Passive RADAR: Deinterleaving and Clustering unknown RADAR pulses

Manon Mottier

01 february 2021

Atos

Content

I. Introduction

II. Data description

III. Proposed approach

IV. Results

V. Conclusion

Introduction

- **Context**

  - Use and development of machine learning tools in the field of Electronic Warfare specifically for passive RADAR intelligence

  - Passive RADAR(s)
    - Emit no electromagnetic waves
    - Only receivers
    - Detect, track and identify RADAR(s)

  - Improvement of RADAR signal processing thanks to ML approaches

  - RADAR signals deinterlacing includes techniques capable of separating, characterizing and identifying all transmitters present in the signal

CentraleSupélec

université PARIS-SACLAY

L2S Laboratoire Signaux & Systèmes

Introduction

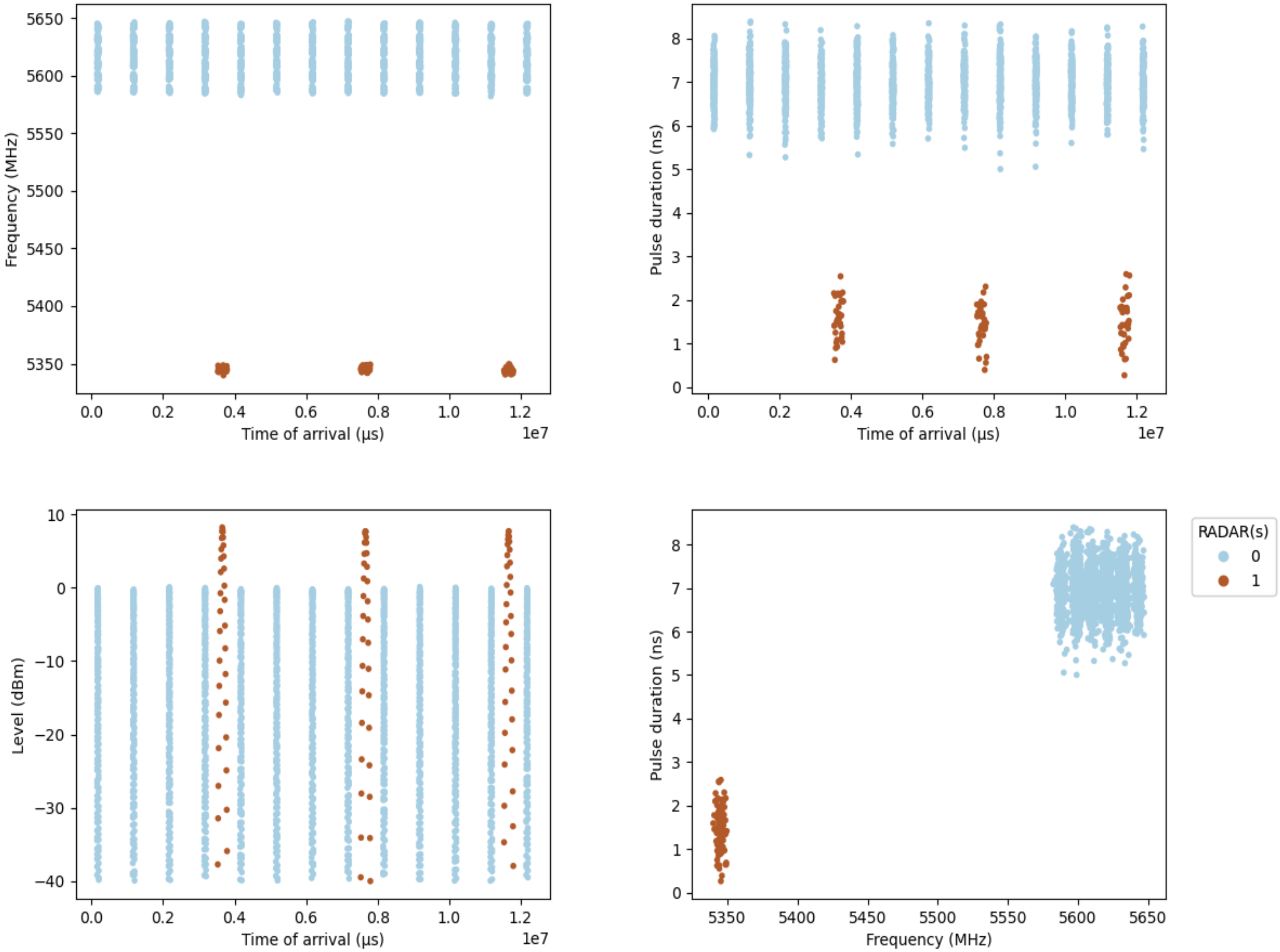
- **Main challenges: data complexity**

  - Some RADARs do not continuously transmit
  - Noisy data, e.g., sensors are not optimal, parasites, weak SNR…
  - Identifiability issues: Different RADARs may have very close, almost identical, characteristics (e.g., in airports or harbours …)
  - Volume and heterogeneity: Signals of different sizes (from 50 observations to more than 1 000 000 observations) that may contain up than 10 RADAR(s)

