Lab Notebook

Photonic Lantern Information Determination

Contents

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 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 [-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

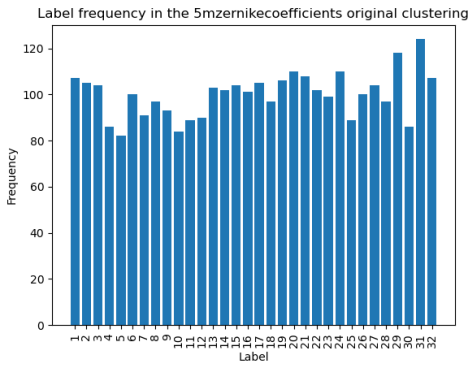
4.1.1 K-Means

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.

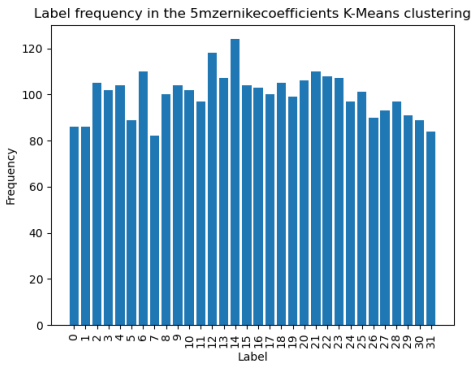
Number of clusters	**Number of initializations**
32	100

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

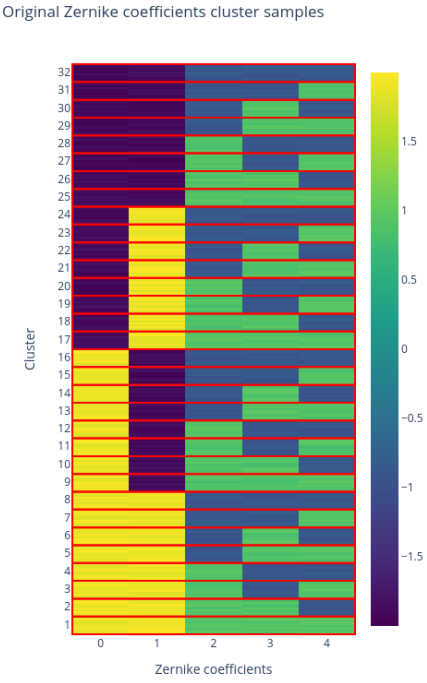
The results are the following:



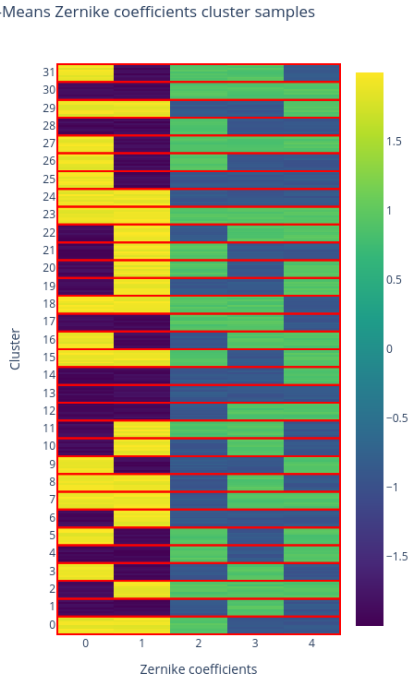
(a) Original cluster densities



(b) K-Means clusters densities



(c) Original cluster samples



(d) K-Means cluster samples

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

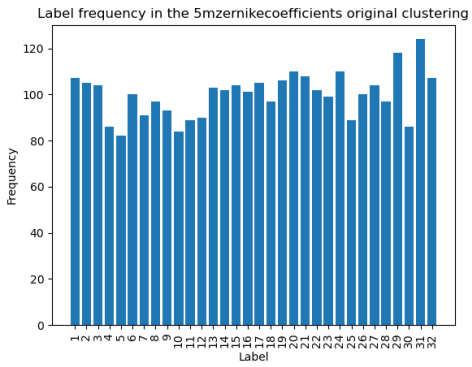
4.1.2 DBSCAN

A configuration that outputs 4 clusters is searched

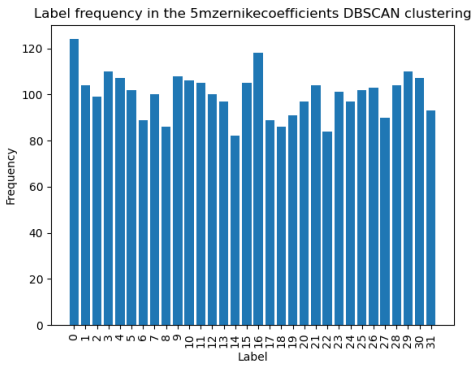
Number of neighbours	**Epsilon**
5	0.14

Table 2: DBSCAN hyperparameter configuration for Zernike coefficients clustering

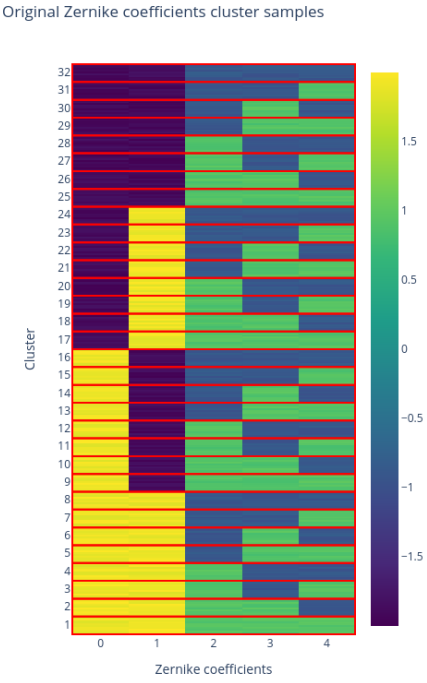
The results are the following:



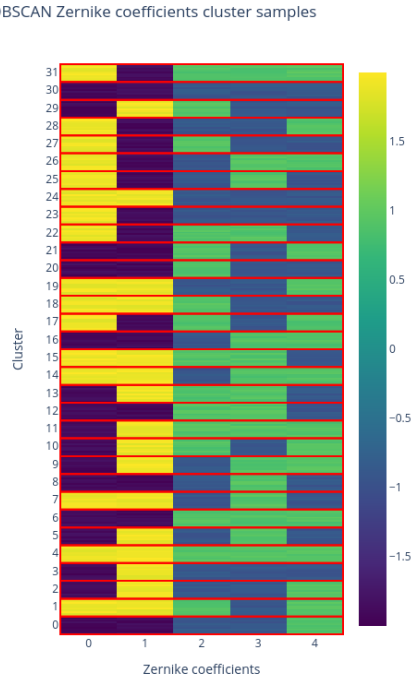
(a) Original cluster densities



(b) DBSCAN clusters densities



(c) Original cluster samples



(d) DBSCAN cluster samples

Figure 2: Comparison between original clustering and DBSCAN clustering

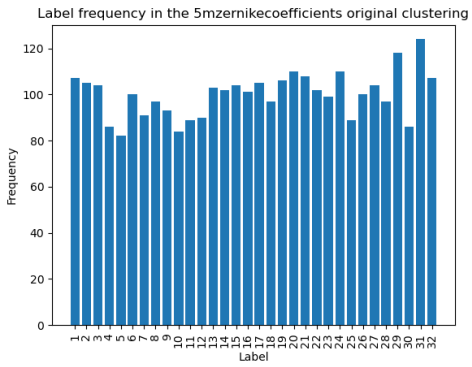
4.1.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

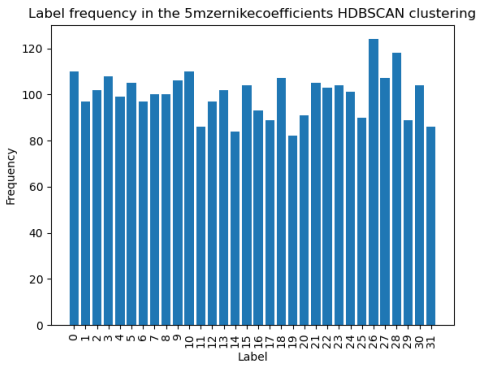
Minimum cluster size
50

Table 3: HDBSCAN hyperparameter configuration for Zernike coefficients clustering

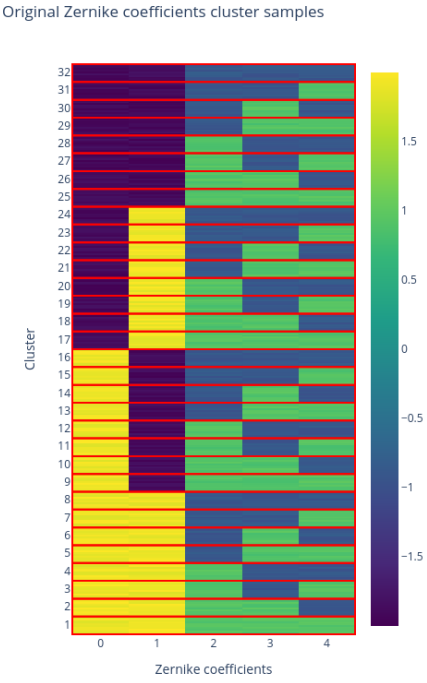
The results are the following:



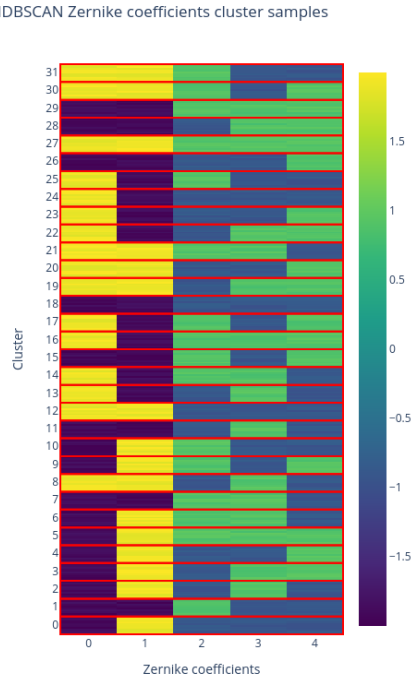
(a) Original cluster densities



(b) HDBSCAN clusters densities



(c) Original cluster samples



(d) HDBSCAN cluster samples

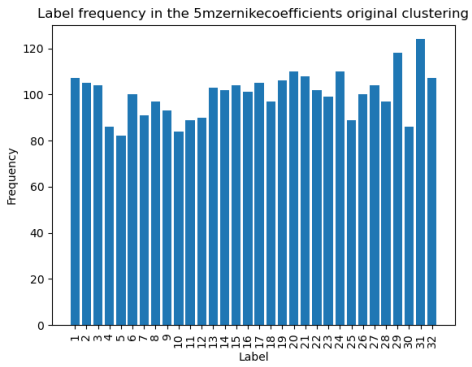
Figure 3: Comparison between original clustering and HDBSCAN clustering

4.1.4 Agglomerative clustering

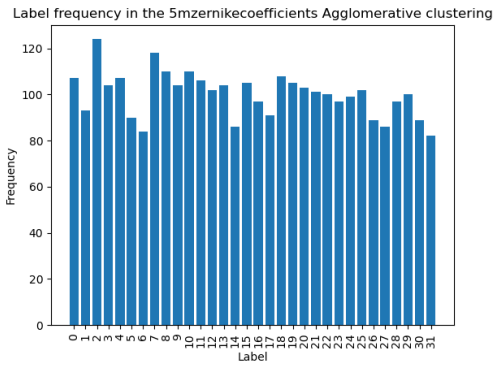
Number of clusters 4
5

Table 4: Agglomerative hyperparameter configuration for Zernike coefficients clustering

