# Lab Notebook

Photonic Lantern Information Determination

# Contents

# 1 Mini Dataset Information Determination

# 2 The data

## 2.1 Zernike coefficients dataset

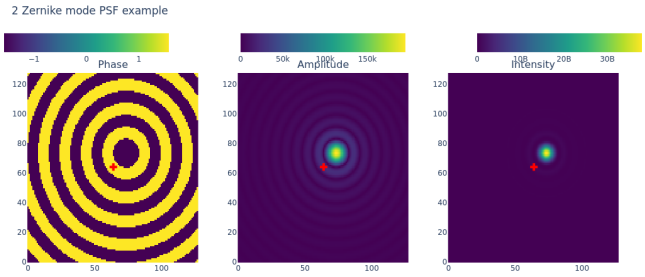
A dataset of 1000 zernike coefficients is created for this report.In particular, each datapoint represent the coefficients of the first 2 Zernike modes, their values ranging between [-2, -1.8] and [1.8, 2].

## 2.2 PSFs coefficients dataset

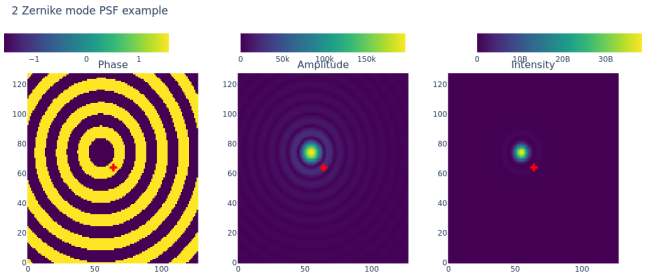
A dataset of 1000 PSFs is created using the Zernike coefficients dataset.


Figure 1: Example original sized PSF

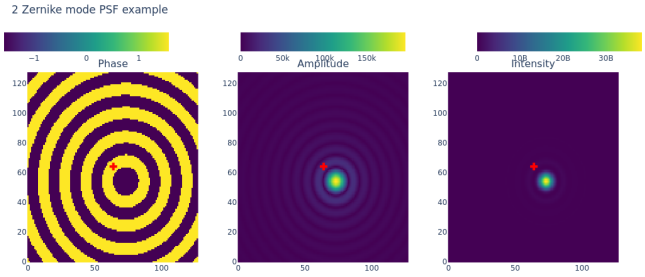
These ranges create 4 original clusters that will be used as reference.



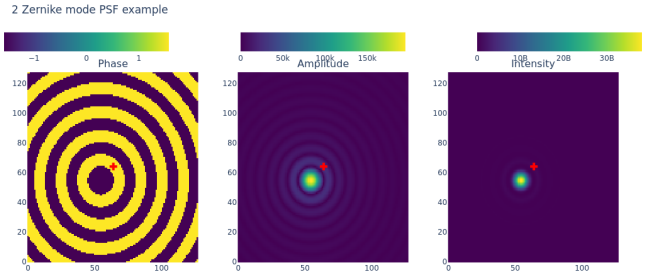

(a) Positive X-tilt and Positive Y-tilt PSF example




(b) Negative X-tilt and Positive Y-tilt PSF example




(c) Positive X-tilt and Negative Y-tilt PSF example




(d) Negative X-tilt and Negative Y-tilt PSF example

Figure 2: 2 Zernike modes PSF examples

### 2.3 LP mode coefficients dataset

A dataset of 1000 LP mode coefficients obtained from computing the overlap integral of the first 19 LP modes with the PSF dataset.

## 3 Preprocessing

### 3.1 LP Coefficients

The 1000x19x2 array is dimensionally reduced using PCA and UMAP both giving an array of 1000x2 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

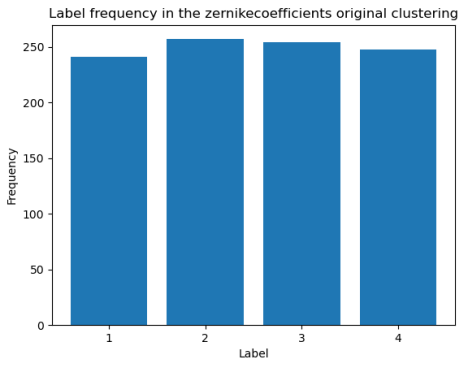
#### 4.1.1 K-Means

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.

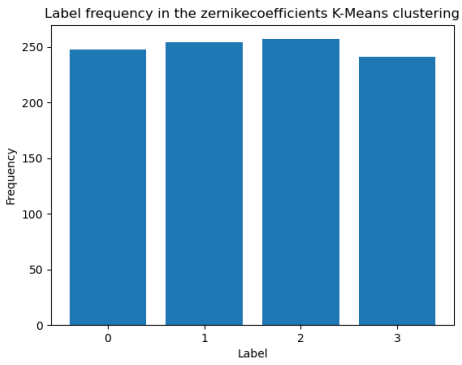
| **Number of clusters** | **Number of initializations** |
|---|---|
| 4 | 10 |

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

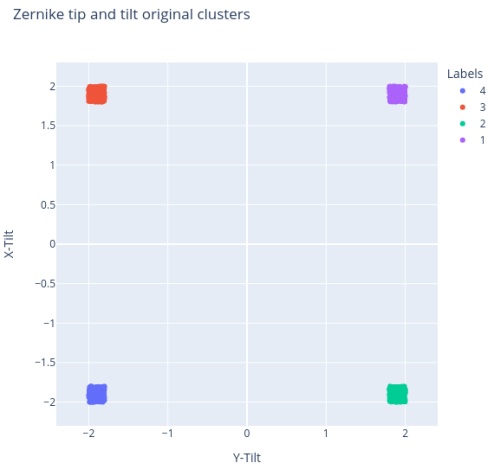
The results are the following:



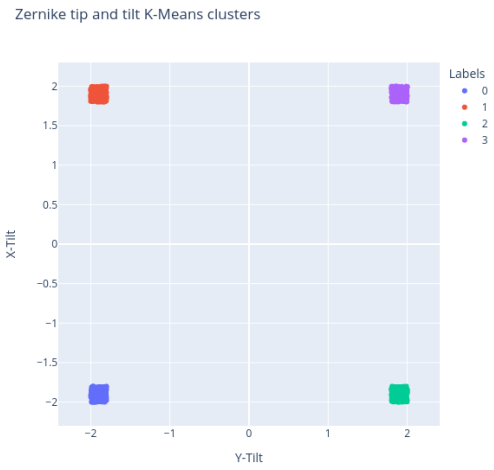

(a) Original cluster densities



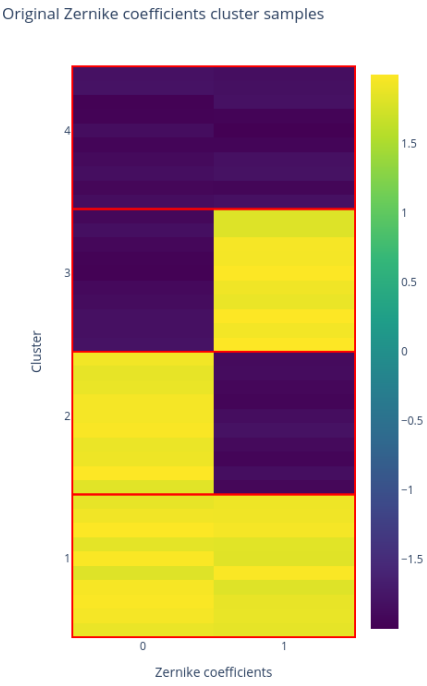

(b) K-Means clusters densities



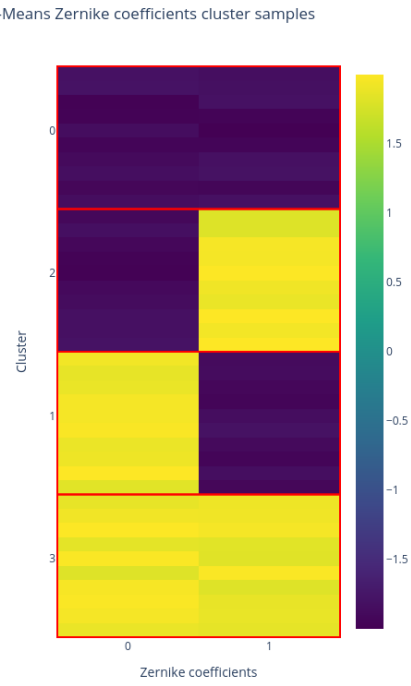

(c) Original clusters




(d) K-Means clusters




(e) Original cluster samples




(f) K-Means cluster samples

Figure 3: Comparison between original clustering and K-Means clustering

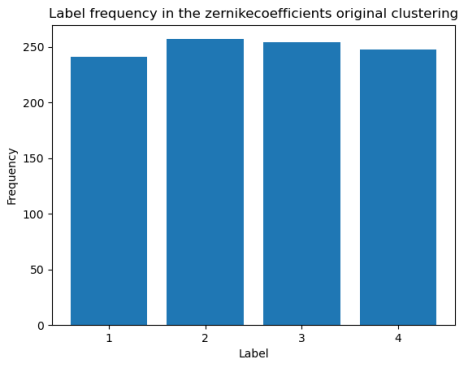
### 4.1.2 DBSCAN

A configuration that outputs 4 clusters is searched

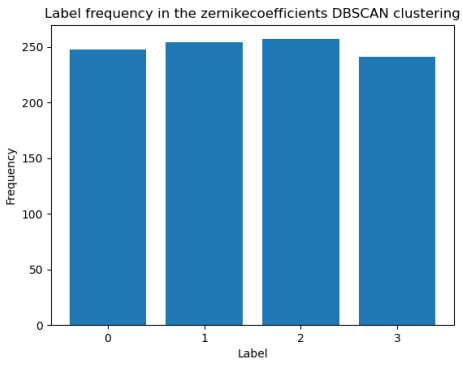
| **Number of neighbours** | **Epsilon** |
|---|---|
| 5 | 0.3 |

Table 2: DBSCAN hyperparameter configuration for Zernike coefficients clustering

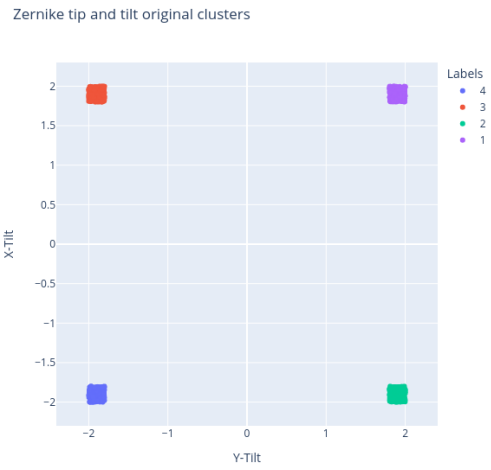
The results are the following:



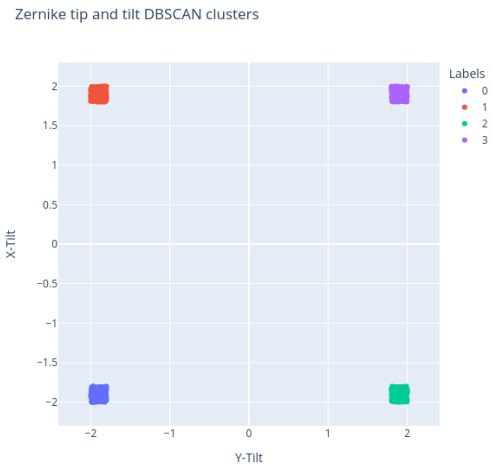

(a) Original cluster densities



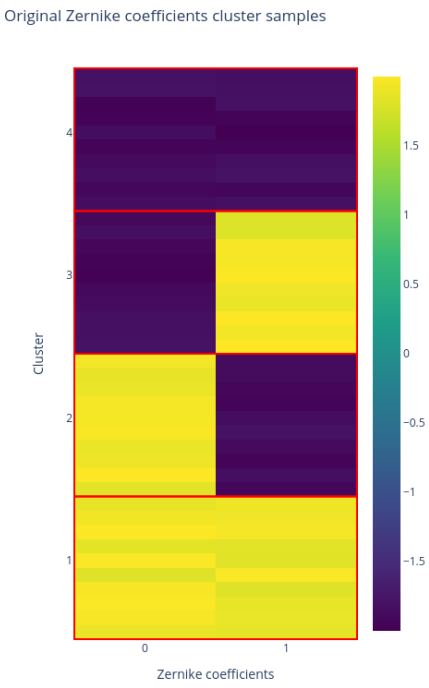

(b) DBSCAN clusters densities



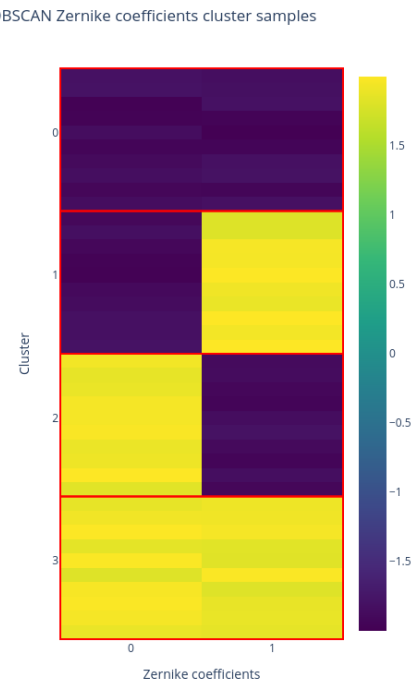

(c) Original clusters




(d) DBSCAN clusters




(e) Original cluster samples




(f) DBSCAN cluster samples

Figure 4: Comparison between original clustering and DBSCAN clustering

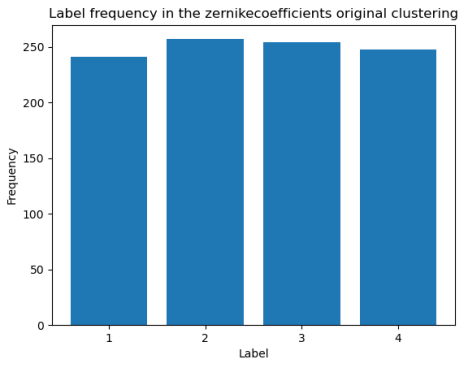
### 4.1.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

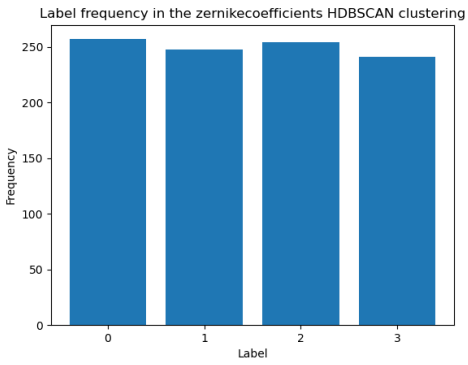
| **Minimum cluster size** |
|---|
| 5 |

Table 3: HDBSCAN hyperparameter configuration for Zernike coefficients clustering

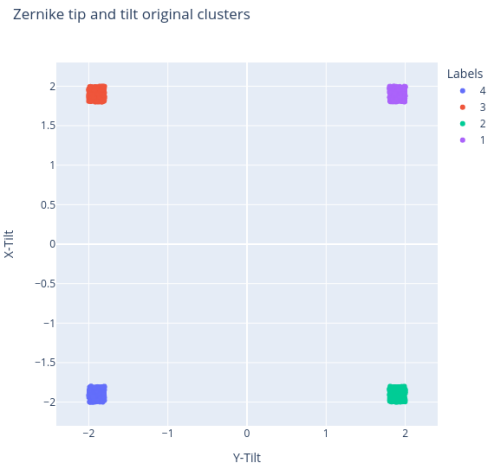
The results are the following:



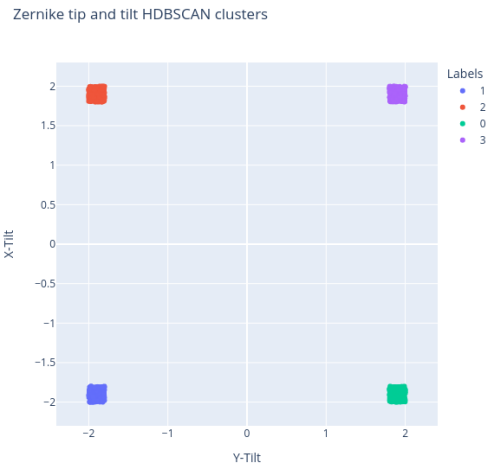

(a) Original cluster densities



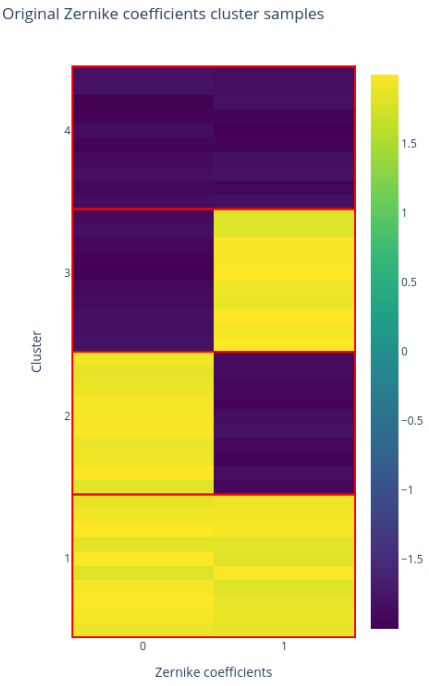

(b) HDBSCAN clusters densities



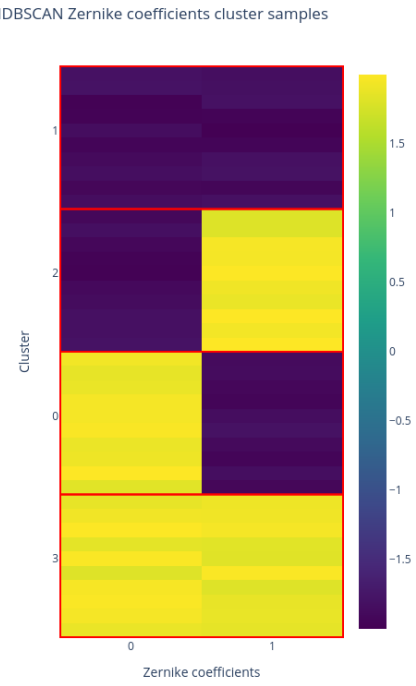

(c) Original clusters




(d) HDBSCAN clusters




(e) Original cluster samples




(f) HDBSCAN cluster samples

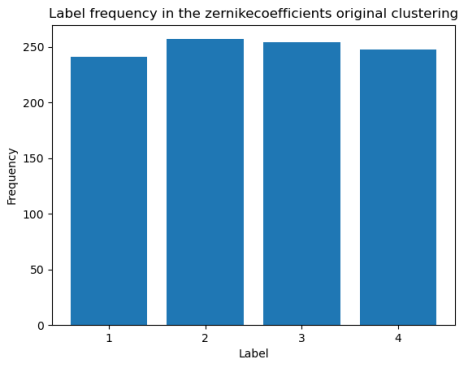
Figure 5: Comparison between original clustering and HDBSCAN clustering

### 4.1.4 Agglomerative clustering

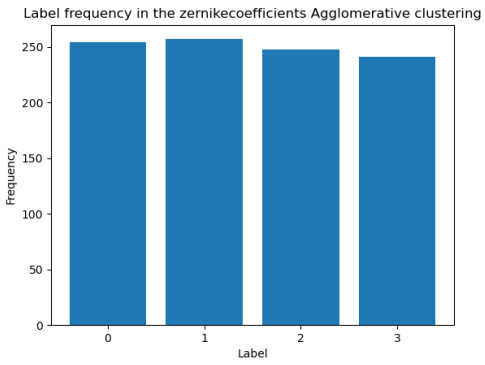
| **Number of clusters 4** |
|---|
| 5 |

Table 4: Agglomerative hyperparameter configuration for Zernike coefficients clustering

The results are the following:

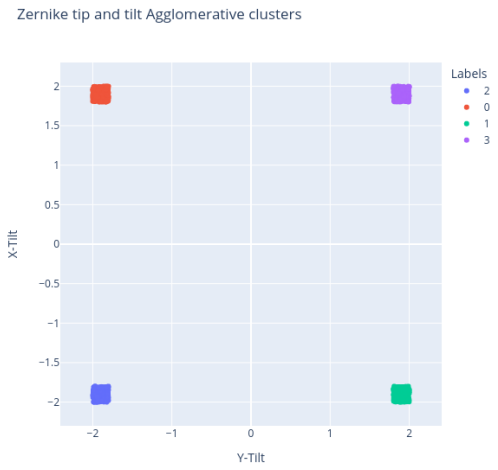



(a) Original cluster densities

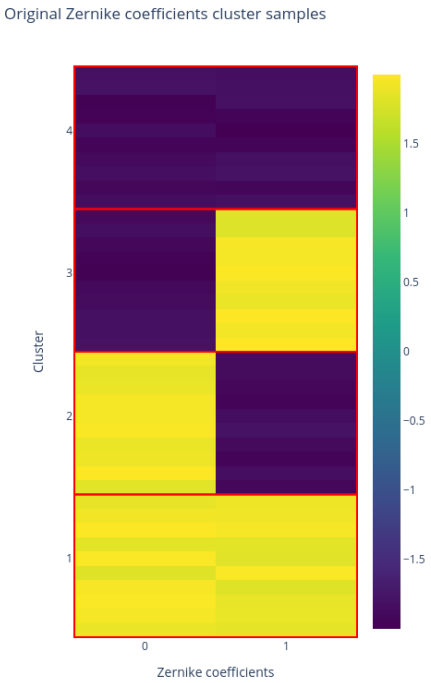



(b) Agglomerative clusters densities

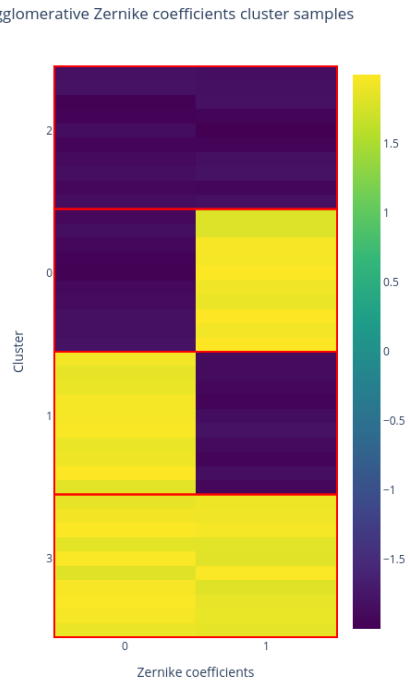

(c) Original clusters




(d) Agglomerative clusters



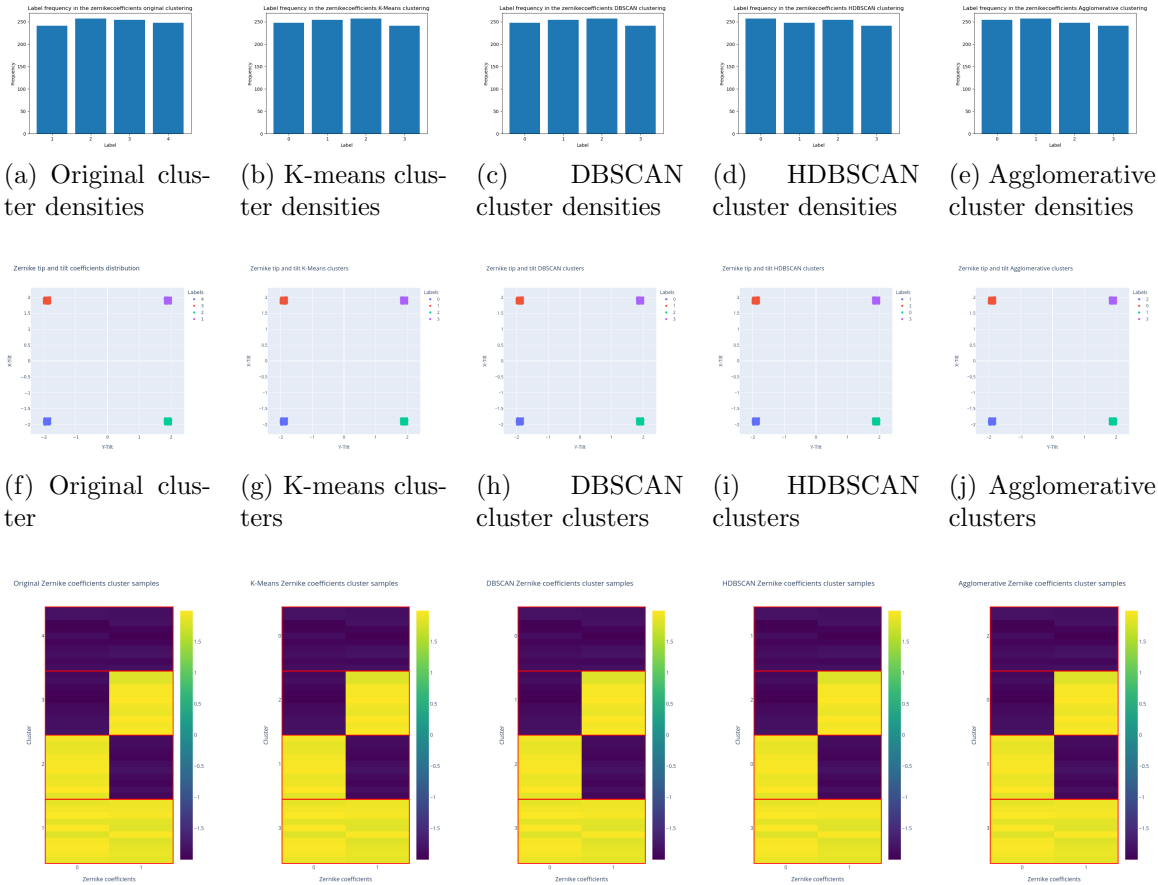

(e) Original cluster samples




(f) Agglomerative cluster samples

Figure 6: Comparison between original clustering and Agglomerative clustering

## 4.2 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 (g) K-means clusters (h) DBSCAN cluster clusters (i) HDBSCAN clusters (j) Agglomerative clusters

(k) Original cluster samples (l) K-means cluster samples (m) DBSCAN cluster samples (n) HDBSCAN cluster samples (o) Agglomerative cluster samples

Figure 7: 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

## 4.3 LP coefficients clustering

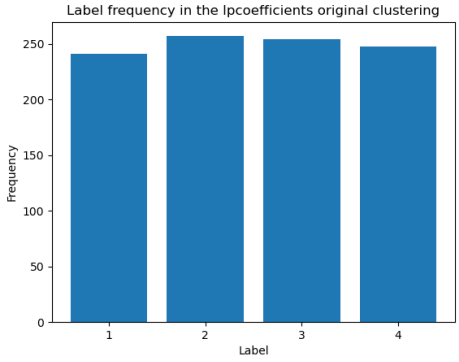
### 4.3.1 K-Means

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.

