Mise en oeuvre de l'analyse non ciblée comme outil de caractérisation chimique des eaux : impact du rejet d'une station d'épuration sur le profil moléculaire d'une rivière

Solène Motteau

▶ **To cite this version:**

Solène Motteau. Mise en oeuvre de l'analyse non ciblée comme outil de caractérisation chimique des eaux : impact du rejet d'une station d'épuration sur le profil moléculaire d'une rivière. Chimie analytique. Université de Poitiers, 2021. Français. NNT : 2021POIT2265 . tel-03516293

HAL Id: tel-03516293
https://theses.hal.science/tel-03516293

Submitted on 7 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

Pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

(ECOLE SUPERIEURE d'INGENIEURS de POITIERS)
(Diplôme National - Arrêté du 25 mai 2016)

Ecole Doctorale : Chimie Ecologie Géosciences Agrosciences
« *Théodore Monod* »

Secteur de Recherche : Chimie théorique, physique, analytique

Présentée par :

Solène MOTTEAU
Maître ès sciences

****

MISE EN ŒUVRE DE L'ANALYSE NON CIBLEE COMME OUTIL DE CARACTERISATION CHIMIQUE DES EAUX : IMPACT DU REJET D'UNE STATION D'EPURATION SUR LE PROFIL MOLECULAIRE D'UNE RIVIERE

****

Directeurs de Thèse : M^me^ Nathalie KARPEL VEL LEITNER
M^r^ Bertrand GOMBERT
M^me^ Marie DEBORDE

****

Soutenue le Mercredi 7 avril 2021

devant la Commission d'Examen

****

JURY

Rapporteurs :	M^r^ Gabriel BILLON
	M^me^ Emmanuelle VULLIET
Examinateurs :	M^me^ Claire ALBASI
Directeurs de thèse :	M^me^ Nathalie KARPEL VEL LEITNER
	M^r^ Bertrand GOMBERT
	M^me^ Marie DEBORDE
Membre invité :	M^me^ Florence GABORIAU

« Vous vous débrouillez comme un chef.

Mais je vous conseille d'économiser vos forces :

il reste encore BEAUCOUP de tests. »

(GLaDOS, Portal 2)

Remerciements

Tout d'abord, je souhaiterais remercier l'Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP, UMR CRNS 7285) ainsi que l'équipe Eaux, Biomarqueurs, Contaminants Organiques, Milieux (E.BiCOM) pour avoir accepté de m'accueillir dans le laboratoire.

Je remercie la région Nouvelle-Aquitaine et Grand Poitiers pour le financement de ma thèse sans lequel ce projet n'aurait pu voir le jour.

Je tiens à remercier les rapporteurs Madame Emmanuelle Vulliet, Monsieur Gabriel Billon ainsi que l'examinatrice Madame Claire Albasi d'avoir accepté d'évaluer mes travaux et d'apporter leur expertise.

Je remercie particulièrement mes directeurs de thèse Madame Nathalie Karpel Vel Leitner, Madame Marie Deborde et Monsieur Bertrand Gombert pour leur patience, leur disponibilité et leur soutien durant ces trois ans. Leurs regards critiques et leurs conseils m'ont permis d'apprendre et d'avancer dans la voie de la recherche.

Merci à l'ensemble des membres de l'équipe pour leur aide, leur soutien et leurs encouragements. Je tiens à remercier plus particulièrement Cristina pour sa gentillesse, sa présence et ses conseils. J'ai pris beaucoup de plaisir à apporter mon aide à l'organisation des Journées Informations Eaux.

Je remercie aussi Quentin pour son aide technique, sa disponibilité et les nombreuses discussions scientifiques que nous avons eu. Merci également à Alexis pour son aide, nos discussions techniques et son apport éclairant sur son domaine de compétence.

J'exprime ma gratitude à mes compagnons de route, mes ordinateurs et cette chère UHPLC-HRMS. Malgré les nombreuses péripéties, plantages, grosses pannes et autres erreurs incongrues, vous m'avez vaillamment accompagnée jusqu'au bout. Au-delà de l'immense frustration que j'ai pu ressentir parfois face aux bugs et aux problèmes, ensemble nous sommes parvenus à produire moult données, qui tiennent sur pas moins de trois disques durs tout de même.

Je remercie chaleureusement les membres de la Team Tea : Julie, Jeannette et Florence. Que de moments partagés au cours de ces années qui ont égayés de nos éclats de rire notre grand bureau. Nous avons eu de riches discussions allant de sujets de recherche sérieux à des interludes musicales. Je suis la dernière de nous quatre mais vous avez été avec moi jusqu'au bout.

Je remercie vivement Pierre, Clément, Tom, Noelline, Alexis, Marie, Antoine, Valérian et Valentin pour leur bonne humeur et leur amitié à chaque instant.

Toute mon affection à mes parents et mon frère qui ont toujours été là pour moi dans mes moments de bonheur et mes moments de détresse. Merci papa pour avoir donné le goût de la science, pour répondre à mes questions même les plus absurdes et me pousser à croire en moi. Merci maman pour tes encouragements, toutes tes petites attentions et tes petits mots.

À celui qui partage ma vie depuis plus de douze ans maintenant, je ne te remercierai jamais assez de ton soutien inconditionnel. Jonas, merci d'avoir partagé tous mes hauts et mes bas sans ne jamais t'arrêter un seul instant d'essayer de me rendre le sourire.

Résumé

La diffusion des micropolluants et plus particulièrement de la diffusion par les rejets de stations d'épuration (STEP) est une problématique récurrente depuis de plusieurs années dans la nécessité de protéger les milieux aquatiques du point de vu de la qualité chimique et écotoxique. Le développement de la spectrométrie de masse haute résolution et de l'accès aux bases de données a permis l'essor de l'analyse non ciblée appliquée aux matrices environnementales dont les eaux de STEP et les eaux de surfaces.

Les objectifs de ce travail de thèse ont été de développer une méthode d'analyse non ciblée, de l'appliquer sur des eaux de stations d'épuration afin de les caractériser et d'étudier l'impact du traitement des eaux usées et du rejet des eaux traitées dans la rivière Clain. Des signaux pertinents et traceurs du rejet de STEP ont également été recherchés par l'utilisation d'analyses différentielles.

Deux méthodes se basant sur les logiciels Mzmine 2 (avec R 3.5.1 et Excel en complément) et Compound Discoverer ont été développées. À la suite de ces développements, le logiciel Compound Discoverer a été conservé pour le reste des travaux de thèse. L'usage de trois réplicas a montré une réduction du nombre de faux positifs. L'effet matrice a également été évaluée de même que l'utilisation de l'extraction sur phase solide qui s'est avérée globalement positive. L'application de l'analyse non ciblée sur des entrées et sortie de STEP a montré une diminution du nombre de signaux détectés en sortie, principalement des signaux plutôt apolaires et/ou de taille importante mais a montré aussi l'apparition de nouveaux signaux. À partir de l'analyse différentielle, 375 signaux pertinents stables ou augmentant en sortie ont été sélectionnés et l'annotation d'une partie d'entre eux a été effectuée. L'étude de l'impact du rejet des eaux traitées a mis en lumière l'apport de nouveaux signaux dans le Clain. Malgré une baisse des valeurs des aires en aval immédiat, il n'a pas été observé de retour à l'état initial en amont. Une sélection de 305 signaux pertinents a été effectuée dont un faible nombre sont communs avec la liste réalisée à partir des échantillons de STEP, témoignant à la fois de l'intérêt de la méthode mais également de la difficulté de déterminer des traceurs apparentés réellement aux rejets de STEP. Des analyses de suspects ont également été réalisées en complément de l'analyse non ciblée.

The dispersion of micropollutants, especially from wastewater treatment plants (WWTP), has been a recurring topic for several years regarding the need of aquatic environments protection i.e. towards chemical quality and ecotoxicity considerations. The development of high-resolution mass spectrometry (UHPLC-HRMS) and the access to databases allowed the development of non-targeted analysis for environmental matrices such as WWTP water and surface water.

The objectives of this PhD work were to develop a non-targeted analysis method and to apply it to water samples from wastewater treatment plants in order to characterize them. The possibly of tracing these waters close to river discharges has been studied. Relevant features and tracers of WWTP discharge were thus sought by the use of differential analyses.

Two methods based on Mzmine 2 software (combined with the use of R 3.5.1 and Excel) and Compound Discoverer were developed. As a result of these developments, the use of Compound Discoverer software was favored for the rest of this work. The use of three injection replicas during the analyses by UHPLC-HRMS resulted in a reduction in the number of false positives. The matrix effect was also evaluated as well as the use of solid phase extraction which was generally positive. The application of non-targeted analysis on samples from WWTP influents and treated waters showed a decrease in the number of features detected in the treated water, especially in the case of features corresponding to apolar molecules and/or with a high molecular mass. However new features were also observed in these waters. From the differential analysis, 375 stable or increasing relevant features in treated waters were selected and a chemical structure was proposed for some of them. The study of the impact of the discharge of treated water highlighted the addition of new features in the river. Despite a decrease in the area values of the features immediately downstream, a return to the initial state of the upstream was not observed. 305 relevant features were selected on the basis of a downstream/upstream comparison whose a small number which were common with the relevant features list made from the analysis of WWTP samples. These results underline both the interest of the method and the difficulty of determining tracers that are really related to WWTP discharges. Analyses of suspects were also carried out in addition to the non-targeted analysis.

SOMMAIRE

REMERCIEMENTS ..... **I**

RESUME ..... **IV**

LISTE DES FIGURES ..... **XII**

LISTE DES TABLEAUX ..... **XVII**

INTRODUCTION GENERALE **1**

CHAPITRE 1 : BIBLIOGRAPHIE **4**

I. LES MICROPOLLUANTS : ORIGINES ET DEVENIR DANS LE CYCLE TECHNIQUE DE L'EAU **5**
I.1. Généralités 5
I.2. Origines de la présence et niveau de contamination en micropolluants dans les eaux. 5
I.2.1. Les vecteurs de pollution de l'eau en milieu anthropisé vers les STEP 7
I.2.1.1. Les eaux domestiques 7
I.2.1.2. Les eaux non-domestiques 8
I.2.1.2.1. Les rejets hospitaliers 8
I.2.1.2.2. Les rejets industriels 8
I.2.1.3. Les eaux de ruissellements urbains 9
I.2.2. Les micropolluants dans les stations de traitement des eaux usées 10
I.2.3. Voies d'élimination des micropolluants organiques dans les STEP conventionnelles 12
I.2.4. Abattement des micropolluants dans les STEP 13
I.2.5. Contamination des eaux de rivières par les micropolluants 14
I.2.5.1. Sources de contamination des rivières 14
I.2.5.2. Niveaux de contamination des eaux de surface 15
I.3. Impact de la présence des micropolluants organiques dans le petit cycle de l'eau 16
I.3.1. Impacts sur l'Homme 17
I.3.2. Impacts sur les milieux 18
I.4. Conclusions 19
II. APPROCHES D'ANALYSE DES MICROPOLLUANTS DANS LES DIFFERENTES MATRICES D'EAUX **19**
II.1. Analyse ciblée 19
II.1.1. Principes de l'analyse ciblée 19
II.1.2. Méthodes de préparations et méthodes instrumentales utilisées lors de l'analyse ciblée de micropolluants dans les eaux 20
II.1.2.1. Méthodes de préparations d'échantillons 20
II.1.2.2. Méthodes instrumentales pour l'analyse ciblée 21
II.1.3. Limites de l'analyse ciblée 22
II.2. La spectrométrie de masse haute résolution 23
II.3. Recherche de suspects ou « screening » 27
II.3.1. Principe de l'analyse de suspects ou suspect screening 27
II.3.2. Développement d'une méthode de screening 27
II.3.2.1. Etapes de l'analyse de suspects et de la confirmation d'identification 27

II.3.2.2. Développement et optimisation de la méthode 28
II.3.2.3. Applications de l'analyse de suspects 28
II.4. Analyse non ciblée 29
II.4.1. Principe de l'analyse non ciblée 29
II.4.2. Étapes de l'analyse non ciblée 29
II.4.3. Les méthodes statistiques d'analyse de données 32
II.4.3.1. Principes généraux et méthodes 32
II.4.3.1.1. Analyses univariées 32
II.4.3.1.2. Couplage analyses univariées – Fold Change 33
II.4.3.1.3. Analyses multivariées 35
II.4.3.2. Exemples d'applications dans le cas d'analyses de l'eau 36
II.4.4. Les approches de priorisation 40
II.4.4.1. Priorisation par les données 40
II.4.4.2. Priorisation par les expériences 42
II.4.5. Stratégies d'identification et de caractérisation 45
II.4.5.1. Identification 45
II.4.5.2. Caractérisation par les formules brutes 46
II.5. Conclusions 49

CHAPITRE 2 : MATERIELS ET METHODES 50

I. CAMPAGNES DE PRELEVEMENTS 51
I.1. Sites de prélèvements 51
I.1.1. Station de traitement des eaux usées de Poitiers 51
I.1.2. Station de traitement des eaux usées de Saint Julien l'Ars 52
I.1.3. Eaux de rivières 53
I.1.3.1. Le Clain 53
I.1.3.2. La Vienne 54
I.2. Campagnes d'échantillonnage et mode de prélèvement 55
I.2.1. Campagnes de prélèvements 55
I.2.2. Mode de prélèvement 56
II. REACTIFS ET ETALONS 56
II.1. Réactifs 56
II.2. Solutions étalons 57
III. PREPARATION D'ECHANTILLONS 59
III.1. Centrifugation et filtration 60
III.1.1. Centrifugation 60
III.1.2. Filtration 60
III.2. Extraction sur phase solide 60
III.2.1. Protocole 60
III.2.2. Estimation des effets matrices 62
III.2.3. Estimation des rendements d'extraction de la méthode SPE 62
III.3. Gestion de la contamination 63
IV. METHODES D'ANALYSE 63
IV.1. Analyse des paramètres usuels 63
IV.2. Analyse par chromatographie liquide couplée à la spectrométrie de masse haute résolution 64
IV.2.1. Appareillage 64
IV.2.2. Conditions Chromatographiques 64

IV.2.3. Paramètres de spectrométrie de masse 65
IV.2.4. Séquence d'analyse 65
V. TRAITEMENT DES DONNEES 65
V.1. Prétraitement des données brutes 66
V.1.1. Méthode par la combinaison des logiciels MZmine 2, R et Excel 66
V.1.2. Méthode par la combinaison des logiciels Compound Discoverer et Excel 69
V.2. Caractérisation et identification 72
V.3. Analyses statistiques 73
V.3.1. Réalisation des empreintes des échantillons 73
V.3.2. Réalisation des box-plots ou boîtes à moustaches 73
V.3.3. Analyse en Composantes Principales 73
V.3.4. Analyses différentielles 73
V.3.5. Génération de la liste de métabolites 74

CHAPITRE 3 : MISE AU POINT DE L'APPROCHE POUR L'ANALYSE NON CIBLEE : DU PRELEVEMENT DES ECHANTILLONS AU TRAITEMENT DES DONNEES 75

I. INTRODUCTION 76
II. OPTIMISATION DES PARAMETRES ET COMPARAISON DES RESULTATS OBTENUS LORS DU PRETRAITEMENT DES DONNEES PAR DEUX METHODES : MZMINE 2 VS COMPOUND DISCOVERER 77
II.1. Traitement des données avec l'ensemble MZmine 2, R et Excel 78
II.1.1. Étapes importantes du traitement des données 78
II.1.2. Optimisation des paramètres de traitement des données et résultats obtenus 78
II.1.2.1. Choix des paramètres de MZmine 2 78
II.1.2.1.1. Détection des pics 78
II.1.2.1.2. Reconstruction des chromatogrammes 81
II.1.2.1.3. Déconvolution des chromatogrammes 84
II.1.2.1.4. Alignement des signaux et « Gap-filling » 85
II.1.2.1.5. Étapes supplémentaires 88
II.1.2.2. Choix des paramètres de Excel et R 88
II.1.2.2.1. Détection des isotopes et des adduits par les logiciels R 3.5.1 et R.studio 88
II.1.2.2.2. Soustraction des blancs avec Excel 89
II.1.2.3. Méthode finale retenue pour le prétraitement des données lors de l'utilisation de MZmine 2, Excel et R 89
II.2. Traitement des données avec la méthode utilisant Compound Discoverer 89
II.2.1. Le schéma d'étapes ou « workflow » 89
II.2.2. Optimisation du traitement des données par Compound Discoverer 90
II.2.2.1. Alignement des temps de rétention 90
II.2.2.2. Détection et regroupement des signaux, recherche des signaux manquants et comparaison avec les blancs 91
II.2.3. Méthode retenue pour le prétraitement des données par Compound Discoverer 94
III.3. Comparaison des résultats et choix de la méthode de retraitement des données 95
III.3.1. Comparaison des résultats obtenus avec les deux méthodes de retraitement des données 95
II.3.2. Choix de la méthode de retraitement des données 98
III. ÉTUDE DE L'IMPACT DES REPETABILITES D'INJECTION ET DE PRELEVEMENT 100
III.1. Impact de la répétabilité d'injection 100
III.2. Impact de la répétabilité de prélèvements 103

IV. IMPACT DE L'APPLICATION D'UNE METHODE D'EXTRACTION/CONCENTRATION DES ECHANTILLONS 106
IV.1. Choix de la méthode d'extraction 106
IV.2. Impact de la méthode d'extraction sur un mélange d'étalons et détermination des rendements d'extraction 107
IV.2.1. Estimation de l'effet matrice 107
IV.2.2. Rendements d'extraction de la méthode SPE 111
IV.3. Comparaison des résultats obtenus pour l'ensemble des signaux avec et sans extraction par SPE 113
V. CONCLUSIONS 116

CHAPITRE 4 : CARACTERISATION DES EAUX USEES BRUTES ET TRAITEES PAR ANALYSE NON CIBLEE ET RECHERCHE DE SUSPECTS 118

I. INTRODUCTION 119
II. CARACTERISATION ET COMPARAISON DES EAUX D'ENTREE DE STEP 120
II.1. Comparaison du nombre total de signaux par échantillon 120
II.2. Empreintes des eaux d'entrée de STEP 121
III. IMPACT DU TRAITEMENT : COMPARAISON DES EAUX D'ENTREE ET DE SORTIE DE STEP 125
III.1. Caractérisation des eaux de sortie de STEP 126
III.2. Comparaison entre les eaux d'entrée et les eaux de sortie de STEP 129
III.2.1. Comparaison du nombre total de signaux 129
III.2.2. Comparaison des entrées et des sorties par leurs statistiques descriptives 130
III.2.3. Comparaison en fonction des tendances de variation des signaux et analyses différentielles 133
IV. CORRELATION ENTRE SIGNAUX ET NATURE DES ECHANTILLONS PAR ANALYSE MULTIVARIEE 139
V. RECHERCHE DE SIGNAUX PERTINENTS DANS LES EAUX DE STEP 144
V.1. Sélection des signaux par analyses différentielles 144
V.2 Approches de caractérisation 146
V.2.1. Attribution de formules brutes par Compound Discoverer 146
V.2.2. Proposition d'annotations par analyses des spectres MS^2 151
V.3. Recherche des signaux pertinents de la liste de Poitiers dans les échantillons de Saint Julien l'Ars 154
VI. RECHERCHE DE SUSPECTS ET DE METABOLITES DANS LES EAUX DE STEP 155
VI.1. Recherche de suspects dans les eaux de STEP 155
VI.1.1. Création et mise en place d'une liste de suspects 155
VI.1.2. Application de l'analyse de suspects dans les eaux de STEP 156
VI.1.3. Confrontation entre les suspects détectés et la liste de signaux pertinents 160
VI.2. Recherche de métabolites complémentaires dans les eaux de STEP 161
VI.2.1. Constitution de la liste de métabolites suspects 161
VI.2.2. Recherche de métabolites suspects 162
VII. CONCLUSIONS 163

CHAPITRE 5 : ANALYSE NON CIBLEE APPLIQUEE AUX EAUX DE RIVIERES ET IMPACT DES REJETS DE LA STEP DE POITIERS SUR LES EAUX DU CLAIN 165

I. INTRODUCTION 166
I.1. Caractérisation d'eaux de rivières : comparaison entre le Clain à l'amont du rejet de la STEP et la Vienne 166

II. VARIABILITE SPATIALE D'UNE EAU DE SURFACE EN ZONE ANTHROPISEE 170
II.1. Comparaison globale du nombre de signaux 170
II.2. Comparaison des échantillons par les statistiques descriptives 171
II.3. Caractérisation d'une zone de rejet de STEP en rivière : comparaison entre le panache du rejet et son aval immédiat 173
II.4. Evaluation de la pression des eaux traitées de la STEP au point de rejet : comparaison des sorties de la STEP et du panache 178
II.5. Impact sur le milieu aquatique : Comparaison Aval/Amont 180
II.6. Corrélation entre signaux et nature des échantillons par analyse multivariée 182
III. RECHERCHE DE SIGNAUX PERTINENTS ENTRE L'AMONT ET L'AVAL 184
III.1. Sélection des signaux par analyse différentielle 184
III.2. Approches de caractérisation 186
III.2.1. Attribution de formules brutes par Compound Discoverer 186
III.2.2. Proposition d'annotations par analyse des spectres MS^2 189
IV. RECHERCHE DE SUSPECTS ET DE METABOLITES 190
IV.1. Recherche de suspects dans les eaux du Clain 190
IV.2. Recherche de métabolites complémentaires dans les eaux du Clain 193
V. CONCLUSION 195

CHAPITRE 6 : BILAN ET ETUDE CRITIQUE 196

I. INTRODUCTION 197
II. COMPARAISON DES SUSPECTS ET METABOLITES SUSPECTS TROUVES DANS LES EAUX DE LA SORTIE DE LA STEP ET DU CLAIN 197
III. ETUDE CRITIQUE SUR LA RECHERCHE DE SIGNAUX PERTINENTS 199
III.1. Comparaison des signaux pertinents obtenus avec les deux approches 202
III.2. Recherche des signaux pertinents obtenus avec l'approche 1 « traiteur d'eau » (sortie/entrée STEP) dans le Clain 203
III.3. Recherche des signaux pertinents obtenus avec l'approche 2 « environnementaliste » (Aval/Amont du Clain) dans la STEP 204
IV. BILAN : ANALYSE MULTIVARIEE SUR L'ENSEMBLE DES POINTS DE PRELEVEMENTS D'UNE CAMPAGNE 205
V. CONCLUSION 206

CONCLUSIONS ET PERSPECTIVES 209

ANNEXES 216

REFERENCES 274

Liste des Figures

Figure 1 : Schéma simplifié du petit cycle de l'eau. 6
Figure 2 : Représentation des sources de contamination des eaux de rivières par les micropolluants. 15
Figure 3 : Illustration du pouvoir de résolution par l'exemple de l'insuline bovine issu de Gross et al., 2011. 24
Figure 4 : Schéma général des étapes à réaliser lors d'une analyse non ciblée pour une application environnementale. 32
Figure 5 : Volcano-plot représentant -log10(p-value) en fonction du log2(Fold Change) Source : Schollée et al., 2016. 34
Figure 6 : Loading plot de l'ACP ayant servi de base au classement des signaux. Repris de Schollée et al. (2015). 37
Figure 7 : Exemples des tendances pouvant être suivies par les intensités des signaux au cours des étapes du traitement des eaux et déterminées par les clusters isolés par la classification hiérarchique. Repris de Schollée et al. (2018). 39
Figure 8 : Classification des niveaux de confiance pour l'indentification des signaux selon Schymanski et al. 2015. 46
Figure 9 : Figure de Van Krevelen réarrangée de Verkh et al. (2018) du rapport H/C en fonction du rapport O/C. 48
Figure 10 : Schéma des trois procédures d'analyses pour la réalisation de l'analyse ciblée, de l'analyse de suspects et de l'analyse non ciblée via la spectrométrie de masse haute résolution. 49
Figure 11 : Image satellite de la station d'épuration de Poitiers avec représentation des points de prélèvements en rouge. 52
Figure 12 : Image satellite de la station d'épuration de Saint Julien de l'Ars avec représentation des points de prélèvements en rouge.. 53
Figure 13 : Images satellites des sites de prélèvements Clain aval, panache et amont représentés par des points de prélèvements en rouge.. 54
Figure 14 : Images satellites du site de prélèvements dans la Vienne 55
Figure 15 : Protocole de la préparation des échantillons pour l'application de l'analyse non ciblée. 59
Figure 16 : Schéma modifié des structures chimiques des phases solides des cartouches Oasis utilisées en SPE. Source : Waters. 61
Figure 17 : Capture d'écran du workflow utilisé pour le traitement des données issues de l'analyse UHPLC-HRMS. 70
Figure 18 : Nombre de chromatogrammes reconstruits après détection des masses en mode positif en fonction du seuil de bruit pour différents types d'échantillons.. 79
Figure 19 : Graphique en radar du nombre d'étalons détectés en mode positif par niveau de bruit paramétré lors de l'étape de détection des masses et pour différentes concentrations 80
Figure 20 : Nombre de chromatogrammes reconstruits en fonction du nombre de scans en mode positif par type d'échantillons. 82

Figure 21 : Graphique en radar du nombre d'étalons détectés en mode positif par nombre de scans minimal paramétré lors de l'étape de reconstruction des chromatogrammes et pour différentes concentrations. 83
Figure 22 : Capture d'écran d'un exemple de déconvolution de deux signaux au m/z 344,2274 selon les réglages finaux..... 85
Figure 23 : Capture d'écran des chromatogrammes d'un exemple de signaux alignés au m/z 272,2570 selon les réglages finaux.. 86
Figure 24 (A et B) : Captures d'écran des chromatogrammes d'un exemple d'un signal d'un m/z 109,1014 présent dans les 3 injections d'un même échantillon d'entrée de STEP selon les réglages finaux..... 87
Figure 25 : Nombre total de signaux en fonction de la valeur d'intensité paramétrée en mode positif. Des valeurs d'intensités ont été testées entre $5x10^4$et $5x10^5$.. 92
Figure 26 : Graphique en radar du nombre d'étalons détectés par seuil d'intensité minimale paramétré lors de l'étape de détection des masses et par concentration.. 93
Figure 27 : Box plots ou boîtes à moustaches résumant les échantillons par leurs statistiques descriptives. 96
Figure 28 : Empreintes d'échantillons de l'entrée en A et de la sortie Nord en B de la STEP de Poitiers et concentrés par SPE..... 97
Figure 29 : Graphique des observations (échantillons) à partir des résultats de Compound Discoverer en A et de MZmine 2 en B. 98
Figure 30 : Nombre de signaux en fonction d'intervalles de coefficients de variation (en %) selon le type d'échantillons (pour 3 injections)..... 101
Figure 31 : Diagrammes de Venn illustrant le nombre de signaux communs entre les trois réplicas d'échantillons pour chaque type de matrice étudiée. 103
Figure 32 : Nombre de signaux en fonction d'intervalles de coefficients de variation (en %) selon le type d'échantillons 105
Figure 33 (A et B) : Diagrammes en barres représentant l'effet matrice en pourcentage en fonction des étalons pour l'entrée de STEP en A et la sortie de STEP en B.. 108
Figure 34 : Diagrammes en barres du rapport des aires des étalons entre l'entrée et la sortie Nord en A et l'amont et l'aval du Clain en B. 110
Figure 35 : Diagrammes en barres des rendements d'extraction dans l'eau ultra-pure, le Clain amont et aval en A, l'eau ultra-pure, l'entrée et la sortie de STEP de Poitiers en B. 112
Figure 36 : Empreinte d'échantillons concentrés de l'entrée et de la sortie de la STEP de Poitiers en A et en B et de l'amont du Clain en C. 114
Figure 37 : Graphique en barres représentant le nombre de signaux en fonction de l'échantillon considéré. . 121
Figure 38 : Empreintes des échantillons d'entrées de la STEP Poitiers pour les campagnes A, B et C. 122
Figure 39 : Empreintes des échantillons d'entrées de la STEP de St Julien l'Ars pour les campagnes A et B annotées 122
Figure 40 : Chromatogrammes d'ions extraits des étalons de metformine, caféine, propranolol, carbamazépine et oxadiazon issus de l'analyse d'un mélange d'étalons à une concentration de 100 µg/L..... 123
Figure 41 : Empreintes des échantillons de sorties de la STEP de Poitiers pour les campagnes A, B et C. 127
Figure 42 : Empreintes des échantillons de sorties de la STEP de St Julien l'Ars pour les campagnes A et B..... 128

Figure 43 : Graphique en barres représentant le nombre de signaux en fonction de l'échantillon considéré. . 129
Figure 44 : Box plots ou boîtes à moustaches résumant les échantillons par leurs statistiques descriptives. ... 131
Figure 45 : Graphique en barres représentant le nombre de signaux en fonction de leur tendance de variation par rapport à l'entrée correspondante. 133
Figure 46 : Empreintes montrant les signaux détectés en fonction de leur tendance associée..... 135
Figure 47 : Empreintes montrant les signaux détectés en fonction de leur tendance associée..... 136
Figure 48 : Volcano-plot des analyses différentielles entre les données des analyses d'entrée et celles des deux sorties par campagne (A, B et C) pour Poitiers..... 138
Figure 49 : Volcano-plot des analyses différentielles entre les données des analyses d'entrée et celles des deux sorties et lagune par campagne (A, B et C) pour St Julien l'Ars 139
Figure 50 : Graphiques des échantillons de Poitiers en A et des échantillons de St-Julien en B 141
Figure 51 : Biplots des échantillons et des signaux détectés selon les facteurs F1 et F2 et les tendances de variation associées pour Poitiers en A et St Julien l'Ars en B. 143
Figure 52 : Empreinte montrant les signaux détectés en fonction de leur tendance associée. 145
Figure 53 : Biplots des échantillons et des signaux détectés selon les facteurs F1 et F2 pour la STEP de Poitiers. 146
Figure 54 : Graphiques du ratio H/C en fonction des ratios O/C et N/C des signaux sélectionnés par analyse différentielle. 147
Figure 55 : Empreinte des 331 signaux pour lesquels une formule brute a été proposée. (A) selon la composition, (B) identification de trois séries d'homologues. 150
Figure 56 : Schéma d'étapes de l'analyse de suspects ayant conduit à l'obtention d'une liste de 198 signaux confrontés à mzCloud ou MetFrag (KEGG, HMDB et Comptox). 158
Figure 57 : Empreintes des échantillons de l'amont du Clain pour les campagnes A, B et C. 167
Figure 58 : Empreintes des échantillons de la Vienne pour les campagnes A et B. 167
Figure 59 : Graphique en barres représentant le nombre de signaux détectés en fonction de l'échantillon considéré 169
Figure 60 : Graphique en barres représentant le nombre de signaux en fonction de l'échantillon considéré. . 170
Figure 61 : Box plots ou boîtes à moustaches résumant les échantillons par leurs statistiques descriptives. ... 172
Figure 62 : Empreintes pour les campagnes A, B et C précisant pour les signaux détectés leur tendance associée correspondant au calcul de la variation de l'aire des signaux de l'aval par rapport au panache 174
Figure 63 : Graphique en barres représentant le nombre de signaux en fonction de leur tendance de variation par échantillon par rapport au panache. 176
Figure 64 : Empreintes précisant pour les signaux détectés leur tendance associée correspondant au calcul de la variation de l'aire des signaux de la sortie Nord de la STEP de Poitiers par rapport au panache du rejet dans le Clain. 178
Figure 65 : Graphique en barres représentant le nombre de signaux en fonction de leur tendance de variation par échantillon par rapport à la sortie de la STEP. 179
Figure 66 : Empreintes précisant pour les signaux détectés leur tendance de variation associée correspondant au calcul de la variation de l'aire des signaux de l'aval par rapport à l'amont. 180

Figure 67 : Graphique en barres représentant le nombre de signaux en fonction de leur tendance de variation par échantillon par rapport à l'amont. 181
Figure 68 : Biplot des échantillons et des signaux détectés selon les facteurs F1 et F2 et les tendances de variation associées calculées entre l'amont et l'aval pour les échantillons du Clain. 183
Figure 69 : Graphiques de type volcano-plot des analyses différentielles entre chaque échantillon de l'aval par rapport à l'amont, par campagne 185
Figure 70 : Empreinte de l'ensemble des signaux considérés comme pertinents, sélectionnés par analyse différentielle entre les échantillons amont et aval selon leur tendance moyenne. 186
Figure 71 : Graphiques du ratio H/C en fonction des ratios O/C et N/C des signaux sélectionnés par analyse différentielle. 187
Figure 72 : Schéma d'étapes de l'analyse de suspects ayant conduit à l'obtention d'une liste de 127 signaux confrontés à mzCloud ou MetFrag (KEGG, HMDB et Comptox). 191
Figure 73 : Diagrammes des Venn des signaux communs entre la sortie de STEP, l'amont et l'aval du Clain dans le cas des suspects et des métabolites suspects. 198
Figure 74 : Diagramme de Venn des signaux pertinents communs obtenus par l'approche 1 (« traiteur d'eau ») et l'approche 2 (« environnementaliste »). 202
Figure 75 : Diagramme de Venn des signaux pertinents sélectionnés par l'approche 1 (« traiteur d'eau ») détectés dans les échantillons amont et aval. 203
Figure 76: Diagramme de Venn des signaux pertinents sélectionnés par l'approche 2 (« environnementaliste ») détectés dans les échantillons d'entrée et de sorties. Les nombres entre parenthèses indiquent le nombre de signaux ayant un spectre MS2. 204
Figure 77 : Biplot des points de prélèvements de la campagne C et des signaux détectés selon les axes F1 et F2 (A) et selon les axes F1 et F3 (B). 205

Figures des Annexes

Figure 78 : Liste des filtres appliqués sur le tableau de données dans Compound Discoverer lors de l'étude d'échantillons de STEP 217
Figure 79 : Liste des filtres appliqués sur le tableau de données dans Compound Discoverer lors de l'étude d'échantillons de rivières 217
Figure 80 : Empreintes des échantillons d'entrée et de sorties de la campagne A avec relief des homologues recherchés par la package « non-target ». 219
Figure 81 : Empreintes des échantillons d'entrée et de sorties de la campagne B avec relief des homologues recherchés par la package « non-target ». 220
Figure 82 : Empreintes des 3 échantillons d'entrées de la campagne C avec relief des homologues recherchés par la package « non-target ». 221
Figure 83 : Empreintes des 3 échantillons de sorties de la campagne C avec relief des homologues recherchés par la package « non-target ». 222

Figure 84 : Chromatogrammes extraits et spectres MS^2 de l'étalon de la caféine, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 234
Figure 85 : Chromatogrammes extraits et spectres MS^2 de l'étalon de la carbamazépine, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 234
Figure 86 : Chromatogrammes extraits et spectres MS2 de l'étalon de la ciprofloxacine, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 235
Figure 87 : Chromatogrammes extraits et spectres MS2 de l'étalon de l'ioméprol, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 235
Figure 88 : Chromatogrammes extraits et spectres MS2 de l'étalon du kétoprofen, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 236
Figure 89 : Chromatogrammes extraits et spectres MS2 de l'étalon de l'oxybenzone, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 236
Figure 90 : Chromatogrammes extraits et spectres MS2 de l'étalon du propranolol, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 237
Figure 91 : Chromatogrammes extraits et spectres MS2 de l'étalon du valsartan, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification. 237
Figure 92 : Empreintes des échantillons de panache du Clain pour les campagnes A, B et C. 245
Figure 93 : Empreintes des échantillons de l'aval du Clain pour les campagnes A, B et C. 246
Figure 94 : Chromatogrammes extraits et spectres MS2 de l'étalon de la carbamazépine et d'échantillons amont et aval du Clain pour la confirmation de l'identification. 253
Figure 95 : Chromatogrammes extraits et spectres MS2 de l'étalon du paracétamol et d'échantillons amont et aval du Clain pour la confirmation de l'identification. 253
Figure 96 : Chromatogrammes extraits et spectres MS2 de l'étalon du propranolol et d'échantillons amont et aval du Clain pour la confirmation de l'identification. 254

Liste des Tableaux

Tableau 1 : Gammes de concentrations minimales et maximales détectées supérieures aux limites de quantifications pour plusieurs familles de micropolluants organiques dans les affluents et les effluents de STEP. 12

Tableau 2 : Gammes de concentrations détectées supérieures aux limites de quantifications de plusieurs familles de micropolluants organiques dans des eaux de surface en Europe. 16

Tableau 3 : Gammes de ratios associées aux différentes classes de molécules (Maizel et Remucal, 2017 ; Hockaday et al., 2009). 47

Tableau 4 : Récapitulatif des échantillons prélevés et des dates de campagne de prélèvements. 56

Tableau 5 : Liste des étalons utilisés dans le cadre des optimisations de traitement de données, de calculs d'effet matrice et des calculs de rendements d'extraction. 57

Tableau 6 : Etapes de la méthode d'extraction sur phase solide à trois cartouches. 61

Tableau 7 : Paramètres de la chromatographie en phase liquide. 64

Tableau 8 : Synthèse des étapes, des paramètres et de leurs valeurs dans le cadre de l'utilisation de l'ensemble de logiciels MZmine 2, R 3.5.1 et Excel. 67

Tableau 9 : Synthèse des étapes, des paramètres et de leurs valeurs pour le prétraitement des données par Compound Discoverer-Excel. 70

Tableau 10 : Nombre d'étalons détectés par nature d'échantillon en fonction du seuil de bruit sélectionné. Concentrations des étalons fixés à 100 µg/L. 80

Tableau 11 : Nombre de signaux détectés en fonction des différentes valeurs de tolérance sur le temps de rétention ou sur la tolérance en masse. 91

Tableau 12 : Avantages et inconvénients pour chacune des deux méthodes développées. 99

Tableau 13 : Nombre de signaux par matrice en fonction du nombre de réplicas d'injection (n = 1 à 3). 102

Tableau 14 : Nombre de signaux pour les trois prélèvements successifs (1, 2 et 3) selon la matrice (campagne C) avec nombre moyen, écart-type, coefficient de variation et estimation d'une incertitude relative par matrice. 104

Tableau 15 : Paramètres physico-chimiques de caractérisation des eaux en entrée de STEP (Poitiers et Saint Julien l'Ars, pour les campagnes A, B et C). 124

Tableau 16 : Nombre de signaux en fonction de la composition, de la catégorie et du nombre d'atomes de carbones parmi les 331 signaux pour lesquels une formule brute a pu être associée. 149

Tableau 17 : Liste des 7 signaux ayant une proposition d'identification par mzCloud avec une déviation en masse inférieure à 5 ppm et une correspondance supérieure à 80%. 152

Tableau 18 : Concentrations estimées pour les 8 suspects confirmées dans un échantillon d'entrée et un échantillon de sortie de STEP de la campagne C. 159

Tableau 19 : Gamme de concentrations des classes de molécules détectées dans les eaux de STEP par Grand Poitiers. 159

Tableau 20 : Liste des signaux présents dans la liste de 198 signaux suspects et la liste des 155 signaux pertinents ayant un spectre MS^2. 160
Tableau 21 : Liste des métabolites suspects ayant une correspondance avec mzCloud. 162
Tableau 22 : Répartition des signaux pour les échantillons du panache et de l'aval du Clain selon les 3 descripteurs : temps de rétention, masse moléculaire et valeur de l'aire des signaux. 175
Tableau 23 : Paramètres physico-chimiques de caractérisation des eaux panache du rejet et aval, pour les campagnes A, B et C. 177
Tableau 24 : Nombre de signaux en fonction de la composition, de la catégorie et du nombre d'atomes de carbones parmi les 273 signaux pour lesquels une formule brute a pu être associée. 188
Tableau 25 : Liste des 4 signaux ayant une proposition d'identification par mzCloud avec une déviation en masse inférieure à 5 ppm et une correspondance supérieure à 80%. 189
Tableau 26 : Liste des signaux présents dans la liste de 127 signaux suspects et la liste des 52 signaux pertinents ayant un spectre MS^2. 192
Tableau 27 : Liste des métabolites suspects ayant une correspondance avec mzCloud. 194
Tableau 28 : Listes des formules développées proposées pour les signaux pertinents obtenus à partir des approches 1 et 2. 199

Tableaux des Annexes

Tableau 29 : Nombre de signaux par tendance calculé entre les échantillons non concentrés et les échantillons concentrés par extraction sur phase solide. 218
Tableau 30 : Liste des propositions de structures attribuées à 25 signaux pertinents par l'utilisation de mzCloud 223
Tableau 31 : Liste des 198 signaux correspondant à un ou plusieurs suspects. 225
Tableau 32 : Liste des 60 molécules parentes utilisées pour générer une liste de métabolites. 238
Tableau 33 : Liste des 89 signaux correspondant à un ou plusieurs métabolites. 240
Tableau 34 : Liste des propositions de structures attribuées à 13 signaux pertinents par l'utilisation de mzCloud. 247
Tableau 35 : Liste des 127 signaux correspondant à un ou plusieurs suspects 248
Tableau 36 : Liste des 34 signaux correspondant à un ou plusieurs métabolites. ..... **Erreur ! Signet non défini.**
Tableau 37 : Liste des 36 signaux communs entre les 2 listes de signaux pertinents déterminés à partir de la comparaison amont-aval et à partir de sorties de STEP, associés à une proposition de nom par mzCloud lorsque celle-ci existe. 259
Tableau 38 : Liste des signaux pertinents issus des approches 1 et 2 selon leur hiérarchisation « très probable » ou « probable ». 260

Introduction générale

L'essor de l'assainissement et des égouts date du XIXème siècle marque en France, impulsé par la volonté d'évacuer les eaux usées des villes pour des raisons d'hygiène. Mais c'est au XXème siècle que se développent les procédés modernes physico-chimiques et par boues activités de traitement des eaux (Duchène, 2005). Ainsi le milieu du XXème siècle voit se développer les stations d'épuration (STEP) y compris dans les zones rurales. Aujourd'hui ces stations d'épuration sont confrontées à une problématique pour laquelle elles n'ont pas été conçues initialement : la présence de micropolluants organiques dans les eaux usées.

Les micropolluants organiques sont pour la plupart des molécules anthropogéniques issues de l'industrie et du mode de vie moderne. Ces micropolluants comprennent des composés de diverses familles de molécules que l'on peut classer selon leur usage ou leur type (produits pharmaceutiques et cosmétiques, pesticides, produits de consommation, composés issus de l'industrie ou encore parabènes, retardateurs de flammes, hormones etc…) et forment une famille hétérogène du point de vue des propriétés physico-chimiques. En effet, les micropolluants organiques peuvent être des molécules très polaires ou au contraire apolaires. Ils peuvent être de petites tailles (inférieures à 100 Da) ou de grandes tailles, tels que des polymères. Si les effets délétères de nombreux micropolluants sont connus, pour une partie importante d'entre eux, on ne connait pas ou très peu leur impact à court et à long terme sur l'environnement et la santé humaine ou animale. De plus, ces molécules subissent des transformations chimiques au cours de leur dissémination dans les milieux ce qui complexifie le problème de la protection des ressources en eaux et la réduction de la pollution.

L'efficacité des STEP sur l'abattement des micropolluants est variable. Certains d'entre eux sont assez bien éliminés (abattement supérieur à 95%) tandis que d'autres sont très peu affectés par les filières classiques, avec un abattement inférieur à 20% (Margot *et al.*, 2015). En conséquence, les STEP sont potentiellement une source de contaminants pour les eaux de surface et les milieux aquatiques. Les micropolluants ont été détectés, à ce jour, dans de nombreux compartiments, à de faibles concentrations dans les eaux de surface (allant du ng/L au µg/l) mais à des concentrations pouvant s'élever à plusieurs centaines de µg/L dans les entrées de STEP (Guerra *et al.*, 2014 ; Kasprzyk-Hordern *et al.*, 2009 ; Loos *et al.*, 2013 ; Singer *et al.*, 2010).

La Directive Cadre sur l'Eau (2000/60/CE) et d'autres directives liées à la protection des eaux souterraines et aux normes de qualité environnementale des eaux fixent le cadre réglementaire pour la protection des ressources en eaux et des écosystèmes. Dans cette directive et dans la législation française sont listées des molécules dangereuses prioritaires et des molécules prioritaires avec un objectif de réduction des émissions. L'absence de ces molécules participe à la qualification de bon état chimique des eaux. Ces listes sont mises à jour en fonction des améliorations réalisées et des nouvelles données scientifiques. Les produits pharmaceutiques et les perturbateurs endocriniens font, par

exemple, partie des nouveaux contaminants étudiés pour la réglementation. Dans un registre plus général, le règlement européen REACH (Registration, Evaluation and Autorisation of Chemicals) s'attache à enregistrer et à évaluer les composés chimiques dans l'Union Européenne dans le but d'améliorer la protection de l'environnement et de la santé humaine. En 2018, plus de 20 000 molécules étaient référencés dans REACH. En ce qui concerne le rejet des contaminants dans l'environnement et les milieux aquatiques, la majorité de ces molécules reste non réglementée à ce jour. Des projets de recherche de micropolluants sont aussi régulièrement menés dans le but de surveiller leurs concentrations et leur dissémination dans les différents compartiments. Les projets AMPERES et SIPIBEL en France sont des exemples de projets ayant, entre autres, étudié la présence et le devenir de micropolluants notamment dans les STEP.

Les progrès analytiques jouent un rôle dans l'acquisition de nouvelles données et dans la détection de très faibles concentrations mettant à jour des contaminations jusqu'alors ignorées. Des méthodes détectant et quantifiant parfois plusieurs centaines de composés dans la même analyse ont été mises au point et opèrent en routine. Ces méthodes dépendent cependant de listes préétablies qui répondent à des choix effectués dans les projets. Un biais de sélection vers des molécules connues est présent, mettant de côté des composés jusqu'à présent inconnus mais pouvant présenter un risque pour le milieu aquatique et/ou la santé humaine. Depuis une quinzaine d'années, les progrès de la spectrométrie de masse haute résolution couplée à la chromatographie liquide et le développement des bases de données ont permis le développement d'un nouveau schéma d'analyse environnementale à l'image de la métabolomique : l'analyse non ciblée appliquée aux eaux (Schymanski *et al.*, 2017). L'un des objectifs de l'analyse non ciblée est d'analyser des échantillons sans *a priori* pour y rechercher des signaux voire des molécules s'avérant pertinents au regard d'une problématique donnée : toxicité, écotoxicité, marqueurs de pollution ou encore recherche de produits de transformations.

Ce sujet de thèse s'inscrit dans ce contexte de l'utilisation de nouvelles méthodes d'analyses pour répondre à la problématique des polluants émergents dans les eaux de STEP et les eaux de rivières par la mise en évidence de signaux traceurs de sorties de STEP ou impactant les rivières.

Les objectifs de cette thèse sont alors :

- le développement d'une méthode d'analyse non ciblée appliquée à des eaux de stations d'épuration et des eaux de rivières,
- l'étude de l'impact du traitement des eaux usées sur la composition globale,
- l'étude de l'impact du rejet des eaux de sortie de STEP sur une rivière,
- la recherche de signaux pertinents marqueurs de ce rejet.

La première partie du manuscrit dresse l'état de l'art sur les micropolluants, leurs sources et leurs niveaux de concentrations dans différents compartiments du cycle technique de l'eau et leurs impacts. Les approches de l'analyse classique des micropolluants puis l'analyse non ciblée et ses étapes sont ensuite abordées.

La seconde partie développe les méthodes et les matériels utilisés au cours des travaux de thèse.

La troisième partie présente les résultats obtenus selon quatre sous-parties :

- Le développement de la méthode de prétraitement des données de l'analyse non ciblée à partir de deux ensembles de logiciels (MZmine 2 couplé à R et Excel d'une part et Compound Discoverer d'autre part) est tout d'abord présenté. Pour chacun des logiciels, les étapes, les paramètres et les réglages testés sont détaillés puis la comparaison des résultats entre les deux logiciels a été effectuée pour déterminer le logiciel utilisé dans la suite des travaux de thèse. Enfin, la répétabilité analytique et de prélèvement a été étudiée ainsi que l'impact de la méthode d'extraction sur phase solide mise en œuvre.
- La seconde sous-partie porte sur l'analyse d'eaux d'entrées et de sorties de STEP selon trois campagnes de prélèvements. La caractérisation des eaux d'entrées a été effectuée et l'impact du traitement des eaux a été étudié. Une recherche de signaux pertinents par analyses différentielles a été réalisée. Les signaux ainsi sélectionnés ont été caractérisés et une tentative d'identification a été effectuée. Des analyses de suspects portant sur un ensemble de composés et de métabolites concluent cette sous-partie.
- La troisième se focalise sur la caractérisation d'eaux de rivières et plus particulièrement sur la rivière Clain. La variabilité spatiale de la composition du Clain dans une zone anthropisée est ensuite exposée et sert de point de départ pour la sélection de signaux pertinents liés à l'impact de la STEP sur le Clain. Les résultats d'analyse de suspects sont ensuite explorés.
- La quatrième et dernière sous-partie compare les différentes listes de suspects et de métabolites suspects. Un bilan critique conclut cette sous-partie en comparant les listes de signaux pertinents des approches développées pour les sélectionner avec les différents échantillons.

Une conclusion générale clôt ce manuscrit, récapitulant les résultats obtenus et détaillant des perspectives.

Chapitre 1 : Bibliographie

I. Les micropolluants : origines et devenir dans le cycle technique de l'eau

I.1. Généralités

Les micropolluants organiques sont des molécules présentes dans l'environnement à de très faibles concentrations (i.e. de l'ordre du nanogramme par litre au microgramme par litre) et pouvant avoir une action toxique ou écotoxique pour les organismes présents dans les milieux contaminés. Ces micropolluants ont des propriétés chimiques différentes les uns des autres ainsi qu'une capacité à se dégrader variable, certains pouvant même se bioaccumuler. Essentiellement d'origine anthropique, les micropolluants appartiennent à différentes familles chimiques ou familles d'utilisation, avec par exemple :

- Produits pharmaceutiques
- Pesticides et Biocides
- Surfactants
- Produits d'hygiène et cosmétiques
- Retardateurs de flammes
- Plastifiants
- Composés perfluorés
- Hydrocarbures aromatiques polycycliques

Cette liste n'est pas exhaustive mais rappelle que ces molécules sont présentes dans de nombreux secteurs des activités humaines, de la production à l'utilisation, en milieu rural, urbain et industriel. Chacune de ces classes est composée de sous-familles comme par exemple les parfums, les conservateurs et les filtres UV pour les produits cosmétiques et d'hygiène.

Les paragraphes suivants s'attacheront à préciser les sources de contamination des eaux par différentes classes de micropolluants, leurs niveaux de concentrations et leur devenir à travers le petit cycle de l'eau.

I.2. Origines de la présence et niveau de contamination en micropolluants dans les eaux.

L'utilisation des ressources en eaux pour les activités humaines suit le petit cycle de l'eau également appelé cycle technique de l'eau dont les étapes sont représentées dans la Figure 1.

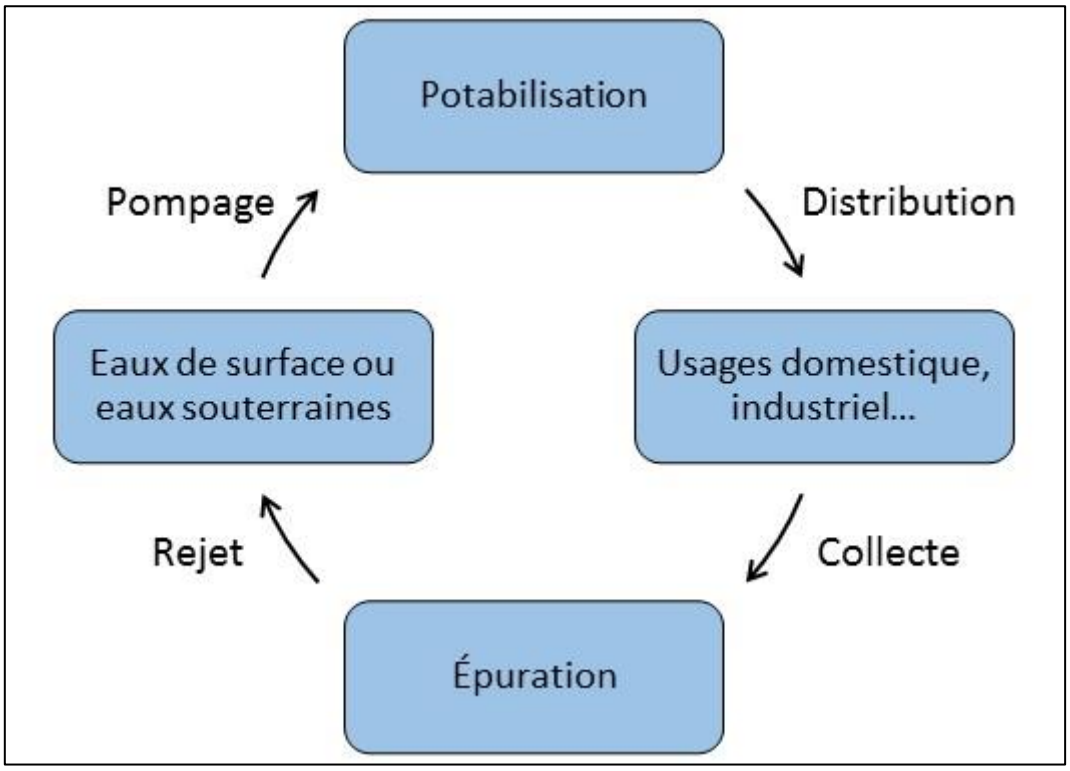


Figure 1 : Schéma simplifié du petit cycle de l'eau.

La plupart des étapes du petit cycle de l'eau sont susceptibles de conduire à la diffusion et à la dispersion de ces molécules de micropolluants dans le milieu naturel voire dans les réseaux de distribution si les traitements des différents usages (épuration et potabilisation) sont insuffisants. Les usages contaminent l'eau utilisée par diverses micropolluants (produits, contact avec les matériaux, procédés industriels…) qui sont collectés par la station d'épuration. Une partie des contaminants peuvent être traités mais d'autres peuvent être rejetés dans le milieu naturel. Les eaux souterraines et eaux de surfaces pouvant être contaminées par diverses sources vont être pompées pour la potabilisation. Les réactions d'oxydation lors de la désinfection avec les micropolluants présents dans la ressource peuvent former d'autres polluants susceptibles de circuler dans le réseau de distribution. Les ruptures accidentelles des canalisations de collecte et de distribution peuvent également participer à la dispersion des polluants.

Les eaux peuvent être polluées de deux manières :

- la pollution diffuse : pollution en continue avec des sources multiples généralement en petites doses et plus difficile à identifier. Ce type de pollution est observé suite à l'utilisation de pesticides, à l'épandage, à l'érosion des sols, aux activités domestiques et industrielles, ou encore aux traitements insuffisants des eaux usées,
- la pollution ponctuelle : pollution accidentelle, locale, parfois massive pouvant généralement être identifiée et survenant par exemple suite à des accidents industriels, des fuites, des déversements directs dans les eaux, des débordements des égouts.

De très nombreuses études se sont intéressées à la présence et au devenir de différentes molécules ou différentes classes de micropolluants dans les différents compartiments du petit cycle de l'eau.

Les paragraphes suivants s'attacheront à préciser les sources de contamination des eaux par différentes classes de micropolluants, leurs niveaux de concentrations et leur devenir à travers le petit cycle de l'eau.

I.2.1. Les vecteurs de pollution de l'eau en milieu anthropisé vers les STEP

Les micropolluants détectés dans les eaux de STEP sont issus de trois sources : les eaux domestiques ou urbaines, les eaux non-domestiques et les eaux pluviales, lorsque le réseau d'eaux usées est imparfaitement séparatif. Ces micropolluants proviennent de pollutions diffuses qui alimentent continuellement les STEP en micropolluants organiques.

I.2.1.1. Les eaux domestiques

L'utilisation de l'eau pour un usage domestique conduit à la présence de micropolluants dans les eaux ménagères et dans les eaux vannes.

Les eaux vannes sont les eaux collectées des toilettes qui entraînent les matières fécales et l'urine. Ces eaux peuvent être vectrices de produits pharmaceutiques et de leurs métabolites. À ce niveau, la présence de médicaments peut provenir soit :

- du rejet des résidus médicamenteux par les voies naturelles. Dans ces conditions, les médicaments absorbés par les humains sont métabolisés par leur organisme pour être excrétés par les selles ou les urines sous leur forme native ou sous une ou plusieurs formes métabolisées.
- de la négligence des utilisateurs qui les jettent dans les toilettes ou dans les éviers. On estime en effet que 12% des médicaments au sein d'une foyer sont jetés dans les toilettes ou les égouts et 66% avec les ordures ménagères (Académie de pharmacie, 2008).

En 2011, plus de 3 000 substances actives et 300 principes actifs à usage vétérinaire étaient disponibles (ANSES, 2011).

A côté des composés pharmaceutiques, d'autres composés peuvent également provenir des excrétions dans les eaux vannes, tels que les additifs alimentaires comme les édulcorants artificiels (acésulfame, aspartame, cyclamate, sucralose…). Ceux-ci ne sont pas conçus pour être métabolisés dans l'optique de ne pas apporter de calories. Par conséquent, la forme native de ces molécules est la plus susceptible d'être diffusée vers les stations d'épuration (Margo *et al.,* 2015).

Les eaux ménagères constituent le reste des eaux usées et sont issues notamment des machines à laver, lave-vaisselles, éviers et douches. Elles sont également potentiellement vectrices d'une grande diversité de micropolluants tels que des produits d'hygiène et cosmétiques, des surfactants et détergents présents par le lavage, des retardateurs de flammes provenant des textiles, des composés perfluorés, des substances anticorrosion et des plastifiants. De nombreux micropolluants dont des phtalates et des alkylphénols ont été détectés dans ces eaux avec des concentrations variant de la limite de quantification à une concentration médiane de 100 µg/L pour l'un des phtalates dans les eaux de machine à laver (Deshayes *et al.,* 2017).

I.2.1.2. Les eaux non-domestiques

Les eaux industrielles et les eaux usées provenant des hôpitaux arrivent dans les STEP. Comme les eaux domestiques elles peuvent donc constituer une source de micropolluants.

I.2.1.2.1. Les rejets hospitaliers

Les rejets d'hôpitaux contiennent en particulier des résidus médicamenteux, des biocides, des détergents et des tensio-actifs. Par exemple, le sulfametoxazole et la ciprofloxacine sont des antibiotiques fréquemment détectés dans ces eaux (Gros *et al.*, 2013 ; Lozano *et al.,* 2013 ; Paulus *et al.,* 2019). Les eaux usées d'hôpitaux seraient aussi d'importants émetteurs d'agents de contraste vers les STEP (Verlichi *et al.,* 2012). Selon un calcul effectué par Santos *et al.* (2013), sur quatre hôpitaux de différentes tailles au Portugal, la contribution des hôpitaux pour ces molécules pourrait aller jusqu'à 50% dans certains cas. Enfin, le projet multipartenaires SIPIBEL, lancé en 2010, a étudié la présence de 15 médicaments de plusieurs familles thérapeutiques. Dans ce travail, douze d'entre eux ont été quantifiés dans les rejets hospitaliers avec une fréquence de détection supérieure à 70% excepté pour une molécule (econazole) (rapport SIPIBEL 201-2015). Dans ce type de rejets, les gammes de concentrations des composés détectés sont généralement comprises entre quelques nanogrammes par litre et quelques dizaines de microgrammes par litre en fonction des composés et des sites d'étude (Rapport SIPIBEL 2011-2015 ; Verlichi *et al.,* 2012).

I.2.1.2.2. Les rejets industriels

Les activités industrielles sont une source de micropolluants dans les eaux usées par la fabrication des polluants eux-mêmes (médicaments, biocides…) avec les réactifs, les sous-produits, les additifs, les déchets et les substances produites. Les composés perfluorés, bien que présents dans les eaux urbaines, sont une source importante de rejets industriels (Alder *et al.,* 2015 ; Ahrens *et al.*, 2009 ; Sinclair et Kannan, 2006). Il en est de même pour des alkylphénols éthoxylates et leurs précurseurs ainsi que des phtalates, détectés dans des eaux usées industrielles avec des concentrations comprises

entre une centaine de nanogrammes par litre et 70 µg/L pour des nonyphénols éthoxylates (Bergé *et al.,* 2012 et 2013).

I.2.1.3. Les eaux de ruissellements urbains

Dans le cas des réseaux unitaires, les eaux de pluie en milieux urbains arrivent dans les STEP suite aux phénomènes de ruissellement sur un nombre varié de surfaces : routes, parkings, bâtiments, mobiliers urbains, véhicules, jardins…

Dès les années 1980, la présence et le devenir des HAP, PCB, polychlorodibenzo-dioxines ou -furanes et autres molécules de mêmes familles dans l'atmosphère ont été étudiés et leur transfert dans les eaux de pluies avec des concentrations mesurées comprises entre le nanogramme par litre et la dizaine de nanogrammes par litre a été rapporté (Granier et Chevreuil, 1997 ; Hart *et al.,* 1993). Les eaux de pluies sont également porteuses d'autres types de micropolluants tels que des nonylphénols et des nonyphénols éthoxylates avec des concentrations comprises entre 0,03 et 3,10 µg/L ou des phthalates avec des concentrations entre 0,01 et 39 µg/L (Bergé *et al.,* 2012, 2013). Parmi ces composés, les nonyphénols et nonylphénols éthoxylates ont aussi été mis en évidence dans des eaux de ruissellement avec des concentrations comprises entre 0,10 et 1,20 µg/L, avec comme origine principale les peintures, laques, huiles de moteurs et bâtiments (Björklund *et al.,* 2009). De même, en élargissant aux alkylphénols éthoxylates, des concentrations entre 0,15 et 190 µg/L ont été déterminées dans des eaux de ruissellement (Bressy *et al.,* 2011). Concernant les pesticides, plus d'une quinzaine de pesticides dont l'atrazine et deux de ses métabolites, le lindane, l'isoproturon, le metolachlore ou le metazachlore ont été détectés dans les eaux de pluie en Europe selon une étude menée par Dubus *et al.* (2000). De même, plus récemment en Chine (Yang *et al.,* 2012), divers pesticides organochlorés ont été quantifiés avec des concentrations allant du ng/L à la centaine de ng/L. Ces composés sont aussi susceptibles d'être lessivés par les eaux de pluie des jardins, espaces publiques et bâtiments (Burkhardt *et al.,* 2012 ; Coutu *et al.,* 2012). Enfin, parmi les autres classes de micropolluants, des phtalates (Bergé *et al.,* 2013 ; Clark *et al.,* 2003 ; Teil *et al.,* 2006), des produits pharmaceutiques, de la caféine, du bisphénol A, des édulcorants et des hormones ont été détectés dans les eaux de ruissellement avec des concentrations médianes allant de 2 ng/L à plus de 1 µg/L pour la caféine (Buran *et al.,* 2018 ; Müller *et al*., 2020) dont les fréquences de détection peuvent aller jusqu'à 100% dans les échantillons considérés.

Ainsi, la littérature existante montre que les eaux de ruissellement sont une source de micropolluants organiques de natures diverses responsables d'une pollution diffuse soit directement par la présence de ces contaminants dans les eaux de pluie soit par le phénomène de ruissellement. Les concentrations

rapportées s'échelonnent du nanogramme par litre à quelques centaines de microgrammes par litre. Les eaux de ruissellement s'avèrent donc être une source non négligeable de contaminants.

I.2.2. Les micropolluants dans les stations de traitement des eaux usées

De nombreux travaux se sont intéressés à l'étude de diverses familles de micropolluants dans les eaux de STEP urbaines tout au long du processus de traitement, pour différents types de procédés d'épuration à travers le monde. Il est difficile d'être exhaustif au regard de la grande variété de micropolluants et des multiples publications portant sur les affluents et les effluents de STEP. Certains types de micropolluants ont été bien plus analysés que d'autres. C'est le cas des produits pharmaceutiques, des pesticides, des surfactants et des produits d'hygiène et cosmétiques dont nombre d'entre eux ont été détectés à maintes reprises dans les eaux de stations d'épuration. D'autres familles de molécules ont fait l'objet d'analyse dans une moindre mesure, comme par exemple les phtalates, les retardateurs de flammes, les hormones ou encore les produits illicites.

Le Tableau 1 montre les gammes de concentrations classiquement retrouvées pour certains micropolluants. Ces concentrations vont du nanogramme par litre au microgramme par litre avec une valeur dépassant le milligramme par litre pour l'acésulfame. Dans ce tableau, les plus grandes concentrations observées concernent les produits pharmaceutiques, les parabènes, les édulcorants, les produits anticorrosion, les composés perfluorés, les phtalates, les retardateurs de flammes et les surfactants. Cependant, les autres familles présentées montrent des concentrations maximales supérieures au microgramme par litre dans les affluents et supérieures à la centaine de nanogrammes par litre dans les effluents. Pour l'ensemble des composés rapportés, il convient de souligner des disparités importantes en fonction de la famille des molécules ou même à l'intérieur de ces familles. Les produits pharmaceutiques sont un bon exemple de cette diversité de concentrations. En effet, la metformine, avec les antidouleurs, figure parmi les concentrations les plus élevées avec des mesures supérieures à 10 µg/L et jusqu'à plus de 130 µg/L pour le paracétamol tandis que d'autres produits ne dépassent pas les 150 ng/L en entrée de STEP (Petrie *et al.,* 2016). Ces différences de concentrations peuvent en partie être liées aux volumes et fréquences d'utilisation : la metformine est en effet très fréquemment utilisée dans le traitement du diabète de type 2 de même que les antidouleurs sont des molécules très usuelles. Les gammes de concentrations peuvent aussi varier, dans des proportions plus ou moins importantes, en fonction des STEP considérées, du pays ou du continent de l'étude en raison des répartitions et des consommations différentes de ces molécules entre les pays. Néanmoins les comparaisons effectuées par Loos *et al.* (2013) entre plusieurs travaux sur les effluents tendent à montrer que les valeurs moyennes et médianes sont assez comparables en Europe notamment pour les produits anticorrosion, les composés perfluorés, les retardateurs de flammes et une partie des composés pharmaceutiques.

Pour ces composés, les fréquences de détection, non représentées dans ce tableau, s'échelonnent de <0,5% à 100% en fonction des familles et des molécules que ce soit dans les affluents ou les effluents et en fonction des STEP étudiées. Les travaux d'Al Aukidy *et al.* (2012) illustrent ces fréquences de détection et ces variations sur les produits pharmaceutiques. Dans cette étude, parmi les quatre beta-bloquants analysées (metoprolol, atenolol, sotalol et timolol) trois d'entre eux ont une fréquence de détection de 100% dans les effluents de deux STEP différentes tandis que le timolol a une fréquence de détection de 33% dans une STEP et 100% dans l'autre (n=3). De la même manière, Stasinakis *et al,.* (2013) ont étudié huit benzotriazoles et benzothiazoles ainsi que 18 composés perfluorés dans les affluents de STEP. Dans ce travail, un benzothiazole et un benzotriazole ont été mesurés avec des fréquences de détection inférieures à 100% (64% et 14% respectivement). Trois composés perfluorés ont été analysés à chaque reprise (dans 14 échantillons) ; huit n'ont jamais été détectés et les sept derniers ont des fréquences comprises entre 21 et 64%. Ainsi, ces exemples montrent la variabilité inhérente à la présence des micropolluants.

La comparaison entre les affluents et les effluents indique une diminution en général des concentrations entre les deux matrices, démontrant un impact des STEP sur les micropolluants. Toutefois, il est important de noter que pour certaines familles de molécules les valeurs relevées dans les effluents sont supérieures aux valeurs dans les affluents. D'une part, cela pourrait en partie s'expliquer parce que les données concernant les effluents de STEP sont plus nombreuses dans la littérature que pour les affluents, notamment pour ce qui est des analyses à grande échelle comme celle de Loos *et al.* (2013) qui regroupe 90 STEP à travers l'Europe. D'autre part, le Tableau 1 regroupe les concentrations maximales observées mettant en évidence les écarts existants à l'intérieur des diverses familles de contaminants et donc la variété de concentrations existantes.

Le Tableau 1 témoigne donc de la contamination non négligeable des STEP par des micropolluants organiques, y compris dans les eaux traitées. La nature de ces polluants est très diverse et reflète la multitude des molécules dispersées dans les eaux parvenant jusqu'aux STEP.

Tableau 1 : Gammes de concentrations minimales et maximales détectées pour plusieurs familles de micropolluants organiques dans les affluents et les effluents de STEP.

		Affluents de STEP	Effluents de STEP	Références
		Concentrations en ng/L		
Produits pharmaceutiques	**Antibiotiques, Beta-bloquants, Analgésiques, Agents de contrastes …**	7,6-500 000	2 – 150 000	1-2-7-18-19-21
Produits d'hygiène et cosmétiques	**Filtres UV**	51 - 6084	3,1 - 6325	1-2-8-20
	Muscs	10 - 6500	75 - 1400	1-9
	Parabènes	38,3 – 30 688	5,8-36	1-2-20
Produits domestiques ou industriels	**Anticorrosion**	18 - 15 971	14 – 221 000	1-10-11-21
	Phtalates	60 - 122 000	0,083 – 69 000	1-4
	Retardateurs de flammes	400 - 6600	0,6 – 43 000	1-12-21
	Composés perfluorés	2,6 - 513	0,07 – 23 800	1-11-13-21
	Edulcorants (Acésulfame, etc)	50 – 93 200	500 – 2 500 000	1-5-6-21
Produits illicites	**Amphétamine, Cocaïne, Héroïne, MDMA …**	10 - 3701	2 - 389	1-2-14
Hormones	**Hormones naturelles ou synthétiques**	0 - 44,5	0,1 - 217	1-2-15
Pesticides	**Herbicides, Fongicides, Insecticides**	<1 - 2526	0,30 – 15 800	1-16-17-21
Surfactants	**Surfactants ioniques et non ioniques**	80 - 430 000	240 – 150 000	1 - 3

Références : 1 : Margot *et al.,* 2015 ; 2 : Petrie *et al.,* 2016 ; 3 : Bergé *et al.,* 2012 ; 4 : Bergé *et al.,* 2014 ; 5 : Baena-Nohueras *et al.,* 2018 ; 6 : Lange *et al.,* 2012 ; 7 : Al Aukidy *et al.,* 2012 ; 8 : Balmer *et al.,* 2005 ; 9 : Clara *et al.,* 2010 ; 10 : Asimakopoulos *et al.,* 2013 ; 11 : Stasinakis *et al.,* 2013 ; 12 : O'Brien *et al.,* 2015 : 13 : Arhens *et al.*, 2009 ; 14 : Bijlsma *et al.,* 2012 ; 15 : Comtois-Marotte *et al.,* 2016 ; 16 : Campo *et al.,* 2013 ; 17 : Singer *et al.,* 2010 ; 18 : Guerra *et al.,* 2014 ; 19 : Gros *et al.,* 2013 ; 20 : Kasprzyk-hordern *et al*., 2009 ; 21 : Loos *et al.,* 2013

I.2.3. Voies d'élimination des micropolluants organiques dans les STEP conventionnelles

Les STEP n'ont, à l'origine, pas été conçues dans le but d'éliminer une pollution par les micropolluants. Néanmoins, les procédés de traitement dans les STEP agissent sur une partie des micropolluants par des réactions de dégradation chimique ou biologique. Les principaux mécanismes d'élimination dans les STEP conventionnelles, c'est-à-dire à boues activées, sont : la volatilisation, les réactions abiotiques,

la biotransformation par des micro-organismes, et l'adsorption sur les matières en suspension ou les boues.

La volatilisation s'opère à la surface des eaux et dépend de la volatilité du contaminant et des conditions de traitement : agitation, température, aération ou encore de la pression atmosphérique (Pomiès *et al.,* 2013). Les réactions abiotiques sont l'ensemble des réactions opérant dans les STEP n'impliquant pas les microorganismes telles que la photolyse par l'exposition aux rayons UV ou les réactions d'hydrolyse. La photolyse reste cependant négligeable, principalement à cause de la turbidité et de la surface d'exposition. Les biotransformations dans les STEP sont dues à la présence de microorganismes dont le développement est favorisé pour permettre l'abattement de la matière organique. Ces mécanismes constitueraient l'une des principales voies d'élimination, particulièrement pour les molécules hydrophiles (Margot *et al.,* 2015). Les phénomènes de biotransformation des micropolluants peuvent se produire de deux manières : soit directement par les microorganismes pour leur développement ou leur survie, soit indirectement suite à des réactions secondaires par des enzymes non spécifiques ou des cofacteurs produits par les microorganismes (Benner *et al.,* 2013 ; Fischer *et al.,* 2014). Enfin, l'adsorption est l'autre mécanisme d'élimination prépondérant, avant tout pour les molécules hydrophobes ou chargées positivement. Les interactions mises en œuvre (électrostatiques et hydrophobiques) dans ce mécanisme impliquent les particules en suspension et certaines parties de microorganismes (Joss *et al.,* 2006). Toutefois, l'adsorption des micropolluants peut aussi s'effectuer sur la matière organique colloïdale en suspension. Dans ce cas, les micropolluants organiques ne sont pas éliminés avec les boues mais restent dans les eaux (Barret *et al.,* 2010 ; Katsoyiannis *et al.,* 2007).

I.2.4. Abattement des micropolluants dans les STEP

Les pourcentages d'élimination des micropolluants sont très variables et sont « molécules-dépendants ». Ils varient également selon la filière de traitement mise en place. Dans le cas d'une STEP conventionnelle à boues activées, les valeurs trouvées dans la littérature vont de quelques pourcents à plus de 90%. Un nombre non négligeable de micropolluants a une diminution de concentration inférieure à 50%. En effet, parmi les 159 micropolluants organiques considérés par Margot *et al.* (2015), 67 d'entre eux sont abattus à hauteur de plus de 80%, 30 autres entre 50 et 80% et enfin 62 ont un pourcentage d'abattement inférieur à 50% soit plus d'un tiers des polluants examinés. Dans cette étude, il ressort également que les surfactants, les hormones, les hydrocarbures aromatiques polycycliques et les produits d'hygiène et cosmétiques font parties des familles de contaminants qui ont le plus de composés éliminés à plus de 85% tandis que les produits pharmaceutiques et les produits utilisés pour un usage domestique présentent des pourcentages d'élimination beaucoup plus variables. Certains retardateurs de flammes phosphorés sont par exemple abattus à plus de 60%

(triphényl phosphate, tris(butoxyéthyl) alors que d'autres composés de la même classe sont éliminés à moins de 5% (tris(2-chloroéthyl)phosphate, tris[2-chloro-1-(chlorométhyl)éthyl]phosphate). Similairement, l'irbersartan, un antihypertenseur, est éliminé à hauteur de 10% tandis que le valsartan, un autre antihypertenseur, est abattu à plus de 50%. Ces résultats sont en accord avec d'autres données montrant que les surfactants, les alkylphénols, les filtres UV et les parabènes sont en général bien éliminés par les STEP conventionnelles mais qu'une variabilité importante est toujours constatée pour les produits pharmaceutiques (Balmer *et al.,* 2005 ; Kasprzyk-Hordern *et al.,* 2009 ; Lara-Martin *et al.,* 2011 ; Petrie *et al.,* 2016).

Ainsi, l'abattement des micropolluants organiques dans les STEP conventionnelles n'est généralement pas complet et dépend de la famille ou de la molécule considérée. Si certains contaminants ont un pourcentage de diminution supérieur à 80% d'autres ne voient leur concentration diminuer que de 10% ou moins. Ces données d'abattement sont à mettre en relation avec les données de concentrations du paragraphe I.2.2. C'est en effet en raison de ces éliminations incomplètes que des micropolluants organiques sont détectés dans les effluents de sorties des STEP.

I.2.5. Contamination des eaux de rivières par les micropolluants

I.2.5.1. Sources de contamination des rivières

La contamination des eaux de rivières provient de plusieurs sources :

- Les rejets des eaux traitées par les stations d'épuration urbaines et industrielles contenant des résidus de micropolluants, comme montré par les concentrations relevées dans le tableau précédent. Ces rejets sont le vecteur principal de contaminants dans les milieux aquatiques pour certaines familles de molécules : c'est le cas des produits pharmaceutiques, des produits cosmétiques et d'hygiène personnelle, des composés perfluorés et des composés anticorrosion entre autres (Kasprzyk-Horden *et al.,* 2008 ; Loss *et al.,* 2013 ; Reemstma *et al.,* 2009).
- Les eaux de ruissellement urbain atteignant les eaux de surfaces par écoulement ou par rejets directs en l'absence de réseau unitaire.
- Le phénomène de lessivage notamment dans le cas des pesticides dans les zones agricoles.
- Les débordements ou fuites en provenance des égouts (Clara *et al.*, 2004 ; Nakada *et al.,* 2008 ; Palmer *et al.,* 2008 ; Phillips and Chalmers, 2009 ; Kuroda *et al.,* 2012 ; Lee *et al.,* 2012 ; Tran *et al.,* 2019 ; Al Aukidy and Verlicchi, 2017).

La Figure 2 schématise ces différentes sources.

Figure 2 : Représentation des sources de contamination des eaux de rivières par les micropolluants. Source reprise et modifiée : www.monreseaudeau.fr

I.2.5.2. Niveaux de contamination des eaux de surface

Similairement aux affluents et effluents de STEP, les eaux de rivières ont fait l'objet de nombreuses analyses couvrant une grande variété de micropolluants. Le Tableau 2 répertorie des gammes de concentrations relevées pour plusieurs familles de micropolluants organiques.

Les concentrations relevées s'échelonnent du nanogramme par litre à plusieurs centaines de microgrammes par litre. Les concentrations les plus élevées concernent les surfactants, les phtalates, les édulcorants et les produits pharmaceutiques. De façon comparable aux affluents et aux effluents de STEP, on constate que selon les contaminants considérés, les gammes de concentrations sont peu à très étendues. Les concentrations détectées sont moins importantes que dans les effluents de STEP, les rejets étant dilués dans les eaux de surface.

Les eaux de surface sont donc contaminées par une vaste gamme de micropolluants organiques à des concentrations pouvant être non négligeables (Tableau 2).

Tableau 2 : Gammes de concentrations détectées plusieurs familles de micropolluants organiques dans des eaux de surface en Europe.

		Eau de surface Concentration en ng/L	Références
Produits pharmaceutiques	**Antibiotiques, Beta-bloquants, Analgésiques, Agents de contrastes …**	<0,3-5970	1-4-8-9-10-11-12
Produits d'hygiène et cosmétiques	**Filtres UV**	3,1 - 345	1-7-11
	Muscs	40 -190	3
	Parabènes	0,3-305	1-11
Produits domestiques ou industriels	**Anticorrosion**	<50-9 230	18
	Phtalates	10-31 200	6-15
	Retardateurs de flammes	10-1 000	19
	Composés perfluorés	7,6-26,4	2
	Edulcorants	0,2-12 700	8-13
Produits illicites	**Amphétamine, Cocaïne, Héroïne, MDMA …**	<0,3-92	1-11
Hormones	**Hormones naturelles ou synthétiques**	6 - 35	4
Pesticides	**Herbicides, Fongicides, …**	4,0-17 500	16-17
Surfactants	**Surfactants ioniques et non ioniques**	2– 490 000	5-14

Références : 1 : Petrie *et al.,* 2016 ; 2 : Arhens *et al.,* 2009 ; 3 : Clara *et al.*, 2011 ; 4 : Pailler *et al.,* 2009 ; 5 : Bergé *et al.,* 2012 ; 6 : Bergé *et al.,* 2014 ; 7 : Balmer *et al.,* 2005 ; 8 : Comtois-Marotte *et al.,* 2016 ; 9 : Ens *et al.,* 2014 ; 10 : Gracia-Lor *et al.,* 2014 ; 11 : Kasprzyk-Hordern *et al.,* 2009 ; 12 : Gros *et al.,* 2012 ; 13 : Lange *et al.,* 2012 ; 14 : Jardak *et al.*, 2015 ; 15 : Sanchez-Avila *et al.,* 2011 ; 16 : Singer *et al.,* 2010 ; 17 : Derbalah *et al.,* 2018 ; 18 : Reemtsma *et al.,* 2010 ; 19 : Reemtsma *et al.,* 2008

I.3. Impact de la présence des micropolluants organiques dans le petit cycle de l'eau

Les micropolluants organiques peuvent avoir un impact à la fois sur l'Homme et sur les milieux. Les produits de transformations des micropolluants sont également à considérer dans les mesures de toxicité et d'écotoxicité non seulement parce qu'ils coexistent avec les molécules parentes parfois à des concentrations voisines mais aussi parce que certains peuvent être aussi voire plus toxiques que

leur précurseur (Michael *et al.,* 2014). On considère actuellement deux types de toxicité : la toxicité aigüe, qui mesure un effet immédiat, et la toxicité chronique qui mesure un effet dans le temps. Si la toxicité aigüe est relativement limitée pour les micropolluants à travers les milieux aquatiques, la toxicité chronique est en revanche sujette à préoccupation en raison de leur présence régulière à l'état de traces dans les milieux.

I.3.1. Impacts sur l'Homme

L'Homme est exposé aux micropolluants par plusieurs voies : cutanée, inhalée et orale *via* l'alimentation. L'eau est donc une voie potentielle d'exposition. Certains micropolluants ont vu leurs effets testés *in vivo* ou *in vitro* et se révèlent être des sources de danger dont les principaux recensés sont les effets cancérigènes, tératogènes et mutagènes. C'est par exemple le cas de retardateurs de flammes organophosphorés considérés comme reprotoxiques et ayant un effet cancérigène (Hou *et al.,* 2016 ; Van der Ween and de Boer *et al.,* 2012). L'exposition à certains pesticides a été par ailleurs liée à des maladies neurologiques telles que la maladie d'Alzheimer ou la maladie de Parkinson. Bien que les micropolluants présentent des dangers, l'exposition, est le facteur à prendre en compte pour estimer les risques et donc les impacts des micropolluants organiques sur la santé humaine. Les agriculteurs sont par exemple bien plus exposés aux pesticides que le reste de la population, exposée à de plus faibles doses. Dans le cas de micropolluants, les risques encourus par l'exposition au sein du cycle technique de l'eau sont mal connus et restent difficiles à évaluer du fait des faibles doses d'exposition et des multiples molécules impliquées susceptibles de générer des effets de synergie (effet cocktail). Néanmoins des limites réglementaires ont été fixées pour les eaux de distribution telles que les teneurs en pesticides, et en métabolites jugés pertinents, qui ne doivent pas excéder 0,1 µg/L par molécule identifiée. Un autre impact possible des micropolluants est le caractère perturbateur endocrinien. En effet, certains contaminants parmi les plastifiants, les hormones synthétiques ou naturelles, certains surfactants, certains pesticides, les PBDE ou encore les PCB sont suspectés ou sont classés comme perturbateurs endocriniens (Esperenza *et al.,* 2007). Plusieurs alkylphénols (nonyphénol, dérivés et isomères) sont ainsi inscrits comme substances préoccupantes en raison de leur effet possible sur le système hormonal (règlement REACH, ECHA). Enfin, le mésusage des antibiotiques et la contamination induite des ressources en eaux par ces produits pharmaceutiques ont pour conséquence la génération de phénomènes d'antibiorésistance. L'organisation mondiale de la santé a classé l'antibiorésistance comme une menace très sérieuse pour la santé mondiale. La présence de bactéries résistantes a été montrée dans des affluents et effluents de stations d'épuration et dans des effluents d'hôpitaux (Paulus *et al.,* 2019 ; Thompson *et al.,* 2013). Les pénicillines, les sulfonamides, les tétracyclines et les macrolides sont parmi les molécules les plus susceptibles de provoquer de l'antibiorésistance (Fisher *et al.,* 2012). À ces effets individuels s'ajoute la problématique

des effets cocktails pouvant exister sous la forme d'effets synergétiques, additionels ou encore antagonistes entre les micropolluants (Laetz *et al*., 2009). Ces effets cocktails font l'objet de travaux de modélisation pour tenter de prédire la toxicité de ces mélanges (D'Almeida *et al*., 2020). De part le nombre de combinaisons possibles, ces effets restent toutefois difficiles à évaluer.

I.3.2. Impacts sur les milieux

Les micropolluants organiques ont un impact « molécule-dépendant » sur la faune et la flore des milieux récepteurs. À ce niveau, une problématique importante des micropolluants est leur persistance dans les différents compartiments des milieux (aquatiques et sols). Les propriétés de biodégradabilité et les propriétés hydrophobes des molécules sont des facteurs jouant sur la bioaccumulation dans les organismes aquatiques et sur l'adsorption dans les sédiments. Ainsi, certains micropolluants peu ou pas dégradés dans l'environnement peuvent se stocker dans ces compartiments. La convention de Stockholm vise à l'élimination et à la restriction de contaminants spécifiques dans le monde. Elle a tout d'abord listé une première liste de polluants (Aldrine, Chlordane, Dieldrine, Endrine, Heptachlore, Hexachlorobenzène, Mirex, Toxaphène et les Polychloro-biphényles) puis a rajouté d'autres polluants (Chlordecone, Decabromodiphenyl ether, Hexabromobiphenyl, Hexabromocyclododecane, Hexabromodiphenyl ether et heptabromodiphenyl ether, Hexachlorobutadiene, Alpha hexachlorocyclohexane, Betahexachlorocyclohexane, Pentachlorobenzene, Pentachlorophenol avec ses sels et esters, Polychlorinated naphthalenes, paraffines chlorés à chaines courtes, Endosulfan et ses isomères, Tetrabromodiphenyl ether et pentabromodiphenyl ether). Pour leur majorité ces contaminants sont des substances bromées ou chlorées et appartiennent à la famille des pesticides ou des produits industriels. Toutefois, d'autres polluants non présents dans cette liste persistent dans l'environnement tels que les surfactants, d'autres retardateurs de flammes bromés et organophosphorés, des produits pharmaceutiques tels que la carbamazépine, ou encore des pesticides tels que l'atrazine et l'alachlore.

Il a été montré que les perturbateurs endocriniens ont un impact sur la faune, notamment aquatique. Kidd *et al.* (2007) ont par exemple montré l'effondrement d'une population de poissons après l'exposition à une hormone synthétique. De même, la « féminisation » des poissons par l'exposition aux perturbateurs endocriniens a mis en lumière ce problème auprès du grand public (Itzel *et al.,* 2018). Enfin, des mesures effectuées au niveau des effluents de station d'épuration par des tests de toxicité ont montré une perturbation résiduelle susceptible d'avoir des effets néfastes (rapport SIPIBEL 2011-2015) démontrant la nécessité d'étudier les conséquences de la présence de ces contaminants. Pour certains composés, des valeurs de concentrations prédites sans effet écotoxicologiques sont disponibles dans la littérature. Mais ces valeurs ne concernent qu'un nombre limité de micropolluants au regard de la diversité des contaminants et de leurs produits de transformations.

I.4. Conclusions

Les micropolluants organiques sont des molécules ubiquitaires ayant comme sources les différents compartiments d'activités humaines : industries, usages domestiques, hôpitaux, agriculture, paysage urbain… Les micropolluants sont présents dans l'ensemble du petit cycle de l'eau avec des concentrations relevées pouvant atteindre plusieurs centaines de microgrammes par litre. Les affluents de STEP sont les eaux les plus chargées en contaminants mais les effluents n'en sont pas pour autant exempts, certains présentant même des niveaux de concentrations du même ordre de grandeur que dans les affluents. Les eaux de surface présentent des niveaux de contamination allant du nanogramme par litre à quelques centaines de microgrammes par litre pour les surfactants. La pollution induite par les micropolluants organiques n'est pas simplement ponctuelle mais plutôt généralement continue, certains polluants étant détectés lors de chaque prélèvement. La dispersion de contaminants dans les milieux aquatiques peut être en lien avec une élimination incomplète par les STEP et varie fortement en fonction du polluant étudié. Cette dissémination implique de potentiels effets négatifs sur les milieux et l'Homme en raison du caractère mutagène, cancérigène, reprotoxique ou perturbateur endocrinien de certains d'entre eux mais également en raison de leur persistance, leur bioaccumulation ou l'antibiorésistance induite.

II. Approches d'analyse des micropolluants dans les différentes matrices d'eaux

II.1. Analyse ciblée

II.1.1. Principes de l'analyse ciblée

L'analyse ciblée a pour but principal l'identification et/ou la quantification de substances répertoriées et donc l'obtention de concentrations fiables, dans une matrice donnée. Ainsi, les résultats d'une analyse ciblée sont qualitatifs (présence/absence i.e. < LD) et/ou quantitatifs. Les objectifs principaux de l'analyse ciblée sont la présence et le suivi de molécules par exemple dans un milieu donné ou tout au long d'un processus. L'appellation « monitoring » peut parfois être utilisée lors de la réalisation de ce type d'analyse.

L'analyse ciblée implique le développement de méthodes sensibles et spécifiques comportant plusieurs étapes : le prélèvement, la préparation des échantillons, la méthode d'analyse et enfin le traitement des résultats. Chacune de ces étapes doit être optimisée spécifiquement pour les composés à doser. Les méthodes développées, du prélèvement à l'analyse, doivent aussi répondre à des critères qualités généralement répertoriés selon des normes ou des bonnes pratiques. C'est notamment le cas des méthodes utilisées par les laboratoires dédiés à l'analyse.

Le développement de méthodes d'analyse ciblée s'appuie sur le choix d'une liste de composés définie en amont. Cette liste peut être orientée vers un nombre limité de composés ou encore vers une ou plusieurs familles de molécules différentes. Lorsque les composés chimiques déterminés présentent des propriétés très diverses, on parle alors de méthodes « multi-résidus ». Ce type de méthodes regroupe généralement plusieurs approches, selon un rapport du programme AQUA-REF de 2012. Ainsi, peuvent être utilisées dans ces conditions :

- plusieurs méthodes de préparations d'échantillons et plusieurs méthodes de détections en parallèle sur un même échantillon,
- plusieurs méthodes de préparations d'échantillons mais une seule méthode de détection,
- un seul procédé de préparation et plusieurs méthodes instrumentales pour la détection,
- une seule méthode de préparation et une seule méthode d'analyse instrumentale.

Pour confirmer l'identité des composés analysés voire les quantifier, l'analyse ciblée nécessite l'utilisation d'étalons de référence pour développer les méthodes.

II.1.2. Méthodes de préparations et méthodes instrumentales utilisées lors de l'analyse ciblée de micropolluants dans les eaux

II.1.2.1. Méthodes de préparations d'échantillons

La littérature sur l'analyse des micropolluants dans l'eau est aujourd'hui abondante et concerne un nombre important de molécules diverses dans les différents compartiments du cycle naturel et du cycle technique de l'eau. Concernant la préparation des échantillons, il est impossible d'être complétement exhaustif sur l'ensemble des méthodes utilisées. Néanmoins, de nombreuses publications (Baker *et al.*, 2011 ; Buchberger *et al.*, 2011 ; Xu *et al.*, 2013 ; Perez-Fernandez *et al.*, 2017 ; Salgueiro-Gonzalez *et al.*, 2017) ont passé en revue les principales méthodes utilisées au cours de ces dernières années, que ce soit pour des familles de molécules précises ou pour des méthodes multi-résidus. Dans les méthodes multi-résidus, lorsqu'une unique méthode de préparation d'échantillons (ou méthode d'extraction) est utilisée, elle doit pouvoir couvrir la large gamme de propriétés inhérente à l'analyse des composés. La revue détaillée de Perez-Fernandez *et al.* (2017) indique deux tendances prédominantes : les analyses de type multi-résidus portant sur un grand nombre de composés et les méthodes de « dilute and shoot » adaptées pour des molécules très polaires. Dans les deux cas, l'extraction sur phase solide (SPE) sur cartouche ou sur disque est la technique de préparation d'échantillons la plus usitée pour un grand nombre de micropolluants (Buchberger *et al.,* 2011 ; Pérez-Fernandez *et al.,* 2017) et les phases de type C18 ou polystyrène-divinylbenzène (PS-DVB) sont préférentiellement utilisées pour les composés peu polaires. Dans le cas des matrices environnementales, les phases combinant une partie hydrophobe et une partie échangeuse d'ions se

sont généralisées. Les phases dites « mixed-mode » ayant une balance hydrophile-hydrophobe avec une structure polymérique sont aujourd'hui les phases les plus répandues, notamment la cartouche Oasis HLB de chez Waters qui utilise un copolymère de divinylbenzene et vinylpyrrolidone. Les mixed-mode du même fabricant, Oasis MAX et WAX, Oasis MCX et WCX, sont également très répandus (Buchberger *et al.,* 2011), notamment en vue de l'analyse des produits pharmaceutiques ainsi que des produits d'hygiène et cosmétiques. La phase Strata-X du fournisseur Phenomenex est une alternative aux phases Oasis. Enfin, des phases solides dédiées à l'analyse de contaminants spécifiques existent sous la forme d'impression moléculaire en polymère (MIP). Ces adsorbants sont dédiés à la reconnaissance de molécules cibles ou de groupements cibles, pouvant apporter une amélioration de l'extraction des analytes cibles. Bien que la SPE sur cartouche ou disques soit la méthode usuelle, d'autres alternatives telles que l'extraction sur phase solide dispersive (d-SPE) et l'extraction sur phase solide magnétique (MSPE) sont également parfois utilisées pour la préparation des échantillons. La d-SPE utilise un adsorbant dispersé en solution et la MSPE emploie également un adsorbant dispersible mais ayant des propriétés magnétiques. Les principaux avantages de ces deux méthodes sont la réutilisation de l'adsorbant et l'absence de phénomène de colmatage pouvant être observé sur les cartouches. La d-SPE a servi notamment pour l'extraction de pesticides, de phtalates et d'antibiotiques. Enfin, si la SPE et ses dérivés sont les techniques couramment utilisées, l'extraction liquide-liquide, la micro-extraction sur phase solide SPME et l'extraction par sorption sur barreau magnétique (SBSE) sont d'autres méthodes existantes (Buchberger *et al.,* 2011) (la SPME et la SBSE servant plutôt pour les molécules hydrophobes avant une analyse par chromatographie gazeuse).

II.1.2.2. Méthodes instrumentales pour l'analyse ciblée

Les méthodes chromatographiques sont les méthodes privilégiées pour la séparation des composés en analyse ciblée et la spectrométrie de masse, particulièrement la spectrométrie de masse en tandem, est la méthode de détection très majoritaire pour la détection des micropolluants en raison de ses capacités de spécificité et de sensibilité. Le choix de la chromatographie gazeuse ou liquide (GC ou LC) dépend essentiellement de la polarité mais surtout de la volatilité des analytes considérés.

La chromatographie liquide couplée à la spectrométrie de masse (LC-MS) est la méthode la plus répandue pour l'analyse de produits pharmaceutiques et de produits d'hygiène ou cosmétiques (Buchberger *et al.,* 2011 ; Fischer *et al.,* 2012). Ce couplage est également utilisé pour les retardateurs de flammes, les surfactants, les produits anticorrosion, les composés perfluorés, leurs produits de transformations, ou encore pour les méthodes multi-résidus (Fischer *et al.,* 2012 ; Pérez-Fernandez *et al.,* 2017). Au niveau du spectromètre de masse, les sources d'ionisation à pression atmosphérique sont très majoritairement employées et la source électrospray (ESI) est la référence dans le cas des composés polaires (Pérez-Fernandez *et al.,* 2017). Cependant d'autres sources à pression

atmosphérique comme la source d'ionisation chimique (APCI) peuvent présenter un intérêt dans le cas des contaminants modérément polaires et est moins soumise aux effets matrices qui constituent le problème majeur rencontré lors de l'utilisation de la source ESI. De même, d'autres sources dites « ambiantes » qui ne s'appuient sur aucune étape préliminaire peuvent être employées. C'est le cas notamment des sources permettant l'ionisation électrospray par désorption (DESI) et l'analyse directe en temps réel (DART). L'interface DART a par exemple été utilisée dans des méthodes semi-quantitatives pour des composés tels que les filtres UV (Haunschmidt *et al.* 2010).

Le couplage chromatographie gazeuse – spectrométrie de masse (GC-MS) est essentiellement utilisé pour les polluants lipophiles et volatiles (Pérez-Fernandez *et al.,* 2017). Les retardateurs de flammes, les produits anticorrosion, les PCB, les dioxines et les HAP sont les principales molécules analysées par GC-MS (Fischer *et al.,* 2012). La GCxGC soit la GC en deux dimensions est une intéressante alternative à la GC simple, notamment dans le cas des analyses multi-résidus (Fischer *et al.,* 2012).

II.1.3. Limites de l'analyse ciblée

L'indisponibilité d'étalons de référence, par absence de fournisseurs ou prix rédhibitoires, est un facteur limitant pour le choix d'analytes cibles et pour confirmer l'identification. Or, la limitation principale de l'analyse ciblée se trouve dans le choix au préalable des analytes cibles pour l'élaboration des méthodes d'analyses. Ces sélections écartent des molécules pouvant être pertinentes au regard des questionnements posés : toxicité, écotoxicité, dispersion de la pollution et devenir des contaminants.

La pertinence des molécules choisies est soumise aux biais de sélection, aux contraintes matérielles et au manque de connaissances sur les produits de transformations. En effet, la sélection des polluants peut se faire à partir soit des listes de molécules prioritaires et déjà réglementées par la législation européenne ou française, soit des molécules préoccupantes qui ne sont pas encore entrées dans le champ réglementaire ou encore soit des substances émergentes issues de la littérature scientifique existante. Le nombre croissant de molécules organiques inscrites sur le Chemical Abstract Service (CAS), qui répertorie désormais plus de 100 millions de substances et prises en compte par le règlement REACH qui en comporte plus de 140 000 produites ou importées en Europe témoigne de l'immensité du nombre de molécules pouvant être potentiellement des contaminants d'intérêt. Et, rien qu'au sein des foyers, Schwarzenback *et al.* (2006) ont estimé qu'entre 30 000 et 70 000 molécules sont utilisées avec entre autres des produits pharmaceutiques, des biocides ou encore des surfactants. Les contaminants réglementés ne représentent qu'une petite partie de ces substances. Par exemple, la directive cadre sur l'eau et ses annexes ne portent aujourd'hui que sur 53 substances prioritaires. De plus, à côté de ces substances connues, certains contaminants n'ont aujourd'hui pas encore été

identifiés tels que de nombreux produits de transformations générés au cours du traitement des eaux usées par exemple (Daughton *et al.,* 2004, European Commission, 2013). Seule une infime fraction de ces produits de transformation est en effet aujourd'hui connue. Ces produits ainsi que les contaminants émergents connus non réglementés constituent la « partie immergée de l'Iceberg ».

Lors d'une analyse ciblée, seuls les contaminants connus peuvent être recherchés. Or, les inquiétudes croissantes liées à la présence des micropolluants et à leurs potentiels effets délétères sur la santé humaine et sur les écosystèmes, couplées à la dynamique de la recherche sur l'analyse des micropolluants dans les eaux suscitent aujourd'hui de plus en plus le besoin d'explorer et d'identifier de nouveaux polluants.

Des progrès technologiques de la spectrométrie de masse ont donc conduit à l'amélioration et l'émergence de nouveaux analyseurs permettant des analyses à haute résolution, adaptés à cette volonté d'identification et participant au développement de nouvelles approches.

II.2. La spectrométrie de masse haute résolution

Les spectromètres de masse possèdent plusieurs caractéristiques qui définissent leur potentiel à résoudre des problèmes analytiques spécifiques : (i) la résolution, (ii) la précision en masse, (iii) la vitesse d'analyse ou d'acquisition, (iv) la gamme de masse et (v) la sensibilité. (i) Le pouvoir de résolution d'un appareil (R) est défini par sa capacité à séparer deux pics voisins. En pratique,

$$R = M/\Delta M \quad (1)$$

avec M, la masse de l'ion et ΔM, la différence de masse entre deux ions pouvant être déterminée par deux méthodes : la première consiste à considérer la différence de masse entre deux pics adjacents bien résolus (i.e. hauteur de la vallée entre les deux pics < 10 % de la hauteur) tandis que la seconde repose sur la largeur à mi-hauteur d'un seul pic (Hernandez *et al*., 2012). Liée à la résolution, (ii) la précision sur la masse est aussi un critère important dans la détermination de formules brutes : plus la précision en masse est élevée plus le nombre de formules brutes possibles est faible. (iii) La vitesse d'acquisition est le nombre de spectres par seconde que l'analyseur peut réaliser. Ce paramètre est important en cas de couplage avec des méthodes séparatives pour l'analyse de mélanges complexes car une vitesse trop faible impactera leur qualité. (iv) La gamme de masse correspond à l'écart entre la plus grande masse sur charge et la plus petite masse sur charge. Enfin, (v) la sensibilité repose sur la plus petite gamme de concentration que l'appareil est théoriquement capable d'atteindre. Ces cinq paramètres ont évolué depuis les débuts de la spectrométrie de masse. Ils ne sont pas uniformes et dépendent du spectromètre de masse choisi.

La spectrométrie de masse de type triple-quadripôle, est un outil puissant pour l'analyse ciblée grâce à sa sensibilité et sa sélectivité. Néanmoins, les inconvénients majeurs de ces instruments restent leur capacité de résolution, leur précision sur la masse et leur sensibilité réduite lorsque tous les ions sont suivis (i.e. analyses réalisées en mode appelé « Full Scan ») ce qui limite la capacité d'identification. Ainsi, les spectromètres de masse à haute résolution sont généralement préférés pour la recherche et l'identification de nouvelles molécules (Krauss *et al.,* 2010 ; Pérez-Fernandez *et al.,* 2017). En effet, plus la résolution est importante, plus le pic est fin et plus il est possible de séparer des pics proches. Ainsi, comme illustré dans la Figure 3, un pic unique détecté à une résolution inférieure à 5000 peut être décomposé en un massif constitué de plusieurs pics lors de l'augmentation de la résolution.

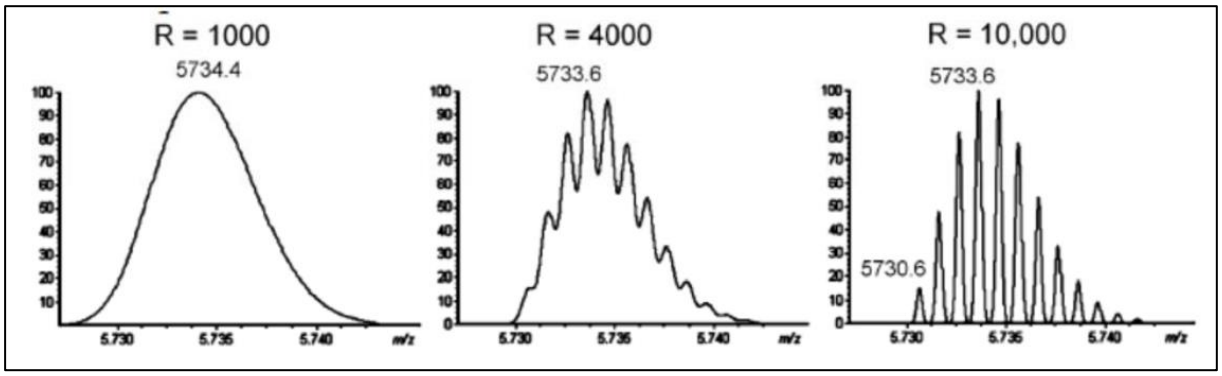


Figure 3 : Illustration du pouvoir de résolution par l'exemple de l'insuline bovine issu de Gross *et al.*, 2011.

Cette aptitude à séparer les ions, dont le rapport masse sur charge (m/z) est très proche, est un avantage pour tenter d'identifier un composé inconnu ou mettre en relief des molécules distinctes. Ces deux paramètres (résolution et précision en masse) sont deux points forts essentiels des appareils de spectrométrie de masse haute résolution, qui concernent aujourd'hui plusieurs types d'analyseurs.

Historiquement, l'augmentation du pouvoir de résolution date des années 50-60 avec le développement des instruments à double-focalisation et secteurs magnétiques. Dans des conditions de vide et de propreté optimales, les premiers instruments dits de haute résolution étaient capables d'atteindre une résolution de 10 000 – 20 000 pour une faible sensibilité (Russel et Edmondson, 1997). Selon ces auteurs avant les années 2000, la précision en masse des appareils pouvait déjà atteindre 1 et 5 ppm.

Le couplage de la chromatographie en phase gazeuse avec un analyseur à secteur magnétique est devenu un instrument de référence pour l'analyse des polluants organiques persistants (POP) et les publications concernant les analyses de polluants organiques sont nombreuses (Gross, 2014). La résolution (à 10% de la vallée) de ces spectromètres de masse est généralement comprise entre 10 000 et 20 000 (Hernandez *et al.,* 2012) même si des résolutions supérieures ont été rapportées (Santos *et al.,* 1996, 1997).

Le deuxième analyseur à haute résolution utilisé en couplage avec la chromatographie en phase gazeuse est le spectromètre de masse à temps de vol (TOF, time-of-flying). Le concept du TOF a été défini en 1946 par William E. Stephens. Les progrès de ces deux technologies (GC et TOF) les ont rendus attractives par rapport aux analyseurs à secteurs magnétiques. En effet, les instruments de type TOF sont aujourd'hui capables d'atteindre des résolutions supérieures à 50 000 voire plus de 80 000 selon les fournisseurs. Leur vitesse d'acquisition est parmi les plus élevées par rapport aux autres analyseurs. En 2003, Santos et Galceran ont conclu dans leur review que le couplage GC-TOF se popularisait grâce à ses capacités (sensibilité, résolution, précision en masse et vitesse d'acquisition) pour l'identification et la quantification en matrices environnementales (Santos et Galceran, 2003). De même, des publications récentes confirment que le couplage GC-TOF est une alternative efficace aux secteurs magnétiques notamment pour l'identification de molécules connues ou inconnues (Plaza-Bolanos *et al.,* 2013). Ces spectromètres de masse de type TOF ont aussi été couplés à la chromatographie en phase liquide. D'après Hernandez *et al.* (2012), les premières publications sur le couplage LC-TOF datent de 1998-2000 avec Hogenboom *et al.* (1999 et 2000). Cette équipe de chercheurs a essayé de déterminer les aptitudes du TOF pour l'analyse de micropolluants dans l'eau et l'identification de pesticides dans une première publication (Hogenboom *et al.,* 1999) puis a identifié des produits de transformations de l'alachlor (Hogenboom *et al., 2000*). A la même période, la première publication avec un système hybride Q-TOF (quadripole-time-of-flying) portant sur le criblage et l'identification de composés modèles dans l'eau potable est apparue. En 2006, Sancho *et al.* se sont penchés sur le potentiel du couplage LC-TOF et LC-(Q)TOF pour l'analyse de pesticides dans les eaux et ont souligné l'intérêt de cette technologie pour le criblage en mode full-scan, la quantification ou encore l'identification. En 2011, un exemple d'application sur les surfactants et leurs métabolites, présents dans des matrices eaux et sols, a montré l'intérêt du couplage LC- (LCT)TOF à la fois pour l'analyse multi-résidus et pour la recherche de nouveaux métabolites (Lara-Martin *et al.,* 2011). Enfin, plusieurs autres études ont également confirmé les capacités du couplage LC-(Q)TOF pour l'identification de composés inconnus (Ibanez *et al.,* 2004 et 2005, Hernandez *et al.,* 2004).

Les spectromètres de masse à résonance cyclonique ionique transformée de Fourier (FT-ICR) possèdent à ce jour le pouvoir de résolution le plus grand (résolution jusqu'à 1 000 000 pour un m/z de 400). Le coût prohibitif, la complexité d'utilisation et la taille de ces instruments ont limité leur « démocratisation ». Plusieurs exemples d'utilisation de ces instruments existent tout de même dans l'analyse de l'eau, notamment pour l'étude de la modification de la matière organique dissoute durant la potabilisation de l'eau ou encore son devenir pendant la chloration. L'analyse de l'eau de mer a fait également l'objet d'études avec ce type de spectromètre de masse FT-ICR (Hansman *et al.*, 2015).

Enfin à côté de ces spectromètres de masse, un équipement mettant en œuvre l'architecture de type Orbitrap a été mis sur le marché en 2005 (Hu *et al.,* 2005 ; Makarov *et al.,* 2006 et 2010). De par ses

performances (résolution supérieure à 100 000, sensibilité, précision en masse inférieure à 2 ppm et gamme de masse satisfaisante), cet appareil est présenté comme une alternative aux analyseurs de type TOF et FT-ICR. Il permet en effet de répondre à des problématiques nécessitant un pouvoir de résolution uniquement atteignable par les FT-ICR mais en ayant un coût d'achat et d'entretien plus accessibles (Gosetti *et al.,* 2015). Krauss et Hollender ont publié en 2008 une étude sur les capacités de l'Orbitrap pour la quantification des nitrosamines dans les eaux usées. Cette étude est considérée comme étant la première application environnementale de l'Orbitrap. Un an après, Hogenboom *et al.* (2009) se sont penchés sur l'utilisation de l'analyseur Orbitrap pour le screening et l'identification de polluants émergents. Plusieurs variantes de l'Orbitrap se sont développées par la suite avec des instruments hydrides (trappe à ions linéaires/Orbitrap ou quadripole-Orbitrap) ou non (Orbitrap seul). Ainsi, Helbling *et al.* (2010) se sont intéressés aux produits de transformations dans des eaux usées et ont utilisé avec succès le LIT-Orbitrap (linear ion Trap-Orbitrap) en modes Full Scan et MS^2 pour l'analyse ciblée et le criblage des produits de transformations avec pour but de les identifier. Wode *et al.* (2012) ont quant à eux exploité un Orbitrap non hydride pour une méthode multi-résidus portant sur 72 micropolluants dans des eaux d'origines variées (eaux potables, eaux usées, diluées ou non) en utilisant deux masses exactes par molécule. Ils en ont conclu que la méthode développée était rapide et sensible. Enfin, d'autres équipes ont employé l'Orbitrap avec succès pour quantifier des composés dans des eaux de différentes natures. Ils ont montré des limites de quantification comprises en 0,01 µg/L et 10 µg/L démontrant l'important champ d'application potentiel de cette technologie (Pinhancos *et al.,* 2011 ; Wille *et al.,* 2011).

En conclusion, une partie importante de la bibliographie faisant référence à la haute résolution est datée d'après 2005, période de démocratisation des spectromètres de masse à haute résolution (HRMS) (Hernandez *et al.,* 2012). Les qualités de ces analyseurs ont permis le développement de méthodes quantitatives pour l'analyse de polluants organiques mais surtout l'obtention d'un outil performant pour la recherche et l'identification de substances non répertoriées (Acena *et al.,* 2015). Ainsi, si les analyseurs de haute résolution ne sont pas exempts de défauts (vitesse d'acquisition limitant les capacités du couplage UHPLC-Orbitrap, instabilité de l'étalonnage, sensibilité en analyse ciblée et gamme dynamique pour l'analyseur TOF), leurs aptitudes actuelles et futures devraient continuer de participer à leur expansion au sein des laboratoires pour l'analyse des micropolluants et l'analyse environnementale en général (Gosetti *et al.,* 2015 ; Hernandez *et al.,* 2012 ; Krauss *et al.,* 2010).

II.3. Recherche de suspects ou « screening »

II.3.1. Principe de l'analyse de suspects ou suspect screening

L'analyse ou la recherche de suspects (suspects screening en anglais) permet la détection d'un très grand nombre de micropolluants sans réaliser de sélection réduite au départ (Acena *et al.,* 2015). Ce type d'analyse permet donc de s'affranchir en partie de l'utilisation d'étalons. L'appellation « suspects screening » a tout d'abord été définie par Krauss *et al.* (2010) et, par la suite, Hernandez *et al.* (2012) ont également proposé le terme additionnel « post-target ». L'objectif de l'analyse de suspects est le monitoring qualitatif ou semi-quantitatif (présence, fréquence) de molécules connues ou de suspects et ce, si possible, avec le niveau de confiance le plus élevé.

Le principe général de l'analyse de suspects ou de l'analyse *post-target* est le suivant : une liste des suspects d'intérêt est établie à partir de données internes, de données de la bibliographie et/ou de la législation. Cette liste comporte, pour chacun des suspects, leur formule brute, leur masse exacte et leur structure. Elle est dépendante des objectifs des travaux associés et peut comporter de plusieurs centaines au millier de molécules. L'analyse des échantillons étudiés est ensuite effectuée par de la spectrométrie de masse haute résolution, généralement couplée à une méthode séparative (en phase liquide ou en phase gaz), avec une méthode d'acquisition en mode Full Scan généralement associée à des analyses MS^2. De ces données sont ensuite extraites les masses exactes des suspects recherchés. En cas de résultats positifs, plusieurs étapes de confirmation d'identité peuvent alors être appliquées (i.e. analyses MS^2, injection de standard…).

II.3.2. Développement d'une méthode de screening

II.3.2.1. Etapes de l'analyse de suspects et de la confirmation d'identification

S'il peut exister des variations propres à chaque étude, le processus de l'analyse de suspects suit une ligne directrice constituée de plusieurs étapes (Krauss *et al.,* 2010 ; Hernandez *et al.,* 2012 ; Acena *et al.,* 2015) dont :

- Création de la liste de suspects,
- Prélèvement des échantillons,
- Préparation et analyse par GC-HRMS ou LC-HRMS,
- Prétraitement des données (détaillé dans le paragraphe II.4.2 de ce chapitre),
- Confrontation masses exactes-liste de suspects,
- Réduction par le massif isotopique, la prédiction des temps de rétention, la prédiction des produits de transformations, le ratio signal sur bruit, un critère de ratio intensité du pic sur aire du pic, la forme du pic…

- Comparaison des fragments mesurés et prédits par la MS^2 des suspects.

Ces étapes sont les plus fréquemment citées dans la littérature et sont susceptibles de varier lors de telles analyses. Toutefois, dans tous les cas, l'identification non équivoque nécessite la comparaison avec un étalon.

