# Feature selection results for big\_dataset and Nmin = 15

1. result: **(signal\_strength, peak\_frequency, max\_frequency)**

**Estimated Epsilon: 0.0079026, Picked Epsilon: 0.037903, MinNumPoints: 15**

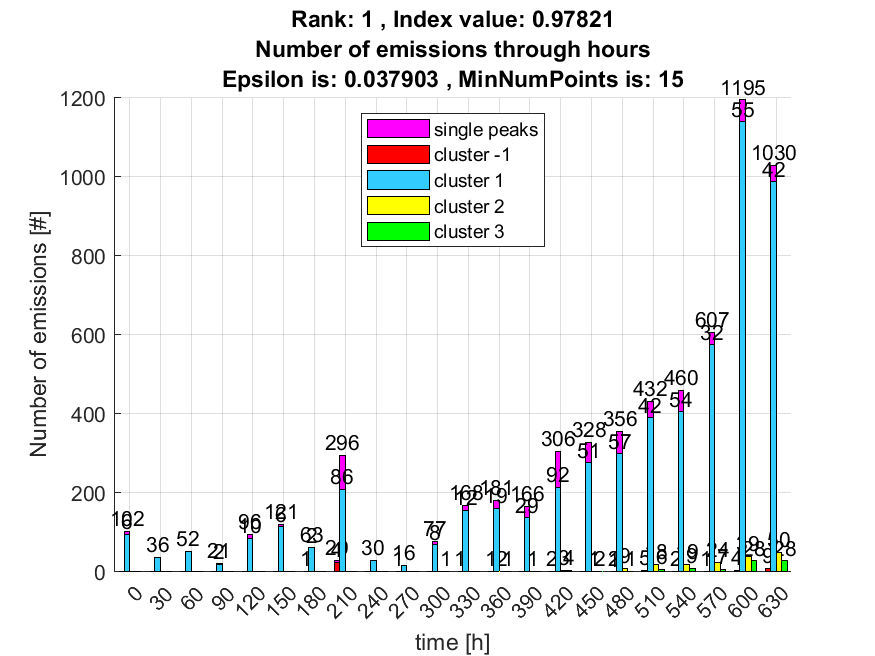
**DBCV index value: 0.97821**

**cluster -1 had 0.81967% single peaks**

**cluster 1 had 99.1803% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**2. result: (peak\_amplitude, peak\_frequency, max\_frequency)**

**Estimated Epsilon: 0.010077, Picked Epsilon: 0.030077, MinNumPoints: 15**

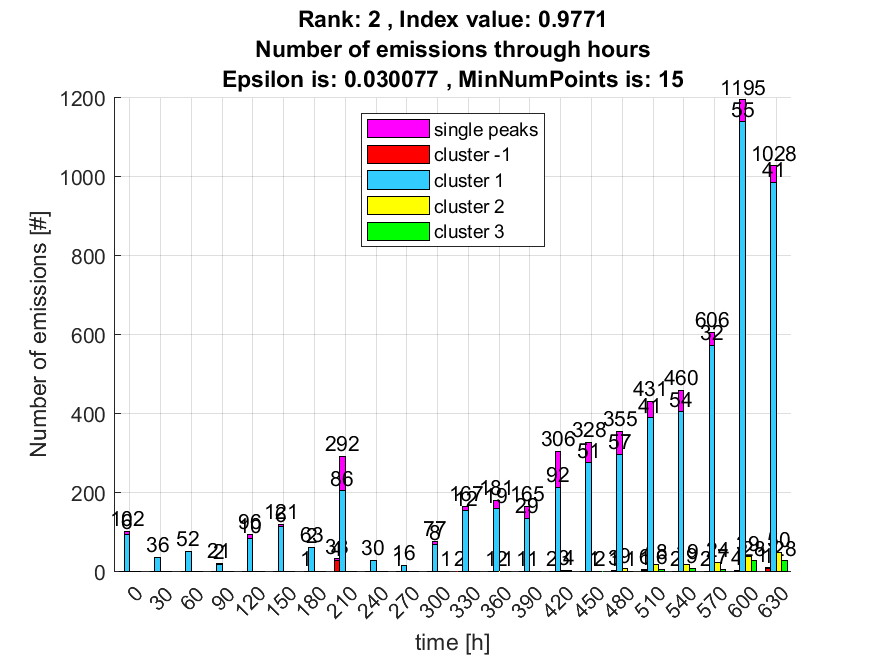
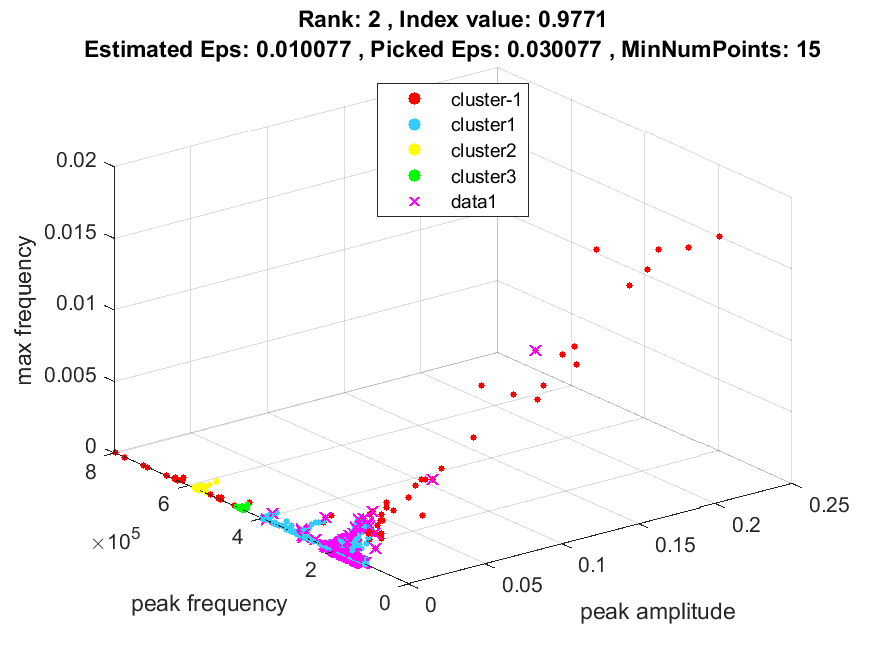
**DBCV index value: 0.9771**

**cluster -1 had 1.1475% single peaks**

**cluster 1 had 98.8525% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**3. result: (peak\_amplitude, pp4, peak\_frequency)**

