Lab Notebook

Photonic Lantern Information Determination

Contents

1 Mini Dataset Information Determination

2 The data

2.1 Zernike coefficients dataset

A dataset of 6400 zernike coefficients is created for this report.In particular, each datapoint represent 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 [0.8, 1]
- Modes 7, 8, 9 and 10 between [-0.5, -0.3] and [0.3, 0.5]

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

2.2 PSFs intensities dataset

A dataset of 6400 PSFs is created using the Zernike coefficients dataset.

2.3 LP mode coefficients dataset

A dataset of 6400 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 6400 PL output fluxes obtained from the PL transfer matrix and LP coefficients.

3 Preprocessing

3.1 PSF Intensities

The 6400x128x128 array is dimensionally reduced using PCA and UMAP both giving an array of 6400x19 projections of the PSF Intensities.

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

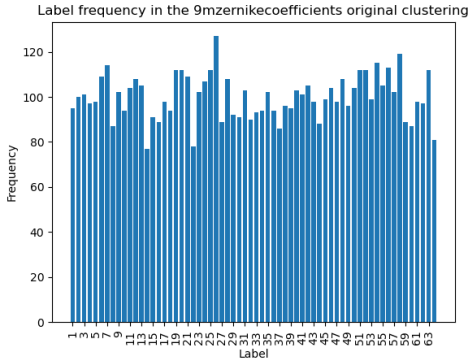
4.1.1 K-Means

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.

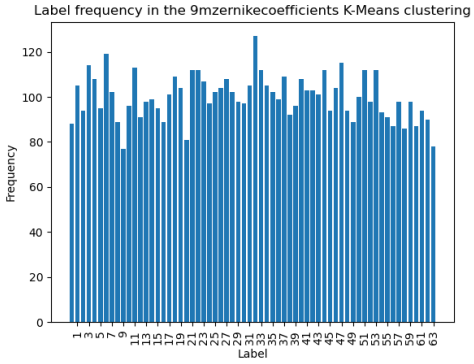
Number of clusters	**Number of initializations**
64	100

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

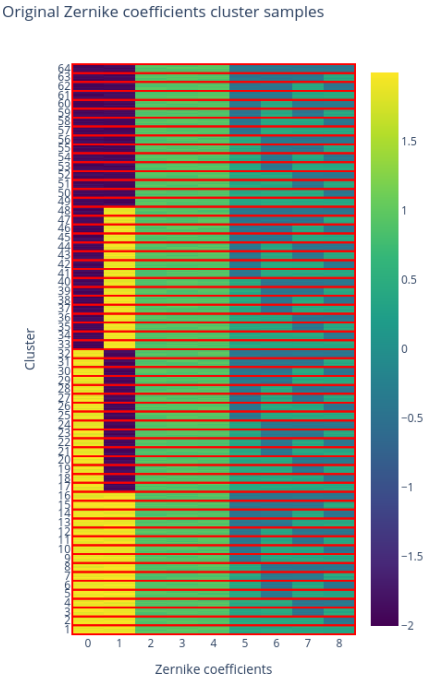
The results are the following:



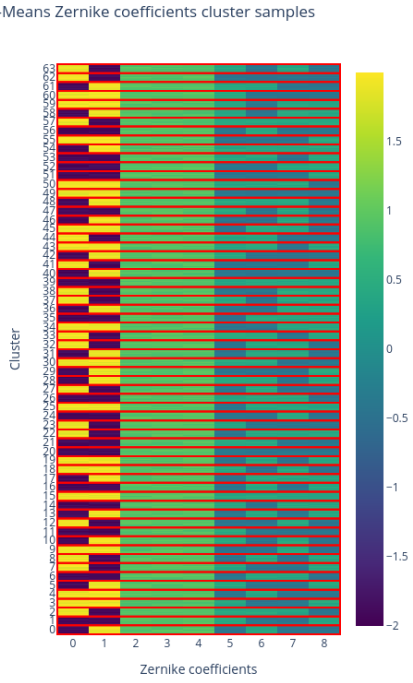
(a) Original cluster densities



(b) K-Means clusters densities



(c) Original cluster samples



(d) K-Means cluster samples

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

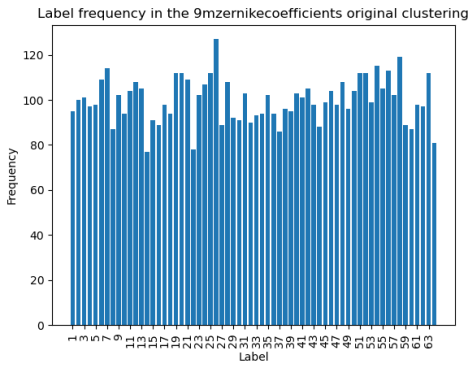
4.1.2 DBSCAN

A configuration that outputs 4 clusters is searched

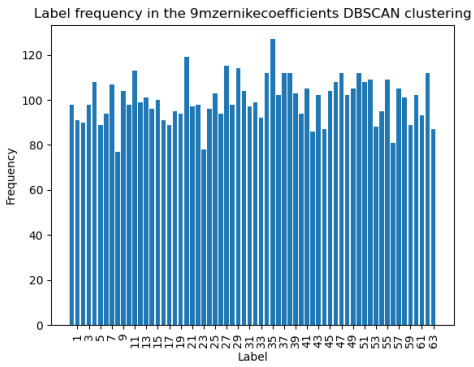
Number of neighbours	**Epsilon**
5	0.2

Table 2: DBSCAN hyperparameter configuration for Zernike coefficients clustering

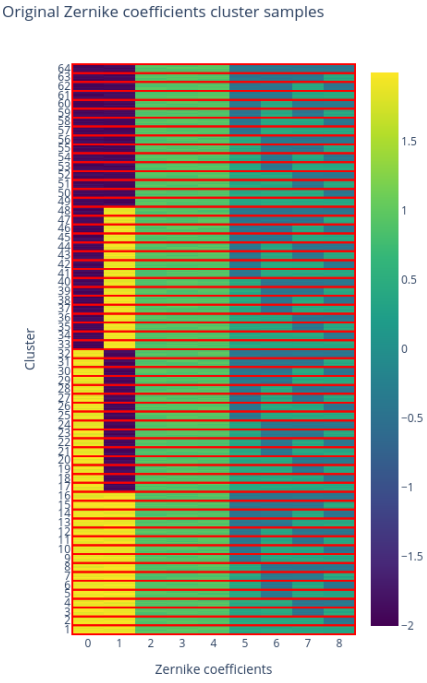
The results are the following:



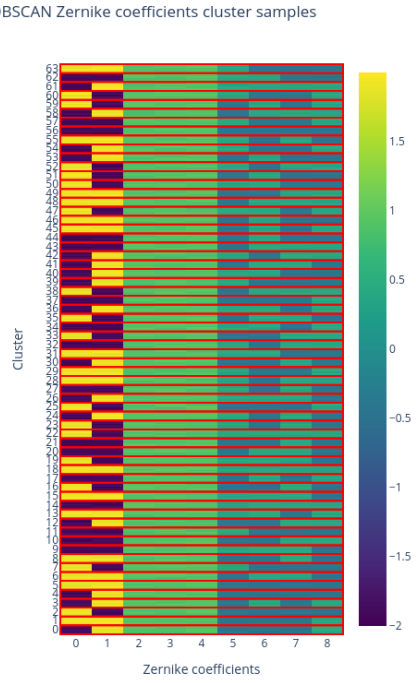
(a) Original cluster densities



(b) DBSCAN clusters densities



(c) Original cluster samples



(d) DBSCAN cluster samples

Figure 2: Comparison between original clustering and DBSCAN clustering

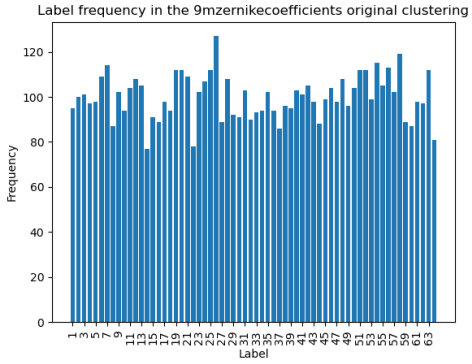
4.1.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

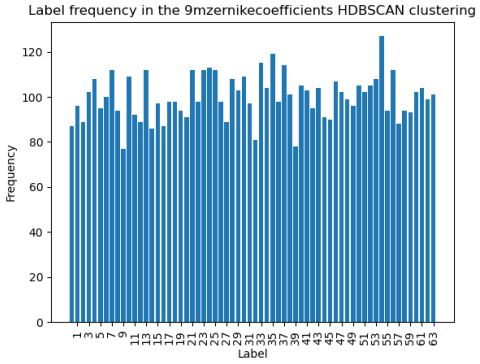
Minimum cluster size
50

Table 3: HDBSCAN hyperparameter configuration for Zernike coefficients clustering