II.3.2.2. Développement et optimisation de la méthode

Le développement de la méthode instrumentale peut être effectué en s'appuyant sur un ensemble d'étalons couvrant une large gamme de propriétés. De même les étapes de traitement des données peuvent être évaluées par l'utilisation d'un mélange d'étalons à une ou plusieurs concentrations. Cette évaluation a pour objectif d'estimer principalement le taux de faux négatifs (c'est-à-dire les composés présents dans l'échantillon mais absents du résultat final) et de faux positifs (c'est-à-dire de composés absents mais identifiés par erreur lors du résultat final). Par exemple, Segura *et al.* (2011) ont validé leur méthode d'analyse de suspects par l'utilisation de dix-sept étalons. De même, Moschet *et al.* (2013) ont optimisé leur analyse de suspects par l'utilisation de quarante-cinq étalons puis ont calculé le taux de succès et estimé le taux de faux négatifs ainsi que de faux positifs. Après application de leur méthode, 70% de leurs étalons ont été détectés, 89% des pics ont été exclus par les filtres automatiques. Les faux négatifs sont principalement issus d'aires trop faibles pour être prises en compte ou de la soustraction d'une contamination présente dans le blanc.

II.3.2.3. Applications de l'analyse de suspects

Plusieurs familles de micropolluants ont été intégrées à des méthodes de screening tels que des produits pharmaceutiques et leurs métabolites ou des pesticides et leurs métabolites et ce, dans les matrices eaux de rivières, affluents (entrées) et effluents (rejets) de stations d'épuration. Ainsi, l'analyse de suspects peut se concentrer sur une famille de composés en particulier (Moschet *et al.,* 2013) ou sur un vaste éventail de micropolluants (Kern *et al.,* 2009 : Segura *et al.,* 2011,).

Kern *et al.* (2009) ont créé une liste de suspects composée de cinquante-deux contaminants connus parmi les produits pharmaceutiques, les pesticides et les biocides avec 1794 produits de transformations listés dans la littérature ou prédits par logiciel et ont utilisé cette liste pour l'analyse d'eaux naturelles. Après l'application d'une série d'étapes comprenant l'exclusion des masses exactes ayant une différence supérieure à 5 ppm avec celles des suspects, l'exclusion des pics d'intensité inférieure à 10^5 et inférieure à dix fois leur intensité dans le blanc, la plausibilité du temps de rétention *via* les log K_{ow}, la correspondance du massif isotopique, la plausibilité de l'ionisation en mode positif ou négatif et l'utilisation de fragmentation, les auteurs ont pu identifier dix-neuf produits de transformation à partir des 1057 pics détectés. De même, Moschet *et al.* (2013) ont appliqué une méthode d'analyse de suspects optimisée sur des échantillons réels avec une liste de 140 suspects de

pesticides et de leurs produits de transformation. Ils ont détecté 60 000 signaux dont 90% ont été éliminés par les étapes d'identification et seize suspects ont été confirmés par injection d'étalons sur les 30 molécules restantes après réduction des données. Segura *et al.* (2001) se sont également penchés sur une liste de 264 suspects issus de diverses familles de micropolluants ionisables en mode positif. La soustraction du blanc, l'intensité, la déviation de la masse exacte et les isotopes ont été les paramètres pris en compte par ces auteurs pour l'analyse de ces suspects. Segura *et al.* ont ainsi identifié la caféine, la désethylatrazine, la simazine et la venlafaxine, confirmés par standards. Enfin, Sjerps *et al.* (2016) se sont appuyés sur une liste de 5219 suspects issus des listes d'autorisation de substances REACH, CMR, pesticides/biocides, produits pharmaceutiques, des règlements des directives cadres de l'eau et des listes de molécules potentiellement pertinentes du réseau NORMAN et IAWR/RIWA pour rechercher des suspects dans divers types d'eaux (i.e. eau potable, eau de surface, effluent de STEP, effluent industriel et eau souterraine). À partir des signaux détectés, 154 signaux ont été reliés à 243 suspects dont 24 ont été confirmés par injection d'étalons, 89 ont pu être reliés à une formule brute grâce au massif isotopique et 45 ont été considérés comme des faux positifs.

Dans la littérature, il a été montré que l'analyse de suspects pouvait être effectuée seule ou en parallèle avec une analyse ciblée classique.

II.4. Analyse non ciblée

II.4.1. Principe de l'analyse non ciblée

L'analyse non ciblée est définie comme l'analyse de tous les signaux (connus ou inconnus) présents dans des échantillons sans *a priori*. Les objectifs de l'analyse non ciblée sont la mise en évidence voire l'identification des signaux par l'utilisation de méthodes dédiées. Fondée sur l'utilisation de la spectrométrie de masse haute résolution, des sources d'ionisation douces voire de l'utilisation de bases de données (si recherche de structure), elle permet d'explorer la complexité des échantillons environnementaux et de générer des connaissances quant à la contamination des milieux par les micropolluants organiques jusqu'alors inconnus et à leur devenir (Schymanski *et al.,* 2017).

II.4.2. Étapes de l'analyse non ciblée

Depuis l'apparition des premières analyses non ciblées pour des problématiques environnementales, un schéma général de l'approche utilisée s'est dégagé des différentes études réalisées jusqu'à présent. Cinq grandes étapes composent ainsi l'analyse non ciblée, elles-mêmes constituées de plusieurs sous-étapes :

- L'échantillonnage ou la récupération de prélèvements est la première étape à réaliser et dépend des objectifs de l'étude conduite. Les échantillons doivent être conservés au frais (4°C)

et analysés rapidement afin de les préserver et d'éviter la dégradation des micropolluants présents.

- L'étape suivante concerne la préparation d'échantillons et l'analyse chromatographique non ciblée. La préparation d'échantillons doit prendre en compte les propriétés physico-chimiques des classes de molécules étudiées. Lors d'une analyse non ciblée, on recherchera généralement la ou les méthodes les plus exhaustives. Il est nécessaire de garder en mémoire que si la préparation d'échantillon a pour but de concentrer les contaminants et diminuer l'effet matrice, une perte de potentiels marqueurs d'intérêts n'est pas à exclure. Un compromis est donc à trouver entre la concentration/purification des échantillons et la conservation de la plus large gamme de micropolluants possible. Pour la préparation des échantillons, l'extraction sur phase solide, avec l'utilisation d'une ou plusieurs phases différentes permettant de couvrir une gamme étendue de composés, est la méthode généralement employée lors d'une analyse non ciblée (Kern *et al.,* 2009 ; Jernberg *et al.,* 2013 ; Schollée *et al.,* 2015 ; Stephan *et al.,* 2016). Les données brutes pour les étapes suivantes sont ensuite obtenues par une méthode séparative couplée à la spectrométrie de masse haute résolution (TOF, Orbitrap ou FT-ICR). Pour cela, une analyse en mode Full Scan est tout d'abord effectuée en haute résolution, habituellement supérieure à 70 000, suivie par une analyse MS^2 si possible à plus faible résolution. Dans le cas de l'analyse de matrice eaux, la chromatographie liquide en phase inverse couplée à un analyseur haute résolution équipé d'une source électrospray est la méthode la plus répandue mais des alternatives telles que la chromatographie HILIC, la chromatographie à deux dimensions, la chromatographie ionique ou l'électrophorèse capillaire existent également (Schymanski *et al.,* 2017).
- L'étape de (pré)traitement des données est une étape totalement absente de l'analyse ciblée mais fondamentale pour la recherche de suspects et l'analyse non ciblée. Les objectifs de cette étape sont la diminution du volume de données et de sa complexité, et l'obtention d'une matrice de données comportant les m/z et temps de rétention unique pour chaque signal détecté avec son intensité ou son aire dans chaque échantillon. L'étape numéro un du (pré)traitement des données est la détection automatique des pics au-dessus d'un seuil de bruit fixé par un algorithme adapté au format des données brutes. Ensuite, la correction des temps de rétention est effectuée afin de compenser les décalages de temps de rétention des pics dans les différents échantillons par la sélection de pics communs entre les échantillons. Puis une étape d'alignement est réalisée entre les matrices obtenues pour les différents échantillons dont le but est de regrouper les signaux identiques ensemble en prenant en compte les décalages de m/z et de temps de rétention possibles. Cette étape est suivie d'une recherche des données manquantes. Elle permet de compléter les données par des signaux

n'ayant pas été détectés dans les étapes précédentes. Enfin, lors du prétraitement des données, les dernières étapes sont le regroupement des signaux correspondants à la même molécule (adduits, isotopes, multichargés) et la soustraction du blanc qui permet d'éliminer des pics n'appartenant pas aux échantillons. La prise en compte de réplicas peut également intervenir dans cette étape servant à l'élimination des pics non présents sur l'ensemble des réplicas et des pics artefacts instrumentaux.

- Une priorisation des données peut ensuite être réalisée pour mettre en évidence des signaux ou des marqueurs pertinents et réduire le volume de données que l'on souhaite identifier. Cette étape, lorsqu'elle est effectuée, s'appuie soit sur les données obtenues par l'analyse non ciblée seule soit en complément d'expériences menées en parallèle ou avant l'analyse non ciblée. La priorisation de données peut s'appuyer sur les données obtenues et certains critères définis par la problématique ou un choix prédéterminé : fréquence de détection à l'échelle de plusieurs campagnes de prélèvements, éléments chimiques spécifiques, intensité des signaux, données (éco) toxicologiques… Cette étape peut aussi se baser sur des méthodes d'analyse de données telles que les méthodes d'analyse univariée portant sur une variable à la fois ou les analyses multivariées portant sur plusieurs voire l'ensemble des variables.
- Enfin, en cas d'identification structurale, l'identification des molécules sélectionnées est la dernière étape de l'analyse non ciblée. Dans ces conditions, l'identification passe par l'emploi de toutes les informations disponibles : masse exacte, défaut de masse, massif isotopique, spectre de masse, fragmentation, bases de données internes et externes et le contexte général du domaine d'application de l'étude (Schymanski *et al.,* 2017). Toutes ces données permettent de donner un degré d'identification au signal considéré. Le plus haut niveau d'identification requiert toutefois l'utilisation d'un étalon qui n'est pas toujours disponible.

Les étapes de priorisation des données et d'identification des signaux d'intérêts reposent sur la nécessité d'avoir des données brutes de qualité, ce qui souligne l'importance de la planification des trois autres étapes, notamment celle de l'analyse et du (pré)traitement des données. La Figure 4 résume ce schéma général avec les étapes qui le composent :

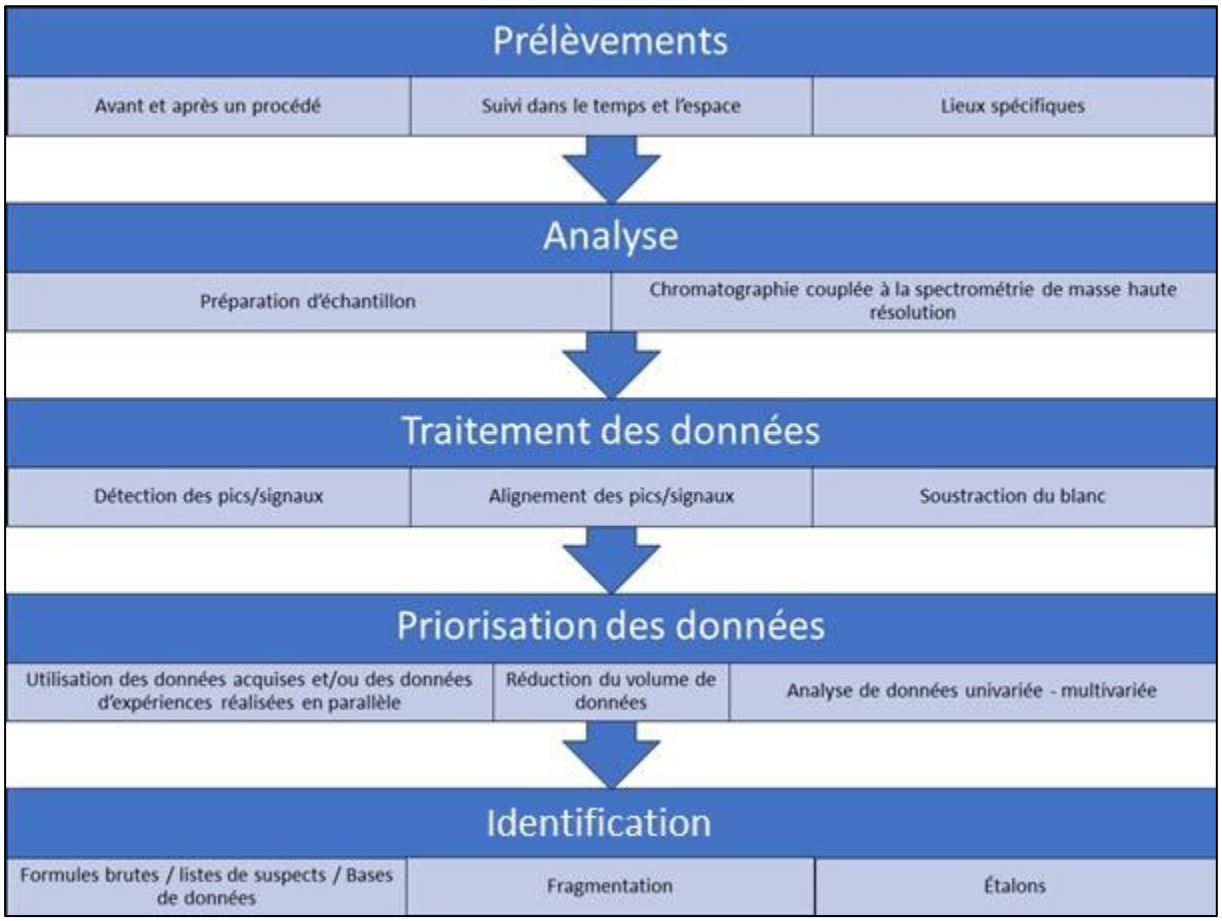


Figure 4 : Schéma général des étapes à réaliser lors d'une analyse non ciblée pour une application environnementale.

II.4.3. Les méthodes statistiques d'analyse de données

Lors de l'analyse non ciblée, il est possible d'étudier l'empreinte globale d'un échantillon pour le comparer à un ou plusieurs autres échantillons. Ces comparaisons peuvent se faire en termes de nombre de signaux observés, de signaux qui apparaissent ou qui disparaissent. De même, des traitements statistiques des données peuvent permettre de préciser si les échantillons sont significativement différents ou non, voire d'indiquer ce qui les différencie.

II.4.3.1. Principes généraux et méthodes

Les méthodes d'analyses statistiques, en plus de l'analyse descriptive, peuvent être classées en deux groupes : les analyses dites univariées et les analyses dites multivariées. Ces groupes sont composés de plusieurs types d'analyses dont le choix dépend de l'objectif et de la nature des données : quantitatives, qualitatives ou mixtes.

II.4.3.1.1. Analyses univariées

Les analyses univariées sont les analyses qui caractérisent la relation entre deux variables pouvant être elles-mêmes composées d'un ou plusieurs groupes de variables. La plupart des méthodes univariées

vont estimer la significativité de la différence entre les deux variables. Ainsi, il est possible de déterminer si la relation entre deux variables est due au hasard ou non. Aussi, la plupart des tests statistiques donnent une p-value qui est une estimation de la probabilité de rejeter l'hypothèse posée (même si elle est vraie).

De nombreux tests statistiques existent selon la normalité ou non de la distribution des variables (tests paramétriques et tests non-paramétriques). Le choix de la méthode dépend aussi du type de question posée (i.e. si on cherche à définir la relation entre deux variables ou à mettre en relief une différence) et de l'indépendance des variables. Les tests paramétriques les plus connus sont le test du Chi2 et le test de Fisher pour des variables qualitatives ou, le test de Student ou l'ANOVA (analyse de la covariance) lorsque l'on considère une variable qualitative et une variable quantitative. Ces deux derniers tests statistiques ont pour équivalents, comme tests non-paramètriques, les tests de Wilcoxon-Mann-Whitney et de Kruskal-Wallis. Les méthodes de corrélation (Pearson ou Spearman par exemple) sont plutôt employées lorsque la détermination de la relation entre deux variables est recherchée. Ces tests peuvent être employés seuls ou combinés avec d'autres paramètres.

II.4.3.1.2. Couplage analyses univariées – Fold Change

Le Fold Change se définie comme le ratio des intensités moyennes des signaux étudiés entre deux types d'échantillons, par exemple des sorties et des entrées de STEP. Le Fold Change peut être utilisé avec des méthodes univariées pour sélectionner des signaux dont les intensités diffèrent significativement entre deux catégories d'échantillons. Le Fold Change peut être combiné avec un test de Student ou une ANOVA (Schollée *et al.,* 2016). Les résultats obtenus sont représentés sous la forme d'un *volcano plot* qui exprime le logarithme en base 10 de la p-value ($-\log_{10}$(p-value)) en fonction du logarithme en base 2 du Fold Change ($\log_2$(Fold Change)). La Figure 5 présente un exemple de *volcano plot* issu de Schollée *et al.* (2016).

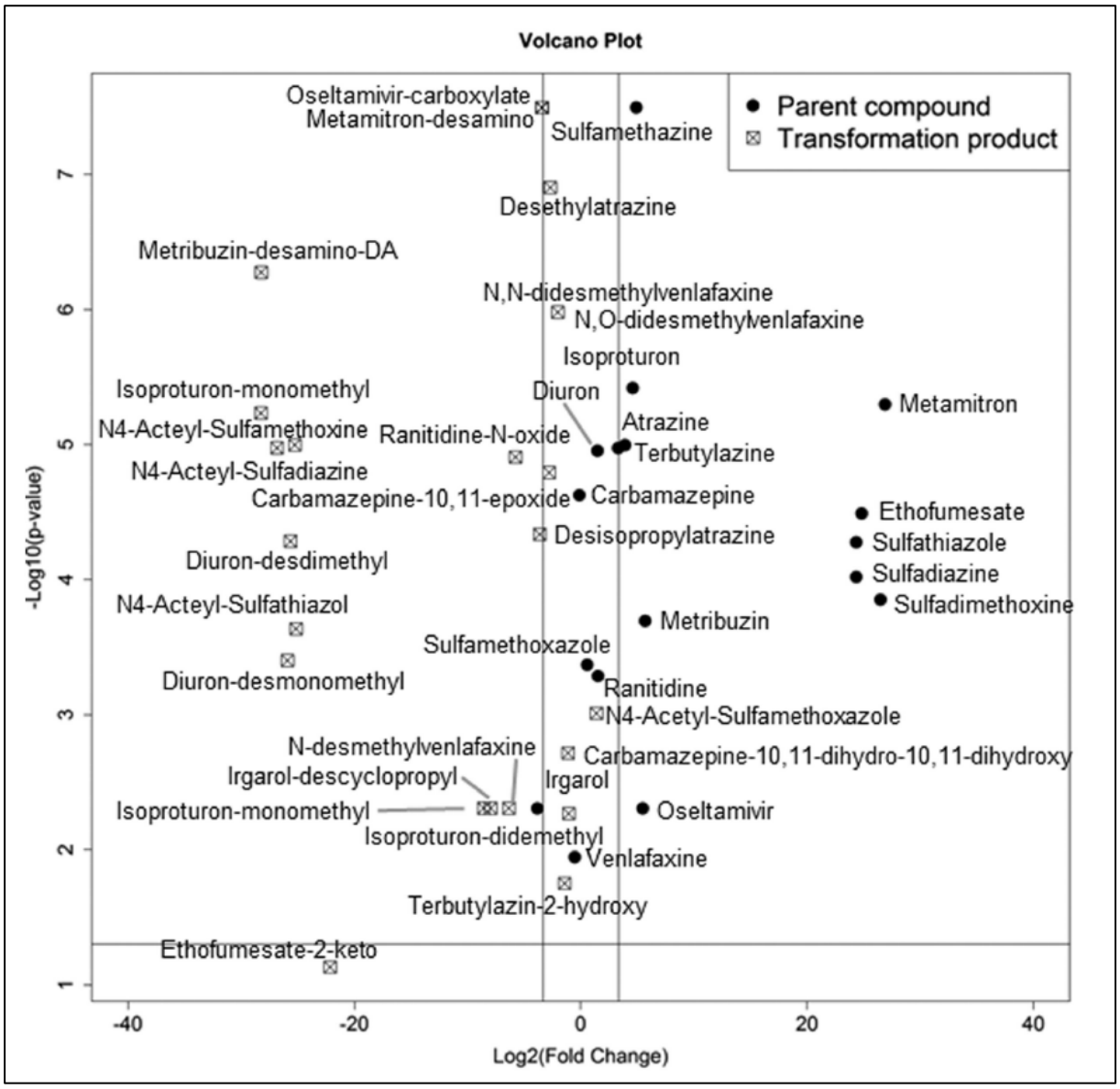


Figure 5 : Volcano-plot représentant -log10(p-value) en fonction du log2(Fold Change) Source : Schollée *et al.*, 2016. Fold Change : rapport entre les intensités des composés testés en entrées de STEP sur les intensités en sortie de STEP.

Sur cette Figure 5, la ligne horizontale correspond à une valeur de 1,30 (donc une p-value de 0,05, soit 5%) et les points en-dessous de cette ligne n'ont pas un Fold Change significatif. Les deux lignes verticales marquent les seuils inférieur et supérieur entre lesquels le Fold Change n'est pas suffisamment grand pour être pertinent. Lors d'une telle approche, le choix des seuils de Fold Change et de la p-value dépendent des objectifs des chercheurs. Les valeurs de Fold Change significatives sont souvent gardées. Toutefois, dans certaines expériences, concernant l'impact d'un traitement par exemple, les faibles Fold Change peuvent également être conservés car ils peuvent correspondre à des signaux réfractaires au traitement.

Le couplage analyses univariées – Fold Change permet donc d'effectuer un premier tri en fonction de la valeur du Fold Change et un second tri en fonction de la significativité de cette valeur. L'une des

finalités de cette méthode dans le cadre de l'analyse non ciblée est la mise en évidence de l'évolution de l'intensité de signaux lors de la comparaison de deux empreintes UHPLC-HRMS.

II.4.3.1.3. Analyses multivariées

Les analyses multivariées portent sur plusieurs variables, parfois en grand nombre. Il existe deux types d'analyse multivariée : les méthodes descriptives et les méthodes explicatives.

Les analyses descriptives décrivent l'ensemble du jeu de données pour mettre en relief sa structure latente. Ces méthodes permettent de regrouper des variables ou des individus partageant des points communs. Elles permettent aussi de réduire la dimensionnalité du jeu données avant l'application de méthodes explicatives. Les méthodes descriptives les plus employées sont l'Analyse en Composantes Principales (ACP) et la Classification Hiérarchique Ascendante (CAH). Néanmoins, peuvent être citées la méthode des nuées dynamiques (k-means), les mélanges gaussiens ou encore l'Analyse Factorielle des Correspondances (AFC). L'analyse en composantes principales repose sur une projection mathématique des variables dans un nouvel espace réduit par combinaison linéaire des variables en nouvelles variables, appelées composantes principales. L'ACP permet de réduire le volume et la redondance des données pour mieux visualiser et décrire le jeu de donnée et sa structure. Elle permet également de détecter des échantillons aberrants. L'interprétation des résultats de l'ACP se fait sur la projection des échantillons (ou individus) et/ou des variables dans des plans composés par deux composantes principales ou axes. Ces plans ou graphiques sont appelés « score plots » lorsque seuls les individus sont étudiés et « loading plot » ou cercle des corrélations pour les variables et leur corrélation vis-à-vis des composantes principales. Un biplot réunissant à la fois les individus et les variables permet de les relier entre eux. La classification hiérarchique est une autre méthode multivariée qui s'appuie sur le regroupement des échantillons et/ou des variables selon des classes ou « clusters ». Cette classification peut se faire selon la similarité ou la dissimilarité calculée selon une méthode d'agrégation. Le résultat est présent sous la forme d'un dendrogramme qui peut être coupé en plusieurs classes distinctes. La classification peut être couplée à une « heatmap » ou carte de chaleur afin de mettre en relief les clusters.

Les méthodes explicatives (et prédictives) ont pour but de relier une ou plusieurs variables à expliquer (Y) aux variables explicatives (X) du jeu de données. Dans le cas de l'utilisation de ces modèles pour la prédiction, il doit y avoir un jeu de données d'apprentissage du modèle et une validation pour éviter les phénomènes de sur-apprentissage conduisant à des erreurs (Triba *et al.*, 2015). Les méthodes explicatives usuelles sont les méthodes de régressions de type régressions des moindres carrées partiels (PLS ou « Partial Least Squares »). La PLS se base sur la combinaison de variables en nouvelles variables appelées variables latentes et peut être utilisée en présence d'un grand nombre de variables

pouvant covariées. Une variante de la PLS est la PLS-DA qui relie des variables explicatives à une variable à expliquer qualitative. Les résultats de la PLS donnent un modèle, un « loading plot » et un « score plot » ainsi qu'un biplot.

Préalablement à l'utilisation des méthodes multivariées, il convient de s'interroger sur la normalisation et/ou la transformation des données pour réduire la variance et/ou limiter le poids des signaux les plus intenses et réduire l'hétéroscédasticité (variance des résidus des variables différentes) due aux variations dans les échantillons et induite par la méthode d'analyse et de préparation (Schollée *et al.,* 2016 ; Van der Berg *et al.,* 2006). Le « centrage » des données (pour centrer la distribution des valeurs non pas autour de la moyenne mais autour de 0, en retranchant à chaque donnée la valeur de la moyenne) est nécessaire pour un grand nombre de méthodes multivariées, dont l'ACP. La mise à l'échelle ou « scaling » pour réduire l'influence des intensités ou aires les plus importantes peut s'effectuer grâce à la division par l'écart-type ou encore la moyenne. Enfin, les transformations des valeurs peuvent passer par une transformation logarithmique ou une transformation basée sur la moyenne des valeurs.

II.4.3.2. Exemples d'applications dans le cas d'analyses de l'eau

Les méthodes statistiques ont déjà été employées dans le cadre de diverses études. La recherche de produis de transformation a, par exemple, exploité plusieurs types de méthodes statistiques.

Negreira *et al.* (2015) ont détecté 19 produits de transformation de l'anticancéreux erlotinib lors d'expérience de chloration. Pour obtenir ce résultat, deux groupes d'échantillons ont été comparés : les échantillons dopés par l'erlotinib ayant subi l'expérience de chloration et ceux d'un groupe contrôle. Les signaux ayant des intensités significativement différentes entre ces deux groupes ont été extraits à l'aide du test de Student. Ces signaux ont alors été considérés par Negreira *et al.* (2015) comme des produits de transformation potentiels. De la même manière, Bourgin *et al.* (2013) ont mis en évidence des produits de transformation du bisphénol A lors d'expériences de chloration d'eaux de boissons. Cette étude a comparé des échantillons chlorés avec ceux d'un groupe contrôle pour mettre en avant les signaux ayant une intensité au moins quatre fois plus intenses que dans le groupe contrôle. La significativité de ces ratios a été testé par un test univarié. Par cette méthode, 21 produits de transformations ont ainsi été mis en évidence. Schollée *et al.* (2015) ont également testé cette approche sur un ensemble d'étalons composés de molécules parentes et certains de leurs produits de transformations ajoutés à une concentration connue dans des échantillons d'entrées et de sorties de STEP. Les Fold Change des intensités entre les échantillons ont été calculés et la significativité de ces Fold Change pour chaque étalon a été obtenue par une combinaison de tests de Welch et de Mann-Whitney. Ces travaux ont montré que si tous les étalons avaient une intensité significativement

différente entre les entrées et les sorties, 17 des 42 étalons testés ne montraient en revanche pas de différence d'intensité supérieure à au moins un ordre de grandeur. Ces exemples témoignent de l'utilité de cette approche pour apprécier les différences entre des échantillons et pour mettre en évidence des signaux (associés à des produits de transformation). Toutefois, un certain nombre d'entre eux peuvent passer inaperçus en raison d'un Fold Change inférieur à un critère seuil donné suite à des effets matrices ou d'autres mécanismes d'élimination ou de reformation de la molécule parente (Schollée *et al.*, 2016).

L'ACP a été utilisée pour visualiser des regroupements de catégories d'échantillons et déterminer des corrélations entre les échantillons et des signaux. Kiss *et al.* (2017) ont par exemple observé des regroupements entre les échantillons d'eaux usées et d'eaux traitées (par différentes étapes de traitements) de part et d'autre du « score plot ». Ils ont mis en évidence la séparation entre les eaux traitées par traitement biologique et ozonation, et les eaux traitées seulement par traitement biologique. La variabilité des eaux usées en fonction de la campagne de prélèvement a aussi été mise en valeur par le « score plot » de l'ACP. Les méthodes multivariées ont également été employées dans la recherche signaux associés à des produits de transformation. Schollée *et al.* (2015) se sont servis de l'ACP pour rechercher des produits de transformations formés entre des échantillons d'entrées et de sorties de STEP. Pour cela, une partie des signaux détectés a été classée comme étant associée soit aux entrées soit aux sorties à partir du loading plot de l'ACP (Figure 6).

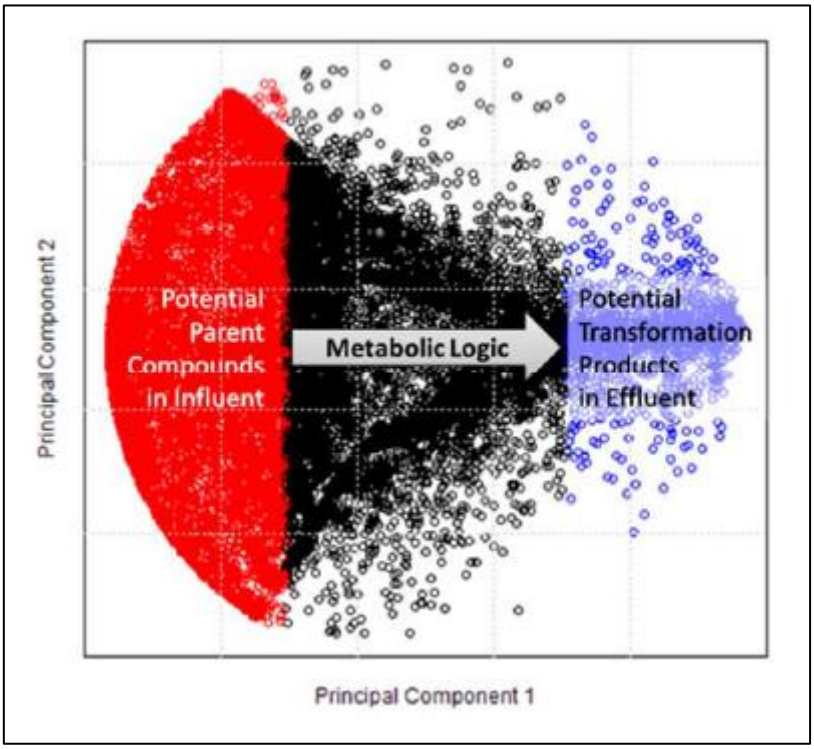


Figure 6 : Loading plot de l'ACP ayant servi de base au classement des signaux. À gauche se trouve en rouge les signaux classés comme associés aux entrées de STEP et à droite en bleu en sorties de STEP. Les signaux en noir n'ont pas été conservés. Repris de Schollée *et al.* (2015).

En faisant l'hypothèse des réactions chimiques pouvant s'opérer pendant le traitement des eaux, et donc en connaissant les variations de masses associées à la perte ou au gain d'un groupement chimique, Schollée *et al.* ont recherché les écarts de masse caractéristiques entre les deux classes de signaux pour extraire des paires molécules parentes-produits de transformations. Vingt paires de molécules parentes-produits de transformations, principalement des surfactants, ont pu être ainsi déterminées. Cette même équipe a étudiée en 2018 le nombre de produits de transformation formés lors de l'ozonation d'effluent de STEP par cette même technique afin de déterminer la dose d'ozone produisant le plus de produits de transformations. Parmi ceux-ci, 69 issus de 27 molécules parentes différentes, ont reçu une proposition de structure (dont 3 confirmés par un étalon). Dans ces mêmes travaux, les intensités maximales des signaux lors de quatre étapes du traitement des eaux comprenant notamment une étape d'ozonation et une étape de post-ozonation ont été relevées. L'application d'une classification hiérarchique sur les intensités des signaux en fonction de l'étape de traitement a permis à Schollée *et al.* d'isoler 30 groupes de signaux pour 11 tendances de variation différentes comme le montre la Figure 7.

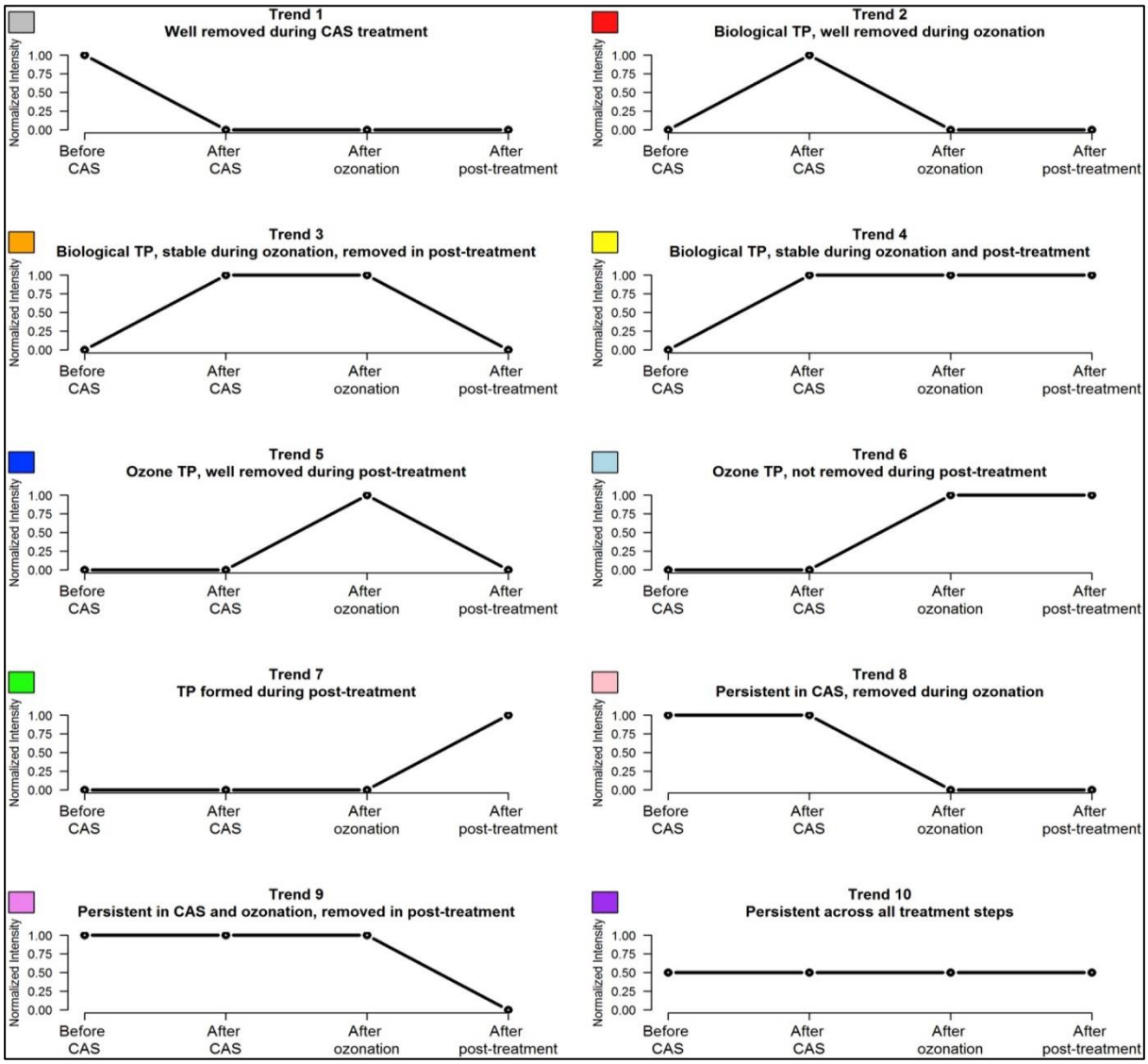


Figure 7 : Exemples des tendances pouvant être suivies par les intensités des signaux au cours des étapes du traitement des eaux et déterminées par les clusters isolés par la classification hiérarchique. CAS : procédé conventionnel des boues activées. Repris de Schollée *et al*. (2018).

De cette façon, Schollée *et al.* (2018) ont pu déterminer plusieurs groupes de signaux suivant la même tendance, dont certains se formant durant le procédé d'ozonation. De la même manière, Albergamo *et al.* (2019) se sont appuyés sur la classification hiérarchique pour isoler des signaux qui suivent les mêmes tendances de variations d'intensité, au cours du temps dans un système de filtration sur berge. Ces travaux ont permis de montrer la persistance et la mobilité de molécules polaires. 72 des signaux associés ont reçu une proposition de structure et 25 d'entre eux ont été confirmés dont certains non encore identifiés à ce jour dans les eaux naturelles.

Enfin, l'équipe de Hug *et al.* (2015) s'est servie de la méthode PLS pour tenter d'identifier des composés inconnus présentant une activité mutagène dans des eaux de STEP. Une analyse non ciblée a été réalisée puis les signaux détectés ont été utilisés comme variables explicatives pour expliquer l'activité mutagène. Des PLS successives ont été effectuées jusqu'à obtenir le meilleur modèle prédictif avec la

meilleure corrélation pour un minimum de signaux, validé par un jeu de signaux sélectionnés au hasard. De cette façon, un modèle optimal sur 205 signaux a été conservé et une tentative d'identification sur ces signaux a été réalisée. Au total, quatre signaux ont été identifiés et des structures (non confirmée par un étalon) ont été proposées pour quatre autres signaux.

II.4.4. Les approches de priorisation

L'identification exhaustive de l'ensemble des signaux présents lors d'une analyse non ciblée est une tâche pratiquement impossible et n'est pas forcément pertinente. Ainsi, lorsque l'on souhaite faire de l'identification structurale, une étape de priorisation est nécessaire afin d'exposer les signaux pertinents à étudier par rapport aux objectifs de l'étude mise en œuvre. Plusieurs stratégies sont possibles pour la réalisation de cette étape dont la priorisation par les données acquises des analyses et la priorisation par les expériences (Schymanski *et al.,* 2017).

Il est important de noter qu'en réalité de nombreuses études se situent entre l'analyse de suspects et l'analyse non ciblée. C'est par exemple le cas des études recherchant des produits de transformations lors d'expériences contrôlées (telle que l'étude de l'action d'un oxydant en traitement des eaux sur une molécule connue pour en déterminer les produits de transformation de structure proche du composé parent) (Krauss *et al.,* 2010).

II.4.4.1. Priorisation par les données

La priorisation par les données consiste à utiliser les données obtenues après traitement et de mettre en évidence des signaux partageant une caractéristique parmi la liste suivante d'exemples :

- Un seuil d'intensité des signaux ou une fréquence de détection,
- Un profil isotopique caractéristique (C, Cl, Br, N, O, S),
- L'appartenance à une série d'homologues,
- La détection de groupements fonctionnels caractéristiques,
- La détection de profils spatiaux ou temporels à travers une série d'échantillons,
- La comparaison avec une liste de suspects.

La sélection par un seuil d'intensité (ou d'aire) des pics est l'approche la plus simple d'utilisation. Elle se base sur l'hypothèse d'une relation concentration-intensité, c'est-à-dire que les pics les plus intenses ont une concentration élevée les rendant donc pertinents pour l'identification. Toutefois, cette hypothèse n'est pas toujours vérifiée et vérifiable. Elle n'est pas non plus forcément liée à un risque significatif. Ainsi, des pics moins intenses mais relatifs à des molécules présentant un risque plus significatif peuvent être exclus (Schollée *et al.,* 2015). C'est néanmoins une méthode souvent appliquée. La sélection par un classement d'intensité ou d'aires de pics peut être associée à d'autres

critères de sélection comme un profil isotopique. C'est la méthode qu'ont choisi Chiaia-Hernandez *et al.* (2014), Hug *et al.* (2013) et Ruff *et al.* (2015), qui ont sélectionné préférentiellement les signaux ayant des massifs isotopiques caractéristiques de la présence de Cl, Br, S ou N et faisant partie des 100 pics présentant la plus forte intensité pour Chiai-Hernandez *et al.* d'une part, ou avec des intensités supérieures à $3x10^6$ pour Hug *et al.* d'autre part. Sans isotopes caractéristiques, les 10 pics les plus intenses, les pics d'intensité supérieure à $2x10^7$ en positif et $7x10^6$ en négatif ou les 150 pics dont les aires sont les plus élevées ont été gardés pour être examinés en détails respectivement par Chiaia-Hernandez *et al.* (2014), Hug *et al* (2013) et Ruff *et al.* (2015). Ainsi, à partir de ces approches de priorisation ces auteurs ont pu sélectionner entre 30 et plus de 700 composés à partir de listes de plusieurs milliers de signaux au départ, permettant une réduction importante du volume de données. Gago-Ferrero *et al.* (2015) se sont également basés sur des filtres caractéristiques d'intensité et d'isotope pour prioriser leurs résultats et ont ainsi retenus plusieurs dizaines de molécules. Hug *et al.* ainsi que Ruff *et al.* ont utilisé comme filtre la répétitivité de la détection des signaux dans plusieurs échantillons pour être retenus comme composés d'intérêts. Cette notion de fréquence peut permettre de se focaliser sur des possibles contaminants réguliers et donc une pollution de type diffuse. Chiai-Hernadez *et al.* (2014) ont quant à eux utilisé le défaut de masse comme critère d'exclusion en plus des filtres indiqués précédemment, pour se concentrer sur un nombre de formules brutes plus restreint. Le défaut de masse, plus particulièrement le défaut de masse de Kendrick, a été également utilisé par Merel *et al.* (2016) pour relier les produits de transformation de type N-oxyde et leurs composés parents après ozonation. En effet, le concept du défaut de masse de Kendrick est de considérer que les molécules ayant la même structure mais qui différent par le nombre de répétition d'un même motif auront le même défaut de masse. Créé au départ pour des séries d'homologues liées par un motif CH_2, le défaut de masse de Kendrick est applicable à d'autres motifs pouvant révéler des séries de molécules homologues ou des contaminants parents-produits de transformations comme dans l'étude de Merel *et al.* qui se sont intéressés à 159 signaux après application du défaut de masse. Thurman *et al.* (2014) ont également adopté une prise en compte du défaut de masse pour mettre en évidence les homologues éthoxylates et ont ainsi fait ressortir deux séries d'homologues. Kiss *et al*. (2017) se sont appuyés sur les représentations des empreintes (m/z en fonction du temps de rétention) pour rechercher des motifs ou des écarts de masse distincts. Kiss *et al*. ont ainsi mis en évidence l'iohéxol et un signal pouvant correspondre à l'ioméprol ou l'iopamidol ainsi que trois potentiels produits de dégradation. Ces auteurs ont également mis ou jour grâce à des différences de masse de 44,02 Da des surfactants de type polyéthoxylates associés à 45 signaux. Enfin, Kiss *et al*. ont utilisés des spectres MS^2 pour recherche des pertes de masses correspondant à des fragments spécifiques pour les relier à des classes de molécules telles que les amines quaternaires. Muz *et al.* (2016) et Zonja *et al.* (2015) se sont eux attachés à mettre en évidence des composés inconnus partageant des groupes

fonctionnels semblables, issus des agents de contrastes iodés pour Zonja *et al.* et des amines aromatiques pour Muz *et al.* (2017). Enfin, une autre forme de priorisation des données a été utilisée par Albergamo *et al.* (2019) qui se sont appuyés sur la méthode de classification hiérarchique ascendante (CAH) pour détecter des groupes de profils de variations d'intensité entre leurs différents échantillons puis sélectionner dans ces profils les signaux les plus intenses pour des tentatives d'identification. Après application de leur méthode de priorisation, Albergamo *et al.* n'ont conservé que 3 764 et 3 845 signaux respectivement en mode positif et négatif sur 10 850 et 7 412 signaux récupérés après (pré)traitement des données, soit une diminution de 65% et 48% du volume initial de données. La dernière approche de priorisation évoquée précédemment est la comparaison avec une liste de suspects déjà traitée dans le paragraphe II.3. La plupart des publications référencées dans le paragraphe II.4.4.1 couplent d'ailleurs l'analyse de suspects et l'analyse non ciblée.

En conclusion, diverses méthodes de priorisation par les données ont été utilisées après (pré)traitement des données afin d'en réduire le volume. Un filtre d'intensité est très souvent appliqué en première attention pouvant être suivi ou précédé par d'autres méthodes telles que la sélection par la fréquence de détection ou des variations d'intensités.

II.4.4.2. Priorisation par les expériences

La priorisation par les expériences se base sur l'exploitation des données d'analyse couplées à la mise en œuvre d'expériences contrôlées ou d'expériences de toxicité. La liste suivante présente des exemples d'approches de priorisation en relation avec ce type d'expérience :

- Recherche de la persistance, de l'élimination/formation en cours d'un procédé,
- Formation de produits de transformations par des réactions biologiques, électrochimiques ou encore oxydatives,
- Réaction avec des composés marqués,
- Analyse dirigée par les effets.

La première méthode de priorisation citée ici a été utilisée par Müller *et al.* (2011) et Parry *et al.* (2016). Müller *et al.* ont mis en relation les empreintes de lixiviats de décharge, celles d'eaux souterraines en aval de ces lixiviats, celles d'eaux puisées dans ces eaux souterraines mais pas encore traitées et enfin celles d'eaux de distribution. Müller *et al.* ont ainsi montré que l'ensemble de ces matrices sont interconnectées et ils ont priorisé les signaux persistants à travers ces matrices depuis les lixiviats jusqu'à l'eau puisée et l'eau traitée en recherchant les pics communs entre les empreintes des différents échantillons. Ainsi, à partir de ce travail, ils ont conservé 13 signaux dans l'objectif de les identifier. Parry *et al.* ont quant à eux mis en place des expériences d'oxydation avancée par UV seul et UV/H_2O_2. Ils ont comparé les performances générales par paires de conditions testées et en

considérant la distribution des signaux significativement éliminés (p-value < 0,05) par condition testée. Ces comparaisons leur ont permis de déterminer les meilleures conditions opératoires et de mettre en évidence les composés difficilement éliminés ou au contraire ceux disparaissant facilement. Au cours de ce travail, ces auteurs se sont également intéressés aux produits de transformation issus des procédés d'oxydation avancée en sélectionnant les signaux uniquement présents dans les effluents et ceux dont l'intensité croît (88 signaux ont ainsi été mis en évidence pour le meilleur traitement). La recherche de produits de transformation a également été l'objet de l'étude de Schollée *et al.* (2015), qui se sont appuyés sur l'analyse en composante principale (ACP) pour souligner les signaux corrélés aux affluents de STEP ainsi que ceux corrélés aux effluents et pour rechercher des liens entre ces catégories de signaux et la prise en compte de treize biotransformations susceptibles de se produire dans un traitement par boues activées. A partir de cette approche, Schollée *et al.* ont ensuite gardé les 300 potentiels produits de transformation les plus significatifs (p-value < 0,01) pour un examen plus approfondi par l'étude de la qualité des pics et des relations structurales entre les composés parents supposés et les produits de transformations suspectés. Helbling *et al.* (2010) se sont aussi penchés sur les produits de biotransformations lors d'un traitement dans une unité pilote de laboratoire en considérant 12 molécules appartenant à la classe des produits pharmaceutiques ou de pesticides. Ils ont confronté les analyses avant et après traitement afin de sélectionner des signaux apparaissant durant l'expérience de biotransformation. Ils sont parvenus à identifier 26 produits de transformations pour ces molécules. De même, dans la volonté d'identifier de nouveaux produits de transformations, Borowska *et al.* (2016) et Schollée *et al.* (2018) se sont penchés sur les produits formés pendant l'ozonation. Pour cela, similairement à Helbling *et al.* (2016), Borowska *et al.* ont sélectionné en amont trois produits pharmaceutiques et ont fait la comparaison entre les échantillons non-ozonés et ozonés sans ajouts d'étalons avec les échantillons contenant les étalons et ozonés. Ces auteurs ont ainsi pu déterminer quinze produits de transformation liés à l'ozonation de ces trois molécules. Schollée *et al.* (2018) ont utilisé quant à eux la méthode combinée de l'ACP avec les réactions de transformations connues. Ils ont pu ainsi conserver un nombre varié de potentiels produits de transformation, en fonction des doses d'ozone utilisées (de 17 à 1 416 pour une dose d'ozone de 0,4 à 1 g d'ozone par gramme de carbone organique dissous) pour une exploration plus poussée. De la même manière qu'Albergamo *et al.* (2019), Schollée et al. se sont aussi servis de la CAH pour comparer cinq méthodes de traitements post-ozonation, notamment par le suivi des profils d'intensité des signaux, identifiant ainsi onze tendances statuant sur la méthode post-ozonation la plus efficace et sur le devenir de signaux formés durant l'ozonation. Enfin, Kolkman *et al.* (2015) ont également travaillé sur un procédé d'oxydation avancée (UV/H_2O_2) avec pour but de mettre en évidence les produits de transformation azotés. Ils ont pour cela employé du nitrate marqué ^{15}N et ont considéré des échantillons de solutions de laboratoire, avec et sans nitrate marqué, traitées à l'UV/H_2O_2. Ils ont alors recherché les signaux

correspondant à des masses exactes séparées de 0,99704 Da et présentant une différence de temps de rétention inférieure à 0,10 minutes. Cette approche leur a permis de mettre en relief 84 produits de transformations azotés dont 22 ont été par la suite détectés dans des échantillons d'eau potable soumises au traitement d'oxydation.

La dernière méthode de priorisation orientée vers les expériences listée précédemment, (i.e. l'analyse dirigée par les effets (effect-directed analysis EDA)), est une approche qui vise à identifier les composés participant à la toxicité d'un milieu ou d'une matrice en couplant des mesures de toxicité (mutagénécité, carcinogénécité, perturbation endocrinienne, toxicité aigüe, écotoxicité…), obtenues par des bioessais ou des bioindicateurs, avec une caractérisation obtenue par des analyses chimiques. L'EDA consiste à mesurer un paramètre de toxicité (perturbation endocrinienne, mutagénicité, carcinogénicité, toxicité aiguë…) sur différentes fractions d'un même échantillon obtenues par une méthode de fractionnement (telle qu'une méthode chromatographique par exemple). Les fractions dont la toxicité est positive sont ensuite de nouveau fractionnées, la toxicité des nouvelles fractions mesurée et ainsi de suite. Ces étapes successives de fractionnement puis de mesure de la toxicité sont répétées jusqu'à l'obtention de mélanges simples dont la nocivité est mesurable et permet l'identification des composés participant à la toxicité. Brack *et al.* (2017) ont publié une analyse en profondeur de l'utilisation de l'EDA allant du développement de l'étude au choix de la méthode d'évaluation de la toxicité en passant par la méthode de préparation d'échantillons et d'analyse jusqu'à l'identification ainsi que sur les outils disponibles pour effectuer l'EDA. Il en ressort que l'EDA est une méthode efficace pour mettre en évidence les molécules responsables de la toxicité dans le milieu ou dans la matrice étudiée. L'EDA peut permettre de s'intéresser à différents marqueurs de toxicité : perturbation endocrinienne, mutagénicité, génotoxicité ou encore réponse à un stress. Cette méthode possède néanmoins des inconvénients non négligeables : les coûts en temps et en argent. En effet, la multiplication des fractionnements et des tests de toxicité en vue de l'obtention de mélanges simples implique un processus long. De plus le succès de l'EDA n'est pas garanti. Weiss *et al.* (2011) ont utilisé comme approche l'EDA en développant une méthode pour identifier les composés ayant une activité androgénique ou anti-androgénique. Pour cela, ces auteurs ont fractionné leurs échantillons et testé séparément chacune des fractions obtenues par un bio essai AR-CALUX® et une analyse par couplage LC-HRMS. La comparaison des fractions actives et non actives a permis de discriminer les signaux d'intérêt à identifier. Trois muscs polycycliques, deux organophosphorés, deux stéroïdes et un hydrocarbure aromatique polycyclique ont ainsi été identifiés. Un autre exemple d'EDA est celle effectuée par Hug *et al.* (2015). Ces auteurs se sont appuyés non pas sur le fractionnement d'échantillons mais sur une régression par les moindres carrés partiels (régression PLS) pour lier la mutagénicité et les signaux détectés. Ainsi, 205 signaux ont covarié avec la mutagénicité dont quatre ont été identifiés et quatre autres ont été proposés sans avoir de confirmation par l'utilisation

d'étalons. Cette méthode, se basant sur une méthode d'analyse multivariée pour mettre en évidence les molécules toxiques, est appelée « virtual-EDA ». C'est une alternative à l'EDA classique et qui ne repose pas sur un fractionnement.

II.4.5. Stratégies d'identification et de caractérisation

II.4.5.1. Identification

L'identification des pics sélectionnés par l'étape de priorisation repose sur plusieurs outils. Ces outils, généralement apportés par des logiciels libres de droits ou commerciaux, reposent sur l'utilisation de bases de données locales ou en ligne. Ce sont les outils de prédiction de formules brutes et les outils de comparaisons des spectres de fragmentation. Il convient de faire la distinction entre les composés « connus-inconnus » et « inconnus-inconnus ». Les premiers ont déjà été reportés dans la littérature et les bases de données, alors que les seconds n'ont jamais été détectés. L'identification des seconds reste la plus ardue car elle repose principalement sur l'étude du spectre de masse, des fragments et de méthodes complémentaires telles que la RMN. Pour les premiers, l'attribution d'une formule brute est en règle générale la première étape pouvant s'appuyer sur le spectre de masse, notamment en présence de massifs isotopiques spécifiques (Cl, Br ou S). La confrontation avec une ou plusieurs bases de données comme ChemSpider ou des bases de données internes vient ensuite en second avec la possibilité de comparer les massifs isotopiques théoriques avec les massifs isotopiques des composés analysés. La fragmentation MS^2 ou MS^n est un complément non négligeable qui permet là aussi de comparer des bases de données de fragmentation avec les fragments obtenus. De même, l'utilisation des fragmentations *in silico*, de prédiction des temps de rétention, d'analyse de la cohérence de l'ionisation par rapport à la structure proposée sont autant d'outils supplémentaires pour proposer un nom, la dernière étape étant de confirmer par un étalon. Schymanski *et al.* (2015) ont proposé 5 niveaux de confiance d'identification. Le niveau 5 est la masse exacte d'intérêt sans autres informations. Le niveau 4 concerne les signaux associés à une formule brute sans équivoque. Les composés de niveau 3 sont des candidats potentiels. Les composés de niveau 2 ont une structure probablement identifiée à l'aide de la base de données et de la fragmentation. Enfin le niveau 1 regroupe les composés identifiés par un étalon. Dans le cas de la non disponibilité d'un étalon, le niveau 2 est le dernier niveau à atteindre. Ces niveaux de confiance sont présentés en Figure 8. Cette classification de niveaux de confiance est une proposition pouvant être amenée à évoluer.

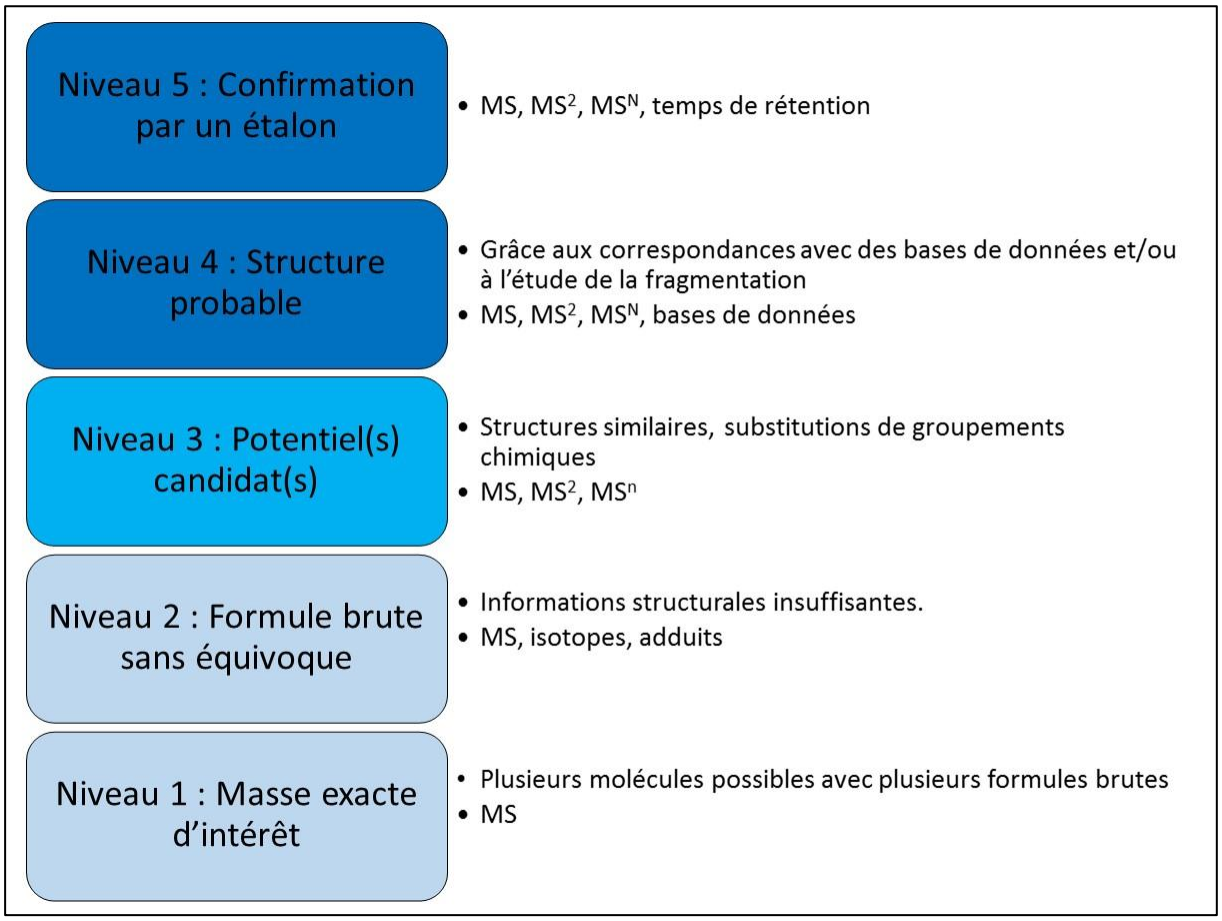


Figure 8 : Classification des niveaux de confiance pour l'indentification des signaux selon Schymanski *et al.* 2015.

II.4.5.2. Caractérisation par les formules brutes

L'attribution d'une formule brute est l'une des premières étapes dans la caractérisation des signaux inconnus. De ces formules brutes peuvent être calculés les ratios H/C, O/C, N/C ou encore S/C. Le ratio H/C peut donner des indications sur le caractère aliphatique ou aromatique du composé. En effet, un ratio H/C élevé indique une probabilité plus élevée d'une structure aliphatique ou saturé tandis que lorsque ce ratio diminue, la structure du composé tend vers une structure insaturée ou aromatique. Ces ratios peuvent être représentés sous la forme de graphiques de Van Krevelen (Van Krevelen, 1950) et ont déjà été utilisés dans la bibliographie pour caractériser la matière organique naturelle et la matière organique dissoute, par la représentation du ratio H/C en fonction du ratio O/C. Kim *et al.* (2003) ont utilisé le graphique de Van Krevelen sur des eaux noires pour en déduire des voies de réactions chimiques entre les structures dans la matière organique naturelle et analyser les classes de molécules présentes. Hockaday *et al.* (2009) ont effectué le même type de travail sur la matière organique dans des eaux naturelles. Ces travaux se sont appuyés sur le graphique du ratio H/C en fonction de O/C et ont été projetés sur le graphique différentes classes de molécules (protéines, lipides, carbohydrates, lignines etc…) déduites par la valeur des ratios (Tableau 3).

Tableau 3 : Gammes de ratios associées aux différentes classes de molécules (Maizel et Remucal, 2017 ; Hockaday *et al.*, 2009).

Classes de molécules	Ratio H/C	Ratio O/C
Lignines (6)	0,7 à 1,5	0,1 à 0,67
Protéines (2)	1,0 à 2,2	0,67 à 1
Lipides (1)	1,5 à 2,0	< 0,3
Carbohydrates (4)	> 1,5	> 0,67
Tannins (7)	0,6 à 0,9	0,5 à 1,3
Sucres aminés (3)	0,6 à 0,7	1,5 à 1,7
Structures condensés (5)	< 0,25	0,5 à 1,1

Les résultats de cette étude ont montré que la méthode d'ionisation (ESI, APPI) et la méthode d'extraction SPE ont un impact sur la nature de la matière organique analysée. D'autres études sur la matière organique se sont également appuyées sur ce type de graphique pour en déduire le type de familles de molécules présentes en majorité dans les échantillons et la nature des structures possibles dans des eaux naturelles et dans des eaux de STEP (Arnold *et al.,* 2014 ; Maizel et Remucal, 2017 ; Phungsai *et al.,* 2016 ; Verkh *et al.,* 2018). Ces études ont montré la prévalence des molécules de type CHO, CHNO, CHOS et CHNOS dans les eaux avec de nombreuses molécules appartenant aux domaines des lignines, protéines, peptides et carbones insaturés. La Figure 9 présente un exemple de graphiques de Van Krevelen avec les ratios H/C en fonction des rapports O/C ou de N/C utilisés par Verkh *et al*. (2018).

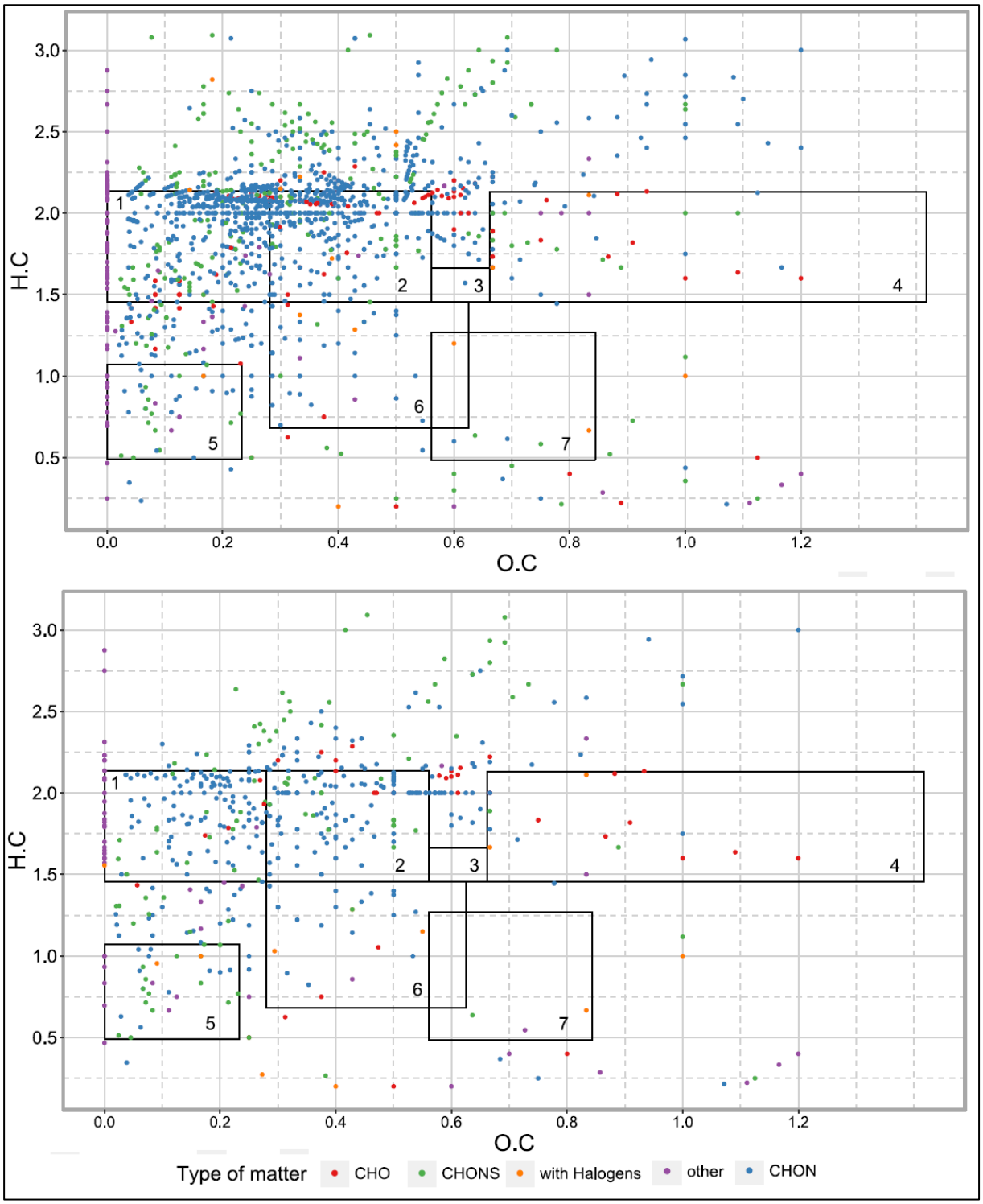


Figure 9 : Figure de Van Krevelen réarrangée de Verkh et al. (2018) du rapport H/C en fonction du rapport O/C de molécules détectées dans des eaux de STEP en entrée de STEP (au-dessus) et après traitement biologique (en-dessous).

La prédiction sur un grand nombre de signaux et particulièrement sur des signaux de hautes masses peut engendrer des fausses attributions dont il est nécessaire d'avoir conscience (Verk *et al.*, 2018). Toutefois, ces travaux montrent l'intérêt de tenter d'attribuer des formules brutes aux signaux même lorsque l'identification n'est pas poussée plus loin.

II.5. Conclusions

On peut aujourd'hui distinguer trois types de méthodes d'analyses des micropolluants : l'analyse ciblée, l'analyse de suspects et l'analyse non-ciblée (Krauss *et al.,* 2010, Hernandez *et al.,* 2012). Ces trois méthodes s'appuient pour la plupart sur des analyses par spectrométrie de masse. Si les méthodes d'analyses ciblées utilisent en majorité la spectrométrie de masse en tandem, la démocratisation des appareils de haute résolution, l'accessibilité à des bases de données et la création de logiciels adaptés ont rendu possible le développement des méthodes d'analyse de suspects et d'analyse non ciblée utilisant un schéma général constitués de plusieurs étapes et qui est aujourd'hui le schéma utilisé dans une grande partie des études. Les étapes de priorisation des données et d'identification sont les étapes les plus longues et sont sans garantie de succès. Plusieurs méthodes de priorisation existent aujourd'hui s'appuyant soit sur les données obtenues soit sur les expériences réalisées. *In fine*, des choix de réduction du volume de données sont communs à ces deux grands groupes. Aujourd'hui la pratique du « tout-en-un » se répand avec l'utilisation des appareils de spectrométrie de masse haute résolution les plus récents (Acena *et al.,* 2015 ; Kaserzon *et al.,* 2017 ; Schymanski *et al.,* 2017). La Figure 10 résume ces trois modes d'analyses.

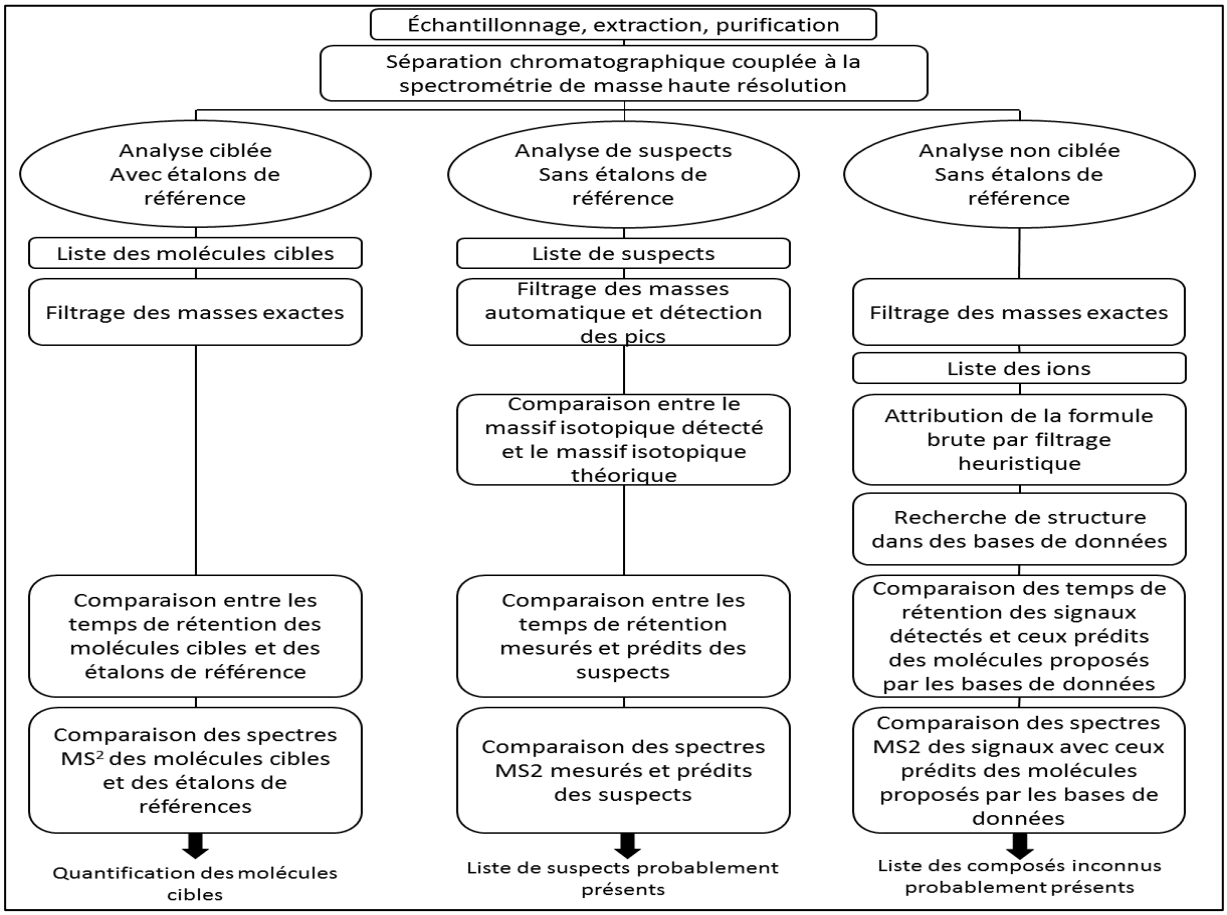


Figure 10 : Schéma des trois procédures d'analyses pour la réalisation de l'analyse ciblée, de l'analyse de suspects et de l'analyse non ciblée via la spectrométrie de masse haute résolution. Schéma traduit provenant de Krauss *et al*. (2010).

Chapitre 2 : Matériels et méthodes

I. Campagnes de prélèvements

I.1. Sites de prélèvements

Plusieurs sites de prélèvements d'origines variées ont été choisis pour la réalisation des campagnes d'échantillonnage : deux stations de traitements des eaux usées par boues activées et deux rivières.

I.1.1. Station de traitement des eaux usées de Poitiers

La station d'épuration des eaux (STEP) de Poitiers, mise en service en 2003, est basée sur la filière de traitement des eaux par boues activées à aération prolongée. D'une capacité nominale de 152 500 équivalents habitants, son débit entrant moyen est de 19 497 m^3 par jour et sa production de boues de 2 005 tMS par an (chiffres de 2018 du site internet sur l'assainissement du Ministère de la transition écologique). Cette station regroupe les eaux usées de 11 des 13 communes de Grand Poitiers. Les eaux usées entrantes proviennent d'un système principalement séparatif à l'exception notable du centre-ville de Poitiers.

Les eaux usées entrantes subissent tout d'abord un traitement primaire composé d'un dégrillage, d'un dessablage et d'un déshuilage. Vient ensuite le traitement secondaire biologique en bassins aérobie, anoxie et anaérobie pour l'élimination des matières organiques, de l'azote et des phosphates puis une clarification est effectuée avant le rejet des eaux traitées dans le Clain. Les eaux d'entrées sont séparées en deux filières ayant les mêmes étapes et les mêmes dimensions. Après traitement, les eaux sont rejetées dans une rivière, le Clain, coulant le long du site.

Trois points de prélèvements ont été choisis sur la STEP de Poitiers :

- Entrée des eaux après dégrillage, dans le bassin de déshuilage (PE),
- Sorties de traitement après clarification avec un point par canal de sortie correspondant aux deux voies de traitement (Nord et Sud : PSN et PSS) qui sont les 2 filières de traitement parallèles.

Ces points sont représentés en rouge sur la Figure 11 suivante.

Figure 11 : Image satellite de la station d'épuration de Poitiers avec représentation des points de prélèvements en rouge. Source : Google Maps.

I.1.2. Station de traitement des eaux usées de Saint Julien l'Ars

Cette station d'épuration est une station par boues activées d'une capacité nominale de 2 333 équivalents habitants, avec un débit entrant moyen de 265 m^3 par jour et une production de boues de 19 tMS par an (chiffres 2018 du site internet sur l'assainissement du ministère de la transition écologique).

Les eaux usées entrantes subissent tout d'abord un traitement primaire composé d'un dégrillage, d'un dessablage et d'un déshuilage. Un traitement secondaire biologique en bassins aérobie, anoxie et anaérobie puis une clarification sont effectués avant le rejet des eaux traitées dans une lagune adjacente.

Trois points de prélèvements ont été choisis sur ce site :

- L'entrée des eaux après dégrillage, dans le bassin de déshuilage (JE),
- En sortie de traitement après clarification (JS),
- Dans la lagune de stockage (JL).

La Figure 12 suivante montre l'image satellite du site de Saint Julien l'Ars.

Figure 12 : Image satellite de la station d'épuration de Saint Julien l'Ars avec représentation des points de prélèvements en rouge. Source : Google Maps.

I.1.3. Eaux de rivières

I.1.3.1. Le Clain

Le Clain est une rivière traversant Poitiers et dont la source se situe en Charente. Le Clain est un affluent de la Vienne. Son débit est compris entre environ 20 m^3/s à 2 m^3/s en période d'étiage, soit un débit moyen d'environ 1 800 000 m^3/j (allant de plus de 8 650 000 m^3/j à moins de 170 000 m^3/j). Des pointes de débit allant jusqu'à 100 m^3/s ont été relevées entre février et mars 2020.

Trois points de prélèvements ont été choisis sur cette rivière pour mesurer l'impact de la station d'épuration :

- En amont à 2 km de la station d'épuration de Poitiers (Cam),
- Dans le panache des eaux traitées rejetées par la STEP dans le Clain (CS),
- En aval immédiat à 30 m de ce panache (Cav).

Le point en amont de la STEP a pour objectif de dresser un « état initial » de l'eau avant rejet et le point du Clain en aval du rejet des eaux traitées permet de faire la comparaison d'une part avec l'amont du Clain et d'autre part avec le panache du Clain et des eaux traitées.

La Figure 13 montre les images satellites des sites de prélèvements sur le Clain.

Figure 13 : Images satellites des sites de prélèvements Clain aval, panache et amont représentés par des points de prélèvements en rouge. Source : Google Maps.

I.1.3.2. La Vienne

La Vienne est une rivière longue de 372 km qui traverse plusieurs départements dont la Haute-Vienne et la Vienne. Elle prend sa source en Corrèze et se jette dans la Loire à Candes-Saint-Martin. Sur la station de Lussac-les-Chateaux à 38 km de Poitiers, les débits relevés s'étendent d'environ 100 m^3/s de débit médian à 14 m^3/s en période d'étiage en 2020. Cela correspond à un débit médian de

8 640 000 m^3/j et à 1 209 600 m^3/j en étiage. Des pointes jusqu'à 400 m^3/s ont été relevées en période de crue entre février et avril 2020.

Le prélèvement dans la Vienne a été effectué dans la ville de Bonneuil-Matours située à 23 km en aval de Poitiers. La Figure 14 présente l'image satellite du site de prélèvement.

Figure 14 : Image satellite du site de prélèvements dans la Vienne (point de prélèvement en rouge). Source : Google Maps.

I.2. Campagnes d'échantillonnage et mode de prélèvement

I.2.1. Campagnes de prélèvements

Trois campagnes de prélèvement ont été réalisées entre l'année 2018 et l'année 2019. Le Tableau 4 résume les campagnes avec la date, les sites prélevés et les conditions météorologiques du jour de prélèvements.

Tableau 4 : Récapitulatif des échantillons prélevés et des dates de campagne de prélèvements *www.historique-meteo.net/france/poitou-charentes/poitiers/ (consulté le 14/10/2020).

Référence de la campagne	Date de la campagne	Site de Poitiers	Site de Saint Julien l'Ars	Site du Clain	Site de la Vienne	Conditions météorologiques précipitations*
Campagne A	25/06/2018	1 entrée, 2 sorties (Nord et Sud)	1 entrée, 2 sorties (sortie et lagune)	1 Amont, 1 Panache, 1 Aval	1 échantillon	Beau temps Absence depuis 2 jours
Campagne B	12/11/2018	1 entrée, 2 sorties (Nord et Sud)	1 entrée, 2 sorties (sortie et lagune)	1 Amont, 1 Panache, 1 Aval	1 échantillon	Pluie modérée 11,1 mm les 11 et 12/11/2018
Campagne C	20/05/2019	3 entrées, 3 sorties (Nord)	Pas de prélèvements	3 Amont, 3 Panache, 3 Aval	Pas de prélèvement	Plutôt dégagé 12,0 mm les 19 et 20/05/2019

I.2.2. Mode de prélèvement

Les différents échantillons ont été récupérés par une méthode de prélèvement dite ponctuelle. Pour les eaux usées et de rivières, les échantillons ont ainsi été prélevés à l'aide d'une perche télescopique équipée d'un bécher en plastique rincé au préalable avec l'eau du milieu échantillonné. Les échantillons ont été récupérés dans des bouteilles en verre d'un volume de 500 mL ayant été passées au four (500°C pendant 4 h) et rincées au moment du prélèvement avec la matrice échantillonnée. Après prélèvement, les échantillons ont été stockés à 4°C jusqu'au protocole de préparation.

II. Réactifs et étalons

II.1. Réactifs

Le méthanol (MeOH) et l'acide formique (AF) de qualité UHPLC-MS ont été fournis par Carlo Erba Reagents (Val de Reuil, France). L'ammoniaque (NH_4OH) 25% de qualité UHPLC-MS LiChropur® a été acheté chez Merck Millipore (Lyon, France). Les composés de référence ont été fournis par la plateforme Sigma-Aldrich de Merck (Saint-Quentin-Fallavier, France) excepté pour l'oxybenzone

acheté chez Alfa Aesar (Thermo Fisher Scientific, Les Ulis, France). Les étalons marqués ont été achetés chez Cluzeau Info Labo (Sainte-Foy-la-Grande, France). L'eau ultra-pure (EUP) a été produite (18,2 MΩ.cm, COD < 0,1 mg/L) par un système de purification PURELAB Flex 2 (ELGA, Veolia) puis un système LC-Pak, Milli-Q IQ 7000 (Merk Millipore, Molsheim, France).

II.2. Solutions étalons

Un ensemble d'étalons a été utilisé dans différents tests et expériences au cours des travaux de thèse. Au total 24 étalons et 5 étalons marqués ont été employés. Le Tableau 5 répertorie ces étalons avec leur formule brute, leur masse exacte, leur numéro CAS et leur catégorie de produit.

Tableau 5 : Liste des étalons utilisés dans le cadre des optimisations de traitement de données, de calculs d'effet matrice et des calculs de rendements d'extraction.

Identité	Formule brute	Masse exacte (g/mol)	Numéro CAS	Catégorie
2-hydroxyatrazine	$C_8H_{15}N_5O$	197,1277	2163-68-0	Métabolite de l'atrazine
Acétochlore	$C_{14}H_{20}ClNO_2$	269,1183	34256-82-1	Herbicide
Alachlore	$C_{14}H_{20}ClNO_2$	269,1183	15972-60-8	Herbicide
Atrazine	$C_8H_{14}ClN_5$	215,0938	1912-24-9	Herbicide
Caféine	$C_8H_{10}N_4O_2$	194,0804	58-08-2	Produit de consommation
Carbamazépine	$C_{15}H_{12}N2O$	236,0950	298-46-4	Antiépileptique
DEET (N,N-Diéthyl-m-toluamide)	$C_{12}H_{17}NO$	191,1310	134-62-3	Insectifuge
Diclofénac	$C_{14}H_{11}Cl_2NO_2$	295,0167	15307-86-5	Anti-inflammatoire
Flunixine	$C_{14}H_{11}F_3N_2O_2$	296,0773	38677-85-9	Anti-inflammatoire vétérinaire
Gemfibrozil	$C_{15}H_{22}O_3$	250,1569	25812-30-0	Hypolipémiant Fibrate
Iopromide	$C_{18}H_{24}I_3N_3O_8$	790,8697	73334-07-3	Agent de contraste
Isoproturon	$C_{12}H_{18}N_2O$	206,1419	34123-59-6	Herbicide
Métazachlore	$C_{14}H_{16}ClN_3O$	277,0982	67129-08-2	Herbicide
Metformine	$C_4H_{11}N_5$	129,1015	1115-70-4	Antidiabétique

Identité	Formule brute	Masse exacte (Da)	Numéro CAS	Catégorie
Methylparaben	$C_8H_8O_3$	152,0473	99-73-3	Conservateur
Métolachlore	$C_{15}H_{22}ClNO_2$	283,1339	51218-45-2	Herbicide
Naproxen	$C_{14}H_{14}O_3$	230,0943	22204-53-1	Anti-inflammatoire
Oxadiazon	$C_{15}H_{18}Cl_2N_2O_3$	344,0695	19666-30-9	Herbicide
Oxybenzone	$C_{14}H_{12}O_3$	228,0786	131-57-7	Filtre UV
Paracétamol	$C_8H_9NO_2$	151,0633	103-90-2	Antalgique Antipyrétique
Propranolol	$C_{16}H_{21}NO_2$	259,1572	318-98-9	Bêta-bloquant
Saflufenacil	$C_{17}H_{17}F_4N_4O_5S$	500,0544	372137-35-4	Herbicide
Sulcotrione	$C_{14}H_{13}ClO_5S$	328,0172	99105-77-8	Herbicide
Tembotrione	$C_{17}H_{16}ClF_3O_6S$	440,0308	335104-84-2	Herbicide
2-hydroxyatrazine-d5	$C_8D_6H_{16}ClNO_2$	289,1716	1276197-25-1	Étalon marqué
Atrazine-d5	$C_8H_9D_5ClN_5$	220,1252	163165-75-1	Étalon marqué
Caféine-d9	C8D9HN4O2	203,1369	72238-85-8	Étalon marqué
Carbamazépine-d10	$C_{15}D_{10}H_2O$	246,1577	132183-78-9	Étalon marqué
Métolachlore-d6	$C_8D_5H_{10}N_5O$	214,1047	1219803-97-0	Étalon marqué

Pour les composés se présentant sous forme solide, des solutions mères individuelles d'étalon ont été préparées par la dissolution d'une masse précise de produit dans un volume de 50 mL d'eau ultra-pure, de méthanol ou d'un mélange EUP/MeOH (80/20, v/v) en fonction de leur solubilité. Dans ces conditions, les concentrations obtenues ont alors été calculées et se sont échelonnées de 100 mg/L à 736 mg/L. Des aliquots de ces solutions mères ont été réalisés et conservés en partie à 4°C et en partie à -20°C pour une conservation de plus longue durée. Pour ces composés, des solutions filles à 1 mg/L ont également été réalisées pour chaque étalon par dilution dans de l'eau ultra-pure ou dans le méthanol et conservées à 4°C.

Pour les composés achetés sous forme de solutions concentrées,

- des solutions à une concentration de 1 mg/L ont directement été préparées par dilution pour les étalons marqués de carbamazépine-d10 et de caféine-d9,

- et des solutions à des concentrations de 0,04 g/L, 0,02 g/L et 0,04 g/L ont été préparées respectivement pour les étalons marqués d'atrazine-d5, de 2-hydroxyatrazine-d5, et de métholachlore-d6.

Les solutions de contrôles qualités contenant un mélange d'étalons et servant à contrôler l'ionisation au cours des séquences d'analyses ont été préparées à une concentration de 100 µg/L à partir des solutions mères par dilution des étalons.

Enfin, pour le dopage des échantillons en étalons à 0,5 µg/L, dans le cadre des expériences destinées au calcul des effets matrices et des rendements d'extraction sur phase solide (SPE), une solution contenant un mélange d'étalons à 1 mg/L a été préparée par dilution des solutions mères lorsque leurs concentrations étaient suffisantes.

III. Préparation d'échantillons

Tous les échantillons ont subi une phase de filtration à laquelle s'est ajoutée une étape de centrifugation dans le cas des échantillons d'entrée de STEP. A ces étapes, un protocole d'extraction/concentration des échantillons a également été étudié et systématiquement réalisé. La vue d'ensemble du protocole de préparation des échantillons est présentée dans la Figure 15. Lors de la préparation des échantillons, un blanc de préparation subissant l'ensemble des étapes de préparations a été systématiquement inclus avec les échantillons analysés.

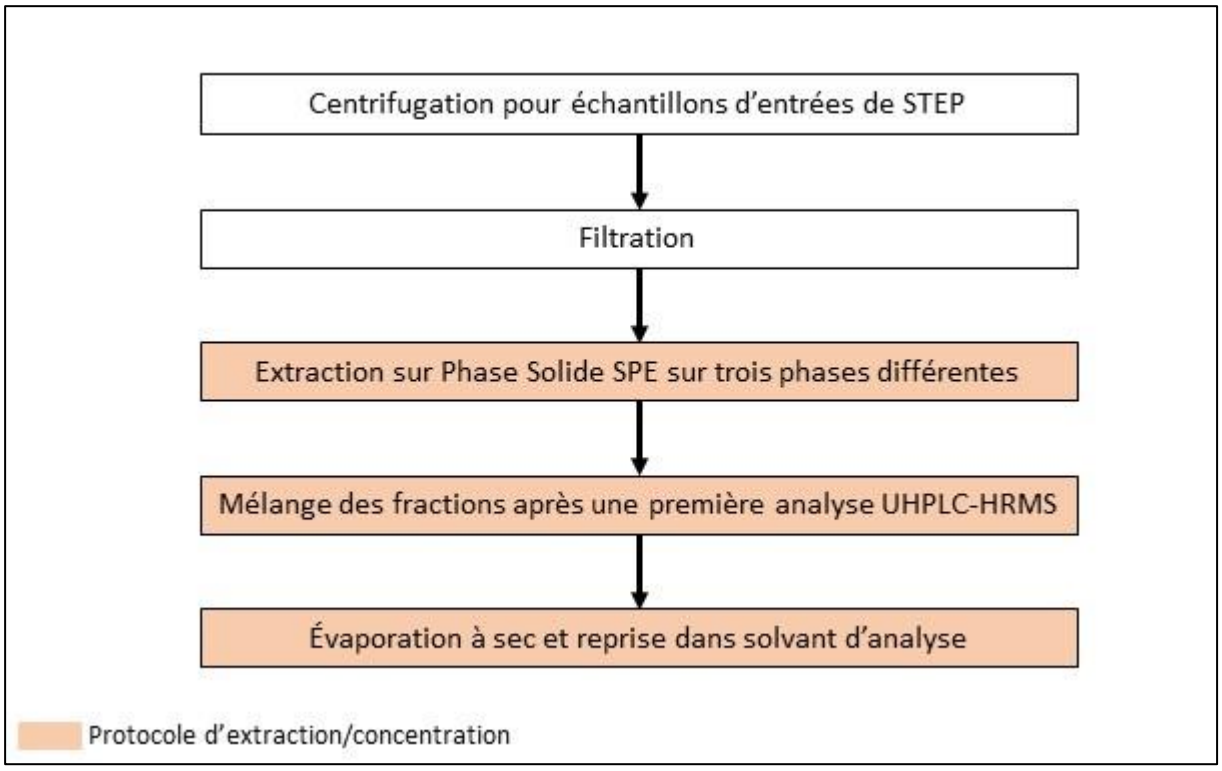


Figure 15 : Protocole de la préparation des échantillons pour l'application de l'analyse non ciblée.

Lors de l'application de ce protocole, la filtration a été réalisée dès le jour du prélèvement et l'ensemble de l'extraction sur phase solide a été réalisée dans les jours suivants. Après préparation, les échantillons ont été stockés à 4°C lorsque l'analyse pouvait être effectuée rapidement (i.e. dans les 48 h) et à -20°C dans le cas contraire jusqu'à l'analyse.

III.1. Centrifugation et filtration

III.1.1. Centrifugation

Les échantillons d'entrées de STEP ont été centrifugés en raison d'une phase particulaire très importante bouchant très rapidement les filtres utilisés lors de la filtration. La centrifugation a été réalisée par une centrifugeuse MF 20-R (Awel International, France) pendant 10 min à 10 000 tr/min.

III.1.2. Filtration

Un volume total de 250 mL de chaque échantillon a été filtré sur Büchner avec un filtre en microfibre de verre GF/F de 0,7 µm (Whatman). Le système de filtration a été rincé trois fois à l'eau osmosée (≈1 MΩ.cm, COD < 0,3 mg/L) entre chaque échantillon. Après filtration, 200 mL ont été conservés pour l'extraction sur phase solide, et 50 mL, pour les analyses usuelles et l'analyse des échantillons non concentrés par chromatographie liquide couplée à la spectrométrie de masse haute résolution.

III.2. Extraction sur phase solide

III.2.1. Protocole

L'extraction sur phase solide (SPE) a été réalisée par un système automatique Rapidtrace®+ (Biotage, Uppsala, Suède). Afin de purifier et concentrer les échantillons avant analyse par spectrométrie de masse, trois cartouches différentes provenant de chez Waters (Guyancourt, France) ont été utilisées dans l'ordre suivant : une phase échangeuse d'anions OASIS MAX® (6 cc, 150 mg), une phase échangeuse de cations OASIS MCX® (6cc, 150 mg) et une phase OASIS HLB® (6cc, 200 mg). Cette méthode a été développée à partir de celle appliquée par Deeb *et al.* (2016) et Stephan *et al.* (2016). Dans cette méthode, la cartouche Oasis HLB a été ajoutée dans le but de retenir les composés non retenus par les cartouches précédentes.

Les structures chimiques des phases solides utilisées sont représentées dans la Figure 16. L'utilisation des trois phases différentes permet de pouvoir effectuer un fractionnement en fonction des propriétés acide, basique ou neutre des analytes. Cette méthode a été choisie pour concentrer les échantillons afin d'augmenter la sensibilité et également retenir le maximum des composés pour rester le plus exhaustif possible.

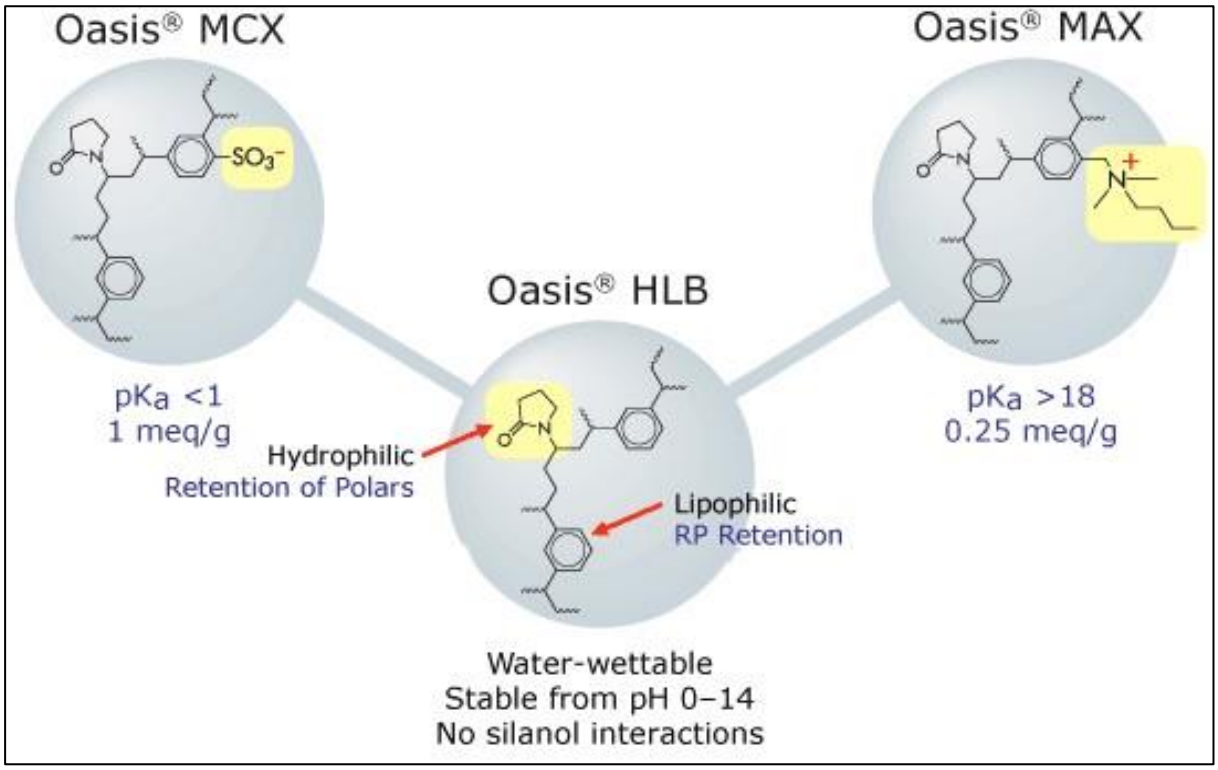


Figure 16 : Schéma modifié des structures chimiques des phases solides des cartouches Oasis utilisées en SPE. Source : Waters.

Lors de la réalisation de ce protocole, les étapes d'extractions suivantes ont été réalisées (Tableau 6) :

Tableau 6 : Etapes de la méthode d'extraction sur phase solide à trois cartouches.

Étapes	Oasis MAX®	Oasis MCX®	Oasis HLB®
Conditionnement	2x3 mL MeOH 2x3 mL EUP	2x3 mL MeOH 2x3 mL EUP	6 mL MeOH 6 mL EUP
Dépôt	200 mL	200 mL	200 mL
Séchage	15 min	15 min	Non effectué
Lavage	2 mL EUP/NH_4OH(95/5, v/v)	2 mL EUP/AF(98/2, v/v)	Non effectué
Elution	6 mL MeOH/Acétate d'éthyl/AF (69/29/2, v/v/v)	6 mL MeOH/Acétate d'éthyl/NH_4OH (67,5/27,5/5, v/v/v)	6 mL MeOH

Les fractions éluées ont été récupérées et le solvant a été évaporé à sec sous flux d'azote à une température contrôlée de 35°C. Les résidus obtenus ont été repris dans 533 µL d'un mélange EUP/MEOH + 0,1% d'acide formique 90/10 v/v (dopé par les étalons internes pour la campagne C). Puis, après une première analyse par UHPLC-HRMS, les fractions ont été mélangées ensemble pour

une seconde analyse. Dans ces conditions, un facteur de concentration de 125 a alors été appliqué sur l'ensemble des campagnes.

III.2.2. Estimation des effets matrices

Pour calculer les effets matrices lors des analyses UHPLC-HRMS, des échantillons de différentes matrices EUP, Clain amont et aval, entrée et sortie de la STEP de Poitiers ont été dopés par le mélange de 24 étalons à une concentration de 100 µg/L après extraction. Le calcul des effets matrices a été réalisé selon la formule suivante :

$$\%EM = \frac{Aire\ (matrice)}{Aire\ (EUP)} \times 100 \quad (2)$$

Avec EM : Effet matrice, Aire (matrice) : Aire de l'intégration de l'étalon dans le chromatogramme pour la matrice, Aire (EUP) : Aire de l'étalon dans l'eau ultra-pure.

Les ratios des aires des étalons entre les matrices Clain amont-aval et entrée-sortie STEP ont également été calculés.

Ainsi, un facteur de correction des effets matrices d'une valeur de 2 déterminé dans le cadre de ce travail (IV.2.1. du chapitre 3 « Mise au point de l'approche ») a été appliqué sur les signaux d'entrées en communs avec les signaux de sorties.

III.2.3. Estimation des rendements d'extraction de la méthode SPE

Les rendements de la méthode SPE ont été estimés par l'extraction d'échantillons des matrice EUP, Clain amont, Clain aval, entrée et sortie de la STEP de Poitiers, dopés par le mélange d'étalons à une concentration finale de 500 ng/L dans les échantillons avant extraction. Après extraction et analyse les rendements ont été calculés selon la formule suivante :

$$\%R = \left(\frac{Aire\ étalon\ (matrice)}{Aire\ étalon\ marqué\ (matrice)} \times \frac{Aire\ étalon\ marqué\ (référence)}{Aire\ étalon\ (référence)} \right) \times 100 \quad (3)$$

Avec %R : rendement en pourcentage, Aire étalon (matrice) : aire de l'étalon dans l'échantillon après extraction, Aire étalon marqué (matrice) : aire de l'étalon marqué dans l'échantillon après extraction, Aire étalon marqué (référence) : aire de l'étalon en solution dans l'eau ultra-pure faisant référence à un rendement de 100% et Aire étalon (référence) : aire de l'étalon en solution dans l'eau ultra-pure, correspondant à un rendement de 100%.

Ces rendements ont été calculés individuellement pour 5 étalons marqués différents (2-hydroxyatrazine d5, atrazine d5, caféine d9, carbamazépine d10 et métolachlore d6) ajoutés dans le

solvant de reprise des échantillons après évaporation à sec. Les rendements calculés ont ensuite été moyennés et les écart-types ont été calculés.

III.3. Gestion de la contamination

La mise en place de mesures pour limiter la contamination des échantillons par l'environnement de travail, le matériel et entre les échantillons, a été nécessaire pour restreindre le nombre de faux positifs et augmenter la qualité des méthodes, du prélèvement à l'analyse chromatographique. Ainsi, toute la verrerie utilisée a été passée, après rinçage préalable à l'eau osmosée, au four avec un cycle de cinq heures jusqu'à une température de 500 °C. Des blancs ont été réalisés au niveau de la préparation d'échantillons et de l'analyse par chromatographie. Ces blancs ont été injectés dans les séquences d'analyses et pris en compte dans les moyennes de blancs lors du retraitement des données.

IV. Méthodes d'analyse

IV.1. Analyse des paramètres usuels

Cinq paramètres usuels ont été mesurés sur les eaux après filtration : le pH, la conductivité, l'absorbance à 254 nm, le carbone organique dissous (COD) et l'azote total.

Le pH a été mesuré par une électrode de verre combinée (KCl) avec un pH-mètre PHM240 MeterLab de Radiometer Analytical (Lyon, France). Le pH-mètre a été étalonné avec des solutions tampons de pH 7 et 10 avant chaque série de mesures.

L'absorbance à 254 nm a été mesurée par un spectrophotomètre UV-visible de la gamme 50 UV-Vis Varian Cary. Le zéro a été effectué avec un blanc d'eau ultrapure avant chaque série de mesures. Les mesures ont été réalisées dans une cuve en quartz de 1 cm.

La conductivité a été mesurée par un conductimètre Cyberscan CON 510 d'Eutech Instruments (Thermo Fisher Scientific, Les Ulis, France) étalonné avant chaque série de mesures par une solution étalon.

Le carbone organique dissous et l'azote total ont été mesurés par un analyseur de carbone organique total (TOC-VCPN) de la société Shimadzu équipé d'un passeur d'échantillon ASI-V. L'analyseur a été étalonné par deux courbes de calibration distinctes correspondant au carbone et à l'azote, allant de 10 à 100 mg/L dans les deux cas. Les échantillons issus de l'entrée de la station d'épuration de Poitiers et de Saint Julien l'Ars ont été dilués par 2 afin d'être dans la gamme de mesure.

IV.2. Analyse par chromatographie liquide couplée à la spectrométrie de masse haute résolution

IV.2.1. Appareillage

Les analyses ont été réalisées par chromatographie liquide ultra haute performance couplée à la spectrométrie de masse haute résolution (UHPLC-HRMS). Pour cela, un système de séparation chromatographique constitué d'une chaine UHPLC Accela 1250, d'un passeur d'échantillons automatique (Accela open auto sampler, Thermo Fisher Scientific, Les Ulis, France) et d'un four pour colonne (Column Oven 300) a été utilisé. Un spectromètre de masse couplé à l'UHPLC de type trappe orbital de la gamme Q-Exactive® (Thermo Fisher Scientific, les Ulis, France) avec une source HESI-II (Heated Electrospray Ionisation-II) a été employé. Lors des analyses, les échantillons ont été placés dans le passeur thermostaté à 4°C jusqu'à l'analyse. L'appareillage d'analyse a été piloté par le logiciel Xcalibur version 3.0.63 (Thermo Fisher Scientific, Les Ulis, France).

IV.2.2. Conditions Chromatographiques

La colonne utilisée pour les analyses a été une colonne Acquity UPLC HSS T3 (2,1 mm x 100 mm, 1,8 µm de taille de particules, Waters). Un support de pré-filtre (2,1 mm de diamètre avec une porosité de 0,2 µm) a été monté sur la colonne pour éviter la contamination et l'obstruction de la colonne. La température du four colonne a été réglée à 30 °C et le débit de phase mobile à 400 µL/min. Le volume d'échantillon injecté a été défini à 20 µL. Concernant la composition de la phase mobile, un mélange d'eau ultra-pure (solvant A) et de méthanol (solvant B) avec ajout de 0,1% d'acide formique dans les deux solvants a été employé. Le Tableau 7 présente le gradient de solvant utilisé pour la séparation chromatographique. Dans ces conditions, le temps d'analyse total s'est élevé à 18 min par injection.

Tableau 7 : Paramètres de la chromatographie en phase liquide.

Temps (min)	EUP + 0,1 AF A (%)	MeOH + 0,1 AF B (%)
0	90	10
1	90	10
6	30	70
8	0	100
10	0	100
13	90	10
18	90	10

IV.2.3. Paramètres de spectrométrie de masse

Les analyses ont été effectuées en mode positif et en mode négatif séparément à l'aide d'une source d'ionisation de type HESI-II. L'étalonnage externe du spectromètre de masse avant chaque analyse a été réalisé en utilisant des solutions Pierce™ LTQ ESI Positive Ion et Pierce™ ESI Negative Ion de chez Thermo Fisher Scientific (Les Ulis, France). Dans ces conditions, la précision en masse après calibration est inférieure à 2 ppm pour un rapport m/z 200.

Les analyses par spectrométrie de masse ont été effectuées en mode Full Scan avec une résolution de 70 000. Les gammes de m/z analysées se sont échelonnées de 100 à 1500. Les débits de gaz de nébulisation et de gaz auxiliaire ont été fixés à 40 et 20 unités arbitraires respectivement. Les températures du capillaire de transfert et de la source ont été réglées à 350 °C et 400 °C respectivement. Le voltage du spray, l'AGC target (Auto Gain Control, nombre d'ions à accumuler) et le temps d'injection maximum des ions ont été paramétrés à +4 kV, $3x10^6$ ($1x10^6$ en négatif) et 200 ms respectivement. Le réglage des lentilles, paramètre S-lens a été fixé à 100 V.

Pour certains échantillons, des analyses MS^2 ont été également réalisées. Pour cela, le mode Full MS-ddMS2 sur les 5 ions les plus intenses a été utilisé avec une résolution de 17 500. La fenêtre d'isolation des ions a été réglée à 1,5 m/z et le temps d'exclusion dynamique a été fixé à 10 s. Lors de ces analyses, trois énergies de collision normalisées (NCE) ont été employées : 15, 30 et 50V.

IV.2.4. Séquence d'analyse

En début de séquence, des blancs puis des contrôles qualités ont été injectés pour équilibrer et vérifier le bon état initial du système. De même, ces blancs et ces contrôles qualités ont été régulièrement injectés lors des séquences afin d'évaluer l'état du système analytique. Concernant les échantillons, l'ordre d'injection a été défini en fonction de la complexité de la matrice : du moins complexe au plus complexe (i.e en commençant par les eaux de surface et en terminant par les entrées de STEP). Chaque échantillon a été analysé trois fois. Les analyses en mode positif ont été réalisées en premier suivi de celles réalisées en mode négatif.

Dans le cadre de ce travail, seuls les résultats obtenus en mode positif ont été exploités.

V. Traitement des données

L'utilisation de l'analyse non ciblée implique la détection de très nombreux signaux comparativement à l'analyse ciblée et il est nécessaire de traiter les données afin d'obtenir la liste des signaux présents dans les échantillons par des logiciels spécifiques à ce type d'approche. Pour cela, de nombreux

logiciels libres ou commerciaux sont disponibles pour traiter automatiquement les données brutes obtenues après analyse.

V.1. Prétraitement des données brutes

Dans le cadre de ce travail, deux méthodes utilisant des ensembles de logiciels ont été étudiées : une première méthode utilisant les logiciels MZmine 2/R/Excel et une seconde méthode utilisant les logiciels Compound Discoverer/Excel.

Les Tableaux 8 et 9 présentent les paramètres utilisés pour les différentes étapes de ces méthodes. L'optimisation de ces paramètres a fait l'objet de ce travail de thèse et est détaillé dans le chapitre 3.

V.1.1. Méthode par la combinaison des logiciels MZmine 2, R et Excel

Lors de cette approche, le traitement des données a été réalisé avec l'ensemble MZmine 2, Excel et R.

MZmine 2 est un logiciel libre basé sur la boîte à outils MZmine initial qui vise à traiter, visualiser et analyser les données obtenues par couplage chromatographie-spectrométrie de masse (Katajamaa *et al*., 2006 ; Pluskal *et al*., 2010). MZmine 2 comporte un ensemble de modules pour réaliser les étapes principales du traitement des données ainsi que des étapes facultatives complémentaires.

L'utilisation d'Excel a permis des opérations de calculs entre les signaux dans le tableau de données de façon automatique grâce à l'utilisation de macro.

Enfin, le logiciel R au travers de l'interface plus simple d'utilisation R.Studio a été employé au cours de cette méthode. Ce logiciel est un logiciel libre conçu pour la réalisation de statistiques et pour l'analyse de données. Il fonctionne grâce à des « packages » exécutant des tâches codées par un développeur. Des scripts réalisant une suite d'opérations peuvent aussi être codés sur R.

Lors de cette approche, dans un premier temps, les données brutes obtenues par analyse UHPLC-HRMS au format raw ont été converties en mode dit « centroid » et au format mzXML par l'outil MSConvert de ProteoWizard 3.0 (Kessner *et al.*, 2008) pour limiter la taille des données et faciliter leur prise en charge par les logiciels utilisés.

A partir des résultats obtenus les étapes principales effectuées sur MZmine 2 ont été les suivantes :

- Détection des m/z significatifs (dont l'intensité est supérieure à un seuil) à partir des données brutes par un algorithme : le seuil et l'algorithme sont définis par l'utilisateur ;
- Reconstruction de chromatogrammes pour chaque m/z détecté lors de l'étape précédente ; chaque chromatogramme représente la détection d'un ou plusieurs ions en fonction du temps pour chaque masse détectée dans l'étape précédente ;

- Déconvolution des chromatogrammes en signaux individuels par un algorithme pour obtenir une liste de signaux par échantillon définis par un m/z, un temps de rétention et une intensité unique ;
- Alignements des signaux à travers l'ensemble des échantillons ;
- « Gap-filling » (recherche des signaux manquants) ;
- Export des données.

Des étapes facultatives de lissage des chromatogrammes reconstruits, de filtrage des signaux ou encore de mise en correspondance des signaux détectés avec des bases de données sont également possibles avec MZmine 2. Chacune de ces étapes comporte un certain nombre de paramètres à régler dont l'impact sur les résultats varie. Il s'agit principalement des tolérances sur la masse et le temps de rétention ainsi que des seuils d'intensité minimales.

À la suite de l'utilisation de MZmine 2, Excel a servi à prendre en compte les blancs analysés et éliminer ainsi les signaux dont l'intensité n'est pas significativement différente à celle mesurée dans le blanc. Pour cela, les blancs ont été regroupés et moyennés par campagne d'analyse. Cette étape est l'un des moyens de réduire le nombre de faux positifs.

Enfin, R a été utilisé avec le package « non-target » qui sert à détecter les adduits et les isotopes dans le tableau de signaux pour pouvoir les regrouper et diminuer le nombre de signaux redondants.

Les étapes avec les différents paramètres et leurs réglages sont résumés dans le Tableau 8.

Tableau 8 : Synthèse des étapes, des paramètres et de leurs valeurs dans le cadre de l'utilisation de l'ensemble de logiciels MZmine 2, R 3.5.1 et Excel.

Logiciel	Étapes	Paramètres
MZmine 2	Détection des signaux	Algorithme centroid Seuil d'intensité : $1x10^4$
	Reconstruction des chromatogrammes	Algorithme ADAP Intensité minimale : $1x10^4$ Taille minimum du groupe (nombre de scans) : 5 Seuil d'intensité minimale contribuant au comptage dans les scans : $1x10^4$ Tolérance maximale sur m/z : 5 ppm

Logiciel	Étapes	Paramètres
MZmine 2	Déconvolution	Algorithme ADAP Seuil Signal / Bruit : 5 Intensité minimale du signal : $1x10^4$ Seuil Coefficient /Aire : 130 Gamme de durée des pics : 0,05-2 min Gamme du pas de temps : 0,05-0,10 min
	Filtrage des pics	Nombre de points minimum : 20 points Largeur de pic à mi-hauteur comprise entre 0,05 et 0,9 min Temps de rétention compris entre 0,5 et 12,3 min
	Normalisation des signaux	Tolérance sur le m/z : 5 ppm Tolérance sur le temps de rétention : 0,2 min Intensité minimum des pics communs à tous les échantillons : $1x10^7$
	Alignements des signaux	Algorithme « Join aligner » Tolérance sur le m/z : 5 ppm Tolérance sur le temps de rétention : 0,4 min
	Gap-filling	Algorithme « Peak finder (multitreaded) » Tolérance sur le m/z : 5 ppm Tolérance sur le temps de rétention : 0,2 min
	Eliminations des doublons	Tolérance sur le m/z : 5 ppm Tolérance sur le temps de rétention : 0,2 min
Excel	Blancs	Elimination des signaux dont l'intensité est inférieure à 5 fois l'intensité du blanc Correction de l'intensité des signaux par la soustraction de l'intensité du blanc Elimination des signaux dont le coefficient de variation est supérieur à 30% sur les trois réplicas
R	Regroupements des adduits	Adduits en mode positif : $[M+H]^+$ $[M+Na]^+$, $[M+NH_4]^+$, $[M+CH_3OH+H]^+$ et $[M+K]^+$ Adduits en mode négatif : $[M-H]^-$, $[M-CHOO]^-$, $[M+FA-H]^-$, $[M-Cl]^-$
	Regroupements des isotopes	Isotopes sélectionnés : ^{13}C, ^{14}N, ^{18}O, ^{37}Cl, ^{81}Br, ^{34}S

Après traitement des données, la liste finale des signaux a par la suite de nouveau été exportée dans Excel sous la forme d'un tableau dont les colonnes sont les échantillons avec leurs réplicas, les lignes sont les couples m/z et temps de rétention des signaux et dont les entrées sont les intensités de ces signaux. Ces différentes étapes ont pour but d'éliminer le bruit et les pics artéfacts, ainsi que de conserver les pics réellement présents dans les chromatogrammes d'analyse des échantillons.

V.1.2. Méthode par la combinaison des logiciels Compound Discoverer et Excel

Le logiciel Compound Discoverer est un logiciel de prétraitement des données développé par Thermo Scientific conçu pour traiter les données brutes obtenues sur les instruments du constructeur. Le logiciel se base sur l'utilisation d'un « workflow » ou bloc d'étapes tel que montré dans la Figure 17.

Lors de l'analyse non ciblée, un premier bloc d'étapes (orange dans la Figure 17) permet de sélectionner les spectres de masses servant de base de travail puis d'aligner l'ensemble des données des différents échantillons considérés.

Une seconde série d'étapes (violet dans la Figure 17) détecte les signaux, établit un tableau regroupant l'ensemble des pics détectés à travers tous les échantillons, recherche les signaux manquants (étape de « gap-filling ») puis compare les signaux obtenus avec ceux du blanc.

Un troisième bloc d'étapes (jaune dans la Figure 17) a pour objectif de tenter d'identifier les signaux à partir de bases de données ou de listes de suspects.

Lors de ces étapes, tous les composés obtenus peuvent aussi faire l'objet d'un calcul de prédiction de formule brute dont les profils isotopiques théoriques sont calculés et comparés avec les profils réels des spectres de masse. Dans ces conditions, les tentatives d'identification sont ensuite classées selon un score.

Enfin, un dernier bloc d'étapes peut également être ajouté pour réaliser des calculs de statistiques descriptives (non présenté dans la Figure 17).

De même, à cet arbre d'étapes correspondant à l'analyse non ciblée peuvent être ajoutées des étapes correspondant à de l'analyse ciblée à partir d'une liste de composés entrée dans le logiciel. Dans ces conditions, les différents types d'analyses peuvent être réalisés dans le même temps.

La Figure 17 montre le schéma d'étapes utilisé dans cette thèse.

