

Directional-iDBSCAN

first look at LIME data

Igor Pains

Igor Abritta and Rafael A Nobrega

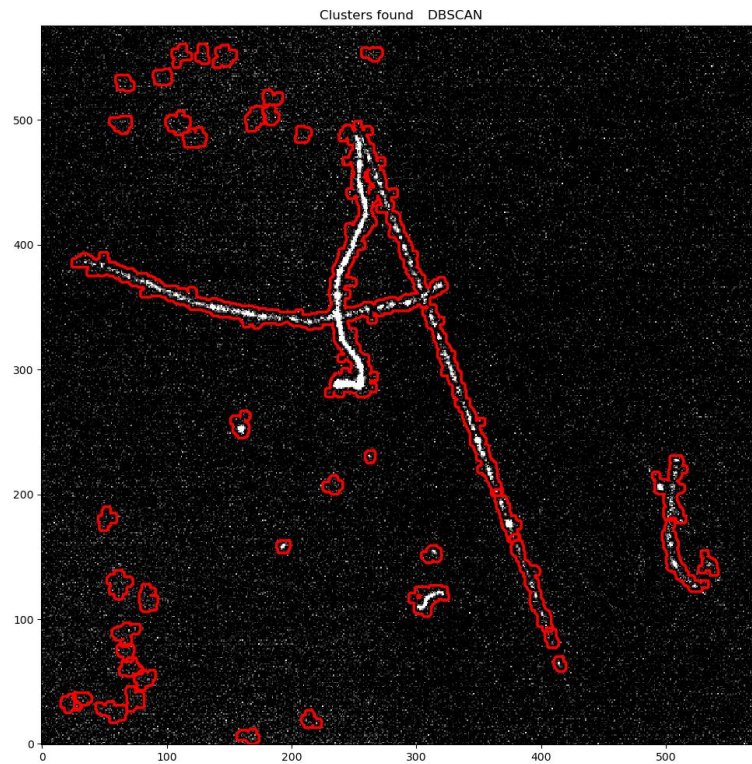
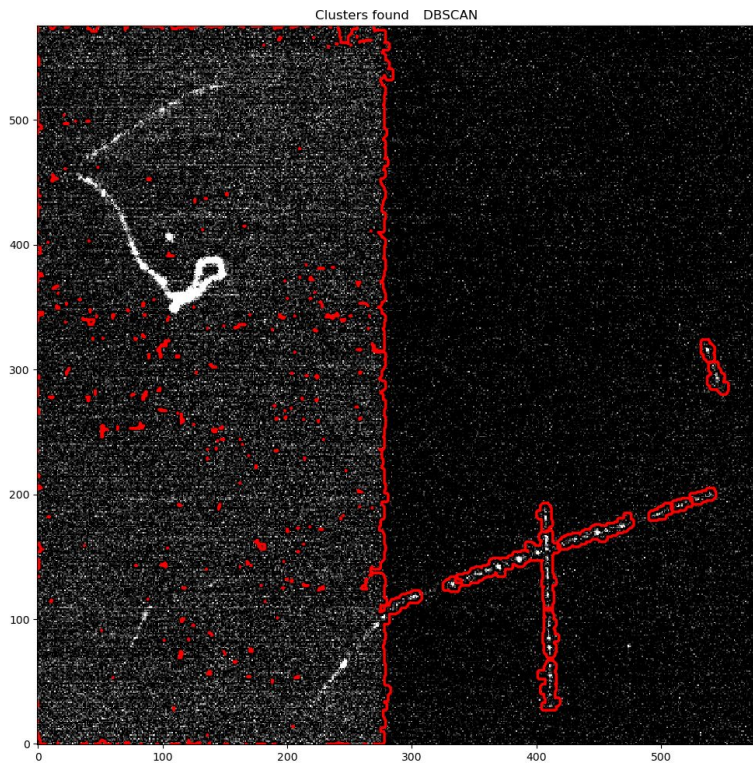
Last presentation

- A quantitative analysis inspired by the iDBSCAN article (^{55}Fe based analysis) was done using the iDDBSCAN algorithm.
 - DOI: 10.1088/1748-0221/15/12/T12003
 - Selected runs: 2054 (EN), 2156 (NRAD) and 2163 (ER)
 - Link of the presentation: [here](#)
- The iDDBSCAN was able to reduce background noise in the low energy region (by reconstructing short, medium and long tracks).
- It was agreed in that meeting that the next step would be to analyse the algorithm with LIME data.

