

Status of clustering

Ben Carls

Where clustering is now

- Want to finalize clustering occurring right after hit reconstruction
- Lots of tools already available (e.g. DBSCAN, Hough line finder, and End Point Finder)
- Goal is to bring them all together

Start off with fuzzy clustering

- Points now have degrees to which they belong to clusters (“fuzzyness”)
- Provides a good starting point to finding electron showers
- Basic idea is:
 1. Assign random cluster memberships to points
 2. Iterate over the memberships to minimize an objective function

The basic algorithm

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k -step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m}$$

3. Update $U^{(k)}, U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. If $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP; otherwise return to step 2.

The objective function we are trying to minimize:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2$$

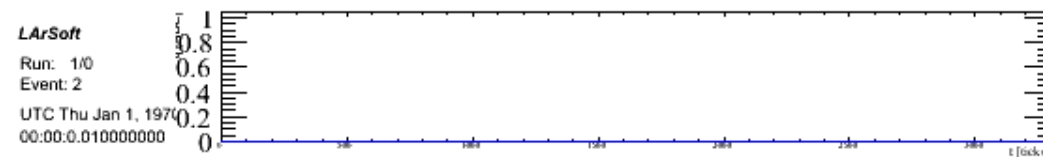
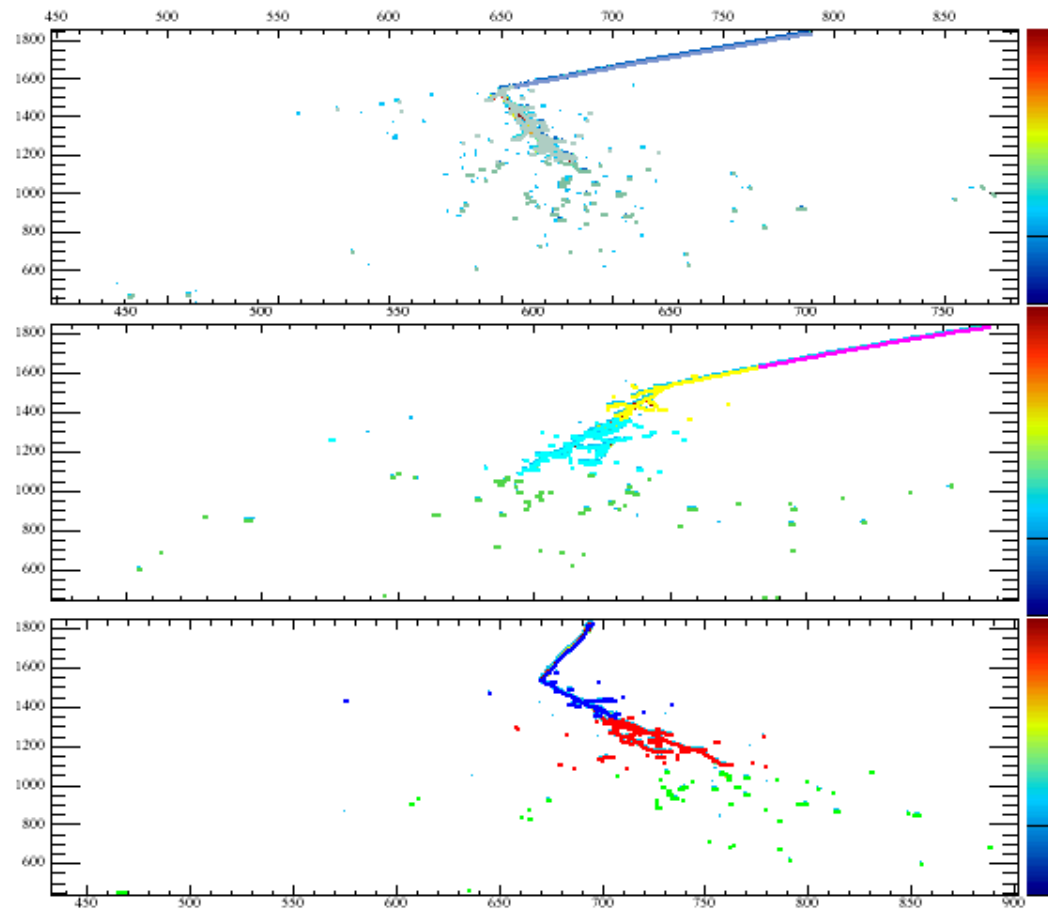
Number of clusters

- The number of clusters needs to be given as an input
- The Xie-Beni index gives us a way to evaluate how well a certain cluster number works

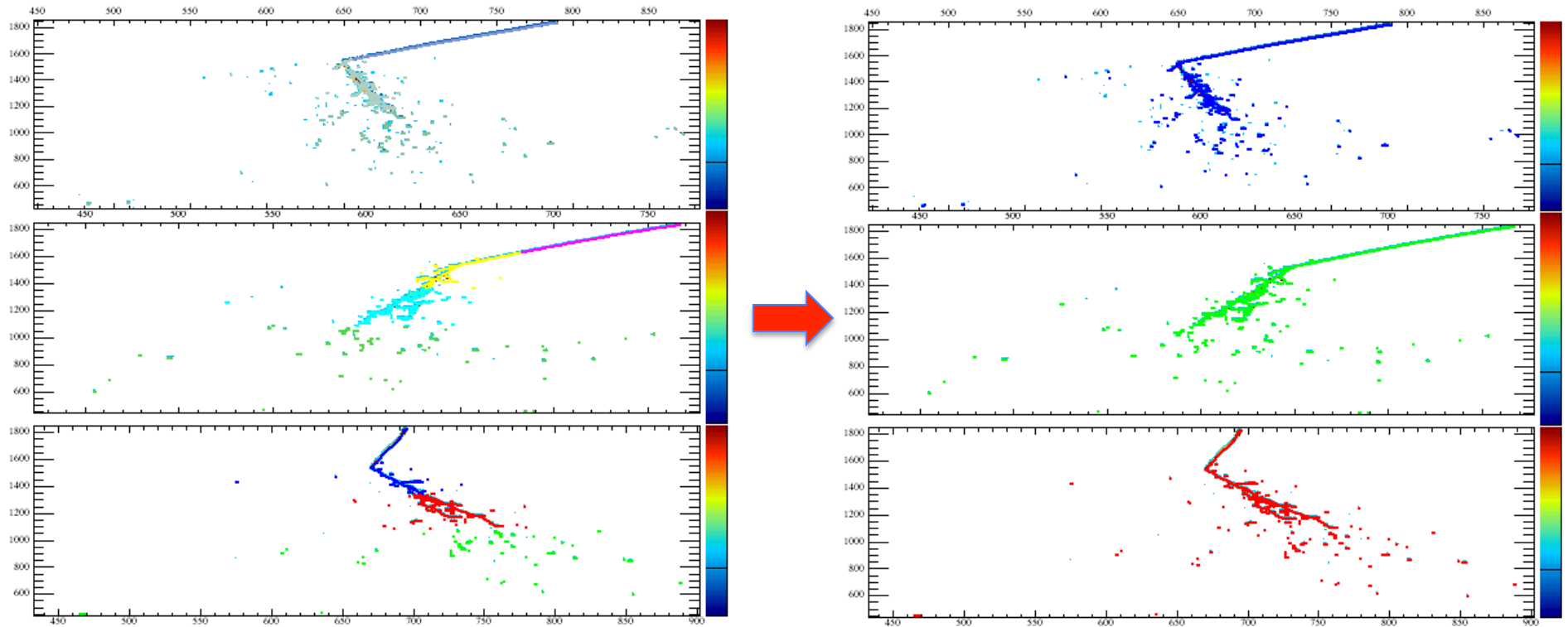
$$\frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K u_{ik}^m \|x_i - c_k\|^2}{\min_{k,\ell} \|c_k - c_\ell\|^2}$$

