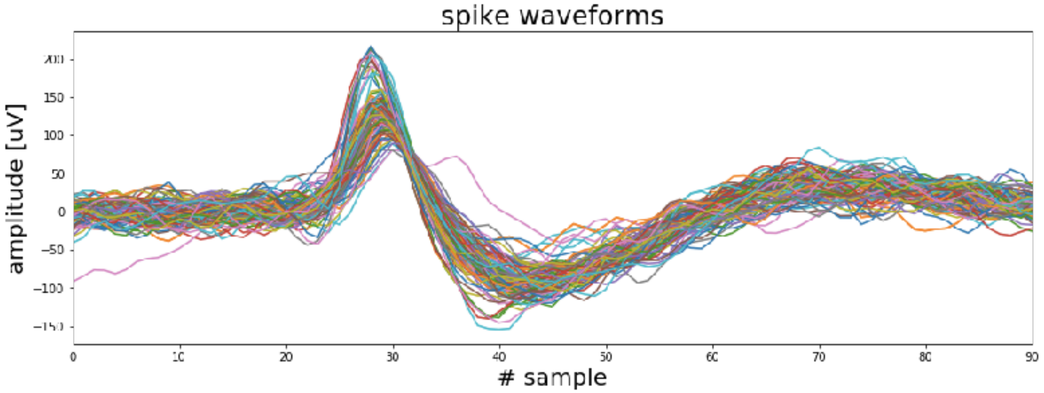
# Understanding Machine Learning models

## Simulation & Analysis

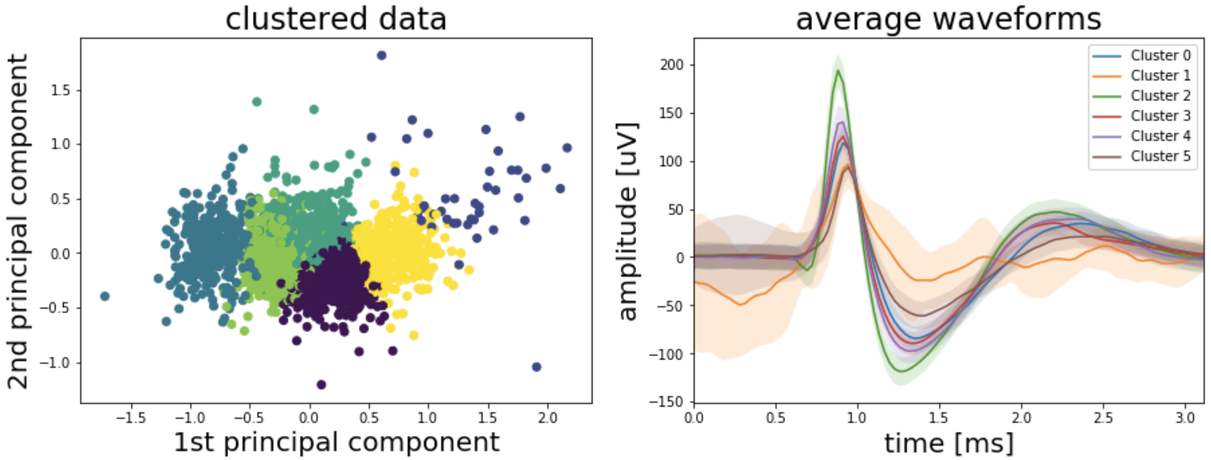
G. Mazzitelli - 28/5/2020

# inspiration & motivation

Epilepsy: find this seizure focus, recording electrodes are inserted into the patients brain with which the neural activity can be monitored in real time




raw date

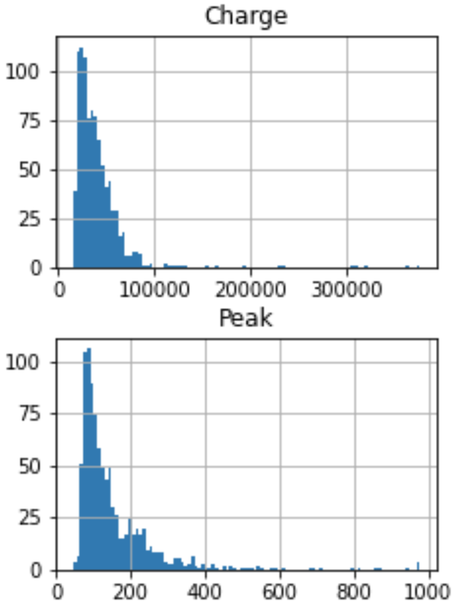


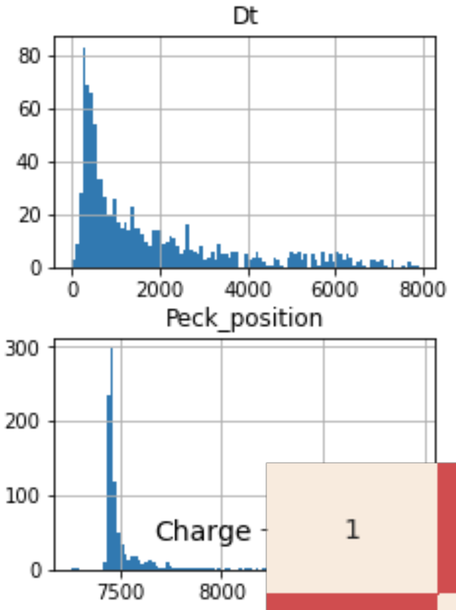

features analysis

https://towardsdatascience.com/whos-talking-using-k-means-clustering-to-sort-neural-events-in-python-e7a8a76f316

