Directional-iDBSCAN

a proposal to CYGNO

Igor Pains

Igor Abritta and Rafael A Nobrega

Last presentation

- An implementation of the iDBSCAN and iDDBSCAN using the *sample_weight* parameter was made aiming to reduce the duration of the algorithm while using the 3D.

- The speed boost was noticeable, especially when the event has several bright tracks.

- After all these updates, we must decide which way to go with the algorithm.
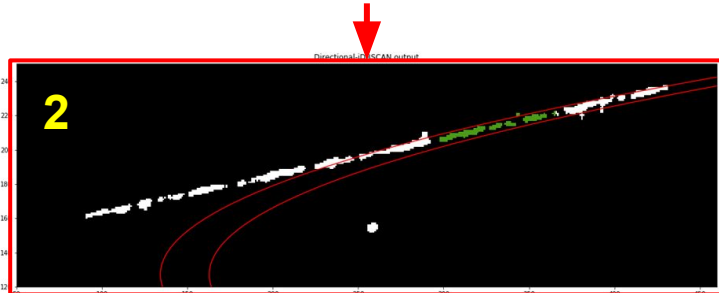
The evolution of DDBSCAN

- The DDBSCAN was conceived primarily as an update of the DBSCAN, aiming the possibility to perform a directional clusterization capable of following long tracks.

- The first implementation of the DDBSCAN was made combining the DBSCAN (seeding) and the RANSAC regressor with a polynomial equation as its base estimator.

The evolution of DDBSCAN

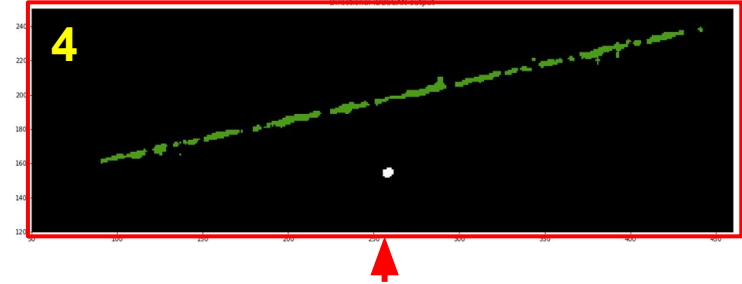**Output of the iDDBSCAN first step**
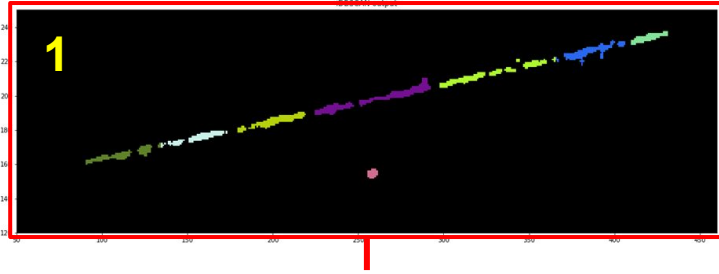
Directionality begins at the highest density cluster and find a first fit model using just that cluster.
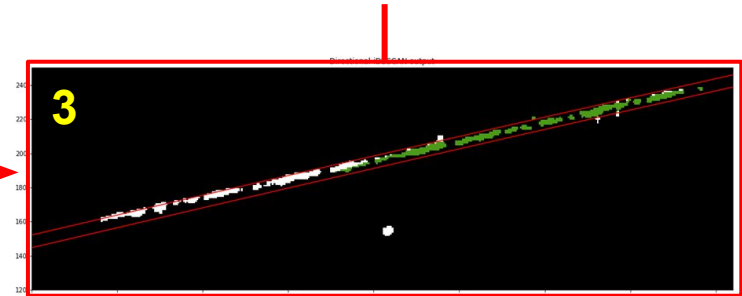