Does not work with single cluster events by definition,
we'll get to that later

There is a problem with the Xie-Beni Index though

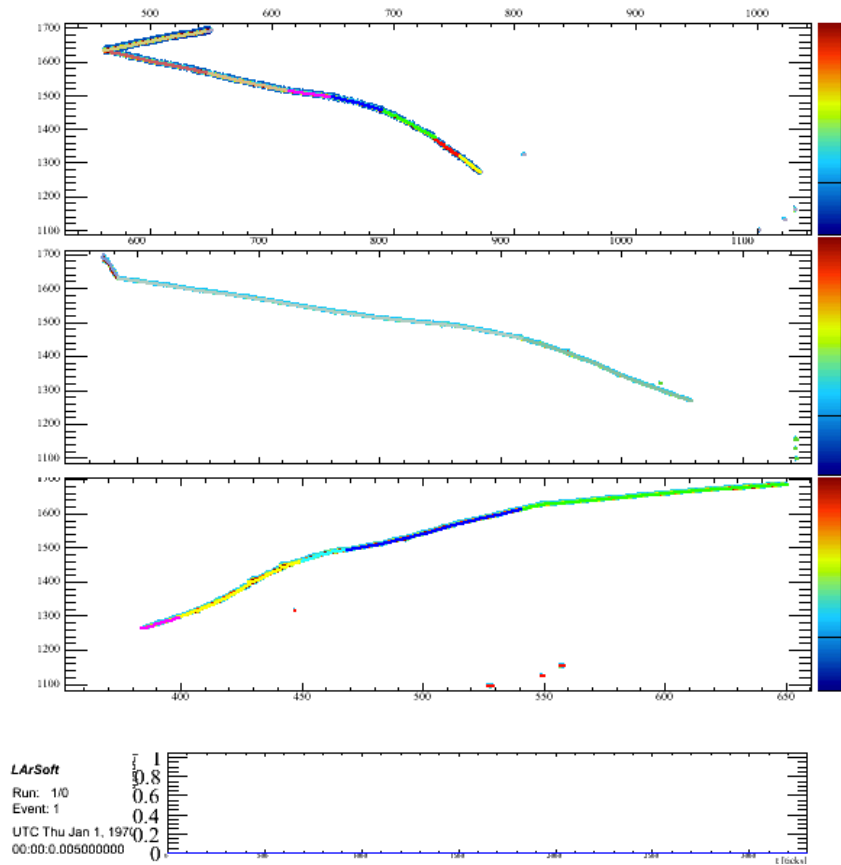


Look to merge clusters sing Euclidean distance



1. Look for point in cluster i closest to centroid in cluster j
2. Compare that point to the point in the cluster j closest to centroid of cluster i
3. Merge the clusters