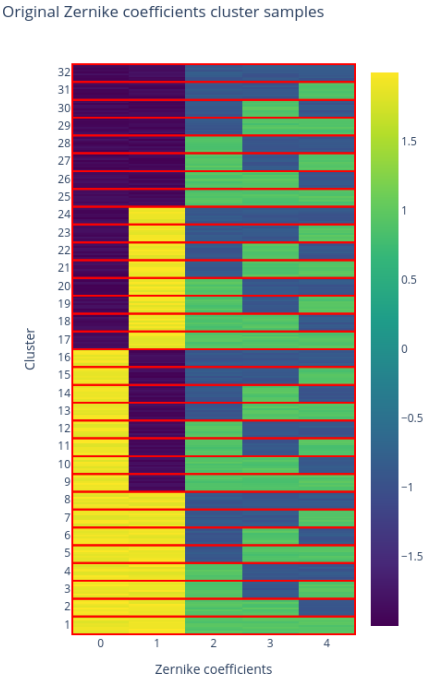
The results are the following:



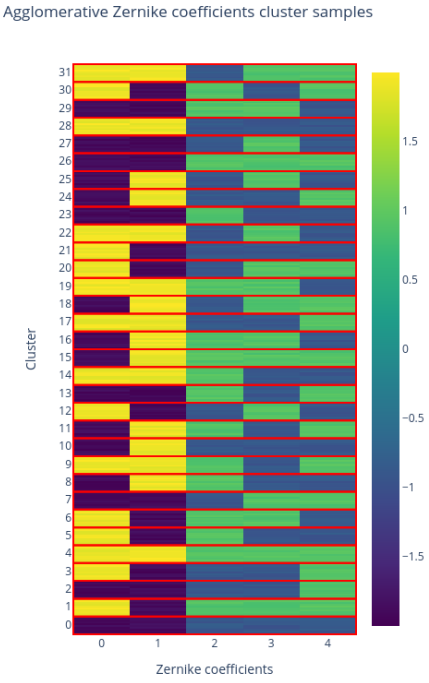
(a) Original cluster densities



(b) Agglomerative clusters densities



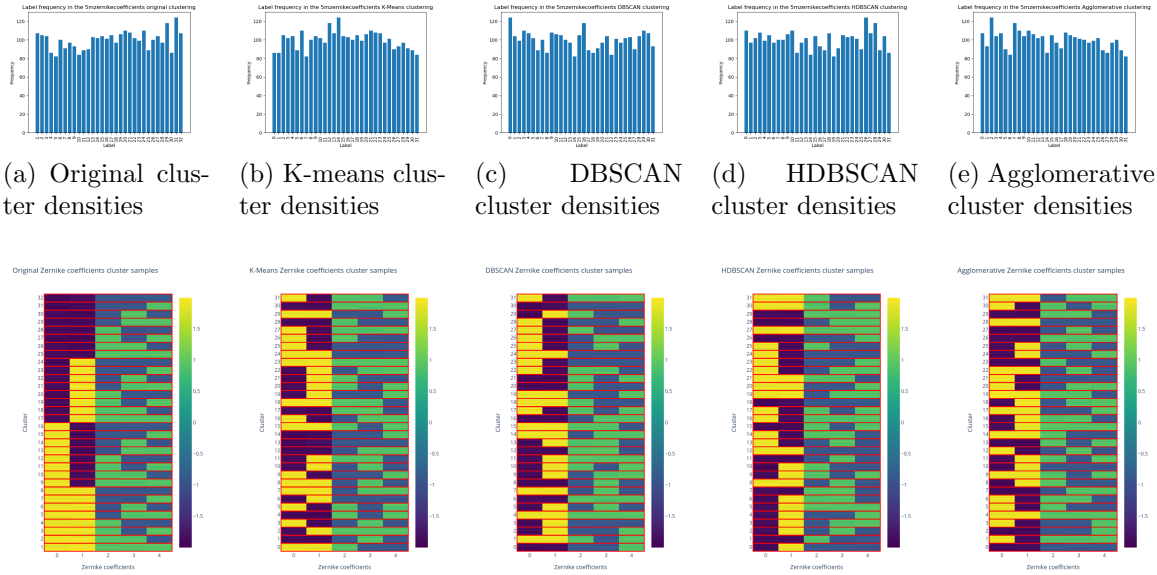
(c) Original cluster samples



(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

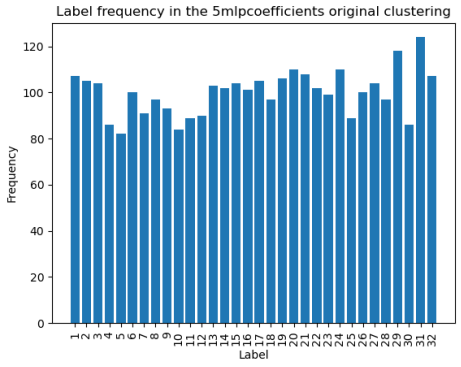
4.2.1 K-Means

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.

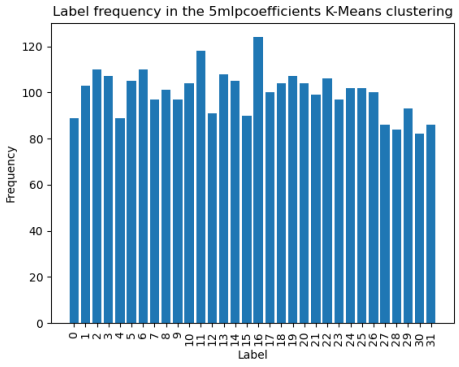
The results are the following:

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

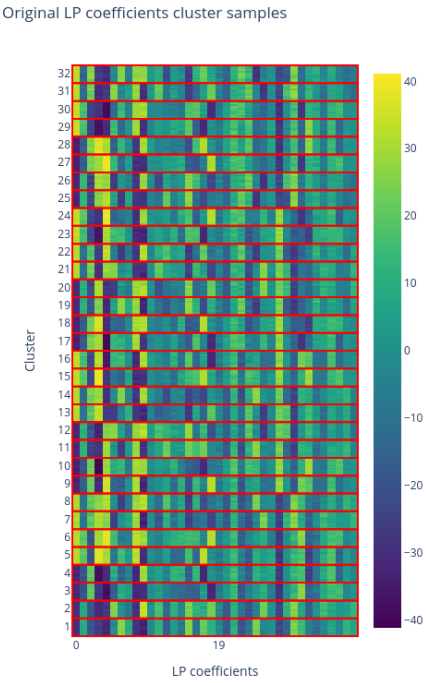
Table 6: K-Means hyperparameter configuration for c coefficients clustering



