

ACADÉMIE DE MONTPELLIER

UNIVERSITÉ DE MONTPELLIER

MASTER  
**STATISTIQUE POUR LES SCIENCES DE LA VIE**

MÉMOIRE sur le stage :

Développement d'outils d'analyses  
pour les données de PTR-ToF-MS

Du 1<sup>er</sup> mars au 31 août 2021  
au Centre d'Écologie Fonctionnelle & Évolutive  
sous la direction de N. Barthes et J.-M. Roger  
par :

JORIS HUGUENIN

soutenu le 10 septembre 2021, devant la commission d'examen :

J.-N. BACRO  
C. REYNES  
R. SABATIER

Professeur  
Maître de Conférence  
Professeur



École Pratique  
des Hautes Études





À travers ce rapport, je dédicace cette année de M2 à Thomas M. Kyle, lauréat du prix Ignobel 1991 pour la découverte de l'élément le plus lourd de l'Univers : l'Administratium. Je salue les efforts cumulés mis en oeuvre pour honorer le trentième anniversaire de ce prix et remercie ceux qui ont su faire  
preuves de patience.



<https://www.cefe.cnrs.fr/fr/>

CEFE : Centre Écologie Fonctionnelle et Évolutive  
PACE : Plateforme d'Analyse Chimique en Écologie

# Table des matières

<b>Table des matières</b>	<b>v</b>
<b>Remerciements</b>	<b>vii</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 Le métier d'ingénieur de plateforme</b>	<b>3</b>
1.1 Contexte . . . . .	3
1.2 Science ouverte . . . . .	4
1.3 Plan d'expérience en écologie chimique . . . . .	5
Détection des traces d'un complément alimentaire . . . . .	6
Caractérisation de COV de l'amandier . . . . .	7
Étude de l'émission journalière des lavandes et des figuiers . . . . .	8
1.4 Conclusion partielle . . . . .	9
<b>2 Présentation de la PTR-ToF-MS</b>	<b>11</b>
2.1 Une brève histoire de la PTR-ToF-MS . . . . .	11
2.2 Instrumentation générale . . . . .	12
présentation des spectres . . . . .	13
2.3 Technique rivale ou alliée de la GC-MS? . . . . .	14
2.4 package proVOC . . . . .	15
Explication général . . . . .	15
analyse AUC . . . . .	16
Calibration, regression linéaire et rapport automatisé . . . . .	17
Perspective . . . . .	18
<b>3 Méthodes Multivariées</b>	<b>21</b>
3.1 Alignement et sélection de variable . . . . .	21
3.2 Analyses Chimiométriques . . . . .	22
Analyse en Composante Principale (ACP) . . . . .	22
Multivariate Curve Resolution (MCR) . . . . .	23
Independent Components Analysis (ICA) . . . . .	24
3.3 Applications . . . . .	25
bourgeon . . . . .	25
lavande . . . . .	25
<b>4 Conclusion</b>	<b>27</b>
<b>Glossaire</b>	<b>29</b>
<b>Bibliographie</b>	<b>31</b>

Table des figures	35
-------------------	----

# Remerciements

Merci à tous



# Introduction générale

À l’instar de l’année 2021, ce stage de M2 a été un peu particulier. Lors de l’élaboration du sujet en janvier, nous avons proposé un sujet résolument tourné vers l’analyse de données. Il s’est avéré que mon temps de stage s’est réparti différemment. Le sujet du stage a évolué afin de mieux couvrir l’ensemble du travail effectué durant ces mois. Le chapitre 1 rend compte de ces aspects en détaillant le contexte de ce stage, l’ouverture qui a été amorcée afin de rendre compatible le travail de la plateforme avec les orientations publiques vers la science ouverte et de la relation entre les analystes et les praticiens.

Le chapitre 2 est dédié à l’instrument de mesure principalement utilisé lors du stage. Le PTR-ToF-MS analyse les molécules volatiles. Après une présentation de cette technique et de notre appareil, je détaillerais les données récupérées ainsi que le package R que j’ai écrit pour procéder à l’analyse. Nous ferons un comparatif entre cette technique et la GC-MS qui est régulièrement utilisé en parallèle.

Enfin, le chapitre 3 détaille les opérations mathématiques utilisées pour l’analyse des données. Je développerais les choix fait quand aux prétraitements et à l’alignement des spectres. Je présenterais également les algorithmes implémentés dans le package, en particulier celui de la MCR-ALS.

À la fin de ce rapport se trouve un glossaire des termes techniques et des acronymes. Avant celui-ci, je livre une conclusion générale sur ce stage ainsi que sur le travail qu’il reste à accomplir. Un soin particulier a été apporté à la découpe des chapitres. Bien que la mode des rapports et articles scientifiques soit au storytelling<sup>1</sup>, les trois chapitres suivant peuvent se lire dans l’ordre d’intérêt du lecteur.

---

<sup>1</sup>mode que j’ai toutefois essayé de respecter.



# Le métier d'ingénieur de plateforme

## 1.1 Contexte

Depuis le 04 novembre 2019, j'occupe un poste d'ingénieur affecté à la Plateforme d'Analyses Chimiques en Écologie (PACE). La PACE, créée en 2000, est spécialisée dans l'analyse chimique pour la communauté de recherche liée à l'écologie, l'environnement et la biodiversité. La PACE comporte six permanents et un CDD. Cette plateforme est un service mutualisé du LabEx CeMEB depuis 2011. L'unité de rattachement est le Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier (CEFE UMR5175 du CNRS). Cette unité mixte possède quatre tutelles; le CNRS, l'Université de Montpellier (UM), l'École Pratique des Hautes Etudes (EPHE) et l'Institut de Recherche pour le Développement (IRD), ainsi que trois partenaires; l'Université Paul Valéry Montpellier 3 (UPVM3), SupAgro Montpellier et l'INRAE. À ce titre, bien que rattachée au CEFE, la PACE accueille des projets de recherche issus des 12 unités du CeMEB (850 personnels permanents) mais aussi de toute la communauté académique nationale ou internationale.

L'Université de Montpellier a été porteuse d'un projet GEPETOs dont l'objet était de financer un projet CPER (Contrat Plan Etat Région) mobilisant des financements FEDER pour l'achat d'un spectromètre de masse en temps réel permettant d'atteindre des mesures de cinétique fine, abrégé en PTR-ToF-MS (Proton Transfert Reaction Time of Flight Mass Spectrometry) pour la PACE. Le chapitre 2 revient en détail sur cet instrument. Ce projet GEPETOs comporte une part dédiée au recrutement d'un ingénieur d'étude devant bénéficier d'une formation en double compétence de niveau Master 2. Avec mes responsables hiérarchiques nous avons choisi le M2 SSV qui me permettait de renforcer mes connaissances statistiques. Le stage du second semestre c'est donc effectué sur la plateforme PACE.

Plusieurs étapes ont été définies pour l'ensemble du contrat :

- mise en place des procédures de recueil et de contrôle des données;
- adaptation des méthodes d'analyse mathématiques pour répondre aux besoins spécifiques du PTR-ToF-MS;
- organisation de la mise en forme et du stockage des données;
- assurer la maintenance des bases de données contenant les data produites par l'instrument et les résultats des analyses.

Il était prévu que le point 2 soit particulièrement mis en avant lors de ce stage. En réalité

l'ensemble des points ci-dessus a été mobilisé. Cette introduction du contexte me permet de décrire l'état d'esprit général lié à ce stage. Effectué sur mon lieu de travail, il y avait une dualité entre la nécessité d'encadrer les utilisateurs de la plateforme et la possibilité offerte d'un temps de réflexion et d'exploration. La première modalité imposait le tempo, du pragmatisme et un calendrier à respecter. Ce printemps-été 2021 était particulièrement attendu pour deux raisons. Il fallait en quelque sorte rattraper le travail qui n'avait pu être effectué lors du confinement général du printemps 2020. De plus, les possibilités de l'appareil commencent à être connues et celui-ci est beaucoup demandé. De plus, l'intitulé initial de ce stage était *"Analyse descriptive des données PTR-TOF-MS des COV émis lors du cycle larvaire de la guêpe de l'amande (Eurytoma amygdali)"*. Je devais accompagner une doctorante dans son travail d'analyse. Nous avons commencé les expériences dès février lors de la période de floraison des amandiers. Après deux mois d'expériences, la quantité de données recueillies était particulièrement importante. La doctorante, ayant commencé sa thèse en janvier et après avoir les expériences avec le timing imposé par la nature, a pu reprendre un rythme normal de thèse et se plonger dans la bibliographie.

De mon côté, j'ai participé à la mise en place d'autres expériences sur lesquelles nous reviendront dans ce rapport. J'ai également pu prendre du recul sur mon activité et la repenser afin qu'elle cadre avec la politique actuellement de science ouverte. Ce stage m'a ainsi permis d'orienter les outils numériques que je met à disposition des chercheurs vers une science reproductible.

## 1.2 Science ouverte

Durant cette période de stage, le gouvernement français a mis en place le Deuxième Plan pour la science ouverte qui définit les actions mises en place sur la période 2021-2024. Le plan est détaillé dans ce [document](#) dont je recommande vivement la lecture. Ce guide de route possède les défauts propres à l'époque<sup>1</sup>. Ceci étant, à l'échelle de la communauté scientifique française, ce plan permettra une large diffusion du savoir et de la culture scientifique grâce à des actions regroupés dans quatre axes :

- généraliser l'accès ouvert aux publications ;
  - généraliser l'obligation de publication en accès ouvert des articles et livres issus de recherches financées par appel à projets sur fonds publics ;
  - soutenir des modèles économiques d'édition en accès ouvert sans frais de publication pour les auteurs ;
  - favoriser le [multilinguisme](#)<sup>2</sup> et la circulation des savoirs scientifiques par la traduction des publications des chercheurs français ;
- structurer, partager et ouvrir les données de la recherche ;
  - mettre en œuvre l'obligation de diffusion des données de recherche financées sur fonds publics ;
  - créer Recherche Data Gouv, la plateforme nationale fédérée des données de la recherche ;

---

<sup>1</sup>pêle-mêle l'utilisation abondante de mots valises, la volonté de récompenser les bons élèves par des badges et des prix ou encore permettre la libre exploitation de la recherche publique à des entreprises qui ne respectent pas leur devoir fiscal

<sup>2</sup>Découvrir l'étrangeté d'une pensée en langue étrangère constitue une expérience herméneutique fondamentale, étant donné qu'une telle rencontre peut déclencher un processus qui contribue à augmenter l'incertitude positive et à remettre en cause les convictions propres pour construire ainsi une barrière contre l'ethnocentrisme dans la pensée scientifique. HAMEL [[Ham13](#)]

- promouvoir l'adoption d'une politique de données sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables (FAIR) ;
- ouvrir et promouvoir les codes sources produits par la recherche ;
  - valoriser et soutenir la diffusion sous licence libre des codes sources issus de recherches financées sur fonds publics ;
  - mettre en valeur la production des codes sources de l'enseignement supérieur, de la recherche et de l'innovation ;
  - définir et promouvoir une politique en matière de logiciels libres ;
- transformer les pratiques pour faire de la science ouverte le principe par défaut.
  - développer et valoriser les compétences de la science ouverte tout au long du parcours des étudiants et des personnels de la recherche ;
  - valoriser la science ouverte et la diversité des productions scientifiques dans l'évaluation des chercheurs et enseignants-chercheurs, des projets et des établissements de recherche ;
  - tripler le budget de la science ouverte en s'appuyant sur le Fonds national pour la science ouverte et le Programme d'investissements d'avenir

Les axes concernant le code et les données me concerne directement. Par mon travail, je produit des données ainsi que le code permettant l'analyse de celle-ci par mes collaborateurs. Le code est écrit en langage R sous la forme d'un package, *proVOC*, déposé sur [gitHub](#). Il génère un workflow qui sauvegarde les paramètres et les données utilisés lors de l'analyse. Le dépôt sur un git permet l'ouverture de ce code. Il est toutefois prévu d'effectuer une migration sur gitLab afin de favoriser l'utilisation des logiciels libres.

