

Plénière – 2 ans de thèse

Manon Mottier

24/06/2021

Trusted partner for your Digital Journey

Atos

Plan

I. Introduction

II. Approche proposée

III. Applications

IV. Conclusions



1

Introduction

Introduction

Contexte :

- Utilisation et développement d'outils d'apprentissage automatique dans le domaine de la guerre électronique, plus particulièrement pour le RADAR passif
- Amélioration du traitement des signaux RADAR grâce aux approches de Machine Learning
- Le dé-entrelacement des signaux RADAR comprend les techniques capables de :
 - Séparer
 - Caractériser
 - Identifier tous les émetteurs présents dans le signal



Introduction

Principaux challenges :

- Certains RADARs ne transmettent pas de façon continue
- Données bruitées
- Problèmes d'identifiabilité : Différents RADARs peuvent avoir des caractéristiques très proches / identiques (par exemple, dans les aéroports ou les ports...).
- Volume et hétérogénéité : Signaux de différentes tailles (de 50 impulsions à plus de 1 000 000 d'impulsions) pouvant contenir jusqu'à 10 RADAR(s).

Introduction

Description des données :

- Simulation des données
- Chaque jeu de données est composé d'impulsions RADAR, décrites par seulement 4 caractéristiques :
 - Frequency (F),
 - Pulse Width (PW)
 - Level (L)
 - Time of Arrival (TOA)
- Pas d'utilisation de la dTOA
- Pas de supposition sur la forme d'onde
- Paramètre non supervisé mais évaluation des méthodes sur des ensembles de données étiquetées.

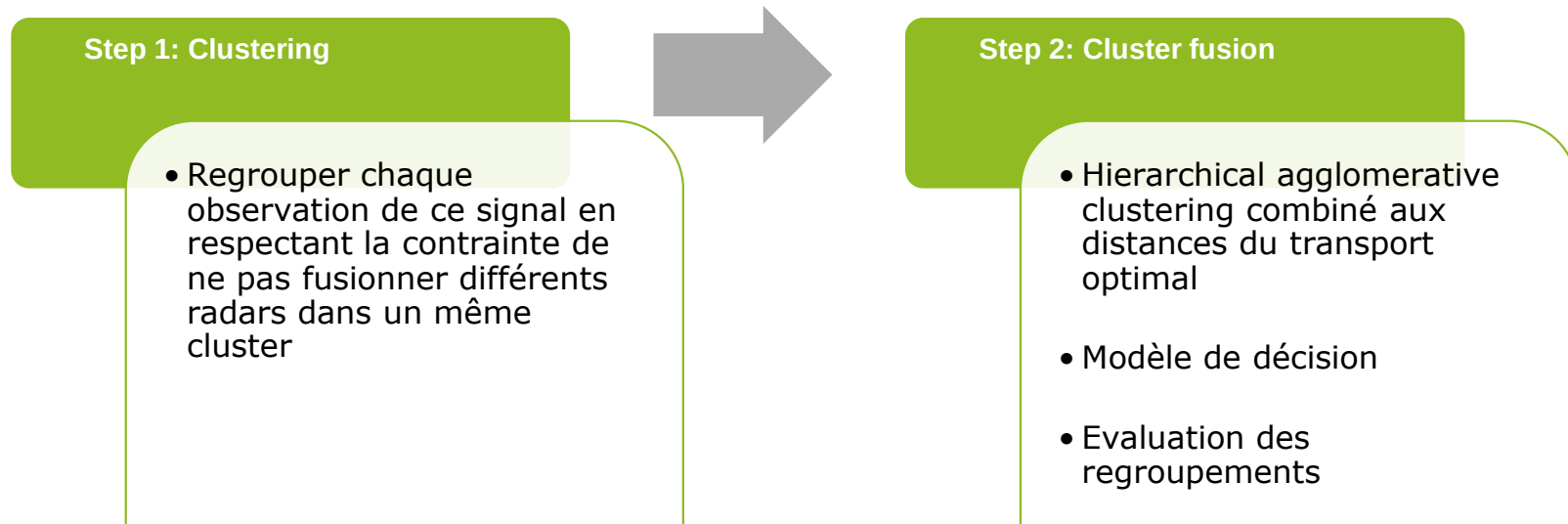
A large, white, stylized outline of the number '2' is positioned on the left side of the slide. The background is a dark blue gradient with a faint, glowing network of white lines and dots, resembling a molecular or digital structure, that extends across the right side of the slide.

2

Approche proposée

Approche proposée

Méthodologie : Implémentation d'une stratégie en 2 étapes



Approche proposée

Étape 1 : Clustering avec HDBSCAN

- Application de l'algorithme de clustering sur deux features : la fréquence et la durée d'impulsion.
- Pourquoi HDBSCAN ?
 - Algorithme non supervisé qui permet de rechercher des clusters de densité(s) variable(s).
 - Peu de paramètres à optimiser
- Sur-estimation du nombre de clusters
 - Permet de recevoir des RADARs qui n'émettent pas en continu
 - Chaque cluster ne doit pas contenir les observations de différents RADARs

Approche proposée

Étape 2 : Regroupement des clusters

Développement d'un algorithme de clustering hiérarchique utilisant les distances de transport optimales :

- 1) Réduction de la dimensionnalité : chaque cluster est représenté par un histogramme de sa TOA,
- 2) Calcul des distances entre les histogrammes des clusters en utilisant le transport optimal.
- 3) Agrégation des 2 clusters les plus proches, c'est-à-dire les 2 clusters ayant les coûts de transport les plus bas.
- 4) Calcul des nouveaux coûts entre les anciens clusters et les nouveaux.
- 5) Répétition des étapes 2 à 4 jusqu'à ce qu'un seul cluster soit obtenu.

➤ **Visualisation des résultats sous forme de dendrogramme**

Approche proposée

Étape 2 : Regroupement des clusters

Construction d'un modèle de décision utilisant plusieurs métriques non supervisées pour couper le dendrogramme et arrêter le regroupement des clusters :

- Score de silhouette : mesure l'homogénéité du cluster en utilisant la différence entre tous les points du même cluster et la distance aux points des autres clusters.
 - Utilisation de 2 distances différentes pour plus de robustesse
- Score de Calinski-Harabasz : mesure la proportion de la variance inter-groupe par rapport à la variance intra-groupe (également appelé critère du ratio de variance).
- Score de Davies-Bouldine : mesure l'homogénéité en utilisant la distance entre un point et son centroïde et la distance entre les autres centroïdes.
 - **Règle de la majorité**

Approche proposée

Evaluation : Qualité de la fusion

Construction d'un modèle d'évaluation pour mesurer la qualité de la fusion sur la base de 3 critères existants :

- Homogénéité : mesure si tous les points d'un cluster donné appartiennent à la même classe (varie entre 0 et 1, 1 indiquant une homogénéité parfaite au sein du cluster).
- Complétude : mesure si tous les points d'une classe donnée sont regroupés dans un seul cluster (varie entre 0 et 1, 1 indiquant qu'il n'y a pas de dispersion des données de la classe entre plusieurs clusters).
- Taux de non-classification : mesure la quantité de données perdues au cours du processus (ie outliers).

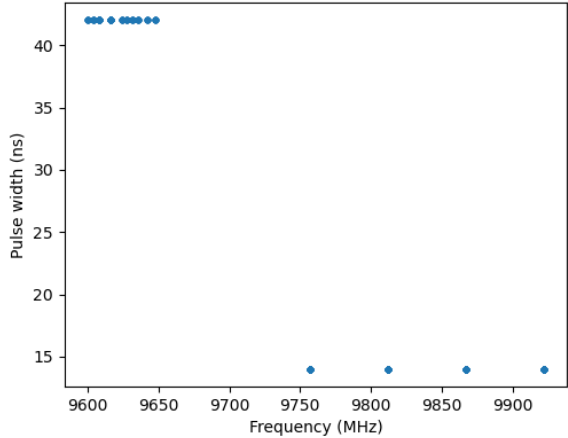
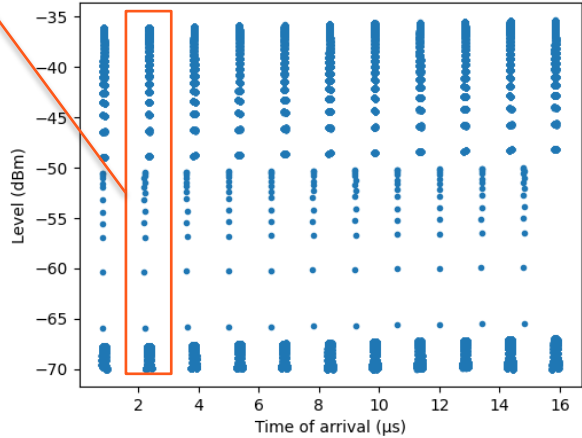
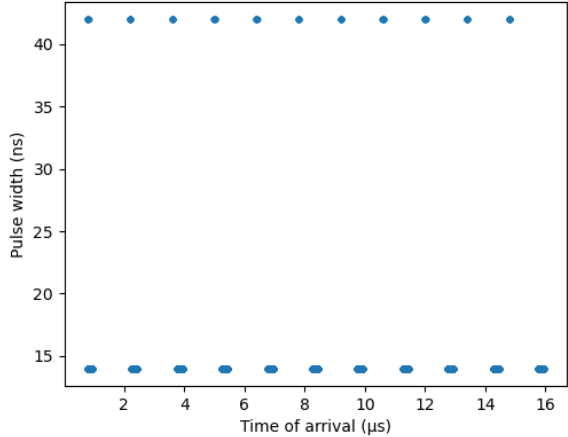
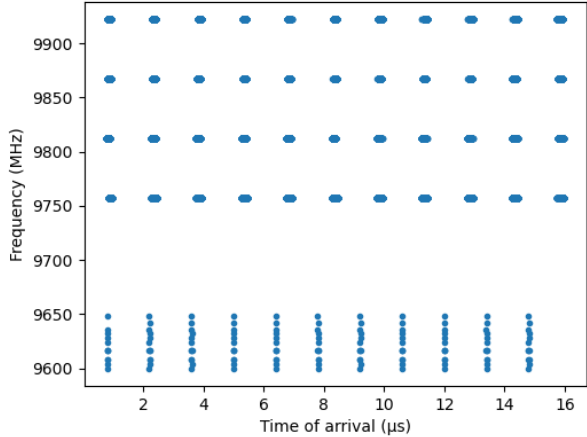
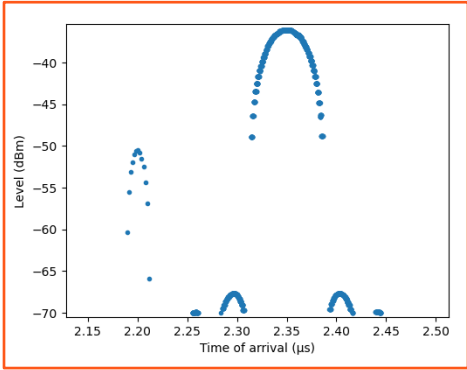


3

Applications

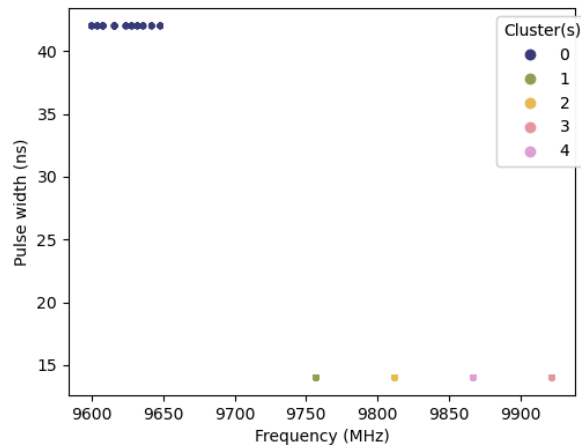
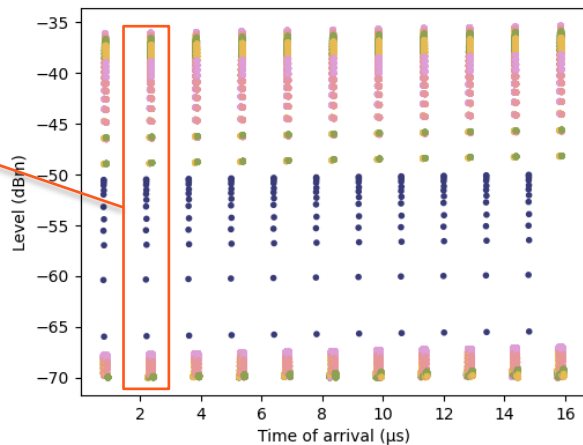
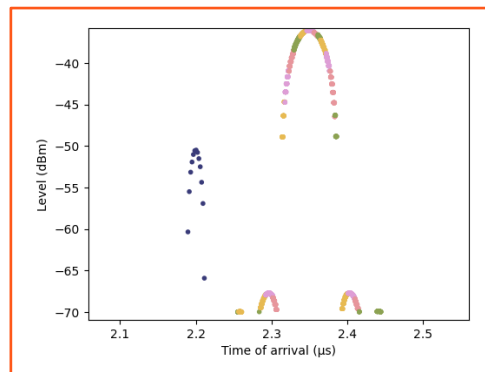
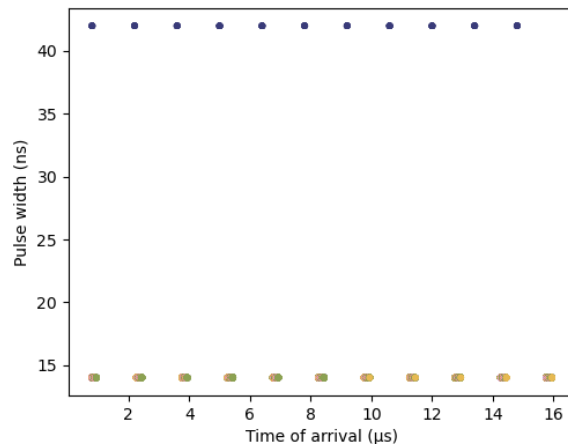
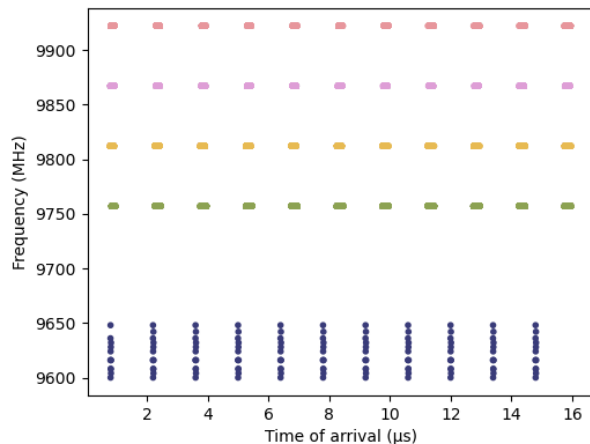
CAS D'APPLICATION 1

