Center for Nuclear Femtography CNF19-15: High-Dimensional Visual Analytics of Particle Kinematics
Project Report

Nicholas Polys, Virginia Tech
Markus Diefenthaler, Jefferson Lab
Srijith Rajamohan, Virginia Tech
JooYoung Whang, Virginia Tech
Dmitry Romanov, Jefferson Lab
Mai Dahshan, Virginia Tech

December, 2019

1 Executive Summary

The goal of this project was to explore the feasibility of Semantic Interaction (SI) methods [SI1, SI2] for Nuclear Femtography. Semantic Interaction is an approach to Human and Machine learning that enables the users to explore and refine their understanding of correlations and inter-relationships within large amounts of multidimensional data. Semantic Interaction combines statistical mathematics and machine learning with real-time scientific visualization.

While a variety of visualization techniques can help scientists to gain a more comprehensive understandings of their data, Semantic Interaction uses the history of the user's interaction to learn about what the user considers as relevant features and allows to map the n-dimensional correlations in a n-dimensional data set.

Toward the exploration of high-dimensional nuclear physics data, we pursued two objectives: 1) adapt our Graphically-Linked Ensemble Explorer (GLEE) to load the results of nuclear physics experiments and 2) evaluate the results with Jefferson Lab scientists and the CNF community.

For the first objective, the team determined the nature and distribution of the data and the appropriate statistics to apply in order to build the internal feature model; these statistics were implemented and validated. This required to extent our Semantic Interaction approach to ensembles and to advance GLEE for data-intensive challenges of nuclear science.

For the second objective, we identified two use cases: We used the properties of selected particles as a toy example to illustrate Semantic Interactions. We then explored the multidimensional dataset of semi-inclusive deep-inelastic scattering and studied the criterion for TMD factorization.

Our efforts have laid the foundation of a new multidisciplinary collaboration in Virginia, resulting in Semantic Interaction approaches and its GLEE technique being sucessfully adapted for Nuclear Femtography. This multidisciplinary collaboration can lay the foundation of a next-generation workspace for nuclear science.

2 Introduction

This project addresses the exploratory visualization problem for Nuclear Physics, finding relationships in with high-dimensional kinematic data. In exploratory visualization, we are examining and comparing 'under-specified' features and correlations in the data set. In other words, the 'features' of interest are unknown a priori, and may span several variables. The analyst may have some intuition about one or two relations in their data set, but no explicit formulation. Through the analytic partnership of Semantic Interaction, user and machine can converge on a formal description of combinations of interest. For our case study, we studied novel phenomena in Nuclear Physics:

Although the building blocks of the nucleon have been known for decades, theoretical and experimental understanding of how the quarks and gluons form the nucleon, and how their strong dynamics determines the nucleon's properties has been elusive. Most of the information about the nucleon's inner structure has emerged from the study of deep-inelastic scattering (DIS) of leptons, traditionally interpreted in the collinear approximation of perturbative Quantum Chromodynamics (pQCD), where the principle variable is the momentum of the quark, expressed as a fraction x of that of the nucleon, in a frame in which the latter is very large.

Advances in pQCD and experimental technologies of polarized beams and targets applied to the same DIS process are revealing correlations among this collinear momentum and the spins of both the parton and parent nucleon, as well as the intrinsic parton momentum component kT transverse to that of the nucleon. The key modern technologies are control of polarizations in the initial state without excessive penalty in luminosity, as well as high beam duty factors permitting detection of not only the scattered leptons but also identified hadrons in the final state with substantial acceptance (SIDIS). These hadrons carry the struck quark's intrinsic transverse momentum, albeit diluted by the fragmentation process, and the type of hadron provides statistical information about the struck quark's flavor. The correlations between x, k_T, and the spins of both the parton and parent nucleon are described in transverse-momentum dependent parton distribution functions (TMDs). The study of TMD observables is an essential part of the Nuclear Physics (NP) programs at the Jefferson Lab and the planned Electron-Ion Collider and will allow the mapping of the motion of quarks and gluons in nuclear matter.

In our proposal, we have started to explore the various kinematic regions in SIDIS as, e.g., described in "Mapping the Kinematical Regimes of Semi-Inclusive Deep Inelastic Scattering" by Boglione et al. [SIDIS]. While our ultimate goal is to study SIDIS measurements from Jefferson Lab's 12 GeV project, we have started our study with the pioneering data set of SIDIS off a transversely polarized proton and studied the criterion for TMD factorization for charged pions and kaons using Semantic Interactions.

Approaches in the vein of Semantic Interaction (SI) have shown promise in the analysis of data as varied as text corpi to simulation ensembles [SI3 , GLEE1]. While these examples provide compelling results that SI can reveal interesting and unknown patterns in high-dimensional data, they are each built and tuned with the nature of their data domain in mind. For example, text and topic modeling involves several assumptions, different from those of geophysics simulations. Crucial to this project was the regularization of terminology and language across our scientific domains and the team's cooperative learning about Nuclear Physics and Human Computer Interaction.

3 Data Domain

Our goal in analyzing these ensembles of observations is to find relationships among sets and subsets of observations, and to quantify, summarize, and analyze the high-dimensional phenomenon.

For our exploratory data set, we have used SIDIS data from Markus Diefenthaler's thesis on "Signals for transversity and transverse-momentum-dependent quark distribution functions studied at the HERMES experiment" [HERM]. The variables in the data set are : Run number, Burst number, Event number, Polarization (beam), Polarization (target), Q2̂, 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, and Phi_h.

Thus, the data provided the kinematic variables and azimuthal angles, Q^2, x, y, ϕ_s, z, $P_{h\perp}, \phi_h$, for pions and kaons which have been detected in coincidence with the scattered lepton. The subscript h denotes the hadron type for the hadrons produced in SIDIS. From the data, we have selected pions and kaons from their event weights and calculated qzp, the ratio of the hard scattering partonin scales:

$$P_{h\perp}^2/z^2 \ll Q^2 \tag{1}$$

and

$$1 \ll Q^2 * z^2/P_{h\perp}^2 := qzp \tag{2}$$

4 Semantic Interaction (SI)

Semantic Interaction provides a method for the machine to learn what features are important to the user. The process is a dialogue as the user expresses a configuration of the set that makes sense to them and the machine reciprocates by returning a new explanatory configuration that takes into account the user's interaction. Through repeated interactions, the machine and user learn from each other and converge to a weighted combination of high-dimensional attributes that explain the 'feature'.

There are three important components of traditional Semantic Interaction, which should be defined: the user's work space, the mapping service, and the similarity model, which references the original source data. First, the user's works pace has two main elements: an Object-level view and a Parametric-level view. The Object-level view is a two dimensional work space where the high-dimensional objects are down-projected by Weighted Multi-Dimensional Scaling (WMDS). In the 2D Object view, each item is represented as thumbnail image that users can directly manipulate. After WMDS projection, the axes in the Object-level view only represent the relative distance between objects in the high-dimensional space.

