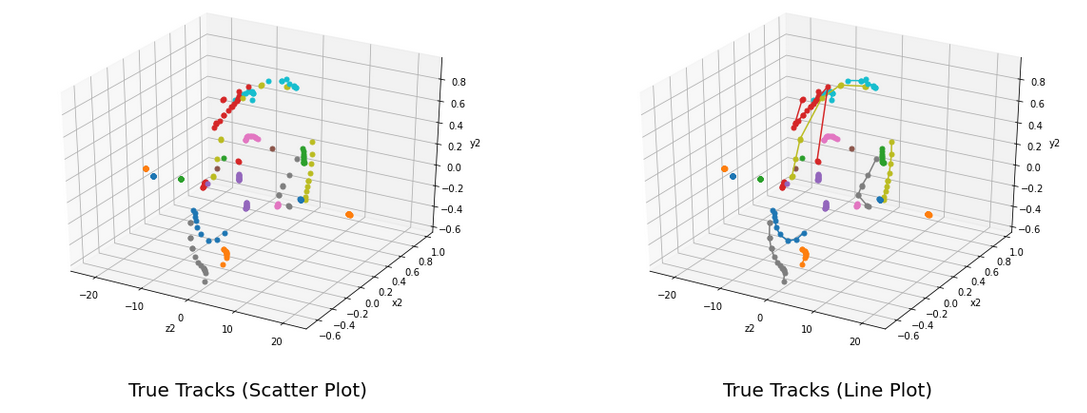
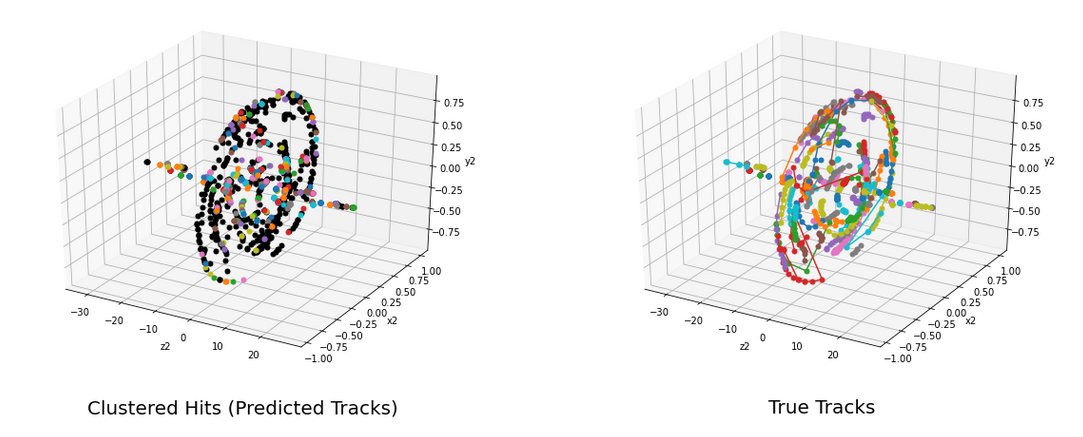
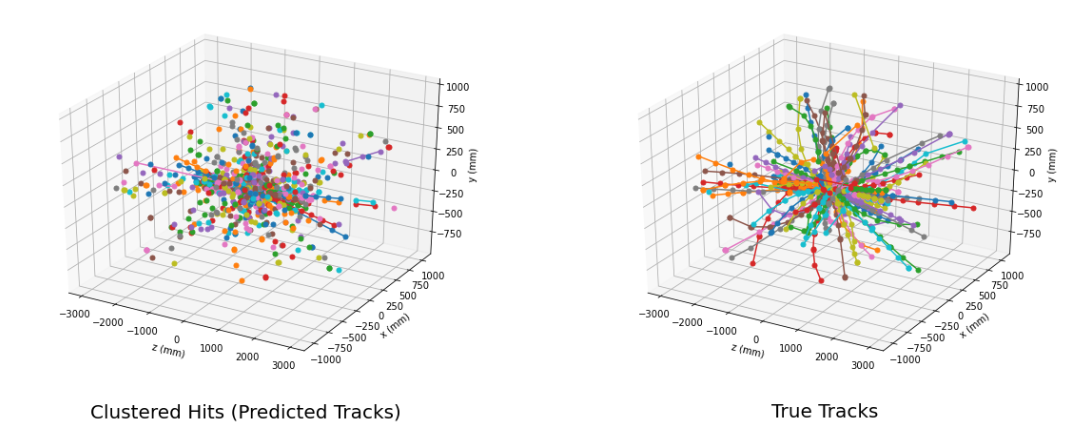
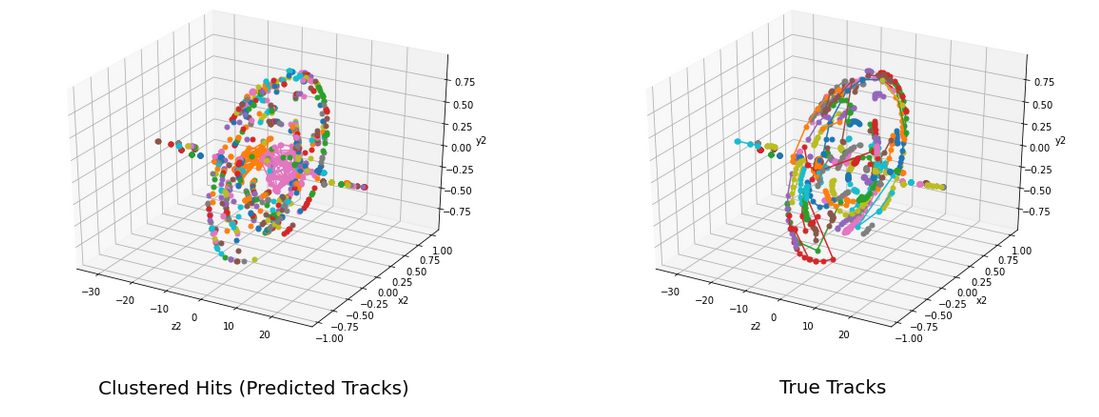
Project Proposal #3

In this first model, we used a clusterer to develop our model. By doing this, it allows for us to find the predicted tracks of the particles that we were wanting to discover and then allows for us to develop a model based around that. The mean score for these events were approximately .199, which is low for this model. A score being close to 1 would produce more accurate results, but since this model contains so much data, it seems as though this model is predicting the wrong values, based on the features input.

We then decided to us a second model of a different number of clusters within the data, this then uses a different number of clusters within the model. This allows for more samples to be clustered and creates a model such that there are more tracks. Our score for this model was also approximately .195, which is lower than the first model, but it does have many more clusters, this could be an example of overfitting because there were over 6,000 clusters created by this one operation.

This visual display the predicted tracks based on the previous model, as it is shown, it does have too many tracks for the true tracks that we have. This could be a result the overfitting as discussed in the previous model.

In this next model, we changed the eps value, meaning there is a less amount of density required to form a cluster, and then also allows for less tracks. As shown, this model does have less tracks and less clusters, but the score is greatly damaged by this, being a 0.0644. This strikingly low number can be attributed to the change in the EPS value, creating a major difference in our model.

The model below shows with different model with scaling and transforming the model based on our coordinates. This then allows for something similar and then creates a score with approximately .1481. This score isn’t too bad, but it seems as the EPS really effects on how the score is produced.