# Waveform & Clustering

## features approach: charge, DT, Peak, Peak_position (RUN2098-AmBe)

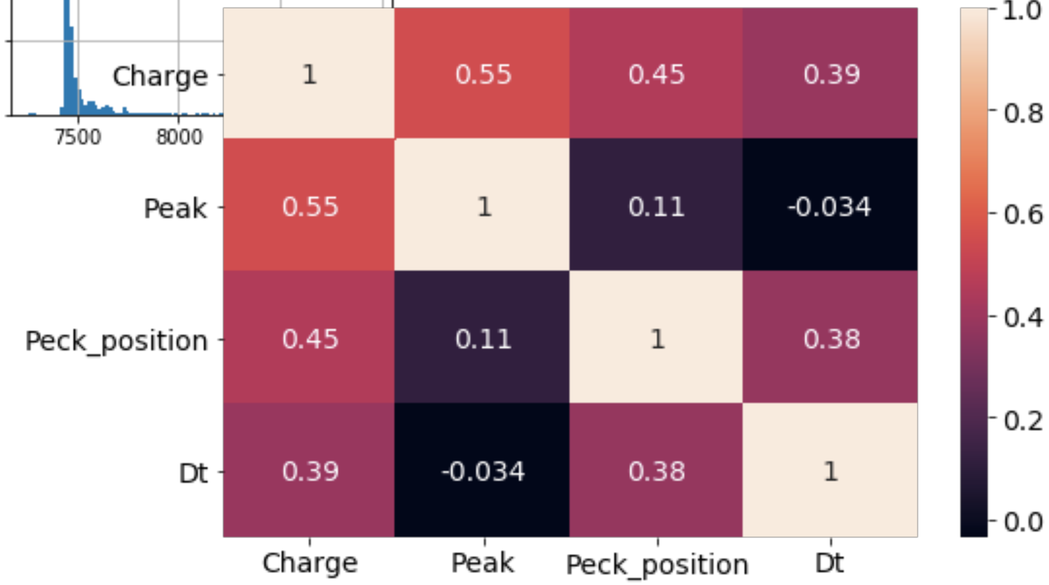


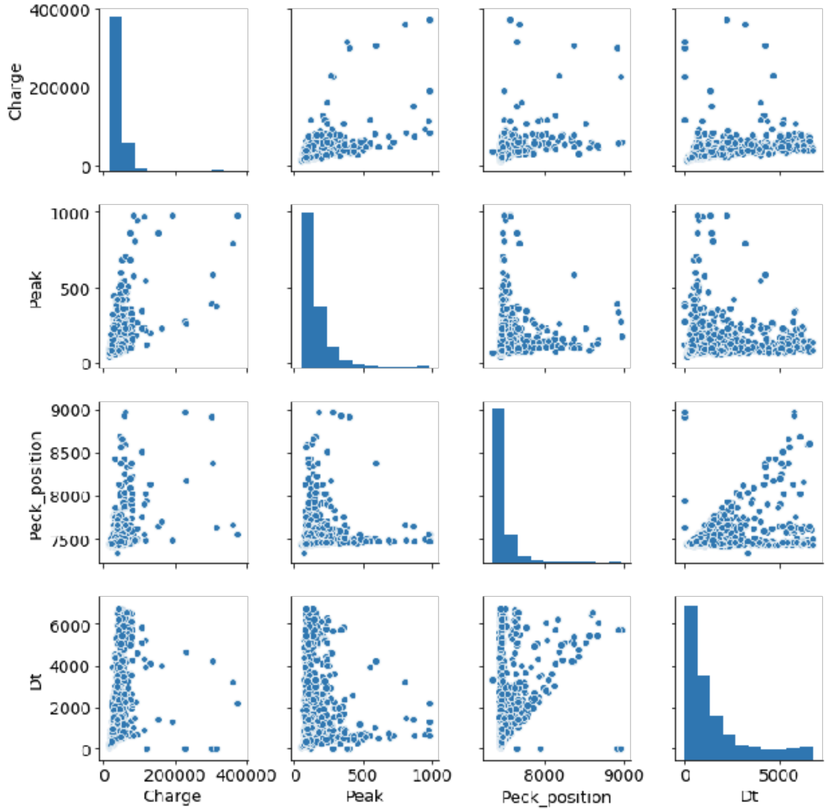







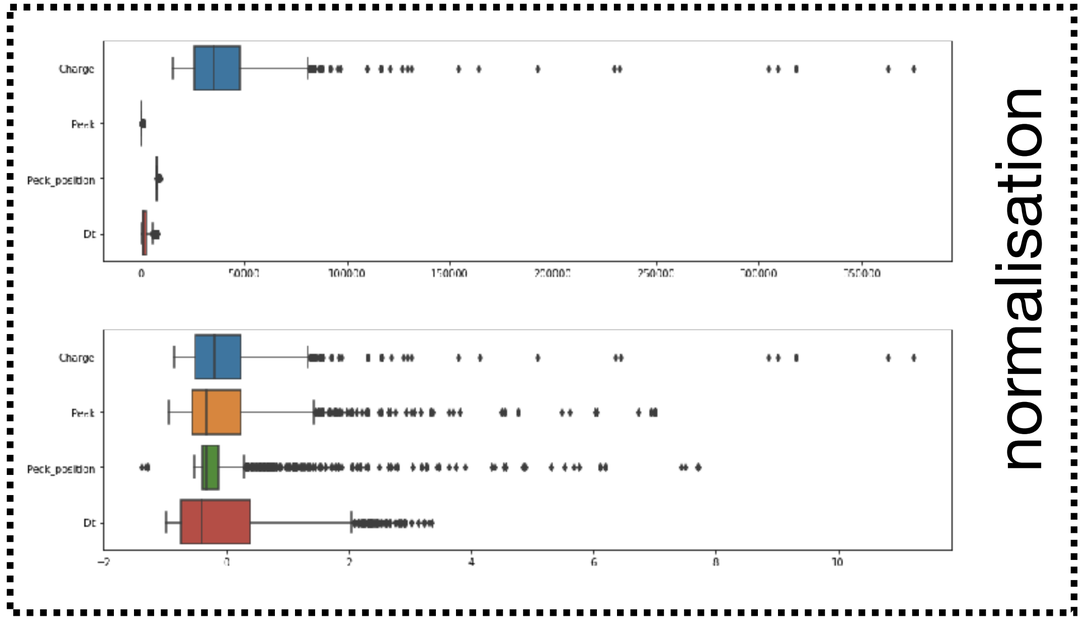


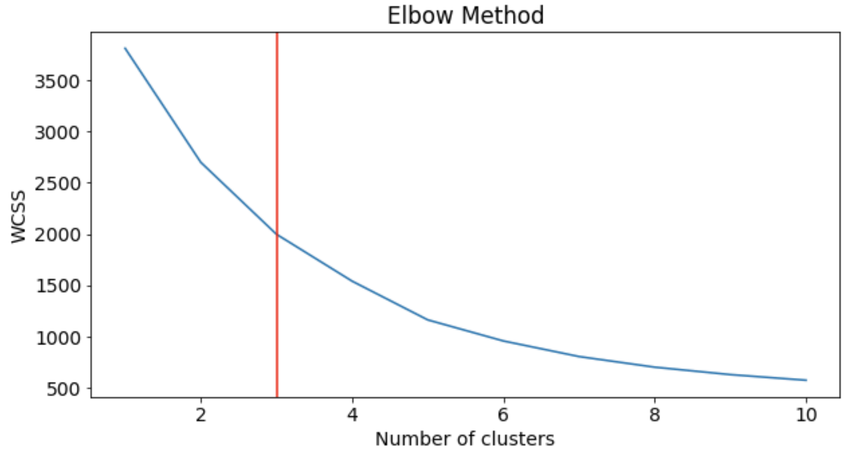



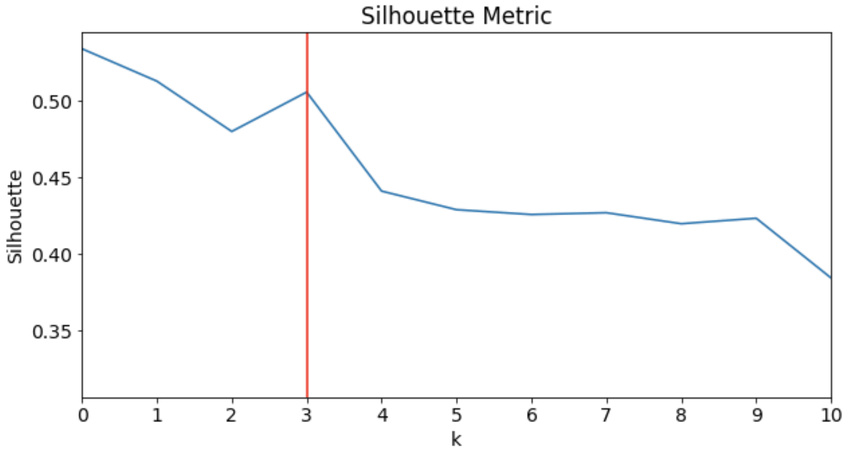
# Waveform & Clustering

## features approach: charge, DT, Peak, Peak_position (RUN2098-AmBe)



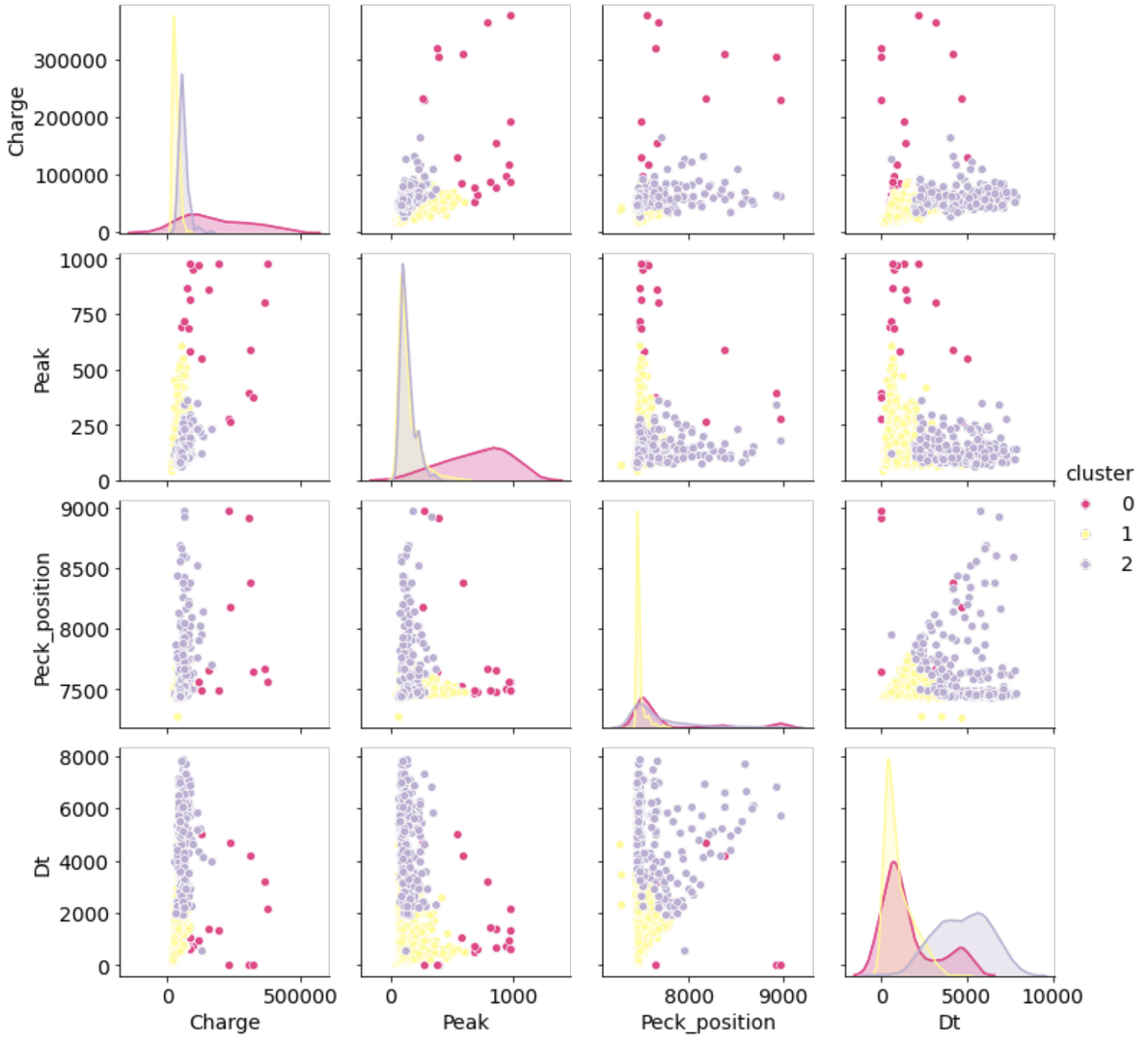

the elbow method is a heuristic approach in determining the number of clusters in a data set