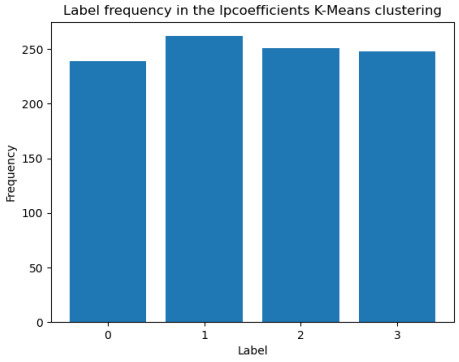
| | **Number of clusters** | **Number of initializations** |
|---|---|---|
| **Original LP coefficients** | 4 | 10 |
| **PCA LP coefficients** | 4 | 10 |
| **UMAP LP coefficients** | 4 | 10 |

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

The results are the following:

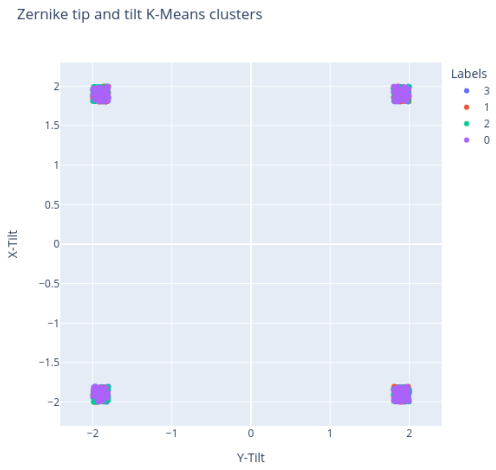



(a) Original cluster densities

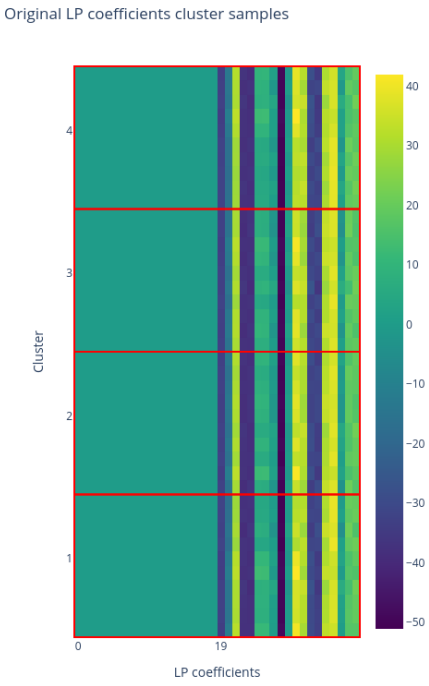



(b) K-Means clusters densities

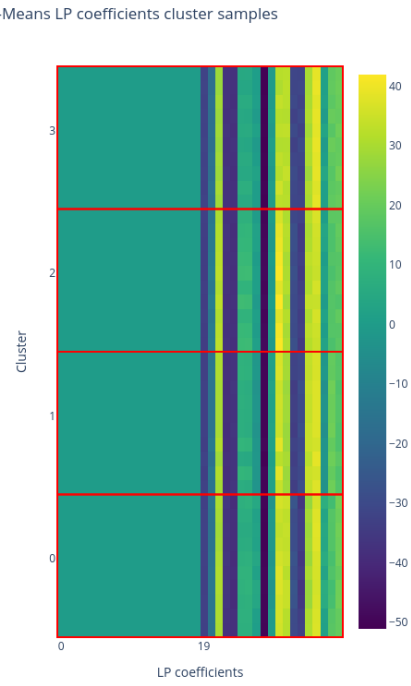

(c) Original clusters




(d) K-Means clusters

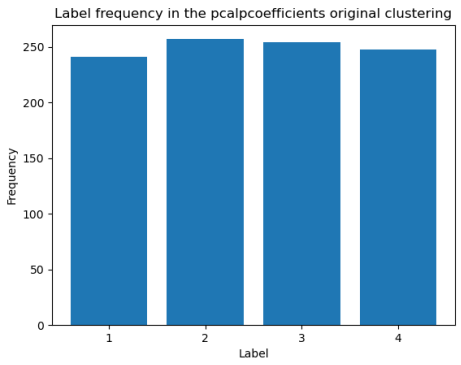



(e) Original cluster samples

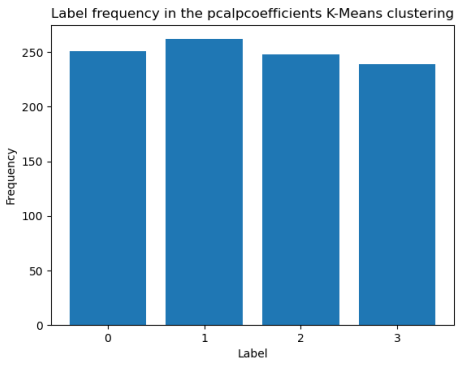



(f) K-Means cluster samples

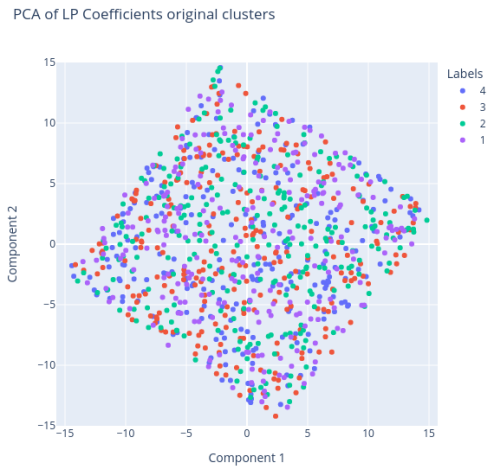
Figure 8: Comparison between original clustering and K-Means clustering from original LP coefficients



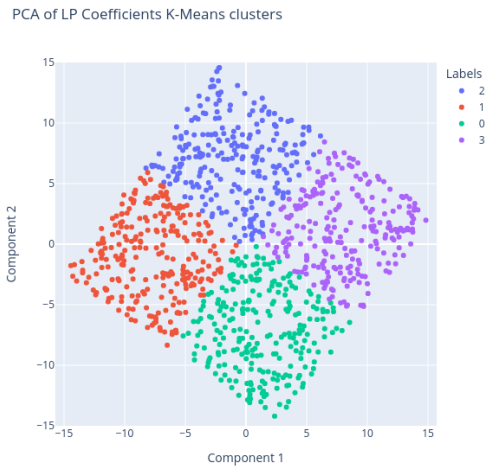

(a) Original cluster densities from PCA



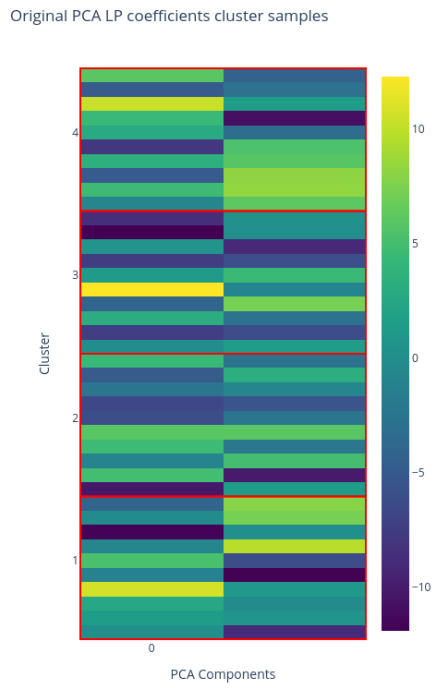

(b) K-Means clusters densities from PCA



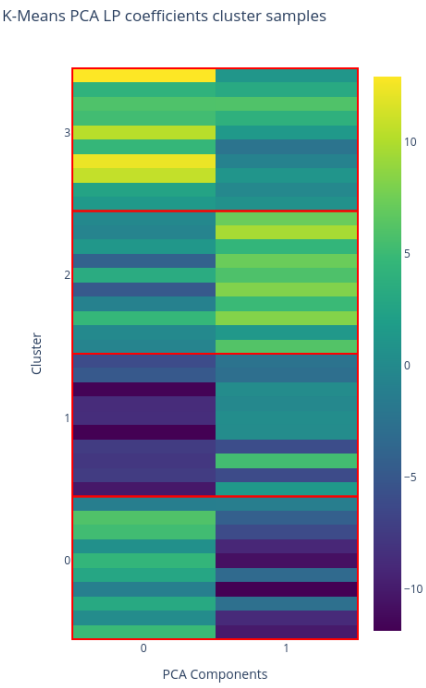

(c) Original clusters




(d) K-Means clusters from PCA

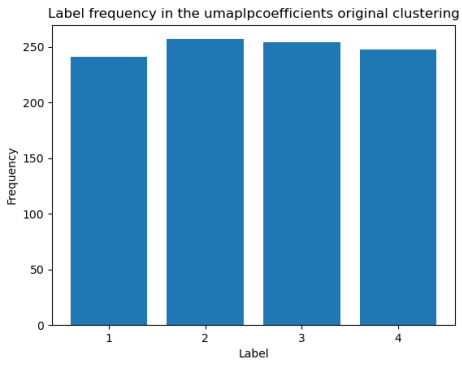



(e) Original cluster samples from PCA

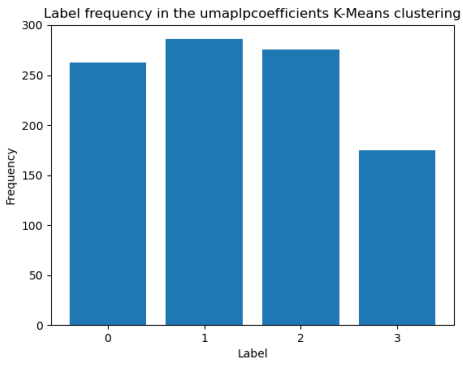



(f) K-Means cluster samples from PCA

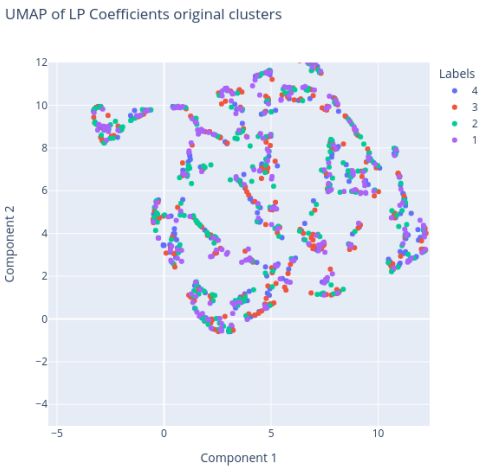
Figure 9: Comparison between original clustering and K-Means clustering from PCA of LP coefficients



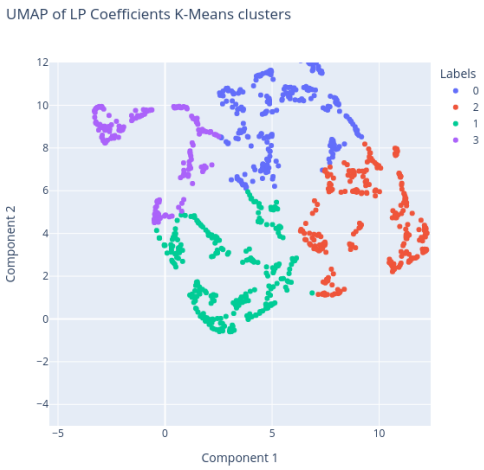

(a) Original cluster densities from UMAP



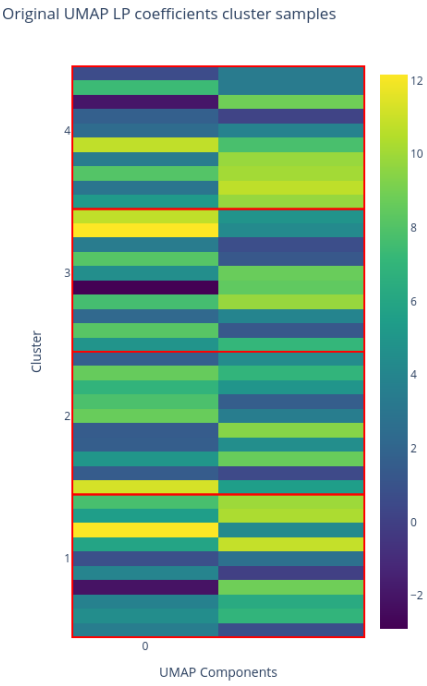

(b) K-Means clusters densities from UMAP



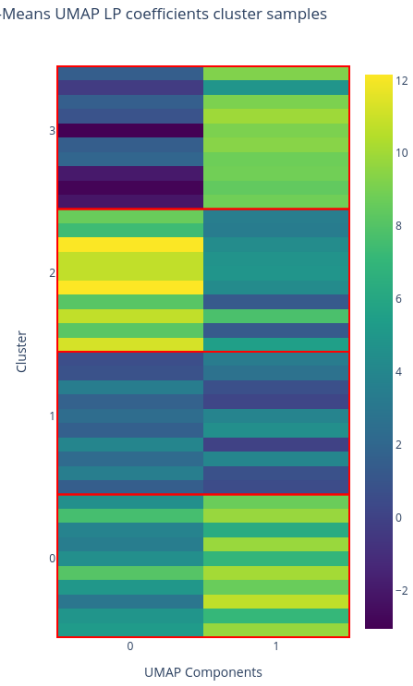

(c) Original clusters from UMAP




(d) K-Means clusters from UMAP




(e) Original cluster samples from UMAP




(f) K-Means cluster samples from UMAP

Figure 10: Comparison between original clustering and K-Means clustering from

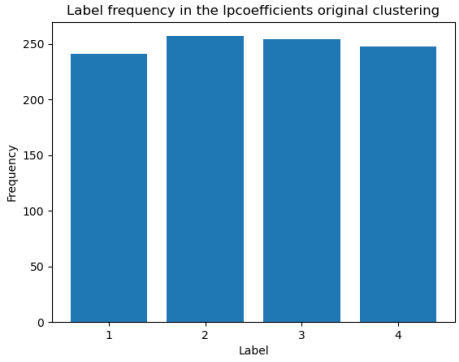
### 4.3.2 DBSCAN

A configuration that outputs 4 clusters is searched

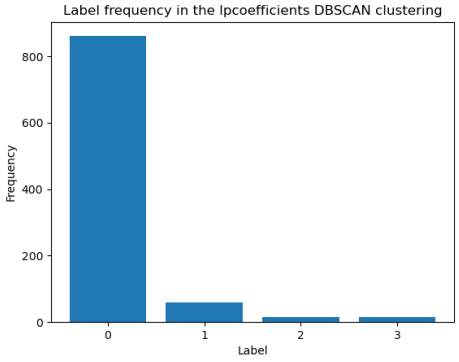
| | **Number of neighbours** | **Epsilon** |
|---|---|---|
| Original LP coefficients | 15 | 1.52 |
| PCA LP coefficients | 15 | 1.52 |
| UMAP LP coefficients | 10 | 0.85 |

Table 7: DBSCAN hyperparameter configuration for LP coefficients clustering

The results are the following:

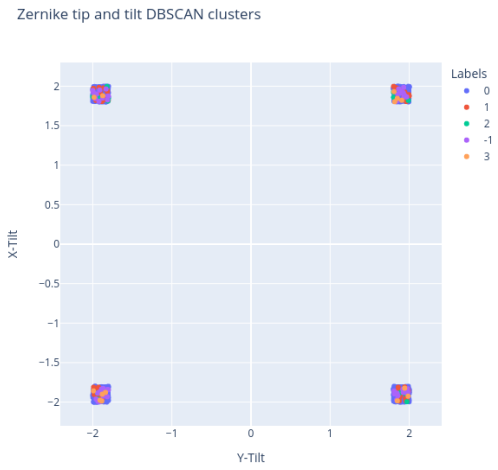



(a) Original cluster densities

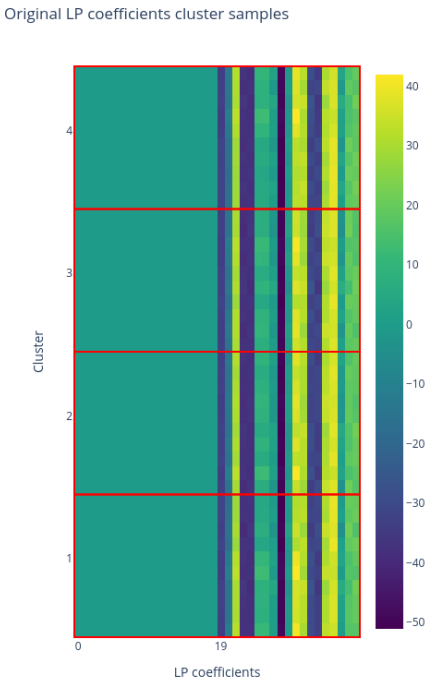



(b) DBSCAN clusters densities

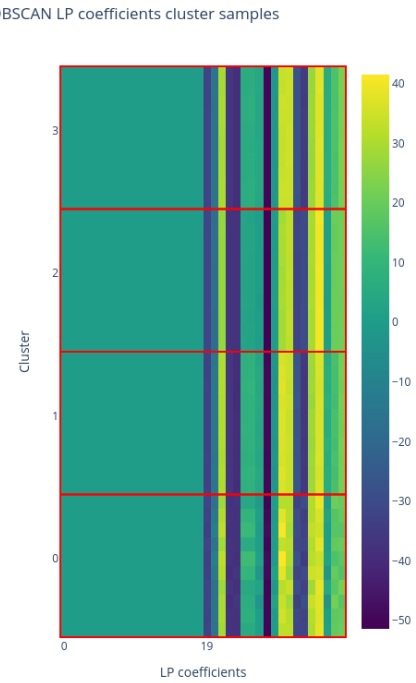

(c) Original clusters




(d) DBSCAN clusters

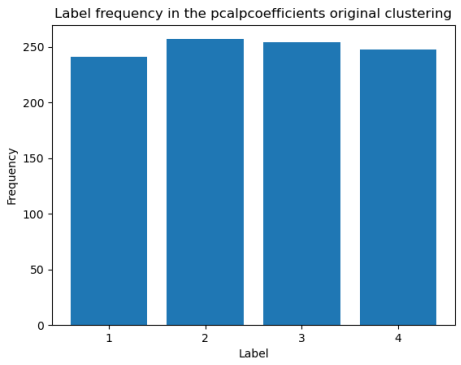



(e) Original cluster samples

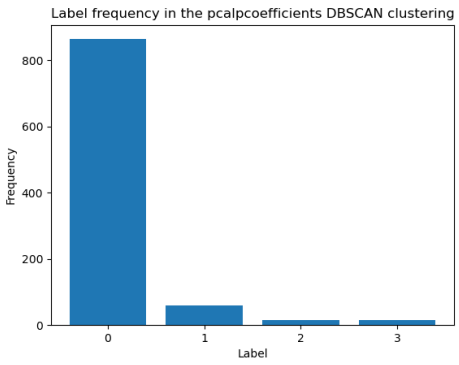



(f) DBSCAN cluster samples

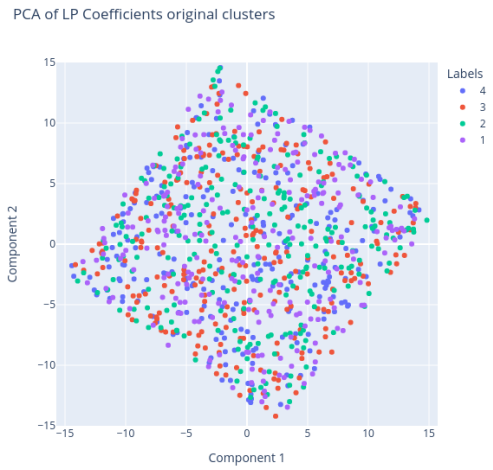
Figure 11: Comparison between original clustering and DBSCAN clustering



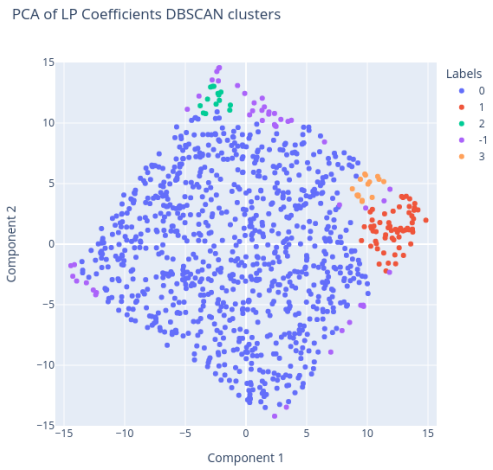

(a) Original cluster densities from PCA



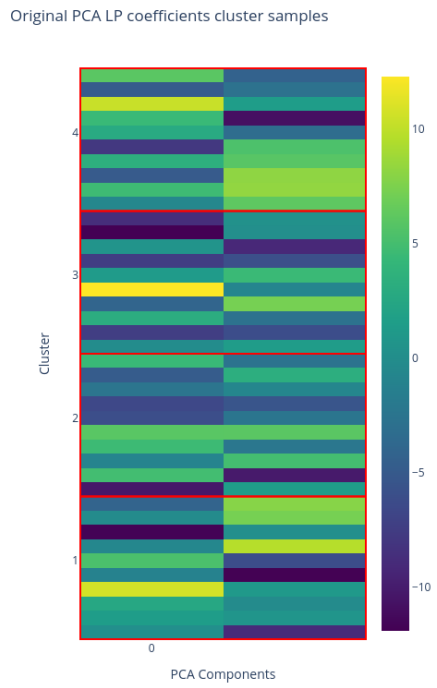

(b) DBSCAN clusters densities from PCA



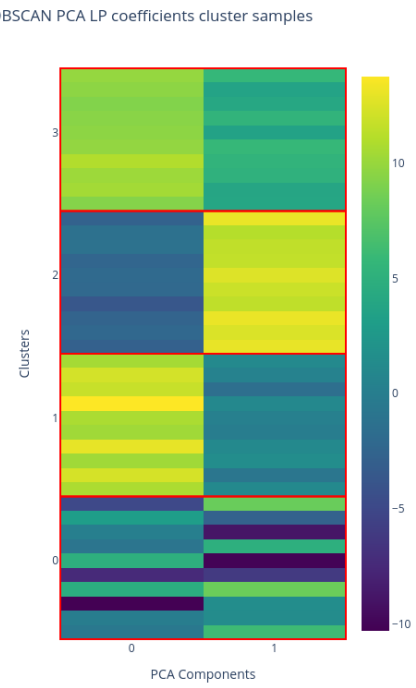

(c) Original clusters from PCA




(d) DBSCAN clusters from PCA

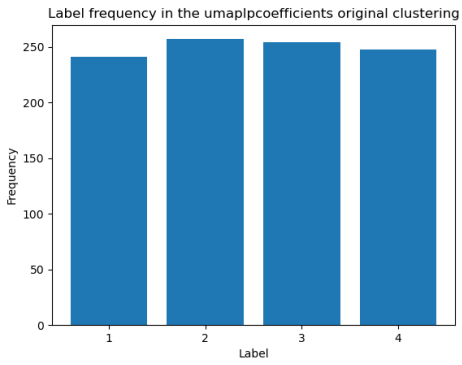



(e) Original cluster samples from PCA

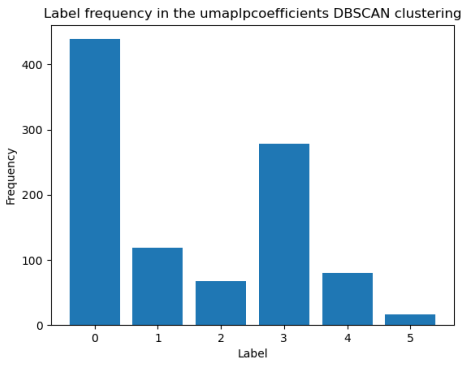



(f) DBSCAN cluster samples from PCA

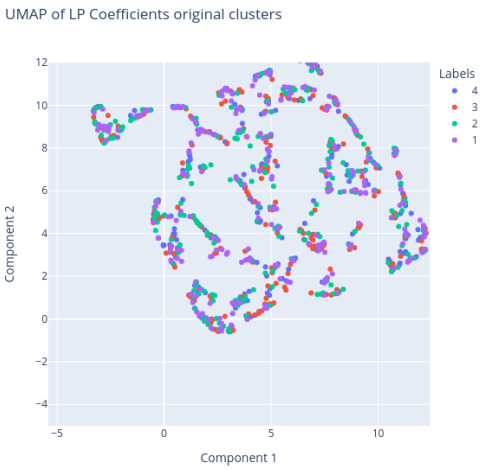
Figure 12: Comparison between original clustering and DBSCAN clustering from PCA of LP coefficients