Par ailleurs, je souhaite développer un module permettant de convertir les data machines dans un format adapté à la futur plateforme Recherche Data Gouv. Ce module facilitera la production de *data paper* dont j'ai découvert l'existence au cours des [journées casuHAL 2021](#). Grâce à deux demi-journées de formation sur Rmarkdown et sur la création de package, j'ai pu intégrer la génération de rapport automatisé dans le package<sup>3</sup> qui est un premier pas vers l'écriture automatisé des data paper.<sup>4</sup>

J'ai également découvert et testé les packages et templates permettant de créer sous R un document découpé en chapitre. J'utilisais déjà Rmarkdown depuis un certains temps mais le passage à [Bookdown](#) et [memoiR](#) a demandé un temps d'adaptation plus long que prévu. Ceci étant, ça n'a pas été en vain puisque j'ai eu plusieurs retours enthousiastes de collègues en vu d'une formation sur ces packages. J'envisage, une fois ce rapport terminé, de développer un clone de [memoiR](#) (avec l'accord de l'auteur) sur le style de [EcoFoG](#) afin de mettre à disposition aux membres du CEFÉ un templates clé en main. Ceci permettra de favoriser l'écriture de rapport par des logiciels libres et d'améliorer la reproductibilité de la science.

## 1.3 Plan d'expérience en écologie chimique

Le chapitre suivant décrit la technique de la PTR-ToF-MS. Nous avons ici juste besoin de savoir que cette technique analyse en continue les molécules volatiles émises par l'échantillon. L'instrument ne peut analyse que un échantillon à la fois mais plusieurs peuvent se succéder au sein de séquence.

<sup>3</sup><https://daranzolin.github.io/2021-03-03-automated-rmarkdown/>

<sup>4</sup>voir également BOETTIGER et al. [Boe+15]

Durant la période de stage, j'ai accompagné plusieurs séries d'expériences. Les modèles biologiques analysés varient à chaque expérience. Les plans d'expériences ont été conçu en dialoguant avec les utilisateurs. Bien souvent, ils avaient en tête l'expérience qui répondait le mieux possible à leur question biologique. J'anticipais les incohérences techniques et j'optimisais les paramètres de l'instrument. Avec le recul, je me suis rendu compte que mes connaissances acquissent en cours de recueil planifié de données n'étaient pas pleinement exploité.

Un des problèmes principales est dû à la contamination de la ligne et de l'instrument par l'échantillon précédent. Nous pouvons analyser ces expériences pour repérer les erreurs commises.

### Détection des traces d'un complément alimentaire

Cette expérience avait pour objectif de mesurer la diffusion de l'odeur de nourriture (pour des poules) dans quatre sites d'un mésocosme constitué de deux compartiments. La figure 1.1 illustre cette prise de mesure. Les sites étaient localisés dans la mangeoire (s1, compartiment 1), au centre du compartiment 1 (s2), à la jonction des deux compartiments (s3, côté compartiment 2, figure 1.2) et au fond du compartiment 2 (s4). La nourriture au sein de la mangeoire était soit de la nourriture témoin (T) soit dopée par de l'huile essentielle (A et B). Il y avait 5 mésocosmes pour chaque modalité.



FIG. 1.1 : PTR-ToF-MS relié à l'un des quatre site de mesure d'un mésocosme

Dans la théorie d'une expérience totalement randomisée, il aurait fallut définir l'ordre aléatoire de chacune des 60 acquisitions<sup>5</sup>. Dans la pratique, cette expérience nécessitait de déplacer l'instrument d'un mésocosme à l'autre entre chaque acquisition. Cette manipulation

<sup>5</sup> 5 (mésocosmes) x 3 (modalité) x 4 (site d'acquisition) = 60

aurait drastiquement augmenté la durée de l'expérience. Nous avons donc tiré au hasard l'ordre de passage de chaque mésocosme, effectué un blanc et les acquisitions sur les quatre sites de mesures puis terminé par un nouveau blanc.

Il aurait été juste de tirer au hasard l'ordre de passage des sites. Pour autant, nous avons opté pour une hypothèse forte de commencer par la site le plus éloigné de la mangeoire (s4, s3, s2, s1) afin que les potentielles traces ne soient pas contaminées par une odeur plus forte. Dans les faits, nous laissons assez de temps entre deux mesures pour renouveler plusieurs fois le volume d'analyse à l'intérieur de l'instrument ce qui permettait d'isoler les mesures. Après coup, je pense qu'il aurait été bon de procédé malgré tout à une randomisation de l'ordre des sites au sein d'un mésocosme.



FIG. 1.2 : Sonde à proximité de la mangeoire

Nous avons également une seule journée pour effectuer l'ensemble des acquisitions, ce qui a été juste suffisant. Nous n'avons pas pu effectuer de répétition sur un même échantillon. L'analyse des data a été effectuée par le post-doc en charge de l'expérience. Je n'ai pas de retour sur des potentiels contaminations intra et inter échantillons.

Bien que la réalisation de cette expérience est statistiquement discutable, elle a été réalisée en tout début de la période de stage et m'a permis la réflexion sur comment intégrer mieux la planification d'expérience.

## Caractérisation de COV de l'amandier

Ce travail s'inscrit dans la première année d'un projet de thèse et consistait à analyser les COV émis par les différents organes (bourgeon, fleur, feuille, fruit) des amandiers. Pour

des raisons de calendrier, j'ai aidé à la mise en place des expériences début février puis à la formation de la doctorante. Une fois cette dernière autonome avec l'instrument, j'ai peu assisté au suivi de l'expérience. J'ai toutefois soumis l'idée de suivre le débourrement d'une fleure avec le PTR-ToF-MS.

Nous avons suivi durant un week-end les COVs émis par un bourgeon que nous avons isolé du reste de la plante. Ce bourgeons avait été sélectionné car il était proche de la floraison. Ainsi, nous avons pu recueillir les odeurs émises avant, pendant et après le débourrage. Cependant, nous n'avons pu obtenir qu'une seule unité expérimentale pour cette expérience. Par ailleurs, nous avons filmé le bourgeons durant l'expérience mais la caméra s'est éteinte après quelques heures. Ainsi, nous n'avons pas de témoin validant l'heure de phénomène. Les données techniques et les résultats de l'analyse sont détaillés à la section 3.3. Nous retenterons probablement cette expérience en février 2022 sur plusieurs unités expérimentales, en privilégiant une série de photo plutôt qu'un film.

Cette expérience a permis de mettre en avant les capacités de l'appareil et a fait germer plusieurs idées. Nous avons fait plusieurs réunions préparatoires pour mettre en œuvre l'une d'elles. Nous souhaitons analyser la réponse chimique d'une fleure après la diffusion du bruit de son pollinisateur.

## Étude de l'émission journalière des lavandes et des figuiers

L'expérience sur les lavandes présentée ici s'inscrit dans un cadre expérimentale plus vaste s'intéressant aux phénomènes liées à l'ozonation. Je ne reviendrais pas dans ce rapport sur la partie ozonée. Cependant, nous nous sommes intéressé aux variations journalières dans des conditions non ozonée (plantes témoins). Nous avons mis en place quatre chambres de mesure permettant chacune soit d'isoler un plant de lavande soit d'être un blanc. Ces chambre de mesure sont des cylindres en plastiques, ventilé, de volume équivalent et relié chacun à un flux d'air propre d'un débit de 5 litres/min. La longueur et le volume des lignes reliant les chambres à l'instrument de mesure n'étaient pas identique. Nous avons lancé des acquisitions sur les quatre chambres en série durant 48h. Nous avons répété ce processus deux fois, pour obtenir neuf plants au total dans notre plan d'expérience. Pour les lavandes, il n'y avait qu'une modalité (plante témoin). Toutefois, un des quatre moteurs permettant la ventilation à l'intérieur de la chambre ne fonctionnait pas. Nous avons décidé que cette chambre serait le blanc. Là encore, cette décision est critiquable. D'un côté, il fallait que le moteur HS soit celui de la chambre "blanc" afin de ne pas créer artificiellement les modalités "plante témoin ventilée" et "plante témoin non-ventilée". D'un autre côté, nous comparons un blanc non-ventilé à des plantes ventilées. Au regard de débit d'air pure de 5L/min, nous avons jugé ce biais négligeable. En revanche, nous aurions pu, au prix de quelques branchements et rebranchements supplémentaires, déplacer le moteur défectueux d'une chambre à l'autre entre les répétitions. Ceci aurait permis de mieux répartir l'erreur expérimentale due aux lignes de mesure. Ceci étant, les résultats présentés dans la section 3.3 sont satisfaisant.

Quelques jours après l'expérience avec les lavandes, nous avons réutilisé le même dispositif avec des figuiers. Le temps d'analyse a été augmenté pour passer à 5 jours. La chambre "blanc" est resté la même (non ventilée). Cette fois, les plantes étaient soit "polinisées" soit "non-polinisées" (témoin). Nous avons décalé la plante témoin d'une chambre à chaque répétition malgré qu'un statisticien puriste aurait probablement tiré au sort à chaque répétition, au risque de tomber trois fois sur la même chambre. Je n'ai pas aidé à l'analyse de cette expérience et ne peut pas me prononcer sur la pertinence des choix effectués.

## 1.4 Conclusion partielle

J’ai commencé ce métier en novembre 2019. En raison des conditions sanitaires et de la formation continue en M2 SSV, j’ai une vision légèrement biaisée d’une année “standard” sur ce poste. Pour autant, je pense avoir trouvé le bon équilibre entre d’un côté le développement d’outils numériques résolument orienté utilisateurs et qui entre dans le Deuxième Plan national pour la science ouverte et de l’autre côté l’accompagnement des utilisateurs de la conception de leurs expériences à l’acquisition[AB20].

Le temps annuel consacré à chacune de ces facettes est plutôt équilibré bien que les saisons chaudes soient bien évidemment plutôt dédiées au terrain. Après ces quelques descriptions d’expériences, je propose une description plus détaillée de l’instrument PTR-ToF-MS et de son environnement dans le champs de la spectroscopie de masse, chapitre 2, mais un lecteur pressé de connaître les conclusions de ces expériences peut sauter au chapitre 3.



## Présentation de la PTR-ToF-MS

De très nombreux projets de recherches issus du CeMEB nécessitent l'analyse de Composés Organiques Volatiles (COV ou VOC en anglais) aussi communément appelés “odeurs”. Les COVs sont omniprésents dans la nature et permettent, avec les autres sens, une organisation du vivant en agissant comme vecteurs d'informations de la médiation chimique. Par l'acquisition d'un PTR-ToF-MS, la communauté souhaitait lever trois verrous techniques :

- l'appareil permet un fonctionnement en flux continu avec une résolution temporelle très fine (fréquence d'analyse allant jusqu'à 10 scans par seconde), permettant ainsi le suivi des cinétiques d'émissions de COV. Cet instrument simplifie drastiquement et affine les expériences cinétiques effectuées par des mesures ponctuelles de GC-MS ;
- le PTR-ToF-MS possède une excellente résolution en masse qui facilite l'identification des molécules. Un spectre est composé de plus de 140 000 mesures de masses couvrant une large gamme des masses des COV biologiques (de 70 à 500 m/z). L'appareil possède donc une résolution environ mille fois supérieure par rapport à un simple quadripôle (PTR-MS) qui fournit des m/z à l'unité de masse atomique (uma ou Dalton Da) ;
- le seuil de détection extrêmement bas, de l'ordre du ppt (part per trillion), permet une sensibilité adaptée à la mesure de trace.

