Analysis Cygnus - 01

10-04-2019

Objetivos e Última Apresentação

- **Objetivo geral**
  - Ajudar na análise e processamento dos dados do experimento Cygnus.
- **Objetivos específicos**
  - Desenvolver um algoritmo capaz de executar uma rápida e eficiente clusterização nas imagens coletadas;
  - Análise e criação das variáveis que serão usadas posteriormente na classificação dos sinais;
  - Desenvolver algoritmo para classificação dos eventos.
    - A ideia é usar o KDE + Likelihood nessa tarefa.

- **Última apresentação**
  - Foi apresentado o momento atual do trabalho;
  - As tarefas a serem executadas.

Progresso das últimas semanas

27 - 03 - 19

- Caracterizando os traços ‘curvos’;
- Avaliando a eficiência do i2DBSCAN;
- Escrevendo artigo sobre o I2DBSCAN para o congresso IBPRIA 2019
  - Faltando só os resultados.

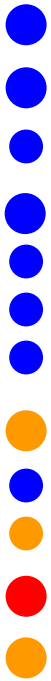
https://www.overleaf.com/read/fmtrrkxvdjyr

10 - 04 - 19

- As tarefas agendadas para essas duas semanas não foram terminadas e estão em fase de execução.

Artigo IbPRIA - Novo deadline 30 de Abril

- ❑ Abstract
- ❑ Introduction
  - ❑ Clusterization methods overview
- ❑ Methodology
  - ❑ Cygno overview
  - ❑ Data set
  - ❑ Pedestal subtraction and steps
- ❑ Development
  - ❑ i2DBSCAN
  - ❑ Evaluation
- ❑ Results
- ❑ Conclusions



- Not started
- In process
- First version
- Done

Sobre a avaliação do i2DBSCAN

Index evaluation

❑ Davies-Bouldin

Zero is the lowest possible score. Values closer to zero indicate a better partition.

