) to constrain probability computation. Double-click box to remove, or double-click background to clear all constraints

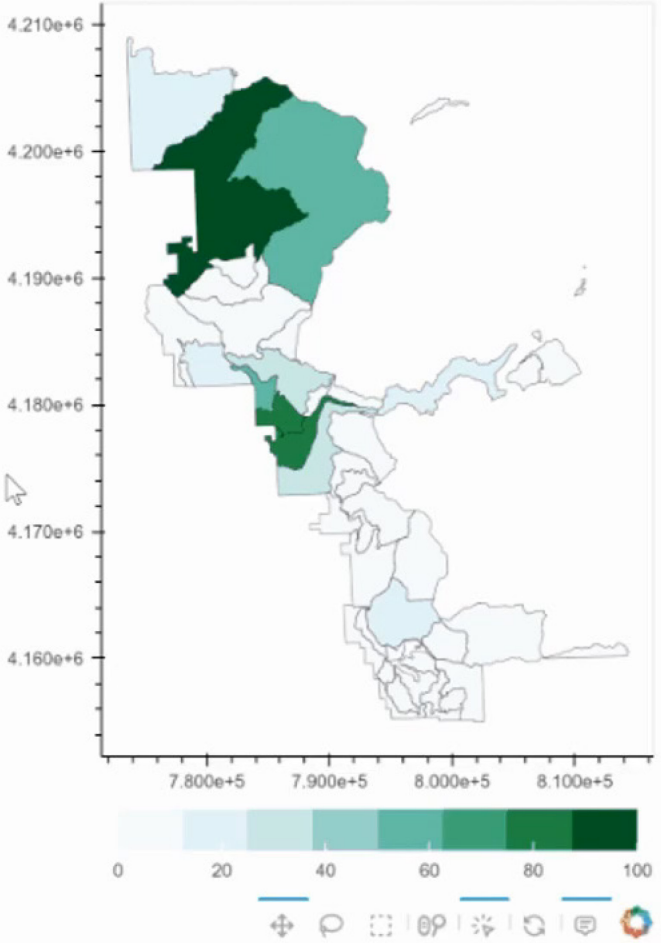
Click-select on map (hold SHIFT for multiple) to restrict view of solutions to those that include selected regions. Click background to clear region constriants.

Pareto Frontier (each line is one solution)



#	burned_area	building_damage	habitat_damage	ids
0	1989772154	4369425	90874449	13,14,21,27,28
1	2021041248	4343495	87386636	13,14,27,28,29
2	2147710446	4801440	72692578	13,23,24,27,28
3	2232898412	4872835	63119445	13,24,27,28,29
4	1902170952	4300050	93533147	8,13,23,27,28
5	1667302455	3007825	100344179	13,14,20,23,28
6	1664282068	2987420	103832024	8,13,14,23,28
7	2143563352	4709795	75735030	13,14,24,27,28
8	1993919248	4461070	87831997	13,21,23,27,28
9	1398802725	606205	108795771	8,13,14,21,23
10	1935853282	4272100	96959769	13,14,23,27,28

The percent of selected solutions each burn unit appea



It is all about the people, their teamwork, and their passion!!