Silhouette: mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample

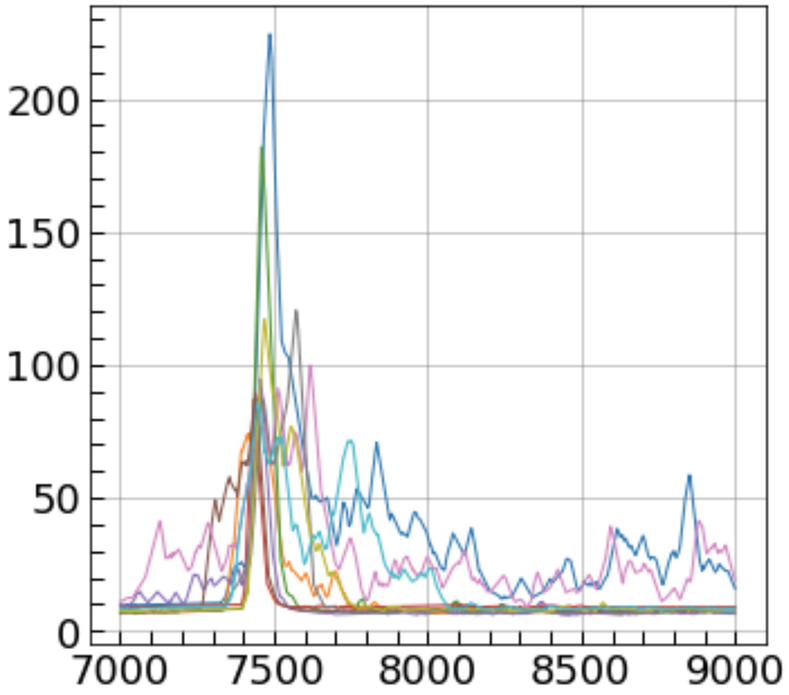
# Waveform & Clustering

## features approach: charge, DT, Peak, Peak_position (RUN2098-AmBe)

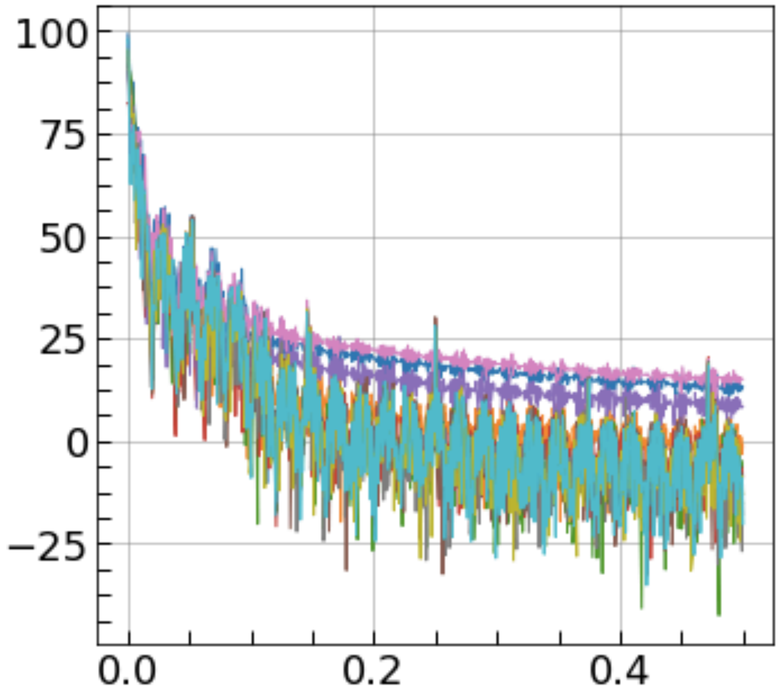


# Waveform

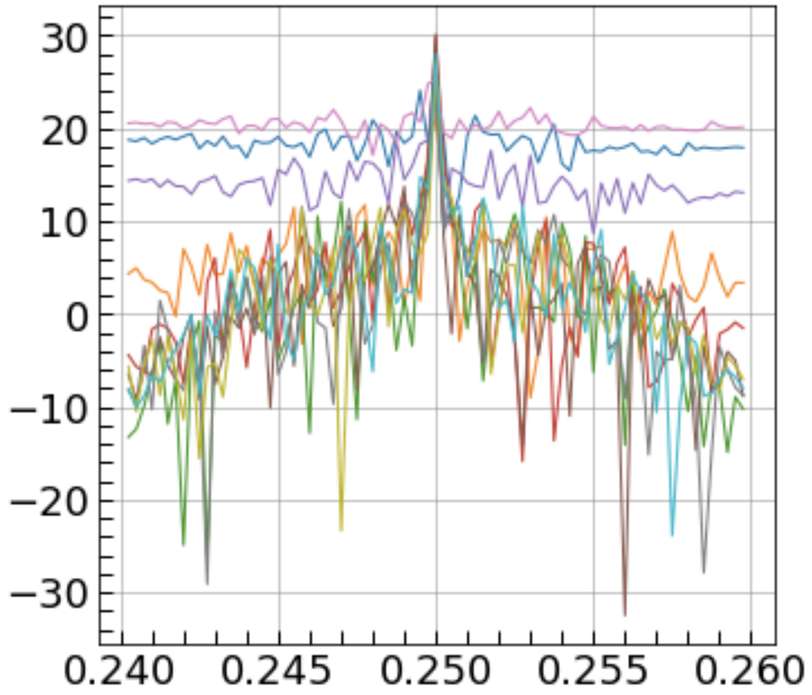
## (RUN2098-AmBe) raw data example - 10 waveform