❑ Calinski-Harabaz

The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster

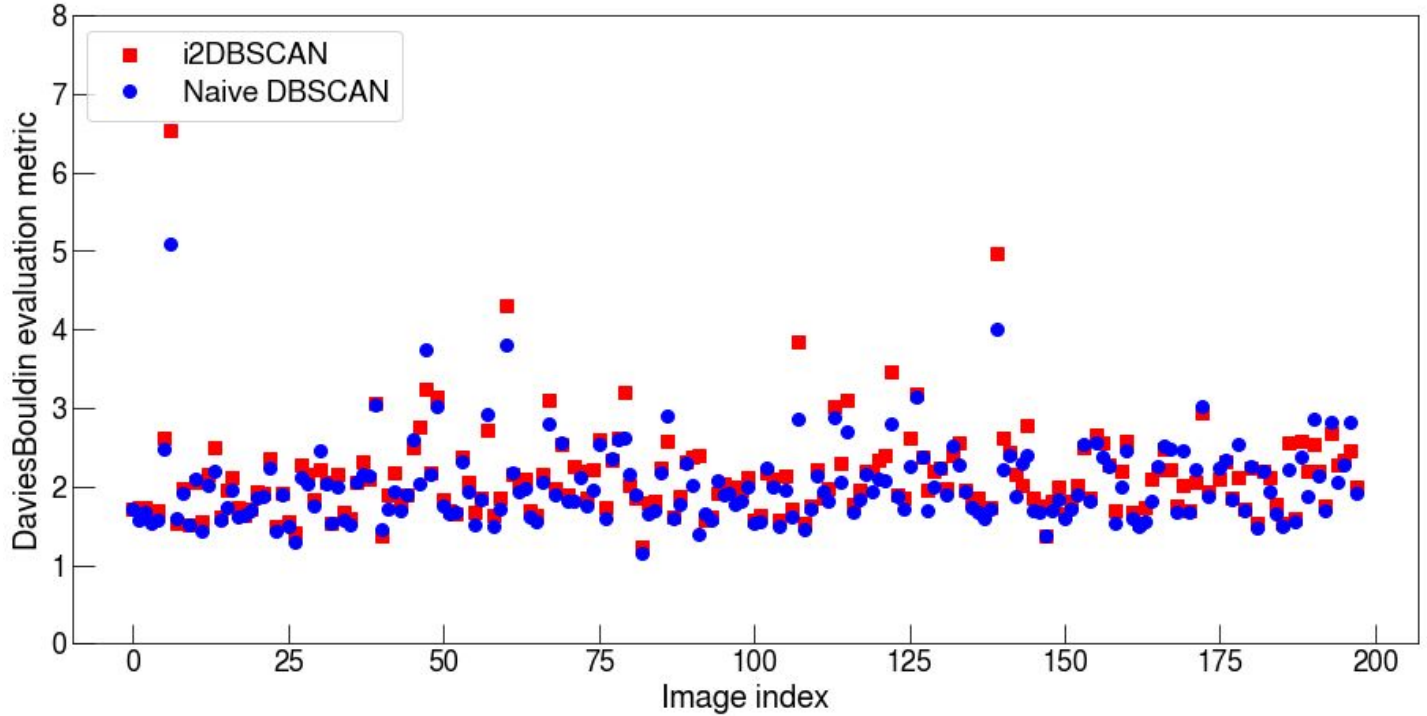
❑ Silhouette

The score is bounded between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters. The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster.

Index evaluation

❑ Davies-Bouldin

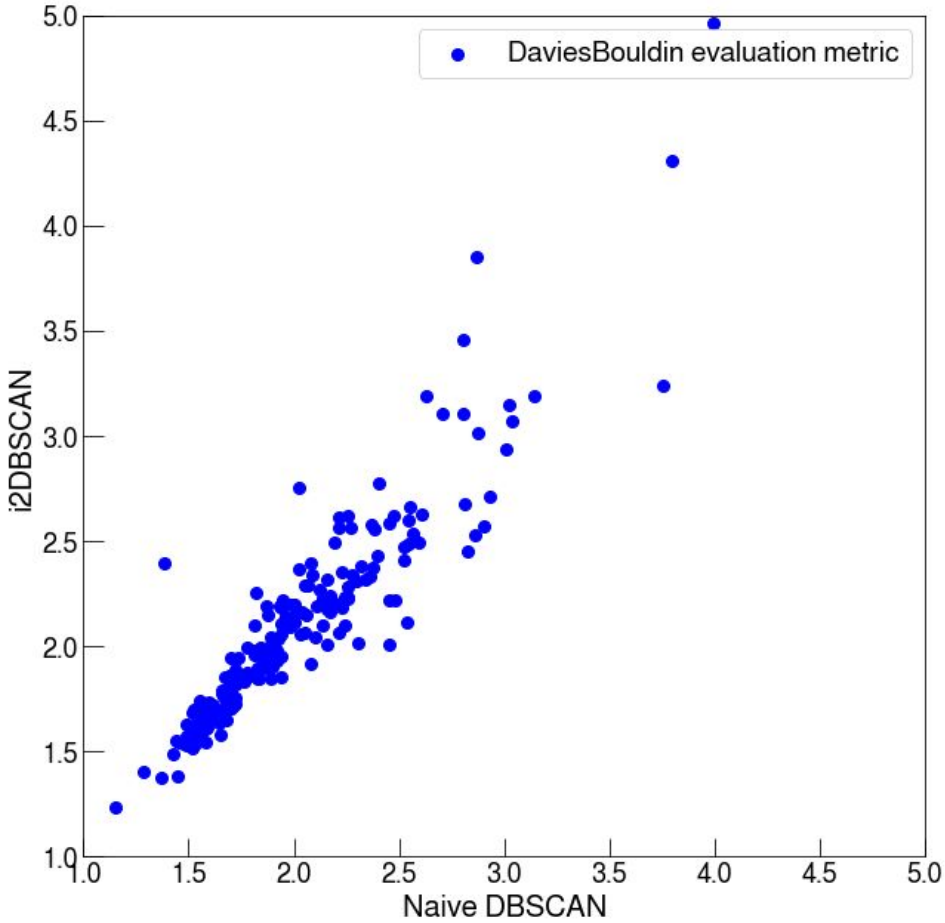
Zero is the lowest possible score. Values closer to zero indicate a better partition.



