

Projet R&T AIMS: Artificial Intelligence for Mass Spectrometry

IJCLab Orsay (laboratoire porteur): Françoise Bouvet, Serge Della Negra, Isabelle Ribaud, **Thomas Zerguerras**

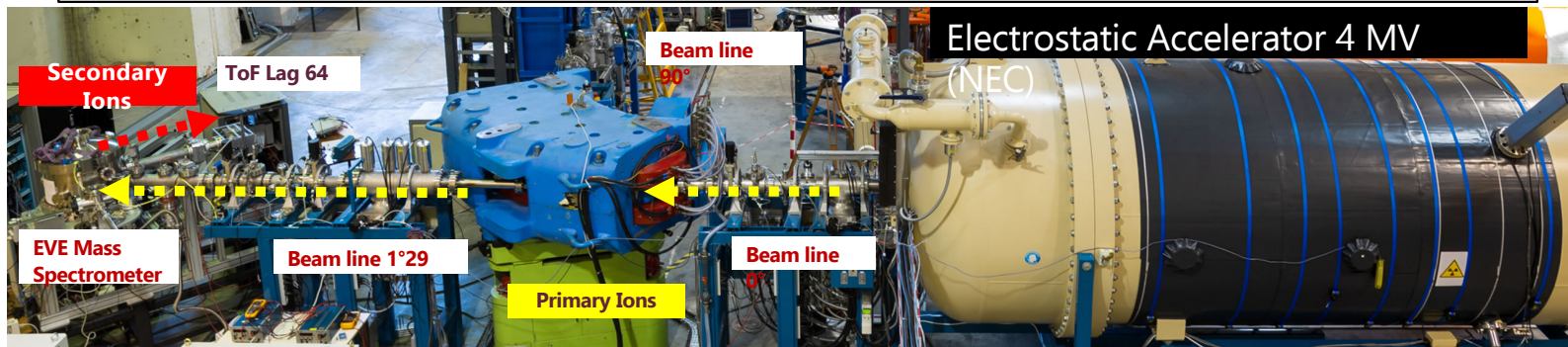
LPSC Grenoble: Julien Angot

Contact : thomas.zerguerras@ijclab.in2p3.fr



Contexte: spectrométrie de masse

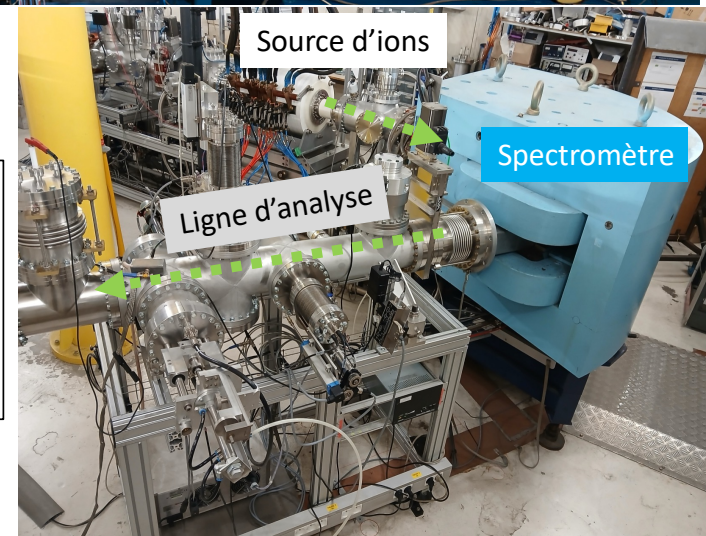
Plateforme MOSAIC Andromède (ANR: 10-EQPX-0023) @ IJCLab : SIMS-TOF