**Estimated Epsilon: 0.024553, Picked Epsilon: 0.044553, MinNumPoints: 15**

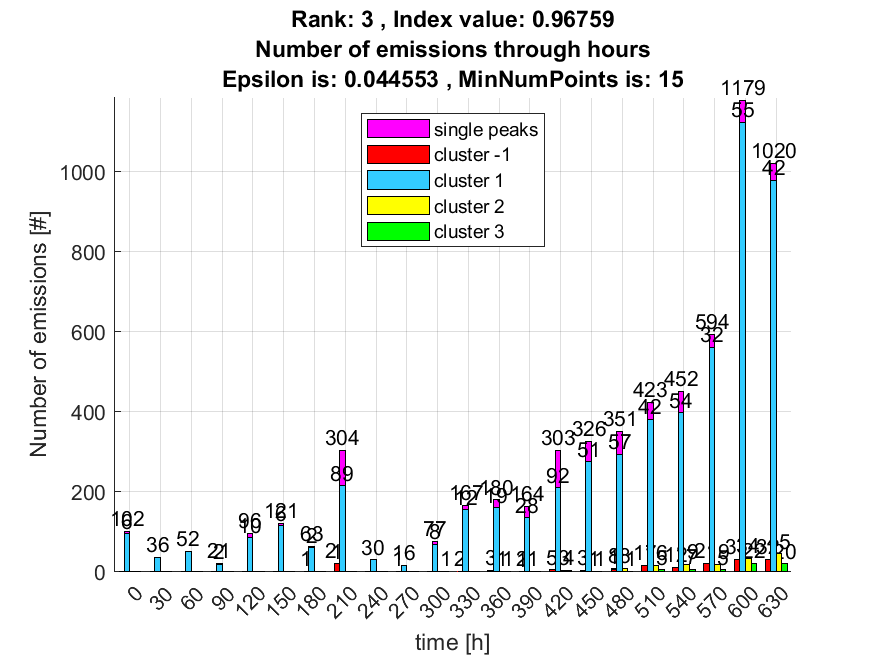
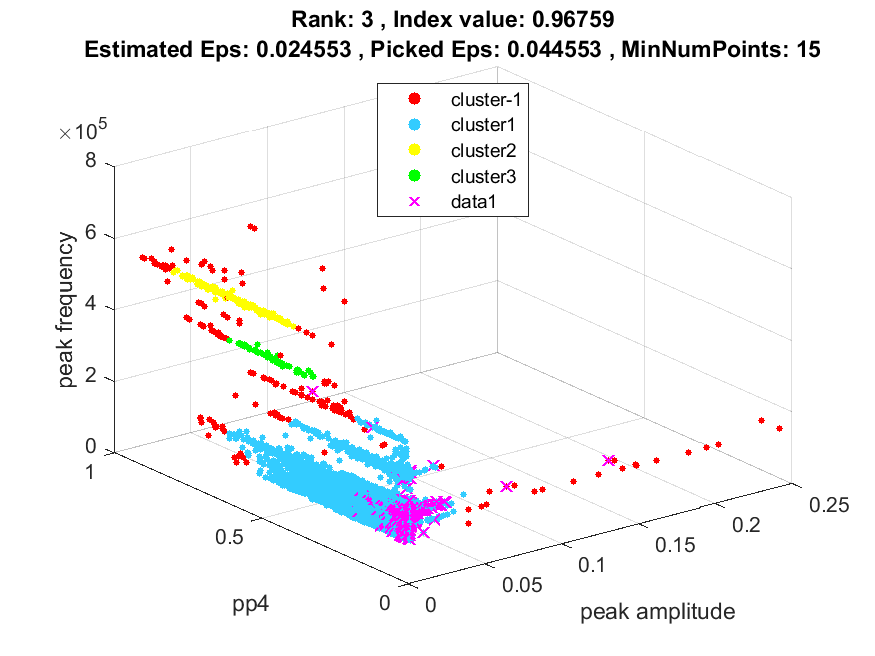
**DBCV index value: 0.96759**

**cluster -1 had 0.4918% single peaks**

**cluster 1 had 99.5082% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**4. result: (signal\_strength, peak\_frequency, total\_counts)**

**Estimated Epsilon: 0.025843, Picked Epsilon: 0.045843, MinNumPoints: 15**

**DBCV index value: 0.96564**

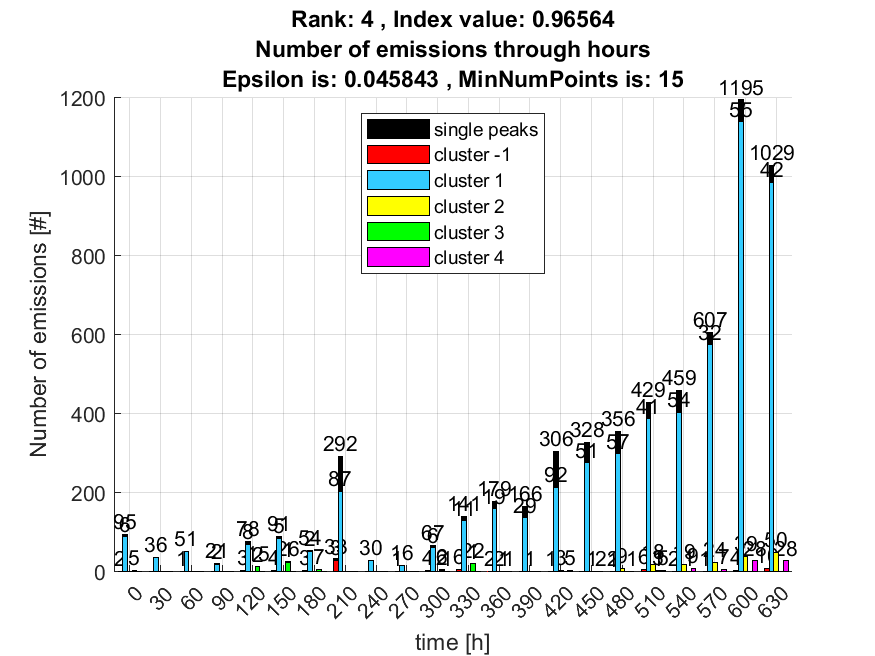
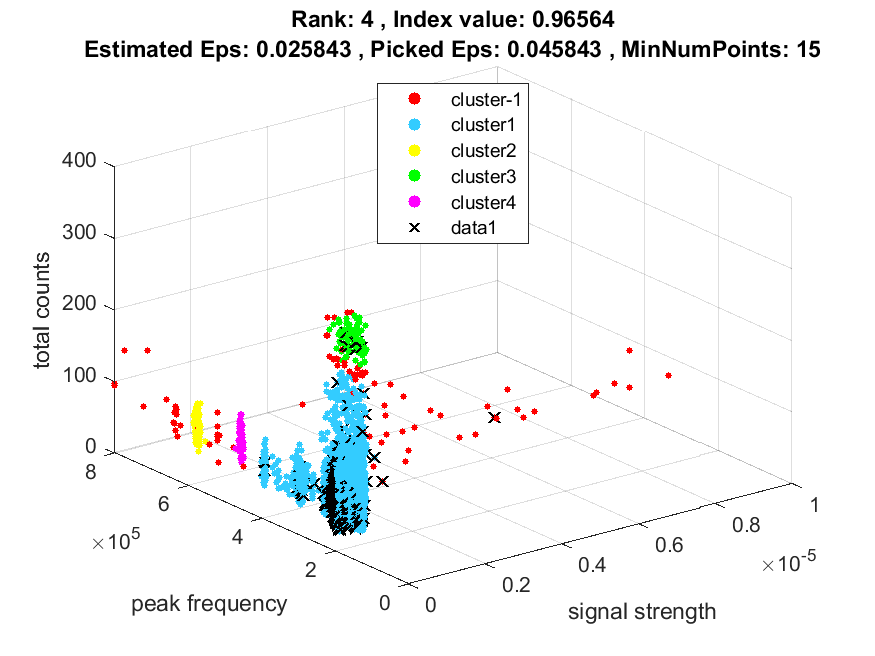
**cluster -1 had 0.65574% single peaks**