Index evaluation

Davies-Bouldin

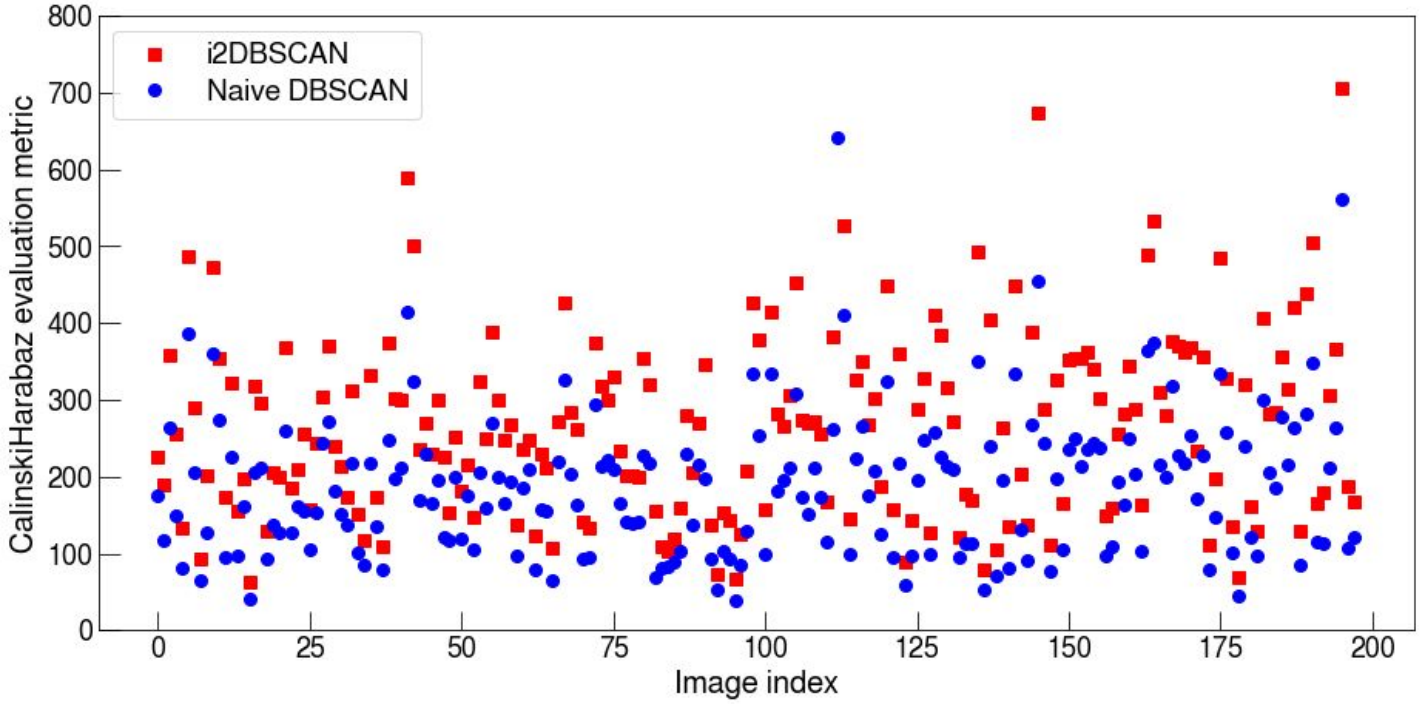
Zero is the lowest possible score.
Values **closer to zero** indicate a **better partition**.