### 2.1 Une brève histoire de la PTR-ToF-MS

Durant la décennie 1990, le *Institut für Ionenphysik der Leopold-Franzens-Universität* d'Innsbruck en Autriche développe un nouvel instrument pour l'analyse chimique des gaz. En travaillant avec la *Universitätsklinik für Innere Medizin*, l'équipe menée par Lindinger développe un spectromètre en phase gazeuse qui ionise les molécules grâce à un gaz neutre ([LLA91] et [LHP93]). Ils comprennent rapidement l'avantage d'utiliser un proton (apporté à la réaction sous la forme  $\text{H}_3\text{O}^+$ ) par rapport aux ions  $\text{Kr}^+$  et  $\text{Xe}^+$ . Ils publient une sorte de preuve de concept en 1994 [Lag+94] et détaillent plus finement l'instrumentation et les réactions chimiques [Han+95]<sup>1</sup>. En 1998 sort l'article qui fait désormais référence, [LHJ98], signé par le trio Lindinger, Hansel et Jordan (et sa version courte [LJ98]).

---

<sup>1</sup>avec des remerciements incroyables

Ionicon commercialise la même année le premier PTR-MS. Cette technique se déploie rapidement dans les sciences de l’atmosphère, médicales et biotech ([BA01]) permettant la détection de traces de COV. Rapide et sensible, la PTR-MS élimine les désavantages de la GC-MS. Cependant, la PTR-MS détecte de la masse nominale des ions. Un échantillon complexe peut contenir plusieurs ions isobares<sup>2</sup>. Il convient alors de gagner en résolution de masse et séparer ces isobares. [Jor+09] revient sur les tentatives les plus concluantes ([Bla+04], [Enn+05], [Ino+06] et [Tan+07] ) avant de proposer son approche : la PTR-ToF-MS. Développé en collaboration avec la précédente équipe, le minutieux article de [GMH10] conclut superbement la vingtaine d’années de développement nécessaire à cette technique. Les années suivantes permettront simplement une amélioration des différentes parties de l’instrument.

Par ailleurs, c’est au tour des mathématiciens d’apporter leur contribution et en particulier à l’équipe de Cappellin avec deux articles. Le premier, [Cap+10], permet de mieux estimer la masse exacte de chaque ion détecté. Le seconde, [Cap+11], offre une méthodologie détaillée de l’utilisation du PTR-TOF-MS de l’acquisition à la fin de l’analyse. Puis en 2015, [Hol15] propose un logiciel pour l’analyse des data mais cette tentative n’a pas été suivie par la communauté d’utilisateur. Il existe désormais un nombre important d’articles utilisant la PTR-TOF-MS. Bien entendu, des outils commun de chimométries ont été utilisés pour renforcer les résultats, [Deu+19]. Mais cela sera abordé dans le chapitre suivant.

Par ailleurs, je suppose que la prochaine avancée majeure viendra d’une technologie permettant de discriminer les isomères<sup>3</sup> [Cla+21]<sup>4</sup>. Par exemple, cela permettrait de séparer les différents monoterpènes. Actuellement, ce désavantage est comblé par le couplage de la PTR-TOF-MS avec une seconde technique d’analyse chimique. Ce court historique montre toutefois que la collaboration entre physiciens, chimistes, biologistes et mathématiciens est fructueuse et promis à un bel avenir.

## 2.2 Instrumentation générale

Comme nous l’avons vu précédemment, le PTR-TOF-MS est conçu pour analyser finement la masse moléculaire des échantillons. Pour cela l’instrument aspire un débit d’air constant de 100 ml/min<sup>5</sup>. Cet *air in* sur la figure 2.1 peut être diluée par un gaz propre (pour éviter la saturation) ou étalon (pour la calibration) puis est injecté dans le réacteur (*FIMR* pour *focusing ion-molecule reactor*). Dans cette chambre, des ions  $H^+$  sont également injecté à débit constant afin d’ioniser les molécules de l’échantillons (*Proton-Transfer-Reaction*).

Il est important de comprendre que, contrairement à une ionisation par électron<sup>6</sup>, il n’y a quasiment pas de fragmentation moléculaire. En revanche les ions ont “une masse molaire de plus”(un proton) que les molécules initiales. Par exemple, le linalol, un COV fortement produit par la lavande, a une masse molaire de 154,249 g/mol mais sera détecté à 155,256 g/mol. Par ailleurs, le nombre d’ions  $MH^+$  dépend du taux de protonation de la molécule  $M$ . Sous forme d’ions, l’échantillon peut être focalisé par un champ électromagnétique généré par un quadripole et envoyé dans la partie TOF.

Une fois dans la colonne, l’énergie potentiel d’un pulse électrique à 25kHz est convertie en énergie cinétique par les ions propulsés. À énergie constante, la masse fait la différence lors de la mesure du temps de vol (TOF, figure 2.2). Ainsi les ions les moins lourds arrivent en

<sup>2</sup>molécules possédant un nombre de nucléons identique.

<sup>3</sup>molécules partageant la même formule brute mais avec une stéréochimie différente.

<sup>4</sup>si un mémoire de master permettait la science fiction, j’aurais évidemment développé quelques idées à propos de capteurs de champs électromagnétiques ou de spectromètres optiques à la place des capteurs MS actuels.

<sup>5</sup>ou sccm, pour *standard cubic centimeters per minute*

<sup>6</sup>comme c’est le cas avec la GC-MS

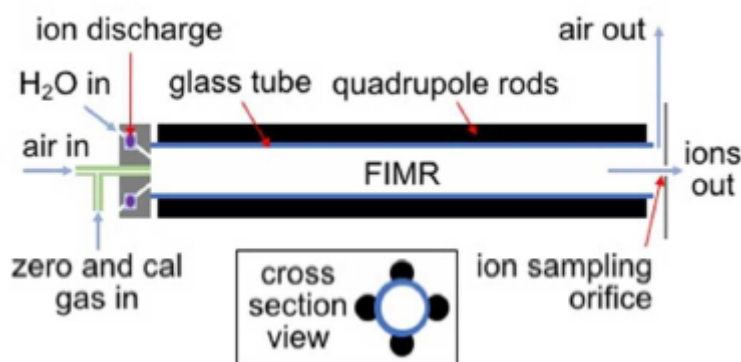


FIG. 2.1 : Schéma de la chambre d'ionisation (Reprinted with permission from KRECHMER et al. [Kre+18]. Copyright 2021 American Chemical Society.)

premiers sur le détecteur. Après calibration, une simple conversion temps de vol/ masse permet de connaître l'intensité de chaque masse d'un échantillon.

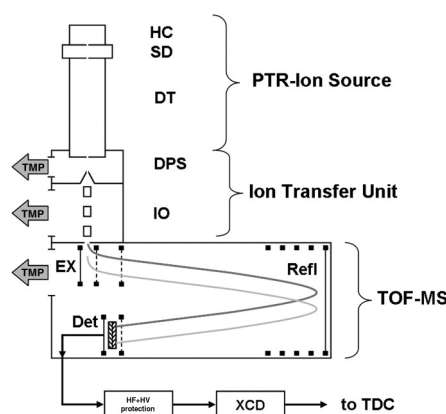


FIG. 2.2 : Schéma de la colonne TOF et du détecteur MS (Reproduit de l'article GRAUS et al. [GMH10], consultable ici : <https://www.sciencedirect.com/science/article/pii/S1044030510001005#fig1>

Pour une compréhension plus exhaustive, je recommande la lecture des articles de HANSEL et al. [Han+95] pour la partie PTR-MS, de GRAUS et al. [GMH10] pour la partie TOF et de KRECHMER et al. [Kre+18] spécifique à l'instrument utilisé au CEFÉ.

## présentation des spectres

Nous pouvons à présent nous intéresser aux spectres à analyser. J'ai représenté sur la figure 2.3 un spectre de trois plants de lavandes acquis autour de 8h du matin fin juin. Une figure identique mais dynamique peut se télécharger [ici](#)<sup>7</sup>. Cette figure dynamique permet d'effectuer des zoom afin d'avoir une compréhension plus précise qu'avec la simple figure 2.3.

Les pics sont extrêmement fins et centrés autour des masses unitaires. Cela s'explique facilement au regard de la masse des atomes composants les VOC (H, C, O pour l'essentiel) sur le [tableau de Mendeleiev](#)<sup>8</sup>. Il y a un plusieurs ordres de grandeur<sup>9</sup> entre les pics très intenses et ceux faibles. Cette intensité ne présage rien quand à l'importance de ce pic dans l'analyse.

<sup>7</sup>il faut télécharger le fichier puis l'ouvrir, par exemple avec un explorateur internet.