Users can arrange similar items closer and pull different items further apart. This interaction triggers an inverse MDS where the system tries to find the set of high-dimensional weights that satisfy the low-dimensional configuration expressed by the user's spatial layout. The work space also includes a Parametric-level view, which provides a slider for each attribute of the high-dimensional space. Using these sliders, users can set each weight individually. When satisfied with their interactions, the user clicks the 'Layout' button to re-project the Object Level space.

The second component is the middle-ware Web service that processes the users' projection requests and delivers new object positions. This bi-directional pipeline incrementally builds a weight vector, which we call the Similarity Model [BID]. In the case of forward-projecting (down from high D to low D), the weights are applied to each attribute in the data set and then MDS is applied. In the backward projection (up from low D to high D), we implemented an inverse MDS algorithm to solve for the weights. The mathematical details of these algorithms are described below.

The third component of Semantic Interaction is the data and the weight vector itself. In the terms of machine learning, the weight vector can be considered as the description of the features and relationships expressed by the user. Saving and reloading Semantic Interaction results (the weight vector, or feature model), enables scientists to re-evaluate new observations through a set of relations they have learned before.

The Graphically-Linked Ensemble Explorer (GLEE) we extended for this project adds additional visualization components and user interface features to the original SI work space. Specifically: in addition to the two linked views of the SI work space, we added observation sub-setting and linking with parallel coordinates and statistical views, including 1D box plots and 2D scatter plots (Figure 1). Semantic Interaction with statistical views provides a way for scientists to explore their hypotheses and data relations with detailed plots and numeric read outs as observations are sub set, interrogated, and compared.

4.1 Simple Example: Particle Zoo

To help illustrate the process of Semantic Interaction, let's consider an example that should resonate with these readers. If we look at the attributes of fundamental particles in the Standard Model, we can cluster and categorize them in several ways. For example, by mass, charge, or year of discovery, among others. Sorting by each of these attributes (weighting them in high-dimensional space by way of SI) yields different configurations in the low-dimensional work space. The feedback between user and model constructs a weighted projection that embodies the user's understanding.

For example, with this data set we can illustrate several aspects of Semantic Interaction, especially its power to show clusters and relationships among observations. Specifically, with Parametric interaction, we can quickly see what observations share a property, such as high spin charge and low mass. On the other hand, a user might group a set of particles together and upon re-projection find they all have some other attribute combination in common (Figure 2, Figure 3. Figure 4) . Recall that the Object-level space has no axes or units: it is simply low-dimensional projection of a high-dimensional data set that attempts to preserve the relative distance between items. Thus, similar items should be closer and different items further apart.

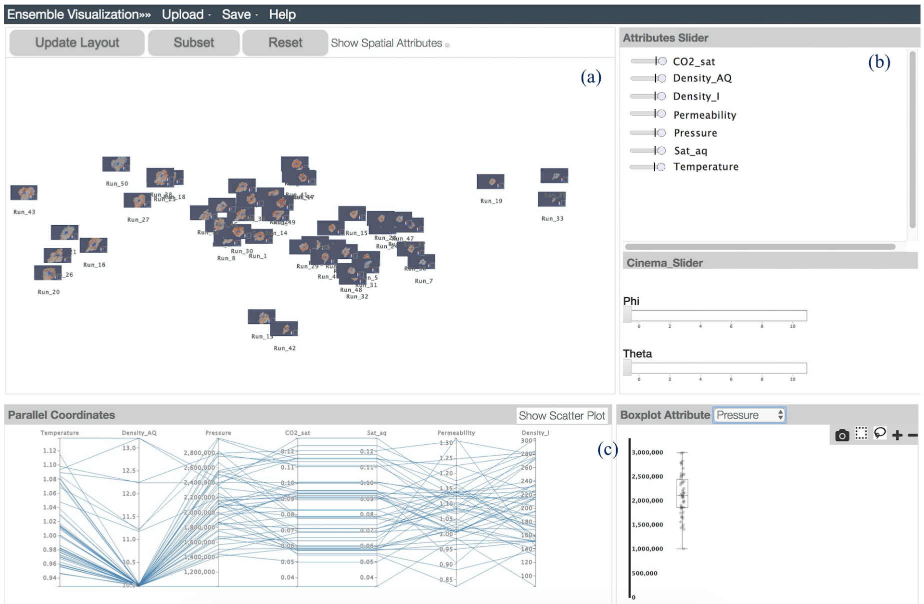

Figure 1: Screenshot of the original GLEE showing a geophysics simulation ensemble's analysis. Top left is the Object-level work space, which has no axes or scale (a); Upper right is the Parametric-level view with a slider to weight for each attribute (b); at bottom are the statistical view components (c).