Données brutes



Clustering HDBSCAN

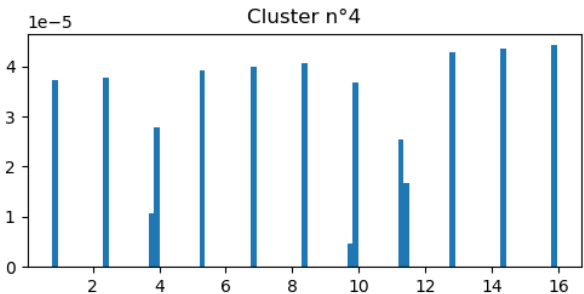
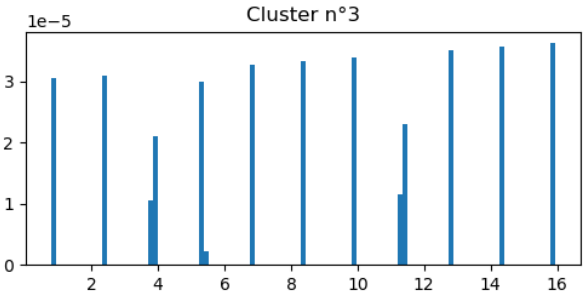
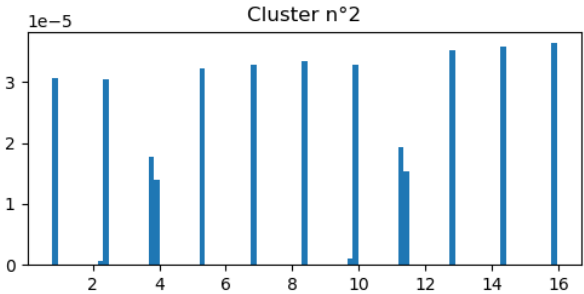
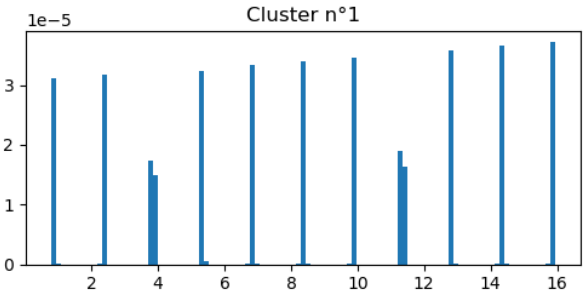
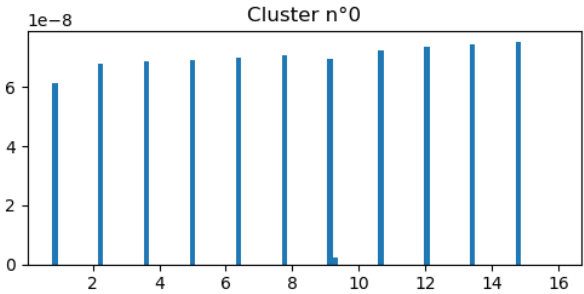
Clusters	Impulsions	%
0	142	0,8
1	3 825	21
2	4 512	24,8
3	4 557	25,1
4	5 152	28,3



Histogrammes

TOA

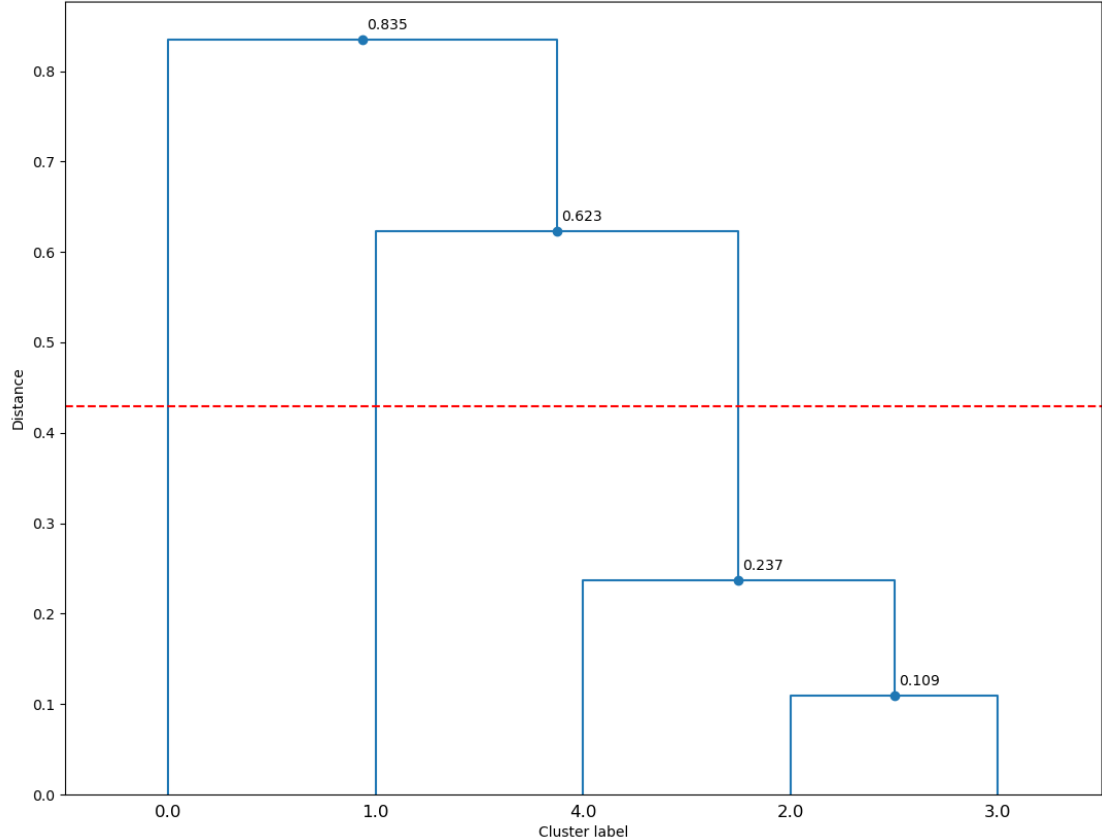
Information	Paramètre
Variable	TOA + LEVEL
Bins	100
Distribution	Uniforme
Clusters	5



Dendrogramme

Méthode 1 : HAC avec distances du Transport Optimal sur la TOA

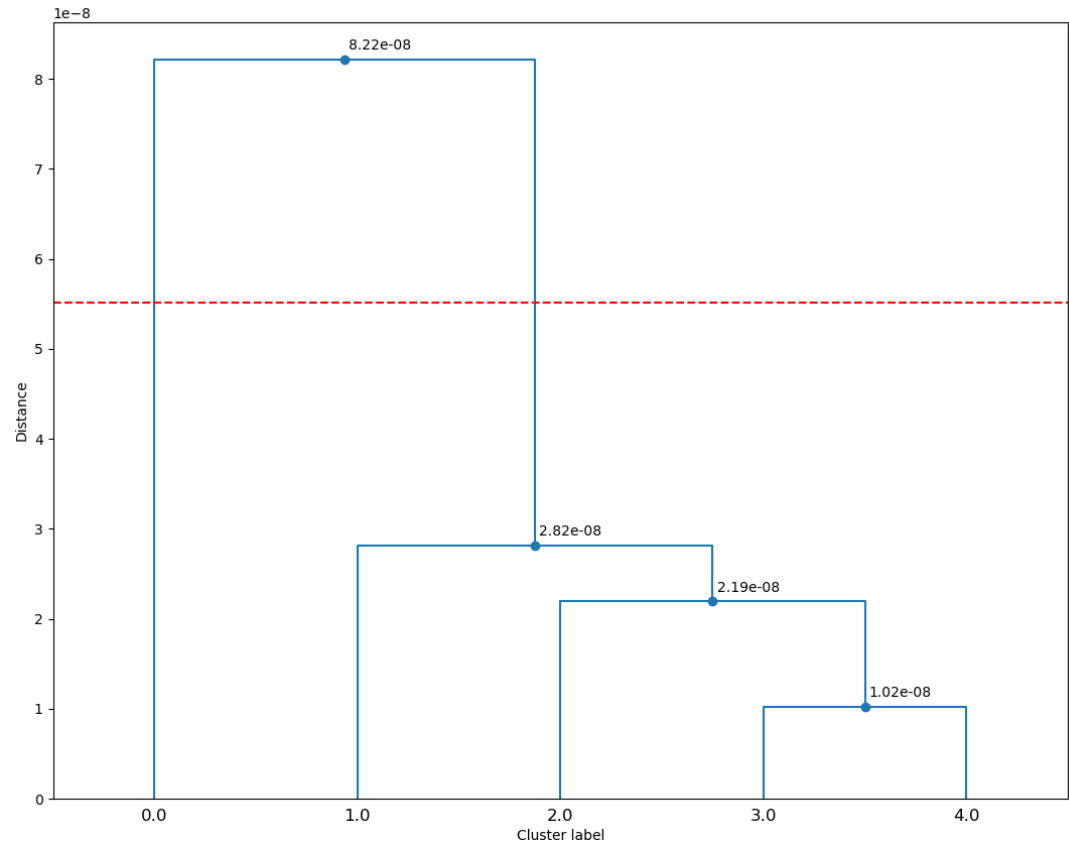
Arguments	Paramètres
Variable	TOA
Cut optimal	3
Type de données	Histogrammes
Spécificités des données	Nombre d'impulsions
Bins	100



Dendrogramme

Méthode 2 : HAC avec distances du Transport Optimal sur le LEVEL

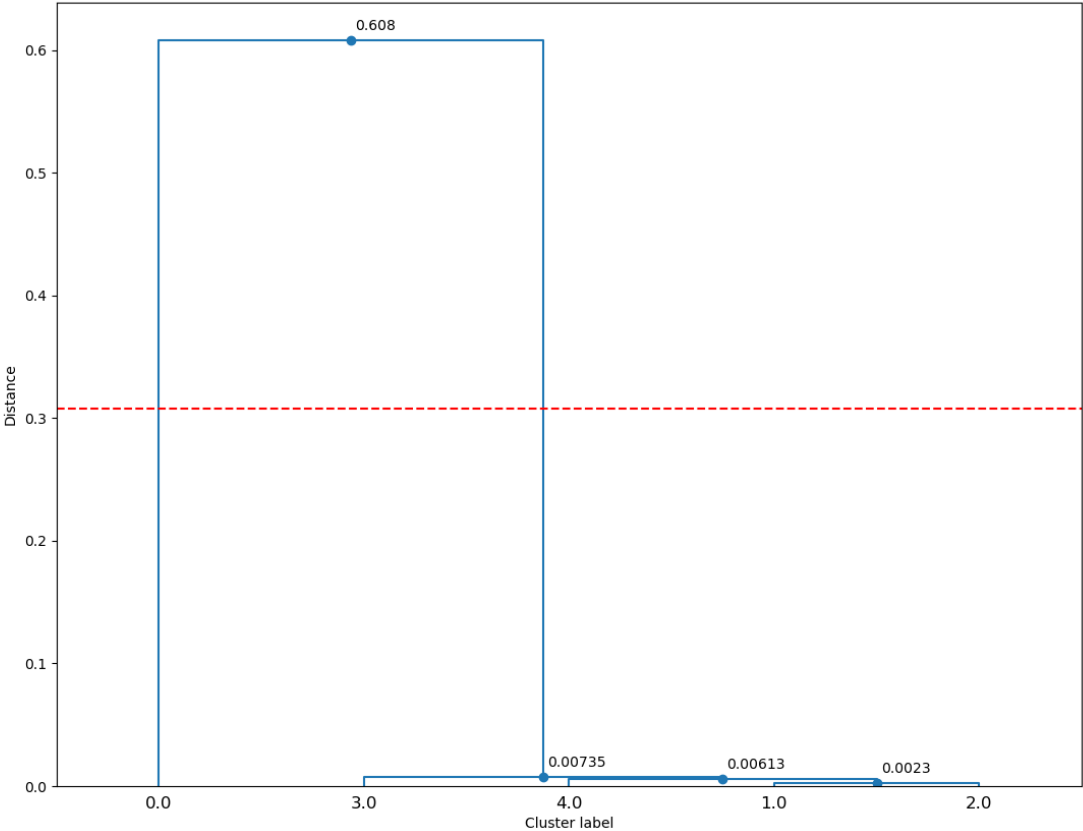
Arguments	Paramètres
Variable	LEVEL
Cut optimal	2
Type de données	Histogrammes
Spécificités des données	Nombre d'impulsions
Bins	100



Dendrogramme

Méthode 3 : HAC avec distances du Transport Optimal sur la TOA avec le LEVEL

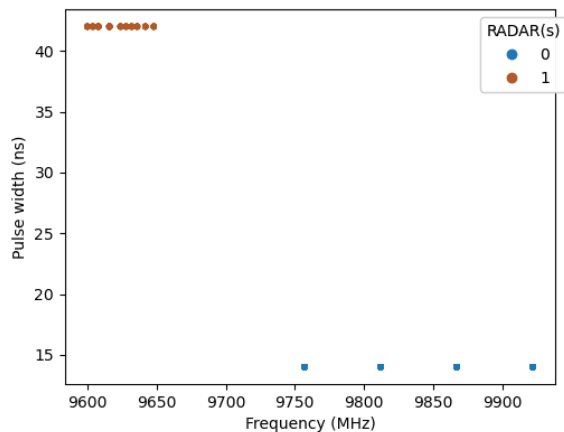
Arguments	Paramètres
Variable	TOA + LEVEL
Cut optimal	2
Type de données	Histogrammes
Spécificités des données	Niveaux cumulés
Bins	100



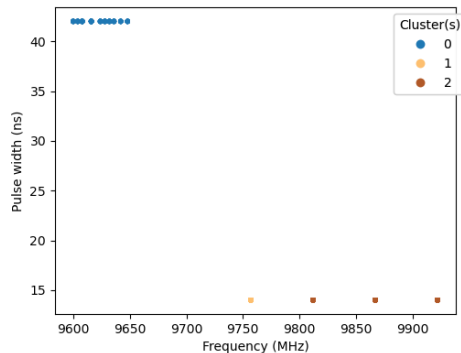
Regroupements finaux

FREQ x DI

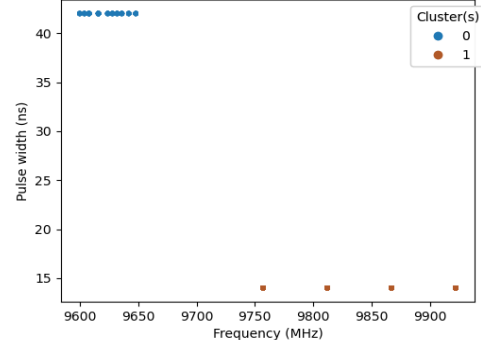
Vérité terrain



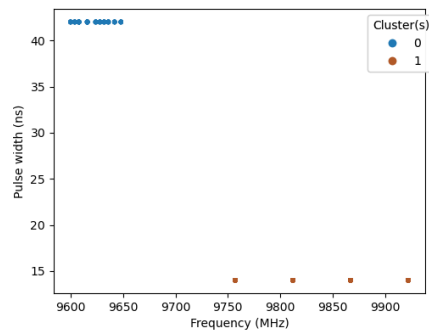
HAC + OT sur la TOA



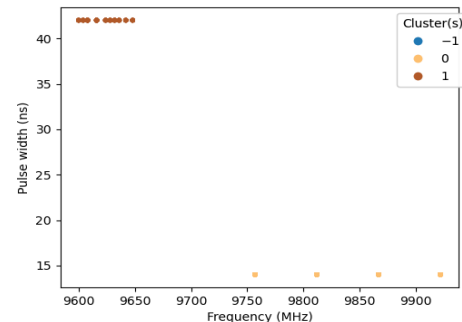
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



