

# **Lab Notebook**

Photonic Lantern Information Determination

# Contents

<b>1</b>	<b>Mini Dataset Information Determination</b>	<b>3</b>
<b>2</b>	<b>The data</b>	<b>3</b>
2.1	Zernike coefficients dataset . . . . .	3
2.2	PSFs intensities dataset . . . . .	3
2.3	LP mode coefficients dataset . . . . .	3
2.4	LP mode coefficients dataset . . . . .	3
<b>3</b>	<b>Preprocessing</b>	<b>4</b>
3.1	PSF Intensities . . . . .	4
3.2	LP Coefficients . . . . .	4
3.3	Output fluxes . . . . .	4
<b>4</b>	<b>Clustering</b>	<b>4</b>
4.1	Zernike coefficients clustering . . . . .	5
4.1.1	K-Means . . . . .	5
4.1.2	DBSCAN . . . . .	6
4.1.3	HDBSCAN . . . . .	8
4.1.4	Agglomerative clustering . . . . .	10
4.1.5	Summary . . . . .	12
4.2	LP coefficients clustering . . . . .	13
4.2.1	K-Means . . . . .	13
4.2.2	DBSCAN . . . . .	15
4.2.3	HDBSCAN . . . . .	16
4.2.4	Agglomerative clustering . . . . .	18
4.2.5	Summary . . . . .	20

---

4.3	Output fluxes clustering . . . . .	21
4.3.1	K-Means . . . . .	21
4.3.2	DBSCAN . . . . .	23
4.3.3	HDBSCAN . . . . .	24
4.3.4	Agglomerative clustering . . . . .	26
4.3.5	Summary . . . . .	28
4.4	PSF Intensities clustering . . . . .	29
4.4.1	K-Means . . . . .	29
4.4.2	DBSCAN . . . . .	31
4.4.3	HDBSCAN . . . . .	33
4.4.4	Agglomerative clustering . . . . .	35
4.4.5	Summary . . . . .	37

# 1 Mini Dataset Information Determination

## 2 The data

### 2.1 Zernike coefficients dataset

A dataset of 3200 zernike coefficients is created for this report. In particular, each datapoint represents the coefficients of the first 5 Zernike modes, their values ranging between:

- The first 2 modes between  $[-2, -1.8]$  and  $[1.8, 2]$ .
- Modes 4, 5 and 6 between  $[-1, -0.8]$  and  $[0.8, 1]$

These ranges create 32 original clusters that will be used as reference.

### 2.2 PSFs intensities dataset

A dataset of 3200 PSFs is created using the Zernike coefficients dataset.

### 2.3 LP mode coefficients dataset

A dataset of 3200 LP mode coefficients obtained from computing the overlap integral of the first 19 LP modes with the PSF dataset.

### 2.4 LP mode coefficients dataset

A dataset of 3200 PL output fluxes obtained from the PL transfer matrix and LP coefficients.

## 3 Preprocessing

### 3.1 PSF Intensities

The 3200x128x128 array is dimensionally reduced using PCA and UMAP both giving an array of 3200x19 projections of the PSF Intensities.

### 3.2 LP Coefficients

The 3200x19x2 array is dimensionally reduced using PCA and UMAP both giving an array of 3200x2 projections of the original LP coefficients.

### 3.3 Output fluxes

The 3200x19 array is dimensionally reduced using PCA and UMAP both giving an array of 3200x2 projections of the original LP coefficients.

## 4 Clustering

A series of different clustering algorithms are used:

- K-Means
- DBSCAN
- HDBSCAN
- Agglomerative clustering

The clusters obtained will be compared the original clusters using NMI

## 4.1 Zernike coefficients clustering

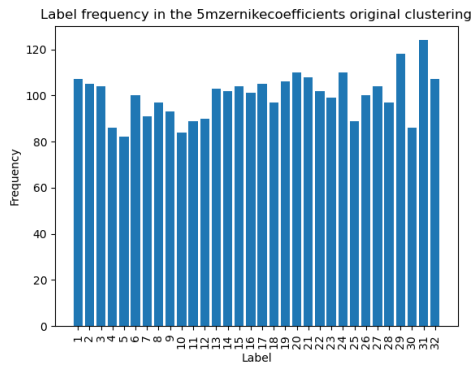
### 4.1.1 K-Means

As K-Means allows for the number of clusters to define, and we know that there are 4 in the original dataset, K-Means is used to find 4 clusters.

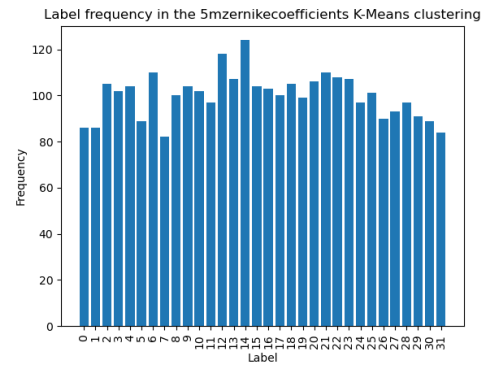
Number of clusters	Number of initializations
32	100

Table 1: K-Means hyperparameter configuration for Zernike coefficients clustering

The results are the following:

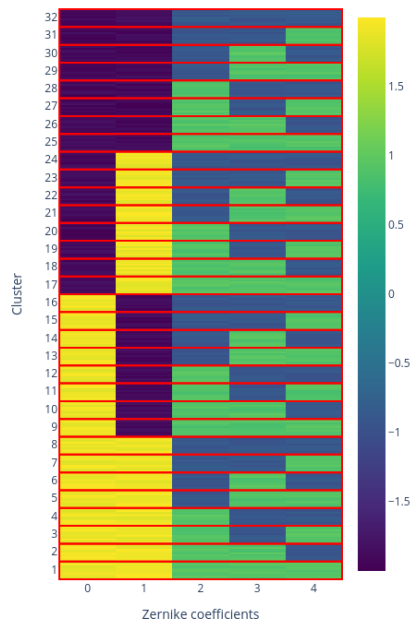


(a) Original cluster densities



(b) K-Means clusters densities

Original Zernike coefficients cluster samples



(c) Original cluster samples

K-Means Zernike coefficients cluster samples



(d) K-Means cluster samples

Figure 1: Comparison between original clustering and K-Means clustering

#### 4.1.2 DBSCAN

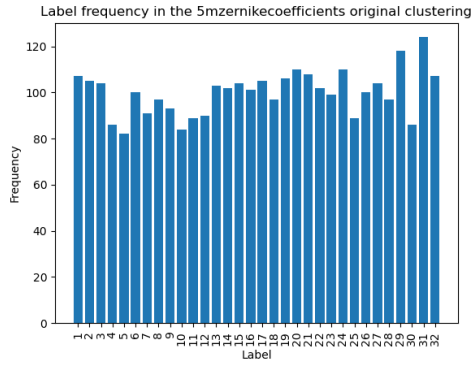
A configuration that outputs 4 clusters is searched

Number of neighbours	Epsilon
5	0.14

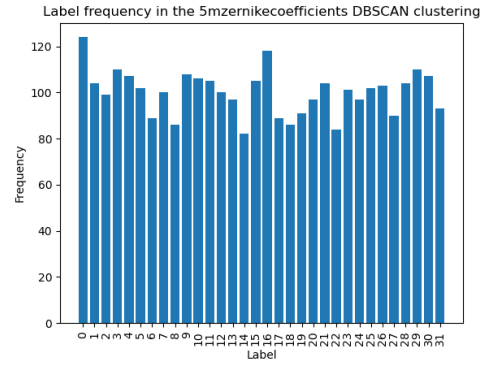
Table 2: DBSCAN hyperparameter configuration for Zernike coefficients clustering

The results are the following:



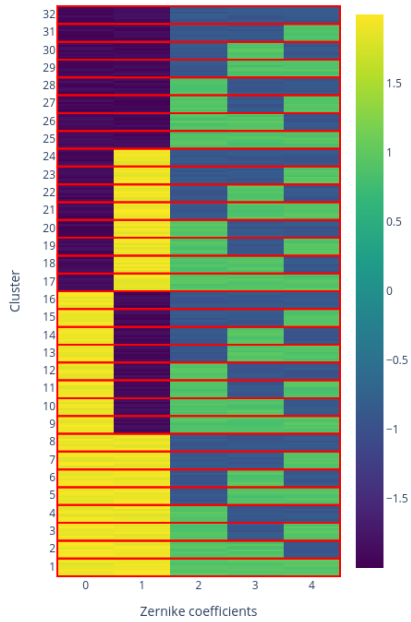


(a) Original cluster densities



(b) DBSCAN clusters densities

Original Zernike coefficients cluster samples



(c) Original cluster samples

DBSCAN Zernike coefficients cluster samples



(d) DBSCAN cluster samples

Figure 2: Comparison between original clustering and DBSCAN clustering

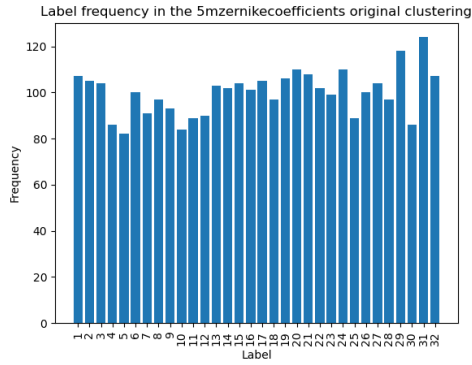
### 4.1.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

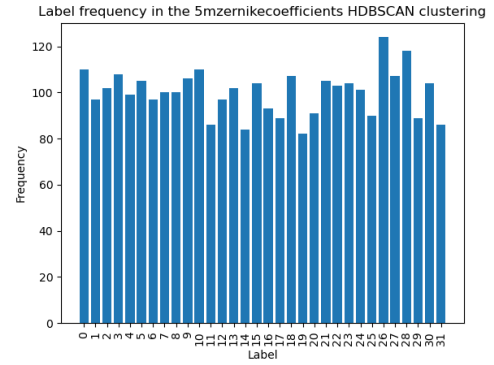
Minimum cluster size
50

Table 3: HDBSCAN hyperparameter configuration for Zernike coefficients clustering

The results are the following:

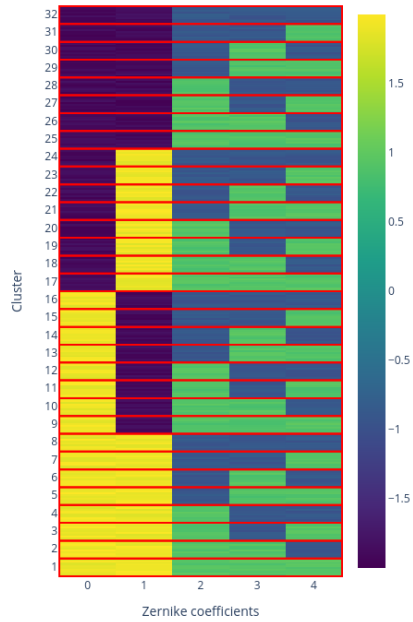


(a) Original cluster densities



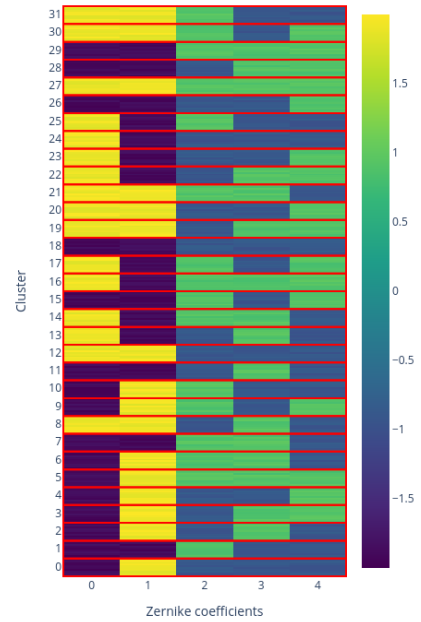
(b) HDBSCAN clusters densities

Original Zernike coefficients cluster samples



(c) Original cluster samples

HDBSCAN Zernike coefficients cluster samples



(d) HDBSCAN cluster samples

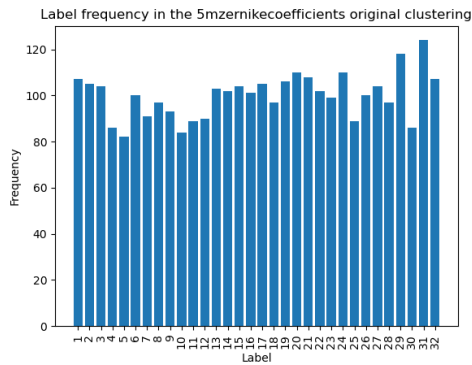
Figure 3: Comparison between original clustering and HDBSCAN clustering

#### 4.1.4 Agglomerative clustering

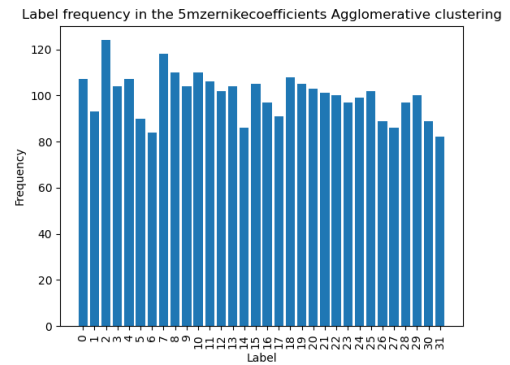
Number of clusters 4
5

Table 4: Agglomerative hyperparameter configuration for Zernike coefficients clustering

The results are the following:

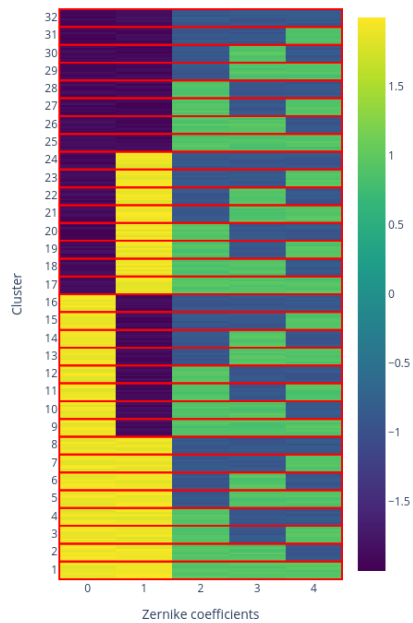


(a) Original cluster densities



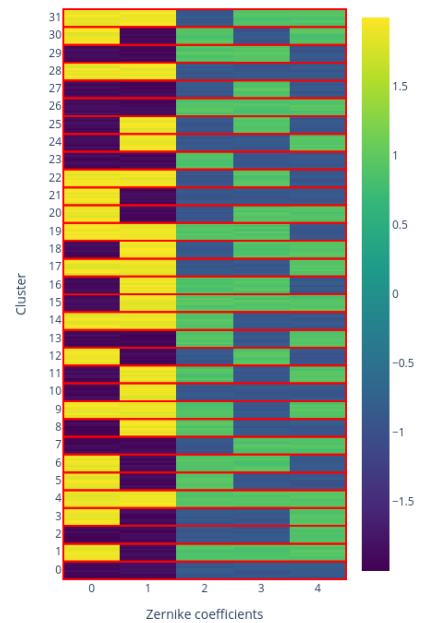
(b) Agglomerative clusters densities

Original Zernike coefficients cluster samples



(c) Original cluster samples

Agglomerative Zernike coefficients cluster samples



(d) Agglomerative cluster samples

Figure 4: Comparison between original clustering and Agglomerative clustering

#### 4.1.5 Summary

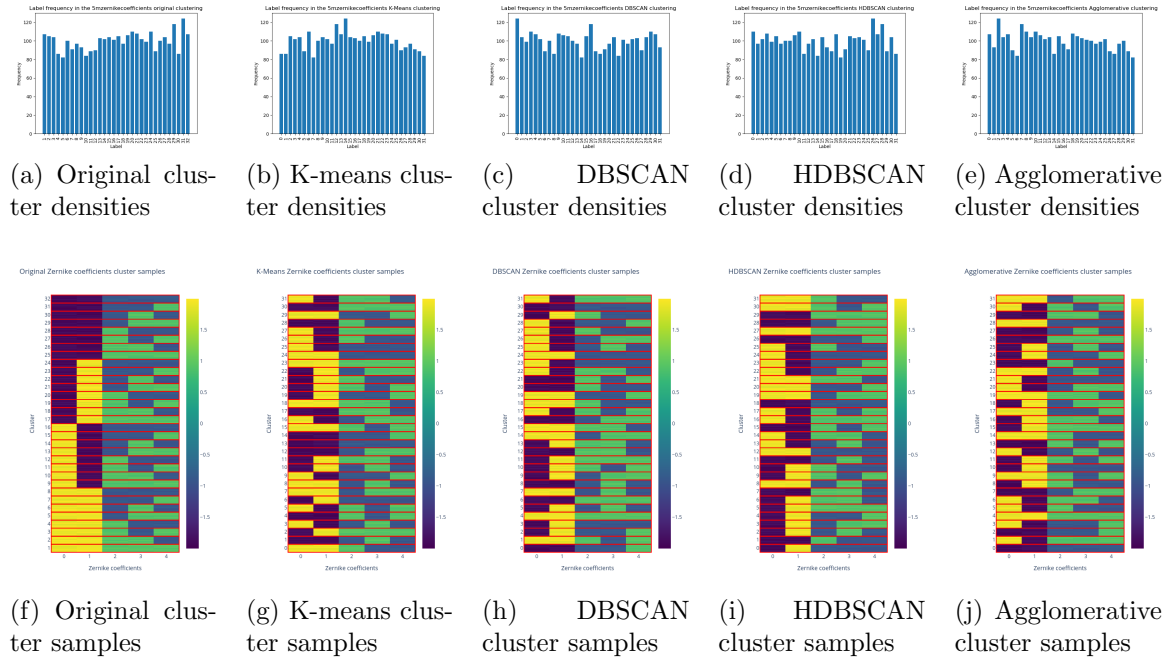


Figure 5: Comparison between clustering algorithms

	Original	K-Means	DBSCAN	HDBSCAN	Agglomerative
Original	\	1	1	1	1
K-Means		\	1	1	1
DBSCAN			\	1	1
HDBSCAN				\	1

Table 5: Normalized Mutual Information between clusters

## 4.2 LP coefficients clustering

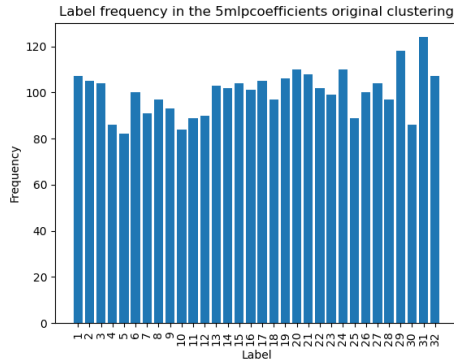
### 4.2.1 K-Means

As K-Means allows for the number of clusters to be defined, and we know that there are 4 in the original dataset, K-Means is used to find 4 clusters.

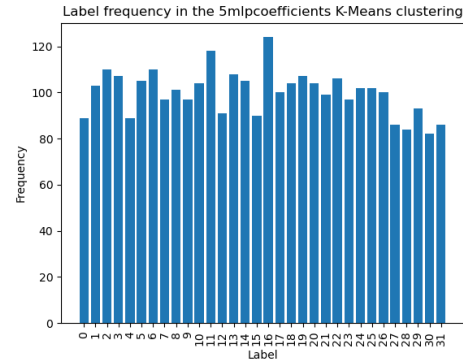
The results are the following:

	Number of clusters	Number of initializations
<b>Original LP coefficients</b>	32	100

Table 6: K-Means hyperparameter configuration for c coefficients clustering

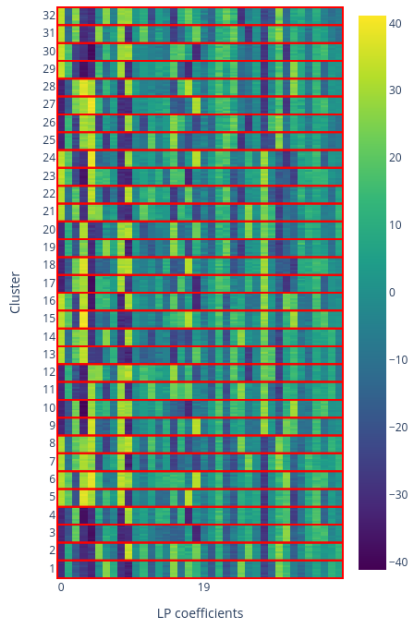


(a) Original cluster densities



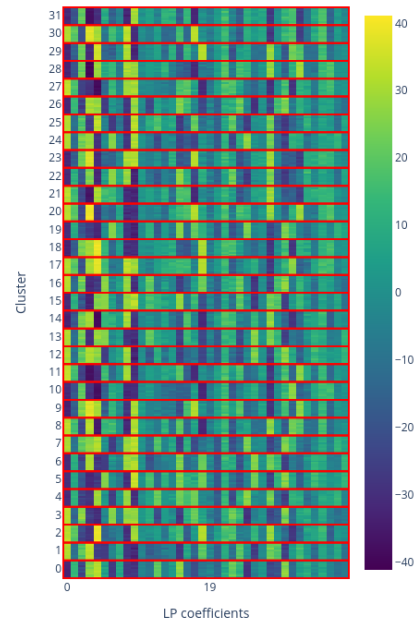
(b) K-Means clusters densities

Original LP coefficients cluster samples



(c) Original cluster samples

K-Means LP coefficients cluster samples



(d) K-Means cluster samples

Figure 6: Comparison between original clustering and K-Means clustering from original LP coefficients

### 4.2.2 DBSCAN

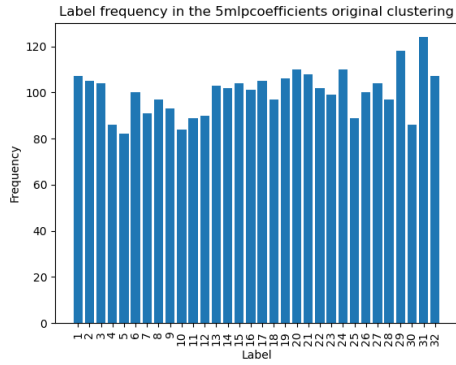
A configuration that outputs 4 clusters is searched

	<b>Number of neighbours</b>	<b>Epsilon</b>
Original LP coefficients	15	11

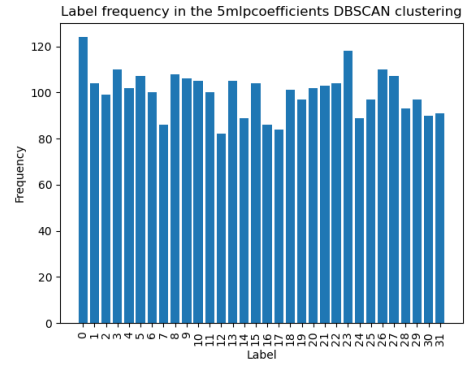
Table 7: DBSCAN hyperparameter configuration for LP coefficients clustering

The results are the following:



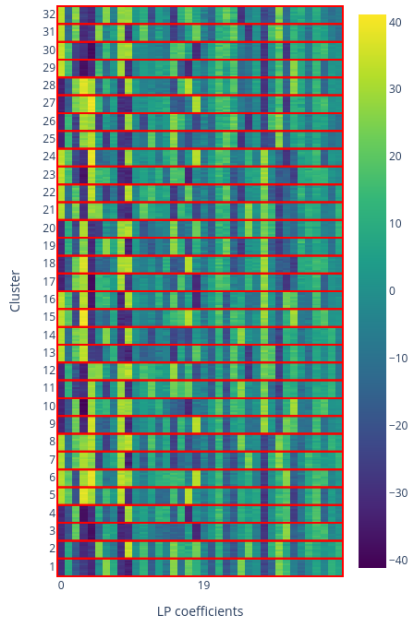


(a) Original cluster densities



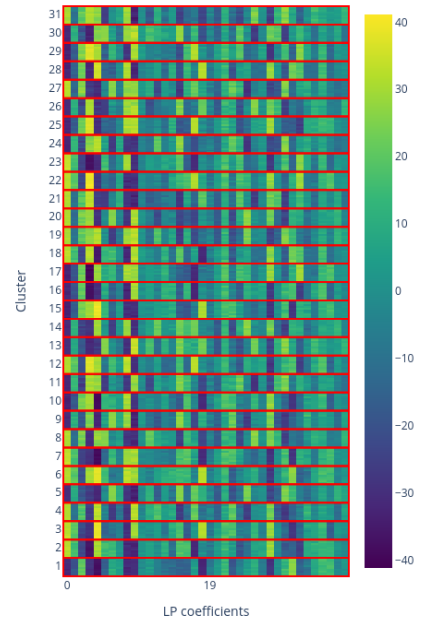
(b) DBSCAN clusters densities

Original LP coefficients cluster samples



(c) Original cluster samples

DBSCAN LP coefficients cluster samples



(d) DBSCAN cluster samples

Figure 7: Comparison between original clustering and DBSCAN clustering

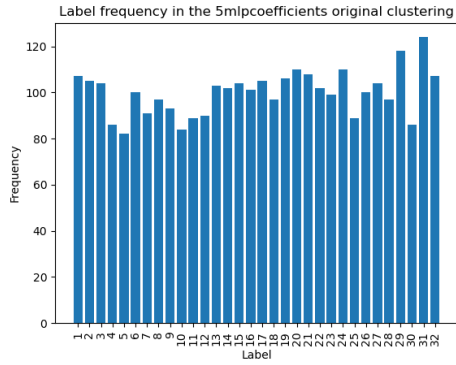
### 4.2.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