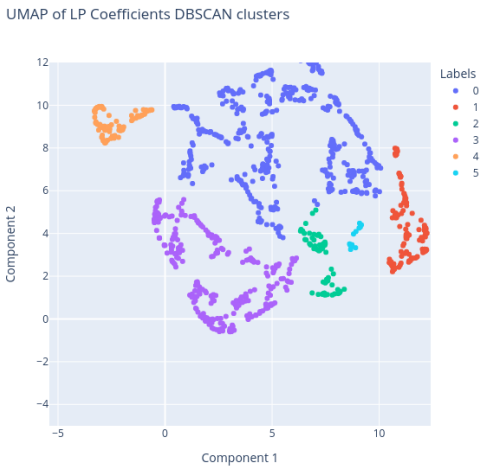

(a) Original cluster densities from UMAP



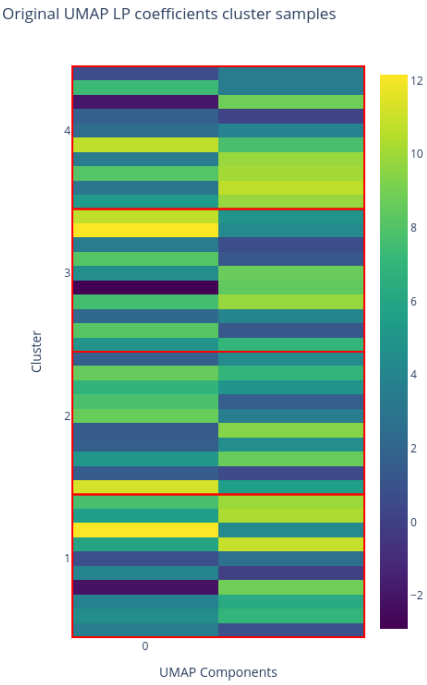

(b) DBSCAN clusters densities from UMAP



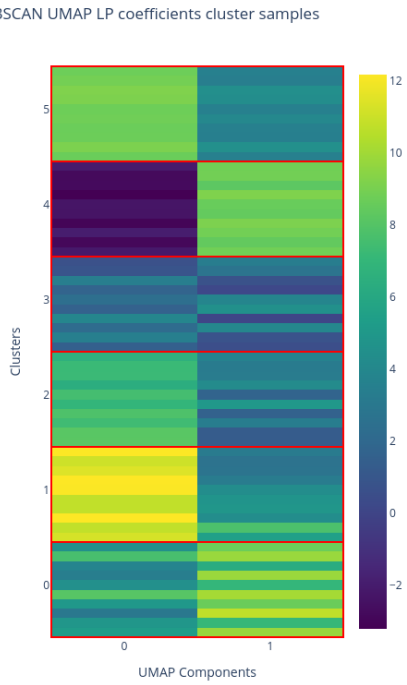

(c) Original clusters from UMAP




(d) DBSCAN clusters from UMAP




(e) Original cluster samples from UMAP




(f) DBSCAN cluster samples from UMAP

Figure 13: Comparison between original clustering and DBSCAN clustering from

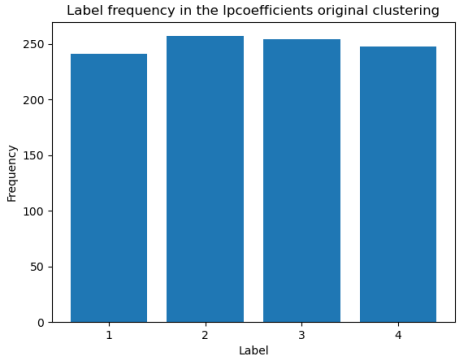
### 4.3.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

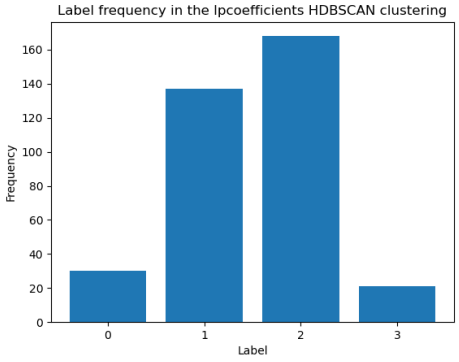
| | **Minimum cluster size** |
|---|---|
| Original LP coefficients | 21 |
| PCA LP coefficients | 21 |
| UMAP LP coefficients | 25 |

Table 8: HDBSCAN hyperparameter configuration for LP coefficients clustering

The results are the following:

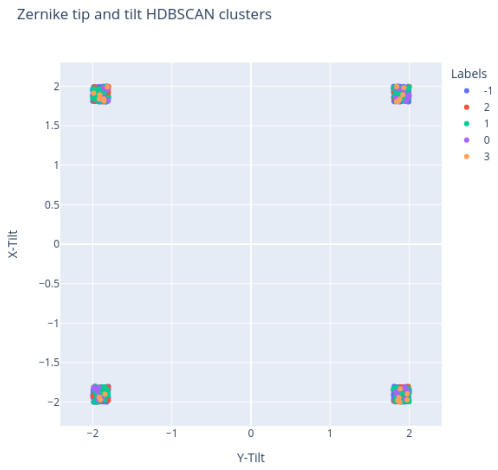



(a) Original cluster densities

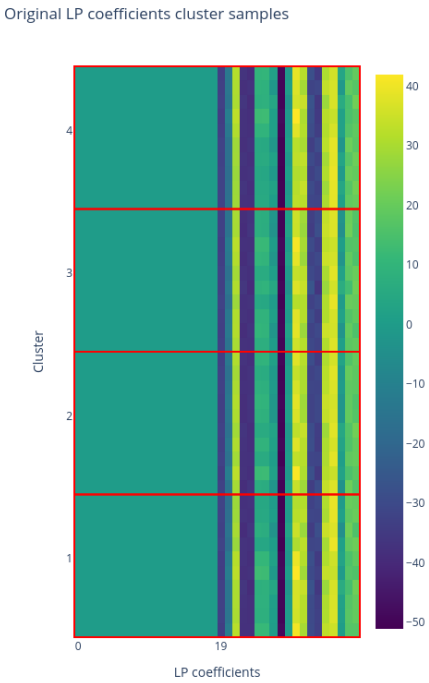



(b) HDBSCAN clusters densities

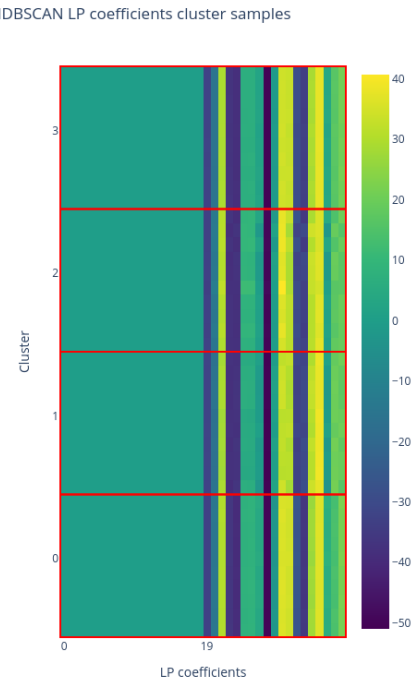

(c) Original clusters




(d) HDBSCAN clusters

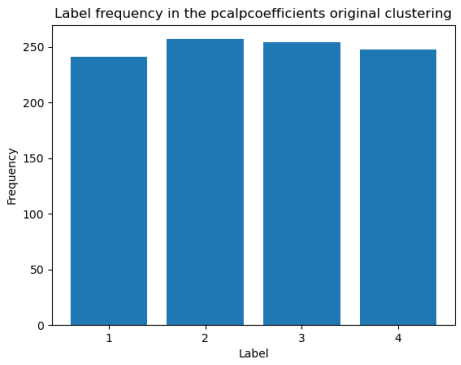



(e) Original cluster samples

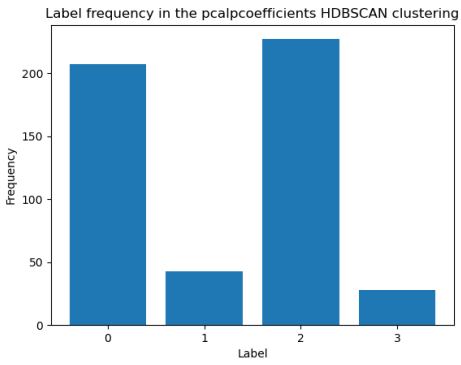



(f) HDBSCAN cluster samples

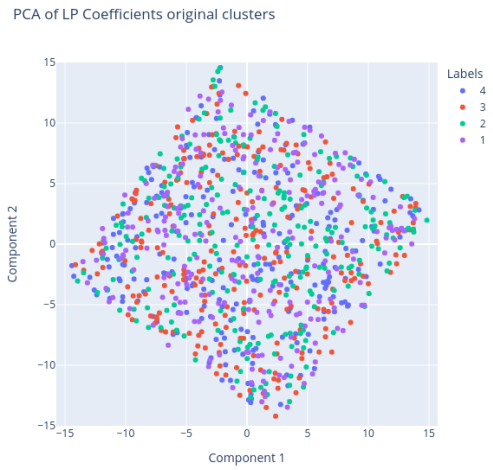
Figure 14: Comparison between original clustering and HDBSCAN clustering



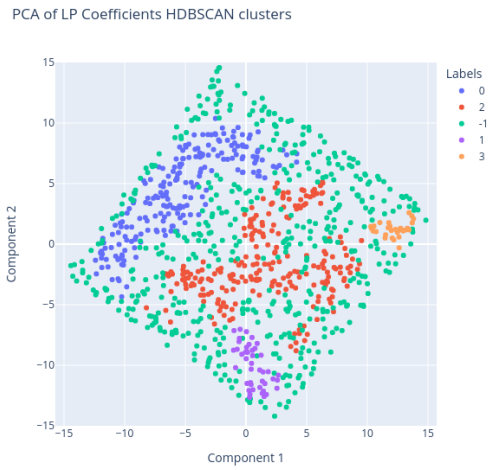

(a) Original cluster densities from PCA



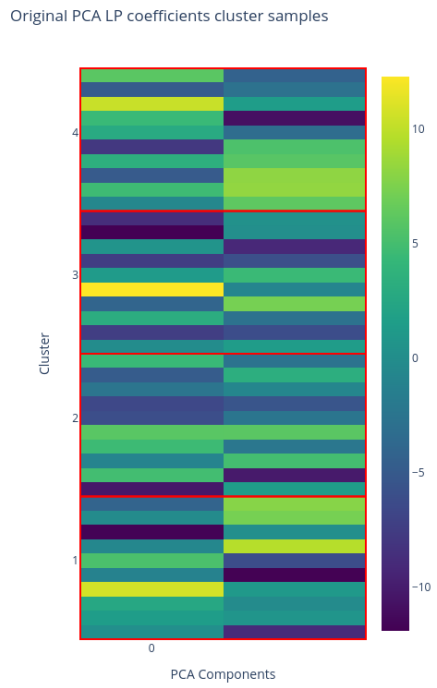

(b) HDBSCAN clusters densities from PCA



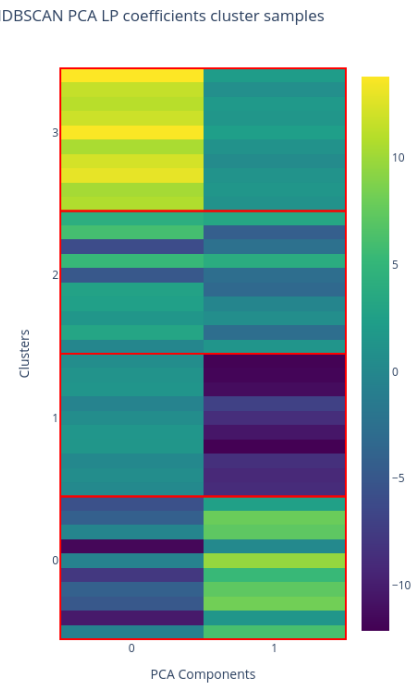

(c) Original clusters from PCA




(d) HDBSCAN clusters from PCA

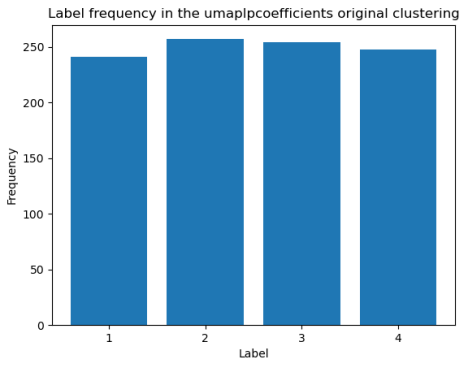



(e) Original cluster samples from PCA

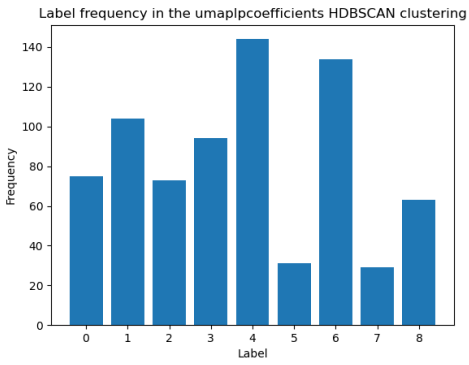



(f) HDBSCAN cluster samples from PCA

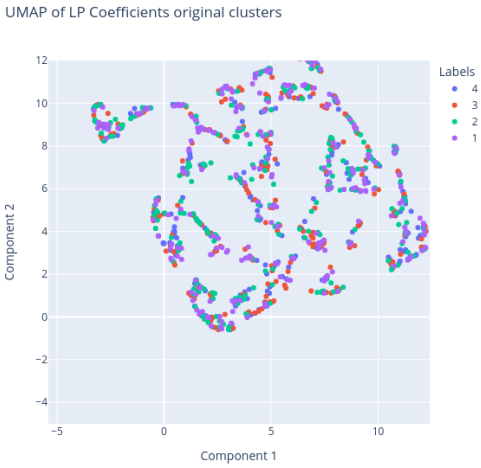
Figure 15: Comparison between original clustering and HDBSCAN clustering from



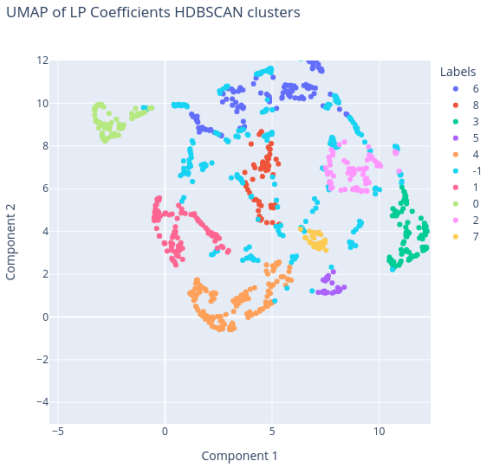

(a) Original cluster densities from UMAP



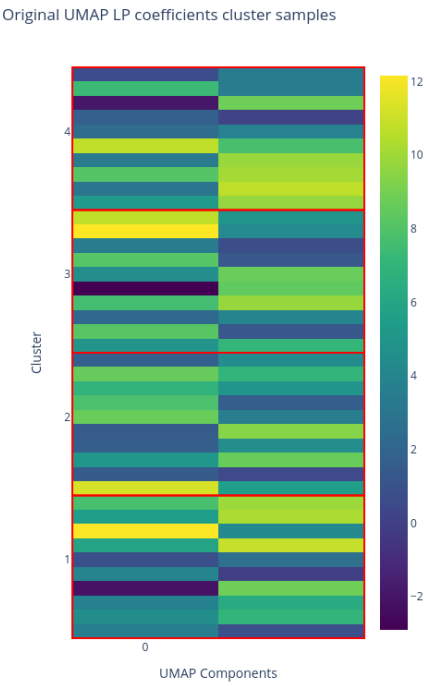

(b) HDBSCAN clusters densities from UMAP



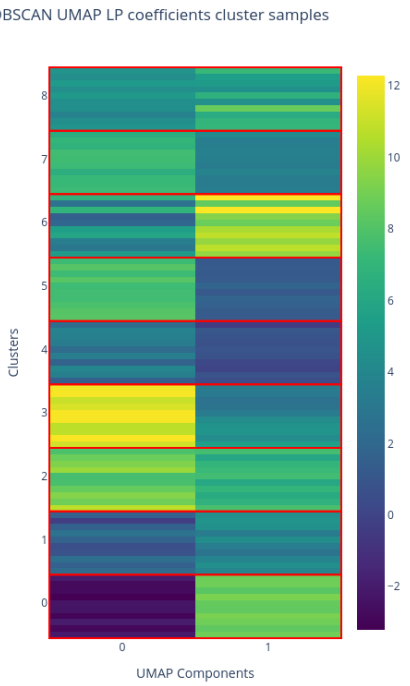

(c) Original clusters from UMAP




(d) HDBSCAN clusters from UMAP




(e) Original cluster samples from UMAP



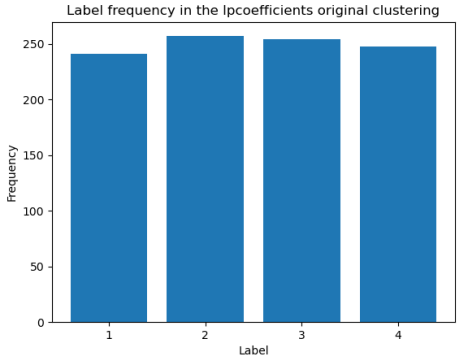

(f) HDBSCAN cluster samples from UMAP

### 4.3.4 Agglomerative clustering

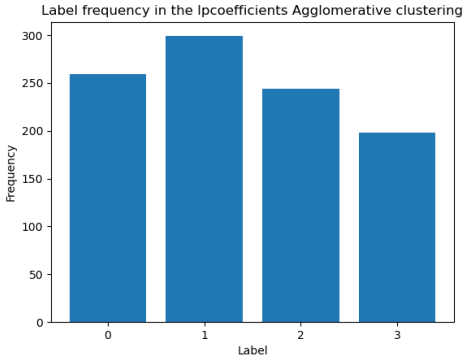
| | **Number of clusters** |
|---|---|
| Original LP coefficients | 4 |
| PCA LP coefficients | 4 |
| UMAP LP coefficients | 4 |

Table 9: Agglomerative hyperparameter configuration for LP coefficients clustering

The results are the following:

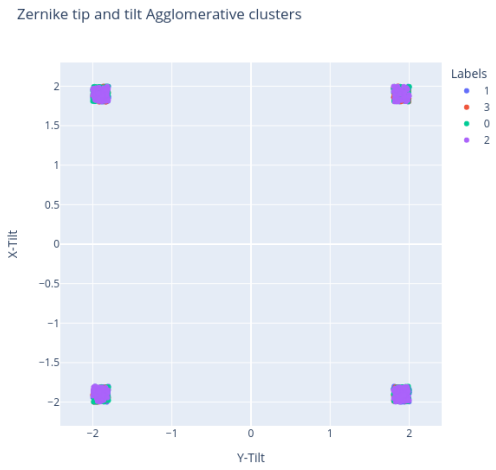



(a) Original cluster densities

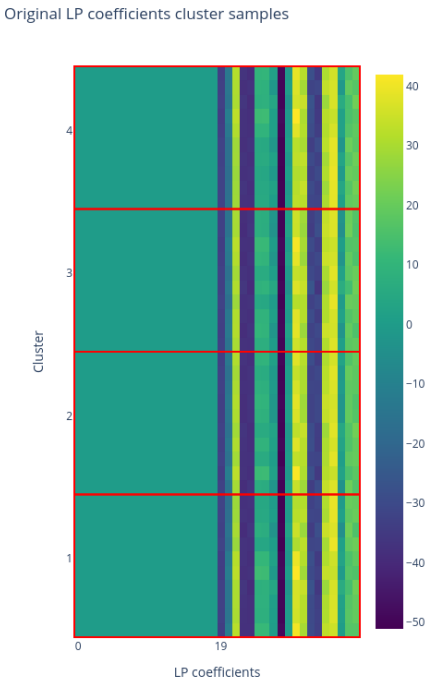



(b) Agglomerative clusters densities

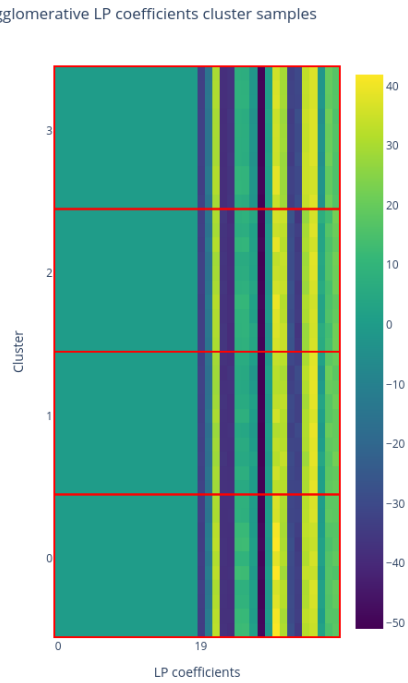

(c) Original clusters




(d) Agglomerative clusters

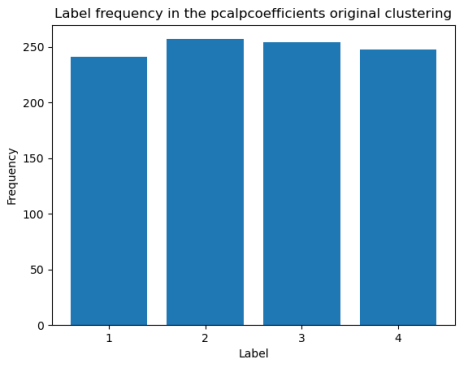



(e) Original cluster samples

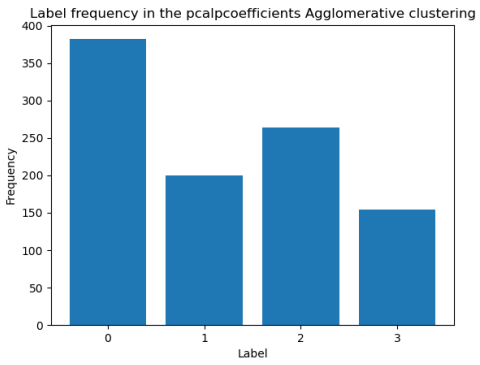



(f) Agglomerative cluster samples

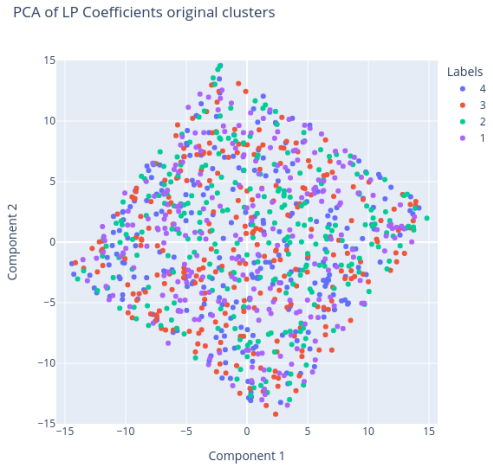
Figure 17: Comparison between original clustering and Agglomerative clustering