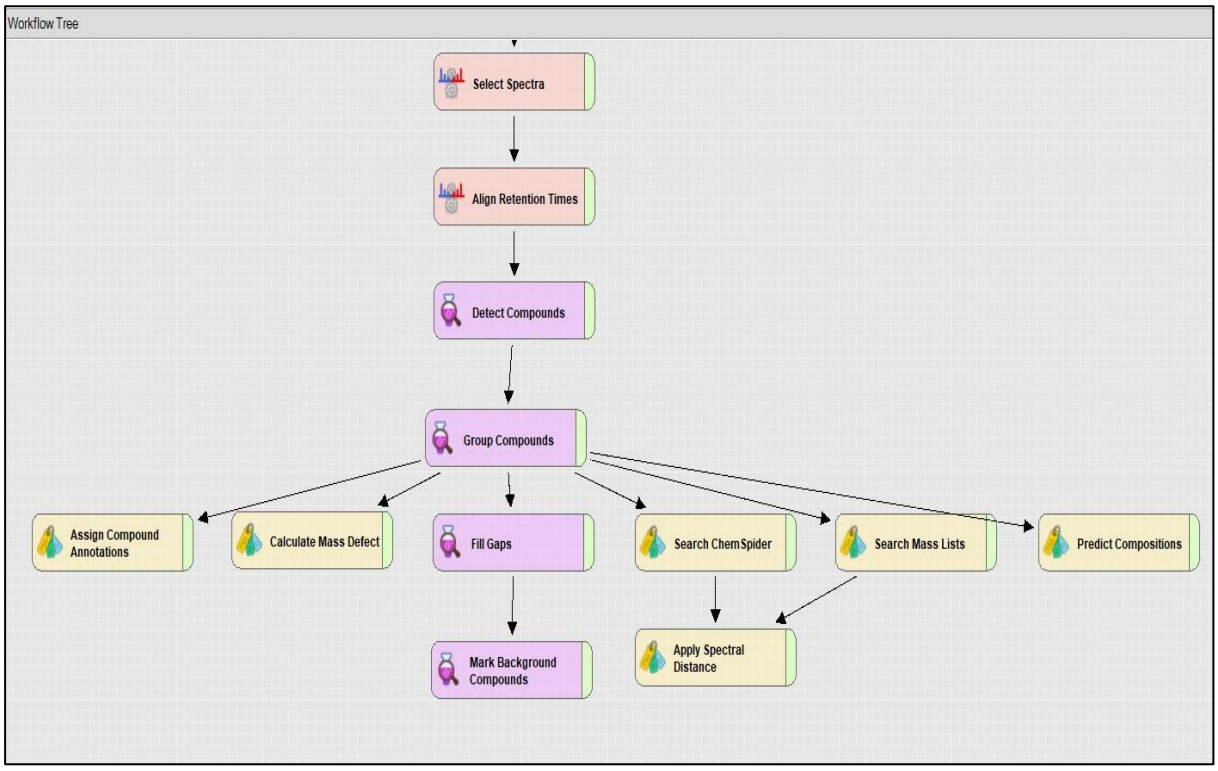


Figure 17 : Capture d'écran du workflow utilisé pour le traitement des données issues de l'analyse UHPLC-HRMS.

Après prétraitement avec le logiciel Compound Discoverer, le tableau des données regroupe les signaux détectés à travers tous les échantillons avec une masse moléculaire, un temps de rétention et une aire par échantillon. Ce tableau est transféré dans Excel pour la prise en compte des blancs. Les blancs sont regroupés et moyennés par campagne d'analyse. Une normalisation-quantile est effectuée par type de matrice afin de corriger les variations dues à l'instrumentation. Les signaux sont ensuite conservés ou éliminés par comparaison avec les intensités des signaux des blancs puis les valeurs des aires du blanc sont soustraites aux aires des signaux. Le Tableau 9 présente les étapes réalisées avec leurs paramètres et réglages. Lorsque les réglages n'ont pas été précisés c'est que la valeur par défaut proposée par le logiciel a été conservée.

Tableau 9 : Synthèse des étapes, des paramètres et de leurs valeurs pour le prétraitement des données par Compound Discoverer-Excel.

Logiciels	Étapes	Paramètres
Compound Discoverer	Sélection des spectres	Conservation des spectres avec des m/z de 100 à 1500
	Alignements des temps de rétention	Tolérance sur le m/z : 5 ppm Tolérance sur le temps de rétention : 1 min

Logiciels	Étapes	Paramètres
Compound Discoverer	Détection des signaux	Intensité minimale : 7,5 x 10^4 Largeur de pic maximal : 0,8 min Enlever les singlets : True Nombre de scans minimal par pic : 8 Seuil Signal/Bruit : 5 Tolérance sur m/z : 5 ppm Isotopes : 100C, 200H, 10Br, 10Cl, 30N, 30O, 3P, 10S Adduits en mode positif : $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M+CH_3OH+H]^+$, $[M+K]^+$, $[2M+H]^+$, $[2M+K]^+$, $[2M+Na]^+$, $[M+2H]^{2+}$, $[M+3H]^{3+}$, $[M+H+K]^{2+}$, $[M+H+Na]^{2+}$, $[M+H+NH4]^{2+}$, $[M+H-H_2O]^+$ Adduits en mode négatif : $[2M+FA-H]^-$, $[M-Cl]^-$, $[M+FA-H]^-$, $[M-2H]^{2-}$, $[M-H]^-$, $[M-H+HAc]^-$, $[M-H-H_2O]^-$.
	Groupement des signaux	Tolérance sur le m/z : 10 ppm Tolérance sur le temps de rétention : 0,4 min Adduit de base : mode positif $[M+H]^+$, mode négatif $[M-H]^-$
	Gap-filling	Tolérance sur le m/z : 5 ppm Rapport signal sur bruit > 5
	Blancs	Ratio aire du signal sur aire du blanc > 5
	Filtres supplémentaires	Temps de rétention : 0,5 – 12 min Background : False Filtres dépendants des matrices étudiés : voir Annexe 1
Excel	Blancs	Normalisation des aires par matrice par la méthode des quantiles Elimination des signaux dont l'aire est inférieure à 5 fois celle du blanc Correction de l'aire des signaux par la soustraction de l'intensité du blanc Elimination des signaux dont le coefficient de variation est supérieur à 30% sur les trois réplicas Application d'un facteur correction de 2 dans les échantillons d'entrées sur les signaux communs entre entrées et sorties

Après traitement des données, un tableau est obtenu dont les colonnes sont les échantillons avec leurs réplicas, les lignes sont les signaux (masses moléculaires et temps de rétention) et dont les entrées sont les aires des pics des signaux. Ce tableau constitue le tableau de travail principal.

V.2. Caractérisation et identification

Les tentatives de caractérisation et d'identification ont été réalisées à l'aide de Compound Discoverer. La prédiction des formules brutes a été effectuée par le module « Predict Composition » en utilisant les paramètres suivants :

- Eléments choisis et nombre maximal : C100 H200 Br5 Cl5 O30 N20 P3 S8.
- Tolérance sur la masse : 5 ppm ou 3 mDa.
- Vérification de la présence de chlore ou de brome par la présence des pics isotopiques ^{17}Cl et ^{81}Br.
- Paramètre « spectral fit » minimal fixé à 30%. Ce paramètre donne la similarité spectrale entre le spectre obtenu et le spectre théorique.

Après exécution de ce module, seules les formules brutes ayant un paramètre « pattern coverage » supérieur à 80% ont été conservées. Ce paramètre calcule le ratio de la somme des intensités des pics isotopiques du spectre MS^1 sur la somme des intensités du spectre théorique. Le logiciel attribue un rang sur les formules brutes. La validité de la formule brute de premier rang a été vérifiée. Dans le cadre de ce travail, seuls les signaux de masses moléculaires inférieures à 800 Da ont été caractérisés afin d'éviter un taux trop important de mauvaises attributions. En cas de résultats très proches en termes de valeurs de « pattern coverage », de «spectral fit » et de nombre de pics isotopiques conservés, la formule brute la plus proche en termes de masse a été conservée.

En plus de la prédiction de formules brutes, à partir des résultats obtenus par analyse MS et MS^2, la base de données de mzCloud a été interrogée pour essayer d'attribuer une formule développée aux composés avec une déviation en masse maximale de 5 ppm. De même, le logiciel de fragmentation *in silico* MetFrag (Ruttkies *et al.*, 2016) qui combine la recherche dans des bases de données et la prédiction de fragmentation a été utilisé dans sa version en ligne. Dans ces conditions, la déviation en masse a été réglée à 5 ppm et la déviation absolue en masse sur les fragments générés a été fixée à 3 mDa. Les bases de données interrogées par MetFrag ont été KEGG (Kyoto Encyclopedia of Genes and Genomes), Comptox (version du 7 mars 2019) et enfin HMDB (Human Metabolome Database).

Enfin, une liste établie a été interrogée au cours des travaux de thèse (comportant plusieurs classes de molécules telles que des produits pharmaceutiques et cosmétiques, des pesticides, des surfactants ou

encore des produits illicites). La tolérance en masse a également été fixée à 10 ppm pour la recherche sur ces listes.

V.3. Analyses statistiques

Les analyses statistiques des données ont été réalisées par le logiciel XLSTAT Ecology 2018.1 sous Excel à partir du tableau contenant les signaux en fonction des échantillons.

V.3.1. Réalisation des empreintes des échantillons

Les empreintes (ou profils moléculaires) représentent les masses moléculaires (ou les m/z) en fonction des temps de rétention des signaux détectés dans les échantillons. Ces empreintes ont été construites à partir de chaque échantillon après élimination des signaux dont la valeur est de 0 (e.g. signaux dont l'intensité est inférieure à 5 fois celle du blanc et/ou CV > 30 % sur les 3 réplicas) dans le tableau de données principales. Les aires des signaux ont été moyennées sur les trois réplicas. En fonction des objectifs de représentation, un code couleur été ajouté aux empreintes.

V.3.2. Réalisation des box-plots ou boîtes à moustaches

Les box-plots ont été réalisées par l'outil « visualisation des données, graphiques univariés » sur les trois paramètres : temps de rétention, masses moléculaires et $\log_{10}$ des aires. Sur ces box-plots ont été représentés les moyennes, les médianes, les premiers et troisièmes quartiles, les valeurs maximales et minimales. Chaque box-plot a été construite sur l'ensemble des signaux par échantillon.

V.3.3. Analyse en Composantes Principales

L'analyse en composante principale (ACP) a été utilisée afin de classer et d'établir des relations entre les échantillons à partir des empreintes chimiques obtenues et donc des signaux détectés. L'ACP a été réalisée sur le tableau de données par le module « analyse des données, analyse en composantes principales ». Avant toute ACP, les données ont été centrées-réduites (normalisation n-1).

V.3.4. Analyses différentielles

Les analyses différentielles ont été effectuées par le module « analyse des données omiques, analyse différentielle ». Les analyses différentielles ont été réalisées par campagne et par groupe entrée-sortie ou aval-amont, en différenciant les sorties nord et les sorties sud. Les aires des signaux ont été moyennées sur les trois réplicas. La variable explicative choisie est la nature de la matrice de l'échantillon (entrée, sortie Nord, sortie Sud, amont, aval ou panache). Le type de test réalisé est un test paramétrique avec un niveau de signification de 5% (risque α de 5%) et des corrections post-hoc de type Benjamini-Hochberg. Ce type de correction fait partie de la famille des False Discovery Rate

(FDR). Elle corrige la valeur des p-values en fonction de leur nombre et du taux de p-value non significatives. La correction de Benjamini-Hochberg est adaptée à la recherche d'un grand nombre de variables pouvant être pertinentes et est moins conservatrice qu'une correction de type Benjamini-Yekutieli. Toutes les p-values ainsi calculées ont été récupérées afin de tracer le graphique des $-\log_{10}$(p-value) en fonction du $\log_2$ des ratios (Fold Change) des échantillons considérés.

La sélection des signaux pouvant s'avérer pertinents a été effectuée à partir des résultats de ces analyses différentielles en prenant comme seuil de significativité une p-value inférieure à 0,05 et un Fold Change supérieur à -1 lors des analyses différentielles entrées-sorties ou supérieur à 1 dans le cas des analyses aval-amont.

V.3.5. Génération de la liste de métabolites

La liste de métabolites utilisée pour une analyse de suspects a été générée par le logiciel KNIME avec le module « SyGMA Metabolites ». La liste a été générée à partir de 60 molécules.

Chapitre 3 : Mise au point de l'approche pour l'analyse non ciblée : du prélèvement des échantillons au traitement des données

I. Introduction

Le développement d'une méthode analytique vise à obtenir une méthode spécifique, sensible et robuste. Dans le cas de l'analyse ciblée, ce développement passe par une validation (limite de détection, limite de quantification, spécificité etc…) assurant la fiabilité des résultats. Pour l'analyse non ciblée, la validation de méthode telle que pensée pour l'analyse ciblée n'est pas applicable en l'état. En effet, l'analyse non ciblée constitue une nouveauté dans le domaine environnemental et les méthodes de traitement des données y manquent encore d'harmonisation. Cependant, des stratégies existent pour améliorer la qualité des données par l'étude et l'optimisation de leur traitement, de la répétabilité générale, de la sensibilité et de la mise en évidence des biais existants.

Concernant le traitement des données, le développement de l'analyse non ciblée passe nécessairement par les étapes de prétraitement, évoquées dans la partie bibliographique précédente. Le prétraitement des données a pour objectif l'obtention d'un tableau regroupant les signaux détectés au travers de l'ensemble des échantillons analysés. Chaque signal extrait possède alors au minimum un m/z, un temps de rétention et une intensité (ou une aire) uniques.

Ces étapes sont réalisées à partir de logiciels libres ou commerciaux par des procédés automatiques. Les algorithmes sur lesquels reposent ces logiciels sont toutefois différents et il n'existe pas à ce jour d'uniformisation bien qu'un ensemble de paramètres de réglages soit commun. Le choix des logiciels et de leurs réglages peut donc potentiellement influencer le résultat du traitement des données. Dans le cadre de cette thèse, deux approches faisant appel à des ensembles de logiciels différents ont été utilisées et comparées. Dans ces approches, les méthodes qui y ont été développées sont présentées dans la première partie de ce chapitre. De même, une comparaison des résultats obtenus par chacune de ces approches au travers de la réalisation d'une campagne d'analyse (campagne C) est effectuée. A partir de ce travail, le choix de la méthode pour la suite de la thèse a pu être réalisé.

L'influence de la répétabilité analytique et de la répétabilité de prélèvements sur les résultats est également soulevée lors de l'analyse non ciblée. Au même titre qu'en analyse ciblée, la répétabilité est un critère de robustesse de la méthode et la répétition des analyses sert à minimiser et à comprendre les variations sur les résultats obtenus. Ainsi, dans la seconde partie de ce chapitre, la répétabilité analytique et la répétabilité des prélèvements sont étudiés à partir des résultats obtenus pour la campagne C.

Enfin, lors de ce type d'approche, les résultats dépendent des seuils d'intensité dont une partie repose sur la sensibilité de la méthode d'analyse. Pour améliorer la sensibilité, la concentration des

échantillons par une méthode d'extraction avant analyse peut être employée. Ainsi, la troisième partie de ce chapitre s'attache à montrer les effets sur les résultats de la concentration préalable par SPE des échantillons avant analyse.

II. Optimisation des paramètres et comparaison des résultats obtenus lors du prétraitement des données par deux méthodes : MZmine 2 vs Compound Discoverer

Les logiciels des deux méthodes suivies dans ce travail ont leur propre mode de fonctionnement et leurs propres algorithmes. Cependant, ces logiciels partagent un ensemble de grandes étapes :

- Détection des signaux,
- Alignement des signaux détectés à travers l'ensemble des échantillons analysés formant le tableau de signaux,
- Recherche des signaux manquants pour compléter les données n'ayant pas été détectées ou ayant été éliminées dans les étapes précédentes pour compléter le tableau de signaux (= étape du « gap-filling »),
- Détection des adduits et des isotopes pour les regrouper avec le pic monoisotopique.

L'ordre de ces étapes varie en fonction de la méthode employée. De même, des étapes supplémentaires sont parfois disponibles et nécessaires suivant le logiciel employé.

Quelle que soit l'approche utilisée, chaque étape possède un ou plusieurs paramètres à régler. Ces réglages ont été évalués et consolidés grâce à l'exploitation des analyses de six types d'échantillons prélevés lors des trois campagnes (A, B et C) :

- des eaux d'entrées de la STEP de Poitiers (17 injections, n=17),
- des eaux de la sortie Nord de la STEP de Poitiers (17 injections, n=17),
- des échantillons d'eau du Clain prélevés en amont et en aval de la STEP (17 injections de chaque, n=17),
- un mélange d'étalons (9 injections, n = 9) à une concentration fixée à 100 µg/L,
- un mélange d'étalons (2 injections, n = 2) à des concentrations allant de 50 ng/L à 1 µg/L,
- des blancs d'eau ultra-pure (20 injections, n=20).

Cet ensemble d'échantillons permet de prendre en considération à la fois les effets de la matrice sur la détection des signaux et les effets du niveau de concentration sur le nombre d'étalons détectés.

Une méthode de prétraitement des données brutes satisfaisante vise à réduire le nombre de faux négatifs et de faux positifs. L'utilisation de critères trop restrictifs peut conduire à augmenter le nombre de faux négatifs tandis que des critères trop larges peuvent quant à eux conduire à augmenter le nombre de faux positifs (Schulz *et al.*, 2019). En conséquence, une évaluation des deux méthodes sur un lot d'étalons de molécules pertinentes a été réalisée pour en évaluer la performance.

II.1. Traitement des données avec l'ensemble MZmine 2, R et Excel

II.1.1. Étapes importantes du traitement des données

Le paragraphe V.1.1 du chapitre 2 « Matériel et méthodes » présente les différentes étapes suivies lors de l'utilisation des différents logiciels de cette méthode. Dans cette partie, la démarche d'optimisation des paramètres de détection des pics, de reconstruction des chromatogrammes, d'alignement des signaux et de gap-filling de MZmine 2 est décrite. De même, le choix des paramètres avec R et Excel est expliqué pour la détection des isotopes, des adduits et la soustraction des blancs.

II.1.2. Optimisation des paramètres de traitement des données et résultats obtenus

La conversion des données brutes en mode dit « centroid » et au format mzXML pouvant entrainer une perte d'informations, une comparaison des fichiers dans le format constructeur, avec des fichiers mzXML a tout d'abord été réalisée. Dans ces conditions, à partir de plusieurs extraits de scans sur plusieurs échantillons aucune perte de signaux n'a été constatée. La conversion a donc été conservée.

II.1.2.1. Choix des paramètres de MZmine 2

L'enchainement des étapes de détection des pics, de reconstruction des chromatogrammes et de déconvolution visent à obtenir la liste de signaux par échantillons.

II.1.2.1.1. Détection des pics

La détection des m/z significatifs à partir des spectres de masses peut s'effectuer grâce à cinq algorithmes différents appelés : « centroid », « exact masse », « local maxima », « recursive treshold » et « waveleth transform». Les données utilisées dans MZmine 2 étant en mode centroid, seul l'algorithme centroid peut être appliqué selon le mode d'emploi du logiciel. Ainsi, l'unique paramètre pouvant être réglé pour cette étape est le seuil de bruit en-dessous duquel la masse exacte détectée n'est pas retenue et est considérée comme du bruit. Pour évaluer la valeur de ce seuil, pour chaque m/z détecté, une étape de reconstruction des chromatogrammes (correspondant à l'ensemble des signaux d'un même m/z) a été entreprise pour différents niveaux de bruit. En effet, le choix du seuil

de bruit a une influence directe sur cette étape qui utilise les m/z significatifs pour construire les chromatogrammes. La Figure 18 présente le nombre de chromatogrammes reconstruits en fonction du niveau de bruit pour chaque type d'échantillon.

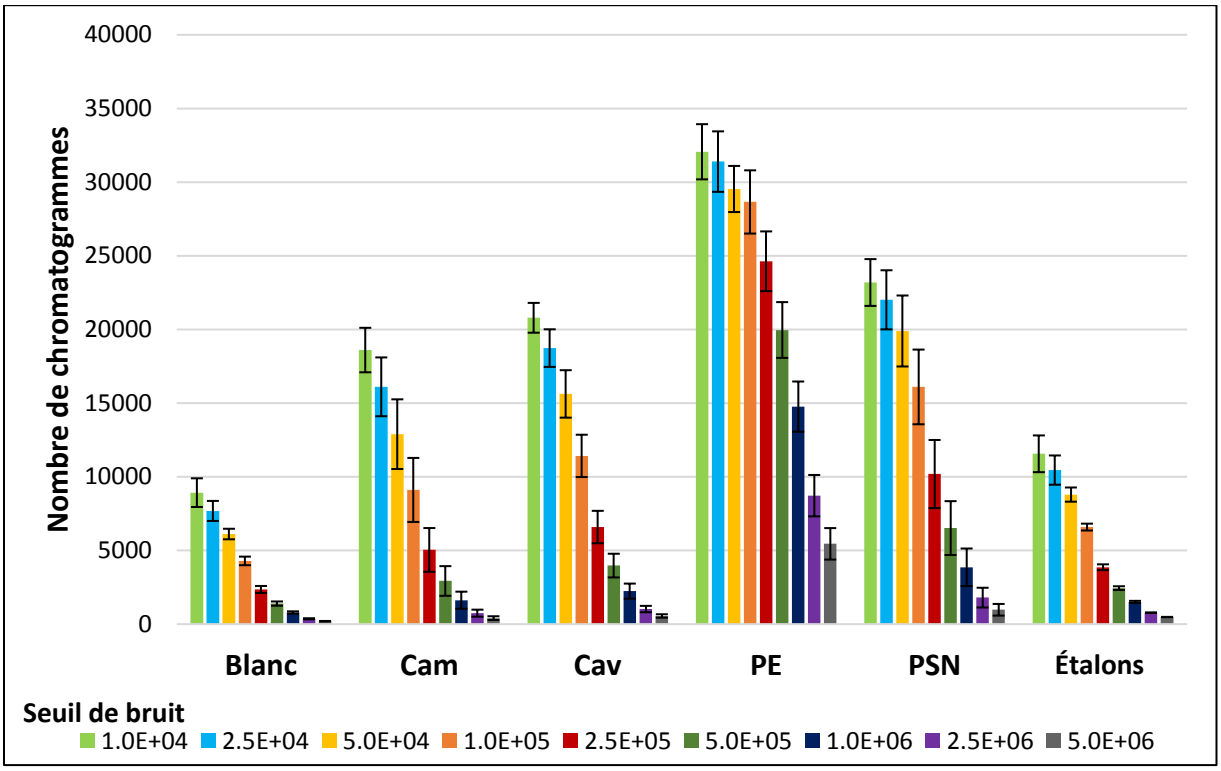


Figure 18 : Nombre de chromatogrammes reconstruits après détection des masses en mode positif en fonction du seuil de bruit pour différents types d'échantillons. Blanc : EUP, Cam : Clain amont, Cav : Clain aval, PE : entrée STEP de Poitiers, PSN : sortie STEP de Poitiers.

Cette figure montre l'augmentation du nombre de chromatogrammes avec la diminution du seuil de bruit et ce pour chacun des types d'échantillons. Cette augmentation est particulièrement significative dans le cas des blancs et des échantillons de rivières (moins chargés) avec un nombre de chromatogrammes jusqu'à 45 fois plus élevé entre les deux valeurs extrêmes. Pour les échantillons d'entrée de STEP, ce nombre est 6 fois plus grand entre les deux valeurs extrêmes. De manière générale, il apparaît que moins l'échantillon est chargé en signaux plus le nombre de chromatogrammes augmente lors de la réduction du seuil de bruit.

Augmenter le seuil de bruit réduit le nombre de données obtenues et par conséquent le nombre potentiel de faux positifs et de données à explorer. Toutefois, il existe un risque d'éliminer des signaux ayant une réponse peu intense lors de l'analyse UHPLC-HRMS (faux négatifs). Pour aider au choix du réglage de ce paramètre, les molécules étalons introduites (au nombre de 22) dans les échantillons ont donc été recherchées avec les différents niveaux de seuil de bruit. Le Tableau 10 rend compte des résultats obtenus dans les différents types de matrices avec une concentration en étalons de 100 µg/L.

Tableau 10 : Nombre d'étalons détectés par nature d'échantillon en fonction du seuil de bruit sélectionné. Concentrations des étalons fixés à 100 µg/L. Données traitées par MZmine 2 (Cam : Clain amont, Cav : Clain aval, PE : entrée STEP de Poitiers, PSN : sortie STEP de Poitiers, Standards : solution d'étalons dans l'EUP).

Seuil de bruit	Cam	Cav	PE	PSN	Standards
1.0E+04	21	21	21	21	22
2.5E+04	21	21	20	20	22
5.0E+04	21	21	20	20	22
1.0E+05	21	20	20	19	21
2.5E+05	21	21	20	20	22
5.0E+05	21	21	20	20	21
1.0E+06	22	21	20	20	21
2.5E+06	21	19	20	20	21
5.0E+06	20	19	20	20	21

Pour une concentration relativement élevée (100 µg/L), peu de différences sont observables sur l'ensemble des valeurs de seuils d'intensité de bruit testées. Cependant, le nombre de molécules détectées baisse pour les seuils de bruits les plus élevés. Des seuils supérieurs à $5x10^4$ semblent en effet être plus défavorables à l'identification des molécules présentes.

Afin de valider la sélection du niveau de bruit, cinq solutions dopées avec 19 étalons présents à des concentrations allant de 50 ng/L à 1 µg/L (dans l'eau ultra-pure) ont également été analysées. La Figure 19 montre le nombre de molécules détectées en fonction du seuil de bruit paramétré.

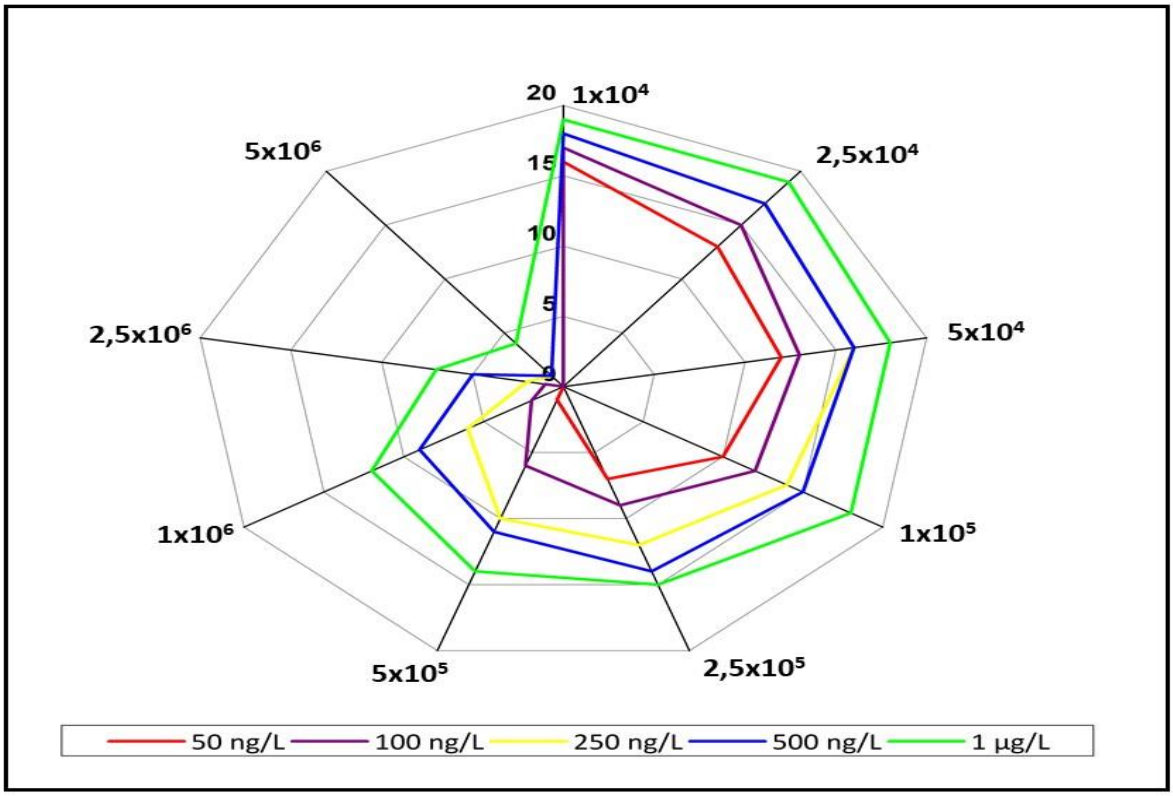


Figure 19 : Graphique en radar du nombre d'étalons détectés en mode positif par niveau de bruit paramétré lors de l'étape de détection des masses et pour différentes concentrations (22 étalons en solution dans l'eau ultra-pure). Données obtenues par le logiciel MZmine 2.

Quelle que soit la concentration en étalon recherchée, cette figure montre la baisse du nombre de molécules détectées avec l'augmentation du niveau de bruit (notamment pour les concentrations les plus faibles). Cette diminution est plus importante à partir du seuil d'intensité de $1x10^5$. Les contaminants présents dans l'environnement aquatique ont généralement des concentrations de l'ordre du ng/L à la centaine de µg/L. Exclure dès les premières étapes de retraitement des données les faibles concentrations lors de la détection des masses ne serait donc pas judicieux avec les objectifs de l'analyse et la limitation du nombre de faux négatifs semble être à privilégier. De plus, le nombre de faux positif pourra être réduit par des étapes ultérieures par exemple à l'aide de comparaison avec les données des blancs (Bader *et al.*, 2016 ; Schulz *et al.*, 2019). Ainsi, au vu de l'ensemble de ces résultats, un seuil d'intensité de $1x10^4$ a été choisi pour cette première étape.

II.1.2.1.2. Reconstruction des chromatogrammes

L'étape de reconstruction des chromatogrammes par masse est réalisée par l'algorithme ADAP lors du traitement des données utilisant MZmine 2. Pour cette étape, quatre critères sont à définir :

- Le nombre minimal de scans consécutifs dans lesquels apparait un pic pour être reconnu comme chromatogramme,
- L'intensité minimale pour intégrer le comptage,
- L'intensité minimale du pic chromatographique pour être pris en compte,
- La tolérance maximale sur le m/z entre les scans lors du comptage.

Le nombre minimal de scans est un paramètre qui dépend à la fois de l'appareil utilisé (spectromètre de masse) et des données étudiées. De la même manière que dans le cas du seuil de bruit, l'influence du nombre de scans a été évaluée sur le nombre d'étalons détectés et sur le nombre de chromatogrammes reconstruits. La Figure 20 montre le nombre de chromatogrammes reconstruits en fonction du nombre de scans réalisés pour chaque type d'échantillon.

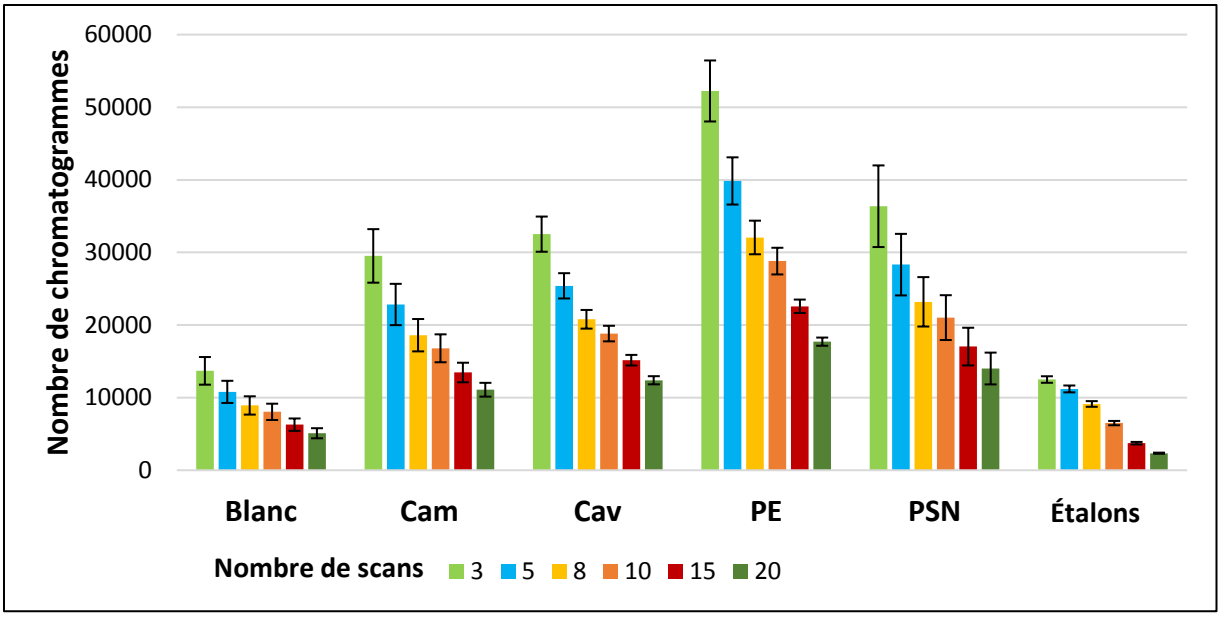


Figure 20 : Nombre de chromatogrammes reconstruits en fonction du nombre de scans en mode positif par type d'échantillons. Blanc : EUP, Cam : Clain amont, Cav : Clain aval, PE : entrée STEP de Poitiers, PSN : sortie STEP de Poitiers, Standards : solution d'étalons dans l'EUP.

Cette figure montre la diminution du nombre de chromatogrammes lors de l'augmentation du nombre de scans minimal requis. Le nombre de chromatogrammes est divisé par 3 pour chaque type d'échantillons entre 3 et 20 scans. Un plus faible nombre minimal de scans s'accompagne d'une augmentation du nombre de chromatogramme et peut induire un accroissement du nombre de faux positifs. Un nombre de scans trop bas peut ainsi mener à conserver des artéfacts ou du bruit de fond dans les données. Comme préalablement pour le seuil de bruit, l'influence de ce paramètre lors de l'analyse de solutions étalons à différentes concentrations a été étudiée dans l'eau ultra-pure. La Figure 21 présente le nombre d'étalons détecté par nombre minimal de scans pour chaque concentration en étalons injectés.

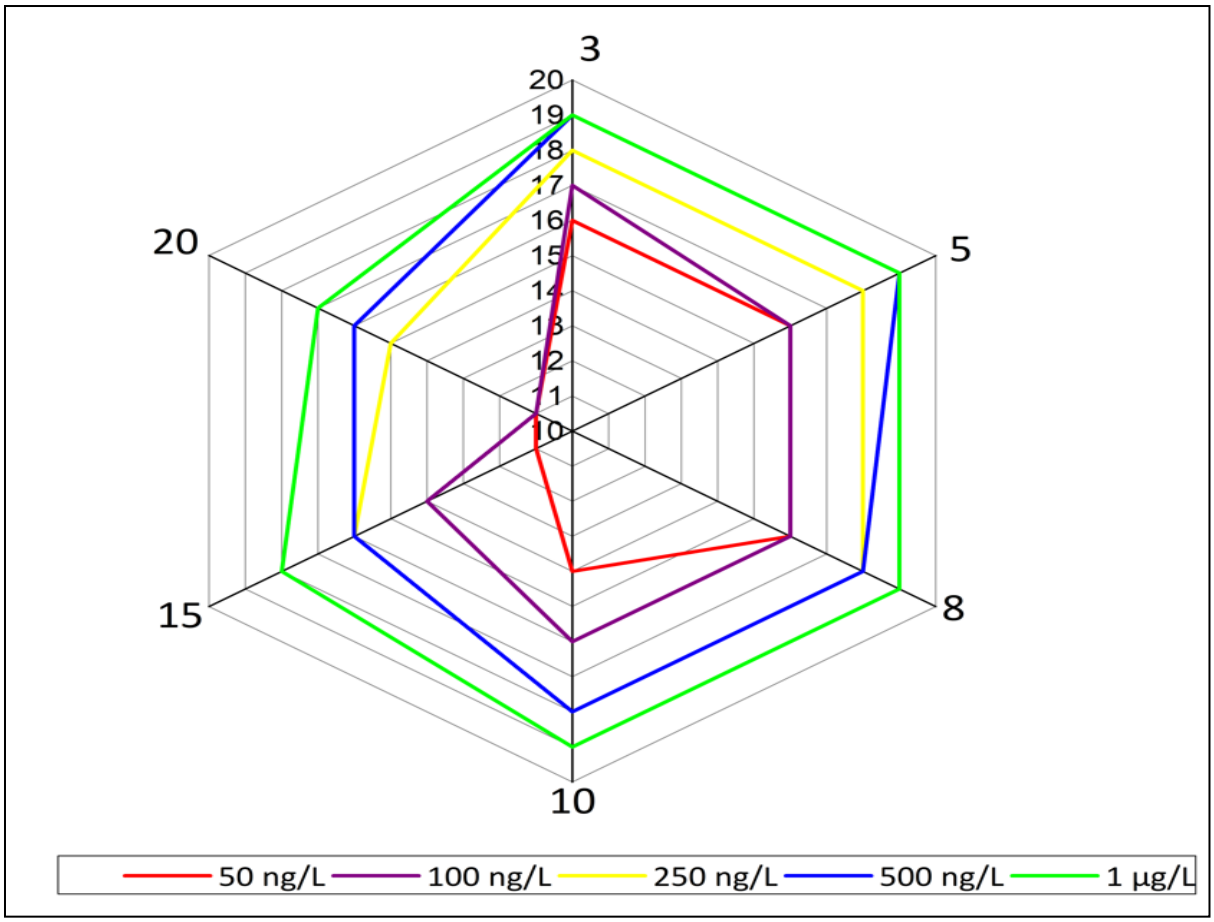


Figure 21 : Graphique en radar du nombre d'étalons détectés en mode positif par nombre de scans minimal paramétré lors de l'étape de reconstruction des chromatogrammes et pour différentes concentrations (22 étalons en solution dans l'eau ultra-pure). L'échelle du nombre d'étalons débute à 10. Données obtenues par le logiciel MZmine 2.

Comme précédemment, le nombre d'étalons reconnus diminue avec l'augmentation du nombre minimal de scans retenus. Cette constatation est aussi plus marquée pour les plus faibles concentrations. Afin de détecter ces concentrations, il semble nécessaire de conserver un nombre minimal de scans compris entre 3 et 8. Au-delà de 8 scans, la baisse du nombre d'étalons détectés devient trop importante pour l'ensemble des concentrations. Un nombre trop élevé de scans va donc limiter le nombre de pics aux signaux les plus intenses et donc diminuer la sensibilité de la méthode. Ainsi, comme compromis, un nombre de scans minimal de 5 a été choisi dans ce travail.

Concernant les autres paramètres, l'intensité minimale pour l'inclusion d'une masse dans le comptage a été fixée au même niveau que le seuil précédent de détection des masses qui est le facteur limitant, c'est-à-dire $1x10^4$. De même, l'intensité minimale du pic a été fixée à cette valeur ce qui permet de garder un nombre de faux négatifs inférieur à 5%. Enfin, la précision en masse de l'appareil servant aux analyses est inférieure à 2 ppm. Toutefois, une tolérance sur le m/z à 5 ppm a été fixée après observations des résultats obtenus avec des solutions étalons et en accord avec ce qui est généralement considéré (Schulz *et al.*, 2019).

II.1.2.1.3. Déconvolution des chromatogrammes

La dernière étape avant d'obtenir une liste de signaux par échantillon est la déconvolution des chromatogrammes en signaux uniques. Cette étape peut être réalisée grâce à six algorithmes différents proposés par MZmine 2 : « Baseline cutt-off », « Noise amplitude », « Stavisky-Golay », « Local minimum search », « Wavelets XCMS » ou « Waveleth ADAP ». Le choix de l'algorithme dépend essentiellement des données utilisées et se fait en les testant successivement, l'objectif étant de distinguer les pics ayant un même m/z mais un temps de rétention proche en incluant le minimum de bruit ou de faux positifs. L'application de chacun des algorithmes sur les données brutes a montré que l'algorithme Waveleth ADAP est le plus adapté pour les échantillons étudiés en raison de sa capacité satisfaisante à moins considérer de bruit de fond (i.e. artéfact) tout en conservant le plus de pics réels. Les autres algorithmes ne sont pas parvenus à détecter des pics réels sans amener à un grand nombre de « faux pics » appartenant au bruit.

L'algorithme ADAP possède six paramètres à régler :

- Le rapport signal sur bruit S/N minimal,
- L'algorithme estimant le rapport signal sur bruit,
- L'intensité minimale du signal à déconvoluer,
- La valeur du coefficient divisée par l'aire du pic,
- La durée du pic,
- La durée de « l'onde » (waveleth) ou pas de temps utilisé pour calculer les coefficients servant de base à la détection.

Ces paramètres ont là encore été déterminés par l'étude des données scan par scan et validés par les mélanges d'étalons. Ainsi, l'intensité minimale du pic a été fixée à $1x10^4$ pour rester cohérent avec la valeur choisie dans les étapes précédentes. Les cinq autres paramètres ont été estimés par l'étude des scans UHPLC-HRMS afin d'obtenir un compromis entre la capacité de l'algorithme à déconvoluer des pics réels et le risque d'intégrer du bruit de fond ce qui contribuerait à augmenter le nombre de faux positifs. Les valeurs suivantes ont alors été retenues : seuil Signal/Bruit 5, intensité minimale du signal $1x10^4$, seuil coefficient/aire 130, gamme de durée des pics 0,05-2 min, durée du pas du temps 0,05-0,10 min. La Figure 22 présente une capture d'écran d'un exemple de déconvolution selon les réglages retenus.

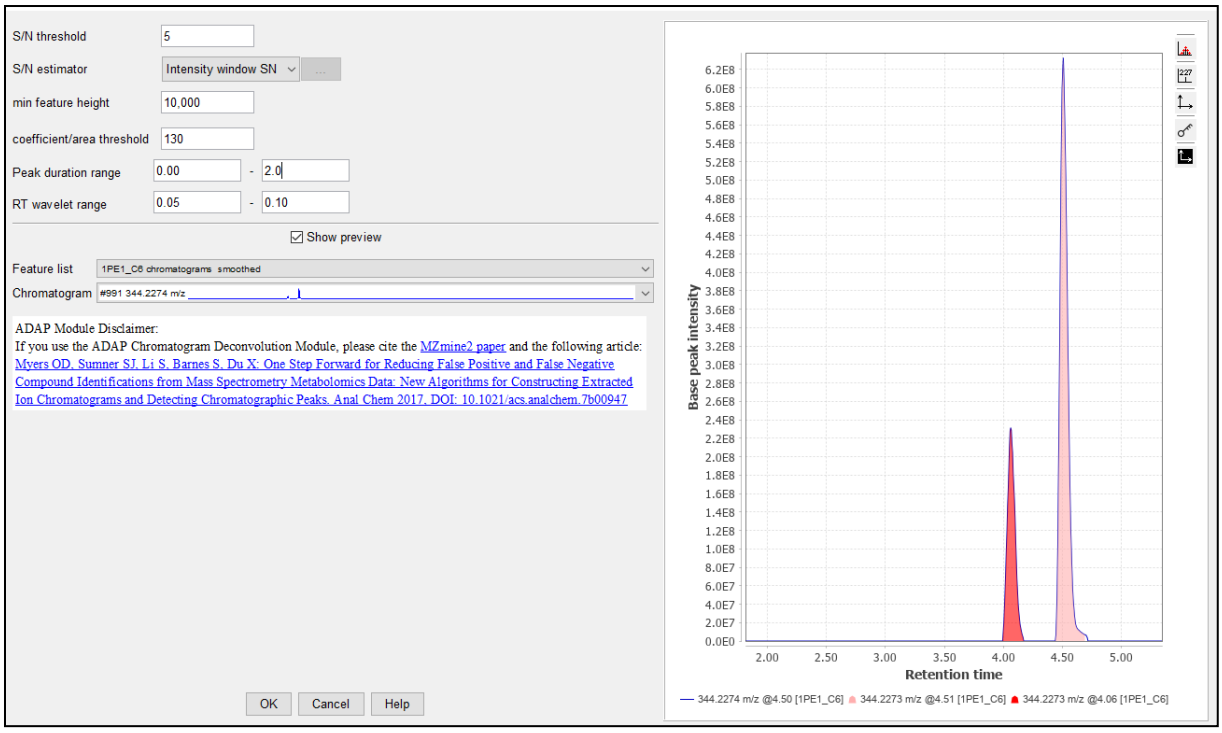


Figure 22 : Capture d'écran d'un exemple de déconvolution de deux signaux au m/z 344,2274 selon les réglages finaux. Logiciel MZmine 2.

Cet exemple illustre une déconvolution réussie : les deux signaux ayant le même m/z mais des temps et des intensités différentes sont reconnus et intégrés comme deux signaux distincts.

II.1.2.1.4. Alignement des signaux et « Gap-filling »

L'alignement des signaux est l'étape qui réunit les signaux à travers l'ensemble des échantillons. Cette étape est critique dans le traitement des données. En effet, un alignement incorrect entraine l'apparition de doublons dans le tableau dans le cas de critères trop restrictifs ou englobe sous la même ligne des signaux différents dans le cas de critères trop larges.

Deux algorithmes sont disponibles dans MZmine 2 pour effectuer cet alignement : « Join aligner » et « RANSAC ». Ces deux algorithmes partagent les tolérances sur le m/z et sur le temps de rétention mais diffèrent sur le mode de fonctionnement. Tous deux ont été testés sur les mêmes groupes d'échantillons et l'algorithme « Join aligner » a été conservé en raison de sa performance (i.e. capacité à correctement aligner les pics et ainsi éviter les doublons) plus satisfaisante que « RANSAC ». En accord avec les paramètres de l'étape de déconvolution des chromatogrammes, une tolérance en masse de 5 ppm a été choisie pour cet algorithme. Concernant le temps de rétention, une tolérance de 0,4 min a été fixée après injections de solutions dopées en étalons et observations des résultats. La Figure 23 présente un exemple de signaux regroupés ensemble après alignement selon les réglages finaux.

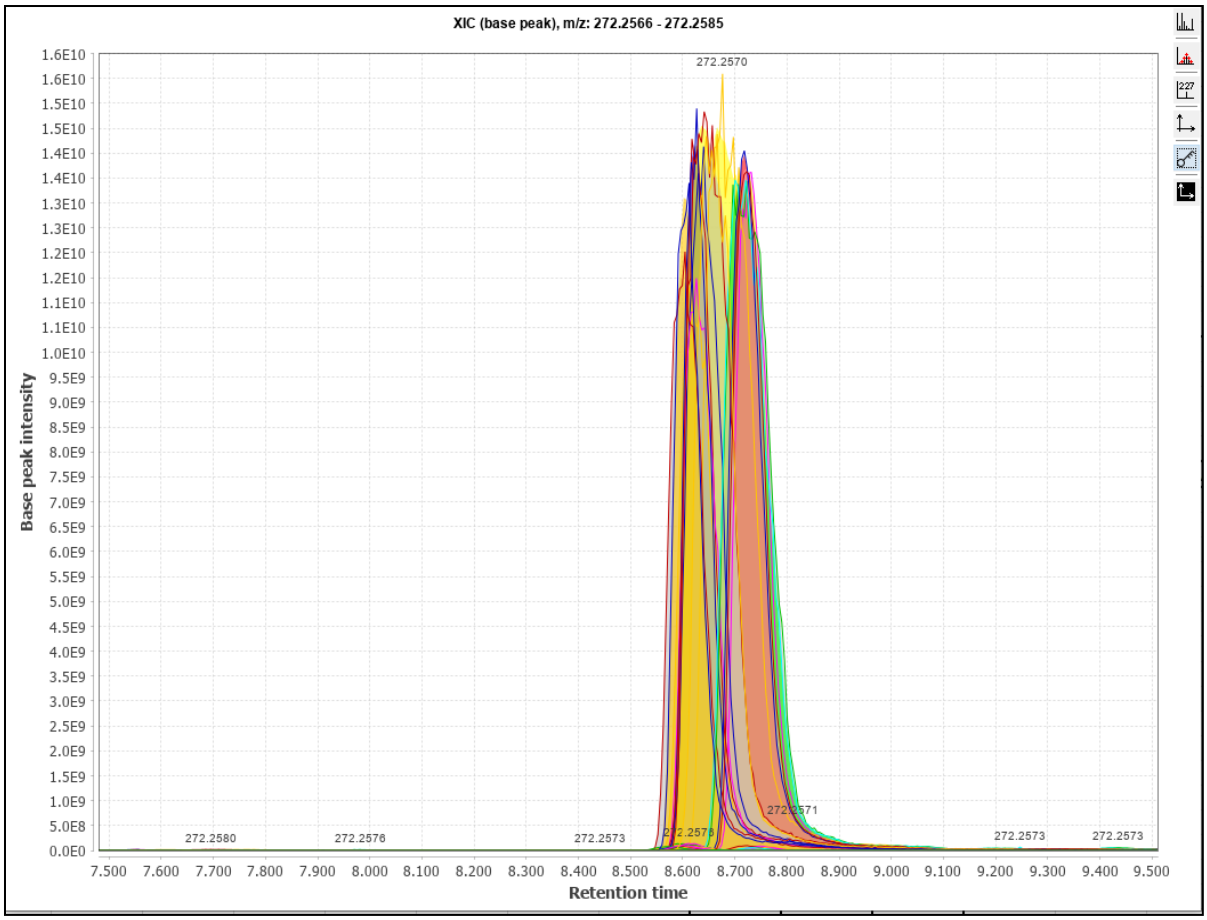


Figure 23 : Capture d'écran des chromatogrammes d'un exemple de signaux alignés au m/z 272,2570 selon les réglages finaux. Logiciel MZmine 2.

Cette figure montre le regroupement d'un même signal présent dans de multiples échantillons et dont le temps de rétention varie légèrement en fonction de l'échantillon.

Pour l'étape du « Gap-filling » deux choix d'algorithme sont également disponibles sur MZmine 2 : « Peak finder » et « Same RT and gap ranger filer ». Le premier recherche des signaux non repérés directement dans les données brutes tandis que le second se base sur la liste de signaux alignés. Afin de choisir une méthode, les deux possibilités ont là encore été testées. Ainsi, l'algorithme « Peak finder » a été conservé pour sa capacité à rechercher les signaux manquants, le second algorithme introduisant trop de variabilité du nombre de signaux sur les réplicas. Comme réglages de cet algorithme, le paramètre du temps de rétention a été fixé à 0,2 min et la tolérance sur la masse a été fixée à 5 ppm. Les Figures 24A et 24B représentent un signal présent dans les 3 injections d'un même échantillon avant et après « gap-filling ».

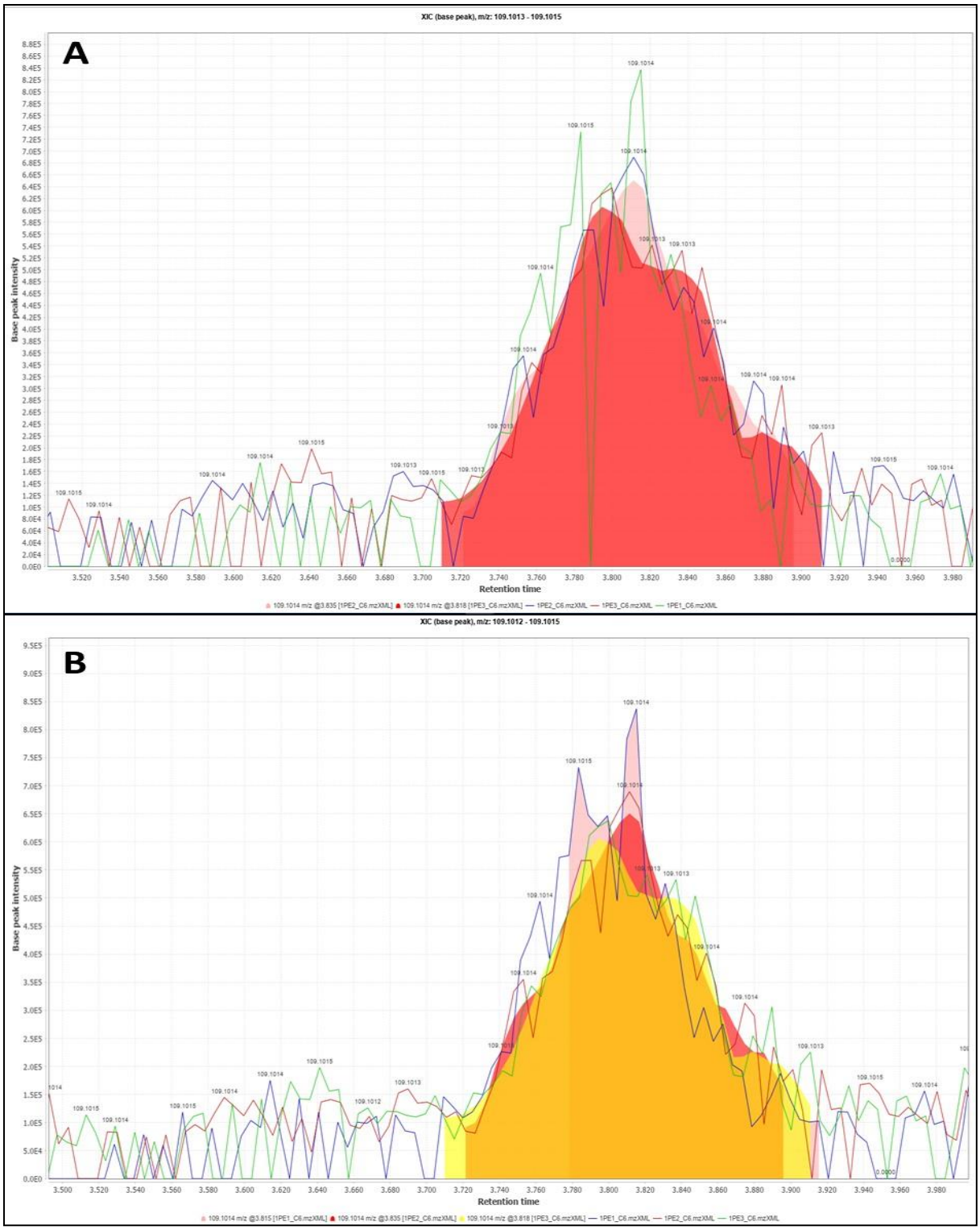


Figure 24 (A et B) : Captures d'écran des chromatogrammes d'un exemple d'un signal d'un m/z 109,1014 présent dans les 3 injections d'un même échantillon d'entrée de STEP selon les réglages finaux. En A avant le « gap-filling », en B après le « gap-filling ». Logiciel MZmine 2.

La Figure 24A indique la présence du signal dans les trois injections (les tracés en vert, bleu et rouge) mais seuls les signaux correspondant à deux injections sont identifiés par la méthode (intégrations matérialisées par les zones rose et rouge) : le troisième signal n'est pas retenu dans la liste de signaux malgré sa présence visible dans le chromatogramme. Après « gap-filling » (Figure 24B), le signal de la

troisième injection (intégré en jaune) a été détecté et ajouté dans la liste des signaux. Cette étape a donc bien corrigé l'absence d'un signal, non repéré durant les étapes précédentes.

II.1.2.1.5. Étapes supplémentaires

En plus des étapes obligatoires précédemment décrites, des étapes ont été ajoutées pour éliminer des signaux de bruits ainsi que des signaux artéfacts et améliorer la qualité de l'alignement.

Comme première étape supplémentaire, un filtrage des signaux après déconvolution a été ajouté. Ainsi, les pics n'ayant pas au moins 20 points, une largeur à mi-hauteur comprise entre 0,05 et 0,9 min et un temps de rétention compris entre 0,50 et 12,30 min ont été éliminés. Ce filtrage a été effectué pour éliminer des faux positifs qui sont soit des pics artéfacts, soit des pics qui apparaissent lors de la déconvolution et qui appartiennent à la ligne de base et ne correspondent pas à de vrais pics de chromatogrammes bien définis.

La seconde étape supplémentaire ajoutée à la méthode a consisté en la normalisation des temps de rétention. Cette seconde étape se base sur les pics communs à l'ensemble des échantillons pour réduire les variations de temps de rétention et améliorer la qualité de l'alignement.

Enfin, la dernière étape supplémentaire a consisté en la suppression des doublons pouvant être présents dans la liste finale suite à un mauvais alignement. En effet, le « gap-filling » qui suit l'alignement comble les pics manquant mais conserve alors des doublons. Ce dernier filtrage des doubles permet d'éviter au maximum ces redondances.

II.1.2.2. Choix des paramètres de Excel et R

II.1.2.2.1. Détection des isotopes et des adduits par les logiciels R 3.5.1 et R.studio

La liste des signaux obtenus par MZmine 2 a ensuite été utilisée par le logiciel R 3.5.1 via R.Studio et le package « non-target ». Ce package permet de détecter les isotopes et les adduits dans le but de les regrouper avec le pic monoisotopique. Dans ces conditions, la détection des isotopes s'est effectuée par la fonction pattern.search et celle des adduits par la fonction adduct.search. Les isotopes pris en compte ont été : ^{13}C, ^{14}N, ^{18}O, ^{37}Cl, ^{81}Br et ^{34}S. En effet, les isotopes du chlore, du brome et du soufre sont abondants et sont utiles pour l'attribution de formules brutes en raison de leurs distributions spécifiques sur les spectres de masse. Les isotopes du carbone, de l'oxygène et de l'azote sont faibles (<1%) mais restent intéressants dans le cas des signaux très intenses ou présentant de nombreux atomes identiques. Les adduits courants ont été recherchés : $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M+CH_3OH+H]^+$ et $[M+K]^+$ en mode positif et $[M-H]^-$, $[M-CHOO]^-$, $[M+FA-H]^-$, $[M-Cl]^-$ en mode négatif.

II.1.2.2.2. Soustraction des blancs avec Excel

La correction des données par l'utilisation de blancs est une étape privilégiée pour l'élimination de faux positifs (Schulz *et al.*, 2019). Cette étape implique la mise en place de blancs d'extraction prenant en compte la préparation d'échantillons et de blancs du système UHPLC-HRMS. La prise en compte des blancs peut s'effectuer de deux manières : par exclusion de tous les pics présents dans les blancs ou par élimination des pics ne respectant pas une valeur choisie portant sur le rapport entre l'intensité dans l'échantillon et l'intensité du blanc. Dans le cas de la méthode développée, c'est la seconde solution qui a été réalisée en éliminant les signaux dont l'intensité n'est pas 5 fois supérieure à celle du blanc. Ce choix a été fait en raison du caractère ubiquitaire des micropolluants pouvant contaminer les blancs par les solvants ou le matériel. Ainsi, grâce à une macro Excel, tous les signaux inférieurs à 5 fois l'intensité du blanc ont donc été éliminés et les signaux restants ont été modifiés par les intensités corrigées en soustrayant le blanc.

II.1.2.3. Méthode finale retenue pour le prétraitement des données lors de l'utilisation de MZmine 2, Excel et R

Après étude des paramètres de réglages des différentes étapes, une méthode finale a été obtenue, combinant les trois logiciels, MZmine 2, Excel et R. Le Tableau 5 du chapitre « Matériel et méthodes » récapitule les réglages utilisés pour chacune des étapes.

Ces différentes étapes ont permis d'éliminer le bruit et les pics artéfacts, de conserver les pics pouvant être pertinents et de réduire le volume de données.

II.2. Traitement des données avec la méthode utilisant Compound Discoverer

II.2.1. Le schéma d'étapes ou « workflow »

La Figure 17 du chapitre « Matériel et méthodes » présente le « workflow » ou blocs d'étapes suivies lors de l'utilisation du logiciel Compound Discoverer.

Les deux premières séries (orange et violet dans la Figure 17 de ce workflow) ont pour objectifs de détecter les signaux, d'établir un tableau regroupant l'ensemble des pics détectés à travers tous les échantillons, de rechercher les signaux manquants (étape de « gap-filling ») puis de comparer les signaux obtenus avec ceux du blanc. Ce présent chapitre a pour objet de décrire la démarche d'optimisation des paramètres de ces étapes.

II.2.2. Optimisation du traitement des données par Compound Discoverer

Chacune des étapes du workflow comporte des paramètres dont les valeurs sont définies en fonction des données utilisées et des choix de l'utilisateur. Si chacun des blocs possède des paramètres qui leur sont propres, trois paramètres sont néanmoins présents sur l'essentiel des étapes considérées : les tolérances appliquées aux masses moléculaires et aux temps de rétention ainsi que le seuil d'intensité. Lors de l'utilisation de Compound Discoverer, comme lors de l'utilisation de MZmine 2, ces paramètres ont une influence prépondérante sur le nombre de signaux obtenus dans le tableau final. Les principales étapes de la méthode ont été évaluées et consolidées de la même manière que lors de l'approche utilisant MZmine 2, avec les mêmes types d'échantillons (introduction du paragraphe II).

Dans ces conditions, pour comparer l'influence des différents paramètres, c'est le tableau final regroupant l'ensemble des signaux détectés à travers tous les échantillons considérés qui a été étudié. Pour chaque valeur test, le nombre d'échantillons et la nature des échantillons ont été les mêmes afin de ne pas fausser les essais pour un même paramètre.

II.2.2.1. Alignement des temps de rétention

Pour l'alignement des données, une première étape de sélection des spectres de masses dans les données brutes est réalisée. Pour cela, dans le cadre de ce travail, seuls les spectres de masse avec des m/z de 100 à 1500 ont été conservés. Pour les autres paramètres, les valeurs par défaut ont été utilisées. Pour les spectres de masse sélectionnés, l'alignement des données est la seconde étape appliquée par le logiciel : cet alignement est réalisé par un modèle non linéaire qui, en cas d'échec, peut passer sur un modèle linéaire. Pour cette étape, comme lors de l'approche utilisant MZmine 2, différentes valeurs de tolérance du temps de rétention ont été testées : 0,2 min ; 0,5 min ; 0,75 min ; 1 min ; 1,2 min ; 1,5 min et 2 min. Dans ces conditions, pour chaque valeur, le logiciel indique si l'alignement est insuffisant. De même, des tolérances en masse de 5 ppm, 10 ppm, 15 ppm et 20 ppm ont été testées. Pour ces paramètres de tolérance, le nombre de signaux obtenus suite aux étapes d'alignement, détection et regroupement des signaux sont présentés dans le Tableau 11. Dans ce tableau, la différence d'ordre de grandeur des nombres de signaux détectés lors de l'étude de ces deux paramètres provient de la différence de seuil minimal d'intensité de détection des pics, fixé à $5{,}5 \times 10^5$ pour les tests de tolérance sur les temps de rétention et à $7{,}5 \times 10^4$ pour les tests de tolérance sur la masse.

Tableau 11 : Nombre de signaux détectés en fonction des différentes valeurs de tolérance sur le temps de rétention ou sur la tolérance en masse.

Tolérance temps de rétention (min) pour une tolérance en masse de 5 ppm	Nombre de signaux	Tolérance masse (ppm) pour une tolérance de temps de rétention de 1 min	Nombre de signaux
0,2	35 859	**5**	328 793
0,5	35 835	**10**	328 153
0,75	35 824	**15**	328 526
1	35 860	**20**	328 450
1,2	35 881		
1,5	35 859		
2	35 868		

Ce tableau indique que les paramètres de tolérance sur l'alignement n'impliquent pas de changements importants puisque les variations constatées sont inférieures à 0,5%. Pour Compound Discoverer cette étape d'alignement, qui repose sur les spectres de masse servant de base, semble donc avoir moins d'impacts que dans le cas de la méthode utilisant MZmine 2. Le logiciel informe par ailleurs de lui-même lorsque l'alignement n'est pas suffisant au regard des échantillons étudiés. La valeur par défaut de 5 ppm a été conservée pour la tolérance sur la masse. En se basant sur les recommandations du constructeur, la tolérance sur les temps de rétention a été fixée à 1 min.

II.2.2.2. Détection et regroupement des signaux, recherche des signaux manquants et comparaison avec les blancs

La détection des signaux dans Compound Discoverer s'effectue sur la base d'un seuil d'intensité défini par l'utilisateur. En se basant sur les recommandations du constructeur les intensités suivantes ont été testées : $5x10^4$, $7,5x10^4$, $1x10^5$, $2,5x10^5$ et $5x10^5$. Dans ces conditions, le nombre total de signaux en fonction de la valeur de l'intensité minimale a été tracée suite aux étapes d'alignement, de détection et de regroupement des signaux (Figure 25).

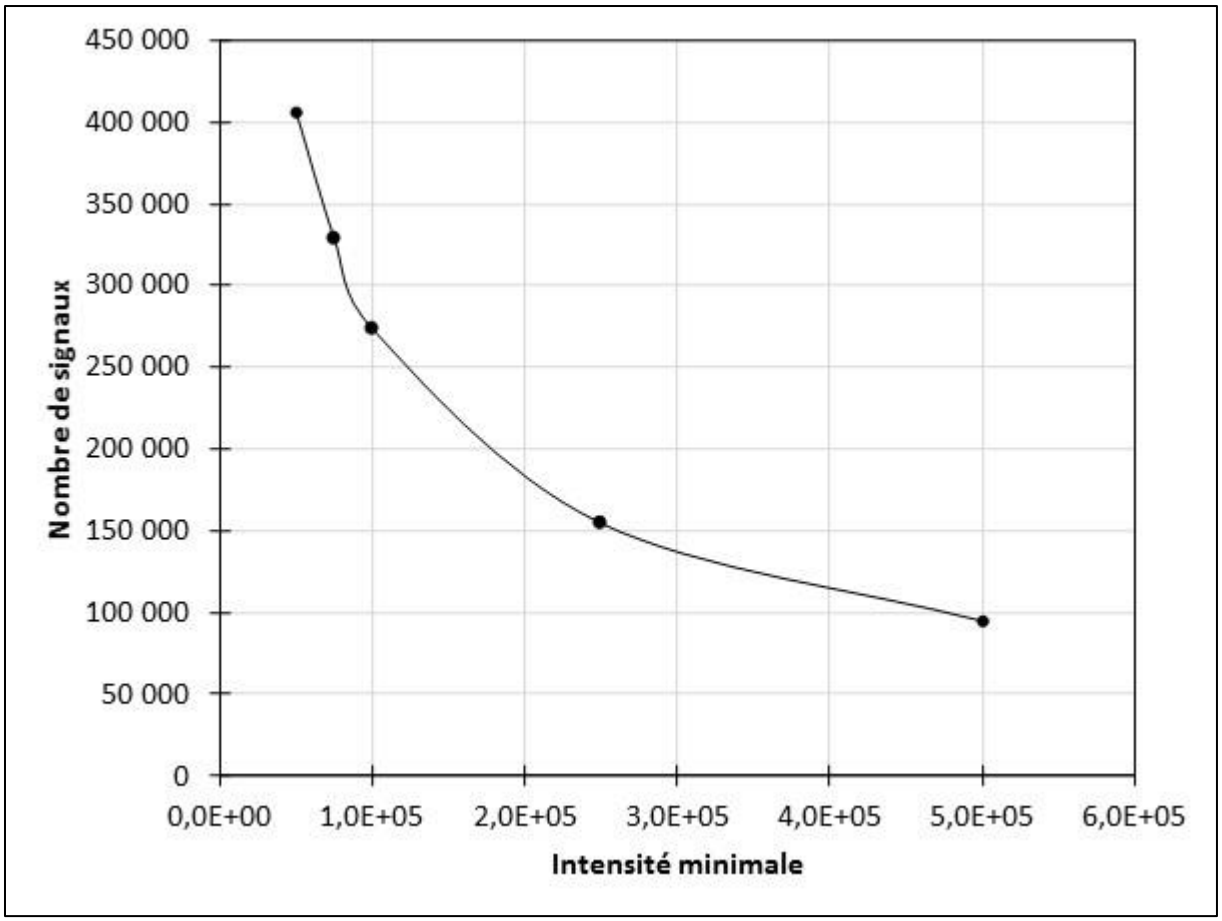


Figure 25 : Nombre total de signaux en fonction de la valeur d'intensité paramétrée en mode positif. Des valeurs d'intensités ont été testées entre $5x10^4$et $5x10^5$. Données issues de Compound Discoverer.

Ces résultats montrent que le nombre de signaux détectés diminue avec l'augmentation du seuil d'intensité. Ces résultats sont cohérents puisqu'un seuil plus grand est logiquement plus restrictif envers les signaux les moins intenses. En changeant ce paramètre, il a pu être constaté qu'une intensité minimale multipliée par 2 ($5x10^4$- $1x10^5$) conduit ici à un nombre total de signaux divisé par environ 1,5 et entrainerait la perte des signaux de plus faibles concentrations.

Pour choisir la valeur de ce paramètre, les résultats obtenus avec des solutions de mélange d'étalons de cinq concentrations différentes ont alors été comparés. Pour chacune de ces solutions, le nombre d'étalons détectés en fonction de l'intensité minimale réglée a été relevé (Figure 26).

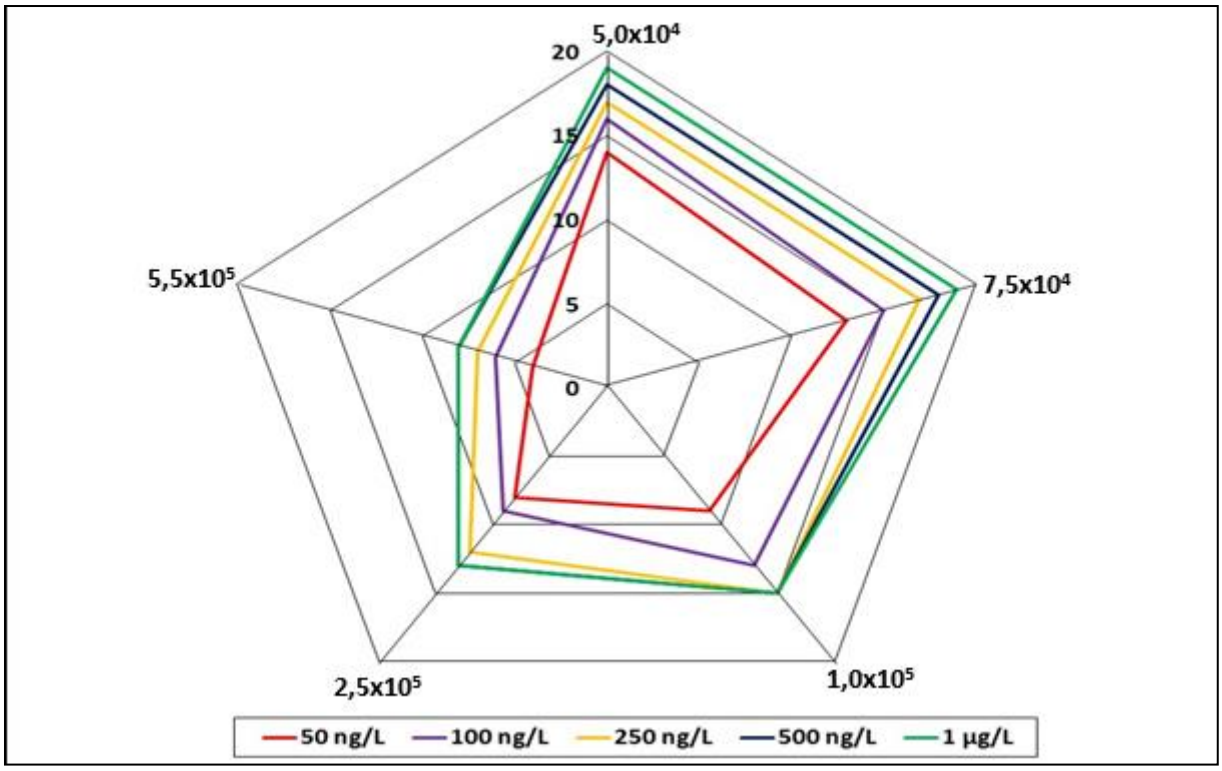


Figure 26 : Graphique en radar du nombre d'étalons détectés par seuil d'intensité minimale paramétré lors de l'étape de détection des masses et par concentration. Données obtenues par le logiciel Compound Discoverer.

Les résultats obtenus dans ces conditions montrent que le nombre de composés détectés diminue avec l'augmentation du seuil d'intensité. Ces résultats vont logiquement dans le même sens que ce qui a été mis en évidence avec MZmine 2. Pour le traitement avec Compound Discoverer, la perte d'étalons semble s'effectuer à partir du seuil de $1x10^5$ et les valeurs inférieures sont satisfaisantes. Ainsi, une intensité minimale pour la détection des signaux de $7,5x10^4$, suffisante pour maintenir dans le tableau final des molécules à faibles concentrations, a été sélectionnée.

Lors de cette même étape, les isotopes et les adduits recherchés doivent également être définis. Aussi, les isotopes du carbone, de l'oxygène, de l'azote, du brome, du chlore, du soufre ont été recherchés. Les adduits pris en compte en mode positif ont été les suivants : $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M+CH_3OH+H]^+$, $[M+K]^+$, $[2M+H]^+$, $[2M+K]^+$, $[2M+Na]^+$, $[M+2H]^{2+}$, $[M+3H]^{3+}$, $[M+H+K]^{2+}$, $[M+H+Na]^{2+}$, $[M+H+NH4]^{2+}$, $[M+H-H_2O]^+$. Les adduits pris en compte en mode négatif ont été $[2M+FA-H]^-$, $[M-Cl]^-$, $[M+FA-H]^-$, $[M-2H]^{2-}$, $[M-H]^-$, $[M-H+HAc]^-$, $[M-H-H_2O]^-$.

L'étape qui suit la détection des signaux est celle du groupement des signaux. Cette étape réunit les signaux détectés par échantillons et réunit sous un même pic d'ion défini, les isotopes, adduits et multichargés. Pour cette étape, trois paramètres sont à régler : les tolérances en masse et en temps de rétention ainsi que l'ion de base. Dans le cadre de ce travail, l'ion de base choisi en mode positif a été le $[M+H]^+$ car cet ion est généralement majoritaire avec l'utilisation d'une source ESI. Pour ce qui

est des tolérances, elles ont été fixées tout d'abord par rapport à MZmine 2 (0,4 min et 5 ppm, respectivement pour le temps de rétention et la masse). Cependant, le retour d'expérience a montré que la tolérance sur la masse était trop stricte puisque certains signaux devant être regroupés ensemble ne l'étaient pas. Par conséquent cette tolérance a été élargie à 10 ppm.

L'étape du « gap-filling » comporte trois paramètres dans Compound Discoverer : la tolérance sur la masse, le rapport signal sur bruit et un paramètre indiquant si l'algorithme doit rechercher directement dans les données non traitées. Pour cette étape, une tolérance de 5 ppm a été choisie pour rester cohérent avec l'étape de détection des signaux. Un rapport signal sur bruit de 5 a été choisi. Enfin, l'algorithme en temps réel a été sélectionné. Dans ces conditions, le temps de calcul est plus long mais les résultats obtenus sont plus satisfaisants.

Pour le blanc, l'étape réalisée sur Compound Discoverer compare pour chaque signal la valeur maximale de l'aire de tous les blancs avec les aires du même signal dans l'ensemble des échantillons. Si aucune aire dans les échantillons n'est supérieure à 5 fois celle du blanc dans la méthode, alors le signal est considéré comme appartenant au blanc. L'étape du blanc, incluse dans Compound Discoverer, permet un premier tri des signaux. Pour affiner les résultats et réaliser les empreintes, un second tri est fait après export dans Excel avec la même macro que celle effectuée dans la méthode utilisant MZmine 2, R et Excel. Dans ces conditions, une soustraction de la valeur des blancs est ainsi également réalisée pour corriger les aires.

Enfin, des filtres supplémentaires peuvent être ajoutés sur le tableau de signaux de Compound Discoverer afin de procéder à l'élimination de faux positifs. Ces filtres peuvent porter sur de nombreux paramètres tels que le temps de rétention, la masse moléculaire, les aires détectées, les coefficients de variation des aires, etc… Des filtres supplémentaires ont été ajoutés en fonction du nombre de données et du type d'échantillons. Cette liste de filtre est détaillée en Annexe 1.

II.2.3. Méthode retenue pour le prétraitement des données par Compound Discoverer

Après cette phase d'optimisation, les paramètres décrits dans le Tableau 9 du chapitre « Matériel et méthode » ont été conservés pour les différentes étapes de la méthode employée avec Compound Discoverer.

III.3. Comparaison des résultats et choix de la méthode de retraitement des données

III.3.1. Comparaison des résultats obtenus avec les deux méthodes de retraitement des données

Les deux méthodes ainsi optimisées donnent des résultats similaires lors de la recherche des étalons présents à diverses concentrations. Toutefois, Compound Discoverer rend des résultats finaux sous la forme de masse moléculaire tandis que MZmine 2 les donne sous la forme de m/z. Pour pouvoir effectuer la comparaison des résultats obtenus avec les deux méthodes, les m/z détectés par Compound Discoverer ont alors été récupérés. Dans ces conditions, l'étape des filtres supplémentaires sur le tableau des signaux avec les masses moléculaires de Compound Discoverer ne peut pas être appliquée sur le tableau des signaux avec les m/z. De plus, les multi-chargés et les adduits sont toujours présents entrainant une redondance. Lors de l'utilisation de MZmine 2, les multichargés sont toujours présents également. En revanche, les adduits et les isotopes ont été annulés. Dans ces conditions, la comparaison est donc en partie biaisée.

Dans cette partie, la comparaison des résultats obtenus avec les deux méthodes de retraitement des données a été effectuée sur trois échantillons d'entrées et trois échantillons de la sortie Nord de la STEP de Poitiers. En utilisant cette approche, la moyenne du nombre de signaux sur les trois échantillons d'entrées est de 19 338 ± 1 470 avec Compound Discoverer et de 49 818 ± 7 790 avec MZmine 2. Pour les sorties, la moyenne est de 11 988 ± 280 avec Compound Discoverer et de 39 131 ± 628 avec MZmine 2. Le nombre de signaux détectés est donc entre 2,6 et 3 fois plus élevée avec MZmine 2. Plusieurs hypothèses peuvent expliquer ces écarts : l'élimination des signaux n'ayant pas au minimum un pic isotopique est effectuée dans Compound Discoverer mais pas avec MZmine 2 ou encore l'algorithme de détection utilisé par MZmine 2 pourrait, par rapport à Compound Discoverer, intégrer plus de pics. Il a été en effet constaté que davantage de pics de mauvaise qualité ou appartenant au bruit de fond étaient généralement conservés lors de l'utilisation de MZmine 2.

La Figure 27 présente les boîtes à moustaches résumant les statistiques descriptives pour six échantillons (3 entrées et 3 sorties Nord) traités avec Compound Discoverer et avec MZmine 2.

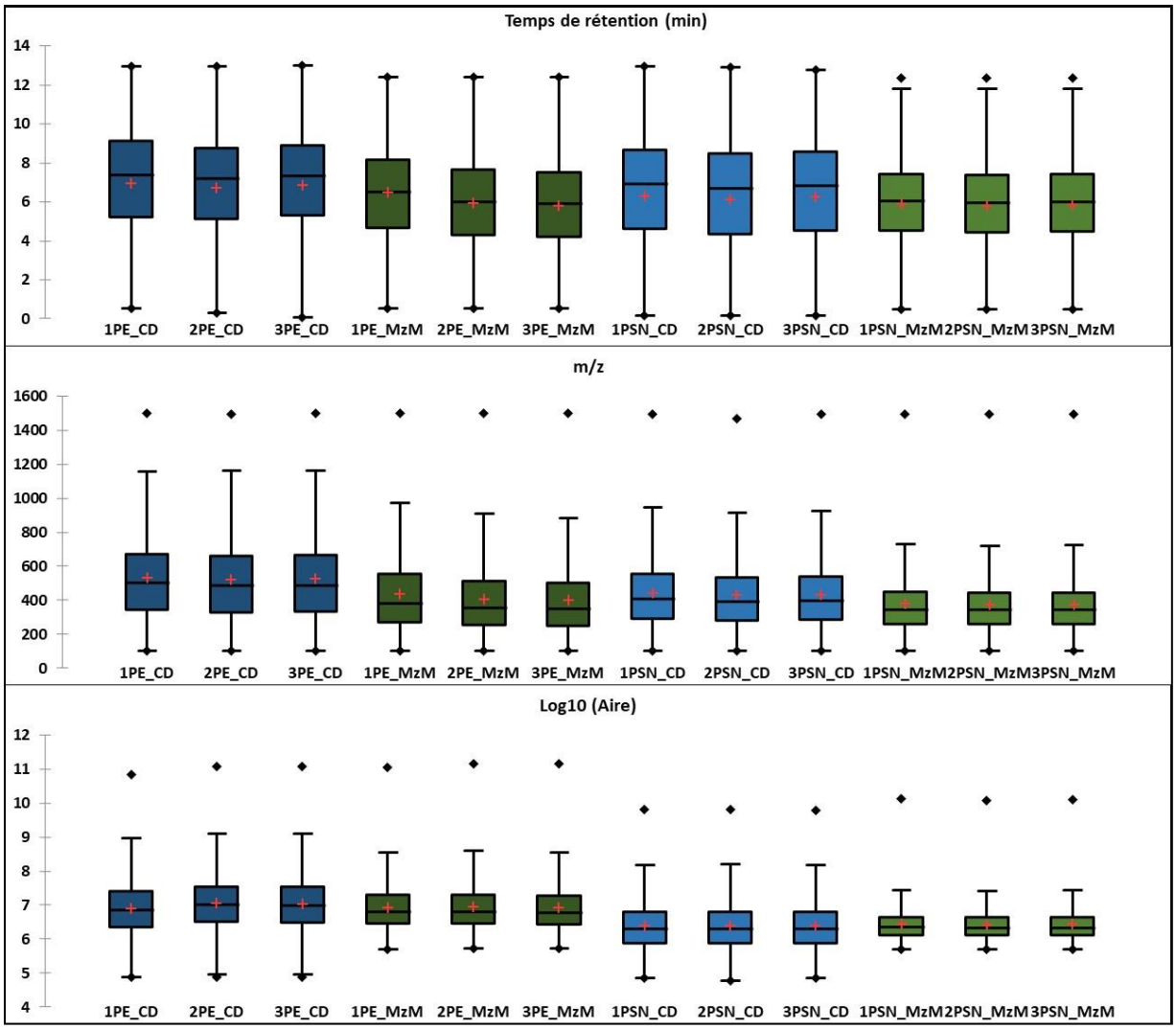


Figure 27 : Box plots ou boîtes à moustaches résumant les échantillons par leurs statistiques descriptives. La croix rouge, les limites inférieures et supérieures de la boîte et les points noirs symbolisent respectivement la moyenne, les premiers et troisièmes quartiles, les valeurs maximales et minimales. Les données traitées par Compound Discoverer sont en Bleu, celles traitées par MZmine 2 en vert. PE : entrée Poitiers, PSN : sortie Nord de Poitiers, CD : Compound Discoverer, MzM : MZmine 2.

Pour les temps de rétention et les m/z, on constate un léger décalage des boîtes vers le bas avec MZmine pour les deux types d'échantillons. En revanche, pour les aires, les moyennes et les médianes sont équivalentes entre les deux logiciels. On voit cependant une dispersion plus faible du log10 des aires avec MZmine 2. Ainsi, les résultats obtenus par Compound Discoverer et par MZmine 2 diffèrent selon les trois paramètres et ce pour les deux échantillons étudiés. Pour aller plus loin, les empreintes obtenues pour un même échantillon d'entrée et un échantillon de sortie ont alors été tracées et superposées.

Les empreintes (ou profils moléculaires) sont la représentation visuelle des signaux détectés dans l'échantillon représentés par le m/z (ou la masse moléculaire) en fonction du temps de rétention. Les

Figures 28A et 28B présentent les empreintes des 10 000 signaux les plus intenses pour un échantillon d'entrée et un échantillon de sortie selon Compound Discoverer ou selon MZmine 2.

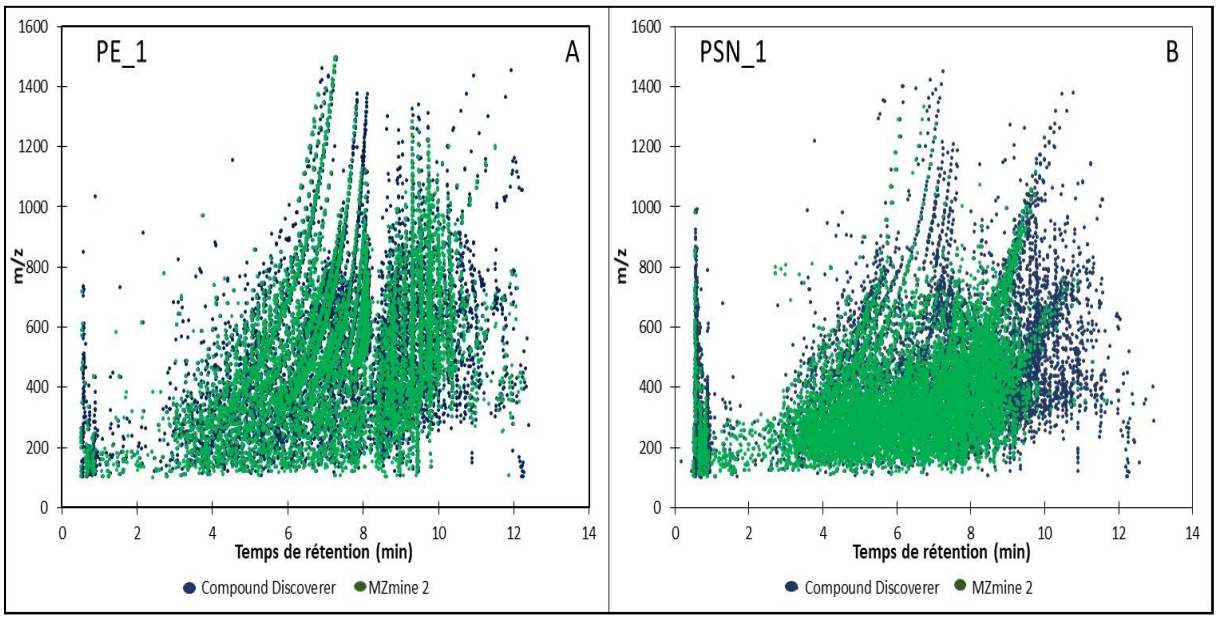


Figure 28 : Empreintes d'échantillons de l'entrée en A et de la sortie Nord en B de la STEP de Poitiers et concentrés par SPE. Prélèvement issu de la campagne C Les points en bleu et vert représentent respectivement les signaux des empreintes obtenue par Compound Discoverer ou obtenue par MZmine 2. PE : entrée Poitiers, PSN : sortie Nord Poitiers.

Sur ces deux figures, on constate une importante similitude entre les empreintes obtenues par les deux méthodes. Toutefois, il ressort également un certain nombre de signaux présents uniquement avec l'un ou l'autre des logiciels : cette observation est plus marquante sur les empreintes de sorties que d'entrées. Les signaux qui diffèrent semblent des signaux plutôt apolaires caractérisés par des forts temps de rétention.

Enfin, une comparaison des résultats obtenus par les deux méthodes a été réalisée en représentant une ACP sur trois échantillons d'entrées et trois échantillons de sorties. Le but de cette comparaison est de savoir si les différences sur les signaux détectés et leur nombre impactent les résultats de l'ACP. Les Figures 29A et 29B montrent les graphiques obtenus pour ces échantillons à partir des deux méthodes de retraitement des données.

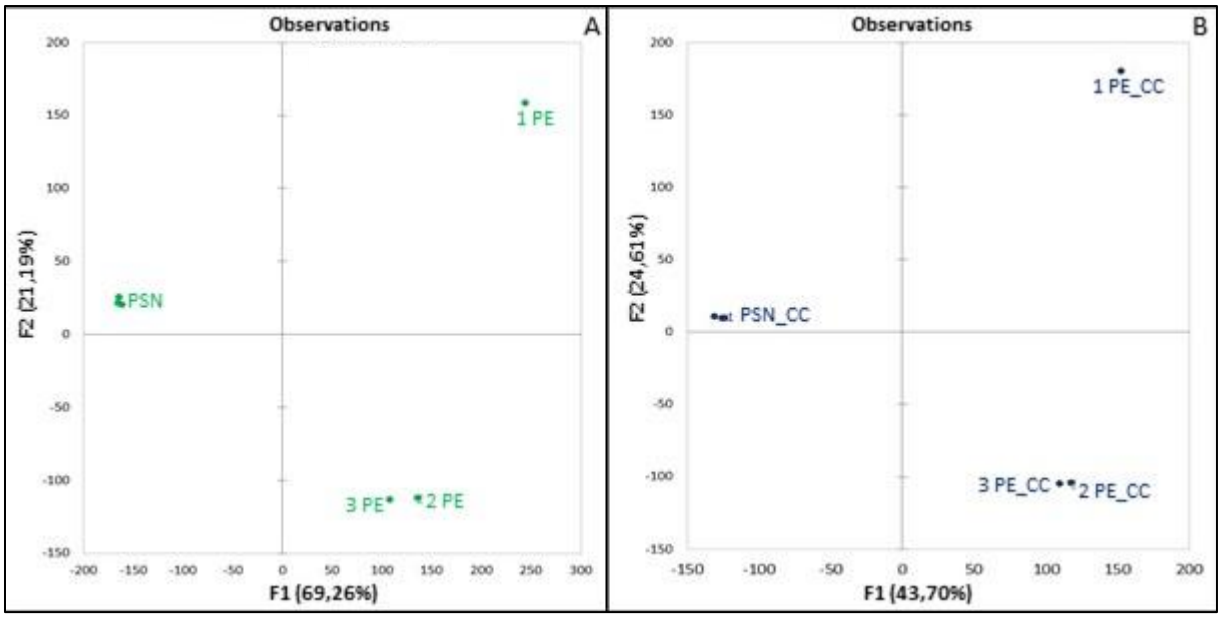


Figure 29 : Graphique des observations (échantillons) à partir des résultats de Compound Discoverer en A et de MZmine 2 en B. PE : entrée Poitiers, PSN : sortie Nord de Poitiers.

Ces figures montrent une séparation très similaire, quelle que soit la méthode utilisée. Cela signifie que malgré les différences au niveau des signaux détectés, les informations sous-jacentes aux tableaux de signaux aboutissent au même résultat : la distinction des entrées et des sorties d'une part, la distinction de l'échantillon d'entrée 1 avec les échantillons d'entrée avec 2 autres échantillons de cette même campagne C.

En conclusion, les deux méthodes rendent des résultats qui différent par le nombre de signaux détectés et la dispersion des signaux en fonction du temps de rétention, du m/z et des aires. Toutefois, les signaux les plus intenses sont en grand partie communs aux deux méthodes et la séparation entre les échantillons par ACP est semblable. Les deux méthodes de retraitement des données sont cohérentes au niveau des résultats de l'ACP et comportent une partie de leurs signaux communs mais de nombreux signaux différents, sans présumer de leur pertinence.

II.3.2. Choix de la méthode de retraitement des données

Afin de choisir un logiciel et par conséquent une méthode de prétraitement des données pour la suite de notre travail, des critères de sélections ont été choisis. Outre la qualité des données obtenues (e.g. nombre de signaux sélectionnés, dispersion de leur m/z, de leur temps de rétention et de leur intensité…), une attention particulière a été portée sur le temps de calcul pour obtenir ces données, la facilité de manipulation du logiciel, le nombre d'échantillons possibles par analyse, la stabilité du logiciel et la souplesse du logiciel.

Au regard de la distribution des gammes de m/z, ces deux méthodes pourraient être complémentaires. Toutefois, la comparaison des logiciels en termes de signaux a montré qu'une partie des signaux les plus intenses est commune aux deux méthodes. De plus, les deux méthodes de retraitement aboutissent aux mêmes tendances lors de la comparaison de groupes d'échantillons. Le temps de calculs et la facilité d'utilisation des logiciels ont donc été prépondérants pour faire notre choix.

Pour ces critères, les avantages et les inconvénients de chacune des méthodes ont été listés dans le Tableau 12. Dans ce tableau, non exhaustif, les points essentiels acquis par retour d'expérience sont présentés.

Tableau 12 : Avantages et inconvénients pour chacune des deux méthodes développées.

Logiciels	Avantages	Inconvénients
MZmine 2 – R – Excel	• Compatible avec tous les formats de données • Informations disponibles sur la plupart des algorithmes • Visibilité à chaque étape	• Temps de calcul long pour ≈ 30 échantillons (> 1 semaine) • Produits des artefacts (Verkh *et al*, 2018) • Blocage intempestif du logiciel avec un très grand nombre d'échantillons
Compound Discoverer	• Temps de calcul rapide pour un gros volume de données : ≈ 30h pour plus de 100 échantillons • Facilité d'utilisation • Couplage facile et rapide avec un grand nombre de bases de données dont celle de Thermo Scientific • Outils d'identification intégrés • Couplage avec des méthodes statistiques • Possibilité de faire de la recherche de produits de transformations directement à partir du logiciel	• Compatible uniquement le format raw • Peu d'informations sur les algorithmes • Manque de visibilité à chaque étape (Comportement « boîte noire »)

Parmi les différences majeures observées entre les deux méthodes, on note le format libre de MZmine 2 et le format propriétaire de Compound Discoverer qui le limite uniquement aux données Thermo Scientific. Une opacité concernant les algorithmes de traitements découle de ce format propriétaire. Lors de l'utilisation de Compound Discoverer, ce comportement « boîte noire » peut donc amener à des difficultés de réglages. Enfin, les mises à jour régulières ne sont pas toujours accessibles pour ce

dernier logiciel et le format propriétaire peut être limitant dans le cas d'essais inter-laboratoires avec des appareils différents.

Cependant, en dépit de son opacité par rapport à MZmine 2, Compound Discoverer présente des avantages très intéressants lors du traitement des données tels que le temps de calcul rapide par rapport au nombre d'échantillons étudiés, l'accessibilité à des bases de données, l'étude rapide des paramètres à régler, le « re-travail » en temps réel sur le tableau de données finale par des filtres et le couplage avec des méthodes statistiques prenant en compte les manipulations effectuées sur le tableau de résultats. En outre, les données peuvent être exportées dans Excel et d'autres traitements statistiques peuvent être facilement appliqués.

Ainsi, pour le reste de nos travaux, le logiciel Compound Discoverer a été conservé pour ses performances, sa facilité d'utilisation, sa rapidité d'utilisation, la possibilité de coupler l'analyse ciblée, la recherche de suspects et l'analyse non ciblée, l'intégration d'outils statistiques prenant en compte les opérations effectuées sur le tableau de résultats et enfin les modules servant à l'identification des signaux.

III. Étude de l'impact des répétabilités d'injection et de prélèvement

La mesure de la répétabilité d'une méthode permet de s'assurer de sa robustesse et de la fiabilité des résultats obtenus. Elle permet également d'estimer les biais de mesure et les faiblesses possibles de la méthode.

III.1. Impact de la répétabilité d'injection

La répétition des analyses réalisée en injectant plusieurs fois un échantillon, a pour objectif de minimiser les effets des variations de mesures, notamment les variations d'intensité. La répétabilité permet d'évaluer les erreurs dues à l'appareillage et est liée à la fidélité de la mesure. Dans le cadre de ce travail, la répétabilité de la méthode a été évaluée après application du traitement des données. L'impact sur les résultats de l'ajout de réplicas sur le nombre de signaux identifiés a été mesuré pour cinq types d'échantillons (correspondant au premier des trois prélèvements successifs de la campagne C) : un échantillon d'entrée de la STEP de Poitiers, un échantillon de la sortie de la STEP de Poitiers (sortie Nord), un échantillon du Clain au niveau du panache du rejet de la STEP, un de l'amont de la STEP ainsi qu'un échantillon du Clain prélevé en aval de la STEP. Dans cette partie, les coefficients de variation de l'aire des signaux ont été calculés en prenant en compte trois réplicas d'injection.

Les Figures 30 A et B représentent la répartition des signaux après traitement des données, en fonction des coefficients de variation obtenus pour les différents types d'échantillons étudiés.

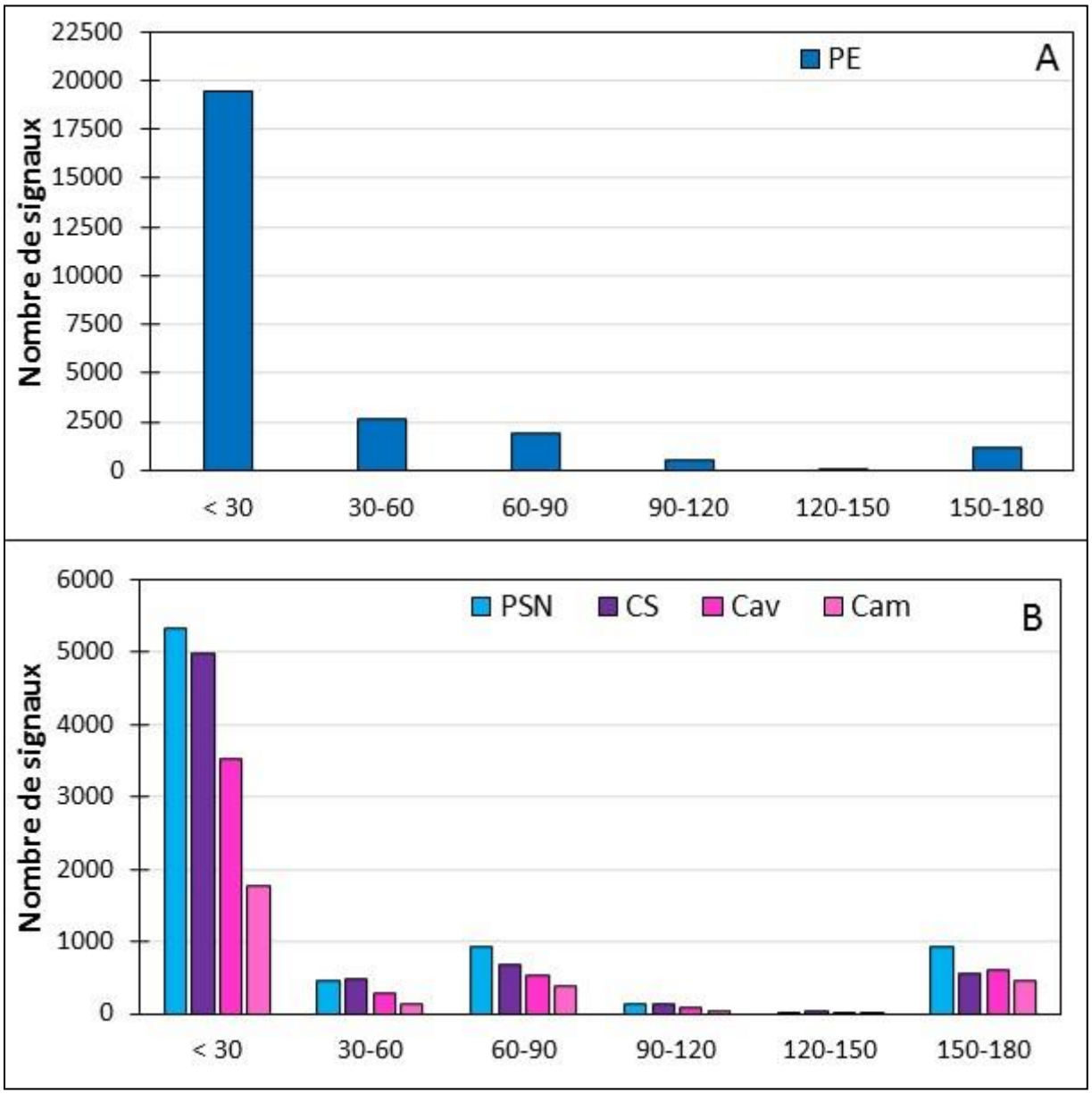


Figure 30 : Nombre de signaux en fonction d'intervalles de coefficients de variation (en %) selon le type d'échantillons (pour 3 injections). Données obtenues par Compound Discoverer. PE : entrée Poitiers, PSN : sortie Nord Poitiers, Cam : Clain amont, Cav : Clain aval, CS : Clain panache.

Ces résultats montrent que la majorité des signaux a un coefficient de variation inférieur à 30% pour tous les types d'échantillons étudiés. Le nombre de signaux décroit avec l'augmentation du coefficient de variation et les signaux très variables (présentant un coefficient de variation supérieur à 60%) ne représentent qu'une fraction limitée (< 31,5%) du total des signaux. Parmi ces signaux, ceux ayant un coefficient de variation supérieur à 150% correspondent à des signaux présents dans une injection d'un échantillon sur les trois. Les signaux présents dans deux injections sur trois ont un coefficient de variation compris entre 86% et 164% en fonction des valeurs d'aires dans les deux injections de

l'échantillon. Enfin, les signaux modérément variables (30-60%), constituent entre 5,2% pour l'amont et jusqu'à 10,2% pour l'entrée de STEP de l'ensemble des signaux.

Afin d'évaluer la réduction du nombre total de signaux en fonction du nombre d'injections, le nombre de signaux détectés après prétraitement des données a été relevé en prenant en compte le nombre d'injection : 1, 2 ou 3. Les résultats obtenus sont présentés dans le Tableau 13. Seuls les signaux présentant un coefficient de variation inférieur à 30% ont été conservés.

Tableau 13 : Nombre de signaux par matrice en fonction du nombre de réplicas d'injection (n = 1 à 3). A partir de la 2ème injection, seuls les signaux présentant un coefficient de variations < 30% ont été comptabilisés (données du premier prélèvement de la campagne C).

Nombre d'injections	Entrée STEP	Sortie STEP	Panache	Aval du Clain	Amont du Clain
1	24 632	6 912	6 311	4 476	2 389
2	20 102	5 751	5 271	3 755	1 923
3	19 412	5 325	4 988	3 529	1 773

Ces résultats montrent une diminution importante du nombre de signaux détectés avec la répétition des injections, plus particulièrement pour les échantillons d'entrées de STEP. La diminution du nombre de signaux avec l'ajout d'une injection supplémentaire s'étend de -16% pour l'aval à -20% pour l'amont du Clain. Avec deux injections supplémentaires, la baisse est comprise entre -21% et -26% La réduction du nombre de signaux est ainsi plus marquée lorsque trois réplicas sont injectés au lieu de deux. De plus, la diminution est moins importante entre deux et trois injections qu'entre une et deux injections. L'impact de l'ajout de réplicas semble donc s'atténuer avec le nombre d'injection. Ce tableau met également en avant la relation entre le nombre de signaux détectés et l'importance de la diminution constatée après injection de réplicas. En effet, plus il y a de signaux détectés, plus la réduction observée est grande lors de l'ajout de réplicas. Ces résultats sont cohérents avec les travaux d'autres auteurs qui ont déterminé que trois réplicas étaient satisfaisants pour apporter une amélioration de la qualité des données par réduction du nombre de faux positifs (Bader *et al.,* 2016 ; Nürenberg *et al.*, 2015). Dans le cas de l'analyse non ciblée, l'utilisation de réplicas amène la réduction de faux positifs, par l'élimination de pics artéfacts ou de pics appartenant au bruit.

Ainsi, au vu de ces résultats, trois réplicas analytiques ont été réalisés systématiquement dans la suite de ce travail. Et sur les signaux obtenus, seuls les signaux communs aux trois réplicas présentant un coefficient de variation < 30% ont été conservés.

III.2. Impact de la répétabilité de prélèvements

Afin d'étudier la répétabilité de prélèvement, trois échantillons prélevés au même moment ont été analysés pour chacune des matrices suivantes : entrée et sortie de la STEP, au niveau du panache du rejet de sortie de la STEP, Clain en amont et en aval de ce panache. Puis, les résultats des analyses obtenues pour chaque groupe d'échantillons ont été traités dans la même séquence de traitement des données par Compounds Discoverer. Seuls les signaux présentant un coefficient de variation inférieur à 30% sur l'ensemble des trois injections ont été conservés pour la comparaison entre les réplicas d'échantillons. La Figure 31 suivante présente les diagrammes de Venn obtenus dans ces conditions, avec le nombre de signaux communs pour les cinq matrices différentes. Les signaux considérés comme communs sont ceux présents dans les trois échantillons sans considération de leur variation d'intensité dans ces trois prélèvements.

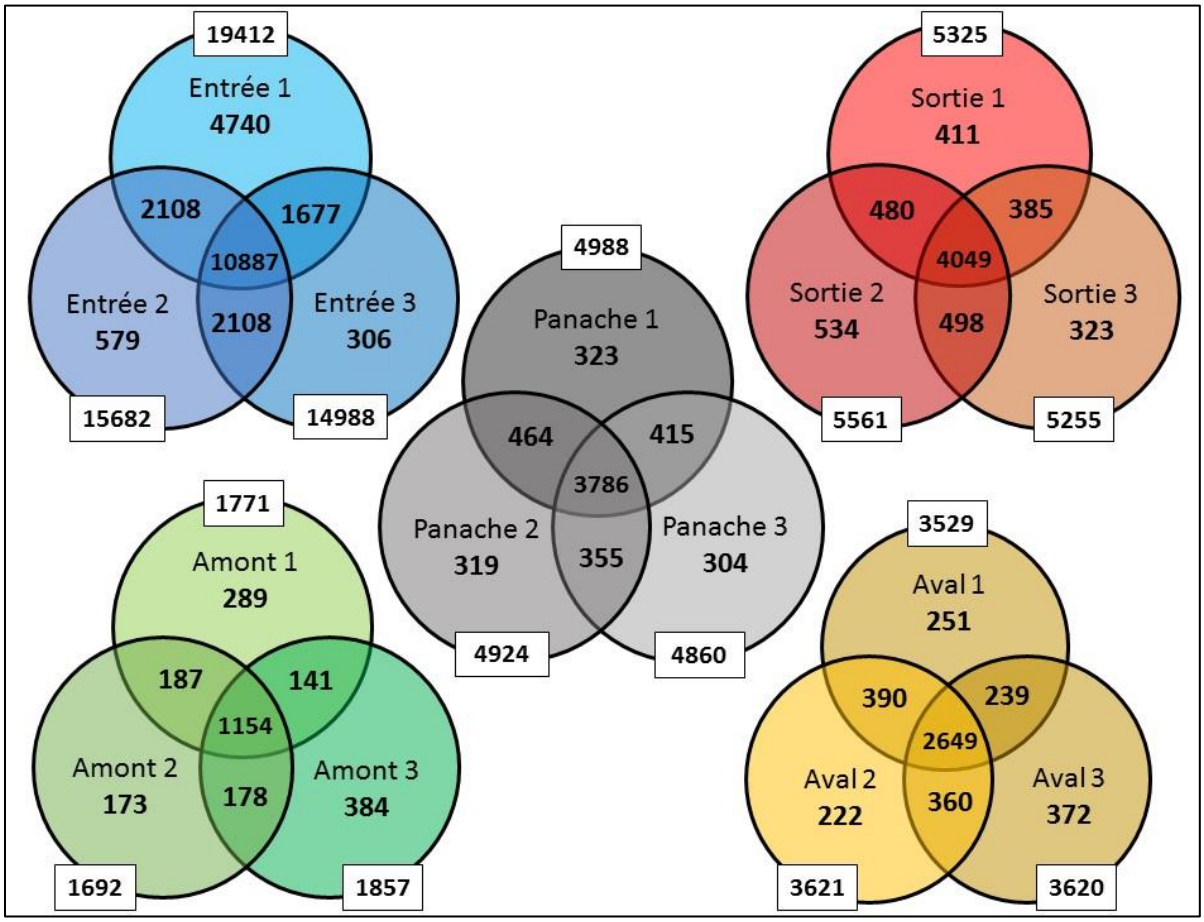


Figure 31 : Diagrammes de Venn illustrant le nombre de signaux communs entre les trois réplicas d'échantillons pour chaque type de matrice étudiée. Le nombre total de signaux de l'échantillon est inscrit dans le cadre blanc. Données traitées par Compound Discoverer.

La première observation tirée de cette figure est la proximité en nombre de signaux de chacun des réplicas. Le Tableau 14 présente les coefficients de variation du nombre de signaux par matrice. Les

valeurs de ces coefficients de variation sont comprises entre 1,3% et 14,3% ce qui est satisfaisant et montre la stabilité de l'ensemble de la méthode (de l'échantillonnage aux données traitées).

Tableau 14 : Nombre de signaux pour les trois prélèvements successifs (1, 2 et 3) selon la matrice (campagne C) avec nombre moyen, écart-type, coefficient de variation et estimation d'une incertitude relative par matrice.

	Entrée	Sortie	Panache	Aval	Amont
Prélèvement 1	19412	5325	4988	3529	1773
Prélèvement 2	15682	5561	4924	3621	1692
Prélèvement 3	14988	5255	4860	3620	1857
Moyenne	16694	5380	4924	3590	1774
Ecart-type	2379	160	64	53	83
CV (%)	14,3	3,0	1,3	1,5	4,7
Erreur relative retenue (%)	15	5	5	5	5

Les résultats obtenus montrent une valeur de pourcentage de signaux présents dans les trois échantillons par rapport à l'empreinte moyenne comprise entre 46,0% et 63,5% quelle que soit la matrice considérée. Les valeurs du nombre de signaux communs sont similaires que l'on considère deux à deux ou l'ensemble des échantillons de même type. Sur des prélèvements échantillonnés à la suite, la répétabilité de la méthode développée est donc satisfaisante si l'on considère ce critère.

En ce qui concerne les aires, la Figure 32 représente la répartition des signaux après traitement des données, en fonction des coefficients de variations obtenus pour les trois prélèvements successifs et les cinq matrices.

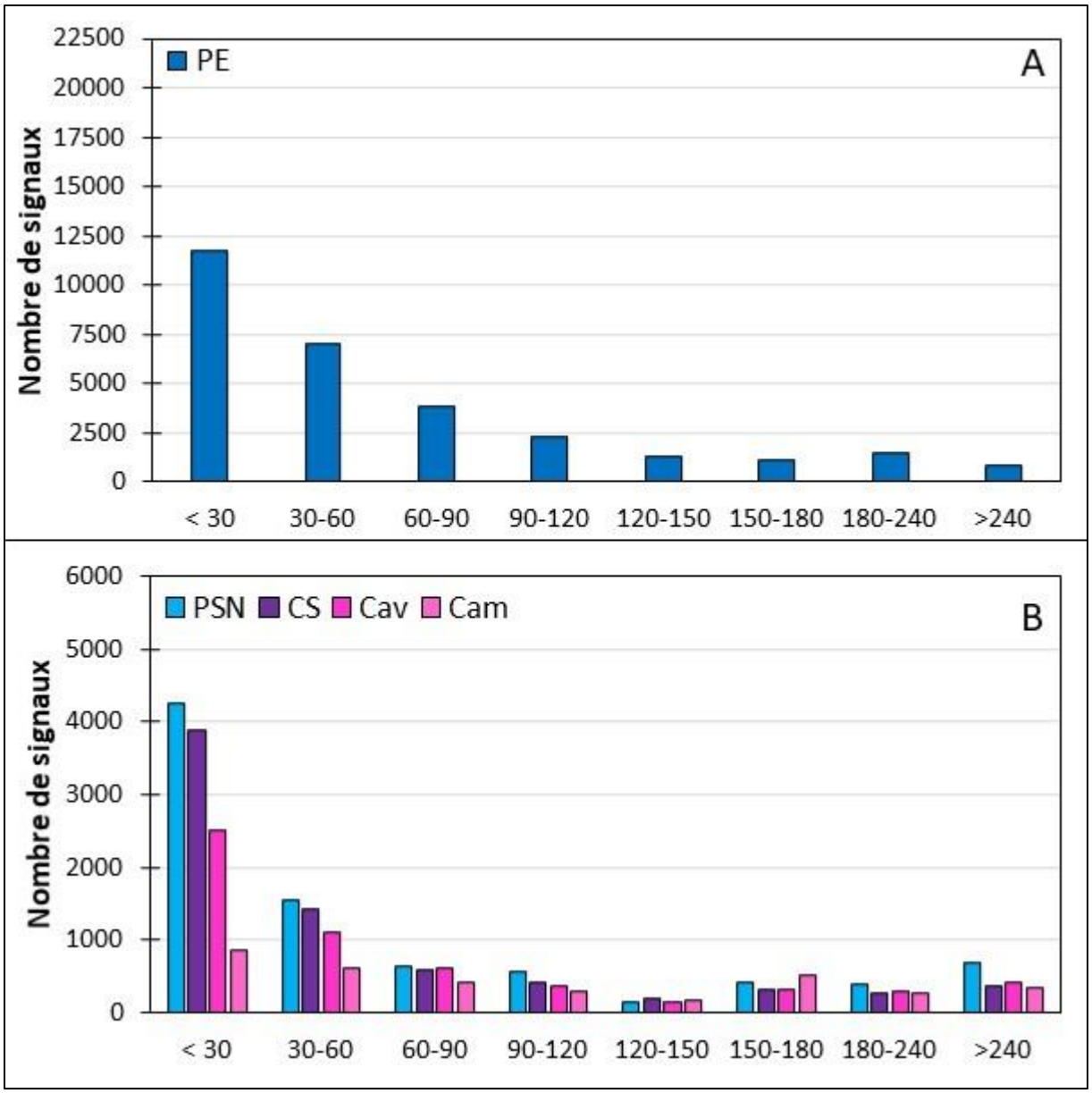


Figure 32 : Nombre de signaux en fonction d'intervalles de coefficients de variation (en %) selon le type d'échantillons. Données obtenues par Compound Discoverer pour les 3 prélèvements successifs (1, 2 et 3) par matrice (Campagne C). PE : entrée Poitiers, PSN : sortie Nord Poitiers, Cam : Clain amont, Cav : Clain aval, CS : Clain panache.

Pour tous les échantillons, les signaux ayant un coefficient de variation inférieur à 30% sur les trois échantillons sont majoritaires. Toutefois, les signaux ayant un coefficient compris entre 30% et 90% représentent également une partie importante du total de signaux détectés sur les trois échantillons, notamment dans le cas des entrées de STEP et de l'amont du Clain. Dans le cas de l'entrée de STEP, les données de la bibliographie montrent généralement une grande proportion de signaux non répétables lorsque la matrice est complexe (Bader *et al.*, 2016). Pour tous les échantillons, à l'exception du l'amont du Clain, les coefficients supérieurs à 90% constituent moins de 8% du total de signaux.