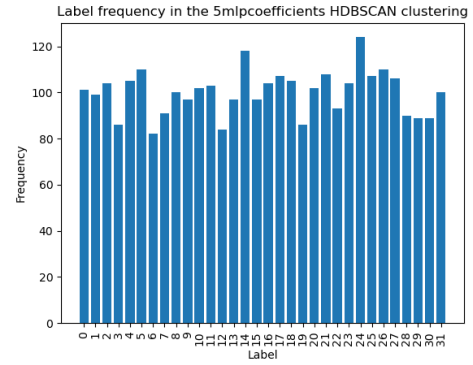
	<b>Minimum cluster size</b>
Original LP coefficients	21

Table 8: HDBSCAN hyperparameter configuration for LP coefficients clustering

The results are the following:

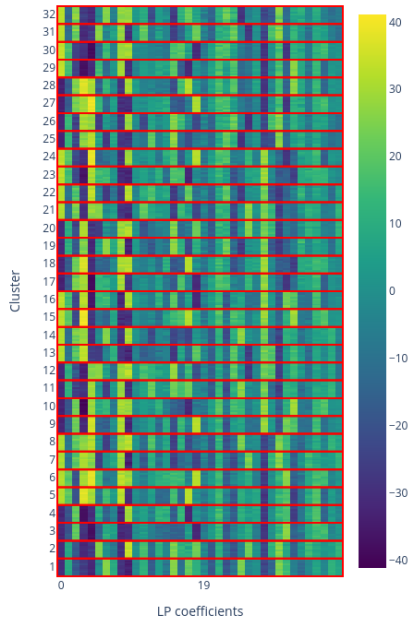


(a) Original cluster densities



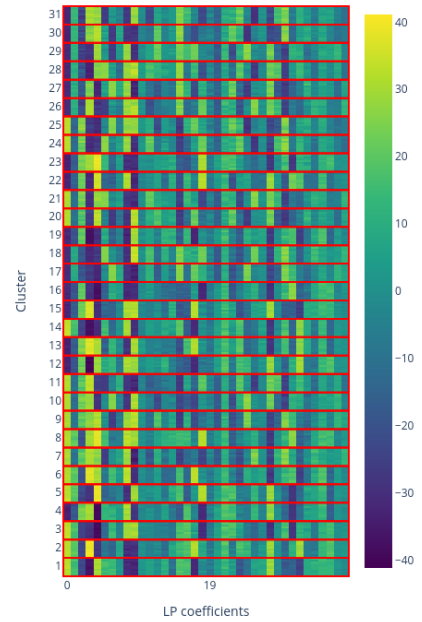
(b) HDBSCAN clusters densities

Original LP coefficients cluster samples



(c) Original cluster samples

HDBSCAN LP coefficients cluster samples



(d) HDBSCAN cluster samples

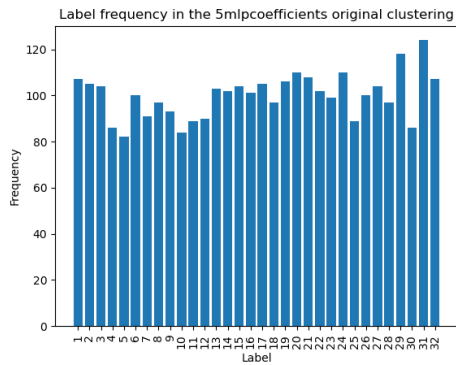
Figure 8: Comparison between original clustering and HDBSCAN clustering

#### 4.2.4 Agglomerative clustering

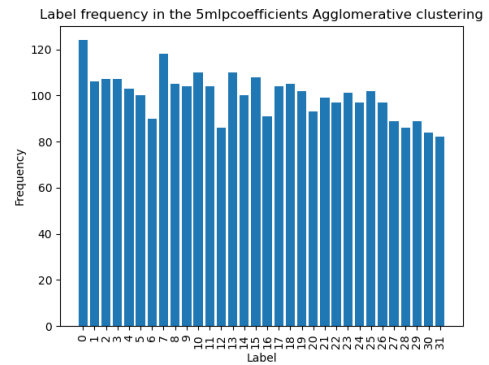
	<b>Number of clusters</b>
Original LP coefficients	4

Table 9: Agglomerative hyperparameter configuration for LP coefficients clustering

The results are the following:

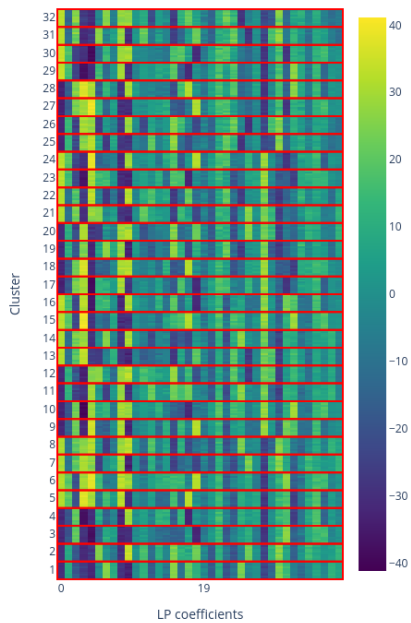


(a) Original cluster densities



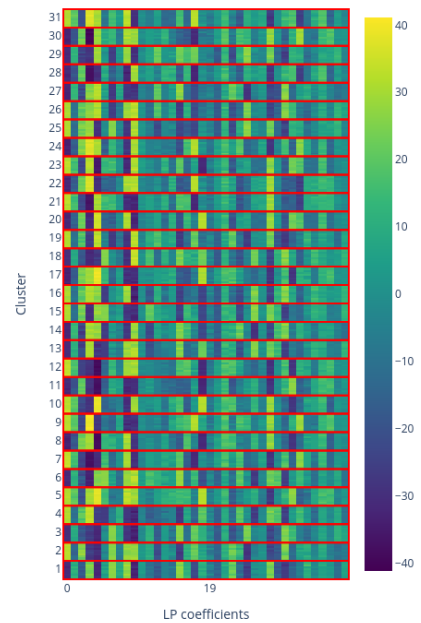
(b) Agglomerative clusters densities

Original LP coefficients cluster samples



(c) Original cluster samples

Agglomerative LP coefficients cluster samples



(d) Agglomerative cluster samples

Figure 9: Comparison between original clustering and Agglomerative clustering

#### 4.2.5 Summary

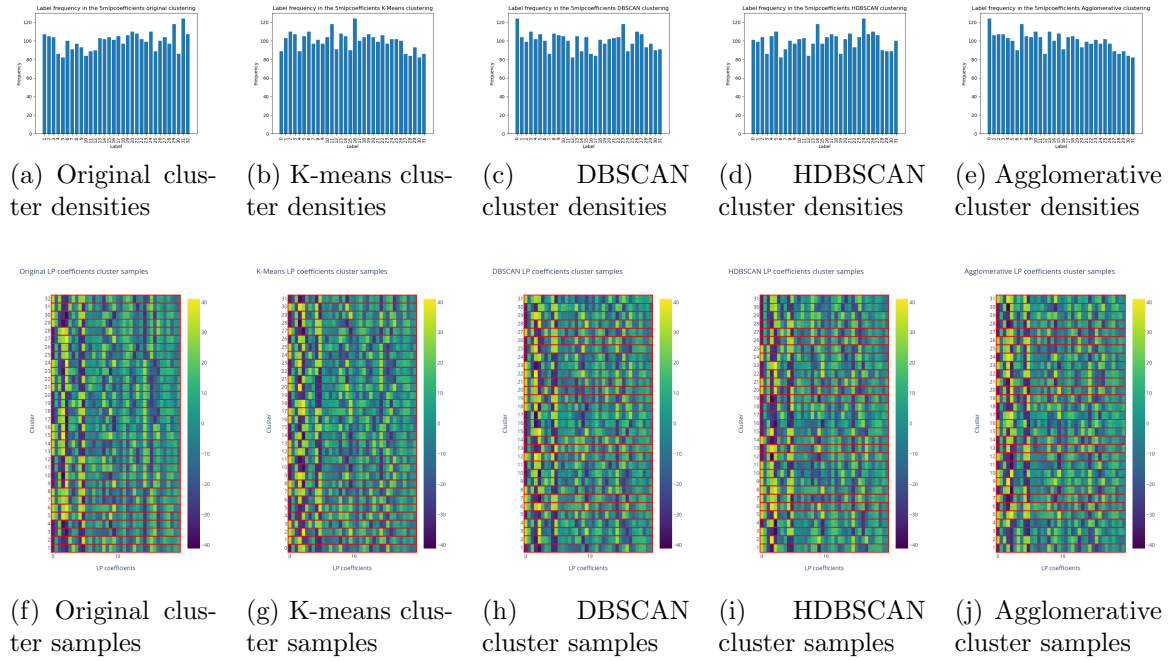


Figure 10: Comparison between clustering LP coefficients algorithms

	Original	K-Means	DBSCAN	HDBSCAN	Agglomerative
Original	\	1	1	1	1
K-Means		\	1	1	1
DBSCAN			\	1	1
HDBSCAN				\	1