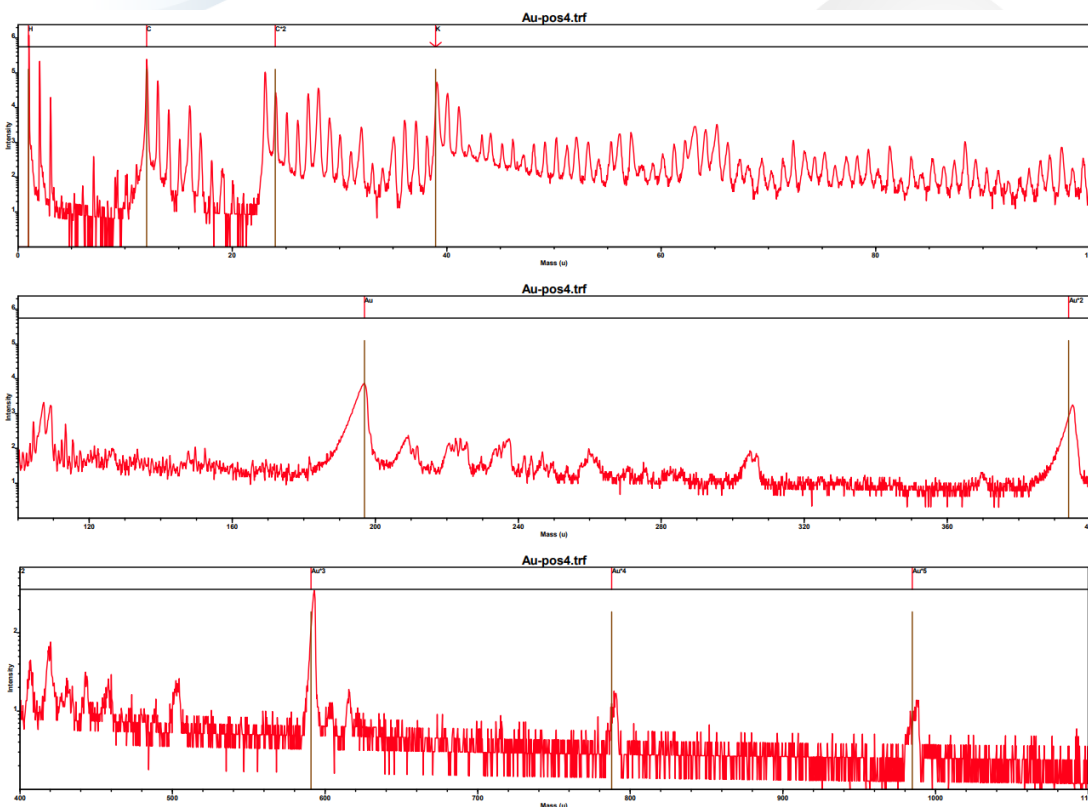
Source d'ions:
LMIS (AuSi)

Ligne 1+-N+: diagnostic pureté de sources pour accélérateurs

- **Source d'ions (Booster):** ECR, 14.5GHz, 100-800W, HT: 20kV
- **Spectromètre:** 120°, rayon 0.75m
- **Mesure de courant:**
 - Coupelle de Faraday avec anneau électrostatique
 - Fentes et channeltron



Applications: physique nucléaire, astrophysique nucléaire, astrochimie, science des matériaux, biologie ...



- Identifier les pics de manière **rapide** et **automatique**
- **Développer des modèles prédictifs d'attribution des pics :**
 - rapport isotopique, reconnaissance de motif ou famille de pics
 - fragmentation, clusters
 - autres contaminants, substrat

Difficultés :

- Obtention de données homogènes depuis des sources diverses
- Complexité des spectres de masse (plusieurs centaines de pics par spectre)
- Collecte et construction de métadonnées, standardisation
- Volume limité de données expérimentales



Données

Données expérimentales

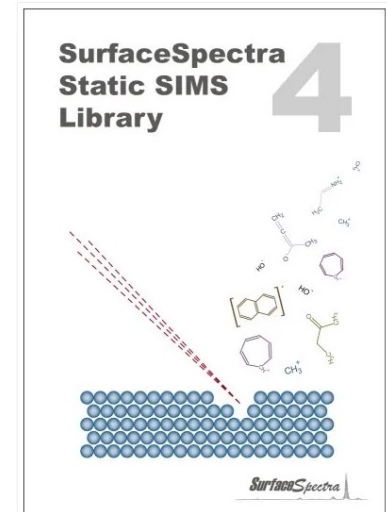
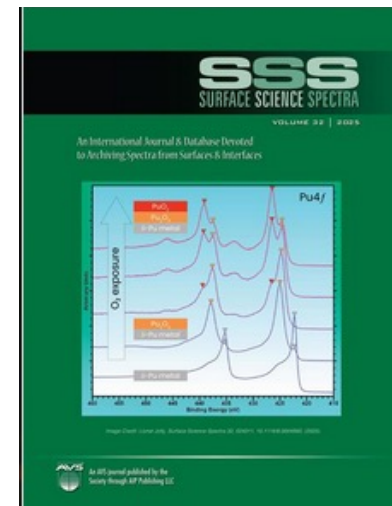
- MOSAIC/ANDROMÈDE
- LPSC Grenoble
- GANIL Caen (discussion en cours)