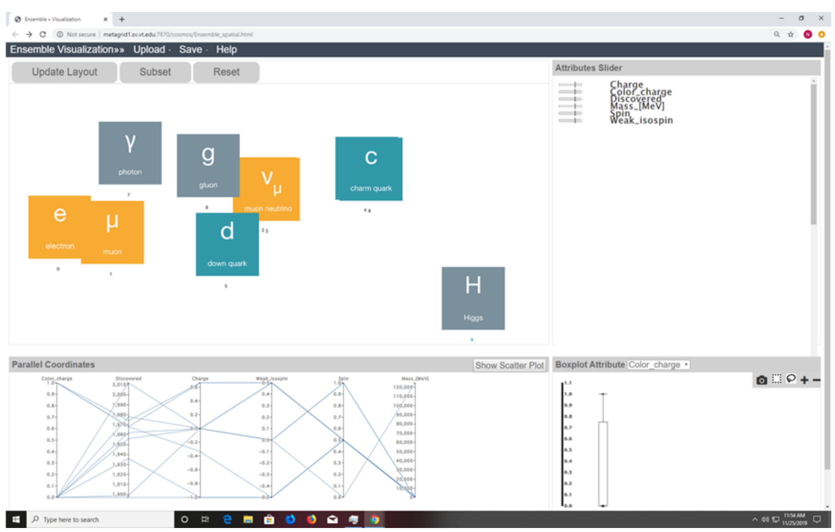

Figure 2: Particle Zoo initial projection

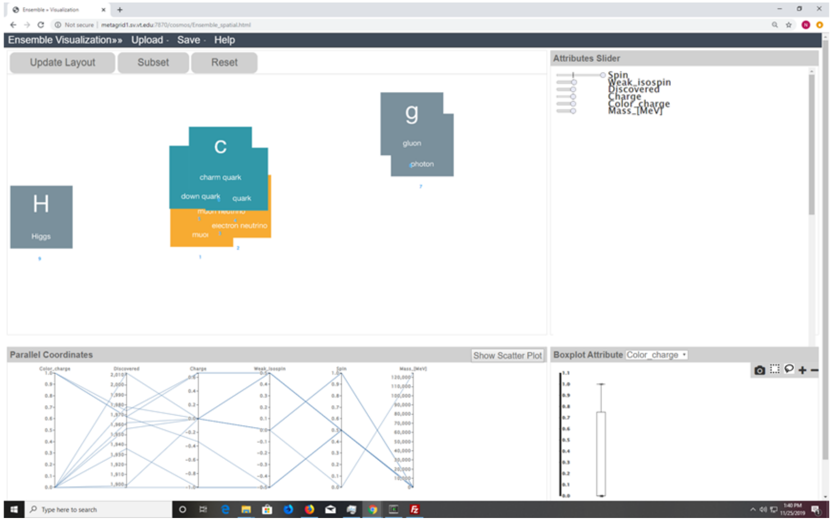

Figure 3: Particle Zoo projected by spin

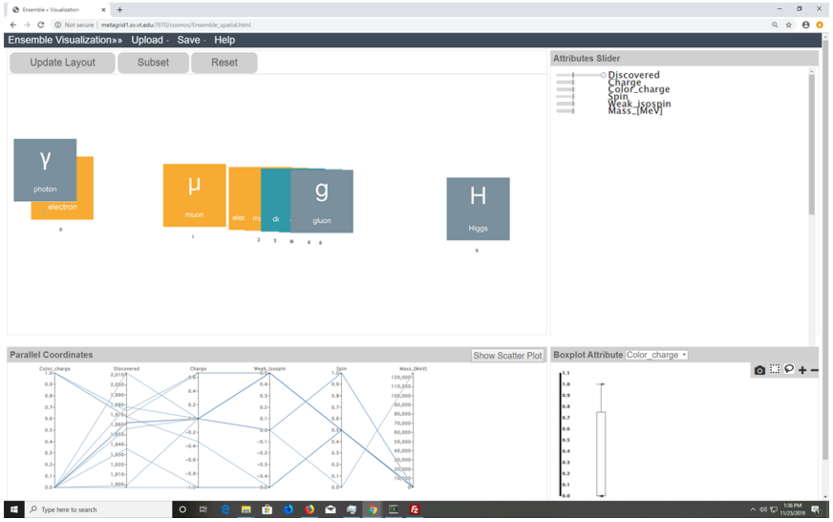

Figure 4: Particle Zoo projected by year of discovery

5 Approach for Particle Kinematics

5.1 Data Extraction

With thousands of rows of particle events, we could not fit a thumbnail of each event into the Object-level SI work space. Even with large screens, the practical limits is a handful of hundreds [LDIS]. Therefore, we needed an aggregation strategy for cutting the observations into groups (ensembles of event observations). Our first strategy was to make cuts of the data by particle type and charge and the further split them based on magnitude of qzp.

Here we describe the data sampling for each thumbnail. Note of Terminology: we consider the raw columns as attributes for each row (an observation, or item). When we aggregate rows, our attributes are be combined into a distribution. For a given set or subset, we describe its distribution by some parameters (e.g. the Mean, Median, Standard Deviation), which are components of a vector (parameters) summarizing each attribute's range and tendency. From a user point of view, sliders in the Parametric-level view are currently named one for each attribute. In the WMDS models however, we apply the attribute weight to each component of that attribute's distribution parameters (V1 uses the Mean and STD).

Originally, there are 2mil rows in the HERMES data. We filter these rows into sets for each particle type: Pi+, Pi-, K+, K-, Pi0. For each particle set, we divide further into 3 groups by the qzp values: qzp below 0.5, qzp between 0.5 and 1.5, qzp above 1.5 We choose to make n thumbnails for each group. Let the number of rows in the smallest group of the three be k. (Smallest in terms of number of rows) We randomly select floor(k / n) rows in a group and use these to form a thumbnail. We 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.) The Mean value and Standard Deviation are calculated and stored for each thumbnail data's individual components.

5.2 Particle Portraits

We used the open-source Paraview codebase to produce Python scripts for data loading and for image and 3D model generation. We created and validated the pipelines by importing and visualizing the known HERMES data. We exported these visualizations as ISO-IEC Extensible 3D (X3D) models (www.web3d.org) and Cinema image databases [CIN1]. the Cinema images were loaded into GLEE with the csv of the data file.

Our initial efforts created 3-dimensional histograms of x, Z, and Ph-perp. We created volumetric visualizations and 3D topographic plots of each particle in the 3 different TMD regimes. However, the size of distinguishing features and the perspective rendering made it difficult for users to perceive distinguishing features among them. Some potential design solutions to this may be: a better scaling for the drawing space or improved colorizing, both of which we applied to our Version 1.0 release (V1).

Our current solution is to use a 2D contour at 3 iso values. The lines were plotted using 1 hue (red) plus luminance for the traced scalar value of frequency. This approach does appear to portray characteristic differences in the distributions of the data summarized in the thumbnail. Our design for the V1 portraits were improved by: adding axes and insuring equivalent proportion and framing of data in the thumbnail (Table 1).

Creation of a thumbnail is achieved by this process: We 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. Using binning with 10000 bins, we estimate the density of the points at each point's location. Next, we create a contour plot of the density map, and color code it from white to red. Finally, we show the data axis and save the screenshot.

6 Problem formulation for MDS

When we attempt to calculate the backward projection, we are looking for the inverse of the forward projection: the weight combinations that minimize the stress of our objective function. The objective function to be minimized or otherwise referred to as the stress functions that is seen for most Euclidean metric MDS takes the following form:

$$\textbf{Stress}^2 = \sum_{i=1}^{nrows} \sum_{j=i+1}^{nrows} \Big(([X_i - X_j][W][X_i - X_j]^T - [Z_i - Z_j][Z_i - Z_j]^T) \Big) \quad (3)$$

Where $[X]$ is the $nrows \times n$ high dimensional matrix of observations that we have and $[Z]$ is a low-dimensional $nrows \times p$ matrix that is the dimension-reduced data we hope to obtain through MDS. The weight matrix $[W]$ is a diagonal matrix of dimension $n \times n$. This weights on the diagonal of this matrix are the weights that the user interacts with in GLEE. A row (or vector) of each of the high-dimensional (HD) and low-dimensional (LD) matrices are represented as X_i and Z_i respectively. Elements of these matrices are represented using the notation X_{ij} and Z_{ij} respectively. However, w_i corresponds to an element of the diagonal on the weight matrix.

Note that 1-based indexing is used for all indices here throughout this document.