Table 10: Normalized Mutual Information between original LP coefficients clusters

### 4.3 Output fluxes clustering

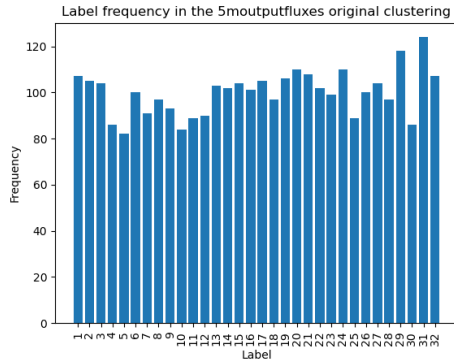
#### 4.3.1 K-Means

As K-Means allows for the number of clusters to be defined, and we know that there are 4 in the original dataset, K-Means is used to find 4 clusters.

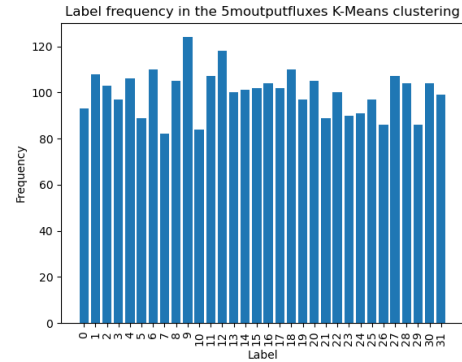
The results are the following:

	Number of clusters	Number of initializations
<b>Original Output fluxes</b>	32	100

Table 11: K-Means hyperparameter configuration for c coefficients clustering

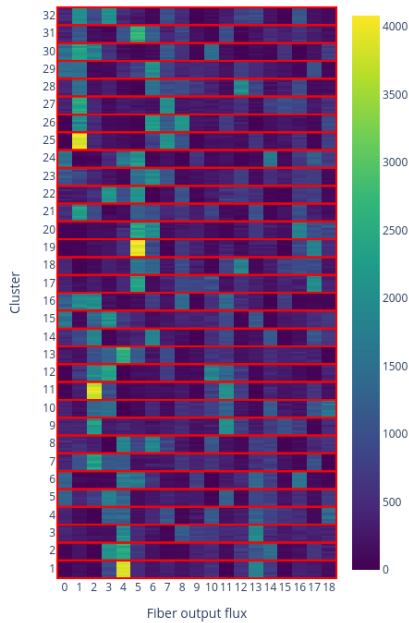


(a) Original cluster densities



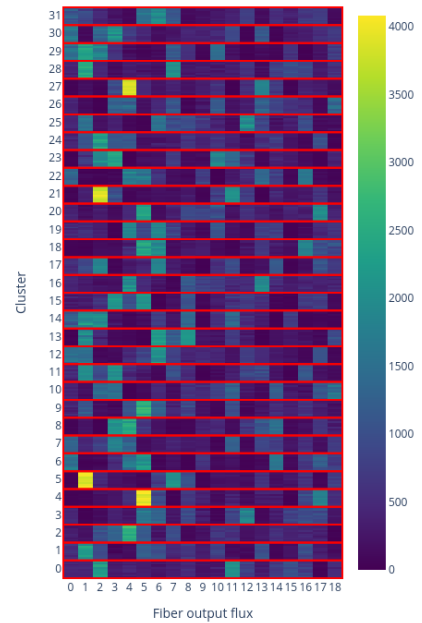
(b) K-Means clusters densities

Original output fluxes cluster samples



(c) Original cluster samples

K-Means output fluxes cluster samples



(d) K-Means cluster samples

Figure 11: Comparison between original clustering and K-Means clustering from original Output fluxes

### 4.3.2 DBSCAN

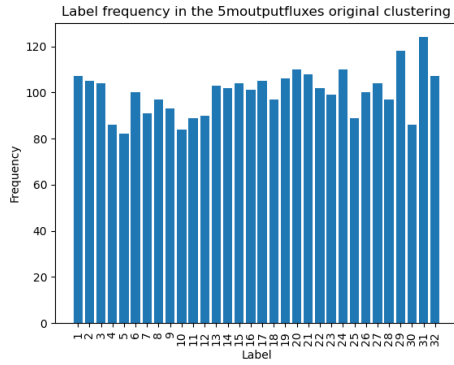
A configuration that outputs 4 clusters is searched

	<b>Number of neighbours</b>	<b>Epsilon</b>
Original Output fluxes	7	300

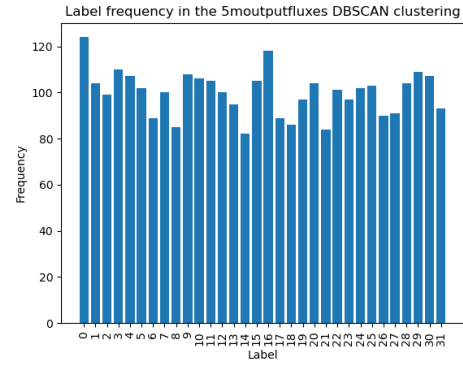
Table 12: DBSCAN hyperparameter configuration for Output fluxes clustering

The results are the following:



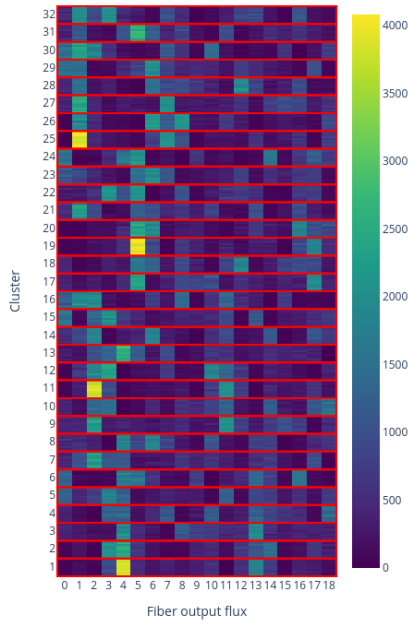


(a) Original cluster densities



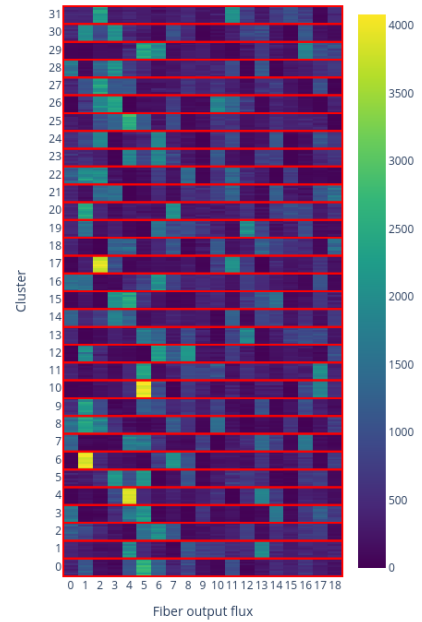
(b) DBSCAN clusters densities

Original output fluxes cluster samples



(c) Original cluster samples

DBSCAN output fluxes cluster samples



(d) DBSCAN cluster samples

Figure 12: Comparison between original clustering and DBSCAN clustering

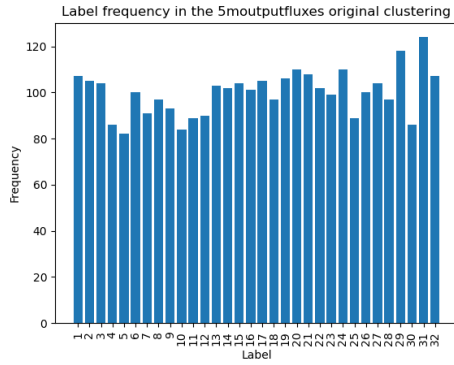
### 4.3.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