raw data



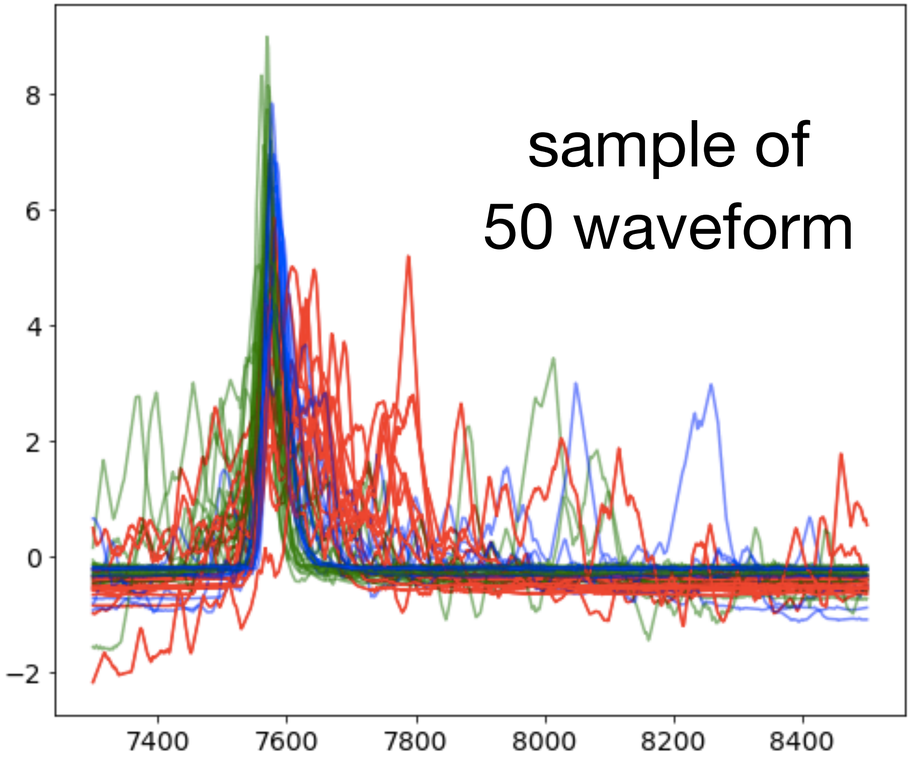

FFT

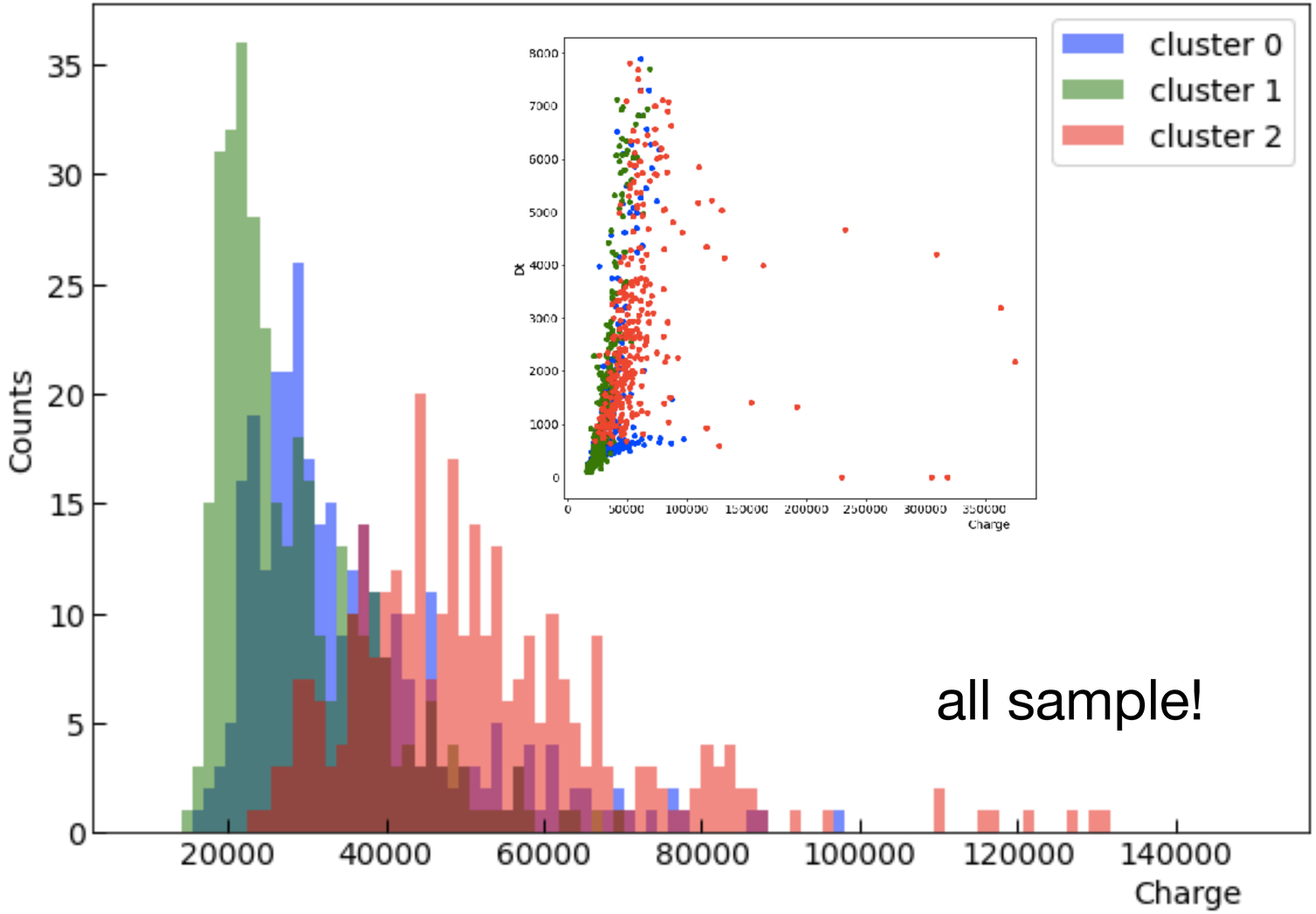



FFT zoom

# Waveform

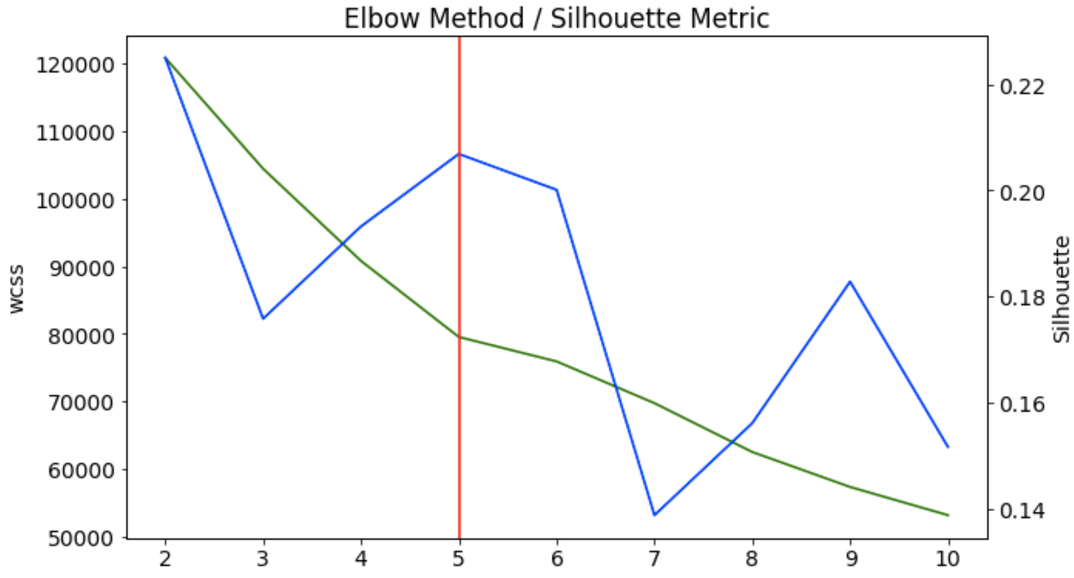
## (RUN2098-AmBe) Kmeans clustering in 3 families