Cluster found on the event 15 of the run 2065.
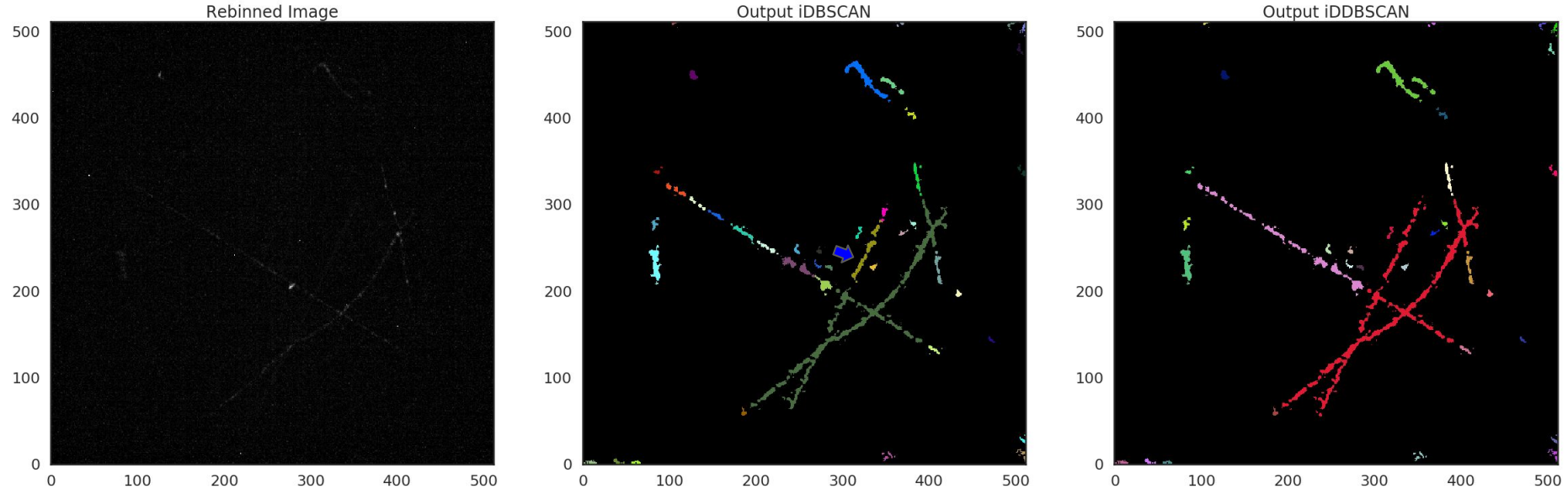
Final cluster

Then, search for new points using the fit, improve the fit with the new information and repeat until nothing else is found.

4

The evolution of DDBSCAN

- After more images from new runs were analyzed (2097, 2098, 2317, 2320), it was noted that most larger clusters with a polynomial behavior could be explained through a polynomial of degree 3.

- Events with overlapped tracks couldn't be properly clustered because the algorithm would put all those points together after the directional search ended.
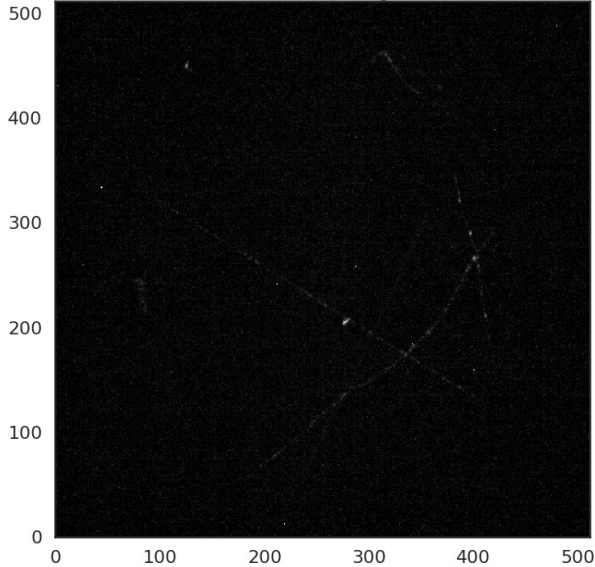
The evolution of DDBSCAN

This is the event 24 of the run 2320, although a good polynomial model was found starting from the cluster marked in blue, the surrounding points would be added to the final cluster because of the overlapped tracks until this version of the DDBSCAN.
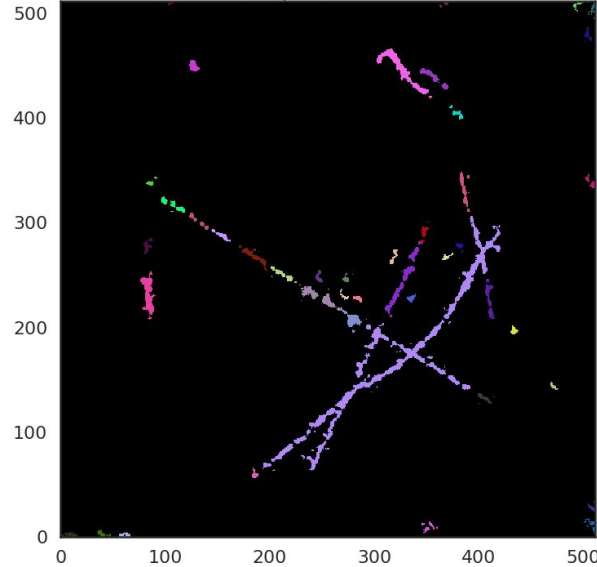
The evolution of DDBSCAN

- An update was implemented to separate overlapped tracks. That was done by removing the extra step that would classify the points very close to the cluster as part of it regardless of the fit.

