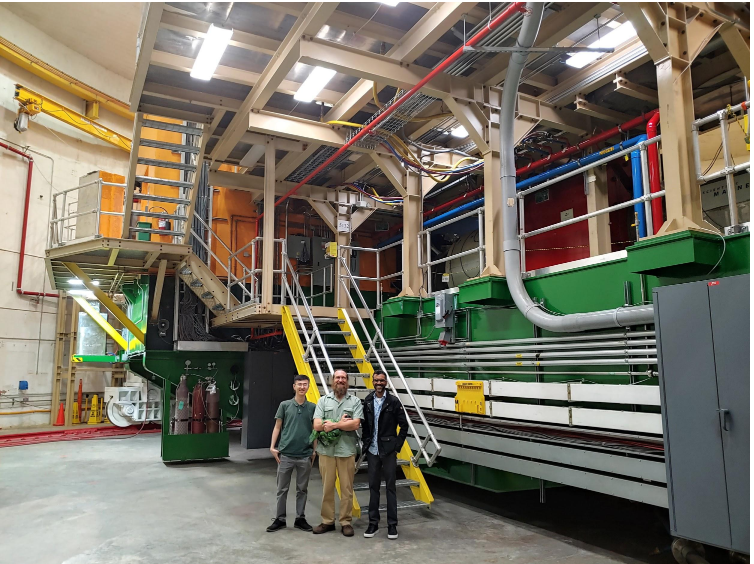
Next-generation Visual Analysis Workspace for Multidimensional Nuclear Femtography Data

CENTER for NUCLEAR FEMTOGRAPHY

EIC² EIC Center at Jefferson Lab



PI Name and Institution: Nicholas F. Polys, Virginia Tech

Markus Diefenthaler, Jefferson Lab

Srijith Rajamohan, Virginia Tech

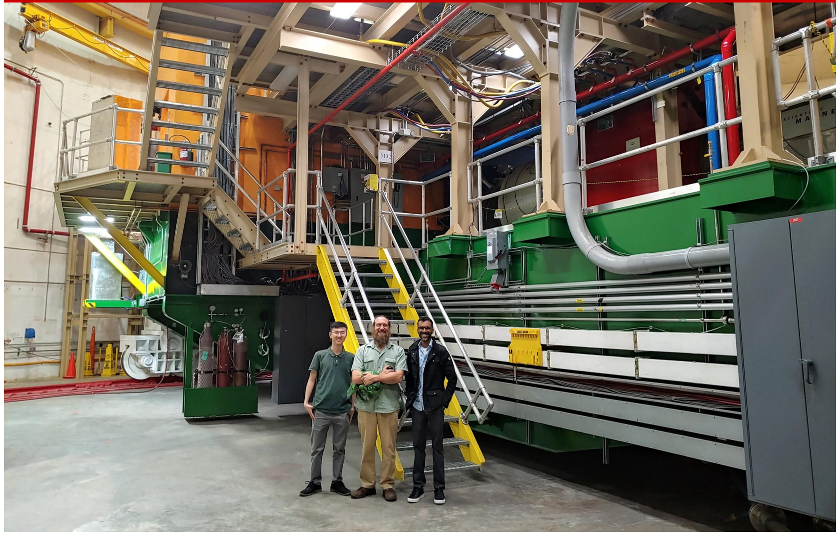
Dmitry Romanov, Jefferson Lab

List of Collaborators: Mai Dahshan, Virginia Tech

JooYoung Whang, Virginia Tech

Virginia Tech

Jefferson Lab



***CNF_19_15** : Our exploratory study demonstrates the synergy between **computer science** and **nuclear physics** in the visualization of semi-inclusive deep-inelastic scattering processes and the study of TMD observables.*

VIRGINIA TECH | Advanced Research Computing

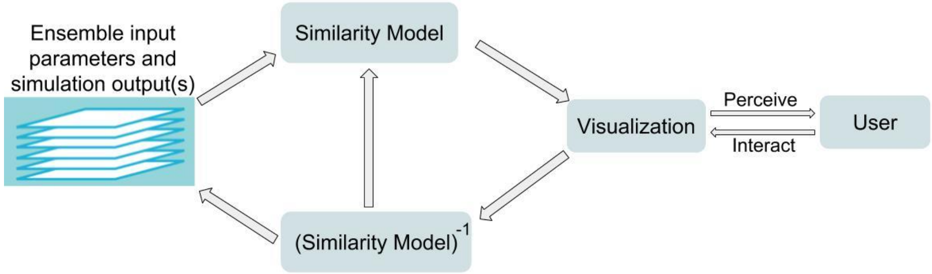
Jefferson Lab
Exploring the Nature of Matter

Next-generation Visual Analysis Workspace for Multidimensional Nuclear Femtography Data