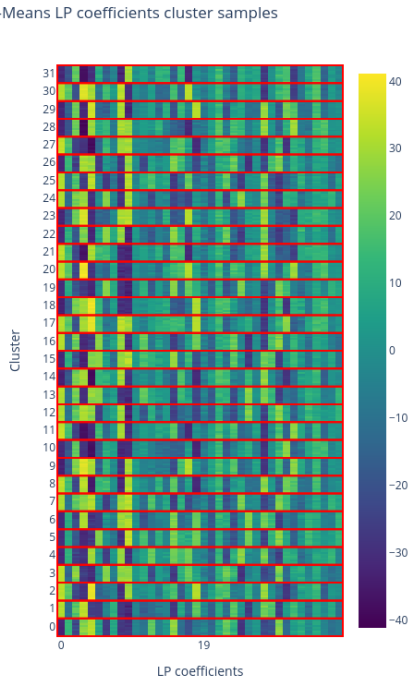
(a) Original cluster densities



(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

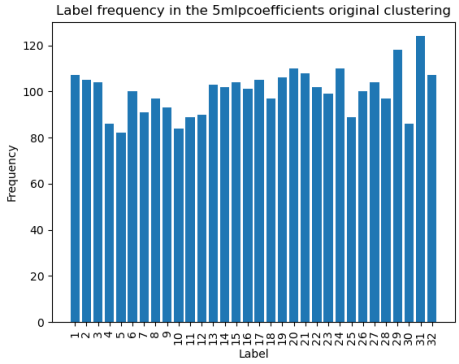
4.2.2 DBSCAN

A configuration that outputs 4 clusters is searched

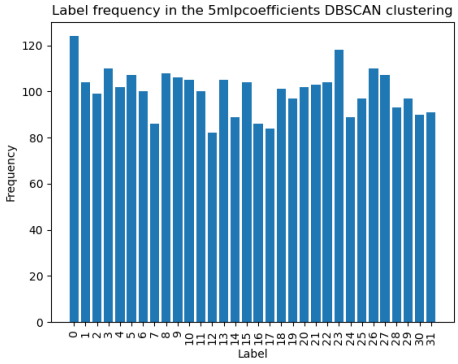
	Number of neighbours	**Epsilon**
Original LP coefficients	15	11

Table 7: DBSCAN hyperparameter configuration for LP coefficients clustering

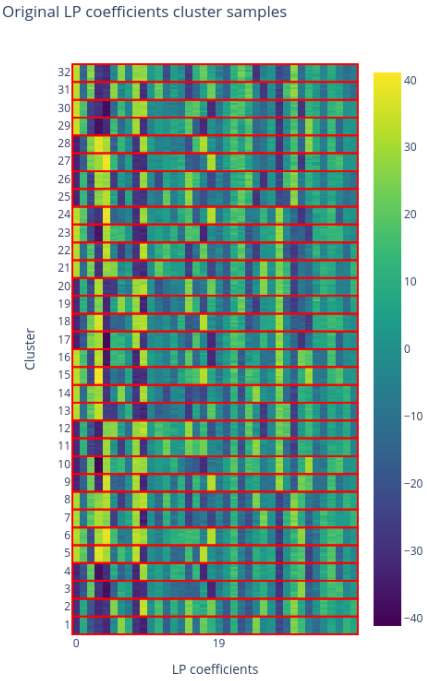
The results are the following:



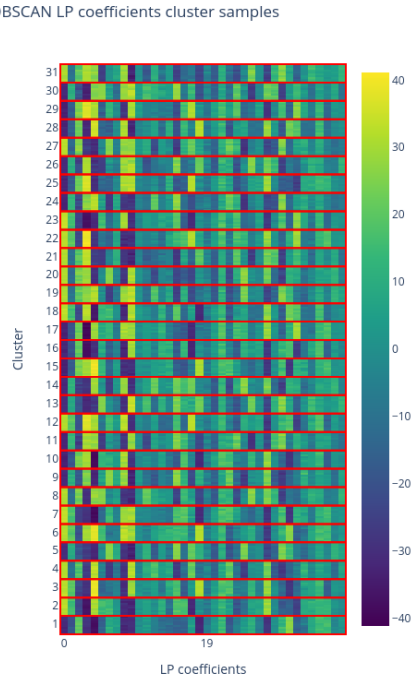
(a) Original cluster densities



(b) DBSCAN clusters densities



(c) Original cluster samples



(d) DBSCAN cluster samples

Figure 7: Comparison between original clustering and DBSCAN clustering

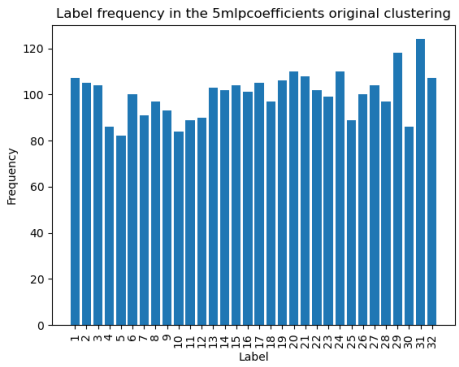
4.2.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

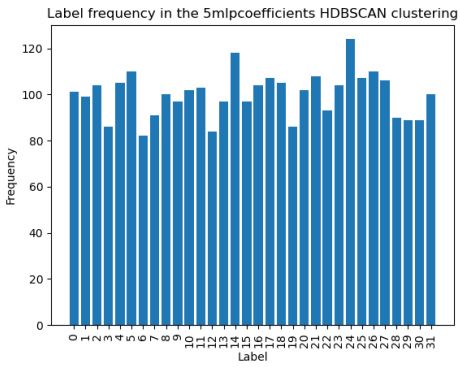
	Minimum cluster size
Original LP coefficients	21

