Anomaly Detection and Clustering – Example Output

Max States Tests

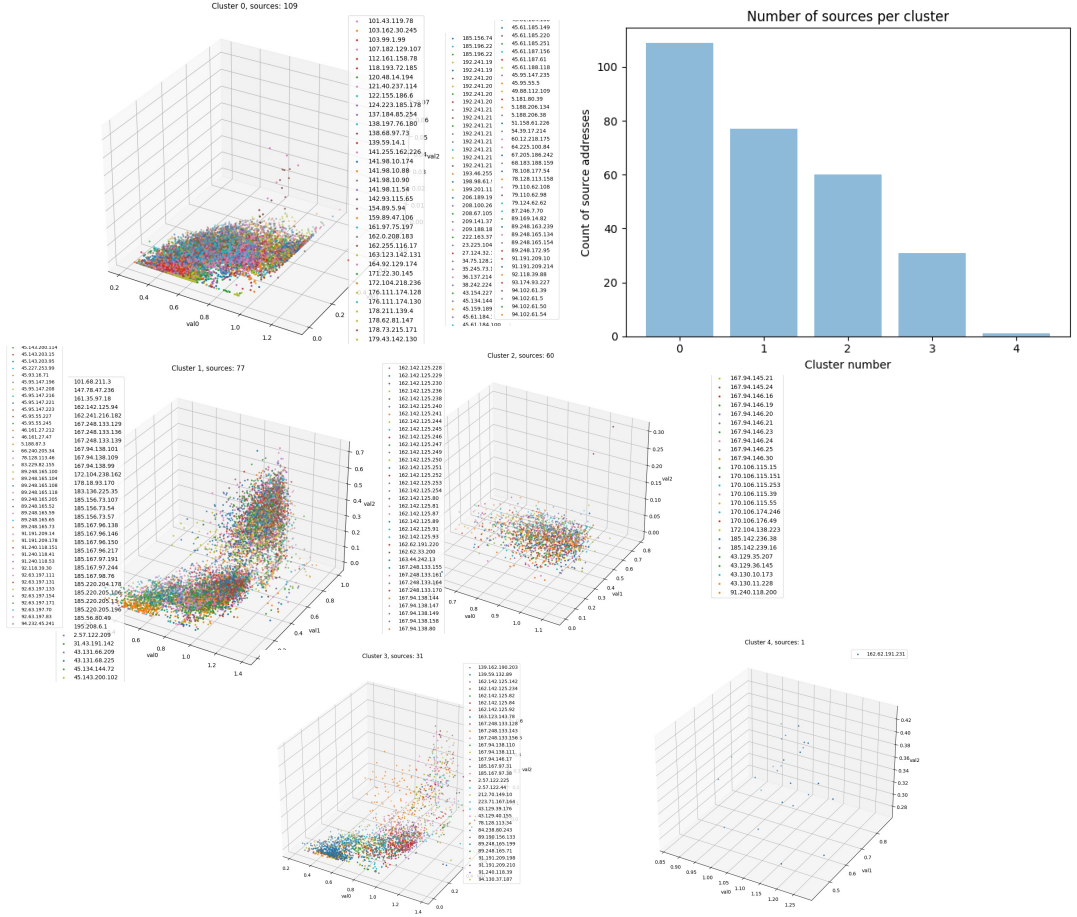
Max States Times

- Max States is the number of packets held in memory and used for interpolation
- The table below shows the amount of time it takes for different pieces of feature extraction depending on different max states
- Higher max states show better groupings of each source when clustering

Max States	Interpolation n	Number of packets	Features time (s)	Interpolation time (s)	PCA time (s)	Total time (s)	Clusters n (mean shift quantile=0.15)
10	10	64473	20.43	45.14	13.91	92.43	5
50	50	64473	22.39	38.17	11.69	86.38	4
64	64	64473	24.24	39.00	11.68	92.90	3
100	100	64473	17.60	32.36	9.64	74.45	4
128	128	64473	23.78	32.95	9.42	80.77	5
256	256	64473	25.55	26.23	6.64	74.28	7

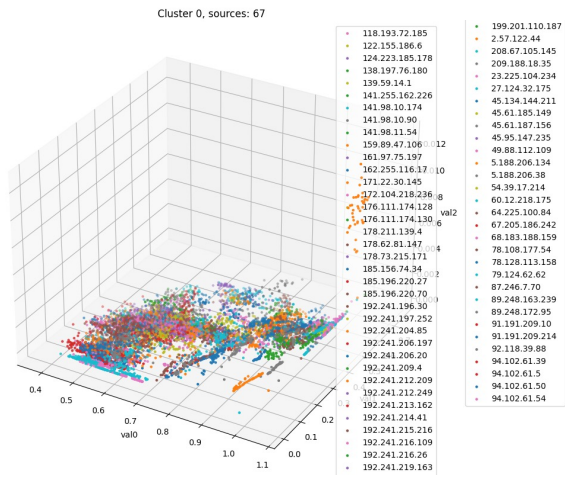
Max States 10

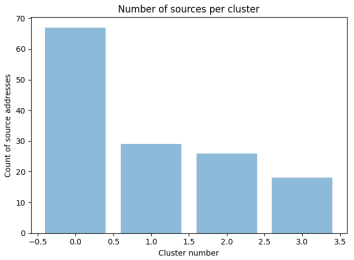
- traffic_type: 6
- max_states: 10
- interpolation_n: 10
- 64473 packets processed
- Features time: 20.43
- Interpolation time: 45.14
- PCA time: 13.91
- Total time: 92.43
- Clustering: Mean Shift, q=0.15

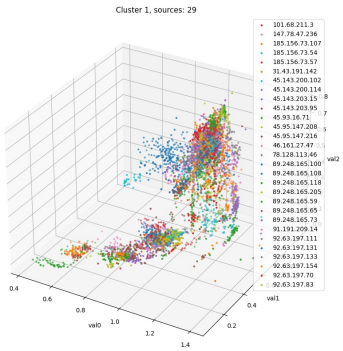


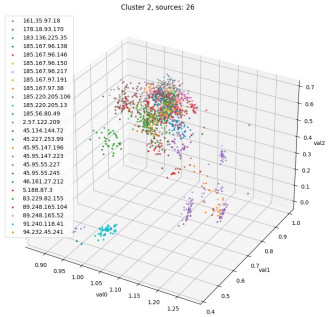
Max States 50

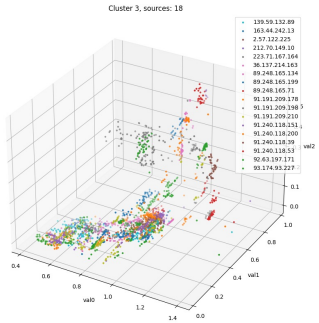
- traffic_type: 6
- max_states: 50
- interpolation_n: 50
- 64473 packets processed
- Features time: 22.39
- Interpolation time: 38.17
- PCA time: 11.69
- Total time: 86.38
- Clustering: Mean Shift, q=0.15





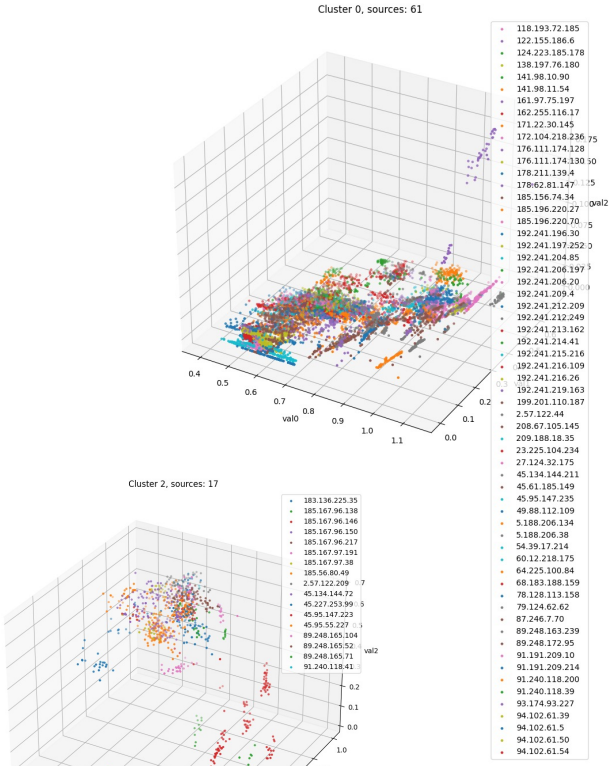


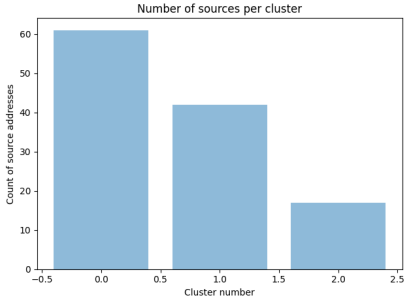


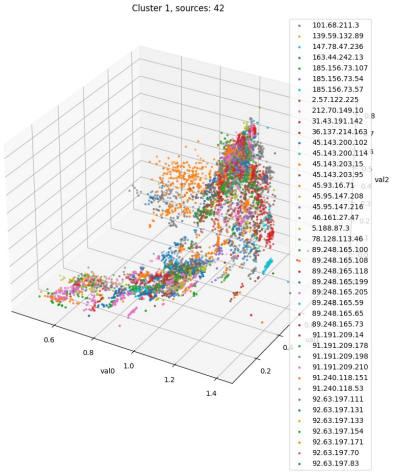


Max States 64

- traffic_type: 6
- max_states: 64
- interpolation_n: 64
- 64473 packets processed
- Features time: 24.24
- Interpolation time: 39.00
- PCA time: 11.68
- Total time: 92.90
- Clustering: Mean Shift, q=0.15