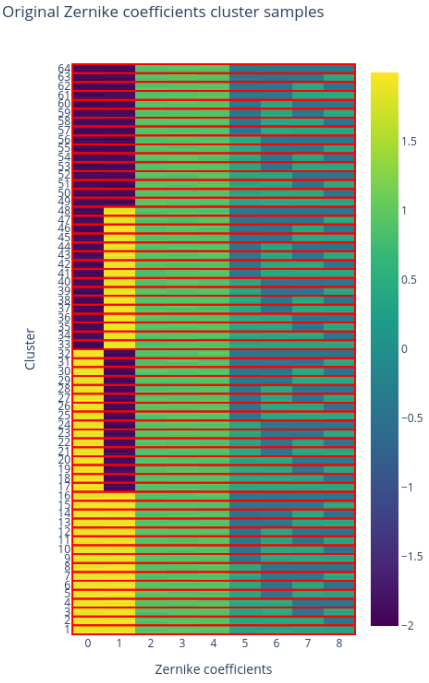
The results are the following:



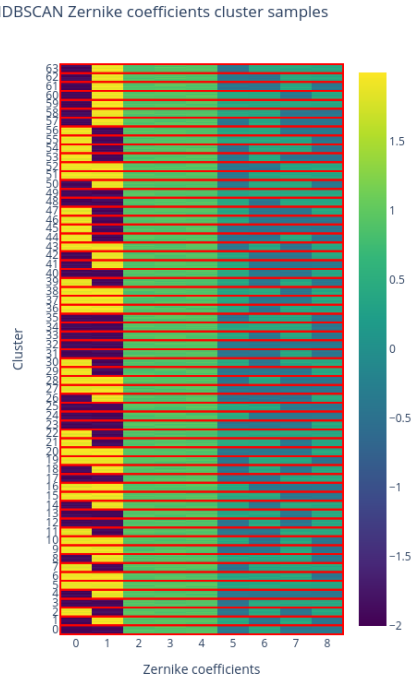
(a) Original cluster densities



(b) HDBSCAN clusters densities



(c) Original cluster samples



(d) HDBSCAN cluster samples

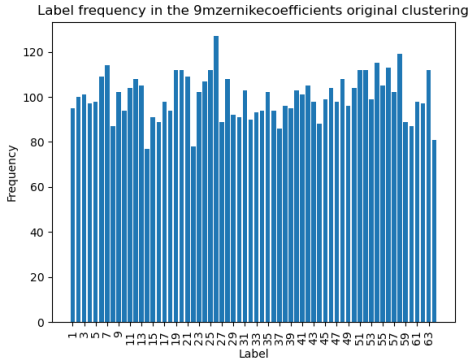
Figure 3: Comparison between original clustering and HDBSCAN clustering

4.1.4 Agglomerative clustering

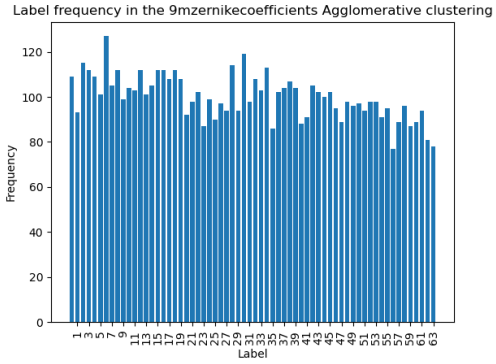
Number of clusters
64

Table 4: Agglomerative hyperparameter configuration for Zernike coefficients clustering

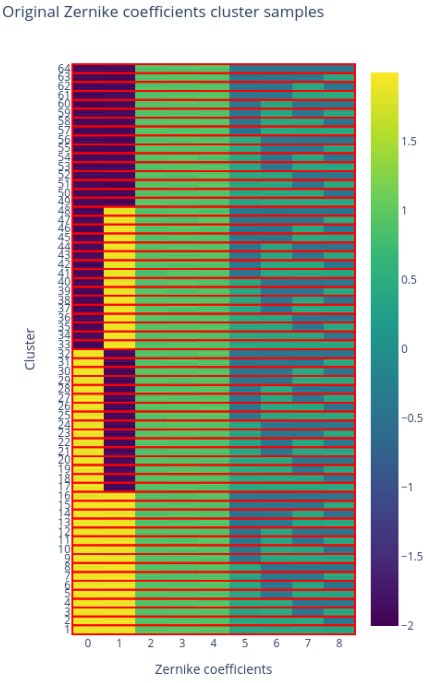
The results are the following:



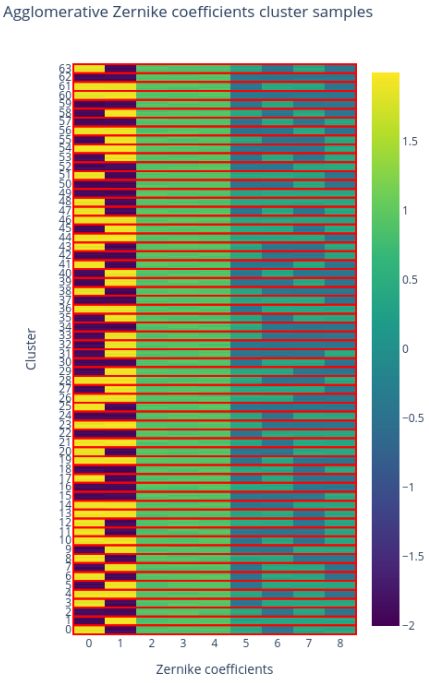
(a) Original cluster densities



(b) Agglomerative clusters densities



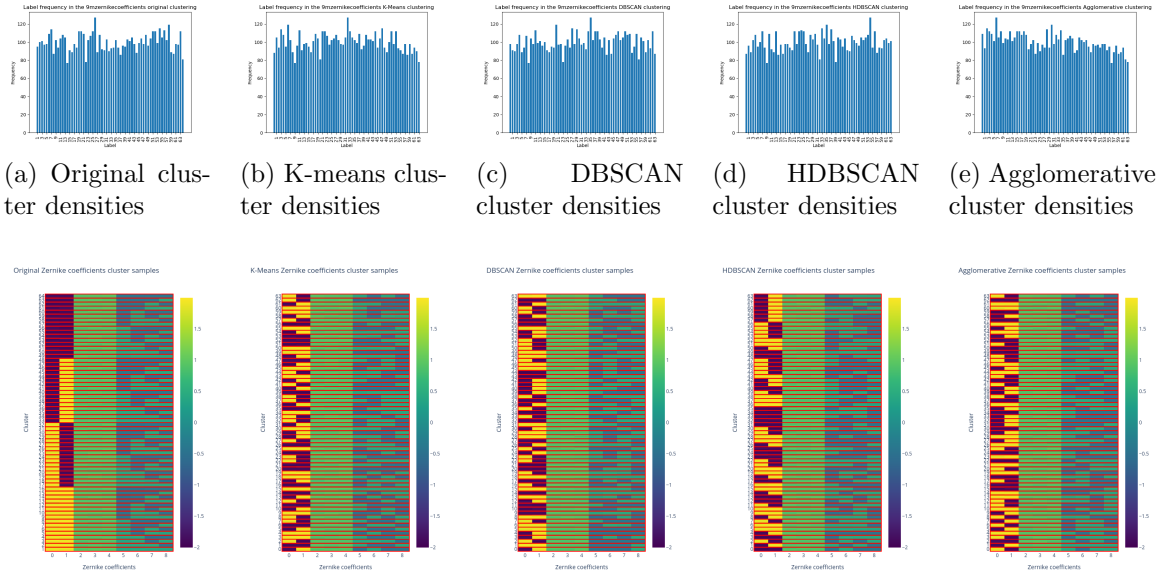
(c) Original cluster samples



(d) Agglomerative cluster samples

Figure 4: Comparison between original clustering and Agglomerative clustering

4.1.5 Summary



(a) Original cluster densities (b) K-means cluster densities (c) DBSCAN cluster densities (d) HDBSCAN cluster densities (e) Agglomerative cluster densities

(f) Original cluster samples (g) K-means cluster samples (h) DBSCAN cluster samples (i) HDBSCAN cluster samples (j) Agglomerative cluster samples

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

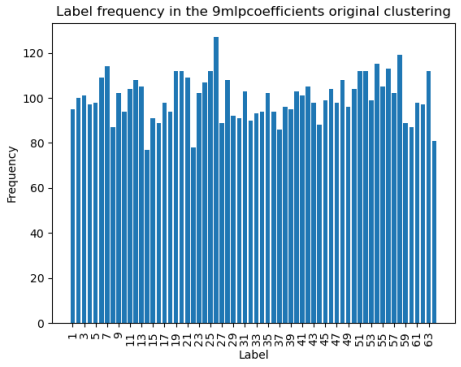
4.2.1 K-Means

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.