Table 8: HDBSCAN hyperparameter configuration for LP coefficients clustering

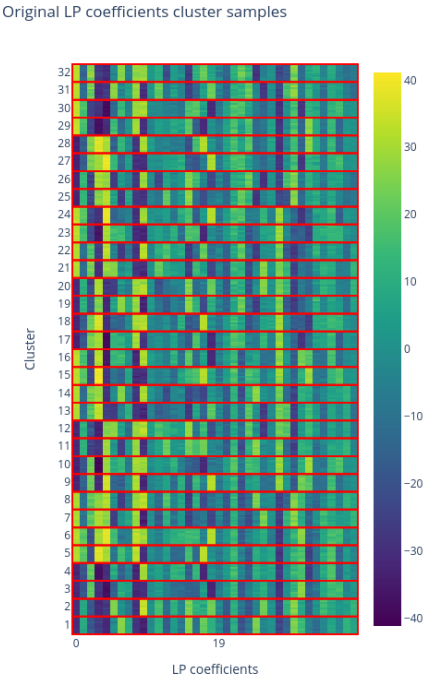
The results are the following:



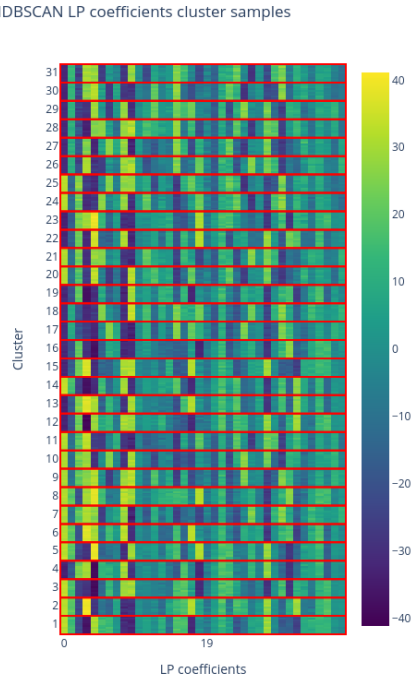
(a) Original cluster densities



(b) HDBSCAN clusters densities



(c) Original cluster samples



(d) HDBSCAN cluster samples

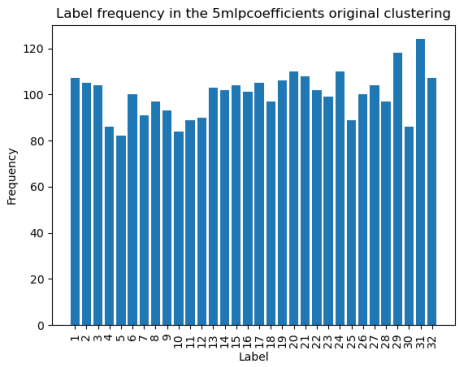
Figure 8: Comparison between original clustering and HDBSCAN clustering

4.2.4 Agglomerative clustering

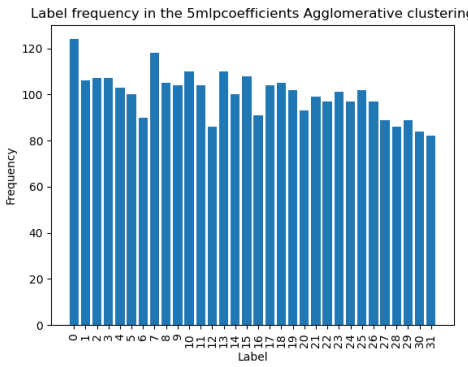
	Number of clusters
Original LP coefficients	4

Table 9: Agglomerative hyperparameter configuration for LP coefficients clustering

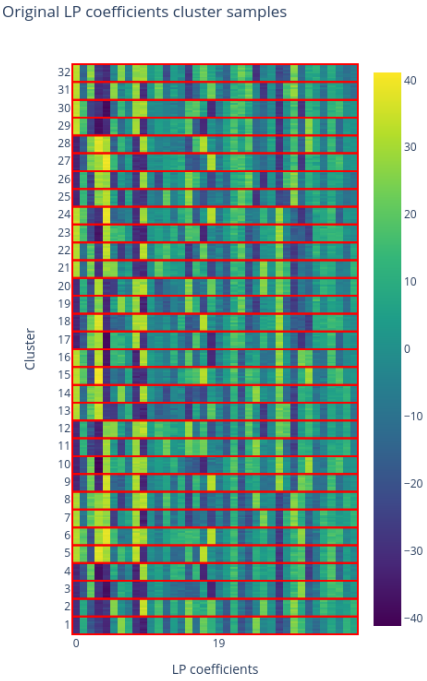
The results are the following:



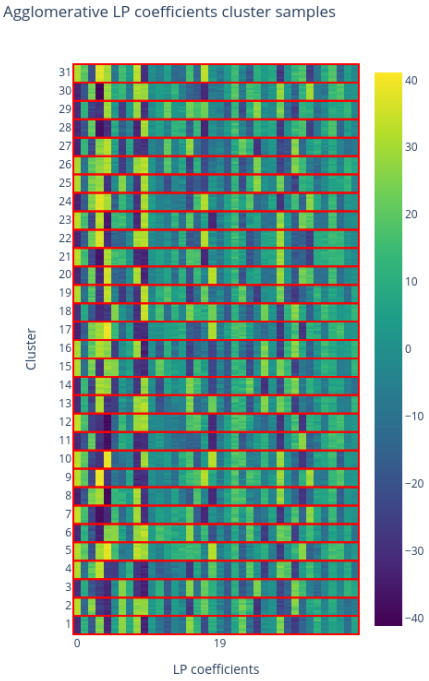
(a) Original cluster densities



(b) Agglomerative clusters densities



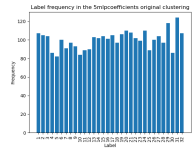
(c) Original cluster samples



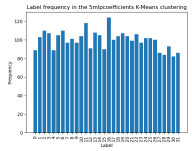
(d) Agglomerative cluster samples

Figure 9: Comparison between original clustering and Agglomerative clustering

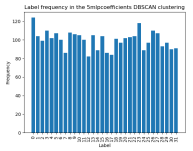
4.2.5 Summary



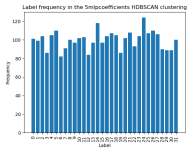
(a) Original cluster densities



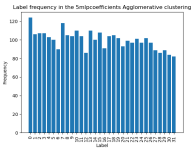
(b) K-means cluster densities



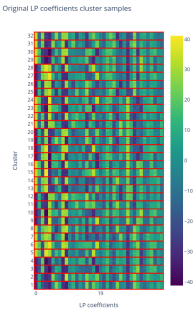
(c) DBSCAN cluster densities



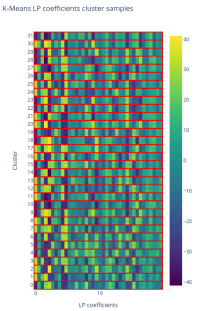
(d) HDBSCAN cluster densities



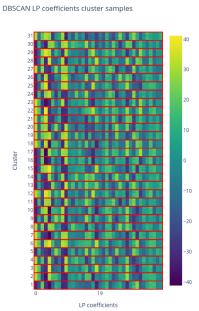
(e) Agglomerative cluster densities



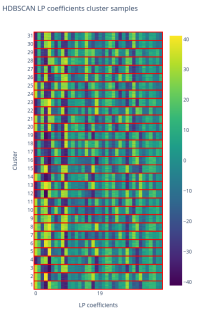
(f) Original cluster samples



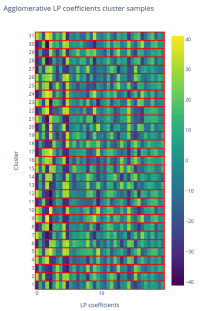
(g) K-means cluster samples



(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

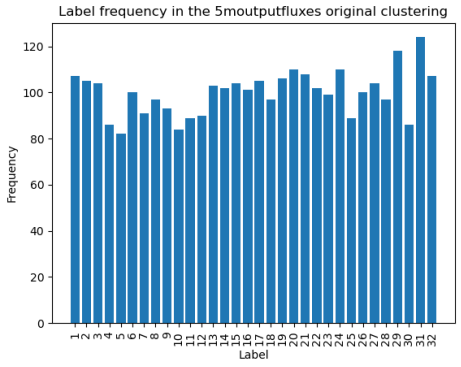
4.3.1 K-Means

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.

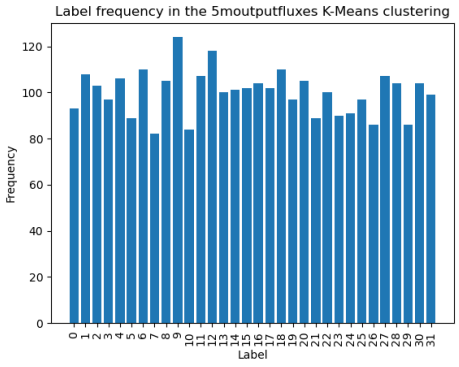
The results are the following:

	Number of clusters	**Number of initializations**
Original Output fluxes	32	100

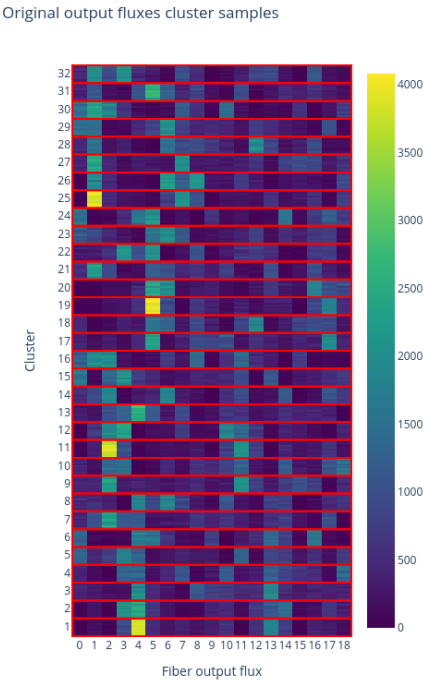
Table 11: K-Means hyperparameter configuration for c coefficients clustering



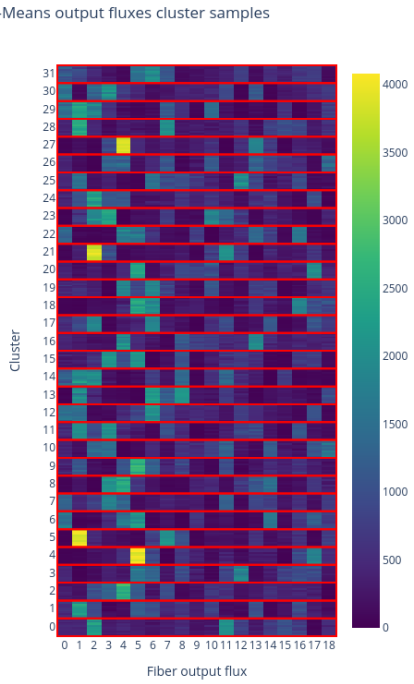
(a) Original cluster densities



(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

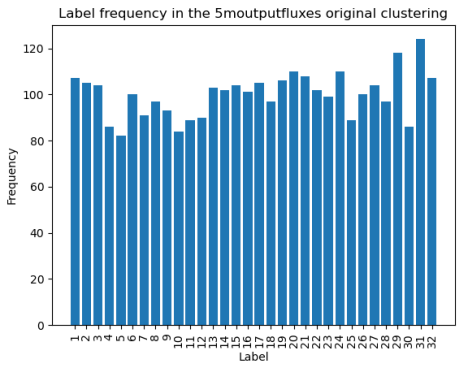
4.3.2 DBSCAN

A configuration that outputs 4 clusters is searched

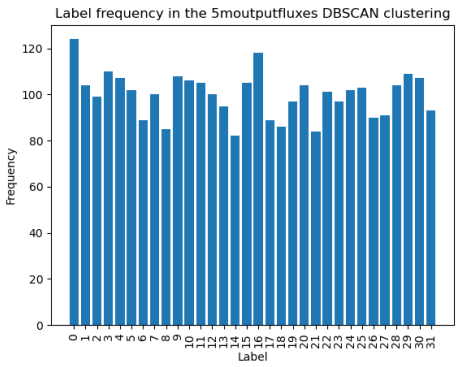
	Number of neighbours	**Epsilon**
Original Output fluxes	7	300