Données bibliographiques

- Surface Science Spectra

Base de données constructeurs

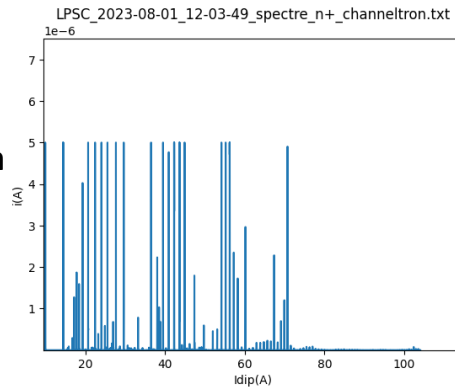
- Surface Spectra Static SIMS Library





Pré-traitement des données

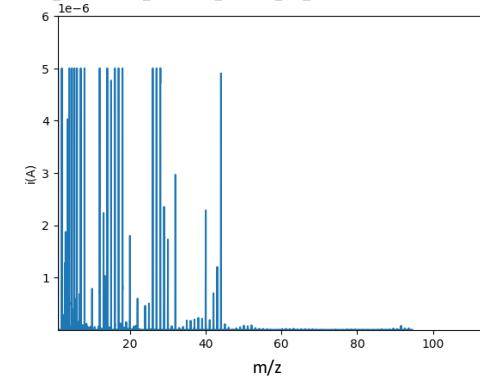
Calibration
m/z



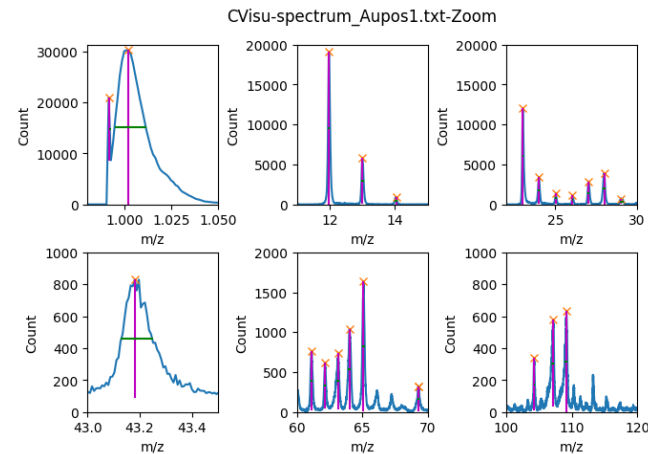
Fichier « Tags »

```
A;qX;1;n;Species;Remark;Interpolation
12;1;36;-0.00411E+0;4.999886E-6;12.000000;C;
12;2;25;-0.00411E+0;5.000000E-6;12.000000;C;
12;3;20;-0.00411E+0;5.000000E-6;12.000000;C;
12;4;17;-0.00411E+0;5.000000E-6;12.000000;C;
12;5;15;-0.00411E+0;5.000000E-6;12.000000;C;
14;1;39;-0.00411E+0;4.000000E-6;14.00387401;N;
14;3;22;-0.00411E+0;5.000000E-6;14.00387401;N;
14;4;19;-0.00411E+0;4.000000E-6;14.00387401;N;
14;5;17;-0.00411E+0;1.277704E-6;14.00387401;N;
16;2;29;-0.00411E+0;5.000000E-6;15.99491464;O;
16;3;24;-0.00411E+0;5.000000E-6;15.99491464;O;
16;5;18;-0.00411E+0;1.000000E-6;15.99491464;O;
16;6;16;-0.00411E+0;2.9132700E-9;15.99491464;O;
40;1;67;-0.00411E+0;2.282377E-6;39.9623831;Ar;
40;2;47;-0.00411E+0;1.000000E-6;39.9623831;Ar;
40;3;38;-0.00411E+0;1.000000E-6;39.9623831;Ar;
40;4;33;-0.00411E+0;7.0350730E-6;39.9623831;Ar;
40;6;26;-0.00411E+0;6.7554603E-6;39.9623831;Ar;
40;7;24;-0.00411E+0;5.8420102E-6;39.9623831;Ar;
40;8;23;-0.00411E+0;3.9000000E-6;39.9623831;Ar;
40;9;21;-0.00411E+0;6.100000E-6;39.9623831;Ar;
```