Polys
Rajamohan
Diefenthaler
Romanov
Dashan
Whang

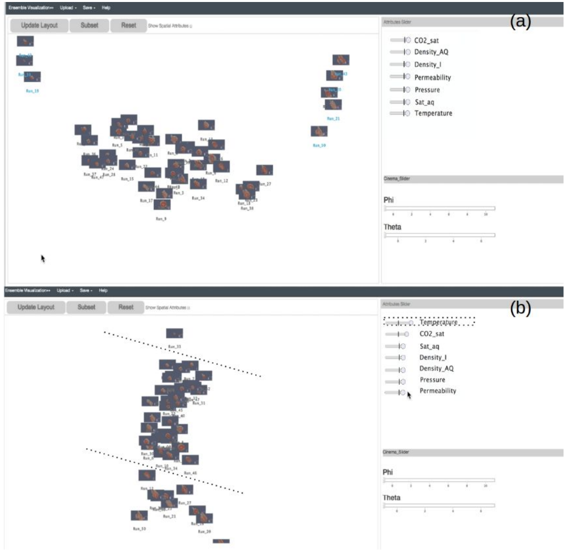
Project Goals:

- Using Semantic Interaction to understand particle kinematics



Approach:

- Map particle kinematics into Paraview, VR, Web3D
  - Demonstrate w HERMES data, extend to 12 GeV data
- Export X3D, Cinema visualization
- Instance GLEE for project data sets and user upload



VIRGINIA TECH | Advanced Research Computing

Goal:
to explore the application of Semantic Interaction methods to particle factorization

Will we learn anything new?

Approach:

1. Use the SIDIS dataset, JLAB 12GeV data
2. Create Particle Portraits (Hermes as proxy)
3. Load into the GLEE workspace @ VT
4. Look for known correlations
5. Find new correlations through Semantic Interaction

Exploring the kinematic phasespace of pions and charged kaons in semi-inclusive deep-inelastic scattering (SIDIS).

Starting point:

- Visual representations in ParaView
- SIDIS data from HERMES for charged pions and kaons

Studied:

- Kinematic correlations (Bjorken-x, z, and Phperp)
- Differences in kinematic correlations between pi+ and K+

Ongoing study:

- Kinematic correlations for various TMD factorization criteria

Semantic Interaction

The user and the machine build a model together:

incrementally formalizing and emphasizing

relationships among data members

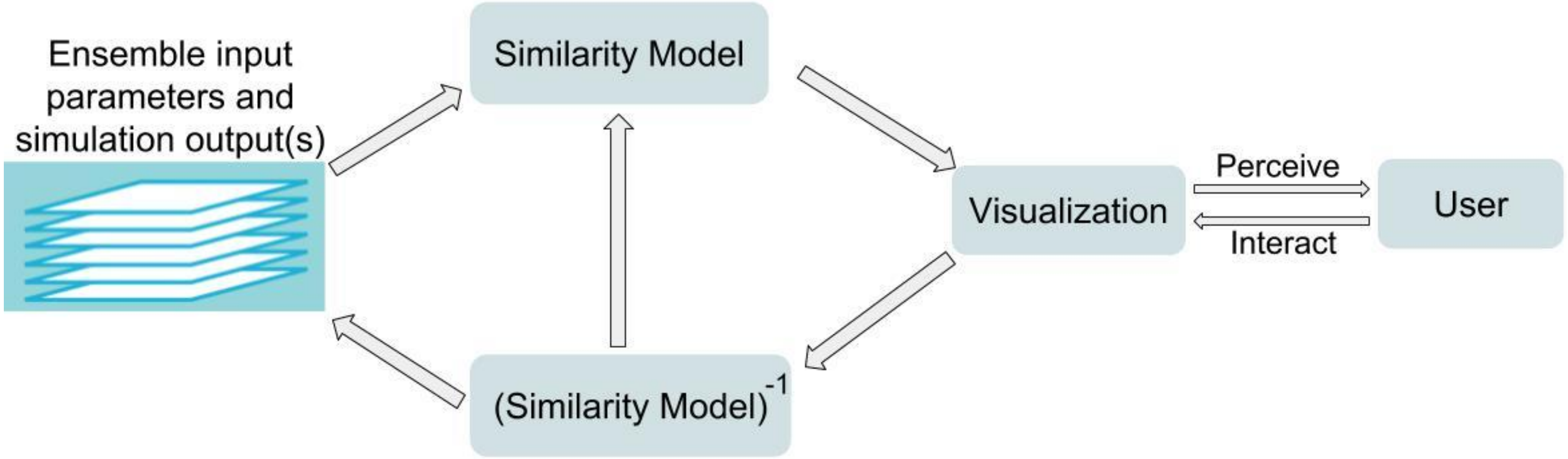
through the user's interactions.

The Graphically-Linked Ensemble Explorer (GLEE) implements this approach in a WWW framework.

Semantic Interaction

For Exploratory visualization =

under-specified problems + human intuition + computational rigor



Data Preparation

The alignment of information to connect disciplines:

Understanding how to adapt the Multi-Dimensional Scaling projection method of Semantic Interaction to this domain?