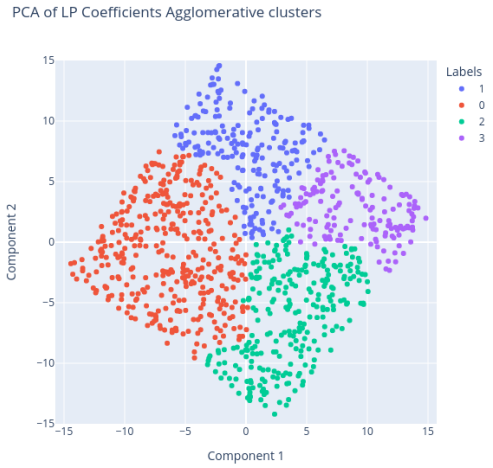

(a) Original cluster densities from PCA



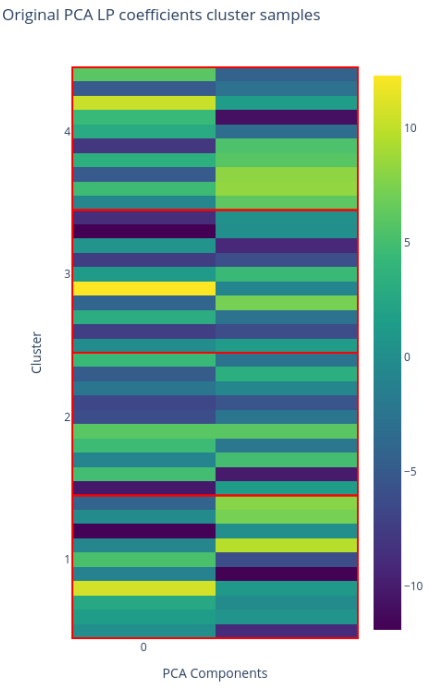

(b) Agglomerative clusters densities from PCA



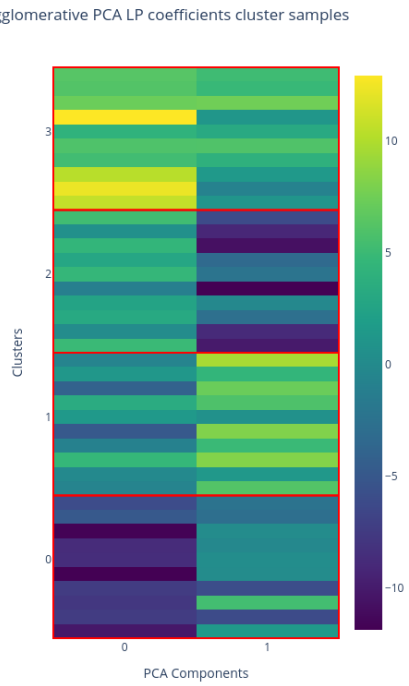

(c) Original clusters from PCA



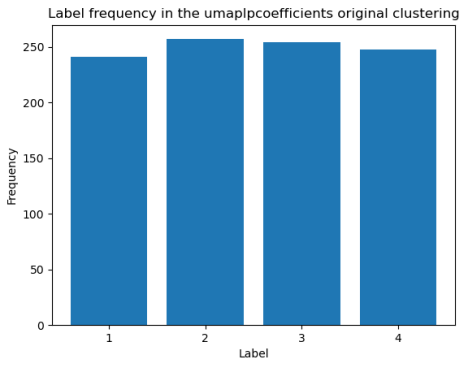

(d) Agglomerative clusters from PCA



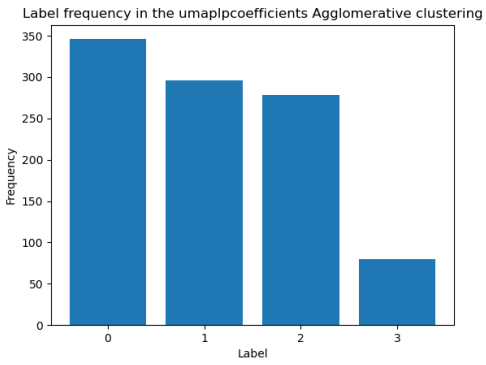

(e) Original cluster samples from PCA



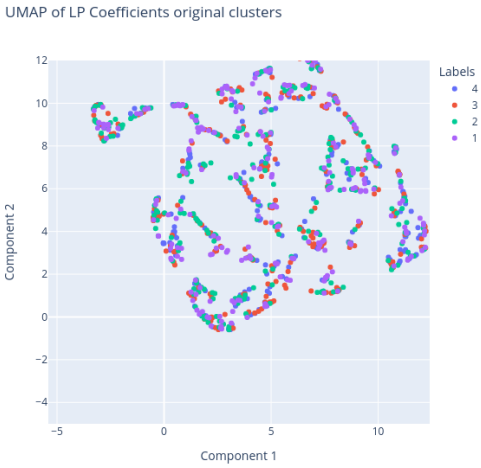

(f) Agglomerative cluster samples from PCA



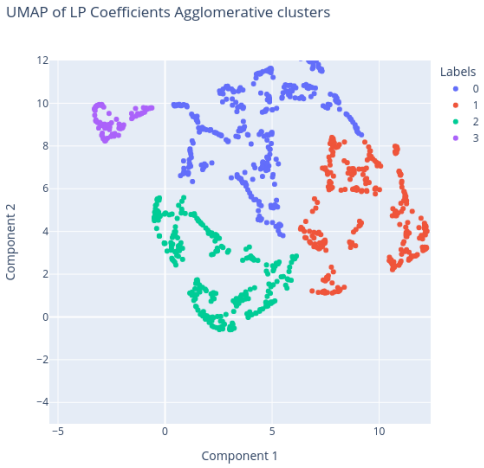

(a) Original cluster densities from UMAP



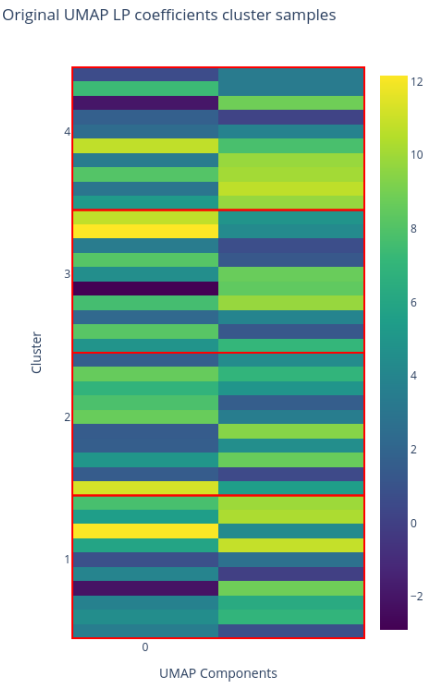

(b) Agglomerative clusters densities from UMAP



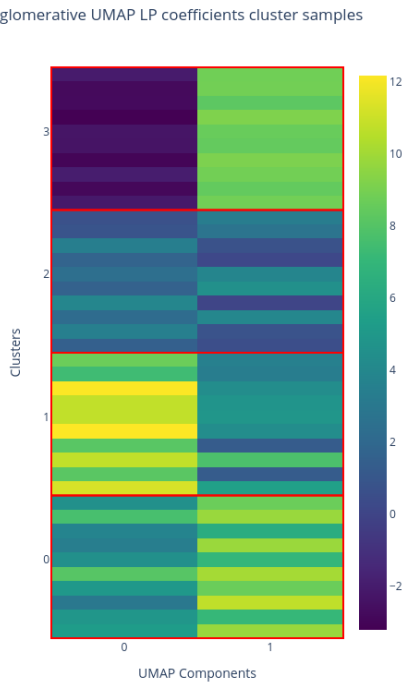

(c) Original clusters from UMAP




(d) Agglomerative clusters from UMAP

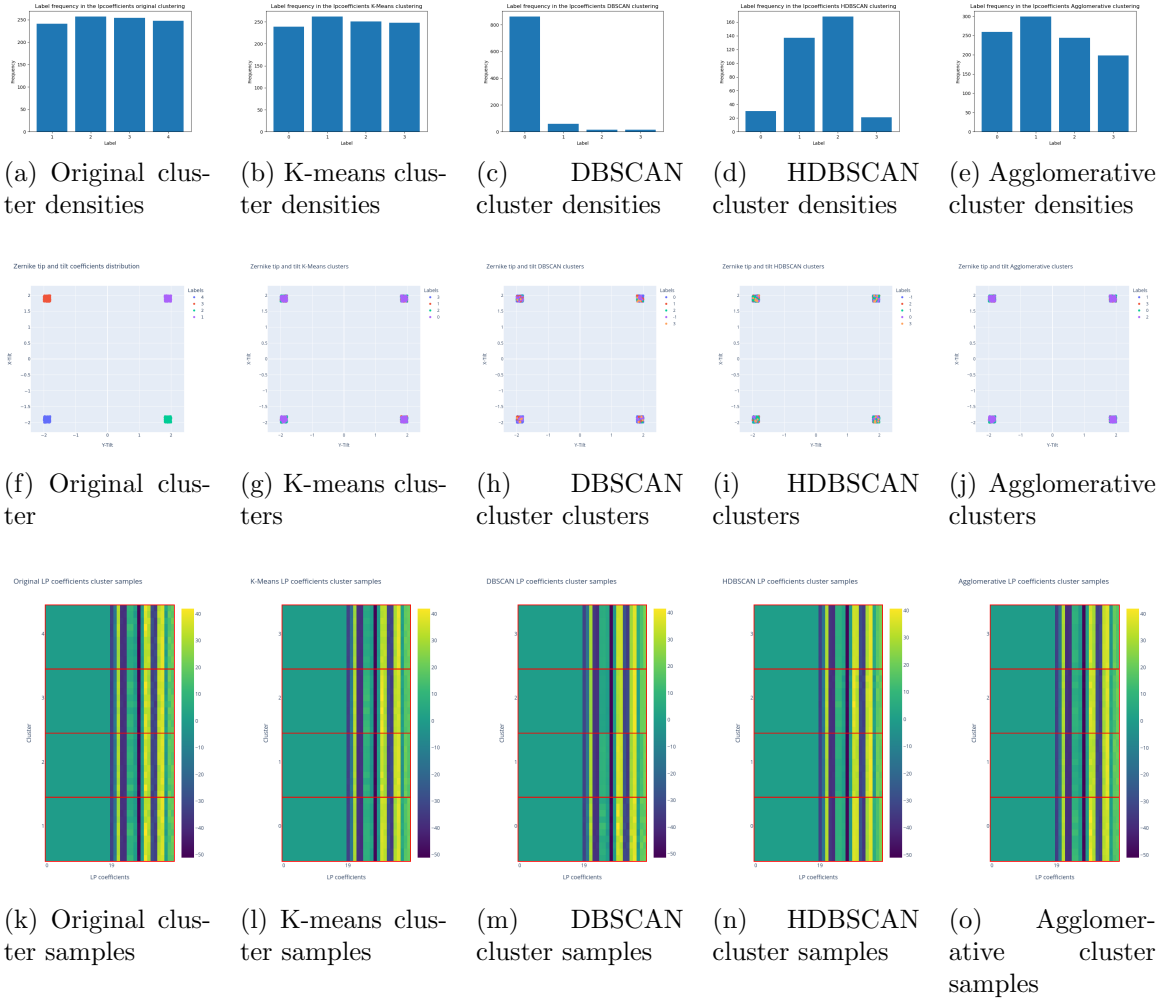



(e) Original cluster samples from UMAP




(f) Agglomerative cluster samples from UMAP

## 4.4 Summary



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

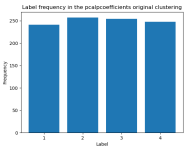
(f) Original cluster (g) K-means clusters (h) DBSCAN cluster clusters (i) HDBSCAN clusters (j) Agglomerative clusters

(k) Original cluster samples (l) K-means cluster samples (m) DBSCAN cluster samples (n) HDBSCAN cluster samples (o) Agglomerative cluster samples

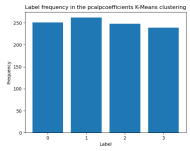
Figure 20: Comparison between clustering LP coefficients algorithms

| | **Original** | **K-Means** | **DBSCAN** | **HDBSCAN** | **Agglomerative** |
|---|---|---|---|---|---|
| **Original** | \ | 0.002 | 0.008 | 0.004 | 0.002 |
| **K-Means** | | \ | 0.154 | 0.143 | 0.695 |
| **DBSCAN** | | | \ | 0.175 | 0.142 |
| **HDBSCAN** | | | | \ | 0.002 |

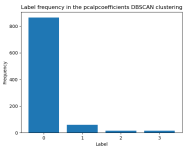
Table 10: Normalized Mutual Information between original LP coefficients clusters



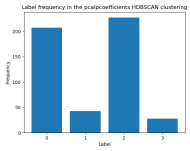
(a) Original cluster densities from PCA



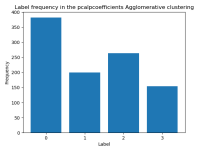
(b) K-means cluster densities from PCA



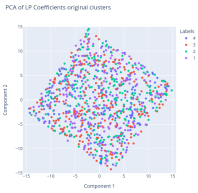
(c) DBSCAN cluster densities from PCA



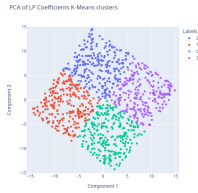
(d) HDBSCAN cluster densities from PCA



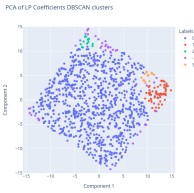
(e) Agglomerative cluster densities from PCA



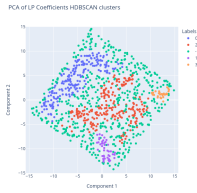
(f) Original cluster from PCA



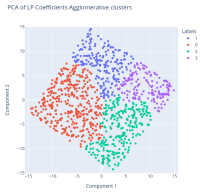
(g) K-means clusters from PCA



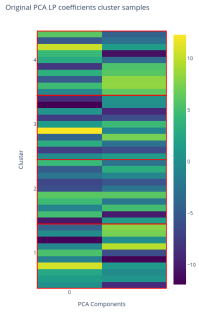
(h) DBSCAN cluster clusters from PCA



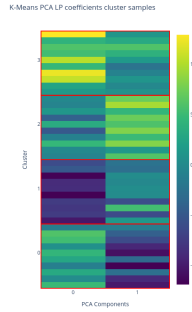
(i) HDBSCAN clusters from PCA



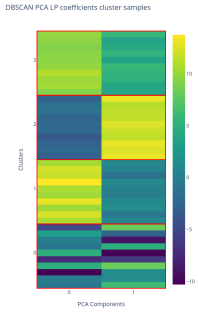
(j) Agglomerative clusters from PCA



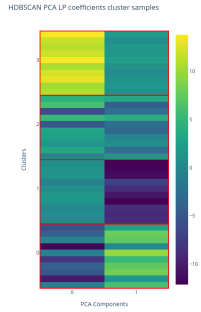
(k) Original cluster samples from PCA



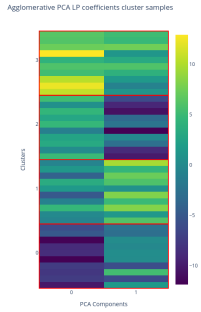
(l) K-means cluster samples from PCA



(m) DBSCAN cluster samples-from PCA



(n) HDBSCAN cluster samples from PCA

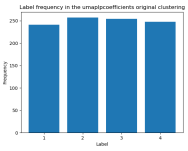


(o) Agglomerative cluster samples from PCA

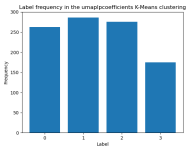
Figure 21: Comparison between clustering PCA LP coefficients algorithms

| | **Original** | **K-Means** | **DBSCAN** | **HDBSCAN** | **Agglomerative** |
|---|---|---|---|---|---|
| **Original** | \ | 0.002 | 0.003 | 0.004 | 0.003 |
| **K-Means** | | \ | 0.150 | 0.143 | 0.642 |
| **DBSCAN** | | | \ | 0.174 | 0.209 |
| **HDBSCAN** | | | | \ | 0.003 |

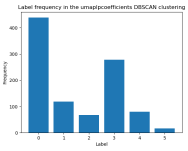
Table 11: Normalized Mutual Information between PCA LP coefficients clusters



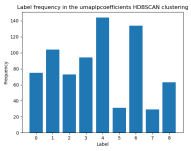
(a) Original cluster densities from UMAP



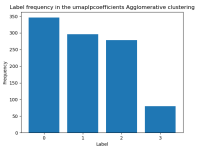
(b) K-means cluster densities from UMAP



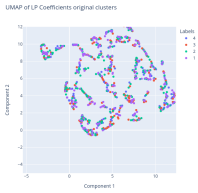
(c) DBSCAN cluster densities from UMAP



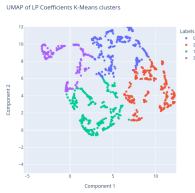
(d) HDBSCAN cluster densities from UMAP



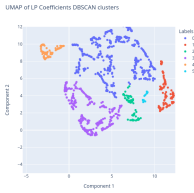
(e) Agglomerative cluster densities from UMAP



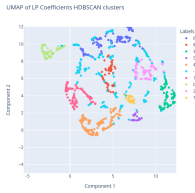
(f) Original cluster from UMAP



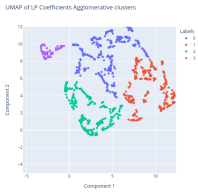
(g) K-means clusters from UMAP



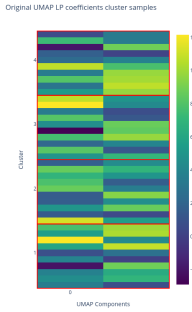
(h) DBSCAN cluster clusters from UMAP



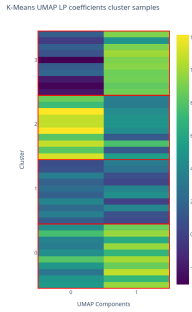
(i) HDBSCAN clusters from UMAP



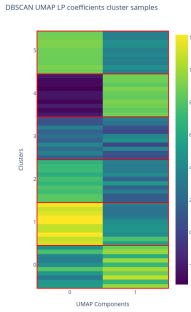
(j) Agglomerative clusters from UMAP



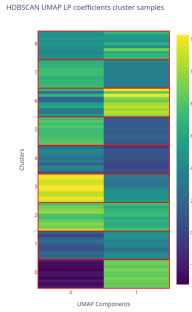
(k) Original cluster samples from UMAP



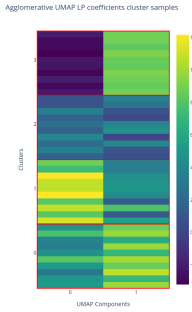
(l) K-means cluster samples from UMAP



(m) DBSCAN cluster samples-from UMAP



(n) HDBSCAN cluster samples from UMAP



(o) Agglomerative cluster samples from UMAP

Figure 22: Comparison between clustering UMAP LP coefficients algorithms

| | **Original** | **K-Means** | **DBSCAN** | **HDBSCAN** | **Agglomerative** |
|---|---|---|---|---|---|
| **Original** | \ | 0.003 | 0.001 | 0.009 | 0.001 |
| **K-Means** | | \ | 0.438 | 0.480 | 0.667 |
| **DBSCAN** | | | \ | 0.487 | 0.529 |
| **HDBSCAN** | | | | \ | 0.001 |

Table 12: Normalized Mutual Information between UMAP LP coefficients clusters

## 4.5 Output fluxes clustering

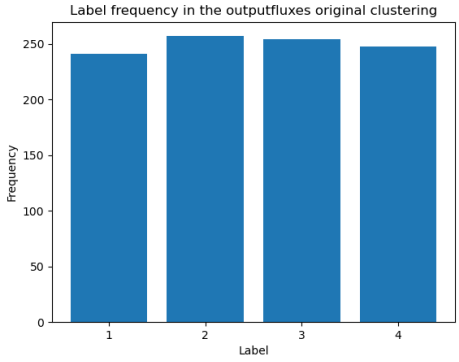
### 4.5.1 K-Means

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.

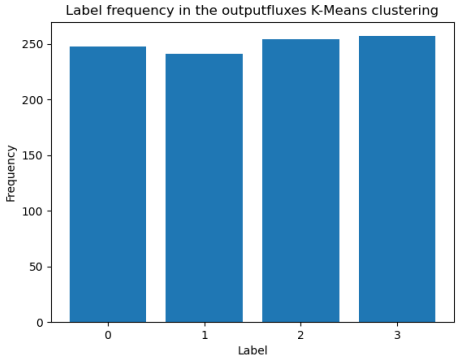
| | **Number of clusters** | **Number of initializations** |
|---|---|---|
| **Original Output fluxes** | 4 | 10 |
| **PCA Output fluxes** | 4 | 10 |
| **UMAP Output fluxes** | 4 | 10 |

Table 13: K-Means hyperparameter configuration for c coefficients clustering

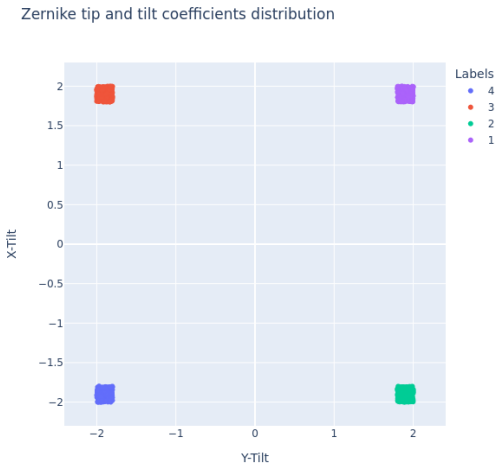
The results are the following:



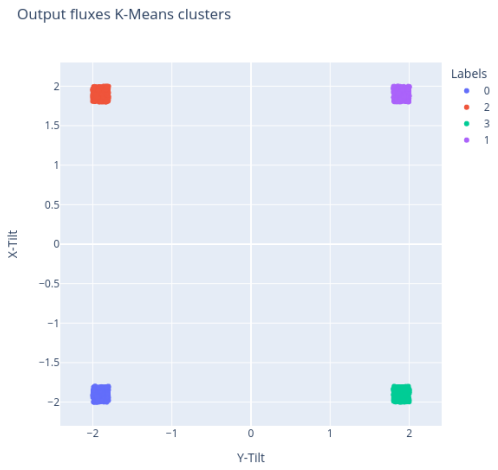

(a) Original cluster densities



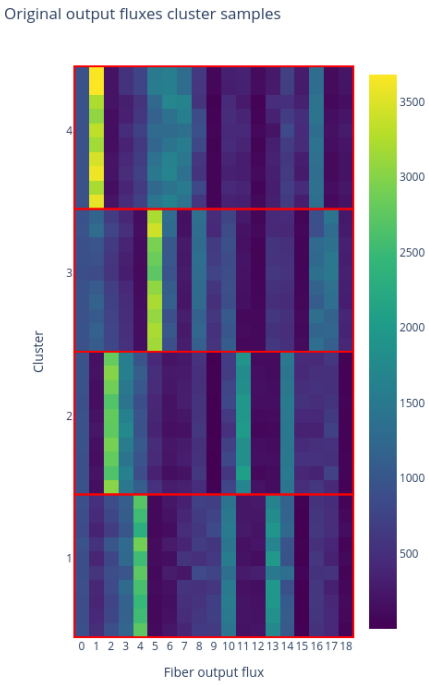

(b) K-Means clusters densities



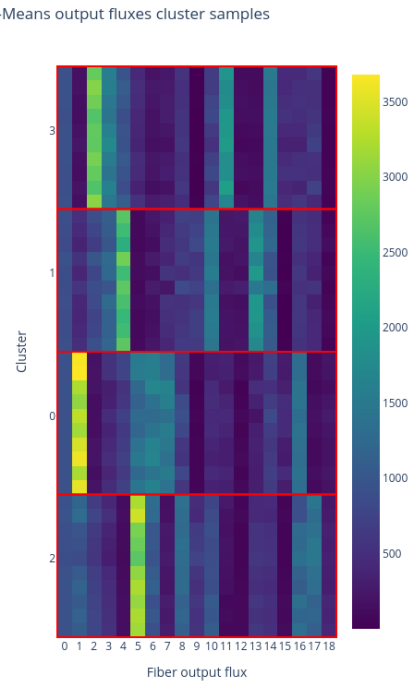

(c) Original clusters




(d) K-Means clusters

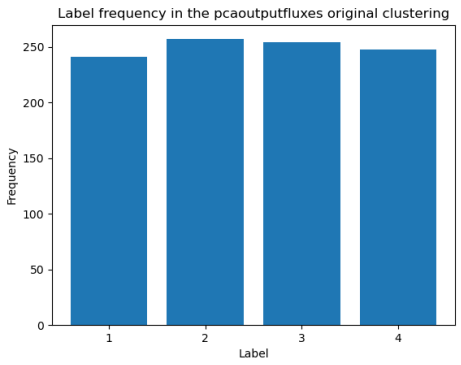



(e) Original cluster samples

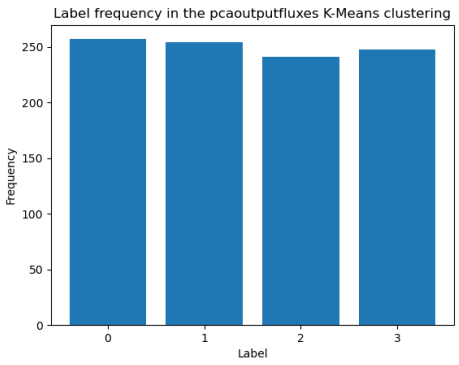



(f) K-Means cluster samples

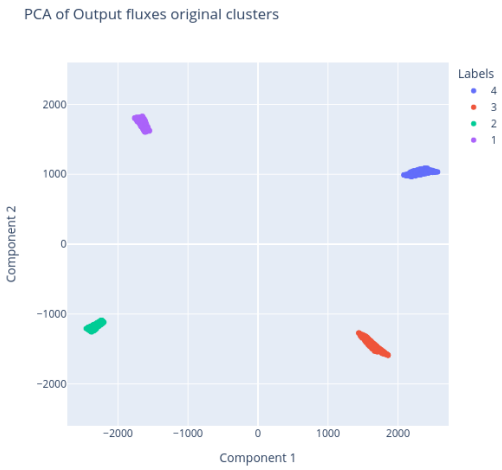
Figure 23: Comparison between original clustering and K-Means clustering from original Output fluxes



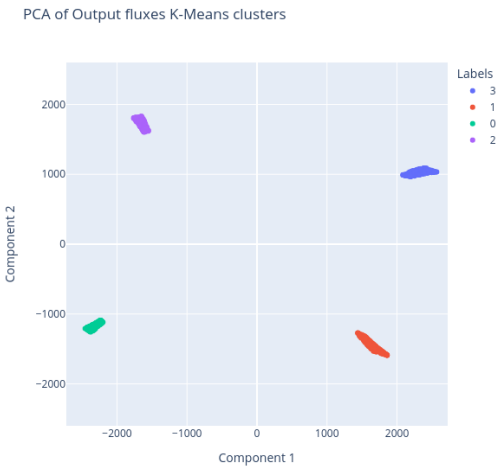

(a) Original cluster densities from PCA



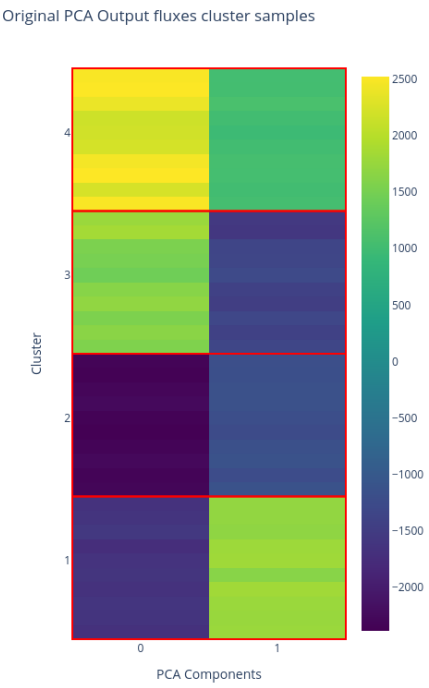

(b) K-Means clusters densities from PCA




(c) Original clusters



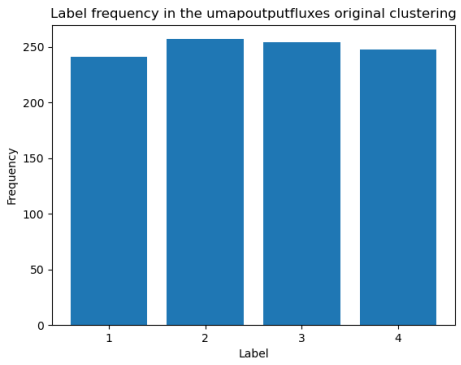

(d) K-Means clusters from PCA



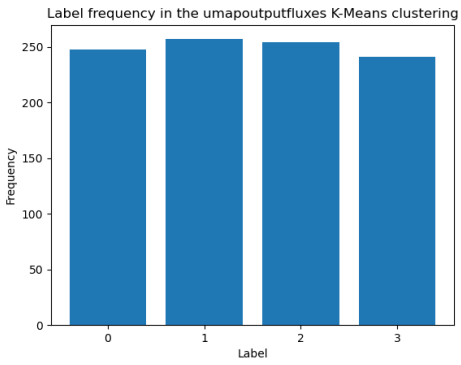

(e) Original cluster samples from PCA


(f) K-Means cluster samples from PCA

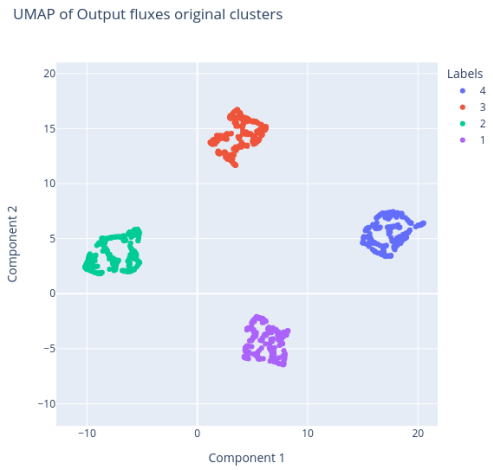
Figure 24: Comparison between original clustering and K-Means clustering from PCA of Output fluxes