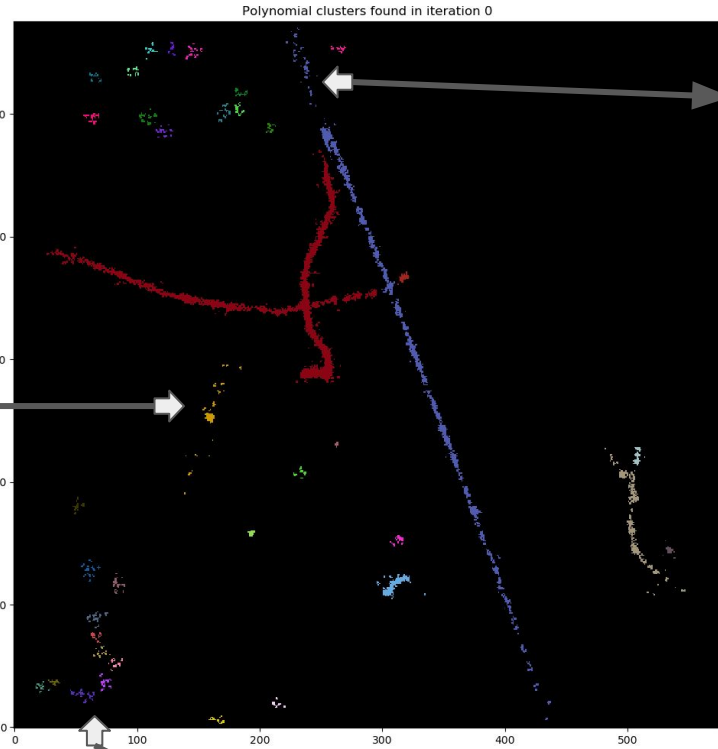
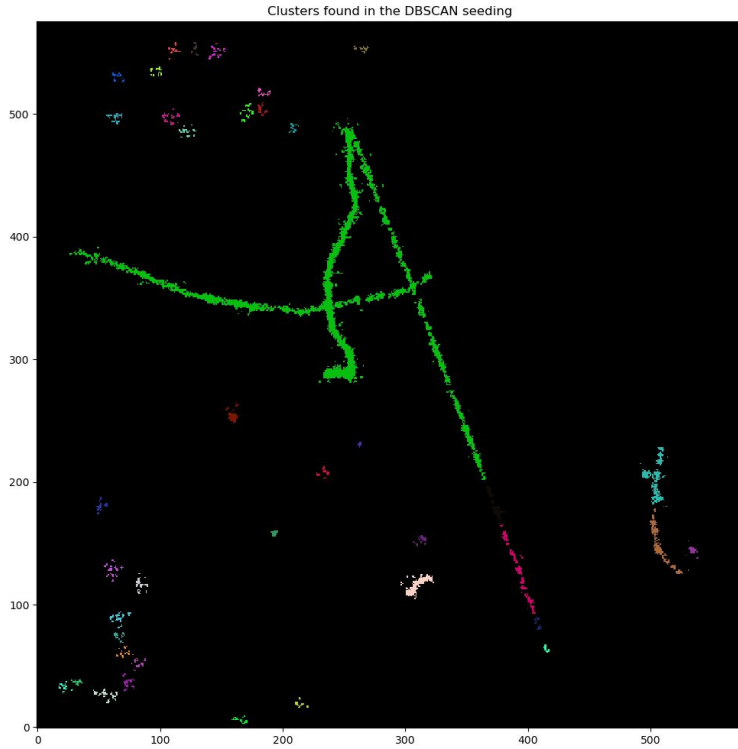
LIME data

- The first look at LIME data was done with the same parameters used in LEMOn data.
- The chosen runs:
 - [3737-3791]: Ambe runs (focus on 3790).
 - [3792-3794]: Cosmics runs soon after source off (focus on 3793).
 - This set of data was affected by the light entering the camera corners (sigma used was 1.6).