Now look for Hough lines and expand them into clusters

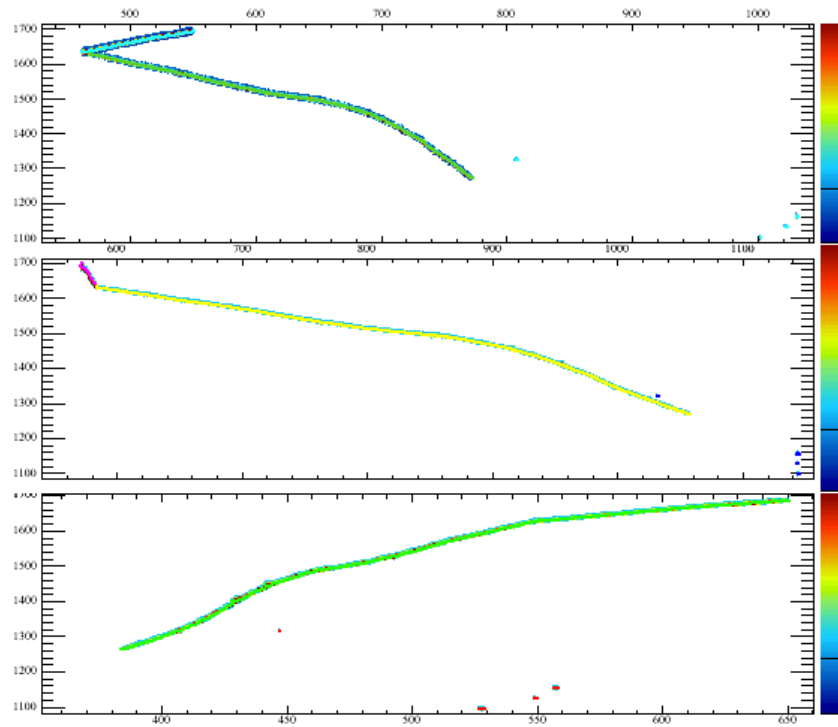


Search for Hough lines in the already identified fuzzy clusters

Hough line finder was sped up to make this practical

An issue remains with needing to merge the Hough lines though

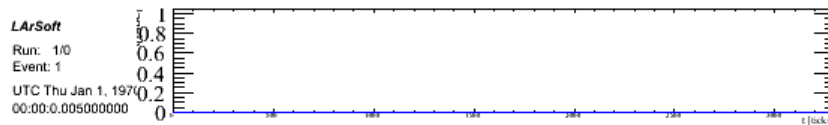
Merging the Hough lines



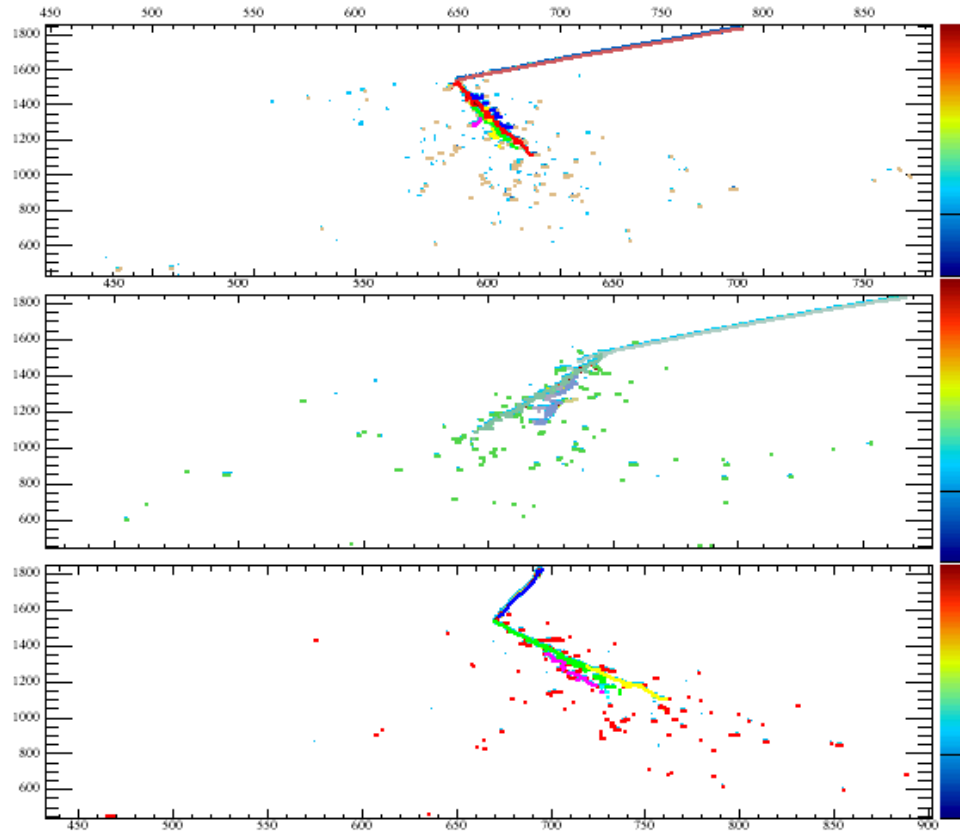
Pick a line and check at its end points

If a nearby line is found, check the angle between the slopes

If the angle between the slopes is $< 30^\circ$, merge the lines



An issue arises with the Hough lines in electron showers

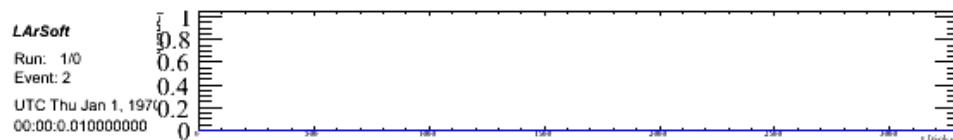


Hough line finder identifies lines in electron showers, need to veto

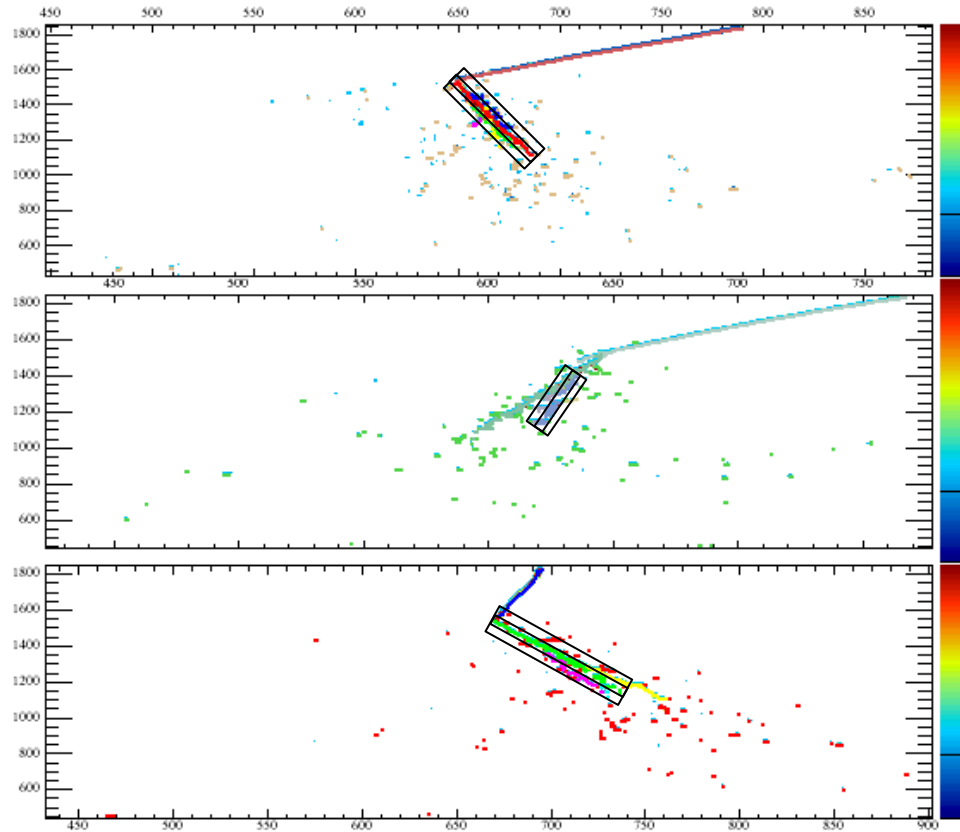
Currently examining area around lines and checking B/S, applying a cut

This cut is applied after the lines have been merged

Open to a better approach



An issue arises with the Hough lines in electron showers

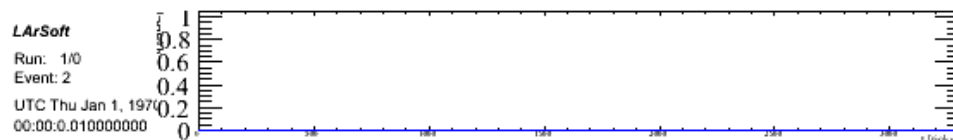


Hough line finder identifies lines in electron showers, need to veto

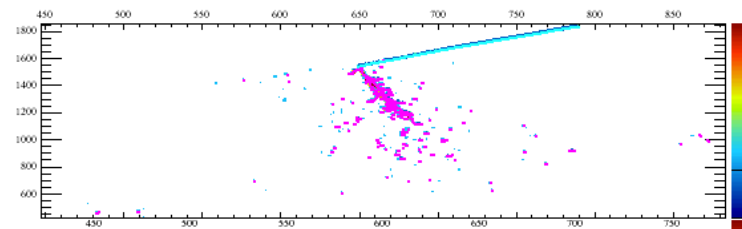
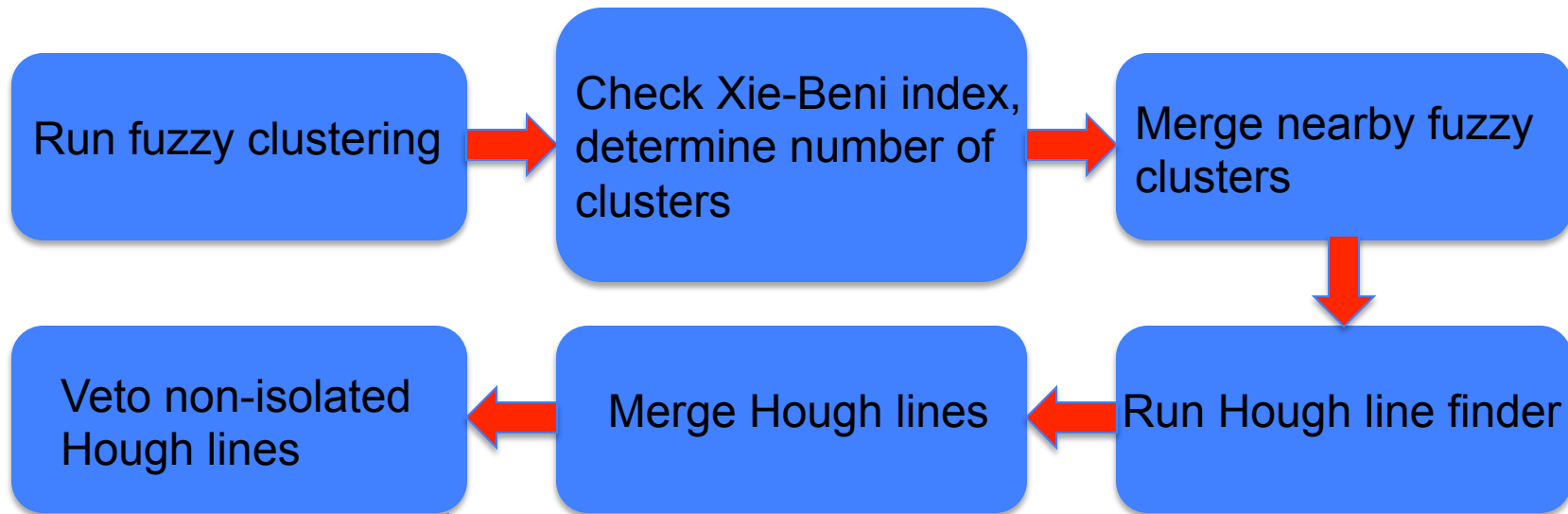
Currently examining area around lines and checking B/S, applying a cut