**cluster 1 had 98.1967% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 1.1475% single peaks**

**cluster 4 had 0% single peaks**



**5. result: (peak\_amplitude, pp1, peak\_frequency)**

**Estimated Epsilon: 0.016395, Picked Epsilon: 0.046395, MinNumPoints: 15**

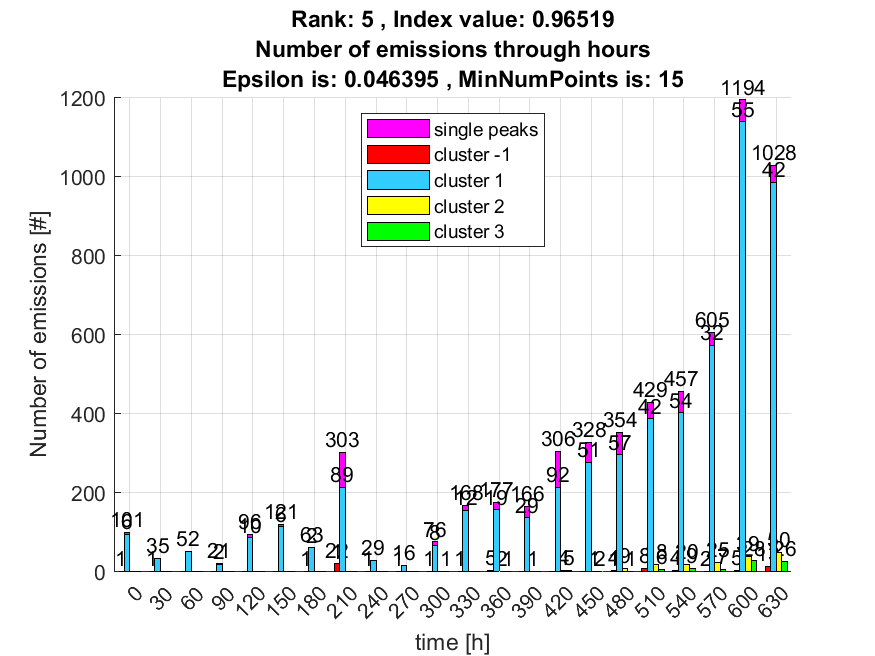
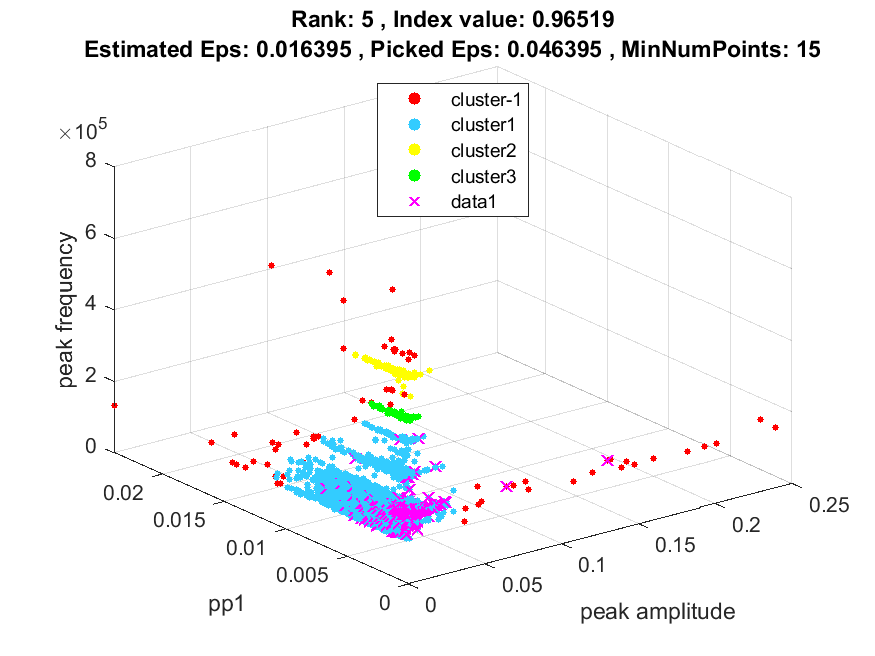
**DBCV index value: 0.96519**

**cluster -1 had 0.32787% single peaks**

**cluster 1 had 99.6721% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**6. result: (peak\_frequency, max\_frequency, total\_counts)**

**Estimated Epsilon: 0.024849, Picked Epsilon: 0.044849, MinNumPoints: 15**

**DBCV index value: 0.96518**

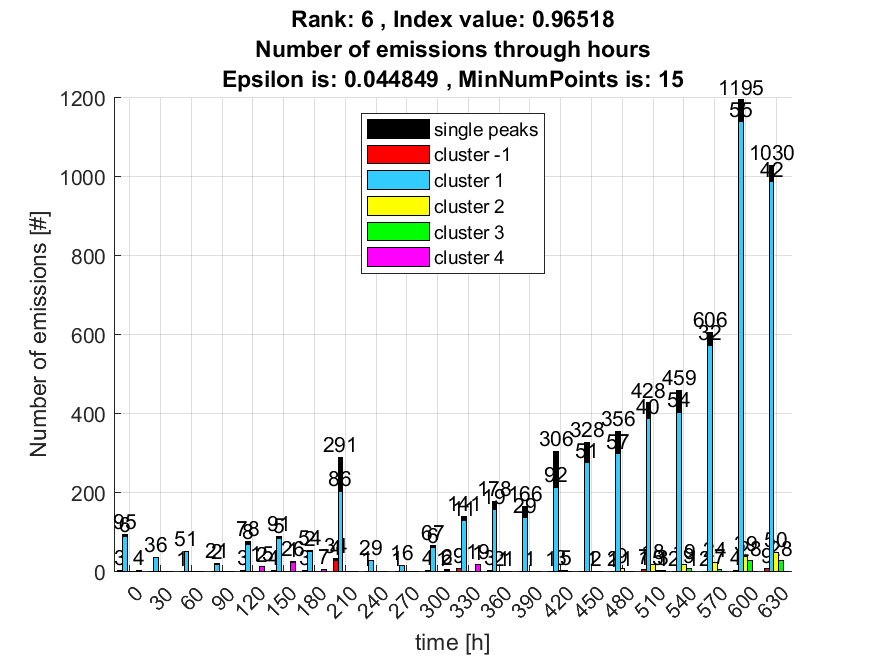
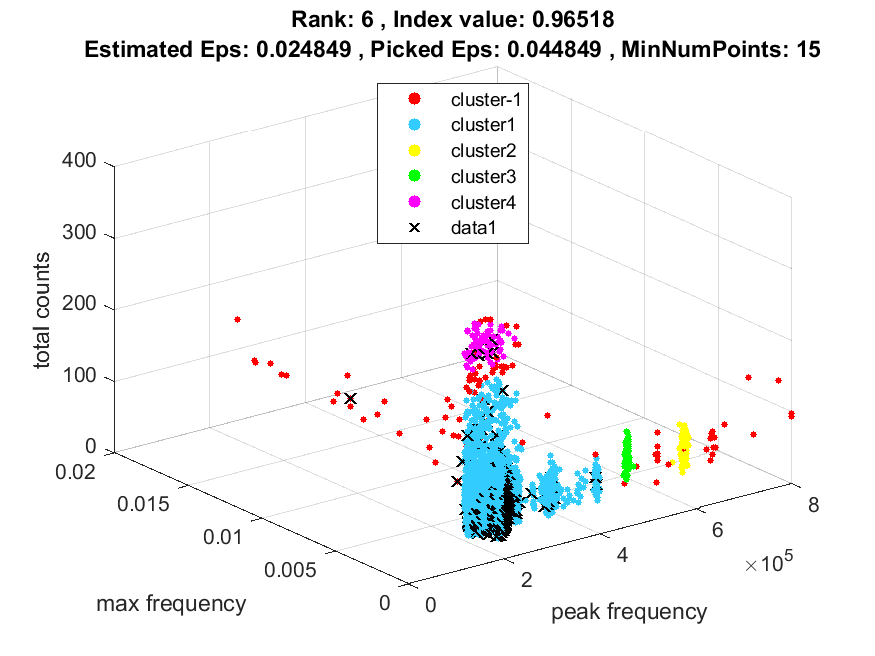
**cluster -1 had 0.98361% single peaks**