LPSC_2023-08-01_12-03-49_spectre_n+_channeltron.txt A/q calibrated



Recherche automatique et caractérisation de pics

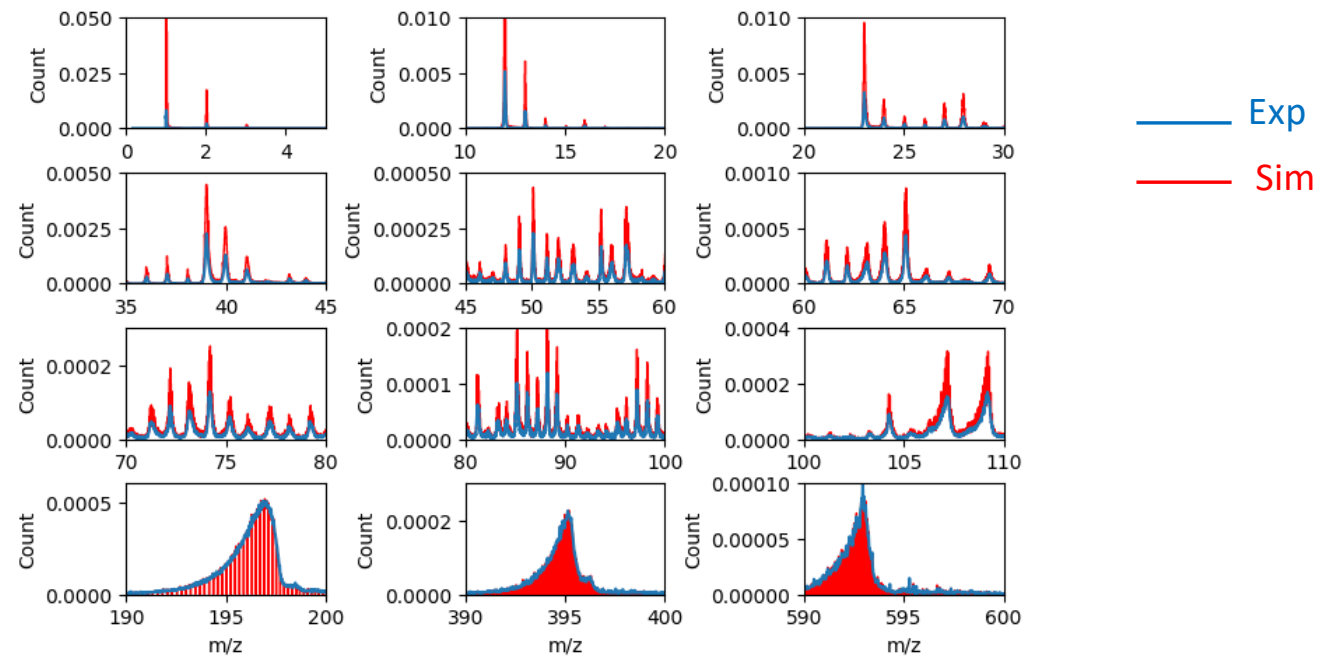




Simulations

- Objectif: augmenter la quantité de données
- Génération à partir de spectres expérimentaux (normalisation -> densité de probabilité)
- Intégration progressive de nouveaux échantillons de référence et introduction de perturbations (bruit, shift ...)

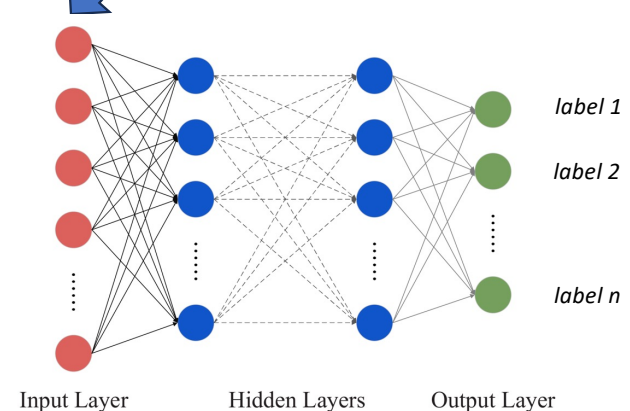
Travail en cours





Données (expérimentales, simulations)