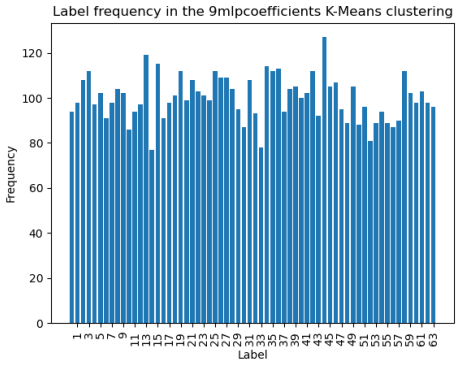
The results are the following:

	Number of clusters	**Number of initializations**
Original LP coefficients	32	100

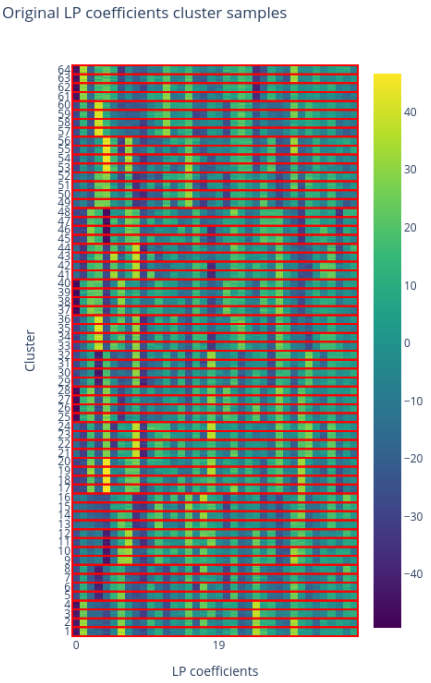
Table 6: K-Means hyperparameter configuration for c coefficients clustering



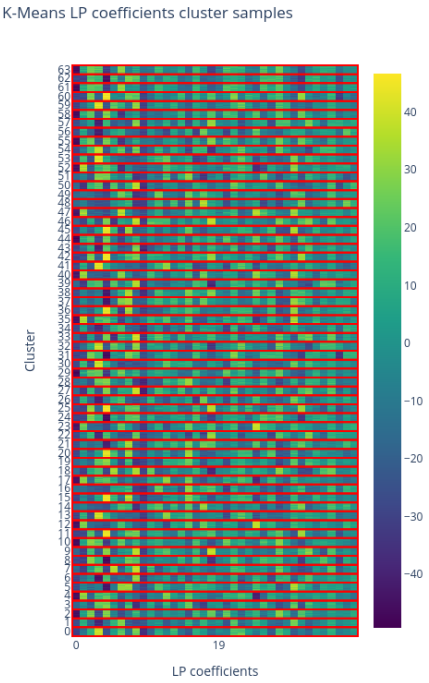
(a) Original cluster densities



(b) K-Means clusters densities



(c) Original cluster samples



(d) K-Means cluster samples

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

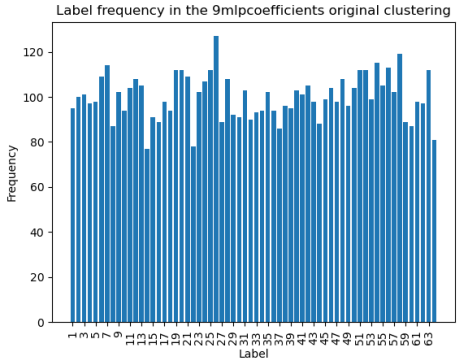
4.2.2 DBSCAN

A configuration that outputs 4 clusters is searched

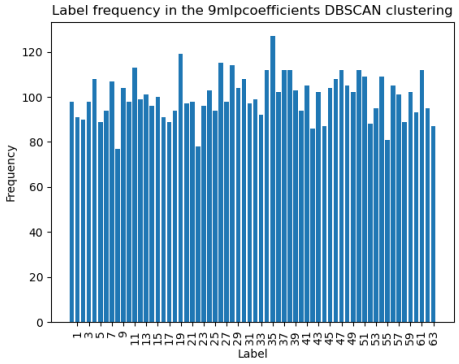
	Number of neighbours	**Epsilon**
Original LP coefficients	15	18

Table 7: DBSCAN hyperparameter configuration for LP coefficients clustering

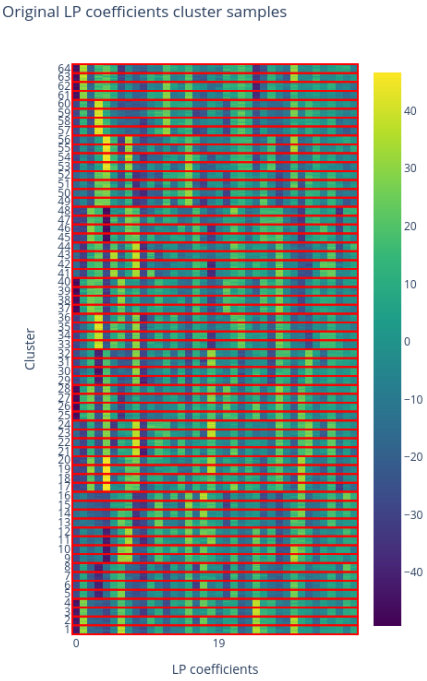
The results are the following:



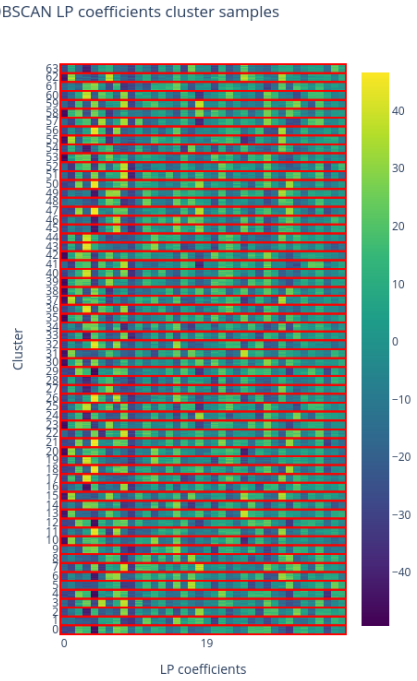
(a) Original cluster densities



(b) DBSCAN clusters densities



(c) Original cluster samples



(d) DBSCAN cluster samples

Figure 7: Comparison between original clustering and DBSCAN clustering

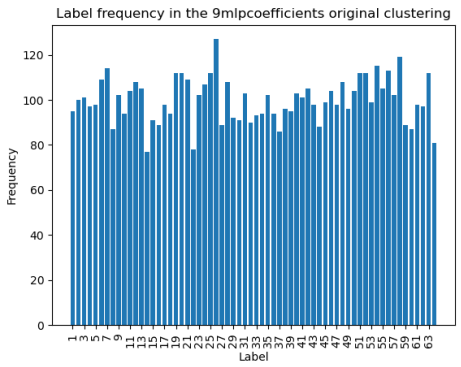
4.2.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

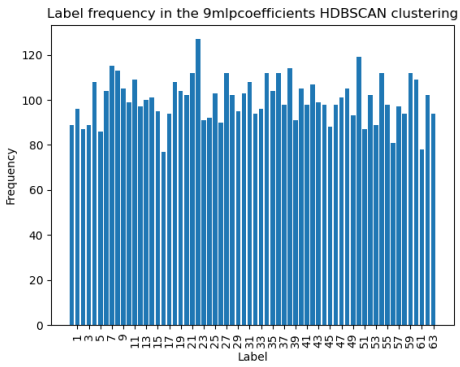
	Minimum cluster size
Original LP coefficients	21

Table 8: HDBSCAN hyperparameter configuration for LP coefficients clustering

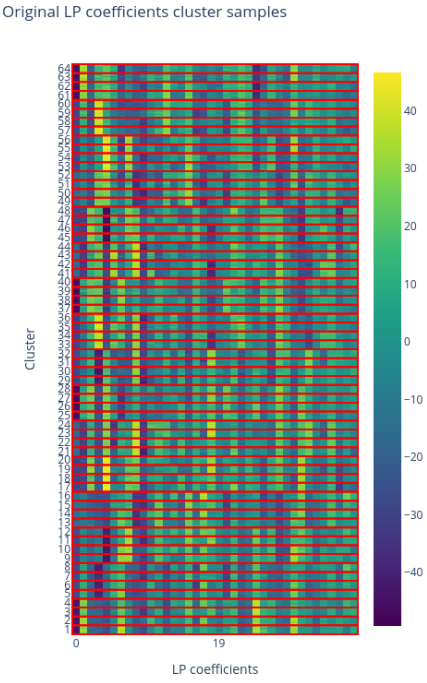
The results are the following:



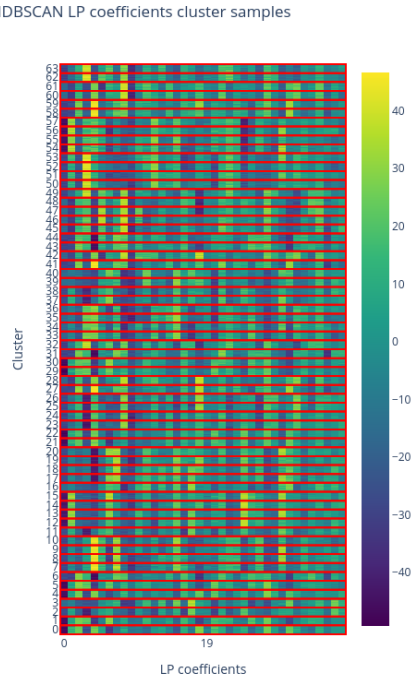
(a) Original cluster densities



(b) HDBSCAN clusters densities



(c) Original cluster samples



(d) HDBSCAN cluster samples

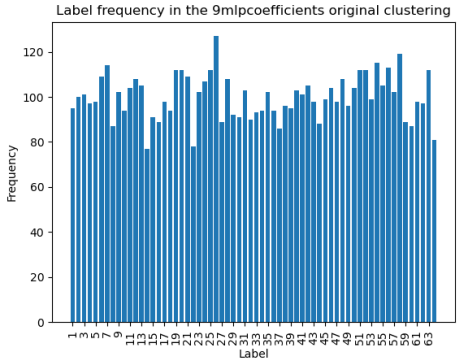
Figure 8: Comparison between original clustering and HDBSCAN clustering