**cluster 1 had 97.8689% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**

**cluster 4 had 1.1475% single peaks**



**7. result: (peak\_amplitude, peak\_frequency, fall\_time)**

**Estimated Epsilon: 0.019445, Picked Epsilon: 0.049445, MinNumPoints: 15**

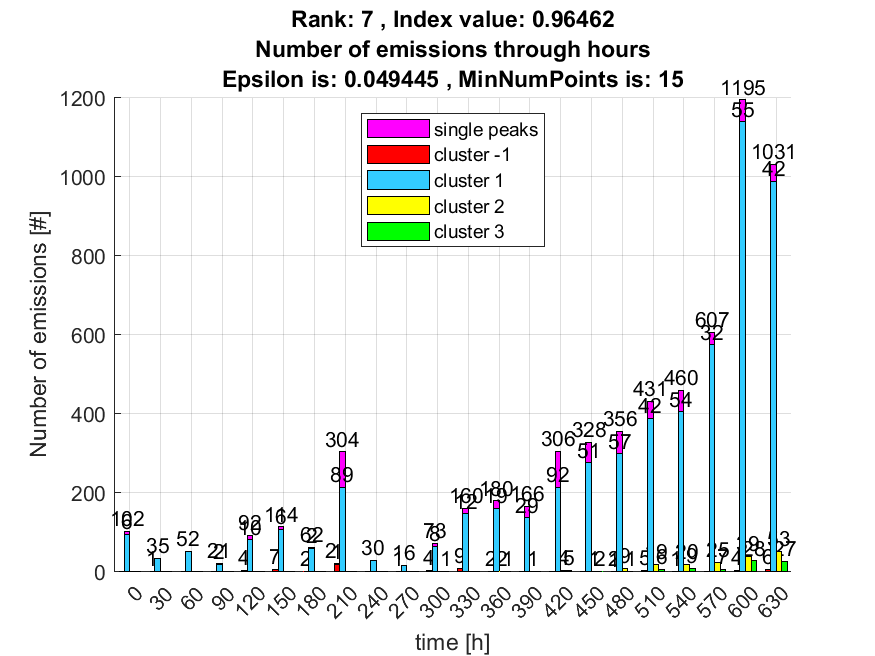
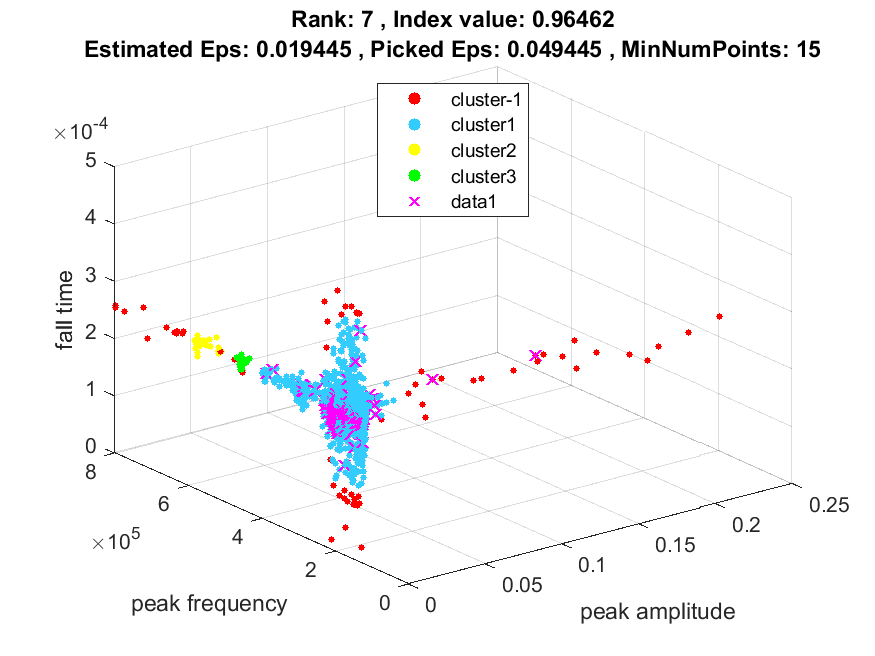
**DBCV index value: 0.96462**

**cluster -1 had 0.32787% single peaks**

**cluster 1 had 99.6721% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**8. result: (peak\_amplitude, peak\_frequency, total\_counts)**

**Estimated Epsilon: 0.026485, Picked Epsilon: 0.056485, MinNumPoints: 15**

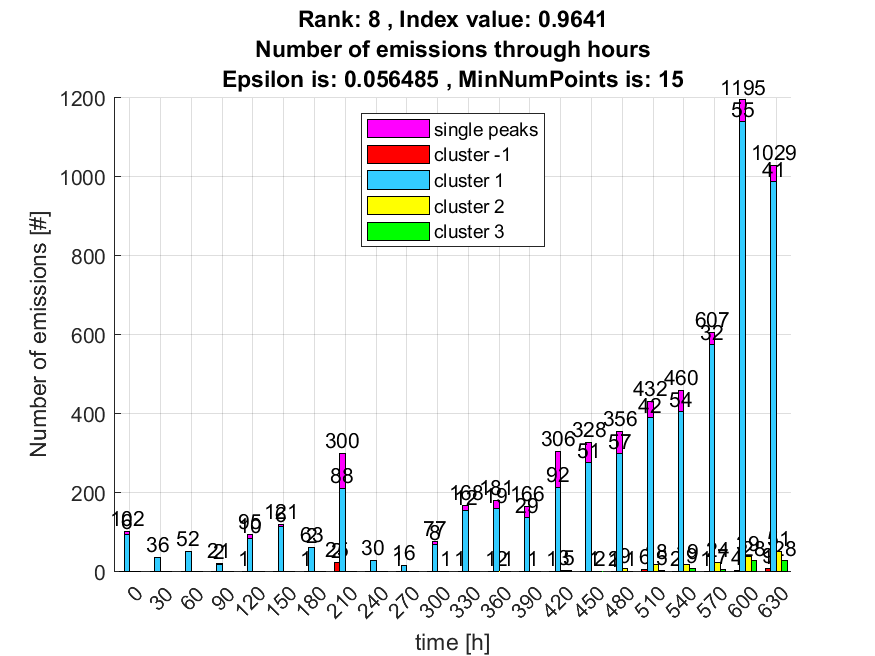
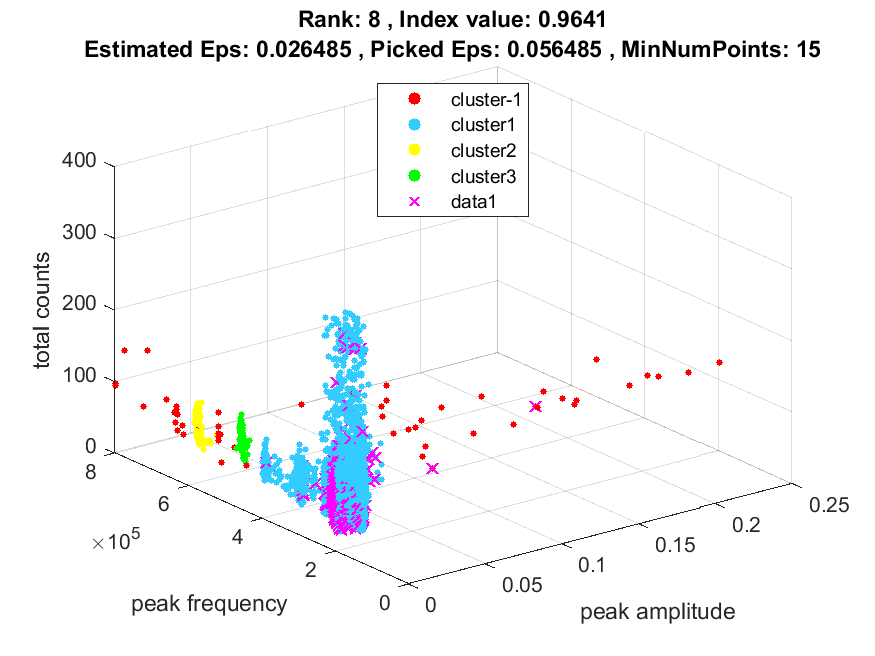
**DBCV index value: 0.9641**

**cluster -1 had 0.65574% single peaks**

**cluster 1 had 99.3443% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**9. result: (peak\_frequency, max\_frequency, fall\_time)**