- How to visualize observations?
- What factors to include?
- How to weight them?, and
- What distance function should be used ?

Starting with HERMES data

- **Raw quantities**: 1n Columns for different observations (rows)
  - X
  - Z
  - Phperp...
- **Derived quantities**, which are distinguishing factors for factorization studies?
  - TMD factorization: Q^2 >> P_hperp^2 / z^2
  - qzp := P_hperp^2 / (Q^2 * z^2),
  - Match known 2D histograms per variable

Variable		
Run number		
Burst number		
Event number		
Polarization (beam)		
Polarization (target)		
Q^2		
x		
y		
Phi_s		
epsilon		
Charge		
Pi +/- weight		
Pi 0 weight		
K+/-		
Proton weight		
Z (Pi +/-)		
Z (Pi 0)		
Z (K +/-)		
Z (proton weight)		
P_hperp		
Phi_h		

Cutting up HERMES (for example)

Splitting criteria to find known correlations:

1. Originally, there are ~2mil rows in the HERMES data, each row representing the kinematics of a final-state hadron in SIDIS
2. filter these rows into *sets* for each particle type:

   Pi+, Pi-, K+, K-
3. For each particle type *set*, divide it into 3 *groups* by the qzp values:

   qzp < 0.5,   0.5 <= qzp <= 1.5,   qzp >1.5

Making Thumbnails

1. Choose to make ***n*** thumbnails for each *group*.
2. Let the number of rows in the smallest *group* of the three be ***k***. (Smallest in terms of number of rows)
3. Randomly select **floor(k / n)** rows in a *group* and use these to form a thumbnail. Repeat **n** times for each *group*. This results in **3n** thumbnails for each particle *set*. (no same row is selected twice in the entire process.)
4. The mean value and standard deviation are calculated and stored for each thumbnail data's individual components.

Paraview and Cinema

- Paraview.org 5.6 and custom Python scripts
  - Data loading
  - Data binning
  - Visualization mapping (i.e. to color, transparency, ...)

- Export Cinema image sets for use in GLEE workspace
  - https://cinemascience.github.io/
  - https://cinemaviewer.org/
  - GLEE server uses tabular data and Cinema image sets to create the Semantic Interaction Workspace (hosted at VT)