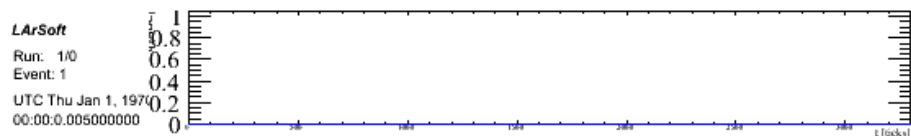
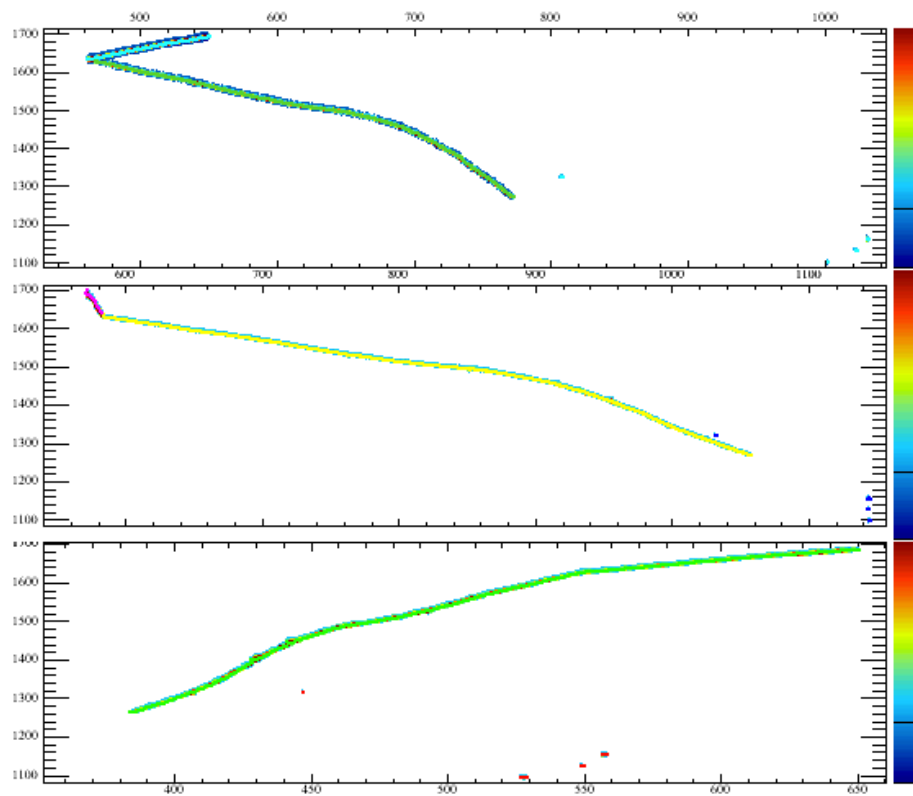
This cut is applied after the lines have been merged

Open to a better approach

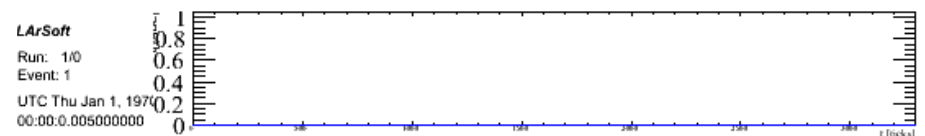
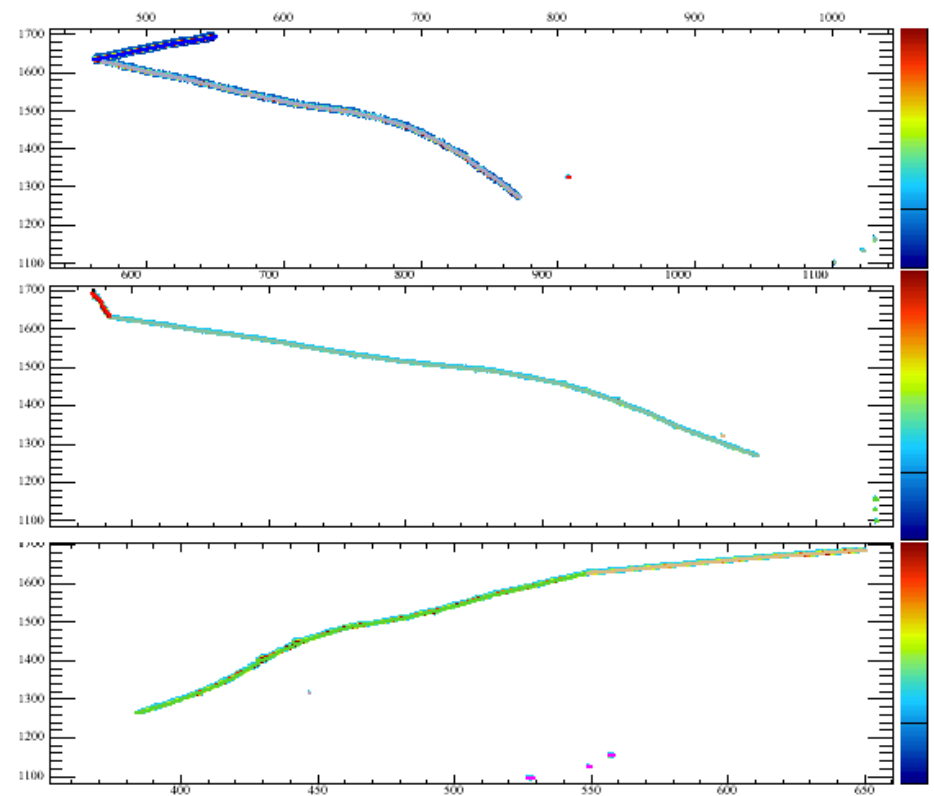