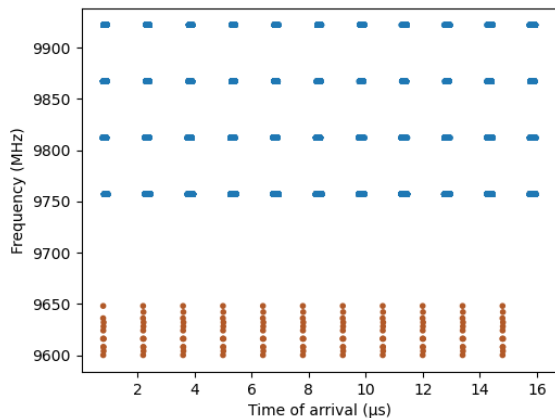
AVANTIX



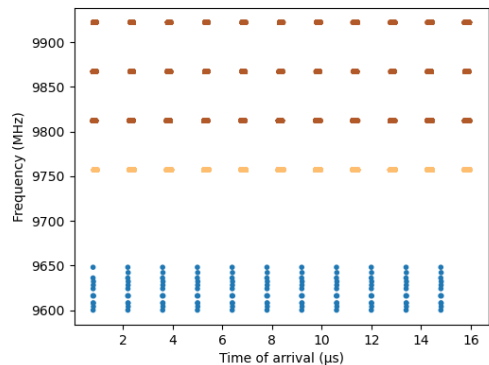
Regroupements finaux

FREQ x TOA

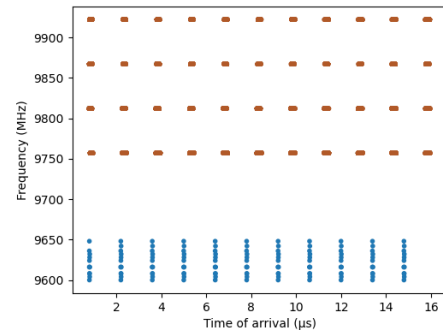
Vérité terrain



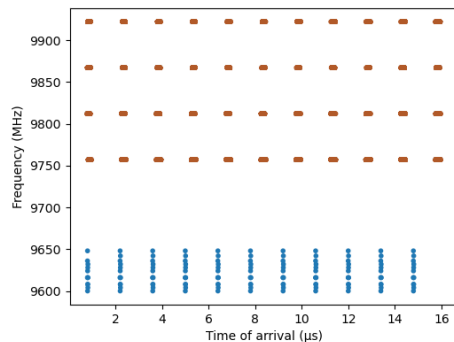
HAC + OT sur la TOA



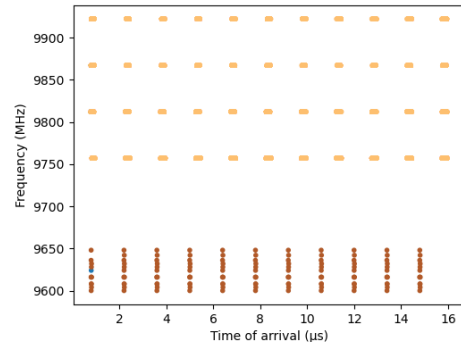
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



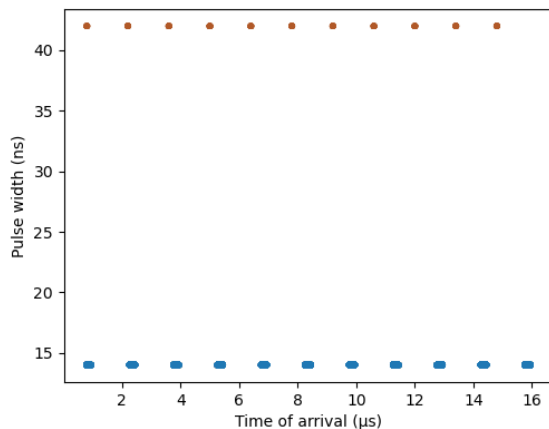
AVANTIX



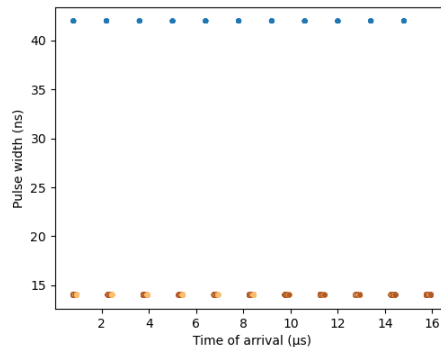
Regroupements finaux

DI x TOA

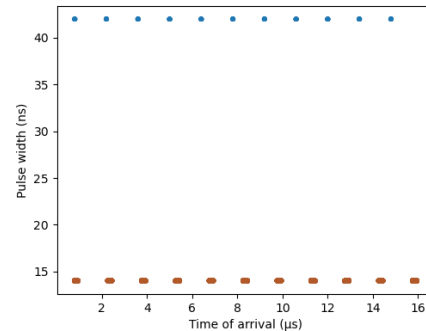
Vérité terrain



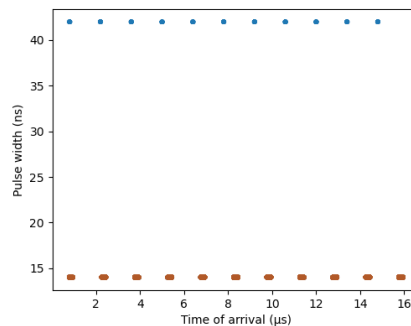
HAC + OT sur la TOA



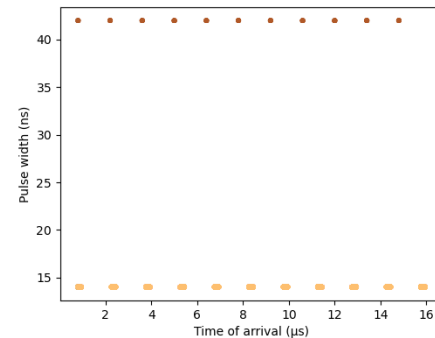
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



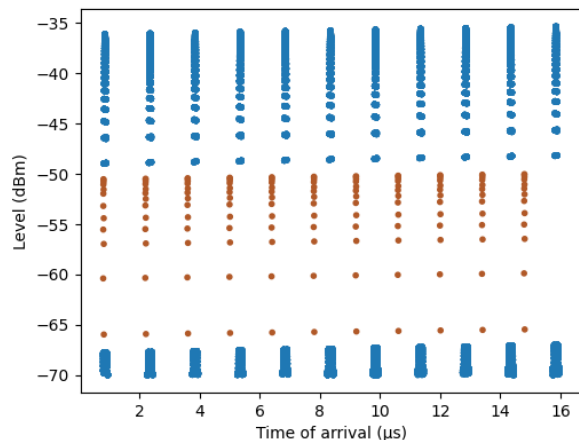
AVANTIX



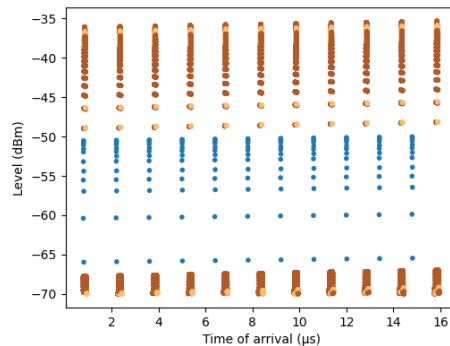
Regroupements finaux

LEVEL x TOA

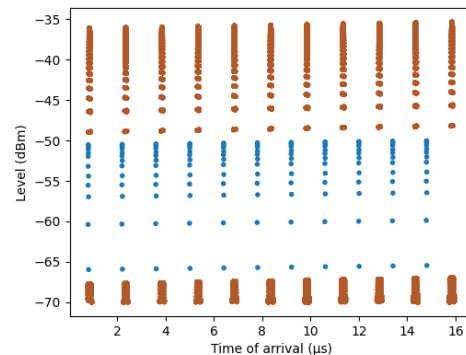
Vérité terrain



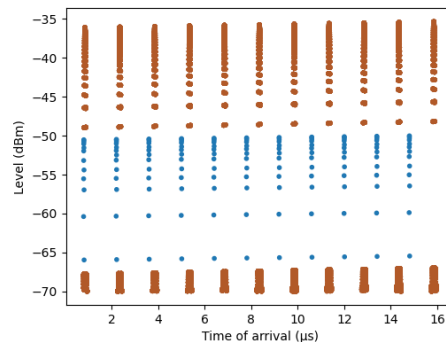
HAC + OT sur la TOA



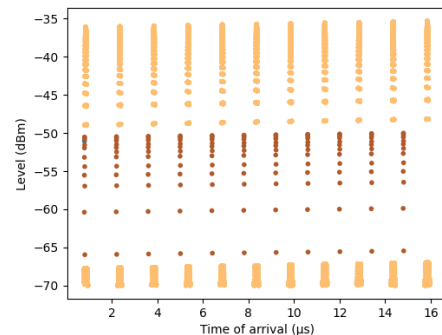
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



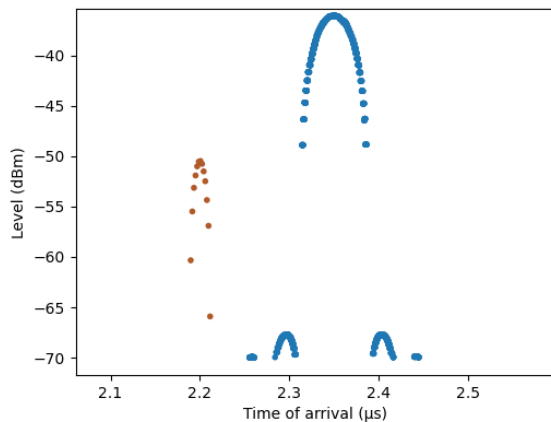
AVANTIX



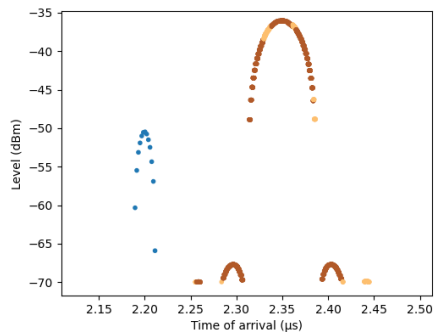
Regroupements finaux

LEVEL x TOA zoom

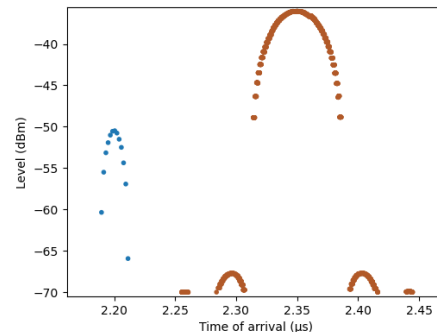
Vérité terrain



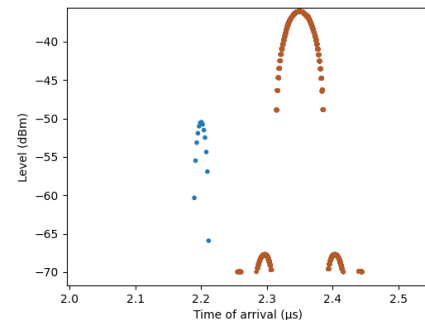
HAC + OT sur la TOA



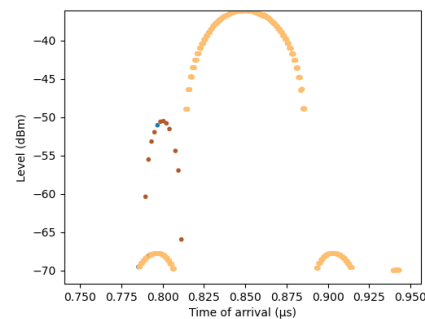
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