Data Description

- Realistic simulator since real labeled data are not easily accessible
- Each dataset is composed of RADAR pulses, described by 4 features: Frequency (F), Pulse Width (PW), Level (L) and Time of Arrival (TOA)
- Example of a simple signal containing 2 RADAR(s)
  - ➢ They emit on very different frequency bands and in a regular way

RADAR	Obs	%
0	34 684	96%
1	1 578	4%
	36 262	



Proposed approach

<u>Methodology</u> : Implementing a two-step strategy for signal deinterlacing

Step 1: Clustering

- Cluster each observation of this signal under the constraint of not merging different radars in one cluster

Step 2: Cluster fusion

- Hierarchical agglomerative clustering combined with optimal transport distance

Proposed approach

Step 1: Clustering with HDBSCAN

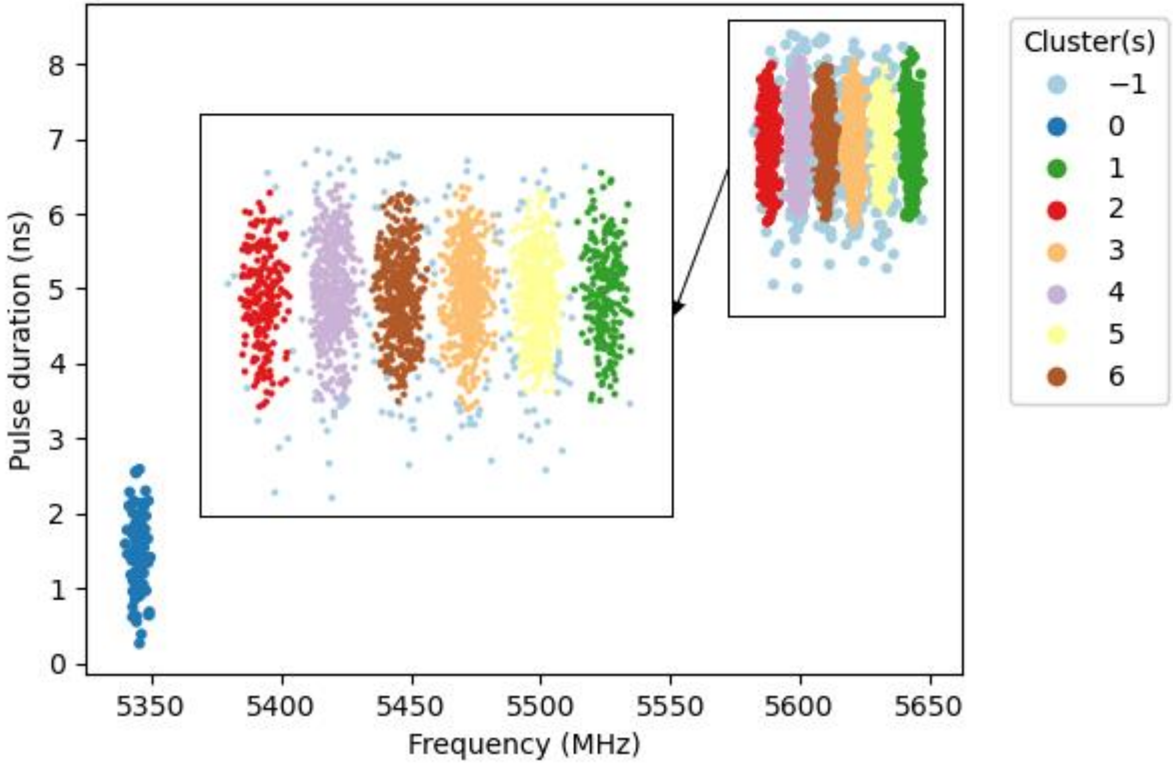
- Application of clustering algorithm on two features: Frequency and Pulse Width
  - Why HDBSCAN ?
    - Unsupervised algorithm that allows us to search for clusters of varying densities
    - Few parameters to optimize
- Over-estimation of the number of clusters
  - Allows to receive RADARs that do not transmit continuously
  - Each cluster **must not** contain the observations of different RADARs

Proposed approach

Step 1: Clustering with HDBSCAN

Clusters label	Obs	%
-1 (outliers)	2 417	7%
0	1 578	4%
1	3 351	9%
2	3 235	9%
3	6 577	18%
4	6 540	18%
5	6 263	17%
6	6 301	17%

Clusters repartition



Results of HDBSCAN clustering in the F-PW plane.