<sup>8</sup><https://www.youtube.com/watch?v=0lNuTSz6KVM>

<sup>9</sup>entre 10 et 10<sup>6</sup>

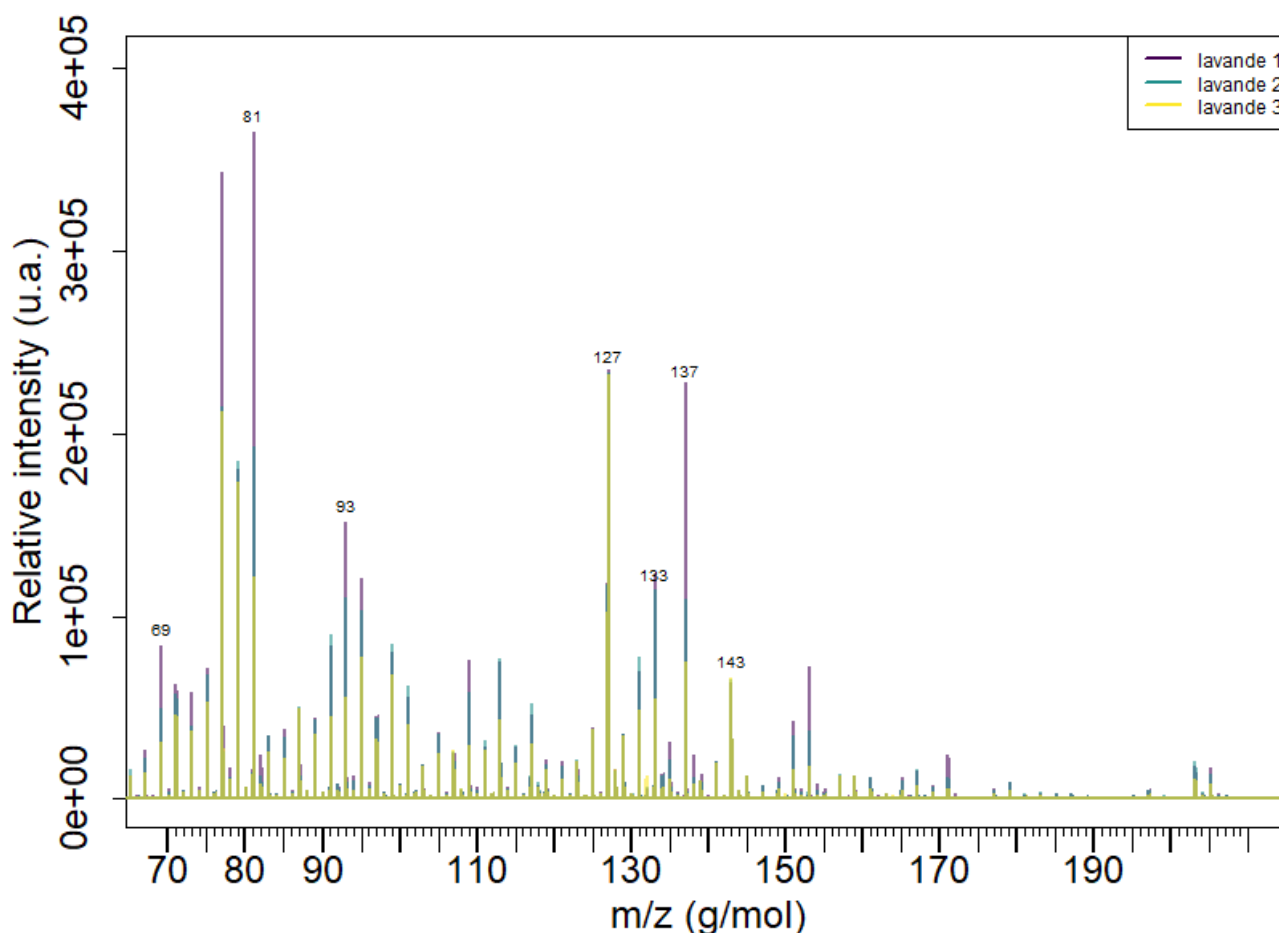


FIG. 2.3 : spectre MS

La plage spectral de l'instrument est de 70 à 500 g/mol. Cela est fixé par le constructeur pour des raisons techniques alors qu'il y a de nombreuses molécules d'intérêts de masses inférieures à 70g/mol et peu de molécules naturelles dépassant les 250 g/mol. Ces considérations prises en compte, nous obtenons pour chaque spectre un vecteur brute de 158 768 points. Ces spectres sont réduits, après alignement (voir section 2.4 et 3.1 ) à une taille inférieure à 50 000 points tout en gardant une résolution de  $10^{-3}$  g/mol. De plus, la PTR-TOF-MS a été développée afin d'analyser des cinétiques d'émissions. Dans nos cas, nous enregistrons des spectres de quelques secondes consécutivement durant plusieurs minutes à plusieurs jours. Pour une expériences, nous générons ainsi des dizaines, voir des milliers, de spectres. Ceci étant, nous pouvons dès à présent percevoir les problèmes liés aux temps de calculs que nous allons rencontrer et la nécessité de développer des outils optimisés pour analyser de tels objets.

## 2.3 Technique rivale ou alliée de la GC-MS ?

La métabolomique est dernier maillon de la chaîne -omic. Il se concentre sur l'analyse des petites molécules produites par un système biologique. Les deux domaines d'analyses sont la résonance magnétique nucléaire (NMR) et la spectroscopie de masse (MS). Pour l'analyse des VOC en spectroscopie de masse, la technique de GC-MS est clairement la plus utilisée à l'heure actuelle. La GC-MS, utilisé depuis des décennies, est largement implanté dans le paysage scientifique international. De plus, de nombreux logiciels performant<sup>10</sup> existent. Il est légitime

<sup>10</sup>MZmine pour n'en citer qu'un

de se poser la question de la nécessité d'utilisation des deux techniques<sup>11</sup>.

Premièrement, la GC-MS accumule les VOC dans des pièges chimiques<sup>12</sup> puis les analyses dans un second temps. La durée d'accumulation dans le piège est de l'ordre de la dizaine de minute. La PTR-TOF-MS a une sensibilité accrue permettant l'acquisition de spectres à une cadence de 25kHz. Afin d'augmenter le ratio signal sur bruit et de drastiquement réduire la quantité de data générée, ces spectres sont additionnés les uns aux autres sur des durées d'acquisitions plus longues<sup>13</sup>. En répétant ces acquisitions, la PTR-TOF-MS permet l'analyse cinétique d'un phénomène. Analyse statique contre analyse dynamique. Le prix de cette sensibilité est comme bien souvent une sélectivité dégradée. Ainsi la GC-MS, *a contrario* de la PTR-TOF-MS, peut discriminer les isomères. Cela est permis grâce à la fragmentation des molécules lors de l'étape d'ionisation par un électron qui intervient après la phase de chromatographie.

Bien que les molécules analysées soient les mêmes, les approches et résultats sont donc bien différents. De plus, il est souvent intelligent de croiser ces deux techniques [Maj+18]. De façon empirique, le PTR-TOF-MS analyse séquentiellement plusieurs échantillons. Il est aisé de collecter les VOC pour le GC-MS juste avant ou après une séquence dédiée au PTR-TOF-MS. Nous pouvons ainsi, par exemple, obtenir deux ou trois analyses ponctuelles de GC-MS lors d'une journée entière d'analyse PTR-TOF-MS. Nous combinons la sensibilité et la spécificité des deux techniques.

Les données obtenues en GC-MS sont donc plus complexes qu'en PTR-TOF-MS. Chaque échantillon est représenté par une matrice ayant une dimension pour le temps de rétention et une dimension pour le spectre de masse. Il n'est donc pas possible de réutiliser tels quels les outils numériques de la GC-MS. Cependant, il serait dommage de ne pas piquer les idées déjà existantes. J'ai donc créé un package R spécifique à l'analyse des données du PTR-TOF-MS mais en m'inspirant largement de ce qui se fait sur d'autres techniques : workflow (idée piquée à MZmine), bucketing (pratique courante dans la RMN), chimiométrie (largement utilisée pour l'analyse des spectres optiques) ; ou en utilisant des packages destinés à d'autres méthodes (MALDIquant dont la transparence du nom laisse peu de doute sur la méthode ciblée par ce package)

## 2.4 package proVOC

Comme expliqué dans la section sur la science ouverte ( 1.2), ce package peut se retrouver sur ce dépôt Git.

### Explication général

L'objectif de proVOC est de permettre une vue d'ensemble rapide<sup>14</sup> pour les utilisateurs pressés d'analyser *leurs* données sous R sans avoir besoin d'avoir une grande connaissance de ce langage. La fonction `import_sp()` constitue la première étape de l'analyse. Les données contenues dans des fichiers `.h5` générés par notre instrument sont importées dans l'espace de travail [Fis+21]. Après cette étape, les spectres sont automatiquement alignés puis réduits. Les détails à propos de l'alignement des spectres sont données dans la section 3.1. Pour l'étape de réduction, j'ai défini empiriquement de supprimer chaque colonne (masse) qui ne possédait aucune ligne (intensité) supérieure à un seuil. Ce seuil est calculé pour chaque jeu de donnée

<sup>11</sup>ou de façon moins prosaïque : pourquoi diable les physiciens ne restent pas sagement avec leurs bosons et viennent régulièrement embêter leurs collègues avec des idées nouvelles ?

<sup>12</sup>qui se présentent sous différentes formes en fonction des utilisations

<sup>13</sup>typiquement de l'ordre d'une à trente secondes

<sup>14</sup>proVOC : **P**erform a **R**apid **O**verview of the **V**olatile **O**rganic **C**ompound

comme égale à 20 fois l'écart-type médian de chaque masse<sup>15</sup>. Il est probable que cette étape de réduction peut être optimisée d'avantage mais je reviens sur les perspectives d'améliorations du package dans la section dédiée, 2.4.

Passé ces étapes, la fonction met en forme les données dans un objet *list*, récupère les metadata (date et heure d'acquisition) puis détecte les pics grâce à la fonction *detectPeaks()* du package **MALDIquant** et calcule les AUC grâce à la fonction *AUC()* du package **DescTools**. Enfin, cet liste nommé “sp” est sauvegardé en .Rdata. Ce travail préliminaire peut être long.

Une fois effectué, l'import permet à l'utilisateur de lancer facilement une série d'analyse pour explorer rapidement son jeu de donnée. Pour cela, il lui suffit de modifier des options dans un workflow :

```
# Workflow pour l'analyse :
mt <- list("h5" = mt_h5,                                # Don't touch
          "acq" = which(mt_h5$use4analysis == TRUE),      # Don't touch
          "wd" = sp$h5$wd,                                # Don't touch
          #-----#

[...]
```

```
# graphe pour visualiser les spectres de masses de chaque acquisition :
  "view_plot" = FALSE,      # Si TRUE, graphe moyen de chaque acquisition,
                           # sinon précisez le numero d'une ou plusieurs
                           # acquisition(s). (ex :
                           # c(200,300,400)
                           # mt_h5$end
                           # c(mt_h5$start, mt_h5$end)

  "view_each.group" = FALSE, # FALSE ou le nom d'une meta colonne
                           # ("XXX" de mt_h5$XXX) pour les graphes
                           # larges ou centres sur un seul pic.

  "view_plot_xmin" = 70,
  "view_plot_xmax" = 250,
  # ou ...
  "view_plot_peak" = FALSE, # c(59,73,81,93,109,127,137,143,153) ou
                           # False ou c(61,...) Permet de ne regarder
                           # qu'un seul pic

  #-----#
```

[...]

```
# ICA
"ICA" = FALSE,
"ICA_AUC" = TRUE,
"nc" = 6
#-----#
)

PTR_MS_analysis(sp, mt)      # debute les analyses
```

L'objet *mt* peut être sauvegardé pour se souvenir des paramètres utilisés lors de l'analyse. Nous pouvons détailler ci-dessous quelques options.

## analyse AUC

```
#-----#

# Aire sous la courbe (AUC) :
"AUC" = TRUE,
"AUC_plot_exp" = FALSE,      # Si TRUE, l'axe 'y' est exponentielle.
"AUC_each.group" = FALSE,    # Si l'argument est le nom d'une meta colonne
```

<sup>15</sup>cf. “calculation of threshold” dans la fonction *reduction2()*

```

# ( "XXX" de mt_h5$XXX) chaque groupe de cette
# colonne aura son graphe. Sinon FALSE.
"AUC_each.Mass" = TRUE,      # Si TRUE, un graphe pour chaque masse
"M.num" = c(81, 137, 153),  # Les masses analysees (utilise pour AUC)
"AUC_plot_dy" = TRUE,       # Si FALSE, plot exporter en .tiff. Si TRUE, plot
                             # exporter en .html (dynamique)
"AUC_format" = "date",      # L'axe du graphe est base sur une duree ("time")
                             # ou sur la date et heure d'acquisition ("date")
# -----#

```

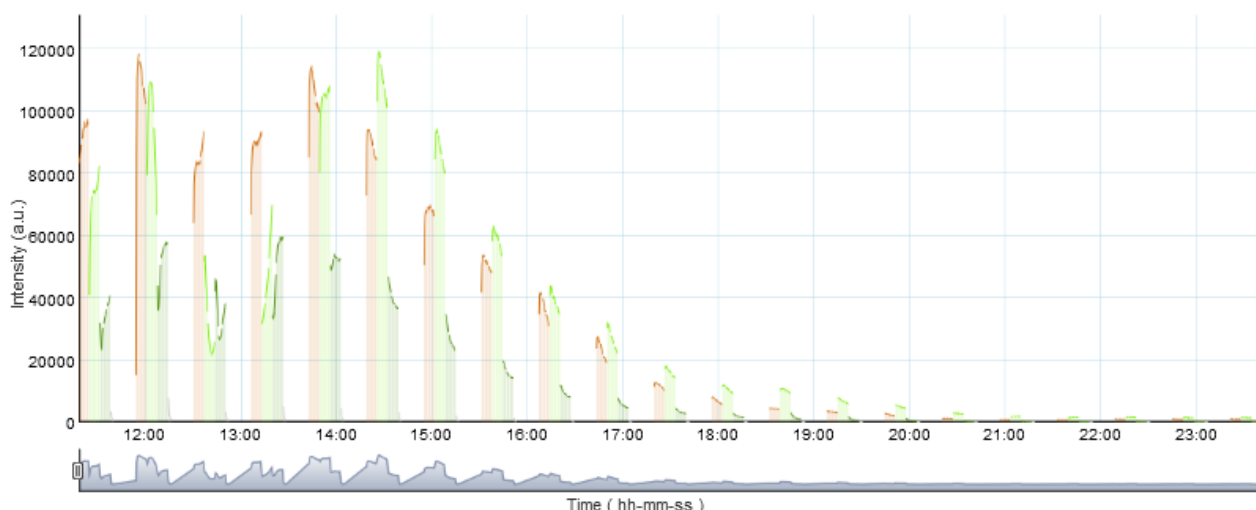


FIG. 2.4 : spaghetti plot

La figure 2.4 montre l'évolution quotidienne de l'émission des molécules de masse 137 (de 136.5 à 137.5) de trois plants de figuiers dont le cadre expérimentale est décrit section 1.3. Chaque plante est analysée par cycle de 384 secondes composé de 24 acquisitions de 16 secondes (regroupées en trois parties pour des raisons techniques). Vous pouvez télécharger cette figure [ici](#) afin de zoomer plus facilement.