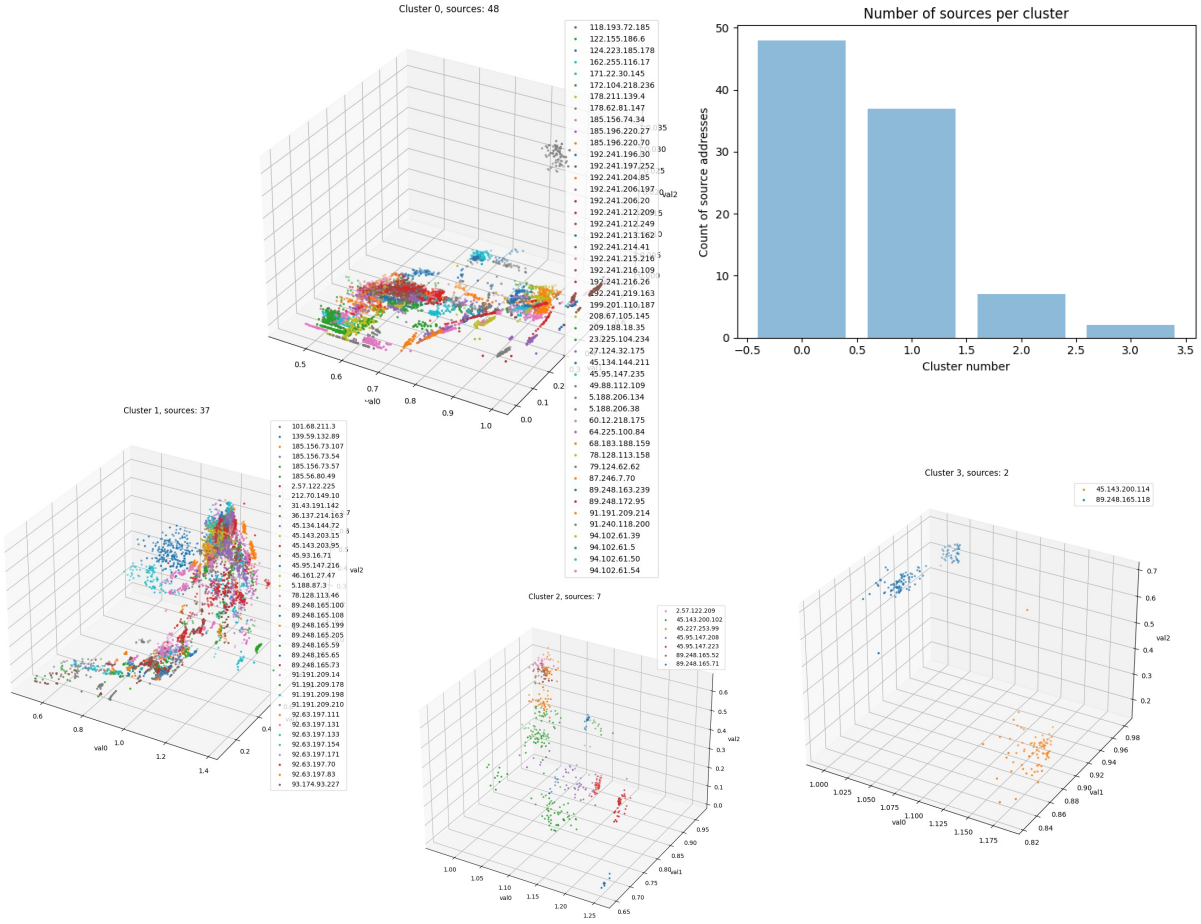






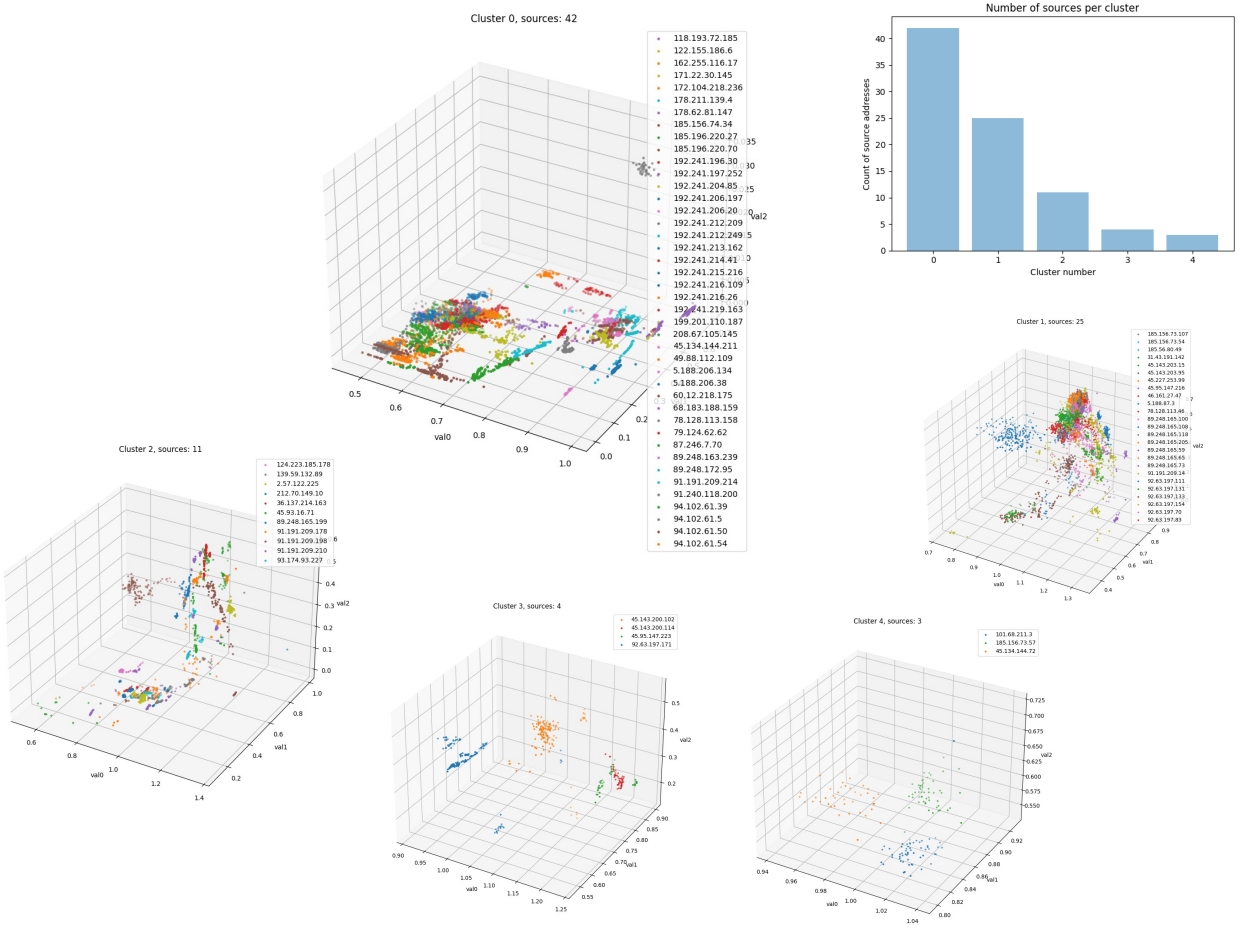
Max States 100

- traffic_type: 6
- max_states: 100
- interpolation_n: 100
- 64473 packets processed
- Features time: 17.60
- Interpolation time: 32.36
- PCA time: 9.64
- Total time: 74.45
- Clustering: Mean Shift, quar



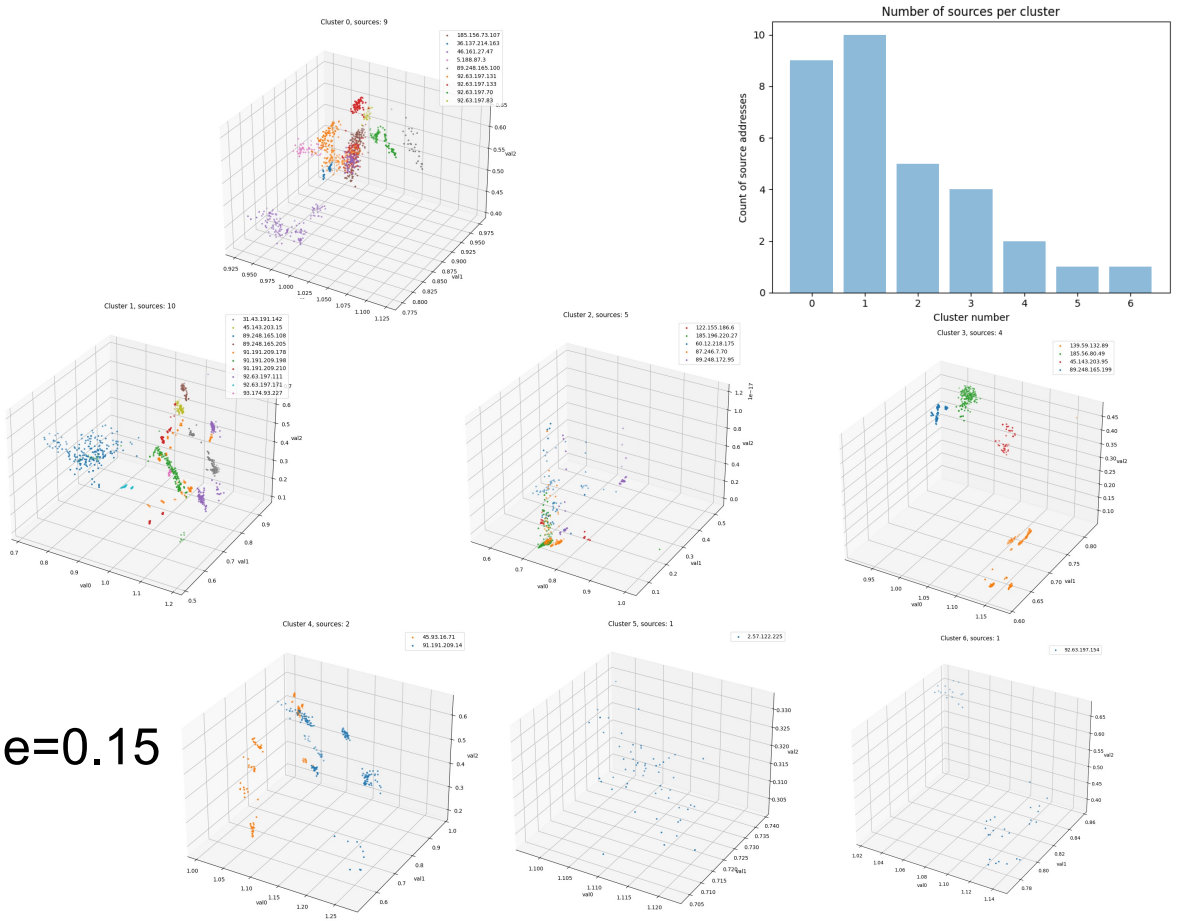
Max States 128

- traffic_type: 6
- max_states: 128
- interpolation_n: 128
- 64473 packets processed
- Features time: 23.78
- Interpolation time: 32.95
- PCA time: 9.42
- Total time: 80.77
- Clustering: Mean Shift, qua