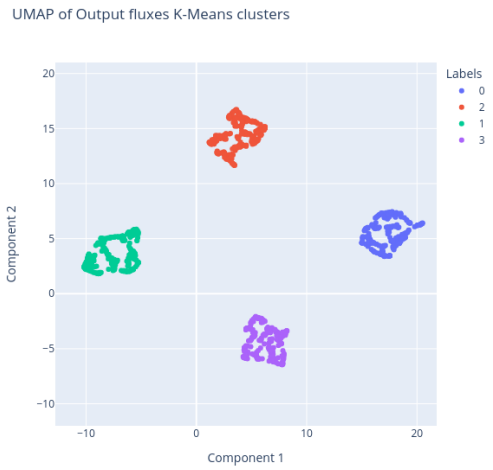

(a) Original cluster densities from UMAP



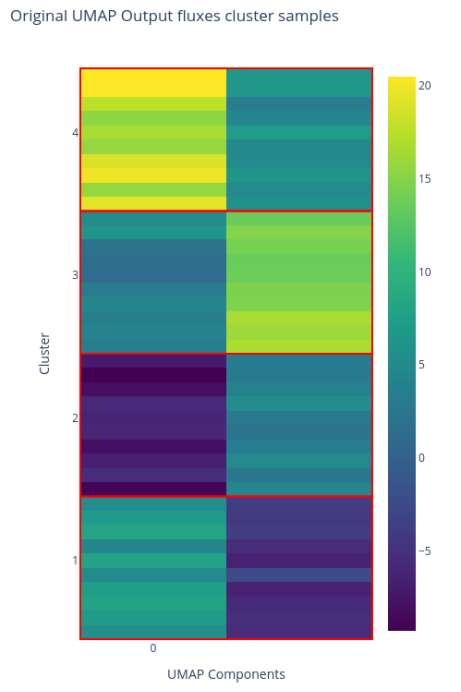

(b) K-Means clusters densities from UMAP



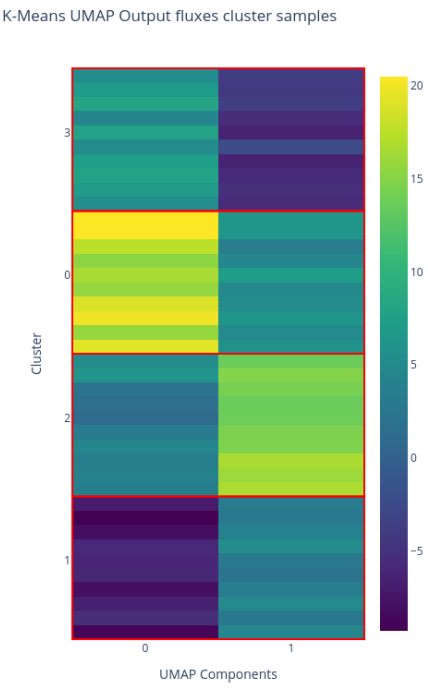

(c) Original clusters from UMAP




(d) K-Means clusters from UMAP




(e) Original cluster samples from UMAP




(f) K-Means cluster samples from UMAP

Figure 25: Comparison between original clustering and K-Means clustering from

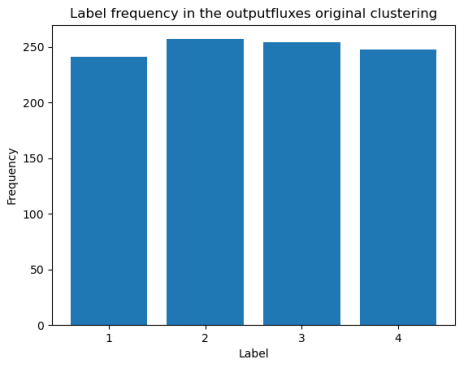
### 4.5.2 DBSCAN

A configuration that outputs 4 clusters is searched

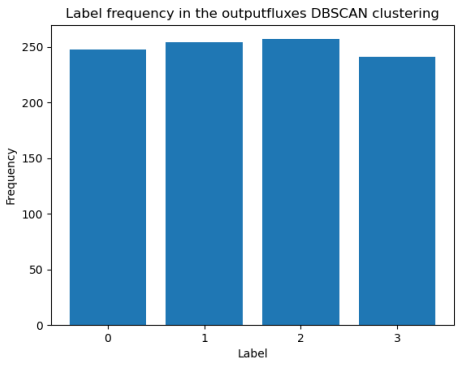
| | **Number of neighbours** | **Epsilon** |
|---|---|---|
| Original Output fluxes | 7 | 120 |
| PCA Output fluxes | 15 | 40 |
| UMAP Output fluxes | 10 | 0.85 |

Table 14: DBSCAN hyperparameter configuration for Output fluxes clustering

The results are the following:

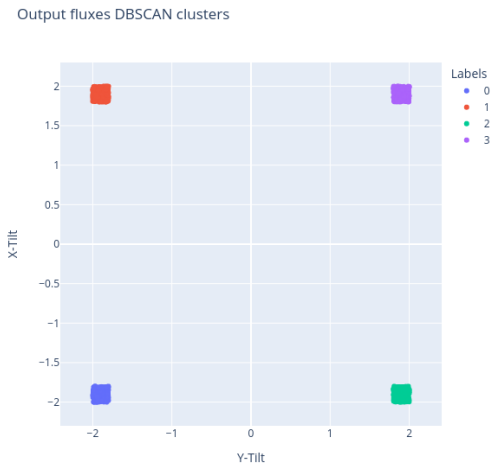



(a) Original cluster densities

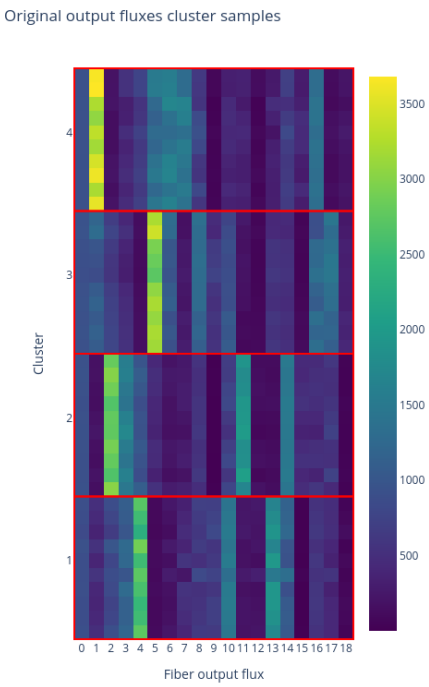



(b) DBSCAN clusters densities

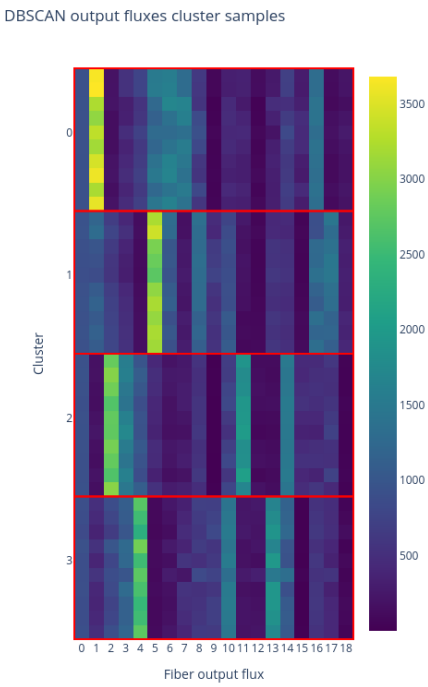

(c) Original clusters




(d) DBSCAN clusters

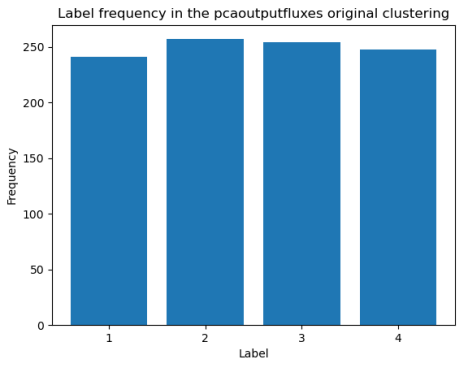



(e) Original cluster samples

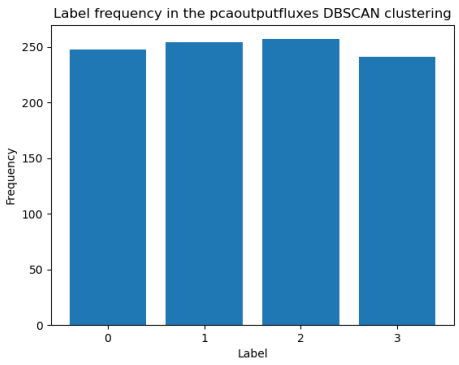



(f) DBSCAN cluster samples

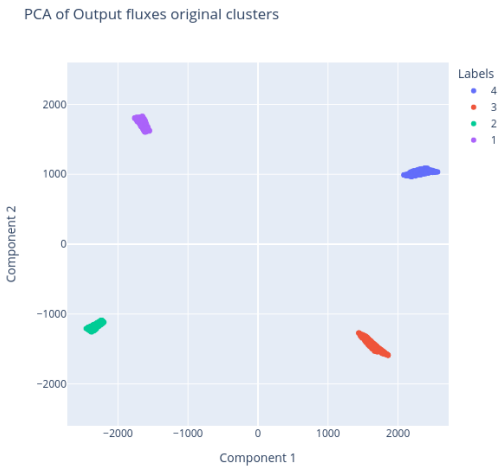
Figure 26: Comparison between original clustering and DBSCAN clustering



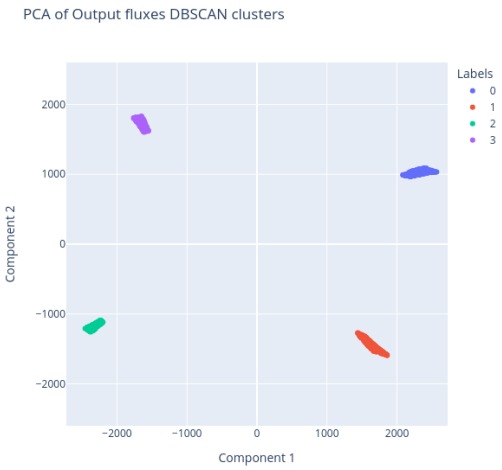

(a) Original cluster densities from PCA



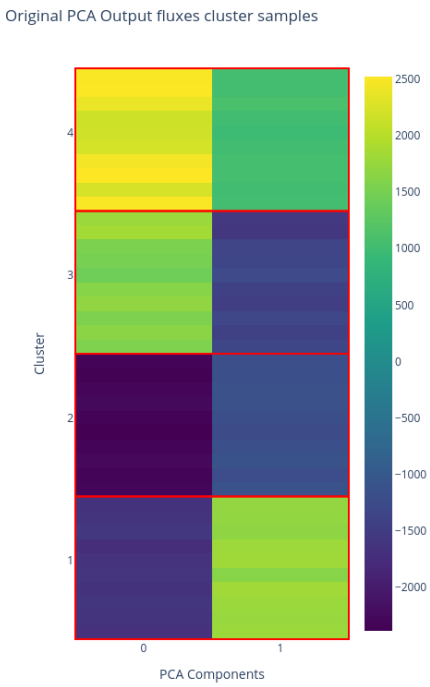

(b) DBSCAN clusters densities from PCA



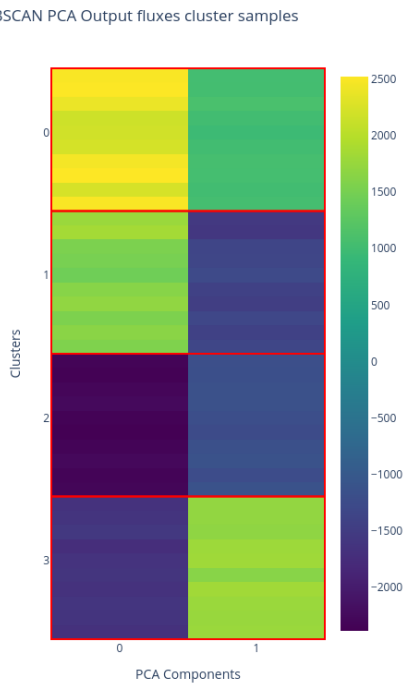

(c) Original clusters from PCA




(d) DBSCAN clusters from PCA

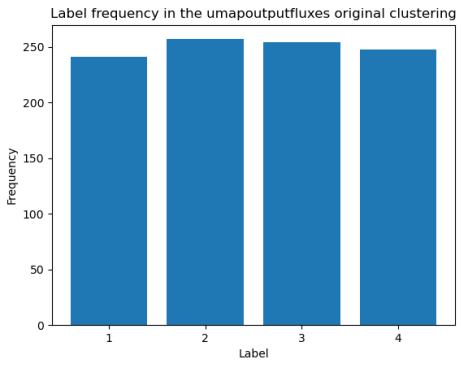



(e) Original cluster samples from PCA

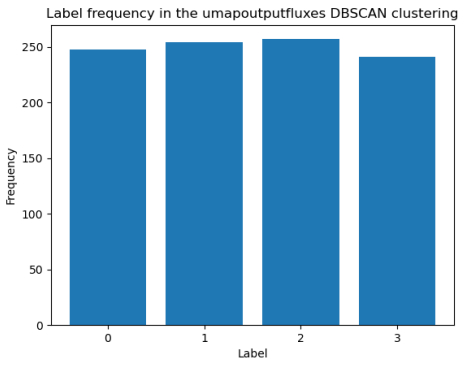



(f) DBSCAN cluster samples from PCA

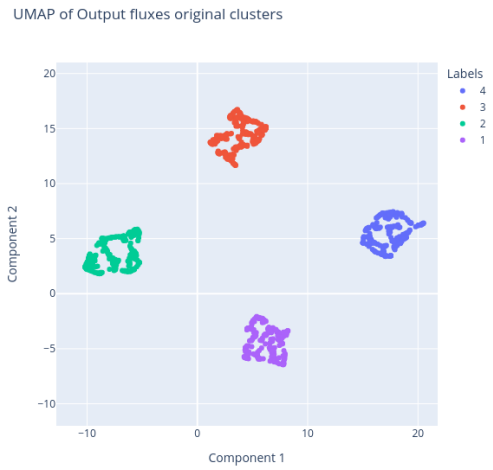
Figure 27: Comparison between original clustering and DBSCAN clustering from PCA of Output fluxes



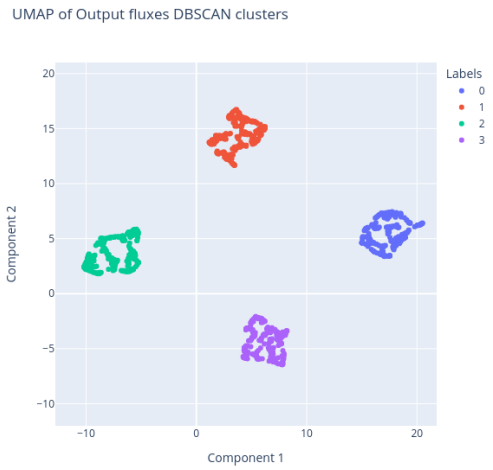

(a) Original cluster densities from UMAP



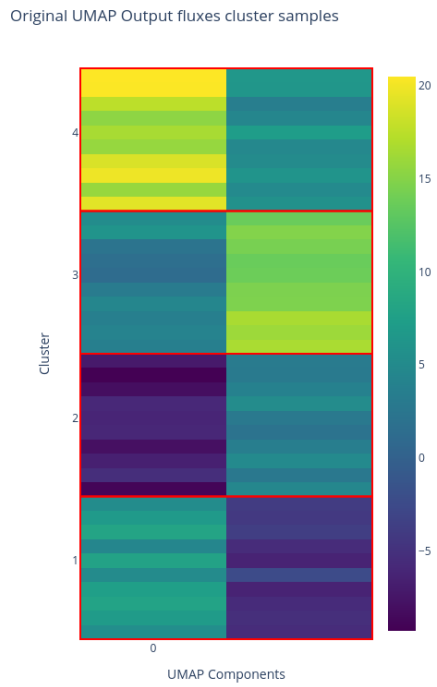

(b) DBSCAN clusters densities from UMAP



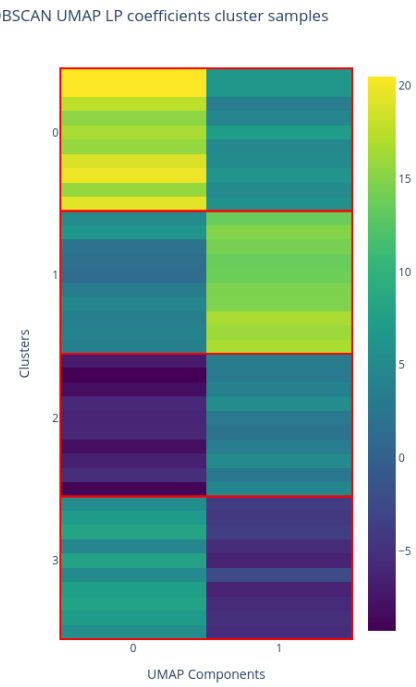

(c) Original clusters from UMAP




(d) DBSCAN clusters from UMAP




(e) Original cluster samples from UMAP




(f) DBSCAN cluster samples from UMAP

Figure 28: Comparison between original clustering and DBSCAN clustering from

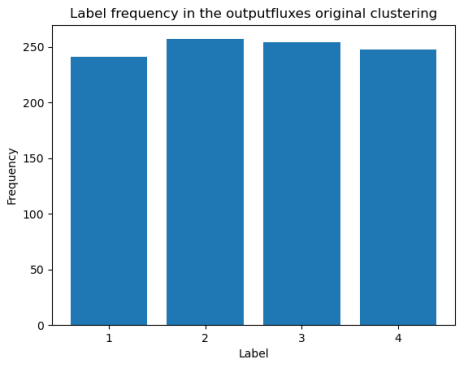
### 4.5.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

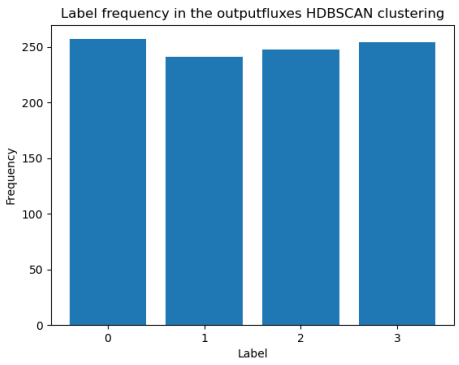
| | **Minimum cluster size** |
|---|---|
| Original Output fluxes | 21 |
| PCA Output fluxes | 21 |
| HDBSCAN Output fluxes | 25 |

Table 15: HDBSCAN hyperparameter configuration for Output fluxes clustering

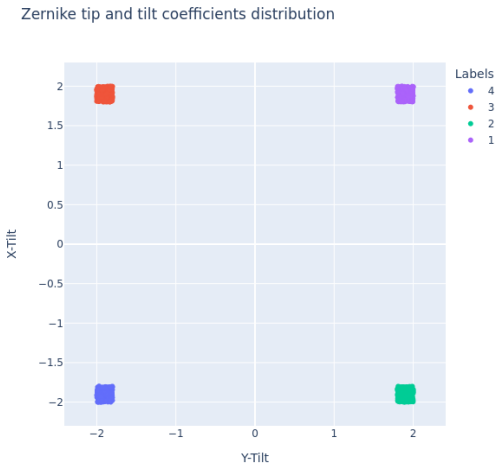
The results are the following:



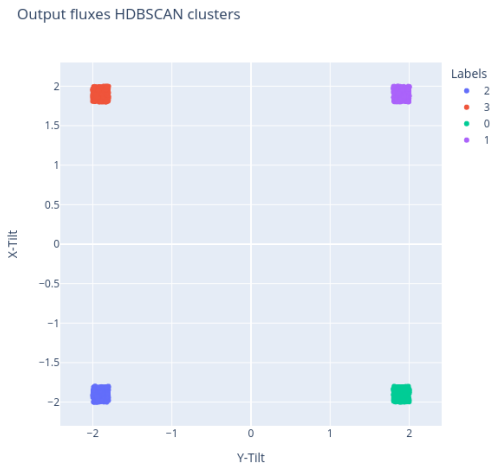

(a) Original cluster densities



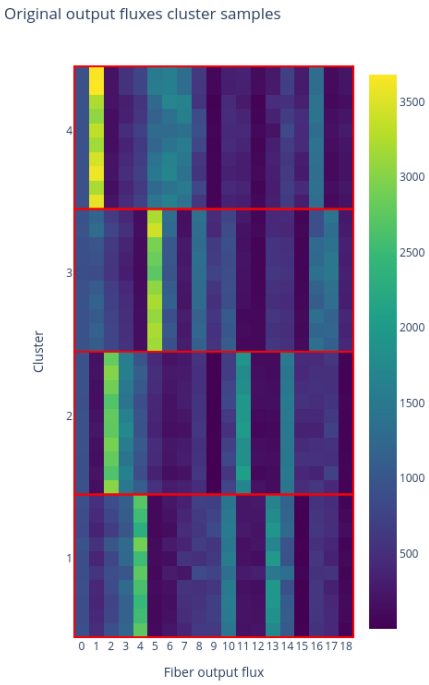

(b) HDBSCAN clusters densities



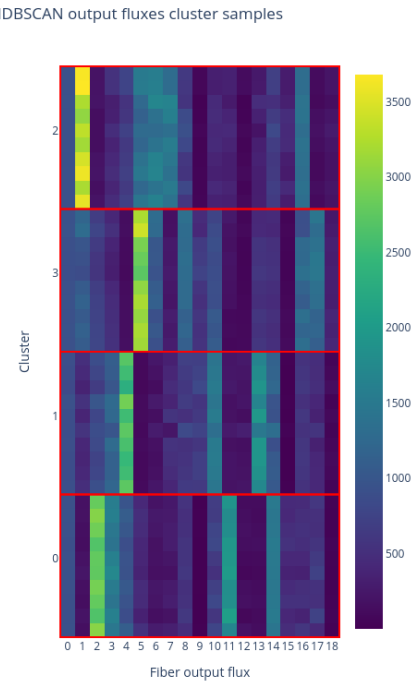

(c) Original clusters




(d) HDBSCAN clusters

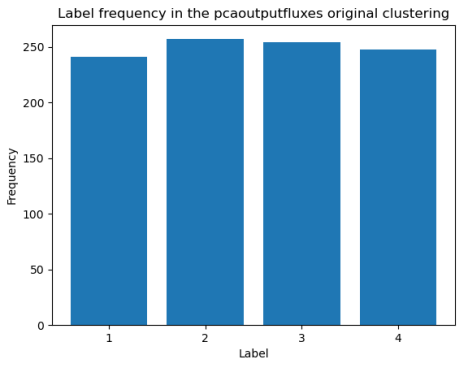



(e) Original cluster samples

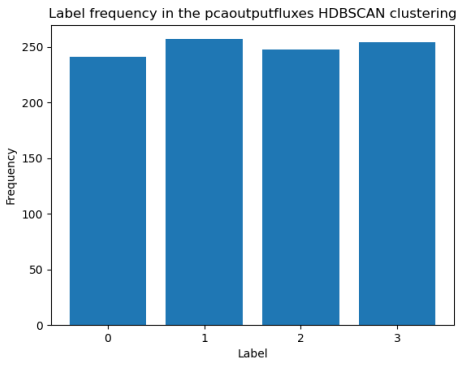



(f) HDBSCAN cluster samples

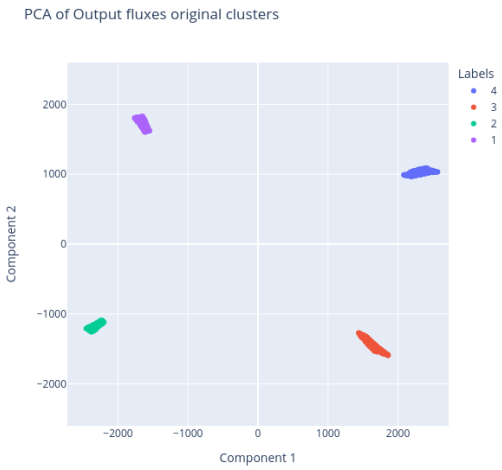
Figure 29: Comparison between original clustering and HDBSCAN clustering



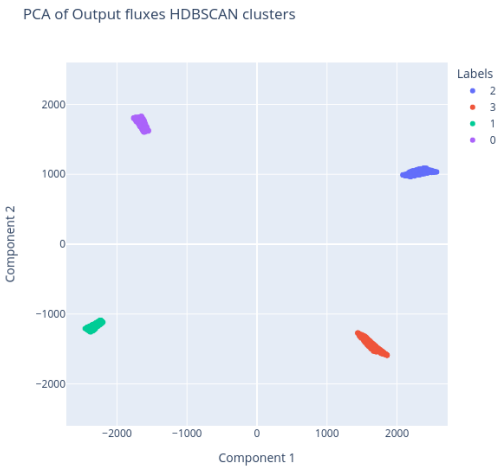

(a) Original cluster densities from PCA



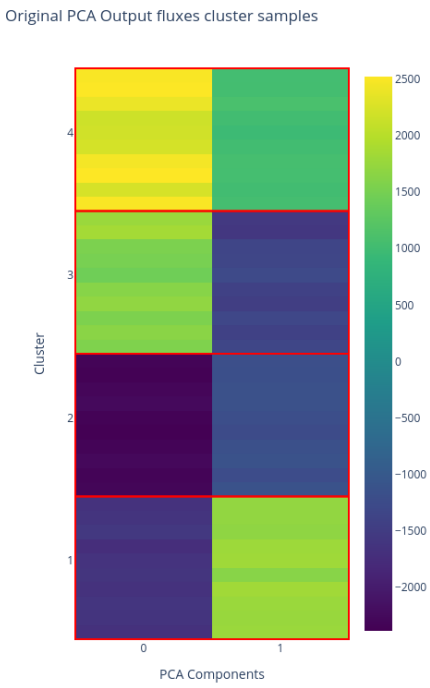

(b) HDBSCAN clusters densities from PCA



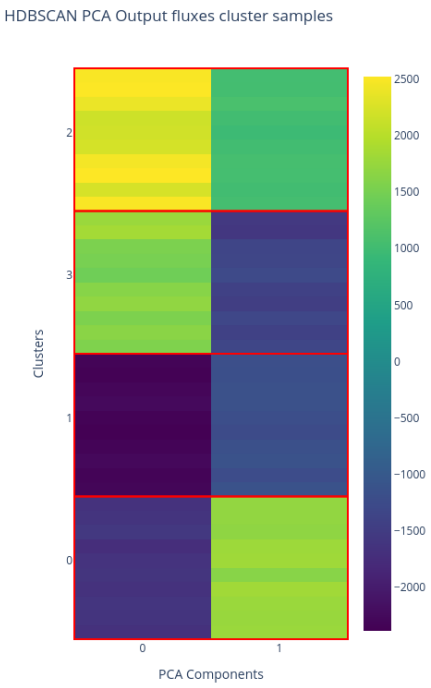

(c) Original clusters from PCA




(d) HDBSCAN clusters from PCA

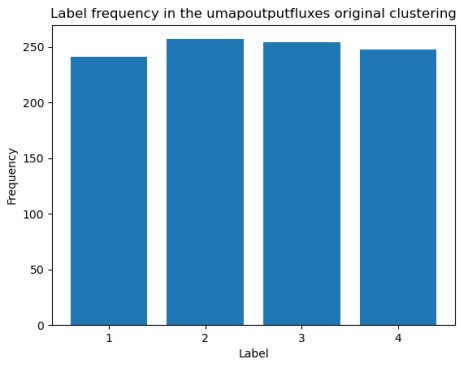



(e) Original cluster samples from PCA

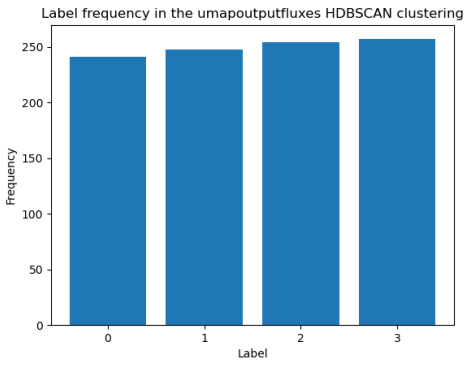



(f) HDBSCAN cluster samples from PCA

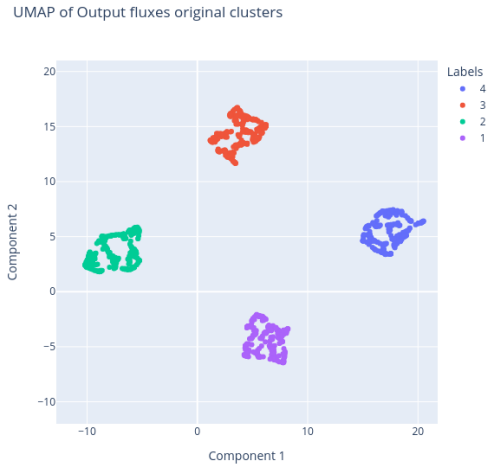
Figure 30: Comparison between original clustering and HDBSCAN clustering from