Proposed approach

Step 2: Cluster fusion

Development of a hierarchical clustering algorithm using optimal transport distances

1. Dimensionality reduction: each cluster is represented by a histogram of its TOA
2. Compute distances between cluster histograms using optimal transport
3. Aggregation of the 2 closest clusters i.e., the 2 clusters with the lowest transport costs
4. Computation of new costs between old clusters and the new ones
5. Repeat

Proposed approach

- Transporting distribution to another has a cost:

$$C(\mathbf{P}) = \sum_{n=1}^{N} \sum_{m=1}^{M} C_{nm} P_{nm} = \langle \mathbf{C}, \mathbf{P} \rangle$$

$C(P)$: Total cost

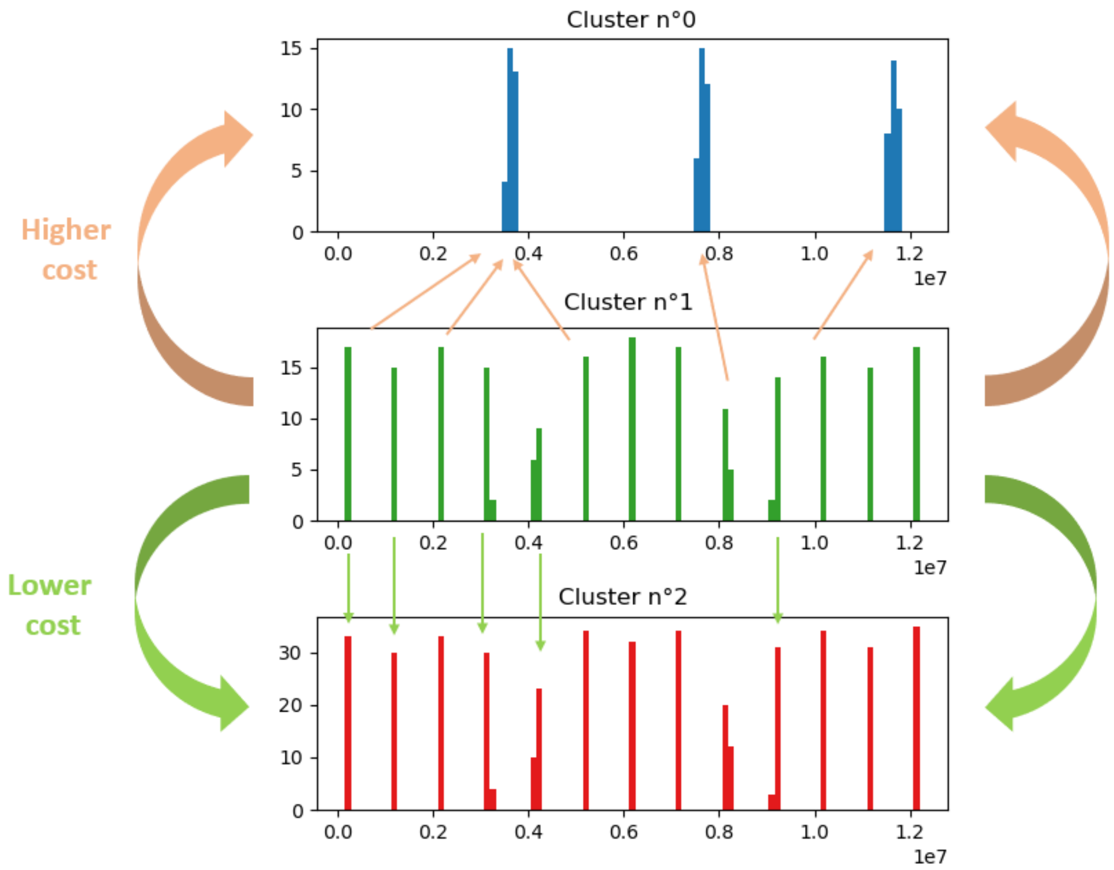
C_{nm} : cost of transporting unit mass from x_n to y_m (transport cost)

P_{nm} : quantity of mass taken from x_n to y_m (transport plan)

- The optimal transport plan is a simple minimization problem:

$$\mathbf{P}^{\star} = \underset{\mathbf{P} \in \mathbf{R}_{+}^{N \times M}}{\operatorname{argmin}} \langle \mathbf{C}, \mathbf{P} \rangle \text{ such that } \mathbf{P}\mathbf{1}_M = \mathbf{a}, \mathbf{P}^T \mathbf{1}_N = \mathbf{b}^T$$

Under this condition which confirms that the transport plan is well adapted to probability densities