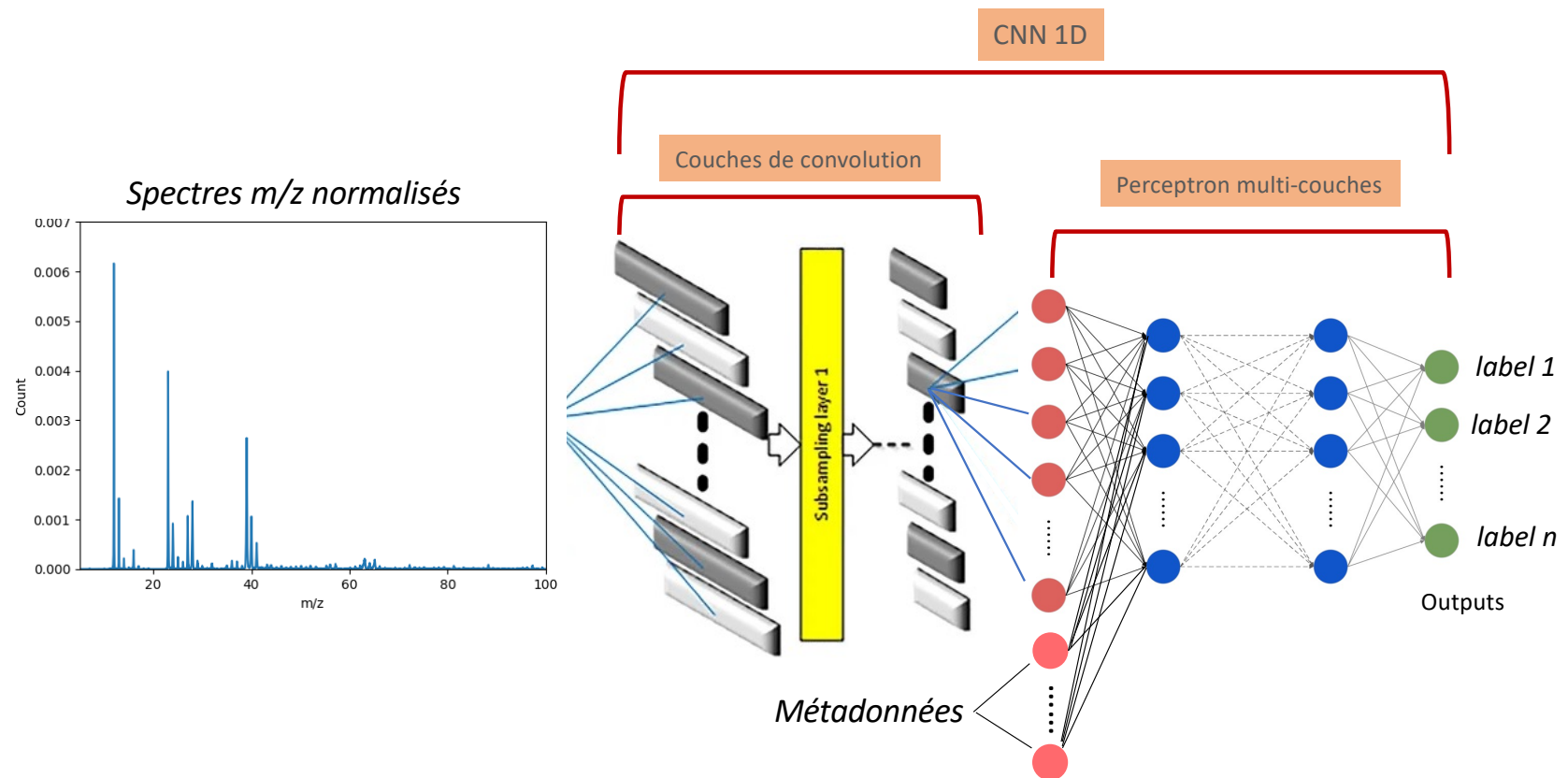
Données pour entraînement et validation



- **Utilisation d'un Perceptron Multi Couche (MLP) pour caractériser les pics**

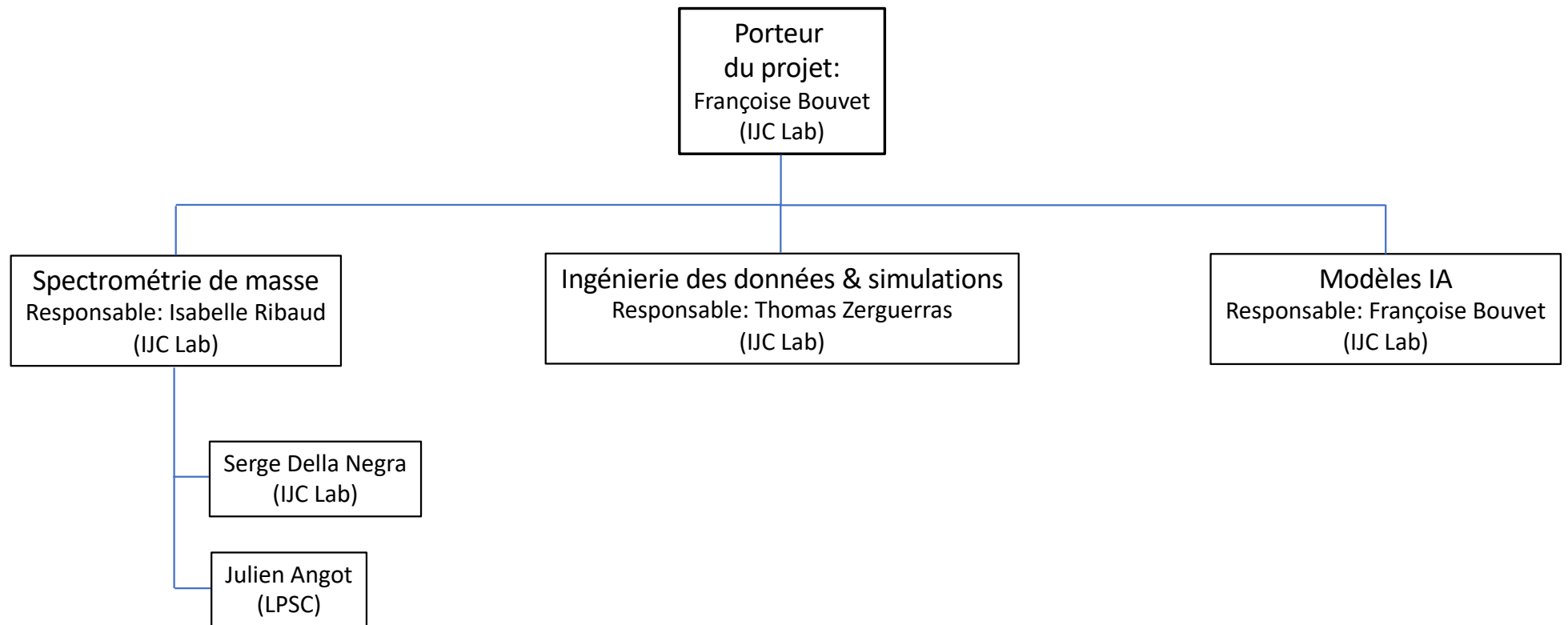


Option 2 : identification des pics à l'aide d'un réseau de convolution 1D





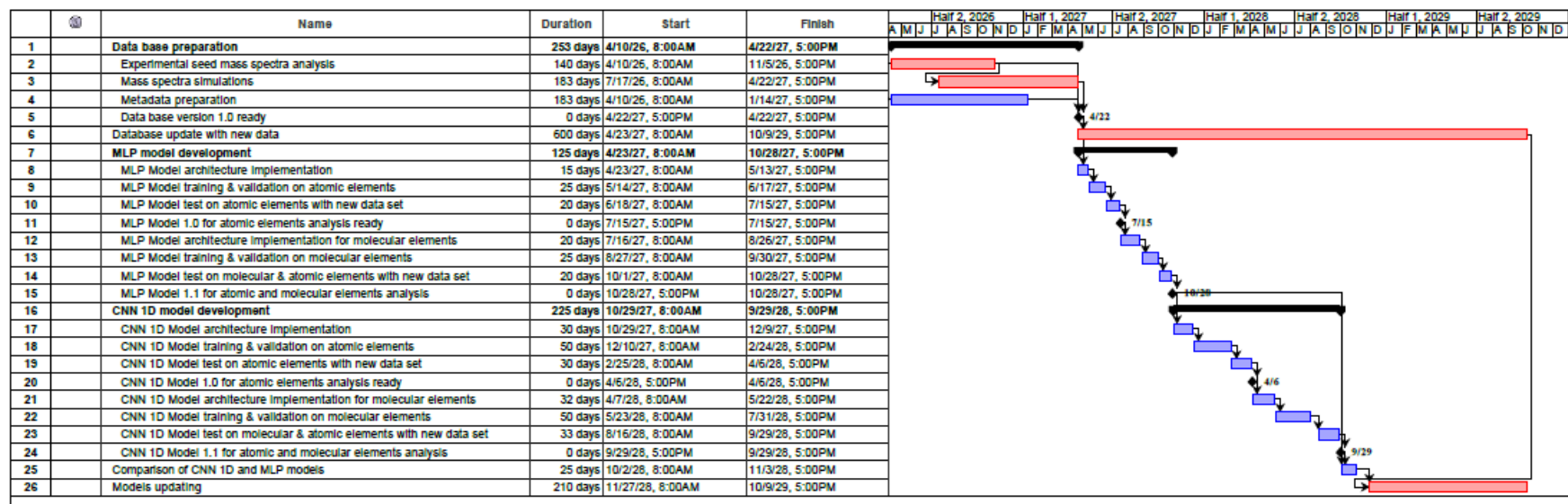
Organisation





Budget et calendrier (préliminaire)

- **Durée du projet R&T AIMS: 3 ans (2027-2029)**
- **Budget: ~40k€ sur 3 ans dont:**
 - ~27k€ : missions (réunions de travail, workshops, réseaux métiers, GDR, conférences ...)
 - ~13k€ :
 - Capacités de stockage
 - PC avec GPU (pré-test modèles IA)
 - Achat / renouvellement licence (accès à base de données TOF constructeurs)
- **Calendrier**





Conclusion - perspectives

Travail en cours

- Sélection des données expérimentales d'intérêt pour démarrer
- Pré-traitement et préparation des données
- Définition des métadonnées (infrastructure expérimentale, résolution du spectromètre, ...)
- Développement de modules Python pour écriture/lecture de métadonnées (format: imzML)
- Préparation fiche R&T AIMS

A venir

- Développement de la méthode de simulation pour augmenter la quantité de données
- Constitution de la BdD (Andromède, LPSC, Surface Science Spectra et/ou données constructeurs)
- Construction et comparaison des modèles :
 - Extraction des données puis MLP (option 1)
 - Réseau CNN 1D (option 2)
- Adaptation des modèles à des nouvelles données : piège à ion (ISOLTRAP @ CERN), MALDI (Matrix Assisted Laser Desorption Ionization) (I2BC @ Gif-sur-Yvette), ...



Merci de votre attention !