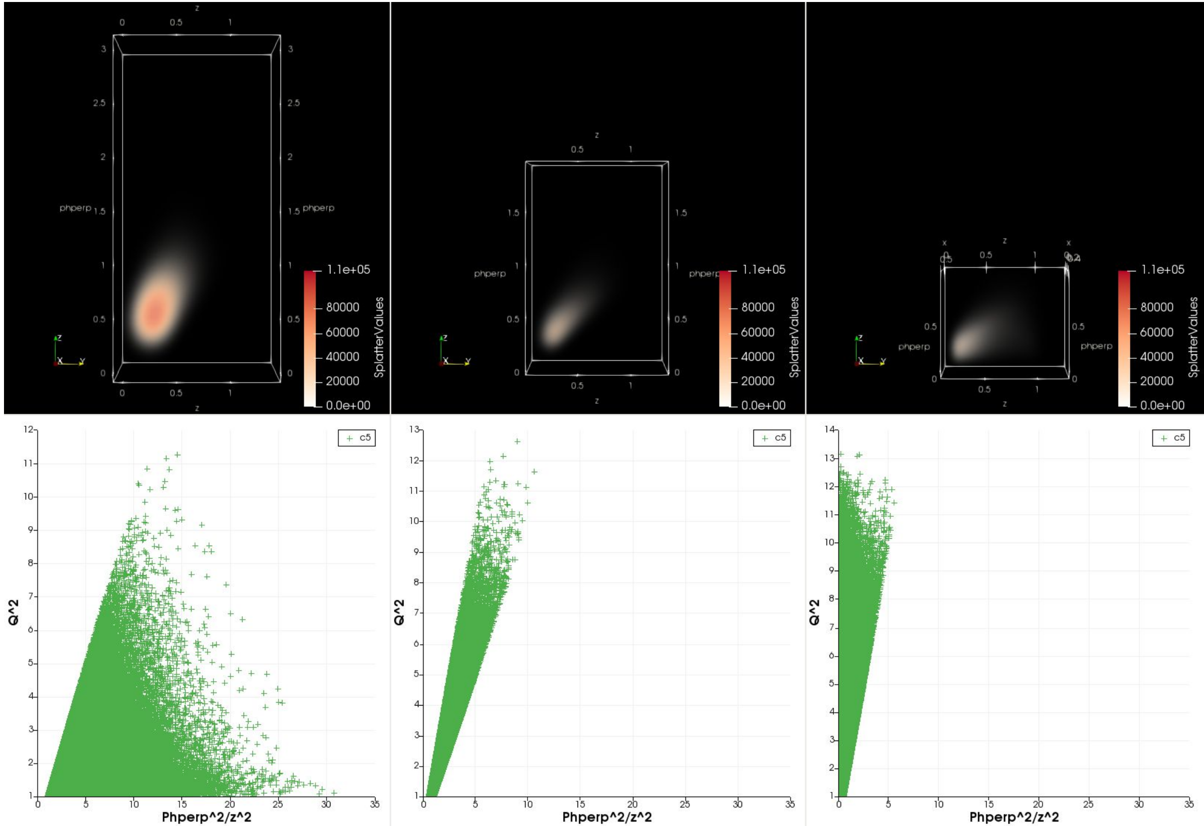

The Super High Momentum Spectrometer
Thomas Jefferson National Accelerator Facility
Before
After
Building and Installing Five New Superconducting Magnets
REMOVE ALL LADDERS BEFORE RUN

Ideas on the Way

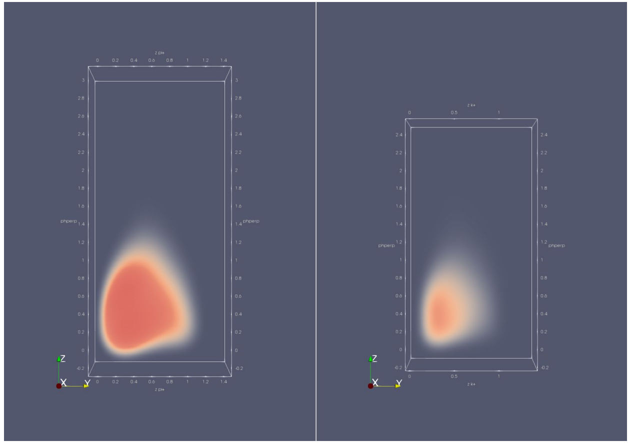
(Pi+)

- Volume rendering
- 3D surface plots
- 2D contours

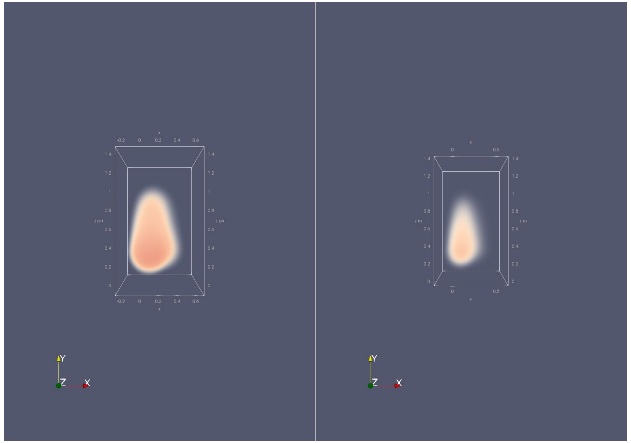
qzp = P_hperp^2 / (Q^2 * z^2)