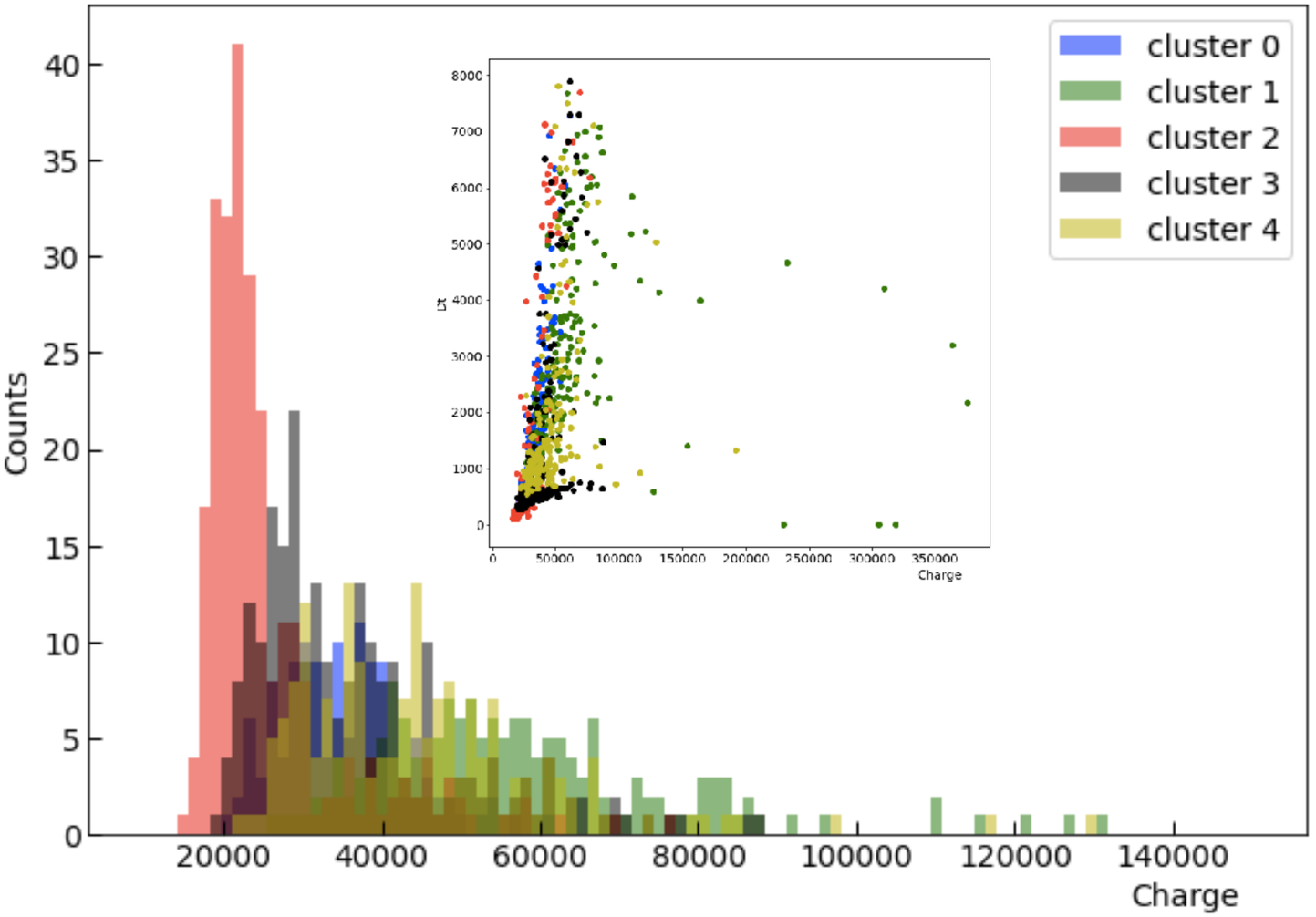


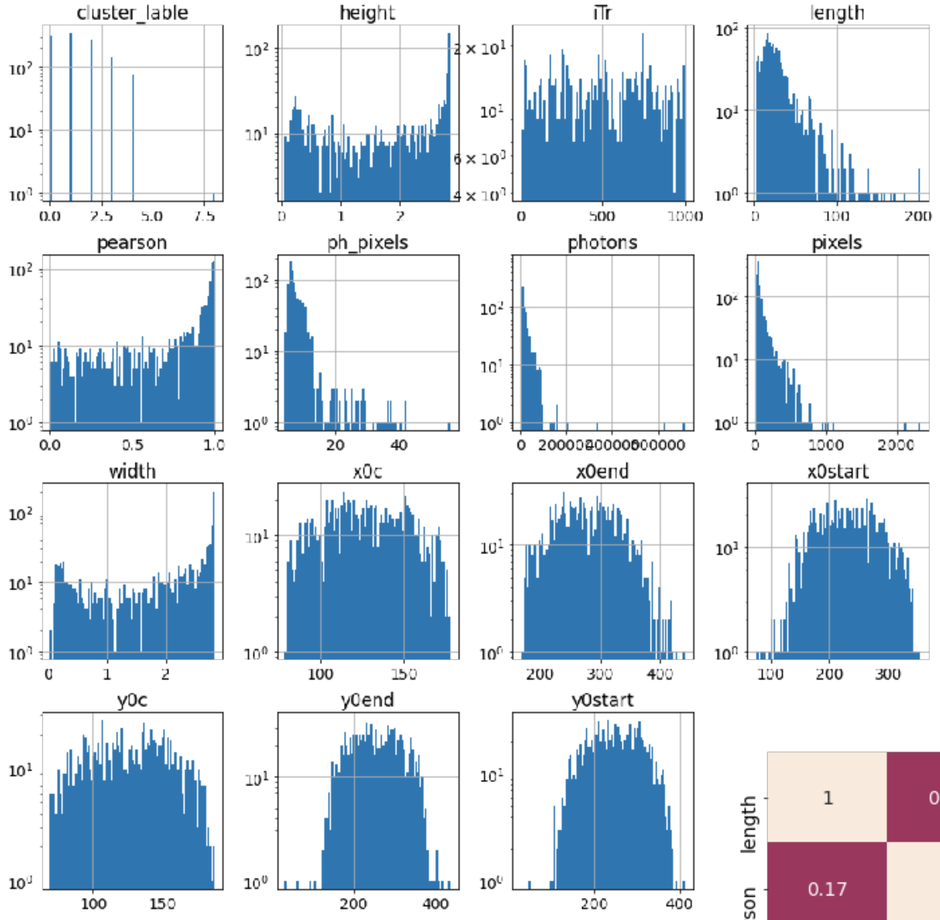
# Waveform

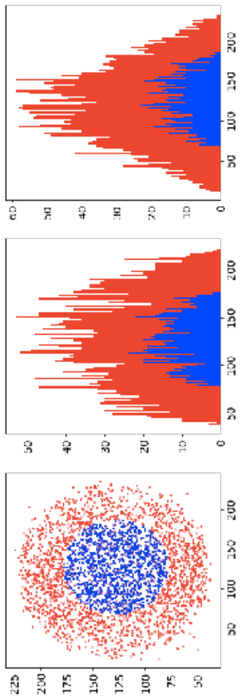
## (RUN2098-AmBe) Kmeans clustering in 5 families