AVANTIX



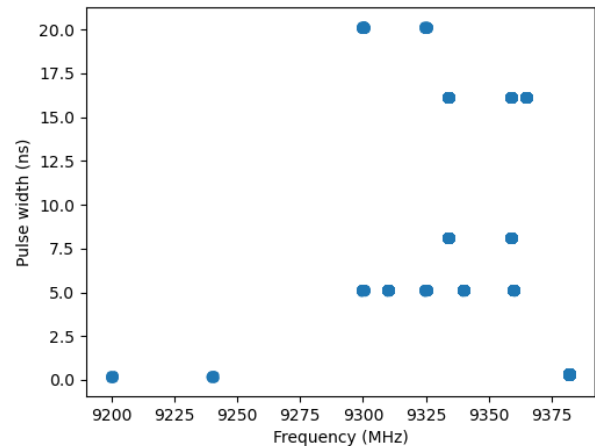
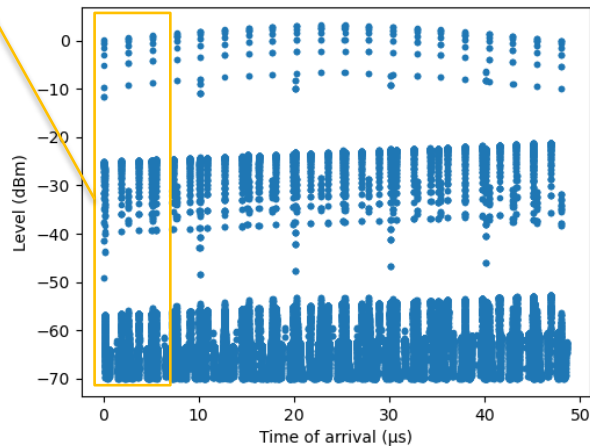
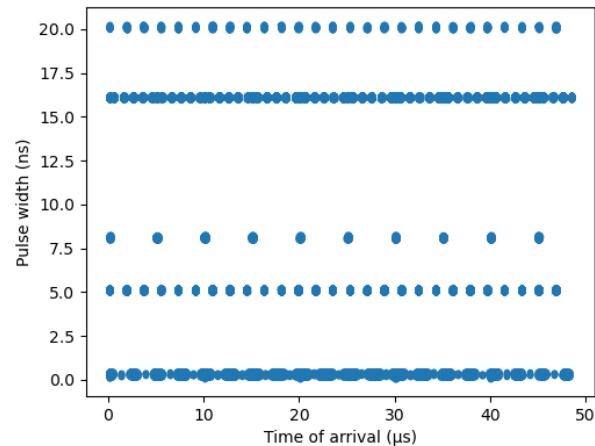
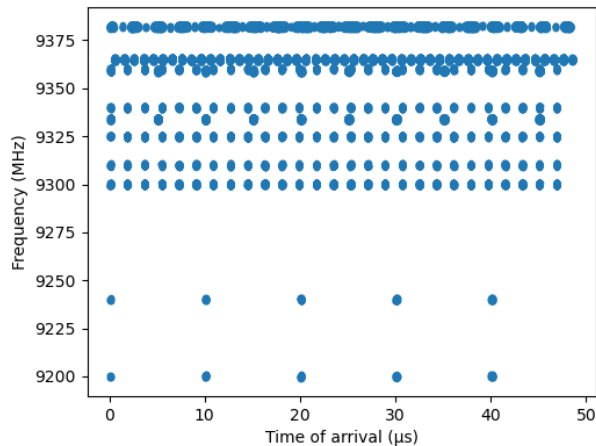
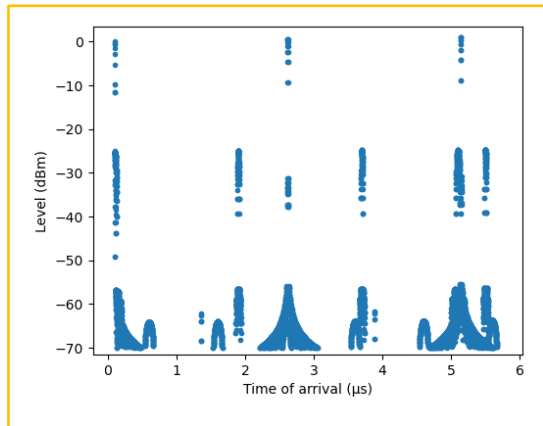
Regroupements finaux

Statistiques

Méthode	Homogénéité	Complétude	Outliers	Clusters
HAC + OT sur la TOA	1	0,08	0	3
HAC + OT sur le LEVEL	1	1	0	2
HAC + OT sur la TOA avec le LEVEL	1	1	0	2
AVANTIX	1	1	5	2

CAS D'APPLICATION 2

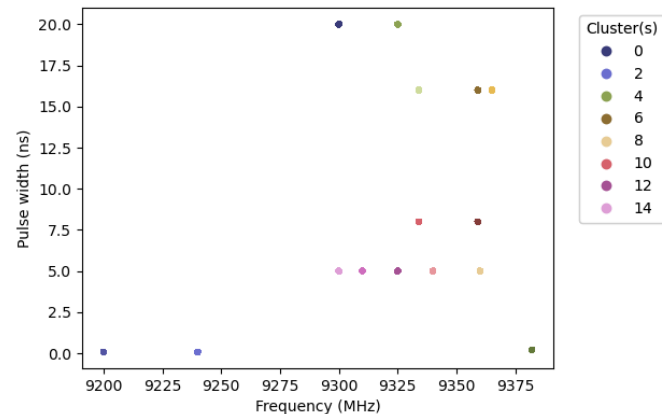
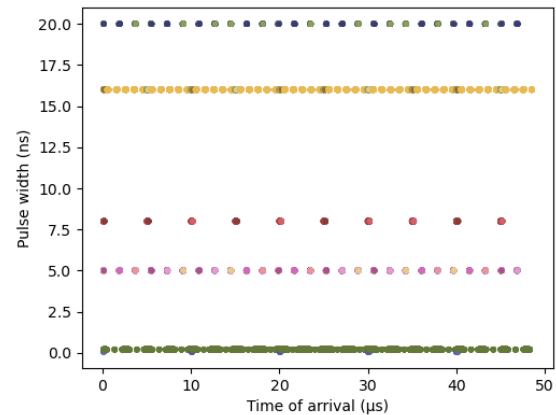
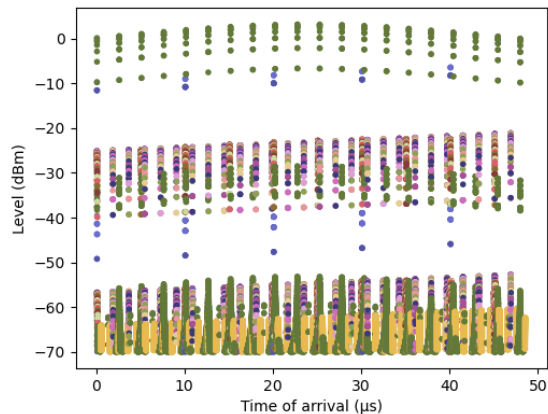
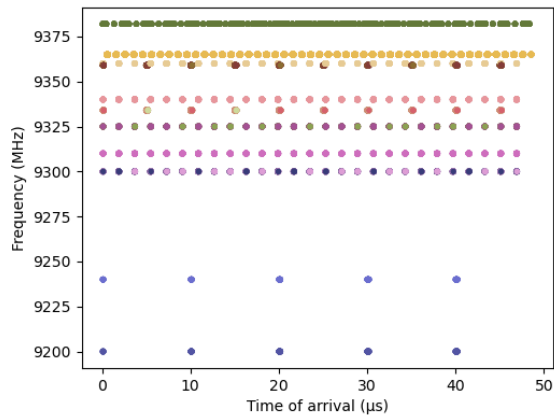
Données brutes



Clustering

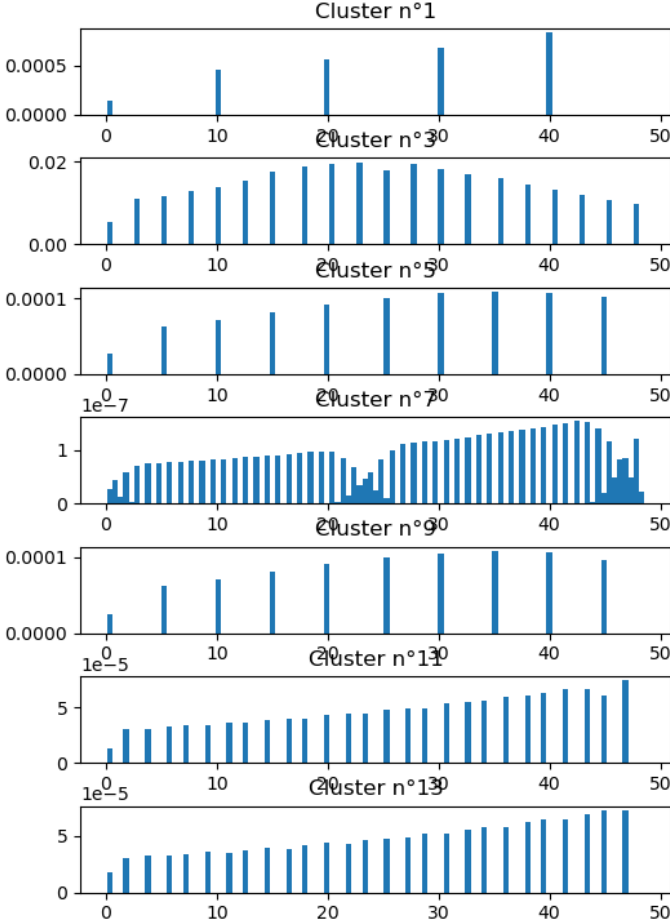
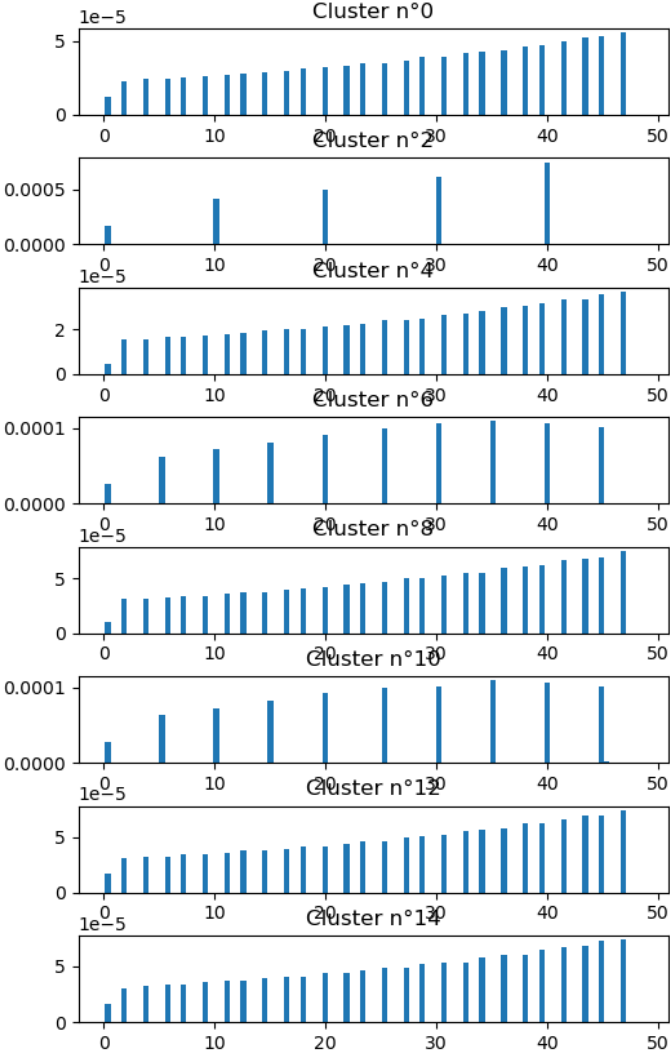
HDBSCAN

Clusters	Impulsions	%
0.0	936	1,9
1.0	176	0,4
2.0	174	0,3
3.0	25 195	50,4
4.0	621	1,2
5.0	895	1,8
6.0	890	1,8
7.0	13 080	26,2
8.0	1 254	2,5
9.0	887	1,8
10.0	890	1,8
11.0	1 255	2,5
12.0	1 246	2,5
13.0	1 256	2,5
14.0	1 244	2,5



Histogrammes TOA

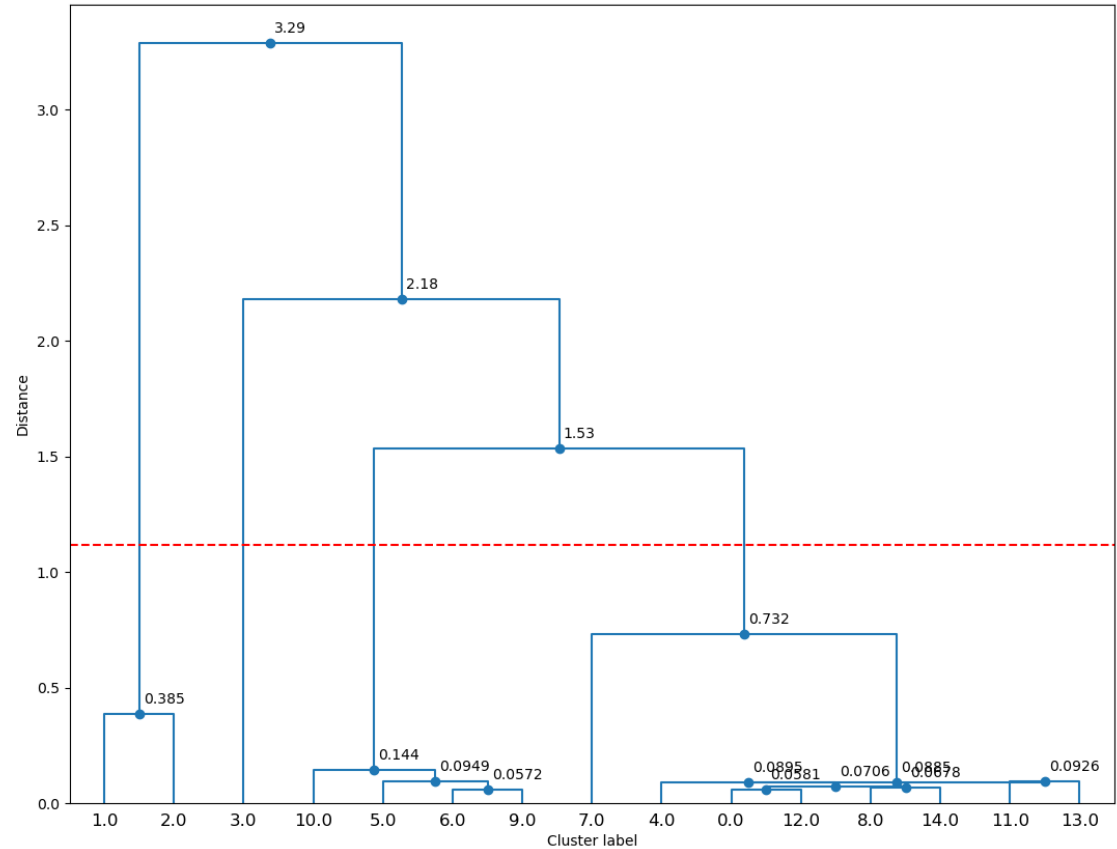
Information	Paramètre
Variable	TOA + LEVEL
Bins	100
Distribution	Uniforme
Clusters	15