Max States 256

- traffic_type: 6
- max_states: 256
- interpolation_n: 256
- 64473 packets processed
- Features time: 25.55
- Interpolation time: 26.23
- PCA time: 6.64
- Total time: 74.28
- Clustering: Mean Shift, quantile=0.15



Clustering Tests

For All Tests:

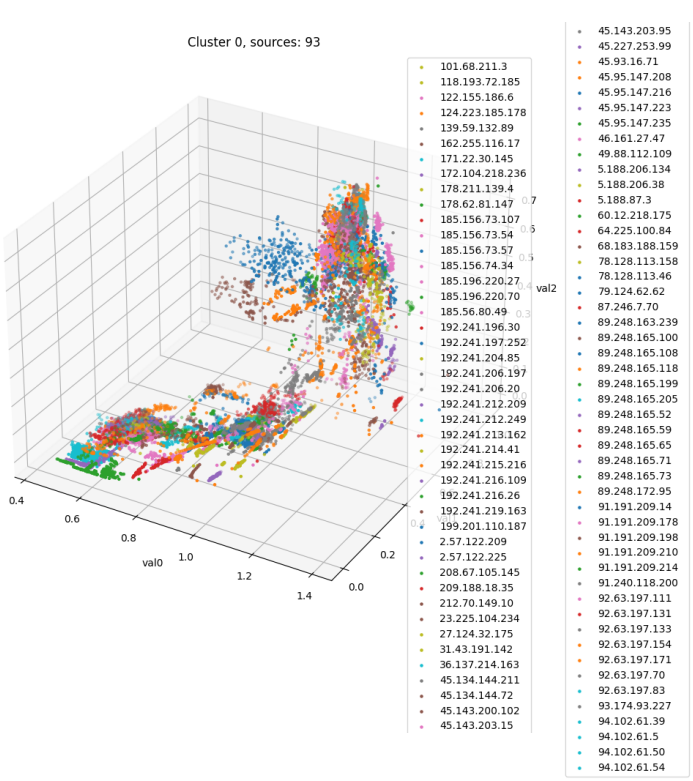
- traffic_type: 6
- max_states: 100
- interpolation_n: 100
- 64473 packets processed
- Features time: 17.602816104888916
- Interpolation time: 32.36164569854736
- PCA time: 9.637251138687134
- Total time: 74.4487636089325

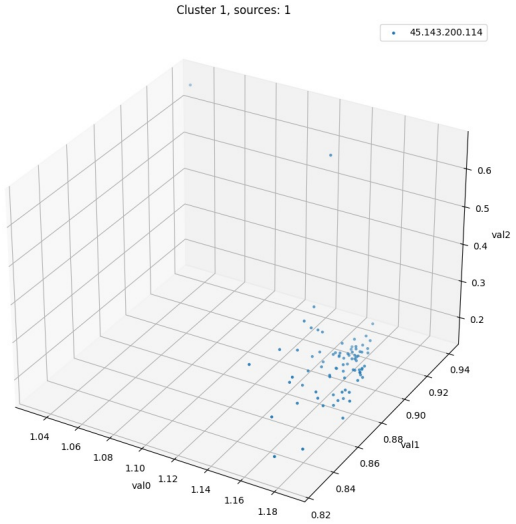
Clustering Metrics

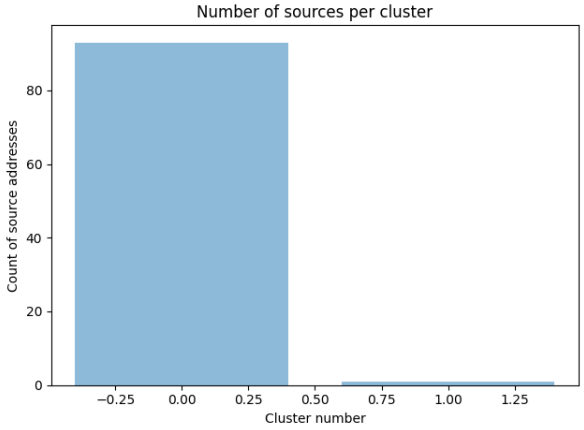
- The table to the right shows the clustering metrics for clustering algorithms with different parameters.
- Assessments show agglomerative clustering separating some sources by themselves when using single linkage regardless of the number of clusters
- Spectral clustering shows least ability to identify individual sources that are separated, whereas the mean shift algorithm shows separation of individual sources when using a smaller quartile value

			Silhouette Coefficient	Davies-Bouldin Index	Clustering Time (s) (w/o calculating matrix)
Agglomerative Clustering	n_clusters	Linkage			
	2	single	0.10010	1.12199	0.0134978
	2	average	0.40322	0.91366	0.0139527
	2	complete	0.26179	1.17829	0.0098705
	4	single	-0.00963	1.03387	0.0031600
	4	average	0.18627	1.25755	0.0146666
	4	complete	0.34936	1.04212	0.0227036
	6	single	-0.02710	0.90239	0.0147417
	6	average	0.13514	1.30577	0.0343530
	6	complete	0.23493	1.88944	0.0142896
Spectral Clustering	n_clusters				
	2		0.41933	0.99926	0.4628527
	4		0.26947	1.36879	0.5373106
	6		0.29994	1.16826	0.4075165
Mean Shift Clustering	quantile				
	0.1		0.28800	1.00302	2.9799180
	0.2		0.38469	1.01291	3.9945939
	0.3		0.41538	0.95008	3.6587830

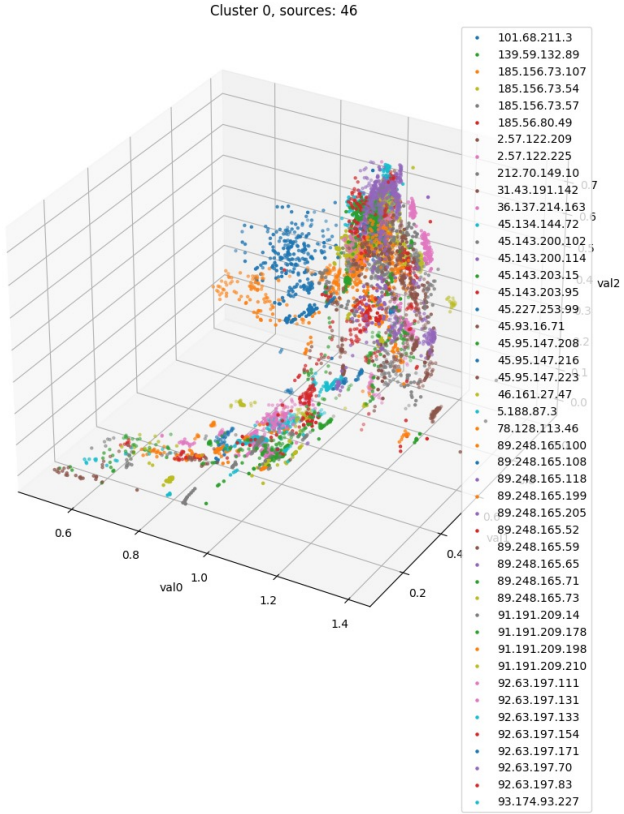
Agglomerative Clustering, n_clusters=2, linkage=single

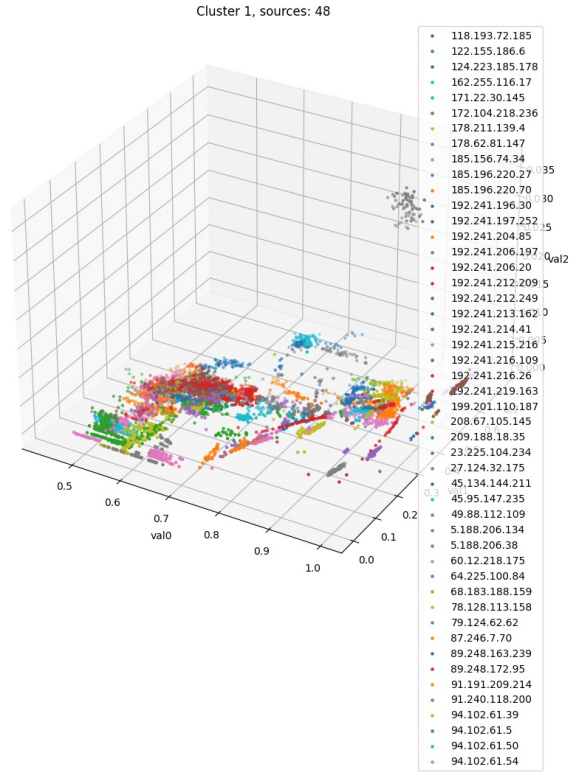


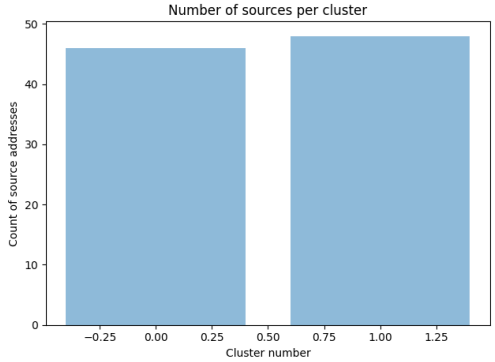




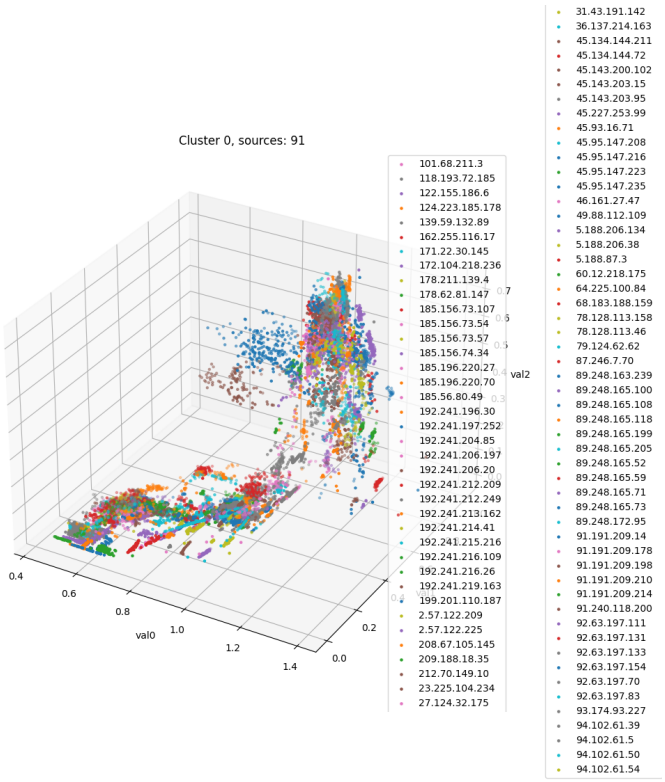
Agglomerative Clustering, n_clusters=2, linkage=complete