Run 3790



Run 3790 - Event 712

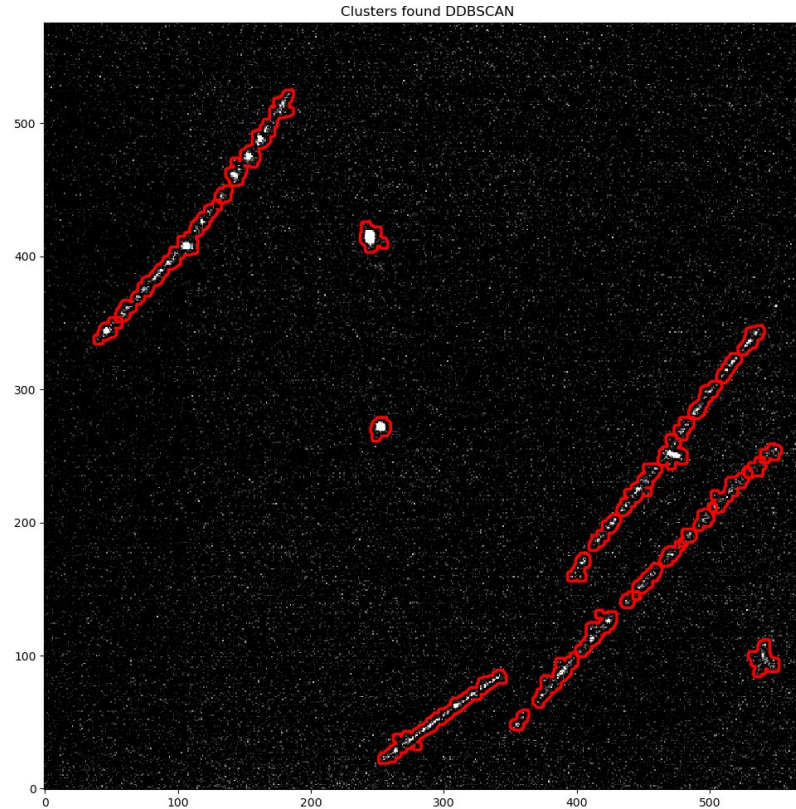


Noise added to a cosmic cluster.

Noise added to a circular cluster. (should be avoided)