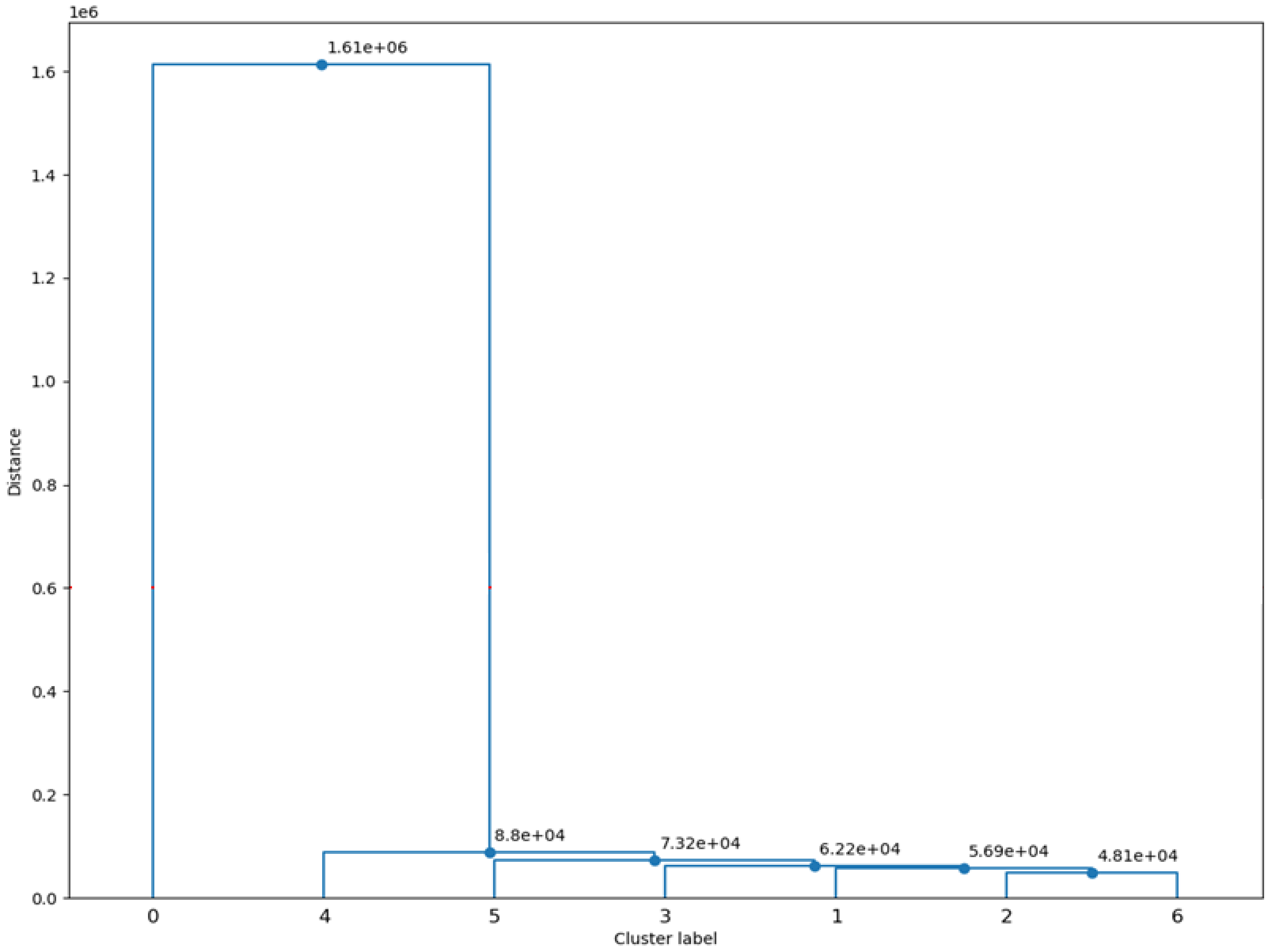
Histogram = proportion of observation distributed in each bin for a cluster

Proposed approach

Step 2: Cluster fusion

- Construction of a decision rule using several unsupervised metrics to cut the dendrogram
  - Silhouette score
  - Calinski-Harabasz score
  - Davies-Bouldin score

-> Majority rule or higher number of clusters



Proposed approach

	Higher score			Lower score
Number of clusters	Silhouette score (euclidean distance)	Silhouette score (cosine distance)	Calinski-Harabasz score	Davies-Bouldin score
2	0,93	0,99	25 425	0,06
3	0,32	0,09	17 157	0,49
4	0,19	-0,03	14 206	0,68
5	0,33	-0,03	11 183	1,35
6	0,56	-0,03	42 746	1,36