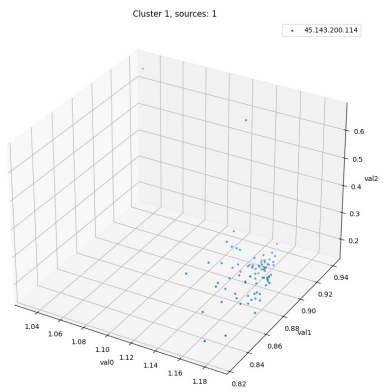


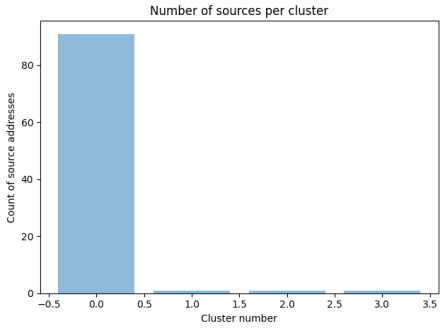


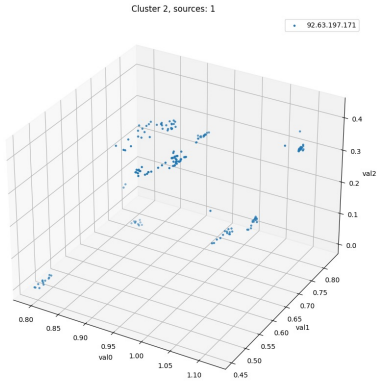


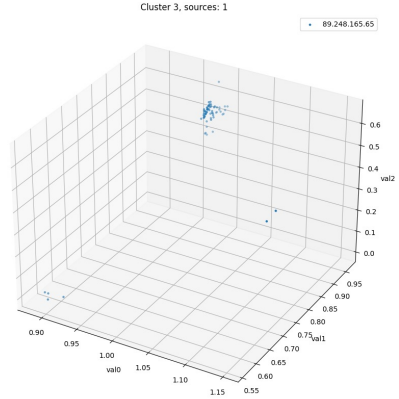
Agglomerative Clustering, n_clusters=4, linkage=single



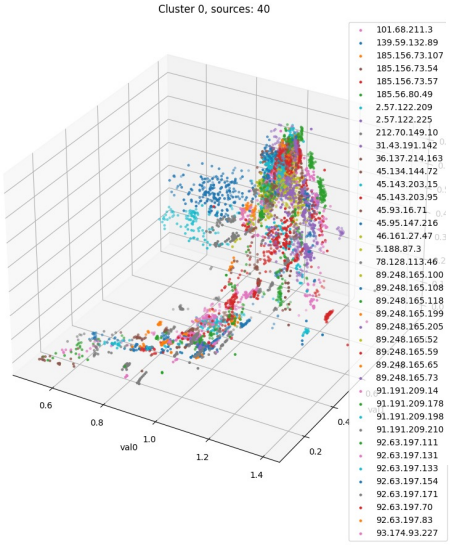


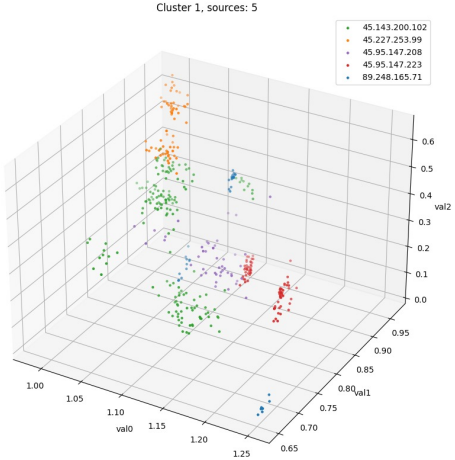


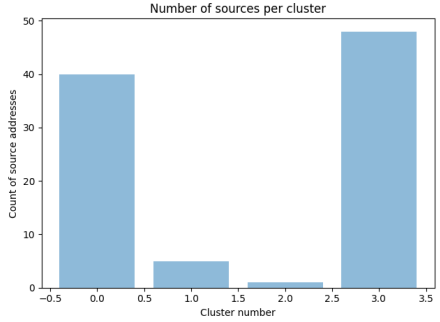


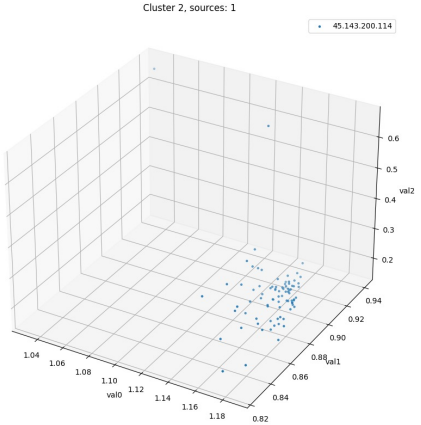


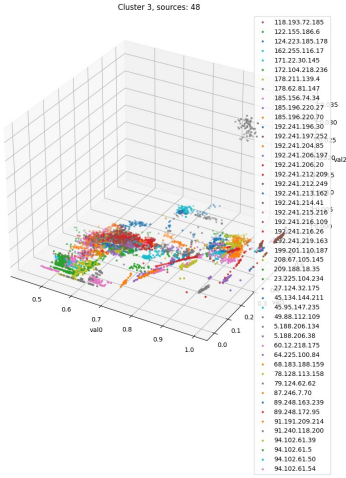
Agglomerative Clustering, n_clusters=4, linkage=average



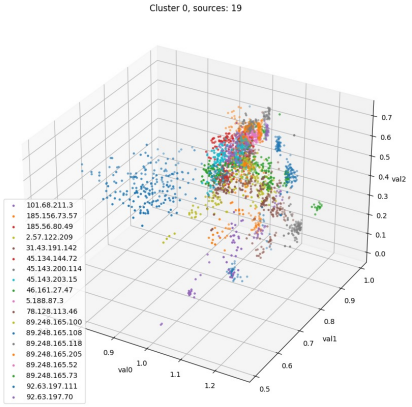


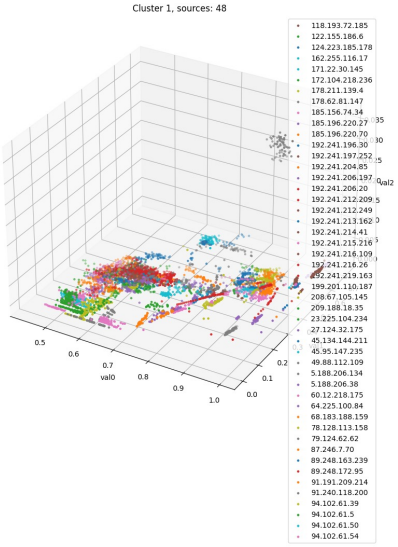


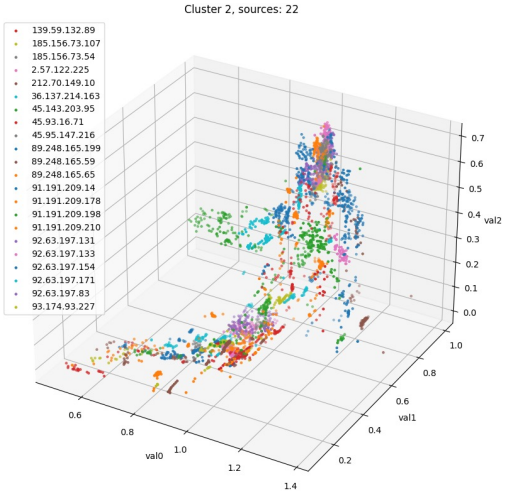


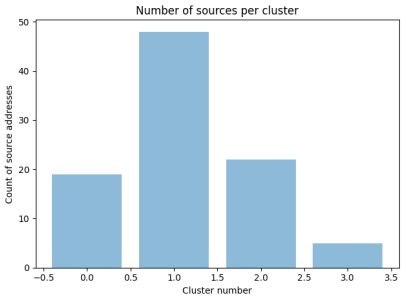


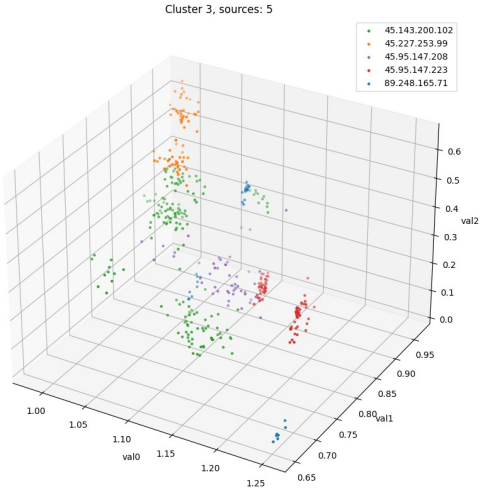
Agglomerative Clustering, n_clusters=4, linkage=complete