En conclusion, une reproductibilité de la méthode est donc observable lors de la réalisation de prélèvements successifs. Les nombres de signaux détectés sont proches mais des écarts sur les valeurs des aires des signaux peuvent être observés.

Pour la suite de ce travail, l'ensemble des signaux de ces trois prélèvements a été considéré pour la campagne C. Pour chaque signal commun aux trois prélèvements, la moyenne des aires a été calculée. Et pour chaque matrice, la moyenne du nombre de signaux et les écarts-types ont été utilisés. Ainsi, à partir de cette étude, une erreur relative sur le nombre de signaux pour chaque matrice a été estimée en arrondissant les valeurs des coefficients de variation. Ces valeurs (appelées « erreur relative retenue » dans le Tableau 14) ont été considérées comme constantes d'une campagne à l'autre pour une matrice donnée dans la suite de ce travail de thèse.

IV. Impact de l'application d'une méthode d'extraction/concentration des échantillons

L'extraction sur phase solide (SPE) est une méthode classique d'extraction et de concentration des analytes à partir d'une matrice donnée. Son utilisation dans le cadre de cette étude implique l'ajout d'une ou plusieurs étapes intermédiaires entre la filtration, effectuée dans tous les cas, et l'analyse (Figure 15). Les objectifs principaux de la SPE sont la concentration des molécules d'intérêts et la réduction des effets matrices. Toutefois, lors d'une analyse non ciblée, l'ensemble des signaux sont regardés et tous ne se comportent pas de la même manière face à la SPE, certains pouvant même disparaître. Lors d'une telle approche, il est donc important de s'interroger sur l'impact de la SPE sur la pertinence de son intégration dans la méthode.

IV.1. Choix de la méthode d'extraction

Classiquement, l'utilisation de la SPE passe par la sélection d'une seule cartouche composée d'une phase solide choisie en fonction des propriétés physico-chimiques des molécules à concentrer. Dans le cas de l'analyse non ciblée, l'objectif est à la fois de concentrer les échantillons pour augmenter la sensibilité et de diminuer les effets matrices tout en conservant le maximum de contaminants avec un large spectre de propriétés physico-chimiques. Dans ces conditions, les méthodes SPE utilisées dans la littérature se basent sur l'utilisation d'une (Sjerps *et al.*, 2016) ou plusieurs phases solides (Moschet *et al.*, 2015 ; Schollée *et al.*, 2018 ; Stephan *et al.*, 2016). Les méthodes utilisant un ensemble de 4 ou 5 phases solides regroupent un large panel d'interactions possibles (hydrophiles, lipophiles, anioniques, cationiques ou encore interactions π-π) afin de toucher une large gamme de polarité. Ces méthodes demandent en revanche plus de matériel en l'absence de cartouche multicouche dédiée à l'analyse non ciblée et coûtent plus cher à mettre en œuvre. Pour ces raisons, dans ce travail, nous avons choisi

d'employer deux phases solides échangeuses d'ions fortes (cationique et anionique), les Oasis® MCX et MAX (voir chapitre 2 : « Matériels et méthodes »). L'utilisation de ces phases, issue des travaux de Deeb *et al.* (2016) a été appliquée par Stephan *et al.* (2016) en analyse non ciblée. À ces cartouches, a été associée une troisième cartouche mixed-mode Oasis® HLB, dont l'utilisation est usuelle pour les analyses ciblées multi-résidus. La cartouche Oasis® HLB est effectivement capable de retenir les composés acides, basiques, neutres, polaires et apolaires ce qui permet de retenir des molécules passées aux travers des autres cartouches notamment en cas de dépassement de la capacité de rétention des phases. Le facteur de concentration théorique de la méthode d'extraction/concentration a été de 125.

IV.2. Impact de la méthode d'extraction sur un mélange d'étalons et détermination des rendements d'extraction

Pour évaluer l'impact de la méthode d'extraction sur les résultats, la méthode SPE comprenant l'ensemble de trois cartouches Oasis® (MCX, MAX et HLB dans cet ordre) a été testée sur un ensemble d'étalons à la fois dans l'eau ultra-pure et dans les quatre matrices étudiées (Clain amont, Clain aval, sortie de STEP et entrée de STEP). Deux expériences ont été réalisées : la première expérience a été de doper chacune des matrices après extraction par un mélange d'étalons à 1 µg/L afin d'évaluer les effets matrices lors de l'analyse par UHPLC-HRMS. La seconde expérience a consisté à ajouter un mélange d'étalons d'une concentration de 0,5 µg/L à tous les types d'échantillons avant de réaliser le protocole de la SPE. Le but de cette seconde expérience a été d'estimer les rendements d'extractions et deux séries ont été effectuées pour cela.

IV.2.1. Estimation de l'effet matrice

L'effet matrice est un phénomène qui va impacter l'ionisation en spectrométrie de masse : l'environnement chimique dans lequel se trouve une molécule va entraîner une amplification ou une surpression de l'ionisation et donc du signal. Cet effet matrice est molécule-dépendante et matrice-dépendante. Pour mesurer cet effet, un mélange de 24 étalons a été ajouté après extraction dans des échantillons d'eau ultra-pure, d'entrée et de sorties Nord de la STEP de Poitiers ainsi que de l'amont et de l'aval du Clain. L'effet matrice a été calculé selon l'équation (2) (voir III.2.2. du chapitre 2 : « Matériels et méthodes »)

Une valeur inférieure à 100% signifie une diminution du signal tandis qu'une valeur supérieure à 100% signifie une augmentation du signal. Les aires dans les matrices ont été corrigées au préalable par les aires initialement présentes dans les échantillons (i.e. avant dopage). Pour certains étalons, dans les échantillons de STEP, les effets matrices calculés sont aberrants en raison de leur présence initiale dans

les échantillons à des concentrations supérieures à celle ajoutée : c'est le cas pour le paracétamol, la metformine, la caféine et le diclofénac. Les Figures 33 A à D montrent les effets matrices calculés pour l'entrée et la sortie Nord de la STEP de Poitiers et pour l'amont et l'aval du Clain.

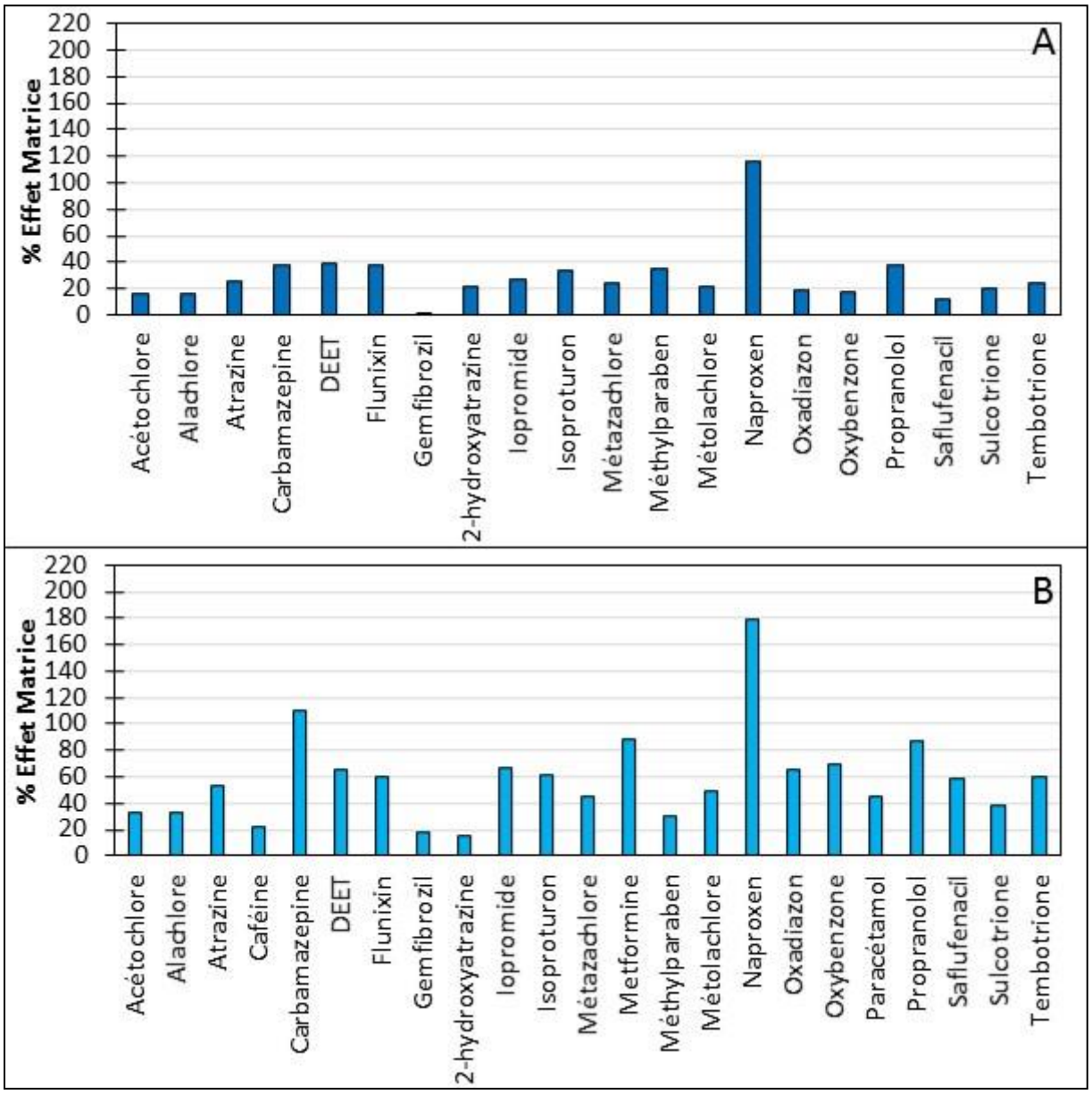


Figure 33 (A et B) : Diagrammes en barres représentant l'effet matrice en pourcentage en fonction des étalons pour l'entrée de STEP en A et la sortie de STEP en B. Effet matrice calculée selon le rapport de l'aire du signal pour l'étalon dans la matrice sur l'aire obtenue pour l'eau ultra-pure. Concentration des étalons : 1 µg/L.

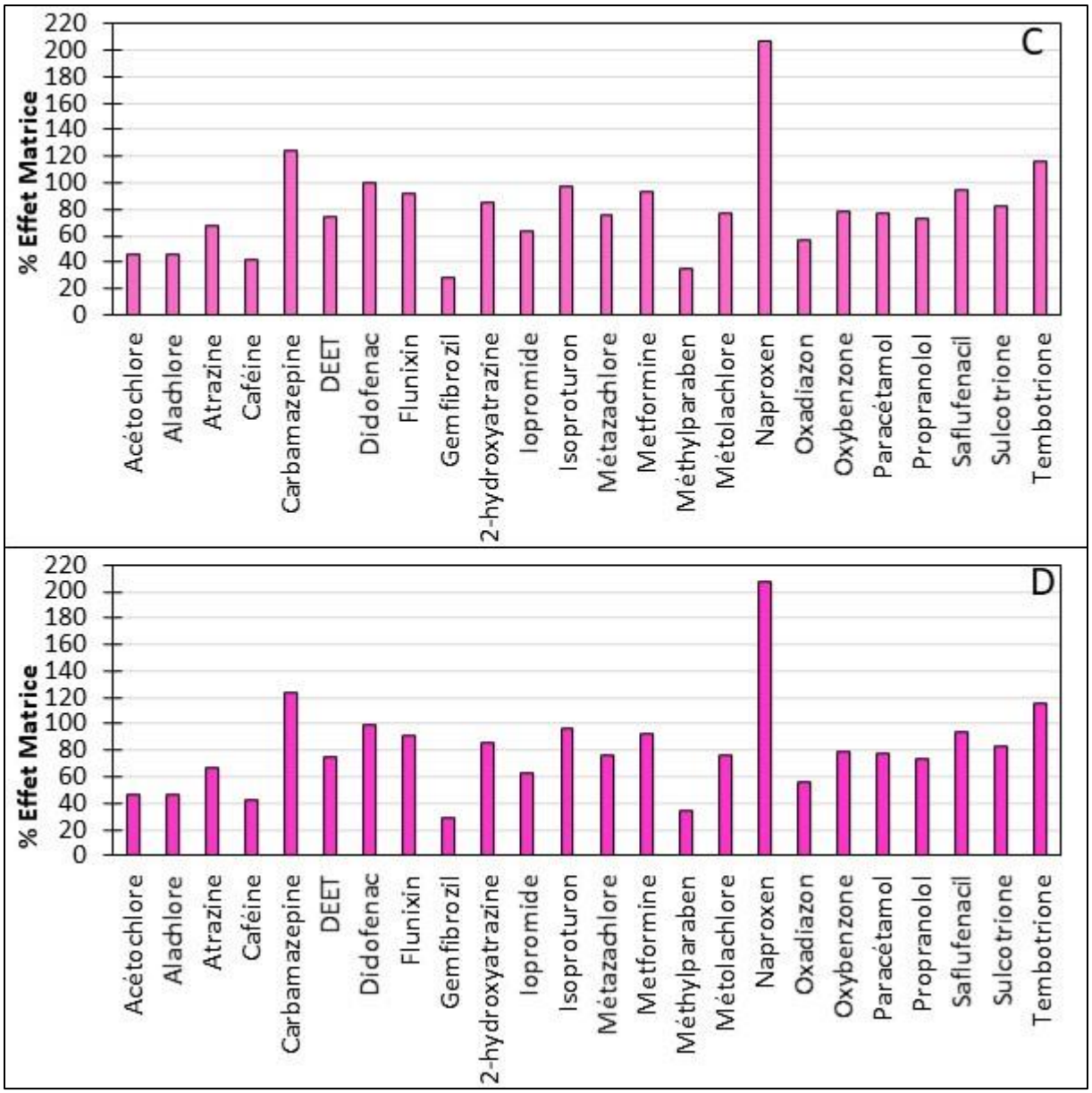


Figure 33 (C et D) : Diagrammes en barres représentant l'effet matrice en pourcentage en fonction des étalons pour l'amont du Clain en C et l'aval du Clain en D. Effet matrice calculée selon le rapport de l'aire du signal pour l'étalon dans la matrice sur l'aire obtenue pour l'eau ultra-pure. Concentration des étalons : 1 µg/L.

Ces figures confirment la présence d'un effet matrice pour tous les types d'échantillons étudiés. L'importance de cet effet dépend de la matrice considérée. En entrée de STEP, comme attendu, l'effet matrice est en majorité plus prononcé avec un pourcentage moyen de 30%. Le naproxen est impacté par un effet matrice entrainant une amplification du signal alors que, à l'inverse, la réponse du gemfibrozil est presque absente (1%). Pour la sortie de STEP, le pourcentage moyen de l'effet matrice est de 60%, soit une diminution de moins de la moitié des aires des étalons. Le naproxen et le gemfibrozil sont là encore les étalons subissant respectivement la plus forte diminution et la plus forte

augmentation du signal. Enfin, les effets matrices pour les échantillons prélevés dans le Clain (amont et aval, avec des effets moyens de 80%) sont moins importants que pour les échantillons de STEP.

Dans le cadre de ce travail, des comparaisons entre les entrées et les sorties de STEP ainsi qu'entre l'amont et l'aval du Clain ont été effectuées. Les rapports des aires mesurées dans les échantillons entre l'entrée et la sortie de la STEP et l'amont et l'aval du Clain ont également été évalués afin de mettre en évidence la nécessité de la prise en compte de l'effet matrice dans les comparaisons (Figure 34).

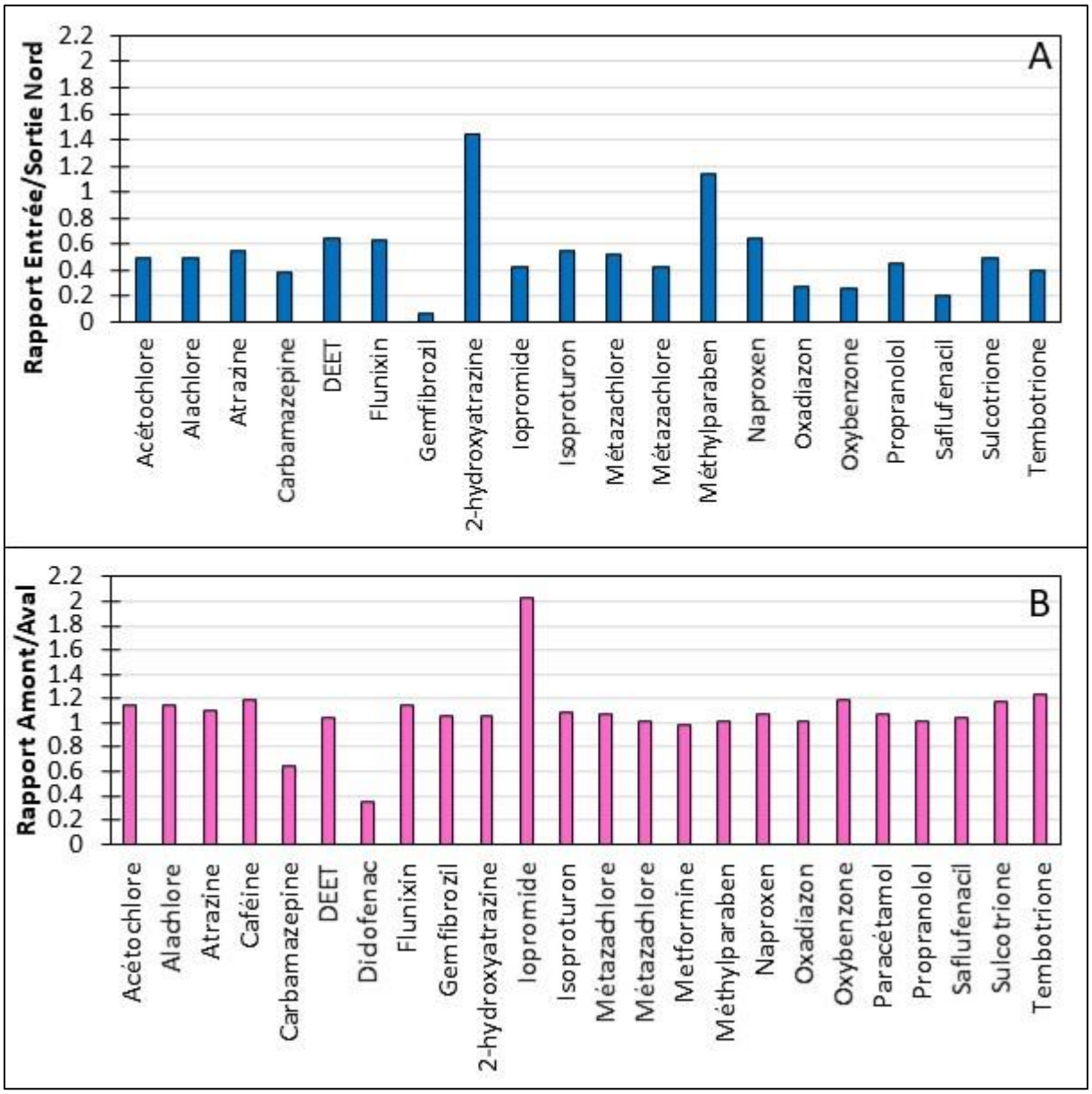


Figure 34 : Diagrammes en barres du rapport des aires des étalons entre l'entrée et la sortie Nord en A et l'amont et l'aval du Clain en B.

Comme suggéré par la figure précédente, les effets matrices en entrée et en sortie de STEP sont différents. En moyenne, le rapport entre ces deux matrices est de 0,48 soit un effet près de 2 fois plus

important en entrée (Figure 34A). À l'exception de trois étalons (carbamazépine, diclofénac, iopromide), les effets matrices sont similaires entre les échantillons prélevés en amont et en aval du Clain : la valeur moyenne du rapport entre les aires mesurées dans les deux matrices est de 1,1 (Figure 34B).

Ainsi, la comparaison entre l'amont et l'aval du Clain ne nécessite pas en règle générale de correction de l'effet matrice. Toutefois, la différence d'effet matrice en STEP altère la qualité de la comparaison entre l'entrée et la sortie de la STEP. La compensation des effets matrices en analyse ciblée peut être réalisée par trois moyens : la dilution des échantillons avant analyse, l'injection d'un plus faible volume d'échantillon ou l'utilisation d'étalons internes. L'application de ces méthodes en analyse non ciblée a plusieurs désavantages. La dilution avant injection réduit le nombre de signaux détectés : par exemple, seuls 52% des signaux en entrée et 41% en sortie ont été conservés pour une dilution par 3 dans l'étude de Nürenberg *et al.* (2015). L'utilisation d'étalons internes requiert des molécules dont le temps de rétention et la structure sont identiques ou proches de ceux de l'analyte subissant l'effet matrice. En analyse non ciblée, la structure des molécules est inconnue et rend le choix d'un étalon interne difficile. En choisir un de façon arbitraire pourrait avoir un effet contraire à celui recherché (Nürenberg *et al.*, 2015). L'extrapolation de l'effet matrice à des signaux inconnus est donc difficile et sans garantie d'amélioration des résultats (Schollée *et al.*, 2016). De plus, le nombre d'étalons internes nécessaires pour couvrir l'ensemble des temps de rétention implique un coût très élevé.

Dans le cadre de ce travail, sur la base des résultats obtenus pour les différents étalons étudiés, le choix a été fait de n'appliquer une correction que sur les signaux communs aux entrées et aux sorties à l'aide d'un facteur correctif de 2. Ce choix nous est apparu comme étant un bon compromis au vu des résultats de la Figure 34. Toutefois, il est à noter que, selon Nürenberg *et al.* (2015), ceci peut mener à la sous-estimation du nombre de signaux éliminés dans les échantillons ayant subi le traitement des eaux et à une légère surestimation des molécules pouvant être formées durant le procédé de traitement.

IV.2.2. Rendements d'extraction de la méthode SPE

Les rendements d'extraction de la SPE ont été évalués à partir d'un mélange d'étalons ajouté pour obtenir des concentrations à 0,5 µg/L (avant extraction), dans les cinq matrices suivantes : eau ultra-pure, Clain amont et aval, entrée et sortie Nord de la STEP de Poitiers. Dans ces conditions, les rendements ont été calculés à partir de l'équation (2) (voir III.2.3. du Chapitre 2 : « Matériels et méthodes »). Et pour chaque calcul, les aires dans les matrices étudiées ont été corrigées par les aires des mêmes molécules sans ajout. Les résultats sont présentés dans les Figure 35A et 35B. Ces résultats

sont la moyenne des rendements calculés avec cinq étalons internes différentes (2-hydroxyatrazine d5, atrazine d5, caféine d9, carbamazépine d10 et métolachlore d6).

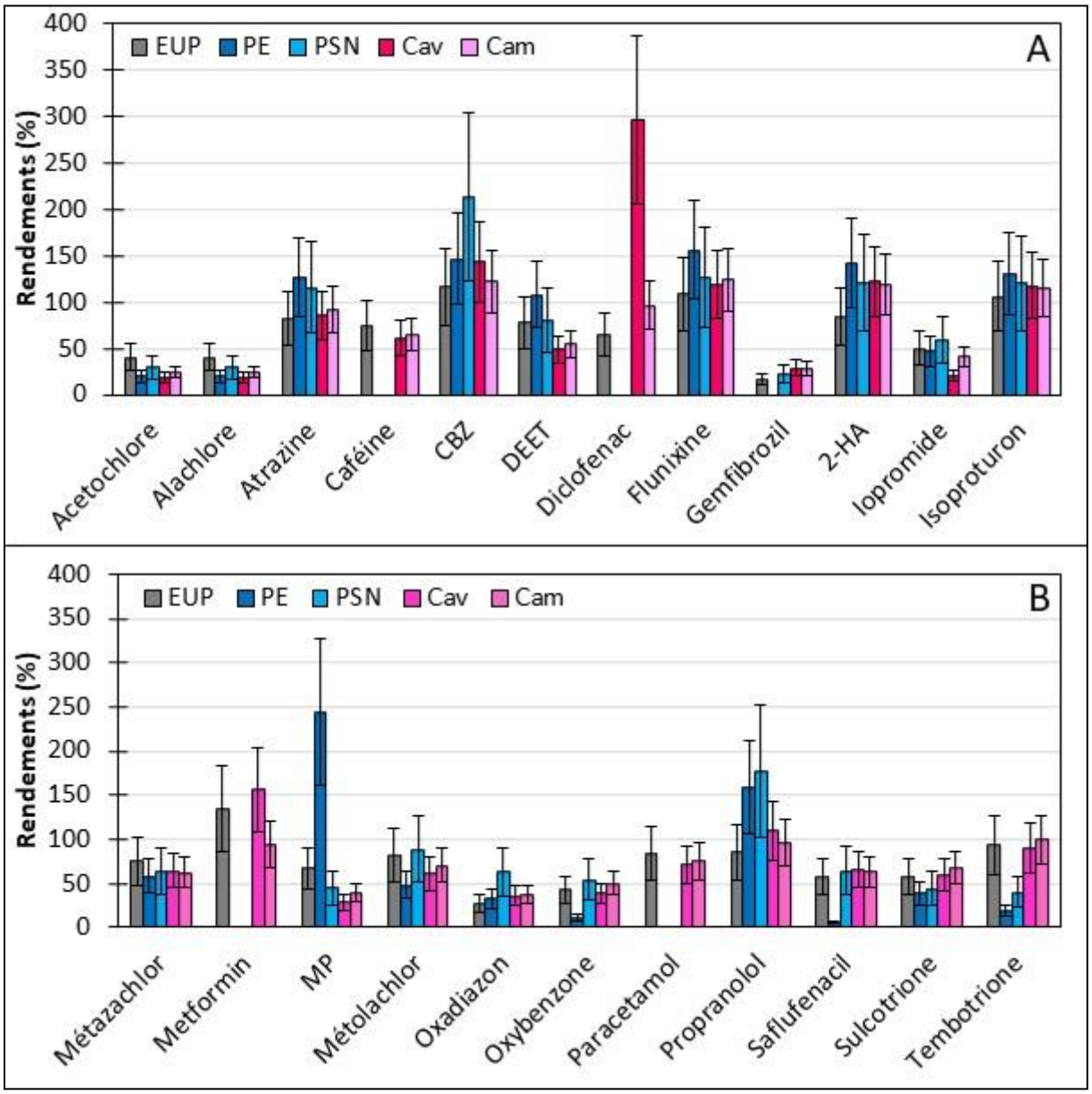


Figure 35 : Diagrammes en barres des rendements d'extraction dans l'eau ultra-pure, le Clain amont et aval en A, l'eau ultra-pure, l'entrée et la sortie de STEP de Poitiers en B. CBZ : Carbamazépine, 2-HA : 2-hydroxyatrazine, MP : Méthylparabène, EUP : eau ultra-pure, Cam : Clain amont, Cam : Clain aval, PE : entrée STEP de Poitiers, PSN : sortie Nord STEP de Poitiers.

Les Figures 35A et 35B montrent des rendements variables pour les molécules suivant la matrice considérée. Dans le cas des échantillons de STEP, la caféine, le diclofénac, le paracétamol et la metformine ont été retirés des graphiques en raison de rendements aberrants (>1000%) vraisemblablement dus à la présence initiale de ces composés dans la matrice à des concentrations plus importantes que les concentrations ajoutées. Le naproxen n'a également pas été représenté sur les graphiques en raison de rendement supérieur à 300% dans l'ensemble des matrices pour une raison

inconnue. A partir de ces résultats, il peut être noté que les rendements d'extraction obtenus sont généralement comparables d'une matrice à l'autre. Les rendements sont supérieurs à 60% pour la majorité des molécules quelle que soit la matrice (i.e. pour 66% des étalons pour le Clain, 54% pour l'entrée de STEP et 73% pour la sortie de STEP).

Dans le cadre de ce travail, il n'a pas été possible d'utiliser un étalon interne spécifique pour chaque molécule pour compenser les effets matrice. Ceci constitue une limite à cette étude des rendements d'extraction. De plus, les barres d'erreurs représentant l'écart-type entre les rendements avec les différents étalons internes montrent d'importantes variations pour la plupart des étalons du mélange. Néanmoins, ces résultats restent corrects pour une méthode directement inspirée de la bibliographie sans optimisation, et qui utilise trois cartouches successives qui augmentent les probabilités de perte d'échantillon au cours de la préparation.

Bien entendu, dans cette partie, les rendements portent sur un nombre limité de molécules connues et ne présage pas de l'impact de la SPE sur la globalité de l'analyse non ciblée. Cette approche donne toutefois un premier aperçu de l'impact de l'extraction pour les différentes matrices.

IV.3. Comparaison des résultats obtenus pour l'ensemble des signaux avec et sans extraction par SPE

L'objectif de l'analyse non ciblée est d'étudier les échantillons de manière la plus globale possible. Il est donc nécessaire de limiter les pertes potentielles de signaux dans les empreintes à chacune des étapes de la méthode. Dans cet objectif, il a été entrepris une évaluation de l'impact de la SPE sur la quantité des signaux gagnés et perdus, et également sur l'intensité des signaux. Pour cela, les empreintes obtenues lors de l'analyse, avant et après extraction (facteur de concentration théorique de 125), pour les échantillons d'entrée, de sortie de STEP et du Clain (en amont des rejets de la STEP) ont été comparées pour les campagnes A, B et C. Pour chacune des empreintes de ces prélèvements, chaque signal a été classé, selon la modification de sa valeur d'aire de pic lors de l'analyse après SPE par rapport à celle obtenue sans SPE, dans l'une des cinq catégories suivantes : apparition, augmentation, stabilité, disparition ou diminution. Dans cette classification, un signal a été considéré comme stable, dès que le rapport des aires des pics entre l'échantillon concentré et l'échantillon non concentré était compris entre 0,5 et 2.

Les empreintes d'échantillons de la campagne C concentrés obtenus dans ces conditions sont présentées dans la Figure 36. Dans cette figure, les signaux qui ont disparu pendant la SPE, bien qu'absents de l'échantillon, ont été inclus dans l'empreinte par nécessité de comparaison.

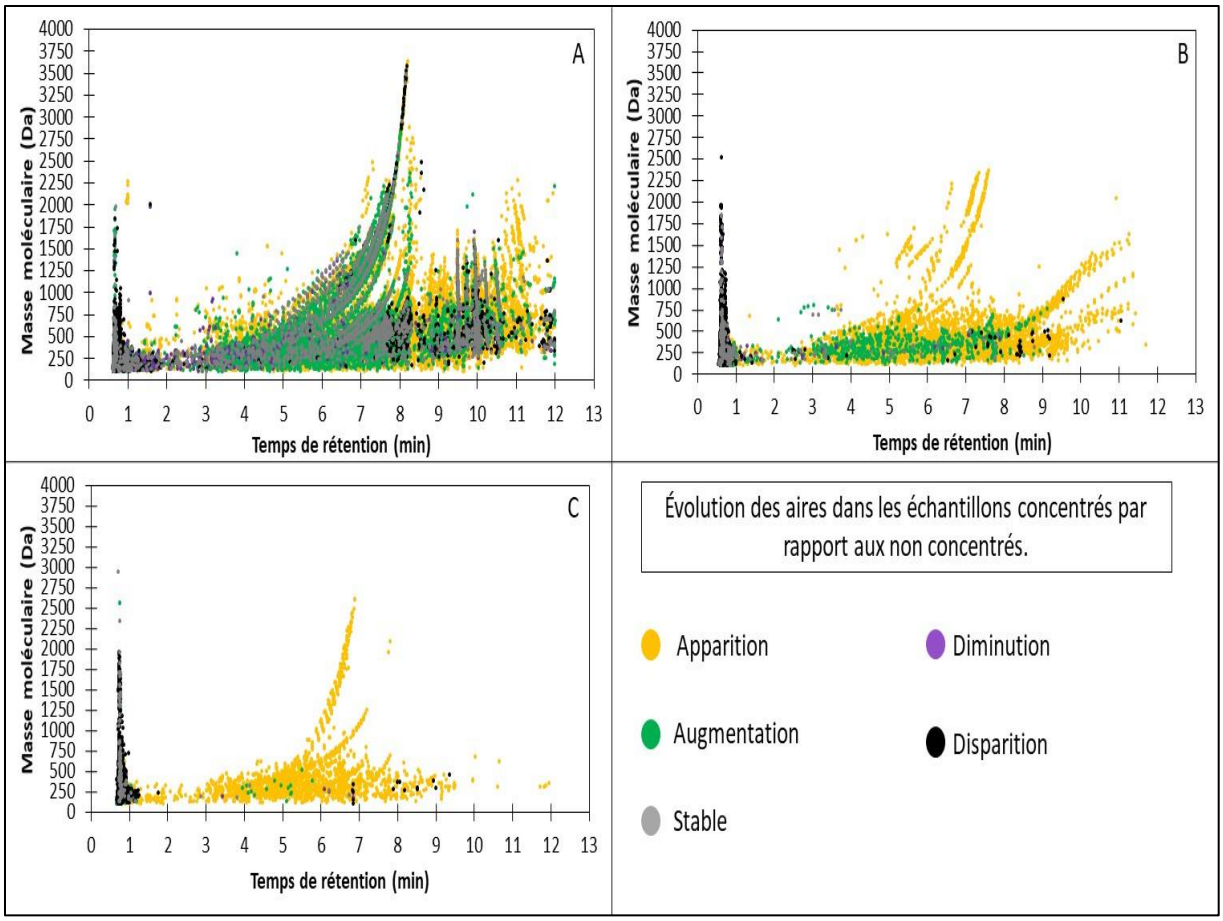


Figure 36 : Empreinte d'échantillons concentrés de l'entrée et de la sortie de la STEP de Poitiers en A et en B et de l'amont du Clain en C. Prélèvements issus de la campagne C. La couleur des points représente l'évolution des aires des signaux par rapport aux aires obtenues pour l'échantillon non concentré correspondant. Empreinte réalisée par Compound Discoverer après analyse par UHPLC-HRMS.

Ces résultats montrent en entrée de STEP (Figure 36A) que :

- Dans leur majorité, les signaux détectés dans l'échantillon concentré sont soit des nouveaux signaux soit des signaux qui ont vu leur aire augmentée par rapport à l'échantillon non concentré. Ces signaux sont dispersés sur l'ensemble de l'empreinte.
- Certains signaux ont disparu ou présentent une aire du pic qui diminue lors de la SPE. Ces deux catégories de signaux sont plutôt disséminées dans l'empreinte. Néanmoins, une concentration plus importante semble observée en début d'analyse.
- Les signaux dits stables sont également présents et distribués le long de l'empreinte.

Le rapport moyen du nombre de signaux apparaissant ou augmentant sur le nombre de signaux disparaissant ou diminuant est de 8,4 ± 0,3. Ce rapport confirme l'observation visuelle de l'empreinte et la mise en évidence de l'impact attendu de la SPE sur cet échantillon. Si l'on prend en compte les signaux stables en plus de ceux diminuant ou disparaissant, ce ratio est de 5,0 ± 0,2.

Pour l'échantillon de sortie de la STEP de Poitiers (Figure 36B) :

- Pour la plupart, les signaux détectés après application de la SPE ont augmenté ou sont apparus tandis qu'un nombre limité d'autres signaux ne sont plus présents ou ont diminué.
- Le rapport du nombre de signaux apparaissant ou augmentant sur le nombre de signaux disparaissant ou diminuant est de 4,9 ± 0,1.
- Une nouvelle fois, les molécules qui ont disparu ou ont diminué semblent majoritairement regroupées au début de l'empreinte (pour les faibles temps de rétention). Contrairement à l'empreinte de l'entrée (Figure 36A) ce phénomène semble également observé dans le cas des signaux stables.

Enfin dans le cas de l'échantillon du Clain (Figure 36C), on peut faire des observations similaires à celles faites sur la sortie : les signaux apparaissant sont répartis le long de l'empreinte tandis que les signaux disparaissant sont situés au début de l'empreinte. Le rapport du nombre de signaux apparaissant ou augmentant sur le nombre de signaux disparaissant ou diminuant est de 1,8 ± 0,1. Cette valeur est plus faible que les autres valeurs de rapports. Ceci peut s'expliquer par le nombre plus important de signaux assez polaires présents dans ces échantillons vis-à-vis du faible nombre de signaux moins polaires. Les molécules très polaires sont en effet reconnues comme étant les moins bien concentrées par SPE.

Les nombres de signaux pour chaque catégorie ont été déterminés et sont répertoriés dans un tableau en Annexe 2 pour chaque point de prélèvement, et les facteurs de concentrations moyens ont été calculés dans chaque matrice pour les signaux qui augmentent : 18 en entrée et en sortie de STEP et de l'ordre de 5 en amont du Clain. Ces facteurs sont bien inférieurs au facteur de concentration théorique de 125. Les rendements calculés à partir d'étalons dans le paragraphe précédent ont également montré que la récupération des composés n'est pas parfaite et donc que le facteur de concentration est inférieur au facteur théorique. Toutefois, en raison des effets matrices pouvant changer entre les échantillons (concentrés et non concentrés) et variables suivant les signaux, il est difficile de réellement conclure si ces facteurs sont représentatifs ou non.

Bien entendu, une partie des constatations effectuées entre la méthode sans extraction et la méthode avec extraction est également dépendantes de la colonne de chromatographie UHPLC qui favorise la rétention des molécules moyennement polaires et apolaires.

Ainsi, au vu de ces résultats, quelle que soit la matrice étudiée, la méthode SPE semble affecter des signaux correspondant à une large gamme de masse et de polarité en cohérence avec l'utilisation de trois cartouches différentes. La comparaison des échantillons concentrés et non concentrés montre un impact majoritairement positif de la SPE et cohérent avec les objectifs de l'analyse non ciblée c'est-

à-dire l'augmentation de la sensibilité de la méthode avec une perte de signaux aussi faible que possible. Cet impact positif (i.e. augmentation ou apparition de signaux) est valable pour toutes les matrices d'intérêt étudiées. Toutefois, il l'est d'autant plus pour les échantillons de rivières qui, sans SPE, n'apporteraient que très peu d'informations. L'impact négatif de la SPE (i.e. diminution ou perte de signaux) semble porter essentiellement sur les signaux très polaires. Ce phénomène est en accord avec la non-rétention de ce type de molécules sur les cartouches. En conclusion, les impacts positifs de la SPE semblent cohérents avec ce qui était attendu. En dépit de la faiblesse de cette méthode SPE sur les composés fortement polaires, cette étape a donc été conservée pour la suite des travaux. Ainsi, les données montrées dans les chapitres suivants seront toutes issues de l'analyse d'échantillons concentrés.

V. Conclusions

Deux méthodes reposant sur l'utilisation de MZmine 2 (complété par R 3.5.1 et Excel) et de Compound Discoverer ont été développées afin de traiter les données brutes issues de l'analyse UHPLC-HRMS. Le développement de ces deux méthodes a été effectué à l'aide de différents types d'échantillons (i.e. entrées et sorties de STEP, panache au niveau du rejet, et Clain prélevés en amont et en aval de la STEP) et d'eau purifiée dopés par un mélange d'étalons à différentes concentrations allant de 50 ng/L à 1 µg/L. Les paramètres principaux de chaque étape des « workflows » de ces logiciels ont été testés à l'aide des données brutes de ces échantillons afin d'en optimiser les méthodes. Pour optimiser le réglage de ces paramètres, un compromis a été recherché pour limiter le nombre de faux négatifs, particulièrement pour les faibles concentrations, et le nombre de faux positifs. Les paramètres apparus comme les plus importants ont été les seuils d'intensité ainsi que les tolérances sur les temps de rétention et sur le rapport m/z. En termes de résultats, au vu de la distribution des gammes de m/z, les deux méthodes optimisées semblent complémentaires. Toutefois, la comparaison des logiciels a montré qu'une grande partie des signaux les plus intenses est commune aux deux méthodes. Pour la suite des travaux, le logiciel Compound Discoverer a été conservé en raison de ses performances satisfaisantes, en termes de rapidité de temps de calculs, de sa facilité et de sa souplesse d'utilisation, du couplage possible entre les trois types d'analyse (analyse ciblée, analyse de suspects et analyse non ciblée) et enfin de sa liaison aisée avec des bases de données variées.

Lors de l'utilisation de cette méthode de traitement des données, l'injection de réplicas analytiques a montré une diminution du nombre de faux positifs par l'élimination de signaux dont l'intensité est décelée comme trop variable. A partir de ce travail, un critère seuil de 30% du coefficient de variation a été défini et conservé comme critère de sélection des signaux pour la suite des travaux. La robustesse de la méthode a également été vérifiée par son application sur trois échantillons de même nature

prélevés au même moment. Dans ces conditions, un pourcentage de signaux communs compris entre 46,0 et 63,5% pour l'ensemble des matrices a été observé avec des coefficients de variations sur le nombre de signaux par échantillon satisfaisants (compris entre 1,5 et 15%). Lors de ce travail, il a été constaté que les échantillons d'entrées de STEP, en raison de leur complexité étaient les matrices les plus sujettes aux variations.

Pour terminer, l'application d'une méthode d'extraction sur phase solide à partir de trois phases différentes (échangeuse d'anion, échangeuse de cation et mixed-mode) a été étudiée. Les rendements d'extraction déterminés pour un mélange d'étalons sont globalement compris entre 50 et 75% en moyenne selon les matrices. La perte de molécules parmi les plus polaires peut être attendue. Néanmoins, cette méthode s'est révélée être intéressante vis-à-vis des objectifs de l'analyse non ciblée pour l'ensemble des matrices d'intérêt. Pour la suite de ce travail, la méthode SPE a donc été conservée pour concentrer les échantillons et les données obtenues après application de cette méthode ont été privilégiées lors de l'étude des résultats des campagnes de prélèvements. Pour tenir des effets matrices pour la comparaison des signaux communs entre les entrées et les sorties de STEP, un facteur de correction de 2 a été observé sur les aires des signaux des entrées.

Chapitre 4 : Caractérisation des eaux usées brutes et traitées par analyse non ciblée et recherche de suspects

I. Introduction

Le traitement des eaux usées répond à la nécessité de dépolluer ces eaux avant de les restituer au milieu naturel. Ce processus passe par plusieurs étapes de traitement dont les objectifs sont l'élimination d'une fraction de la matière organique, de la matière azotée voire des phosphates. Les procédés mis en œuvre impliquent des réactions chimiques et biochimiques permettant la dégradation et la métabolisation de la matière organique présente dans les eaux usées. Les eaux traitées rejetées doivent ainsi répondre à des normes portant essentiellement sur la demande chimique ou biochimique en oxygène, les matières en suspension, l'azote et le phosphore dans les zones sensibles. Ces normes fixent des contraintes de concentrations de rejet et de rendements épuratoires minimaux qui dépendent de la taille de la station d'épuration (STEP). Les eaux rejetées doivent ainsi avoir une qualité proche de celle du milieu récepteur et induire un impact aussi faible que possible sur ce milieu. Or, si les stations d'épuration répondent aux normes, les critères de qualité ne renseignent que très partiellement sur les changements s'opérant dans les eaux et sur la nature des molécules en sortie de STEP. La problématique des micropolluants et des molécules inconnues constitue un domaine d'étude récent. De même, à ce jour, les produits de transformations des micropolluants au cours des traitements des eaux connaissent un intérêt croissant. Pour apporter des éléments de réponses à ces interrogations, l'analyse non ciblée pourrait être un outil adapté notamment pour caractériser les eaux et les comparer, mettre en valeur des signaux pertinents, rechercher des suspects et des métabolites.

Dans ce chapitre, seront étudiées les eaux de deux stations d'épuration, situées dans les villes de Poitiers et de St Julien l'Ars (département de la Vienne). Dans un premier temps, les eaux d'entrées de ces stations seront caractérisées afin de dresser un état initial des eaux avant traitement. Puis, les eaux de sorties seront elles aussi caractérisées avant d'être comparées aux eaux d'entrées. Dans un second temps, des corrélations entre la nature de la matrice et les signaux détectés par analyses multivariées seront effectuées. Puis, des signaux reliés à des impacts potentiels de la STEP sur le milieu récepteur seront recherchés. Ces signaux seront nommés « signaux pertinents » dans la suite de ce travail. Enfin, l'application d'une analyse de suspects sera détaillée pour rechercher des micropolluants et des produits de transformations : l'étape de création de la liste sera décrite, les résultats de l'analyse dans plusieurs matrices seront présentés et une estimation semi-quantitative avec l'étude de l'évolution de l'intensité des signaux sera effectuée sur les suspects identifiés. Pour terminer, une conclusion récapitulera l'ensemble des résultats obtenus dans ce chapitre.

II. Caractérisation et comparaison des eaux d'entrée de STEP

La première étape de l'étude des eaux de STEP a porté sur la caractérisation des eaux d'entrées de STEP afin d'établir l'état des lieux avant les procédés d'épuration. L'utilisation de la méthode d'analyse non ciblée précédemment développée puis l'exploitation et l'étude des données obtenues sont mises en œuvre dans cet objectif.

Les eaux des deux stations d'épuration (i.e. STEP de Poitiers et STEP de St Julien l'Ars) ont ainsi été caractérisées. Pour cela, deux et trois campagnes de prélèvements ont été réalisées respectivement au niveau de la STEP de Saint Julien et de celle de Poitiers. Lors de chaque campagne, trois points de prélèvement ont été considérés au niveau de chaque STEP (i.e. 1 point en entrée et 2 points en sortie, Nord et Sud, au niveau de la STEP de Poitiers et, 1 point en entrée, 1 point en sortie et 1 point au niveau de la lagune pour Saint Julien l'Ars). Au niveau de la STEP de Poitiers, les trois échantillons prélevés successivement lors de la campagne C ont pour but d'étudier la répétabilité de la méthode et son impact sur les résultats. Le Tableau 4 du chapitre « Matériels et méthodes » récapitule ces échantillonnages. Les données HRMS d'analyse des eaux des deux STEP ont été traitées, par Compound Discoverer, séparément afin d'éviter l'ajout de faux positifs potentiels dus à des différences entre les matrices.

II.1. Comparaison du nombre total de signaux par échantillon

Après analyse et obtention des tableaux de données, la première information rapidement accessible est le nombre de signaux en fonction de l'échantillon. Ce nombre a été déterminé à partir de la liste des signaux obtenue par Compound Discoverer en retirant ceux dont les aires de pic sont 5 fois inférieures à celles du blanc. La Figure 37 visualise ces nombres par échantillon pour les eaux d'entrées de STEP des deux sites de prélèvements.

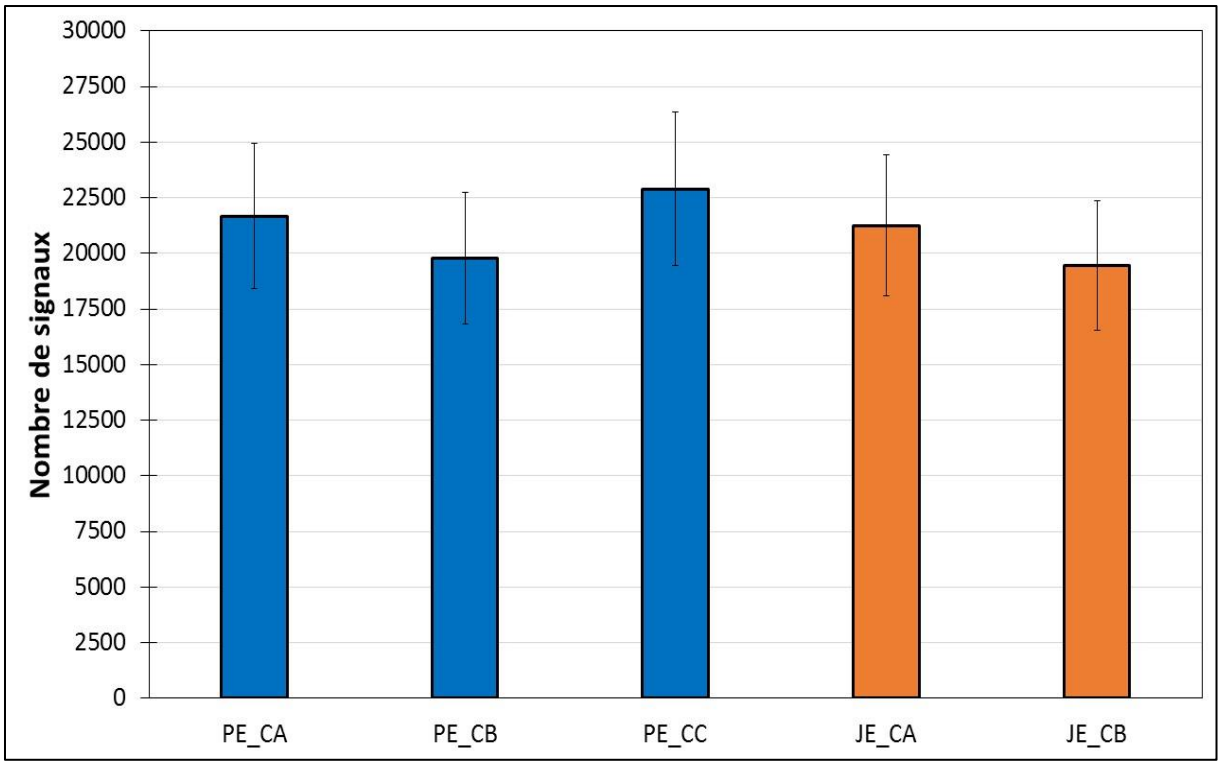


Figure 37 : Graphique en barres représentant le nombre de signaux en fonction de l'échantillon considéré. Les échantillons d'entrée de la STEP de Poitiers (PE) et de la STEP de St Julien l'Ars (JE) sont respectivement en couleur bleu et en couleur orange.

Cette figure montre un nombre de signaux semblables (21 450 ± 1 570) pour l'ensemble des échantillons. Le nombre de signaux mesuré sur les échantillons d'entrée (Poitiers et Saint Julien l'Ars) lors de la campagne B est cependant de l'ordre de 9 à 14 % plus faible que pour les deux autres, ce qui est peu significatif au regard des incertitudes.

II.2. Empreintes des eaux d'entrée de STEP

Pour la caractérisation des eaux d'entrée de STEP, les empreintes de chacun des échantillons ont été établies pour avoir une vision globale de la répartition des signaux (seuls les signaux dont les aires de pic sont 5 fois supérieures à celles du blanc ont été conservés). Ces empreintes permettent de comparer le type de molécules présentes dans les échantillons en termes de polarité, de taille et d'intensité des signaux. Cette comparaison des empreintes entre les échantillons peut aussi permettre de détecter d'éventuelles disparités entre les prélèvements.

Les Figures 38 et 39 présentent les empreintes des entrées de Poitiers et de St Julien l'Ars.

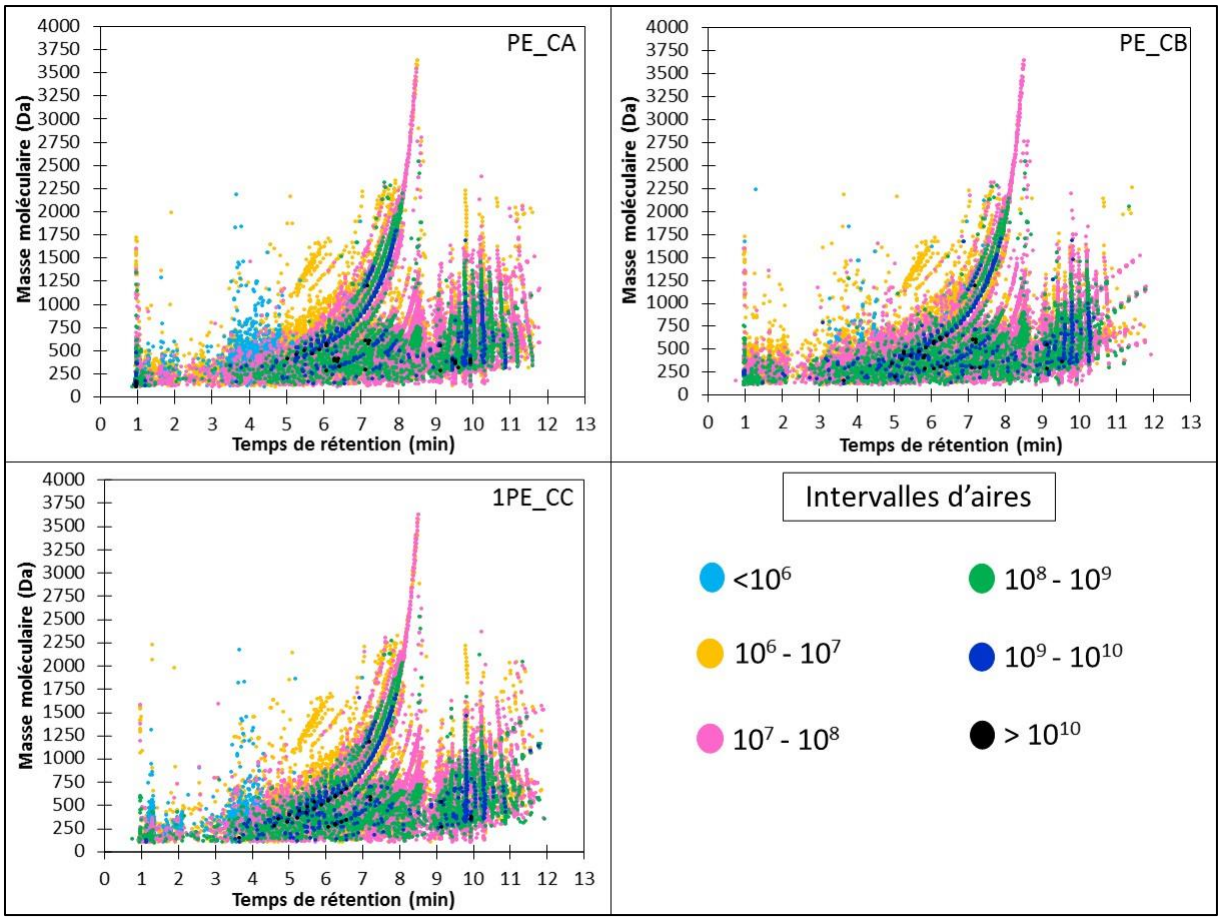


Figure 38 : Empreintes des échantillons d'entrées de la STEP Poitiers pour les campagnes A, B et C. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. PE : entrée Poitiers, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI +).

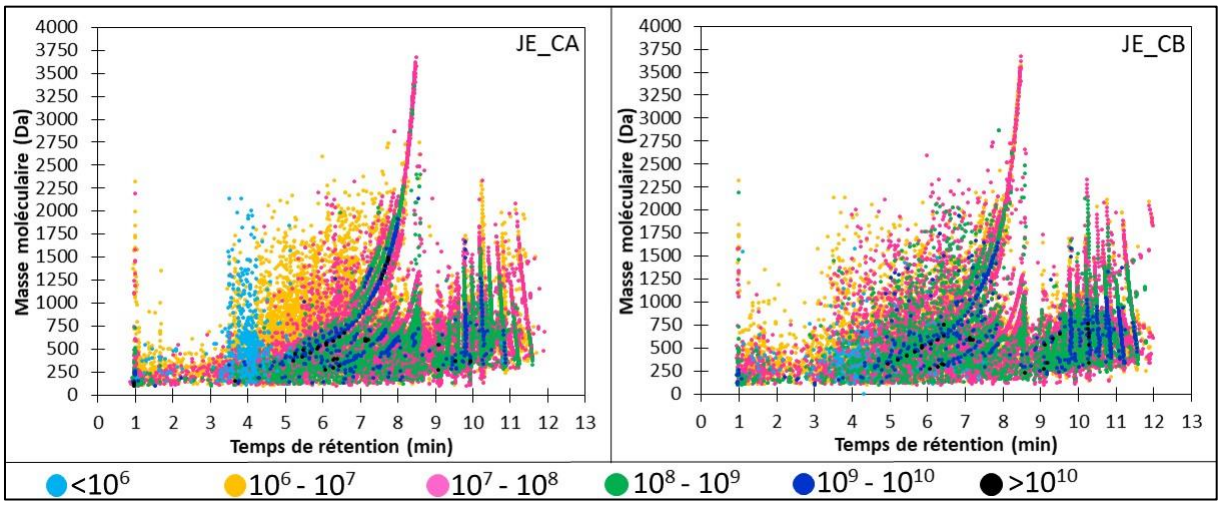


Figure 39 : Empreintes des échantillons d'entrées de la STEP de St Julien l'Ars pour les campagnes A et B annotées. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. JE : entrée St Julien l'Ars, CA : campagne A, CB : campagne B. Mode positif (HESI +).

Ces figures montrent une répartition des signaux détectés sur l'ensemble de la durée d'analyse. Toutefois, la majorité des signaux est détectée après 4 minutes pour les deux STEP. Afin de relier les

temps de rétention observés avec la gamme de polarités des molécules associées aux signaux, les logP ont été déterminés pour des étalons analysés. La Figure 40 suivante présente les chromatogrammes d'ions extraits pour cinq étalons couvrant la majorité de la durée d'analyse. Dans ces conditions, la metformine et l'oxadiazon sont respectivement le premier et le dernier des étalons détectés.

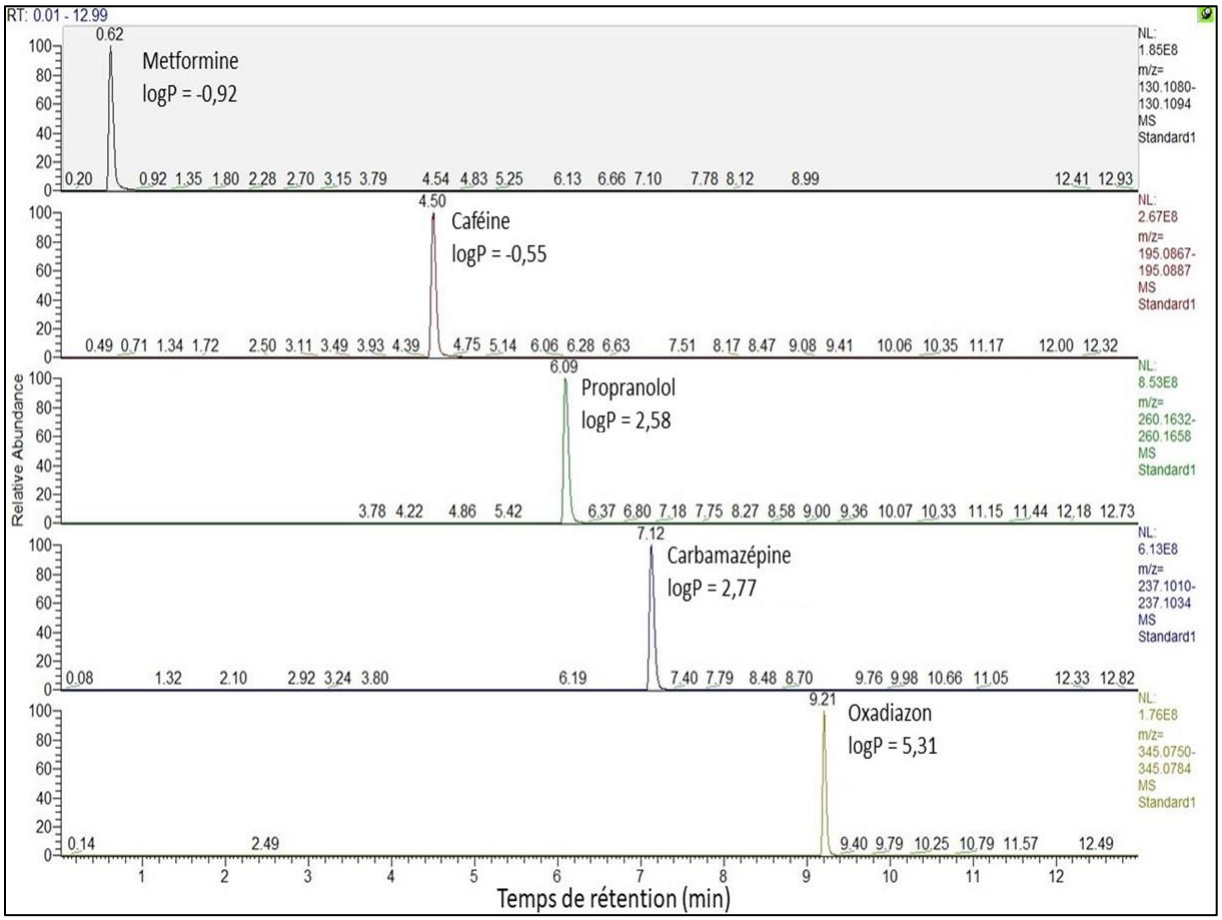


Figure 40 : Chromatogrammes des ions moléculaires des étalons de metformine, caféine, propranolol, carbamazépine et oxadiazon issus de l'analyse d'un mélange d'étalons à une concentration de 100 µg/L.

Les valeurs de logP associées aux temps de rétention indiquées sur cette figure montrent que les composés très polaires (logP < -1) ne sont pas retenus par la colonne (en accord avec sa nature de type C18). La majorité des signaux détectés est donc plutôt apolaire. Néanmoins au regard des valeurs concernant la metformine et la caféine, une partie des composés polaires reste détectable par l'analyse.

Les empreintes montrent une large palette de masses moléculaires allant de 100 Da à presque 4 000 Da. Pour les deux STEP, entre 5 minutes et 8 minutes, plusieurs ensembles de points se suivent les uns et les autres, formant comme des motifs, avec des masses de plus en plus importantes. Ces enchainements de signaux peuvent correspondre à des homologues, c'est-à-dire à des molécules possédant une structure apparentée avec un motif qui se répète. Les surfactants et la famille

polyéthylène glycol (PEG) font partie de ce type de molécules (Thurman *et al.*, 2014). Une recherche d'homologues a été effectuée sur ces empreintes. Les résultats sont présentés en Annexe 3. Ainsi les différences de masse 14,02 ; 44,03 et 58,04 Da correspondent respectivement aux groupements CH_2, C_2H_4O et CH_2+C_2H_4O, la différence de 44,03 Da pouvant typiquement correspondre à des PEG. En dehors de ces enchainements assez visibles, d'autres homologues peuvent également être présents dans le reste de l'empreinte. La variété des masses moléculaires indique la présence d'un éventail de composés de familles différentes pouvant aller des acides aminés aux protéines et aux polymères.

Sur les Figures 38 et 39, la représentation des intervalles d'intensité, par les diverses couleurs, révèle que les signaux les moins intenses, en turquoise, semblent se concentrer au début des empreintes. Les signaux les plus intenses (>10^9) semblent plutôt présents en milieu et fin d'empreintes. Dans le cas de l'échantillon de la campagne B de St Julien l'Ars, la répartition des intervalles d'aires semble différente par rapport aux résultats obtenus pour l'échantillon de la première campagne, notamment pour les temps de rétention compris entre 3 et 7 minutes. En effet, beaucoup de signaux de ces temps de rétention sont présents sur la Figure 39B avec un intervalle d'aires de niveau supérieur à celui des signaux de la même zone sur la Figure 39A (ainsi beaucoup de signaux en jaune en A semblent en rose en B ou encore des signaux rose en A sont en vert en B). Pour Poitiers, l'échantillon de la campagne B montre également, dans une moindre mesure, des changements de même nature pour des temps de rétention autour de 4 min : les aires inférieures à 10^6 y sont moins présentes que dans les empreintes relatives aux prélèvements des campagnes A et C. Ces empreintes sont à mettre en perspectives avec le tableau ci-dessous des données physico-chimiques de caractérisation de ces eaux (Tableau 15).

Tableau 15 : Paramètres physico-chimiques de caractérisation des eaux en entrée de STEP (Poitiers et Saint Julien l'Ars, pour les campagnes A, B et C).

	pH	Conductivité (µS/cm 25°C)	COD (mg/L)	NT (mg/L)	UV 254 nm (sur 1 cm)
Poitiers					
Campagne A	8,13	3 800	27,5	78,0	0,749
Campagne B	8,30	635	19,2	34,6	0,322
Campagne C	8,28	1 709	42,6	62,9	0,622
Saint Julien l'Ars					
Campagne A	8,09	2 105	38,1	104,8	1,058
Campagne B	8,05	434	19,4	15,3	0,357

De la même manière, les valeurs de ces paramètres montrent une certaine singularité de la campagne B vis-à-vis des deux autres campagnes d'échantillonnage. Les eaux des entrées des deux STEP apparaissent moins concentrées lors de la campagne B que lors des deux autres (conductivité et charge organique significativement plus faibles). La plus forte dilution des échantillons de la campagne B, générant probablement moins d'effets matrices lors des analyses par spectrométrie de masse pourrait donc en partie expliquer les différences observées sur les empreintes.

Ainsi, Les empreintes des trois campagnes sont globalement très similaires à la fois par la répartition des signaux et la distribution des aires de pics, soulignant une part de stabilité dans la composition de ces échantillons. Les différences observées des empreintes des échantillons de la campagne B pourraient être liées au caractère plus dilué des échantillons de cette campagne pouvant générer moins d'effet matrice lors des analyses UHPLC-HRMS. Par ailleurs, les empreintes des entrées de STEP de Poitiers et de St Julien l'Ars ont également une allure proche. On peut noter toutefois la présence de plus de signaux entre 3 et 7 minutes pour les échantillons provenant de St Julien l'Ars.
Bien entendu, ces empreintes sont toujours incomplètes en raison de l'impossibilité de la méthode d'analyse de détecter l'ensemble des molécules présentes dans les échantillons (fonction de la nature de la colonne UHPLC). Notons d'autre part que des faux positifs peuvent être présents dans les résultats en dépit des différents filtres mis en place pour leur réduction. La composition de ces empreintes conserve donc un degré d'incertitude à prendre en compte lorsqu'on les étudie plus en détail. Ces empreintes témoignent cependant de la complexité chimique des eaux usées en entrée de STEP.

III. Impact du traitement : comparaison des eaux d'entrée et de sortie de STEP

L'objectif de l'épuration des eaux usées est de dépolluer les eaux avant de les restituer au milieu naturel. Les rendements journaliers des abattements en demande biochimique en oxygène (DBO) et en demande chimique en oxygène (DCO) constituent deux mesures d'efficacité des STEP généralement employées. Ces rendements sont liés à l'abattement des matières organiques par le procédé d'épuration. Pour les STEP de Poitiers et de St Julien l'Ars les rendements épuratoires attendus sont supérieurs respectivement à 80% et à 70%. Ces mesures reflètent l'action du traitement des eaux mais elles ne renseignent pas ou très peu sur les changements pouvant s'opérer à plusieurs niveaux : polarité, taille et concentration des molécules présentes ou encore production de nouveaux composés par le procédé de traitement.

Ainsi, pour tenter de répondre à ces interrogations, les échantillons de sorties ont tout d'abord été caractérisés. Puis, les données relatives aux entrées et aux sorties de la STEP de Poitiers et de St Julien l'Ars ont été confrontées ensemble dans le but de mettre en évidence :

- l'impact du traitement sur le nombre de signaux et les statistiques descriptives des échantillons,
- les types de molécules sur lesquels le traitement des eaux agit en priorité,
- l'existence d'écarts entre les sorties Nord et Sud de Poitiers,
- des différences entre les eaux de STEP de Poitiers et de St Julien l'Ars.

III.1. Caractérisation des eaux de sortie de STEP

Dans un premier temps, les eaux de sorties de STEP ont été caractérisées par leurs empreintes. Les Figures 41 et 42 montrent les résultats obtenus, pour les sorties Nord et Sud de Poitiers ainsi que pour la sortie et la lagune de St Julien l'Ars pour chacune des campagnes.

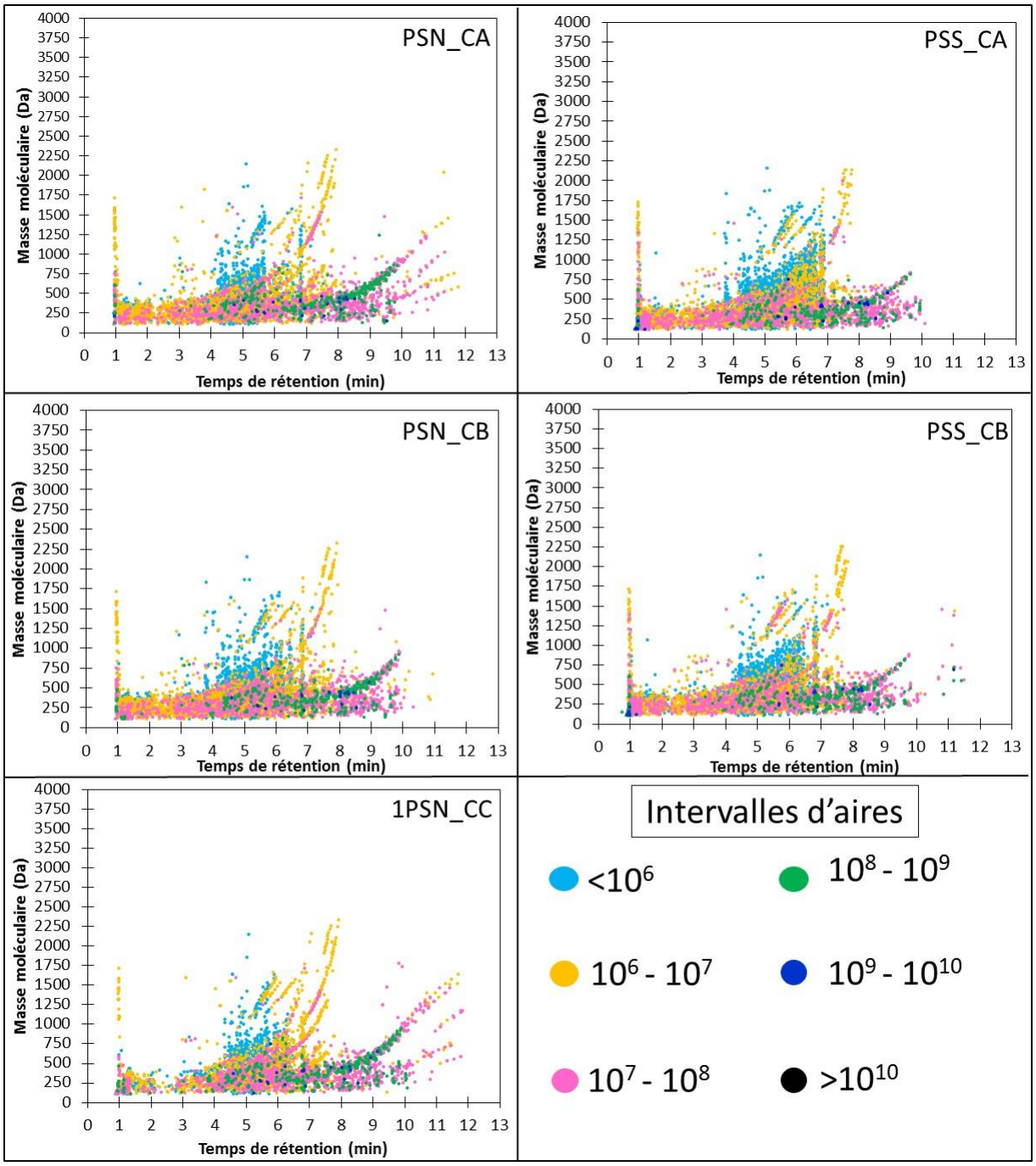


Figure 41 : Empreintes des échantillons de sorties de la STEP de Poitiers pour les campagnes A, B et C. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. PSN : sortie Nord Poitiers, PSS : Sortie Sud Poitiers, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI +).

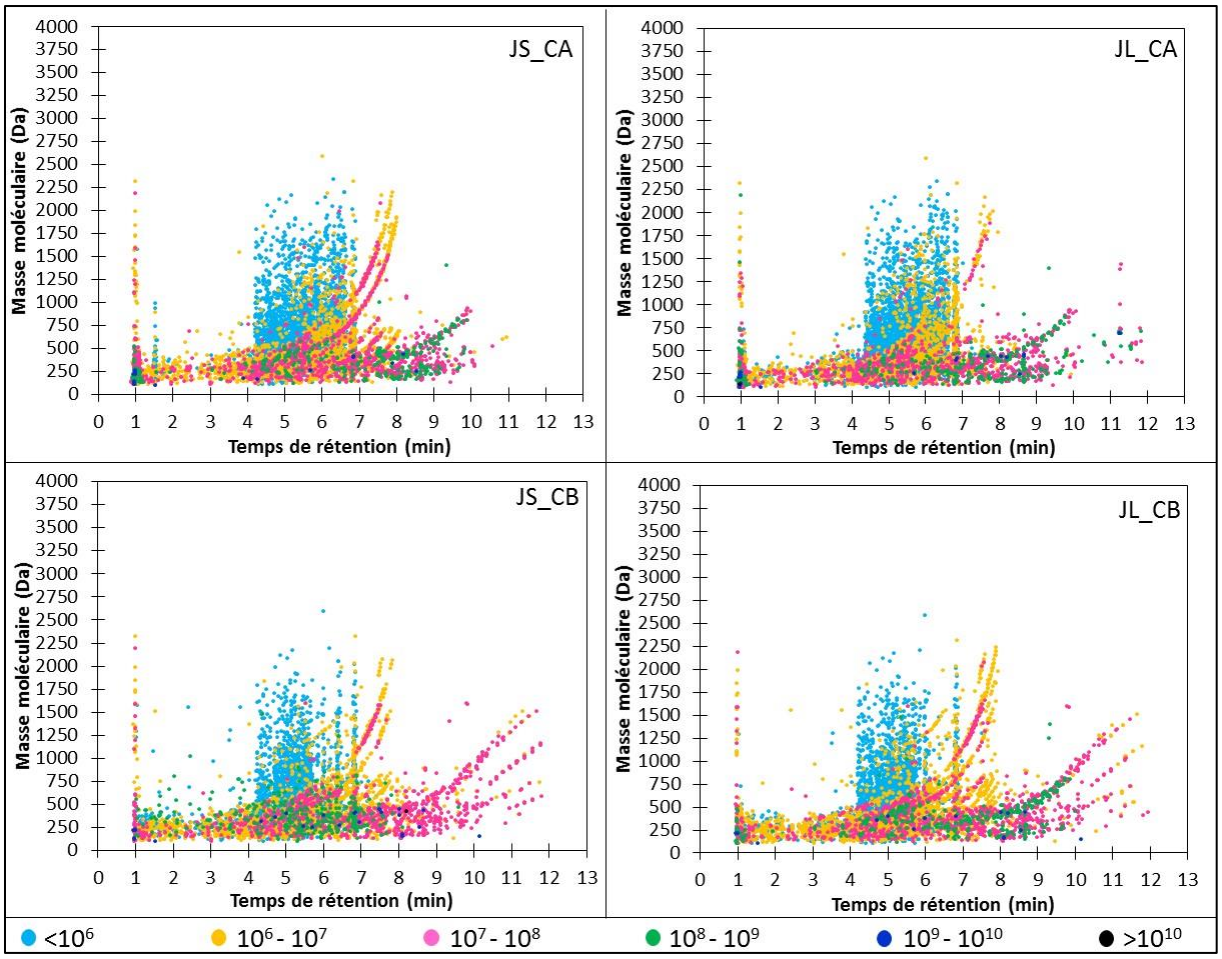


Figure 42 : Empreintes des échantillons de sorties de la STEP de St Julien l'Ars pour les campagnes A et B. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. JS : sortie St Julien l'Ars, JL : Lagune de St Julien l'Ars, CA : campagne A, CB : campagne B. Mode positif (HESI +).

Pour les deux STEP, l'essentiel des signaux est compris entre 4 et 8 minutes. Toutefois, en plus de cet intervalle, la gamme de signaux détectés s'étend d'une minute jusqu'à 12 minutes. L'échelle de polarité des composés présents dans ces échantillons comprend donc des molécules polaires et apolaires.

Les masses moléculaires sont pour la plupart inférieures à 750 Da. Des signaux de masses plus élevées sont cependant visibles, principalement entre 4 et 8 minutes et après 9 minutes. Ces échantillons semblent donc composés en majorité par des molécules de petites et moyennes tailles. Cette gamme de masses laisse de nombreux types de molécules possibles : micropolluants organiques, acides aminées, peptides ou encore polymères pour les plus hautes masses.

Les signaux présentant les aires les moins importantes ($<10^6$) sont concentrés entre les temps de rétention 4 et 7 minutes et avant 2 minutes. Particulièrement sur l'empreinte de St Julien l'Ars, ces

signaux forment comme « un voile » et la vérification manuelle d'un panel aléatoire de ces signaux a montré une surreprésentation de faux positifs. Ce phénomène, moins marqué dans le cas de Poitiers, illustre la problématique des faux positifs restant dans les données malgré l'application de blancs et de filtres supplémentaires. Concernant les autres intervalles d'aires, pour les deux STEP, les aires les plus intenses (> 10^8) se trouvent essentiellement après 5 minutes ou en tout début d'analyse. Les autres gammes d'aires sont reparties tout au long de l'empreinte. Les signaux les plus intenses (> 10^8) constituent une minorité par rapport aux autres signaux : ceux dont l'aire est supérieure à 10^9 sont même quasiment absents des empreintes. D'une manière générale, les valeurs d'aires des signaux ne semblent pas spécifiquement reliées à certaines gammes de masses moléculaires ou de temps de rétention (excepté pour les aires inférieures à 10^6).

III.2. Comparaison entre les eaux d'entrée et les eaux de sortie de STEP

III.2.1. Comparaison du nombre total de signaux

Pour comparer les eaux d'entrée et de sortie de STEP, le nombre total de signaux a tout d'abord été calculé pour l'ensemble des échantillons. Les pourcentages de variation de ces valeurs ont également été calculés pour chacune des sorties par rapport à leurs entrées respectives. Les résultats obtenus sont présentés dans la Figure 43.

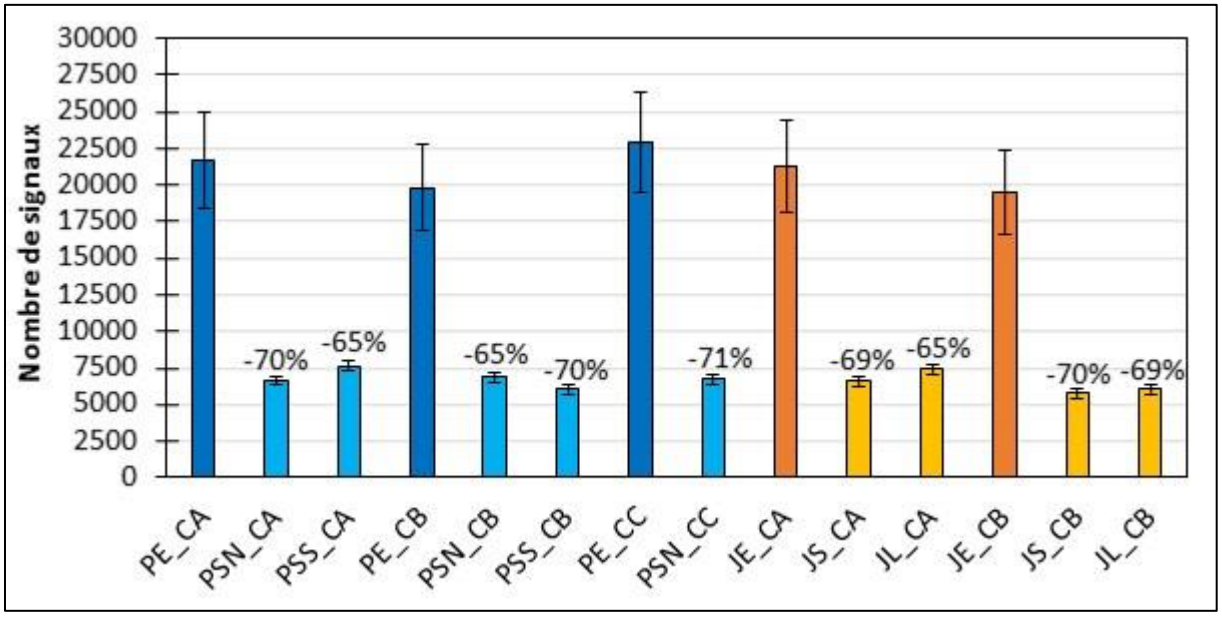


Figure 43 : Graphique en barres représentant le nombre de signaux en fonction de l'échantillon considéré. Les échantillons de la STEP de Poitiers sont représentés en bleu foncé pour les entrées (PE) et en turquoise pour les sorties Nord (PSN) et les sorties sud (PSS). Les échantillons de la STEP de St Julien l'Ars sont représentés en orange pour les entrées (JE) et en jaune pour la sortie (JS) et la lagune (JL). CA : campagne A ; CB : campagne B, CC : campagne C.

Ces résultats montrent la diminution du nombre de signaux détectés pour tous les échantillons de sortie quelle que soit la campagne (entre -65 et -71%, pour la STEP de Poitiers comme celle de St Julien l'Ars). Trois informations ressortent des valeurs d'abattement de cette figure :

- les deux sorties de Poitiers (Nord et Sud) ont des valeurs d'abattement équivalentes,
- les deux STEP ont des performances similaires si on considère ce paramètre,
- les fluctuations de diminution du nombre de signaux sont faibles entre les campagnes.

Le procédé d'épuration des eaux réduit ainsi fortement le nombre de signaux détectés dans les échantillons et est reproductible sur deux filières dans une même STEP et entre deux STEP différentes.

III.2.2. Comparaison des entrées et des sorties par leurs statistiques descriptives

Pour tous les échantillons, les statistiques descriptives (moyennes, médianes, quartiles, maxima, minima) ont été calculées pour les trois paramètres (temps de rétention, masses moléculaires et aires des pics). Ces statistiques ont été reportées dans la Figure 44 qui montre les boîtes à moustaches des échantillons d'entrées et de sorties pour les STEP de Poitiers et de St Julien l'Ars pour les trois campagnes de prélèvements.

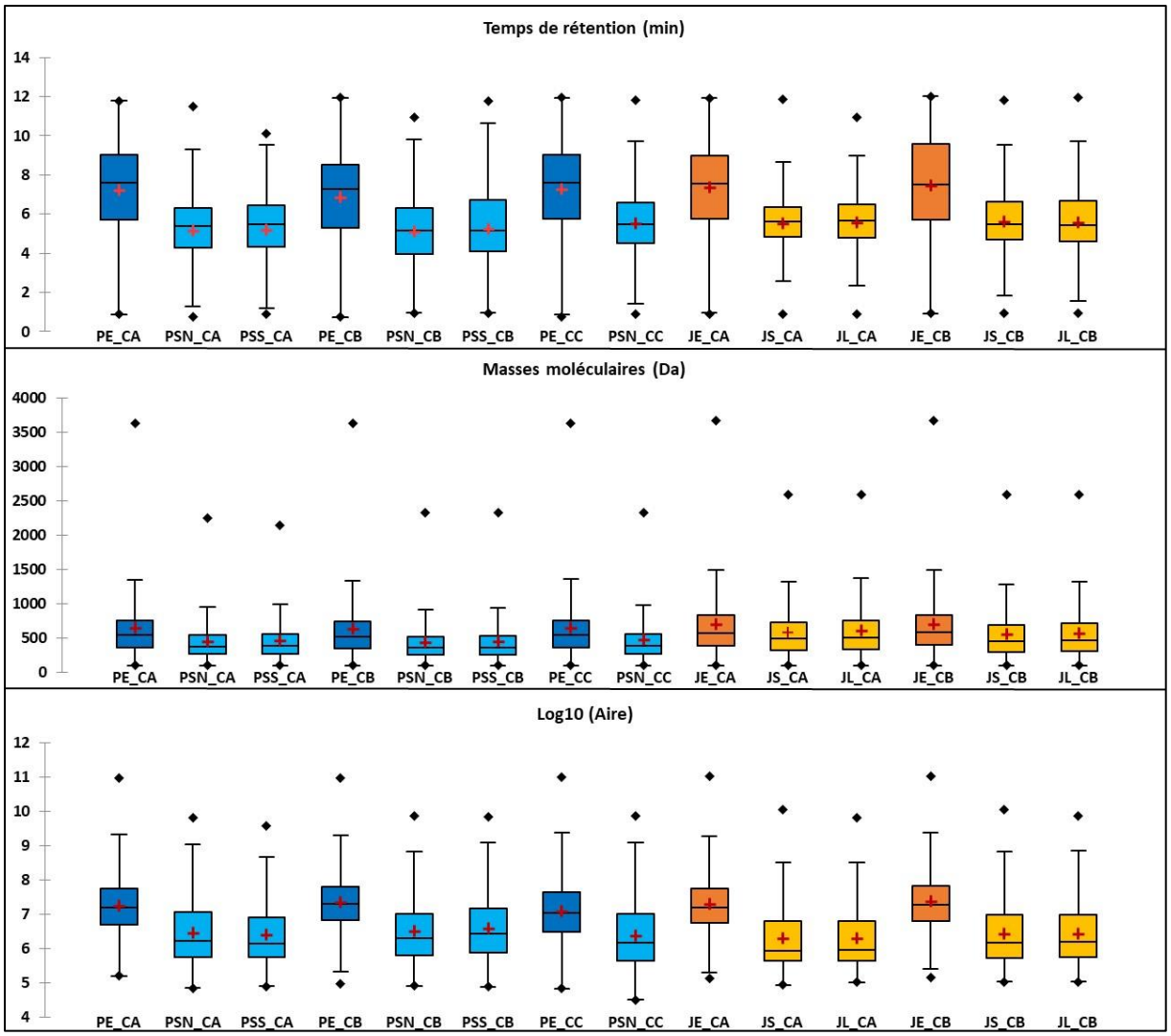


Figure 44 : Box plots ou boîtes à moustaches résumant les échantillons par leurs statistiques descriptives. La croix rouge, les limites inférieures et supérieures de la boîte et les points noirs symbolisent respectivement la moyenne, les premiers et troisièmes quartiles, les valeurs maximales et minimales. Les échantillons de Poitiers sont présentés en bleu (foncé pour les entrées et turquoise pour les sorties) et ceux de St Julien l'Ars en orange (foncé pour les entrées et plus clair pour les sorties).

Cette figure montre la différence entre les échantillons d'entrées et de sorties pour les trois paramètres et pour les deux sites de prélèvements.

Pour les temps de rétention, on observe une diminution significative de la médiane (7,5 ± 0,2 min pour les empreintes des échantillons d'entrées et 5,4 ± 0,2 min pour les sorties) pour les eaux des deux STEP. La polarité globale de la composition des échantillons augmente donc lors du traitement d'épuration. La dispersion observée est très similaire pour les temps de rétention entre les deux sorties de Poitiers de même que celle entre les eaux de sortie de la STEP de Poitiers et les eaux de la STEP de St Julien l'Ars.

Dans le cas des masses moléculaires, les valeurs maximales détectées semblent fortement impactées par le traitement avec des masses maximales supérieures à 3 500 Da en entrées et inférieures à 2 500 Da en sorties, quelle que soit la sortie considérée. Concernant ce point, il peut être noté que plus de 80% et plus de 70% des signaux ont une masse inférieure à 600 Da dans les sorties de Poitiers et dans les eaux de sorties de St Julien l'Ars, respectivement. En entrées, ces pourcentages sont inférieurs à 75% pour Poitiers et inférieurs à 66% pour St Julien l'Ars. Au niveau de St Julien l'Ars, la plus faible diminution de la présence de grandes molécules est certainement sous-évaluée en raison des faux positifs de faible intensité remarqués dans les empreintes (Figure 42). Ceci peut également influer sur la dispersion des signaux. Enfin, si on considère les masses, les résultats sont voisins entre les campagnes et entre les deux types de sorties. Le traitement des eaux réduit donc la proportion de molécules de grandes tailles dans les échantillons des deux STEP.

Pour ce qui est de l'intensité, la Figure 44 indique une forte diminution de l'intensité globale des signaux dans toutes les sorties étudiées. Cela se traduit par des aires ayant des valeurs en moyenne 36-37 % et 58 % en dessous des valeurs initiales de 10^6 et entre 10^6-10^8 respectivement pour les sorties de Poitiers. Pour les signaux de ces intensités, les deux sorties sont donc équivalentes avec des pourcentages moyens de 36% ($<10^6$), 58% (10^6-10^8). Les signaux les plus intenses ($>10^8$) ne représentent qu'entre 5 et 7% du total de signaux dans les sorties. Pour ces mêmes intervalles d'intensités, les entrées ont des pourcentages moyens de 3% ($<10^6$), 77% (10^6-10^8) et 18% ($>10^8$). Une diminution, entre les entrées et les sorties, de la proportion de signaux pour les aires les plus grandes et une augmentation de celle pour les aires les plus faibles sont donc notées au niveau de la STEP de Poitiers. Dans le cas de St Julien l'Ars ces pourcentages moyens sont de 48% ($<10^6$), 45% (10^6-10^8) et 5% ($>10^8$) en sorties et 2% ($<10^6$), 80% (10^6-10^8) et 18,2% ($>10^8$) en entrées. Si la distribution des signaux (les pourcentages) diffère entre les deux STEP, les tendances sont donc identiques entre Poitiers et St Julien l'Ars. Ces résultats sont à modérer par la part de faux positif dans les signaux d'aires inférieures à 10^6, plus particulièrement dans le cas de St Julien l'Ars.

Au vu de ces résultats portant sur les statistiques descriptives, les impacts des procédés d'épuration se traduisent par des variations des empreintes des eaux et sont visibles sur les trois paramètres décrivant les signaux. Ainsi, les eaux de sorties sont composées de molécules plus polaires, plus petites et moins intenses que les eaux d'entrée de STEP. Les deux sorties de Poitiers et les deux STEP montrent entre elles des résultats similaires. On constate également une homogénéité entre les campagnes.

III.2.3. Comparaison en fonction des tendances de variation des signaux et analyses différentielles

Le nombre total de signaux et les statistiques descriptives ne renseignent pas complètement sur les tendances que suivent les signaux lors de l'épuration des eaux. Le nombre de signaux en fonction de ces tendances d'évolution a donc été étudié. Pour cela, cinq tendances ont été définies en fonction de la valeur des aires en sortie par rapport à celle en entrée : élimination, formation, abattement supérieur à 50%, augmentation supérieure à 50% et variation comprise entre -50 et 50%. La Figure 45 montre le nombre de signaux pour les différents échantillons de sorties en fonction de ces tendances par rapport à leur entrée respective.

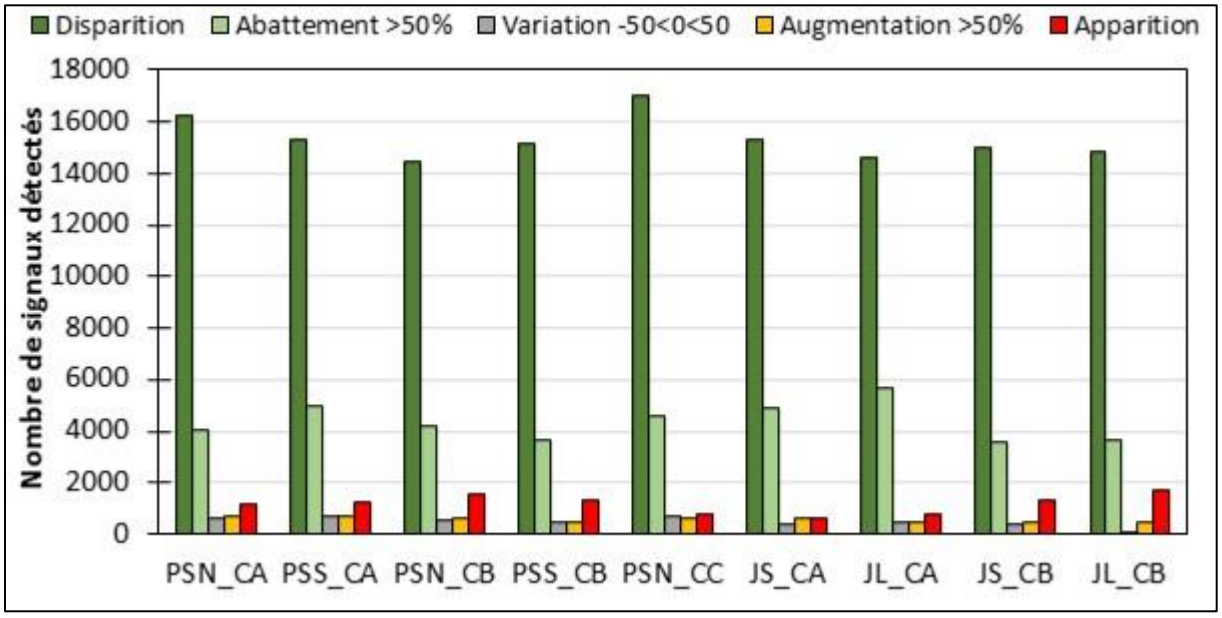


Figure 45 : Graphique en barres représentant le nombre de signaux en fonction de leur tendance de variation par rapport à l'entrée correspondante. PSN : Poitiers Sortie Nord, PSS : Poitiers Sortie Sud, JS : St Julien l'Ars Sortie, JL : St Julien l'Ars lagune, CA : campagne A ; CB : campagne B, CC : campagne C.

Cette figure montre que la disparition de signaux en sortie représente la tendance majoritaire parmi ces cinq tendances. Elle est suivie par l'abattement supérieur à 50% de l'aire des signaux. Ces deux tendances concernent à elles-seules entre 94% et 96% du total des signaux d'entrée. Cet impact est donc très largement majoritaire. A côté de ces signaux, les signaux dont la tendance est stable (variation entre -50% et +50%) constituent environ 3% du total des signaux d'entrée et entre 6 et 10% du total en sortie pour l'ensemble des échantillons. L'hypothèse principale qui peut expliquer cette tendance est le caractère réfractaire au traitement des composés représentés par ces signaux. Enfin, une augmentation (>50%) et une apparition de signaux ont également été observées dans les sorties. Ces signaux constituent entre 17% et 33% du nombre total de signaux en sortie. Ces valeurs sont donc

significatives. Ces résultats illustrent que l'épuration s'accompagne de processus de transformation conduisant à la formation de nouveaux composés qui se traduit par l'augmentation ou l'apparition de signaux. Ces observations sont en accord avec les travaux de Verkh *et al.* (2018) qui ont constaté la disparition de 65% des signaux étudiés avec l'apparition de 24% de nouveaux signaux dans les eaux traitées après le traitement biologiques des eaux (traitement secondaire).

Afin d'explorer la nature des signaux affectés par le traitement d'épuration, les empreintes ont également été tracées en prenant en compte les cinq tendances décrites ci-dessus. Les empreintes d'entrées ont été prises comme références auxquelles ont été ajoutés les signaux apparaissant dans les sorties. Les Figure 46 et 47 montrent les empreintes des échantillons de Poitiers et de St Julien l'Ars obtenues de cette manière.

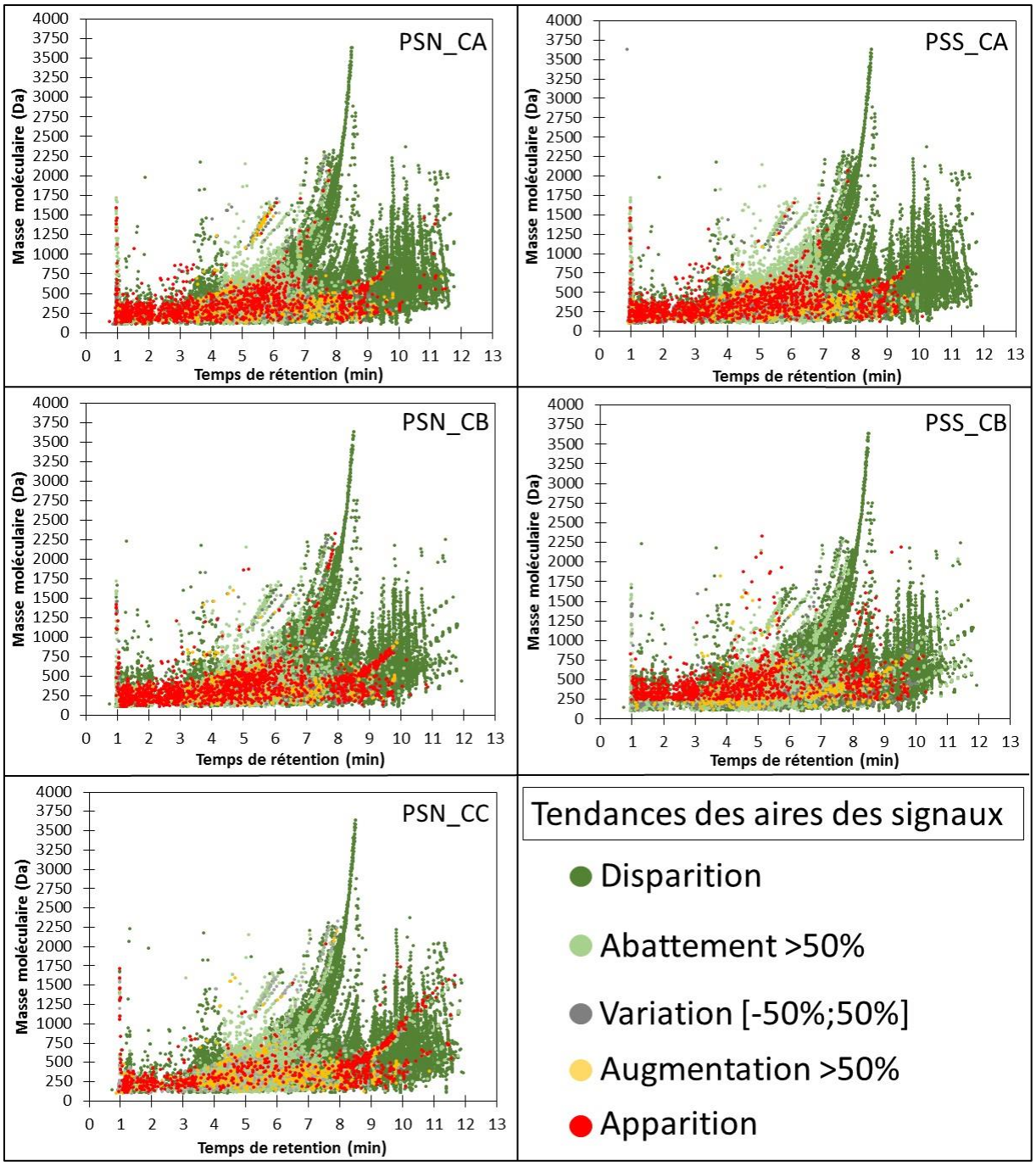


Figure 46 : Empreintes montrant les signaux détectés en fonction de leur tendance associée. PSN : Poitiers Sortie Nord, PSS : Poitiers Sortie Sud, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI +).

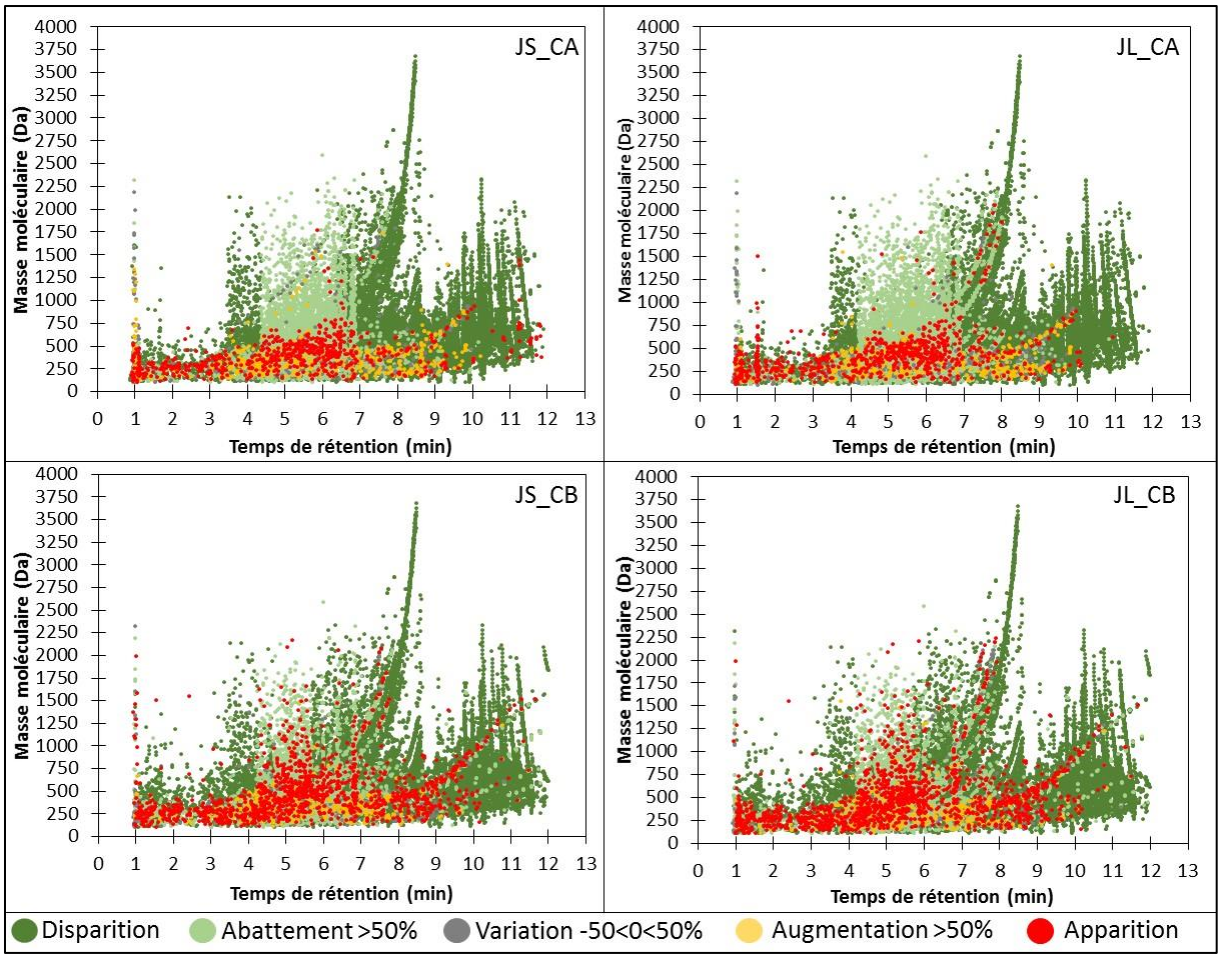


Figure 47 : Empreintes montrant les signaux détectés en fonction de leur tendance associée. JS : sortie St Julien l'Ars, JL : Lagune St Julien l'Ars, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI +).

Les Figures 46 et 47 indiquent que les signaux disparaissant en sortie sont en majorité des signaux dont le temps de rétention est supérieur à 6 minutes et/ou dont la masse est supérieure à 600 Da. Ces observations visuelles rejoignent les commentaires effectués dans les paragraphes III.1 et III.2.2 mentionnant que les composés plutôt apolaires et de hautes masses sont majoritairement abattus ou fortement réduits, indépendamment de la STEP ou de la campagne d'échantillonnage.

Les signaux apparaissant ou augmentant sont dispersés sur l'ensemble de l'empreinte et sont préférentiellement des signaux de masse inférieure à 600 Da polaires à moyennement polaires. Parmi les signaux, des masses moléculaires plus élevées sont toutefois également visibles entre 6-8 min et après 9 min. Un enchainement de signaux alignés (formant des motifs) entre 8 et 12 min augmentant en masse sur la plupart des empreintes présentées peut également être observé. Sur ce type de signaux, des modifications des empreintes sont observables entre les deux sorties de Poitiers et entre les campagnes. Si les gammes de masses et de polarité sont équivalentes, les signaux ne sont en effet pas toujours les mêmes. Deux hypothèses peuvent être avancées pour expliquer ce phénomène : une

variation des concentrations de ces signaux sous la limite de détection ou la modification de la composition en composés parents de ces nouveaux signaux avant le procédé de traitement (Evgenidou *et al.*, 2015 ; Gros *et al.*, 2013 ; Guerra *et al.*, 2014). Dans le cas de St Julien l'Ars, la sortie et la lagune sont très similaires. Ce résultat est cohérent puisque la classification des signaux est effectuée par rapport à l'entrée dans les deux cas et la lagune est une zone de stockage des eaux de sortie. Effectuer le classement des signaux entre la sortie et la lagune mettrait en relief les modifications induites par le stockage sur les eaux de sortie mais cette comparaison ne rentre pas pleinement dans le champ de la présente étude. Comme pour Poitiers, les deux campagnes de St Julien l'Ars montrent des changements dans les signaux apparaissant ou augmentant. La variabilité de ce type de signaux semble donc être indépendante de la STEP.

Pour poursuivre ces analyses différentielles, le calcul du rapport de l'aire des signaux des sorties sur l'aire des signaux des entrées puis la représentation du logarithme en base 2 de ce ratio (Fold Change) en fonction de $-\log_{10}$ de la p-value calculée pour chaque signal (graphique de type volcano-plot) ont également été réalisés. Un logarithme binaire du Fold Change (FC) supérieur à 1 signifie au moins un doublement de la valeur de l'aire tandis qu'un logarithme binaire du Fold Change inférieur à -1 signifie au moins une diminution de la moitié de la valeur de l'aire. Entre ces deux bornes sont présents des signaux variant peu entre les entrées et les sorties. La valeur de la p-value choisie comme seuil de significativité a été de 0,05 soit 5%. Les valeurs en-dessous de ce seuil ont donc été considérées significatives.

Des analyses différentielles ont été effectuées :

- entre les données des analyses de l'entrée et des deux sorties (Nord et Sud) pour toutes les campagnes d'échantillonnage de la STEP de Poitiers. Pour la campagne C, l'analyse différentielle a été effectuée sur la moyenne des aires des entrées et la moyenne des aires des sorties.
- entre l'entrée et la sortie de la STEP de St Julien l'Ars (campagnes A et B).

Les valeurs à 0 correspondant à l'absence du signal dans l'empreinte de l'échantillon ont été remplacées par la valeur du blanc pour garantir la visibilité de tous les points sur les volcano-plots. Les volcano-plot de chacune des analyses différentielles sont représentés dans la Figure 48 pour Poitiers et la Figure 49 pour St Julien l'Ars.

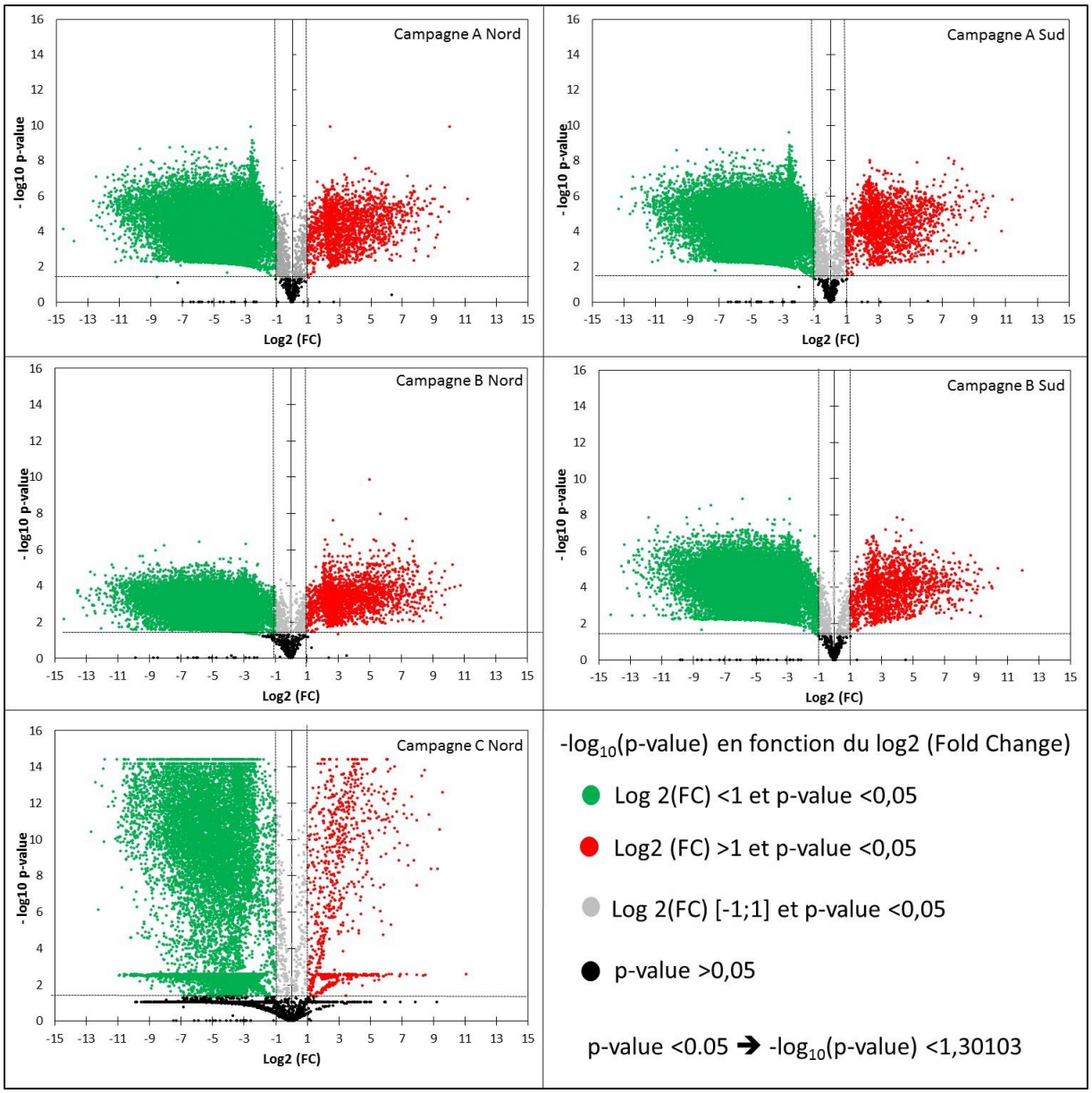


Figure 48 : Volcano-plot des analyses différentielles entre les données des analyses d'entrée et celles des deux sorties par campagne (A, B et C) pour Poitiers. Analyses différentielles réalisées par XLSTAT.

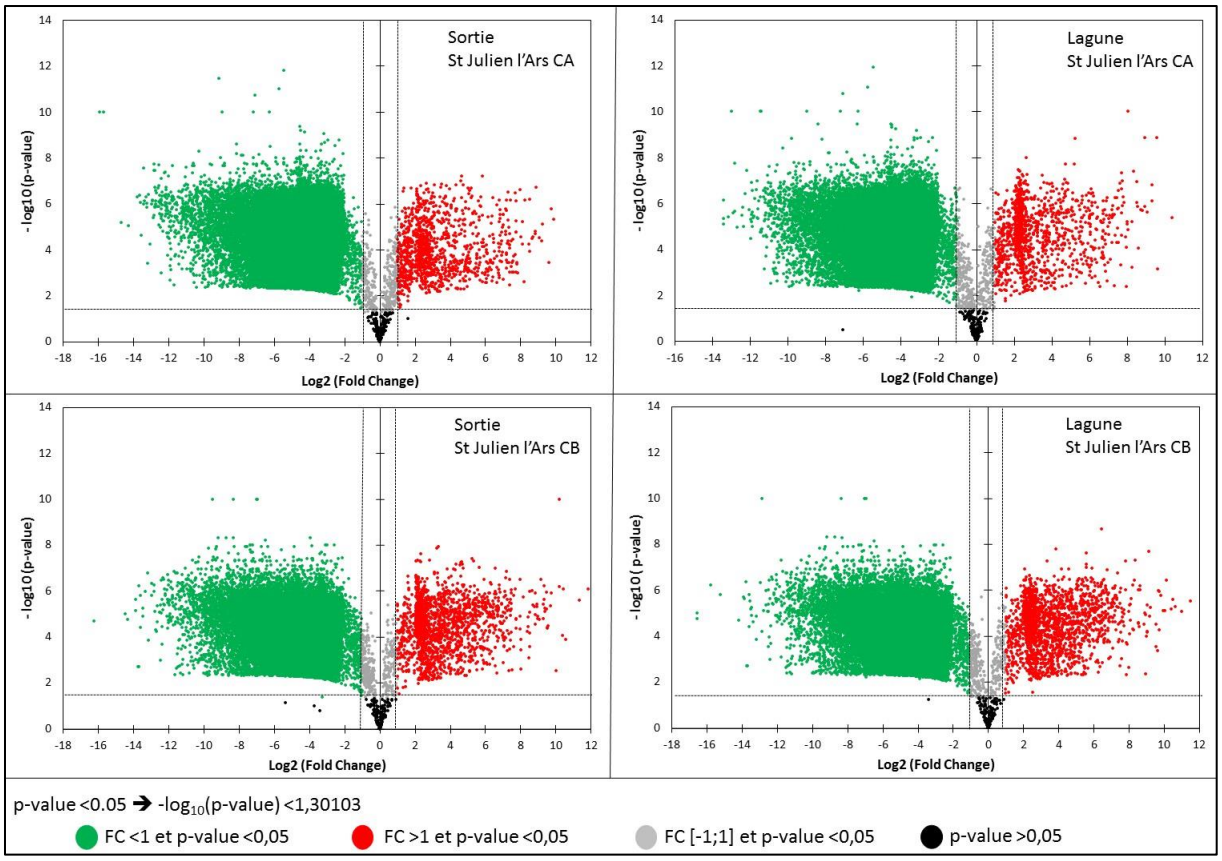


Figure 49 : Volcano-plot des analyses différentielles entre les données des analyses d'entrée et celles des deux sorties et lagune par campagne (A, B et C) pour St Julien l'Ars. Analyses différentielles réalisées par XLSTAT.