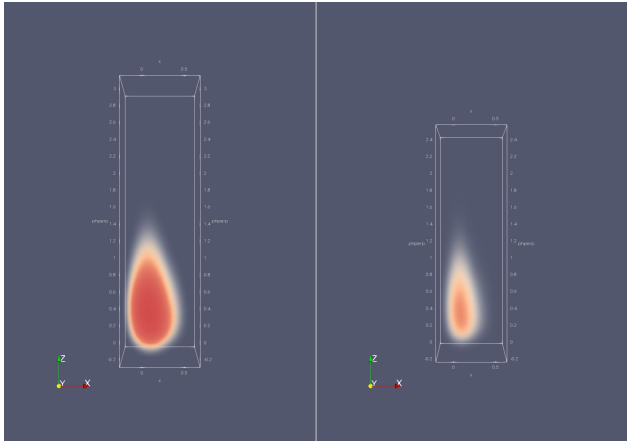
Volume rendering Histograms in 3 space



Z vs P_hperp



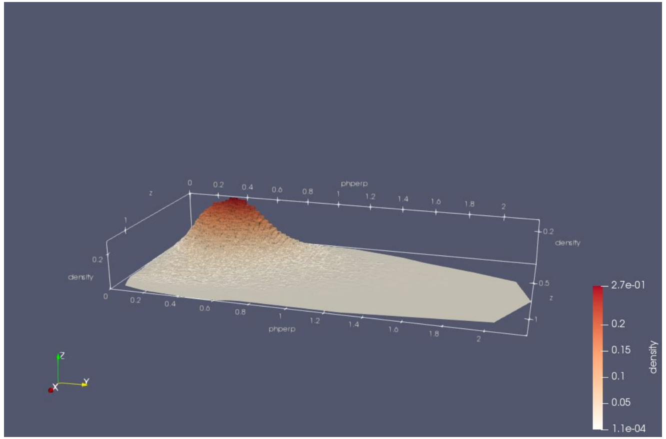
X vs z



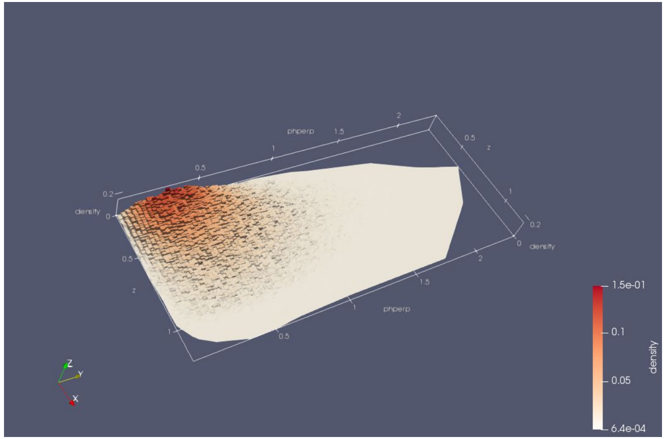
X vs P_hperp

Pi+ (left) vs K+ (right)

3-D surface histogram



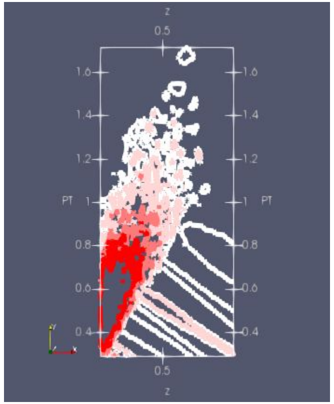
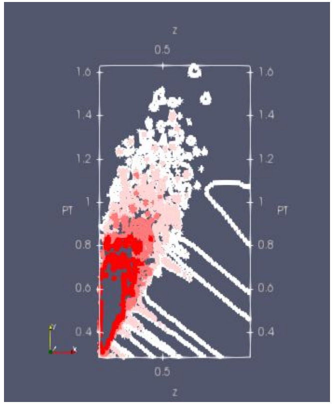
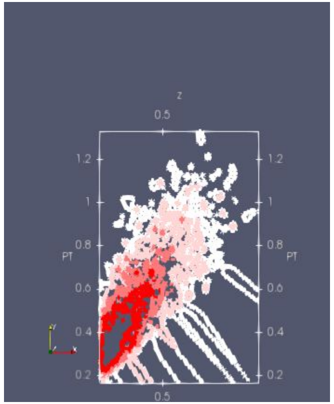
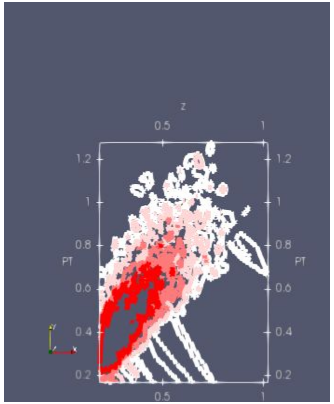
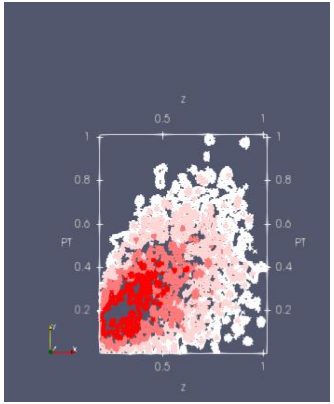
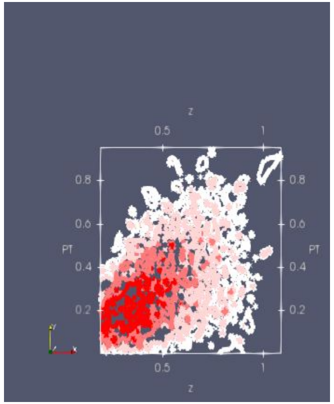
Pi+ density plot