**Estimated Epsilon: 0.018014, Picked Epsilon: 0.048014, MinNumPoints: 15**

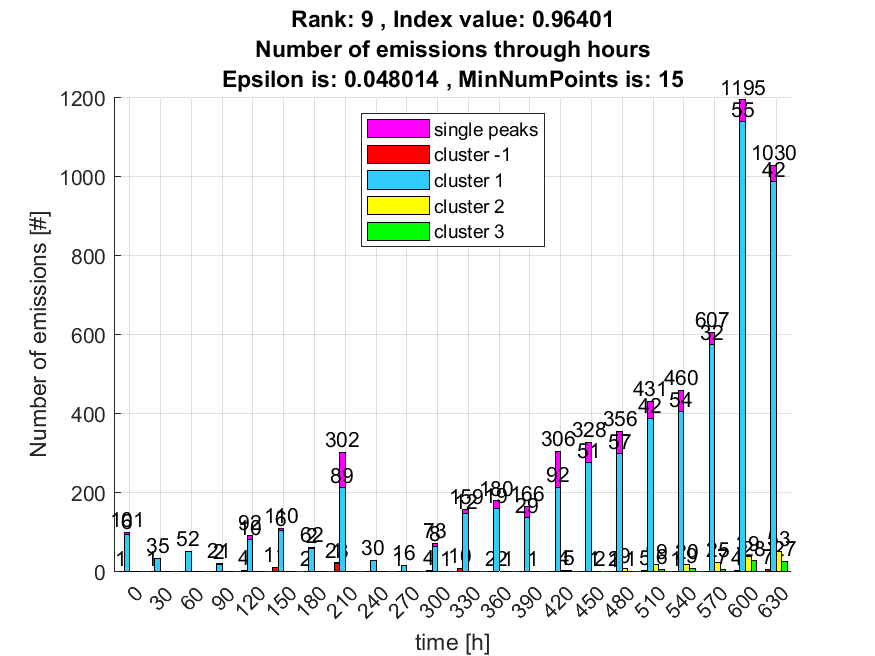
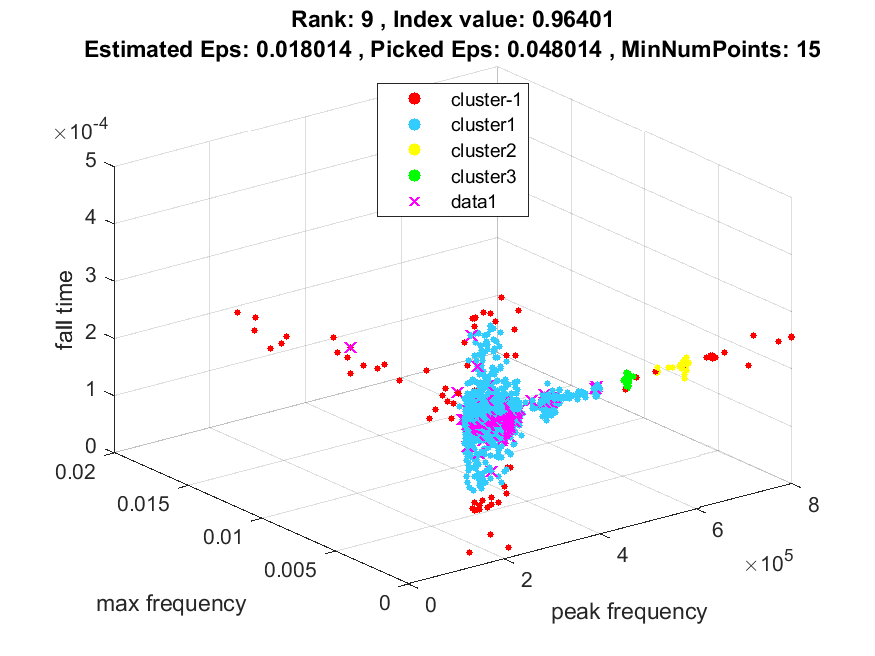
**DBCV index value: 0.96401**

**cluster -1 had 0.32787% single peaks**

**cluster 1 had 99.6721% single peaks**

**cluster 2 had 0% single peaks**

**cluster 3 had 0% single peaks**



**10. result: (pp1, peak\_frequency, max\_frequency)**

**Estimated Epsilon: 0.015981, Picked Epsilon: 0.045981, MinNumPoints: 15**

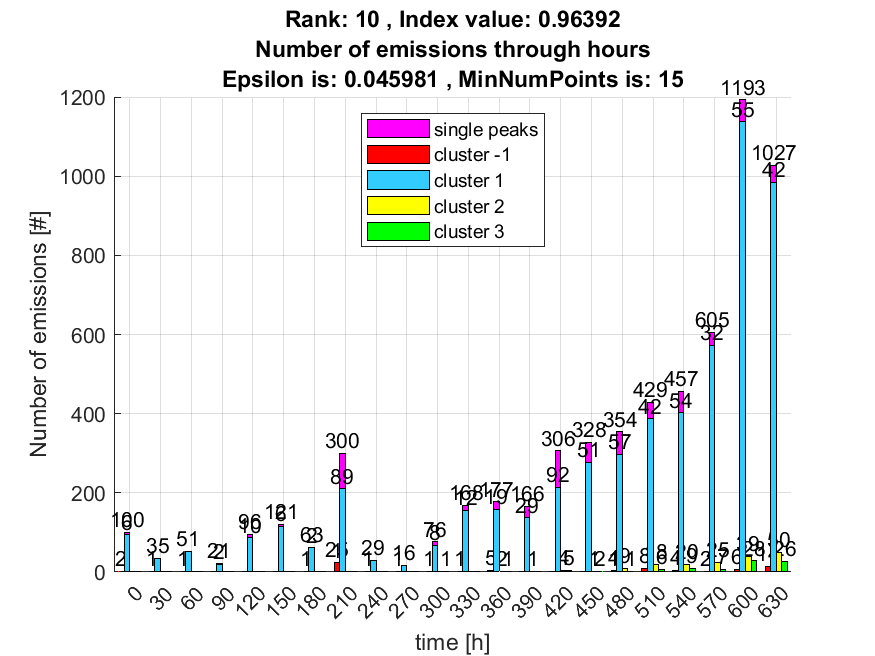
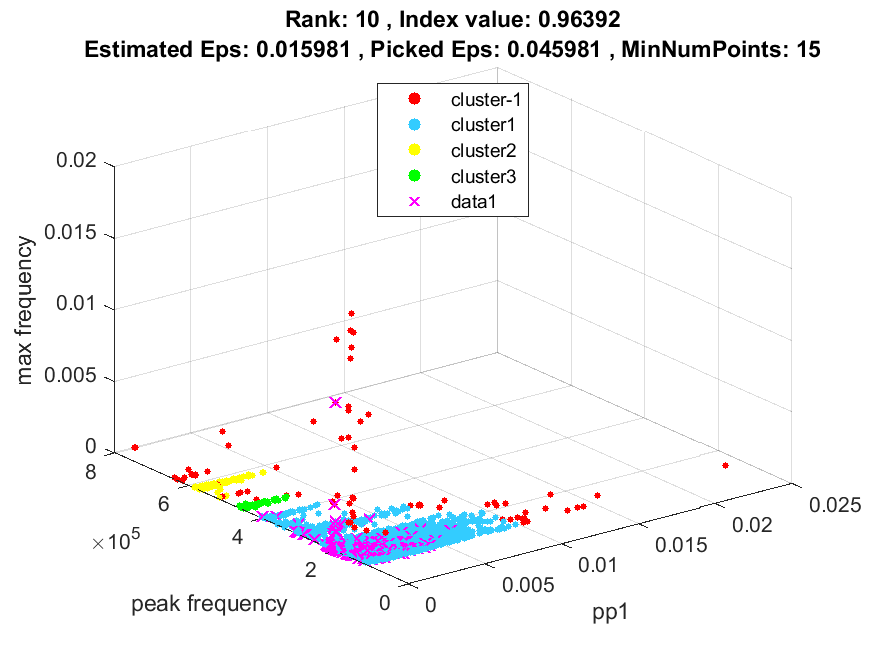
**DBCV index value: 0.96392**