Index evaluation

❑ Calinski-Harabaz

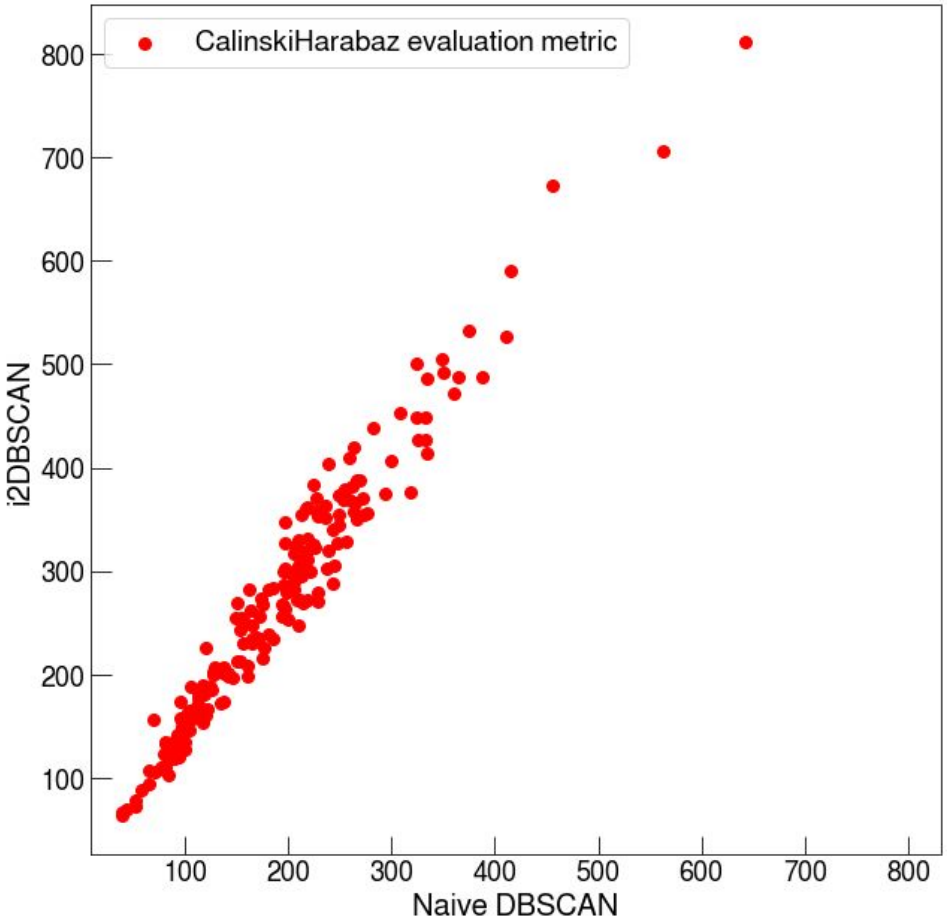
The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster



Index evaluation

Calinski-Harabaz

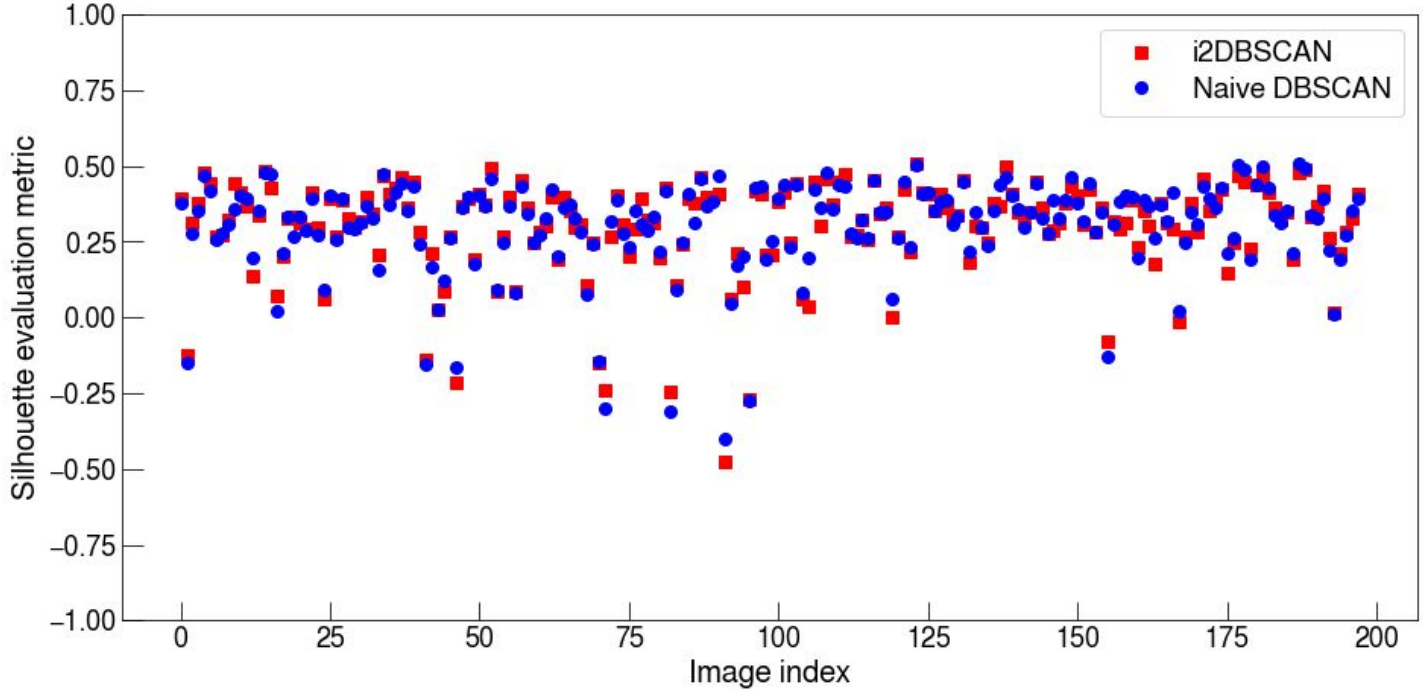
The score is **higher** when clusters are **dense and well separated**, which relates to a standard concept of a cluster.



Index evaluation

❑ Silhouette

The score is bounded between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters. The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster.