- The algorithm was successful in separating overlapped tracks after this update, but in some cases there were *'halos'* close to some clusters.
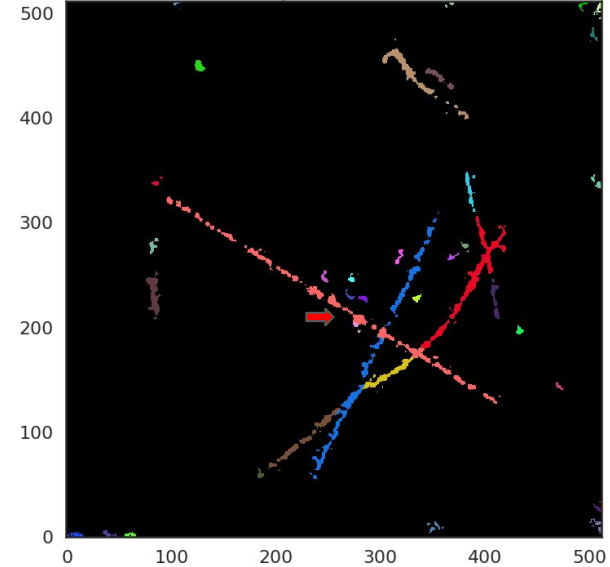
The evolution of DDBSCAN
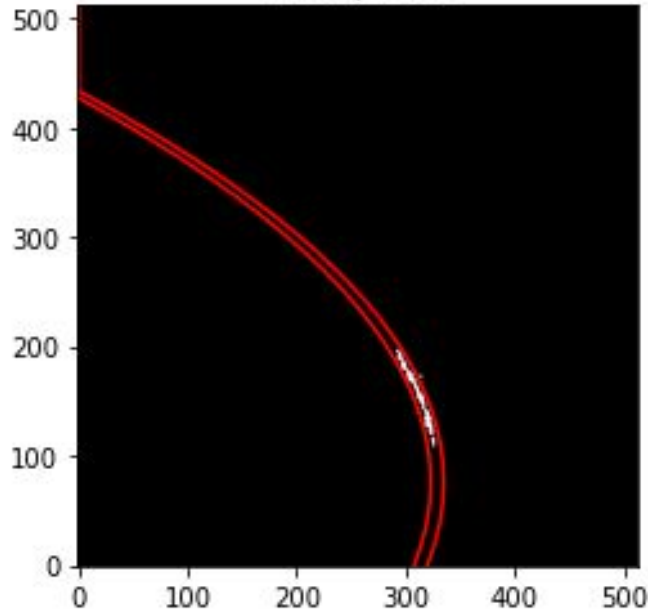

Run 2320 - Event 24

The evolution of DDBSCAN

- The last major update in the DDBSCAN was the modification of the RANSAC to keep the fits with a constant width.

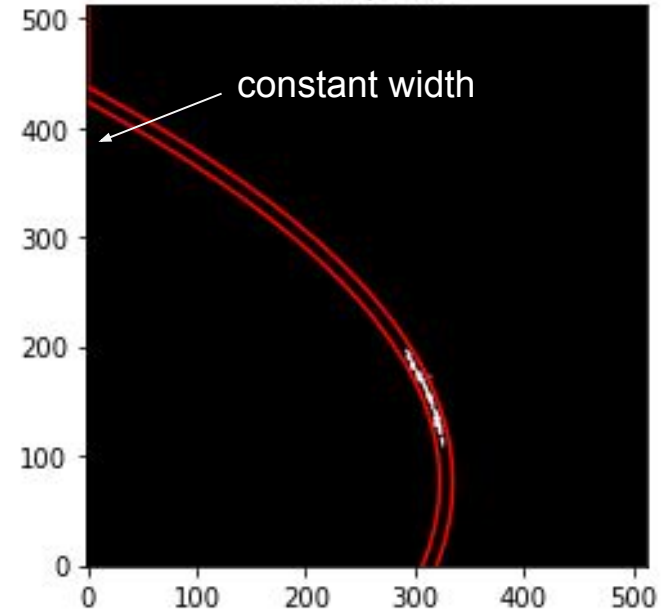- This new RANSAC avoided possible mistakes with the fits and maintained the good results that its older version had.