We can write a general form of MDS for extending this to different distance norms and ensembles of data.

$$\textbf{Stress}^2 = \sum_{i=0}^{nrows} \sum_{j=i+1}^{nrows} \Big(\delta_{HD}(X_i[W], X_j[W]) - \delta_{LD}(Z_i, Z_j) \Big)^2 \quad (4)$$

where term A and term B are defined as the distance between two rows using all the attributes/columns of those rows as:

$$\textbf{term A} = \delta_{HD}(X_i[W], X_j[W]), \textbf{term B} = \delta_{LD}(Z_i, Z_j) \quad (5)$$

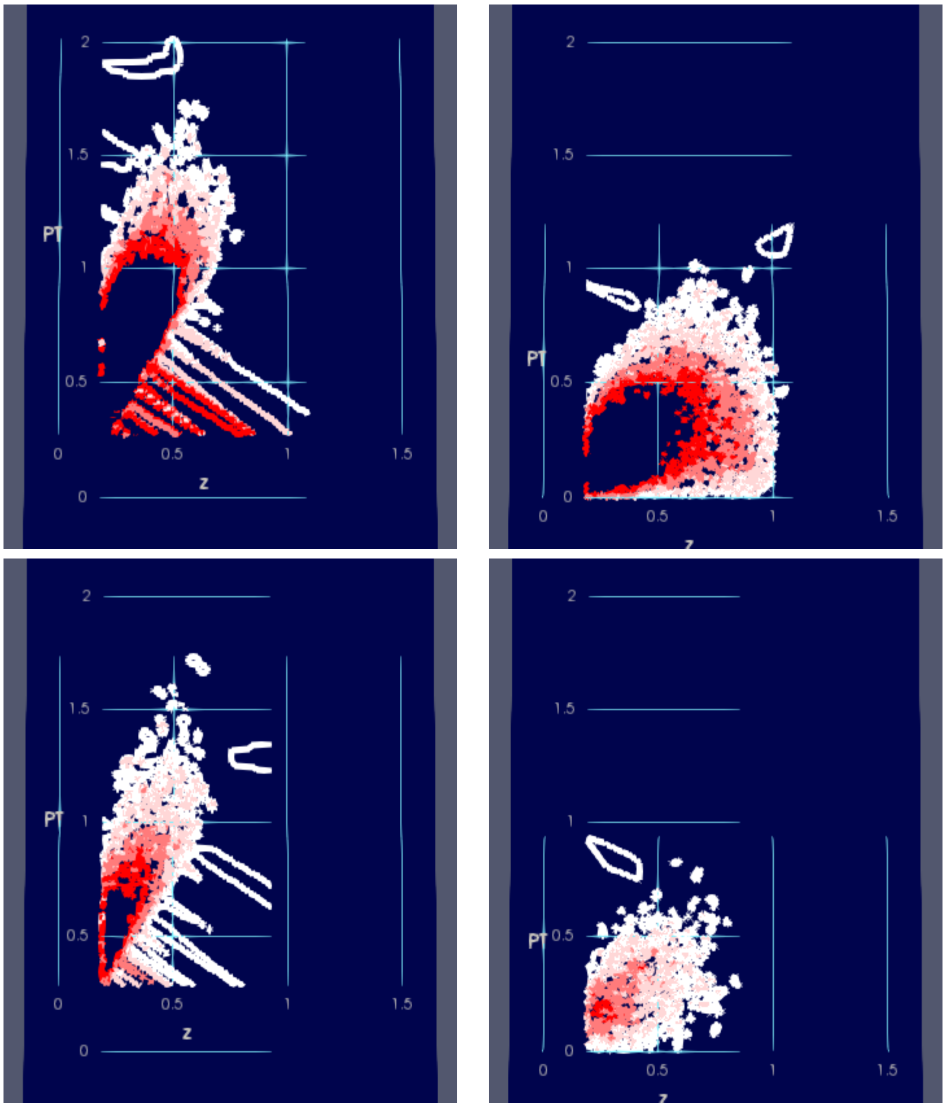

Table 1: Example Particle Portraits: Pions (top; positive on left, negative on right) and Kaons (bottom; positive on left, negative on right)

6.1 Euclidean norm distance/ L2 norm distance

$$\delta_{HD}(X_i, X_j, W) = \sqrt{\sum_{k=1}^{n} (X_{ik} - X_{jk})^2 w_k^2} \tag{6}$$

$$\delta_{LD}(Z_i, Z_j) = \sqrt{\sum_{k=1}^{p} (Z_{ik} - Z_{jk})^2} \tag{7}$$

We used the Euclidean distance here although we can use a squared euclidean distance as well to avoid having to take the square root.

6.2 Manhattan distance/ L1 norm distance

$$\delta_{HD}(X_i, X_j, W) = \sum_{k=1}^{n} |X_{ik} - X_{jk})| w_k \tag{8}$$

$$\delta_{LD}(Z_i, Z_j) = \sum_{k=1}^{p} |Z_{ik} - Z_{jk}| \tag{9}$$

6.3 Inner product norm for distance

$$\delta_{HD}(X_i, X_j, W) = (X_i[W]) \cdot (X_j[W])^T \tag{10}$$

$$\delta_{LD}(Z_i, Z_j) = Z_i \cdot Z_j^T \tag{11}$$

7 MDS for Ensembles

Term A currently only uses the mean of an ensemble right now as it approximates the ensembles with a point estimate. Keep in mind that term A computes the distance between 2 rows of observations whether it is a point or a point estimate from an ensemble.

We have two options to compute the distance between the 2 rows modelled as a distribution.

1. Consider each bin as a multivariate Gaussian distribution from all the attributes and compute the distances between the multivariate distributions.

2. Consider each bin and each attribute separately. Compute the distance between each bin for each attribute separately and sum them up.

Both of these approaches involve computing the distances between Gaussian distributions. The first one would compute the distances between multivariate distributions whereas the second one involves the sum of distances between several univariate distributions.

If we consider option 2, we can write the high-dimensional distance as:

$$\delta_{HD}(X_i, X_j, W) = \sqrt{\sum_{k=1}^{n} D_b(G(X_{ik}), G(X_{jk}))^2 w_k^2} \tag{12}$$

$$= \sqrt{\sum_{k=1}^{n} D_b(G(\mu_{ik}, \sigma_{ik}), G(\mu_{jk}, \sigma_{jk}))^2 w_k^2} \tag{13}$$

This can be simplified to to give an approximate distance as

$$D_{bk} \approx (\mu_{ik} - \mu_{jk})^2 + (\sigma_{ik} - \sigma_{jk})^2 \tag{14}$$

8 Challenges

Applying Semantic Interaction to the data domain of Nuclear Physics requires the deliberate formulation of the statistics used in the SI weighting scheme, use of the appropriate distance functions, and convergence criteria appropriate to real-time interaction. Through this collaboration, we were able to define each of these and provide the scientific confidence that features observed in GLEE had physical significance. This bodes well for future research and application of this method for data exploration and discovery.