K+ density plot

X-axis = Z
Y-axis = P_hperp

2-D Contour

2-D contour plot
of **K+** where
X-axis = z
Y-axis = P_hperp

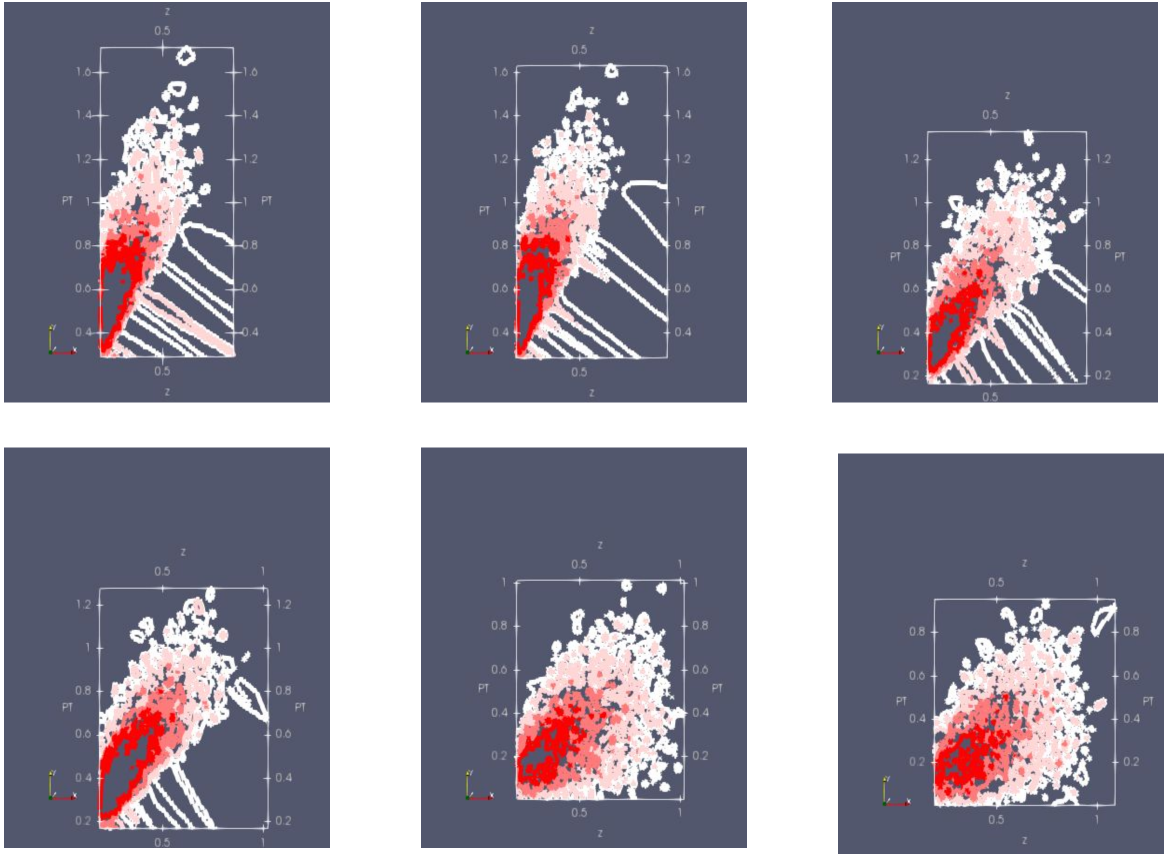
10799 pts in each thumbnail

Visual Forms: *Particle Portraits*

Approach: visualize known particle profiles as frequency distributions

1. Convert the row data into 2-D point data by letting the **x-axis** as the **Z value** of the particle, and the **y-axis** as the **P_hperp value** of the particle.
2. Using binning with 10000 bins, estimate the density of the points at each point's location.
3. Create a contour plot of the density map, and color code it from white to red.
4. Show the data axis and save the screenshot.

Sample thumbnails of K+



Semantic Interaction

The user and the machine build a model together:

incrementally formalizing and emphasizing

relationships among data members

through the user's interactions.

The Graphically-Linked Ensemble Explorer (GLEE) implements this approach in a WWW framework.

Similarity Models (MDS)

Weighted MDS - what gets weighted?

(attributes, components, parameters)?

Thus, we must investigate the distribution of values in across an ensemble and several cuts… and several distance functions

Semantic Interaction w/ GLEE for Hadronic Physics:

- When projecting from High D to Low D
- When projecting from Low D to High D

Terminology clarification

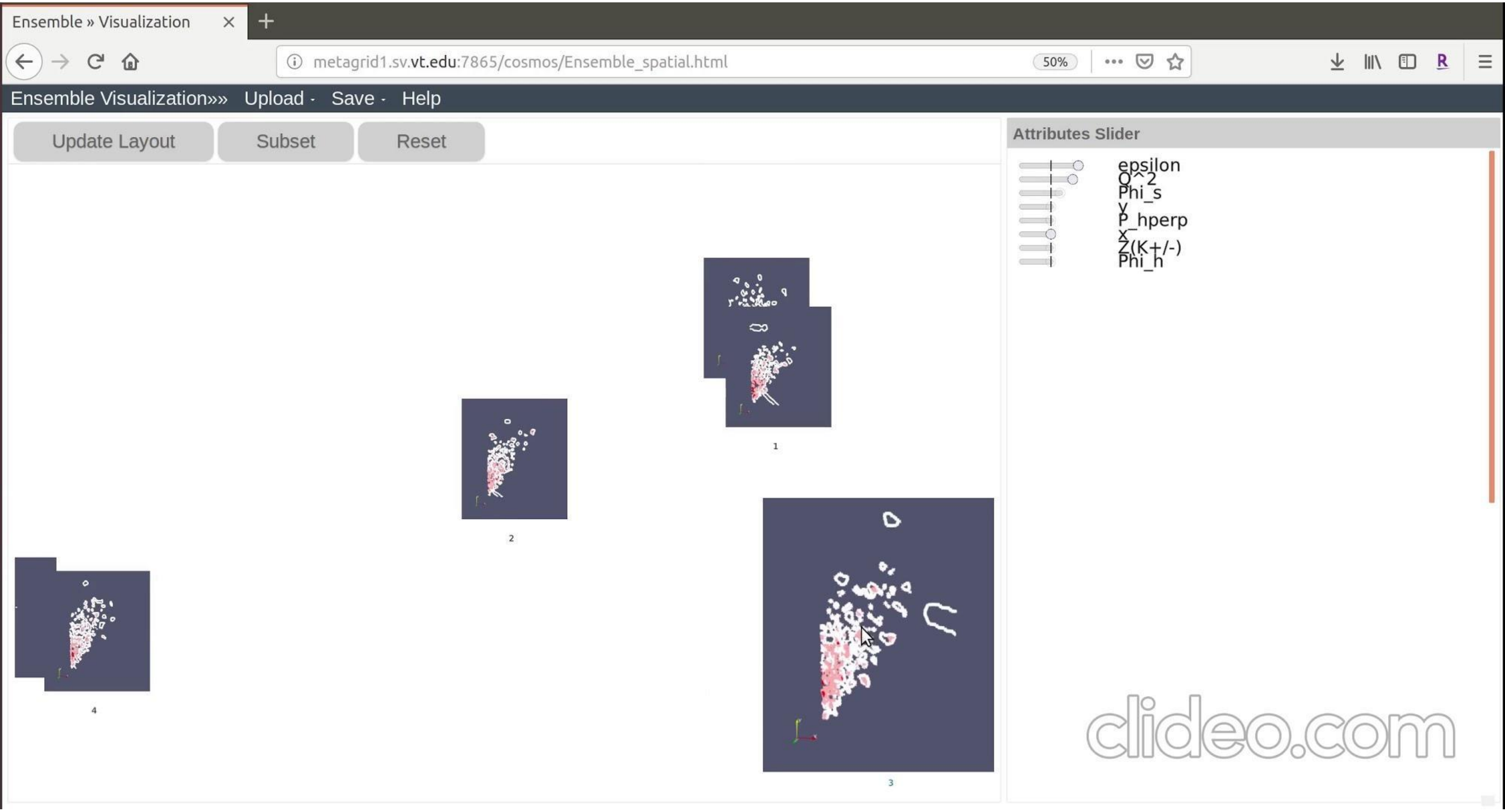
An ***Attribute*** can have multiple ***Components***, and each components have multiple ***Parameters*** that describe the distribution of that Component across the ensemble

	Attribute 1	**Attribute** *m*
Observation1	**Component** 1, Component 2, … Mean and standard deviation of each *Component* = **parameters** of that components' distributions	
Observation*n*		

Model and View

Weighted Multi-Dimensional Scaling (WMDS) can be used to up and down project parameters based on user interaction:

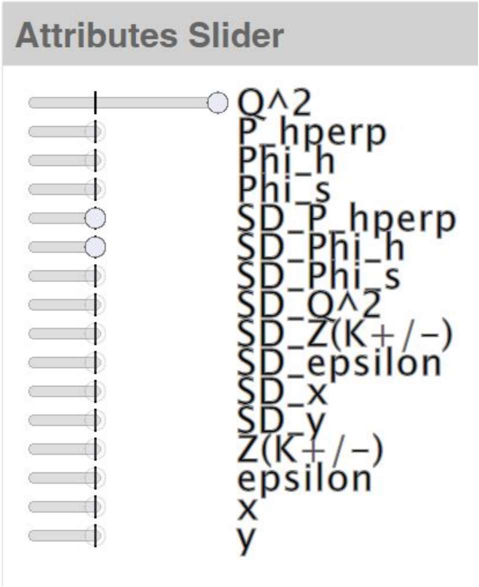
- Include components' Mean and StDev in the High-D parameter model
- Users see sliders for attributes in the GLEE PI window
- Examine appropriate distance function (e.g. Euclidean distance vs. Cosine distance)


Ensemble » Visualization
metagrid1.sv.vt.edu:7865/cosmos/Ensemble_spatial.html
50%
Ensemble Visualization»»
Upload
Save
Help
Update Layout
Subset
Reset
Attributes Slider
epsilon
Q^2
Phi_s
y
P_hperp
x
Z(K+/-)
Phi_h
1
2
3
4
clideo.com

Column Q^2 & GLEE

Q^2 seems to be quadratically increasing according to qzp

Q^2
1.559312
1.562237
2.023418
2.027454
2.906983
2.916709

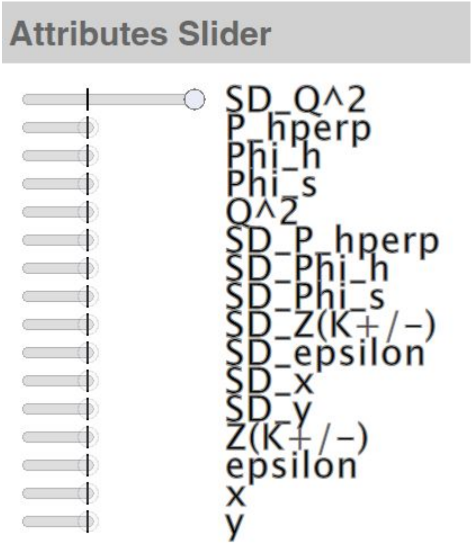


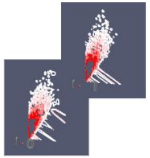
Increasing the Weight for Q^2 separates the thumbnails quadratically according to qzp