Nous pouvons observer que le début de chaque cycle d'enregistrement est systématiquement impacté par la mesure précédente et le volume mort des lignes de mesures. Nous pouvons justifier ce biais expérimentale et donc le supprimer en ne prenant pas les 8 premières acquisitions de chaque cycle.

Ce bloc d'options est vraiment très simple d'utilisation. En jouant sur les masses et les modalités incluses dans le graphe, les utilisateur arrivent facilement à appréhender l'évolution des VOC de leur système. Pour autant, nous reviendrons en 2.4 sur l'utilisation de l'AUC et sa méthode de calcul qui impact grandement ce module.

## Calibration, regression linéaire et rapport automatisé

```

# -----#
# Calibration
"calibration" = FALSE,
"cal_rapport" = TRUE,      # Redige un rapport automatise pour evaluer la qualite de la calibration
"cal_plot_exp" = TRUE,     # Les axes x et y sont exponentielles ?
"M.conc" = c(1.008, 1.018, 1.013, 1.014, 1.013, 1.013, 1.015, 1.006, 1.007,
              0.994, 1.013), # La concentration, en ppm, de chaque masse,
                             # par defaut C = 1.000 ppm. (1 ou FALSE)
# -----#

```

Afin de vérifier la calibration de l'appareil, j'utilise un gaz étalon que je dilue avec un gaz pure lors des mesures de calibration. J'ai développé un script pour analyser cette calibration.

La sortie de ce script se fait sous forme de **rapports automatisés** basé sur des régressions linéaires simples. J'ai eu pour projet de développer cela en combinant les différentes masses et en proposant des tests de validation. La base de ce script peut également être reprise pour mesurer la concentration de différentes molécules. Ceci étant, le manque de temps et le côté monovariable ont largement freiné ce projet. D'autant que nous allons le voir, le travail ne manque pas.

## Perspective

Dans cette partie, je n'ai pas développé les analyses plus poussées (alignement et chimiométrie) qui sont expliqués dans le chapitre suivant. Nous avons simplement vu le principe de base de la PTR-TOF-MS ainsi que les modules de base de l'analyse : obtenir le graphe des spectres de masses, le suivi des AUC et le calcul de concentration. Le package proVOC permet d'effectuer ces opérations de façon simple et sans rien coder. De plus, il met à disposition les jeux de données pour que les utilisateurs puissent soit appliquer leur propre traitement soit utiliser des modules avancés.

Le package a été conçu pour assurer une reproductibilité des analyses. Lors de la conception de la structure du package, nous avons peu de recule sur l'utilisation du PTR-TOF-MS. Désormais, je suis convaincu qu'il me faut développer la version 2 très prochainement. Trois grosses modifications vont être effectuées.

La première est assez technique et concerne l'importation de donnée :

```
# Data importation and reduction ####
f_h5 <- dir("h5") %>%
  grep(".h5", .) # localise h5 files
ls_h5 <- vector("list", length(f_h5)) # the acquisitions listes
names(ls_h5) <- nm_ls(f_h5) # and the name
for (i in 1:length(ls_h5)) {
  ls_h5[[i]] <- paste0("h5/", dir("h5")[f_h5[i]]) %>%
    H5Fopen() # files import
}
mt_h5 <- meta_h5(ls_h5) # make the meta file
sp <- sp_red(ls_h5, mt_h5) # spectra reduction
sp$h5 <- list(f_h5 = f_h5, ls_h5 = ls_h5, mt_h5 = mt_h5, wd = getwd())
h5closeAll() # close the connexion with the h5 files.
# ls_h5 become NULL
```

Les fonctions *H5open()* et *H5close()* du package **rhdf5** permettent d'ouvrir les fichiers h5 afin de les importer puis de les fermer. La fonction *sp\_red()*, que j'ai écrite, est une meta fonction qui permet d'importer les spectres, les meta data, d'aligner les spectres, de supprimer les données inutiles et des détecter les peaks. Comme nous pouvons le voir, elle se trouve en sandwich entre l'ouverture de toutes les datas et la fermeture des données. Cette maladresse occupe de la RAM inutilement et ralentie beaucoup cette étape. Pour autant, toucher à la fonction qui importe les données et les mets en forme est une opération risquée qui met en péril la fonctionnalité des fonctions ultérieures.

La seconde modification sera sur le calcul de l'AUC. Initialement, je voulais trouver un moyen simple d'exprimer les 50 000 points des spectres en une valeur plus acceptable pour de l'analyse chimiométrique. J'avais donc piqué l'idée du bucketing des communautés de RMN en calculant l'AUC par masse unitaire (de M-0.5 à M+0.5). Cela permet d'obtenir une matrice de seulement 400 colonnes environ. En réalité, cela masque les cinétiques des molécules isobares. De plus, j'ai découvert après coup une fonction très performante pour détecter les pics (2.4). Je peux donc remplacer les AUC unitaires par l'intensité des pics. Il n'y aura alors plus qu'une confusion sur les isomères mais qui correspond aux limites techniques de l'instrument. J'aimerais également travailler sur l'AUC des pics et leur déconvolution afin de savoir si de l'information y est cachée.

La dernière modification se fera sur la structure :

```
PTR_MS_analysis(sp, mt) # debute les analyses
```

Une seule fonction permet de lancer la série d'analyses précisé dans le workflow *mt* sur les data de l'élément *sp*. La fonction *PTR\_MS\_analysis()* ne retourne rien, si ce n'est des graphiques, des fichiers *.csv* ou des rapports automatiques. J'aimerais inclure le workflow à l'intérieur de l'objet *sp* et le compléter au fur et à mesure. Ainsi l'utilisateur pourrait lancer des opérations à la volée, par exemple (en reprenant le block AUC de la section 2.4) :

```
# Aire sous la courbe (AUC) :
sp <- sp_AUC(sp, AUC_plot_exp = FALSE, AUC_each.group = FALSE,
  AUC_each.Mass = TRUE, M.num = c(81, 137, 153), AUC_plot_dy = TRUE,
  AUC_format = "date")
```

Cette forme franchement plus digeste permettra d'inclure plus simplement une analyse PTR-TOF-MS dans un script ou un rapport Rmarkdown.

J'espère avec ces modifications importantes rendre l'analyse plus optimisée en temps et répondre mieux aux attentes des utilisateurs. Ce package est sur GitHub depuis septembre 2020. J'ai fait peu de pub car je connais ses limites et je souhaiterais corriger ses principales défauts avant de le proposer aux collègues des autres centres de recherche avec qui nous avons des contacts. Les retours sur les quelques démonstrations effectuées sont très positives et proVOC permettra (je l'espère) de répondre positivement à un réel besoin. De plus proVOC est complètement cohérent avec les principes de l'open science développés au chapitre 1 et intègre des analyses chimiométriques qui sont détaillées au chapitre 3.

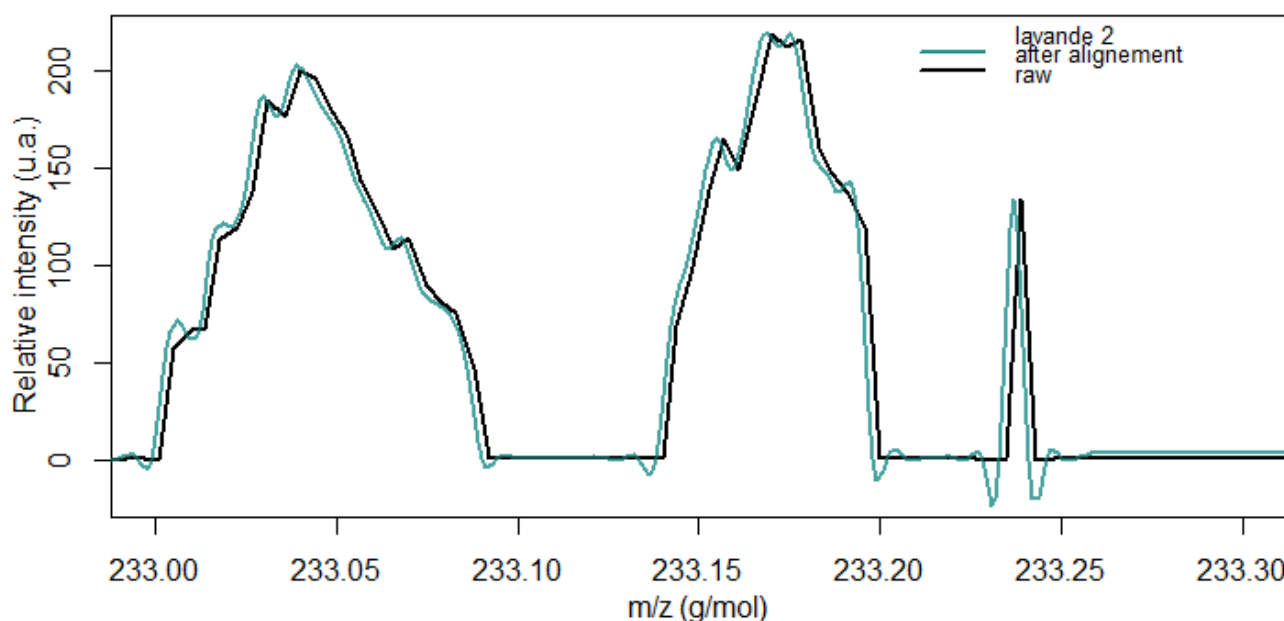


## Méthodes Multivariées

Nous allons à présent nous intéresser aux algorithmes utilisés pour traiter les jeux de données générés par le PTR-TOF-MS. Ce chapitre est divisé en deux parties. Lors de l'écriture du package, j'avais travaillé sur différentes méthodes d'alignement avant de m'arrêter sur l'une d'elle. Le début de ce stage m'a permis de reprendre ce travail avec des idées nouvelles que je présenterais dans une première partie. La deuxième partie présente l'utilisation de la chimiométrie pour l'analyse de donnée, d'abord avec la PCA puis avec deux méthodes moins populaires, l'ICA et la MCR qui ont l'avantage chacune d'essayer de décomposer les données en produits purs.