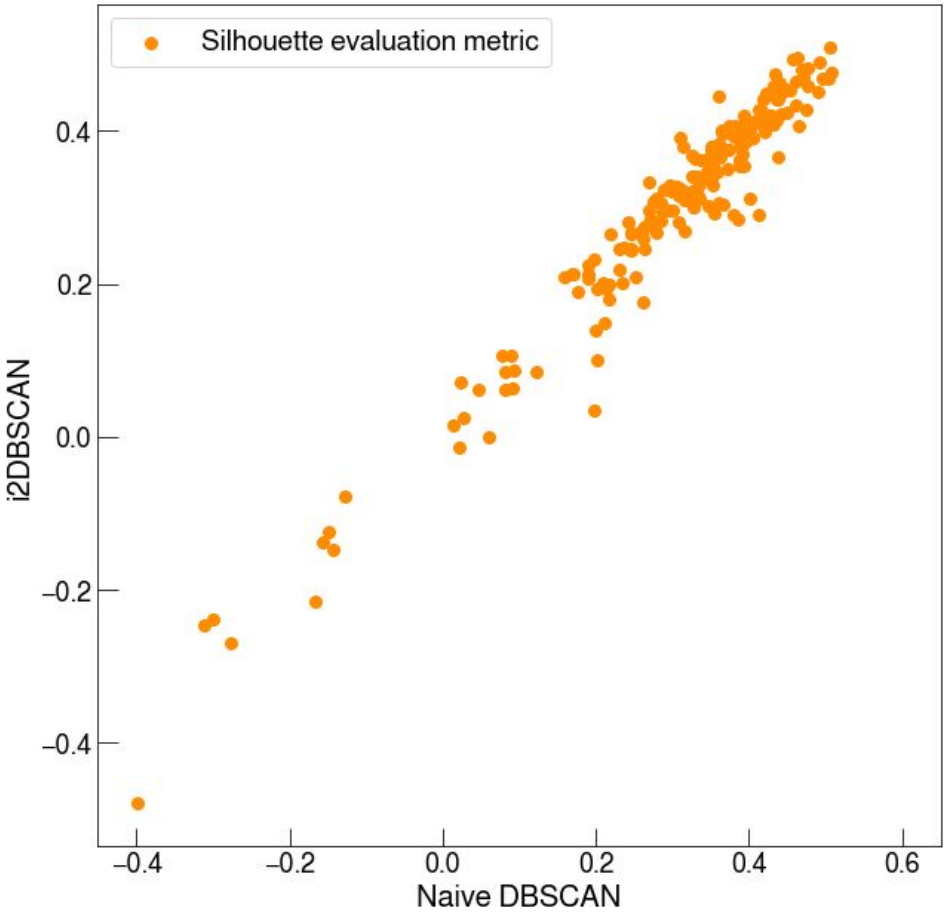
Index evaluation

Silhouette

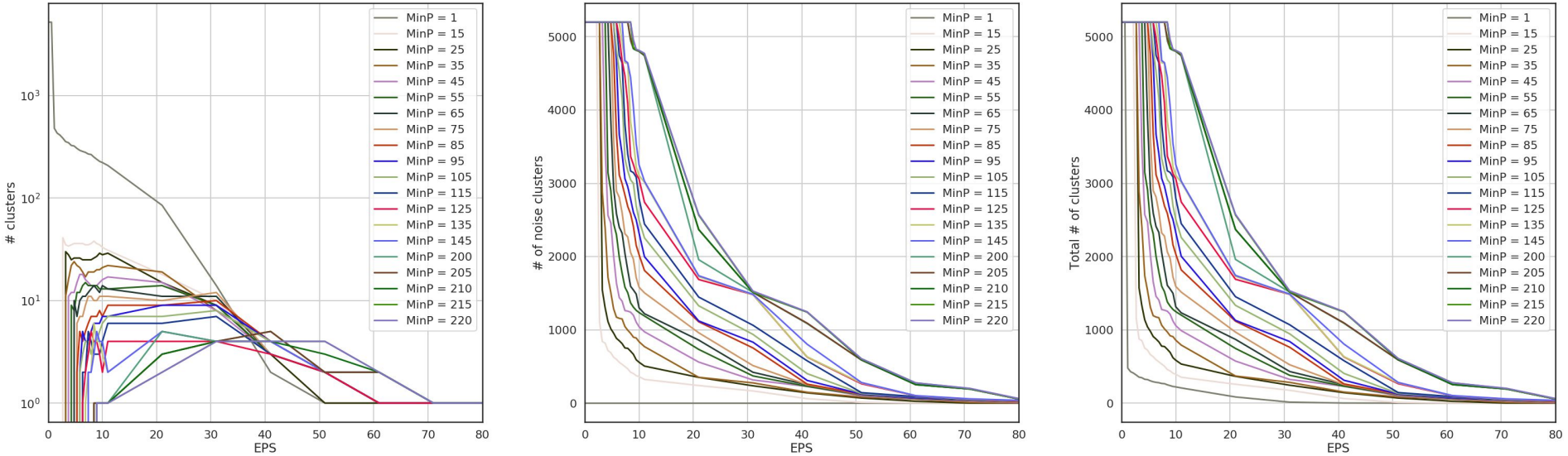
The score is bounded between **-1 for incorrect clustering** and **+1 for highly dense clustering**. Scores around zero indicate overlapping clusters.

The score is **higher** when clusters are **dense and well separated**, which relates to a standard concept of a cluster.

Usando esses métodos de avaliação não foi possível chegar a nenhum resultado conclusivo.