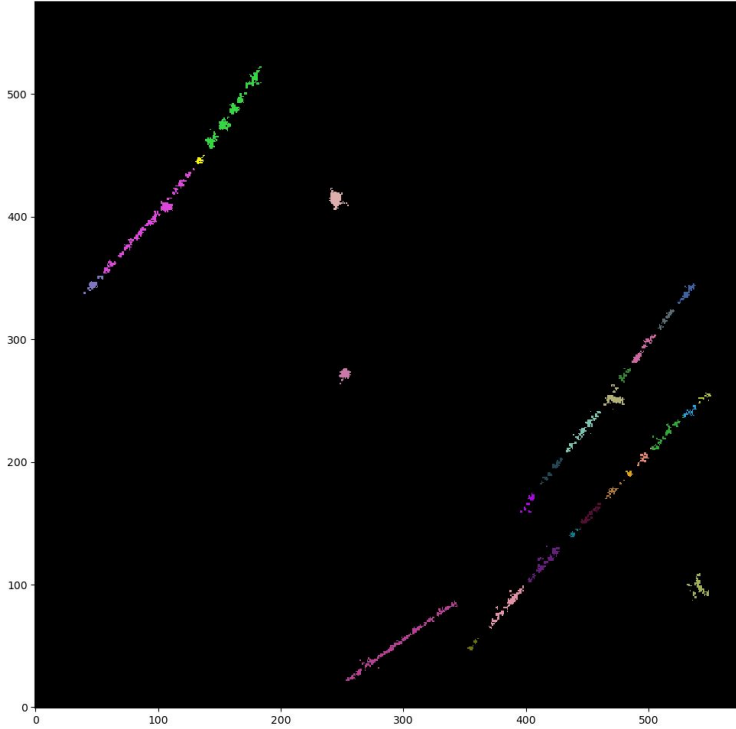
Noise clusters

Run 3790 - Event 7

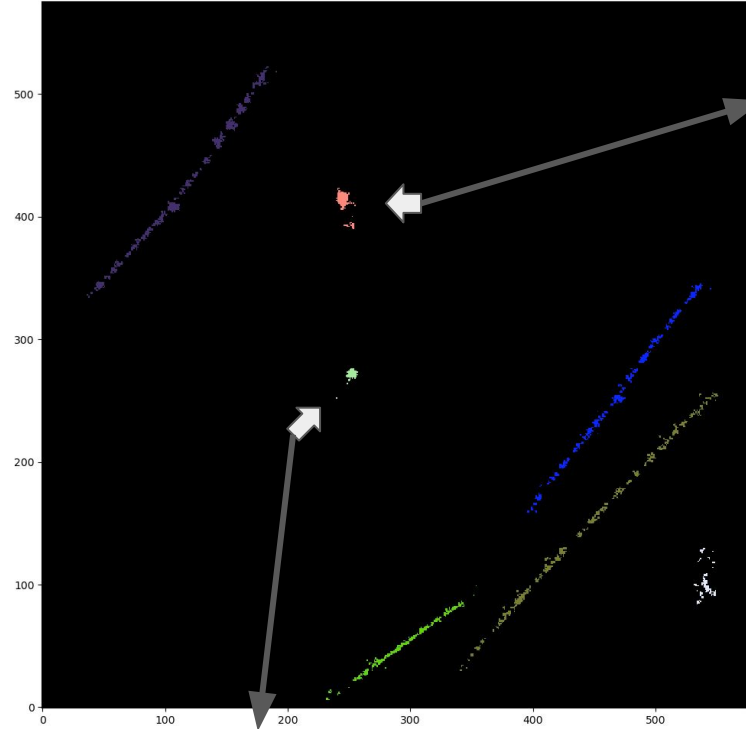


Run 3790 - Event 7

Clusters found in the DBSCAN seeding



Polynomial clusters found in iteration 0



Noise added to these clusters.

Noise added to a circular cluster. (should be avoided)

Solutions

- Use the `expand_noncore` parameter as *False* might solve the noise problem.
 - It is advisable to use it as *True* for a better directional clustering, especially when the tracks are faint.
- Tune the parameters to the LIME data.
 - `dir_minsamples` may be raised so small clusters won't be tested in the DBSCAN seeding.
- Improve the DBSCAN seeding selection.