Les Figures 48 et 49 montrent que la grande majorité des signaux ont un logarithme binaire du Fold Change inférieur à -1 et donc diminuent ou sont éliminés dans les échantillons de sorties (en vert sur la figure). On constate également l'augmentation (ou l'apparition) de signaux (Log2 Fold Change >1, en rouge sur la figure) dans les sorties ainsi que la présence d'autres signaux dont la valeur d'aire varie peu entre l'entrée et les sorties (en gris sur la figure). Enfin, très peu de signaux ont une p-value non significative (p-value > 0,05).

IV. Corrélation entre signaux et nature des échantillons par analyse multivariée

Les résultats précédents ont montré l'impact du traitement des eaux sur les signaux détectés dans les échantillons. Les empreintes et les boîtes à moustaches sont des outils de visualisation des données mettant en avant les différences les plus évidentes entre les échantillons. Elles ne permettent pas néanmoins de faire le lien entre les signaux qui discriminent significativement les eaux avant et après traitements. Pour réaliser cet objectif, les analyses statistiques univariées ou multivariées sont généralement employées et constituent les principaux outils de priorisation des données. Dans cette

partie, on a étudié la capacité d'analyses multivariées à distinguer les deux catégories d'échantillons (entrées et sorties).

Les analyses multivariées peuvent être descriptives ou explicatives. Les méthodes descriptives permettent d'explorer les données sans hypothèse de départ pour diminuer le nombre de données à prendre en considération, rechercher des corrélations entre des variables et des échantillons, mettre en évidence des prélèvements singuliers ou encore mettre en évidence des groupes d'échantillons. Les méthodes explicatives cherchent quant à elles à tester et mettre en lumière les liens entre le tableau de données et une ou plusieurs variables à expliquer : toxicité ou niveau de pollution par exemple. Dans le cadre de ce travail, l'analyse en composantes principales ACP a été utilisée. Cette méthode permet de réduire le nombre de variables décrivant les données par projection sur de nouvelles variables appelées facteurs. Elle permet de réduire la dimensionnalité des données initiales, d'exclure des données aberrantes et de mettre en relief des corrélations entre les variables étudiées et les échantillons analysés.

Les données des échantillons de Poitiers et des échantillons de St Julien l'Ars ont été soumis séparément à une ACP par XLSTAT. Les données ont été préalablement normalisées (n-1) afin que les signaux les plus intenses n'aient pas un poids prépondérant dans les résultats par rapport aux signaux moins intenses.

On s'est intéressé tout d'abord aux graphiques représentant les échantillons selon les axes des facteurs 1 (F1) et 2 (F2). Ces graphiques des individus sont utilisés pour relier les individus (échantillons) entre eux et donc déterminer des groupements, dont leur position relative entre eux et par rapport aux axes est interprétable d'un point de vue physique ou chimique. Les Figures 50A et 50B correspondent aux graphiques des individus pour Poitiers et St Julien l'Ars respectivement.

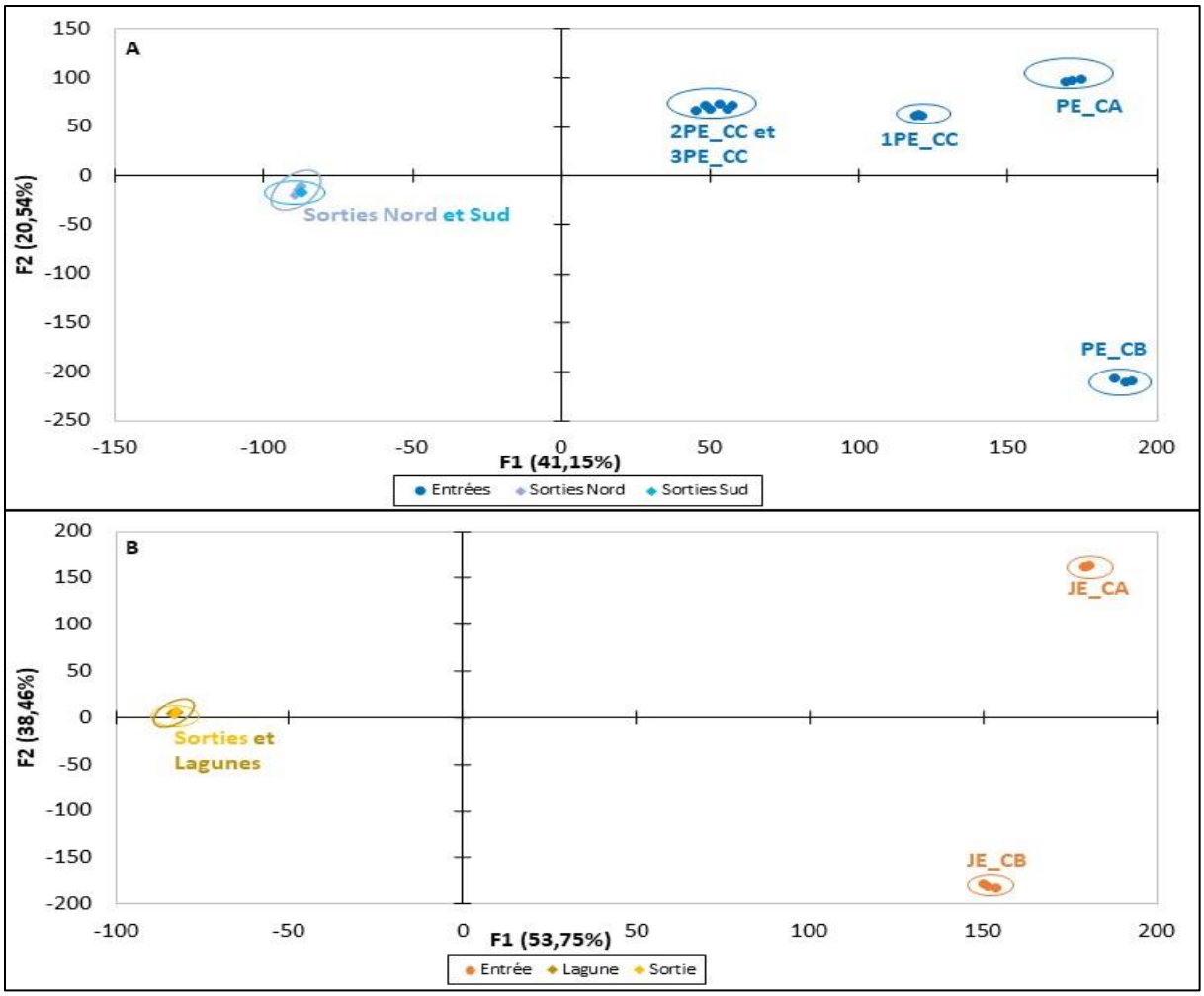


Figure 50 : Graphiques des échantillons de Poitiers en A et des échantillons de St-Julien en B. PE : Entrée Poitiers, JE : Entrée St Julien l'Ars, CA : campagne A, CB : campagne B, CC : campagne C.

D'après ces figures, les deux facteurs présentés comprennent 62% et 92% de l'information totale pour Poitiers et St Julien l'Ars respectivement. L'essentiel de l'information est donc visible par ces deux axes pour St Julien l'Ars, ce qui est peu surprenant en raison du nombre d'échantillons analysés inférieur à celui de Poitiers. En effet, l'augmentation du nombre d'échantillons conduit à disperser l'information sur davantage d'axes. Pour les deux sites de prélèvements, les résultats obtenus montrent une première séparation des entrées et des sorties le long de l'axe F1. Une seconde séparation par l'axe F2 détache la campagne B des autres campagnes. Ainsi, les sorties sont regroupées ensemble dans les deux figures et la campagne B diverge donc des autres campagnes pour les deux STEP. Cette séparation pourrait s'expliquer par des changements dans la composition des eaux d'entrée de STEP en fonction des dates de prélèvements et sont en accord avec les observations faites en entrée de STEP pour la campagne B (paragraphe II.2 précédent). Ces écarts seraient atténués en sorties par le procédé d'épuration des eaux. Dans le cas de Poitiers, on observe également que l'échantillon 1 de la campagne C se distingue des deux autres échantillons de cette campagne. Cet écart pourrait en partie provenir de variations apportées lors de la préparation ou de l'analyse par UHPLC-HRMS de l'échantillon.

Ainsi, dans ces figures, l'axe F1 représenterait le « niveau de pollution » des eaux pour les deux STEP tandis que l'axe F2 symboliserait les variations temporelles de composition des eaux d'entrées. Ces figures sont en accord avec les résultats de Schollée *et al.* (2015) qui ont observé la même séparation entre entrées et sorties et entre des entrées dont les prélèvements ont été effectuées à des dates différentes. La superposition des sorties peut s'expliquer par le plus faible nombre de signaux de ces échantillons dans le tableau de données initiales et qui apportent donc moins d'informations que les entrées. Une séparation des sorties n'est en effet visible, pour Poitiers, qu'à partir de l'axe F6 de cette ACP qui ne représente que 2,6% de la variance et qu'à partir de l'axe F3 pour St Julien l'Ars qui représente 3,7% de la variance.

Dans le but de déterminer la corrélation d'ensemble de signaux avec les échantillons, des biplots comprenant à la fois les échantillons et les signaux ont été construits. Pour cela, en raison de la séparation entre les entrées et les sorties sur le graphique des échantillons, l'hypothèse de la corrélation des signaux avec les échantillons en fonction de leur tendance de variation a été examinée. Par conséquent les cinq tendances déterminées précédemment (disparition, Abattement >50%, Variation comprise entre -50% et 50%, augmentation >50% et apparition) ont été représentées dans les biplots par un jeu de couleurs. Les Figures 51A et 51B suivantes montrent ces biplots pour les deux sites de prélèvements. Dans ces figures, les tendances de variation utilisées sont les tendances moyennes (c'est-à-dire établies à partir de la moyenne des aires dans les entrées et dans les sorties) sur l'ensemble des campagnes et non pas par campagne.

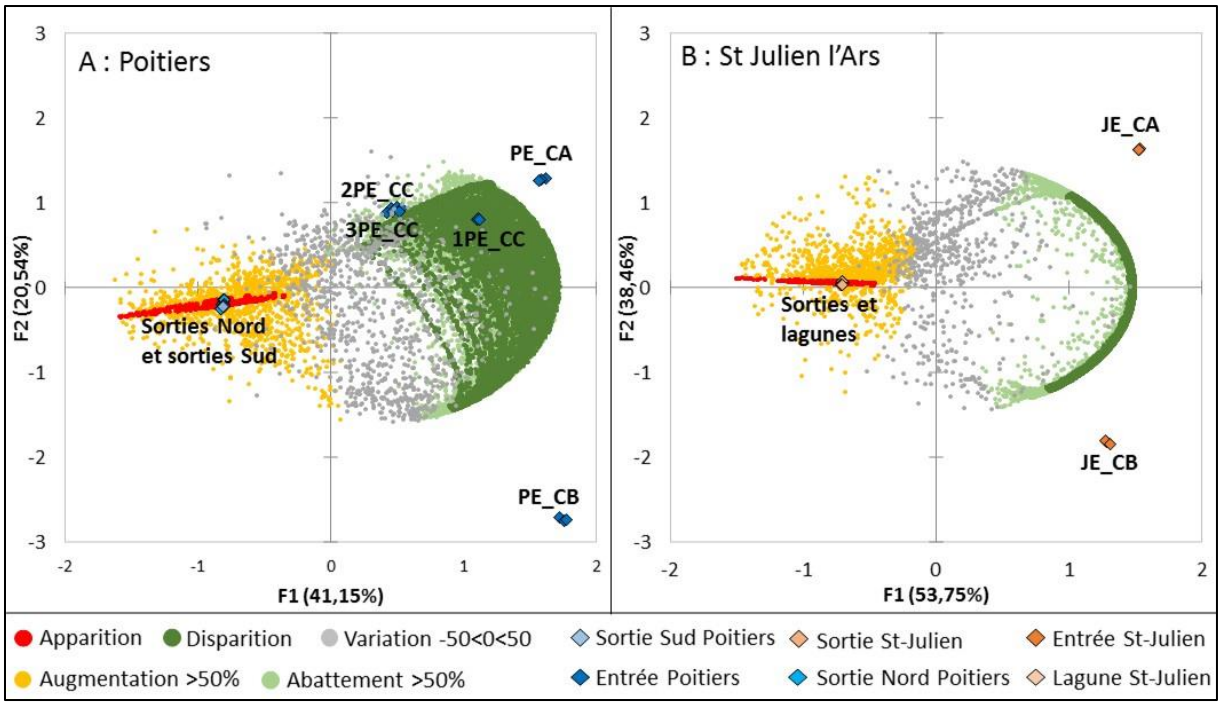


Figure 51 : Biplots des échantillons et des signaux détectés selon les facteurs F1 et F2 et les tendances de variation associées pour Poitiers en A et St Julien l'Ars en B. PE : entrée Poitiers, JE : entrée St Julien l'Ars, CA : campagne A, CB : campagne B, CC : campagne C.

Sur ces biplots, on observe pour Poitiers et St Julien l'Ars des groupements de signaux vers les entrées d'une part et vers les sorties d'autre part. Il y a donc une corrélation entre certains signaux et les catégories d'échantillons. Ces résultats sont en accord avec les résultats de Schollée *et al*. (2015) qui ont observé les mêmes regroupements de signaux. Concernant les variations, pour les deux STEP, la représentation des tendances de variation témoigne d'une relation entre la tendance et la catégorie d'échantillon. En effet, les signaux apparaissant (en rouge) ou augmentant (en jaune) sont corrélés avec les sorties tandis que les signaux disparaissant (en vert foncé) ou diminuant (en vert clair) sont corrélés avec les entrées. Les signaux plutôt stables se retrouvent quant à eux entre les deux catégories et ne sont donc pas liés fortement avec l'une ou l'autre des catégories. Néanmoins, le nombre d'échantillons étudiés dans le cadre de ce travail était restreint ce qui pourrait d'ailleurs expliquer l'allure particulière du biplot de St Julien l'Ars. La reproductibilité de ce lien entre tendances moyennes et échantillons devra donc être vérifiée par la prise en compte de davantage d'échantillons. De plus, une limite de cette classification des signaux est l'effet matrice qui peut impacter la réponse (aire) observée lors des analyses UHPLC-HRMS et mener ainsi à une corrélation inadéquate du signal avec une catégorie d'échantillons. L'application du facteur de correction estimé dans le chapitre 3 « Mise au point de l'approche » ne permet pas totalement d'écarter cette possibilité.

V. Recherche de signaux pertinents dans les eaux de STEP

La recherche de signaux pertinents peut être orientée en fonction de la problématique considérée : toxicité, rémanence ou apparition de nouveaux composés, recherche de produits de transformations, etc. Cette recherche peut également être effectuée à partir des données obtenues par l'analyse non ciblée ou encore par des expériences complémentaires. Dans le cas de l'utilisation des résultats obtenus par l'analyse non ciblée, plusieurs points de départ sont possibles : les résultats des analyses multivariées, les ratios ou Fold Change entre les entrées et les sorties, l'intensité des signaux ou encore la recherche d'atomes reconnaissables sur un spectre de masse. Plusieurs de ces méthodes peuvent être combinées entre elles pour réduire le nombre de signaux à prendre en considération.

La recherche de signaux pertinents dépend de l'objectif qui y est associé et les signaux ainsi sélectionnés peuvent différer entre deux objectifs différents et deux méthodes de sélection différentes. Dans le cadre de ce travail, l'objectif a été de sélectionner des signaux pertinents en lien avec l'impact des STEP sur le milieu récepteur des eaux traitées, par l'analyse différentielle et donc sans *a priori* sur les signaux intenses ou les massifs isotopiques spécifiques (de type Cl ou Br).

V.1. Sélection des signaux par analyses différentielles

La recherche de signaux pertinents peut s'appuyer sur l'utilisation du Fold Change. Dans cette démarche, les volcano-plots présentés précédemment (Figure 48) ont été utilisés.

Pour cela, lors de cette sélection, les signaux présents avec une p-value supérieure à 0,05 ont tout d'abord été écartés. La valeur de la p-value permet en effet de sélectionner selon leur significativité statistiques les ratios des aires des signaux entre l'entrée et les sorties. Ensuite, les signaux ayant un logarithme binaire du Fold Change supérieur à -1 ont été sélectionnés pour conserver les signaux qui augmentent et apparaissent en sortie mais également les signaux qui varient peu (Log2 FC compris entre -1 et 1 car ils peuvent correspondre à des molécules potentiellement réfractaires au traitement des eaux. Ce sont ainsi les signaux représentés en rouge et en gris (Figure 48) auxquels nous nous sommes intéressés. En effet, ces signaux correspondent aux molécules les plus susceptibles d'impacter le milieu récepteur.

Ces catégories de signaux (apparaissant, augmentant et stables) constituent entre 1 407 (Campagne C) et 2 442 signaux (Campagne B Nord). Parmi ces signaux, seuls ceux communs aux échantillons des trois campagnes ont été conservés. Ainsi, 375 signaux ont été sélectionnés. L'approche choisie est donc une approche restrictive mais focalisée sur des signaux correspondant potentiellement à des molécules impactant régulièrement le milieu récepteur des eaux traitées, la rivière Clain. Pour ces

signaux, les tendances moyennes sur les données de l'ensemble des campagnes d'échantillonnage ont été recalculées. La Figure 52 représente l'empreinte des signaux conservés après la sélection détaillée précédemment selon leur tendance moyenne obtenue pour l'ensemble des campagnes.

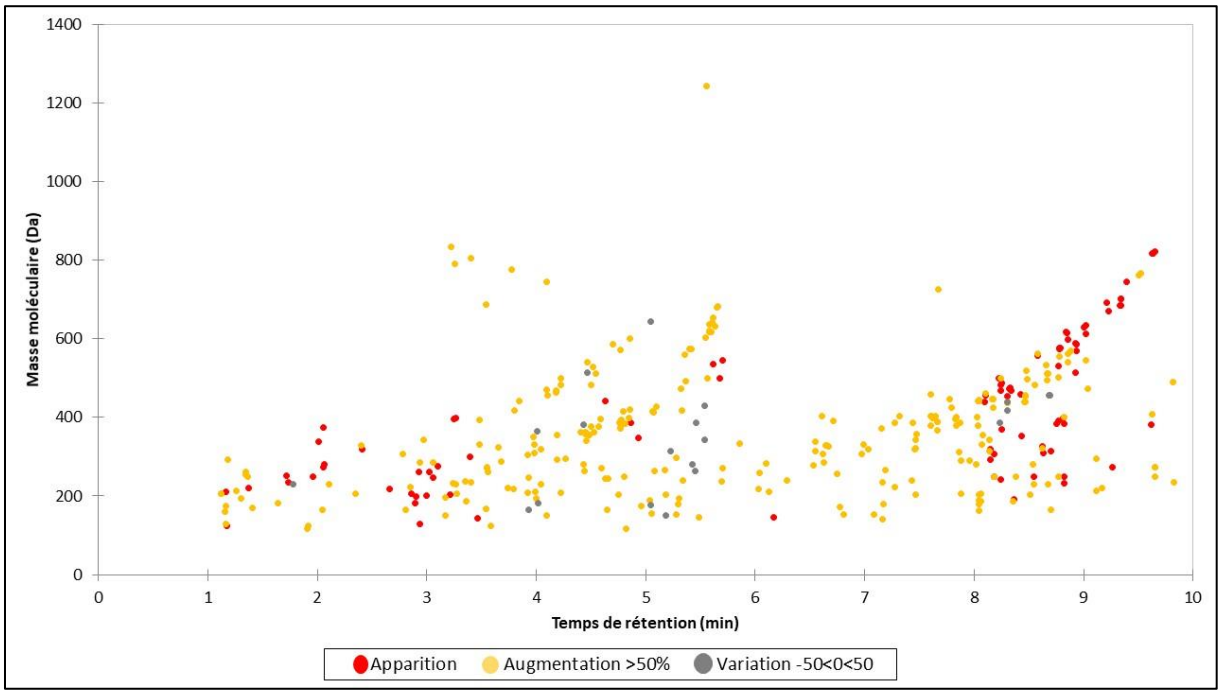


Figure 52 : Empreinte montrant les signaux détectés en fonction de leur tendance associée. Les grise, jaune et rouge représentent respectivement les signaux qui varient entre 0-50%, augmentent et apparaissent.

Cette figure rejoint les observations déjà effectuées dans les Figures 46 et 47 : les signaux dont la valeur d'aire augmente ou qui apparaissent en sortie correspondent à des molécules de polarités variées, généralement de petites et moyennes tailles (< 800 Da) pour la majorité d'entre elles. Cela peut s'expliquer par la relation de dégradation, de déconjugaison ou de transformation entre les molécules présentant des signaux diminuant (ou disparaissant) et les molécules correspondant aux nouveaux signaux (ou ceux dont la valeur d'aire augmente). Les réactions évoquées entrainent un changement de masse ou de polarité par perte ou réarrangement de groupements chimiques. Ces signaux communs entre les empreintes sont à la fois des signaux correspondant à des molécules non totalement éliminées en sortie et aussi à des signaux nouveaux ou dont l'intensité augmente. Ce sont donc des signaux correspondant à des molécules ayant un impact potentiel sur les milieux récepteurs. Toutefois sans une étude plus approfondie de l'identité des molécules correspondant à ces signaux, il est difficile de présager de leur identité : acides aminés, produits de dégradation de la matière organique, micropolluants réfractaires ou encore produits de transformation de micropolluants.

Les méthodes, ACP et analyse différentielle, sont des méthodes complémentaires qui peuvent être toutes deux utilisées pour la sélection de signaux pertinents. La position des signaux pertinents sélectionnés précédemment a donc ensuite été mise en évidence sur le biplot des échantillons-signaux

pour voir si une sélection par l'ACP aurait abouti à des résultats similaires. La Figure 53 présente ainsi le même biplot que la Figure 51A mais avec le code couleur suivant : signaux sélectionnés en rouge et non sélectionnés en gris.

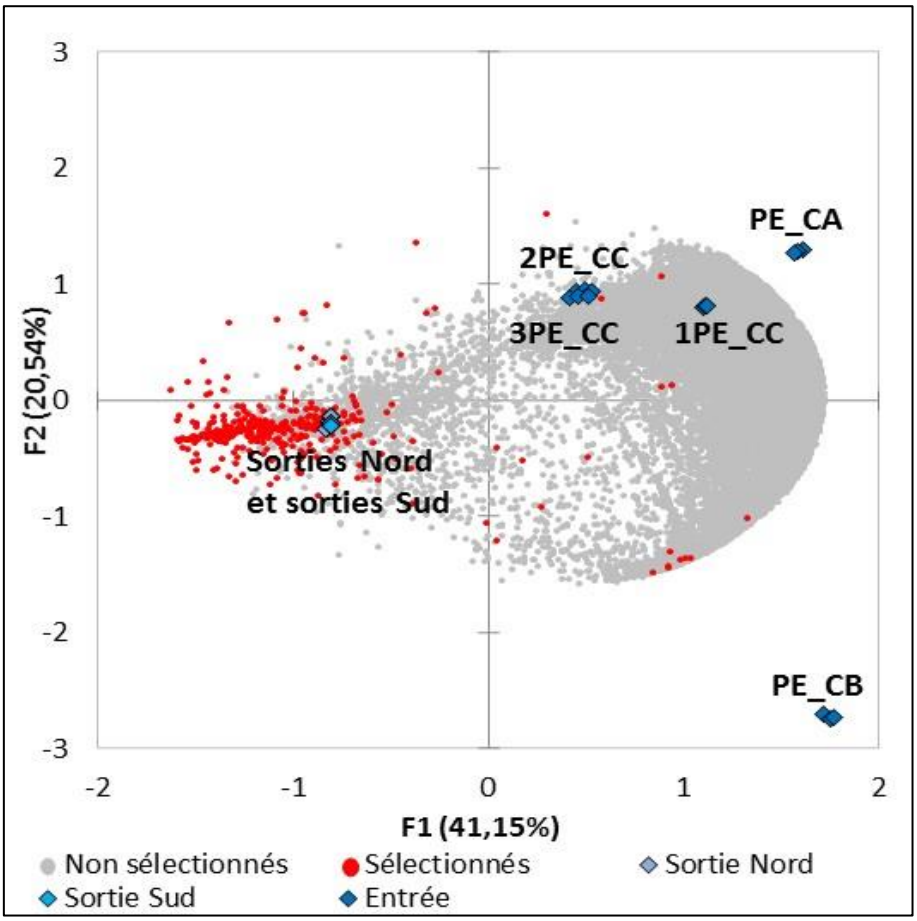


Figure 53 : Biplots des échantillons et des signaux détectés selon les facteurs F1 et F2 pour la STEP de Poitiers. PE : entrée Poitiers, CA : campagne A, CB : campagne B, CC : campagne C. Les points des sorties sont confondus.

Ces résultats montrent une bonne correspondance entre les signaux sélectionnés et les signaux corrélés avec les sorties sur l'ACP bien que quelques rares signaux soient plutôt corrélés avec les entrées. Une sélection par l'ACP aurait donc abouti à des résultats proches.

V.2 Approches de caractérisation

Afin de mieux caractériser les signaux pertinents sélectionnés, des attributions de formules brutes et des tentatives d'annotations ont été réalisées.

V.2.1. Attribution de formules brutes par Compound Discoverer

La prédiction des formules brutes pour les signaux sélectionnés a été réalisée par Compound Discoverer de telle sorte que les formules brutes calculées aient une déviation sur la masse maximale de ± 5 ppm ou de ± 3 mDa. Dans le cas ou plusieurs formules brutes peuvent être proposées, un

classement a été réalisé automatiquement par Compound Discoverer. Dans ces conditions, en cas de formules brutes ayant un « spectral fit » (similarité spectrale entre le spectre obtenu et le spectre théorique) et un « pattern coverage » (ratio de la somme des intensités des pics isotopiques du spectre MS1 sur la somme des intensités du spectre théorique) très proches (moins de 5% de différence pour le spectral fit et moins de 2% de différence pour le pattern coverage), la formule brute la plus proche en terme de déviation de masse a été conservée. Il est important de noter que plus la masse moléculaire considérée augmente plus le taux de mauvaise attribution est susceptible de s'accroître. Ainsi, les signaux de masse supérieure à 800 Da ont été exclus lors de cette approche. Une formule brute a ainsi été attribuée à 331 signaux parmi les 375 signaux sélectionnés. À partir de ces formules brutes, les ratios H/C, O/C, N/C ont alors été calculés car ces atomes constituent les atomes majoritaires. La Figure 54 présente les graphiques du ratio H/C en fonction des ratios O/C et N/C.

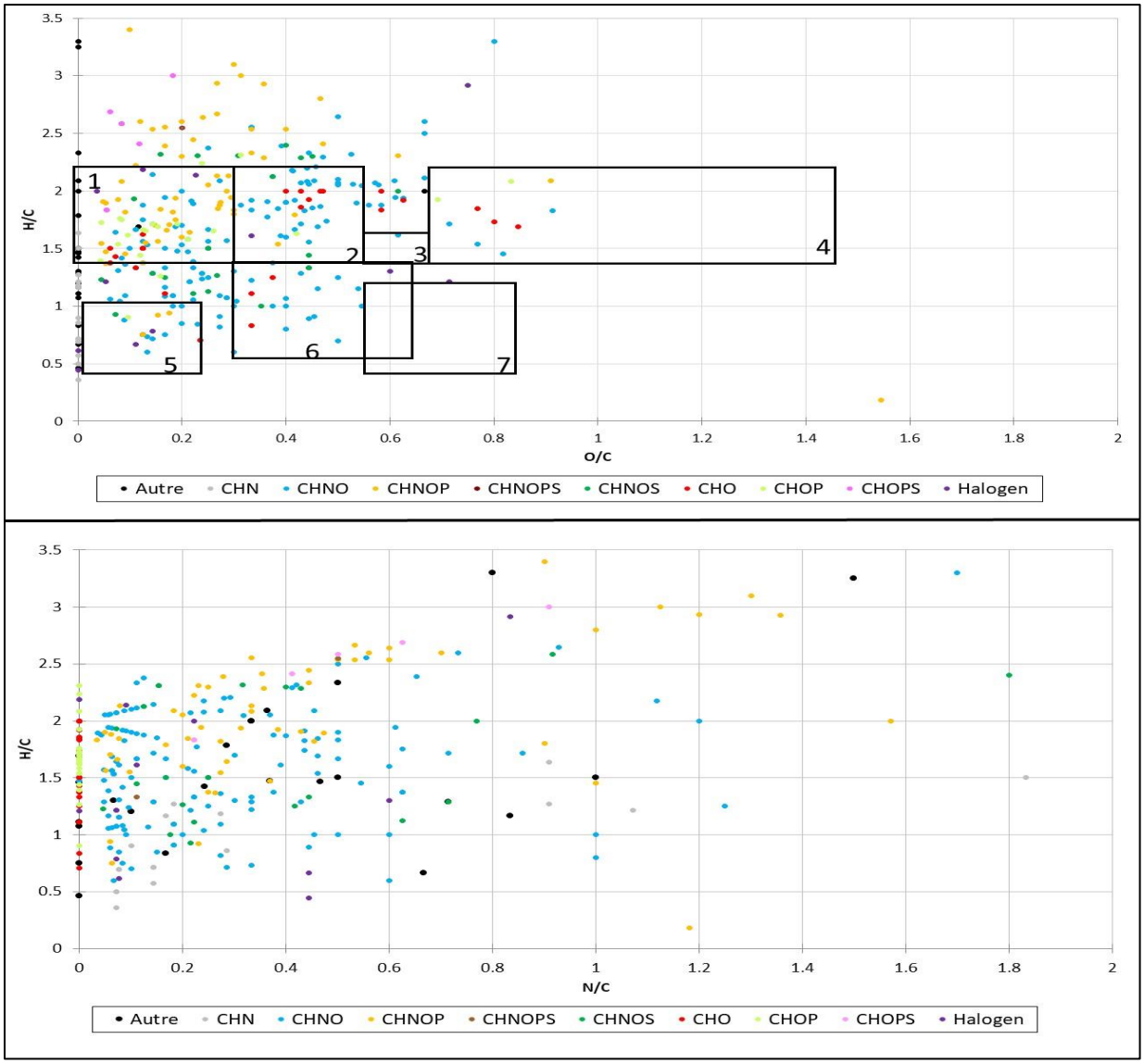


Figure 54 : Graphiques du ratio H/C en fonction des ratios O/C et N/C des signaux sélectionnés par analyse différentielle. 1 : lipides, 2 : protéines, 3 : sucres aminés, 4 : carbohydrates, 5 : structures aromatiques condensées, 6 : lignines, 7 : tannins.

Sur le graphique du ratio H/C en fonction du ratio O/C ont été précisées des zones regroupant, selon les travaux de Maizel et Remucal (2017), certaines familles de molécules. Ainsi, la plupart des signaux ont un rapport O/C inférieur à 0,6 et un rapport H/C inférieur à 2,5. Les classes « lipides » (1), « protéines » (2), « structures condensés » (5) et « lignines » (6) ainsi délimitées sont les classes qui englobent le plus de signaux dans la figure. Un nombre significatif des signaux sélectionnés pourraient donc correspondre à des molécules issues des matières organiques naturelles ou biologiques mais cela reste des indications qui seraient à valider par une démarche d'identification. L'apparition ou l'augmentation de ce type de molécules pourrait ainsi être due à la dégradation de la matière organique. La majorité des types de formules brutes est répartie sur le graphique à l'exception des formules de type CHOP dont la valeur du ratio H/C est plutôt comprise 1,2 et 1,8. Cette figure montre que les signaux pertinents regroupent à la fois des signaux plutôt saturés et/ou aliphatiques (H/C > 2) et des signaux pouvant correspondre à une structure insaturée (H/C < 2). Les signaux de type CHNO et CHNOP sont les plus représentés. Les signaux de type CHNO ont des ratio H/C ayant des plus faibles valeurs que ceux de types CHNOP. Les rares signaux de type CHOPS (en rose) sont globalement en dehors des catégories définies. Ce type de signaux pourrait correspondre à des molécules de nature anthropique. Ces résultats rejoignent les résultats de Verkh *et al.* (2018) qui ont observé des répartitions proches dans les signaux détectés. Les travaux de Maizel et Remucal (2017) et Phungsai *et al.* (2016) ont quant à eux montré un regroupement des signaux plus resserrés sur les zones « protéines » (2), « lignines » (6) et « tannins » (7). Il faut noter toutefois que leurs analyses ont été effectuées en mode négatif et que le traitement des données appliqué n'était pas aussi exclusif que la sélection faite ici. Par ailleurs, on peut également noter que le phosphore a été inclus dans nos travaux alors qu'il ne l'était pas dans l'étude de Verkh *et al.*

Le graphique du ratio H/C en fonction du ratio N/C montre une majorité de signaux retenus avec un ratio N/C inférieur à 0,8 (91% des signaux) et correspondant à des molécules insaturées ou peu saturées (H/C <2). Les valeurs de ratio N/C semblent augmenter avec le ratio H/C mais cette tendance n'est pas clairement marquée.

Le Tableau 16 résume le nombre de signaux par composition, par catégorie et par nombre d'atomes de carbones pour les signaux pour lesquels une formule brute a été proposée. Les ratios H/C et O/C correspondant aux catégories sont présentés dans la bibliographie dans le paragraphe II.4.5.2 (Tableau 3).

Tableau 16 : Nombre de signaux en fonction de la composition, de la catégorie et du nombre d'atomes de carbones parmi les 331 signaux pour lesquels une formule brute a pu être associée.

Composition	Nombre de signaux	Catégorie	Nombre de signaux	Nombre d'atomes de C	Nombre de signaux
CHO	27	Lipides (1)	85	1 à 5	6
CHOP	21	Protéines (2)	52	5 à 10	70
CHNO	139	Sucres aminées (3)	1	10 à 15	112
CHNOP	60	Carbohydrates (4)	11	15 à 20	73
CHOPS	6	Structures condensées (5)	38	20 à 50	70
CHNOS	23	Lignines (6)	25	> 50	0
CHN	17	Tannins (7)	2		
CHNOPS	2	« Hors catégories »	117		
Halogènes	12				
Autres	24				

D'après ces résultats, un total de 27% des formules brutes contiendrait au moins un atome de phosphore élément fréquent dans les molécules organiques comme le soufre mais moins présent que les atomes C, H, N, O qui sont généralement majoritaires. Les formules brutes les plus représentées seraient les formules de type CHNO (42%) et CHNOP (18%) suivies par CHO (8%), CHNOS (7%) et CHOP (6%). Ces proportions sont cohérentes avec l'abondance des atomes C, H, N et O généralement observée dans les molécules organiques ainsi que celles du S et du P dans une moindre importance.

Enfin, la Figure 55 l'empreinte des 331 signaux pour lesquels une formule brute a été proposée. Sur cette figure, pour chaque signal, une couleur différente a été attribuée en fonction de la composition des formules brutes sur le graphique A et trois séries d'homologues sont mis en évidence sur le graphique B (Figure 55).

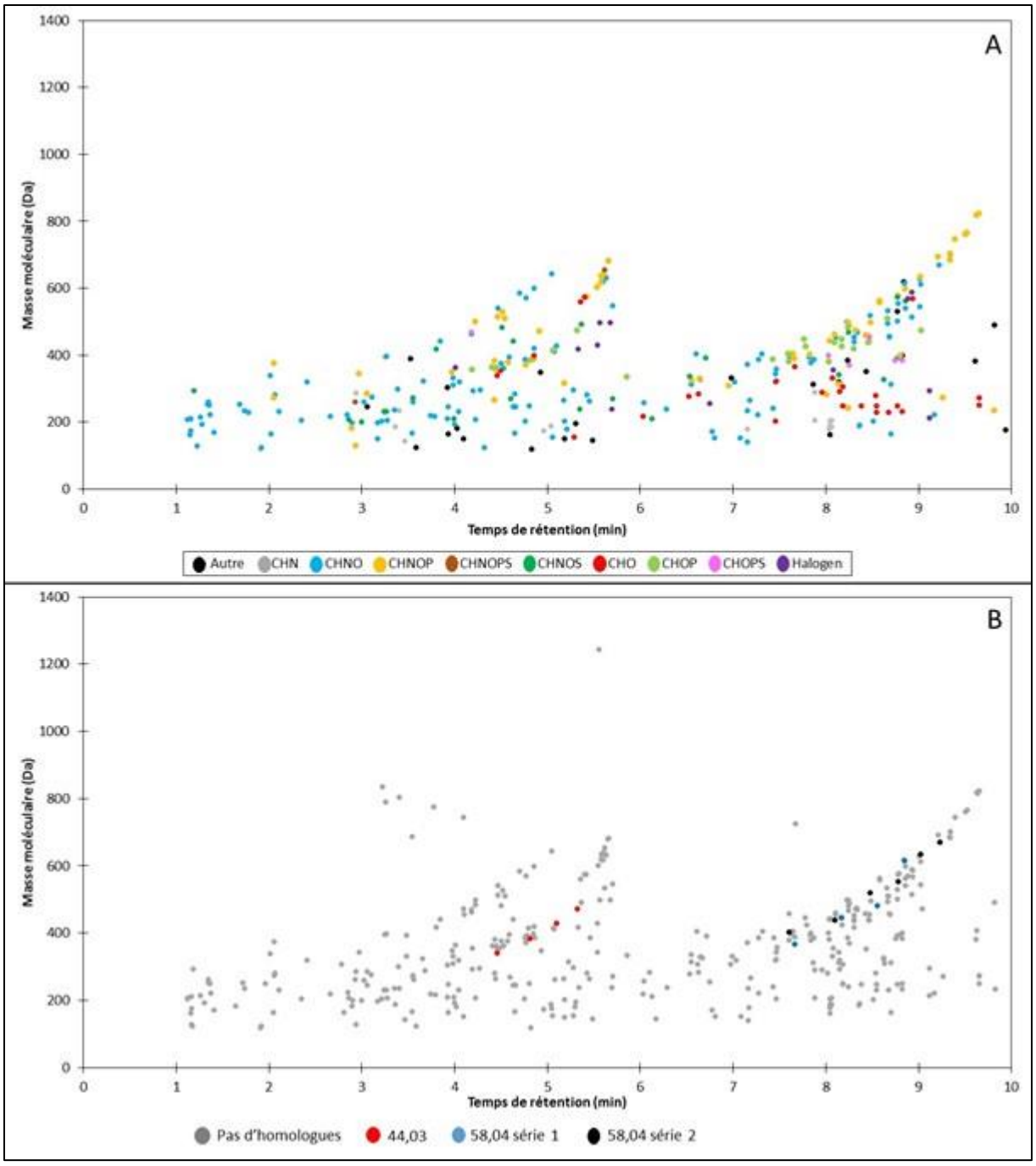


Figure 55 : Empreinte des 331 signaux pour lesquels une formule brute a été proposée. (A) selon la composition, (B) identification de trois séries d'homologues.

La Figure 55A montre que les signaux ayant une composition de type CHNO, CHN, CHONS, CHNOP et autre sont répartis sur l'ensemble de la gamme de temps de rétention. Les signaux dont la composition est de type CHOPS, CHOP, CHO et ceux ayant un halogène sont essentiellement présents après 4 minutes excluant donc une partie des molécules très polaires. Les signaux de type CHNOP et CHOP et CHOPS ont en général des masses plus élevées, montrant une cohérence avec la Figure 54 qui indique qu'une partie d'entre eux sont potentiellement des lipides. Enfin, la composition la plus représentée, CHNO, est répartie sur l'ensemble de l'empreinte pour les temps de rétention et les masses couvrant

ainsi une large gamme de molécules possibles (que ce soit de la matière organique naturelle ou des composés anthropiques). Si des tendances sont observables, les différentes compositions ne concernent alors pas des signaux associés à des propriétés spécifiques (que ce soit en termes de polarité ou en termes de masse). Il n'y a pas de zones caractéristiques associées à un type précis de structure.

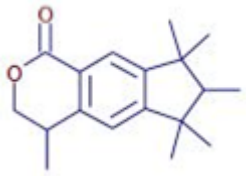
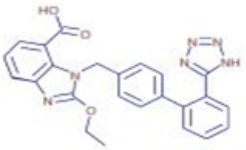
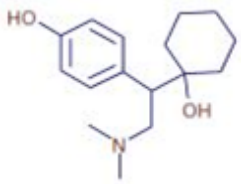
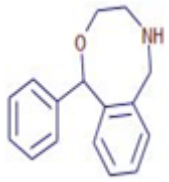
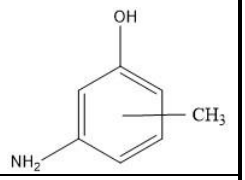
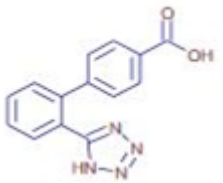
Trois séries différentes d'homologues ont été spécifiquement identifiées et sont représentées dans l'empreinte reprenant l'ensemble des signaux sélectionnés (Figure 55B) : une série séparée par une différence de masse 44,03 correspondant à C_2H_4O et deux séries caractérisées par une différence de masse de 58,04 correspondant à CH_2+C_2H_4O. Les molécules de ces séries pourraient ainsi faire partie de la même famille ou de la même classe de composés.

En conclusion, les signaux pertinents sélectionnés présentent une gamme variée de ratio O/C, N/C et H/C reflétant des structures diverses (aliphatiques, insaturés…) pouvant correspondre pour certains à la matière organique naturelle dissoute et pour d'autres à des molécules d'origine anthropique. Les signaux sélectionnés ne correspondent donc pas à des classes bien spécifiques de molécules.

V.2.2. Proposition d'annotations par analyses des spectres MS^2

Parmi la liste de signaux pertinents sélectionnés, 155 d'entre eux ont été fragmentés avec l'obtention d'un spectre MS^2 lors des analyses UHPLC-HRMS de la campagne C. Pour ces 155 signaux, une recherche de formule développée a alors été réalisée grâce à la base de données mzCloud par comparaison des spectres MS^2. 41 structures pour 25 signaux, présentant une déviation en masse inférieure à 5 ppm, ont ainsi été proposées (Annexe 4). Parmi ces structures, 7 ont présenté un pourcentage de correspondance supérieur à 80% et sont listés dans le Tableau 17 (dont un composé associé à deux signaux différents, et deux structures associées au même signal).

Tableau 17 : Liste des 7 signaux ayant une proposition d'identification par mzCloud avec une déviation en masse inférieure à 5 ppm et une correspondance supérieure à 80%.

Propositions de structures	Formule Brute	Masse exacte (Da)	Tr (min)	Famille	Structure
Galaxolidone	$C_{18}H_{24}O_2$	272,1767	8,87 9,23	Métabolite d'un musc synthétique	
Candesartan	$C_{24}H_{20}N_6O_3$	440,1586	7,92	Produit pharmaceutique	
O-desmethyl venlafaxine	$C_{16}H_{25}NO_2$	263,1881	5,06	Métabolite de produit pharmaceutique	
Desmethyl nefopam	$C_{16}H_{17}NO$	239,1307	5,90	Métabolite de produit pharmaceutique	
5-amino-o-crésol 2-amino-p-crésol	C_7H_9NO	123,0685	1,52	Produit de synthèse	
Métabolite du Valsartan	$C_{14}H_{10}N_4O_2$	266,0800	6,80	Métabolite de produit pharmaceutique	

Parmi ces molécules, deux sont des produits de synthèse et quatre sont des produits pharmaceutiques (1) ou des métabolites de produits pharmaceutiques (3). La dernière molécule de cette liste est un métabolite du musc galaxoline. La présence dans les STEP de certains de ces produits a déjà été décrite dans la bibliographie. C'est le cas notamment du métabolite du musc galaxoline dont la biotransformation dans une station d'épuration a déjà été reporté par Tasselli et Guzzela (2020), du o-desmethylvenlafaxine qui a été détecté dans deux STEP par Xing *et al.* (2018), ou encore du candesartan qui, lui aussi a été quantifiée dans des sorties de STEP (Castro *et al.*, 2019). La présence du desmethyl nefopam n'a, à notre connaissance, pas été encore reportée dans des eaux de STEP mais seulement dans des échantillons de plasma humain (Hoizey *et al.*, 2006). Dans le cadre de cette étude,

ces annotations n'ont pas été confirmées par l'injection d'étalons. Toutefois dans le cadre de ce travail, à côté de ces structures, le diclofénac a également été suspecté comme étant associé à un signal pertinent suite à l'injection d'une solution étalon. Un pic chromatographique avec un temps de rétention et un spectre de masse MS^1 similaire à celui du diclofénac a en effet été observé lors de l'analyse des échantillons non dopés.

Néanmoins, dans le cas de l'échantillon non dopé, le signal d'intérêt n'a pas fait partie des cinq signaux les plus intenses passant à la fragmentation (principe de la ddMS2 top-N) lors des analyses MS^2. Ceci ne permet donc pas de conclure formellement sur l'identité de ce signal mais seulement de donner un faisceau d'indices orientant vers l'identification du diclofénac.

Cette molécule a été détectée à de nombreuses reprises dans les eaux usées (Loos *et al.*, 2013). Dans le cas où sa présence serait avérée ici, le diclofénac serait alors un composé dont la valeur d'aire augmente dans les échantillons de sorties sans que cela soit réellement significatif (ratio entre sorties et entrées compris entre 1,4 et 1,7 pour les trois campagnes). Dans ces conditions, trois hypothèses seraient possibles :

- le diclofénac est réfractaire au traitement ;
- les métabolites du diclofénac, issus de la métabolisation dans le corps humain, seraient déconjugués dans la filière d'épuration et libèreraient alors la molécule parente ;
- ou un effet matrice bien plus important que celui estimé en entrée de STEP malgré l'application du facteur de correction (facteur d'environ 2) serait observé et indiquerait ainsi une fausse augmentation.

Cet exemple, à confirmer, du diclofénac illustre la problématique de l'incertitude liée à l'effet matrice qui ne peut être levé que par la quantification avec l'étalon marqué correspondant ou par la méthode des ajouts dosés. Cependant, à partir de cette approche, il est toujours possible de conclure sur une absence d'élimination complète de la molécule par le traitement. Dans le cas d'autres produits pharmaceutiques (par exemple le candesartan) dont les identités restent également à confirmer, le même cas de figure serait observé.

Ainsi sur les 375 signaux sélectionnés initialement, une formule brute a été attribuée pour 88% d'entre eux (331 signaux), pour 41% (155 signaux) un spectre MS^2 a été obtenu, 7% (25 signaux) présentent un spectre MS^2 ayant des similitudes avec des spectres MS^2 de mzCloud et, seuls 1,9% (7 signaux) ont un pourcentage de correspondance supérieur à 80%.

L'appel à d'autres bases de données telles que PubChem mènerait probablement à d'autres propositions de structures. Toutefois, cela ne garantirait pas un plus grand nombre de molécules identifiées. De plus, le facteur limitant resterait toujours la disponibilité d'étalons de référence. La prise en compte de l'environnement de prélèvements et du type de molécules pouvant y être présentes peut aider à la prise de décision quant à la pertinence des propositions de structures (Schymanski *et al.*, 2017). En ce sens la proposition de produits pharmaceutiques ou de métabolites de produits pharmaceutiques dans des eaux de station d'épuration urbaines est cohérente. Si l'identification est nécessaire pour le développement de la connaissance, la quantification et une réduction de l'incertitude, l'absence d'identité de ces signaux n'empêche néanmoins pas le suivi de ces signaux dans les milieux récepteurs lorsque l'on dispose des spectres MS^2.

V.3. Recherche des signaux pertinents de la liste de Poitiers dans les échantillons de Saint Julien l'Ars

La liste des signaux pertinents obtenus à partir des échantillons de Poitiers a ensuite été intégrée dans un nouveau traitement par Compound Discoverer des données issues de l'analyse des échantillons de Saint Julien l'Ars. L'objectif était de déterminer si les signaux relevés pour les échantillons d'une STEP donnée sont spécifiques à cette STEP ou non et de lister les signaux communs entre les deux STEP. Les paramètres pour la recherche de ces signaux ont été les tolérances en masse et le temps de rétention (réglés respectivement à 5 ppm et 0,2 min). A partir de cette approche, sur les 375 signaux de départ retenus pour pertinents au travers des analyses des échantillons de la STEP de Poitiers, 203 ont été suspectés comme étant présents dans les échantillons de Saint Julien l'Ars (soit 54%). Les mécanismes de réactions, s'effectuant lors de l'épuration par boues activées de filières similaires, sont partiellement identiques entre deux STEP différentes. Il semble alors cohérent de retrouver à la STEP de Saint Julien l'Ars un grand nombre des signaux retenus comme pertinents pour la STEP de Poitiers.

Ces résultats montrent la contribution possible des stations d'épuration dans la dissémination dans les milieux aquatiques de contaminants, de produits de transformations ou de dégradations. Toutefois en l'absence d'identification ou de données de toxicité, on ne peut pas présumer de l'impact négatif sur les milieux du rejet des molécules correspondant aux signaux retenus.

VI. Recherche de suspects et de métabolites dans les eaux de STEP

VI.1. Recherche de suspects dans les eaux de STEP

La détection de la présence de micropolluants aurait pu être effectuée en partant de la liste de spectres MS^2 de mzCloud ayant une correspondance avec un ou plusieurs signaux. Cependant de nombreuses molécules répertoriées dans mzCloud ne sont pas en cohérence avec le milieu étudié.

De plus de nombreux composés de cette base de données n'ont pas de nom commun usuel et sont uniquement identifiés par leur nom IUPAC ce qui rend difficile d'établir *a posteriori* si leur présence est possible dans le milieu considéré. Utiliser directement la liste des spectres MS^2 de mzCloud (comme précédemment pour la recherche de signaux pertinents) peut donc prendre un temps conséquent si on prend en compte la totalité des signaux des échantillons.

Dans le but de rechercher la présence potentielle de micropolluants dans les échantillons analysés, une analyse de suspects, telle que décrite dans le chapitre « Bibliographie » (paragraphe II.3.) a alors été réalisée. L'analyse de suspects permet de sélectionner un large panel de composés susceptibles de se trouver dans cette catégorie d'échantillons et suit plusieurs grandes étapes dont :

- création de la liste de suspects,
- prétraitement des données permettant la détection automatique des signaux dans les échantillons,
- confrontation des masses exactes détectées et des masses exactes de la liste de suspects,
- réduction du nombre de suspects par l'élimination des faux positifs, par la forme du pic, par les profils isotopiques, ou encore par la prédiction des temps de rétention etc,
- confrontation entre les spectres MS^2 obtenus et ceux répertoriés dans des bases de données.

Ces étapes excluent au fur et à mesure des suspects pour ne conserver que ceux respectant les critères (temps de rétention et correspondance du spectre MS^2) ou ceux confirmés par l'injection d'un étalon de référence.

VI.1.1. Création et mise en place d'une liste de suspects

La création d'une liste de suspects permet l'inclusion d'un grand nombre de composés, supérieur au millier par exemple. Cette liste se construit en fonction de la problématique étudiée pour y inclure des composés en cohérence avec le domaine d'étude. Elle peut être construite à partir de la bibliographie, de retours d'expériences, de listes ou de bases de données déjà existantes.

Les eaux de STEP étudiées proviennent très majoritairement d'usages domestiques et hospitaliers. De nombreuses molécules sont donc susceptibles de les contaminer. Par conséquent, la liste de suspects appliquée à ce type de matrice peut être composée d'une grande variété de molécules.

Les eaux de rivières sont exposées à de multiples sources de contamination : lessivage des routes, ruissellements urbains, rejets de STEP, déversoir d'orages, accidents ponctuels etc. Tout comme les eaux de STEP, les eaux de rivières reçoivent donc potentiellement une large gamme de polluants.

En raison de ces multiples sources et types de contaminants susceptibles de se retrouver dans les échantillons étudiés, une liste constituée de plusieurs classes de composés (1 487 molécules) a été considérées dans cette partie : produits pharmaceutiques, produits cosmétiques, composés poly ou perfluorés, retardateurs de flammes, plastifiants, produits anticorrosion, pesticides, produits dérivés de la cigarette, produits illicites, surfactants/désinfectants, hormones, produits de consommation, produits industriels, et certains métabolites ou produits de transformation de ces molécules. Pour établir cette liste, plusieurs autres listes ont été fusionnées. Ces listes proviennent à la fois des listes du réseau NORMAN et de méthodes d'analyse multi-résidus décrites dans la bibliographie.

Après création de cette liste, des étapes de recherche de suspects ont été intégrées dans le logiciel Compound Discoverer.

VI.1.2. Application de l'analyse de suspects dans les eaux de STEP

La recherche de suspects a été réalisée dans toutes les matrices eaux de la STEP de Poitiers, entrée et sorties.

Pour cela, la liste initiale comprenant 1 487 molécules et élaborée dans la présente étude a été considérée. L'analyse de suspects a été effectuée par un traitement des données réalisé directement avec Compound Discoverer. Lors de ce traitement, après application de cette liste sur le tableau de signaux détectés, 1 207 signaux ont présenté une masse exacte correspondant à au moins un suspect (différence de masse < 10 ppm) et 956 d'entre eux ont été fragmentés lors des analyses MS^2, soit 64% du nombre initial.
Après vérification de la compatibilité entre les temps de rétentions des signaux avec les logP et logD calculés par ChemAxon puis élimination des faux positifs manuellement à partir des chromatogrammes, 198 signaux correspondant à un ou plusieurs suspects (soit 13% du nombre initial) étaient alors encore présents. La liste des 198 suspects associées à ces signaux (présentant une correspondance avec une base de données avec une déviation en masse < 5 ppm) est présentée en Annexe 5 avec le pourcentage de correspondance. Dans cette liste, un suspect peut correspondre à

plusieurs signaux. La présence de nombreux produits pharmaceutiques parmi les suspects, est cohérente avec l'origine des eaux analysées qui proviennent d'une station d'épuration urbaine connectée au Centre Hospitalier Universitaire de Poitiers.

Parmi ces 198 suspects, la base de données de spectres MS^2 de mzCloud englobe les spectres MS^2 pour 115 d'entre eux et 74 ont montré une correspondance supérieure à 80% avec les spectres expérimentaux des échantillons (soit 58% des 198 et 8% de la liste initiale). Pour les 83 autres suspects, les spectres de masse ont été soumis à MetFrag (Ruttkies *et al.*, 2016) avec les bases de données KEGG, HMDB et CompTox du 7 mars 2019. 31 d'entre eux ont alors montré un spectre de fragmentation MS^2 compatible à plus de 50% avec les spectres de MetFrag. 31 autres ont un spectre de fragmentation MS^2 compatible à moins de 50%.

Enfin, pour 9 suspects, des injections de standard ont été réalisées pour confirmer leur présence. Par cette dernière approche, 8 composés ont ainsi été confirmés dont la caféine, la ciprofloxacine, la carbamazépine, le propranolol, le ketoprofène, l'oxybenzone, l'ioméprol et le valsartan. Six des huit suspects confirmés ont des pourcentages de correspondance supérieurs à 80% par mzCloud. Cependant, un des suspects présente un pourcentage inférieur à 80 % et un autre n'a de réponse que dans MetFrag, avec une correspondance inférieure à 50%. Pour ces composés, les chromatogrammes, les spectres MS^2 des étalons et d'échantillons sont présentés en Annexe 6. Ces molécules font parties des molécules les plus fréquemment détectées dans les eaux de STEP (Loos *et al.*, 2013). Dans le cas du tramadol (présentant une correspondance <80% dans mzCloud), un temps de rétention correspondant bien à celui de l'étalon a été observé mais un spectre MS^2 différent a été noté.

Ces résultats indiquent donc, que, dans le cas de produits pharmaceutiques, la base de données mzCloud ne tend à donner que des indications sur la fragmentation et que les résultats doivent être confirmés par l'utilisation d'étalons. Les deux suspects confirmés dans le cadre de ce travail avec des pourcentages d'équivalence inférieurs à 50% dans MetFrag ou 80% dans mzCloud montrent que ces seuils restent subjectifs et que les suspects ayant un pourcentage inférieur à ces seuils doivent également être considérés.

La Figure 56 suivante résume les étapes effectuées ainsi que le nombre de signaux restants à chaque étape.

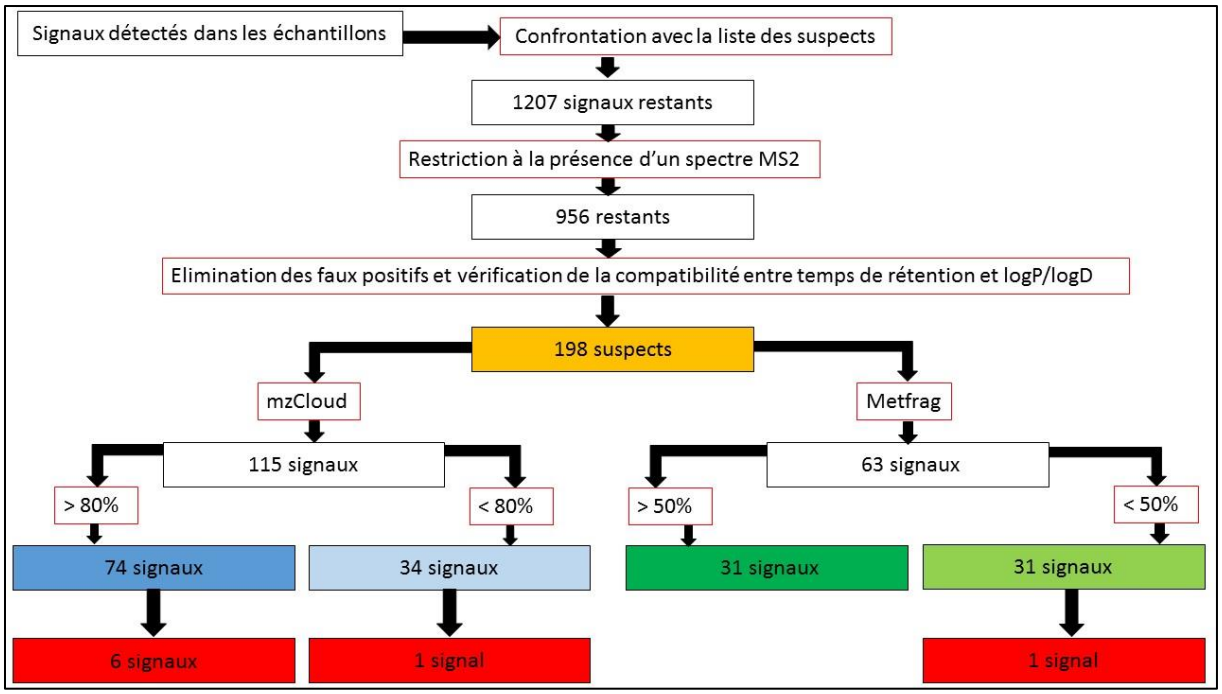


Figure 56 : Schéma d'étapes de l'analyse de suspects ayant conduit à l'obtention d'une liste de 198 signaux confrontés à mzCloud ou MetFrag (KEGG, HMDB et Comptox).

La confirmation d'un faible nombre de suspects, observée dans cette étude, par rapport à la taille de la liste initiale n'est pas surprenante. Ce nombre restreint d'identifications est en effet courant lors de ce type d'approche (Chiaia-Hernandez *et al.*, 2014 ; Sjerps *et al.*, 2016).

Parmi les 198 suspects, il est à noter que 148 sont présents à la fois dans les entrées et les sorties de STEP, 40 sont uniquement présents dans les entrées et 10 sont observés uniquement dans les sorties de STEP. La majorité des suspects (74%) sont donc présents dans les deux types d'échantillons. Ainsi, dans le cas où ces suspects seraient confirmés, cela signifie que de nombreux micropolluants ne sont pas éliminés par la STEP et sont relargués dans le milieu aquatique. Cela rejoint les résultats des analyses effectués par Grand Poitiers qui montrent que des dizaines de micropolluants sont détectés dans les eaux de sorties, notamment des pesticides. Cela est également cohérent avec la compilation d'études de Margot *et al.* (2015) qui a répertorié de nombreux polluants dont les taux d'abattements sont inférieurs à 70%.

Les concentrations des 8 suspects confirmés ont été estimées dans un des échantillons de la campagne C (Tableau 18), sur la base de l'analyse d'un unique point d'étalonnage en tenant compte des rendements d'extractions et des effets matrices déterminés précédemment (chapitre 3 « Mise au point de l'approche », paragraphe VI.2.).

Tableau 18 : Concentrations estimées pour les 8 suspects confirmées dans un échantillon d'entrée et un échantillon de sortie de STEP de la campagne C.

Molécule	Concentration en entrée (µg/L)	Concentration en sortie (µg/L)
Caféine	200	0,37
Carbamazépine	4,9	3,8
Ciprofloxacine	6,2	0,58
Ioméprol	100	11
Kétoprofène	26	4,2
Oxybenzone	160	100
Propranolol	0,29	-
Valsartan	92	31

Les concentrations déterminées sont du même ordre de grandeur que celles souvent mesurées pour les micropolluants dans les eaux de stations d'épuration. Les concentrations diminuent pour toutes les molécules entre l'entrée et la sortie de la STEP. Ces concentrations sont à considérer avec précaution. Néanmoins cela donne une indication de la présence de ces molécules à des concentrations significatives au vu de l'impact environnemental.

Dans cette partie, les suspects identifiés et la liste de suspects utilisée sont plutôt orientés vers les produits pharmaceutiques bien que d'autres familles de molécules soient présentes. Grand Poitiers, qui gère la station d'épuration, procède régulièrement, dans le cadre de la démarche RDSE (Recherche et réduction des rejets de substances dangereuses dans l'eau) à l'analyse d'une liste de micropolluants par des laboratoires prestataires. Ils ont ainsi analysé plusieurs pesticides, composés poly ou perfluorés, des polychlorobiphényles, des phénols ou encore hydrocarbures polycycliques aromatiques. Les gammes de concentrations détectées en entrée et en sortie de STEP sont présentées dans le Tableau 19.

Tableau 19 : Gamme de concentrations des classes de molécules détectées dans les eaux de STEP par Grand Poitiers.

Classe de molécules	Concentration en entrée (µg/L)	Concentration en sortie (µg/L)
Pesticides	<0,015 – 2,5	<0,005 – 0,56
Perfluorés	<0,05 - <45	<0,05
Polychlorobiphényles	<0,005 - <10	<0,05 - <0,01
Phénols	<0,05 - 190	<0,05 - <0,10
Hydrocarbures polycycliques aromatiques	<0,005 - 116	<0,0005 – 0,01

Ces données, combinées aux résultats obtenus dans ce travail lors de l'analyse de suspects, montrent bien la variété des molécules présentes dans les eaux de STEP et les ordres de grandeur des concentrations retrouvées. Ces données sont donc complémentaires. L'analyse de suspects peut servir à étendre un champ de recherche de molécules non analysés habituellement en recherchant rapidement une grande liste de composés dans divers échantillons. Une recherche en considérant l'ensemble des spectres MS^2 de mzCloud aurait probablement permis de compléter la liste des molécules présentes au niveau de la STEP. Une première recherche *via* cette approche (sans vérification de la pertinence des structures proposées par rapport à l'échantillon analysé) a en effet montré que 764 molécules supplémentaires, non présentes dans notre liste de suspects, seraient proposées. Des produits pharmaceutiques, des hormones et leurs dérivés ou des pesticides sont retrouvés parmi ces molécules ainsi que de nombreux composés (50% environ) n'ayant pas de noms communs/usuels (uniquement identifiées par leur nom IUPAC).

VI.1.3. Confrontation entre les suspects détectés et la liste de signaux pertinents

La liste des 198 suspects isolés a été confrontée à la liste des 155 signaux pertinents présentant un spectre MS^2. Ces 2 listes sont composées de signaux possédant un spectre MS^2 permettant d'appuyer la présence d'un même signal dans les deux listes. Par cette approche, six signaux communs ont été trouvés et sont répertoriés dans le Tableau 20.

Tableau 20 : Liste des signaux présents dans la liste de 198 signaux suspects et la liste des 155 signaux pertinents ayant un spectre MS^2. * : pourcentage de correspondance inférieur à 80% dans mzCloud. Tr : Temps de rétention.

Nom(s) suspect(s)	Formule brute	Masse exacte (Da)	Tr (min)	Base de données
O-Desmethylvenlafaxine	$C_{16}H_{25}NO_2$	263,1881	5,06	mzCloud
N-Desmethylvenlafaxine				MetFrag
Oxprenolol	$C_{15}H_{23}NO_3$	265,1681	4,76	MetFrag
Galaxolidone	$C_{18}H_{24}O_2$	272,1767	9,27	mzCloud
β-Estradiol*				mzCloud
19-Norandrostenedione*				mzCloud
Cycloheximide	$C_{15}H_{23}NO_4$	281,1623	4,05	MetFrag
Kresoxim méthyle	$C_{18}H_{19}NO_4$	313,1310	8,29	MetFrag
Candesartan	$C_{24}H_{20}N_6O_3$	440,1586	7,92	mzCloud

Dans ce tableau, 5 structures (dont 3 avec un pourcentage de correspondance > 80 %) des 41 propositions faites précédemment pour des signaux pertinents (Annexe 4) sont présentes. Dans le cas

des propositions multiples O-desmethylvenlafaxine/N-desmethylvenlafaxine et estradiol/19-Norandrostenedione, les structures des suspects sont très proches. Les structures du galaxolidone et de l'estradiol sont en revanche plus éloignées. Les signaux associés au kresoxim méthyle, cyclohexemide et oxprenolol ont été listés comme présents uniquement dans les échantillons de sorties.

Ainsi, l'analyse de suspects a enrichi le nombre de propositions pour les signaux pertinents. Ce résultat indique l'intérêt de mener une analyse de suspects et une analyse non ciblée en même temps pour croiser les résultats comme le souligne par ailleurs Krauss *et al.* (2010) qui parle d'analyses « all-in-one » (ciblé, suspects et non-ciblé).

VI.2. Recherche de métabolites complémentaires dans les eaux de STEP

Les métabolites sont les molécules issues des transformations chimiques ou biochimiques s'opérant à différents niveaux dans le milieu naturel ou en amont. On peut distinguer d'une part les métabolites provenant de la métabolisation humaine ou animale et d'autre part les produits de transformation s'effectuant dans les STEP ou dans le milieu naturel par divers mécanismes. Une partie très importante de ces composés reste inconnue à ce jour (Li *et al.*, 2017) mais leur analyse ne doit pas être négligée car la toxicité d'un métabolite peut parfois être significative voire supérieure à celle de la molécule parente (Farré *et al.*, 2008). La combinaison de l'analyse non ciblée, de l'analyse de suspects et de la prédiction *in silico* de métabolites peut apporter de nouvelles informations sur l'existence et la présence de métabolites.

VI.2.1. Constitution de la liste de métabolites suspects

La recherche de métabolites peut s'effectuer de deux façons principales : l'une par les données acquises avec l'analyse de données et, la seconde, par la prise en compte d'une liste de molécules de départ et des réactions susceptibles de se produire dans le milieu. La seconde méthode se rapporte plus à de l'analyse de suspects que purement à de l'analyse non ciblée. Dans le cadre de ce travail, la deuxième approche a été retenue. Une liste de métabolites a ainsi été générée à partir de 60 molécules régulièrement retrouvées dans les STEP (Loos *et al*, 2013 ; Rapport Sipibel 2011-2015) (Annexe 7). Pour cela, la méthode SyGMA (Systematic Generation of potential Metabolites) (Ridder et Wagener, 2008) a été employée pour conserver les métabolites les plus probables en termes de stabilité et de formation lors de réactions chimiques de type oxydation, hydrolyse-, réduction ou condensation. Une liste de 233 potentiels métabolites a ainsi été générée. Pour chacun de ces métabolites, les logP et logD (pH 3) ont été calculés par ChemAxon.

VI.2.2. Recherche de métabolites suspects

A partir de cette liste, une recherche des métabolites suspects a alors été effectuée sur les échantillons de la STEP de Poitiers par Compound Discoverer. Dans un premier temps, il est apparu qu'un total de 826 signaux possédait une masse exacte correspondant à celle d'un métabolite suspect (différence de masse < 10 ppm). Parmi ces signaux, 202 ont été fragmentés lors des analyses MS^2 et après élimination des faux positifs, 89 signaux étaient encore présents. Ces 89 signaux sont présentés en Annexe 8. 19 de ces derniers signaux ont montré un spectre MS^2 similaire à celui de la base de données mzCloud (dont 16 signaux avec des pourcentages supérieurs à 80%). 25 autres ont montré un spectre MS^2 compatible à une base de données sur MetFrag (dont 7 avec plus de 50% de correspondance). Le Tableau 21 liste les 19 signaux ayant une correspondance avec mzCloud.

Tableau 21 : Liste des métabolites suspects ayant une correspondance avec mzCloud. N : nature de l'échantillon dans lequel le signal a été détecté, Tr : temps de rétention. * : molécule parente présente dans la liste de 198 signaux suspects.

Nom du métabolite	Formule brute	Masse exacte	Tr (min)	Molécule parente	N
p-phenylenediamine	$C_6H_8N_2$	109,0763	0,71	4'-Aminoacetanilide	E
			0,92		S
4-aminophenol	C_6H_7NO	110,0610	0,89	Paracétamol*	E/S
Théophylline	$C_7H_8N4O_2$	180,0689	3,38	Caféine*	E/S
Théobromine			3,93		
Paraxanthine					
Ecgonine méthyle ester	$C_{10}H_{17}NO_3$	199,1212	0,76	Cocaïne*	S
			6,33		E/S
4-acetaminoantipyrine	$C_{13}H_{15}N_3O_2$	245,1168	4,34	4-Formylaminoantipyrine*	E/S
O-desméthyltramadol	$C_{15}H_{23}NO_2$	249,1726	5,30	Tramadol	E/S
			4,16		
N-desméthyltramadol			4,73		
N-desméthylvenlafaxine	$C_{16}H_{25}NO_2$	263,1882	5,06	Venlafaxine*	E/S
O-desméthylvenlafaxine					
Aténolol acide	$C_{14}H_{21}NO_4$	268,1559	4,03	Aténolol*	E/S
			4,17		
			4,23		
			4,38		
Benzoylecgonine	$C_{16}H_{19}NO_4$	289,1312	4,95	Cocaïne*	E/S
N-acetylsulfamethoxazole	$C_{12}H_{13}N_3O_4S$	295,0630	5,75	Sulfamethoxazole*	E/S

Ces résultats laissent envisager la présence de métabolites dont les molécules parentes concernées seraient la caféine, l'aténolol, la venlafaxine, la cocaïne, le tramadol, le sulfamethoxazole et le 4'-aminoacetanilide. En l'absence d'étalon, aucune confirmation n'a cependant pu être effectuée. Parmi les molécules parentes suspectées, la caféine a été identifiée comme présents dans des échantillons. L'aténolol, la venlafaxine, la cocaïne et le sulfamethoxazole restent quant à eux des suspects potentiels

à confirmer avec des spectres MS^2 existants dans la base de données mzCloud. La présence de ces métabolites est donc cohérente avec ces résultats. Toutefois, 16 de ces signaux sont présents à la fois dans les entrées et dans les sorties et un n'est présent qu'en entrée. Pour ces métabolites, une formation en amont de la station d'épuration (i.e. métabolites humains, ou formation préalable dans le réseau d'assainissement) semble envisageable. Dans le cas de l'ecgnonine méthyle ester et du p-phenylenediamine uniquement présents en sortie de STEP, leur formation pourrait provenir directement des réactions liées aux procédés d'épuration biologique.

Lors d'une telle approche, l'absence de réponse des bases de données interrogées peut s'expliquer soit par l'absence de données spectrales ou parce que le signal ne correspond pas à la molécule. Dans ces conditions, le processus d'identification devient alors plus complexe. Ceci montre donc la nécessité de compléter les bases de données en signaux jugés pertinents et l'intérêt des bases de données internes dans le cas de répétition de détection de certains signaux. La prédiction de métabolites *in silico* peut faciliter et constituer un point de départ dans la recherche de nouveaux métabolites. Cependant, le manque de références dans les bases données implique généralement un très long travail d'identification qui, bien souvent, reste limité par l'absence d'étalon de référence.

La confrontation entre la liste de métabolites suspects dans cette partie et la liste de signaux pertinents n'a montré aucun signal commun.

VII. Conclusions

L'analyse non ciblée appliquée aux eaux de stations d'épuration de Poitiers et de Saint Julien l'Ars a permis de mettre en évidence que les eaux d'entrées contiennent de nombreuses molécules de tailles et de polarités diversifiées, confirmant la nature très complexe de ces matrices. Au cours des trois campagnes de prélèvements, effectuées sur un an, il a été montré une certaine homogénéité des entrées de STEP, quelle que soit la station d'épuration étudiée (celle de Poitiers ou de Saint Julien l'Ars).

Il a été observé que le traitement des eaux a un impact sur le nombre, la taille, la polarité et la concentration des molécules correspondant aux signaux. Ces quatre paramètres ont en effet vu leur valeur diminuer en sortie de STEP lors des trois campagnes. Au cours du traitement de l'eau, plus de 90% du nombre total de signaux disparait ou est abattu à plus de 50% en sortie. Néanmoins, les empreintes des eaux ont également témoigné de l'augmentation de l'aire de certains signaux et de l'apparition de nouveaux signaux au cours du traitement des eaux (ces signaux compteraient pour 17 à 33% du total de signaux en sortie). Enfin, d'autres signaux semblent peu impactés par le traitement

et peuvent donc être considérés comme correspondant à des molécules réfractaires au traitement. Ces phénomènes d'augmentation et d'apparition de signaux et de signaux réfractaires ont été observés sur l'ensemble des échantillons (pour les deux STEP et les trois campagnes). Si les signaux concernés ne sont pas toujours les mêmes, la gamme de masses et la gamme de polarités semblent en revanche similaires.

Des analyses en composantes principales ont montré des regroupements d'échantillons de même type (entrées et sorties) et la corrélation des signaux avec les échantillons qui peut s'expliquer par le statut (présence ou absence) du signal dans l'échantillon ou par la différence de valeur d'aire du signal entre entrées et sorties. Ainsi, il a été observé que les signaux diminuants ou disparaissant sont corrélés avec les entrées et les signaux augmentant ou apparaissant sont corrélés avec les sorties. Les signaux réfractaires semblent quant à eux peu corrélés avec l'un ou l'autre type d'échantillons. Ces résultats sont valables pour les deux STEP mais d'autres campagnes d'échantillonnages et d'analyses seraient nécessaires pour confirmer cette conclusion.

Grâce à l'analyse différentielle, 375 signaux ont été sélectionnés à partir des analyses des échantillons de Poitiers. Ces signaux correspondent à des molécules qui pourraient impacter le milieu récepteur en raison de leur présence dans les eaux traitées de l'ensemble des campagnes et pourraient ainsi générer une pollution Ces signaux sont corrélés avec les sorties sur l'ACP. Pour 331 d'entre eux, une formule brute a été attribuée et les rapports O/C, N/C et H/C ont été calculés montrant une diversité de structures possibles. Ces signaux pourraient ainsi correspondre à des constituants de la matière organique naturelle (lipides, acides aminées, etc…) ou correspondre à des molécules anthropiques. Parmi ces signaux, 25 (dont 7 produits pharmaceutiques) ont un spectre MS^2 compatible avec les spectres MS^2 de la base de données mzCloud et ont été associés à une formule développée.

Enfin, dans ce chapitre de thèse, la recherche de suspects a permis d'envisager la présence dans les eaux de STEP de 8 molécules, dont 6 produits pharmaceutiques. D'autres suspects, présentant des spectres MS^2 compatibles avec des spectres de bases de données ont également été proposés. Toutefois, leurs structures restent à confirmer par l'injection d'étalons. Dans le cas des métabolites, il a été suspecté, la présence de composés générés à partir de la caféine, de l'aténolol, de la venlafaxine, de la cocaïne, sulfamethoxazole et du 4'-Aminoacetanilide, possédant un spectre de masse MS^2 cohérent avec la base de données mzCloud, et dont l'existence peut se justifier dans le contexte de la présence de produits pharmaceutiques dans les stations d'épuration.

Chapitre 5 : Analyse non ciblée appliquée aux eaux de rivières et impact des rejets de la STEP de Poitiers sur les eaux du Clain

I. Introduction

Les eaux de surface constituent le principal milieu récepteur des eaux de rejets des stations d'épuration. Ce milieu aquatique est déjà soumis à la pression des eaux de pluies, des eaux de ruissellement et des pollutions accidentelles. Par conséquent, le rejet des eaux traitées de STEP peut constituer une pression supplémentaire sur la qualité des eaux de surface par l'apport de nouveaux polluants. Si la qualité des eaux traitées et rejetées par les STEP fait l'objet de réglementations, le chapitre précédent a montré qu'il reste de nombreuses molécules organiques dans les eaux traitées caractérisées par de nombreux signaux dans les empreintes de ces eaux, dont de nouveaux signaux formés durant le traitement. Dans le cas de la STEP de Poitiers, ces molécules sont présentes dans les effluents rejetés dans le Clain adjacent. L'utilisation de l'analyse non ciblée devrait contribuer à mieux cerner l'impact potentiel de la STEP sur le profil moléculaire du Clain et permettre d'examiner si cet impact s'atténue ou non immédiatement en aval.

Dans ce chapitre, la rivière Clain qui reçoit les eaux de la STEP de Poitiers sera principalement étudiée. Le Clain en amont de la STEP sera tout d'abord caractérisé et comparé avec la rivière de la Vienne. La Vienne et le Clain sont deux rivières qui parcourent le département de la Vienne, le Clain traversant Poitiers et se déversant dans la Vienne à une trentaine de kilomètres de Poitiers. La variabilité spatiale du Clain sera ensuite étudiée. La comparaison de la zone de rejet des eaux traitées de la STEP avec des eaux prélevées dans le Clain à l'aval immédiat de cette zone, puis avec les échantillons de sortie de STEP sera effectuée. De même l'amont et l'aval au rejet dans le Clain seront comparés et une analyse multivariée sera réalisée. A partir de ces résultats, des signaux jugés pertinents seront mis en évidence entre l'amont et l'aval puis caractérisés. Enfin, seront présentés les résultats obtenus lors de la recherche de suspects (réalisée à partir d'une liste préétablie de molécules et de propositions de métabolites générées *in silico*). Pour terminer, une conclusion reprendra l'ensemble des résultats obtenus dans ce chapitre.

I.1. Caractérisation d'eaux de rivières : comparaison entre le Clain à l'amont du rejet de la STEP et la Vienne

Au cours des trois campagnes de prélèvements, cinq échantillons d'eaux du Clain prélevés à l'amont du rejet de la STEP ont été effectués (dont 3 lors de la campagne C). Pour la Vienne, un échantillon lors de la campagne A et un échantillon lors de la campagne B ont été prélevés (voir Tableau 4 du chapitre 2 « Matériels et méthodes »). Les empreintes des échantillons de l'amont du Clain et de la Vienne ont

été réalisées après analyse par UHPLC-HRMS et traitements des données par Compound Discoverer et Excel. Ces empreintes sont présentées dans les Figures 57 et 58.

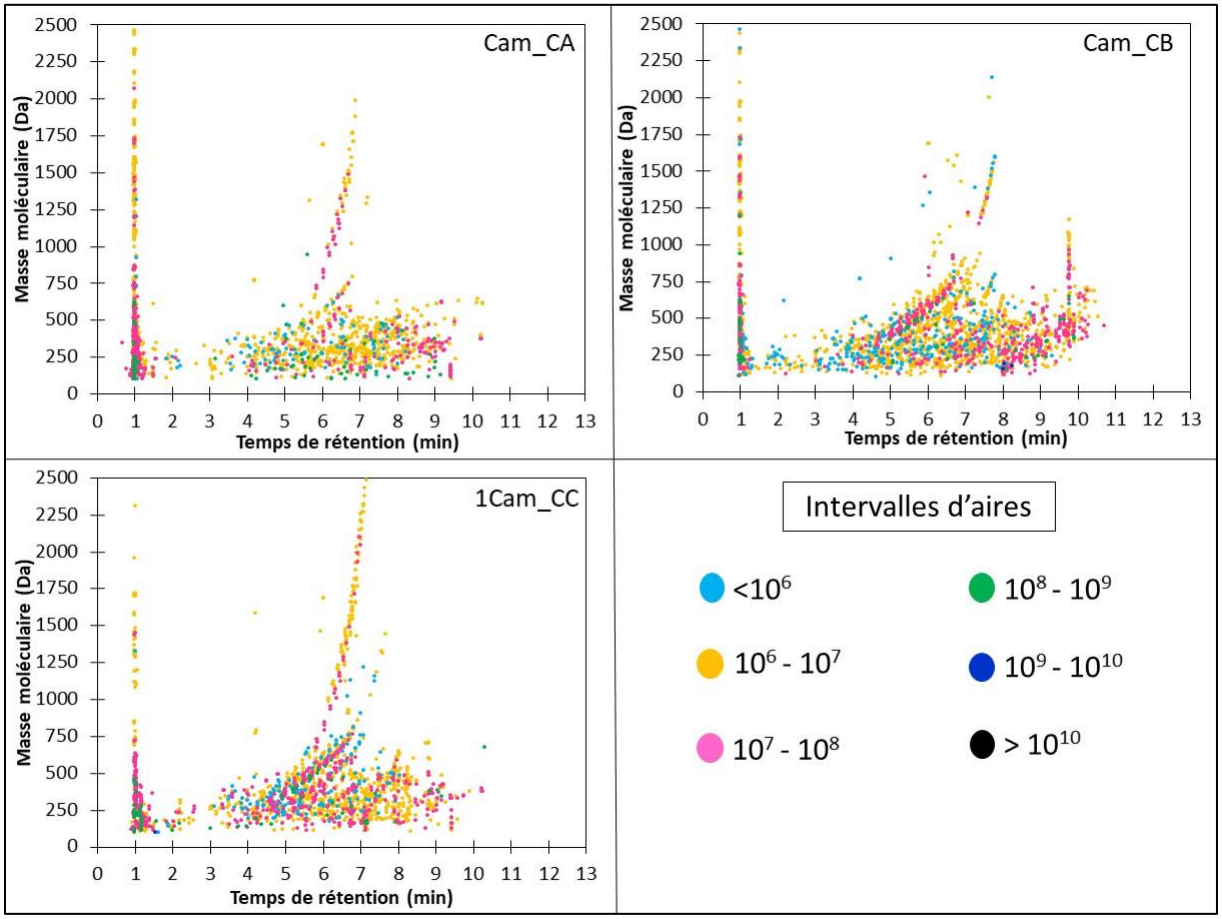


Figure 57 : Empreintes des échantillons de l'amont du Clain pour les campagnes A, B et C. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. Cam : amont du Clain, CA : campagne A, CB : campagne B, CC : campagne C (échantillon 1). Mode positif (HESI+).

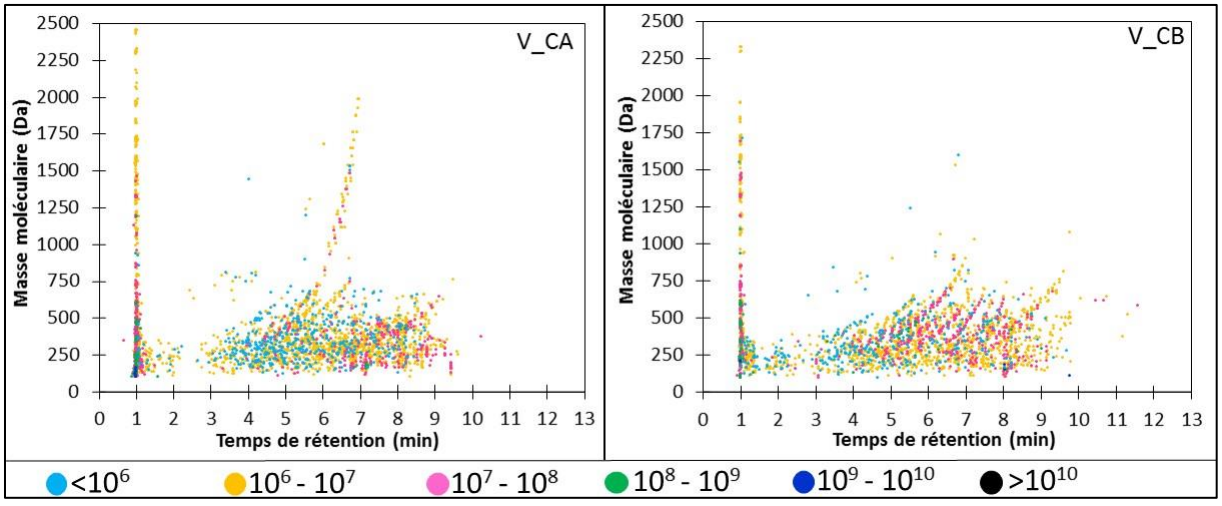


Figure 58 : Empreintes des échantillons de la Vienne pour les campagnes A et B. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. V : Vienne, CA : campagne A, CB : campagne B. Mode positif (HESI+).

Au niveau du Clain, la Figure 57 montre des empreintes composées majoritairement de signaux dont la masse moléculaire est inférieure à 700 Da, correspondant donc, pour la plupart, à de petites molécules. Quelques signaux de plus hautes masses moléculaires sont cependant visibles, surtout pour des faibles temps de rétention et entre les temps de rétention 5 et 7 minutes, pouvant correspondre à des homologues. La plupart des signaux se concentrent entre les temps de rétention 3 et 9 min, couvrant une gamme de polarité de moyennement polaire à apolaire. Des signaux caractéristiques de molécules très polaires, non ou peu retenues par la colonne chromatographique, sont également détectés et concentrés pour les faibles temps de rétention (< 1,5 min). Les signaux ayant une aire très intense (> 10^8) sont très peu représentés dans les empreintes et la majorité des aires sont comprises entre 10^6 et 10^7. Ces signaux les plus intenses sont répartis sur l'ensemble de l'empreinte, ne semblant pas indiquer de gammes de masses moléculaires ou de polarités privilégiées. Au niveau du Clain, les trois campagnes partagent des points communs sur le type de signaux présents, cependant les empreintes ne sont pas tout à fait analogues montrant des variations au sein de la matrice de l'échantillon en fonction de la campagne. L'analyse UHPLC-HRMS des échantillons de la campagne B se montre ainsi plus chargée en signaux mais avec légèrement moins de signaux de hautes masses.

La Figure 58 des empreintes des échantillons de la Vienne montre une répartition des signaux, donc de leurs propriétés (taille et polarité), proche de celle du Clain mais avec un nombre de signaux qui semble plus important pour la campagne A. Concernant les aires des signaux, la majorité est comprise entre 10^6 et 10^8 mais on observe cependant une plus grande proportion de signaux de faibles intensités (< 10^6) et de plus haute intensité (> 10^8), ce qui différencie les empreintes de la Vienne par rapport à celle du Clain. Sur l'échantillon de la Vienne, prélevé lors de la campagne A, est observé un enchainement de signaux entre 5 et 7 min allant jusqu'à des masses de plus de 1000 Da. Cet enchainement est également visible sur les échantillons du Clain. Il est en revanche absent de l'échantillon de la Vienne de la campagne B.

La Figure 59 permet de confronter le nombre de signaux détectés en amont du Clain et dans la Vienne aux valeurs de carbone organique dissous (COD) mesurées.

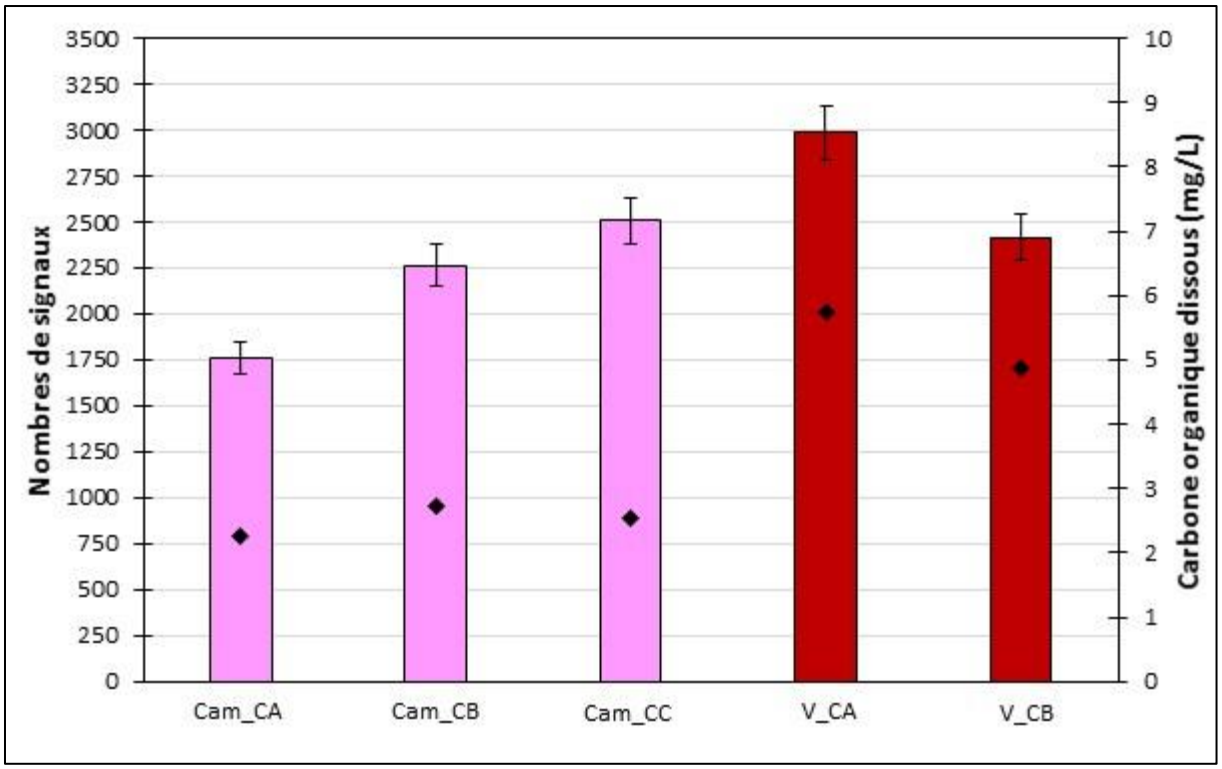


Figure 59 : Graphique en barres représentant le nombre de signaux détectés en fonction de l'échantillon considéré. Le losange noir (ordonnée de droite) correspond à la mesure du carbone organique dissous (COD). Cam (couleur rose) : Clain amont, V (couleur rouge) : Vienne, CA : campagne A, CB : campagne B, CC : campagne C.

Cette figure montre un nombre de signaux plus élevé dans l'échantillon de la Vienne par rapport à celui du Clain pour la campagne A. L'écart est en revanche peu significatif entre les échantillons de la campagne B. Par ailleurs, des mesures de conductivité et de COD ont montré des valeurs plus faibles dans la Vienne pour la conductimétrie (moyennes de 311 et de 573 µS/cm, respectivement pour les échantillons prélevés dans la Vienne et dans le Clain) mais plus grandes pour le COD (5,3 mg/L contre 2,5 mg/L pour le Clain). Il ne semble pas y avoir de relation entre les valeurs de COD et le nombre de signaux détectés.

En conclusion, les empreintes des échantillons de ces deux rivières présentent à la fois des caractéristiques proches, avec une majorité de signaux de faibles masses moyennement ou peu intenses, mais aussi des différences au niveau du nombre de signaux détectés plus important dans le cas de la campagne A de la Vienne, mais sans lien évident avec les valeurs de COD.

II. Variabilité spatiale d'une eau de surface en zone anthropisée

Dans cette partie, les quatre types d'échantillons : sortie de la STEP, panache du rejet dans le Clain, amont à ce rejet et aval immédiat, ont été comparés afin d'étudier la variabilité spatiale du Clain à proximité du rejet des eaux traitées de la STEP de Poitiers.

Le panache constitue le mélange entre les eaux du Clain et les eaux traitées par la STEP de Poitiers. A ce niveau, le rejet des eaux traitées forme en effet un panache coloré et turbide visible en surface. Le point en aval immédiat, à environ 30 m du panache, est prélevé dans le but d'évaluer les changements immédiats le long de l'écoulement du Clain. L'aspect visuel de l'eau prélevée à ce point est similaire à celui de l'eau prélevée à deux kilomètres en amont du rejet.

II.1. Comparaison globale du nombre de signaux

Le nombre de signaux a tout d'abord été déterminé pour l'ensemble des échantillons prélevés dans le Clain pour apprécier la variabilité spatiale d'une eau de surface en zone anthropisée. À titre de comparaison, ce nombre de signaux, et les valeurs de carbone organique dissous mesurées, sont représentés dans la Figure 60. Dans cette figure, à titre de comparaison les résultats obtenus ont été comparés à ceux de la sortie de la STEP.

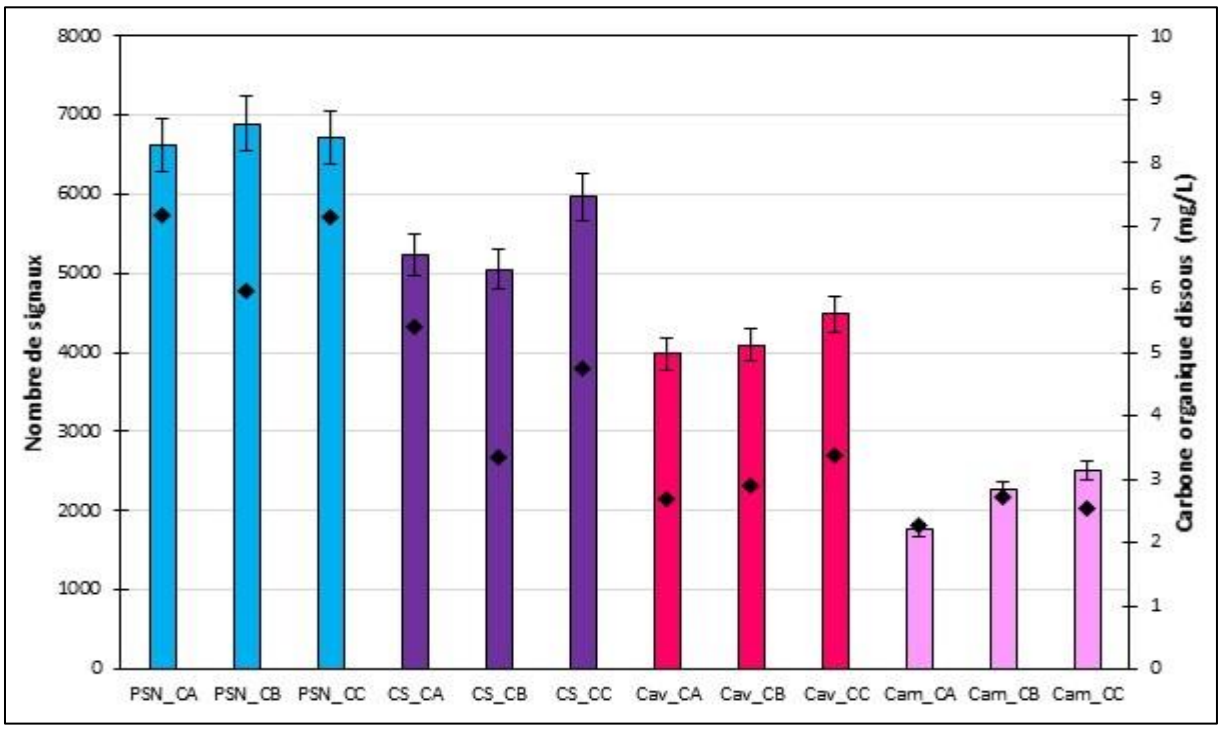


Figure 60 : Graphique en barres représentant le nombre de signaux en fonction de l'échantillon considéré. Le losange noir (ordonnée de droite) correspond à la mesure du carbone organique dissous (COD). PSN (couleur bleue) : Poitiers sortie Nord, Cam (couleur rose) : Clain amont, CS (couleur mauve) : Clain panache, Cav (couleur rouge) : Clain aval, CA : campagne A, CB : campagne B, CC : campagne C.

Cette figure montre une diminution du nombre de signaux des échantillons des sorties de STEP jusqu'aux échantillons de l'aval. Une diminution selon la même tendance est observée pour la teneur en COD des eaux prélevées. Un nombre de signaux significativement plus élevé est ainsi observé en aval par rapport à l'amont (augmentation moyenne de 95 ± 27%). Par ailleurs, une réduction de l'ordre de 20% en moyenne est observée entre les échantillons prélevés dans le panache et l'aval (-23 ± 3%), de même qu'entre les sorties de la STEP et le panache (-20 ± 8%). Ainsi, ces résultats montrent que :

- le rejet des eaux usées traitées apportent de nouveaux signaux dans le Clain,
- le nombre de signaux diminue entre la sortie de la STEP et l'aval,
- le nombre de signaux en aval immédiat reste bien supérieur à celui constaté en amont (près de deux fois plus de signaux sont observés entre l'aval et l'amont).

Entre les campagnes, la variation du nombre de signaux est de ± 2% pour la sortie de STEP, ± 9% pour le panache, ± 6% pour l'aval et de ± 18% pour l'amont. Une « reproductibilité » plus importante est donc visible pour les échantillons provenant de la sortie de la STEP. Ceci peut s'expliquer par la stabilité du traitement des eaux tandis que les échantillons du Clain sont plus exposés aux variations (de pluviométrie et de débit notamment). L'amont correspond au point de prélèvement présentant la plus grande variabilité.

II.2. Comparaison des échantillons par les statistiques descriptives

Les statistiques descriptives (quartiles, moyenne, médiane, maximum et minimum) ont été calculées pour chaque échantillon selon les trois paramètres décrivant les signaux : temps de rétention, masse moléculaire et valeur de l'aire du signal. La Figure 61 résume ces informations sous la forme de boîtes à moustaches.

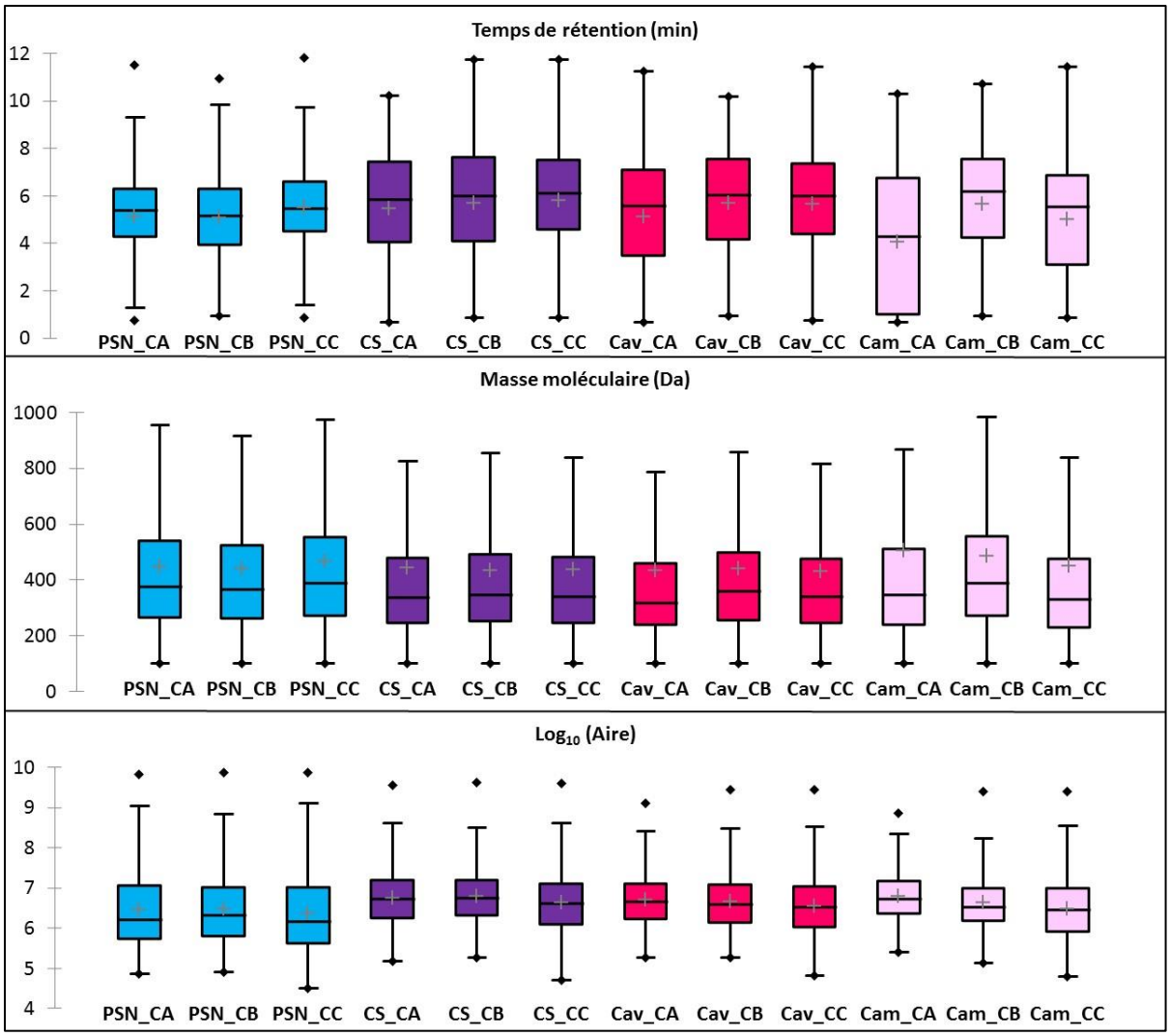


Figure 61 : Box plots ou boîtes à moustaches résumant les échantillons par leurs statistiques descriptives. La croix rouge, les limites inférieures et supérieures de la boîte et les points noirs symbolisent respectivement la moyenne, les premiers et troisièmes quartiles, les valeurs maximales et minimales. Cam : Clain amont, CS : Clain panache, Cav : Clain aval, PSN : Poitiers sortie Nord, CA : campagne A, CB : campagne B et CC : campagne C.

La dispersion des temps de rétention augmente dans les échantillons du Clain par rapport aux échantillons de la sortie de STEP. Pour chaque point de prélèvement, les médianes de l'ensemble des campagnes sont proches, (excepté pour l'échantillon amont de la campagne A). L'impact de la STEP sur le Clain est peu visible au niveau de ce paramètre.

Concernant les masses moléculaires, les médianes sont proches pour l'ensemble des points de prélèvement. La position des box plot des échantillons du Clain est cependant légèrement plus basse que celles des sorties de STEP. La dispersion des masses moléculaires couvre donc des gammes de valeurs légèrement plus faibles dans ces échantillons. Entre les échantillons du Clain eux-mêmes, les dispersions sont équivalentes. L'impact de la STEP n'est donc là encore pas visible pour ce paramètre.

Enfin, pour ce qui concerne les aires des signaux, les valeurs maximales sont les plus élevées pour les sorties de STEP suivies par le panache et l'aval. Les sorties de STEP apportent en effet des signaux dont les aires sont plus élevées que dans les échantillons amont.

Les résultats de cette figure montrent un impact limité du rejet des eaux traitées sur les dispersions des paramètres temps de rétention, masses moléculaires et valeurs d'aires. L'impact le plus visible concerne les valeurs d'aires, avec les valeurs maximales qui sont plus élevées dans les échantillons du panache avant de diminuer plus à l'aval.

II.3. Caractérisation d'une zone de rejet de STEP en rivière : comparaison entre le panache du rejet et son aval immédiat

Si les rejets de STEP sont conformes aux normes établies, des interrogations subsistent quant à l'impact des eaux traitées sur le milieu récepteur. Dans le but d'étudier cet impact des eaux traitées de la STEP de Poitiers dans le Clain, la caractérisation de la zone de rejet de la STEP a tout d'abord été réalisée à partir des prélèvements effectués au niveau :

- du panache du rejet des eaux traitées dans le Clain,
- et de l'aval immédiat près de la rive.

L'étude de ces deux points de prélèvements (panache et aval) a porté sur :

- la comparaison des empreintes,
- les variations des valeurs d'aires des signaux.

Les empreintes des échantillons du panache, de l'aval et les empreintes résultant du calcul de la variation de l'aire des signaux entre le panache et l'aval pour les trois campagnes sont présentées respectivement dans l'Annexe 9 et dans la Figure 62.

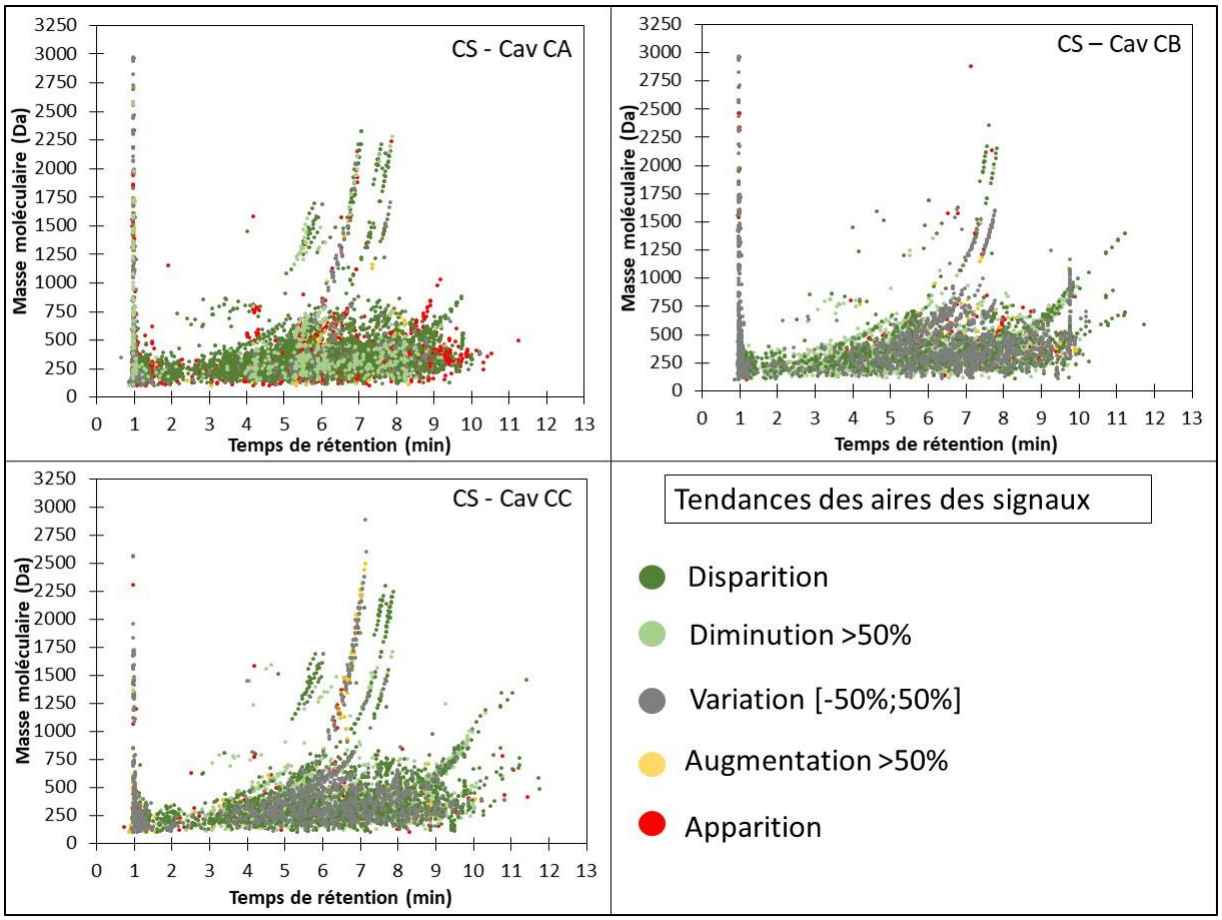


Figure 62 : Empreintes pour les campagnes A, B et C précisant pour les signaux détectés leur tendance associée correspondant au calcul de la variation de l'aire des signaux de l'aval par rapport au panache. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. CS : panache du Clain, Cav : aval du Clain, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI+).

En comparant directement l'ensemble des empreintes du panache et de l'aval (Annexe 9), on constate que les deux types d'échantillons montrent une gamme de polarité des signaux de moyennement polaire à plutôt apolaire et que la plupart des signaux correspondent à des petites et moyennes molécules. Ces signaux sont répartis sur l'ensemble des temps de rétention des empreintes et sur des masses moléculaires inférieures à 700 Da. Pour les deux types d'échantillons, le Tableau 22 présente la dispersion du nombre de signaux entre quelques intervalles de valeurs pour les trois descripteurs.

Tableau 22 : Répartition des signaux pour les échantillons du panache et de l'aval du Clain selon les 3 descripteurs : temps de rétention, masse moléculaire et valeur de l'aire des signaux.

Panache					Aval				
	Nombre total de signaux	Nombre de signaux par intervalle				Nombre total de signaux	Nombre de signaux par intervalle		
Temps de rétention (min)		<3	3-9	>9	**Temps de rétention (min)**		<3	3-9	>9
CS CA	5238	976	4105	157	**Cav CA**	3984	950	2911	123
CS CB	5045	858	3760	427	**Cav CB**	4091	670	3117	304
CS CC	5966	807	4754	405	**Cav CC**	4483	684	3547	252
Gamme de masse (Da)		<700		>700	**Gamme de masse (Da)**		<700		>700
CS CA	5238	4687		551	**Cav CA**	3984	3558		426
CS CB	5045	4527		518	**Cav CB**	4091	3655		436
CS CC	5966	5364		602	**Cav CC**	4483	4059		424
Gamme d'aire		$<10^6$	10^6-10^8	$>10^8$	**Gamme d'aire**		$<10^7$	10^6-10^8	$>10^8$
CS CA	5238	798	4178	262	**Cav CA**	3984	615	3212	157
CS CB	5045	488	4315	242	**Cav CB**	4091	679	3256	156
CS CC	5966	1256	4479	231	**Cav CC**	4483	1054	3288	141

Le nombre de signaux est plus de 20% plus faible dans les échantillons de l'aval du rejet que dans ceux du panache (4 200 ± 300 signaux à l'aval contre 5 400 ± 500 dans le panache). En revanche, en accord avec ce qui est observé dans l'Annexe 9, ce tableau montre que les empreintes du panache et de l'aval partagent de nombreux points communs sur les trois descripteurs :

- une majorité de signaux a des temps de rétention compris entre 3 et 9 minutes (73 à 80% des signaux),
- une grande majorité des signaux a une masse inférieure à 700 Da (environ 90% des signaux),
- les valeurs d'aires les plus représentées sont comprises entre 10^6 et 10^8 (pour 73 à 86% des signaux).

En ce qui concerne l'évolution des signaux des échantillons de l'aval par rapport au panache, la Figure 62 montre que les diminutions et disparitions (représentées en vert sur) se retrouvent sur l'ensemble de l'empreinte et ne concernent donc pas une catégorie de signaux particuliers. A côté de ces disparitions ou diminutions, on peut voir également de nombreux signaux variant peu entre les deux types d'échantillons, pour la plupart des signaux inférieurs à 700 Da.

Pour appuyer ces observations, la Figure 63 représente le nombre de signaux par catégories de tendance (disparition, diminution, stagnation, augmentation et apparition) pour les trois campagnes de prélèvement. Dans cette figure, le pourcentage du nombre de signaux présents dans l'échantillon

de l'aval est exprimé par rapport au nombre total de signaux de l'échantillon du panache pour chaque tendance et chaque campagne (la somme des pourcentages relatifs aux quatre tendances : disparition, diminution, variation faible et augmentation est donc de 100%).

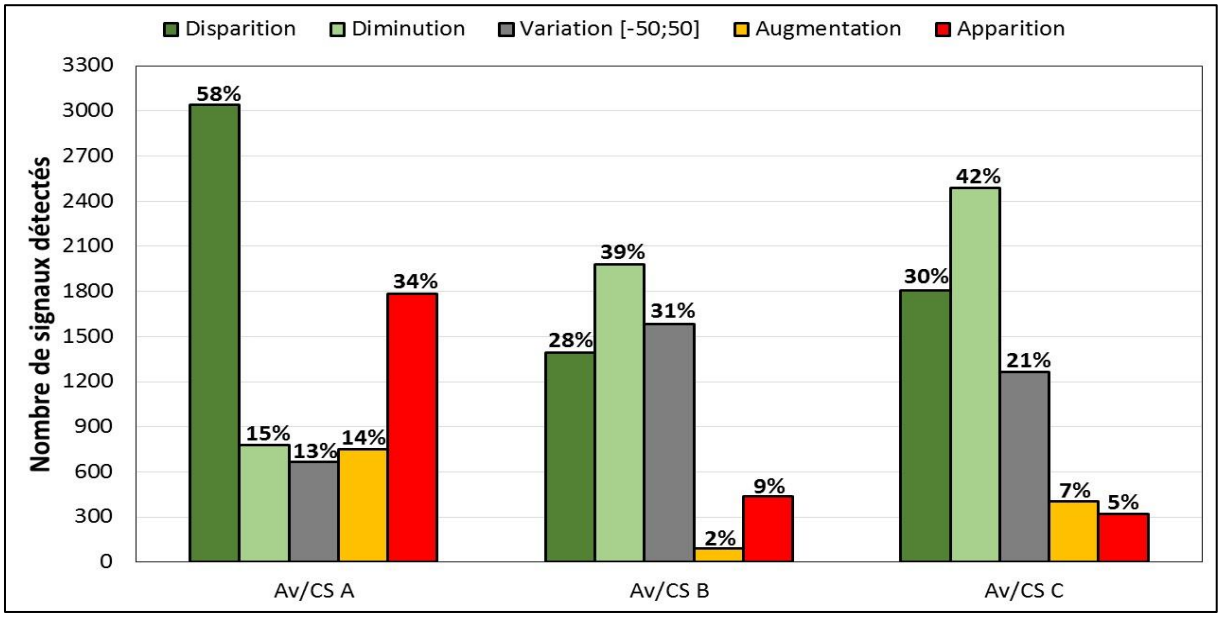


Figure 63 : Graphique en barres représentant le nombre de signaux de l'aval en fonction de leur tendance de variation par échantillon par rapport au panache. Les pourcentages inscrits représentent le nombre de signaux de la tendance sur le nombre total de signaux du panache. CS : Clain panache, Cav : Clain aval, CA : campagne A ; CB : campagne B, CC : campagne C.