CEDMAV


Xuan Huang

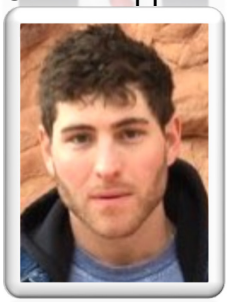

Osei Kobe

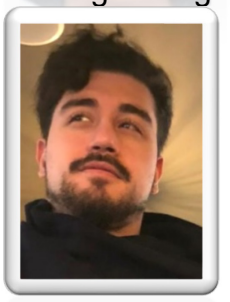

Owen Koppe


Duong Hoang


Aashish Panta


Aniketh Venkat


Sam Leventhal


Alper Sahistan


Zhimin Li


Nate Morrical

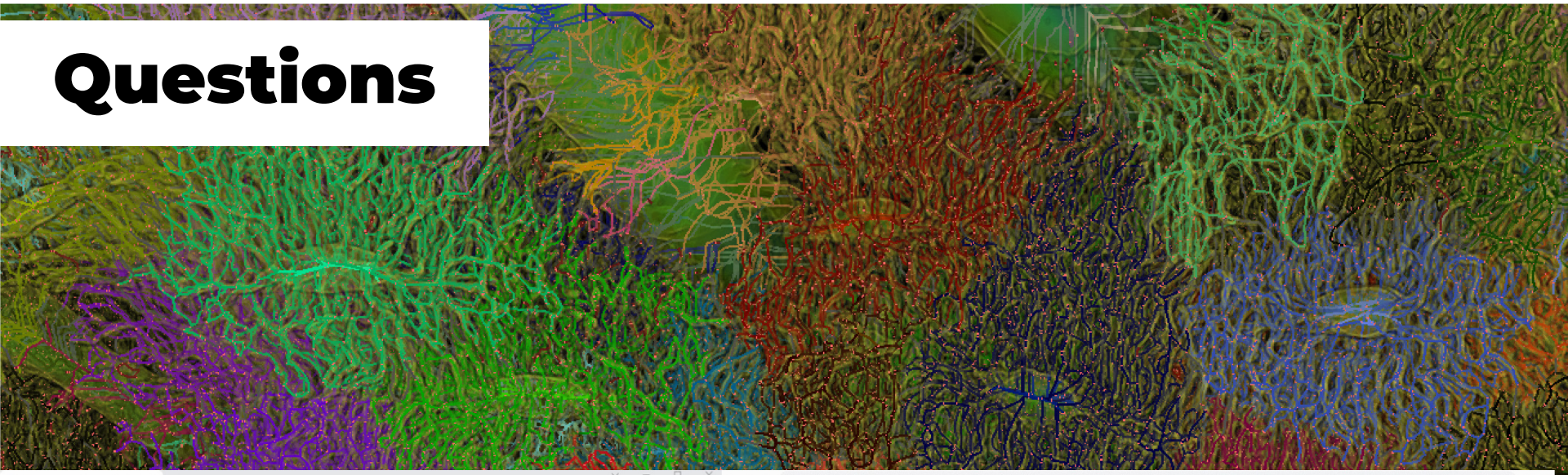

Valerio Pascucci

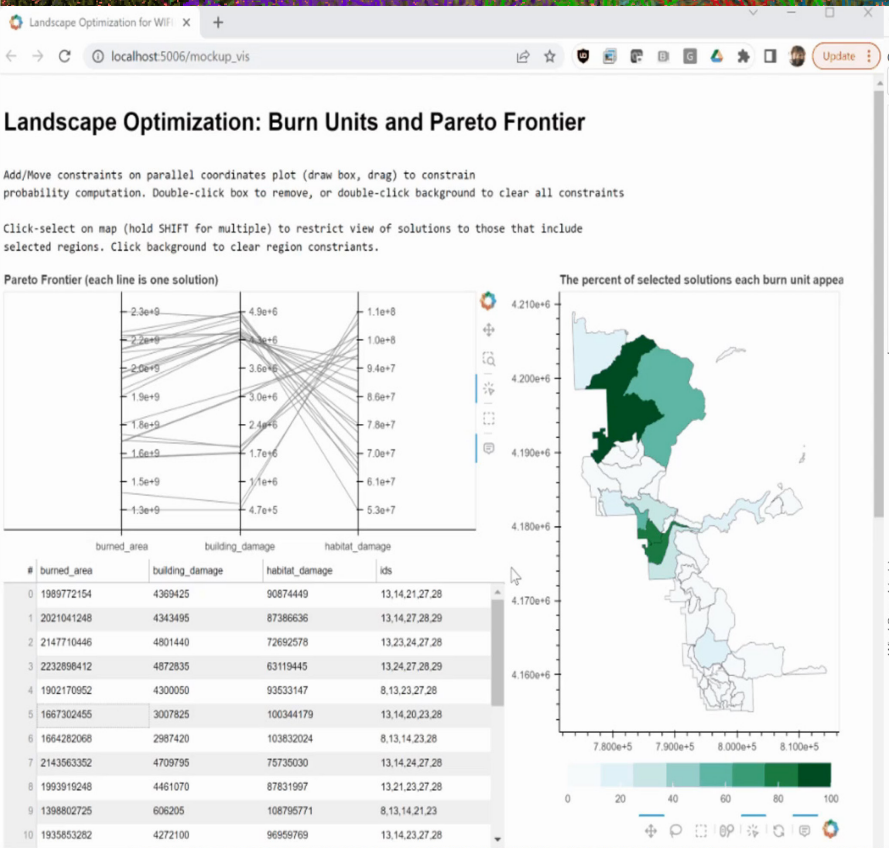
Recipients of:

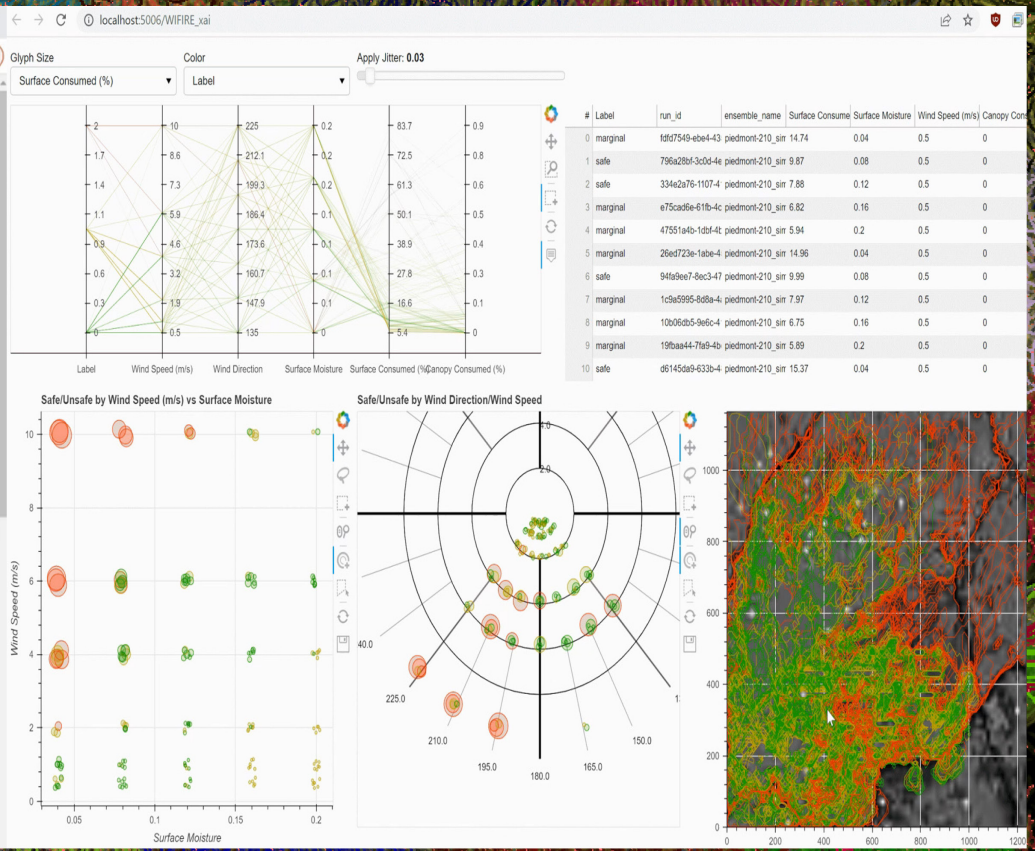
- Distinguished Research Award
- IEEE Technical achievement award
- IEEE Visualization Academy
- 2 Test of time awards
- 11 Best Paper awards
- Distinguished Mentor Award
- DOE Excellence in Technology Transfer

Looking for new students!
contact Pascucci@sci.utah.edu

Questions




Landscape Optimization for WIFI
localhost:5006/mockup_vis
Update
Landscape Optimization: Burn Units and Pareto Frontier
Add/Move constraints on parallel coordinates plot (draw box, drag) to constrain
probability computation. Double-click box to remove, or double-click background to clear all constraints
Click-select on map (hold SHIFT for multiple) to restrict view of solutions to those that include
selected regions. Click background to clear region constriants.
Pareto Frontier (each line is one solution)
burned_area
building_damage
habitat_damage
The percent of selected solutions each burn unit appea
burned_area building_damage habitat_damage ids
0 1989772154 4369425 90874449 13,14,21,27,28
1 2021041248 4343495 87386636 13,14,27,28,29
2 2147710446 4801440 72692578 13,23,24,27,28
3 2232898412 4872835 63119445 13,24,27,28,29
4 1902170952 4300050 93533147 8,13,23,27,28
5 1667302455 3007825 100344179 13,14,20,23,28
6 1664282068 2987420 103832024 8,13,14,23,28
7 2143563352 4709795 75735030 13,14,24,27,28
8 1993919248 4461070 87831997 13,21,23,27,28
9 1398802725 606205 108795771 8,13,14,21,23
10 1935853282 4272100 96959769 13,14,23,27,28


localhost:5006/WIFIRE_xai
Glyph Size
Surface Consumed (%)
Color
Label
Apply Jitter: 0.03
Label
Wind Speed (m/s)
Wind Direction
Surface Moisture
Surface Consumed (%)
Canopy Consumed (%)
Label run_id ensemble_name Surface Consume Surface Moisture Wind Speed (m/s) Canopy Cons
0 marginal fdfd7549-ebe4-43 piedmont-210_sir 14.74 0.04 0.5 0
1 safe 796a28bf-3c0d-4e piedmont-210_sir 9.87 0.08 0.5 0
2 safe 334e2a76-1107-4 piedmont-210_sir 7.88 0.12 0.5 0
3 marginal e75cad6e-61fb-4c piedmont-210_sir 6.82 0.16 0.5 0
4 marginal 47551a4b-1dbf-4t piedmont-210_sir 5.94 0.2 0.5 0
5 marginal 26ed723e-1abe-4 piedmont-210_sir 14.96 0.04 0.5 0
6 safe 94fa9ee7-8ec3-47 piedmont-210_sir 9.99 0.08 0.5 0
7 marginal 1c9a5995-8d8a-4 piedmont-210_sir 7.97 0.12 0.5 0
8 marginal 10b06db5-9e6c-4 piedmont-210_sir 6.75 0.16 0.5 0
9 marginal 19fbaa44-7fa9-4b piedmont-210_sir 5.89 0.2 0.5 0
10 safe d6145da9-633b-4 piedmont-210_sir 15.37 0.04 0.5 0
Safe/Unsafe by Wind Speed (m/s) vs Surface Moisture
Wind Speed (m/s)
Surface Moisture
Safe/Unsafe by Wind Direction/Wind Speed