4.2.4 Agglomerative clustering

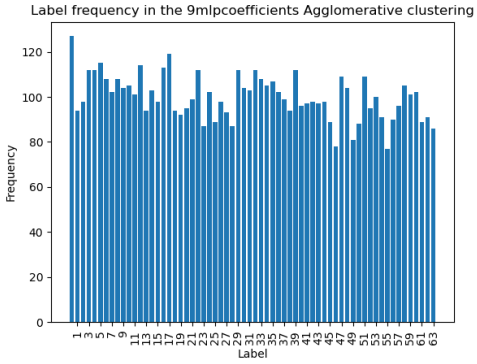
	Number of clusters
Original LP coefficients	64

Table 9: Agglomerative hyperparameter configuration for LP coefficients clustering

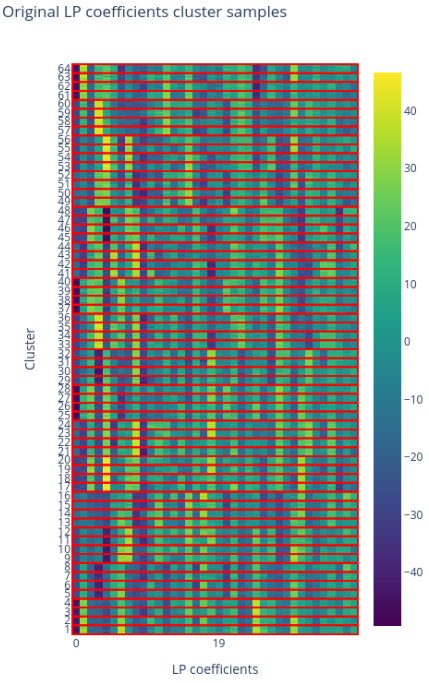
The results are the following:



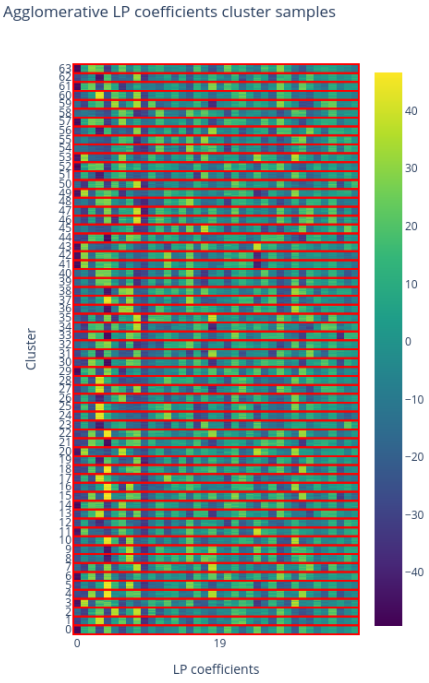
(a) Original cluster densities



(b) Agglomerative clusters densities



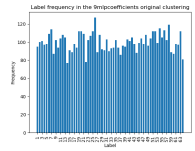
(c) Original cluster samples



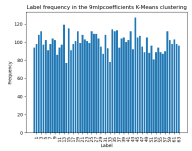
(d) Agglomerative cluster samples

Figure 9: Comparison between original clustering and Agglomerative clustering

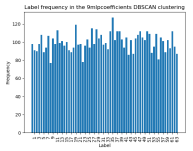
4.2.5 Summary



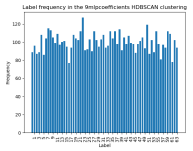
(a) Original cluster densities



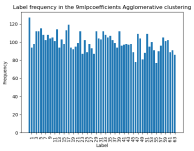
(b) K-means cluster densities



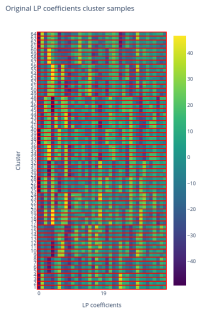
(c) DBSCAN cluster densities



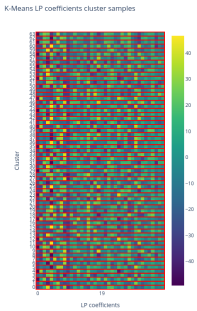
(d) HDBSCAN cluster densities



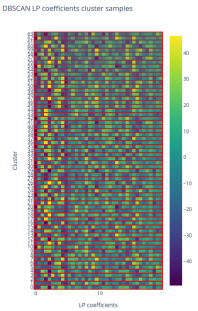
(e) Agglomerative cluster densities



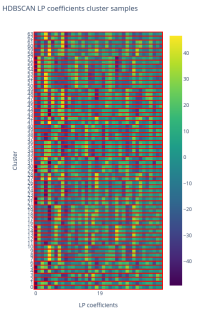
(f) Original cluster samples



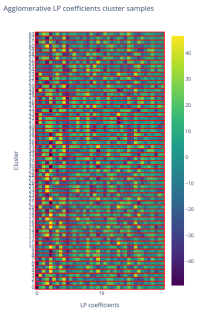
(g) K-means cluster samples



(h) DBSCAN cluster samples



(i) HDBSCAN cluster samples



(j) Agglomerative cluster samples

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

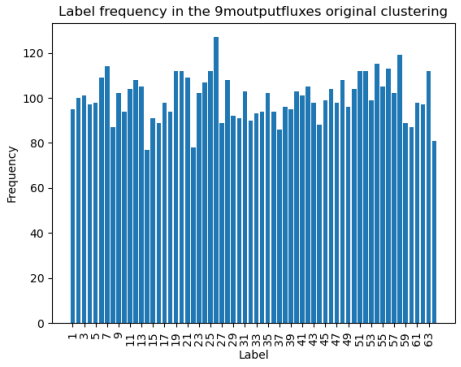
4.3.1 K-Means

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.

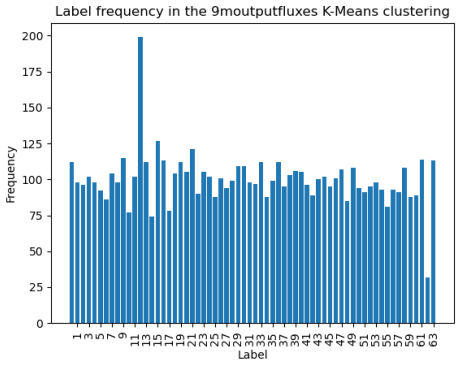
The results are the following:

	Number of clusters	Number of initializations
Original Output fluxes	64	100

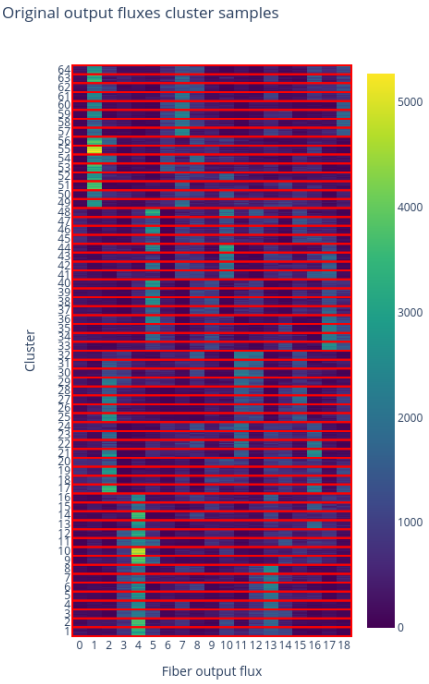
Table 11: K-Means hyperparameter configuration for c coefficients clustering