There is also the issue of the scalability of SI work spaces like GLEE; there are challenges in the analytics and the visualization. For example, the number of items on the screen, the number of items, the number of attributes are all challenges of both design and computation. In this project, we had to tackle both of these challenges of formalization and scalability.

Specifically, we had to adapt the GLEE platform to work with the assumptions carried by our event aggregation strategies (e.g. cutting above). While GLEE was built to analyze multiple individual simulation runs in an ensemble, our case aggregates multiple observations into a thumbnail (an object). Thus, our objects are really summaries of several observations and have descriptive statistics like Mean and Standard Deviation, among others. In Version 1, we use Mean and Standard Deviation as equally-weighted components of the attribute to describe the distribution of observations in each set. Thus, we are using Semantic Interaction to ask, 'What is similar or dis-similar about the distribution of the variables in each set of observations?'

9 Results

This project supported time for VT faculty and an undergraduate wage student (CS Major). With weekly video conference and screen-sharing meetings, we developed a strong cross-disciplinary collaboration for the Center for Nuclear Femtography. This grant also provided travel support for each team to travel to the other and mix their perspectives and communities through guest lectures and dedicated face-to-face time at the respective labs. The value of this time

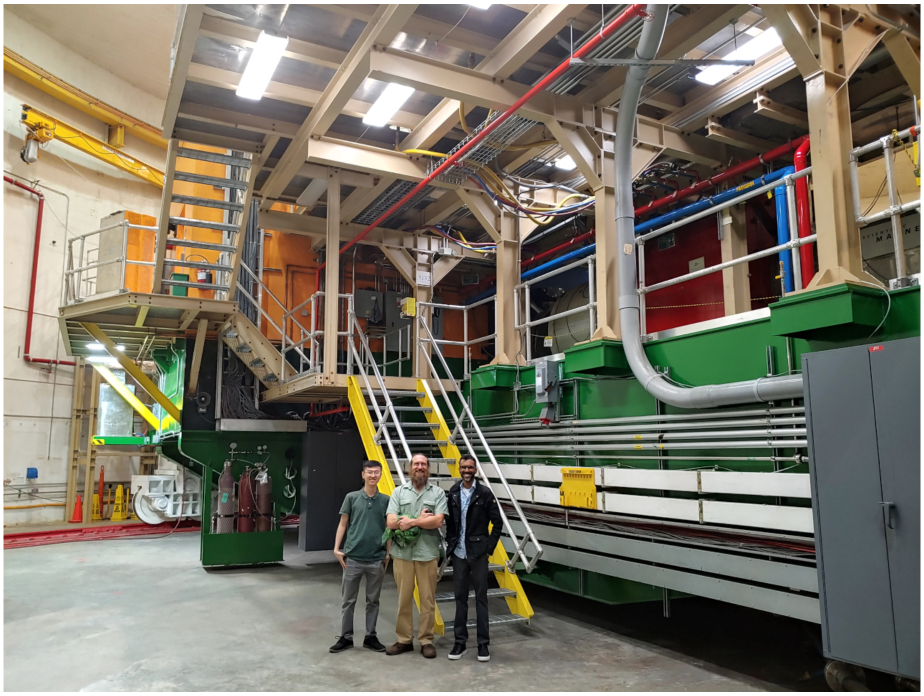

Figure 5: VT at JLab.

cannot be underestimated (Figure 5, Figure 6). At Jefferson Lab, Dr. Polys presented at the monthly Computing Round Table, "Perspectives on Visualization and Science: Communication and Discovery" [SCI1]. At Virginia Tech, Dr. Diefenthaler presented "Exploring the heart of matter at Jefferson Lab" to a packed room of faculty and students in Torgersen Hall.

Through this project, we adapted the GLEE environment to support particle kinematics data and demonstrated the mathematical and statistical foundations of SI with kinematic data. Remember that the Object-level space has no axes or units: it is simply low-dimensional projection of a high-dimensional data set that attempts to preserve the relative distance between items. Thus, similar items should be closer and different items further apart. Crucial to the success of our project was the regularization of our different terminology and language across computer science, statistical mathematics, and Nuclear Physics; this enabled us to dive deeply into the mathematics and statistical assumptions of GLEE, revising the distance metric and WMDS algorithm to work with distributions of observations, rather than singleton observations.

There was a strong effort in this project for implementation as several software components (Python, Paraview, X3D) needed to be integrated for the Nuclear Physics use cases and new features were required to manifest our revised statistics into GLEE. We created an instance on a Virginia Tech server:

Figure 6: JLab at VT.

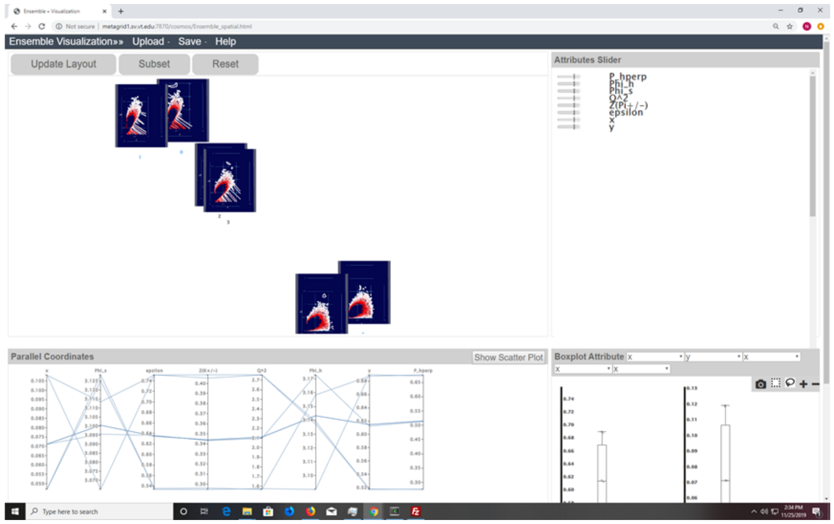

Figure 7: Positive Pions initial projection

GLEE for Particle Kinematics. By putting this online at a url, were were able to distribute the platform and gather input from the scientific community. We also included two smaller example data sets (Particle Zoo and Toy Kinematics) that demonstrate how SI works. We also created a project website that uses video, text, and graphics to explain Semantic Interaction using the Particle Zoo data set: Project Website

9.1 Augmented Kinematic Test

In addition to the Particle Zoo example, we did another Proof-of-Principle by validation of the projection and weighting of our particle kinematics tables. We added several columns with known correlations, and were able to see these known features and relationship cluster in GLEE as expected (the 'Toy Kinematics' data set). Following this milestone, we proceeded to put each particle population into GLEE as an example data set.

9.2 Pions

We made portraits for positive and negative Pions. We noted that Phi-s and Phi-h change cluster membership by WMDS down projection. Positive Pions are shown in Figure 7, Figure 8, Figure 9; Negative Pions are shown in Figure 10, Figure 11, and Figure 12.