**cluster -1 had 0.32787% single peaks**

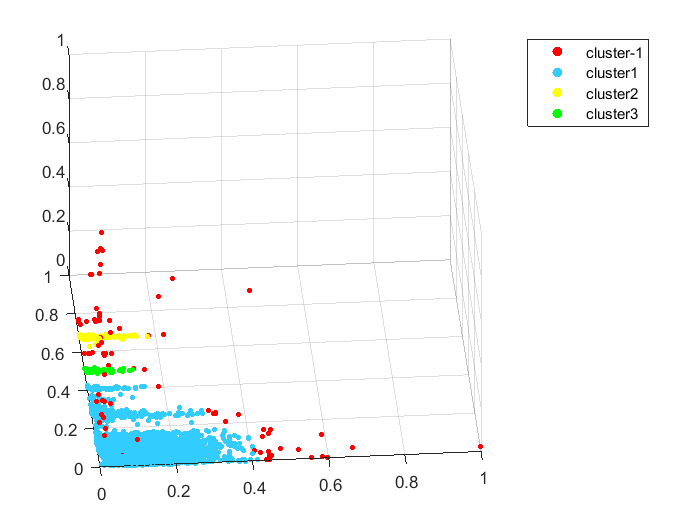
**cluster 1 had 99.6721% single peaks**

**cluster 2 had 0% single peaks**

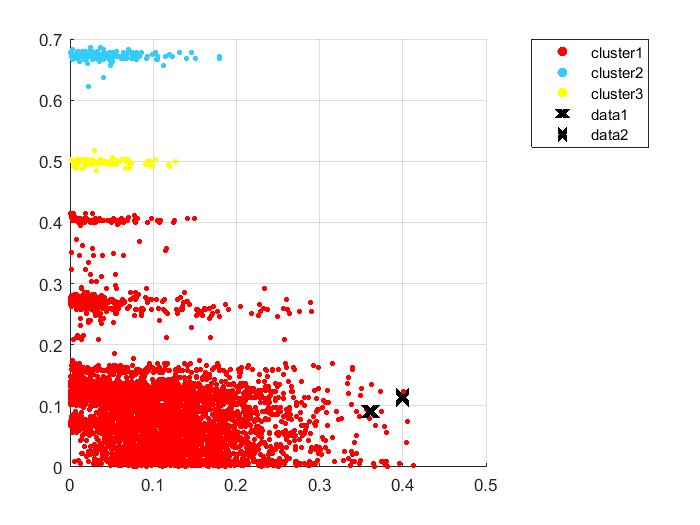
**cluster 3 had 0% single peaks**



**Wrong cluster:**



**DSC of 1 cluster:**



**DSPC: DSC**

**0 0.2560 0.0820 0.002014 0.0008492 0.000334397**

**0.2560 0 0.15716**

**0.0820 0.15716 0**

**min\_DSPC:**

**0.0820 0.15716 0.0820**

**V\_index:**

**0.97543 0.9945 0.9959**

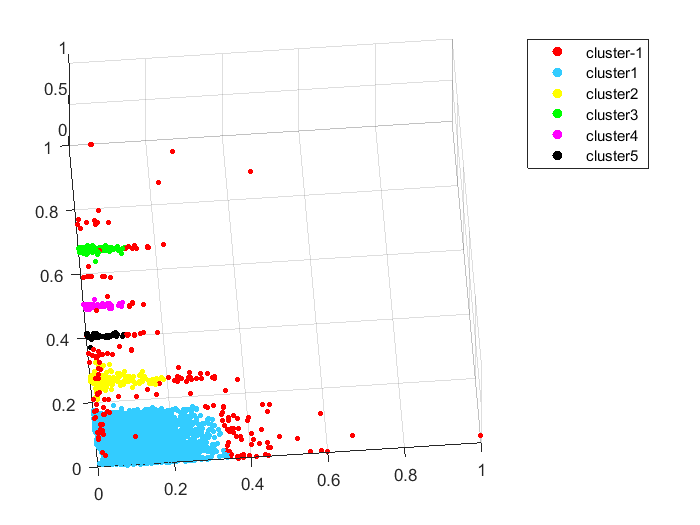
**Length:**

**6138 165 86**

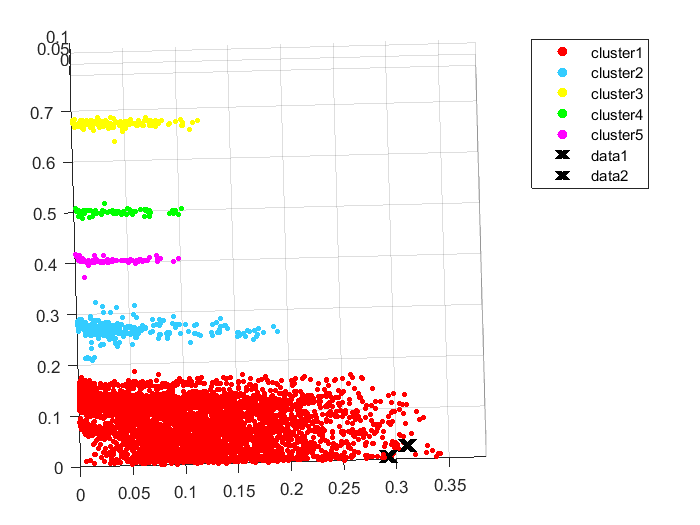
**DBCV\_index:**

**0.9638**

**Good cluster:**



**DSC of 1 cluster:**



**DSPC:**

**0 0.04575 0.4986 0.32573 0.23428**

**0.045758 0 0.36072 0.18876 0.097148**

**0.49860 0.36072 0 0.1571 0.25335**

**0.3257 0.18876 0.157164 0 0.07740**

**0.23428 0.097148 0.2533 0.07740 0**

**DSC:**

**0.000735767 0.00056122 0.00014739 0.0003245 0.0002142**

**Min\_DSPC:**

**V\_index:**

**0.98392 0.98773 0.9990 0.99580 0.9972**

**Length:**

**5636 339 157 82 70**

**DBCV\_index:**

**0.9563**

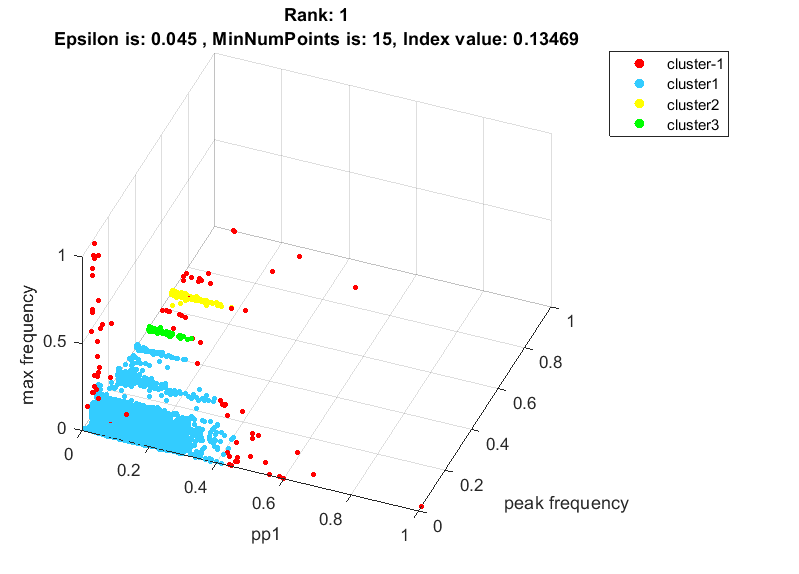
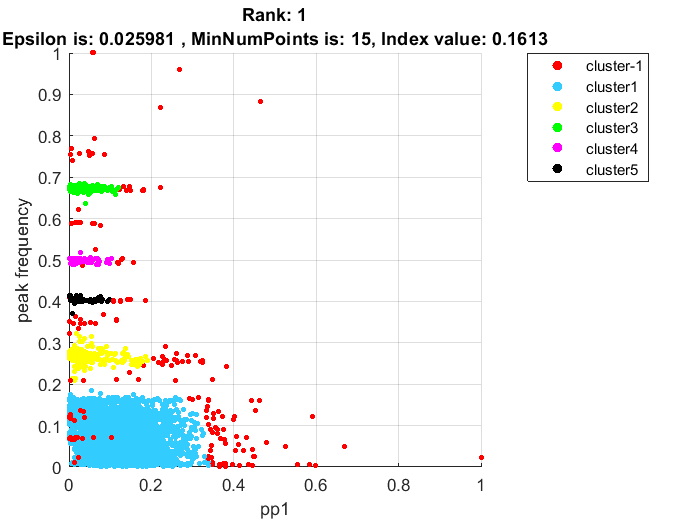
max\_density = max([min\_DSPC,DSC(clust\_i)]);

% If a cluster has better density compactness than density separation we obtain positive values of the validity index.