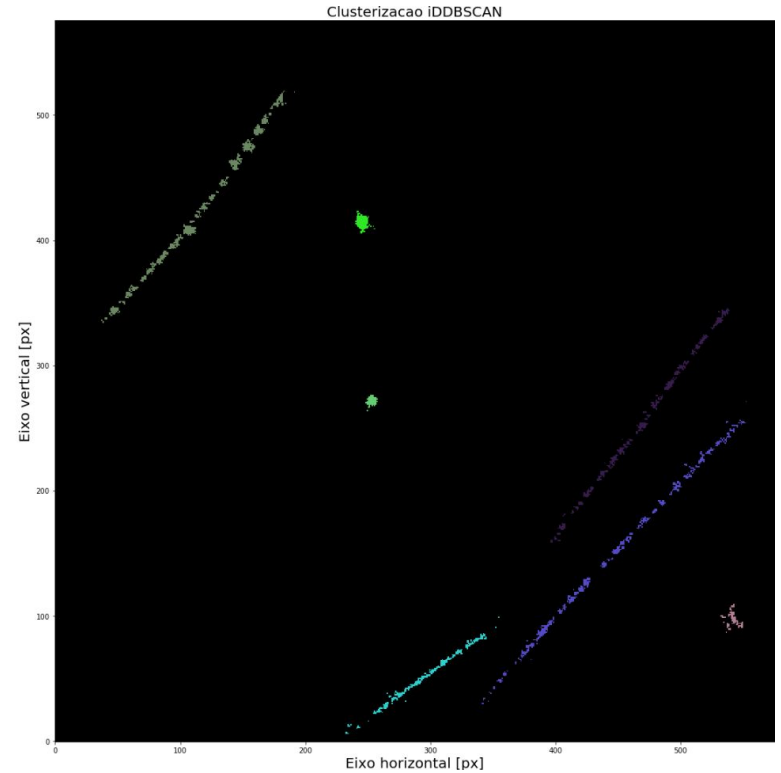
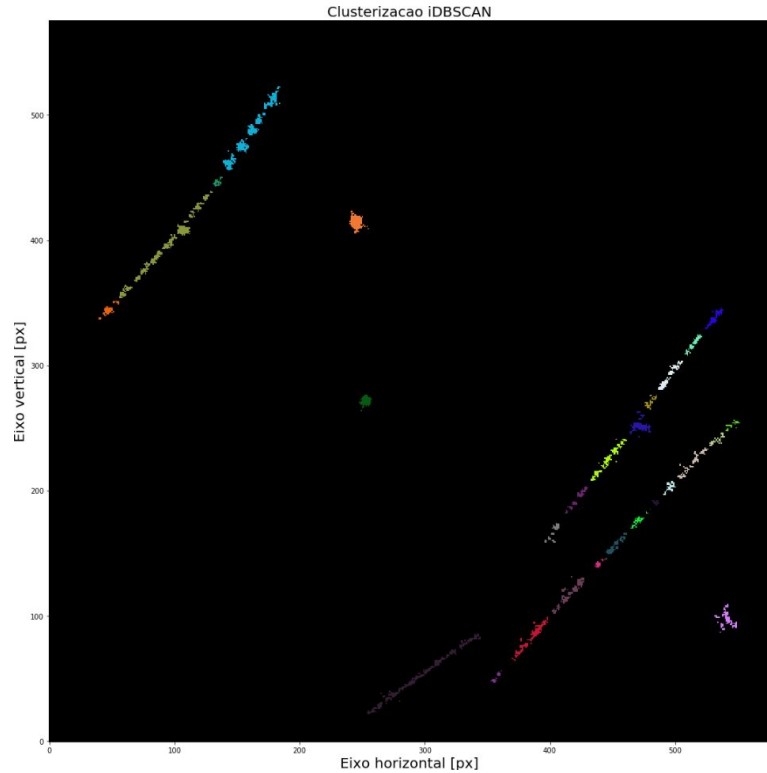
DBSCAN seeding

- The clusters found by the iDBSCAN are seeds.
 - Each seed is tested with the RANSAC if it has more points than the `dir_minsamples` parameter.
 - The seeds with a good RANSAC model (the accuracy is compared with the `dir_min_accuracy` parameter) have one of their core points saved for the directional clustering.
- The solution is to change the RANSAC to the one used before Emanuele's implementation in "*lime21*". (7 months ago)

RANSAC - DBSCAN seeding

- RANSAC (scikit-learn library):
 - `min_samples` = 50% of the set.
 - `residual_threshold` = $\text{median}(y - \text{median}(y))$.
 - The `residual_threshold` should be tested (if it is equal to 0 the function raises an error).
- It has difficulties to select horizontal/vertical cosmic clusters and is good to reject circular clusters.
 - The `residual_threshold` calculated by the function is pretty small on horizontal and circular tracks.
- Changes needed to select horizontal/vertical tracks while reject circular tracks:
 - Set a minimum value of residual threshold.
 - Rotate the clusters that did not have good RANSAC fits and test again.

Run 3790 - Event 7



Conclusions

- The two modifications worked well to reject circular tracks on the DBSCAN seeding.
- The second modification (rotate clusters) is better at selecting horizontal/vertical tracks.
- Performance at highly curved tracks still to be tested.
 - Possible modification is to use RANSAC with polynomial regression and keep using the standard `residual_threshold`.

Next steps

- Do speed tests to choose the fastest modification to be implemented.
- Tune the iDDBSCAN parameters in order to do a quantitative analysis.
 - Which dataset should be used?
- Any ideas or suggestions?