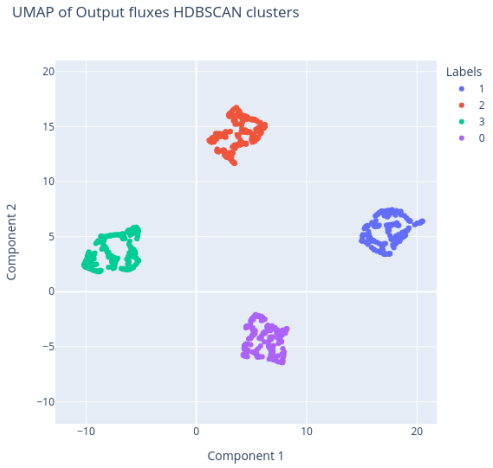

(a) Original cluster densities from UMAP



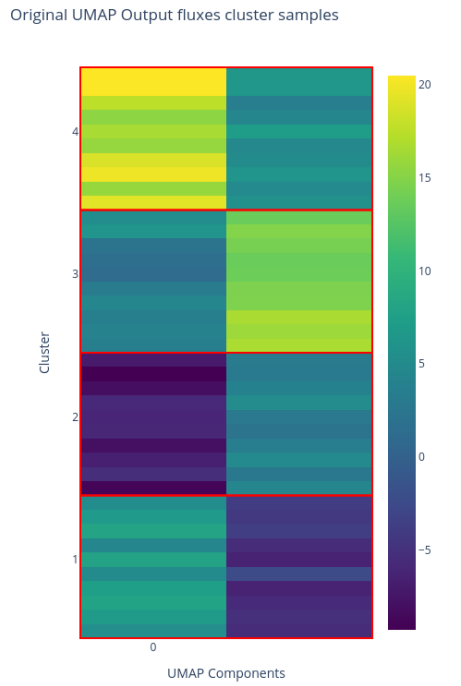

(b) HDBSCAN clusters densities from UMAP



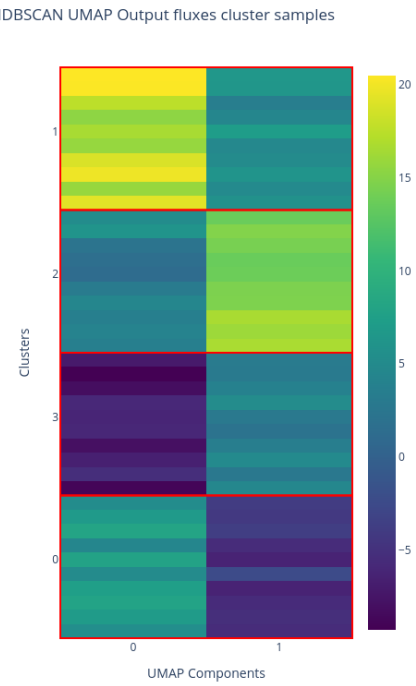

(c) Original clusters from UMAP




(d) HDBSCAN clusters from UMAP




(e) Original cluster samples from UMAP



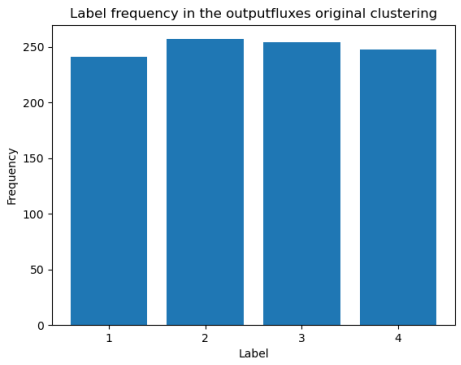

(f) HDBSCAN cluster samples from UMAP

### 4.5.4 Agglomerative clustering

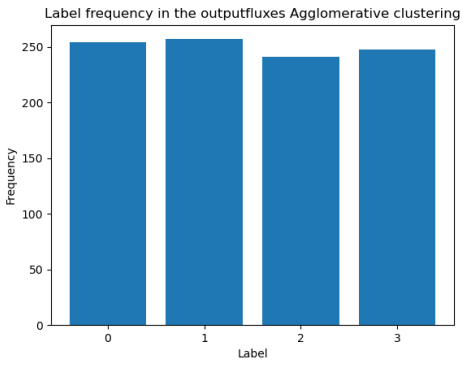
| | **Number of clusters** |
|---|---|
| Original Output fluxes | 4 |
| PCA Output fluxes | 4 |
| UMAP Output fluxes | 4 |

Table 16: Agglomerative hyperparameter configuration for Output fluxes clustering

The results are the following:

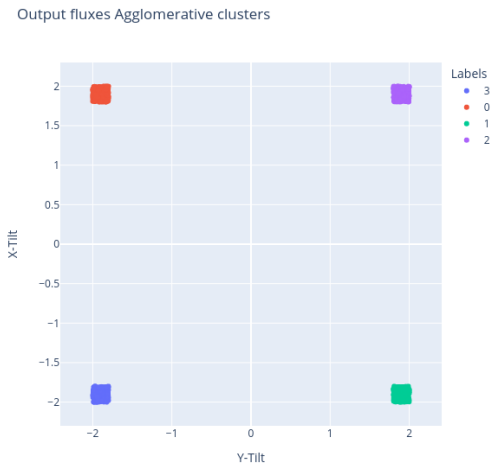



(a) Original cluster densities

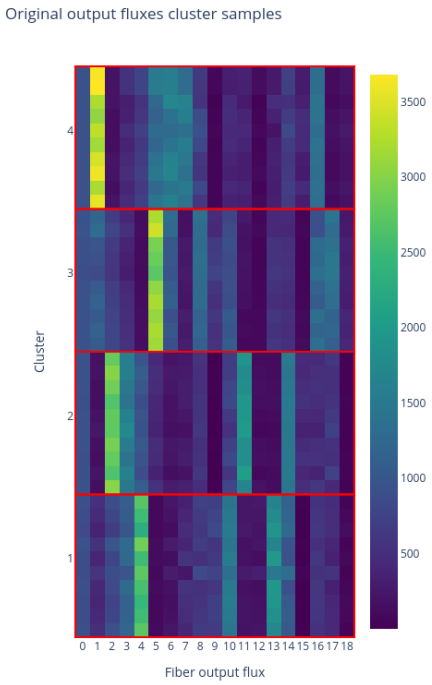



(b) Agglomerative clusters densities

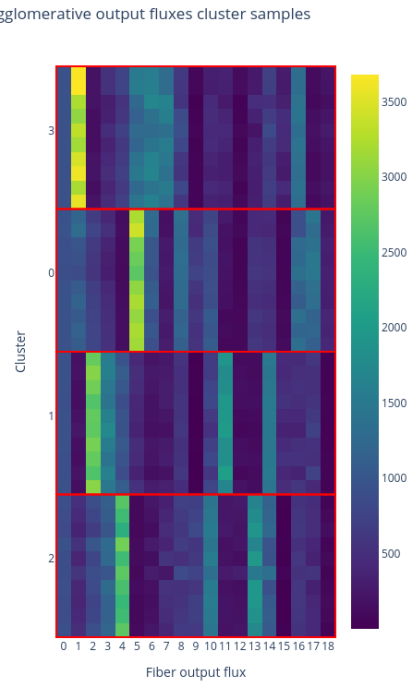

(c) Original clusters




(d) Agglomerative clusters

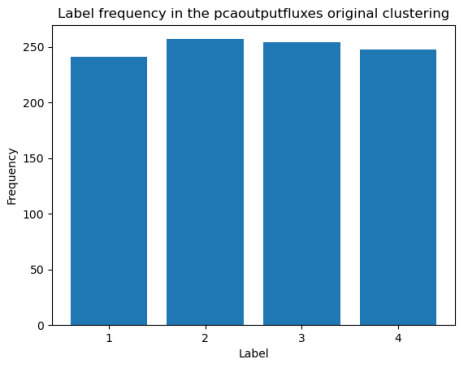



(e) Original cluster samples

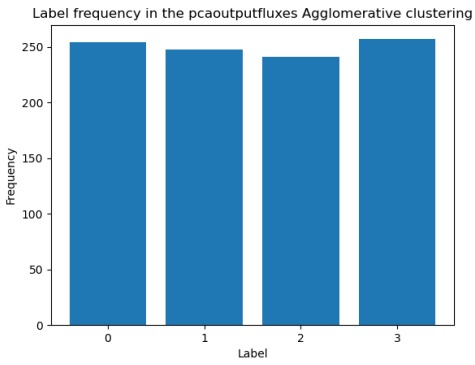



(f) Agglomerative cluster samples

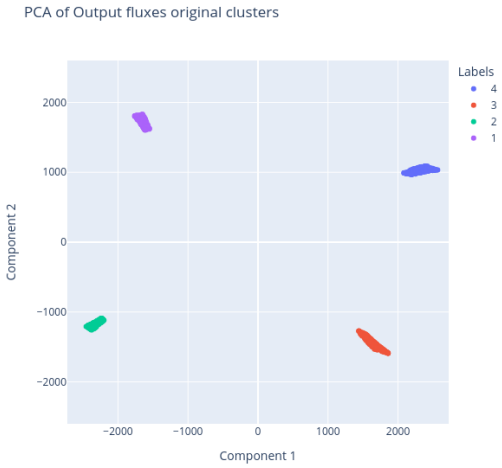
Figure 32: Comparison between original clustering and Agglomerative clustering



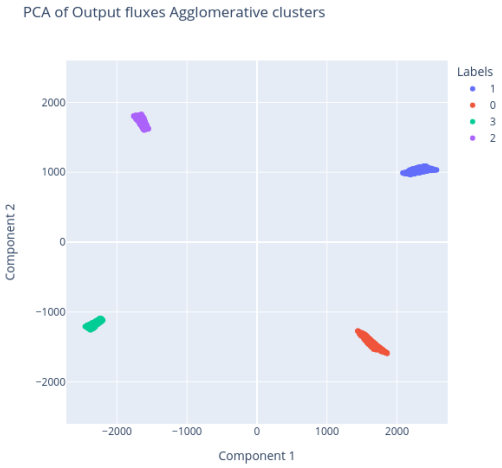

(a) Original cluster densities from PCA



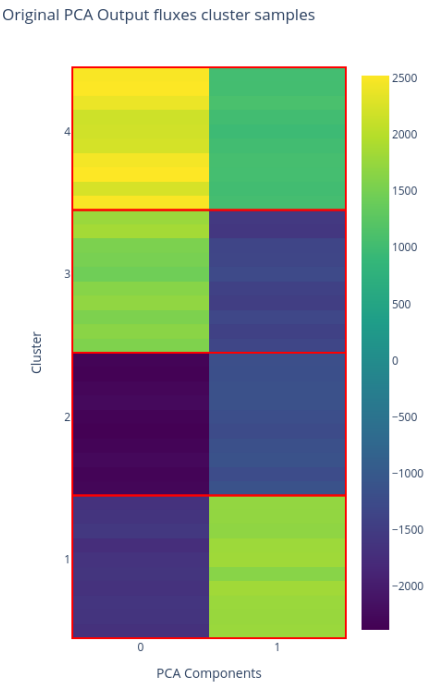

(b) Agglomerative clusters densities from PCA



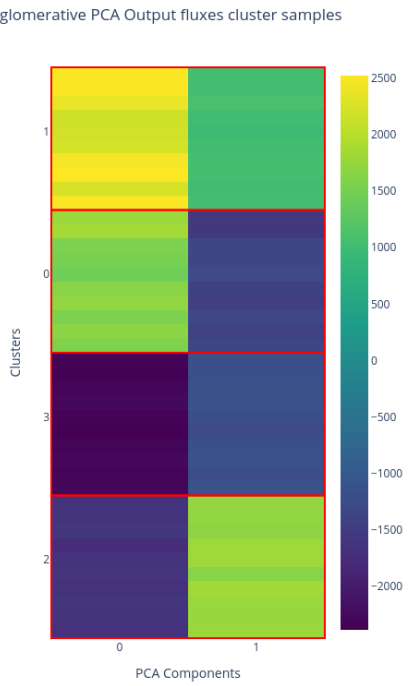

(c) Original clusters from PCA



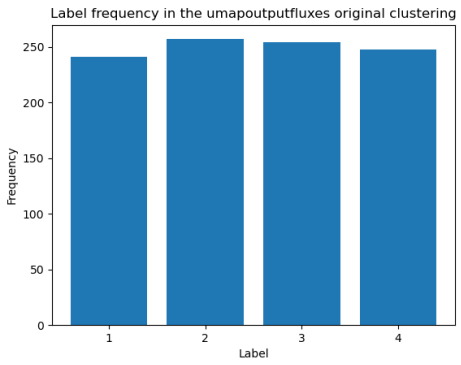

(d) Agglomerative clusters from PCA



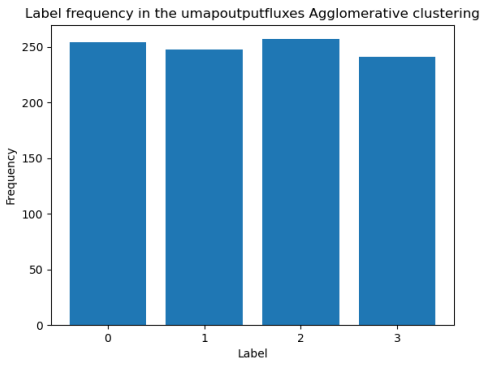

(e) Original cluster samples from PCA



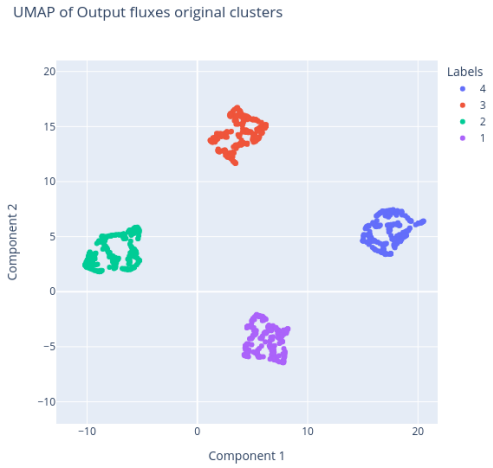

(f) Agglomerative cluster samples from PCA



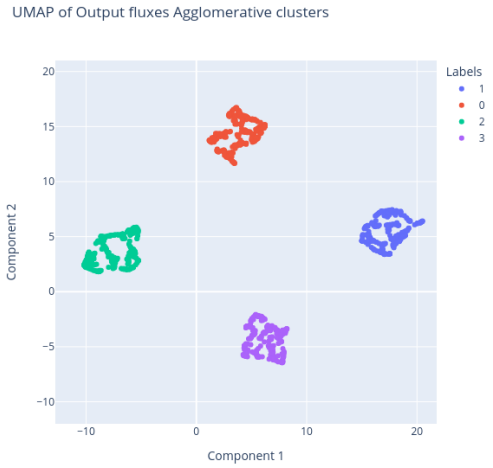

(a) Original cluster densities from UMAP



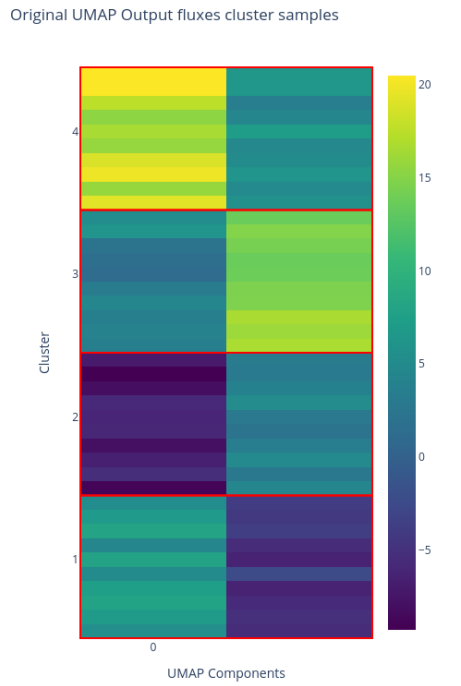

(b) Agglomerative clusters densities from UMAP



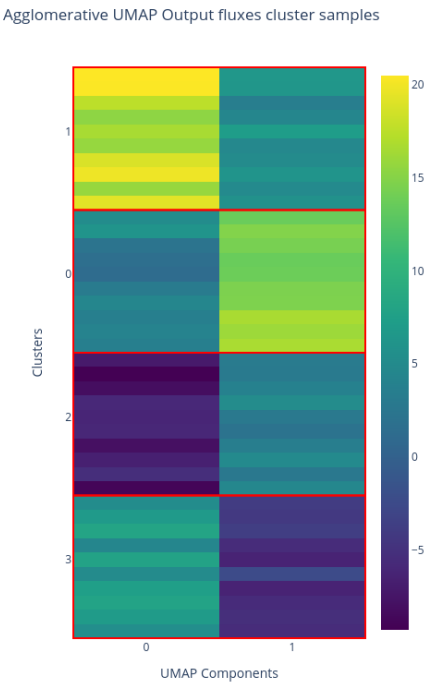

(c) Original clusters from UMAP




(d) Agglomerative clusters from UMAP

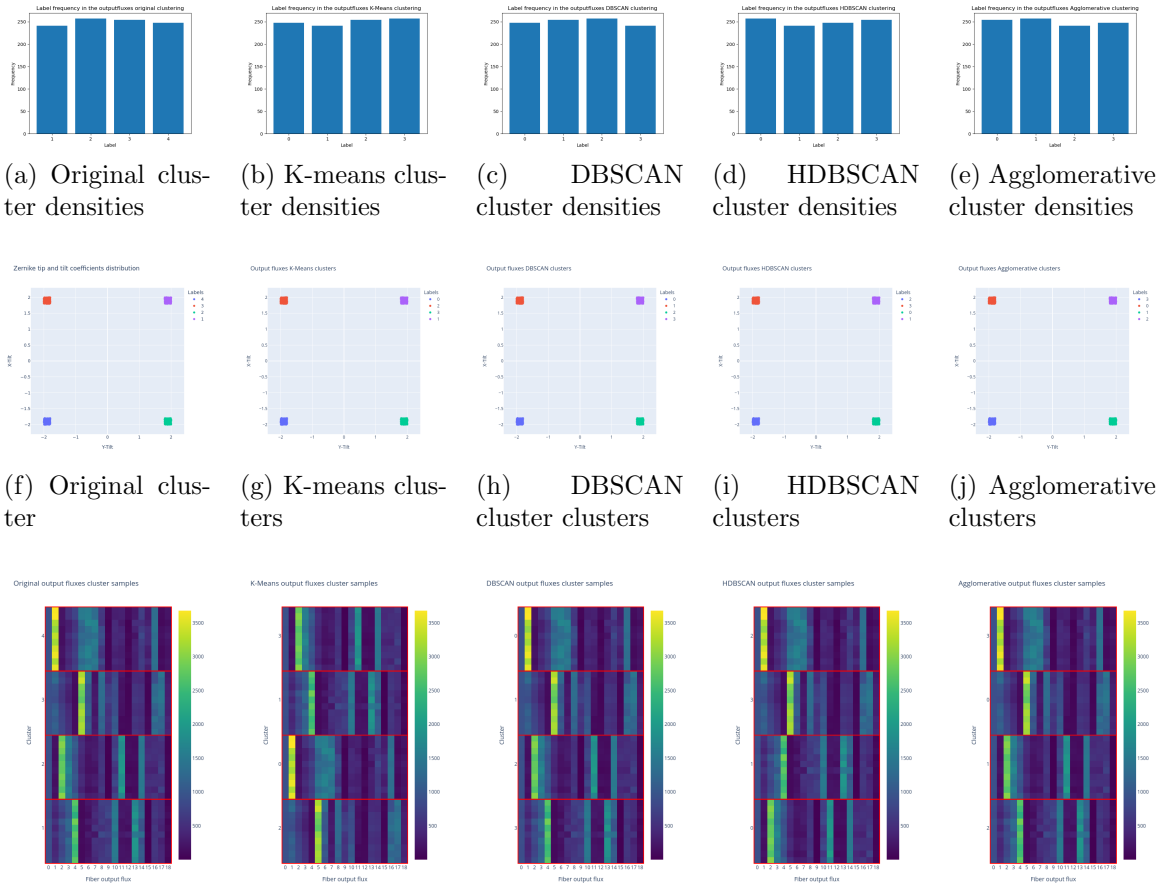



(e) Original cluster samples from UMAP




(f) Agglomerative cluster samples from UMAP

## 4.6 Summary



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

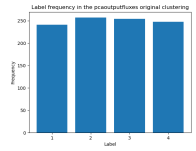
(f) Original cluster (g) K-means clusters (h) DBSCAN cluster clusters (i) HDBSCAN clusters (j) Agglomerative clusters

(k) Original cluster samples (l) K-means cluster samples (m) DBSCAN cluster samples (n) HDBSCAN cluster samples (o) Agglomerative cluster samples

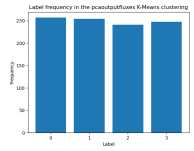
Figure 35: Comparison between clustering Output fluxes algorithms

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

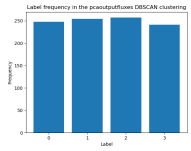
Table 17: Normalized Mutual Information between original Output fluxes clusters



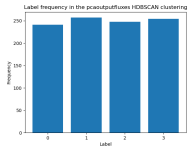
(a) Original cluster densities from PCA



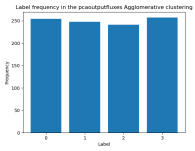
(b) K-means cluster densities from PCA



(c) DBSCAN cluster densities from PCA



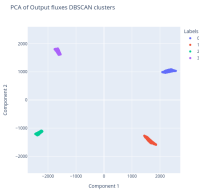
(d) HDBSCAN cluster densities from PCA



(e) Agglomerative cluster densities from PCA

(f) Original cluster from PCA

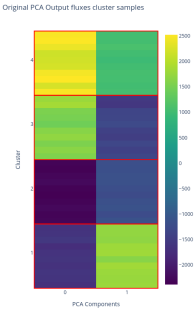
(g) K-means clusters from PCA



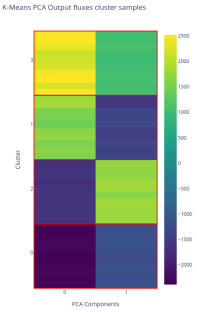
(h) DBSCAN cluster clusters from PCA

(i) HDBSCAN clusters from PCA

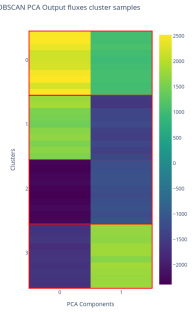
(j) Agglomerative clusters from PCA



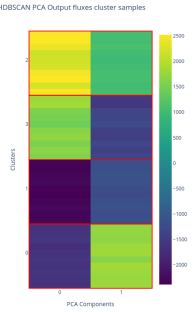
(k) Original cluster samples from PCA



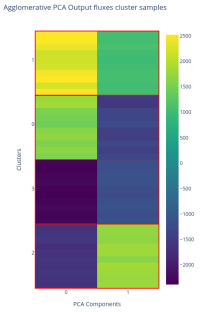
(l) K-means cluster samples from PCA



(m) DBSCAN cluster samples-from PCA



(n) HDBSCAN cluster samples from PCA

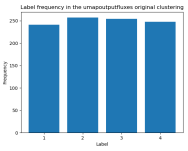


(o) Agglomerative cluster samples from PCA

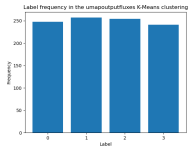
Figure 36: Comparison between clustering PCA Output fluxes algorithms

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

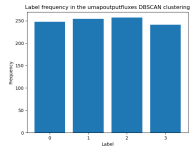
Table 18: Normalized Mutual Information between PCA Output fluxes clusters



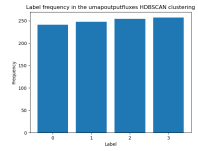
(a) Original cluster densities from UMAP



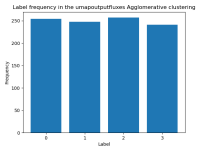
(b) K-means cluster densities from UMAP



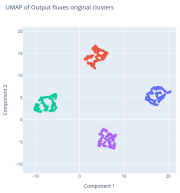
(c) DBSCAN cluster densities from UMAP



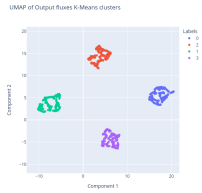
(d) HDBSCAN cluster densities from UMAP



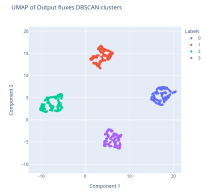
(e) Agglomerative cluster densities from UMAP