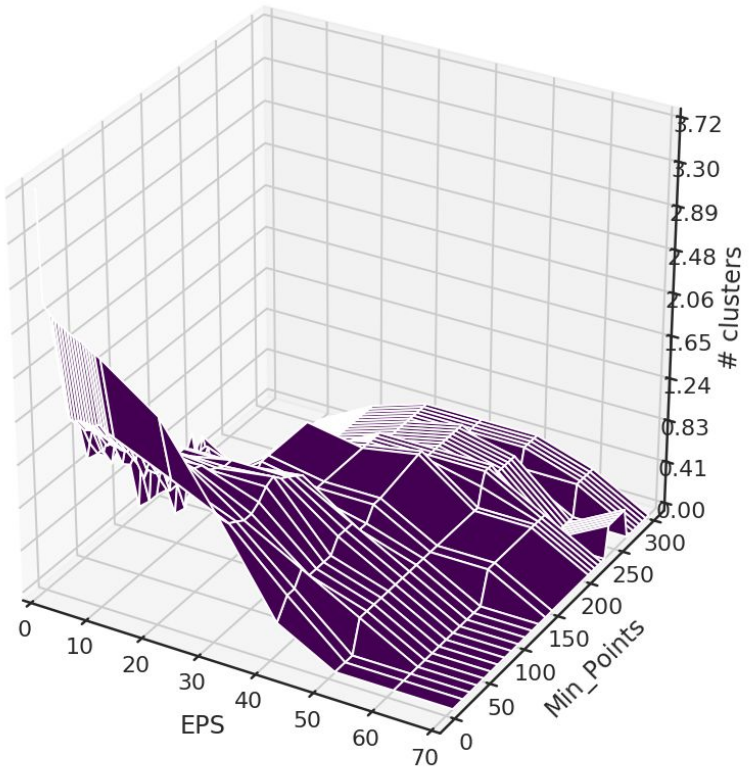
Naive DBSCAN - run494 img 16

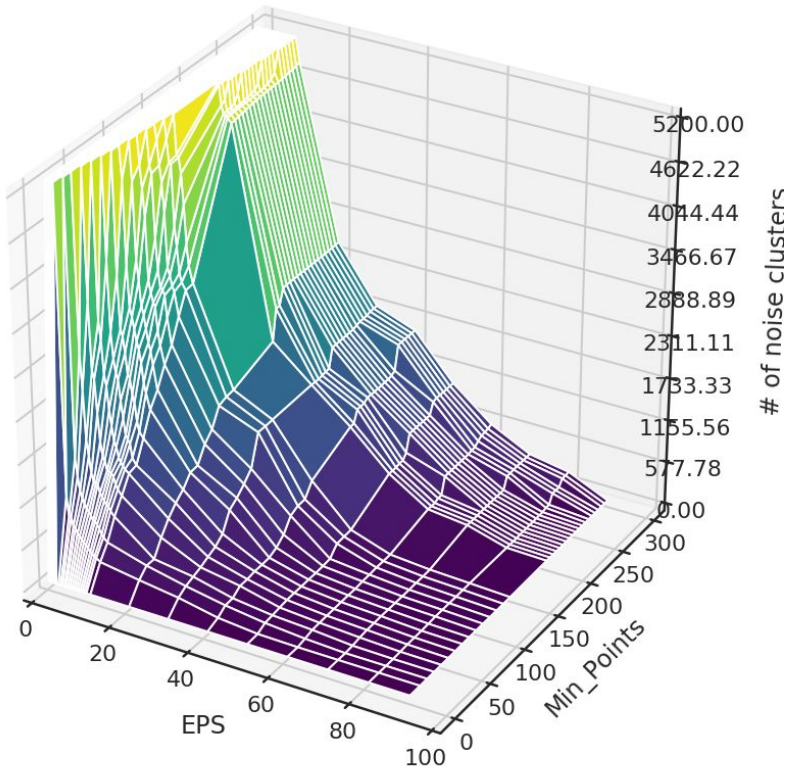


Using DBSCAN in a Cygno image and making a scan over **eps** and **minP**, we can see:

- There is an inflection point when EPS is increasing.
  - Due to the fact that when **EPS** is **low** some samples cannot be clustering, becoming noise and when **EPS** is **high** some nearby clusters are clustering together;
- When **MinP** increase more samples are clustering as noise, so the number of clusters decrease.