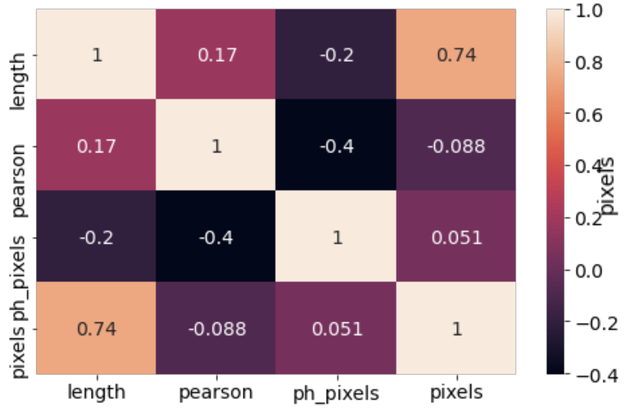


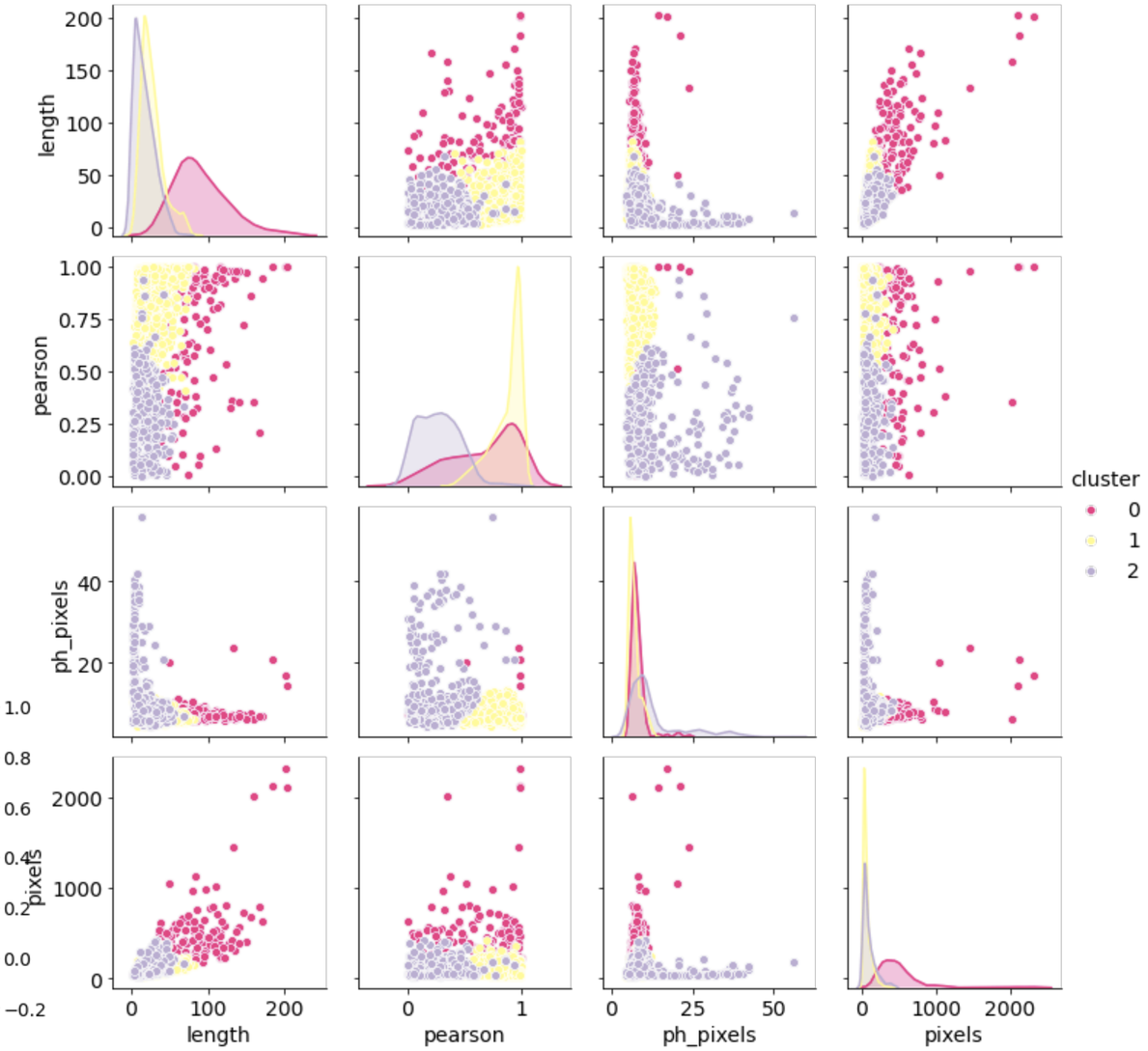



# DBSCAN futures

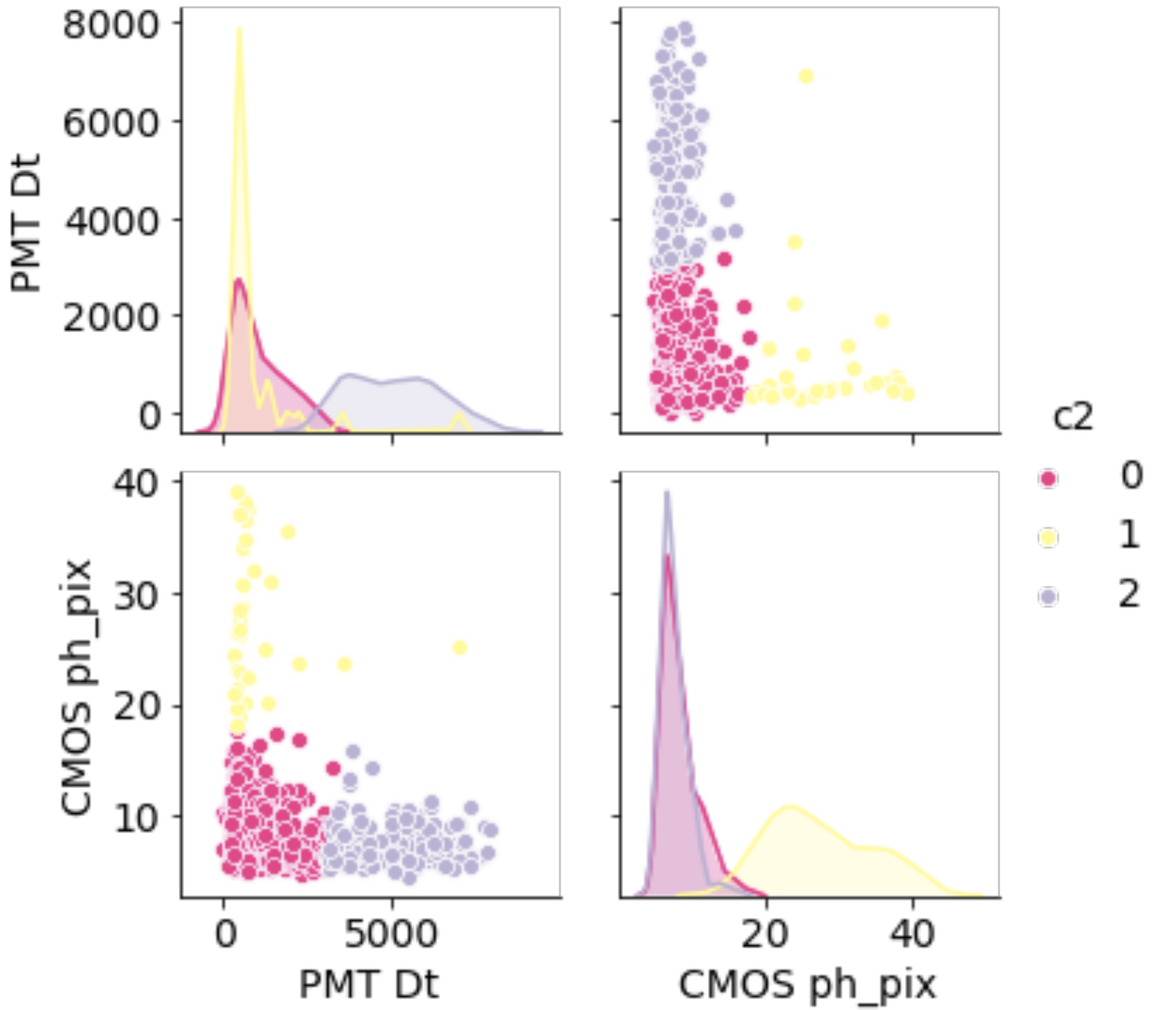




# Conclusion 1/2

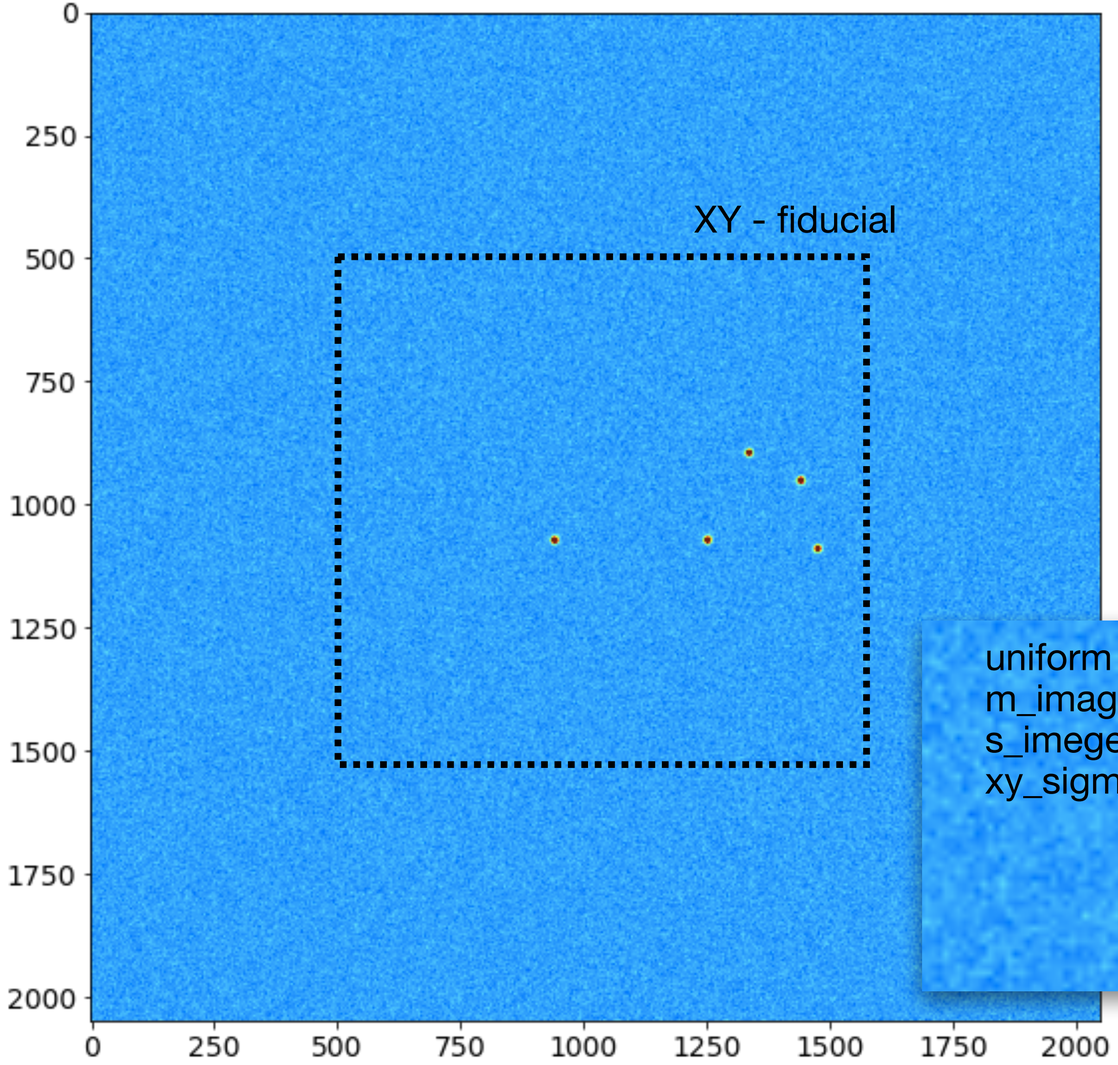


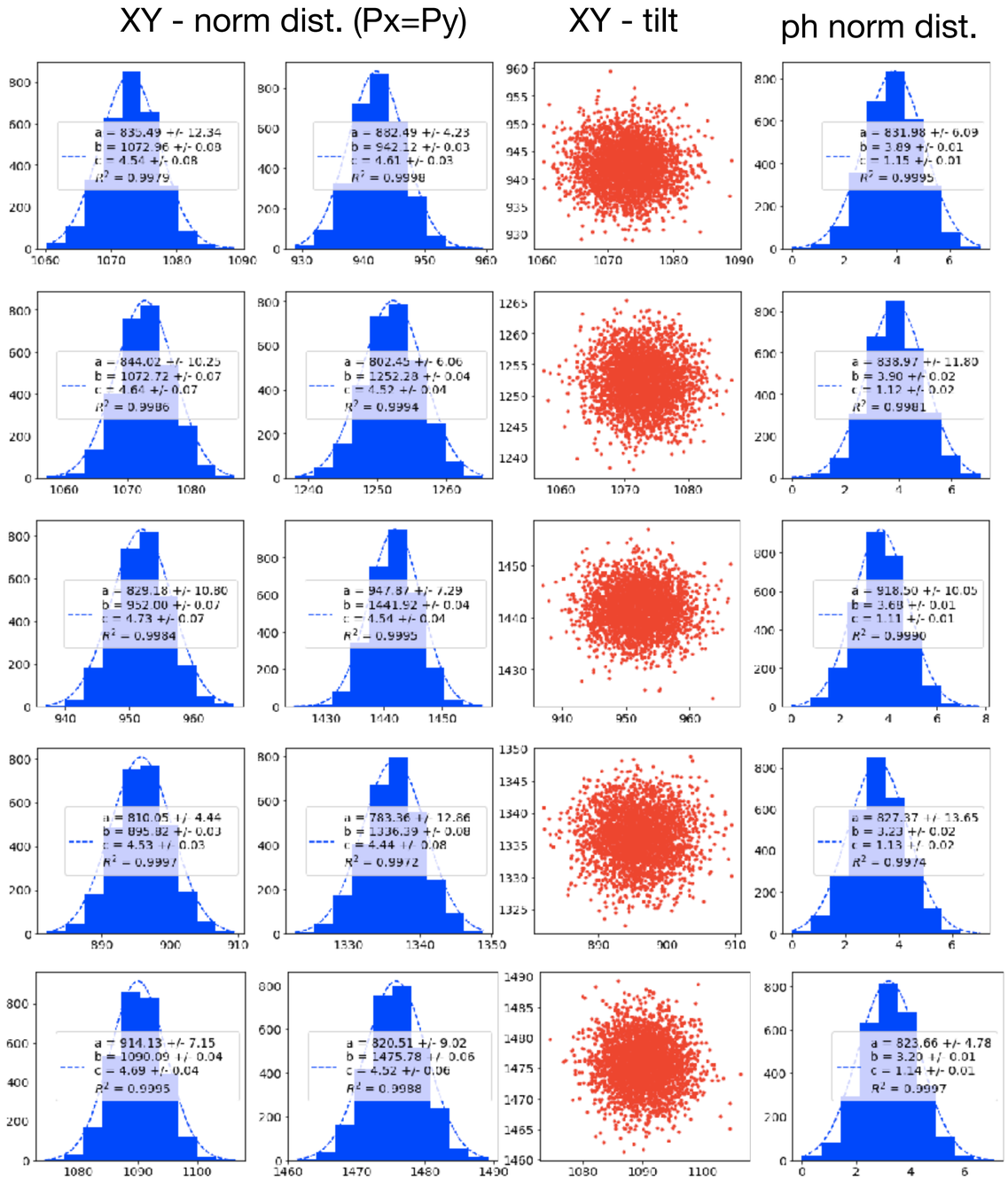

- looks that a clustering algorithms work fine on PMT waveform
- the FFT could also be use to improve features