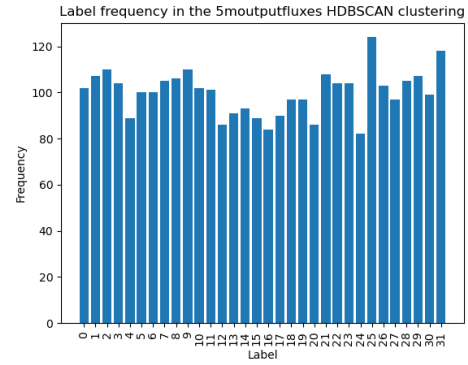
	<b>Minimum cluster size</b>
Original Output fluxes	21

Table 13: HDBSCAN hyperparameter configuration for Output fluxes clustering

The results are the following:

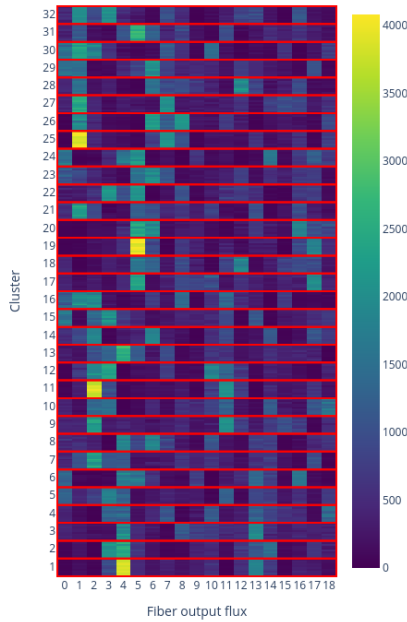


(a) Original cluster densities



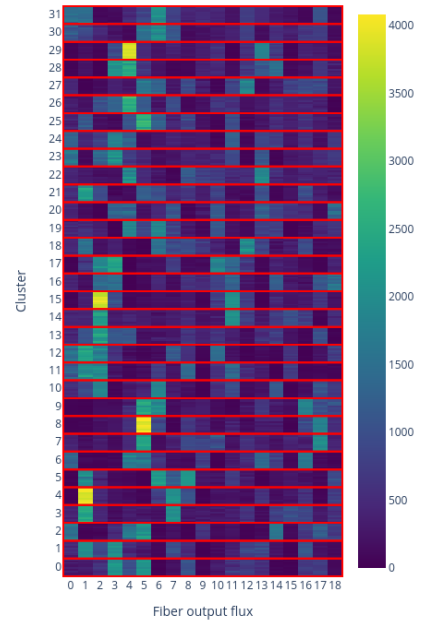
(b) HDBSCAN clusters densities

Original output fluxes cluster samples



(c) Original cluster samples

HDBSCAN output fluxes cluster samples



(d) HDBSCAN cluster samples

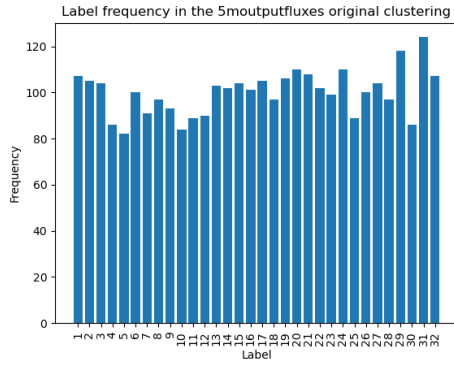
Figure 13: Comparison between original clustering and HDBSCAN clustering

#### 4.3.4 Agglomerative clustering

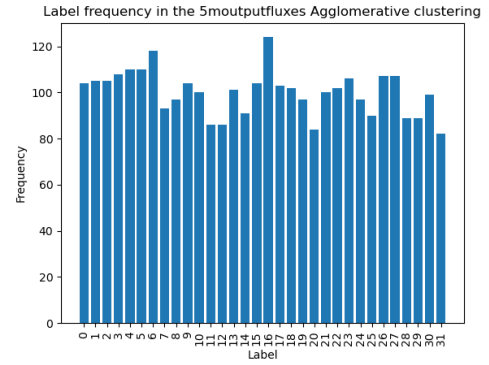
	<b>Number of clusters</b>
Original Output fluxes	32

Table 14: Agglomerative hyperparameter configuration for Output fluxes clustering

The results are the following:

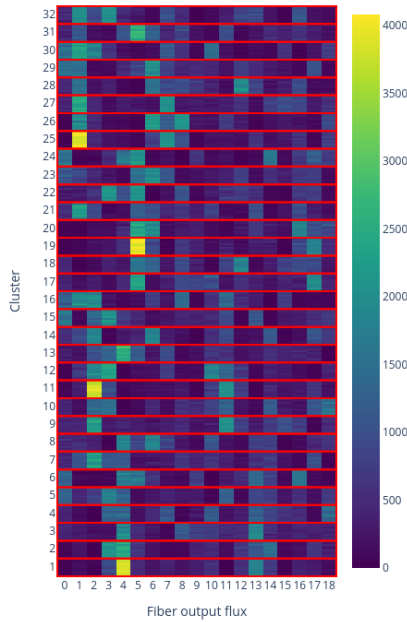


(a) Original cluster densities



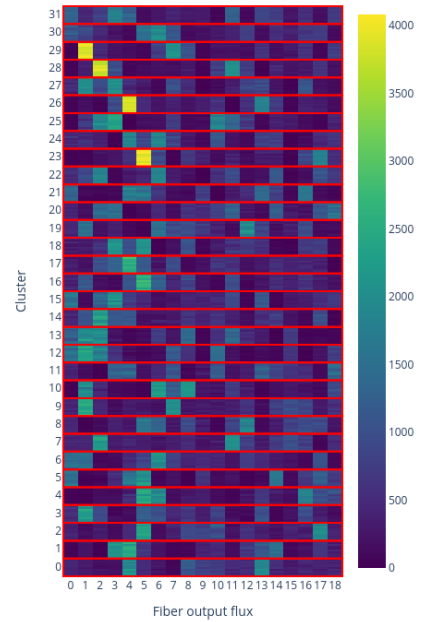
(b) Agglomerative clusters densities

Original output fluxes cluster samples



(c) Original cluster samples

Agglomerative output fluxes cluster samples



(d) Agglomerative cluster samples

Figure 14: Comparison between original clustering and Agglomerative clustering

#### 4.3.5 Summary

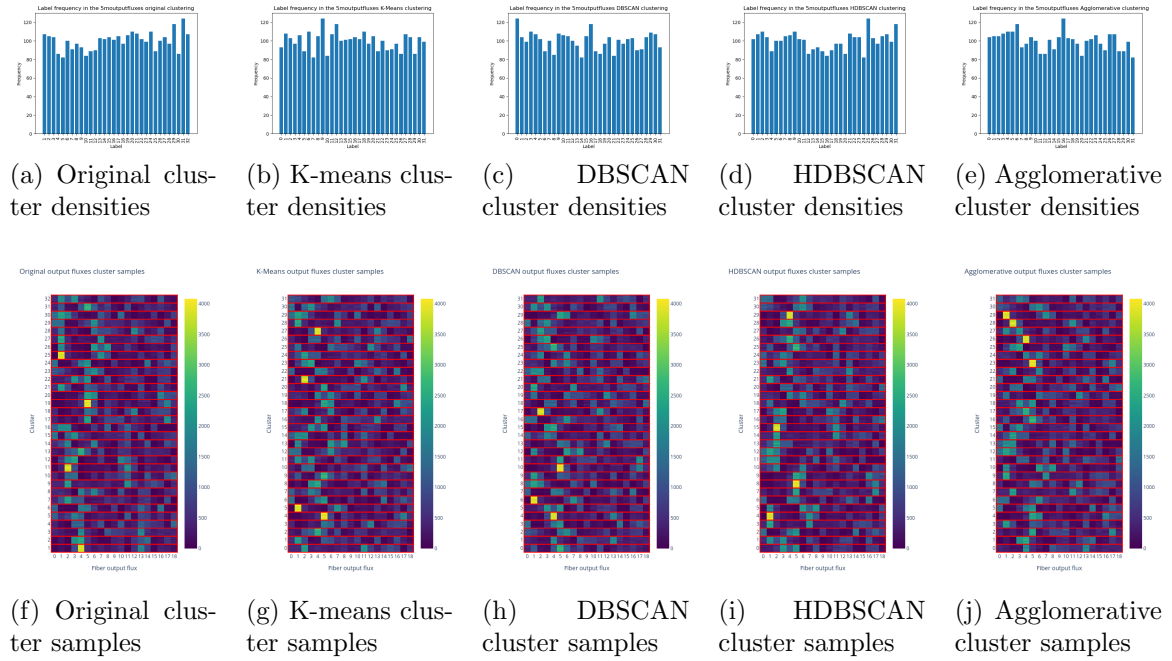


Figure 15: Comparison between clustering Output fluxes algorithms

	Original	K-Means	DBSCAN	HDBSCAN	Agglomerative
Original	\	1	0.998	1	1
K-Means		\	0.998	1	1
DBSCAN			\	0.998	0.998
HDBSCAN				\	1