Dendrogramme

Méthode 1 : HAC avec distances du Transport Optimal sur la TOA

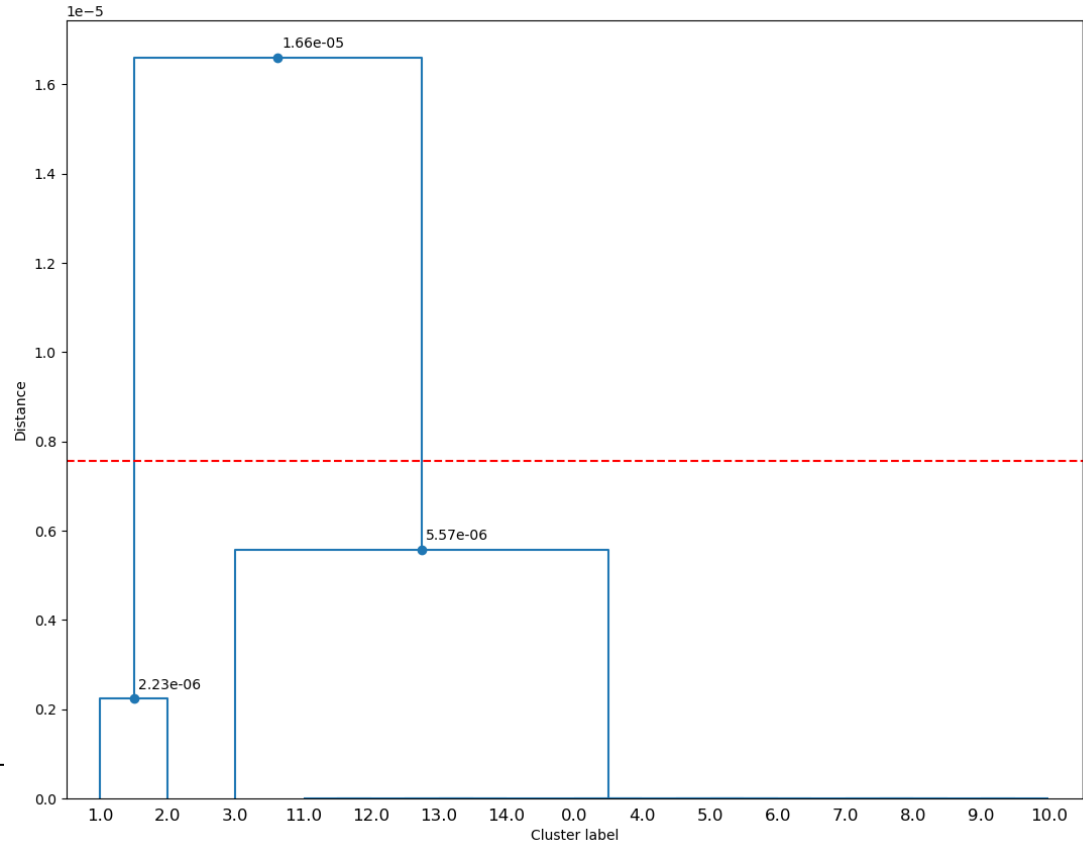
Arguments	Paramètres
Variable	TOA
Cut optimal	4
Type de données	Histogrammes
Spécificités des données	Nombre d'impulsions
Bins	100



Dendrogramme

Méthode 2 : HAC avec distances du Transport Optimal sur le LEVEL

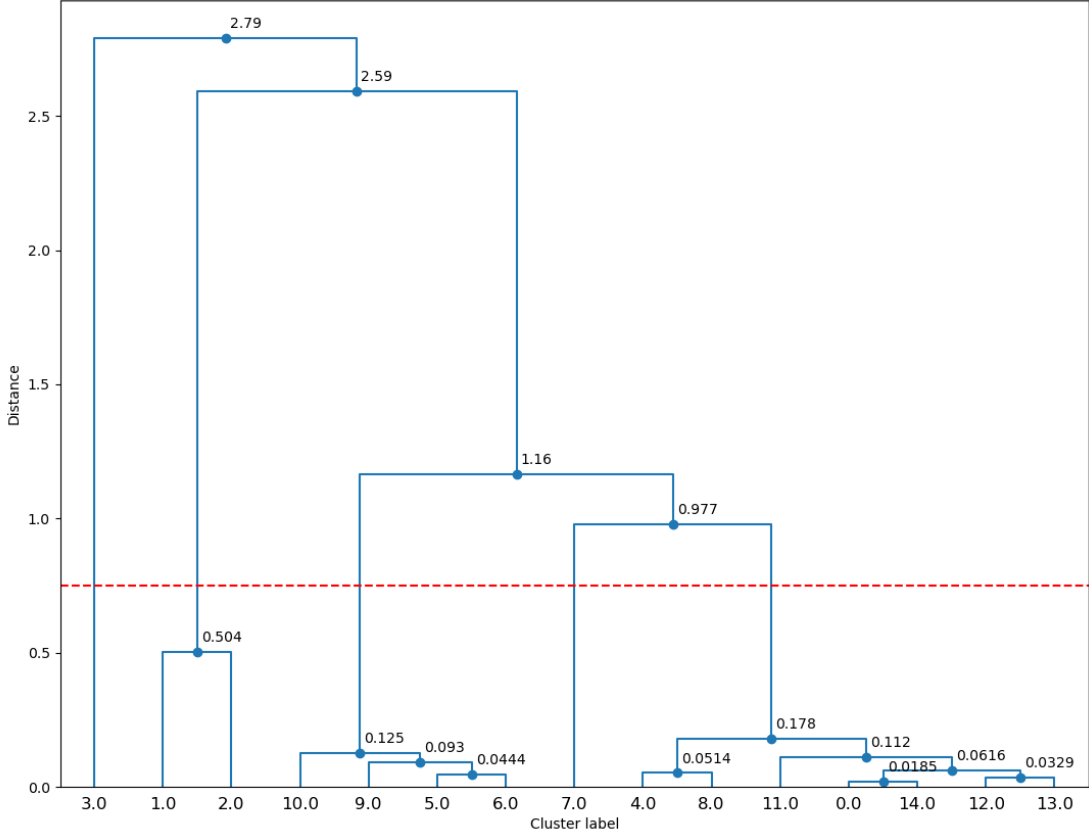
Arguments	Paramètres
Variable	LEVEL
Cut optimal	2
Type de données	Histogrammes
Spécificités des données	Nombre d'impulsions
Bins	100



Dendrogramme

Méthode 3 : HAC avec distances du Transport Optimal sur la TOA avec le LEVEL

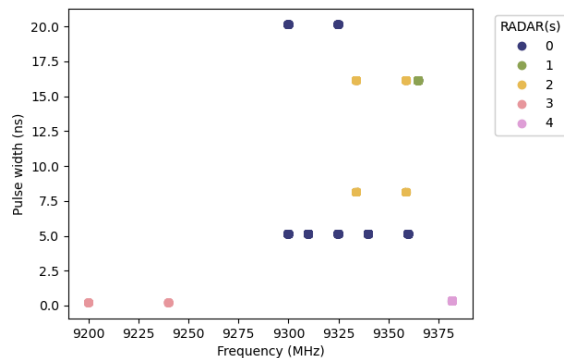
Arguments	Paramètres
Variable	TOA + LEVEL
Cut optimal	5
Type de données	Histogrammes
Spécificités des données	Niveaux cumulés
Bins	100



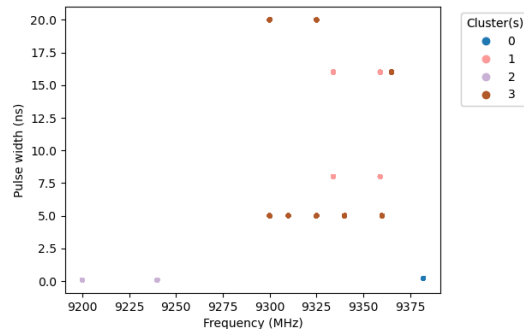
Regroupements finaux

FREQ x DI

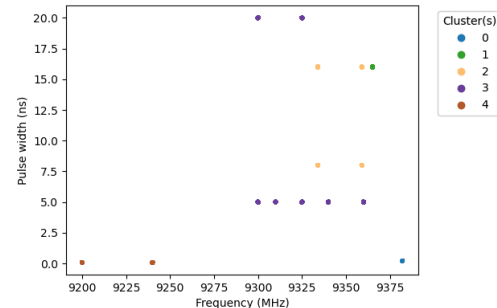
Vérité terrain



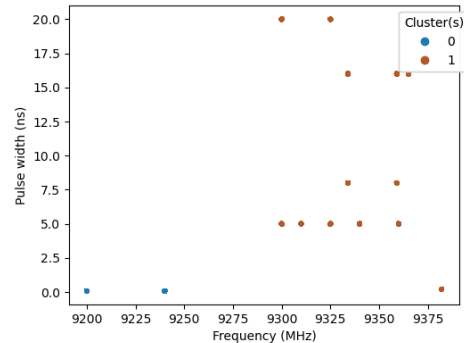
HAC + OT sur la TOA



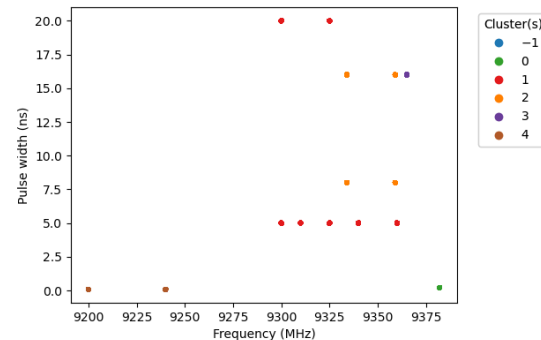
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



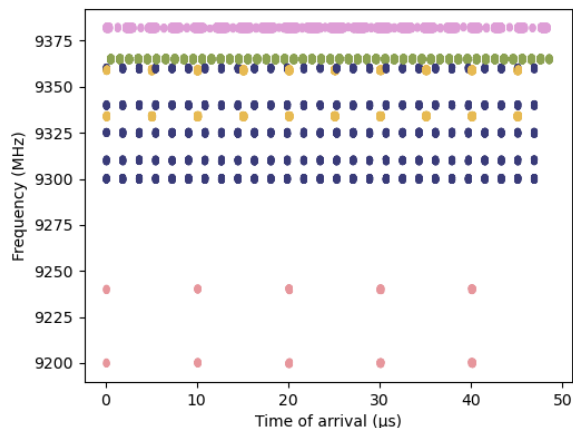
AVANTIX



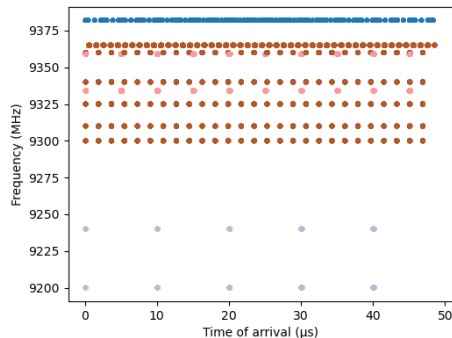
Regroupements finaux

FREQ x TOA

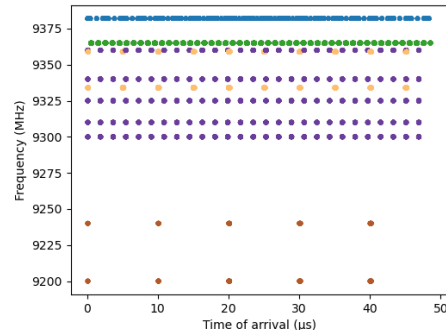
Vérité terrain



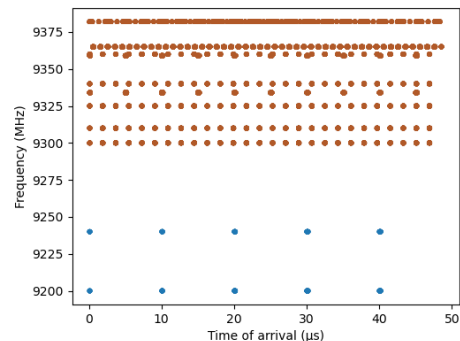
HAC + OT sur la TOA



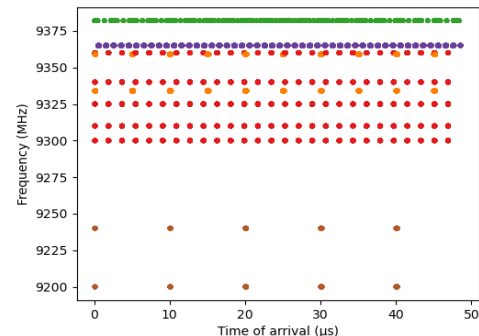
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



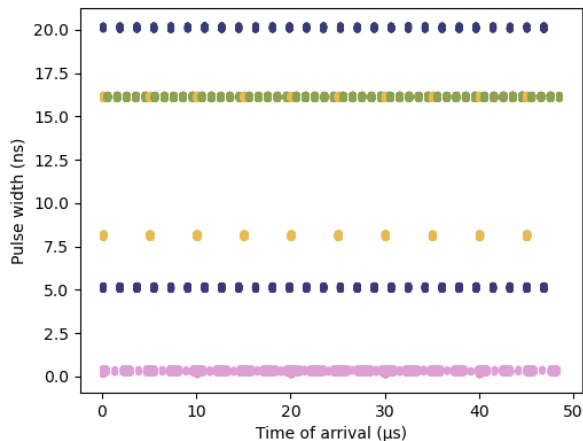
AVANTIX



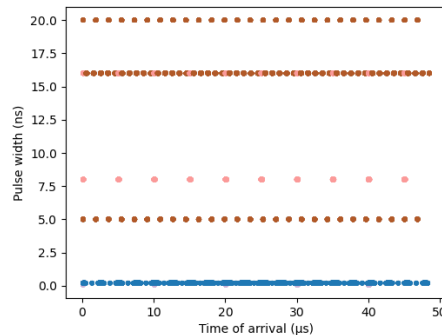
Regroupements finaux

DI x TOA

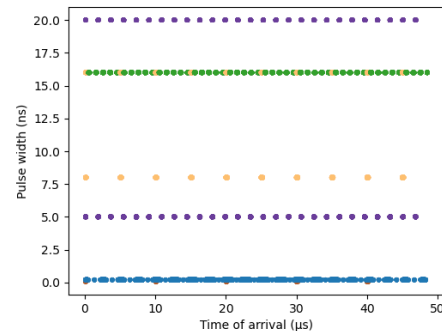
Vérité terrain



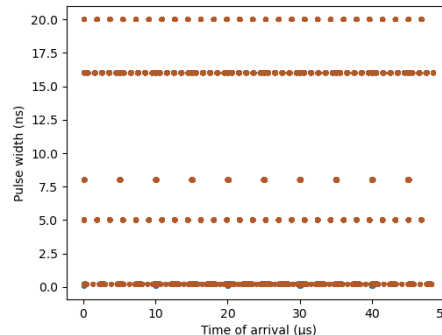
HAC + OT sur la TOA



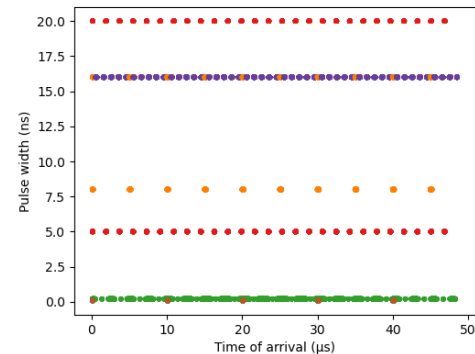
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