The evolution of DDBSCAN

Attempt one

Here is the difference of the RANSAC fit in a curve with and without constant width.
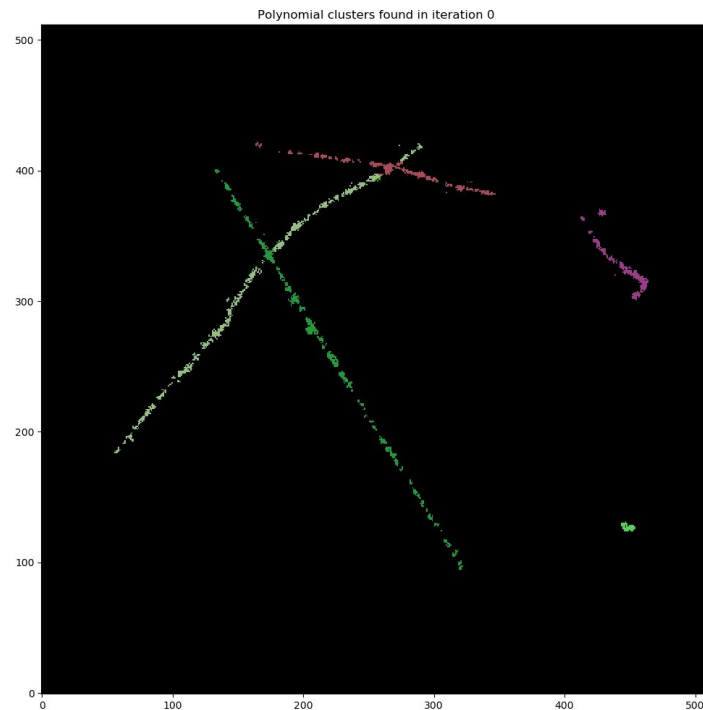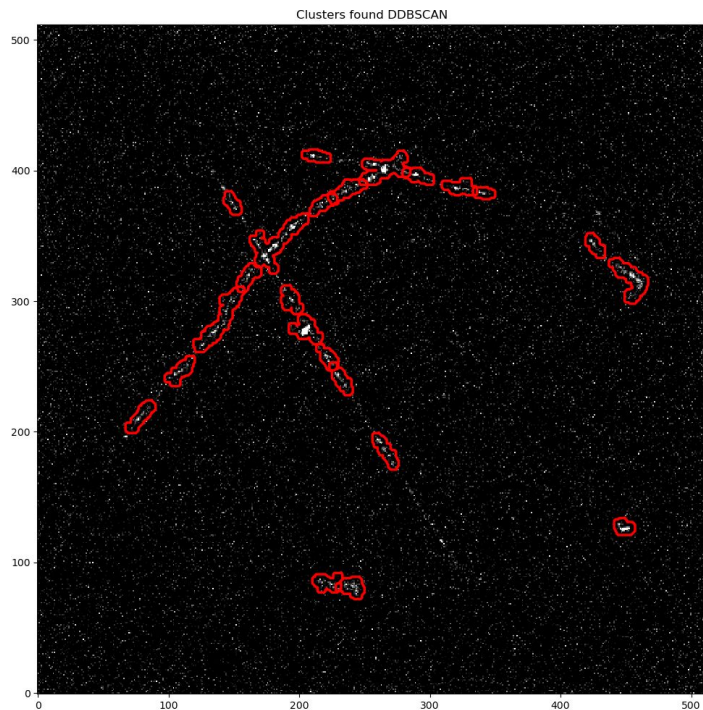

Attempt two

constant width

The evolution of DDBSCAN

- After these updates, some attempts were made to speed up the algorithm like using cython and the '*sample_weight*' parameter.

- The attempt to use cython was not successful, while using the '*sample_weight*' parameter was.

The evolution of DDBSCAN

Run 2320 - Event 24

Next steps

- Tune the DDBSCAN parameters with quantitative results.
- Propose an article about DDBSCAN applied to CYGNO.
  - These two proposals would need some kind of simulation that has long tracks to evaluate the results of the DDBSCAN.

- Look for faster implementations of the RANSAC to speed up the algorithm even more.