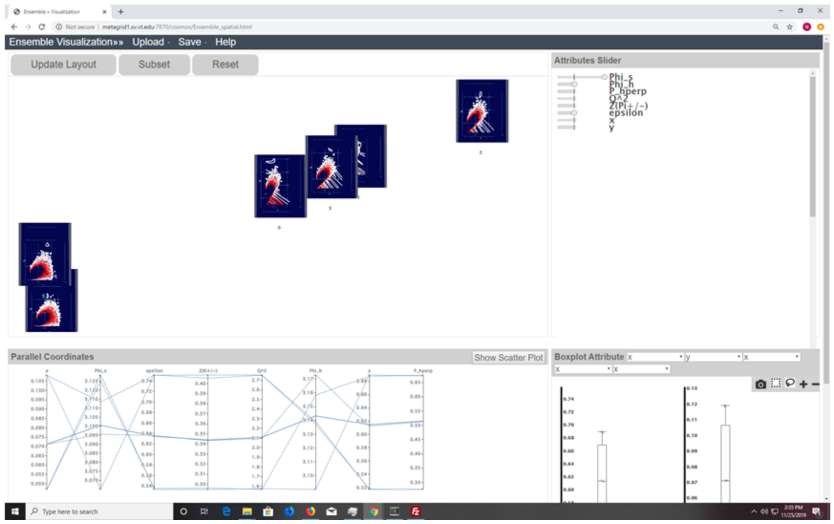

Figure 8: Positive Pions projected with Phi-h weighted

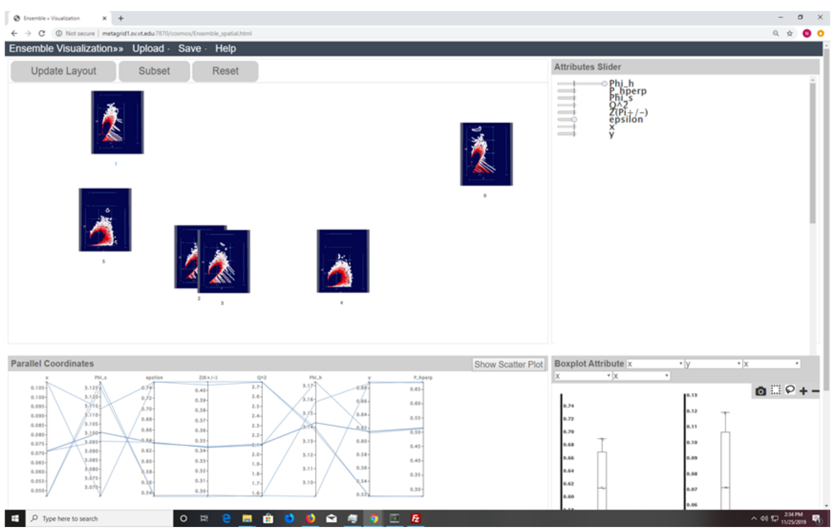

Figure 9: Positive Pions projected with Phi-s weighted

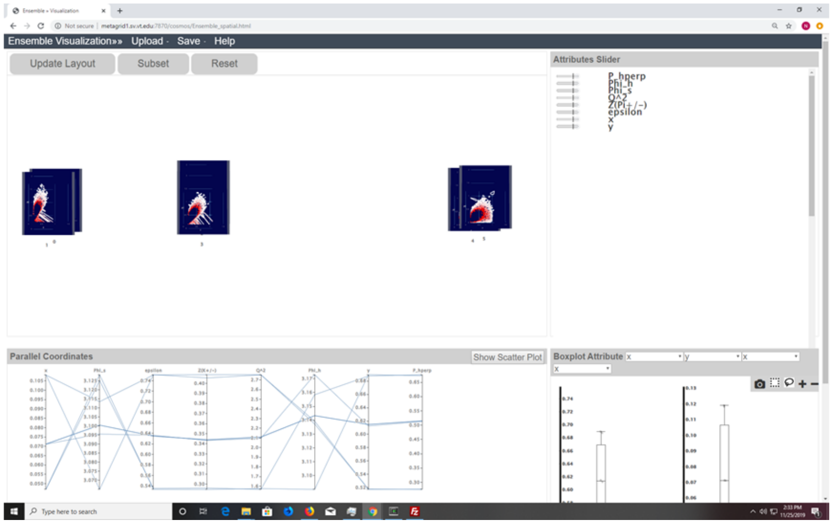

Figure 10: Negative Pions initial projection

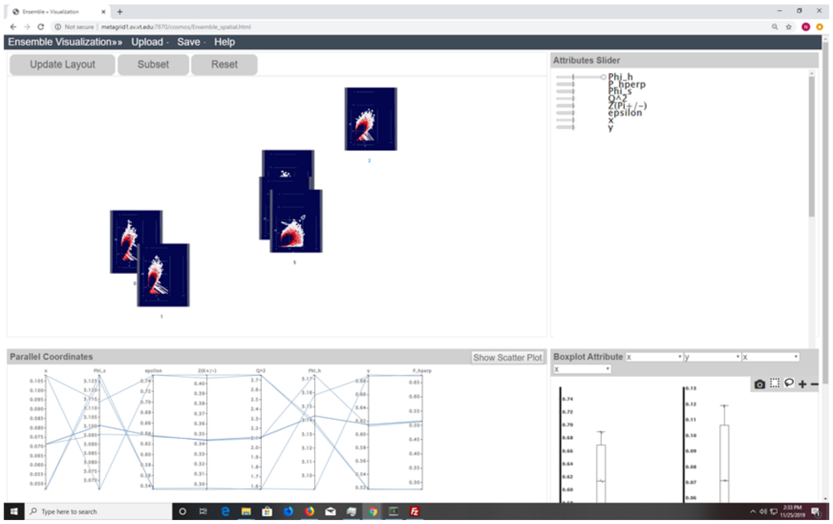

Figure 11: Negative Pions projected with Phi-h weighted

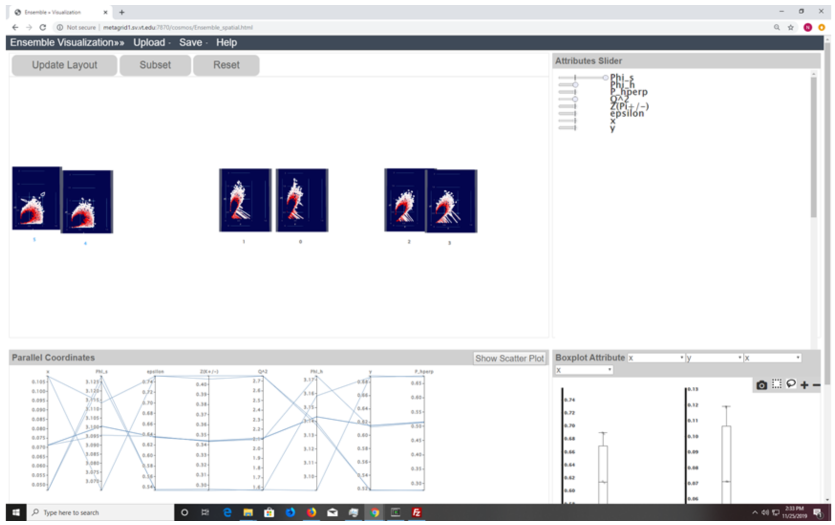

Figure 12: Negative Pions projected with Phi-s weighted

9.3 Kaons

We made portraits for positive and negative Kaons. We noted that Phi-s and Phi-h and Y change cluster membership by WMDS down projection. Positive Kaons are shown in Figure 9.3, Figure 14, and Figure 15. Negative Kaons are shown in Figure 16, Figure 17, and Figure 18. Note that Phi-s and Phi-h weighting changes cluster membership. In the case of Positive Kaons, the three initial groups could be re-sorted into two in the OLI work space where the portraits had 'tails' or not; re-clustering from three groups into two, and then updating the layout significantly up-weighted 'y'.

10 Conclusions and Future Work

This project has given us confidence in the application of the Semantic Interaction method, and a path to further development with the GLEE platform adapted to Nuclear Physics. The results are illustrated with a website, video, and a live online demo. The collaborations from this Center for Nuclear Femtography have spawned new questions and ideas and future proposals that will include members of the larger CNF community.

During the supplemental period remaining in the 2019 calendar year, we will collect our lessons into a 1.1 Version release, which will include: a) any identified bug fixes, b) new data cuts and portrait methods, and c) the ability to directly upload new data sets into GLEE.

Through this project, we have shown Semantic Interaction's capacity to tease out interesting known and unknown relationships among high-dimensional Nuclear Physics data. The Semantic Interaction method provides fertile ground for