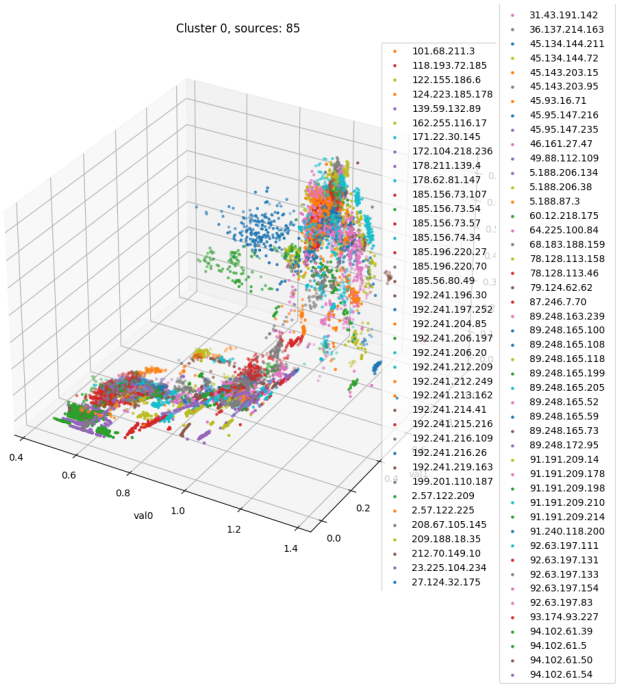


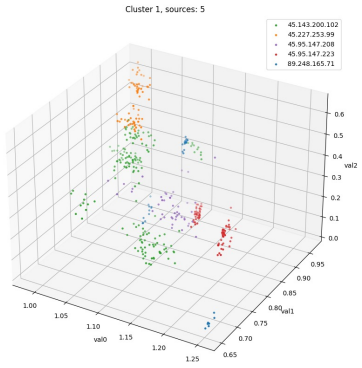


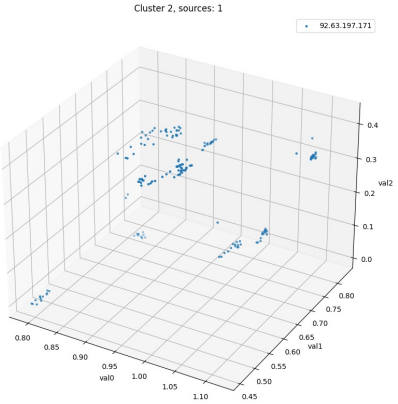


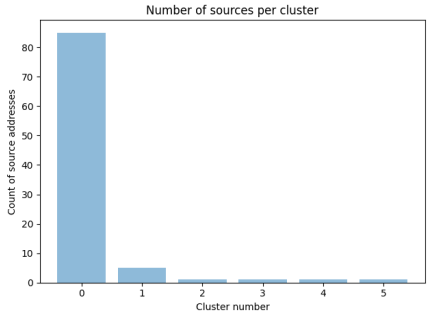


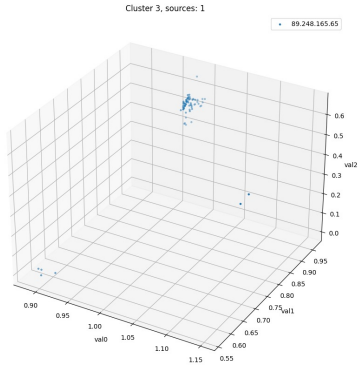
Agglomerative Clustering, n_clusters=6, linkage=single