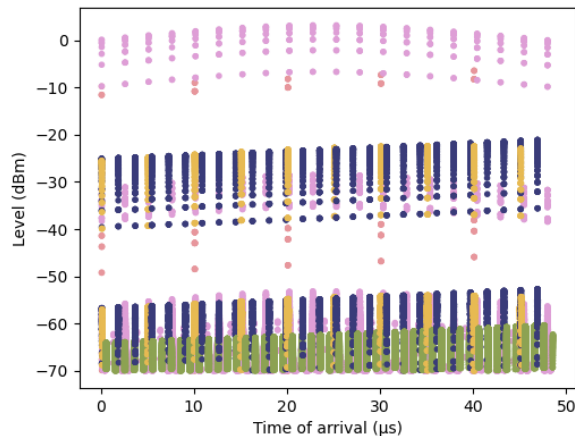
AVANTIX



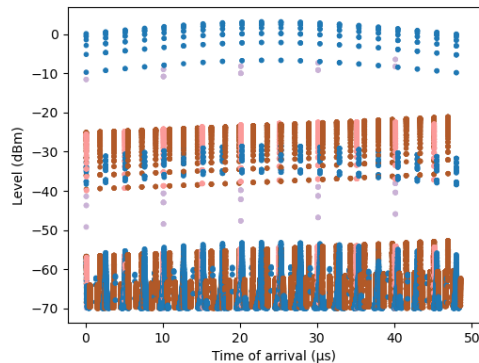
Regroupements finaux

LEVEL x TOA

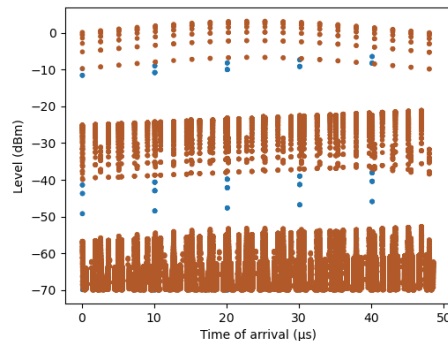
Vérité terrain



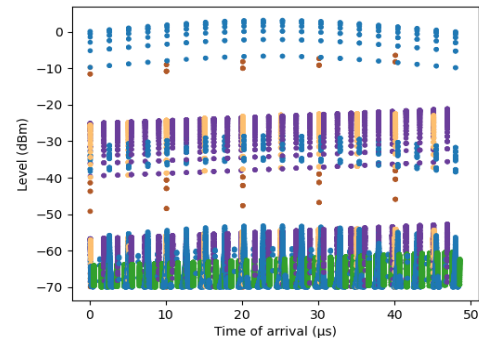
HAC + OT sur la TOA



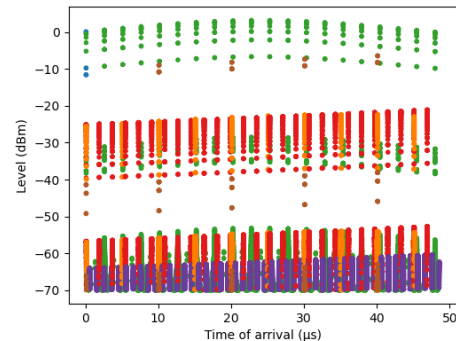
HAC + OT sur le LEVEL



HAC + OT sur la TOA avec le LEVEL



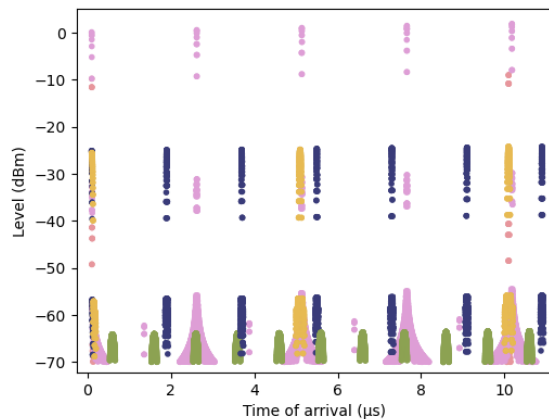
AVANTIX



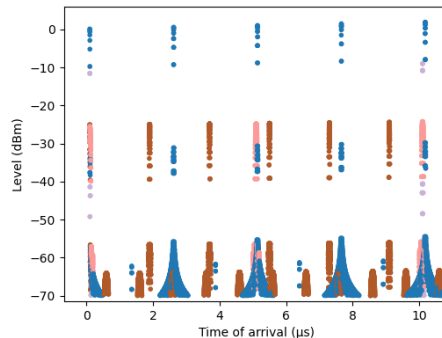
Regroupements finaux

LEVEL x TOA zoom

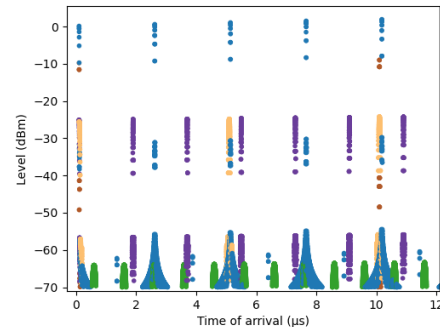
Vérité terrain



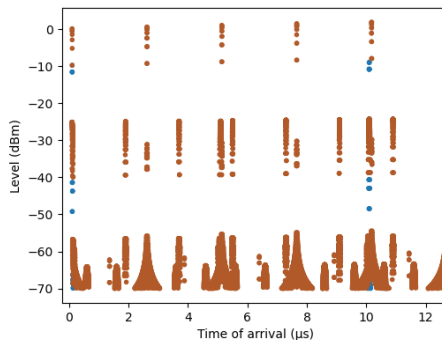
HAC + OT sur la TOA



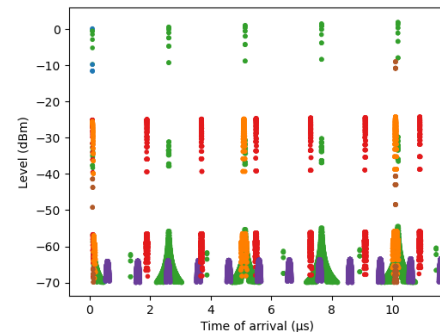
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



AVANTIX



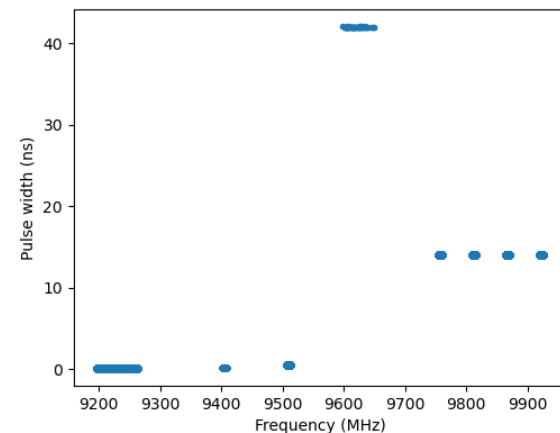
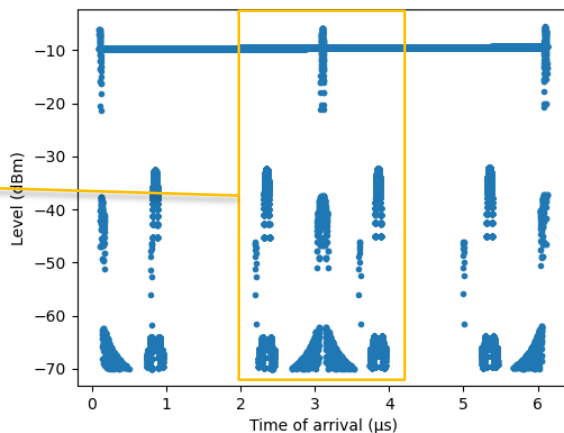
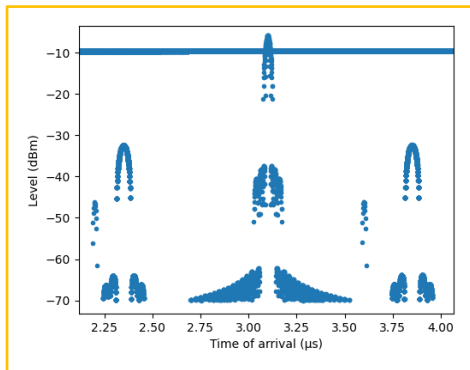
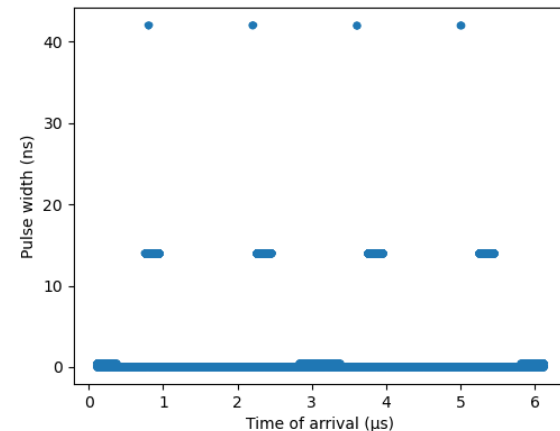
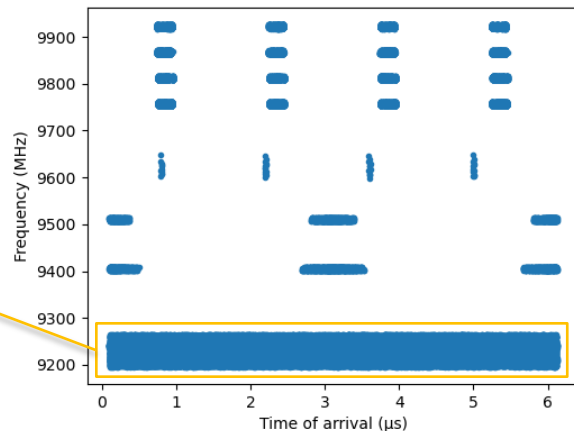
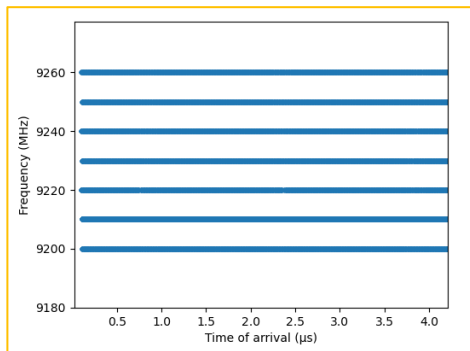
Regroupements finaux

Statistiques

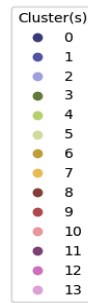
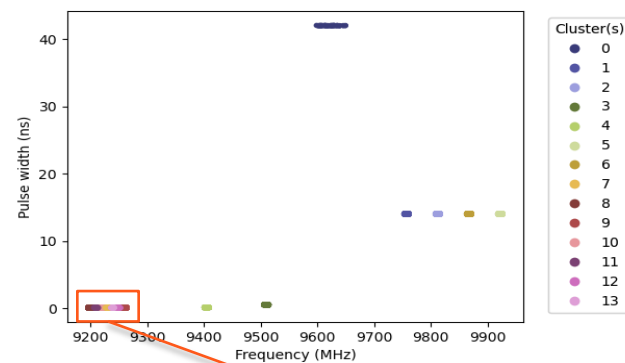
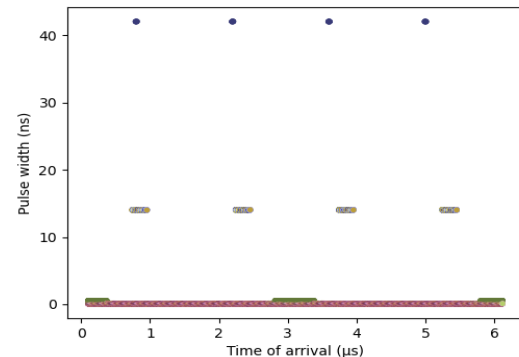
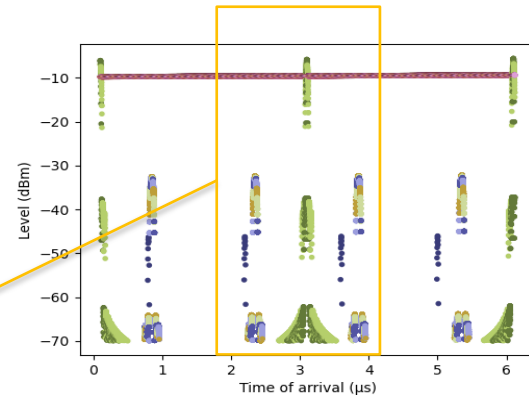
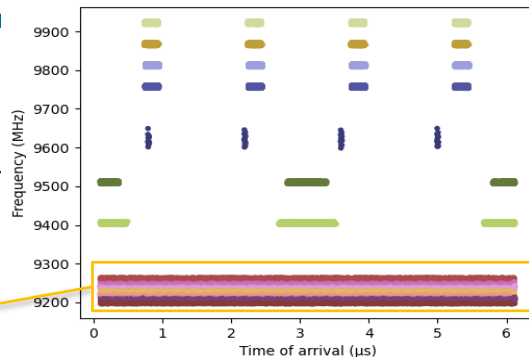
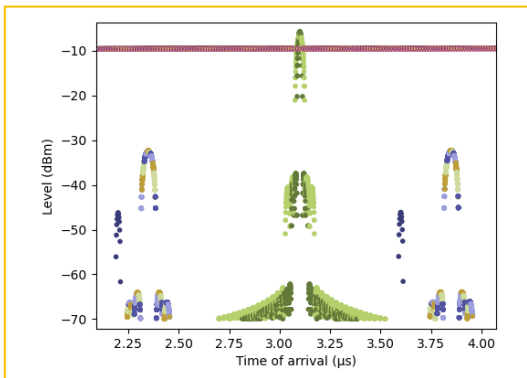
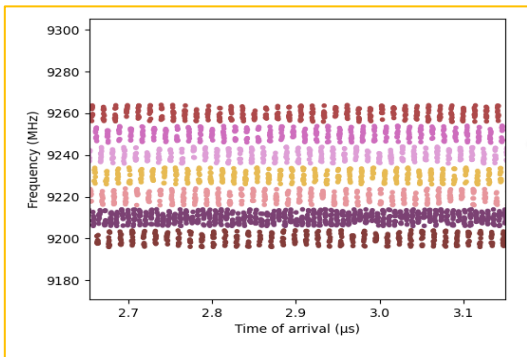
Méthode	Homogénéité	Complétude	Outliers	Clusters
HAC + OT sur la TOA	0,77	1	0	4
HAC + OT sur le LEVEL	0,03	1	0	2
HAC + OT sur la TOA avec le LEVEL	1	1	0	5
AVANTIX	1	1	13	5