Table 12: DBSCAN hyperparameter configuration for Output fluxes clustering

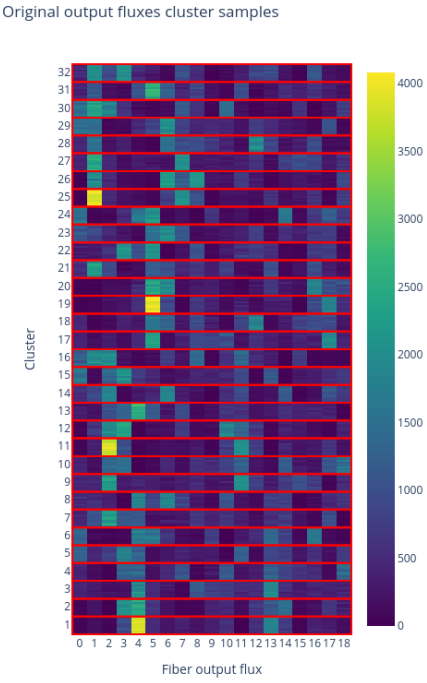
The results are the following:



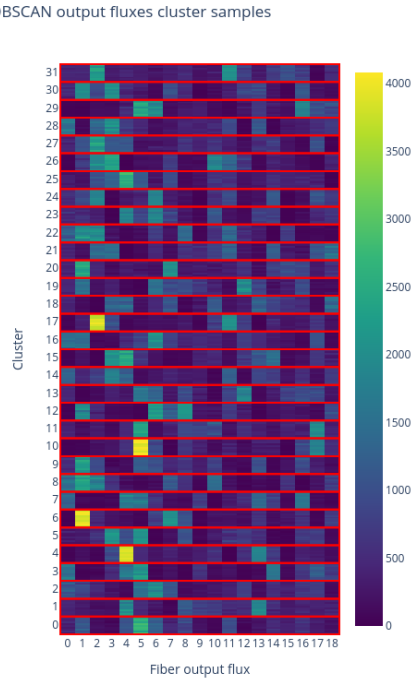
(a) Original cluster densities



(b) DBSCAN clusters densities



(c) Original cluster samples



(d) DBSCAN cluster samples

Figure 12: Comparison between original clustering and DBSCAN clustering

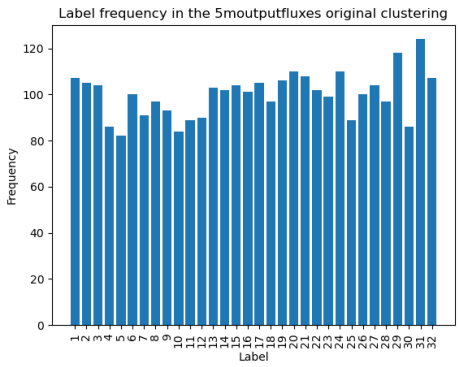
4.3.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

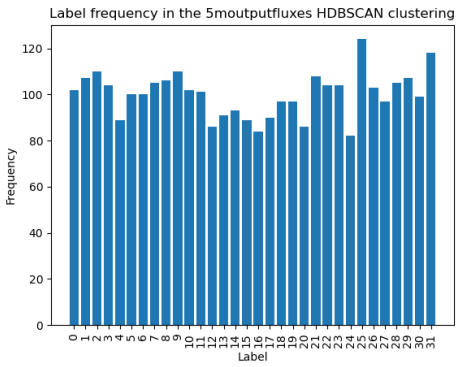
	Minimum cluster size
Original Output fluxes	21

Table 13: HDBSCAN hyperparameter configuration for Output fluxes clustering

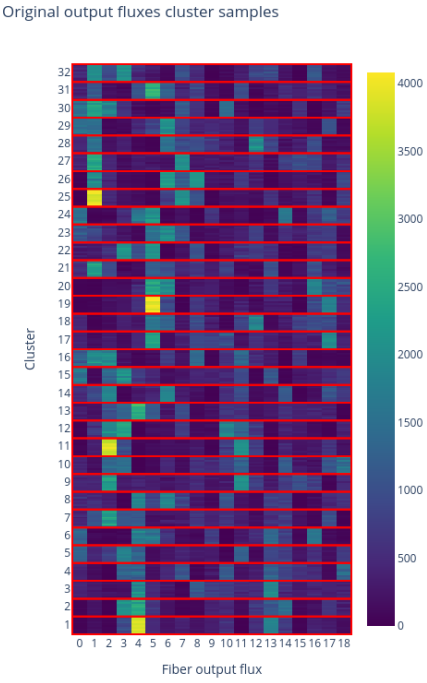
The results are the following:



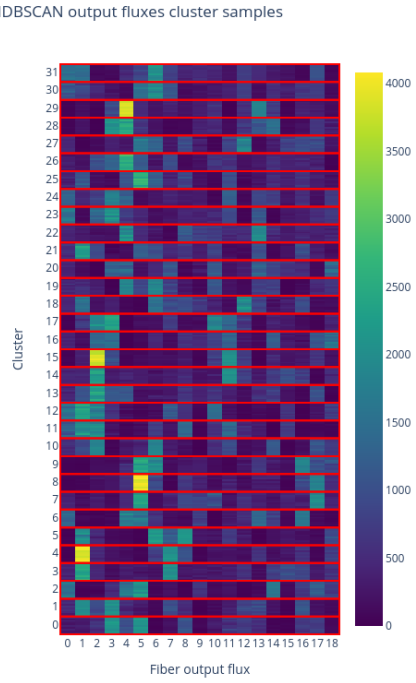
(a) Original cluster densities



(b) HDBSCAN clusters densities



(c) Original cluster samples



(d) HDBSCAN cluster samples

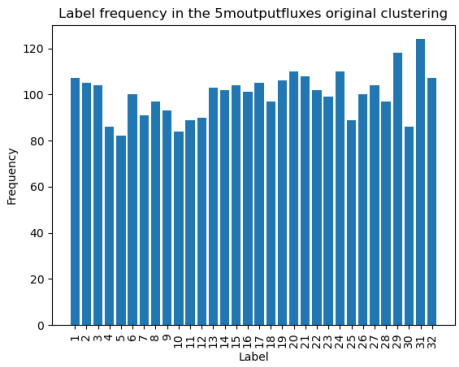
Figure 13: Comparison between original clustering and HDBSCAN clustering

4.3.4 Agglomerative clustering

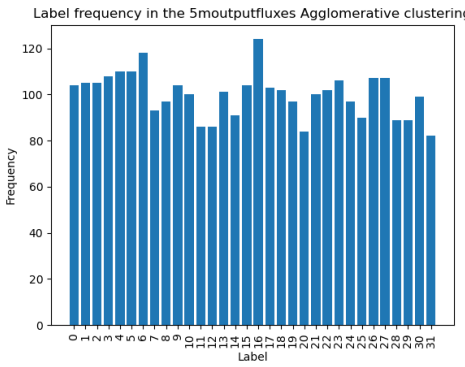
	Number of clusters
Original Output fluxes	32

Table 14: Agglomerative hyperparameter configuration for Output fluxes clustering

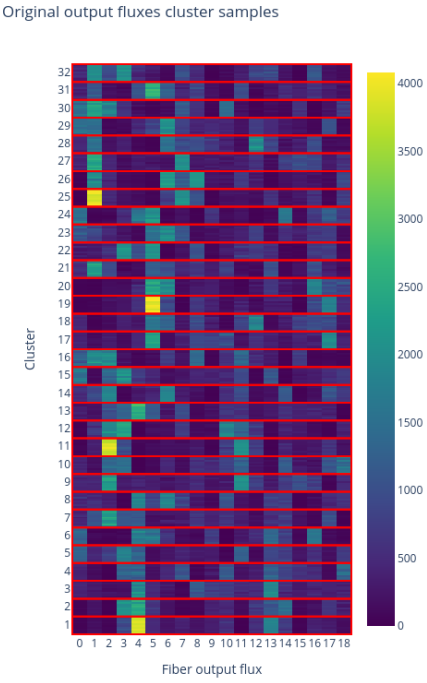
The results are the following:



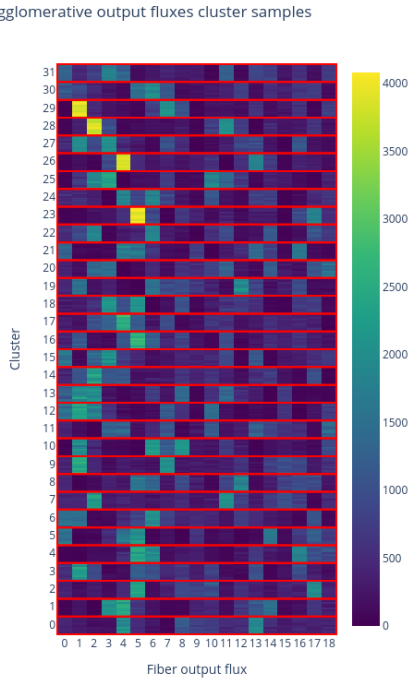
(a) Original cluster densities



(b) Agglomerative clusters densities



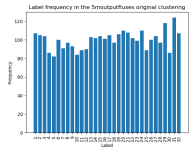
(c) Original cluster samples



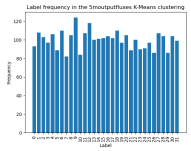
(d) Agglomerative cluster samples

Figure 14: Comparison between original clustering and Agglomerative clustering

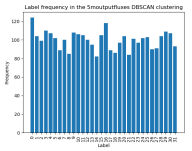
4.3.5 Summary



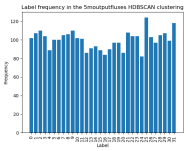
(a) Original cluster densities



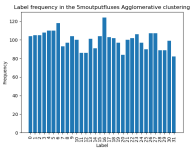
(b) K-means cluster densities



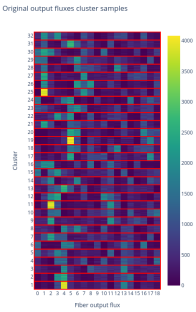
(c) DBSCAN cluster densities



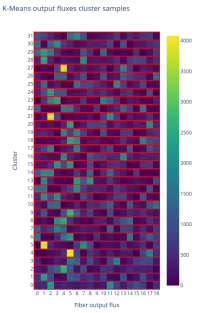
(d) HDBSCAN cluster densities



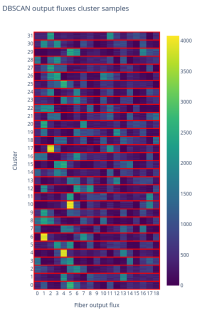
(e) Agglomerative cluster densities



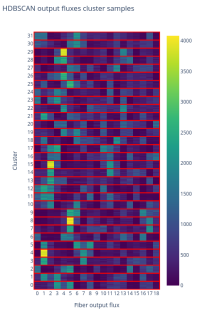
(f) Original cluster samples



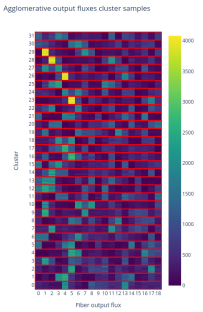
(g) K-means cluster samples



(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	\	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