% If the density inside a cluster is lower than the density that separates it from other clusters, the index is negative.

VC\_index(clust\_i) = (min\_DSPC-DSC(clust\_i))/max\_density;

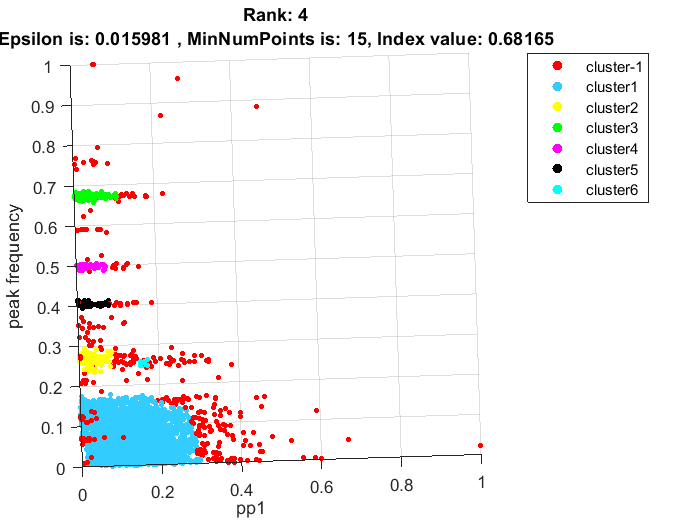
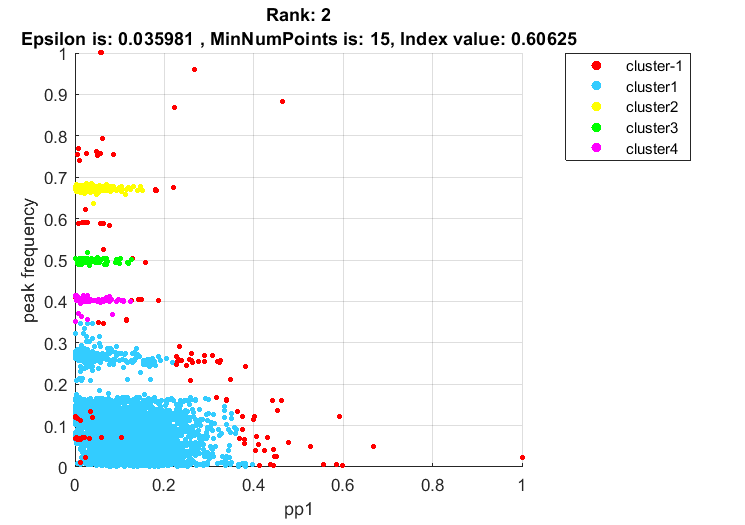
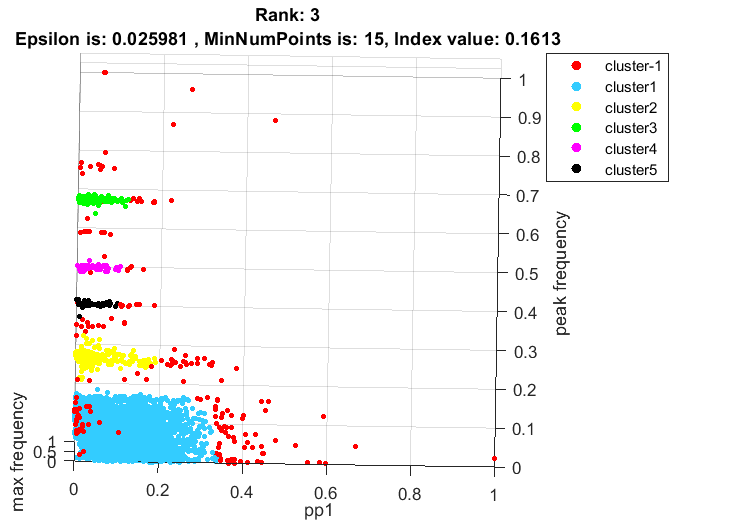
**Opet si ostavio squared euclidean distance, promjena izgleda ovako:**



**Indeks je manji , ali bolje bira!**

**Novi problem: rank 3 i 2 imaju isti DSC prvog clustera zbog velikog core distance-a točke na rubu custera koja oboma ulazi u min span tree!**

**Mjere DBCV nije prikladna:** [automl\_camera.pdf (in2p3.fr)](https://indico.ijclab.in2p3.fr/event/2914/contributions/6483/subcontributions/177/attachments/6054/7179/automl_camera.pdf)