Overview of the algorithm

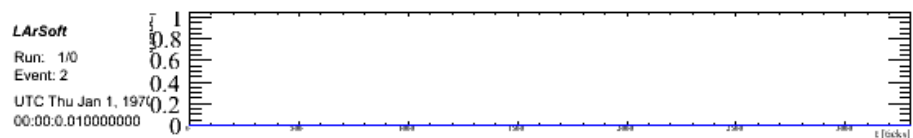
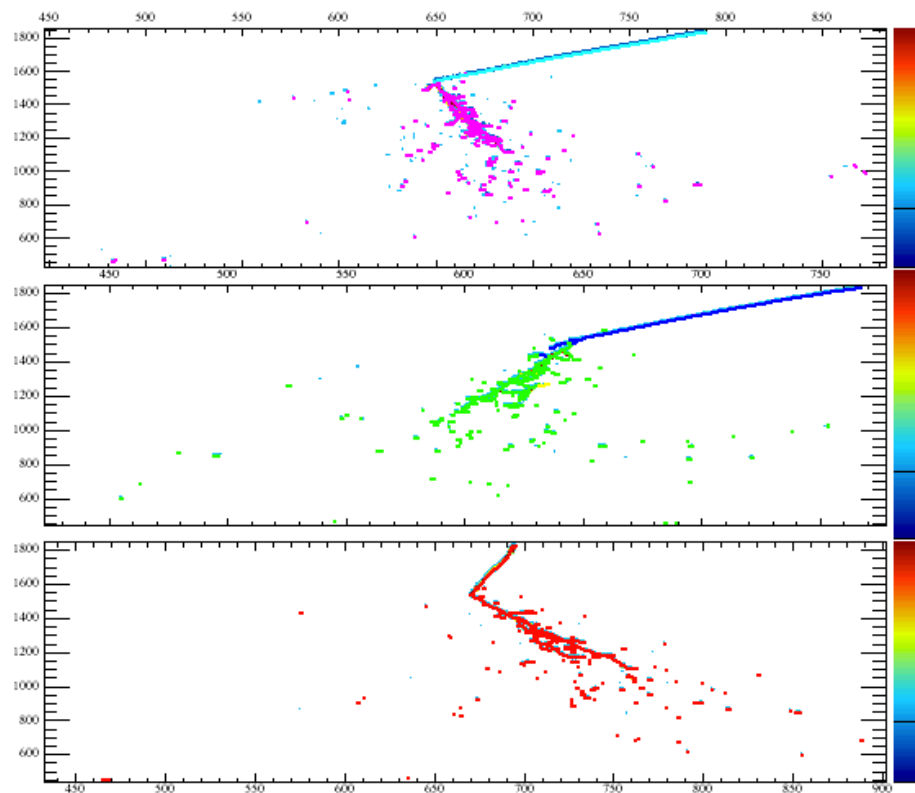




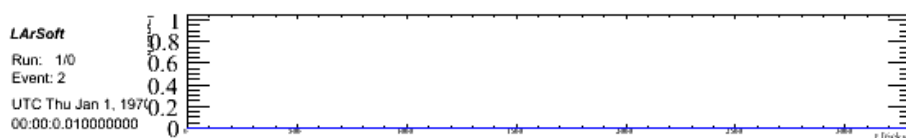
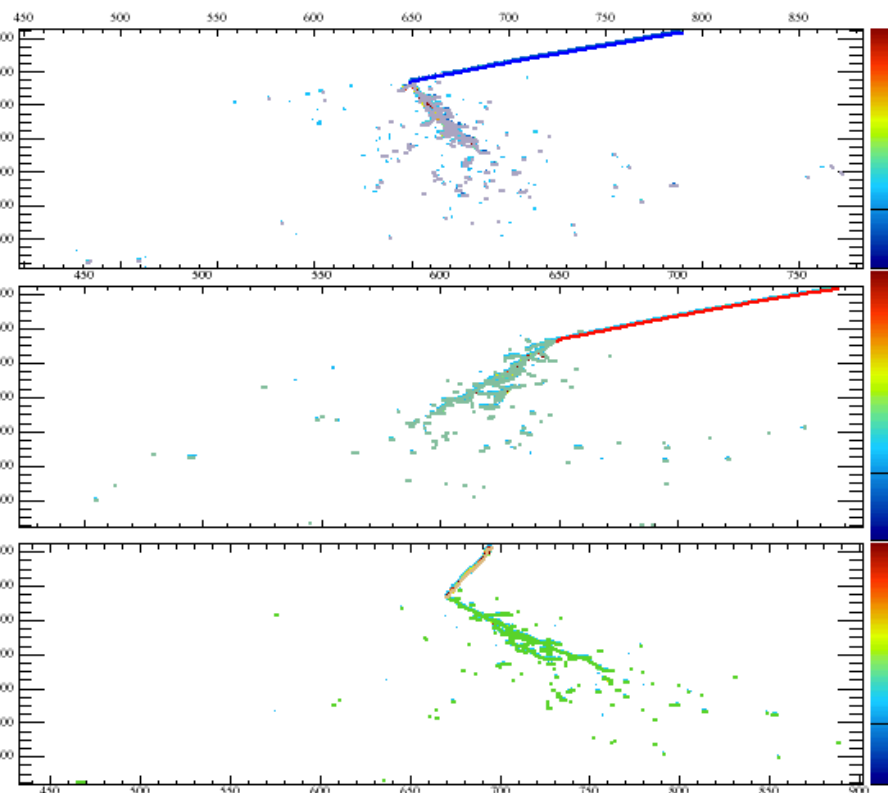
Fuzzy and Hough line clustering