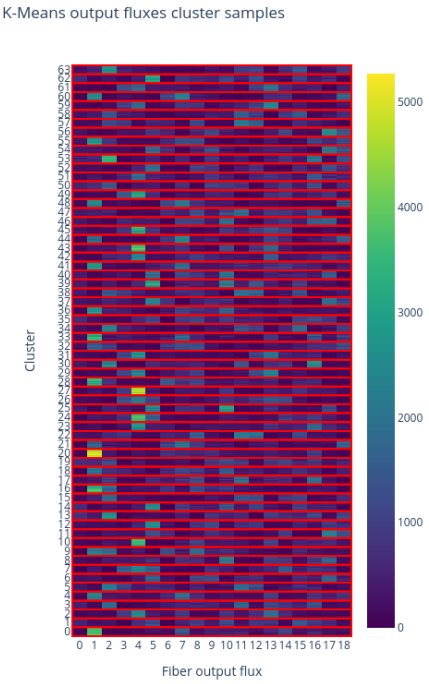
(a) Original cluster densities



(b) K-Means clusters densities



(c) Original cluster samples



(d) K-Means cluster samples

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

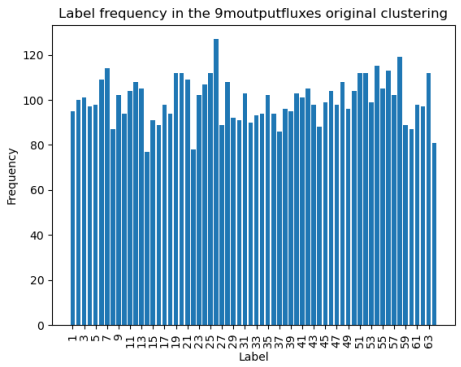
4.3.2 DBSCAN

A configuration that outputs 4 clusters is searched

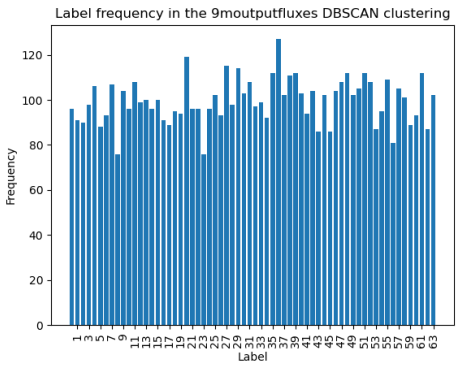
	Number of neighbours	**Epsilon**
Original Output fluxes	3	400

Table 12: DBSCAN hyperparameter configuration for Output fluxes clustering

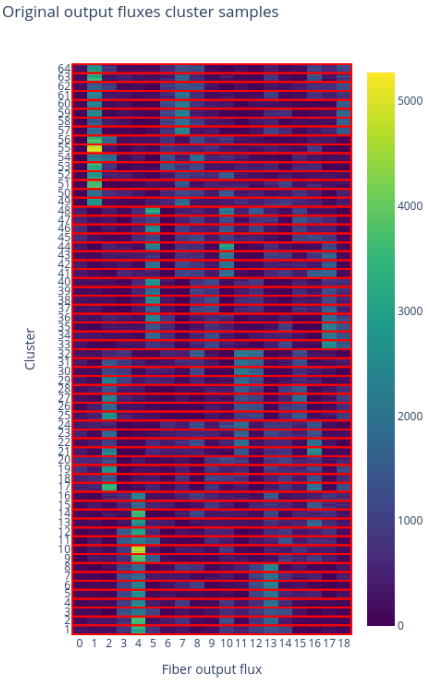
The results are the following:



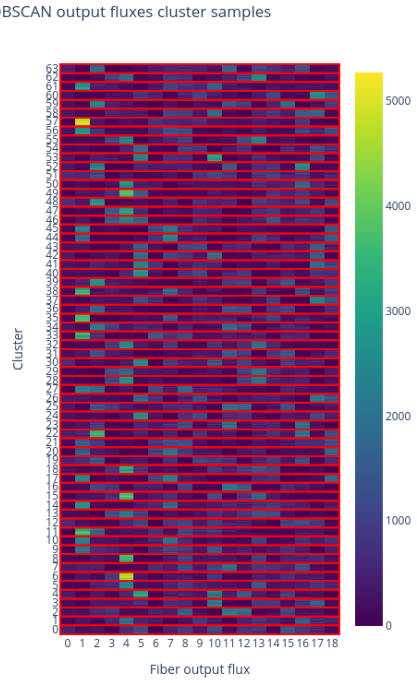
(a) Original cluster densities



(b) DBSCAN clusters densities



(c) Original cluster samples



(d) DBSCAN cluster samples

Figure 12: Comparison between original clustering and DBSCAN clustering

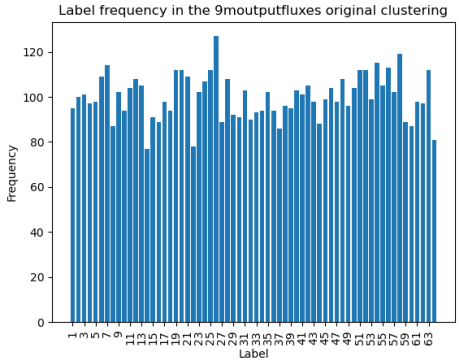
4.3.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

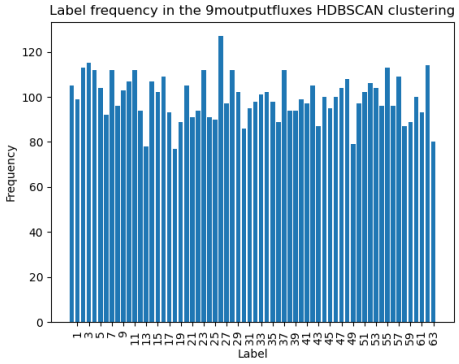
	Minimum cluster size
Original Output fluxes	10

Table 13: HDBSCAN hyperparameter configuration for Output fluxes clustering

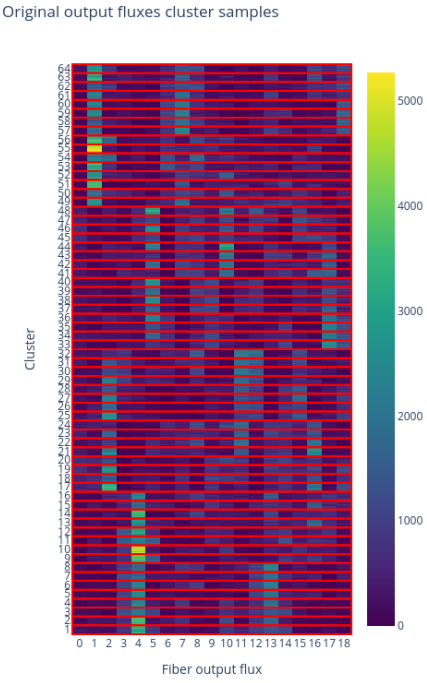
The results are the following:



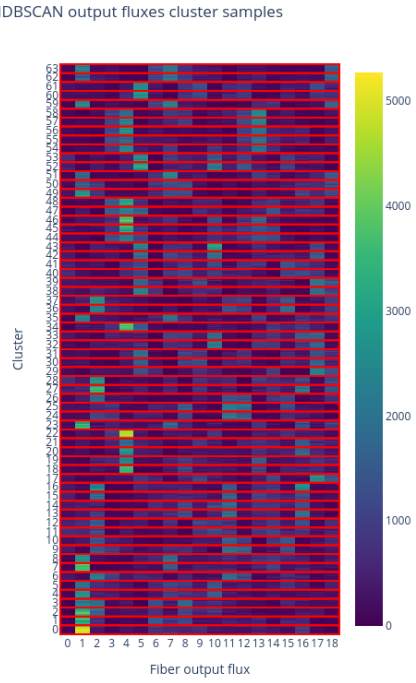
(a) Original cluster densities



(b) HDBSCAN clusters densities



(c) Original cluster samples



(d) HDBSCAN cluster samples

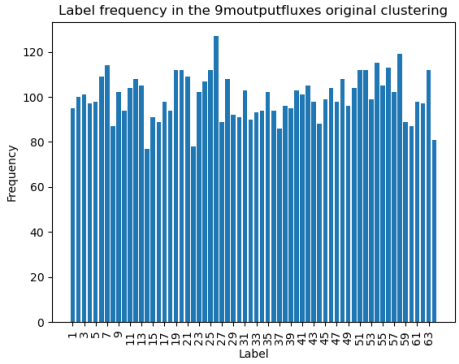
Figure 13: Comparison between original clustering and HDBSCAN clustering

4.3.4 Agglomerative clustering

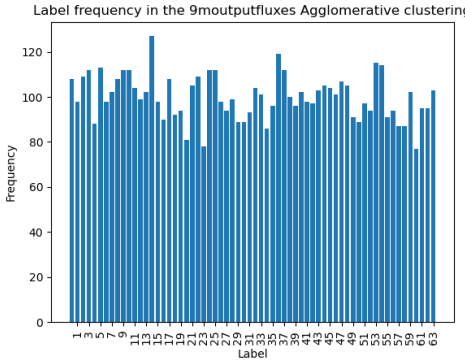
	Number of clusters
Original Output fluxes	64

Table 14: Agglomerative hyperparameter configuration for Output fluxes clustering

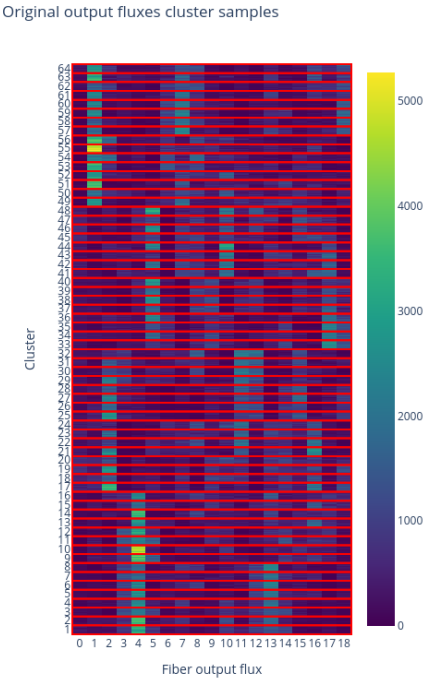
The results are the following:



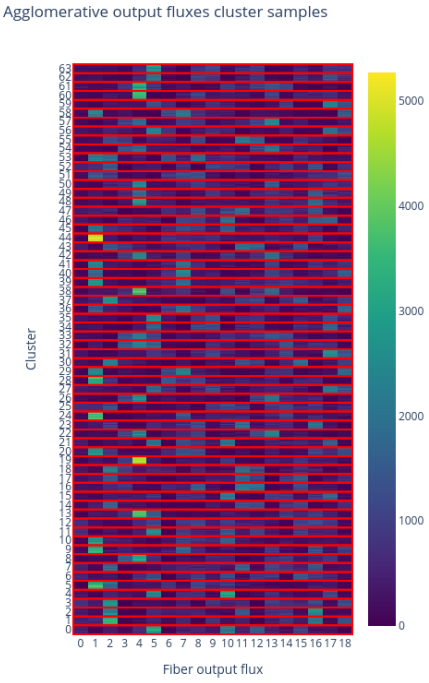
(a) Original cluster densities



(b) Agglomerative clusters densities



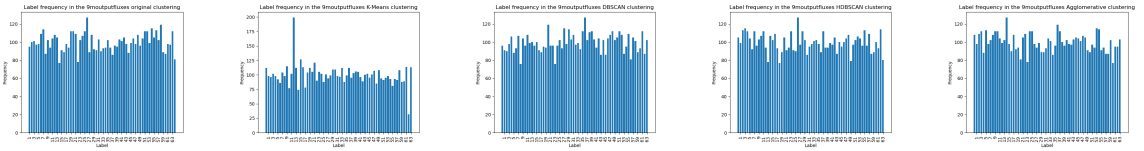
(c) Original cluster samples



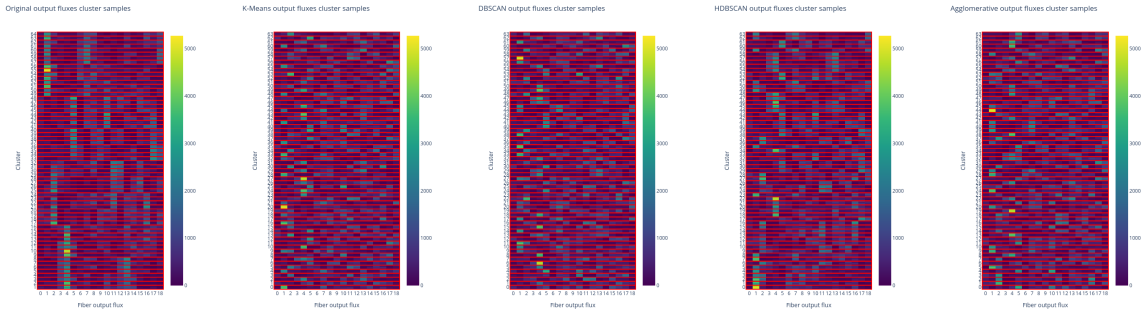
(d) Agglomerative cluster samples

Figure 14: Comparison between original clustering and Agglomerative clustering

4.3.5 Summary



(a) Original cluster densities (b) K-means cluster densities (c) DBSCAN cluster densities (d) HDBSCAN cluster densities (e) Agglomerative cluster densities



(f) Original cluster samples (g) K-means cluster samples (h) DBSCAN cluster samples (i) HDBSCAN cluster samples (j) Agglomerative cluster samples

Figure 15: Comparison between clustering Output fluxes algorithms

	Original	**K-Means**	**DBSCAN**	**HDBSCAN**	**Agglomerative**
Original	\	0.994	0.995	0.995	1
K-Means		\	0.990	0.991	0.994
DBSCAN			\	0.993	0.995
HDBSCAN				\	0.995

Table 15: Normalized Mutual Information between original Output fluxes clusters

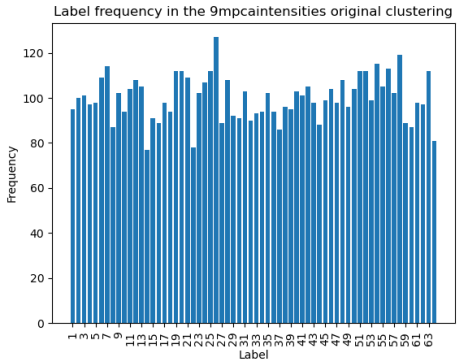
4.4 PSF Intensities clustering

4.4.1 K-Means

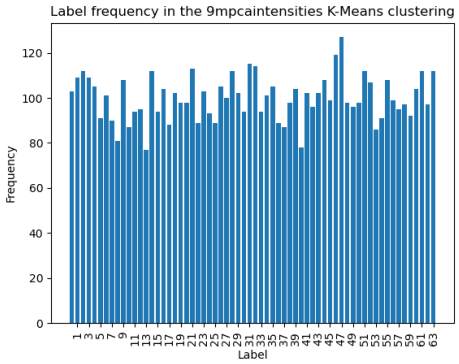
As K-Means allows for the number of clusters to be defined, and we know that there are 4 in the original dataset, K-Means is used to find 4 clusters.

	Number of clusters	**Number of initializations**
PCA PSF Intensities	64	100

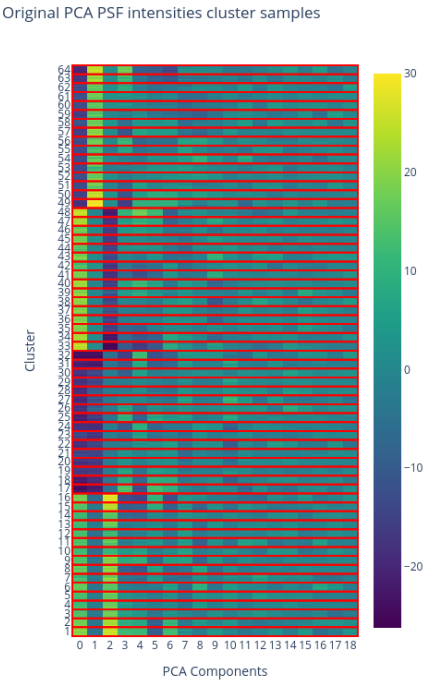
Table 16: K-Means hyperparameter configuration for c coefficients clustering



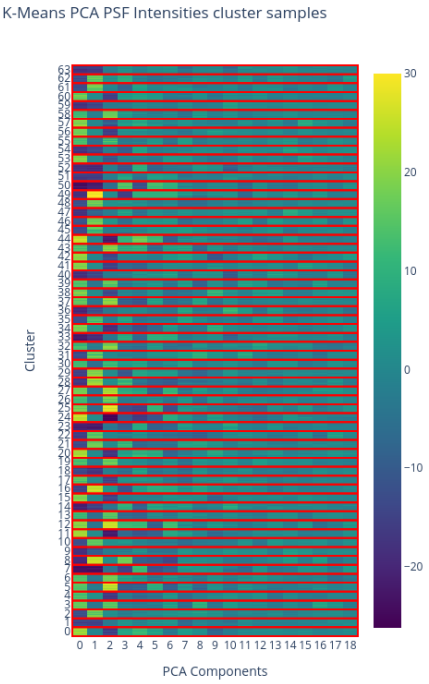
(a) Original cluster densities from PCA



(b) K-Means clusters densities from PCA



(c) Original cluster samples from PCA



(d) K-Means cluster samples from PCA

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

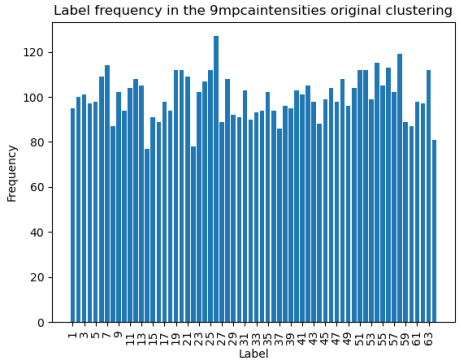
4.4.2 DBSCAN

A configuration that outputs 4 clusters is searched

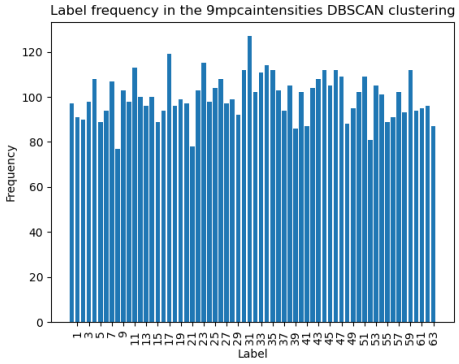
	Number of neighbours	**Epsilon**
PCA PSF Intensities	10	4

Table 17: DBSCAN hyperparameter configuration for PSF Intensities clustering

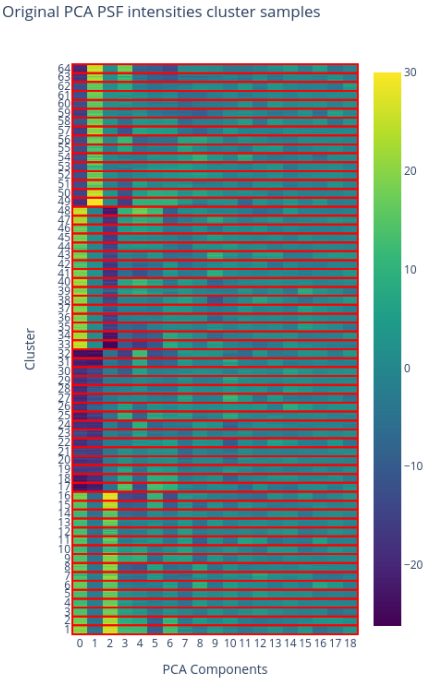
The results are the following:



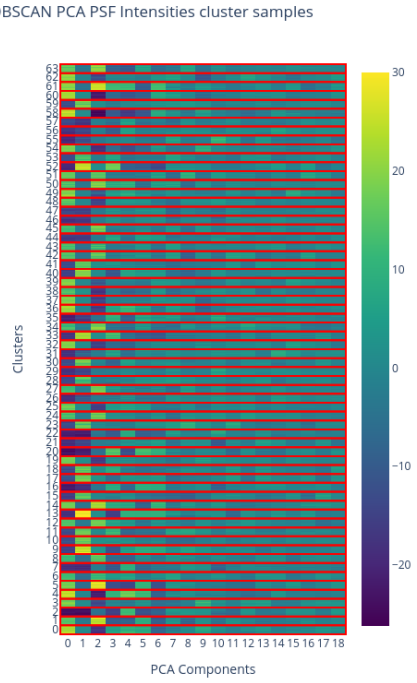
(a) Original cluster densities from PCA



(b) DBSCAN clusters densities from PCA



(c) Original cluster samples from PCA



(d) DBSCAN cluster samples from PCA

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

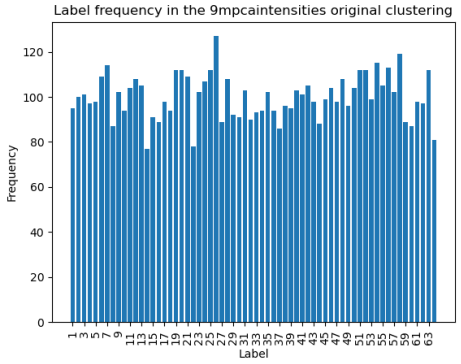
4.4.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

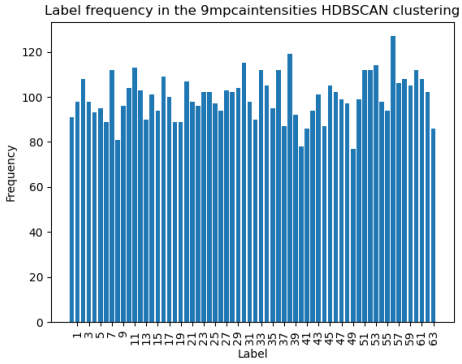
	Minimum cluster size
PCA PSF Intensities	15

Table 18: HDBSCAN hyperparameter configuration for PSF Intensities clustering

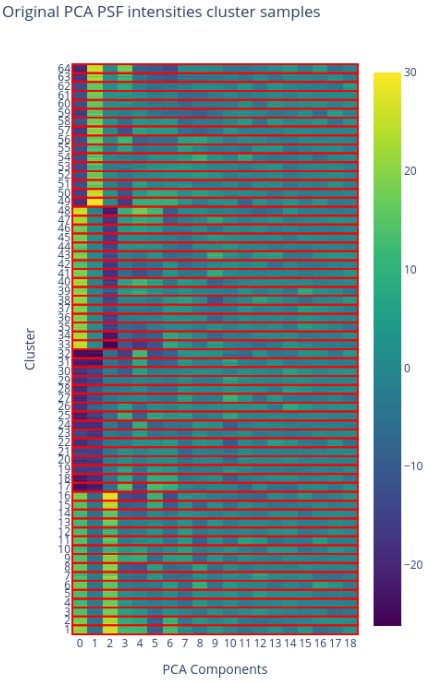
The results are the following:



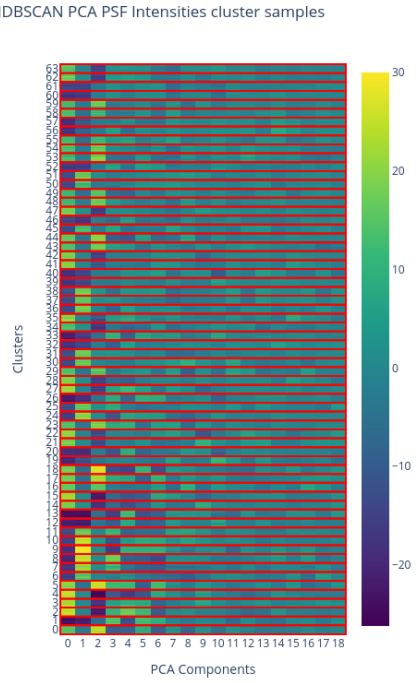
(a) Original cluster densities from PCA



(b) HDBSCAN clusters densities from PCA



(c) Original cluster samples from PCA



(d) HDBSCAN cluster samples from PCA

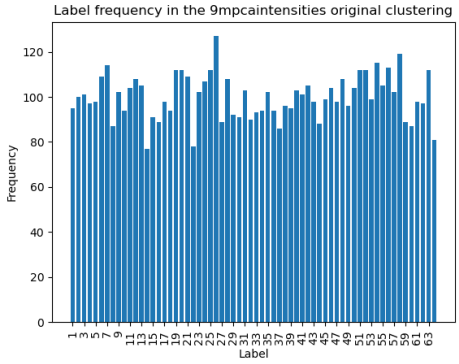
Figure 18: Comparison between original clustering and HDBSCAN clustering from PCA of PSF Intensities

4.4.4 Agglomerative clustering

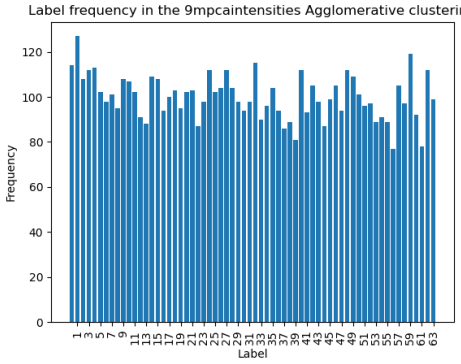
	Number of clusters
PCA PSF Intensities	64

Table 19: Agglomerative hyperparameter configuration for PSF Intensities clustering

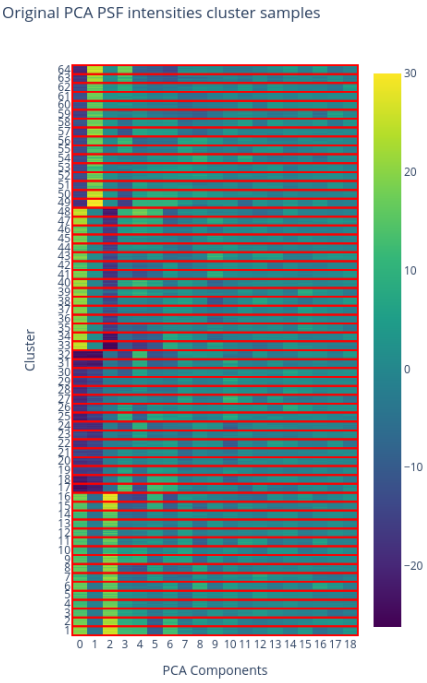
The results are the following:



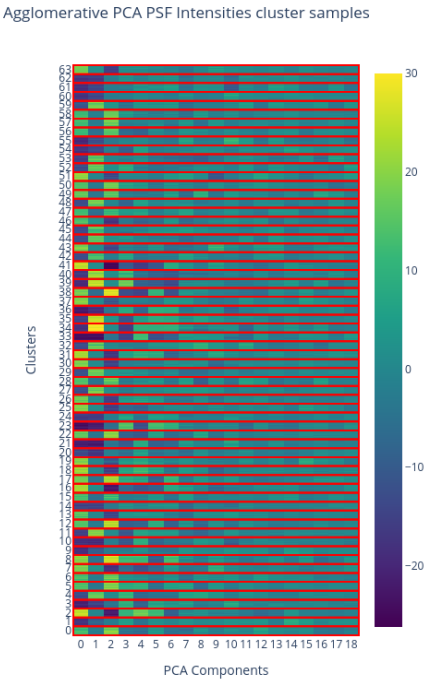
(a) Original cluster densities from PCA



(b) Agglomerative clusters densities from PCA



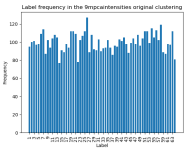
(c) Original cluster samples from PCA



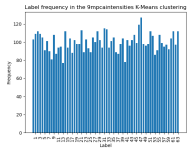
(d) Agglomerative cluster samples from PCA