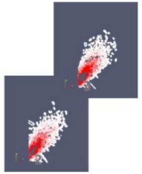
Column SD_Q^2 & GLEE

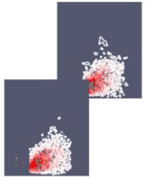
SD_Q^2 proportionally increases according to qzp

SD_Q^2
0.576975
0.595719
1.029479
1.029089
1.728903
1.729812



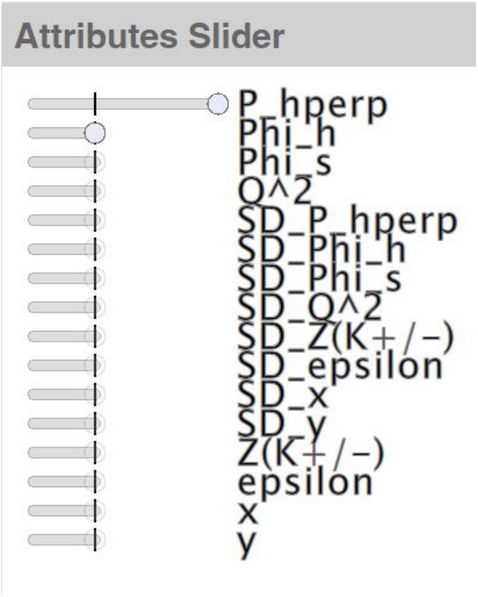






Increasing the Weight for SD_Q^2 Separates the thumbnails according to qzp

Column P_hperp & GLEE



P_hperp proportionally decreases according
To qzp values

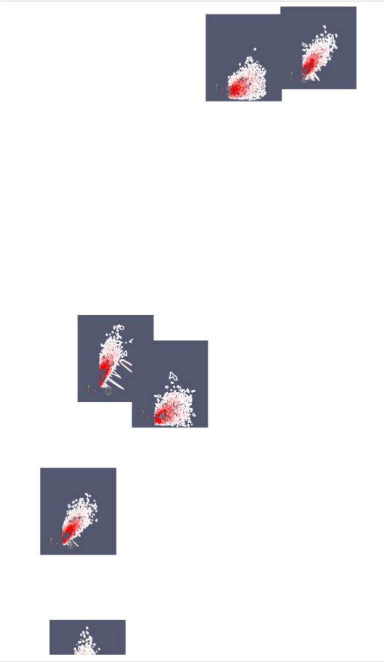
P_hperp
0.710593
0.709122
0.559037
0.554866
0.314239
0.31656

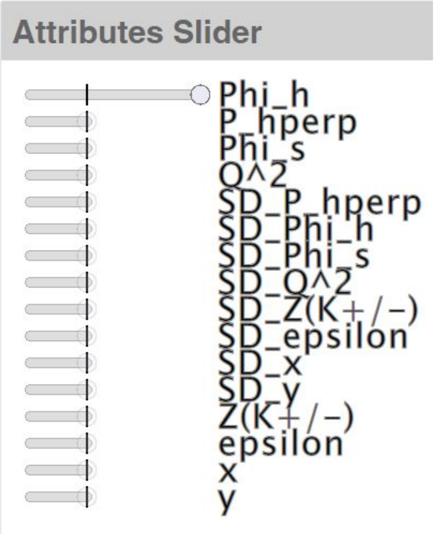
Increasing the Weight for P_hperp separates the thumbnails apart

Column Phi_h & GLEE

Phi_h varies throughout all qzp values

Phi_h
3.145927
3.168318
3.119855
3.15746
3.146924
3.122512





Increasing the Weight for Phi_h scatters the thumbnails

Takeaways

- Novel collaboration across disciplines shows promise; ‘Regularization’ is the biggest investment
- ***The Human IS the Loop***
- Semantic Interaction is a semi-supervised approach and we continue to explore and evaluate appropriate factors and distance functions

Next Steps

- Iterate portrait designs to project wrap:
  - **Pi+, Pi-, K+, K-**
- Iterate MDS models to project wrap
- Study early data from JLAB 12 GeV experiments (e.g., Hall C factorization tests)
  - Study more kinematic correlations and regimes (JLAB-THY-19-2920)
- Evaluate visual and statistical models with domain experts
- Publish
  - Paraview State Files for collaborator use
  - GLEE webservice for user-uploaded data
- Next stage CNF Proposal, EIC proposal