Table 15: Normalized Mutual Information between original Output fluxes clusters

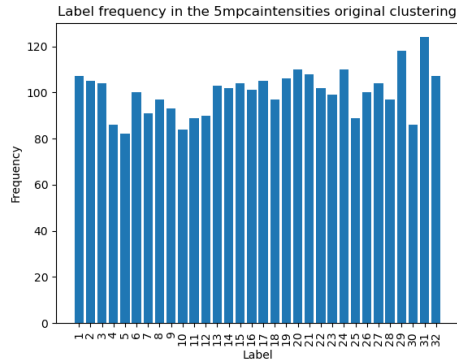
## 4.4 PSF Intensities clustering

### 4.4.1 K-Means

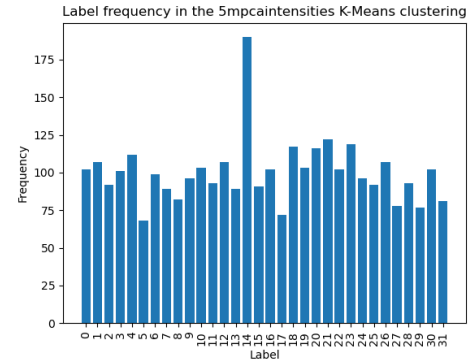
As K-Means allows for the number of clusters to be defined, and we know that there are 4 in the original dataset, K-Means is used to find 4 clusters.

	Number of clusters	Number of initializations
<b>PCA PSF Intensities</b>	4	10

Table 16: K-Means hyperparameter configuration for c coefficients clustering



(a) Original cluster densities from PCA



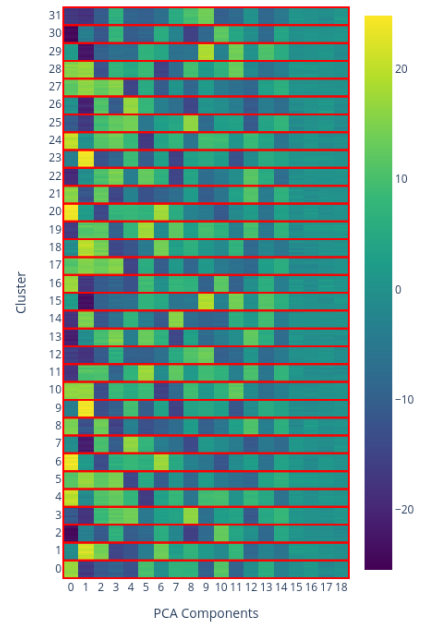
(b) K-Means clusters densities from PCA

Original PCA PSF intensities cluster samples



(c) Original cluster samples from PCA

K-Means PCA PSF Intensities cluster samples



(d) K-Means cluster samples from PCA

Figure 16: Comparison between original clustering and K-Means clustering from PCA of PSF Intensities

#### 4.4.2 DBSCAN

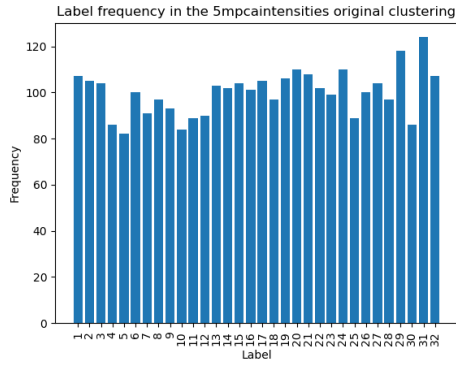
A configuration that outputs 4 clusters is searched

	Number of neighbours	Epsilon
PCA PSF Intensities	100	5

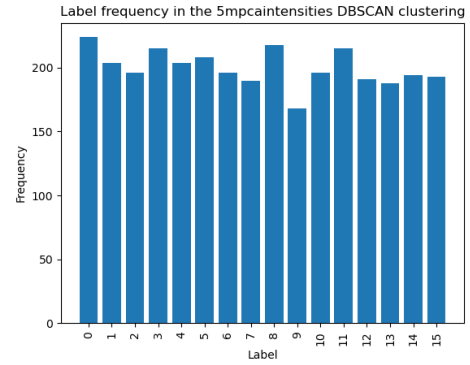
Table 17: DBSCAN hyperparameter configuration for PSF Intensities clustering

The results are the following:



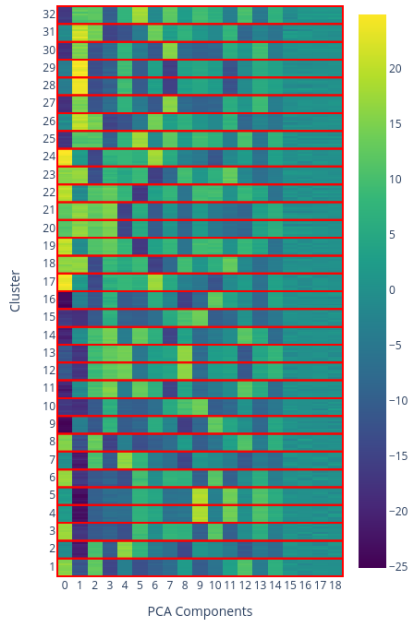


(a) Original cluster densities from PCA



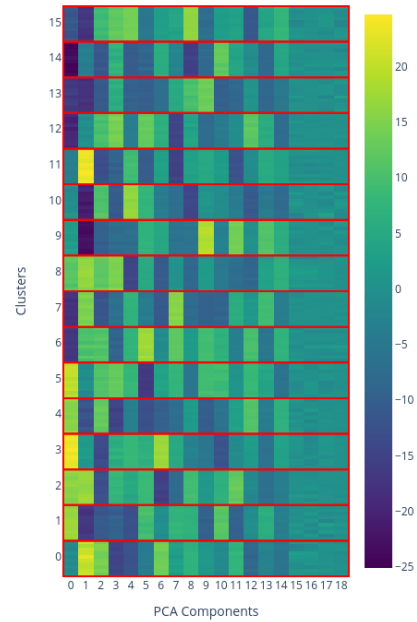
(b) DBSCAN clusters densities from PCA

Original PCA PSF intensities cluster samples



(c) Original cluster samples from PCA

DBSCAN PCA PSF Intensities cluster samples



(d) DBSCAN cluster samples from PCA

Figure 17: Comparison between original clustering and DBSCAN clustering from PCA of PSF Intensities

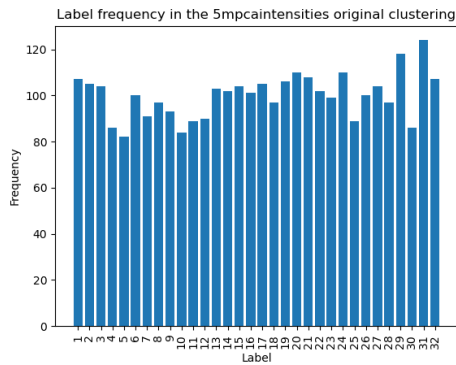
#### 4.4.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

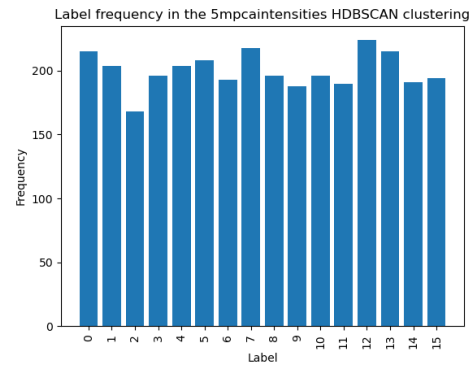
	Minimum cluster size
PCA PSF Intensities	100

Table 18: HDBSCAN hyperparameter configuration for PSF Intensities clustering

The results are the following:

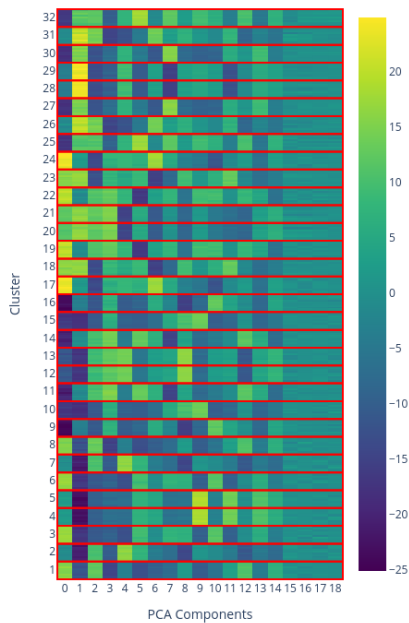


(a) Original cluster densities from PCA



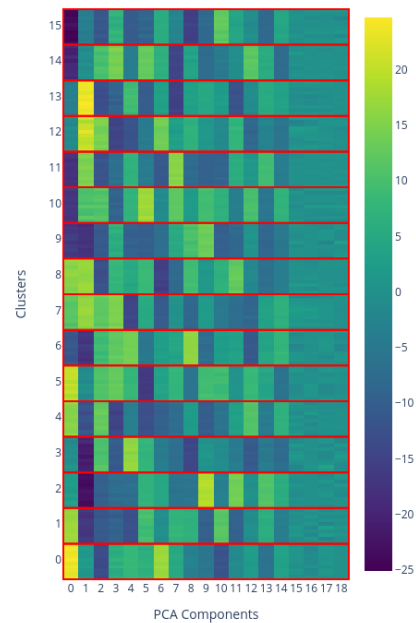
(b) HDBSCAN clusters densities from PCA

Original PCA PSF intensities cluster samples



(c) Original cluster samples from PCA

HDBSCAN PCA PSF Intensities cluster samples



(d) HDBSCAN cluster samples from PCA

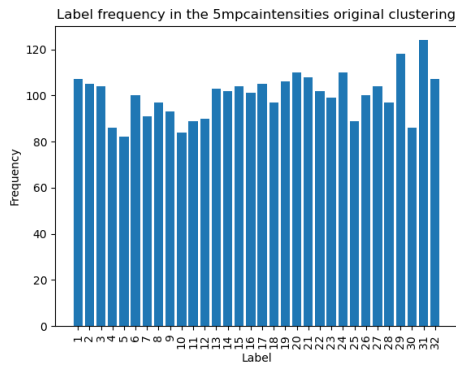
Figure 18: Comparison between original clustering and HDBSCAN clustering from PCA of PSF Intensities

#### 4.4.4 Agglomerative clustering

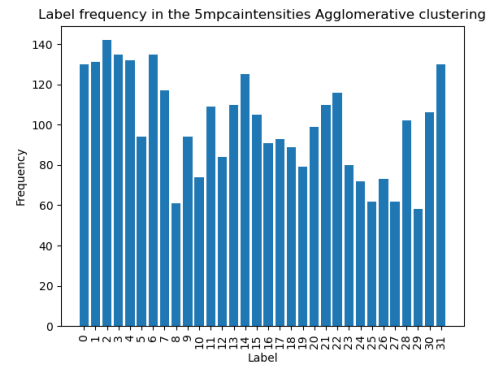
	<b>Number of clusters</b>
PCA PSF Intensities	32

Table 19: Agglomerative hyperparameter configuration for PSF Intensities clustering

The results are the following:

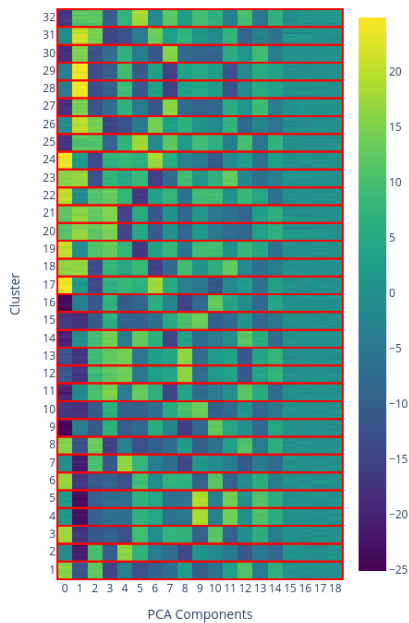


(a) Original cluster densities from PCA



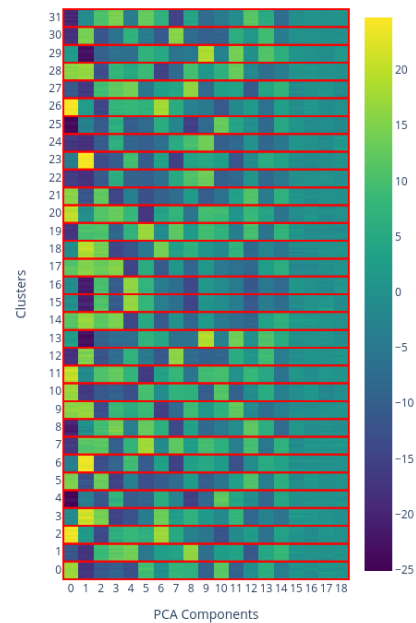
(b) Agglomerative clusters densities from PCA

Original PCA PSF intensities cluster samples



(c) Original cluster samples from PCA

Agglomerative PCA PSF Intensities cluster samples



(d) Agglomerative cluster samples from PCA

Figure 19: Comparison between original clustering and Agglomerative clustering

#### 4.4.5 Summary

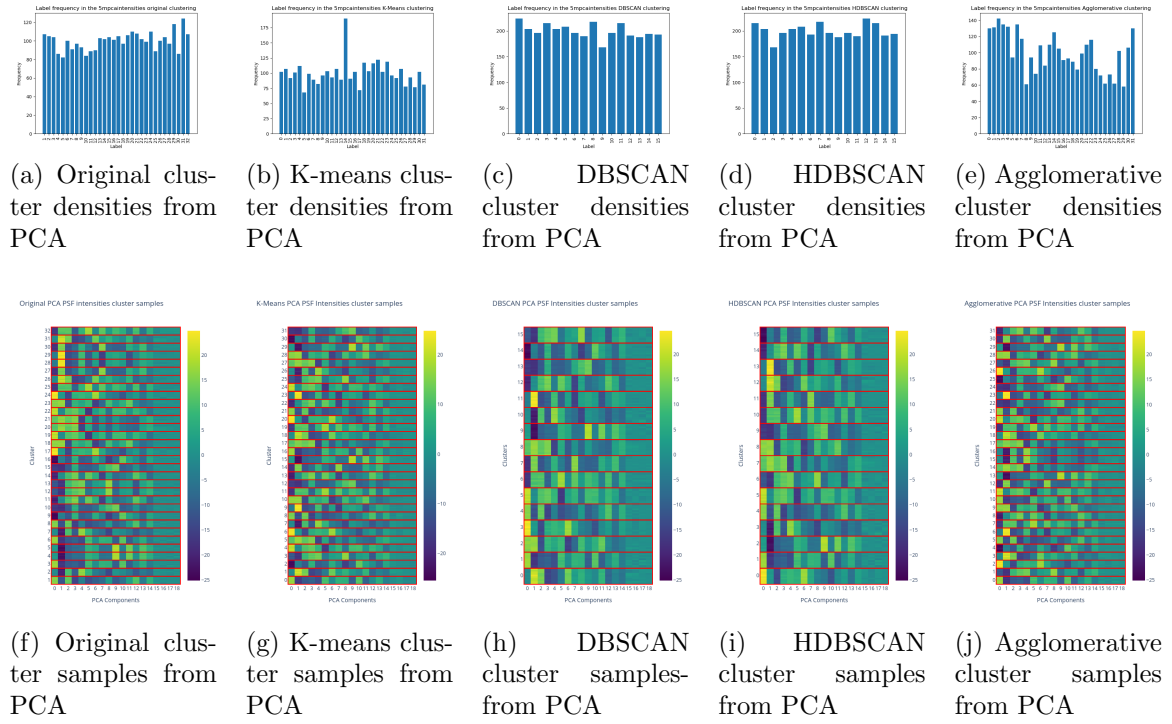


Figure 20: Comparison between clustering PCA PSF Intensities algorithms

	Original	K-Means	DBSCAN	HDBSCAN	Agglomerative
Original	\	0.802	0.889	0.889	0.803
K-Means		\	0.891	0.891	0.909
DBSCAN			\	1	0.892
HDBSCAN				\	0.892

Table 20: Normalized Mutual Information between PCA PSF Intensities clusters