**Za loši rezultat:**

**DSC:**

**0.0365 0.0093 0.0110 0.0116**

**DSPC:**

**0 0.3997 0.2298 0.1430**

**0.3997 0 0.1698 0.2568**

**0.2298 0.1698 0 0.0882**

**0.1430 0.2568 0.0882 0**

**min\_DSPC:**

**0.0882**

**VC\_index:**

**0.7447 0.9454 0.8753 0.8690**

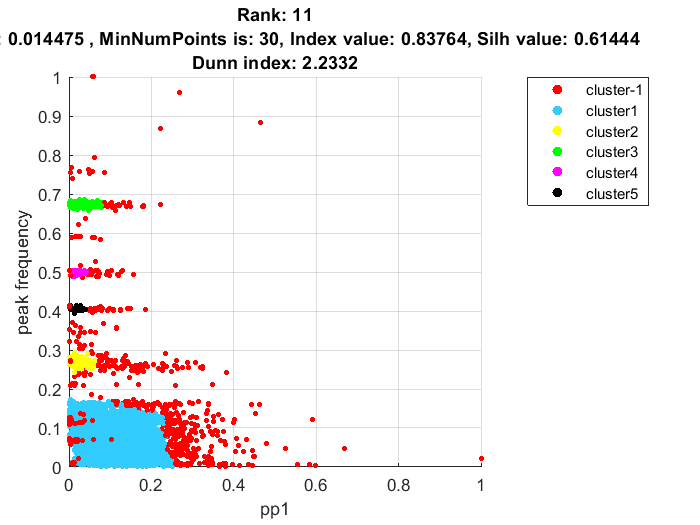
**DB\_index:**

**0.7387**

**Za dobar rezultat:**

**Novi indeks baziran na optornosti od spajanja clustera koji imaju rupe:**

**Problem uzimanja manjih gušćih clustera nad većim manje gušćim:**



**DSC:**

**0.0065 0.0047 0.0287 0.0049 0.0052**

**DSPC:**

**0 0.0961 0.5053 0.3350 0.2386**

**0.0961 0 0.3838 0.2130 0.1156**

**0.5053 0.3838 0 0.1657 0.2632**

**0.3350 0.2130 0.1657 0 0.0935**

**0.2386 0.1156 0.2632 0.0935 0**

**VC\_index:**

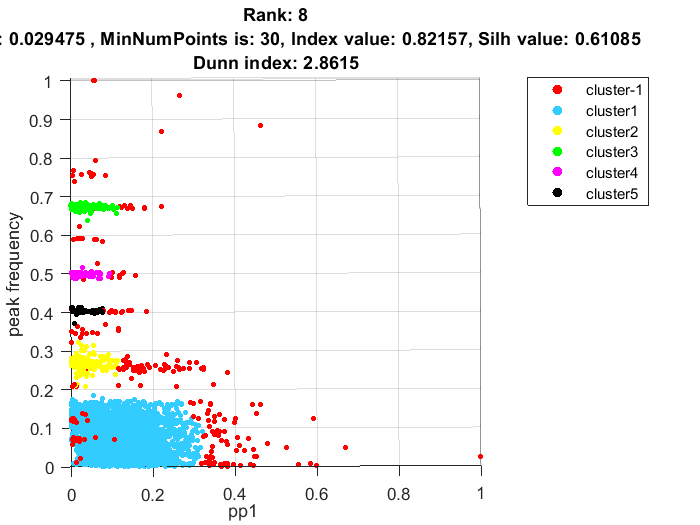
**0.9319 0.9516 0.8268 0.9474 0.9446**

**clust\_i\_len:**

**5343 247 143 49 44**

**DB\_index:**

**0.8376**



**DSC:**

**0.0095 0.0096 0.0057 0.0078 0.0069**

**DSPC:**

**0 0.0631 0.5049 0.3327 0.2386**

**0.0631 0 0.3836 0.2110 0.1157**

**0.5049 0.3836 0 0.1657 0.2563**

**0.3327 0.2110 0.1657 0 0.0842**

**0.2386 0.1157 0.2563 0.0842 0**

**VC\_index:**

**0.8493 0.8486 0.9653 0.9074 0.9182**

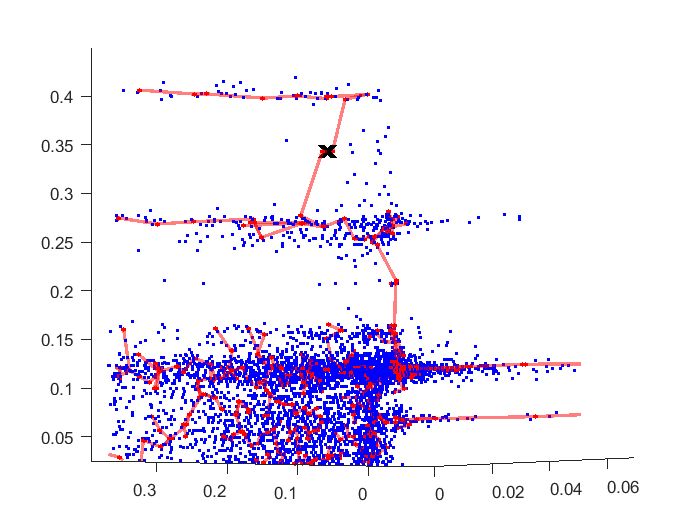
**clust\_i\_len:**

**5621 304 156 80 67**

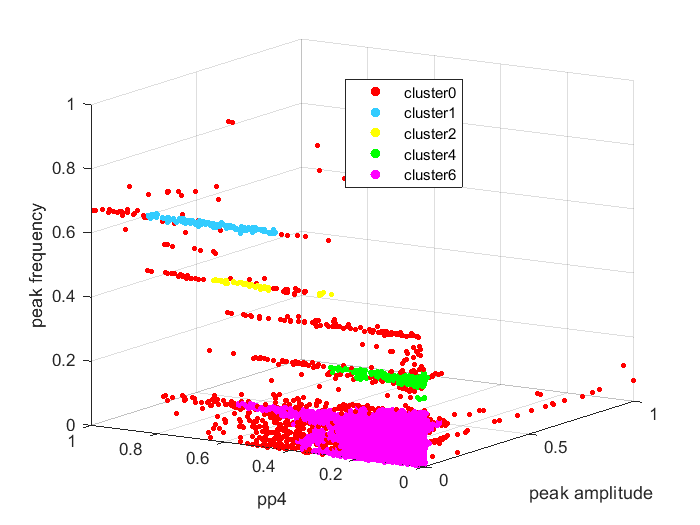
**DB\_index:**

**0.8216**

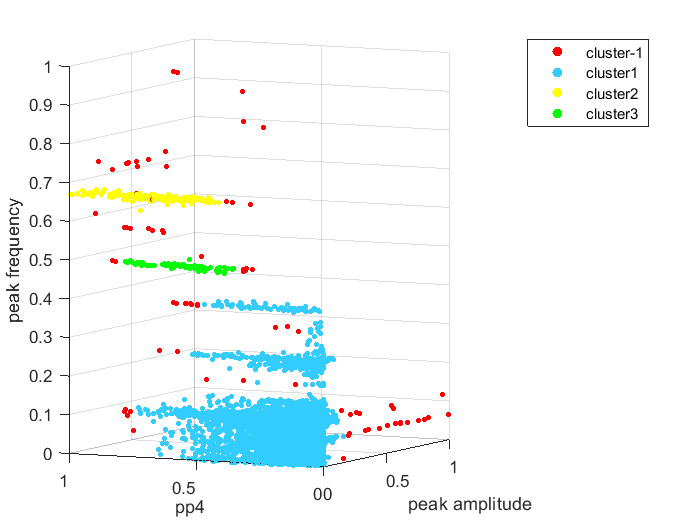
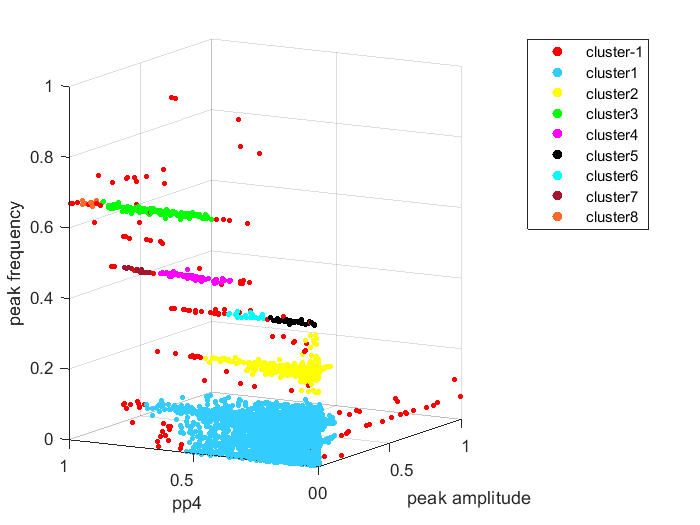
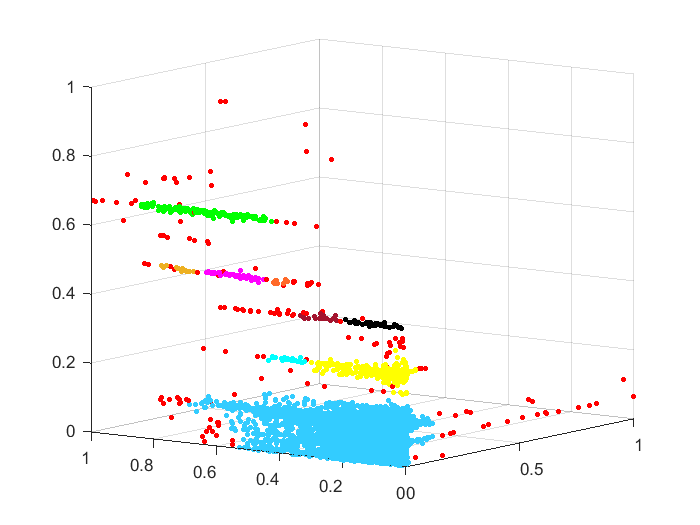
**Kod određivanja rupa ako je više internih čvorova točke rupa onda je potrebno pratit stablo do prve točke veće gustoće gdje je 70% veće gustoće. Prati svaki izlaz iz zadanih start i end čvorova i ako uspiješ iz obaju točka nekako pomoću susjeda doći do točke veće gustoće za 70% onda je zadana točka rupa. Ako je na rubu točka onda će možda imati zadovoljen uvijet ,ali zato postoji dodan uvijet da mora odvajati MST na barem više stabla od 15 ili više točaka.**



**HDBSCAN najbolji cluster :**



**OPTICS CLUSTER:**



**Je li i redu s rupama?**