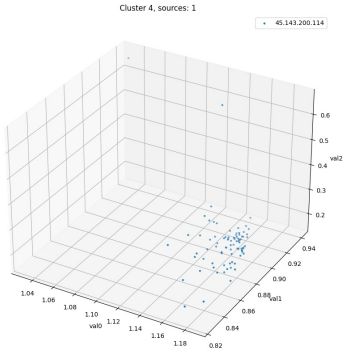


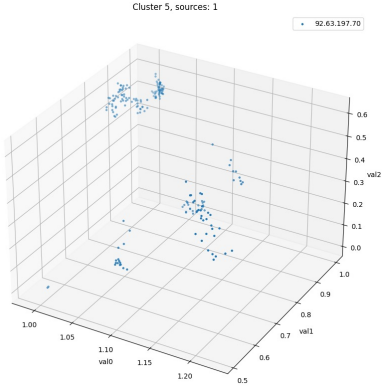




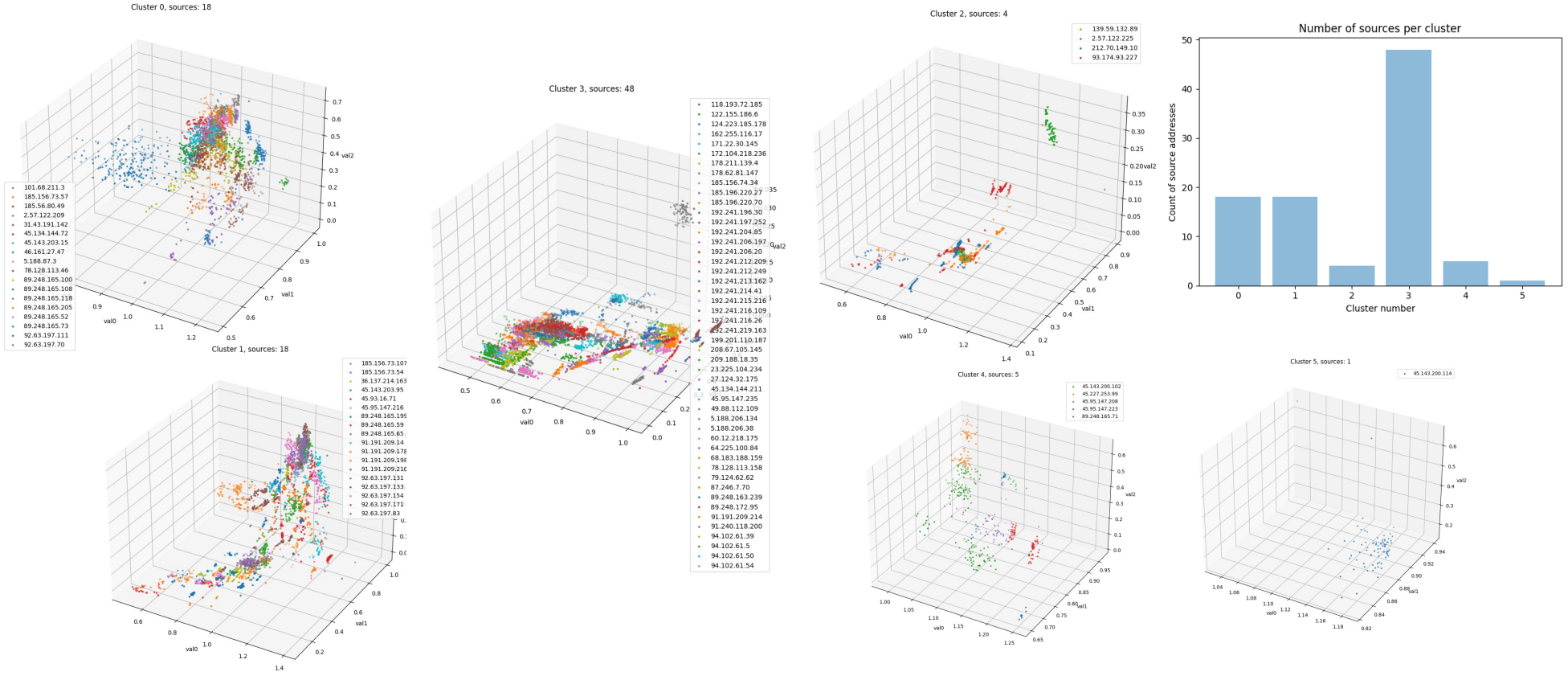




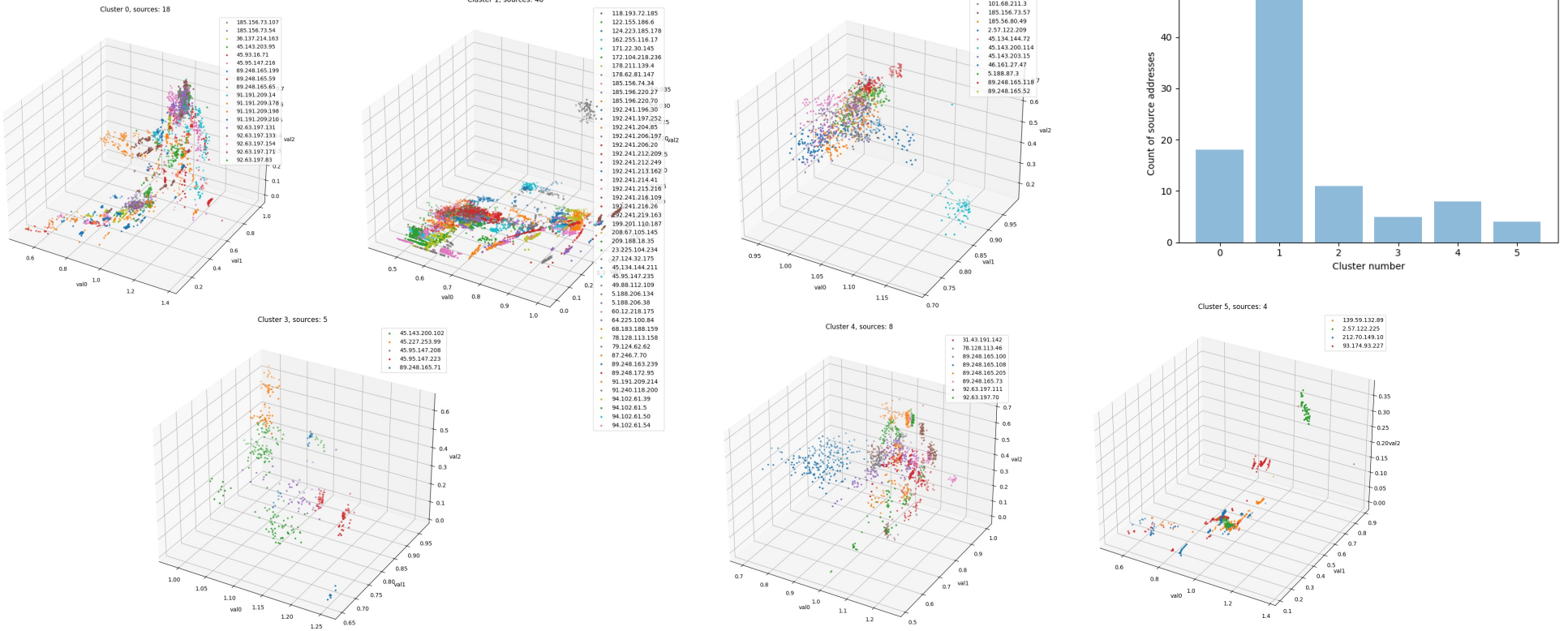




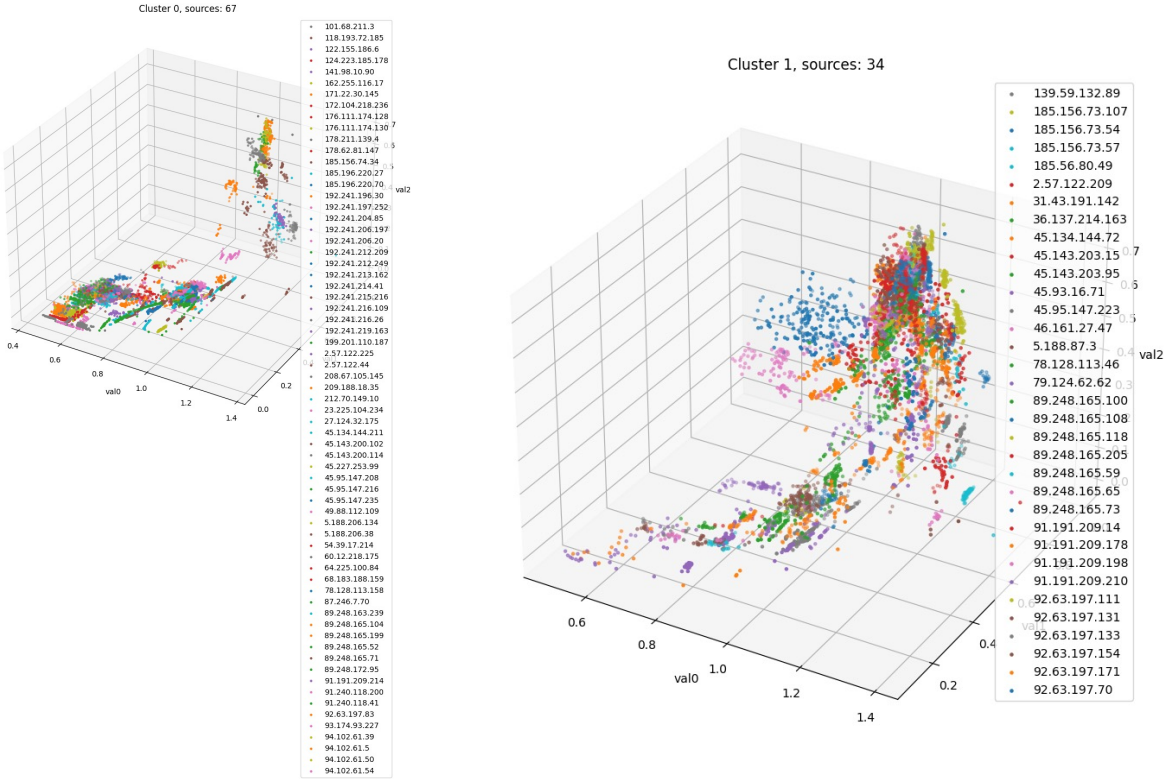
Agglomerative Clustering, n_clusters=6, linkage=average