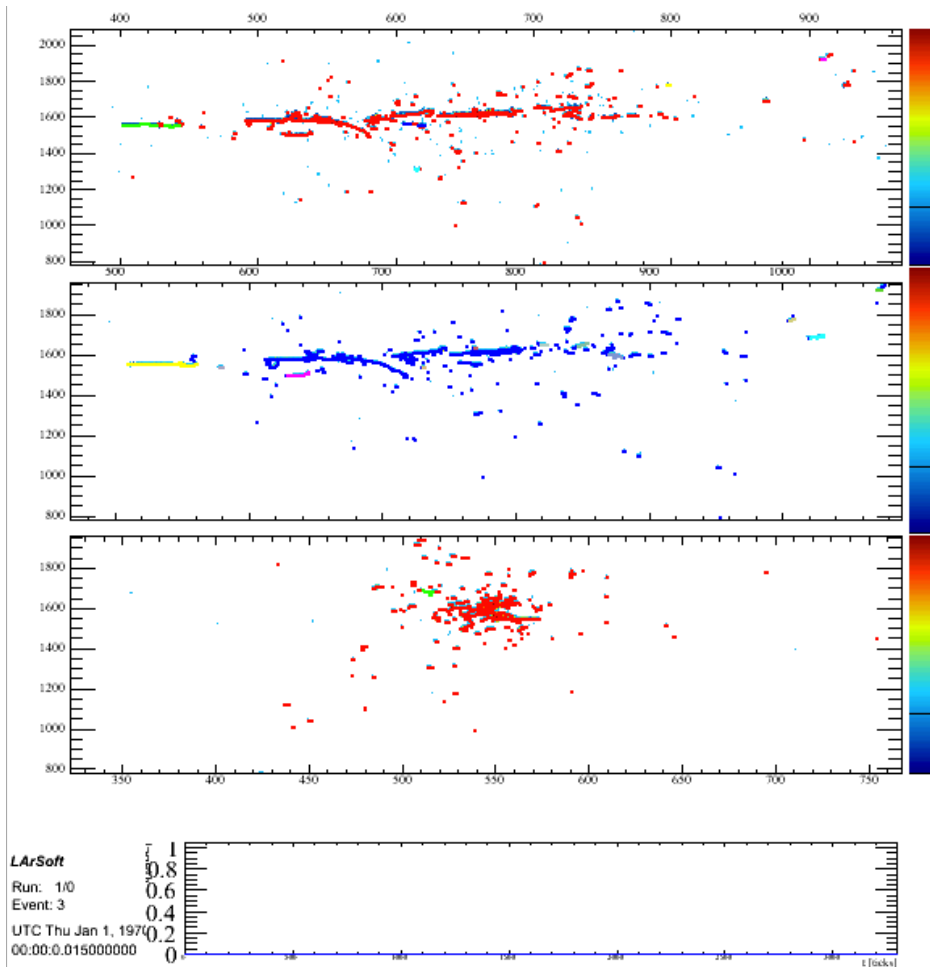
Cluster cheater



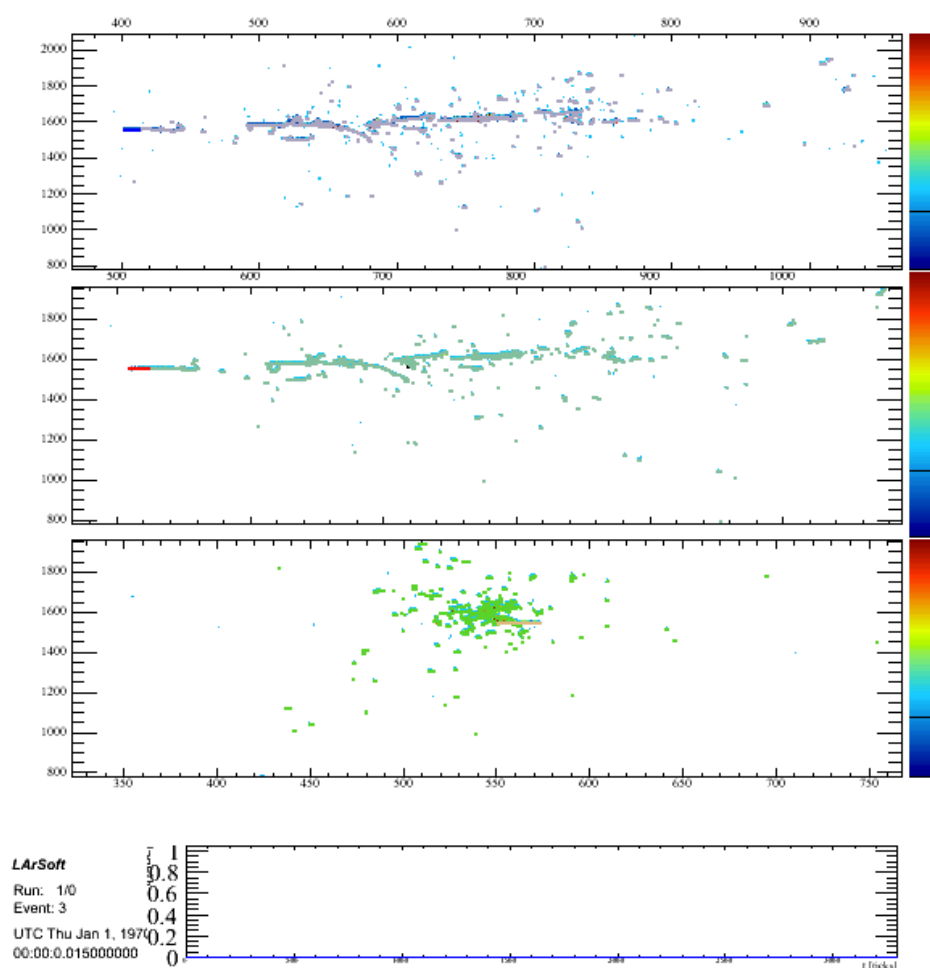
Fuzzy and Hough line clustering



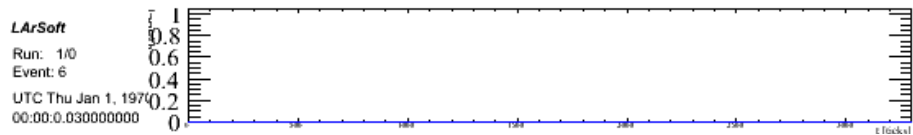
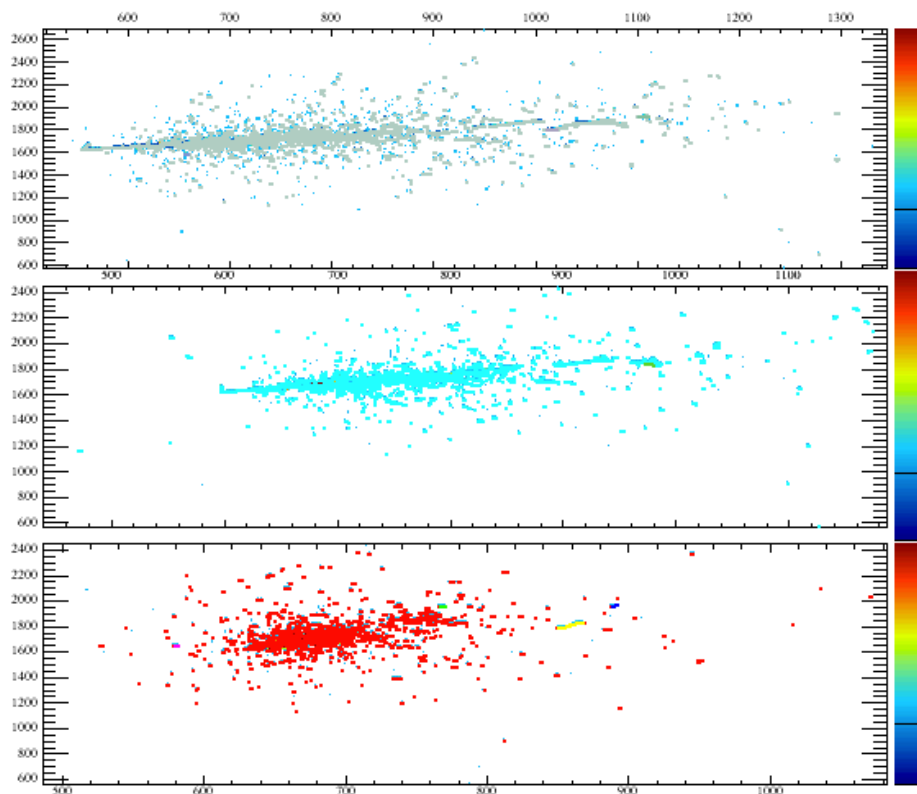
Cluster cheater