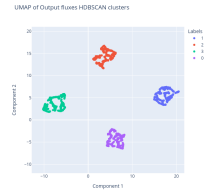
(f) Original cluster from UMAP



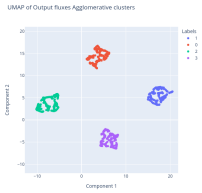
(g) K-means clusters from UMAP



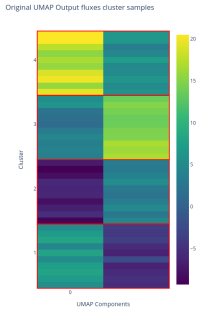
(h) DBSCAN cluster clusters from UMAP



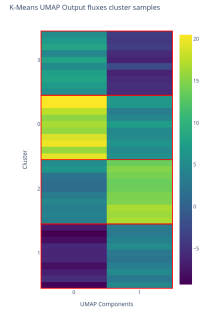
(i) HDBSCAN clusters from UMAP



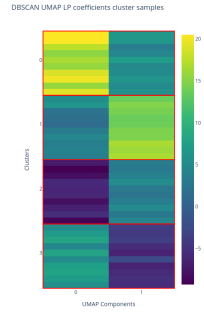
(j) Agglomerative clusters from UMAP



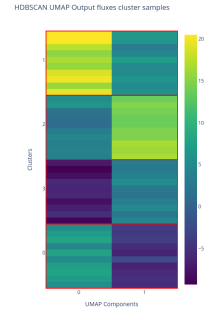
(k) Original cluster samples from UMAP



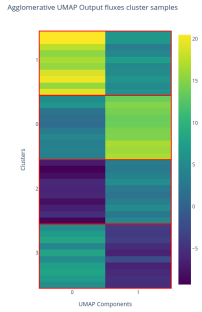
(l) K-means cluster samples from UMAP



(m) DBSCAN cluster samples-from UMAP



(n) HDBSCAN cluster samples from UMAP



(o) Agglomerative cluster samples from UMAP

Figure 37: Comparison between clustering UMAP Output fluxes algorithms

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

Table 19: Normalized Mutual Information between UMAP Output fluxes clusters

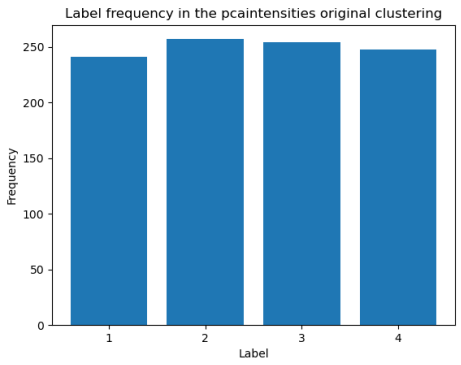
## 4.7 PSF Intensities clustering

### 4.7.1 K-Means

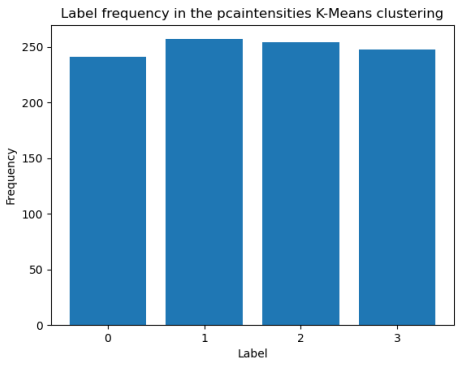
As K-Means allows for the number of clusters to be defined, and we know that there are 4 in the original dataset, K-Means is used to find 4 clusters.

| | **Number of clusters** | **Number of initializations** |
|---|---|---|
| **PCA PSF Intensities** | 4 | 10 |
| **UMAP PSF Intensities** | 4 | 10 |

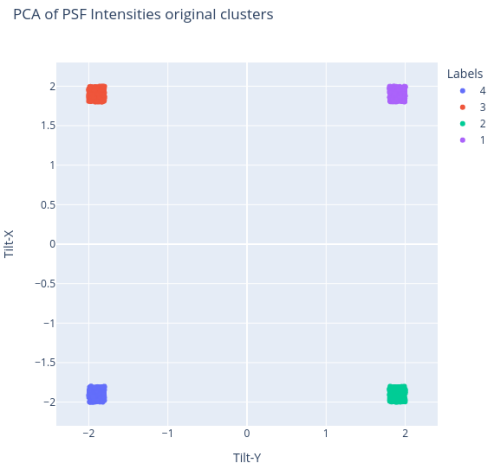
Table 20: K-Means hyperparameter configuration for c coefficients clustering




(a) Original cluster densities from PCA

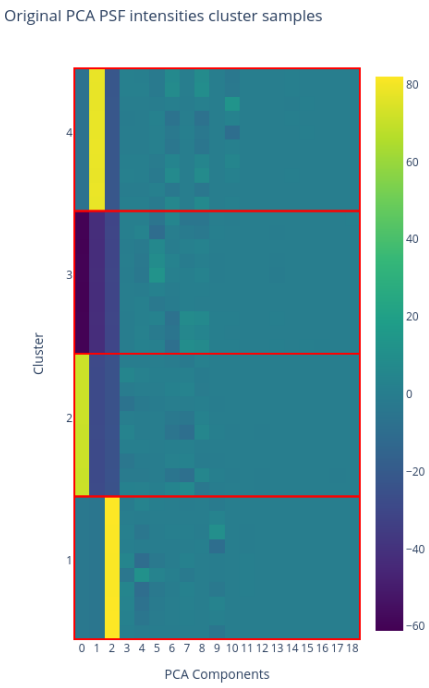



(b) K-Means clusters densities from PCA

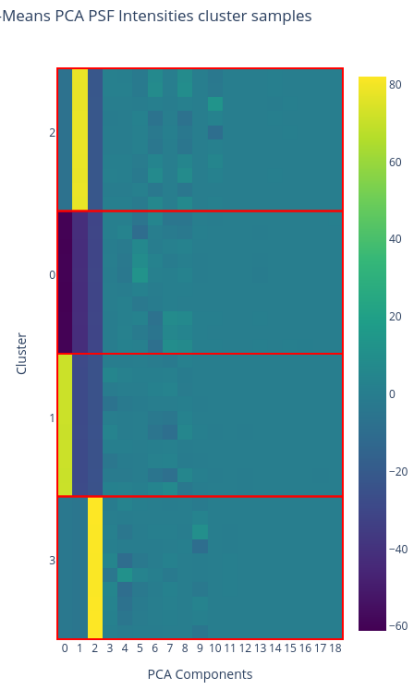



(c) Original clusters


(d) K-Means clusters from PCA

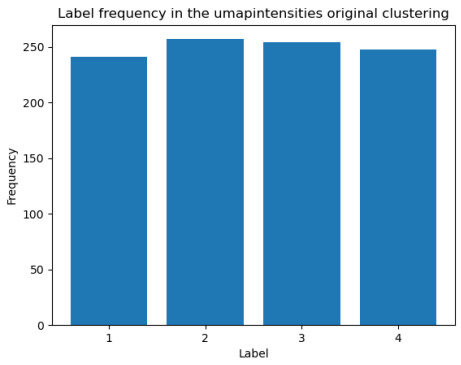



(e) Original cluster samples from PCA

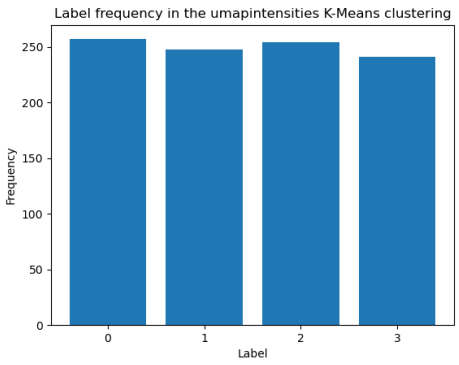



(f) K-Means cluster samples from PCA

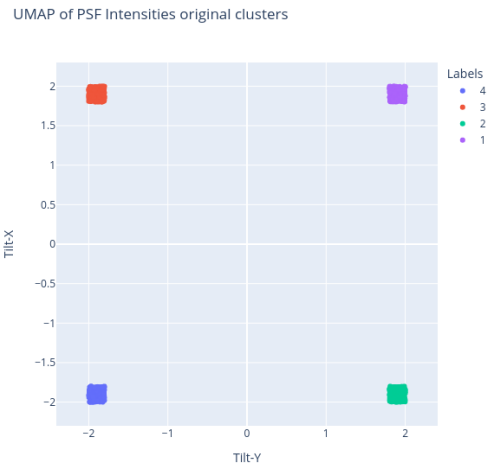
Figure 38: Comparison between original clustering and K-Means clustering from PCA of PSF Intensities



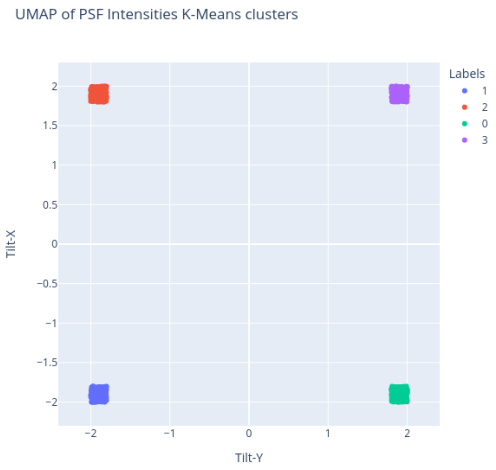

(a) Original cluster densities from UMAP



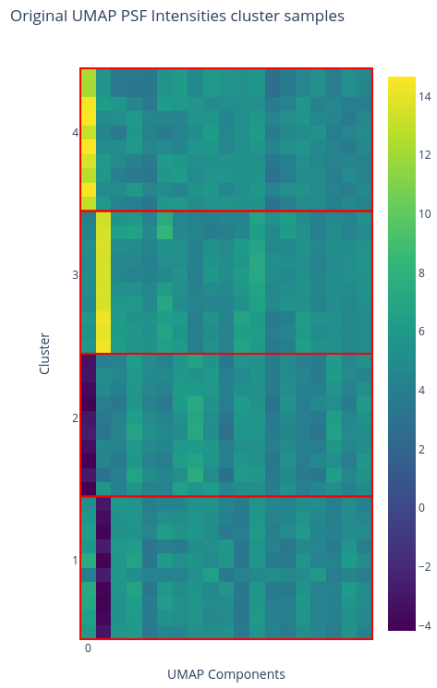

(b) K-Means clusters densities from UMAP



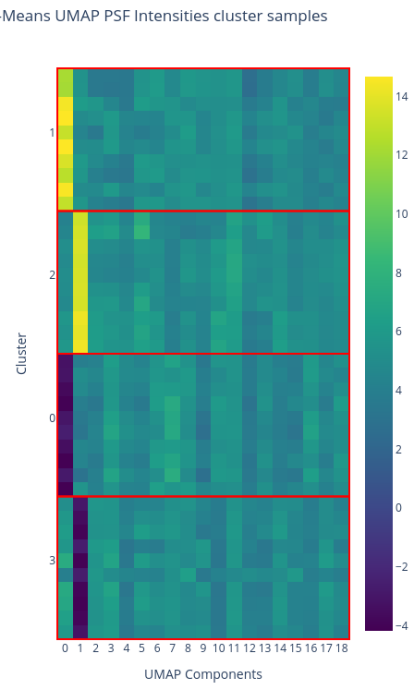

(c) Original clusters from UMAP




(d) K-Means clusters from UMAP




(e) Original cluster samples from UMAP




(f) K-Means cluster samples from UMAP

Figure 39: Comparison between original clustering and K-Means clustering from

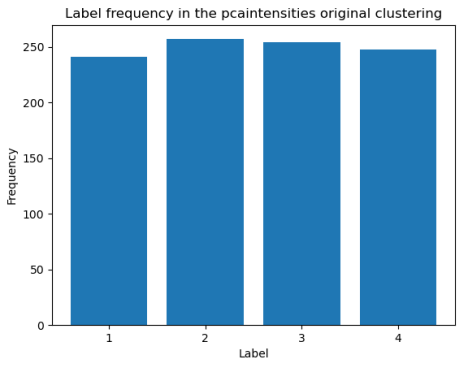
### 4.7.2 DBSCAN

A configuration that outputs 4 clusters is searched

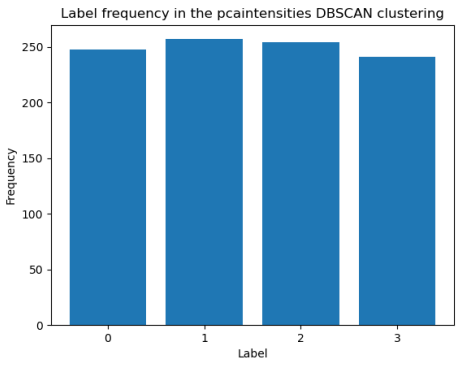
| | **Number of neighbours** | **Epsilon** |
|---|---|---|
| PCA PSF Intensities | 15 | 4.5 |
| UMAP PSF Intensities | 10 | 0.85 |

Table 21: DBSCAN hyperparameter configuration for PSF Intensities clustering

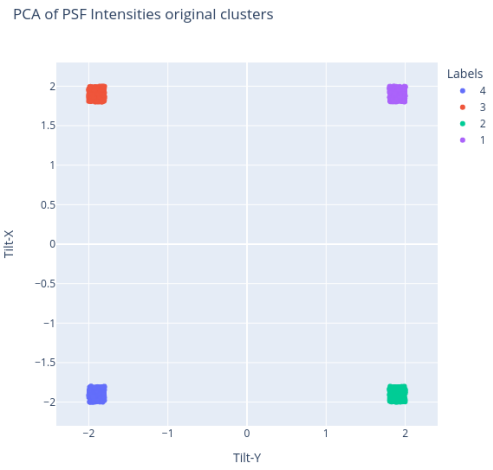
The results are the following:



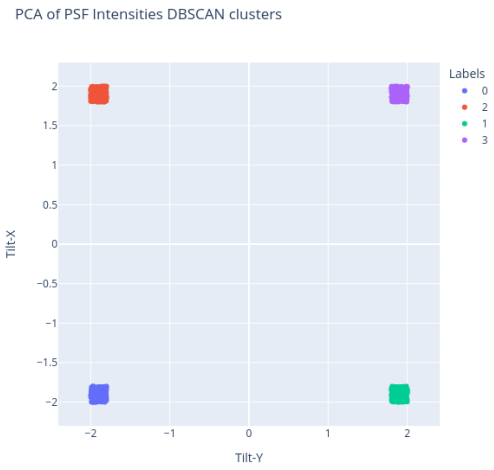

(a) Original cluster densities from PCA



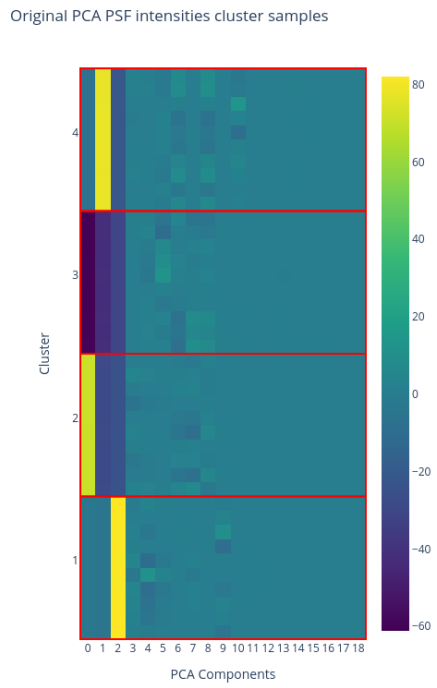

(b) DBSCAN clusters densities from PCA



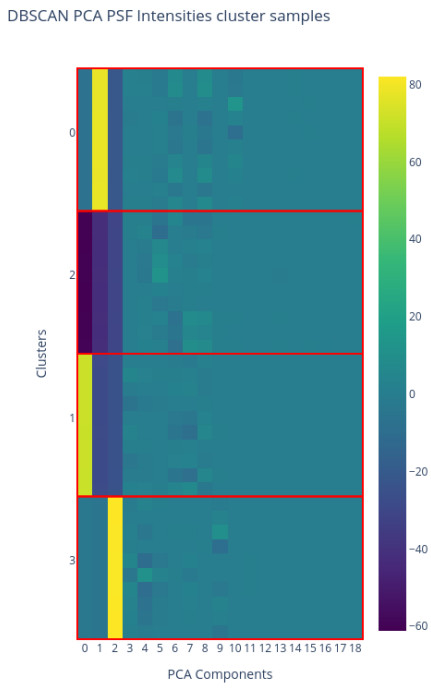

(c) Original clusters from PCA




(d) DBSCAN clusters from PCA

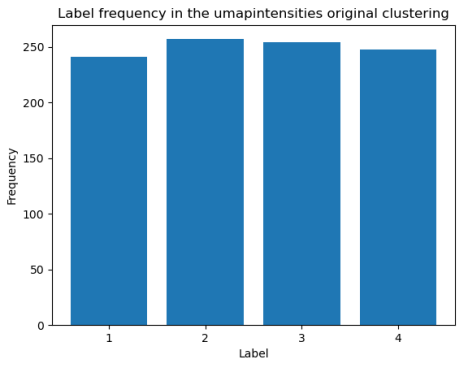



(e) Original cluster samples from PCA

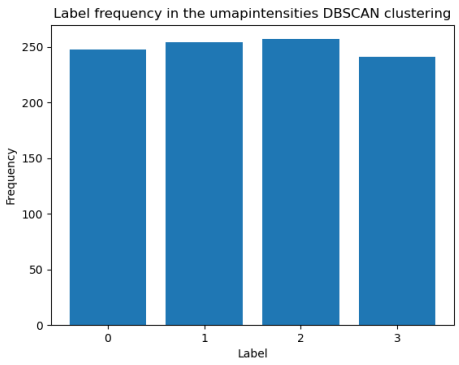



(f) DBSCAN cluster samples from PCA

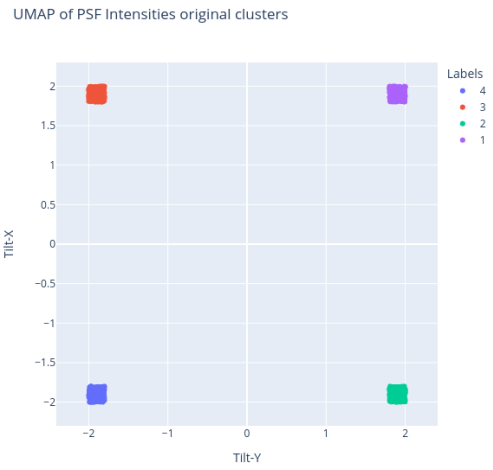
Figure 40: Comparison between original clustering and DBSCAN clustering from PCA of PSF Intensities



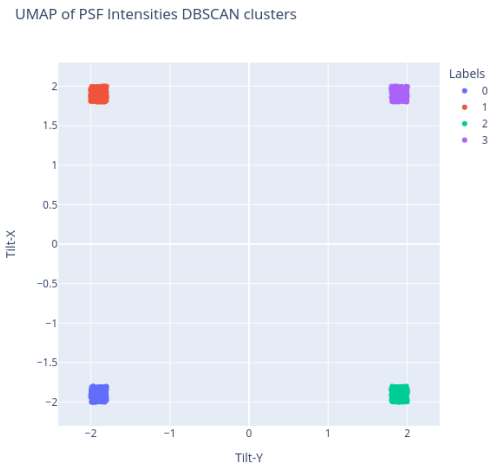

(a) Original cluster densities from UMAP



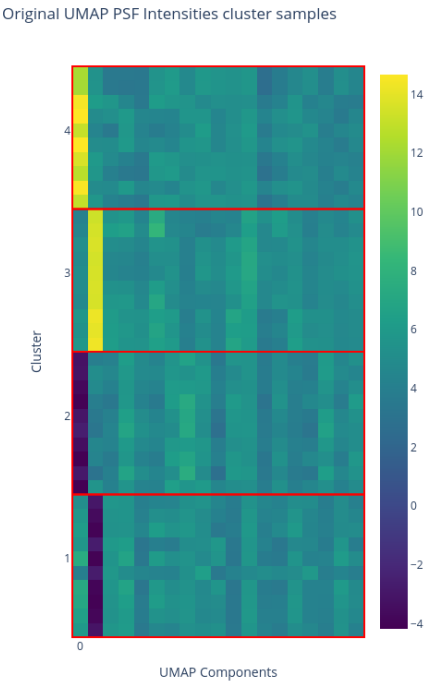

(b) DBSCAN clusters densities from UMAP



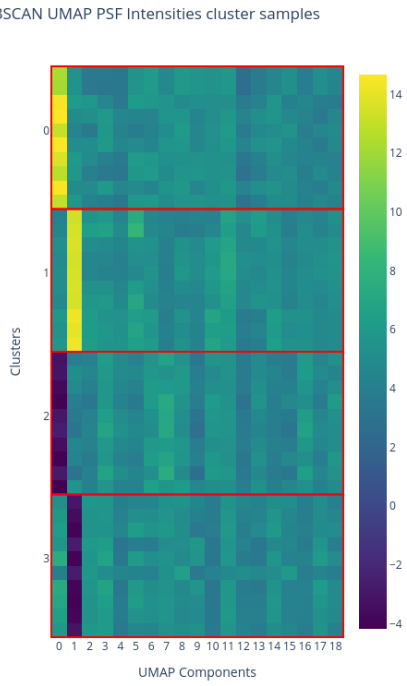

(c) Original clusters from UMAP




(d) DBSCAN clusters from UMAP




(e) Original cluster samples from UMAP




(f) DBSCAN cluster samples from UMAP

Figure 41: Comparison between original clustering and DBSCAN clustering from

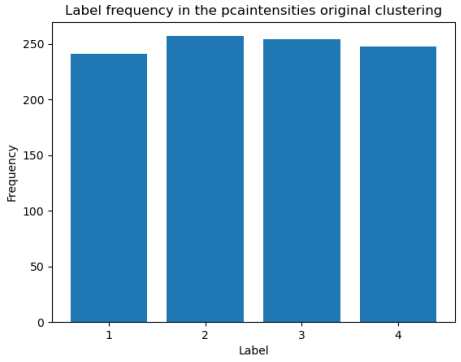
### 4.7.3 HDBSCAN

A configuration that outputs 4 clusters is searched.

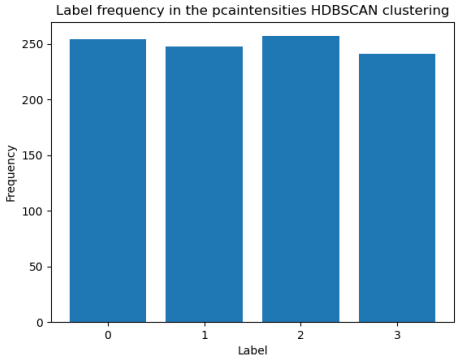
| | **Minimum cluster size** |
|---|---|
| PCA PSF Intensities | 21 |
| UMAP PSF Intensities | 25 |

Table 22: HDBSCAN hyperparameter configuration for PSF Intensities clustering

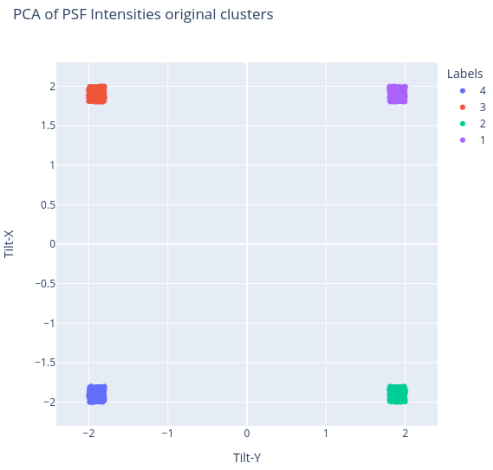
The results are the following:



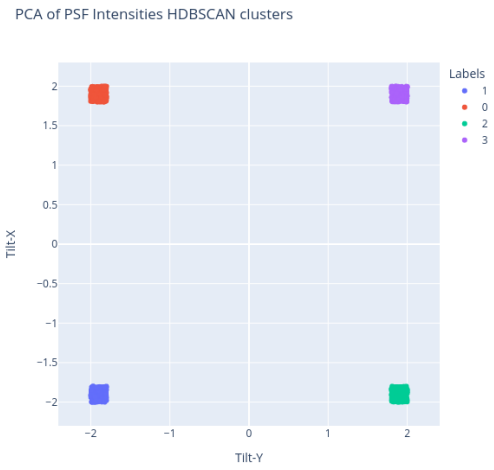

(a) Original cluster densities from PCA



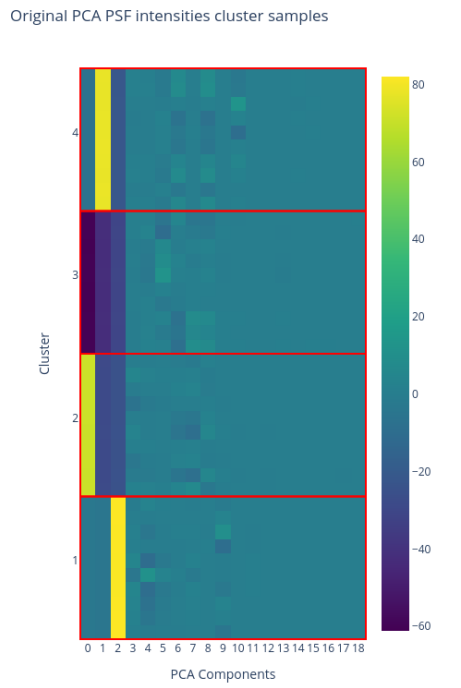

(b) HDBSCAN clusters densities from PCA



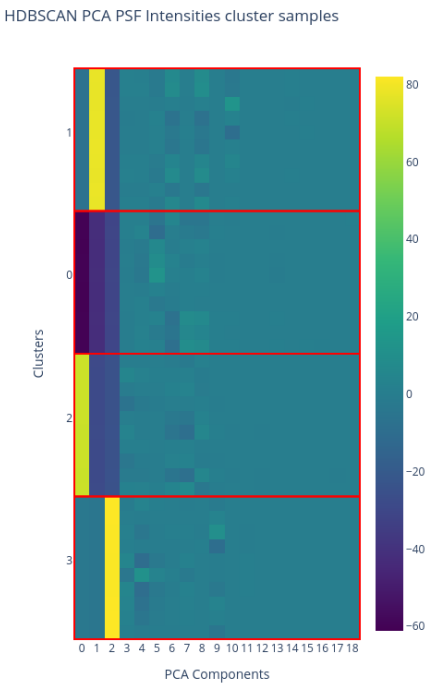

(c) Original clusters from PCA




(d) HDBSCAN clusters from PCA

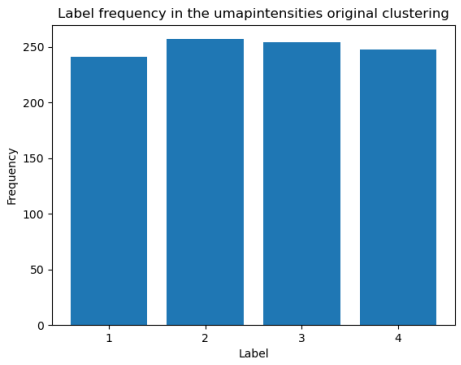



(e) Original cluster samples from PCA

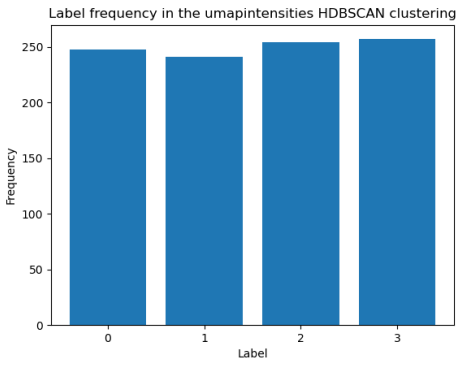



(f) HDBSCAN cluster samples from PCA

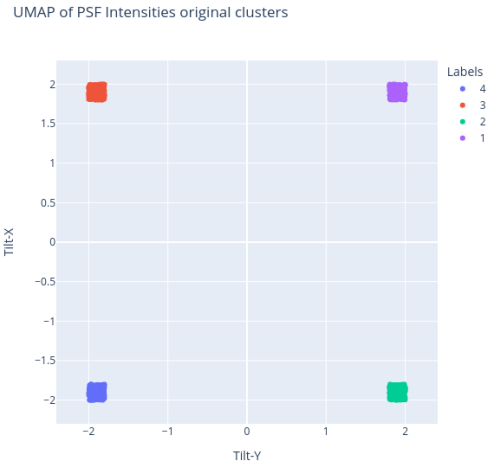
Figure 42: Comparison between original clustering and HDBSCAN clustering from