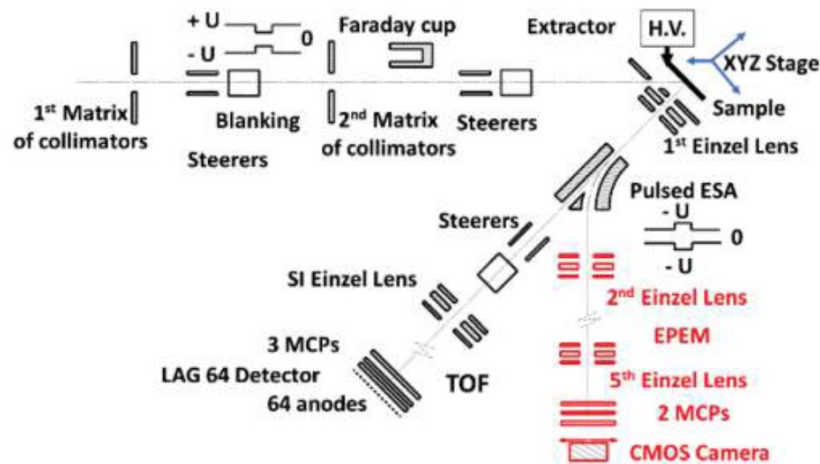


Back-up

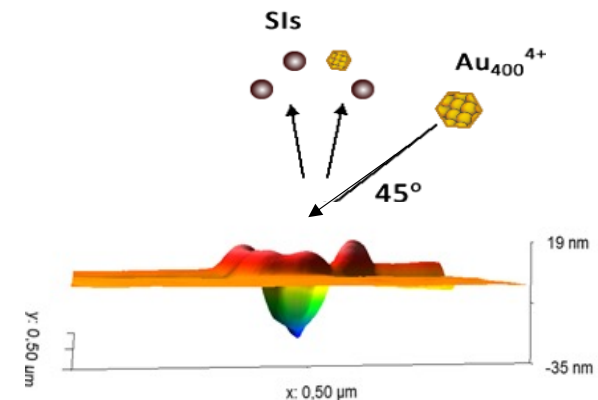


Spectrométrie de masse d'ions secondaires à temps de vol (ToF-SIMS)

Analyse de surface d'échantillons complexes à l'échelle nanométrique (composition et caractérisation structurale)



Experimental set-up



Ions primaires : Nanoparticules d'or multichargées accélérées dans le domaine du MeV , Au_{400}^{4+} 12MeV

Ions secondaires : atomes, molécules, fragments, clusters, adduits

Echantillons variés : organiques, inorganiques, métalliques

Applications : matériaux, biologie, micrométéorites, microélectronique



Prétraitement des données

Identification des pics sur le spectre en masse/charge

1. Conversion du temps de vol en m/z

$$\frac{1}{2}mv^2 = zV$$

$$\frac{m}{z} = \frac{2Vt^2}{L^2}$$

$$v = \frac{L}{t}$$

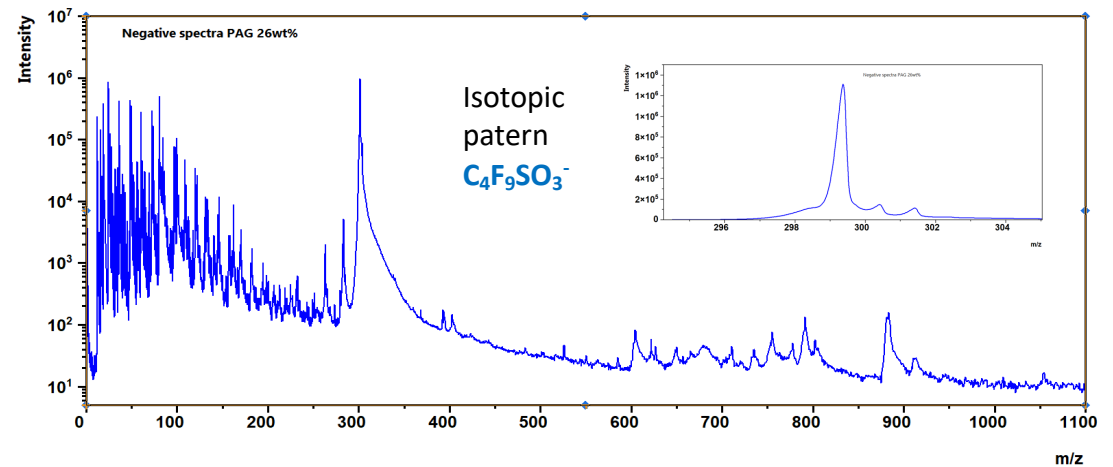
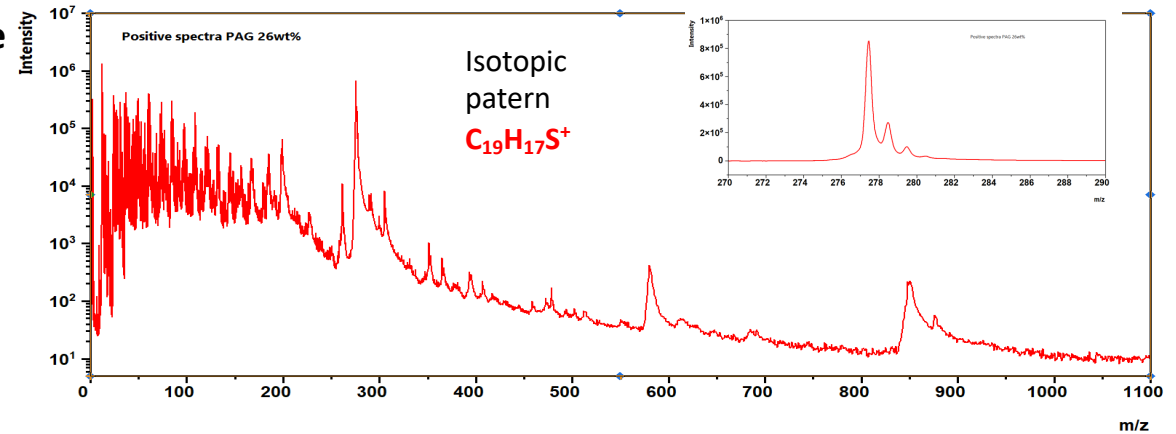
où :