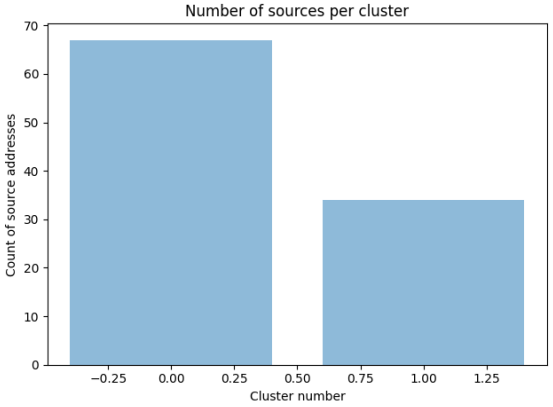


Agglomerative Clustering, n_clusters=6, linkage=complete

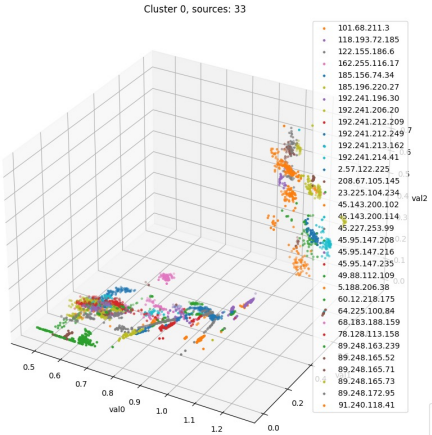


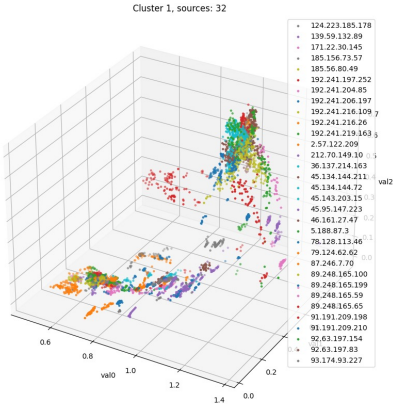
Spectral Clustering, n_clusters=2

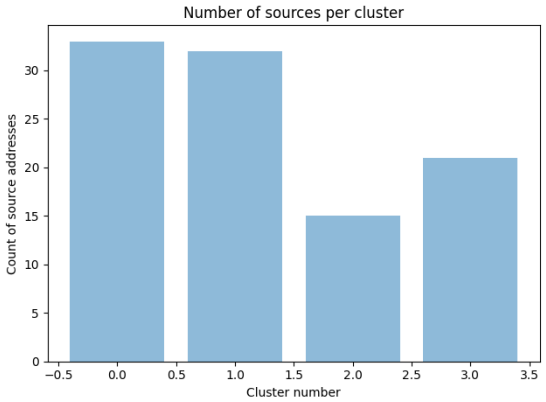


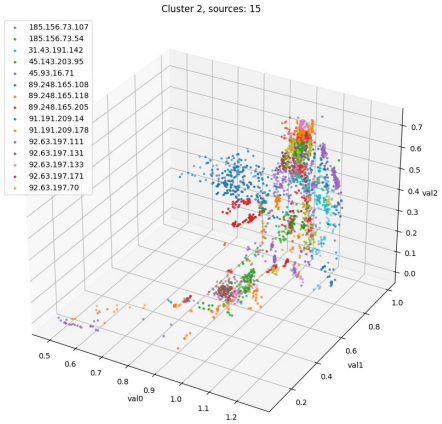


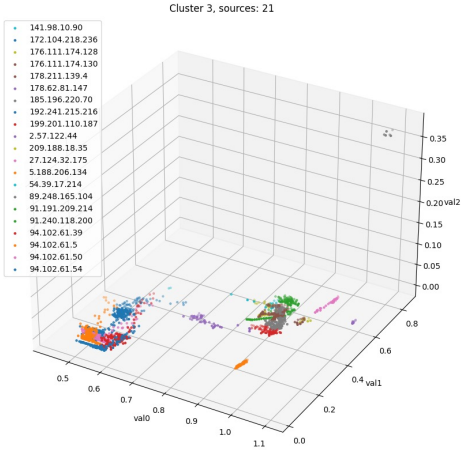
Spectral Clustering, n_clusters=4



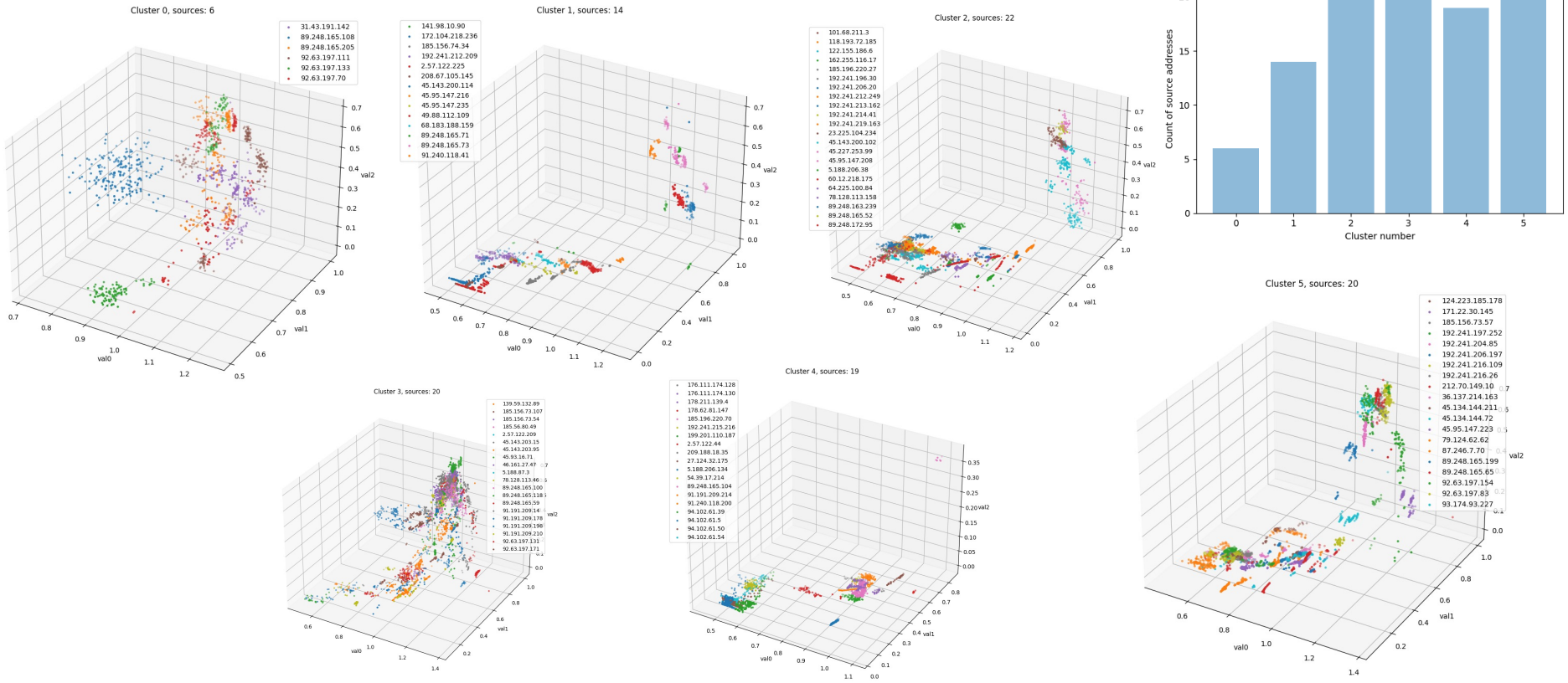




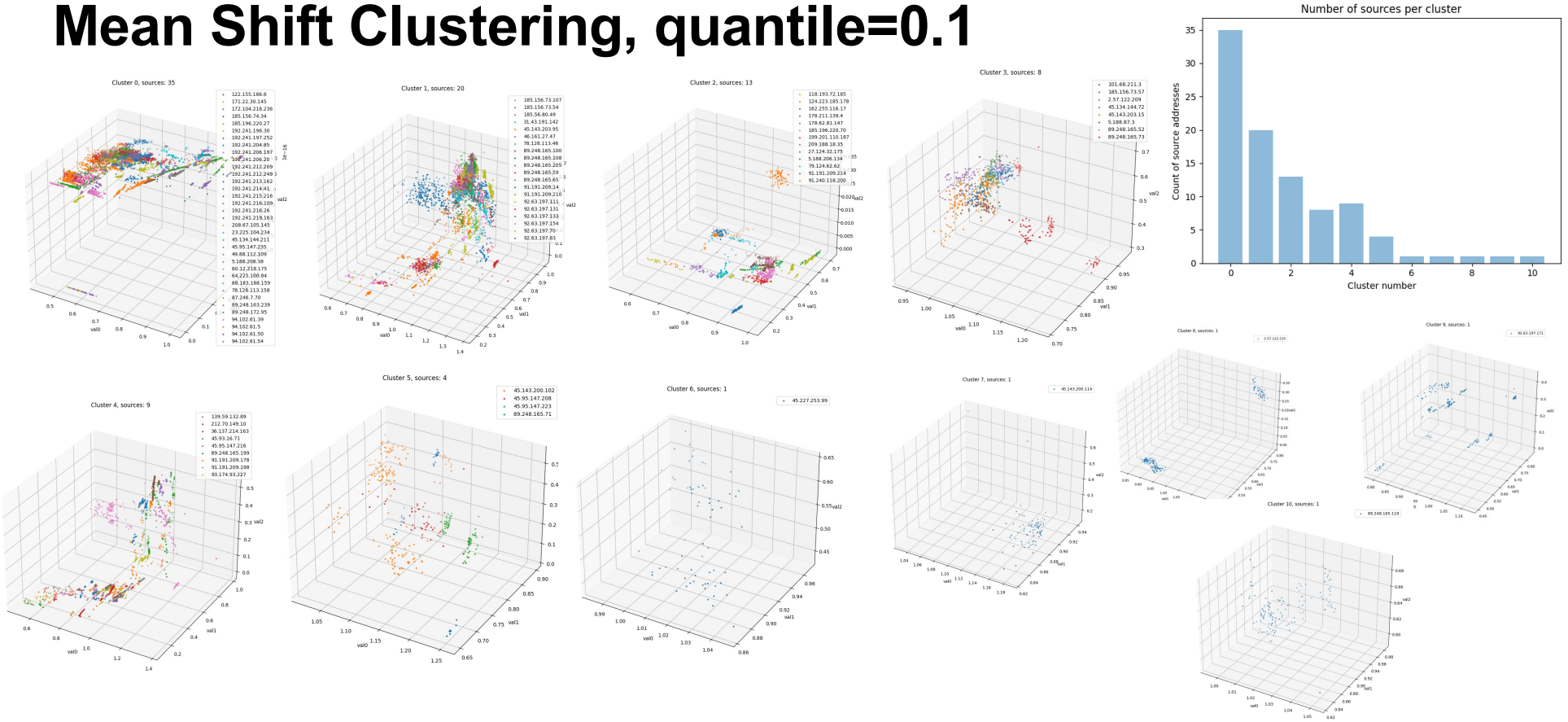




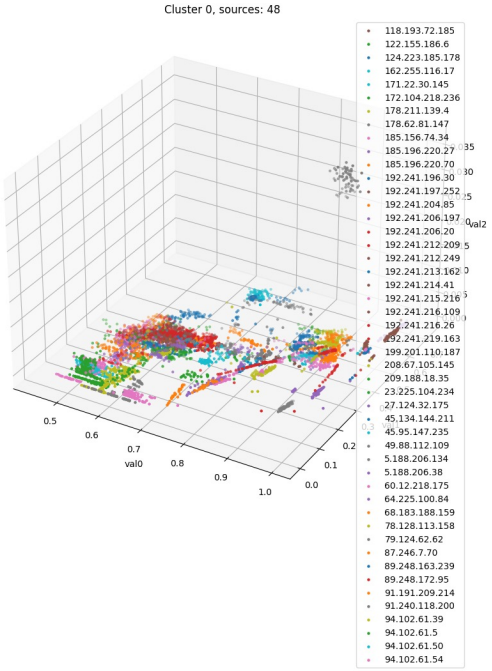
Spectral Clustering, n_clusters=6

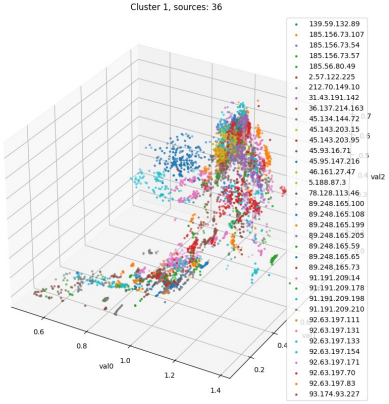


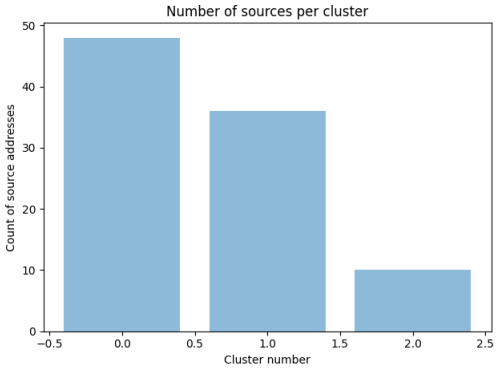
Mean Shift Clustering, quantile=0.1

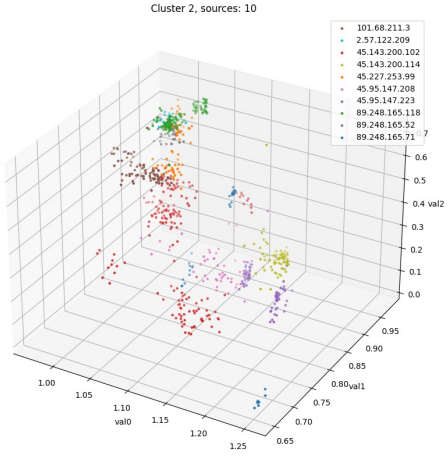


Mean Shift Clustering, quantile=0.2

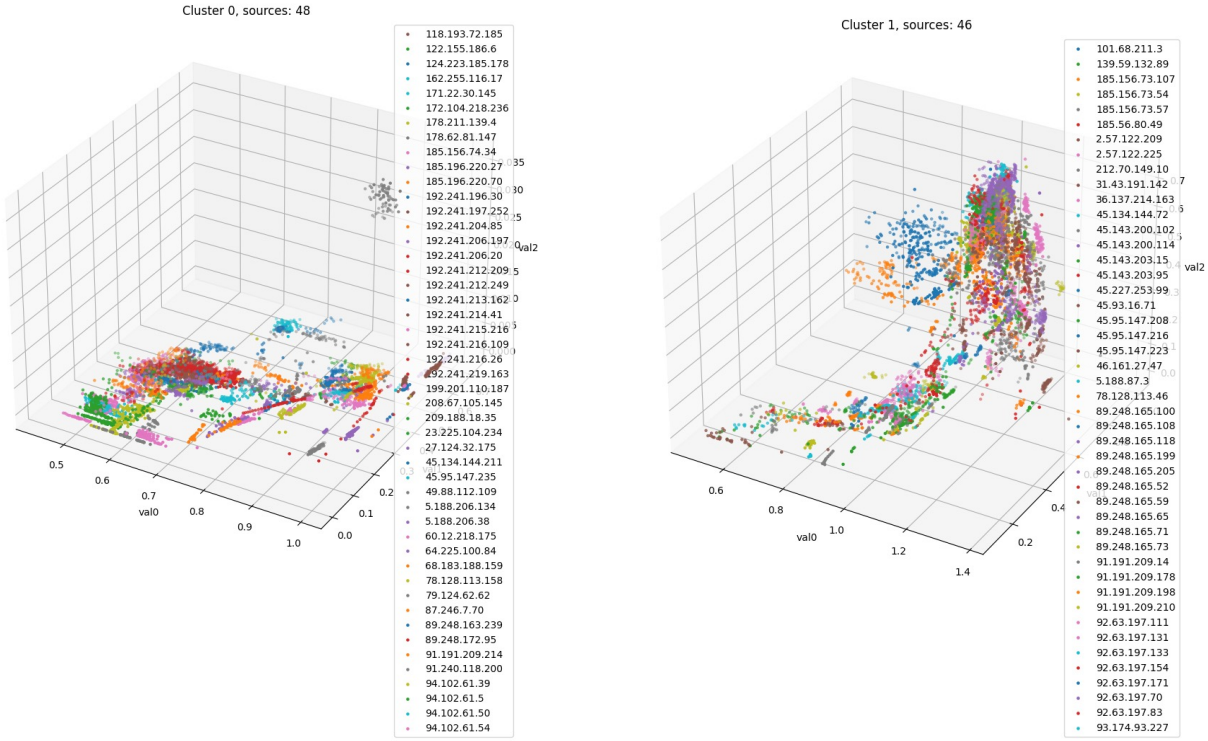


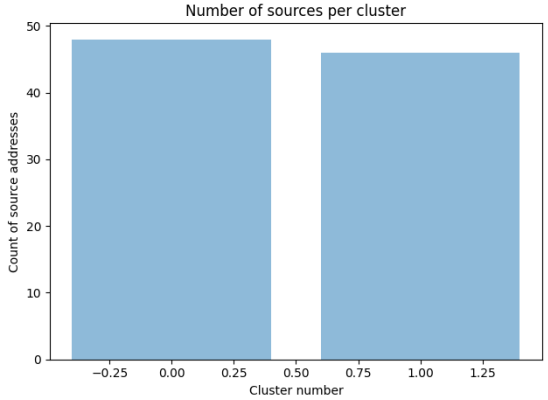






Mean Shift Clustering, quantile=0.3





PCA Explained Variances

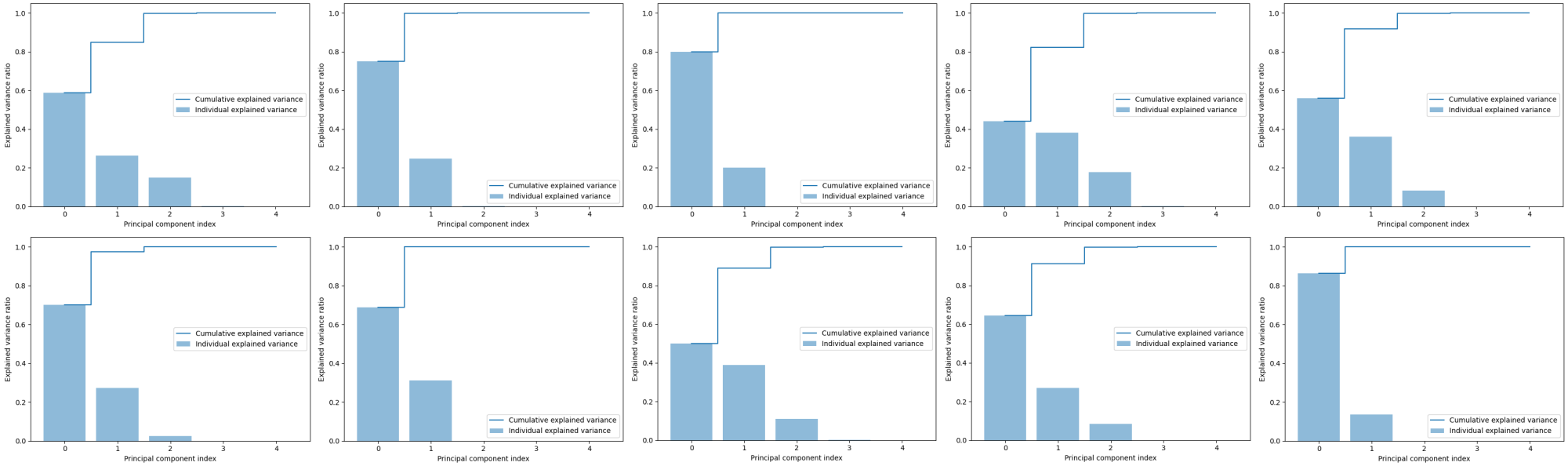
Calculating PCA Explained Variances – 5 PC Values

- traffic_type: 6
- max_states: 100
- interpolation_n: 100
- 64473 packets processed
- Features time: 20.58
- Interpolation time: 40.10
- PCA time: 11.57
- Total time: 88.99
- Dissimilarity matrix time: 4.25
  - For mean shift and agglomerative
- Spectral distance matrix time: 9.28
  - For spectral clustering
- Mean shift clustering time: 0.74
  - quantile=0.1
- Agglomerative clustering time: 0.01
  - n_clusters=10, linkage=complete
- Spectral clustering time: 0.18
  - n_clusters=10