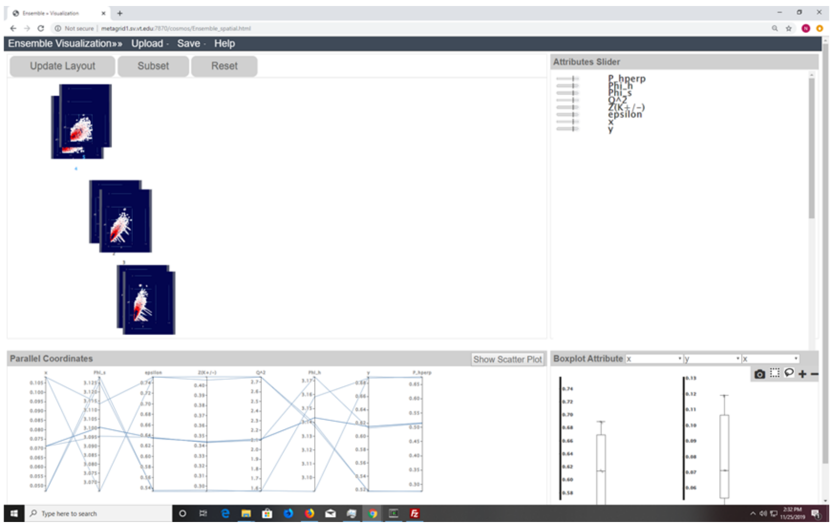

Figure 13: Positive Kaons initial projection

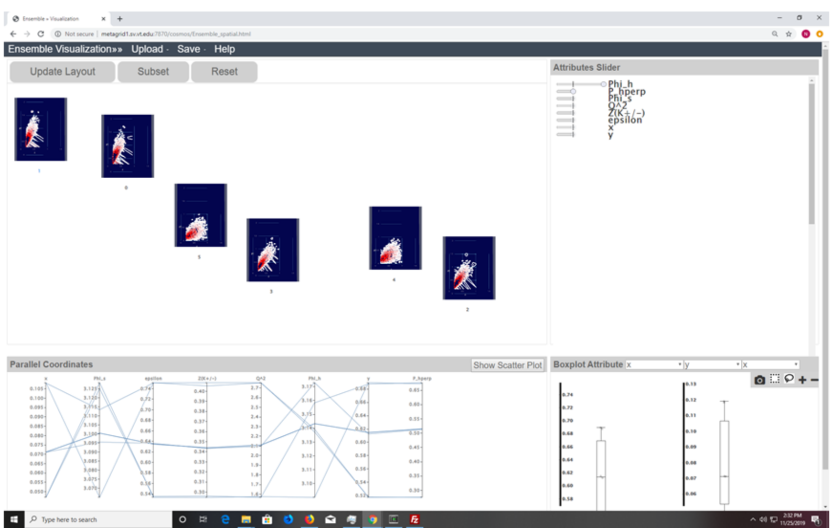

Figure 14: Positive Kaons projected with Phi-h weighted

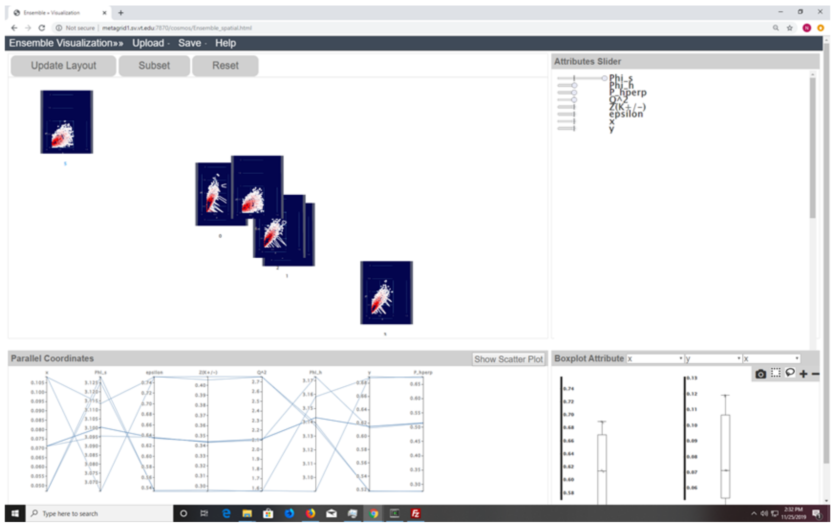

Figure 15: Positive Kaons projected with Phi-s weighted

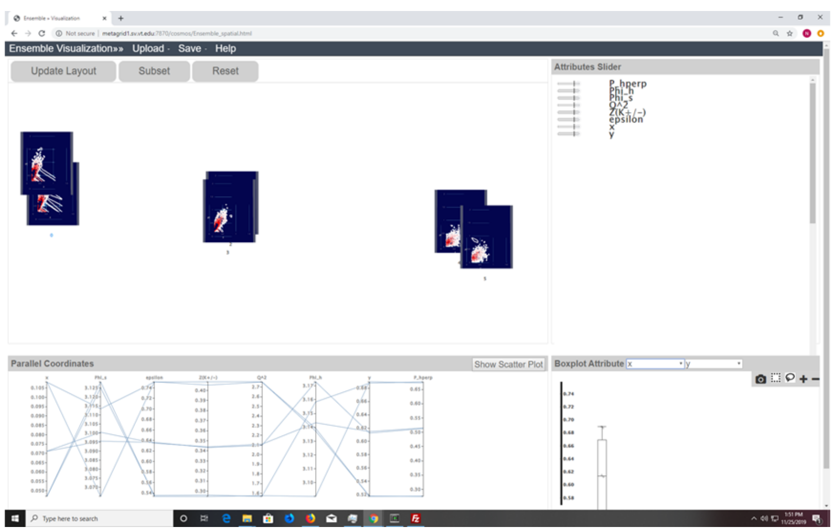

Figure 16: Negative Kaons initial projection

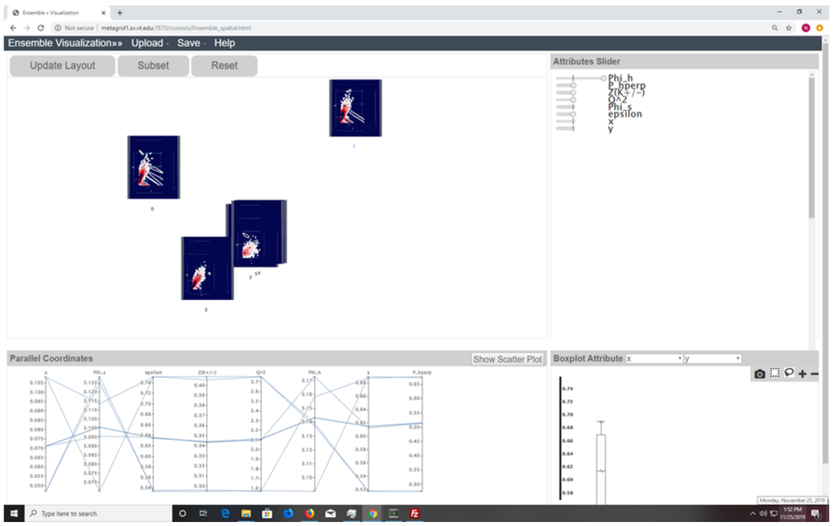

Figure 17: Negative Kaons projected with Phi-h weighted

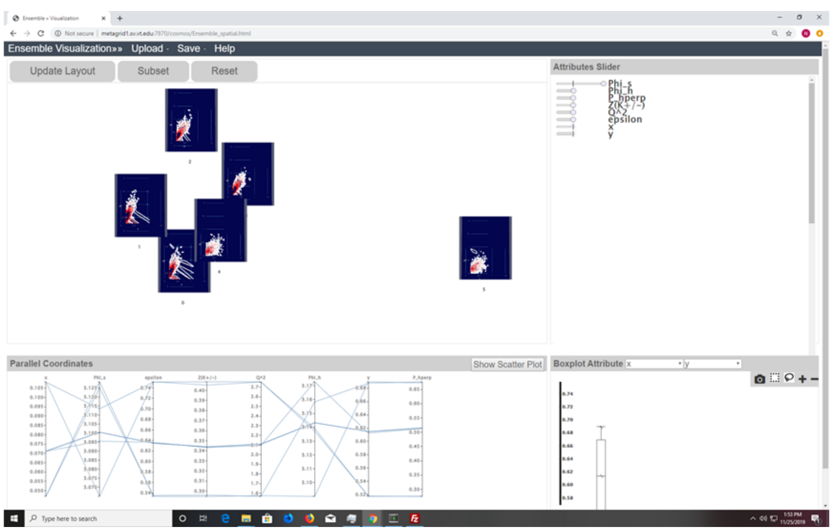

Figure 18: Negative Kaons projected with Phi-s weighted

future work that should include research into the wider possible combinations of cognitive (human) and machine (statistical) models. Specifically, great progress could be made with additional expertise on the team such as Bayesian Statistics and Optimization, High-Performance Computing (HPC), and a detailed mapping of the kinematic region of SIDIS, including not only the HERMES data but also first data from the SIDIS measurements at the 12 GeV science program.

11 References

[SI1] Endert, A., Fiaux, P., North, C.: Semantic interaction for sensemaking: inferring analytical reasoning for model steering. IEEE Transactions on Visualization and Computer Graphics 18, 12 (2012),2879–2888.

[SI2] Endert, A., Chang, R., North, C., Zhou, M.: Semantic interaction: Coupling cognition and computation through usable interactive analytics. IEEE Computer Graphics and Applications 35, 4 (2015),94–99.

[SI] Dowling, M., Wycoff, N., Mayer, B., Wenskovitch, J., House, L., Polys, N., North, C. and Hauck, P., 2019. "Interac-tive Visual Analytics for Sensemaking with Big Text". Big Data Research, 16, pp.49-58.