### 3.1 Alignement et sélection de variable

Actuellement, le package utilise la fonction *alignSpectra()* du package **MALDIquant**. L'alignement est fortement lié à la détection de pic. Soit le modèle utilisé n'est pas optimum et ajoute ou supprime des pics. Soit il crée un décalage trop important en alignant mal les pics se qui créer de la confusion lors de l'étape d'identification.



*alignSpectra()* est performant mais rencontre des limites sur les pics de faible ratio signal sur bruit, comme nous pouvons le voir sur la figure ???. Le bruit intensifie l'effet vaguelette et génère de l'ambiguïté sur le nombre de pics présent. Par ailleurs, la fonction crée des pics négatifs qui faussent le calcul de l'AUC.

J'ai trouvé un article très intéressant de PICAUD et al. [Pic+18] qui propose un algorithme pour supprimer la ligne de base, aligner les pics et les détecter. L'article explique les problèmes de faire ces trois étapes successivement et propose donc une alternative. Nous avons discuté avec l'auteur de sa solution. Malheureusement, le script est écrit en C++. Il est possible d'utiliser ce langage sous R, via Rcpp par exemple, mais cela aurait demandé un peu trop de temps. Pour l'instant, cette solution a été mise de côté mais pourra être exploré plus tard.

De plus, Jean-Michel Roger m'a proposé une approche basé sur l'algorithme COVSEL [Rog+11] pour résoudre notre problème. Cette méthode permet de recalculer les masses en recalculant l'abscisse pour chaque spectre. On observe cependant dans la théorie (équations 4 et 5) que l'on va devoir diagonaliser une matrice d'environ 150000x150000 points, essentiellement composée de zéro. Cette opération est impossible à effectuer avec un ordinateur de bureau. Pour autant, la solution que j'ai trouvée est d'utiliser l'algorithme COVSEL sur une toute petite partie du spectre et de le faire glisser sur toute la longueur. Cette technique fonctionne mais n'est pas du tout optimisée et prend un temps non acceptable (environ une minute par spectre).

Finalement, nous avons fait le choix de garder la fonction *alignSpectra()* de **MALDIquant**. De plus, ce package nous permet d'avoir une liste de tous les pics détectés. Ainsi, nous avons pour un même jeu de donnée, trois matrices à utiliser pour les analyses chimiométriques :

- la matrice alignée (composée d'environ 50 000 masses); ( $M_i$ )
- les AUC (~340 masses unitaires non nulles); ( $M_j$ )
- les pics (~300 pics, qui peuvent être des isobares). ( $M_j$ )

Pour une étude chimiométrique, nous pouvons discuter de la pertinence des jeux de données. Avec une matrice alignée  $M_i$ , on peut estimer que les algorithmes puissants des chimiométriciens vont détecter de petites variations négligées par les méthodes de détection de pics. Les  $M_i$  peuvent donc détecter les traces de VOC. Cependant les matrices  $M_j$  (AUC ou pics) possèdent un nombre de colonnes réduit de presque deux ordres de grandeur. De plus, le grand avantage du PTR-TOF-MS est d'obtenir des spectres ayant une excellente résolution de masse ( $10^{-3}$  Dalton). Nous pouvons donc utiliser une matrice  $M_j$  pour les analyses chimiométriques.

Maintenant que les outils pour détecter les pics et les outils d'analyses sont fonctionnelles pour les jeux de données du PTR-TOF-MS, il serait très intéressant d'analyser (via une MCR par exemple) comment la modification des paramètres de la fonction *detectPeaks()* (de MALDIquant) influe sur les résultats de chimiométrie.

## 3.2 Analyses Chimiométriques

### Analyse en Composante Principale (ACP)

L'ACP<sup>1</sup> est une méthode bien connue et enseignée dans toutes les bonnes formations d'analyses. Elle permet d'obtenir rapidement une vision de la variance des données. ProVOC permet d'effectuer une ACP sur le jeu de données de l'utilisateur. Un exemplaire peut être consulté ici. Il décrit le jeu de données utilisé dans la section 1.3 et y sera commenté dans

---

<sup>1</sup>Wikipédia est un haut lieu de la science ouverte

cette même section. L'ACP peut être appliqué à la matrice alignée avec un temps de calcul acceptable.

Cependant, le résultat de l'ACP est un pur objet mathématique qui est souvent très compliqué à interpréter correctement. Je me suis donc intéressé à deux algorithmes que je souhaitais utiliser depuis [longtemps](#)<sup>2</sup>.

## Multivariate Curve Resolution (MCR)

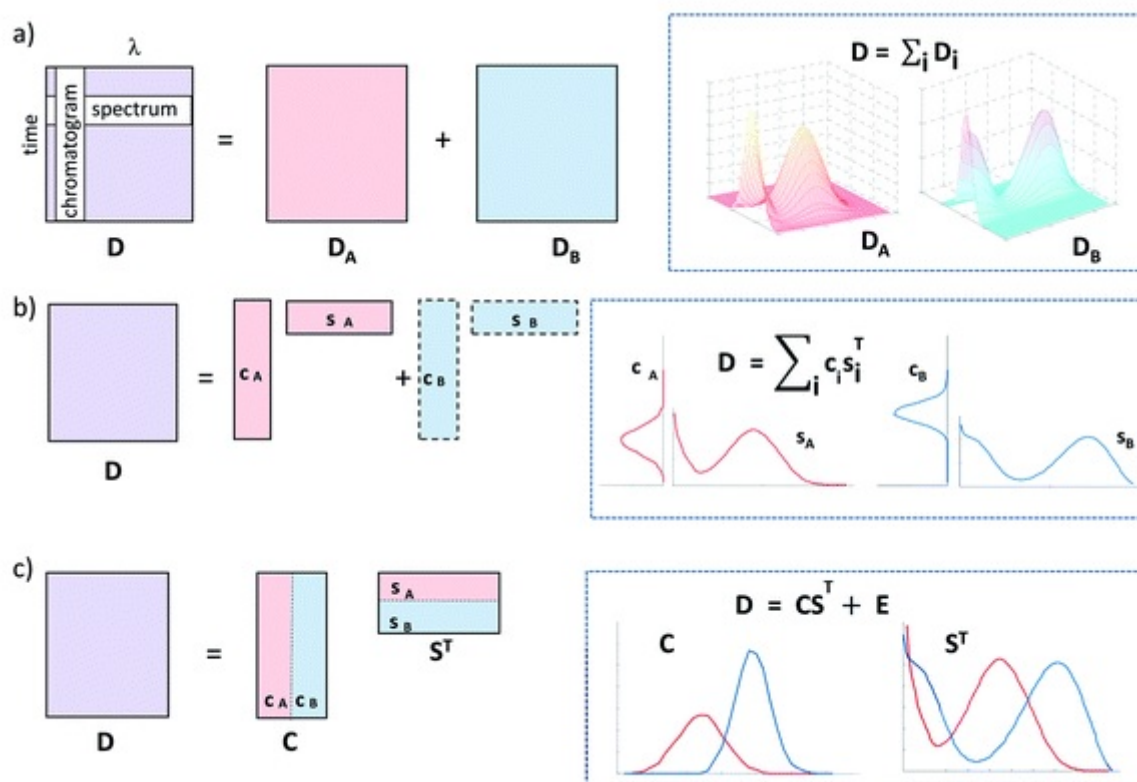


FIG. 3.1 : Bilinear model obtained from MCR for an HPLC-DAD data set. Expressed as (a) sum of pure signal contributions; (b) sum of the dyads of pure concentration profile and spectra; and (c) product of matrices of pure concentration profiles and spectra. (Image reproduit de JUAN et al. [JJT14] avec l'autorisation de l'éditeur.)

La MCR est une famille de méthodes qui cherche à décomposer l'ensemble des spectres en plusieurs sous-ensembles de produits purs. La figure 3.1 issue de l'article de JUAN et al. [JJT14] permet de bien comprendre l'objectif d'une MCR. La matrice  $D$ , obtenue par l'expérience, est l'expression de de deux produits purs  $S_A^T$  et  $S_B^T$  (deux bouquets de VOC pour notre PTR-TOF-MS) de concentration  $C_A$  et  $C_B$ , qui sont les quatre vecteurs que nous cherchons à connaître<sup>3</sup>. Les étapes de l'équation sont décrites dans l'article précédemment citée.

De plus, la MCR permet l'ajout de contraintes telles que la “non-négativité” qui permet de ne pas avoir de composantes négatives dans les  $S_i^T$  (fréquent en ACP) ou la contrainte “d'égalité” qui permet par exemple de renseigner un spectre pure de référence.

Un autre aspect très important de la MCR est de comprendre les trois “ambiguïtés” liées à l'algorithme :

<sup>2</sup>très probablement depuis le GFSV 2014

<sup>3</sup>et dont nous ne connaissons rien.

- l'ambigüité de **permutation**. Il n'y a pas d'ordre dans spectres purs et leur concentration associée. Le spectre  $S_i^T$  ne représente pas plus ou moins l'ensemble du jeu de donnée que le spectre  $S_j^T$ .
- l'ambigüité d'**intensité**. L'intensité du spectre pur  $S_i$  et sa concentration  $C_i$  peuvent être multipliées conjointement par les facteurs  $k_i$  et  $k_i^{-1}$ . La MCR ne peut donner qu'une concentration relative.
- l'ambigüité de **rotation**. Il est possible d'introduire une matrice inversible  $\mathbf{R}$  dans les équations :

$$D = CS^T + E$$

$$D = (CR)(R^{-1}S^T) + E$$

Puisque l'algorithme de la MCR cherche à minimiser l'erreur  $\mathbf{E}$ , la matrice  $\mathbf{R}$  ne fait pas varier  $\mathbf{E}$  mais change la composition des résultats.

Les contraintes permettent de lever en partie ces ambigüités.

La MCR permet donc d'initier les matrices  $\mathbf{C}$  et  $\mathbf{S}^T$ . Cette opération est utilisé dans un ensemble itératif MCR-ALS (alternating least squares) qui permet d'optimiser les résultats en fonction des contraintes utilisées. Par ailleurs, le nombre de composantes est un paramètre essentiel qui correspond véritablement à un nombre de phénomènes naturels. Il ne peut pas être déterminé arbitrairement à l'inverse du nombre de CP dans une ACP. Théoriquement, une MCR peut être effectué avec le jeu de donnée de la matrice alignée. Cependant, avec un ordinateur de bureau, les calculs sont assez long.

Un excellent complément d'information se trouve dans le livre RUCKEBUSCH [Ruc16] et en particulier les chapitres 2 [JT16], 3 [KWB16] et 14 [HDR16]. De plus, le chapitre 7 [RR16] présente l'algorithme de l'ICA. Ces deux algorithmes cherchent tous deux à extraire des spectres purs. Pour autant, les deux approches sont différentes et génèrent parfois des incompréhensions entre les communautés.