CAS D'APPLICATION 3

Données brutes



Clustering HDBSCAN

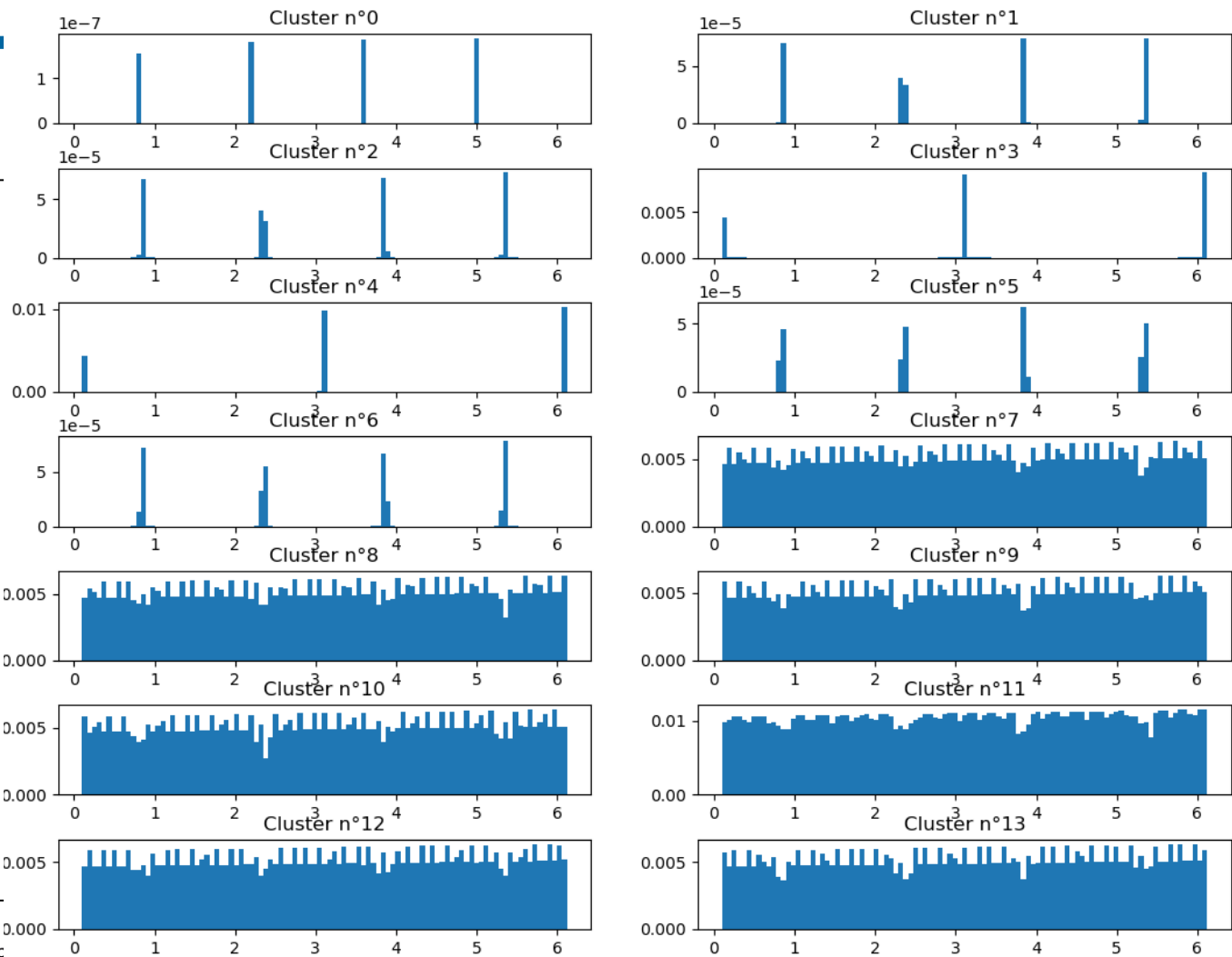


Cluster s	Impulsions	%		Clusters	Impulsio ns	%
0	50	0,1		7	4681	9,4
1	2304	4,6		8	4669	9,3
2	2066	4,1		9	4653	9,3
3	1576	3,2		10	4648	9,3
4	1725	3,5		11	9332	18,7
5	2378	4,8		12	4675	9,4
6	2571	5,1		13	4671	9,3



Histogrammes TOA

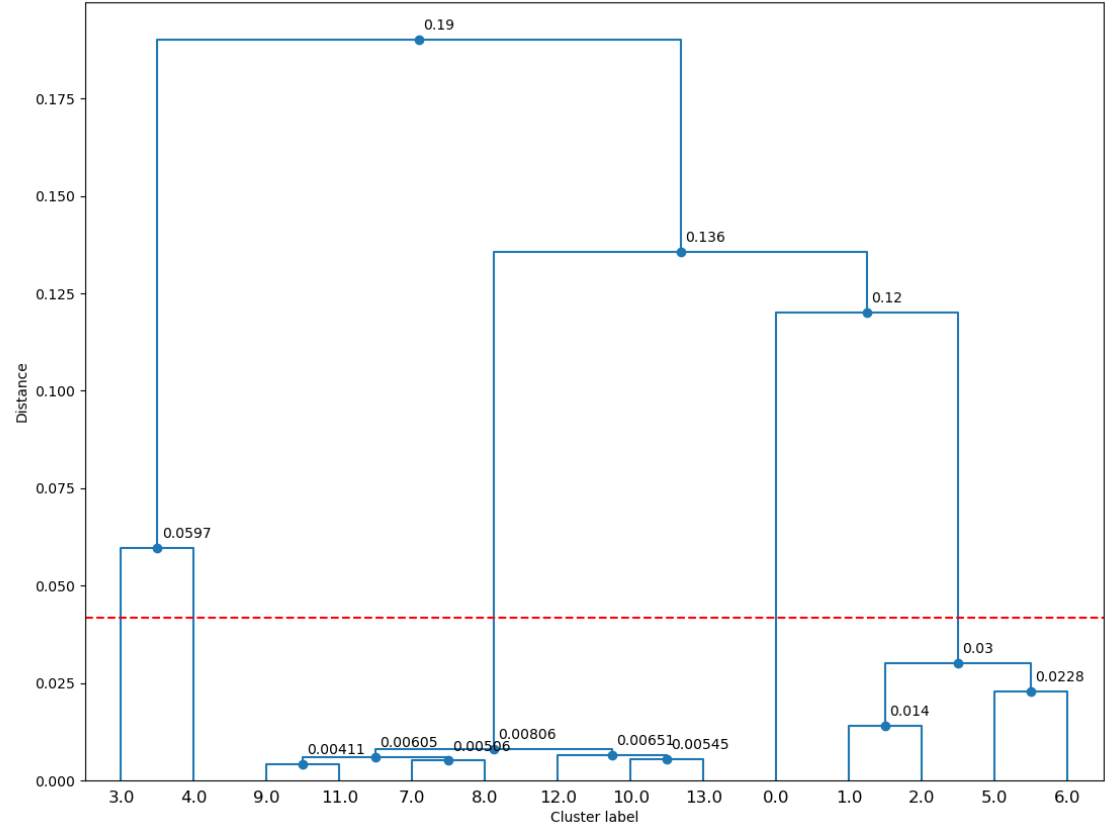
Information	Paramètre
Variable	TOA + LEVEL
Bins	100
Distribution	Uniforme
Clusters	14



Dendrogramme

Méthode 1 : HAC avec distances du Transport Optimal sur la TOA

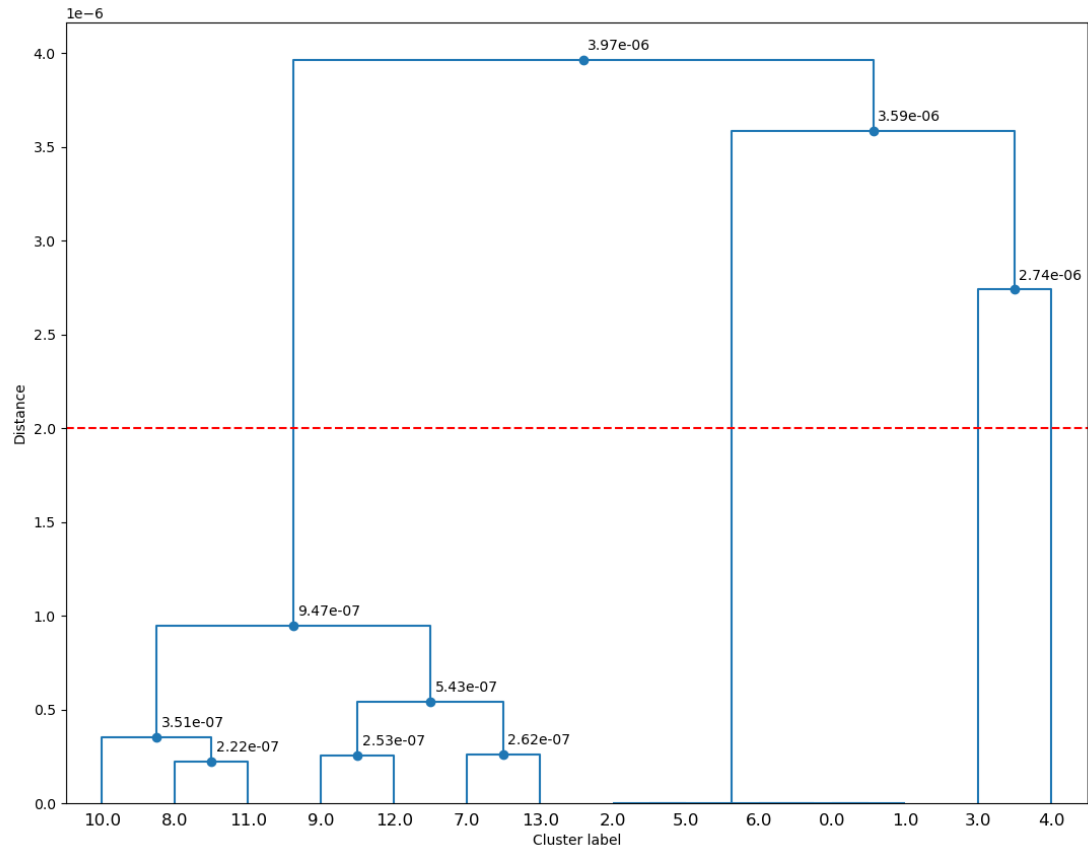
Information	Paramètre
Variable	TOA
Cut optimal	5
Type de données	Histogrammes
Spécificités des données	Nombre d'impulsions
Bins	100



Dendrogramme

Méthode 2 : HAC avec distances du Transport Optimal sur le LEVEL

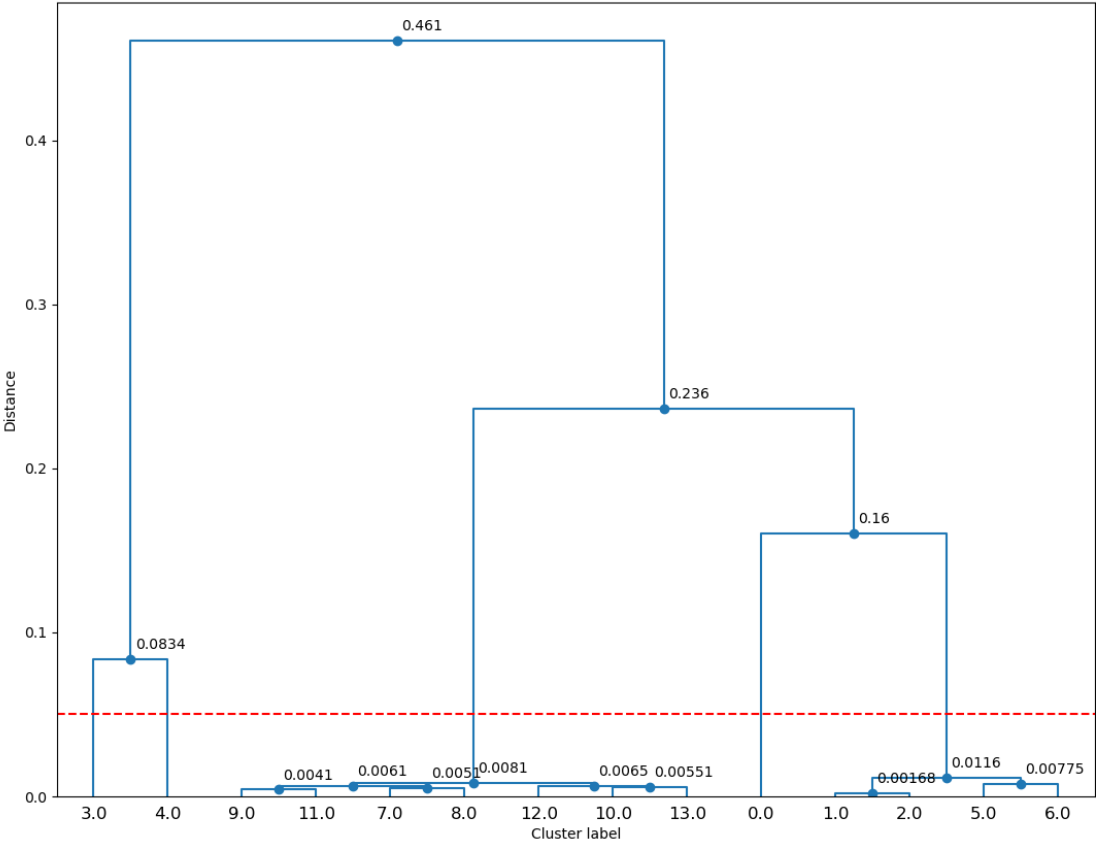
Information	Paramètre
Variable	LEVEL
Cut optimal	4
Type de données	Histogrammes
Spécificités des données	Nombre d'impulsions
Bins	100



Dendrogramme

Méthode 3 : HAC avec distances du Transport Optimal sur la TOA avec le LEVEL

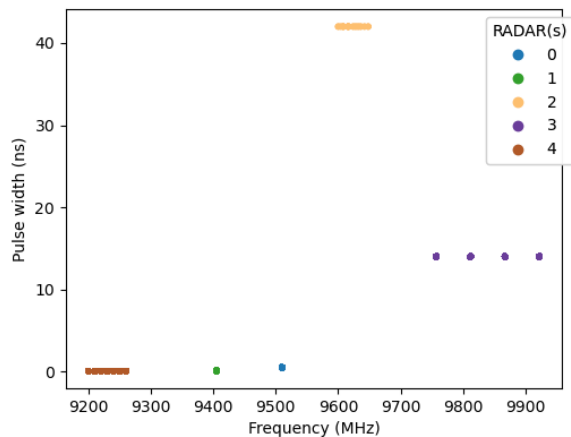
Information	Paramètre
Variable	TOA + LEVEL
Cut optimal	5
Type de données	Histogrammes
Spécificités des données	Niveaux cumulés
Bins	100