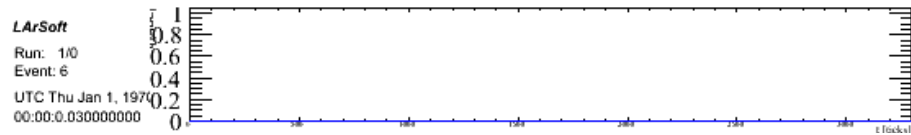
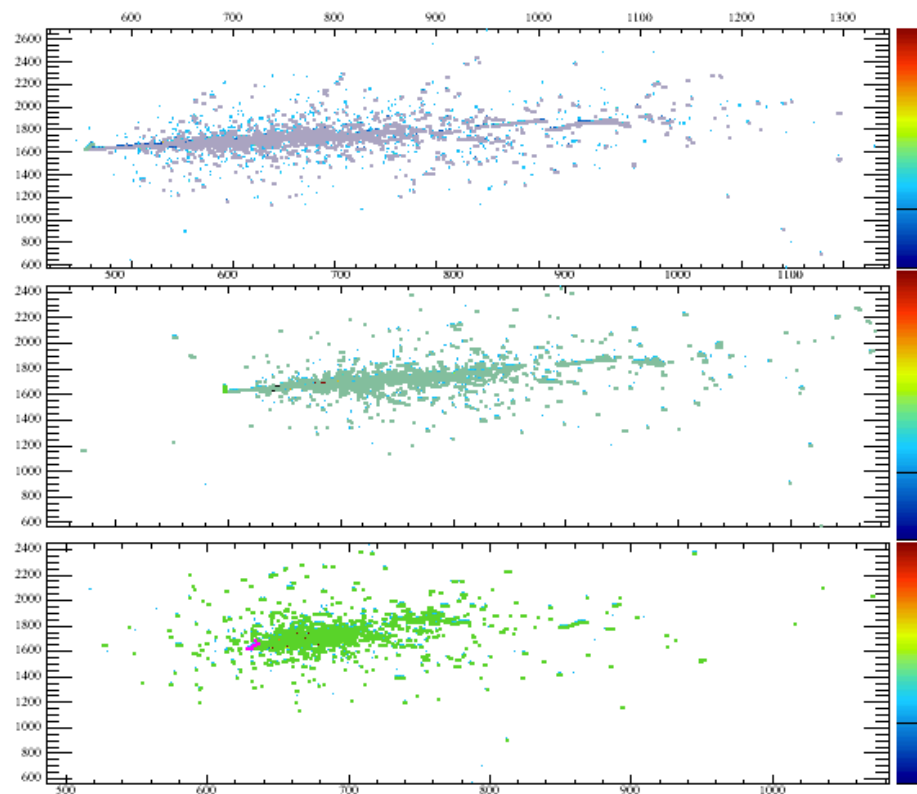
Fuzzy and Hough line clustering



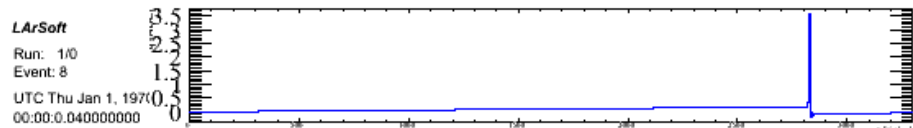
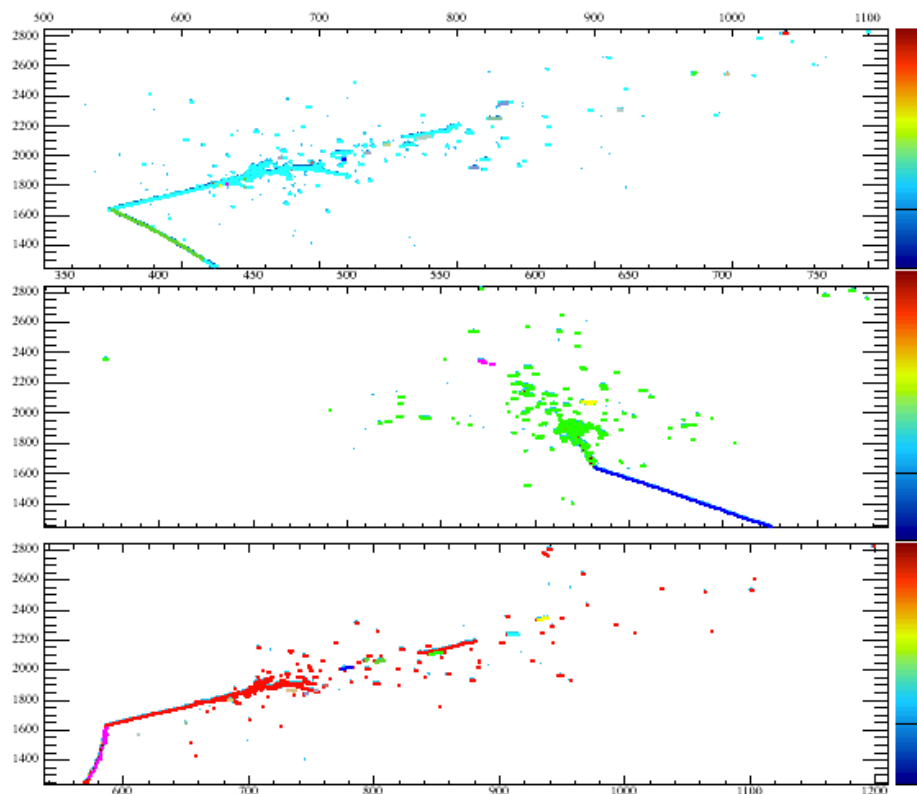
Cluster cheater



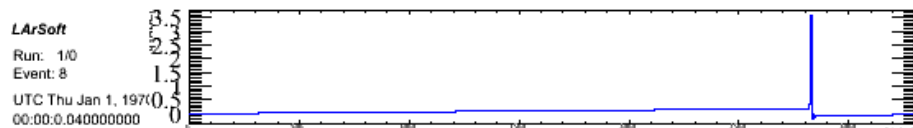
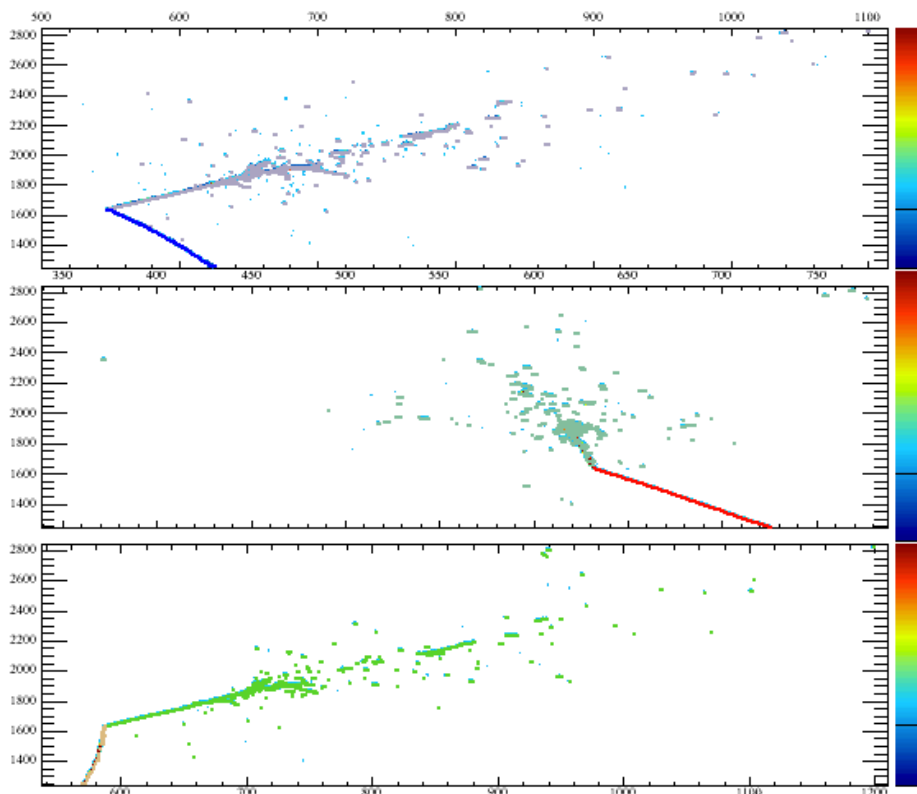
Fuzzy and Hough line clustering