- m est la masse
- v : vitesse
- z : charge
- V tension d'accélération
- L : longueur du temps de vol

$$\frac{m}{z} = (A \times tm + B)^2$$

2. Interpolation -> points régulièrement espacés

3. Normalisation par le nombre d'ions primaires

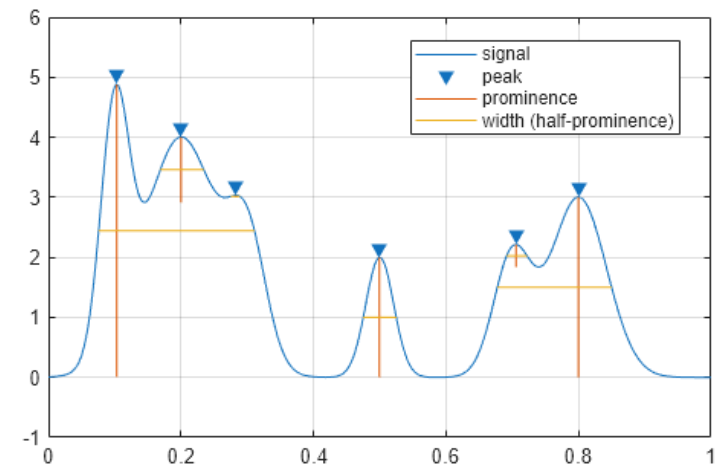




Option 1 : extraction et caractérisation des pics

Caractérisation d'un pic :

- **Position**
- **Hauteur**
- **Largeur à mi-hauteur**
- **Proéminence** : mesure à quel point un sommet se démarque de la ligne de base environnante du signal
- **Largeur** à une distance relative donnée de la hauteur du pic par rapport à la proéminence



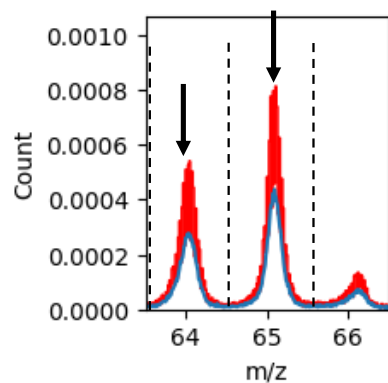
Recherche automatique de pics (bibliothèque SciPy)



Simulations: peaks ratio

Peak m/z = 64: [63.5; 64.5]

Peak m/z = 65: [64.5; 65.5]



— Measurement
— Simulation

Random generated histogram number of bins: 100 000

Peak ratio : 64/65

m/z	Sum Aupos1 Exp	Integral Aupos1 Exp	Sum Aupos1 Rnd	Integral Aupos1 Rnd
64	8,4656E-03	8,0126E-05	8,4872E-03	1,3008E-04
65	1,0823E-02	1,0329E-04	1,0789E-02	1,6549E-04
Ratio 64/65	0,782	0,776	0,787	0,786

Random generated histogram number of bins: 70 000

Peak ratio : 64/65

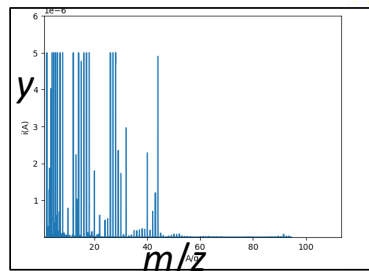
m/z	Sum Aupos1 Exp	Integral Aupos1 Exp	Sum Aupos1 Rnd	Integral Aupos1 Rnd
64	8,4656E-03	8,0126E-05	8,5437E-03	1,8704E-04
65	1,0823E-02	1,0329E-04	1,0791E-02	2,3647E-04
Ratio 64/65	0,782	0,776	0,792	0,791

- Integral calculations following the sum and the trapeze method are consistent.
- The peak ratio in the random generated spectrum histogram is consistent with that of the experimental spectrum



CNN 1D filter

Filter f



m/z spectrum

y

Convolution: $y * f$



Intérêt des utilisateurs de la plateforme / Collaborations

Collaborations en cours

- LPSC (Accélérateur et source d'ions)
- IonTOF (mode réflecteur)
 - Placamat (Plateforme Aquitaine de Caractérisation des Matériaux , université de Bordeaux)
- Piège à ions
 - ISOLTRAP (ISOLDE, CERN)

Utilisateurs d'Andromède

- IonTOF (mode réflecteur)
 - LAEC (Liban, Biologie et Astrochimie, collaboration avec IAS Orsay)
- MALDI (Matrix Assisted Laser Desorption Ionization)
 - I2BC (Biologie, CNRS-UPSay, Gif/Yvette)