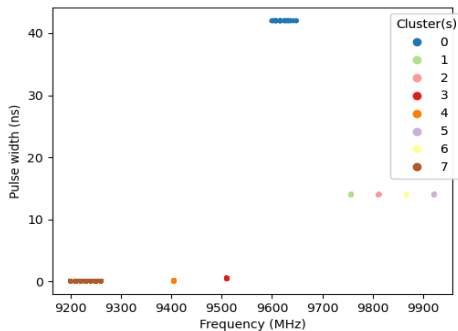
Regroupements finaux

FREQ x DI

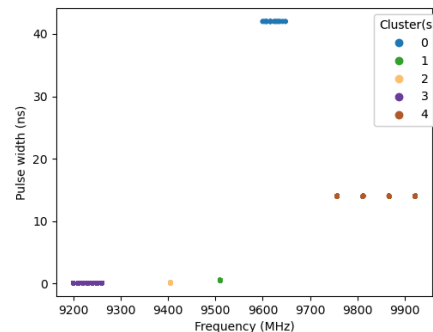
Vérité terrain



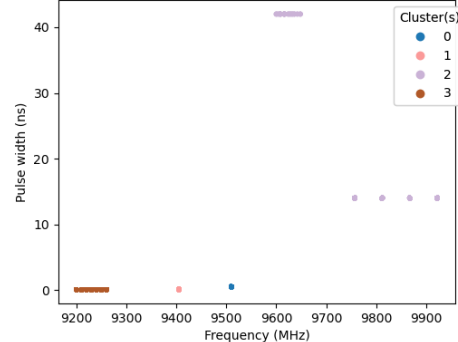
HAC + OT sur la TOA



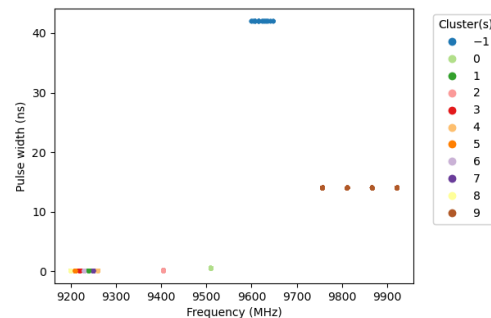
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



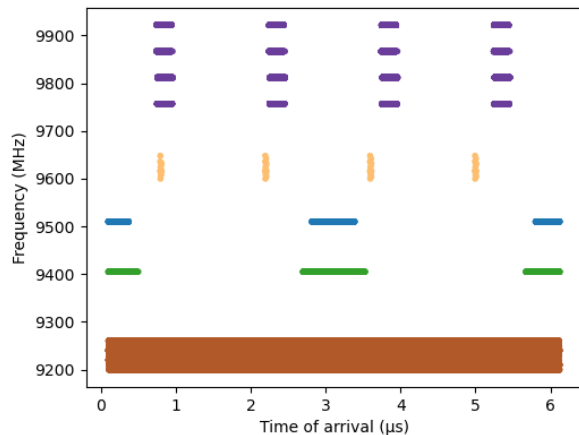
AVANTIX



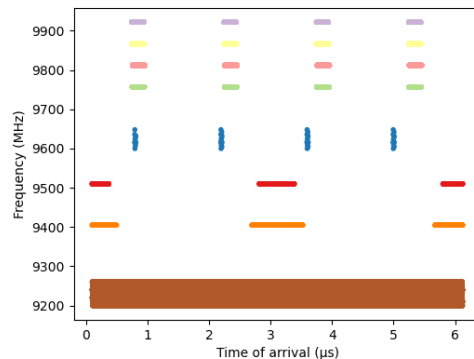
Regroupements finaux

FREQ x TOA

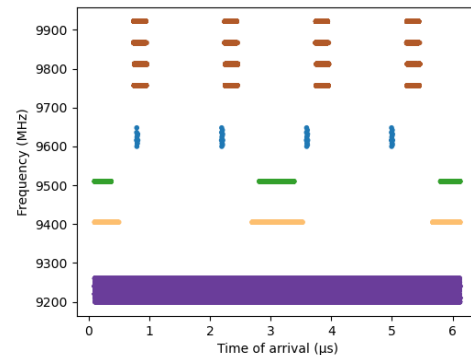
Vérité terrain



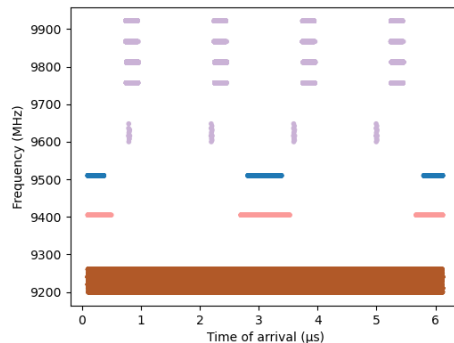
HAC + OT sur la TOA



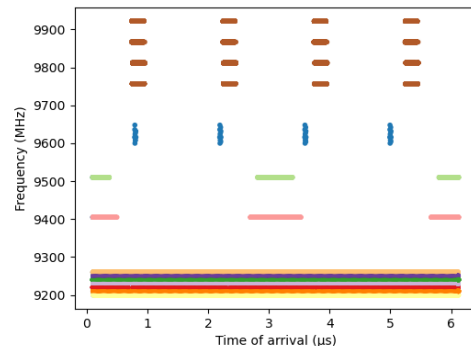
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



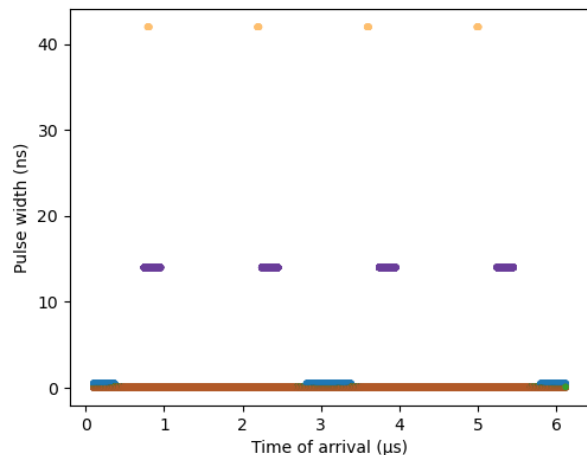
AVANTIX



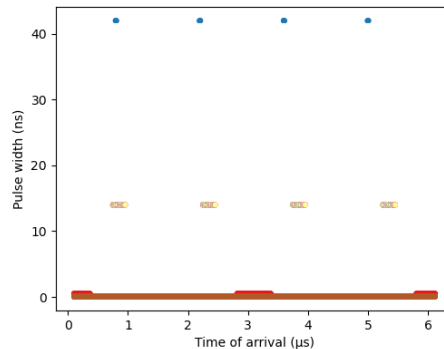
Regroupements finaux

DI x TOA

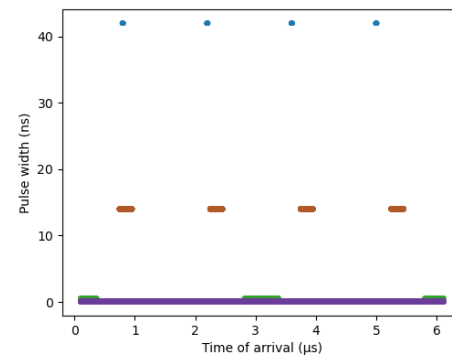
Vérité terrain



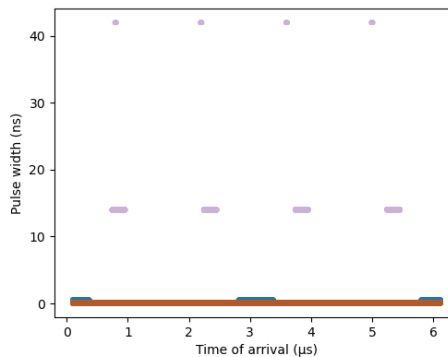
HAC + OT sur la TOA



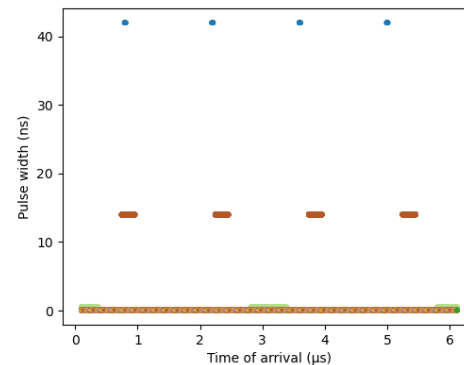
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



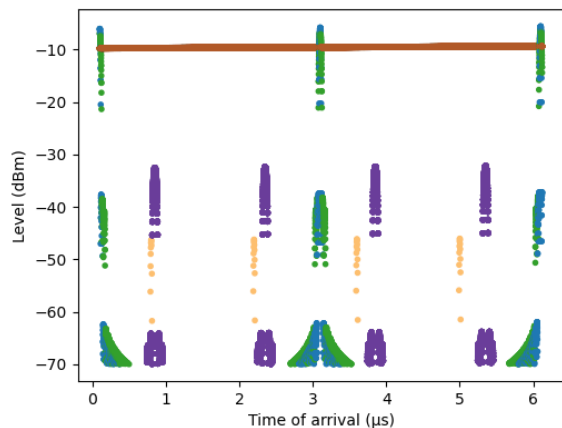
AVANTIX



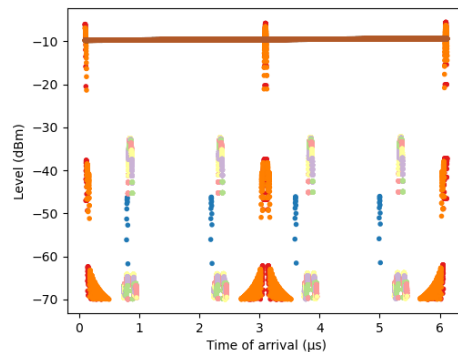
Regroupements finaux

LEVEL x TOA

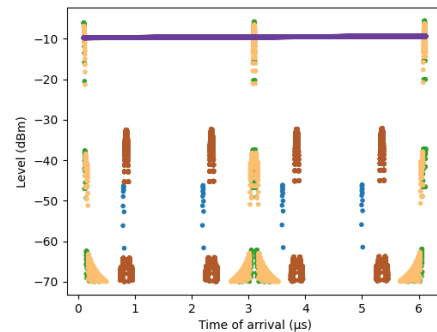
Vérité terrain



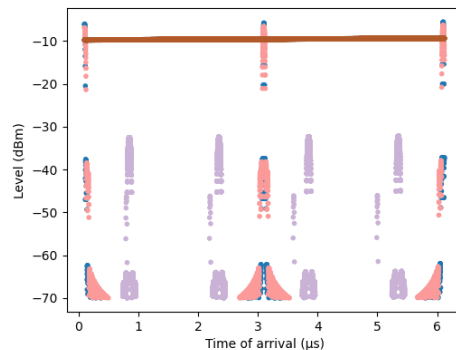
HAC + OT sur la TOA



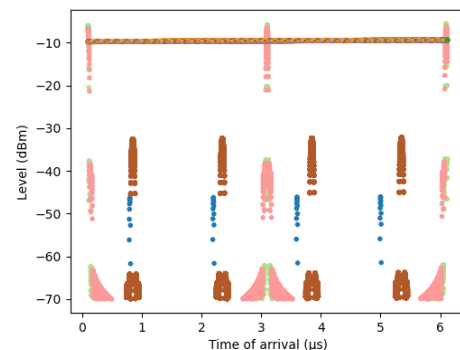
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



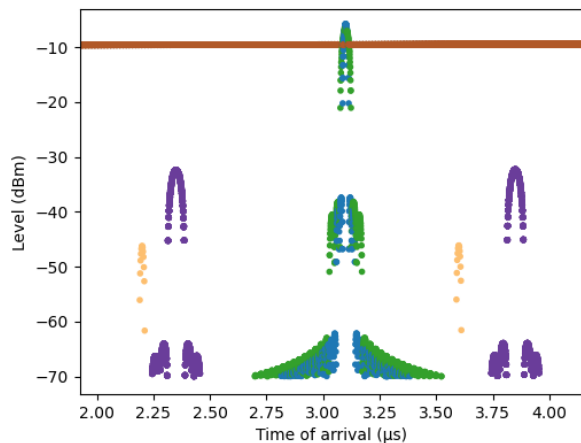
AVANTIX



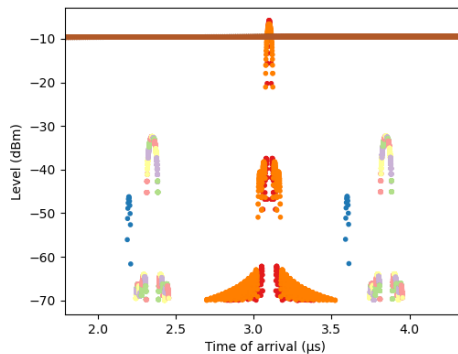
Regroupements finaux

LEVEL x TOA zoom

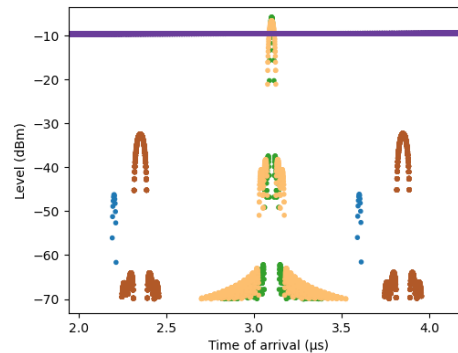
Vérité terrain



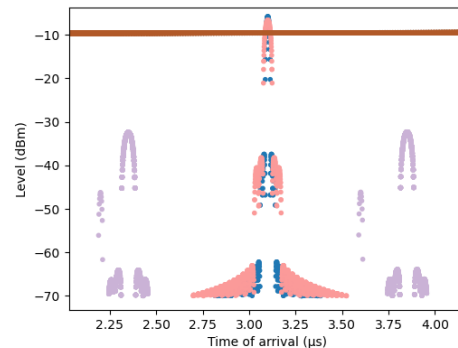
HAC + OT sur la TOA



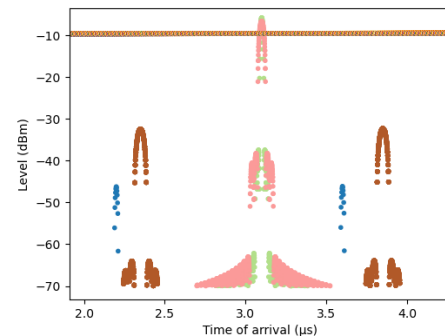
HAC + OT sur la TOA avec le LEVEL



HAC + OT sur le LEVEL



AVANTIX



Regroupements finaux

Statistiques

Méthode	Homogénéité	Complétude	Outliers	Clusters
HAC + OT sur la TOA	1	0,74	0	8
HAC + OT sur le LEVEL	0,99	1	0	5
HAC + OT sur la TOA avec le LEVEL	1	1	0	4
AVANTIX	0,99	0,34	70	10



4

Conclusion

Conclusion

- Publication d'un article et participation à la conférence IEEE Radar Conference 2021
- Prochaines étapes :
 - Création d'une métrique pour évaluer les regroupements entre les clusters
 - Création d'une méthode pour normaliser les données

Thank you

For more information please contact:

Manon MOTTIER
manon.mottier@atos.net