## Independent Components Analysis (ICA)

L'objectif principale de l'ICA est de retrouver les *signal sources* mélangés à l'intérieur du jeu de données et la proportion de chacun d'eux. Pour cela, l'ICA s'intéresse à la distribution de l'intensité du signal sur un histogramme (cf. Fig2 de D. N. RUTLEDGE et al. [RR16]). Selon la loi central limite, plus l'histogramme tend vers une répartition gaussienne, plus il y a de chances que ce signal soit du bruit. Inversement, moins la répartition est gaussienne, plus le signal correspond à un signal source. De plus, comme pour une ACP, l'ICA cherche à obtenir des signaux sources indépendant, et donc orthogonaux. Contrairement à la MCR, L'ICA n'impose pas de contrainte. La phrase de [RR16] permet de bien comprendre la différence entre les deux méthodes : *This is because while MCR aims to extract the signals of pure compounds, ICA extracts signals reflecting underlying **independent phenomena**, which may in fact be combinations of the signals of several pure compounds.*<sup>4</sup>

Il existe plusieurs algorithmes qui utilise l'ICA, [WDH08] et [Al-15]. Cependant, l'algorithme JADE [CS93] est probablement le plus utilisé pour l'ICA, [RJ13] et [RJ15]. Un package JADE pour R est déposé et maintenu sur le CRAN. Le principe de JADE est de décomposer la matrice

<sup>4</sup>En effet, alors que la MCR vise à extraire les composés purs, l'ICA extrait les signaux reflétant des **phénomènes indépendants** sous-jacents, qui peuvent en fait être des combinaisons des signaux de plusieurs composés purs.

$\mathbf{X}$  (le jeu de donnée expérimentale) en utilisant les loadings d'une ACP. Ces loadings sont centrés et normés puis arrangés pour former un tenseur d'ordre 4. De façon itérative, JADE va ensuite optimiser ces loadings pour les rendre indépendants.

Cette description succincte permet très vite de comprendre que réaliser un tenseur d'ordre 4 à base de vecteur de 55000 points risque d'être problématique. L'ICA ne peut se faire que sur une matrice réduite comme l'AUC ou les pics détectés.

## 3.3 Applications

...

### bourgeon

blabla voir le détail section [1.3](#)

### lavande

blabla voir le détail section [1.3](#)



# CHAPITRE 4

## Conclusion

c'était chouette



# Glossaire

CEFE (UMR 5175) : Centre d'Écologie Fonctionnelle et Évolutive. Laboratoire d'accueil situé sur le site de la DR13 du CNRS, route de Mende à Montpellier.

COV (ou VOC) : Composé Organique Volatile. Molécules produites par des systèmes biologique et analyse par le PTR-ToF-MS.

GC-MS : Gaz Chromatographie Mass Spectroscopy. Autre méthode d'analyse des VOC, moins sensible mais plus discriminante que la PTR-TOF-MS.

PACE : Plateforme d'Analyse Chimique en Écologie. Plateforme situé au sein de l'UMR 5175.

PTR-ToF-MS : Proton Transfert Reaction Time of Flight Mass Spectroscopy. Acronyme utilisé pour l'instrument de mesure (le PTR-ToF-MS) et pour la méthode d'analyse (la PTR-ToF-MS). Par abus de langage, l'acronyme PTR-MS est régulièrement utilisé.



# Bibliographie

- [AB20] J. ARTHO et R. BATRO. *Rotting Christ*. Batro'Games. T. 1. Nov. 2020. URL : <http://www.legrog.org/jeux/rotting-christ/rotting-christ-fr> (cf. p. 9).
- [BA01] L. BABCOCK et N. ADAMS. *Advances in Gas Phase Ion Chemistry, Volume 4 - 1st Edition*. 1<sup>re</sup> éd. T. 4. 2001. URL : <https://www-elsevier-com.insee.bib.cnrs.fr/books/advances-in-gas-phase-ion-chemistry/babcock/978-0-444-50929-1> (visité le 12/08/2021) (cf. p. 12).
- [Bla+04] R. S. BLAKE, C. WHYTE, C. O. HUGHES, A. M. ELLIS et P. S. MONKS. « Demonstration of proton-transfer reaction time-of-flight mass spectrometry for real-time analysis of trace volatile organic compounds ». eng. In : *Anal Chem* 76.13 (juill. 2004), p. 3841-3845. DOI : [10.1021/ac0498260](https://doi.org/10.1021/ac0498260) (cf. p. 12).
- [Boe+15] C. BOETTIGER, S. CHAMBERLAIN, E. HART et K. RAM. « Building Software, Building Community : Lessons from the rOpenSci Project ». en. In : *Journal of Open Research Software* 3.1 (nov. 2015). Number : 1 Publisher : Ubiquity Press, e8. DOI : [10.5334/jors.bu](https://doi.org/10.5334/jors.bu). URL : <http://openresearchsoftware.metajnl.com/articles/10.5334/jors.bu/> (visité le 23/08/2021) (cf. p. 5).
- [Cap+10] L. CAPPELLIN, F. BIASIOLI, A. FABRIS, E. SCHUHFRIED, C. SOUKOULIS, T. D. MÄRK et F. GASPERI. « Improved mass accuracy in PTR-TOF-MS : Another step towards better compound identification in PTR-MS ». en. In : *International Journal of Mass Spectrometry* 290.1 (fév. 2010), p. 60-63. DOI : [10.1016/j.ijms.2009.11.007](https://doi.org/10.1016/j.ijms.2009.11.007). URL : <https://linkinghub.elsevier.com/retrieve/pii/S1387380609003571> (visité le 03/05/2021) (cf. p. 12).
- [Cap+11] L. CAPPELLIN, F. BIASIOLI, P. M. GRANITTO, E. SCHUHFRIED, C. SOUKOULIS, F. COSTA, T. D. MÄRK et F. GASPERI. « On data analysis in PTR-TOF-MS : From raw spectra to data mining ». en. In : *Sensors and Actuators B : Chemical* 155.1 (juill. 2011), p. 183-190. DOI : [10.1016/j.snb.2010.11.044](https://doi.org/10.1016/j.snb.2010.11.044). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0925400510009135> (visité le 03/05/2021) (cf. p. 12).
- [CS93] J. CARDOSO et A. SOULOUMIAC. « Blind beamforming for non-gaussian signals ». en. In : *IEEE Proc. F Radar Signal Process. UK* 140.6 (1993), p. 362. DOI : [10.1049/ip-f-2.1993.0054](https://doi.org/10.1049/ip-f-2.1993.0054). URL : <https://digital-library.theiet.org/content/journals/10.1049/ip-f-2.1993.0054> (visité le 24/08/2021) (cf. p. 24).
- [Cla+21] M. S. CLAFLIN, D. PAGONIS, Z. FINEWAX, A. V. HANDSCHY, D. A. DAY, W. L. BROWN, J. T. JAYNE, D. R. WORSNOP, J. L. JIMENEZ, P. J. ZIEMANN, J. de GOUW et B. M. LERNER. « An in situ gas chromatograph with automatic detector switching between PTR- and EI-TOF-MS : isomer-resolved measurements of indoor air ». In : *Atmospheric Measurement Techniques* 14.1 (2021), p. 133-152. DOI : [10.5194/amt-14-133-2021](https://doi.org/10.5194/amt-14-133-2021). URL : <https://amt.copernicus.org/articles/14/133/2021/> (cf. p. 12).

- [Deu+19] Z. DEUSCHER, I. ANDRIOT, E. SÉMON, M. REPOUX, S. PREYS, J.-M. ROGER, R. BOULANGER, H. LABOURÉ et J.-L. LE QUÉRÉ. « Volatile compounds profiling by using proton transfer reaction-time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates organoleptic differences ». en. In : *J Mass Spectrom* 54.1 (jan. 2019), p. 92-119. DOI : [10.1002/jms.4317](https://doi.org/10.1002/jms.4317). URL : <http://doi.wiley.com/10.1002/jms.4317> (visité le 17/05/2021) (cf. p. 12).
- [Enn+05] C. J. ENNIS, J. C. REYNOLDS, B. J. KEELY et L. J. CARPENTER. « A hollow cathode proton transfer reaction time of flight mass spectrometer ». en. In : *International Journal of Mass Spectrometry* 247.1 (déc. 2005), p. 72-80. DOI : [10.1016/j.ijms.2005.09.008](https://doi.org/10.1016/j.ijms.2005.09.008). URL : <https://www.sciencedirect.com/science/article/pii/S1387380605002472> (visité le 12/08/2021) (cf. p. 12).
- [Fis+21] B. FISCHER, M. SMITH, G. PAU, M. MORGAN et D. v. TWISK. *rhdf5 : R Interface to HDF5*. 2021. DOI : [10.18129/B9.bioc.rhdf5](https://doi.org/10.18129/B9.bioc.rhdf5). URL : <https://bioconductor.org/packages/rhdf5/> (visité le 18/08/2021) (cf. p. 15).
- [GMH10] M. GRAUS, M. MÜLLER et A. HANSEL. « High resolution PTR-TOF : Quantification and formula confirmation of VOC in real time ». en. In : *J Am Soc Mass Spectrom* 21.6 (juin 2010), p. 1037-1044. DOI : [10.1016/j.jasms.2010.02.006](https://doi.org/10.1016/j.jasms.2010.02.006). URL : <https://pubs.acs.org/doi/10.1021/jasms.8b03774> (visité le 03/05/2021) (cf. p. 12, 13).
- [Ham13] R. HAMEL. « L'anglais, langue unique pour les sciences? Le rôle des modèles plurilingues dans la recherche, la communication scientifique et l'enseignement supérieur ». In : *Synergies Europe* 8 (jan. 2013), p. 53-66 (cf. p. 4).
- [Han+95] A. HANSEL, A. JORDAN, R. HOLZINGER, P. PRAZELLER, W. VOGEL et W. LINDINGER. « Proton transfer reaction mass spectrometry : on-line trace gas analysis at the ppb level ». en. In : *International Journal of Mass Spectrometry and Ion Processes*. Honour Biography David Smith 149-150 (nov. 1995), p. 609-619. DOI : [10.1016/0168-1176\(95\)04294-U](https://doi.org/10.1016/0168-1176(95)04294-U). URL : <https://www.sciencedirect.com/science/article/pii/016811769504294U> (visité le 12/08/2021) (cf. p. 11, 13).
- [Hol15] R. HOLZINGER. « PTRwid : A new widget tool for processing PTR-TOF-MS data ». en. In : *Atmos. Meas. Tech.* 8.9 (sept. 2015), p. 3903-3922. DOI : [10.5194/amt-8-3903-2015](https://doi.org/10.5194/amt-8-3903-2015). URL : <https://amt.copernicus.org/articles/8/3903/2015/> (visité le 17/05/2021) (cf. p. 12).
- [HDR16] S. HUGELIER, O. DEVOS et C. RUCKEBUSCH. « Chapter 14 – A smoothness constraint in multivariate curve resolution-alternating least squares of spectroscopy data ». In : *Data Handling in Science and Technology*. T. 30. Resolving spectral mixtures - With applications from ultrafast time-resolved spectroscopy to super-resolution imaging. 2016, p. 453-476. DOI : [10.1016/B978-0-444-63638-6.00014-0](https://doi.org/10.1016/B978-0-444-63638-6.00014-0). URL : <https://hal.archives-ouvertes.fr/hal-01386913> (visité le 26/05/2021) (cf. p. 24).
- [Ino+06] S. INOMATA, H. TANIMOTO, N. AOKI, J. HIROKAWA et Y. SADANAGA. « A novel discharge source of hydronium ions for proton transfer reaction ionization : design, characterization, and performance ». en. In : *Rapid Communications in Mass Spectrometry* 20.6 (2006). \_eprint : <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/rcm.2405>. URL : <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/rcm.2405> (visité le 12/08/2021) (cf. p. 12).