Cette figure montre un écart entre la campagne A et les deux autres campagnes sur la répartition des signaux. En effet dans la campagne A, la disparition et l'apparition sont les deux tendances majoritaires tandis que dans les campagnes B et C, ce sont les tendances de disparition, de diminution et de stagnation (variation [-50 ; 50%]) qui sont majoritaires. La surreprésentation des signaux apparaissant et disparaissant lors de la campagne A concerne des signaux couvrant toute la gamme de temps de rétention. Il semble difficile de trouver une explication à cet écart par rapport aux autres échantillons. Toutefois, en plus de biais possibles pouvant être générés lors de la préparation des échantillons ou du traitement des données, ceci pourrait en partie s'expliquer par une dilution différente des rejets au niveau du Clain et/ou par une composition différente des échantillons (nature et/ou intensité des signaux) pouvant générer des effets matrices différents lors des analyses UHPLC-HRMS de cette campagne. Entre la campagne B et la campagne C, des différences sont également visibles mais les tendances majoritaires sont les mêmes, tout comme la proportion de signaux partageant des tendances similaires.

Pour les signaux dont l'aire diminue entre le panache et l'aval, les ratios entre les valeurs d'aires des deux types d'échantillons ont été calculés. Ces calculs montrent que pour la campagne B et C, entre 73% et 95% des ratios sont compris entre 2 et 5, semblant indiquer un effet de dilution assez homogène. Pour la campagne A en revanche, 40% des signaux qui diminuent ont une valeur d'aire

réduite entre 2 et 5 fois en aval et 60% ont une diminution supérieure à 5 fois. Sachant que cette campagne A compte plus de signaux disparaissant et aussi moins de signaux stagnants, l'hypothèse d'une dilution plus forte du rejet de la STEP dans le Clain durant cette campagne pourrait être ainsi confortée.

Le Tableau 23 présente le débit du Clain et les principales caractéristiques physico-chimiques (conductivité, COD, NT et absorbance UV) des eaux du panache et à l'aval du rejet lors des campagnes A, B et C. Dans ce tableau, le panache apparaît plus concentré en termes de conductivité et de charge organique lors de la campagne A alors que les valeurs de ces paramètres à l'aval sont sensiblement plus proches entre les trois campagnes. Ces éléments appuient les hypothèses avancées précédemment pour expliquer la singularité des résultats obtenus pour la campagne A.

Tableau 23 : Paramètres physico-chimiques de caractérisation des eaux panache du rejet et aval, pour les campagnes A, B et C.

Campagne	Type d'échantillon	Débit du Clain*	Conductivité (µS 25°C)	COD (mg/L)	NT (mg/L)	UV 254 nm (sur 1 cm)
Campagne A	Panache	8,18 m^3/s	991	5,41	11,66	0,152
	Aval		603	2,69	7,37	0,083
Campagne B	Panache	5,42 m^3/s	649	3,36	4,26	0,080
	Aval		681	2,91	4,98	0,074
Campagne C	Panache	7,42 m^3/s	712	4,74	3,44	0,123
	Aval		611	3,37	5,26	0,084

* Débit mesuré en amont du Clain, station du Pont St Cyprien, L2341620

Quelle que soit la campagne, les signaux stagnants pourraient correspondre à des signaux de la matrice Clain ainsi qu'à des signaux issus des eaux traitées encore modérément diluées dans les eaux de la rivière, ou encore à des signaux impactés différemment par la matrice lors des analyses UHPLC-HRMS (l'effet matrice étant un effet molécule dépendant).

Ces résultats illustrent donc une certaine capacité de la rivière, immédiatement en aval du rejet des eaux épurées à atténuer l'impact de la STEP, notamment par un effet de dilution des concentrations exprimé par la diminution des valeurs d'aires jusqu'à la disparition de signaux. Toutefois, si cette diminution des valeurs d'aires témoigne d'une diminution des concentrations, elle ne rend cependant pas compte de l'impact de ces concentrations sur le plan toxicologique qui reste à évaluer.

II.4. Evaluation de la pression des eaux traitées de la STEP au point de rejet : comparaison des sorties de la STEP et du panache

Afin d'évaluer la pression des eaux traitées de la STEP au point de rejet du Clain, l'évolution des aires des signaux a été étudiée entre les échantillons de sortie de la STEP et ceux du panache dans le rejet.

Les Figures 64 et 65 illustrent les variations des aires des signaux dans le panache du Clain par rapport aux sorties de STEP, pour les trois campagnes d'échantillonnage.

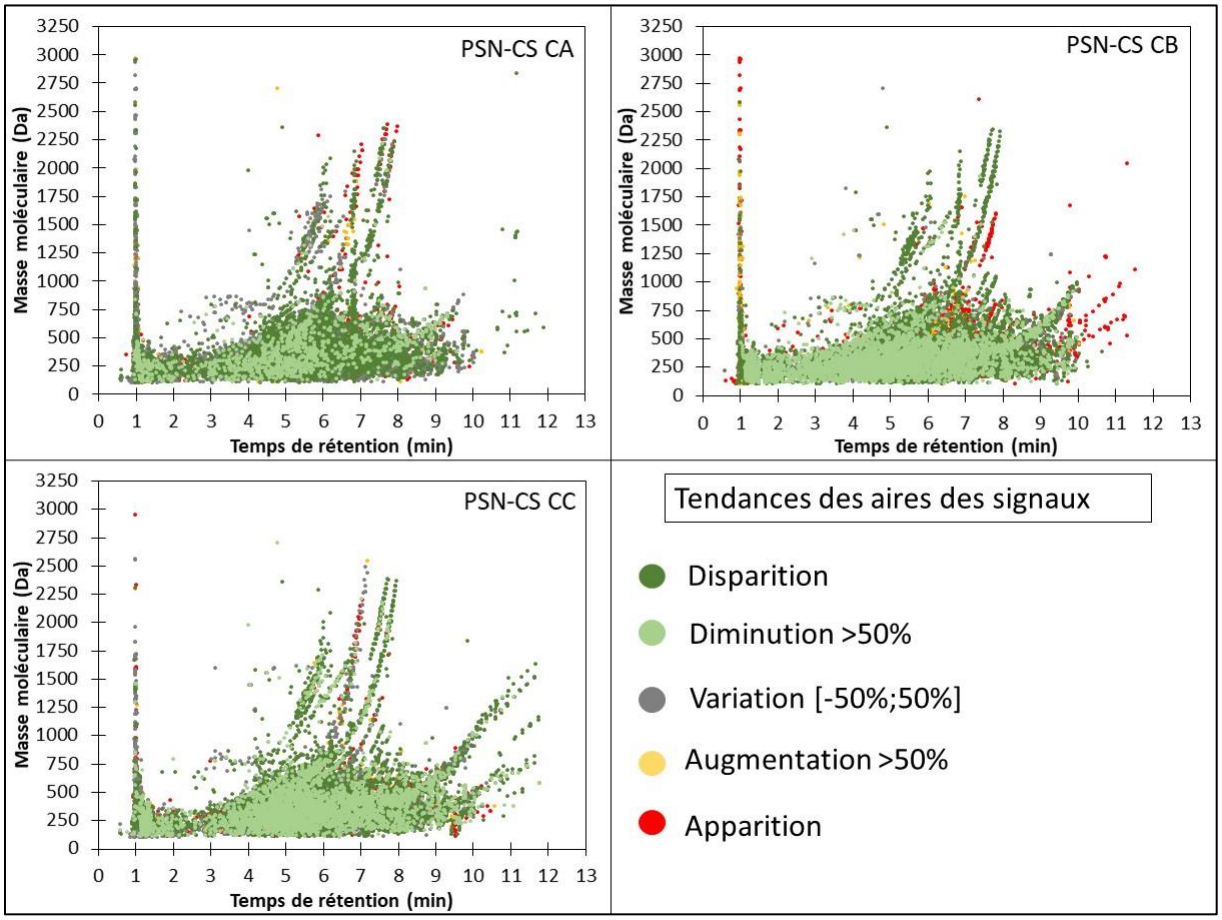


Figure 64 : Empreintes précisant pour les signaux détectés leur tendance associée correspondant au calcul de la variation de l'aire des signaux de la sortie Nord de la STEP de Poitiers par rapport au panache du rejet dans le Clain. PSN : Poitiers sortie Nord, C CS : panache Clain, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI+).

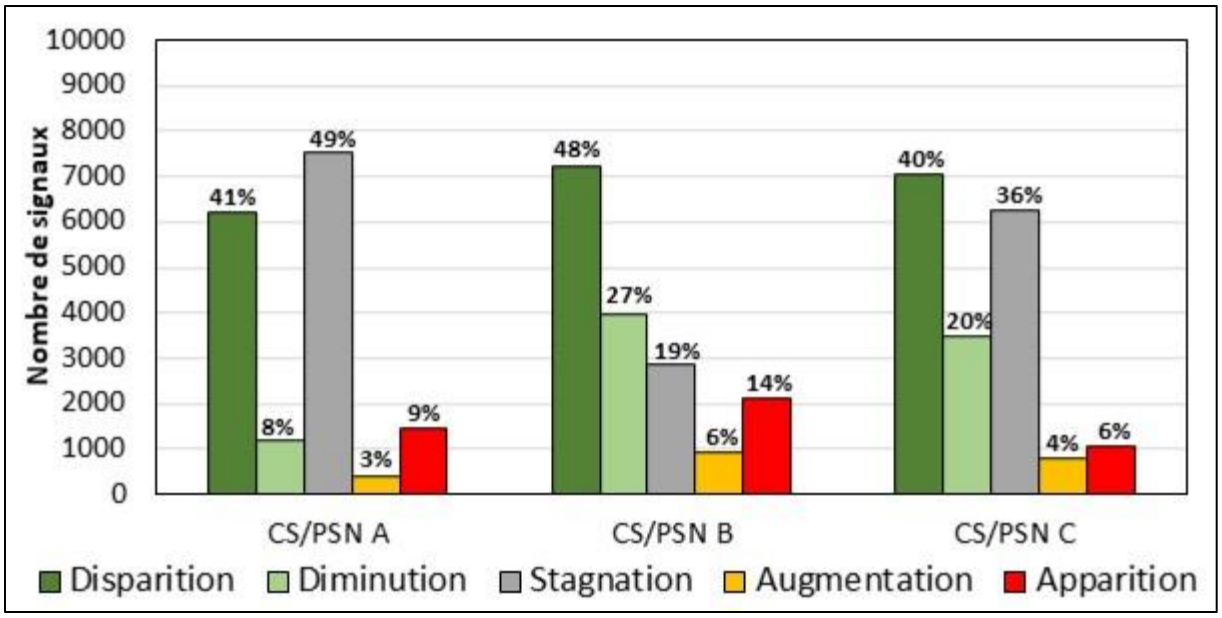


Figure 65 : Graphique en barres représentant le nombre de signaux du panache en fonction de leur tendance de variation par échantillon par rapport à la sortie de la STEP. Les pourcentages inscrits représentent le nombre de signaux de la tendance sur le nombre total de signaux de la sortie de la STEP. CS : Clain panache, PSN : Sortie Nord STEP Poitiers, CA : campagne A ; CB : campagne B, CC : campagne C.

La représentation des tendances entre les sorties de STEP et le panache du rejet montre que les disparitions, diminutions, et stagnations des signaux sont les tendances majoritaires entre les deux types d'échantillons (avec des pourcentages moyens respectivement de 43%, 18% et 35% par rapport au nombre total de signaux en sortie de STEP).

L'exploration des empreintes montre que les signaux liés aux sorties de STEP (après 9 minutes et de masse supérieure à 500 Da) diminuent tandis que des signaux plus liés à l'amont du Clain pourraient apparaître au mélange entre les eaux du Clain et le rejet des eaux traitées (panache). Le calcul des ratios a démontré que plus de 90% des signaux ont une aire 2 à 5 fois inférieure dans le panache pour les trois campagnes. A titre de comparaison, la dilution du rejet de la STEP dans le Clain estimée au regard des débits varie au cours de l'année entre 1/10 et 1/500. Les différences observées entre ces taux de dilution et l'impact sur les aires proviennent en grande partie du caractère partiel du mélange des eaux du rejet et du Clain au niveau du panache. Des effets matrices différents entre la sortie de la STEP et le panache pourraient également y contribuer.

Au vu de ces résultats, et en accord avec l'aspect visuel des eaux au niveau du rejet, le panache constitue une zone de mélange partiel localisé, mélange qui se poursuit au-delà du point de prélèvement aval.

II.5. Impact sur le milieu aquatique : Comparaison Aval/Amont

Dans cette partie, une comparaison entre les échantillons de l'amont et ceux de l'aval a été réalisée pour évaluer l'impact des rejets de STEP sur le milieu aquatique. Les variations des aires des signaux, entre ces deux types d'échantillons (aval/amont), ont été calculées et sont représentées sur les empreintes combinées pour chaque campagne dans la Figure 66.

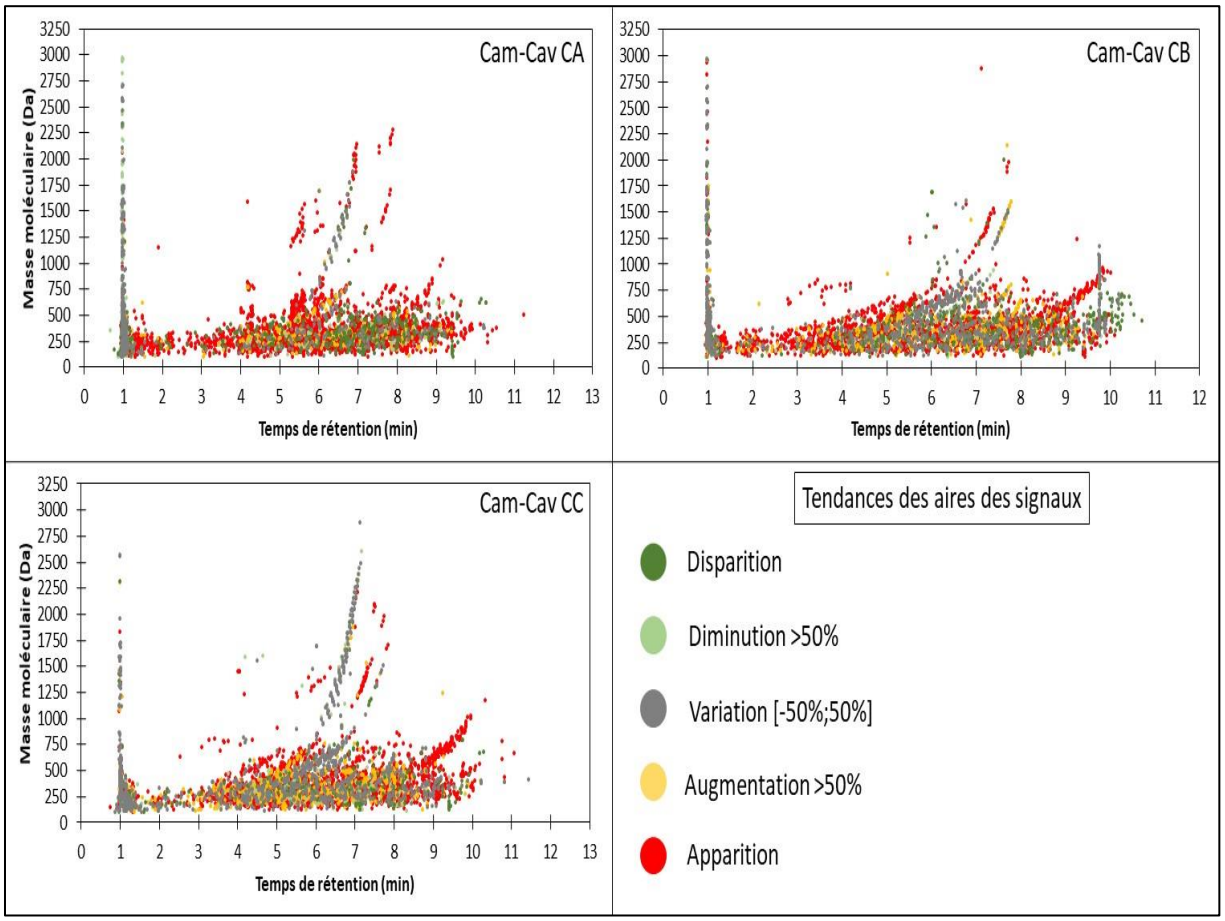


Figure 66 : Empreintes précisant pour les signaux détectés leur tendance de variation associée correspondant au calcul de la variation de l'aire des signaux de l'aval par rapport à l'amont. Cam : Clain amont, Cav : Clain aval, CA : campagne A, CB : campagne B, CC : campagne C. Mode positif (HESI+).

Cette figure met en évidence l'apparition d'un grand nombre de signaux entre l'amont et l'aval et ceci pour chaque campagne. Ces signaux apparaissant correspondent vraisemblablement en grande partie aux signaux apparus sur l'empreinte du panache *via* les rejets de la STEP. Ces nouveaux signaux ont des polarités et des masses variées. Ils peuvent ainsi s'apparenter à de nombreux types de molécules. D'autres signaux voient leur aire augmenter entre l'amont et l'aval : les molécules correspondant à ces signaux sont donc déjà présentes en amont, mais des apports complémentaires sont possibles, que ce soit par le rejet des eaux traitées de la STEP ou par d'autres contributions plus diffuses (comme le ruissellement par exemple). Enfin, on note aussi la présence :

- de signaux « stagnants » dont la variation est inférieure à ± 50% : c'est le cas par exemple des signaux formant un enchainement (motif) de temps de rétention compris entre 5,5 et 7 min avec des masses supérieures à 700 Da. Ces signaux « stagnants » sont présents dans l'ensemble des empreintes du Clain et représentent une partie des constituants qui restent stables sur l'ensemble de la distance entre les points de prélèvements.
- et de signaux diminuant ou disparaissant.

Afin d'estimer la proportion des signaux de chaque tendance de variation (aval/amont) dans les empreintes, le nombre de signaux de chaque tendance a été calculé pour chaque campagne et est présenté dans la Figure 67.

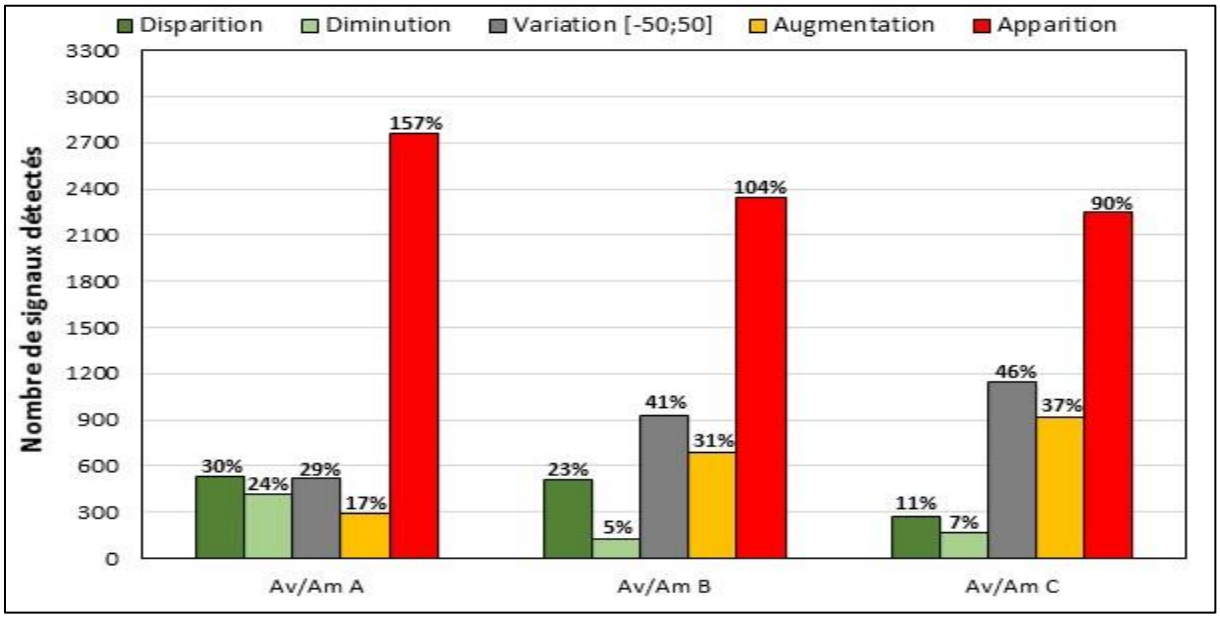


Figure 67 : Graphique en barres représentant le nombre de signaux de l'aval en fonction de leur tendance de variation par échantillon par rapport à l'amont. Les pourcentages inscrits représentent le nombre de signaux de la tendance sur le nombre de signaux total de l'amont. Cam : Clain amont, Cav : Clain aval, CA : campagne A ; CB : campagne B, CC : campagne C.

Les empreintes aval sont ainsi composées pour plus de la moitié de nouveaux signaux qui étaient absents en amont. L'aire de près de 30% des signaux déjà présents en amont augmente en aval. L'impact du rejet de la STEP sur le Clain est donc significatif à l'aval immédiat du rejet. Ces résultats sont en accord avec les travaux de Ruff *et al* (2015) qui ont montré que les polluants détectés dans le Rhin proviennent majoritairement des STEP. Les signaux variant peu (< 50% de variation) entre les deux types d'échantillons représentent entre 29% et 46% du nombre total de signaux présents en amont (entre 7% et 24% du total de signaux en aval). Ces signaux pourraient correspondre pour partie à de la matière organique naturelle présente dans les eaux du Clain. Comme pour les résultats précédents, les tendances d'évolution des signaux des échantillons prélevés lors de la campagne A apparaissent sensiblement différents de ceux des échantillons prélevés lors des deux autres campagnes, avec plus de signaux apparaissant et moins de signaux augmentant ou stagnant.

L'impact des rejets de la STEP entre l'amont et l'aval se traduit donc par une augmentation du nombre de signaux avec un apport de nouveaux signaux localement au niveau du panache du rejet, sans modification globale des propriétés moyennes des signaux présents dans les empreintes de ces eaux (temps de rétention, masse moléculaire, intensité). Cet impact s'atténue fortement immédiatement en aval, sans toutefois correspondre à l'état observé à l'amont.

II.6. Corrélation entre signaux et nature des échantillons par analyse multivariée

L'ensemble des empreintes des échantillons prélevés dans le Clain (amont, panache et aval) a été soumis à une analyse en composante principale avec normalisation (n-1) préalable. La Figure 68 présente le biplot, montrant à la fois les signaux et les échantillons. Les signaux apparaissent avec un code couleur correspondant à la variation entre l'empreinte moyenne des échantillons de l'aval par rapport à celle des échantillons de l'amont. L'objectif est de déterminer si un type d'échantillon donné peut être associé sur le biplot à l'absence, à la présence ou à une variation des valeurs d'aire des signaux entre l'amont et l'aval.

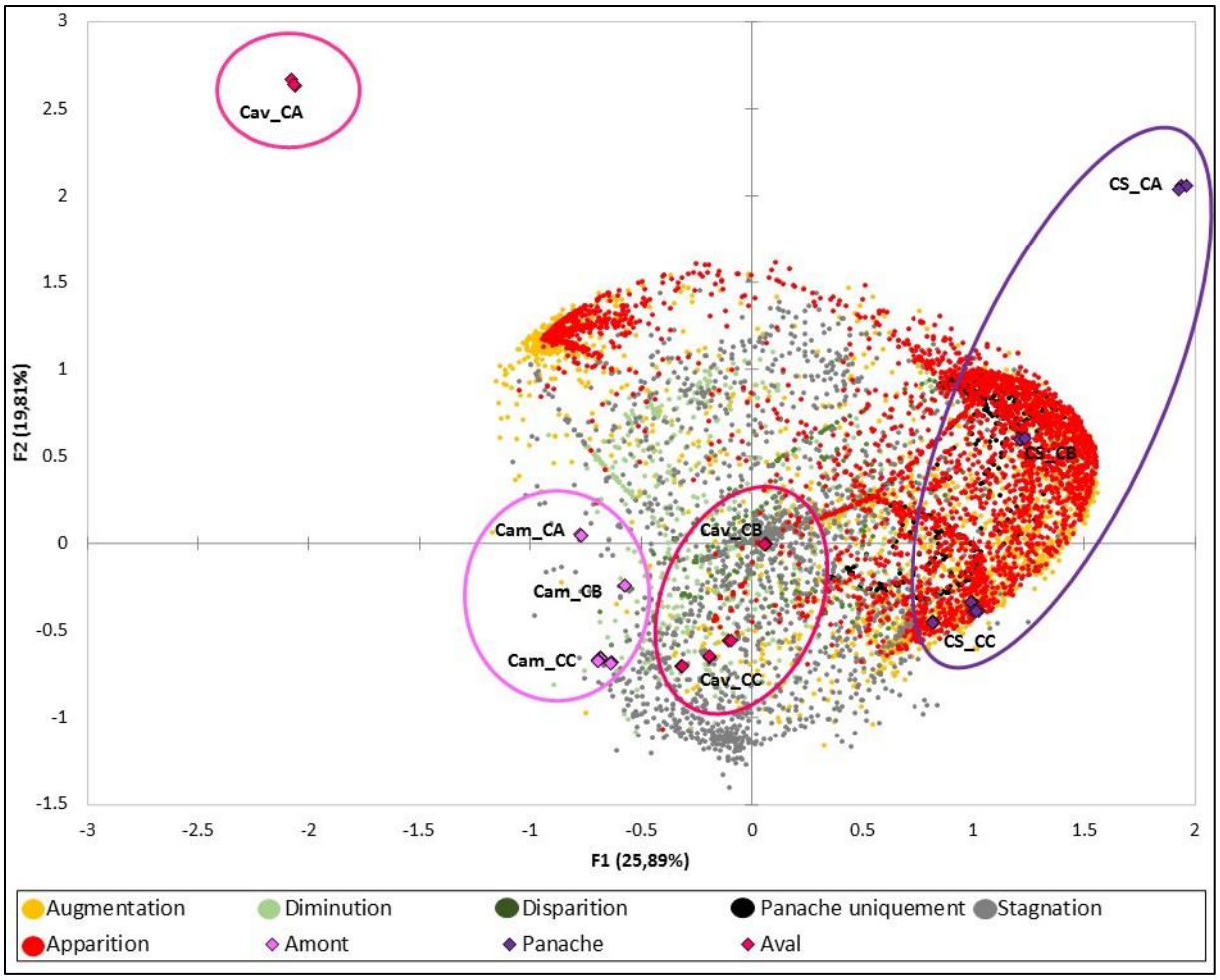


Figure 68 : Biplot des échantillons et des signaux détectés selon les facteurs F1 et F2 et les tendances de variation associées calculées entre l'amont et l'aval pour les échantillons du Clain. Cam : Clain amont, Cav : Clain aval, CS : Clain STEP (panache), CA : campagne A, CB : campagne B, CC : campagne C.

Concernant les échantillons, cette figure montre une séparation par l'axe F1 des échantillons de l'amont du Clain et de ceux du panache. Entre ces deux types d'échantillons se trouvent ceux issus de l'aval. La composante F1 fait donc la séparation entre les types d'échantillons. On remarque aussi que les échantillons de l'aval sont globalement plus proches des échantillons de l'amont que de ceux du panache. Les échantillons d'amont sont proches les uns des autres. Ceux du panache semblent plus dispersés. Le panache est le mélange entre les eaux du Clain et le rejet des eaux traitées de la STEP. Davantage de variabilité lors du prélèvement peut donc être observée. De plus, il est possible que selon la période de l'année, la qualité de ce mélange soit fluctuante, dépendant de facteurs tels que la température, le débit du Clain ou encore la différence de conductivité des eaux. Pour la campagne A, l'échantillon aval, et dans une moindre mesure l'échantillon du panache, sont plus éloignés des deux autres échantillons, en cohérence avec la singularité observée de la campagne A vis-à-vis des deux autres campagnes. L'axe F1 sépare donc les échantillons selon leur nombre de signaux et finalement selon leur « degré de pollution » tandis que l'axe F2 fait la séparation entre les campagnes et donc les fluctuations entre les campagnes.

Concernant les signaux, le code couleur montre une corrélation préférentielle des signaux apparaissant et augmentant entre l'amont et l'aval avec les échantillons du panache. Ainsi, une partie de ces signaux semble plutôt liée aux échantillons du panache. Ceci appuie donc l'hypothèse que les signaux qui apparaissent en aval proviennent principalement du panache *via* le rejet des eaux de STEP. Enfin, les signaux disparaissant, diminuant ou stagnant sont répartis sur l'ensemble du biplot. Ils ne sont donc pas spécifiquement reliés à l'effet du panache, et donc de la STEP.

III. Recherche de signaux pertinents entre l'amont et l'aval

L'étude de la variabilité spatiale a montré des changements dans le Clain dus aux rejets de la STEP et visibles à partir de l'évolution des signaux entre l'amont et l'aval. Ces changements de tendances correspondent pour partie à des variations de concentrations de molécules présentes dans l'eau de surface et peuvent être utilisés pour la recherche de signaux pertinents traceurs d'apports anthropiques. Les travaux de Carpenter *et al.* (2018 ; 2019) et de Beckers *et al.* (2020) ont montré des changements d'intensité de signaux et des clusters liés à des évènements particuliers (présence d'une STEP ou de pollutions diffuses) puis ils s'en sont servis pour tenter de mettre en évidence des micropolluants pertinents liés à une source de diffusion donnée. Dans le cadre des présents travaux, il a été décidé de rechercher des signaux pertinents par la réalisation d'analyses différentielles.

III.1. Sélection des signaux par analyse différentielle

Similairement au travail réalisé entre les entrées et les sorties de STEP (voir le chapitre 4, « Caractérisation des eaux usées brutes et traitées »), une recherche de signaux pertinents a été effectuée à partir de la comparaison entre les empreintes des échantillons de l'amont et de l'aval afin de sélectionner des marqueurs du rejet de la STEP dans le Clain. Pour cela, une analyse différentielle utilisant le Fold Change a été effectuée, pour chaque campagne, entre l'amont et l'aval. Dans le cas de la campagne C, l'analyse différentielle a été faite sur l'ensemble des échantillons prélevés lors de cette campagne (trois échantillons pour chaque point de prélèvement). Les valeurs des signaux absents dans un échantillon ont été remplacées par les valeurs du blanc afin de les rendre exploitables sur les volcano-plots pour effectuer la comparaison entre tous les signaux détectés. La Figure 69 présente les résultats des volcano-plots obtenus dans ces conditions.

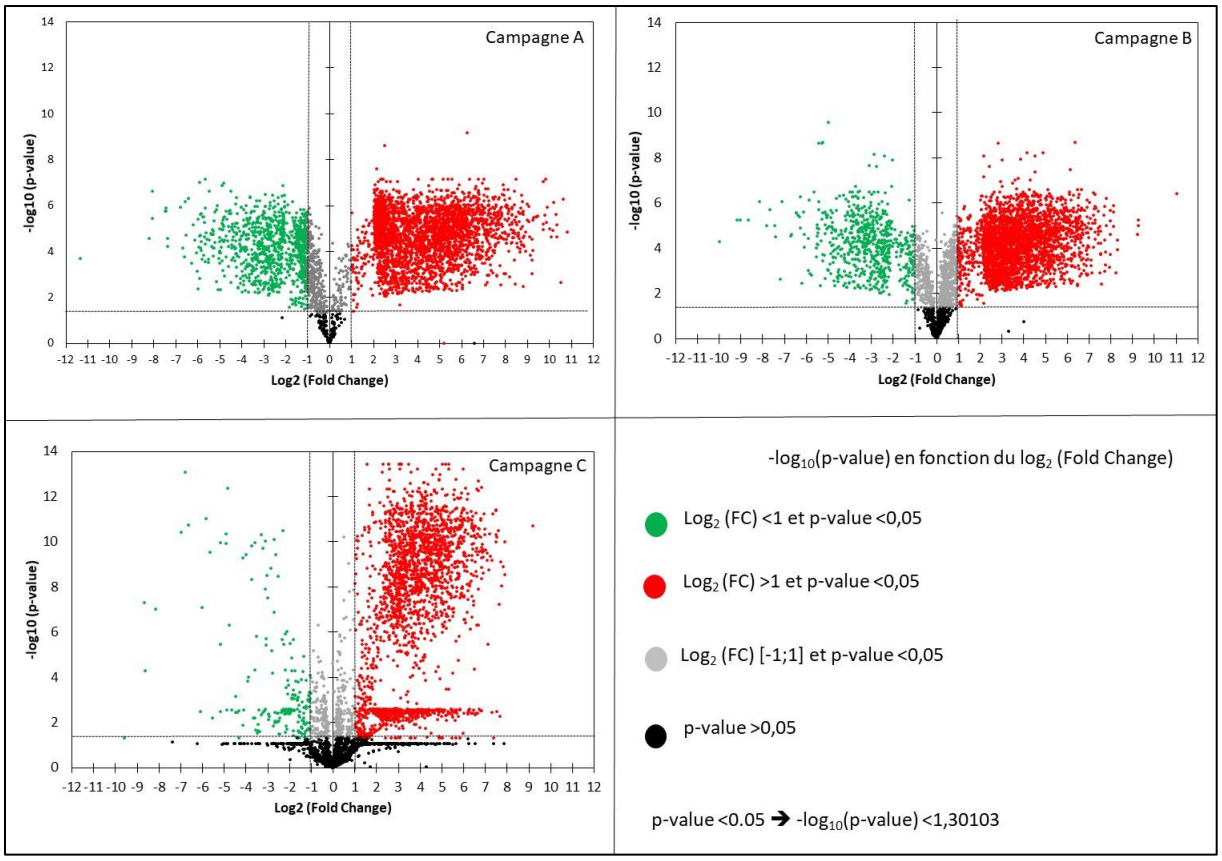


Figure 69 : Graphiques de type volcano-plot des analyses différentielles entre chaque échantillon de l'aval par rapport à l'amont, par campagne. Analyses différentielles réalisées par XLSTAT.

Ces résultats montrent, comme dans la Figure 67, le nombre important de signaux dont la valeur d'aire augmente en aval par rapport à l'amont (en rouge) en comparaison avec les signaux diminuant (en vert). Les signaux variant peu sont considérés ici comme de moindre intérêt car il a été estimé que globalement ils font partie de la « matrice Clain » et constituent ainsi le bruit de fond de l'analyse.

Ainsi, pour la recherche de signaux pertinents, les signaux dont le Fold Change est supérieur à 1 ont été sélectionnés lorsque leur p-value est inférieure à 0,05. De cette façon les signaux apparaissant et augmentant par rapport à l'amont sont conservés. De façon à ne prendre en compte que les signaux impactant régulièrement le milieu, comme dans l'approche du chapitre 4, seuls les signaux présents dans les trois campagnes ont été conservés. Ainsi, 305 signaux ont été retenus. La Figure 70 montre l'empreinte des signaux ainsi conservés.

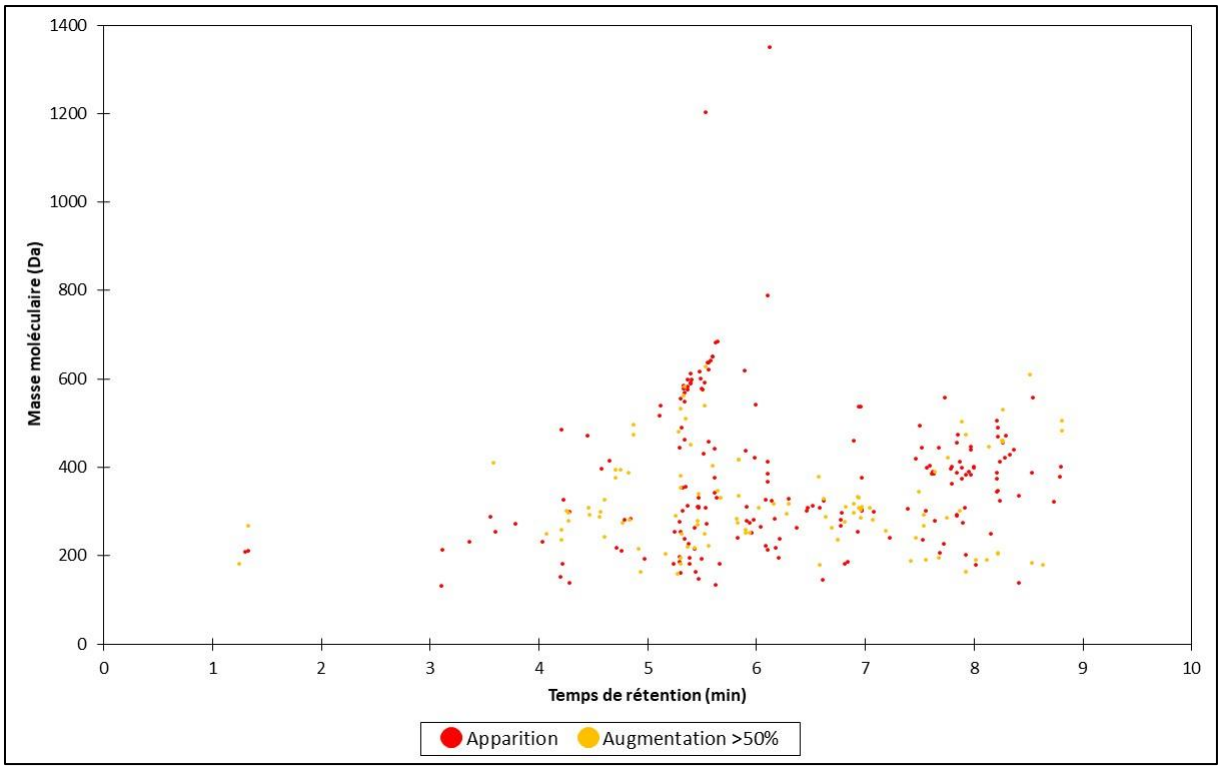


Figure 70 : Empreinte de l'ensemble des signaux considérés comme pertinents, sélectionnés par analyse différentielle entre les échantillons amont et aval selon leur tendance moyenne.

L'empreinte des signaux pertinents sélectionnés montre que la majorité d'entre eux (> 85% des signaux) ont une masse moléculaire inférieure à 500 Da et un temps de rétention compris entre 3 et 9 minutes. Ces signaux pertinents couvrent donc une gamme de polarité allant de signaux peu polaires à plutôt apolaires

III.2. Approches de caractérisation

III.2.1. Attribution de formules brutes par Compound Discoverer

Les formules brutes des composés correspondant aux signaux pertinents ainsi sélectionnés ont été prédites par Compound Discoverer en considérant une tolérance de déviation en masse de ± 5 ppm ou de ± 3 mDa. Plus la masse augmente plus le risque d'une erreur d'attribution augmente, ainsi les signaux de masses moléculaires supérieures à 800 Da ont été exclus lors de cette approche afin de ne pas augmenter significativement le taux de mauvaise attribution. Une formule brute a ainsi été attribuée pour 273 signaux (sur les 305) et les ratios H/C, O/C et N/C ont été calculés. La Figure 71 montre le ratio H/C en fonction des ratios O/C et N/C avec un code couleur attribué selon le type de formule brute parmi les types suivants : CHO, CHN, CHNO, CHNOP, CHNOPS, CHNOS, CHOP, CHOPS, présence d'un halogène et autre.

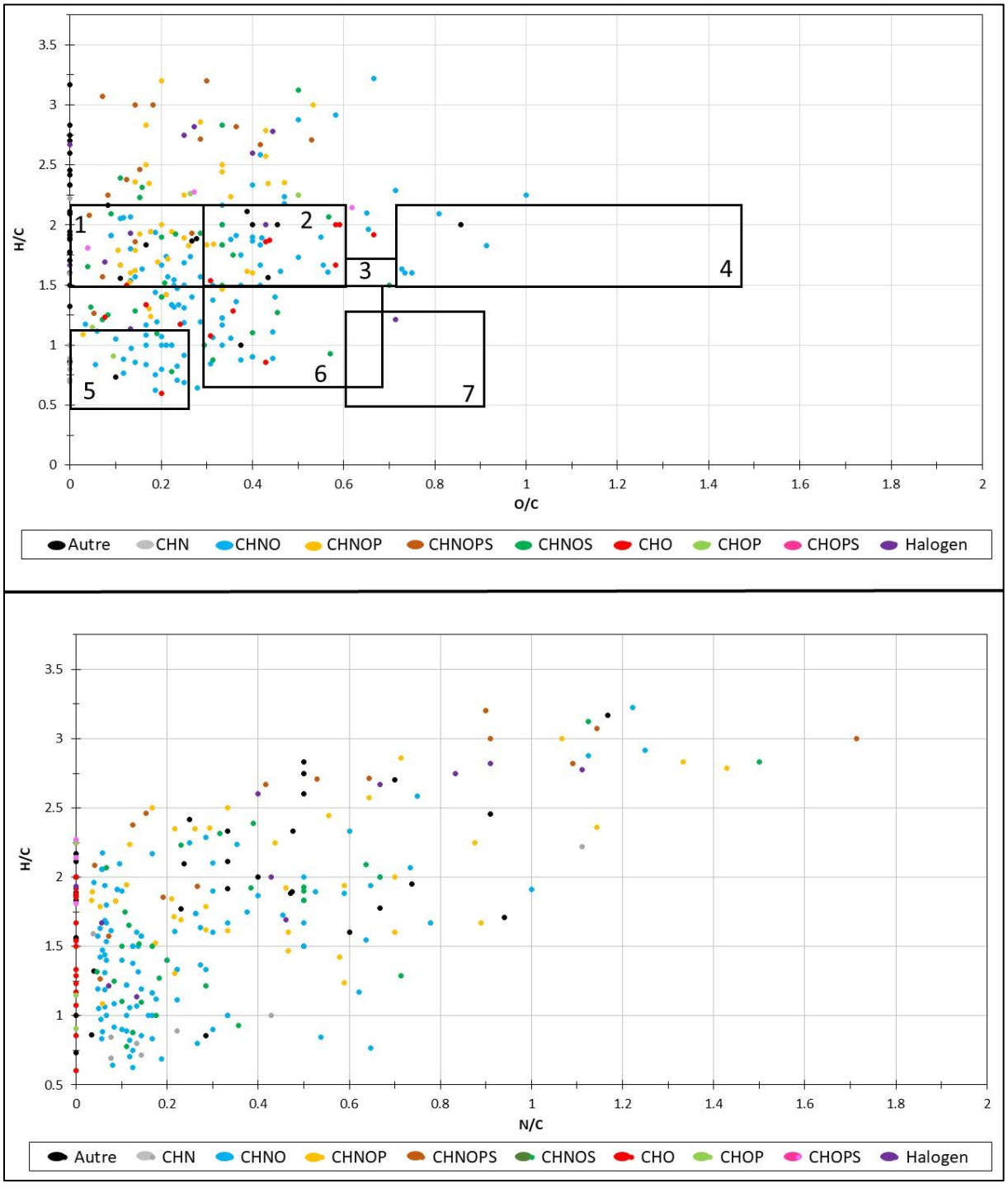


Figure 71 : Graphiques du ratio H/C en fonction des ratios O/C et N/C des signaux sélectionnés par analyse différentielle. 1 : lipides, 2 : protéines, 3 : sucres aminés, 4 : carbohydrates, 5 : hydrocarbures condensés, 6 : lignines, 7 : tannins.

De même, pour l'ensemble des signaux pour lesquels une formule brute a été proposée, le Tableau 24 résume le nombre de signaux par type de formule brute, par catégorie de composés et selon le nombre d'atomes de carbones.

Tableau 24 : Nombre de signaux en fonction de la composition, de la catégorie et du nombre d'atomes de carbones parmi les 273 signaux pour lesquels une formule brute a pu être associée.

Composition	Nombre de signaux	Catégorie	Nombre de signaux	Nombre d'atomes de C	Nombre de signaux
CHO	15	Lipides (1)	69	1 à 5	3
CHOP	4	Protéines (2)	41	5 à 10	61
CHNO	111	Sucres aminées (3)	1	10 à 15	86
CHNOP	39	Carbohydrates (4)	9	15 à 20	70
CHOPS	3	Structures condensées (5)	36	20 à 50	53
CHNOS	11	Lignines (6)	28	> 50	0
CHN	30	Tannins (7)	1		
CHNOPS	9	« Hors catégories »	88		
Halogènes	16				
Autres	35				

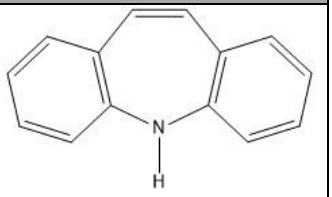
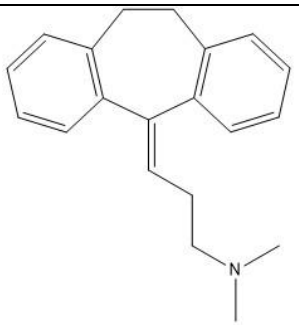
Selon la classification de Hockaday *et al.* (2009), ces résultats montrent une concentration importante des signaux pertinents dans les zones des classes de lipides (1), protéines (2), structures condensés (5) et lignines (6). En comparaison, les familles des sucres aminés (3), carbohydrates (4) et tannins (7) sont peu ou moins représentées. Une fraction importante des signaux sélectionnés pourrait donc être liée à la matière organique naturelle présente dans le Clain et pourrait aussi être apportée par la STEP. Aucun type de formules brutes n'est relié spécifiquement à une classe de molécules mais on peut noter certaines associations. Ainsi, les formules brutes de type CHNOP sont plutôt reliées aux lipides ou sont en dehors des catégories définies (hors catégories). Ces molécules semblent être des molécules saturées en H, en raison du ratio H/C supérieur à 1,5 pour la majorité d'entre eux. Les formules brutes de type CHNO sont plus reliées aux classes (1)-(2)-(5)-(6). Elles pourraient correspondre en partie à des molécules saturées mais aussi insaturées et donc à des structures variées tout comme les composés de type CHO. Les formules brutes avec un halogène présentent un rapport H/C faible indiquant soit des molécules ayant un cycle aromatique soit des doubles ou triples liaisons. Le reste des signaux, représentant 32% du nombre total, ne font partie d'aucune des sept familles identifiées.

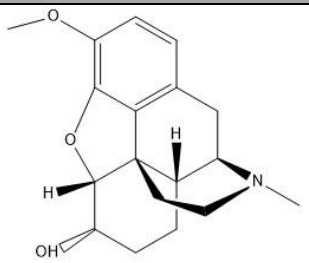
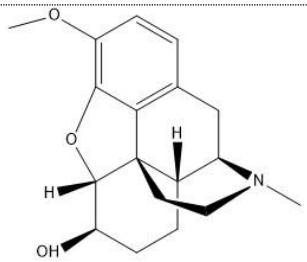
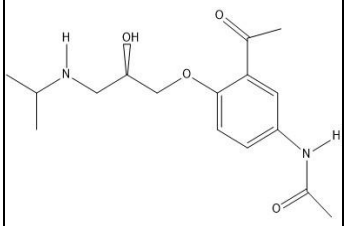
Le graphique représentant les ratios H/C en fonction de N/C montre une tendance d'augmentation conjointe des deux ratios : le ratio N/C augmente inversement au degré d'insaturation, comme cela avait déjà été indiqué lors de la sélection des signaux pertinents par comparaison des eaux de sortie avec les eaux d'entrées de STEP (chapitre 4 « Caractérisation des eaux usées brutes et traitées », paragraphe V.2.1.). L'ensemble des types de formules brutes est réparti sur la gamme de ratio de N/C et aucun type de formule brute spécifique ne semble lié à une gamme précise de ce ratio. Il peut être noté sur ce graphique que la majeure partie (91%) des signaux a un ratio N/C inférieur à 0,8. Ce résultat est en accord avec les résultats de Kind et Fiehn (2007) qui ont montré que 99,7% des petites molécules ont un ratio N/C inférieur à 1,3.

III.2.2. Proposition d'annotations par analyse des spectres MS^2

Parmi les 305 signaux sélectionnés, 52 d'entre eux ont été fragmentés avec l'obtention d'un spectre MS^2 lors des analyses UHPLC-HRMS. Pour ces signaux une confrontation des spectres MS^2 avec ceux de la base de données mzCloud a été réalisée pour faire une proposition de formule développée. Ainsi, 13 signaux ont reçu une ou plusieurs propositions de structures pour un total de propositions (dont 5 avec un pourcentage de correspondance supérieur à 80% et 16 avec un pourcentage de correspondance inférieur à 80%). Le Tableau 25 présente la liste des formules développées pour les 4 signaux ayant reçu des propositions (dont 2 pour un même signal) dont le pourcentage de correspondance est supérieur à 80% signaux. La liste complète est présentée en Annexe 10.

Tableau 25 : Liste des 4 signaux ayant une proposition d'identification par mzCloud avec une déviation en masse inférieure à 5 ppm et une correspondance supérieure à 80%.

Propositions de structures	Formule brute	Masse exacte (Da)	Tr (min)	Famille	Structure
Iminostilbene	$C_{14}H_{11}N$	193,0897	7,64	Métabolite de produit pharmaceutique	
Amitriptyline	$C_{20}H_{23}N$	277,1829	7,63	Produit pharmaceutique	

Propositions de structures	Formule brute	Masse exacte (Da)	Tr (min)	Famille	Structure
Dihydrocodeine	$C_{18}H_{23}NO_3$	301,1675	3,74	Produit pharmaceutique	
6β-Hydrocol				Produit pharmaceutique	
Diacetalol	$C_{16}H_{24}N_2O_4$	308,1736	4,43	Métabolite de produit pharmaceutique	

Ainsi, par comparaison des échantillons de l'aval et de l'amont, 305 signaux pertinents susceptibles de provenir de la STEP ont été mis en évidence. Une formule brute a été attribuée à 90% d'entre eux (273), 17% (52) seulement ont été fragmentés lors des analyses MS^2 et 4% (13) ont reçu une ou plusieurs propositions de structure ayant une déviation en masse inférieure à 5 ppm.

IV. Recherche de suspects et de métabolites

IV.1. Recherche de suspects dans les eaux du Clain

Une recherche de suspects a été réalisée dans les échantillons prélevés dans le Clain à l'amont et à l'aval du rejet de la STEP. Pour cela, la liste de suspects utilisée (1 487 molécules) a été la même que celle appliquée dans le chapitre 4 « Caractérisation des eaux usées brutes et traitées » pour la STEP de Poitiers et un même schéma d'étapes (chapitre 4 « Caractérisation des eaux usées brutes et traitées » VI.1.) a ainsi été adopté :

- confrontation entre la liste de suspects et le tableau des signaux détectés dans les échantillons,
- conservation des signaux ayant un spectre MS^2,
- vérification de la compatibilité des temps de rétention avec les logP et logD calculés par ChemAxon puis élimination des faux positifs,

- confrontation entre les spectres MS^2 et la base de données mzCloud incluse dans Compound Discoverer,
- recherche dans MetFrag lorsqu'il n'y pas de résultat dans mzCloud dans la base de données KEGG puis CompTox (07/03/2019).

Dans ces conditions, la confrontation entre la liste de suspects et le tableau des signaux des échantillons du Clain a réduit le tableau à 2 000 signaux pouvant correspondre à un ou plusieurs suspects. Puis après élimination des signaux n'ayant pas de spectres MS^2, des faux positifs et vérification de l'adéquation des temps de rétention, 127 signaux ont été conservés. La liste de ces signaux est présentée en Annexe 11. Parmi ces 127 signaux, 61 d'entre eux ont présenté un spectre MS^2 compatible avec un spectre de la base de données mzCloud dont 38 avec une correspondance calculée par mzCloud supérieure à 80%. Les spectres des 66 autres signaux ont été soumis à Metfrag avec dans un premier temps une recherche dans les bases de données KEGG, HMDB et CompTox (7 mars 2019). A partir de cette approche, des structures ont été proposées pour 53 d'entre eux (dont 13 présentant une correspondance supérieure à 50%). Enfin parmi les 127 signaux pouvant correspondre à un ou plusieurs suspects, trois suspects (paracétamol, carbamazépine et propranolol) ont été confirmés par les temps de rétention, les spectres MS et les spectres MS^2, lors de l'injection d'étalons. Les spectres MS2 et les chromatogrammes de ces 3 composés sont présentés en Annexe 12. La Figure 72 récapitule le nombre de signaux restant à chaque étape.

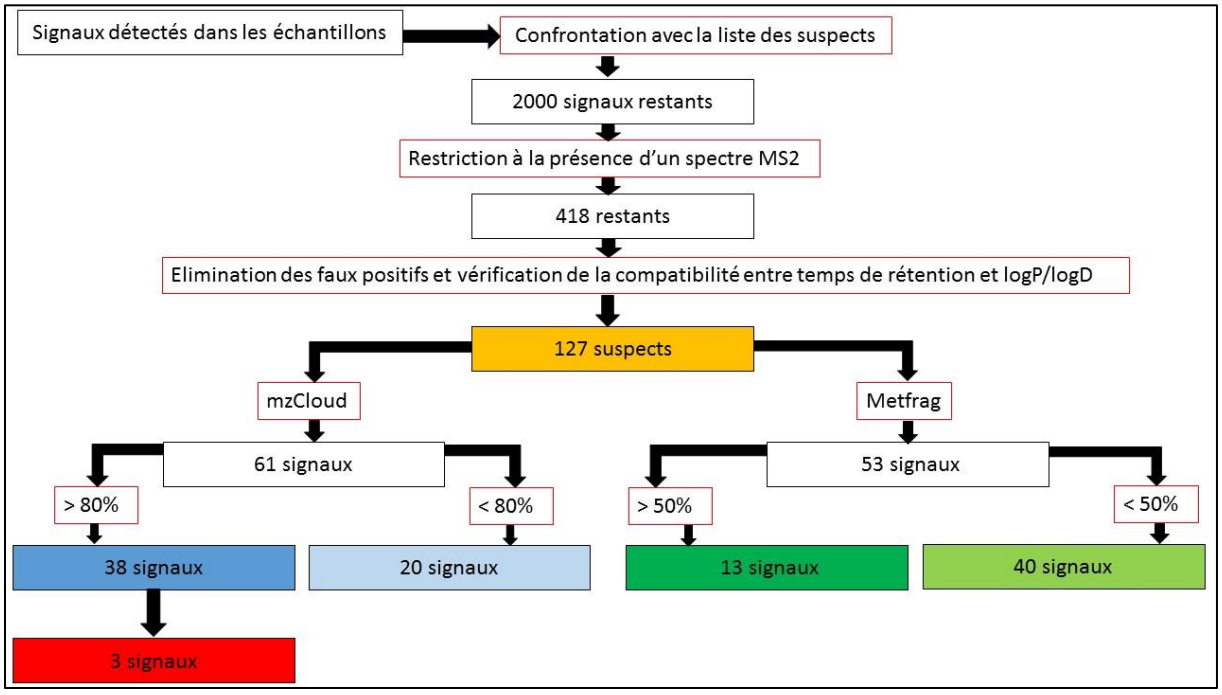


Figure 72 : Schéma d'étapes de l'analyse de suspects ayant conduit à l'obtention d'une liste de 127 signaux confrontés à mzCloud ou MetFrag (KEGG, HMDB et Comptox).

Parmi ces 127 signaux suspects, la carbamazépine epoxide, le galaxolidone, le 4'-formylaminoantypiridine ont déjà été détectés par Kern *et al.* (2009) dans des échantillons d'eaux de surface. D'autres suspects tels que les bisoprolol, celiprolol, fexofenadine, lamotrigine ou encore la sulfapyridine ont été également déjà observés par Li *et al.* (2017) dans des eaux de surface.

La présence de chacun de ces 127 signaux, dans au moins un des échantillons de chaque type (amont et aval), a été contrôlée et le résultat est indiqué dans la colonne « catégorie » du tableau en Annexe 11. Ainsi 49% des signaux seraient présents seulement en aval, 5% seulement en amont et 46% à la fois en amont et en aval. Les signaux uniquement présents en aval proviennent vraisemblablement en grande partie du rejet de la STEP de Poitiers. Dans les cas des signaux présents dans les deux échantillons, 62% des signaux ont au moins une aire 2 fois plus élevées en aval qu'en amont. Il semble possible d'envisager la présence de ces suspects en amont, liée à d'autres sources de contamination, puis un nouvel apport de ces molécules avec les rejets de la STEP de Poitiers, augmentant ainsi l'aire des signaux en aval. Ces résultats sont cohérents avec ceux de Ruff *et al.* (2015) qui ont montré une augmentation de la concentration de certains contaminants principalement liée à la présence de STEP au niveau de plusieurs points d'échantillonnage dans le Rhin.

La liste de 127 signaux suspects a ensuite été confrontée à la liste des 52 signaux pertinents retenus pour la comparaison amont/aval ayant un spectre MS^2 afin de rechercher des signaux communs et compléter si besoin la liste des structures proposées pour les signaux pertinents (Tableau 25 et Annexe 10). Au total, 8 signaux communs ont été trouvés, présentant les mêmes masses exactes, temps de rétention et spectres MS^2. Ces 8 signaux sont listés dans le Tableau 26.

Tableau 26 : Liste des signaux présents dans la liste de 127 signaux suspects et la liste des 52 signaux pertinents ayant un spectre MS^2. * : pourcentage de correspondance inférieur à 80% dans mzCloud. Tr : Temps de rétention.

Nom(s) suspect(s)	Formule brute	Masse exacte (Da)	Tr (min)	Base de données
Sulfapyridine*	$C_{11}H_{11}N_3O_2S$	249,0576	3,96	mzCloud
O-desmethyltramadol	$C_{15}H_{23}NO_2$	249,1724	5,32	MetFrag
Cimetidine*	$C_{10}H_{16}N_6S$	252,11537	3,52	mzCloud
Atenolol*	$C_{14}H_{22}N_2O_3$	266,16267	1,25	mzCloud
Amitriptyline/Maprotiline*	$C_{20}H_{23}N$	277,18286	7,54	mzCloud
Venlafaxine	$C_{17}H_{27}NO_2$	277,20562	5,36	MetFrag
Bestatin	$C_{16}H_{24}N_2O_4$	308,17363	4,35	MetFrag
Desmethylcitalopram	$C_{19}H_{19}FN_2O$	310,14814	6,69	mzCloud

Ces 8 signaux pourraient correspondre à des produits pharmaceutiques ou à des métabolites de pharmaceutiques. Dans le cas où ces suspects seraient confirmés et sachant que ces signaux sont également des signaux considérés comme pertinents, cela confirmerait le rôle des STEP dans la contamination des rivières par ce type de composés.

IV.2. Recherche de métabolites complémentaires dans les eaux du Clain

Une recherche de métabolites supplémentaires dans les eaux du Clain a été effectuée à partir de la même liste (233 molécules) générée *in silico* (à partir de 60 molécules) que celle utilisée pour les échantillons de la STEP de Poitiers (chapitre 4 « Caractérisation des eaux usées brutes et traitées », paragraphe VI.2).

Cette liste de métabolites potentiels a été confrontée au tableau des signaux des échantillons du Clain et 419 d'entre eux ont ainsi été conservés. Seuls 88 de ces signaux ont été fragmentés lors des analyses MS^2. Après élimination des faux positifs et des suspects dont le logP/logD est incompatible avec le temps de rétention du signal obtenu, 34 signaux listés en Annexe 13 étaient encore présents. Pour ces 34 signaux des propositions de formules développées ont été réalisées *via* mzCloud ou MetFrag (par les bases de données KEGG puis CompTox puis HMDB) par comparaison des spectres MS^2 du signal avec ceux obtenus *in silico.* Sur les 34 signaux listés, 9 ont une correspondance avec mzCloud, dont 4 avec un pourcentage supérieur à 80% et 10 ont une correspondance avec MetFrag. Le Tableau 27 présente les 9 signaux correspondant à un ou plusieurs métabolites suspects ayant une correspondance avec mzCloud. Dans le cas du desméthyltramadol et de la desméthylvenlafaxine, ces structures correspondent à plusieurs signaux.

Tableau 27 : Liste des métabolites suspects ayant une correspondance avec mzCloud. N : nature de l'échantillon dans lequel le signal a été détecté, Tr : Temps de rétention. * : molécule parente présente dans la liste de 198 signaux suspects.

Nom du métabolite	Formule brute	Masse exacte (Da)	Tr (min)	Molécule parente	N
para-phenyldiamine	$C_6H_8O_2$	108,0690	1,21	4'-Aminoacetanillide	Cav
para-aminophenol	C_6H_7NO	109,0531	1,27	Paracétamol*	Cam/Cav
4-acetaminoantipyrine	$C_{13}H_{15}N_3O_2$	245,1167	4,63	4-Aminoantipyrine 4-Formylaminoantipyrine	Cam/Cav
N-desméthyltramadol	$C_{15}H_{23}NO_2$	249,1724	5,32	Tramadol	Cam/Cav
O-desméthyltramadol			4,45 3,84		
N-desméthylvenlafaxine	$C_{16}H_{25}NO_2$	263,1880	5,35	Venlafaxine*	Cam/Cav
O-Desméthylvenlafaxine					
Acide aténolique	$C_{14}H_{21}NO_4$	267,1472	4,43	Aténolol*	Cam/Cav
Norcodéine	$C_{17}H_{19}NO_3$	285,1362	3,64	Codéine*	Cav

En plus de ces candidats, la desoaminylazithromycine et le O-desmethylomeprazole présentent une correspondance dans MetFrag. Cependant, la stéréochimie de ces suspects proposée dans MetFrag n'est pas similaire à celle des métabolites de la liste établie dans ce travail.

Sept des 9 suspects sont présents dans les deux types d'échantillons tandis que deux sont présents uniquement en aval. Pour ces deux derniers métabolites, une probable origine par la STEP de Poitiers peut donc être envisagée. Dans le cas de la norcodéine et du paraphenyldiamine, ces composés ont en effet été retrouvés à la sortie de la STEP. De même, pour ces deux métabolites les molécules parentes ont été observées au niveau de la STEP (entrée et sortie). Toutefois, la codéine, molécule parente, fait également partie des suspects conservés dans le paragraphe IV.1 précédent et susceptibles de se retrouver en amont. La dégradation de cette molécule dans le Clain pourrait donc également justifier la présence de la norcodéine en aval. Pour les sept autres suspects, leur présence dans les deux types d'échantillons peut s'expliquer par une contamination en amont par le métabolite. De plus, cinq d'entre eux ont une valeur d'aire supérieure en aval à celle en amont, indiquant leur présence probable dans les rejets de la STEP. Pour plusieurs de ces métabolites, les molécules parentes sont également présentes dans la liste de suspects détectés précédemment et susceptibles d'être observées en amont.

Parmi les 34 signaux correspondant aux métabolites présents dans le Clain (pour lesquels un spectre MS^2 a été obtenu et ayant un temps de rétention cohérent), un unique signal appartient à la liste des

signaux pertinents sélectionnés. Ce signal associé à une masse de 249,1724 Da et un temps de rétention de 5,32 min pourrait correspondre au N-desméthlyltramadol ou au O-desmethyltramadol.

V. Conclusion

L'étude et la comparaison des échantillons du Clain ont mis en lumière l'impact du rejet de la STEP sur l'empreinte moléculaire de la rivière. Un effet de dilution a été constaté entre le panache et l'aval immédiat au rejet. La comparaison entre les empreintes des échantillons des sorties de STEP, du panache, de son aval et de son amont, a en effet montré d'une part : un plus faible nombre de signaux à l'amont et une diminution de ce nombre, des sorties de la STEP jusqu'à l'aval, et d'autre part : une diminution parallèle des aires des signaux. De plus, la comparaison directe entre l'amont et l'aval a montré l'apparition de nouveaux signaux qui composent plus de 50% des signaux des empreintes en aval. Ainsi, si l'effet de la dilution du rejet dans le Clain est important, l'impact des rejets de STEP reste toutefois bien visible en aval immédiat du Clain et ce pour les 3 campagnes.

Des analyses en composantes principales ont montré une séparation entre les différents types d'échantillons (amont, aval et panache). Il a été observé que les échantillons de l'aval sont situés entre ceux de l'amont et ceux du panache (mais globalement plus proches des échantillons de l'amont). Concernant les signaux, une corrélation préférentielle des signaux apparaissant et augmentant entre l'amont et l'aval avec les échantillons du panache a été observée. Pour ces signaux, un apport principal *via* le rejet des eaux de STEP peut donc être envisagé.

Afin de mettre en évidence des signaux pertinents témoignant de l'impact de la STEP, une analyse différentielle entre les échantillons de l'amont et ceux de l'aval a été réalisée. Ainsi, 305 signaux pertinents communs aux trois campagnes de prélèvements ont été sélectionnés Ces signaux sont majoritairement des signaux de masse inférieure à 500 Da couvrant une large gamme de polarité et de formules de type CHNO, CHNOP et CHNOS associées aux catégories des lipides, des protéines et des structures condensées.

Enfin, une recherche de suspects et de métabolites suspects a été réalisée et 161 signaux (127 suspects et 34 métabolites suspects) ont été conservés.

Chapitre 6 : Bilan et étude critique

I. Introduction

Dans le cadre de ce travail, la sélection de signaux pertinents et la recherche de suspects ont été effectuées à partir des eaux de la STEP puis à partir des eaux du Clain de manière séparée. Dans ces conditions, la recherche de signaux pertinents a conduit à la sélection de plus de 300 signaux dans les deux cas et certains d'entre eux ont reçu une ou plusieurs propositions de structures. La recherche de suspects a permis la mise en évidence de différents suspects et métabolites suspects potentiels dans ces échantillons.

Cependant, les eaux de la STEP et du Clain sont liées par le rejet des eaux traitées dans la rivière. Ainsi, l'objectif de ce chapitre a été de comparer les résultats obtenus séparément à partir des deux catégories d'échantillons. Dans le cas de la recherche de signaux pertinents, ce croisement a permis de prêter un regard critique sur les deux approches employées, d'étudier leurs limites, mais également de proposer une hiérarchisation des signaux sélectionnés en termes de traceurs des rejets de la STEP dans la rivière.

Une analyse multivariée prenant en considération l'ensemble des échantillons d'une campagne a ensuite été réalisée.

II. Comparaison des suspects et métabolites suspects trouvés dans les eaux de la sortie de la STEP et du Clain

Des suspects et des métabolites suspects ont été recherchés dans les échantillons d'eaux de la STEP de Poitiers et d'eaux de la rivière Clain dans les chapitres 4 et 5. Dans les échantillons de la STEP, 198 suspects et 89 métabolites ont été détectés (répertoriés dans les annexes 5 et 8). Pour les eaux du Clain, 127 suspects et 34 métabolites suspects ont été listés dans les annexes 11 et 13. Les listes de suspects et les listes de métabolites ont été croisées entre elles afin de rechercher les signaux communs, présents dans les différentes catégories d'échantillons : eaux de sortie de la STEP et eaux de la rivière (amont et aval au rejet de la STEP). La Figure 73 présente ainsi les signaux communs entre les trois types d'échantillons pour les suspects et les métabolites suspects.

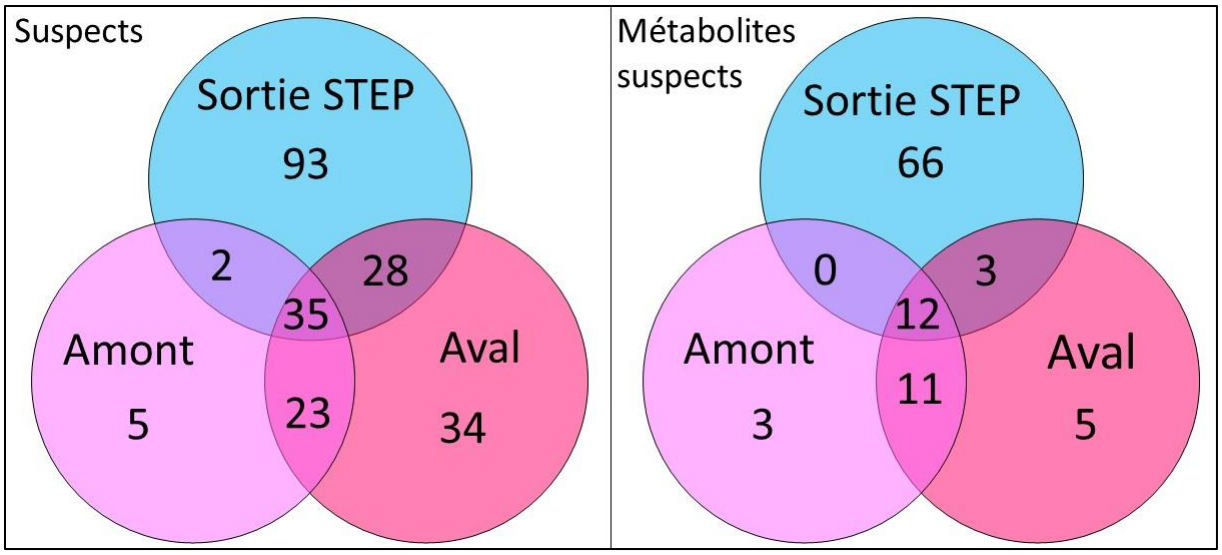


Figure 73 : Diagrammes de Venn des signaux communs entre la sortie de STEP, l'amont et l'aval du Clain dans le cas des suspects et des métabolites suspects.

La Figure 73 montre que l'amont du Clain possède moins de suspects (métabolites compris) communs avec la sortie de la STEP que son aval. Après confrontation, 37 suspects et 12 métabolites suspects sont en effet communs entre la sortie de STEP et l'amont tandis que 63 suspects et 15 métabolites suspects sont communs entre la sortie et l'aval.

Des effets matrices et des faux négatifs ou positifs doivent bien entendu rester envisageables. Toutefois, lors de l'absence de ces suspects en amont, la contamination principale du Clain par le rejet de la STEP semble probable. Lorsque les signaux sont présents à la fois en amont et en aval, 77% des signaux des suspects et 100% de ceux des métabolites suspects ont une valeur d'aire au moins deux fois plus importante en aval qu'en amont, ce qui semble montrer un apport complémentaire par le rejet de la STEP de ces contaminants déjà présents en amont. Dans ce dernier cas, la détection en amont pourrait provenir d'autres sources de contamination. Enfin, pour les suspects et métabolites suspects présents en sortie de STEP mais pas en aval, la dilution du rejet dans le Clain pourrait expliquer en grande partie l'absence de ces molécules en aval (l'aire de leurs signaux devenant inférieure au seuil de détection).

A noter que parmi les métabolites suspects présents à la sortie de la STEP et en aval, l'acide aténolique, la desvenlafaxine, la norcodéine, la N-desmethylvenlafaxine, le para-aminophénol ont pour molécules parentes des suspects également présents dans les mêmes échantillons.

III. Etude critique sur la recherche de signaux pertinents

Les signaux « pertinents », c'est-à-dire pouvant être considérés comme traceurs du rejet de la STEP dans le Clain, ont été recherchés dans le cadre de ce travail. Pour cela, deux approches ont été utilisées. La première approche, celle du « traiteur d'eau » en quelque sorte, a consisté à comparer les entrées et les sorties de STEP pour rechercher les signaux réfractaires au traitement d'épuration ou les signaux augmentant ou apparaissant en sortie de STEP. De cette manière, 375 signaux ont été sélectionnés comme pertinents. Une formule brute a été attribuée pour 331 de ces signaux et des formules développées ont été proposées pour 28 d'entre eux.

La seconde approche, celle de l' « environnementaliste », a consisté à comparer l'amont et l'aval du Clain pour rechercher les signaux augmentant ou apparaissant en aval. Il s'agit ici de caractériser l'impact d'un rejet d'eaux traitées sur la rivière. Lors de cette seconde approche, 305 signaux ont été sélectionnés. Une formule brute a été attribuée pour 273 d'entre eux et des formules développées ont été proposées pour 15 signaux.

Le Tableau 28 présente les signaux pertinents ayant reçu une ou plusieurs propositions de structures et classés en fonction de l'approche utilisées pour les mettre en évidence : approche du « traiteur d'eau » (approche 1) et approche « environnementaliste » (approche 2).

Tableau 28 : Listes des formules développées proposées pour les signaux pertinents obtenus à partir des approches 1 et 2. Tr : temps de rétention, MC (%) : pourcentage de correspondance du spectre MS^2 expérimental avec la base de données mzCloud ou MetFrag*.

Approche(s)	Proposition(s) de structure(s)	Formule Brute	Masse exacte (Da)	Tr (min)	MC (%)
Approche 1	5-amino-o-crésol	C_7H_9NO	123,0685	1,52	83,9
	2-amino-4-crésol				82,5
	2-anisidine				76,9
	2-pyridine ethanol				74,9
	6-methyl-2-pyridinemethanol				73,7
Approche 1	Tropine	$C_8H_{15}NO$	141,1153	6,77	77,9
	3,3,5,5-Tetramethylpyrroline-N-oxide				77,2
Approche 1	N-Cyanoimido-S,S-dimethyl-dithiocarbonate	$C_4H_6N_2S_2$	145,9972	5,10	61,8
Approche 1	NP-007065	$C_8H_{10}O_3$	154,0630	4,85	67,2
	2,6-Dimethoxyphenol				65,2
Approche 1	2-(Cyclopropylcarbonyl)-3-(dimethylamino)acrylonitrile	$C_9H_{12}N_2O$	164,0949	1,64	68,2
Approche 1	Edrophonium	$C_{10}H_{15}NO$	165,1154	8,28	77,0
	Perillartine				73,2

Approche(s)	Proposition(s) de structure(s)	Formule Brute	Masse exacte (Da)	Tr (min)	MC (%)
Approche 1	(4E)-3-Hydroxy-2,4-dimethyl-4-heptenamide	$C_9H_{17}NO_2$	171,1268	6,23	77,7
Approche 1	α-Methyltryptamine	$C_{12}H_{13}NO_2$	174,1161	4,53	71,5
	1(1H-indol-5-yl)-2-propanamine				65,9
	1(1H-indol-6-yl)-2-propanamine				61,7
Approche 1	2-Amino-6-methylmercaptopurine	$C_6H_7N_5S$	181,0425	3,62	79,9
	2-(methylthio)[1,2,4]triazolo[1,5-a]pyrimidin-7-amine				66,3
Approche 1	4-Methyl-N,N-dimethylcathinone	$C_{12}H_{17}NO$	191,1310	7,87	61,9
Approche 1	5-Phenylnicotinic acid	$C_{12}H_9NO_2$	199,0634	2,50	78,6
Approche 1	4-morpholinobenzoic acid	$C_{11}H_{13}NO_3$	207,0930	2,46	66,6
Approche 1	Tapentadol	$C_{14}H_{23}NO$	221,1773	8,78	62,4
Approche 1	Desmethyl nefopam	$C_{16}H_{17}NO$	239,1307	5,90	85,2
Approche 1	Alprenolol	$C_{15}H_{23}NO_2$	249,1741	4,36	70,1
Approche 1	O-desmethylvenlafaxine	$C_{16}H_{25}NO_2$	263,1881	5,06	83,2
	N-Desmethylvenlafaxine				79,8*
Approche 1	Oxprenolol	$C_{15}H_{23}NO_3$	265,1681	4,76	30,6*
Approche 1	Métabolite du Valsartan	$C_{14}H_{10}N_4O_2$	266,0800	6,8	86,6
Approche 1	Galaxolidone	$C_{18}H_{24}O_2$	272,1767	8,87/ 9,23	95,9
	Estradiol				76 ,4
	19-Norandrostenedione				64,7
Approche 1	Cycloheximide	$C_{15}H_{23}NO_4$	281,1623	4,05	18,4*
Approche 1	Codeine	$C_{18}H_{21}NO_3$	299,1524	2,95	78,7
	Hydrocodone				78,8
Approche 1	Flubendazole	$C_{16}H_{12}FN_3O_3$	313,0876	7,54	66,5
Approche 1	Kresoxim méthyle	$C_{18}H_{19}NO_4$	313,1310	8,29	41,4*
Approche 1	N,N-Dimethyl-4-{5-[(3S)-1-(phenylsulfonyl)-3-pyrrolidinyl]-1,3,4-oxadiazol-2-yl}aniline	$C_{20}H_{22}N_4O_3S$	398,1416	4,42	71,9
Approche 1	AB-FUBINACA 2A metabolite	$C_{20}H_{19}FN_4O_4$			65,9
Approche 1	Candesartan	$C_{24}H_{20}N_6O_3$	440,1586	7,92	83,7
Approche 1-2	N-Cyanoimido-S,S-dimethyl-dithiocarbonate	$C_4H_6N_2S_2$	145,9972	5,10	61,8
Approche 1-2	Indole-3-butyric acid	$C_{12}H_{13}NO_2$	203,0949	4,78	76,0
	3-ethyl-4-hydroxy-1-methyl-1,2-dihydroquinolin-2-one				71,6
	N1-(5-oxo-5,6,7,8-tetrahydronaphthalen-2-yl)acetamide				65,0
	Methsuximide				64,4
Approche 2	1,3-Divinyl-2-imidazolidinone	$C_7H_{10}N_2O$	138,0792	4,28	70,2

Approche(s)	Proposition(s) de structure(s)	Formule Brute	Masse exacte (Da)	Tr (min)	MC (%)
Approche 2	Iminostilbene	$C_{14}H_{11}N$	193,0897	7,64	82,4
	2-Phenylindole				71,3
Approche 2	Sulfapyridine	$C_{11}H_{11}N_3O_2S$	249,0576	4,05	75,9
Approche 2	N-desmethyltramadol	$C_{15}H_{23}NO_2$	249,1724	5,32	61,3
	O-desmethyltramadol				31,9*
	Alprenolol				71,3
Approche 2	Cimetidine	$C_{10}H_{16}N_6S$	252,1154	3,60	72,2
Approche 2	Atenolol	$C_{14}H_{22}N_2O_3$	266,1627	1,34	75,6
Approche 2	Amitriptyline	$C_{20}H_{23}N$	277,1829	7,63	85,1
	Maprotiline				70,8
Approche 2	Venlafaxine	$C_{17}H_{27}NO_2$	277,2056	5,36	30,8*
Approche 2	5-Hydroxy-7-(hydroxymethyl)-2-methyl-2-(5-oxotetrahydro-2-furanyl)-2,3-dihydro-4H-chromen-4-one	$C_{15}H_{16}O_6$	292,0974	7,51	65,5
	3-(3-Acetyl-2,6-dihydroxy-5-methylbenzyl)-4-hydroxy-5-methyl-2(5H)-furanone				63,0
Approche 2	Dihydrocodeine	$C_{18}H_{23}NO_3$	301,1675	3,74	82,3
	6β-Hydrocol				82,3
Approche 2	Diacetalol	$C_{16}H_{24}N_2O_4$	308,1736	4,43	85,0
	Bestatin				50,0*
Approche 2	Desmethylcitalopram	$C_{19}H_{19}FN_2O$	310,1481	6,78	71,3
Approche 2	Disopyramide	$C_{21}H_{29}N_3O$	339,2312	5,46	79,2

Dans cette partie, afin d'apporter un regard critique sur les deux approches ayant mené à la sélection des signaux pertinents et étudier leurs limites, une comparaison des signaux communs aux deux approches puis un croisement entre les listes de pertinents et les signaux des échantillons ont été effectués dans l'ordre suivant :

- recherche des signaux pertinents communs aux deux approches,
- recherche des signaux pertinents de l'approche 1 (« traiteur d'eau ») dans les échantillons amont et aval du Clain,
- recherche des signaux pertinents de l'approche 2 (« environnementaliste ») dans les échantillons de sorties et d'entrées de la STEP.

A partir de cette démarche une proposition de hiérarchisation des signaux pertinents en termes de traceurs des rejets de la STEP dans la rivière a ensuite été réalisée et est présentée dans la conclusion de ce chapitre.

III.1. Comparaison des signaux pertinents obtenus avec les deux approches

Les listes de signaux pertinents établies par les deux approches ont ainsi été confrontées afin de rechercher les signaux communs. La Figure 74 montre le diagramme de Venn des signaux communs aux deux listes.

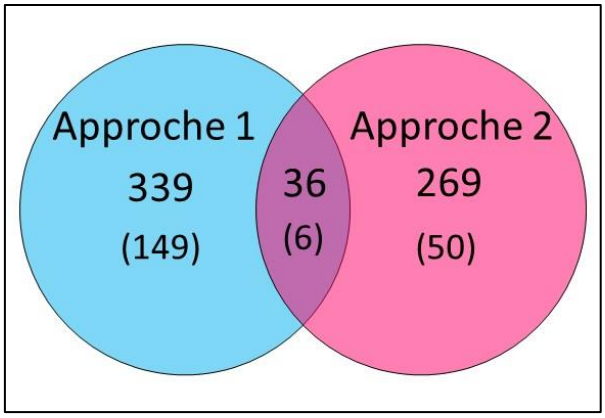


Figure 74 : Diagramme de Venn des signaux pertinents communs obtenus par l'approche 1 (« traiteur d'eau ») et l'approche 2 (« environnementaliste »). Les nombres entre parenthèses indiquent le nombre de signaux ayant un spectre MS^2.

Ainsi, 36 signaux pertinents communs ont été déterminés (soit 9,6% des 375 signaux de la liste des pertinents de l'approche 1 et 11,8% des 305 signaux pertinents de l'approche 2). La liste des signaux pertinents communs aux deux approches est présentée en Annexe 14. Dans cette liste, seuls deux signaux ont une ou plusieurs propositions de structures.

Seule une faible partie des signaux pertinents est donc commune aux deux approches. Les deux approches aboutissent donc dans l'ensemble à la sélection de signaux différents. Pour expliquer ce résultat, plusieurs hypothèses sont possibles. Par exemple :

- Des signaux sélectionnés par l'approche 1 ne seraient pas retenus par l'approche 2, si après dilution du rejet dans le Clain, ils ne sont plus détectés en aval,
- Des signaux sélectionnés par l'approche 2 ne seraient pas sélectionnés par l'approche 1 si :
    - des effets matrices plus importants au sein de la STEP empêchent leur détection,
    - ils proviennent d'autres sources de contamination (ou de la dégradation de molécules) entre l'amont et l'aval du Clain,
    - ces signaux issus du rejet sont suffisamment intenses pour pouvoir être observés en aval mais correspondent à des signaux diminuants entre l'entrée et la sortie de STEP.

Le croisement de ces listes a donc permis de mettre en lumière des signaux communs aux deux approches et d'en souligner les limites.

III.2. Recherche des signaux pertinents obtenus avec l'approche 1 « traiteur d'eau » (sortie/entrée STEP) dans le Clain

Les signaux pertinents issus de l'approche 1 ont été recherchés dans les échantillons amont et aval du Clain. La Figure 75 présente le diagramme de Venn du nombre de signaux pertinents détectés dans les échantillons amont et aval.

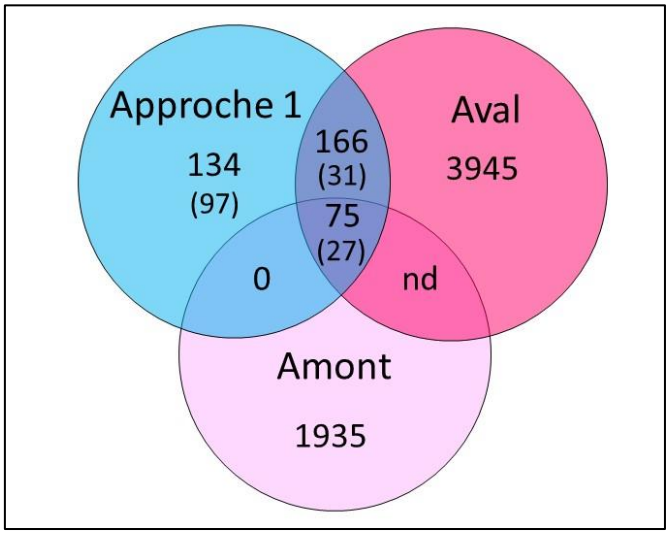


Figure 75 : Diagramme de Venn des signaux pertinents sélectionnés par l'approche 1 (« traiteur d'eau ») détectés dans les échantillons amont et aval. Les nombres entre parenthèses indiquent le nombre de signaux ayant un spectre MS^2. Le nombre de signaux dans les cercles amont et aval représente la moyenne du nombre de signaux sur les trois campagnes. nd : nombre de signaux communs non déterminé

Pour que les signaux sélectionnés avec l'approche 1 soient réellement pertinents comme traceurs de rejet de la STEP dans la rivière, ces signaux doivent être détectés en aval immédiat.

Cette figure montre que 241 signaux pertinents sélectionnés avec l'approche 1 (44 %) sont retrouvés en aval immédiat. Parmi ces signaux, 75 ont également été détectés en amont et 20 (10 en aval et 10 communs aval/amont) ont reçu une proposition de structure. Aucun signal pertinent n'a été détecté exclusivement en amont.

Les signaux présents exclusivement en aval proviennent *a priori* uniquement du rejet de la STEP tandis que les signaux aussi retrouvés en amont peuvent provenir de la STEP mais également d'autres sources de contamination. Pour ces signaux, les aires moyennes sont plus élevées en aval qu'en amont montrant une probable contamination complémentaire du Clain par le rejet.

Pour expliquer l'absence de certains signaux pertinents sélectionnés avec l'approche 1 en aval, au moins deux hypothèses sont possibles : la première pourrait provenir de la dilution du rejet dans le Clain rendant indétectables ces signaux en aval et la seconde pourrait être un biais de sélection en faveur d'un faux positif lors de la mise en œuvre de l'approche 1.

La détection des signaux dans l'amont du Clain pourrait être considérée comme une limite de l'approche 1, le but étant de rechercher des signaux traceurs du rejet de STEP. Toutefois, cela n'exclut pas complétement une origine *via* la STEP si une variation significative des aires de ces signaux est observée entre l'amont et l'aval. La STEP constituerait dans ce cas une source de contamination supplémentaire dans le Clain.

L'ensemble des résultats (notamment la présence de signaux pertinents de l'approche 1 dans l'aval) conforte globalement la capacité de la sélection des signaux par la réalisation d'analyses différentielles.

III.3. Recherche des signaux pertinents obtenus avec l'approche 2 « environnementaliste » (Aval/Amont du Clain) dans la STEP

De la même manière, de façon à confirmer la pertinence des signaux sélectionnés, les signaux pertinents obtenus avec l'approche 2 ont été recherchés dans les échantillons de STEP, entrées et sorties. La Figure 76 montre le nombre de signaux pertinents obtenus avec l'approche 2, trouvés en sortie, en entrées et dans les deux types d'échantillons.

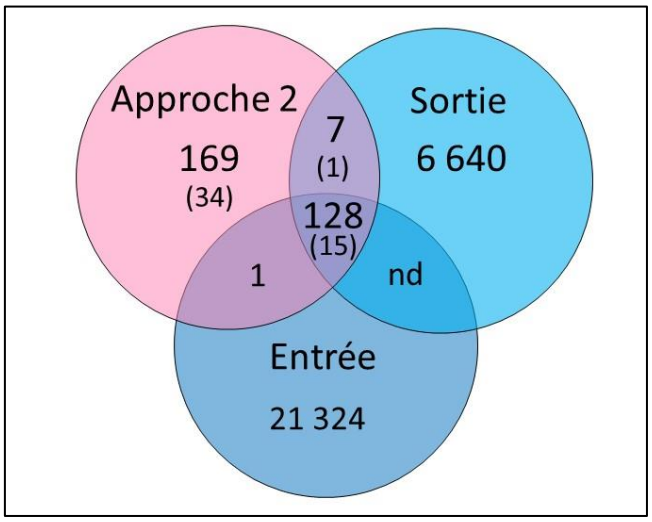


Figure 76 : Diagramme de Venn des signaux pertinents sélectionnés par l'approche 2 (« environnementaliste ») détectés dans les échantillons d'entrée et de sorties. Les nombres entre parenthèses indiquent le nombre de signaux ayant un spectre MS^2. Le nombre de signaux dans les cercles amont et aval représente la moyenne du nombre de signaux sur les trois campagnes. nd : : nombre de signaux communs non déterminé

Pour que les signaux sélectionnés avec l'approche 2 soient réellement pertinents, il faudrait qu'ils soient également retrouvés dans la sortie de STEP. Cette Figure montre que 135 signaux sur 305 signaux pertinents (44%) sont présents dans les sorties de STEP (dont 128 également présents dans les entrées). Parmi ces signaux, 11 ont reçu une proposition de structure.

Ainsi, 56% des signaux pertinents sélectionnés avec l'approche 2 ne sont pas présents en sortie de STEP. Pour ces signaux, en plus de biais de sélection possibles lors de l'utilisation de l'approche 2, plusieurs hypothèses peuvent expliquer leur absence dans les échantillons de STEP :

- d'autres apports de contamination que la STEP entre l'amont et l'aval,
- la dégradation de composés parents présents en amont conduisant à l'apparition de nouveaux signaux observés en aval.

IV. Bilan : Analyse multivariée sur l'ensemble des points de prélèvements d'une campagne

Afin de déterminer si les différents types de prélèvements peuvent être regroupés par catégorie, une ACP a été réalisée sur l'ensemble des échantillons de la campagne C. La Figure 77 présente les biplots selon les axes F1 et F2 mais aussi F1 et F3. Sur cette figure, les signaux en couleur correspondent aux signaux pertinents retenus par chacune des deux approches ainsi que ceux communs aux deux approches.

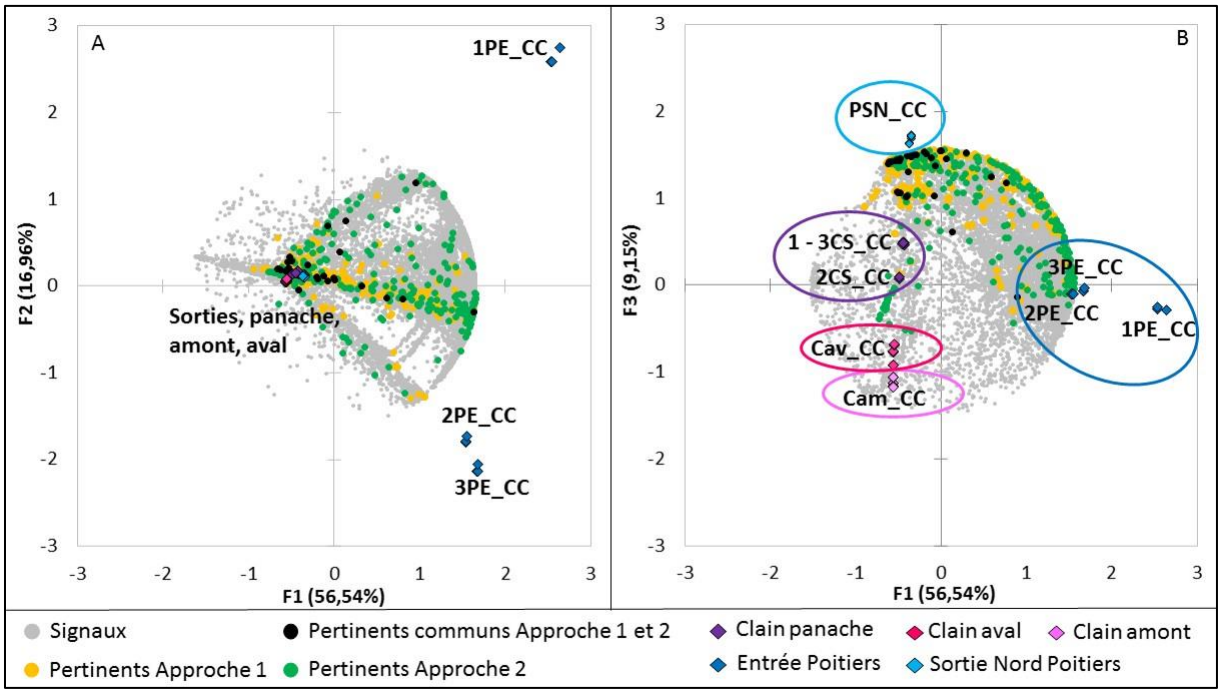


Figure 77 : Biplot des points de prélèvements de la campagne C et des signaux détectés selon les axes F1 et F2 (A) et selon les axes F1 et F3 (B). PE : entrée de STEP, PSN : sortie Nord de STEP, Cam : Clain Amont, Cav : Clain aval, CS : panache du Clain.

Cette figure montre une séparation nette des entrées et des sorties de STEP. Une séparation des échantillons de sortie, de panache et d'amont-aval est également visible dans le plan F1-F3. L'axe F1 sépare les échantillons d'eaux usées dites « sales » des autres échantillons d'eaux dites « propres » tandis que l'axe F3 sépare les eaux « propres » entre elles (en partant des eaux de sortie de STEP aux eaux de rivières : panache du rejet, aval et amont du Clain). L'ACP permet donc de séparer les échantillons en fonction de leur nature, signifiant la présence de différences au sein des échantillons.

Ces résultats sont à relier avec les empreintes des chapitres précédents qui ont mis en lumière ces différences entre les prélèvements.

On retrouve sur le plan F1-F2 la corrélation d'un grand nombre de signaux avec les entrées de STEP tel que cela a été décrit auparavant. Dans le plan F1-F3 cette corrélation est toujours existante mais moins marquée. On peut voir également un plus grand nombre de signaux corrélés avec les sorties de STEP qu'avec les échantillons du Clain. Enfin, la mise en relief des signaux pertinents en fonction de l'approche utilisée montre la dispersion des signaux provenant d'une seule approche. La représentation dans le plan F1-F3 (Figure 77 B) permet cependant de souligner le peu de signaux pertinents corrélés avec les eaux de rivières ce qui est en accord avec la provenance de ces signaux dans l'aval *via* la STEP. Dans le cas des signaux communs aux deux approches, ces signaux sont regroupés vers les eaux traitées pour leur grande majorité (excepté 4 signaux). Et la séparation des échantillons par l'axe F3 permet plus particulièrement de les corréler aux sorties de STEP. Ce résultat est intéressant et va dans le sens de l'objectif initial de ce travail consistant à rechercher des traceurs de rejets de STEP. Cela appuie également l'intérêt de croiser les approches pour sélectionner des signaux pertinents.

V. Conclusion

La confrontation des listes de suspects et des métabolites suspects des échantillons de sortie de STEP et du Clain (amont et aval) a mis en évidence 65 suspects et 15 métabolites suspects communs à ces deux catégories d'échantillons. Cette confrontation a montré que les échantillons prélevés dans le Clain à l'aval du rejet possèdent plus de signaux en communs avec les sorties de STEP que ceux prélevés en amont. Une partie des suspects communs est à la fois présente en amont et en aval. Lors de l'absence de ces suspects en amont, une contamination principale du Clain *via* la STEP a été envisagée. La présence en amont pourrait s'expliquer par une contamination des eaux par d'autres sources. Enfin, les suspects présents en sortie de STEP, mais absents an aval pourraient provenir d'un effet de dilution du rejet des eaux de STEP dans le Clain.

Concernant la recherche des signaux pertinents, les croisements réalisés entre les listes de signaux obtenus par les deux approches et les signaux des échantillons de la STEP et de la rivière ont mis en évidence les limites de chacune de ces démarches. En effet, l'approche 1 du « traiteur d'eau » ne prend pas en compte l'effet de dilution des rejets d'eaux traitées dans la rivière et donc leur non-détection possible en aval. L'approche 2 de « l'environnementaliste » peut quant à elle conduire à retenir des signaux issus d'autres sources de contamination ou de la dégradation de composés parents présents en amont.

Il a été montré que la présence des signaux pertinents de l'approche 1 (« traiteur d'eau ») en aval, comme la présence de ceux de l'approche 2 (« environnementaliste ») dans les sorties de STEP, permet d'appuyer en partie la pertinence de ces signaux. Toutefois, cela n'exclut pas complétement la possibilité de retenir un signal qui ne serait pas réellement pertinent. Un signal sélectionné par l'approche 1 qui est présent en aval peut en effet ne pas être pertinent s'il est aussi détecté en amont avec une intensité égale voire supérieure à celle observée en aval.

Enfin, il a pu être observé que la présence des signaux dans les listes obtenues par les 2 approches permet de réduire les biais sur la pertinence du signal.

Ainsi, lors de la recherche de signaux pertinents, un même niveau de confiance ne peut pas être attribué à tous les signaux sélectionnés suivant l'approche utilisée et les niveaux de confirmation obtenus (i.e. s'ils ont été sélectionnés avec une ou deux approches, ou s'ils ont été retrouvés en aval ou en sortie de STEP). Grâce à ces résultats, une hiérarchisation de la pertinence des signaux pourrait donc être envisagée suivant le nombre d'éléments de confirmation utilisés voire validés lors de la recherche de signaux pertinents : nombre d'approches de sélection, présence en aval ou en sorties de STEP.

Plusieurs niveaux de « pertinence » pourraient être proposés dans cette classification :

- les signaux sélectionnés par une seule approche sans confirmation supplémentaire pourraient être considérés comme « pertinents possibles »,
- les signaux sélectionnés par une seule approche mais également présents en aval ou en sorties de STEP pourraient être considérés comme « pertinents probables »,
- les signaux pertinents validés par les deux approches pourraient être considérés comme des « pertinents très probables ».

Cette classification pourrait permettre de garder un regard critique sur les signaux pertinents d'une liste établie dans le cadre d'un travail, en affichant un niveau de « pertinence » pour chaque signal de la liste.

Dans le cadre de ce travail de thèse, à partir des deux approches utilisées, 36 signaux ont été classés « très probables », 318 (218 approche 1 et 100 approche 2) comme « probables » et enfin 290 (121 approche 1, 169 approche 2) ont été classés comme « possibles ». Les signaux « très probables » et « probables » sont présentés en Annexe 15.

Enfin, la réalisation d'une ACP sur l'ensemble des échantillons d'une campagne a montré la séparation des échantillons d'eaux usées et des autres types d'eaux dits « propres » (rivières et sorties de STEP) dans le plan F1-F2 mais également la séparation des eaux « propres » entre elles dans le plan F1-F3 (en partant des eaux de sortie de STEP aux eaux de rivière : panache du rejet, aval et amont du Clain). La mise en évidence des signaux pertinents sur les biplots a permis de montrer que les signaux pertinents présents dans une seule approche sont corrélés avec les eaux de STEP (entrées et les sorties) tandis que ceux communs aux deux approches sont plutôt corrélés avec les sorties de STEP. Ce résultat est intéressant et va dans le sens de l'objectif initial de ce travail consistant à rechercher des traceurs de rejets de STEP.

Conclusions et perspectives

La grande diversité des micropolluants contaminant les eaux a complexifié la problématique de la protection des milieux aquatiques. Les rejets de STEP sont une source de contamination de ces milieux notamment car les procédés d'épuration n'ont pas été initialement conçus pour traiter ce type de pollution. À ce jour, la réglementation reste limitée à la surveillance, l'identification des sources et la réduction des émissions pour un nombre limité de micropolluants. De plus, des micropolluants encore inconnus, par exemple issus de transformations chimiques ou biologiques, peuvent avoir eux aussi des effets négatifs mais ne pas être détectés par les méthodes de routines. Pour tenter de faire face à ces problèmes et étendre le champ des connaissances à propos de cette pollution émergente, l'utilisation de la méthode d'analyse non ciblée a progressé dans les laboratoires de recherche. L'analyse non ciblée s'est développée dans le contexte environnemental grâce à l'essor de la spectrométrie de masse haute résolution et aux bases de données. Elle s'avère être un outil pour la caractérisation et la comparaison d'échantillons, la recherche et l'identification de composés.

Au cours de ces travaux de thèse, le développement d'une méthode d'analyse non ciblée a été réalisé et appliqué principalement aux eaux de la STEP de Poitiers et à celles de la rivière Clain. Les objectifs ont ainsi été de :

- développer la méthode, comprenant les étapes de préparation des échantillons et de traitement des données,
- caractériser les échantillons d'entrées et de sorties de STEP, au travers de leurs empreintes UHPLC-HRMS (profils moléculaires), et de les comparer pour évaluer l'impact l'épuration sur les eaux puis rechercher des signaux pertinents marqueurs de ce traitement,
- étudier les changements observés dans le profil moléculaire des échantillons de rivière au niveau d'une zone de rejets d'eaux traitées et rechercher des signaux marqueurs de ce rejet (signaux considérés comme pertinents).

Pour l'étape de prétraitement des données de l'analyse non ciblée, deux méthodes ont été développées utilisant soit MZmine 2, soit Compound Discoverer. L'optimisation de chaque paramètre a été effectuée, à l'aide d'échantillons de chaque matrice et d'eau ultra-pure dopés par un mélange d'étalons dont les concentrations se sont échelonnées de 50 ng/L à 1 µg/L, dans l'objectif de minimiser tout à la fois le nombres de faux négatifs et de faux positifs. Sur l'ensemble des paramètres réglés, les tolérances sur les temps de rétention et sur le rapport m/z (ou les masses moléculaires) sont apparues comme les paramètres les plus critiques. Après développement et optimisation, les deux méthodes ont été comparées. Les caractéristiques des signaux en termes de temps de rétention et de masses moléculaires ont présenté des valeurs médianes plus faibles avec MZmine 2 qu'avec Compound

Discoverer. En revanche, les aires des signaux se sont montrées équivalentes (avec cependant une dispersion plus faible avec MZmine 2) et une grande partie des signaux les plus intenses est apparue commune aux deux méthodes. De plus, malgré ces différences, les analyses en composantes principales effectuées séparément sur les données traitées par les deux méthodes conduisent à des résultats très similaires. Ainsi, la comparaison des résultats des deux logiciels a montré que le choix d'un logiciel par rapport à un autre peut avoir des conséquences sur la nature des signaux détectés même si les résultats des deux outils semblent permettre d'apporter des conclusions similaires. Pour la suite des travaux, il a été décidé d'utiliser Compound Discoverer en raison de ses performances satisfaisantes, de sa facilité d'utilisation, de son accès direct à mzCloud et à d'autres bases de données, du couplage possible entre analyse ciblée, analyse de suspects et analyse non ciblée, et enfin du temps de traitement des données comparativement plus rapide même avec un grand nombre d'échantillons. Pour réduire le nombre de faux positifs, l'utilisation de réplicas analytiques s'est avérée efficace et un nombre de trois réplicas avec un coefficient de variation inférieur à 30% entre les trois répliquas a été considéré pour retenir un signal. Enfin, l'impact de l'extraction des échantillons par SPE s'est révélé positif vis-à-vis des objectifs de l'analyse non ciblée (avec dans l'ensemble un fort gain du nombre de signaux détectés et un nombre limité de signaux ayant disparu ou diminué). Pour corriger les effets matrices, un facteur de correction voisin de 2 a été déterminé entre les échantillons d'entrées et de sorties de STEP. L'ensemble de ce développement a donc conduit à l'obtention d'une méthode d'analyse non ciblée mettant en œuvre une étape d'extraction/concentration préalable puis l'utilisation du logiciel Compound Discoverer pour le traitement des données.

La seconde partie des résultats a montré que le traitement d'épuration des eaux conduit à l'abattement et à l'élimination d'une majorité de signaux. Toutefois, certains signaux semblent réfractaires au traitement et de nouveaux signaux dont l'impact est inconnu apparaissent également. L'impact du traitement des eaux sur la composition des échantillons est visible par analyse en composantes principales. Les résultats de l'ACP ont montré le regroupement des échantillons de même type (entrées ou sorties de STEP). De même, il a été observé que les signaux diminuants ou disparaissant sont corrélés avec les entrées et les signaux augmentant ou apparaissant sont corrélés avec les sorties. Ainsi, une corrélation des signaux avec les échantillons, qui peut s'expliquer par le statut (présence ou absence) du signal dans l'échantillon ou par la différence de valeur d'aire du signal entre entrées et sorties, a été notée. Afin de rechercher des signaux pertinents marqueurs de l'épuration et pouvant impacter le milieu récepteur (i.e. formés au cours du procédé ou réfractaires au procédé), une analyse différentielle a été réalisée. De cette manière, 375 signaux ont été sélectionnés. Pour 331 d'entre eux, une formule brute a été attribuée et les rapports O/C, N/C et H/C ont été calculés montrant une diversité de structures possibles. Ces signaux pourraient ainsi

correspondre à des constituants de la matière organique naturelle (lipides, acides aminées, etc…) ou correspondre à des molécules anthropiques. Parmi ces signaux, 25 (dont 7 produits pharmaceutiques) ont été associés à une formule développée. Une recherche de suspects a également été réalisée et 198 suspects ont été mis en évidence (dont la caféine, l'oxybenzone et six produits pharmaceutiques confirmés par l'injection d'un étalon). Dans le cas des métabolites, dont la structure a été simulée *in sillico*, la présence de plusieurs de ces composés liés à la caféine, l'aténolol, la venlafaxine, la cocaïne, le sulfamethoxazole et le 4'-Aminoacetanilide a été suspectée. Pour certains d'entre eux, les molécules parentes sont également susceptibles d'être présentes dans les échantillons. La présence possible de ces molécules dans les échantillons de STEP (qui reste à confirmer par l'injection d'étalons) souligne que la contamination des eaux par les micropolluants est une problématique d'importance dans l'amélioration du traitement des eaux usées.

Lors de l'étude de la variabilité du Clain en zone anthropisée, la comparaison des profils moléculaires a montré une première diminution des valeurs d'aires des signaux entre les sorties de STEP et le panache, puis une seconde diminution entre le panache et l'aval au rejet. Toutefois, entre les différents types d'échantillons, une stagnation des aires de certains signaux a également été observée. Enfin, si le nombre de signaux détectés entre la sortie de la STEP jusqu'à l'aval a diminué, un nombre de signaux environ deux fois plus élevé qu'en amont qu'en aval est observé. Aussi, si les effets de dilution ont permis d'atténuer l'effet de la STEP, un retour à l'état initial de l'amont du Clain n'a pas été observé en aval immédiat sur les trois campagnes de prélèvements. Par conséquent, une recherche de signaux pertinents, traceurs de l'impact des rejets de la STEP de Poitiers, a été effectuée entre l'amont et l'aval par analyses différentielles et 305 signaux ont été sélectionnés. Parmi ces signaux, 15 ont une ou plusieurs propositions de structures. Enfin, la réalisation d'analyse de suspects a montré la contamination possible des échantillons du Clain par 127 suspects et 34 métabolites suspects. Ces résultats constituent une indication de la présence de micropolluants dans le Clain.

Dans le dernier chapitre de ce manuscrit, les listes de suspects et de métabolites suspects ont été croisées entre elles, permettant la mise en évidence de 65 suspects et 15 métabolites suspects communs entre les échantillons de la sortie de STEP et ceux du Clain (amont et aval). La présence de certains suspects uniquement en aval et dans les sorties de STEP indique le rôle de la station d'épuration dans la dissémination de micropolluants dans le Clain. Une étude critique sur la recherche des signaux pertinents traceurs du rejet a été réalisée pour évaluer les limites de chaque approche utilisée (i.e. approche 1 « traiteur d'eau » et approche 2 « environnementaliste »). Pour cela, les signaux pertinents issus de la STEP (approche 1 « traiteur d'eau ») et ceux du Clain (approche 2 « environnementaliste ») ont été croisés. Puis les signaux de l'approche 1 ont été recherchés dans les échantillons du Clain tandis que ceux de l'approche 2 ont été recherchés dans les eaux de STEP. Ainsi,

36 signaux communs aux deux listes ont été mis en évidence, 241 signaux de l'approche 1 ont été détectés dans le Clain et enfin 136 signaux de l'approche 2 ont été détectés en sortie de STEP. Il a été constaté, à partir de ce travail, qu'un même niveau de confiance ne pouvait pas être attribué à tous les signaux sélectionnés suivant l'approche utilisée et les niveaux de confirmation obtenus (i.e. s'ils ont été sélectionnés avec une ou deux approches, ou s'ils ont été retrouvés en aval ou en sortie de STEP). Grâce à ces résultats, une hiérarchisation des signaux en trois catégories a donc été établie : « pertinents possibles », « pertinents probables » et « pertinents très probables » suivant si la présence du signal considéré comme pertinent a été mis en évidence par une seule approche, une approche mais est également présent en sortie de STEP (pour l'approche 1) ou dans l'aval du Clain (pour l'approche 2) ou par les deux approches. Cette proposition de hiérarchisation permet d'instaurer un premier niveau de confiance quant à la pertinence des signaux sélectionnés. Ainsi, parmi les signaux sélectionnés, 36 seraient classés comme de « très probables » signaux pertinents, 318 comme « probables » (218 approche 1 et 100 approche 2) et enfin 290 (121 approche 1, 169 approche 2) comme « possibles ». Le classement de signaux comme de « très probables » et « probables » signaux pertinents conforte la démarche développée au cours des travaux de thèse et permet donc de proposer un premier ensemble de signaux à suivre comme traceurs de rejets de STEP.

La partie prétraitement des données est une partie essentielle de l'analyse non ciblée impliquant l'utilisation de logiciels tiers. Cette étape est à l'origine d'une partie des biais que comporte l'analyse non ciblée par la génération de faux positifs ou de faux négatifs. Comme montré lors de la comparaison des logiciels utilisés, le choix du logiciel peut impacter le nombre et les signaux détectés. Des tentatives d'harmonisation des étapes, des paramètres de réglage utilisés et du niveau de confiance des résultats obtenus sont en cours d'élaboration par différentes équipes de recherche moteurs de l'analyse non ciblée environnementale et il sera nécessaire de réévaluer et modifier la méthode avec les progrès du domaine. De manière générale, les questions de la robustesse, de la justesse et de la validation des méthodes non ciblées sont encore en cours de réflexion dans les laboratoires de recherche. Une publication de Schulze *et al.*, parue en décembre 2020, propose une évaluation des méthodes de contrôle qualité et des recommandations qu'il serait intéressant d'étudier pour renforcer la méthode développée. À plus long terme, l'utilisation de façon routinière d'une méthode « all-in-one » couplant l'analyse ciblée, l'analyse de suspects et l'analyse non ciblées serait un objectif à atteindre. Une telle méthode combinerait ainsi le suivi des molécules d'intérêts et l'exploration des signaux détectés pour : (1) mettre en évidence des signaux pertinents, (2) détecter des pollutions accidentelles, et (3) rechercher les sources de contamination.

La méthode développée lors des présents travaux de thèse s'est appuyée sur l'utilisation d'une colonne C18 avec une source d'ionisation HESI. L'emploi d'une autre colonne chromatographique ou d'une autre source d'ionisation pourrait permettre d'étendre la gamme des molécules analysées, par exemple vers des composés de polarités différentes. La chromatographie à interaction hydrophile (HILIC) est une des options de séparation possible et l'APCI (Atmospheric Pressure Chemical Ionization) est une autre source d'ionisation encore peu employée. Ainsi des méthodes complémentaires ou orthogonales pourraient être développées. Dans le même esprit, lors de la préparation des échantillons, d'autres phases stationnaires pourraient être testées afin d'élargir le champ des molécules extraites. L'utilisation de plusieurs phases stationnaires différentes au sein d'une même cartouche serait un moyen d'augmenter les rendements d'extractions et la robustesse de la méthode. Les échantillonneurs passifs seraient quant à eux un moyen d'augmenter la sensibilité de la méthode dans les rivières par la rétention *in situ* des contaminants

Les travaux menés ont montré que, si la station d'épuration de Poitiers remplit son rôle d'un point de vue global et règlementaire, cependant une fraction des composés associés au profil moléculaire des eaux d'entrée est peu ou pas abattue. La présence de signaux augmentant ou apparaissant peut être un témoin des transformations s'opérant dans les STEP. La présence de potentiels métabolites dans les signaux sélectionnés et les résultats des analyses de suspects appuient l'intérêt d'étudier ce type de signaux. La réalisation d'études de plus grande ampleur en termes d'échantillons pourrait permettre d'isoler ces composés diffusés dans le milieu aquatique de façon répétée.

L'étude du Clain au niveau de la zone de rejet des eaux traitées de la station d'épuration de Poitiers a montré un impact du rejet sur le profil moléculaire de la rivière. Un effet de dilution immédiatement en aval atténue cet impact mais le Clain n'y est pas restauré à son état initial amont. De façon à compléter ce travail de thèse, des campagnes de plus grandes ampleurs sur un ensemble de points le long du Clain en amont et aval, proche et éloigné, de zones de rejets seraient à réaliser. Des prélèvements plus éloignés des zones de rejets permettraient en effet d'établir si le Clain retourne à un état proche de l'amont au fur et à mesure de l'éloignement de ces zones. En multipliant les points de prélèvements le long du Clain un autre objectif serait d'établir une cartographie des sources de contaminations de la rivière.

Si les suspects détectés sont confirmés, leur nature montre l'importance de la problématique de certains micropolluants tels que les produits pharmaceutiques dans le traitement des eaux et la nécessité d'obtenir plus d'informations quant aux mécanismes de transformation de ces molécules ainsi qu'à leurs possibles toxicité et écotoxicité.

La recherche des signaux pertinents traceurs de rejets de STEP s'est appuyée sur les variations d'aires et la répétition de détection. Cette recherche à partir des profils moléculaires des eaux de la STEP et de ceux d'échantillons d'eau prélevés dans le Clain à l'amont et à l'aval du rejet a permis de sélectionner plusieurs centaines de signaux dont une fraction d'entre eux ont reçu une proposition d'identification qui reste à confirmer. Ces échantillons et les listes obtenues possèdent des signaux communs qui confortent la capacité de la méthode utilisée. Toutefois, cette méthode n'est pas exempte de biais en raison des variations dues aux effets matrices ou encore à des faux positifs ou négatifs liés au traitement des données ou aux blancs. D'autre part, les signaux pertinents sélectionnés dans cette étude sont le témoin d'une pollution diffuse dans les eaux du Clain. Néanmoins, les travaux effectués ne renseignent pas sur l'activité biologique de ces composés et sur l'impact écotoxique du rejet des eaux. Réaliser un couplage entre des tests de toxicité et la méthode d'analyse non ciblée permettrait de sélectionner des signaux correspondant à des molécules présentant une toxicité significative et pourrait également déterminer si les signaux déjà sélectionnés correspondent à des composés toxiques. Ce couplage peut se faire sous la forme d'une analyse dirigée par les effets (EDA) ou une « virtual EDA » reposant en partie sur les statistiques. Aussi, des données de toxicité pourraient constituer un élément supplémentaire dans la hiérarchisation des signaux pertinents.