[Cin1] Woodring, J., Ahrens, J. P., Patchett, J., Tauxe, C., Rogers, D. H.: High-dimensional scientific data exploration via cinema. In 2017 IEEE Workshop on Data Systems for Interactive Analysis(DSIA)(2017), IEEE, pp. 1–5.

[GLEE1] Dashan, Polys, Pollyea, Jayne (under review). "Making Sense of Scientific Simulation Ensembles with Semantic Interaction" Computer Graphics Forum.

[BID] Dowling, M., Wenskovitch, J., Hauck, P., Binford, A., Polys, N. and North, C., (2018). A Bidirectional Pipeline for Semantic Interaction. In Proceedings of the Workshop on Machine Learning from User Interaction for Visualization and Analytics (IEEE VIS 2018) (Vol. 11).

[LDIS] Andrews C, Endert A, North C. Space to think: large high-resolution displays for sensemaking. InProceedings of the SIGCHI conference on human factors in computing systems 2010 Apr 10 (pp. 55-64). ACM.

[HERM] Diefenthaler, M., 2010. Signals for transversity and transverse momentum dependent quark distribution functions studied at the HERMES experiment (No. DESY-THESIS-2010-032).

[SIDIS] Boglione, M., Dotson, A., Gamberg, L., Gordon, S., Gonzalez-Hernandez, J.O., Prokudin, A., Rogers, T.C. and Sato, N., 2019. Mapping the Kinematical Regimes of Semi-Inclusive Deep Inelastic Scattering. arXiv preprint arXiv:1904.12882.

[SCI1] Polys, Nicholas "Perspectives on Visualization and Science: Communication and Discovery Jefferson Lab June 11, 2012

https://www.jlab.org/indico/event/307/session/5/contribution/13/material/slides/0.pdf