Figure 19: Comparison between original clustering and Agglomerative clustering

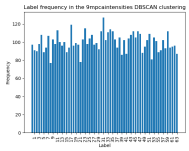
4.4.5 Summary



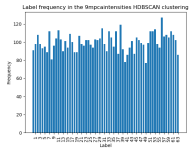
(a) Original cluster densities from PCA



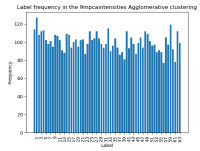
(b) K-means cluster densities from PCA



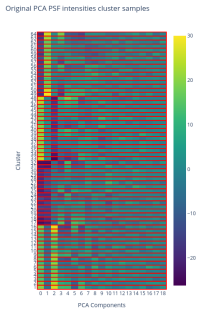
(c) DBSCAN cluster densities from PCA



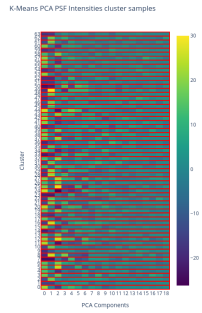
(d) HDBSCAN cluster densities from PCA



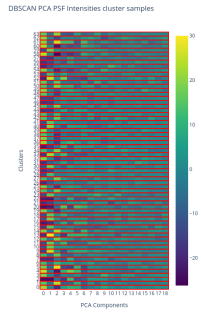
(e) Agglomerative cluster densities from PCA



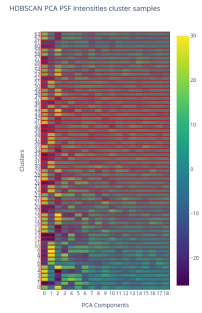
(f) Original cluster samples from PCA



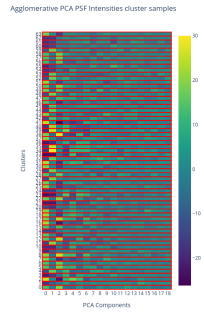
(g) K-means cluster samples from PCA



(h) DBSCAN cluster samples-from PCA



(i) HDBSCAN cluster samples from PCA



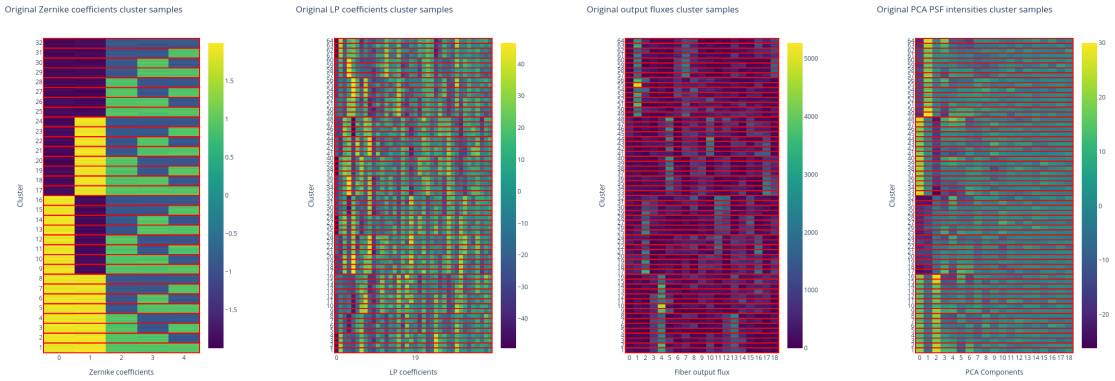
(j) Agglomerative cluster samples from PCA

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

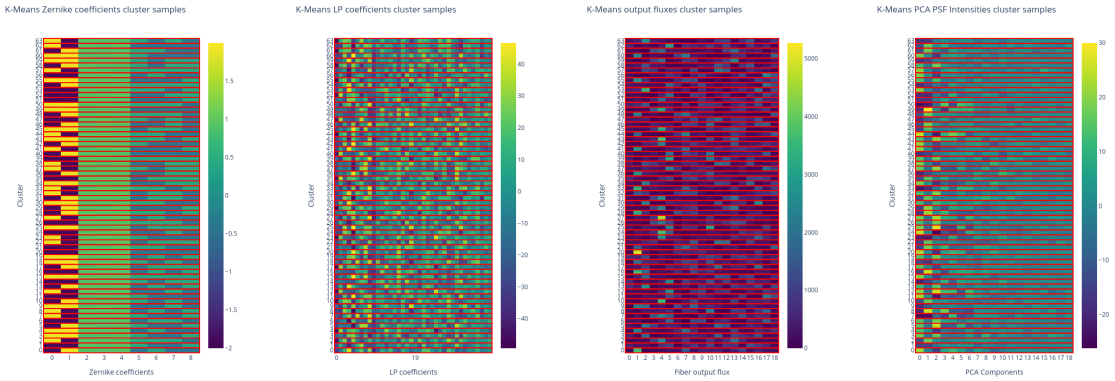
Table 20: Normalized Mutual Information between PCA PSF Intensities clusters

5 Dataset clusters comparison



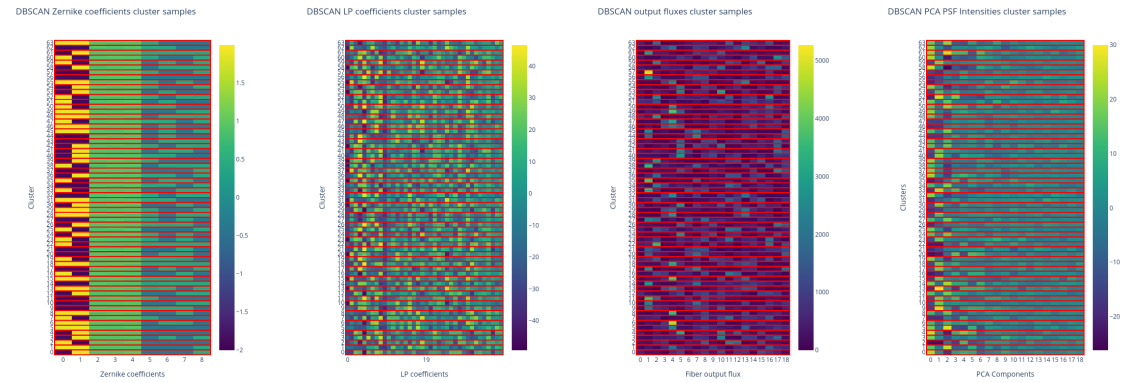
(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) PCA PSF Intensities

Figure 21: Original clusters from the datasets



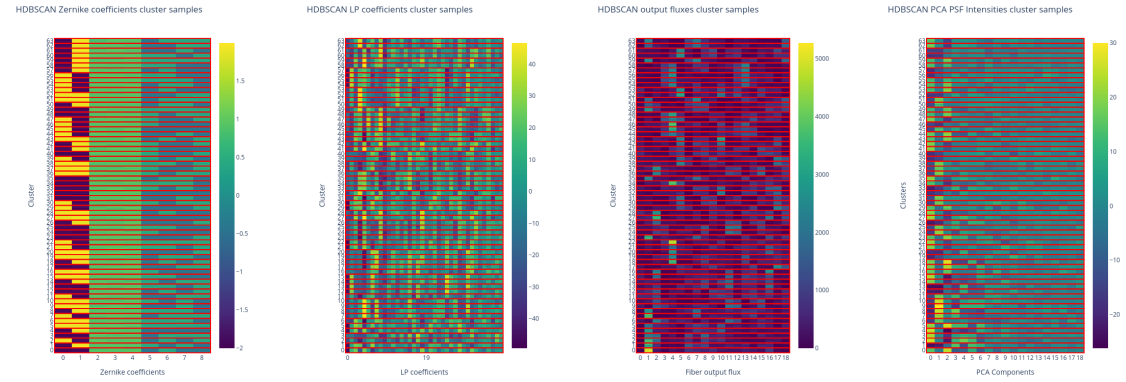
(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) PCA PSF Intensities

Figure 22: K-Means clusters from the datasets



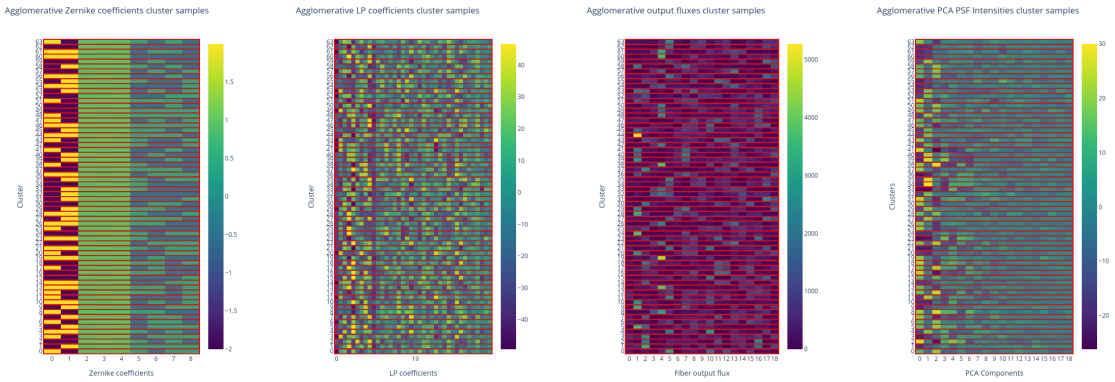
(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) PCA PSF Intensities

Figure 23: DBSCAN clusters from the datasets



(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) PCA PSF Intensities

Figure 24: HDBSCAN clusters from the datasets



(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) PCA PSF Intensities

Figure 25: Agglomerative clusters from the datasets