Naive DBSCAN - run494 img 16





i2DBSCAN evaluation - Noise rejection iteration over 30 images

With noise rejection iteration

Number of clusters: 2469

Number of noise pixels: 8148

Without noise rejection iteration

Number of clusters: 2864

Number of noise pixels: 6256

If we considered that the ‘noise rejection iteration’ is only labeling as noise the true noise pixels, this approach is able to reject about **14%** of the noise clusters.

And **30%** more pixels are labelled as noise pixels.

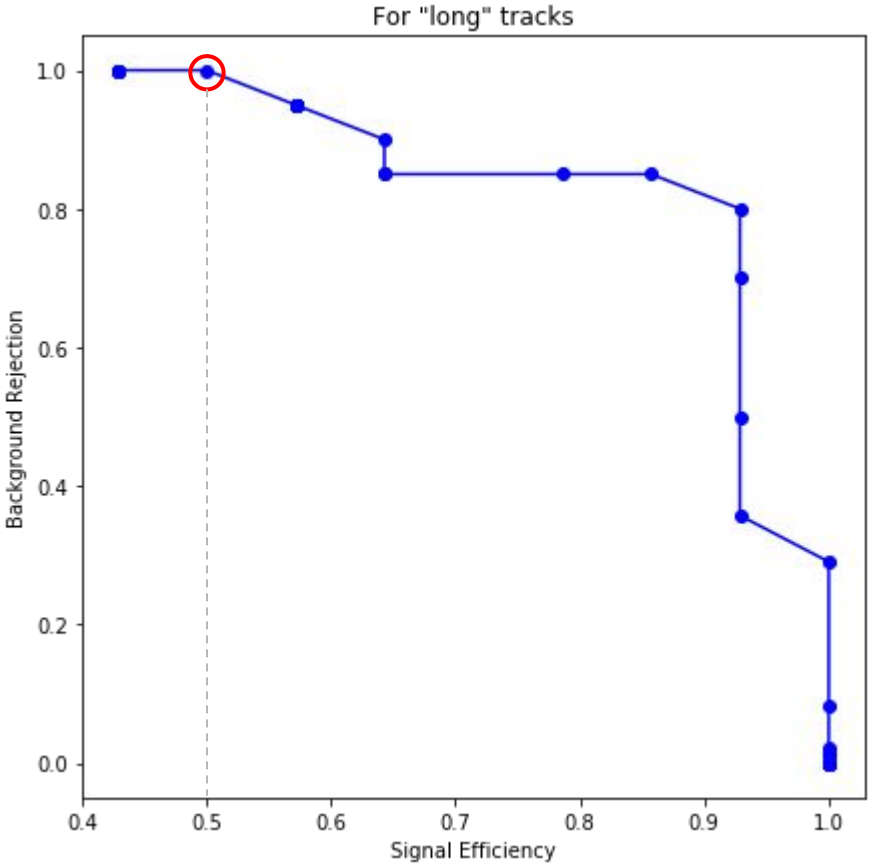
Therefore, using this method it is possible to use tight parameters in the first step, then relax them in the last step, without getting several noise events as a signal.

i2DBSCAN evaluation - Long Tracks

In the ROC curve the threshold is the ‘cut’ to determine what is the ‘long’ tracks.

So, we can work with 100% of background rejection and with a 50% of signal efficiency.

As a first approach to identify the particles.



Próximos passos e novidades

- Como já comentado o próximo passo seria a análise de um novo banco de dados, onde novos problemas vão aparecer.

- Além disso, está sendo falado em conseguir pessoas para ajudar na análise de dados e desenvolvimento do ambiente de análise.
- A equipe CYGNUS está fechando parcerias com outras equipes de matéria escura e descobrindo análises que precisam ser feitas.
- Essas duas últimas semana foram feitas aquisições longas de dados, para entender a interferência da temperatura e/ou pressão no detector