- [Jor+09] A. JORDAN, S. HAIDACHER, G. HANEL, E. HARTUNGEN, L. MÄRK, H. SEEHAUSER, R. SCHOTTKOWSKY, P. SULZER et T. MÄRK. « A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS) ». en. In : *International Journal of Mass Spectrometry* 286.2-3 (sept. 2009), p. 122-128. DOI : [10.1016/j.ijms.2009.07.005](https://doi.org/10.1016/j.ijms.2009.07.005). URL : <https://linkinghub.elsevier.com/retrieve/pii/S1387380609002371> (visité le 30/04/2021) (cf. p. 12).
- [JT16] A. de JUAN et R. TAULER. « Chapter 2 - Multivariate Curve Resolution-Alternating Least Squares for Spectroscopic Data ». en. In : *Data Handling in Science and Technology*. Sous la dir. de C. RUCKEBUSCH. T. 30. Resolving Spectral Mixtures. Elsevier, jan. 2016, p. 5-51. DOI : [10.1016/B978-0-444-63638-6.00002-4](https://doi.org/10.1016/B978-0-444-63638-6.00002-4). URL : <https://www.sciencedirect.com/science/article/pii/B9780444636386000024> (visité le 26/05/2021) (cf. p. 24).
- [JJT14] A. de JUAN, J. JAUMOT et R. TAULER. « Multivariate Curve Resolution (MCR). Solving the mixture analysis problem ». en. In : *Anal. Methods* 6.14 (2014), p. 4964-4976. DOI : [10.1039/C4AY00571F](https://doi.org/10.1039/C4AY00571F). URL : <http://xlink.rsc.org/?DOI=C4AY00571F> (visité le 19/05/2021) (cf. p. 23).
- [Kre+18] J. KRECHMER, F. LOPEZ-HILFIKER, A. KOSS, M. HUTTERLI, C. STOERMER, B. DEMING, J. KIMMEL, C. WARNEKE, R. HOLZINGER, J. JAYNE, D. WORSNOP, K. FUHRER, M. GONIN et J. de GOUW. « Evaluation of a New Reagent-Ion Source and Focusing Ion-Molecule Reactor for Use in Proton-Transfer-Reaction Mass Spectrometry ». In : *Anal. Chem.* 90.20 (oct. 2018). Publisher : American Chemical Society, p. 12011-12018. DOI : [10.1021/acs.analchem.8b02641](https://doi.org/10.1021/acs.analchem.8b02641). URL : <https://doi.org/10.1021/acs.analchem.8b02641> (visité le 13/08/2021) (cf. p. 13).
- [KWB16] S. KUCHERYAVSKIY, W. WINDIG et A. BOGOMOLOV. « Chapter 3 - Spectral Unmixing Using the Concept of Pure Variables ». en. In : *Data Handling in Science and Technology*. Sous la dir. de C. RUCKEBUSCH. T. 30. Resolving Spectral Mixtures. Elsevier, jan. 2016, p. 53-99. DOI : [10.1016/B978-0-444-63638-6.00003-6](https://doi.org/10.1016/B978-0-444-63638-6.00003-6). URL : <https://www.sciencedirect.com/science/article/pii/B9780444636386000036> (visité le 27/05/2021) (cf. p. 24).
- [Lag+94] A. LAGG, J. TAUCHER, A. HANSEL et W. LINDINGER. « Applications of proton transfer reactions to gas analysis ». en. In : *International Journal of Mass Spectrometry and Ion Processes* 134.1 (juin 1994), p. 55-66. DOI : [10.1016/0168-1176\(94\)03965-8](https://doi.org/10.1016/0168-1176(94)03965-8). URL : <https://www.sciencedirect.com/science/article/pii/0168117694039658> (visité le 12/08/2021) (cf. p. 11).
- [LLA91] W. LINDINGER, K. LEITER et M. ANDRIOLLO. « Industrial Multicomponent Gas Analysis ». In : *Ionen-Technik GmbH* 20.7 (1991), p. 24-25 (cf. p. 11).
- [LHJ98] W. LINDINGER, A. HANSEL et A. JORDAN. « On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research ». en. In : *International Journal of Mass Spectrometry and Ion Processes* 173.3 (fév. 1998), p. 191-241. DOI : [10.1016/S0168-1176\(97\)00281-4](https://doi.org/10.1016/S0168-1176(97)00281-4). URL : <https://www.sciencedirect.com/science/article/pii/S0168117697002814> (visité le 12/08/2021) (cf. p. 11).
- [LHP93] W. LINDINGER, J. HIRBER et H. PARETZKE. « An ion/molecule-reaction mass spectrometer used for on-line trace gas analysis ». en. In : *International Journal of Mass Spectrometry and Ion Processes* 129 (nov. 1993), p. 79-88. DOI : [10.1016/0168-1176\(93\)87031-M](https://doi.org/10.1016/0168-1176(93)87031-M). URL : <https://www.sciencedirect.com/science/article/pii/016811769387031M> (visité le 12/08/2021) (cf. p. 11).

- [LJ98] W. LINDINGER et A. JORDAN. « Proton-transfer-reaction mass spectrometry (PTR-MS) : on-line monitoring of volatile organic compounds at pptv levels ». en. In : *Chem. Soc. Rev.* 27.5 (jan. 1998). Publisher : The Royal Society of Chemistry, p. 347-375. DOI : [10.1039/A827347Z](https://doi.org/10.1039/A827347Z). URL : <https://pubs.rsc.org/en/content/articlelanding/1998/cs/a827347z> (visité le 30/04/2021) (cf. p. 11).
- [Maj+18] T. MAJCHRZAK, W. WOJNOWSKI, M. LUBINSKA-SZCZYGEŁ, A. RÓŻAŃSKA, J. NAMIEŚNIK et T. DYMERSKI. « PTR-MS and GC-MS as complementary techniques for analysis of volatiles : A tutorial review ». en. In : *Analytica Chimica Acta* 1035 (déc. 2018), p. 1-13. DOI : [10.1016/j.aca.2018.06.056](https://doi.org/10.1016/j.aca.2018.06.056). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0003267018308213> (visité le 17/05/2021) (cf. p. 15).
- [Pic+18] V. PICAUD, J.-F. GIOVANNELLI, C. TRUNTZER, J.-P. CHARRIER, A. GIREMUS, P. GRANGEAT et C. MERCIER. « Linear MALDI-ToF simultaneous spectrum deconvolution and baseline removal ». In : *BMC Bioinformatics* 19.1 (avr. 2018), p. 123. DOI : [10.1186/s12859-018-2116-3](https://doi.org/10.1186/s12859-018-2116-3). URL : <https://doi.org/10.1186/s12859-018-2116-3> (visité le 29/04/2021) (cf. p. 22).
- [Rog+11] J. ROGER, B. PALAGOS, D. BERTRAND et E. FERNANDEZ-AHUMADA. « CovSel : Variable selection for highly multivariate and multi-response calibration ». en. In : *Chemometrics and Intelligent Laboratory Systems* 106.2 (avr. 2011), p. 216-223. DOI : [10.1016/j.chemolab.2010.10.003](https://doi.org/10.1016/j.chemolab.2010.10.003). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0169743910001978> (visité le 03/05/2021) (cf. p. 22).
- [Ruc16] C. RUCKEBUSCH. *Resolving Spectral Mixtures, Volume 30 - 1st Edition*. 1<sup>re</sup> éd. T. 30. Août 2016. URL : <https://www.sciencedirect.com/bookseries/data-handling-in-science-and-technology/vol/30/suppl/C> (visité le 24/08/2021) (cf. p. 24).
- [RR16] D. N. RUTLEDGE et D. RIMBAUD BOUVERESSE. « Chapter 7 - Independent Components Analysis : Theory and Applications ». en. In : *Data Handling in Science and Technology*. Sous la dir. de C. RUCKEBUSCH. T. 30. Resolving Spectral Mixtures. Elsevier, jan. 2016, p. 225-277. DOI : [10.1016/B978-0-444-63638-6.00007-3](https://doi.org/10.1016/B978-0-444-63638-6.00007-3). URL : <https://www.sciencedirect.com/science/article/pii/B9780444636386000073> (visité le 26/05/2021) (cf. p. 24).
- [RJ13] D. RUTLEDGE et D. JOUAN-RIMBAUD BOUVERESSE. « Independent Components Analysis with the JADE algorithm ». en. In : *TrAC Trends in Analytical Chemistry* 50 (oct. 2013), p. 22-32. DOI : [10.1016/j.trac.2013.03.013](https://doi.org/10.1016/j.trac.2013.03.013). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0165993613001222> (visité le 24/08/2021) (cf. p. 24).
- [RJ15] D. RUTLEDGE et D. JOUAN-RIMBAUD BOUVERESSE. « Corrigendum to 'Independent Components Analysis with the JADE algorithm' ». en. In : *TrAC Trends in Analytical Chemistry* 67 (avr. 2015), p. 220. DOI : [10.1016/j.trac.2015.02.001](https://doi.org/10.1016/j.trac.2015.02.001). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0165993615000345> (visité le 24/08/2021) (cf. p. 24).
- [Al-15] A. AL-SAEH. « Independent Component Analysis for Separation of Speech Mixtures : A Comparison Among Thirty Algorithms ». en. In : *IJEEE* 11.1 (juin 2015), p. 1-9. DOI : [10.37917/ijeee.11.1.1](https://doi.org/10.37917/ijeee.11.1.1). URL : <http://ijeee.edu.iq/Papers/Vol11-Issue1/102709.pdf> (visité le 24/08/2021) (cf. p. 24).
- [Tan+07] H. TANIMOTO, N. AOKI, S. INOMATA, J. HIROKAWA et Y. SADANAGA. « Development of a PTR-TOFMS instrument for real-time measurements of volatile organic compounds in air ». en. In : *International Journal of Mass Spectrometry* 263.1 (mai 2007), p. 1-11. DOI : [10.1016/j.ijms.2007.01.009](https://doi.org/10.1016/j.ijms.2007.01.009). URL : <https://www.sciencedirect.com/science/article/pii/S1387380607000231> (visité le 12/08/2021) (cf. p. 12).

- [WDH08] G. WANG, Q. DING et Z. HOU. « Independent component analysis and its applications in signal processing for analytical chemistry ». en. In : *TrAC Trends in Analytical Chemistry* 27.4 (avr. 2008), p. 368-376. DOI : [10.1016/j.trac.2008.01.009](https://doi.org/10.1016/j.trac.2008.01.009). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0165993608000101> (visit  le 24/08/2021) (cf. p. 24).



# Table des figures

1.1	PTR-ToF-MS relié à l'un des quatre site de mesure d'un mésocosme . . . . .	6
1.2	Sonde à proximité de la mangeoire . . . . .	7
2.1	Schéma de la chambre d'ionisation (Reprinted with permission from KRECHMER et al. [Kre+18]. Copyright 2021 American Chemical Society.) . . . . .	13
2.2	Schéma de la colonne TOF et du détecteur MS (Reproduit de l'article GRAUS et al. [GMH10], consultable ici : <a href="https://www.sciencedirect.com/science/article/pii/S1044030510001005#fig1">https://www.sciencedirect.com/science/article/pii/S1044030510001005#fig1</a> . . . . .	13
2.3	spectre MS . . . . .	14
2.4	spaghettis plot . . . . .	17
3.1	Bilinear model obtained from MCR for an HPLC-DAD data set. Expressed as (a) sum of pure signal contributions ; (b) sum of the dyads of pure concentration profile and spectra ; and (c) product of matrices of pure concentration profiles and spectra. (Image reproduit de JUAN et al. [JJT14] avec l'autorisation de l'éditeur.) . . . . .	23

**Abstract** English abstract, on the last page. This is a bookdown template based on LaTeX memoir class.

**Keywords** Chimie-métrie, PTR-ToF-MS, Spectroscopie, Environnement.

**Résumé** Résumé en Français

**Mots-Clés** Chimie-métrie, PTR-ToF-MS, Spectroscopie, Environnement.



École Pratique  
des Hautes Études