Proposed approach

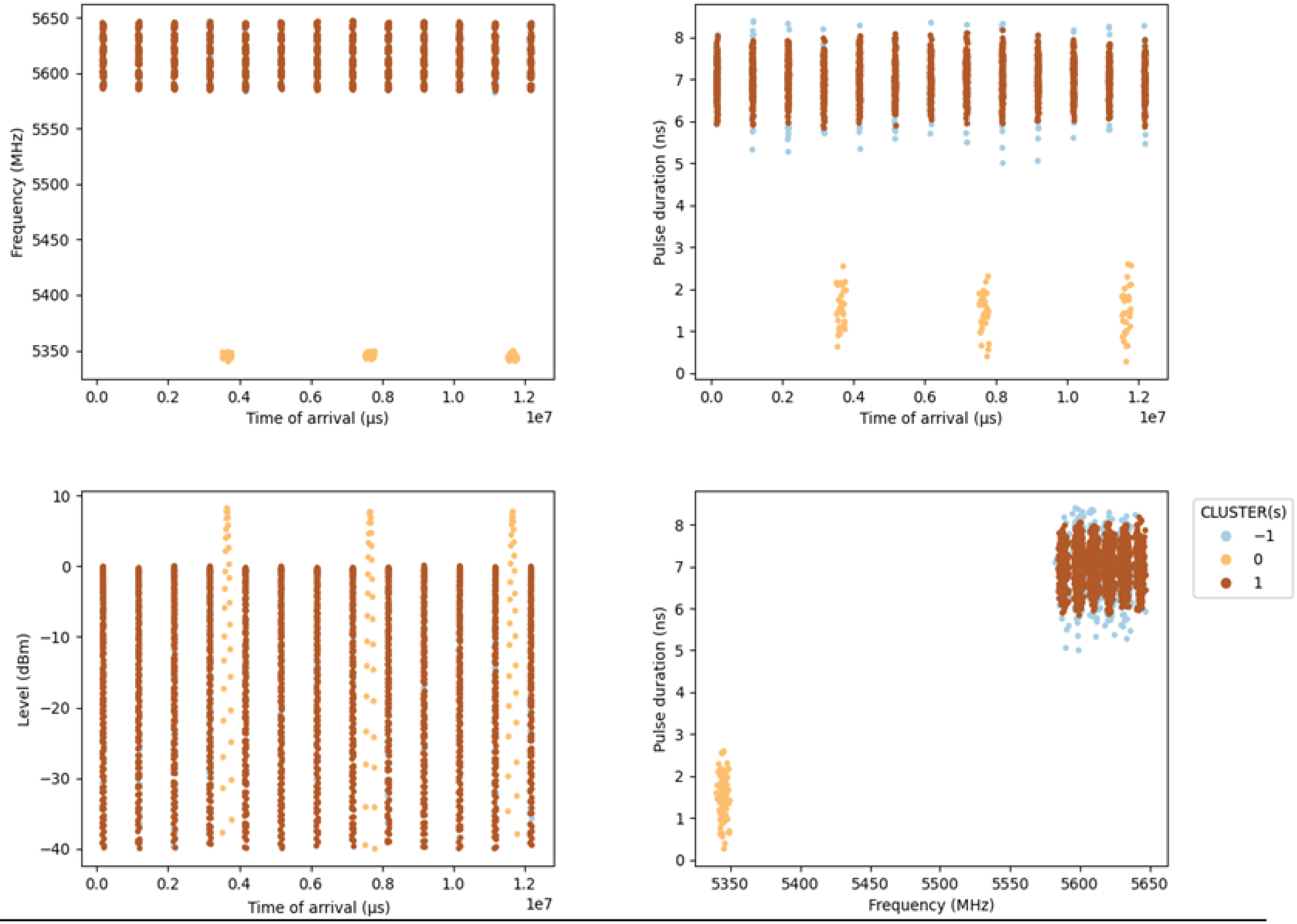
- Perfect clustering and grouping

- Outlier class

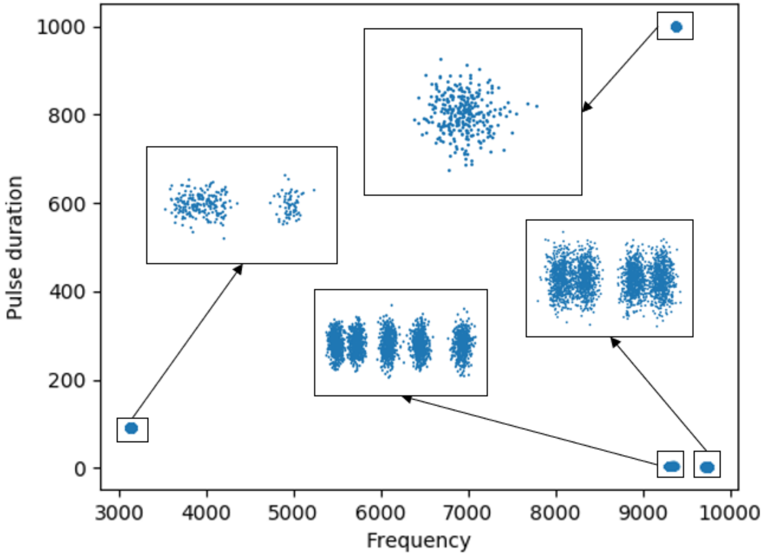
RADAR	Obs	%
-1	2 417	7%
0	1 578	4%
1	32 267	89%