- what and how to correlate DBSCAN clusters data with PMT sum of many different contribution?

# Simulation

# Fe work around

## sklearn

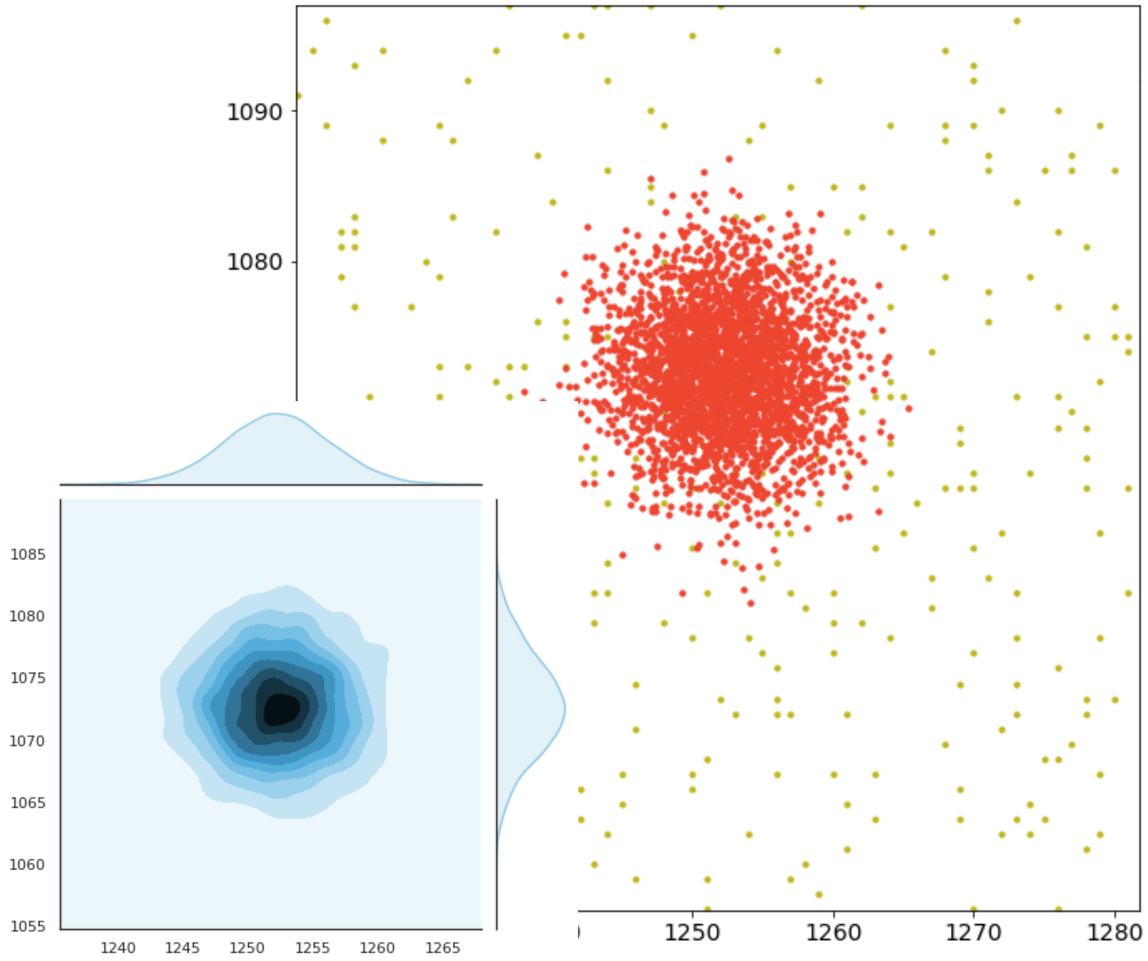


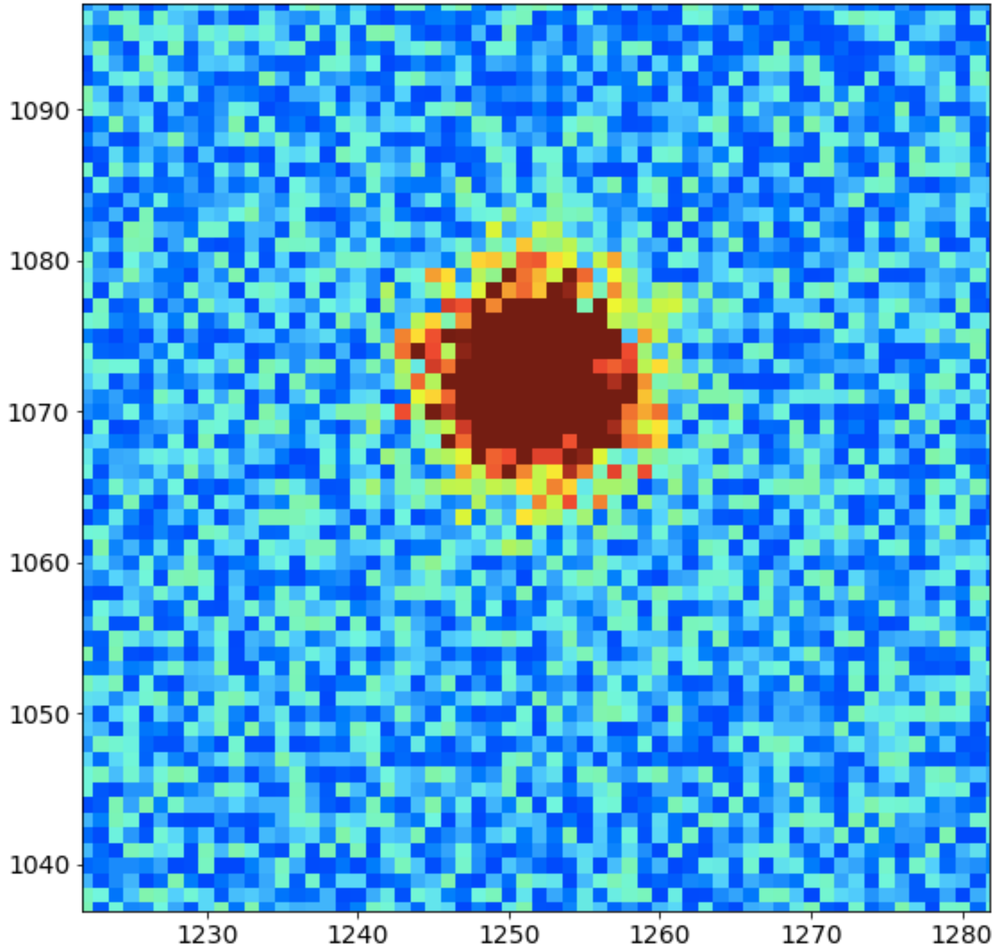
# imaging

5 spot * 100 image ~ 17 s on colab

noise-m_image > 3.5




vmin=98,vmax=110




! all float

# DBSCAN

SIM(OUT)