Enfin, la proposition de hiérarchisation est un moyen d'introduire des niveaux de pertinence et de confiance dans la sélection de traceurs de STEP. Toutefois, cette classification doit être consolidée d'une part par la réalisation de campagnes de prélèvements plus longues, avec plus d'échantillons, et d'autre part avec l'extension de l'approche à d'autres stations d'épuration. La répétition de la détection des signaux pourrait ainsi renforcer ces niveaux de confiance. De plus, la création d'une base de données interne de spectres de fragmentation, sans nécessairement aller jusqu'à l'identification, serait un plus pour le suivi de ces signaux. L'ensemble de ces démarches pourrait permettre, à terme, de compléter les normes actuelles de surveillance de la qualité des rejets par la mise en œuvre d'un suivi des profils moléculaires et notamment celui des nouveaux traceurs de rejets de STEP ainsi identifiés selon la méthode développée au cours de cette thèse.

Annexes

Annexe 1

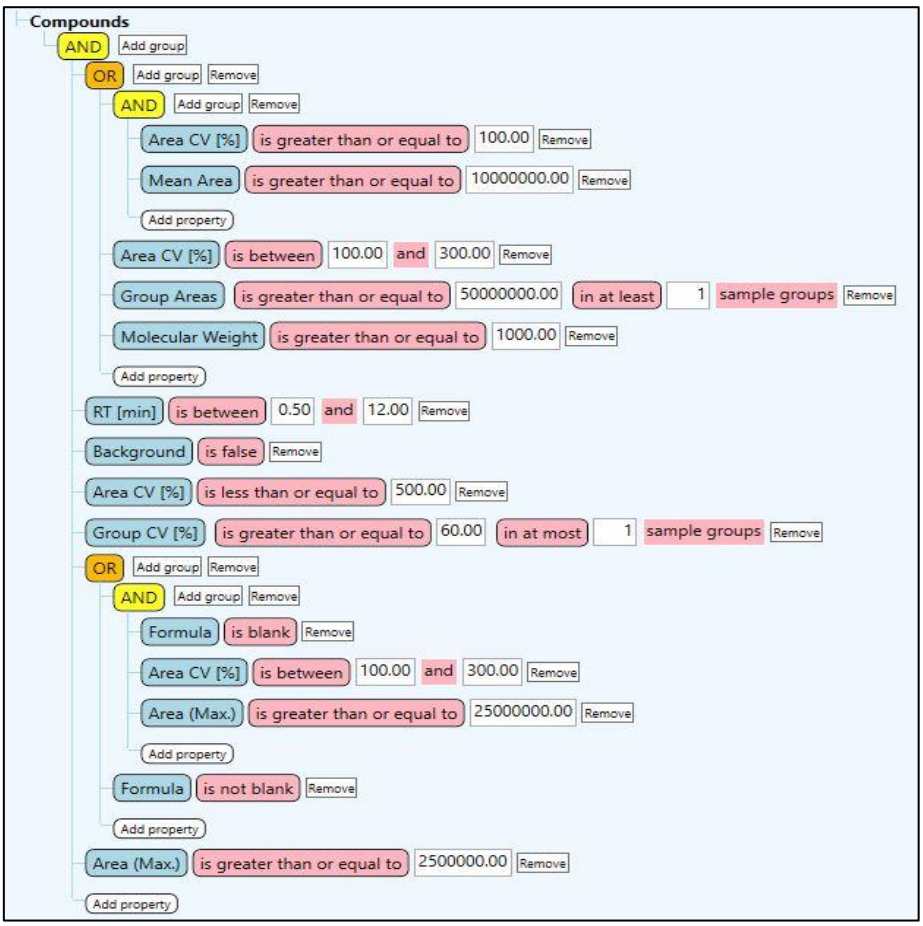


Figure 78 : Liste des filtres appliqués sur le tableau de données dans Compound Discoverer lors de l'étude d'échantillons de STEP

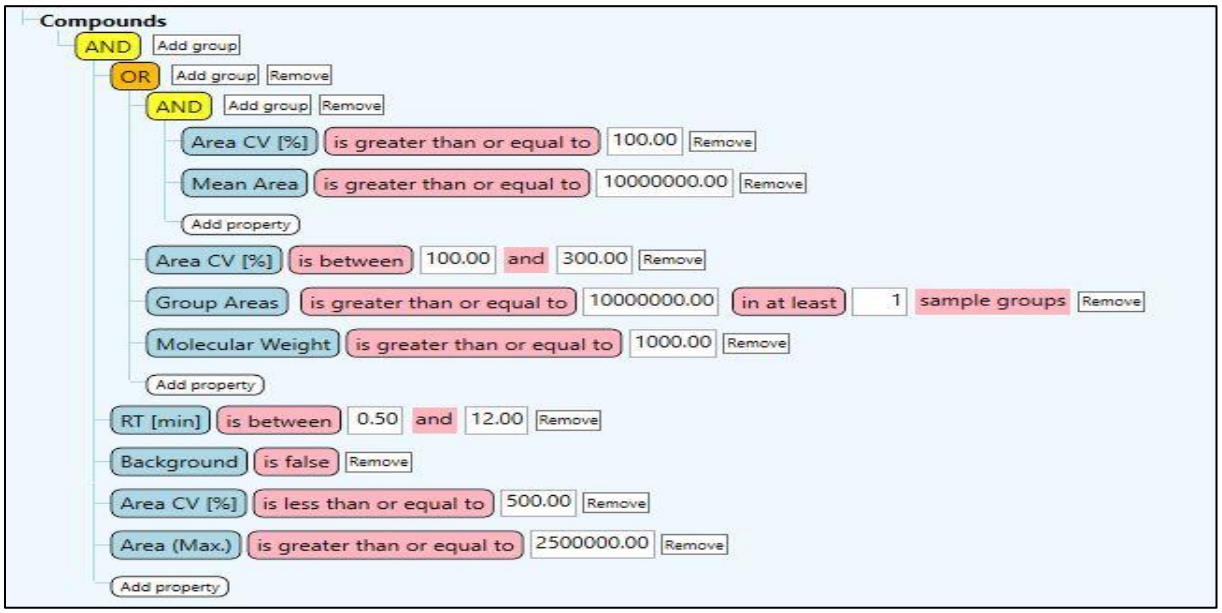


Figure 79 : Liste des filtres appliqués sur le tableau de données dans Compound Discoverer lors de l'étude d'échantillons de rivières

Annexe 2

Tableau 29 : Nombre de signaux par tendance calculé entre les échantillons non concentrés et les échantillons concentrés par extraction sur phase solide.

Tendance	Entrée	Sortie	Clain Amont
Apparition	18 098 ± 1943	4 835 ± 122	2 158 ± 165
Augmentation	10 557 ± 360	1 026 ± 10	272 ± 28
Stable	2 353 ± 73	300 ± 17	145 ± 8
Diminution	1 145 ± 81	223 ± 5	192 ± 13
Disparition	2 259 ± 13	982 ± 9	1 147 ± 17

Annexe 3

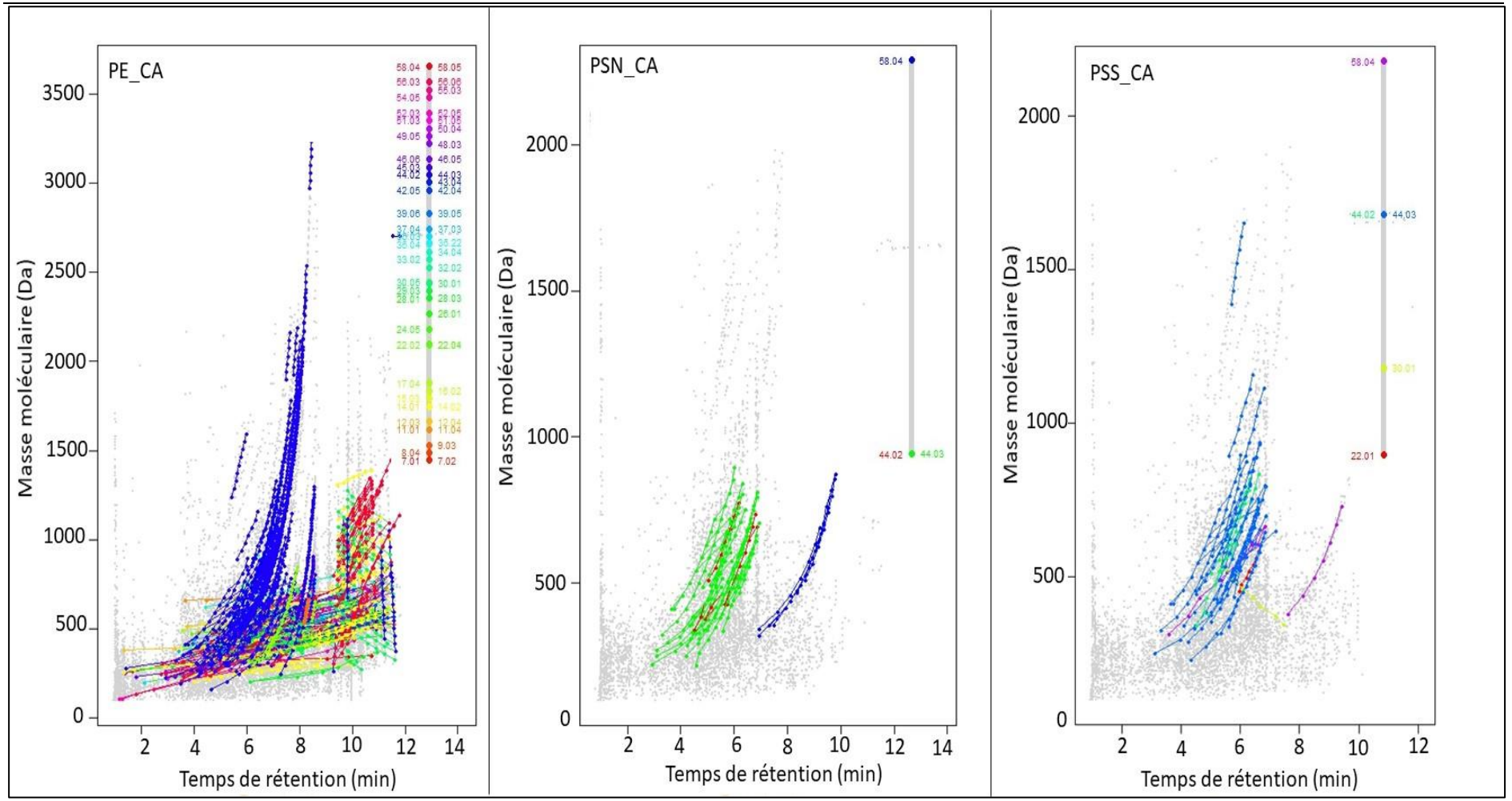


Figure 80 : Empreintes des échantillons d'entrée et de sorties de la campagne A avec relief des homologues recherchés par la package « non-target ». PE : entrée, PSN : entrée Nord, PSS : Sortie sud CA : campagne A. Mode positif (HESI+).

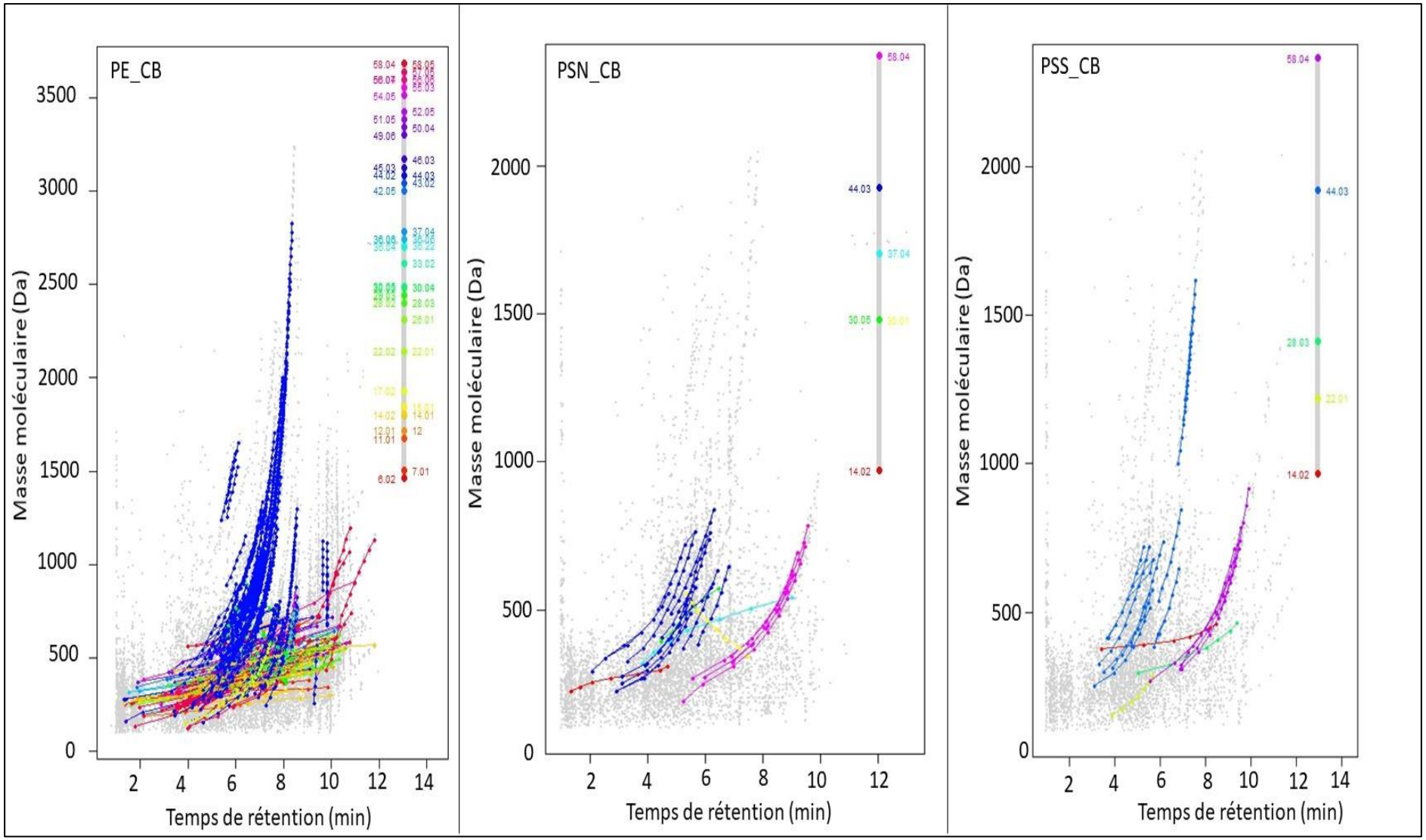


Figure 81 : Empreintes des échantillons d'entrée et de sorties de la campagne B avec relief des homologues recherchés par la package « non-target ». PE : entrée, PSN : entrée Nord, PSS : Sortie sud CB : campagne B. Mode positif (HESI+).

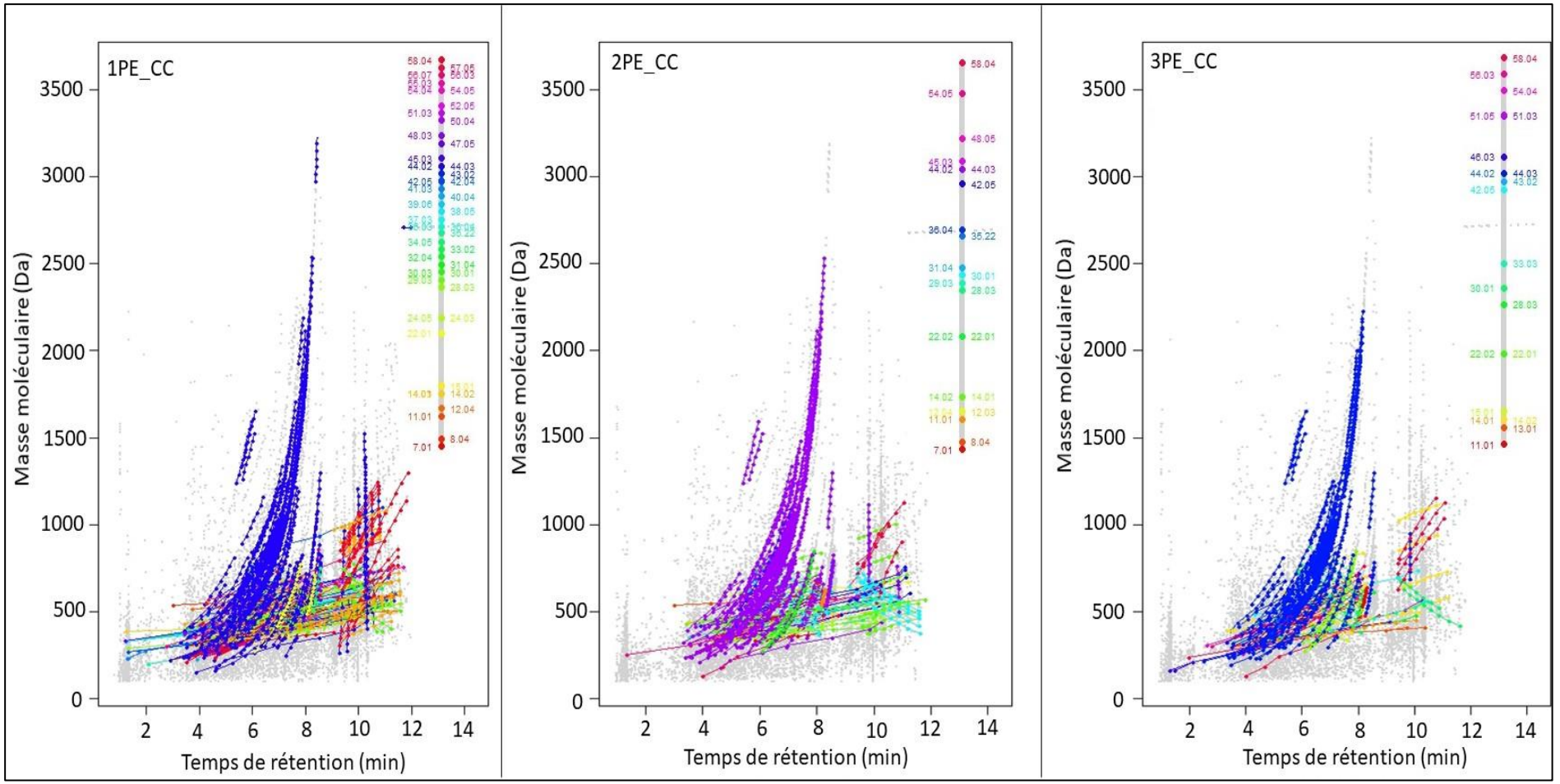


Figure 82 : Empreintes des 3 échantillons d'entrées de la campagne C avec relief des homologues recherchés par la package « non-target ». PE : entrée Poitiers, CC : campagne C. Mode positif (HESI+).

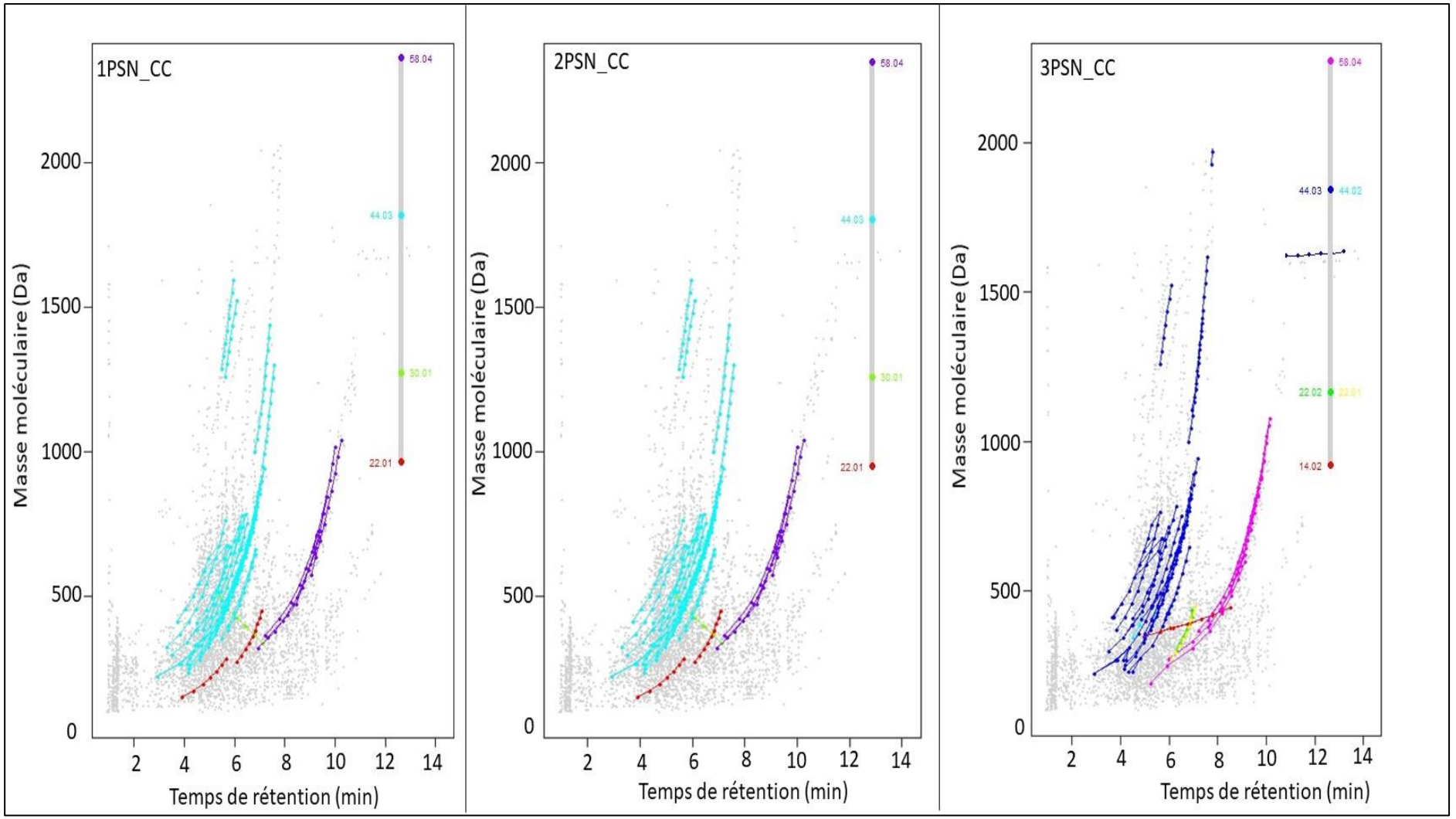


Figure 83 : Empreintes des 3 échantillons de sorties de la campagne C avec relief des homologues recherchés par la package « non-target ». PSN : entrée nord de Poitiers, CC : campagne C. Mode positif (HESI+).

Annexe 4

Tableau 30 : Liste des propositions de structures attribuées à 25 signaux pertinents par l'utilisation de mzCloud. Tr : temps de rétention, MC : pourcentage de correspondance entre le spectre MS^2 expérimental et celui de mzCloud.

Proposition(s) de structure(s)	Formule Brute	Masse exacte	Tr (min)	MC (%)
5-amino-o-crésol	C_7H_9NO	123,0685	1,52	83,9
2-amino-4-crésol				82,5
2-anisidine				76,9
2-pyridine ethanol				74,9
6-methyl-2-pyridinemethanol				73,7
Tropine	$C_8H_{15}NO$	141,1153	6,77	77,9
3,3,5,5-Tetramethylpyrroline-N-oxide				77,2
N-Cyanoimido-S,S-dimethyl-dithiocarbonate	$C_4H_6N_2S_2$	145,9972	5,10	61,8
NP-007065	$C_8H_{10}O_3$	154,0630	4,85	67,2
2,6-Dimethoxyphenol				65,2
2-(Cyclopropylcarbonyl)-3-(dimethylamino)acrylonitrile	$C_9H_{12}N_2O$	164,0949	1,64	68,2
Edrophonium	$C_{10}H_{15}NO$	165,1154	8,28	77,0
Perillartine				73,2
(4^E)-3-Hydroxy-2,4-dimethyl-4-heptenamide	$C_9H_{17}NO_2$	171,1268	6,23	77,7
α-Methyltryptamine	$C_{12}H_{13}NO_2$	174,1161	4,53	71,5
1(1H-indol-5-yl)-2-propanamine				65,9
1(1H-indol-6-yl)-2-propanamine				61,7
2-Amino-6-methylmercaptopurine	$C_6H_7N_5S$	181,0425	3,62	79,9
2-(methylthio)[1,2,4]triazolo[1,5-a]pyrimidin-7-amine				66,3
4-Methyl-N,N-dimethylcathinone	$C_{12}H_{17}NO$	191,1310	7,87	61,9
5-Phenylnicotinic acid	$C_{12}H_9NO_2$	199,0634	2,50	78,6
Indole-3-butyric acid	$C_{12}H_{13}NO_2$	203,0949	4,78	76,0
3-ethyl-4-hydroxy-1-methyl-1,2-dihydroquinolin-2-one				71,6
N1-(5-oxo-5,6,7,8-tetrahydronaphthalen-2-yl)acetamide				65,0
Methsuximide				64,4
4-morpholinobenzoic acid	$C_{11}H_{13}NO_3$	207,0930	2,46	66,6
Tapentadol	$C_{14}H_{23}NO$	221,1773	8,78	62,4
Desmethyl nefopam	$C_{16}H_{17}NO$	239,1307	5,90	85,2
Alprenolol	$C_{15}H_{23}NO_2$	249,1741	4,36	70,1
O-desmethylvenlafaxine	$C_{16}H_{25}NO_2$	263,1881	5,06	83,2
Métabolite du Valsartan	$C_{14}H_{10}N_4O_2$	266,0800	6,8	86,6

Proposition(s) de structure(s)	Formule Brute	Masse exacte	Tr (min)	MC (%)
Galaxolidone	$C_{18}H_{24}O_2$	272,1767	8,87	95,9
β-Estradiol				76 ,4
19-Norandrostenedione			9,23	64,7
Codeine	$C_{18}H_{21}NO_3$	299,1524	2,95	78,7
Hydrocodone				78,8
Flubendazole	$C_{16}H_{12}FN_3O_3$	313,0876	7,54	66,5
N,N-Dimethyl-4-{5-[(3S)-1-(phenylsulfonyl)-3-pyrrolidinyl]-1,3,4-oxadiazol-2-yl}aniline	$C_{20}H_{22}N_4O_3S$	398,1416	4,42	71,9
AB-FUBINACA 2A metabolite	$C_{20}H_{19}FN_4O_4$			65,9
N'',N'''-di[2-chloro-4-(dimethylamino)benzylidene]carbonic dihydrazide	$C_{19}H_{22}Cl_2N_6O$	420,1237	4,39	67,6
Candesartan	$C_{24}H_{20}N_6O_3$	440,1586	7,92	83,7

Annexe 5

Tableau 31 : Liste des 198 signaux correspondant à un ou plusieurs suspects avec leurs formules brutes, la masse exacte et le temps de rétention (Tr) du signal détecté, le pourcentage de la meilleure correspondance avec mzCloud, les logP et logD (pH = 3), la nature de l'échantillon dans lequel le suspect est présent, le pourcentage de correspondance entre les spectres de MS^2 dans la base de données considérée (MC %) et un code couleur correspondant à : une correspondance avec mzCloud supérieure à 80% en bleu foncé et inférieure à 80% en bleu clair, une correspondance avec les bases de données de MetFrag (KEGG, CompTox et HMDB) supérieure à 50% en vert foncé et inférieure à 50% en vert clair, pas de correspondance avec une base de données en jaune, et la confirmation du suspect en rouge. Les signaux en gras sont également des signaux de la liste des signaux pertinents. E : entrée STEP, S : sortie STEP, C : code couleur, MC : pourcentage de correspondance entre le spectre MS^2 expérimental et celui de mzCloud, Tr : Temps de rétention.

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Benzotriazole	$C_6H_5N_3$	119.0484	4,85	E/S	85,0	1,30	1,30	
Phenethylamine/ 2,6-Dimethylaniline	$C_8H_{11}N$	121,0892	2,78	E	87,9	1,39/2,17	-1,64/0,89	
Phenethylamine/ 2,6-Dimethylaniline	$C_8H_{11}N$	121,0893	1,71	E/S	87,9	1,39/2,17	-1,64/0,89	
Phenethylamine/ 2,6-Dimethylaniline	$C_8H_{11}N$	121,0894	0,96	E/S	78,8	1,39/2,17	-1,64/0,89	
Nicotinamide	$C_6H_6N_2O$	122,0481	0,96	E/S	62,8	-0,37	-0,37	
Chlormequat	$C_5H_{13}ClN$	122,0739	1,11	E/S	9,6	-3,31	-3,31	
Acide nicotinique	$C_6H_5NO_2$	123,0322	0,95	E/S	75,5	-0,17	-0,17	
Melamine	$C_3H_6N_6$	126,0656	0,76	E/S	81,4	-0,60	-2,81	
2-Aminobenzimidazol/ 5-Methyl-1H-benzotriazole/ 4-Methylbenzotriazole	$C_7H_7N_3$	133,0640	6,07	E/S	92,3	1,11/1,82/ 1,82	-0,32/1,81/1,81	
4-Hydroxybenzotriazole	$C_6H_5N_3O$	135,0433	3,70	E/S		1,00	1,00	
Amphetamine	$C_9H_{13}N$	135,1047	2,26	E/S	80,6	1,80	1,80	
Amphetamine	$C_9H_{13}N$	135,1055	2,04	E	79,3	1,80	1,80	
N-Methylnicotinamide	$C_7H_8N_2O$	136,0636	0,70	E	89,4	-0,17	-0,73	
N-Methylnicotinamide	$C_7H_8N_2O$	136,0637	1,49	E		-0,17	-0,73	
Dimetridazole	$C_5H_7N_3O_2$	141,0541	3,15	E/S	65,6	0,23	0,01	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Piracetam	$C_6H_{10}N_2O_2$	142,0743	0,80	E/S		-1,69	-1,69	
2-Naphthylamine	C_10H_9N	143,0734	4,37	E/S	87,7	2,13	0,90	
8-Hydroxyquinoline	C_9H_7NO	145,0526	5,63	E	82,3	1,83	0,60	
8-Hydroxyquinoline	C_9H_7NO	145,0527	5,08	E/S	82,3	1,83	0,60	
8-Hydroxyquinoline	C_9H_7NO	145,0527	4,16	E/S	82,3	1,83	0,60	
N-methylacetanilide	$C_9H_{11}NO$	149,0843	0,90	E/S	68,6	1,08	1,08	
4`-Aminoacetanilide	$C_8H1_0N_2O$	150,0794	0,98	E/S		0,38	-0,67	
Paracetamol	$C_8H_9NO_2$	151,0633	3,28	E/S	86,7	0,91	0,91	
Norephedrine	$C_9H_{13}NO$	151,1000	1,35	E	65,9	0,89	-2,15	
Norephedrine	$C_9H_{13}NO$	151,1000	0,82	E/S	52,9	0,89	-2,15	
Acide aminosalicylique/ Mesalazine	$C_7H_7NO_3$	153,0427	1,48	E/S	80	0,82/-0,29	0,83/-032	
Dopamine	$C_8H_{11}NO_2$	153,0793	1,31	E/S	62,2	0,03	-2,25	
Dopamine	$C_8H_{11}NO_2$	153,0793	0,80	E/S	48,6	0,03	-2,25	
2',3'-Bipyridine	$C_8H_{10}N_2$	156,0688	3,27	E/S		1,57	0,10	
Pregabalin	$C_8H_{17}NO_2$	159,1259	0,77	E/S		-1,35	-2,08	
Nicotine	$C_{10}H_{14}N_2$	162,1160	0,92	E	85,9	1,17	1,17	
Benzocaine	$C_9H_{11}NO_2$	165,0794	2,53	E	54,2	1,51	1,30	
Pseudoephedrine/ Ephedrine	$C_{10}H_{15}NO$	165,1153	3,85	E/S	43,2	1,32	-1,92	
Pseudoephedrine/ Ephedrine	$C_{10}H_{15}NO$	165,1156	3,45	E/S	83	*1,32*	*-1,92*	
Ethylparabène	$C_9H_{10}O_3$	166,0631	5,52	E	72,0	2,03	2,03	
Phenylephrine	$C_9H_{13}NO_2$	167,0946	0,76	E/S	61,6	-0,07	-2,65	
Levetiracetam	$C_8H_{14}N_2O_2$	170,1059	0,77	E/S	50,5	-0,59	-0,59	
Metronidazole	$C_6H_9N_3O_3$	171,0644	2,89	E/S	79,6	-0,46	-0,74	
Gabapentine	$C_9H_{17}NO_2$	171,1259	0,88	E/S	97,1	-1,27	-2,00	
Cotinine/ Nornicotine,N-formyl	$C_{10}H_{12}N_2O$	176,0953	1,12	E	94,3	0,21/0,34	-0,72/-0,59	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Phenacetine/ Acide fusarique	$C_{10}H_{13}NO_2$	179,0951	5,25	E/S	32,3	1,41/1,14	1,41/1,14	
Paraxanthine/Theophylline/ Theobromine	$C_7H_8N_4O_2$/$C_7H_8N_4O_2$/$C_7H_8N_4O_2$	180,0645	3,38	E/S	92,9	0,24/-0,77/-0,77	0,24/-0,77/-0,77	
Paraxanthine/Theophylline/ Theobromine	$C_7H_8N_4O_2$/$C_7H_8N_4O_2$/$C_7H_8N_4O_2$	180,0645	3,93	E/S	92,9	0,24/-0,77/-0,77	0,24/-0,77/-0,77	
Propylparabène	$C_{10}H_{12}O_3$	180,0785	6,26	E	71,2	2,55	2,55	
Phenazone/Antipyrine	$C_{11}H_{12}N_2O$	188,0955	5,02	S	85,9	1,21/1,21	1,21/1,21	
Propamocarb	$C_9H_{20}N_2O_2$	188,1523	0,82	E/S	93,4	0,77	-2,73	
DEET	$C_{12}H_{17}NO$	191,1306	7,68	E/S	96,4	2,50	2,50	
DMP/Mono-éthyl-phtalate	$C_{10}H_{10}O_4$	194,0584	4,15	E	71,7	1,98/1,99	1,98/1,73	
Caféine	$C_8H_{10}N_4O_2$	194,0800	4,65	E/S	98,5	-0,55	-0,55	
PEG EO4 (HO-PEG4-HO)	$C_8H_{18}O_5$	194,1150	2,11	E	9,10	-1,35	-1,35	
PEG EO4 (HO-PEG4-HO)	$C_8H_{18}O_5$	194,1153	3,76	E	12,8	-1,35	-1,35	
Acide salycilique	$C_{12}H_{16}O_3$	208,1092	8,77	E	72,4	4,04	3,54	
Aminocarb	$C_{11}H_{16}N_2O_2$	208,1210	3,76	S	75,1	2,38	0,53	
Aminocarb	$C_{11}H_{16}N_2O_2$	208,1210	4,12	E/S	72,2	2,38	0,53	
Aminocarb	$C_{11}H_{16}N_2O_2$	208,1222	4,33	S	37,5	2,38	0,53	
Atraton/ Propazine-2-hydroxy/ Terbutylazine-2-hydroxy	$C_9H_{17}N_5O$	211,1434	5,24	E/S	78,1	1,81/0,37/ 0,23	-0,12/-0,14/0,18	
Trans-Zeatin	$C_{10}H_{13}N_5O$	219,1118	1,02	E/S	90,3	-0,15	-1,18	
E/Smetaxalone/Carbofuran	$C_{12}H_{15}NO_3$	221,1051	6,46	E/S	70,4	2,37/2,05	2,37/2,05	
Acycloguanosine/Aciclovir	$C_8H_{11}N_5O_3$	225,0866	1,41	E/S	73,7	-1,03	-1,32	
Oxybenzone	$C_{14}H_{12}O_3$	228,0782	8,69	E/S	80	3,62	3,62	
4-Formylaminoantipyrine	$C_{12}H_{13}N_3O_2$	231,1011	4,26	E/S	77,6	0,11	0,11	
Diuron	$C_9H_{10}Cl_2N_2O$	232,0176	7,76	E/S	67,1	2,53	2,53	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Acide Nalidixique/ Fluometuron	$C_{12}H_{12}N_2O_3$/ $C_{10}H_{11}F_3N_2O$	232,0846	3,59	E	72,5	0,79/2,20	0,79/2,20	
Lidocaine	$C_{14}H_{22}N_2O$	234,1731	4,41	E/S	75,3	2,84	-0,64	
Carbamazepine	$C_{15}H_{12}N_2O$	236,0945	7,18	E/S	98,5	2,77	2,77	
Salbutamol	$C_{13}H_{21}NO_3$	239,1533	7,34	E	22,9	0,35	-2,36	
Terbutryne/Prometryne	$C_{10}H_{19}N_5S$	241,1361	7,59	E/S	84,8	2,88/3,01	0,92/1,06	
Biotine	$C_{10}H_{16}N_2O_3S$	244,0888	1,72	E/S	39,4	0,32	0,30	
Sulfapyridine	$C_{11}H_{11}N_3O_2S$	249,0575	3,69	E/S	73,9	1,01	0,95	
Alprenolol/ O-Desmethyl-cis-tramadol	$C_{15}H2_3NO_2$	249,1724	5,33	E/S		3,10/1,72	3,10/1,72	
Alprenolol/O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1726	4,16	E/S	73,8	3,10/1,72	3,10/1,72	
Alprenolol /O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1738	4,03	E/S	70,1	3,10/1,72	3,10/1,72	
Alprenolol/ O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1741	4,73	E/S	67,3	3,10/1,72	3,10/1,72	
Carbamazepine 10,11-epoxide	$C_{15}H_{12}N_2O_2$	252,0897	6,22	E/S	92,5	1,97	1,97	
Cimetidine	$C_{10}H_{16}N_6S$	252,1160	3,17	E/S	78,7	-0,11	-1,17	
Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	253,0522	4,92	E/S	87,9	0,79	0,75	
Kétoprofen	$C_{16}H_{14}O_3$	254,0937	7,80	E/S	85,4	3,61	3,56	
Lamotrigine	$C_9H_7Cl_2N_5$	255,0074	5,29	E/S	95,6	1,93	-0,66	
Propranolol	$C_{16}H_{21}NO_2$	259,1569	6,14	E/S	96,8	2,58	-0,66	
O-Desmethylvenlafaxine/ N-Desmethylvenlafaxine/	**$C_{16}H_{25}NO_2$**	**263,1881**	**5,06**	**E/S**	**83,2**	**2,45**	**-1,05**	
O-Desmethylvenlafaxine/ N-Desmethylvenlafaxine/	**$C_{16}H_{25}NO_2$**	**263,1882**	**4,99**	**E/S**	**83,2**	**2,45**	**-1,05**	
O-Desmethylvenlafaxine/ N-Desmethylvenlafaxine/	$C_{16}H_{25}NO_2$	263,1889	4,92	E/S	79,8	2,45	-1,05	
Mianserin	$C_{18}H_{20}N_2$	264,1627	6,40	E/S	77,8	3,83	0,47	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Mirtazapine	$C_{17}H_{19}N_3$	265,1583	4,80	E/S	78,6	3,21	-1,25	
Oxprenolol	**$C_{15}H_{23}NO_3$**	**265,1681**	**4,76**	**S**	**30,6**	**2,17**	**-1,07**	
Honokiol	$C_{18}H_{18}O_2$	266,1306	8,49	E/S	55,8	5,21	5,21	
Atenolol/Tributylphosphate	$C_{14}H_{22}N_2O_3$/ $C_{12}H_{27}O_4P$	266,1631	3,08	E/S	88,7	0,4/4,09	-2,82/4,09	
Zidovudine	$C_{10}H_{13}N_5O_4$	267,0969	1,43	E/S	73	-0,30	-0,41	
Metoprolol	$C_{15}H_{25}NO_3$	267,1828	5,11	E/S	96,9	1,76	-1,48	
Estrone	$C_{18}H_{22}O_2$	270,1613	8,84	E/S	69,9	4,31	4,31	
Sotalol	$C_{12}H_{20}N_2O_3S$	272,1190	2,66	E/S	97,4	-0,40	-3,19	
Galaxolidone/Estradiol	**C18H24O2**	**272,1767**	**9,27**	**E/S**	**95,9**	**4,65/3,75**	**4,65/3,75**	
Amitriptyline/ Maprotiline	$C_{20}H_{23}N$	277,1830	6,37	E/S	88,8	4,81/4,63/ 4,81/4,37	1,31/1,13/1,31/1,13	
Venlafaxine	$C_{17}H_{27}NO_2$	277,2039	6,11	E/S	90,9	2,74	-0,76	
Oxadixyl	$C_{14}H_{18}N_2O_4$	278,1267	5,06	E/S	43,1	1,79	1,79	
Doxepin	$C_{19}H_{21}NO$	279,1620	5,95	E/S	28,5	3,84	0,34	
Tramadol N-Oxide	$C_{16}H_{25}NO_3$	279,1831	4,04	E/S	37,1	1,33	1,24	
Mianserin-N-Oxide	$C_{18}H_{20}N_2O$	280,1574	4,73	E/S		2,71	2,63	
Cycloheximide	**$C_{15}H_{23}NO_4$**	**281,1623**	**4,05**	**S**		**0,91**	**0,91**	
Morphine/Norcodéine/ Hydromorphone	$C_{17}H_{19}NO_3$	285,1364	1,54	E/S	88,3	0,90/0,96/ 1,62	-2,3/-2,3/-2,3	
Morphine/Norcodéine/ Hydromorphone	$C_{17}H_{19}NO_3$	285,1367	3,48	E/S	88,3	0,90/0,96/ 1,62	-2,3/-2,3/-2,3	
Oxazepam	$C_{15}H_{11}ClN_2O_2$	286,0502	7,71	E/S	91,1	2,92	2,92	
16α-Hydroxyestrone	$C_{18}H_{22}O_3$	286,1577	7,98	S	62,7	3,44	3,44	
Androstenedione	$C_{19}H_{26}O_2$	286,1928	8,31	E	88	3,93	3,93	
Tenofovir	$C_9H_{14}N_5O_4P$	287,0786	1,20	E/S	73,8	-3,44	-3,44	
Trimethoprim	$C_{14}H_{18}N_4O_3$	290,1379	4,21	E/S	88,9	1,28	-0,19	
Nonivamide/Pramoxine	$C_{17}H_{27}NO_3$	293,1988	7,44	S	32,2	3,83/2,84	3,83/-0,55	
Sulfacytine/Sulfametomidine	$C_{12}H_{14}N_4O_3S$	294,0786	2,65	E/S	31,1	0,06/1,39	-0,01/-0,25	
Acide méclofénamique	$C_{14}H_{11}Cl_2NO_2$	295,0164	8,69	E/S	85,7	4,26/6,09	4,22/6,03	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Metoclopramide	$C_{14}H_{22}ClN_3O_2$	299,1399	4,62	E/S	85,6	1,40	-2,12	
Codeine	$C_{18}H_{21}NO_3$	299,1514	3,39	E/S	94,6	1,34	-2,16	
Codeine	$C_{18}H_{21}NO_3$	299,1517	3,80	S	94,6	1,34	-2,16	
Codeine	$C_{18}H_{21}NO_3$	299,1518	3,42	E/S	94,6	1,34	-2,16	
Acide Nordihydroguaiarétique	$C_{18}H_{22}O_4$	302,1514	8,32	E/S	59,2	4,76	4,76	
Acide Nordihydroguaiarétique	$C_{18}H_{22}O_4$	302,1517	7,91	E/S	52,0	4,76	4,76	
Acide Nordihydroguaiarétique	$C_{18}H_{22}O_4$	302,1526	8,23	E/S	52,7	4,76	4,76	
C14AEO2 (C14E2)	$C_{18}H_{38}O_3$	302,2809	10,48	E		5,16	5,16	
Cocaine/Hydromorphinol	$C_{17}H_{21}NO_4$	303,1471	5,02	E	92,5	2,28/	-1,22	
Vildagliptin	$C_{17}H_{25}N_3O_2$	303,1952	3,04	E/S	89,4	-0,22	-3,46	
Fluconazole	$C_{13}H_{12}F_2N_6O$	306,1041	5,47	E/S	86,7	0,56	0,49	
Methadone	$C_{21}H_{27}NO$	309,2094	7,16	E/S	88	5,01	1,51	
Desmethylcitalopram	$C_{19}H_{19}FN_2O$	310,1482	6,41	E/S	80,5	5,23	4,65	
Kresoxim méthyle	**$C_{18}H_{19}NO_4$**	**313,1310**	**8,29**	**S**	**41,4**	**4,34**	**4,37**	
Acide fenofibrique	$C_{17}H_{15}ClO_4$	318,0657	8,69	E/S	77,6	3,9	3,9	
Pazufloxacine	$C_{16}H_{15}FN_2O_4$	318,0996	3,17	E/S		-1,09	-1,90	
Imiprothrine	$C_{17}H_{22}N_2O_4$	318,1575	4,76	E		1,51	1,51	
Imiprothrine	$C_{17}H_{22}N_2O_4$	318,1576	6,88	E/S		1,51	1,51	
Imiprothrine	$C_{17}H_{22}N_2O_4$	318,1582	7,06	E/S		1,51	1,51	
Acide mycophénolique	$C_{17}H_{20}O_6$/ $C_{15}H_{17}FN_4O_3$	320,1256	8,00	E/S	82,6	3,53	3,43	
Citalopram	$C_{20}H_{21}FN_2O$	324,1632	6,43	E/S	57,7	3,77	0,26	
Bisoprolol	$C_{18}H_{31}NO_4$	325,2253	5,97	E/S	87,4	2,20	-1,04	
Clozapine	$C_{18}H_{19}ClN_4$	326,1300	6,10	E/S	84,4	3,40	-0,96	
Labetalol	$C_{19}H_{24}N_2O_3$	328,1783	5,98	E/S	42,5	1,89	-0,27	
Labetalol	$C_{19}H_{24}N_2O_3$	328,1807	5,64	E	76,5	1,89	-0,27	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
17α-Hydroxyprogesterone/ THC-OH, 11-OH-THC	$C_{21}H_{30}O_3$	330,2217	9,08	E/S	67,2	4,67	4,67	
Ciprofloxacine	$C_{17}H_{18}FN_3O_3$	331,1333	4,73	E/S	75,9	-0,86	-1,68	
Acebutolol	$C_{18}H_{28}N_2O_4$	336,2041	5,14	E/S	91,6	1,53	1,53	
Linézolide	$C_{16}H_{20}FN_3O_4$	337,1421	4,52	E/S	88,3	0,67	0,67	
Bitertanol	$C_{20}H_{23}N_3O_2$	337,1776	10,87	E/S	88	4,32	4,29	
Propiconazole	$C_{15}H_{17}Cl_2N_3O_2$	341,0695	8,83	E/S	82,5	4,32	4,29	
Omeprazole	$C_{17}H_{19}N_3O_3S$	345,1141	6,09	E/S	67,9	2,43	1,13	
C14AEO3 (C14E3)	$C_{20}H_{42}O_4$	346,3071	10,49	E	78,3	5,11	5,11	
Griseofulvin	$C_{17}H_{17}ClO_6$	352,0717	7,57	E/S	75,4	2,17	2,17	
Nitrendipine	$C_{18}H_{20}N_2O_6$	360,1317	5,65	E/S	39,4	2,17	2,17	
Ofloxacine/Levofloxacine	$C_{18}H_{20}FN_3O_4$	361,1432	4,56	E/S	84,9	0,09/0,09	-1,72/-1,72	
Celiprolol	$C_{20}H_{33}N_3O_4$	379,2464	5,58	E/S	95,9	1,50	-1,74	
Quetiapine	$C_{21}H_{25}N_3O_2S$	383,1664	6,46	E/S	67,1	2,81	-0,79	
Cetirizine	$C_{21}H_{25}ClN_2O_3$	388,1555	7,55	E/S	86,9	0,87	0,39	
Alfuzosine	$C_{19}H_{27}N_5O_4$	389,2064	5,74	E/S	85,5	1,53	1,53	
Mevastatin	$C_{23}H_{34}O_5$	390,2375	6,85	E	41,4	3,62	3,62	
DEHP/DNOP	$C_{24}H_{38}O_4$	390,2756	9,08	E	11,8	8,03/8,19	8,03/8,19	
DEHP/DNOP	$C_{24}H_{38}O_4$	390,2758	8,87	E	11,8	8,03/8,19	8,03/8,19	
Sparfloxacine	$C_{19}H_{22}F_2N_4O_3$	392,1654	4,39	E/S	26,1	-0,08	-0,91	
Pholcodine	$C_{23}H_{30}N_2O_4$	398,2215	0,89	E/S	81,5	1,14	-5,48	
TBEP	$C_{18}H_{39}O_7P$	398,2432	9,24	E/S		3,95	3,95	
Lovastatine	$C_{24}H_{36}O_5$	404,2555	8,47	E	53,9	3,90	3,90	
Risperidone	$C_{23}H_{27}FN_4O_2$	410,2148	4,75	E/S		2,63	-2,46	
Diltiazem	$C_{22}H_{26}N_2O_4S$	414,1603	6,85	E/S	82,1	2,73	-0,77	
PEG EO9 (HO-PEG9-OH)	$C_{18}H_{38}O_{10}$	414,2459	5,09	E/S	56,7	-1,58	-1,58	
PEG EO9 (HO-PEG9-OH)	$C_{18}H_{38}O_{10}$	414,2467	4,66	E	50,8	-1,58	-1,58	
DINP	$C_{26}H_{42}O_4$	418,3076	11,76	E/S	86,5	8,76	8,76	
Talmetoprim	$C_{22}H_{20}N_4O_5$	420,1461	7,41	E/S		2,72	2,69	
Losartan	$C_{22}H_{23}ClN_6O$	422,1618	7,75	E/S	83,7	5,00	4,22	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Clindamycine	$C_{18}H_{33}ClN_2O_5S$	424,1792	6,51	E/S	83,9	1,04	-2,43	
Irbesartan	$C_{25}H_{28}N_6O$	428,2325	7,77	E/S	89,3	5,39	4,37	
Mycophenolate mofetil	$C_{23}H_{31}NO_7$	433,2092	6,71	E	42,0	3,47	0,45	
Cloxacilline	$C_{19}H_{18}ClN_3O_5S$	435,0658	7,05	E/S	82,2	2,30	2,23	
Valsartan	$C_{24}H_{29}N_5O_3$	435,2260	8,20	E/S	90,5	5,27	5,25	
Candesartan	**$C_{24}H_{20}N_6O_3$**	**440,1586**	**7,92**	**E/S**	**83,7**	**5,31**	**5,18**	
C12AEO6 (C12E6)	$C_{24}H_{50}O_7$	450,3540	10,90	E	67,7	4,08	4,08	
C12AEO6 (C12E6)	$C_{24}H_{50}O_7$	450,3541	9,93	E	84,4	4,08	4,08	
C12AEO6 (C12E6)	$C_{24}H_{50}O_7$	450,3547	11,92	E	68,9	4,08	4,08	
C12AEO6 (C12E6)	$C_{24}H_{50}O_7$	450,3547	11,81	E	67,4	4,08	4,08	
Repaglinide	$C_{27}H_{36}N_2O_4$	452,2644	7,89	S	56,3	3,95	3,42	
Verapamil	$C_{27}H_{38}N_2O_4$	454,2820	6,85	E/S	85,4	5,04	1,54	
Fumagillin	$C_{26}H_{34}O_7$	458,2288	6,69	E/S	42,1	4,05	4,05	
Fumagillin	$C_{26}H_{34}O_7$	458,2322	10,91	E	29,2	4,05	4,05	
PEG EO10 (HO-PEG10-HO)	$C_{20}H_{42}O_{11}$	458,2733	4,89	E/S	50,0	-1,63	-1,63	
Bacampicillin	$C_{21}H_{27}N_3O_7S$	465,1544	3,90	E	21,2	1,47	-1,53	
Staurosporine	$C_{28}H_{26}N_4O_3$	466,2031	4,80	E	25,0	3,97	0,73	
C14AEO6 (C14E6)	$C_{26}H_{54}O_7$	478,3855	10,48	E	55,9	4,97	4,97	
Fexofenadine	$C_{32}H_{39}NO_4$	501,2877	7,14	E/S	88,5	2,94	2,32	
Telmisartan	$C_{33}H_{30}N_4O_2$	514,2357	7,63	E/S	95,3	6,13	5,19	
PEG EO12 (HO-PEG12-HO)	$C_{24}H_{50}O_{13}$	546,3240	5,25	E/S	51,2	-1,72	-1,72	
PEG EO12 (HO-PEG12-HO)	$C_{24}H_{50}O_{13}$	546,3242	5,69	E/S	82,6	-1,72	-1,72	
Darunavir	$C_{27}H_{37}N_3O_7S$	547,2351	7,63	E/S	76	2,82	2,72	
PEG EO13 (HO-PEG13-HO)	$C_{26}H_{54}O_{14}$	590,3502	5,88	E/S	69,2	-1,77	-1,77	
C16AEO8 (C16E8)	$C_{32}H_{66}O_9$	594,4696	11,11	E/S	75,0	5,76	5,76	
Clarithromycine	$C_{38}H_{69}NO_{13}$	747,4769	7,68	E/S	83,1	3,24	-0,26	
Azithromycine	$C_{38}H_{72}N_2O_{12}$	748,5078	5,96	E/S	94,3	2,45	-4,56	
Azithromycine	$C_{38}H_{72}N_2O_{12}$	748,5105	5,71	E/S	94,3	2,45	-4,56	
Oligomycine C	$C_{45}H_{74}O_{10}$	774,5303	9,43	E	53,7	8,27	8,27	
Iomeprol	$C_{17}H_{22}I_3N_3O_8$	776,8530	2,70	E/S	13,6	-1,45/-0,74	-1,45/-0,74	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Dolastatin 10	$C_{42}H_{68}N_6O_6S$	784,4852	6,05	E	51,6	5,07	1,53	
Diproleandomycine	$C_{41}H_{69}NO_{14}$	799,4782	6,00	E		5,26	1,76	
Roxithromycine	$C_{41}H_{76}N_2O_{15}$	836,5227	7,79	E/S	90,5	3,00	-0,58	
Megalomicine A	$C_{44}H_{80}N_2O_{15}$	876,5489	5,79	E/S	50,7	2,84	-4,16	
Spiramycine III	$C_{46}H_{78}N_2O_{15}$	898,5327	6,69	E/S	56,0	3,64	-3,36	

Annexe 6

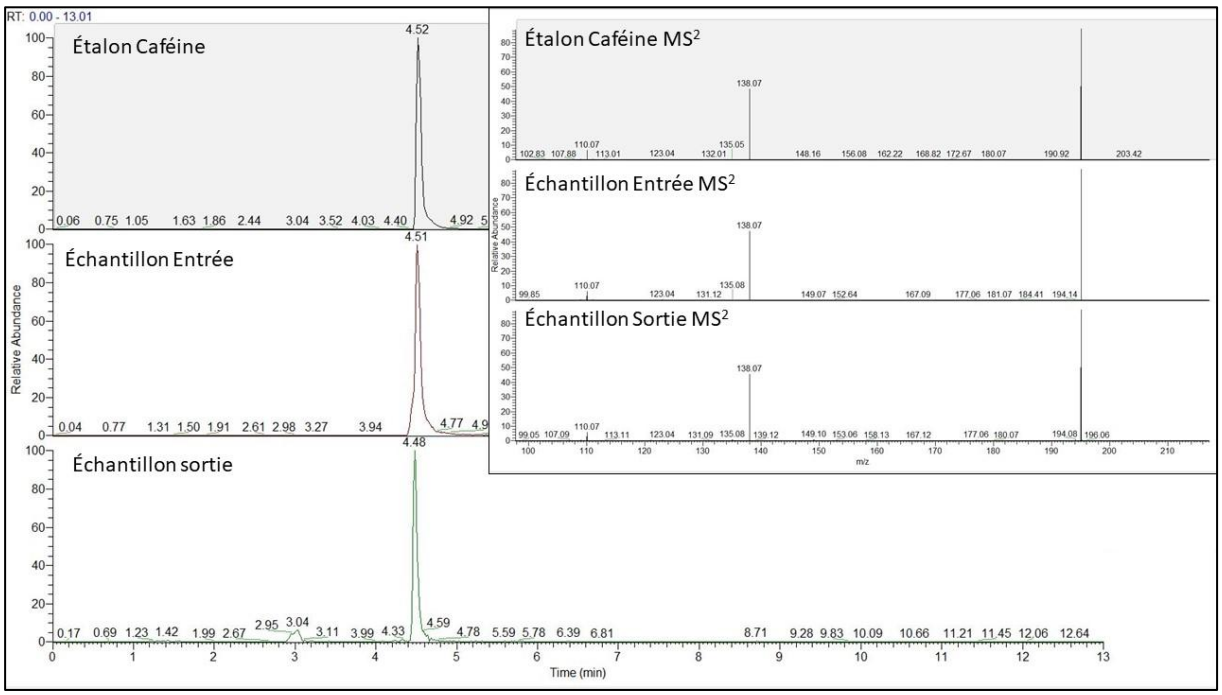


Figure 84 : Chromatogrammes de l'ion moléculaire de la caféine et spectres MS² de l'étalon de la caféine, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

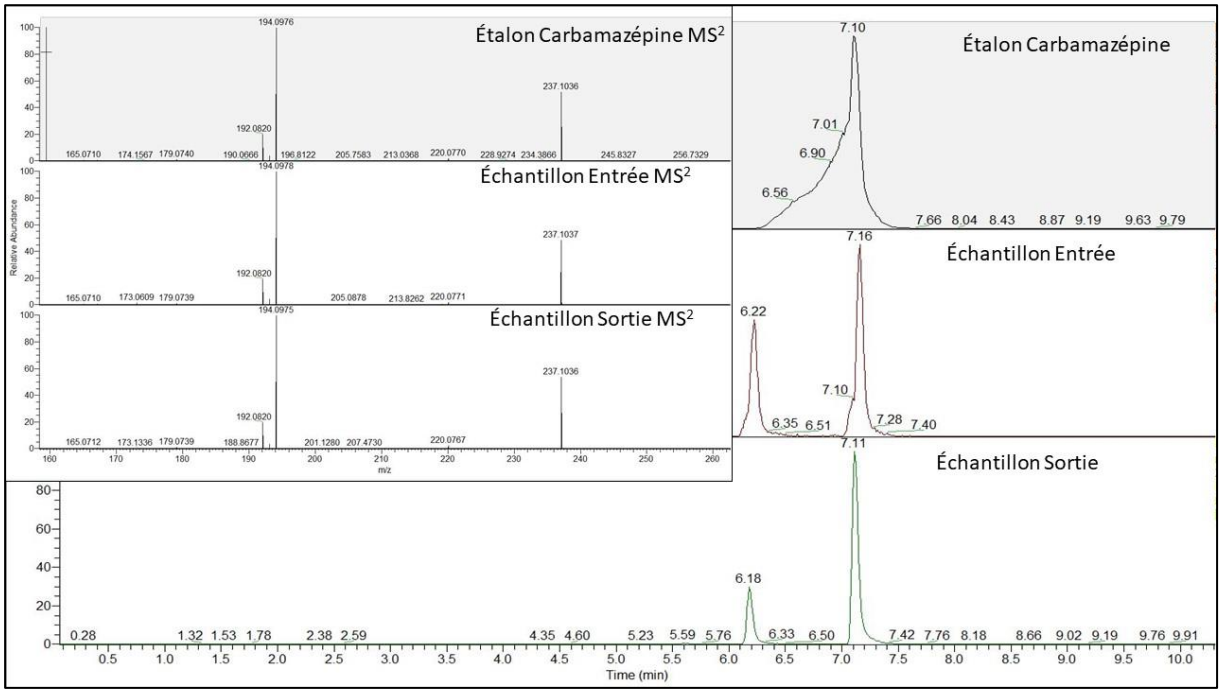


Figure 85 : Chromatogrammes de l'ion moléculaire de la carbamazépine et spectres MS² de l'étalon de la carbamazépine, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

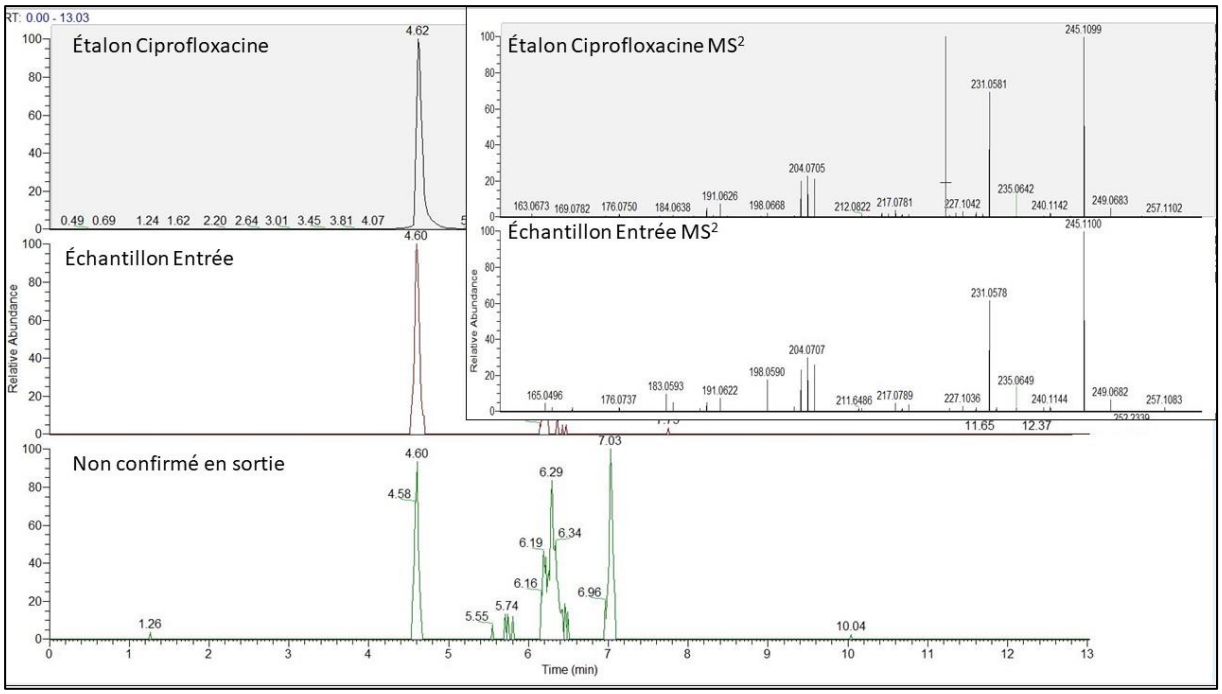


Figure 86 : Chromatogrammes de l'ion moléculaire de la ciprofloxacine et spectres MS^2 de l'étalon de la ciprofloxacine, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

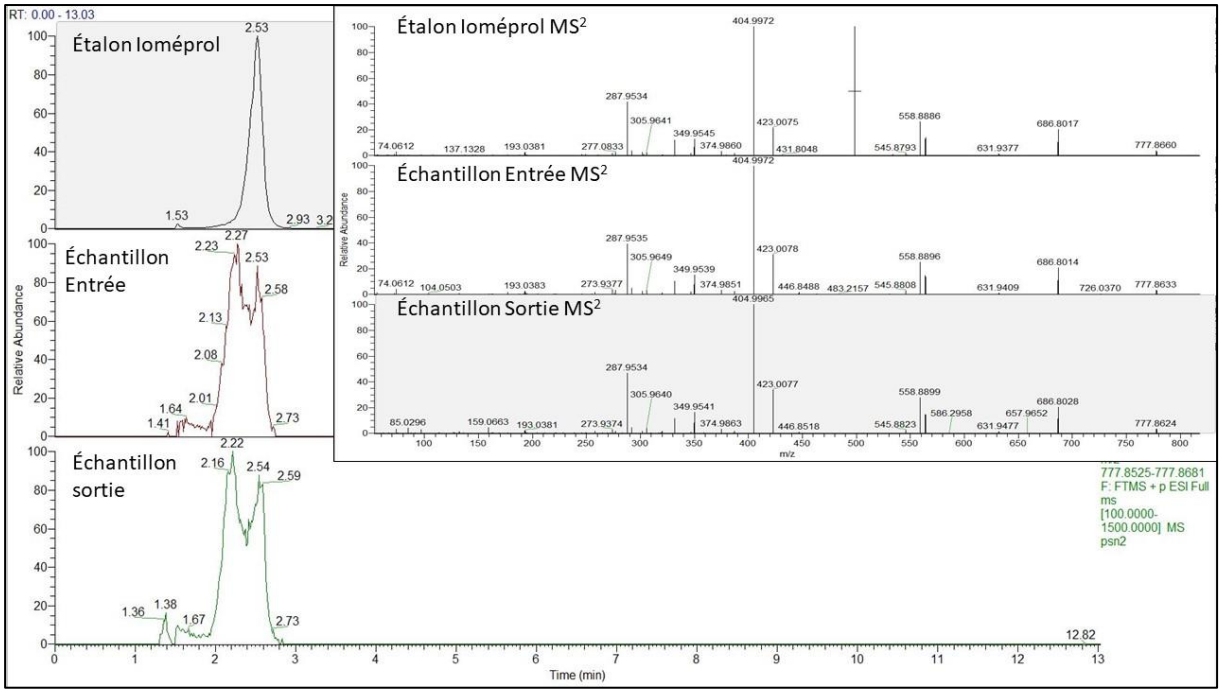


Figure 87 : Chromatogrammes de l'ion moléculaire de l'ioméprol et spectres MS^2 de l'étalon de l'ioméprol, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

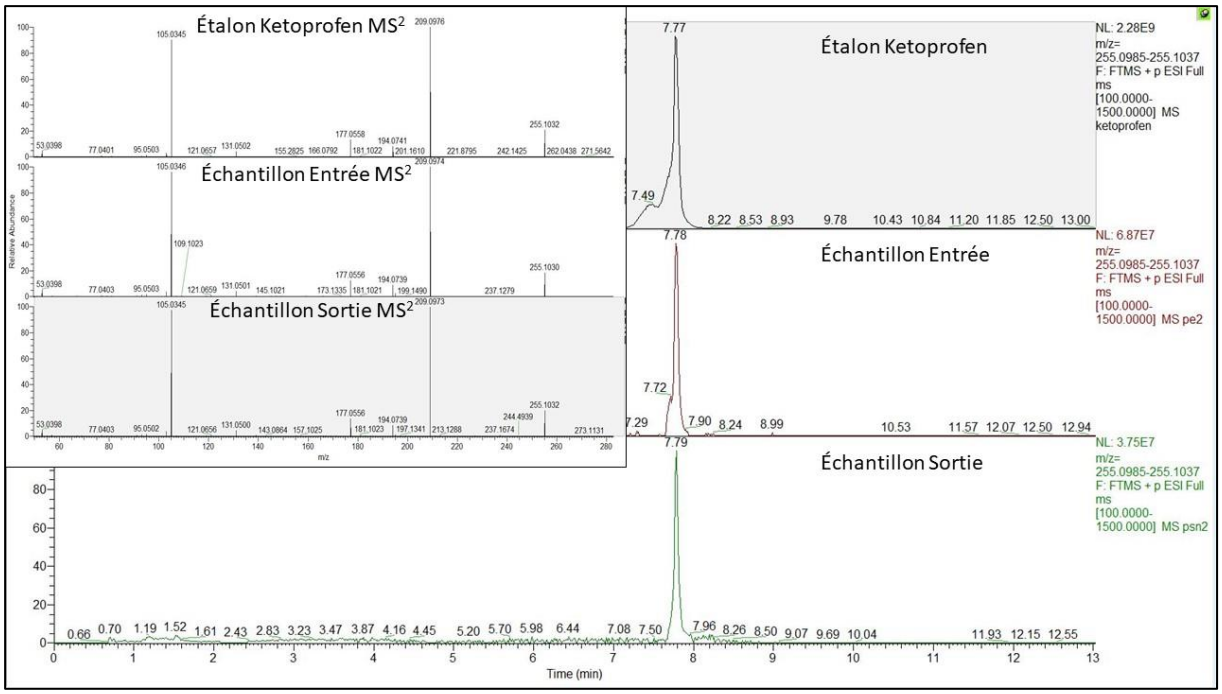


Figure 88 : Chromatogrammes de l'ion moléculaire du kétoprofen et spectres MS^2 de l'étalon du kétoprofen, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

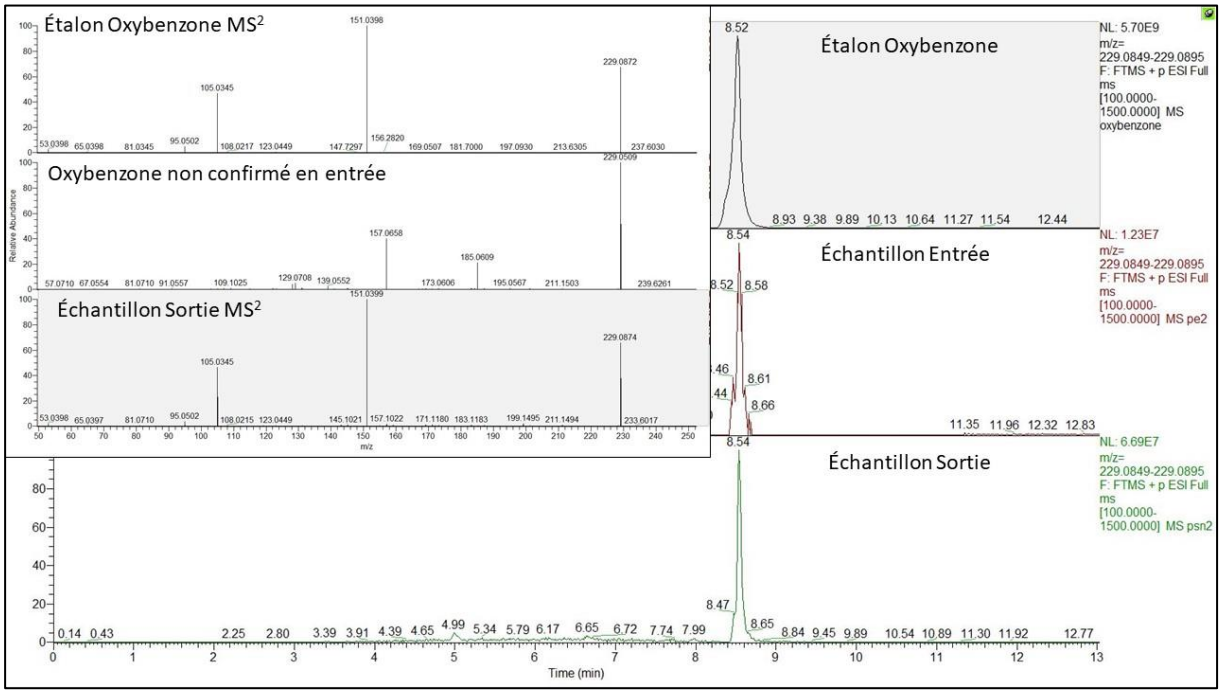


Figure 89 : Chromatogrammes de l'ion moléculaire de l'oxybenzone et spectres MS^2 de l'étalon de l'oxybenzone, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

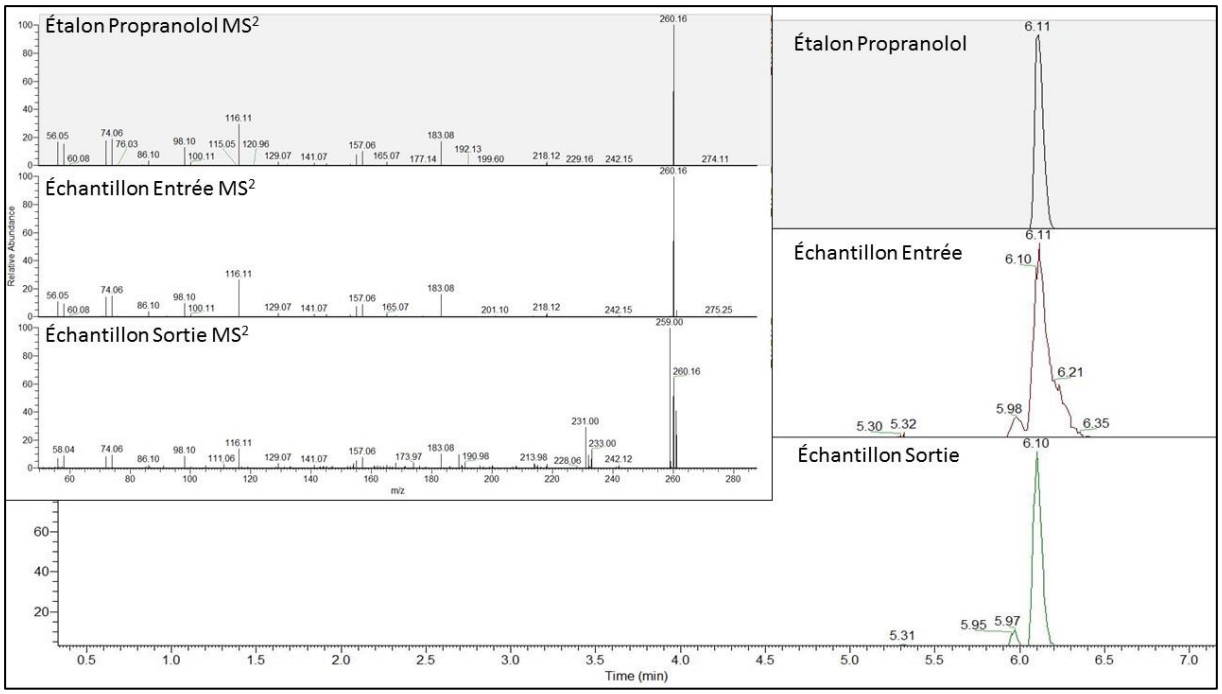


Figure 90 : Chromatogrammes de l'ion moléculaire du propranolol et spectres MS^2 de l'étalon du propranolol, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

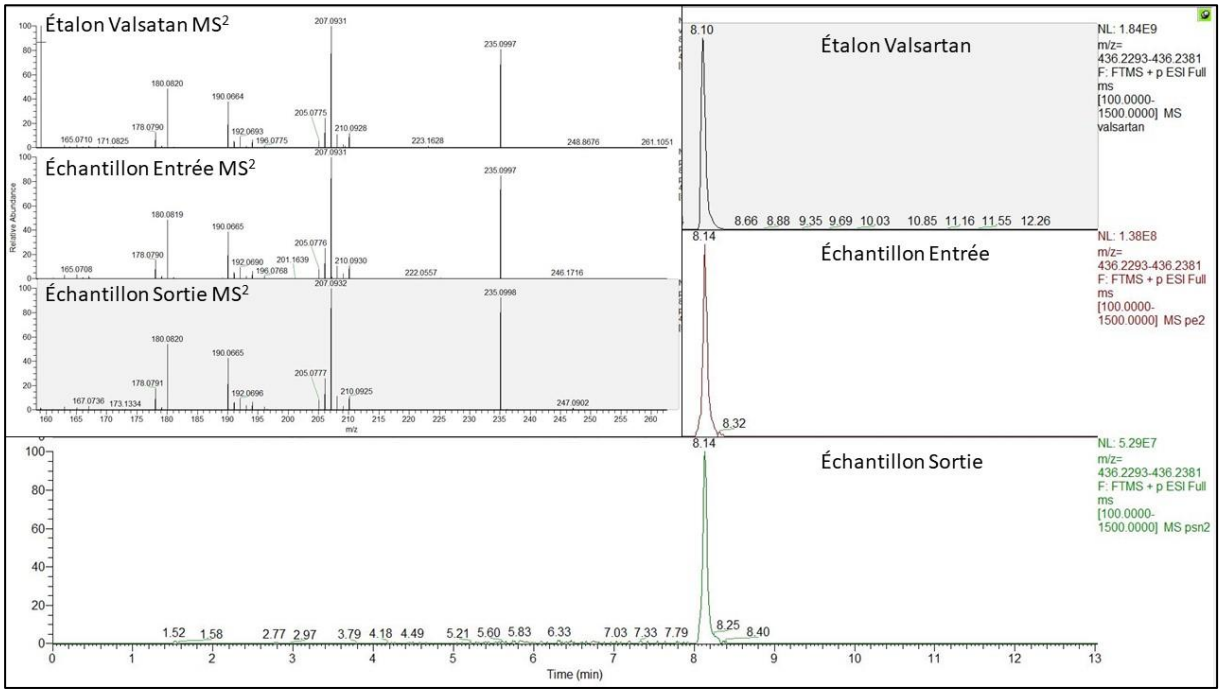


Figure 91 : Chromatogrammes de l'ion moléculaire du valsartan et spectres MS^2 de l'étalon du valsartan, d'un échantillon d'entrée et d'un échantillon de sortie pour la confirmation de l'identification.

Annexe 7

Tableau 32 : Liste des 60 molécules parentes utilisées pour générer une liste de métabolites.

Nom	Formule	Nombre de métabolites générés
2-(Methylthio)benzothiazole	$C_8H_7NS_2$	2
4'-Aminoacetanilide	C_8H10N_2O	2
4-Aminoantipyrine	$C_{11}H_{13}N_3O$	4
4-Formylaminoantipyrine	$C_{12}H_{13}N_3O_2$	5
4-Methylbenzotriazole	$C_7H_7N_3$	3
4-Toluenesulfonamide	$C_7H_9NO_2S$	2
5-Amino-2-chlorotoluene-4-sulfonic acid	$C_7H_8ClNO_3S$	2
5-Methyl-1H-benzotriazole	$C_7H_7N_3$	4
Acesulfam	$C_4H_5NO_4S$	1
Acetylaminoantipyrine	$C_{13}H_{15}N_3O_2$	4
Aspartame	$C_{14}H_{18}N_2O_5$	4
Atenolol	$C_{14}H_{22}N_2O_3$	4
Azithromycin	$C_{38}H_{72}N_2O_{12}$	16
Benzophenone-4	$C_{13}H_{10}O$	3
Benzothiazole	C_7H_5NS	1
Benzothiazole-2-sulfonic acid	$C_7H_5NO_3S_2$	2
Benzotriazole	$C_6H_5N_3$	1
Cafeine	$C_8H_{10}N_4O_2$	4
Candesartan	$C_{24}H_{20}N_6O_3$	3
Carbamazepine	$C_{15}H_{12}N_2O$	1
Ciprofloxacine	$C_{17}H_{18}FN_3O_3$	2
Cocaine	$C_{17}H_{21}NO_4$	5
Codeine	$C_{18}H_{21}NO_3$	5
Diclofenac	$C_{14}H_{11}Cl_2NO_2$	4
Ensulizole	$C_{13}H_{10}N_2O_3S$	1
Fosphomycin	$C_3H_5O_4P$	1
Furosemide	$C_{12}H_{11}ClN_2O_5S$	3
Gabapentine-Lactam	$C_9H_{15}NO$	2
HCTZ	$C_7H_8ClN_3O_4S_2$	1
Ibuprofen	$C_{13}H_{18}O_2$	2
Iohexol	$C_{19}H_{26}I_3N_3O_9$	10
Iomeprol	$C_{17}H_{22}I_3N_3O_8$	8
Iopromide	$C_{18}H_{24}I_3N_3O_8$	12
Irbesartan	$C_{25}H_{28}N_6O$	4
Kétoprofen	$C_{16}H_{14}O_3$	2
Lamotrigine	$C_9H_7Cl_2N_5$	2
Methylparaben	$C_8H_8O_3$	4
Metoprolol	$C_{15}H_{25}NO_3$	4
Morphine	$C_{17}H_{19}NO_3$	7
m-Xylene-4-sulphonic acid	$C_8H_{10}O_3S$	2
Nicotine	$C_{10}H_{14}N_2$	3

Nom	Formule	Nombre de métabolites générés
Ofloxacin	$C_{18}H_{20}FN_3O_4$	5
Omeprazole	$C_{17}H_{19}N_3O_3S$	7
Oxazepam	$C_{15}H_{11}ClN_2O_2$	1
Paracetamol	$C_8H_9NO_2$	3
Phthalate-bis-(2-ethylhexyl) (DEHP)	$C_{24}H_{38}O_4$	4
Propanolol	$C_{16}H_{21}NO_2$	3
Propylparaben	$C_{10}H_{12}O_3$	4
p-Toluidine-m-sulfonic acid	$C_7H_9NO_3S$	2
Sucralose	$C_{12}H_{19}Cl_3O_8$	9
Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	2
Tamoxifen	$C_{26}H_{29}NO$	3
TCPP	$C_9H_{18}Cl_3O_4P$	2
Telmisartan	$C_{33}H_{30}N_4O_2$	7
Tramadol	$C_{16}H_{25}NO_2$	7
Triclosan	$C_{12}H_7Cl_3O_2$	2
Trimethoprim	$C_{14}H_{18}N_4O_3$	4
Valsartan	$C_{24}H_{29}N_5O_3$	4
Vancomycin	$C_{66}H_{75}Cl_2N_9O_{24}$	5
Venlafaxine	$C_{17}H_{27}NO_2$	7

Annexe 8

Tableau 33 : Liste des 89 signaux correspondant à un ou plusieurs métabolites avec leurs formules brutes, la masse exacte et le temps de rétention (Tr) du signal détecté, le pourcentage de la meilleure correspondance avec mzCloud, les logP et logD (pH = 3), la nature de l'échantillon dans lequel le suspect est présent et un code couleur correspondant à : une correspondance avec mzCloud supérieure à 80% en bleu foncé et inférieure à 80% en bleu clair, une correspondance avec les bases de données de MetFrag (KEGG, CompTox, HMDB) supérieure à 50% en vert foncé et inférieure à 50% en vert clair et pas de correspondance avec une base de données en jaune. Les signaux en gras sont également des signaux de la liste des signaux pertinents. C : code couleur.

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
p-phenylenediamine	$C_6H_8N_2$	108,0689	0,71	E/S	85,1	0,32	-2,22	
p-phenylenediamine	$C_6H_8N_2$	108,0690	0,92	E/S	85,1	0,32	-2,22	
4-Aminophenol	C_6H_7NO	109,0530	0,89	E/S	90,8	0,84	-1,19	
4-Aminophenol	C_6H_7NO	109,0531	0,77	E/S	32,0	0,84	-1,19	
Acide benzoïque	$C_7H_6O_2$	122,0369	4,73	E	46,2	1,33	1,31	
7-hydroxy-2-azaspiro[4.5]decan-3-one	$C_9H_{15}NO_2$	169,1107	4,38	E/S		-0,35	-0,35	
(2R,3R,4S,6R)-4-(dimethylamino)-6-methyloxane-2,3-diol	$C_8H_{17}NO_3$	175,1207	0,78	E/S		-0,49	-3,99	
(2R,4R,5S,6S)-4-methoxy-4,6-dimethyloxane-2,5-diol	$C_8H_{16}O_4$	176,1047	3,76	E/S		-0,06	-0,06	
(2R,4R,5S,6S)-4-methoxy-4,6-dimethyloxane-2,5-diol	$C_8H_{16}O_4$	176,1047	3,07	E/S		-0,06	-0,06	
paraxanthine	C7H8N4O2	180,0645	3,38	E/S	90,1	0,24	0,24	
paraxanthine	C7H8N4O2	180,0645	3,93	E/S	90,1	0,24	0,24	
2-(methylsulfanyl)-1,3-benzothiazol-6-ol	C8H7NOS2	196,9973	6,21	E/S	18,2	3,12	3,11	
ecgonine methyl ester	C10H17NO3	199,1207	1,06	E/S	81,4	-0,21	-3,71	
ecgonine methyl ester	C10H17NO3	199,1212	0,76	E/S	58,1	-0,21	-3,71	
(2S)-3-methyl-2-pentanamidobutanoic acid	C10H19NO3	201,1362	0,77	S		1,72	1,70	
(2S)-3-methyl-2-pentanamidobutanoic acid	C10H19NO3	201,1372	7,41	E		1,72	1,70	
1,3,7-trimethyluric acid	C8H10N4O3	210,0750	4,30	E/S	39,7	0,33	0,33	
1,3,7-trimethyluric acid	C8H10N4O3	210,0751	3,95	E/S	21,3	0,33	0,33	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
1,3,7-trimethyluric acid	C8H10N4O3	210,0757	4,24	E/S	53,8	0,33	0,33	
N-(4-methylbenzenesulfonyl)acetamide	C9H11NO3S	213,0463	3,85	E/S		1,08	1,06	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)formamide	C11H11N3O2	217,0850	2,78	E/S		0,16	0,16	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)formamide	C11H11N3O2	217,0850	3,68	E/S		0,16	0,16	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)formamide	C11H11N3O2	217,0855	1,42	E/S		0,16	0,16	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)formamide	C11H11N3O2	217,0861	2,66	E/S		0,16	0,16	
atenolol-desisopropyl	C11H16N2O3	224,1159	4,05	E		-0,78	-3,81	
atenolol-desisopropyl	C11H16N2O3	224,1162	5,91	E/S		-0,78	-3,81	
atenolol-desisopropyl	C11H16N2O3	224,1164	3,28	E/S		-0,78	-3,81	
1-amino-3-[4-(2-methoxyethyl)phenoxy]propan-2-ol	C12H19NO3	225,1362	2,94	S		0,55	-2,48	
1-amino-3-[4-(2-methoxyethyl)phenoxy]propan-2-ol	C12H19NO3	225,1373	1,64	E/S		0,55	-2,48	
acetaminophen sulfate	C8H9NO5S	231,0208	2,69	E		0,43	-1,94	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)acetamide	C12H13N3O2	231,1008	2,93	E/S		0,21	0,21	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)acetamide	C12H13N3O2	231,1011	4,26	E/S		0,21	0,21	
4-acetaminoantipyrine	C13H15N3O2	245,1168	4,34	E/S	80	0,15	0,15	
N-desmethyltramadol / O-desmethyltramadol	C15H23NO2	249,1724	5,33	E/S	33,3	1,72	-1,20	
N-desmethyltramadol / O-desmethyltramadol	C15H23NO2	249,1726	4,16	E/S	73,8	1,72	-1,20	
N-desmethyltramadol / O-desmethyltramadol	C15H23NO2	249,1738	4,03	E/S	52,6	1,72	-1,20	
N-desmethyltramadol / O-desmethyltramadol	C15H23NO2	249,1741	4,73	E/S	67,3	1,72	-1,20	
bis[(1-chloropropan-2-yl)oxy]phosphinic acid	C6H13Cl2O4P	249,9928	8,21	E/S	2,3	2,85	2,85	
2-hydroxycarbamazepine	C15H12N2O2	252,0897	6,22	E/S	24,0	2,46	2,46	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
N-desmethylvenlafaxine/O-desmethylenlafaxine	C16H25NO2	263,1881	5,06	E/S	83,2	2,85	2,85	
N-desmethylvenlafaxine/O-desmethylenlafaxine	C16H25NO2	263,1882	4,99	E/S	83,2	2,85	2,85	
N-desmethylvenlafaxine/O-desmethylenlafaxine	C16H25NO2	263,1889	4,92	E/S	38,9	2,85	2,85	
2-{4-[3-(isopropylamino)-2-oxopropoxy]phenyl}acetamide	C14H20N2O3	264,1473	4,11	E/S		0,74	-2,50	
2-{4-[3-(isopropylamino)-2-oxopropoxy]phenyl}acetamide	C14H20N2O3	264,1473	4,48	E/S		0,74	-2,50	
2-{4-[3-(isopropylamino)-2-oxopropoxy]phenyl}acetamide	C14H20N2O3	264,1486	4,27	E/S		0,74	-2,50	
1-(isopropylamino)-3-[4-(2-methoxyethyl)phenoxy]propan-2-one	C15H23NO3	265,1677	3,48	E/S		2,07	-1,17	
1-(isopropylamino)-3-[4-(2-methoxyethyl)phenoxy]propan-2-one	C15H23NO3	265,1681	4,76	E/S		2,07	-1,17	
1-(isopropylamino)-3-[4-(2-methoxyethyl)phenoxy]propan-2-one	C15H23NO3	265,1689	2,35	E/S		2,07	-1,17	
Acide aténolique	C14H21NO4	267,1466	4,17	E/S	95,2	-1,24	-1,69	
Acide aténolique	C14H21NO4	267,1467	4,23	E/S	95,2	-1,24	-1,69	
Acide aténolique	C14H21NO4	267,1488	4,03	E/S	95,2	-1,24	-1,69	
Acide aténolique	C14H21NO4	267,1488	4,38	E/S	95,2	-1,24	-1,69	
9-hydroxy-10-oxo-2-azatricyclopentadeca-1(15),3,5,7,11,13-hexaene-2-carboxamide	C15H12N2O3	268,0847	7,64	E/S		2,81	2,81	
9-hydroxy-10-oxo-2-azatricyclopentadeca-1(15),3,5,7,11,13-hexaene-2-carboxamide	C15H12N2O3	268,0850	7,17	E/S		2,81	2,81	
2-chloro-N-(2-ethyl-6-methylphenyl)-N-(1-hydroxypropan-2-yl)acetamide	C14H20ClNO2	269,116	3,98	E/S		1,10	1,10	
normorphine	C16H17NO3	271,1203	6,95	E/S	47,4	0,26	-2,43	
normorphine	C16H17NO3	271,1203	6,69	E/S	63,6	0,26	-2,43	
5-[(2,4-diaminopyrimidin-5-yl)methyl]-2,3-dimethoxyphenol	C13H16N4O3	276,1209	7,01	E/S		3,80	3,80	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
5-[(2,4-diaminopyrimidin-5-yl)methyl]-2,3-dimethoxyphenol	C13H16N4O3	276,1219	3,24	E/S		1,14	-0,34	
5-[(2,4-diaminopyrimidin-5-yl)methyl]-2,3-dimethoxyphenol	C13H16N4O3	276,1220	3,65	E/S		1,14	-0,34	
4-[(2,4-diaminopyrimidin-5-yl)methyl]-2,6-dimethoxyphenol	C13H16N4O3	276,1239	3,53	E/S	48,8	1,14	-0,34	
Mono(2-ethylhexyl)phtalate	$C_{16}H_{22}O_4$	278,1508	9,15	E/S	19,2	4,66	4,40	
Mono(2-ethylhexyl)phtalate	$C_{16}H_{22}O_4$	278,1512	11,03	E	76,5	4,66	4,40	
(1R,2R)-2-[(dimethylamino)methyl]-1-(3-methoxyphenyl)cyclohexane-1,4-diol	$C_{16}H_{25}NO_3$	279,1831	4,04	E/S	37,1	0,99	-2,51	
(1R,2R)-2-[(dimethylamino)methyl]-1-(3-methoxyphenyl)cyclohexane-1,4-diol	$C_{16}H_{25}NO_3$	279,1831	6,70	E/S		0,99	-2,51	
norcodeine	$C_{17}H_{19}NO_3$	285,1364	1,54	E/S	51,3	0,96	-2,28	
norcodeine	$C_{17}H_{19}NO_3$	285,1367	3,48	E/S	82,3	0,96	-2,28	
benzoylecgonine	$C_{16}H_{19}NO_4$	289,1312	4,95	E/S	87,8	-0,59	-0,88	
2-(1-hydroxycyclohexyl)-2-(4-methoxyphenyl)-N,N-dimethylethanamine oxide	$C_{17}H_{27}NO_3$	293,1988	7,44	E/S		1,61	1,53	
1-[2-(dimethylamino)-1-(4-methoxyphenyl)ethyl]cyclohexane-1,3-diol	$C_{17}H_{27}NO_3$	293,1988	4,51	E/S		1,28	-2,22	
acetylsulfamethoxazole	$C_{12}H_{13}N_3O_4S$	295,0630	5,74	E/S	87,3	0,86	0,86	
morphine N-oxide	$C_{17}H_{19}NO_4$	301,1308	4,71	E	43,2	0,07	0,00	
morphine N-oxide	$C_{17}H_{19}NO_4$	301,1324	3,71	E/S	20,2	0,07	0,00	
morphine N-oxide	$C_{17}H_{19}NO_4$	301,1332	4,63	E/S	22,4	0,07	0,00	
4'-hydroxydiclofenac	$C_{14}H_{11}Cl_2NO_3$	311,0113	8,04	E/S	20	3,96	3,89	
codeine N-oxide	$C_{18}H_{21}NO_4$	315,1469	3,59	E/S	51,2	0,22	0,14	
codeine N-oxide	$C_{18}H_{21}NO_4$	315,1471	3,65	S	30,2	0,22	0,14	
codeine N-oxide	$C_{18}H_{21}NO_4$	315,1479	3,20	E/S	53,5	0,22	0,14	
2-[(5-methoxy-1H-1,3-benzodiazol-2-ylsulfinyl)methyl]-3,5-dimethylpyridin-4-ol	$C_{16}H_{17}N_3O_3S$	331,0984	4,75	E/S		2,29	0,98	

Nom du ou des suspects	Formule(s) brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
2-[(5-methoxy-1H-1,3-benzodiazol-2-ylsulfinyl)methyl]-3,5-dimethylpyridin-4-ol	$C_{16}H_{17}N_3O_3S$	331,0992	5,41	E/S		2,29	0,98	
2-[(5-methoxy-1H-1,3-benzodiazol-2-ylsulfinyl)methyl]-3,5-dimethylpyridin-4-ol	$C_{16}H_{17}N_3O_3S$	331,0995	5,29	E/S		2,29	0,98	
1-(2-ethyl-5-hydroxyhexyl) 2-(2-ethylhexyl) phthalate	$C_{24}H_{38}O_5$	406,2717	8,59	E/S		6,56	6,56	
2-butyl-7-hydroxy-3-{[2'-(2H-1,2,3,4-tetrazol-5-yl)-[1,1'-biphenyl]-4-yl]methyl}-1,3-diazaspiro[4.4]non-1-en-4-one	$C_{25}H_{28}N_6O_2$	444,2264	7,36	E		4,01	3,13	
6-[(7-chloro-2-oxo-5-phenyl-1,3-dihydro-1,4-benzodiazepin-3-yl)oxy]-3,4,5-trihydroxyoxane-2-carboxylic acid	$C_{21}H_{19}ClN_2O_8$	462,0848	6,99	E/S		4,86	3,90	
4'-{[6-(6-hydroxy-1-methyl-1,3-benzodiazol-2-yl)-4-methyl-2-propyl-1,3-benzodiazol-1-yl]methyl}-[1,1'-biphenyl]-2-carboxylic acid	$C_{33}H_{30}N_4O_3$	530,2279	8,90	E		5,68	4,75	
4'-{[6-(6-hydroxy-1-methyl-1,3-benzodiazol-2-yl)-4-methyl-2-propyl-1,3-benzodiazol-1-yl]methyl}-[1,1'-biphenyl]-2-carboxylic acid	$C_{33}H_{30}N_4O_3$	530,2310	7,59	S		5,75	4,79	
desosaminylazithromycin	$C_{30}H_{58}N_2O_9$	590,4151	4,02	E/S		1,34	-5,66	
N1,N3-bis(2,3-dihydroxypropyl)-5-(2-hydroxyacetamido)-2,4,6-triiodo-N1-methylbenzene-1,3-dicarboxamide	$C_{17}H_{22}I_3N_3O_8$	776,8530	2,70	E/S		-1,09	-1,09	
3-({3-[(2,3-dihydroxypropyl)carbamoyl]-5-(2-hydroxy-N-methylacetamido)-2,4,6-triiodophenyl}formamido)-2-hydroxypropanoic acid	$C_{17}H_{20}I_3N_3O_9$	790,8320	2,84	E/S		-1,13	-2,40	

Annexe 9

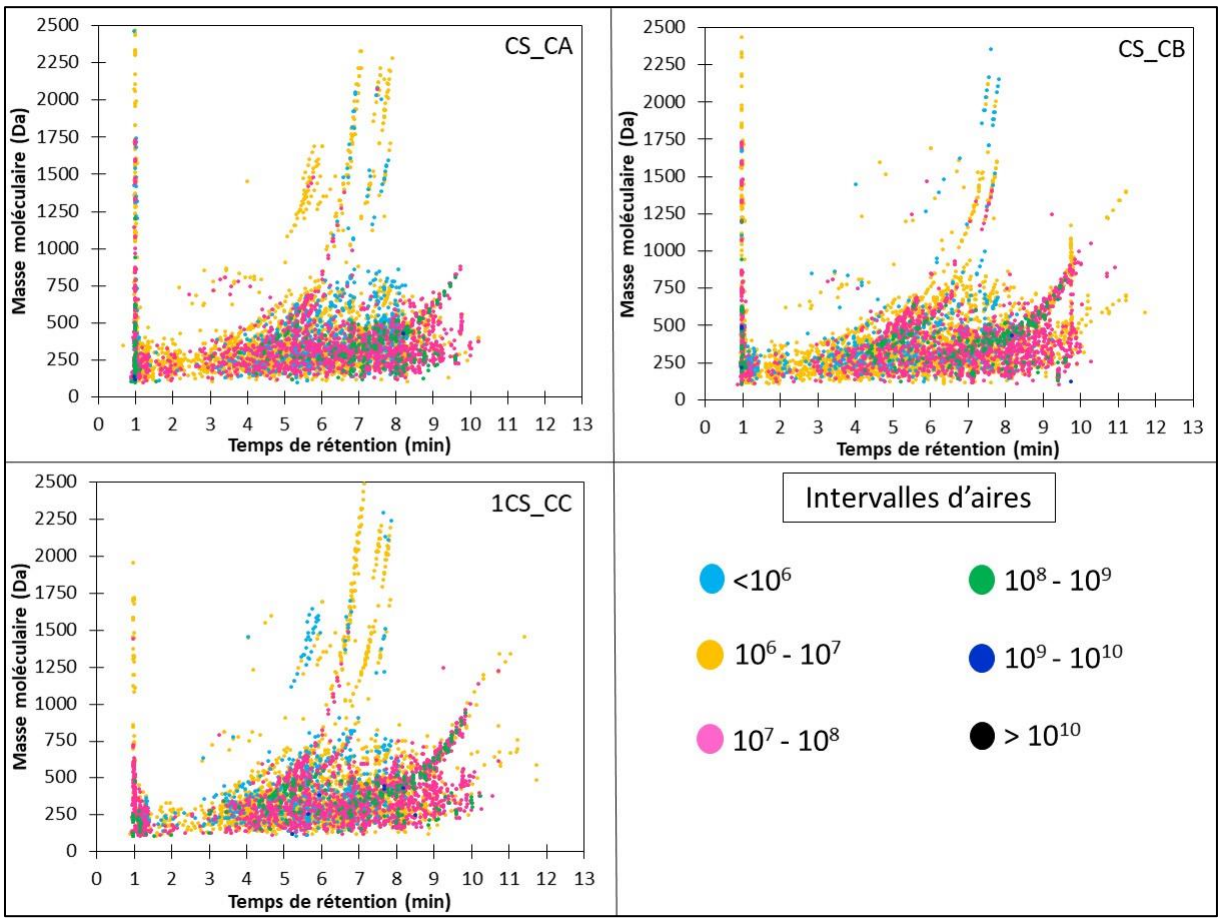


Figure 92 : Empreintes des échantillons de panache du Clain pour les campagnes A, B et C. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. CS : panache du Clain, CA : campagne A, CB : campagne B, CC : campagne C (échantillon 1). Mode positif (HESI+).

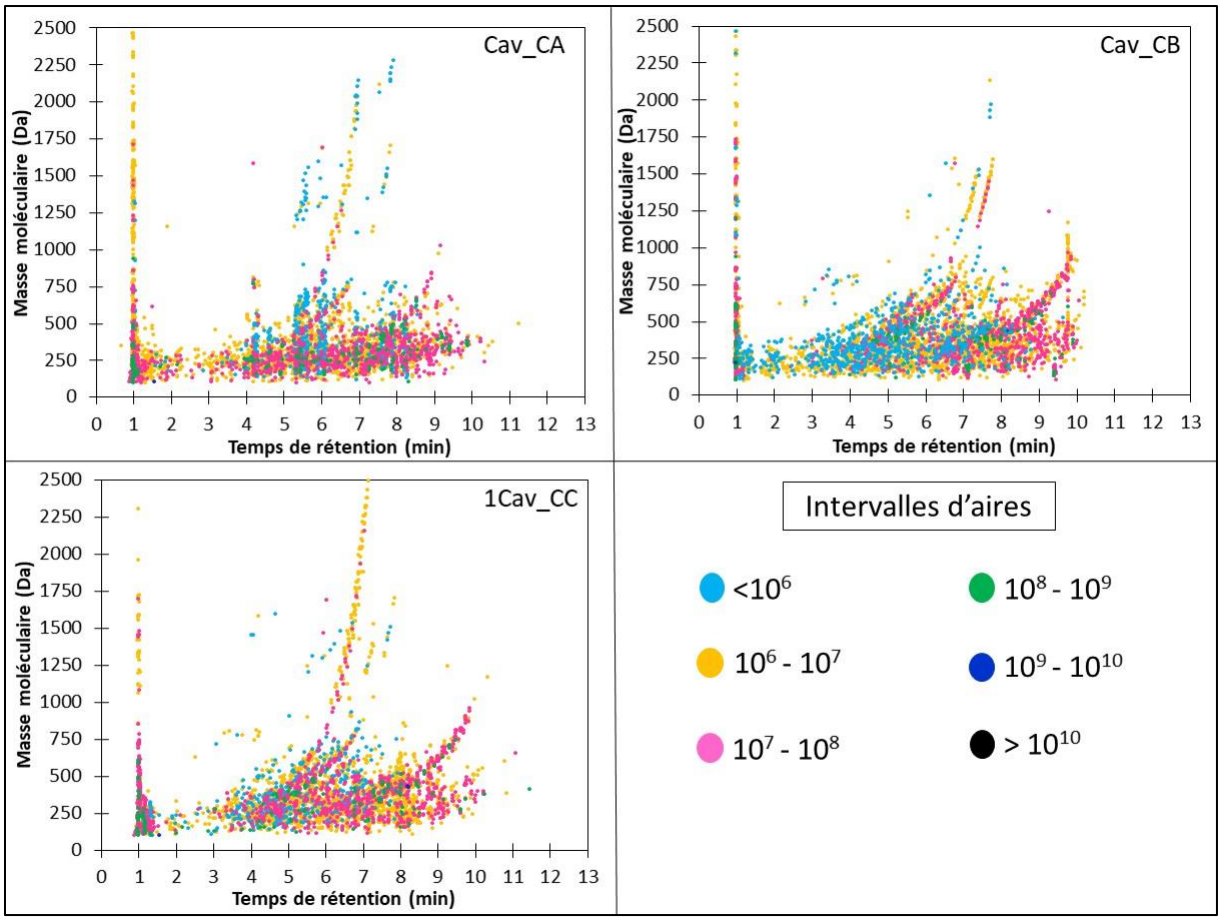


Figure 93 : Empreintes des échantillons de l'aval du Clain pour les campagnes A, B et C. Chaque point représente un signal détecté dans l'échantillon avec une masse, un temps de rétention et une aire de pic. Cav : aval du Clain, CA : campagne A, CB : campagne B, CC : campagne C (échantillon 1). Mode positif (HESI+).

Annexe 10

Tableau 34 : Liste des propositions de structures attribuées à 13 signaux pertinents par l'utilisation de mzCloud. Tr : temps de rétention, MC : pourcentage de correspondance entre le spectre MS2 expérimental et celui de mzCloud.

Nom du suspect	Formule Brute	Masse exacte (Da)	Tr (min)	MC(%)
1,3-Divinyl-2-imidazolidinone	$C_7H_{10}N_2O$	138,0792	4,28	70,2
Iminostilbene/2-Phenylindole	$C_{14}H_{11}N$	193,0897	7,64	82,4/71,3
Indole-3-butyrique acide/Methsuximide/ 3-ethyl-4-hydroxy-1-methyl-1,2-dihydroquinolin-2-one/ N1-(5-oxo-5,6,7,8-tetrahydronaphthalen-2-yl)acetamide	$C_{12}H_{13}NO_2$	203,0945	5,18	70,9/68,1/70,3/66,6
Sulfapyridine	$C_{11}H_{11}N_3O_2S$	249,0576	4,05	75,9
N-desmethyltramadol/Alprenolol	$C_{15}H_{23}NO_2$	249,1724	5,32	61,3/71,9
Cimetidine	$C_{10}H_{16}N_6S$	252,1154	3,60	72,2
Atenolol	$C_{14}H_{22}N_2O_3$	266,1627	1,34	75,6
Amitriptyline/Maprotiline	$C_{20}H_{23}N$	277,1829	7,63	85,1/70,8
5-Hydroxy-7-(hydroxymethyl)-2-methyl-2-(5-oxotetrahydro-2-furanyl)-2,3-dihydro-4H-chromen-4-one/ 3-(3-Acetyl-2,6-dihydroxy-5-methylbenzyl)-4-hydroxy-5-methyl-2(5H)-furanone	$C_{15}H_{16}O_6$	292,0974	7,51	65,5/63,0
Dihydrocodeine/6β-Hydrocol	$C_{18}H_{23}NO_3$	301,1675	3,74	82,3/82,3
Diacetalol	$C_{16}H_{24}N_2O_4$	308,1736	4,43	85
Desmethylcitalopram	$C_{19}H_{19}FN_2O$	310,1481	6,78	71,3
Disopyramide	$C_{21}H_{29}N_3O$	339,2312	5,46	79,2

Annexe 11

Tableau 35 : Liste des 127 signaux correspondant à un ou plusieurs suspects avec leurs formules brutes, la masse exacte et le temps de rétention (Tr) du signal détecté, le pourcentage de la meilleure correspondance avec mzCloud, les logP et logD (pH = 3), la nature de l'échantillon dans lequel le suspect est présent et un code couleur correspondant à : une correspondance avec mzCloud supérieure à 80% en bleu foncé et inférieure à 80% en bleu clair, une correspondance avec les bases de données de MetFrag (KEGG, CompTox, HMDB) supérieure à 50% en vert foncé et inférieure à 50% en vert clair, pas de correspondance avec une base de données en jaune, et la confirmation du suspect en rouge. Les signaux en gras sont également des signaux de la liste des signaux pertinents et ceux écrits en rouge sont également présents en sorties de STEP. Cam : Clain amont, Cav : Clain aval, C : code couleur.