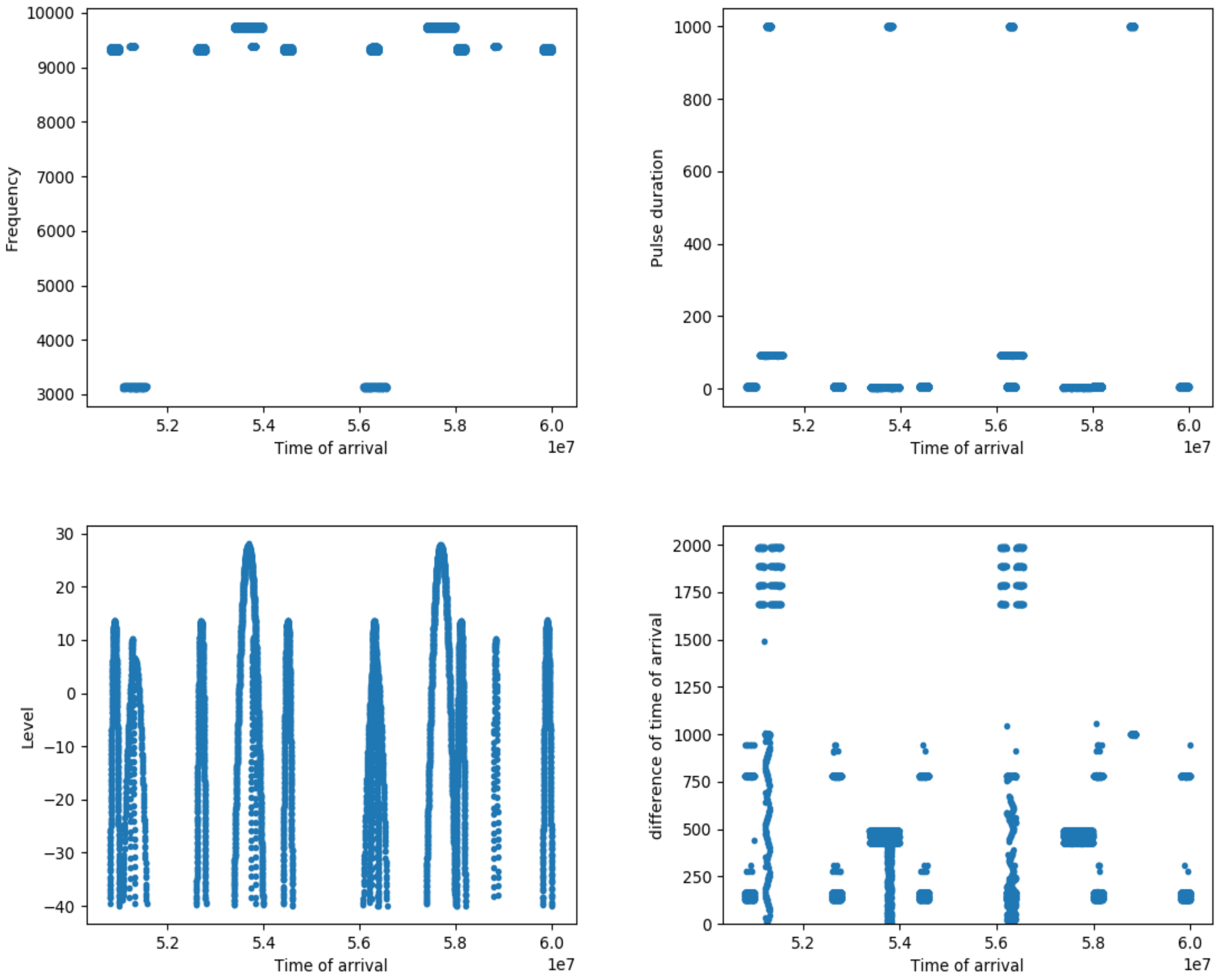
- 7%

Emmiters repartition



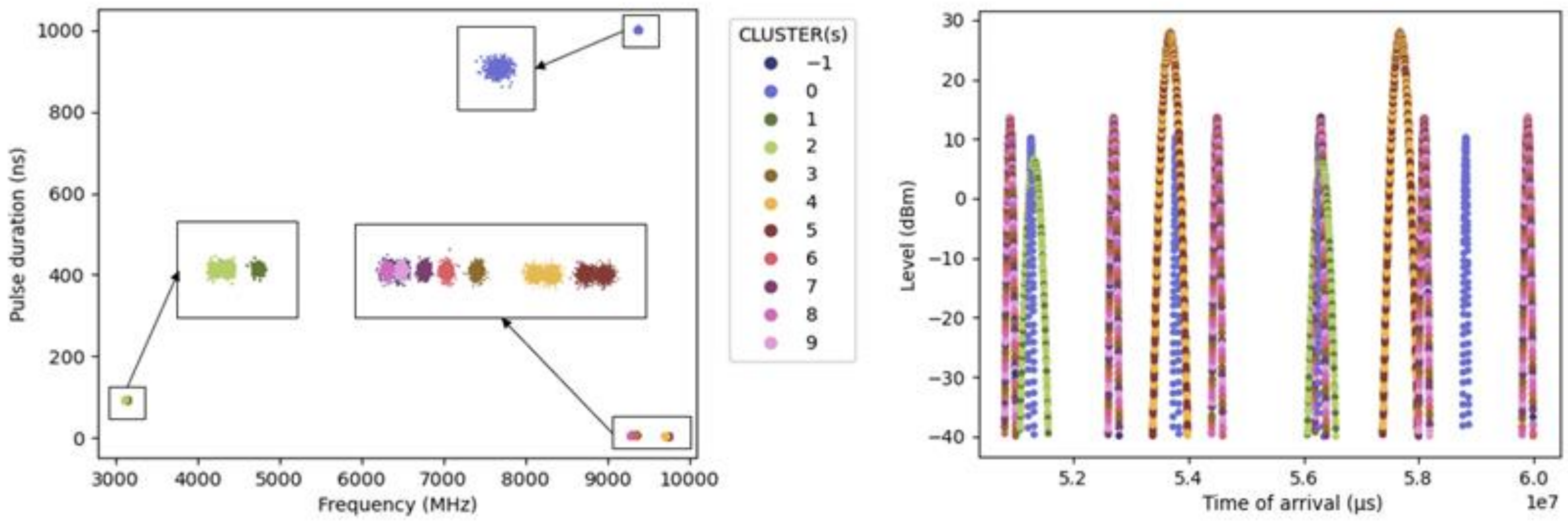
Results on more realistic data





Results on more realistic data

Clusters	Obs	%
-1	954	1%
0	8 199	5%
1	2 703	2%
2	8 094	5%
3	13 907	9%
4	32 704	21%
5	32 726	21%
6	13 892	9%
7	13 891	9%
8	13 502	9%
9	13 368	9%

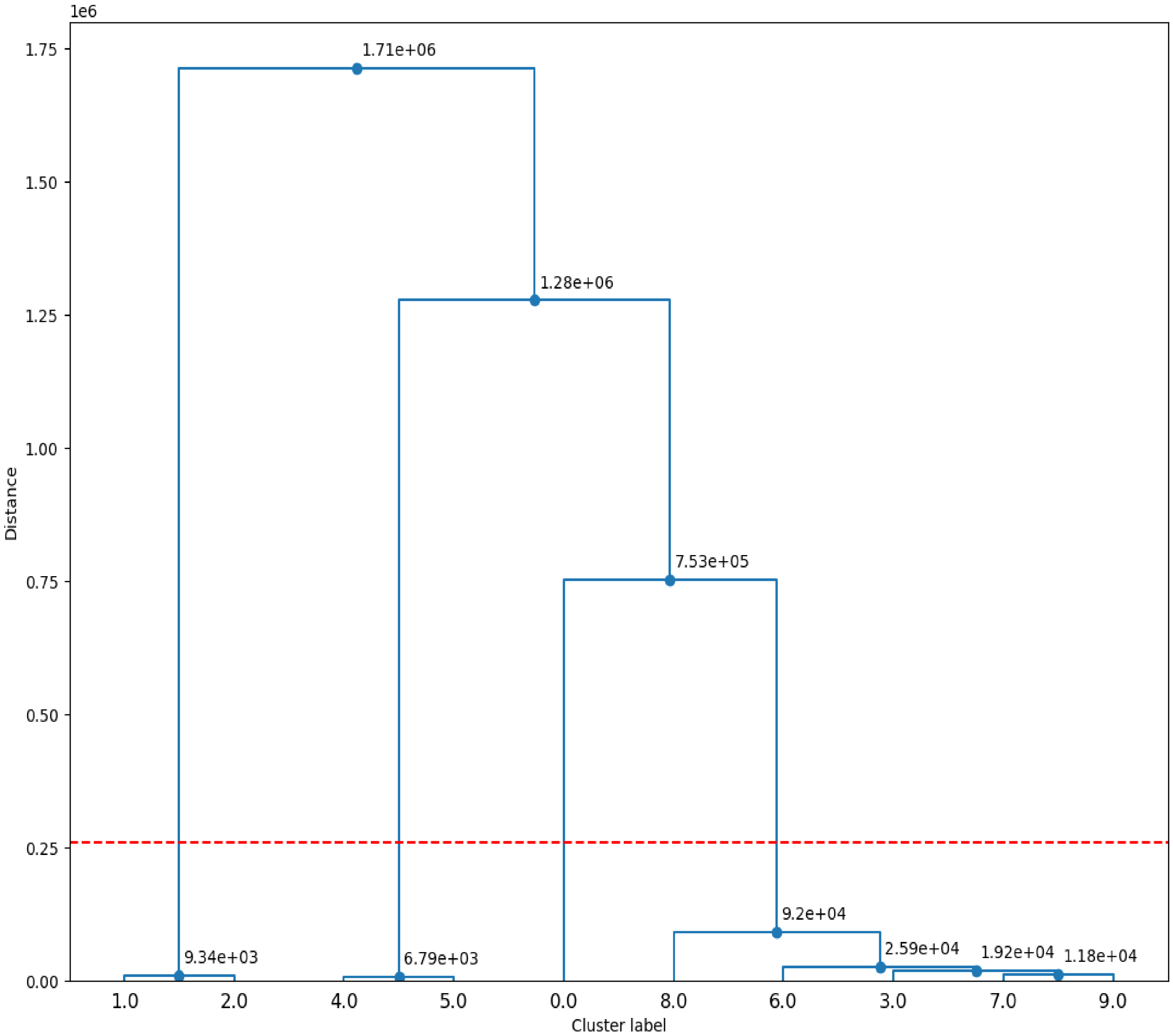


Results of HDBSCAN clustering

Results on more realistic data

Higher score — Lower score

Number of clusters	Silhouette score (euclidean)	Silhouette score (cosine)	Calinski-Harabasz score	Davies-Bouldin score
2	0,96	0,20	225 665	0,044
3	0,79	-0,10	199 414	0,323
4	0,95	0,89	22 699 186	0,041
5	0,66	0,42	24 383 075	0,229
6	0,56	0,42	19 966 460	0,985
7	0,79	0,41	40 515 795	0,248
8	0,89	0,40	40 049 740	0,169
9	0,87	0,35	39 582 412	0,203