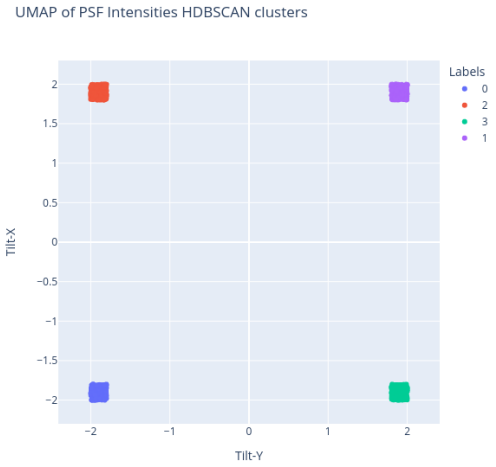

(a) Original cluster densities from UMAP



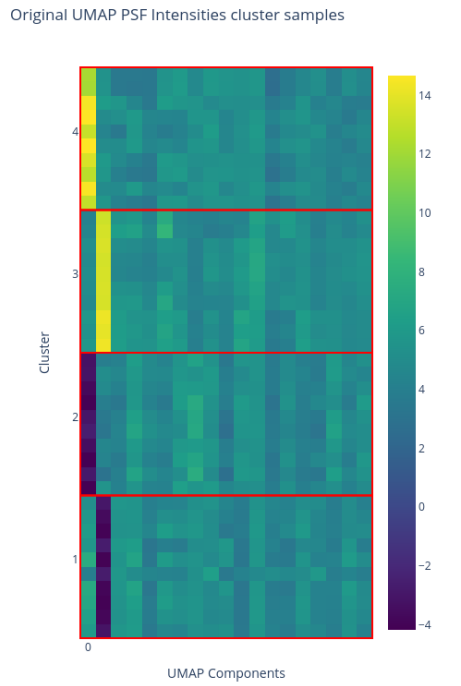

(b) HDBSCAN clusters densities from UMAP



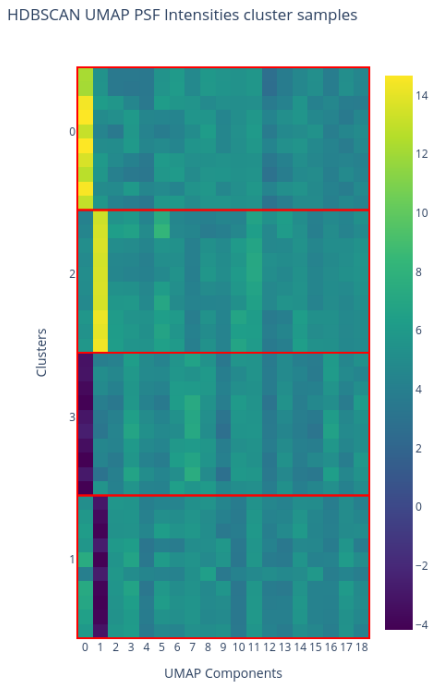

(c) Original clusters from UMAP




(d) HDBSCAN clusters from UMAP




(e) Original cluster samples from UMAP



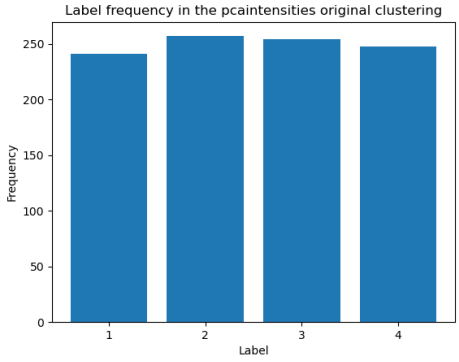

(f) HDBSCAN cluster samples from UMAP

### 4.7.4 Agglomerative clustering

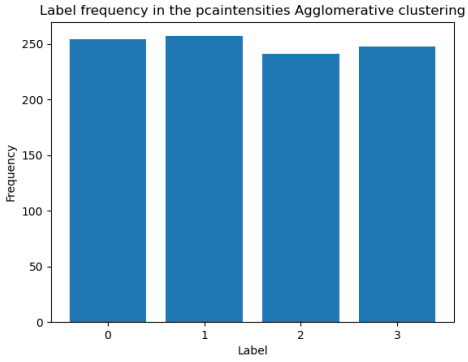
| | **Number of clusters** |
|---|---|
| PCA PSF Intensities | 4 |
| UMAP PSF Intensities | 4 |

Table 23: Agglomerative hyperparameter configuration for PSF Intensities clustering

The results are the following:



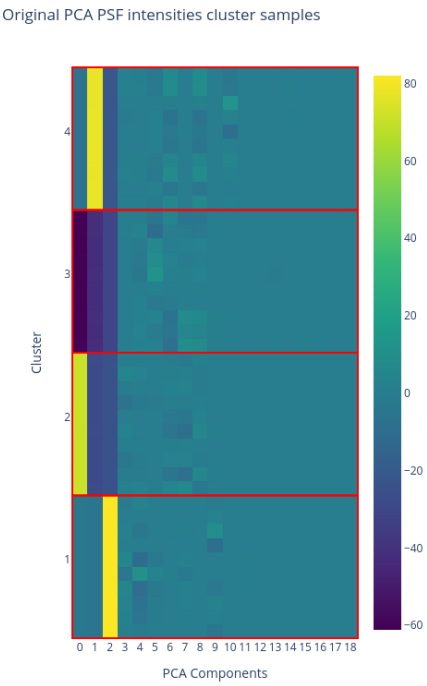

(a) Original cluster densities from PCA



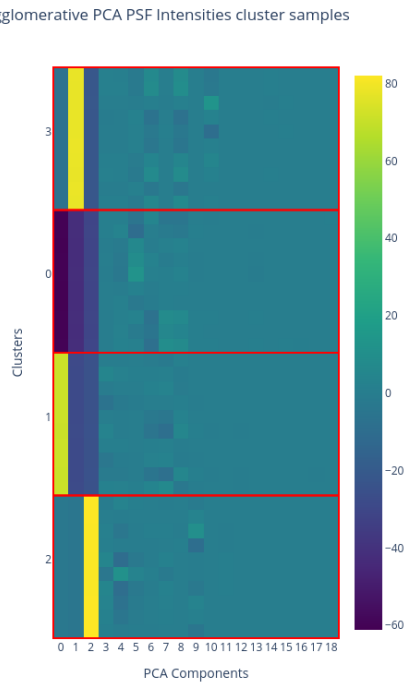

(b) Agglomerative clusters densities from PCA


(c) Original clusters from PCA

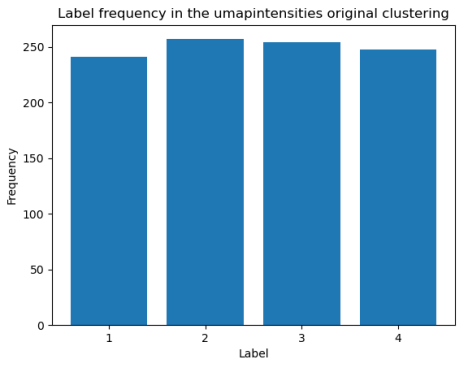

(d) Agglomerative clusters from PCA



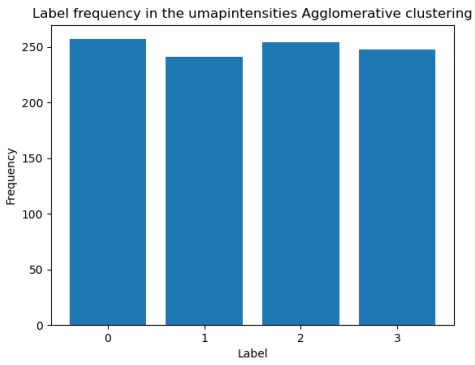

(e) Original cluster samples from PCA



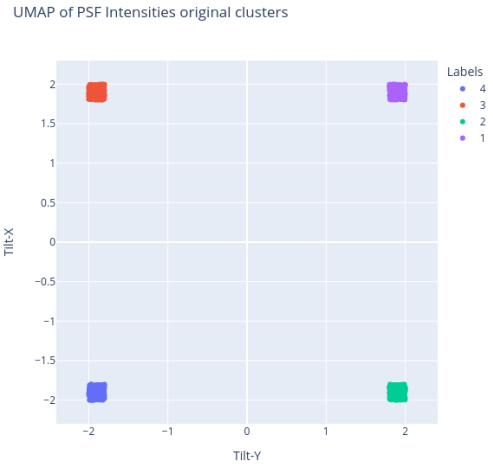

(f) Agglomerative cluster samples from PCA



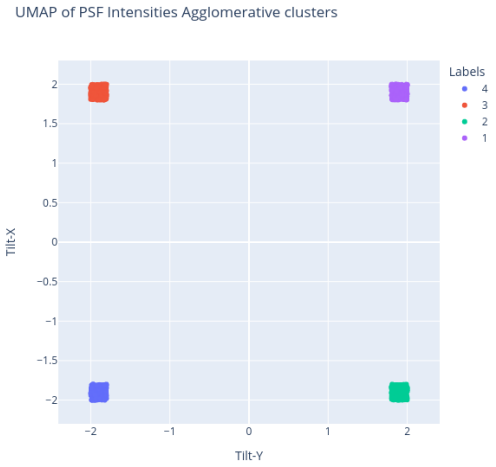

(a) Original cluster densities from UMAP



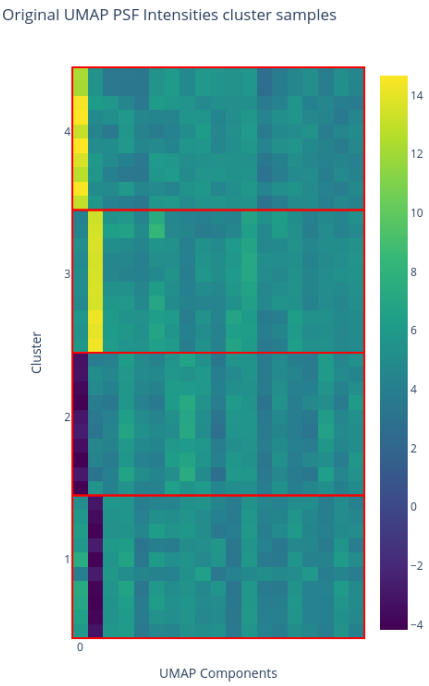

(b) Agglomerative clusters densities from UMAP



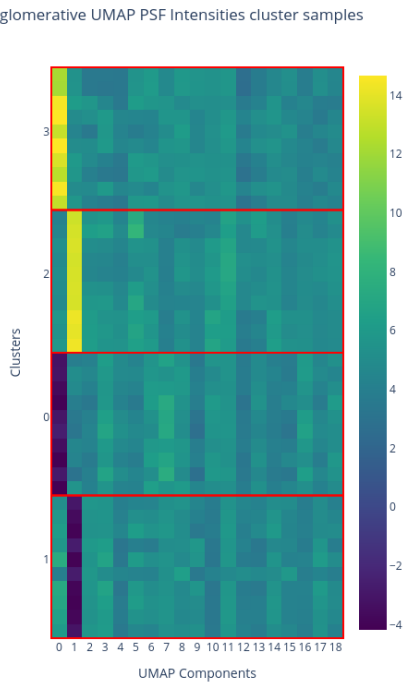

(c) Original clusters from UMAP




(d) Agglomerative clusters from UMAP

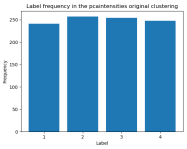



(e) Original cluster samples from UMAP

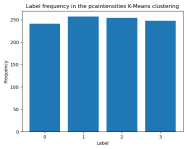



(f) Agglomerative cluster samples from UMAP

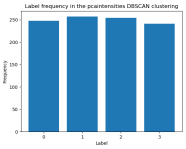
## 4.8 Summary



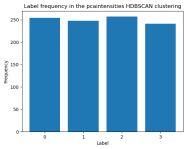

(a) Original cluster densities from PCA



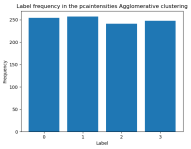

(b) K-means cluster densities from PCA



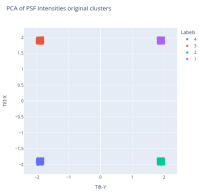

(c) DBSCAN cluster densities from PCA




(d) HDBSCAN cluster densities from PCA

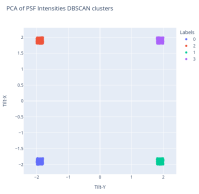



(e) Agglomerative cluster densities from PCA




(f) Original cluster from PCA

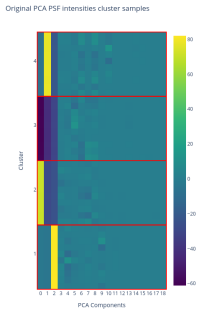

(g) K-means clusters from PCA



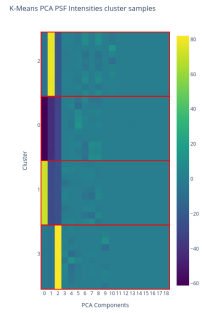

(h) DBSCAN cluster clusters from PCA


(i) HDBSCAN clusters from PCA

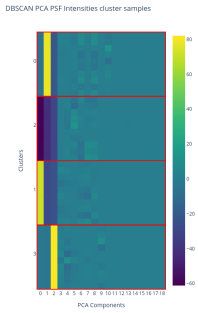

(j) Agglomerative clusters from PCA



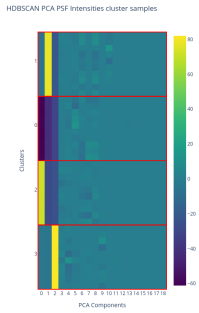

(k) Original cluster samples from PCA



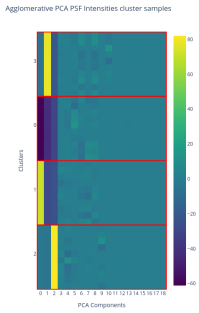

(l) K-means cluster samples from PCA




(m) DBSCAN cluster samples-from PCA




(n) HDBSCAN cluster samples from PCA

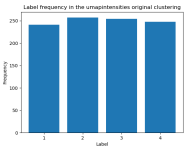



(o) Agglomerative cluster samples from PCA

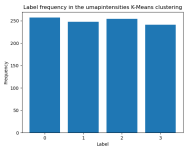
Figure 46: Comparison between clustering PCA PSF Intensities algorithms

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

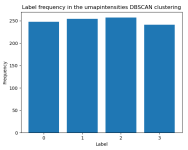
Table 24: Normalized Mutual Information between PCA PSF Intensities clusters



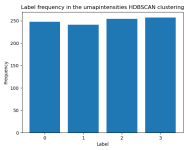

(a) Original cluster densities from UMAP



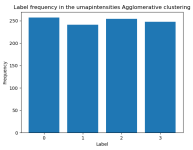

(b) K-means cluster densities from UMAP




(c) DBSCAN cluster densities from UMAP

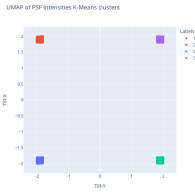



(d) HDBSCAN cluster densities from UMAP

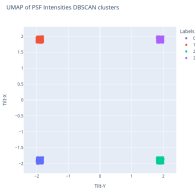



(e) Agglomerative cluster densities from UMAP

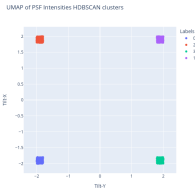

(f) Original cluster from UMAP



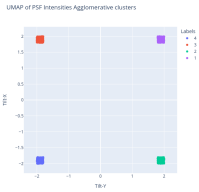

(g) K-means clusters from UMAP



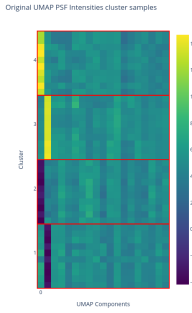

(h) DBSCAN cluster clusters from UMAP



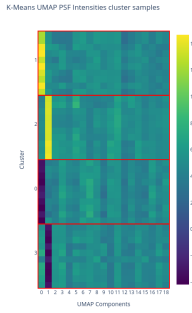

(i) HDBSCAN clusters from UMAP



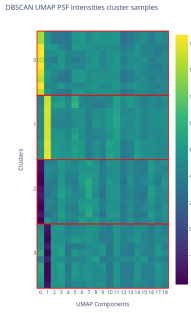

(j) Agglomerative clusters from UMAP



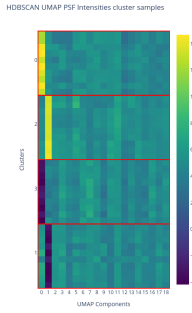

(k) Original cluster samples from UMAP



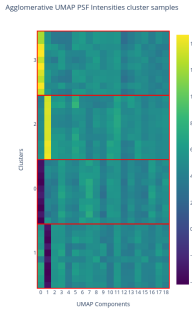

(l) K-means cluster samples from UMAP




(m) DBSCAN cluster samples-from UMAP




(n) HDBSCAN cluster samples from UMAP



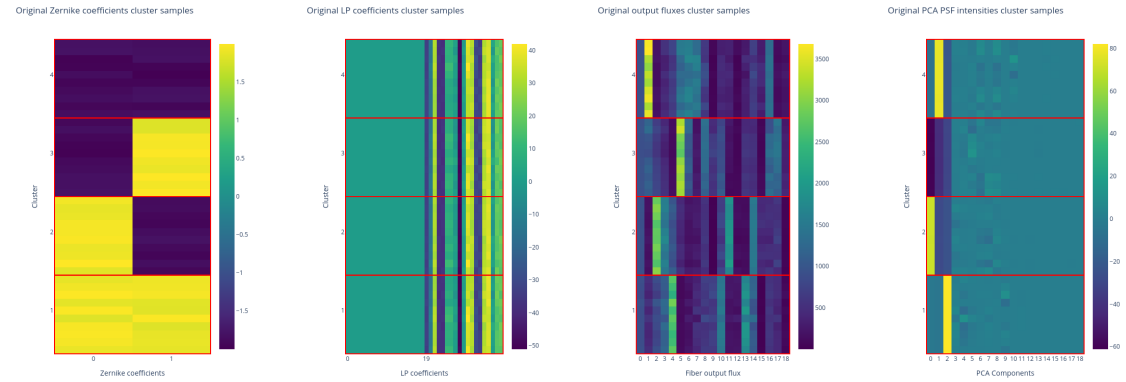

(o) Agglomerative cluster samples from UMAP

Figure 47: Comparison between clustering UMAP PSF Intensities algorithms

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

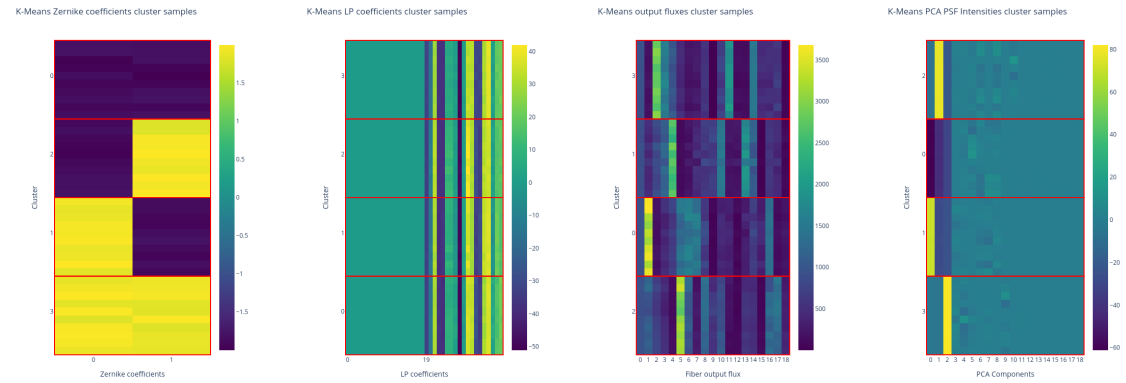
Table 25: Normalized Mutual Information between UMAP PSF Intensities clusters

# 5 Dataset clusters comparison



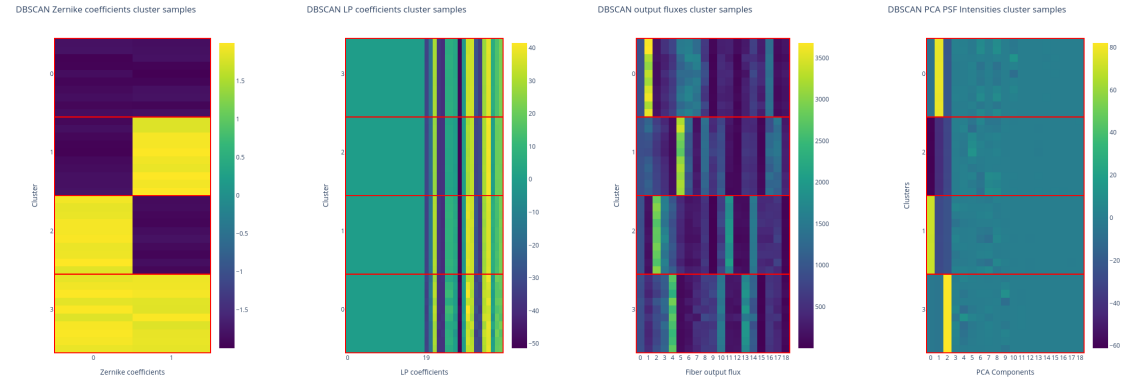

(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) Intensities

Figure 48: Original clusters from the datasets



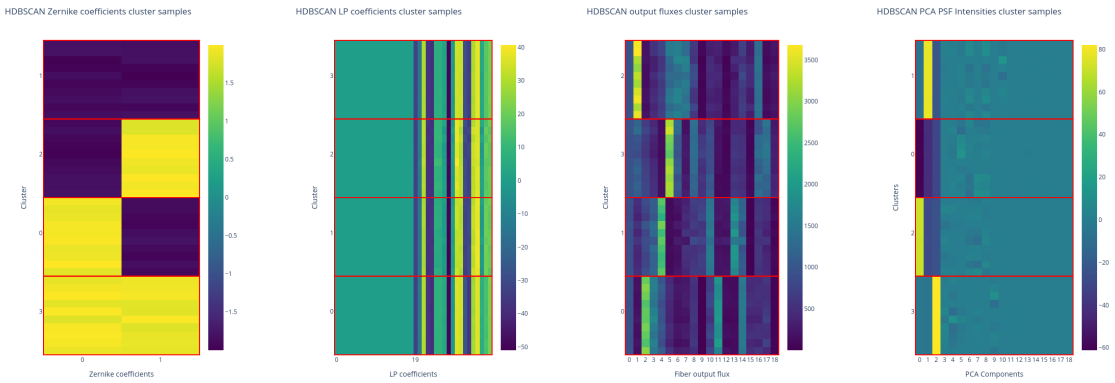

(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) Intensities

Figure 49: K-Means clusters from the datasets



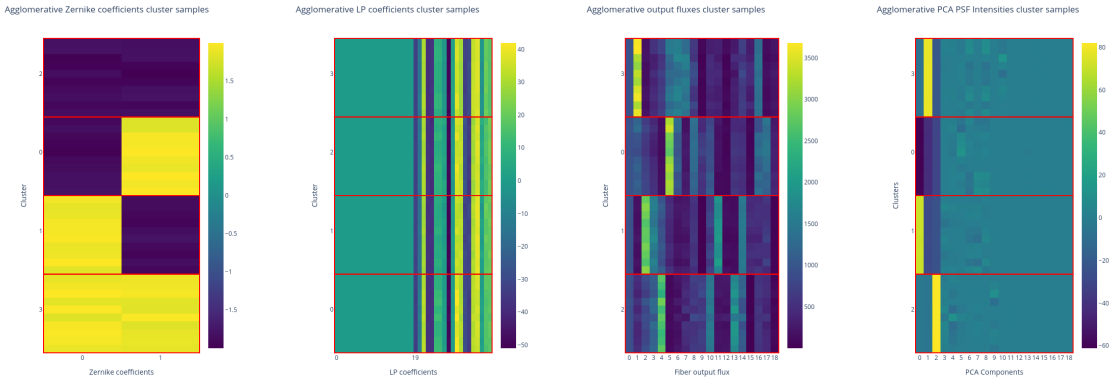

(a) Zernike coefficients (b) LP coefficients (c) Output fluxes (d) Intensities

Figure 50: DBSCAN clusters from the datasets




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

Figure 51: HDBSCAN clusters from the datasets




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

Figure 52: Agglomerative clusters from the datasets