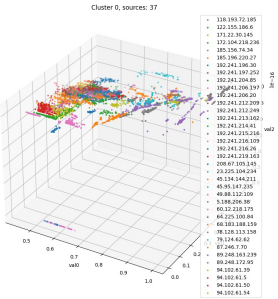
Samples of PCA Explained Variances – 5 PC Values

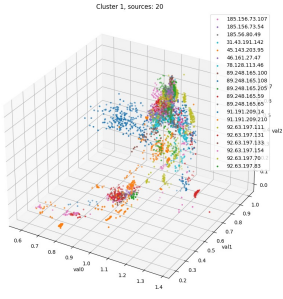
Explained variance – how much variance is contained in each principal component value

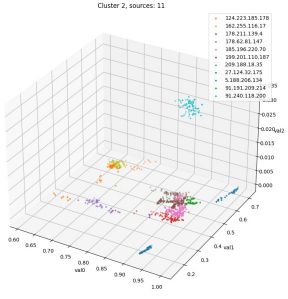
Almost all variance is contained in the first 3 principal component values

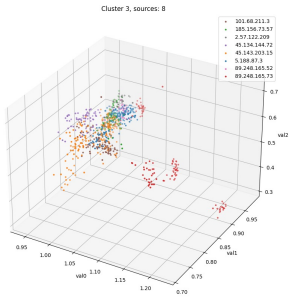


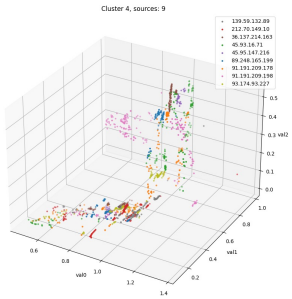
Clusters – 5 PC Values

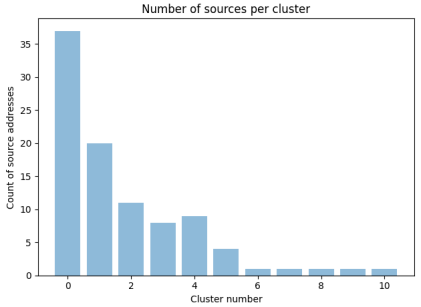


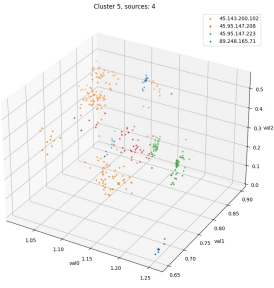


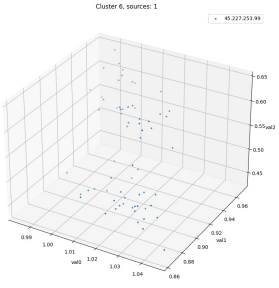


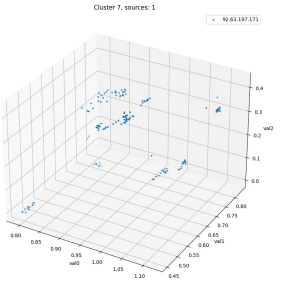


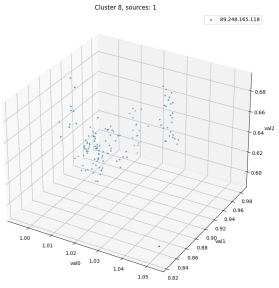


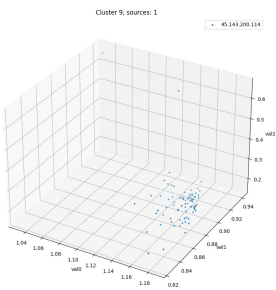


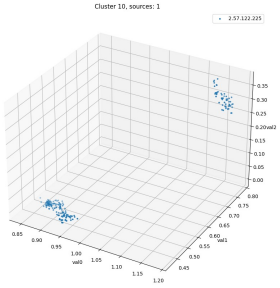










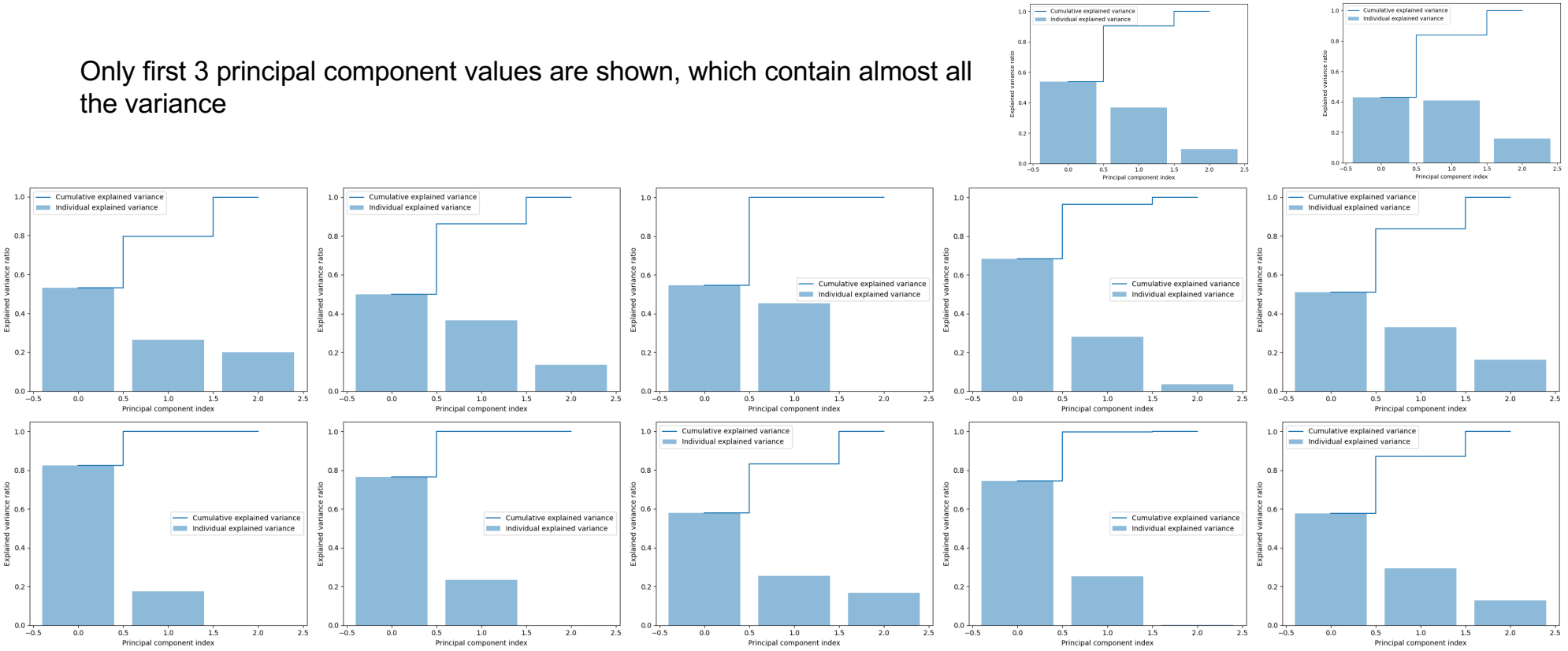


Calculating PCA Explained Variances – 3 PC Values

- traffic_type: 6
- max_states: 100
- interpolation_n: 100
- 64473 packets processed
- Features time: 20.19
- Interpolation time: 38.76
- PCA time: 11.10
- Total time: 90.80
- Dissimilarity matrix time: 4.94
  - For mean shift and agglomerative
- Spectral distance matrix time: 15.69
  - For spectral clustering
- Mean shift clustering time: 3.53
  - quantile=0.1
- Agglomerative clustering time: 0.02
  - n_clusters=10, linkage=complete
- Spectral clustering time: 0.22
  - n_clusters=10

Samples of PCA Explained Variances – 3 PC Values

Only first 3 principal component values are shown, which contain almost all the variance



Clusters – 3 PC Values