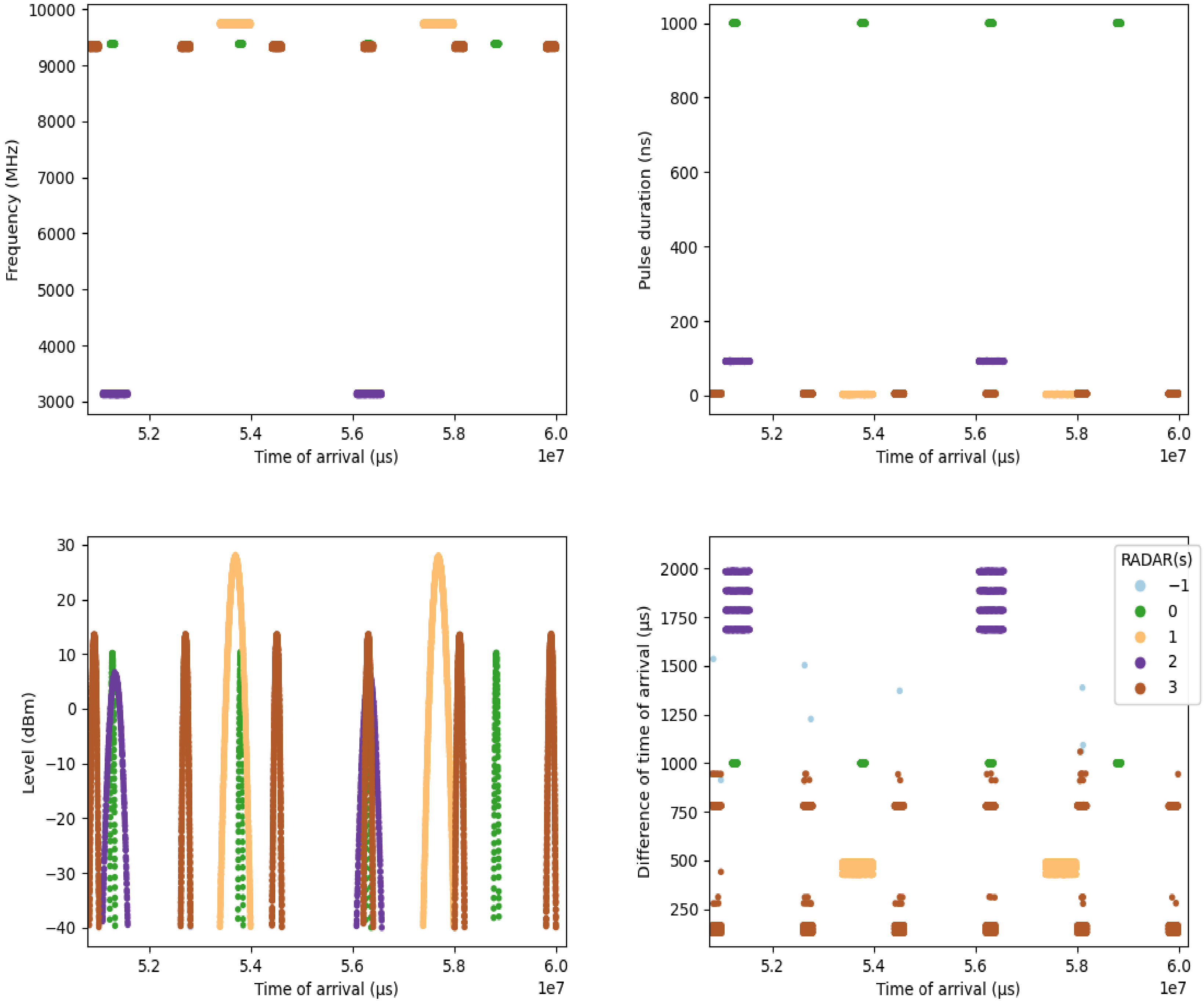


Results on more realistic data

RADAR(s)	Obs	%
0	8 199	5,3%
1	65 430	42,5%
2	10 797	7%
3	68 560	44,5%
-1 (outliers)	954	0,6%

- 0,01%

- 1,4%



Conclusion

- Works on moderately complex cases
- Improvements in progress to differentiate between similar RADAR(s) in Frequency and Pulse Width
- Next objectives:
  - Cluster representativeness: size, distribution, etc.
  - RADAR with a strong frequency sweep
  - Construction of a decision model to cut the dendrogram

Thank you for your attention