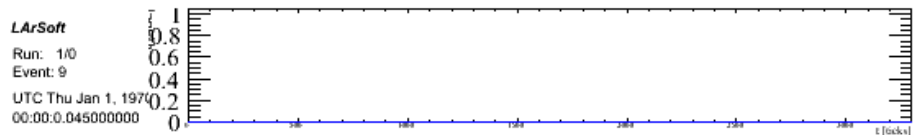
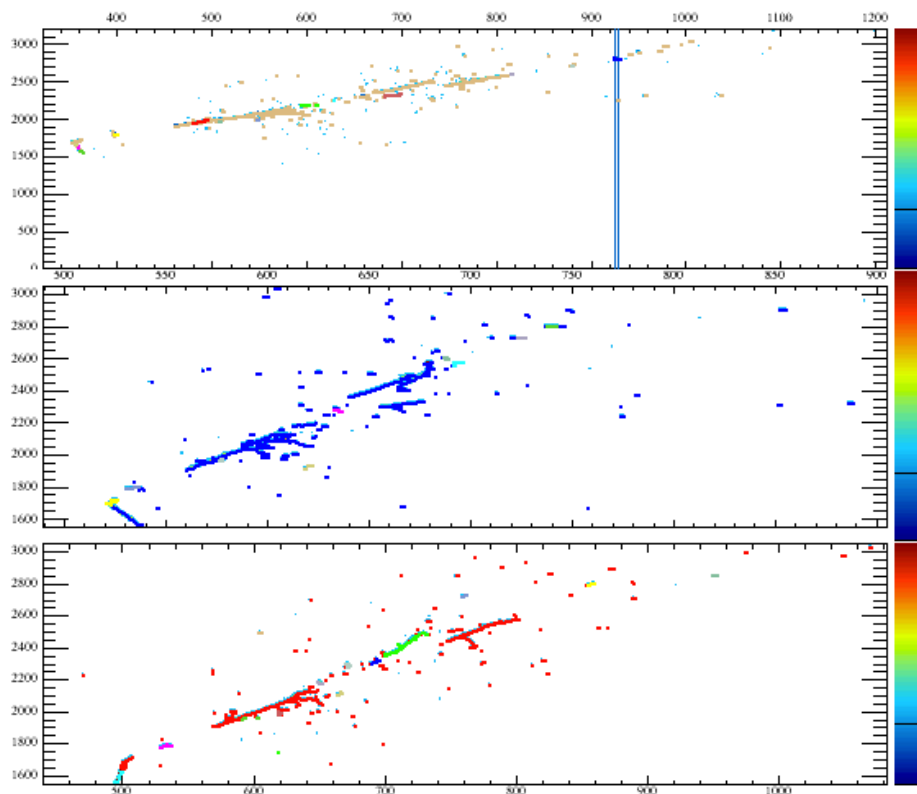
Cluster cheater



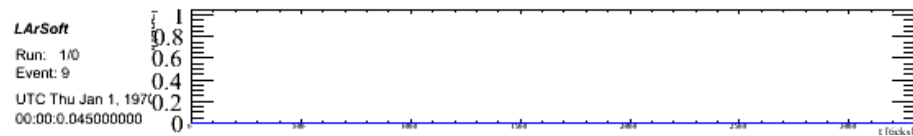
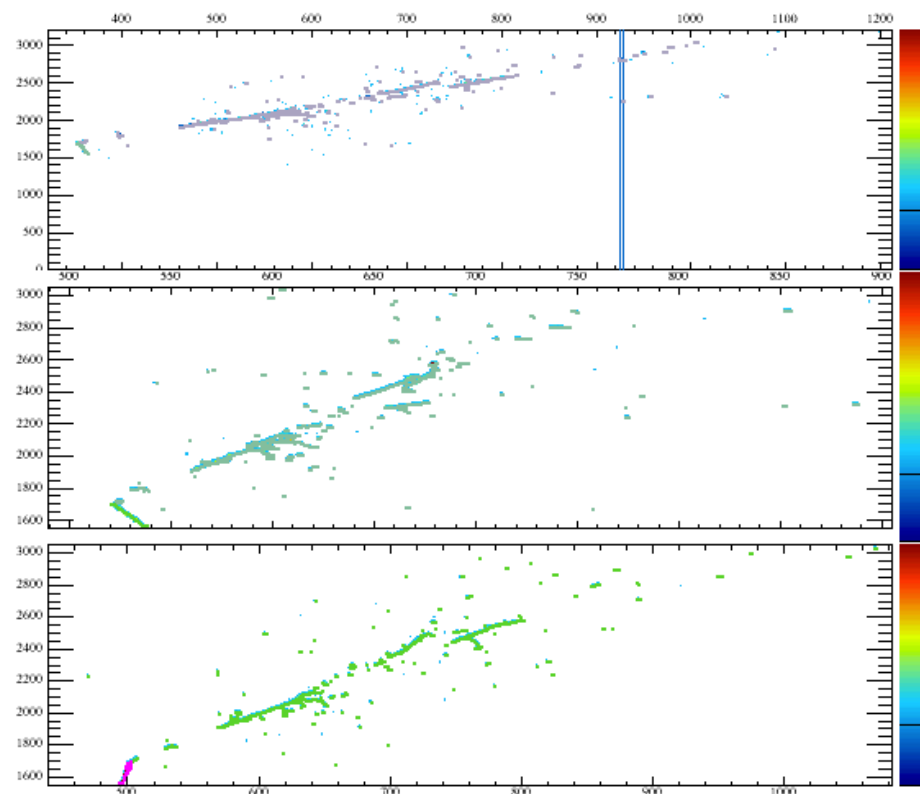
Fuzzy and Hough line clustering



Cluster cheater



Fuzzy and Hough line clustering



Cluster cheater

Next steps

- Finish optimization
- Examine more events (only CCQE so far) to improve robustness
- Finalize the needed modules