| | iTr | image |
|---|---|---|
| **95** | 95 | [[99.73538564343858, 101.77535204841473, 102.6... |
| **96** | 96 | [[99.84488393284403, 102.52351488762105, 99.76... |
| **97** | 97 | [[102.54013475607381, 101.47048471208953, 102.... |
| **98** | 98 | [[99.57021351709082, 103.16033324534018, 100.4... |
| **99** | 99 | [[101.88468871854188, 102.51683772727921, 103.... |

plk2root (Emanule)

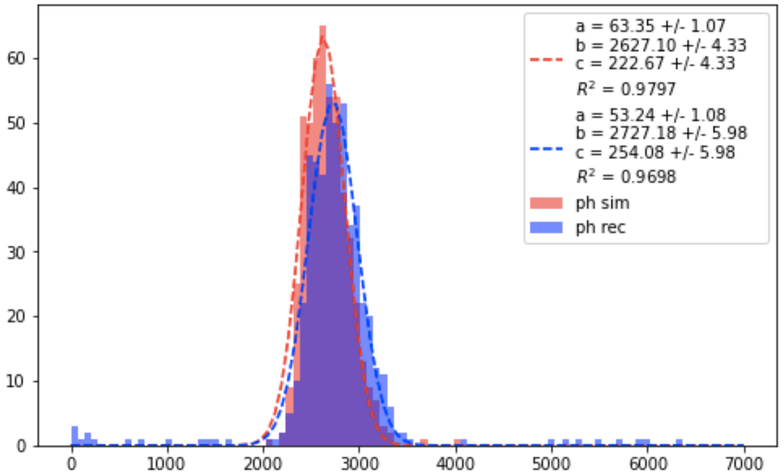
| | iTr | photons | blobs_points |
|---|---|---|---|
| **495** | 99.0 | 2452.577832 | 0.0 |
| **496** | 99.0 | 2604.912876 | 1.0 |
| **497** | 99.0 | 2660.719968 | 2.0 |
| **498** | 99.0 | 2529.341997 | 3.0 |
| **499** | 99.0 | 2724.850270 | 4.0 |

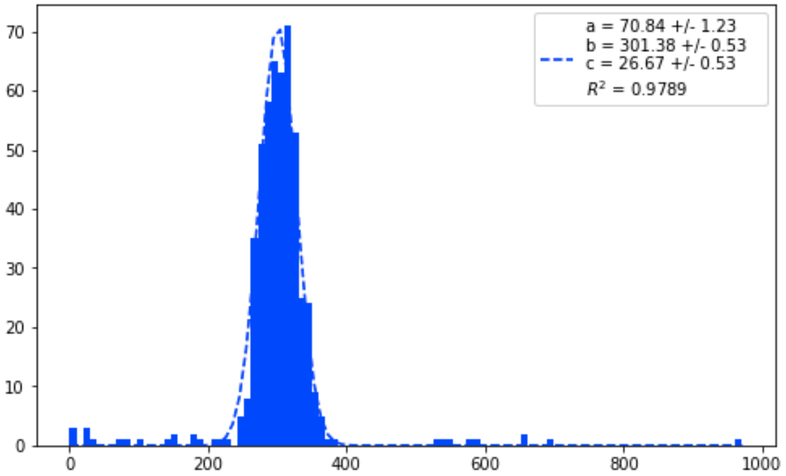
plk2CYGNODB (??)

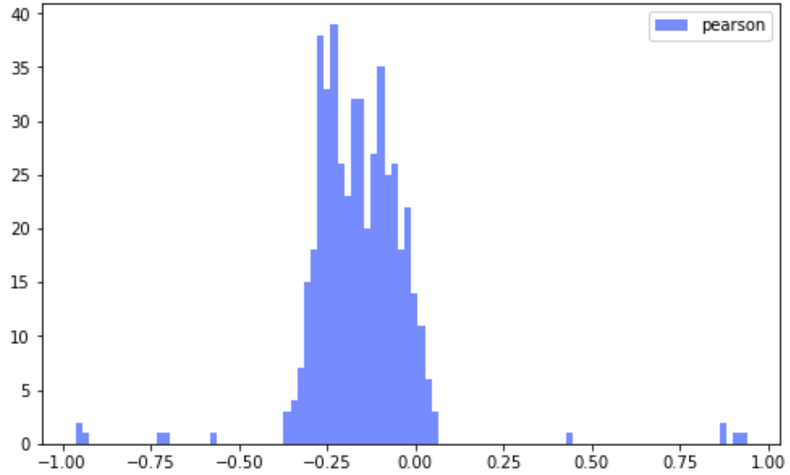
DBSCAN (OUT)

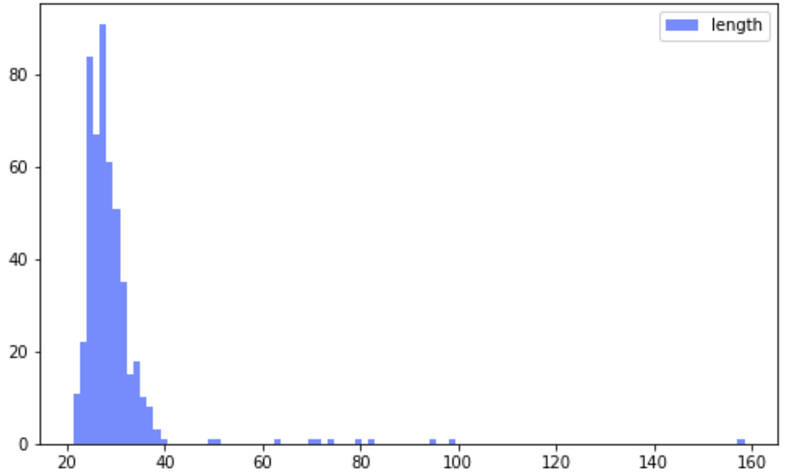
| | iTr | cluster_lable | pixels | photons | ph_pixels | x0start | y0start | x0end | y0end | width | height | pearson | length | x0c | y0c |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **481** | 98.0 | 4.0 | 325.0 | 2523.933871 | 7.765950 | 1666.0 | 712.0 | 1698.0 | 706.0 | 1.584169 | 2.343162 | -0.372602 | 32.557641 | 1682.0 | 709.0 |
| **482** | 99.0 | 0.0 | 491.0 | 5305.053363 | 10.804589 | 576.0 | 1248.0 | 628.0 | 1282.0 | 2.780871 | 0.516483 | 0.933311 | 62.128898 | 602.0 | 1265.0 |
| **483** | 99.0 | 1.0 | 233.0 | 2668.744005 | 11.453837 | 780.0 | 571.0 | 804.0 | 582.0 | 2.189313 | 1.790784 | 0.198273 | 26.400758 | 792.0 | 576.5 |
| **484** | 99.0 | 2.0 | 245.0 | 2776.820931 | 11.333963 | 903.0 | 1014.0 | 925.0 | 1020.0 | 2.146689 | 1.841665 | 0.152068 | 22.803509 | 914.0 | 1017.0 |
| **485** | 99.0 | 3.0 | 227.0 | 2462.578364 | 10.848363 | 1184.0 | 1035.0 | 1207.0 | 1042.0 | 2.177378 | 1.805276 | 0.185244 | 24.041631 | 1195.5 | 1038.5 |

# DBSCAN (2D)


a = 63.35 +/- 1.07
b = 2627.10 +/- 4.33
c = 222.67 +/- 4.33
$R^2$ = 0.9797
a = 53.24 +/- 1.08
b = 2727.18 +/- 5.98
c = 254.08 +/- 5.98
$R^2$ = 0.9698
ph sim
ph rec


a = 70.84 +/- 1.23
b = 301.38 +/- 0.53
c = 26.67 +/- 0.53
$R^2$ = 0.9789


pearson


length

# Conclusion 2/2

## simulation

- round numbers, realistic ph distribution
- set correct simulation parameter
- ! COMS noise (Telegraph)/radioactivity and cosmic
- produce large statics (parallelise production/analysis)
- select a list of interesting input/output