Nom du suspect	Formule(s) Brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Benzotriazole	$C_6H_5N_3$	119,0485	5,13	Cam/Cav	85,9	1,30	1,3	
Nicotinamide	$C_6H_6N_2O$	122,0481	1,28	Cav	51,1	-0,37	-0,37	
Dimetridazole	$C_5H_7N_3O_2$	141,0537	3,46	Cam/Cav	70,6	0,23	0,01	
Piracetam	$C_6H_{10}N_2O_2$	142,0745	1,07	Cav	51,7	-1,69	-1,69	
8-Hydroxyquinoline	C_9H_7NO	145,0526	4,42	Cam/Cav	81,6	1,83	0,60	
Cinnamamide	C_9H_9NO	147,0685	1,29	Cav		-0,47	-0,50	
Methamphetamine	$C_{10}H_{15}N$	149,1214	5,07	Cav	20,2	2,23	-1,01	
Paracetamol	$C_8H_9NO_2$	151,0637	3,15	Cav	86,2	0,91	0,91	
L-(-)-Norephedrine	$C_9H_{13}NO$	151,0997	1,12	Cam/Cav	53,2	0,89	-2,15	
L-(-)-Norephedrine	$C_9H_{13}NO$	151,1000	3,32	Cam	51,1	0,89	-2,15	
Dopamine	$C_8H_{11}NO_2$	153,0794	1,09	Cav	33,7	0,03	-2,25	
Patulin	$C_7H_6O_4$	154,0278	4,19	Cav	12,8	-0,38	-0,38	
Pregabalin	$C_8H_{17}NO_2$	159,1259	1,65	Cam/Cav	31,8	-1,35	-2,08	
Pregabalin	$C_8H_{17}NO_2$	159,1262	1,08	Cam/Cav	24,4	-1,35	-2,08	
Fenuron	$C_9H_{12}N_2O$	164,0950	1,95	Cav		1,33	1,33	
Phenylephrine	$C_9H_{13}NO_2$	167,0945	1,07	Cav	34,6	-0,07	-2,64	
Phenylephrine	$C_9H_{13}NO_2$	167,0951	1,30	Cav	30,7	-0,07	-2,64	
Desethylhydroxyatrazine	$C_6H_{11}N_5O$	169,0971	1,96	Cam	65,6	-0,6	-0,6	
Levetiracetam	$C_8H_{14}N_2O_2$	170,1058	1,08	Cam/Cav	50,5	-0,59	-0,59	

Nom du suspect	Formule(s) Brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Gabapentin	$C_9H_{17}NO_2$	171,1258	1,18	Cam/Cav	87,6	-1,27	-2,00	
Epinephrine	$C_9H_{13}NO_3$	183,0899	1,18	Cav	51,4	-0,43	-2,95	
Terbutylazine-desethyl-2-hydroxy	$C_7H_{13}N_5O$	183,1119	4,10	Cam/Cav		-0,40	-0,42	
Terbutylazine-desethyl-2-hydroxy	$C_7H_{13}N_5O$	183,1127	2,47	Cav		-0,40	-0,42	
Phenazone/Antipyrine	$C_{11}H_{12}N_2O$	188,0949	5,35	Cav	74,3	1,22/1,22	1,22/1,22	
Carbendazim	$C_9H_9N_3O_2$	191,0692	5,42	Cav	26,7	1,80	0,73	
DEET	$C_{12}H_{17}NO$	191,1318	7,88	Cam/Cav	95,2	2,50	2,50	
MDMA	$C_{11}H_{15}NO_2$	193,1102	3,39	Cav	51,3	1,86	-1,38	
Aurantiogliocladin	$C_{10}H_{12}O_4$	196,0731	5,34	Cam/Cav		1,00	1,00	
Simazine	$C_7H_{12}ClN_5$	201,0797	3,72	Cav	13,3	1,78	0,60	
Levamisole	$C_{11}H_{12}N_2S$	204,0722	3,87	Cav	79,8	2,36	-0,35	
MDEA	$C_{12}H_{17}NO_2$	207,1259	4,23	Cam/Cav		2,22	-1,03	
MDEA	$C_{12}H_{17}NO_2$	207,1269	4,04	Cav		2,22	-1,03	
Aminocarb	$C_{11}H_{16}N_2O_2$	208,1212	4,08	Cav	40,2	2,09	0,53	
Dibutylphosphate	$C_8H_{18}O_4P$	209,0954	6,03	Cam/Cav		2,38	1,31	
Terbacil	$C_9H_{13}ClN_2O_2$	216,0645	4,45	Cav	19,4	1,29	1,29	
Primidone	$C_{12}H_{14}N_2O_2$	218,1055	4,82	Cav	**53,7**	1,12	1,12	
Trans-zeatin	$C_{10}H_{13}N_5O$	219,1124	1,31	Cav	86,5	-0,15	-1,18	
Carbofuran/Metaxalone	$C_{12}H_{15}NO_3$	221,1052	6,74	Cav	53,3	2,05/2,38	2,05/2,38	
4-Formylaminoantypiridine	$C_{12}H_{13}N_3O_2$	231,1008	3,24	Cav		0,11	0,11	
4-Formylaminoantypiridine	$C_{12}H_{13}N_3O_2$	231,1010	4,56	Cav		0,11	0,11	
Norfentanyl	$C_{14}H_{20}N_2O$	232,1581	3,45	Cam	29,0	1,42	-1,82	
Lenacil	$C_{13}H_{18}N_2O_2$	234,1376	4,74	Cam/Cav	29,7	1,83	1,83	
Lidocaine	$C_{14}H_{22}N_2O$	234,1732	4,68	Cam/Cav	74,9	2,84	-0,64	
Carbamazepine	$C_{15}H_{12}N_2O$	236,0946	7,15	Cam/Cav	97,9	2,77	2,77	
Salbutamol	$C_{13}H_{21}NO_3$	239,1518	2,97	Cam/Cav	30,6	0,35	-2,36	
Cyanazine	$C_9H_{13}ClN_6$	240,0896	5,33	Cav	26,4	1,96	1,96	
Terbutryne/Prometryne	$C_{10}H_{19}N_5S$	241,1359	7,86	Cam	87,5	2,88/3,01	0,92/1,01	
Pindolol	$C_{14}H_{20}N_2O_2$	248,1535	3,48	Cav	25,0	1,69	-1,55	

Nom du suspect	Formule(s) Brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Sulfapyridine	**$C_{11}H_{11}N_3O_2S$**	**249,0576**	**3,96**	**Cav**	75,9	1,01	0,95	
Alprenolol/O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1724	4,72	Cam/Cav	71,9	3,10/1,72	3,10/1,72	
Alprenolol/O-Desmethyl-cis-tramadol	**$C_{15}H_{23}NO_2$**	**249,1724**	**5,24**	**Cam/Cav**	71,9	3,10/1,72	3,10/1,72	
Alprenolol/O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1724	4,45	Cam/Cav	71,9	3,10/1,72	3,10/1,72	
Alprenolol/O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1725	5,61	Cam/Cav	26,2	3,10/1,72	3,10/1,72	
Alprenolol/O-Desmethyl-cis-tramadol	$C_{15}H_{23}NO_2$	249,1726	3,84	Cam/Cav	53,5	3,10/1,72	3,10/1,72	
Carbamazepine 10,11- Epoxyde	$C_{15}H_{12}N_2O_2$	252,0894	6,74	Cav	87,5	1,97	1,97	
Carbamazepine 10,11- Epoxyde	$C_{15}H_{12}N_2O_2$	252,0894	7,06	Cav	87,5	1,97	1,97	
Cimetidine	**$C_{10}H_{16}N_6S$**	**252,1154**	**3,52**	**Cav**	**72,2**	-0,11	-1,17	
Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	253,0522	5,20	Cam/Cav	86	0,79	0,75	
Lamotrigine	$C_9H_7Cl_2N_5$	255,0073	5,57	Cam/Cav	97,1	1,93	-0,66	
Imidacloprid	$C_9H_{10}ClN_5O_2$	255,0517	5,39	Cav	23,3	0,87	-1,53	
Propranolol	$C_{16}H_{21}NO_2$	259,1569	6,11	Cam/Cav	97,6	2,58	-0,66	
Tramadol/Desvenlafaxine	$C_{16}H_{25}NO_2$	263,1880	4,87	Cam/Cav	41,7	2,45/2,29	-1,05/-0,91	
Tramadol/Desvenlafaxine	$C_{16}H_{25}NO_2$	263,1880	5,35	Cam/Cav	81,1	2,45/2,29	-1,05/-0,91	
Anisomycin	$C_{14}H_{19}NO_4$	265,1311	1,59	Cav	4,3	0,73	-2,51	
Mirtazapine	$C_{17}H_{19}N_3$	265,1580	5,13	Cav	73,5	3,21	-1,25	
Oxprenolol	$C_{15}H_{23}NO_3$	265,1674	3,78	Cam/Cav	60,4	2,17	-1,07	
Oxprenolol	$C_{15}H_{23}NO_3$	265,1677	5,07	Cam/Cav	34,0	2,17	-1,07	
Atenolol	**$C_{14}H_{22}N_2O_3$**	**266,1627**	**1,25**	**Cam/Cav**	**87,0**	0,43	-2,82	
Zidovudine	$C_{10}H_{13}N_5O_4$	267,0963	1,78	Cam/Cav	33,9	-0,30	-0,41	
Acide aténolique	$C_{14}H_{21}NO_4$	267,1472	4,43	Cam/Cav	85,3	-1,24	-1,69	
Metoprolol	$C_{15}H_{25}NO_3$	267,1828	5,39	Cam/Cav	94,9	1,76	-1,48	
Sotalol	$C_{12}H_{20}N_2O_3S$	272,1200	2,84	Cam/Cav	89,7	-0,39	-3,19	
Galaxolidone/Estradiol	$C_{18}H_{24}O_2$	272,1767	9,53	Cav	89,7	4,65/3,75	4,65/3,75	
Amitriptyline/Maprotiline	$C_{20}H_{23}N$	277,1827	6,67	Cam/Cav	**53,3**	4,81/4,63	1,31/1,13	
Amitriptyline/Maprotiline	**$C_{20}H_{23}N$**	**277,1829**	**7,54**	**Cav**	**85,1**	4,81/4,63	1,31/1,13	
Venlafaxine	$C_{17}H_{27}NO_2$	277,2037	6,39	Cam/Cav	91,7	2,74	-0,76	
Venlafaxine	**$C_{17}H_{27}NO_2$**	**277,2056**	**5,36**	**Cav**	**30,8**	2,74	-0,76	

Nom du suspect	Formule(s) Brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Sulfamethazine	$C_{12}H_{14}N_4O_2S$	278,0822	5,57	Cav	33,9	0,65	0,61	
Pentoxifylline	$C_{13}H_{18}N_4O_3$	278,1366	5,01	Cam	32,3	0,23	0,23	
Tramadol N-Oxide	$C_{16}H_{25}NO_3$	279,1831	4,32	Cam/Cav		1,33	1,24	
Tramadol N-Oxide	$C_{16}H_{25}NO_3$	279,1839	5,03	Cav		1,33	1,24	
Cycloheximide	$C_{15}H_{23}NO_4$	281,1628	4,31	Cam/Cav	**20,3**	0,91	0,91	
Metolachlor	$C_{15}H_{22}ClNO_2$	283,1334	8,86	Cam/Cav	96,5	3,45	3,45	
Morphine/Norcodeine/Hydromorphe	$C_{17}H_{19}NO_3$	285,1362	3,64	Cav	68,9	0,90	-2,30	
Oxazepam	$C_{15}H_{11}ClN_2O_2$	286,0499	7,98	Cam/Cav	90,2	2,92	2,92	
Sulfacytine	$C_{12}H_{14}N_4O_3S$	294,0787	3,00	Cam/Cav	2,3	0,06/1,39	-0,01/-0,25	
Diclofenac	$C_{14}H_{11}Cl_2NO_2$	295,0169	8,97	Cav	88,2	4,26	4,22	
Metoclopramide	$C_{14}H_{22}ClN_3O_2$	299,1396	4,93	Cav	86,7	1,40	-2,12	
Codeine	$C_{18}H_{21}NO_3$	299,1516	3,33	Cam/Cav	85,8	1,34	-2,16	
Codeine	$C_{18}H_{21}NO_3$	299,1525	4,38	Cav	30,0	1,34	-2,16	
Cocaine/Hydromorphinol	$C_{17}H_{21}NO_4$	303,1490	3,90	Cav	83,0	2,28	-1,22	
Bestatin	**$C_{16}H_{24}N_2O_4$**	**308,1736**	**4,35**	**Cam/Cav**	**50,0**	-1,1	-1,62	
Nadolol	$C_{17}H_{27}NO_4$	309,1945	4,18	Cav	**44,8**	0,87	-2,38	
Desloratadine	$C_{19}H_{19}ClN_2$	310,1257	4,57	Cam/Cav	6,3	3,97	-0,38	
Desmethylcitalopram	**$C_{19}H_{19}FN_2O$**	**310,1481**	**6,69**	**Cav**	**71,3**	5,24	4,65	
Kresoxim-methyl	$C_{18}H_{19}NO_4$	313,1306	8,58	Cav	49,3	4,34	4,34	
Imiprothrin	$C_{17}H_{22}N_2O_4$	318,1598	7,02	Cav		1,51	1,51	
Gliclazide	$C_{15}H_{21}N_3O_3S$	323,1304	8,13	Cav	85,3	1,73	1,69	
Citalopram	$C_{20}H_{21}FN_2O$	324,1631	6,71	Cav	53,3	3,76	0,26	
Bisoprolol	$C_{18}H_{31}NO_4$	325,2250	6,27	Cam/Cav	86,6	2,20	-1,05	
Clozapine	$C_{18}H_{19}ClN_4$	326,1294	6,44	Cav	81,6	3,40	-0,96	
PEG EO7 (HO-PEG7-HO)	$C_{14}H_{30}O_8$	326,1934	4,85	Cam/Cav	82,5	-1,49	-1,49	
DCHP	$C_{20}H_{26}O_4$	330,1808	9,16	Cam	37,7	5,58	5,58	
Acebutolol	$C_{18}H_{28}N_2O_4$	336,2041	5,42	Cam/Cav	85,8	1,53	1,53	
Linezolid	$C_{16}H_{20}FN_3O_4$	337,1424	4,80	Cav	43,1	0,64	0,64	
Linezolid	$C_{16}H_{20}FN_3O_4$	337,1440	6,16	Cav	80,9	0,64	0,64	

Nom du suspect	Formule(s) Brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Log P	Log D (pH 3)	C
Propiconazole	$C_{15}H_{17}Cl_2N_3O_2$	341,0690	9,13	Cam/Cav	66,4	4,33	4,29	
Ciclacillin	$C_{15}H_{23}N_3O_4S$	341,1409	3,29	Cam/Cav	51,6	-1,92	-2,26	
Griseofulvin	$C_{17}H_{17}ClO_6$	352,0712	7,88	Cav	73,8	2,17	2,17	
Tebufenozide	$C_{22}H_{28}N_2O_2$	352,2155	8,25	Cam/Cav	32,9	5,35	5,35	
Ofloxacin/Levofloxacine	$C_{18}H_{20}FN_3O_4$	361,1435	4,81	Cav	78,5	0,09/0,09	-1,72/-1,72	
PEG EO8 (HO-PEG8-OH)	$C_{16}H_{34}O_9$	370,2196	5,13	Cam/Cav	82,4	-1,54	-1,54	
Celiprolol	$C_{20}H_{33}N_3O_4$	379,2469	5,83	Cam/Cav	89,7	1,50	-1,74	
Alfuzosine	$C_{19}H_{27}N_5O_4$	389,2058	6,04	Cav	75,2	1,53	1,53	
PEG EO9 (HO-PEG9-OH)	$C_{18}H_{38}O_{10}$	414,2459	5,36	Cam/Cav	10,0	-1,58	-1,58	
Irbesartan	$C_{25}H_{28}N_6O$	428,2320	8,04	Cam	97,6	5,39	4,37	
Cloxacillin	$C_{19}H_{18}ClN_3O_5S$	435,0651	7,35	Cav	12,9	2,30	2,23	
Tembotrione	$C_{17}H_{16}ClF_3O_6S$	440,0323	8,16	Cav	66,1	2,59	2,59	
Candesartan	$C_{24}H_{20}N_6O_3$	440,1585	8,18	Cav	65,2	5,31	5,18	
PEG EO10 (HO-PEG10-HO)	$C_{20}H_{42}O_{11}$	458,2718	5,58	Cam/Cav	28,2	-1,63	-1,63	
Fexofenadine	$C_{32}H_{39}NO_4$	501,2874	7,41	Cam/Cav	88,2	2,94	2,32	
PEG EO11 (HO-PEG11-HO)	$C_{22}H_{46}O_{12}$	502,2980	5,76	Cam/Cav	79,3	-1,68	-1,68	
Telmisartan	$C_{33}H_{30}N_4O_2$	514,2373	7,90	Cav	89,1	6,13	5,19	
Clarithromycin	$C_{38}H_{69}NO_{13}$	747,4744	7,97	Cav	81,4	3,24	-0,26	
Azithromycin	$C_{38}H_{72}N_2O_{12}$	748,5080	6,12	Cam/Cav	86,5	2,45	-4,56	
Roxithromycin	$C_{41}H_{76}N_2O_{15}$	836,5268	7,97	Cav	81,8	3,00	-0,58	
Megalomicin A	$C_{44}H_{80}N_2O_{15}$	876,5490	6,09	Cam/Cav	25,9	2,84	-4,16	

Annexe 12

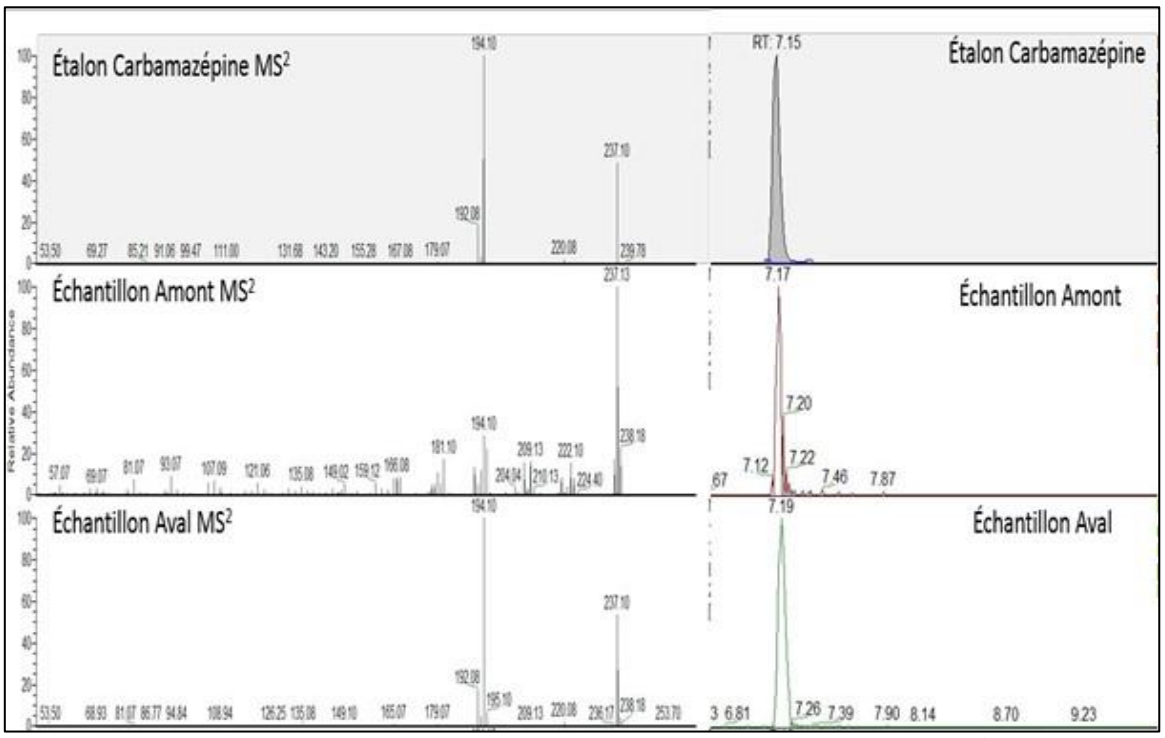


Figure 94 : Chromatogrammes de l'ion moléculaire de la carbamazépine et spectres MS² de l'étalon de la carbamazépine et d'échantillons amont et aval du Clain pour la confirmation de l'identification.

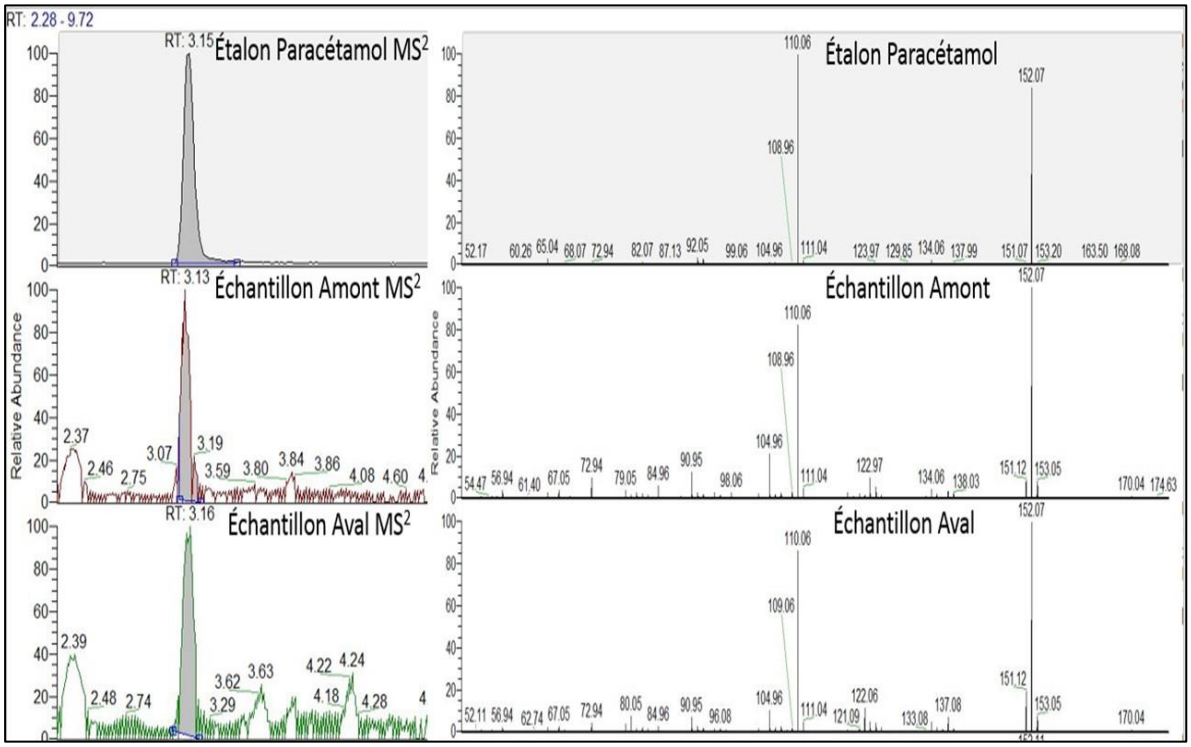


Figure 95 : Chromatogrammes de l'ion moléculaire de la carbamazépine et spectres MS² de l'étalon du paracétamol et d'échantillons amont et aval du Clain pour la confirmation de l'identification.

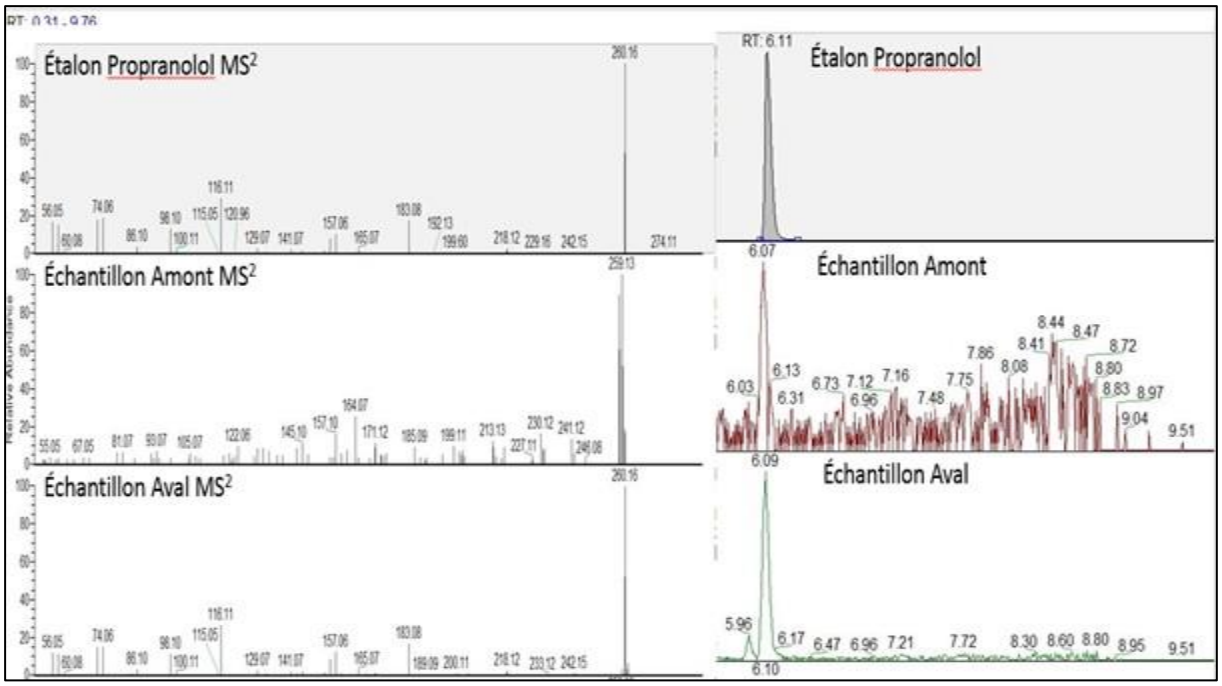


Figure 96 : Chromatogrammes de l'ion moléculaire du propranolol et spectres MS^2 de l'étalon du paracétamol et d'échantillons amont et aval du Clain pour la confirmation de l'identification.

Annexe 13

Tableau 36 : Liste des 34 signaux correspondant à un ou plusieurs métabolites avec leurs formules brutes, la masse exacte et le temps de rétention (Tr) du signal détecté, le pourcentage de la meilleure correspondance avec mzCloud, les logP et logD (pH = 3), la nature de l'échantillon dans lequel le suspect est présent et un code couleur correspondant à : une correspondance avec mzCloud supérieure à 80% en bleu foncé et inférieure à 80% en bleu clair, une correspondance avec les bases de données de MetFrag (KEGG, CompTox, HMDB) inférieure à 50% en vert clair (aucun signal n'a présenté de correspondance supérieure à 50%), pas de correspondance avec une base de données en jaune. Les signaux en gras sont également des signaux de la liste des signaux pertinents. C : code couleur. Aucun suspect n'a été confirmé.

Nom du suspect	Formule(s) Brute	Masse exacte (Da)	Tr (min)	Nature	MC (%)	Molécule parente	Log P	Log D (pH 3)	C
para-phenylenediamine	$C_6H_8N_2$	108,0690	1,21	Cam/Cav	84,4	4'-Aminoacetanilide	0,32	-2,22	
Para-aminophenol	C_6H_7NO	109,0531	1,27	Cam/Cav	83,8	Paracétamol	0,84	-1,19	
para-hydroxybenzoic acid	$C_7H_6O_3$	138,0315	6,69	Cav	19,5	Méthylparaben/ Propylparaben	1,33	1,33	
7-hydroxy-2-azaspiro[4.5]decan-3-one	$C_9H_{15}NO_2$	169,1100	1,12	Cam/Cav		Gabapentine-Lactam	-0,35	-0,35	
(2R,3R,4S,6R)-4-(dimethylamino)-6-methyloxane-2,3-diol	$C_8H_{17}NO_3$	175,1205	1,08	Cam/Cav		Azithromycin	-0,49	-3,99	
(2R,3R,4S,6R)-4-(dimethylamino)-6-methyloxane-2,3-diol	$C_8H_{17}NO_3$	175,1208	1,39	Cam		Azithromycin	-0,06	-0,06	
nicotine 1-N-oxide	$C_{10}H_{14}N_2O$	178,1104	4,10	Cam/Cav	36,7	Nicotine	0,04	-0,90	
2-hydroxypropyl 4-hydroxybenzoate	$C_{10}H_{12}O_4$	196,0731	5,34	Cam/Cav		Propylparaben	1,40	1,40	
1-amino-3-[4-(2-methoxyethyl)phenoxy]propan-2-ol	$C_{12}H_{19}NO_3$	225,1364	2,03	Cav		Métoprolol	0,55	-2,48	

N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)acetamide	$C_{12}H_{13}N_3O_2$	231,1008	3,24	Cav		Acetylaminoantipyrine	0,21	0,21	
N-(3-methyl-5-oxo-1-phenyl-2H-pyrazol-4-yl)acetamide	$C_{12}H_{13}N_3O_2$	231,1010	4,56	Cav		Acetylaminoantipyrine	0,21	0,21	
Nom du suspect	**Formule(s) Brute**	**Masse exacte (Da)**	**Tr (min)**	**Nature**	**MC (%)**	**Molécule parente**	**Log P**	**Log D (pH 3)**	**C**
4-acetaminoantipyrine	$C_{13}H_{15}N_3O_2$	245,1167	4,63	Cam/Cav	79,9	4-Aminoantipyrine/ 4-Formylaminoantipyrine	0,15	0,15	
N-desmethyltramadol/ O-desmethyltramadol	$C_{15}H_{23}NO_2$	249,1724	4,72	Cam/Cav	42,8	Tramadol	2,07/ 1,72	-1,18/-1,20	
N-desmethyltramadol/ O-desmethyltramadol	**$C_{15}H_{23}NO_2$**	**249,1724**	**5,32**	**Cam/Cav**	**61,3**	**Tramadol**	**2,07/ 1,72**	**-1,18/-1,20**	
N-desmethyltramadol/ O-desmethyltramadol	$C_{15}H_{23}NO_2$	249,1724	4,45	Cam/Cav	76,2	Tramadol	2,07/ 1,72	-1,18/-1,20	
N-desmethyltramadol/ O-desmethyltramadol	$C_{15}H_{23}NO_2$	249,1725	5,61	Cam/Cav	49,2	Tramadol	2,07/ 1,72	-1,18/-1,20	
N-desmethyltramadol/ O-desmethyltramadol	$C_{15}H_{23}NO_2$	249,1740	3,84	Cam	61,3	Tramadol	2,07/ 1,72	-1,18/-1,20	
2-hydroxycarbamazepine	$C_{15}H_{12}N_2O_2$	252,0894	6,74	Cam/Cav	29,2	Carbamazépne	2,46	2,46	
2-hydroxycarbamazepine	$C_{15}H_{12}N_2O_2$	252,0894	7,06	Cav	28,3	Carbamazépine	2,46	2,46	
2-hydroxycarbamazepine	$C_{15}H_{12}N_2O_2$	252,0898	6,52	Cam/Cav	30,2	Carbamazépine	2,46	2,46	
N-Desmethylvenlafaxine/ O-Desmethylvenlafaxine	$C_{16}H_{25}NO_2$	263,1880	4,87	Cam/Cav	13,3	Venlafaxine	2,36/ 2,29	-0,89/-0,91	
N-desmethylvenlafaxine/ O-Desmethylvenlafaxine	$C_{16}H_{25}NO_2$	263,1880	5,35	Cam/Cav	83,3	Venlafaxine	2,36/ 2,29	-0,89/-0,91	
1-(isopropylamino)-3-[4-(2-methoxyethyl)phenoxy]propan-2-one	$C_{15}H_{23}NO_3$	265,1674	3,78	Cam/Cav		Métoprolol	2,07	-1,17	

1-(isopropylamino)-3-[4-(2-methoxyethyl)phenoxy]propan-2-one	$C_{15}H_{23}NO_3$	265,1677	5,07	Cam/Cav		Métoprolol	2,07	-1,17	
Nom du suspect	**Formule(s) Brute**	**Masse exacte (Da)**	**Tr (min)**	**Nature**	**MC (%)**	**Molécule parente**	**Log P**	**Log D (pH 3)**	**C**
Acide aténolique – acide métropolol	$C_{14}H_{21}NO_4$	267,1472	4,43	Cam/Cav	85,3	Métoprolol/Aténolol	-1,24	-1,69	
(1R,6R)-6-[(dimethylamino)methyl]-1-(3-methoxyphenyl)cyclohexane-1,3-diol	$C_{16}H_{25}NO_3$	279,1831	4,32	Cam/Cav		Tramadol	1,06	-2,44	
1-[(1R,2R)-2-hydroxy-2-(3-methoxyphenyl)cyclohexyl]-N,N-dimethylmethanamine oxide	$C_{16}H_{25}NO_3$	279,1834	6,96	Cav		Tramadol	1,33	1,24	
(1R,6R)-6-[(dimethylamino)methyl]-1-(3-methoxyphenyl)cyclohexane-1,3-diol	$C_{16}H_{25}NO_3$	279,1837	3,87	Cam/Cam		Tramadol	0,99	-2,51	
1-[(1R,2R)-2-hydroxy-2-(3-methoxyphenyl)cyclohexyl]-N,N-dimethylmethanamine oxide	$C_{16}H_{25}NO_3$	279,1839	5,03	Cam/Cav		Tramadol	1,33	1,24	
morphinone	$C_{17}H_{17}NO_3$	283,1215	5,07	Cam/Cav	39,1	Morphine	1,49	-1,69	
norcodeine	$C_{17}H_{19}NO_3$	285,1362	3,64	Cav	68,6	Codéine	0,96	-2,28	
1-[2-(dimethylamino)-1-(4-methoxyphenyl)ethyl]cyclohexane-1,3-diol	$C_{17}H_{27}NO_3$	293,1989	4,78	Cam/Cav		Venlafaxine	1,28	-2,22	
codeine N-oxide	$C_{18}H_{21}NO_4$	315,1467	3,94	Cav	35,6	Codéine	0,22	0,14	
desosaminylazithromycin	$C_{30}H_{58}N_2O_9$	590,4157	4,29	Cam		Azithromycin	1,34	-5,66	

Annexe 14

Tableau 37 : Liste des 36 signaux communs entre les 2 listes de signaux pertinents déterminés à partir de la comparaison amont-aval et à partir de sorties de STEP, associés à une proposition de nom par mzCloud lorsque celle-ci existe.

Proposition de structure(s)	Formule(s) Brute	Masse exacte (Da)	Tr (min)	MC (%)
N-Cyanoimido-S,S-dimethyl-dithiocarbonate	$C_4H_6N_2S_2$	145,9970	5,09	
Commun 2	$C_9H_{12}N_2O_2$	180,0893	4,86	
Commun 3	$C_6H_{11}N_3O_2S$	189,0569	7,64	
Commun 4	$C_6H_{11}N_3O_2S$	189,0570	7,74	
Commun 5	$C_7H_9N_5S$	195,0576	4,91	
Indole-3-butyric acid	$C_{12}H_{13}NO_2$	203,0944	4,79	76,0
3-ethyl-4-hydroxy-1-methyl-1,2-dihydroquinolin-2-one				71,6
N1-(5-oxo-5,6,7,8-tetrahydronaphthalen-2-yl)acetamide				65,0
Methsuximide				64,4
Commun 7	$C_{14}H_{10}N_2$	206,0834	7,84	
Commun 8	$C_{14}H_{10}N_2$	206,0834	7,31	
Commun 9	$C_7H_9N_5OS$	211,0524	5,73	
Commun 10	$C_{11}H_{21}NO_4$	231,1467	3,66	
Commun 11	$C_{16}H_{24}O_2$	248,1766	7,77	
Commun 12	$C_{14}H_{26}O_6$	290,1721	7,47	
Commun 13	$C_{16}H_{25}NO_4$	295,1777	3,90	
Commun 14	$C_{12}H_{20}N_6O_5$	328,1492	6,24	
Commun 15	$C_{13}H_{25}N_5O_3S$	331,1681	6,55	
Commun 16	$C_{21}H_{19}O_2P$	334,1117	5,46	
Commun 17	$C_{15}H_{28}N_6O_6$	388,2062	7,26	
Commun 18	$C_{10}H_{32}N_9O_3PS$	389,2094	7,25	
Commun 19	$C_{21}H_{33}NO_6$	395,2297	4,20	
Commun 20	$C_{18}H_{37}NO_8$	395,2512	7,41	
Commun 21	$C_{18}H_{36}N_6S_2$	400,2422	8,43	
Commun 22	$C_{16}H_{31}N_5O_3P_2$	403,1892	7,22	
Commun 23	$C_{23}H_{35}N_4O_3P$	446,2474	7,76	
Commun 24	$C_{18}H_{35}N_{11}O_3$	453,2918	7,89	
Commun 25	$C_{19}H_{44}N_6O_3S_2$	468,2920	7,84	
Commun 26	$C_{17}H_{39}N_7O_8$	469,2874	7,92	
Commun 27	$C_{19}H_{36}N_9P_3$	483,2306	3,83	
Commun 28	$C_{17}H_{41}N_6O_8P$	488,2708	7,85	
Commun 29	$C_{31}H_{48}N_3O_4P$	557,3389	8,17	
Commun 30	$C_{24}H_{48}O_{14}$	560,3029	4,96	
Commun 31	$C_{20}H_{42}N_6O_{13}$	574,2821	4,99	
Commun 32	$C_{17}H_{37}N_{19}O_7$	619,3123	5,18	
Commun 33	$C_{30}H_{55}NO_9P_2$	635,3349	5,18	
Commun 34	$C_{22}H_{36}N_6O_{16}$	640,2180	5,20	
Commun 35	$C_{11}H_2N_{13}O_{17}P_3$	680,8943	5,25	
Commun 36	$C_{23}H_{42}N_2O_{21}$	682,2274	5,27	

Annexe 15

Tableau 38 : Liste des signaux pertinents issus des approches 1 et 2 selon leur hiérarchisation « très probable » ou « probable ». Une ou des propositions de structures et formules sont associées aux signaux lorsqu'elles existent. Tr : temps de rétention

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approches 1-2	N-Cyanoimido-S,S-dimethyl-dithiocarbonate	$C_4H_6N_2S_2$	145,9970	5,47	Très probable
Approches 1-2		$C_9H_{12}N_2O_2$	180,0893	5,24	Très probable
Approches 1-2		$C_6H_{11}N_3O_2S$	189,0569	8,02	Très probable
Approches 1-2		$C_6H_{11}N_3O_2S$	189,0570	8,12	Très probable
Approches 1-2		$C_7H_9N_5S$	195,0576	5,29	Très probable
Approches 1-2		$C_{12}H_{13}NO_2$	203,0944	5,17	Très probable
Approches 1-2		$C_{14}H_{10}N_2$	206,0834	8,22	Très probable
Approches 1-2		$C_{14}H_{10}N_2$	206,0834	7,69	Très probable
Approches 1-2		$C_7H_9N_5OS$	211,0524	6,11	Très probable
Approches 1-2		$C_{11}H_{21}NO_4$	231,1467	4,04	Très probable
Approches 1-2		$C_{16}H_{24}O_2$	248,1766	8,15	Très probable
Approches 1-2		$C_{14}H_{26}O_6$	290,1721	7,85	Très probable
Approches 1-2		$C_{16}H_{25}NO_4$	295,1777	4,28	Très probable
Approches 1-2		$C_{12}H_{20}N_6O_5$	328,1492	6,62	Très probable
Approches 1-2		$C_{13}H_{25}N_5O_3S$	331,1681	6,93	Très probable
Approches 1-2		$C_{21}H_{19}O_2P$	334,1117	5,84	Très probable
Approches 1-2		$C_{15}H_{28}N_6O_6$	388,2062	7,64	Très probable
Approches 1-2		$C_{10}H_{32}N_9O_3PS$	389,2094	7,63	Très probable
Approches 1-2		$C_{21}H_{33}NO_6$	395,2297	4,58	Très probable
Approches 1-2		$C_{18}H_{37}NO_8$	395,2512	7,79	Très probable
Approches 1-2		$C_{18}H_{36}N_6S_2$	400,2422	8,80	Très probable
Approches 1-2		$C_{16}H_{31}N_5O_3P_2$	403,1892	7,60	Très probable
Approches 1-2		$C_{23}H_{35}N_4O_3P$	446,2474	8,14	Très probable
Approches 1-2		$C_{18}H_{35}N_{11}O_3$	453,2918	8,27	Très probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approches 1-2		$C_{19}H_{44}N_6O_3S_2$	468,2920	8,22	Très probable
Approches 1-2		$C_{17}H_{39}N_7O_8$	469,2874	8,30	Très probable
Approches 1-2		$C_{19}H_{36}N_9P_3$	483,2306	4,21	Très probable
Approches 1-2		$C_{17}H_{41}N_6O_8P$	488,2708	8,23	Très probable
Approches 1-2		$C_{31}H_{48}N_3O_4P$	557,3389	8,55	Très probable
Approches 1-2		$C_{24}H_{48}O_{14}$	560,3029	5,34	Très probable
Approches 1-2		$C_{20}H_{42}N_6O_{13}$	574,2821	5,37	Très probable
Approches 1-2		$C_{17}H_{37}N_{19}O_7$	619,3123	5,56	Très probable
Approches 1-2		$C_{30}H_{55}NO_9P_2$	635,3349	5,56	Très probable
Approches 1-2		$C_{22}H_{36}N_6O_{16}$	640,2180	5,58	Très probable
Approches 1-2		$C_{11}H_2N_{13}O_{17}P_3$	680,8943	5,63	Très probable
Approches 1-2		$C_{23}H_{42}N_2O_{21}$	682,2274	5,65	Très probable
Approche 1		$C_7H_6N_2$	118,0533	1,91	Probable
Approche 1		C_9H_{10}	118,0784	4,83	Probable
Approche 1		C_6H_5NS	123,0145	3,59	Probable
Approche 1	5-amino-o-crésol	C_7H_9NO	123,0685	1,52	Probable
	2-amino-4-crésol				
	2-anisidine				
	2-pyridine ethanol				
	6-methyl-2-pyridinemethanol				
Approche 1		C_5H_8NOP	129,0339	2,94	Probable
Approche 1		$C_7H_{15}NO$	129,1153	1,17	Probable
Approche 1		$C_8H_{15}NO$	141,1153	6,77	Probable
Approche 1		$C_{10}H_9N$	143,0735	3,47	Probable
Approche 1	Tropine	$C_8H_{15}NO$	145,1117	6,17	Probable
	3,3,5,5-Tetramethylpyrroline-N-oxide				
Approche 1		C_8H_6OS	150,0139	5,19	Probable
Approche 1		$C_7H_9N_3O$	151,0745	3,17	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_9H_{15}NO$	153,1152	6,81	Probable
Approche 1		$C_9H_{15}NO$	153,1153	7,09	Probable
Approche 1		$C_7H_9NO_3$	155,0581	5,06	Probable
Approche 1		$C_8H_{19}NO_2$	161,1414	1,16	Probable
Approche 1		$C_{13}H_6$	162,0468	8,05	Probable
Approche 1	2-(Cyclopropylcarbonyl)-3-(dimethylamino)acrylonitrile	$C_9H_{12}N_2O$	164,0949	1,64	Probable
Approche 1	Edrophonium Perillartine	$C_{10}H_{15}NO$	165,1154	8,28	Probable
Approche 1		$C_8H_{10}N_2O_2$	166,0740	4,65	Probable
Approche 1		$C_5H_5N_5O_2$	167,0443	3,55	Probable
Approche 1	(4E)-3-Hydroxy-2,4-dimethyl-4-heptenamide	$C_9H_{17}NO_2$	171,1268	6,23	Probable
Approche 1	α-Methyltryptamine 1(1H-indol-5-yl)-2-propanamine 1(1H-indol-6-yl)-2-propanamine	$C_{12}H_{13}NO_2$	174,1161	4,53	Probable
Approche 1		$C_8H_{17}NO_3$	175,1207	1,18	Probable
Approche 1		$C_{13}H_9N$	179,0733	8,04	Probable
Approche 1	2-Amino-6-methylmercaptopurine 2-(methylthio)[1,2,4]triazolo[1,5-a]pyrimidin-7-amine	$C_6H_7N_5S$	181,0425	3,62	Probable
Approche 1		$C_{12}H_{14}N_2$	186,1154	8,36	Probable
Approche 1		$C_{14}H_5N$	187,0416	8,06	Probable
Approche 1		$C_{11}H_{13}N_3$	187,1108	3,36	Probable
Approche 1	4-Methyl-N,N-dimethylcathinone	$C_{12}H_{17}NO$	191,1310	7,87	Probable
Approche 1		$C_8H_{11}N_5O$	193,0963	4,00	Probable
Approche 1		$C_8H_{11}N_3O_3$	197,0799	3,17	Probable
Approche 1	5-Phenylnicotinic acid	$C_{12}H_9NO_2$	199,0634	2,50	Probable
Approche 1		$C_{13}H_{17}NO$	203,1306	8,51	Probable
Approche 1		$C_{11}H_{12}N_2O_2$	204,0896	3,21	Probable
Approche 1		$C_{11}H_{12}N_2O_2$	204,0897	4,76	Probable
Approche 1		$C_{14}H_{20}O$	204,1514	7,46	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_{11}H_{15}N_3O$	205,1213	2,35	Probable
Approche 1		$C_{14}H_{10}N_2$	206,0689	3,28	Probable
Approche 1	4-morpholinobenzoic acid	$C_{11}H_{13}NO_3$	207,0930	2,46	Probable
Approche 1		$C_9H_{21}NO_4$	207,1470	1,12	Probable
Approche 1		$C_8H_8N_4O_3$	208,0593	4,23	Probable
Approche 1		$C_9H_{11}N_3O_3$	209,0799	3,93	Probable
Approche 1		$C_9H_{10}N_2O_2S$	210,0463	3,99	Probable
Approche 1		$C_{10}H_{17}N_3O_2$	211,1318	1,17	Probable
Approche 1		$C_{13}H_8ClN$	213,0338	9,12	Probable
Approche 1		$C_{10}H_{19}NO_4$	217,1311	3,79	Probable
Approche 1		$C_{11}H_{10}N_2O_3$	218,0691	2,66	Probable
Approche 1		$C_{11}H_{12}N_2O_3$	220,0846	3,74	Probable
Approche 1		$C_{10}H_7NO_5$	221,0321	1,37	Probable
Approche 1	Tapentadol	$C_{14}H_{23}NO$	221,1773	8,78	Probable
Approche 1		$C_{12}H_{18}N_2O_2$	222,1365	7,28	Probable
Approche 1		$C_{10}H_{13}N_3O_3$	223,0954	2,85	Probable
Approche 1		$C_{16}H_{22}O$	230,1664	8,68	Probable
Approche 1		$C_{16}H_{22}O$	230,1665	8,54	Probable
Approche 1		$C_9H_{13}NO_4S$	231,0564	3,26	Probable
Approche 1		$C_{11}H_9N_3O_3$	231,0641	2,11	Probable
Approche 1		$C_8H_{12}N2O_2S_2$	232,0337	3,24	Probable
Approche 1		$C_{16}H_{24}O$	232,1822	8,82	Probable
Approche 1		$C_{15}H_9NO_2$	235,0631	7,17	Probable
Approche 1		$C_9H_8N_4O_4$	236,0542	3,35	Probable
Approche 1		$C_8H_9N_5O_2S$	239,0474	5,34	Probable
Approche 1	Desmethyl nefopam	$C_{16}H_{17}NO$	239,1307	5,90	Probable
Approche 1		$C_{11}H_{11}N_5O_2$	245,0910	4,63	Probable
Approche 1		$C_{14}H_{17}NO_3$	247,1207	3,93	Probable
Approche 1		$C_{16}H_{24}O_2$	248,1768	8,38	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_{16}H_{24}O_2$	248,1768	8,54	Probable
Approche 1		$C_{16}H_{24}O_2$	248,1769	8,77	Probable
Approche 1	Alprenolol	$C_{15}H_{23}NO_2$	249,1726	4,36	Probable
Approche 1		$C_{11}H_{10}N_2O_5$	250,0587	1,36	Probable
Approche 1		$C_{16}H_{26}O_2$	250,1927	9,65	Probable
Approche 1		$C_{11}H_{11}NO_6$	253,0582	1,35	Probable
Approche 1		$C_9H_6Cl_2N_4O$	255,9915	6,75	Probable
Approche 1		$C_7H_{12}N_6O_5$	260,0869	3,02	Probable
Approche 1		$C_9H_{12}NO_4PS$	261,0215	2,93	Probable
Approche 1		$C_{14}H_{15}NO_4$	261,0997	3,56	Probable
Approche 1		$C_{14}H_{15}NO_4$	261,1001	1,35	Probable
Approche 1			263,1053	5,08	Probable
Approche 1	O-desmethylvenlafaxine N-Desmethylvenlafaxine	$C_{16}H_{25}NO_2$	263,1881	5,06	Probable
Approche 1	Oxprenolol	$C_{15}H_{23}NO_3$	265,1676	4,76	Probable
Approche 1		$C_{12}H_{18}N_2O_3S$	270,1032	4,60	Probable
Approche 1		$C_{14}H_{13}N_3OS$	271,0774	5,70	Probable
Approche 1		$C_9H_{12}N_4O_4S$	272,0575	3,55	Probable
Approche 1	Galaxolidone Estradiol 19-Norandrostenedione	$C_{18}H_{24}O_2$	272,1767	9 ,23 8,87	Probable
Approche 1		$C_{13}H_{12}N_3O_2P$	273,0669	2,05	Probable
Approche 1			274,1062	3,55	Probable
Approche 1		$C_{14}H_{17}N_3O_3$	275,1265	3,11	Probable
Approche 1		$C_{17}H_{12}O_4$	280,0727	8,53	Probable
Approche 1		$C_{16}H_{12}NO_2P$	281,0593	8,01	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_{13}H_{15}NO_6$	281,0895	2,07	Probable
Approche 1	Cycloheximide	$C_{15}H_{23}NO_4$	281,1622	4,05	Probable
Approche 1			287,0826	3,68	Probable
Approche 1		$C_{14}H_{28}O_6$	292,1878	8,15	Probable
Approche 1		$C_{15}H_{11}N_5O_2$	293,0944	1,19	Probable
Approche 1		$C_{14}H_{11}Cl_2NO_2$	295,0157	9,12	Probable
Approche 1		$C_{13}H_{15}NO_7$	297,0858	5,28	Probable
Approche 1	Codeine Hydrocodone	$C_{18}H_{21}NO_3$	299,1524	2,95	Probable
Approche 1		$C_{11}H_{23}N_4P_3$	304,1132	3,92	Probable
Approche 1			306,1675	6,62	Probable
Approche 1		$C_{15}H_{30}O_6$	306,2031	8,18	Probable
Approche 1			307,1163	2,78	Probable
Approche 1		$C_{10}H_{26}N_7O_2P$	307,1890	6,97	Probable
Approche 1			309,1433	8,63	Probable
Approche 1		$C_8H_{10}N_{10}O_4$	310,0891	3,98	Probable
Approche 1		$C_{14}H_{25}N_4PS$	312,1542	7,86	Probable
Approche 1	Kresoxim méthyle	$C_{18}H_{19}NO_4$	313,1310	8,29	Probable
Approche 1		$C_{12}H_{22}N_6O_4$	314,1703	6,55	Probable
Approche 1		$C_7H_{14}N_{11}O_2P$	315,1076	5,23	Probable
Approche 1		$C_{13}H_{21}NO_8$	319,1276	2,41	Probable
Approche 1		$C_{18}H_{25}NO_4$	319,1781	7,03	Probable
Approche 1			321,0581	8,62	Probable
Approche 1		$C_{15}H_{30}O_7$	322,1985	7,47	Probable
Approche 1		$C_{14}H_{29}NO_7$	323,1938	3,66	Probable
Approche 1		$C_{14}H_{23}N_4O_3P$	326,1495	6,67	Probable
Approche 1		$C_{13}H_{25}N_5OP_2$	329,1531	6,64	Probable
Approche 1		$C_{10}H_8N_{10}O_4$	332,0716	3,98	Probable
Approche 1		$C_{18}H_{20}O_6$	332,1248	8,07	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1			332,1731	3,48	Probable
Approche 1		$C_{14}H_{29}NO_8$	339,1890	2,01	Probable
Approche 1		$C_{13}H_{24}O_{10}$	340,1361	4,46	Probable
Approche 1		$C_{13}H_{30}N_2O_4S_2$	342,1648	8,14	Probable
Approche 1		$C_{13}H_{24}N_6O_5$	344,1806	7,46	Probable
Approche 1		$C_{11}H_{16}N_{11}OP$	349,1268	3,97	Probable
Approche 1		$C_{13}H_{22}O_{11}$	354,1152	4,49	Probable
Approche 1		$C_{17}H_{29}NO_3P_2$	357,1626	4,45	Probable
Approche 1		$C_{19}H_{24}O_3P_2$	362,1183	4,41	Probable
Approche 1		$C_{17}H_{34}O_8$	366,2246	7,67	Probable
Approche 1		$C_{18}H_{24}N_6O_3$	372,1909	7,16	Probable
Approche 1		$C_{11}H_{16}N_6O_9$	376,0978	4,5	Probable
Approche 1		$C_{15}H_{32}N_5O_4P$	377,2189	4,57	Probable
Approche 1		$C_{21}H_{34}O_2P_2$	380,2041	7,60	Probable
Approche 1		$C_{22}H_{38}OP_2$	380,2398	8,04	Probable
Approche 1		$C_{26}H_{38}S$	382,2684	9,62	Probable
Approche 1		$C_{19}H_{30}O_4P_2$	384,1625	4,82	Probable
Approche 1		$C_{18}H_{33}N_4OPS$	384,2114	8,75	Probable
Approche 1			385,1935	5,46	Probable
Approche 1		$C_{19}H_{28}N_7P$	385,2139	8,24	Probable
Approche 1		$C_{21}H_{26}N_2O_5$	386,1848	7,28	Probable
Approche 1		$C_{23}H_{32}OP_2$	386,1912	7,43	Probable
Approche 1		$C_{16}H_{30}N_6O_5$	386,2276	7,87	Probable
Approche 1			387,2095	4,77	Probable
Approche 1		$C_{17}H_{17}N_3O_6S$	391,0833	6,71	Probable
Approche 1			393,1685	4,77	Probable
Approche 1		$C_{14}H_{17}N_{15}$	395,1786	3,25	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_{21}H_{33}NO_6$	397,2305	7,62	Probable
Approche 1	N,N-Dimethyl-4-{5-[(3S)-1-(phenylsulfonyl)-3-pyrrolidinyl]-1,3,4-oxadiazol-2-yl}aniline	$C_{20}H_{22}N_4O_3S$	398,1413	4,42	Probable
	AB-FUBINACA 2A metabolite	$C_{20}H_{19}FN_4O_4$			
Approche 1		$C_{11}H_{33}N_{10}O_2PS$	400,2246	8,03	Probable
Approche 1		$C_{19}H_{28}N_7OP$	401,2097	7,84	Probable
Approche 1		$C_{16}H_{37}O_5P_3$	402,1860	7,61	Probable
Approche 1		$C_{22}H_{24}N_6O_2$	404,1957	7,32	Probable
Approche 1		$C_{21}H_{36}O_4P_2$	414,2095	5,06	Probable
Approche 1		$C_{10}H_{13}Cl_3N_6O_6$	417,9957	5,33	Probable
Approche 1		$C_{13}H_{20}N_6O_{10}$	420,1233	4,85	Probable
Approche 1		$C_{23}H_{38}O_3P_2$	424,2303	7,79	Probable
Approche 1		$C_{14}H_{24}N_{10}O_6$	428,1885	5,10	Probable
Approche 1		$C_{14}H_{17}Cl_2NO_{10}$	429,0237	5,54	Probable
Approche 1			436,2662	8,31	Probable
Approche 1		$C_{24}H_{40}O_3P_2$	438,2451	8,10	Probable
Approche 1		$C_{25}H_{44}O_2P_2$	438,2817	8,46	Probable
Approche 1	Candesartan	$C_{24}H_{20}N_6O_3$	440,1586	7,92	Probable
Approche 1		$C_{18}H_{35}NO_{11}$	441,2205	3,84	Probable
Approche 1		$C_{25}H_{26}N_6O_2$	442,2103	8,04	Probable
Approche 1		$C_{10}H_{31}N_{13}O_3P_2$	443,2132	8,05	Probable
Approche 1			445,58569	8,17	Probable
Approche 1		$C_{25}H_{36}O_3P_2$	446,2121	7,78	Probable
Approche 1		$C_{20}H_{41}NO_{10}$	455,2718	8,10	Probable
Approche 1		$C_{18}H_{46}N_6O_3P_2$	456,3114	8,46	Probable
Approche 1		$C_{21}H_{27}N_7O_5$	457,2082	8,68	Probable
Approche 1		$C_{22}H_{34}N_6OP_2$	460,2271	8,10	Probable
Approche 1		$C_{20}H_{37}N_3O_9$	463,2524	4,18	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_{17}H_{41}N_7O_2P_2S$	469,2519	4,18	Probable
Approche 1		$C_{15}H_{38}N_9O_6P$	471,2676	4,10	Probable
Approche 1		$C_{23}H_{38}O_6P_2$	472,2147	5,32	Probable
Approche 1		$C_{21}H_{47}O_5P_3$	472,2637	9,03	Probable
Approche 1		$C_{18}H_{38}N_2O_{12}$	474,2430	8,33	Probable
Approche 1		$C_{13}H_{26}N_{10}O_8S$	482,2709	8,24	Probable
Approche 1		$C_{30}H_{39}N_2PS$	490,2555	9,81	Probable
Approche 1		$C_{24}H_{46}O_{10}$	494,3078	8,67	Probable
Approche 1		$C_{27}H_{46}O_4P_2$	496,2871	8,48	Probable
Approche 1		$C_{22}H_{47}BrN_2O_5$	498,2669	5,57	Probable
Approche 1		$C_{28}H_{48}O_4P_2$	510,3022	8,67	Probable
Approche 1			511,3341	8,67	Probable
Approche 1		$C_{20}H_{46}N_5O_4P_3$	513,2778	4,47	Probable
Approche 1		$C_{34}H_{30}N_2O_3$	514,2252	8,92	Probable
Approche 1		$C_{10}H_{33}N_{17}O_8$	519,2719	8,47	Probable
Approche 1		$C_{21}H_{40}N_9OP_3$	527,2570	4,52	Probable
Approche 1		$C_{26}H_{44}O_3P_2S_2$	530,2203	8,77	Probable
Approche 1		$C_{15}H_{39}N_{11}O_{10}$	533,2871	8,66	Probable
Approche 1		$C_{19}H_{39}N_7O_{11}$	541,2724	4,47	Probable
Approche 1		$C_{22}H_{46}N_6O_{10}$	554,3285	8,78	Probable
Approche 1		$C_{26}H_{48}N_2O_7P_2$	562,2951	8,58	Probable
Approche 1		$C_{27}H_{52}O_{12}$	568,3438	8,93	Probable
Approche 1		$C_{18}H_{29}Cl_3N_2O_6S_3$	570,0260	8,88	Probable
Approche 1		$C_{24}H_{45}NO_{14}$	571,2832	4,77	Probable
Approche 1		$C_{26}H_{40}O_{10}P_2$	574,2095	5,42	Probable
Approche 1		$C_{20}H_{46}N_8O_9S$	574,3113	8,78	Probable
Approche 1			586,37395	8,936	Probable
Approche 1		$C_{22}H_{45}N_7O_{12}$	599,3143	4,854	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 1		$C_{25}H_{52}N_6O_{11}$	612,37027	9,022	Probable
Approche 1		$C_{24}H_{53}N_7O_{11}$	615,38119	8,851	Probable
Approche 1		$C_{23}H_{42}N_{10}O_{10}$	618,3088	5,601	Probable
Approche 1		$C_{33}H_{47}N_8PS$	618,33733	8,838	Probable
Approche 1		$C_{25}H_{55}N_7O_{11}$	629,39598	9,006	Probable
Approche 1		$C_{23}H_{40}N_{10}O_{11}$	632,28805	5,635	Probable
Approche 1		$C_{16}H_{48}N_{18}O_5P_2$	634,35242	9,022	Probable
Approche 1		$C_{20}H_{51}N_{10}O_4PS_4$	654,27018	5,613	Probable
Approche 1		$C_{28}H_{58}N_6O_{12}$	670,41178	9,225	Probable
Approche 1		$C_{39}H_{61}N_2O_6P$	684,42687	9,337	Probable
Approche 1			686,78443	3,544	Probable
Approche 1			744,79004	4,098	Probable
Approche 1		$C_{28}H_{71}N_{14}O_4P_3$	760,49827	9,507	Probable
Approche 1		$C_{25}H_{66}N_{15}O_6P_3$	765,45378	9,524	Probable
Approche 1			774,80028	3,777	Probable
Approche 1			790,83157	3,261	Probable
Approche 1			804,81089	3,4	Probable
Approche 1			817,53539	9,627	Probable
Approche 1			822,49227	9,65	Probable
Approche 1			1242,71149	5,552	Probable
Approche 2			144,0686	6,61	Probable
Approche 2		$C_6H_9NO_2S$	159,0351	5,31	Probable
Approche 2		$C_5H_{10}N_2O_2S$	162,0466	7,92	Probable
Approche 2		$C_9H_9NO_2$	163,0631	5,44	Probable
Approche 2		$C_{13}H_9N$	179,0731	8,02	Probable
Approche 2		$C_{13}H_9N$	179,0732	8,63	Probable
Approche 2		$C_{13}H_9N$	179,0732	6,58	Probable
Approche 2		$C_9H_{12}N_2O_2$	180,0896	5,67	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 2		$C_9H_{12}N_2O_2$	180,0897	5,39	Probable
Approche 2		$C_9H_{11}NO_3$	181,0734	6,81	Probable
Approche 2		$C_9H_{11}NO_3$	181,0736	1,25	Probable
Approche 2		$C_{13}H_{11}N$	181,0889	4,22	Probable
Approche 2		$C_9H_{15}N_3O$	181,1212	5,31	Probable
Approche 2	Iminostilbene 2-Phenylindole	$C_{14}H_{11}N$	193,0897	7,64	Probable
Approche 2		$C_{10}H_{14}N_2O_2$	194,1053	5,39	Probable
Approche 2		$C_5H_{13}ClN_2O_2S$	200,0387	7,93	Probable
Approche 2		$C_9H_{10}N_2O_4$	210,0637	1,33	Probable
Approche 2		$C_{12}H_{12}N_2O_2$	216,0895	4,72	Probable
Approche 2		$C_7H_{14}O_6S$	226,0519	5,38	Probable
Approche 2		$C_{11}H_{21}NO_4$	231,1004	3,37	Probable
Approche 2		$C_{10}H_9N_3O_4$	235,0589	4,21	Probable
Approche 2		$C_8H_{22}N_4S_2$	238,1293	7,47	Probable
Approche 2		$C_{10}H_{11}NO_4S$	241,0400	4,61	Probable
Approche 2		$C_6H_{17}N_8OP$	248,1254	5,53	Probable
Approche 2	N-desmethyltramadol O-desmethyltramadol Alprenolol	$C_{15}H_{23}NO_2$	249,1724	5,32	Probable
Approche 2		$C_{12}H_{26}OS_2$	250,1410	5,91	Probable
Approche 2	Cimetidine	$C_{10}H_{16}N_6S$	252,1153	3,60	Probable
Approche 2		$C_9H_{18}N_6O_3$	258,1436	5,91	Probable
Approche 2	Atenolol	$C_{14}H_{22}N_2O_3$	266,1627	1,34	Probable
Approche 2		$C_9H_{20}N_{10}$	268,1860	5,47	Probable
Approche 2		$C_8H_{18}N_2O_8$	270,1073	5,54	Probable
Approche 2		$C_{12}H_{18}N_2O_5$	270,1211	3,79	Probable
Approche 2		$C_{15}H_{16}N_2O_3$	272,1155	5,83	Probable
Approche 2		$C_{16}H_{23}NO_3$	277,1672	4,27	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 2			279,1829	7,07	Probable
Approche 2		$C_{17}H_{19}N_3O$	281,1523	4,85	Probable
Approche 2		$C_9H_{15}N_8OP$	282,1108	5,82	Probable
Approche 2		$C_{16}H_{31}NO_3$	285,2299	6,96	Probable
Approche 2		$C_{11}H_{14}N_2O_5S$	286,0618	3,56	Probable
Approche 2		$C_{12}H_{23}N_4PS$	286,1387	6,64	Probable
Approche 2		$C_{16}H_{19}NO_4$	289,1308	5,26	Probable
Approche 2		$C_{18}H_{30}ClN$	295,2068	6,90	Probable
Approche 2		$C_{18}H_{37}NO_2$	299,2818	7,87	Probable
Approche 2			302,1260	6,96	Probable
Approche 2		$C_{16}H_{17}NO_5$	303,1101	6,98	Probable
Approche 2		$C_{15}H_{28}O_4S$	304,1699	7,40	Probable
Approche 2		$C_{20}H_{21}NO_2$	307,1565	7,05	Probable
Approche 2		$C_{17}H_{12}N_2O_4$	308,0789	7,92	Probable
Approche 2	Diacetalol Bestatin	$C_{16}H_{24}N_2O_4$	308,1736	4,43	Probable
Approche 2		$C_{16}H_{24}N_2O_4$	308,1731	6,47	Probable
Approche 2		$C_{18}H_{28}O_2S$	308,1827	6,94	Probable
Approche 2	Desmethylcitalopram	$C_{19}H_{19}FN_2O$	310,1481	6,78	Probable
Approche 2		$C_{12}H_{24}N_6O_4$	316,1492	6,30	Probable
Approche 2		$C_{16}H_{22}N_2O_5$	322,1523	6,14	Probable
Approche 2		$C_{15}H_{21}N_3O_3S$	323,1291	8,24	Probable
Approche 2		C11H21N11O	323,1940	6,62	Probable
Approche 2		$C_{10}H_{15}NO_7S_2$	325,0296	4,61	Probable
Approche 2		$C_{12}H_{22}N_6O_5$	330,1648	6,95	Probable
Approche 2	Disopyramide	$C_{21}H_{29}N_3O$	339,2312	5,46	Probable
Approche 2		$C_{17}H_{18}N_2O_6$	346,1155	5,65	Probable
Approche 2		$C_{19}H_{19}N_3O_4$	353,1366	5,31	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 2		$C_9H_{25}ClN_{10}O_4$	372,1743	8,21	Probable
Approche 2		$C_{17}H_{37}NO_8$	383,2510	7,63	Probable
Approche 2		$C_{16}H_{38}N_2O_2P_2S$	384,2113	7,62	Probable
Approche 2		$C_{12}H_{32}ClN_8PS$	386,1906	8,22	Probable
Approche 2		$C_{11}H_{31}ClN_{10}O_3$	386,2271	8,54	Probable
Approche 2		$C_9H_{29}N_{11}O_6$	387,2306	7,84	Probable
Approche 2		$C_{15}H_{31}N_11O_2$	397,2661	8,00	Probable
Approche 2		$C_{21}H_{44}N_5P$	397,3337	7,89	Probable
Approche 2		$C_{11}H_{33}N_{10}O_2PS$	400,2246	8,00	Probable
Approche 2		$C_{22}H_{29}N_3O_5$	415,2098	5,84	Probable
Approche 2		$C_{16}H_{36}N_7O_4P$	421,2560	5,99	Probable
Approche 2		$C_{14}H_{17}Cl_2NO_{10}$	429,0234	5,52	Probable
Approche 2		$C_{17}H_{21}N_{10}O_3P$	444,1545	7,68	Probable
Approche 2		$C_{17}H_{32}N_8S_3$	444,1911	7,53	Probable
Approche 2		$C_{21}H_{28}N_6O_5$	444,2111	5,30	Probable
Approche 2		$C_{19}H_{43}O_5P_3$	444,2321	7,97	Probable
Approche 2		$C_{17}H_{33}N_{10}O_3P$	456,2468	8,26	Probable
Approche 2		$C_{14}H_{36}N_9O_6P$	457,2531	8,27	Probable
Approche 2		$C_{14}H_{38}N_9O_4PS$	459,2505	8,26	Probable
Approche 2		$C_{21}H_{34}N_6O_3P_2$	480,2168	5,29	Probable
Approche 2		$C_{24}H_{50}NOP_3S$	493,2847	7,51	Probable
Approche 2		$C_{29}H_{34}O_7$	494,2326	4,87	Probable
Approche 2		$C_{12}H_{35}N_{15}O_7$	501,2866	7,89	Probable
Approche 2		$C_{23}H_{42}N_2O_6P_2$	504,2522	8,21	Probable
Approche 2		$C_{19}H_{35}N_4O_6P_3$	508,1757	5,35	Probable
Approche 2		$C_{22}H_{44}O_{13}$	516,2771	5,11	Probable
Approche 2		$C_{19}H_{36}N_{10}O_8$	532,2718	5,31	Probable
Approche 2		$C_{19}H_{27}N_{11}O_4P_2$	535,1725	6,94	Probable
Approche 2		$C_{23}H_{36}O_{10}S_2$	536,1755	6,96	Probable

Approche	Proposition de structure(s)	Formule Brute	Masse moléculaire (Da)	Tr (min)	Hiérarchisation
Approche 2		$C_{20}H_{38}N_6O_{11}$	538,2587	5,12	Probable
Approche 2		$C_{23}H_{54}N_5O_4P_3$	557,3547	7,74	Probable
Approche 2		$C_{29}H_{44}N_4O_6S$	576,2982	5,50	Probable
Approche 2		$C_{26}H_{44}N_6O_5P_2$	582,2850	5,34	Probable
Approche 2		$C_{17}H_{46}N_9O_9PS$	583,2882	5,34	Probable
Approche 2		$C_{23}H_{37}N_5O_{13}$	591,2363	5,40	Probable
Approche 2		$C_{20}H_{32}N_6O_{15}$	596,1917	5,40	Probable
Approche 2		$C_{21}H_{44}N_2O_{17}$	596,2641	5,37	Probable
Approche 2		$C_{23}H_{54}N_6O_{10}P_2$	636,3380	5,55	Probable
Approche 2			1348,7159	6,12	Probable

Références

Aceña, J, Stampachiacchiere, S, Pérez, S, Barceló, D, 2015. Advances in liquid chromatography–high resolution mass spectrometry for quantitative and qualitative environmental analysis. Anal Bioanal Chem 407, 6289–6299. https://doi.org/10.1007/s00216-015-8852-6

Ahrens, L, Felizeter, S, Sturm, R, Xie, Z, Ebinghaus, R, 2009. Polyfluorinated compounds in waste water treatment plant effluents and surface waters along the River Elbe, Germany. Marine Pollution Bulletin 58, 1326–1333. https://doi.org/10.1016/j.marpolbul.2009.04.028

Al Aukidy, M, Verlicchi, P, Jelic, A, Petrovic, M, Barcelò, D, 2012. Monitoring release of pharmaceutical compounds: Occurrence and environmental risk assessment of two WWTP effluents and their receiving bodies in the Po Valley, Italy. Science of The Total Environment 438, 15–25. https://doi.org/10.1016/j.scitotenv.2012.08.061

Albergamo, V, Schollée, J.E, Schymanski, E.L, Helmus, R, Timmer, H, Hollender, J, de Voogt, P, 2019. Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. Environ. Sci. Technol. 53, 7584–7594. https://doi.org/10.1021/acs.est.9b01750

Alder, A.C, van der Voet, J, 2015. Occurrence and point source characterization of perfluoroalkyl acids in sewage sludge. Chemosphere 129, 62–73. https://doi.org/10.1016/j.chemosphere.2014.07.045

Al-Odaini, N.A, Zakaria, M.P, Yaziz, M.I, Surif, S, 2010. Multi-residue analytical method for human pharmaceuticals and synthetic hormones in river water and sewage effluents by solid-phase extraction and liquid chromatography–tandem mass spectrometry. Journal of Chromatography A 1217, 6791–6806. https://doi.org/10.1016/j.chroma.2010.08.033

Arnold, W. A.; Longnecker, K.; Kroeger, K. D.; Kujawinski, E. B, 2014. Molecular Signature of Organic Nitrogen in Septic-Impacted Groundwater. Environ. Sci.: Processes Impacts, 16, 2400–2407. https://doi.org/10.1039/C4EM00289J.

Asimakopoulos, A.G, Ajibola, A, Kannan, K, Thomaidis, N.S, 2013. Occurrence and removal efficiencies of benzotriazoles and benzothiazoles in a wastewater treatment plant in Greece. Science of The Total Environment 452–453, 163–171. https://doi.org/10.1016/j.scitotenv.2013.02.041

Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2016. General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. Analytica Chimica Acta 935, 173–186. https://doi.org/10.1016/j.aca.2016.06.030

Baena-Nogueras, R.M, Traverso-Soto, J.M, Biel-Maeso, M, Villar-Navarro, E, Lara-Martín, P.A, 2018. Sources and trends of artificial sweeteners in coastal waters in the bay of Cadiz (NE Atlantic). Marine Pollution Bulletin 135, 607–616. https://doi.org/10.1016/j.marpolbul.2018.07.069

Baker, D.R., Kasprzyk-Hordern, B., 2011. Critical evaluation of methodology commonly used in sample collection, storage and preparation for the analysis of pharmaceuticals and illicit drugs in surface water and wastewater by solid phase extraction and liquid chromatography–mass spectrometry. Journal of Chromatography A 1218, 8036–8059. https://doi.org/10.1016/j.chroma.2011.09.012

Balmer, M.E, Buser, H.-R, Müller, M.D, Poiger, T, 2005. Occurrence of Some Organic UV Filters in Wastewater, in Surface Waters, and in Fish from Swiss Lakes. Environ. Sci. Technol. 39, 953–962. https://doi.org/10.1021/es040055r

Barret, M, Carrère, H, Latrille, E, Wisniewski, C, Patureau, D, 2010a. Micropollutant and Sludge Characterization for Modeling Sorption Equilibria. Environ. Sci. Technol. 44, 1100–1106. https://doi.org/10.1021/es902575d

Barret, M, Patureau, D, Latrille, E, Carrère, H, 2010b. A three-compartment model for micropollutants sorption in sludge: Methodological approach and insights. Water Research 44, 616–624. https://doi.org/10.1016/j.watres.2009.08.029

Beckers, L.-M., Brack, W., Dann, J.P., Krauss, M., Müller, E., Schulze, T., 2020. Unraveling longitudinal pollution patterns of organic micropollutants in a river by non-target screening and cluster analysis. Science of The Total Environment 727, 138388. https://doi.org/10.1016/j.scitotenv.2020.138388

Burant, A., Selbig, W., Furlong, E.T., Higgins, C.P., 2018. Trace organic contaminants in urban runoff: Associations with urban land-use. Environmental Pollution 242, 2068–2077. https://doi.org/10.1016/j.envpol.2018.06.066

Benner, J, Helbling, D.E, Kohler, H.-P.E, Wittebol, J, Kaiser, E, Prasse, C, Ternes, T.A, Albers, C.N, Aamand, J, Horemans, B, Springael, D, Walravens, E, Boon, N, 2013. Is biological treatment a viable alternative for micropollutant removal in drinking water treatment processes? Water Research 47, 5955–5976. https://doi.org/10.1016/j.watres.2013.07.015

Bergé, A, Cladière, M, Gasperi, J, Coursimault, A, Tassin, B, Moilleron, R, 2013. Meta-analysis of environmental contamination by phthalates. Environ Sci Pollut Res 20, 8057–8076. https://doi.org/10.1007/s11356-013-1982-5

Bergé, A, Cladière, M, Gasperi, J, Coursimault, A, Tassin, B, Moilleron, R, 2012. Meta-analysis of environmental contamination by alkylphenols. Environ Sci Pollut Res 19, 3798–3819. https://doi.org/10.1007/s11356-012-1094-7

Bijlsma, L, Emke, E, Hernández, F, de Voogt, P, 2012. Investigation of drugs of abuse and relevant metabolites in Dutch sewage water by liquid chromatography coupled to high resolution mass spectrometry. Chemosphere 89, 1399–1406. https://doi.org/10.1016/j.chemosphere.2012.05.110

Björklund, K., Cousins, A.P., Strömvall, A.-M., Malmqvist, P.-A., 2009. Phthalates and nonylphenols in urban runoff: Occurrence, distribution and area emission factors. Science of The Total Environment 407, 4665–4672. https://doi.org/10.1016/j.scitotenv.2009.04.040

Borowska, E, Bourgin, M, Hollender, J, Kienle, C, McArdell, C.S, von Gunten, U, 2016. Oxidation of cetirizine, fexofenadine and hydrochlorothiazide during ozonation: Kinetics and formation of transformation products. Water Research 94, 350–362. https://doi.org/10.1016/j.watres.2016.02.020

Bourgin, M., Beck, B., Boehler, M., Borowska, E., Fleiner, J., Salhi, E., Teichler, R., von Gunten, U., Siegrist, H., McArdell, C.S., 2018. Evaluation of a full-scale wastewater treatment plant upgraded with ozonation and biological post-treatments: Abatement of micropollutants, formation of transformation products and oxidation by-products. Water Research 129, 486–498. https://doi.org/10.1016/j.watres.2017.10.036

Brack, W, Dulio, V, Ågerstrand, M, Allan, I, Altenburger, R, Brinkmann, M, Bunke, D, Burgess, R.M, Cousins, I, Escher, B.I, Hernández, F.J, Hewitt, L.M, Hilscherová, K, Hollender, J, Hollert, H, Kase, R, Klauer, B, Lindim, C, Herráez, D.L, Miège, C, Munthe, J, O'Toole, S, Posthuma, L, Rüdel, H, Schäfer, R.B, Sengl, M, Smedes, F, van de Meent, D, van den Brink, P.J, van Gils, J, van Wezel, A.P, Vethaak, A.D, Vermeirssen, E, von der Ohe, P.C, Vrana, B, 2017. Towards the review of the European Union Water Framework Directive: Recommendations for more efficient assessment and management of chemical contamination in European surface water resources. Science of The Total Environment 576, 720–737. https://doi.org/10.1016/j.scitotenv.2016.10.104

Bressy, A., Gromaire, M.-C., Lorgeoux, C., Chebbo, G., 2011. Alkylphenols in atmospheric depositions and urban runoff. Water Science and Technology 63, 671–679. https://doi.org/10.2166/wst.2011.121

Buchberger, W.W, 2011. Current approaches to trace analysis of pharmaceuticals and personal care products in the environment. Journal of Chromatography A 1218, 603–618. https://doi.org/10.1016/j.chroma.2010.10.040

Burkhardt, M, Zuleeg, S, Vonbank, R, Bester, K, Carmeliet, J, Boller, M, Wangler, T, 2012. Leaching of Biocides from Façades under Natural Weather Conditions. Environ. Sci. Technol. 46, 5497–5503. https://doi.org/10.1021/es2040009

Campo, J, Masiá, A, Blasco, C, Picó, Y, 2013. Occurrence and removal efficiency of pesticides in sewage treatment plants of four Mediterranean River Basins. Journal of Hazardous Materials 263, 146–157. https://doi.org/10.1016/j.jhazmat.2013.09.061

Carpenter, C.M.G., Helbling, D.E., 2018. Widespread Micropollutant Monitoring in the Hudson River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources. Environ. Sci. Technol. 52, 6187–6196. https://doi.org/10.1021/acs.est.8b00945

Carpenter, C.M.G., Wong, L.Y.J., Johnson, C.A., Helbling, D.E., 2019. Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream. Environ. Sci. Technol. 53, 77–87. https://doi.org/10.1021/acs.est.8b05320

Castro, G., Rodríguez, I., Ramil, M., Cela, R., 2019. Selective determination of sartan drugs in environmental water samples by mixed-mode solid-phase extraction and liquid chromatography tandem mass spectrometry. Chemosphere 224, 562–571. https://doi.org/10.1016/j.chemosphere.2019.02.137

Chiaia-Hernandez, A.C, Schymanski, E.L, Kumar, P, Singer, H.P, Hollender, J, 2014. Suspect and nontarget screening approaches to identify organic contaminant records in lake sediments. Analytical and Bioanalytical Chemistry, 406, 7323–7335. https://doi.org/10.1007/s00216-014-8166-0

Clara, M., Strenn, B., Kreuzinger, N., 2004. Carbamazepine as a possible anthropogenic marker in the aquatic environment: investigations on the behaviour of Carbamazepine in wastewater treatment and during groundwater infiltration. Water Research 38, 947–954. https://doi.org/10.1016/j.watres.2003.10.058

Clara, M, Windhofer, G, Hartl, W, Braun, K, Simon, M, Gans, O, Scheffknecht, C, Chovanec, A, 2010. Occurrence of phthalates in surface runoff, untreated and treated wastewater and fate during wastewater treatment. Chemosphere 78, 1078–1084. https://doi.org/10.1016/j.chemosphere.2009.12.052

Clark K, Cousins IT, Mackay D, 2003. Assessment of critical exposure pathways. In Staples CA (ed), The Handbook of Environmental Chemistry, Vol, 3, Part Q: Phthalate Esters. New York, Springer, pp. 227-262.

Comtois-Marotte, S, Chappuis, T, Vo Duy, S, Gilbert, N, Lajeunesse, A, Taktek, S, Desrosiers, M, Veilleux, É, Sauvé, S, 2017. Analysis of emerging contaminants in water and solid samples using high resolution mass spectrometry with a Q Exactive orbital ion trap and estrogenic activity with YES-assay. Chemosphere 166, 400–411. https://doi.org/10.1016/j.chemosphere.2016.09.077

Coutu, S, Del Giudice, D, Rossi, L, Barry, D.A, 2012. Modeling of facade leaching in urban catchments: Water Resour. Res. 48. https://doi.org/10.1029/2012WR012359

D'Almeida, M., Sire, O., Lardjane, S., Duval, H., 2020. Development of a new approach using mathematical modeling to predict cocktail effects of micropollutants of diverse origins. Environmental Research 188, 109897. https://doi.org/10.1016/j.envres.2020.109897

Daughton, C.G., 2004. Non-regulated water contaminants: emerging research. Environmental Impact Assessment Review 24, 711–732. https://doi.org/10.1016/j.eiar.2004.06.003

Deeb, A.A., Schmidt, T.C., 2016. Tandem anion and cation exchange solid phase extraction for the enrichment of micropollutants and their transformation products from ozonation in a wastewater treatment plant. Analytical and Bioanalytical Chemistry 408, 4219–4232. https://doi.org/10.1007/s00216-016-9523-y

Deshayes, S, Eudes, V, Bigourie, M, Droguet, C, Moilleron, R, 2017. Alkylphenol and phthalate contamination of all sources of greywater from French households. Science of The Total Environment 599–600, 883–890. https://doi.org/10.1016/j.scitotenv.2017.05.038

Dubus, I.G., Hollis, J.M., Brown, C.D., 2000. Pesticides in rainfall in Europe. Environmental Pollution 110, 331–344. https://doi.org/10.1016/S0269-7491(99)00295-X

Duchène, P., 2005. Cent ans de procédés d'épuration des eaux résiduaires. Techniques Sciences et Méthodes, p. 177-187

Eide, I, Neverdal, G, Thorvaldsen, B, Arneberg, R, Grung, B, Kvalheim, O.M, 2004. Toxicological evaluation of complex mixtures: fingerprinting and multivariate analysis. Environmental Toxicology and Pharmacology 18, 127–133. https://doi.org/10.1016/j.etap.2004.01.011

Eide, I, Neverdal, G, Thorvaldsen, B, Shen, H, Grung, B, Kvalheim, O, 2001. Resolution of GC-MS Data of Complex PAC Mixtures and Regression Modeling of Mutagenicity by PLS. Environ. Sci. Technol. 35, 2314–2318. https://doi.org/10.1021/es000154e

Esperanza, M., Suidan, M.T., Marfil-Vega, R., Gonzalez, C., Sorial, G.A., McCauley, P., Brenner, R., 2007. Fate of sex hormones in two pilot-scale municipal wastewater treatment plants: Conventional treatment. Chemosphere 66, 1535–1544. https://doi.org/10.1016/j.chemosphere.2006.08.020

Evgenidou, E.N., Konstantinou, I.K., Lambropoulou, D.A., 2015. Occurrence and removal of transformation products of PPCPs and illicit drugs in wastewaters: A review. Science of The Total Environment 505, 905–926. https://doi.org/10.1016/j.scitotenv.2014.10.021

Farré, M. la, Pérez, S., Kantiani, L., Barceló, D., 2008. Fate and toxicity of emerging pollutants, their metabolites and transformation products in the aquatic environment. TrAC Trends in Analytical Chemistry 27, 991–1007. https://doi.org/10.1016/j.trac.2008.09.010

Fischer, K, Majewsky, M, 2014. Cometabolic degradation of organic wastewater micropollutants by activated sludge and sludge-inherent microorganisms. Applied Microbiology and Biotechnology 98, 6583–6597. https://doi.org/10.1007/s00253-014-5826-0

Gago-Ferrero, P, Schymanski, E.L, Bletsou, A.A, Aalizadeh, R, Hollender, J, Thomaidis, N.S, 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. Environ. Sci. Technol. 49, 12333–12341. https://doi.org/10.1021/acs.est.5b03454

Gosetti, F., Mazzucco, E., Gennaro, M.C., Marengo, E., 2016. Contaminants in water: non-target UHPLC/MS analysis. Environmental Chemistry Letters 14, 51–65. https://doi.org/10.1007/s10311-015-0527-1

Gros, M, Rodríguez-Mozaz, S, Barceló, D, 2013. Rapid analysis of multiclass antibiotic residues and some of their metabolites in hospital, urban wastewater and river water by ultra-high-performance liquid chromatography coupled to quadrupole-linear ion trap tandem mass spectrometry. Journal of Chromatography A 1292, 173–188. https://doi.org/10.1016/j.chroma.2012.12.072

Gross, J. H.; Roepstorff, P., 2014. Mass Spectrometry: A Textbook. Springer Berlin Heidelberg: p.76; 112. https://doi.org/10.1007/978-3-642-10711-5

Guerra, P, Kim, M, Shah, A, Alaee, M, Smyth, S.A, 2014. Occurrence and fate of antibiotic, analgesic/anti-inflammatory, and antifungal compounds in five wastewater treatment processes. Science of The Total Environment 473–474, 235–243. https://doi.org/10.1016/j.scitotenv.2013.12.008

Hansman, R. L.; Dittmar, T.; Herndl, G. J, 2015. Conservation of Dissolved Organic Matter Molecular Composition during Mixing of the Deep-Water Masses of the Northeast Atlantic Ocean. Marine Chemistry, 177, 288–297. https://doi.org/10.1016/j.marchem.2015.06.001.

Haunschmidt, M., Klampfl, C.W., Buchberger, W., Hertsens, R., 2010. Determination of organic UV filters in water by stir bar sorptive extraction and direct analysis in real-time mass spectrometry. Analytical and Bioanalytical Chemistry 397, 269–275. https://doi.org/10.1007/s00216-009-3438-9

Hart, K.M, Tremp, J, Molnar, E, Giger, W, 1993. The occurrence and the fate of organic pollutants in the atmosphere. Water Air Soil Pollut 68, 91–112. https://doi.org/10.1007/BF00479395

Helbling, D.E, Hollender, J, Kohler, H.-P.E, Singer, H, Fenner, K, 2010. High-Throughput Identification of Microbial Transformation Products of Organic Micropollutants. Environ. Sci. Technol. 44, 6621–6627. https://doi.org/10.1021/es100970m

Hernández, F, Ibáñez, M, Sancho, J.V, Pozo, Ó.J, 2004. Comparison of Different Mass Spectrometric Techniques Combined with Liquid Chromatography for Confirmation of Pesticides in Environmental Water Based on the Use of Identification Points. Anal. Chem. 76, 4349–4357. https://doi.org/10.1021/ac049768i

Hernández, F, Sancho, J.V, Ibáñez, M, Abad, E, Portolés, T, Mattioli, L, 2012. Current use of high-resolution mass spectrometry in the environmental sciences. Analytical and Bioanalytical Chemistry, 403, 1251–1264. https://doi.org/10.1007/s00216-012-5844-7

Hockaday, W. C.; Purcell, J. M.; Marshall, A. G.; Baldock, J. A.; Hatcher, P. G, 2009. Electrospray and Photoionization Mass Spectrometry for the Characterization of Organic Matter in Natural Waters: A Qualitative Assessment. Limnology and Oceanography: Methods, 7, 81–95. https://doi.org/10.4319/lom.2009.7.81.

Hogenboom, A. C.; Niessen, W. M. A.; Little, D.; Brinkman, U. A. Th, 1999. Accurate Mass Determinations for the Confirmation and Identification of Organic Microcontaminants in Surface Water Using On-Line Solid-Phase Extraction Liquid Chromatography Electrospray Orthogonal-Acceleration Time-of-Flight Mass Spectrometry. Rapid Communications in Mass Spectrometry, 13, 125–133. https://doi.org/10.1002/(SICI)1097-0231(19990130)13:2<125::AID-RCM473>3.0.CO;2-S.

Hogenboom, A. C.; Niessen, W. M. A.; Brinkman, U. A. Th, 2000. Characterization of Photodegradation Products of Alachlor in Water by On-Line Solid-Phase Extraction Liquid Chromatography Combined with Tandem Mass Spectrometry and Orthogonal-Acceleration Time-of-Flight Mass Spectrometry. Rapid Communications in Mass Spectrometry, 14, 1914–1924. https://doi.org/10.1002/1097-0231(20001030)14:20<1914::AID-RCM112>3.0.CO;2-R.

Hogenboom, A.C, van Leerdam, J.A, de Voogt, P, 2009. Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography–hybrid linear ion trap Orbitrap mass spectrometry. Journal of Chromatography A 1216, 510–519. https://doi.org/10.1016/j.chroma.2008.08.053

Hoizey, G., Goglin, A., Malinovsky, J.-M., Robinet, A., Binet, L., Kaltenbach, M.L., Millart, H., Lamiable, D., 2006. Specific and sensitive analysis of nefopam and its main metabolite desmethyl-nefopam in human plasma by liquid chromatography–ion trap tandem mass spectrometry. Journal of Pharmaceutical and Biomedical Analysis 42, 593–600. https://doi.org/10.1016/j.jpba.2006.05.024

Hollender, J, Zimmermann, S.G, Koepke, S, Krauss, M, McArdell, C.S, Ort, C, Singer, H, von Gunten, U, Siegrist, H, 2009. Elimination of Organic Micropollutants in a Municipal Wastewater Treatment Plant Upgraded with a Full-Scale Post-Ozonation Followed by Sand Filtration. Environ. Sci. Technol. 43, 7862–7869. https://doi.org/10.1021/es9014629

Hou, R., Xu, Y., Wang, Z., 2016. Review of OPFRs in animals and humans: Absorption, bioaccumulation, metabolism, and internal exposure research. Chemosphere 153, 78–90. https://doi.org/10.1016/j.chemosphere.2016.03.003

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., Graham Cooks, R., 2005. The Orbitrap: a new mass spectrometer. Journal of Mass Spectrometry 40, 430–443. https://doi.org/10.1002/jms.856

Hug, C, Sievers, M, Ottermanns, R, Hollert, H, Brack, W, Krauss, M, 2015. Linking mutagenic activity to micropollutant concentrations in wastewater samples by partial least square regression and subsequent identification of variables. Chemosphere 138, 176–182. https://doi.org/10.1016/j.chemosphere.2015.05.072

Hug, C, Ulrich, N, Schulze, T, Brack, W, Krauss, M, 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. Environmental Pollution 184, 25–32. https://doi.org/10.1016/j.envpol.2013.07.048

Ibáñez, M, Sancho, J.V, Pozo, Ó.J, Hernández, F, 2006. Use of liquid chromatography quadrupole time-of-flight mass spectrometry in the elucidation of transformation products and metabolites of pesticides. Diazinon as a case study. Anal Bioanal Chem 384, 448–457. https://doi.org/10.1007/s00216-005-0167-6

Ibáñez, M, Sancho, J.V, Pozo, Ó.J, Hernández, F, 2004. Use of Quadrupole Time-of-Flight Mass Spectrometry in Environmental Analysis: Elucidation of Transformation Products of Triazine Herbicides in Water after UV Exposure. Anal. Chem. 76, 1328–1335. https://doi.org/10.1021/ac035200i

Itzel, F., Jewell, K.S., Leonhardt, J., Gehrmann, L., Nielsen, U., Ternes, T.A., Schmidt, T.C., Tuerk, J., 2018. Comprehensive analysis of antagonistic endocrine activity during ozone treatment of hospital wastewater. Science of The Total Environment 624, 1443–1454. https://doi.org/10.1016/j.scitotenv.2017.12.181

Jardak, K.; Drogui, P.; Daghrir, R, 2016. Surfactants in Aquatic and Terrestrial Environment: Occurrence, Behavior, and Treatment Processes. Environmental Science and Pollution Research, 23, 3195–3216. https://doi.org/10.1007/s11356-015-5803-x.

Jernberg, J, Pellinen, J, Rantalainen, A.-L, 2013. Identification of organic xenobiotics in urban aquatic environments using time-of-flight mass spectrometry. Science of The Total Environment 450–451, 1–6. https://doi.org/10.1016/j.scitotenv.2013.02.006

Joss, A, Zabczynski, S, Göbel, A, Hoffmann, B, Löffler, D, McArdell, C.S, Ternes, T.A, Thomsen, A, Siegrist, H, 2006. Biological degradation of pharmaceuticals in municipal wastewater treatment: Proposing a classification scheme. Water Research 40, 1686–1696. https://doi.org/10.1016/j.watres.2006.02.014

Katsoyiannis, A., Samara, C., 2007. The fate of dissolved organic carbon (DOC) in the wastewater treatment process and its importance in the removal of wastewater contaminants. Environmental Science and Pollution Research - International 14, 284–292. https://doi.org/10.1065/espr2006.05.302

Kaserzon, S.L, Heffernan, A.L, Thompson, K, Mueller, J.F, Gomez Ramos, M.J, 2017. Rapid screening and identification of chemical hazards in surface and drinking water using high resolution mass spectrometry and a case-control filter. Chemosphere 182, 656–664. https://doi.org/10.1016/j.chemosphere.2017.05.071

Kasprzyk-Hordern, B, Dinsdale, R.M, Guwy, A.J, 2009. The removal of pharmaceuticals, personal care products, endocrine disruptors and illicit drugs during wastewater treatment and its impact on the quality of receiving waters. Water Research 43, 363–380. https://doi.org/10.1016/j.watres.2008.10.047

Katajamaa, M., Miettinen, J., Orešič, M., 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 22, 634–636. https://doi.org/10.1093/bioinformatics/btk039

Khalil Granier, L., Chevreuil, M., 1997. Behaviour and spatial and temporal variations of polychlorinated biphenyls and lindane in the urban atmosphere of the Paris area, France. Atmospheric Environment 31, 3787–3802. https://doi.org/10.1016/S1352-2310(97)00210-0

Kern, S, Fenner, K, Singer, H.P, Schwarzenbach, R.P, Hollender, J, 2009. Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. Environ. Sci. Technol. 43, 7039–7046. https://doi.org/10.1021/es901979h

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536. https://doi.org/10.1093/bioinformatics/btn323

Kidd, K.A., Blanchfield, P.J., Mills, K.H., Palace, V.P., Evans, R.E., Lazorchak, J.M., Flick, R.W., 2007. Collapse of a fish population after exposure to a synthetic estrogen. Proc Natl Acad Sci USA 104, 8897. https://doi.org/10.1073/pnas.0609568104

Kim, S.; Kramer, R. W.; Hatcher, P. G, 2003. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. Anal. Chem, 75, 5336–5344. https://doi.org/10.1021/ac034415p.

Kiss, A.; Bergé, A.; Domenjoud, B.; Gonzalez-Ospina, A.; Vulliet, E, 2018. Chemometric and High-Resolution Mass Spectrometry Tools for the Characterization and Comparison of Raw and Treated Wastewater Samples of a Pilot Plant on the SIPIBEL Site. Environmental Science and Pollution Research, 25, 9230–9242. https://doi.org/10.1007/s11356-017-0748-x.

Kolkman, A, Martijn, B.J, Vughs, D, Baken, K.A, van Wezel, A.P, 2015. Tracing Nitrogenous Disinfection Byproducts after Medium Pressure UV Water Treatment by Stable Isotope Labeling and High Resolution Mass Spectrometry. Environ. Sci. Technol. 49, 4458–4465. https://doi.org/10.1021/es506063h

Krauss, M, Singer, H, Hollender, J, 2010. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. Analytical and Bioanalytical Chemistry 397, 943–951. https://doi.org/10.1007/s00216-010-3608-9

Kuroda, K., Murakami, M., Oguma, K., Muramatsu, Y., Takada, H., Takizawa, S., 2012. Assessment of Groundwater Pollution in Tokyo Using PPCPs as Sewage Markers. Environ. Sci. Technol. 46, 1455–1464. https://doi.org/10.1021/es202059g

Laetz Cathy A., Baldwin David H., Collier Tracy K., Hebert Vincent, Stark John D., Scholz Nathaniel L., 2009. The Synergistic Toxicity of Pesticide Mixtures: Implications for Risk Assessment and the Conservation of Endangered Pacific Salmon. Environmental Health Perspectives 117, 348–353. https://doi.org/10.1289/ehp.0800096

Lange, F. T.; Scheurer, M.; Brauch, H.-J, 2012. Artificial Sweeteners—a Recently Recognized Class of Emerging Environmental Contaminants: A Review. Analytical and Bioanalytical Chemistry, 403, 2503–2518. https://doi.org/10.1007/s00216-012-5892-z.

Lara-Martín, P.A., González-Mazo, E., Brownawell, B.J., 2011. Multi-residue method for the analysis of synthetic surfactants and their degradation metabolites in aquatic systems by liquid chromatography–time-of-flight-mass spectrometry. Journal of Chromatography A 1218, 4799–4807. https://doi.org/10.1016/j.chroma.2011.02.031

Lee, S., Lee, J., Kim, M., 2012. The influence of storm-water sewer overflows on stream water quality and source tracking of fecal contamination. KSCE Journal of Civil Engineering 16, 39–44. https://doi.org/10.1007/s12205-012-1198-0

Li, H, Yi, X, Cheng, F, Tong, Y, Mehler, W.T, You, J, 2019. Identifying Organic Toxicants in Sediment Using Effect-Directed Analysis: A Combination of Bioaccessibility-Based Extraction and High-Throughput Midge Toxicity Testing. Environ. Sci. Technol. 53, 996–1003. https://doi.org/10.1021/acs.est.8b05633

Li, Z., Kaserzon, S.L., Plassmann, M.M., Sobek, A., Gómez Ramos, M.J., Radke, M., 2017. A strategic screening approach to identify transformation products of organic micropollutants formed in natural waters. Environ. Sci.: Processes Impacts 19, 488–498. https://doi.org/10.1039/C6EM00635C

Loos, R, Carvalho, R, António, D.C, Comero, S, Locoro, G, Tavazzi, S, Paracchini, B, Ghiani, M, Lettieri, T, Blaha, L, Jarosova, B, Voorspoels, S, Servaes, K, Haglund, P, Fick, J, Lindberg, R.H, Schwesig, D, Gawlik, B.M, 2013. EU-wide monitoring survey on emerging polar organic contaminants in wastewater treatment plant effluents. Water Research 47, 6475–6487. https://doi.org/10.1016/j.watres.2013.08.024

Lozano, N, Rice, C.P, Ramirez, M, Torrents, A, 2013. Fate of Triclocarban, Triclosan and Methyltriclosan during wastewater and biosolids treatment processes. Water Research 47, 4519–4527. https://doi.org/10.1016/j.watres.2013.05.015

Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., Horning, S., 2006. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. Anal. Chem. 78, 2113–2120. https://doi.org/10.1021/ac0518811

Makarov, A, Scigelova, M, 2010. Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A 1217, 3938–3945. https://doi.org/10.1016/j.chroma.2010.02.022

Margot, J, Rossi, L, Barry, D.A, Holliger, C, 2015. A review of the fate of micropollutants in wastewater treatment plants: Micropollutants in wastewater treatment plants. WIREs Water 2, 457–487. https://doi.org/10.1002/wat2.1090

Maizel, A. C.; Remucal, C. K, 2017. The Effect of Advanced Secondary Municipal Wastewater Treatment on the Molecular Composition of Dissolved Organic Matter. Water Research, 122, 42–52. https://doi.org/10.1016/j.watres.2017.05.055.

Merel, S, Lege, S, Yanez Heras, J.E, Zwiener, C, 2017. Assessment of *N* -Oxide Formation during Wastewater Ozonation. Environ. Sci. Technol. 51, 410–417. https://doi.org/10.1021/acs.est.6b02373

Michael, I., Vasquez, M., Hapeshi, E., Haddad, T., Baginska, E., Kümmerer, K., Fatta-Kassinos, D., 2014. Metabolites and transformation products of pharmaceuticals in the aquatic environment as contaminants of emerging concern 413–458.

Moschet, C, Piazzoli, A, Singer, H, Hollender, J, 2013. Alleviating the Reference Standard Dilemma Using a Systematic Exact Mass Suspect Screening Approach with Liquid Chromatography-High Resolution Mass Spectrometry. Anal. Chem. 85, 10312–10320. https://doi.org/10.1021/ac4021598

Müller, A, Schulz, W, Ruck, W.K.L, Weber, W.H, 2011. A new approach to data evaluation in the non-target screening of organic trace substances in water analysis. Chemosphere 85, 1211–1219. https://doi.org/10.1016/j.chemosphere.2011.07.009

Müller, A., Österlund, H., Marsalek, J., Viklander, M., 2020. The pollution conveyed by urban runoff: A review of sources. Science of The Total Environment 709, 136125. https://doi.org/10.1016/j.scitotenv.2019.136125

Muz, M, Ost, N, Kühne, R, Schüürmann, G, Brack, W, Krauss, M, 2017. Nontargeted detection and identification of (aromatic) amines in environmental samples based on diagnostic derivatization and LC-high resolution mass spectrometry. Chemosphere 166, 300–310. https://doi.org/10.1016/j.chemosphere.2016.09.138

Nakada, N., Kiri, K., Shinohara, H., Harada, A., Kuroda, K., Takizawa, S., Takada, H., 2008. Evaluation of Pharmaceuticals and Personal Care Products as Water-soluble Molecular Markers of Sewage. Environ. Sci. Technol. 42, 6347–6353. https://doi.org/10.1021/es7030856

Negreira, N., Regueiro, J., López de Alda, M., Barceló, D., 2015. Degradation of the anticancer drug erlotinib during water chlorination: Non-targeted approach for the identification of transformation products. Water Research 85, 103–113. https://doi.org/10.1016/j.watres.2015.08.005

Nürenberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on liquid chromatography – high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. Journal of Chromatography A 1426, 77–90. https://doi.org/10.1016/j.chroma.2015.11.014

O'Brien, J.W, Thai, P.K, Brandsma, S.H, Leonards, P.E.G, Ort, C, Mueller, J.F, 2015. Wastewater analysis of Census day samples to investigate per capita input of organophosphorus flame retardants and plasticizers into wastewater. Chemosphere 138, 328–334. https://doi.org/10.1016/j.chemosphere.2015.06.014

Pailler, J.-Y., Krein, A., Pfister, L., Hoffmann, L., Guignard, C., 2009. Solid phase extraction coupled to liquid chromatography-tandem mass spectrometry analysis of sulfonamides, tetracyclines, analgesics and hormones in surface water and wastewater in Luxembourg. Science of The Total Environment 407, 4736–4743. https://doi.org/10.1016/j.scitotenv.2009.04.042

Palmer, P.M., Wilson, L.R., O'Keefe, P., Sheridan, R., King, T., Chen, C.-Y., 2008. Sources of pharmaceutical pollution in the New York City Watershed. Science of The Total Environment 394, 90–102. https://doi.org/10.1016/j.scitotenv.2008.01.011

Parry, E, Young, T.M, 2016. Comparing targeted and non-targeted high-resolution mass spectrometric approaches for assessing advanced oxidation reactor performance. Water Research 104, 72–81. https://doi.org/10.1016/j.watres.2016.07.056

Paulus, G.K, Hornstra, L.M, Alygizakis, N, Slobodnik, J, Thomaidis, N, Medema, G, 2019. The impact of on-site hospital wastewater treatment on the downstream communal wastewater system in terms of antibiotics and antibiotic resistance genes. International Journal of Hygiene and Environmental Health 222, 635–644. https://doi.org/10.1016/j.ijheh.2019.01.004

Pérez-Fernández, V, Mainero Rocca, L, Tomai, P, Fanali, S, Gentili, A, 2017. Recent advancements and future trends in environmental analysis: Sample preparation, liquid chromatography and mass spectrometry. Analytica Chimica Acta 983, 9–41. https://doi.org/10.1016/j.aca.2017.06.029

Petrie, B, Youdan, J, Barden, R, Kasprzyk-Hordern, B, 2016. Multi-residue analysis of 90 emerging contaminants in liquid and solid environmental matrices by ultra-high-performance liquid chromatography tandem mass spectrometry. Journal of Chromatography A 1431, 64–78. https://doi.org/10.1016/j.chroma.2015.12.036

Pinhancos, R., Maass, S., Ramanathan, D.M., 2011. High-resolution mass spectrometry method for the detection, characterization and quantitation of pharmaceuticals in water. Journal of Mass Spectrometry 46, 1175–1181. https://doi.org/10.1002/jms.2005

Phillips, P., Chalmers, A., 2009. Wastewater Effluent, Combined Sewer Overflows, and Other Sources of Organic Compounds to Lake Champlain1. JAWRA Journal of the American Water Resources Association 45, 45–57. https://doi.org/10.1111/j.1752-1688.2008.00288.x

Phungsai, P., Kurisu, F., Kasuga, I., Furumai, H., 2016. Molecular characterization of low molecular weight dissolved organic matter in water reclamation processes using Orbitrap mass spectrometry. Water Research 100, 526–536. https://doi.org/10.1016/j.watres.2016.05.047

Plaza-Bolaños, P., Valera-Tarifa, N.M., Frenich, A.G., 2013. Chapter 18 - Current Applications of GC-(Q)TOF and GC–HRMS for the Determination of Persistent Organic Pollutants in Water and Sediments Samples, in: Ferrer, I., Thurman, E.M. (Eds.), Comprehensive Analytical Chemistry. Elsevier, pp. 431–454. https://doi.org/10.1016/B978-0-444-62623-3.00018-6

Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11, 395. https://doi.org/10.1186/1471-2105-11-395

Pomiès, M, Choubert, J.-M, Wisniewski, C, Coquery, M, 2013. Modelling of micropollutant removal in biological wastewater treatments: A review. Science of The Total Environment 443, 733–748. https://doi.org/10.1016/j.scitotenv.2012.11.037

Rapport Académie nationale de Pharmacie. Médicaments et Environnement. 2008, 105p.

Rapport ANSES. Campagne nationale d'occurrence des résidus de médicaments dans les eaux destinées à la consommation humaine. Ressources en eaux brutes et eaux traitées. Convention G2799. 18/03/2011.

Rapport SIPIBEL 2011-2015 : résultats de quatre années de suivi, d'études et de recherches, sur le site pilote de Bellecombe, octobre 2016, 174p. http://www.graie.org/Sipibel/publications/sipibel-rapport-effluentshospitaliersmedicaments-oct16.pdf

Reemtsma, T.; Miehe, U.; Duennbier, U.; Jekel, M, 2010. Polar Pollutants in Municipal Wastewater and the Water Cycle: Occurrence and Removal of Benzotriazoles. Water Research, 44, 596–604. https://doi.org/10.1016/j.watres.2009.07.016.

Reemtsma, T.; Quintana, J. B.; Rodil, R.; Garcı´a-López, M.; Rodrı´guez, I, 2008. Organophosphorus Flame Retardants and Plasticizers in Water and Air I. Occurrence and Fate. TrAC Trends in Analytical Chemistry, 27, 727–737. https://doi.org/10.1016/j.trac.2008.07.002.

Ridder, L., Wagener, M., 2008. SyGMa: Combining Expert Knowledge and Empirical Scoring in the Prediction of Metabolites. ChemMedChem 3, 821–832. https://doi.org/10.1002/cmdc.200700312

Ruff, M, Mueller, M.S, Loos, M, Singer, H.P, 2015. Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry – Identification of unknown sources and compounds. Water Research 87, 145–154. https://doi.org/10.1016/j.watres.2015.09.017

Russell, D.H, Edmondson, R.D, 1997. High-resolution Mass Spectrometry and Accurate Mass Measurements with Emphasis on the Characterization of Peptides and Proteins by Matrix-assisted Laser Desorption/Ionization Time-of-Ñight Mass Spectrometry. Journal of Mass Spectrometry, 32, 14.

Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. Journal of Cheminformatics 8, 3. https://doi.org/10.1186/s13321-016-0115-9

Sancho, J. V.; Pozo, Ó. J.; Ibáñez, M.; Hernández, F, 2006. Potential of Liquid Chromatography/Time-of-Flight Mass Spectrometry for the Determination of Pesticides and Transformation Products in Water. Analytical and Bioanalytical Chemistry, 386, 987. https://doi.org/10.1007/s00216-006-0532-0.

Santos, F.J., Galceran, M.T., Caixach, J., Huguet, X., Rivera, J., 1996. Characterization of Polychlorinated Terphenyl Mixtures by High Resolution Gas Chromatography and Tandem Mass Spectrometry. Rapid Communications in Mass Spectrometry 10, 1774–1780. https://doi.org/10.1002/(SICI)1097-0231(199611)10:14<1774::AID-RCM705>3.0.CO;2-I

Santos, F.J., Galceran, M.T., Caixach, J., Rivera, J., Huguet, X., 1997. Characterization of Toxaphene by High Resolution Gas Chromatography Combined with High Resolution Mass Spectrometry and Tandem Mass Spectrometry. Rapid Communications in Mass Spectrometry 11, 341–348. https://doi.org/10.1002/(SICI)1097-0231(19970228)11:4<341::AID-RCM832>3.0.CO;2-N

Santos, F. J.; Galceran, M. T, 2003. Modern Developments in Gas Chromatography–Mass Spectrometry-Based Environmental Analysis. Journal of Chromatography A, 1000, 125–151. https://doi.org/10.1016/S0021-9673(03)00305-4.

Santos, L.H.M.L.M, Gros, M, Rodriguez-Mozaz, S, Delerue-Matos, C, Pena, A, Barceló, D, Montenegro, M.C.B.S.M, 2013. Contribution of hospital effluents to the load of pharmaceuticals in urban wastewaters: Identification of ecologically relevant pharmaceuticals. Science of The Total Environment 461–462, 302–316. https://doi.org/10.1016/j.scitotenv.2013.04.077

Schollée, J.E, Bourgin, M, von Gunten, U, McArdell, C.S, Hollender, J, 2018. Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. Water Research 142, 267–278. https://doi.org/10.1016/j.watres.2018.05.045

Schollée, J.E, Schymanski, E.L, Avak, S.E, Loos, M, Hollender, J, 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. Anal. Chem. 87, 12121–12129. https://doi.org/10.1021/acs.analchem.5b02905

Schulz, W., Lucke, T., Achten, C., Oberleitner, D., Balsaa, P., Hinnenkamp, V., Brüggen, S., Duennbier, U., Liebmann, D., Fink, A., Götz, S., Geiß, S., Hohrenk, L., Härtel, C., Letzel, T., Liesener, A., Reineke, A., Logemann, J., Petri, M., Zwiener, C., 2019. Use of non-target screening by means of LC-ESI-HRMS in water analysis (Guideline - edition 1.0 2019).

Schulze, B., Jeon, Y., Kaserzon, S., Heffernan, A.L., Dewapriya, P., O'Brien, J., Gomez Ramos, M.J., Ghorbani Gorji, S., Mueller, J.F., Thomas, K.V., Samanipour, S., 2020. An assessment of quality assurance/quality control efforts in high resolution mass spectrometry non-target workflows for analysis of environmental samples. TrAC Trends in Analytical Chemistry 133, 116063. https://doi.org/10.1016/j.trac.2020.116063

Schymanski, E.L, Hollender, J, Singer, H.P, Ferguson, P.L, 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? Environ. Sci. Technol. 51, 11505–11512. https://doi.org/10.1021/acs.est.7b02184

Schymanski, E.L, Singer, H.P, Longrée, P, Loos, M, Ruff, M, Stravs, M.A, Ripollés Vidal, C, Hollender, J, 2014. Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. Environ. Sci. Technol. 48, 1811–1818. https://doi.org/10.1021/es4044374

Schymanski, E.L, Singer, H.P, Slobodnik, J, Ipolyi, I.M, Oswald, P, Krauss, M, Schulze, T, Haglund, P, Letzel, T, Grosse, S, Thomaidis, N.S, Bletsou, A, Zwiener, C, Ibáñez, M, Portolés, T, de Boer, R, Reid, M.J, Onghena, M, Kunkel, U, Schulz, W, Guillon, A, Noyon, N, Leroy, G, Bados, P, Bogialli, S, Stipaničev, D, Rostkowski, P, Hollender, J, 2015. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Analytical and Bioanalytical Chemistry, 407, 6237–6255. https://doi.org/10.1007/s00216-015-8681-7

Segura, P.A, MacLeod, S.L, Lemoine, P, Sauvé, S, Gagnon, C, 2011. Quantification of carbamazepine and atrazine and screening of suspect organic contaminants in surface and drinking waters. Chemosphere 84, 1085–1094. https://doi.org/10.1016/j.chemosphere.2011.04.056

Sinclair, E., Kannan, K., 2006. Mass Loading and Fate of Perfluoroalkyl Surfactants in Wastewater Treatment Plants. Environ. Sci. Technol. 40, 1408–1414. https://doi.org/10.1021/es051798v

Singer, H, Jaus, S, Hanke, I, Lück, A, Hollender, J, Alder, A.C, 2010. Determination of biocides and pesticides by on-line solid phase extraction coupled with mass spectrometry and their behaviour in wastewater and surface water. Environmental Pollution 158, 3054–3064. https://doi.org/10.1016/j.envpol.2010.06.013

Sjerps, R.M.A, Vughs, D, van Leerdam, J.A, ter Laak, T.L, van Wezel, A.P, 2016. Data-driven prioritization of chemicals for various water types using suspect screening LC-HRMS. Water Research 93, 254–264. https://doi.org/10.1016/j.watres.2016.02.034

Stasinakis, A.S, Thomaidis, N.S, Arvaniti, O.S, Asimakopoulos, A.G, Samaras, V.G, Ajibola, A, Mamais, D, Lekkas, T.D, 2013. Contribution of primary and secondary treatment on the removal of benzothiazoles, benzotriazoles, endocrine disruptors, pharmaceuticals and perfluorinated compounds in a sewage treatment plant. Science of The Total Environment 463–464, 1067–1075. https://doi.org/10.1016/j.scitotenv.2013.06.087

Stephan, S, Hippler, J, Köhler, T, Deeb, A.A, Schmidt, T.C, Schmitz, O.J, 2016. Contaminant screening of wastewater with HPLC-IM-qTOF-MS and LC+LC-IM-qTOF-MS using a CCS database. Anal Bioanal Chem 408, 6545–6555. https://doi.org/10.1007/s00216-016-9820-5

Schwarzenbach, R.; Escher, B.; Fenner, K.; Hofstetter, T.; Johnson, C.; Gunten, U.; Wehrli, B, 2006. The Challenge of Micropollutants in Aquatic Systems. Science (New York, N.Y.), 313, 1072–1077. https://doi.org/10.1126/science.1127291.

Tasselli, S., Guzzella, L., 2020. Polycyclic musk fragrances (PMFs) in wastewater and activated sludge: analytical protocol and application to a real case study. Environmental Science and Pollution Research 27, 30977–30986. https://doi.org/10.1007/s11356-019-06767-7

Teil, M.J., Blanchard, M., Chevreuil, M., 2006. Atmospheric fate of phthalate esters in an urban area (Paris-France). Science of The Total Environment 354, 212–223. https://doi.org/10.1016/j.scitotenv.2004.12.083

Thompson, J.M., Gündoğdu, A., Stratton, H.M., Katouli, M., 2013. Antibiotic resistant Staphylococcus aureus in hospital wastewaters and sewage treatment plants with special reference to methicillin-resistant Staphylococcus aureus (MRSA). Journal of Applied Microbiology 114, 44–54. https://doi.org/10.1111/jam.12037

Thurman, E.M, Ferrer, I, Blotevogel, J, Borch, T, 2014. Analysis of Hydraulic Fracturing Flowback and Produced Waters Using Accurate Mass: Identification of Ethoxylated Surfactants. Anal. Chem. 86, 9653–9661. https://doi.org/10.1021/ac502163k

Tran, N.H., Reinhard, M., Khan, E., Chen, H., Nguyen, V.T., Li, Y., Goh, S.G., Nguyen, Q.B., Saeidi, N., Gin, K.Y.-H., 2019. Emerging contaminants in wastewater, stormwater runoff, and surface water: Application as chemical markers for diffuse sources. Science of The Total Environment 676, 252–267. https://doi.org/10.1016/j.scitotenv.2019.04.160

Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P, 2015. PLS/OPLS Models in Metabolomics: The Impact of Permutation of Dataset Rows on the K-Fold Cross-Validation Quality Parameters. Mol. BioSyst., 11, 13–19. https://doi.org/10.1039/C4MB00414K.

Van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 7, 142. https://doi.org/10.1186/1471-2164-7-142

Van der Veen, I.; de Boer, J, 2012. Phosphorus Flame Retardants: Properties, Production, Environmental Occurrence, Toxicity and Analysis. Chemosphere, 88, 1119–1153. https://doi.org/10.1016/j.chemosphere.2012.03.067

Van Krevelen, D. W, 1950. Studies of Gas Absorption. VI. A Graphical Representation for the Efficiency of Physical Absorption. Recueil des Travaux Chimiques des Pays-Bas, 69, 503–508. https://doi.org/10.1002/recl.19500690416.

Verkh, Y., Rozman, M., Petrovic, M., 2018. A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. Chemosphere 200, 397–404. https://doi.org/10.1016/j.chemosphere.2018.02.095

Verlicchi, P, Al Aukidy, M, Zambello, E, 2012. Occurrence of pharmaceutical compounds in urban wastewater: Removal, mass load and environmental risk after a secondary treatment—A review. Science of The Total Environment 429, 123–155. https://doi.org/10.1016/j.scitotenv.2012.04.028

Weiss, J.M, Simon, E, Stroomberg, G.J, de Boer, R, de Boer, J, van der Linden, S.C, Leonards, P.E.G, Lamoree, M.H, 2011. Identification strategy for unknown pollutants using high-resolution mass spectrometry: Androgen-disrupting compounds identified through effect-directed analysis. Anal Bioanal Chem 400, 3141–3149. https://doi.org/10.1007/s00216-011-4939-x

Wille, K, Claessens, M, Rappé, K, Monteyne, E, Janssen, C.R, De Brabander, H.F, Vanhaecke, L, 2011. Rapid quantification of pharmaceuticals and pesticides in passive samplers using ultra high performance liquid chromatography coupled to high resolution mass spectrometry. Journal of Chromatography A 1218, 9162–9173. https://doi.org/10.1016/j.chroma.2011.10.039

Wode, F, van Baar, P, Dünnbier, U, Hecht, F, Taute, T, Jekel, M, Reemtsma, T, 2015. Search for over 2000 current and legacy micropollutants on a wastewater infiltration site with a UPLC-high resolution MS target screening method. Water Research 69, 274–283. https://doi.org/10.1016/j.watres.2014.11.034

Xing, Y., Yu, Y., Men, Y., 2018. Emerging investigators series: occurrence and fate of emerging organic contaminants in wastewater treatment plants with an enhanced nitrification step. Environ. Sci.: Water Res. Technol. 4, 1412–1426. https://doi.org/10.1039/C8EW00278A

Xu, W., Wang, X., Cai, Z., 2013. Analytical chemistry of the persistent organic pollutants identified in the Stockholm Convention: A review. Analytica Chimica Acta 790, 1–13. https://doi.org/10.1016/j.aca.2013.04.026

Yang, G., Ma, L., Xu, D., Liu, L., Jia, H., Chen, Y., Zhang, Y., Chai, Z., 2012. Temporal variations of organochlorine pesticides in precipitation in Beijing, China. Atmospheric Environment 61, 614–619. https://doi.org/10.1016/j.atmosenv.2012.08.026

Zonja, B, Delgado, A, Pérez, S, Barceló, D, 2015. LC-HRMS Suspect Screening for Detection-Based Prioritization of Iodinated Contrast Media Photodegradates in Surface Waters. Environ. Sci. Technol. 49, 3464–3472. https://doi.org/10.1021/es505250q